Assessment of ECMWF SEAS5 Seasonal Forecast Performance over South America

S. Gubler,[a] K. Sedlmeier,[a] J. Bhend,[a] G. Avalos,[b] C. A. S. Coelho,[c] Y. Escajadillo,[b]
M. Jacques-Coper,[d] R. Martinez,[e] C. Schwierz,[a] M. de Skansi,[f] and Ch. Spirig[a]

[a] *Federal Office of Meteorology and Climatology, MeteoSwiss, Zürich, Switzerland*
[b] *Servicio Nacional de Meteorología e Hidrología del Perú, Lima, Peru*
[c] *Centro de Previsão de Tempo e Estudos Climáticos, Instituto Nacional de Pesquisas Espaciais, Cachoeira Paulista, Brazil*
[d] *Departamento de Geofísica and (CR)2, Universidad de Concepción, Concepción, Chile*
[e] *Centro Internacional para la Investigación del Fenómeno de El Niño, Guayaquil, Ecuador*
[f] *Servicio Meteorológico Nacional, Buenos Aires, Argentina*

ABSTRACT

Seasonal predictions have a great socioeconomic potential if they are reliable and skillful. In this study, we assess the prediction performance of SEAS5, version 5 of the seasonal prediction system of the European Centre for Medium-Range Weather Forecasts (ECMWF), over South America against homogenized station data. For temperature, we find the highest prediction performances in the tropics during austral summer, where the probability that the predictions correctly discriminate different observed outcomes is 70%. In regions lying to the east of the Andes, the predictions of maximum and minimum temperature still exhibit considerable performance, while farther to the south in Chile and Argentina the temperature prediction performance is low. Generally, the prediction performance of minimum temperature is slightly lower than for maximum temperature. The prediction performance of precipitation is generally lower and spatially and temporally more variable than for temperature. The highest prediction performance is observed at the coast and over the highlands of Colombia and Ecuador, over the northeastern part of Brazil, and over an isolated region to the north of Uruguay during DJF. In general, Niño-3.4 has a strong influence on both air temperature and precipitation in the regions where ECMWF SEAS5 shows high performance, in some regions through teleconnections (e.g., to the north of Uruguay). However, we show that SEAS5 outperforms a simple empirical prediction based on Niño-3.4 in most regions where the prediction performance of the dynamical model is high, thereby supporting the potential benefit of using a dynamical model instead of statistical relationships for predictions at the seasonal scale.

1. Introduction

Seasonal climate forecasts are increasingly sought after to support decisions in a variety of sectors. Their potential has been demonstrated with applications in agriculture (WMO 2007; Hansen 2002) through the prediction of droughts (NIDIS 2004; Schubert et al. 2007; Pozzi et al. 2013; Shafiee-Jood et al. 2014; Yuan and Wood 2013) and crop yield modeling (Cantelaube and Terres 2005; Challinor et al. 2005; Ceglar et al. 2018), in the energy sector (Troccoli 2010; Brayshaw et al. 2011; De Felice et al. 2015; Clark et al. 2017; Svensson et al. 2015), in the insurance sector through weather derivatives (Jewson and Brix 2005), in the transport sector (Palin et al. 2016; Karpechko et al. 2015), in seasonal hurricane prediction (Emanuel et al. 2012), and in the health sector through malaria predictions (Morse et al. 2005). Seasonal predictions are based on statistical tools, for example based on linear regression, or on dynamical models such as ECMWF Systems 4 and 5 (Molteni et al. 2011; Johnson et al. 2019), NCEP CFS (Saha et al. 2006), and many others [see Troccoli (2010) for an overview], or a combination of both (Coelho et al. 2006a).

The quality of monthly and seasonal predictions has been regionally assessed within a variety of studies (e.g., Palmer 2002; Palmer et al. 2004; Saha et al. 2006; Wang et al. 2009; Alessandri et al. 2011; Lee et al. 2010, 2011;

Kim et al. 2012). Within the European Provision of Regional Impacts Assessments on Seasonal and Decadal Time Scales (EUPORIAS) project for example, verification metrics of the European Centre for Medium-Range Weather Forecasts (ECMWF) forecasts Systems 4 and 5 (Molteni et al. 2011; Johnson et al. 2019) against ERA-Interim (Dee et al. 2011) were globally calculated on a 2° regular grid and published online (Wehrli et al. 2017). However, the interpretations of the verification analyses often focused on regions in the Northern Hemisphere. For South America, two studies by Coelho et al. (2006a,b) based on both empirical and multimodel approaches showed that the best seasonal precipitation prediction quality during austral summer is found in the tropics and the region around southern Brazil, Uruguay, Paraguay, and northern Argentina. A recent study by Osman and Vera (2017) on both temperature and precipitation confirms the highest performance in the tropics during DJF for temperature, and states that precipitation prediction ability over South America is similar in DJF and JJA.

In South America, national weather services (NWSs) typically use the IRI Climate Predictability Tool (CPT) (Mason and Baddour 2008; Mason and Tippett 2016) to issue seasonal forecasts. CPT is a statistical downscaling tool and can be used for statistical prediction using differing predictor fields. In practice, the predictor field is determined individually for each region, forecast variable, and time period of interest, often a tedious and subjective task. Further, many weather services qualitatively modify the CPT output based on a consensus discussion before publicly issuing the seasonal forecasts. The subjectivity in the procedure and the fact that the forecasts are not stored makes retrospective verification of these predictions difficult. Dynamical models are currently much less used in operational seasonal forecasting in South America. However, there is a tendency for dynamical model forecasts to be used more extensively in the future by the NWSs in South America, particularly because they have a variety of advantages over statistical models: the physical consistency between different variables is assured to the extent represented by the model, predictions are issued globally meaning that no abrupt changes occur at country borders, and there is no need to define predictors individually. Since the ECMWF seasonal forecasting system has proven to be among the best models predicting ENSO (Barnston et al. 2012) and is openly available now through the Copernicus Climate Change Service,[1] there is considerable interest of the NWSs in South America in this system, and consequently in its verification.

El Niño–Southern Oscillation (ENSO) is one of the principal modes influencing climate at the seasonal scale in South America (Grimm et al. 2000; Pezzi and Cavalcanti 2001; Zhou and Lau 2001; Brönnimann 2007; Troccoli 2010; Shimizu et al. 2017; Sulca et al. 2018), and the predictability of ENSO and its teleconnection has been identified as the main source of predictability at the seasonal scale in that region (e.g., Manzanas et al. 2014). It is widely known that ENSO has a large influence on extreme events in the Andean region. For instance, a strong El Niño event may result in economic losses due to increased rainfall, landslides and floods, mainly in the lowlands of Ecuador and Peru, and rainfall deficits in Colombia and the highlands of Peru (e.g., Vicente-Serrano et al. 2017; Erfanian et al. 2017). Given ENSO's large impacts in South America and considering the ability of seasonal forecasts to predict it at least to some extent, there is obvious potential in socioeconomic benefits with seasonal prediction in this region. Within the second phase of Climandes (Rosas et al. 2016), a project aiming at developing climate services for the agricultural sector, we investigated the potential benefit of seasonal forecasts for smallholder farmers. To ensure that forecasts are trusted and applied by farmers, the performance of the forecast and the forecast's relevance for decision making at the local scale are crucial (Ash et al. 2007). For instance, Ziervogel et al. (2005) state that seasonal climate forecasts may only benefit smallholder farmers if they are correct in more than 60%–70% of the cases, otherwise they do more harm than good. Furthermore, the local focus implies that forecast performance should be ideally evaluated against representative ground observations rather than quasi-observations such as reanalyses that have traditionally been used to analyze seasonal forecast quality. The use of reanalyses as ground truth in South America is further complicated as they show deficiencies in representing the local climate conditions for both temperature (Hofer et al. 2012) and precipitation (Imfeld et al. 2019) in regions of complex terrain.

For these reasons, this study verifies SEAS5, the latest seasonal prediction system of the ECMWF, against high-quality homogenized in situ precipitation and air temperature observations over South America. Furthermore, we extend the existing analyses on seasonal forecast performance cited above by comparing forecast quality for all seasons and regions in South America. The performance of SEAS5 is then interpreted with an analysis of the relationship of the ENSO index of region 3.4 (Niño-3.4) with air temperature and precipitation observations. Further, SEAS5 prediction performance is compared to a simple statistical model based on Niño-3.4 to study whether high prediction

---

[1] https://climate.copernicus.eu/.

performance observed for SEAS5 is purely due to ENSO as the main source of predictability.

2. Data

The ECMWF seasonal prediction system SEAS5, released in 2017, is the operational seasonal forecasting system of ECMWF at time of publication (Johnson et al. 2019). We verify the hindcasts (or reforecasts) of ECMWF SEAS5 that are available for the period of 1981–2018. Like the operational seasonal forecast, hindcasts are initialized on the first day of every month, but in contrast to the operational forecast ensemble with 51 members, the hindcasts consist of 25 members. SEAS5 has a spatial resolution of $\sim$35 km^2 and forecasts run out to 7 months. This study focuses on forecasts with a 1-month lead time, referring to a forecast for December–February that is issued in November, for example.

The observations used as ground truth for verification consist of more than 200 meteorological stations measuring daily precipitation (Prec) as well as maximum and minimum temperature (Tmax and Tmin, respectively). The stations cover the whole region of South America from Colombia and Brazil to southern Chile and Argentina (Fig. 1), with the exception of Venezuela, Guiana, Suriname, and French Guiana. In a joint effort described in Skansi et al. (2013), all measurements were quality controlled applying the RClimDex software (Zhang and Yang 2004) and an additional quality control software developed by Aguilar et al. (2010). The observations were further homogenized using the RHtestV3 (Wang et al. 2010) and the RSNHT softwares (Alexandersson and Moberg 1997; Aguilar 2010), however without gap filling. The homogenized observations are available from 1965 to 2010. Note that homogenization and quality control are important prerequisites for verification since erroneous measurements or artificial break points may alter the outcomes of a verification procedure. In this study, all data are averaged on a monthly/seasonal basis.

The relationship between ENSO (e.g., Trenberth 1997) and the meteorological variables is analyzed based on the seasonal El Niño index[2] (Barnston and Ropelewski 1992) as a 3-month running mean of ERSST.v5 sea surface temperature (SST) anomalies in the Niño-3.4 region (5°N–5°S, 120°–170°W) issued by NOAA (Huang et al. 2017), referred to as Niño-3.4 hereafter.

[2] Downloaded from http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml.
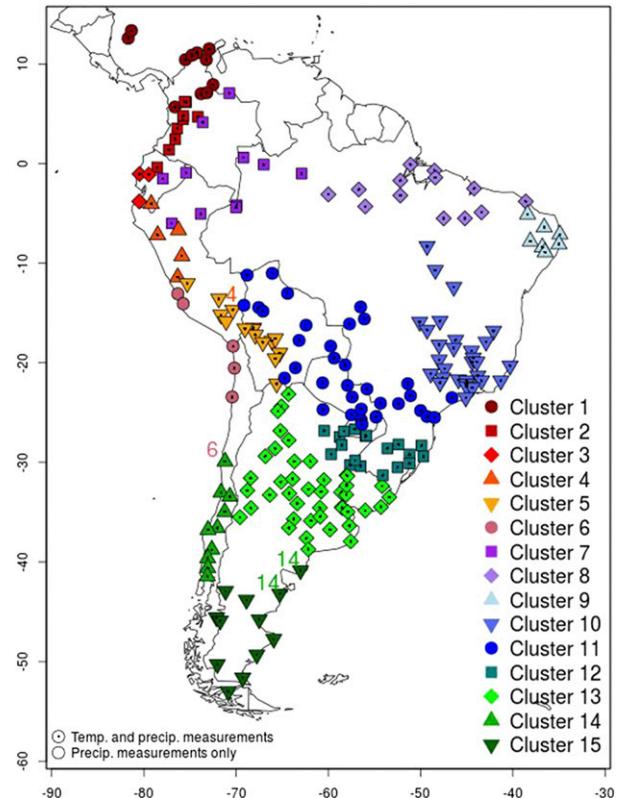
FIG. 1. Station locations and clusters, indicated through distinct colors. Empty points indicate stations for which only precipitation is measured, and points with a dot indicate stations with both temperature and precipitation measurements. The four stations that were manually moved to another cluster are marked with the number of their original cluster, at the upper left of the station.

3. Methods

a. Clustering—Climatology of South America

As a first step, the observation sites are clustered based on the correlations of the standardized monthly anomalies of Tmin, Tmax, and precipitation. A hierarchical distance clustering of the Fisher-transformed and averaged correlation matrices of the three variables was performed using the "agnes" algorithm (Kaufman and Rousseeuw 1990; Struyf et al. 1996, 1997; Lance and Williams 1967) of the "cluster" R-package. An analysis of differing cluster sizes (2–20) showed that a relatively large number of clusters is needed to represent the differing climate zones in South America. As a compromise between representing all climate zones and avoiding having a cluster number that is too large for analysis, a total of 15 clusters was selected. From these 15 objective clusters, four stations were manually moved to another cluster for obvious geographic considerations such as horizontal distance and altitude (Fig. 1). All subsequent analyses are performed on the

resulting clusters (referred to by an uppercase letter "C" followed by the cluster number), which are briefly described here in terms of their climatology. The climatologies depicted in Fig. 2 are averages of the measurements of all stations lying within a cluster.

Four distinct rainfall classes can be distinguished in South America (Fig. 2). Many regions show a unimodal rainfall distribution peaking either in austral summer (C3–C5 and C10–C12), in austral spring (C8 and C9), or in austral winter (C14 and C15). A clearly bimodal rainfall pattern is observed in the Colombian and Ecuadorian Andes and lowlands (C1, C2, and C7) peaking during the transition seasons. No rainfall at all is observed in the desert region at the northern Chilean/southern Peruvian coast (C6). The highest rainfall amounts occur in the Amazonian regions in Brazil, while the lowest amounts occur in the desert region mentioned above, in Argentina, and in the Altiplano region in Bolivia and Peru.

In the northern tropical regions (C1–C3, C6, and C7), temperature is almost constant throughout the year, and the daily temperature range (i.e., the difference between day and night temperatures) reaches around 5°–10°C on average. Farther toward the south, the annual cycle becomes apparent and increasingly more pronounced, with a mean amplitude (i.e., mean differences between winter and summer temperature) of up to 15°C in the south.

b. Verification

The verification was carried out for 3-monthly means of the forecasts with 1-month lead time (i.e., forecast months 2–4). It focuses on ensemble averages and on categorical forecasts in the form of tercile probabilities. This means that both the forecasts and verifying observations are categorized into three climatologically equiprobable classes, the boundaries of which are derived from the hindcast climatology and observation climatology, respectively. This forecast format implicitly contains a calibration step and verification results are much less affected by systematic forecast errors (e.g., time-mean biases and variance errors). In the following, the lower (upper) terciles will be referred to as dry (wet) for precipitation and cold (warm) for temperature for the sake of simplicity (even though, for example, "dry" does not necessarily imply dry in an absolute sense, but drier-than-normal conditions). The closest grid point of the SEAS5 hindcasts was bias-corrected at each station by quantile mapping using "qmap" (Gudmundsson et al. 2012) implemented within the "biascorrection" R-package (Bhend 2017). The verification was performed for the overlapping time period of the two datasets (i.e., 1981–2010).

It is widely recognized that verification should be done based on a variety of measures to assess the different aspects of forecast quality (Jolliffe and Stephenson 2012). Therefore, diverse prediction quality metrics measuring the association, the accuracy, the discrimination, and the reliability of the forecasts were applied here (see Murphy 1993 for an overview, Table 1). These metrics assess the following aspects of the forecasts (described in the same order of appearance as in the listing above): the linear relationship between the variation of the forecast ensemble mean and the variation of the observations (measured by the Pearson correlation coefficient); the level of agreement between forecasts and observation (measured by the root-mean-squared error of the forecast ensemble mean and the observation, and by the ranked probability skill score of the tercile forecasts and observations); the ability of the forecast to distinguish outcomes for which the observations differ (measured by the discrimination score and the area under the receiver operating curve); and the correspondence of the forecasted probability and the observed frequency of an event (measured by the Weisheimer reliability scores).

All metrics were calculated using R (R Core Team 2012) applying the "easyVerification" (MeteoSwiss 2016) and SpecsVerification (Siegert et al. 2017) R-packages. Except for the Weisheimer score, all metrics were calculated individually at each station and then summarized for each cluster using the median. The Weisheimer reliability score—referred to as Weisheimer score in the following—is based on the slope of the line of the reliability diagram and its associated uncertainty (e.g., Weisheimer and Palmer 2014). Note that a slope of one refers to a perfect correspondence of the forecasted probabilities and the observed frequencies. To ensure a large sample size, the Weisheimer score was determined individually for each cluster by calculating the score based on an "artificial" time series consisting of the pooled hindcasts and observations of all stations lying within a cluster. The score ranges from "dangerous to use" (score 1) to "almost perfect" (score 5).

To give an overview on the verification results, we focus on Pearson correlations and the Weisheimer score. Other scores are included in the discussion of the clusters in case they show significant differences to these main scores and provide additional insights for interpretation. The results for all scores can be found in the appendix.

c. Comparison with ENSO

The statistical relationship between ENSO and the meteorological variables is analyzed based on the Pearson correlation of the seasonal Niño-3.4 index
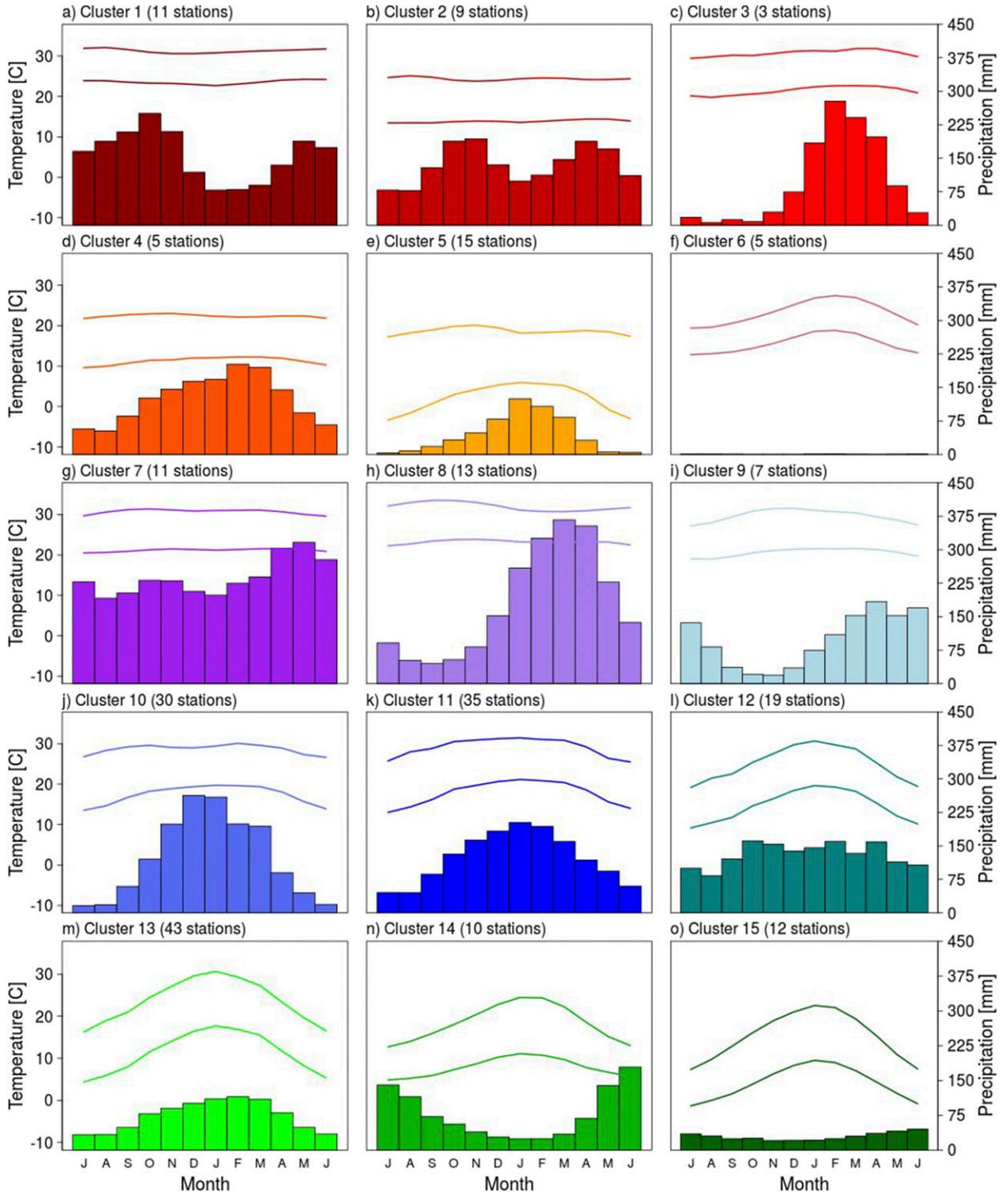
FIG. 2. Climatology (monthly means) of maximum and minimum temperature (lines), as well as precipitation (bars) for the different station clusters (see Fig. 1). The number in the parentheses indicates the number of stations on which the climatologies are based.

TABLE 1. Overview of the verification metrics assessed in this study.

Score	References	Abbreviation	What it measures	Based on	Limits	Optimal value
Correlation	Pearson (1895)	CORR	Association of the model ensemble mean and the observations	Ensemble mean	$[-1; 1]$	1
Root-mean-squared error skill score	Hyndman and Koehler (2006)	RMSESS	Accuracy of the model ensemble mean	Ensemble mean	$[-\infty; 1]$	1
Rank probability skill score	Weigel et al. (2007)	RPSS	Accuracy of tercile predictions	Tercile predictions	$[-\infty; 1]$	1
Area under the receiver operating curve skill score	Mason and Graham (1999)	ROCSS	Discrimination	Tercile predictions	$[-1; 1]$	1
Generalized discrimination score	Weigel and Mason (2011)	DISCR	Discrimination	Tercile predictions	$[0; 1]$	1
Weisheimer categories	Weisheimer and Palmer (2014)	Weisheimer score	Reliability	Tercile predictions	$[1; 5]$	5

and the observations. The correlations are first determined for different lags from 0 to 11 months. In the analyses, a special focus is put on a lag-1 comparison, referring for example to the Niño-3.4 index in NDJ correlated against the observations of DJF (where January and February refer to the following year). This focus was made since the relationship between ENSO and the observations are strongest for the shortest lag over the whole continent, and also because the NWSs in South America using CPT often apply an ENSO index at lag 1 for their seasonal forecasts.

Furthermore, the prediction skill of SEAS5 was compared to this lag-1 Niño-3.4 benchmark. The reader should note that this benchmark is not applicable as a prediction in practice since two-thirds of the predicted months will already have passed when using a seasonal index at lag-1 for prediction. Here, we use it as a theoretical reference and express both SEAS5 and Niño-3.4 lag-1 coefficients of determination with respect to the observations as a fraction. Fraction values above one indicate a higher performance of SEAS5 while values below one stand for higher performance of a simple linear model using Niño-3.4 as predictor. It is widely known that the different types of ENSO (i.e., different regions of anomalous sea surface temperature) influence the atmosphere in different ways and may have opposite effects depending on where the sea surface temperature is maximized (Waylen and Poveda 2002; Takahashi et al. 2011; Penalba and Rivera 2016; Tedeschi et al. 2015, 2016; Garreaud 2018). However, studying the influence of ENSO in regions other than Niño-3.4 goes beyond the scope of this publication.

d. Usefulness of seasonal forecasts for small-scale applications

One main goal of the Climandes project was to determine the usefulness of SEAS5 for small-scale applications for agriculture in South America. Ziervogel et al. (2005) state that forecasts need to be correct at least 60%–70% of the time in order to be of use for smallholder farmers. The generalized discrimination score, measuring the percentage of correct forecasts (Weigel and Mason 2011), is a suitable measure of the forecasts' usefulness. To enable analyses based on the Pearson correlation as done in this analysis, the percentages determined by Ziervogel et al. (2005), which can be related to the discrimination score, were empirically converted to thresholds regarding the Pearson correlation. Therefore, a linear model on the correlation and the discrimination score was established based on all the data points available in this study. By applying a Fisher transformation (i.e., the inverse hyperbolic tangent function) to the correlation and the discrimination

score (after first being transformed to fall into the interval $[-1; 1]$), it is assured that the two verification metrics are approximately normally distributed. The fitted linear model resulted in the following relationship between the correlation and the discrimination score (DISCR):

$$CORR = tanh[0.032 + 1.496 \times tanh^{-1}(2 \times DISCR - 1)].$$

From the formula, it follows that the case of forecasts being correct at least 60% of the time (i.e., a discrimination of 0.6) corresponds to correlations of at least 0.32, while forecasts that are correct at least 70% relate to correlations lying above 0.58. Therefore, we chose 0.3 as the correlation threshold below which forecasts are potentially harmful and 0.6 as the correlation threshold above which forecasts are useful even for small-scale applications. Note that the uncertainty with regard to these thresholds is quite large, and that the values 0.032 and 1.496 are estimated empirically and hold for the data underlying this study. Whether the parameters would strongly differ in another region of the world cannot be determined here.

4. Results and discussion

a. Forecast quality of ECMWF SEAS5 and the influence of ENSO

Given the strong influence of ENSO on the climate of South America and its important role in determining predictability at the seasonal scale (see references in the introduction), a high prediction performance is to be expected in South America, with potential even for small-scale applications. Detailed analyses however show that the prediction performance at the seasonal scale is limited to specific regions and seasons of the year, and highly dependent on the variable of interest. In general, the performance of seasonal temperature predictions is higher than that of precipitation predictions (Fig. 3), while differences between minimum and maximum temperature predictions are less pronounced. This is not surprising since precipitation predictions generally exhibit far less skill than temperature predictions due to the intermittent nature of precipitation and due to precipitation formation being strongly influenced by local processes. In the following sections, the spatiotemporal differences of temperature and precipitation prediction performances are assessed in more detail and put into relation with the Niño-3.4 index.

1) NORTHERN ANDES AND PACIFIC COAST

At the stations lying close to the Pacific Ocean and in the Andes north of 5°S (C1–C3), high prediction performance of SEAS5 is found, with correlations of the temperature hindcasts ranging on average above the usefulness threshold of 0.6 determined in section 3d (Figs. 3a,c) and high reliability classes ranging between 4 and 5 (Figs. 4b,c,e,f). This is also the region with the strongest relationship between Niño-3.4 and air temperatures (Fig. 5). As they are strongly influenced by the sea surface temperature, air temperatures in this region rise (decrease) during El Niño (La Niña). This correlation is strongest from austral winter to fall, with correlations between Niño-3.4 and observations ranging between 0.4 and 0.8 for Tmax and between 0.3 and 0.6 for Tmin (Figs. 5a,b). In the tropical regions of the Andes (as well as in the Amazon described below), the relationship of Niño-3.4 with precipitation is of type warm (positive ENSO index)—dry/cold (negative ENSO index)—wet, indicated by the negative correlations of the precipitation observations with this ENSO index (Fig. 5c), while at the Ecuadorian coast, the influence of ENSO is of opposite sign. These findings corroborate previous studies by Vuille et al. (2000), Poveda et al. (2011), Waylen and Poveda (2002), Córdoba-Machado et al. (2015a,b), Recalde-Coronel et al. (2014), and Sulca et al. (2018). In accordance with Vuille et al. (2000), for instance, the strongest negative correlations of Niño-3.4 and precipitation in the Colombian/Ecuadorian highlands occur during December–March (Fig. 5c) as well as in June–September (Fig. A1). In contrast, El Niño is related to heavy precipitation events due to strong convection in austral summer along the coast of northern Peru and Ecuador (Aceituno 1988; Takahashi 2004; Lagos et al. 2008; Bazo et al. 2013; among others). This increase in excessive precipitation is due to enhanced water vapor availability and convection because of anomalously high SSTs (e.g., Lavado-Casimiro and Espinoza 2014) and is confined to a narrow band at the Ecuadorian coast (Vuille et al. 2000). Similarly as for temperature, the prediction performance for precipitation is relatively high in these tropical regions with correlations above 0.45 (C1 and C2) and even above 0.8 (C3) (Fig. 3e), and Weisheimer scores ranging between 4 and 5 for clusters 1 and 2 (Figs. 4h,i). However, note that cluster number 3 at the Ecuadorian coast exhibits a Weisheimer score of only 3.

At the example of this cluster, it can be illustrated that analyzing several performance measures is required to fully describe forecast quality. Despite the strong association of rainfall forecasts and corresponding observations at the Ecuadorian coast (C3), accuracy and discrimination of tercile category forecasts are clearly lower than for the other clusters close by (e.g., Fig. A1). A detailed analysis showed that the strong El Niño years with heavy precipitation events (1982/83 and 1997/98) are well captured by the ECMWF SEAS5
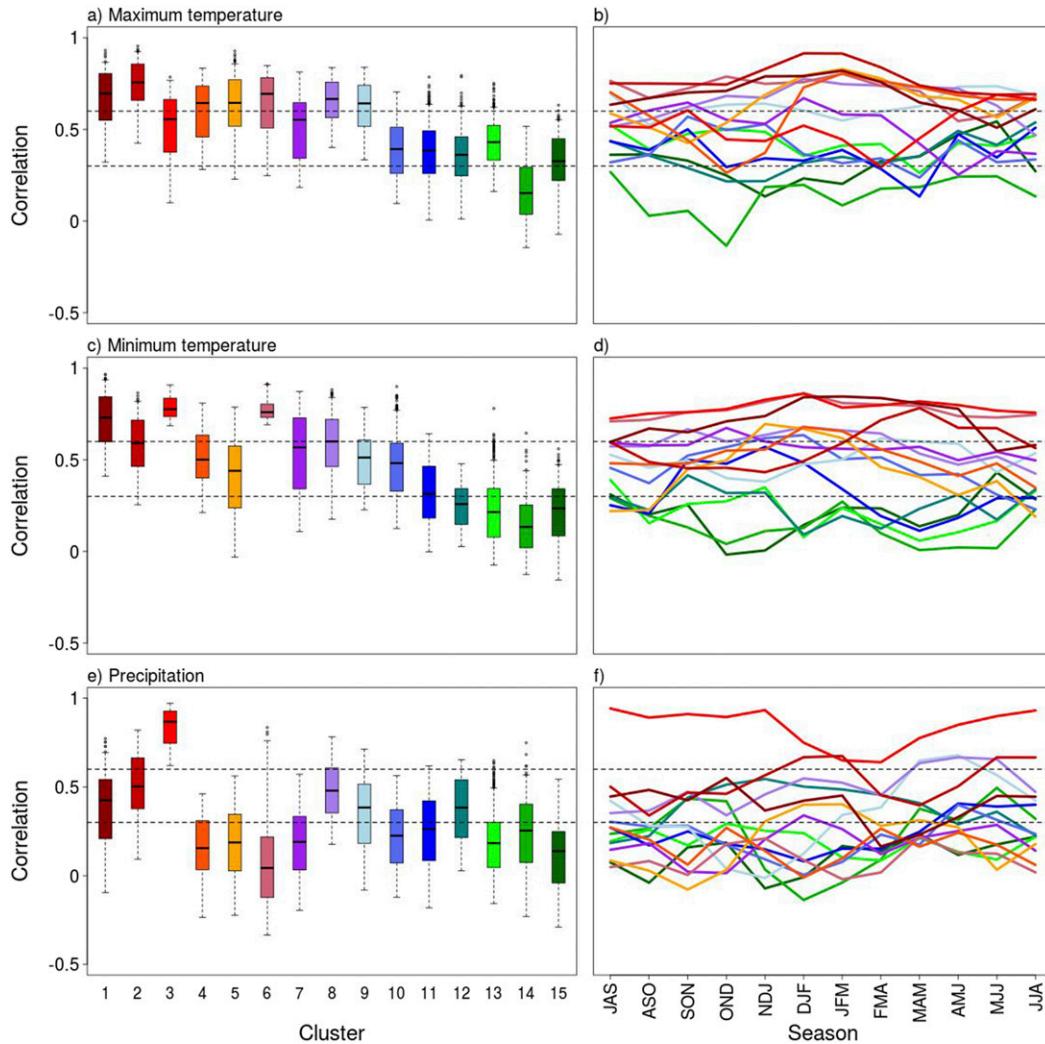
FIG. 3. (left) Correlations of hindcasts with observations as boxplots over all stations and seasons in a cluster and (right) as the median over all stations per cluster for different seasons. The black horizontal lines indicate the thresholds below or above which the forecasts are considered as potentially harmful (0.3) or useful (0.6), respectively, for small-scale applications.

resulting in high correlations and a small RMSE compared to the climatological forecast. The low RPSS and discrimination as well as the Weisheimer score of 3 (Figs. 4h,i), however, indicate that the year-to-year variability is not well captured in years when El Niño is weaker or absent. This reflects the general challenge of issuing seasonal predictions during neutral ENSO phases, or, in this region, during La Niña.

2) CENTRAL ANDES

The prediction performance of temperature farther south in the Andes [i.e., in the central Andes (C4 and C5)], is still high especially during austral summer (Figs. 3a,c and A2). This is seen both in correlations ranging above 0.6 for Tmax and between 0.3 and 0.6 for

Tmin on average, and the Weisheimer scores that range between 4 and 5 for both Tmin and Tmax in DJF (Figs. 4b,c,e,f). In contrast however, the negative rainfall anomalies associated with positive ENSO phases in the Central Andes during austral summer reported previously (Vuille 1999; Silva et al. 2008; Lagos et al. 2008; Lavado-Casimiro et al. 2012; Lavado-Casimiro and Espinoza 2014) and confirmed by this study (correlation values around −0.3, Fig. 5) do not lead to a prediction performance exceeding climatological information for precipitation, indicated by the correlations lying below the usefulness threshold of 0.3 (Fig. 3c) and the generally low reliability scores (Fig. 4h). An exception is found with regard to a reliability score of 4 for dry episodes in the Peruvian and Bolivian Altiplano.
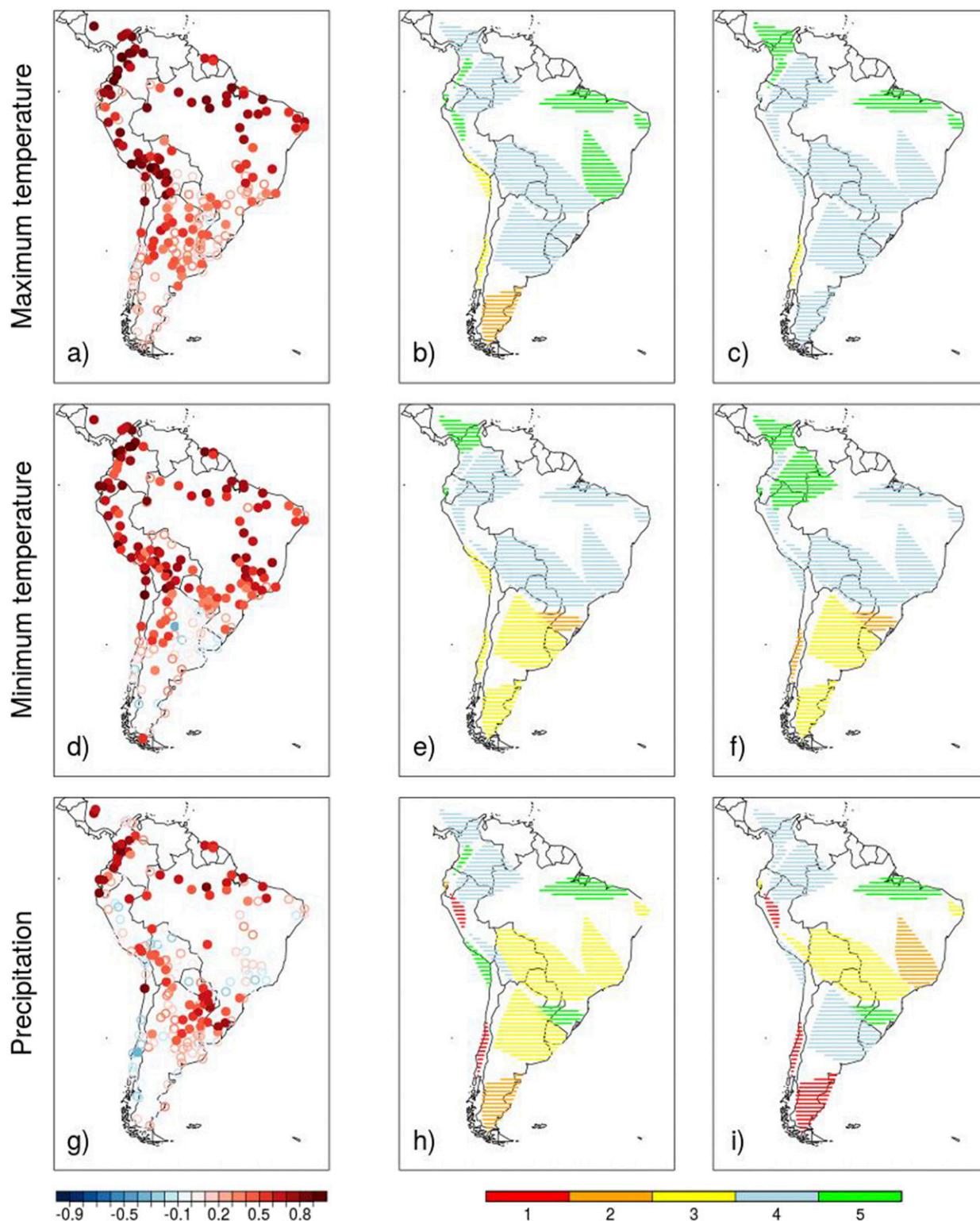
FIG. 4. (left) Correlation and (center),(right) Weisheimer reliability maps calculated between hindcasts and station observations of (a)–(c) maximum temperature, (d)–(f) minimum temperature, and (g)–(i) precipitation for DJF. Correlations are shown from blue (negative correlations) to red (positive correlations). Filled circles indicate correlations that are significant at the 0.05 significance level. The Weisheimer reliability categories are given from 1 in red (dangerous forecast) to 5 in green (perfect forecast) for the cold/dry category in (b), (e), and (h) and the warm/wet category in (c), (f), and (i) (see color legend).
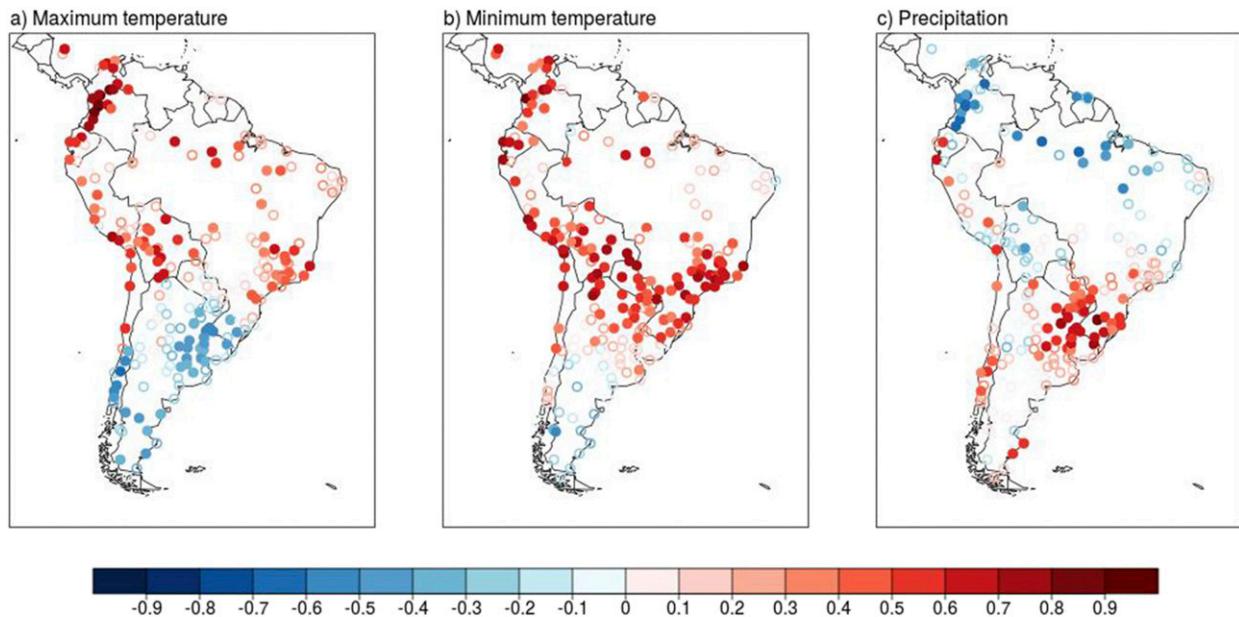
FIG. 5. Correlation maps between the seasonal Niño-3.4 index at lag-1 and the station observations in DJF for (a) maximum temperature, (b) minimum temperature, and (c) precipitation. Filled circles indicate correlations that are significant at the 0.05 significance level.

The low correlations of SEAS5 with precipitation observations may be due to different reasons, for instance the independence of the onset of the rainy season from ENSO (Silva et al. 2008) or the relation of dry spells in the Peruvian Andes with wet anomalies in northeastern Brazil through the Bolivian high (Sulca et al. 2016). Furthermore, other processes such as the upper-tropospheric zonal wind anomalies influence precipitation in the central Andes (Imfeld et al. 2019). Investigating whether these phenomena are well represented in SEAS5 would require further examinations going beyond the scope of this study.

As a side effect, the verification analysis pointed us to single remaining quality problems in the observations. For example, the temperature forecast skill at one station in the Altiplano region differed strongly in comparison to other stations in that region. A closer look at the temperature record revealed obvious issues of the temperature measurements (e.g., Hunziker et al. 2017), making it necessary for the station to be excluded from the analysis. It is known that model evaluation can be used to assess the quality of observational data (see Massonnet et al. (2016) for a more comprehensive approach and in-depth discussion).

3) NORTHERN AND CENTRAL AMAZON

Toward the east of the northern Andes [i.e., in the western Amazon basin (C7)], temperature prediction performance is still relatively high with correlation

values clearly above 0.3 despite a weaker relationship with Niño-3.4 than in the Andes (Figs. 3a–c, 5a,b, and A3). The high prediction quality of SEAS5 in this region indicates that modes other than ENSO influence the predictability of temperature at the seasonal scale, for instance, North Atlantic SST (Marengo et al. 2008; Coelho et al. 2012; Panisset et al. 2018), which are presumably well represented by SEAS5. In contrast, precipitation performance is very low throughout the year in that specific region. Although it has been shown that the SST of the Atlantic influences precipitation in the Amazon region (Yoon and Zeng 2010), specifically a north–south tropical Atlantic SST dipole-like structure (Vuille et al. 2000; Ronchail et al. 2002), this study indicates no prediction performance for precipitation in the region, with correlation values below 0.3 (Figs. 3e and A3).

In contrast, the prediction performance of SEAS5 in northern Brazil (C8) is quite high for many 3-month periods with correlation values ranging clearly above 0.3 for precipitation and even up to 0.6 for temperature. The highest correlations for temperature occur during DJF (Figs. 3b,d and A3) and from MAM to MJJ for precipitation (Fig. 3e). The influence of ENSO on both temperatures and precipitation in the region is relatively high (Fig. 5) (e.g., Uvo et al. 1998 and Coelho et al. 2002, 2006a). Aceituno (1988) showed that drier-than-normal conditions prevail during negative phases of the Southern Oscillation in northeastern Brazil in late austral summer,

which might be a reason for the relatively high prediction performance for precipitation in comparison to its neighboring regions.

In eastern Brazil (C9), temperature shows only a weak positive correlation with Niño-3.4 around MAM (Fig. A3). Nevertheless, the temperature prediction scores of SEAS5 are positive and quite high throughout the whole year and range around 0.6 with no indications of seasonality. It has previously been shown that processes other than ENSO influence the predictability of temperature in the region such as the tropical Atlantic dipole (Moura and Shukla 1981), possibly being the reason for the relatively high year-round prediction performance. With regard to precipitation, the highest prediction performance is observed from AMJ to MJJ similar as for C8 (Fig. 3e), indicating in general that climate modes other than ENSO influence predictions of precipitation at the seasonal scale in this region. Pezzi and Cavalcanti (2001) reported that precipitation anomalies in northeastern Brazil are influenced both by the SST conditions in the Central Pacific and the Tropical Atlantic SST Dipole, resulting in drier-than-normal conditions if El Niño is combined with a positive dipole, and wetter-than-normal for negative dipoles

4) EXTRATROPICS

Between the latitudes of 20° and 30°S (C10–C15), the temperature forecast performance is only marginal. Only in JJA, the scores reach values that are slightly above zero. Similarly, precipitation performance is low, except for the region to the north of Uruguay (C12), which stands out as a local peak with reliability category 5 during DJF (Figs. 4h,i). This isolated region of higher prediction performance with correlation values above 0.3 has already been detected in previous studies (e.g., Coelho et al. 2006a,b). This increased precipitation prediction performance in the region with respect to its surrounding regions stands in relationship with teleconnections of ENSO through a warm–wet/cold–dry relationship (Figs. 5c and A4) (see also Diaz et al. 1998; Montecinos et al. 2000).

No significant forecast performance was found for regions farther to the south (C13–C15, Figs. 3, 4, and A5), although various studies exist highlighting certain relationships of both temperature and precipitation with ENSO (e.g., Garbarini et al. 2016; Aceituno 1988; Montecinos et al. 2000; Montecinos and Aceituno 2003; González and Vera 2010; Rutllant and Fuenzalida 1991; Garreaud et al. 2009; Schneider and Gies 2004), some of which were also found here (Fig. 5). However, the correlations were weaker than the usefulness threshold of 0.3 and are therefore not further discussed here.

b. Is the identified prediction performance solely due to ENSO?

The comparison of a statistical forecast based on Niño-3.4 alone and SEAS5 shows that SEAS5 predictions for temperatures outperform the simple statistical model in most regions (Figs. 6a,b), providing indication that prediction performance in SEAS5 is not solely due to a simple lagged response to ENSO. For precipitation, the differences between SEAS5 and a simple statistical ENSO model are less pronounced (Fig. 6c). Especially in Peru, the highlands of Bolivia, as well as in the region to the north of Uruguay, the simple statistical model yields similar correlations on average for precipitation. In these regions, ENSO and its teleconnections are probably the only modes of variability that are well represented in SEAS5. An analysis of other modes of variability and their representation in SEAS5 could yield further insights and possibilities of improvement of the model in this region.

Certainly, more sophisticated statistical tools such as CPT also make use of additional predictors. It cannot be derived and was not the goal of these analyses that SEAS5 outperforms any statistical prediction. A thorough verification of CPT forecasts would be required to assess the potential extra benefit of one method over the other. However, the generally higher performance of SEAS5 over a simple empirical model using solely Niño-3.4 as the predictor variable for South America seasonal climate variables supports a potential benefit of using dynamical models, in addition to their advantages mentioned in the introduction of this study.

c. On the effect of spatial aggregation on forecast quality

In general, it is widely accepted that spatiotemporal aggregation of seasonal forecasts increases their performance (Buizza and Leutbecher 2015). In this section, the verification results are compared to a study by Weisheimer and Palmer (2014) done at the global scale, where regions such as South America are summarized in two areas (i.e., the continent of South America is partitioned into two parts roughly divided by the 18°S latitude). The study by Weisheimer and Palmer (2014) shows that the reliability of ECMWF System 4 of both precipitation and temperature is high over South America [see Figs. 4 and 5 in Weisheimer and Palmer (2014)], except for the lower tercile categories (cold respectively dry) in DJF. In DJF, the southern part of South America falls into the medium or ''marginally useful'' of the five reliability categories.

Despite two differences in the datasets used in Weisheimer and Palmer (2014) compared to the
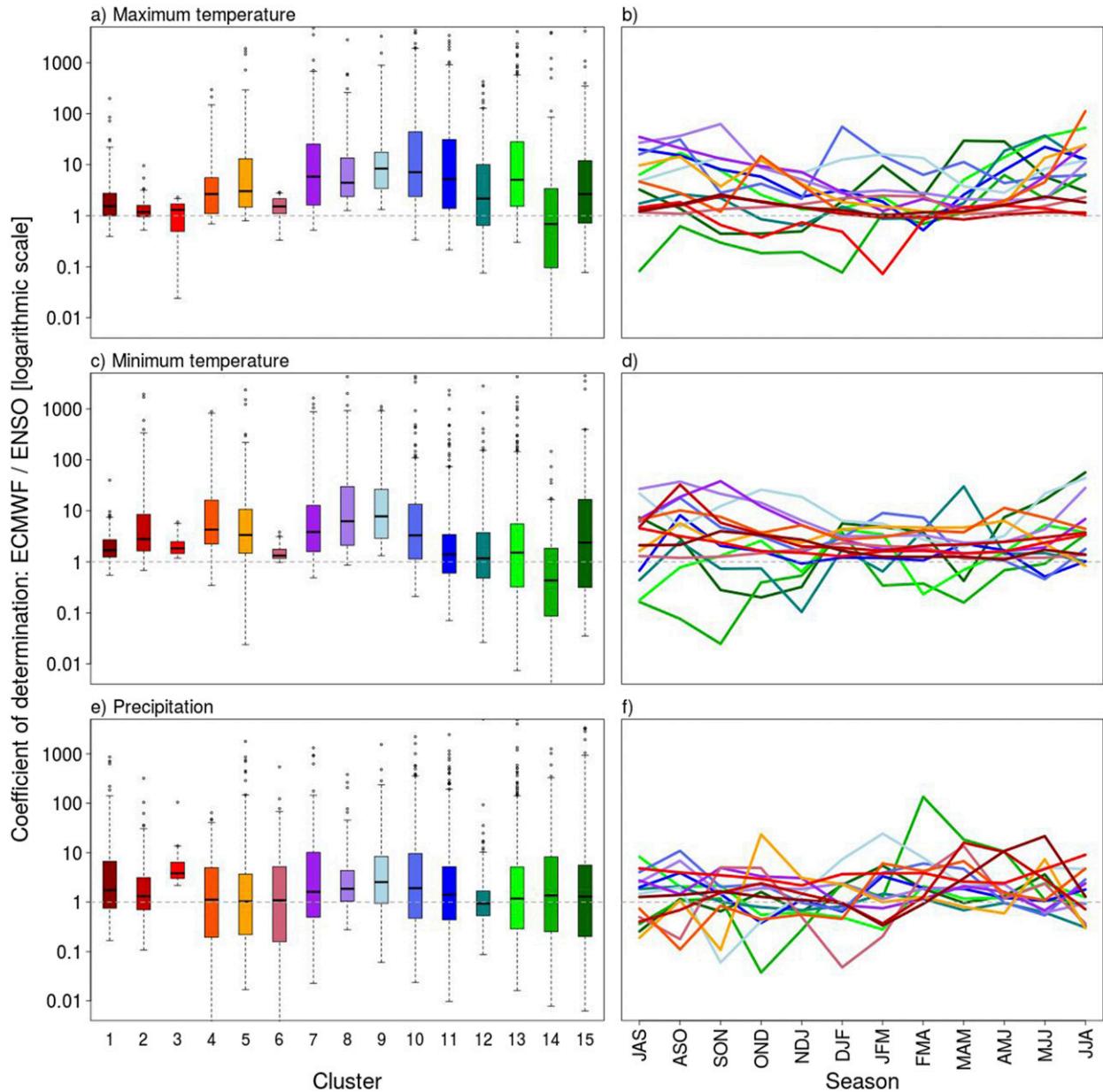
FIG. 6. Ratio of the coefficients of determination between (i) hindcasts with observations and (ii) ENSO with observations (note the logarithmic scale). Values above 1 indicate that a higher coefficient of determination was obtained by the hindcasts, meaning that the ECMWF hindcast outperform a simple statistical model using ENSO. On the other hand, values below 1 indicate that a simple lag-1 ENSO model outperforms the dynamical model.

present study (i.e., an updated version of the ECMWF seasonal prediction model and the use of station data instead of reanalysis data as ground truth), we think that new insights are gained by a comparison of the studies. For instance, more detailed spatial differences, as well as a greater complexity of the patterns are suggested by this study. It is clear that the global study by Weisheimer and Palmer (2014), aiming at introducing a simple categorization of the prediction performance and at

providing a broad global picture on reliability, does not aim at resolving individual features as done by the present study.

The reliability scores determined by Weisheimer and Palmer (2014) are mostly higher than the scores observed in this study, especially in the southern part of the continent. In regions south of 25°S, the reliability of average temperature ranges between 2 and 3 in DJF in this study (not shown), while classified between 3 and 4
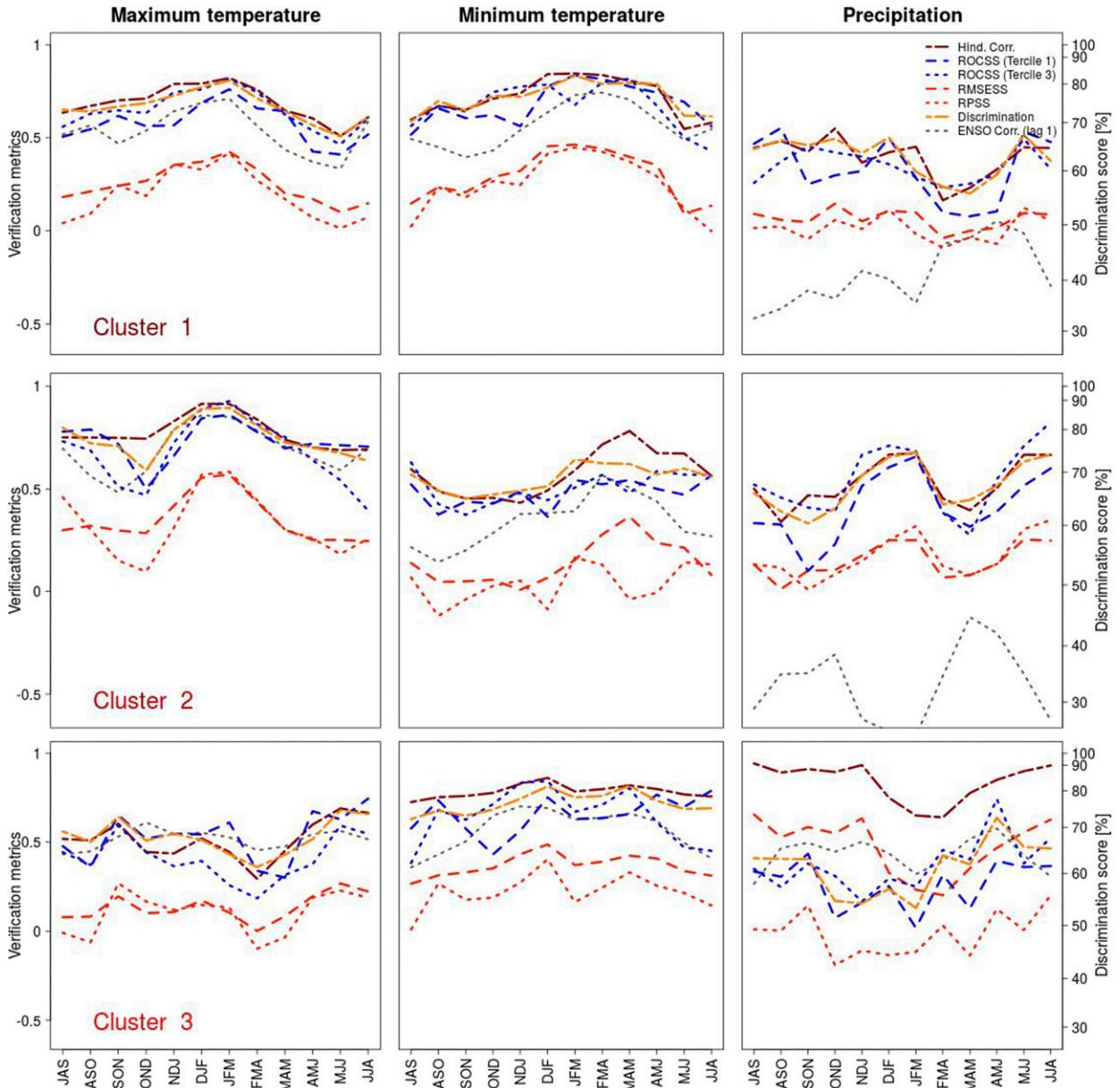
FIG. A1. Verification measures for C1–C3 for (left) maximum temperature, (center) minimum temperature, and (right) precipitation.

in Weisheimer and Palmer (2014). Higher reliabilities can be attributed to the effect of the improvement of prediction performance due to spatial aggregation (Buizza and Leutbecher 2015). Improving performance by aggregation (spatially or temporally) however limits the potential use of the seasonal forecasts to stakeholders operating at larger scales and excludes for instance smallholder farmers.

On the other hand, the opposite also occurs (i.e., we observed higher reliability at lower spatial aggregations). The most prominent example in this regard is related to precipitation forecasts in the region to the north of Uruguay (C12) in DJF [see section 4a(4)]. In this quite small cluster, "perfect" reliability for both wet and dry terciles is obtained in DJF, while the reliability score in all but one of the surrounding regions ranges between 1 and 3 (Figs. 4h,i). Obviously, this local spot of high prediction performance is spatially not resolved in Weisheimer and Palmer (2014). It is however possible that the reliability estimated by Weisheimer and Palmer (2014) is affected. The example indicates that spatial aggregation might result in the loss of information and performance, and thereby conceal potential opportunities also for stakeholders acting at smaller scales.
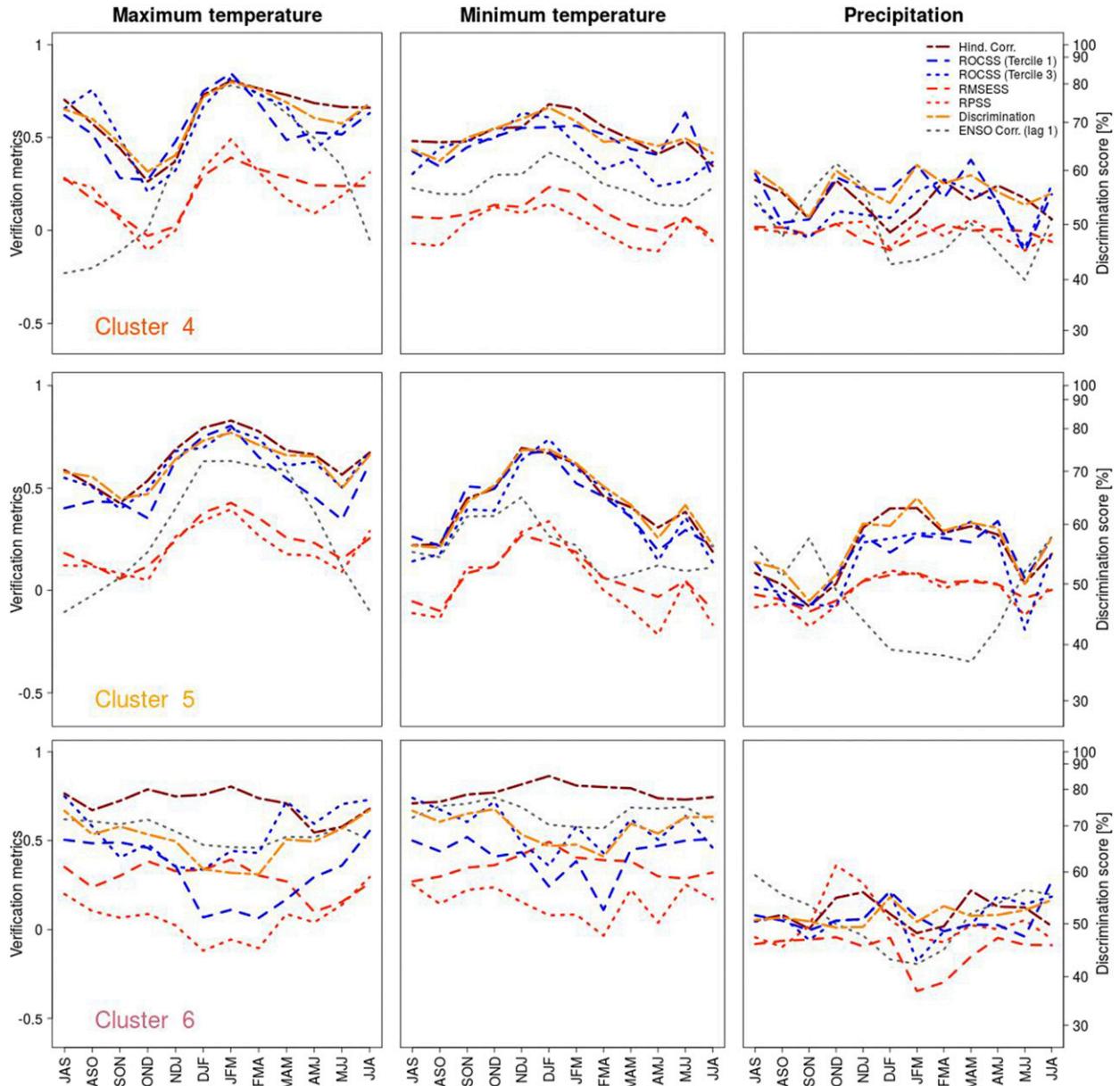
FIG. A2. As in Fig. A1, but for C4–C6.

Another discrepancy in the spatial structure is observed for the temperature reliability in JJA. In Weisheimer and Palmer (2014), the southern part of the continent has a higher reliability (5) than the northern part (4). In contrast, this study finds no differences between the southern and the northern part of the continent during JJA (Fig. A7). The reliability scores found here range around the "still useful" category (category 4) with very few individual clusters being lower or higher. In contrast to Weisheimer and Palmer (2014), the pattern observed in this study does not suggest a higher seasonal prediction quality for average temperature in JJA in the extratropical regions of South America compared to the tropical ones.

5. Conclusions

This study is the first to investigate temperature and precipitation prediction performance of a state-of-the-art dynamical seasonal forecast model against homogenized station observations over South America for all seasons. Thereby, biases in the verification result due to known biases or errors in reanalysis or
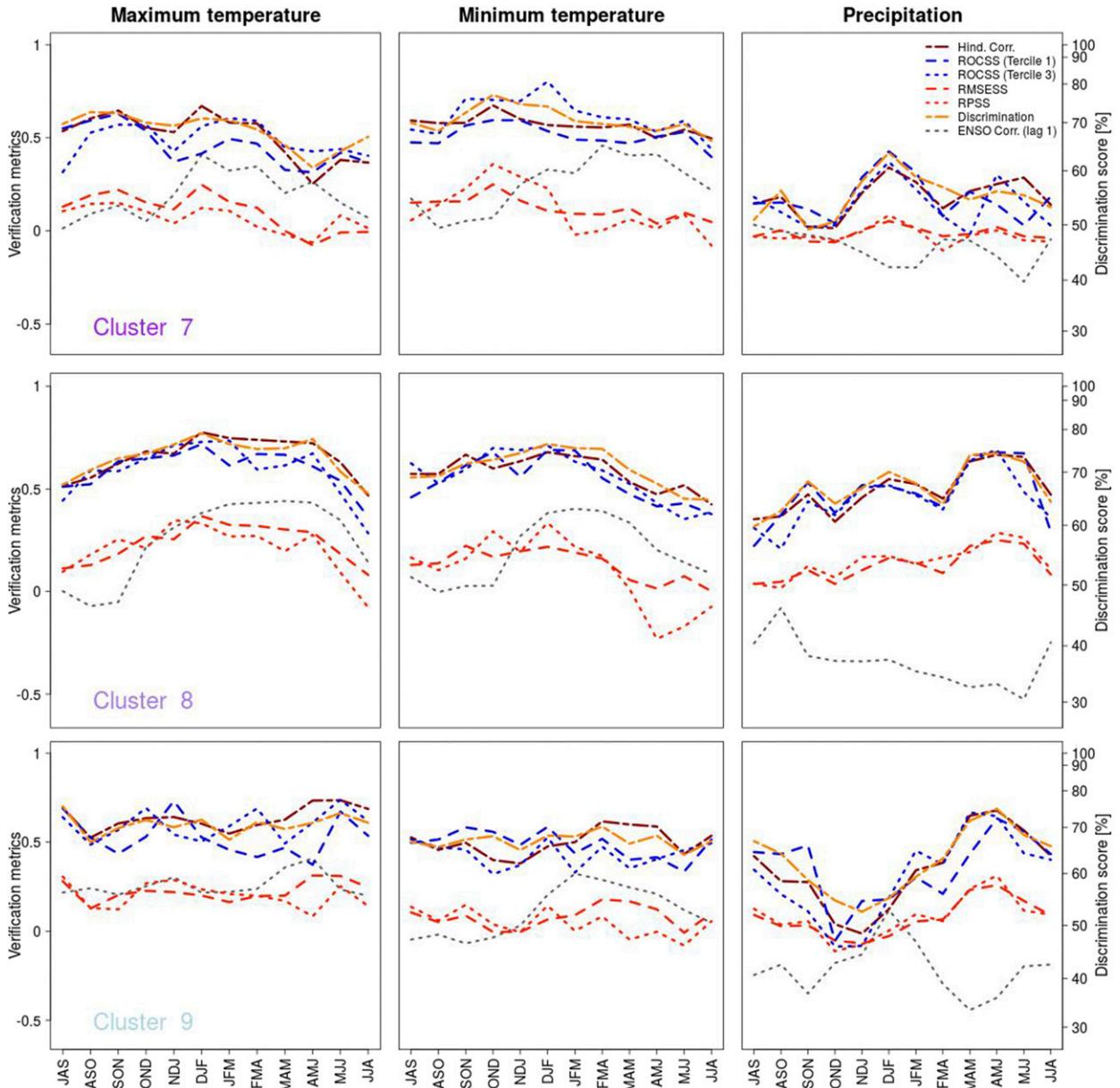
FIG. A3. As in Fig. A1, but for C7–C9.

other datasets such as those based on satellites were avoided.

In accordance with previous studies (e.g., Coelho et al. 2006a,b; Manzanas et al. 2014; Weisheimer and Palmer 2014), the highest performances for precipitation and temperature were found in those regions most strongly affected by ENSO variability (i.e., in the tropics during DJF). In the southern extratropics, generally characterized by low seasonal prediction performance, an isolated region of high precipitation prediction performance is found to the north of Uruguay in DJF, possibly due to ENSO teleconnections. Here, the prediction

performance of SEAS5 was found to be on a level that is potentially useful even for applications by smallholder farmers.

It is widely recognized that the potential of predictions can be increased through spatial and temporal aggregation. This study showed, however, that regions of high potential prediction performance with limited extent can be identified, such as the mentioned isolated subtropical region to the north of Uruguay. These findings are possibly relevant for operational prediction at smaller scales. Furthermore, the results presented here do contradict previous studies in some cases.
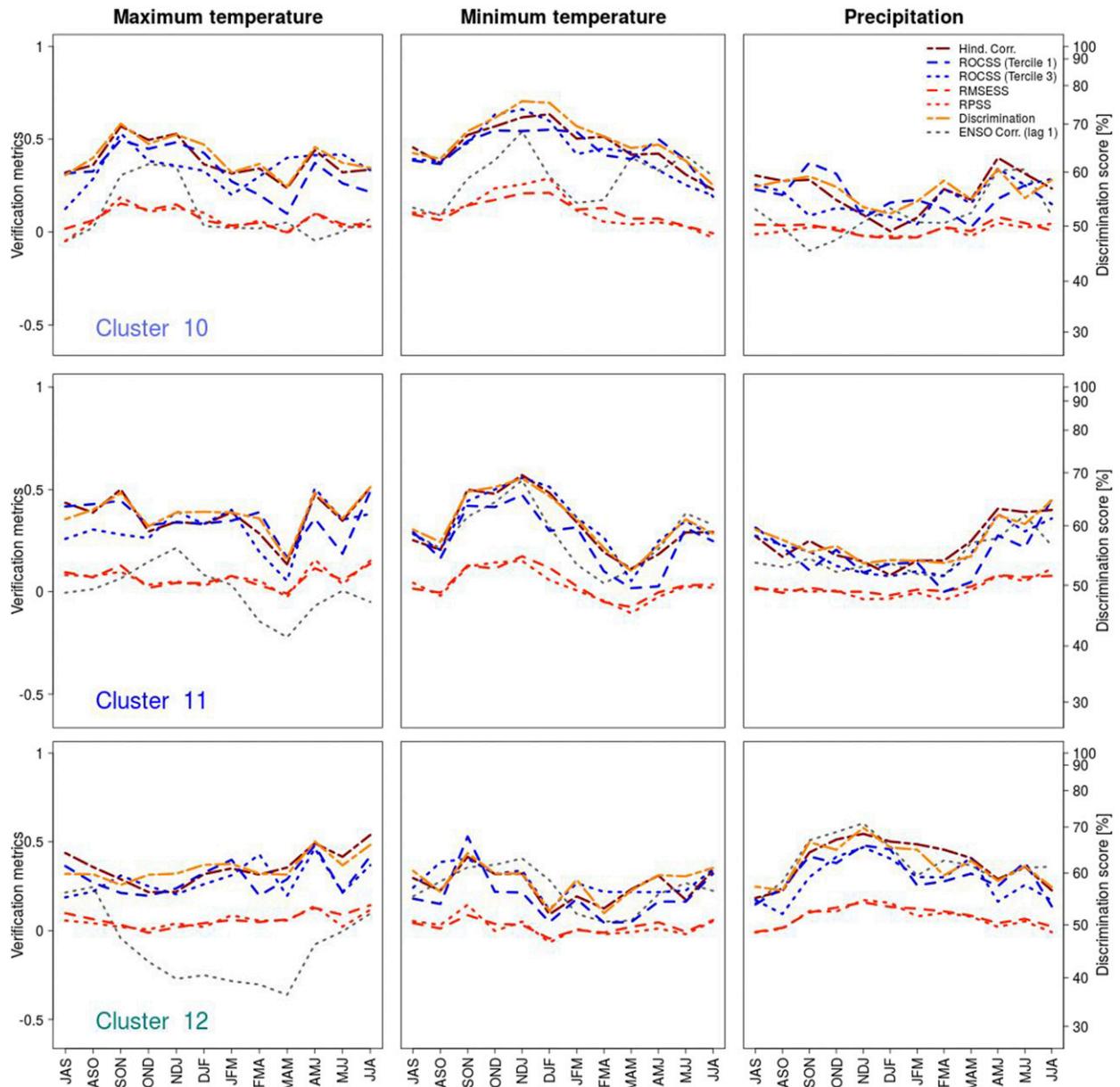
FIG. A4. As in Fig. A1, but for C10–C12.

Discrepancies were observed with regard to both spatial and temporal patterns of the prediction performance, indicating that differing prediction models and/or observation data used for verification can result in different findings.

The example of the verification metrics for precipitation at the Ecuadorian coast showed that the analysis of more than one skill metric is required to assess the performance of a model. The very high correlations of SEAS5 with precipitation observations mainly stemmed from individual strong El Niño events. Forecast quality measures based on tercile category forecasts, however, exhibited much lower values, illustrating that the model was not able to discriminate precipitation events well except for these strong El Niño events. While the prediction of heavy precipitation events during El Niño is certainly relevant for end users, this case shows that analyzing correlations alone is not sufficient to evaluate the model performance.

We conclude that the seasonal forecasts from ECMWF SEAS5 perform adequately and are reliable enough to be usefully applied in many regions. Further, we found evidence that the prediction performance of SEAS5 does not solely stem from ENSO, but also from other sources of
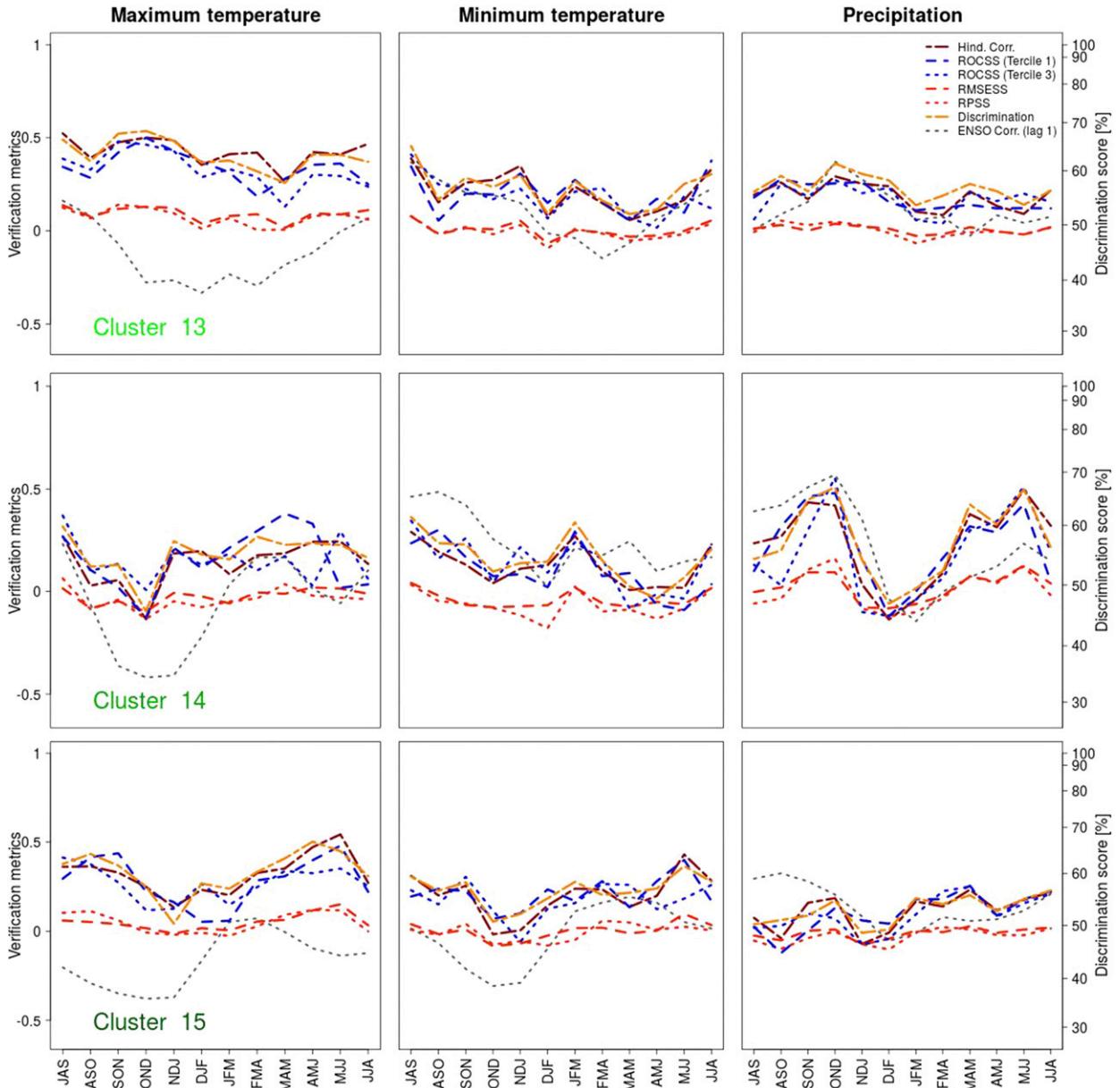
FIG. A5. As in Fig. A1, but for C13–C15.

predictability that contributed to a higher performance in all regions where high predictability was identified. Due to this benefit, we strongly encourage national weather services in South America to complement or replace their seasonal forecasts based on empirical models with dynamical model predictions, or to combine the predictions from these two modeling approaches.

FIG. A6. As in Fig. 4, but for MAM.
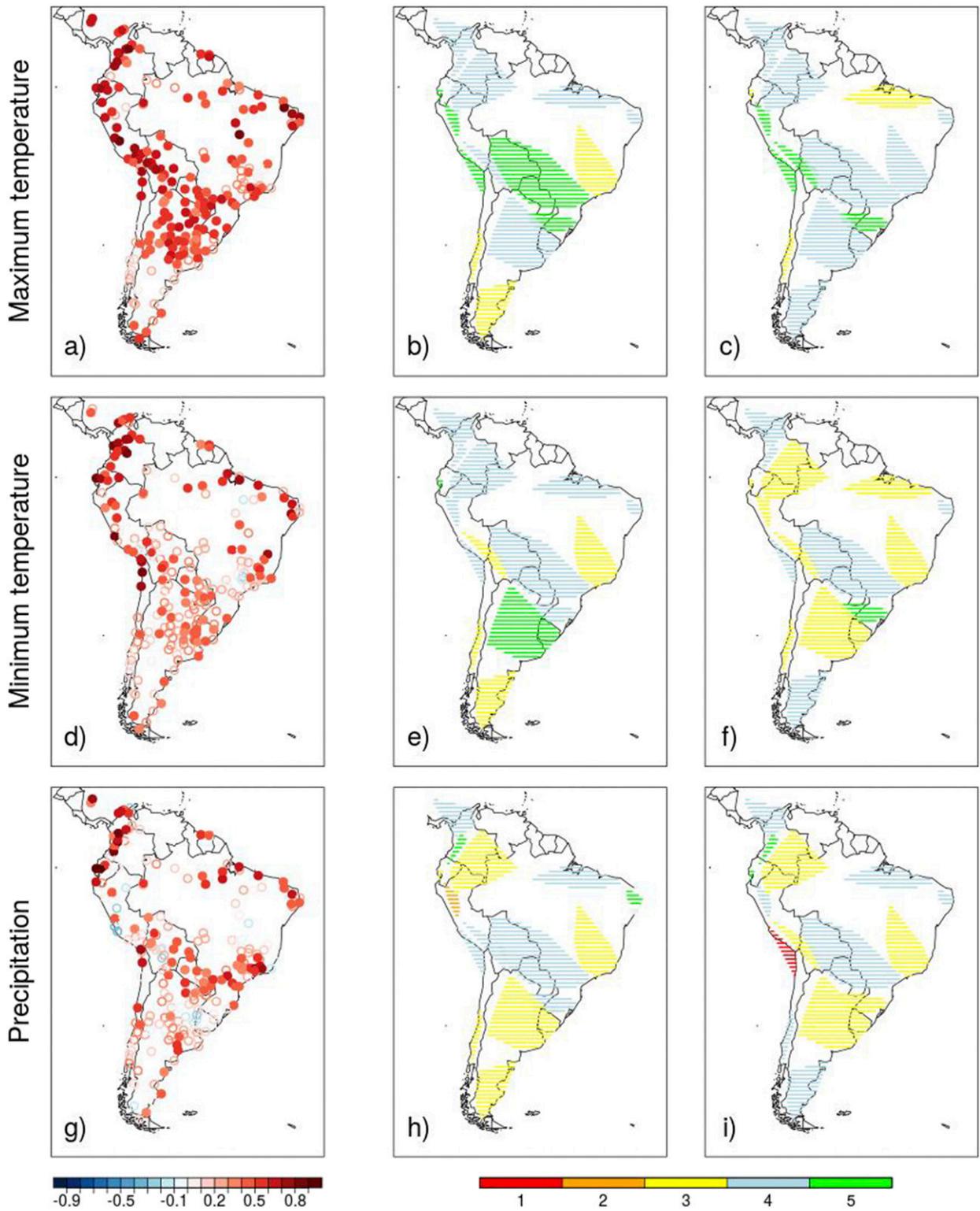
FIG. A7. As in Fig. 4, but for JJA.
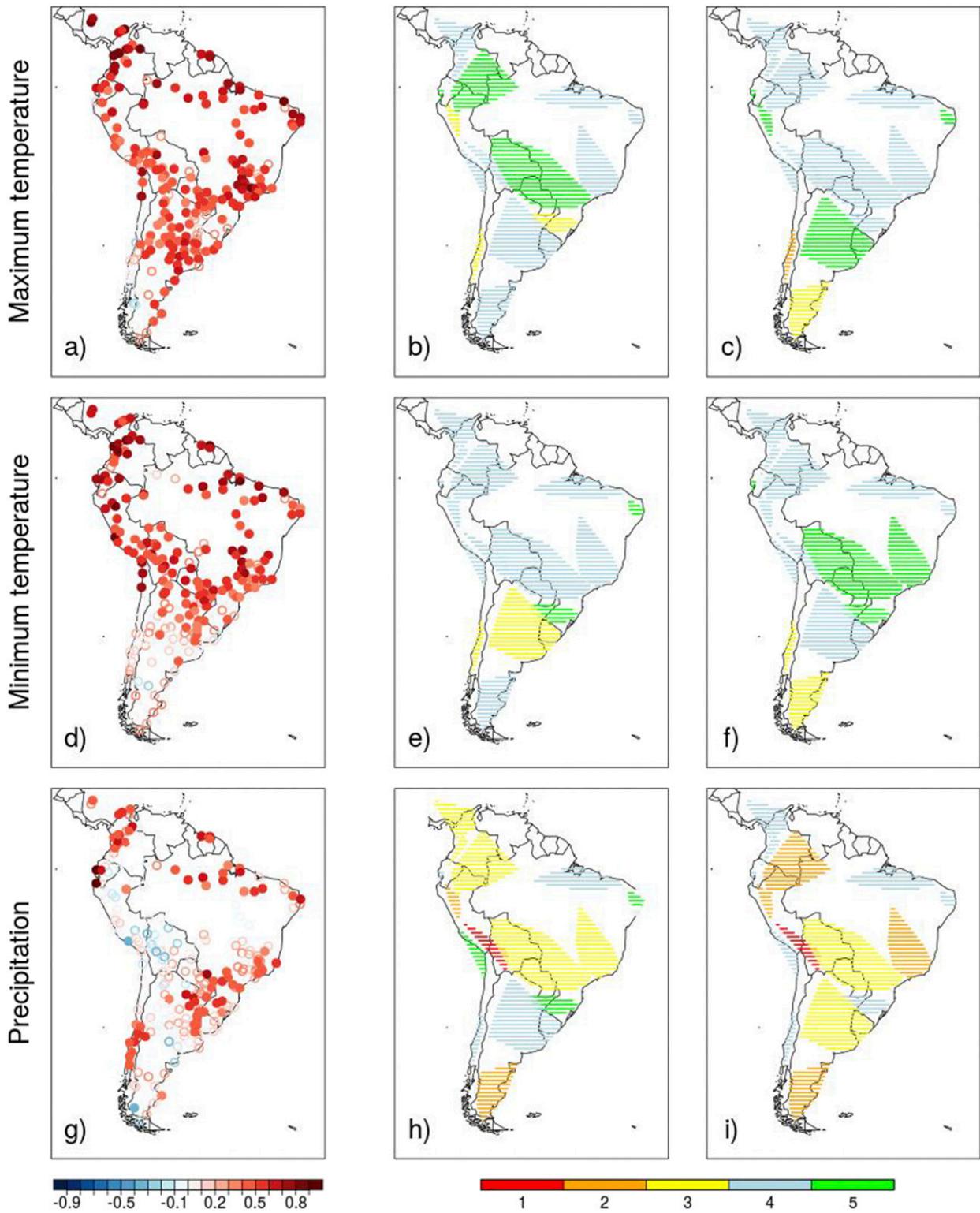
FIG. A8. As in Fig. 4, but for SON.

(CLIMAX Project) for the support received. MJC thanks projects FONDAP 15110009, FONDECYT 11170486, and PAI 79160105 (all CONICYT-Chile).

APPENDIX

Verification Metrics for Individual Clusters and Seasons

The figures in the appendix show the verification metrics for all individual clusters (Figs. A1–A5), as well as the correlation and Weisheimer categories for MAM (Fig. A6), JJA (Fig. A7), and SON (Fig. A8) complementing Fig. 4.

REFERENCES

Aceituno, P., 1988: On the functioning of the Southern Oscillation in the South American sector. Part I: Surface climate. *Mon. Wea. Rev.*, **116**, 505–524, https://doi.org/10.1175/1520-0493(1988)116<0505:OTFOTS>2.0.CO;2.

Aguilar, E., 2010: Manual de uso de RSNHT.R (software de homogenización del Center for Climate Change, URV, Tarragona, España). Centre for Climate Change self-editing, accessed 20 July 2012, http://www.c3.urv.cat/data/manual/manual_rsnht.

——, J. Sigró, and M. Brunet, 2010: RCLIMDEX con funcionalidades extras de Control de Calidad. Manual de Uso, versión 1.0, Centre for Climate Change self-editing, accessed 20 July 2012, http://www.c3.urv.cat/data/manual/Manual_rclimdex_extraQCr.

Alessandri, A., A. Borrelli, A. Navarra, A. Arribas, M. Déqué, P. Rogel, and A. Weisheimer, 2011: Evaluation of probabilistic quality and value of the ENSEMBLES multimodel seasonal forecasts: Comparison with DEMETER. *Mon. Wea. Rev.*, **139**, 581–607, https://doi.org/10.1175/2010MWR3417.1.

Alexandersson, H., and A. Moberg, 1997: Homogenization of Swedish temperature data. Part I. Homogeneity test for linear trends. *Int. J. Climatol.*, **17**, 25–34, https://doi.org/10.1002/(SICI)1097-0088(199701)17:1<25::AID-JOC103>3.0.CO;2-J.

Ash, A., P. McIntosh, B. R. Cullen, P. Carberry, and M. S. Smith, 2007: Constraints and opportunities in applying seasonal climate forecasts in agriculture. *Aust. J. Agric. Res.*, **58**, 952–956, https://doi.org/10.1071/AR06188.

Barnston, A. G., and C. F. Ropelewski, 1992: Prediction of ENSO episodes using canonical correlation analysis. *J. Climate*, **5**, 1316–1345, https://doi.org/10.1175/1520-0442(1992)005<1316:POEEUC>2.0.CO;2.

——, M. K. Tippett, M. L. L'Heureux, S. Li, and G. DeWitt, 2012: Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? *Bull. Amer. Meteor. Soc.*, **93**, 631–651, https://doi.org/10.1175/BAMS-D-11-00111.1.

Bazo, J., M. de las Nieves Lorenzo, and R. Porfirio da Rocha, 2013: Relationship between monthly rainfall in NW Peru and tropical sea surface temperature. *Adv. Meteor.*, **2013**, 1–9, https://doi.org/10.1155/2013/152875.

Bhend, J., 2017: biascorrection: Bias correct ensemble forecasts, version 0.5.11. R package, accessed 26 February 2020, http://github.com/jonasbhend/biascorrection.

Brayshaw, D. J., A. Troccoli, R. Fordham, and J. Methven, 2011: The impact of large scale atmospheric circulation patterns on wind power generation and its potential predictability: A case study over the UK. *Renewable Energy*, **36**, 2087–2096, https://doi.org/10.1016/j.renene.2011.01.025.

Brönnimann, S., 2007: Impact of El Niño-Southern Oscillation on European climate. *Rev. Geophys.*, **45**, RG3003, https://doi.org/10.1029/2006RG000199.

Buizza, R., and M. Leutbecher, 2015: The forecast skill horizon. *Quart. J. Roy. Meteor. Soc.*, **141**, 3366–3382, https://doi.org/10.1002/qj.2619.

Cantelaube, P., and J.-M. Terres, 2005: Seasonal weather forecasts for crop yield modelling in Europe. *Tellus*, **57A**, 476–487, https://doi.org/10.1111/j.1600-0870.2005.00125.x.

Ceglar, A., A. Toreti, C. Prodhomme, M. Zampieri, M. Turco, and F. J. Doblas-Reyes, 2018: Land-surface initialisation improves seasonal climate prediction skill for maize yield forecast. *Sci. Rep.*, **8**, 1322, https://doi.org/10.1038/s41598-018-19586-6.

Challinor, A. J., J. M. Slingo, T. R. Wheeler, and F. J. Doblas-Reyes, 2005: Probabilistic simulations of crop yield over western India using the DEMETER seasonal hindcast ensembles. *Tellus*, **57A**, 498–512, https://doi.org/10.1111/j.1600-0870.2005.00126.x.

Clark, R. T., P. E. Bett, H. E. Thornton, and A. A. Scaife, 2017: Skillful seasonal predictions for the European energy industry. *Environ. Res. Lett.*, **12**, 024002, https://doi.org/10.1088/1748-9326/aa57ab.

Coelho, C. A. S., C. B. Uvo, and T. Ambrizzi, 2002: Exploring the impacts of the tropical Pacific SST on the precipitation patterns over South America during ENSO periods. *Theor. Appl. Climatol.*, **71**, 185–197, https://doi.org/10.1007/s007040200004.

——, D. B. Stephenson, M. Balmaseda, F. J. Doblas-Reyes, and G. J. van Oldenborgh, 2006a: Toward an integrated seasonal forecasting system for South America. *J. Climate*, **19**, 3704–3721, https://doi.org/10.1175/JCLI3801.1.

——, ——, F. J. Doblas-Reyes, and M. Balmaseda, 2006b: The skill of empirical and combined/calibrated coupled multi-model South American seasonal predictions during ENSO. *Adv. Geosci.*, **6**, 51–55, https://doi.org/10.5194/ADGEO-6-51-2006.

——, and Coauthors, 2012: Climate diagnostics of three major drought events in the Amazon and illustrations of their seasonal precipitation predictions. *Meteor. Appl.*, **19**, 237–255, https://doi.org/10.1002/met.1324.

Córdoba-Machado, S., R. Palomino-Lemus, S. R. Gámiz-Fortis, Y. Castro-Díez, and M. J. Esteban-Parra, 2015a: Influence of tropical Pacific SST on seasonal precipitation in Colombia: Prediction using El Niño and El Niño Modoki. *Climate Dyn.*, **44**, 1293–1310, https://doi.org/10.1007/s00382-014-2232-3.

——, ——, ——, ——, and ——, 2015b: Assessing the impact of El Niño Modoki on seasonal precipitation in Colombia. *Global Planet. Change*, **124**, 41–61, https://doi.org/10.1016/j.gloplacha.2014.11.003.

De Felice, M., A. Alessandri, and F. Catalano, 2015: Seasonal climate forecasts for medium-term electricity demand forecasting. *Appl. Energy*, **137**, 435–444, https://doi.org/10.1016/j.apenergy.2014.10.030.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/10.1002/qj.828.

Diaz, A. F., C. D. Studzinski, and C. R. Mechoso, 1998: Relationships between precipitation anomalies in Uruguay and southern Brazil and sea surface temperature in the Pacific and Atlantic Oceans. *J. Climate*, **11**, 251–271, https://doi.org/10.1175/1520-0442(1998)011<0251:RBPAIU>2.0.CO;2.

Emanuel, K., F. Fondriest, and J. Kossin, 2012: Potential economic value of seasonal hurricane forecasts. *Wea. Climate Soc.*, **4**, 110–117, https://doi.org/10.1175/WCAS-D-11-00017.1.

Erfanian, A., G. Wang, and L. Fomenko, 2017: Unprecedented drought over tropical South American in 2016: Significantly under-predicted by tropical SST. *Sci. Rep.*, **7**, 5811, https://doi.org/10.1038/s41598-017-05373-2.

Garbarini, E., M. M. Skansi, M. H. Gonzalez, and A. Rolla, 2016: ENSO influence over precipitation in Argentina. *Advances in Environmental Research*, J. A. Daniels, Ed., Vol. 52, NOVA Publisher, 223–246.

Garreaud, R. D., 2018: A plausible atmospheric trigger for the 2017 coastal El Niño. *Int. J. Climatol.*, **38**, e1296–e1302, https://doi.org/10.1002/joc.5426.

——, M. Vuille, R. Compagnucci, and J. Marengo, 2009: Present-day South American climate. *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, **281**, 180–195, https://doi.org/10.1016/j.palaeo.2007.10.032.

González, M. H., and C. S. Vera, 2010: On the interannual wintertime rainfall variability in the Southern Andes. *Int. J. Climatol.*, **30**, 643–657, https://doi.org/10.1002/JOC.1910.

Grimm, A., V. Barros, and M. Doyle, 2000: Climate variability in southern South America associated with El Niño and La Niña events. *J. Climate*, **13**, 35–58, https://doi.org/10.1175/1520-0442(2000)013<0035:CVISSA>2.0.CO;2.

Gudmundsson, L., J. B. Bremnes, J. E. Haugen, and T. Engen-Skaugen, 2012: Downscaling RCM precipitation to the station scale using statistical transformations—A comparison of methods. *Hydrol. Earth Syst. Sci.*, **16**, 3383–3390, https://doi.org/10.5194/hess-16-3383-2012.

Hansen, J. W., 2002: Realizing the potential benefits of climate prediction to agriculture: Issues, approaches, challenges. *Agric. Syst.*, **74**, 309–330, https://doi.org/10.1016/S0308-521X(02)00043-4.

Hewitt, C., S. Mason, and D. Walland, 2012: The global framework for climate services. *Nat. Climate Change*, **2**, 831–832, https://doi.org/10.1038/nclimate1745.

Hofer, M., B. Marzeion, and T. Mölg, 2012: Comparing the skill of different reanalyses and their ensembles as predictors for daily air temperature on a glaciated mountain (Peru). *Climate Dyn.*, **39**, 1969–1980, https://doi.org/10.1007/s00382-012-1501-2.

Huang, B., and Coauthors, 2017: Extended Reconstructed Sea Surface Temperature, version 5 (ERSSTv5), upgrades, validations, and intercomparisons. *J. Climate*, **30**, 8179–8205, https://doi.org/10.1175/JCLI-D-16-0836.1.

Hunziker, S., and Coauthors, 2017: Identifying, attributing, and overcoming common data quality issues of manned station observations. *Int. J. Climatol.*, **37**, 4131–4145, https://doi.org/10.1002/joc.5037.

Hyndman, R. J., and A. B. Koehler, 2006: Another look at measures of forecast accuracy. *Int. J. Forecasting*, **22**, 679–688, https://doi.org/10.1016/j.ijforecast.2006.03.001.

Imfeld, N., C. Barreto Schuler, K. M. Correa Marrou, M. Jacques-Coper, K. Sedlmeier, S. Gubler, A. Huerta, and S. Brönnimann, 2019: Summertime precipitation deficits in the southern Peruvian highlands since 1964. *Int. J. Climatol.*, **39**, 4497–4513, https://doi.org/10.1002/joc.6087.

Jewson, S., and A. Brix, 2005: *Weather Derivative Valuation: The Meteorological, Statistical, Financial and Mathematical Foundations*. Cambridge University Press, 373 pp.

Johnson, S. J., and Coauthors, 2019: SEAS5: The new ECMWF seasonal forecast system. *Geosci. Model Dev.*, **12**, 1087–1117, https://doi.org/10.5194/gmd-12-1087-2019.

Jolliffe, I. T., and D. B. Stephenson, 2012: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. 2nd ed. John Wiley & Sons, 292 pp.

Karpechko, A. Y., K. A. Peterson, A. A. Scaife, J. Vainio, and H. Gregow, 2015: Skilful seasonal predictions of Baltic Sea ice cover. *Environ. Res. Lett.*, **10**, 044007, https://doi.org/10.1088/1748-9326/10/4/044007.

Kaufman, L., and P. J. Rousseeuw, 1990: *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, 342 pp.

Kim, H.-M., P. J. Webster, and J. A. Curry, 2012: Seasonal prediction skill of ECMWF system 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere winter. *Climate Dyn.*, **39**, 2957–2973, https://doi.org/10.1007/s00382-012-1364-6.

Lagos, P., Y. Silva, E. Nickl, and K. Mosquera, 2008: El Niño–related precipitation variability in Perú. *Adv. Geosci.*, **14**, 231–237, https://doi.org/10.5194/adgeo-14-231-2008.

Lance, G. N., and W. T. Williams, 1967: A general theory of classificatory sorting strategies: 1. Hierarchical systems. *Comput. J.*, **9**, 373–380, https://doi.org/10.1093/comjnl/9.4.373.

Lavado-Casimiro, W., and J. C. Espinoza, 2014: Impactos de El Niño y La Niña en las lluvias del Perú (Impacts of El Niño and La Niña in the precipitation over Peru (1965–2007). *Rev. Bras. Meteor.*, **29**, 171–182, https://doi.org/10.1590/S0102-77862014000200003.

Lavado-Casimiro, W. S., J. Ronchail, D. Labat, J. C. Espinoza, and J. L. Guyot, 2012: Basin-scale analysis of rainfall and runoff in Peru (1969–2004), Pacific, Titicaca and Amazonas drainages. *Hydrol. Sci. J.*, **57**, 625–642, https://doi.org/10.1080/02626667.2012.672985.

Lee, J.-Y., and Coauthors, 2010: How are seasonal prediction skills related to models' performance on mean state and annual cycle? *Climate Dyn.*, **35**, 267–283, https://doi.org/10.1007/s00382-010-0857-4.

——, B. Wang, Q. Ding, K.-J. Ha, J.-B. Ahn, A. Kumer, B. Stern, and O. Alves, 2011: How predictable is the Northern Hemisphere summer upper-tropospheric circulation? *Climate Dyn.*, **37**, 1189–1203, https://doi.org/10.1007/s00382-010-0909-9.

Manzanas, R., M. D. Frías, A. S. Cofiño, and J. M. Gutiérrez, 2014: Validation of 40 year multimodel seasonal precipitation forecasts: The role of ENSO on the global skill. *J. Geophys. Res. Atmos*, **119**, 1708–1719, https://doi.org/10.1002/2013JD020680.

Marengo, J. A., and Coauthors, 2008: The drought of Amazonia in 2005. *J. Climate*, **21**, 495–516, https://doi.org/10.1175/2007JCLI1600.1.

Mason, J. S., and N. E. Graham, 1999: Conditional probabilities, relative operating characteristics, and relative operating levels. *Wea. Forecasting*, **14**, 713–725, https://doi.org/10.1175/1520-0434(1999)014<0713:CPROCA>2.0.CO;2.

——, and O. Baddour, 2008: Statistical modelling. *Seasonal Climate: Forecasting and Managing Risk*, A. Troccoli et al., Eds., NATO Science Series, Vol. 82, Springer, 163–185.

——, and M. K. Tippett, 2016: Climate predictability tool version 15.3.9. Columbia University Academic Commons, accessed 26 February 2020, https://doi.org/10.7916/D8668DCW.

Massonnet, F., O. Bellprat, V. Guemas, and F. J. Doblas-Reyes, 2016: Using climate models to estimate the quality of global observational data sets. *Science*, **354**, 452–455, https://doi.org/10.1126/science.aaf6369.

MeteoSwiss, 2016: EasyVerification: Ensemble verification for large data sets. Accessed 2 August 2016, https://CRAN.R-project.org/package=SpecsVerification.

Molteni, F., and Coauthors, 2011: The new ECMWF seasonal forecast system (System 4). ECMWF Tech. Memo. 656, 49 pp., https://www.ecmwf.int/sites/default/files/elibrary/2011/11209-new-ecmwf-seasonal-forecast-system-system-4.pdf.

Montecinos, A., and P. Aceituno, 2003: Seasonality of the ENSO related rainfall variability in central Chile and associated circulation anomalies. *J. Climate*, **16**, 281–296, https://doi.org/10.1175/1520-0442(2003)016<0281:SOTERR>2.0.CO;2.

——, A. Díaz, and P. Aceituno, 2000: Seasonal diagnostic and predictability of rainfall in subtropical South America based on tropical Pacific SST. *J. Climate*, **13**, 746–758, https://doi.org/10.1175/1520-0442(2000)013<0746:SDAPOR>2.0.CO;2.

Morse, A. P., F. Doblas-Reyes, M. Hoshen, R. Hagedorn, and T. N. Palmer, 2005: A forecast quality assessment of an end-to-end probabilistic multi-model seasonal forecast system using a malaria model. *Tellus*, **57A**, 64–475, https://doi.org/10.3402/tellusa.v57i3.14668.

Moura, A. D., and J. Shukla, 1981: On the dynamics of droughts in Northeast Brazil: Observations, theory and numerical experiments with a general circulation model. *J. Atmos. Sci.*, **38**, 2653–2675, https://doi.org/10.1175/1520-0469(1981)038<2653:OTDODI>2.0.CO;2.

Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281–293, https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2.

NIDIS, 2004: Creating a drought early warning system for the 21st century: The National Integrated Drought Information System (NIDIS). National Oceanic and Atmospheric Administration (NOAA) and the Western Governor's Association (WGA) Rep., 16 pp., https://www.drought.gov/drought/sites/drought.gov.drought/files/media/whatisnidis/Documents/200406_WGA_NIDIS_Report.pdf.

Osman, M., and C. S. Vera, 2017: Climate predictability and prediction skill on seasonal time scales over South America from CHFP models. *Climate Dyn.*, **49**, 2365–2383, https://doi.org/10.1007/s00382-016-3444-5.

Palin, E. J., A. A. Scaife, E. Wallace, E. C. Pope, A. Arribas, and A. Brookshaw, 2016: Skillful seasonal forecasts of winter disruption to the UK transport system. *J. Appl. Meteor. Climatol.*, **55**, 325–344, https://doi.org/10.1175/JAMC-D-15-0102.1.

Palmer, T. N., 2002: The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Quart. J. Roy. Meteor. Soc.*, **128**, 747–774, https://doi.org/10.1256/0035900021643593.

——, and Coauthors, 2004: Development of a European Multimodel Ensemble System for Seasonal to Interannual Prediction (DEMETER). *Bull. Amer. Meteor. Soc.*, **85**, 853–872, https://doi.org/10.1175/BAMS-85-6-853.

Panisset, J. S., R. Libonati, C. M. P. Gouveia, F. Machado-Silva, D. A. Franca, J. R. A. Franca, and L. F. Peres, 2018: Contrasting patterns of the extreme drought episodes of 2005, 2010 and 2015 in the Amazon Basin. *Int. J. Climatol.*, **38**, 1096–1104, https://doi.org/10.1002/joc.5224.

Pearson, K., 1895: Notes on regression and inheritance in the case of two parents. *Proc. Roy. Soc. London*, **58**, 240–242, https://doi.org/10.1098/rspl.1895.0041.

Penalba, O. C., and J. A. Rivera, 2016: Precipitation response to El Niño/La Niña events in Southern South America—Emphasis in regional drought occurrences. *Adv. Geosci.*, **42**, 1–14, https://doi.org/10.5194/adgeo-42-1-2016.

Pezzi, L. P., and I. F. A. Cavalcanti, 2001: The relative importance of ENSO and tropical Atlantic sea surface temperature anomalies for seasonal precipitation over South America: A numerical study. *Climate Dyn.*, **17**, 205–212, https://doi.org/10.1007/s003820000104.

Poveda, G., D. M. Álvarez, and Ó. A. Rueda, 2011: Hydroclimatic variability over the Andes of Colombia associated with ENSO: A review of climatic processes and their impact on one of the earth's most important biodiversity hotspots. *Climate Dyn.*, **36**, 2233–2249, https://doi.org/10.1007/s00382-010-0931-y.

Pozzi, W., and Coauthors, 2013: Toward global drought early warning capability: Expanding international cooperation for the development of a framework for monitoring and forecasting. *Bull. Amer. Meteor. Soc.*, **94**, 776–785, https://doi.org/10.1175/BAMS-D-11-00176.1.

R Core Team, 2012: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, accessed 26 February 2020, http://www.R-project.org/.

Recalde-Coronel, G. C., A. G. Barnston, and Á. G. Muñoz, 2014: Predictability of December–April rainfall in coastal and Andean Ecuador. *J. Appl. Meteor. Climatol.*, **53**, 1471–1493, https://doi.org/10.1175/JAMC-D-13-0133.1.

Ronchail, J., G. Cochonneau, M. Molinier, J. Guyot, A. G. De Miranda Chaves, V. Guimarães, and E. de Oliveira, 2002: Interannual rainfall variability in the Amazon basin and seasurface temperatures in the equatorial Pacific and the tropical Atlantic Oceans. *Int. J. Climatol.*, **22**, 1663–1686, https://doi.org/10.1002/joc.815.

Rosas, G., and Coauthors, 2016: Towards implementing climate services in Peru—The project Climandes. *Climate. Serv.*, **4**, 30–41, https://doi.org/10.1016/j.cliser.2016.10.001.

Rutllant, J., and H. Fuenzalida, 1991: Synoptic aspects of the central Chile rainfall variability associated with the Southern Oscillation. *Int. J. Climatol.*, **11**, 63–76, https://doi.org/10.1002/joc.3370110105.

Saha, S., and Coauthors, 2006: The NCEP Climate Forecast System. *J. Climate*, **19**, 3483–3517, https://doi.org/10.1175/JCLI3812.1.

Schneider, C., and D. Gies, 2004: Effects of El Niño-Southern Oscillation on southernmost South America precipitation at 53°S revealed from NCEP-NCAR reanalyses and weather station data. *Int. J. Climatol.*, **24**, 1057–1076, https://doi.org/10.1002/joc.1057.

Schubert, S., R. Koster, M. Hoerling, R. Saeger, D. Lettenmaier, A. Kumar, and D. Gutzler, 2007: Supplement to predicting drought on seasonal-to-decadal time scales: A national drought attribution and prediction consortium. *Bull. Amer. Meteor. Soc.*, **88**, S9–S10, https://doi.org/10.1175/BAMS-88-10-Schubert.

Shafiee-Jood, M., X. Cai, L. Chen, X.-Z. Liang, and P. Kumar, 2014: Assessing the value of seasonal climate forecast information through an end-to-end forecasting framework: Application to U.S. 2012 drought in central Illinois. *Water Resour. Res.*, **50**, 6592–6609, https://doi.org/10.1002/2014WR015822.

Shimizu, M. H., T. Ambrizzi, and B. Liebmann, 2017: Extreme precipitation events and their relationship with ENSO and

MJO phases over northern South America. *Int. J. Climatol.*, **37**, 2977–2989, https://doi.org/10.1002/joc.4893.

Siegert, S., J. Bhend, I. Kroener, and M. de Felice, 2017: Package 'SpecsVerification': Forecast verification routines for ensemble forecasts of weather and climate. R-project, 44 pp., https://cran.r-project.org/web/packages/SpecsVerification/SpecsVerification.pdf.

Silva, Y., K. Takahashi, and R. Chávez, 2008: Dry and wet rainy seasons in the Mantaro river basin (Central Peruvian Andes). *Adv. Geosci.*, **14**, 261–264, https://doi.org/10.5194/adgeo-14-261-2008.

Skansi, M., and Coauthors, 2013: Warming and wetting signals emerging from analysis of changes in climate extreme indices over South America. *Global Planet. Change*, **100**, 295–307, https://doi.org/10.1016/j.gloplacha.2012.11.004.

Struyf, A., M. Hubert, and P. J. Rousseeuw, 1996: Clustering in an object-oriented environment. *J. Stat. Software*, **1**, 16223, https://doi.org/10.18637/jss.v001.i04.

——, ——, and ——, 1997: Integrating robust clustering techniques in S-PLUS. *Comput. Stat. Data Anal.*, **26**, 17–37, https://doi.org/10.1016/S0167-9473(97)00020-0.

Sulca, J., M. Vuille, Y. Silva, and K. Takahashi, 2016: Teleconnections between the Peruvian Central Andes and Northeast Brazil during extreme rainfall events in austral summer. *J. Hydrometeor.*, **17**, 499–515, https://doi.org/10.1175/JHM-D-15-0034.1.

——, K. Takahashi, J.-C. Espinoza, M. Vuille, and W. Lavado-Casimiro, 2018: Impacts of different ENSO flavors and tropical Pacific convection variability (ITCZ, SPCZ) on austral summer rainfall in South America, with a focus on Peru. *Int. J. Climatol.*, **38**, 420–435, https://doi.org/10.1002/joc.5185.

Svensson, C., and Coauthors, 2015: Long-range forecasts of UK winter hydrology. *Environ. Res. Lett.*, **10**, 064006, https://doi.org/10.1088/1748-9326/10/6/064006.

Takahashi, K., 2004: The atmospheric circulation associated with extreme rainfall events in Piura, Peru, during 1997–1998 and 2002 El Niño events. *Ann. Geophys.*, **22**, 3917–3926, https://doi.org/10.5194/angeo-22-3917-2004.

——, A. Montecinos, K. Goubanova, and B. Dewitte, 2011: ENSO regimes: Reinterpreting the canonical and Modoki El Niño. *Geophys. Res. Lett.*, **38**, L10704, https://doi.org/10.1029/2011GL047364.

Tedeschi, R. G., A. M. Grimm, and I. F. A. Cavalcanti, 2015: Influence of central and east ENSO on extreme events of precipitation in South America during austral spring and summer. *Int. J. Climatol.*, **35**, 2045–2064, https://doi.org/10.1002/joc.4106.

——, ——, and ——, 2016: Influence of central and east ENSO on precipitation and its extreme events in South America during austral autumn and winter. *Int. J. Climatol.*, **36**, 4797–4814, https://doi.org/10.1002/joc.4670.

Trenberth, K. E., 1997: The definition of El Niño. *Bull. Amer. Meteor. Soc.*, **78**, 2771–2777, https://doi.org/10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2.

Troccoli, A., 2010: Seasonal climate forecasting. *Meteor. Appl.*, **17**, 251–268, https://doi.org/10.1002/met.184.

Uvo, C. B., C. A. Repelli, S. E. Zebiak, and Y. Kushnir, 1998: The relationships between tropical Pacific and Atlantic SST and northeast Brazil monthly precipitation. *J. Climate*, **11**, 551–562, https://doi.org/10.1175/1520-0442(1998)011<0551:TRBTPA>2.0.CO;2.

Vicente-Serrano, S. M., and Coauthors, 2017: The complex influence of ENSO on droughts in Ecuador. *Climate Dyn.*, **48**, 405–427, https://doi.org/10.1007/s00382-016-3082-y.

Vuille, M., 1999: Atmospheric circulation over the Bolivian Altiplano during dry and wet periods and extreme phases of the Southern Oscillation. *Int. J. Climatol.*, **19**, 1579–1600, https://doi.org/10.1002/(SICI)1097-0088(19991130)19:14<1579::AID-JOC441>3.0.CO;2-N.

——, R. S. Bradley, and F. Keimig, 2000: Climate variability in the Andes of Ecuador and its relation to tropical Pacific and Atlantic sea surface temperature anomalies. *J. Climate*, **13**, 2520–2535, https://doi.org/10.1175/1520-0442(2000)013<2520:CVITAO>2.0.CO;2.

Wang, B., and Coauthors, 2009: Advance and prospectus of seasonal prediction: Assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004). *Climate Dyn.*, **33**, 93–117, https://doi.org/10.1007/s00382-008-0460-0.

Wang, X. L., H. Chen, Y. Wu, Y. Feng, and Q. Pu, 2010: New techniques for the detection and adjustment of shifts in daily precipitation data series. *J. Appl. Meteor. Climatol.*, **49**, 2416–2436, https://doi.org/10.1175/2010JAMC2376.1.

Waylen, P., and G. Poveda, 2002: El Niño-Southern Oscillation and aspects of western South American hydro-climatology. *Hydrol. Processes*, **16**, 1247–1260, https://doi.org/10.1002/hyp.1060.

Wehrli, K., J. Bhend, and M. A. Liniger, 2017: Systematic quality assessment of an operational seasonal forecasting system. Tech. Rep. 263, MeteoSwiss, 52 pp.

Weigel, A. P., and S. J. Mason, 2011: The generalized discrimination scores for ensemble forecasts. *Mon. Wea. Rev.*, **139**, 3069–3074, https://doi.org/10.1175/MWR-D-10-05069.1.

——, M. Liniger, and C. Appenzeller, 2007: The discrete Brier and ranked probability skill scores. *Mon. Wea. Rev.*, **135**, 118–124, https://doi.org/10.1175/MWR3280.1.

Weisheimer, A., and T. N. Palmer, 2014: On the reliability of seasonal climate forecasts. *J. Roy. Soc. Interface*, **11**, 20131162, https://doi.org/10.1098/rsif.2013.1162.

WMO, 2007: Use of seasonal forecasts and climate prediction in operational agriculture. WMO/TD-1344, WMO, 81 pp., http://www.wamis.org/agm/pubs/CAGMRep/CAGM102.pdf.

——, 2011: Climate knowledge for action: A global framework for climate services and empowering the most vulnerable. WMO TD-1065, WMO, 248 pp.

Yoon, J.-H., and N. Zeng, 2010: An Atlantic influence on Amazon rainfall. *Climate Dyn.*, **34**, 249–264, https://doi.org/10.1007/s00382-009-0551-6.

Yuan, X., and E. F. Wood, 2013: Multimodel seasonal forecasting of global drought onset. *Geophys. Res. Lett.*, **40**, 4900–4905, https://doi.org/10.1002/grl.50949.

Zhang, X., and F. Yang, 2004: RClimDex (1.0) user manual. Climate Research Branch, Environment Canada, 23 pp., accessed 26 February 2020, http://etccdi.pacificclimate.org/soft.

Zhou, J., and K.-M. Lau, 2001: Principal modes of interannual and decadal variability of summer rainfall over South America. *Int. J. Climatol.*, **21**, 1623–1644, https://doi.org/10.1002/joc.700.

Ziervogel, G., M. Bithell, R. Washington, and T. Downing, 2005: Agent-based social simulation: A method for assessing the impact of seasonal climate forecast applications among smallholder farmers. *Agric. Syst.*, **83**, 1–26, https://doi.org/10.1016/j.agsy.2004.02.009.