AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE

10.1002/2013JD020680

Validation of 40 year multimodel seasonal precipitation forecasts: The role of ENSO on the global skill

R. Manzanas[1], M. D. Frías[2], A. S. Cofiño[2], and J. M. Gutiérrez[1]

[1]Grupo de Meteorología, Instituto de Física de Cantabria, CSIC-Universidad de Cantabria, Santander (Spain), [2]Grupo de Meteorología, Departamento Matemática Aplicada y Ciencias de la Computación, Universidad de Cantabria, Santander (Spain)

Key Points:
- Global user-oriented validation of seasonal rainfall forecasts for 1961–2000
- Assessment of the seasonal and spatial distribution of the models' skill found
- Proper attribution of the models' performance to the ENSO phenomenon

Correspondence to:
R. Manzanas,
rmanzanas@ifca.unican.es

Citation:
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, doi:10.1002/2013JD020680.

Received 22 AUG 2013
Accepted 16 JAN 2014
Accepted article online 21 JAN 2014
Published online 25 FEB 2014

Abstract The skill of seasonal precipitation forecasts is assessed worldwide—grid point by grid point—for the 40 year period 1961–2000, considering the ENSEMBLES multimodel hindcast and applying a tercile-based probabilistic approach in terms of the relative operating characteristic skill score (ROCSS). Although predictability varies with region, season, and lead time, results indicate that (1) significant skill is mainly located in the tropics—20 to 40% of the total land areas; (2) overall, September–October (March–May) is the most (least) skillful season; and (3) the skill weakens (with respect to the 1 month lead case) at 4 months lead—especially in June–August—although the ROCSS spatial patterns are broadly preserved—particularly in Northern South America and the Malay Archipelago. The contribution of El Niño–Southern Oscillation (ENSO) events to this 40 year skill is also analyzed, based on the idea that the seasonal predictability may be mainly driven by El Niño and La Niña precipitation teleconnections and, consequently, limited by the ability of the different seasonal forecasting models to accurately reproduce them. Results show that the ROCSS spatial patterns for (1) the full period 1961–2000 and (2) El Niño and La Niña events are highly correlated—over 0.85. Moreover, the observed teleconnection patterns are properly simulated (predicted)—with spatial correlations around 0.8—by most of the models at both 1 and 4 months lead time.

1. Introduction

Seasonal forecasting is a promising research field with enormous impact on different socioeconomic sectors such as water resources, agriculture, energy, and health [see *Doblas-Reyes et al.*, 2013, and references therein]. Nowadays, seasonal forecasts are routinely produced by several institutions around the world using different global ocean-atmosphere coupled models. Moreover, these products are collected by a number of regional focal points worldwide to produce operational consensus seasonal forecasts with socioeconomic potential; see, e.g., the Regional Climate Outlook Forum sponsored by the World Meteorological Organization. However, there are still several limiting factors which hinder the practical use of seasonal forecasts [see, e.g., *Goddard et al.*, 2010]. For instance, it is known that seasonal predictability strongly varies with the target variable, region, and season [*Halpert and Ropelewski*, 1992; *van Oldenborgh*, 2004; *Barnston et al.*, 2010; *Doblas-Reyes et al.*, 2010].

Therefore, in order to properly communicate the uncertainties related to seasonal predictions, it is needed to develop a comprehensive assessment of the performance of the different forecasting models worldwide, especially for those variables most widely used by the stakeholders and end users. In particular, precipitation is the most challenging case for being less skillfully predicted than surface temperatures [see, e.g., *Barnston et al.*, 2010; *Doblas-Reyes et al.*, 2010; *Bundel et al.*, 2011]. However, the majority of verification studies for seasonal forecasts of this variable have been conducted over limited areas of the world and for concrete seasons [see, e.g., *Batté and Déqué*, 2011; *Lim et al.*, 2011; *Kim et al.*, 2012a; *Landman and Beraki*, 2012]. A few studies have also been conducted worldwide [*van Oldenborgh et al.*, 2005; *Wang et al.*, 2009; *Barnston et al.*, 2010; *Doblas-Reyes et al.*, 2010], using a number of validation scores—correlation, ranked probability skill score (RPSS), and Brier skill score (BSS). However, the limited hindcast period available in the latter works does not ensure a robust statistical validation. For instance, *Doblas-Reyes et al.* [2010] analyzed the ENSEMBLES multimodel seasonal data set, computing averaged scores over six large-scale regions of the world for the period 1991–2005.

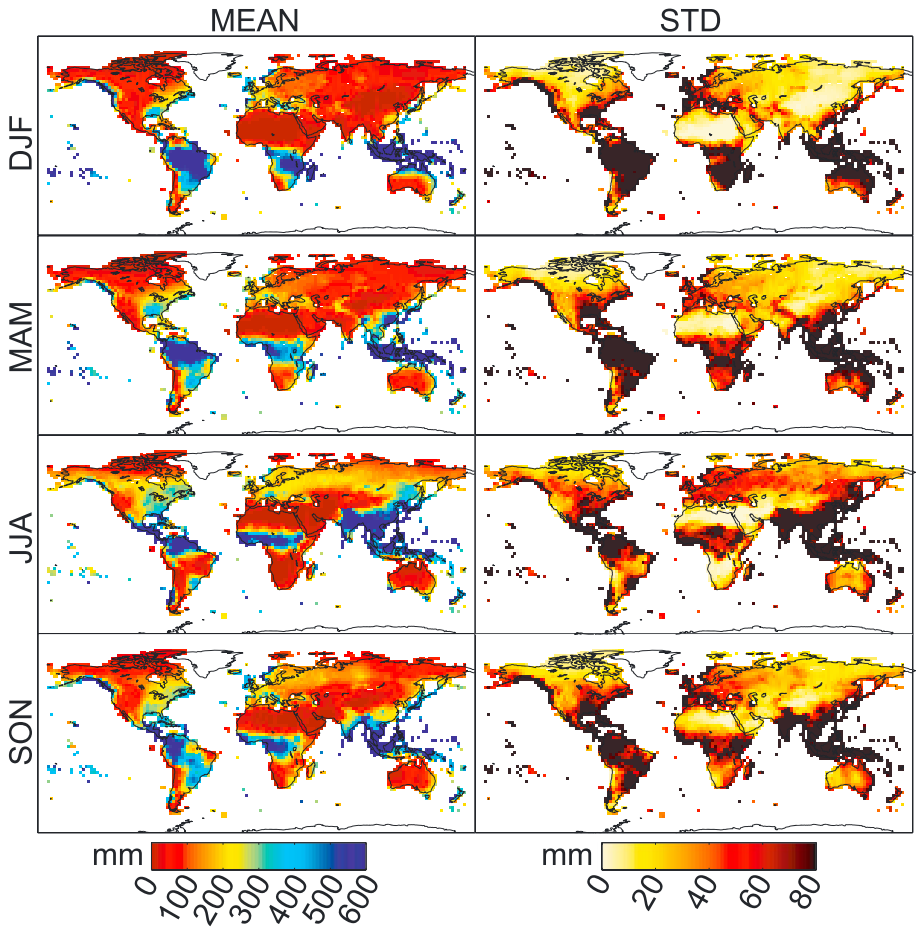

Figure 1. (left) Mean and (right) interannual standard deviation of seasonal accumulated precipitation from VASClimO v1.1 for the four seasons considered (in rows) in the period 1961–2000.

In this paper we present a global—grid point by grid point—40 year (1961–2000) validation of the ENSEMBLES multimodel seasonal hindcast—the longest-to-date available data set of retrospective forecasts—by applying a simple tercile-based probabilistic validation scheme, obtaining a simple and easy to interpret (adequate for communication with decision makers) measure of skill, the relative operating characteristic skill score (ROCSS), which is recommended by the Lead Centre for the Standardized Verification System of Long Range Forecasts (http://www.bom.gov.au/wmo/lrfvs/index.html) for the verification of probabilistic seasonal forecasts. One and four month lead predictions are considered for each of the four standard boreal seasons. Additionally, since El Niño–Southern Oscillation (ENSO) is known to be the major driving factor for seasonal predictability [*Goddard and Dilley*, 2005], we also analyze its direct (through the Sea Surface Temperature (SST) anomalies in El Niño 3.4 region) and indirect (through the associated El Niño and La Niña teleconnections with precipitation) effects in the different regions of the world, assessing to which extent each of them contributes to the skill. Thus, the two main goals of this study are (1) to fill the lack of an up-to-date user-oriented global validation of seasonal precipitation forecasts considering a long (40 year) period—identifying those regions of the world with significant skill—and (2) to analyze the role that ENSO plays on this skill.

The paper is organized as follows: The data used are described in section 2. The methodology applied is explained in section 3. Results are presented and discussed through sections 4 and 5. Finally, the main conclusions are given in section 6.

2. Data

VASClimO v1.1 [*Beck et al.*, 2005] was considered as the reference data set for validation. This gauge-based product provides monthly precipitation totals on a 2.5° resolution grid for the global land areas (except

Table 1. Main Components of the Five State-of-the-Art Atmosphere-Ocean Coupled Models Contributing to the ENSEMBLES Multimodel Seasonal Hindcast

Center	Atmospheric Model and Resolution	Ocean Model and Resolution
ECMWF	IFS CY31R1 (T159/L62)	HOPE (0.3°–1.4°/L29)
IFM-GEOMAR	ECHAM5 (T63/L31)	MPI-OM1 (1.5°/L40)
CMCC-INGV	ECHAM5 (T63/L19)	OPA8.2 (2.0°/L31)
MF	ARPEGE4.6 (T63)	OPA8.2 (2.0°/L31)
UKMO	HadGEM2-A (N96/L38)	HadGEM2-O (0.33°–1.0°/L20)

the Antarctica) for the period 1951–2000. Figure 1 shows the mean seasonal totals and the corresponding interannual standard deviation (STD) for this data set for the period of study 1961–2000.

In order to test the sensitivity to the reference data in the validation process, all calculations were also done for an alternative precipitation data set, the Global Precipitation Climatology Centre (GPCC) full data reanalysis version 6 [*Becker et al.*, 2013]. The results obtained in both cases were very similar; thus, only VASClimO v1.1 is considered hereafter.

Predictions were obtained from the longest-to-date multimodel seasonal hindcast, provided by the European project ENSEMBLES [*Weisheimer et al.*, 2009], which comprises five state-of-the-art atmosphere-ocean coupled models from the following centers: The UK Met Office (UKMO), Météo France (MF), the European Centre for Medium-Range Weather Forecasts (ECMWF), the Leibniz Institute of Marine Sciences (IFM-GEOMAR), and the Euro-Mediterranean Centre for Climate Change (CMCC-INGV). Table 1 summarizes the main components of these models.

The atmosphere and the ocean were initialized using realistic estimates of their observed states, and each model was run from an ensemble of nine initial conditions (nine equiprobable members). For each model, 7 months long runs were issued 4 times a year within the period 1960–2005, starting the first of February, May, August, and November (see *Weisheimer et al.* [2009] for more details about the experiment). Thus, the seasons considered for validation were the standard boreal winter December–February (DJF), spring March–May (MAM), summer June–August (JJA) and autumn September–November (SON), since this allows to analyze 1 and 4 months lead predictions; e.g., the initializations of August and May can be used to forecast SON. Note that although alternative 3 months' seasons could be more informative in particular regions of the world, there would be a single lead time available for them, thus limiting the study. The validation period considered was 1961–2000, common to VASClimO v1.1 observations and the ENSEMBLES models. All the models were bilinearly interpolated to the grid of the observations—similar results were obtained using the nearest grid point interpolation technique (not shown).

3. Methodology

The validation methodology used in this work is a tercile-based probabilistic approach previously applied in other studies [see, e.g., *Frías et al.*, 2010; *Vellinga et al.*, 2013]. Thus, for each particular grid point and each particular model, member, and season, the 40 year interannual series of predicted seasonal precipitation were categorized into three categories (dry, normal, and wet), according to their respective climatological terciles within the period 1961–2000. Then, a probabilistic forecast was computed year by year by considering the number of members falling within each category, out of a total of $n = 9$ members. The terciles were defined independently for each model, considering the interannual series of its nine members (a total of $40 \times 9 = 360$ values). Terciles were not computed at a member level since no significant overlap among the dry and wet terciles of the nine members was found applying a Student's t test. In the case of the multimodel (denoted hereafter as MM), $n = 45$ members were used to compute the probabilistic forecasts, thus assuming equal weights for all the models. The terciles for the MM were computed independently for each model. Note that working with precipitation categories instead of with raw values implicitly entails a bias correction grid point by grid point, which is required for a fair validation since the different models exhibit diverse season and region-dependent biases (not shown).

Rather than using deterministic scores [e.g., *van Oldenborgh et al.*, 2005; *Batté and Déqué*, 2011; *Lim et al.*, 2011; *Li et al.*, 2012; *Singh et al.*, 2012], the forecast performance is assessed in terms of the probabilistic ROCSS [see, e.g., *Kharin and Zwiers*, 2003], which is a reasonable first choice to communicate the value of a

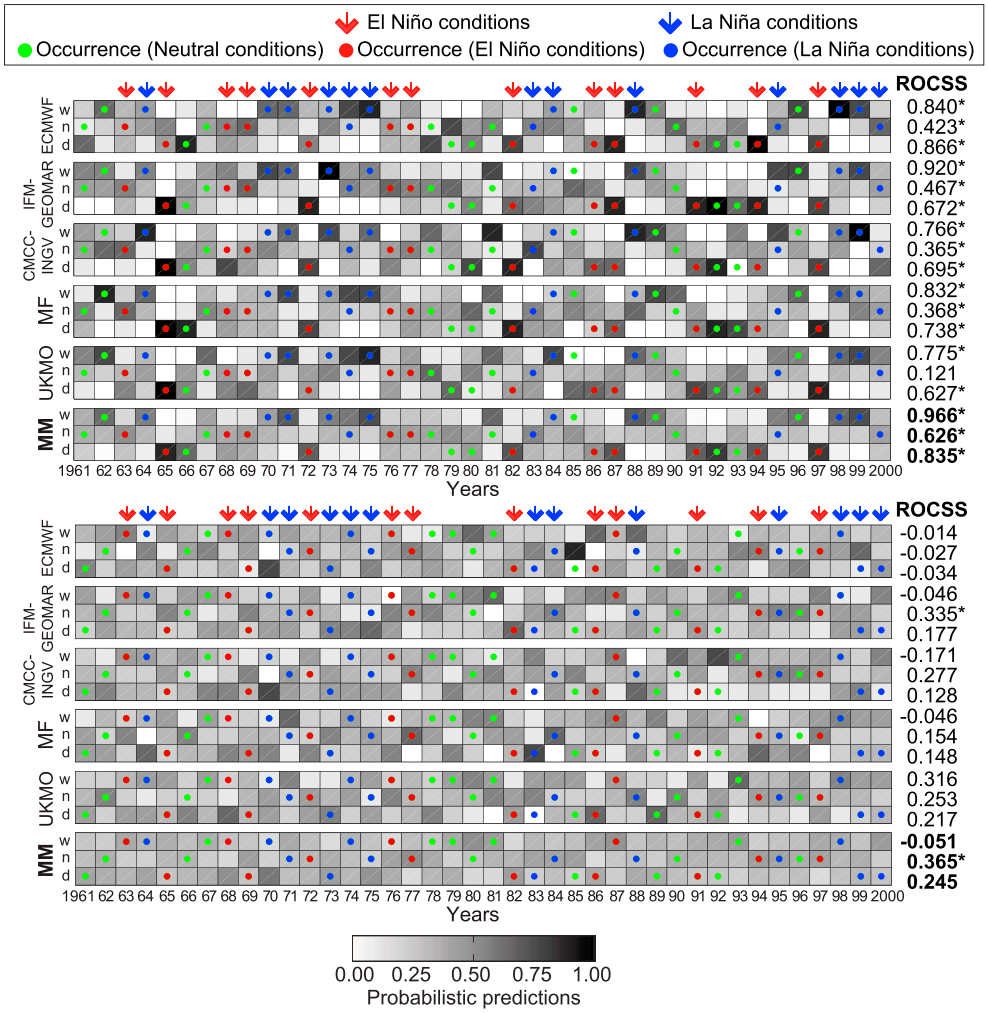

Figure 2. One month lead probabilistic predictions from the five models and the MM (in bold) for SON in an illustrative grid point in (top) the Malay Archipelago—11.25°S, 151.25°E—and (bottom) Europe—48.75°N, 16.25°E. For each tercile—*d*, *n*, and *w* stand for dry, normal, and wet, respectively. Probabilities are displayed in a white (0)-to-black (1) scale. Red/blue/green points mark the observed tercile in El Niño/La Niña/neutral years. Numbers on the right show the ROCSS for each model and each tercile. Asterisks indicate significant values (at a 0.05 level).

forecast to the end users [see, e.g., *Thiaw et al.*, 1999]. For each tercile category (e.g., dry events), the ROCSS is computed as $2A - 1$, where A is the area under the ROC curve (commonly used to evaluate the performance of probabilistic systems). The ROCSS ranges from 1 (perfect forecast system) to −1 (perfectly wrong forecast system). A value zero indicates no skill with respect to a climatological prediction. In this work, the statistical significance of the ROCSS was obtained by bootstrapping [*Mason and Graham*, 2002] with 1000 samples; i.e., by generating 1000 time series of probabilistic forecasts by randomly resampling the original 1961–2000 sequence.

As an illustrative example of the validation scheme followed, Figure 2 shows the 1961–2000 interannual time series of probabilistic predictions from the five models and the MM and the binary occurrence/non occurrence for the three terciles in two particular grid points—one in the Malay Archipelago (Figure 2, top) and the other in Europe (Figure 2, bottom)—at 1 month lead time for SON. Although varying from year to year and from model to model, predictions exhibit a higher resolution (probabilities far from 1/3) in the former point. Furthermore, resolution in this case increases, in general, in El Niño and La Niña conditions (marked with red and blue arrows, respectively) suggesting the existence of a predictability signal linked to ENSO in this region of the world for this season. Numbers on the right correspond to the ROCSS for the different models and terciles. High skill—over 0.7 in most of the cases—is found for the dry and wet terciles

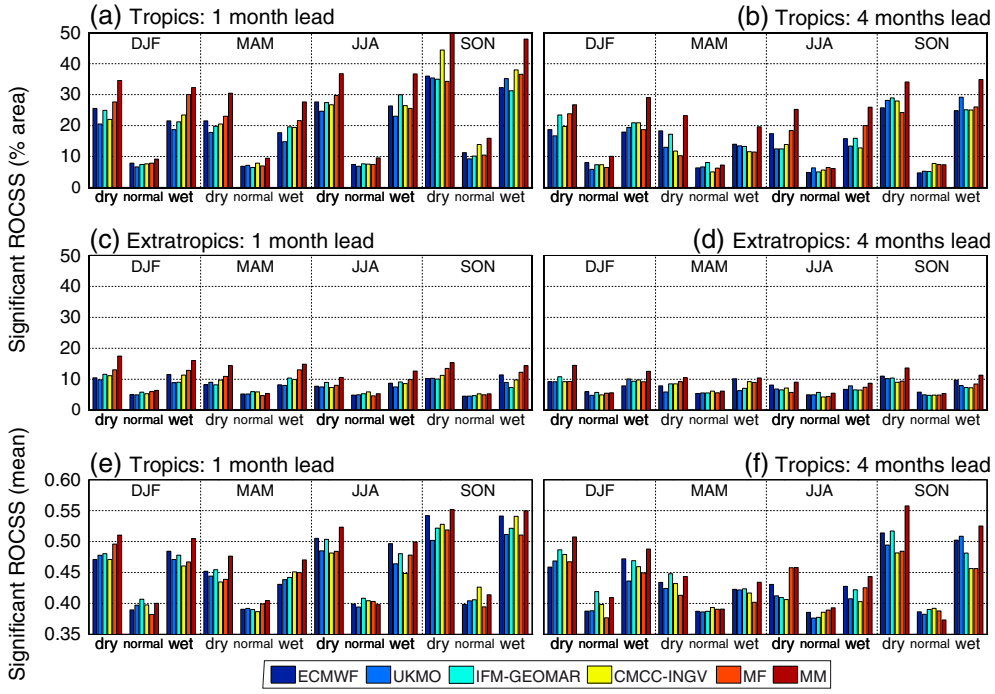

Figure 3. Percentage of land areas with significant—at a 0.05 level—ROCSS for (a, and c) 1 and (b, and d) 4 months lead predictions from the five models and the MM (see colors in legend) in (a, and b) the tropics and (c, and d) the extratropics. The mean value of the significant ROCSS in the tropics at 1 and 4 months lead time is shown in (e) and (f), respectively.

for the point in the Malay Archipelago. On the contrary, almost no skill—non significant ROCSS—is found for the point in Europe.

4. Overall Skill

The above-described methodology was applied globally—grid point by grid point—in order to compute the ROCSS (and its corresponding significance) for the five models and the MM in the period 1961–2000, thus obtaining a measure of overall skill. As a summary of the results obtained, Figures 3a–3d show the percentage of grid points with significant (at a 0.05 level) skill in the tropical (region between 23.5°N and 23.5°S latitudes) and the extratropical land areas, for both 1 and 4 months lead predictions. Although predictability varies with region, season, model, and lead time, several general conclusions can be obtained. First, the skill concentrates in the extreme (wet and dry) terciles, whereas almost no skill is obtained for normal conditions (note that the percentage of significant grid points is around 5% in this case, which can be explained by chance according to the significance level considered). This lack of skill for the near normal category is in agreement with previous studies [see, e.g., *van den Dool and Toth*, 1991]. Second, predictability is mainly located in the tropics (with 20 to 40% of total land areas showing significant skill) rather than in the extratropics (only 10%), which is also in agreement with previous studies [see, e.g., *van Oldenborgh et al.*, 2005]. Furthermore, SON (MAM) is overall the most (least) skillful season. Third, all models yield similar results for a concrete region, season, and lead time, with the MM outperforming any of the single models in all cases, which is also in agreement with previous studies [see, e.g., *Doblas-Reyes et al.*, 2009; *Bundel et al.*, 2011; *Ma et al.*, 2012]. Finally, the spatial coverage of the skillful areas decays at 4 months lead time—particularly in JJA—although not sharply. This general low decrease in skill with lead time was also found by *Barnston* [1994], who attributed it to a persistent ENSO signal. For a full interpretation of the previous results, Figures 3a and 3b should be analyzed jointly with Figures 3e and 3f, which display the mean value of the significant ROCSS in the tropics at 1 and 4 months lead time. Note that there is a clear correspondence between Figures 3e and 3f and Figures 3a and 3b, so all the previous comments apply.

In order to further analyze the above results in the different regions of the world, global spatial maps of ROCSS were obtained for all the models and the MM. For conciseness, only results for the MM are reported in the following (note its better performance). Figures 4 and 5 show the significant skill for the dry (left)

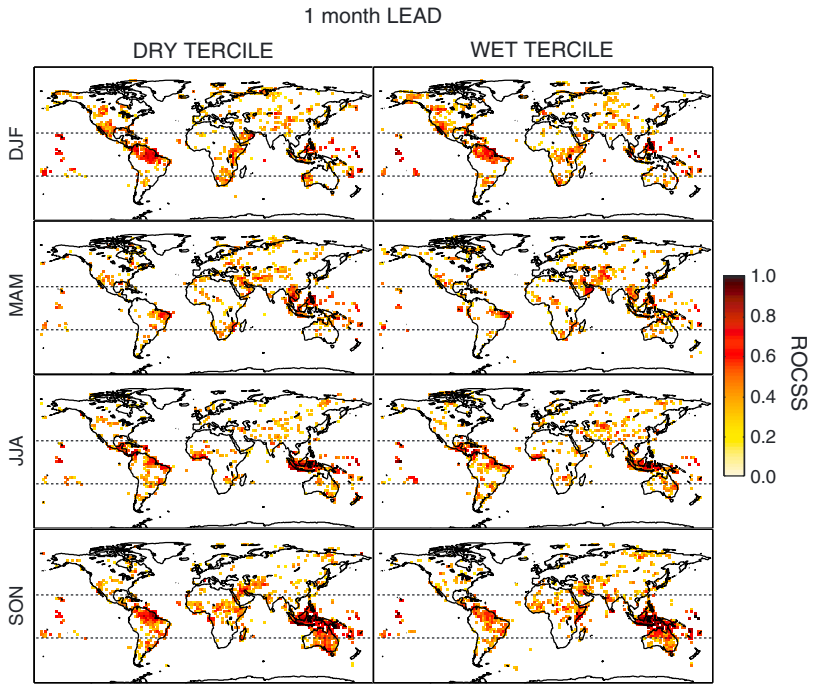

Figure 4. MM skill for the (left) dry and (right) wet terciles at 1 month lead time for the period 1961–2000, by seasons (in rows). Only significant—at a 5% level—ROCSS are shown. Dashed lines indicate the tropics/extratropics division.

and wet (right) terciles at 1 and 4 months lead time, respectively, by seasons (in rows). From these figures it can be seen that significant skill—there is clear symmetry for dry and wet terciles—is mainly located over the tropics. Furthermore, although both the signal and spatial coverage of the skillful areas slightly reduce at 4 months lead time—with respect to the 1 month lead case—the skill patterns are broadly preserved (particularly in DJF and SON).

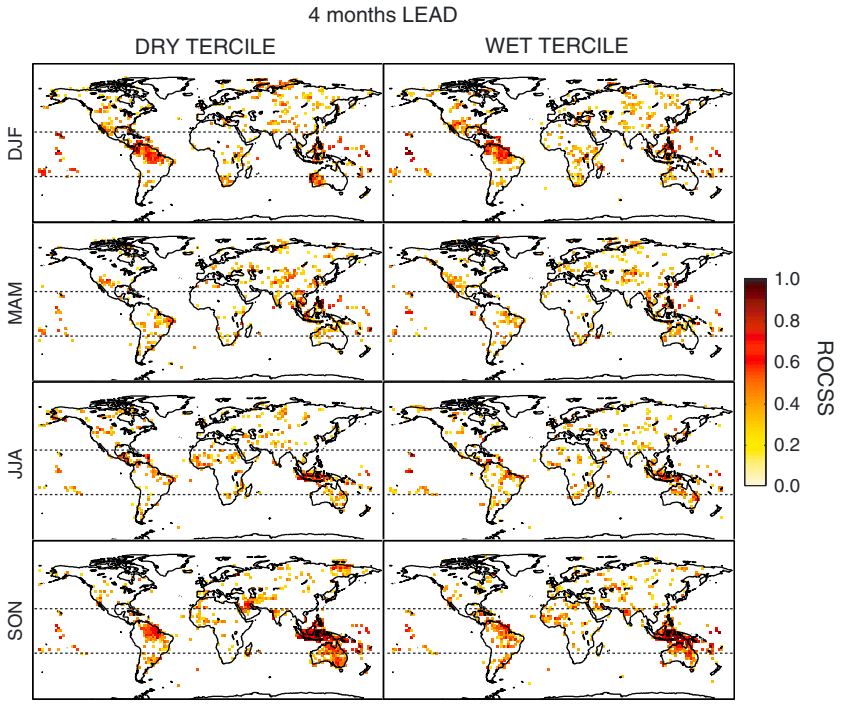

Figure 5. As Figure 4 but for the 4 months lead predictions.

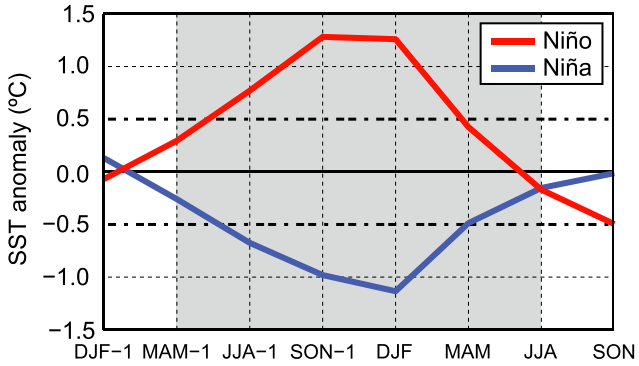

Figure 6. Mean observed SST anomaly in El Niño 3.4 region for the El Niño (red) and La Niña (blue) events considered. Observations come from the last version of the Extended Reconstructed Sea Surface Temperature data set (ERSST v3b) [*Smith et al.*, 2008].

By seasons, the main skillful regions at 1 month lead time in DJF are the Gulf of California, Northern South America, Central and Southern Africa, Western Australia, and the Pacific islands of Oceania in Melanesia, Micronesia, and Polynesia. Except in Africa, where the predictability signal weakens, most of this skill remains at 4 months lead time. In MAM, skill at 1 month lead time is located over parts of Western U.S., Northeastern Brazil, Southern Africa, parts of the Arabian Peninsula, Indochina, and the Malay Archipelago. Most of this predictability vanishes at 4 months lead time over Africa and the Arabian Peninsula. In JJA, Central America, Northern Brazil, the Gulf of Guinea, the Malay Archipelago, Eastern Australia, and the Pacific islands of Oceania are the main skillful regions at 1 month lead time. However, most of this skill is only maintained in the Malay Archipelago and the Pacific islands of Oceania at 4 months lead time. Finally, 1 month lead skill in SON is located over Northern South America, a belt in Central Africa (especially in the Somali Peninsula), parts of Middle East, the Malay Archipelago, Australia, and the Pacific islands of Oceania. Moreover, this skill remains almost unaltered at 4 months lead time for all the aforementioned regions except the Somali Peninsula, thus indicating a persisting predictability signal.

In the light of the previous results, Northern South America and the Malay Archipelago seem to be the most skillful regions of the world for seasonal forecasting of precipitation. Note that seasonal predictability in these regions has been analyzed in previous studies [*Aldrian et al.*, 2007; *Haylock and McBride*, 2001], considering also its derived socioeconomic impacts [*Kirono and Tapper*, 1999].

5. ENSO-Driven Skill

Despite the important achievements reached in seasonal forecasting in the last 10 years, significant levels of skill for precipitation are only generally found over regions connected with ENSO [see, e.g., *Coelho et al.*, 2006; *Barnston et al.*, 2010; *Arribas et al.*, 2011; *Lim et al.*, 2011; *Kim et al.*, 2012a, 2012b; *Landman and Beraki*, 2012], which is known to be the dominant mode of seasonal variability [*Doblas-Reyes et al.*, 2010]. Therefore, in this section we analyze both the direct (through the SST anomalies in El Niño 3.4 region) and indirect (through its associated atmospheric teleconnections) influences of ENSO on the skill obtained for precipitation for 1961–2000 (section 4).

5.1. SST in El Niño 3.4 Region

Although alternative indices for the definition of warm (El Niño) and cool (La Niña) ENSO events have been proposed, the Oceanic Niño Index—based on the SST anomalies in El Niño 3.4 region (5°N-5°S, 120°W-170°W)—has become the de facto standard used by the National Oceanic and Atmospheric

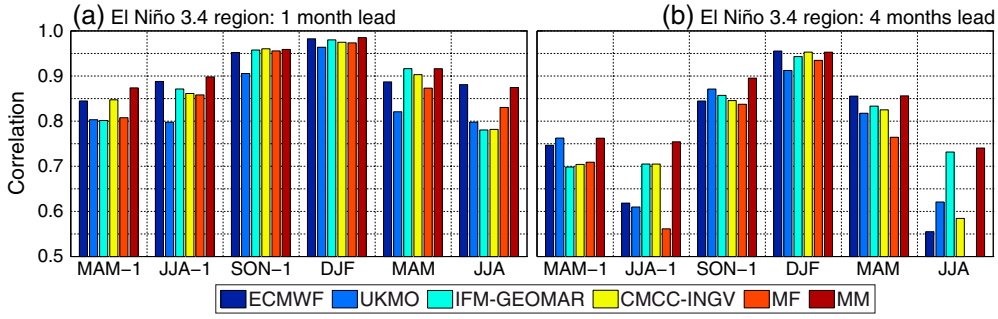

Figure 7. Correlation between the observed (ERSST v3b) and simulated SST in El Niño 3.4 region during El Niño and La Niña episodes for the five models and the MM (see colors in legend), at (a) 1 and (b) 4 months lead time.

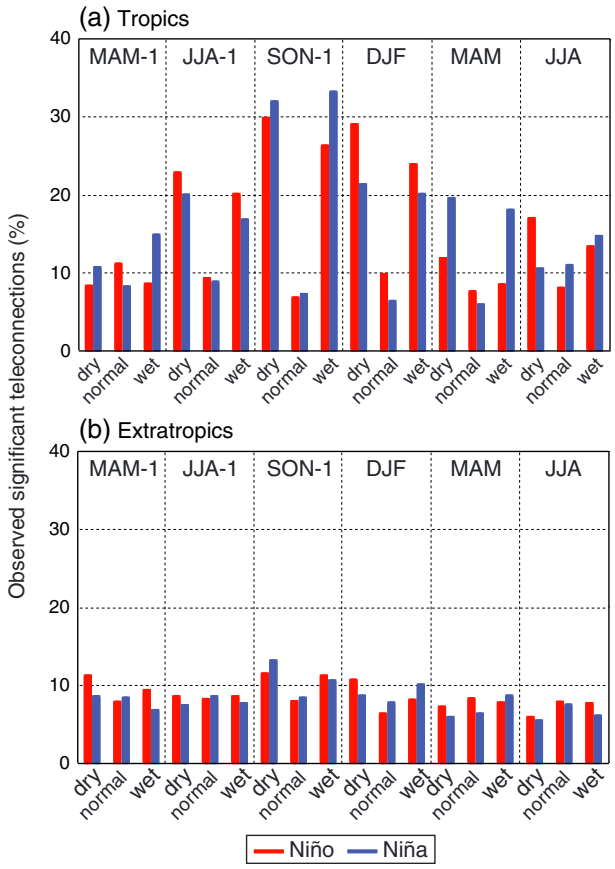

Figure 8. Percentage of areas showing significant (red) El Niño and (blue) La Niña teleconnections with precipitation in (a) the tropics and (b) the extratropics, by seasons.

Administration. According to this index, an El Niño (La Niña) event is defined when the SST anomaly in five consecutive overlapping 3 month seasons remains equal or above (equal or below) the 0.5°C (−0.5°C) threshold. In this work, we adopted this definition and analyzed the 18 month period spanning from the spring of the onset year to the summer of the decay year (as indicated by the shading in Figure 6). Note that this period is centered in SON and DJF, when the SST anomalies reach their maximum (minimum) value. The following (decay) years were considered for El Niño (La Niña) events: 1964, 1966, 1969, 1970, 1973, 1977, 1978, 1983, 1987, 1988, 1992, 1995, and 1998 (1965, 1971, 1972, 1974, 1975, 1976, 1984, 1985, 1989, 1996, 1999, 2000, and 2001). In the following, we use the notation MAM-1 and JJA-1 (MAM and JJA) to refer to the seasons of onset (decay) year of the event.

In order to assess the performance of the different seasonal forecasting models to predict ENSO, we computed the correlation between the observed (ERSST v3b) and the simulated SST in El Niño 3.4 region during the above El Niño and La Niña episodes. Figure 7 shows the results for 1 (left) and 4 (right) months lead predictions. The poorest skill is obtained for summer (both JJA-1 and JJA), when correlations decrease substantially from 1 to 4 months lead time. Note that this is in agreement wit the ENSO spring predictability barrier documented in previous studies [see, e.g., *Zheng and Zhu*, 2010; *Tippett et al.*, 2011; *Yan and Yu*, 2012; *Duan and Wei*, 2013].

However, in order to properly disentangle the role of ENSO in the overall skill found for precipitation (section 4), it is needed to analyze not only the models' ability to forecast the phenomenon itself (as characterized by the SST in El Niño 3.4 region) but also the remote effect of its associated atmospheric teleconnections, carrying (time lagged in some cases) the predictability signal to the different regions of the world.

5.2. ENSO Teleconnections

ENSO teleconnections with precipitation were calculated, following a tercile-based approach, in terms of the frequency of occurrence of each category (dry, normal, and wet) in the El Niño and La Niña events considered—as compared to the expected climatological frequency 1/3. A chi-square test for equality of proportions was applied to detect those frequencies significantly higher (lower) than 1/3, which were considered as significant positive (negative) ENSO teleconnections.

Figure 8 shows the percentage of areas exhibiting significant—at a 0.05 level—El Niño (red) and La Niña (blue) teleconnections in the tropics (Figure 8a) and the extratropics (Figure 8b) by seasons. Tropical teleconnections are stronger than extratropical ones, with different influence of El Niño and La Niña for different seasons—particularly in MAM, when La Niña has a greater effect than El Niño. Thus, in the following, we restrict the analysis to the tropics.

Figure 9 shows the maps of ENSO teleconnections in this region for the different seasons (in rows). Red/blue colors indicate high/low frequency of occurrence of the corresponding dry (left) and wet (right) terciles

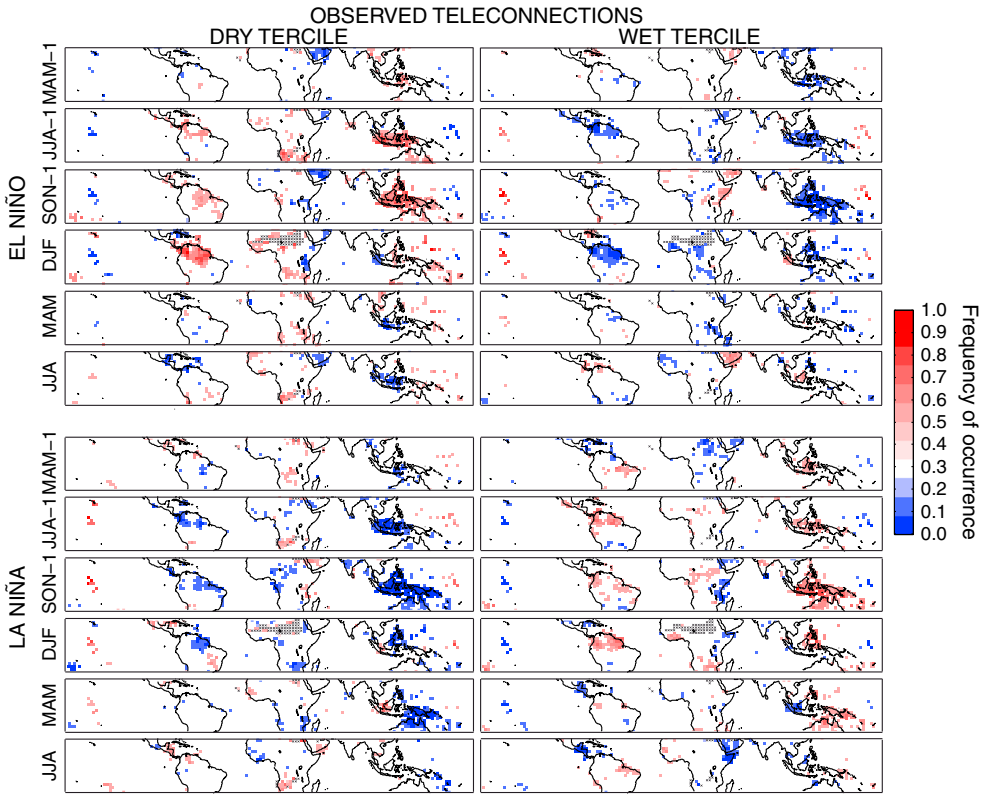

Figure 9. Relative frequency of occurrence for the (left) dry and (right) wet terciles in the tropics in (top) El Niño and (bottom) La Niña periods, for the different seasons (in rows). Red (blue) color corresponds to values above (below) 1/3—the expected climatological frequency. Only significant (at a 5% level) teleconnections, according to a chi-square test, are displayed. Black crosses indicate grid points where precipitation categories cannot be properly defined (series with less than three different values).

during El Niño (Figure 9, top) and La Niña (Figure 9, bottom) events. Notice that El Niño and La Niña—as opposite phases of the same underlying phenomenon—tend to yield similar patterns but with opposite signals, although there are some exceptions; e.g., the Malay Archipelago and Northern Australia are teleconnected in MAM with La Niña but not with El Niño. Overall, the results are in agreement with previous studies [see, e.g., *Ropelewski and Halpert*, 1987; *van Oldenborgh et al.*, 2000; *Kayano et al.*, 2009; *Shaman and Tziperman*, 2011; *Zhang et al.*, 2012; *Yadav et al.*, 2013; *Zhang et al.*, 2013].

Comparison of Figure 9 with Figures 4 and 5 reveals that, in general, the skillful zones are significantly teleconnected with ENSO; e.g., Northern South America in DJF and SON and the Malay Archipelago in JJA and SON. This suggests that the seasonality and the spatial distribution of the overall skill could be explained by this phenomenon through its associated teleconnections. Therefore, it is important to assess the ability of the models to properly reproduce El Niño and La Niña precipitation teleconnections. Figure 10 shows the spatial correlation between the observed and simulated El Niño and La Niña teleconnections patterns (as given by the maps of terciles frequencies) considering 1 (left) and 4 months (right) lead predictions from the five single models and the MM (see colors in legend), over their corresponding skillful regions within the tropics. Note that although a similar analysis has been done in *Yang and Delsole* [2012], who compared observed and simulated ENSO teleconnections using a field regression analysis, they did not assess the predictive skill of the teleconnections found, which is the aim here. As can be seen, the agreement is good for both lead times, with correlations over 0.8 in some seasons. On the one hand, the observed patterns (for both El Niño and La Niña) are best reproduced in SON-1 and DJF, the central seasons of the phenomenon. On the other hand, they are worst reproduced in MAM-1 in all cases.

Interestingly, note that in spite of the aforementioned spring barrier for ENSO (Figure 7), the models exhibit a relative good performance in reproducing the existing—both El Niño and La Niña—teleconnections in summer (particularly in JJA-1), which could explain the overall skill found for precipitation in this season (Figure 3).

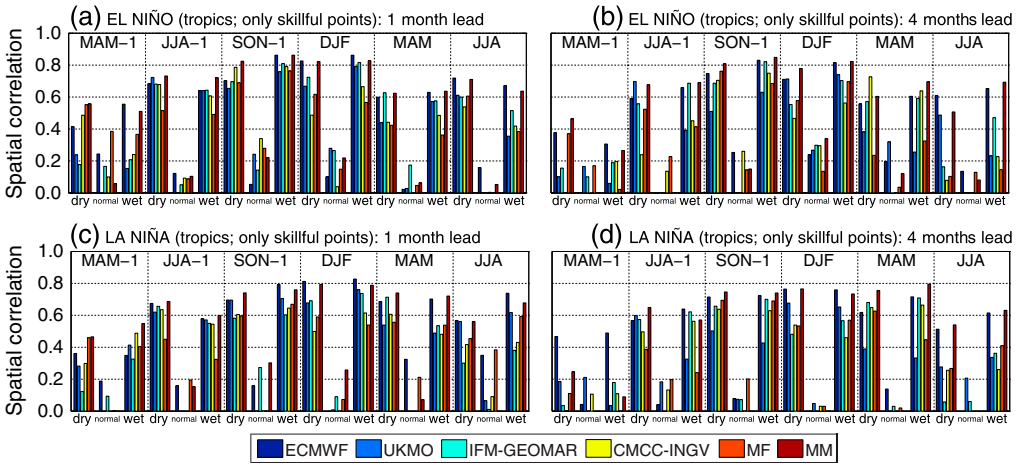

Figure 10. Spatial correlation between the observed and predicted (a and b) El Niño and (c and d) La Niña teleconnection patterns for the tropical regions showing significant skill at (left) 1 and (right) 4 months lead time, for the five models and the MM (see colors in legend).

5.3. Contribution of ENSO to the Overall Skill

All the previous results suggest the idea that the seasonality and the spatial distribution of the overall (1961–2000) skill found in section 4 may be mainly driven by the indirect effect of ENSO through its associated atmospheric teleconnections and therefore limited by the ability of the different models to reproduce the observed El Niño and La Niña teleconnections with precipitation. Thus, to further assess the role of ENSO on the global skill for each particular model (and the MM) we computed the spatial correlation between the corresponding ROCSS maps for (1) the full period 1961–2000—shown in Figures 4 and 5 for the MM—and (2) the 26 El Niño and La Niña events considered. Figure 11 shows the results obtained for 1 (left) and 4 (right) months lead predictions in the tropics for JJA-1, SON-1, DJF, and MAM, the seasons with the strongest teleconnections (Figure 8).

Correlations are very high (over 0.85) in most of the seasons at both 1 and 4 months lead, what confirms that the overall (1961–2000) skill attained in this region may be mainly explained by the contribution of El Niño and La Niña years.

6. Conclusions

The skill of seasonal precipitation forecasts has been assessed worldwide—grid point by grid point—for the 40 year period 1961–2000, considering the ENSEMBLES multimodel hindcast and applying a tercile-based probabilistic approach in terms of the ROC skill score (ROCSS). Although predictability varies with region, season, and lead time, results indicate that (1) significant skill is mainly located in the tropics—20 to 40% of the total land areas; (2) overall, SON (MAM) is the most (least) skillful season; and 3) the skill weakens

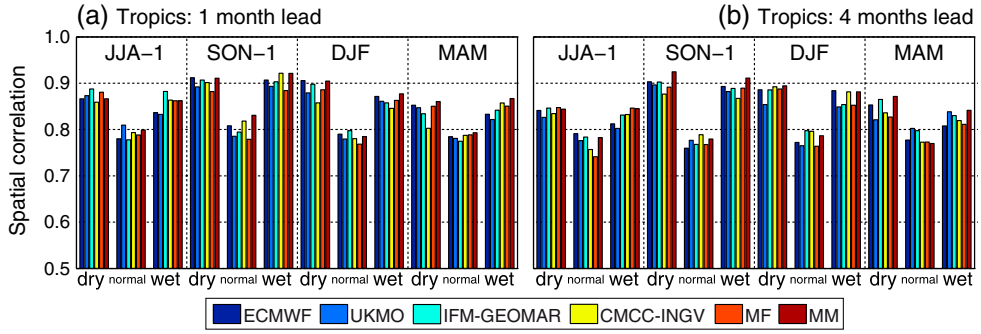

Figure 11. Spatial correlation between the ROCSS maps for the 40 year period (1961–2000) and for the 26 El Niño and La Niña events at (a) 1 and (b) 4 months lead for JJA-1, SON-1, DJF, and MAM (see Figure 6 for further details on the definition of seasons). Results are shown for each model and the MM (see colors in legend).

(with respect to the 1 month lead case) at 4 months lead—especially in JJA, —although the ROCSS spatial patterns are broadly preserved—particularly in Northern South America and the Malay Archipelago.

Results from a conditioned—restricted to El Niño and La Niña events—validation (also in terms of the ROCSS) and a study of El Niño and La Niña teleconnections with precipitation suggest that the seasonality and the spatial distribution of the overall (1961–2000) skill found for this variable may be not only determined by the direct effect of ENSO and therefore by the skill of the different forecasting systems to predict the SST in El Niño 3.4 region but rather by its indirect effect through its associated El Niño and La Niña teleconnections and consequently limited by the models' ability to accurately simulate (predict) the observed teleconnection patterns. In addition, it is found that whereas the overall skill in DJF and SON—the peak seasons of the phenomenon—might be uniquely related to El Niño and La Niña teleconnections in the corresponding current season, both the onset and decay seasons contribute to the overall skill in MAM and JJA.

For instance, high overall skill is found for JJA, which seems to be in contradiction with the well-known ENSO spring predictability barrier; in particular, predictability at 1 month lead is higher in JJA than in DJF despite the SST in El Niño 3.4 region is better predicted in the latter season. However, this could be explained by the models' ability found to reproduce the observed precipitation teleconnection patterns in both the onset and the decay (especially in the former) summers, which are found to exhibit significant teleconnections. Finally, and as another illustrative example of the results found, even though the SST in El Niño 3.4 region is better predicted in DJF than in SON (especially at 4 months lead) and the different models are shown to reproduce similarly the observed teleconnection patterns in both seasons, the overall skill is higher in SON than in DJF. One explanation for this could be that the observed teleconnections—especially La Niña ones—are stronger in the former season.

Acknowledgments

This study was supported by the EU projects QWeCI and SPECS, funded by the European Commission through the Seventh Framework Programme for Research under grant agreements 243964 and 308378, respectively. The authors want also to acknowledge the EU project ENSEMBLES, financed by the European Commission through the Sixth Framework Programme for Research under contract GOCE-CT-2003-505539, for the data provided, which were retrieved from the Meteorological Archival and Retrieval System (MARS) of the ECMWF, http://www.ecmwf.int/services/archive/. RM also acknowledges the EU project EUPORIAS, funded by the European Commission through the Seventh Framework Programme for Research under grant agreement 308291. Finally, the authors would like to thank the two anonymous reviewers for their useful comments which helped to improve the manuscript substantially.

References

Aldrian, E., L. Dümenil Gates, and F. H. Widodo (2007), Seasonal variability of Indonesian rainfall in ECHAM4 simulations and in the reanalyses: The role of ENSO, *Theor. Appl. Climatol.*, *87*(1–4), 41–59, doi:10.1007/s00704-006-0218-8.

Arribas, A., et al. (2011), The GloSea4 ensemble prediction system for seasonal forecasting, *Mon. Weather Rev.*, *139*(6), 1891–1910, doi:10.1175/2010MWR3615.1.

Barnston, A. G. (1994), Linear statistical short-term climate predictive skill in the Northern Hemisphere, *J. Clim.*, *7*(10), 1513–1564, doi:10.1175/1520-0442(1994)007<1513:LSSTCP>2.0.CO;2.

Barnston, A. G., S. Li, S. J. Mason, D. G. Dewitt, L. Goddard, and X. Gong (2010), Verification of the first 11 years of IRI's seasonal climate forecasts, *J. Appl. Meteorol. Climatol.*, *49*(3), 493–520, doi:10.1175/2009JAMC2325.1.

Batté, L., and M. Déqué (2011), Seasonal predictions of precipitation over Africa using coupled ocean-atmosphere general circulation models: Skill of the ENSEMBLES project multimodel ensemble forecasts, *Tellus A*, *63*(2), 283–299, doi:10.1111/j.1600-0870.2010.00493.x.

Beck, C., J. Grieser, and B. Rudolf (2005), A new monthly precipitation climatology for the global land areas for the period 1951 to 2000, *Tech. Rep.*, Klimastatusbericht.

Becker, A., P. Finger, A. Meyer-Christoffer, B. Rudolf, K. Schamm, U. Schneider, and M. Ziese (2013), A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present, *Earth Syst. Sci. Data*, *5*(1), 71–99, doi:10.5194/essd-5-71-2013.

Bundel, A. Y., V. N. Kryzhov, Y. M. Min, V. M. Khan, R. M. Vilfand, and V. A. Tishchenko (2011), Assessment of probability multimodel seasonal forecast based on the APCC model data, *Russ. Meteorol. Hydrol.*, *36*(3), 145–154, doi:10.3103/S1068373911030010.

Coelho, C. A. S., D. B. Stephenson, F. J. Doblas-Reyes, and M. Balmaseda (2006), The skill of empirical and combined/calibrated coupled multi-model South American seasonal predictions during ENSO, *Adv. Geosci.*, *6*, 51–55, doi:10.5194/adgeo-6-51-2006.

Doblas-Reyes, F. J., A. Weisheimer, A. Déqué, N. Keenlyside, M. MacVean, J. M. Murphy, P. Rogel, D. Smith, and T. N. Palmer (2009), Addressing model uncertainty in seasonal and annual dynamical ensemble forecasts, *Q. J. R. Meteorolog. Soc.*, *135*(643), 1538–1559, doi:10.1002/qj.464.

Doblas-Reyes, F. J., J. García-Serrano, F. Lienert, A. P. Biescas, and L. R. L. Rodrigues (2013), Seasonal climate predictability and forecasting: Status and prospects, *Wiley Interdiscip. Rev. Clim. Change*, *4*(4), 245–268, doi:10.1002/wcc.217.

Doblas-Reyes, F. J., A. Weisheimer, T. N. Palmer, J. M. Murphy, and D. Smith (2010), Forecast quality assessment of the ENSEMBLES seasonal-to-decadal Stream 2 hindcasts, *Tech. Rep. 621*, European Centre for Medium-Range Weather Forecasts (ECMWF).

Duan, W., and C. Wei (2013), The 'spring predictability barrier' for ENSO predictions and its possible mechanism: Results from a fully coupled model, *Int. J. Climatol.*, *33*(5), 1280–1292, doi:10.1002/joc.3513.

Frías, M. D., S. Herrera, A. S. Cofiño, and J. M. Gutiérrez (2010), Assessing the skill of precipitation and temperature seasonal forecasts in Spain. Windows of opportunity related to ENSO events, *J. Clim.*, *23*(2), 209–212, doi:10.1175/2009JCLI2824.1.

Goddard, L., and M. Dilley (2005), El Niño: Catastrophe or opportunity, *J. Clim.*, *18*(5), 651–665, doi:10.1175/JCLI-3277.1.

Goddard, L., Y. Aitchellouche, W. Baethgen, M. Dettinger, R. Graham, P. Hayman, M. Kadi, R. Martínez, and H. Meinke (2010), Providing seasonal-to-interannual climate information for risk management and decision-making, *Procedia Environ. Sci.*, *1*, 81–101, doi:10.1016/j.proenv.2010.09.007.

Halpert, M. S., and C. F. Ropelewski (1992), Surface temperature patterns associated with the Southern Oscillation, *J. Clim.*, *5*(6), 577–593, doi:10.1175/1520-0442(1992)005<0577:STPAWT>2.0.CO;2.

Haylock, M., and J. McBride (2001), Spatial coherence and predictability of Indonesian wet season rainfall, *J. Clim.*, *14*(18), 3882–3887.

Kayano, M. T., C. P. de Oliveira, and R. V. Andreoli (2009), Interannual relations between South American rainfall and tropical sea surface temperature anomalies before and after 1976, *Int. J. Climatol.*, *29*(10), 1439–1448.

Kharin, V. V., and F. W. Zwiers (2003), On the ROC score of probability forecasts, *J. Clim.*, *16*(24), 4145–4150, doi:10.1175/1520-0442(2003)016<4145:OTRSOP>2.0.CO;2.
Kim, H. M., P. J. Webster, and J. A. Curry (2012a), Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere winter, *Clim. Dyn.*, *39*(12), 2957–2973, doi:10.1007/s00382-012-1364-6.
Kim, H. M., P. J. Webster, J. A. Curry, and V. E. Toma (2012b), Asian summer monsoon prediction in ECMWF System 4 and NCEP CFSv2 retrospective seasonal forecasts, *Clim. Dyn.*, *39*(12), 2975–2991, doi:10.1007/s00382-012-1470-5.
Kirono, D. G. C., and N. J. Tapper (1999), ENSO rainfall variability and impacts on crop production in Indonesia, *Phys. Geogr.*, *20*(6), 508–519, doi:10.1080/027236461999.10642693.
Landman, W. A., and A. Beraki (2012), Multi-model forecast skill for mid-summer rainfall over southern Africa, *Int. J. Climatol.*, *32*(2), 303–314, doi:10.1002/joc.2273.
Li, C., R. Lu, and B. Dong (2012), Predictability of the western North Pacific summer climate demonstrated by the coupled models of ENSEMBLES, *Clim. Dyn.*, *39*(1–2), 329–346, doi:10.1007/s00382-011-1274-z.
Lim, E. P., H. H. Hendon, D. L. T. Anderson, A. Charles, and O. Alves (2011), Dynamical, statistical-dynamical, and multimodel ensemble forecasts of Australian spring season rainfall, *Mon. Weather Rev.*, *139*(3), 958–975, doi:10.1175/2010MWR3399.1.
Ma, S., X. Rodó, and F. J. Doblas-Reyes (2012), Evaluation of the DEMETER performance for seasonal hindcasts of the Indian summer monsoon rainfall, *Int. J. Climatol.*, *32*(11), 1717–1729, doi:10.1002/joc.2389.
Mason, S. J., and N. E. Graham (2002), Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation, *Q. J. R. Meteorol. Soc.*, *128*, 2145–2166, doi:10.1256/003590002320603584.
Ropelewski, C. F., and M. S. Halpert (1987), Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation, *Mon. Weather Rev.*, *115*(8), 1606–1626, doi:10.1175/1520-0493.
Shaman, J., and E. Tziperman (2011), An atmospheric teleconnection linking ENSO and Southwestern European precipitation, *J. Clim.*, *24*(1), 124–139, doi:10.1175/2010JCLI3590.1.
Singh, A., N. Acharya, U. C. Mohanty, A. W. Robertson, and G. Mishra (2012), On the predictability of Indian summer monsoon rainfall in general circulation model at different lead time, *Dyn. Atmos. Oceans*, *58*, 108–127, doi:10.1016/j.dynatmoce.2012.09.004.
Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore (2008), Improvements to NOAA's historical merged land–ocean surface temperature analysis (1880–2006), *J. Clim.*, *21*(10), 2283–2296, doi:10.1175/2007JCLI2100.1.
Thiaw, W. M., A. G. Barnston, and V. Kumar (1999), Predictions of African rainfall on the seasonal timescale, *J. Geophys. Res.*, *104*(D24), 31,589–31,597, doi:10.1029/1999JD900906.
Tippett, M. K., A. G. Barnston, and S. Li (2011), Performance of recent multimodel ENSO forecasts, *J. Appl. Meteorol. Climatol.*, *51*(3), 637–654, doi:10.1175/JAMC-D-11-093.1.
van den Dool, H. M., and Z. Toth (1991), Why do forecasts for 'near normal' often fail?, *Weather Forecasting*, *6*(1), 76–85, doi:10.1175/1520-0434(1991)006<0076:WDFFNO>2.0.CO;2.
van Oldenborgh, G. J. (2004), Assessing the skill of seasonal forecasts, *Tech. Rep.*, KNMI Research Biennial Reports.
van Oldenborgh, G. J., G. Burgers, and A. K. Tank (2000), On the El Niño teleconnection to spring precipitation in Europe, *Int. J. Climatol.*, *20*(5), 565–574, doi:10.1002/(SICI)1097-0088(200004)20:5<565::AID-JOC488>3.0.CO;2-5.
van Oldenborgh, G. J., M. A. Balmaseda, L. Ferranti, T. N. Stockdale, and D. L. T. Anderson (2005), Evaluation of atmospheric fields from the ECMWF seasonal forecasts over a 15 year period, *J. Clim.*, *18*(16), 3250–3269, doi:10.1175/JCLI3421.1.
Vellinga, M., A. Arribas, and R. Graham (2013), Seasonal forecasts for regional onset of the West African monsoon, *Clim. Dyn.*, *40*(11–12), 3047–3070, doi:10.1007/s00382-012-1520-z.
Wang, B., et al. (2009), Advance and prospectus of seasonal prediction: Assessment of the APCC/ CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004), *Clim. Dyn.*, *33*(1), 93–117, doi:10.1007/s00382-008-0460-0.
Weisheimer, A., F. J. Doblas-Reyes, T. N. Palmer, A. Alessandri, A. Arribas, M. Déqué, N. Keenlyside, M. MacVean, A. Navarra, and P. Rogel (2009), ENSEMBLES: A new multi-model ensemble for seasonal-to-annual prediction. Skill and progress beyond DEMETER in forecasting tropical pacific SSTs, *Geophys. Res. Lett.*, *36*, L21711, doi:10.1029/2009GL040896.
Yadav, R. K., D. A. Ramu, and A. P. Dimri (2013), On the relationship between ENSO patterns and winter precipitation over North and Central India, *Global Planet. Change*, *107*, 50–58, doi:10.1016/j.gloplacha.2013.04.006.
Yan, L., and Y. Yu (2012), The spring prediction barrier in ENSO hindcast experiments using the FGOALS–g model, *Chin. J. Oceanol. Limnol.*, *30*(6), 1093–1104, doi:10.1007/s00343-012-1271-z.
Yang, X., and T. Delsole (2012), Systematic comparison of ENSO teleconnection patterns between models and observations, *J. Clim.*, *25*(2), 425–446, doi:10.1175/JCLI-D-11-00175.1.
Zhang, Q., J. Li, V. P. Singh, C. Y. Xu, and J. Deng (2013), Influence of ENSO on precipitation in the East River basin, South China, *J. Geophys. Res. Atmos.*, *118*, 2207–2219, doi:10.1002/jgrd.50279.
Zhang, Y., Y. Qian, V. Dulière, E. P. Salathé Jr., and L. R. Leung (2012), ENSO anomalies over the Western United States: Present and future patterns in regional climate simulations, *Clim. Change*, *110*(1–2), 315–346, doi:10.1007/s10584-011-0088-7.
Zheng, F., and J. Zhu (2010), Spring predictability barrier of ENSO events from the perspective of an ensemble prediction system, *Global Planet. Change*, *72*(3), 108–117, doi:10.1016/j.gloplacha.2010.01.021.