Tellus
SERIES A
DYNAMIC
METEOROLOGY
AND OCEANOGRAPHY
PUBLISHED BY THE INTERNATIONAL METEOROLOGICAL INSTITUTE IN STOCKHOLM

Tellus (2011), 63A, 283–299

TELLUS

Seasonal predictions of precipitation over Africa using coupled ocean-atmosphere general circulation models: skill of the ENSEMBLES project multimodel ensemble forecasts

By L. BATTÉ* and M. DÉQUÉ, *Météo France, CNRM-GAME, 42 avenue G. Coriolis, FR-31057 Toulouse Cedex, France*

(Manuscript received 22 June 2010; in final form 28 October 2010)

ABSTRACT

ENSEMBLES stream 2 single-model and multimodel ensemble seasonal precipitation forecasts are evaluated over the African continent with respect to Global Precipitation Climatology Centre (GPCC) precipitation data for the 1960–2005 time period using deterministic and probabilistic skill scores. Focus is set on three regions of Africa during main precipitation seasons: West Africa during boreal summer, southern Africa during austral summer and the Greater Horn of Africa during the 'long rains' and 'short rains' transition seasons. The 45-member multimodel improves the ensemble spread-skill ratio over all regions, which translates into enhanced skill in terms of anomaly correlation and ranked probability skill scores for climatological precipitation deciles over West Africa and southern Africa. Results are contrasted depending on the region and probabilistic formulations of the ensemble predictions after a quantile–quantile calibration give valuable information essentially over areas where deterministic skill is found. Probabilistic skill scores illustrate the range of possibilities for more user-related applications of ensemble seasonal forecasts. A simple illustration using a cost-loss model shows that model potential economic values can reach over 10% depending on the regions and occurrences studied.

1. Introduction

Multimodel ensemble forecasting is now a mainstream method used for medium-range to seasonal predictions as well as climate modelling. This method addresses two sources of uncertainty in dynamical models: ensemble forecasting accounts for uncertainties due to initial conditions while the use of several models with different physical formulations and parameterizations helps widening the ensemble spread and estimating single-model uncertainties (Palmer et al., 2004).

Seasonal-to-annual predictions from five state-of-the-art European global coupled climate models were run as part of the EU-funded ENSEMBLES project. Recent assessment results for different variables (sea-surface temperature, near-surface temperature, 500 hPa geopotential height, precipitation) in Doblas-Reyes et al. (2009) show that multimodel ensemble seasonal predictions (at least up to 4 month lead) have higher skill than other methods used to reduce model uncertainties such as perturbed-parameter (Collins et al., 2006) or stochastic-physics (Berner et al., 2008) approaches. These alternative methods will therefore not be discussed in this paper.

The focus in this paper is set on seasonal precipitation forecasts over the African continent, and more specifically over subregions with noteworthy seasonal precipitation cycles. These precipitation forecasts are of interest in regions subject to persisting drought if used as early warning and decision-making tools in water supply management. At a seasonal time scale, the chaotic nature of atmospheric dynamics make it difficult for GCM precipitation forecasts to provide chronological information and the long-term predictability relies mostly on ocean conditions (Shukla, 1998). As such, GCM precipitation outputs are not used for prediction of monsoon onset dates without further analysis of other fields, such as circulation patterns. Quantitative precipitation seasonal forecasts show little skill when compared to persistence or climatology and extracting useful information from such forecasts often requires elaborate post-processing methods such as statistical adaptation (Garric et al., 2002; Bouali et al., 2008). These research efforts were led as a consequence of poor seasonal precipitation forecasting skills in past ensemble forecast experiments. Another way of

*Corresponding author.
e-mail: lauriane.batte@meteo.fr
DOI: 10.1111/j.1600-0870.2010.00493.x

dealing with ensemble forecast information is to use a probabilistic formulation. This approach avoids quantitative forecast issues by yielding probabilities for different possible outcomes of a given event (Déqué and Royer, 1992). The evaluation of such forecasts is therefore more user-related, in the sense that some probabilistic scores can be highly dependent on the model actual usefulness in a given situation. Evaluation of the ENSEMBLES stream 1 (1990–2001) seasonal forecasts skill over West Africa during the monsoon season showed that direct model precipitation outputs can be improved using model-output statistics based on simulated atmospheric dynamics (Philippon et al., 2010). Our goal here is to evaluate seasonal forecast skill for the ENSEMBLES stream 2 monthly ensemble re-forecasts over the extended 1960–2005 period using various deterministic and probabilistic scores, focusing on the precipitation outputs of the models. Additional analysis of precipitation forecast skill related to sea-surface temperatures (SSTs) is also provided.

This paper is organized as follows. The ENSEMBLES stream 2 forecasts and comparison data sets are described in the following section, along with precipitation climatologies over the subregions studied in this paper. Section 3 details the methodology and skill scores calculated throughout our study. Main results are presented in Section 4 and Section 5 summarizes general conclusions.

2. Data and regions of interest

2.1. ENSEMBLES project seasonal re-forecasts

The ENSEMBLES seasonal re-forecasts used in this study are from the Stream 2 experiment. In this experiment five coupled global circulation ensemble models were run. Participitating research centers are the European Centre for Medium-Range Weather Forecasts (ECMWF), the Leibniz Institute of Marine Sciences at Kiel University (IFM-GEOMAR), Météo France (MF), the UK Met Office (UKMO) and the Euro-Mediterranean Centre for Climate Change in Bologna (CMCC-INGV). Table 1 shows basic characteristics of each model. Each ensemble comprises nine runs initialized with different sets of ocean reanalyses generated from wind stress and SST perturbations. The full multimodel ensemble (MME) is therefore made of 45 members. Further details may be found in Weisheimer et al. (2009).

Stream 2 seasonal re-forecasts were run from February, May, August and November over the 1960–2005 period. Seasonal forecasts for each month 2–4 average (March–April–May or MAM, June–July–August or JJA, September–October–November or SON and December–January–February or DJF) were used in this study. The multimodel ensemble combination method used for deterministic forecasts is a simple 45-ensemble mean, without individual model weighting.

Data were retrieved from a ECMWF server (http://www.ecmwf.int/research/EU_projects/ENSEMBLES/data/data_dissemination.html). The precipitation fields are available for each model on a global $2.5^\circ \times 2.5^\circ$ grid.

2.2. Reference data: the GPCC monthly global data set

The Global Precipitation Climatology Centre (GPCC) Full Data Reanalysis version 4 data set for monthly mean precipation was used as the reference data set for climatology and verification of the ENSEMBLES stream 2 hindcasts. This data set, described in Schneider et al. (2008), is a reanalysis of monthly rain-gauge data from up to 45000 land stations around the globe using GPCC climatology as analysis background and covers the 1901–2007 period. It is the most appropriate GPCC product for hydrometeorological model verification. The choice of GPCC data was also based on station availability over the regions of interest and during the entire ENSEMBLES stream 2 period (1960–2005). The $1^\circ \times 1^\circ$ gridded data was used, so the ENSEMBLES predictions were regridded onto the corresponding grid for comparison purposes.

Table 1. Basic characteristics of the main components of the five ENSEMBLES stream 2 models used for seasonal forecasting

Centre	Atmosphere component	Ocean component
ECMWF	IFS CY31R1	HOPE
	T159/L62	0.3°–1.4°/L62
IFM-Geomar	ECHAM5	MPI-OMI
	T63/L31	1.5°/L40
Météo France	ARPEGE4.6	OPA8.2
	T63/L31	2°/L31
UKMO	HadGEM2-A	HadGEM2-O
	N96/L38	$0.33^\circ - 1^\circ$
CMCC-INGV	ECHAM5	OPA8.2
	T63/L19	2°/L31

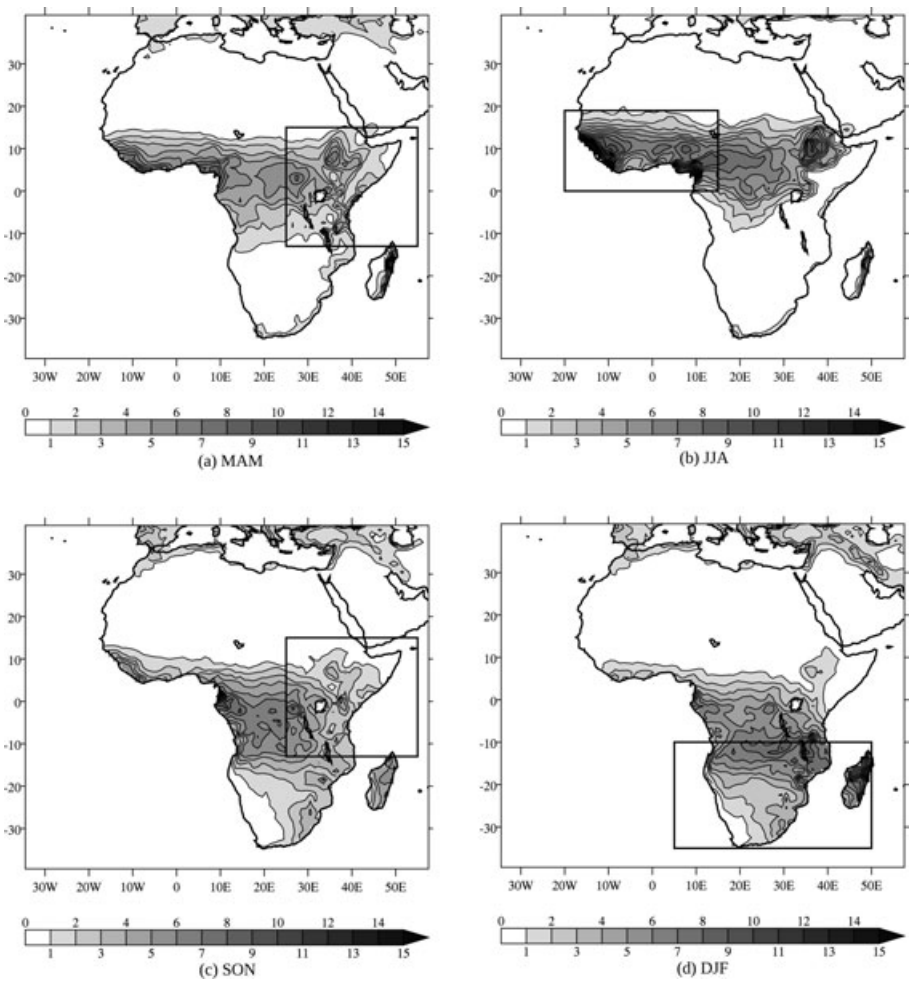


Fig. 1. GPCC seasonal mean precipitation (mm per day) over the African continent for four three-month seasons, averaged over the 1960–2005 period. Boxes show the regions of interest for the given seasons.

2.3. African subregions

Subregions of the African continent were defined so as to focus on areas subjected to significant seasonal precipitation amounts. Figure 1 shows GPCC mean precipitation from 1960 to 2005 over the African continent for the four three-month seasons defined earlier. The impact of SST over the ENSO region and the tropical Atlantic on African rainfall is thoroughly discussed in Camberlin et al. (2001). Other teleconnections between East-African rainfall and Indian Ocean SST have also been demonstrated (Bowden and Semazzi, 2007).

The quantitatively most important precipitation event over the African continent is the West African Monsoon. High mean daily precipitation rates are found for GPCC data during the JJA season along latitudes ranging from approximately 2° to 15° North, with a peak locally exceeding 16 mm per day over Sierra Leone.

Other areas of Africa subjected to seasonal rainfall phenomena are the Greater Horn of Africa during the MAM and SON interseasons and southern Africa during austral summer (DJF). More details on the geographical bounds of these regions in this study and the meteorological phenomena observed there are given below.

2.3.1. The West African region. Although JJA rainfall affects most of the longitudinal extent of the African continent, the region used in this study is restricted to West Africa, ranging from the Gulf of Guinea and the Bight of Biafra to the Atlantic coast alongside Senegal. Grid points of the domain referred to hereafter as 'West Africa' (WA) cover latitudes from 0° to 19°N and longitudes from 20°W to 15°E.

Seasonal means cannot illustrate the spatial shift of precipitation during the West African Monsoon, however comparison between the MAM, JJA and SON precipitation rates over the sub-sahelian regions as well as a clear latitudinal gradient of precipitation rates are consistent with the northward shift of the ITCZ during the monsoon season documented by Sultan et al. (2003). Several studies have shown evidence of strong positive correlation between Gulf of Guinea precipitation and tropical

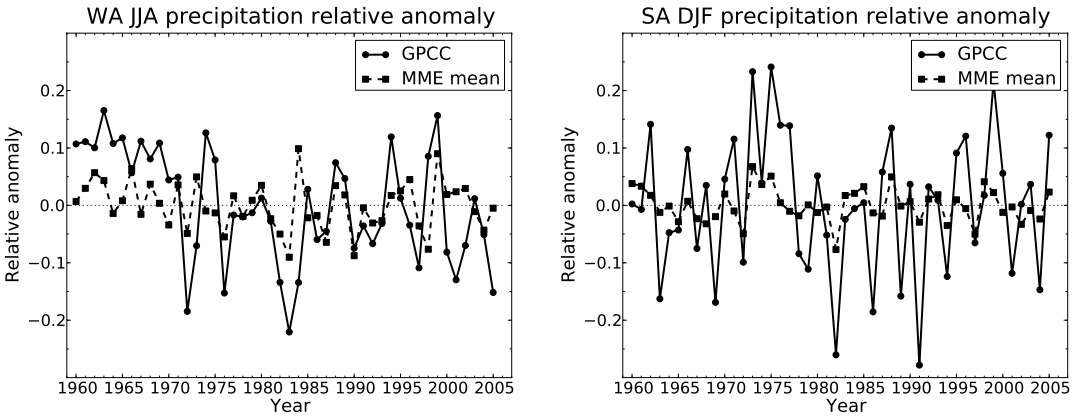


Fig. 2. Seasonal mean precipitation relative anomaly with reference to the 1960–2005 period for GPCC data and the ENSEMBLES project MME ensemble mean, for spatially averaged West Africa JJA precipitation (a) and southern Africa DJF precipitation (b).

South Atlantic SSTs as well as negative correlation between Sahelian precipitation and Niño 3 SSTs during July, August and September (Camberlin et al., 2001; Vizy and Cook, 2001; Giannini et al., 2005).

Figure 2(a) shows a preliminary evaluation of how the ENSEMBLES MME ensemble mean replicates interannual variability of spatially averaged JJA precipitation amounts over land in West Africa. The GPCC anomalies illustrate the observed transition between a 'wet' period before 1970 and abnormally dry years in the 70s and 80s. The MME fails to clearly exhibit this downward trend of precipitation anomalies. Correlation between the MME relative anomalies and GPCC relative anomalies is 0.382. The ability of the MME to represent the precipitation anomaly varies during the 1960–2005 period: for instance, summers from 1987 to 1993 are fairly well represented whereas in 1984, the MME is totally off-course. Observed relative anomalies range from -17 to $+22\%$. Looking at each individual member, predicted relative anomalies range from -33 to $+35\%$ (not shown). Anomalies are compensated by averaging all members in the MME ensemble mean. This explains the smaller range for predicted anomalies in Fig. 2.

2.3.2. The South African region. As shown in Fig. 1(d), DJF precipitation in Africa affects mainly areas south of the Equator, with precipitation means over Madagascar exceeding 10 mm. The southern part of the African continent shows a longitudinal gradient in precipitation rates with little to no precipitation on the Atlantic coast and more than 3 mm per day along the Indian Ocean. This is related to a south-eastwards extension of the ITCZ off the Indian Ocean coast of southern Africa. In order to encompass the entire south African region and Madagascar, the area chosen in this paper ranges from 35°S to 10°S and 5°E to 50°E.

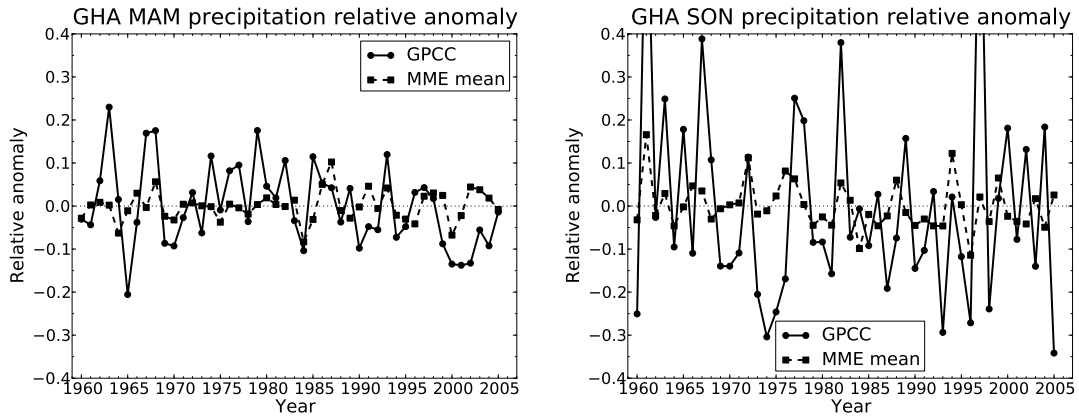
Southern Africa precipitation values and spatial distribution strongly vary from one year to the next, partly due to high variability of SSTs in the surrounding oceans. Previous studies have also shown strong links between ENSO and drought in this region (Rouault and Richard, 2005).

Interannual variability of spatially averaged DJF precipitation over southern Africa for GPCC data and modelled by the ENSEMBLES MME is shown in Fig. 2(b). Correlation between both relative anomaly curves equals 0.624 and is notably higher than for boreal summer in West Africa. The MME duplicates well year-to-year sign and variation of the relative anomaly. Once again, although individual member relative anomalies span the -27% to $+25\%$ range (to be compared with the observed relative anomaly variation between -28% and $+25\%$), MME relative anomalies vary at most from -8% to 7%.

2.3.3. Greater Horn of Africa long rains and short rains seasons. Figure 1 shows that the structure and intensity of precipitation over the Horn of Africa region varies greatly during the year due to a shift of the ITCZ between the northern and southern hemispheres. Spring (MAM) and fall (SON) rainy seasons appear to be transition seasons between the West African monsoon regime in boreal summer and the South African rains in austral summer. During both of the aforementioned summer events precipitation rates over the Horn of Africa north of the Equator are close to zero. In MAM and SON mean precipitation rates remain small and barely reach 2 to 3 mm per day. MAM is the 'long rains' season and shows a lower interannual variability than the SON 'short rains' which are strongly influenced by ENSO and the Indian Ocean dipole (IOD). Both seasons have been thoroughly documented in previous papers (see for instance Camberlin and Philippon, 2002; Bowden and Semazzi, 2007).

The region chosen in this study, hereafter referred to as 'GHA' is a Greater Horn of Africa area including regions affected by both rainy seasons, with latitudes ranging from 13°S to 15°N and longitudes from 25° to 55°E.

The GPCC observed interannual variability of spatially averaged MAM and SON precipitation rates is consistent with



(a) (b)

Fig. 3. Seasonal mean precipitation relative anomaly with reference to the 1960–2005 period for GPCC data and the ENSEMBLES project MME ensemble mean, for spatially averaged Greater Horn of Africa MAM (a) and SON (b) precipitation.

the characteristics stated above (see Fig. 3). MAM precipitation anomalies are smaller than for the SON season, during which positive anomalies reach over 30 % four times during the 1960–2005 period. Three of these strong positive anomalies in precipitation could be due to the 1961, 1982 and 1997 positive ENSO-IOD events described in Bowden and Semazzi (2007). The link between SSTs and SON precipitation is a possible explanation for the higher correlation between observed GPCC precipitation anomalies and those predicted by the ENSEMBLES MME. Correlation is worth approximately 0.424 for SON instead of 0.249 for MAM, illustrating the higher predictability of short versus long rains. For MAM, individual members of each model predict higher anomalies than observed, covering a -40% to $+44\%$ range which is twice as large as the observed range. As for SON anomalies, individual members generally underestimate the positive anomalies, which leads to a covered range of -34% to $+60\%$ instead of the observed -34% to $+86\%$. Again, as shown by Fig. 3, the MME ensemble mean predicts much smaller anomalies due to member error compensation.

Results over the GHA region should be interpreted with caution, since fewer rain-gauge stations were available per grid point for the GPCC reanalysis than for the other two regions discussed in this paper (see for instance Fig. 8 in Schneider et al., 2008).

3. Methodology

The aim of this paper is to provide a range of scores of the ENSEMBLES MME and individual models, as an assessment of ensemble precipitation forecasting skill over the African continent. Alongside currently-used deterministic scores such as root mean square error or correlation coefficients, probabilistic formulations of predictions will be evaluated.

3.1. Deterministic scores

Deterministic scores are calculated for the MME ensemble mean and individual models. Aside from the anomaly correlation coefficient (ACC) scores calculated for every grid point and season of the 1960–2005 period, a 'mean-ACC' score is calculated over each subregion for the seasons of interest according to Déqué and Royer (1992). If o_i represents GPCC observed seasonal precipitation over a certain grid point for year i of the time period and p_i the corresponding model prediction, mACC is formulated as follows:

$$m\mathrm{ACC} = \frac{\overline{\langle (o_i - \overline{o})(p_i - \overline{p}) \rangle}}{\sqrt{\overline{\langle (o_i - \overline{o})^2 \rangle} \overline{\langle (p_i - \overline{p})^2 \rangle}}} \tag{1}$$

$\overline{X}$ and $\langle X \rangle$ are averages of variable X, respectively over time and space.

A spread-skill ratio was also examined for bias-corrected individual models and the MME to quantify the increase of spread incurred by the multimodel.

3.2. Probabilistic forecast and calibration

The basic principle of probabilistic forecasts is to use the spread of the different ensembles members, in addition to their mean, in order to issue a forecast. Events chosen here are precipitation exceeding (or remaining below) a certain quantile of climatological precipitation. Threshold quantiles for a given season are defined using the other years of the 1960–2005 period. The way we define here probabilistic forecasts from ensembles is straightforward: the predicted probability is the fraction of individual forecast outputs (out of m months by n ensemble members) which yield precipitation amounts above (or below) the corresponding threshold.

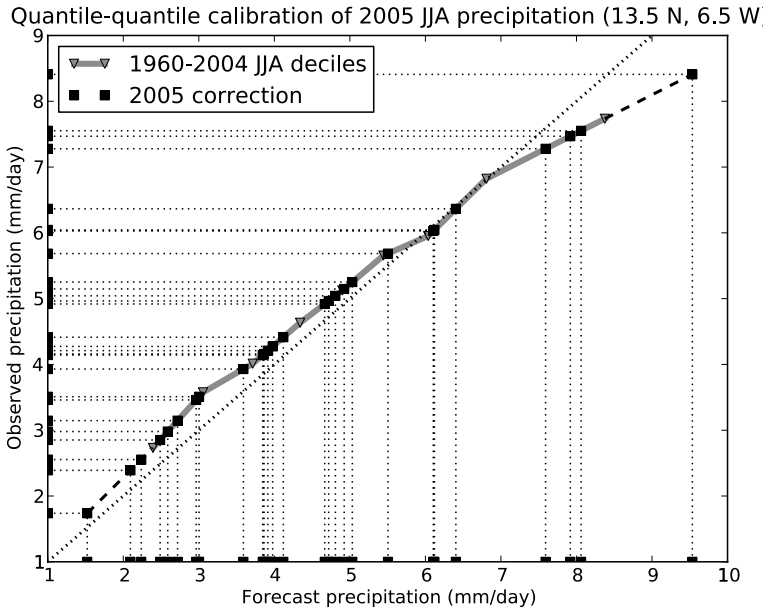


Fig. 4. Illustration of the quantile–quantile calibration technique used for probabilistic forecasts in the case of JJA 2005 precipitation over grid point (13.5° N, 6.5° W) for an individual model of the ENSEMBLES project. The thick grey curve represents the quantile–quantile plot of forecast (x-axis) and observed (y-axis) JJA monthly precipitation values for the other years of the period (1960 to 2004). Forecast precipitation values for the 9 ensemble members and 3 months of JJA 2005 are placed along the x-axis and corrected values are read along the y-axis (black squares), using the projection (dotted lines) on the quantile–quantile curve and linear interpolation between values (and extrapolation for values beyond quantiles, dashed lines).

In order to extract as much information from the different forecasts as possible, a quantile–quantile calibration technique was used to correct each forecast datum. The choice of such a technique is based on the hypothesis that models are able to provide information on ranks in precipitation rather than accurate amounts. Outputs for a given season are corrected using information from forecasts and observations of the other years of the 1960–2005 period (leave-but-one method). For each grid point, corrected predictions for a given model, season and year are inferred from the deciles of the forecast and observed monthly precipitation values of every other corresponding season of the time period, using linear interpolation between deciles. An example of this correction is given in Fig. 4, which illustrates the calibration method for JJA 2005 monthly precipitation single-model forecasts over a given grid point of WA. Note that this technique corrects the climatology of the members, but not their predictability. In particular, if a model underestimates spread between members, this is not corrected here.

3.3. Ranked probability skill score

One of the most frequently-used probability scores in meteorology and climatology is the Brier Score, which is calculated as follows for a binary event (Brier, 1950):

$$BS_i = (y_i - o_i)^2, \tag{2}$$

where y_i is the probability forecast of event A during season i and o_i is equal to one if A occurs for season i and zero otherwise. Then BS can be averaged over the n seasons of the period studied. As suggested in Murphy (1973) and Wilks (2006), the Brier Score can be written as the sum of three terms by sorting the forecast-observation pairs according to the J possible forecast values $\{y_j, j \in 1, .., J\}$ for a single event. If N_j is the number of times the forecast probability equals y_j the following decomposition is obtained:

$$BS = \frac{1}{n}\sum_{j=1}^{J} N_j(y_j - \overline{o_j})^2 - \frac{1}{n}\sum_{j=1}^{J} N_j(\overline{o_j} - \overline{o})^2 + \overline{o}(1 - \overline{o}) \tag{3}$$

$$BS = \text{Reliability} - \text{Resolution} + \text{Uncertainty}. \tag{4}$$

The reliability term measures a weighted distance between forecast probabilities and observed frequency of the event, giving information on the calibration of the model. It should therefore be as small as possible and can be minimized by forecast post-processing. Meanwhile, the resolution term is meant to be as large as possible and estimates the model ability to separate events that have different probabilities of occurrence. A small resolution would mean that for each forecast value y_j the corresponding relative observed frequency of the event $\overline{o_j}$ is close to the overall observed frequency $\overline{o}$, suggesting that forecast probability values give very little information on the actual outcome of the event. The uncertainty term depends only on observations and their variability: it corresponds to the Brier Score of a forecast based on the climatological distribution of the event. Indeed the reliability and resolution of such a forecast are both worth zero, meaning that climatology forecasts have perfect reliability but no resolution whatsoever.

Toth et al. (2003) show how adding Brier scores for events $A_k = \{X \leq x_k\}$ for K increasing thresholds of a variable X amounts to calculating a discrete ranked probability score, or RPS (Epstein, 1969). In this paper we chose this method, adding Brier scores for the $K = 9$ precipitation events $A_k = \{\text{seasonal precipitation of year } i \leq x_{k,i}\}$, where $x_{k,i}$ is the kth

decile of observations for the other years of the period:

$$RPS = \frac{1}{K}\sum_{k=1}^{K} BS_k = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{n}\sum_{i=1}^{n}(y_{k,i} - o_{k,i})^2. \quad (5)$$

Events with 0 and 1 probability are deliberately omitted. RPS ranges from 0 (perfect score) to 1. Calculating this score for the ENSEMBLES models ensemble forecasts (RPS_m) and for the observed climatological distribution over the other years (RPS_c) gives an indication of the relative benefit of forecast use. This relative score is called the RPSS or ranked probability skill score:

$$\mathrm{RPSS} = 1 - \frac{\mathrm{RPS_m}}{\mathrm{RPS_c}} \quad (6)$$

RPSS values range from $-\infty$ to 1. Negative values indicate that climatology outperforms the evaluated model in forecasting precipitation quantiles, whereas positive values can be interpreted as the model improvement rate with respect to climatology.

Using the Brier Score decomposition shown in eq. (4) and the definition of the discrete RPS in eq. (5) yields a decomposition of the RPS into reliability, resolution and uncertainty terms.

3.4. Forecast potential economic value

The RPSS compares model forecasts to using the climatological distribution and in doing so estimates the added value of the model. Another approach to probabilistic forecasts evaluation goes even further by computing a model potential 'economic value' (EV) in an idealized case study (Palmer, 2002). In this framework, potential economic value for a given event (say, precipitation higher or lower than the k th K-quantile) depends on the hit rate a, false alarm rate b and miss rate c of a model as well as a cost-loss ratio C/L.

In this paper, economic value of a model is defined according to Richardson (2003) as the ratio between savings made by using the actual model instead of climatology and hypothetical possible savings made using a 'perfect' model. The mean expense M_m for a model user is given by eq. (8). The climatology mean expense $M_{\mathrm{climatology}}$ is calculated in leave-but-one mode and the 'perfect' model gives an exact probability forecast of the event.

$$\mathrm{EV} = \frac{M_m - M_{\mathrm{climatology}}}{M_{\mathrm{perfect}} - M_{\mathrm{climatology}}} \quad (7)$$

$$M_m = (a+b)C + cL. \quad (8)$$

For each forecast (real, climatology or 'perfect'), the probability for which the user deems preventive action necessary is called probability threshold and is optimized in our study using results from the other years of the time period studied over each grid point. a, b and c therefore depend on the probability threshold and the value of C/L, or cost-loss ratio.

The potential economic value of ENSEMBLES single model forecasts and the MME was evaluated for a set of events and cost-loss ratio (C/L) situations. Events chosen were precipitation exceeding the third quartile of climatology (75 th percentile), referred to as E^+ and precipitation remaining below the first quartile (25 th percentile), abbreviated as E^-. We focused on C/L values close to the climatological probability of such events occurring $\overline{o}$, in both cases worth 0.25. From a user's point of view, a model forecast is most useful in decision-making for C/L values close to $\overline{o}$. In other cases, users can rely on their own common sense to make their mind on whether to spend C on preventive measures or not. Results shown in this paper are overall results for the different regions studied. Rates a, b and c are calculated over all grid points to give an economic value of the models over the region as a whole.

4. Results and discussion

Main results for the ENSEMBLES individual models and MME are presented here, starting with general results over the entire African region and then focusing on the different regions and seasons of interest.

4.1. General results over the African continent

Figure 5 shows seasonal precipitation ACC over the 1960–2005 period for the MME and all four seasons studied in this paper.

In MAM (Fig. 5a), the ENSEMBLES multimodel failures are located mainly over the Sahara and Central Africa. ACCs over the Greater Horn of Africa region are mainly positive but small. Spatial distribution of ACCs differs between the MAM and JJA seasons (Fig. 5b). For the latter, scores are higher over the West African region, particularly on the Gulf of Guinea coast. Over the eastern half of the continent ACCs remain positive but weak and negative scores are found in southern Africa around latitudes from 10 to 20° South where observed mean precipitation is close to zero (see Fig. 1b). In SON (Fig. 5c), South Africa is the region most affected by negative ACCs. GHA region ACCs are distinctively higher for this season (short rains) than for MAM long rains, most certainly due to the stronger teleconnections with SST indexes discussed earlier. ACCs are higher in South Africa during the DJF season (Fig. 5d), when the area is affected by the largest precipitation rates (see Fig. 1d). Note however that south of 20°S, correlations are higher where precipitation rates are lower. Positive ACCs are also found over Mauritania and the GHA region for this season.

Generally speaking precipitation ACCs are positive over the regions described earlier during the seasons of interest, even if correlations are weak in the case of MAM precipitation over the GHA region.

Table 2 shows the mACC values for individual models and the MME calculated over each region for the season(s) of higher precipitation rates. Calculated mACCs are all positive but reach at most 0.28. Scores for MAM precipitation over the GHA region confirm results shown for ACCs and are lower than mACCs

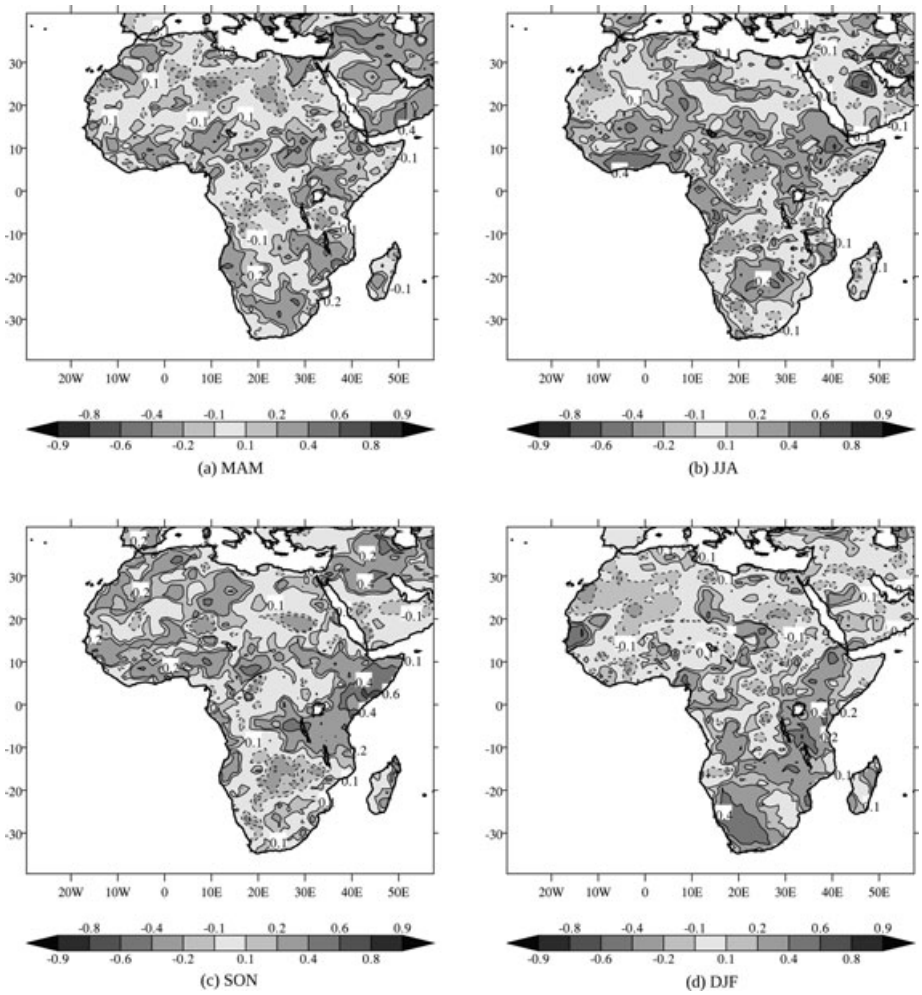


Fig. 5. Anomaly correlation coefficients over the African continent for the four seasons of interest. Solid and dashed lines indicate positive and negative values, respectively.

Table 2. mACC values for the regions of study and corresponding seasons of interest over the 1960–2005 period and for three 15 or 16-year subperiods (*) see text for more details

Model	WA JJA	SA DJF	GHA MAM	GHA SON
ECMWF	0.104	0.181	0.030	0.034
IFM-G	0.142	0.138	0.032	0.246
MF	0.130	0.031	0.063	0.285
UKMO	0.227	0.124	0.065	0.101
CMCC	0.149	0.121	0.027	0.056
10-member MM (*)	0.198	0.132	0.057	0.201
MME	0.223	0.189	0.080	0.244
MME 1960–1974	0.300	0.257	0.093	0.353
MME 1975–1989	0.169	0.170	0.097	0.134
MME 1990–2005	0.134	0.190	0.068	0.242

for the SON season for each individual model and the MME. mACCs obtained for the MME illustrate the interest of combining single model ensembles into a 45-member multimodel ensemble, since scores are higher than single-model scores for all regions and models save for three cases for which a single-model ensemble performs outstandingly well.

Results are also shown for a multimodel 10-member ensemble to test the influence of ensemble size on the MME scores. Each

model provides two randomly-drawn members for this reduced multimodel. The score shown in Table 2 is the mean of 1000 random draws. It gives an indication of score enhancement due to the size increase in the MME. Indeed, over WA the reduced multimodel ranks second when compared to the 9-member models as did the MME and over GHA and SA it lags behind two individual models with a clearly reduced mACC when compared to the MME. This illustrates that the multimodel performance in precipitation seasonal forecasts resides mainly in the increased number of members of the ensemble, especially when one of the single-model ensembles yields poor anomaly correlation coefficients over the region. An apparent drawback in using multimodels is to include poorer models in the ensemble, however a closer look at results over different regions shows that no model is clearly superior to the others in terms of mACC scores and each model ranks in the top two single models for one region at least.

The last three lines in Table 2 show MME scores for 15 or 16-year subperiods of the 1960–2005 time period so as to illustrate a possible variation of skill over time. These scores can only be compared with each other since the length of the time period is in these cases much shorter. They will be discussed in the following paragraphs.

4.2. *Prediction of West African JJA precipitation*

More specific results for the West African region are presented here. Fig. 6(a) shows a scatter plot of the ensemble spread and ensemble-mean RMSE skill values of each model for every JJA season of the 1960–2005 period. The thin dashed line corresponds to a spread-skill ratio equal to one, the ideal value when ensemble spread is consistent with model error. Individual models have higher errors than their ensemble spread and unlike RMSE, ensemble spread varies very little from one season to the other. Results are much more satisfactory for the MME (black squares), since for most years ensemble spread is close to RMSE. For some years the MME is slightly over-dispersive,

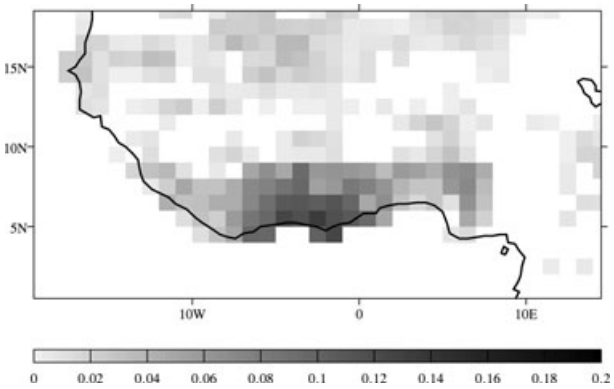


Fig. 7. RPSS over West Africa for JJA precipitation—only grid points with positive values are shown.

which is never the case for individual models over West Africa. As found in previous studies, multimodel ensembles have a better spread-skill ratio (closer to one) than single model ensembles. However, as stated in Doblas-Reyes et al. (2009) this does not guarantee actual model skill over the region.

On average, the mACC scores shown in Table 2 are highest for the WA region and range from 0.10 to 0.23, which suggests that a probabilistic formulation can add value to the forecasts by using the entire spectra of information from the different ensemble members. The mACC scores for the subperiods show a deterioration of mACC for WA JJA precipitation between the 1960 to 1974 period and the following periods. As explained earlier, the MME fails to capture the downward trend in precipitation anomalies from wet conditions in the 1960s to drought in the 1980s. 1960–1974 is a relatively wet period during which the MME performs better in terms of mACC.

Figure 7 shows the ranked probability skill score for the 45-member MME in its probabilistic formulation and the A_k events defined earlier. Values are shown at gridpoints where the RPSS is positive, meaning that the MME provides added value with respect to a simple climatological seasonal forecast. As for the

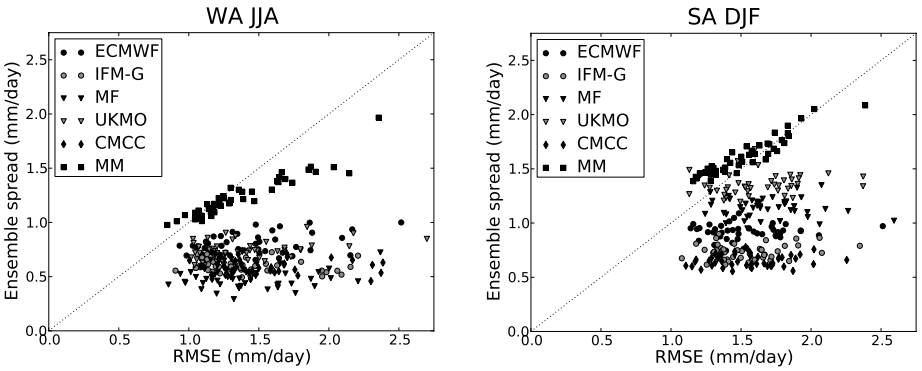


Fig. 6. Spread-skill scatter plot for the West African region JJA (a) and southern Africa DJF (b) precipitation over the 1960–2005 period. Each point represents spread and ensemble-mean RMSE skill values for a given season and model.

Table 3. Single model and MME spatial average RPSS (%), RPS and reliability–resolution decomposition (values divided by uncertainty) for the different regions and seasons of interest over the 1960–2005 period

Model		**WA JJA**				**SA DJF**		
	RPSS (%)	RPS	Rel	Res	RPSS (%)	RPS	Rel	Res
ECMWF	−2.71	0.188	0.29	0.26	−0.45	0.184	0.27	0.27
IFMG	−2.21	0.187	0.26	0.23	−2.12	0.187	0.27	0.25
MF	−5.65	0.194	0.36	0.30	−6.00	0.194	0.34	0.28
UKMO	−0.38	0.184	0.28	0.28	−2.18	0.187	0.27	0.25
CMCC-INGV	−8.32	0.199	0.39	0.30	−2.25	0.187	0.29	0.26
MME	0.98	0.182	0.41	0.42	1.44	0.180	0.41	0.42
		GHA MAM				**GHA SON**		
ECMWF	−4.16	0.191	0.36	0.31	-8.08	0.198	0.41	0.33
IFMG	−3.96	0.191	0.36	0.32	−4.59	0.192	0.37	0.33
MF	−4.70	0.192	0.37	0.32	−3.95	0.191	0.39	0.35
UKMO	−3.53	0.190	0.35	0.32	−6.78	0.196	0.41	0.34
CMCC-INGV	−4.48	0.192	0.37	0.32	−6.97	0.196	0.39	0.32
MME	−1.25	0.186	0.44	0.43	−2.86	0.189	0.49	0.46

ACCs calculated over the region, the best scores are achieved along the northern coast of the Gulf of Guinea. Most grid points of the Sahel region have a positive RPSS although values are very small. Between 9 and 14°N as well as in the eastern part of the area, the MME often ranks below the climatological forecast in terms of RPS. A glance at ACCs calculated for these grid points (see Fig. 5b) shows that they are generally very small and sometimes negative, which illustrates that probabilistic formulation of ensemble forecasts cannot create value over grid points where models have no skill in the deterministic sense. Generally speaking, RPSS values are small and range from −10% to 13.5% with an average over WA of 1.0%.

Table 3 shows the RPSS and RPS spatial average and for each single model and the MME as well as the reliability-resolution terms of the RPS discussed earlier. These terms are divided by the RPS uncertainty term. Scores for WA are shown in the upper left corner of the table. The MME is the only model for which the spatial average of RPSS is positive over the West African region, meaning that single models alone are less efficient than climatology in predicting precipitation ranks. When compared to single model RPSS maxima (13.1%) and minima (−62.4%) over the area, the multimodel clearly improves the minimum RPSS over the region whereas the maximum RPSS value is only slightly larger than the highest single model maximum score. This shows that multimodel additional information is particularly useful over grid points where single models experience difficulties and won't further improve scores over grid points where precipitation predictions are already skillful. This conclusion is corroborated by comparing figures of RPSS values over the area for single models and the MME (not shown).

The reliability-resolution terms in Table 3 show that the multimodel improves the single-model RPSS by dramatically increasing model resolution. This positive effect compensates a smaller increase in the reliability term, meaning that the multimodel calibration is poorer than that of single models.

Economic value of the ENSEMBLES single models and MME for events E^- and E^+ over West Africa during the JJA season for different cost-loss ratio values is shown in Fig. 8. C/L values range between 0.05 and 0.45 with a 0.025 step. For both events, EV is maximum for $C/L = \overline{o}$ and reaches at most 10.4% for E^+ and 6.1% for E^-. In most cases $\overline{o}$ is the only C/L value for which models reduce expenses with respect to climatological forecasts. The three models with highest values are the same for both events and values are generally higher for E^+ than E^-. Unlike results shown for RPSS, the multimodel does not

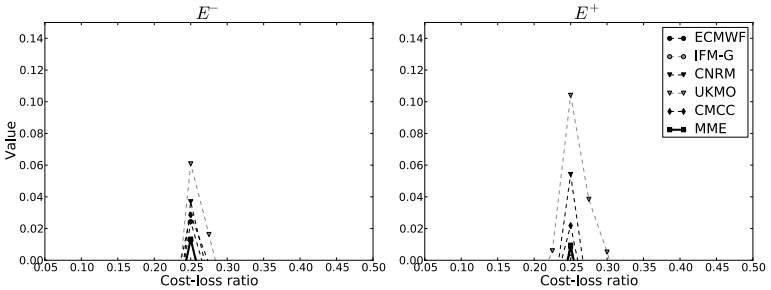


Fig. 8. Economic value of ENSEMBLES project single models and MME (thick line) for events E^- (left) and E^+ (right) and different cost-loss ratio values (see text for more details) for predictions of JJA rainfall over West Africa.

improve economic value with respect to individual models for both events.

4.3. Results over the Greater Horn of Africa

Results over GHA vary between the MAM and SON seasons. Mean ACC scores shown earlier suggested that seasonal precipitation predictability was higher for SON than MAM. A comparison of Figs. 9(a) and (b) does not lead necessarily to the same conclusion: both scatter plots show similar spread-skill ratios for both seasons, with spread and RMSE generally lower during the MAM season. As for WA in JJA, the multimodel helps broaden model ensemble spread while maintaining similar RMSE levels. For both seasons and for most years of the 1960–2005 period the multimodel is slightly over-dispersive.

The evolution of mACC for three subperiods shown in Table 2 differs between both seasons. In MAM, no striking difference is found between the subperiods, although 1990–2005 scores are slightly lower: the three subperiods show low anomaly correlation coefficients for the MME over the GHA region. In SON, mACCs are notably higher for the 1960–1974 subperiod (0.353), when mean precipitations are higher than average and lower for the 1990–2005 subperiod (0.134), when mean precipitations are slightly lower, although strong interannual variations in precipitation anomalies make further interpretation of these results quite difficult (see Fig. 3b).

RPSS calculations over GHA are shown in Figs. 10(a) and (b) for the MAM and SON seasons, respectively. An immediate conclusion when looking at both figures is that the MME precipitation deciles predictions are much better for SON than for MAM. There are many more grid points with positive RPSS in the case of SON (69% vs. 31.9% for MAM). The spatial distribution of positive RPSS is once again consistent with results for anomaly correlation coefficients. However, RPSS values shown in Table 3 (lower half) show that average scores are lower for SON than MAM. This is due to highly negative RPSS scores over the northwest corner of the GHA domain. All models tend to always predict positive precipitation rates over these gridpoints where little to no precipitation is actually observed. In spite of the quantile–quantile correction each model exhibits very poor skill over this subregion, penalizing the overall score over GHA for SON.

Individual models and MME economic values for events E^- and E^+ were calculated over GHA for both seasons. Results for E^+ during SON are shown in Fig. 11 . Each model has a positive EV for $C/L = \overline{o}$, and values reach over 15% for one single model. The multimodel ranks second among all models, contrasting with results over WA.

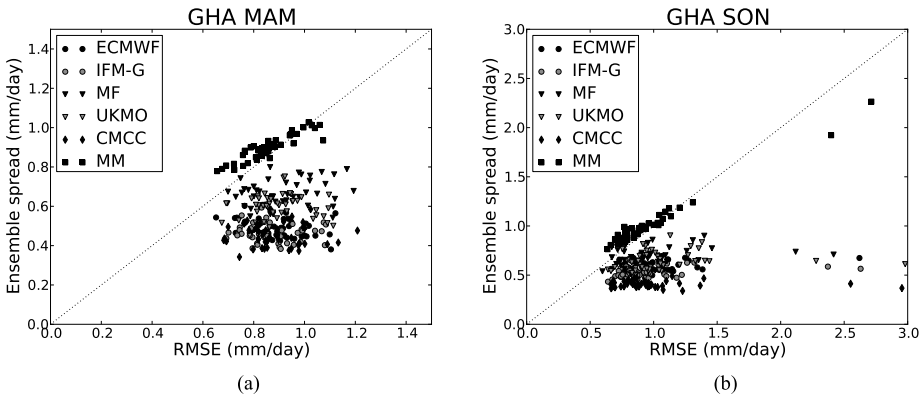


Fig. 9. Same as Fig. 6 but for Greater Horn of Africa MAM (a) and SON (b) precipitation. Note that axis scales are not the same for both figures.

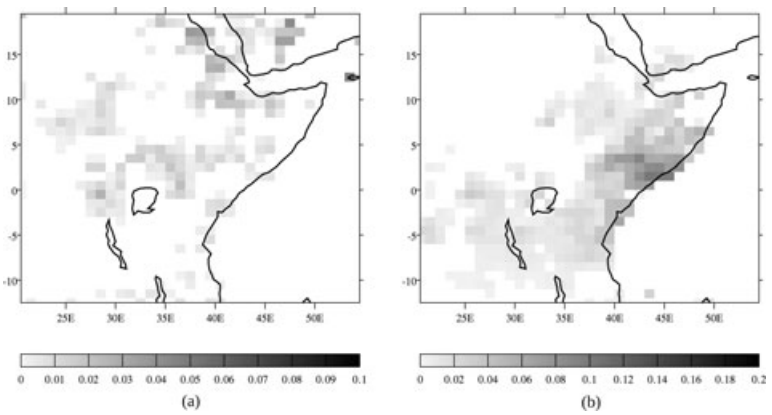


Fig. 10. Same as Fig. 7 but for Greater Horn of Africa MAM (a) and SON (b) precipitation.

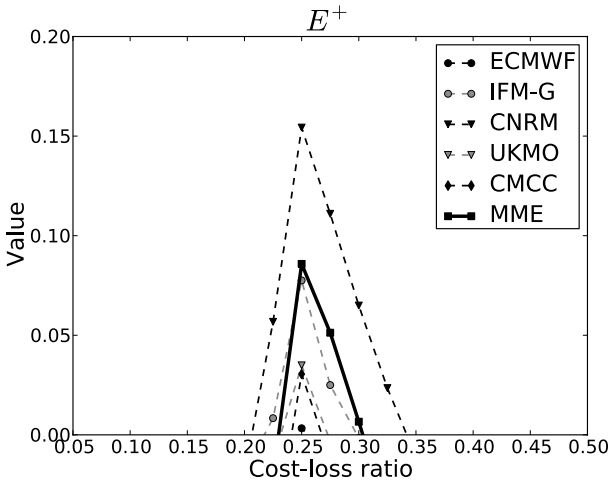


Fig. 11. Economic value of ENSEMBLES stream 2 single models and MME (thick line) for event E^+ and different cost-loss ratio values (see text for more details) for predictions of SON rainfall over GHA.

For E^- in SON only one model has a positive EV (of only 3%) over GHA. Similarly disappointing results are found over GHA during MAM where very few models show hints of economic value and multimodel values remain negative for all cost-loss ratios. These results are not surprising given the poor performance of probabilistic and deterministic forecasts over the region, and are therefore not shown here.

4.4. South African DJF precipitation forecast skill

A spread-skill scatter plot for DJF precipitation over the SA region is shown in Fig. 6(b). Results are quite similar to those observed over WA in boreal summer, even if ensemble spread and RMSE values are generally higher. Ensemble spread interannual variations for individual models are small, leading to a mainly horizontal dispersion of points in the scatter plot. This is corrected by using the MME for which the alignment of points is closer to the idealistic diagonal line. As for GHA, switching from single models to MME leads to a better spread-skill ratio and slightly over-dispersive ensembles.

When compared to WA JJA, mACC values for SA DJF (Table 2) are poorer for 4 out of 5 single models and the multimodel, yet the MME mACC value (0.189) gives hope that a probabilistic formulation may yield positive results over the region. It is worth noticing that SA DJF ACC values are less noisy in space than WA JJA (see Figs. 5b and d) although spatial distribution of precipitation amounts is heterogeneous in both cases. Table 2 shows that mACC over SA is higher for 1960–1974 than for the other two subperiods examined, as for WA JJA and GHA SON. However, unlike WA JJA no clear trend in precipitation can be seen in observations or in the MME predictions (see Fig. 2b).

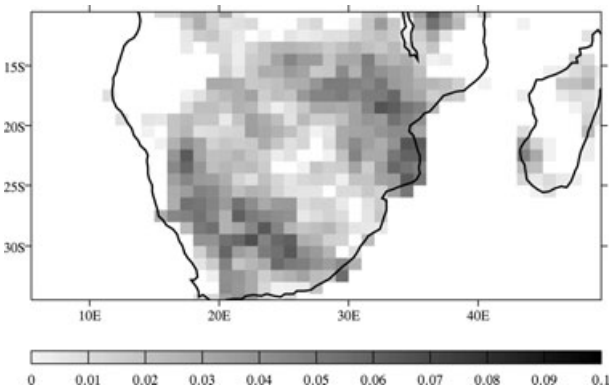


Fig. 12. Same as Fig. 7 but for southern Africa DJF precipitation.

Figure 12 shows the ranked probability skill scores of the MME over each grid point for SA DJF precipitation. Results averaged over the area are shown in the upper right-hand corner of Table 3. The MME RPSS averaged over land gridpoints is positive and even slightly larger than for WA JJA, although maximum and minimum values over the area are quite smaller than those calculated for West Africa (not shown). More strikingly, positive RPSS values are found over approximately 75% of the SA region (for WA JJA, RPSS is positive for only 57% of the grid points).

Results for model EV over southern Africa for dry (E^-) and wet (E^+) events are given in Fig. 13. Unlike over WA, the multimodel improves economic value of most single models for E^-, with maximum value reaching 7.3% of a perfect model score. Values for E^- are slightly higher than over WA and the best models are not the same. For E^+ results are quantitatively similar for single models, yet the multimodel EV is 50% lower than that of the highest ranking model. Results for the multimodel are still much better than over West Africa.

4.5. Links between SST and precipitation

Results shown in the previous paragraphs are generally better for regions and seasons where precipitation is strongly linked to oceanic surface conditions, as illustrated for instance by differences in mACC between MAM and SON precipitation forecasts over the GHA region.

To further investigate links between sea-surface temperature and precipitation prediction by the MME, several additional score calculations and model-observation comparisons were led. The ERA-40 prescribed SST data was used for years from 1960 to 1988 included and ERA-Interim prescribed SST data covered the remaining 1989–2005 period. Note that ERA-40 and ERA-Interim SST data are almost the same from 1989 to 2001. Prescribed sea-surface temperatures are based on the HADISST1 data set prior to November 1981 and NOAA/NCEP SST products thereafter (Uppala et al., 2005).

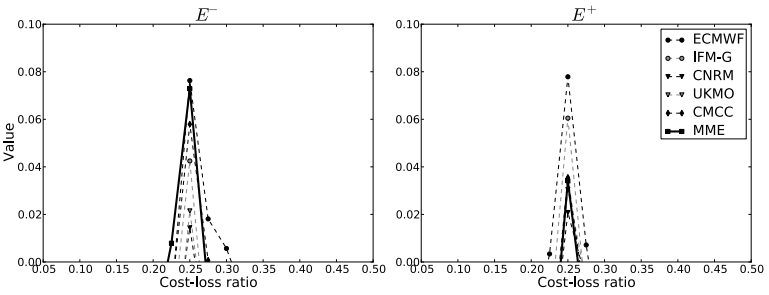


Fig. 13. Same as Fig. 8 but for predictions of DJF precipitation over SA.

4.5.1. Covariance between precipitation anomalies and simultaneous SSTs. As in Guérémy et al. (2005), covariances between seasonal SSTs and standardized anomalies of precipitation over the regions studied were calculated to exhibit the SST anomalies related to precipitation anomalies and therefore possible forcing regions. Results are shown in Fig. 14 and can indicate the ability of the model to replicate observed links between seasonal precipitation anomalies over the GHA, WA and SA regions and simultaneous sea-surface temperatures around the globe.

Covariances between MAM SSTs and MAM GHA precipitation anomalies show no distinct patterns in the observation data sets (Fig. 14a). Camberlin and Philippon (2002) suggested that MAM GHA precipitation links with SSTs over the Tropical Pacific were difficult to assess at a seasonal scale since phase shifts in ENSO often occur during this season. The MME covariances shown in Fig. 14(b) are positive over the Eastern Tropical Pacific, meaning that higher precipitation over GHA is correlated with higher SSTs over this region in the model. Covariances over the Atlantic and Indian Oceans are also inconsistent with observations, illustrating possible sources of model error.

Results for other regions and seasons are generally better. Figures 14(c) and (d) show that for JJA WA the MME finds the same sign of covariances as the observations over the tropical Pacific and the Gulf of Guinea, although values differ. As in the observed data, higher-than-average SSTs over the Gulf of Guinea and lower-than-average SSTs over the Eastern Tropical Pacific lead to positive anomalies in precipitation over the WA box. SON SST covariance with GHA precipitation anomalies shows a strong positive pattern over the Eastern Tropical Pacific (Fig. 4e) which also appears for the MME but with a smaller amplitude (Fig. 4f). Negative covariances just west of Indonesia are well replicated. Out of all seasons and regions, the one for which the MME depicts observed covariances best is DJF SA. A Pacific Ocean SST latitudinal quadripole appears in both Figs. 4(g) and (h), and the MME shows similar structures for South Atlantic and Indian Ocean covariances.

These results can be linked to the MME performances observed earlier. GHA in MAM is the region for which scores are lowest, and the MME seems to yield unobserved patterns between SSTs and precipitation anomalies. On the other hand, for DJF SA, the MME depicts well both patterns and values of covariances, and this is the region and season for which the best probabilistic scores are found.

4.5.2. Dependency of results on ENSO. Previous studies have shown that seasonal prediction skill is closely related to ENSO amplitude (Wang et al., 2009). In order to study the dependency of the earlier presented results with ENSO, the 46 years of the time period were split into three sets depending on the tercile of DJF Niño 3.4 mean observed SST in the current year. DJF 1959–1960 was also included in the time period to study 1960 seasonal forecast skill.

The three sets obtained are listed in Table 4. These sets are the same if the Niño 3 region is used instead of Niño 3.4 and are quite similar to recorded El Niño and La Niña episodes. The choice of DJF terciles as a criterion instead of usual ENSO indices is dictated by the need to obtain three sets of same size, so scores such as ACC are comparable.

mACC results were calculated for each region and season for these three categories, using for WA JJA and GHA MAM the preceding DJF season, and for GHA SON and SA DJF the following or synchronous DJF season as a classification criterion (for instance, for DJF 1997–1998, SON 1997, DJF 1997–1998, MAM 1998 and JJA 1998 scores are related to the 'warm years' category). Results for these mACC are presented in Table 5. For WA JJA, results are similar for all three categories, indicating that the same level of correlation is maintained in non-ENSO years. This tends to illustrate the predominance of local and synchronous ocean conditions such as JJA Gulf of Guinea SSTs over large-scale teleconnections in MME forecast skill over the WA region as a whole.

Results over GHA and SA are more contrasted. MME mACCs are higher over the average years subperiod than for warm and cold years for GHA SON precipitation prediction, but the opposite results are found for GHA MAM and SA DJF. Figures 14(g) and (h) showed that anomalies in DJF precipitation over SA were negatively correlated to DJF SSTs over the tropical Pacific and that the MME exhibited a covariance pattern similar to observations. The improved mACC results obtained for SA DJF during warm and cold years are consistent with these former results.

Results for MAM mACCs are somewhat surprising since no clear connection between ENSO and seasonal-scale precipitation over the entire GHA domain has been documented.

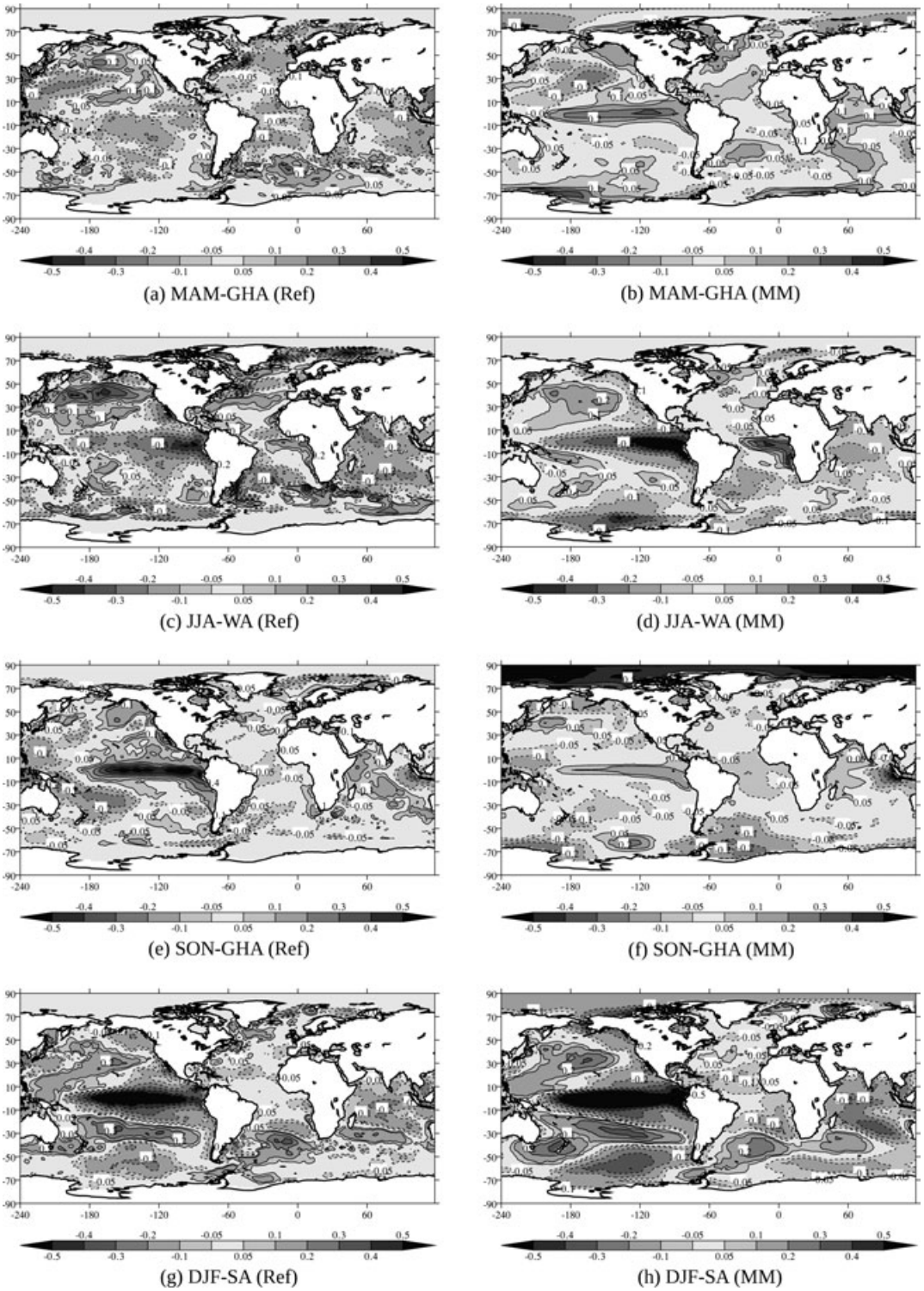


Fig. 14. Covariance between SSTs and standardized precipitation anomalies over GHA (for MAM, SON) WA (for JJA) et SA (for DJF). The left column shows covariances between ERA SSTs and GPCC precipitation and the right column shows the MME covariances. Values are in Kelvin, solid and dashed lines indicate positive and negative values, respectively.

Table 4. Classification of DJF seasons in warm, average and cold years categories for Niño 3.4 SST. (The year corresponds to the month of December)

Warm years	1963 1965 1968 1969 1972 1976 1977 1979 1982 1986 1987 1991 1994 1997 2002 2004
Average years	1959 1960 1961 1962 1966 1974 1978 1980 1981 1989 1990 1992 1993 1996 2001 2003
Cold years	1964 1967 1970 1971 1973 1975 1983 1984 1985 1988 1995 1998 1999 2000 2005

Table 5. mACC values for the regions of study and corresponding seasons of interest over the 1960–2005 period for the MME over warm, average and cold DJF SST years

ENSO category	WA JJA	SA DJF	GHA MAM	GHA SON
Warm years	0.214	0.174	0.068	0.108
Average years	0.229	0.047	−0.050	0.381
Cold years	0.201	0.151	0.190	0.155

Grid-point temporal ACCs for the three subcategories of years over GHA reveal that strong positive ACCs are found during cold years over equatorial East Africa and during warm years over the northwestern corner of the GHA domain and north of Kenya and Uganda (not shown). For average years, a negative mACC is found, consistent with scores found over most regions of the GHA domain. For SON, high ACCs (over 0.7) for average years are found over Somalia and along the coast, and positive ACCs mainly south of 4° N for inland areas (not shown). During warm and cold years, ACC patterns differ, with positive values located mainly north of the Equator for cold years and over the Horn of Africa and south of the Equator (but with smaller values) for warm years. Lower ACCs for warm and cold years in SON could be explained by the fact that the MME forecasts are initialized in August and early signs of warmer or colder than average DJF SST conditions may not be well captured by the MME. It is interesting to note that in the case of warm years (generally giving rise to higher SON precipitation rates over GHA), lower mACCs for the MME ensemble mean do not translate into absence of skill of the probabilistic MME in forecasting wet events (such as E^+ studied for economic value scores).

5. Summary and conclusions

Seasonal forecast skill is often assessed over tropical regions for surface temperature variables. This paper gives an overview of the ENSEMBLES stream 2 seasonal forecast project: single models and multimodel ensemble seasonal precipitation forecast skill is assessed over the African continent. The aim of this study is to illustrate potential predictability of seasonal precipitation using only simple and robust model calibrations.

Results over specific regions during higher-than-average precipitation or monsoon seasons are quite contrasted. For all regions in this study the 45 member MME considerably increased ensemble spread without deteriorating model errors. Over West Africa during JJA, results are encouraging. The ENSEMBLES MME manages to grasp some interannual variability and has added value with respect to climatology in terms of RPS not only over the Gulf of Guinea but also over the Sahel (although with much lower scores), which is not the case for most individual models. In southern Africa for the DJF season, positive RPS skill is found over most grid points of the region. Scores are much more disappointing over the Greater Horn of Africa. For the SON 'short rains' season, more grid points with positive RPSS are found than for the MAM 'long rains' season. However, average scores for SON are lower due to strongly negative RPSS over the northeastern part of the GHA domain. Generally speaking, results are thus better for regions and seasons where precipitation is strongly linked with oceanic surface conditions.

In terms of EV, positive values were found for both single models and the MME over the West African and southern African regions for cost-loss ratios equal to the probability of occurrence of the precipitation events studied. This illustrates that even for a variable highly dependent on parameterizations such as precipitation, GCMs can show additional skill when compared to climatology. These results are encouraging and contrast with the reputation GCMs have been given over the years when it comes to precipitation forecasts outside the Tropical Pacific area. They also suggest room for improvement using adaptation techniques such as model-output statistics.

Maps of MME global SST covariance with precipitation anomalies over the regions studied were compared to observations and showed that the ENSEMBLES MME generally replicated the main SST patterns linked with simultaneous precipitation anomalies, save for the GHA region in MAM where no clear observed patterns were found. Deterministic scores calculated for warm, average and cold years of DJF SSTs over the Niño 3.4 region illustrate possible links between ENSO and MME skill, especially over the SA region where mACCs were higher for warm and cold years than for average years. No clear difference in skill was found between the three sets of years for the WA box in JJA. This could be due to the fact that the WA box

studied in this paper comprises both the Gulf of Guinea, strongly influenced by local SSTs, and the Sahel for which precipitation is known to be linked to ENSO.

Probabilistic evaluations of seasonal precipitation predictability are more user-related than general deterministic scores such as anomaly correlation. It is worth noticing however that positive results for probabilistic forecasts were found only over regions were deterministic skill had been demonstrated, suggesting that the use of information from all of the ensemble members is fruitful only where the ensemble mean forecast shows hints of performance. Results might depend on the event studied and the number of quantiles chosen for calibration, but additional calculations using terciles instead of quartiles for EV showed similar results. The fairly long time period during which data is evaluated (46 years) argues for robustness of results, however it is worth noting that most scores are low (especially the RPSS) compared to SST prediction in the tropical Pacific. The usefulness of MME seasonal forecasts must therefore be qualified in application studies implying for instance water management or health-related decision-making.

6. Acknowledgments

The ENSEMBLES data used in this work was produced by the European Commission FP6 Integrated Project ENSEMBLES (Contract number 505539). The Global Precipitation Climatology Centre (GPCC) is operated by Deutscher Wetterdienst (DWD) under the auspices of the World Meteorological Organization (WMO). ERA-40 and ERA-Interim SST data were supplied by ECMWF. We are grateful to two anonymous reviewers who contributed to improve significantly the manuscript.

References

Berner, J., Doblas-Reyes, F., Palmer, T., Shutts, G. and Weisheimer, A. 2008. Impact of a quasi-stochastic cellular automaton backscatter scheme on the systematic error and seasonal prediction skill of a global climate model. *Philos. Trans. R. Soc. London, Ser. A* **366**(1875), 2559–2577. doi:10.1098/rsta.2008.0033.

Bouali, L., Philippon, N., Fontaine, B. and Lemond, J. 2008. Performance of DEMETER calibration for rainfall forecasting purposes: application to the July – August Sahelian rainfall. *J. Geophys. Res.* **113**(D15),doi:10.1029/2007JD009403.

Bowden, J. and Semazzi, F. 2007. Empirical analysis of intraseasonal climate variability over the Greater Horn of Africa. *J. Clim.* **20**(23), 5715–5731.

Brier, G. 1950. Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.* **78**, 1–3.

Camberlin, P., Janicot, S. and Poccard, I. 2001. Seasonality and atmospheric dynamics of the teleconnection between African rainfall and tropical sea-surface temperature: Atlantic VS. ENSO. *Int. J. Climatol.* **21**(8), 973–1005.

Camberlin, P. and Philippon, N. 2002. The East-African March-May rainy season: Associated atmospheric dynamics and predictability over the 1968–1997 period. *J. Clim.* **15**(9), 1002–1019.

Collins, M., Booth, B., Harris, G., Murphy, J., Sexton, D. and co-authors. 2006. Towards quantifying uncertainty in transient climate change. *Clim. Dyn.* **27**, 127–147, doi:10.1007/s00382-006-0121-0.

Déqué, M. and Royer, J.-F. 1992. The skill of extended-range extratopical winter dynamical forecasts. *J. Clim.* **5**, 1346–1356.

Doblas-Reyes, F., Weisheimer, A., Déqué, M., Keenlyside, N., MacVean, M. and co-authors. 2009. Addressing model uncertainty in seasonal and annual dynamical ensemble forecasts. *Q. J. R. Meteorol. Soc.* **135**, 1538–1559, doi:10.1002/qj.464.

Epstein, E. 1969. A scoring system for probability forecasts of ranked categories. *J. Appl. Meteorol.* **8**, 985–987.

Garric, G., Douville, H. and Déqué, M. 2002. Prospects for improved seasonal predictions of monsoon precipitation over Sahel. *Int. J. Climatol.* **22**, 331–345, doi:10.1002/joc.736.

Giannini, A., Saravanan, R. and Chang, P. 2005. Dynamics of the boreal summer African monsoon in the NSIPP1 atmospheric model. *Clim. Dyn.* **25**, 517–535. doi:10.1007/s00382-005-0056-x.

Guérémy, J.-F., Déqué, M., Braun, A. and Piedelièvre, J.-P. 2005. Actual and potential skill of seasonal predictions using the CNRM contribution to DEMETER: coupled versus uncoupled model. *Tellus* **57A**, 308–319.

Murphy, A. 1973. A new vector partition of the probability score. *J. Appl. Meteorol.* **12**, 595–600.

Palmer, T. 2002. The economic value of ensemble forecasts as a tool for risk assessment : from days to decades. *Q. J. R. Meteorolog. Soc.* **128**, 747–774.

Palmer, T., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M. and co-authors. 2004. Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). *Bull. Am. Meteorol. Soc.* **85**, 853–872.

Philippon, N., Doblas-Reyes, F. and Ruti, P. 2010. Skill, reproducibility and potential predictability of the West African monsoon in coupled GCMs. *Clim. Dyn.* **35**, 53–74, doi:10.1007/s00382-010-0856-5.

Richardson, D. 2003. Economic value and skill. In: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, (eds. I. Joliffe and D. Stephenson), John Wiley & Sons Ltd, Chichester, UK, pp. 165–187.

Rouault, M. and Richard, Y. 2005. Intensity and spatial extent of droughts in southern Africa. *Geophys. Res. Lett.* **32**, doi:10.1029/2005GL022436.

Schneider, U., Fuchs, T., Meyer-Christoffer, A. and Rudolf, B. 2008. Global precipitation analysis products of the GPCC. Global Precipitation Climatology Centre (GPCC) Internet Publication, Offenbach, Germany, pp. 1–12.

Shukla, J. 1998. Predictability in the midst of chaos: a scientific basis for climate forecasting. *Science* **282**, 728–731, doi:10.1126/science.282.5389.728.

Sultan, B., Janicot, S. and Diedhiou, A. 2003. The West African Monsoon dynamics. Part I: documentation of intraseasonal variability. *J. Clim.* **16**(21), 3389–3406.

Toth, Z., Talagrand, O., Candille, G. and Zhu, Y. 2003. Probability and ensemble forecasts. In: *Forecast Verification, A Practitioner's Guide in Atmospheric Science*, (eds. I. Joliffe and D. Stephenson), John Wiley & Sons Ltd, Chichester, UK, pp. 137–163.

Uppala, S. M., Kallberg, P., Simmons, A., Andrae, U., Da Costa Bechtold, V. and co-authors. 2005. The ERA-40 reanalysis. *Q. J. R. Meteorol. Soc.* **131**, 2961–3012, doi:10.1256/qj.04.176.

Vizy, E. and Cook, K. 2001. Mechanisms by which Gulf of Guinea and Eastern North Atlantic sea surface temperature anomalies can influence African Rainfall. *J. Clim.* **14**(5), 795–821.

Wang, B., Lee, J.-Y., Kang, I.-S., Shukla, J., Park, C.-K. and co-authors. 2009. Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004). *Clim. Dyn.* **33**, 93–117, doi:10.1007/s00382-008-0460-0.

Weisheimer, A., Doblas-Reyes, F., Palmer, T., Alessandri, A., Arribas, A. and co-authors. 2009. ENSEMBLES: a new multi-model ensemble for seasonal-to-annual predictions—skill and progress beyond DEMETER in forecasting tropical Pacific SSTs. *Geophys. Res. Lett.* **36**, doi:10.1029/2009GL040896.

Wilks, D. 2006. Statistical Methods in the Atmospheric Sciences. In: Chapter 7: Forecast Verification, 2nd Edition, Academic Press, London, UK, pp. 255–335.