Seasonal forecast quality of the West African monsoon rainfall regimes by multiple forecast systems

Luis Ricardo Lage RODRIGUES[1], Javier GARCÍA-SERRANO[1,2] and Francisco DOBLAS-REYES[1,3]

(1) Institut Català de Ciències del Clima (IC3), Barcelona, Spain

(2) LOCEAN-IPSL, Université Pierre et Marie Curie (UPMC), Paris, France

(3) Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

luis.rodrigues@ic3.cat

Key points:

- Forecast quality assessment of the West African monsoon rainfall regimes
- A simple statistical model is a difficult benchmark to be outperformed
- Multi-model ensemble is not always better than the best single forecast system

Abstract

A targeted methodology to study the West African monsoon (WAM) rainfall variability is considered where monthly rainfall is averaged over 10ºW-10ºE to take into account the latitudinal migration and temporal distribution of the WAM summer rainfall. Two observational rainfall datasets and a large number of quasi-operational forecast systems, among them two systems from the European EUROSIP initiative and six systems from the North American Multi-Model Ensemble (NMME) project, are used in this research. The two leading modes of the WAM rainfall variability, namely the Guinean and Sahelian regimes, are estimated by applying principal component analysis (PCA) on the

longitudinally averaged precipitation. The PCA is performed upon the observations and each forecast system and lead time separately. A statistical model based on simple linear regression using sea surface temperature indices as predictors is considered both as a benchmark and an additional forecast system. The combination of the dynamical forecast systems and the statistical model is performed using different methods of combination. It is shown that most forecast systems capture the main features associated with the Guinean regime, that is, rainfall located mainly south of 10ºN and the northward migration of rainfall over the season. On the other hand, only a fraction of the forecast systems capture the characteristics of the rainfall signal north of 10ºN associated with the Sahelian regime. A simple statistical model proves to be of great value and outperforms most state-of-the-art dynamical forecast systems when predicting the principal components associated with the Guinean and Sahelian regimes. Combining all forecast systems do not lead to improved forecasts when compared to the best single forecast system, the ECMWF System 4 (S4). In fact, S4 is far better than any forecast system when predicting the variability of the WAM rainfall regimes several months ahead. This suggests that in some special occasions like this one a multi-model approach is not necessarily better than an especially skillful model.

Key words: West African monsoon, seasonal prediction, operational forecast systems, combination, forecast quality assessment

1. Introduction

Associated with the apparent motion of the sun, the Intertropical Convergence Zone (ITCZ) experiences a latitudinal shift along the year that plays a fundamental role in determining the West African monsoon (WAM) rainfall variability [*Motha et al*., 1980; *Sylla et al*., 2013]. The WAM rainfall variability spans a wide range of timescales, from intraseasonal [*Sultan et al*., 2003] to interdecadal [*Nicholson*, 1993], and is influenced by both local and remote oceanic forcings and associated changes in the atmospheric circulation [*Folland et al*., 1986; *Fontaine et al*., 1995, 1998; *Fontaine and Janicot*, 1996; *Janicot et al*., 1998, 2001; *Joly and Voldoire*, 2009, 2010; *Hourdin et al*., 2010; *Mohino et al*., 2011a, 2011b; *Rodríguez-Fonseca et al*., 2011].

Motha et al. [1980] analyzed long-term rainfall data in Nigeria and found two distinct rainfall patterns. In one of them, rainfall anomalies of opposite signs are observed in the Sahelian and Guinean regions. They suggested that this was associated with the latitudinal migration of the ITCZ such as that above (below) normal rainfall in the Sahelian (Guinean) region is observed when the ITCZ is placed further north of its climatological position. The opposite takes place when the ITCZ does not penetrate into the Sahelian region with its normal intensity. In the second pattern, rainfall anomalies with the same sign are experienced throughout the WAM region. These patterns show a low-frequency modulation of their spatial extent [see *Rodríguez-Fonseca et al*., 2011 for a review].

The two leading modes of WAM rainfall variability, extracted by using principal component analysis (PCA), correspond to the rainfall variability along the Sahelian and

Guinean regions [*Giannini et al*., 2003, 2005; *Tippet and Giannini*, 2006; *Philippon et al*., 2010]. While the Guinean rainfall regime is mostly explained by interannual variations, the variability in the semi-arid Sahelian region occurs mostly on decadal time scales, although interannual variations also play a role in this region, specially linked to the El Niño-Southern Oscillation (ENSO) [*Fontaine et al*., 1998; *Janicot et al*., 2001; *Giannini et al*., 2003, 2005; *Tippet and Giannini*, 2006]. It is worth noting that the internal variability may be important for the WAM rainfall variability, even at decadal time scales [*Caminade and Terray*, 2010].

Forecasting the WAM summer rainfall is of great importance, especially taking into account that a large part of this region employs rain fed agriculture [*Sylla et al*., 2013]. On the other hand, farmers find that forecasting the total amount of seasonal rainfall is of limited usefulness [*Ingram et al*., 2002]. Instead, they would benefit from having information such as the duration and distribution of rainfall over time and space or the timing of the monsoon onset [*Ingram et al*., 2002; *Vellinga et al*., 2013; *Sylla et al*., 2013]. This kind of information has been hardly taken into account in predictability studies over the WAM region. In this study, the seasonal evolution of the WAM summer rainfall is taken into account through the meridional evolution of rainfall from June to October (i.e. one month prior to and one month after the July, August and September (JAS) period). Latitude-time diagrams of longitudinally averaged rainfall are considered as this approach provides a suitable representation of the integrated atmospheric dynamics of the WAM system, which is related to shifts in the local ITCZ [e.g. *Sultan and Janicot*, 2000, 2003; *Sultan et al.*, 2003].

Atmospheric general circulation models (AGCMs) forced with observed SSTs are able to simulate successfully the two WAM rainfall regimes [*Giannini et al*., 2003, 2005; *Tippet and Giannini*, 2006]. However, *Goddard and Mason* [2002] compared the ensemble-mean anomaly correlation simulated and predicted by an AGCM using persisted SST anomalies and found that errors in the predicted SST could lead to a significant degradation of the predictive skill. They showed that the WAM rainfall during the July-August season is one of the most severe examples of this loss of prediction skill. In a different study, *Tompkins and Feudale* [2010] noticed that a dipole bias in the WAM rainfall prediction by the European Centre for Medium-Range Weather Forecasts (ECMWF) climate forecast system 3 (S3), with dry (wet) conditions over the Sahel (Gulf of Guinea). A warm bias in the Equatorial Atlantic SST predictions by S3 would affect the observed northward migration of the ITCZ. When S3 is run with observed SST as boundary forcing the dipole bias disappears, and an overall reduction in rainfall bias is found [*Tompkins and Feudale,* 2010]. Capture the interannual variability of the Equatorial Atlantic SST using simulations from the Coupled Model Intercomparison Project 3 (CMIP3) is still an issue; and consequently, its influence on the rainfall over the Western African continent is hardly reproduced [*Joly and Voldoire*, 2010].

Cook and Vizy [2006] studied the ability of eighteen climate models to simulate the climatology and the dipole mode of WAM variability associated with the meridional migration of the ITCZ. They found that all of them have positive SST bias in the Gulf of Guinea, only ten could simulate the main observed climatological features (e.g., some of the forecast system put the maximum rainfall over the ocean due to the warm SST biases)

and only eight the dipole mode of variability. An analysis of the recently available CMIP5 historical simulations shows that dynamical forecast systems still have substantial SST biases in the Equatorial Atlantic [*Roehrig et al*., 2013]. *Zuo et al*. [2013] used the newest version of the National Centers for Environmental Prediction (NCEP) operational forecast system, the climate forecast system version 2 (CFSv2), to assess the predictability of the modes of interannual rainfall variability of three northern hemisphere monsoon systems: the Asian and Indo-Pacific, the West African and the North American monsoon systems. They found that the low predictability of the principal components (PC) associated with the two main modes of the WAM rainfall variability could be probably due to the link between the WAM and the equatorial Atlantic SST, which is poorly predicted by the CFSv2.

In addition, predictive skill can be negatively affected if the model used to take advantage of SST information does not properly describe the mechanisms responsible for the WAM rainfall [*Krishna Kumar et al*., 2005]. *Im et al.* [2014] used a regional climate model with observations and reanalysis as initial and boundary conditions to show the sensitivity of the WAM rainfall, surface energy balance and circulation to the land surface and convection schemes. They show that predictability of these parameters over the WAM can be significantly improved when the land surface and convection are better represented in the model. *Zuo et al*. [2013] found that the poor representation of land surface processes in the CFSv2 could in part explain the low predictability of this forecast system when predicting the WAM rainfall regimes. Improving the representation of land surface and rainfall processes in dynamical forecast system is difficult and the skill

139 improvement in one region is usually followed by a degradation in another one so that the
140 overall improvement is usually small [*Tompkins and Feudale,* 2010].
141
142 When systematic errors are important, several studies have shown how the combination
143 of several dynamical forecast systems yields on average better deterministic and
144 probabilistic forecast skill than any of the single systems [*Coelho et al*., 2004; *Doblas-*
145 *Reyes et al*., 2005; *Hagedorn et al*., 2005; *Stephenson et al*., 2005; *Batté and Dequé*,
146 2011; *Rodrigues et al*., 2014]. It has been shown that combining statistical and dynamical
147 forecast systems could enhance forecast skill even further [*Coelho et al*., 2004;
148 *Stephenson et al*., 2005]. *Coelho et al*. [2004] and *Stephenson et al*. [2005] used the
149 Forecast Assimilation (FA) technique, a Bayesian method for calibrating and combining
150 several dynamical forecast systems taking into account historical (observed) information,
151 to forecast SST over the Pacific region. The FA technique assigns weights to each
152 forecast system in the combination procedure based on each systems' forecast error (i.e.
153 more weight to forecast systems with less forecast error). *Stephenson et al*. [2005] found
154 that the FA technique could improve forecasts not only over the single systems but also
155 over the simple multi-model (SMM) combination, where all forecast systems are
156 combined assigning equal weights. *Rodrigues et al*. [2014] studied the benefits of
157 combing three operational dynamical forecast systems and a simple statistical model to
158 predict SST over three ocean basins. They found that on average the SMM is better than
159 the single forecast systems and the combination methods that assign weight to each
160 forecast system, including the FA. On the other hand, assigning different weights could

reduce low skill when most forecast systems perform badly, which is typically the case for the WAM precipitation.

Previous multi-model assessments, however, showed limited benefit of merging different sources of information. *Bouali et al*. [2008] found that the Development of a European Multi-model Ensemble System for Seasonal to Interannual Prediction (DEMETER) multi-model system has only modest skill when predicting the Sahelian rainfall. *Philippon et al*. [2010] studied the skill of the ENSEMBLES stream 1 multi-model when forecasting key parameters of the WAM and found that the Guinean rainfall regime could be accurately predicted by these systems, but not the Sahelian regime. *Batté and Dequé* [2011] used the ENSEMBLES stream 2 forecast systems to study the precipitation seasonal forecast skill over Africa and found that the SMM improves on average forecast skill over the single systems. They also found that probabilistic forecasts were more skilful in the Guinean region than in the Sahelian region. *Vellinga et al*. [2013] used several forecast systems, including the ones from the ENSEMBLES project and the UK Met Office operational seasonal forecast system GloSea4, to study the skill of these systems when forecasting the onset of the WAM rainy season. They found that these forecast systems have modest probabilistic skill when forecasting the onset of the Sahelian rainfall. This was attributed to the difficulty of such systems to capture the mean rainfall amount in the Sahel and the influence of a diversity of intraseasonal phenomena that usually have little or no predictability at this timescale.

New aspects of seasonal climate prediction of the WAM are addressed in this study. Firstly, a targeted methodology to assess both the seasonal evolution of the WAM rainfall within a rainy season and its interannual variability simultaneously is considered. Secondly, the two leading modes of the WAM rainfall variability are estimated using the seasonal evolution diagrams over the whole hindcast period. The robustness of the methodology was estimated using two different datasets to assess the uncertainty associated with the observations. Thirdly, several quasi-operational forecast systems were used to estimate the leading modes of WAM rainfall variability. The aim is to assess the ability of the forecast systems to predict the seasonal evolution of the latitudinal migration of rainfall over West Africa. A simple statistical model that uses SST indices as predictors for the WAM rainfall regimes is considered as both a benchmark and an additional forecast system. Finally, several methods of combination are used to combine the dynamical and empirical seasonal predictions.

To illustrate these objectives, the paper is organized as follows. In Section 2, the observations and forecast systems are described. Section 3 describes how the two leading modes of WAM rainfall variability are estimated, the combination of the predictions and the forecast quality assessment. Section 4 presents the results and Section 5 describes the main conclusions.

2. Data and forecast systems

Two observational precipitation datasets have been used in this study: the version 2.2 of the Global Precipitation Climatology Project (GPCP) monthly satellite-gauge combined

[*Huffman and Bolvin*, 2013] and the Global Precipitation Climatology Center (GPCC) version 6.0 monthly gridded gauge analysis derived from quality controlled station data [*Schneider et al*., 2011]. The GPCP dataset has a 2.5º resolution and covers both land and ocean. The period of the GPCP data is from 1979 onwards. On the other hand, the 1º resolution GPCC dataset is available only over land and the period from 1901 onwards. The GPCP dataset is used for the validation of the forecast systems while both datasets are used to assess the observational uncertainty. The GPCC dataset is also used to estimate the linear regression coefficients of the statistical model.

The Extended Reconstructed Sea Surface Temperature analysis version v3b (ERSSTv3b) is used to estimate the SST indices that are used as predictors in the statistical model [*Smith et al*., 2008]. ERSSTv3b is generated using *in situ* SST data and improved statistical methods that allow stable reconstruction using sparse data and it has a 2º resolution. ERSSTv3b covers the period from January 1854 onwards.

An unusually large number of quasi-operational dynamical forecast systems is used in this study, among them two European Seasonal to Interannual Prediction system (EUROSIP) and six North American Multi-Model Ensemble (NMME) forecast systems. Besides, a simple statistical model is used as a benchmark for comparison with the dynamical forecast systems. These systems are described below.

The atmospheric component of the ECMWF climate forecast system 4 (S4) is the cycle 36r4 of the ECMWF Integrated Forecast System (IFS) [*Molteni et al*., 2011; *Kim et al*.,

2012]. It has a horizontal resolution of about 80 km and 91 vertical levels, extending up to about 0.01 hPa. The ocean component of S4, the Nucleus for European Modelling of the Ocean (NEMO) version 3.0, has a horizontal resolution of about 1º with equatorial refinement and 42 vertical levels, 18 of which are in the upper 200 m. S4's hindcasts have 15 ensemble members, all starting in burst mode on the first day of every month at 0 UTC. The simulations are seven-month long and cover the period 1981-2011.

CFSv2 uses the NCEP Global Forecast System (GFS), with horizontal resolution of about 100 km and 64 vertical levels, as its atmospheric component [*Saha et al*., 2013; *Yuan et al*., 2011; *Kim et al*., 2012; *Kirtman et al*., 2013]. Its ocean component is the Geophysical Fluid Dynamics Laboratory Modular Ocean Model version 4 (MOM4) and it has maximum horizontal resolution of 0.25º within 10º of the equator and 0.5º poleward and 40 vertical levels. CFSv2 hindcasts have 24 ensemble members, except those starting in November, which have 28 members. The hindcasts are initialized in different days and times, being the ones initialized after the day 7 used as the lead time zero ensemble members of the next month. For example, the ensemble members for the target month of February at lead time zero have start dates in January 11th, 16th, 21st, 26th, 31st, and the February 5th (at the synoptic times 00, 06, 12 and 18 UTC) of the same year. The simulations are ten-month long and cover the period 1982-2011.

MF3 uses the Action de Recherche Petite Echelle Grande Echelle (ARPEGE) version 4 as its atmospheric component [*Alessandri et al*., 2011]. It has a horizontal resolution of about 300 km and 91 vertical levels, reaching high into the stratosphere. Its ocean

252 component, is the global version of the Océan PArallélisé (OPA) model version 8.2, has
253 horizontal resolution of about 2º and 31 vertical levels. MF3´s hindcasts have 11
254 ensemble members, all starting in burst mode on the first day of every month at 0 UTC.
255 The simulations are seven-month long and cover the period 1981-2011.
256
257 CCSM3 uses the Community Atmosphere Model (CAM) version 3, with horizontal
258 resolution of approximately 150 km and composed of 26 vertical levels [*Kirtman and*
259 *Min*, 2009; *Kirtman et al*., 2013; *Yoshikatsu et al*., 2008], as its atmospheric component.
260 The Parallel Ocean Program (POP) with 1º horizontal resolution and 40 vertical levels is
261 the ocean component [*Yoshikatsu et al*., 2008]. CCSM3 hindcasts have 6 ensemble
262 members, all starting in burst mode on the first day of every month at 0 UTC. The
263 simulations are twelve-month long and cover the period 1982-2011.
264
265 GFDL uses the GFDL Atmospheric Model AM2p12 with horizontal resolution of
266 roughly 200 km and 24 vertical levels as its atmospheric component and the MOM4 with
267 maximum horizontal resolution of about 0.3º near the Equator (1º elsewhere) and 50
268 vertical levels as it ocean component [*Zhang et al*., 2007; *Kirtman et al*., 2013]. The
269 hindcasts have 10 ensemble members, all starting in burst mode on the first day of every
270 month at 0 UTC. The simulations are twelve-month long and cover the period 1982-
271 2011.
272

IRI-ECHAM anomaly and IRI-ECHAM direct use the coupled forecast system described in *DeWitt* [2005] with some updated parameterizations[1]. The atmospheric component is the European Centre-Hamburg model (ECHAM) version 4.5 with horizontal resolution of about 300 km and 19 vertical levels. The ocean component is the GFDL MOM version 3 with zonal resolution of 1.5º and meridional resolution of 0.5º between 10ºS and 10ºN, gradually increasing to 1.5º keeping constant at this value north of 30ºN and south of 30ºS. There are 25 layers in the vertical with 17 layers in the upper 450 m. Both forecast systems produce hindcasts with 12 ensemble members, all of them starting in burst mode on the first day of every month at 0 UTC and are nine-month long. They cover the common period 1982-2011. The difference between the two versions of the IRI system is that the IRI-ECHAM direct employs direct coupling while the IRI-ECHAM anomaly employs anomaly coupling.

CMC2 uses the Canadian Centre for Climate Modelling and Analysis (CCCam) atmospheric circulation model version 4 (CanAM4) as its atmospheric component [*Kirtman et al*., 2013; *Merryfield et al*., 2013]. CanAM4 has a horizontal resolution of about 200 km and 35 vertical levels. The ocean component is the CCCma ocean model version 4 (CanOM4) with horizontal resolution of approximately 100 km and 40 vertical levels. CMC hindcasts have 10 ensemble members, all starting in burst mode on the first day of every month at 0 UTC. The simulations are twelve-month long and cover the period 1981-2011.

[1] http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/.IRI-ECHAM4p5-DirectCoupled/.MONTHLY/.dataset_documentation.html

The statistical model is based on simple linear regression where the predictand is the PC associated with a WAM rainfall regime and the predictor is an SST index. The equations are described in *Coelho et al*. [2004]. The two SST indices that have been considered are described below. The statistical model is estimated in retrospective mode, that is, only the period prior to the target year is used in the estimation of the regression coefficients as in an operational context [*Mason and Baddour*, 2008]. The first training period is 1951-1981 and it is increased by one year at a time. The GPCC and ERSSTv3b datasets are used to estimate the regression coefficients while the GPCP dataset is used for the validation. The May Atlantic 3 (Atl3), i.e. the SST averaged over 20ºW-0ºE/3ºS-3ºN [*Zebiak*, 1993], monthly-mean anomaly is used as predictor for the PC associated with the Guinean rainfall regime and the previous-year December AMO monthly-mean anomaly is used as predictor for the PC associated with the Sahelian regime. The months of May and December of the previous year are the lead time 0 and 4 months, respectively, relative to the target period June through October. They are chosen as predictors because they have the highest correlation (i.e. when only months prior to the target period are considered) with their respective predictands when using the dataset for the calibration of the statistical model. The aim is to emulate an operational forecast system where only months prior to the target period would be considered as predictors. Detailed information about the choice of the predictors is given in the Appendix A.

3. Methods

The deterministic skill of the dynamical forecast systems described above has been assessed at each grid point over the WAM region (22ºW - 22ºE; 0º - 22ºN) to evaluate to

what measure they are able to simulate the total WAM rainfall. The correlation coefficient is used to assess the degree of linear correspondence between the predicted ensemble-mean and observed JAS rainfall. The mean and variability errors have been also assessed on a grid-point basis. All dynamical forecast systems were interpolated into the GPCP grid prior to computing the correlation coefficient and the systematic errors.

A novelty of this work resides in the approach used to assess the forecast quality of the WAM rainfall hindcasts. A targeted methodology to evaluate how the forecast systems predict both the seasonal evolution of the WAM rainfall and its interannual variability simultaneously is considered. In this technique, monthly rainfall is averaged over the 10ºW-10ºE African Monsoon Multidisplinary Analysis transect [e.g. *Hourdin et al.*, 2010; *Losada et al.*, 2010; *Roehrig et al.*, 2013]. Averaging rainfall zonally allows taking into account two relevant features of the WAM variability: the latitudinal migration and the seasonal distribution of the summer rainfall [*Hourdin et al*., 2010]. The latitudinal range of the study extends from the Equator to 20ºN and the period between June and October of each year. The southernmost limit is intended to capture the inland penetration of monsoonal rainfall over the Guinean region, while the northernmost limit tries to capture the Sahelian rainfall, which usually reaches 18ºN in the observations. The period chosen to characterize the seasonal variability is from June to October, where June and October are one month prior to and one month after the JAS target summer season. With the forecast systems analyzed in this paper, the longest forecast time that can be considered is seven months, the longest forecast time of both S4 and MF3. With seven forecast months and covering the period June to October, three start dates can be

considered to estimate the intraseasonal evolution of the WAM rainfall: June (lead 0), May (lead 1 month) and April (lead 2 months). Most users, such as farmers, request receiving information about seasonal rainfall about 1-2 months before the climatological monsoon onset in July, that is, the information should be available in late April or early May [*Ingram et al*., 2002]. For this reason, most figures displayed are for lead time 1, that is, a prediction starting in the first of May or late April (as is the case for the CFSv2 and the statistical model).

The GPCP and GPCC precipitation datasets for the period 1982-2011 and 1951-2011, respectively, are used. The selection of the previous two periods has different motivations. While the year 1982 is the first year available for all the dynamical seasonal hindcasts, 1951 is the year from which a large number of stations are used in the GPCC dataset, and therefore, makes it a more trustworthy period for this dataset [*Schneider et al*., 2011]. The GPCP dataset, which starts in 1979, is also used with a mask over the ocean for comparison with the GPCC data, which has values only over land. GPCP is considered only from 1982 to agree with the common period of the hindcasts, 1982-2011. Climatologies for the observed rainfall are hence computed with four different samples: GPCP 1982-2011, GPCP land-only 1982-2011, GPCC 1951-2011 and GPCC 1982-2011. The systematic error of the predicted rainfall is computed against GPCP for the period 1982-2011 and for the three start dates, April, May and June.

Principal component analysis [PCA; *Wilks*, 2006] of the covariance matrix is performed upon the observed and predicted zonally averaged rainfall to estimate the leading modes

of WAM rainfall variability. The three-dimensional data matrix used to estimate the covariance matrix contains the longitudes, the months from June to October and the number of years, which will identify the modes of interannual variability taking into account at the same time the seasonal variability. For the hindcasts, the third dimension is the number of ensemble members times the number of years. The anomalies, estimated for both observations and forecast systems prior to applying the PCA, were computed using three-year out cross-validation (i.e., the target year and the years prior to and after it were removed prior to computing the climatology instead of just the target year) to avoid artificial skill in the forecast quality assessment [*Mason and Baddour*, 2008]. The leading modes of the WAM rainfall variability are described as a set of spatial patterns (empirical orthogonal functions, EOFs) and associated standardized time series (PCs) that are associated to specific modes of variability. PCA is performed upon the observations and each forecast system and lead time separately to take into account the fact that the hindcasts might represent the variability in a way different to the observations, while this representation also depends on the lead time [*Doblas-Reyes et al*., 2003; *Philippon et al*., 2010].

Most decision makers need a reliable probabilistic prediction instead of a set of forecasts performed by either statistical or dynamical methods to take action [*Doblas-Reyes et al*., 2013]. Therefore, an important aim of this work is to combine the dynamical forecast systems and the statistical model described above to estimate a single forecast for the WAM rainfall regimes. In order to understand the benefits of the combination on the forecast quality of the WAM rainfall variability modes, the forecast PCs were combined

using two different approaches. In the first approach, all forecast systems are combined with equal weight; hereinafter this method is referred to as the simple multi-model (SMM). In this combination, the ensemble-mean prediction is the arithmetic mean of the ensemble mean of all forecast systems, whereas the probabilistic prediction is the arithmetic mean of the probabilistic predictions of all forecast systems. The second combination approach assigns weights to each forecast system based on their past performance. The method used to assign the weights is the FA [*Stephenson et al*., 2005]. The FA is a Bayesian method for calibrating and combining predictions from several sources with prior (historical) empirical information. The FA was applied using either the climatology (FAC) or the statistical model (FAS) as the prior information. A thorough description of these combination methods is available in *Rodrigues et al*. [2014] and references therein. The combinations are performed using three-year out cross-validation to avoid artificial skill in the forecast quality assessment [*Mason and Baddour*, 2008]. It is important to bear in mind that the FA method is expected to provide reliable predictions as it also calibrates them, that is, it provides probabilistic predictions of a specific event whose average frequency of actual occurrence equals the probability.

The forecast quality assessment where the predicted and observed values are compared is an important step in climate prediction. It assesses to what measure the combination of different forecast systems leads to an improved forecast or if a forecast system improves when compared to previous versions. Due to the high dimensionality of the problem of forecast verification, it is very important to take into account multiple verification measures to obtain richer and more robust conclusions about the quality and/or value of

410 the forecast systems [*Murphy*, 1991]. The correlation coefficient between the predicted
411 ensemble mean, as a deterministic prediction, and observed PCs for each forecast system
412 and lead time is computed.
413
414 Several measures have been used to assess the quality of the probabilistic predictions. A
415 generalized version of the Brier Skill Score (BSS) and its reliability and resolution
416 components are used to assess the forecast quality for binary events [*Stephenson et al*.,
417 2008; *Doblas-Reyes et al*., 2009]. The two binary events are the probability of the WAM
418 rainfall regime being above the median and the upper quartile of the climatological
419 distribution. The generalized version of the BSS takes into account the within-bin
420 variance of the forecasts and the within-bin covariance between forecasts and
421 observations to make the BSS components less sensitive to the arbitrary number of bins
422 used in the BSS decomposition [*Stephenson et al*., 2008].
423
424 The other probabilistic scores used are the Continuous Ranked Probability skill score
425 (CRPSS) and the ignorance skill score. The CRPS is defined as the integrated squared
426 difference between the cumulative forecast and observation distribution [*Jolliffe and*
427 *Stephenson*, 2012]. It is defined on a continuous scale so that there is no need to reduce
428 forecasts into discrete probabilities of binary or categorical events as in the BSS. On the
429 other hand, reducing the forecasts into discrete probabilities using binary events is
430 important to verify how the forecast systems behave when predicting different binary
431 events. The CRPS can be computed in two ways: a) when an ensemble of forecasts is
432 available as in the case of the dynamical forecast systems, the CRPS is estimated using a

433 frequentist approach as in *Hersbasch* [2000] and *Jolliffe and Stephenson* [2012], and b)
434 when the predicted mean and standard deviation are available as in the case of the
435 statistical model and the combination of all systems, the CRPS is computed assuming that
436 the cumulative distribution function (CDF) is Gaussian as in *Gneiting et al*. [2005]. The
437 ignorance score is the negative of the logarithm of the predictive density function at the
438 verifying value [*Gneiting et al*., 2005; *Jolliffe and Stephenson*, 2012]. The ignorance
439 score is computed either by assuming a Gaussian probability density function (PDF) as in
440 the case of the statistical model and the combinations and using a generic kernel density
441 estimate when an ensemble of forecasts is available.
442
443 The reference forecast for all these scores is the climatological forecast defined as the
444 climatological PDF estimated from the historical observations. The climatological PDF is
445 used to estimate the median and the upper quartile necessary as thresholds to compute the
446 BS of the climatological forecast or the CDF used to estimate the ignorance score of the
447 climatology. A particular verification dataset is just one of many possible samples from a
448 population and therefore verification measures need to be shown together with an
449 indication of the sampling uncertainty [*Jolliffe and Stephenson*, 2012]. The sampling
450 uncertainty in the verification measures is quantified using 95% confidence intervals
451 [*Nicholls*, 2001; *Mason*, 2008; *Jolliffe and Stephenson*, 2012]. The only exception to this
452 is the grid-point correlation coefficient displayed as a map where the use of confidence
453 intervals would result in a very complex map [*Nicholls*, 2001]. In this case, p-values are
454 used to quantify the sampling uncertainty. Both the confidence intervals and the p-values
455 are estimated using a non-parametric bootstrap method [*Mason*, 2008; *Jolliffe and*

456 *Stephenson*, 2012]. In this procedure, the forecast-observation pairs are randomly
457 resampled with replacement, keeping the forecast and observation pairs together [*Mason*,
458 2008]. The bootstrap size is chosen to be 1,000. From these 1,000 resamples, the 2.5%
459 and 97.5% quantiles, which represent the lower and upper confidence interval limits,
460 respectively, are estimated. On the other hand, the null hypothesis used to estimate the p-
461 values is that the correlation coefficient is zero, while the alternative hypothesis is that
462 the correlation coefficient is larger than zero (i.e. one-tailed test).
463
464
465 **4. Results**
466 Figure 1 shows the correlation between the predicted ensemble mean and the observed
467 JAS rainfall at each grid point over the WAM region for the period 1982-2011. The
468 correlation is computed for all dynamical forecast systems at lead time 2 months (i.e.
469 predictions starting in May). The aim is to assess the ability of these systems to predict
470 the spatial distribution of the seasonal WAM rainfall. S4 shows positive correlations in
471 almost all grid points at the three start dates analysed (results for the two start dates of
472 June and April are not shown), most of them with p-values smaller than 0.10. On the
473 other hand, CFSv2 has low, and in several instances, negative correlation values. Most of
474 the positive correlation values in this forecast system appear north of 10ºN, over the
475 Sahel. Correlation values below 0.1 are found more often in the region south of 10ºN.
476 MF3 also has low correlation skill when compared to S4, but contrary to CFSv2, most of
477 the positive correlation appears south of 10ºN in the Guinean region over the longitudinal
478 range 20ºW-10ºE. CCSM3 performs generally worse than the previous three forecast

systems but, as MF3, it performs better over the Guinean region. GFDL performs well in almost all grid points and lead times, except for the Guinean region at lead time 3 (i.e. predictions starting in April), where it has correlation values below 0.1 more often than above it (not shown). As in the CFSv2 case, GFDL performs better over the Sahel than over the Guinean region. The IRI-ECHAM systems perform poorly over the WAM region, especially over the Sahel where they have negative values more often than positive ones. CMC2 shows positive correlations all over the WAM region at the three start dates analyzed (i.e., lead times 1, 2 and 3 months).

Figure 1 illustrates that S4 has the highest overall correlation skill at all lead times, followed by CMC2 and GFDL, respectively. This is a feature that will appear in many other of the diagnostics described in this paper. S4 seems to represent a leap forward in the seasonal prediction of the WAM precipitation with respect to previous versions of this system and to other contemporaneous operational systems. This leap forward can be measured when compared with the performance of the previous ECMWF forecast system, which had similar skill to other European systems [*Batté and Déqué*, 2011]. The grid-point correlation over the WAM region does not substantially change with lead time in any of the forecast systems.

S4, GFDL and CMC2 have smaller mean systematic errors when compared to the other systems (not shown). Therefore, even though a direct link between mean biases and forecast skill could not be established, one can expect that improving the physical processes that are at the origin of the model drift and the systematic error, and that

hamper the conversion of predictability into skill, could lead to improvements in forecast quality [*DeWitt,* 2005]. *Molteni* et al. [2011] found that S4 has improved the simulation and prediction of ocean/atmosphere variability in the tropical Atlantic and adjacent regions when compared to S3, which could benefit the prediction of the WAM precipitation. They highlight that some of the improvements S4 has achieved when compared to its predecessor might be due to higher horizontal and vertical resolution, a more accurate initialization of the land-surface variables, and improved physical parameterization, among other reasons. In fact, it is observed that the higher the model resolution of a system is, the smaller its biases are, with CCSM3 being the only exception to this simple rule. However, as pointed out by *Kirtman* et al. [2013], CCSM3 is generally worse when compared to the other NMME systems in terms of root mean square error (RMSE) of the tropical SST for September start dates at leads 0-5 months. As a consequence, it is planned to be replaced by CCSM4 in the second phase of the NMME project [*Kirtman et al*., 2013]. On the other hand, not always a small bias leads to a high correlation. For instance, CFSv2 shows a relatively small bias, while at the same time it has low correlation.

The WAM rainfall displays a strong monthly variability, which is illustrated by considering the latitudinal migration of the zonally-averaged rainfall between the months of June and October. Figure 2 shows the climatology of monthly-mean rainfall averaged over 10ºW- 10ºE and displayed over the latitudes between the Equator and 20ºN. The climatology is computed using the GPCP dataset for the period 1982-2011 (upper left panel), GPCP after applying a mask over the ocean for the same period (upper right

525 panel), GPCC for the same period (lower right panel) and the GPCC dataset for the
526 period 1951-2011 (lower left panel). The climatologies of the zonally averaged monthly
527 rainfall have similar patterns in both GPCP and GPCC. They show a northward migration
528 of the rainfall that reaches its northernmost position at 18ºN in July and August, moving
529 southward later in the year. Some differences between GPCC and GPCP can be found
530 over the common period. These differences already point at the observational uncertainty
531 of the WAM precipitation.
532
533 Every dynamical forecast system successfully simulates the meridional shift of the
534 rainfall for the three lead times analyzed (not shown). However, they all fail in simulating
535 the correct position and magnitude of the rainfall maxima and therefore have substantial
536 biases, suggesting that these systems do not fully reproduce the physical processes
537 associated with the WAM rainfall. As an illustration, Figure 3 shows the systematic error
538 of the dynamical forecast systems at lead time 1. CCSM3 has a larger bias than the other
539 forecast systems. It not only fails to simulate the rainfall maxima in August but it is the
540 only forecast system that simulates rainfall above 2 mm/day north of 18ºN. MF3 also has
541 substantial biases. In particular, it has a positive bias south of 10ºN and negative north of
542 it indicating that in this forecast system the ITCZ does not penetrate as far north as in the
543 observations, which creates a dipole-like bias pattern (i.e. excessive precipitation at lower
544 latitudes and a deficit at higher latitudes). This pattern is also observed in S4 and CFSv2
545 but with smaller magnitude when compared to MF3. The IRI-ECHAM systems and
546 CMC2 have a dipole-like pattern with inverse sign (i.e. excessive precipitation at higher
547 latitudes), while GFDL has a positive bias overall. The forecast systems could be ranked

548 in decreasing order of the mean bias (i.e. sum of the mean bias over the whole domain) at
549 lead time 1 to give S4, CMC2, CFSv2, IRI-ECHAM direct, GFDL, MF3, IRI-ECHAM
550 ano and CCSM3. As it was also found in the analysis without the longitudinal averaging
551 (not shown), the systems with lower (higher) systematic errors are the systems with
552 higher (lower) resolution, CCSM3 being the only exception to this.
553
554 The two leading modes of the observational WAM rainfall, obtained with the PCA
555 method described above, are shown in Figure 4. The aim of the longitudinal averaging
556 applied to the data prior to the PCA is to concentrate in both the latitudinal migration and
557 the seasonal distribution of the WAM rainfall. The first EOF (EOF1) in the GPCP dataset
558 shows positive values south of 10ºN, in the Guinean region, while the second EOF
559 (EOF2) shows positive values north of 10ºN, in the Sahelian region. The variance
560 associated with these two EOFs is 29% and 23%, respectively (Table 1). This is in
561 agreement with the WAM patterns described in the literature using different
562 methodologies [*Motha et al*., 1980; *Fontaine et al*., 1995; *Fontaine and Janicot*, 1996;
563 *Janicot et al*., 1998; *Giannini et al*., 2003, 2005; *Mohino et al*., 2011b; *Rodríguez-*
564 *Fonseca et al*., 2011]. The same analysis has been performed on the GPCP dataset after
565 applying a mask over the ocean and on the GPCC dataset with a common period 1982-
566 2011 and an extended period 1951-2011 to assess the observational uncertainty. The
567 GPCP land-only and the GPCC datasets have a reverse order of the leading modes when
568 compared to the GPCP land-ocean (Figure 4). This reverse pattern when land-ocean and
569 land-only data are used has been documented previously [*Giannini et al*., 2005]. This
570 reversal is probably due to the variance maximization of inland precipitation, where the

latitudinal migration from the ocean into the Guinean region early in the season is not considered. The variance explained by these two EOFs varies, being 31% (EOF1) and 24% (EOF2) in the GPCP land-only, 27% and 20% in the GPCC, and 30% and 18% in the GPCC for the common period 1982-2011. The difference between the smallest and largest values of the variance explained in the observational datasets is 4% and 6% for the first and second EOF, respectively. As previously with the mean bias, this uncertainty in the observations will be taken into account when interpreting the EOFs from the hindcasts.

To illustrate that the Guinean regime is captured in the EOF1 (EOF2) when the dataset have values over both land and ocean (land only) and vice-versa for the Sahelian regime, the PCs associated with these EOFs are displayed in Figure 5. The first PC (PC1) of the GPCP dataset is highly correlated with the second PC (PC2) of the GPCP land-only and GPCC datasets, and vice versa (see figure caption). The GPCC PCs show that the Guinean regime is characterized mainly by interannual variability while the Sahelian regime is associated with substantial interdecadal variations, although interannual variations also play a role in the latter as described in previous studies [*Fontaine et al*., 1998; *Giannini et al*., 2003, 2005]. The PCA has been also performed over the full spatial field (i.e. without longitudinal averaging) of the GPCP JAS rainfall with longitudes 10ºW-10ºE and latitudes between the Equator and 20ºN. The aim is to compare the modes of variability of the WAM rainfall computed in a conventional way by applying PCA on the seasonally-averaged spatial field with the ones computed by applying the PCA on the longitudinally-averaged seasonal evolution diagrams shown above. The first

and second EOFs of the JAS full spatial field are also associated with the Guinean and Sahelian regimes, respectively (not shown). The lower panels of Figure 5 show the PCs associated with the Guinean and Sahelian regimes estimated by applying the PCA on both the full spatial field and the seasonal evolution diagram. As expected, the PCs are highly correlated in both cases, being the correlation 0.91 for the Guinean regime and 0.90 for the Sahelian regime. Even so, the zonally-averaged rainfall approach allows a better characterization of the intraseasonal evolution of the rainfall regimes because the rainy seasons associated with the two modes are not simultaneous.

The first EOF of the dynamical forecast systems reproduces the overall features associated with the observed Guinean regime, as they locate the positive values south of 10ºN and capture the northward migration of the rainfall (Figure 6 illustrates the results for lead time 1 month). This is similar to what is found in the GPCP land and ocean dataset. However, the forecast systems fail to simulate the accurate magnitude and location of the maxima of the observed EOF, and some of the forecast systems even reproduce a pattern different to the one found for the observations in Figure 4. S4's EOF1 closely resembles the GPCP EOF1 pattern. The variance explained by S4's EOF1 varies considerably with lead time, from 25% at lead time 0 (underestimated when compared to GPCP) to 34% and 41% at lead times 1 and 2 months, an important overestimation when compared to all the observational estimates (Table 1). This could be explained by the fact that S4 underestimates (overestimates) the Guinean rainfall at lead time 0 (2) months with respect to GPCP. This is likely due to the increasing SST bias with forecast time in the Equatorial Atlantic [*Doblas-Reyes et al.*, 2013]. CFSv2 also captures well the Guinean regime's pattern, albeit overestimates the role of the rainfall in September and

618 October. MF3 captures the anomalous rainfall in June, July and August as in the GPCP
619 dataset, but overestimates it in several latitudes and target months. Surprisingly, despite
620 its large systematic errors (Fig. 3), CCSM3 captures the rainfall evolution anomaly in
621 June, July and August, but overestimates the duration of the anomalous rainy season. In
622 addition, CCSM3 overestimates the variance explained by the EOF1 at all lead times and
623 has the largest difference when compared to GPCP (Table 1). GFDL generally
624 overestimates the rainfall anomalies, but differently from the previous forecast systems it
625 yields rainfall above 10ºN and in several latitudes in the target months of September and
626 October. Both IRI systems place the rainfall maximum in June and thus, overestimate the
627 rainfall at this target month. IRI-ECHAM anomaly underestimates the observed rainfall
628 anomaly maxima in July and August and overestimates the rainfall latitudinal extent later
629 in the season, while IRI-ECHAM direct simulates better than IRI-ECHAM anomaly the
630 rainfall maxima, but overestimates the signal in September and October. The IRI-
631 ECHAM systems overestimate the variance explained by the first EOF at all lead times,
632 except for the IRI-ECHAM anomaly at lead time 2 (Table 1). CMC2 generally
633 underestimates the amplitude of the pattern, although it shifts the pattern north of 10ºN,
634 contrary to what is found in GPCP. CFSv2, MF3 and CMC2 underestimate the variance
635 explained by the first EOF at all lead times.
636
637 Contrary to the Guinean regime, the Sahelian regime is only well simulated by S4, yet the
638 amplitude of the pattern is generally underestimated when compared to GPCP. CFSv2
639 captures the pattern north of 10ºN, but gives an unrealistic pattern with a signal of
640 opposite sign south of 10ºN in August. MF3 also captures the pattern north of 10ºN in

July and August, but has a pattern of opposite sign in June. CCSM3 completely fail to simulate any signal north of 10ºN. GFDL captures the pattern in the Sahelian region in August, but shows a pattern of similar sign in June and of opposite sign in October, which are not found in the GPCP pattern. Both IRI-ECHAM systems completely fail to capture the Sahelian regime. CMC2 captures the Sahelian signal but, as other systems do, also simulates a pattern of opposite sign south of 10ºN. All forecast systems underestimate the variance explained by the second EOF when compared to GPCP EOF2 at the three lead times (Table 1), which is supposed to be related to the problems all the systems have to timely shift the precipitation over the Sahel during the rainy season.

Figure 7 illustrates the indices for the Guinean rainfall regime predicted by the statistical model, the dynamical forecast systems and their combinations. The predictions shown are for lead time 1 month (i.e. predictions starting in May). Several deterministic and probabilistic scores are also displayed. The zero line is shown for reference. The statistical model, which is based on the May Atl3 index as predictor, captures well the interannual variability associated with the Guinean regime. The correlation coefficient of the statistical model is the third largest among the single forecast systems, being outperformed only by S4 and MF3 and it is one of the few systems that has a positive BSS. In addition, the statistical model outperforms all forecast systems and combinations in terms of reliability skill score when predicting the Guinean rainfall above the median at lead times 1 (Figure 7) and 2 months (not shown). This illustrates that simple linear regression models are still difficult to beat by state-of-the-art dynamical forecast systems, especially in the tropical Atlantic basin.

Following on its excellent representation of the Guinean rainfall spatial-temporal pattern, S4 captures the interannual variability associated with the Guinean regime and its ensemble-mean correlation is 0.66. S4 is also skillful probabilistically, with most of the observations falling inside the 95% predicted interval (a sign of reliability). The resulting positive BSS values are among the three largest for the two binary events described in this study. Additionally, in most cases it shows the best resolution skill score for the Guinean regime above the median and upper quartile at the three lead times. MF3 has lower skill than S4, but still shows a high ensemble-mean correlation, while the BSS ranges between negative values (event above the median) and low positive ones (0.18 for the event above the upper quartile). CFSv2 and GFDL have positive correlation of 0.26 and 0.25, respectively, but no positive skill in terms of BSS. Finally, CCSM3, the IRI-ECHAM systems and CMC2 have no deterministic or probabilistic skill when predicting the Guinean regime with 1 month lead time.

It has been considered difficult to improve the SMM forecasts using combination methods that assign different weights to the forecast systems based on the past performance [*DelSole et al*., 2012]. In this case, when the different models are brought together, the SMM performs worse than the weighted combinations FAC and FAS.. This can be explained because weighting methods can provide more skillful forecasts than the SMM when most systems perform badly and there is a small subset that stands out [*Rodrigues et al*., 2014]. When comparing with all the forecast systems available, the FAC has the best correlation coefficient (Figure 7), which is slightly higher than S4 and

687 FAS. On the other hand, both FAC and FAS are outperformed by S4 in terms of BSS,
688 reflecting the difficulty that combination methods have to conserve the forecast
689 resolution when producing more reliable predictions.
690
691 A summary of the forecast quality measures for both the Guinean and Sahelian regimes
692 and the three lead times considered can be found in Figure 8. The statistical model has
693 only one correlation value for each WAM regime (the correlation does not vary with lead
694 time) as it takes advantage of using the best SST predictor for each regime (see Appendix
695 A for detailed information). Interestingly, a statistical model based on simple linear
696 regression still provides useful information and beats most of dynamical forecast systems
697 when predicting the Guinean and the Sahelian regime. Only S4 and MF3 outperform the
698 statistical model when predicting the Guinean regime, and S4 and GFDL (for lead time
699 0), S4, GFDL, IRI-ECHAM direct, and CMC2 (for lead time 1) and only S4 (for lead
700 time 2) when predicting the Sahelian regime.
701
702 S4 has the highest correlation when predicting both rainfall regimes at all lead times, with
703 two exceptions in the prediction of the Guinean regime: FAS is the best at lead time 0,
704 while FAC is the best at lead time 1 (Figure 8). As mentioned above, S4 has improved
705 when compared to its predecessor when predicting the WAM variability [*Molteni et al*.,
706 2011]. S4 has correlation above 0.6 in all cases, except for the Guinean regime at lead
707 time 2 months. Interestingly, the S4 correlation for the Sahelian regime does not vary
708 much with lead time. MF3 (GFDL) is only competitive when predicting the Guinean
709 (Sahelian) regime with average correlation of about 0.45. On the other hand, CFSv2 has

no skill when predicting the Guinean regime and low correlation when predicting the Sahelian regime. CCSM3, CMC2 and both IRI-ECHAM systems perform generally worse than the other dynamical forecast systems. As pointed out previously, the SMM usually outperforms unequal methods of combination when most single forecast systems have skill, as in the Sahelian regime at lead time 1. The opposite would happen when only a fraction of the forecast systems have skill as in most cases in Figure 8. However, in this study S4 is an outlier when predicting the WAM rainfall variability modes as this system is far better than any other single forecast system. Therefore, combining it with the other forecast systems will hardly improve the forecast quality of the WAM rainfall regimes.

Formulating skilful and reliable probabilistic predictions, which are the main requirements for decision making [*Jolliffe and Stephenson*, 2012; *Doblas-Reyes et al*., 2013], is still an issue for most of the forecast systems analyzed here for the WAM rainfall regimes (Figure 7, 9 and 10). S4 has the best probabilistic prediction in terms of BSS (considering the events "rainfall regime above the median" or "above the upper quartile"; not shown), the CRPSS (Figure 9) and the ignorance skill score (Figure 10) more often than not. S4 is clearly an outlier as it is the only forecast system that has skill in terms of CRPSS and ignorance skill score. Another outlier is the CCSM3, which is the worst forecast system in almost all cases. Two reasons could explain this behavior in CCSM3 concerning the probabilistic scores: the small number of ensemble members (six members), which makes its forecasts overconfident, and the low accuracy and large systematic error, as described above (Figure 7). As in the case of the correlation

coefficient, the negative skill of most forecast systems makes the combinations to perform worse than S4 alone.

5. Summary and conclusions

A targeted methodology to assess the year-to-year variations of the WAM rainfall variability has been illustrated in this paper. This method estimates the main regimes of the WAM rainfall using monthly data averaged over 10ºW-10ºE covering the latitudes between the Equator and 20ºN and the period from June to October. The aim of the longitudinal averaging is to take into account the latitudinal migration and temporal distribution of the summer rainfall over the WAM region. This approach represents a process-oriented assessment of both the variability and predictability of the ITCZ-related WAM rainfall. Principal component analysis (PCA) is applied on the seasonal evolution diagrams to estimate the leading modes of the WAM rainfall variability. PCA is performed upon the observations and each forecast system and lead time separately to take into account the fact that the hindcasts might represent the variability in a way different to the observations, while this representation also depends on the lead time [*Doblas-Reyes et al*., 2003; *Philippon et al*., 2010]. The spatial patterns (EOFs) and the associated time series (PCs) related to the leading modes are used to describe the WAM rainfall regimes.

Two observational datasets (GPCP and GPCC) and a large number of quasi-operational forecast systems, among them the two systems from the EUROSIP initiative and six from the NMME project, are used in this research. The aim of using two different

observational datasets is twofold: firstly, to assess the observation uncertainty, and secondly, to build a statistical model using a dataset different from the one used for the forecast quality assessment. A simple statistical model built in retrospective mode as in an operational context [*Mason and Baddour*, 2008] is also used to predict the PCs associated with the Guinean and Sahelian regimes. Another aim of this research is to combine all the dynamical forecast systems and the statistical model to provide a single source of forecast information, something needed by the stakeholders [*Doblas-Reyes et al*., 2013].

The forecast systems are combined using combination methods with both equal and unequal weights. In the first case the predicted mean of each forecast system is averaged assigning equal weights to the forecast systems (i.e. simple average of the predicted mean). The second way of combining the forecast systems consists in assigning a larger weight to the systems that have smaller errors. The FA method [*Coelho et al*., 2004; *Stephenson et al*., 2005] is used to assign the weights. Finally, a forecast quality assessment is performed upon both combinations and forecast systems. Several deterministic and probabilistic verification scores have been used to take into account the high dimensionality of the forecast quality assessment [*Murphy*, 1991; *Jolliffe and Stephenson*, 2012]. To the best of our knowledge, this work offers an unprecedented probabilistic evaluation of the seasonal prediction forecast quality of the WAM rainfall variability.

778 The main results of this study, which are innovative for the use of a large set of forecast
779 systems and the way the seasonal variations of the WAM rainfall have been taken into
780 account, are:

781 • As in previous studies [*Motha et al*., 1980; *Fontaine et al*., 1995; *Fontaine and*
782 *Janicot*, 1996; *Janicot et al*., 1998; Giannini et al., 2003, 2005; *Mohino et al*.,
783 2011b; *Rodríguez-Fonseca et al*., 2011], the two leading modes of the WAM
784 rainfall variability are associated with the Guinean and Sahelian rainfall regimes.
785 The Guinean and Sahelian regimes appear in the EOF1 and EOF2, respectively,
786 when data are available over land and ocean (i.e. GPCP and the dynamical
787 forecast systems). The Guinean (Sahelian) regime is found in the EOF2 (EOF1)
788 when the data are available only over land (GPCP after applying a mask over the
789 ocean and GPCC). For the common period 1982-2011, the variance explained by
790 the Guinean mode varies from 29% (GPCP) to 20% (GPCC) and by the Sahelian
791 mode from 31% (GPCP land-only) to 23% (GPCP) (Table 1).

792 • The PCs associated with the Guinean and Sahelian regimes estimated from GPCP
793 are highly correlated with the ones estimated from GPCC. In addition, the PCs
794 associated with the Guinean and Sahelian regimes estimated using a more
795 traditional way, i.e. by applying a PCA on the spatial rainfall field, are highly
796 correlated with the ones used in this study (Figure 5). This suggests that the
797 seasonal variability does not modify the interannual nature of these regimes and
798 that the substantial observational uncertainty is not as large as to substantially
799 modify the characteristics of these regimes. The innovative component of the

800 analysis presented in this paper is that the modes offer information about the
801 intraseasonal variations of the rainfall regimes.

802 • Most forecast systems capture the main features associated with the Guinean
803 regime (EOF1), that is, rainfall located south of 10ºN and the seasonal northward
804 migration of rainfall. However, they are all biased and several of the forecast
805 systems simulate the rainfall anomalies in the wrong location. On the other hand,
806 only a fraction of the forecast systems capture the rainfall signal north of 10ºN
807 associated with the Sahelian regime as observed in the GPCP dataset (EOF2).

808 • A fraction of the forecast systems have significant positive correlation (i.e., when
809 the lower limit of the confidence interval is above zero) between the predicted
810 mean and observed PC associated with the WAM regimes. However, only S4 has
811 significant correlation when predicting both WAM regimes. MF3 performs well
812 when predicting the Guinean regime and GFDL when predicting the Sahelian
813 regime. The deterministic and probabilistic forecast quality assessment show two
814 outliers: S4 and CCSM3. On the one hand, S4 is clearly the best forecast system
815 for all scoring measures in most occasions. On the other hand, CCSM3 is clearly
816 the worst system in most cases. Not surprisingly, it is shown that CCSM3 has the
817 largest rainfall systematic errors over continental West Africa (Figure 3). CCSM3
818 has been identified as an outlier when compared to other NMME forecast systems
819 in terms of root mean square error (RMSE) of tropical SST for September start
820 dates and will be replaced by CCSM4 in the next phase of the NMME project
821 [*Kirtman et al*., 2013].

- The simple statistical model outperforms several state-of-the-art dynamical forecast systems when predicting the PCs associated with the Guinean and Sahelian regimes (Figure 8). This result emphasises the importance of using empirical benchmarks to compare with the dynamical forecast systems, particularly in an operational context.
- Combining all forecast systems do not lead to improved forecasts when compared to the best single forecast system, S4. In fact, S4 is far better than any forecast system when predicting the WAM rainfall regimes. This suggests that in some occasions, a multi-model approach is not necessarily better than an especially skillful model that is clearly identified.

Apart from showing that current operational or quasi-operational seasonal forecast systems can skillfully and reliably predict the interannual variations of the WAM rainfall regimes, which is an important result for the emerging climate services, the example described here illustrates that not always the SMM should be the preferred option in seasonal prediction. S4 is clearly the best forecast system when predicting both WAM rainfall regimes. The equal-weighting combination, with much lower skill than S4, does not improve the forecast quality of the resulting multi-model. At the same time, the two unequal-weighting combination approaches used here also do not improve the quality of the predictions with respect to S4. This suggests that the multi-model approach should not be automatically considered the best option in a prediction context and that a detailed analysis of the single systems should be carried out in each specific instance. Furthermore, given the important investment in model and initial-condition development

undertaken by ECMWF, it is clear that multi-model predictions will only improve if a sufficient number of single systems are continuously improved.

Appendix A

Figure A1 shows the correlation coefficient between indices of the Guinean and Sahelian regimes and three SST indices for all months of the year for the period 1951-2011. SST indices representing the main SST variability over several ocean regions are obtained via spatial averaging. These indices represent the main patterns of climate variability and are widely used as predictive tools in statistical models [*Doblas-Reyes et al*., 2013]. The Equatorial Pacific, North and Equatorial Atlantic ocean basins are known to play an important role on the WAM rainfall variability [*Folland et al*., 1986; *Fontaine and Janicot*, 1996; *Fontaine et al*., 1998; *Joly and Voldoire,* 2009, 2010; *Mohino et al*., 2011a, 2011b; *Rodríguez-Fonseca et al*., 2011]. Therefore, the Niño3.4 (SST anomalies averaged over 170ºW-120ºW and 5ºS-5ºN; *Trenberth*, 1997), the Atlantic Multidecadal Oscillation (AMO; SST anomalies averaged over 80ºW-0ºW and 0º-60ºN minus global SST anomalies over 60ºS-60ºN; *Trenberth and Shea*, 2006) and the Atlantic 3 (Atl3; SST anomalies averaged over 20ºW-0ºW and 3ºS-3ºN; *Zebiak*, 1993) indices are used. The SST over other regions, such as in the Mediterranean basin that might also play a role on the WAM rainfall variability [*Fontaine et al*., 2010], is not taken into account for the sake of simplicity. The correlation between the rainfall regimes and the SST indices is computed using the period 1951-2011. The Niño3.4 SST index is not well correlated either with the Guinean or the Sahelian regime (the maximum absolute correlation values are 0.31 and 0.25, respectively). This might be either because the Niño3.4-WAM rainfall

relationship is not stationary [*Mohino et al*., 2011b; *Rodríguez-Fonseca et al*., 2011] or because not all ENSO events can be linked to WAM rainfall anomalies [*Joly and Voldoire*, 2009]. The time series associated with the AMO (black line) and the Atl3 (blue line) are almost of opposite sign when comparing the Guinean (left panel) and the Sahelian (right panel) regimes. As shown previously, there is positive correlation between the Atl3 and the Guinean regime [*Joly and Voldoire*, 2010] and the AMO and the Sahelian regime [*Mohino et al*., 2011a; *Rodríguez-Fonseca et al*., 2011]. Therefore, we used the Atl3 and the AMO as predictors for the Guinean and Sahelian regimes, respectively. The PCs associated with the WAM rainfall regimes are computed for the target months between June and October (see Section 2 for detailed information) and, as a consequence, only the months prior to June of the target year may be considered as predictors when trying to mimic an operational forecasting approach. Figure A1 shows that the best predictor for the Guinean regime is the Atl3 of May of the target year while the best predictor for the Sahelian regime is the AMO of December of the year prior to the target year.

Acknowledgements

This study was supported by the Spanish MINECO-funded RUCSS project (CGL2010-20657), the European Union's FP7-funded QWeCI (GA 243964) and SPECS (GA 308378) projects, and the Catalan Government. J. G-S was additionally supported by the European Union's FP7-funded NACLIM (GA 308299) project. The authors are grateful to all the institutions that performed the hindcasts for making their data available. We would like to thank Belén Rodríguez-Fonseca (UCM/IGEO, Madrid, Spain) for the

original discussions about applying the PCA to longitudinally-averaged WAM rainfall. The authors would also like to thank three anonymous reviewers for the useful comments and suggestions.

References

Alessandri, A., A. Borrelli, A. Navarra, A. Arribas, M. Déqué, P. Rogel, and A. Weisheimer (2011), Evaluation of Probabilistic Quality and Value of the ENSEMBLES Multi-model Seasonal Forecasts: Comparison with DEMETER, *Mon. Wea. Rev*., 139, 581-607, doi: 10.1175/2010MWR3417.1.

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

Bouali, L., N. Philippon, B. Fontaine, and J. Lemond (2008), Performance of DEMETER calibration for rainfall forecasting purposes: Application to the July - August Sahelian rainfall, *J. Geophys. Res*., 113, D15111, doi: 10.1029/2007JD009403.

Caminade, C., and L. Terray (2010), Twentieth century Sahel rainfall variability as simulated by the ARPEGE AGCM, and future changes, *Clim. Dyn*., 35, 75-94, doi: 10.1007/s00382-009-0545-4.

914 Coelho, C. A. S., S. Pezzulli, M. Balmaseda, F. J. Doblas-Reyes, and D. B. Stephenson
915 (2004), Forecast Calibration and Combination: A Simple Bayesian Approach for ENSO,
916 *J. Clim.*, 17, 1504-1516, doi: 10.1175/1520-0442(2004)017<1504:FCACAS>2.0.CO;2.
917
918 Cook, K. H., E. K. Vizy (2006), Coupled Model Simulations of the West African
919 Monsoon System: Twentieth- and Twenty-First-Century Simulations, *J. Clim.*, 19, 3681-
920 3703, doi: 10.1175/JCLI3814.1.
921
922 DelSole, T., X. Yang, and M. K. Tippett (2012) Is unequal weighting significantly better
923 than equal weighting for multi-model forecasting? *Q. J. R. Meteorol. Soc*., 139, 176-183.
924 doi: 10.1002/qj.1961.
925
926 DeWitt, D. G. (2005), Retrospective Forecasts of Interannual Sea Surface Temperature
927 Anomalies from 1982 to Present Using a Directly Coupled Atmosphere-Ocean General
928 Circulation Model, *Mon. Wea. Rev.*, 133, 2972-2995, doi: 10.1175/MWR3016.1.
929
930 Doblas-Reyes, F. J., V. Pavan, and D. B. Stephenson (2003), The skill of multi-model
931 seasonal forecasts of the wintertime North Atlantic Oscillation, *Clim. Dyn*., 21, 501-514,
932 doi: 10.1007/s00382-003-0350-4.
933
934 Doblas-Reyes, F. J., R. Hagedorn, and T. N. Palmer (2005), The rationale behind the
935 success of multi-model ensembles in seasonal forecasting - II. Calibration and
936 combination, *Tellus A*, 57, 234-252, doi: 10.1111/j.1600-0870.2005.00104.x.

937

938 Doblas-Reyes, F. J., A. Weisheimer, M. Déqué, N. Keenlyside, M. McVean, J. M.
939 Murphy, P. Rogel, D. Smith, and T. N. Palmer (2009), Addressing model uncertainty in
940 seasonal and annual dynamical ensemble forecasts, *Q. J. R. Meteorol. Soc.*, 135, 1538-
941 1559. doi: 10.1002/qj.464.

942

943 Doblas-Reyes, F. J., J. García-Serrano, F. Lienert, A. P. Biescas, and L. R. L. Rodrigues
944 (2013), Seasonal climate predictability and forecasting: status and prospects, *WIREs*
945 *Clim. Change*, 4, 245-268, doi: 10.1002/wcc.217.

946

947 Folland, C. K., T. N. Palmer, and D. E. Parher (1986), Sahel rainfall and worldwide sea
948 surface temperature, *Nature*, 320, 602-607, doi: 10.1038/320602a0.

949

950 Fontaine, B., S. Janicot, and V. Moron (1995), Rainfall anomaly patterns and wind field
951 signals over West Africa in August (1958-1989), *J. Clim*., 8, 1503-1510, doi:
952 10.1175/1520-0442(1995)008<1503:RAPAWF>2.0.CO;2.

953

954 Fontaine, B., and S. Janicot (1996), Near-global sea surface temperature variability
955 associated with West African rainfall anomaly types, *J. Clim*., 9, 2935-2940.

956

957 Fontaine, B., S. Trzaska, and S. Janicot (1998), Evolution of the relationship between
958 near global and Atlantic SST modes and the rainy season in West Africa: statistical

959 analyses and sensitivity experiments, *Clim. Dyn.*, 14, 353-368, doi:
960 10.1007/s003820050228.
961
962 Giannini, A., R. Saravanan, and P. Chang (2003), Oceanic forcing of Sahel rainfall on
963 interannual to interdecadal timescales, *Science*, 302, 1027-1030, doi:
964 10.1126/science.1089357.
965
966 Giannini, A., R. Saravanan, and P. Chang (2005), Dynamics of the boreal summer
967 African monsoon in the NSIPP1 atmospheric model, *Clim. Dyn.*, 25, 517-535, doi:
968 10.1007/s00382-005-0056-x.
969
970 Gneiting, T., A. E. Raftery, A. H. Westveld, T. Goldman (2005), Calibrated Probabilistic
971 Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation,
972 *Mon. Wea. Rev.*, 133, 1098-1118, doi: 10.1175/MWR2904.1.
973
974 Goddard, L., and S. J. Mason (2002), Sensitivity of seasonal climate forecasts to persisted
975 SST anomalies, *Clim. Dyn*., 19, 619-632, doi: 10.1007/s00382-002-0251-y.
976
977 Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer (2005), The rationale behind the
978 success of multi-model ensembles in seasonal forecasting - I. Basic concept, *Tellus 57A*,
979 219-233, doi: 10.1111/j.1600-0870.2005.00103.x.
980

981 Hersbach, H. (2000), Decomposition of the Continuous Ranked Probability Score for
982 Ensemble Prediction Systems, *Wea. Forecasting*, 15, 559-570. doi: 10.1175/1520-
983 0434(2000)015<0559:DOTCRP>2.0.CO;2.
984
985 Hourdin, F., I. Musat, F. Guichard, P. M. Ruti, F. Favot, M.-A. Filiberti, M. Pham, J.-Y.
986 Grandpeix, J. Polcher, P. Marquet, A. Boone, J.-P Lafore, J-L Redelsperger, A
987 Dell'aquila, T. Losada, A. K. Traore, and H. Gallée (2010), AMMA-Model
988 Intercomparison Project, *Bull. Amer. Meteor. Soc.*, 91, 95-104, doi:
989 10.1175/2009BAMS2791.1.
990
991 Huffman, G. J., and D. T. Bolvin (2013), GPCP Version 2.2 Combined Precipitation Data
992 Set Documentation, Internet Publication, 1-46, (Available at:
993 http://www1.ncdc.noaa.gov/pub/data/gpcp/gpcp-v2.2/doc/V2.2_doc.pdf), Accessed: 16
994 November 2012.
995
996 Ingram, K. T., M. C. Roncoli, and P. H. Kirshen (2002), Opportunities and constraints for
997 farmers of west Africa to use seasonal precipitation forecasts with Burkina Faso as a case
998 study, *Agricultural Systems*, 74, 331-349, doi: 10.1016/S0308-521X(02)00044-6.
999
1000 Janicot, S., A. Harzallah, B. Fontaine, and V. Moron (1998), West African monsoon
1001 dynamics and eastern equatorial Atlantic and Pacific SST anomalies, *J. Clim.*, 11, 1874-
1002 1882.
1003

1004 Janicot, S., S. Trzaska, and I. Poccard (2001), Summer Sahel-ENSO teleconnection and
1005 decadal time scale SST variations, *Clim. Dyn.*, 18, 303-320, doi:
1006 10.1007/s003820100172.
1007
1008 Jolliffe, I. T., and Stephenson, D. B. (2012), *Forecast Verification: A Practitioner's*
1009 *Guide in Atmospheric Science*, Second Edition, John Wiley & Sons, Ltd, Chichester, doi:
1010 10.1002/9781119960003.ch1.
1011
1012 Joly, M., and A. Voldoire (2009), Influence of ENSO on the West African Monsoon:
1013 Temporal Aspects and Atmospheric Processes, *J. Clim.*, 22, 3193-3210, doi:
1014 10.1175/2008JCLI2450.1.
1015
1016 Joly, M., and A. Voldoire (2010), Role of the Gulf of Guinea in the inter-annual
1017 variability of the West African monsoon: what do we learn from CMIP3 coupled
1018 simulations?, *Int. J. Climatol*., 30, 1843-1856, doi: 10.1002/joc.2026.
1019
1020 Kim, H. M., P. J. Webster, and J. A. Curry (2012), Seasonal prediction skill of ECMWF
1021 System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter,
1022 *Clim. Dyn*., 39, 2957-2973, doi: 10.1007/s00382-012-1364-1366.
1023
1024 Kirtman, B. P., D. Min (2009), Multimodel Ensemble ENSO Prediction with CCSM and
1025 CFS, *Mon. Wea. Rev*., 137, 2908-2930, doi: 10.1175/2009MWR2672.1.
1026

1027 Kirtman, B. P., D. Min, J. M. Infanti, J. L. Kinter III, D. A. Paolino, Q. Zhang, H. van
1028 den Dool, S. Saha, M. P. Mendez, E. Becker, P. Peng, P. Tripp, J. Huang, D. G. DeWitt,
1029 M. K. Tippett, A. G. Barnston, S. Li, A. Rosati, S. D. Schubert, M. Rienecker, M. Suarez,
1030 Z. E. Li, J. Marshak, Y.-K. Lim, J. Tribbia, K. Pegion, W. J. Merryfield, B. Denis, E. F.
1031 Wood (2013), The North American Multi-Model Ensemble (NMME): Phase-1 Seasonal
1032 to Interannual Prediction, Phase-2 Toward Developing Intra-Seasonal Prediction, doi:
1033 10.1175/BAMS-D-12-00050.1.
1034
1035 Krishna Kumar, K., M. Hoerling, and B. Rajagopalan (2005), Advancing dynamical
1036 prediction of Indian monsoon rainfall, *Geophys. Res. Lett*., 32, L08704, doi:
1037 10.1029/2004GL021979.
1038
1039 Losada, T., B. Rodríguez-Fonseca, S. Janicot, S. Gervois, F. Chauvin, and P. Ruti (2010),
1040 A multi-model approach to the Atlantic Equatorial mode: impact on the West African
1041 monsoon, *Clim. Dyn*., 35, 29-43, doi: 10.1007/s00382-009-0625-5.
1042
1043 Mason, S. J. (2008), Understanding forecast verification statistics, *Meteorol. Appl*., 15,
1044 31-40, doi: 10.1002/met.51.
1045
1046 Mason, S. J., O. Baddour (2008), Statistical modelling, in Seasonal climate: forecasting
1047 and managing risk, edited by A. Troccoli, M. S. J. Harrison, D. L. T. Anderson, S. J.
1048 Mason, pp. 167-206, Springer, Dordrecht.
1049

Merryfield, W. J., W.-S. Lee, G. J. Boer, V. V. Kharin, J. F. Scinocca, G. M. Flato, R. S. Ajayamohan, J. C. Fyfe, Y. Tang, and S. Polavarapu (2013), The Canadian Seasonal to Interannual Prediction System. Part I: Models and Initialization, *Mon. Wea. Rev.*, 141, 2910-2945, doi: 10.1175/MWR-D-12-00216.1.

Mohino, E., S. Janicot, J. Bader (2011a), Sahel rainfall and decadal to multi-decadal sea surface temperature variability, *Clim. Dyn*., 37, 419-440, doi: 10.1007/s00382-010-0867-2.

Mohino, E., B. Rodriguez-Fonseca, T. Losada, S. Gervois, S. Janicot, J. Bader, P. Ruti, and F. Chauvin (2011b), Changes in the interannual SST-forced signals on West African rainfall: AGCM intercomparison, *Clim. Dyn*., 37, 1707-1725, doi: 10.1007/s00382-011-1093-2.

Molteni, F., T. Stockdale, M. Balmaseda, G. Balsamo, R. Buizza, L. Ferranti, L. Magnusson, K. Mogensen, T. Palmer, and F. Vitart (2011), The new ECMWF seasonal forecast system (System 4), ECMWF Technical Memorandum 656, pp. 51, (Available at http://www.ecmwf.int/publications/library/do/references/list/14), accessed on 20 December 2012.

Motha, R. P., S. K. Leduc, L. T. Steyaert, C. M. Sakamoto, and N. D. Strommen (1980), Precipitation Patterns in West Africa, *Mon. Wea. Rev*., 108, 1567-1578, doi: 10.1175/1520-0493(1980)108<1567:PPIWA>2.0.CO;2.

1073

1074 Murphy, A. H. (1991), Forecast verification: Its Complexity and Dimensionality, *Mon.*
1075 *Wea. Rev.*, 119, 1590-1601, doi: 10.1175/1520-
1076 0493(1991)119<1590:FVICAD>2.0.CO;2.

1077

1078 Nicholls, N. (2001), Commentary and analysis: The Insignificance of Significance
1079 Testing, *Bull. Amer. Meteor. Soc.*, 82, 981-986, doi: 10.1175/1520-
1080 0477(2001)082<0981:CAATIO>2.3.CO;2.

1081

1082 Nicholson, S. E. (1993), An overview of African rainfall fluctuations of the last decade,
1083 *J. Clim.*, 6, 1463-1466, doi: 10.1175/1520-0442(1993)006<1463:AOOARF>2.0.CO;2.

1084

1085 Philippon, N., F. J. Doblas-Reyes, and P. M. Ruti (2010), Skill, reproducibility and
1086 potential predictability of the West African monsoon in coupled GCMs, *Clim. Dyn.*, 35,
1087 53-74, doi: 10.1007/s00382-010-0856-5.

1088

1089 Rodrigues, L. R. L., Doblas-Reyes, F. J., Coelho, C. A. S. (2014), Multi-model
1090 calibration and combination of tropical seasonal sea surface temperature forecasts, *Clim.*
1091 *Dyn.*, 42, 597-616, doi: 10.1007/s00382-013-1779-8.

1092

1093 Rodríguez-Fonseca, B., S. Janicot, E. Mohino, T. Losada, J. Bader, C. Caminade, F.
1094 Chauvin, B. Fontaine, J. García-Serrano, S. Gervois, M. Joly, I. Polo, P. Ruti, P. Roucou,

and A. Voldoire (2011), Interannual and decadal SST-forced responses of the West African Monsoon, *Atmos. Sci. Lett*., 12, 67-74, doi: 10.1002/asl.308.

Roehrig, R., D. Bouniol, F. Guichard, F. Hourdin, J.-L. Redelsperger (2013), The Present and Future of the West African Monsoon: A Process-Oriented Assessment of CMIP5 Simulations along the AMMA Transect, *J. Climate*, 26, 6471-6505, doi: 10.1175/JCLI-D-12-00505.1.

Saha, S., S. Moorthi, X. Wu, J. Wang, S. Nadiga, P. Tripp, D. Behringer, Y.-T. Hou, H. Chuang, M. Iredell, M. Ek, J. Meng, R. Yang, M. P. Mendez, H. van den Dool, Q. Zhang, W. Wang, M. Chen, and E. Becker (2013), The NCEP Climate Forecast System version 2, *J. Clim.*, doi: 10.1175/JCLI-D-12-00823.1.

Schneider, U., A. Becker, A. Meyer-Christoffer, M. Ziese, and B. Rudolf (2011), Global Precipitation Analysis Products of the GPCC. Global Precipitation Climatology Centre (GPCC), DWD, Internet Publikation, 1-13. (Available at: http://www.dwd.de/bvbw/generator/DWDWWW/Content/Oeffentlichkeit/KU/KU4/KU42/en/Reports__Publications/GPCC__intro__products__2011,templateId=raw,property=publicationFile.pdf/GPCC_intro_products_2011.pdf), Accessed: 16 November 2012.

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, 2283-2296, doi: 10.1175/2007JCLI2100.1.

Stephenson, D. B., C. A. S. Coelho, F. J. Doblas-Reyes, and M. Balmaseda (2005), Forecast assimilation: a unified framework for the combination of multi-model weather and climate predictions, *Tellus A*, 57: 253–264, doi: 10.1111/j.1600-0870.2005.00110.x.

Stephenson, D. B., C. A. S. Coelho, I. T. Jolliffe (2008), Two Extra Components in the Brier Score Decomposition, *Wea. Forecasting*, 23, 752-757, doi: 10.1175/2007WAF2006116.1.

Sultan, B., and S. Janicot (2000), Abrupt shift of the ITCZ over West Africa and intra-seasonal variability, *Geophys. Res. Lett.*, 27, 3353-3356, doi: 10.1029/1999GL011285.

Sultan, B., S. Janicot, and A. Diedhiou (2003), The West African monsoon dynamics. Part I: documentation of intraseasonal variability, *J. Clim.*, 16, 3389-3406, doi: 10.1175/1520-0442(2003)016<3389:TWAMDP>2.0.CO;2.

Sultan, B., and S. Janicot (2003), The West African monsoon dynamics. Part II: the 'preonset' and 'onset' of the summer monsoon, *J. Clim.*, 16, 3407-3427, doi: 10.1175/1520-0442(2003)016<3407:TWAMDP>2.0.CO;2.

Sylla, M. B., I. Diallo, and J. S. Pal (2013), West African Monsoon in State-of-the-Science Regional Climate Models, in Climate Variability - Regional and Thematic Patterns, edited by Aondover Tarhule, doi: 10.5772/55140, (available from:

http://www.intechopen.com/books/climate-variability-regional-and-thematic-patterns/west-african-monsoon-in-state-of-the-science-regional-climate-models).

Tippet, M. K., and A. Giannini (2006), Potentially predictable components of African summer rainfall in an SST-forced GCM simulation, *J. Clim*., 19, 3133-3144, doi: 10.1175/JCLI3779.1.

Tompkins, A. M., and L. Feudale (2010), Seasonal Ensemble Predictions of West African Monsoon Precipitation in the ECMWF System 3 with a Focus on the AMMA Special Observing Period in 2006, *Wea. Forecasting*, 25, 768-788, doi: 10.1175/2009WAF2222236.1.

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

Trenberth, K. E. and D. J. Shea (2006), Atlantic hurricanes and natural variability in 2005, *Geophys. Res. Lett*., 33, doi: 10.1029/ 2006GL026894.

Vellinga, M., A. Arribas, and R. Graham (2013), Seasonal forecasts for regional onset of the West African monsoon, *Clim. Dyn.*, 40, 3047-3070, doi: 10.1007/s00382-012-1520-z.

Wilks, D. (2006), *Statistical methods in the atmospheric sciences*, Second edition, International Geophysics Series, vol. 59, Elsevier, Oxford.

Yoshikatsu, Y., M. Koki, T. Hiroshi (2008), Global Warming Projections Using the Community Climate System Model, CCSM3, *Nec Technical J.*, 3, 73-76.

Yuan, X., E. F. Wood, L. Luo, and M. Pan (2011), A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction, *Geophys Res Lett*, 38, doi: 10.1029/ 2011GL047792.

Zhang, S., M. J. Harrison, A. Rosati, and A. Wittenberg (2007), System Design and Evaluation of Coupled Ensemble Data Assimilation for Global Oceanic Climate Studies, *Mon. Wea. Rev*., 135, 3541-3564, doi: 10.1175/MWR3466.1.

Zebiak, S. E. (1993), Air-sea interaction in the equatorial Atlantic region, *J. Clim*., 6, 1567-1586, doi: 10.1175/1520-0442(1993)006<1567:AIITEA>2.0.CO;2.

Zuo, Z., S. Yang, Z.-Z. Hu, R. Zhang, W. Wang, B. Huang, and F. Wang (2013), Predictable patterns and predictive skills of monsoon precipitation in Northern Hemisphere summer in NCEP CFSv2 reforecasts, *Clim. Dyn*., 40, 3071-3088, doi: 10.1007/s00382-013-1772-2.

Tables

Table 1: Variance explained (%) by the first and second modes of the WAM rainfall variability by the GPCP, GPCC, and the dynamical forecast systems. For the predicted modes of variability, the variance is displayed for each lead time.

	Variance (%): First mode			Variance (%): Second mode		
	Lead 0	Lead 1	Lead 2	Lead 0	Lead 1	Lead 2
GPCP		29			23	
GPCP land-only		31			24	
GPCC		27			20	
GPCC (1951-2011)		30			18	
S4	25	34	41	15	14	11
CFSv2	15	19	18	09	09	08
MF3	27	20	16	11	11	11
CCSM3	46	49	51	10	09	09
GFDL	24	22	30	19	18	18
IRI-ECHAM ano	34	31	29	14	15	15
IRI-ECHAM direct	32	33	31	11	12	11
CMC2	18	18	15	12	12	13

Figures

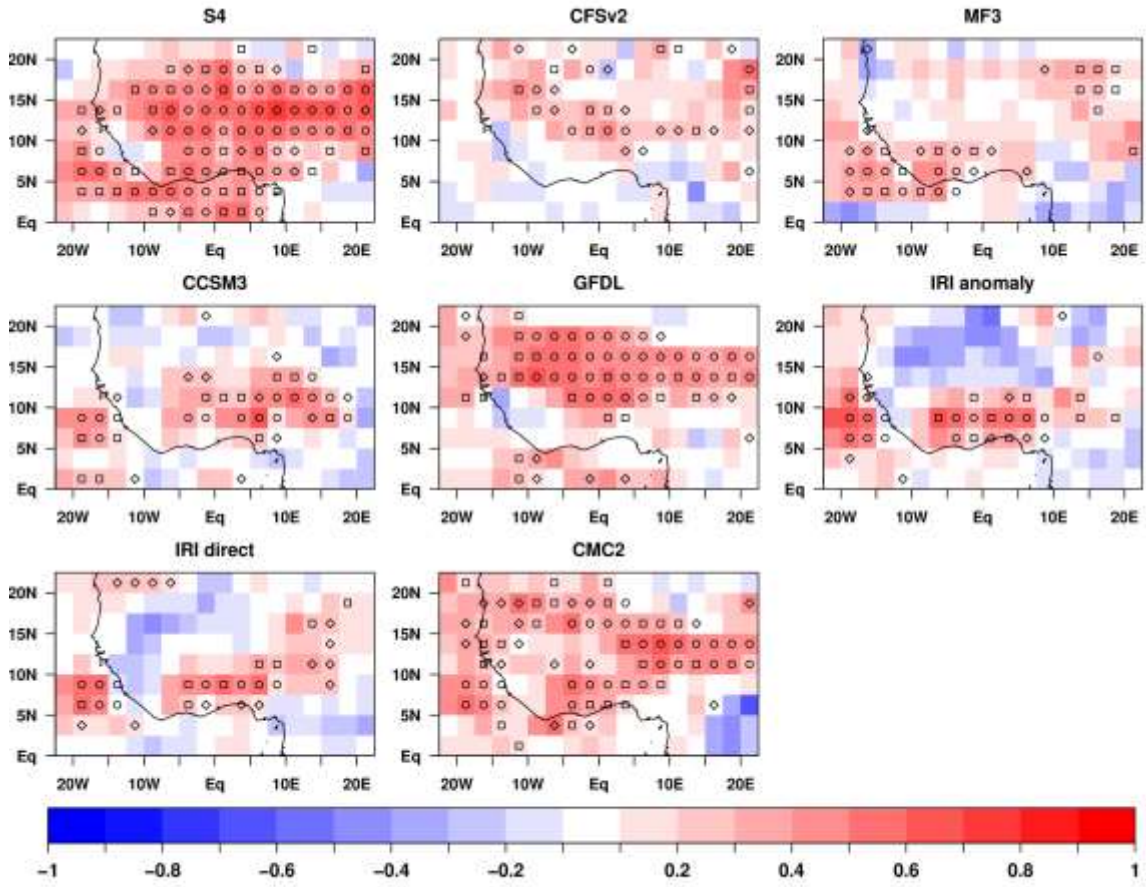


Figure 1. Correlation coefficient between the predicted ensemble mean and observed summer (JAS) rainfall at each grid-point over the WAM region for the period 1982-2011. The GPCP dataset was used as the reference data. Forecasts are for lead time 1 month and interpolated into the GPCP grid prior to computing the correlation coefficient. Circles are for p-values smaller than or equal 0.01, squares for p values between 0.05 and 0.01, and diamonds for p values between 0.10 and 0.05.

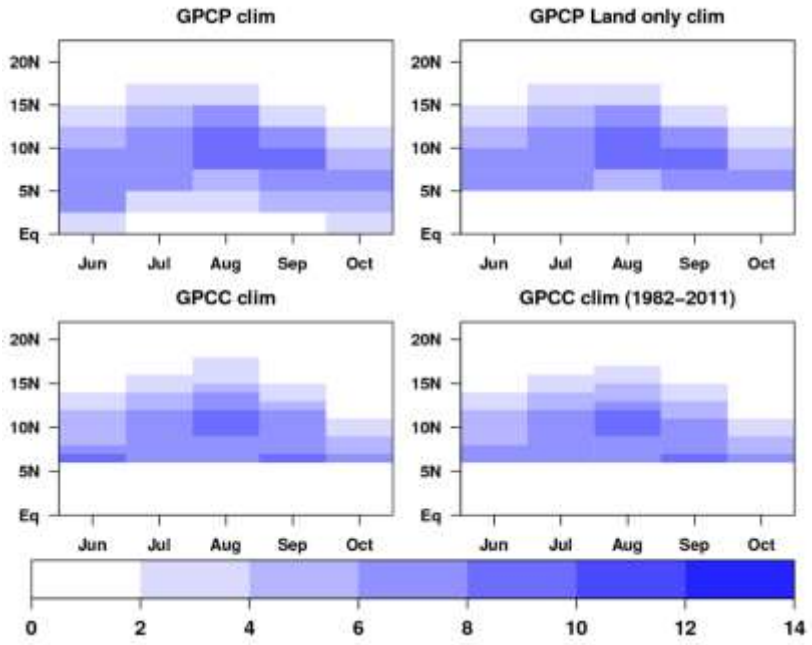


Figure 2. Monthly rainfall (mm/day) averaged over 10ºW-10ºE as a function of month, from June to October, and of latitude. Climatologies of the two analysed observational datasets, GPCP and GPCC were computed using the period 1982-2011 and 1951-2011, respectively, except when indicate otherwise. For comparison, the GPCP climatology was also computed masking the ocean and the GPCC using only the common period 1982-2011.

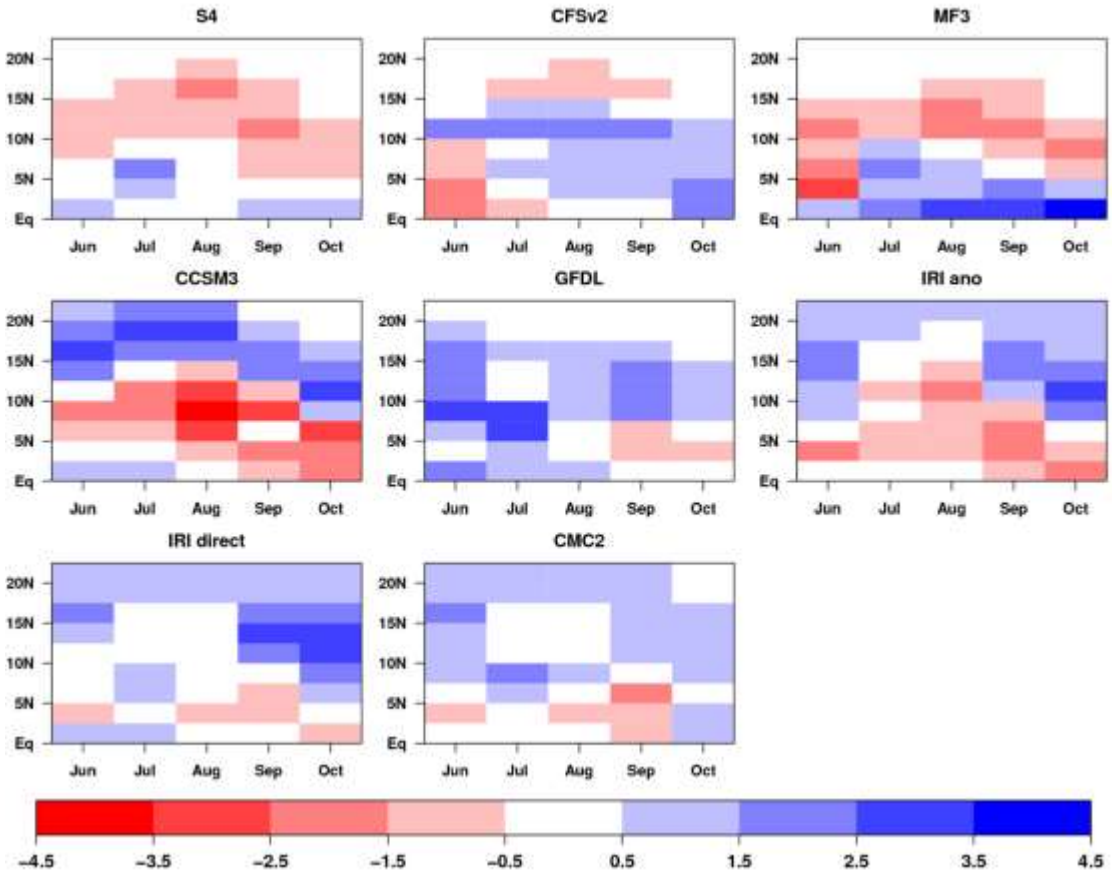


Figure 3: Mean precipitation bias (mm/day) of the dynamical forecast systems over the WAM region for the period 1982-2011 is computed as the difference between the one-month-lead hindcasts and the GPCP mean climatological estimates. The hindcasts were interpolated into the GPCP grid prior to computing the systematic error.

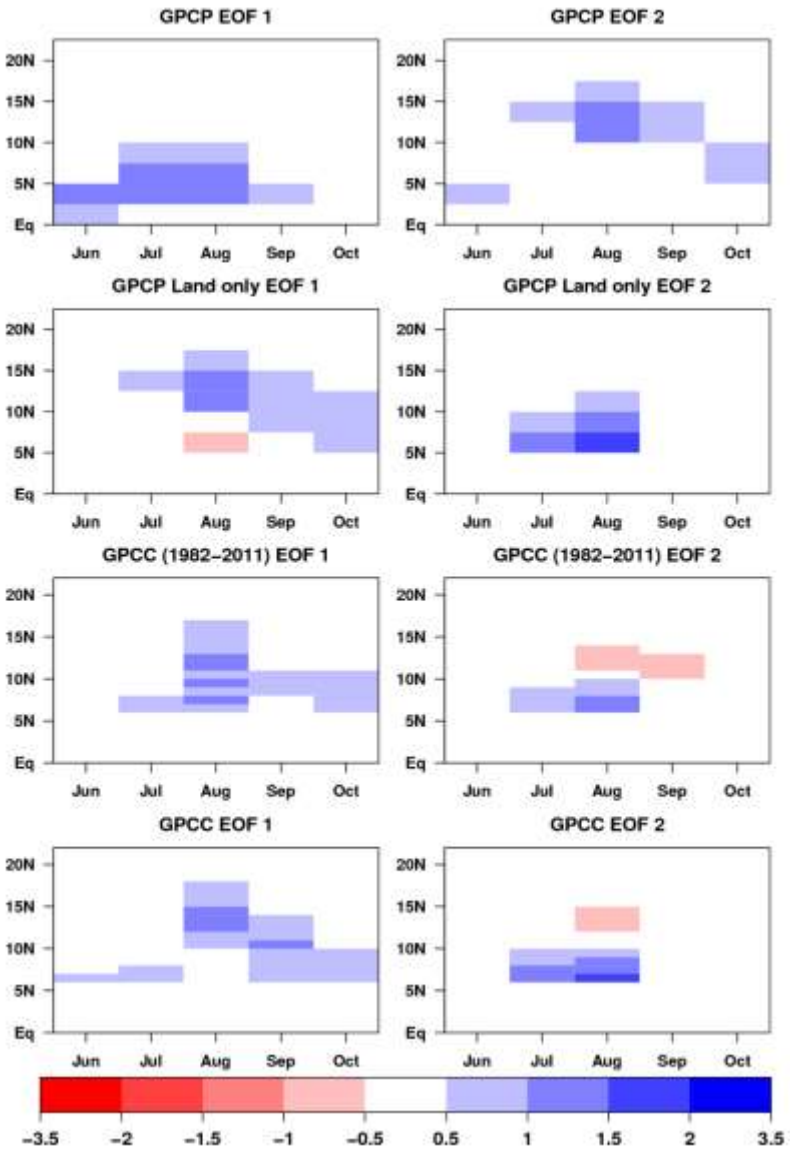


Figure 4. Leading two EOFs of the longitudinally-averaged precipitation datasets of Figure 2.

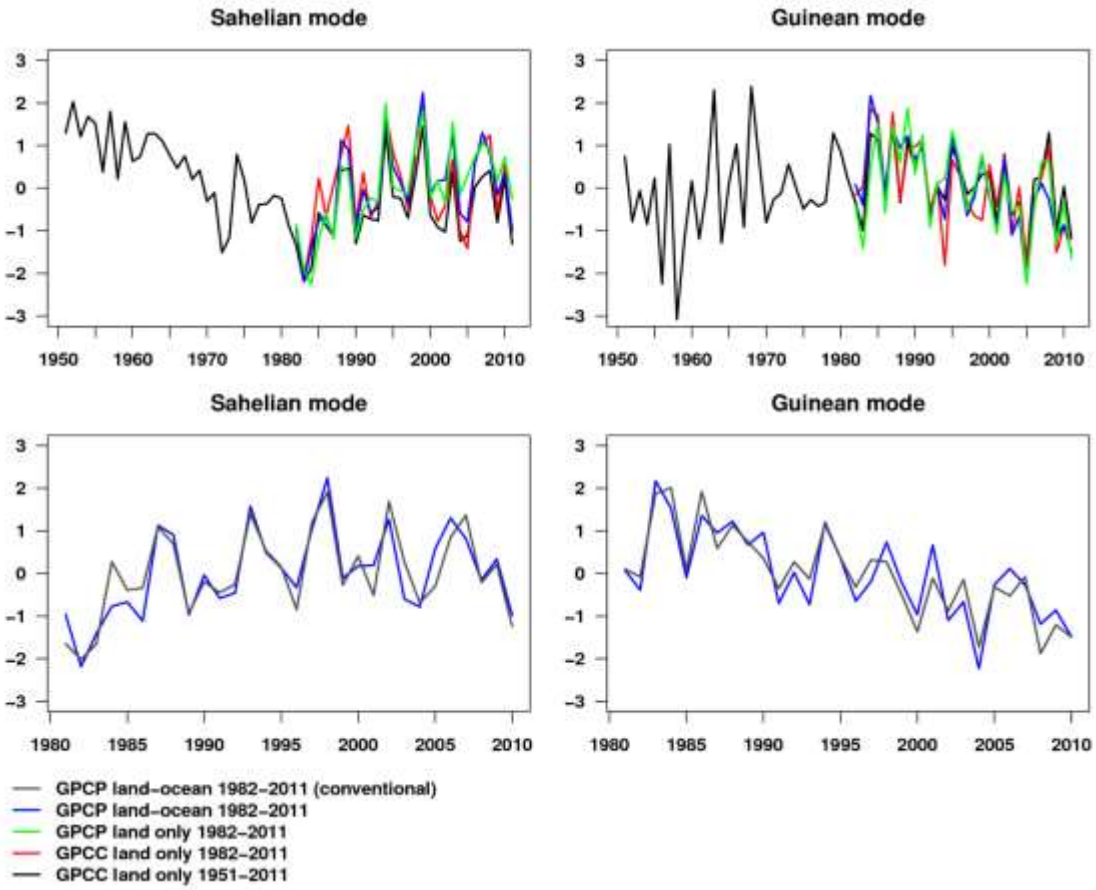
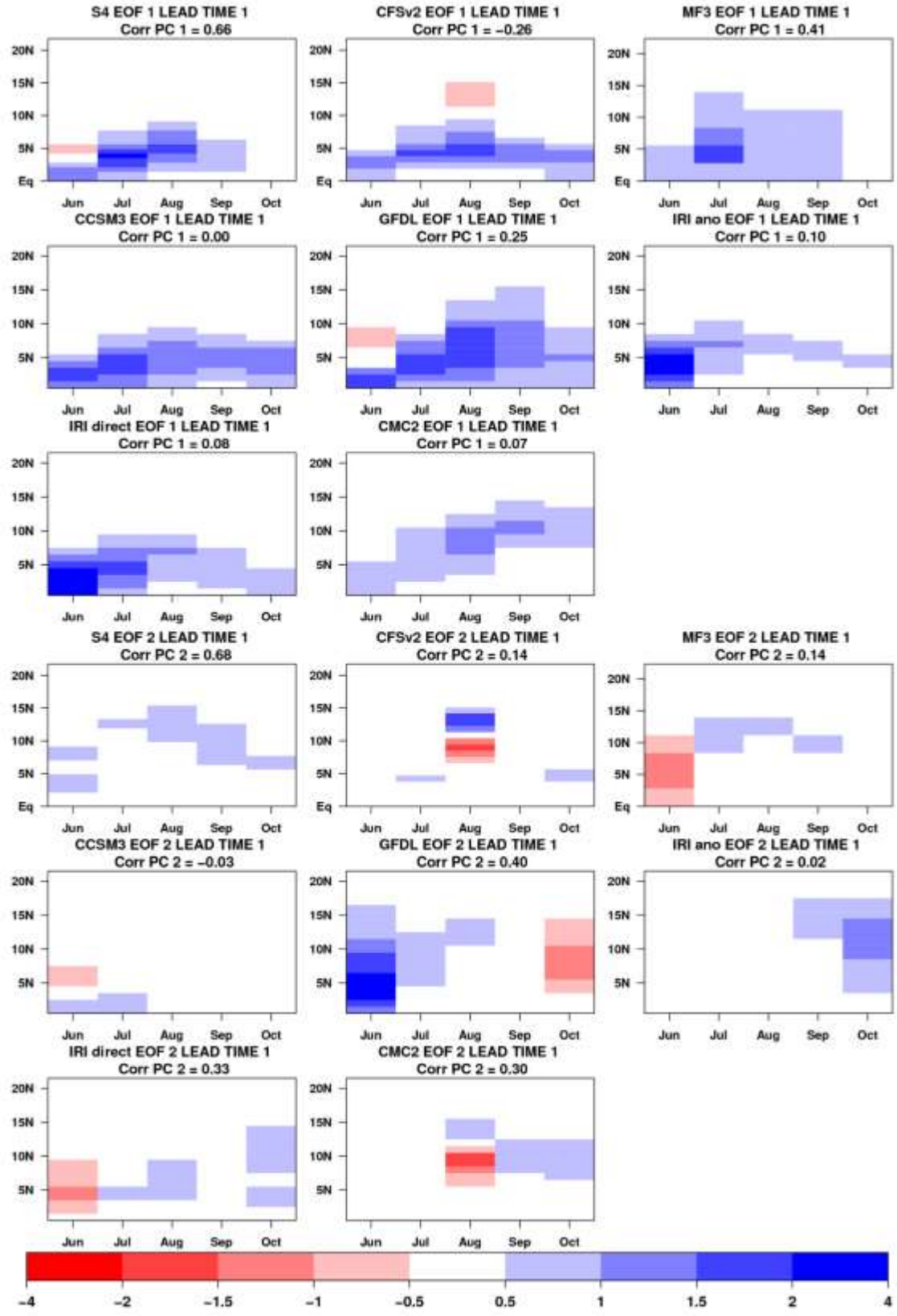


Figure 5. Principal components associated with the EOFs shown in Figure 4. The blue line is the PC of the GPCP land and ocean, the green line is the PC of the GPCP land only and the red line is the PC of the GPCC land only. These three PC are computed for the common period 1982-2011. The black line is the PC computed using the GPCC land only for the period 1951-2011. These PCs are estimated using the seasonal evolution diagrams averaged over 10ºW-10ºE, covering the latitudes between the Equator and 20ºN and the period between June and October. For comparison, the PCs are also estimated using the traditional way with the full spatial field (i.e., without applying the longitudinal averaging) over 10ºW-10ºE and between the Equator and 20ºN on the JAS rainfall (gray line, bottom panels). The blue lines are the same in the top and bottom panels. The correlation between GPCP land and ocean PC1 (blue line in the upper right panel) and the GPCC land only PC2 (black line in the upper right panel) is 0.84 while the correlation between the GPCP land and ocean PC2 (blue line in the upper left panel) and the GPCC land only PC1 (black line in the upper left panel) is 0.95. The correlation between the WAM rainfall regimes estimated using the seasonal

evolution diagrams and the spatial field is 0.91 for the Guinean regime (lower right panel) and 0.90 for the Sahelian regime (lower left panel).


S4 EOF 1 LEAD TIME 1
Corr PC 1 = 0.66
CFSv2 EOF 1 LEAD TIME 1
Corr PC 1 = −0.26
MF3 EOF 1 LEAD TIME 1
Corr PC 1 = 0.41
CCSM3 EOF 1 LEAD TIME 1
Corr PC 1 = 0.00
GFDL EOF 1 LEAD TIME 1
Corr PC 1 = 0.25
IRI ano EOF 1 LEAD TIME 1
Corr PC 1 = 0.10
IRI direct EOF 1 LEAD TIME 1
Corr PC 1 = 0.08
CMC2 EOF 1 LEAD TIME 1
Corr PC 1 = 0.07
S4 EOF 2 LEAD TIME 1
Corr PC 2 = 0.68
CFSv2 EOF 2 LEAD TIME 1
Corr PC 2 = 0.14
MF3 EOF 2 LEAD TIME 1
Corr PC 2 = 0.14
CCSM3 EOF 2 LEAD TIME 1
Corr PC 2 = −0.03
GFDL EOF 2 LEAD TIME 1
Corr PC 2 = 0.40
IRI ano EOF 2 LEAD TIME 1
Corr PC 2 = 0.02
IRI direct EOF 2 LEAD TIME 1
Corr PC 2 = 0.33
CMC2 EOF 2 LEAD TIME 1
Corr PC 2 = 0.30
20N
15N
10N
5N
Eq
Jun
Jul
Aug
Sep
Oct
−4
−2
−1.5
−1
−0.5
0.5
1
1.5
2
4

1229

Figure 6. As Figure 4 but for the lead time 1 month (start date in May) dynamical hindcasts. EOF1 is displayed in the upper set of panels and EOF2 in the lower set of panels. The correlation between the predicted and observed PCs is included in the second line of the panel title.

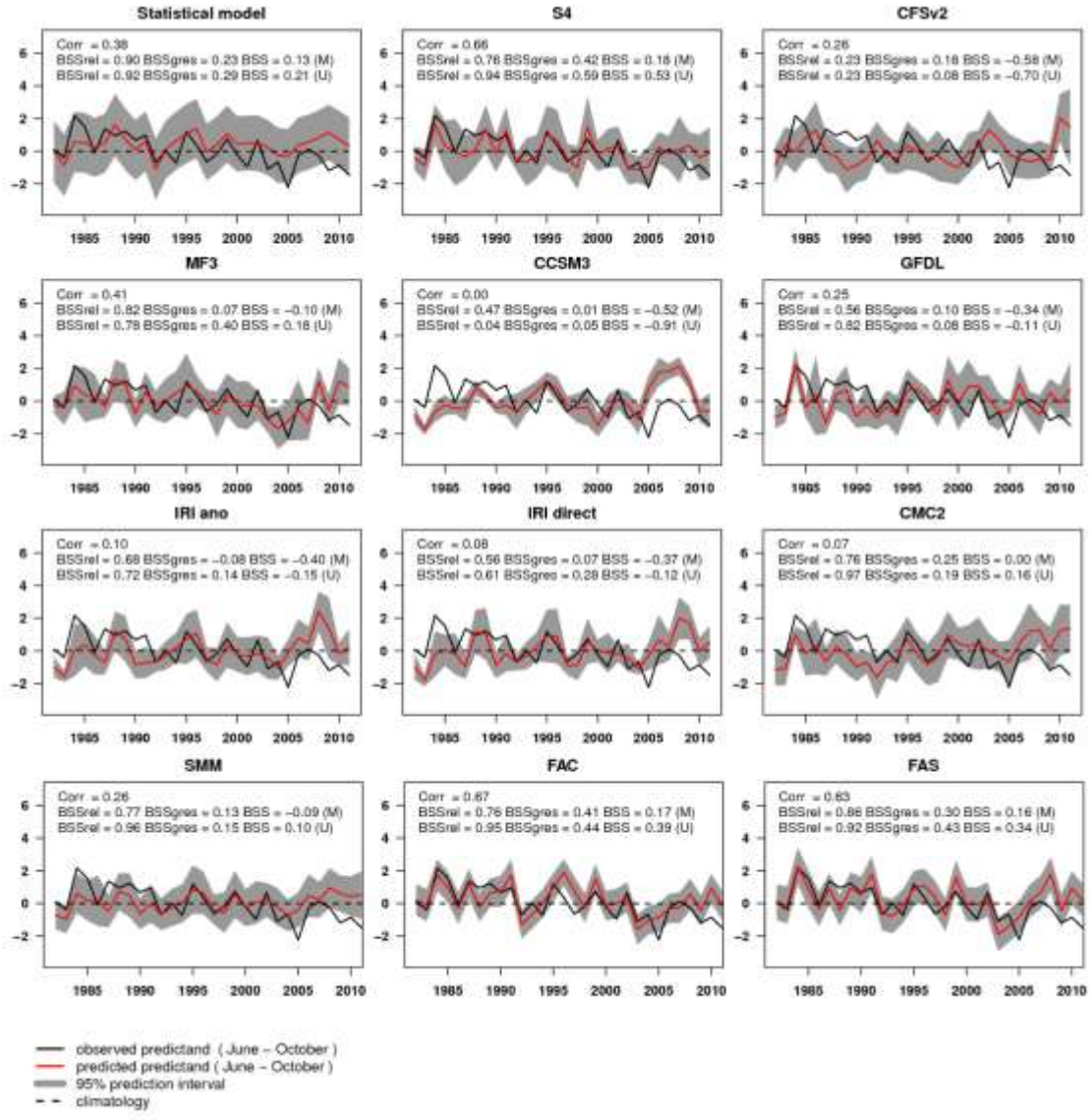


Figure 7. Leading principal component (Guinean regime) predicted by the statistical model, the dynamical forecast systems and their combinations. Predictions are for lead time 1 (start date in May). Observed values (black solid line), predicted values (red solid line), 95% predicted interval (grey area) and the zero line (black dashed line) are displayed. The values displayed are anomalies. The correlation coefficient, the

BSS, BSSrel and BSSgres for probabilities of rainfall regime being above the median (M) and the upper quartile (U) are displayed in each panel.

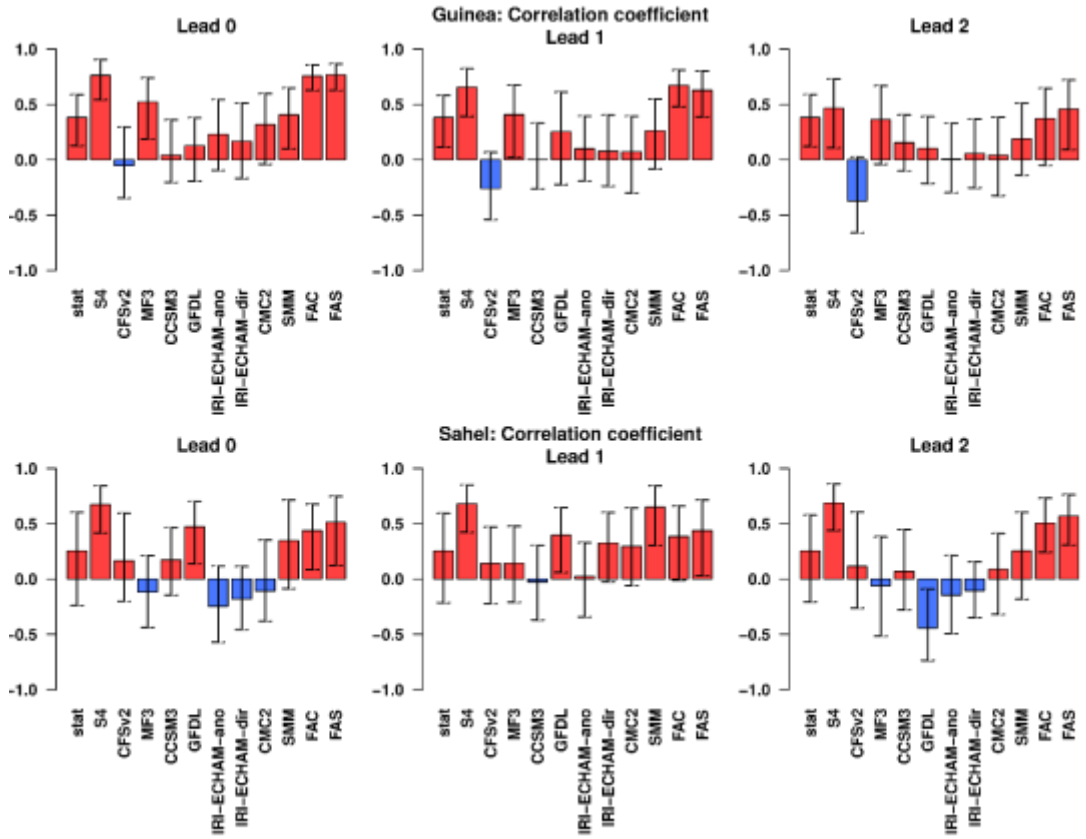


Figure 8. Correlation coefficient between the observed and predicted ensemble mean PCs for the period 1982-2011. The correlation was computed for the Sahelian (lower panel) and Guinean (upper panel) rainfall regimes and for lead times zero, one and two. The bars in each histogram represent the forecast systems. The lower and upper bound of the bootstrapped confidence interval is displayed as vertical bars.

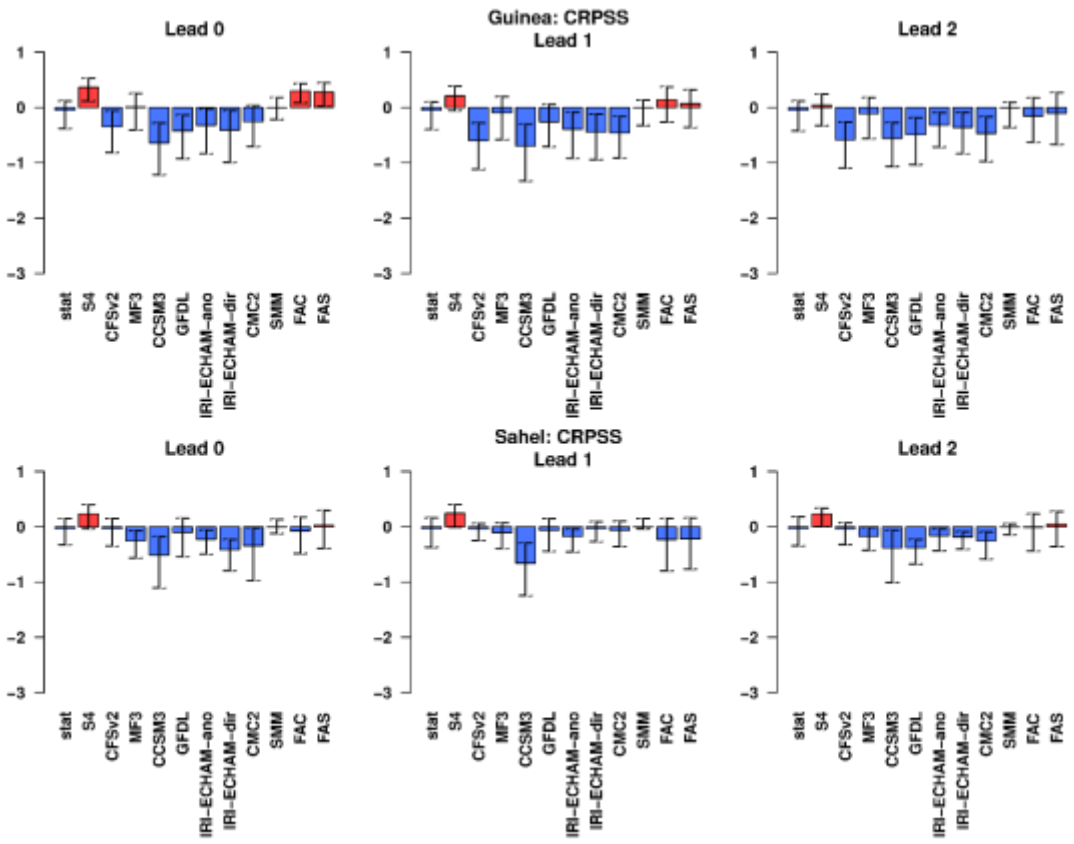


Figure 9. Same as Figure 8, but for the CRPSS.

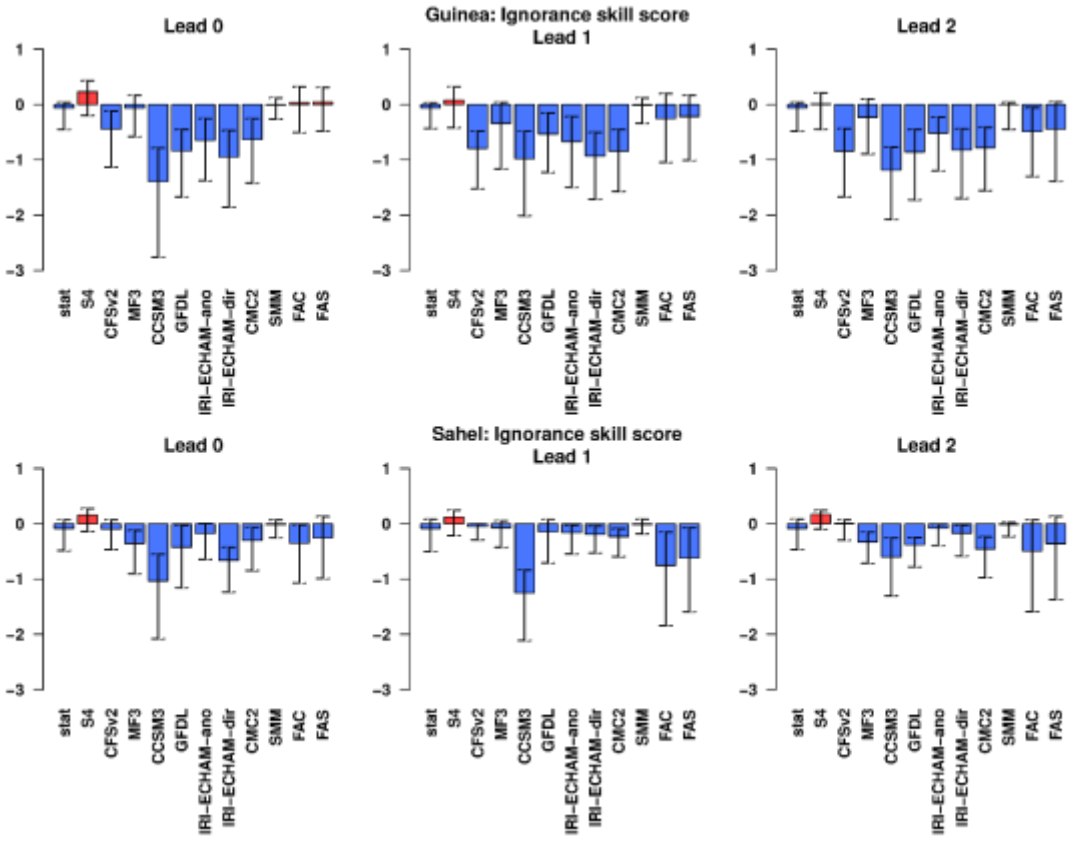


Figure 10. Same as Figure 8, but for the ignorance skill score.

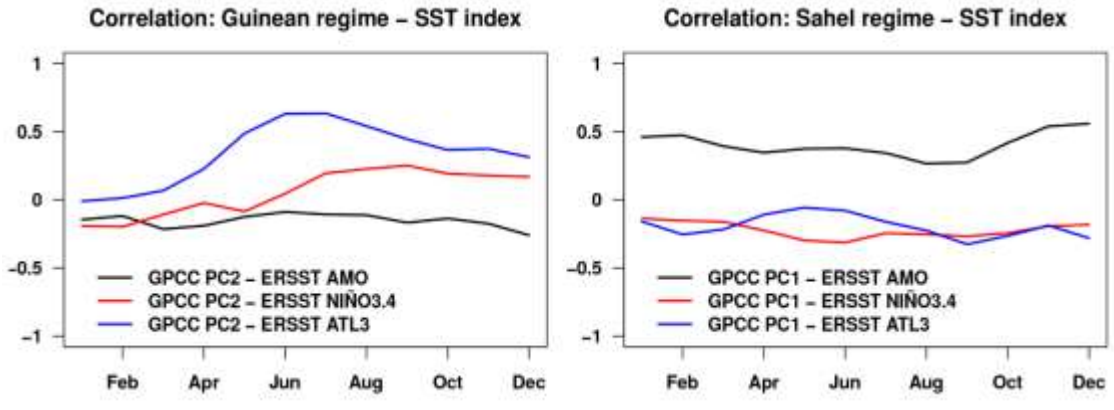


Figure A1: Correlation coefficient between the Guinean and Sahelian regimes (estimated from the GPCC seasonal evolution diagram described above) and three ERSSTv3b SST indices: AMO, Niño3.4 and Atl3. The correlation is computed for each month of the year and for the period 1951-2011.