Advances in the Lead Time of Sahel Rainfall Prediction With the North American Multimodel Ensemble

Alessandra Giannini, A. Ali, C P Kelley, B L Lamptey, B Minoungou, O Ndiaye

Advances in the lead time of Sahel rainfall prediction with the North American Multi-Model Ensemble

A. Giannini[1,2], A. Ali[3], C. P. Kelley[1], B. L. Lamptey[4,5], B. Minoungou[3], O. Ndiaye[1,6]

[1]International Research Institute for Climate and Society, The Earth Institute at Columbia University, Palisades NY, U.S.A.
[2]Now at Laboratoire de Météorologie Dynamique/IPSL, Ecole Normale Supérieure, PSL Research University, Sorbonne Université, École Polytechnique, IP Paris, CNRS, Paris, France
[3]Centre Régional AGRHYMET, Niamey, Niger
[4]Cheney Fellow, University of Leeds, School of Earth and Environment, UK
[5]Formerly at the African Centre of Meteorological Applications for Development, Niamey, Niger
[6]Agence Nationale de l'Aviation Civile et de la Météorologie, Dakar, Senegal

Key Points:

- The North American Multi-Model Ensemble predicts July-September Sahel-wide precipitation as skillfully in February/March as in June.
- Skill comes from the ability to predict tropical Pacific and North Atlantic surface temperatures, attributable to 2 models in particular.
- Skill in predicting the spatial average is significantly higher than the spatial average of local/gridpoint skill.

Corresponding author: A. Giannini, alessandra.giannini@lmd.ens.fr

Abstract

We assess the deterministic skill in seasonal climate predictions of Sahel rainfall made with the North American Multi-Model Ensemble (NMME). We find that skill for a regionally averaged rainfall index is essentially the same for forecasts for the July-September target season made as early as February/March and as late as June. The two dominant influences on the climate of the Sahel, the North Atlantic and the global tropical oceans, shape this predictability. Multi-model ensemble skill hinges on the combination of skillful predictions of the El Niño-Southern Oscillation made with one model (CMC2-CanCM4) with those of North Atlantic sea surface temperatures made with another (NASA-GEOSS2S).

Plain Language Summary

The seasonal climate outlook forum for the Sudano-Sahelian region of West Africa convenes in mid/late-April at the earliest, because the statistical models currently in use to make predictions for the July-September rainy season have little skill before then. Here we show that the North American Multi-Model Ensemble (NMME), a seasonal climate prediction system based on dynamical models, predicts Sahel-wide July-September rainfall anomalies in February/March with essentially the same skill as in June. An earlier, by 2-3 months, outlook is consequential to decisions that can exploit it for better preparedness, such as purchasing, stocking and distributing adapted seed varieties, or triggering humanitarian intervention to prevent regional food insecurity.

The NMME prediction system owes its skill to the correct characterization of oceanic influence on Sahel rainfall, which is achieved by combining output from two models particularly skillful at predicting North Atlantic and tropical Pacific sea surface temperature anomalies respectively. Recognition that the oceanic source of predictability is the same for the entire region means that whether the forecast for the regional average holds in a given year, at a specific location, largely depends on the strength of oceanic influence in that year, rather than on any local condition or consideration.

Keywords: Sahel, precipitation, seasonal climate prediction, PRESA-SS, El Niño-Southern Oscillation, climate services.

1 Introduction

Rainfall in the Sahel, the semi-arid southern edge of the Sahara, is characterized by high spatio-temporal variability. Variability in time is evident in Figure 1, an update of Ali and Lebel's (2009) analysis, in the multi-decadal swings between the anomalously wet 1950s and 1960s and the anomalously dry 1970s and 1980s. Interannual variability is particularly marked in the current epoch, which has been labelled a *partial recovery* (Nicholson 2005; AGRHYMET 2010). Variability in space is an intrinsic property of convective precipitation (Le Barbé and Lebel 1997; Rio *et al.* 2019). Despite this apparent complexity, seasonal and sub-continental anomalies are coherent. The strength of observed regionally averaged precipitation anomalies is proportional to the area characterized by anomalies of consistent sign, that is to say, that the larger the anomaly in the regional average, the more extensive the area of anomaly of the same sign (Ali and Lebel 2009). A leading Empirical Orthogonal Function of sub-Saharan African rainfall variability defines the Sahel as the poleward edge of the northern hemisphere summer monsoon (Giannini *et al.* 2005). In models, this pattern is present in atmospheric simulations run over climatological sea surface temperature (SST), and amplified in the presence of observed SST variability.

For societies where a large fraction of the population finds employment in the agricultural sector, skillful seasonal prediction is a valuable tool to manage crop risk related to climate variation (Tall *et al.* 2018; Ouedraogo *et al.* 2018). Indeed, the West African climate outlook forum has met annually since 1998 to produce consensus forecasts (Ogallo *et al.* 2000; Traoré *et al.* 2014). Initially a single forum, referred to as PRESAO from the French acronym for *Prévisions Saisonnières en Afrique de l'Ouest* (Seasonal Predictions in West Africa), the process has recently split into two, PRESA-GG and PRESA-SS, involving the Gulf of Guinea and Sudano-Sahelian countries, respectively, in recognition of differences in seasonality. The original prediction methodology was statistical, and exploited multi-linear regression to relate the predictand, that is, precipitation at broad subnational scales, with predictors chosen among a small set of SST indices (e.g., Folland *et al.* 1991; Ward *et al.* 1993; Baddour 1998). In current practice, during the *pre-forum* experts from the National Meteorological Services present their predictions, still largely statistical in nature. Discussions combining these quantitative assessments with the qualitative assessment of predictions made by research and operational centers worldwide are distilled into a consensus forecast, which is communicated to stakeholders in the *forum*.

Globally, seasonal prediction has evolved from a 2-tier to a 1-tier approach. In the 2-tier approach of the mid-1990s, SSTs were predicted first, usually with a combination of statistical and dynamical models, and used as boundary conditions for atmospheric models (e.g., Barnston *et al.* 2003). In the current 1-tier approach, a coupled ocean-atmosphere model is used to simultaneously predict SST and atmospheric variables of interest, typically temperature and precipitation. Operational centers, labeled Global Producing Centres of Long-range Forecasts by the World Meteorological Organization, make predictions with dynamical models. In fact, current prediction systems combine repeated simulations—termed *ensembles*, made up of *members* started from slightly different initial conditions—with different coupled models into a Multi-Model Ensemble (MME). These efforts started with the "Development of a European Multimodel Ensemble system for seasonal to inTERannual prediction" (DEMETER), a European project (Palmer *et al.* 2004). Efforts to increase access to the output from dynamical forecasts are more recent. The North American Multi-Model Ensemble (NMME; Kirtman *et al.* 2014), the prediction system exploited here, is one such system. It started sharing real-time forecasts in 2011. These are updated monthly and are openly accessible through the IRI Data Library (see Supporting Information for a brief tutorial).

To facilitate the production of national forecasts at PRESAO, IRI developed the *Climate Predictability Tool* (CPT; Mason and Tippett 2017). Using dynamical model output as the predictor field, Ndiaye *et al.* (2008) demonstrated the improvement in skill when using the 925 hPa wind field for Sahel rainfall instead of rainfall itself. The use of a dynamical prediction system over a statistical one is advantageous, because once a robust model-output-statistics (MOS) routine is put in place, such routine is independent of lead time. In contrast, because there is no guarantee that the predictors extracted from observations for a given lead time be the same for all lead times, development of statistical routines requires that a model be developed for each lead time. To illustrate this difference, let's presume that predictability in our region of interest is defined by ENSO. The skill of a statistical prediction system is constrained by the ability to identify the signature associated with ENSO evolution in observations at the desired lead time. In contrast, a dynamical prediction system relies on the system's ability to predict ENSO with the desired lead time. Ndiaye *et al.* (2011) first demonstrated the potential for increasing forecast lead time in the Sahel using dynamical models, highlighting the ability of the NCEP Climate Forecast System (CFS) to capture the El Niño-Southern Oscillation (ENSO). Sheen *et al.* (2017) showed skill in forecasts initialized in November for the following June-August season in the UK Met Office forecast system DePreSys.

Because it still relies primarily on statistical schemes, the PRESAO process does not attempt predictions earlier than April or May for the June-August and July-September seasons in the Sudano-Sahel. Here we report on the breakthrough in increased lead time of a skillful prediction for Sahel precipitation, which is made all the more robust by exploitation of a multi-model ensemble, the NMME. Secondly, we reflect on the spatio-temporal nature of predictability, specifically, its oceanic origin and implications for the provision of local information typically demanded by real-world decisions.

2 Data and Methods

Predictors are derived from the precipitation fields output by 5 models in the North American Multi-Model Ensemble (NMME), one per modeling group (Environment Canada, NASA/Goddard Space Flight Center/Global Modeling and Assimilation Office, NCAR/Center for Ocean-Land-Atmosphere Studies/Rosenstiel School for Marine and Atmospheric Sciences, NOAA/Geophysical Fluid Dynamics Laboratory, and NOAA/National Centers for Environmental Prediction/Climate Prediction Center). The model versions selected were current as of the 2018 PRESA-SS, which was held in Abidjan, Côte d'Ivoire from April 30 to May 4. Details of the simulations are reported in Table S1 in Supporting Information. The predictors are:

1. regional rainfall averages, over two domains of extremely different size: a sub-continental Sahel (10-20°N, 20°W-40°E), and a rectangular domain encompassing a single country (Senegal: 13-16°N, 17-12°W)
2. the full precipitation fields over the same domains specified in (1).

Predictands are derived from CHIRPS (University of California, Santa Barbara's Climate Hazards group InfraRed Precipitation with Station data; Funk *et al.* 2015). They mirror the two types of predictors defined above, that is, spatial average or explicit gridpoint-by-gridpoint field. Specifically, in the left and middle columns of the rows in Figure 2, discussed in the Results section, predictor and predictand quantities are the same. In addition, to measure the spatial coherence of predictions, we consider a third type of predictand:

3. the fraction of area characterized by abundant rains, that is, the portion of gridpoints in a region with rainfall anomalies greater than 0.5 times the local standard deviation.

We assess predictions at two different spatial scales, namely, sub-continental and country-wide, in part to highlight the large-scale nature of predictability of Sahelian climate, in part to give a sense of how such predictability is affected when national perspectives are taken into account. The specific choice of Senegal is relevant to discussions of the heterogeneity of climatic variations between the western and central Sahel (Lamptey 2008; Lebel and Ali 2009; Biasutti 2013; Salack *et al.* 2014; Panthou *et al.* 2018). We provide recipes to download predictor fields from the IRI Data Library in Supporting Information.

We are interested in seasonal prediction of precipitation accumulation for the core monsoon season, that is, July-September (JAS). We refer to the shortest lead time, that of a prediction for JAS made at the beginning of June, as a 1-month lead, and to the longest lead time, that of a prediction made for the same JAS target at the beginning of January, as a 6-month lead. For each model we compute the average of all available ensemble members, with ensemble size varying with model and/or between hindcasts (1981-2010) and real-time forecasts (2011-2016) as reported in Table S1. Single model ensemble means are weighed equally when averaged into the multi-model mean. We use *deterministic* measures of skill, that is, both Pearson and Spearman correlations (Becker *et al.* 2014; Barnston *et al.* 2017), meaning that we only take into consideration information derived from the ensemble mean, not from the ensemble spread.

3 Results

The top panel of Figure 2 shows time series of NMME predictions of Sahel-wide rainfall anomaly at different lead times, in color, and compares them to observations, in black. Qualitatively, it is possible to associate successes in prediction with the recurrence of La Niña events and abundant rainfall, for example, in 1988, 1998-99 and 2010, and of El Niño events and deficient rainfall, for example in 1987, 1997 and 2009. It is also possible to identify calamitous forecast failures, most notably, in 1984, the driest year in the 20th century.

3.1 Prediction skill as a function of lead time and spatial extent

Panels in the lower portion of Figure 2 quantify skill dependence on forecast start date. Skill is measured by correlation between predictions and observations of rainfall over the 35-year period of study (1982-2016: the 5% significance level with 33 degrees of freedom is 0.33, plotted in the thick grey dotted line). In each panel, the forecast made for the shortest lead time, at the beginning of June, is on the left, with lead time increasing to the right. Panels on the top of two rows in Figure 2 are for the sub-continental Sahel, panels on the bottom of two rows, for a box including Senegal. In each panel, solid lines denote Spearman correlation, dashed lines Pearson correlation, in the thick red line for the multi-model mean, and in thinner lines of different color for the single models.

The panels in the left column of the rows in Figure 2 represent correlations of regional averages, that is, the skill in predicting the spatially averaged anomaly in accumulation for the region under consideration. Predictor and predictand are the same. For the sub-continental Sahelian average (top row, left) multi-model mean values are remarkably consistent across lead times, varying between 0.5 and 0.6. Two models show skill comparable to the multi-model mean, those in the orange and turquoise lines. Orange model correlations are lowest for forecasts with start dates in January and February, and increase as lead time decreases. Turquoise model correlations are lowest for forecasts with an April start date. When Senegal-average rainfall is used to predict itself (bottom row, left) the situation is more unstable: (i) values are overall lower, (ii) the multi-model mean is surpassed by two models, in purple and especially in orange, and (iii) there is greater variation with start date, with a tendency for skill to increase as lead time decreases (with the notable exception of the orange model).

The panels in the middle column of the two rows in Figure 2 represent anomaly correlation as defined in Becker *et al.* (2014): correlations in time between the ensemble-mean predicted and observed fields, regridded to the same 1°x 1°grid in longitude and latitude, are first computed locally, at each gridpoint, then averaged over all gridpoints in the domain, again for the entire sub-continental Sahel in the top row, and for Senegal in the bottom row. Again, predictor and predictand are the same. The loss of skill when comparing gridpoint value (middle column) and regional average (left column) predictions is large. The average of local correlation values varies around 0.3 for the multi-model mean of sub-continental Sahel rainfall, against values between 0.5 and 0.6 for the regional average, and is consistently lower for single-model forecasts. The loss of skill is smaller in the case of Senegal, where it was lower to begin with.

To further characterize the nature of local predictability, the panels in the right column in the rows in Figure 2 depict the skill in predicting measures of spatial coherence using the Sahel regional average as predictor. In the top row, right column, Sahel average accumulation is used to predict the fraction of Sahelian domain covered by a positive anomaly 0.5 times the local standard deviation or greater. Ali and Lebel (2009) found consistency in the relationship between the magnitude of a regionally averaged anomaly and its spatial coherence, that is, the spatial extent of anomalies of the same sign. We interpret the comparison of skill in predicting the area with significant positive precipitation, in the top row, right column of Figure 2, with that in predicting the regional average, in the top row, left column, consistently with Ali and Lebel (2009). The regional average is a good measure of the strength of the signal: the stronger the signal, the larger the number of points behaving consistently with it. However, the loss of skill in the top row, middle column implies that exactly which points or locations will behave as predicted, and which will deviate from prediction, is unpredictable. In the bottom row, right column, Sahel average accumulation is used to predict Senegal average accumulation. Comparison of the bottom row, left and right columns in Figure 2 shows that the Sahel average is a better predictor of Senegal-average rainfall than the Senegal average itself, and that, once more, the orange model is more skillful than the multi-model mean.

3.2 Oceanic sources of predictability

To characterize the oceanic origin of NMME predictability, we regress predictions of sub-continental July-September average Sahel rainfall onto the simultaneous SST field. In Figure 3, the predicted rainfall index is regressed against predicted SST fields. In Figure S1 in Supporting Information, the same predicted rainfall index is regressed against observed SST fields. Panels in each row of Figure 3 (and Figure S1 in Supporting Information) are ordered by forecast start date, with the shortest lead, June, on the left, and the longest, January, on the right. The single row at the top of Figure 3 (and Figure S1) represents the multi-model mean. The following 5 rows represents each model's ensemble mean.

ENSO is the strongest source of predictability (Ndiaye *et al.* 2011). As expected, above-average Sahel rainfall is associated with the negative phase of ENSO, or La Niña conditions in the tropical Pacific (Janicot *et al.* 1996; Ward 1998; Giannini *et al.* 2003). The ENSO signature is present in the multi-model mean and in most models at all lead times. The long-lead skill in ENSO prediction of CMC2-CanCM4, the turquoise model in the rows in Figure 2 and the third row in the bottom of Figure 3, is well known (e.g., Gonzalez and Goddard 2016). This model's skill in predicting Sahel rainfall rivals that of the multi-model mean for all start dates except April, as noted in the previous subsection. This loss of skill in predicting Sahel rainfall in view of the model's skill in predicting ENSO can be interpreted as a relic of the spring predictability barrier.

The North Atlantic Ocean is a complementary source of predictability. Its contribution is best captured in two models, in the middle and especially in the bottom of two rows in Figure 3, which shape the multi-model mean picture. These correspond to the turquoise and orange lines in the rows in Figure 2, respectively. In the turquoise model, CMC2-CanCM4, the North Atlantic warming that is positively correlated with Sahel rainfall is extratropical in winter, and *propagates*, for lack of a better word, along the eastern boundary toward the tropics as the Sahelian rainy season approaches. In the orange model, NASA-GEOSS2S, the 5% statistical significance of positive regression values colors the entire North Atlantic basin starting in January, with largest values in extratropical latitudes, between 30 and 60°N. Regression values in the tropical North Atlantic weaken in March-May, and strengthen in June, while regression values with ENSO strengthen, peaking in May. In the multi-model mean, the fact that extra-tropical North Atlantic Ocean anomalies are strongest in winter (January and February start dates) supports relating these to wintertime North Atlantic Oscillation (NAO) forcing of SST anomalies. The late-spring strengthening of tropical North Atlantic anomalies is suggestive of re-emergence mediated by the response of trade winds to higher latitude SST anomalies (Seager *et al.* 2000; Chiang *et al.* 2003; Czaja *et al.* 2002; Clement *et al.* 2015), and is worthy of more detailed research.

3.3 Translating insights into practice at PRESA-SS

Finally, to relate directly to the practice of making seasonal predictions at PRESA-SS, we run CPT to test our insights about the spatio-temporal predictability of Sahel rainfall. As an illustration, we consider predictions made at the beginning of April, the start date most closely preceding the current PRESA-SS calendar. The predictor field is NMME multi-model mean precipitation in the 10-20°N, 20°W-40°E region. The predictand field is CHIRPS precipitation in the same region. To filter out spatial noise, when running Canonical Correlation Analysis (CCA; Bretherton *et al.* 1992), CPT extracts the dominant spatio-temporal pattern(s) applying Principal Component Analysis (PCA; Preisendorfer 1988) to the predictor and predictand fields. The resulting leading patterns of variability in each field are correlated in CCA to predict the best correlated pattern(s). CPT conveniently automates this routine, and provides cross-validated measures of skill. One such summary measure is the *goodness index*, defined as the spatial average of Kendall's tau rank correlation (Alfaro *et al.* 2018; Wilks 2011). CPT computes this index for all combinations of predictor and predictand Principal Components (PCs) and retains as the predictive model the one associated with the highest goodness index.

In our case, when we test retaining a maximum of 10 PCs of the predictor and predictand fields to predict a single precipitation pattern, the best model is composed of all 10 predictor and only 1, the first, predictand PCs. These retain respectively 96% of the total variance of the predictor field, and 46% of the predictand field. This model essentially predicts Sahel-wide rainfall. The goodness index varies between 0.247 and 0.359, when 1 and 10 predictor PCs are retained respectively. In comparison, when we use the Sahel-wide precipitation average as the single predictor, CPT computes a goodness skill of 0.226. These values are consistent with the multi-model mean anomaly correlation values plotted in Figure 2, in the top row, middle column, varying between 0.25 and 0.35. As a cross-check, to see whether we missed any potential sources of predictability, we extract the 10 predictor (precipitation) PCs from CPT and correlate them with predictions of precipitation and SST in the NMME multi-model mean. The first 5 patterns are shown in Figure 4. The first is a Sahelian pattern. Not surprisingly, over the 1982-2016 period it correlates strongly with ENSO. The second is a Gulf of Guinea pattern which strongly correlates with local SSTs, but has no projection on the Sahel (Giannini *et al.* 2003, 2005). Despite the intriguing SST patterns of PCs 3 and 4 in the North Atlantic, the projection onto Sahel rainfall of the remaining patterns is non-existent. This behavior raises the concern that a model that essentially takes as many predictor PCs as are available may be contaminated with artificial skill—statistical skill that has no physical counterpart—and strengthens our conclusion that the predictability is all in a Sahel-wide pattern.

4 Conclusions

We assessed the deterministic skill of the North American multi-model ensemble (NMME), an operational, state-of-the-art seasonal climate prediction system, in predicting July-September precipitation over the Sahel. We found skill in predicting regionally and seasonally averaged rainfall anomalies as early as February/March. Skill in the multi-model mean is the combination of skillful predictions of ENSO in one model, and of North Atlantic sea surface temperatures in another. That such distinct behaviors simply add up to shape multi-model mean skill is exemplary of the value of multi-model ensemble prediction systems. Interestingly, the skill of a system composed of only these two *best* models, which correspond to the turquoise and orange lines in the rows in Figure 2, is more variable: it is higher in some instances, lower in others, than the skill of the system based on all models considered here (see Figure S2 in Supporting Information). At the smaller scale of Senegal, the greater skill of the model that best predicts North Atlantic temperatures is indicative of the greater relative influence of the adjacent basin on the westernmost portion of the Sahel. This behavior points to east-west differences in sub-regional dynamics that merit following up.

We emphasize that this level of skill, consistent with the large-scale, oceanic origin of predictability, is achieved at the very largest spatial scales, that is, the sub-continental scale of the entire Sahel. As illustrated in the rows in Figure 2, the skill in regionally averaged precipitation (in the left column) is different from, and significantly larger than the regional average of local skill (in the middle column). Further, the stronger the predictable signal, captured in SST anomalies, the greater the spatial coherence of the outcome. A local forecast scheme could be envisioned that calculated probabilities based solely on the regionally averaged signal weighted by its strength.

The skill at lead times of 3-4 months on the start of the rainy season that is described here is a significant advancement. Its practical implications are profound, considering that the current regional climate outlook forum process, largely based on statistical prediction models, convenes in May, or April at the earliest. Awaiting a further quantitative assessment, this level of skill should be sufficient for the timely communication of an early qualitative outlook. This may be relevant for national governments to assure timely approval of the budget items supporting the agricultural sector, in the form of purchasing and stocking for inputs that are best adapted to the predicted character of the upcoming season. It may be even more relevant for institutions concerned with regional food security, such as the CILSS (Comité Permanent Inter-États de Lutte contre la Sécheresse dans le Sahel, or Permanent Interstate Committee for Drought Control in the Sahel) and its global partners, including the UN World Food Programme, the Famine Early Warning Systems Network, and the Réseau de Prévention des Crises Alimentaires, because it could buy them more time to secure donor funding ahead of a potential large-scale crisis, such as a repeat drought year.

Finally, we find confirmation that Sahelian variability is shaped by the interplay of independent, North Atlantic and global tropical, sources of predictability, encapsulated in the North Atlantic Relative Index (Giannini *et al.* 2013). Indeed, the competition in warming between North Atlantic and global tropical oceans under the influence of greenhouse gases is one way to interpret the increased interannual variability that is qualitatively manifest in any Sahelian rainfall time series since the mid-1990s, including that in Figure 1—behavior which makes seasonal prediction all the more valuable.

Acknowledgments

The data used in this study were accessed via the IRI Data Library, at:

http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/

and *http://iridl.ldeo.columbia.edu/SOURCES/.UCSB/.CHIRPS/*

The original data repositories can be found at:

https://www.nws.noaa.gov/ost/CTB/nmme.htm

and *https://chc.ucsb.edu/data/chirps/*

AG acknowledges the support of the National Aeronautics and Space Administration through grant NNX16AN29G (AST-2/SERVIR) and of the French Programme d'Investissements d'Avenir de l'Agence Nationale de la Recherche through contract ANR-17-MPGA-0015 (PRODUCT). AG and CPK acknowledge the support of Columbia University's World Project *ACToday*.

References

AGRHYMET. (2010). Le Sahel face aux changements climatiques. *Bulletin mensuel*. (Numéro Spécial; www.agrhymet.ne)

Alfaro, E. J., Chourio, X., Muñoz, A. G., & Mason, S. J. (2018). Improved seasonal prediction skill of rainfall for the Primera season in Central America. *Int. J. Climatol.*, *38*, e255–e268. (DOI:10.1002/joc.5366)

Ali, A., & Lebel, T. (2009). The Sahelian standardized rainfall index revisited. *Int. J. Climatol.*, *29*, 1705–1714.

Baddour, O. (1998). Prévision climatique en Afrique; climate forecast in Africa. *WMO/TD*, *No. 927*. (World Meteorological Organization; African Centre of Meteorological Application for Development; Climate Information and Prediction Services)

Barnston, A. G., Mason, S. J., Goddard, L., DeWitt, D. G., & Zebiak, S. E. (2003). Multimodel ensembling in seasonal climate forecasting at IRI. *Bull. Amer. Meteor. Soc.*, *84*, 1783–1796.

Barnston, A. G., Tippett, M. K., Ranganathan, M., & L'Heureux, M. L. (2019). Deterministic skill of ENSO predictions from the North American Multimodel Ensemble. *Clim. Dyn.*, *53*, 7215-7234. (https://doi.org/10.1007/s00382-017-3603-3)

Becker, E., van den Dool, H., & Zhang, Q. (2014). Predictability and forecast skill in NMME. *J. Climate*, *27*, 5891–5906.

Biasutti, M. (2013). Forced Sahel rainfall trends in the CMIP5 archive. *J. Geophys. Res.*, *118*, 1613–1623.

Bretherton, C. S., Smith, C., & Wallace, J. M. (1992). An intercomparison of methods for finding coupled patterns in climate data. *J. Climate*, *5*, 541–560.

Chiang, J. C. H., Biasutti, M., & Battisti, D. S. (2003). Sensitivity of the Atlantic ITCZ to Last Glacial Maximum boundary conditions. *Paleoceanography*, *18*, 1094. (doi: 10.1029/2003PA000916)

Clement, A., Bellomo, K., Murphy, L. N., Cane, M. A., Mauritsen, T., Rädel, G., & Stevens, B. (2015). The Atlantic Multidecadal Oscillation without a role for ocean circulation. *Science*, *350*, 320–324.

Czaja, A., van der Vaart, P., & Marshall, J. (2002). A diagnostic study of the role of remote forcing in tropical Atlantic variability. *J. Climate*, *15*, 3280–3290.

Folland, C., Owen, J., Ward, M. N., & Colman, A. (1991). Prediction of seasonal rainfall in the Sahel region using empirical and dynamical methods. *J. Forecasting*, *10(1-2)*, 21–56.

Funk, C., Nicholson, S. E., Landsfeld, M., Klotter, D., Peterson, P., & Harrison, L. (2015). The centennial trends Greater Horn of Africa precipitation dataset. *Scientific Data*, *2*, 150050. (DOI: 10.1038/sdata.201)

Giannini, A., Salack, S., Lodoun, T., Ali, A., Gaye, A. T., & Ndiaye, O. (2013). A unifying view of climate change in the Sahel linking intra-seasonal, interannual and longer time scales. *Environ. Res. Lett.*, *8*, 024010.

Giannini, A., Saravanan, R., & Chang, P. (2003). Oceanic forcing of Sahel rainfall on interannual to interdecadal time scales. *Science*, *302*, 1027–1030. (doi:10.1126/science.1089357)

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

Gonzalez, P. L. M., & Goddard, L. (2016). Long-lead ENSO predictability from CMIP5 decadal hindcasts. *Clim. Dyn.*, *46*, 3127–3147. (DOI 10.1007/s00382-015-2757-0)

Janicot, S., Moron, V., & Fontaine, B. (1996). Sahel droughts and ENSO dynamics. *Geophys. Res. Lett.*, *23*, 515–518.

Kirtman, B. P., Min, D., Infanti, J. M., Kinter III, J. L., Paolino, D. A., Zhang, Q., ... Peng, P. (2014). The North American Multi-Model Ensemble: phase 1 seasonal to interannual prediction; phase 2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, *95(4)*, 585–601.

Lamptey, B. L. (2008). Comparison of gridded multisatellite rainfall estimates with gridded gauge rainfall over West Africa. *J. Appl. Meteor. Clim.*, *47*, 185–205.

Le Barbé, L., & Lebel, T. (1997). Rainfall climatology of the HAPEX-Sahel region during the years 1950-1990. *J. Hydrology*, *188–189*, 43–73.

Lebel, T., & Ali, A. (2009). Recent trends in the Central and Western Sahel rainfall regime (1990–2007). *J. Hydrol.*, *375*, 52–64. (doi:10.1016/j. jhydrol.2008.11.030)

Mason, S. J., & Tippett, M. K. (2017). Climate predictability tool version 15.5.10. (https://academiccommons.columbia.edu/doi/10.7916/D8668DCW)

Ndiaye, O., Goddard, L., & Ward, M. N. (2008). Using regional wind fields to improve general circulation model forecasts of JulySeptember Sahel rainfall. *Int. J. Climatology*, *29*, 1262–1275.

Ndiaye, O., Ward, M., & Thiaw, W. M. (2011). Predictability of seasonal Sahel rainfall using GCMs and lead-time improvements through the use of a coupled model. *J. Climate*, *24*, 1931–1949.

Nicholson, S. E. (2005). On the question of the "recovery" of the rains in the West African Sahel. *J. Arid Environ.*, *63*, 615–641.

Ogallo, L. A., Boulahya, M. S., & Keane, T. (2000). Applications of seasonal to interannual climate prediction in agricultural planning and operations. *Agricultural and Forest Meteorology*, *103*, 159–166.

Ouedraogo, I., Diouf, N. S., Ouédraogo, M., Ndiaye, O., & Zougmoré, R. B. (2018). Closing the gap between climate information producers and users: assessment of needs and uptake in Senegal. *Climate*, *6*, 13. (doi:10.3390/cli6010013)

Palmer, T. N., & Coauthors. (2004). Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). *Bull. Amer. Meteor. Soc.*, *85*, 853–872.

Panthou, G., Lebel, T., Vischel, T., Quantin, G., Sane, Y., Ba, A., ... Diop-Kane, M. (2018). Rainfall intensification in tropical semi-arid regions: the Sahelian case. *Environ. Res. Lett.*, *13*, 064013. (https://doi.org/10.1088/1748-9326/aac334)

Preisendorfer, R. W. (1988). Principal component analysis in meteorology and oceanography. , 418 pp. (posthumously compiled and edited by C. Mobley. In: Developments in Atmospheric Science, v. 17. Amsterdam, New York: Elsevier. xviii, 425 pp. Elsevier)

Rio, C., DelGenio, A. D., & Hourdin, F. (2019). Ongoing breakthroughs in convective parameterization. *Current Climate Change Reports*, *5*, 95–111. (https://doi.org/10.1007/s40641-019-00127-w)

Salack, S., Giannini, A., Diakhaté, M., Gaye, A. T., & Muller, B. (2014). Oceanic influence on the sub-seasonal to interannual timing and frequency of extreme dry spells over the West African Sahel. *Clim. Dyn.*, *42*, 189–201.

Seager, R., Kushnir, Y., Visbeck, M., Naik, N., Miller, J., Krahmann, G., & Cullen, H. (2000). Causes of Atlantic Ocean climate variability between 1958 and 1998. *J. Climate*, *13*, 2845–2862.

Sheen, K. L., Smith, D. M., Dunstone, N. J., Eade, R., Rowell, D. P., & Vellinga, M. (2017). Skilful prediction of Sahel summer rainfall on interannual and multi-year timescales. *Nature Communications*, *8*, 14966. (DOI:10.1038/ncomms14966)

Tall, A., Coulibaly, J. A., & Diop, M. (2018). Do climate services make a difference? A review of evaluation methodologies and practices to assess the value of climate information services for farmers: implications for Africa. *Climate Services*, *11*, 1–12. (https://doi.org/10.1016/j.cliser.2018.06.001)

Traoré, S. B., Ali, A., Tinni, S. H., Samaké, M., Garba, I., Maigari, I., . . . Dieye, P. O. (2014). AGRHYMET: A drought monitoring and capacity building center in the West Africa Region. *Weather and Climate Extremes*, *3*, 22–30.

Ward, M. N. (1998). Diagnosis and short-lead time prediction of summer rainfall in tropical North Africa at interannual and multidecadal timescales. *J. Climate*, *11*, 3167–3191.

Ward, M. N., Folland, C. K., Maskell, K., Colman, A. W., Rowell, D. P., & Lane, K. B. (1993). Experimental seasonal forecasting of tropical rainfall at the UK Meteorological Office. In *Prediction of Interannual Climate Variations* (pp. 197–216). Springer-Verlag, Berlin, Heidelberg. (J. Shukla, Ed.)

Wilks, D. S. (2011). Statistical methods in the atmospheric sciences. Academic Press: New York.

Figure Captions

Figure 1 Standardized Sahelian rainfall index from rain gauge observations. Updated from and calculated over the same region as Ali and Lebel (2009), covering the countries of the CILSS, from the Atlantic coast to Chad included.

Figure 2 (Top panel) Sahel rainfall time series, the average over the "sub-continental Sahel" region in 10-20°N, 20°W-40°E, in the multi-model mean of 5 NMME models, in colors according to prediction start time (see legend insert), and in observations (CHIRPS) in black. (Two rows of panels) Skill of NMME predictions for the July-September season. Predictions are started from the previous January, on the right in each panel, to the June immediately before the season, on the left, corresponding to lead times from 6 months to 1 month. Skill is measured by Spearman (solid line) and Pearson (dashed line) correlations over 1982-2016: the thick, red line is for the multi-model mean, the thinner lines of different colors are for single models, with the thick grey dotted line representing the 5% significance level. (Top row) Sahel-wide predictions. (Bottom row) predictions over Senegal. (Left column) prediction of the spatial average. (Middle column) gridpoint prediction, averaged over the area. (Right column) predictions based on Sahel-average rainfall of (top row) the fraction of Sahel area under positive rainfall anomaly and (bottom row) Senegal average rainfall, based on Sahel-average rainfall.

Figure 3 Regressions of predicted July-September Sahel rainfall with simultaneously predicted sea surface temperatures, for start dates from January on the right to June on the left. The separate, top row is for the multi-model mean. Rows below are for single models. Values are in degrees Celsius: contour starts at 0.1 degrees and is every 0.2 degrees. Color, red for positive values and blue for negative values, indicates statistical significance of the regression values at the 5% level.

Figure 4 Correlation maps of the 5 leading NMME precipitation predictor fields in the region 10-20°N, 20°W-40°E extracted from CPT with (left) precipitation and (right) SST. Predictions are made in April for the July-September season over 1982-2016. Only values statistically significant at the 5% level, corresponding to a value of 0.33, are plotted. Contour is every 0.2, starting at 0.4.

Supporting Information for "Advances in the lead time of Sahel rainfall prediction with the North American Multi-Model Ensemble"

A. Giannini[1,2], A. Ali[3], C. P. Kelley[1], B. L. Lamptey[4,5], B. Minoungou[3], O. Ndiaye[1,6]

[1]International Research Institute for Climate and Society, The Earth Institute at Columbia University, Palisades NY, U.S.A.

[2]Now at Laboratoire de Météorologie Dynamique/IPSL, École Normale Supérieure, PSL Research University, Sorbonne Université, École Polytechnique, IP Paris, CNRS, Paris, France

[3]Centre Régional AGRHYMET, Niamey, Niger

[4]Cheney Fellow, University of Leeds, School of Earth and Environment, UK

[5]Formerly at the African Centre of Meteorological Applications for Development, Niamey, Niger

[6]Agence Nationale de l'Aviation Civile et de la Météorologie, Dakar, Senegal

Corresponding author: Alessandra Giannini, International Research Institute for Climate and Society, The Earth Institute at Columbia University, Palisades NY, U.S.A. and Laboratoire de Météorologie Dynamique/IPSL, École Normale Supérieure, PSL Research University, Sorbonne Université, École Polytechnique, IP Paris, CNRS, Paris, France. (alessandra.giannini@lmd.ens.fr)

Contents of this file

Introduction

This Supplementary Information includes additional Figures and a Table. Figures S1 and S2 are extensions of figures 3 and 2, respectively. Table S1 provides detailed information about the model simulations analyzed. It also includes a template recipe to define and download rainfall predictors from the NMME archive maintained in the IRI Data Library.

Figure S1. The Sahel precipitation index derived from NMME predictions is correlated with the *observed* patterns of sea surface temperature (SST). The intent is to compare these correlation patterns with those that result from correlating the same index with *predicted* SSTs, in Figure 3, to gauge the extent to which NMME captures the SST-Sahel precipitation relationship.

Figure S2. The left panel is repeated from the top row, left column of Figure 2, which details the skill dependence on lead time for the 5-model ensemble. It is compared to the same skill for a 2-model ensemble based on the *best* models, in the right panel. As discussed in the Conclusions section, while the skill of the smaller ensemble is at times larger, the skill of the larger ensemble is more stable across lead times.

Table S1. This table contains details about the simulations analyzed, made with the 5 models named in the Data and Methods section of the article.

Accessing the NMME in the IRI Data Library

The IRI Data Library (IRIDL) maintains a regularly updated archive of NMME model output, including hindcasts and proper real-time forecasts (see Table S1) at `http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/`. In this note we document the steps to (1) select model output, *e.g.*, choose models and variables, and define spatial domain, start date and lead time, (2) combine models into the multi-model mean, and (3) download in a format compatible with the Climate Predictability Tool (CPT).

Working behind the scenes is *ingrid* (See `http://iridl.ldeo.columbia.edu/dochelp/Documentation/funcindex.html` for function documentation), the coding language germane to the IRIDL, developed to select, analyze and visualize data in a web browser environment. The coding becomes visible by clicking on the `Expert mode` tab. This action opens a window, to which the lines of code described below, in `Courier` font, can be copied and pasted directly. (Clicking on the `OK` button below the `Expert mode` window executes the code.)

Let us start with an NMME model that archives hindcasts and forecasts in the same directory, and select the precipitation (`prec`) variable. From `http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/`, first select the model, `COLA-RSMAS-CCSM4`, by clicking on its name, then the `MONTHLY` directory, which contains the archive of predictions, and finally the variable, `prec`. These actions are explicited in `Expert mode` as:

```
SOURCES .Models .NMME .COLA-RSMAS-CCSM4 .MONTHLY .prec
```

We specify the geographic domain of interest by applying the function `RANGEEDGES` to the `X` and `Y` grids:

```
X -20 40 RANGEEDGES

Y 10 20 RANGEEDGES
```

where `X` and `Y` are typically longitude and latitude respectively. In *ingrid* longitudes west of the Greenwich meridian and latitudes south of the Equator are identified by negative values, longitudes east and latitudes north, by positive values. The ranges set above correspond to the sub-continental Sahel as defined in this study: 10-20°N, 20°W-40°E.

We average gridpoints in the domain into a regional index with the command line

```
[X Y] average
```

Start date and lead time of prediction are typically indicated by `S` and `L` respectively. The following:

```
S (0000 1 Apr ) VALUES

L (3.5) (5.5) RANGEEDGES
```

denotes a prediction made on 1 April, with lead times comprised between 3.5 and 5.5 months. Since months are typically identified by the mid-month date, *e.g.*, 16 Jan, 16 Feb, 16 Mar, etc., this combination of selections on `S` and `L` identifies the July-September period. A prediction made on 1 April for the June-August period looks like this:

```
S (0000 1 Apr) VALUES

L (2.5) (4.5) RANGEEDGES
```

In addition, `M` typically denotes ensemble member. Therefore, `[M] average` denotes the ensemble mean over all members available.

The combined specifications for the case of a model archiving hindcasts and forecasts in the same directory, resulting in April predictions for July-September Sahel average rainfall over the entire period covered by the NMME (1982 to present), can be copied and pasted directly to the `Expert mode` window, where they look like this:

```
SOURCES .Models .NMME .COLA-RSMAS-CCSM4 .MONTHLY .prec
X -20 40 RANGEEDGES
Y 10 20 RANGEEDGES
S (0000 1 Apr ) VALUES
L (3.5) (5.5) RANGEEDGES
[L] /keepgrids average
[M X Y]average
```

The same specifications for the case of a model archiving hindcasts and forecasts in separate directories look like this:

```
SOURCES .Models .NMME .CMC2-CanCM4 .HINDCAST .MONTHLY .prec
  X -20 40 RANGEEDGES Y 10 20 RANGEEDGES
  S (0000 1 Apr ) VALUES L (3.5) (5.5) RANGEEDGES
  [L] /keepgrids average
  [M X Y]average
SOURCES .Models .NMME .CMC2-CanCM4 .FORECAST .MONTHLY .prec
  X -20 40 RANGEEDGES Y 10 20 RANGEEDGES
  S (0000 1 Apr ) VALUES L (3.5) (5.5) RANGEEDGES
  [L] /keepgrids average
  [M X Y]average
  appendstream
```

In other words, the specifications are repeated for hindcasts and forecasts, and the two data streams are combined using the function `appendstream`.

To combine more models into the multi-model mean, models are added together and divided by their number. Using the two models described thus far:

```
SOURCES .Models .NMME .CMC2-CanCM4 .HINDCAST .MONTHLY .prec
  X -20 40 RANGEEDGES Y 10 20 RANGEEDGES
  S (0000 1 Apr ) VALUES L (3.5) (5.5) RANGEEDGES
  [L] /keepgrids average [M X Y]average
SOURCES .Models .NMME .CMC2-CanCM4 .FORECAST .MONTHLY .prec
  X -20 40 RANGEEDGES Y 10 20 RANGEEDGES
  S (0000 1 Apr ) VALUES L (3.5) (5.5) RANGEEDGES
  [L] /keepgrids average [M X Y]average
  appendstream
  SOURCES .Models .NMME .COLA-RSMAS-CCSM4 .MONTHLY .prec
  S (0000 1 Apr ) VALUES L (3.5) (5.5) RANGEEDGES
  X -20 40 RANGEEDGES Y 10 20 RANGEEDGES
  [L] /keepgrids average
  [M X Y]average
  add 2 div
```

yields the result sought, with the last line, `add 2 div`, signifying that the two data streams are first added up and then divided by 2. Note that in *ingrid* space and line break are equivalent.

Manipulations to consider in order to tailor data selection to user needs include:

- selecting a different spatial domain. The domain is set using the functions `RANGE` or `RANGEEDGES` operating on the `X` and `Y` grids. For example, in the country case investigated in this study, the rectangular domain comprising Senegal, defined by 13-16°N, 17-12°W, is rendered as `Y 13 16 RANGE X -17 -12 RANGE`;

- including the line `[X Y]average` computes a regional average. Eliminating it results in longitude/latitude fields, or maps of the predictor variable;

- start date (`S`) and lead time (`L`) need to be adjusted consistently, depending on the time that the prediction is made, and the period to be predicted.

The end result is an up-to-date time series of Sahel average precipitation predictions, concatenating hindcasts and forecasts starting in 1982, and using the 5 models analyzed in this study. In `Expert mode` it looks like this:

```
SOURCES .Models .NMME .CMC2-CanCM4 .HINDCAST .MONTHLY .prec
X -20 40 RANGEEDGES Y 10 20 RANGEEDGES
S (0000 1 Apr ) VALUES L (3.5) (5.5) RANGEEDGES
[L] /keepgrids average
[M X Y]average
SOURCES .Models .NMME .CMC2-CanCM4 .FORECAST .MONTHLY .prec
X -20 40 RANGEEDGES Y 10 20 RANGEEDGES
S (0000 1 Apr ) VALUES L (3.5) (5.5) RANGEEDGES
[L] /keepgrids average
```

```
[M X Y]average
appendstream
SOURCES .Models .NMME .NCEP-CFSv2 .HINDCAST .MONTHLY .prec
X -20 40 RANGEEDGES Y 10 20 RANGEEDGES
S (0000 1 Apr ) VALUES L (3.5) (5.5) RANGEEDGES
[L] /keepgrids average
[M X Y]average
SOURCES .Models .NMME .NCEP-CFSv2 .FORECAST .EARLY_MONTH_SAMPLES .MONTHLY
.prec
X -20 40 RANGEEDGES Y 10 20 RANGEEDGES
S (0000 1 Apr ) VALUES L (3.5) (5.5) RANGEEDGES
[L] /keepgrids average
[M X Y]average
appendstream
add
SOURCES .Models .NMME .NASA-GEOSS2S .HINDCAST .MONTHLY .prec
X -20 40 RANGEEDGES Y 10 20 RANGEEDGES
S (0000 1 Apr ) VALUES L (3.5) (5.5) RANGEEDGES
[L] /keepgrids average
[M X Y]average
SOURCES .Models .NMME .NASA-GEOSS2S .FORECAST .MONTHLY .prec
X -20 40 RANGEEDGES Y 10 20 RANGEEDGES
```

```
S (0000 1 Apr ) VALUES L (3.5) (5.5) RANGEEDGES

[L] /keepgrids average

[M X Y]average

appendstream

add

SOURCES .Models .NMME .GFDL-CM2p1-aer04 .MONTHLY .prec

X -20 40 RANGEEDGES Y 10 20 RANGEEDGES

S (0000 1 Apr ) VALUES L (3.5) (5.5) RANGEEDGES

[L] /keepgrids average

[M X Y]average

add

SOURCES .Models .NMME .COLA-RSMAS-CCSM4 .MONTHLY .prec

X -20 40 RANGEEDGES Y 10 20 RANGEEDGES

S (0000 1 Apr ) VALUES L (3.5) (5.5) RANGEEDGES

[L] /keepgrids average

[M X Y]average

add

5 div

/missing_value -999.0 def

SLtoT L removeGRID
```

When needed, forecasts are appended to hindcasts using `appendstream`, then each model is added to the previous, using `add`, and finally, the total sum is divided by the number of models, in this case 5, using `div`.

When model output is ready for download, the user clicks first on the `Data files` tab above the `Expert mode` window, then chooses the relevant format on the following web page. In addition to a format compatible with CPT, frequently used formats include formats for direct input into NCL or Matlab scripts, and downloads into netCDF files.

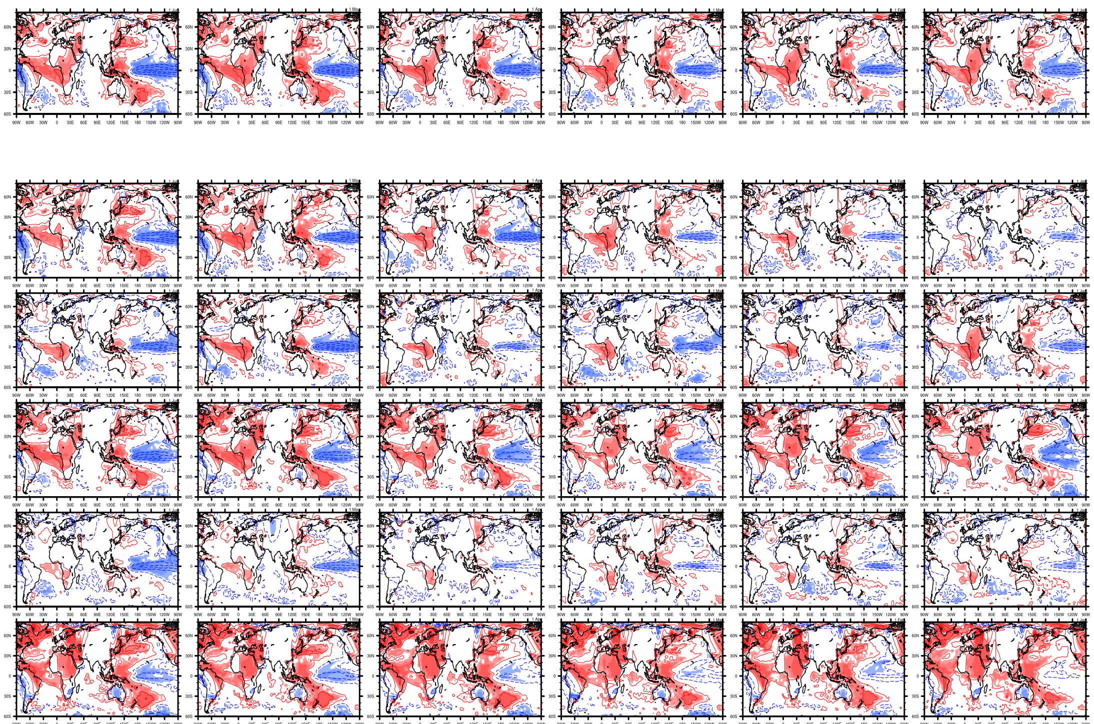

Figure S1. Regressions of predicted July-September Sahel rainfall with observed July-September sea surface temperatures, for start dates from January on the right to June on the left. The figure structure is the same as in Figure 3. The separate, top row is for the multi-model mean. Rows below are for single models. Values are in degrees Celsius: contour starts at 0.1 degrees and is every 0.2 degrees. Color, red for positive values and blue for negative values, indicates statistical significance of the regression values at the 5% level.

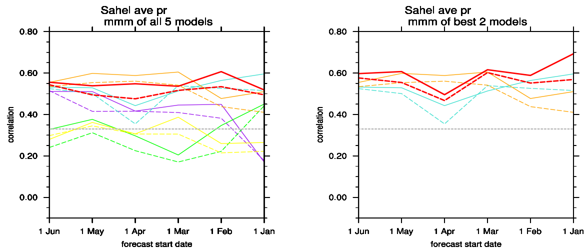

Figure S2. Skill of NMME predictions for the July-September season. Predictions are started from the previous January, on the right in each panel, to the June immediately before the season, on the left, corresponding to lead times from 6 months to 1 month. Skill is measured by Spearman [solid line] and Pearson [dashed line] correlations over 1982-2016: the thick, red line is for the multi-model mean, the thinner lines of different colors are for single models, with the thick grey dotted line representing the 5% significance level. The left panel is the same as in Figure 2, for the multi-model system of all 5 NMME models considered in this study. The right panel is for the 2-model system using only the *best* models.

Model	*Lead time*	*Hindcast period*	*Ens Size*	*Forecast period*	*Ens Size*
CMC2-CanCM4	0-11	Jan 1981-Dec 2010	10	Jan 2011 to present	10
COLA-RSMAS-CCSM4	0-11	Jan 1982-Dec 2010	10	Jan 2011 to present	10
NASA-GEOSS2S	0-8	Feb 1981-Jan 2017	4	Nov 2017 to present	10
NCEP-CFSv2	0-9	Jan 1982-Dec 2010	24	Mar 2011 to present	32
GFDL-CM2p1-aer04	0-11	Jan 1982-Dec 2010	10	Jan 2011 to present	10

Table S1. Specifics of the NMME simulations used. The *lead time* is in months.