Chapter 1

Assessing Seasonal Climate Forecasts Over Africa to Support Decision-Making

Niko Wanders* and Eric F. Wood†

**Department of Physical Geography,*
Utrecht University, Utrecht, The Netherlands
n.wanders@uu.nl
†Department of Civil and Environmental Engineering,
Princeton University, Princeton, NJ 08544, USA

Recent drought events like in the 2011 Horn of Africa and the ongoing drought in California have an enormous impact on nature and society. Reliable seasonal weather outlooks are critical for drought management and other applications like, crop modelling, flood forecasting and planning of reservoir operation, and would help reduce the potential economic damage from extremes as well as help optimize crop yields during more normal weather years from improved agricultural management. However, most seasonal forecasts are limited by low spatial and temporal model resolutions. The newly released North American Multi-Model Ensemble phase 2 (NMME-2), provides subseasonal forecast that increase the temporal resolution from monthly to daily, enabling subseasonal forecasting for end-user applications that rely on a daily temporal resolution. In this study we give an overview of the current status of the NMME subseasonal forecasts ensembles, their skill over the African continent and the forecast skill of the ensembles for applications related to agriculture and hydrology. We show that the NMME-2 subseasonal forecasts are significantly skilful for both precipitation and 2 m air temperature for large parts of Africa. The precipitation forecasts are skilful up to a lead time of two months, while temperature anomaly forecasts have a significant skill beyond the three months lead for most of Africa. Potential applications that would benefit from the new NMME-2 ensemble were studied in more detail for West Africa. We show that the models have a significant skill in forecasting the onset of the annual rain season in West Africa, and thereby the start of the growing season. Additionally, the models have a significant forecast skill to predict the onset and peak of the high flow season for most parts of West Africa. The low uncertainty in the forecasts compared to the observed anomalies indicates that local stakeholders will benefit from the high temporal resolution that the NMME-2 provides. Results encourage future research into the potential of the new subseasonal NMME-2 forecast ensemble to forecast more specific end-user applications and climate services, which require skilful high temporal resolution forecasts.

1. Introduction

Floods and drought events occur in all regions of the world with large societal impact (Kundzewicz and Kaczmarek 2000). Recent drought events like in the Horn of Africa (United Nations 2011; Sida *et al.* 2012) and the on-going drought in California (Howitt *et al.* 2014) have an enormous impact on nature and society. The 2011/2012 drought over the Horn of Africa was an example where food aid and other international help were not deployed until it was too late and the drought was already established. This led to an increased number of fatalities and economical damage (e.g. crop and livestock losses). With adequate planning supported by long-term meteorological forecasts and hydrological modelling, the severe impact of this drought could have been reduced. Preventive measures could have been taken and valuable resources could have been reserved for the extreme drought conditions. These drought

Bridging Science and Policy Implication for Managing Climate Extremes
Edited by Hong-Sang Jung and Bin Wang

events clearly show the need for an early warning system that allows accurate monitoring and forecasting of drought conditions and other natural disasters at a continental to global scale. This also calls for improved high resolution (both in space and time) long-term meteorological forecasts.

For short-range meteorological forecasts (up to 2 weeks in advance) weather model forecasts are available at high spatial and temporal resolution from a number of centers (Yan and van der Dool 2011; Magnusson and Källén 2013). However, to increase the time that is needed to optimally use weather information for a range of end-user applications, it is important to extend the forecast range beyond the two-week period. Applications that will benefit from information from these extended forecasts include, amongst others, crop modelling, flood and drought forecasting, and planning of reservoir operations. Seasonal forecast models produce weather outlooks that range from 14 days to one year, thereby, bridging the gap between weather and climate models. They are available at a coarser resolution and lower temporal resolution (typically monthly timescale and 1° spatial resolution) compared to the high resolution weather models. Studies by Yuan *et al.* (2013) and the DEWFORA project (Dutra *et al.* 2014) showed that although they have a lower resolution, the seasonal forecasts still have the potential to detect anomalies in precipitation, surface air temperature and soil moisture over Africa up to several months in advance. However, it remains unknown how this seasonal forecast signal can impact the forecasts of the crop growing season or other hydrological variables. This is due to the fact that the monthly resolution is too coarse for many applications (e.g., onset of the growing season and start of river flows). To overcome this existing gap between the monthly scale data and climate services that rely on higher temporal resolution forecasts (i.e. daily), the North American Multi-Model Ensemble (NMME) provides subseasonal forecasts. These new higher temporal resolution data could open new potential applications for end users that want to use subseasonal forecasts in their decision support systems and for their climate services.

In this study we provide an overview of the current status of the NMME subseasonal forecasts models, their skill over the African continent and the forecast skill of the ensemble set for applications related to agriculture and hydrology. First, we look at the subseasonal forecast skill of the models and validate the forecasts of precipitation and air temperature over Africa. Thereafter, we study the ability of the models to forecast the onset of the growing season over Western Africa. Finally, the skill in reproducing the river discharge over Western Africa is discussed.

2. Material and Methods

2.1. *Seasonal forecast models*

In this study we use the forecasts from the North American Multi-Model Ensemble phase 2 (NMME-2) project that provides subseasonal forecasts for a 31-year period at a daily temporal resolution (Kirtman *et al.* 2014). NMME-2 provides the first multi-model ensemble that produces seasonal forecast with a daily temporal resolution. Due to the improved temporal resolution this newly available forecast ensemble can help to enhance the added value of seasonal forecasts for climate service.

We assess the subseasonal forecast skill of the NMME-2 ensemble for daily precipitation and 2 m air temperature forecasts over Africa using the output from four readily available NMME-2 models for over the period 1982–2012 (Kirtman *et al.* 2014). The models used in the analysis are: CanCM3 (Merryfield *et al.* 2013), CanCM4 (Merryfield *et al.* 2013), FLOR-B01 (Vecci *et al.* 2014) and CCSM4 (Hurrell *et al.* 2013), where for CCSM part of the forecasts is not available (Table 1). Two additional models are part of the NMME-2 ensemble, however, at the time of writing this paper their precipitation forecasts

Table 1. Number of ensemble member of the individual NMME-2 models, including data availability and applied reprocessing.

	CanCM3	CanCM4	FLOR-B01	CCSM4
Temperature data availability	100%	100%	100%	71.4%
Precipitation data availability	100%	100%	100%	76.5%
Resampling	None	None	Bilinear	Bilinear
Ensemble members	10	10	12	10

were not available. The added value of these incomplete models would be hampered by the absence of important meteorological variables like precipitation.

2.2. *Reference datasets*

To validate the subseasonal forecasts over Africa made by the NMME-II ensemble, we use the Princeton Global Forcing (PGF, Sheffield *et al.* 2006) dataset. This dataset covers the NMME-2 re-forecast period 1982–2012 and is available with a daily temporal resolution at a 0.25° spatial resolution at the global scale, and was aggregated up to 1° to match the spatial resolution of NMME-2. Details on the dataset can be found in the respective publication.

2.3. *Hydrological modelling*

To evaluate the ability to use the NMME-2 for a hydrological decision support system over Africa, we used the hydrological model VIC (Liang *et al.* 1994; 1996). This model allows us to propagate the precipitation and temperature forecasts and make forecasts of hydrological variables (e.g. soil moisture, river discharge and baseflow). VIC is widely used in many hydrological studies (e.g. Sheffield *et al.* 2014; Wanders *et al.* 2015) and has been calibrated globally against observed discharge from the Global Runoff Data Centre (GRDC) to ensure accurate hydrological simulations. Using the daily temperature, VIC is able to resolve the surface energy balance, thereby providing additional valuable information on important variables like the surface temperature and evaporation flux.

To evaluate the performance of the subseasonal forecasts using VIC, we created a baseline where VIC has been forced by observational data from the Princeton Global Forcing dataset. This reference simulation is used both for the verification of the forecasts as well as the initialization of the forecast. To ensure that the hydroclimatology of the forecast and the reference match, the NMME-2 forecast meteorological variables are bias corrected against the PGF. The bias correction adjusts the rainfall intensities and number of rain days based on matching the cumulative density functions (CDF-matching) of the forecasts and the observational data. For daily 2 m air temperature, the time series were bias-corrected against the PGF using CDF matching.

2.4. *Canonical event analysis*

We follow the approach developed by Roundy *et al.* 2015 to perform a canonical event analysis over a range of temporal and spatial scales. Canonical events are specifically defined spatially and temporally aggregated forecasts, with the finest event being the model resolution (1° spatially and daily) and alternative events being aggregated forecasts — for example 3° spatially and seasonally averaged forecasts. Model forecast skill can vary with canonical events. When a seasonal forecast model shows skill over a range of canonical events (i.e. a range of temporal and spatial scales), it shows signs of robust forecast skill and is therefore more

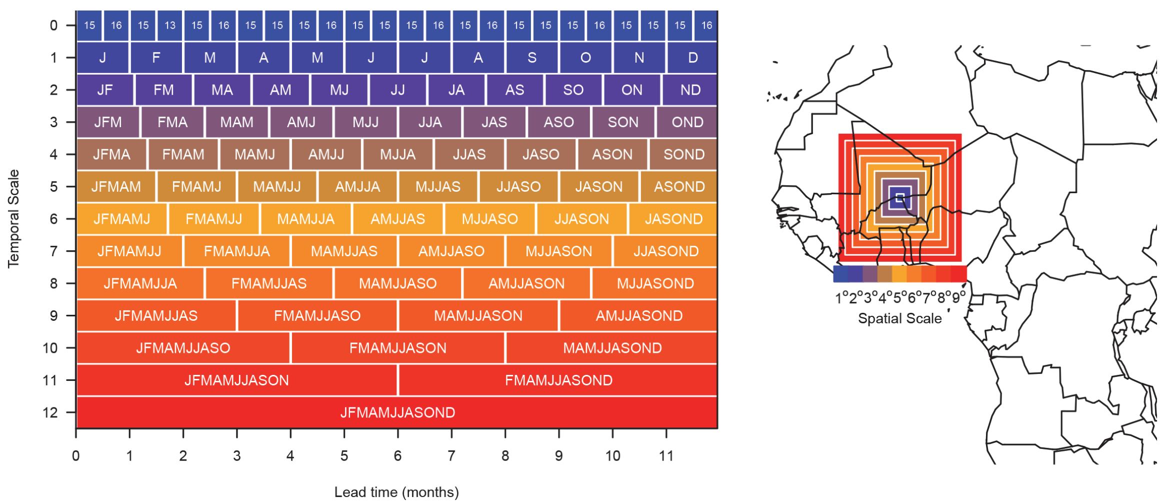

Fig. 1. Temporal (*left*) aggregation setup of a forecast initialized in January, and the spatial aggregation (*right*) of the same January forecast.

reliable when applied in real-life forecasting situations. Forecast skill for canonical events is computed using the Spearman ranked correlation between the seasonal forecast and observed data. When at a location the correlation is statistically significant ($p < 0.05$) and positive ($R > 0.0$) we find that the model has skill for that particular temporal and spatial scale at that location. This statistically significant event is used to calculate the probabilistic predictability metric (PPM) that is given by:

$$\mathrm{PPM} = E_S/E_t \tag{1}$$

where E_s is the number of events that show skill and E_t is the total number of events. Following Eq. (1), it can be shown that a PPM of 0 indicates no skill for any temporal or spatial scale at that location for that initialization month, while a PPM of 1 indicates skill across all scales.

The temporal and spatial scales under consideration in this work are provided in Fig. 1, where we show the scales for a January forecast. We show that the January forecast can be used to obtain a monthly average of a meteorological variable for the coming month. However, it can also be used to get the same monthly average of an 11 month lead period. Similarly for a two monthly temporal aggregation we can obtain the 0 to 10 month lead forecasts, while the 12 month forecast can only be obtained at a lead time of 0 months (Fig. 1). The same January forecast can also be used to produce forecasts for different temporal aggregation periods, ranging from subseasonal two week periods to 12 month periods. In total this provides us with 102 unique temporal timescales for the January forecast and 8 unique spatial scales. This leads to a combined total of 816 unique combinations of temporal and spatial scale. When the model shows a PPM of 1 this would indicate that the correlation is significantly positive for all these 816 events.

2.5. *Detection of start of agricultural and high flow season*

The start of the growing season in the sub-Saharan regions of Africa is largely dominated by large scale annual rain events coming from the South around April. Here, we used the normalized cumulative density function of precipitation to estimate the onset of the growing season. By using a normalized CDF the systematic over- or underestimation of the precipitation by the seasonal forecasts model is accounted for and

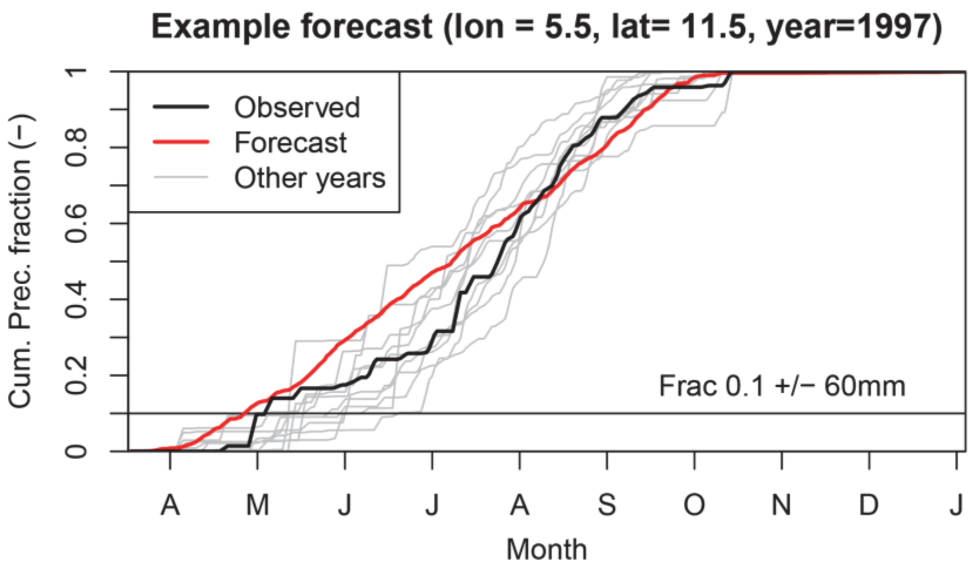

Fig. 2. Detection of the growing season as applied in this study. The normalized cumulative precipitation is shown for a seasonal model forecast from CanCM3 (red), the observed precipitation (black) and the spread over different years. The onset of the growing season is defined as the day that the precipitation exceeds the 0.1 threshold for the first time.

the percentiles can be directly compared to the observations. For all forecasts, the day that the exceedance of the 10^{th} percentile is forecasted is selected as the day that the growing season starts and the absolute difference between this day and the reference (observed day) is calculated (Fig. 2). When the mean absolute difference between the forecasts and observations is smaller than the randomly selected year and the actual year. To assess the skill of the model to forecast the onset of the growing season, we only used the forecast issued on the 1^{st} of March. This forecast was selected based on local information from The Centre Regional de Formation et d'Application en Agrométéorologie et Hydrologie Opérationnelle in Niger (AGRHYMET) that the onset of the growing season starts after the 15^{th} of March and they use a seasonal forecast from the 1^{st} of March in their operational system. Earlier forecasts lack forecast skill so the agricultural planners at AGRHYMET do not require information before the 1^{st} of March. To ensure that the assessment of the model forecasts skill is in agreement with the demands from local end users and farmers, we used the same forecasts as they would in their operational system.

2.6. *Detection of start of agricultural and high flow season*

The start of the growing season in the sub-Saharan regions of Africa is largely dominated by large scale annual rain events coming from the South around April. Here, we used the normalized cumulative density function of precipitation to estimate the onset of the growing season. By using a normalized CDF the systematic over- or underestimation of the precipitation by the seasonal forecasts model is accounted for and the percentiles can be directly compared to the observations. For all forecasts, the day that the exceedance of the 10^{th} percentile is forecasted is selected as the day that the growing season starts and the absolute difference between this day and the reference

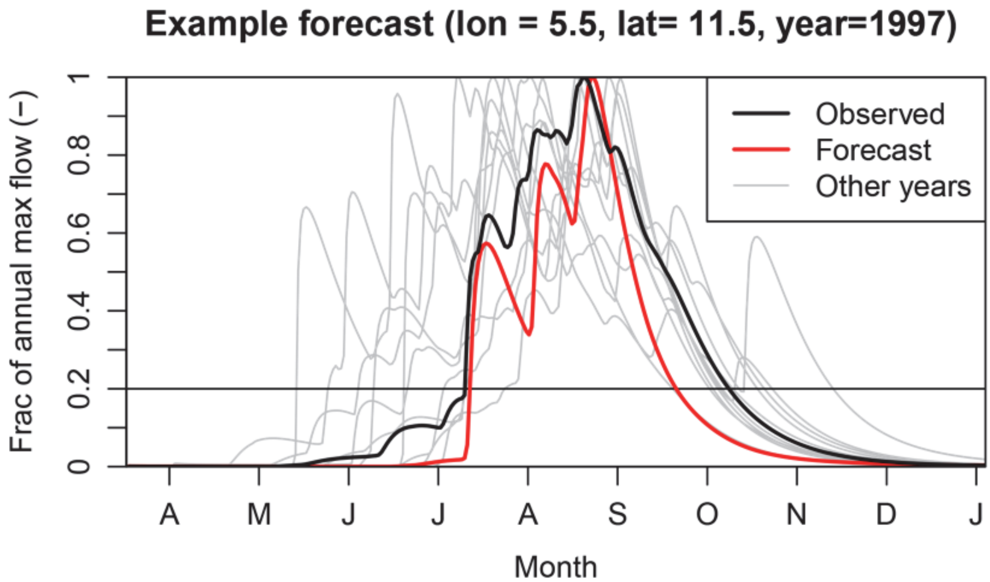

Fig. 3. Detection of the high flow season as applied in this study. The start of the high flow season is defined as the first exceedance of the 20th percentile threshold. The peak of the high flow season is the day that the forecast hits the maximum annual discharge.

(observed day) is calculated (Fig. 2). When the mean absolute difference between the forecasts and observations is smaller than the interannual spread between the observed days, the forecast is deemed skilful. This implies that the model has a significant forecast skill when the differences between model forecast and observations are smaller than the differences between a random selected year and the actual year. To assess the skill of the model to forecast the onset of the growing season, we only used the forecast issued on the 1st of March. This forecast was selected based on local information from The Centre Regional de Formation et d'Application en Agrométéorologie et Hydrologie Opérationnelle in Niger (AGRHYMET) that the onset of the growing season starts after the 15th of March and they use a seasonal forecast from the 1st of March in their operational system. Earlier forecasts lack forecast skill so the agricultural planners at AGRHYMET do not require information before the 1st of March. To ensure that the assessment of the model forecasts skill is in agreement with the demands from local end users and farmers, we used the same forecasts as they would in their operational system.

To detect the start of the high flow season, we used the hydrological model VIC to provide baseflow estimates. We define two metrics to quantify the skill of the model to accurately forecast the start and peak of the high flow season. First, the high flow season is normalized where the peak discharge is used to normalize the discharge (Fig. 3). The start of the high flow season is detected by looking at the first exceedance of the 20th percentile of the annual maximum flow. The mean absolute difference between the forecast and observed start of the high flow season is used as a metric for the model performance. Similar to the detection of the growing season, the offset is compared to the interannual spread between the observed onsets to identify if the model has a lower signal to noise ratio than would be obtained with using historical information. The start of the high flow season is important to determine for end-user applications that depend on the presence of open water (e.g., irrigation), while the peak flow is important for the replenishment of water resources for the upcoming dry season in this region (e.g., reservoir operations).

3. Results

3.1. *Meteorological forecast skill over Africa*

To analyze the skill of each seasonal forecast model, we first computed the Spearman ranked correlation for all forecast months separately. The obtained correlations have been used to calculate the PPM (Eq. (1)) for all initialization months, and all temporal and spatial scales. Significant positive correlations have been identified as events that show skill (E_s). The anomaly correlation for precipitation in the first week is high for the CanCM and CCSM models (Fig. 4). The forecast skill fades quickly after the first weeks, leading to low or negligible skill after a one month lead. In general, the precipitation is difficult to predict due to its high natural variability — something that is confirmed for the subseasonal forecasts. The forecast skill of FLOR is additionally hampered by the poor initial conditions that result in a poor overall forecast skill of the subseasonal precipitation. The FLOR model uses the AMIP climatological runs to initialize the land surface states in the model (Gates *et al.* 1998), which limits the forecasts skill for the first weeks. The forecast skill in these weeks largely dependent on the initial conditions of the land surface and are therefore, strongly hampered by the initialization procedure currently used by FLOR, hence the lower forecast skill.

In general, higher anomaly correlations are found for temperature forecasts compared to precipitation forecasts (Fig. 5). The decrease in the predictability metric as a function of increased lead time is not as strong as found for precipitation forecasts. The predictability over

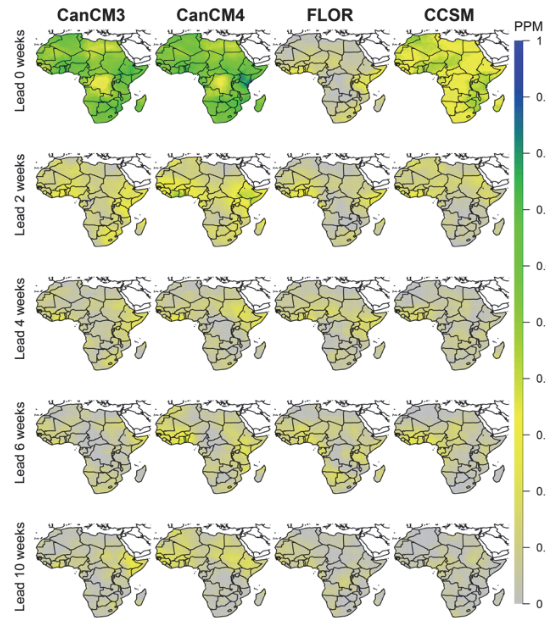

Fig. 4. Skill in precipitation forecasts for different leads and subseasonal forecast models from the NMME-2 ensemble. A PPM of 1 indicates all forecasts are skilful, for the given lead time and location.

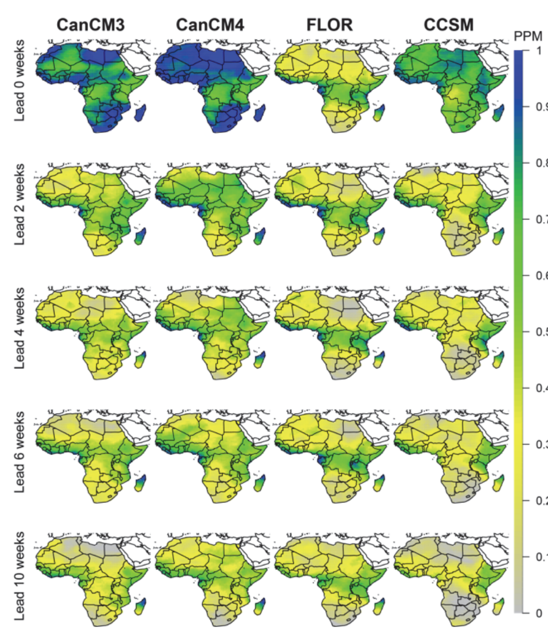

Fig. 5. Skill in temperature forecasts for different leads and subseasonal forecast models from the NMME-2 ensemble. A PPM of 1 indicates all forecasts are skilful, for the given lead time and location.

Africa (especially Equatorial Africa) is very persistent and high, with around 80–90% (PPM between 0.8 and 0.9) of the forecast scenarios showing significant skill. In contrast to the precipitation forecast, significant skill can be found beyond 3 month lead time. In addition, we see that the FLOR model outperforms the CanCM models for the longer lead times. The model still suffers from the AMIP climatological initialization method, however, the skill for the longer leads is in general higher than can be found for the other models.

For all seasonal forecast models, the seasonal precipitation forecast skill reduces with increased lead times (Fig. 6). Lead times beyond a 2 month lead time have no, to negligible prediction skill regarding precipitation anomalies at small temporal aggregation periods. This feature is not unique for Africa. Larger parts of the world are characterized by a low predictability, which quickly disappears with increasing lead time of one or two months. In general, a lower prediction skill for the higher resolution temporal anomalies is found in CanCM and CCSM models, while FLOR shows a high skill at the longer leads at longer temporal scales.

3.2. *Forecast crop growing season*

The growing season in Sub-Saharan Africa is limited by the availability of rain water that falls in the wet season between April and October because of the predominance of rain-fed agriculture. For an optimal yield, it is important that the planting of the fields be done just with the onset of the seasonal rains. If planting is too early, then the seedlings may suffer drought before the real rain season onset, planted too late and yields tend to be significantly lower. We looked at the skill of the seasonal forecast models to reproduce the onset of the rain season

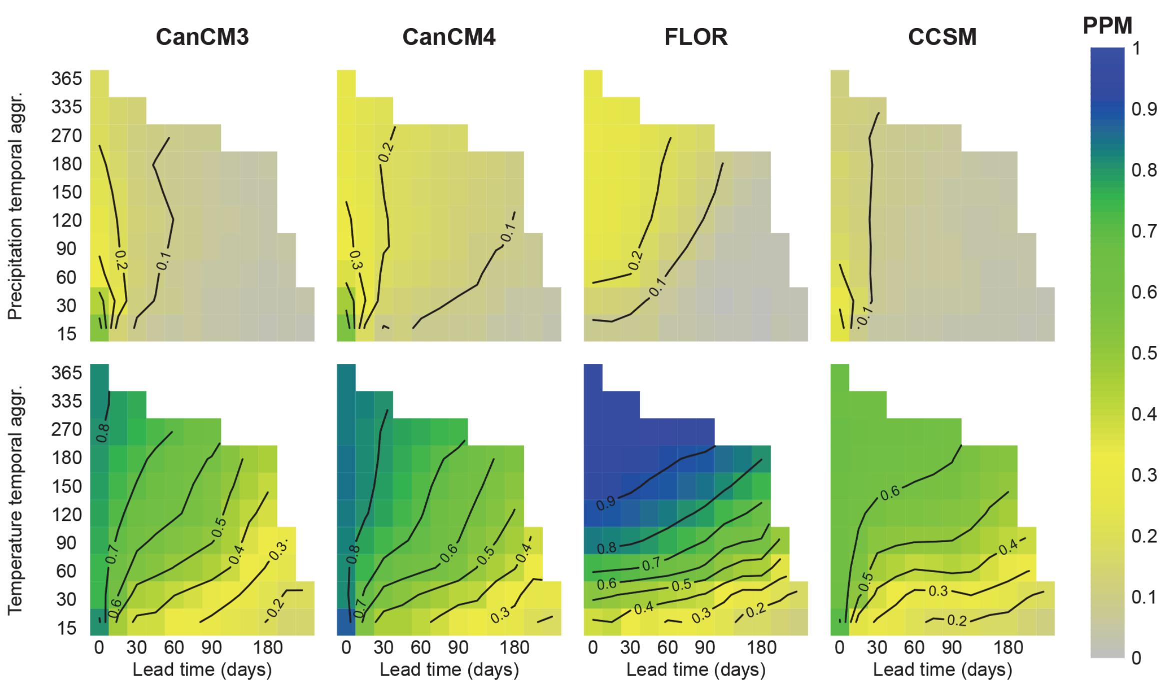

Fig. 6. Skill of African subseasonal forecast across different temporal scales and lead times. A PPM of 1 indicates all forecasts are skilful in the given combination of temporal scale and lead time.

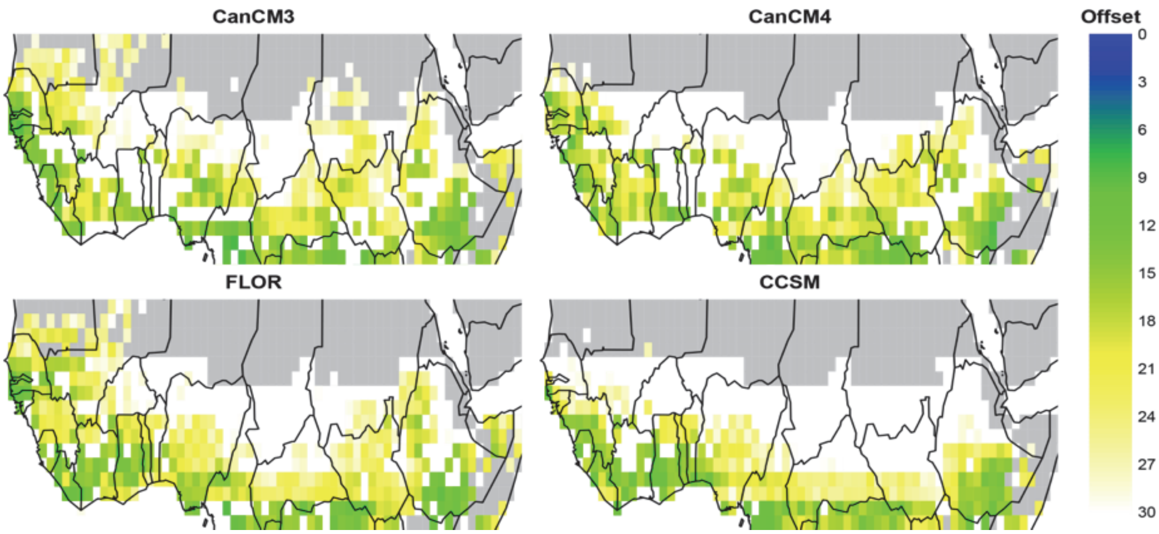

Fig. 7. Forecasts skill for the start of the growing season. Grey regions indicate where the signal to noise ratio is higher than the interannual spread in the start of the growing season. White areas indicate an offset of 30 days or more.

with the forecast issued on the 1st of March, roughly 8 weeks ahead (depending on the geographical location) of the start of the growing season (Fig. 7). We found that the forecast skill of the model is high for the coastal regions of West Africa, where the precipitation onset is early in the year and more abundant. The offset in the forecast increases in the Sahel, where rain is less predictable and the spread in the observations is larger. The models all show skill in the wetter parts of Chad, Niger and Mali, while the northern parts of these countries (with an annual precipitation of less than 200 mm) show a low predictability of the rain onset, partly due to the higher interannual variability. Small differences are found between models, indicating that the choice of model is of less importance.

3.3. *Forecast high flow season*

Accurate forecast of the start of the high flow season could help in water resources management at the end of a dry season. When the onset of the high flow season is accurately forecasted one could optimize the utilisation of the remaining water in reservoirs, lakes and the groundwater. Here we show that the forecasting of the start of the high flow season can be done with an accuracy of around 1 month (Fig. 8). The observed spread and onset of the start of the high flow season is larger than was observed for the start of the growing season (e.g. comparing Fig. 2 and Fig. 3). This wider initial spread makes it more difficult to produce accurate forecasts of the exact onset date. Additionally, the onset of the high flow season is later than that of the growing season. As a result, the 1st of March forecast has to accurately forecast the start of the high flow season 2–4 months ahead. It is shown that large discrepancies exist between the models, where CCSM lacks the required skill to provide forecasts with a higher accuracy than a 60 days offset. Of the remaining models, CanCM4 and FLOR show the highest skill in most regions and have similar spatial patterns. For some regions in the Sahara desert, an offset of zero days is found; this is caused by the

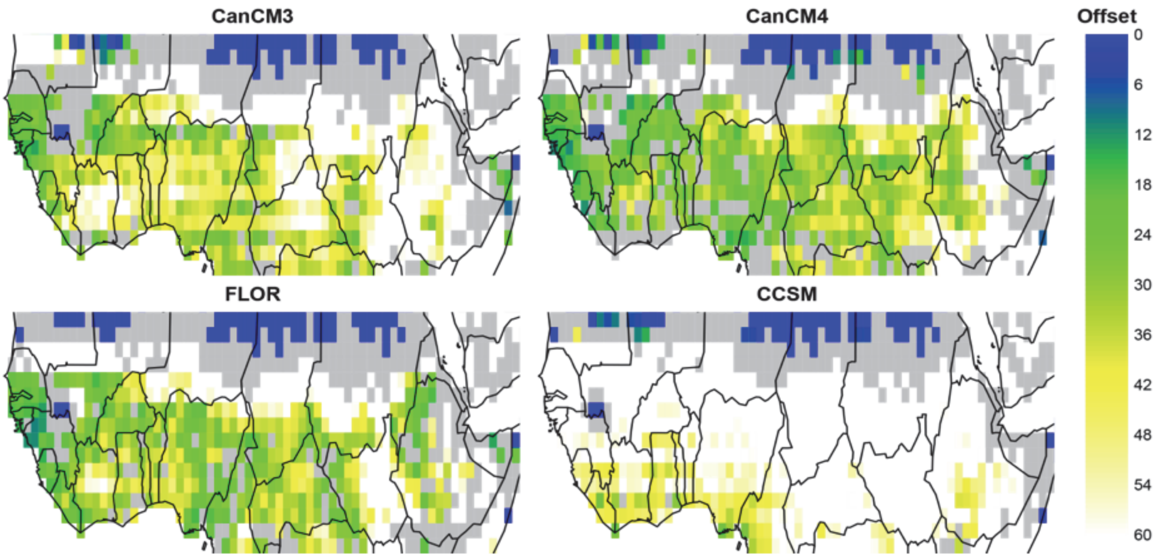

Fig. 8. Forecast skill for the start of the high flow season. Grey regions indicate where the signal to noise ratio is higher than the interannual spread in the start of the high flow season. White areas indicate an average offset of 60 days or more.

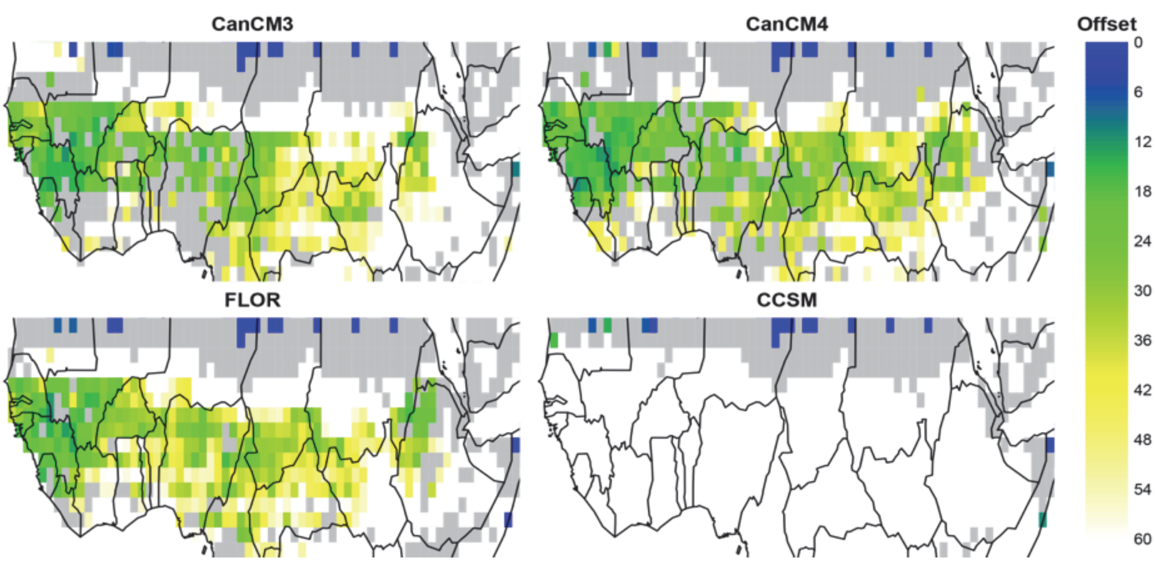

Fig. 9. Forecast skill for the peak of the high flow season. Grey regions indicate where the signal to noise ratio is higher than the interannual spread in the peak of the high flow season. White areas indicate an average offset of 60 days or more.

complete absence of river discharge in the model forecast and observations.

In addition to the start of the high flow season, it is important for water resource management to have forecast knowledge of the peak and end of the high flow season. We show that the ability of the models to accurately forecast the peak of the high flow season is almost identical to their skill in forecasting the start of the high flow season. In general the peak discharge is found 4–6 months after the 1st of March forecast initialization and hence is more difficult to accurately forecast. Nonetheless, the models show a similar offset in the forecast of the peak flow timing as they do for the onset of the high flow season. The highest skill is found in the transition regions between the wet and dry climatic regions of West Africa. These regions also have a very distinct peak flow season, which reduces the false detection of this event in the model forecasts. A low skill is found for CCSM, which is hampered by lower data availability, leading to more uncertainty in the detection skill which results in a low forecast skill. The other models again show similar patterns in terms of forecast skill and the forecasts offset compared to the observations.

4. Discussion

4.1. *Canonical event analysis*

Most assessments of model forecast skill are done based on other metrics than with the canonical event analysis combined with the Probabilistic Predictability Metric (PPM). Anomaly correlations or Brier scores are methods often used to evaluate seasonal forecast. The downside of these methods is that they either provide information at set initialization months, for set lead times (e.g., using correlations) or require the definition of specific events (e.g., exceedance over threshold, Brier score). This is not a problem when scores are required for a fixed forecast scenario (e.g. April rain, forecasted in January), however, it does not inform us on the general performance of the model. Therefore, we argue that unless these clear goals exist, it is beneficial to use the PPM as a score that informs the end user on the average model performance at a specific location or a decomposition of the performance as a function of temporal aggregation and lead time.

A question that remains is how the obtained forecasts skills in terms of PPM could be translated towards applications. From the description of the PPM, it can be derived that a PPM of 0.1 indicates significantly skilful forecast in 10% of the scenarios. This information may be more intuitive to end users than an anomaly correlation for a forecast. However, it does not inform the user on the actual strength of the skill in these 10% — this could be anything between a correlation of 0.35 (lower significance boundary) to 0.75 (highest correlation found). On the other hand, it has the advantage that it provides a linear skill metric that can be used to determine the average skill of the model over a specific location or season in a more comprehensive way.

From the PPM analysis (e.g., Fig. 6), it is shown that different forecast models should be used for different purposes. The CanCM and CCSM models clearly provide the best subseasonal short range, high temporal resolution forecast that can inform end users on events that are happening within the coming month. When more extensive planning is required (e.g. water resource management), FLOR seems to do a better job at predicting the long-term precipitation and temperature anomalies. The improved forecast skill of FLOR with longer leads could be strongly related to the ability to have a more accurate forecast of global sea surface temperatures (SST). Strong teleconnections have been found between SST anomalies and observed rainfall variability over the African continent (Rowell 2013). Strong teleconnections are found for the El Nino, Atlantic and Indian Ocean Dipole indices. Currently, FLOR is the only model in this ensemble that doesn't have advanced land initialization scheme embedded in their forecast system. This suggests that a potential improvement in the first months after forecast initialization could be expected when a more advanced land initialization is implemented. This is currently a topic that is under study and but no definitive results have been obtained so far.

The results with regard to differences in forecast skill in both short-term and long-term forecast for different models can help to improve the quality of decision support system and their role in providing valuable forecasts for climate services. Additionally, the PPM analysis can inform end users when not to use seasonal forecasts for their decision-making process. This prevents high-impact measures to be taken based on overall poor model forecasts quality or a poor forecast quality for the lead time and temporal scale of interest.

4.2. *Growing season*

To accurately detect the start of the growing season, multiple factors need to be taken into account. First of all, the start of the growing season is largely determined by the local climatology and whether the start of the growing season is limited by the amount of available energy (e.g., large parts of Europe) or the available water for crop growth (e.g., West Africa). For large parts of Africa the start of the growing season is largely dominated by precipitation patterns and amounts, and therefore has a more uncertain start than can be found for energy limited regions that are heavily dependent on incoming solar radiation. The added value of having accurate forecasts for the start of the rain season is significant since they dominate much of the agricultural activity in West Africa and the Horn of Africa (Barbé *et al.* 2002). These are regions that can clearly benefit from accurate subseasonal forecasts to optimize the potential crop yield.

In general, the detection of the growing season is hampered by multiple factors. Firstly, it is important to match the modelled start of the growing season with the perception of the farmers in these regions. Due to an existing collaboration with AGRHYMET, we could determine the guidelines used for the onset of the growing season in these regions in Africa. In general, AGRYHMET applies a threshold of 3 days of consecutive rain exceeding a total of 20 mm of accumulated precipitation. Since the seasonal forecast models are known to have too much drizzle (low rainfall intensities) in their forecast we transformed this 20 mm threshold to the 10^{th} percentile of the annual rainfall. Using the observational dataset we validated this threshold and found that these two criteria are mostly fulfilled in a time period of one week. This ensured that we have a more objective threshold that could be applied for a large region. The second problem for detecting the growing season is related to the different crops that are planted in different regions. Every crop has specific demands and therefore depends on different planting dates with relation to the rainfall. This information is not available with the forecast and is therefore difficult to implement in the growing season detection methodology. We assumed here that the methodology supplied by AGHRYMET would be the best benchmark for the start of the growing season.

From the results it is clear that the detection of the growing season clearly benefits from the daily temporal resolution of the NMME-2 forecasts. By using daily resolution forecasts, we can now actually address these more detailed questions related to local climate services. It is shown that we can use these models to forecast the onset of the growing season roughly two months in advance. This could help local farmers and authorities to prepare for the coming growing season and make sure resources are in place in time. Additionally, when a very late onset of the growing season is forecasted, farmers can switch to crops that have lower water demand or produce their yield in a shorter growing season (Brumbelow and Georgakakos 2001).

This is only possible due to the higher temporal resolution of the data. This leads to a major improvement in the usability of seasonal forecasts in agricultural planning.

4.3. *High flow season*

The high flow season in most parts of Africa is highly seasonal and strongly linked with the precipitation season (e.g. Fig. 3). In West Africa this very strong seasonality creates a high need for water resource planning and management, since available resources are low during large parts of the year. The added value of subseasonal forecasts could be substantial if the duration and magnitude of the flow season can be accurately forecasted. The timing of the peak discharge is important for water resources management, where the peak flow determines potential to replenish reservoirs after long dry periods, the timing of the peak discharge is important. A large number of reservoirs depends on a distinct wet season to refill their storage for the upcoming drier periods. Additionally, the remaining water in the reservoirs can be optimally utilized when the onset of the high flow season and peak discharge is known. We show here that the seasonal forecast models are capable of accurately forecasting the timing of the annual discharge peak in West Africa and could therefore, be a valuable tool for water resource planning in early spring.

A valuable next step would be to perform a more detailed assessment of the added value of subseasonal forecasts for hydrological modelling and end users that require detailed hydrological subseasonal forecasts. These results provide an initial result on the potential of the new subseasonal forecast models for application-based climate services.

5. Conclusions

In this study, we show that subseasonal forecasting from the NMME-2 ensembles can substantially improve the climate services that depend on high temporal resolution forecasts. We show that the forecast skill for both precipitation and 2 m air temperature is significantly skilful for many regions in Africa. Most notably, there is a high skill for both subseasonal precipitation and temperature over West Africa and the Horn of Africa. In both regions, we find relatively high population densities in combination with highly seasonal rainfall and a high dependency of the agriculture on these seasonal rainfall events (Barbé *et al.* 2002). The skill in the precipitation forecast fades quickly with increased forecast lead, while the forecast skill for temperature anomalies remains high for lead times beyond 3 months. It is shown that the skill of the models varies across the temporal aggregation window and lead times. For example, the FLOR model shows a relatively low skill on the subseasonal scale for short leads. However, this model shows a stronger forecast skill for longer lead times and larger temporal aggregation windows. This indicates that models should be used accordingly for applications that rely on forecast at a longer lead or larger temporal resolution.

To study the impact of the NMME-2 ensemble for application driven scientific questions, we looked at two important end-user applications in West Africa. This region is of particular interest since we have information on the end-user requirements. The weather patterns have a large impact on the agricultural system and the seasonal forecast models show significant skill in this region. Results show that the models have significant skill in forecasting the onset of the annual rain season and thereby the start of the agricultural growing season. This is a promising result, where the increased temporal resolution of the seasonal climate models, from a monthly to a daily temporal resolution, helps meet end user requirements.

In addition to the detection of the growing season, we analyzed the skill of the NMME-2 ensemble to forecast important hydrological events. First, we looked at the predictability of the onset of the high flow season. This is important moment for agriculture and water resource management, since resources are needed for irrigation and replenishing the lost water resources over the dry winter period. In West Africa there is a clear seasonal pattern, going from none to negligible river discharge to a high flow season 1 to 2 months after the first rainfall events. The models have a significant forecast skill to predict the onset of the high flow season for most parts of West Africa. The average onset is within the sub-monthly domain, indicating that the there is an additional gain by using the subseasonal forecast. Similar results were found for the onset of the peak flow season, where the uncertainty in the forecast is often lower than 30 days for many regions in West Africa. This indicates that local stakeholders will benefit from the NMME-2 ensemble for their decision support systems.

Overall, it is concluded that with the introduction of NMME-2, progress has been made going from seasonal to subseasonal forecasting. We argue that additional progress can be made, by extending the existing the ensemble with more models and by looking more in-depth at specific climate services.

Acknowledgments

This research was supported by the NOAA Climate Program Office under grant NA15OAR431 0075 (Assessing Phase 2 NMME Forecasts for Improved Predictions of Drought and Water Management) to EFW, and NW was supported by a NWO Rubicon Fellowship 825.15.003. (Forecasting to Reduce Socio-Economic Effects of Droughts).

References

Brumbelow, K. and A. Georgakakos, 2001: Agricultural planning and irrigation management: The need for decision support. *The Climate Report*, **1**(4), 2–6.

Dutra, E., W. Pozzi, F. Wetterhall, F. Di Giuseppe, L. Magnusson, G. Naumann, P. Barbosa, J. Vogt, and F. Pappenberger, 2014: Global meteorological drought — Part 2: Seasonal forecasts. *Hydrol. Earth Syst. Sci. Discuss.*, **11**, 919–944. doi:10.5194/hessd-11-919-2014.

Fan, Y. and H. van den Dool, 2011: Bias correction and forecast skill of NCEP GFS ensemble week-1 and week-2 precipitation, 2-m surface air temperature, and soil moisture forecasts. *Weather Forecast.*, **26**, 355–370. doi:10.1175/WAF-D-10-05028.1.

Gates, W. L., J. Boyle, C. Covey, C. Dease, C. Doutriaux, R. Drach, M. Fiorino, P. Gleckler, J. Hnilo, S. Marlais, T. Phillips, G. Potter, B. Santer, K. Sperber, K. Taylor, and D. Williams, 1998: An overview of the results of the atmospheric model intercomparison project (AMIP I). *Bull. Am. Meteorol. Soc.*, **73**, 1962–1970.

Howitt, R., J. Medelln-Azuara, D. MacEwan, J. Lund, and D. Sumner, 2014: Economic analysis of the 2014 drought for California agriculture, Report of UC Davis, Center for Watershed Sciences.

Hurrell, J. W., M. M. Holland, P. R. Gent, S. Ghan, J. E. Kay, P. J. Kushner, J.-F. Lamarque, W. G. Large, D. Lawrence, K. Lindsay, W. H. Lipscomb, M. C. Long, N. Mahowald, D. R. Marsh, R. B. Neale, P. Rasch, S. Vavrus, M. Vertenstein, D. Bader, W. D. Collins, J. J. Hack, J. Kiehl, and S. Marshall, 2013: The community earth system model: A framework for collaborative research. *Bull. Am. Meteorol. Soc.*, **94**, 1339–1360. doi:10.1175/BAMS-D-12-00121.1.

Kirtman, B. P., Dughong Min, Johnna M. Infanti, James L. Kinter, III, Daniel A. Paolino, Qin Zhang, Huug van den Dool, Suranjana Saha, Malaquias Pena Mendez, Emily Becker, Peitao Peng, Patrick Tripp, Jin Huang, David G. DeWitt, Michael K. Tippett, Anthony G. Barnston, Shuhua Li, Anthony Rosati, Siegfried D. Schubert, Michele Rienecker, Max Suarez, Zhao E. Li, Jelena Marshak, Young-Kwon Lim, Joseph Tribbia, Kathleen Pegion, William J. Merryfield, Bertrand Denis, and Eric F. Wood, 2014: The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Am. Meteorol. Soc.*, **95**, 585–601. doi:10.1175/BAMS-D-12-00050.1.

Kundzewicz, Z. W. and Z. Kaczmarek, 2000: Coping with hydrological extremes. *Water International*, **25**(1), 66–75. doi:10.1080/02508060008686798.

Luc Le Barbé, Thierry Lebel, and Dominique Tapsoba, 2002: Rainfall variability in West Africa during the years 1950–90. *J. Climate*, **15**, 187–202. doi:http://dx.doi.org/10.1175/1520-0442(2002)015<0187:RVIWAD>2.0.CO;2.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res.*, **99**(D7), 14415–14428.

Liang, X., E. F. Wood, and D. P. Lettenmaier, 1996: Surface soil moisture parameterization of the VIC-2 L model: Evaluation and modification. *Global and Planetary Change*, **13**(14), pp. 195–206.

Magnusson, L. and E. Källén, 2013. Factors influencing skill improvements in the ECMWF forecasting system. *Mon. Weather Rev.*, **141**, 3142–3153. doi:10.1175/MWR-D-12-00318.1.

Merryfield, William J., Woo-Sung Lee, George J. Boer, Viatcheslav V. Kharin, John F. Scinocca, Gregory M. Flato, R. S. Ajayamohan, John C. Fyfe, Youmin Tang, and Saroja Polavarapu, 2013: The Canadian seasonal to interannual prediction system, Part I: Models and Initialization. *Mon. Weather Rev.*, **141**, 2910–2945.

Roundy, Joshua K., Xing Yuan, John Schaake, and Eric F. Wood, 2015: A framework for diagnosing seasonal prediction through canonical event analysis. *Mon. Weather Rev.*, **143**, 2404–2418. doi: http://dx.doi.org/10.1175/MWR-D-14-00190.1.

Rowell, D. P., 2013: Simulating SST Teleconnections to Africa: What is the state of the art?. *J. Climate*, **26**, 5397–5418. doi:10.1175/JCLI-D-12-00761.1.

Sheffield, J., G. Goteti, and E. F. Wood, 2006: Development of a 50-yr high-resolution global dataset of meteorological forcings for land surface modeling. *J. Climate*, **19**(13), 3088–3111.

Sheffield, J., E. F. Wood, N. Chaney, K. Guan, S. Sadri, X. Yuan, L. Olang, A. Amani, A. Ali, S. Demuth, and L. Ogallo, 2014: A drought monitoring and forecasting system for sub-Sahara African water resources and food security. *Bull. Am. Meteorol. Soc.*, **95**(6), 861–882.

Sida, L., B. Gray, and E. Asmare, 2012: Real-time evaluation of the humanitarian response to the horn of African drought crises. Tech. report, Inter-Agency Standing Committee.

United Nations, 2011: Humanitarian requirements for the horn of Africa drought 2011. Tech. report, Office for the Coordination of Humanitarian Affairs (OCHA), New York and Geneva.

Vecchi, G. A., T. Delworth, R. Gudgel, S. Kapnick, A. Rosati, A. T. Wittenberg, F. Zeng, W. Anderson, V. Balaji, K. Dixon, L. Jia, H.-S. Kim, L. Krishnamurthy, R. Msadek, W. F. Stern, S. D. Underwood, G. Villarini, X. Yang, and S. Zhang, 2014: On the seasonal forecasting of regional tropical cyclone activity. *J. Climate*, **27**, 7994–8016. doi:10.1175/JCLI-D-14-00158.1.

Wanders, N., M. Pan, and E. F. Wood, 2015: Correction of real-time satellite precipitation with multi-sensor satellite observations of land surface variables. *Remote Sens. Environ.*, **160**, 206–221. http://doi.org/http://doi.org/10.1016/j.rse.2015.01.016.

Yuan, X., E. F. Wood, N. W. Chaney, J. Sheffield, J. Kam, M. Liang, and K. Guan, 2013: Probabilistic seasonal forecasting of African drought by dynamical models. *J. Hydrometeorol.*, **14**, 1706–1720.