Direct and indirect seasonal rainfall forecasts for East Africa using global dynamical models

Andrew W. Colman1 | Richard J. Graham1 | Michael K. Davey2

1Met Office, Exeter, UK
2Department of Applied Mathematics, University of Cambridge, Cambridge, UK

Abstract
Regional-scale seasonal climate outlooks are typically produced using forecast information either local to the region or from another area with teleconnections to the region. Dynamical global long-range forecast (LRF) systems can provide both types of information, and these two approaches are compared in the context of seasonal rainfall forecasts for two adjoining areas in the Greater Horn of Africa region in tropical East Africa. The direct method utilizes the unprocessed LRF outputs for the region. For the “indirect” method, canonical correlation analysis is used and works by identifying patterns in historical LRF predictions over large tropical domains that relate to observed variability in the East Africa regions. For a given year, projections of the LRF forecasts onto those patterns can then be employed to construct the regional forecasts. This approach takes advantage of the tendency for LRF systems to have greater skill for large-scale variability than for smaller regional-scale features. Using case studies, it is found the two approaches contain complementary information: the indirect approach can provide notable skill benefits in years with strong large-scale forcing while in some years, particularly when large-scale signals are weak, the “direct” forecast are superior. Skill comparisons over many years found that, although results are region/season dependent, in general the indirect approach has higher skill overall—with improvements equal to or greater than those afforded by a 1-month reduction in lead time with the direct approach. Results for both methods used separately and in combination are provided for the March–April–May and September-to-December seasons in the two regions, using data from currently operational dynamical LRF systems. Skill was best for the September-to-December season in the southern region, using “indirect” forecasts. The “direct” approach was better than “indirect” for the March-to-May season in the northern region. In general, combination did not produce a substantial benefit.

KEYWORDS
canonical correlation analysis, dynamical long-range forecast systems, Greater Horn of Africa, seasonal rainfall prediction
1 | INTRODUCTION

The Greater Horn of Africa (GHA) is prone to recurrent drought and flood events often with devastating impacts on local societies. To facilitate preparedness, seasonal forecasting of rainfall is a key activity for National Meteorological Services in the region, and is supported by the GHA Regional Climate Outlook Forum (GHACOF) which includes contributions from global producers of seasonal forecasts (see http://rcc.icpac.net/index.php/long-range-forecast/regional-consensus-climate-outlook-statement).

Historically, statistical methods have been used for seasonal forecasting, often based on the association of seasonal precipitation with pre-season climate indicators such as sea surface temperature (SST) anomalies (e.g., Mutai et al., 1998; Mwale et al., 2004 for East Africa). More recently, physically based dynamical models of the global ocean and atmosphere are playing an increasing role in the production of seasonal outlooks (e.g., Graham et al., 2011). The output from such models in the region of interest can be used directly to produce a forecast. Methods employed include objective averaging of outputs from several models or, as has been typical of most forecasts generated as part of the GHACOF activity, through use of consensus-based expert judgement to blend model outputs with those of statistical methods. Walker et al., 2019 have shown for the GHA that objective use of dynamical model forecasts (in their case from a single model) can provide better prediction skill than the expert judgement approach and also avoid undesirable biases arising from subjectivity associated with the consensus-based forecasts. However, dynamical model forecasts are also subject to systematic errors such as local biases and distortions in teleconnected responses to, for example, El Niño–Southern Oscillation (ENSO) events, which can inhibit the prediction performance of the direct method. The global forecasts also contain data for climate features that are more predictable than local features, such as ENSO events, and statistical methods can be used to link large scale features to the local region of interest, to make what we shall refer to as “indirect” forecasts. A key objective of this paper is to describe and compare direct and indirect seasonal rainfall forecasts in the GHA from various dynamical systems, and consider any benefit of combining such forecasts. A probabilistic forecast format for three categories is used, with assessments concentrating on the “Wet” and “Dry” categories.

A detailed review of the climate of the GHA region, including seasonal forecasting aspects, can be found in Nicholson (2017). For this study, the GHA region is divided into two parts, GHA South (GHAS) and GHA North (GHAN) as indicated in Figure 1. GHAS covers the region 5°N–12°S, 28°E–42°E and includes the countries Kenya, Tanzania, Uganda, Rwanda and Burundi, where there are two distinct wet seasons: the “long rains” in March–April–May (MAM), and the “short rains” in September–October–November–December (SOND). GHAN covers the region 25°–5°N, 20°–53°E. GHAN also receives rain in MAM and SOND, although the wettest time of year for much of GHAN is the June–September period when the region is affected by the eastern reaches of the Sahel rainy season. Predictions for GHAS and GHAN are investigated, as these two regions are equivalent to the area considered in GHACOF outlooks for MAM and SOND rainfall.

In Section 2, the overall experiment configurations are summarized and the model and observational data are discussed. The methodologies for calculating direct and indirect forecasts are described in Section 3. Canonical correlation analysis (CCA) is employed in the construction of forecasts by the indirect method, and an example for the SOND season in the GHAS region is provided.

Using hindcast data from a relatively long period (1960–2005) from two dynamical systems, the performance of direct and indirect forecasts is compared in Section 4, for the MAM season in GHAS. In individual years the two methods can provide contrasting predictions, and some particular cases are discussed. The effect of combining the direct and indirect forecasts is described.

Hindcast data from recent operational dynamical forecast systems are available for a shorter period. In Section 5 a systematic assessment of the forecast skill for the MAM and SOND seasons in GHAS and GHAN is described, using data for the period 1981–2017 from four such forecast systems. A comparison with the consensus forecasts issued by GHACOF is included. Summary and conclusions are in the final section.
2 | OVERVIEW OF EXPERIMENT CONFIGURATIONS AND THE DYNAMICAL MODEL AND OBSERVATIONAL DATA USED

2.1 | Experiment configurations

Three example sets of predictions, each with a different “indirect” configuration, are investigated. The first two examples (Sections 3.3 and 4) illuminate the mechanisms by which constructed forecast fields using CCA can improve on direct model output. The third example (Section 5.1) then focuses on evaluating the skill benefits of construction and combination using four currently operational models. Each example uses a different strategy in training and verifying the indirect forecast models, selected to make best use of the hindcast period available—details are provided in Table 1.

Example 1 has 14 years of hindcast data from the GloSea5 version used, which is insufficient to have independent training and verification periods—so these periods are held the same and independence sought by using cross-validation with a 1-year window when training the CCA-based models. Example 2 examines predictions for individual years to illuminate the mechanisms behind successful (i.e., better than the direct method) and unsuccessful CCA based constructions. Hindcasts of 45 years from two model versions from the ENSEMBLES project are used, allowing independent verification and training periods and no need for cross-validation for verification. Although the ENSEMBLES versions of these models have now been superseded, for GHA these versions have similar skill to modern versions and their use here serves to demonstrate the mechanisms by which local forecast errors can be corrected by reference to better-predicted variability in teleconnected regions. In the third example we conduct a skill comparison of the direct and indirect approaches for four currently (2019) operational models with similar hindcast lengths, to estimate skill benefits that calibration can bring to modern-day systems. An approach mirroring a possible operational configuration in which the training and verification periods are independent, and a sequence of sub-periods is used for training as shown in Table 1.

2.2 | Model data

Hindcasts (retrospective forecasts) are typically produced for dynamical seasonal forecast systems for a period of 20 years or more to enable assessment of predictive skill, and also for calibration purposes such as correction of systematic biases. In order to sample a range of outcomes and account for uncertainty in the model formulation and in the initial conditions, both hindcasts and real-time forecasts are produced as ensembles by running the systems many times with slightly differing initial conditions and sometimes different model parameters to reflect the uncertainty in these factors.

| Table 1 | Summary of the indirect (CCA-based) forecast examples produced using dynamical model hindcasts and observational datasets |
|---------|-------------------------------------------------------------------------------------------------|
| Example 1 (Section 3) | ENSEMBLES stream 2: Operational in 2019 |
| Dynamical LRF models | GloSea5 | ECMWF, Meteo-France |
| Predictor domain | 40°N–40°S; 60°W–100°E | 40°N–40°S; 0°–360°E |
| Predictand domain, season, initialisation month(s) | GHAS, SOND, Aug | GHAS, MAM, Feb |
| Training observations | GPCP | PREC/L |
| Training cross-validation | 1 year window | 1 year window |
| Training: Forecast periods | 1996–2009:1996–2009 | 1960–1988:1989–2005 |
| | | 1981–2011:2002–2004 |
| | | 1981–2013:2005–2007 |
| | | 1981–2016:2010–2017 |
| | | 1981–2010:2011–2013 |
| | | 1981–2013:2014–2016 |
| | | 1981–2016:2017 |
| Verification period | 1996–2009 | 1989–2005 |
| Verification cross-validation? | 1 year window | Independent |
| Verification observations | GPCP | PREC/L |

Abbreviations: CCA, canonical correlation analysis; ECMWF, European Centre for Medium-Range Weather Forecasts; GHAN, Greater Horn of Africa North; GHAS, Greater Horn of Africa South; GPCP, Global Precipitation Climatology Project; MAM, March–April–May; SOND, September–October–November–December.
For this study hindcast data from several dynamical long-range forecast systems and centres have been used, as listed in Table 1. For the example in Section 3 illustrating the use of CCA to construct indirect forecasts, the Met Office GloSea5 system (MacLachlan et al., 2015) was selected. For the example in Section 4 comparing direct and indirect forecasts over a long period, Stream 2 data from the ENSEMBLES project (Weisheimer et al., 2009) from the European Centre for Medium-Range Weather Forecasts (ECMWF) and Météo France systems were used. (Data obtained from the EU FP6 ENSEMBLES project archive at http://www.ensembles-eu.org.) For the results in Section 5, four systems operational in 2019 with ~37 years of hindcast data were selected: ECMWF (system5, Stockdale et al., 2018); the Canadian Meteorological Centre (CMC1-CanCM3 and CMC2-CanCM4 (referred to as CMC1 and CMC2); and NOAA CPC CFSv2. The CMC1, CMC2 and CFS systems are part of the North American Multi-Model Ensemble (NMME, Kirtman et al., 2014). ECMWF data were obtained from the ECMWF archive, and NMME data were obtained from the IRI data library. In each case hindcast precipitation data were obtained. Hindcast SST data were also retrieved and used to develop indirect forecasts of rainfall, but as the results were generally less skilful these are not included in this article.

2.3 Observational data

Observational data are required for predictive skill assessment, and also for the construction of indirect forecasts. We make use of two gridded monthly precipitation datasets based on observations. The Global Precipitation Climatology Project (GPCP) dataset (Adler et al., 2012) is based on combined satellite and gauge data, with near global coverage over both ocean and land from 1979 onwards. The satellite information greatly improves spatial and temporal coverage, but relies on the interpretation of remote observing information from various instruments. Version GPCP2.3, with 2.5° resolution in latitude and longitude, was used in Examples 1 and 3.

For the construction of the indirect forecast models in Section 4, requiring data from 1960 to 1988, the NCEP PREC/L dataset (Chen et al., 2002) was selected. This is a gridded analysis of gauge data alone which contains a complete record of half-degree gridded monthly precipitation over land dating back to 1948. Its main weaknesses are reduced station coverage in recent years (since around 1990), thus some gridded values are estimated from what can be quite distant stations but it does not contain the errors that can arise from blending gauge data with satellite soundings. Note that PREC/L is only used in Example 2.

The observational and forecast system datasets, and details of the examples where they are used, are summarized in Table 1.

3 | DIRECT AND INDIRECT METHODOLOGIES

Long-range forecasts are often presented in the form of probabilities for various categories. As is common, we shall calculate probabilities for three categories which are equiprobable over some reference period. The categories are referred to as Above, Near and Below Normal—for precipitation, they are also often called Wet, Average and Dry. The reference values that define the category boundaries are referred to as the upper and lower terciles.

3.1 Direct forecasts

For dynamical systems the data from ensemble members can be used straightforwardly to produce probability forecasts. The probability of an event, such as seasonal rainfall in a “Wet” category, can be estimated as the number of ensemble members predicting that event divided by the total number of ensemble members: we will use this method for direct forecasts. When the category boundaries are defined using hindcast data from the same system this approach includes an element of bias correction. (If the hindcast data period does not match a particular reference period, then observed values for the two periods can be used to adjust the terciles.) We refer here to such ensemble-based probabilities as “direct” forecasts.

3.2 Indirect forecasts—The use of CCA

Indirect forecasts make use of statistical relationships between the quantity to be predicted and some predictor variable(s), determined by analysing predictor and predictand data in a reference (training) period. We use CCA for this purpose. CCA is a procedure that identifies pairs of spatial patterns, in two space–time datasets, that are optimally correlated in time (see, e.g., Glahn, 1968 in a meteorological context, and references therein). For our seasonal forecast application, one is the predictor dataset with ensemble-mean output from dynamical forecast systems (here restricted to large-scale tropical domains), and the other is a predictand dataset of precipitation observations in GHAS or GHAN for the same time period. The predictor and predictand variables and domains do not need to be the same: the aim here is to utilize teleconnection relationships between large-scale tropical variability and regional precipitation (cf. Barnston and Smith, 1996). We chose precipitation as the predictor variable to benefit from the dynamical
ocean–atmosphere systems’ ability to represent the potentially nonlinear relationship between SST related forcing and atmospheric response (Feddersen et al., 1999) and we chose a large (tropicwide) domain as the LRF systems tend to be better at representing seasonal predictability on a large scale (Arribas et al., 2011).

To reduce the size of large datasets, and to address data redundancy arising from high correlation in data at nearby grid points, empirical orthogonal function (EOF) analysis—or equivalently principal component analysis—is applied as a pre-filtering process to represent the predictor and/or predictand fields by a few dominant (and linearly independent) spatial patterns and their associated time series, before applying CCA. In the context of long-range forecasting, this approach was introduced by Barnett and Preisendorfer (1987), wherein a detailed description can be found. The leading EOF and CCA patterns tend to be relatively large scale, and generally a small number of such patterns are selected for use in a forecast model. This pre-filtering approach is utilized here.

The relationships identified using CCA can then be used to construct a statistical forecast model (Barnett and Preisendorfer, 1987; Barnston, 1994). For both the training and forecast steps we use the CCA facility available as part of the Climate Predictability Tool (CPT) developed by the International Research Institute for Climate and Society (IRI) (Source: https://iri.columbia.edu/our-expertise/climate/tools/cpt/).

We made use of the option in CPT that finds the number (NX) of leading EOFs for the predictor, NY for the predictand and NCCA for the number of leading paired CCA modes that optimize the prediction skill according to a user-selected skill metric. Starting with NX = NY = NCCA = 1, a forecast model is constructed using linear regression and the skill metric evaluated, the process is then repeated, incrementing NY, NX and NCCA progressively up to user-selected maxima and evaluating skill for all combinations (with the proviso that, on any iteration, NCCA is less than or equal to min [NX, NY]). We set the maximum number of EOFs and CCA modes to be considered to 10. The values of NX, NY and NCCA that deliver the best performance according to the skill metric are retained. In this study the skill metric used was the average cross-validated Pearson’s correlation over all grid points in the Y domain.

Ideally the skill assessment period is separate from the CCA “training” period, but often data are not available to allow separate multi-decadal training and assessment. In that case, cross-validation is used, by which the CCA training is done over all available years except the year to be predicted (and optionally an even number of years either side). Examples of both approaches are provided in the following sections (see also Table 1).

The CPT package has been widely used for seasonal forecasting: see, for example, Landman et al. (2012), Manzano and Ines (2015), Kipkogei et al. (2017), and references therein. As precipitation data distributions are often not Gaussian and linear methods assume Gaussian data, CPT includes an option to apply a transformation to the Y field such that the temporal distributions are Gaussian prior to the CCA analysis—and this option was used in our CCA operations.

Although CCA-based forecasts are by nature deterministic, the forecasts can be converted to probabilities, usually by assuming a Gaussian distribution centred on the deterministic value, the width of which is calculated from the standard error between hindcasts using the statistical model and corresponding observations in an assessment period. Category probabilities can then be calculated from this distribution, and this procedure is part of the functionality of the CPT package. The net result of the process is a statistical model that uses gridded dynamical forecast model data in the predictor domain as input and produces a gridded categorical probability forecast in the predictand domain.

### 3.3 Example 1: A CCA-based indirect forecast for GHAS in the SOND season

A simple example illustrating indirect forecasts using CCA is presented here, using data from GloSea5 (horizontal resolution ~80 km). The predictor data are the GloSea5 ensemble-mean SOND gridded precipitation for a Tropical Atlantic/Africa/Indian Ocean (TAAI) domain (40°N–40°S, 60°W–100°E) for 1996–2009, from hindcasts initialised in August. The predictand is GPCP precipitation in GHAS for the SOND season. Applying the procedure described in the previous sub-section, with a 1-year cross-validation window and training and assessment periods both 1996–2009, leads to an optimal choice of NX = 4, NY = 4 and NCCA = 2. Most skill is associated with the leading CCA mode in this case.

The leading pair of CCA spatial patterns and the associated time series are illustrated in Figure 2a–c. The X pattern (Figure 2a) shows negative values over tropical East Africa and the west Indian Ocean with a tongue of negative weights stretching further across the Indian Ocean roughly between the equator and 10°N. Two areas of positive weights lie to the north and south of this tongue. Over the Atlantic region there is a weak negative equatorial tongue with positive values to the north and south. The loading pattern over the Indian Ocean resembles the precipitation response to the Indian Ocean Dipole (IOD) mode of SST variability (Saji et al., 1999), known to influence the GHA region in this season (Owiti et al., 2008; Nicholson, 2017).
The corresponding observational Y pattern (Figure 2b) shows negative weights over all of the GHAS region. The corresponding time series (Figure 2c) are very similar indicating a strong relationship between the large-scale predictor and the regional predictand, such that the X pattern as shown in Figure 2a is associated with widespread lower than normal precipitation in GHAS. (An opposite-signed X pattern would likewise be associated with higher than normal GHAS precipitation.) Note that the negative signal in the X pattern near the African East coast only covers the eastern half of GHAS. This discrepancy suggests that the GloSea5 model may be systematically failing to extend the negative precipitation anomalies (associated with the large-scale IOD related X pattern) far enough inland over GHAS.

The GloSea5 August-initialised precipitation forecast anomaly for TAAI for SOND 1998 is shown in Figure 2d. The pattern resembles the X CCA pattern in Figure 2a, and consequently the corresponding Y pattern provides a strong deterministic dry prediction for GHAS (not shown). As described above, this is converted to a probabilistic forecast for each predictand gridbox, and in this case high probabilities result for the “below” tercile category as shown in Figure 2e, exceeding 75% (dark red) over most of the region, particularly south of the equator.

The observed precipitation in SOND 1998 was indeed below average for most of GHAS (not shown). The “direct” probability forecast, based on the GloSea5 precipitation forecast ensemble, favoured dry conditions in the south but wet in the northwest of the GHAS region, and in this case the indirect forecast provided better guidance.

Note that the CCA process often leads to a leading highly correlated pair of time series, as in this example. However this relationship may sometimes only account for a small part of the regional precipitation variance, and CCA pairs with lower correlation may contribute more to forecast skill and thus be selected in the construction of an indirect forecast model.
4 | COMPARISON AND COMBINATION OF DIRECT AND INDIRECT FORECASTS

4.1 | Forecast comparisons

In this section we compare direct and indirect forecasts for the MAM season in GHAS, and describe performance in individual years to gain insight into how the CCA-based method can improve forecast skill (Example 2 in Table 1). In this case hindcast data from a relatively long time period (1960–2005) is used, to accommodate a training period of 1960–1988 and verification for the independent period 1989–2005. (The training and verification periods are each comparable to the length of hindcast datasets typically used in current operational seasonal prediction systems.) Direct predictions for the MAM “long rains” in the GHAS region generally have lower forecast skill than for the SOND season (Rourke, 2011; Mwangi et al., 2014; Nicholson, 2014, 2015) related to weaker associations with ENSO (Ogallo, 1988) and the IOD (Owiti et al., 2008) and so potential to improve skill through the indirect approach is of high interest.

ENSEMBLES hindcast data from the ECMWF (EC) and Météo-France (MF) systems were used, using MAM precipitation hindcasts with February start dates. The predictor domain for indirect forecasts was extended to 40°N–40°S, 0°E to 360°E, to permit potential capture of teleconnection impact from all tropical ocean basins. Optimal NX, NY and NCCA values were determined from the training period data, then gridded probability forecasts were produced for the assessment period. (Note that cross-validation is used in the training period as part of the optimisation process.) For EC, two CCA modes were retained (NX = 3, NY = 2, NCCA = 2), while for MF one mode was optimal (NX = 2, NY = 3, NCCA = 1).

An illustration of the forecast performance in 1989–2005 over the GHAS region as a whole is provided in Figure 3 by comparing the difference between Wet and Dry category probabilities, averaged over the 2.5 × 2.5° grid boxes in the GHAS region, with the fraction of Wet minus the fraction of Dry grid boxes observed in the region in PREC/L data for each year. The EC direct forecasts (Figure 3a, purple) capture the observed year-to-year variability in 1991–1995 and 1998–2000. However, there are major differences centred on 1997 and 2002 where the direct forecast year-to-year variability is in the opposite sense to that observed. Performance in these years results in a low overall temporal correlation of 0.09. In contrast the EC indirect forecasts (blue) correctly capture the direction of the observed year-to-year changes centred on 1997 and 2002, raising the overall correlation to

![Figure 3](image-url)
0.57, which is a substantial improvement. However the year to year variability of the CCA-calibrated values is lower than that of the direct values.

The direct forecasts from the MF system (Figure 3b, purple) are able to capture the observed year-to-year changes better than the EC direct forecasts, with a correlation of 0.40. The direct MF forecast misses the dry 1992 season however, in contrast to the EC direct forecast. The MF indirect forecasts also improve on the MF direct forecasts (note, e.g., the much better representation of the dry signal in 1998) increasing the correlation to 0.70.

On occasions the direct forecast is better than the indirect version: in particular in 1999, when Wet conditions prevailed in the observations and direct forecasts, the indirect version was neutral in EC and MF, reflecting a low signal in the large-scale predictor pattern, as investigated in more detail later.

To explore why performance of the indirect method varies in different years more closely, we first look at differences in precipitation between MAM 1998 (generally dry in GHAS) and 1997 (generally wet in GHAS) over the circumtropical predictor domain used in the CCA-based method, as shown in Figure 4 for the EC and MF hindcasts and GPCP observations. In GPCP observations (Figure 4 lower panel) there are large differences in the tropical Pacific with 1998 wetter than 1997 over a band along and mainly south of the equator in the eastern tropics and drier to the north, south and west of this band. The wetter conditions over the East Pacific in MAM 1998 are related to warm SSTs in this region following the large 1997 El Niño event, which developed after MAM 1997 and was present but weakening in MAM 1998. The second largest coherent area of differences is over the southern Indian Ocean, where 1998 was wetter than 1997 along an east–west band from northern Madagascar to Java with a southward broadening of these differences in the east. A wide horseshoe of drier conditions sits around this area including over southern Africa and the Indian Ocean from the Horn of Africa eastward to southeast Asia. The wet signal in the southern equatorial Indian Ocean is consistent with a general warming of SST in this region following and associated with the 1997 El Niño (Alexander et al., 2002). These large-scale differences are generally well represented in both the EC and MF hindcast difference fields (Figure 4 upper and middle panels respectively).

Returning attention to the target GHAS region and vicinity, it may be seen that in observed differences the band of substantially greater precipitation in 1998 versus 1997 over the Indian Ocean is truncated at the coast of equatorial east Africa with generally drier conditions over the GHAS region forming part of the broad horseshoe shape of negative differences over Africa and the northern and southern Indian Ocean. In contrast both the EC and MF difference fields show a westward extension of the Indian Ocean positive band into equatorial east Africa. This places the erroneous wetter MAM 1998 versus 1997 values in the EC and MF local forecasts, seen in Figure 3, in the context of larger scale differences.

**FIGURE 4** 1998 minus 1997 March–May seasonal precipitation in standardized units. Ensemble mean hindcasts from ECMWF (top panel) and Météo-France (middle). Observed difference in GPCP observations (bottom). ECMWF, European Centre for Medium-Range Weather Forecasts; GPCP, Global Precipitation Climatology Project.
We next relate these patterns to the indirect forecasts. The leading CCA pattern pairs derived from the MF and EC precipitation hindcasts and PRECL observations in the 1960–1989 training period are shown in Figure 5. The leading CCA pair using MAM GPCP observations for the X field over the more recent period 1979–2013 is also provided for comparison (bottom panel). (GPCP does not extend back prior to 1979.) The three sets of CCA pairs are very similar on the large scale. This indicates that MAM GHAS precipitation variability is connected to a well-defined circum-tropical precipitation pattern, and that, conditional on observed GHAS precipitation, this co-variability is well represented in the EC and MF systems.

The X patterns in Figure 5 closely resemble, both in pattern and sign, the large-scale aspects of the 1998–1997 precipitation differences (Figure 4), and given the association of these patterns with negative anomalies over GHAS (Figure 5, Y patterns), hence the EC and MF CCA-calibrated forecasts are drier in 1998, as observed, in contrast to the forecasts based on local data (Figure 3). Specific similarities between the X loadings in Figure 5 and the difference pattern of Figure 4 include positive values along and south of the equator in the tropical Pacific, the horseshoe of dry conditions around these areas, and positive values over the Maritime continent and in the Indian Ocean south of the equator.

The X patterns in Figure 5 are somewhat similar to the correlation patterns between gridded Indo-Pacific rainfall and SST and East African rainfall identified by Shukla et al. (2014) and used to predict (indirectly) MAM equatorial East African rainfall using dynamical hindcast data and a constructed analogue method.

The indirect approach does not always help. In 1999 the EC and MF direct forecasts capture the observed very wet conditions well, and in 2000 the MF direct forecast indicates very dry conditions as observed and EC direct is neutral; however the EC and MF indirect forecasts indicate near-equal chances of wet and dry conditions in both these years (Figure 3). The failure of the indirect forecasts is likely related to the smaller and less spatially coherent large-scale precipitation fields in these years that result in weak

---

**FIGURE 5** Predictor (left) and predictand (right) patterns for the leading CCA mode. The predictor domain is the entire tropics between 40°N and 40°S and the predictand domain is the GHAS region. Top: ECMWF hindcast and PRECL observations (1960–1988); Centre: Meteo-France hindcast and PRECL observations (1960–1988); bottom: GPCP and PRECL observations (1979–2013). CCA, canonical correlation analysis; ECMWF, European Centre for Medium-Range Weather Forecasts; GHAS, Greater Horn of Africa South; GPCP, Global Precipitation Climatology Project.
projections on the predictor X patterns. The precipitation differences between 2000 and 1999, are shown in Figure 6. Particularly over the Pacific, the differences are much weaker than for 1998 versus 1997 (cf. Figure 4).

The example suggests that the circum-tropical X-domain, while giving best overall results in verification calculated over the hindcast period, may not be the best choice in all years. Experiments with CCA using a domain restricted to the Indian Ocean, where 2000 versus 1999 precipitation differences are larger than in the Pacific, gave better results for the EC calibration. It is beyond the scope of the paper to explore use of case dependent input domains, or to investigate the source of predictability in the “raw” model outputs in these years. Rather, we note that drivers of regional precipitation anomalies will not in every year be strongly related to large-scale patterns. In such cases, calibration using the statistics of the large-scale model response over many years may well obscure any dynamical model regional predictability intrinsic to the year in question, as might derive for example from information introduced in initialization.

### 4.2 Combination

These examples indicate that direct and indirect forecasts contain complementary information and thus that forecast skill might be improved by a combination of the two methods. A linear combination approach for deterministic seasonal forecasts was described by Fraedrich and Smith (1989), who obtained a simple formula for the combination that optimized the correlation score. In the context of combining forecasts from different systems there have been many articles since (e.g., Kug et al., 2008 in the context of multi-model ensembles). For probabilistic forecasts the optimal combination is not obvious, so we adopt a pragmatic approach of linearly combining direct and indirect category probabilities in different proportions and establishing empirically the optimum proportion for the region considered.

Combined forecasts were calculated by first making weighted averages of direct and indirect probabilities for each gridbox in the GHSAS domain and then averaging over the domain. Figure 7 is a repeat of Figure 3 with additional results from combined forecasts with indirect: direct category probabilities weighted 50:50 and 80:20.

According to the correlation measure, for EC the 50:50 combination (correlation 0.33) outperforms the direct forecasts (0.09) but not the indirect forecasts (0.57). The 80:20 combination (0.59) has very similar correlation to the indirect forecasts. For the MF forecasts combination has a clearer benefit, with the 80:20 combination yielding a higher correlation (0.78) than either the direct (0.40) or indirect (0.70) or 50:50 combination (also 0.70). The combined forecasts have the advantage of less extreme “misses” at the expense of weaker signals, a typical consequence of combining forecasts. As noted before, 1997 and 1999 were relatively wet seasons. In 1997 the direct EC and MF forecasts had higher

---

**FIGURE 6** As Figure 4, but for March–May 2000 minus 1999 precipitation in standardized units.
probabilities for dry conditions, while the indirect forecasts indicated wet. By contrast, in 1999 the direct forecasts correctly favoured wet while the indirect forecasts were somewhat dry. The combined forecasts produce an intermediate result: the EC and MF 80:20 combinations both weakly indicate wet conditions in 1997 and 1999.

To check if there is any further advantage to be gained by using the EC and MF information together as a multimodel, similar forecasts were produced by first averaging the EC and MF category grid-box probabilities together with equal weight, and then combining the direct and indirect results. Just using EC + MF direct forecasts resulted in a correlation of 0.28, intermediate between the separate EC and MF values. For the indirect forecasts and the 80:20 combination the results were very similar to the better (MF) of the separate EC and MF values.

The correlation measure adopted above is intended as a convenient indicator of relative performance. A more rigorous view of forecast skill is provided by the relative operating characteristic (ROC) skill measure, which has been adopted by World Meteorological Organisation (WMO) as one of the standard measures for seasonal forecast performance. ROC scores can be calculated for categorical probabilistic forecasts, and are based on the rates of occurrences
of events in forecasts and observations: see, for example, Broecker (2012) for details. A ROC score of 50% represents a chance level and 100% represents perfect skill. ROC scores for gridded probabilistic category forecasts were calculated for various direct and indirect combinations for the EC + MF multi-model. The spatial variation of skill across the GHAS region is indicated in Figure 8. Typically, the spatial distribution of skill is quite patchy. For both the Dry and Wet categories, the indirect method enhances skill over the region (compare the number of grid squares with ROC >65%—orange to red shading). For the Dry category the area-average ROC score is 56% (direct), increasing to around 65% for the 80:20 and 100:0 combinations. Similarly for the Wet category the area-average ROC is 56% (direct), increasing to about 63% for 80:20 and 100:0 combinations.

5 | FORECAST SKILL FOR GHAS AND GHAN, MAM AND SOND

5.1 | ROC scores from a multi-model

The strategy of combining direct and CCA-based indirect forecasts is applied here to assess skill for GHAS and GHAN in the MAM and SOND seasons, using multi-model hindcast data from the four currently operational dynamical systems described in section 2 and Table 1 (Example 3 in Table 1).

As in Example 2, precipitation hindcast data in the domain 40 N–40S, 0–360°E were used for the CCA predictor patterns. GPCP data in the GHAS and GHAN regions were used for training and assessment. Direct and indirect hindcasts were produced from each system, then the multi-model hindcasts were calculated by averaging Wet and Dry category probabilities at the grid box level with equal weight for each system.

An assessment period of 2002–2017 was selected. As the hindcast sets are not long, a strategy to produce indirect forecasts with a sequence of independent training periods was adopted, as in Table 1, which reflects the practise of using

![Figure 8](image-url)
training periods up to a recent year to make real-time forecasts. In this case 5-year cross-validation windows were used in each training period. For consistency, tercile boundaries for 1981–2001 were used throughout the sequence. (For the CFS2 model, training periods start in 1982.) To investigate the skill for longer lead times as well, predictions were produced using model hindcast data with start times in each of the 2 months before the target season.

The numbers of EOF and CCA modes (NX, NY, NCCA) used in the CCA-calibrated forecasts for MAM GHAS are listed in Table 2. The numbers vary substantially from case to case, as is typical with this methodology. Often only a single CCA mode is selected (NCCA = 1), However this single CCA mode frequently involves the use of several EOF modes of the predictor (X) and predictand (Y) fields (e.g., CMC2 Feb predictions over 1981–2004 for GHAS uses NX = 9, NY = 6 and NCCA = 1). Forecasts for GHAN and SOND show similar variations in numbers of modes to those displayed in Table 2.

ROC scores for Wet and Dry categories were calculated for each predictand gridbox, then averaged over the region. To analyse the effect of combining direct and indirect forecasts, results are presented with combinations ranging from pure direct (0%) to pure indirect (100%) weighting in 10% increments.

Results for GHAS for the MAM season are provided in Figure 9. Confidence intervals have been evaluated using bootstrapping (Efron and Tibshirani, 1986). Results for start dates in January and in February are indicated. The most obvious feature is that the pure indirect forecasts (ROC between 60 and 65%) are both clearly better than the pure direct forecasts (all below 56%, near chance levels). Combination produces no clear benefit. The skill levels are quite similar to those found in Example 2 (Figure 8), indicating that advances in dynamical systems may not have substantially improved the skill of direct forecasts in this region and season.

Skill for the Wet category is consistently slightly (<5%) above that for the Dry category. The skill for forecasts with January start dates is only slightly (<5%) below that for February starts, indicating that useful forecasts could be issued with the longer lead. Moreover, it is notable that pure indirect forecasts from January are more skillful than pure direct forecasts from February and thus the skill benefit is at least equal to that obtained by a 1-month reduction in lead-time for direct forecasts.

Results for GHAN for the MAM season are likewise provided in Figure 10. Skill is generally lower than that for GHAS in MAM. In contrast to GHAS, for the indirect forecasts skill is low, near or below 55%, while for the direct forecasts skill is generally somewhat higher, up to 61% for

| MODEL | CFS | CMC1 | CMC2 | EC | CFS | CMC1 | CMC2 | EC |
|-------|-----|------|------|----|-----|------|------|----|
| PERIOD | JAN | JAN | JAN | JAN | FEB | FEB | FEB | FEB |
| 81–01 | 4101 | 1051 | 931 | 961 | 821 | 781 | 961 | 1032 |
| 81–04 | 5101 | 391 | 353 | 1052 | 742 | 782 | 961 | 881 |
| 81–07 | 792 | 10107 | 332 | 1071 | 691 | 10102 | 1066 | 591 |
| 81–10 | 883 | 5101 | 342 | 951 | 882 | 851 | 10108 | 651 |
| 81–13 | 961 | 482 | 352 | 942 | 871 | 542 | 1032 | 1071 |
| 81–16 | 963 | 332 | 443 | 863 | 562 | 942 | 1051 | 871 |

Notes: “Period” indicates the training period (see Table 1). The entries indicate values for NX (predictor EOFs, left number), NY (predictand EOFs, middle number) NCCA (CCA modes selected, right number).

Abbreviations: CCA, canonical correlation analysis; EOF, empirical orthogonal function; GHAS, Greater Horn of Africa South.

Skill for the Wet category is consistently slightly (<5%) above that for the Dry category. The skill for forecasts with January start dates is only slightly (<5%) below that for February starts, indicating that useful forecasts could be issued with the longer lead. Moreover, it is notable that pure indirect forecasts from January are more skillful than pure direct forecasts from February and thus the skill benefit is at least equal to that obtained by a 1-month reduction in lead-time for direct forecasts.

Results for GHAN for the MAM season are likewise provided in Figure 10. Skill is generally lower than that for GHAS in MAM. In contrast to GHAS, for the indirect forecasts skill is low, near or below 55%, while for the direct forecasts skill is generally somewhat higher, up to 61% for

FIGURE 9 GHAS region and MAM season, verification period 2002–2017. Regionally averaged ROC skill for Dry and Wet tercile categories, using indirect and direct forecasts combined in the proportion indicated below the bars (e.g., 40 = 40%/60% indirect: direct). In each pair of bars the right bar is for the Wet category and the left bar Dry. The letters Ja (Fe) below the pairs of bars indicate use of hindcast data with start dates in January (February). Vertical black lines indicate 95% (2.5 to 97.5%) confidence intervals (calculated from 22,000 bootstrap analyses) for January start Dry category: Intervals for other cases are very similar. GHAS, Greater Horn of Africa South; MAM, March–April–May

TABLE 2 Numbers of EOF and CCA modes selected for indirect (CCA-based) GHAS forecasts for MAM in Example 3, using hindcasts with start dates in JAN (January) or FEB (February)
the Wet category with February start dates. (The exception is the Dry category with February starts, for which the direct skill is slightly below the indirect skill.) Combining the direct and indirect forecasts is not clearly beneficial. Overall, for the Wet category the direct forecasts are best, and January starts are nearly as good as February starts. For the Dry category the indirect forecasts provide slightly better skill than direct forecasts for February starts, but the reverse is true for January starts.

In the SOND season for GHAS (Figure 11) the skill of both direct and indirect forecasts is substantially higher than in the MAM season. (This is consistent with, e.g., Shukla et al., 2016, who noted higher skill for OND versus MAM in their assessment of multi-model direct forecasts in a similar region). For the Dry category the pure indirect forecasts have ROC skill approaching 80%, versus just over 65% for pure direct forecasts. The July and August starts have very similar skill for the Dry category. For the Wet category the indirect forecasts again outperform the direct forecasts, and skill is somewhat lower than for the Dry category. The indirect forecasts have 73% skill for August starts, 69% for July starts. For both categories combination has no clear benefit. Again, the skill increment gained by the pure indirect approach is at least as good as a 1-month reduction in lead time with the direct approach.

Skills for the SOND season in GHAN are provided in Figure 12. As for GHAS skill is better than for MAM. In contrast to MAM forecasts for this region, the SOND indirect forecasts have higher skill than the direct forecasts. For August starts the ROC skill for pure indirect forecasts is 69% for Dry and 66% for Wet, with skill about 5% lower for July starts. In this case combination provides a slight advantage, for Wet and Dry categories and both lead times, but less than 5% increase on pure indirect skill.

**FIGURE 10** As Figure 9 but for the GHAN region, MAM season. For clarity, bootstrap confidence interval bars are not shown but are similar to those in Figure 9. GHAN, Greater Horn of Africa North; MAM, March–April–May

**FIGURE 11** As Figure 10 but for the GHAS region, SOND season. The letters Ju (Au) below the pairs of bars indicate hindcast data with start dates in July (August). GHAS, Greater Horn of Africa South; SOND, September–October–November–December

**5.2 | Comparison with issued GHACOF forecasts**

Mason and Chidzambwa (2009) undertook a verification of African RCOF consensus rainfall forecasts issued between 1998 and 2007, including GHACOF MAM and SOND seasons, constructing regionally averaged ROC skill scores for three categories. The GHACOF region is equivalent to GHAS-plus-GHAN. The multi-model system in Section 5.1 was used to make forecasts for the same periods and region for comparison, with a range of combinations of direct and indirect forecasts as before. The multi-model ROC skills and GHACOF estimates are summarized in Table 3 for the Wet and Dry categories.

For the MAM season and the Wet category, the best multi-model skill obtained was 64% (for both January and February starts, using an 80% indirect method combination), which is an improvement on the 53% skill estimate for
For the MAM season in GHAS, for the indirect forecasts a methodology involving CCA was adopted, as described in Section 3. CCA temporally related spatial patterns in a large tropical region and in a GHA region can be identified: see, for example, Figure 2a–c. Such patterns can be used to generate probability forecasts for rainfall categories, as illustrated in Figure 2e for the GHAS region SOND season.

Making use of data for a relatively long hindcast period (1960–2005) for two dynamical seasonal forecast systems, in Section 3 direct and indirect forecasts were examined in more detail for the MAM season in GHAS. For the indirect forecasts tropic-wide precipitation extending to 40°N and 40°S was used as the predictor variable: similar results (not shown) were obtained with SST as the predictor. The dominant CCA patterns (Figure 5) were similar to those obtained using contemporaneous observations, demonstrating the ability of the dynamical systems to represent such patterns and relationships.

Overall the indirect forecasts matched the observed seasonal rainfall variability better than the direct forecasts in the 1989–2005 assessment period (Figure 3). The indirect forecasts captured the strong difference between the 1997 and 1998 GHAS rainfall, when there was a relatively strong match between the predictor patterns and observed tropic-wide rainfall. In some individual seasons (most notably 1999) the direct forecasts were more in line with the observed conditions, when the CCA predictor patterns were weak.

A simple strategy to combine the direct and indirect forecasts was tested, by averaging the categorical probabilities at the gridpoint level. As there is not a theoretical best combination for this approach, a range of different weights was used. According to the measure used in Figure 7, such combination was beneficial. Using a more conventional ROC score, combination did not clearly improve on the score with purely indirect forecasts. We conclude that simple combination is a strategy worth considering, though results will depend on the particular forecast quantity of interest. The contrasting performance of direct and indirect forecasts in individual years also suggests the conditional use of indirect forecasts (perhaps weighting according to predictor projection strength) should be explored, but this was beyond the scope of our investigation.

Following this direct and indirect forecast comparison which demonstrated the benefits of the indirect approach, we undertook a methodical assessment of skill for the GHAS and GHAN regions for the MAM and SOND seasons, using a multi-model with hindcast data for 1981–2017 from four dynamical systems that were operational in 2019. Regionally averaged ROC scores for Wet and Dry categories, for a range of direct and indirect combinations, and two lead times, are provided in Figures 9–12. The highest skill was found for GHAS SOND, lowest for GHAN MAM. In most cases the indirect method clearly outperformed the direct method, with the skill benefit exceeding that obtained by a 1-month reduction in lead time. The indirect method did not show benefit for the GHAN MAM season, suggesting that teleconnections are less effective for that location and season. Recent findings by Vellinga and Milton (2018) and others indicate that the East Africa MAM rainfall season is not coherent over the whole season and different teleconnections apply at different times within the season. Thus it may be possible to improve forecasts for MAM by

**TABLE 3** ROC skill scores for the period 1998–2007

| GHACOF   | Multi-model |
|----------|-------------|
| MAM      |             |
| Wet      | 53          |
| Dry      | 56          |
| SOND     |             |
| Wet      | 47          |
| Dry      | 65          |

Notes: The GHACOF values are as estimated by Mason and Chidzambwa (2009) for the GHACOF region. The multi-model values are for the GHAN-plus-GHAS region.
Abbreviations: GHACOF, GHA Regional Climate Outlook Forum; GHAN, Greater Horn of Africa North; GHAS, Greater Horn of Africa South; ROC, relative operating characteristic.

GHACOF. For the MAM Dry category the best score was 57% for February starts, which is effectively no better than the 56% GHACOF consensus skill estimate.

For the SOND season the best multi-model skills were obtained using August start times and 100% indirect method, with skill 70% for Wet (substantially above the 47% estimate for GHACOF) and 74% for Dry (slightly above the 65% GHACOF value). While the comparison is not precise, the results provide further evidence of the potential benefits of using indirect and combination methods for seasonal outlooks.

**6 | SUMMARY AND DISCUSSION**

Dynamical seasonal prediction systems that use global coupled ocean-atmosphere-land models produce ensemble forecast data that can be used to generate regional forecasts. The most straightforward way to make such a forecast is to use data for the variable of interest (e.g., precipitation) just in the region of interest—a direct forecast. As variability in a region is also influenced by various large-scale climate events, such as ENSO, regional forecasts can also be produced by using large-scale information to infer regional behaviour—that is, by indirect forecasts.

In this article we have described and compared direct and indirect rainfall forecasts in northern and southern regions of the GHA, for March-to-May and September-to-December seasons. For the indirect forecasts a methodology involving CCA was adopted, as described in Section 3. CCA temporally related spatial patterns in a large tropical region and in a GHA region can be identified: see, for example, Figure 2a–c. Such patterns can be used to generate probability forecasts for rainfall categories, as illustrated in Figure 2e for the GHAS region SOND season.

Making use of data for a relatively long hindcast period (1960–2005) for two dynamical seasonal forecast systems, in Section 3 direct and indirect forecasts were examined in more detail for the MAM season in GHAS. For the indirect forecasts tropic-wide precipitation extending to 40°N and 40°S was used as the predictor variable: similar results (not shown) were obtained with SST as the predictor. The dominant CCA patterns (Figure 5) were similar to those obtained using contemporaneous observations, demonstrating the ability of the dynamical systems to represent such patterns and relationships.

Overall the indirect forecasts matched the observed seasonal rainfall variability better than the direct forecasts in the 1989–2005 assessment period (Figure 3). The indirect forecasts captured the strong difference between the 1997 and 1998 GHAS rainfall, when there was a relatively strong match between the predictor patterns and observed tropic-wide rainfall. In some individual seasons (most notably 1999) the direct forecasts were more in line with the observed conditions, when the CCA predictor patterns were weak.

A simple strategy to combine the direct and indirect forecasts was tested, by averaging the categorical probabilities at the gridpoint level. As there is not a theoretical best combination for this approach, a range of different weights was used. According to the measure used in Figure 7, such combination was beneficial. Using a more conventional ROC score, combination did not clearly improve on the score with purely indirect forecasts. We conclude that simple combination is a strategy worth considering, though results will depend on the particular forecast quantity of interest. The contrasting performance of direct and indirect forecasts in individual years also suggests the conditional use of indirect forecasts (perhaps weighting according to predictor projection strength) should be explored, but this was beyond the scope of our investigation.

Following this direct and indirect forecast comparison which demonstrated the benefits of the indirect approach, we undertook a methodical assessment of skill for the GHAS and GHAN regions for the MAM and SOND seasons, using a multi-model with hindcast data for 1981–2017 from four dynamical systems that were operational in 2019. Regionally averaged ROC scores for Wet and Dry categories, for a range of direct and indirect combinations, and two lead times, are provided in Figures 9–12. The highest skill was found for GHAS SOND, lowest for GHAN MAM. In most cases the indirect method clearly outperformed the direct method, with the skill benefit exceeding that obtained by a 1-month reduction in lead time. The indirect method did not show benefit for the GHAN MAM season, suggesting that teleconnections are less effective for that location and season. Recent findings by Vellinga and Milton (2018) and others indicate that the East Africa MAM rainfall season is not coherent over the whole season and different teleconnections apply at different times within the season. Thus it may be possible to improve forecasts for MAM by
blending separate forecasts for different times within the season.

Skills using data from hindcasts that started 1-month earlier than those typically used (e.g., July rather than August for SOND) were generally just below those of hindcasts from the normal start-dates. For some applications, the benefit of a longer lead time could outweigh the small reduction in skill, and it would be worth exploring yet longer lead times.

The results also suggest that the use of dynamical models and an indirect approach can yield forecast skills at levels comparable to those of consensus forecasts from GHACOF.

It is likely that higher indirect skills could be attained through a more extensive investigation of predictor domains, as these were not optimized in this study. Note also that CCA might also be used with a predictor domain centred on the GHA rather than a tropics-wide domain: CCA would operate to correct within-region systematic biases which may be strongly connected with the larger scale and thus carry much of the large-scale information.

Simple combination of direct and indirect forecasts generally did not improve the ROC scores, though small gains can be seen in Figure 12 (GHAN SOND). The lack of clear benefit of combination may be because one method (usually indirect) had substantially higher skill than the other. Another reason may be that the two methods already have some commonality, in that the direct forecast values in the region are already influenced by the teleconnections that underlie the indirect forecasts. Further, the indirect predictor domains used here included the GHA regions, though given the disparity in size this is likely a minor effect.

The availability of hindcast data covering over 30 years from four of the dynamical systems among the WMO-designated Global Producers of Long-Range Forecasts was essential to the work described in Section 5, with separate training and verification periods. Extension of such availability to more systems would aid similar future regional skill exploration. Similarly, extension of the length of hindcast periods would be beneficial, and in particular would aid the investigation of conditional strategies.

ACKNOWLEDGEMENTS

The CCA forecasts were produced using the IRI Climate Predictability Tool (CPT) http://iri.columbia.edu/our-expertise/climate/tools/cpt/. We thank Simon Mason for advice on its use and methodology. The work was greatly facilitated by the availability of hindcast data in the IRI data library (http://iridl.ldeo.columbia.edu/). We thank the IGAD Climate Prediction and Applications Centre (ICPAC) for collaboration in exploring the use of combination as a contributing tool for consensus seasonal forecasting for the GHA region, and the reviewers for their constructive comments. Andrew Colman and Richard Graham acknowledge funding from the EU Framework 7 SPECS project and from the UK government via the NERC/DIFD ForPAc project—grant number NE/P000428/1.

ORCID

Andrew W. Colman https://orcid.org/0000-0003-0863-3358

REFERENCES

Adler, R.F., Gu, G. and Huffman, J.G. (2012) Estimating climatological bias errors for the Global Precipitation Climatology Project (GPCP). Journal of Applied Meteorology and Climatology, 51, 84–99. https://doi.org/10.1175/JAMC-D-11-052.1.

Alexander, M.A., Blade, I., Newman, M., Lanzante, J.R., Lau, N.-C. and Scott, J.D. (2002) The atmospheric bridge: the influence of ENSO teleconnections on air–sea interaction over the global oceans. Journal of Climate, 15, 2205–2231.

Arribas, A., Glover, M., Maidens, A., Peterson, A.K., Gordon, M., MacLachlan, C., Graham, R., Fereday, D., Camp, J., Scaife, A.A., Xavier, P., McLean, P., Colman, A. and Cusack, S. (2011) The GloSea4 ensemble prediction system for seasonal forecasting. Monthly Weather Review, 139, 1891–1910.

Barnston, T.P. and Preisendorfer, R. (1987) Origins and levels of monthly and seasonal forecast skill for the United States surface air temperatures determined by canonical correlation analysis. Monthly Weather Review, 115, 1825–1850.

Barnston, A.G. (1994) Linear statistical short-term climate predictive skill in the northern hemisphere. Journal of Climate, 7, 1513–1564.

Barnston, A.G. and Smith, T.M. (1996) Specification and prediction of global surface temperature and precipitation from global SST using CCA. Journal of Climate, 9, 2660–2697.

Broecker, J. (2012) Probability forecasts. In: Forecast Verification: A practitioner’s Guide in Atmospheric Science. Wiley: Jolliffe IT and Stephenson DB Eds, pp. 119–140 ISBN: 978-0-470-66071-3.

Chen, M.P., Xie, P., Janowiak, J.E. and Arkin, P.A. (2002) Global land precipitation: a 50-yr monthly analysis based on gauge observations. Journal of Hydrometeorology, 3, 249–266.

Efron, B. and Tibshirani, R. (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science, 1, 54–75.

Feddersen, H., Navarra, A. and Ward, M.N. (1999) Reduction of model systematic error by statistical correction for dynamical seasonal predictions. Journal of Climate, 12, 1974–1989.

Fraedrich, K. and Smith, N.R. (1989) Combining predictive schemes in long-range forecasting. Journal of Climate, 2, 291–294.

Glahn, H.R. (1968) Canonical correlation and its relationship to discriminant analysis and multiple regression. Journal of the Atmospheric Sciences, 25, 23–31.

Graham, R.J., Yun, W.-T., Kim, J., Kumar, A., Jones, D., Bettio, L., Gagnon, N., Kolli, R.K. and Smith, D. (2011) Long-range forecasting and the global framework for climate services. Climate Research, 47, 47–55. https://doi.org/10.3354/cr00963.
Kipkogei, O., Mwanthi, A.M., Mwesigwa, J.B., Athuru, Z.K.K., Wanzala, M.A. and Artan, G. (2017) Improved seasonal prediction of rainfall over East Africa for application in agriculture: statistical downscaling of CFSv2 and GFDL-FLOR. *Journal of Applied Meteorology and Climatology*, 56, 3229–3243. https://doi.org/10.1175/JAMC-D-16-0365.1.

Kirtman, B.P., Min, D., Infanti, J.M., Kinter, J.L., Ill, Paolino, D.A., Zhang, Q., van den Dool, H., Saha, S., Mendez, M.P., Becker, E., Peng, P., Tripp, P., Huang, J., DeWitt, D.G., Tippett, M.K., Barnston, A.G., Li, S., Rosati, A., Schubert, S.D., Rienecker, M., Suarez, M., Li, Z.E., Marshall, J., Lim, Y.K., Tribbia, J., Pegion, K., Merryfield, W.J., Denis, B. and Wood, E.F. (2014) The north American multi-model ensemble: phase-1 seasonal to interannual prediction: phase-2 toward developing intra-seasonal prediction. *Bulletin of the American Meteorological Society*, 95, 585–601.

Kug, J.S., Lee, J.Y., Kang, I.S., Wang, B. and Park, C.K. (2008) Optimal multi-model ensemble method in seasonal climate prediction. *Asia-Pacific Journal of Atmospheric Sciences*, 44, 259–267.

Landman, W.A., DeWitt, D., Lee, D., Beraki, A. and Lotter, D. (2012) Seasonal rainfall prediction skill over South Africa: one versus two-tiered forecasting systems. *Weather and Forecasting*, 27, 489–501. https://doi.org/10.1175/WAF-D-11-00078.1.

MacLachlan, C., Arribas, A., Peterson, K.A., Maidens, A., Fereray, D., Scaife, A.A., Gordon, M., Vellinga, M., Williams, A., Comer, R.E., Camp, J., Xavier, P. and Madec, G. (2015) Global seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system. *Quarterly Journal of the Royal Meteorological Society*, 141, 1072–1084. https://doi.org/10.1002/qj.2396.

Manzano, V.J.P. and Ines, V.M. (2015) Predictability of may to august (MJJA) seasonal rainfall in northern Philippines. *Journal of Nature Studies*, 14(2), 32–39.

Mason, S. and Chidzambwa, S. (2009) Verification of RCOF forecasts. International Research Institute for Climate and Society, Palisades, New York. Technical Report number: 10.7916/D8ST3SB0

Muali, C.C., Ward, M.N. and Colman, A.W. (1998) Towards the prediction of East Africa short rains based on sea surface temperature-atmosphere coupling. *International Journal of Climatology*, 18, 975–997.

Mwale, D., Gan, T.Y. and She, S.S.P. (2004) A new analysis of variability and predictability of seasonal rainfall of central southern Africa for 1950-94. *International Journal of Climatology*, 24, 1509–1530.

Mwangi, E., Wetterhall, F., Dutra, E., Di Giuseppe, F. and Pappenberger, F. (2014) Forecasting droughts in East Africa. *Hydrology and Earth System Sciences*, 18, 611–620.

Nicholson, S.E. (2014) The predictability of rainfall over the Greater Horn of Africa. Part I: prediction of seasonal rainfall. *Journal of Hydrometeorology*, 15, 1011–1027. https://doi.org/10.1175/JHM-D-13-062.1.

Nicholson, S.E. (2015) The predictability of rainfall over the Greater Horn of Africa. Part II: prediction of monthly rainfall during the long rains. *Journal of Hydrometeorology*, 16, 2001–2012. https://doi.org/10.1175/JHM-D-14-0138.1.

Nicholson, S.E. (2017) Climate and climatic variability of rainfall over eastern Africa. *Reviews of Geophysics*, 55, 590–635.

Ogallo, L.A. (1988) Relationships between seasonal rainfall in East Africa and the Southern Oscillation. *Journal of Climatology*, 8, 31–43.

Owiti, Z., Ogallo, L.A. and Mutembi, J.N. (2008) Linkages between the Indian Ocean Dipole and East African seasonal rainfall anomalies. *Journal of Kenya Meteorological Society*, 2, 3–17.

Rourke, J.M.A. (2011) Seasonal prediction of African rainfall with a focus on Kenya. PhD Thesis, University College London, http://discovery.ucl.ac.uk/1302403/

Saji, N.H., Goswami, B.N., Vinayachandran, P.N. and Yamagata, T. (1999) A dipole mode in the tropical Indian Ocean. *Nature*, 401, 360–363.

Shukla, S., Funk, C. and Hoell, A. (2014) Using constructed analogs to improve the skill of National Multi-Model Ensemble March-April precipitation forecasts in equatorial East Africa. *Environmental Research Letters*, 9, 094009. https://doi.org/10.1088/1748-9224/9/9/094009.

Shukla, S., Roberts, J., Hoell, A., Funk, C.C., Robertson, F. and Kirtman, B. (2016) Assessing north American multimodel ensemble (NMME) seasonal forecast skill to assist in the early warning of anomalous hydrometeorological events over East Africa. *Climate Dynamics*, 1–17. https://doi.org/10.1007/s00382-016-3296-z.

Stockdale, T., Johnson, S., Ferranti, L., Balmaseda, M. and Briceag, S. (2018) ECMWF's new long-range forecasting system SEAS5. European Centre for Medium-Range Weather Forecasts Newsletter No. 154. Available at https://www.ecmwf.int/en/elibrary/17938-newsletter-no-154-winter-2017-18

Vellinga, M. and Milton, S.F. (2018) Drivers of interannual variability of the east African long rains. *Quarterly Journal of the Royal Meteorological Society*, 144, 861–876.

Walker, D.P., Birch, C.E., Marsham, J.H., Scaife, A.A., Graham, R.J. and Segale, Z.T. (2019) Skill of dynamical and GHACOF consensus seasonal forecasts of east African rainfall. *Climate Dynamics*. https://doi.org/10.1007/s00382-019-04835-9.

Weisheimer, A., Doblas-Reyes, F.J., Palmer, T.N., Alessandri, A., Arribas, A., Déqué, M., Keenlyside, N., MacVean, M., Navarra, A. and Rogel, P. (2009) ENSEMBLES: a new multi-model ensemble for seasonal-to-annual predictions — skill and progress beyond DEMETER in forecasting tropical Pacific SSTs. *Geophysical Research Letters*, 36, L21711. https://doi.org/10.1029/2009GL040896.