

## Seasonal-to-Interannual Variability of Ethiopia/Horn of Africa Monsoon. Part II: Statistical Multimodel Ensemble Rainfall Predictions

ZEWDU T. SEGELE

*Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma*

MICHAEL B. RICHMAN, LANCE M. LESLIE, AND PETER J. LAMB

*Cooperative Institute for Mesoscale Meteorological Studies, and School of Meteorology, University of Oklahoma, Norman, Oklahoma*

(Manuscript received 8 July 2014, in final form 23 October 2014)

### ABSTRACT

An ensemble-based multiple linear regression technique is developed to assess the predictability of regional and national June–September (JJAS) anomalies and local monthly rainfall totals for Ethiopia. The ensemble prediction approach captures potential predictive signals in regional circulations and global sea surface temperatures (SSTs) two to three months in advance of the monsoon season. Sets of 20 potential predictors are selected from visual assessments of correlation maps that relate rainfall with regional and global predictors. Individual predictors in each set are utilized to initialize specific forward stepwise regression models to develop ensembles of equal number of statistical model estimates, which allow quantifying prediction uncertainties related to individual predictors and models. Prediction skill improvement is achieved through error minimization afforded by the ensemble.

For retroactive validation (RV), the ensemble predictions reproduce well the observed all-Ethiopian JJAS rainfall variability two months in advance. The ensemble mean prediction outperforms climatology, with mean square error reduction ( $SS_{\text{Clim}}$ ) of 62%. The skill of the prediction remains high for leave-one-out cross validation (LOOCV), with the observed–predicted correlation  $r(SS_{\text{Clim}})$  being +0.81 (65%) for 1970–2002. For tercile predictions (below, near, and above normal), the ranked probability skill score is 0.45, indicating improvement compared to climatological forecasts. Similarly high prediction skill is found for local prediction of monthly rainfall total at Addis Ababa ( $r = +0.72$ ) and Combolcha ( $r = +0.68$ ), and for regional prediction of JJAS standardized rainfall anomalies for northeastern Ethiopia ( $r = +0.80$ ). Compared to the previous generation of rainfall forecasts, the ensemble predictions developed in this paper show substantial value to benefit society.

### 1. Introduction

Rainfall is the most important climate element affecting the Ethiopian population (96.6 million). The main rainy season (called Kiremt) occurs during June–September (JJAS), when the northern two-thirds of the country receives 65%–95% of Ethiopia's total annual rainfall (Segele and Lamb 2005) and produces 85%–95% of its food (Degefu 1987). Because all Ethiopian agricultural

activities and resulting crop production are dependent upon on the distribution and amount of JJAS rainfall, its seasonal prediction is of great importance for agricultural planning and socioeconomic disaster mitigation.

Several previous studies have examined aspects of the intraseasonal-to-interannual variability of Ethiopian rainfall. These studies have considered the following: Kiremt onset, cessation, and resulting growing season variability (Segele and Lamb 2005); the abrupt latitudinal rainfall changes involved and their representation in climate model simulations (Riddle and Cook 2008; Segele et al. 2009c); associated large-scale atmospheric circulation characteristics and global sea surface temperature patterns (Beltrando and Camberlin 1993; Shanko and Camberlin 1998; Segele et al. 2009a; Berhane et al. 2014); and connections to the Indian monsoon

 Denotes Open Access content.

*Corresponding author address:* Dr. Zewdu T. Segele, CIMMS, University of Oklahoma, Suite 2100, 120 David L. Boren Boulevard, Norman, OK 73072-7304.  
E-mail: zewdu@ou.edu

DOI: 10.1175/JCLI-D-14-00476.1

(Camberlin 1997; Vizi and Cook 2003). Also investigated have been the hydrological responses to rainfall (Conway 2000), temporal trends and decadal shifts of the rain (Seleshi and Zanke 2004; Bowden and Semazzi 2007; Cheung et al. 2008; Williams et al. 2012; Viste et al. 2013; Berhane et al. 2014), and seasonal spring (Belg) and Kiremt rainfall predictability (Gissila et al. 2004; Korecha and Barnston 2007; Block and Rajagopalan 2007; Diro et al. 2008; Korecha and Sorteberg 2013; Jury 2013; Nicholson 2014).

Segele et al. (2009b, hereinafter Part I) used wavelet analysis to identify temporal and spectral characteristics of the seasonal-to-interannual variability of 5-day average Ethiopian JJAS rainfall, and to present a time–frequency quantification of teleconnections between the rainfall and large-scale atmospheric circulation and global sea surface temperature (SST) patterns. The results linked Ethiopian monsoon rainfall variations principally with seasonal-to-interannual time-scale atmospheric circulation patterns that involve fluctuations in the major components of the monsoon system, including the monsoon trough, tropical easterly jet (TEJ), and Somali low level jet (SLLJ). Part I thus provided the physical basis for understanding the mechanisms of Ethiopian rainfall variability and identifying regional rainfall drivers. The current study builds on that foundation by developing new statistical methods for assessing the predictability of monthly-to-seasonal Ethiopian rainfall on national and local scales, as a step toward an operational seasonal prediction capability for Ethiopia.

On interannual time scales, monsoon rainfall variability now is widely recognized as being governed by slowly varying surface boundary conditions, such as SSTs, land surface albedo, and soil moisture. For Africa, the role of basin- and global-scale SST anomalies for such rainfall variability received considerable attention over the last 35 years (e.g., Lamb 1978a,b; Folland et al. 1986, 1991; Lamb and Pepler 1992; Barnston et al. 1996; Ward 1998; Mutai and Ward 2000; Camberlin et al. 2001; Giannini et al. 2003; Segele et al. 2009a; Part I; Ndiaye et al. 2011). Using diagnostic relationships developed in those studies, several other investigations further assessed the predictability of Ethiopian Belg (Diro et al. 2008) and Kiremt (Gissila et al. 2004; Block and Rajagopalan 2007; Korecha and Barnston 2007; Jury 2013; Nicholson 2014) seasonal rainfall. The possible role of land surface conditions for African rainfall was noted first by Charney (1975) and has been investigated in many subsequent studies (e.g., Charney et al. 1977; Charney and Shukla 1981; Xue and Shukla 1993; Clark and Arritt 1995; Clark et al. 2001).

The surface boundary focus of the present Ethiopian study is SST. However, El Niño–Southern Oscillation

(ENSO)-related “predictability barrier” in Northern Hemisphere spring (e.g., Goswami and Shukla 1991; Webster and Yang 1992; Webster et al. 1998) can pose a major challenge to providing seasonal rainfall forecasts two or more months in advance in the tropics (Goddard et al. 2001; Korecha and Barnston 2007). Also, because the effects of slowly evolving global SST variations on Ethiopian rainfall must be transmitted through changes in local and regional atmospheric circulations and SST patterns, development of monthly-to-seasonal Ethiopian rainfall predictions can be enhanced by identifying local and regional predictors and aggregating their effects into ensemble-based statistical prediction schemes.

The present Kiremt predictability investigation involves 2–3-month lead time forecasts of local, regional, and national rainfall for Ethiopia. Individual predictions employ a set of 20 predetermined regional and global predictor variables to construct equal number ensembles of initial statistical model estimates that capture potentially valuable and unique signals from every variable in the 20-member predictor set. Final models are selected objectively from the 20 initial models after a series of tests are applied that address regression assumptions and statistical significance requirements. The averaging of the final ensembles of statistical estimates is shown to improve cross-validation skill compared to a traditional, single-model based technique. The ensemble provides a novel approach to quantify the envelope of uncertainty created by the model building process, reflecting the predictive uncertainties associated with individual predictors and models. Cross-validation skill improvement is achieved through error minimization afforded by the ensemble.

## 2. Data and methodology

Accurate prediction of all-Ethiopian Kiremt rainfall would provide valuable national-scale information. Stakeholders would benefit from accurate monthly rainfall predictions at local and regional levels, especially for highly populated areas and those with large interannual rainfall variability. To achieve a high level of accuracy and a measure of forecast uncertainty, ensemble-based statistical prediction tools were developed and assessed for 1) all-Ethiopian standardized Kiremt rainfall anomalies, 2) August total rainfall at Addis Ababa (the most populated city in the African Union) and Combolcha (center of the 1984 Ethiopian drought), and 3) northeastern Ethiopia standardized JJAS rainfall anomalies. Addis Ababa and Combolcha have daily rainfall records, with Combolcha’s high interannual rainfall variability challenging traditional

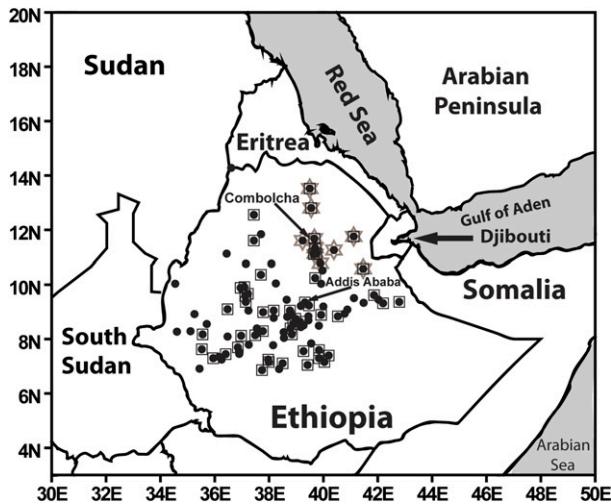


FIG. 1. Regional map showing locations of 100 Ethiopian rain gauge stations (dots) with data for 1970–99. Squares enclose 52 stations for which 2000–02 data were added to those for 1970–99 to extend use of the retroactive verification (RV) and leave-one-out cross validation (LOOCV) methods for all-Ethiopian June–September (JJAS) rainfall anomaly prediction through 2002. Stars locate 11 northeastern Ethiopia stations with 1970–99 data, with 7 collocated squares added to extend use of LOOCV for northeastern Ethiopia regional prediction of JJAS rainfall anomaly through 2002. Thin arrows locate Addis Ababa and Combolcha, which were used for local August rainfall total prediction for 1970–99 using the LOOCV approach.

single-model-based seasonal forecast methods (e.g., Gissila et al. 2004; Korecha and Barnston 2007). Northeastern Ethiopia was selected for regional prediction because of the region’s susceptibility to drought and high rainfall variability (e.g., Lanckriet et al. 2015).

#### a. Data

This study utilizes sets of Ethiopian rainfall (predictand) and large-scale atmospheric and global SST (predictor) data for 1970–99 that were described in Part I and now have been extended through 2000–02. For seasonal prediction at a national level, a standardized all-Ethiopian JJAS rainfall index originally was constructed for 1970–99 using daily totals from 100 rain gauge stations (Fig. 1, dots). The present extension of that index for 2000–02 employed seasonal totals for 52 of those stations (Fig. 1, squares) analyzed in Korecha and Barnston (2007; data obtained from A. G. Barnston in 2008). For local prediction examples, August monthly rainfall totals for 1970–99 were used for Addis Ababa and Combolcha (Fig. 1). For regional prediction, 11 stations from three regional states (Amahara, Afar, and Tigray) were used to construct a standardized regional JJAS rainfall index for 1970–99 for northeastern Ethiopia (Fig. 1, stars). This index was extended for 2000–02 using

seven stations from Korecha and Barnston (2007) seasonal total rainfall for northeastern Ethiopia (Fig. 1, collocated squares/stars).

The set of atmospheric predictors for 1970–99 was constructed from monthly NCEP–NCAR reanalysis data (50°N–40°S, 30°W–90°E and 2.5° latitude  $\times$  2.5° longitude spatial resolution; Kalnay et al. 1996). These predictors included daily-averaged fields of mean sea level pressure (MSLP), geopotential height, temperature, horizontal wind, vertical velocity, and specific humidity at standard pressure levels. SST predictors for 1970–99 were obtained from the Met Office Hadley Centre global monthly SST dataset (HadISST1; 1° latitude  $\times$  1° longitude resolution; Rayner et al. 2003). Both these datasets were extended to 2002.

#### b. Prediction design

In Part I, wavelet analysis identified the temporal and spectral characteristics of the intraseasonal-to-interannual variability of 5-day average JJAS Ethiopian rainfall, and their strong contemporary teleconnections with large-scale atmospheric circulation and global SST patterns. Although variability on quasi-biennial (QB; 5%) and ENSO (2%) time scales accounted for much less rainfall variance than the annual mode (66%), they modulated the season-long persistence of major regional monsoon components and their remote teleconnection linkages. However, time-lagged relationships between regional rainfall and such large-scale variables generally do not remain strong or statistically significant (e.g., DelSole and Shukla 2002; Segele et al. 2009a, their Fig. 14), which limits their predictive potential.

The present study therefore develops and applies an ensemble-based multiple linear regression technique to predict 1) standardized JJAS rainfall anomalies for all of Ethiopia for 1970–2002, 2) monthly rainfall totals for Addis Ababa and Combolcha for 1970–99, and 3) regional standardized rainfall anomalies for northeastern Ethiopia for 1970–2002 (Fig. 1). This approach assumes that slowly evolving global SST variations affect Ethiopian rainfall through changes in more regional and local atmospheric circulation and SST anomalies, even when the influential ENSO state is unpredictable or its phase has minimal effect (e.g., DelSole and Shukla 2002; Nicholson 2014). The methodology involves three distinct steps: 1) building a set of potential atmospheric circulation and SST predictors from a series of correlation maps; 2) constructing initial ensemble model members by specifying each potential predictor as the first predictor in a multiple regression model; and 3) selecting final ensemble model members based on statistical significance tests.

## 1) STEP 1: BUILD A SET OF POTENTIAL PREDICTORS

Guided by [Part I](#), the full three-dimensional state of the regional atmosphere and global SST was represented by a set of potential March predictors of Kiremt (JJAS) rainfall. Raw atmospheric predictors consisted of grid-point values for 50°N–40°S, 30°W–90°E of geopotential height ( $\Phi$ ), temperature ( $T$ ), horizontal wind ( $u$ ,  $v$ ), vertical velocity ( $w$ ), and specific humidity ( $q$ ) at standard pressure levels, and MSLP. To accommodate potential interaction effects between these predictors, additional second-order predictors were derived by their cross multiplication—that is,  $\Phi u$ ,  $\Phi v$ ,  $Tu$ ,  $Tv$ ,  $uw$ ,  $uv$ , and  $vw$  at 12 pressure levels from 1000 to 100 hPa, and  $qu$  and  $qv$  at eight levels from 1000 to 300 hPa. Also, following [Part I](#), the horizontal wind components and vertical velocity for 1000–100 hPa within 30°N–20°S, 30°–50°E were used to construct zonal and meridional vertical cross sections of  $uw$  and  $vw$ . Raw global (60°N–60°S) SST predictors for March were used in their original form. Each gridpoint variable was standardized. Correlations were computed at all grid points and levels and, following [Gissila et al. \(2004\)](#), [Block and Rajagopalan \(2007\)](#), [Korecha and Barnston \(2007\)](#), and [Nicholson \(2014\)](#), coherent areas of highly significant correlation were selected and the average values of individual variables within the coherent areas were taken as candidate predictors. Here, it was required that the correlations remain statistically significant at 5% level according to a nonparametric bootstrap test for 1970–90 and 1970–99. In addition, the physical significance of the correlations was assessed for all selected predictors. Representative samples of predictors are discussed in [section 3a](#). The number of candidate predictors for a given predictand was limited to a maximum of 20. These predictors were then used for both retroactive/retrospective verification (RV) and leave-one-out cross validation (LOOCV) approaches. In RV, the 20 selected March predictors were used to generate all-Ethiopian Kiremt (JJAS) rainfall predictions for the 1990–2002 verification period, based on 1970–89 model development data. In the LOOCV, 20 March predictors were used for each year during 1) 1970–2002 for all of Ethiopia national and northeastern regional predictions of seasonal rainfall anomalies and 2) 1970–99 for local monthly rainfall prediction at Addis Ababa and Combolcha, based on model development data for all other years in the corresponding training period.

The above predictors may not be mathematically independent. However, the problem of multicollinearity that arises in a given predictor matrix, owing to lack of independence between levels and variables, was evaluated for individual regression models during the final model selection stage, as described in step 3 below.

## 2) STEP 2: CONSTRUCTION OF MULTIMODEL ENSEMBLE SET

Model development used a forward selection stepwise multiple linear regression procedure ([S-PLUS 2013](#)) that began with base models containing only one predictor from the preselected set and an intercept. Other preselected predictors then were added individually in turn, until no further improvement resulted according to the Akaike information criterion (AIC; [Venables and Ripley 1997](#), 218–222). This process was repeated until all preselected predictors ( $N_p$ ) variables for a given prediction/verification year were used as a first model parameter, giving  $N_p$  individual prediction equations. This ensures that a potential predictive signal uniquely associated with a predictor is not unduly discarded. The forward selection procedure was necessary because the alternative backward elimination stepwise approach starts from a full (complex) model and does not guarantee that the final model will contain a particular predictor as its first parameter.

## 3) STEP 3: SELECTION OF FINAL MODEL MEMBERS

Final multimodel ensemble members were selected from candidate models identified in step 2, based on predictor independence, model assumption validity, model coefficient statistical significance, and overall model significance. Several procedures were applied to ensure that the multiple regression assumption of independent predictors was not seriously violated. Correlations between all predictors were determined for the entire time series, and models for which correlation magnitudes between any two predictors exceeded 0.5 were discarded. Further scrutiny for multicollinearity used the variance inflation factor (VIF; [Tamhane and Dunlop 2000](#), p. 417) to eliminate remaining models with VIFs  $>10$ . Remaining models were considered for further analysis only when the null hypotheses of normality, zero autocorrelation, and homoscedasticity were not rejected at the 5% ( $\alpha = 0.05$ ) significance level according to the following tests: the Shapiro–Wilk normality test ([Steinskog et al. 2007](#)), the Durbin–Watson test for serial independence ([Wilks 2006](#), 192–193), and the Breusch–Pagan test for conditional heteroscedasticity ([Breusch and Pagan 1979](#)). A model then was admitted to a final multimodel ensemble set only when its individual regression coefficients, as well as the overall model fit, were statistically significant at the 5% level according to an  $F$  test (e.g., [Draper and Smith 1981](#), 31–33; [Tamhane and Dunlop 2000](#), 407–410). In the above significance tests, if the cost of type I error is much larger (smaller) than expected, the  $\alpha$  value can be adjusted to a smaller (larger) value. The cost of type II error needs to be considered simultaneously when adjusting  $\alpha$ . Following other regression-based prediction development

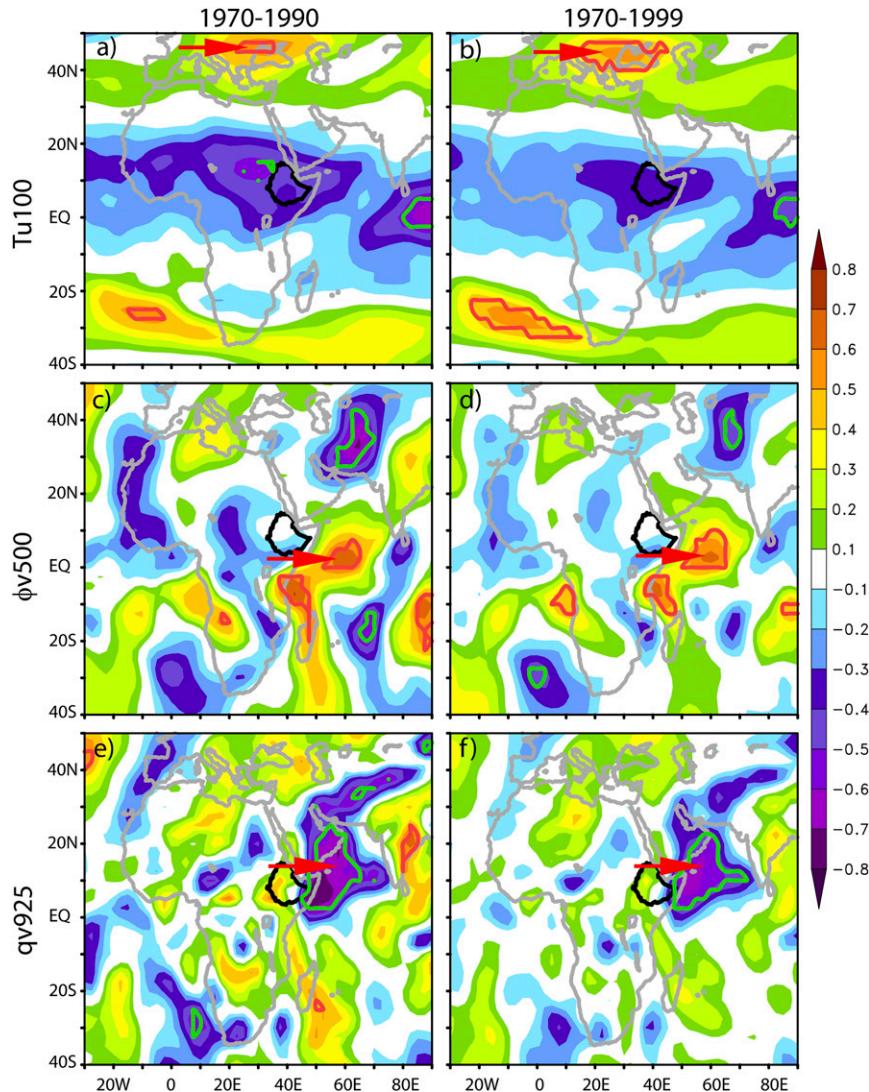


FIG. 2. Correlation analyses involving selected March atmospheric predictors and Kiremt (JJAS) standardized all-Ethiopian rainfall anomalies. Correlation patterns between gridpoint predictors Tu100,  $\phi v500$ , and  $qv925$  (shown at left) for March and standardized all-Ethiopian JJAS rainfall anomalies for (left) 1970–90 and (right) 1970–99. Correlation values are color-coded at right, with solid (dashed) red isopleths enclosing positive (negative) correlation values with achieved significance levels (ASLs) of  $\leq 1\%$  according to a nonparametric bootstrap test. A thick black line encloses Ethiopia. Arrows indicate locations of selected predictors (Table 1).

(e.g., Hastenrath et al. 1995; DelSole and Shukla 2002; Saunders and Lea 2005; Block and Rajagopalan 2007; Korecha and Barnston 2007), nonparametric tests were not performed for regression coefficient significance because of the considerable computational requirements associated with the large number of regression models involved.

### c. Forecast verification

Comprehensive model evaluation was performed by comparing predicted and observed rainfall values for

mutually exclusive/independent verification periods, using goodness-of-fit and relative/absolute error measures with supporting estimates of reliability (Willmott et al. 1985; Legates and McCabe 1999, hereinafter LM99). Mean absolute error (MAE) and root-mean-square error (RMSE) statistics quantified prediction errors. To assess the models' goodness of fit or relative error, the Pearson's product-moment correlation coefficient, coefficient of efficiency [ $E_2$ ; Eq. (2) of LM99], index of agreement [ $d_2$ ; Eq. (3) of LM99], modified index of agreement [ $d_1$ ; Eq. (4) of LM99], and modified

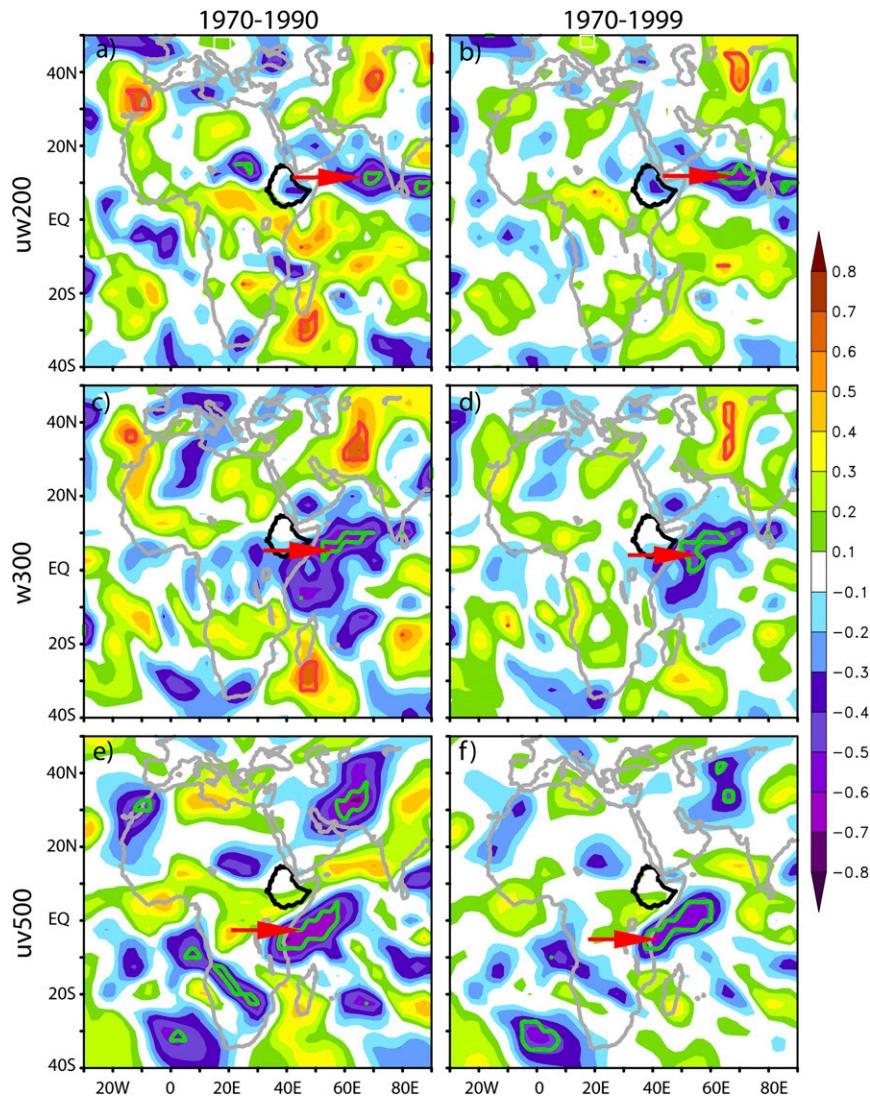


FIG. 3. As in Fig. 2, but for  $uw_{200}$ ,  $w_{300}$ , and  $uv_{500}$ . See Table 1 for predictor details.

coefficient of efficiency [ $E_1$ ; Eq. (5) of LM99] were all calculated. Willmott et al. (1985) and LM99 explain  $E_2$ ,  $d_2$ ,  $d_1$ , and  $E_1$ , while DelSole and Shukla (2002) and Wilks (2006, 280–281) further discuss  $E_2$ .

The aforementioned deterministic model evaluation metrics were complemented by the ranked probability score (RPS), a commonly used ensemble forecast performance measure (Mason and Mimmack 1992; Wilks 2006, 299–302; Block and Rajagopalan 2007), defined as

$$\text{RPS} = \sum_{m=1}^J (Y_m - O_m)^2, \quad (1)$$

where  $J$  is the number of prediction categories, and  $Y_m$  and  $O_m$  are cumulative predicted and observed probabilities through category  $m$ , respectively. The RPS was converted

into a ranked probability skill score (RPSS) by computing it relative to climatological probabilities, where

$$\text{RPSS} = 1.0 - \frac{\text{RPS}}{\text{RPS}_{\text{Clim}}}. \quad (2)$$

Three equiprobable categories (terciles) were used, and were obtained by ranking and grouping predictand time series into the bottom (below normal), middle (near normal), and top (above normal) thirds.

Because the aforementioned measures follow no known distribution, no standard parametric approach can determine the reliability and significance; therefore, significance and confidence intervals were assessed through two nonparametric bootstrap methods that permit any data distribution to generate statistics (Efron

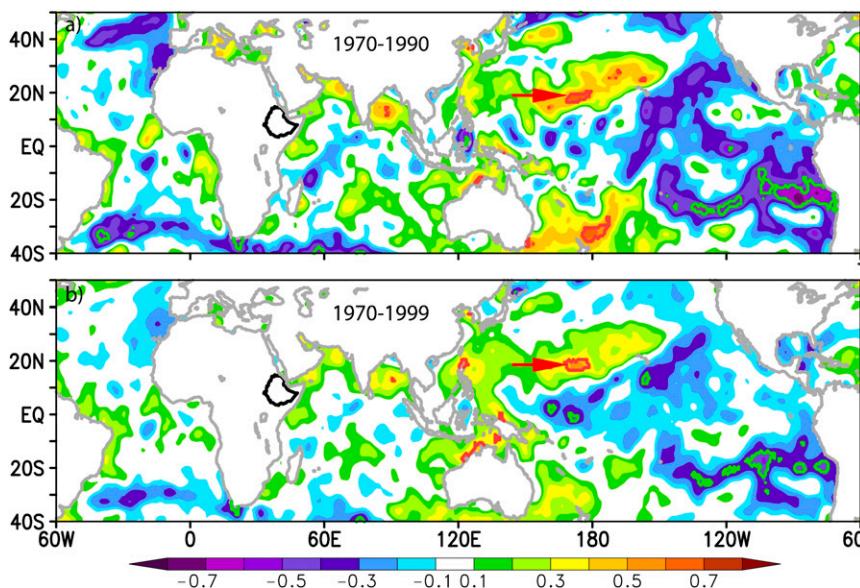


FIG. 4. Correlation analyses involving standardized gridpoint SST predictors for March and Kiremt (JJAS) standardized all-Ethiopian rainfall anomalies for (a) 1970–90 and (b) 1970–99. Correlation values are color-coded at bottom, with solid (dashed) red isopleths enclosing positive (negative) correlation values with ASLs of  $\leq 1\%$  according to a nonparametric bootstrap test. Thick black line encloses Ethiopia. Arrows indicate locations of selected predictors (Table 1).

and Gong 1983; Willmott et al. 1985; Solow 1985; LM99). First, and following Part I, the reliability of each sample statistic (Mason and Mimmack 1992; Nicholls 2001) was assessed by calculating its confidence

intervals, using the bias-corrected and accelerated (BCa) bootstrap technique of Efron and Tibshirani (1993, p. 178). To establish confidence interval accuracy, the BCa method was applied to sample statistics

TABLE 1. Predictors selected for the prediction of standardized Kiremt (JJAS) all-Ethiopian rainfall anomalies (Fig. 1; 100 stations for 1970–99, dots; 52 stations for 2000–02, squares) using both retroactive verification (RV, section 2b) and leave-one out cross validation (LOOCV, section 2b) approaches. Predictors were selected based on their high correlations with rainfall (sections 2b and 3a). All predictand–predictor correlations are statistically significant with ASLs  $\leq 5\%$  according to a nonparametric bootstrap test.

Predictor description	Symbol	Location	Correlation ( $\times 100$ )	
			1970–90	1970–99
1 Air temperature at 100 hPa	T100	Southern Europe (45°–50°N, 10°–15°E)	–46	–53
2 Zonal geopotential height flux at 300 hPa	$\phi u_{300}$	Eastern Indian Ocean (5°S–0°, 80°–85°E)	–59	–51
3 Zonal geopotential height flux at 500 hPa	$\phi v_{500}$	Western Indian Ocean (0°–5°N, 55°–60°E)	+64	+60
4 Zonal moisture flux at 700 hPa	qu700	South Atlantic (25°–30°S, 10°–15°W)	–61	–55
5 Meridional moisture flux at 925 hPa	qv925	Arabian Sea (10°–15°N, 60°–65°E)	+56	+57
6 Meridional moisture flux at 1000 hPa	qv1000	Somali coast (10°–15°N, 50°–55°E)	–74	–65
7 Sea surface temperature	SST	North tropical Pacific (16°–18°N, 171°–175°E)	–63	–55
8 Zonal temperature flux at 100 hPa	Tu100	Southeastern Europe (40°–45°N, 25°–30°E)	+45	+50
9 Meridional temperature flux at 600 hPa	Tv600	Central Indian Ocean (0°–5°N, 60°–65°E)	–72	–67
10 Meridional temperature flux at 925 hPa	Tv925	Northern Arabian Sea (20°–25°N, 60°–65°E)	+55	+58
11 Zonal/meridional wind flux at 500 hPa	uv500	Western Indian Ocean (5°S–0°, 40°–45°E)	+62	+58
12 Zonal/meridional wind flux at 925 hPa	uv925	Somali coast (10°–15°N, 50°–55°E)	–49	–49
13 Zonal/vertical wind flux at 100 hPa	uw100a	Asia (40°–45°N, 65°–70°E)	+52	+55
14 Zonal/vertical wind flux at 100 hPa	uw100b	Asia (45°–50°N, 80°–85°E)	+51	+50
15 Zonal/vertical wind flux at 200 hPa	uw200	Arabian Sea (10°–15°N, 65°–70°E)	–48	–48
16 Meridional/vertical wind flux at 850 hPa	vw850	South Atlantic Ocean (25°–30°S, 10°–15°W)	+46	+48
17 Meridional/vertical wind flux at 925 hPa	vw925	Central Indian Ocean (5°–10°S, 70°–75°E)	+64	+63
18 Meridional/vertical wind flux at 1000 hPa	vw1000	Somali coast (5°–10°N, 50°–55°E)	–56	–53
19 Vertical wind at 150 hPa	w150	Asia (45°–50°N, 65°–70°E)	+71	+59
20 Vertical wind at 300 hPa	w300	Western Indian Ocean (5°S–0°, 50°–55°E)	+66	–59

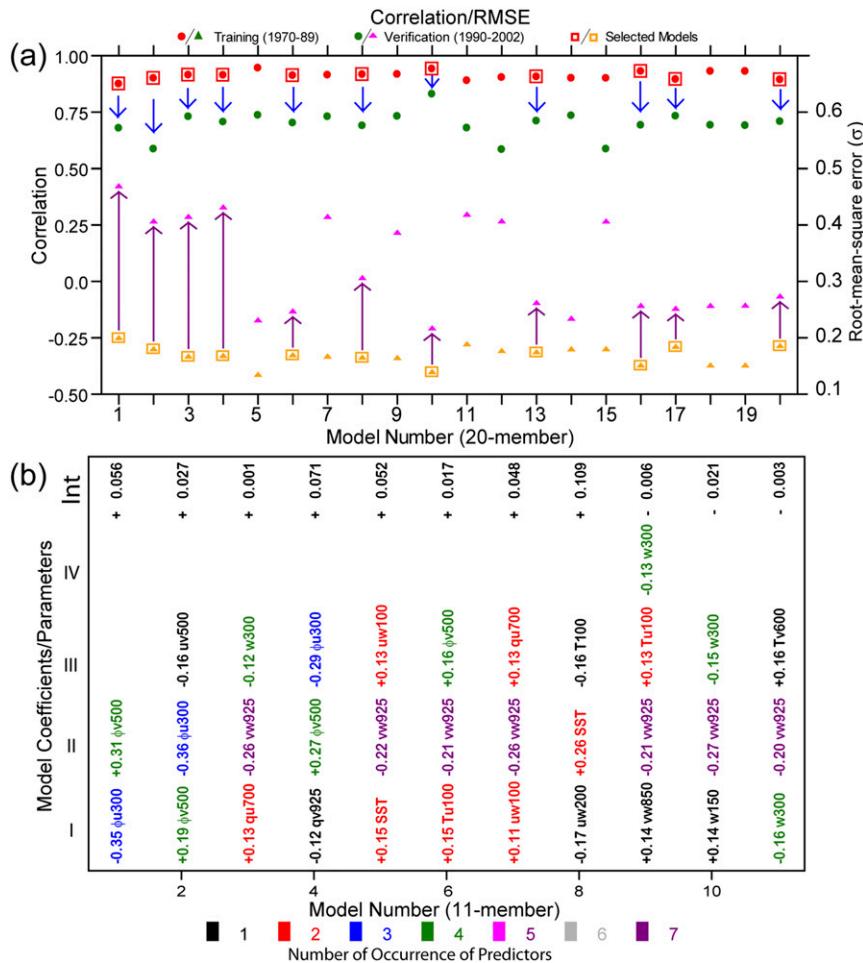


FIG. 5. RV of model performance initiated from average March atmospheric and SST conditions. (a) Correlations/root-mean-square errors (RMSEs) between predicted and observed Kiremt (JJAS) standardized all-Ethiopian rainfall anomalies (Fig. 1) are shown for the 1970–89 training (red/orange) and 1990–2002 independent verification (green/magenta) periods for each of the initial 20 regression models (model sequence number along abscissa). Red/orange squares show 11 models that were selected as final ensemble members, with blue (purple) arrows indicating the decreased (increased) correlations (RMSE) between their predicted values and observations for 1990–2002 independent verification period. (b) Regression equations for the final 11-member multimodel ensemble (abscissa) and the frequency of occurrence of March predictors (color coded at bottom). The ordinate shows model coefficients/parameters, with roman numerals I–IV on right giving the order of predictor appearances in that model. Variable name abbreviations are given in Table 1. See sections 2b, 2c, and 3b for details.

computed from 5000 bootstrap samples of  $N$  pairs of observed/predicted rainfall time series members. Each of the  $N$  pairs was obtained by reshuffling and randomly choosing, with replacement, single-pair values at a time from the  $N$  pairs of observed/predicted rainfall time series members. In the second bootstrap approach, an achieved significance level (ASL)—the probability of exceeding the observed correlation by chance—was obtained for correlations computed from the same 5000 bootstrap samples of  $N$  pairs of observed/predicted rainfall time series members (Efron and Tibshirani 1993,

p. 203). Each observed–predicted correlation was ranked among the bootstrap sample correlations to determine the fraction of the random correlations that were at least as large as the observed correlation.

### 3. Prediction of Kiremt rainfall

#### a. Examples of March predictors for all-Ethiopia Kiremt rainfall

This section documents the physical importance of certain selected atmospheric and SST predictors for

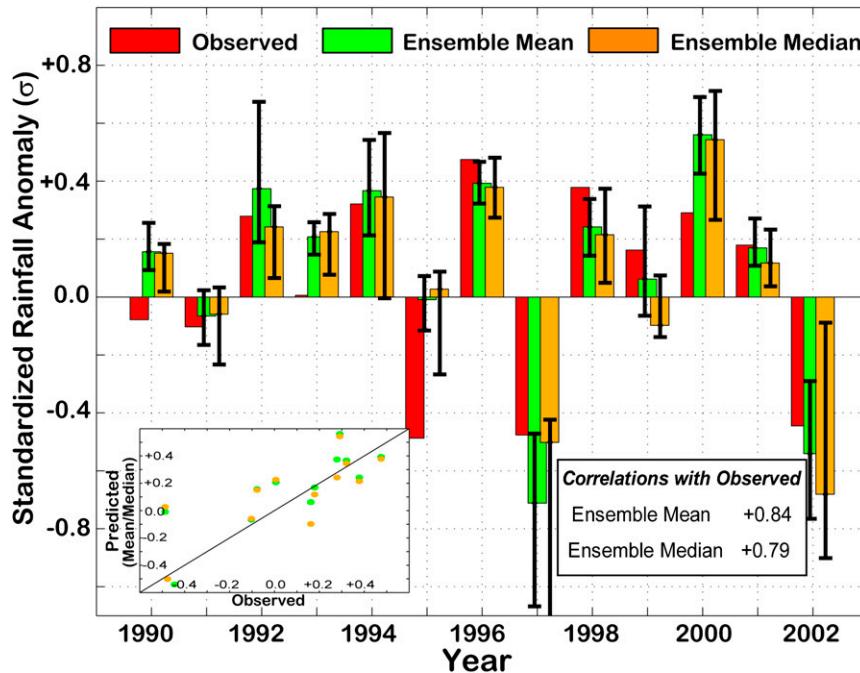


FIG. 6. Two-month lead time prediction of observed (red) Kiremt (JJAS) standardized all-Ethiopian rainfall anomalies (Fig. 1) for 1990–2002 using the RV approach initiated from average March atmospheric and SST conditions. Predictions are mean (green) and median (orange) of final 11-member multimodel ensemble developed using 1970–89 training data. Black vertical bars give 95% confidence intervals for final multimodel-predicted values, computed by applying the bias-corrected and accelerated (BCa) bootstrap method described in section 2c.

Kiremt rainfall prediction. Figures 2 and 3 (atmospheric circulation) and Fig. 4 (SST) document the March predictors that were most strongly correlated with standardized all-Ethiopian Kiremt rainfall for the 1970–90 and 1970–99 periods. The discussion below focuses on

regions of strong correlations indicated by red arrows in Figs. 2–4.

Zonal temperature advection at 100 hPa (Tu100) over southern Europe during March correlated strongly with Kiremt rainfall, with achieved significance levels (ASLs)

TABLE 2. Verification (section 2c) of 2-month lead time model predictions of standardized Kiremt (JJAS) all-Ethiopian rainfall anomalies (Fig. 1, squares, 52 stations) for 1990–2002 using the RV approach (section 2b, steps 2 and 3), based on atmospheric and SST predictors observed in March. Predictions are means for final 11-member multimodel ensemble developed using 1970–89 training data (section 3b). Details on statistics and confidence intervals (BCa bootstrap method) appear in section 2c. The median column gives the median values of the statistics obtained for the individual 11 regression estimates without pooling and the mean column under the bootstrap heading contains average of 5000 bootstrap statistic replicates.

Statistic	Median	Predicted vs obs values	Bootstrap		
			95% confidence intervals		Mean
			Lower limit	Upper limit	
Standard deviation (obs, $\sigma$ )	—	+0.34	+0.25	+0.42	+0.32
Standard deviation (predicted, $\sigma$ )	+0.18	+0.36	+0.20	+0.51	+0.34
Mean absolute error (MAE; $\sigma$ )	+0.22	+0.16	+0.10	+0.24	+0.16
Root-mean-square error (RMSE; $\sigma$ )	+0.27	+0.20	+0.13	+0.29	+0.19
Skill score (vs climatology; $SS_{\text{Clim}}$ )	+0.28	+0.62	+0.26	+0.86	+0.57
Skill score (vs persistence; $SS_{\text{Pers}}$ )	+0.76	+0.87	+0.80	+0.96	+0.79
RPSS (vs climatology; $RPSS_{\text{Clim}}$ )	+0.20	+0.55	+0.29	+0.75	+0.55
Pearson's correlation ( $r$ )	+0.71	+0.84	+0.62	+0.95	+0.84
Index of agreement ( $d_2$ )	+0.82	+0.91	+0.73	+0.97	+0.89
Modified index of agreement ( $d_1$ )	+0.60	+0.72	+0.57	+0.84	+0.70
Modified coefficient of efficiency ( $E_1$ )	+0.20	+0.44	+0.20	+0.67	+0.40

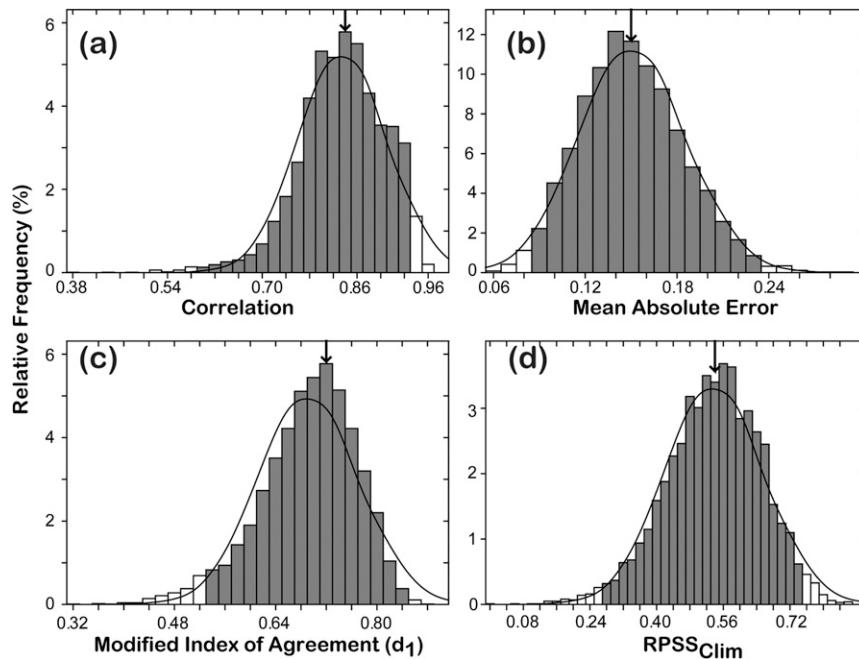


FIG. 7. Frequency distributions of bootstrap samples of selected verification statistics (section 2c) for 2-month lead time prediction of standardized all-Ethiopian Kiremt (JJAS) rainfall anomalies (Fig. 1) for 1990–2002, using the RV approach initiated from March atmospheric and SST conditions. Predictions are means for final 31-member multimodel ensemble developed using 1970–89 training data. (a) Pearson’s product-moment correlation coefficient, (b) mean absolute error, (c) modified index of agreement, and (d) ranked probability skill score over climatology. The 95% BCa bootstrap confidence intervals (section 2c) span the shaded bars. Thin lines depict probability density functions for normal distributions using the means and standard deviations of bootstrap replicates. Arrows locate the original statistic values (Table 1).

of  $\leq 1\%$  according to a nonparametric bootstrap test for both the 1970–90 and 1970–99 periods (Fig. 2, top). The association between north–south variations in zonal westerlies and warm lower stratospheric midlatitude air is manifested by strong correlations. The correlation magnitude decreased appreciably (from +0.50 to +0.40) at 150 hPa, although similar correlation patterns continue in the zonal wind at pressure levels greater than 250 hPa. The positive correlation indicates a likelihood of enhanced Kiremt rainfall with increased Tu100 over southern Europe.

Another potential predictor involving the mass field was found over the western equatorial Indian Ocean in association with the meridional flux of geopotential height at 500 hPa ( $\phi_{v500}$ ) during March (Fig. 2, middle). The main correlation signal was linked to the meridional wind ( $v_{500}$ ), but including variability in the midtropospheric tropical geopotential height field increased the correlation with Kiremt rainfall. Perturbation in the northerly wind occurs in response to the east–west (mainly) oscillation of a weak tropical high over Africa and the eastward passage of subtropical westerlies,

which break the weak zonal midtropospheric easterlies over the equatorial Indian Ocean. The positive  $v_{500}$ –rainfall correlations indicate that the strengthening of equatorial southerlies over the western Indian Ocean during March is likely to enhance Ethiopian rainfall in summer.

In the lower troposphere, meridional moisture flux in the western Indian Ocean in March is associated strongly with Ethiopian Kiremt rainfall. March meridional moisture flux is affected by a transitory subtropical anticyclone over the Arabian Peninsula, which brings moisture for the short rains in Ethiopia from mid-February to mid-May (locally called Belg) when it moves eastward over the Arabian Sea. In association with the eastward displacement of the subtropical high/anticyclone to the Arabian Sea (usually in response to a southward intrusion of westerlies), the dry continental northeasterlies are replaced by maritime northeasterly to easterly humid air flowing into Ethiopia from the Arabian Sea. The negative rainfall and meridional moisture flux at 925 hPa ( $qv_{925}$ ) association indicates that increased (decreased) northerly moisture flux over

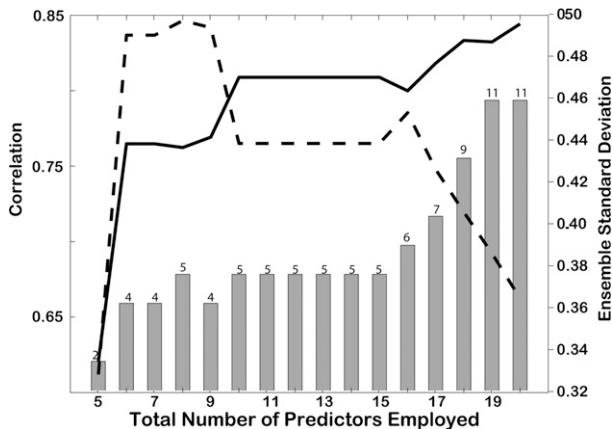


FIG. 8. Sensitivity of cross-validated model performance to ensemble size. Correlation between observed and ensemble predicted (solid black line, left y axis) and ensemble mean standard deviation (dashed line, right y axis) for two-month lead time prediction of standardized all-Ethiopian rainfall anomalies (Fig. 1) for 1990–2002 using the RV approach as a function of the total number of predictors (abscissa) used in developing models as described in section 2b (step 2). Vertical bars show corresponding final ensemble size (values given on top of each bar) for each simulation determined by step 3 (section 2b). The first final model ensemble (2) was developed from a 5-member predictor set, while the last final ensemble (11) was developed from 20 initial predictors.

the northern Indian Ocean during March can lead to enhanced (reduced) Ethiopian seasonal rainfall during JJAS.

Predictors in the horizontal wind and vertical velocity fields include 500–200-hPa zonal and meridional winds and vertical velocity. In the upper troposphere, the product of 200 hPa  $u$  and  $w$  ( $uw_{200}$ ) during March possessed strong negative correlation with Kiremt rainfall over northeastern Arabian Sea. This correlation arises from the passage of a ridge–trough system that perturbs a normally weak zonal and subsiding flow over the northern Indian Ocean. The negative correlation indicates that increasing westerlies/ascending motion (easterlies/descending motion) during March likely lead to reduced (enhanced) Ethiopian Kiremt rainfall. Nicholson (2014) also found strong correlation between Horn of Africa rainfall and zonal wind at 200 hPa for May. The 200-hPa vertical velocity association with Ethiopian rainfall strengthens at 300 hPa with its footprint extending farther west to the western Indian Ocean (Fig. 3, middle).

In the middle troposphere, the zonal and meridional wind for March showed strong correlation with Ethiopian rainfall off the coast of Kenya (Fig. 3, bottom). The predictability signal was obtained from both the zonal and meridional components, with the zonal and meridional wind interaction shifting the region of influence near the African coast compared to Fig. 2 (middle). The

negative correlation reflects the simultaneous strengthening of northerlies (southerlies) and westerlies (easterlies) during March over the western Indian Ocean can enhance Ethiopian rainfall during JJAS.

Much of Ethiopian Kiremt rainfall predictability from SST is generated from tropical Pacific SST, where there are strong statistically significant correlations for both the 1970–90 and 1970–99 periods. Previous studies also found the northern tropical Pacific to be a source of Ethiopian rainfall predictability (e.g., Block and Rajagopalan 2007; Korecha and Barnston 2007; Nicholson 2014). After demonstrating the circulation patterns associated with the correlations, a physical basis for retention of these predictors is confirmed and a total of 20 predictors were selected for modeling of all-Ethiopian Kiremt rainfall (Table 1). These predictors were used for both RV and LOOCV approaches.

#### b. Retroactive verification of all Ethiopian predictions using March predictors

For RV of all-Ethiopian Kiremt rainfall predictions, 20 individual regression models (with intercepts) initially were developed for the 1970–89 training period, following the multimodel ensemble construction (step 2) described in section 2b. The numbers of model coefficients (not counting the intercepts) stepped into the initial regression models were 2 (5% of models), 3 (65%), and 4 (30%). The final multimodel ensemble set (11 models) was identified as described in step 3 (section 2b) using model statistics for the 1970–89 training period (regression coefficients and residuals, numbers of coefficients, multiple  $R^2$ , VIF, and correlations between predictors).

Model performance is summarized in Fig. 5, which displays the correlations between model-predicted and observed standardized all-Ethiopian Kiremt rainfall anomalies for the training (1970–89) and independent verification (1990–2002) periods for the initial 20 and final 11 models (Fig. 5b). All 20 models reproduced well the observations for the training period ( $R^2 > 0.77$ ), but their cross-validated skill decreased significantly for the independent 1990–2002 period, with 55% of models having less than 50% explained variance and RMSE increasing by 39%–193% (Fig. 5a). Figure 5b shows the regression equations for the final 11 selected models that satisfied the assumptions of linear regression technique objectively and attained statistical significance of 5% (section 2b). All models have a maximum of four non-intercept coefficients. With the exception of the  $vw_{925}$  predictor, which appeared in 7 of the 11 final models as the second parameter, the other predictors are fairly uniformly distributed with  $\leq 4$  out of 11 frequency of occurrence in the ensemble. Comparison of Figs. 5a and 6

TABLE 3. Characteristics of regression models developed using a LOOCV approach (section 2b) for 2-month lead time prediction of standardized Kiremt (JJAS) all-Ethiopia rainfall anomalies (Fig. 1) for 1970–2002, based on atmospheric and SST predictors observed in March. The columns under the heading number of models having indicated number of coefficients (3–6) exclude the intercept. Models for which the predictors were strongly intercorrelated (i.e.,  $|r_{px}| \geq 0.5$ , where  $r_{px}$  is the correlation between any two predictors in a model) were removed. The column under the heading number of models significant at 5% level gives results of tests for 1) normality, serial independence, and constant variance according to the Shapiro–Wilk, Durbin–Watson, and Breusch–Pagan tests, respectively, and 2) individual model coefficients and overall model fit according to an  $F$  test. The column under final ensemble model size gives the number of statistically significant (at 5% level) models selected having a max of five coefficients excluding the intercept.

Year	No. of models having indicated number of coefficients (3–6)				No. of models removed ( $ r_{px}  \geq 0.5$ )	No. of models significant at 5% level	Final ensemble model size
	3	4	5	6 (excluded)			
1970	—	5	12	3	10	17	8
1971	—	9	11	—	3	20	14
1972	2	8	10	—	7	18	10
1973	—	3	14	3	8	16	7
1974	—	1	14	5	5	14	8
1975	—	5	12	3	7	14	8
1976	—	7	12	1	5	18	9
1977	—	5	14	1	7	18	9
1978	1	7	10	2	5	17	11
1979	—	8	11	1	4	19	12
1980	—	11	9	—	4	20	12
1981	—	5	12	3	5	16	11
1982	—	8	8	4	8	15	7
1983	—	3	13	4	5	16	8
1984	—	9	9	2	7	15	9
1985	—	4	14	2	6	14	9
1986	—	7	12	1	4	19	11
1987	—	9	9	2	6	17	9
1988	—	12	7	1	4	18	11
1989	—	6	14	—	3	19	12
1990	—	7	11	2	8	18	11
1991	—	6	12	2	6	18	11
1992	—	8	11	1	6	19	9
1993	—	10	9	1	5	17	11
1994	—	7	12	1	7	17	12
1995	—	8	12	—	4	15	10
1996	2	9	6	3	7	17	11
1997	—	8	9	3	8	17	10
1998	—	5	14	1	7	19	8
1999	—	5	15	—	6	20	10
2000	—	7	12	1	8	19	10
2001	—	8	10	2	4	18	13
2002	—	10	7	3	4	12	8

shows that significant prediction improvement resulted from averaging the ensemble members, consistent with findings of Krishnamurti et al. (2000) and Bohn et al. (2010) for dynamic multimodel ensemble forecasting. This benefit primarily arises from cancelling (reinforcing) disagreements among (common features between) ensemble members (Wilks 2006, p. 235).

Figure 6 and Table 2 document the RV performance using the mean and median of the final 11-member model ensemble. To highlight the skill advantages that arise from the ensemble averaging, the median column in Table 2 gives the median values of the statistics under the observed versus predicted column but obtained for

the final 11 regression models individually prior to ensemble averaging. Pearson's correlations between observed and predicted ensemble means (+0.84) and medians (+0.79) for the 1990–2002 verification period have ASLs of <1% according to a nonparametric bootstrap test (Fig. 6). The Spearman's rank correlation increased slightly for the ensemble mean (+0.87) but remained nearly the same for the median (+0.78), both with similarly determined ASLs  $\leq 1\%$ , indicating strong monotonic association between the ranking of ensemble forecasts and observations. Predictions have similar variability as the observations, as evidenced by their standard deviations (Table 2, Fig. 6 inset). Absolute

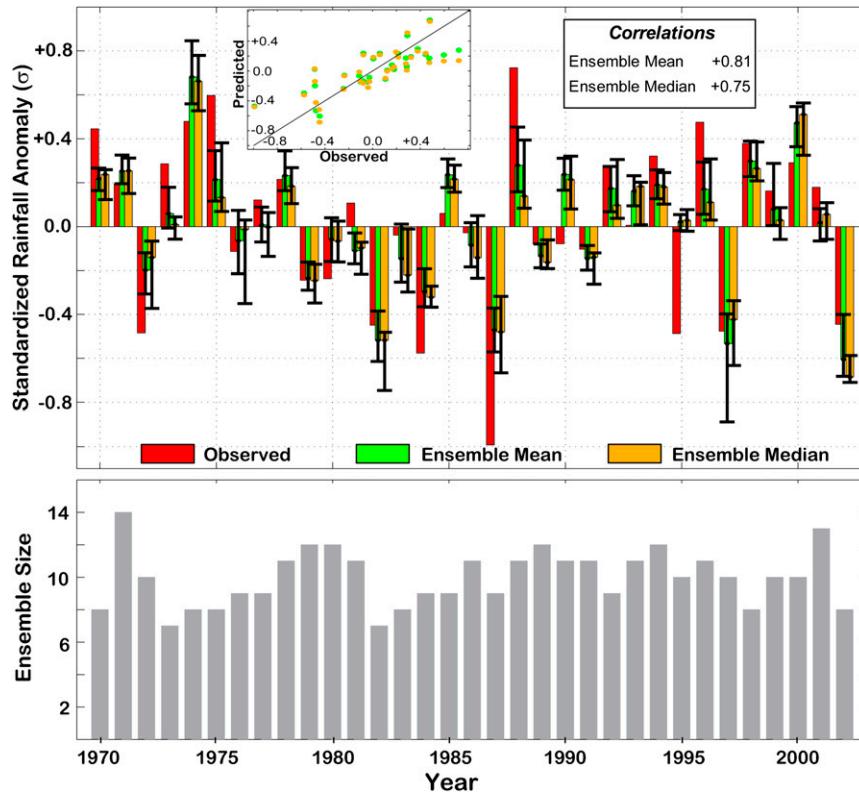


FIG. 9. Two-month lead time prediction of standardized all-Ethiopian Kiremt (JJAS) rainfall anomalies (Fig. 1, 100 stations for 1970–99, dots; 52 stations for 2000–02, squares) for 1970–2002 using a LOOCV approach that was initiated from average March atmospheric and SST conditions. (top) Time series of observed (red) vs mean (green) and median (orange) of ensemble model–predicted Kiremt rainfall anomalies. Inset shows corresponding scatter diagram of observed vs ensemble–predicted mean (green) and median (orange). (bottom) Number of final ensemble model members selected. Black vertical bars in top panel give 95% confidence intervals for the final multimodel–predicted values, computed by applying the BCa bootstrap method described in section 2c.

prediction error statistics (MAE and RMSE) are 46%–59% of the observed standard deviation. All relative error measures in Table 2 indicate that the model ensemble mean satisfactorily reproduced the observed standardized rainfall anomalies. The prediction showed conventional skill score ( $SS_{\text{Clim}}$ ; Wilks 2006, p. 281) improvement over 1990–2002 climatology (62% MSE reduction) and persistence (87%). For tercile predictions (below, near, and above normal), the ranked probability skill score ( $RPSS_{\text{Clim}}$ ) exceeded that of climatology (+0.55). Further, by comparing the ensemble mean statistics with the median average values for the individual regression model counterparts, there emerges a clear increase in skill for all statistics for the ensemble mean: increased prediction variance, reduced absolute errors, and much increased goodness-of-fit statistics.

Figure 7 documents the reliability of the above statistics in the context of empirically derived frequency

distributions of bootstrapped samples and corresponding confidence intervals. Because the  $RPSS_{\text{Clim}}$  [skewness  $b_1 = -0.3$  and kurtosis  $b_2 = +3.2$ , defined in Tamhane and Dunlop (2000, p. 118) and computed from sample replicates used in Fig. 7] and (especially) the Pearson's correlation  $r$  ( $b_1 = -0.7$ ,  $b_2 = +4.3$ ) are negatively skewed and leptokurtic compared to the standard normal distribution, the BCa method makes appropriate adjustments to correct the skewness bias in the estimation of their confidence intervals. The modified index of agreement  $d_1$  also is negatively skewed and leptokurtic compared to the standard normal distribution ( $b_1 = -0.7$ ,  $b_2 = +3.6$ ), while MAE is positively skewed ( $b_1 = 0.3$ ) and mesokurtic ( $b_2 = +3.0$ ).

To examine the cross-validated performance of the ensemble prediction as a function of ensemble sizes, the number of predictors, and hence initial ensemble model sizes, was varied from five to the maximum available

TABLE 4. Summary statistics and 95% confidence intervals for model verifications (section 2c) of 2-month lead time prediction of standardized Kiremt (JJAS) all-Ethiopian rainfall anomalies (Fig. 1) for 1970–2002 using a LOOCV approach (section 2b, steps 2 and 3), based on atmospheric and SST predictors observed in March. Ensemble models were developed by excluding a single verification year in turn (section 3c). Confidence intervals were computed by applying the BCa bootstrap method to 5000 verification statistic replicates obtained by bootstrapping the observed and predicted rainfall anomalies (section 2c). The median column gives the median values of the statistics obtained for the individual regression estimates without pooling, and the mean column under the bootstrap heading contains the average of 5000 bootstrap statistic replicates. Asterisks indicate values were computed using the least common number of models available in all verification years (i.e., 7; Table 3).

Statistic	Median	Predicted vs obs values	Bootstrap		
			95% confidence intervals		Mean
			Lower limit	Upper limit	
Standard deviation (obs, $\sigma$ )	—	+0.39	+0.32	+0.50	+0.38
Standard deviation (predicted, $\sigma$ )	+0.15	+0.29	+0.24	+0.38	+0.28
Mean absolute error (MAE; $\sigma$ )	+0.22	+0.18	+0.14	+0.24	+0.18
Root-mean-square error (RMSE; $\sigma$ )	+0.27	+0.23	+0.18	+0.29	+0.22
Skill score (vs climatology; $SS_{\text{Clim}}$ )	+0.49*	+0.65	+0.50	+0.74	+0.63
Skill score (vs persistence; $SS_{\text{Pers}}$ )	+0.80*	+0.86	+0.83	+0.92	+0.81
RPSS (vs climatology; $RPSS_{\text{Clim}}$ )	-0.02*	+0.45	+0.22	+0.62	+0.45
Pearson's correlation ( $r$ )	+0.72*	+0.81	+0.66	+0.86	+0.81
Index of agreement ( $d_2$ )	+0.83*	+0.88	+0.80	+0.92	+0.87
Modified index of agreement ( $d_1$ )	+0.61*	+0.66	+0.57	+0.74	+0.66
Modified coefficient of efficiency ( $E_1$ )	+0.27*	+0.41	+0.26	+0.53	+0.40

number (20) and the ensemble performance was assessed following the model development and selection procedures described in section 2b (steps 2 and 3). Using cross-validated correlation between observed values and predicted ensemble mean as a measure of model

performance, Fig. 8 shows that as the ensemble size increases, the cross-validated correlation skill also increases from as low as +0.62 for a two-member model ensemble to the maximum of +0.84 for the final 11-member ensemble model set. At the same time, the

TABLE 5. Predictors selected for the prediction of August total rainfall for Addis Ababa (Fig. 1) using a LOOCV approach (section 2b). Predictors were selected based on their high correlations with rainfall, with predictor–predictor correlations being statistically significant with ASLs  $\leq 5\%$  according to a nonparametric bootstrap test.

Predictor description	Symbol	Location	Correlation ( $\times 100$ )	
			1970–90	1970–99
1 Air temperature at 100 hPa	T100	Middle East (30°–35°N, 45°–50°E)	-52	-48
2 Air temperature at 400 hPa	T400	Southern Atlantic Ocean (20°–25°S, 25°–30°W)	+59	+55
3 Meridional moisture flux at 500 hPa	qu500	Southern Africa (20°–25°N, 25°–30°E)	-49	-46
4 Meridional moisture flux at 850 hPa	qv850	Northern Arabian Sea (20°–25°N, 55°–60°E)	-55	-58
5 Sea surface temperature	SST1	Bay of Bengal (11°–16°N, 88°–93°E)	+55	+49
6 Sea surface temperature	SST2	Southeastern Indian Ocean (46°–51°S, 128°–133°E)	+51	+56
7 Meridional temperature flux at 300 hPa	Tv300	Central Indian Ocean (5°–10°S, 70°–75°E)	-48	-49
8 Zonal/meridional wind flux at 500 hPa	uw500	Gulf of Guinea (0°–5°N, 0°–5°W)	+63	+58
9 Zonal/vertical wind flux at 200 hPa	uw200	Eastern Indian Ocean (5°–10°N, 80°–85°E)	-60	-47
10 Zonal/vertical wind flux at 850 hPa	uw850a	Western Indian Ocean (5°–10°N, 55°–60°E)	+51	+53
11 Zonal/vertical wind flux at 850 hPa	uw850b	Asia (40°–45°N, 60°–65°E)	-67	-56
12 Zonal wind at 400 hPa	u400	South Atlantic Ocean (15°–20°S, 25°–30°W)	-45	-46
13 Meridional/vertical wind flux at 150 hPa	vw150	Central Indian Ocean (5°–10°S, 80°–85°E)	-53	-48
14 Meridional/vertical wind flux at 200 hPa	vw200a	Southern Indian Ocean (35°–40°S, 15°–20°E)	+52	+47
15 Meridional/vertical wind flux at 200 hPa	vw200b	Southern Atlantic Ocean (35°–40°S, 10°–15°W)	+58	+50
16 Meridional/vertical wind flux at 600 hPa	vw600	Ethiopia (5°–10°N, 40°–45°E)	+50	+46
17 Meridional wind at 250 hPa	v250	Central Indian Ocean (0°–5°S, 70°–75°E)	-45	-52
18 Vertical velocity at 300 hPa	w300	Asia (40°–45°N, 65°–70°E)	+61	+60
19 Vertical velocity at 400 hPa	w400	Western Indian Ocean (0°–5°N, 55°–60°E)	-45	-49
20 Vertical velocity at 500 hPa	w500	South Atlantic Ocean (10°–15°S, 0°–5°W)	-54	-51

TABLE 6. Predictors selected for the prediction of August total rainfall for Combolcha (Fig. 1) using a LOOCV approach (section 2b). Predictors were selected based on their high correlations with rainfall, with predictor–predictor correlations being statistically significant with ASLs  $\leq 5\%$  according to a nonparametric bootstrap test.

	Predictor description	Symbol	Location	Correlation ( $\times 100$ )	
				1970–90	1970–99
1	Meridional moisture flux at 925 hPa	qv925	Central Indian Ocean (0°–5°N, 85°–90°E)	+66	+65
2	Sea surface temperature	SST1	Southern Indian Ocean (51°–56°S, 171°–176°E)	+45	+49
3	Sea surface temperature	SST2	South Pacific Ocean (50°–55°S, 120°–125°W)	–65	–57
4	Sea surface temperature	SST3	East Pacific Ocean (15°–20°S, 80°–85°W)	–45	–45
5	Zonal/meridional wind flux at 500 hPa	uv500	South Atlantic Ocean (10°–15°S, 10°–15°W)	–56	–57
6	Zonal/meridional wind flux at 600 hPa	uv600a	Southeastern Indian Ocean (10°–15°S, 55°–60°E)	–58	–52
7	Zonal/meridional wind flux at 600 hPa	uv600b	South Atlantic Ocean (10°–15°S, 10°–15°W)	–63	–60
8	Zonal/meridional wind flux at 925 hPa	uv925	Gulf of Guinea (0°–5°N, 0°–5°E)	+45	+54
9	Zonal/vertical wind flux at 150 hPa	uw150	Eastern Indian Ocean (30°–35°S, 65°–70°E)	–52	–52
10	Zonal/vertical wind flux at 200 hPa	uv200	Central Africa (0°–5°N, 25°–30°E)	+46	+42
11	Zonal/vertical wind flux at 500 hPa	uw500	East Africa (0°–5°N, 35°–40°E)	+46	+45
12	Zonal wind at 150 hPa	u150a	Central Africa (0°–5°N, 25°–30°E)	–50	–48
13	Zonal wind at 150 hPa	u150b	Central Africa (0°–5°N, 30°–35°E)	–48	–46
14	Zonal/vertical wind flux at 150 hPa averaged between 30° and 50°E	uw150x	Cross section (0°–10°S)	+45	+46
15	Meridional/vertical wind flux at 300 hPa	vw300	India (25°–30°N, 80°–85°E)	–47	–46
16	Meridional/vertical wind flux at 400 hPa	vw400a	Central Indian Ocean (0°–5°S, 75°–80°E)	–53	–48
17	Meridional/vertical wind flux at 400 hPa	vw400b	South Atlantic Ocean (10°–15°S, 5°–10°W)	+46	+46
18	Meridional/vertical wind flux at 500 hPa	vw500	Central Indian Ocean (0°–5°S, 80°–85°E)	–61	–53
19	Meridional wind at 1000 hPa	v1000	Central Indian Ocean (0°–5°N, 85°–90°E)	+64	+60
20	Vertical velocity at 250 hPa	w250	Mediterranean Sea (30°–35°N, 30°–35°E)	+46	+46

variance of the prediction also decreases as the ensemble size increases, as shown by the dashed line in Fig. 8. However, the ensemble standard deviation for the final ensemble is still higher than the observed standard deviation (Table 2). It can also be seen from the figure that an increase in the number of predictors may not increase the ensemble size, because some of the models associated with the added predictors do not make it to the final ensemble set as they fail to satisfy preset conditions of statistical significance and regression requirements. In general, diversifying the predictor set can increase the likelihood of capturing regional atmospheric predictive signals that improve the prediction of Kiremt rainfall.

### c. Leave-one-out cross validation of all Ethiopian predictions using March predictors

All-Ethiopian Kiremt rainfall predictability also was assessed for 1970–2002 using March predictors via the LOOCV approach described in section 2b. The procedures described in section 2b were followed for model construction (step 2) and final ensemble model member selection (step 3). After the passage of all statistical tests, 7–14 models were accepted as the final ensemble set from the initial 20 models (Table 2). In addition to the intercept, the initial models had 3–6 predictor coefficients, with the majority of the models having 5 coefficients. The maximum number of nonintercept

coefficients allowed in the final ensemble set was five, because reducing the number of regression parameters to four led to a single model for 1974 (Table 3) that did not satisfy the regression assumptions and statistical significance requirements.

The mean and median of the final ensemble model members were used as verification metrics for the forecasts. Figure 9 and Table 3 document the performance of the LOOCV approach for the predictions of standardized all-Ethiopian Kiremt rainfall anomalies. The Pearson's correlations between the observed and predicted ensemble mean and median respectively are +0.81 and +0.75 (ASLs  $< 1\%$  according to a nonparametric bootstrap test) for 1970–2002. The Spearman's rank correlation was +0.81 for the ensemble mean and +0.77 for the median.

Table 4 provides additional prediction verification statistics (section 2c) for the final multimodel ensemble shown in Table 3 (last column). The performance statistics under the median column were obtained from individual regression estimates. For verification statistics that require pairwise observed–predicted time series, only the least common number of models in the 33 verification years (7 for 1973 and 1982) was used so that no data pairs are excluded or repeatedly used in cross validating individual regression estimates. This issue arises in the LOOCV approach because the number of

TABLE 7. Predictors selected for the prediction of standardized JJAS northeastern Ethiopia (Fig. 1; 11 stations for 1970–99, stars; 7 stations for 2000–02, squares) using a LOOCV approach (section 2b). Predictors were selected based on their high correlations with rainfall, with predictand–predictor correlations being statistically significant with ASLs  $\leq 5\%$  according to a nonparametric bootstrap test.

	Predictor description	Symbol	Location	Correlation ( $\times 100$ )	
				1970–90	1970–99
1	Air temperature at 100 hPa	T100a	Russia (45°–50°N, 50°–55°E)	–56	–52
2	Air Temperature at 100 hPa	T100b	Southern Indian Ocean (20°–25°S, 75°–80°E)	–56	–49
3	Air Temperature at 150 hPa	T150	India (20°–25°N, 80°–85°E)	+53	+48
4	Meridional moisture flux at 500 hPa	qv500	Asia (35°–40°N, 65°–70°E)	+62	+55
5	Zonal/meridional wind flux at 100 hPa	uv100a	Southern Africa (20°–25°S, 25°–30°E)	+60	+59
6	Zonal/meridional wind flux at 100 hPa	uv100b	Mozambique Channel (20°–25°S, 35°–40°E)	+48	+51
7	Zonal/meridional wind flux at 500 hPa	uv500	Western Indian Ocean (5°–10°S, 40°–45°E)	–73	–67
8	Zonal/meridional wind flux at 600 hPa	uv600	South Atlantic Ocean (10°–15°S, 10°–15°W)	–59	–53
9	Zonal/vertical velocity flux at 100 hPa	uw100	Somali coast (0°–5°N, 45°–50°E)	–59	–53
10	Zonal wind at 100 hPa	u100	Southern Europe (45°–50°N, 25°–30°E)	+45	+46
11	Zonal wind at 400 hPa	u400	South Atlantic (20°–25°S, 10°–15°W)	–48	–49
12	Zonal/vertical wind flux at 150 hPa averaged between 5° and 10°N	uw150x	Cross section (65°–70°E)	–49	–44
13	Meridional/vertical wind flux at 100 hPa	vw100	Asia (45°–50°N, 85°–90°E)	–57	–63
14	Meridional/vertical wind flux at 300 hPa	vw300	India (25°–30°N, 75°–80°E)	–51	–58
15	Meridional/vertical wind flux at 700 hPa	vw700	South Atlantic Ocean (5°–10°S, 15°–20°W)	+51	+49
16	Vertical velocity at 100 hPa	w100a	Eastern Europe (35°–40°N, 45°–50°E)	–49	–49
17	Vertical velocity at 100 hPa	w100b	Central Indian Ocean (5°–10°N, 55°–60°E)	–58	–49
18	Vertical velocity at 100 hPa	w100c	Asia (45°–50°N, 80°–85°E)	+48	+63
19	Vertical velocity at 150 hPa	w150	South Atlantic Ocean (15°–20°N, 20°–25°W)	+61	+51
20	Vertical velocity at 700 hPa	w700	West Africa (5°–10°N, 0°–5°W)	–67	–63

final models (and hence the number of predicted values) varies from year to year (Table 3). To be as objective as possible in selecting the least common number of models, seven models were randomly sampled (without replacement) for all verification years with larger final model ensembles and were used for cross validation. As seen from Table 4, the LOOCV ensemble mean prediction proved more skillful than the individual regression estimates for all verification measures including prediction variance (standard deviation),  $SS_{\text{Clim}}$ ,  $RPSS_{\text{Clim}}$ ,  $r$ , and  $E_1$ . In particular, the  $RPSS_{\text{Clim}}$  skill for the individual regression estimates was far inferior to the skill obtained for the ensemble mean. To ensure that this low skill was not an artifact of a one-time sampling issue,  $RPSS_{\text{Clim}}$  was computed for 10 more samples obtained for different initializations of the random number generator. In all cases, the median of the  $RPSS_{\text{Clim}}$  for the individual regression estimates was less than 0, indicating worse than climatology skill.

The prediction variability for the ensemble mean is close to the observed variation, with the prediction standard deviation being close to 74% of the standard deviation of the observations (Table 4). The MAE (0.18 standardized units) and RMSE (0.23 standardized units) are smaller than the observed standard deviation (0.39 standardized units). The LOOCV prediction approach showed substantial reduction in MSE compared to

climatology (65%) and persistence (86%). Generally, the skill of the LOOCV remains close to the skill obtained for the RV approach and provided skillful all-Ethiopian Kiremt rainfall forecasts with the average regional atmospheric and global SST conditions observed in March (e.g.,  $r = +0.81$ ,  $d_1 = 66\%$ , and  $RPSS = 45\%$ ).

#### d. Leave-one-out cross validation for local and regional predictions using March predictors

The LOOCV approach was applied on finer temporal and spatial scales for the prediction of August monthly rainfall totals at Addis Ababa and Combolcha (Fig. 1) and standardized JJAS rainfall anomalies for northeastern Ethiopia based on atmospheric and SST predictors observed in March. Tables 5–7 list the 20-predictor sets used for local and regional predictions. Following previous studies (e.g., Gissila et al. 2004; Block and Rajagopalan 2007; Korecha and Barnston 2007; Nicholson 2014), these predictors were selected based on a careful assessment of maps of predictand–predictor correlations. All selected predictors have statistically significant correlations with ASLs  $\leq 5\%$  according to a nonparametric bootstrap test for both the 1970–90 and 1970–99 periods. The predictors for Addis Ababa (Table 5) include upper tropospheric air temperature, meridional moisture flux over the northern Arabian Sea, SST over the Bay of

TABLE 8. Characteristics of regression models developed using a LOOCV approach (section 2b, steps 2 and 3) approach for 3-month lead time localized prediction of August monthly total rainfall at Addis Ababa (Bole International Airport; Fig. 1) for 1970–99, based on atmospheric and SST predictors observed in March. The columns under the heading number of models having indicated number of coefficients (3–7) exclude the intercept. Models for which the predictors were strongly intercorrelated (i.e.,  $|r_{px}| \geq 0.5$ , where  $r_{px}$  is the correlation between any two predictors in a model) were removed. The column under the heading number of models significant at 5% level gives results of tests for 1) normality, serial independence, and constant variance according to the Shapiro–Wilk, Durbin–Watson, and Breusch–Pagan tests, respectively, and 2) individual model coefficients and overall model fit according to an  $F$  test. The “Final ensemble model size” column gives the number of statistically significant (at 5% level) models selected with a max of five nonintercept coefficients.

Year	No. of models having indicated number of coefficients (3–7)				No. of models removed ( $ r_{px}  \geq 0.5$ )	No. of models significant at 5% level	Final ensemble model size
	3	4	5	$\geq 6$ (excluded)			
1970	—	5	6	9	5	11	7
1971	—	7	6	7	2	13	10
1972	9	8	2	1	3	19	14
1973	—	8	7	5	5	15	9
1974	3	9	7	1	3	17	12
1975	2	10	6	2	2	18	12
1976	—	9	7	4	5	16	13
1977	2	6	7	5	4	14	8
1978	2	9	9	—	1	16	13
1979	3	5	5	7	5	13	10
1980	—	7	8	5	5	15	9
1981	—	6	7	7	4	13	8
1982	2	5	8	5	3	13	9
1983	—	8	9	3	5	16	8
1984	—	7	6	7	5	9	5
1985	—	6	9	5	3	14	11
1986	1	8	5	6	3	14	11
1987	2	5	11	2	1	16	12
1988	2	6	8	4	4	16	10
1989	1	6	9	4	2	16	12
1990	—	6	10	4	4	12	7
1991	3	5	7	5	2	14	10
1992	1	6	8	5	3	15	12
1993	1	12	4	3	1	10	9
1994	1	7	6	6	2	14	11
1995	1	7	9	3	2	17	12
1996	3	6	5	6	3	14	10
1997	4	5	10	1	3	19	11
1998	1	7	7	5	4	14	11
1999	1	5	8	6	4	14	7

Bengal, and products of zonal and meridional winds with vertical velocities across the southern oceans, most of which are strongly correlated with the performance of the Belg rains in Ethiopia and the development of the monsoon systems in summer. For Combolcha (Table 6), lower-tropospheric meridional wind and specific humidity over the central Indian Ocean and SST over the southern Pacific possessed strong predictive signals, while for regional predictions (Table 7) midtropospheric zonal and meridional wind over the western Indian Ocean and vertical velocity over West Africa and/or the Gulf of Guinea in March are major predictors of JJAS rainfall anomalies for the northeastern Ethiopia.

The procedures presented in section 2b (steps 2 and 3) were followed for model construction and final ensemble model member selection for local predictions for Addis Ababa and Combolcha for 1970–99 and regional prediction for northeastern Ethiopia for 1970–2002. In all cases, models with a maximum of only five nonintercept coefficients were allowed in the final ensemble set.

For Addis Ababa, the final ensemble models range from 5 for 1984 to 14 for 1972, with the majority of the models having 4–5 nonintercept coefficients (Table 8). Training period degrees-of-freedom-adjusted  $R^2$  (coefficient of multiple determination) varied from 56% to 88% (from 61% to 91%) for all verification years. The

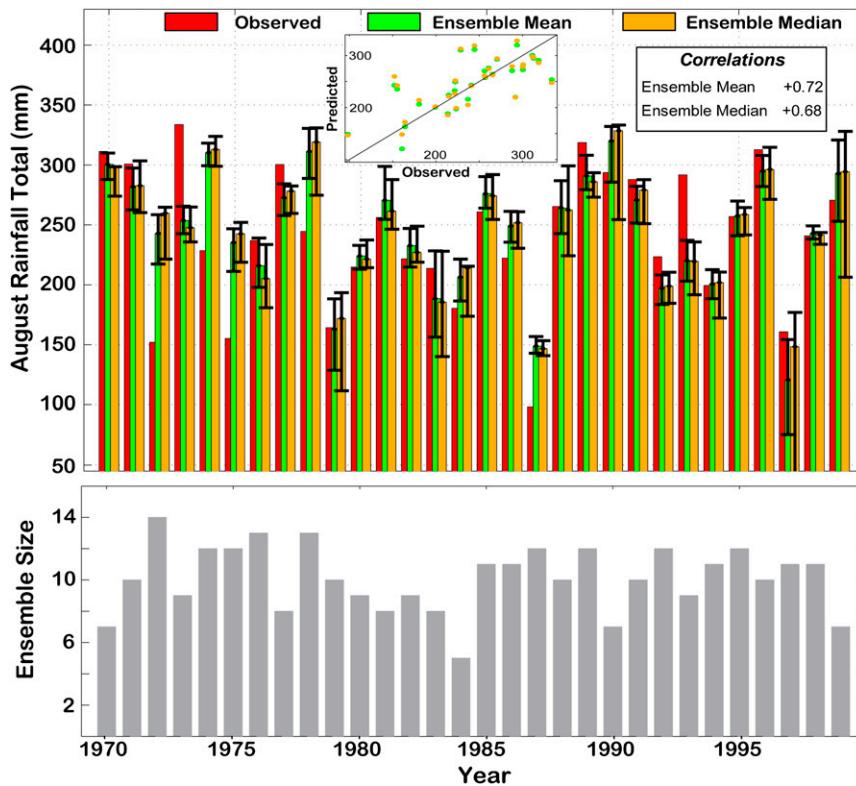


FIG. 10. Three-month lead time localized prediction of August monthly total rainfall at Addis Ababa (Bole International Airport; Fig. 1) for 1970–99 using a LOOCV approach that was initiated from average April atmospheric and SST conditions. (top) Time series of observed (red) vs mean (green) and median (orange) of ensemble model–predicted August rainfall total amounts. Inset shows corresponding scatter diagram of observed vs ensemble–predicted mean (green) and median (orange). (bottom) Number of final ensemble model members selected. Black vertical bars in top panel give 95% confidence intervals for the final multimodel–predicted values, computed by applying the BCa bootstrap method described in section 2c.

LOOCV ensemble mean overestimated the three driest years (1987, 1972, and 1975) by 51–91 mm and underestimated the wettest three years by 18–80 mm (Fig. 10). However, differences between predicted and observed rainfall amounts were less than 50 (33) mm for 77% (73%) of the 1970–99 verification years. The observed and LOOCV ensemble means for 1970–99 differed by <5 mm, with the year-to-year variability of the ensemble mean being 86% of the observed standard deviation. The  $RPSS_{Clim}$  and all MSE-based skill scores ( $SS_{Clim}$ ,  $SS_{Pers}$ ,  $E_1$ ,  $d_1$ , and  $d_2$ ) showed improved forecast skill relative to the 1970–99 climatology and persistence (Table 9). The Pearson’s correlations between observed and predicted ensemble mean and median respectively are +0.72 and +0.68 (ASLs <1% according to a conventional nonparametric bootstrap test), which respectively decreased to +0.64 and +0.58 when the maximum number of nonintercept coefficients was reduced to four. In the latter case, however, the final ensemble sizes were small, ranging from 3 or 4 models for

many verification years to 12 models for 1972, which can affect the quantification of prediction uncertainty for years with small ensemble sizes. Compared to the median values of individual regression estimates, there is a substantial improvement of the ensemble averaging as evidenced by a smaller predicted-minus-observed mean difference, higher prediction variance, and smaller MAE and RMSE values. In addition, the ensemble mean provided higher skill values (Table 8) for all verification metrics including  $RPSS_{Clim}$ ,  $SS_{Clim}$ , and  $E_1$  for five randomly selected model ensembles for all years except 1984, which only had five ensemble members.

For the prediction of August rainfall total for Combolcha, from 7 (for 1974) to 17 (for 1986) final ensemble members were selected from the initial 20 models (Table 10). In addition to the intercept, the initial models stepped in 3–6 predictors, with the majority of the models having four regression coefficients. The LOOCV ensemble mean (median) forecasts (Fig. 11) for the final ensembles produced an observed–predicted Pearson’s

TABLE 9. As in Table 3, but for 3-month lead time localized prediction of August monthly total rainfall at Addis Ababa (Bole International Airport; Fig. 1) for 1970–99, based on atmospheric and SST predictors observed in March. Asterisks indicate shown values were computed using the least common number of models available for all verification years (i.e., 5; Table 5).

Statistic	Median	Obs statistic value	Bootstrap		Mean
			Confidence interval		
			Lower limit	Upper limit	
Obs mean (mm)	—	240.6	218.8	260.2	240.5
Predicted mean (mm)	+251.3	245.2	226.9	261.6	245.3
Standard deviation (obs, mm)	—	58.1	47.8	75.0	56.7
Standard deviation (predicted, mm)	+25.6	49.8	39.5	64.9	48.5
Mean absolute error (MAE; mm)	+41.2	30.4	22.1	41.3	30.4
Root-mean-square error (RMSE; mm)	+52.5	40.5	30.4	51.8	40.0
Skill score (climatology; $SS_{Clim}$ )	+0.15*	0.50	0.04	0.76	0.46
Skill score (persistence; $SS_{Pers}$ )	+0.73*	0.84	0.74	0.93	0.79
RPSS (climatology; $RPSS_{Clim}$ )	-0.13*	0.40	0.02	0.65	0.39
Correlation ( $r$ )	+0.62*	0.72	0.49	0.85	0.72
Index of agreement ( $d_2$ )	+0.78*	0.84	0.69	0.92	0.83
Modified index of agreement ( $d_1$ )	+0.57*	0.65	0.52	0.76	0.64
Modified coefficient of efficiency ( $E_1$ )	+0.12*	0.35	0.08	0.54	0.33

correlation of +0.68 (+0.63), both with  $\leq 1\%$  ASLs according to a nonparametric bootstrap test. The Spearman's rank correlation decreased for both the ensemble mean (+0.63) and median (+0.56), indicating a reduced monotonic association between ensemble-mean predicted and observed time series. Reducing the maximum number of coefficients from five to four did not change the correlation skills, but lowered the ensemble sizes to between 5 (for many verification years) and 12 for 1996. The observed and prediction ensemble means have comparable values (Table 11), but the predicted year-to-year variability was 19% lower than the observed standard deviation. The prediction outperformed climatology and persistence, with a 45% (76%) reduction of the MSE obtained from 1970–99 climatology (persistence). Note that  $RPSS_{Clim}$  and  $E_1$  are not statistically different from zero. Despite the ensemble's poor  $RPSS_{Clim}$  and  $E_1$  skill values, the ensemble averaging showed modest improvements over the median values of individual regression estimates (excluding  $RPSS_{Clim}$  and  $E_1$ ), with percentage skill increases of 2%–37% for relative error measures, and reductions in absolute error measures of 5%–10%.

For regional JJAS rainfall prediction for northeastern Ethiopia, from 6 (for 1986) to 13 (1994) final model ensembles were selected. The majority of models again had 4–5 coefficients, but none of the final ensemble models had more than five nonintercept coefficients (Table 12). The Pearson's correlations between observed and predicted ensemble mean and median respectively are +0.80 and +0.75 (Fig. 12; ASLs  $< 1\%$  according to a conventional nonparametric bootstrap test), while the Spearman's rank correlation counterpart

for the ensemble mean (median) is +0.76 (+0.75). The Pearson's correlation skill values for the ensemble mean and median slightly decreased respectively to +0.78 and 0.73 when the maximum number of nonintercept coefficient was reduced from five to four. However, the small change in correlation values was accompanied by significant reductions in ensemble sizes (4 and 5 for many years and 11 for 1982). The small ensemble sizes in turn affect the quantification of prediction uncertainty. Although the ensemble mean overestimated the driest four years and underestimated the wettest two anomalies, the prediction exhibited sufficiently high year-to-year variability (90% of the observed standard deviation). Unlike for Combolcha, the  $RPSS_{Clim}$  and all MSE-based skill scores ( $SS_{Clim}$ ,  $SS_{Pers}$ ,  $E_1$ ,  $d_1$ , and  $d_2$ ) showed improved forecast skill relative to the 1970–2002 climatology and persistence (Table 13). Compared to the median values of individual regression estimates, the ensemble mean showed 20%–22% decreases in MAE and RMSE and 7%–62% increases in  $SS_{Clim}$ ,  $SS_{Pers}$ ,  $r$ ,  $E_1$ ,  $d_1$ , and  $d_2$  for six randomly selected model ensembles for all years except for 1986, which had only six ensemble model members. Thus, the ensemble-based statistical technique provided substantively high skill scores for regional prediction of JJAS standardized rainfall anomalies two months in advance of the onset of Kiremt in northeastern Ethiopia.

#### 4. Summary and discussion

This study evaluates the predictability of Ethiopian Kiremt monthly to seasonal rainfall at local, regional, and national levels based on the climate system

TABLE 10. As in Table 4, but for 3-month lead time localized prediction of August monthly total rainfall at Combolcha (Fig. 1) for 1970–99, based on atmospheric and SST predictors observed in March. The max number of coefficients excluding the intercept stepped in was 6, and all models with six coefficients were excluded from the final ensemble models.

Year	No. of models having indicated number of coefficients (3–6)				No. of models removed ( $ r_{px}  \geq 0.5$ )	No. of models significant at 5% level	Final ensemble model size
	3	4	5	6 (excluded)			
1970	1	12	6	1	6	18	10
1971	1	12	6	1	6	18	10
1972	—	6	12	2	2	15	10
1973	—	12	8	—	6	17	8
1974	1	11	8	—	9	18	7
1975	1	11	8	—	3	20	12
1976	1	13	5	1	4	18	12
1977	—	5	14	1	4	18	12
1978	1	7	12	—	5	20	12
1979	—	16	4	—	5	19	11
1980	3	17	—	—	10	19	8
1981	—	10	9	1	4	18	11
1982	1	15	4	—	5	19	11
1983	1	13	6	—	5	19	11
1984	1	6	12	1	4	15	12
1985	1	13	6	—	4	18	13
1986	—	13	7	—	1	20	17
1987	—	12	8	—	1	19	16
1988	1	12	6	1	5	19	12
1989	1	14	5	—	6	19	9
1990	3	6	11	—	5	20	12
1991	—	10	8	2	5	18	10
1992	1	12	6	1	4	18	12
1993	—	9	11	—	3	20	13
1994	—	10	10	—	7	19	10
1995	1	10	9	—	2	19	13
1996	3	17	—	—	4	20	12
1997	1	11	7	1	6	18	10
1998	—	9	11	—	3	18	12
1999	—	11	8	1	6	19	11

causation of Kiremt rainfall variability over Ethiopia. Time-scale analysis in Part I showed the contemporaneous linkages of Ethiopian rainfall variability with both atmospheric and sea surface temperature (SST) forcing during June–September (JJAS). The monsoon’s response to both slowly varying SST variability and higher-frequency regional atmospheric changes are found to be the primary climate system variability causation to assess Ethiopian seasonal and monthly rainfall predictability one to three months in advance.

A pool of predictors (20) deemed relevant to the rainfall was selected by careful assessments of a series of historical correlation maps that relate rainfall with individual regional atmospheric and global SST variables for March. Regression models were constructed using a forward stepwise model-fitting procedure, for which each selected predictor was specified as a first model parameter along with an intercept to ensure that a potential predictive signal uniquely associated with

a predictor is not unduly discarded. This produces an ensemble of models. Forecast skill was assessed using several verification metrics.

In the retroactive verification (RV) approach, the ensemble prediction for 1990–2002 reproduced well the observed all-Ethiopian Kiremt rainfall variability two months in advance, with a Pearson’s correlation [mean-square skill score over climatology ( $SS_{Clim}$ )] of +0.84 (62%). The leave-one-out cross verification (LOOCV) for 1970–2002, which is a fairer test of the predictive capability of the models, has a Pearson’s correlation ( $SS_{Clim}$ ) of +0.81 (65%). For probability forecasts of below normal, near normal, and above normal, the ensemble mean showed improvement compared to climatological forecasts, with a ranked probability skill score ( $RPSS_{Clim}$ ) of 0.45 for the LOOCV approach. Results of LOOCV-based localized predictions of August rainfall at Addis Ababa and Combolcha for 1970–99 showed that the predictions captured the relative

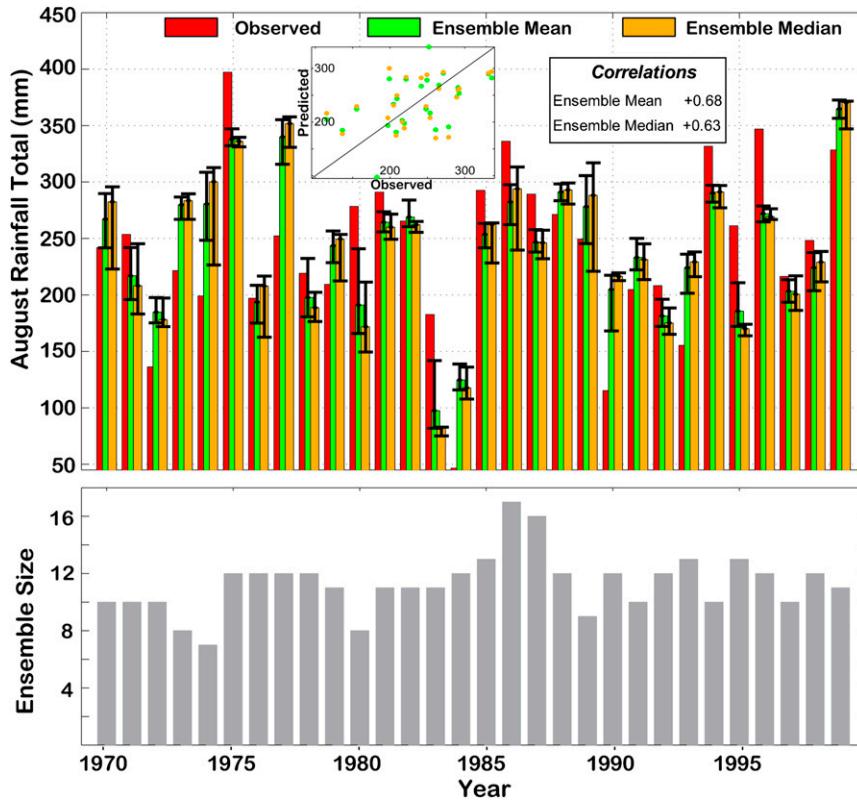


FIG. 11. As in Fig. 10, but for Combolcha (Fig. 1).

interannual variability well. The correlations between observed and ensemble means for Addis Ababa (+0.72) and Combolcha (+0.68) were high. Consistent with decreasing Spearman’s rank correlation, the  $RPSS_{Clim}$  skill is low especially for Combolcha (+0.12) and not

statistically significant at 5% level. For regional prediction of JJAS standardized rainfall for northeastern Ethiopia, strong linear correlation (+0.80) was found.

There is a marked improvement in the performance of the current empirical ensemble prediction method

TABLE 11. As in Table 3, but for 3-month lead time localized prediction of August monthly total rainfall at Combolcha (Fig. 1) for 1970–99, based on atmospheric and SST predictors observed in March. Asterisks indicate shown values were computed using the least common number of models available for all verification years (i.e., 7; Table 10).

Statistic	Median	Obs statistic value	Bootstrap		Mean
			Confidence interval		
			Lower limit	Upper limit	
Obs mean (mm)	—	241.5	215.2	266.6	241.7
Predicted mean (mm)	245.2	240.8	219.3	262.1	240.1
Standard deviation (obs, mm)	—	73.3	56.7	101.4	71.3
Standard deviation (predicted, mm)	24.0	59.6	47.3	79.0	58.2
Mean absolute error (MAE; mm)	49.0	46.7	37.5	56.2	46.6
Root-mean-square error (RMSE; mm)	59.0	53.5	44.9	62.8	53.3
Skill score (climatology; $SS_{Clim}$ )	0.33*	0.45	0.10	0.71	0.39
Skill score (persistence; $SS_{Pers}$ )	0.71*	0.76	0.57	0.87	0.75
RPSS (climatology; $RPSS_{Clim}$ )	-0.50*	0.12	-0.25	0.42	0.11
Correlation ( $r$ )	0.64*	0.68	0.49	0.83	0.68
Index of agreement ( $d_2$ )	0.79*	0.81	0.71	0.89	0.80
Modified index of agreement ( $d_1$ )	0.55*	0.54	0.43	0.65	0.53
Modified coefficient of efficiency ( $E_1$ )	0.11*	0.16	-0.12	0.37	0.13

TABLE 12. As in Table 3, but for a LOOCV approach (section 2b) for 2-month lead time prediction of standardized JJAS northeastern Ethiopia (Fig. 1; 11 stations for 1970–99, stars; 7 stations for 2000–02, collocated squares) rainfall anomalies for 1970–2002, based on atmospheric and SST predictors observed in March. The columns under the heading number of models having indicated number of coefficients (3–7) exclude the intercept. The column final ensemble model size gives the number of statistically significant (at 5% level) models selected with a max of five nonintercept coefficients.

Year	No. of models having indicated number of coefficients (3–7)				No. of models removed ( $ r_{pk}  \geq 0.5$ )	No. of models significant at 5% level	Final ensemble model size
	3	4	5	$\geq 6$ (excluded)			
1970	—	8	12	—	3	20	12
1971	1	7	10	2	4	18	11
1972	—	9	9	2	5	17	7
1973	—	12	6	2	1	18	11
1974	—	10	9	1	4	19	10
1975	—	7	7	6	8	14	7
1976	—	7	10	3	5	16	8
1977	—	9	8	3	6	17	8
1978	2	9	8	1	4	18	11
1979	—	9	9	2	4	17	10
1980	—	9	6	5	4	15	7
1981	—	9	8	3	4	17	11
1982	—	12	7	1	4	19	13
1983	—	9	9	2	6	17	8
1984	1	6	7	6	3	14	8
1985	—	7	9	4	2	16	9
1986	—	12	5	3	6	15	6
1987	—	10	9	1	4	19	10
1988	1	8	9	2	8	18	9
1989	1	13	5	1	2	18	11
1990	—	12	7	1	4	19	12
1991	—	11	7	2	3	18	10
1992	1	11	6	2	5	18	11
1993	1	7	9	3	6	17	10
1994	1	11	8	—	4	18	13
1995	—	8	7	5	5	14	8
1996	—	9	8	3	3	17	10
1997	—	8	11	1	4	18	10
1998	3	4	9	4	4	12	7
1999	—	8	10	2	4	16	8
2000	—	11	8	1	5	18	10
2001	—	8	8	4	6	16	9
2002	—	9	8	3	4	17	9

compared to prediction skill found in previous studies. For example, the correlation skill derived herein between predicted and observed standardized all-Ethiopian Kiremt rainfall anomalies was +0.81 (+0.84) for the LOOCV (RV) for models initiated from average March atmospheric and SST conditions. This can be compared to Korecha and Barnston (2007), who found a correlation skill of +0.64 (+0.51) between predicted and observed standardized all-Ethiopian Kiremt rainfall anomalies for LOOCV (RV) verification approach for models developed from March–May SSTs (for improvements of 27%–65%). For regional northwestern Ethiopia prediction, Block and Rajagopalan (2007) found cross-validated correlation (RPSS<sub>Clim</sub>) skill of +0.69 (+0.39), based on local polynomial regression models developed

using March–May predictors. The current ensemble prediction technique has about 16%–44% gain in explained variance compared to the corresponding  $R^2$  in these studies. In addition, the ensemble prediction provided longer lead times compared to the above studies (May versus March). However, Nicholson (2014) found a high correlation skill of +0.75 for LOOCV-based July–September rainfall predictions for the larger Horn of Africa region encompassing Ethiopia and Sudan. Although enlargement of the domain could have reduced the rainfall variability [e.g., negative 1996 rainfall anomaly for the larger Horn of Africa in Fig. 12 of Nicholson (2014) versus the locally wet all-Ethiopian Kiremt in Figs. 6 and 8] and affect the comparison of results, the current ensemble-based prediction still

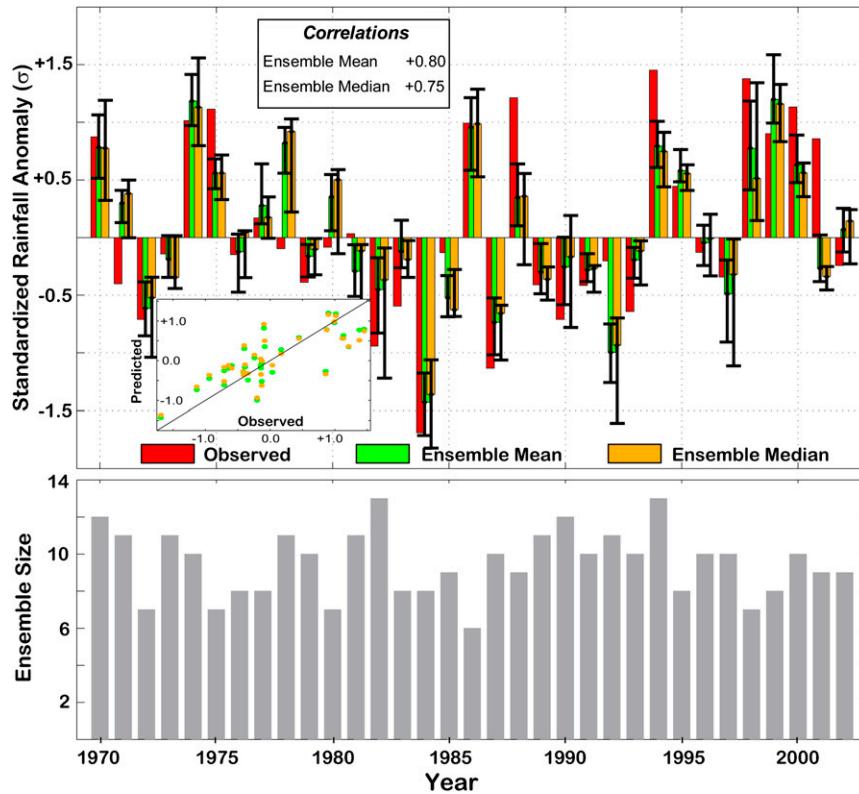


FIG. 12. As in Fig. 9, but for 2-month lead time prediction of standardized JJAS northeastern Ethiopia rainfall anomalies (Fig. 1; 11 stations for 1970–99, stars; 7 stations for 2000–02, collocated squares) for 1970–2002 using a LOOCV approach that was initiated from average March atmospheric and SST conditions.

shows improved skill at a longer lead times (March versus May) and finer spatial scales, with 8%–9% gain in explained variance compared to the corresponding  $R^2$  of 56% (for LOOCV) and 62% (for RV) in Nicholson (2014).

The RV (using a model developed for 1970–89 for upcoming season prediction) and LOOCV (using a model developed for all years up to the current year for upcoming season prediction) approaches can be applied

TABLE 13. As in Table 4, but for 2-month lead time prediction of standardized JJAS northeastern Ethiopia (Fig. 1, 11 stations for 1970–99, stars; 7 stations for 2000–02, collocated squares) rainfall anomalies for 1970–2002, based on atmospheric and SST predictors observed in March. Asterisks indicate values were computed using the least common number of models available in all verification years (i.e., 6; Table 12).

Statistic	Median	Predicted vs obs values	Bootstrap		
			95% confidence intervals		Mean
			Lower limit	Upper limit	
Standard deviation (obs, $\sigma$ )	—	+0.79	+0.67	+0.98	+0.78
Standard deviation (predicted, $\sigma$ )	+0.35	+0.63	+0.53	+0.80	+0.62
Mean absolute error (MAE; $\sigma$ )	+0.48	+0.38	+0.29	+0.48	+0.38
Root-mean-square error (RMSE; $\sigma$ )	+0.59	+0.47	+0.38	+0.59	+0.47
Skill score (vs climatology; $SS_{Clim}$ )	+0.43*	+0.64	+0.38	+0.81	+0.61
Skill score (vs persistence; $SS_{Pers}$ )	+0.68*	+0.80	+0.57	+0.89	+0.80
RPSS (vs climatology; $RPSS_{Clim}$ )	−0.13*	+0.36	+0.02	+0.60	+0.36
Pearson’s correlation ( $r$ )	+0.69*	+0.80	+0.63	+0.90	+0.80
Index of agreement ( $d_2$ )	+0.83*	+0.88	+0.79	+0.94	+0.87
Modified index of agreement ( $d_1$ )	+0.61*	+0.68	+0.58	+0.77	+0.67
Modified coefficient of efficiency ( $E_1$ )	+0.26*	+0.43	+0.21	+0.58	+0.40

for real-time predictions of all-Ethiopian Kiremt rainfall anomalies a few days after the end of March, as the necessary initialization data (real-time reanalysis, SST) are available 5–10 days after the end of the month. The predictors selected in this study may not pose a problem in the near-term climate. However, the predictor pool should be reevaluated every few years to account for 1) predictors that remain skillful but have decreasing frequency, 2) predictors that are decreasing in skill, 3) predictors that were not included in the selected sets that are starting to show additional skill, and 4) changes in the likelihood functions due to climate change. Updating the RV and LOOCV prediction models using the entire data with follow up real-time forecast evaluation realistically could lead to successful real-time operational use of the approaches. The simultaneous use of both approaches builds confidence in the value of the prediction if they yield similar forecasts. Moreover, to our knowledge, the methodology of building observational ensembles for statistical prediction is unique. The high-quality local and national prediction capability should have significant beneficial societal implications. In particular, the forecasting of seasonal anomalies at regional and national scales and monthly rainfall totals at specific localities with usable skill could play a key role in risk management to help minimize the damaging effects of recurring droughts in Ethiopia.

*Acknowledgments.* This research was supported by the NOAA Cooperative Institute for Mesoscale Meteorological Studies (CIMMS). The computations for this project were performed at the OU Supercomputing Center for Education and Research (OSCER) at The University of Oklahoma. The primary research data (1970–99) were provided by the National Meteorological Agency of Ethiopia in the early part of this research. Ethiopian seasonal rainfall station data for 2000–02 were kindly provided by Dr. A. G. Barnston. The comments and suggestions of three anonymous reviewers are appreciated. In particular, Reviewer 1's detailed analysis of the manuscript greatly improved the final product. The first three authors express their deep and lasting gratitude to the fourth author, the late Peter J. Lamb, who dedicated much of his life to investigation of climate variability in Africa. Without his support and expert assistance in preparing this manuscript, it would not have been possible.

#### REFERENCES

- Barnston, A. G., W. Thiao, and V. Kumar, 1996: Long-lead forecasts of seasonal precipitation in Africa using CCA. *Wea. Forecasting*, **11**, 506–520, doi:10.1175/1520-0434(1996)011<0506:LLFOSP>2.0.CO;2.
- Beltrando, G., and P. Camberlin, 1993: Interannual variability of rainfall in the eastern Horn of Africa and indicators of atmospheric circulation. *Int. J. Climatol.*, **13**, 533–546, doi:10.1002/joc.3370130505.
- Berhane, F., B. Zaitchik, and A. Dezfuli, 2014: Subseasonal analysis of precipitation variability in the Blue Nile river basin. *J. Climate*, **27**, 325–344, doi:10.1175/JCLI-D-13-00094.1.
- Block, P., and B. Rajagopalan, 2007: Interannual variability and ensemble forecast of Upper Blue Nile Basin Kiremt season precipitation. *J. Hydrometeorol.*, **8**, 327–343, doi:10.1175/JHM580.1.
- Bohn, T. J., M. Y. Sonessa, and D. P. Lettenmaier, 2010: Seasonal hydrologic forecasting: Do multimodel ensemble averages always yield improvements in forecast skill? *J. Hydrometeorol.*, **11**, 1358–1372, doi:10.1175/2010JHM1267.1.
- Bowden, J. H., and F. H. M. Semazzi, 2007: Empirical analysis of intraseasonal climate variability over the Greater Horn of Africa. *J. Climate*, **20**, 5715–5731, doi:10.1175/2007JCLI1587.1.
- Breusch, T. S., and A. R. Pagan, 1979: A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, **47**, 1287–1294, doi:10.2307/1911963.
- Camberlin, P., 1997: Rainfall anomalies in the source region of the Nile and their connection with the Indian summer. *J. Climate*, **10**, 1380–1392, doi:10.1175/1520-0442(1997)010<1380:RAITSR>2.0.CO;2.
- , S. Janicot, and I. Pocard, 2001: Seasonality and atmospheric dynamics of the teleconnection between African rainfall and tropical sea-surface temperature: Atlantic vs. ENSO. *Int. J. Climatol.*, **21**, 973–1005, doi:10.1002/joc.673.
- Charney, J. G., 1975: Dynamics of deserts and drought in the Sahel. *Quart. J. Roy. Meteor. Soc.*, **101**, 193–202, doi:10.1002/qj.49710142802.
- , and J. Shukla, 1981: Predictability of monsoons. *Monsoon Dynamics*, J. Lighthill and R. P. Pearce, Eds., University Press, 99–109.
- , W. J. Quirk, S.-H. Chow, and J. Kornfield, 1977: A comparative study of the effects of albedo change on drought in semi-arid regions. *J. Atmos. Sci.*, **34**, 1366–1385, doi:10.1175/1520-0469(1977)034<1366:ACSOLE>2.0.CO;2.
- Cheung, W. H., G. B. Senay, and A. Singh, 2008: Trends and spatial distribution of annual and seasonal rainfall in Ethiopia. *Int. J. Climatol.*, **28**, 1723–1734, doi:10.1002/joc.1623.
- Clark, C. A., and R. W. Arritt, 1995: Numerical simulations of the effect of soil moisture and vegetation cover on the development of deep convection. *J. Appl. Meteor.*, **34**, 2029–2045, doi:10.1175/1520-0450(1995)034<2029:NSOTEO>2.0.CO;2.
- Clark, D. B., Y. Xue, R. J. Harding, and P. J. Valdes, 2001: Modeling the impact of land surface degradation on the climate of tropical North Africa. *J. Climate*, **14**, 1809–1822, doi:10.1175/1520-0442(2001)014<1809:MTIOLS>2.0.CO;2.
- Conway, D., 2000: The climate and hydrology of the Upper Blue Nile River. *Geogr. J.*, **166**, 49–62, doi:10.1111/j.1475-4959.2000.tb00006.x.
- Degefu, W., 1987: Some aspects of meteorological droughts in Ethiopia. *Drought and Hunger in Africa: Denying Famine a Future*, M. H. Glantz, Ed., Cambridge University Press, 23–36.
- DelSole, T., and J. Shukla, 2002: Linear prediction of Indian monsoon rainfall. *J. Climate*, **15**, 3645–3658, doi:10.1175/1520-0442(2002)015<3645:LPOIMR>2.0.CO;2.
- Diro, G. T., E. Black, and D. I. F. Grimes, 2008: Seasonal forecasting of Ethiopian spring rains. *Meteor. Appl.*, **15**, 73–83, doi:10.1002/met.63.

- Draper, N. R., and H. Smith, 1981: *Applied Regression Analysis*. Wiley, 724 pp.
- Efron, B., and G. Gong, 1983: A leisurely look at the bootstrap, jackknife, and cross-validation. *Amer. Stat.*, **37**, 36–48.
- , and R. J. Tibshirani, 1993: *An Introduction to the Bootstrap*. Chapman and Hall, 436 pp.
- Folland, C. K., T. N. Palmer, and D. E. Parker, 1986: Sahel rainfall and world-wide sea temperatures. *Nature*, **320**, 602–607, doi:10.1038/320602a0.
- , J. Owen, M. N. Ward, and A. Colman, 1991: Prediction of seasonal rainfall in the Sahel region using empirical and dynamical methods. *J. Forecasting*, **10**, 21–56, doi:10.1002/for.3980100104.
- Giannini, A., R. Saravanan, and P. Chang, 2003: Oceanic forcing of Sahel rainfall on interannual to interdecadal time scales. *Science*, **302**, 1027–1030, doi:10.1126/science.1089357.
- Gissila, T., E. Black, D. I. F. Grimes, and J. M. Slingo, 2004: Seasonal forecasting of the Ethiopian summer rains. *Int. J. Climatol.*, **24**, 1345–1358, doi:10.1002/joc.1078.
- Goddard, L., S. J. Mason, S. E. Zebiak, C. F. Ropelewski, R. Basher, and M. A. Cane, 2001: Current approaches to seasonal-to-interannual climate predictions. *Int. J. Climatol.*, **21**, 1111–1152, doi:10.1002/joc.636.
- Goswami, B., and J. Shukla, 1991: Predictability of a coupled ocean–atmosphere model. *J. Climate*, **4**, 3–22, doi:10.1175/1520-0442(1991)004<0003:POACOA>2.0.CO;2.
- Hastenrath, S., L. Greischar, and J. van Heerden, 1995: Prediction of summer rainfall over South Africa. *J. Climate*, **8**, 1511–1518, doi:10.1175/1520-0442(1995)008<1511:POTSRO>2.0.CO;2.
- Jury, M. R., 2013: Ethiopian highlands crop-climate prediction: 1979–2009. *J. Appl. Meteor. Climatol.*, **52**, 1116–1126, doi:10.1175/JAMC-D-12-0139.1.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.
- Korecha, D., and A. G. Barnston, 2007: Predictability of June to September rainfall in Ethiopia. *Mon. Wea. Rev.*, **135**, 628–650, doi:10.1175/MWR3304.1.
- , and A. Sorteberg, 2013: Validation of operational seasonal rainfall forecast in Ethiopia. *Water Resour. Res.*, **49**, 7681–7697, doi:10.1002/2013WR013760.
- Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. Larow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran, 2000: Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate*, **13**, 4196–4216, doi:10.1175/1520-0442(2000)013<4196:MEFFWA>2.0.CO;2.
- Lamb, P. J., 1978a: Large-scale tropical Atlantic surface circulation patterns associated with sub-Saharan weather anomalies. *Tellus*, **30A**, 240–251, doi:10.1111/j.2153-3490.1978.tb00839.x.
- , 1978b: Case studies of tropical Atlantic surface circulation patterns during recent sub-Saharan weather anomalies: 1967 and 1968. *Mon. Wea. Rev.*, **106**, 482–491, doi:10.1175/1520-0493(1978)106<0482:CSOTAS>2.0.CO;2.
- , and R. A. Peppler, 1992: Further case studies of tropical Atlantic surface atmospheric and oceanic patterns associated with sub-Saharan drought. *J. Climate*, **5**, 476–488, doi:10.1175/1520-0442(1992)005<0476:FCSOTA>2.0.CO;2.
- Lanczkiet, S., A. Frankl, E. Adgo, P. Termonia, and J. Nyssen, 2015: Droughts related to quasi-global oscillations: A diagnostic teleconnection analysis in North Ethiopia. *Int. J. Climatol.*, doi:10.1002/joc.4074, in press.
- Legates, D. R., and G. J. McCabe, 1999: Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.*, **35**, 233–241, doi:10.1029/1998WR900018.
- Mason, S. J., and G. M. Mimmack, 1992: The use of bootstrap confidence intervals for the correlation coefficient in climatology. *Theor. Appl. Climatol.*, **45**, 229–233, doi:10.1007/BF00865512.
- Mutai, C. C., and M. N. Ward, 2000: East African rainfall and the tropical circulation/convection on intraseasonal to inter-annual timescales. *J. Climate*, **13**, 3915–3939, doi:10.1175/1520-0442(2000)013<3915:EARATT>2.0.CO;2.
- Ndiaye, O., M. N. Ward, and W. M. Thiaw, 2011: Predictability of seasonal Sahel rainfall using GCMs and lead-time improvements through the use of a coupled model. *J. Climate*, **24**, 1931–1949, doi:10.1175/2010JCLI3557.1.
- Nicholls, N., 2001: Commentary and analysis: The insignificance of significance testing. *Bull. Amer. Meteor. Soc.*, **82**, 981–986, doi:10.1175/1520-0477(2001)082<0981:CAATID>2.3.CO;2.
- Nicholson, S. E., 2014: The predictability of rainfall over the Greater Horn of Africa. Part I. Prediction of seasonal rainfall. *J. Hydrometeorol.*, **15**, 1011–1027, doi:10.1175/JHM-D-13-062.1.
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, doi:10.1029/2002JD002670.
- Riddle, E. E., and K. H. Cook, 2008: Abrupt rainfall transitions over the Greater Horn of Africa: Observations and regional model simulations. *J. Geophys. Res.*, **113**, D15109, doi:10.1029/2007JD009202.
- Saunders, M. A., and A. S. Lea, 2005: Seasonal prediction of hurricane activity reaching the coast of the United States. *Nature*, **434**, 1005–1008, doi:10.1038/nature03454.
- Segele, Z. T., and P. J. Lamb, 2005: Characterization and variability of Kiremt rainy season over Ethiopia. *Meteor. Atmos. Phys.*, **89**, 153–180, doi:10.1007/s00703-005-0127-x.
- , —, and L. M. Leslie, 2009a: Large-scale atmospheric circulation and global sea surface temperature associations with Horn of Africa June–September rainfall. *Int. J. Climatol.*, **29**, 1075–1100, doi:10.1002/joc.1751.
- , —, and —, 2009b: Seasonal-to-interannual variability of Ethiopian/Horn of Africa monsoon. Part I: Associations of wavelet-filtered large-scale atmospheric circulation and global sea surface temperature. *J. Climate*, **22**, 3396–3421, doi:10.1175/2008JCLI2859.1.
- , L. M. Leslie, and P. J. Lamb, 2009c: Evaluation and adaptation of a regional climate model for the Horn of Africa: Rainfall climatology and interannual variability. *Int. J. Climatol.*, **29**, 47–65, doi:10.1002/joc.1681.
- Seleshi, Y., and U. Zanke, 2004: Recent changes in rainfall and rainy days in Ethiopia. *Int. J. Climatol.*, **24**, 973–983, doi:10.1002/joc.1052.
- Shanko, D., and P. Camberlin, 1998: The effects of the southwest Indian Ocean tropical cyclones on Ethiopian drought. *Int. J. Climatol.*, **18**, 1373–1388, doi:10.1002/(SICI)1097-0088(1998100)18:12<1373::AID-JOC313>3.0.CO;2-K.
- Solow, A. R., 1985: Bootstrapping correlated data. *J. Int. Assoc. Math. Geol.*, **17**, 769–775, doi:10.1007/BF01031616.
- S-PLUS, 2013: TIBCO Spotfire S+ 8.2. Tibco Software, Inc. [Available online at <http://tibco-spotfire-s.software.informer.com/8.2/>.]
- Steinskog, D. J., D. B. Tjøstheim, and N. G. Kvamstø, 2007: A cautionary note on the use of the Kolmogorov–Smirnov test

- for normality. *Mon. Wea. Rev.*, **135**, 1151–1157, doi:[10.1175/MWR3326.1](https://doi.org/10.1175/MWR3326.1).
- Tamhane, A. C., and D. D. Dunlop, 2000: *Statistics and Data Analysis: From Elementary to Intermediate*. Prentice Hall, 736 pp.
- Venables, W. M., and B. D. Ripley, 1997: *Statistics and Computing: Modern Applied Statistics with S-Plus*. Springer, 548 pp.
- Viste, E., D. Korecha, and A. Sorteberg, 2013: Recent drought and precipitation tendencies in Ethiopia. *Theor. Appl. Climatol.*, **112**, 535–551, doi:[10.1007/s00704-012-0746-3](https://doi.org/10.1007/s00704-012-0746-3).
- Vizy, E. K., and K. H. Cook, 2003: Connection between the summer East African and Indian rainfall regimes. *J. Geophys. Res.*, **108**, 4510, doi:[10.1029/2003JD003452](https://doi.org/10.1029/2003JD003452).
- Ward, M. N., 1998: Diagnosis and short-lead time prediction of summer rainfall in tropical Africa at interannual and multi-decadal timescales. *J. Climate*, **11**, 3167–3191, doi:[10.1175/1520-0442\(1998\)011<3167:DASLTP>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<3167:DASLTP>2.0.CO;2).
- Webster, P. J., and S. Yang, 1992: Monsoon and ENSO: Selectively interactive systems. *Quart. J. Roy. Meteor. Soc.*, **118**, 877–925, doi:[10.1002/qj.49711850705](https://doi.org/10.1002/qj.49711850705).
- , V. O. Magaña, T. N. Palmer, J. Shukla, R. A. Tomas, M. Yanai, and T. Yasunari, 1998: Monsoons: Processes, predictability, and prospects for prediction. *J. Geophys. Res.*, **103**, 14 451–14 510, doi:[10.1029/97JC02719](https://doi.org/10.1029/97JC02719).
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 91, Academic Press, 627 pp.
- Williams, A. P., and Coauthors, 2012: Recent summer precipitation trends in the Greater Horn of Africa and the emerging role of Indian Ocean Sea surface temperature. *Climate Dyn.*, **39**, 2307–2328, doi:[10.1007/s00382-011-1222-y](https://doi.org/10.1007/s00382-011-1222-y).
- Willmott, C. J., S. G. Ackleson, R. E. Davis, J. J. Feddema, K. M. Klink, D. R. Legates, J. O'Donnell, and C. M. Rowe, 1985: Statistics for the evaluation and comparison of models. *J. Geophys. Res.*, **90**, 8995–9005, doi:[10.1029/JC090iC05p08995](https://doi.org/10.1029/JC090iC05p08995).
- Xue, Y., and J. Shukla, 1993: The influence of land surface properties on Sahel climate. Part I: Desertification. *J. Climate*, **6**, 2232–2245, doi:[10.1175/1520-0442\(1993\)006<2232:TIOLSP>2.0.CO;2](https://doi.org/10.1175/1520-0442(1993)006<2232:TIOLSP>2.0.CO;2).