

# Investigating Long-Range Seasonal Predictability of East African Short Rains: Influence of the Mascarene High on the Indian Ocean Walker Cell

XIAO PENG

*School of Civil and Environmental Engineering, Cornell University, Ithaca, New York*

SCOTT STEINSCHNEIDER

*Department of Biological and Environmental Engineering, Cornell University, Ithaca, New York*

JOHN ALBERTSON

*School of Civil and Environmental Engineering, Cornell University, Ithaca, New York*

(Manuscript received 8 May 2019, in final form 11 November 2019)

## ABSTRACT

We investigate the predictability of East African short rains at long (up to 12 month) lead times by relating seasonal rainfall anomalies to climate anomalies associated with the predominant Walker circulation, including sea surface temperatures (SST), geopotential heights, zonal and meridional winds, and vertical velocities. The underlying teleconnections are examined using a regularized regression model that shows two periods of high model skill (0–3-month lead and 7–9-month lead) with similar spatial patterns of predictability. We observe large-scale circulation anomalies consistent with the Walker circulation at short lead times (0–3 months) and dipoles of SST and height anomalies over the Mascarene high region at longer lead times (7–9 months). These two patterns are linked in time by anticyclonic winds in the dipole region associated with a perturbed meridional circulation (4–6-month lead). Overall, these results suggest that there is potential to extend forecast lead times beyond a few months for drought impact mitigation applications.

## 1. Introduction

Food security has been and continues to be a major challenge in eastern Africa. This region is particularly vulnerable to extreme weather events like droughts and floods, because the regional economy and food production are largely dependent on rainfed agriculture. As an example, the 2011–12 drought caused a disastrous food crisis that impacted more than 13 million people in the Horn of Africa, killing between 50 000 and 100 000 people (Hillier and Dempsey 2012). These impacts perpetuate a cycle of poverty (Barnett et al. 2008) and

highlight the critical need for improved climate resiliency in the region. Toward this end, this study seeks to improve seasonal predictions of rainfall in East Africa, with a particular focus on extended lead times (>6 months), to enable early warning systems and mitigative actions that can help reduce the impacts of major droughts and floods.

Eastern African rainfall seasonality is bimodal, with two distinct rainy periods termed the “long rains” (March–May) and the “short rains” [October–December (OND) or October–November (ON)]. Interannual variability of rainfall from the two seasons is relatively independent (Clark et al. 2003), with higher variability in the short rains but more total rainfall during the long rain season (Mutai et al. 1998). Further, interannual variability in the short rains is more closely related to external forcing (Hastenrath et al. 1993, 2004). This motivates our focus on improving the predictability of the East African short rains (EASR) in the study presented here.

There is a well-established teleconnection between EASR and forcing from large-scale circulation over the

---

Denotes content that is immediately available upon publication as open access.

---

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JAMC-D-19-0109.s1>.

---

Corresponding author: John Albertson, [albertson@cornell.edu](mailto:albertson@cornell.edu)

DOI: 10.1175/JAMC-D-19-0109.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

Indian Ocean basin. Under normal conditions, the short rains are influenced by a Walker circulation cell characterized by surface westerlies over the Indian Ocean, ascending air over Indonesia, upper-level easterlies, and descending air over East Africa (Hastenrath 2007). Intensification of this circulation can result in drier conditions in East Africa. This cell, and by extension East African rainfall, can be modulated by underlying sea surface temperature anomalies (SSTAs) in the Indian Ocean. Saji et al. (1999) related tropical East African rainfall anomalies to the Indian Ocean dipole (IOD), a gradient in SSTAs with positive events defined by warm waters off the coast of East Africa and cold waters near Indonesia. Black et al. (2003) identified a similar relationship but found it to be strongly nonlinear, with rainfall anomalies only associated with extreme and persistent SSTA events that reverse the normal zonal SST gradient for several months. Under these situations, strong easterly winds in the northern-central Indian Ocean weaken the westerly surface flow that normally transports moisture away from East Africa. The anomalous easterlies can even reverse the full Walker cell, enhancing convergence and convection over East Africa and leading to greater rainfall in that region during boreal autumn.

The dynamical pathways were further explored by Ummenhofer et al. (2009), who used ensemble simulations to show that the warm western SST pole of the IOD has a larger role than the eastern cold pole in promoting stronger easterly winds from the Indian Ocean and moisture convergence over East Africa. Bahaga et al. (2015) also confirmed the dominant role of the western pole of the IOD using AGCM simulations, which demonstrated that the warm western pole initiated a Gill-type response. In Hastenrath et al. (2011), the important role of upper-level atmospheric motion was addressed: the summer warming and high stand of upper-tropospheric topography over South Asia enables strong upper-tropospheric easterlies over the northern-central Indian Ocean in the following boreal autumn, while lack of such a mechanism in boreal spring leads to less predictability of the long rains. A more recent study by Nicholson (2015) utilizing a 139-year rainfall record showed that the links between the Walker cell and the East African short rains were significantly weaker during certain historical periods (e.g., 1920–60) and stressed the time dependence of the links. The nonstationary correlation between Indian Ocean SSTs (e.g., IOD) and EASR was further investigated by Bahaga et al. (2019), which suggests that more interannual variability of the EASR was explained by the Indian Ocean SSTs in recent years.

Many studies have demonstrated the potential for seasonal prediction of the EASR based on the above

teleconnection patterns (Mutai et al. 1998; Hastenrath et al. 2004; Behera et al. 2005; Nicholson 2014; Walker et al. 2019). Much of the prediction skill appears linked to dynamics internal to the Indian Ocean, and in particular, measures of the Walker cell across the Indian Ocean. While there is some predictability associated with dynamical fields, the east–west SSTA gradient provides superior prediction skill up to three months lead time. Mutai et al. (1998) developed a regression model using the July–September and OND principal components associated with “varimax”-rotated empirical orthogonal functions (EOFs) of the global SSTs, which yielded correlation coefficients  $r$  of [0.56–0.78] in the testing period at a lead time of 1 month. Hastenrath et al. (2004) used a stepwise regression model with multiple predictors that are measures of the Walker cell (e.g., surface and upper-level zonal winds, vertical air motion, SST gradient, and pressure difference), and produced cross-validated correlation coefficients within the range of [0.02, 0.74] at a lead time of 1 month. Nicholson (2014) used a similar regression framework and achieved cross-validated  $r$  scores of 0.7 at a lead time of 2 months primarily using atmospheric predictors, and  $r$  scores within [0.6, 0.71] at a lead time of 5 months using mainly SST predictors. Other work (Behera et al. 2005; Tierney et al. 2013) also suggests that ENSO plays a limited role in controlling modulations in the Walker cell over the Indian Ocean, and thus has less value for seasonal forecast development compared to direct observations of the Indian basin circulation. Also, Bahaga et al. (2019) disentangled the IOD and ENSO influence on the EASR, showing that the ENSO influence is mediated by an in-phase occurrence of IOD.

To the best of our knowledge, the longest lead times of skillful seasonal forecasts have been limited to around 3–5 months. These lead times are related to measures of the Walker cell used in forecast development, which are first observed as SSTAs over the eastern Indian Ocean that develop in June (Tozuka et al. 2016). These lead times, while useful, may not be long enough to enable decision makers to effectively execute mitigative actions. For example, if water managers knew that there was a high likelihood of drought in the short rains prior to or during the previous long rain season (6–8-month lead), they could prepare by reducing the allocation of water supplies provided by the long rains to mitigate water shortages later in the year. Similarly, long lead times are needed for use in cropping decisions. This motivates the question of whether we can extend seasonal forecast lead times of the East African short rains by focusing on large-scale climate processes that trigger subsequent perturbations in the Walker cell.

We posit that a focus on the Mascarene high and its relationship to interannual variability of EASR presents a promising approach for this extension. The Mascarene high (MH) is a semipermanent subtropical ridge in the Southern Hemisphere off the coast of Madagascar and has been shown to have great impacts on variability of the Asian Monsoon (Turner and Annamalai 2012; Xue et al. 2004). Black et al. (2003) proposed that anomalies in the MH could initiate a perturbed meridional circulation cell, that in turn could influence the Walker circulation and by extension EASR. The link between the MH and EASR was further studied by Manatsa et al. (2014), who used correlation and composite analyses for flood and drought events based on observations and reanalysis data to explore the connection and found that EASR variability is strongly linked to the zonal displacement of the MH. When the MH eastern ridge is anomalously displaced to the west, the southeast trade winds over the southern Indian Ocean strengthen, which alters the SST pattern and leads to suppressed convection over East Africa. Conversely, when the MH eastern ridge is anomalously displaced to the east, the southeast trade winds over the southern Indian Ocean weaken, which again alters the SST pattern and leads to enhanced convection over East Africa. However, the seasonally lagged relationship between anomalies in the MH, the Walker circulation, and EASR has not been sufficiently studied to determine whether there is potential to extend forecast lead times of EASR. This important research gap is a major motivation of this paper.

In this study, we first explore potential predictors from a predictor pool consisting of multiple climate fields over the Indian Ocean using a regularized regression model. Instead of only using spatially averaged rainfall as the predictand, the regression analysis is repeated for rainfall at every grid cell to examine spatial variability of the predictive skill. Then, regression coefficient maps are compared with Indian Ocean basin composites prior to historical extreme EASR events to better understand the predictive signals being identified by the regression models. Using a cross-validated procedure and additional lead-lag analyses on large-scale climate fields, we assess the robustness of the regression models and the resulting insights. On the basis of these results, we propose a theory on the potential for extended predictability of EASR.

## 2. Data

Monthly rainfall data between 1981 and 2017 are collected from the Climate Hazards Group Infrared

Precipitation with Station data (CHIRPS) (Funk et al. 2015) for their fine spatial resolution and relatively long temporal coverage. The gridded rainfall data incorporate satellite imagery ( $0.05^\circ \times 0.05^\circ$  resolution) and in situ station data and are reinterpolated to grids of  $0.5^\circ$  in this study. The CHIRPS rainfall performed very well when compared with in situ observations, likely because of its direct inclusion of rain gauge data and microwave images during calibration (Kimani et al. 2017). This study focuses on the equatorial East African region (land area in  $5^\circ\text{S}$ – $5^\circ\text{N}$ ,  $33^\circ$ – $47^\circ\text{E}$ ), similar to the short rains area used in Bahaga et al. (2019). This region has a well-defined rainfall regime with consistent, bimodal seasonality (Yang et al. 2015). Average rainfall between October and November is used to characterize the short rains in this study. Previous studies of the short rain season have either used rainfall over the OND (Mutai et al. 1998; Clark et al. 2003; Bahaga et al. 2019) or ON (Hastenrath et al. 2004, 1993; Behera et al. 2005; Nicholson 2015) period to represent the short rains. Our preliminary analysis (Fig. S1 in the online supplemental material) suggested that predictability was slightly higher for the ON rainfall than for the OND rainfall, and when regressions were built separately using October, November, and December rainfall there was little connection between December rainfall and large-scale climate fields. Therefore, we selected ON rainfall as our short rain index to better isolate the predictive signal.

The distribution of the EASR is highly skewed, which can overly emphasize a few extreme wet events in statistical analysis of the data. Therefore, the standardized precipitation index (SPI) is used on the basis of a fitted gamma distribution (McKee et al. 1993). The EASR series at all grid cells were first averaged and then the SPI was calculated using the spatial average. Extreme events are defined as the events with absolute values of SPIs larger than 1 as based on the spatial-averaged EASR SPI, which is shown in Fig. 1.

The monthly gridded SSTs ( $1^\circ$  latitude by  $1^\circ$  longitude) over 1980–2017 are collected from the Hadley Centre Sea Ice and Sea Surface Temperature dataset (Rayner et al. 2003). Only SSTs in the Indian Ocean basin are used here since previous studies suggest the influence of SSTs on the EASR is “local” (Ummenhofer et al. 2009; Bahaga et al. 2015). The Indian Ocean basin domain is based on the ocean boundary definition from the National Oceanic and Atmospheric Administration ([https://www.nodc.noaa.gov/woce/woce\\_v3/woce\\_data\\_1/woce-uot/summary/bound.htm](https://www.nodc.noaa.gov/woce/woce_v3/woce_data_1/woce-uot/summary/bound.htm)). Several other climate fields ( $2.5^\circ \times 2.5^\circ$  resolution) over the region of  $65^\circ\text{S}$ – $35^\circ\text{N}$ ,  $20^\circ$ – $150^\circ\text{E}$  are also taken from the NCEP–NCAR Reanalysis II product over the period from January 1980 to December 2017, including 850-hPa geopotential

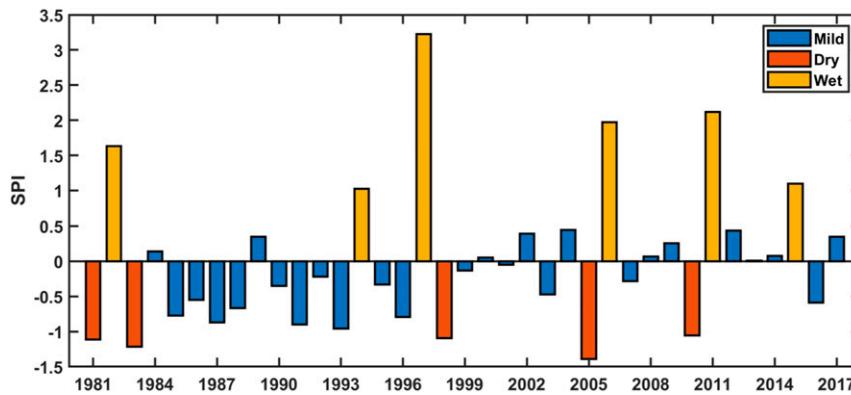


FIG. 1. The spatial-averaged EASR SPI series from 1981 to 2017.

height (hereinafter height or HGT), 850-hPa (low) and 200-hPa (upper) zonal/meridional winds ( $u/v$  winds), and 500-hPa vertical motion in the atmosphere ( $\omega$ ). Standardized anomalies are computed for all variables except rainfall by first subtracting the monthly climatology and then dividing by the local standard deviation to reduce the effects of the dominant seasonal cycle and varying amplitudes at different latitudes. This step is done locally at each grid cell.

### 3. Methods

Linear models between the EASR SPI and several gridded climate variables (SST, height,  $u/v$  wind, and  $\omega$  anomalies) at different lead times are developed using elastic net regression (Zou and Hastie 2005). Elastic net is a penalized regression that is a balance between ridge (Hoerl and Kennard 1970) and the Least Absolute Shrinkage and Selection Operator (LASSO; Tibshirani 1996) regression and promotes group selection, that is, the inclusion or exclusion of groups of strongly correlated variables in the model. This approach helps manage collinearity and avoid overfitting by selecting predictors among a large candidate set but retains groups of potentially important predictors instead of arbitrarily assigning the effect to a single variable (i.e., LASSO regression). A general form is given by

$$(\alpha, \boldsymbol{\beta}) = \operatorname{argmin} \left[ \frac{1}{2N} \sum_i \left( y_i - \alpha - \sum_j \beta_j x_{i-t,j} \right)^2 - \lambda P(\boldsymbol{\beta}) \right],$$

where the penalty term  $P(\boldsymbol{\beta})$  is given by

$$P(\boldsymbol{\beta}) = \frac{1-k}{2} \sum_j \beta_j^2 + k \sum_j |\beta_j|.$$

Here,  $y_i$  is the EASR SPI at time  $i$ ,  $x_{i-t,j}$  is the value of a climate variable at grid cell  $j$  at time  $i-t$ ,  $(\alpha, \boldsymbol{\beta})$  are the

estimated regression coefficient, and  $\lambda$  is a nonnegative regularization parameter. The parameter  $k$  controls the balance between ridge and LASSO penalties and is set to 0.01 in this study after preliminary testing of model performance with different  $k$  values. Models with different values of  $k$  (i.e., 1, 0.1, 0.01) yield similar model skill, and a smaller  $k$  is used for better visualization (models with larger  $k$  would assign effects to very few predictors, leading to very sparse coefficient maps). On the basis of this parameterization, the model selects several groups of gridded climate variables from the original data fields that are representative of the larger spatial patterns of climate related to the short rains.

The hyperparameter  $\lambda$  is typically chosen using a  $J$ -fold cross validation (CV) based on the training data (Tibshirani 1996; Breiman and Spector 1992). However, to repeat the CV for every climate variable and every lead time is not feasible given the computation expense. In addition, a constant  $\lambda$  helps ensure the degree of shrinkage is the same across each variable and lead time, making it easier to compare the results. Therefore, for each of the climate variables we fix the value of  $\lambda$  for all lead times. We first select  $\lambda$  using a 10-fold cross validation for 0 lead time based on all of the data between 1981 and 2017. This step is repeated separately 100 times for each climate variable to account for variability introduced through the random selection of fold. The values of  $\lambda$  are then collected and the median values are selected and used in the following analysis. We note that sensitivity tests suggest that the regression model has low sensitivity to the value of  $\lambda$ .

With the hyperparameters (i.e.,  $k$  and  $\lambda$ ) determined, the models are fitted and tested 200 times using shuffled data, where in each iteration 27 randomly selected years are used for model training and the remaining 10 years are used for model testing. This cross-validation procedure is used to assess the robustness of regression model results. The spatial-averaged EASR SPI is used as the

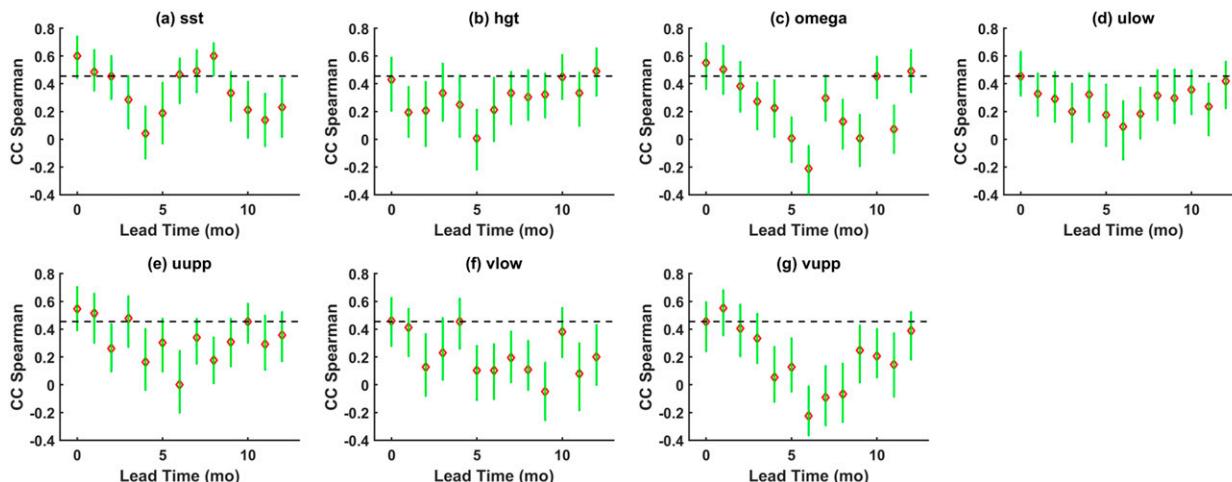


FIG. 2. Model skill, shown here as a Spearman correlation between model predictions and observations, as a function of lead time for the testing period using shuffled spatially averaged EASR SPI. The model predictors are anomalies of (a) SST, (b) 850-hPa HGT, (c) 500-hPa omega, (d) 850-hPa *u* wind, (e) 200-hPa *u* wind, (f) 850-hPa *v* wind, and (g) 200-hPa *v* wind. The 25th–75th-percentile ranges are plotted as green lines; the median CCs are shown with red diamonds. The horizontal black dashed lines indicate the one-tailed  $p = 0.1$  significance level.

predictand and the analysis in this step is conducted separately for each individual climate variable using the corresponding  $\lambda$ s. Also, to account for spatial variability of the predictability, the regression is repeated using SPI at every grid by fitting the model using only SST as the predictor. These models are only fitted over 1981–2009 and tested over 2008–17. All results are evaluated in terms of the distribution of model skill, reported as the Spearman’s correlation coefficient (CC). The statistical significance of the regression coefficients is not straightforward to calculate using a regularized regression and so is not reported (Javanmard and Montanari 2014).

The elastic net regression will lead to a small number of selected grid cells for a subset of climate predictors that are most related to the EASR at different lead times. The spatial patterns of these selected climate predictors are also compared against climate composites to ensure the models are consistent with those composite patterns. According to our definition of EASR extremes, there are six wet extreme events (1982, 1994, 1997, 2006, 2011, and 2015) and five dry extreme events (1981, 1983, 1998, 2005, 2010) in the 1981–2017 record. We composite SST, height, *u/v* wind, and omega anomalies 0–12 months prior to these extreme dry and extreme wet events. We examine composites on the difference between wet and dry events and separately for each type of event, to consider asymmetric controls on wet and dry events that are driven by different processes (Black et al. 2003; Nicholson 2015). A bootstrapping significance test is applied in the composite analysis, in which 1000 composite means of six randomly selected years (for the wet composite) and five randomly

selected years (for the dry composite) are collected and their difference are calculated (i.e., 1000 composite differences). On the basis of these bootstrapped composites, only the results that are statistically significant at the 90% level are shown.

Last, a lead–lag correlation analysis is conducted to further examine the patterns found in the regression and composite analyses. We define indices of large-scale climate fields for several of our covariates in key regions of the Indian Ocean basin identified in the previous analyses, and then examine the lead–lag correlations of these indices with each other and with other indices from previous studies used to characterize the Walker circulation. The goal of this analysis is to help develop a better understanding of the causal chain of climate processes that lead to long-lead predictability of the EASR.

#### 4. Results and discussions

##### a. Model skill with shuffled data using spatially averaged SPI

Regression model skill for the testing period using shuffled data is shown in Fig. 2. For all climate variables, consistent decreasing trends in accuracy are observed for out-of-sample predictions as lead time increases from 0 to 5/6 months, with median CC values above or around  $p = 0.1$  significance for concurrent (lag 0) predictions to around zero skill at a lead time of 6 months. We note the level of skill exhibited by the elastic net regression at lead times of 0–6 months is similar to that from the multiple linear regression by Nicholson (2014). Interestingly though, at longer leads (6–10 months) and

for some climate predictors, out-of-sample prediction performance increases to levels comparable to lead times of 0–2 months, before dropping again to non-significant levels at lead times of 11–12 months. When SSTAs are used as predictors and at a lead time of 8 months, even the 25th percentile of prediction skill is above the 0.1 significance level. At a lead time of 10 months, the median CCs are around the significance level for models using height, omega, and 200-hPa  $u$  wind; at a lead time of 12 months, the median CCs are above the significance level for models using height and omega.

### b. Spatial variability of predictive skill

Spatial variability of the predictive skill is examined by repeating the regression analysis for SPI at each grid cell. Only SST is used as the predictor, and the data are not shuffled in this analysis. Model skill as a function of lead time is shown in Fig. 3. Consistent patterns are observed: at lead times of 0–5 months, model skill drops from above the 0.1 significant level to around zero skill; at a lead time of 8 months, model skill rises above that observed at a 0 lead. To make visualization of the results easier, we divide the lead times into ranges roughly based on behavior of the prediction skill: 1) S1: lead times of 0–3 months; 2) S2: lead time of 4–6 months; 3) S3: lead times of 7–9 months; and 4) S4: lead times of 10–12 months. Prediction skill maps for our study area during the 4 seasonal lead times are shown in Fig. 4. During S1 and S3, at most grid cells (83% and 80% for S1 and S3, respectively), significant prediction skill is observed for least one month within the season. However, only 54% and 62% of the grid cells are found to have significant prediction skill during S2 and S4, respectively. More grid cells with significant prediction skill are distributed along the coast for S4 and farther inland for S2. Similar spatial patterns are observed for S1 and S3, in which only some grid cells around the central region are found to not have significant prediction skill. The gap region with relatively poor predictive skill overlaps with the eastern part of the East African Highlands. While the poor predictive skill could be physically based, we cannot rule out the possibility that it is resulted from the elevation-associated bias in the rainfall data (Kimani et al. 2017). We do note that consistent patterns of predictive skill were observed when the analysis was repeated using rainfall from the Global Precipitation Climatology Project (GPCP) (Fig. S2 of the online supplemental material).

### c. Comparison between climate composites and regression predictor patterns

The predictive skill at lead times of 7–9 months is notable because the anomalous Walker circulation has

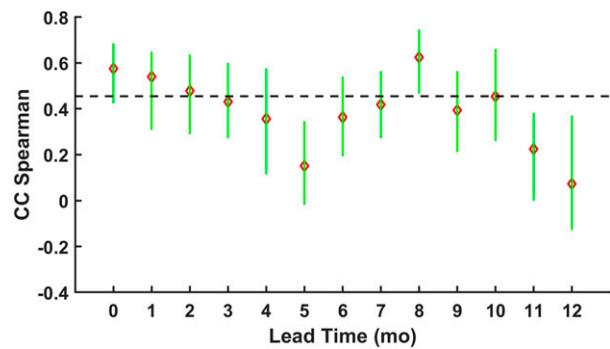


FIG. 3. Similar to Fig. 2, but showing the distribution of regression model skill at each EASR SPI grid cell. The model is fit to SST data between 1981 and 2007, and correlation coefficients are shown for the testing period (2008–17).

not yet been initiated (Tozuka et al. 2016), suggesting there is some other mechanism besides the tropical Walker cell acting as a source of predictability. To better understand this source and determine how it propagates into shorter lead times, spatial distributions of regression coefficients for each period are examined in Figs. 5 and 6. The coefficients are fitted using the full series of the spatially averaged ESAR SPI between 1981 and 2017 and the regression analysis is done individually for each climate variable. Results from the  $u$  wind and  $v$  wind models are combined to produce coefficient maps for lower (850 hPa) and upper (200 hPa) winds. The coefficients represent sparse selection of predictor variables under elastic net regression from the large candidate set for the entire domain and highlight those regions that provide the best source of predictability at different lead times. It is worth mentioning here that the regularization procedure is designed to select only a small subset of grid cells to be included in the final model as a way to control for collinearity among the full covariate set. Therefore, the set of regression coefficients can be used to identify “hot spots” where a certain climate variable most strongly contributes to the teleconnection, but the coefficients should not be interpreted as representing the true spatial extent of significant relationships between the climate fields and EASR. Hence, the coefficient maps are compared with composite maps (Figs. 7 and 8) of wet–dry years that include no regularization and therefore present the teleconnection patterns more continuously.

During S4, negative (cold) SST coefficients are observed off the coast of the Horn of Africa while positive (warm) SST coefficients are observed near the Sumatra region, resembling a negative-phase Walker circulation (Fig. 5d). Other characteristics of this negative phase are also observed: positive (subsiding) omega coefficients are observed over East Africa (Fig. 5l) and westerlies

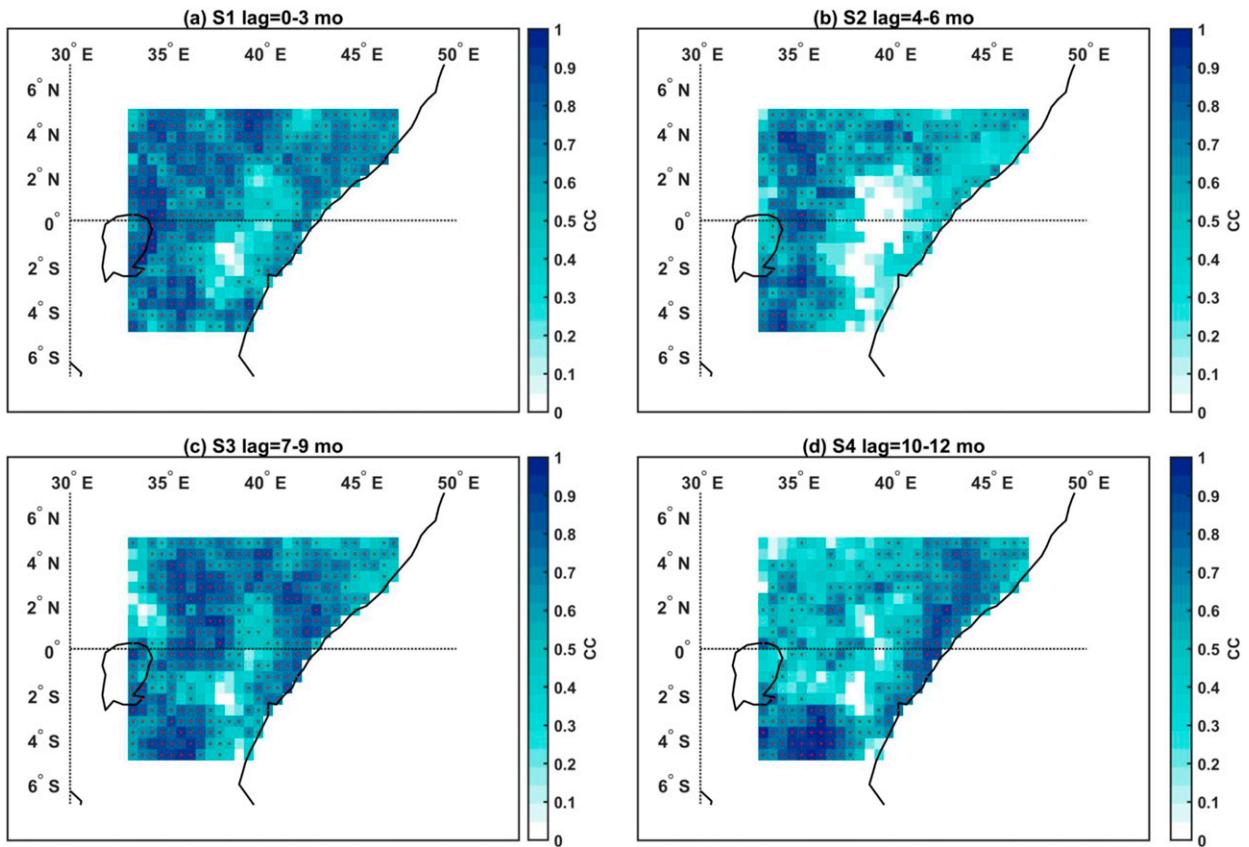


FIG. 4. Maps of the best intraseason prediction skill for different lead-time ranges. The regression analysis is conducted for periods of (a) S1 (lead times = 0–3 months), (b) S2 (lead times = 4–6 months), (c) S3 (lead times = 7–9 months), and (d) S4 (lead times = 10–12 months), and the best intraseason skill is defined as the highest CC score for the testing period within the corresponding season. The grid cells with significant prediction skill (one-tailed  $p = 0.1$ ) are marked with dots.

are observed over the tropical Indian Ocean for the 850 hPa winds (Fig. 6d). All these factors contribute to less convection over East Africa and less transport of water vapor from ocean to land, which result in a drier short rain event. This suggests that dry short rain seasons tend to precede wetter seasons in the following year, which is supported by the negative autocorrelation of the EASR SPI at a 1-year lead time (Spearman’s rank correlation yields  $-0.47$  for 36 samples; one-tailed  $p < 0.005$ ).

Range S3 has notably high prediction skill observed before the initiation of the perturbed Walker circulation. In the SST coefficient map (Fig. 5c), clusters of positive (warm) coefficients are observed around the Mascarene high area over the southern Indian Ocean and negative (cold) coefficients are sparsely distributed over the central and northern Indian Ocean. Positive height coefficients are also observed in areas slightly farther to the north (Fig. 5g). During S2, the model ceases to select SST predictors, as the SSTA and height features associated with the perturbed MH fades out

and the modulated Walker cell is not fully developed (Fig. 5b). Instead, the model emphasizes southerlies around the eastern Indian Ocean near the Sumatra region for 850 hPa winds (Fig. 6b). A plausible explanation for the drop in model skill during the transition period of S2 is that the climate variables chosen by the model (i.e., winds) are noisy, both in terms of measurement (i.e., reanalysis model) error and possibly with respect to geographic location, and therefore are more difficult for the regression model to use when predicting EASR. Patterns of coefficients during S1 are very consistent and well described by the Walker circulation theory: positive (warm) SST coefficients (Fig. 5a) and negative (uplifting) omega coefficients (Fig. 5i) are observed in the western basin and opposite coefficients for the eastern basin; negative coefficients are observed over the tropical Indian Ocean for low-level winds (Fig. 6a). The eastern Indian Ocean is dominated by positive height coefficients (Fig. 5e). Another relevant point is that the southerlies for low-level winds around the eastern Indian Ocean persists from S2 to S1 (Figs. 6b,a).

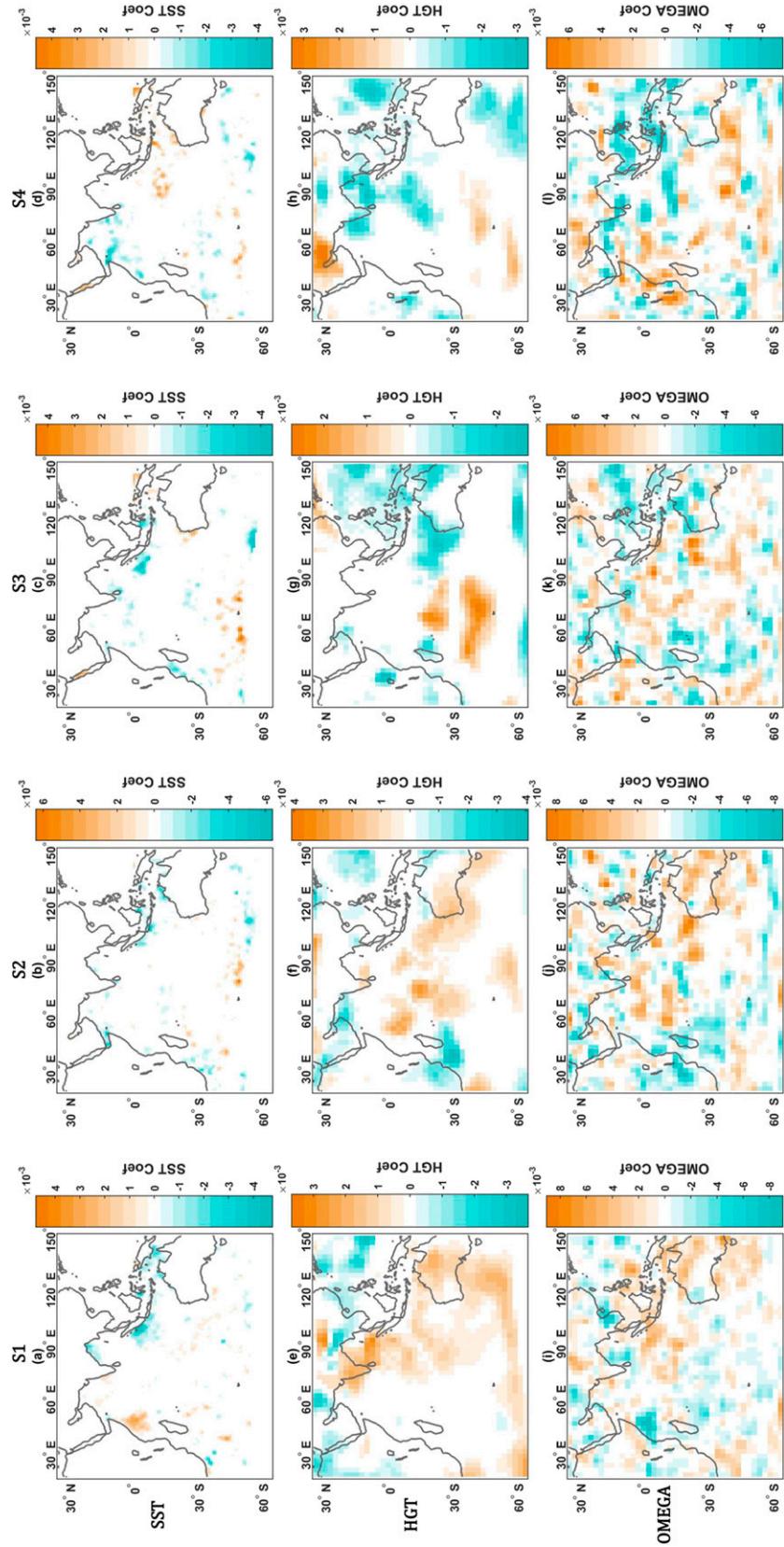


FIG. 5. Coefficient maps for anomalies of (top) SST, (middle) HGT, (bottom) omega during periods of (a),(e),(i) S1; (b),(f),(j) S2; (c),(g),(k) S3; and (d),(h),(l) S4. Coefficients values are shown as colors.

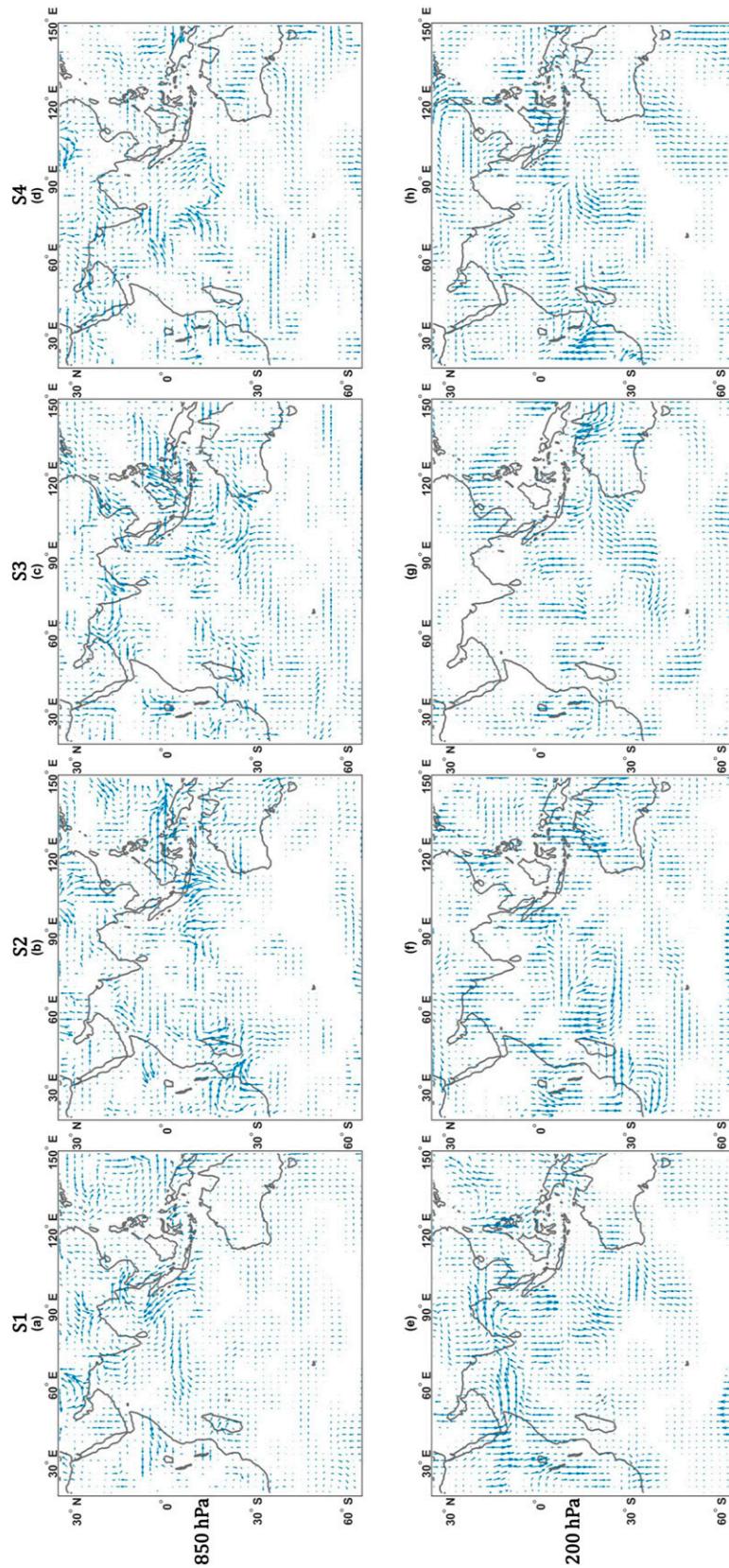


FIG. 6. Coefficient maps for anomalies of (top) 850-hPa wind and (bottom) 200-hPa wind during periods of (a), (e) S1, (b), (f) S2, (c), (g) S3, and (d), (h) S4. Coefficients of the  $u$  winds are visualized using zonal arrows, with an eastward arrow denoting a positive coefficient; coefficients of the  $v$  winds are visualized using meridional arrows, with a northward arrow denoting a positive coefficient. Amplitudes of the coefficients are represented using the lengths of the arrows. For each atmosphere level (i.e., 850 and 200 hPa), the  $u$  and  $v$  wind coefficient maps are combined by adding the two vector fields.

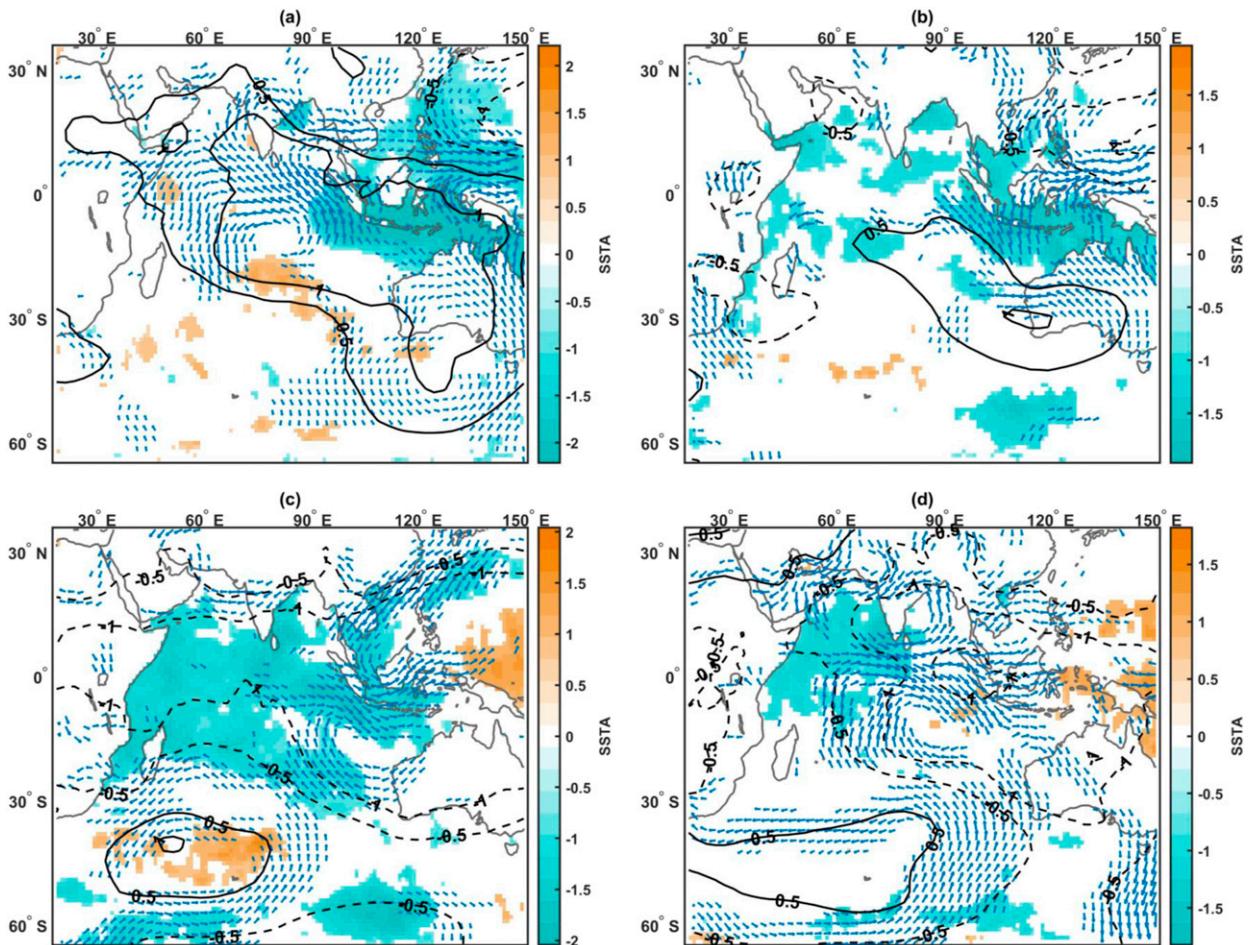


FIG. 7. Composite maps for the difference between historical extreme wet and dry events for SST (shading), 850-hPa HGT (contours), and 850-hPa winds (arrows) for (a) S1, (b) S2, (c) S3, and (d) S4. Results are only shown if significant when compared with a bootstrapped 90% confidence interval.

We note that, for all seasons, regression coefficients for omega are noisy (Figs. 5i–l) and predictive skill using this variable is erratic (Fig. 2c). Additional analysis (not shown) suggests these effects are a result of the spatial heterogeneity in the omega field, which can lead to instability in the regression model fit and predictions (as compared to smoother fields like SST and HGT).

Results from the composite analysis for the difference between historical extreme wet and dry events are shown in Figs. 7 and 8. These are consistent with the patterns seen in the individual coefficient maps but aid in highlighting the important features. During the previous short rain season (S4; Figs. 7d and 8d), anomalous tropical westerlies at low levels and easterlies at high levels are observed and associated with a tropical Indian Ocean SSTA gradient, with cold waters in the west and warm anomalies in the eastern Indian Ocean. The Sumatra region is dominated by negative omega anomalies, or enhanced uplifting, which is consistent

with the local warm-water pool and negative height anomalies at that time. The low-level pressure gradient and uplifting over Sumatra are also associated with anomalous surface westerlies that advect moist air away from East Africa, reducing the short rains.

During S3 (Figs. 7c and 8c), a second dipole of SST anomalies appears over the central and southern Indian Ocean southeast of Madagascar (the Mascarene high area). A similar dipole pattern of height is observed. The warm SST pole is also dominated by anomalously positive height with a low-level anticyclonic wind pattern forming around it. During S2 (Figs. 7b and 8b), the dipole of SST dissipates, and cold water continues to accumulate over the tropical eastern Indian Ocean.

During S1 (Figs. 7a and 8a), the anomalous easterly winds at low levels on the northern extent of the anticyclone reach the equator and strengthen. Over this same timeframe, cold water keeps accumulating over the tropical eastern Indian Ocean near the Sumatra

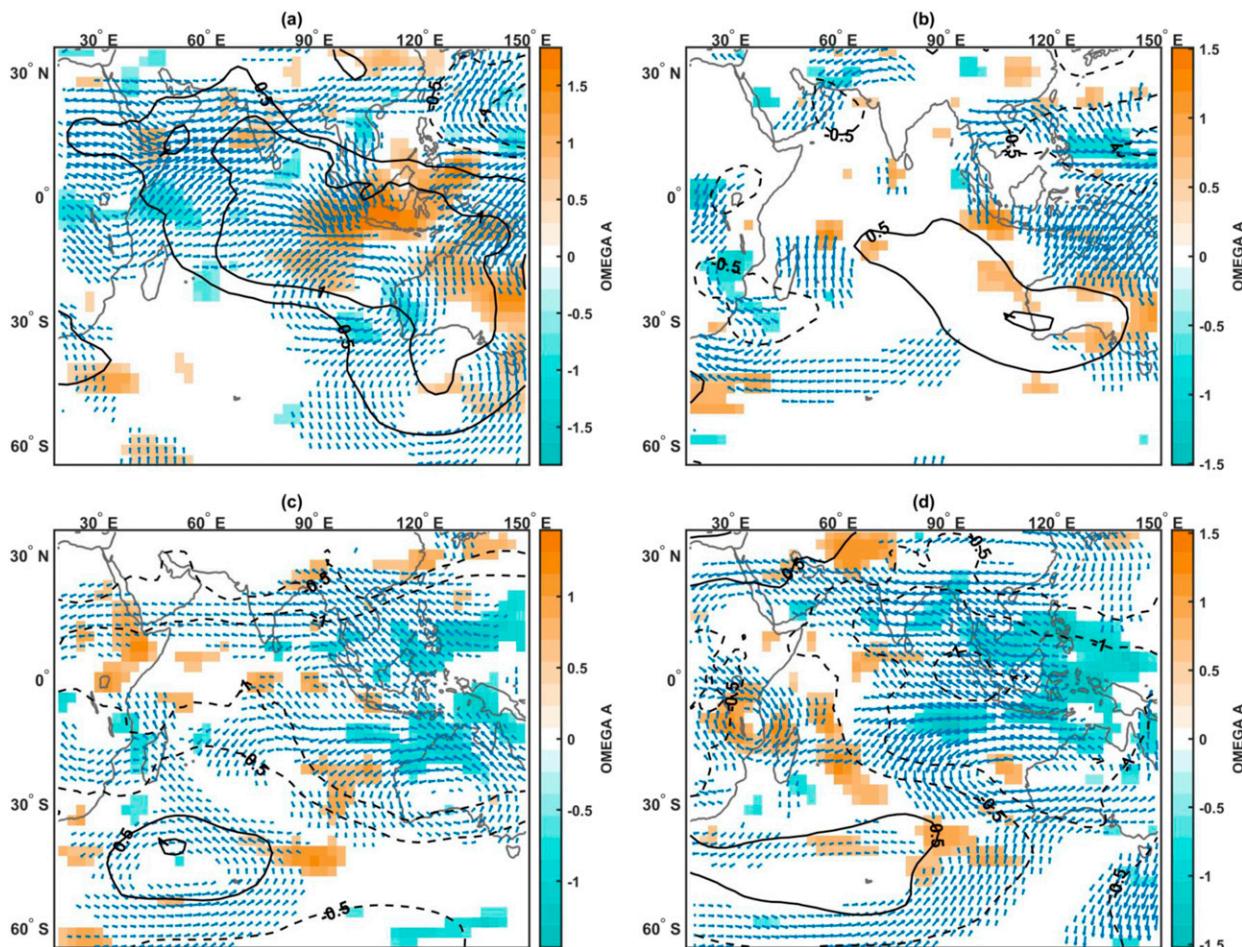


FIG. 8. As in Fig. 7, but for 500-hPa omega (shading), 850-hPa HGT (contours), and 200-hPa winds (arrows).

region, with mild, warm SSTAs in the western basin. Anomalous positive height anomalies dominate the tropical eastern Indian Ocean around Sumatra prior to the short rain season. Positive omega anomalies (subsiding air) are also observed over Sumatra while negative omega anomalies (uplifting air) are observed over the Horn of Africa. At high levels, the tropical region is dominated by anomalous westerlies. These anomalous conditions are all consistent with perturbations to the Walker circulation that have been proposed previously as controlling factors for the short rains and useful measures for seasonal forecast development (Mutai et al. 1998; Hastenrath et al. 2004; Behera et al. 2005; Nicholson 2014).

Similar composite maps are also examined for wet extreme and dry extreme composites separately (Figs. S3–S6 in the online supplemental material). Overall, the average structures of the wettest and driest seasons appear relatively symmetric and represent opposite phases of the “wet–dry” composite maps. Some noticeable

differences are that in the wet-only events, the dipole of SSTAs is much more significant than that in the dry-only events, when the SSTAs over the Mascarene high area are relatively muted. In general, the long-lead progressions averaged over wet and dry short rain seasons tend to mirror each other.

From the analysis given above, extreme short rain events appear to be associated with dipoles of SST and height anomalies and an anticyclone-like wind pattern at long lead times (S3, 7–9 months in advance). We interpret the dipole of SST and height anomalies that peak during S3 as the manifestation of an intensified Mascarene high. The eastern branch of that circulation advects cold water to the eastern Indian Ocean near Sumatra from high latitudes, while the western branch of the low-level high advects equatorial warm water farther south. As the dipole mode dissipates during S2, we draw on the theory proposed by Black et al. (2003) to explain how processes maintain or even enhance the cold-water pool over the eastern Indian Ocean.

We hypothesize that the local cold-water pool in that region modulates the Hadley circulation, perturbing meridional winds at a larger scale. This perturbed meridional circulation induces anomalous southerlies that keep transporting cold water to the eastern Indian Ocean from high latitudes. This process is also consistent with our observations during S2 and S1 when anomalous southerlies are observed over the eastern tropical Indian Ocean in the wet–dry composites, along with subsiding air over Australia near 30°S. The enhanced cold pool around Sumatra then exerts a positive feedback to the anomalous Walker circulation, which provides the strongest source of predictability for the short rains 0–3 months (S1) in advance.

*d. The lead–lag analysis between large-scale climate indices*

To further test the proposed causal chain, three indices are defined based on key features of the regression coefficient maps and climate composites from the analyses above (Fig. 9). These include a long-lead SST dipole index (SL), a long-lead pressure dipole index (HL), and a long-lead meridional wind index (VL). The SL is defined as the difference between spatially averaged SSTs in two regions: 50°–35°S, 40°–80°E ( $SL_p$ ) and 0°–20°N, 50°–100°E ( $SL_n$ ). The HL is defined as the difference between spatially averaged 850-hPa geopotential heights in those same two regions. The SL and HL are used to represent dipoles in SSTs and heights that emerge with an anomalous Mascarene high during S3. The VL is defined as the spatially averaged 850-hPa meridional wind in the region 15°S–5°N, 90°–100°E (VL) and is used to quantify anomalous southerlies that transport cold water from higher latitudes during S2–S1. As a measure of the Walker circulation during S1, a short-lead SST dipole index (SS) based on Hastenrath et al. (2004) is defined as the difference between averaged SSTs in 5°S–10°N, 45°–55°E ( $SS_p$ ) and 15°S–5°S, 90°–110°E ( $SS_n$ ).

Figure 10 shows the lead–lag correlations of these different climate indices. Here, correlations in the upper left or lower right are associated with situations in which the index on the vertical-axis leads or lags, respectively, the index on the horizontal axis. Figure 10a demonstrates a strong statistical coupling between the Mascarene high–related dipole in SSTs (SL) and heights (HL) at lead times between 6 and 10 months prior to the short rains. The correlations indicate that anomalies in heights precede those in SSTs by between 1 and 3 months, suggesting that the SST anomalies are forced by atmospheric anomalies. The comparison between SL and VL in Fig. 10b highlights that the long-lead dipole in SSTs for much of the year preceding the short rains, but

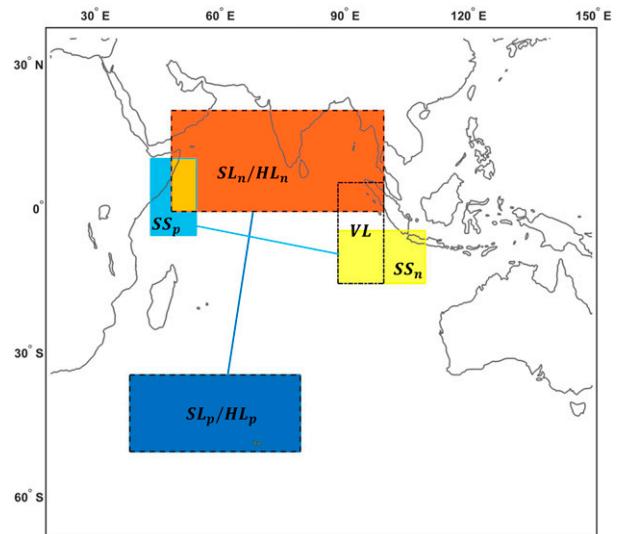


FIG. 9. Regions used to define large-scale climate indices used in the lead–lag analysis, including the long-lead SST dipole index ( $SL_p$  and  $SL_n$ ), the long-lead height dipole index ( $HL_p$  and  $HL_n$ ), the long-lead meridional wind index (VL), and the short-lead SST dipole index ( $SS_p$  and  $SS_n$ ).

especially at a lag of 7–10 months (December–March), is related to anomalous southerly winds in August and May (2 and 5 months prior to the start of the short rains, respectively). Figure 10c then shows that those southerly winds exhibit strong concurrent or slightly lagged relationships with the SS that provides a measure of the anomalous Walker cell. This is consistent with the theory that the Walker circulation is enhanced during S1 by persistent southerlies transporting cold water from higher latitudes to the eastern tropical Indian Ocean. We also see that the SS and VL indices are correlated in August and May (for VL leading SS and vice versa), suggesting a link between the southerly winds and the enhanced Walker cell during the transitional S2 season. We also note that the long-lead SL index is highly correlated with the short-lead SS index, with the correlation reaching 0.58 when the SL index leads the SS index by 8 months (not shown).

## 5. Conclusions

In short, we propose a causal chain of physical processes that explains the high predictive skill of EASR anomalies at lead times of 7–9 months (S3) (explained here for wet events):

- 1) A strengthened Mascarene high first appears and forms dipoles of SST and height anomalies during lead times of 7–9 months (Figs. 5c,g and 7c).
- 2) The associated dipole mode triggers an anticyclone-like wind anomaly that advects cold water to the

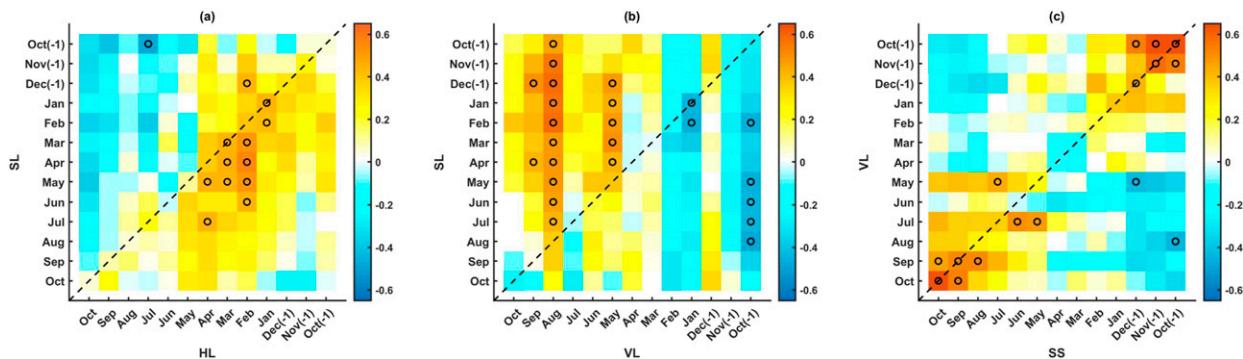


FIG. 10. Correlation coefficients (Pearson’s) at varying lead times prior to the short-rain season between (a) SL and HL, (b) SL and VL, and (c) VL and SS. The significant correlations are marked with open black circles (one-tailed  $p = 0.001$ ). A “(-1)” in the axis label indicates the month from the previous calendar year.

- eastern Indian Ocean region from high latitudes during lead times of 4–6 months (Figs. 6b,c and 7b,c).
- 3) The cold water over the eastern Indian Ocean perturbs the local Hadley circulation, which further perturbs the meridional circulation at larger scales, enhancing the cold-water pool at lead times of 4–6 months (Black et al. 2003).
  - 4) The cold-water pool modulates the Walker circulation by inducing anomalous easterlies, bringing moist air to land, and enhancing uplift and convection over East Africa at lead times of 0–3 months (Figs. 6a and 7a).

These patterns are identified by a regularized regression model, in which high predictive skill is observed for lead times of 0–3 months (S1) and lead times of 7–9 months (S3). Maps of regression coefficients show that certain groups of predictors were selected for those two lead times: at lead times of 0–3 months (S1), the model selected cold water over the Sumatra region and anomalous easterlies at 850 hPa, while at lead times of 7–9 months (S3), the model selected the dipole of SSTs with a warm pole over the Mascarene high area and a cold pole over the northern and central Indian Ocean. Consistent patterns are also observed in the composite analysis, in which large-scale climate fields prior to historical extremes are examined and compared. A set of lead–lag analyses also supported the proposed causal chain. Our findings of high predictability at lead times of 7–9 months suggest that there is significant potential to extend lead times in forecast models for East Africa, which could prove valuable for water management in the region.

While the results of this empirical study are promising, this study was limited to 37 years of observations. There is evidence that teleconnection between the Indian Ocean and EASR are nonstationary (Nicholson 2015; Bahaga et al. 2019). Therefore, further analysis using

general circulation model (GCM)-based experiments is needed to confirm the dynamic pathways linking the two stages of high predictability and different climate fields. Such an experiment would require a modeling exercise that activates or suppresses SST and pressure anomalies associated with the Mascarene high, and then determines whether the activation of Mascarene high anomalies in the model are associated with the dynamic pathways in the Indian Ocean basin we identify in this study. This effort is left for future work.

*Acknowledgments.* The rainfall data were collected from the CHIRPS (<http://chg.geog.ucsb.edu/data/chirps/>), and data for large-scale climate fields including geopotential heights,  $u/v$  winds, and vertical velocity of the atmosphere were collected from the NCEP–NCAR Reanalysis II, which is provided by the NOAA/OAR/ESRL PSD (<https://www.esrl.noaa.gov/psd/>). The SST data were collected from the HadISST dataset (<https://www.metoffice.gov.uk/hadobs/hadisst/>). This work was partially supported by the National Science Foundation under Grant ECCS-1556900.

REFERENCES

Bahaga, T. K., G. Mengistu Tsidu, F. Kucharski, and G. T. Diro, 2015: Potential predictability of the sea-surface temperature forced equatorial East African short rains interannual variability in the 20th century. *Quart. J. Roy. Meteor. Soc.*, **141**, 16–26, <https://doi.org/10.1002/qj.2338>.

—, A. H. Fink, and P. Knippertz, 2019: Revisiting interannual to decadal teleconnections influencing seasonal rainfall in the greater horn of Africa during the 20th century. *Int. J. Climatol.*, **39**, 2765–2785, <https://doi.org/10.1002/joc.5986>.

Barnett, B. J., C. B. Barrett, and J. R. Skees, 2008: Poverty traps and index-based risk transfer products. *World Dev.*, **36**, 1766–1785, <https://doi.org/10.1016/j.worlddev.2007.10.016>.

Behera, S. K., J.-J. Luo, S. Masson, P. Delecluse, S. Gualdi, A. Navarra, and T. Yamagata, 2005: Paramount impact of the Indian Ocean dipole on the East African short rains: A CGCM study. *J. Climate*, **18**, 4514–4530, <https://doi.org/10.1175/JCLI3541.1>.

- Black, E., J. Slingo, and K. R. Sperber, 2003: An observational study of the relationship between excessively strong short rains in coastal East Africa and Indian Ocean SST. *Mon. Wea. Rev.*, **131**, 74–94, [https://doi.org/10.1175/1520-0493\(2003\)131<0074:AOSOTR>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<0074:AOSOTR>2.0.CO;2).
- Breiman, L., and P. Spector, 1992: Submodel selection and evaluation in regression. The  $X$ -random case. *Int. Stat. Rev.*, **60**, 291–319, <https://doi.org/10.2307/1403680>.
- Clark, C. O., P. J. Webster, and J. E. Cole, 2003: Interdecadal variability of the relationship between the Indian Ocean zonal mode and East African coastal rainfall anomalies. *J. Climate*, **16**, 548–554, [https://doi.org/10.1175/1520-0442\(2003\)016<0548:IVOTRB>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<0548:IVOTRB>2.0.CO;2).
- Funk, C., and Coauthors, 2015: The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Sci. Data*, **2**, 150066, <https://doi.org/10.1038/SDATA.2015.66>.
- Hastenrath, S., 2007: Circulation mechanisms of climate anomalies in East Africa and the equatorial Indian Ocean. *Dyn. Atmos. Oceans*, **43**, 25–35, <https://doi.org/10.1016/j.dynatmoce.2006.06.002>.
- , A. Nicklis, and L. Greischar, 1993: Atmospheric-hydro-spheric mechanisms of climate anomalies in the western equatorial Indian Ocean. *J. Geophys. Res.*, **98**, 20 219–20 235, <https://doi.org/10.1029/93JC02330>.
- , D. Polzin, and P. Camberlin, 2004: Exploring the predictability of the ‘short rains’ at the coast of East Africa. *Int. J. Climatol.*, **24**, 1333–1343, <https://doi.org/10.1002/joc.1070>.
- , —, and C. Mutai, 2011: Circulation mechanisms of Kenya rainfall anomalies. *J. Climate*, **24**, 404–412, <https://doi.org/10.1175/2010JCLI3599.1>.
- Hillier, D., and B. Dempsey, 2012: A dangerous delay: The cost of late response to early warnings in the 2011 drought in the Horn of Africa. Oxfam International Briefing Paper, 34 pp., <http://policy-practice.oxfam.org.uk/publications/a-dangerous-delay-the-cost-of-late-response-to-early-warnings-in-the-2011-droug-203389>.
- Hoerl, A. E., and R. W. Kennard, 1970: Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, **12**, 55–67, <https://doi.org/10.1080/00401706.1970.10488634>.
- Javanmard, A., and A. Montanari, 2014: Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.*, **15**, 2869–2909.
- Kimani, M. W., J. C. B. Hoedjes, and Z. Su, 2017: An assessment of satellite-derived rainfall products relative to ground observations over East Africa. *Remote Sens.*, **9**, 430, <https://doi.org/10.3390/rs9050430>.
- Manatsa, D., Y. Morioka, S. K. Behera, C. H. Matarira, and T. Yamagata, 2014: Impact of Mascarene high variability on the East African ‘short rains.’ *Climate Dyn.*, **42**, 1259–1274, <https://doi.org/10.1007/s00382-013-1848-z>.
- McKee, T. B., N. J. Doesken, and J. Kleist, 1993: The relationship of drought frequency and duration to time scales. *Proc. Eighth Conf. on Applied Climatology*, Anaheim, CA, Amer. Meteor. Soc., 179–184.
- Mutai, C. C., M. N. Ward, and A. W. Colman, 1998: Towards the prediction of the East Africa short rains based on sea-surface temperature–atmosphere coupling. *Int. J. Climatol.*, **18**, 975–997, [https://doi.org/10.1002/\(SICI\)1097-0088\(199807\)18:9<975::AID-JOC259>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1097-0088(199807)18:9<975::AID-JOC259>3.0.CO;2-U).
- 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, <https://doi.org/10.1175/JHM-D-13-062.1>.
- , 2015: Long-term variability of the East African ‘short rains’ and its links to large-scale factors. *Int. J. Climatol.*, **35**, 3979–3990, <https://doi.org/10.1002/joc.4259>.
- 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, <https://doi.org/10.1029/2002JD002670>.
- Saji, N. H., B. N. Goswami, P. N. Vinayachandran, and T. Yamagata, 1999: A dipole mode in the tropical Indian Ocean. *Nature*, **401**, 360–363, <https://doi.org/10.1038/43854>.
- Tibshirani, R., 1996: Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc.*, **58B**, 267–288, <https://doi.org/10.1111/J.2517-6161.1996.TB02080.X>.
- Tierney, J. E., J. E. Smerdon, K. J. Anchukaitis, and R. Seager, 2013: Multidecadal variability in East African hydroclimate controlled by the Indian Ocean. *Nature*, **493**, 389–392, <https://doi.org/10.1038/nature11785>.
- Tozuka, T., S. Endo, and T. Yamagata, 2016: Anomalous walker circulations associated with two flavors of the Indian Ocean dipole. *Geophys. Res. Lett.*, **43**, 5378–5384, <https://doi.org/10.1002/2016GL068639>.
- Turner, A. G., and H. Annamalai, 2012: Climate change and the South Asian summer monsoon. *Nat. Climate Change*, **2**, 587–595, <https://doi.org/10.1038/nclimate1495>.
- Ummenhofer, C. C., A. Sen Gupta, M. H. England, and C. J. C. Reason, 2009: Contributions of Indian Ocean sea surface temperatures to enhanced East African rainfall. *J. Climate*, **22**, 993–1013, <https://doi.org/10.1175/2008JCLI2493.1>.
- Walker, D. P., C. E. Birch, J. H. Marsham, A. A. Scaife, R. J. Graham, and Z. T. Segele, 2019: Skill of dynamical and GHACOF consensus seasonal forecasts of East African rainfall. *Climate Dyn.*, **53**, 4911–4935, <https://doi.org/10.1007/S00382-019-04835-9>.
- Xue, F., H. J. Wang, and J. H. He, 2004: Interannual variability of Mascarene high and Australian high and their influences on East Asian summer monsoon. *J. Meteor. Soc. Japan*, **82**, 1173–1186, <https://doi.org/10.2151/jmsj.2004.1173>.
- Yang, W., R. Seager, M. A. Cane, and B. Lyon, 2015: The annual cycle of East African precipitation. *J. Climate*, **28**, 2385–2404, <https://doi.org/10.1175/JCLI-D-14-00484.1>.
- Zou, H., and T. Hastie, 2005: Regularization and variable selection via the elastic net. *J. Roy. Stat. Soc.*, **67B**, 301–320, <https://doi.org/10.1111/j.1467-9868.2005.00503.x>.