Predictability of the Normalized Difference Vegetation Index in Kenya and Potential Applications as an Indicator of Rift Valley Fever Outbreaks in the Greater Horn of Africa

MATAYO INDEJE AND M. NEIL WARD

International Research Institute for Climate and Society, The Earth Institute at Columbia University, Palisades, New York

LABAN J. OGALLO

IGAD Climate Prediction and Application Centre, Nairobi, Kenya

GLYN DAVIES

Late SVRO, Veterinary Research Laboratory, Kabete, Kenya

MAXX DILLEY

International Research Institute for Climate and Society, The Earth Institute at Columbia University, Palisades, New York

ASSAF ANYAMBA

UMBC Goddard Earth Sciences Technology Center, and Biospheric Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland

(Manuscript received 24 June 2004, in final form 31 August 2005)

ABSTRACT

In this paper the progress made in producing predictions of the Normalized Difference Vegetation Index (NDVI) over Kenya in the Greater Horn of Africa (GHA) for the October–December (OND) season is discussed. Several studies have identified a statistically significant relationship between rainfall and NDVI in the region. Predictability of seasonal rainfall by global climate models (GCMs) during the OND season over the GHA has also been established as being among the best in the world. Information was extracted from GCM seasonal prediction output using statistical transformations. The extracted information was then used in the prediction of NDVI. NDVI is a key variable for management of various climate-sensitive problems. For example, it has been shown to have the potential to predict environmental conditions associated with Rift Valley Fever (RVF) viral activity and this is referred to throughout the paper as a motivation for the study. RVF affects humans and livestock and is particularly economically important in the GHA. The establishment of predictability for NDVI in this paper is therefore part of a methodology that could ultimately generate information useful for managing RVF in livestock in the GHA. It has been shown that NDVI can be predicted skillfully over the GHA with a 2–3-month lead time. Such information is crucial for tailoring forecast information to support RVF monitoring and prediction over the region, as well as many other potential applications (e.g., livestock forage estimation). More generally, the Famine Early Warning System (FEWS), a project of the U.S. Agency for International Development (USAID) and the National Aeronautics and Space Administration (NASA) and other specialized technical centers routinely use NDVI images to monitor environmental conditions worldwide. The high predictability for NDVI established in this paper could therefore supplement the routine monitoring of environmental conditions for a wide range of applications.

Corresponding author address: Matayo Indeje, International Research Institute for Climate and Society (IRI), 61 Route 9W, Palisades, NY 10964.
E-mail: mindeje@iri.columbia.edu

1. Introduction

The initial motivation for the current investigation of Normalized Difference Vegetation Index (NDVI) predictability is its potential for contributing to the stabilization of the livestock trade between the Greater Horn of Africa (GHA) and the Middle East. Cases of Rift Valley Fever (RVF), a vector-borne disease, have been reported in parts of Africa since the 1950s (Davies et al. 1985; Swanepoel 1981). The disease was first identified in Kenya in 1931, when a target flock of exotic sheep kept in the Rift Valley suffered severe losses. The RVF virus is transmitted by mosquitoes of the genus *Aedes*, which breed in flooded low-lying habitats known as dambos (Meegan and Bailey 1989). Dambo depressions are common in many parts of Africa (Davies et al. 1985). RVF causes epizootics (large-scale transmission) in domestic animals and epidemics in human populations closely associated with infected animals. Outbreaks of RVF in recent years have been accompanied by bans on livestock trade between the GHA and the Middle East. RVF outbreaks and trade bans since the 1997/98 El Niño event have cost the GHA \$300–\$500 million annually [Organization of African Union/Inter-African Bureau of Animal Resources (OAU/IBAR) 2003, personal communication). Livestock accounts for a significant percentage of the gross domestic product (GDP) in many GHA countries. RVF-related trade stoppages have had severely detrimental impacts on livelihoods of many pastoralists.

Outbreaks of RVF in Africa are characterized by distinct spatial and temporal patterns that are directly related to specific environmental parameters associated with mosquito vectors. Depending on their status, environmental variables such as vegetation, soil moisture, and temperature, maintain endemic levels of the virus and/or promote epizootic level of transmission (Linthicum et al. 1990). The NDVI is a measure of vegetation greenness and is a good proxy for rainfall and soil moisture (Tucker 1979; Tucker et al. 1986).

RVF disease has been reported in Kenya at intervals of 3–12 yr, mostly in the Eastern and Western Highland plateau areas at the coast. Outbreaks were reported in the years 1951–53, 1961–63, 1967/68, 1977–79, 1982/83, and 1997/98 (Davies et al. 1985; Linthicum et al. 1999). The most recent major outbreak during late 1997 to early 1998 has been linked to the heavy and prolonged rains associated with El Niño–Southern Oscillation (ENSO; Trenberth 1998; Linthicum et al. 1999). Most reported cases during this period were confined to the semiarid zones in the north and northeast of Kenya, similar to the 1961/62 episode.

Efforts have been made to monitor and forecast possible locations of RVF epizootics in order to undertake surveillance and control measures in advance. Remotely sensed NDVI data have been used to monitor RVF in restricted areas in Kenya where epizootics occur. Surveillance conducted within one ecological zone in the country (Linthicum et al. 1990) concluded that NDVI values of 0.43 and above corresponded to at least short-term flooding of mosquito-breeding dambo habitats. A synthetic aperture radar (SAR) instrument that provides high-resolution and sensitive ground moisture assessment was used to identify these mosquito-breeding habitats.

Various studies have reported on the relationship in various parts of Africa between rainfall and NDVI (Davenport and Nicholson 1993; Tucker et al. 1991; Hielkema et al. 1986) and between NDVI and ENSO (Anyamba and Eastman 1996; Verdin et al. 1999; among others). The NDVI in the semiarid zones of eastern Africa has been identified to have a good relationship with rainfall (Davenport and Nicholson 1993). Moreover in these semiarid zones, the use of the NDVI is optimized according to the advantages and shortcomings of the rainfall data (Anyamba et al. 2002). Davenport and Nicholson (1993) showed strong similarity between temporal and spatial patterns of NDVI and rainfall when annual rainfall was below l000 mm and monthly rainfall below 200 mm. In this range, the good relationship of rainfall and NDVI was maintained for interannual variability, as well as resolving the mean annual cycle and spatial patterns. They further established that the overall relation between NDVI and rainfall was log linear and the correlation between annually integrated NDVI and the log of annual rainfall was about 0.89. The best association on a monthly scale was between December NDVI and the average of rainfall in the concurrent (December) and the two previous months (October and November). The higher correlations with 3-month averages than individual month indicates that NDVI is likely a better integrator of soil moisture conditions than of rainfall alone.

Periods of RVF epizootic activity have correlated with persistent and excessive rainfall, with an apparent lag that allows 1–2 months of warning of the disease through monitoring of rainfall trends (Davies et al. 1985). This relationship was observed at one study site (1°12′S, 37°E) in Kenya. Linthicum et al. (1999) used an Autoregressive Integrated Moving Average (ARIMA) model to determine the best predictors of RVF activity among various combinations of the Southern Oscillation index (SOI), equatorial Pacific and Indian Ocean sea surface temperatures (SSTs), and NDVI. The best associations to the outbreak data were achieved when equatorial Pacific and Indian Ocean

SST and NDVI anomaly data were used in the model. This approach could have been used to successfully predict each of the three RVF outbreaks that occurred between 1982 and 1998 at a study site in Kenya. However, the approach of using multiple dependent variables, which are highly correlated, has a problem of overfitting and poorly estimating the statistical models and hence generating false skill.

In our approach to forecasting NDVI values, we use a method based on empirical orthogonal function (EOF) analysis that enables fields of highly correlated data to be represented adequately by a small number of uncorrelated fields, which account for much of the variance in the spatial and temporal variability of the original data. Using the EOF predictor variables will substantially eliminate the problem of overfitting and allow identification of more robust statistical models that can be expected to perform comparably when applied to historic data or in future real-time situations. We further perform a cross validation to test the stability of the constructed regression model. Cross validation ensures that observations from the forecast period do not directly influence predictions while allowing for efficient use of limited data (Stone 1974). The aim is for these forecasting approaches to complement ongoing monitoring activities. For example, the National Aeronautics and Space Administration (NASA) routinely monitors the RVF episodes in Africa using the evolution of NDVI (Anyamba et al. 2003; more information available online at http://www.geis.fhp.osd.mil/GEIS/SurveillanceActivities/RVFWeb/infopages/updateRVF.asp).

Seasonal climate forecasting is an emerging science with the potential to inform decision makers of adverse seasonal climates (e.g., drought, excess rainfall) months prior to their actual occurrence. Forecast accuracy and reliability vary according to geographic region, season, and year and their usefulness is dependent on whether or not the limited information they provide can be understood and used effectively for specific sectoral decisions (Thomson et al. 2003). Improvements in the understanding of interactions between the atmosphere and sea and land surfaces, advances in modeling the global climate system, and substantial investment in monitoring the tropical oceans now provide a degree of predictability of climate fluctuations at a seasonal (i.e., a few months) lead time in many parts of the world, especially in the Tropics (Latif et al. 1998; Goddard et al. 2001). Climate predictability is highly dependent on the extent to which the regional climate is determined by SST patterns of the global, and particularly the tropical oceans.

In this study we use rainfall and circulation output from a state-of-the-science global climate model (GCM) driven with observed and predicted SST patterns. The output of the model is used to predict NDVI. We show that climate model output can be used to skillfully predict NDVI in Kenya with a lead time of a few months. NDVI is highly dependent on soil moisture conditions that in turn, are dependent on rainfall and other factors including soil type and elevation. The ability to predict atmospheric circulation patterns and associated rainfall with good skill over the region, and "downscale" it into high-resolution NDVI information, will build confidence in predictions of the patterns of RVF risk a few months in advance. It will also provide the basis for exploring the enhanced management of other problems related to NDVI variations in the region.

2. Data

The GHA region exhibits three rainfall seasons of March–May (MAM; long rains), June–August (JJA; experienced over the northern parts of the subregion), and October–December (OND; short rains). These rainfall patterns are characterized by the north–south movement of the intertropical convergence zone (ITCZ), a zone of confluence of air currents from the north and south over the African continent. This rainfall belt follows the position of the sun during the annual cycle.

We selected the short rains season of OND for several reasons: 1) it is predicted with better skill over Kenya and the entire GHA compared to the other two seasons, and is also highly related to ENSO (Ropelewski and Halpert 1987; Ogallo et al. 1988; Farmer 1988; Nicholson 1996; Indeje et al. 2000; Camberlin et al. 2001), and 2) rainfall reliability in this season is crucial and more important to the communities living in the arid and semiarid lands (ASALs) of Kenya (Fig. 1) who mostly depend on livestock for subsistence. Vegetation production is highly variable in this semiarid environment of Kenya due in part to its sensitivity to year-to-year variability in the amount and timing of rainfall. Figure 1 shows NDVI variability over Kenya, expressed as the interannual standard deviation of the NDVI values. The map pattern shows areas of low variability of NDVI (highlands), and high variability (ASALS). Sufficiently high rainfall amounts observed over the highland areas of Kenya sustain vegetation cover by maintaining soil moisture conditions, which results in low NDVI variability over the region. Previous studies have also indicated high (low) correlation between rainfall and NDVI in the ASALS (highlands) areas of Kenya (Davenport and Nicholson 1993). Davenport and Nicholson (1993) postulated that as rainfall increases,

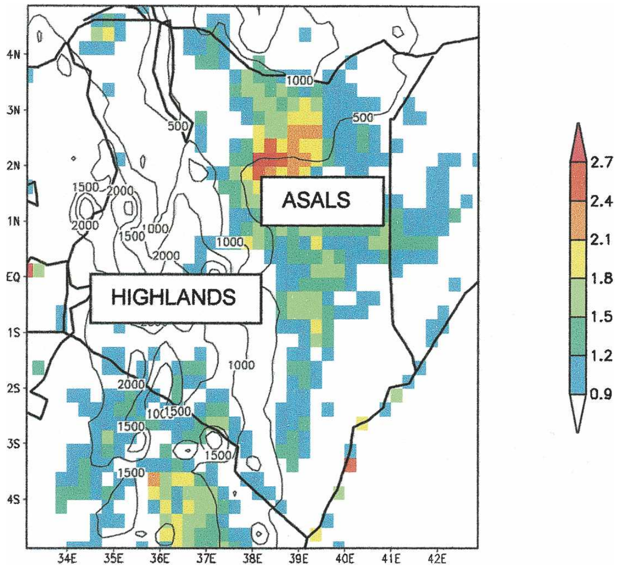

FIG. 1. NDVI variability (1982–98) for December over Kenya. NDVI variability values (shaded) are dimensionless. Contours are elevation in meters. ASALS receive below 500 mm of rainfall yr^{-1}. Highland areas receive more that 2000 mm yr^{-1}.

NDVI increases until some threshold is reached and then remains constant thereafter despite any further increase in rainfall. We used the new updated and recalibrated NDVI (version 3) dataset for Africa processed by the Global Monitoring and Modeling Systems (GIMMS) group at NASA's Goddard Space Flight Center (available online at http://islscp2.sesda.com/ISLSCP2_1/html_pages/groups/veg/gimms_ndvi_monthly_xdeg.html). The data have been derived from measurements made by the Advanced Very High-Resolution Radiometer (AVHRR) instrument on polar-orbiting meteorological satellites operated by the National Oceanic and Atmospheric Administration (NOAA). NDVI is mapped on Albers equal-area projection and has been calibrated for intrasensor differences and intrasensor degradation. The data have been corrected for El Chichón and Mt. Pinatubo volcanic events to remove bias associated with volcanic aerosol contamination. Spatial and temporal correction has been applied to the data to remove cloud pixels, which is then updated in the historical archives by the U.S. Geological Survey (USGS). The new NASA Global Inventory Modeling and Mapping Studies (GIMMS) operational data have a greater maximum NDVI value of 0.90 versus 0.75 for the predecessor dataset. The temporal range of the NDVI data we used is from January 1982 to December 1998. Monthly data series were derived from the 10-day data by maximum value compositing (Holben 1986), and then interpolated from 8-km pixel resolution to a 25-km grid. Bicubic (Matlab interpolation function) was used in transforming NDVI from 8 to 25 km. The scheme generates each target grid by interpolation from the nearest 16 mapped source grid boxes. The algorithm performance has been proved to be quite satisfactory for computer execution time and memory usage. It also yields a good balance between accuracy in detail preservation and smoothness. The 25-km grid spacing was chosen in order to avoid overloading the processing software while at the same time retaining a resolution capable of describing NDVI variability in the region of interest.

Monthly averages were calculated from the NDVI time series for the 25-km grid boxes coinciding with locations over eastern Kenya where RVF cases have been reported. A total of 40 grid boxes were used in the averaging. December NDVI is used to infer the effect of OND rainfall on the vegetation development over the region. There are several cases of RVF reports in the GHA that coincided with the OND rainfall season with some cases reported in the following year. This analysis of end-of-year NDVI values could provide evidence on expected reports of RVF cases in the same year or early in the following year. Linthicum et al. (1999) reported elevated NDVI anomalies over East Africa starting in October 1997 (the start of the normal short rainy period) and extending to April 1998 (through the normal dry season of January and February), which were significantly correlated with RVF activity. Likewise, investigation of NDVI images of Kenya for OND 1982–84 (Linthicum et al. 1987), showed intense green vegetation in the December 1982 in the central portion of Kenya that was substantially higher than the corresponding period in the two subsequent years. RVF disease was also reported in the country during this period (OND 1982–84). Locations of RVF epizootics and NDVI values for December 1982 and 1997 are shown in Fig. 2. The figure also shows the location near Nairobi where long records of RVF epizootics exist (Davies et al. 1985; Linthicum et al. 1999) and where the 0.43 NDVI threshhold was identified. The NDVI resolution at 25-km grid spacing (a total of 40 by 40 grid points) was used in this figure. High values of NDVI were observed during this period over the ASALS and parts of the highland areas of Kenya. Interestingly, the NDVI values over the highlands where there were RVF outbreaks are higher than in the surrounding ones. Similar observations are reported in Linthicum et al. (1987), with the inference that NDVI is an indicator of areas of potential RVF outbreaks.

3. Methods

An investigation into predictability is conducted in order to establish the degree to which GCM output can

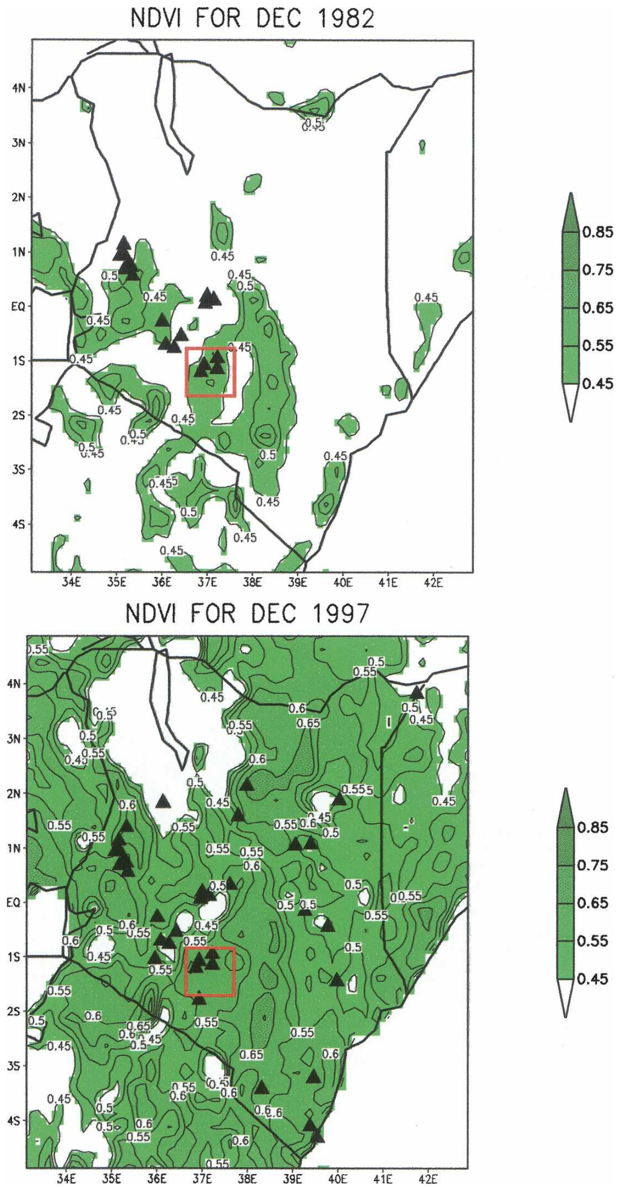

FIG. 2. NDVI values for December (a) 1982 and (b) 1997. Filled triangles are locations where RVF cases were reported. The square box is the location near Nairobi with a long record of RVF epizootics (Davies et al. 1985; Linthicum et al. 1999). A suggestion of an environmental threshold for RVF of 0.43 NDVI was based on this one location.

identify NDVI conditions in advance. Statistical transformation of GCM output is necessary to correct systematic biases between the real world and its modeled presentation. Such an approach could be utilized to construct relationships between the desired forecast quantity such as NDVI- and GCM-simulated variables such as precipitation, large-scale circulation, etc. In the current study, NDVI data are forecast using their statistical association with GCM-predicted rainfall and circulation variables. Climate forecast fields are provided from the ECHAM v4.5 (Roeckner et al. 1996) GCM. We used output from simulations utilized by the International Research Institute for Climate and Society (see online at http://iri.columbia.edu/) as input to their operational seasonal forecasts. We analyzed the mean of an ensemble of 24 GCM integrations, each run with different initial atmospheric conditions but the same SST boundary conditions, over the 17-yr period (1982–98). The method is repeated by using GCM forecasts made from persisted SST anomaly conditions. In the latter case, the GCM is forced by the September SST anomaly conditions that are then persisted through the forecast period (until the end of December). The persisted SST GCM forecasts enable a 0-month lead time for OND seasonal rainfall and a 2–3-month lead time in December NDVI forecast. The GCM nonlinearly transforms SST information from around the globe to produce a dynamic solution of atmospheric variables over a given region.

One of the shortcomings of the GCM is the coarse resolution of about 280 km that is used. The use of the coarse resolution in the GCM, or other approximations contained in the model equations, could often result in systematic shifts in the location of spatial rainfall patterns to an extent of reducing the overall prediction skill. We therefore apply a correction to the GCM output using the model output statistics (MOS) approach. The MOS concept, widely used in weather forecasting, objectively interprets numerical model output and produces site-specific forecast guidance (Wilks 1995). The MOS method involves matching observed data such as station seasonal precipitation or NDVI with output from numerical model predictions. Forecast equations for a specific region (location) are then derived by statistical techniques including various forms of regression. In this way the bias and spatial inaccuracy of the numerical model, as well as the local climatology, are built into the forecast system. When transforming to a variable such as NDVI, the MOS approach also implicitly represents physical processes that connect the GCM output variable (such as rainfall) to the target variable (such as NDVI). Analysis of GCM circulations patterns (figure not shown) indicates that low-level (700 hPa) zonal wind correlates significantly with rainfall over Kenya. The skill obtained from GCM circulation indices is however lower than that of GCM precipitation. Statistical methods have previously been used to relate GCM output to finer scales suitable for applications (Wilby et al. 2002; Landman and Goddard 2002; Hansen and Indeje 2004). For seasonal predictions, the three commonly used methods (that are methodologically related) are EOF, singular value decomposition (SVD), and canonical correlation analysis (CCA). EOF analysis enables fields of highly correlated data to be represented adequately by a small number of orthogonal functions and corresponding orthogonal time coefficients, which account for much of the variance in their spatial and temporal variability (Kutzbach 1967; Kidson and Thompson 1998). Each principal component (PC) pattern represents a predictor field with spatial coherence, but in a way that efficiently deals with the risk of overfitting the empirical model. The SVD method decomposes a cross-covariance matrix of simulated and observed fields into singular vectors and expansion coefficients. Details of this method and its application to geophysical data analysis are discussed in Feddersen et al. (1999). CCA is a multivariate statistical technique that calculates linear combinations of a set of predictors that maximizes least square relationships to similarly calculated linear combinations of a set of predictands. A limited number of leading principal components of the covariance or correlation matrix that represent sufficient percentage of the variance from the original datasets are retained to further the analysis. CCA can also be viewed as a special form of EOF analysis where the correlation structure between predictor and predictand datasets is described more completely with each successive canonical mode (Graham et al. 1994). The CCA statistical technique was used in this study to relate GCM precipitation and 700-hPa zonal wind predictor to NDVI predictand fields. The SVD technique was tried and gave similar results as those obtained by using the CCA method. The CCA time series obtained from the statistical decomposition of the GCM fields are related to NDVI using multiple regression. The most recommended multilinear regression approach when a relatively long time series is available is to build the model using an independent dataset (the training set) and use the remaining sample in model verification. However, for short length time series (17 yr in our case), model performance is validated using a jackknife or cross-validation technique in which data from one (or more) point(s) in time are systematically withheld from the dataset. A specification model is then derived from the remaining part of the dataset, and the specification is tested on the withheld data. The computational drawback of this method is that instead of performing the model construction only once, we have to repeat it as each year is withheld (Feddersen et al. 1999). The multiple linear regression models were trained with the time series of the leading modes through stepwise screening calibration (Wilks 1995; Kidson and Thompson 1998), in which the contribution of each predictor was evaluated through a cross-validation analysis. Only those that contributed to the cross-validation skill were included in the predictor dataset. This approach reduces the risk of overfitting the models and simultaneously extracts as much useful information as possible from the predictor data (Feddersen et al. 1999).

4. Results

The variables that we analyze in our model include NDVI (predictand) and a combined GCM precipitation and 700-hPa zonal wind (predictor). We performed EOF analysis on the combined GCM rainfall and 700-hPa zonal wind for the period of 1982–98 (17 yr). The EOF analysis was performed on a covariance matrix (of a combined GCM precipitation and 700-hPa zonal wind) with each anomaly field standardized. The resulting standardized values were dimensionless with a zero mean and unit standard deviation. The GCM domain (10°S–10°N, 10°–60°E) was chosen as to include the known regional climate forcing mechanisms from the Indian Ocean and the Congo tropical forest. The EOF method can be sensitive to changes in domain size. Analyses were performed on various domain sizes before settling on the one that we used. The first four principal component (PC1–PC4) time series for the GCM's OND combined precipitation and 700-hPa zonal wind are shown in Fig. 3. The first four modes explain 32.8%, 18.3%, 11.6%, and 8.4% of the variance, respectively—a total of 71.1%. PC1 shows two major peaks in 1982 and 1997, which were the major El Niño years during the study period. PC3 indicate positive peaks during 1982/83, 1988, and 1991–93 and negative peaks during 1984–87, 1989, 1994–98. Years 1982/83, 1991/92, and 1987/88 were associated with El Niños and 1984/85 and 1995/96 with La Niñas. A feature of PC2 is a strong positive loading in 1985 followed by a relatively strong negative loading in 1986. PC4 indicates a positive trend with negative indices in 1980s turning to positive from 1990. Recent studies by Schreck and Semazzi (2004) have identified a similar time series in the East African rainfall variability and have associated it with a global warming trend.

Correlation between NDVI and the PC1–PC4 time series are shown in Fig. 4. Areas of substantial correlation indicate where each of the PCs has predictive power for the NDVI values. Strongest predictive potential is found for PC1. High positive correlations (>0.6) with PC1 are shown over northern Kenya, parts of the coastal strip, and the southern parts of the country, decreasing toward the highland areas. The spatial variations in skill may reflect variations in climate predictability or climate–NDVI coupling over the region. This hypothesis could be investigated further using a high-resolution regional climate model coupled with

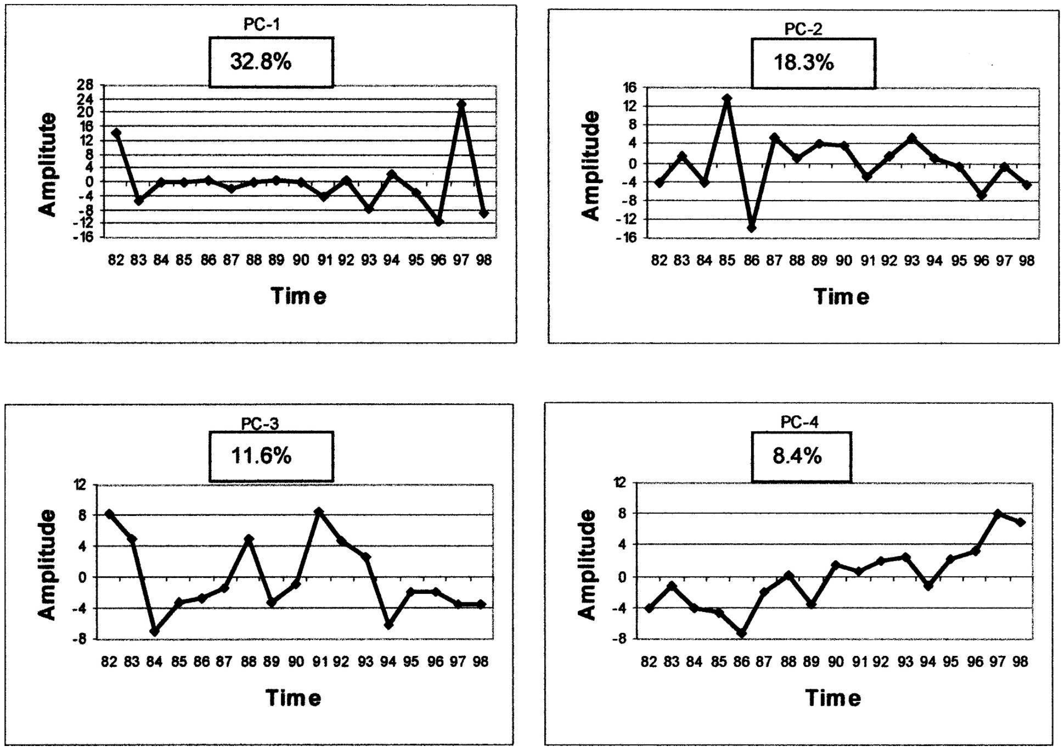

Fig. 3. First four PC time series of OND GCM simulations of precipitation and circulation over 10°S–10°N, 10°–60°E.

station observation data. The subsequent three PCs each show some additional predictive potential. The second PC time series correlates moderately over northern Tanzania. The third PC time series shows moderate correlations over the highland areas, and the southern and coastal strip of Kenya. The PC4 correlation map includes an indication of predictability over the extreme southwest of the domain.

a. Linkage to large-scale SST forcing

The correlation between the PC1 time series and SST is high (>0.4, significant at the 90% level) over the western Indian Ocean and eastern Pacific Ocean (Fig. 5a). This suggests that a tropical Pacific ENSO effect, combined with Indian Ocean forcing, is contributing to high NDVI predictability over Kenya through teleconnection linkages. An east–west gradient is also evident over the Indian Ocean with positive correlations (>0.4) to the west and negative correlations (<−0.4) in the central and eastern parts of the basin. The 1997/98 West Indian Ocean warming during the El Niño of that period could dominate this pattern. However, analysis obtained by omitting 1997 and 1982 in the predictor field through ranked correlation analysis (figure not shown), yielded similar but slightly weaker correlation patterns over the eastern Pacific and western Indian Ocean. The remote forcing of the regional climate is through shifts in the Walker circulation and regionally through warm western Indian Ocean SSTs enhancing near-surface moisture that is transported inland by an easterly wind regime. The advected moisture interacts with the local orography to result in rainfall and corresponding NDVI patterns (Anyamba et al. 2002).

The PC2 time series and SST correlation indicates high positive values (>0.3) to the southern tip of the African continent, and the equatorial and parts of the northern Atlantic (Fig. 5b). This pattern is hypothesized to explain part of the ITCZ's variability and location that is controlled by the north–south pressure gradients that are in turn modulated by surface SST conditions. For instance, warming in the southern Indian and Atlantic Oceans would diffuse the semipermanent Mascarene and St. Helena high pressure cells, restricting the ITCZ to south of the equator, affecting variability in northern Tanzania. High positive correlation is also indicated over the equatorial and the northern Atlantic Ocean with PC4 (Fig. 5d). The PC4 time series matches a positive trend in the NDVI in northern Tanzania. Correlation between GCM precipitation simulation PC3 and SST shows negative values in the

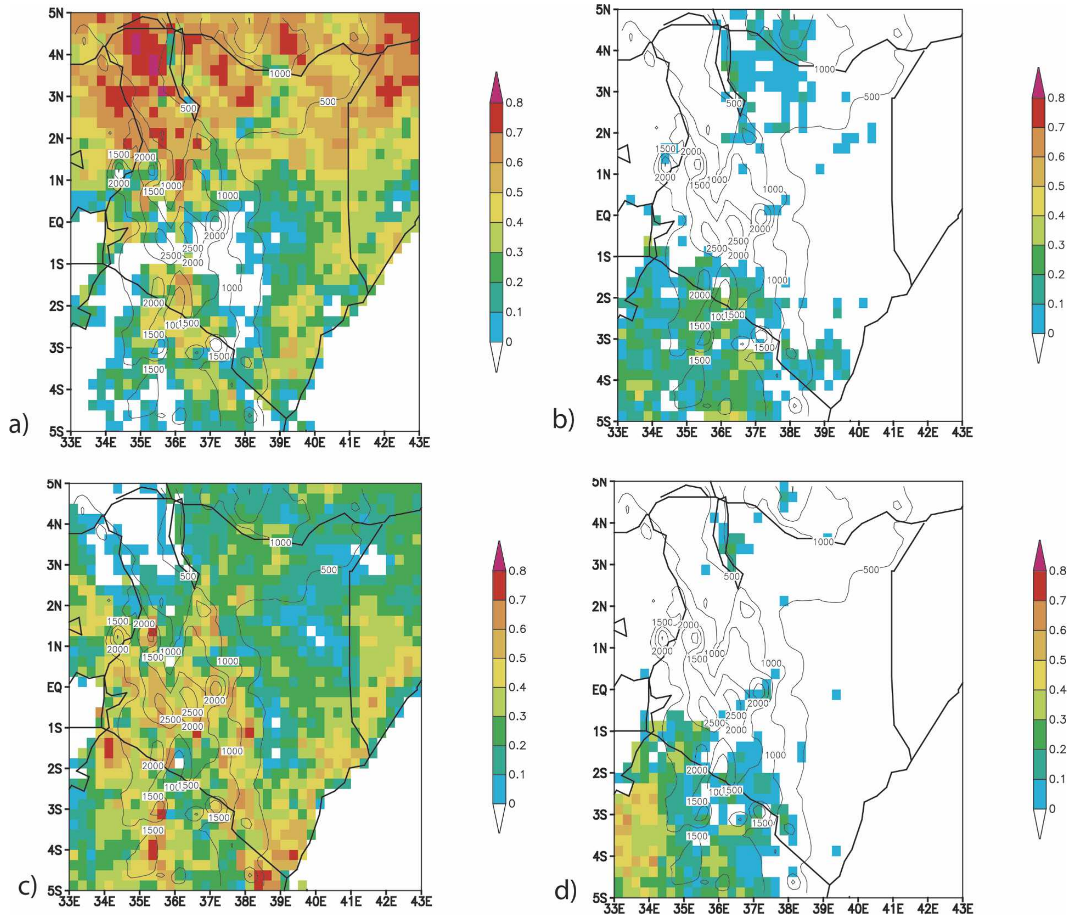

FIG. 4. Correlation between NDVI and the first four PC time series of OND GCM precipitation and circulation (GCM forced by observed SSTs): (a) PC1, (b) PC2, (c) PC3, and (d) PC4. Contours are elevation in meters.

tropical Atlantic Ocean, and pockets of positive values over the Indian Ocean and also negative values over the western Pacific Ocean (Fig. 5c). High negative correlations are also observed over the location of high pressure cells that modulate regional climate and weather, namely, the Azores (North Atlantic), Mascarene (south Indian Ocean), and St. Helena (South Atlantic Ocean). Reduction in pressures over these regions would restrict the ITCZ to the south and west of Kenya. This third pattern could also be related to the "dipole" between Madagascar and the southwest part of the Indian Basin. Figure 4 shows high skill between this mode and NDVI over the western and coastal areas of Kenya. Warming over the equatorial Atlantic Ocean has also been associated with frequent westerly waves across the equatorial African continent that penetrate as far as western Kenya, modulating weather and climate over these regions (Mutai and Ward 2000).

In summary, the second, third, and fourth modes of the GCM precipitation simulations are related to the impact of the adjacent oceans on the climate over Kenya, whereas the first mode describes the large-scale ENSO teleconnections.

b. Skill of NDVI predictions from GCM hindcasts

Figure 6 shows the simultaneous correlation between observed NDVI and cross-validated predictor fields for the OND season. The contribution of each month is significant in the overall seasonal (OND) correlation. November contributes about 40% of the seasonal rainfall with October and December contributing about 30% each. The substantial contribution of each month to the seasonal total, and the known similarity of monthly teleconnections with SST and the large-scale atmosphere, justify including the three months in the analysis. The CCA method draws on the first four EOF modes, which explain about 71% of the total variance. The correlation pattern shown in Fig. 6 indicates high skill (substantial areas with correlation values >0.6) mapped over the northern, eastern, and southern parts

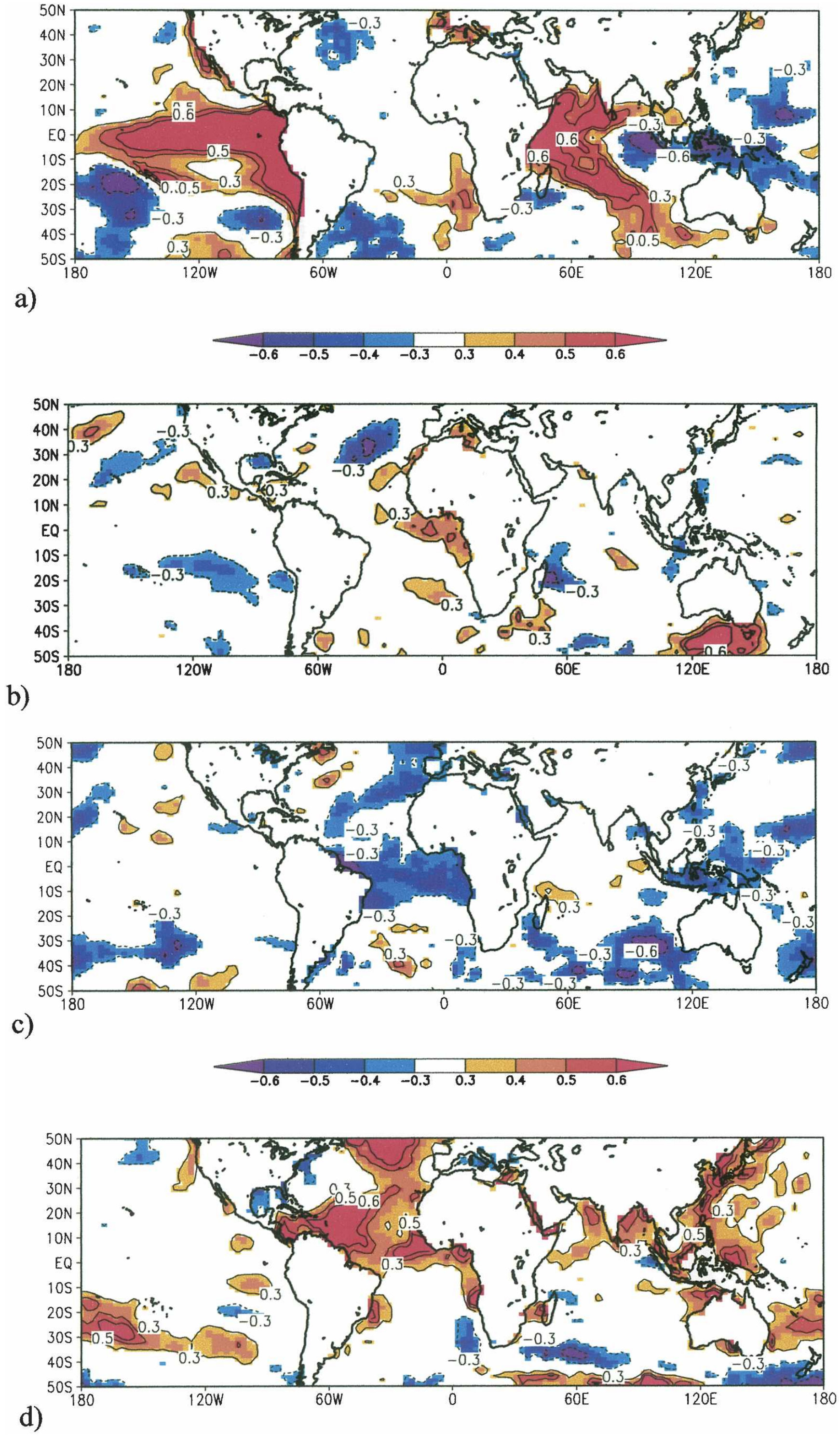

FIG. 5. Correlation between PC time series of OND GCM precipitation and circulation and SSTAs: (a) PC1, (b) PC2, (c) PC3, and (d) PC4.

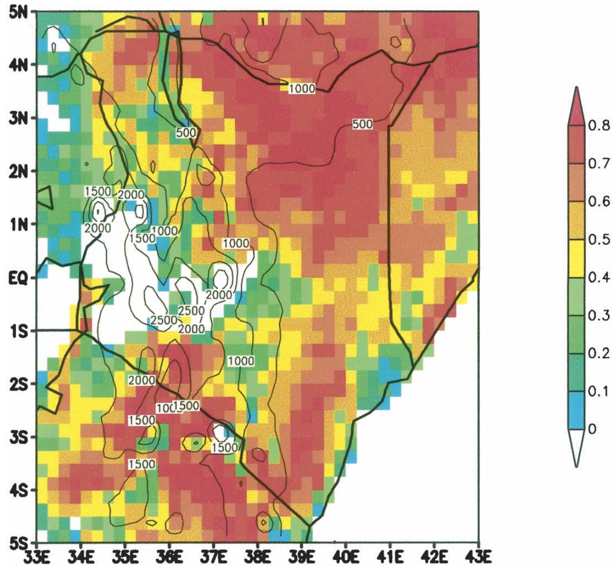

FIG. 6. Simultaneous correlation between observed OND NDVI (predictand) and OND cross-validated first canonical GCM precipitation and circulation (predictor) time series. Contours are elevation in meters.

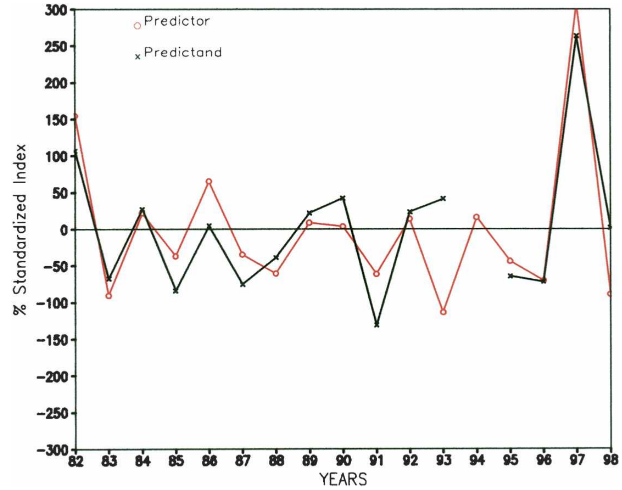

FIG. 7. Time series of observed OND NDVI (predictand) averaged over eastern Kenya and cross-validated first canonical (predictor) time series of OND GCM precipitation and circulation. Correlation coefficient $r = 0.83$.

of Kenya. Low skill is shown over the highland areas of the country. There is high predictive skill for NDVI over ASALS of Kenya, which are mainly occupied by the pastoralists and these are areas where livestock for export are reared. Timely prediction of NDVI in these areas would benefit the local community in the management of fodder, which is dependent on variability in green biomass. NDVI time series averaged over eastern parts of Kenya (0.5°–3°N, 38.5°–39.5°E) and cross-validated CCA1 time series for OND GCM precipitation are shown in Fig. 7 (several grid pixels for October and November 1994 were missing). The CCA1 time series is capable of replicating the NDVI with a high correlation skill of 0.82 (significant at the 99% level). The GCM is able to predict high NDVI values for the years 1982 and 1997 that were associated with the RVF outbreaks in the country. Predictions of the NDVI should not be expected to be perfect for a number of reasons. The climate prediction is probabilistic, and it will always be the case that in some years, the outcome is in the outlying wings of the distribution of those possible given the prevailing SST conditions. It is also the case that the GCM itself is an approximation, and may sometimes misrepresent the effects of SST on the regional climate. A further candidate for errors in the predictions is possible observational error on the NDVI values. Thus, the opposite sign in the predictor and predictand time series during 1993 could be attributed to any of the above sources of error. However, despite the possible sources of inaccuracy, most years are well represented, indicating a fairly robust cascade of predictive information from global SST fields into regional climate patterns and subsequently into the greenness of vegetation. The good skill is not solely confined to the periods when the strongest El Niño conditions prevailed (1982/1997), notably also including the years 1983, 1984, 1995, and 1996.

Figure 8 shows the spatial correlation between the cross-validated OND CCA1 predictor and the observed December predictand field. A visual check shows most

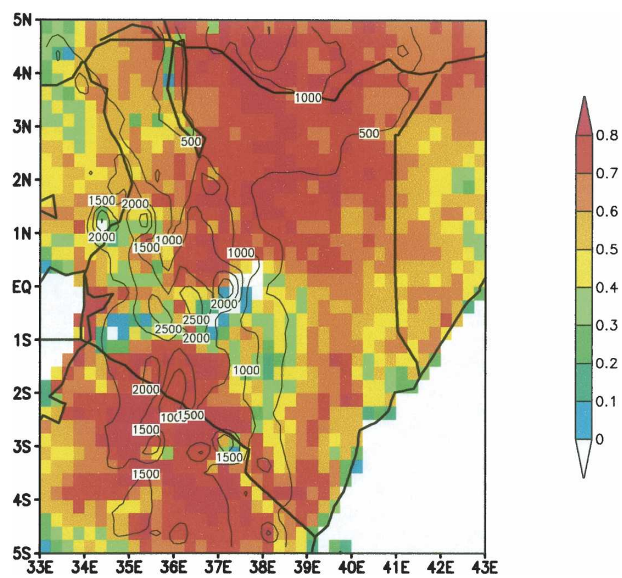

FIG. 8. Correlation between observed December NDVI (predictand) and cross-validated first canonical time series of OND GCM precipitation and circulation (predictor). Contours are elevation in meters.

of the grid points in the domain with a correlation coefficient greater that 0.4 (significant at the 90% level). Figure 8 shows increased correlation skill for December NDVI compared with that shown in Fig. 6, which implies that we can use seasonal GCM information to predict the vegetation conditions that will prevail at the end of the OND rainfall season. There is high correlation skill (>0.6, significant at the 99% level) over most parts of Kenya, decreasing over the highlands areas. The low skill over the highland areas could be linked to low NDVI variability (Fig. 1), which is consistent with generally abundant rainfall patterns over these areas. In contrast, time series of the December NDVI averaged over eastern Kenya and cross-validated predictor time series (Fig. 9) indicates particularly high prediction skill for these areas (correlation coefficient r = 0.83, significant at the 99% level).

c. Skill of NDVI predictions from GCM forecasts

The previous section establishes the basis for predicting NDVI, given observed SST conditions. Advances in capacity for operational prediction of NDVI are demonstrated in this section. The aim is to demonstrate the ability to predict NDVI in real time based on the GCM model output. Here we use GCM output obtained by forcing the model using persisted SST anomalies. The SST anomaly conditions for September are used to force the GCM though to the end of the season (i.e., OND). Skill levels of ECHAM GCM rainfall forecasts for different seasons and locations over the GHA region are shown by the IRI (see online at http://iri.columbia.edu). Skill in predicting December values

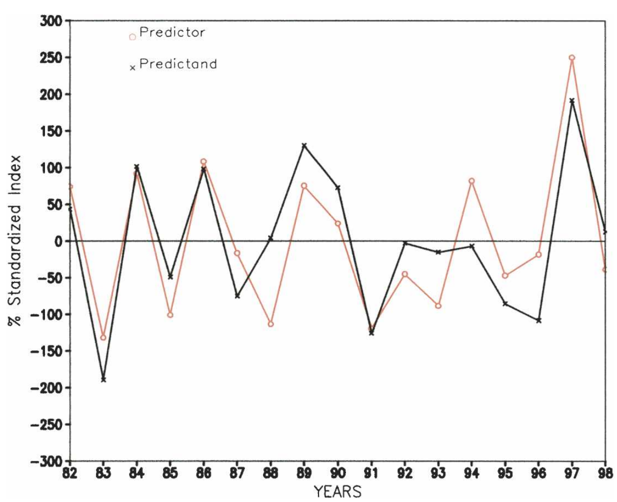

Fig. 9. Cross-validated time series of observed December NDVI (predictand) averaged over eastern Kenya and first canonical (predictor) time series of OND GCM precipitation and circulation. Correlation coefficient r = 0.83.

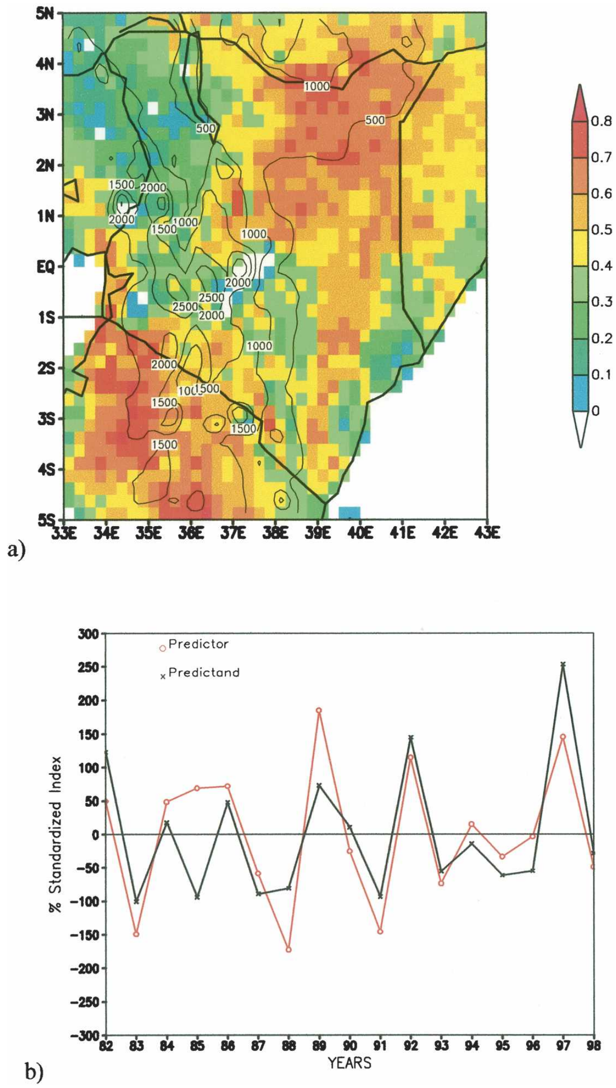

Fig. 10. Predicting December values of the NDVI in Kenya using output from the GCM. The GCM uses September SSTAs so real-time forecasts from this system would be available in early October. (a) Correlation between cross-validated prediction and observed NDVI. Contours show land elevation in meters. (b) Time series of the predicted and observed NDVI for an area average across eastern Kenya (Correlation = 0.76).

of the NDVI in East Africa using output from the ECHAM4.5 GCM is shown in Fig. 10a. Areas of skill >0.5 are widespread with some pockets >0.7. Figure 10b gives a graphical presentation of the accuracy of the forecasts: time series of the predicted and observed NDVI for an area average across eastern Kenya (correlation = 0.76, significant at the 99% level). Predictions are made using large-scale GCM fields of rainfall and low-level winds. The GCM experiments are based on persisted September SST information, so the forecast information would be available in early October. A

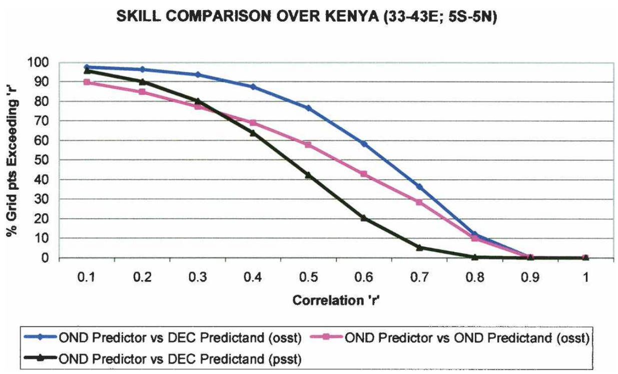

FIG. 11. Comparison of NDVI prediction skill over Kenya. Skill levels obtained by using cross-validated statistical models with predictors derived from GCM output. The output is from forecast experiments with persisted SST anomalies (psst) and simulation experiments with observed SST (osst). The figure shows the percentage number of grid points (within the domain enclosed by Kenya: 5°S–5°N, 33°–43°E) exceeding a given threshold value of correlation skill.

total of four PCs are used and each of predictors is evaluated through cross validation. The figure shows skillful correlations (>0.6) over the ASALS covering the northeastern, eastern, and southern Kenya. Thus, a skillful forecast is obtained 2–3 months in advance for the end of season (December) NDVI conditions.

Figure 11 provides a broader summary, showing skill levels obtained by using cross-validated persisted SST (psst) and observed SST (osst) GCM predictor and observed NDVI predictand at simultaneous and lagged times. The figure shows the percentage number of grid points (within the domain enclosed by Kenya, 5°S–5°N, 33°–43°E) exceeding given thresholds of correlation values. The best overall skill is shown between the OND predictor and the December predictand for the GCM forced by observed SST. Only a modest drop in the correlation skill values is shown when using persisted SST to force the GCM, which implies that we could predict NDVI with a lead time of 2–3 months over Kenya with considerable skill. Goddard and Mason (2002) have indicated that the use of forecast SST to force the GCM improves on the rainfall skill over that obtained by using persisted SST to force the GCM, so the persisted skill results in this paper can be viewed as a conservative estimate of that currently achievable in real time.

Initial analysis of different lag predictor series for NDVI during the OND season showed the best relationships at one month prior to the target season. Davenport and Nicholson (1993) showed similar results on their analysis of rainfall and NDVI over the same region.

5. Discussion

NDVI, a measure of vegetation greenness, is often highly climate dependent. Rainfall amount coupled with types of soils that have high capacity of retaining moisture create suitable conditions for high NDVI values. NDVI also increases in low-lying dambo areas when they are flooded for about 60 days (Davies et al. 1985). This fast growth of NDVI values in dambo areas due to the availability of water in these depressions provides suitable conditions for mosquito breeding. Animals are attracted to these areas for fodder and water sources making them highly exposed to mosquito bites that can result in RVF epizootics. Our study has shown that NDVI can be predicted skillfully over Kenya with a lead time of 2–3 months. Such forecasts of NDVI are thus a potential input to an RVF forecast model.

The physics and dynamics of the MOS predictors are related to the large-scale climatic forcing derived from the air–sea interactions. The first GCM precipitation PC time series is significantly correlated with SST over tropical areas of the western Indian Ocean and eastern Pacific Ocean, which indicate ENSO coupling through local and remote teleconnections. The second, third, and fourth GCM rainfall PC time series are related to the local ITCZ variability that is controlled by regional pressure gradients modulated by SST conditions in the neighboring oceans. GCM forecasts of precipitation and circulation fields show, in turn, good skill as predictors for NDVI over most parts of Kenya, with a notable exception in the highland areas. Variation in skill may also be related to variations in NDVI sensitivity to climate, or to spatial variations in climate predictability, or to sampling noise. Research is needed to better understand the skill variations, to give more confidence in the detailed spatial output of experimental forecast systems. Indeed, the spatial variations of skill in NDVI over some parts of Kenya require careful diagnostic analysis of the high-resolution climate and its interaction with the land surface. The use of regional models with high resolution capable of resolving the diverse topographic features over the GHA may provide better understanding of the physical mechanisms responsible for the low NDVI forecast skill over some parts of Kenya, and particularly high levels in other parts. Previous regional model studies have already provided better understanding of local climatic features over the region (Sun et al. 1999; Indeje et al. 2001).

In many geographic locations around the world, there is substantial loss of skill in seasonal rainfall when using persisted SSTs to force the GCM, as compared to using actual observed SSTs. For GHA, using September SSTAs still permits reasonable skill levels for OND rainfall. Furthermore, Goddard and Mason (2002) have shown that by using a system to forecast SSTs that includes the influence of the tropical Pacific on the Indian Ocean at this time of the year, could result in even better forecast skill for the GHA. Thus, GCM results obtained from persisted SST forcing could be considered a "lower limit."

The other climate factor that has been associated with RVF spread is related to the wind speed and direction. Mosquitoes may be transported by wind to remote areas away from the breeding origin. Movement of animals is also a factor that may influence the transfer of the virus. Convection and local air currents may be important vehicles for the transport of infected mosquitoes or other vectors of RVF during epizootics, and this could produce local or distant extension from the original foci of epizootics. It should be noted, for example, that the RVF outbreak in Egypt in 1977 was coincident with that in East Africa (Davies et al. 1985). The use of output from high-resolution regional climate models that simulate climate patterns in some of the years when RVF outbreaks have occurred could provide deeper understanding of the prevailing climatic conditions. Use of observational data could provide better understanding on the statistics of the intraseasonal weather. Statistics of intraseason weather (such as dry spells, rainy spells, and wind patterns) may provide the best predictors for NDVI or soil moisture and circulation patterns that could affect virus transmission within the region. We have demonstrated the capacity to produce predictions for NDVI over Kenya for the OND season. MAM is a period of lower predictability of rainfall, and predictability of NDVI for that period is yet to be established.

An increase in human population and the expansion of agriculture is destroying some of the natural ecosystems that support dambo mosquito-breeding sites. In recent years many dambos have been extensively used for agricultural purposes, where their higher water table allows areas of successful cultivation in a generally drier ecotype. This could cause changes in NDVI patterns and reduction in the dambo-breeding habitats of the mosquitoes and hence RVF epizootics in the region. The objective of the current work was to develop a methodology of projecting NDVI in space and time that could be used as a tool in applications sensitive to NDVI variability. The established relationships are strong enough to warrant applying the model in near-real-time situations. For optimal operational purposes, there is need to merge the system with ones using monitored climate information such as observed rainfall. We are therefore exploring the same scheme using observed pre-October rainfall amounts as additional predictors to the GCM output. To form a basis for real-time interventions by decision makers in the RVF problem, a prediction system must be able to quantify in a reliable way the areas of high risk for viral activity. Quantifying the linkage between a variable like NDVI and the risk of viral activity is therefore a further requirement for application of the methods in this paper to the RVF problem. RVF control efforts, which include the implementation of mosquito-control strategies and livestock immunization, could then be more confidently and effectively put in place. Work continues on ways of incorporating this model as a contribution to an operational early warning system for monitoring RVF over the GHA and the Middle East.

6. Conclusions

Statistical transformation procedures have been applied to the output from global climate model (GCM) seasonal predictions, in order to derive prediction information that more closely matches the needs of a societal problem. The methodology is based on the premise that climate variability, and especially precipitation, drive substantial variability in NDVI, which is a key indicator for the management of a range of environmentally related social problems. One such problem that has been identified is livestock Rift Valley Fever (RVF) and its effects on pastoral livelihoods both directly and through trade implications. Demonstration of our ability to predict seasonal precipitation variability with good skill, and then "downscale" it into high-resolution NDVI information, suggests a reliable scheme for predicting the risk of RVF outbreaks a few months in advance is feasible across much of this region, at least to the extent that RVF is linked to NDVI.

NDVI is highly dependent on soil moisture conditions, and precipitation fields from the GCM were expected to provide a good proxy. NDVI is also dependent on other factors that include soil type, elevation, etc., which may induce a varying responses of NDVI to climate forcing. The approach of applying a statistical transformation to the output of the GCM to fit it to NDVI variability provides an empirical methodology that factors in these complexities.

In the absence soil moisture data, NDVI has been used to monitor RVF episodes over the GHA. We have demonstrated from this study that NDVI can be skillfully predicted with a 2–3-month lead time using GCM forecasts of regional rainfall and circulation patterns. Our results show good skill for NDVI over most parts of Kenya, decreasing in the highland areas.

Results were initially derived using GCM simulations with observed SST. When the GCM is run mimicking a true operational forecast situation (i.e., using persisted September SSTAs) skill declines only marginally as shown in Fig. 11. For an NDVI index in northeast Kenya, correlation skill falls marginally from 0.82 to 0.76. These results are based on a sample of 17 yr. The climate predictability for the region has been demonstrated over a much larger set of years and is considered robust. It will be useful to extend the analysis with NDVI over recent years to continue to increase the robustness of this part of the forecast system.

Although RVF cases have been reported both over the Kenya highlands as well as the lowlands, our analysis has shown better NDVI predictive skill over the lowlands, suggesting at this stage, higher confidence is expected in predicting RVF outbreaks over the lowlands of Kenya and Tanzania. These results provide potentially useful information for factoring in the detection and monitoring of suitable conditions for RVF outbreaks, and developing strategies for mosquito control and disease prevention. An early warning system could be tailored to provide guidance on the remedies to be taken including vaccination of the animals, stabilizing the movement of the animals, and mosquito control measures such as treating of the mosquito-breeding dambos.

In addition to the possible contribution managing RVF outbreaks, the demonstration of a capability to predict NDVI using seasonal climate forecasts brings a range of additional opportunities. The Famine Early Warning System (FEWS), a project of the U.S. Agency for International Development (USAID) and the National Aeronautics and Space Administration (NASA) routinely use Normalized Difference Vegetation Index (NDVI) images to monitor environmental conditions worldwide. Thus, other opportunities can be subsequently pursued in areas where NDVI provides useful information, such as in the prediction of livestock fodder.

Acknowledgments. We greatly appreciate comments by Drs. L. Goddard, A. Giannini, and other IRI staff. The authors acknowledge support from NOAA through its Cooperative Agreement NA050AR431004, and the United States Agency for International Development (USAID)/Office of Disaster Management (OFDA). The views expressed herein are those of the author and do not necessarily reflect the views of NOAA/USAID/OFDA or any of its subagencies.

REFERENCES

Anyamba, A., and J. R. Eastman, 1996: Interannual variability of NDVI over Africa and its relation to El Niño/Southern Oscillation. *Int. J. Remote Sens.,* **17,** 2533–2548.

——, C. J. Tucker, and R. Mahoney, 2002: El Niño and La Niña: Vegetation response patterns over East and Southern Africa 1997–2000. *J. Climate,* **15,** 3096–3103.

——, K. J. Linthicum, R. Mahoney, and C. J. Tucker, 2003: Mapping potential risk of rift valley fever outbreaks in African savannas using vegetation index time series data. *Photogramm. Eng. Remote Sens*., **68,** 137–145.

Camberlin, P., S. Janicot, and I. Poccard, 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.

Davenport, M. L., and S. E. Nicholson, 1993: On the relation between rainfall and the Normalized Difference Vegetation Index for diverse vegetation types in East Africa. *Int. J. Remote Sens.,* **14,** 2369–2389.

Davies, F. G., K. J. Linthicum, and A. D. James, 1985: Rainfall and epizootic Rift Valley fever. *Bull. World Health Org.,* **63,** 941–943.

Farmer, G., 1988: Seasonal forecasting of the Kenya Coast shortrains. *Int. J. Climatol.,* **8,** 489–497.

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

Goddard, L., and S. J. Mason, 2002: Sensitivity of seasonal climate forecasts to persisted SST anomalies. *Climate Dyn.,* **19,** 619–632.

——, 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.

Graham, N. E., P. Barnett, R. Wilde, M. Ponater, and S. Schubert, 1994: On the roles of tropical and midlatitude SSTs in forcing interannual to interdecadal variability in the winter Northern Hemisphere circulation. *J. Climate,* **7,** 1416–1441.

Hansen, J. W., and M. Indeje, 2004: Linking dynamic seasonal climate forecasts with crop simulation for maize yield prediction in semi-arid Kenya. *Agric. For. Meteor.,* **125,** 143–157.

Hielkema, J. U., S. D. Prince, and W. L. Astle, 1986: Rainfall and vegetation monitoring in the Savanna Zone of the Democratic Republic of Sudan using the NOAA Advanced Very High Resolution Radiometer. *Int. J. Remote Sens.,* **7,** 1499–1514.

Holben, B. N., 1986: Characteristics of maximum-value composite images for temporal AVHRR data. *Int. J. Remote Sens.,* **7,** 1417–1434.

Indeje, M., F. H. M. Semazzi, and L. J. Ogallo, 2000: ENSO signals in East African rainfall and their prediction potentials. *Int. J. Climatol.,* **20,** 19–46.

——, ——, L. Xie, and L. J. Ogallo, 2001: Mechanistic model simulations of the East African climate using NCAR regional climate model: Influence of large-scale orography on the Turkana low-level jet. *J. Climate,* **14,** 2710–2724.

Kidson, J. W., and C. S. Thompson, 1998: A comparison of statistical and model-based downscaling techniques for estimating local climate variations. *J. Climate,* **11,** 735–753.

Kutzbach, J. E., 1967: Empirical eigenvectors of sea level pressure, surface temperature and precipitation complexes over North America. *J. Appl. Meteor.,* **6,** 791–802.

Landman, W. A., and L. Goddard, 2002: Statistical recalibration of GCM forecasts over southern Africa using model output statistics. *J. Climate,* **15,** 2038–2055.

Latif, M., and Coauthors, 1998: A review of the predictability and prediction of ENSO. *J. Geophys. Res.,* **103,** 14 375–14 393.

Linthicum, K. J., C. L. Bailey, F. G. Davies, and C. J. Tucker, 1987: Detection of Rift Valley fever viral activity in Kenya by satellite remote sensing imagery. *Science,* **235,** 1656–1659.

——, and Coauthors, 1990: Application of polar-orbiting meteorological satellite data to detect flooding of Rift Valley fever virus vector mosquito habitats in Kenya. *Med. Vet. Entomol.,* **4,** 433–438.

——, A. Anyamba, C. J. Tucker, P. W. Kelley, M. F. Myers, and C. J. Peters, 1999: Climate and satellite indicators to forecast Rift Valley fever epidemics in Kenya. *Science,* **285,** 397–400.

Meegan, J. M., and C. L. Bailey, 1989: Rift Valley fever. *The Arbo Viruses: Epidemiology and Ecology,* T. P. Monath, Ed., Vol. IV, CRC Press, 51–76.

Mutai, C. C., and M. N. Ward, 2000: East African rainfall and the tropical circulation/convection on intraseasonal to interannual timescales. *J. Climate,* **13,** 3915–3939.

Nicholson, S. E., 1996: A review of climate dynamics and climate variability in Eastern Africa. *The Limnology, Climatology and Paleoclimatology of the East African Lakes*, T. C. Johnson and E. O. Odada, Eds., Gordon and Breach, 25–56.

Ogallo, L. J., J. E. Janowiak, and M. S. Halpert, 1988: Teleconnection between seasonal rainfall over East Africa and Global sea surface temperature anomalies. *J. Meteor. Soc. Japan,* **66,** 807–822.

Roeckner, E., J. M. Oberhuber, A. Bacher, M. Christoph, and I. Kirchner, 1996: ENSO variability and atmospheric response in a global coupled atmosphere-ocean GCM. *Climate Dyn.,* **12,** 737–754.

Ropelewski, C. F., and M. Halpert, 1987: Global and regional scale precipitation patterns associated with the El Niño/ Southern Oscillation. *Mon. Wea. Rev.,* **115,** 1606–1626.

Schreck, C. J., III, and F. H. M. Semazzi, 2004: Variability of the recent climate of eastern Africa. *Int. J. Climatol.,* **24,** 681–701.

Stone, M., 1974: Cross-validatory choice and assessment of statistical predictions (with discussion). *J. Roy. Stat. Soc.,* **B36,** 111–147.

Sun, L., F. H. M. Semazzi, F. Giorgi, and L. J. Ogallo, 1999: Application of the NCAR Regional Climate Model to Eastern Africa. Part I: Simulations of autumn rains of 1988. *J. Geophys. Res.,* **104** (D6), 6529–6548.

Swanepoel, R., 1981: Observations on Rift Valley Fever in Zimbabwe. *Contrib. Epidemiol. Biostat.,* **3,** 83–91.

Thomson, M., M. Indeje, S. Connor, M. Dilley, and N. Ward, 2003: Malaria early warning in Kenya and seasonal climate forecasts. *The LANCET*, **362** (9383), 580. [Available online at http://www.thelancet.com.]

Tucker, C. J., 1979: Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.,* **8,** 127–150.

——, C. O. Justice, and S. D. Prince, 1986: Monitoring the grasslands of the Sahel 1984-1985. *Int. J. Remote Sens.,* **7,** 1571–1581.

——, H. E. Dregne, and W. W. Newcomb, 1991: Expansion and contraction of the Sahara Desert from 1980 to 1990. *Science,* **253,** 299–301.

Trenberth, K. E., 1998: Forecasts of the development of the 1997-98 El Niño event. *CLIVAR Exchanges,* Vol. 3, No. 2/3, 4–14. [Available online at http://www.clivar.ucar.edu.]

Verdin, J., C. Funk, R. Klaver, and D. Roberts, 1999: Exploring the correlation between Southern Africa NDVI and Pacific sea surface temperatures: Results for the 1998 maize growing season. *Int. J. Remote Sens.,* **20,** 2117–2124.

Wilby, R., L. Dawsonb, and E. M. Barrow, 2002: SDSM—A decision support tool for the assessment of regional climate change impacts. *Environmental Modelling & Software,* D. S. Wilks, Ed., Vol. 17, *Statistical Methods in the Atmospheric Sciences*, Academic Press, 145–157.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.