Wavelet Analysis of Seasonal Rainfall Variability of the Upper Blue Nile Basin, Its Teleconnection to Global Sea Surface Temperature, and Its Forecasting by an Artificial Neural Network

MOHAMED HELMY ELSANABARY
Civil Engineering Department, Port Said University, Port Said, Egypt, and University of Alberta, Edmonton, Alberta, Canada

THIAN YEW GAN
Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta, Canada

(Manuscript received 1 March 2013, in final form 30 June 2013)

ABSTRACT

Rainfall is the primary driver of basin hydrologic processes. This article examines a recently developed rainfall predictive tool that combines wavelet principal component analysis (WPCA), an artificial neural networks-genetic algorithm (ANN-GA), and statistical disaggregation into an integrated framework useful for the management of water resources around the upper Blue Nile River basin (UBNB) in Ethiopia. From the correlation field between scale-average wavelet powers (SAWPs) of the February–May (FMAM) global sea surface temperature (SST) and the first wavelet principal component (WPC1) of June–September (JJAS) seasonal rainfall over the UBNB, sectors of the Indian, Atlantic, and Pacific Oceans where SSTs show a strong teleconnection with JJAS rainfall in the UBNB ($r \approx 0.4$) were identified. An ANN-GA model was developed to forecast the UBNB seasonal rainfall using the selected SST sectors. Results show that ANN-GA forecasted seasonal rainfall amounts that agree well with the observed data for the UBNB [root-mean-square errors (RMSEs) between 0.72 and 0.82, correlation between 0.68 and 0.77, and Hansen–Kuipers (HK) scores between 0.5 and 0.77], but the results in the foothills region of the Great Rift Valley (GRV) were poor, which is expected since the variability of WPC1 mainly comes from the highlands of Ethiopia. The Valencia and Schaake model was used to disaggregate the forecasted seasonal rainfall to weekly rainfall, which was found to reasonably capture the characteristics of the observed weekly rainfall over the UBNB. The ability to forecast the UBNB rainfall at a season-long lead time will be useful for an optimal allocation of water usage among various competing users in the river basin.

1. Introduction

The primary water supply of countries downstream of the Nile River, Egypt and Sudan, comes from the Nile itself, where one of the most important tributaries is located at the upper Blue Nile basin (UBNB), in Ethiopia, which contributes about 60% of the Nile’s total streamflow reaching the Aswan High Dam in Egypt (Seleshi and Demaree 1995; Yates and Strzepek 1998; Sutcliffe and Parks 1999; Conway 2000). Rainfall plays a major role in basin hydrology, for too much (little) rainfall will result in undesirable floods (droughts). Further, to effectively manage the water resources of the UBNB, it will be useful to develop a predictive tool that can forecast the rainfall of the UBNB at weekly to seasonal time scales. Roughly 70% of the annual rainfall of the UBNB occurs during the kiremt season, which is from June to September (hereafter the 4-month seasons will be abbreviated by the first letter of each month, e.g., JJAS) (Conway 2000), when about 85%–95% of the Ethiopian crops are grown annually (Degefu 1987). Therefore, the UBNB provides the largest and most economically significant water resource for the country. While Ethiopia has been planning irrigation and hydropower projects using the Blue Nile River, other neighboring countries also need to increase their allocations of Nile’s water resources to sustain their populations and economic development. Therefore, to

DOI: 10.1175/MWR-D-13-00085.1

© 2014 American Meteorological Society
avoid potential conflicts between competing countries, it will be necessary to establish equitable transboundary water-sharing agreements and honorable implementation of such agreements.

Since the UBNB contributes a significant amount of water to the Nile River, it is important to forecast its seasonal precipitation reliably. However, this can be challenging given that the UBNB had experienced hydrologic extremes, both floods and droughts, in the past few decades. Droughts and famines are endemic because agriculture, which accounts for approximately 50% of Ethiopia’s gross domestic product and employs 80% of the country’s population, is heavily dependent on rainfall (Abegaz et al. 2007). However, climate is a multifractal process of high fractal dimensions (e.g., Gan et al. 2002), and most forecasts simulated by climate models are not reliable beyond a week because climate is highly sensitive to the perturbation of its initial conditions, and numerical round-off errors grow exponentially as numerical weather prediction (NWP) models continue simulating the climate processes. Further, there is a general lack of good quality, long record climate data in the UBNB and its surrounding regions, which could be useful for fine-tuning NWP simulations. Therefore, it is a challenge to reliably forecast weekly precipitation at one-season lead time even though such information will be useful for water resources management and planning purposes in the UBNB.

Segele et al. (2009) showed that Ethiopia’s rainfall is nonstationary, particularly over the UBNB. Because of the nonstationary characteristics inherent in most rainfall data, nonstationary techniques are generally needed to accurately analyze the variability of rainfall data, which because of teleconnections, has been found to be related to the variability of sea surface temperatures (SSTs) (Diro et al. 2011; Elsanabary et al. 2014). Degefu (1987) and Wolde-Georgis et al. (2001) found that the Indian and Atlantic Oceans provide primary sources of moisture for most rainfall occurring over Ethiopia. Camberlin (1996) showed that monsoonal activity over India is a major trigger for the July–September rainfall variability in the East African highlands. Jury (2010) found that rainfall in northern Ethiopia that impacted the Nile River flow had been linked to the Atlantic zonal overturning circulation pattern and that rainfall modes exhibited interdecadal (10–12 yr) cycles throughout much of the twentieth century.

Therefore, for Ethiopia, and in particular the UBNB, it will be useful to teleconnect the variability of its precipitation to oceanic SSTs. Knowledge of this teleconnection between selected oceanic SSTs and the rainfall of the UBNB would enhance Ethiopia’s rainfall forecasting ability, which would be beneficial to its agriculture practices and for effective management of its water resources, especially if hydrologic extremes such as droughts are expected. Furthermore, knowledge of the variability of rainfall across the Blue Nile basin has the potential to assist downstream countries such as Egypt and Sudan that rely heavily on the water of the Nile River. Currently, the National Meteorological Agency (NMA) of Ethiopia issues seasonal summer rainfall forecasting for the country (Gissila et al. 2004). In Sudan, the Early Warning and Humanitarian Information Center (EWHIC) issues flood forecasts at key gauge stations along the Blue Nile at up to 4-day lead time (Thiemig et al. 2011).

With the motivation of developing a robust framework for forecasting JJAS rainfall at weekly time steps, the objectives of this study are outlined in section 2. The characteristics and climatology of the study site (UBNB) are presented in section 3. Climate data are examined in section 4. The research procedure and methodology are reviewed in section 5. Discussion of results is presented in section 6. Our summary and conclusions are detailed in section 7, with recommendations for future work offered in section 8.

2. Research objectives

With the above statement of the problem, the primary objective of this study is to develop a framework for forecasting the weekly seasonal rainfall of the UBNB driven by selected sectors of oceanic SSTs at one-season lead time. This framework is developed on the basis of wavelet-based principal component analysis (WPCA), artificial neural networks calibrated by a genetic algorithm (ANN-GA), and a statistical disaggregation algorithm according to the following procedures: 1) to analyze the nonstationary variability of the February–May (FMAM) SST of the global oceans, 2) to identify sectors of SSTs that are strongly teleconnected to the seasonal rainfall (JJAS) of the upper Blue Nile basin, 3) to drive ANN-GA to forecast the JJAS rainfall over the UBNB from the above-selected SST sectors of the oceans at one-season lead time, and 4) to disaggregate the seasonal rainfall forecasted by ANN-GA to weekly rainfall.

3. Study site

a. Upper Blue Nile basin

The UBNB (Fig. 1), 176 000 km² in area and occupying 17% of Ethiopia while forming the western part of the Ethiopian highlands, on average receives 1600 mm of precipitation annually (Sutcliffe and Parks 1999), and has a mean annual discharge volume of 48.5 km³ (Conway...
The Blue Nile (Abay) begins at Tana Lake (2150 km²) at an elevation of about 1800 m MSL, then leaves the lake at the Tississat Falls, which drop over 50 m vertically (Shahin 1985). The UBNB generally receives more precipitation (1000–2400 mm yr⁻¹) than the eastern, northern, and southern parts of Ethiopia, and due to the orographic effect, the precipitation generally increases with altitude (Kloos and Legesse 2010). This orographic effect is considered to be a significant challenge for all rainfall-forecasting models. Therefore, rainfall modeling in this region will be useful for the optimal allocation of water usage in the basin.

b. Climatic regime of the UBNB, Ethiopia

Over Ethiopia, the spatial distribution of seasonal rainfall is related to the migration of the intertropical convergence zone from south to north, which causes the rainfall to be more concentrated west of the Great Rift Valley (GRV) (Sutcliffe and Parks 1999, Degefu 1987; Gamachu 1977). In the western highlands where the UBNB is located, rainfall occurs over much of the year, but the wet season occurs predominantly from June to September. In the southern, central, and northern parts of Ethiopia, the dry season (called belg) occurs in February–May (FMAM) while the wet season (called kiremt) occurs in JJAS (Kloos and Legesse 2010) (Fig. 2). Essentially, from February to May, the southeasterly winds that carry moisture from the Indian Ocean give rise to the first rainy season over most of Ethiopia, while from June to September the southwesterly and southeasterly winds are responsible for the second rainy season over Ethiopia. Seasonal rainfall and wind variations are different between 3-month (January–March (JFM), April–June (AMJ), July–September (JAS), and October–December (OND)) or 4-month (FMAM, JJAS, and October–January (ONDJ)) sequences. Figure 3 shows the wind speed variation from season to season and for different time periods (i.e., considering 3- or 4-month seasons). Figures 3a–d show the wind patterns for 3-month seasons (1970–90), while Figs. 3f–h show the wind patterns for 4-month seasons during 1970–90, and Figs. 3i–k show the wind patterns for 4-month seasons from 1990 to 2002. Monthly wind data in Fig. 3 were taken from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis dataset (www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.surface.html).

4. Climate data

a. Rainfall data

Historical observed monthly precipitation data (1900–98) were taken from Hulme’s (1992) dataset at the Climatic Research Unit, University of East Anglia (UEA), United Kingdom. The data, gridded at a resolution of
2.5° × 3.75°, were extracted from the region 18°–2°N to 33°–48°E over Ethiopia and the surrounding areas (Fig. 1). Data were extracted for 23 grids. These data are part of a historical monthly precipitation dataset for global land areas. These gridded data were constructed from station data using the Thiessen polygon method. No topographic weighting has been applied to the interpolation scheme. Because the method interpolates anomalies and not precipitation values, it is reasonable to exclude the effects of elevation. The data quality control process for these gridded data is described in Hulme (1992, 1994), Hulme and New (1997), and Hulme et al. (1998). Given that the focus of this study is the kiremt season, the seasonal data for JJAS were computed. The 6-h (1961–2002) gridded data from the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) at 2.5° × 2.5° resolution were collected from the ECMWF website (http://data-portal.ecmwf.int/data/d/era40_daily/) and aggregated to weekly data for JJAS for comparison with the forecasted weekly seasonal rainfall.

b. Sea surface temperature

Global mean monthly SST data from 1870 to 2008, at 1° × 1° grid resolution, were extracted for about 65,000 grids from version 1.1 of the Hadley Centre Sea Ice and Sea Surface Temperature dataset (HadISST). The monthly SST anomaly grid data were aggregated to seasonal FMAM SSTs to be used as the SST predictors for forecasting the seasonal precipitation of the UBNB at one-season lead time.

5. Research procedure and methodology

a. Wavelet analysis

A wavelet function is a “little” wave with limited duration in comparison to waves such as sinusoidal waves. A wavelet function can be expressed in several ways, but it is mainly divided into two main categories: continuous (e.g., Morlet, Mexican hat) and orthogonal (e.g., Haar, Daubechies). A Morlet wavelet consists of a sinusoidal wave modulated by a Gaussian distribution, which is ideal for detecting climate signals. It has the ability to detect
FIG. 3. 850-hPa winds over Africa showing the prevailing atmospheric circulation and moisture transport in terms of wind speed (m s$^{-1}$) and direction throughout different seasons as indicated for a particular time period (ERA-40 data): (a)–(d) OND, JFM, AMJ, and JAS for the period 1970–90; (f)–(h) ONDJ, FMAM, and JJAS for the period 1970–90; and (i)–(k) ONDJ, FMAM, and JJAS for the period 1990–2002.
both time-dependent amplitude and phase changes for
different frequencies in a time series (Lau and Weng 1995). In other words, a Morlet wavelet can effectively
delineate characteristics of precipitation, temperature,
and SST data (Mwale et al. 2009). Therefore, a Morlet
wavelet, $c_0$, based on a nondimensional time pa-
rameter $h$, is chosen for this study (Torrence and Compo
1998):

$$c_0(h) = \frac{\sqrt{2}}{\pi} \frac{\sin(\pi h)}{h}. $$ (1)

In this paper, the continuous wavelet transform (CWT)
is used, which is a convolution of the input data sequence
with a set of functions generated by the mother Morlet
wavelet. By stretching a wavelet in time over the $x$ axis
(along a localized time index $n$), its scale $s$ varies over the
$y$ axis continuously. By so doing, we could generate a set
of transforming wavelets (Addison 2002; Mallat 1999):

$$\psi_{s,n}(t) = \frac{1}{\sqrt{s}} \psi \left( \frac{t-n}{s} \right). $$ (2)

The CWT or the wavelet spectrum $W_n(s)$ of a discrete
(i.e., rainfall or SST) time series $x(t)$ of time step $\delta t$ using
a scaled and translated version of $c_0(\eta)$ is as follows:

$$W(t) = \sum_{n=0}^{N-1} x(t) \psi^{*} \left[ \frac{(t-n)\delta t}{s} \right], $$ (3)

where (*) indicates the complex conjugate of $\psi$, $t = 0, \ldots, N - 1$, and $N$ is the sample size. Also, the subscript of $\psi_0$ in Eq. (1) has been dropped to indicate that the
wavelet has been normalized so that it has unit energy.

**Figure 4** shows an example of the wavelet spectrum ($x$–$y$
image) generated from the convolution represented by
Eq. (3). For convenience, the scales are expressed as
fractional power of two (Torrence and Compo 1998).

The magnitude of the $W_n(s)$ coefficients represents how
well the wavelet matches with the precipitation time se-
ries. To investigate fluctuations in power over a range of
scales (band), the scale-averaged wavelet power (SAWP),
for either the rainfall or SST time series, which represents
the mean variance of the wavelet coefficients over a range

---

**Fig. 4.** Examples of continuous wavelet and global spectra of Ethiopian rainfall during JJAS at (a) 17.5°N, 37.5°E and (b) 12.5°N, 45°E. Each example presents the Morlet wavelet power spectrum, where the dashed line is the CI, beyond which the energy is contaminated by the effect of zero padding, and the thick black contours represent the 95% confidence level of local power relative to a red-noise background; the global wavelet power spectrum (solid line) with the 95% confidence level (dashed line) are shown to the right.
of scales from $s_1$ to $s_2$ (e.g., from 2 to 8 years) was also computed for each grid point \cite{T98}:
\begin{equation}
W_k^2 = \frac{\delta_1 \delta_i}{C_{\delta i}} \sum_{j=1}^{l_k} \frac{|W(s_j)|^2}{s_j},
\end{equation}
where $C_{\delta i}$ is 0.776 for the Morlet wavelet, $\delta_i$ is a factor for scale averaging, and $\delta_1$ is the sampling period \cite{T98}. The correlation between a time series and a climate anomaly within certain scale bands and time periods can be identified from the wavelet coherence defined as below \cite{T99}:
\begin{equation}
R_k^2(s) = \frac{|s^{-1}W^X(s,n)W^Y(s,n)|^2}{|s^{-1}|W^X(s,n)|^2 |s^{-1}|W^Y(s,n)|^2},
\end{equation}
where $W^X(s,n)$ and $W^Y(s,n)$ are the wavelet transforms of the $X$ and $Y$ time series, respectively, where $t$ is the time index and $s$ is the scale. The cross-wavelet spectrum of $X$ and $Y$ is $W^{XY}(s,n)$, \cite{T98} is a smoothing in both time and scale, and $R_k^2(s)$ is the wavelet squared coherency where $0 < R_k^2(s) \leq 1$. Since the wavelet transform conserves variance, Eq. (5) is an accurate representation of the normalized covariance between two time series (see Figs. 12 and 13).

b. Principal component analysis

Principal component analysis (PCA), a multivariate technique that orthogonally transforms a set of possibly correlated variables into a set of linearly uncorrelated variables called principal components (Abdi and Williams 2010), is widely used in atmospheric sciences \cite{W06}. In other words, PCA transforms multivariate data into independent PCs such that several leading PCs generally explain a large proportion of the total variance.

The rainfall or SST data matrix $X$ of $N$ variants or climate stations and each with $n$ observations ($n \times K$) were converted to anomalies ($X'$), by subtracting the mean and standardized by the standard deviation. PCA transforms $X$ into another matrix of principal components $U$ that accounts for all the variability of $x_i$, so that a few PCs (designated as $U_m$) account for the majority of the joint variation of $X$. To examine the teleconnection between the rainfall of the UBNB and SST, PCA was applied to the SAWP (2–8 yr) of rainfall and SSTs \cite{Mwale2010, Elsanabary2014}, so that a few $U_m$ account for the majority of the SAWP variation. The SAWP signals ($U_m$) are computed as follows:
\begin{equation}
U_m = e_m^{T}X' = \sum_{k=1}^{K} e_{km}X'_k, \quad m = 1, \ldots, M \quad \text{(where } M \ll K),
\end{equation}
where $e_{km}$ are the eigenvectors, $X'_k$ are the $k$th SAWP anomalies of $X$, and $M$ represents a small subset of the $K$ possible signals. The signals $U_m$ are usually the major spatial and temporal patterns that account for the majority of the variations in the SAWP, and can be used to delineate rainfall or SST variations spatially into independent zones.

c. ANN-GA for forecasting seasonal precipitation

An artificial neural network is comprised of a large number of simple, highly interconnected processing elements known as neurons, typically arranged in layers, to relate inputs and outputs of a system \cite{Hsu1995, Chau2005, Chen2005, Mutill2006, Wu2009, Taormina2012}. ANNs are data driven and nonparametric, and they do not necessarily require the enforcement of constraints or a priori solution structures \cite{DeVos2005}. The proposed model receives signals from the input nodes and transforms these signals through the network until they reach the output nodes \cite{Mwale2010, Kuo2010a, Mwale2004}. Through a trial and error approach, a five-node hidden layer was adopted, where the number of input nodes depends on the number of input variables or predictors and SST sectors, and the output layer has only one node, the predictand, which is the seasonal rainfall. The input data were normalized to zero mean and scaled to unit variance.

The ANN model’s parameters, which are essentially weights and biases in each of the input, hidden, and output layers, were calibrated by an ANN-GA. The GA iteratively searched for optimal model parameters on the basis of maximizing an objective function that represents the correlation between simulated and observed seasonal rainfall amounts. The GA consists of three processes: selection, crossover, and mutation. First, all of the neural networks are ranked according to their respective performance and evaluated in terms of objective function values computed in descending order. Only the top 85% of the population (set at 200 for this study) is retained for further selection. Second, the crossover process is performed by selecting pairs of neural networks from the former population and then swapping their weights and biases using a one-point crossover scheme \cite{Chau2005, Mutill2006}. In that scheme, a randomly chosen location is selected in the hidden layer and then the weights and biases are changed on either side of that location between pairs of the neural networks. The crossover process is replicated for the whole population between all other pairs of neural networks. Third, mutation is performed as a trial to restore “good” weights and eliminate biases that happened during the selection process.
A small percentage of the population (1%) is randomly chosen for mutation and a proportion of their weight is randomly muted. The three processes are repeated for several generations (Mwale et al. 2004; Kuo 2010a,b). At each generation, the best neural network is kept until a better solution is found in successive generations. To ensure the calibrated ANN model is valid or dependable, the weights and biases of the best surviving network found by the GA at the calibration stage were further tested by driving the model with input data independent of the calibration experience. If the performance of the ANN model at the validation stage is comparable to that of the calibration stage, there is a basis for accepting the calibrated ANN model as being physically sound and useful for rainfall forecasting purposes.

d. Disaggregation of seasonal precipitation

The Blue Nile River has a length of 1400 km and a slope between its source (1800 m) and the gauging station (500 m) (Shahin 1985) of about 10.77%. For a natural channel not well defined with a slope of 8%–11%, the approximate average flow velocity is between 1.22 and 2.13 m s\(^{-1}\) (Chow et al. 1988, 164–165). Therefore, the “time of concentration,” which indicates the time needed for the whole UBNB to contribute runoff at the basin outlet near El Diem station, Sudanese–Ethiopian border, is about 7–8 days or about a week. Therefore, it is reasonable to disaggregate the seasonal rainfall (JJAS) of the UBNB into weekly rainfall, which can then be used to model the hydrology of the UBNB at weekly time steps. For simplicity’s sake, each month was divided into 4 weeks. For example, for June and September, the first 2 weeks were averaged over 8 days, while the last 2 weeks were averaged over 7 days. For July and August, the first 3 weeks of each month were averaged over 8 days and the last week of each month was averaged over 7 days.

The disaggregation of seasonal rainfall in the UBNB may be carried out using the temporal disaggregation model of Valencia and Schaake (1973), the Lane model of Lane and Frevert (1990), the canonical random cascade model of Gupta and Waymire (1993), or the analog approach of Lorenz (1969), assuming the correlation structure between the JJAS and weekly rainfall data remains more or less stationary over the study period. If the analog approach is used, long datasets are required but the UBNB only has a limited amount of hydroclimatic data. Among the three disaggregation models, Kuo et al. (2010a) found that the Valencia and Schaake (1973) approach has been more effective in disaggregating seasonal rainfall to 3-day time-scale rainfall data. Therefore, the disaggregation method of Valencia and Schaake (1973) has been used in this study. Statistical properties between seasonal and weekly precipitation are used to disaggregate the seasonal precipitation to four weekly rainfall totals for each of the four months (i.e., for a total of 16 weekly precipitation totals). The disaggregation model of Valencia and Schaake (1973) is of the following form:

\[
Z_t = AX_t + B\varepsilon_t, \tag{7}
\]

where \(Z_t\) is the \((m \times 1)\) vector of the weekly rainfall of the \(t\)th year; \(X_t\) is the forecasted seasonal rainfall of the \(t\)th year; \(m\) is the number of weeks for the rainy season (JJAS), which in this case is 16 weeks; and \(\varepsilon_t\) is the \((m \times 1)\) vector of the standard normal deviates (i.e., zero mean and unity variance). In addition, \(A\) is a \((m \times 1)\) vector of coefficients that sum to one and can be considered as weekly contributions to the seasonal rainfall vector \(X_t\), while \(B\) is a \((m \times m)\) matrix of coefficients. Both \(A\) and \(B\) were estimated using the Stochastic Analysis, Modeling, and Simulation (SAMS) software developed by the U.S. Bureau of Reclamation and Colorado State University (Sveinsson et al. 2007).

e. Statistical measures

To assess the forecasting skill of ANN-GA, five goodness-of-fit statistics were computed: Pearson correlation \(r\), root-mean-square error (RMSE), Nash–Sutcliffe coefficient of efficiency (\(N_S\)), bias, and Hanssen–Kuipers (HK):

\[
r = \frac{1}{k} \sum_{i=1}^{k} \frac{(X_i - \bar{X})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{k} (X_i - \bar{X})^2 \sum_{i=1}^{k} (O_i - \bar{O})^2}}, \tag{8}
\]

\[
\text{RMSE} = \frac{1}{k} \sqrt{\sum_{i=1}^{k} (X_i - O_i)^2}, \tag{9}
\]

\[
N_S = 1 - \frac{\sum_{i=1}^{k} (X_i - O_i)^2}{\sum_{i=1}^{k} (X_i - \bar{X})^2}, \tag{10}
\]

\[
\text{bias} = \frac{1}{k} \sum_{i=1}^{k} (X_i - O_i), \quad \text{and} \tag{11}
\]

\[
\text{HK} = \frac{H - E_c}{T - E_m}, \tag{12}
\]

where \(O_i\) and \(X_i\) are the observed and simulated rainfall amounts, respectively; \(\bar{X}\) and \(\bar{O}\) are their respective means; and \(K\) is the total number of data points. Here,
r ranges from −1 to 1 with a correlation of 1 (−1) being a perfect positive (negative) relationship between $O_i$ and $X_i$, while $N_S$ may range from $-\infty$ to 1, with 1 corresponding to a perfect match, an $N_S$ of 0 indicating model predictions that are only as good as the mean of $O_i$, and an $N_S$ that is less than zero, indicating that the prediction is as good as random.

To compute the HK skill score, the forecasted and observed rainfall data are grouped into three categories: dry, near normal, and wet. The percentages of the JJAS rainfall below 33%, 33%–67%, and over 67% are used to compute the contingency table categories. Where $H$ is the total number of correct forecasts, $T$ is the total number of correct forecasts obtainable with a perfect forecast model, $N_c$ is the number of correct hits expected by chance, and $N_m$ is the marginal number of correct (observation) hits expected by chance. HK values range from $-1$ to $+1$. Perfect forecasts receive a score of 1, random forecasts receive a score of 0, and forecasts inferior to random forecasts receive a score of $-1$.

6. Results

a. Wavelet analysis of seasonal rainfall

The Morlet wavelet analysis and the SAWP computed for the gridded seasonal rainfall of Ethiopia are herein discussed. Elsanabary et al. (2014) show that the JJAS (kiremt) seasonal rainfall data of Ethiopia exhibit interannual oscillations at 2–4- and 5–7-yr cycles. The scale-average wavelet power at different scales at a particular time scale is called the local wavelet spectra, while the power summation at each scale for the entire time domain is called the global wavelet spectra. Both the local and the global wavelet spectra for the JJAS seasonal rainfall in Ethiopia were computed (see an example in Fig. 4). The same figures were produced for the SST (not shown). The local wavelet spectra represent the changes in wavelet power at various scales (periods) over time, while the global wavelet spectrum shows the averaging of all local wavelet spectra over time for each periodicity. Thick black contours in Fig. 4 indicate that the power is statistically significant at the 95% confidence level of the red-noise process. The dashed line encompassing the wavelet spectrum depicts the cone of influence (CI); outside the CI the spectra may be affected by zero padding and should be treated with caution. Figure 4 shows that in the 1950s, 1960s, and 1980s, strong power was detected at the interannual cycles of about 2–8 yr. A strong interdecadal oscillation of about 16 yr was also detected during 1950–80 for western parts of Ethiopia (UBNB; see Fig. 4a). Given that part of the interdecadal oscillation was outside the CI, and three episodes of statistically significant interannual oscillations of 2–8 yr were detected, the SAWP for rainfall and SST are estimated between the 2- and 8-yr cycles.

Figure 5 shows the spatial correlation between WPC1 of the JJAS rainfall with the SAWP for all grid points over Ethiopia. The strong positive correlation over the UBNB indicates that the variability of WPC1 mainly comes from this area of Ethiopia, which likely implies that the prediction skill of the ANN-GA model (section 5c) driven by the SST of the surrounding oceans (selected on the basis of a strong correlation with the WPC1 of JJAS rainfall) should be higher in the UBNB than in other parts of Ethiopia. From Fig. 5, we also subjectively select a correlation coefficient of 0.5 as the threshold, whereby we expect the prediction skill of the ANN-GA model will likely be adequate.

To highlight the rainfall signals that come from the UBNB, the power Hovmöller (PH) plot of 2–8-yr cycles is presented within a time–longitude framework (Fig. 6) at a latitude of 10.25°N (where the UBNB is located) for the twentieth century. It shows that in 1910–20, 1950–60, and 1980–90, the SAWP of the JJAS precipitation between 37° and 40.8°E had been statistically significant at the 5% significance level over the white noise (e.g., solid contours). To draw the PH plot, we used University of Delaware (UoD) rainfall data, which are highly correlated with the UEA data (refer to Elsanabary and Gan 2012; Elsanabary et al. 2014). The PH of the JJAS rainfall of the UoD at 2–8 yr for all the grid points (the $x$ axis represents longitudes) over a section passing by the middle of the UBNB (e.g., 10.25°N, where UoD and UEA are highly correlated) is presented in Fig. 6. Figure 6a
generally shows a strong spatial coherence across the UBNB (37° and 40°E) at three different time periods, while Fig. 6b shows zonally averaged power over time, and Fig. 6c shows the time-averaged SAWP as a function of longitude. Overall, on the basis of three local maxima detected between 37° and 40.8°E, significant interannual oscillations appear to have occurred over the UBNB in the last century.
b. Wavelet analysis of seasonal SST

From a Morlet wavelet analysis applied to the global, seasonal SST (FMAM), some sectors of the global SST exhibit statistically significant interannual oscillations at 2–8-yr cycles. Figures 7a and 7b show, respectively, the spatial correlation patterns between WPC1 and WPC2 of the SSTs and the SST SAWP of the FMAM season at 2–8-yr spectral band. WPC1 (WPC2) accounted for 31% (12%) of the total SST variance of the FMAM season.
c. Teleconnection between WPC1 of JJAS rainfall and SAWP of global FMAM SST

Figures 8a and 8b show the spatial correlation patterns between WPC1 of gridded precipitation during kiremt (JJAS) and each of the 65,000 SAWPs of gridded global SST during JFM (FMAM). As expected, WPC1 of rainfall is more strongly correlated to the SAWP of SST at one season than at two-season lead time, and so the JJAS rainfall is linked to the FMAM SST (see Fig. 8).
Figure 8a shows that JJAS rainfall is correlated with the El Niño region (equatorial Pacific Ocean (EPO), and the northern Atlantic Ocean (NAO, west of the Sahara Desert)). Some other areas include areas east and west of the South Pacific Ocean (SPO) of South America and areas south of the Indian Ocean (south Indian Ocean (SIO)). The correlation between rainfall WPC1 and JFM SST SAWP in the El Niño region appears to be relatively

![Wavelet power spectra for JFM seasons](image1)

**Fig. 9.** Wavelet power spectra of the leading first PC for the JFM seasonal oceanic SST sectors: (a) EPO, (b) SPO, (c) NAO, and (d) SIO. The thick black contours enclose the statistically significant wavelet power at the 5% level of a red-noise process, and the thin black curve is the CI.

![Wavelet power spectra for FMAM seasons](image2)

**Fig. 10.** Wavelet power spectra of the leading first PC for the FMAM seasonal oceanic SST sectors: (a) EIO, (b) SIO, (c) SAO, and (d) SPO. The thick black contours enclose the statistically significant wavelet power at the 5% level of a red-noise process, and the thin black curve is the CI.
weak (Elsanabary et al. 2014). Figure 8b shows that JJAS rainfall is correlated to FMAM SST of the northern and southern Indian Oceans, the southern Atlantic Ocean, and some sectors of the Pacific Ocean west of South America. From Fig. 8b and from the correlation between SAWP of FMAM SST and WPC1 of JJAS rainfall WPCs, we found that SSTs explain up to 64% of the rainfall variability in the 2–8-yr frequency band. Since JFM has not yet surpassed the spring barrier, we have to focus on the FMAM season as the forecasting season for the JJAS rainfall over the UBNB.

By linking wavelet transformed FMAM SSTs to JJAS rainfall, we can use the former as predictors in the ANN-GA model for forecasting the JJAS rainfall at one-season lead time. Areas where the correlation between FMAM SST and JJAS rainfall are above 0.4 are located in the equatorial Indian Ocean (EIO; 20°S–11°N, 61°–85°E), the SIO (59°–43°S, 34°–80°E), the southern Atlantic Ocean (SAO; 48°–35°S, 48°–27°W), and the SPO (41°–26°S, 112°–91°W) and are presented later (see Fig. 10b). Therefore, the FMAM SST data of the above ocean sectors were used as the predictors. The selected predictors are near the four climate indices [Pacific decadal oscillation (PDO), the Southern Oscillation index (SOI), the Indian Ocean dipole (IOD), and the Atlantic multi-decadal oscillation (AMO)] that Taye and Willems (2012) used to explain anomalous rainfall over Ethiopia through temporal variability of extreme high and low flows and rainfall.

d. Wavelet coherence and phase differences

The wavelet coherence between seasonal rainfall and PC1 of the identified SST sectors for JFM and FMAM are presented to provide more detailed information about their relationships, which change over time. Figures 9 and 10 show stronger wavelet coherence between them in the 2–8-yr cycle except at the SPO during JFM, where a 8–16-yr cycle is evident (see Fig. 9b).
The wavelet coherence between seasonal rainfall and PC1 of the identified SST sectors for JFM and FMAM are presented to provide more detailed information about their changing relationships. Figures 11 and 12 show the wavelet coherence and phase difference between WPC1 of the seasonal rainfall (JJAS) at UBNB and the PC1 of JFM (FMAM) SST for each of the four identified sectors (see Figs. 7a,b). In Figs. 12 and 13, the phase differences between WPC1 of JJAS rainfall and each of the four SST PC1s are plotted as vectors, where the arrows pointing toward the right indicate the two signals (rainfall and SST) are in phase while arrows pointing toward the left indicate the signals are out of phase. When arrows point upward (downward), it means that the rainfall WPC1 leads (lags) the PC1 of SST by 90°.

If SST leads the rainfall, it likely means that SST can be a predictor for forecasting seasonal rainfall at one-season lead time. The thick contours enclose periods of statistically significant coherence with respect to a red-noise process simulated by a Monte Carlo experiment of Jevrejeva et al. (2003).

From Fig. 11a, WPC1 of the JJAS rainfall and the selected JFM SST in the EPO region exhibits statistically significant coherence on an interannual scale (e.g., 1910s, 1930s, and 1960s), on a 4–8-yr scale in the 1920s, and on a 8–16-yr scale from 1930 to 1960. Notice that the existence of significant coherence between two signals does not necessarily depend on the existence of significant wavelet power in the two signals. The phase difference shows WPC1 of the UBNB rainfall generally changes from 90° to 0° during the 1920s (i.e., the rainfall leads the SST) and is in phase with the EPO SST from 1930 to 1960.

In spite of the generally weak SPO activities between the 1900s and 1960s, rainfall and SPO show high coherence on the 2–8-yr scale. The inconsistency in the relationship between rainfall and SPO is clearly evident from the phase distribution on the interannual scale, where
the phase difference changes from near 200° to 270° during the 1920s and 1950s to near 90° after the 1960s with a 10–20-yr cycle, which means that the JJAS rainfall is leading the JFM SST from that sector of the Pacific Ocean (Fig. 11b). Figure 11c shows that the phase distribution between JJAS rainfall and the NAO appears to be more inconsistent. From Fig. 11d, the SST from the SIO leads the rainfall by 90° during the 1950s and 1960s, which is a promising sector for forecasting the rainfall. The rainfall season was not quite affected by the JFM SST.

Since the JFM season has not yet surpassed the spring barrier, we have focused on the FMAM season as the forecasting season for the JJAS rainfall over the UBNB. Therefore, we tried to explore the SST just prior to the JJAS rainfall season (i.e., FMAM SST). Figure 12a shows that the WPC1 of the JJAS rainfall season exhibits statistically significant coherence on a 2–8-yr scale in the 1910s and 1920s with SST leading the rainfall, on a 25-yr scale between 1940 and 1960 with rainfall leading the SST by 90°, and on a 2–4-yr scale in the 1980s with rainfall lagging the SST by 250°. From Fig. 11b, WPC1 of the JJAS rainfall shows strong coherence with PC1 of the selected FMAM SST from SIO at a 2–4-yr scale during the 1910s, 1920s, 1950s, and 1980s (Fig. 12b).

Also, Fig. 12b shows a 15-yr cycle from 1920 to 1940, a 4–8-yr cycle from the 1950s to the 1990s, and a 2–4-yr cycle during only the 1980s. For these three cycles the rainfall was almost lagging the SST by 90°. In addition, Fig. 11c shows a promising indication for successful rainfall forecasting where the SST was leading the rainfall by 90° with a 4–10-yr cycle from 1910 to the 1920s and a 2–4-yr cycle in the 1930s, 1950s, and 1970s. Finally, the WPC1 of the JJAS rainfall season exhibits statistically significant coherence on a 4-yr scale in the 1910s and on a 2–4-yr scale during the 1930s with SST leading the rainfall.

The above results reveal that WPC1 of JJAS rainfall and the JFM SST sectors have lower coherence at 2–8-yr scales, while the WPC1 of JJAS rainfall has a higher level of coherence with the selected FMAM SST at 2–8-yr scales. The results show a generally high correlation between WPC1 of JJAS rainfall and SAWP of selected SST of the previous season. Therefore, the above sectors of FMAM of the global oceans are selected as predictors for the statistical model (ANN-GA) for forecasting the JJAS rainfall.

e. Seasonal rainfall forecasting

Given that the JJAS rainfall over the UBNB was strongly correlated with FMAM SST in EIO, SIO, SAO, and SPO at interannual time scales (Fig. 7b), it seems feasible to use these SST data to forecast the seasonal rainfall (JJAS) of the UBNB at one-season lead time. As an attempt to eliminate unnecessary input data, only data that explain the largest amount of variance are retained. From a scree plot (figure not shown) showing the
percentage of explained variances of various principal components (PCs), only the first PC of FMAM SST, which explains ≈90% of the total variance, was retained to forecast the JJAS rainfall. All input (SST PC1s) and output data (seasonal rainfall) of the ANN-GA were normalized before the analysis.

For the two grid points (points 4 and 7) located in the UBNB (Fig. 1), the ANN-GA was first calibrated using 21 yr of data (1975–95) to forecast the seasonal rainfall of the two grid points and independently validated using 6 yr of data (1996–2001). The time series of the forecasted and the observed seasonal rainfall of JJAS for the UBNB for both the calibration and validation stages are presented in Fig. 13. Summary statistics related to the performance of the ANN-GA at both stages are also given in Table 1. With a correlation coefficient $r$ of about 0.7 (i.e., $R^2 = 0.49$) and an HK score of 0.5 at calibration stage, ANN-GA is considered adequately calibrated, which is confirmed at the validation stage where both the $r$ and HK scores are maintained or even higher. However, in the foothills of the Great Rift Valley, $r$ for the validation stage was only about 0.3. Apparently, even though the forecasted JJAS rainfall for the central part of the UBNB, near the foothills of the GRV, shows strong correlation with the observed data, the forecasted JJAS rainfall for stations located at the edges of the UBNB are not, which is expected because WPC1 of JJAS rainfall in the central UBNB is strongly correlated with the SAWP of individual grid points ($r$ of 0.6–0.8), but it is poorly correlated ($r$ of 0–0.5) with the SAWP of individual stations along the basin edges (Fig. 5). Similar results are also found in terms of RMSE and HK scores.

Apparently, ANN-GA driven by the selected sectors of SST of some oceans has good predictability in the central UBNB. Next, the forecasted JJAS rainfall of the UBNB is disaggregated to weekly rainfall data, which will be of an adequate time scale for hydrologic modeling of the UBNB.

### f. Seasonal rainfall disaggregation

Observed seasonal (JJAS) rainfall amounts of the two grid points were disaggregated by the algorithm of Valencia and Schaeke (1973, hereafter VS) (Figs. 14a and 14b, respectively, for grid points 4 and 7). The rainfall data of the first 21 yr (1975–1995) were used to calibrate parameters of the disaggregation model and the rainfall data of the last six yr (1996–2001) were used to validate the calibrated disaggregation model. The results in terms of Pearson correlation, RMSE, $N_S$, and bias are given in Table 2. The disaggregated rainfall at point 4 accounted for 45% of the observed weekly rainfall variability, while the RMSE was 1.28 mm week$^{-1}$, $N_S$ was −0.23, and bias

<table>
<thead>
<tr>
<th>Grid point</th>
<th>$R^2$</th>
<th>RMSE</th>
<th>$N_S$</th>
<th>Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.45</td>
<td>1.28</td>
<td>−0.23</td>
<td>0.00</td>
</tr>
<tr>
<td>7</td>
<td>0.44</td>
<td>1.58</td>
<td>0.03</td>
<td>0.01</td>
</tr>
</tbody>
</table>

### Table 1. Summary statistics for the forecasted JJAS rainfall season.

<table>
<thead>
<tr>
<th>Grid point</th>
<th>Bias</th>
<th>RMSE</th>
<th>$r$</th>
<th>HK</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>−1.60</td>
<td>0.78</td>
<td>0.68</td>
<td>0.50</td>
</tr>
<tr>
<td>Calibration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validation</td>
<td>1.56</td>
<td>0.72</td>
<td>0.79</td>
<td>0.50</td>
</tr>
<tr>
<td>7</td>
<td>−0.20</td>
<td>0.82</td>
<td>0.71</td>
<td>0.50</td>
</tr>
<tr>
<td>Calibration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validation</td>
<td>0.19</td>
<td>0.75</td>
<td>0.77</td>
<td>0.77</td>
</tr>
</tbody>
</table>

### Table 2. Summary statistics for the statistically disaggregated weekly rainfall during the JJAS rainy season.
was zero. In addition, the disaggregated rainfall at point 7 accounted for 44% of the weekly rainfall variance, with RMSE slightly higher (1.58 mm week$^{-1}$), $N_S$ (0.03), and bias (0.01 mm). The disaggregated weekly seasonal rainfall amounts at grid points 4 and 7 are shown in Figs. 15 and 16, respectively.

7. Summary and conclusions

The June–September (JJAS) seasonal rainfall of the upper Blue Nile basin (UBNB) of Ethiopia is essential to the agriculture and water resource management of western Ethiopia because it constitutes a significant proportion of the country’s annual rainfall. Therefore, to effectively manage the water resources of the UBNB, it will be important to accurately forecast its weekly JJAS rainfall, preferably at a seasonal lead time. Three key findings of this study include the following.

1) WPCs and SAWP of seasonal FMAM SST of the Indian, Atlantic, and Pacific Oceans were computed. Then, from the correlation field between the SAWPs of the FMAM SST of the three oceans and the WPC1 of JJAS seasonal rainfall over the UBNB, sectors of the Indian, Atlantic, and Pacific Oceans where SST shows a strong teleconnection with the JJAS rainfall of UBNB ($r \geq 0.4$) were identified. High wavelet coherence was also found between WPC1 of selected sectors of FMAM SST and WPC1 of seasonal rainfall of the following season (JJAS), which is consistent with the correlation analysis between WPC1 of JJAS rainfall and SAWP of the selected sectors of FMAM SST.
2) Seasonal FMAM SST data in these identified sectors of the three oceans were used as predictors in a calibrated ANN-GA model to forecast the seasonal JJAS rainfall of the UBNB at one-season lead time. The sectors of FMAM SST fields chosen were located in the equatorial Indian Ocean (20°S–11°N, 61°–85°E), the southern Indian Ocean (59°–43°S, 34°–80°E), the southern Atlantic Ocean (48°–35°S, 48°–27°W), and the southern Pacific Ocean (41°–26°S, 112°–91°W). Results show that ANN-GA forecasted seasonal rainfall amounts agree well with the observed data for the highlands of the UBNB, specifically the UBNB (RMSEs between 0.72 and 0.82, correlation between 0.68 and 0.77, and HK scores between 0.5 and 0.77). The model shows limitations when forecasting in the foothills, especially for the GRV region, which is expected since the variability of WPCI mainly comes from the highlands of Ethiopia. To improve the model’s performance, the moving average technique, a preprocessing technique, could be used, as recommended by Wu et al. (2009).

3) The disaggregation model of VS was used to disaggregate the seasonal rainfall to weekly rainfall, which can then be used to drive a hydrologic model to forecast the streamflow of the UBNB at one-season lead time. Even though there are discrepancies between disaggregated and observed weekly rainfall, and an $R^2$ of 0.44, as seen in scatterplots in Fig. 14, the data points generally scatter around the 45° line. In addition, the weekly, disaggregated rainfall reasonably captures the characteristics of the observed weekly rainfall, as shown in the time series plots in
Figs. 15 and 16. The ability to forecast the rainfall of the UBNB at seasonal lead time will be useful for an optimal allocation of water usage among various competing users in the river basin. Our future research will focus on forecasting the weekly streamflow of the UBNB in Ethiopia at a seasonal lead time. Also, using other rainfall forecasting models, such as crisp distributed artificial neural network (CDANN) and crisp distributed support vectors regression (CDSVR), will be explored.

8. Recommendations for future work

Given that SSTs are generally predictable by climate models at up to 6-month lead time, we have developed a wavelet-based, ANN-GA model driven by selected SSTs as predictors to predict the seasonal precipitation of the UBNB in Ethiopia. Using the model of VS, we disaggregated the predicted seasonal rainfall to weekly time-scale rainfall data. As an extension to this study, we will use the disaggregated, weekly rainfall data to drive a basin-scale hydrologic (rainfall–runoff) model, such as the Swedish Hydrologiska Byrån Vattenbalansavdelning (HBV) or the National Weather Service’s Sacramento model, to simulate the streamflow data of the UBNB. If the predicted streamflow results agree well with observed streamflow data from the UBNB, it will be possible to apply this combined climate–hydrologic system, driven by SST data predicted by climate models, to forecast the streamflow of the UBNB at up to one-season lead time, which will be very useful for the future seasonal planning and management of the water resources of the UBNB, especially during dry years.

Acknowledgments. This research was funded by the Egyptian Ministry of Higher Education (MHE). The first author was also partly funded by a graduate teaching assistantship of the University of Alberta. The authors thank Dr. Mike Hulme for the observed precipitation dataset, gu23wld0098.dat (version 1.0), constructed at and supplied by the Climatic Research Unit, UEA, Norwich, United Kingdom. We acknowledge Dr. Chun-Chao Kuo for his valuable discussions on ANN-GA. The Wavelet software was provided by Torrence and Compo, and may be downloaded online (http://atoc.colorado.edu/research/wavelets/). The Wavelet coherence software was provided by A. Grinsted and may be found online (http://www.pol.ac.uk/home/research/waveletcoherence/). The SAMS 2007 software is downloaded from the Colorado State University website (http://www.sams.colostate.edu/download.html). Great appreciation is expressed to the two anonymous reviewers of this paper for their comments and suggestions.

REFERENCES


