Theoretical and Applied Climatology
https://doi.org/10.1007/s00704-017-2358-4

ORIGINAL PAPER

Prediction of onset and cessation of austral summer rainfall and dry spell frequency analysis in semiarid Botswana

Jimmy Byakatonda[1,2] · B. P. Parida[1] · Piet K. Kenabatho[3] · D. B. Moalafhi[3]

Received: 14 November 2016 / Accepted: 25 December 2017

Abstract
Uncertainties in rainfall have increased in the recent past exacerbating climate risks which are projected to be higher in semiarid environments. This study investigates the associated features of rainfall such as rain onset, cessation, length of the rain season (LRS), and dry spell frequency (DSF) as part of climate risk management in Botswana. Their trends were analysed using Mann-Kendall test statistic and Sen's Slope estimator. The rainfall-evapotranspiration relationships were used in formulating the rain onset and cessation criteria. To understand some of the complexities arising from such uncertainties, artificial neural network (ANN) is used to predict onset and cessation of rain. Results reveal higher coefficients of variation in onset dates as compared to cessation of rain. Pandamatenga experiences the earliest onset on 28th of November while Tsabong the latest on 14th of January. Likewise, earliest cessation is observed at Tshane on 22nd of February and the latest on 30th of March at Shakawe. The shortest LRS of 45 days is registered at Tsabong whereas the northern locations show LRS greater than 100 days. Stations across the country experience strong negative correlation between onset and LRS of − 0.9. DSF shows increasing trends in 50% of the stations but only significant at Mahalapye, Pandamatenga, and Shakawe. Combining the LRS criteria and DSF, Kasane, Pandamatenga, and Shakawe were identified to be suitable for rainfed agriculture in Botswana especially for short to medium maturing cereal varieties. Predictions of onset and cessation indicate the possibility of delayed onset by 2–5 weeks in the next 5 years. Information generated from this study could help Botswana in climate risk management in the context of rainfed farming.

Keywords Artificial neural network · Climate variability · Inter tropical convergence zone · Length of the rain season · Rainfed farming · Trend analysis

1 Introduction

In semiarid Botswana where approximately 80% of the populace is engaged in rainfed agriculture and the country's main source of fresh water is rainfall (Statistics Botswana 2009, 2015; Batisani and Yarnal 2010), uncertainties in rainfall are injurious to livelihoods and national economy as a whole. Rainfall has been identified as one single factor that influences agricultural yields as well as quantities of surface, ground water and soil moisture storage (Sivakumar 1988; Mugalavai et al. 2008; Amekudzi et al. 2015). In many instances, average conditions are used in planning and management of both agricultural and water resources (Usman and Reason 2004). However, a rain season classified as near normal may still record poor harvest due to high variability and short rainfall episodes punctuated with relatively longer dry spells. For better crop yields, well distributed rains just enough to meet crop water requirements is sufficient than intermittent heavy showers that registers high rainfall totals (Usman and Reason 2004; Kebede et al. 2016).

Various studies in regards to analysis of onset and cessation of rain conducted in both West and East Africa showed high association between rain onset and length of rain season (Sivakumar 1988; Odekunle 2006; Mugalavai et al. 2008;

✉ Jimmy Byakatonda
byakatondaj@hotmail.com

[1] Department of Civil Engineering, University of Botswana, P/Bag, 0061 Gaborone, Botswana

[2] Department of Biosystems Engineering, Gulu University, P.O.Box 166, Gulu, Uganda

[3] Department of Environmental Science, University of Botswana, P/Bag 00704, Gaborone, Botswana

Published online: 07 January 2018

Amekudzi et al. 2015). They associated early rain onset with longer seasonal rains. However in these studies, trends in onset and cessation of rain were not investigated to determine the direction and magnitude of any possible changes in these features of rainfall over time.

Since variability in seasonal rainfall totals alone may not explain the relationship between rainfall and crop yields, dry spell frequency analysis has usually been undertaken for different parts of the globe as a means of further investigating the implications of changing climate on food security (Usman and Reason 2004; Araya et al. 2010; Ngetich et al. 2014). A number of definitions exist for dry spells, but for this study, it is taken as extended periods with no rainfall within a rain season (Ngetich et al. 2014). Dry spells are crucial in this study since they are directly linked to crop moisture stress. Studies in semiarid environments in Sub-Saharan Africa have indicated that dry spells range between 5 and 15 days (Usman and Reason 2004; Ngetich et al. 2014). Regionally, studies analysing dry spells have been conducted in the southern Africa using gridded data by Usman and Reason (2004). However, gridded data is sometimes coarse and lacks site-specific attributes. For this reason, this study proposes the use of locally observed ground data for Botswana.

Rain water in Botswana is harvested and stored mainly in surface reservoirs spread across the country. The stored water is used for domestic, industrial and agricultural purposes (mainly for watering animals). Agriculture in Botswana is predominantly rainfed and mainly confined in the east to northeast. This implies that if uncertainties associated with rainfall are not well managed, both agricultural and water resources will be severely affected. The study area receives 90% of her rainfall in the austral summer between November and March (GOB-MMEWR 2006; Batisani and Yarnal 2010) which is the season proposed for analysis in this study. This season coincides especially with the southward shifting of the inter tropical convergence zone (ITCZ) and westerlies that are responsible for most of the rains over southern Africa (Nicholson et al. 2001; Usman and Reason 2004). On the other hand, rainfall patterns are also commonly reported to be affected by climate variability and change which are closely associated with El Niño southern oscillation (ENSO) that reaches its peak during austral summer (Nicholson et al. 2001; Recha et al. 2012; Byakatonda et al. 2018).

For successful rainfed farming, correct timing of agricultural activities such as seed bed preparation, sowing and harvesting is of paramount importance (Araya and Stroosnijder 2011; Amekudzi et al. 2015; Kebede et al. 2016). Therefore, prediction of onset and cessation dates would go a long way in maximising returns from rainfed farming. Research carried out in semiarid areas of West Africa and South Africa by Odekunle (2006) and Moeletsi and Walker (2012) respectively indicate high variation in onset and cessation dates sometimes exceeding 10 weeks making accurate predictions a challenging task. To mitigate this problem, use of artificial intelligence that mimics human learning and reasoning abilities (Masinde 2014) is proposed. Many forms of artificial intelligence exist but the most prominent of all are artificial neural network (ANNs) which is found to perform well in nonlinear multivariate complex environments (Chen et al. 2010; Machiwal and Jha 2012; Masinde 2014; Byakatonda et al. 2016, 2018). With continued uncertainties in climate patterns and frequent extreme weather conditions globally, the use of ANNs in climate variability studies has increased in the recent past in an attempt to understand these unmeasurable nonlinear conditions (Chen et al. 2005; Mishra and Desai 2006; Liu et al. 2010). ANNs have been found to give reasonable results when applied to short-term predictions. It is a tool capable of learning any data combination regardless of the process that generated it (Chang et al. 2015; Byakatonda et al. 2016, 2018).

This study thus investigates onset and cessation of rain during summer season including their trends for semiarid Botswana between the years 1960 and 2014. The study specifically attempts to: (1) determine rain onset, cessation and dry spell frequency, (2) determine direction and magnitude of trends in onset, rain cessation and dry spell frequency, and (3) develop a 5 year ahead prediction model for rainfall onset and cessation dates and provide demonstrations to aid timely farming operations. With declining crop yields in the recent past across the study area mainly attributed to erratic rainfall, it is necessary to generate information that can reduce uncertainties associated with climate variability.

2 Study area

Botswana, the study area, is located in southern Africa bordering Zambia in the north, Zimbabwe in the east, Namibia in the west and South Africa in the south. It occupies a land area of 580,000 km^2 with a human population of 2,210,000 (Statistics Botswana 2015). The topography of the study area is mainly characterised by flat terrain ranging from 900 to 1100 m above mean sea level (AMSL). The main economic activities in Botswana include trade, mining, tourism and agriculture. Botswana's economy largely depends on the mining sector that contributes more than 20% of the national GDP (Statistics Botswana 2015). Agriculture which is mainly subsistence employs close to 80% of the population but only contributes 3% to the national growth domestic product (GDP) (Byakatonda et al. 2018). The agricultural sector has persistently suffered poor yields lower than the continental average attributed to uncertainties in rainfall amounts and patterns. These uncertainties are mainly due to absence of information on rainfall variables such as dry spell frequency, onset and cessation of rain dates which are crucial for any successful

rainfed agriculture venture. This study therefore seeks to determine and analyse these rainfall variables in an effort to improve agricultural productivity.

Botswana's climate is predominantly influenced by two important features viz.: the Okavango delta in the north and Kalahari desert in the west. Its climate is classified as semiarid based on the Koppen's climate classification (Alvares et al. 2013). The study area is located in the subtropics with annual rainfall totals ranging from 250 mm in the southwest to 600 mm in the northeast. Rainfall distribution mostly depends on spatial location and season of the year. There are three different rainfall formation mechanisms over the study area with rainfall in the north mainly resulting from shifting of the ITCZ between November and March (GOB-MEWT 2012). The central and southern locations for most of the time depend on the easterlies and westerlies from the Indian and Atlantic Oceans. Most of the rain is received during the summer period between November and March (Batisani and Yarnal 2010). During the winter months (May to September) which are characterised by cold temperatures and dry winds, rainfall is largely depressed.

3 Data and methods

3.1 Data

Determination of onset and cessation of rain dates requires that both rainfall and temperature time series are available. To achieve this, 14 synoptic stations spread across the study area as shown in Fig. 1 were selected based on the following criteria;

i. A station was required to have one continuous data with less than 5% missing values and
ii. Have records of rainfall and maximum and minimum temperature of equal record length not less than 15 years.

The climatological year according to the Department of Meteorological Services (DMS) of Botswana starts on 1st of July and ends on 30th of June. For purposes of demarcating the rainfall season, rainfall mass curves for the 14 synoptic stations were plotted for the July–June climatological year. To enable comparison, the plots were made for a common record period between 1998 and 2014.

Daily rainfall and temperature data recorded at the 14 synoptic stations as shown in Table 1 were obtained from the Department of Meteorological Services (DMS) of Botswana. Rainfall and temperature time series were of varying length with the longest record between 1960 and 2014 and the shortest ranged from 1998 to 2014. One of the challenges of this study was finding stations with uniform record length of data sets and for this reason observed changes are limited to climate variability. Data for 2015 and 2016 were available but only used in evaluation of the artificial neural network (ANN) model predictions. For easier data management, rainfall time series are arranged according to the climatological year. This enabled aggregating austral summer rainfall as one continuous time series.

The FAO/WMO-recommended Penman Monteith formula for computing potential evapotranspiration (ET_0) was not applied due to absence of required data on sunshine hours, relative humidity, wind speed and solar radiation. For this reason, the Hargreaves method described in Allen et al. (1998) and applied in Byakatonda et al. (2016) was used to compute ET_0. This method has been applied in semiarid areas and found satisfactory in comparison with other temperature methods (Beguería et al. 2014; Stagge et al. 2014).

3.2 Identification of onset and cessation of rain

Numerous methods exist for determining onset and cessation of rain dates as presented in Sivakumar (1988), Recha et al. (2012), Ngetich et al. (2014) and Amekudzi et al. (2015); however, most of these methods rely only on rainfall data to determine the onset and cessation dates. Due to the important role ET_0 plays in plant moisture deficit, the rainfall-evapotranspiration relationship as described in Araya and Stroosnijder (2011), Kebede et al. (2016) and Byakatonda et al. (2016) is used in this study to formulate a criteria for identifying onset and cessation of rain dates.

Onset is assumed to occur when pentad rainfall equals or exceeds half of the pentad ET_0 (Araya and Stroosnijder 2011; Byakatonda et al. 2016). This is true under the following objective criteria: (1) these conditions are maintained for the next 10 days with rainfall totals exceeding 25 mm and (2) no dry spell exceeding two pentads occur within 30 days of identified onset date.

Conversely cessation is assumed to occur after the onset date when the pentad rainfall total falls below half the pentad ET_0 under the following objective criteria: a dry spell exceeding two pentads occurs after the deficit is recorded and cessation date is then taken as the 7th day from the day the deficit started. The length of the rain season is computed as the number of days between onset and cessation dates. Number of rain days in a season is determined as the total number of days that registers rainfall more than 1 mm.

Time series of onset, cessation dates and length of growing season are generated for further analysis of trends and determination of coefficient of variation (CV). Mean rain onset and cessation dates are determined for each station over the study area. Frequency analysis is also performed on the generated time series of onset and cessation dates at each of the 14 synoptic stations using histograms. This is done in order to determine possible distinct dates over the

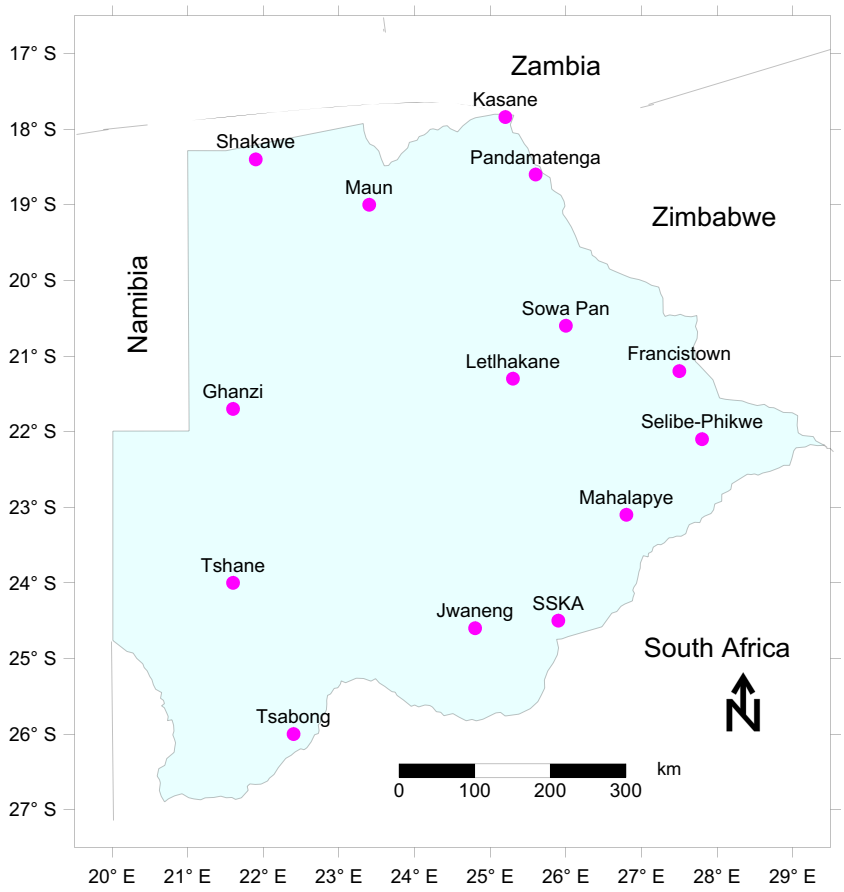

Fig. 1 Spatial distribution of synoptic stations used in the study

period of study that can be used to inform farmers of sowing dates. For locations where onset and cessation dates are spread over a long period of time, ANN is used to predict and identify the probable dates.

3.3 Dry spell frequency analysis

Knowledge on the frequency of dry spells is crucial for successful rainfed venture farming. Dry spells were aggregated

Table 1 Station locations and length of meteorological variable record

SN	Station ID	Station name	Latitude	Longitude	Elevation.	Period of record		Seasonal rainfall	
			°S	°E	Amsl (m)	Rainfall	Temperature	Totals (mm)	CV (%)
1	033-FRAN	Francistown	21.2	27.5	968	1960–2014	1960–2014	405.9	43
2	039-GANT	Ghanzi	21.7	21.6	1131	1960–2014	1961–2014	371.3	43
3	053-JWAN	Jwaneng	24.6	24.8	935	1988–2014	1989–2014	354.8	32
4	064-KASA	Kasane	17.8	25.2	960	1968–2014	1983–2014	539.9	29
5	093-LET2	Letlhakane	21.3	25.3	991	1993–2014	1994–2014	347.2	42
6	106-MAHA	Mahalapye	23.1	26.8	1005	1960–2013	1971–2014	389.4	39
7	130-MAUN	Maun	19.0	23.4	945	1960–2014	1965–2014	407.1	39
8	183-PAN2	Pandamatenga	17.8	28.6	1071	1998–2014	1998–2014	466.4	33
9	213-SEL2	Selibe-Phikwe	23.1	37.8	892	1998–2014	2000–2014	291.3	40
10	223-SHAK	Shakawe	18.4	21.9	1030	1960–2014	1965–2014	480.9	37
11	035-SSKA	SSKA	24.7	25.9	975	1985–2014	1985–2014	374.0	40
12	033-SUAP	Sowa Pan	20.6	26.0	908	1992–2014	2000–2014	400.1	50
13	244-TSAB	Tsabong	26.0	22.4	960	1960–2014	1961–2014	229.0	50
14	251-TSHA	Tshane	24.0	21.6	1118	1960–2014	1961–2014	275.7	49

at different intervals of 10 and 15 days in West Africa by Sivakumar (1992) and 5, 10 and 15 days by Ngetich et al. (2014) in East Africa. Aggregation at 5 day (Pentads) intervals is made for this study to include finer and lower bounds of dry spells. Dry spell frequency is determined as the number of pentads with rainfall less than 5 mm. Time series of dry spell frequency is generated at each station for further analysis.

The homogeneity of summer rainfall variables time series in form of onset, cessation, length of the rain season and dry spell frequency is tested using four absolute tests at each station used in the study. These are the standard normal homogeneity test (Alexandersson 1986), Pettit test (Pettit 1979), Buishand range test (Buishand 1982) and the Von Neumann ratio test (Von Neumann 1941). None of the stations exhibit statistically significant inhomogeneity in time series at 1% significant level and hence this allowed them for further trend and variability analysis (Wijngaard et al. 2003; Costa and Soares 2009).

3.4 Trend analysis of rainfall variables

Non parametric Mann-Kendall (MK) monotonic trend test is used in this study to investigate the direction of trend in the summer rainfall variables. This method has found wide application in hydro meteorological studies (Kampata et al. 2008; Tabari et al. 2011; Some'e et al. 2013; Akinsanola and Ogunjobi 2015; Kebede et al. 2016). The MK is given by

$$S = \sum_{q=1}^{k-1} \sum_{r=1}^{k} \mathrm{Sgn}\left(P_r - P_q\right) \quad (1)$$

where P_r and P_q represents a rainfall variable e.g. onset dates at time intervals t_r and t_q ($r > q$) and k is the total number of data points available in the time series for analysis. Details of the Sgn function and the MK-Z statistic can be found in Kampata et al. (2008).

Positive values of the Z-statistic indicate increasing trend in a given seasonal rainfall variable while negative Z values represent decreasing trend. The null hypothesis (H0) of no trend is accepted if the computed Z value is less than a critical value obtained from a standard normal table. The alternative hypothesis (H1) assumes existence of trend in the time series (Akinsanola and Ogunjobi 2015; Some'e et al. 2013). For this study a 95% confidence level is applied for test of significance. The effect of serial correlation on the trend is investigated through testing for significance of the lag 1 correlation. For stations with significant lags at 95%, the effective sample size technique as described in Gocic and Trajkovic (2013) is used.

3.5 Sen's slope estimator

While the MK test in Section 3.4 gives the trend direction and Z statistic, it does not give the rate of change. In view of this, the Sen's slope estimator is used to quantify the slope. The Sen's slope method has found application in various trend analysis studies (Yue and Wang 2004; Some'e et al. 2013; Akinsanola and Ogunjobi 2015; Sabzevari et al. 2015). The slope of m-paired time series according to the Sen's slope estimator given in Sen (1968) and applied in Sabzevari et al. (2015) and Akinsanola and Ogunjobi (2015) is given by

$$Q_k = \frac{P_r - P_q}{t_r - t_q} \quad \text{for } k = 1, 2 \ldots . m \quad (2)$$

The median of the Q_k time series arranged in ascending order results is the Sen's slope estimate.

3.6 Predicting onset and cessation of rain dates using artificial neural network (ANN)

Stations with not well defined period of rain onset or cessation as analysed with histograms in Section 3.2 are further investigated with ANN to identity the probable dates. ANN have various combinations of different configurations but the race is between static ANN (back propagation neural network) and the dynamic ANN (recurrent neural network) (Menezes and Barreto 2008; Lahmiri 2016). Results from various prediction models have revealed that dynamic ANN which incorporates a time delay through feedback connections mimics accurately neural-biological information an important aspect in time series modelling (Maier and Dandy 2000; Chang et al. 2014, 2015). This study proposes to develop a multistep ahead rain onset and cessation date prediction model with training, testing and validation performed over a period of historical record for each station as shown in Table 1. Multistep ahead predictions are more complex than a one step ahead and as such require more model memory (Sorjamaa et al. 2007; Chang et al. 2014). For this reason, a nonlinear autoregressive with exogenous input (NARX) network, a class of recurrent neural networks (RNN), is selected for use in this study since it has been found to outperform static ANNs in hydrological time series modelling (Gao and Meng Joo 2005; Menezes and Barreto 2008; Byakatonda et al. 2016; Lahmiri 2016).

NARX artificial neural network is a class of dynamic networks that have been found to perform well in prediction of nonlinear and randomised time series due to their flexibility (Diaconescu 2008; Anari et al. 2011; Nasr and Zahran 2014). The NARX model construction process and prediction criteria can be summarised in the following steps:

1. Data preprocessing by standardising both input and target series using the Mapminmax option
2. Building the network in form of exogenous input nodes and autoregressive target nodes.
3. Choosing the number of hidden layers (in this case a single layer was used).

4. Training of the network in the open loop mode using Levenberg-Marquardt optimisation function with termination based on multi criteria including performance, gradient and number of epoch.
5. Prediction is then conducted in the closed loop mode and its performance evaluated through coefficient of correlation (R) and mean square error (MSE).

3.6.1 Data preprocessing

In an effort to optimise predictions, time series of target and input predictors are inspected for consistency. The inputs are selected based on the fact that they are known factors that can externally affect the outcome of the target series (Huang et al. 2006; Diaconescu 2008). In this study, seasonal rainfall totals, dry spell frequency, number of rain days and length of the rain season are selected as exogenous inputs. Onset and cessation dates are the target series which at the same time act as the autoregressive component of the model.

The selected data are then rescaled to a standardised range between − 1 and 1 for easier learning of the network. This is achieved through the Mapminmax function in the ANN toolbox through the following equation by Morid et al. (2007) and applied in Masinde (2014);

$$X_{\mathrm{p}} = C_{\min} + \frac{(X_{\mathrm{t}} - X_{tmin})}{(X_{tmax} - X_{tmin})}.(C_{\max} - C_{\min}) \tag{3}$$

where X_{p} is the rescaled data series, X_{t} is the original series, $C_{\min} = -1$ and $C_{\max} = 1$, X_{tmin} = minimum value of X_{t} series and X_{tmax} is maximum value of X_{t} series.

3.6.2 NARX construction, training and learning

The NARX model consists of input, hidden and output layers with two tapped delay elements originating from input and output regressors. Hidden layer neurons are chosen based on various network combinations with the final number following the procedure proposed by Hecht-Nielsen (1987), applied in Maier and Dandy (2001), Stathakis (2009) and Byakatonda et al. (2016). The procedure states that $\mathrm{H_L} \leq 2\mathrm{X_p} + 1$ where $\mathrm{H_L}$ is the number of neurons in the hidden layer and X_{p} number of predictors. The hidden neurons arrived at for this study is 9 embedded in a single hidden layer. This dynamic recurrent neural network with feed forward back propagation utilises a sigmoid transfer function in the hidden layer and a linear transfer function in the output layer. The Levenberg-Marquardt optimisation function is used to update the weights and bias values during the training mode. This function keeps to minimum the squared errors and weights while the right combination that generalises well the network is being generated (Demuth et al. 2009; Byakatonda et al. 2016). Different training algorithms are used and the one adopted in this study (Levenberg-Marquardt) gives the best performance with the lowest mean squared error (MSE) and highest regression coefficient between the predicted and actual targets.

The model topology is presented in Fig. 2. The recurrent connections from the outputs may delay a number of unit times to form new inputs. The nonlinear system for N-steps predictions ($N \geq 1$) can be represented mathematically as

$$\hat{y}(t+N) = \left\{ \left[X_1(t+N-1), \ldots, X_1(t+N-d); X_2(t+N-1), \ldots, X_2(t+N-d) \ldots; .. X_{\mathrm{p}}(t+N-1), \ldots, X_{\mathrm{p}}(t+N-d) \right]; \left[\hat{y}(t+N-1), \ldots, \hat{y}(t+N-d) \right] \right\} \tag{4}$$

where d is the input and output delay (memory orders) ≥ 1, p is the number of predictors available at the input regressor. $y(t+N)$ is output value at $(t+N)$ time step, $\hat{y}$ are estimated target series, $X_{\mathrm{p}}(t)$ input vector with p arrays where $p = 1,2,3,\ldots$ whereas $f(.)$ is a nonlinear mapping function that is approximated by the learning algorithm. The network has two regressors: the regressor $y(t+N-d)$ acts as an autoregressive model while $X_{\mathrm{p}}(t+N-d)$ is the implicit exogenous variable in the time series. In this model, the target series are predicted using their past values to generate future values. Their prediction is assisted by exogenous inputs $X_{\mathrm{p}}(t+N-d)$ which are external predictors that influence the outcome of the target series.

Training of the network and its application in prediction are achieved in two ways. The training mode is the series parallel connection (Open loop), where the output regressor at the input layer is comprised of only actual values of target series and the network only makes a one-step ahead prediction. This is presented mathematically as

$$\hat{y}(t+1) = \left\{ \left[X_1(t), \ldots, X_1(t+1-d); X_2(t), \ldots, X_2(t+1-d) \ldots; .. X_{\mathrm{p}}(t), \ldots, X_{\mathrm{p}}(t+1-d) \right]; \left[y(t), \ldots, y(t+1-d) \right] \right\} \tag{5}$$

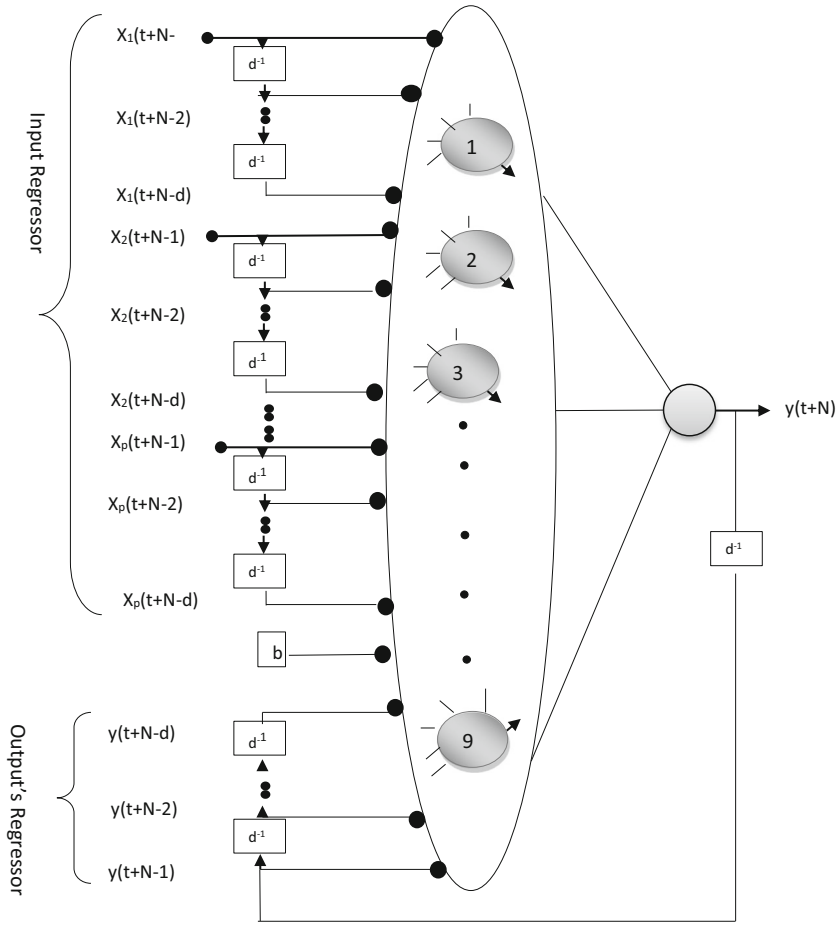

Fig. 2 Nonlinear autoregressive with exogenous input (NARX) network prediction model with d^{-1} unit time delays

where $y(t)$ are actual measured target values at the output's regressor.

3.6.3 NARX prediction and performance criteria

In a multistep prediction mode, the parallel mode (closed loop) is used. In this mode, the estimated outputs are fed back to the output's regressor as presented mathematically in (4). When the network is in the prediction mode, the input and target time series are divided into two groups. The first group is without the most recent time series data whose length is equivalent to the prediction horizon. For input series $X_p(t)(1,2,3,\ldots,M) \in X$ and M is the total number of data series, the length of the first group is considered as $(M–N)$ used in training and testing the network. This arrangement allows the model to learn the input-output relationship without recorded data. The second group of input data set is equivalent to N most recent historical values used for simulation of the model and generating of new target series, whereas the actual target series is responsible for validation of the predicted values. This study makes 5 years ahead prediction of onset and cessation dates for 2015 to 2019.

Model performance is evaluated based on errors from training, validation, and testing phases. In this study, the mean square error (MSE) and correlation coefficient (R) between the targets and outputs from the neural network model are used together with the autocorrelations (Diaconescu 2008; Chang et al. 2014, 2015) as follows:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N} e_i^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{6}$$

$$r_{y\hat{y}} = \frac{\sum_{i=i}^{N}(y_i - \bar{y}_i)\left(\hat{y}_i - \bar{\hat{y}}_i\right)}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2}\sqrt{\sum_{i=1}^{N}\left(\hat{y}_i - \bar{\hat{y}}_i\right)^2}} \tag{7}$$

where y_i and $\hat{y}_i$ are the observed and predicted values respectively and N is the horizon of prediction. $\bar{y}_i$ and $\bar{\hat{y}}_i$ are average values of observed and predicted data respectively. Historical records of 2015 and 2016 exist and they are used to compare with predicted values as part of evaluating the prediction ability of the model.

Table 2 Seasonal rainfall characteristics

Station	Onset date		Cessation date		Length of rain season		Dry spell frequency	
	Avg.	CV(%)	Avg.	CV(%)	Avg.	CV(%)	Avg.	CV(%)
Francistown	29-Nov	17.6	10-Mar	13.4	102	33.6	18	19.6
Ghanzi	12-Dec	25.9	14-Mar	17.0	92	46.6	17	17.6
Jwaneng	15-Dec	23.0	20-Mar	12.3	95	45.8	16	15.9
Kasane	2-Dec	12.6	19-Mar	8.1	107	21.4	13	26.7
Letlhakane	1-Jan	19.4	16-Mar	12.5	74	38.6	18	14.9
Mahalapye	3-Dec	21.4	2-Mar	15.5	90	44.7	17	17.1
Maun	17-Dec	21.4	10-Mar	14.0	83	41.4	15	24.9
Pandamatenga	28-Nov	13.2	27-Mar	8.2	120	14.5	14	19.8
Selibe-Phikwe	5-Dec	29.8	11-Mar	20.9	96	47.4	20	12.6
Shakawe	10-Dec	16.8	30-Mar	9.5	112	29.3	14	24.5
SSKA	2-Dec	21.3	7-Mar	16.5	95	40.5	18	14.9
Sowa Pan	3-Dec	23.3	13-Mar	8.1	101	36.7	17	22.6
Tsabong	14-Jan	27.5	28-Feb	22.9	45	55.1	20	17.3
Tshane	28-Dec	23.7	22-Feb	22.3	56	58.3	19	14.7

4 Results and discussion

Results comprises of averages of seasonal rainfall totals, rain onset dates, cessation of rain dates, length of the rain season and dry spell frequency. Also their coefficients of variation (CV) are analysed and presented in Tables 1 and 2. Long term changes in these rainfall variables have been investigated using MK trend test statistic and Sen's slope estimator, their results are also presented in Table 3. This section also presents rainfall mass curves for the 14 synoptic stations.

4.1 Rainfall mass curves

For purposes of demarcating the summer rainfall season, rainfall mass curves were plotted for the July–June climatological year. The results from this analysis are presented in Fig. 3. From the plots, it is evident that rainfall contribution is registered from October–November and ceases during March–April at most of the stations. The biggest contribution is observed between December and February when the curves exhibit the steepest positive gradient that

Table 3 MK trends Z-statistic and Sen's slope estimator for rainfall characteristics

Station	Onset date		Cessation date		Length of rain season		Dry spell frequency	
	MK-Z	Sen-S	MK-Z	Sen-S	MK-Z	Sen-S	MK- Z	Sen-S
Francistown	0.11	0.00	1.56	0.48	1.18	0.40	− 0.68	− 0.10
Ghanzi	− 0.06	0.00	0.59	0.29	1.18	0.38	0.73	0.13
Jwaneng	*2.29	2.63	− 1.15	− 0.76	*-2.45	− 3.12	− 0.66	− 0.32
Kasane	− 0.29	− 0.10	*2.23	1.00	*2.11	0.86	− 1.70	− 0.60
Letlhakane	0.03	0.16	− 0.45	− 0.76	− 0.68	− 1.06	0.45	0.55
Mahalapye	− 0.10	0.00	− 0.84	− 0.51	− 1.22	− 0.64	*2.07	0.43
Maun	− 1.31	− 0.38	0.31	0.14	0.90	0.39	0.41	0.10
Pandamatenga	− 0.95	− 1.29	− 0.09	− 0.17	0.09	0.40	*2.07	1.72
Selibe-Phikwe	0.18	1.10	0.55	2.27	1.22	4.06	− 0.93	− 0.90
Shakawe	− 0.44	− 0.15	*−2.01	− 0.44	− 1.28	− 0.43	*2.74	0.50
SSKA	0.15	0.15	− 0.64	− 0.66	− 0.34	− 0.27	1.41	0.43
Sowa Pan	− 0.49	− 0.86	− 1.57	− 1.40	0.08	0.04	− 0.46	− 0.38
Tsabong	− 1.75	− 1.10	− 1.19	− 0.62	1.54	0.33	− 0.18	− 0.03
Tshane	− 0.04	− 0.03	0.46	0.26	1.07	0.21	− 0.40	− 0.05

*Significant trend at 95% confidence level

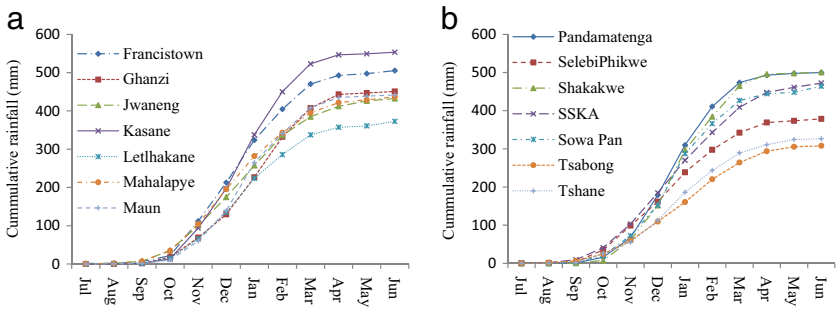

Fig. 3 Rainfall mass curves for a common period between 1998 and 2014

coincides with mid austral summer season in Botswana. It is on this premise that the rainfall season was demarcated to fall between November and March.

4.2 Seasonal rainfall distribution and characteristics

Seasonal rainfall characteristics such as rainfall totals and coefficients of variation (CVs) are presented in Table 1 and their spatial distribution in Fig. 4. From these results, the highest rainfall of 540 mm is observed at Kasane in the northeast and the lowest of 230 mm registered at Tsabong in the southeast. Generally northern and eastern locations exhibit higher seasonal rainfall totals compared to the south and western locations bordering the Kalahari desert. These results also show that Botswana's seasonal rainfall is characterised by high CVs greater than 25%. Locations that recorded higher rainfall also registered lower CVs as the case at Kasane which recorded the lowest CV of 29%. Similarly, locations that registered low rainfall also recorded the highest CVs such as Tsabong with a CV of 50%. Nsubuga et al. (2014) in their study indicated that CVs greater than 20% in rainfall is an indication of unreliability. This high variability is further evidence of erratic nature of rainfall in Botswana which is injurious to rainfed agriculture. Therefore, it is necessary that other rainfall characteristics such as dry spell frequency and onset and cessation of rain dates are determined and further

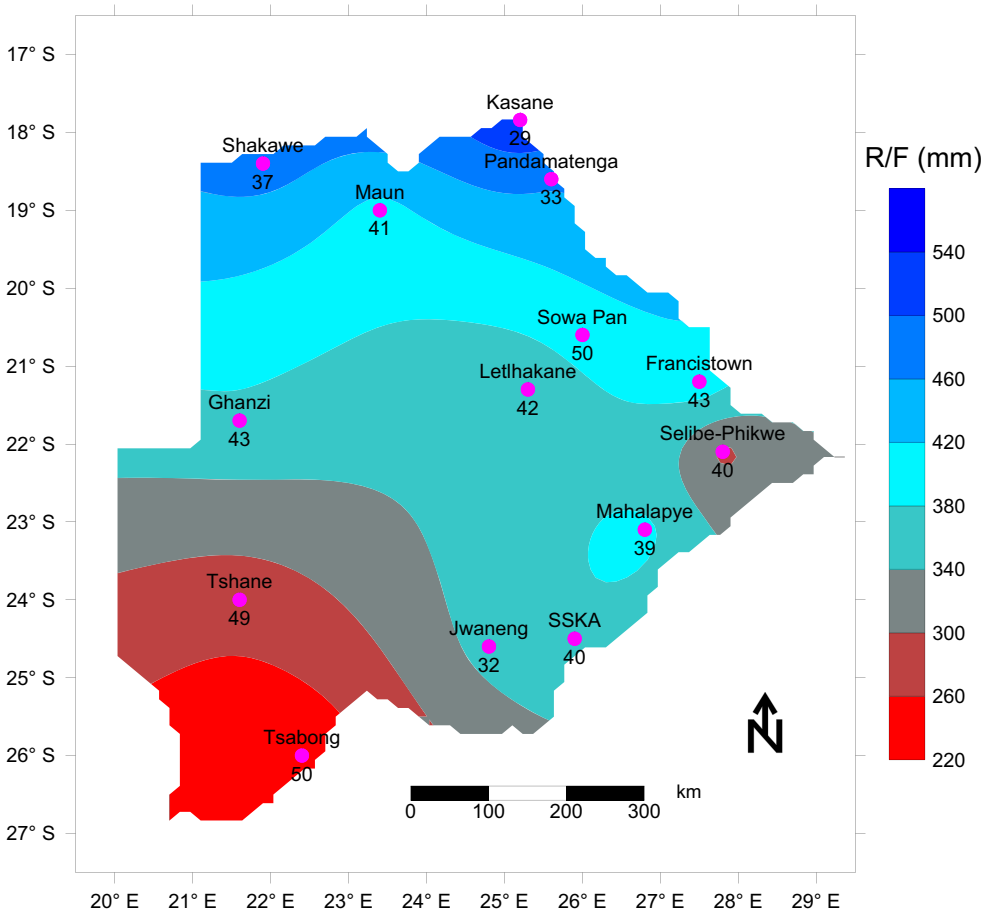

Fig. 4 Seasonal rainfall distribution with coefficients of variation (%) shown at individual stations

investigated. Spatial distribution of rainfall shown in Fig. 4 indicates a southwest to northeast increasing gradient in seasonal rainfall totals. The high rainfall recorded in the northern locations could be as a result of the shifting of the ITCZ whose influence is only felt at latitudes less than 20° S (Usman and Reason 2004). The high CVs registered in this study is in agreement with results from a similar study on climate variability in semiarid locations of Iran by Modarres and da Silva (2007) who reported that the drier the climate, the higher the variability. This high variability will require adaptation measures in order to improve productivity of rainfed agriculture. This could be achieved by transferring information on onset and cessation of rain dates to the various stakeholders in a timely manner.

4.3 Rain onset characteristics

Table 2 presents the mean dates for rain onset from the 14 synoptic stations. The onset dates start in the final week of November through to the second week of January. The earliest onset is recorded at Pandamatenga on 28th of Nov and Francistown on 29th of Nov. Late onset is recorded at Tsabong on 14th of Jan. From scrutiny of the CVs, stations that recorded early onset also show lower CVs of 18% at Francistown and 13% at Pandamatenga. The highest CV of 30% in onset is recorded at Selibe-Phikwe. Stations with late onset generally show high CVs as is the case of Tsabong with CV of 28%. Southern drier locations exhibited higher CVs as compared to northern areas, showing almost same progression as seasonal rainfall totals. At all the stations, onset dates exhibited higher CVs compared to cessation. This could be due to the fact that onset is spread over 3 months of November, December and January while cessation occurs only between February and March.

Onset exhibiting high variability as compared to cessation dates points to increased risks associated with insufficient and unrealistic moisture at the start of the planting season as well as for surface and ground water replenishments. The variability in onset is in agreement with findings of Moeletsi and Walker (2012) who reported a similar scenario in South Africa's Free state which is in the same climatic zone as most locations in the study area. Early onset of rain in November in the northeastern region which are high rainfall locations in Botswana could also be associated with the southward shifting of the ITCZ which approaches the latitudes of 15–20° S during the month of November (Usman and Reason 2004).

Trends in onset presented in Table 3 indicate no major changes except at Jwaneng where onset is significantly getting late at a rate of 2.6 days/season. The rest of the stations recorded marginal changes with 64% of the stations showing earlier tendencies in onset dates though none is significant. Arid locations in the southwest of Botswana could be expanding affecting neighbouring locations such as Jwaneng which is presenting with shortening LRS. Recent studies indicate that arid and semiarid areas are on the increase globally due to effects from large scale climate predicators and local feedbacks that lead to extended dry periods (Dai 2011; Huang et al. 2016). Hence, warming tendencies at the fringes of the Kalahari desert could be part of a global expansion of arid lands.

4.3.1 Frequency analysis of rain onset dates

Rain onset dates plotted on histograms reveal that onset is spread across austral summer with no distinct dates that can be used for informing users and managers of rainfed farming. Six representative stations (Fig. 5) were selected in the west at Ghanzi where analysis of onset reveals that the most likely rain onset is spread from 8th of Oct to18th of Jan a period amounting to 13 weeks. In the central region at Letlhakane, onset is highly likely between 30th of Nov and 1st of Feb, a period of 8 weeks. At Jwaneng in the south, the most probable onset period is spread from 30th of Oct to 27th of Jan totaling 9 weeks of onset date uncertainty. In the north at Maun, the likely period of onset is from 2nd of Nov to 8th of Feb, a period of 12 weeks for which rains are expected. In the northeast at Pandamatenga, farmers expect rains for a period between 18th of Nov and 23rd of Dec totaling 5 weeks. In the eastern region at Selibe-Phikwe, rain onset is from 2nd of Oct to 31st of Dec covering a period of 9 weeks. In the southwest at Tsabong, the onset is spread from 14th of Nov to 9th of Apr, a combined period of 12 weeks of uncertainty for the rainfed farmers.

At all the 14 synoptic stations, the period for which onset is highly likely is shortest in the high rainfall areas in the northeast of 3 weeks at Kasane and highest in the drier locations in the southwest and west of 20 weeks at Tsabong and 13 weeks at Ghanzi, respectively. Since onset dates are spread over a long period of time (up to 20 weeks at some stations), it becomes a challenge for managers in the agricultural sector to make decisions on sowing dates. It is therefore necessary to apply ANN to identify probable onset dates to facilitate decision-making on future sowing and harvest dates. The disparity in rain onset dates across the study area could be associated with the erratic nature of the rainfall typical for semiarid areas. Similar studies in semiarid areas in West Africa by Odekunle (2006) also revealed variations in rain onset dates exceeding 10 weeks which is in agreement with the above findings.

4.4 Rain cessation characteristics

From Table 2, the mean dates for cessation of rain across Botswana start from last week of February to the end of March. Early cessations are recorded on 22nd and 28th of Feb at Tshane and Tsabong, respectively. Pandamatenga and

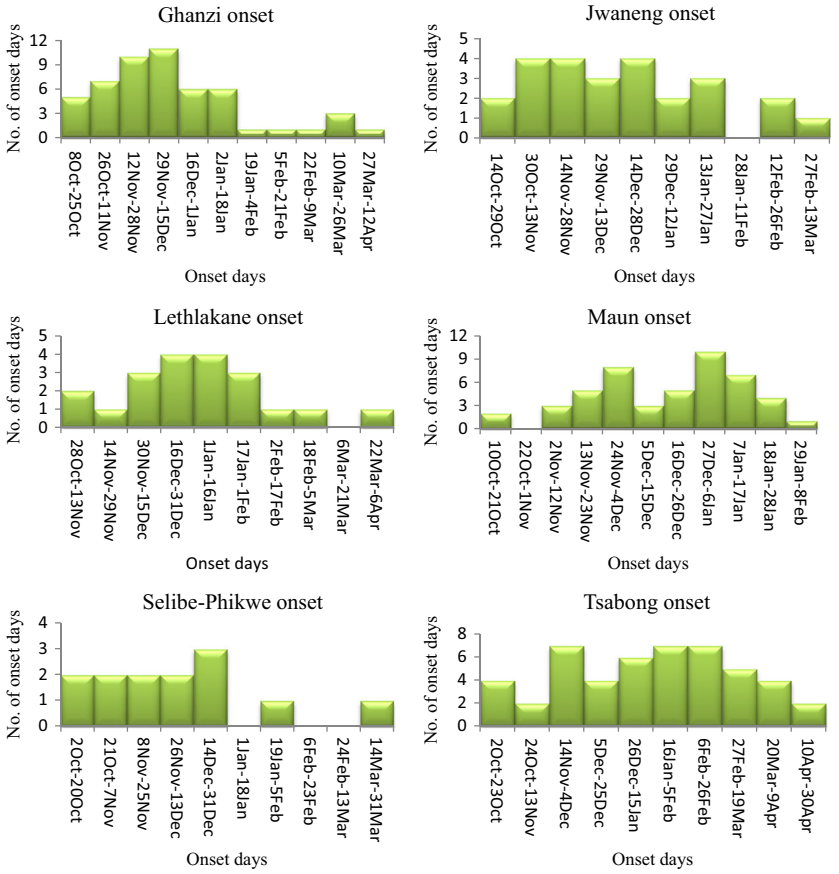

Fig. 5 Rain onset histograms at selected stations during observed period

Shakawe recorded the latest cessation on 27th and 30th March, respectively. Stations with early cessation such as Tsabong and Tshane show higher CVs of 23 and 22%, respectively. Kasane and Shakawe show the lowest CVs in cessation of 8 and 10%, respectively. Generally, CVs in cessation are lower than those of rain onset. Still the drier locations show higher CVs just as the case of rain onset.

Cessation of rain dates covers a mean period between February and March with individual stations showing varying periods across austral summer. Dry locations in the west and southwest showed early cessation as compared to the northeastern areas with high summer rainfall totals. Cessation dates were found to have less variability which could imply that the length of the rain season may be mainly affected by rain onset rather than cessation of rain. These findings are in agreement with results by Recha et al. (2012) who reported higher variability in onset dates compared to cessation dates while investigating seasonal rainfall variability in Kenya. Cessation of rain in March in the northern areas could also be attributed to the retreating ITCZ during the same time period. The southwestern locations of Tsabong and Tshane with early cessation dates are not influenced by the movement of the ITCZ. The southwestern dry areas manifesting with late onset and early cessation of rain could be as a result of the influence of weather patterns in the neighbouring Kalahari Desert.

Trends in cessation dates presented in Table 3 do indicate that rains at Kasane are significantly ceding later at a rate of 1 day/season. The results also indicate that rains are getting shorter at 57% of the stations though the trends are not significant except at Shakawe where cessation dates are significantly ceding earlier at a rate of 0.4 days/season. This trend in the long run is likely to shorten and shift the rain season at these locations.

To determine the degree of association between onset and cessation of rain, a Pearson correlation analysis is undertaken at which seemingly weak negative correlation of $R = -0.42$ results between these rainfall variables. Negative correlation between onset and cessation, though weak, confirms the inverse relationship between early onset and late cessation of rain across the country. This means that delayed onset of rains brings early cessation of rains which could shorten the rain season. Therefore, onset dates may give an indication of the length of rain season rather than cessation.

4.4.1 Frequency analysis of cessation dates

Equally as the case of rain onset, stations were selected from all the regions of the study area as shown in Fig. 6. In the east at Francistown, cessation of rain is highly anticipated between 12th Feb and 22nd Apr, a period of 9 weeks, at Ghanzi in the

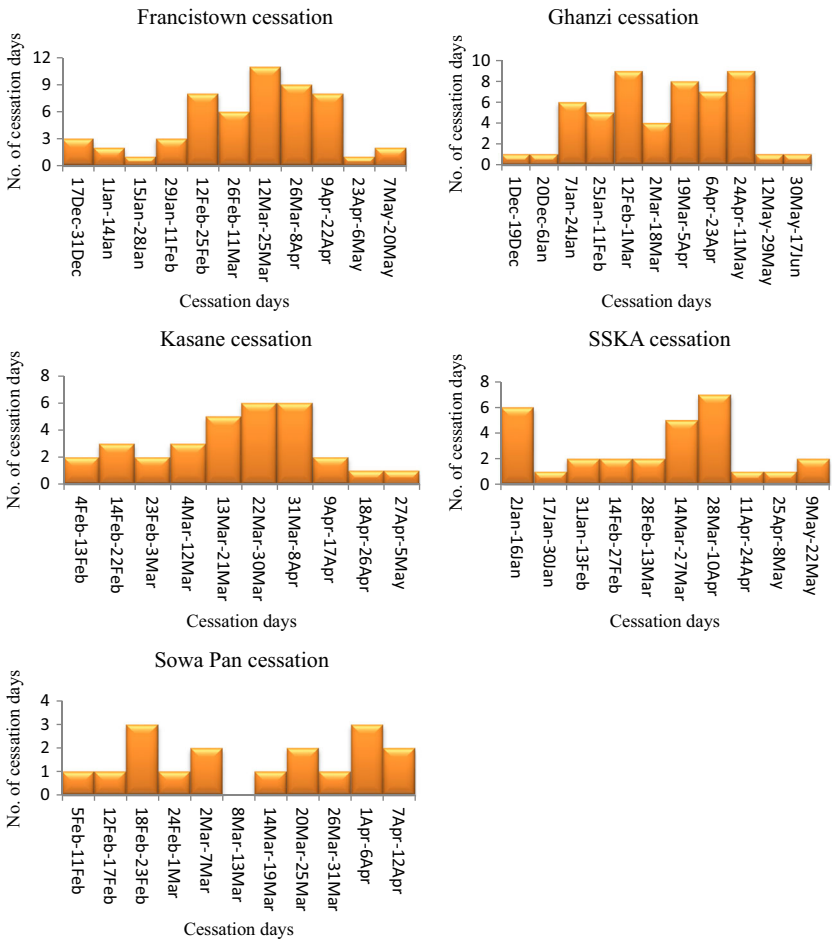

Fig. 6 Cessation of rain histograms at selected stations during observed period

west, it was not clear when the rains cedes. The rains were expected to terminate between 7th Jan and 29th May, a period of almost 20 weeks. This long period affects planning especially for rainfed farming. At Kasane in the northeast, cessation is highly likely between 13th Mar and 8th Apr, a period of 4 weeks. In the south at SSKA, it is not clear which period is highly likely for rain cessation. Two distinct time periods are observed, from 2nd Jan to 16th Jan (2 weeks) and 14th Mar and 10th Apr (4 weeks). It is not possible to identify a particular period for rain cessation at this location. In the central at Sowa Pan, there is equally no well-defined period of cessation. Also, two likely cessation periods emerged from 18th Feb to 7th Mar and 20th Mar to 12th Apr. For similar reasons as onset, ANN is used to predict probable cessation dates at all the 14 synoptic stations.

4.5 Length of the rain season (LRS)

Mean length of the rain season presented in Table 2 indicates that the longest duration of rain is recorded at Pandamatenga in the northeast at 120 days and the least of 45 days at Tsabong southwest of the study area. The eastern to northeastern locations which show earlier onset dates are experiencing longer rains as compared to the southern locations that have relatively late onset dates. The resulting CVs from the analysis indicate that the length of the rain season is highly variable at locations with short rains such as Tsabong and Tshane with CVs of 55 and 58%, respectively. Pandamatenga which has the longest rains shows the lowest CV of 15%.

The southwestern locations show highly variable and short LRS signalling high risks to rainfed farming with Tsabong and Tshane posting 45 and 56 days, respectively. Only five locations were identified with LRS of more than 100 days. These locations are Kasane, Pandamatenga, Sowa Pan, Shakawe and Francistown which are located in the northern and eastern part of the country. Based on the criteria of LRS only, these stations are able to support short to medium season cereal varieties as reported by Moeletsi and Walker (2012). However Sowa Pan due to presence of saline soils may not adequately support arable farming. In Kenya the medium season varieties are recommended for LRS up to 140 days (Mugalavai et al. 2008). The same variety however may not mature at the same time in different locations due to varying heat flux conditions that affect the phonological processes of cereals (Neog et al. 2008; Moeletsi and Walker 2012).

Trends in the length of rain season (LRS) shown in Table 3 indicate that Jwaneng station where rains are getting later is having significant reduction in the LRS at a rate of 3 days/season. Similarly at Kasane which reported later cessation shows a significant increase in the LRS at a rate of 0.86 days/season. The shortening LRS at Jwaneng could be linked to global aridity changes while Kasane is known to be a high rainfall area linked to the movement of the ITCZ.

For purposes of evaluating the effect of onset and cessation on the LRS, Pearson correlation was also used. The results show a strong negative correlation of − 0.9 between onset and the LRS. Similarly, for cessation of rain and LRS, a strong positive correlation of 0.8 resulted. High negative correlations between LRS and rain onset is an indication that early onset mostly results in long rainfall season. This is in agreement with findings from similar studies in Kenya by Mugalavai et al. (2008) and Ngetich et al. (2014) who found correlations in the same order and direction. The strong degree of association implies that onset date can be used to predict the LRS and hence the cessation date. The findings could be helpful in advising farmers on the correct cultivar at the beginning of the growing season which coincides with rain onset dates. The high positive association between LRS and cessation is also an indication that late cessation corresponds to long rain and vice versa.

4.6 Dry spell frequency analysis

Results from dry spell frequency analysis in Table 2 show a high number of dry spells in the southwestern locations of the study area at Tsabong and Tshane at 20 and 19 episodes per season, respectively. Pandamatenga and Kasane in the northeast show the lowest episodes of 14 and 13, respectively. Results also indicate that drier locations have lower CVs; an indication that dry spells are more consistent in these locations.

Dry spell frequency is another rainfall variable that directly relates to plant water stress. It measures reliability of seasonal rainfall. A season with good rains but higher than average number of dry spells will have plants suffering from water stress and hence poor yields (Usman and Reason 2004). The low variability in dry spells is an indication that they are a common occurrence in the study area. A normal season for rainfed farming has been defined as one with utmost average number of dry spells. A season in the austral summer is then required to have not more than 15 dry spells (average number) to be categorised as normal (Usman and Reason 2004). From the LRS criteria, four locations were identified to be suitable for rainfed farming, combining the LRS criteria and DSF; only Kasane, Pandamatenga and Shakawe are found suitable for rainfed agriculture in Botswana.

Trends in dry spells in Table 3 reveal increase in dry spells at 50% of the locations though only significant at Mahalapye, Pandamatenga and Shakawe at a rate of 0.43, 1.72 and 0.50 spells/season, respectively. The remaining stations have dry spell frequency trends decreasing though none of them are significant. The southwestern locations which are already experiencing high DSF are showing decreasing trends though not significant. Similarly, the east to northeastern areas are showing decreasing tendencies in DSF while the central region shows marginal changes in DSF. Mahalapye in the south and Shakawe in the northwest are two locations showing substantial increase in DSF. Significantly increasing trends at Shakawe which has been identified as a potential location for rainfed farming may pose a threat to ongoing efforts of alleviating the food security situation in Botswana.

Trends in dry spell frequency indicating a significant increase at Pandamatenga and Shakawe are rather a shortcoming in attempts to improve productivity of rainfed farming over the study area. On a positive note, the northern high rainfall locations are still registering the lowest number of dry spells; this could also be influenced by the ITCZ movement over these locations. The southwestern locations which are still experiencing the highest number of dry spell could be because the ITCZ has little influence at latitudes greater than 20° S. These locations mainly depend on westerlies for moisture supply (Nicholson et al. 2001; Usman and Reason 2004). High number of dry spells in the drier locations of Botswana is an indication that those areas are not recovering but rather getting worse.

4.7 Prediction of rain onset and cessation dates with ANN

Results from the performance of ANNs and 5-year ahead predictions are presented in Table 4. Performance is presented for both open loop (one step ahead prediction) and closed loop (multistep ahead prediction). At 57% of the stations, the open loop returned lower mean square error (MSE) as compared to closed loop. This is a clear manifestation that the network was able to learn well the predictors before the simulations were made. Correlation coefficients (R) were significantly high for all the stations under study which gives credence to the predictions. The best performing network for onset is recorded at Sowa Pan whereas for cessation is realised at Letlhakane with both locations registering an R of 99%. Figure 7 presents graphical comparison of observed and predicted onset and cessation days at Kasane. It demonstrates how the NARX-ANN model is able to learn the prevailing patterns in order to make reasonable predictions. Predictions of onset reveal that in 57% of the stations, the onset dates are likely to come later than in the historical period by between 2 and 5 weeks.

The predictions further show that at 78% of the stations, cessation dates have a possibility of moving ahead by up to 8 weeks. At Shakawe, the cessation dates have remained the same on 30th of March. Another observation from this study

Table 4 ANN performance and onset and cessation dates predictions

Station	MSE				R		Historical mean date		Predicted 5-year mean date (2015–2019)	
	Open loop network		Closed loop network							
	Onset	Cessn	Onset	Cessn	Onset	Cessn	Onset	Cessn	Onset	Cessn
Francistown	0.123	0.140	0.149	0.117	0.85	0.76	29-Nov	10-Mar	18-Nov	27-Feb
Ghanzi	0.138	0.104	0.141	0.197	0.78	0.78	12-Dec	14-Mar	7-Jan	15-Mar
Jwaneng	0.017	0.091	0.065	0.453	0.96	0.92	15-Dec	20-Mar	18-Dec	30-Mar
Kasane	0.049	0.045	0.019	0.232	0.86	0.89	2-Dec	19-Mar	1-Dec	22-Mar
Letlhakane	0.065	0.080	0.357	0.63	0.80	0.99	1-Jan	16-Mar	7-Dec	26-Mar
Mahalapye	0.236	0.241	0.183	0.197	0.83	0.83	3-Dec	2-Mar	28-Dec	2-Apr
Maun	0.099	0.033	0.169	0.302	0.80	0.82	17-Dec	10-Mar	13-Dec	3-Mar
Pandamatenga	0.009	0.007	0.288	0.77	0.94	0.90	28-Nov	27-Mar	4-Dec	28-Mar
Selibe-Phikwe	0.098	0.132	0.321	0.301	0.93	0.98	5-Dec	11-Mar	5-Jan	28-Mar
Shakawe	0.065	0.068	0.128	0.403	0.77	0.75	10-Dec	30-Mar	29-Dec	30-Mar
SSKA	0.035	0.098	0.229	0.632	0.85	0.88	2-Dec	7-Mar	8-Jan	5-Apr
Sowa Pan	0.031	0.078	0.436	0.775	0.99	0.85	3-Dec	13-Mar	13-Dec	28-Mar
Tsabong	0.160	0.145	0.259	0.321	0.77	0.79	14-Jan	28-Feb	4-Dec	8-Feb
Tshane	0.123	0.094	0.198	0.121	0.77	0.81	28-Dec	22-Feb	27-Dec	7-Mar

is that cessation dates are now spread over 3 months of February, March and April. This could point at higher variability in cessation dates and hence affecting the length of the rain season. Since it is already indicated in Section 4.5 that early onset is closely associated with long rainfall season, the delayed onset of rain could be a signal for shortening and eventual shift of the rain season over the study area in the next 5 years (2015–2019). The capability of the ANN model to make 5 years ahead predictions is evaluated and presented in Table 5. This is achieved by comparing the predicted dates of onset and cessation with historical values for years 2015 and 2016. From the results, the deviations are generally within 5 days at most stations except at Selibe-Phikwe, Sowa Pan, SSKA in 2015 and Jwaneng in 2016. More stations showing higher than 5 days deviation in 2015 could be attributed to the strongest El Niño in 100 years that was experienced over the

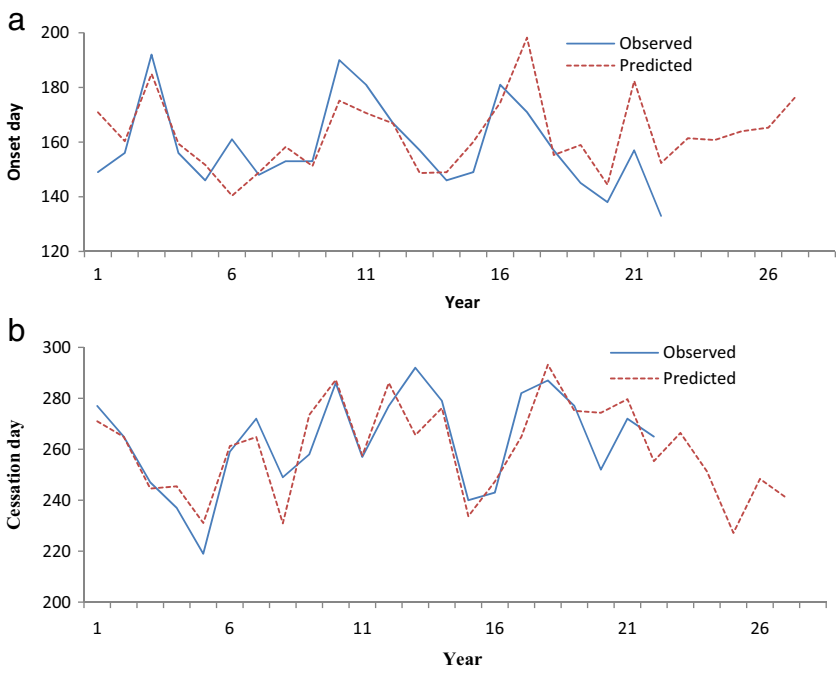

Fig. 7 Nonlinear autoregressive with exogenous input (NARX) model performance evaluation of actual targets and model outputs: **a** Onset and **b** cessation days at Kasane

Table 5 ANN prediction performance evaluation for the years 2015 and 2016

Station	Historical period				Prediction period			
	Onset		Cessation		Onset		Cessation	
	2015	2016	2015	2016	2015	2016	2015	2016
Francistown	12-Nov	23-Feb	15-Feb	28-Mar	15-Nov	20-Feb	10-Feb	26-Mar
Ghanzi	27-Mar	30-Dec	20-Apr	15-Mar	29-Mar	31-Dec	25-Apr	13-Mar
Jwaneng	22-Nov	10-Jan	6-Apr	28-Mar	25-Nov	3-Jan	6-Apr	23-Mar
Kasane	11-Dec	13-Nov	31-Jan	18-Apr	12-Dec	11-Nov	31-Jan	12-Apr
Letlhakane	28-Jan	12-Jan	13-Apr	8-Mar	25-Jan	12-Jan	6-Apr	8-Mar
Mahalapye	26-Dec	12-Dec	7-Apr	25-Jan	31-Dec	11-Dec	9-Apr	22-Jan
Maun	14-Dec	23-Dec	28-Feb	3-Feb	10-Dec	23-Dec	27-Feb	7-Feb
Pandamatenga	11-Dec	26-Jan	9-Apr	21-Mar	9-Dec	28-Jan	4-Apr	27-Mar
Selibe-Phikwe	11-Dec	25-Feb	19-Mar	21-Mar	13-Dec	21-Feb	26-Mar	26-Mar
Shakawe	26-Dec	23-Dec	22-Apr	14-Apr	25-Dec	27-Dec	19-Apr	14-Apr
SSKA	10-Dec	1-Dec	17-Apr	28-Mar	14-Dec	1-Dec	10-Apr	30-Mar
Sowa Pan	10-Dec	25-Dec	14-Apr	28-Mar	2-Dec	29-Dec	16-Apr	29-Mar
Tsabong	2-Nov	16-Feb	3-Dec	26-Feb	1-Nov	13-Feb	8-Dec	3-Mar
Tshane	15-Dec	21-Jan	12-Apr	3-Feb	20-Dec	26-Jan	13-Apr	3-Feb

region during that year. However, even for these deviations, none exceeded 10 days which has been identified as the maximum allowable dry spell for optimal plant growth (Mugalavai et al. 2008; Araya and Stroosnijder 2011). For the locations of Kasane, Pandamatenga and Shakawe which have been identified as being suitable for agricultural production, predictions at these locations are within 5 days deviations from historical record. This further demonstrates the NARX model ability in making 5 years ahead prediction that could support agricultural planning in Botswana.

Realistic prediction of onset and cessation of rain are essential for successful rainfed farming. The NARX artificial neural network (ANN) model has demonstrated its capability to predict onset and cessation dates. Since it has been demonstrated in Sections 4.5 above that delayed onset leads to early cessation and with predictions indicating a late onset, it is projected that rains may get shorter in the near future. The shortening and shifting of the rain season could be part of increased incidences of climate variability that have been projected by the IPCC scenarios to get worse in drier locations of the arid and semiarid areas (Stocker et al. 2013; Pachauri et al. 2014).

5 Conclusions

This study investigated rain onset, cessation and dry spell frequency during austral summer for semiarid Botswana over the study period of 1960 to 2014. Predictions of onset and cessation were made using artificial neural network to provide insights into optimisation of moisture conditions for agricultural and water resources management as part of climate risk mitigation. Based on the results and discussion in Sections 4, the following conclusions and summary have been drawn;

1. The study has revealed that Pandamatenga in the northeast registered the earliest rain onset date on 28th of November while Tsabong located in the drier region of southwest recorded the latest onset date on 14th of January. The patterns of cessation dates were observed to take a reverse spatial direction to those of onset dates with Tshane in the southwest recording the earliest cessation on 22nd of February. Shakawe in the northwest registered the latest cessation date on 30th of March. Investigation of the association between onset and cessation of rain revealed a negative correlation. This study also found out that onset dates were less stable characterised by higher coefficient of variations compared to cessation dates.
2. Investigations of the length of the rain season (LRS) revealed that the longest season was 120 days recorded at Pandamatenga and the shortest of 45 days at Tsabong. The other two stations with LRS greater than 100 days were Kasane and Shakawe. This implies that only short to medium varieties of cereals can be grown at these locations. The LRS was found to be highly influenced by the date of rain onset with a degree of association of -0.9 which links early onset with long rainfall season.
3. Using below average dry spell frequency (DSF) as a measure of a normal agricultural season, Kasane, Pandamatenga and Shakawe were identified as low risk areas for rainfed farming. However, results

revealed significantly increasing trends in DSF at Pandamatenga and Shakawe which may rather affect rainfed farming in the long run.

4. In an effort to reduce uncertainties in rainfed agriculture, predictions of onset and cessation dates of rain were made. These predictions revealed that the onset of rain is projected to be getting delayed by 2–5 weeks in 57% of the stations over the next 5 years (2015–2019). These predictions also indicate the possibility of shortening and shifting of the rainfall season in the near future.

The above findings from this study can facilitate better planning for rainfed agriculture in Botswana while minimising related climate risks.

Acknowledgments The authors appreciate support provided by the Mobility for Engineering Graduates in Africa (METEGA) and Carnegie Cooperation of New York through RUFORUM in form of research funds. We also acknowledge the Department of Meteorological Services (DMS) of Botswana for their valuable meteorological data. Special gratitude to Gulu University which granted the first author study leave to enable him focus on this study. The authors are also grateful to the anonymous reviewers for their valuable comments on the manuscript.

References

Akinsanola AA, Ogunjobi KO (2015) Recent homogeneity analysis and long-term spatio-temporal rainfall trends in Nigeria. Theor Appl Climatol 128(1-2):275–289. https://doi.org/10.1007/s00704-015-1701-x

Alexandersson H (1986) A homogeneity test applied to precipitation data. J Climatol 6(6):661–675. https://doi.org/10.1002/joc.3370060607

Allen RG, Pereira LS, Raes D, Smith M (1998) FAO irrigation and drainage paper no. 56. Rome Food Agric Organ United Nations 56:97–156

Alvares CA, Stape JL, Sentelhas PC, de Moraes Gonçalves JL, Sparovek G (2013) Köppen's climate classification map for Brazil. Meteorol Zeitschrift 22(6):711–728. https://doi.org/10.1127/0941-2948/2013/0507

Amekudzi L, Yamba E, Preko K, Asare E, Aryee J, Baidu M, Codjoe S (2015) Variabilities in rainfall onset, cessation and length of rainy season for the various agro-ecological zones of Ghana. Climate 3(2):416–434. https://doi.org/10.3390/cli3020416

Anari PL, Darani HS, Nafarzadegan AR (2011) Application of ANN and ANFIS models for estimating total infiltration rate in an arid rangeland ecosystem. Res J Environ Sci 5:236

Araya A, Stroosnijder L (2011) Assessing drought risk and irrigation need in northern Ethiopia. Agric For Meteorol 151(4):425–436. https://doi.org/10.1016/j.agrformet.2010.11.014

Araya A, Keesstra SD, Stroosnijder L (2010) A new agro-climatic classification for crop suitability zoning in northern semi-arid Ethiopia. Agric For Meteorol 150(7-8):1057–1064. https://doi.org/10.1016/j.agrformet.2010.04.003

Batisani N, Yarnal B (2010) Rainfall variability and trends in semi-arid Botswana: implications for climate change adaptation policy. Appl Geogr 30(4):483–489. https://doi.org/10.1016/j.apgeog.2009.10.007

Beguería S, Vicente-Serrano SM, Reig F, Latorre B (2014) Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. Int J Climatol 34(10):3001–3023. https://doi.org/10.1002/joc.3887

Buishand TA (1982) Some methods for testing the homogeneity of rainfall records. J Hydrol 58(1-2):11–27. https://doi.org/10.1016/0022-1694(82)90066-X

Byakatonda J, Parida BP, Kenabatho PK, Moalafhi DB (2016) Modeling dryness severity using artificial neural network at the Okavango Delta, Botswana. Glob Nest J 18:463–481

Byakatonda J, Parida BP, Kenabatho PK, Moalafhi DB (2018) Influence of climate variability and length of rainy season on crop yields in semiarid Botswana. Agric For Meteorol 248:130–144. https://doi.org/10.1016/j.agrformet.2017.09.016

Chang F-J, Chen P-A, Y-R L et al (2014) Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control. J Hydrol 517:836–846. https://doi.org/10.1016/j.jhydrol.2014.06.013

Chang F-J, Tsai Y-H, Chen P-A, Coynel A, Vachaud G (2015) Modeling water quality in an urban river using hydrological factors—data driven approaches. J Environ Manag 151:87–96. https://doi.org/10.1016/j.jenvman.2014.12.014

Chen X, JL W, Wang L (2005) Prediction of climate change impacts on streamflow of lake bosten using artificial neural network model. J Lake Sci 3:4

Chen C-S, Chen BP-T, Chou FN-F, Yang C-C (2010) Development and application of a decision group back-propagation neural network for flood forecasting. J Hydrol 385(1-4):173–182. https://doi.org/10.1016/j.jhydrol.2010.02.019

Costa AC, Soares A (2009) Trends in extreme precipitation indices derived from a daily rainfall database for the south of Portugal. Int J Climatol 29(13):1956–1975. https://doi.org/10.1002/joc.1834

Dai A (2011) Drought under global warming: a review. Wiley Interdiscip Rev Clim Chang 2(1):45–65. https://doi.org/10.1002/wcc.81

Demuth H, Beale M, Hagan M (2009) Neural network toolbox (version 4)

Diaconescu E (2008) The use of NARX neural networks to predict chaotic time series. Wseas Trans Comput Res 3:182–191

Gao Y, Meng Joo E (2005) NARMAX time series model prediction: feedforward and recurrent fuzzy neural network approaches. Fuzzy Sets Syst 150(2):331–350. https://doi.org/10.1016/j.fss.2004.09.015

GOB-MEWT (2012) Second National Communication to the United Nations Framework Convention on Climate Change (UNFCCC). Gaborone

GOB-MMEWR (2006) National water mater plan review, vol 3. Gaborone

Gocic M, Trajkovic S (2013) Analysis of changes in meteorological variables using Mann-Kendall and Sen's slope estimator statistical tests in Serbia. Glob Planet Change 100:172–182. https://doi.org/10.1016/j.gloplacha.2012.10.014

Hecht-Nielsen R (1987) Kolmogorov's mapping neural network existence theorem. Proceedings of the international conference on Neural Networks:11–13

Huang W, Wang S, Yu L et al (2006) A new computational method of input selection for stock market forecasting with neural networks. Comput Sci 2006:308–315

Huang J, Ji M, Xie Y, Wang S, He Y, Ran J (2016) Global semi-arid climate change over last 60 years. Clim Dyn 46(3-4):1131–1150. https://doi.org/10.1007/s00382-015-2636-8

Kampata JM, Parida BP, Moalafhi DB (2008) Trend analysis of rainfall in the headstreams of the Zambezi River Basin in Zambia. Phys Chem Earth, Parts A/B/C 33(8-13):621–625. https://doi.org/10.1016/j.pce.2008.06.012

Kebede A, Diekkrüger B, Edossa DC (2016) Dry spell, onset and cessation of the wet season rainfall in the Upper Baro-Akobo Basin, Ethiopia. Theor Appl Climatol 129(3-4):1–10. https://doi.org/10.1007/s00704-016-1813-y

Lahmiri S (2016) On simulation performance of feedforward and NARX networks under different numerical training algorithms. In: Handbook of research on computational simulation and modeling in engineering. IGI global, pp 171–183. https://doi.org/10.4018/978-1-4666-8823-0.ch005

Liu Z, Peng C, Xiang W et al (2010) Application of artificial neural networks in global climate change and ecological research: an overview. Chinese Sci Bull 55(34):3853–3863. https://doi.org/10.1007/s11434-010-4183-3

Machiwal D, Jha MK (2012) Hydrologic time series analysis: theory and practice. Springer Science & Business Media. https://doi.org/10.1007/978-94-007-1861-6

Maier HR, Dandy GC (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ Model Softw 15(1):101–124. https://doi.org/10.1016/S1364-8152(99)00007-9

Maier HR, Dandy GC (2001) Neural network based modelling of environmental variables: a systematic approach. Math Comput Model 33(6-7):669–682. https://doi.org/10.1016/S0895-7177(00)00271-5

Masinde M (2014) Artificial neural networks models for predicting effective drought index: factoring effects of rainfall variability. Mitig Adapt Strateg Glob Chang 19(8):1139–1162. https://doi.org/10.1007/s11027-013-9464-0

Menezes JMP, Barreto GA (2008) Long-term time series prediction with the NARX network: an empirical evaluation. Neurocomputing 71(16-18):3335–3343. https://doi.org/10.1016/j.neucom.2008.01.030

Mishra AK, Desai VR (2006) Drought forecasting using feed-forward recursive neural network. Ecol Model 198(1-2):127–138. https://doi.org/10.1016/j.ecolmodel.2006.04.017

Modarres R, da Silva VPR (2007) Rainfall trends in arid and semi-arid regions of Iran. J Arid Environ 70(2):344–355. https://doi.org/10.1016/j.jaridenv.2006.12.024

Moeletsi M, Walker S (2012) Rainy season characteristics of the Free State Province of South Africa with reference to rain-fed maize production. Water SA 38(5):775–782. https://doi.org/10.4314/wsa.v38i5.17

Morid S, Smakhtin V, Bagherzadeh K (2007) Drought forecasting using artificial neural networks and time series of drought indices. Int J Climatol 27(15):2103–2111. https://doi.org/10.1002/joc.1498

Mugalavai EM, Kipkorir EC, Raes D, Rao MS (2008) Analysis of rainfall onset, cessation and length of growing season for western Kenya. Agric For Meteorol 148(6-7):1123–1135. https://doi.org/10.1016/j.agrformet.2008.02.013

Nasr M, Zahran HF (2014) Using of pH as a tool to predict salinity of groundwater for irrigation purpose using artificial neural network. Egypt J Aquat Res 40(2):111–115. https://doi.org/10.1016/j.ejar.2014.06.005

Neog P, Bhuyan J, Baruah N (2008) Thermal indices in relation to crop phenology and seed yield of soybean (Glycine max L. Merrill). J Agrometeorol 10:388–392

Ngetich KF, Mucheru-Muna M, Mugwe JN, Shisanya CA, Diels J, Mugendi DN (2014) Length of growing season, rainfall temporal distribution, onset and cessation dates in the Kenyan highlands. Agric For Meteorol 188:24–32. https://doi.org/10.1016/j.agrformet.2013.12.011

Nicholson SE, Leposo D, Grist J (2001) The relationship between El Niño and drought over Botswana. J Clim 14(3):323–335

Nsubuga FWN, Botai OJ, Olwoch JM, Rautenbach CJW, Bevis Y, Adetunji AO (2014) The nature of rainfall in the main drainage sub-basins of Uganda. Hydrol Sci J 59(2):278–299. https://doi.org/10.1080/02626667.2013.804188

Odekunle TO (2006) Determining rainy season onset and retreat over Nigeria from precipitation amount and number of rainy days. Theor Appl Climatol 83(1-4):193–201. https://doi.org/10.1007/s00704-005-0166-8

Pachauri RK, Allen MR, Barros VR, et al (2014) Climate change 2014: synthesis report. Contribution of working groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change. IPCC

Pettit AN (1979) Anon-parametric approach to the change-point detection. Appl Stat 28(2):126–135. https://doi.org/10.2307/2346729

Recha CW, Makokha GL, Traore PS, Shisanya C, Lodoun T, Sako A (2012) Determination of seasonal rainfall variability, onset and cessation in semi-arid Tharaka district, Kenya. Theor Appl Climatol 108(3-4):479–494. https://doi.org/10.1007/s00704-011-0544-3

Sabzevari AA, Zarenistanak M, Tabari H, Moghimi S (2015) Evaluation of precipitation and river discharge variations over southwestern Iran during recent decades. J Earth Syst Sci 124:335–352

Sen PK (1968) Estimates of the regression coefficient based on Kendall's tau. J Am Stat Assoc 63(324):1379–1389. https://doi.org/10.1080/01621459.1968.10480934

Sivakumar MVK (1988) Predicting rainy season potential from the onset of rains in Southern Sahelian and Sudanian climatic zones of West Africa. Agric For Meteorol 42(4):295–305. https://doi.org/10.1016/0168-1923(88)90039-1

Sivakumar MVK (1992) Empirical analysis of dry spells for agricultural applications in West Africa. J Clim 5(5):532–539

Some'e BS, Ezani A, Tabari H (2013) Spatiotemporal trends of aridity index in arid and semi-arid regions of Iran. Theor Appl Climatol 111(1-2):149–160. https://doi.org/10.1007/s00704-012-0650-x

Sorjamaa A, Hao J, Reyhani N, Ji Y, Lendasse A (2007) Methodology for long-term prediction of time series. Neurocomputing 70(16-18):2861–2869. https://doi.org/10.1016/j.neucom.2006.06.015

Stagge JH, Tallaksen LM, Xu CY, Van Lanen HAJ (2014) Standardized precipitation-evapotranspiration index (SPEI): sensitivity to potential evapotranspiration model and parameters. Proc FRIEND-water 367–373

Stathakis D (2009) How many hidden layers and nodes? Int J Remote Sens 30(8):2133–2147. https://doi.org/10.1080/01431160802549278

Statistics Botswana (2009) Botswana water statistics, vol 3. Gaborone

Statistics Botswana (2015) Annual Agricultural Survey Report 2013. Gaborone

Stocker TF, Qin D, Plattner GK, et al (2013) Climate change 2013: the physical science basis. Intergovernmental panel on climate change, working group I contribution to the IPCC fifth assessment report (AR5)

Tabari H, Somee BS, Zadeh MR (2011) Testing for long-term trends in climatic variables in Iran. Atmos Res 100(1):132–140. https://doi.org/10.1016/j.atmosres.2011.01.005

Usman MT, Reason CJC (2004) Dry spell frequencies and their variability over southern Africa. Clim Res 26:199–211. https://doi.org/10.3354/cr026199

Von Neumann J (1941) Distribution of the ratio of the mean square successive difference to the variance. Ann Math Stat 12(4):367–395. https://doi.org/10.1214/aoms/1177731677

Wijngaard JB, Klein Tank AMG, Können GP (2003) Homogeneity of 20th century European daily temperature and precipitation series. Int J Climatol 23(6):679–692. https://doi.org/10.1002/joc.906

Yue S, Wang C (2004) The Mann-Kendall test modified by effective sample size to detect trend in serially correlated hydrological series. Water Resour Manag 18(3):201–218. https://doi.org/10.1023/B:WARM.0000043140.61082.60