Hydrology and
Earth System
Sciences

Seasonal forecasting of hydrological drought in the Limpopo Basin: a comparison of statistical methods

Mathias Seibert[1], Bruno Merz[1,2], and Heiko Apel[1]

[1]GFZ – German Centre for Geosciences, Section 5.4: Hydrology, Potsdam, Germany
[2]University of Potsdam, Institute of Earth and Environmental Science, Potsdam, Germany

Correspondence to: Mathias Seibert (mathias.seibert@gmail.com)

Abstract. The Limpopo Basin in southern Africa is prone to droughts which affect the livelihood of millions of people in South Africa, Botswana, Zimbabwe and Mozambique. Seasonal drought early warning is thus vital for the whole region. In this study, the predictability of hydrological droughts during the main runoff period from December to May is assessed using statistical approaches. Three methods (multiple linear models, artificial neural networks, random forest regression trees) are compared in terms of their ability to forecast streamflow with up to 12 months of lead time. The following four main findings result from the study.

1. There are stations in the basin at which standardised streamflow is predictable with lead times up to 12 months. The results show high inter-station differences of forecast skill but reach a coefficient of determination as high as 0.73 (cross validated).

2. A large range of potential predictors is considered in this study, comprising well-established climate indices, customised teleconnection indices derived from sea surface temperatures and antecedent streamflow as a proxy of catchment conditions. El Niño and customised indices, representing sea surface temperature in the Atlantic and Indian oceans, prove to be important teleconnection predictors for the region. Antecedent streamflow is a strong predictor in small catchments (with median 42 % explained variance), whereas teleconnections exert a stronger influence in large catchments.

3. Multiple linear models show the best forecast skill in this study and the greatest robustness compared to artificial neural networks and random forest regression trees,

despite their capabilities to represent nonlinear relationships.

4. Employed in early warning, the models can be used to forecast a specific drought level. Even if the coefficient of determination is low, the forecast models have a skill better than a climatological forecast, which is shown by analysis of receiver operating characteristics (ROCs). Seasonal statistical forecasts in the Limpopo show promising results, and thus it is recommended to employ them as complementary to existing forecasts in order to strengthen preparedness for droughts.

1 Introduction

Drought is a slowly progressing phenomenon which is challenging to detect ahead. As a result, drought management frequently remains crisis management, which is limited to fighting drought when impacts have already started to unfold. A more desirable reaction is the conversion of crisis management to risk management (Wilhite and Hayes, 2000). This is a challenging process in which drought forecasting with a long lead time is recommended in order to adopt mitigation actions and raise preparedness (Vicente-Serrano et al., 2012a). Forecasting products should be tailored to the end users' needs, such as water resources managers (Winsemius et al., 2014; Masih et al., 2014). The forecast information must satisfy the need for a thorough drought assessment without overwhelming end users with high complexity. Seasonal forecasts have a high uncertainty; therefore, it is an important task to convey the skill of the forecast system,

and for end users to include uncertainty information in the decision-making process. This can be achieved, for example, by providing probabilistic drought forecast information.

The Limpopo Basin in southern Africa (ca. $408\,000\,km^2$) is strongly affected by droughts, which had severe impacts on agriculture, economy and food security in southern Africa, for example, in the early 1980s and 1990s (Love et al., 2010; Rouault and Richard, 2003; Rouault, 2005; Masih et al., 2014; FAO, 2004). The Limpopo Basin is a highly modified catchment, where irrigation demands by agriculture are high and may even exceed supply in parts of the basin (FAO, 1997). Southern African water resources are regarded as being highly affected by seasonal variability, a fact that is likely to be exacerbated by climate change (Kusangaya et al., 2014). Zhu and Ringler (2010) estimated a decrease in Limpopo streamflow by 2030 due to climate change, which is contradictory to studies that found increases of precipitation (Tadross et al., 2005) and runoff (Li et al., 2015) in parts of southern Africa, including vast parts of the Limpopo Basin.

Hence, it seems that the climate change impact in the Limpopo Basin remains very uncertain and might exhibit a stronger effect on precipitation variability than on average precipitation (Tadross et al., 2005). The high inter-annual variability of precipitation and the tense condition of water resources require improvements in water management in the riparian states (South Africa, Botswana, Zimbabwe and Mozambique) who cooperate within the Limpopo Watercourse Commission since 2003. This study analyses the annual to seasonal predictability of (seasonal) hydrological drought in the Limpopo Basin using statistical methods which could improve the preparedness and help mitigate drought disasters.

In order to understand hydrological droughts and to cope with them properly, an appropriate drought indicator has to be selected and forecasted (Wetterhall et al., 2015; Winsemius et al., 2014). In the Limpopo Basin, the main rainy season runoff lasts from December to May, and the total streamflow of the period is an adequate indicator for hydrological droughts. In this study, the standardised streamflow index (SSI) is used as hydrological drought index (Vicente-Serrano, 2012b). Standardisation of streamflow is less common than for precipitation (Mishra and Singh, 2010), but nevertheless useful for two reasons. First, it facilitates the comparison of droughts at different stations. Second, the standardised indicator is normally distributed and has, therefore, a higher sensitivity to droughts compared to original streamflow which is often strongly positively skewed.

Statistical streamflow forecasting is challenging due to the complexity of the signal and the underlying processes, especially in highly modified catchments such as the Limpopo Basin (FAO, 2004). The streamflow signal integrates meteorological, hydrological and anthropogenic effects, such as irrigation and water storage, thus interlacing hydrological drought and water scarcity (Van Loon and Van Lanen, 2013).

Anthropogenic effects (e.g. operation of dams and irrigation) are typically time-varying and can be considered in hydrological models (Trambauer et al., 2014). Thereby, it is possible to separate drought from water scarcity (Van Loon and Van Lanen, 2013) by simulating naturalised streamflow. In a statistical approach as presented here, the anthropogenic effect is not accounted for and therefore increases prediction uncertainties.

Atmospheric circulation processes have a chaotic component that is not susceptible to prediction, but predictability can be deduced from the land–atmosphere and land–ocean interactions. The latter can be represented by teleconnections to sea surface temperatures (SSTs), which is a common approach in both tropical and humid climates. In more dry climate zones, the land–atmosphere interaction, and therefore the land surface moisture condition, is likely to be more important (Koster et al., 2000), since atmospheric moisture is recycled over the land surface (Gimeno et al., 2010). It can be expected that both SSTs and land surface conditions are important factors in the Limpopo Basin, because it extends from the ocean to very arid regions in Botswana.

The term "teleconnection" refers to the influence of sometimes remote ocean regions on atmospheric variables, such as moisture content or precipitation. Past studies on southern African precipitation found predictability based on El Niño, the Indian Ocean and the Atlantic Ocean (Reason et al., 2006; Landman et al., 2005; Landman and Mason, 1999). However, the atmospheric circulation is very complex, sometimes having the effect that even strong El Niño events do not propagate to the region (Thomson et al., 2003). A reason might be that the ocean region south of Africa is the major source for precipitation in southern Africa (Gimeno et al., 2010). This region is characterised by the chaotic collision system of the warm Agulhas and the cold Antarctic circumpolar ocean current (see Fig. 1) (Peterson and Stramma, 1991). In the collision process, warm Agulhas eddies can form, maintaining higher evaporation until they dissipate. There are more complex effects such as the Darwin sea level pressure (Manatsa et al., 2007), the linkage of the El Niño–Southern Oscillation (ENSO) with the Indian Ocean dipole (Yuan and Li, 2008) or the stratospheric quasi-biennial oscillation (Jury, 1996) and even the Antarctic ozone depletion (Manatsa et al., 2013). Despite that complexity, SST teleconnections remain the preferred choice of predictors in seasonal forecasting (Landman et al., 2005; Landman and Mason, 1999; Funk et al., 2014). In this study, widely used climate indices are complemented with customised indices resulting from a composite and correlation analysis of SSTs in the Indian and Atlantic oceans.

Many methods have been applied in drought forecasting (Mishra and Singh, 2011). Three models are chosen for comparison in this study. The first are multiple linear models (MLMs), which are widely used in similar studies (Diro et al., 2011, e.g.). They are, however, limited to linear combinations of predictors. Artificial neural networks (ANNs) are applied as a second method. They are flexible nonlinear

models and have been applied successfully in several seasonal prediction studies (Mwale et al., 2004; Morid et al., 2007; Mishra and Desai, 2006). In addition, we develop random forest regression (RFOR) tree models (Breiman, 2001). These are particularly suited for representing conditional relationships in complex data including nonlinearities. Random forest regression trees have only rarely been applied for seasonal drought forecasting (Chen et al., 2012). These data-driven approaches are useful for seasonal forecasting in regions where hydrological observations are available, but additional data characterising the catchments are limited.

A recent publication by Trambauer et al. (2015) presented forecasting results for the Limpopo Basin achieved by a chain of process-based models, namely the hydrological model PCR-GLOBWB (Van Beek and Bierkens, 2009) with input from the seasonal forecasting system S4 (Molteni et al., 2011) and reanalysis data ERA-Interim (Dee et al., 2011) by ECMWF. The skill of the forecasting system for total streamflow between December and May (DJFMAM) exceeded climatological forecasts (climatology) with "moderate skill for all lead times" up to 5 months (forecast in December) (Trambauer et al., 2015). To parameterise such models is challenging in data-sparse regions such as southern Africa (Trambauer et al., 2014). Compared to forecasts based on simulation models, statistical forecast models require less input data and computational power. The main requirement for model development is a sufficiently long record of relevant drought indicators. In summary, both approaches have their advantages and disadvantages. Here, we evaluate the predictability of hydrological drought in a data-driven approach, which can serve as a baseline for other seasonal forecast systems. Special care is taken of the predictor selection, model validation and forecast verification process for the use of the forecast models in a drought early warning system. We present the forecasting skill for hydrological drought during the main rainy season runoff from December to May achieved with the three selected statistical models.

2 Data and methods

2.1 Study area: Limpopo Basin

The Limpopo Basin is located in southern Africa with the riparian states South Africa, Botswana, Zimbabwe and Mozambique, where it flows into the Indian Ocean and covers an area of approximately $400\,000\,\text{km}^2$ (Fig. 1). The area is dominated by hot steppe climate, while the southernmost regions reach into the warm temperate climate zone of South Africa and the eastern region comprises parts of the savanna climate in Mozambique. The highest mountains of the Waterberg mountain range in South Africa reach ca. 2300 m in elevation. The rainy season usually lasts from October to March. The average annual rainfall ranges from ca. 250 to 1050 mm with 530 mm on average, but with high inter-annual varia-
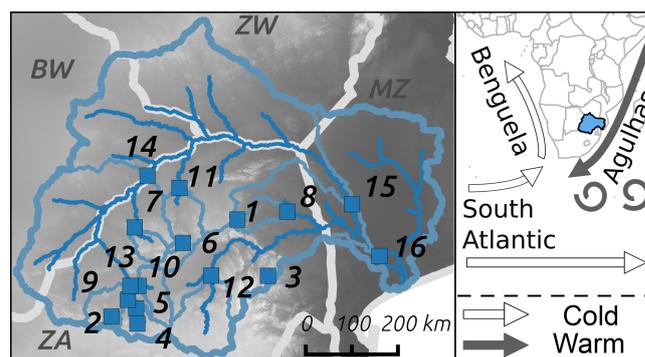
Figure 1. Location of the Limpopo Basin and streamflow stations. Streamflow stations with sub-basins are numbered according to Table 1 and elevation (max. 2300 m) as background (left). Location of the Limpopo Basin within sub-Saharan Africa and schematic of ocean currents with the warm Agulhas Current and the Benguela, South Atlantic and Antarctic Circumpolar (South Atlantic) currents shown on the right.
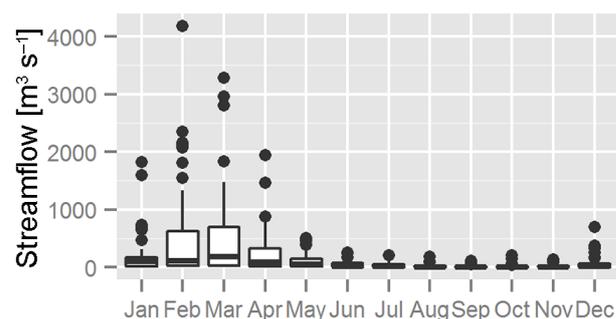
Figure 2. Box plot of monthly streamflow at station Chokwe.

tion which makes drought a common natural hazard. Mean annual runoff is approximately 4550 million $\text{m}^3\,\text{a}^{-1}$ (station Chokwe) and the main rainy season runoff lasts from December to May (Fig. 2). Rainfed farming and grazing is very common, but commercial irrigation farming is also widespread, so that irrigation is the most important water usage with about 50 % of total water use (FAO, 2004). Intra-basin and inter-basin water transfers exist in South Africa (inter-basin transfers from Incomati, Usutu and Orange rivers) and Botswana (intra-basin transfers). Water use and storage heavily affect streamflow with the effect that, for example, in the Matlabas subcatchment only approximately 5 % of the naturalised mean annual runoff is recorded (FAO, 2004, Table 8). The total dam capacity is ca. 2500 million m^3, and the dam capacity per subcatchment often exceeds mean annual streamflow (Table 1). Hence, many of the streamflow time series are heavily affected by water use and management.

Table 1. Streamflow stations included in the analysis: observation period, average annual streamflow volume (Q_{ann}) (in $M m^3 a^{-1}$), dam capacity (V_{Dam}) (in $M m^3$), dam capacity relative to mean annual flow volume (V_{Drel}), catchment area (area) (in km^2). Dam capacities are estimations based on available information.

	Station	Time period	Q_{ann}	V_{Dam}	V_{Drel}	Area
1	Woodbush	June 1977–February 2012 (34.2 years)	8.53	1.94	0.23	12
2	Rietvallei	January 1971–March 2012 (39.6 years)	2.90	0.00	0.00	15
3	Naauwpoort	March 1957–October 2010 (47.1 years)	10.57	14.22	1.35	87
4	Hartbeeshoek	October 1964–March 2012 (44.8 years)	4.52	0.00	0.00	101
5	Krokodilrivier	March 1972–March 2012 (40.1 years)	185.46	0.00	0.00	215
6	Doorndraai	September 1954–February 2012 (57.4 years)	8.24	44.20	5.36	409
7	Mokolo	August 1980–February 2012 (31.4 years)	119.50	145.92	1.22	4315
8	Letaba Ranch	October 1959–March 2012 (45.4 years)	100.77	235.60	2.34	4724
9	Beestekraal	March 1951–March 2012 (59.9 years)	153.67	268.79	1.75	6032
10	Klipvoor	April 1970–March 2012 (42 years)	118.70	134.16	1.13	6159
11	Glen Alpine	May 1970–February 2012 (41.8 years)	101.40	67.47	0.67	11 246
12	Loskop Noord	September 1938–May 2011 (44.7 years)	230.14	960.68	4.17	16 542
13	Buffelspoort	September 1955–March 2012 (56.5 years)	131.27	487.45	3.71	20 383
14	Botswana	April 1971–February 2012 (36.8 years)	475.09	945.19	1.99	100 977
15	Combomume	March 1966–August 2011 (41.1 years)	3084.29	1311.16	0.43	259 214
16	Chokwe	July 1951–May 2011 (56.8 years)	4552.25	3252.13	0.71	343 225

2.2 Data

The database of this study comprises streamflow data, which serve as both predictand and predictor, climate indices and gridded sea surface temperature anomalies (potential predictors). The Global Runoff Data Centre (GRDC, 2011) provides streamflow from all countries in the Limpopo Basin. These data are extended by the runoff observations available from the Department of Water Affairs of the Republic of South Africa (DWAF) and Mozambique Regional Administration of Waters in the South (ARA-Sul). A subset of 16 stations (Fig. 1 and Table 1) satisfies the following conditions:

- a record length of at least 30 years (360 observations at monthly resolution);

- completeness of at least 90 % in the observation period.

The stations selected for this study are mainly located in the South African part of the basin where data availability is better and the conditions are met (Fig. 1). The HydroSHEDS data set (Lehner et al., 2008) is used to derive catchment outlines per station and catchment areas. Dam capacities are collated from the DWAF database (http://www.dwaf.gov.za/Hydrology/).

Several prominent atmospheric indices are acquired from the online resources provided by the Climate Prediction Center of the National Oceanic and Atmospheric Administration (NOAA). To represent the influence of the ENSO, the SST indices of regions Niño-1 and -2, Niño-3, Niño-4, Niño-3.4 are compiled in addition to the Trans-Niño and the Oceanic Niño index (see Table 2 for a detailed list). These indices form the basis of the potential predictors. This set of

widely used indices is augmented with customised predictors based on analysing the sea surface temperature data set HadISST~1.1 (Rayner et al., 2003) provided by the British Met office.

2.3 Hydrological drought predictand: standardised streamflow index

In streamflow standardisation, a time series is transformed to a normally distributed time series, which can be applied at different temporal scales. The discharge is averaged for the chosen period (for example, January–February) for every year, and this series of annual averages is then standardised. The use of standardised streamflow translates into a drought metric independent of catchment size, climatology and streamflow characteristics (Lorenzo-Lacruz et al., 2013). Furthermore, the strength of event anomalies can easily be compared between very different catchments. For the purpose of seasonal drought forecasting, the inter-annual variability of low flow is more important than in a general streamflow forecast. The distribution of streamflow is usually right skewed; hence, high extremes can have a large effect in the model training process. Standardisation transforms the original flow distribution into a normal distribution with a mean of 0 and standard deviation of 1. Thus, it is likely that the models are more sensitive to low flow variability, when trained with standardised streamflow. However, only a few hydrological studies use standardised streamflow, e.g. Modarres (2006). In meteorological studies, forecasting of standardised precipitation is more frequent, e.g. Mishra and Desai (2005), Morid et al. (2007) and Belayneh et al. (2014). Another beneficial aspect of forecasting standardised indices is that the transformed variables are normally dis-

Table 2. Climate indices used as potential predictors.

Variable/data set	Start	Source
Southern Oscillation Index (SOI)	January 1951	Climate Prediction Center of NOAA
Darwin sea level pressure	January 1951	Climate Prediction Center of NOAA
Tahiti sea level pressure	January 1951	Climate Prediction Center of NOAA
ENSO indices (ERSST)	January 1950	Climate Prediction Center of NOAA
ENSO indices (OISST)	January 1982	Climate Prediction Center of NOAA
North Atlantic Oscillation (NAO)	January 1950	Climate Prediction Center of NOAA
Indian Ocean dipole mode index (DMI)	November 1981	Based on OISST v.2 (Reynolds et al., 2007)
Oceanic Niño index (ONI)	February 1950	Based on ERSST v.3b (Smith et al., 2008)
Trans-Niño index (TNI)	March 1870	HadSST 1.1 and OISST v.2
Niño 3.4 (HadSST)	January 1871	Climate Prediction Center of NOAA

tributed and defined for $\mathbb{R}$, whereas precipitation and streamflow are defined for $\mathbb{R}^{\geq 0}$, only. Thus, corrections normally applied to prevent undefined forecasts (such as precipitation below zero) are not required.

Streamflow standardisation (Shukla and Wood, 2008; Vicente-Serrano, 2012b) is conducted using the algorithms implemented in the R package SPEI version 1.6. This algorithm uses the gamma distribution and unbiased probability weighted moments as the fitting method. The SSIs are calculated for each station at the scale of 6 months. SSI_6^{May} of May at that scale covers the desired main runoff period from December to May, henceforth named SSI_{DJFMAM}. The streamflow conditions are classified as drought when $SSI < -0.5$, which has a 30.9 % probability (given the normal distribution of SSI) and thus is an approximation of the lower tercile extremes.

2.4 Potential predictors: customised climate indices based on SSTs

Besides widely used climate indices, specific predictors are derived based on an analysis of past droughts and streamflow variability. SST fields and SSI are compared to detect ocean regions with predictive potential for streamflow in the Limpopo Basin. The analysis is limited to streamflow of the station Chokwe which has the largest catchment area of all stations. The ocean region is restricted to an area extending from latitudes 50° S to 25° N and longitudes 65° W to 115° E.

Composite analysis is used to identify ocean regions with predictive potential for droughts. Composites are generated for SST anomalies preceding drought from December to May defined by the drought threshold of $SSI_{DJFMAM} < -0.5$. Hence, composites are calculated for every month from the November before DJFMAM to the previous year's December. Composite maps are constructed by calculating the average SST field for the selected years. The resulting map shows the SST anomalies associated to droughts in the Limpopo Basin and the respective significance levels tested with the Mann–Whitney test for two samples.

Additionally, correlation analysis is conducted with SSI_{DJFMAM} and the SST field to identify ocean regions with predictive potential for streamflow variability. The significance of the Pearson correlations is calculated with $t = r\sqrt{\frac{N-2}{1-r^2}}$ using the Student's t distribution with $df = N - 2$ degrees of freedom, sample size N and observed correlation r by testing the null hypothesis $\rho = 0$ (correlation of the general population).

Ocean regions that show correlations and composite anomalies with a significance level of 0.05 are chosen for the construction of potential predictors. The region outlines are manually specified and defined rather generously so as to cover the anomaly regions resulting from different analyses. Then, indices are calculated by spatially aggregating the SST data to obtain time series with monthly means. Every index is calculated at three aggregation levels: 1, 3 and 6 months. A longer aggregation period indicates longer-lasting anomalies of SST, while the 1-monthly anomaly might capture short-term effects.

2.5 Forecast model setup

The objective for the modelling is the predictability analysis of standardised streamflow using teleconnections and catchment conditions as predictors in data-driven approaches. Suitable statistical methods are compared by assessing the prediction performance and robustness for drought early warning with a leave-one-out cross-validation scheme. The adopted statistical methods are MLMs, ANNs coupled to the genetic algorithm (ANN-GAs) and RFORs. MLMs are very common in modelling systems with linear relationships between predictors and predictands (see details in Sect. 2.5.1). ANN-GAs and RFORs are established data-mining methods. Both have the advantage of allowing nonlinear relationships. ANNs are applied in this work in order to evaluate if the forecast quality of the MLM predictor combinations can be improved by allowing nonlinear relations. In a similar study, where Australian rainfall was forecasted, Mekanik et al. (2013) achieved even better generalisation properties with ANNs compared to MLMs. ANN-GAs and RFORs dif-

Table 3. Lead time definition of SSI_{DJFMAM} forecast: time of input parameters (months abbreviated) and warning issue time.

Lead	12	9	6	3	2	1	0	−1
Input	prev. Nov	Feb	May	Aug	Sep	Oct	Nov	Dec
Issue	prev. Dec	Mar	Jun	Sep	Oct	Nov	Dec	Jan

fer, among other aspects, in the type of results which are deterministic for ANN-GAs and probabilistic for RFORs (see details in Sect. 2.5.2 and 2.5.3). An overview of data flow, model validation and forecast verification is presented in Fig. 3, and details are discussed in Sect. 2.6.

The models are set up to predict the standardised total streamflow of December–May (SSI_{DJFMAM}), which is one value per year, at the lead times of 1, 2, 3, 6, 9 and 12 months. We apply a strict definition of lead time as the difference (in months) between the availability of the forecast and the start of the predicted period. The resulting dates of forecast issue are presented in Table 3. Some time is lost due to the dependency on external predictor data sources, which are not available immediately due to collation and processing operations. Thus, the forecast based on month m would be available in the following month $m + 1$.

2.5.1 Multiple linear models

The first type of the data-driven models is the MLM. In MLMs, the dependent variable y is related to linear combinations of the intercept β_0, the predictors x_1 to x_p with slope factors β_1 to β_p and the error term ε:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon. \tag{1}$$

In this study, ordinary least squares regression is applied to estimate β_0 to β_p. This method requires independent predictors and normally distributed residuals. Collinearity in the predictor data set can cause overfitting effects. Despite these limitations, MLM is relatively robust to outliers, can produce good approximations and has been successfully applied in similar studies relating atmospheric teleconnections to drought indicators (Mishra and Singh, 2010), for example, in the prediction of Ethiopian rains (Diro et al., 2008, 2011). The strength is the simplistic and reductionistic approach, employing only a few significant predictors.

Predictor selection is performed by the automated both-ways stepwise selection algorithm. Starting from an empty model (intercept only), predictors are added to a model one at a time and the model quality is calculated. The predictor resulting in the highest model quality is retained and more predictors are added likewise in the following iterations. In both-ways stepwise selection, once added, predictors can still be removed from the model at a later iteration step. The selection process continues until the addition or removal of predictors does not lead to an increase in model quality. The measure for model quality has to balance the goodness of
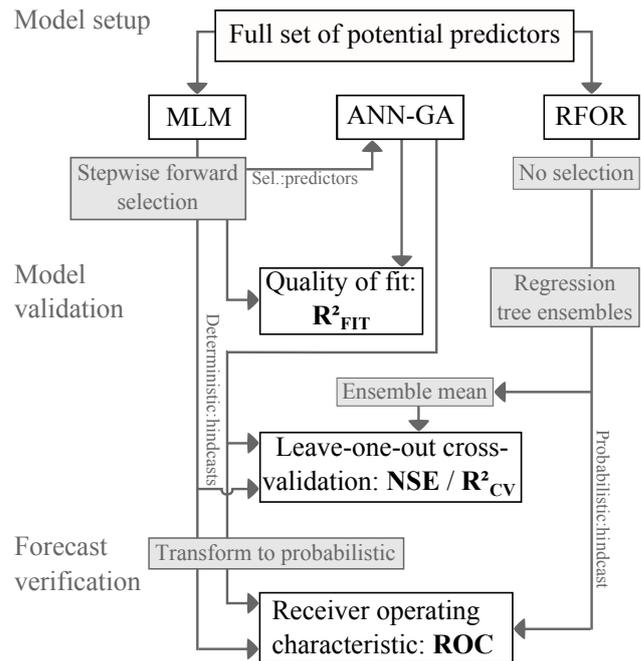
Figure 3. Modelling and validation scheme to account for the differences in models: MLMs, ANN-GAs and RFORs. The predictors selected for the MLMs are also used in the ANN-GAs, whereas RFOR works with the complete predictor set. MLMs and ANN-GAs are deterministic and require transformation to probabilistic form, whereas RFOR is probabilistic and is transformed to a deterministic value for the purpose of forecast comparison. Deterministic forecast skill is assessed using the coefficient of determination R^2_{CV}; the probabilistic properties are analysed using the ROC score.

fit with model complexity, i.e. the number of model parameters. Two measures are applied: Akaike's information criterion (AIC) and Bayesian information criterion (BIC), with the latter resulting in more conservative models.

The models' degree of overfitting is tested by performing leave-one-out cross validation. The difference by which the cross-validated root mean square error ($RMSE_{CV}$) exceeds the RMSE of the model fit residuals is used as a measure of generalisation properties. Furthermore, forecast uncertainty is estimated based on $RMSE_{CV}$. Models are developed for three sets of input aggregation levels and two information criteria (AIC/BIC), resulting in six models per lead time and station. For each lead time, the model achieving maximum generalisation properties and minimal error configuration is selected (only the selected models are presented in the results). The selected models contain varying numbers of predictors for which the relative importance is calculated. The contribution to the total explained variance, i.e. the predictor importance, depends on the order of the predictors. Predictor importance is calculated with the R package "relaimpo" according to Lindeman et al. (1980). Drought probability is calculated under the assumption of a normally distributed forecast error estimated using $RMSE_{CV}$ (Diro et al., 2011).

2.5.2 Artificial neural network models

The ANN is trained with the genetic algorithm (ANN-GA) which was successfully applied in forecasting rainfall in eastern Africa by Mwale et al. (2007). The genetic algorithm employs the processes of population growth for the model learning process. The network is designed with at least three layers, each containing a number of nodes. The nodes contain the data and are connected to the nodes of the next layer by transformation functions. The first layer is the input layer, where each node represents a predictor variable. The last layer is the output layer which contains the response variable. There can be several hidden layers in between, but in this study the models are set up with a single hidden layer. The number of nodes in the hidden layer is varied over four model setups with increasing complexity containing 3, 5, 7 or 10 nodes. All nodes of one layer are connected to the nodes of the next layer. The nodes are parameterised by so-called biases and weights, which define how the input from other nodes are weighted. The values of a node j in the hidden layer H are calculated based on the N input nodes x_i by

$$H_j = \sum_{i=1}^{N} W_{ji} x_i + B_{jo}, \tag{2}$$

where W_{ji} are the weights for the input nodes and B_{jo} are the biases of the hidden nodes. Then, the H_j value is translated by the nonlinear function

$$f(H_j) = \frac{1}{1 + e^{-H_j}}, \tag{3}$$

and combined by the weights and biases assigned in the output layer to calculate the prediction value.

The next step is the model learning with the genetic algorithm (GA) to determine the best parameterisation. The learning process starts with a random generation of ANN parameters. The GA is an iterative learning algorithm that regards model parameterisations as chromosomes in a genetic population (here 3000), which is subjected to evolutionary processes, as with every iteration step, a new generation is created undergoing mutation and crossover. A total of 15 % of chromosomes in the new generation are assigned with random parameterisations. First, chromosomes are ranked by forecast skill, called "fitness". Then, the best 85 % of the chromosomes are retained in the genetic pool. However, these are subjected to mutation and crossover processes. During the mutation process, in a small part of the chromosomes (here 5 %), some of the weights and biases are mutated, i.e. values are randomised. Thereby, small changes in the skilful chromosomes are triggered, which will be retained for the next generation, if forecast skill is improved. In the crossover process, pairs of chromosomes are chosen from the retained chromosomes, and weights and biases are exchanged at one point of the pair of chromosomes (here, with a crossover rate of 0.6). The crossover makes sure that skilful configurations stay in the population and slowly converge to one solution. The procedure is iterated until the root mean square error (fitness) is smaller than 0.005, or a maximum of 1500 iterations is reached. The ANN-GA method is applied as implemented in the R package "ANN" (Roy-Desrosiers, 2012).

The generalisation properties of the forecast models are evaluated by leave-one-out cross validation in the same way as with the MLM. The ANN-GA result is deterministic and is transformed to a probabilistic drought forecast with the same approach as with the MLM by assuming a normal distribution with a standard deviation estimated from the cross-validated $RMSE_{CV}$.

2.5.3 Random forest regression tree models

Regression tree modelling is a multivariate data-driven method and a special version of a decision tree tailored to predict continuous variables. Regression trees are used to predict a single variable based on multiple predictors by performing recursive partitioning on the training data. Hereby, the data set is partitioned into homogeneous groups, the so-called branches, which are identified by a specific condition and could be the occurrence of an El Niño event, which would lead to lower streamflow, for example. By repeating the partitioning process, a sequence of conditions based on several predictors leads to small homogenous groups, the so-called leaves. Regression trees are strict data-driven multivariate models able to map nonlinearity and interactions between predictors. This is a promising feature particularly for atmospheric and hydrological sciences, but up to now the method is not common in these disciplines. Hall et al. (2011) is one of the rare examples where the forecasting performance of random forests and other observation-based methods was evaluated. RFOR modelling fulfils several desirable characteristics (Breiman, 2001), including ease of parallelisation, robustness to outliers, fast calculation, internal estimates of error, strength and variable importance. The method extends regression tree modelling by introducing a tree model ensemble (here 500) which can be used to represent forecast uncertainty. Single-regression trees easily suffer from overfitting, which is improved by RFOR by training the trees in a bagging approach. Hereby, every tree is trained with a different data set, created by sampling from the original data set with replacement (in-bag samples). All observations not selected are referred to as "out-of-bag" and are used for validation and estimation of variable importance, which is described in Sect. 2.7. The random forest models are set up with 500 regression trees that have a minimum final node size of five observations. The implementation of the algorithm in the R package "randomForest" by Liaw and Wiener (2002) is applied. Although random forest provides an internal error estimation, leave-one-out cross validation is also applied for the random forest models for the sake of exact comparability with the other approaches.

2.6 Model and forecast validation

The wealth of potential predictors, some even showing a weak correlation (e.g. ENSO-related predictors), increases the risk for overfitting. Overfitting is the effect that a model can start fitting the noise contained in the predictor data instead of the signal. Robust statistical learning methods minimise the risk of overfitting. Every model learning algorithm is facing a tradeoff between data fit and generalisation. Model validation serves the purpose to identify the models with the best generalisation properties. Comparison of the models' forecasting results is performed using the independent forecasts resulting from a leave-one-out cross validation (LOO-CV), which results in a more realistic estimation of the real forecast uncertainty. The LOO-CV prediction time series resembles a hindcast series. The deterministic forecast performance of the models is assessed by the coefficient of determination, which is equivalent to the Nash–Sutcliffe efficiency (NSE).

A drought-specific forecast verification was performed with the receiver operating characteristic (ROC). The ROC score assesses the forecasts' skill to distinguish between occurrence and non-occurrence of drought which required probabilistic forecast transformation. A moderate level of drought was tested following the definition of drought below -0.5 (SSI < -0.5). In a ROC analysis, a diagram is constructed that presents the hit rate H in dependency of the probability of detection (POD) for a range of early warning thresholds. According to Wilks (2006), the first step in ROC analysis is the calculation of 2×2 contingency tables $C(I)$ for I warning thresholds with $0 < I < 1$. Applied to a probabilistic drought hindcast series, the hit rate

$$H = \frac{N_{\text{correct}}}{N_{\text{drought}}} \tag{4}$$

is calculated from the number of correct forecasts N_{correct} and the total number of occurred droughts N_{drought}. Consequently, the probability of false detection

$$\text{POD} = \frac{N_{\text{false}}}{N_{\text{nodrought}}} \tag{5}$$

is calculated from the number of false alarms N_{false} and the number of non-drought events $N_{\text{nodrought}}$ (wet or normal). The ROC score is the area under the curve and is used for model comparison. A perfect forecast reaches a ROC score of 1, while a score of 0.5 has no skill and is equivalent to a random forecast. The score is calculated with the R package "verification" (NCAR, 2012). This package employs a method by Mason (2008) who showed that the ROC score can be estimated from the Mann–Whitney U statistic. In order to estimate the uncertainty of the ROC score calculation, ROC score confidence intervals (95 % level) are estimated by 100-fold bootstrapping of the hindcast series and subsequent ROC score calculation.

In summary, all models are validated using the same leave-one-out cross-validation scheme, the result of which is used for skill assessment with the NSE and ROC scores. Therefore, comparability of model skills is assured also between different methods.

## 2.7 Analysis of predictor importance

Predictor importance is analysed for MLM and RFOR. The MLM predictor importance is calculated with the "lmg" method by Lindeman et al. (1980), which estimates a partial coefficient of determination for every predictor. These are affected by the order and combination of predictors in the model. The lmg method minimizes these effects and gives a robust estimate of the true coefficient. Due to the high number of potential predictors, the analysis is focussed on the following predictor groups: Atlantic, Indian Ocean, ENSO, DMI, NAO and streamflow.

Predictor importance in RFOR models is assessed differently. First of all, it is important to be aware that in regression tree models the interdependency of variables is an essential property of the method. The combinations of variables are of higher interest than single-variable importance only, which results in a different approach for importance analysis. Predictor importance for random forest models is based on the out-of-bag classification errors. The out-of-bag samples are randomised one variable after the other and the percentage increase in the prediction error is calculated. The understanding is that the more the randomisation of a predictor causes an increase in prediction error, the more important it is.

Collinearity of predictors can affect the importance estimation, since predictors might easily replace each other in the regression trees if they have a similar predictive strength. This can cause several effects. On the one hand, the importance per single predictor might be underestimated if it is not located at an important position in all regression tree models. On the other hand, in the presence of collinearity, there would be multiple predictors with underestimated predictor importance. Therefore, the results of RFOR predictor importance are summarised for comparison with the MLM partial coefficient of determination. Closely related predictors are merged as relative group importance, calculated as

$$I_g = \frac{\sum_{i=1}^{m} I_{g,i}}{\sum_{i=1}^{p} I_i},$$

which estimates the importance of g predictor groups with m group members and p total predictors. The relative importance per group can be displayed in a similar manner as the partial coefficients of determination available for MLM, allowing for comparison of predictor importance between the methods. When comparing the results, one has to consider the method-specific differences between partial R^2 and RFOR predictor importance.
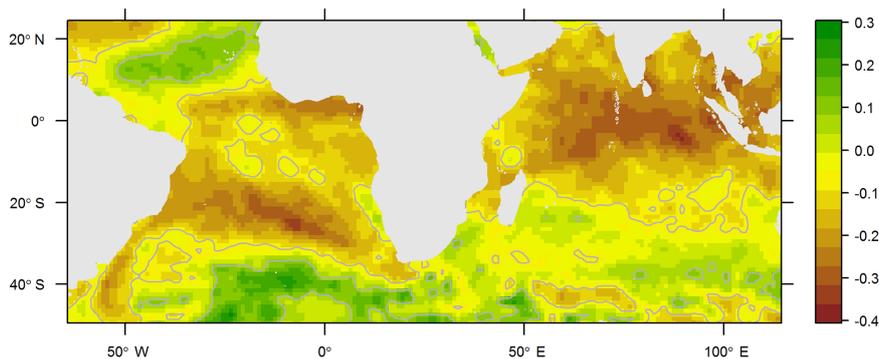
Figure 4. Correlation of standardised streamflow (6-month scale) of station Chokwe and sea surface temperature anomalies; grey contours indicate significance at 0.05.

3 Results and discussion

3.1 Identification of customised potential predictors

The list of potential predictors contains several well-established climate indices. These cover climate anomalies in the Atlantic, Indian and Pacific oceans but might not capture the effects in the proximity of southern Africa. Therefore, complementary customised climate predictors are deduced from correlation and composite analysis of SSTs in the southern Atlantic and Indian oceans for drought in the Limpopo Basin. SST is correlated with SSI_{DJFMAM} of station Chokwe, which has the largest basin of the stations. The Spearman correlation coefficient ranges from -0.36 to 0.26 with a median of -0.08 (Fig. 4). The correlations can be considered low but a large share of the correlations is still significant given a large sample size of 724 observations (months). Negative correlations are found in the northern Indian Ocean and the Atlantic from $10°$ N to $30°$ S. They indicate that warm anomalies are related to drought in the Chokwe signal. Correlations are strong in the Gulf of Guinea and the central southern Atlantic. Positive correlations are located in the southern ocean regions dominated by the circumpolar current. South of Cape Horn, small-scale random patterns are found. At the Namibian and Angolan shorelines, a relatively thin zone exhibits positive correlations as well, which might be attributed to upwelling cold water from the Benguela Current in the region.

The composite analysis for conditions preceding SSI_{DJFMAM} drought shows more restricted regions with significant anomalies (Fig. 5). Droughts are associated with positive anomalies in the north-western Indian Ocean during the preceding October and November. During June and July, positive anomalies occur in the south-eastern region south of Madagascar. In the Atlantic, positive anomalies are significant in the Gulf of Guinea at longer lead times of 10–12 months. In October, positive anomalies also appear south of the African continent. Similar to the correlation analysis, these show a high small-scale variability. This

ocean region is characterised by a complex system of currents with upwelling cold water of the Benguela Current in the west and the Agulhas warm water current in the east, which collides with the South Atlantic Current. The anomalies in the mixing region of the South Atlantic Current and the Agulhas Current are very small and might be related to warm- or cold-water eddies (see Fig. 1), which form under the special conditions of the two mixing currents (Peterson and Stramma, 1991). Due to the small extent of the anomalies and the chaotic nature of the eddy formation, these anomalies are not included as predictors. Nevertheless, the currents themselves are represented by customised predictors based on other ocean regions in the Indian Ocean (predictor named "Agu") and the southern Atlantic (predictors named "SWAtl", "SEAtl" and "BC" in Fig. 6).

As a result of the correlation and composite analysis, in total, 10 ocean regions are defined (see Table 4 and Fig. 6) and potential SST predictor indices are calculated with three aggregation levels (1, 3, 6 months), resulting in 30 customised SST indices. The total set of potential predictors comprises 55 variables. Of these, 16 are well-known climate indices, of which 14 are related to El Niño and SOI. In addition to NAO, the influence of the Atlantic region is represented by 18 customised predictors (6 regions by 3 aggregation levels). The Indian Ocean, represented by the climate index DMI, is complemented by 12 additional customised predictors (4 regions and 3 aggregations levels). Furthermore, there are three predictors representing the antecedent catchment conditions in the form of current standardised streamflow at aggregation levels of 1, 3 and 6 months.

### 3.2 Inter-model comparison of predictor selection and importance

MLM and ANN models consist of specifically reduced predictors sets, whereas RFOR relies on the complete predictor set. Therefore, predictor selection frequencies are presented for MLM only. Predictor importance is compared for MLM and RFOR, for which different estimators of predictor im-
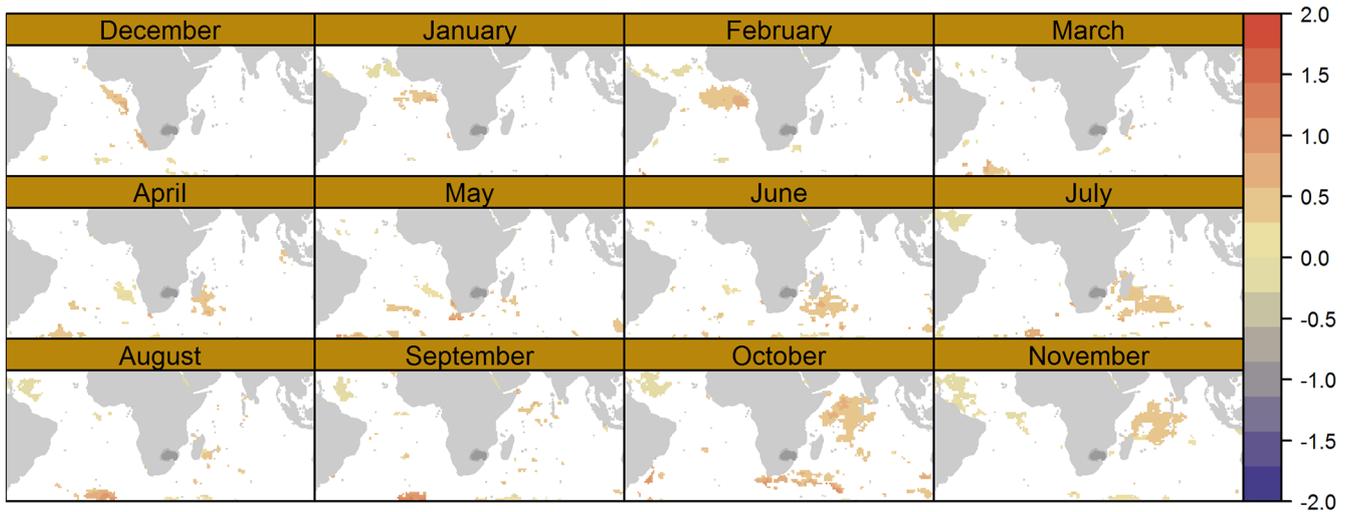
Figure 5. Composites: anomalies of sea surface temperature preceding drought during DJFMAM (significant anomalies presented only.)

Table 4. Potential customised predictors from ocean regions tele-connected to drought in the Limpopo Basin. Region selection is based on correlation and composite analysis. Coordinates indicate extents of the polygons (minimum, maximum).

| Ocean region	Latitude	Longitude	Abbreviation
Atlantic Ocean			
Benguela Current	−34, −6	−12, 13	BC
Southern African coast	−38, −20	8, 25	SACoast
South-western	−48, −40	−40, −15	SWAtl
South-eastern	−48, −40	−15, 15	SEAtl
Gulf of Guinea	−5, 7	−25, 12	GuiGulf
Guinea coast	7, 20	−40, −10	GuiCoast
Indian Ocean			
Eastern equatorial	−18, 2	75, 120	eeqIO
Western equatorial	−10, 5	40, 60	weqIO
Agulhas Current	−35, −20	28, 45	Agu
South of Madagascar	−37, −27	45, 60	SMad
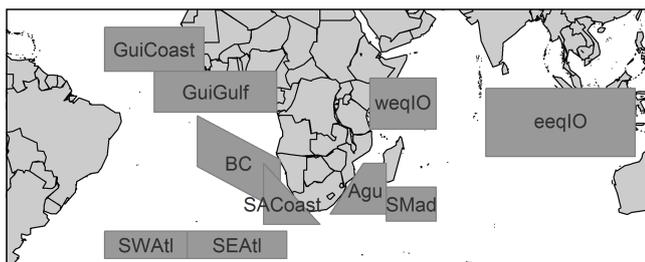
**Figure 6.** Regions of customised SST indices. Predictor region abbreviations are listed in Table 4.

portance exist: partial coefficient of determination and RFOR predictor importance.

The proportion of selection of the predictors in the MLM models (Fig. 7) shows which predictors are frequently part of the final MLM (and ANN) models. Antecedent streamflow is selected with highest frequency (Fig. 7), followed by several customised indices of the Atlantic and DMI. The first of the El Niño-related parameters (ENSO1.2) is in seventh place. Every ENSO-related predictor has a selection frequency of less than 0.21, which is rather low given the relevance of ENSO in the region. The indices are based on different SST ocean regions (ENSO 1.2, 3, 3.4) or are calculated based on SLP (SOI); however, they are correlated and the indexes might easily replace each other in different selection runs. As a result, the proportion of selection might be low for specific ENSO indices but not for the ENSO anomaly altogether. Additional ENSO-related predictors are tested originating from two different data sets: ERSST and OISST. ERSST would be the recommended data set for time series modelling due to its greater length of record. On the other hand, the OISST data set is a shorter data set with higher quality achieved by the inclusion of new and improved types of SST observations such as satellite imagery. However, the results show that the ERSST ENSO indices are selected more often than the OISST ENSO indices.

Collinearity between predictors might have affected ENSO predictors, but also other factors such as DMI and Darwin SLP, which show some correlation with ENSO signals. Darwin SLP was found to be superior to ENSO for drought prediction in Zimbabwe by Manatsa et al. (2007). Due to the correlation of these signals, they cannot be distinguished by a linear model as used here. However, Darwin SLP is only selected in 12 % of the models, which does not support the finding by Manatsa et al. (2007), since ENSO predictors were selected more frequently.
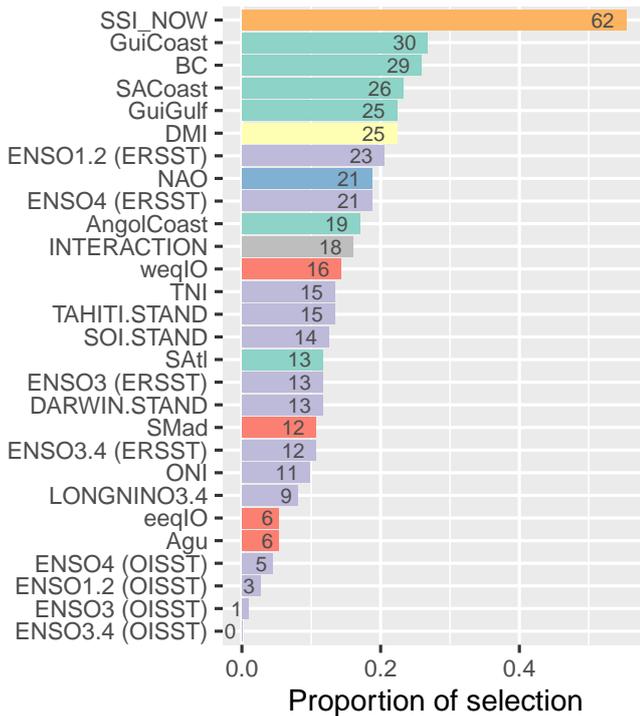
Figure 7. Predictor proportion of selection out of 112 (7 lead times for 16 stations) multiple linear models, with absolute number as text. The presented predictors include customised and predefined teleconnection indices (for abbreviations, see Table 4), current streamflow (SSI_NOW) and interactions of predictors (no specific interactions presented). Bars are coloured according to groups: customised Atlantic (green), DMI (yellow), ENSO (purple), customised Indian Ocean (red), NAO (blue), current streamflow (orange), interactions of selected predictors (grey).
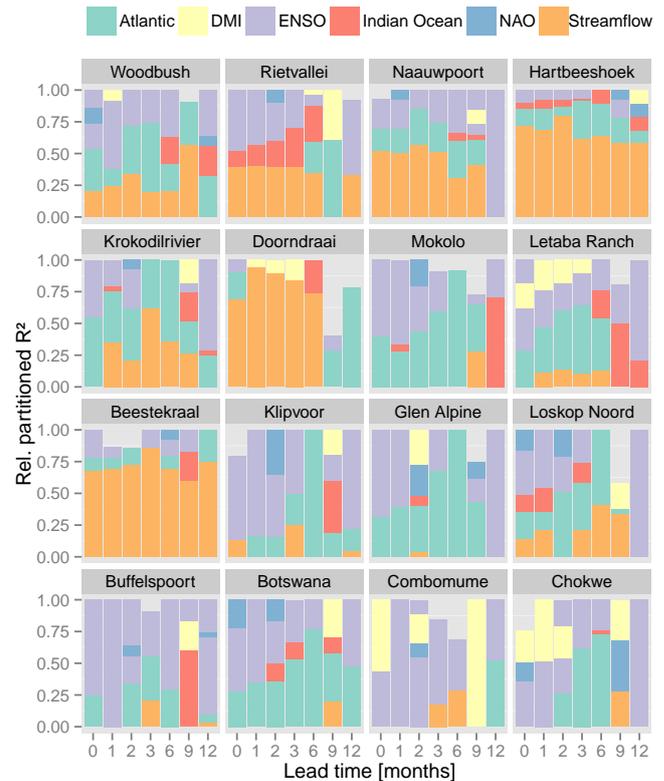
Figure 8. Relative partial R^2 by the predictors in the MLMs summarised as follows: antecedent streamflow conditions, Indian Ocean dipole mode index (DMI), El Niño–Southern Oscillation (ENSO) indices, North Atlantic Oscillation index (NAO) and the customised indices of the southern hemispheric Atlantic (Atlantic) and the Indian Ocean (Indian Ocean). Parameter interactions are excluded from the plot and account for the white gaps. Stations are ordered by catchment area from small (top left) to large (bottom right).

In fact, over all stations and lead times, ENSO-related predictors form the most important group of SST predictors. The group includes ENSO indices but also SOI, as well as Darwin and Tahiti SLP. Predictors from this group are selected in 80 % of the linear models (Fig. 8). The ENSO predictors' contribution to overall explained variability differs widely between models but the median is rather high with 41.2 %. However, in contrast to the majority of stations, the stations Hartbeeshoek, Beestekraal and Doorndraai show only a weak or even absent effect of ENSO predictors (Fig. 8).

The customised predictors for the Atlantic are incorporated in many linear models (73 %) and the median of the relative partial coefficients of determination reaches 31.8 %, which is only slightly below the contribution of the ENSO indices. The contribution is particularly strong at the stations Glen Alpine and Botswana. Besides El Niño and ocean regions in the Indian Ocean, parts of the Atlantic are also described as an important factor for southern African rainfall by Reason et al. (2006). In their study, the predictability of rainfall is attributed to the influence of the Benguela Current and the SST of the south-eastern Atlantic, which is related to

the South Atlantic Current. In addition, our results indicate a connection to the SST of the Gulf of Guinea (GuiGulf) and equatorial Atlantic (GuiCoast) (customised indices in Fig. 7). The NAO index is also included in the MLM models, but only at a rate of 18 %. The contribution of NAO to the explained variance is 13 % (median).

Antecedent streamflow (SSI_NOW) is selected in 55 % of the linear models and is very important for many of them. In half of the models where antecedent streamflow is included, the predictor contributes at least 42 % of the explained variance (median rel. part. R^2). The importance of antecedent catchment state is supported by van Dijk et al. (2013), who found that initial conditions provided most skill as opposed to meteorological forcing in a forecasting experiment with a global ensemble streamflow prediction system. Antecedent streamflow is particularly prominent in smaller catchments (Fig. 8). In statistical models, antecedent streamflow is a common predictor which exploits signal autocorrelation that is caused by the delayed rainfall–runoff response in the hydrological system (Robertson and Wang, 2012). As a param-

eter representing catchment memory and other autocorrelation properties, it could be expected that the importance of the predictor decreases with higher lead times. This effect is observed at stations Hartbeeshoek, Doorndraai, Krokodilrivier and Naauwpoort, but the decrease is not strong. The most obvious effect is present for a lead time of 12 months, where antecedent streamflow is selected only in 4 of 16 stations.

The dipole mode index (DMI) of the Indian Ocean is selected less often. Of all models, 22 % include the index as predictor and its median share in the models' explained variance is 20.1 %. This is of particular interest given the selection rate of the customised indices of the Indian Ocean, which reached 30 % with a median relative partitioned R^2 of 0.15. The Indian Ocean predictors are selected particularly at longer lead times. In short, DMI is seldom selected, but has a comparatively high importance in the models.

The predictor importance of the MLM models differs strongly between lead times and stations. Several cases exhibit very different parameter selections than foregoing lead times (for example, a lead time of 9 months of station Rietvallei). These cases result in an impression of randomness in predictor selection which might indicate that these observations are statistical artefacts. One possible reason might be that, on the one hand, these statistical artefacts could occur when selection is performed under non-ideal conditions, e.g. collinearity. On the other hand, different predictor configurations might lead to very similar AIC/BIC values, but only the model with the highest value is chosen. As a result, the estimated predictor importance for the MLM models is highly specific to the selected models and can be inconsistent between lead times. In addition, it has to be kept in mind that most of these models only achieve a low R^2 below 0.3. Therefore, a predictor reaching 10 % in relative partitioned R^2 at a total explained variance of only 30 % still only explains about 3 % and has very low influence on the forecast. Thus, one should try to find the overarching pattern and not overinterpret specific contributions at certain lead times.

In contrast, the results of the RFOR models can provide a more general picture, as they always include all predictors and use a randomisation process to estimate predictor importance, shown in Fig. 9. For the ease of comparability, Fig. 9 is designed similarly to the relative partial R^2 of the MLMs presented in Fig. 8, but it is important to note that the importance measures are different (see Sect. 2.7). RFOR predictor importance shows the sensitivity of the model error to the individual predictors. RFOR produces a more even pattern of predictor importance than MLM. This is caused by the fact that RFOR encompasses all predictors. The randomised RFOR ensembles are then compared to single MLM realisations, which are bound to one specific selection.

The RFOR predictor importance shows four main differences and two confirming features in comparison to the MLM results. First, the Atlantic has a more constant and stronger importance. Second, in contrast to the MLM results, a stronger effect by the lead time is observed, for example, at
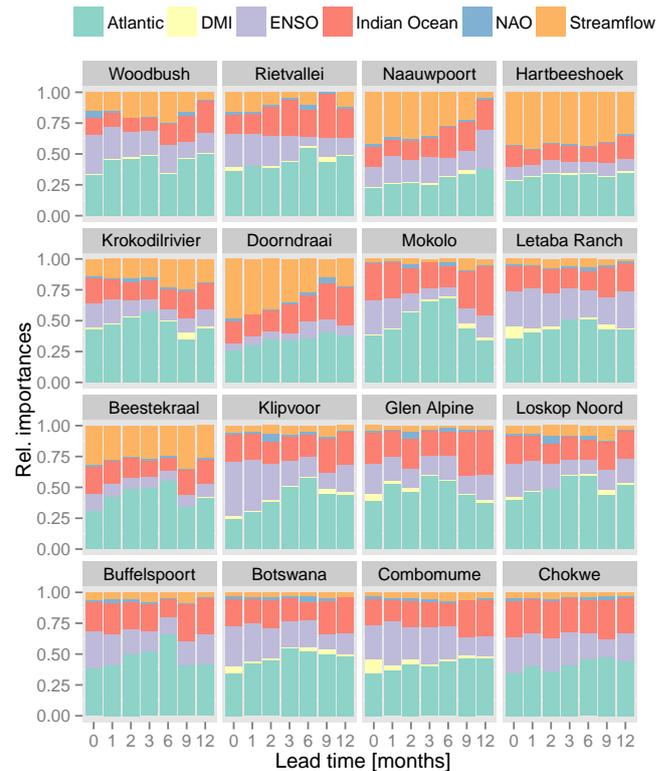
Figure 9. Relative importance of predictors in the RFOR models grouped as follows: antecedent streamflow conditions, Indian Ocean dipole mode index (DMI), El Niño–Southern Oscillation (ENSO) indices, North Atlantic Oscillation index (NAO) and the customised indices of the southern hemispheric Atlantic (Atlantic) and the Indian Ocean (Indian Ocean). Stations are ordered increasingly by catchment area from small (top left) to large (bottom right).

station Doorndraai, where the importance of streamflow decreases with higher lead times (Fig. 9). At several stations the importance of the Atlantic Ocean predictors increases from a lead time of 0 to 6 months and drops thereafter, which is also observed in the MLM models. Third, the predictors from the Indian Ocean are more important at all stations and are more constant over all lead times. Fourth, ENSO predictors are less important compared to the MLM models, where they are the dominant predictor group.

The RFOR results confirm the major relevance of antecedent streamflow at different lead times, and produces a very similar pattern compared to relative partial R^2 of the MLMs, where streamflow is a strong predictor at stations with smaller catchment areas. Furthermore, the effect of decreasing importance of streamflow with longer lead times was strongest at the stations Naauwpoort and Doorndraai. The results also confirm the lower value of NAO and DMI. Overall, the customised SST indices in the Indian Ocean are more emphasized and more persistent for all stations and lead times compared to the MLM results. This might be an effect of the forced selection of only a single final MLM that
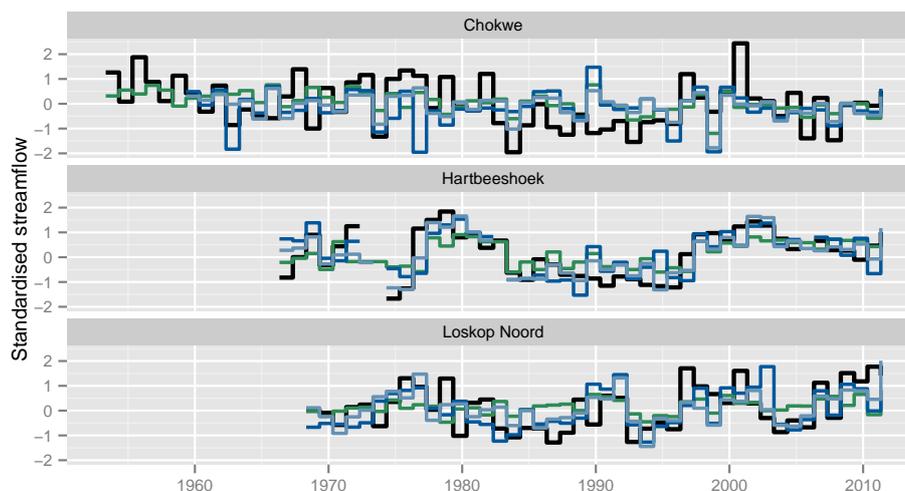
Figure 10. Hindcasted time series of DJFMAM forecast models at a lead time of 1 month (November forecast) of stations Chokwe with low skill, Hartbeeshoek with good skill and Loskop Noord with medium skill. The presented time series are the cross-validated, thus independent, results of the MLM (light blue), ANN (dark blue) and RFOR (green) models. Observed standardised streamflow is shown by the underlying black line.

causes the Indian Ocean indices to be dropped from some models. However, it might also indicate that indices in the Indian Ocean in particular have a conditional relationship with other indices, which could only be represented by RFOR and not MLM. The low forecasting skill achieved by RFOR does not encourage further investigation in this matter.

In summary, the study shows that customised SST indices can contribute substantially and even outperform global climate indices in predictor importance. It is well known that ENSO has an impact on drought occurrence in southern Africa, but the strength of the relationship has been questioned since other indices like the Darwin SLP or DMI also exhibit a strong – and supposedly stronger – influence as suggested by Manatsa et al. (2007). Distinguishing the importance of these correlated predictors is difficult but our results do not strengthen that finding. In our study, DMI and Darwin do not exceed the importance achieved by ENSO. Instead, specifically customised indices in the Atlantic and Indian oceans show good potential in statistical drought forecasting. There is no doubt about the strong effect of ENSO on global circulation patterns. However, our study suggests that predictability studies are advised to create customised indices that refer to the drought-specific SST anomalies in the region surrounding the area of interest to supplement the global indices. By doing so, it is more likely to capture all factors leading to drought events.

3.3 Forecast skill for drought early warning

The forecast skills vary widely between stations, models, lead times and time periods. For example, the forecast of the largest subcatchment (station Chokwe) achieves low skill (see Fig. 10). A few drought events are forecasted well at a

lead time of 1 month (e.g. 1974), but often event magnitudes are not met (after 2000) or overestimated (several events between 1977 and 2000). The forecast of Hartbeeshoek at the same lead time presents a more promising picture, where models are able to represent the long-term variability of the streamflow but the inter-annual variation is overestimated by the models. The forecast at the station Loskop Noord is of medium skill where, for example, the dry period during the mid-1990s is forecasted but several other extreme events are not. These three examples give an impression of the range of forecast results achieved by a lead time of 1 month.

The forecast skill can be assessed in many ways (Wilks, 2006), both deterministically and probabilistically. In this study, the skill of the different forecast models is deterministically evaluated with the leave-one-out cross-validation NSE, and probabilistically with the ROC score. The former measures the accuracy in forecasting the exact deterministic SSI value, whereas the latter assesses the discriminative skill of a probabilistic drought forecast in early warning mode.

An analysis of the deterministic skill for all stations and lead times, using LOO-CV NSE, reveals that MLM produces the most robust forecasts and achieves the highest forecast skills with a maximum of 0.73 and a median of 0.30 (Fig. 11). The maximum skill reached by the ANN is a little bit lower with 0.61 but the median is very low with 0.03 which is caused by the absent skill at many stations. The ANN forecasts only achieved considerable skill at the stations Naauwpoort, Hartbeeshoek, Krokodilrivier, and a few more cases. The ANN models strongly suffer from overfitting. Regardless of the number of hidden neurons, the difference between fit and cross-validated error is higher than for MLM (data not shown). Employing ANN with the predictors selected for the MLM does not lead to improved forecast skill in this study.
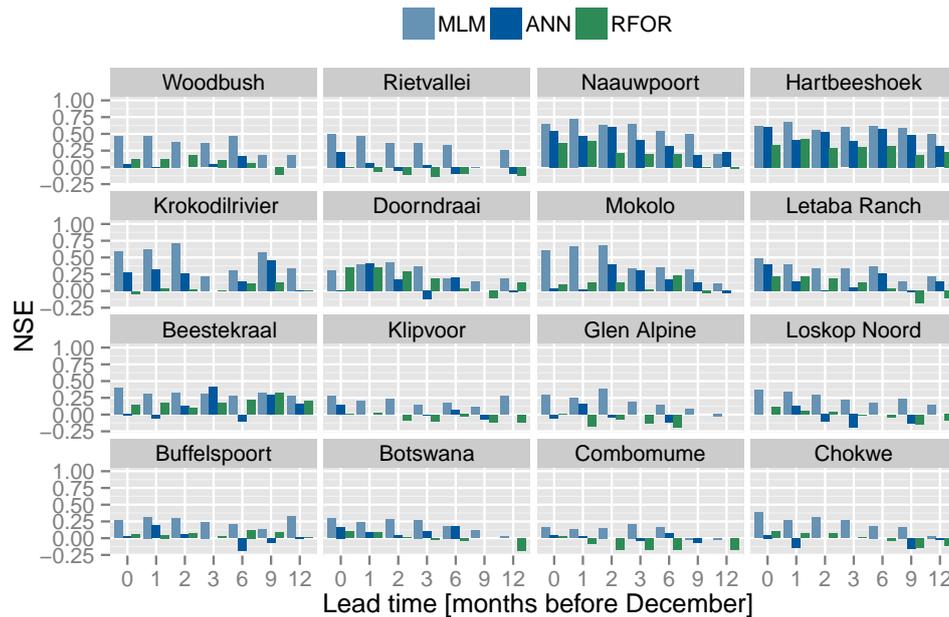
Figure 11. Validation of DJFMAM drought forecasts with LOO-CV Nash–Sutcliffe model efficiency: MLMs (bright blue), ANNs (dark blue) and RFORs (green). Stations are ordered increasingly by catchment area from small (top left) to large (bottom right).

At stations Naauwpoort and Hartbeeshoek the predictability by all model systems is highest. The model skill shows strong inter-station variability, which is not unusual in streamflow forecasting (Robertson and Wang, 2012). While MLM achieves skills like it does in Naauwpoort and Hartbeeshoek at a few more stations, ANN and RFOR only rarely reach that level. The skill is highest in the smaller sub-basins (upper two rows of Fig. 11) and lowest in the bigger catchments (lower two rows of Fig. 11).

The skill of the probabilistic drought forecasts is analysed with the ROC score presented in Fig. 12. The forecasts have more skill than a climatological forecast, i.e. ROC > 0.5, at almost all stations and lead times. As in the deterministic evaluation, skill levels vary strongly between stations. Prediction skill decreases with higher lead times at most stations. The reduction in skill is most pronounced for the longest lead times of 9 and 12 months. Exceptions to this are the stations Beestekraal and Hartbeeshoek having almost constant skill at all lead times. The skill at stations Buffelspoort and Chokwe is generally at a lower level and exhibits an unusual pattern where longer lead times have higher skill.

The three statistical methods achieve different median ROC scores. When models are ranked per station and lead time, the majority of first ranks is achieved by MLM models, most second ranks are ANNs and most third ranks are RFOR models (73, 63 and 63 % of the models per rank, respectively). MLM often reaches the highest skill and is therefore ranked in first place, but it has to be noted that often the differences are minor. Also, there are a few instances where this pattern does not hold. For example, RFOR is much more skilful in forecasting station Chokwe at lead times of 0, 1

and 2 months. These results are similar to the strong inter-station variability found by Robertson and Wang (2012) in a streamflow forecast study for Australian catchments.

Another interesting feature of the results is the variability of the error bars associated to the ROC scores. The error bars are derived by resampling the hindcast series and indicate the influence of individual observations on the robustness of the model predictions. It can be seen that for stations with generally good skill the errors are also small, while for the stations with lower skill the errors tend to be larger. Large errors indicate that a few observations have strong influence on the forecast skill, implying that drought forecasts in these basins are generally difficult with the presented models. A likely explanation for this is the limited length of the observation data time series. For stations with large errors, the observations might not be representative for the underlying real distribution of discharges; thus, robust forecast models cannot be achieved.

Relating the ROC skill with the selected predictors reveals that stations with high ROC scores coincide with a high influence of antecedent streamflow. This is an indication that catchment conditions play an important role for both the development and predictability of droughts. Interestingly, our results show a much higher skill in forecasting droughts in smaller catchments. This might be explained by the degree of human interferences. A plausible hypothesis is that the degree of human interference with streamflow is lower in smaller catchments. The low prediction skill in large catchments might be related to the complexity of the large catchments, where many dams, irrigation schemes and groundwater extractions interact and thus cause lower predictabil-
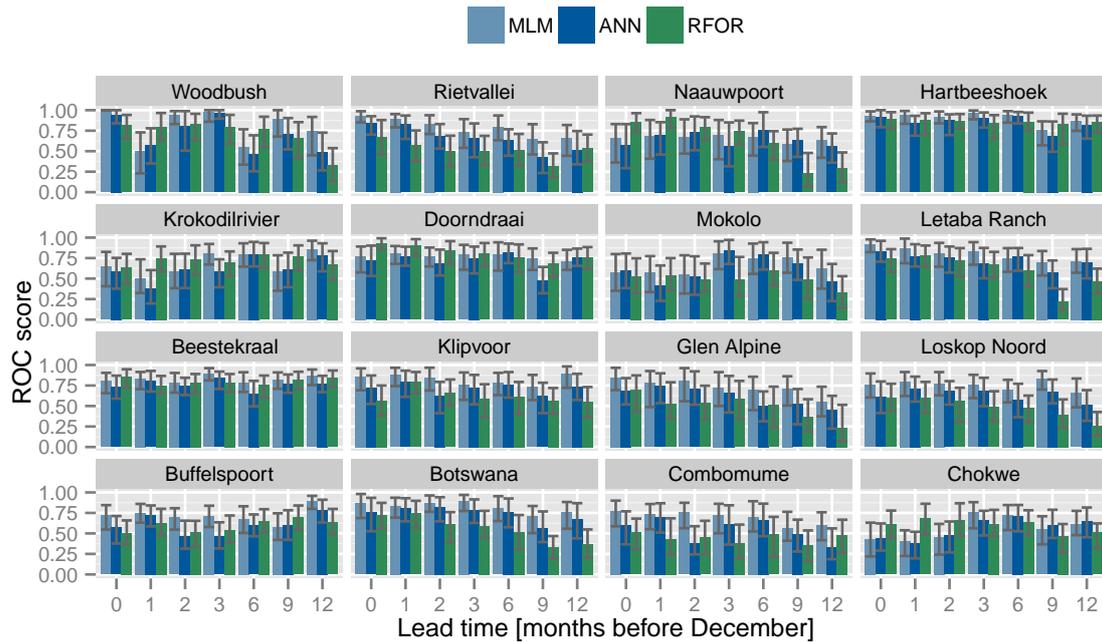
Figure 12. Validation of DJFMAM drought forecasts with ROC scores: ANNs (dark blue), MLMs (bright blue) and RFORs (green). Bars indicate median ROC scores, error bars show bootstrapped 0.95 confidence intervals of the scores. Stations are ordered increasingly by catchment area from small (top left) to large (bottom right).

ity with the adopted methods. Another reason for the better forecast skill for smaller catchments might be a regional bias, since most of the smaller headwater catchments are located in the south of the Limpopo Basin. A definite answer must be left for further analyses focussing on the role of human interferences in the Limpopo Basin.

In order to evaluate the forecast skill of the proposed models in relation to other approaches, a benchmark forecasting system would be necessary, but is not available. However, Dutra et al. (2013) published seasonal forecasts of SPI-6 for the Limpopo (Chokwe sub-basin) but did not present ROC scores specific for a SPI_{DJFMAM} forecast in December and earlier. The studies' lead times correspond to lead times -1 to -5 according to the definition used here (see Table 3) and result in forecasts issued in January–May. The forecast skill (ROC scores) in Dutra et al. (2013) decreases strongly approaching 0.5 (no skill) at a lead time corresponding to January for the SPI_{DJFMAM} forecast (Fig. 13). In comparison, the results presented here for hydrological drought forecasting exceed the skills of the highest presented lead time in Dutra et al. (2013), as does station Chokwe, which covers the same area. It has to be noted, though, that Dutra et al. (2013) present continuous forecasts for all months of a year, and intra-annual variation of predictability is not shown.

In summary, there is reason to regard the seasonal prediction of hydrological droughts in the Limpopo Basin as challenging. For example, the resulting NSE is on a low level at most stations. In addition, the achieved skill of the MLM models sometimes appear random despite the thorough mod-
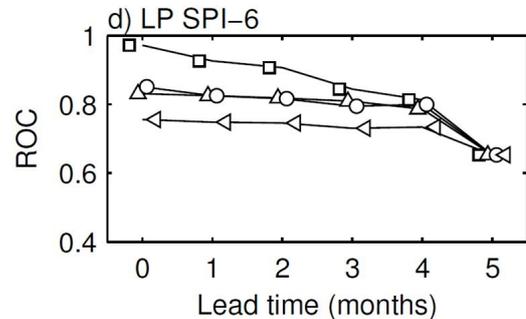
Figure 13. ROC scores of seasonal forecasts of SPI-6 of the Limpopo Basin by Dutra et al. (2013). Lead times have been defined differently in the following way: the lead time represents the number of months before the last of the 6 months of SPI-6, meaning a lead time of 0 months would correspond to a forecast of SPI_{DJFMAM} in May; hence, a lead time of 5 months would be equivalent to a lead time of 0 months in this study. The lines indicate different precipitation data sets. (Source: Fig. 11d from Dutra et al., 2013.)

elling design. For example, the selected predictor combinations can change completely from one lead time to the other. Although this might be an artefact caused by the collinearity of the predictors and the crisp selection of the predictors (Sect. 3.2), these results indicate a generally low predictability of SSI_{DJFMAM} at the lead times presented here. The predictability as tested in this study can build on meteorological influences, for example, sequences of weather patterns,

represented by SST anomalies. The hydrological influences, for example, catchment memory effects, are represented by antecedent streamflow. The hydrological signal contained in the streamflow series, however, is composed of many more influencing factors than these (hydrological processes and anthropogenic interference, in particular). Hence, only a minor part of the signal contained in the streamflow series is predictable by the adopted methods. Therefore, the forecasts exhibit a high uncertainty, a part of which is naturally high given the long lead times and represents the aleatory part of the uncertainty. Furthermore, streamflow observation in general is subject to high uncertainty. Yet, a considerable part of the uncertainty is epistemic, i.e. it is likely to be reduced by model improvements or additional data. There are several factors unaccounted for that could reduce the epistemic error. Examples are the introduction of anthropogenic interference with streamflow (abstraction, storage) or the introduction of further hydrological and meteorological parameters. These factors heavily influence the hydrology in some parts of the Limpopo Basin, where massive abstractions take place or where inter- and intra-basin water transfer schemes are facilitated. Considerable parts of these influences were increasing over the last decades, adding to the non-stationarity of the analysed streamflow signals. Therefore, despite all these unaccounted factors and the generally low forecast skill, the results represent respectable skill at high lead times at several locations within the Limpopo Basin.

4   Conclusions

This study presents the predictability of hydrological droughts in the Limpopo Basin, transferring methodologies predominantly used in meteorology to hydrology. The results show that hydrological drought in the Limpopo can be predicted based on climate indices, SST teleconnections and antecedent streamflow, although the predictability varies between catchments and lead times. Seasonal forecasting is a demanding task in a catchment with very high anthropogenic interference with the hydrological processes. Nevertheless, seasonal to annual predictability is still present in the streamflow signal, which is shown by forecasting skills that exceed climatology. This study has four main findings.

First, although standardised indices are less common in hydrology than in meteorology, we recommend their use in hydrological drought studies. Our results show massive inter-station differences in achieved skill, reaching up to 0.73 (NSE) at a 1-month lead time, which indicates that seasonal forecasting standardised streamflow with statistical methods can prove to be successful. At some stations skill is present up to a 12-month lead time, but many stations (larger catchments in particular) only achieve little skill. However, forecasting SSI enables a better identification of droughts using a common drought definition in basins of different characteristics and size.

Second, the most important climate predictors are ENSO-related as well as customised drought predictors in the southern Atlantic based on sea surface temperatures. In addition to the climate indicators, antecedent streamflow as a proxy for catchment state proves to be another important predictor. Regarding antecedent catchment conditions, the catchments are separated in two groups: one group with a strong importance of antecedent streamflow and comparatively high forecast skill and one with low (or absent) influence and lower forecast skill. Most smaller catchments fell in the first group. The reason for that pronounced effect could not be answered in this study and must be left for future work. Possible causes are the degree of human interference in the catchments, which is likely to be lower in smaller catchments, or a regional bias in the predictability caused by climatology. However, the importance of antecedent streamflow underlines the relevance of catchment conditions for hydrological drought prediction.

Third, the best forecasting skill within this study is achieved with multiple linear models. Based on the results of the cross validation, linear relationships are more robust than the nonlinear models derived by artificial neural networks and random forest methods. ANNs and RFORs are likely to suffer from overfitting of the models, which is in turn a consequence of the limited data set used for training. These results are specific to this region and data set, but they underline the necessity to benchmark more advanced methods with simple methods.

Fourth, in order to determine a forecast's value for early warning, a thorough forecast system verification is imperative. Verification must incorporate both the deterministic properties using, e.g. the NSE and the probabilistic properties using skill scores like ROCs. The deterministic forecast skill shown by NSE is low at many stations, but the analysis of the discrimination properties of the probabilistic forecast shows that the forecasts still exceeds a pure climatological forecast and therefore should not be neglected. Up to now, climatology as a benchmark is adequate, since water management in the basin is typically relying on it instead of seasonal forecasting as a basis for decision making.

This study shows that hydrological drought can be predicted using statistical methods and teleconnection indices and catchment condition as parameters. The methods can be applied in places with available observed streamflow. It is useful when station-specific forecasts have value for water management and decision makers. Its simplicity and low computational demand make it even adoptable as a customised forecast system at an end-user level for dam or sub-basin management. Thus, it is suited for a bottom-up early warning system at the local level, where decisions are set in action.

5 Data availability

The research data underlying this study are mostly publicly accessible and free of charge for research purposes. However, the authors do not have the right to redistribute the data.

Input runoff data and dam capacities of South Africa were accessed directly from the data base of the department of water affairs (DWAF) hydrology section. A reference publication could not be found but the data is available from this web page: https://www.dwa.gov.za/hydrology/.

Further runoff data was acquired from the Global Runoff Data Base by the Global Runoff Data Centre. Input runoff data of GRDC is freely available on request, but cannot be downloaded directly, hence here is only the link to the webpage, an official publication and doi is not available: http://www.bafg.de/GRDC/EN/01_GRDC/13_dtbse/database_node.html

The SST data set HadISST1.1 (Rayner et al., 2003), which was used for the creating of customized SST indexes, is available from http://hadobs.metoffice.com/hadisst/data/download.html (doi:10.1029/2002JD002670).

Catchment outlines and were created using the data set HYDROSHEDS, which is available from: http://hydrosheds.cr.usgs.gov/dataavail.php. Various atmospheric indeces were acquired from the National Oceanic and Atmospheric Administration (NOAA).

Furthermore, several atmosheric indeces were collected from the NOAA webpage and others.

This is the complete list including variable/data set and its source.

– Southern Oscillation Index (SOI),
  Climate Prediction Center of NOAA:
  ftp://ftp.cpc.ncep.noaa.gov/wd52dg/data/indices/soi

– Darwin sea level pressure,
  Climate Prediction Center of NOAA:
  http://www.cpc.ncep.noaa.gov/data/indices/darwin

– Tahiti sea level pressure,
  Climate Prediction Center of NOAA:
  http://www.cpc.ncep.noaa.gov/data/indices/tahiti

– ENSO indices (ERSST),
  Climate Prediction Center of NOAA:
  http://www.cpc.ncep.noaa.gov/data/indices/

– ENSO indices (OISST),
  Climate Prediction Center of NOAA:
  http://www.cpc.ncep.noaa.gov/data/indices/

– North Atlantic Oscillation (NAO),
  Climate Prediction Center of NOAA:
  ftp://ftp.cpc.ncep.noaa.gov/wd52dg/data/indices/

– Indian Ocean Dipole Mode Index (DMI),
  derived from NOAA OISST 2 by JAMSTEC:
  http://www.jamstec.go.jp/frcgc/research/d1/iod/DATA/

– Oceanic Nino Index (ONI),
  based on ERSST.v3b of Climate Prediction Center of NOAA:
  http://www.cpc.ncep.noaa.gov/products/analysis

– Trans Nino Index (TNI),
  HadSST1.1 (until November 1981) and NCEP NOAA OI:
  http://www.esrl.noaa.gov/psd/gcos wgsp/Timeseries/

– NINO3.4 (HadSST),
  Physical Sciences Division of NOAA:
  http://www.esrl.noaa.gov/psd/gcos wgsp/Timeseries/

References

Belayneh, A., Adamowski, J., Khalil, B., and Ozga-Zielinski, B.: Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models, J. Hydrol., 508, 418–429, doi:10.1016/j.jhydrol.2013.10.052, 2014.

Breiman, L.: Random Forests, Mach. Learn., 45, 5–32, doi:10.1023/A:1010933404324, 2001.

Chen, J., Li, M., and Wang, W.: Statistical Uncertainty Estimation Using Random Forests and Its Application to Drought Forecast, Math. Probl. Eng., 2012, 915053, doi:10.1155/2012/915053, 2012.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteorol. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

Diro, G. T., Black, E., and Grimes, D. I. F.: Seasonal forecasting of Ethiopian spring rains, Meteorol. Appl., 83, 73–83, doi:10.1002/met.63, 2008.

Diro, G. T., Grimes, D. I. F., and Black, E.: Teleconnections between Ethiopian summer rainfall and sea surface temperature: part II. Seasonal forecasting, Clim. Dynam., 37, 121–131, doi:10.1007/s00382-010-0896-x, 2011.

Dutra, E., Di Giuseppe, F., Wetterhall, F., and Pappenberger, F.: Seasonal forecasts of droughts in African basins using the Standardized Precipitation Index, Hydrol. Earth Syst. Sci., 17, 2359–2373, doi:10.5194/hess-17-2359-2013, 2013.

FAO: Irrigation Potential in Africa: A basin approach, Technical report, FAO Land and Water Development Division, Rome, Bulletin No. 4., 1997.

FAO: Drought impact mitigation and prevention in the Limpopo River Basin, A situation analysis, FAO Subregional Office for Southern and East Africa Harare, Rome, Land And Water Discussion Paper, No. 4, 2004.

Funk, C., Hoell, A., Shukla, S., Bladé, I., Liebmann, B., Roberts, J. B., Robertson, F. R., and Husak, G.: Predicting East African spring droughts using Pacific and Indian Ocean sea surface temperature indices, Hydrol. Earth Syst. Sci., 18, 4965–4978, doi:10.5194/hess-18-4965-2014, 2014.

Gimeno, L., Drumond, A., Nieto, R., Trigo, R. M., and Stohl, A.: On the origin of continental precipitation, Geophys. Res. Lett., 37, L13804, doi:10.1029/2010GL043712, 2010.

GRDC (Global Runoff Data Centre), World Meteorological Organization (WMO), Federal Institute of Hydrology (BfG), Koblenz, Germany, available at: http://www.bafg.de/GRDC, last access: 18 December 2016, 2011.

Hall, T. J., Mutchler, C. N., Bloy, G. J., Thessin, R. N., Gaffney, S. K., and Lareau, J. J.: Performance of Observation-Based Prediction Algorithms for Very Short-Range, Probabilistic Clear-Sky Condition Forecasting, J. Appl. Meteorol. Clim., 50, 3–19, doi:10.1175/2010JAMC2529.1, 2011.

Jury, M. R.: Regional teleconnection patterns associated with summer rainfall over South Africa, Namibia and Zimbabwe, Int. J. Climatol., 16, 135–153, doi:10.1002/(SICI)1097-0088(199602)16:2<135::AID-JOC4>3.0.CO;2-7, 1996.

Koster, R. D., Suarez, M. J., and Heiser, M.: Variance and predictability of precipitation at seasonal-to-interannual timescales, J. Hydrometeorol., 1, 26–46, doi:10.1175/1525-7541(2000)001<0026:VAPOPA>2.0.CO;2, 2000.

Kusangaya, S., Warburton, M. L., Archer van Garderen, E., and Jewitt, G. P. W.: Impacts of climate change on water resources in southern Africa: A review, Phys. Chem. Earth, 67–69, 47–54, doi:10.1016/j.pce.2013.09.014, 2014.

Landman, W. A. and Mason, S. J.: Operational long-lead prediction of South African rainfall using canonical correlation analysis, Int. J. Climatol., 19, 1073–1090, doi:10.1002/(SICI)1097-0088(199908)19:10<1073::AID-JOC415>3.0.CO;2-J, 1999.

Landman, W. A., Botes, S., Goddard, L., and Shongwe, M.: Assessing the predictability of extreme rainfall seasons over southern Africa, Geophys. Res. Lett., 32, L23818, doi:10.1029/2005GL023965, 2005.

Lehner, B., Verdin, K., and Jarvis, A.: New global hydrography derived from spaceborne elevation data, T. Am. Geophys. Un., 89, 93–94, 2008.

Li, L., Diallo, I., Xu, C.-Y., and Stordal, F.: Hydrological projections under climate change in the near future by RegCM4 in Southern Africa using a large-scale hydrological model, J. Hydrol., 528, 1–16, doi:10.1016/j.jhydrol.2015.05.028, 2015.

Liaw, A. and Wiener, M.: Classification and Regression by random-Forest, R News, 2, 18–22, available at: https://cran.r-project.org/doc/Rnews/Rnews_2002-3.pdf, last access: 18 Dezember 2016, 2002.

Lindeman, R. H., Merenda, P. F., and Gold, R. Z.: Introduction to Bivariate and Multivariate Analysis, Scott, Foresman, Glenview IL, 1980.

Lorenzo-Lacruz, J., Morán-Tejeda, E., Vicente-Serrano, S. M., and López-Moreno, J. I.: Streamflow droughts in the Iberian Peninsula between 1945 and 2005: spatial and temporal patterns, Hydrol. Earth Syst. Sci., 17, 119–134, doi:10.5194/hess-17-119-2013, 2013.

Love, D., Uhlenbrook, S., Twomlow, S., and van der Zaag, P.: Changing hydroclimatic and discharge patterns in the northern Limpopo Basin, Zimbabwe, Water SA, 36, 335–350, 2010.

Manatsa, D., Chingombe, W., Matsikwa, H., and Matarira, C. H.: The superior influence of Darwin Sea level pressure anomalies over ENSO as a simple drought predictor for Southern Africa, Theor. Appl. Climatol., 92, 1–14, doi:10.1007/s00704-007-0315-3, 2007.

Manatsa, D., Morioka, Y., Behera, S. K., Yamagata, T., and Matarira, C. H.: Link between Antarctic ozone depletion and summer warming over southern Africa, Nat. Geosci., 6, 934–939, doi:10.1038/ngeo1968, 2013.

Masih, I., Maskey, S., Mussá, F. E. F., and Trambauer, P.: A review of droughts on the African continent: a geospatial and long-term perspective, Hydrol. Earth Syst. Sci., 18, 3635–3649, doi:10.5194/hess-18-3635-2014, 2014.

Mason, S. J.: Understanding forecast verification statistics, Meteorol. Appl., 15, 31–40, doi:10.1002/met.51, 2008.

Mekanik, F., Imteaz, M. A., Gato-Trinidad, S., and Elmahdi, A.: Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes, J. Hydrol., 503, 11–21, doi:10.1016/j.jhydrol.2013.08.035, 2013.

Mishra, A. K. and Desai, V. R.: Drought forecasting using stochastic models, Stoch. Env. Res. Risk A., 19, 326–339, doi:10.1007/s00477-005-0238-4, 2005.

Mishra, A. K. and Desai, V. R.: Drought forecasting using feed-forward recursive neural network, Ecological Model., 198, 127–138, doi:10.1016/j.ecolmodel.2006.04.017, 2006.

Mishra, A. K. and Singh, V. P.: A review of drought concepts, J. Hydrol., 391, 202–216, doi:10.1016/j.jhydrol.2010.07.012, 2010.

Mishra, A. K. and Singh, V. P.: Drought modeling – A review, J. Hydrology, 403, 157–175, doi:10.1016/j.jhydrol.2011.03.049, 2011.

Modarres, R.: Streamflow drought time series forecasting, Stoch. Env. Res. Risk A., 21, 223–233, doi:10.1007/s00477-006-0058-1, 2006.

Molteni, F., Stockdale, T., Balmaseda, M., Balsamo, G., Buizza, R., Ferranti, L., Magnunson, L., Mogensen, K., Palmer, T., and Vitart, F.: The new ECMWF seasonal forecast system (System 4), ECMWF, Technical Memorandum No. 656, 2011.

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

Mwale, D., Yew Gan, T., and Shen, S. S. P.: A new analysis of variability and predictability of seasonal rainfall of central southern Africa for 1950–94, Int. J. Climatol., 24, 1509–1530, doi:10.1002/joc.1062, 2004.

Mwale, D., Gan, T. Y., Shen, S. S. P., Shu, T. T., and Kim, K.-M.: Wavelet empirical orthogonal functions of space-time-frequency regimes and predictability of southern Africa summer

rainfall, J. Hydrol. Eng., 12, 513–523, doi:10.1061/(ASCE)1084-0699(2007)12:5(513), 2007.

NCAR – Research Applications Laboratory: verification: Forecast verification utilities, http://cran.r-project.org/package=verification, last access: 18 Dezember 2016, 2012.

Peterson, R. G. and Stramma, L.: Upper-level circulation in the South Atlantic Ocean, Prog. Oceanogr., 26, 1–73, 1991.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, J. Geophys. Res., 108, 4407, doi:10.1029/2002JD002670, 2003.

Reason, C. J. C., Landman, W., and Tennant, W.: Seasonal to Decadal Prediction of Southern African Climate and Its Links with Variability of the Atlantic Ocean, B. Am. Meteorol. Soc., 87, 941–955, doi:10.1175/BAMS-87-7-941, 2006.

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S,. and Schlax, M. G.: Daily 5 High-Resolution-Blended Analyses for Sea Surface Temperature, J. Climate, 20, 5473–5496, doi:10.1175/2007JCLI1824.1, 2007.

Robertson, D. E. and Wang, Q. J.: A Bayesian Approach to Predictor Selection for Seasonal Streamflow Forecasting, J. Hydrometeorol., 13, 155–171, doi:10.1175/JHM-D-10-05009.1, 2012.

Rouault, M.: Intensity and spatial extent of droughts in southern Africa, Geophys. Res. Lett., 32, L15702, doi:10.1029/2005GL022436, 2005.

Rouault, M. and Richard, Y.: Intensity and spatial extension of drought in South Africa at different time scales, Water SA, 29, 489–500, 2003.

Roy-Desrosiers, F.: ANN: Feedforward Artificial Neural Network optimized by Genetic Algorithm, available at: https://cran.r-project.org/src/contrib/Archive/ANN/, last access: 27 November 2016, 2012.

Shukla, S. and Wood, A. W.: Use of a standardized runoff index for characterizing hydrologic drought, Geophys. Res. Lett., 35, L02405, doi:10.1029/2007GL032487, 2008.

Smith, T. M., Reynolds, R. W., Peterson, T. C., and Lawrimore, J.: Improvements to NOAA's Historical Merged Land–Ocean Surface Temperature Analysis (1880–2006), J. Climate, 21, 2283–2296, doi:10.1175/2007JCLI2100.1, 2008.

Tadross, M., Jack, C., and Hewitson, B.: On RCM-based projections of change in southern African summer climate, Geophys. Res. Lett., 32, L23713, doi:10.1029/2005GL024460, 2005.

Thomson, M. C., Abayomi, K., Barnston, A. G., Levy, M., and Dilley, M.: El Niño and drought in southern Africa, The Lancet, 361, 437–438, doi:10.1016/S0140-6736(03)12421-X, 2003.

Trambauer, P., Maskey, S., Werner, M., Pappenberger, F., van Beek, L. P. H., and Uhlenbrook, S.: Identification and simulation of space–time variability of past hydrological drought events in the Limpopo River basin, southern Africa, Hydrol. Earth Syst. Sci., 18, 2925–2942, doi:10.5194/hess-18-2925-2014, 2014.

Trambauer, P., Werner, M., Winsemius, H. C., Maskey, S., Dutra, E., and Uhlenbrook, S.: Hydrological drought forecasting and skill assessment for the Limpopo River basin, southern Africa, Hydrol. Earth Syst. Sci., 19, 1695–1711, doi:10.5194/hess-19-1695-2015, 2015.

Van Beek, L. P. H. and Bierkens, M. F. P.: The Global Hydrological Model PCR-GLOBWB: Conceptualization, Parameterization and Verification, PhD thesis, Utrecht University, Utrecht, The Netherlands, 2009.

van Dijk, A. I. J. M., Peña-Arancibia, J. L., Wood, E. F., Sheffield, J., and Beck, H. E.: Global analysis of seasonal streamflow 35 predictability using an ensemble prediction system and observations from 6192 small catchments worldwide, Water Resour. Res., 49, 2729–2746, doi:10.1002/wrcr.20251, 2013.

Van Loon, A. F. and Van Lanen, H. A. J.: Making the distinction between water scarcity and drought using an observation-modeling framework, Water Resour. Res., 49, 1483–1502, doi:10.1002/wrcr.20147, 2013.

Vicente-Serrano, S. M., Beguería, S., Gimeno, L., Eklundh, L., Giuliani, G., Weston, D., El Kenawy, A., López-Moreno, J. I., Nieto, R., Ayenew, T., Konte, D., Ardö, J., and Pegram, G. G. S.: Challenges for drought mitigation in Africa: The potential use of geospatial data and drought information systems, Appl. Geogr., 34, 471–486, doi:10.1016/j.apgeog.2012.02.001, 2012a.

Vicente-Serrano, S. M., López-Moreno, J. I., Beguería, S., Lorenzo-Lacruz, J., Azorin-Molina, C., and Morán-Tejeda, E.: Accurate computation of a streamflow drought index, J. Hydrol. Eng., 17, 318–332, doi:10.1061/(ASCE)HE.1943-5584.0000433, 2012b.

Wetterhall, F., Winsemius, H. C., Dutra, E., Werner, M., and Pappenberger, E.: Seasonal predictions of agro-meteorological drought indicators for the Limpopo basin, Hydrol. Earth Syst. Sci., 19, 2577–2586, doi:10.5194/hess-19-2577-2015, 2015.

Wilhite, D. A. and Hayes, M. J.: Planning for drought: Moving from crisis to risk management[1], J. Am. Water Resour. As., 36, 697–710, doi:10.1111/j.1752-1688.2000.tb04299.x, 2000.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Elsevier Inc., 2nd Edn., 2006.

Winsemius, H. C., Dutra, E., Engelbrecht, F. A., Archer Van Garderen, E., Wetterhall, F., Pappenberger, F., and Werner, M. G. F.: The potential value of seasonal forecasts in a changing climate in southern Africa, Hydrol. Earth Syst. Sci., 18, 1525–1538, doi:10.5194/hess-18-1525-2014, 2014.

Yuan, Y. and Li, C.: Decadal variability of the IOD-ENSO relationship, Chinese Science Bulletin, 53, 1745–1752, doi:10.1007/s11434-008-0196-6, 2008.

Zhu, T. and Ringler, C.: Climate Change Implications for Water Resources in the Limpopo River Basin, IFPRI Discussion Paper 961, 2010.