Hydrological drought forecasting and skill assessment for the Limpopo river basin, southern Africa

P. Trambauer[1], M. Werner[1,2], H.C. Winsemius[2], S. Maskey[1], E. Dutra[3], S. Uhlenbrook[1,4]

[1] UNESCO-IHE, Department of Water Science and Engineering, P.O. Box 3015, 2601 DA Delft, The Netherlands

[2] Deltares, P.O. Box 177, 2600MH Delft, The Netherlands

[3] ECMWF, Shinfield Park, RG2 9AX Reading, United Kingdom

[4] Delft University of Technology, Water Resources Section, P.O. Box 5048, 2600 GA Delft, The Netherlands

Correspondence to: P. Trambauer (p.trambauer@unesco-ihe.org)

Abstract

Ensemble hydrological predictions are normally obtained by forcing hydrological models with ensembles of atmospheric forecasts produced by numerical weather prediction models. To be of practical value to water users, such forecasts should not only be sufficiently skilful, they should also provide information that is relevant to the decisions end users make. The semi-arid Limpopo basin in southern Africa has experienced severe droughts in the past, resulting in crop failure, economic losses and the need for humanitarian aid. In this paper we address the seasonal prediction of hydrological drought in the Limpopo river basin by testing three proposed forecasting systems (FS) that can provide operational guidance to dam operators and water managers at the seasonal time scale. All three FS include a distributed hydrological model of the basin, which is forced with either; (i) a global atmospheric model forecast (ECMWF seasonal forecast system - S4), (ii) the commonly applied Ensemble Streamflow Prediction approach (ESP) using resampled historical data, or (iii) a conditional ESP approach (ESPcond) that is conditional on the ENSO signal. We determine the skill of the three systems in predicting streamflow and commonly used drought indices. We also assess the skill in predicting indicators that are meaningful to local end users in the basin. FS_S4 shows moderate skill for all lead times (3, 4, and 5 months) and aggregation periods. FS_ESP also performs better than climatology for the shorter lead times, but with lower skill than FS_S4. FS_ESPcond shows intermediate skill compared to the other two FS, though its skill is shown to be more robust. The skill of FS_ESP and FS_ESPcond is found to reduce rapidly with increasing lead time when compared to FS_S4. The results show that both FS_S4 and FS_ESPcond have good potential for seasonal hydrological drought forecasting in the Limpopo river basin, which is encouraging in the context of providing better operational guidance to water users.

1. Introduction

Climate change studies show evidence of an intensification of the global water cycle (Huntington, 2006; IPCC, 2007; Hansen et al., 2012; Trenberth, 2012; Coumou and Rahmstorf, 2012), with extreme events including floods and droughts expected to become more frequent. The UNISDR Hyogo Framework of Action 2005-2015 (UNISDR, 2005) describes early warning systems and action plans triggered on the issuing of a warning as one of the most effective strategies to mitigate the impacts of natural hazards. Operational forecasting of streamflow to inform early warning is already commonplace in several parts of the world, but the main focus is often on flood prediction. Operational forecasting of streamflow for drought prediction has to date not been applied as widely, despite the widespread recognition of the relevance and importance of drought forecasting in the research community.

There are several Drought Early Warning Systems (DEWS) currently in existence in the world, though due to the complexity of drought these are arguably less developed than many flood early warning systems. Grasso (2009) reports that only three institutions provide information on the occurrence of major droughts at the global scale; FAO's Global Information and Early Warning System on Food and Agriculture (GIEWS), the Humanitarian Early Warning Service (HEWS) operated by the World Food Programme (WFP), and the Benfield Hazard Research Centre at University College London.

In the United States the U.S. Drought Monitor (http://droughtmonitor.unl.edu/) was set up in collaboration between the US Department of Agriculture (USDA), NOAA, the Climate Prediction Centre, and the University of Nebraska. It provides insight to current drought conditions and impacts at the national and state level through an interactive map, presenting multiple drought indicators combined with field information and expert input. It also includes 6- to 10 day outlooks and monthly and seasonal forecasts of precipitation, temperature, soil moisture and streamflow. The National Weather Service's National Center for Environmental Prediction's (NCEP) also has a (multi-model) drought monitoring system, as well as a seasonal hydrological forecasting system running at the Environmental Modeling Center (Ek et al., 2010). Additionally, the North American Multi-Model Ensemble (NMME), which became an experimental real-time system in August 2011, is mainly focused on seasonal prediction of meteorological drought (Kirtman et al., 2013).

In Europe the European Commission Joint Research Centre (JRC) has established the European Drought Observatory (EDO, http://edo.jrc.ec.europa.eu/), which includes an interactive map viewer with drought-relevant information. It includes real-time maps of different drought indicators, including the Standardized Precipitation Index (SPI), snow and soil moisture anomaly, and vegetation productivity anomaly. These indicators are combined in an overall indicator that is used to provide warnings and alerts. A one week forecast of the expected soil moisture anomaly is also provided. The Beijing Climate Center (BCC) of the China Meteorological Administration (CMA) similarly monitors the development of drought across China, with maps on current drought conditions being updated daily on their website.

The FEWS Net for Eastern Africa, Afghanistan, and Central America reports on current famine conditions, including droughts, by providing monthly bulletins that are accessible on the FEWS Net webpage. However, a drought forecast is not provided. Other drought warning systems over Africa include the Botswana national early warning system (EWS) for drought (Morgan, 1985) and the Regional Integrated Multi-Hazard Early Warning System for Africa and Asia (RIMES). In the latter a drought early warning system is being adapted to identify climate and water supply trends in order to detect the probability and potential severity of drought (RIMES, 2014).

Advances regarding drought early warning systems in Africa in the last few years are remarkable. There is an increasing availability of drought monitoring and forecasting tools for decision making that can provide real time monitoring and forecasting of drought across the continent. The Land Surface Hydrology Group at Princeton University, USA, has recently established an African Flood and Drought Monitor (http://stream.princeton.edu/) with support from the International Hydrology Program of UNESCO. The system provides near real time monitoring of land surface hydrological conditions based on the Variable Infiltration Capacity (VIC) model. The monitor is updated every day at 2 days behind real time. The database provides the daily conditions of precipitation, temperature, wind speed, soil moisture, evaporation, radiation, and different components of runoff in the continent, as well as historic hydrological records in Eastern, Southern and Western sub-regions up to 10 antecedent years, and derived products such as current drought conditions. They also provide precipitation, temperature and SPI forecast (Sheffield et al., 2014). Recently Barbosa et al. (2013) developed a Pan-African map viewer for drought within the framework of the DEWFORA project, following the main features of the earlier developed EDO. The African Drought Observatory (ADO) is a web application hosted by JRC (http://edo.jrc.ec.europa.eu/ado/ado.html) that provides historical and near-real time monitoring information, as well as seasonal forecasts describing meteorological, agricultural and hydrological droughts (Barbosa et al., 2013).

Yuan et al. (2013) applied the NCEP's Climate Forecast System version 2 (CFSv2) combined with the Variable Infiltration Capacity (VIC) land surface model for seasonal drought prediction over Africa. They used both the Standardized Precipitation Index (SPI) and soil moisture as indices and the Brier Skill Score (BSS) to assess the probabilistic drought hindcast for 1982-2002. Their results show relatively good skill in the dry season but only limited skill in the rainy season. They indicate that CFSv2 precipitation is correlated with the observed precipitation over southern Africa, but only accounts for 44-45% of the variance of observations. They point out that for two extreme droughts CFSv2 predicted neutral conditions or only a weak anomaly. Our study focuses on the Limpopo river basin and follows a similar type of analysis, although it does so at a higher resolution and in a more detailed manner. Additionally we present different skill scores for different hydrological drought indicators during the rainy season, and compare different forecasting systems in the basin. We focus on assessing the skill of the forecast in predicting indicators that are meaningful to the local end users in the basin.

The semi-arid Limpopo river basin, located in southern Africa, has experienced severe droughts in the past, which have led to crop failures, high economic losses and the need for humanitarian aid. An effective drought early warning system for this basin is of prime importance. Current practices for drought forecasting in the Limpopo river basin involve three forms of seasonal climate forecasts ranging from regional to local scale: The Southern Africa Regional Climate Outlook Forum (SARCOF) climate outlooks, seasonal climate outlooks prepared by meteorological departments, and forecasts based on local knowledge applied in rural communities. Despite these seasonal forecasts being available in the basin, farmers seem to prefer to rely on drought forecasting systems based on indigenous and traditional knowledge. Such forecasts include signs in (i) the sun, moon and wind; (ii) trees and plants; (iii) and insects, birds and animals (DEWFORA, 2013). For seasonal forecasts to be accepted by the local community there are several challenges that need to be addressed. End users should receive the information in a suitably understandable format at the time they need it for the forecast to be useful. The highly technical information that is typically contained in the forecasts should then be translated to a comprehensible form before being disseminated and delivered to decision makers and farmers. Moreover, end users should be involved in the product verification by providing feedback to the forecasters (DEWFORA, 2012).

Seasonal hydrological drought forecasts aim for high hydrological predictability at a seasonal time scale. Shukla et al. (2013) quantified the contribution of a good representation of initial hydrologic conditions (IHCs) and seasonal meteorological forecast (MF) to seasonal hydrological predictability at different forecast dates and lead times (1, 3, and 6 months) globally. They quantified the contributions of two components of the IHCs (soil moisture and snow water content) through Ensemble Streamflow Prediction (ESP) and Reverse-ESP. Their results show that for the region of the Limpopo river basin the MF dominates the hydrological predictability during the wet season (forecasts starting in October and January) for almost every lead time considered. Only for the 1-month lead time forecasts issued in October the IHCs appeared to some extent to have a higher influence. For the dry season the IHCs dominate the hydrological forecast at all lead times. These results suggest that to improve the seasonal hydrologic forecast skill in the Limpopo river basin, efforts should focus on improving the MF. However, the contribution of other IHCs (surface water and groundwater level) to hydrological predictability should also be assessed.

Yossef et al. (2013) also investigated the relative contribution of initial conditions and meteorological forcing to the skill of the global seasonal streamflow forecasting system FEWS-World, using the global hydrological model PCR-GLOBWB. They use ESP and Reverse-ESP to determine the critical lead time for different locations at which the importance of the initial conditions is surpassed by that of the meteorological forcing. They indicate that for semi-arid regions such as the Limpopo basin the initial conditions do not contribute much to the skill given the high sensitivity of the runoff coefficient to rainfall variability. This would suggest that the predictability in semi-arid basis such as the Limpopo using ESP is limited, with seasonal meteorological forecasts potentially offering better skill.

In this study we introduce three dynamic forecasting systems based on a distributed hydrological model for the seasonal prediction of hydrological droughts for the semi-arid Limpopo basin in southern Africa. All three forecasting systems include a distributed hydrological model of the basin, and are forced by either; (i) a global atmospheric model (ECMWF seasonal forecast system S4), (ii) the Ensemble Streamflow Prediction approach (ESP) using resampled historical data, or (iii) a conditional ESP approach (ESPcond) that is conditioned on the ENSO signal. The aim of this study is to assess the skill of the three systems in predicting meaningful drought indices for the Limpopo basin.

2. Methods and data

The approach followed in this study is summarised in Fig. 1. It starts with obtaining the meteorological seasonal forecast and pre-processing the data. This is then used to force the hydrological model (embedded in the Delft-FEWS forecasting shell (Werner et al., 2013)), thus obtaining seasonal forecasts of streamflow and other hydrological variables.

2.1 Ensemble hydrological forecasting in the Limpopo river basin

2.1.1 Study area - Limpopo river basin

The Limpopo river basin is one of the larger basins in southern Africa, with a drainage area of approximately 415,000 km^2. It is shared by four riparian countries (see Fig. 2); South Africa (45%), Botswana (20%), Mozambique (20%) and Zimbabwe (15%). The climate in the basin is quite diverse. The upper part of the basin lies in the Kalahari Desert and is particularly arid. Towards the Indian Ocean the climate then changes to a hot dry steppe and finally to a tropical dry savannah. In the mountainous regions the climate is markedly cooler. Rainfall in the basin is seasonal, influenced by the movement of the intertropical convergence zone. Moreover, rainfall is highly variable causing frequent droughts, though floods can also occur during the rainy season. In the period 1980-2000, the southern African region was stuck by four major droughts in the seasons 1982/83, 1986/87, 1991/92 and 1994/95. This corresponds to an average frequency of a drought every four or five years, although the periodicity of droughts is not necessarily predictable. It is estimated that during the 1991/92 drought in southern Africa 86 million people were affected, 20 million of whom were considered to be at serious risk of starvation (DEWFORA, 2011).

The annual rainfall in the basin averages some 530 mm yr^{-1}, though the spatial variation is significant, ranging from 200 to 1,200 mm yr^{-1}. Rainfall occurs mainly in the austral summer months (October to April) (LBPTC, 2010). As is common with semi-arid and arid basins, runoff coefficients in the Limpopo are very low, being only 4.3% for the naturalised discharge and a mere 1.7% for the observed discharge at Chókwe station in Mozambique, which is the most downstream station considered in this study. If abstractions are included then only 23 mm yr^{-1} of the 539 mm yr^{-1} precipitation average for the basin upstream of Chókwe runs off. Consequently, hydrological modelling in the Limpopo basin is extremely challenging. Even a small error in precipitation or evaporation estimates could result in quite a large error in runoff estimation. Moreover, the uncertainty in the rainfall input could easily be larger than the runoff coefficient (4.3%) of the basin. Fig. 3 shows the location of selected runoff stations and reservoirs in the Limpopo basin.

2.1.2 The forecasting system

Regional hydrological model

A finer resolution version (0.05 x 0.05°) of the 0.5 x 0.5° resolution global PCR-GLOBWB hydrological model is used. This is a continuous-time simulation, process based distributed model applied on a raster basis. PCR-GLOBWB is in many ways similar to other global hydrological models, but it has several improved features, such as improved schemes for sub-grid parameterization of surface runoff, interflow and baseflow, a kinematic wave based routing for the surface water flow, dynamic inundation of floodplains and a reservoir scheme (van Beek and Bierkens, 2009; van Beek, 2008). The model is set up for the Limpopo basin with a spatial resolution of 0.05 x 0.05° and the simulation is carried out with a daily time step. As the scope of this study is on the skill of the hydrological forecast, reservoirs are considered in a simple way. Cells with reservoirs in the model are considered as having a maximum storage volume. Releases to irrigation are taken into consideration as a fixed monthly value and subject to availability, and the reservoir will spill when full. The reservoirs in the basin are mainly used for irrigation. For a more detailed description of the model set up for the Limpopo river basin the reader is referred to Trambauer et al. (2014b).

Delft-FEWS shell

The hydrological model is embedded in the Delft-FEWS (Flood Early Warning System) open shell for forecasting purposes. The shell provides a sophisticated collection of modules designed for building a hydrological forecasting system customised to the specific requirements of an individual organisation. The philosophy is to provide an open shell for managing the data handling and forecasting process. This shell incorporates a comprehensive library of general data handling utilities, allowing a wide range of external models to be integrated in the system through a published open interface. This allows existing simulation models and data streams to be incorporated into a comprehensive and reliable forecasting system (Werner et al., 2013).

Reference run for the period 1979-2010

The hydrological model is run in simulation mode for a 32 year period (1979-2010) with the ERA-Interim forcing meteorological data at a daily time step. ERA-Interim (ERAI) is the latest global atmospheric reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) and covers the period from January 1979 to the present date with a horizontal resolution of approximately 0.7 degrees and 62 vertical levels. A complete description of the ERAI product is available in Dee et al. (2011). The ERA-Interim precipitation data used in this study was corrected using the GPCP v2.1 product of the Global Precipitation Climatology Project to reduce the bias when compared to measured products (Balsamo et al., 2010). The GPCP v2.1 data is a monthly climatology provided globally at 2.5˚ × 2.5˚ resolution, covering the period from 1979 through to September 2009. It combines the precipitation data available from several sources (satellite data, rain gauge data, etc.) into a merged product (Huffman et al., 2009; Szczypta et al., 2011). From September 2009 to December 2010, the mean monthly ERAI precipitation was corrected using a mean bias coefficient based on the climatology of the bias correction coefficients that were established for the period 1979-2009. While this only corrects for systematic biases, this was the only option available at the time, as a new version of GPCP (version 2.2) was not available. This corrected version of precipitation was also used in the production of the ERA-Interim/Land data set (Balsamo et al., 2013).

Additional to the precipitation, other meteorological parameters from the ERA-Interim reanalysis data that are used to force the model include the 2 metre daily temperature (minimum, maximum and average). Temperature data is used for the computation of the reference potential evaporation that is required to force the hydrological model. In this study the Hargreaves formula is used. This method requires less parameterization than the Penman-Monteith formula, though it has the disadvantage that it is less sensitive to (uncertain) climatic input data, with a possible reduction of the dynamics and accuracy of the potential evaporation as a consequence. However, this also means that it is less sensitive to errors in climatic inputs (Hargreaves and Allen, 2003) that are inherent to any meteorological forecast. Moreover, the choice of the method used for the computation of the reference potential evaporation was shown to have minor effects on the results of the actual evaporation for southern Africa, where actual evaporation is dominated by soil moisture availability (Trambauer et al., 2014a). The ERA-Interim data for the 32-year period from 1979 to 2010, corrected using the GPCP v2.1 dataset are converted to the same spatial resolution as the continental-scale version of the PCR-GLOBWB model. ERAI is archived on an irregular grid (reduced Gaussian) with an approximate resolution of 0.7° over the domain. The data is downscaled from the ERAI grid to the original 0.5˚ model grid using bilinear interpolation and assumed to be constant over the 0.5˚ grid cell. No further downscaling of the meteorological forcings is carried out.

Initial conditions

The reference run provides the initial conditions for all forecasts. Initial conditions at the beginning of each month are saved in the Delft-FEWS data base, and subsequently used as "warm states" to start the forecasts when doing the retroactive forecast (also referred to as hindcasts).

Time period of the simulations

An ensemble of meteorological hindcasts is first tested for the summer rainfall season over southern Africa for the period 1981 to 2010. Seasonal forecasts in this study are issued for only seven months of the year so as to capture the rainy season and main runoff season (meaning the there are five months where we do not issue a forecast). The predictive skill for drought is expected to be higher during the dry season and lower during the wet season given that the hydro-climate has longer persistence during the dry season (Yuan et al., 2013). Yuan et al. (2013) show the high contrast in skill between the dry and wet seasons in southern Africa.

In the hindcast, the first forecast of each season is issued in August and includes the seasonal (6-months) forecast from August to January. The forecast is updated at the beginning of each month from September to February. The last forecast of the season is issued in February, covering the period from February to July (see Fig. 4). All simulations are done at a daily time step.

2.2 Seasonal forecasting systems

All three forecasting systems considered use the same hydrological model of the basin, but are forced with different meteorological forecasts. In the first system (FS_S4) the PCR-GLOBWB hydrological model is forced with the output of a global atmospheric model, the ECMWF seasonal forecast system S4 (atmosphere-ocean coupled). The second forecasting system (FS_ESP) is based on the Ensemble Streamflow Prediction (ESP, Day (1985)) procedure. In the ESP procedure the ensemble meteorological forecast is generated with re-sampled historical meteorological data. The hydrological model is then forced with this re-sampled data. A third system (FS_ESPcond) is proposed given that the El Niño-Southern Oscillation (ENSO) has a clear influence on the inter-annual climate variability over the Limpopo river basin (Landman and Mason, 1999). This is equivalent to the second system but the weights of the ESP ensemble members are conditioned on the ENSO signal (Oceanic Niño Index, ONI). This is explained in full in section 2.2.3.

2.2.1 ECMWF S4 meteorological forecasts (FS_S4)

Meteorological ensemble forecasts

Seasonal meteorological forecasts from the most recent seasonal forecasting system at ECMWF (S4) are used to force the hydrological model. The S4 ensemble seasonal forecasts are initialised on the 1st of each month and the ensemble is generated by perturbations in the initial conditions and by the use of stochastic physics in the atmosphere during the model integration (out to 6 months lead time) (Molteni et al., 2011). The atmospheric resolution is about 79 km with 91 vertical levels, and is fully coupled with an ocean model with a horizontal resolution of 1°. S4 has been in operational use since November 2011, issuing 51 ensemble members with six months lead time. A hindcast set is provided for calibration and verification purposes, covering a period of 30 years (1981 to 2010) with the same configuration as the operational forecasts but with only 15 ensemble members. Molteni et al. (2011) presents an overview of S4 model biases and forecasts performance, and Dutra et al. (2013) and (2014) present an evaluation of S4 in seasonal forecasts of meteorological droughts. They found that S4 derived meteorological drought forecasts over southern Africa were skilful up to four months lead time for SPI-6 in April. In the setup of FS_S4, the hydrological model is forced with the re-forecasts of the ECMWF seasonal system S4, with 15 ensemble members. A (hydrological) re-forecast is made to coincide with the 1st of each month in the 30 year hindcast set. Precipitation inputs to the hydrological model are accumulated from the 6-hourly S4 model values, while evaporation was calculated using the daily maximum and minimum temperatures directly archived by the meteorological model.

Climatological bias correction of seasonal forecasts of precipitation

Mean biases and drifts in the seasonal forecasts of precipitation can have a detrimental influence on the hydrological forecasts. Therefore, a simple climatological bias correction, based on monthly means, is applied to the seasonal forecasts in the form:

$$P'_{m,l} = \alpha_{m,l}\, P_{m,l} \qquad (1)$$

where P and P' are the original and corrected seasonal forecasts of precipitation respectively, α is a multiplicative correction factor and the subscripts m and l are the calendar month (1 to 12, of the initial forecast date) and lead time (0 to 5 months) respectively. The correction factor is given by the ratio:

$$\alpha_{m,l} = \overline{P}_{m*}^{base} / \overline{P}_{m,l} \qquad (2)$$

where $\overline{P}_{m*}^{base}$ is the climatological long term mean of precipitation of the base dataset for a particular calendar month $m*$ ($m*=m+l$), and $\overline{P}_{m,l}$ is the long term ensemble mean of the forecasts for a particular month m and lead time l. The base data set used was ERA-Interim corrected with GPCP to be consistent with the baseline simulation period. The correction factor α is limited to a reasonable range (0.1 and 10), and is linearly interpolated from monthly values to daily values by assuming that it corresponds to day 15 of the particular month. Equation (1) is applied to the daily precipitation values. This is a simple bias correction that only guarantees that the mean forecast climate is similar to the climate of the base data set. It does not address other problems of the forecasts, common to all coupled atmosphere-ocean models, such as inter-annual variability, ensemble spread or daily variability.

2.2.2 ESP meteorological forecasts (FS_ESP)

A widely used approach to seasonal forecasting is the Ensemble Streamflow Prediction (ESP) procedure. ESP predicts future streamflow from the current initial conditions (warm state) in the hydrological model with re-sampled historical meteorological data (ERA-Interim corrected with GPCP observed meteorology from the last 31 years in this study). The procedure assumes that meteorological events that occurred in the past are representative of events that may occur in the future (Day, 1985). Although ESP is normally used in the absence of a seasonal forecast, in this study we use it to compare the skill of the FS_ESP with that of the FS_S4. Moreover, a comparison of these two forecasts may give an indication of what influences the predictability. ESP represents forecast uncertainty due to boundary forcing uncertainties only (Wood and Lettenmaier, 2008) and thus allows measuring the skill that can be expected only from initial states. In the FS_ESP hindcast, the sample of the year in which the forecast starts is excluded from the ensemble to allow for a fair estimate of the forecast uncertainty. The FS_ESP therefore includes 30 (31 minus 1) years in the ensemble.

2.2.3 Conditional ESP meteorological forecasts (FS_ESPcond)

The El Niño-Southern Oscillation (ENSO) is clearly related to inter-annual climate variability over the Limpopo river basin. In southern Africa meteorological droughts tend to happen in the December to March rainy season after onset of an El Niño event (Thomson et al., 2003). However, it is not always the case that this happens. Thomson et al. (2003) recorded a 120% increase in probability of drought disaster in the year after an El Niño onset. To account for the relationship between ENSO and the occurrence of drought, this system is similar to FS_ESP but the weights of ensemble members sampled through the ESP procedure are conditioned on the ENSO signal.

We use the post-ESP weighting technique described in Werner et al. (2004). This approach uses the El Niño - 3.4 index averaged over the 3 month-period immediately prior to the issue date of the forecast to weight ensemble members from ESP. The technique is summarized here for the forecast of the six month Standardised Runoff Index (SRI-6):

1) Compute a vector (**X**) of absolute differences (x_i) between the value of the Niño - 3.4 index (Oceanic Niño Index, ONI) in the forecast year and those of all the other years and sort the vector (**X**) from lowest to highest.

$$\boldsymbol{X} = (x_1, x_2, \dots, x_n) \tag{3}$$

The sorted vector (ℵ) is,

$$\aleph = \left[x_{(1)}, x_{(2)}, \dots, x_{(n)}\right],\ x_{(1)} \leq\ x_{(1)} \dots \leq x_{(n)} \tag{4}$$

2) Compute a vector of weights (**W**) for each member of the ESP ensemble by defining two parameters: a distance-sensitive weighting parameter (λ) and a parameter (α) that defines the k nearest neighbours used to calculate the weight of each member. Higher λ gives more weight to ensemble members with values of ONI closer to that of the forecast year. Higher α restricts attention to the n/α elements in the sorted vector. The ensemble member with the same year as the forecast year is assigned a weight of zero.

$$W = (w_1, w_2, \dots, w_n) \tag{5}$$

$$w_i = \left[1 - \frac{x_{(i)}}{x_{(k)}}\right]^{\lambda - 1},\ x_{(i)} \leq\ x_{(k)} \tag{6}$$

$$w_i = 0,\ x_{(i)} >\ x_{(k)} \tag{7}$$

$$k = NINT\left(\frac{n}{\alpha}\right) \tag{8}$$

3) Calculate the probability (p_i) assigned to each ensemble member *i* by rescaling the weights.

$$p_i = \frac{w_i}{\sum_{j=1}^{n} w_j} \tag{9}$$

The parameters λ and α can be optimised for each case study or sub-basin. The case with $\lambda = \alpha = 1$ is the traditional equal weighting scheme applied to ESP forecasts, with all ensemble members considered to have equal weight. If $\alpha = 1$ and λ varies, all ensemble members are considered, but these have non-zero weights that depend on the absolute distance between the ONI of the forecast year and the ONI of the year of the ensemble member. If $\lambda = 1$ and α varies only the nearest k ensemble members to the forecast year are considered in the ensemble, but they are all weighted equally. This case is similar to the approach applied by Hamlet and Lettenmaier (1999) for the Columbia river, where they restricted the ensemble members to those years that were similar in terms of the ENSO phase and the Pacific decadal oscillation phase. However, this restriction may result in ensembles with only few members, resulting in forecasts that are very sensitive to sampling errors (Brown et al., 2010). In the last case, where both α and λ vary, weights are assigned only to the k nearest ensemble members based on the distance of the index to the index of the forecast year (Werner et al., 2004). Werner et al. (2004) found this last case where both α and λ vary to show the best improvements for forecast skills.

For the FS_ESPcond we chose to keep the parameters constant ($\lambda = 2$ and $\alpha = 1$) given that the optimal selection of parameters would vary for each sub-basin. Performing an in-depth selection of parameters for each sub basin is out of the scope of this study. Here we use $\lambda = 2$ and $\alpha = 1$, meaning that all ensemble members have a non-zero probability of being included in the ensemble, with that probability based on the distance between the ENSO indexes and the distance sensitive weighting parameter (linear for $\lambda = 2$). For each forecast start date, we construct an ensemble meteorological forecast of 30 members to be consistent with FS_ESP. The selection of the members is based on a resampling with replacement procedure given the probability assigned to each member. From the 30 possible ensemble members to be included, those with an ONI index closer to that of the forecast year, have a higher probability of being included in the ensemble. This means that some ensemble members are included more than once, and some are not included at all. The ONI indexes for the period 1979 - 2010 were retrieved from NOAA (2014).

We also use this procedure for the forecast of SRI-4 (JFMA SRI). FS_ESPcond always uses the latest ONI index available prior to the start date of the forecast. This means that for the forecast issued in January, which corresponds to a three months lead time, FS_ESPcond uses the ONI values for October, November and December (OND). Similarly, for the forecast issued in December, which corresponds to a four months lead time, FS_ESPcond uses the SON ONI, and the forecast issued in November (five months lead time) makes use of the ASO ONI.

2.3 Assessing skill of the forecasts

2.3.1 Skill scores

Standard verification skill scores are selected to measure the skill of the forecast ensembles in predicting drought indicators. In this study we use the Standardized Runoff Index (SRI) for the characterization of hydrological droughts. This indicator is explained in the following section. Forecasts are verified against the reference run and the resulting skills are established relative to sample climatology. Cloke and Pappenberger (2008) recommend the use of several verification measures in the same analysis so that the quality of the forecast can be assessed rigorously. We selected three verification scores that measure slightly different properties of the forecast skill. The ROC curve measures discrimination but not bias, the rank histogram measures reliability or bias, and the brier score (BS) accounts both for reliability and sharpness (Renner et al., 2009).

The **ROC** (relative operating characteristic, or receiver operating characteristic) diagram measures the ability of the forecast to discriminate between two alternative outcomes. It plots the hit rate or probability of detection (POD) versus the false alarm rate or probability of false detection (POFD). It is not sensitive to bias in the forecast, so says nothing about the reliability. It is conditioned to the observations. In summary, it indicates the ability of the forecast to discriminate between events and non-events given a certain event threshold (WWRP/WGNE, 2013). The area under a ROC curve (**ROCS**) is used as a score. ROCS can take values from 0 to 1, with a value of 0.5 indicating no skill and a value of 1 representing a perfect score. Values lower than 0.5 indicate negative skill. ROC curves measure how good forecasts are in the context of a very simple decision-making model, and are thus better suited to measure how good forecasts are from the perspective of the user than many other commonly used measures (Tveito et al., 2008).

The **Brier Score** (BS [0-1]) measures the mean squared probability error and represents the magnitude of the probability forecast errors, with a perfect score of zero. The **Brier Skill Score** (BSS [-∞ to 1]) measures the improvement of the probabilistic forecast relative to sample climatology and indicates what the relative skill of the probabilistic forecast is over that of the climatology, in terms of predicting whether or not an event occurred (WWRP/WGNE, 2013).

The **rank histogram** is used to evaluate whether the forecast ensembles are from the same underlying population as the observations, which implies that the observed would have the same probability of occurrence as any of the ensemble member. This would result in a uniform distribution in histogram that plots the frequency of the rank of the observation in the ensemble, while deviations from the uniform distribution reveal deficiencies in ensemble calibration, or reliability (Wilks, 2011).

2.3.2 Standardized Runoff Index (SRI)

The hydrological drought indicator SRI follows the same concept as the Standardized Precipitation Index (SPI) and is defined as a "unit standard normal deviate associated with the percentile of hydrologic runoff accumulated over a specific duration" (Shukla and Wood, 2008). To compute SRI the runoff time series is fitted to a probability density function (a gamma distribution) and the function is used to estimate the cumulative probability of the runoff of interest for a specific month and temporal scale. The cumulative probability is then transformed to the standardised normal distribution with mean zero and variance one (Shukla and Wood, 2008).

2.3.3 Skill assessment

Forecasted streamflow is transformed to the hydrological drought indicator SRI and forecasts of drought are analysed by considering drought conditions to occur for SRI $\leq$ -0.5 (mild to moderate drought). The value of -0.5 was chosen as it corresponds to the 30th percentile in runoff and it is therefore a good compromise between not capturing all negative anomalies and having a sufficient amount of samples for the analysis. The forecasting system is thus evaluated on the skill of predicting SRI falling below the -0.5 threshold.

However, as we also want to analyse the ability of the system to forecast distributed variables (for agricultural droughts) and water levels in the reservoirs (for irrigation curtailments), we also evaluated the skill of the forecast system in predicting these variables.

2.3.4 Estimating uncertainty in the skill scores

Given the small sample size resulting from applying the verification over the 30 year hindcast period, a bootstrap approach is used to estimate the confidence intervals around the ROCS. The idea behind the bootstrap is to treat a finite sample at hand as similarly as possible to the unknown distribution from which it was drawn, which in practice leads to resampling with replacement (Wilks, 2011). The uncertainty of the ROCS is estimated by applying a bootstrap resampling with replacement procedure.

For the FS_S4 and FS_ESP forecasts, we randomly replace (allowing repetition) the original forecast and verification pair to produce a new sample of the same size as our original sample. We then calculate the ROCS from the new sample. We repeat this procedure to create 1000 new samples from which we generate an empirical distribution of the ROCS. The 90% confidence interval is estimated from the 5th and 95th percentiles of this empirical distribution.

For the FS_ESPcond the bootrstrap procedure follows the same theory but is computed slightly differently. In this case the bootstrap is achieved by recreating the ensemble forecasts for the hindcast period 1000 times based on the computed probability vector and computing the skill score from each created ensemble.

A limitation of this bootstrap procedure is that statistics computed from discrete bootstrap samples may differ from the ones based on continuous data, and this might lead to overestimation of the confidence. However, this method is widely used in the literature (Dutra et al., 2014; Friederichs and Thorarinsdottir, 2012; Wilks, 2011) to estimate confidence intervals as it does not require assumptions on the distribution.

2.4 Assessing spatial hydrological indicators

ROCS and BSS are computed for the spatially distributed indicator Root Stress (RS) to assess the skill of the forecast in predicting agricultural drought indicators. The RS is an indicator of the available (or the lack of) soil moisture in the root zone, which can be calculated for each grid cell. The RS varies from 0 to 1, where 0 indicates that the soil water availability in the root zone is at field capacity and 1 indicates that the soil water availability in the root zone is zero and the plant is under maximum water stress. For each grid cell, a drought is defined to occur when the Root stress is higher than the 70th percentile of the observed values for that month. An advantage of defining the threshold as a percentile of the observed sample as proposed by Roulin (2007) is that it assures a sufficiently large enough number of events to verify and also allows for comparison of verification statistics at different locations (Renner et al., 2009).

Additional to indicators such as RS, it is interesting to evaluate the skill of the model in predicting indicators that are meaningful to the end users in the basin. Irrigation is the major water user in the Limpopo basin. The amount of water made available to the irrigation sector may, however, be restricted depending on the water level in the reservoirs in the basin as a percentage of their full capacity (DWA, 2013). The forecasted anomaly of the water level in the reservoir is a decision variable that can give an indication to the water managers of the percentage of irrigation demand that can be covered during the season.

An analysis of the historical time series of water level for Tzaneen reservoir together with the curtailment rules of the reservoir (DWA, 2013) indicate that a 20% curtailment to the irrigation sector is applied when water levels in the reservoir fall below the 50th percentile in the water levels (in % of the capacity of the reservoir). Similarly, a 65% curtailment to the irrigation sector is applied when water levels in the reservoir fall below the 37.5th percentile and a 90% curtailment in the irrigation sector along with a 30% curtailment in the urban sector when the water levels are below the 12th percentile. ROCS and BSS are then computed to assess the skill of the forecast in predicting the water levels in the reservoirs to be lower than these threshold percentages of the full capacity. Although the actual operation of the reservoirs is quite a bit more complex, this can be interpreted as an assessment of the skill of the forecast in predicting curtailments to the irrigation sector.

3. Results

The following section outlines the results when applying the different types of forcing to the hydrological model over the 30 year hindcast period from 1981 to 2010. The analysis is carried out for different verification periods and lead times as the forecast quality may vary significantly with temporal scales and lead times. While the rainy season in the Limpopo river basins spans from October to March, the main rains typically take place from November to February. The main runoff season and the high runoff season, however, lag behind the rainy season by one or two months, occurring in general from December to May and from January to April respectively (see Fig. 4).

3.1 Skill of seasonal streamflow prediction

This section presents the skill expressed in the selected skill scores of the seasonal streamflow prediction for the three forecast systems described (FS_S4, FS_ESP, and FS_ESPcond) for Station 24 (Chókwe), Station 1, Station 18 and Station 20 in the Limpopo river basin (see Fig. 3 for the station locations). Station 24 is the one with the largest drainage area in the basin with available discharge data. Four stations (highlighted in Fig. 3) with diverse drainage areas were selected to assess the influence of the spatial scale and forecast location on the quality of the forecasts. Table 1 presents the main characteristics of these stations, such as drainage area, mean annual runoff and observed runoff coefficient (RC = runoff/precipitation). In these stations the performance of the hydrological model is found to be satisfactory, based on the evaluation measures and ranges proposed by Moriasi et al. (2007), which comprise the Nash-Sucliffe efficiency (NSE), and the ratio of the root mean square error to the standard deviation of the measured data (RSR). The coefficient of determination (R^2) is also included. These results are presented by Trambauer et al. (2014b) and are summarized in Table 1.

Fig. 5 (upper plots) presents the ROC diagram for the 6-months SRI-6 $\leq$ -0.5. For calculating the SRI-6 the verification period is from December to May and the SRI-6 value is recorded at the end of the period in May. The figure shows three of the four stations considered, for a lead time of five months (the forecast is issued in December). December is the only start time of the forecast that captures the whole 6 months main runoff season (from December to May) in the seasonal forecast. The ROC diagram for Station 18 is not presented given that it has a similar behaviour to Station 1. The ROC curves are presented for each forecasting system, and the ROC of FS_ESPcond is represented by the ensemble that results in the median ROCS. Results from the FS_ESPcond show for all stations a narrower 90% confidence interval when compared with the other two forecasting systems considered (see middle and lower plots in Fig. 5), thus suggesting that FS_ESPcond is more robust. Histograms of ROCS for FS_ESP are not shown as these are similar to those of FS_S4.

The ROCS of the FS_S4 in predicting SRI-6 $\leq$ -0.5 are generally quite high (around 0.8), but some lower values such as 0.72 (this for the station with largest contributing area) are observed (Fig. 5). The lower skills for the station with the largest contributing area for FS_S4 might be attributed to the shift from an arid to a more tropical climate, which means that the persistence of initial conditions would be lower. Also, given that this is mostly the case for the FS_S4 and less so for the FS_ESP and FS_ESPcond, we can speculate the ECMWF S4 seasonal forecast might have a better skill for the northern (more arid) part of the basin (area corresponding to the sub-basin draining to station1), than for the southern part of the basin. FS_ESP generally shows the lowest skills, with the skills of FS_ESPcond in between FS_ESP and FS_S4. The verification was also done for forecasts issued for the 4 months period JFMA (high-runoff season) with forecasts issued from November to January respectively. Figure 6 present the ROCS for the 4-monthly SRI (SRI-4 in April) $\leq$ -0.5 for three different lead times (three to five months) and two stations.

For the high runoff season SRI-4, similar results to those of SRI-6 are observed. In almost every case FS_S4 shows higher skill than FS_ESP and FS_ESPcond. The skill of the forecasts tend to decrease with lead time, especially for FS_ESP and FS_ESPcond, which do not show any skill at the 5 months lead time. In contrast, the skill of FS_S4 for the 5 months lead time is still good. The skill score verification for SRI-1 for the same 4-months period Jan-Apr (not shown) shows once more that the FS_S4 is more skilful than the other two forecasting systems. The smaller sub-basins (18 and 20) present lower skill for SRI-4 for the three forecasting systems, and while sub-basin 18 still presents some skill for all the three FS, sub-basin 20 only does so for the FS_S4 for all lead times. In general for all locations, the skill of the FS_S4 reduces slightly with lead time, while the skill of both FS_ESP and FS_ESPcond reduce more rapidly with lead time. A curious fact is that for a few stations (1, 18) for both SRI-4 and SRI-1, the FS_S4 shows a higher skill for a lead time of 4 months than for a lead time of 3 months. For the SRI-4, this means that the forecast is more skilful in predicting the April SRI-4 when issued in December than when issued in January. However, the differences are not statistically significant and this can be due to sampling errors.

Rank histograms for every station and lead time together with the results of the Kolmogorov-Smirnov test show that for the three forecasting systems, uniformity of the distribution cannot be rejected, indicating the forecasts are reliable. Fig. 7 presents the rank histograms of SRI-6 for Station 1 for the three forecasting systems as an example.

3.2 Skill of spatial hydrological indicators

Figure 8 shows the ROCS and the BSS of the FS_S4 in predicting agricultural drought conditions, i.e. in predicting aggregated Root Stress (RS) during the 6-monthly period DJFMAM to be higher than the 70th percentile. Yuan et al. (2013) show that the annual cycle of soil moisture in southern Africa (simulated by the VIC model) lags behind the precipitation. Figure 8 shows that the skill of the FS_S4 forecast in predicting agricultural droughts is higher than climatology (ROCS > 0.5, BSS > 0) throughout the entire basin.

To assess the skill of the seasonal forecast in predicting a specific decision variable in the Limpopo river basin, we calculate the skill of the forecast in predicting water levels thresholds in the reservoir that would result in curtailment to the irrigation sector. The availability of water is represented in each cell by the water level. In the cells corresponding to reservoirs, the water level is a surrogate for the storage and is described as a percentage of the full storage capacity of the reservoir. Figure 9 presents the ROCS and the BSS of the FS_S4 in predicting water levels during the 6-monthly period DJFMAM to be lower than the 50th and 37.5th percentiles, based on the analysis described in section 2.4. The figure shows that the skill of the FS_S4 forecast in predicting low water levels is higher than climatology (ROCS > 0.5, BSS > 0) throughout the basin. The spatial distribution across the basin does show the skill to be higher in the northern basin than in the southern basin, which may contribute to the lower skill found at Station 24 close to the basin outlet than at Station 1 in the upper (northern) basin.

The skill scores in cells that contain the reservoirs are represented by a circle to enhance visibility. It is clear from the figure that the skill of the forecast in predicting low water levels is higher in the reservoirs than in nearby streams. This can of course be expected due to the higher memory introduced by the reservoir's storage capacity with respect to the streams. Figure 10 presents the forecast probability of water levels to be lower than the 50th and 37.5th percentiles during the Dec 1991- May 1992 season as an example. This was the driest season in the last 30 years. The forecast is issued in December 1991.

The forecast probability of water levels in the reservoirs being lower than the 50th and 37.5th percentiles can be interpreted as the forecast probability of a curtailment of 20% and 65% respectively in the irrigation sector during the season. For several reservoirs in the basin the FS_S4 forecast issued in December 1991 predicted a high probability of curtailment to the irrigation sector during the Dec1991-May1992 season. Records confirm the lower than normal water levels during this season, with irrigation quota indeed being curtailed (DWA, 2013).

3.3 Analysis of a specific event

Yuan et al. (2013) note that "The major source of seasonal forecast predictability comes from the ocean, and the strongest signal is the El Niño Southern Oscillation (ENSO)". Given that the ECMWF S4 is influenced by the ENSO signal, it is interesting to analyse how the FS_S4 predicts streamflow in the onset of two clear El Niño years. The 1997/98 El Niño year is described in Thomson et al. (2003) as the largest this century, predicted with a high degree of certainty. Although many of the climate anomalies typical of an El Niño event took place around the globe, the devastating drought that was feared for southern Africa did not happen (Thomson et al., 2003). For this analysis another year was selected that had a less strong ONI but that did result in a severe drought (1982/83). Figure 11 presents the ensemble seasonal streamflow prediction from FS_S4 for both the 1997/98 and the 1982/83 seasons issued in October and updated in December for Station 24. The plots also show the climatology of the streamflow, and the 30th percentile i.e. the value below which 30 percent of the observations are found. The reference streamflow for that season and the forecast ensemble mean are also shown.

Figure 11 shows that in October the predictions from the forecasting system FS_S4 for El Niño seasons of 1982/83 and 1997/98 were relatively similar (see Fig. 11 upper panels) even though the 1997 JAS ONI was notably higher than the 1982 JAS ONI. The updated forecast in December, however, shows a different situation: While the forecast for the 1982/83 season point towards very dry conditions, the forecast of the 1997/98 season indicate near-normal conditions. Yet, the 1997 SON ONI is markedly higher than the 1982 SON ONI. Thus, in spite of the strong ONI conditions, the S4 system correctly forecasted the no-drought condition in 1997/98 season. This indicates that even though the S4 forecasting system is influenced by the Sea Surface Temperatures over the Niño-3.4 region, the precipitation and temperature forecasts over the Limpopo region is not only constrained by the SST evolution, but results from the atmospheric circulation response to different climate forcing.

4. Discussion

The performance of the three hydrological forecasting systems constructed with the same hydrological model and different meteorological ensemble forecasts are evaluated by means of widely used probabilistic verification skill scores, including the ROC diagram and the rank histogram. Among the forecasting systems considered in this study, FS_ESP is considered the most traditional. Such traditional approaches for hydrological forecasts rely on historical observations of the meteorological conditions, without considering meteorological forecasts. In ensemble probabilistic forecasting, the ESP approach, implicitly accounting for hydrologic persistence and historical variability of climate is normally used (Brown et al., 2010). FS_S4 is a more complex forecasting system as it requires as forcing the outputs of a seasonal meteorological forecast system, which are complex numerical models and resource intensive. FS_ESPcond, a modification of the ESP approach, conditions its ensemble on past years that had similar climate conditions to the year in which the forecast is made (Brown et al., 2010). Given that the Limpopo region is known to be affected by ENSO and droughts tend to occur during El Niño years, the forecast ensemble was constructed by assigning weights to the different ensemble traces based on the El Niño index.

The skill evaluation of the seasonal forecasts is limited by the use of model data as verification, i.e. we verify our forecasts against the baseline simulation, which was also used to provide the initial conditions to the forecasts. This is the same approach as taken in Yossef et al.(2013), Winsemius et al. (2014), Shukla et al. (2013) and Renner et al. (2009), and while it allows for the detailed (spatial) evaluation of the skill of the forecasts, it can potentially hide limitations of the modeling system. Therefore, these skill results should be interpreted as the upper limit of real predictability of the current system. Results of the seasonal streamflow prediction show that for every lead time FS_S4 is skilful in predicting SRI-6, SRI-4, and SRI-1 during the summer rainy season, while for both other FS the skill is lower and reduces more rapidly with lead time. This means that the complex S4 seasonal forecasting system adds value to the hydrological predictions compared to the climatology-based forecasting systems, as well as the ENSO mode conditioned climatology forecast systems. This was also observed during a specific event where expected anomalies due to El Niño did not materialise, but FS_S4 detected this. The skill reduces when going from SRI-6 to SRI-4 and SRI-1. This is as expected given the higher variability of the predictand for shorter aggregation periods. The skill from FS_ESP is lower than that of FS_S4 in almost every case, while the skill of FS_ESPcond is in general between the other two. For SRI-4, FS_ESP and to a lesser extent FS_ESPcond do not show any skill for 5 months lead time at any of the stations considered.

As expected, the skill of all forecasts tends to decrease with lead time. This is, however, especially the case for FS_ESP and FS_ESPcond where the decrease in skill with lead time is larger than for FS_S4. For the smaller aggregation periods (SRI-4 and SRI-1) FS_ESP deteriorates to climatology already at a lead time of 3 months for stations 18 and 20, the upstream basins of which are smaller in size. In the larger basins FS_ESP shows predictability up to 4 months lead time, probably due to the spatial aggregation taking place over larger basins, smoothing out uncertainties in space. This indicates that the memory in the hydrology (storage in groundwater, reservoirs, channels and wetlands) contributes to the predictability up to a lead time of 2 to 4 months. For longer lead times, the meteorological forcing dominates the predictability of the system. The critical lead time after which the importance of the meteorological forecast exceeds that of the initial conditions depends on the location and size of the basin and should be analysed for each sub-basin of interest. Rank histograms for every station and lead time indicate that the three forecast systems are reliable given that uniformity of the distribution cannot be rejected.

4.1 What does the analysis mean to end users?

The high predictability of FS_S4 for all lead times and aggregation periods of SRI is encouraging given that such a system, if made operational, may provide end users with sufficient time to decide upon measures to take in anticipation. For example, they might decide to change the cropping date or the cropping area if they expect not to have enough water to fulfil the crop requirements. Therefore, there is added value to using a seasonal meteorological forecast (ECMWF S4) to force the hydrological forecasting system when compared to the conventional ESP. The higher skill of the FS_S4 and FS_ESPcond compared to that of the FS_ESP for every lead time is in line with the study of Shukla et al. (2013), who show that for the region of the Limpopo river basin the meteorological forecast dominates the hydrological predictability for the wet season for almost every lead time considered. Only for the 1-month lead time forecasts issued in October they found a higher influence of the hydrological initial conditions to some extent. Moreover, Yossef et al. (2013) indicate that for semi-arid regions the initial conditions do not contribute much to the skill given the high sensitivity of the runoff coefficient to rainfall variability.

The FS_S4 was also evaluated regarding its ability to predict agricultural droughts and curtailments in irrigation (water levels lower than the 50th and 37.5th percentiles). Maps of spatially distributed ROCS and BSS (Figures 8 and 9) show that the skill of the FS_S4 forecast in predicting these conditions is higher than climatology (ROCS > 0.5, BSS > 0) throughout the basin. Indicating the probability of curtailment to the irrigation sector during the following season is an example of providing a forecast in an understandable format that is useful to the end users. If they are informed that there is a high probability of a high curtailment to the water available for irrigating their crops during the following season, users would have a clear idea of what is the best practice for that situation. Further improvements in forecasting skill could be achieved through better meteorological predictions or better estimation of initial conditions (Yossef et al., 2013). Whether the forecasts indeed have value will depend on the costs of decisions made in response to the forecast, losses in case of a wrong decision and the gain in case of a good decision. This should be further analysed in a continuation of this study.

As a next step, it is recommended that the forecast skill of the FS_S4 and FS_ESPcond be assessed in an actual forecasting mode for a following summer season. The seasonal meteorological forecast from S4 can be obtained in real-time for research purposes. To test a pre-operational system, the forecasting system ought to be statistically post-processed in order to remove biases in streamflow predictions. Moreover, the initial conditions for the forecasts could be better estimated through data assimilation of water levels in reservoirs and streams. This data could be obtained from the water managers of the basin. Despite the limitations of FS_S4 (access to real time atmospheric-ocean seasonal forecasts for non-ECMWF member-states, and their quality) and FS_ESPcond (depending on the calibration and reduced skill at long lead times) both systems show potential for seasonal hydrological drought forecasting in the Limpopo river basin to provide operational guidance to users.

5. Conclusions

We evaluate the performance of three forecasting systems (FS_S4, FS_ESP, and FS_ESPcond) in the Limpopo river basin. These systems make use of the same hydrological model and are forced with three different meteorological ensemble forecasts (two of which are based on resampled climatological records, FS_ESP and FS_ESPcond; and one based on seasonal meteorological forecasts, FS_S4). Results of the seasonal streamflow prediction show that the three forecasting systems show moderate skill in predicting SRI-6 (DJFMAM) $\leq$ -0.5. Moreover, the three forecasting systems are unbiased as suggested by the rank histograms.

For every lead time and aggregation period considered, FS_S4 is found to be skilful in predicting hydrological droughts represented by SRI$\leq$-0.5 during the summer rainy season. The skill reduces when going from SRI-6 to SRI-4 and SRI-1, as well as with increasing lead time. The skill of FS_ESP is lower than that of FS_S4 in almost every case and deteriorates rapidly with lead time, showing no skill after a lead time of 4-5 months for SRI-4 and SRI-1. This indicates that the memory in the hydrology contributes to the predictability up to 2 to 4 months but for longer lead times the predictability of the system is dominated by the meteorological forcing. FS_ESPcond shows in general lower skills than FS_S4 but it becomes comparable and can even outperform the latter for smaller lead times if the parameters for selection and weighting of ensemble members are carefully calibrated for each basin. Moreover, the skill of FS_ESPcond is more robust than that of the other forecasting systems as suggested by the narrower confidence intervals of ROCS. As with FS_ESP, the skill of FS_ESPcond also decreases faster than that of FS_S4 with lead time.

The high predictability of drought of FS_S4 for all lead times and aggregation periods of SRI and for the spatial drought indicators is encouraging given that such a system, if made operational, may provide end users with sufficient time to decide upon measures to take in anticipation. Moreover, FS_ESPcond shows promising results. This forecasting system only requires the ONI index previous to the forecast to weight the ensembles traces to include in the forecast. This system is relatively simple and presents the advantage that it can be coupled with the forecast of the ONI index that is available with a long lead time. Naturally, in this situation the uncertainties of both forecasts need to be considered.

Acknowledgements

This study was carried out in the scope of the DEWFORA (Improved Drought Early Warning and Forecasting to strengthen preparedness and adaptation to droughts in Africa) project which is funded by the Seventh Framework Programme for Research and Technological Development (FP7) of the European Union (grant agreement no: 265454).

References

Balsamo, G., Boussetta, S., Lopez, P., and Ferranti, L.: Evaluation of ERA-Interim and ERA-Interim-GPCP-rescaled precipitation over the USA, ECMWF ERA Report Series 5, 1-25, available at: http://old.ecmwf.int/publications/library/ecpublications/_pdf/era/era_report_series/RS_5.pdf, last access: August 2014, 2010.

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Cloke, H., Dee, D., Dutra, E., Muñoz-Sabater, J., Pappenberger, F., de Rosnay, P., Stockdale, T., and Vitart, F.: ERA-Interim/Land: a global land water resources dataset, Hydrol. Earth Syst. Sci. Discuss., 10, 14705-14745, doi:10.5194/hessd-10-14705-2013, 2013.

Barbosa, P., Naumann, G., Valentini, L., Vogt, J., Dutra, E., Magni, D., and De Jager, A.: A Pan-African map viewer for drought monitoring and forecasting, 14th Waternet Symposium, Dar es Salaam, Tanzania, 30 October to 1 November 2013,

Brown, C., Baroang, K. M., Conrad, E., Lyon, B., Watkins, D., Fiondella, F., Kaheil, Y., Robertson, A., Rodriguez, J., Sheremata, M., and Ward, M. N.: Managing climate risk in water supply systems, IRI technical report 10-15, International Research Institute for Climate and Society, Palisades, NY, available online at: http://iri.columbia.edu/resources/publications/pub_id/1048/, last access: 7 February 2014, 133, 2010.

Cloke, H. L., and Pappenberger, F.: Evaluating forecasts of extreme events for hydrological applications: an approach for screening unfamiliar performance measures, Meteorological Applications, 15, 181-197, doi: 10.1002/met.58, 2008.

Coumou, D., and Rahmstorf, S.: A decade of weather extremes, Nature Clim. Change, 2, 491-496, doi: 10.1038/nclimate1452, 2012.

Day, G.: Extended Streamflow Forecasting Using NWSRFS, Journal of Water Resources Planning and Management, 111, 157-170, 1985.

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, Quarterly Journal of the Royal Meteorological Society, 137, 553-597, doi: 10.1002/qj.828, 2011.

DEWFORA: WP2-D2.2 - Inventory of institutial frameworks and drought mitigation and adaptation practices in Africa, DEWFORA project - EU FP7, www.dewfora.net, last access: August 2014, 2011.

DEWFORA: WP5-D5.1 - Concept report describing the outline of a framework for drought warning and mitigation in Africa DEWFORA project - EU FP7 www.dewfora.net, last access: July 2014, 2012.

DEWFORA: WP6-D6.2 - Limpopo Case Study, Application of the DEWFORA Drought Early Warning Framework DEWFORA project - EU FP7 www.dewfora.net, last access: July 2014, 2013.

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.

Dutra, E., Pozzi, W., Wetterhall, F., Di Giuseppe, F., Magnusson, L., Naumann, G., Barbosa, P., Vogt, J., and Pappenberger, F.: Global meteorological drought - Part 2: Seasonal forecasts, Hydrol. Earth Syst. Sci. , 18, 2669-2678, doi:10.5194/hess-18-2669-2014, 2014.

DWA: Tzaneen Dam, Department of Water Affairs, South Africa, 2013.

Ek, M., Xia, Y., and the NLDAS team: NCEP/EMC NLDAS Support for Drought Monitoring and Seasonal Prediction, US National Oceanic and Atmospheric Administration, Climate Test Bed Joint Seminar Series, NASA, Goddard Visitor Center, Greenbelt, Maryland, http://www.nws.noaa.gov/ost/climate/STIP/FY10CTBSeminars/mek_041410.pdf, access 01/12/2014, 2010.

Friederichs, P., and Thorarinsdottir, T. L.: Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction, Environmetrics, 23, 579-594, 2012.

Grasso, V. F.: Early Warning Systems: State-of-Art Analysis and Future Directions - Draft Report, United Nations Environment Programme (UNEP), 2009.

Hamlet, A., and Lettenmaier, D.: Columbia River Streamflow Forecasting Based on ENSO and PDO Climate Signals, Journal of Water Resources Planning and Management, 125, 333-341, doi:10.1061/(ASCE)0733-9496(1999)125:6(333), 1999.

Hansen, J., Sato, M., and Ruedy, R.: Perception of climate change, Proceedings of the National Academy of Sciences, 109, E2415-E2423, doi:10.1073/pnas.1205276109, 2012.

Hargreaves, G. H., and Allen, R. G.: History and Evaluation of Hargreaves Evapotranspiration Equation, Journal of Irrigation and Drainage Engineering, 129, 53-63, 2003.

Huffman, G. J., Adler, R. F., Bolvin, D. T., and Gu, G.: Improving the global precipitation record: GPCP version 2.1, Geophys. Res. Lett, 36, L17808, doi:10.1029/2009GL040000, 2009.

Huntington, T. G.: Evidence for intensification of the global water cycle: Review and synthesis, Journal of Hydrology, 319, 83-95, http://dx.doi.org/10.1016/j.jhydrol.2005.07.003, 2006.

IPCC: Summary for Policymakers. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., van den Dool, H., Saha, S., Mendez, M. P., Becker, E., Peng, P., Tripp, P., Huang, J., DeWitt, D. G., Tippett, M. K., Barnston, A. G., Li, S., Rosati, A., Schubert, S. D., Rienecker, M., Suarez, M., Li, Z. E., Marshak, J., Lim, Y.-K., Tribbia, J., Pegion, K., Merryfield, W. J., Denis, B., and Wood, E. F.: The North American Multimodel Ensemble: Phase-1 Seasonal-to-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction, Bulletin of the American Meteorological Society, 95, 585-601, doi: 10.1175/bams-d-12-00050.1, 2013.

Landman, W. A., and Mason, S. J.: Operational long-lead prediction of South African rainfall using canonical correlation analysis, International Journal of Climatology, 19, 1073-1090, doi: 10.1002/(sici)1097-0088(199908)19:10<1073::aid-joc415>3.0.co;2-j, 1999.

LBPTC: Joint Limpopo River Basin Study Scoping Phase. Final Report. BIGCON Consortium., Limpopo Basin Permanent Technical Committee, http://www.limcom.org/_system/writable/DMSStorage/1031en/LIMCOM2010_ScopingStudy_Eng.pdf, last access: 26 August 2014, 2010.

Molteni, F., Stockdale, T., Balmaseda, M. A., BALSAMO, G., Buizza, R., Ferranti, L., Magnunson, L., Mogensen, K., Palmer, T., and Vitart, F.: The new ECMWF seasonal forecast system (System 4), ECMWF Tech. Memo., 656, 49 pp., 2011.

Morgan, R.: The development and applications of a Drought Early Warning System in Botswana, Disasters, 9, 44-50, 10.1111/j.1467-7717.1985.tb00909.x, 1985.

Moriasi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., and Veith, T.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, Transactions of the ASABE, 50, 885-900, 2007.

NOAA Historical El Nino/ La Nina episodes (1950-present): http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml, access: Jan 2014, 2014.

Renner, M., Werner, M. G. F., Rademacher, S., and Sprokkereef, E.: Verification of ensemble flow forecasts for the River Rhine, Journal of Hydrology, 376, 463-475, doi: http://dx.doi.org/10.1016/j.jhydrol.2009.07.059, 2009.

RIMES, Regional Integrated Multi-Hazard Early Warning System for Africa and Asia, Water Related Hazard - Drought: http://www.rimes.int/wrh/drought, access: 11 December 2014, 2014.

Roulin, E.: Skill and relative economic value of medium-range hydrological ensemble predictions, Hydrol. Earth Syst. Sci., 11, 725-737, doi: 10.5194/hess-11-725-2007, 2007.

Sheffield, J., Wood, E. F., Chaney, N., Guan, K., Sadri, S., Yuan, X., Olang, L., Amani, A., Ali, A., Demuth, S., and Ogallo, L.: A drought monitoring and forecasting system for sub-Sahara African water resources and food security, Bulletin of the American Meteorological Society, 95, 861-882, doi: 10.1175/bams-d-12-00124.1, 2014.

Shukla, S., and Wood, A. W.: Use of a standardized runoff index for characterizing hydrologic drought, Geophysical research letters, 35, L02405, doi: 10.1029/2007gl032487, 2008.

Shukla, S., Sheffield, J., Wood, E. F., and Lettenmaier, D. P.: On the sources of global land surface hydrologic predictability, Hydrol. Earth Syst. Sci., 17, 2781-2796, doi: 10.5194/hess-17-2781-2013, 2013.

Szczypta, C., Calvet, J. C., Albergel, C., Balsamo, G., Boussetta, S., Carrer, D., Lafont, S., and Meurey, C.: Verification of the new ECMWF ERA-Interim reanalysis over France, Hydrology and Earth System Sciences, 15, 647-666, 2011.

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., Dutra, E., Maskey, S., Werner, M., Pappenberger, F., van Beek, L. P. H., and Uhlenbrook, S.: Comparison of different evaporation estimates over the African continent, Hydrol. Earth Syst. Sci., 18, 193-212, 10.5194/hess-18-193-2014, 2014a.

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. Discuss., 11, 2639-2677, doi:10.5194/hessd-11-2639-2014, 2014b.

Trenberth, K.: Framing the way to relate climate extremes to climate change, Climatic Change, 115, 283-290, doi: 10.1007/s10584-012-0441-5, 2012.

Tveito, O. E., Wegehenkel, M., van der Wel, F., and Dobesch, H.: COST Action 719: The Use of Geographic Information Systems in Climatology and Meteorology: Final Report, EUR-OP, Luxembourg: Office for Official Publications of the European Communities, 2008.

UNISDR: International strategy for disaster reduction ‘‘Hyogo framework for action 2005–2015: Building the Resilience of Nations and Communities to Disasters, World conference on disaster reduction, Kobe, Japan, 2005,

van Beek, L. P. H.: Forcing PCR-GLOBWB with CRU data, Utrecht University, Utrecht, Netherlands: http://vanbeek.geo.uu.nl/suppinfo/vanbeek2008.pdf, last access: December 2013, 2008.

van Beek, L. P. H., and Bierkens, M. F. P.: The Global Hydrological Model PCR-GLOBWB: Conceptualization, Parameterization and Verification, Utrecht University, Faculty of Earth Sciences, Department of Physical Geography, Utrecht, The Netherlands, 2009.

Werner, K., Brandon, D., Clark, M., and Gangopadhyay, S.: Climate Index Weighting Schemes for NWS ESP-Based Seasonal Volume Forecasts, Journal of Hydrometeorology, 5, 1076-1090, doi: 10.1175/jhm-381.1, 2004.

Werner, M., Schellekens, J., Gijsbers, P., van Dijk, M., van den Akker, O., and Heynert, K.: The Delft-FEWS flow forecasting system, Environmental Modelling & Software, 40, 65-77, http://dx.doi.org/10.1016/j.envsoft.2012.07.010, 2013.

Wilks, D. S.: Statistical methods in the atmospheric sciences - 3rd edition, International Geophysics series, vol 100, 676 pp., 2011.

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.

Wood, A. W., and Lettenmaier, D. P.: An ensemble approach for attribution of hydrologic prediction uncertainty, Geophysical research letters, 35, L14401, doi: 10.1029/2008gl034648, 2008.

WWRP/WGNE, Forecast verification: Issues, Methods and FAQ. Joint Working Group on Verification sponsored by the World Meteorological Organization, access: 26 August 2014, 2013.

Yossef, N. C., Winsemius, H., Weerts, A., van Beek, R., and Bierkens, M. F. P.: Skill of a global seasonal streamflow forecasting system, relative roles of initial conditions and meteorological forcing, Water Resources Research, 49, 4687-4699, doi: 10.1002/wrcr.20350, 2013.

Yuan, X., Wood, E. F., Chaney, N. W., Sheffield, J., Kam, J., Liang, M., and Guan, K.: Probabilistic Seasonal Forecasting of African Drought by Dynamical Models, Journal of Hydrometeorology, 14, 1706-1720, doi: 10.1175/jhm-d-13-054.1, 2013.

Table 1 Model evaluation measures for runoff for selected stations, ordered by basin size

Station number	Sub-basin area (km^2)	Mean annual observed runoff (m^3/s)	RCobs (%)	R^2	NSE	RSR
24	342,000	96.9	1.7	0.92	0.90	0.32
1	201,001	39.5	1.2	0.69	0.57	0.65
18	98,240	12.2	0.7	0.68	0.62	0.62
20	12,286	14.8	5.3	0.70	0.65	0.59

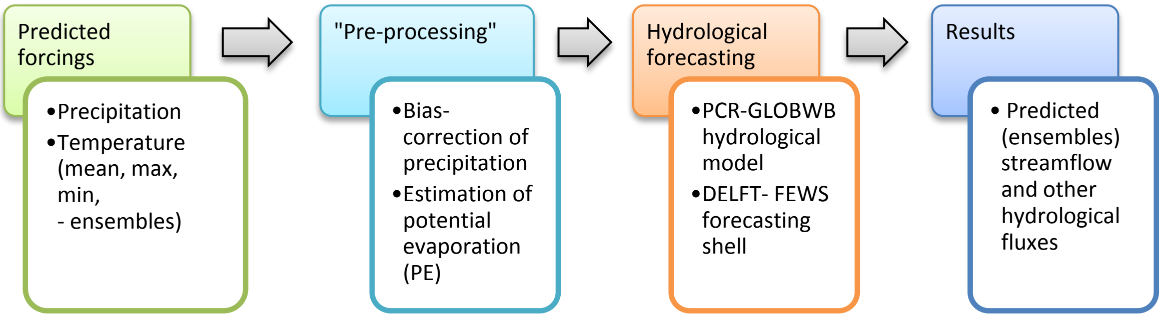

Fig. 1 Approach followed in the forecasting system for the Limpopo river basin

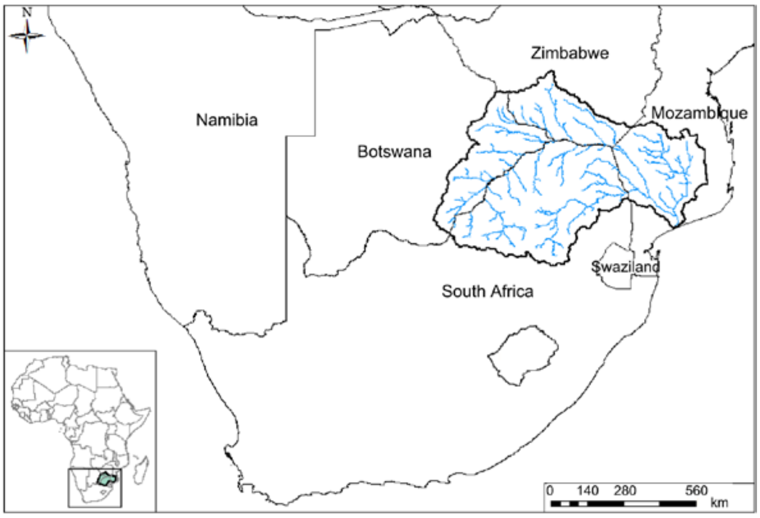

Fig. 2 Location of the Limpopo river basin

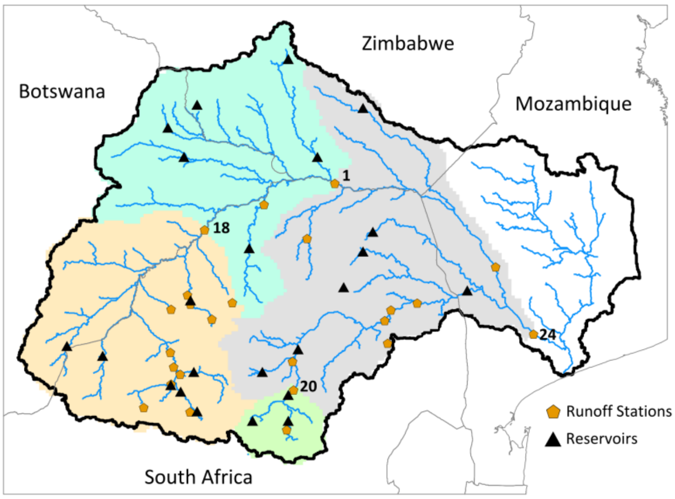

Fig. 3 Locations of selected hydrometric stations and reservoirs in the Limpopo basin

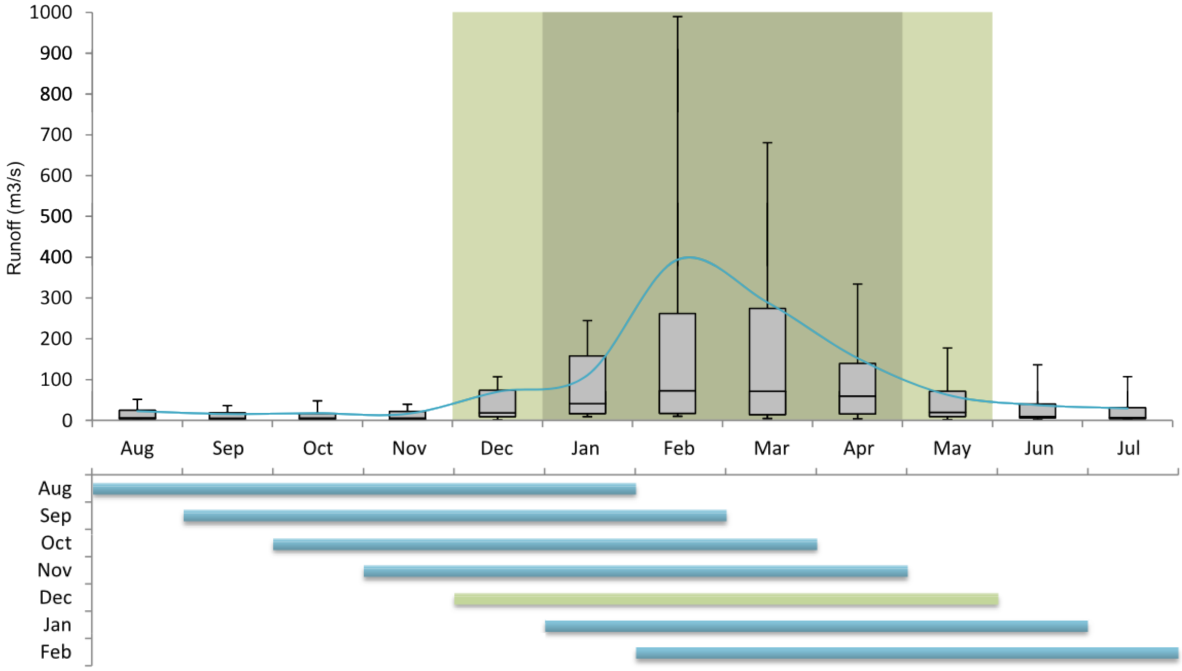

Fig. 4 Upper plot: Limpopo river flow regime for Station 24 at Chókwe. The blue line represents the average observed runoff, and the whiskers of the boxplots represent the 10th percentile and the 90th percentile. The lighter and darker shaded areas represent the main runoff period and high runoff period, respectively. Lower plot: Initialization dates and length of forecasts during the year. The forecast issued in December is highlighted as the one that captures the main runoff season

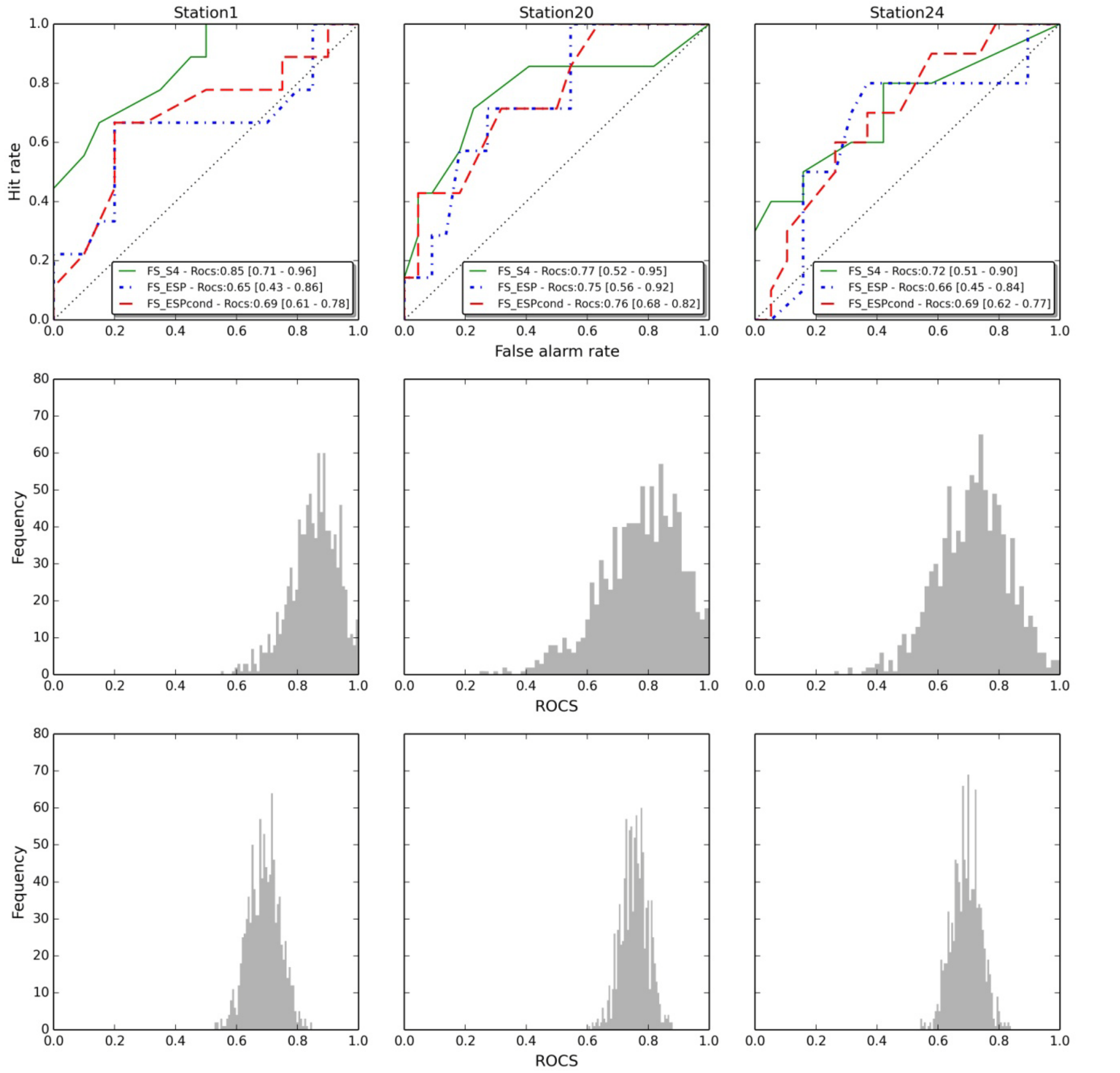

Fig. 5 Upper plots: Relative operating characteristic (ROC) diagram representing false alarm rate versus hit rate for the 6 months SRI (DJFMAM) ≤ -0.5 given by FS_S4, FS_ESP and FS_ESPcond for three stations (1, 20, and 24). The ROCS for each forecasting system together with the 90% confidence interval (5-95th percentiles) resulting from the bootstrap are indicated in the legend. Middle and bottom plots: histogram of the bootstrapped ROCS for FS_S4 (middle) and FS_ESPcond (lower) respectively for the same three stations.

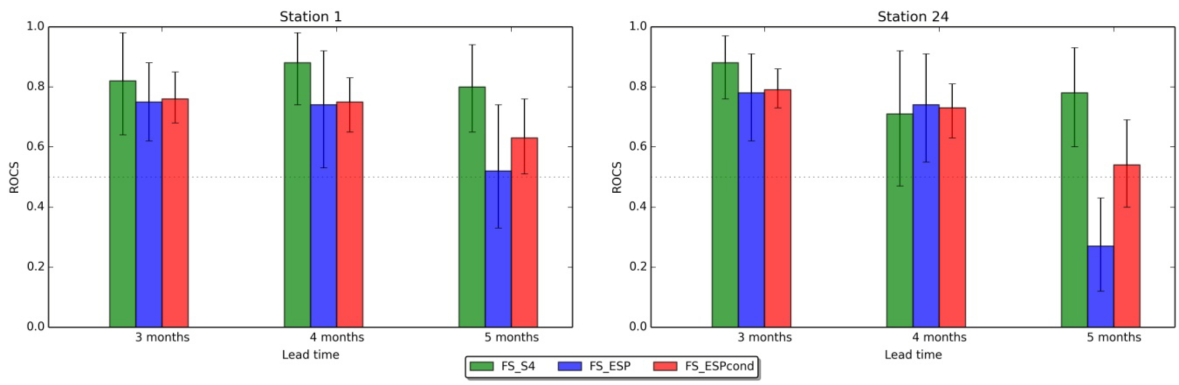

Fig. 6 ROCS for the SRI-4 (JFMA) ≤ -0.5 given by FS_S4, FS_ESP, and FS_ESPcond for different lead times, for two of the stations (1, 24). The error bars represent the 5-95th percentiles of the bootstrapped ROCS values.

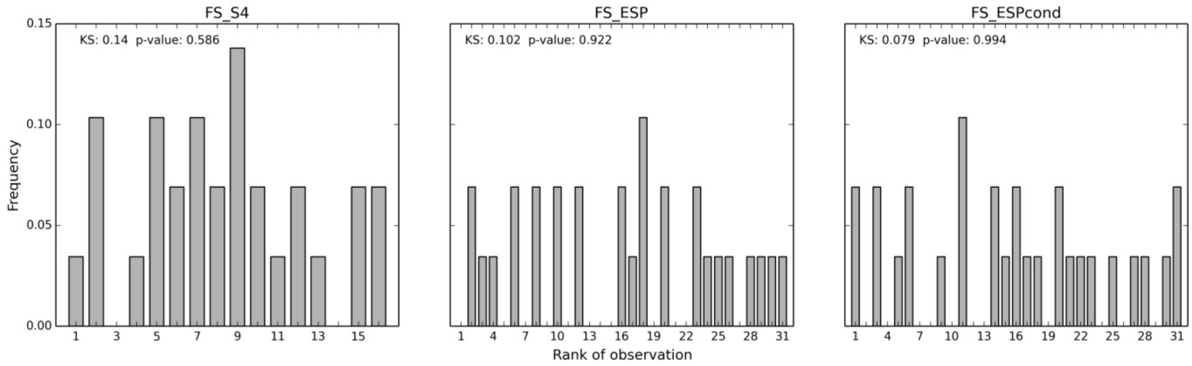

Fig. 7 Rank histograms of SRI-6 for Station 1 for the three forecasting systems (FS_S4, FS_ESP, and FS_ESPcond). The results of the Kolmogorov-Smirnov test for uniformity are presented in each plot.

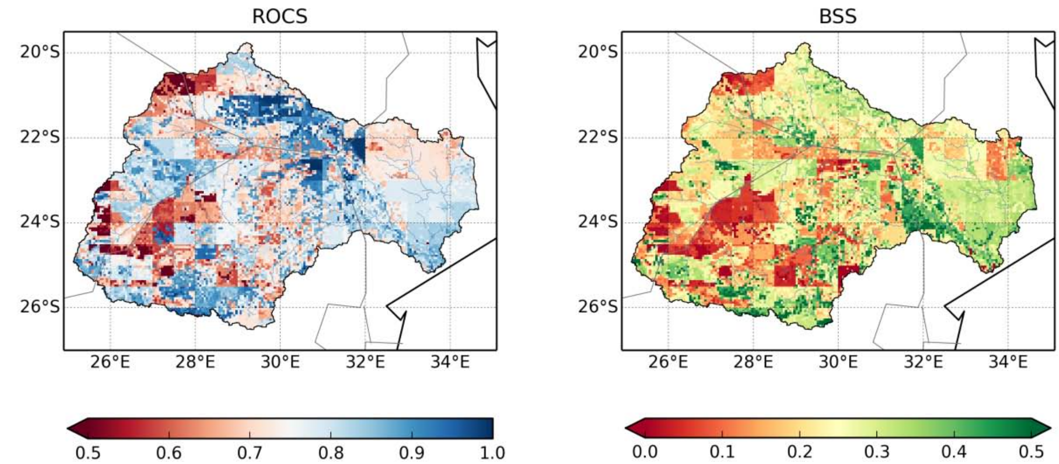

Fig. 8 ROCS and BSS for DJFMAM Root Stress (RS) > 70th percentile for the FS_S4

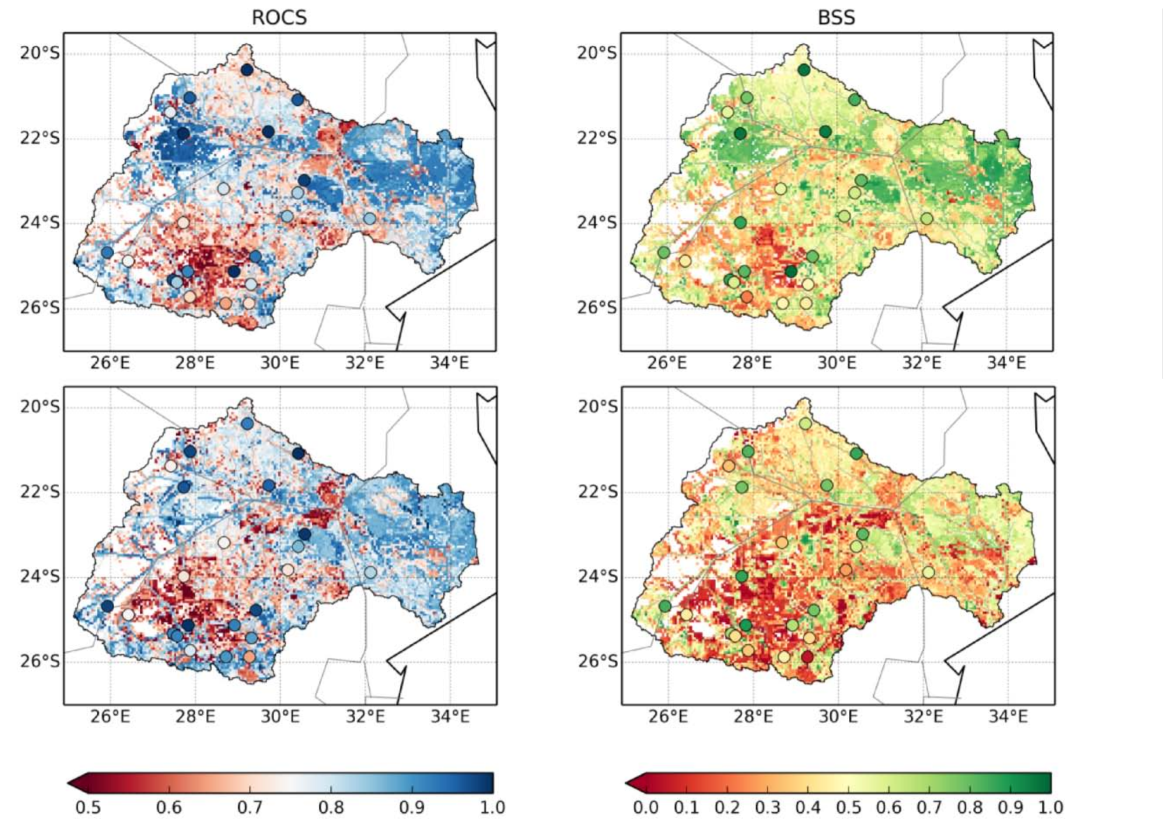

Fig. 9 ROCS and BSS for: DJFMAM Water Level (WL) < 50th percentile (upper plots), and WL < 37.5th percentile (lower plots) for the FS_S4

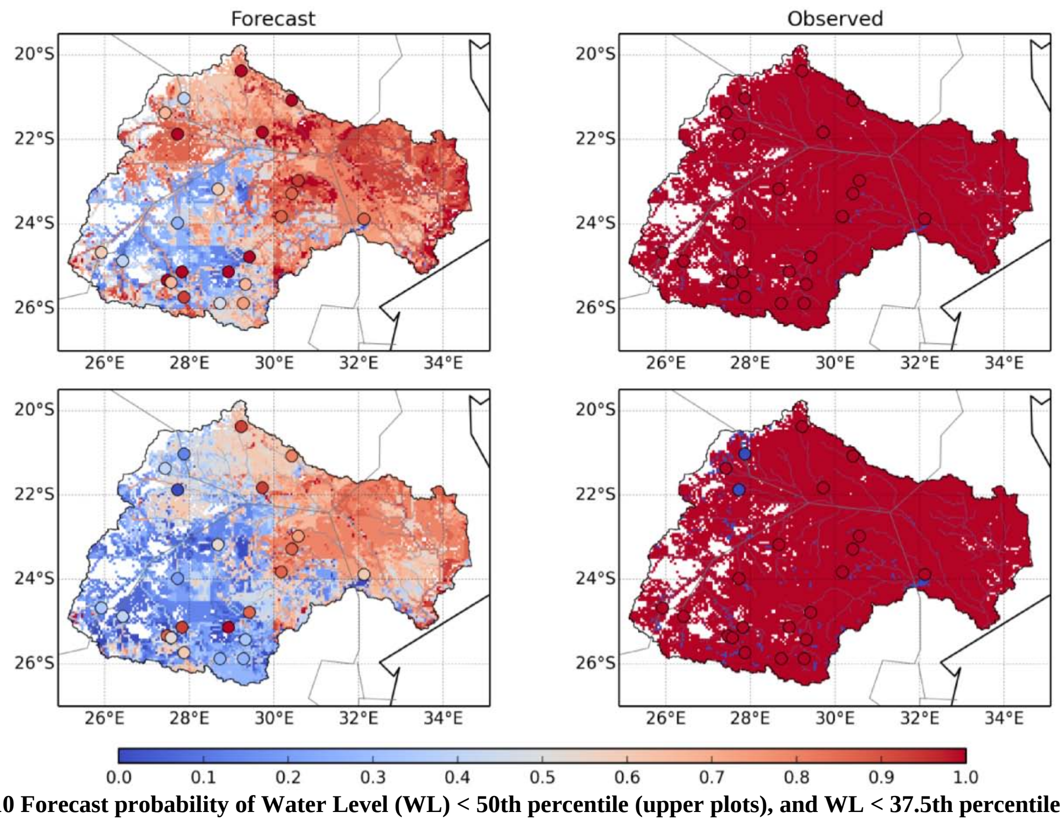

Fig. 10 Forecast probability of Water Level (WL) < 50th percentile (upper plots), and WL < 37.5th percentile (lower plots) for the FS_S4 during the season Dec 1991- May 1992 issued in Dec 1991 (left panels), and what actually occurred: 1=yes, 0=no (right panels).

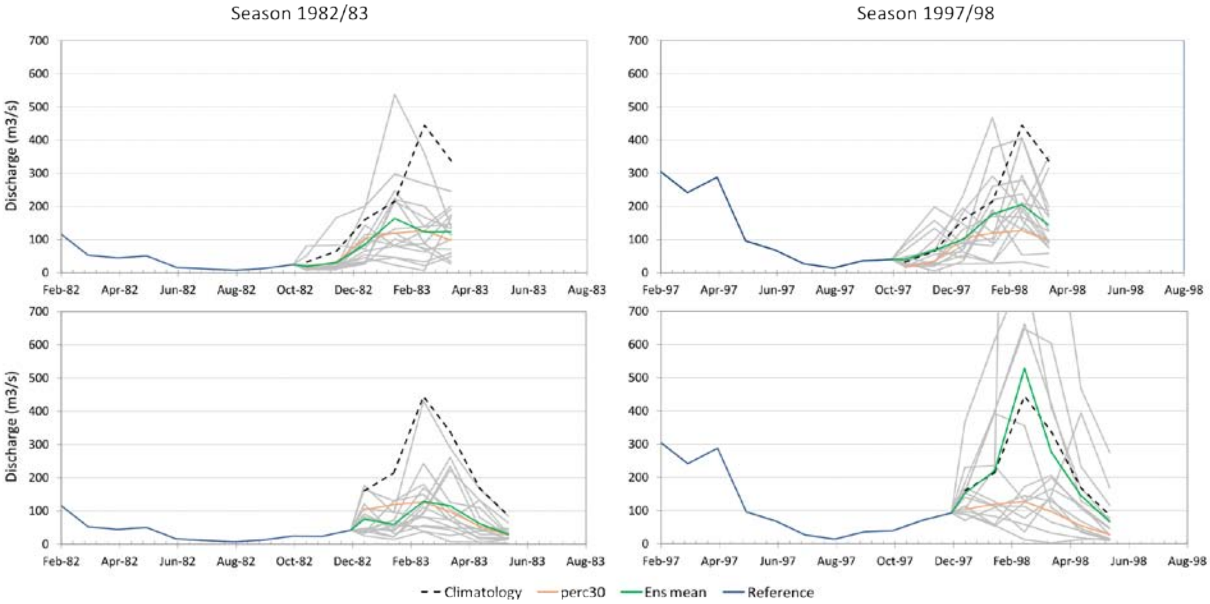

Fig. 11 Seasonal forecast FS_S4 for two seasons issued in October (upper panel) and December (lower panel)