We thank the editor and one reviewer for insightful and constructive comments and provide answers to the comments below. The editor/reviewer comments are in normal font, our replies are in bold font.

Editor

Dear Authors,

thank you indeed for your careful review, where you adequately address all the comments raised by the Referees and myself.

I sent the revised version to one of the previous referees, who acknowledges that you satisfyingly modified the paper according to the suggestions; he/she still have doubts on the opportunity to publish this kind of work on HESS, but, as I have already written in my comments in December, I believe that the wide interest that the adequate use of open data and open models has for all the hydrologists that attempt to study catchments located all over the world but in particular in developing countries and/or at large regional scale, may fully justify the publication of the manuscript.

I would ask you to answer to the last three points raised by the Referee and submit a second revision (that I will revise myself).

Thank you. We have addressed these three points as outlined below.

Reviewer 1

Dear Authors,

You did not address my main criticism/comments regarding the objective of the work (both in abstract/manuscript, it is also away things are written down). The objective of a scientific study cannot be to develop a forecasting system (however you can build and use a flood forecasting system to investigate predictability....)

Page 13 line 291 => using 10% was earlier suggested by Georgakakos (1986?) and later by Weerts and Serafy (2006?)

These references were added to the manuscript.

You discuss and conclude that the forecast might be extended beyond 7 days although the results show that climatology is doing better than all experiments So I don't think you can conclude this based on your study (it is pure speculation)

We agree and have modified the discussion section. The item was removed from the conclusions.

I wondered why CRPSS is not calculated in table6 although I can live with the table as is.

We have not included this to keep the table as compact as possible. If CRPSS is included, the table would need 2 additional columns, i.e. CRPSS computed with climatology as the reference and CRPSS with persistence as the reference. The table is already quite large and complex and we believe that the reader can easily compute various CRPSS scores from the information given in the table. Because the reviewer can live with the table as it is, we suggest keeping it in the present format.

Operational river discharge forecasting in poorly gauged basins: the Kavango River Basin case study

Peter Bauer-Gottwein[1*], Iris. H. Jensen[1], Radoslaw Guzinski[2], Gudny K. T. Bredtoft[1], Sidsel Hansen[1], and Claire I. Michailovsky[1#]

1) Department of Environmental Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
2) DHI GRAS, DK-2970 Hørsholm, Denmark

*: corresponding author, pbau@env.dtu.dk

#: now at Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA

Abstract

Operational probabilistic forecasts of river discharge are essential for effective water resources management. Many studies have addressed this topic using different approaches ranging from purely statistical black-box approaches to physically-based and distributed modelling schemes employing data assimilation techniques. However, few studies have attempted to develop operational probabilistic forecasting approaches for large and poorly gauged river basins. The objective of this study is to develop open-source software tools to support hydrologic forecasting and integrated water resources management in Africa. We present an operational probabilistic forecasting approach which uses public-domain climate forcing data and a hydrologic-hydrodynamic model which is entirely based on open-source software. Data assimilation techniques are used to inform the forecasts with the latest available observations. Forecasts are produced in real time for lead times of 0 to 7 days. The operational probabilistic forecasts are evaluated using a selection of performance statistics and indicators and the performance is compared to persistence and climatology benchmarks. The forecasting system delivers useful forecasts for the Kavango River, which are reliable and sharp. Results indicate that the value of the forecasts is greatest for intermediate lead times between 4 and 7 days.

Introduction

Operational probabilistic hydrological modelling and river discharge forecasting is an active research topic in water resources engineering and applied hydrology (Pagano et al., 2014). Sharp and reliable forecasts of river discharge are required over a range of forecasting horizons for flood and drought management. A state of the art river discharge forecasting system consists of a weather forecast or an ensemble of weather forecasts (Cloke and Pappenberger, 2009), a hydrologic-hydrodynamic modelling system and a data assimilation approach to inform the forecasts with all available in situ and remote sensing observations. Alternatively, in the absence of resources, data and computing power, simpler solutions can be implemented which disregard more and more of the physics and rely on past observations to parameterize black-box type models such as, for instance, artificial neural networks (Maier et al., 2010).

Many studies have shown that operational hydrological models can benefit from the assimilation of in-situ or satellite remote sensing observations. Different techniques and approaches have been presented (Liu et al., 2012). They differ both in terms of the type of data that are assimilated to the models, the assimilation algorithms used and in terms of the assimilation strategy, i.e. which model components, states and/or parameters are updated. Some hydrological data assimilation studies update the internal states of rainfall-runoff models (e.g. Clark et al., 2008; Pauwels and De Lannoy, 2009) while other approaches focus on updating the hydrodynamic parts of the model (Biancamaria et al., 2011; Neal et al., 2009) or combinations of rainfall-runoff and routing state variables (e.g.

Rakovec et al., 2012). One of the most popular algorithms used in hydrologic data assimilation is the ensemble Kalman filter (e.g. Clark et al., 2008). Alternatively, the particle filter (Moradkhani et al., 2005) can be used, which does not require the assumption of Gaussian model errors. Variational data assimilation has also been used in a number of hydrologic studies (e.g. Seo et al., 2009, 2003). Some studies use filtering approaches where the gain is determined heuristically from offline simulations and then used operationally in forecasting mode (Madsen and Skotner, 2005). As pointed out by Liu et al., 2012, despite the large body of literature on hydrologic data assimilation, few studies evaluate the benefit of data assimilation for actual forecasting and practical application of data assimilation by operational agencies is rare.

In many river basins the performance of operational hydrological modelling and forecasting is limited because in-situ observations of precipitation and river discharge are scarce or unavailable. This is also the case for many of Africa's large river basins which are poorly gauged (e.g. Zambezi, Volta, Congo). Consistent, long-term and spatially resolved in-situ observations of precipitation and river discharge are unavailable for large portions of Africa. Moreover, the number of operational meteorological stations and river discharge stations has been decreasing consistently around the world since the 1970s (Fekete and Voeroesmarty, 2007; Peterson and Vose, 1997). Remote sensing techniques have the potential to fill critical data gaps in the observation of the global hydrological cycle. All major components of the water balance, except river discharge, can now be estimated based on various types of remote sensing data. However, the available techniques are still limited by coarse spatial and temporal resolution as well large and/or poorly understood error characteristics (Tang et al., 2009). From a management perspective one of the most important components of the hydrological cycle is river discharge. Extremely high flows in rivers cause flooding which can have severe consequences in terms of fatalities and economic damage. Low flows cause conflicts in the allocation of scarce water resources between economic sectors and/or the environment. Therefore, in many river basins there is a need for hydrological models to provide operational estimates of river discharge based on remotely sensed observations and limited available in-situ measurements.

The TIGER-NET project addresses the demand for free, up-to-date and spatially resolved water information for the African continent. The project is funded by the European Space Agency (ESA) and aims to support integrated water resources management in Africa by (i) providing access to ESA Earth observation (EO) data, (ii) developing an open-source Water Observation and Information System (WOIS) and (iii) implementing capacity building actions in collaboration with African partner institutions (Guzinski et al., 2014).

The WOIS includes a hydrological modelling component, which supports long-term scenario analysis (e.g. impact of climate change, deforestation etc.) as well as operational probabilistic forecasting. The specific objective for the operational modelling capability is to provide reliable and sharp probabilistic forecasts of river discharge over time horizons of up to one week. In addition to hydrological modelling, WOIS includes functionality for operational flood monitoring, basin characterization at high (~30 m) and medium (~1 km) spatial resolutions and derivation of other products requiring EO data processing and analysis (Guzinski et al., 2014). It was designed for use in African organizations, where budgetary and technical constraints often limit the use of EO data for integrated water resources management. Therefore, WOIS is based purely on free, open-source software components and was created as an easy to use tool for both capacity building and operational use. Among the partner institutions engaged in the TIGER-NET project is the Namibian Ministry of Agriculture, Water and Forestry. The Ministry has an interest in forecasting the discharge of the Kavango River.

Based on these requirements, this study has four specific objectives:

1. Development of a robust and simple probabilistic river discharge forecasting system for poorly gauged river basins, based solely on open source software and public-domain data.
2. Informing the forecasting system with in-situ discharge observations in real time.
3. Operational demonstration of the system for the Kavango River case study.
4. Comprehensive evaluation of the operational probabilistic forecasts using a selection of performance statistics and indicators as well as comparison with persistence and climatology benchmarks.

The entire system has been implemented in an open-source GIS environment (QGIS, GDAL, Python). Installation and source code are available for download from the TIGER-NET webpage (www.tiger-net.org).

Materials and Methods

Study Area

The Kavango River originates in the highlands of central Angola and flows south to the border between Angola and Namibia. The Cuito River joins the Kavango River just before the river enters into Namibia's Caprivi Strip. It terminates in the Okavango Delta, a large wetland system in Northern Botswana (Milzow et al., 2009). An overview of the basin is provided in Figure 1. The basin is located on the Southern fringes of the inter-tropical convergence zone. A strong south-to-north precipitation gradient is observed. The climate is highly seasonal and large inter-annual variations are typical, which are controlled by a number of climate time scales (McCarthy et al., 2000; Wolski et al., 2014). The Kavango River is an important resource for all riparian countries and forms the basis of many people's livelihoods (Kgathi et al., 2006). While water scarcity and water allocation between economic sectors and the environment have been in focus for some time, flood risk has recently become a major concern because the northern part of Namibia has experienced increased magnitude and frequency of flooding events since 2008 (Wolski et al., 2014). Water managers need accurate and reliable forecasting tools to deal with both floods and droughts.

Three hydrological modelling efforts have been reported in the literature for the Kavango River basin. Folwell and Farqhuarson, 2006 used the Global Water Availability Assessment (GWAVA) model to assess climate change impacts in the basin. Hughes et al., 2011, 2006 calibrated a Pitman model for the basin and were able to reproduce in-situ observations satisfactorily. Milzow et al., 2011 developed a SWAT (Soil and Water Assessment Tool) model of the Kavango basin and calibrated the model with water levels from radar altimetry, soil moisture from Envisat-ASAR and total water storage change from GRACE.

Long-term in-situ observations of river discharge are available from two hydrometric stations in the basin, Rundu and Mohembo (Figure 1). Table 1 summarizes the main characteristics of the Kavango river basins and the two sub-basins contributing to the stations Rundu and Mohembo.

Hydrologic and hydrodynamic modelling

The modelling approach implemented in this study consists of a hydrologic (rainfall-runoff) model which is coupled to a simple routing model for channel flow. A one-way coupling between the two model compartments is implemented, i.e. once runoff has entered the river channel, the water cannot move back into the land phase of the hydrological cycle.

We use the well-known SWAT hydrological model, version 2009 (Gassman et al., 2005; Neitsch et al., 2011) for rainfall-runoff modelling. SWAT is a semi-distributed, physically based hydrological model which operates at a daily time step. The river basin is divided into a number of sub-basins. Each sub-basin is in turn divided into hydrological response units (HRU), which are defined as portions of the sub-basin with similar terrain slope, land use and soil type. The Kavango SWAT model consists of 12 subbasins with outlets located at the confluences of major tributaries as well as at in-situ discharge station locations (Figure 1).

The hydrodynamic model used in this study is a simple Muskingum routing scheme, which is implemented outside of the SWAT simulator to allow efficient updating in the data assimilation scheme. Muskingum parameters are computed from river widths, assumed cross section geometry and channel Manning numbers (which are calibration parameters). The river is divided into 12 primary individual river reaches. The primary reaches are further sub-divided if required to meet the numerical stability criteria of the Muskingum routing scheme (Chow et al., 1988). The hydrodynamic model state vector consists of the simulated discharges in each individual reach. In the Muskingum routing scheme, the model operator propagating the discharge forward in time is linear, i.e. the simulated discharges at time step *t+1* are a linear function of the simulated discharges at time step *t* and the runoff forcings at time steps *t* and *t+1*:

$$\mathbf{q}^{t+1} = \mathbf{A}\mathbf{q}^{t} + \mathbf{B}\mathbf{r}^{t} + \mathbf{C}\mathbf{r}^{t+1} \quad (1)$$

In this equation, **q** is the vector of simulated discharges and **r** is the vector of runoff forcings, **A**, **B** and **C** are linear operators which depend on the configuration of the river channels and network connectivity and the superscripts indicate time steps. For details on the implementation of the Muskingum routing scheme the reader is referred to Chow et al., 1988 and Michailovsky et al., 2013.

Input data

SWAT requires the following input datasets: elevation, land cover, soil type and climate forcings. The elevation dataset is used for automatic watershed and river network delineation as well as for the determination of terrain slope. We use the ACE2 (Altimeter Corrected Elevation, version 2, Berry et al., 2010) global elevation dataset at a resolution of 30 arc-seconds. The parameterization of vegetation processes in the SWAT model is based on the land cover input dataset. We use the USGS Global Land Cover Characterization (GLCC) dataset, version 2.0 with a spatial resolution of 1 km (USGS, 2008). The soil dataset forms the basis for parameterizing soil hydraulic processes in SWAT. We use the FAO/UNESCO digital soil map of the world and derived soil properties, revision 1, with a spatial resolution of 5 arc-minutes (FAO-Unesco, 1974). Look-up tables translating GLCC land cover classes and FAO/UNESCO soil types into SWAT parameters have been developed by the WaterBase project (George and Leon, 2007).

The model is forced with daily precipitation and daily minimum and maximum temperature from the National Oceanic and Atmospheric Administration's Global Forecast System (NOAA-GFS) which provides up to seven days of forecast at a six hourly temporal resolution and 0.5 degree spatial resolution (NOAA, 2014). Real-time and recent historical forecasts can be downloaded from the NOMADS server (http://nomads.ncdc.noaa.gov/data.php#hires_weather_datasets, last accessed: 14.01.2015). Historical forecasts older than a few months have to be ordered for FTP download. NOAA-GFS data was aggregated to daily precipitation prior to its use in the hydrological model. For historical simulation periods and model calibration, forcing time series consisting of the 1-day ahead forecasts are used. In operational mode, long-term forecasts are successively replaced with short-term forecasts as time proceeds. In order to assess the performance of the NOAA-GFS precipitation forecast for the Kavango region, the 1-day ahead forecasts were compared to FEWS-RFE rainfall estimates (Herman et al., 1997). FEWS-RFE was previously found to be one of the most accurate remote sensing precipitation products for Africa (Milzow et al., 2011; Stisen and Sandholt, 2010).

Calibration and validation of the hydrologic-hydrodynamic model

Calibration and validation of the hydrologic-hydrodynamic model were performed against observed in situ river discharge using a split-sample approach. The years 2005-2011 were used for calibration, while the years 2012-2014 served as validation period. Mean observed flows in the validation period are higher than in the calibration period (Table 2). After a series of dry years in the beginning of the century, the region has experienced much higher amounts of precipitation and river flow since 2008 (Wolski et al., 2014). In order to ensure a balanced representation of both wet and dry years in the calibration period, we had to use a major portion of the entire data record for calibration and could only reserve three years for validation. Particularly for the station Mohembo, only very few observations are available in the validation period (Table 2). The objective function which was minimized in the calibration was formulated as

$$\varphi = (1 - NSE)^2 + RME^2$$
$$RME = \frac{1}{\overline{Q_{obs}}} \frac{1}{n} \sum_{i=1}^{n} (Q_i - Q_{obs,i}) \qquad (2)$$

where *NSE* is the Nash-Sutcliffe model efficiency (Nash and Sutcliffe, 1970) and *RME* is the relative water balance error (relative mean error). The symbols Q and Q_{obs} denote simulated and observed river discharge, respectively, n is the number of available discharge observations and the overbar indicates temporal averaging. This formulation ensured a reasonable trade-off between fitting the observed hydrographs and matching the observed water balance of the catchment. A sequential calibration strategy was implemented: First, the subcatchments upstream of Rundu were calibrated using Rundu observations and subsequently the subcatchments between Rundu and Mohembo were calibrated using Mohembo observations.

Calibration was performed using the model-independent parameter estimation programme PEST (Doherty et al., 2014). Because of the strongly non-linear response of the SWAT rainfall-runoff model, global derivative-free search strategies are the preferred option for calibration of SWAT models (Arnold et al., 2012). We use the shuffled complex evolution (SCE) algorithm (Duan et al., 1992) which performs a global search over the entire allowed parameter space. The SCE algorithm is included in the PEST package (SCEUA_P).

The selection of calibration parameters was the result of an iterative procedure including extensive sensitivity analysis and repeated trial model runs. The final selection was based on the following principles: (i) spatial variation of vegetation and soil parameters is determined by the input datasets and should be left unchanged during calibration. The corresponding SWAT parameters were either not changed at all or multiplied with a global factor. (ii) The water balance of the rainfall-runoff model should be maintained. Therefore the fraction of the recharge entering the deep aquifer was set to zero. (iii) SWAT groundwater parameters are highly uncertain a priori but at the same time very sensitive. Enough spatial variation in groundwater parameters must be allowed in order to reproduce the various recession time scales in the observed hydrographs. (iv) SWAT has two threshold values of the shallow groundwater storage, one controlling the onset of baseflow and one controlling the onset of phreatic evapotranspiration. The absolute magnitudes of the two threshold values are less important because they mainly control the length of the required model warm-up period. However, the difference between these two threshold values has significant control over the water balance of the catchment: If the baseflow threshold is below the phreatic ET threshold, more water will leave the catchment as baseflow and less as actual ET and vice versa. In order to reduce parameter correlation and non-uniqueness, the baseflow threshold was generally fixed at 100 mm in the Kavango SWAT model.

Table 3 provides an overview of the calibration parameters and their allowed ranges. For the groundwater parameters, spatial variation was allowed between the Rundu and Mohembo regions, the upstream and downstream catchments within each region and the high slope and low slope portions of the land surface. This resulted in a total number of 19 calibration parameters for the Rundu region and 20 calibration parameters for the Mohembo region. We chose 8 complexes in the SCE calibration run and the number of complexes remained the same throughout the run. Both the number of parameter sets in each complex and the number of evolution steps before complex shuffling were set to 39 and 41 for the Rundu and Mohembo regions respectively. The convergence criterion was set to a relative improvement of the best objective function of 1% over 10 shuffling loops. A total of 50000 model runs were allowed, however the calibration converged after 14711 and 18373 model runs for the Rundu and Mohembo regions respectively. After completion of the SCE run, the evolution of the parameter values over the course of the shuffling loops was evaluated. All parameter values converged to a stable solution away from the a priori parameter bounds.

Assimilation strategy

The objective of data assimilation is to combine, at each point in time, the model-based estimate of the state of the system as well as the most recent observations of the state, in order to produce the best possible estimate of the current and future states, taking into account the respective uncertainties of simulated states and observations. The assimilation strategy chosen in this study consists of updating the simulated discharge in the Muskingum routing model only, because the objective was to generate probabilistic river discharge forecasts with lead times of up to 7 days. Updates of the rainfall-runoff model states would probably improve long-term forecasts significantly but may have limited effect on forecasts with short lead times in large basins such as the Kavango basin. Moreover, updating the rainfall-runoff model would require ensemble-based assimilation approaches. For the intended user group of the TIGER-NET products, simplicity and efficiency are key criteria.

Observed in-situ discharge at the station Rundu was assimilated to the model in the operational runs. Because the Muskingum routing operator is linear and the measurement operator is linear too, we could use the standard Kalman filter for state updating, since it is the optimal sequential assimilation method for linear dynamics (Kalman, 1960). The Kalman filter simultaneously updates discharge at all basin outlets. If instead of river discharge, water level measurements from space-borne or ground-based instruments are assimilated, the measurement operator becomes non-linear and the extended Kalman filter can be used (Michailovsky et al., 2013). The reader is referred to the literature (e.g. Jazwinski, 1970) for a detailed discussion of the Kalman filter equations and to Michailovsky et al., 2013 for a detailed description of the assimilation approach.

Description of the model error

Runoff is assumed to be the dominant source of error in the routing model. While the routing model parameters, which depend on reach geometries and Manning's friction factors, are uncertain, runoff uncertainty can be expected to be much more significant due to the error in the NOAA-GFS rainfall forcing as well as structural deficiencies and/or parameterization errors in the SWAT model. In order to find a reasonable representation of the model error, the magnitude, auto-correlation and spatial cross-correlation of the runoff error had to be assessed. No direct measurements of runoff are available within the river basin. To derive an operational error model, we assume, in the baseline experiment, that magnitude and autocorrelation of the relative runoff error are the same as magnitude and autocorrelation of the relative model residuals at the available in-situ discharge stations:

$$\mathbf{w_t} = \frac{(\mathbf{Q_{sim,t}} - \mathbf{Q_{obs,t}})}{\mathbf{Q_{obs,t}}} \tag{3}$$

where w_t is the relative model residual (-), $Q_{sim,t}$ is the modelled discharge at the in-situ discharge station at time step t and $Q_{obs,t}$ is the in-situ discharge as time step t. The autocorrelation of the residuals was assumed to be represented by a first order autoregressive (AR1) model:

$$w_t = \delta w_{t-1} + \varepsilon_t \tag{4}$$

where δ is the AR1 parameter and ε is a sequence of white Gaussian noise with a spatial covariance *Q'*. Due to the correlated meteorological inputs the runoff forcing error was assumed to be spatially correlated between the various subcatchments of the model. In the baseline experiment, we assume that the spatial correlation of the runoff forcing error is equivalent to the spatial correlation of the runoff forcing itself. The correlation matrix of the runoff inputs was computed and Q' was set to:

$$\mathbf{Q}' = \mathbf{C}\,\sigma(\epsilon)^2 \tag{5}$$

where $\mathbf{C}$ is the runoff correlation matrix and $\sigma(\epsilon)^2$ is the variance of the white noise component of the AR1 model. The auto-correlated runoff error state was integrated in the Kalman filter updating scheme by augmenting the model state vector with the correlated noise term (Jazwinski, 1970; Michailovsky et al., 2013). This ensures persistence of assimilation benefits in time.

The major source of error in in-situ discharge observations is the rating curve, which is used to transform readings of river stage into river discharge. Rating curves are particularly unreliable for extreme flow rates and, depending on the channel characteristics, the rating curve changes over time and requires frequent updating. In the absence of detailed information on the in-situ measurement procedure, we assumed the measurement error to be uncorrelated in time and proportional to the discharge. In the baseline experiment, the relative error was assumed to be 10 %, which is a typical value for in-situ discharge derived from rating curves (Di Baldassarre, 2009) and comparable to other hydrologic data assimilation studies (e.g. Clark et al., 2008; Georgakakos, 1986; Weerts and El Serafy, 2006).

In order to evaluate the impact of model error and observation error specifications on the performance of the probabilistic discharge forecasts, four additional forecasting experiments were conducted. Table 4 presents an overview of the experiments. In the baseline experiment, the autocorrelation of the relative runoff error was set equal to the autocorrelation of the relative model error at Rundu (0.9942), as described above. The magnitude of the relative runoff error was set to 4.38%, which is the same as the relative model error at Rundu. The spatial correlation of relative runoff error was set equal to the spatial correlation of runoff and the relative observation error was set to 10%. In experiment 1, the autocorrelation of the runoff error was set equal to the autocorrelation of the spatially aggregated runoff (0.9934) while the other specifications are the same as in the baseline run. In experiment 2, the spatial correlation of the runoff error was set to zero and all other specifications are as in the baseline run. In experiment 3, the runoff error specifications are the same as in the baseline and the relative observation error was set to 20%. Finally, in experiment 4, the white noise component of the relative runoff error was increased from 4.38% to 6% and all other specifications are as in the baseline run.

Operational forecasting and performance evaluation

Operational forecasts have been issued at the daily basis for the validation period and supplied to Namibia's Ministry of Agriculture Water and Forestry for web-based dissemination. A set of criteria were used to assess the performance of the probabilistic river discharge forecasts. Performance assessment was done separately for the open loop model and the 0 to 7-day forecasting horizons. The criteria assess the performance of the central model forecast, as well as the reliability and sharpness of the probabilistic forecasts. The following criteria were used to assess the performance of the central model forecast: Nash-Sutcliffe model efficiency (NSE), root-mean square error (RMSE), mean error (ME) and persistence index. The persistence index (PI, Bennett et al., 2013) is defined analogous to the NSE:

$$PI = \frac{\frac{1}{n}\sum_{i=1}^{n}(Q_i - Q_{obs,i})^2 - \frac{1}{n}\sum_{i=1}^{n}(Q_i - Q_{last})^2}{-\frac{1}{n}\sum_{i=1}^{n}(Q_i - Q_{last})^2} \tag{6}$$

where n is the number of forecasted observations, Q are the forecasts, Q_{obs} are the observations and Q_{last} is the latest available observation before the forecasted observation. While the NSE uses the average of the observations as the benchmark (i.e. a forecast that performs as good as the long-term average of the available observations scores an NSE of 0), the PI uses the last available observation as the benchmark (i.e. a forecast that performs as good as the latest available observation scores a PI of 0).

Reliability and sharpness of the probabilistic forecasts were assessed with the coverage of the 95% confidence interval (i.e. percentage of observations that fall within the predicted nominal 95% confidence interval), the sharpness of the 95% confidence interval (width of predicted 95% confidence interval), the Interval Skill Score (ISS) of the 95% confidence interval as well as the continuous ranked probability score (CRPS). The ISS is defined according to Gneiting and Raftery, 2007 as:

$$ISS_\alpha = \sum_{i=1}^{n} iss_\alpha\left(l_i, u_i, Q_{obs,i}\right)$$

$$iss_\alpha(l, u, Q_{obs}) = \begin{cases} (u-l) & \text{if} \quad l < Q_{obs} < u \\ (u-l) + 2/\alpha\,(l-x) & \text{if} \quad Q_{obs} < l \\ (u-l) + 2/\alpha\,(x-u) & \text{if} \quad Q_{obs} < u \end{cases} \tag{7}$$

where α is the level of the confidence interval (0.05 in our case), l is the lower and u the upper bound of the confidence interval.

The CRPS is a verification tool for probabilistic forecasts and can be interpreted as the area between the cumulative distribution function of the forecast and the cumulative distribution function of the observation, which is a Heaviside step function. The CRPS thus compares the full distribution function of the forecast with the observation and not only selected confidence intervals. For normally distributed forecasts, a closed-form expression for the CRPS exists (Gneiting et al., 2004):

$$\begin{aligned} CRPS &= \frac{1}{n}\sum_{i=1}^{n} crps(Q_{obs,i}, Q_i, \sigma_i) \\ crps(Q_{obs}, Q, \sigma) &= \sigma\left[\frac{Q_{obs}-Q}{\sigma}\left(2\Phi\left(\frac{Q_{obs}-Q}{\sigma}\right)-1\right)+2\phi\left(\frac{Q_{obs}-Q}{\sigma}\right)-\frac{1}{\sqrt{\pi}}\right] \end{aligned} \tag{8}$$

where σ is the standard deviation of the probabilistic forecast, Φ is the cumulative distribution function and ϕ the probability density function of the standard normal distribution. For a deterministic forecast, the CRPS is equivalent to the mean absolute error (Boucher et al., 2011; Schellekens et al., 2011). This allows for a systematic and objective comparison between deterministic and probabilistic forecasts.

The performance of operational forecasts was compared to two benchmark forecasts which can be produced with minimal effort: persistence and climatology. Persistence forecasts the flow as equal to the last available observation, while climatology forecasts the flow as equal to the historical average flow for this day of the year.

Results

Comparison of precipitation products

Comparison of the FEWS-RFE and NOAA-GFS precipitation products showed large deviations between the two products. Figure 2 shows a double mass plot for the average precipitation over the entire Kavango River catchment for the period 2005-2012. Obviously, there is a significant bias and the timing of precipitation events is inconsistent too, as evidenced by the wiggles in the double mass curve. The FEWS-RFE product is based on both satellite observations and in-situ gauging stations, while NOAA-GFS is derived from a global weather model. Moreover, FEWS-RFE has been shown to perform well in previous studies on the African continent (Milzow et al., 2011; Stisen and Sandholt, 2010). We therefore assume that the FEWS-RFE product is closer to the unknown true precipitation than NOAA-GFS and bias correct the NOAA-GFS data to match the long-term average precipitation for both products. A spatially and temporally constant precipitation correction factor of 0.67 was therefore used throughout the study. Figure 2 also presents a quantitative comparison of the NOAA-GFS precipitation forecasts for various forecasting horizons. As a general trend, the longer the forecasting horizon, the lower the predicted precipitation compared to the 1-day ahead forecasts. These effects are particularly pronounced for the rainy seasons 2008/2009 and 2011/2012. However, for the most recent years, the double mass plots show slopes close to unity. We therefore did not implement variable bias correction for the different forecasting horizons. Because the NOAA-GFS system is continuously updated and modified (process parameterization, spatial resolution etc.), performance of precipitation forecasts should be regularly checked during operational application of the hydrologic forecasting system. Changes in the quantitative precipitation forecasts may require adjustments in the bias correction and/or recalibration of the hydrological model.

Clearly, the quality of the precipitation forcing is a critical issue, which has significant control over the performance of the forecasting system. Within the TIGER-NET framework, we are dependent on public domain datasets and NOAA-GFS was the only free source of operational weather forecasts for the African continent available to the project. Potentially, model performance could be improved if NOAA-GFS data was corrected dynamically, for instance by continuously benchmarking it against real-time or near real-time precipitation products such as FEWS-RFE or TRMM-3B42 (Huffman et al., 2007) for the recent past and estimating a time-variable bias correction. An even better solution would be to merge NOAA-GFS data with in-situ precipitation data. However, no operational dataset of in-situ precipitation observations is available for this part of Africa.

Performance of the calibrated model

Table 3 provides an overview of the calibrated parameter values. All parameter values are physically reasonable and calibrated parameter values do not stick to the bounds of a-priori parameter intervals.

Model residuals were analysed and tested for normality and autocorrelation. Figure 5 summarizes the results of the model error analysis for the station Rundu. Figure 5a plots the relative error of the hydrologic-hydrodynamic model versus the observed discharge. Obviously, the relative error is not independent of discharge; it is higher for low discharge than for high discharge. The Q-Q plot in Figure 5b shows that the empirical distribution of model errors significantly deviates from a normal distribution. The empirical distribution of the model errors is narrower than the normal distribution and a larger portion of the data is clustered around the mean. The correlogram in Figure 5c shows highly significant auto-correlation of the model errors. Figure 5d shows the residual model errors (ε) after application of the AR1 model (equation 4), plotted against the observed discharge. This distribution looks more even than the distribution of the primary model residuals in Figure 5a. A test for normality using the Q-Q plot shows significant deviations and again a narrower distribution than the normal distribution (Figure 5e). Temporal correlations have been effectively removed from the model errors and no significant correlations remain as shown in Figure 5f. We conclude from this analysis that the relative error of the hydrologic-hydrodynamic model can be reasonably represented with an AR1 model. The time correlation of the AR1 model is δ=0.9942 on the daily time step. The random error contribution is ε=0.0438. As explained in the methods section, we assume, in the baseline experiment, that the same AR1 model parameters can represent the relative error of the runoff forcing and we use this result to parameterize the model error in the Kalman filter assimilation scheme.

Discharge forecasting and data assimilation

Table 5 reports the performance statistics for the probabilistic model runs. We report results for the open-loop run without assimilation, the assimilation run ("now-casting") as well as the 1-7 day ahead forecasts. The various forecasting horizons use different precipitation forcings (forecasts available at the simulated issue date) and in-situ data are assimilated up to simulated issue date. We only assimilate data from the station Rundu, because (i) no real-time observations are available for Mohembo and (ii) this enables us to assess the effect of upstream assimilation on a downstream station. The indicators are reported for both in-situ stations and for the calibration and the validation period. We are well aware that the observations in the calibration period have been used already for model calibration and are now used again for assimilation. Still, we feel that it is useful to present the statistics for information. Figure 6 shows the open-loop and assimilation run for the station Rundu during calibration and validation periods. We first assess the performance of the probabilistic open-loop run. Generally, the chosen error model seems to be appropriate. The forecasts produced by the open-loop run are reliable; the coverage of the nominal 95% confidence interval does not fall below 84% at any of the stations during any of the periods. However, the open-loop forecasts are not very sharp, as evidenced by the wide confidence intervals in Figure 6. This results in a relatively high ISS score.

The assimilation run is much sharper for all stations and periods but we observe a significant loss of reliability in the validation period. This can again be explained by the relatively low number of observations, particularly at the station Mohembo during the validation period as well as relative over-sampling of the high-flow period. ISS scores of the forecasting runs are much lower than for the open-loop run, which indicates massive improvement. The 1-7 day ahead forecast runs show degrading performance for increasing lead times. However, even the 7-day ahead forecast generally has a lower ISS than the open-loop run, except for Rundu during the validation period. Clearly, the central forecast is better for all lead times than the central run in the open-loop simulation. All three indicators (NSE, RMSE and ME) show significant improvement. Coverage decreases rapidly with increasing lead time for the station Rundu but is more or less independent of lead time for the station Mohembo. This can be explained by the routing time lag between the two stations.

Improvements due to assimilation of Rundu data travel down to Mohembo and are still visible at this station after many days. For the station Rundu, increased sharpness is over-compensated by loss of reliability, which leads to increasing ISS scores with increasing lead time. For the validation period, only the 0-3 ahead forecasts are better than the open-loop run, if evaluated with the ISS score.

Table 6 summarizes the performance of the operational forecasts produced in the different forecasting experiments for the validation period and the station Rundu. Results are reported for the baseline and experiments 1, 3 and 4. Experiment 2 produced results that are very similar to the baseline results and those are therefore not separately reported. Table 6 also includes the performance indicators for the persistence and climatology benchmarks.

Experiment 4 generally shows the best performance. According to the CRPS score, the forecasts are superior to the open-loop run for all forecasting horizons. Forecasts are also better than the persistence benchmarks for forecasting horizons between 4 and 7 days. For forecasting horizons between 1 and 6 days, the model outperforms the climatology benchmark. The persistence index indicates that the forecasting system performs worse than the persistence benchmark. However, it is important to note that the PI does not assess the quality of probabilistic forecasts in terms of sharpness and reliability but only takes the central forecast into account and compares two deterministic predictions.

Figure 7 graphically presents the forecasts produced in experiment 4 for the station Rundu during the validation period and Figure 8 shows predictive quantile-quantile plots for these forecasts.

Discussion

The presented approach for the generation of probabilistic river discharge forecasts is simple and robust and designed to work in data-sparse and poorly gauged basins. A key factor for the performance of the system is the rainfall forcing. While the NOAA-GFS rainfall can produce reasonably reliable and sharp forecasts for the Kavango River, the product should be further compared against other operational precipitation products. Promising avenues for future research may be dynamic bias correction using other precipitation or soil moisture products and/or the extension of the forecast lead time beyond 7 days. NOAA-GFS does provide forecasts up to 16 days into the future. However, the spatial resolution is reduced by a factor of 2 for forecasting horizons beyond one week. To further improve the reliability and sharpness of the forecasts, an ensemble of weather forecasts should be used to drive the forecasting system (Cloke and Pappenberger, 2009). One potential source of free ensemble weather forecasts for the African continent is the Global Ensemble Forecasting System (GEFS, http://www.emc.ncep.noaa.gov/?branch=GEFS).

As in other hydrologic data assimilation studies (e.g. Clark et al., 2008), parameterization of the model error is a fundamental issue for the performance of the assimilation scheme. Generally, model error terms can be added to the forcings, the states, and the parameters of a model. Here, we assign all model error to the runoff forcing and quantify magnitude and auto-correlation of the error based on the comparison of simulated and observed river discharge. Unlike other authors, we do not apply error terms to the states and parameters of the routing model, because we assume that these error contributions are minor compared to the runoff error. While this approach is robust and efficient, it clearly represents a strong simplification of reality. It is clear that the simple Muskingum routing model has significant structural error, for instance due to the fact that floodplains and surface water / groundwater interactions are not simulated.

Comparison of the various forecasting experiments shows that assumptions about the model and observation errors have a large impact on the performance of the forecasting system. The magnitude of the relative runoff error is particularly sensitive, as evidenced by the improved performance of experiment 4 compared to the baseline. It is reasonable to assume a higher relative error for the runoff than the relative error computed from the model residuals at Rundu, because the routing model has a smoothing effect on the runoff response. Experiment 3 and the baseline show a comparable performance in terms of CRPS. Basically the higher assumed observation error in experiment 3 results in predictions that are less sharp but more reliable. Comparison of experiment 1 and baseline results shows that even small differences in the assumed autocorrelation of the runoff error result in significant differences in the forecast performance. Higher error autocorrelation leads to increased sharpness, but lower reliability. CRPS indicates that experiment 1 forecasts marginally outperform the baseline forecasts. Experiment 2 results are very close to the baseline, because the spatial correlation of runoff between the different subcatchments is low, due to the variable hydrologic characteristics of the subcatchments. Predictive Q-Q plots for experiment 4 (Figure 8) indicate significant deviations of the empirical distribution of normalized forecast errors from the normal distribution.

As is common for studies dealing with probabilistic river discharge forecasting, we find that our probabilistic forecasts are over-reliable during low flow periods and under-reliable during high-flow periods. This issue can be addressed by separating the total runoff forcing generated by the SWAT model into its components, i.e. overland flow, interflow and baseflow, and developing separate error representations for the various runoff components. However, given the sparse availability of in-situ observations in the basins, it may be difficult to find robust parameters for these error representations.

We generally observe weaker performance of the forecasting system in the beginning of the rainy season, i.e. after the long dry season during the onset of the annual high-flow season. This may be due to deficiencies in the precipitation forecasts and/or due to weaknesses in the representation of hydrological processes in the SWAT model. It appears that in reality, the first rains in the early rainy season already lead to increased river flow, while in the model, these precipitation events are completely absorbed in the various simulated hydrological storage compartments.

In this study, focus has been on the final output of the modelling chain, i.e. river discharge. However, SWAT simulates a multitude of intermediate states and fluxes in the land phase of the hydrological cycle, which could be analysed and compared to observations, if such observations were available. There is an obvious opportunity to inform the modelling system with other types of in-situ and remote sensing observations such as radar altimetry, soil moisture and total water storage from time-variable gravity (Milzow et al., 2011). However, if such data were to be formally assimilated to the modelling system, an ensemble approach would have to be chosen because of the highly non-linear responses inherent in the SWAT model. Many studies have addressed ensemble-based streamflow forecasting with lumped-conceptual or distributed hydrological models. Rakovec et al., 2012 found that rainfall-runoff model states were less sensitive compared to routing states in their hydrologic data assimilation study with the Ensemble Kalman Filter and suggested time lags between the rainfall-runoff model states and streamflow response as the likely reason. Alternative updating strategies that use several previous time steps instead of the last time step only (e.g. Ensemble Kalman Smoother) can potentially solve these problems. Other recurring issues in such studies are high computational demand, and model error parameterization (e.g. Clark et al., 2008).

Conclusions

We have presented an operational probabilistic river discharge forecasting system for poorly gauged basins which relies exclusively on public-domain, open-source software and data. The forecasting system is specifically adapted to the conditions prevailing in many African basins, such as weak in-situ monitoring infrastructure, budget constraints for operational monitoring and management as well as weak institutional capacity. We demonstrated the performance of the forecasting system for the Kavango River and obtained encouraging results. Zero to 7-day ahead probabilistic forecasts produced by the system are sharp and reliable. The system may benefit from ingestion of other types of in-situ or remotely sensed observations such as radar altimetry and soil moisture. The TIGER-NET project and its Water Observation and Information System (WOIS) provide an ideal platform to combine remote sensing observations and hydrological models to generate accurate estimates of hydrological states as well as sharp and reliable forecasts for operational water resources management.

Acknowledgements: We acknowledge funding from the European Space Agency (ESA) through the TIGER-NET project. Real-time and historical in-situ observations for the station Rundu were provided by Namibia's Ministry of Agriculture Water and Forestry. Historical in-situ observations for the station Mohembo were provided by Botswana's Department of Water Affairs.

References

Arnold, J.G., Moriasi, D.N., Gassman, P.W., Abbaspour, K.C., White, M.J., Srinivasan, R., Santhi, C., Harmel, R.D., van Griensven, A., Van Liew, M.W., Kannan, N., Jha, M.K., 2012. SWAT: MODEL USE, CALIBRATION, AND VALIDATION. Trans. ASABE 55, 1491–1508.

Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H. a., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S. a., Robson, B., Seppelt, R., Voinov, A. a., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. Environ. Model. Softw. 40, 1–20. doi:10.1016/j.envsoft.2012.09.011

Berry, P.A.M., Smith, R.G., Benveniste, J., 2010. ACE2: The New Global Digital Elevation Model, in: Mertikas, SP (Ed.), GRAVITY, GEOID AND EARTH OBSERVATION, International Association of Geodesy Symposia. pp. 231–237. doi:10.1007/978-3-642-10634-7_30

Biancamaria, S., Durand, M., Andreadis, K.M., Bates, P.D., Boone, a., Mognard, N.M., Rodríguez, E., Alsdorf, D.E., Lettenmaier, D.P., Clark, E. a., 2011. Assimilation of virtual wide swath altimetry to improve Arctic river modeling. Remote Sens. Environ. 115, 373–381. doi:10.1016/j.rse.2010.09.008

Boucher, M. -a., Anctil, F., Perreault, L., Tremblay, D., 2011. A comparison between ensemble and deterministic hydrological forecasts in an operational context. Adv. Geosci. 29, 85–94. doi:10.5194/adgeo-29-85-2011

Chow, V. Te, Maidment, D.R., Mays, L.W., 1988. Applied Hydrology, Water Resources and Environmental Engineering. McGraw-Hill, New York.

Clark, M.P., Rupp, D.E., Woods, R. a., Zheng, X., Ibbitt, R.P., Slater, A.G., Schmidt, J., Uddstrom, M.J., 2008. Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. Adv. Water Resour. 31, 1309–1324. doi:10.1016/j.advwatres.2008.06.005

Cloke, H.L., Pappenberger, F., 2009. Ensemble flood forecasting: A review. J. Hydrol. 375, 613–626. doi:10.1016/j.jhydrol.2009.06.005

Di Baldassarre, G., 2009. Uncertainty in river discharge observations: a quantitative analysis. Hydrol. EARTH Syst. Sci. 13, 913 – 921.

Doherty, J., Muffels, C., Rumbaugh, J., Tonkin, M., 2014. PEST. Model independent parameter estimation and uncertainty analysis [WWW Document]. URL http://www.pesthomepage.org/Home.php (accessed 7.16.14).

Duan, Q.Y., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resour. Res. 28, 1015–1031. doi:10.1029/91WR02985

FAO-Unesco, 1974. Soil map of the world 1:5000000. Paris.

Fekete, B.M., Voeroesmarty, C.J., 2007. The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements, in: Proceedings of the PUB Kick-off Meeting. IAHS publication 309, Brasilia.

Folwell, S., Farqhuarson, F., 2006. The impacts of climate change on water resources in the Okavango basin, in: Demuth, S and Gustard, A and Planos, E and Scatena, F and Servat, E (Ed.), Climate Variability and Change - Hydrological Impacts, IAHS PUBLICATION. pp. 382–388.

Gassman, P.W., Reyes, M.R., Green, C.H., Arnold, J.G., 2005. SWAT Peer-Reviewed Literature : A Review. Hydrol. Process. 13, 1–17.

Georgakakos, K.P., 1986. A generalized stochastic hydrometeorological model for flood and flash-flood forecasting. 2. case studies. Water Resour. Res. 22, 2096–2106. doi:10.1029/WR022i013p02096

George, C., Leon, L.F., 2007. WaterBase : SWAT in an open source GIS. Open Hydrol. J. 1, 19–24.

Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. J. Am. Stat. Assoc. 102, 359–378. doi:10.1198/016214506000001437

Gneiting, T., Westveld III, A.H., Raftery, A.E., Goldman, T., 2004. Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation ∗. Seattle, Washington.

Guzinski, R., Kass, S., Huber, S., Bauer-Gottwein, P., Jensen, I.H., Naeimi, V., Doubkova, M., Walli, A., Tottrup, C., 2014. A Water Observation and Information System for Integrated Water Resource Management in Africa. Remote Sens. under revi. doi:10.3390/rs60x000x

Herman, A., Kumar, V.B., Arkin, P.A., Kousky, J. V, 1997. Objectively determined 10-day African rainfall estimates created for famine early warning systems. Int. J. Remote Sens. 18, 2147–2159. doi:10.1080/014311697217800

Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Wolff, D.B., Adler, R.F., Gu, G., Hong, Y., Bowman, K.P., Stocker, E.F., 2007. The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales. J. Hydrometeorol. 8, 38–55. doi:10.1175/JHM560.1

Hughes, D. a., Andersson, L., Wilk, J., Savenije, H.H.G., 2006. Regional calibration of the Pitman model for the Okavango River. J. Hydrol. 331, 30–42. doi:10.1016/j.jhydrol.2006.04.047

Hughes, D. a., Kingston, D.G., Todd, M.C., 2011. Uncertainty in water resources availability in the Okavango River basin as a result of climate change. Hydrol. Earth Syst. Sci. 15, 931–941. doi:10.5194/hess-15-931-2011

Jazwinski, A.H., 1970. Stochastic Processes and Filtering Theory. Academic Press, New York.

Kalman, R.E., 1960. A New Approach to Linear Filtering and Prediction Problems. J. basic Eng. 82, 35–45.

Kgathi, D.L., Kniveton, D., Ringrose, S., Turton, a. R., Vanderpost, C.H.M., Lundqvist, J., Seely, M., 2006. The Okavango; a river supporting its people, environment and economic development. J. Hydrol. 331, 3–17. doi:10.1016/j.jhydrol.2006.04.048

Liu, Y., Weerts, a. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, a. I.J.M., van Velzen, N., He, M., Lee, H., Noh, S.J., Rakovec, O., Restrepo, P., 2012. Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. Hydrol. Earth Syst. Sci. 16, 3863–3887. doi:10.5194/hess-16-3863-2012

Madsen, H., Skotner, C., 2005. Adaptive state updating in real-time river flow forecasting—a combined filtering and error forecasting procedure. J. Hydrol. 308, 302–312. doi:10.1016/j.jhydrol.2004.10.030

Maier, H.R., Jain, A., Dandy, G.C., Sudheer, K.P., 2010. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. Environ. Model. Softw. 25, 891–909. doi:10.1016/j.envsoft.2010.02.003

McCarthy, T.S., Cooper, G.R.J., Tyson, P.D., Ellery, W.N., 2000. Seasonal flooding in the Okavango Delta, Botswana - recent history and future prospects. S. Afr. J. Sci. 96, 25–33.

Michailovsky, C.I., Milzow, C., Bauer-Gottwein, P., 2013. Assimilation of radar altimetry to a routing model of the Brahmaputra River. Water Resour. Res. 49, 4807–4816. doi:10.1002/wrcr.20345

Milzow, C., Kgotlhang, L., Bauer-Gottwein, P., Meier, P., Kinzelbach, W., 2009. Regional review: the hydrology of the Okavango Delta, Botswana—processes, data and modelling, Hydrogeology Journal. doi:10.1007/s10040-009-0436-0

Milzow, C., Krogh, P.E., Bauer-Gottwein, P., 2011. Combining satellite radar altimetry, SAR surface soil moisture and GRACE total storage changes for hydrological model calibration in a large poorly gauged catchment. Hydrol. Earth Syst. Sci. 15, 1729–1743. doi:10.5194/hess-15-1729-2011

Moradkhani, H., Hsu, K.-L., Gupta, H., Sorooshian, S., 2005. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. Water Resour. Res. 41, n/a–n/a. doi:10.1029/2004WR003604

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I - a discussion of principles. J. Hydrol. 10, 282–290.

Neal, J., Schumann, G., Bates, P., Buytaert, W., Matgen, P., Pappenberger, F., 2009. A data assimilation approach to discharge estimation from space 3649, 3641–3649. doi:10.1002/hyp

Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., 2011. Soil & Water Assessment Tool Theoretical Documentation Version 2009.

NOAA, 2014. GFS Global Forecast System [WWW Document]. URL http://www.emc.ncep.noaa.gov/index.php?branch=GFS (accessed 7.16.14).

Pagano, T.C., Wood, A.W., Ramos, M.-H., Cloke, H.L., Pappenberger, F., Clark, M.P., Cranston, M., Kavetski, D., Mathevet, T., Sorooshian, S., Verkade, J.S., 2014. Challenges of Operational River Forecasting. J. Hydrometeorol. 15, 1692–1707. doi:10.1175/JHM-D-13-0188.1

Pauwels, V.R.N., De Lannoy, G.J.M., 2009. Ensemble-based assimilation of discharge into rainfall-runoff models: A comparison of approaches to mapping observational information to state space. Water Resour. Res. 45, n/a–n/a. doi:10.1029/2008WR007590

Peterson, T.C., Vose, R.S., 1997. An overview of the global historical climatology network temperature database. Bull. Am. Meteorol. Soc. 78, 2837–2849. doi:10.1175/1520-0477(1997)078<2837:AOOTGH>2.0.CO;2

Rakovec, O., Weerts, a. H., Hazenberg, P., Torfs, P.J.J.F., Uijlenhoet, R., 2012. State updating of a distributed hydrological model with Ensemble Kalman Filtering: effects of updating frequency and observation network density on forecast accuracy. Hydrol. Earth Syst. Sci. 16, 3435–3449. doi:10.5194/hess-16-3435-2012

Schellekens, J., Weerts, a. H., Moore, R.J., Pierce, C.E., Hildon, S., 2011. The use of MOGREPS ensemble rainfall forecasts in operational flood forecasting systems across England and Wales. Adv. Geosci. 29, 77–84. doi:10.5194/adgeo-29-77-2011

Seo, D.-J., Cajina, L., Corby, R., Howieson, T., 2009. Automatic state updating for operational streamflow forecasting via variational data assimilation. J. Hydrol. 367, 255–275. doi:10.1016/j.jhydrol.2009.01.019

Seo, D.-J., Koren, V., Cajina, N., 2003. Real-Time Variational Assimilation of Hydrologic and Hydrometeorological Data into Operational Hydrologic Forecasting. J. Hydrometeorol. 4, 627–641.

Stisen, S., Sandholt, I., 2010. Evaluation of remote-sensing-based rainfall products through predictive capability in hydrological runoff modelling. Hydrol. Process. 24, 879–891. doi:10.1002/hyp.7529

Tang, Q., Gao, H., Lu, H., Lettenmaier, D.P., 2009. Remote sensing: hydrology. Prog. Phys. Geogr. 33, 490–509. doi:10.1177/0309133309346650

USGS, 2008. Global Land Cover Characteristics Data Base Version 2.0 [WWW Document]. URL http://edc2.usgs.gov/glcc/glcc.php (accessed 7.16.14).

Weerts, A.H., El Serafy, G.Y.H., 2006. Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. Water Resour. Res. 42, n/a–n/a. doi:10.1029/2005WR004093

Wolski, P., Stone, D., Tadross, M., Wehner, M., Hewitson, B., 2014. Attribution of floods in the Okavango basin, Southern Africa. J. Hydrol. 511, 350–358. doi:10.1016/j.jhydrol.2014.01.055

Tables

Table 1: Characteristics of the Kavango River basin and the Rundu and Mohembo sub-basins

Sub-basin	Catchment area (km^2)	Mean elevation (mamsl)	Mean annual precipitation (bias-corrected 1-day ahead NOAA-GFS, mm)
Kavango	162050	1320	847
Rundu	101520	1341	843
Mohembo	60530	1286	853

Table 2: Model performance for calibration and validation periods. Numbers in brackets are percent of mean observed flow.

In-situ station	NSE (-)	RMSE (m^3/s)	ME (m^3/s)	Mean of observations (m^3/s)	No. of simulated observations
Calibration Period (2005-2011)					
Rundu	0.73	105.6 (42.5%)	-5.4 (-2.2%)	248.4	2440
Mohembo	0.69	97.1 (32.8%)	6.8 (2.3%)	295.9	1935
Validation Period (2012-2014)					
Rundu	0.74	94.6 (35.0%)	-55.0 (-20.6%)	249.0	572
Mohembo	0.33	144.0 (30.7%)	-119.0 (-25.4%)	469.1	46

Table 3: Model calibration parameters. Subcatchment IDs for the various regions: r = 2+3+5+6+7+9+10; m = 1+4+8+11+12; ru = 2+3; rd = 5+6+7+9+10; mu = 1; md = 4+8+11+12; ruh = HRUs in region ru with terrain slope above 2%; rul = HRUs in region ru with terrain slope below 2%; rdh = HRUs in region rd with terrain slope above 2%; rdl = HRUs in region rd with terrain slope below 2%; muh = HRUs in region mu with terrain slope above 2%; mul = HRUs in region mu with terrain slope below 2%; mdh = HRUs in region md with terrain slope above 2%; mdl= HRUs in region md with terrain slope below 2%.

Parameter	Description and unit	Lower bound	Calibrated value		Upper bound
CN2_m	Multiplier on the SCS curve number for moisture condition II (dimensionless)	0.6	r	0.63	1.2
			m	0.65	
ESCO	Soil evaporative compensation factor (dimensionless)	0.5	r	0.95	1
			m	0.80	
EPCO	Plant uptake compensation factor (dimensionless)	0.5	r	0.89	1
			m	0.92	
CH_N1	Manning's n for tributary channels ($sm^{-1/3}$)	0.02	r	0.185	0.2
			m	0.023	
CH_N2	Manning's n for main reaches ($sm^{-1/3}$)	0.02	r	0.023	0.2
			m	0.104	
GW_DELAY	Groundwater delay (days)	30	ru	81.3	120
			rd	43.4	
			mu	101.6	
			md	112.8	
ALPHA_BF	Base flow recession constant (dimensionless)	0.05	ruh	0.676	1
			rul	0.177	
			rdh	0.221	
			rdl	0.730	
			muh	0.846	
			mul	0.264	
			mdh	0.161	
			mdl	0.080	
GW_REVAP	Groundwater re-evaporation coefficient (dimensionless)	0	ruh	0.81	1
			rul	0.90	
			rdh	0.68	
			rdl	0.53	
			muh	0.75	
			mul	0.86	
			mdh	0.90	
			mdl	0.26	
REVAPMN	Threshold depth of water in shallow aquifer for re-evaporation to occur (mm)	0	ruh	103	200
			rul	29	
			rdh	75	
			rdl	31	
			muh	15	
			mul	100	
			mdh	97	
			mdl	26	
LOSS_11	Fractional loss from the Kavango River between Rundu and Mohembo, due to evaporation, infiltration and abstraction (dimensionless)	0	0.011		0.2

Table 4: Overview of the different forecasting experiments

Experiment	**Autocorrelation of relative runoff error**	**Relative runoff error**	**Spatial correlation of relative runoff error**	**Relative observation error**
Baseline	Same as autocorrelation of model error at Rundu (0.9942)	4.38%	Same as spatial correlation of runoff	10%
Experiment 1	Same as autocorrelation of total runoff (0.9934)	4.38%	Same as spatial correlation of runoff	10%
Experiment 2	Same as autocorrelation of model error at Rundu (0.9942)	4.38%	Zero	10%
Experiment 3	Same as autocorrelation of model error at Rundu (0.9942)	4.38%	Same as spatial correlation of runoff	20%
Experiment 4	Same as autocorrelation of model error at Rundu (0.9942)	6%	Same as spatial correlation of runoff	10%

Table 5: Performance of the operational model in the calibration and validation periods

Period	In-situ station	Run	NSE (-)	RMSE (m^3/s)	ME (m^3/s)	Coverage (%)	Sharpness (m^3/s)	Interval Skill Score (m^3/s)	Mean of predicted observations (m^3/s)	No. of predicted observations
Calibration Period (2005-2011)	Rundu	Open-Loop	0.73	105.6	-5.4	90.0	423.5	654.9	248.4	2440
		Assimilation	0.99	22.9	-0.9	88.6	54.1	147.1	248.4	2440
		1-day ahead	0.98	29.2	-0.3	86.7	64.4	196.3	248.5	2440
		2-day ahead	0.97	36.5	0.5	85.8	75.6	250.8	248.7	2439
		3-day ahead	0.95	44.0	1.3	84.5	86.7	307.5	248.9	2438
		4-day ahead	0.94	51.2	2.2	83.6	97.2	362.0	249.1	2437
		5-day ahead	0.92	57.9	3.1	83.3	106.9	415.2	249.3	2436
		6-day ahead	0.90	64.1	4.0	82.6	115.8	465.5	249.4	2435
		7-day ahead	0.88	69.9	4.9	81.9	124.0	511.5	249.6	2434
	Mohembo	Open-Loop	0.69	97.1	6.8	93.3	478.2	638.1	295.9	1935
		Assimilation	0.93	45.1	-11.3	93.3	154.5	251.2	295.9	1935
		1-day ahead	0.93	45.2	-11.2	93.3	154.5	251.7	295.9	1935
		2-day ahead	0.93	45.1	-11.1	93.4	154.6	249.3	296.0	1934
		3-day ahead	0.93	45.0	-11.0	93.4	154.7	246.9	296.0	1933
		4-day ahead	0.93	44.9	-10.9	93.5	154.8	244.7	296.1	1932
		5-day ahead	0.93	44.8	-10.8	93.5	154.9	242.4	296.2	1931
		6-day ahead	0.93	44.8	-10.6	93.4	155.2	240.2	296.3	1930
		7-day ahead	0.93	45.0	-10.4	93.3	155.5	238.4	296.4	1929
Validation Period (2012-2014)	Rundu	Open-Loop	0.74	94.6	-55.0	83.9	224.6	515.9	249.0	572
		Assimilation	0.97	31.7	-0.5	81.8	43.6	265.7	249.0	572
		1-day ahead	0.96	39.3	0.5	78.8	49.3	351.4	252.5	556
		2-day ahead	0.94	47.3	1.5	75.9	54.9	442.4	254.1	547
		3-day ahead	0.92	54.8	2.3	74.6	60.1	527.4	254.0	544
		4-day ahead	0.89	61.6	3.1	72.4	65.1	609.9	254.2	540
		5-day ahead	0.87	67.5	3.7	70.8	69.9	687.6	254.9	534
		6-day ahead	0.86	72.3	4.2	69.5	74.2	750.4	254.8	531
		7-day ahead	0.84	76.0	4.4	69.0	78.2	799.6	254.4	529
	Mohembo	Open-Loop	0.33	144.0	-119	93.5	498.4	686.7	469.1	46
		Assimilation	0.92	48.4	-9.0	80.4	176.3	206.5	469.1	46
		1-day ahead	0.92	48.7	-7.6	81.8	178.3	209.5	478.9	44
		2-day ahead	0.92	49.0	-8.0	82.2	177.3	208.2	473.4	45
		3-day ahead	0.92	49.9	-7.4	81.8	178.5	210.6	480.4	44
		4-day ahead	0.91	51.2	-7.5	79.5	178.6	213.6	481.4	44
		5-day ahead	0.91	52.3	-6.9	79.5	178.9	218.0	481.1	44
		6-day ahead	0.91	52.7	-7.8	76.6	176.4	233.0	464.2	47
		7-day ahead	0.92	52.1	-8.4	79.2	175.2	255.7	449.0	48

Table 6: Performance indicators for the forecasts issued for the station Rundu in the validation period, excluding model "warm-up" periods

Run	NSE (-)	RMSE (m^3/s)	Coverage (%)	Sharpness (m^3/s)	Interval Skill Score (m^3/s)	Persistence index (-)	CRPS (m^3/s)	No. of predicted observations
Benchmarks								
Persistence, 1-day ahead	1.00	10.3					6.3	556
Persistence, 2-day ahead	0.99	18.4					12.1	547
Persistence, 3-day ahead	0.98	26.7					17.6	544
Persistence, 4-day ahead	0.97	34.7					23.2	540
Persistence, 5-day ahead	0.95	42.6					28.5	534
Persistence, 6-day ahead	0.93	50.2					33.6	531
Persistence, 7-day ahead	0.91	57.4					38.5	529
Climatology	0.82	78.5	100	346.1	346.1		28.2	580
Baseline								
Open-Loop	0.74	94.6	83.9	224.6	515.9		40.0	572
Assimilation	0.97	31.7	81.8	43.6	265.7		13.1	572
1-day ahead	0.96	39.3	78.8	49.3	351.4	-13.7	16.7	556
2-day ahead	0.94	47.3	75.9	54.9	442.4	-5.6	20.3	547
3-day ahead	0.92	54.8	74.6	60.1	527.4	-3.2	23.8	544
4-day ahead	0.89	61.6	72.4	65.1	609.9	-2.1	27.1	540
5-day ahead	0.87	67.5	70.8	69.9	687.6	-1.5	30.1	534
6-day ahead	0.86	72.3	69.5	74.2	750.4	-1.1	32.7	531
7-day ahead	0.84	76.0	69.0	78.2	799.6	-0.8	34.9	529
Experiment 1								
Open-Loop	0.74	94.6	89.9	295.6	473.0		38.4	572
Assimilation	0.98	25.8	87.6	49.5	189.4		10.0	572
1-day ahead	0.97	33.9	84.5	57.7	261.2	-9.9	13.4	556
2-day ahead	0.95	42.6	83.4	66.2	339.7	-4.4	16.9	547
3-day ahead	0.93	50.9	82.2	74.2	416.1	-2.7	20.4	544
4-day ahead	0.90	58.5	81.5	81.8	485.6	-1.8	23.7	540
5-day ahead	0.88	65.2	80.9	89.1	549.1	-1.3	26.8	534
6-day ahead	0.86	70.5	79.5	95.6	599.5	-1.0	29.4	531
7-day ahead	0.85	74.7	78.8	101.5	635.6	-0.7	31.6	529
Experiment 3								
Open-Loop	0.74	94.6	91.3	315.5	464.8		38.1	572
Assimilation	0.96	39.2	85.8	74.9	261.1		15.6	572
1-day ahead	0.94	46.1	83.8	82.2	323.1	-19.2	18.7	556
2-day ahead	0.92	53.2	82.8	89.2	385.9	-7.3	21.8	547
3-day ahead	0.90	59.7	81.6	95.7	441.1	-4.0	24.6	544
4-day ahead	0.88	65.7	81.1	101.9	493.5	-2.6	27.2	540
5-day ahead	0.86	70.9	80.9	108.0	539.4	-1.8	29.7	534
6-day ahead	0.84	75.1	80.0	113.4	571.2	-1.2	31.7	531
7-day ahead	0.83	78.5	79.6	118.5	595.9	-0.9	33.3	529
Experiment 4								
Open-Loop	0.74	94.6	95.3	432.2	525.3		38.6	572
Assimilation	0.99	20.5	91.1	55.6	141.6		7.7	572
1-day ahead	0.98	29.0	89.4	67.5	202.4	-7.0	10.8	556
2-day ahead	0.96	38.4	88.5	80.1	269.0	-3.4	14.3	547
3-day ahead	0.94	47.7	88.6	92.1	335.2	-2.2	17.8	544
4-day ahead	0.91	56.2	87.8	103.6	397.8	-1.6	21.1	540
5-day ahead	0.89	63.8	86.5	114.4	454.0	-1.2	24.2	534
6-day ahead	0.87	69.8	85.7	123.9	497.8	-0.9	26.8	531
7-day ahead	0.85	74.6	85.6	132.7	531.6	-0.7	29.0	529

Figures

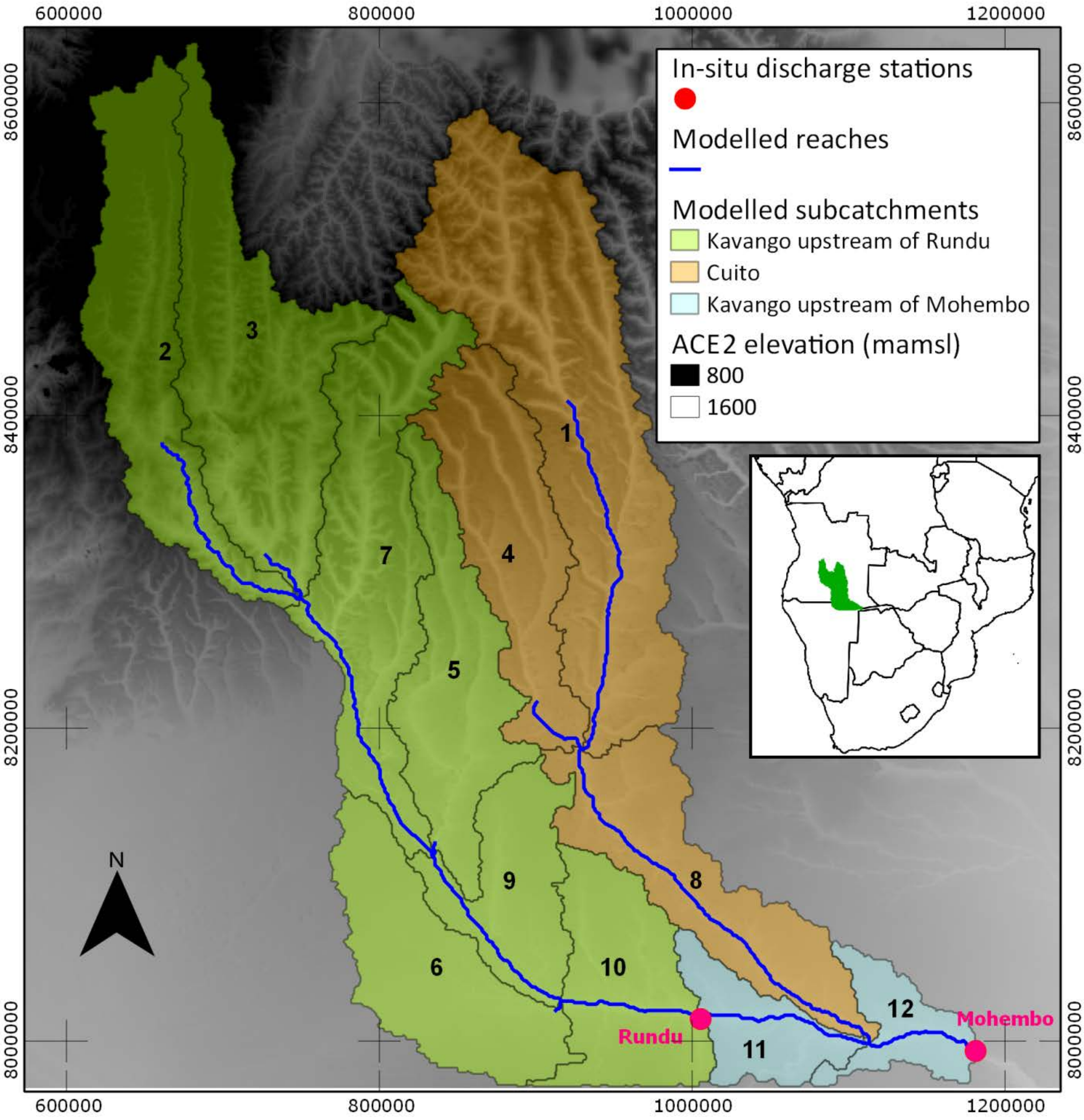

Figure 1: Basemap for the Kavango River Basin with location of in-situ discharge stations. The coordinate system is UTM 33S, WGS84 datum. Inset map shows the location of the basin in Southern Africa.

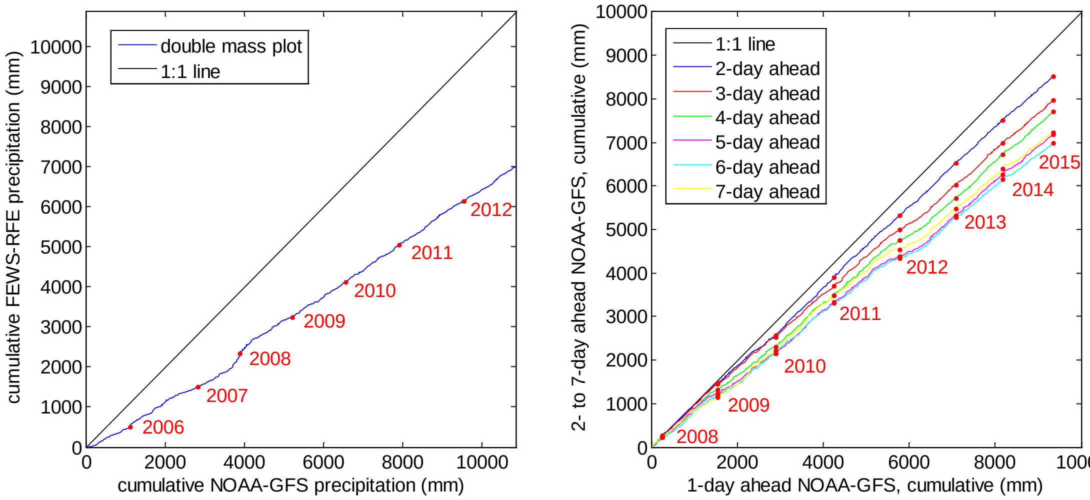

Figure 2: Left: Double mass plot of the FEWS-RFE and NOAA-GFS precipitation products averaged over the entire Kavango River basin. Right: Double mass plots of the 1-day ahead forecasted NOAA-GFS precipitation and the 2-7 day ahead forecasted NOAA-GFS precipitation averaged over the entire Kavango River basin.

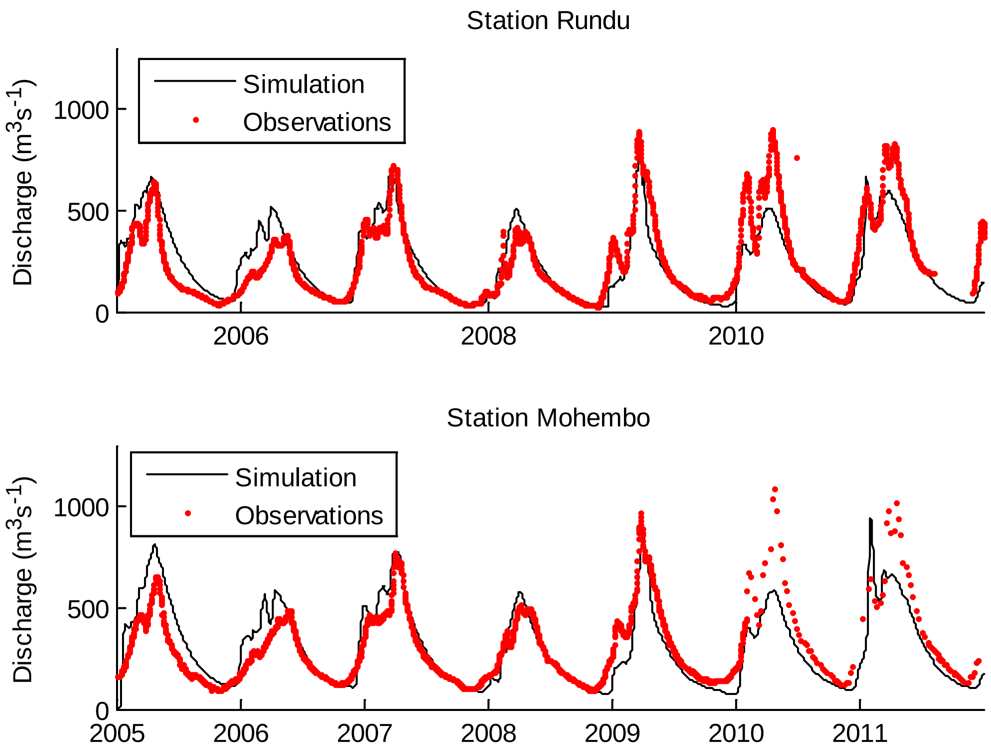

Figure 3: Observed (red dots) and simulated (black lines) hydrographs for the calibration period for Rundu (top) and Mohembo (bottom).

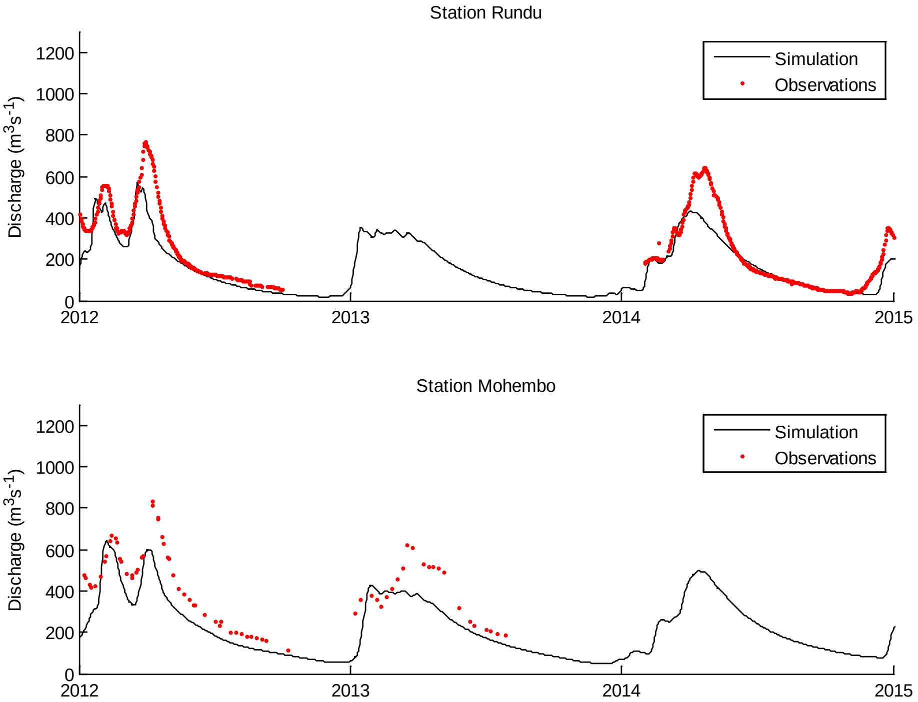

Figure 4: Observed (red dots) and simulated (black lines) hydrographs for the validation period for Rundu (top) and Mohembo (bottom).

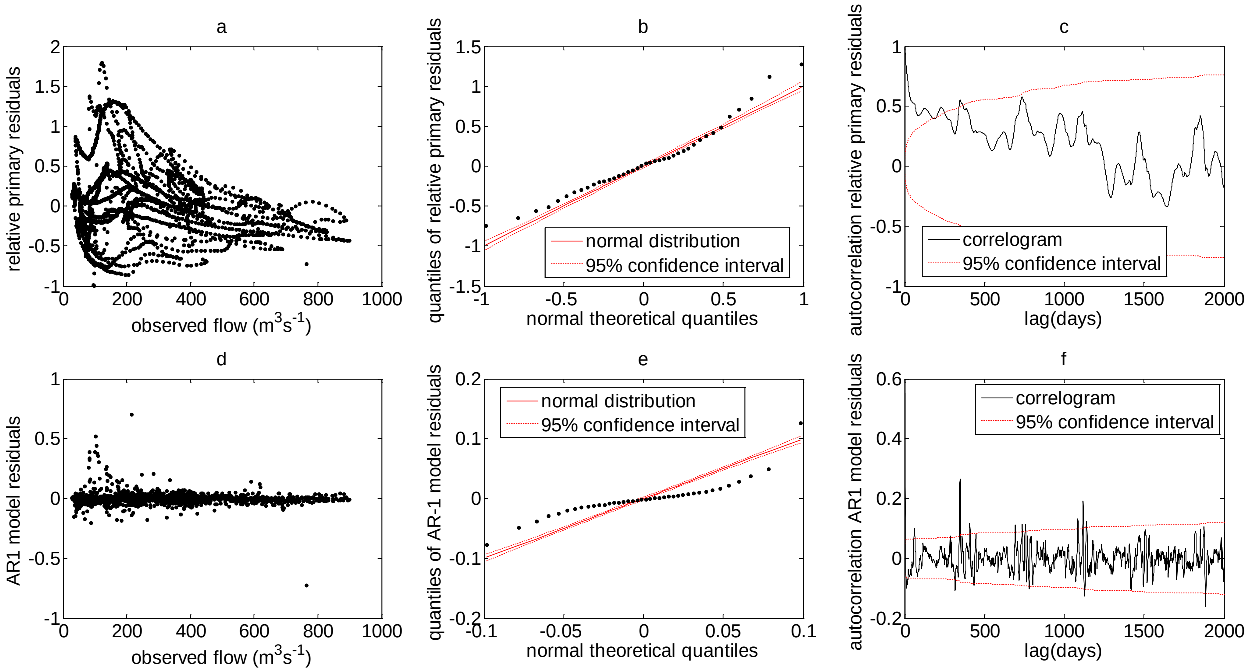

Figure 5: a) Relative error of the hydrologic-hydrodynamic model vs observed discharge. b) Q-Q plot of the relative errors shown in a). c) Correlogram of the relative errors shown in a). d) Relative errors of hydrologic-hydrodynamic model after removal of the time-correlated part plotted vs observed discharge. e) Q-Q plot of the relative errors shown in d). f) Correlogram of the relative errors shown in d).

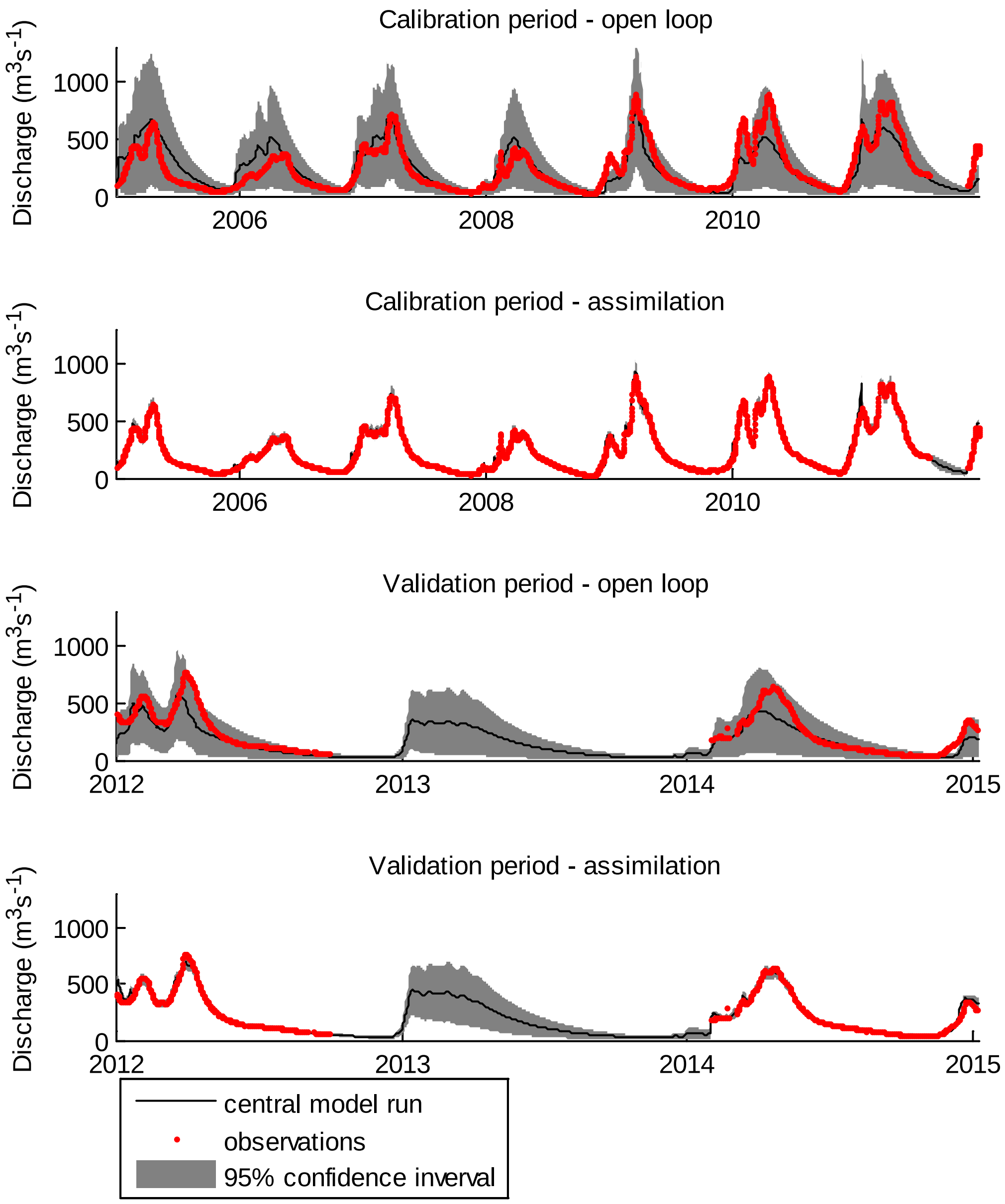

Figure 6: Probabilistic simulation of river discharge in the open-loop and assimilated run for the calibration and the validation periods for the station Rundu.

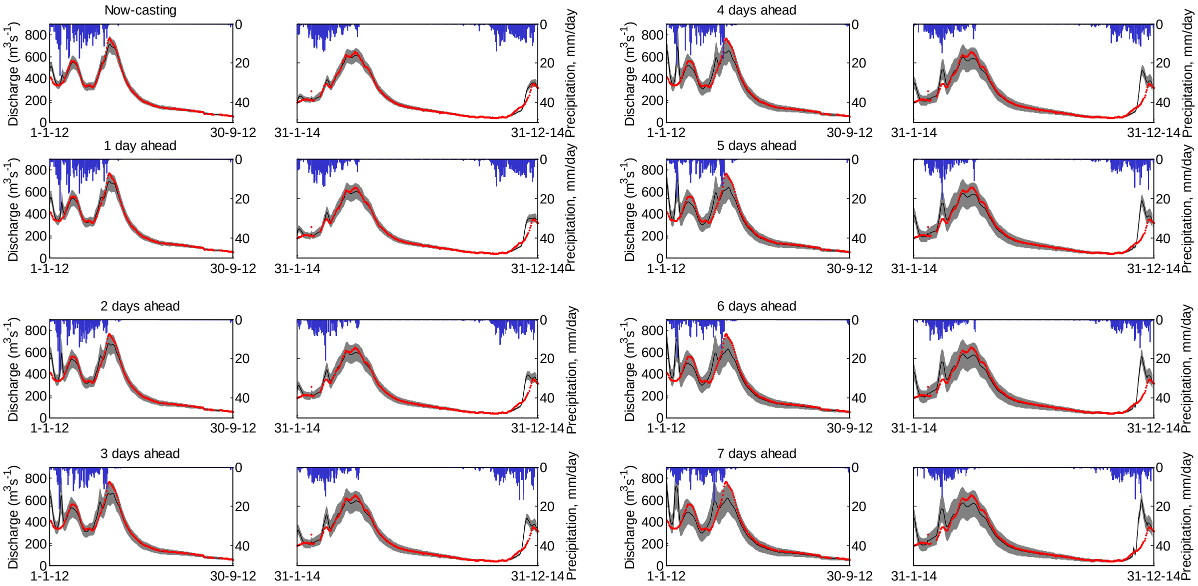

Figure 7: Performance of the 0-7 day ahead probabilistic forecasts in the validation period at Rundu station for experiment 4. The black solid line is the central forecast. Grey shading indicates the 95% confidence interval of the forecast and red dots are observations. Blue bars indicate daily forecasted precipitation from NOAA-GFS.

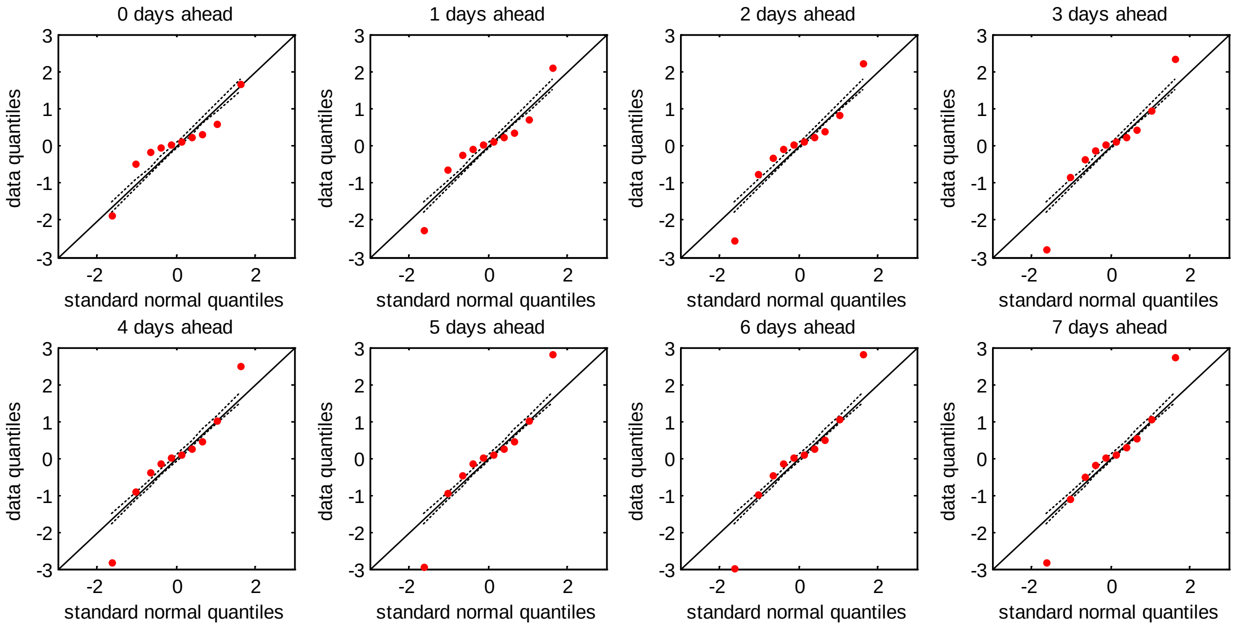

Figure 8: Predictive Q-Q plots for the station Rundu and the validation period for experiment 4.