

1 **Operational river discharge forecasting in poorly gauged basins: the** 2 **Kavango River Basin case study**

3 Peter Bauer-Gottwein^{1*}, Iris. H. Jensen¹, Radoslaw Guzinski², Gudny K. T. Bredtoft¹, Sidsel
4 Hansen¹, and Claire I. Michailovsky^{1#}

5 1) Department of Environmental Engineering, Technical University of Denmark, 2800 Kgs.
6 Lyngby, Denmark

7 2) DHI GRAS, DK-2970 Hørsholm, Denmark

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

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

10 **Abstract**

11 Operational probabilistic forecasts of river discharge are essential for effective water resources
12 management. Many studies have addressed this topic using different approaches ranging from
13 purely statistical black-box approaches to physically-based and distributed modelling schemes
14 employing data assimilation techniques. However, few studies have attempted to develop
15 operational probabilistic forecasting approaches for large and poorly gauged river basins. The
16 objective of this study is to develop open-source software tools to support hydrologic forecasting
17 and integrated water resources management in Africa. We present an operational probabilistic
18 forecasting approach which uses public-domain climate forcing data and a hydrologic-
19 hydrodynamic model which is entirely based on open-source software. Data assimilation techniques
20 are used to inform the forecasts with the latest available observations. Forecasts are produced in real
21 time for lead times of 0 to 7 days. The operational probabilistic forecasts are evaluated using a

22 selection of performance statistics and indicators and the performance is compared to persistence
23 and climatology benchmarks. The forecasting system delivers useful forecasts for the Kavango
24 River, which are reliable and sharp. Results indicate that the value of the forecasts is greatest for
25 intermediate lead times between 4 and 7 days.

26 **Introduction**

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

37 Many studies have shown that operational hydrological models can benefit from the assimilation of
38 in-situ or satellite remote sensing observations. Different techniques and approaches have been
39 presented (Liu et al., 2012). They differ both in terms of the type of data that are assimilated to the
40 models, the assimilation algorithms used and in terms of the assimilation strategy, i.e. which model
41 components, states and/or parameters are updated. Some hydrological data assimilation studies
42 update the internal states of rainfall-runoff models (e.g. Clark et al., 2008; Pauwels and De Lannoy,
43 2009) while other approaches focus on updating the hydrodynamic parts of the model (Biancamaria
44 et al., 2011; Neal et al., 2009) or combinations of rainfall-runoff and routing state variables (e.g.

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

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

69 the environment. Therefore, in many river basins there is a need for hydrological models to provide
70 operational estimates of river discharge based on remotely sensed observations and limited
71 available in-situ measurements.

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

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

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

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

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

102 **Materials and Methods**

103 **Study Area**

104 The Kavango River originates in the highlands of central Angola and flows south to the border
105 between Angola and Namibia. The Cuito River joins the Kavango River just before the river enters
106 into Namibia's Caprivi Strip. It terminates in the Okavango Delta, a large wetland system in
107 Northern Botswana (Milzow et al., 2009). An overview of the basin is provided in Figure 1. The
108 basin is located on the Southern fringes of the inter-tropical convergence zone. A strong south-to-
109 north precipitation gradient is observed. The climate is highly seasonal and large inter-annual
110 variations are typical, which are controlled by a number of climate time scales (McCarthy et al.,
111 2000; Wolski et al., 2014). The Kavango River is an important resource for all riparian countries
112 and forms the basis of many people's livelihoods (Kgathi et al., 2006). While water scarcity and
113 water allocation between economic sectors and the environment have been in focus for some time,

114 flood risk has recently become a major concern because the northern part of Namibia has
115 experienced increased magnitude and frequency of flooding events since 2008 (Wolski et al., 2014).
116 Water managers need accurate and reliable forecasting tools to deal with both floods and droughts.
117 Three hydrological modelling efforts have been reported in the literature for the Kavango River
118 basin. Folwell and Farquharson, 2006 used the Global Water Availability Assessment (GWAVA)
119 model to assess climate change impacts in the basin. Hughes et al., 2011, 2006 calibrated a Pitman
120 model for the basin and were able to reproduce in-situ observations satisfactorily. Milzow et al.,
121 2011 developed a SWAT (Soil and Water Assessment Tool) model of the Kavango basin and
122 calibrated the model with water levels from radar altimetry, soil moisture from Envisat-ASAR and
123 total water storage change from GRACE.
124 Long-term in-situ observations of river discharge are available from two hydrometric stations in the
125 basin, Rundu and Mohebo (Figure 1). Table 1 summarizes the main characteristics of the
126 Kavango river basins and the two sub-basins contributing to the stations Rundu and Mohebo.

127 **Hydrologic and hydrodynamic modelling**

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

132 We use the well-known SWAT hydrological model, version 2009 (Gassman et al., 2005; Neitsch et
133 al., 2011) for rainfall-runoff modelling. SWAT is a semi-distributed, physically based hydrological
134 model which operates at a daily time step. The river basin is divided into a number of sub-basins.
135 Each sub-basin is in turn divided into hydrological response units (HRU), which are defined as
136 portions of the sub-basin with similar terrain slope, land use and soil type. The Kavango SWAT

137 model consists of 12 subbasins with outlets located at the confluences of major tributaries as well as
138 at in-situ discharge station locations (Figure 1).

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

$$149 \quad \mathbf{q}^{t+1} = \mathbf{A}\mathbf{q}^t + \mathbf{B}\mathbf{r}^t + \mathbf{C}\mathbf{r}^{t+1} \quad (1)$$

150 In this equation, \mathbf{q} is the vector of simulated discharges and \mathbf{r} is the vector of runoff forcings, \mathbf{A} , \mathbf{B}
151 and \mathbf{C} are linear operators which depend on the configuration of the river channels and network
152 connectivity and the superscripts indicate time steps. For details on the implementation of the
153 Muskingum routing scheme the reader is referred to Chow et al., 1988 and Michailovsky et al.,
154 2013.

155 **Input data**

156 SWAT requires the following input datasets: elevation, land cover, soil type and climate forcings.
157 The elevation dataset is used for automatic watershed and river network delineation as well as for
158 the determination of terrain slope. We use the ACE2 (Altimeter Corrected Elevation, version 2,
159 Berry et al., 2010) global elevation dataset at a resolution of 30 arc-seconds. The parameterization

160 of vegetation processes in the SWAT model is based on the land cover input dataset. We use the
161 USGS Global Land Cover Characterization (GLCC) dataset, version 2.0 with a spatial resolution of
162 1 km (USGS, 2008). The soil dataset forms the basis for parameterizing soil hydraulic processes in
163 SWAT. We use the FAO/UNESCO digital soil map of the world and derived soil properties,
164 revision 1, with a spatial resolution of 5 arc-minutes (FAO-Unesco, 1974). Look-up tables
165 translating GLCC land cover classes and FAO/UNESCO soil types into SWAT parameters have
166 been developed by the WaterBase project (George and Leon, 2007).

167 The model is forced with daily precipitation and daily minimum and maximum temperature from
168 the National Oceanic and Atmospheric Administration's Global Forecast System (NOAA-GFS)
169 which provides up to seven days of forecast at a six hourly temporal resolution and 0.5 degree
170 spatial resolution (NOAA, 2014). Real-time and recent historical forecasts can be downloaded from
171 the NOMADS server (http://nomads.ncdc.noaa.gov/data.php#hires_weather_datasets, last accessed:
172 [14.01.2015](#)). Historical forecasts older than a few months have to be ordered for FTP download.
173 NOAA-GFS data was aggregated to daily precipitation prior to its use in the hydrological model.
174 For historical simulation periods and model calibration, forcing time series consisting of the 1-day
175 ahead forecasts are used. In operational mode, long-term forecasts are successively replaced with
176 short-term forecasts as time proceeds. In order to assess the performance of the NOAA-GFS
177 precipitation forecast for the Kavango region, the 1-day ahead forecasts were compared to FEWS-
178 RFE rainfall estimates (Herman et al., 1997). FEWS-RFE was previously found to be one of the
179 most accurate remote sensing precipitation products for Africa (Milzow et al., 2011; Stisen and
180 Sandholt, 2010).

181 **Calibration and validation of the hydrologic-hydrodynamic model**

182 Calibration and validation of the hydrologic-hydrodynamic model were performed against observed
183 in situ river discharge using a split-sample approach. The years 2005-2011 were used for

184 calibration, while the years 2012-2014 served as validation period. Mean observed flows in the
185 validation period are higher than in the calibration period (Table 2). After a series of dry years in
186 the beginning of the century, the region has experienced much higher amounts of precipitation and
187 river flow since 2008 (Wolski et al., 2014). In order to ensure a balanced representation of both wet
188 and dry years in the calibration period, we had to use a major portion of the entire data record for
189 calibration and could only reserve three years for validation. Particularly for the station Mohembo,
190 only very few observations are available in the validation period (Table 2). The objective function
191 which was minimized in the calibration was formulated as

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

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

201 Calibration was performed using the model-independent parameter estimation programme PEST
202 (Doherty et al., 2014). Because of the strongly non-linear response of the SWAT rainfall-runoff
203 model, global derivative-free search strategies are the preferred option for calibration of SWAT
204 models (Arnold et al., 2012). We use the shuffled complex evolution (SCE) algorithm (Duan et al.,
205 1992) which performs a global search over the entire allowed parameter space. The SCE algorithm
206 is included in the PEST package (SCEUA_P).

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

224 Table 3 provides an overview of the calibration parameters and their allowed ranges. For the
225 groundwater parameters, spatial variation was allowed between the Rundu and Mohembo regions,
226 the upstream and downstream catchments within each region and the high slope and low slope
227 portions of the land surface. This resulted in a total number of 19 calibration parameters for the
228 Rundu region and 20 calibration parameters for the Mohembo region. We chose 8 complexes in the
229 SCE calibration run and the number of complexes remained the same throughout the run. Both the
230 number of parameter sets in each complex and the number of evolution steps before complex

231 shuffling were set to 39 and 41 for the Rundu and Mohembo regions respectively. The convergence
232 criterion was set to a relative improvement of the best objective function of 1% over 10 shuffling
233 loops. A total of 50000 model runs were allowed, however the calibration converged after 14711
234 and 18373 model runs for the Rundu and Mohembo regions respectively. After completion of the
235 SCE run, the evolution of the parameter values over the course of the shuffling loops was evaluated.
236 All parameter values converged to a stable solution away from the a priori parameter bounds.

237 **Assimilation strategy**

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

249 Observed in-situ discharge at the station Rundu was assimilated to the model in the operational
250 runs. Because the Muskingum routing operator is linear and the measurement operator is linear too,
251 we could use the standard Kalman filter for state updating, since it is the optimal sequential
252 assimilation method for linear dynamics (Kalman, 1960). The Kalman filter simultaneously updates
253 discharge at all basin outlets. If instead of river discharge, water level measurements from space-
254 borne or ground-based instruments are assimilated, the measurement operator becomes non-linear

255 and the extended Kalman filter can be used (Michailovsky et al., 2013). The reader is referred to the
256 literature (e.g. Jazwinski, 1970) for a detailed discussion of the Kalman filter equations and to
257 Michailovsky et al., 2013 for a detailed description of the assimilation approach.

258 **Description of the model error**

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

$$269 \quad w_t = \frac{(Q_{sim,t} - Q_{obs,t})}{Q_{obs,t}} \quad (3)$$

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

$$273 \quad w_t = \delta w_{t-1} + \varepsilon_t \quad (4)$$

274 where δ is the AR1 parameter and ε is a sequence of white Gaussian noise with a spatial covariance
275 Q' . Due to the correlated meteorological inputs the runoff forcing error was assumed to be spatially
276 correlated between the various subcatchments of the model. In the baseline experiment, we assume

277 that the spatial correlation of the runoff forcing error is equivalent to the spatial correlation of the
278 runoff forcing itself. The correlation matrix of the runoff inputs was computed and Q' was set to:

$$279 \quad \mathbf{Q}' = \mathbf{C} \sigma(\epsilon)^2 \quad (5)$$

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

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

293 In order to evaluate the impact of model error and observation error specifications on the
294 performance of the probabilistic discharge forecasts, four additional forecasting experiments were
295 conducted. Table 4 presents an overview of the experiments. In the baseline experiment, the
296 autocorrelation of the relative runoff error was set equal to the autocorrelation of the relative model
297 error at Rundu (0.9942), as described above. The magnitude of the relative runoff error was set to
298 4.38%, which is the same as the relative model error at Rundu. The spatial correlation of relative
299 runoff error was set equal to the spatial correlation of runoff and the relative observation error was

300 set to 10%. In experiment 1, the autocorrelation of the runoff error was set equal to the
 301 autocorrelation of the spatially aggregated runoff (0.9934) while the other specifications are the
 302 same as in the baseline run. In experiment 2, the spatial correlation of the runoff error was set to
 303 zero and all other specifications are as in the baseline run. In experiment 3, the runoff error
 304 specifications are the same as in the baseline and the relative observation error was set to 20%.
 305 Finally, in experiment 4, the white noise component of the relative runoff error was increased from
 306 4.38% to 6% and all other specifications are as in the baseline run.

307 **Operational forecasting and performance evaluation**

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

$$317 \quad PI = \frac{\frac{1}{n} \sum_{i=1}^n (Q_i - Q_{obs,i})^2 - \frac{1}{n} \sum_{i=1}^n (Q_i - Q_{last})^2}{-\frac{1}{n} \sum_{i=1}^n (Q_i - Q_{last})^2} \quad (6)$$

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

322 as the benchmark (i.e. a forecast that performs as good as the latest available observation scores a PI
 323 of 0).

324 Reliability and sharpness of the probabilistic forecasts were assessed with the coverage of the 95%
 325 confidence interval (i.e. percentage of observations that fall within the predicted nominal 95%
 326 confidence interval), the sharpness of the 95% confidence interval (width of predicted 95%
 327 confidence interval), the Interval Skill Score (ISS) of the 95% confidence interval as well as the
 328 continuous ranked probability score (CRPS). The ISS is defined according to Gneiting and Raftery,
 329 2007 as:

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

$$330 \quad iss_{\alpha}(l, u, Q_{obs}) = \begin{cases} (u - l) & \text{if } l < Q_{obs} < u \\ (u - l) + 2/\alpha (l - x) & \text{if } Q_{obs} < l \\ (u - l) + 2/\alpha (x - u) & \text{if } Q_{obs} < u \end{cases} \quad (7)$$

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

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

$$338 \quad CRPS = \frac{1}{n} \sum_{i=1}^n crps(Q_{obs,i}, Q_i, \sigma_i) \quad (8)$$

$$crps(Q_{obs}, Q, \sigma) = \sigma \left[\frac{Q_{obs} - Q}{\sigma} \left(2\Phi \left(\frac{Q_{obs} - Q}{\sigma} \right) - 1 \right) + 2\phi \left(\frac{Q_{obs} - Q}{\sigma} \right) - \frac{1}{\sqrt{\pi}} \right]$$

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

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

348 **Results**

349 **Comparison of precipitation products**

350 Comparison of the FEWS-RFE and NOAA-GFS precipitation products showed large deviations
351 between the two products. Figure 2 shows a double mass plot for the average precipitation over the
352 entire Kavango River catchment for the period 2005-2012. Obviously, there is a significant bias and
353 the timing of precipitation events is inconsistent too, as evidenced by the wiggles in the double
354 mass curve. The FEWS-RFE product is based on both satellite observations and in-situ gauging
355 stations, while NOAA-GFS is derived from a global weather model. Moreover, FEWS-RFE has
356 been shown to perform well in previous studies on the African continent (Milzow et al., 2011;
357 Stisen and Sandholt, 2010). We therefore assume that the FEWS-RFE product is closer to the
358 unknown true precipitation than NOAA-GFS and bias correct the NOAA-GFS data to match the
359 long-term average precipitation for both products. A spatially and temporally constant precipitation
360 correction factor of 0.67 was therefore used throughout the study. Figure 2 also presents a

361 quantitative comparison of the NOAA-GFS precipitation forecasts for various forecasting horizons.
362 As a general trend, the longer the forecasting horizon, the lower the predicted precipitation
363 compared to the 1-day ahead forecasts. These effects are particularly pronounced for the rainy
364 seasons 2008/2009 and 2011/2012. However, for the most recent years, the double mass plots show
365 slopes close to unity. We therefore did not implement variable bias correction for the different
366 forecasting horizons. Because the NOAA-GFS system is continuously updated and modified
367 (process parameterization, spatial resolution etc.), performance of precipitation forecasts should be
368 regularly checked during operational application of the hydrologic forecasting system. Changes in
369 the quantitative precipitation forecasts may require adjustments in the bias correction and/or
370 recalibration of the hydrological model.

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

381 **Performance of the calibrated model**

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

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

404 **Discharge forecasting and data assimilation**

405

406 Table 5 reports the performance statistics for the probabilistic model runs. We report results for the
407 open-loop run without assimilation, the assimilation run (“now-casting”) as well as the 1-7 day
408 ahead forecasts. The various forecasting horizons use different precipitation forcings (forecasts

409 available at the simulated issue date) and in-situ data are assimilated up to simulated issue date. We
410 only assimilate data from the station Rundu, because (i) no real-time observations are available for
411 Mohembo and (ii) this enables us to assess the effect of upstream assimilation on a downstream
412 station. The indicators are reported for both in-situ stations and for the calibration and the validation
413 period. We are well aware that the observations in the calibration period have been used already for
414 model calibration and are now used again for assimilation. Still, we feel that it is useful to present
415 the statistics for information. Figure 6 shows the open-loop and assimilation run for the station
416 Rundu during calibration and validation periods. We first assess the performance of the
417 probabilistic open-loop run. Generally, the chosen error model seems to be appropriate. The
418 forecasts produced by the open-loop run are reliable; the coverage of the nominal 95% confidence
419 interval does not fall below 84% at any of the stations during any of the periods. However, the
420 open-loop forecasts are not very sharp, as evidenced by the wide confidence intervals in Figure 6.
421 This results in a relatively high ISS score.

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

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

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

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

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

453 **Discussion**

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

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

476 Comparison of the various forecasting experiments shows that assumptions about the model and
477 observation errors have a large impact on the performance of the forecasting system. The magnitude
478 of the relative runoff error is particularly sensitive, as evidenced by the improved performance of
479 experiment 4 compared to the baseline. It is reasonable to assume a higher relative error for the
480 runoff than the relative error computed from the model residuals at Rundu, because the routing
481 model has a smoothing effect on the runoff response. Experiment 3 and the baseline show a
482 comparable performance in terms of CRPS. Basically the higher assumed observation error in
483 experiment 3 results in predictions that are less sharp but more reliable. Comparison of experiment
484 1 and baseline results shows that even small differences in the assumed autocorrelation of the runoff
485 error result in significant differences in the forecast performance. Higher error autocorrelation leads
486 to increased sharpness, but lower reliability. CRPS indicates that experiment 1 forecasts marginally
487 outperform the baseline forecasts. Experiment 2 results are very close to the baseline, because the
488 spatial correlation of runoff between the different subcatchments is low, due to the variable
489 hydrologic characteristics of the subcatchments. Predictive Q-Q plots for experiment 4 (Figure 8)
490 indicate significant deviations of the empirical distribution of normalized forecast errors from the
491 normal distribution.

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

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

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

520 **Conclusions**

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

533

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

538

539 **References**

- 540 Arnold, J.G., Moriasi, D.N., Gassman, P.W., Abbaspour, K.C., White, M.J., Srinivasan, R., Santhi,
541 C., Harmel, R.D., van Griensven, A., Van Liew, M.W., Kannan, N., Jha, M.K., 2012. SWAT:
542 MODEL USE, CALIBRATION, AND VALIDATION. *Trans. ASABE* 55, 1491–1508.
- 543 Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H. a., Hamilton, S.H., Jakeman, A.J.,
544 Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S. a., Robson, B., Seppelt,
545 R., Voinov, A. a., Fath, B.D., Andreassian, V., 2013. Characterising performance of
546 environmental models. *Environ. Model. Softw.* 40, 1–20. doi:10.1016/j.envsoft.2012.09.011
- 547 Berry, P.A.M., Smith, R.G., Benveniste, J., 2010. ACE2: The New Global Digital Elevation Model,
548 in: Mertikas, SP (Ed.), GRAVITY, GEOID AND EARTH OBSERVATION, International
549 Association of Geodesy Symposia. pp. 231–237. doi:10.1007/978-3-642-10634-7_30
- 550 Biancamaria, S., Durand, M., Andreadis, K.M., Bates, P.D., Boone, a., Mognard, N.M., Rodríguez,
551 E., Alsdorf, D.E., Lettenmaier, D.P., Clark, E. a., 2011. Assimilation of virtual wide swath
552 altimetry to improve Arctic river modeling. *Remote Sens. Environ.* 115, 373–381.
553 doi:10.1016/j.rse.2010.09.008
- 554 Boucher, M. -a., Anctil, F., Perreault, L., Tremblay, D., 2011. A comparison between ensemble and
555 deterministic hydrological forecasts in an operational context. *Adv. Geosci.* 29, 85–94.
556 doi:10.5194/adgeo-29-85-2011
- 557 Chow, V. Te, Maidment, D.R., Mays, L.W., 1988. *Applied Hydrology, Water Resources and*
558 *Environmental Engineering.* McGraw-Hill, New York.
- 559 Clark, M.P., Rupp, D.E., Woods, R. a., Zheng, X., Ibbitt, R.P., Slater, A.G., Schmidt, J., Uddstrom,
560 M.J., 2008. Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow
561 observations to update states in a distributed hydrological model. *Adv. Water Resour.* 31,
562 1309–1324. doi:10.1016/j.advwatres.2008.06.005
- 563 Cloke, H.L., Pappenberger, F., 2009. Ensemble flood forecasting: A review. *J. Hydrol.* 375, 613–
564 626. doi:10.1016/j.jhydrol.2009.06.005
- 565 Di Baldassarre, G., 2009. Uncertainty in river discharge observations: a quantitative analysis.
566 *Hydrol. EARTH Syst. Sci.* 13, 913 – 921.
- 567 Doherty, J., Muffels, C., Rumbaugh, J., Tonkin, M., 2014. PEST. Model independent parameter
568 estimation and uncertainty analysis [WWW Document]. URL
569 <http://www.pesthomepage.org/Home.php> (accessed 7.16.14).
- 570 Duan, Q.Y., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization for
571 conceptual rainfall-runoff models. *Water Resour. Res.* 28, 1015–1031.
572 doi:10.1029/91WR02985

- 573 FAO-Unesco, 1974. Soil map of the world 1:5000000. Paris.
- 574 Fekete, B.M., Voeroesmarty, C.J., 2007. The current status of global river discharge monitoring and
575 potential new technologies complementing traditional discharge measurements, in:
576 Proceedings of the PUB Kick-off Meeting. IAHS publication 309, Brasilia.
- 577 Folwell, S., Farquharson, F., 2006. The impacts of climate change on water resources in the
578 Okavango basin, in: Demuth, S and Gustard, A and Planos, E and Scatena, F and Servat, E
579 (Ed.), Climate Variability and Change - Hydrological Impacts, IAHS PUBLICATION. pp.
580 382–388.
- 581 Gassman, P.W., Reyes, M.R., Green, C.H., Arnold, J.G., 2005. SWAT Peer-Reviewed Literature :
582 A Review. *Hydrol. Process.* 13, 1–17.
- 583 Georgakakos, K.P., 1986. A generalized stochastic hydrometeorological model for flood and flash-
584 flood forecasting. 2. case studies. *Water Resour. Res.* 22, 2096–2106.
585 doi:10.1029/WR022i013p02096
- 586 George, C., Leon, L.F., 2007. WaterBase : SWAT in an open source GIS. *Open Hydrol. J.* 1, 19–24.
- 587 Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. *J. Am.*
588 *Stat. Assoc.* 102, 359–378. doi:10.1198/016214506000001437
- 589 Gneiting, T., Westveld III, A.H., Raftery, A.E., Goldman, T., 2004. Calibrated Probabilistic
590 Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation *.
591 Seattle, Washington.
- 592 Guzinski, R., Kass, S., Huber, S., Bauer-Gottwein, P., Jensen, I.H., Naeimi, V., Doubkova, M.,
593 Walli, A., Tottrup, C., 2014. A Water Observation and Information System for Integrated
594 Water Resource Management in Africa. *Remote Sens.* under revi. doi:10.3390/rs60x000x
- 595 Herman, A., Kumar, V.B., Arkin, P.A., Kousky, J. V, 1997. Objectively determined 10-day African
596 rainfall estimates created for famine early warning systems. *Int. J. Remote Sens.* 18, 2147–
597 2159. doi:10.1080/014311697217800
- 598 Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Wolff, D.B., Adler, R.F., Gu, G., Hong, Y., Bowman,
599 K.P., Stocker, E.F., 2007. The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-
600 Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales. *J. Hydrometeorol.*
601 8, 38–55. doi:10.1175/JHM560.1
- 602 Hughes, D. a., Andersson, L., Wilk, J., Savenije, H.H.G., 2006. Regional calibration of the Pitman
603 model for the Okavango River. *J. Hydrol.* 331, 30–42. doi:10.1016/j.jhydrol.2006.04.047
- 604 Hughes, D. a., Kingston, D.G., Todd, M.C., 2011. Uncertainty in water resources availability in the
605 Okavango River basin as a result of climate change. *Hydrol. Earth Syst. Sci.* 15, 931–941.
606 doi:10.5194/hess-15-931-2011
- 607 Jazwinski, A.H., 1970. Stochastic Processes and Filtering Theory. Academic Press, New York.

- 608 Kalman, R.E., 1960. A New Approach to Linear Filtering and Prediction Problems. *J. basic Eng.*
609 82, 35–45.
- 610 Kgathi, D.L., Kniveton, D., Ringrose, S., Turton, a. R., Vanderpost, C.H.M., Lundqvist, J., Seely,
611 M., 2006. The Okavango; a river supporting its people, environment and economic
612 development. *J. Hydrol.* 331, 3–17. doi:10.1016/j.jhydrol.2006.04.048
- 613 Liu, Y., Weerts, a. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J.,
614 Schwanenberg, D., Smith, P., van Dijk, a. I.J.M., van Velzen, N., He, M., Lee, H., Noh, S.J.,
615 Rakovec, O., Restrepo, P., 2012. Advancing data assimilation in operational hydrologic
616 forecasting: progresses, challenges, and emerging opportunities. *Hydrol. Earth Syst. Sci.* 16,
617 3863–3887. doi:10.5194/hess-16-3863-2012
- 618 Madsen, H., Skotner, C., 2005. Adaptive state updating in real-time river flow forecasting—a
619 combined filtering and error forecasting procedure. *J. Hydrol.* 308, 302–312.
620 doi:10.1016/j.jhydrol.2004.10.030
- 621 Maier, H.R., Jain, A., Dandy, G.C., Sudheer, K.P., 2010. Methods used for the development of
622 neural networks for the prediction of water resource variables in river systems: Current status
623 and future directions. *Environ. Model. Softw.* 25, 891–909. doi:10.1016/j.envsoft.2010.02.003
- 624 McCarthy, T.S., Cooper, G.R.J., Tyson, P.D., Ellery, W.N., 2000. Seasonal flooding in the
625 Okavango Delta, Botswana - recent history and future prospects. *S. Afr. J. Sci.* 96, 25–33.
- 626 Michailovsky, C.I., Milzow, C., Bauer-Gottwein, P., 2013. Assimilation of radar altimetry to a
627 routing model of the Brahmaputra River. *Water Resour. Res.* 49, 4807–4816.
628 doi:10.1002/wrcr.20345
- 629 Milzow, C., Kgotlhang, L., Bauer-Gottwein, P., Meier, P., Kinzelbach, W., 2009. Regional review:
630 the hydrology of the Okavango Delta, Botswana—processes, data and modelling,
631 *Hydrogeology Journal.* doi:10.1007/s10040-009-0436-0
- 632 Milzow, C., Krogh, P.E., Bauer-Gottwein, P., 2011. Combining satellite radar altimetry, SAR
633 surface soil moisture and GRACE total storage changes for hydrological model calibration in a
634 large poorly gauged catchment. *Hydrol. Earth Syst. Sci.* 15, 1729–1743. doi:10.5194/hess-15-
635 1729-2011
- 636 Moradkhani, H., Hsu, K.-L., Gupta, H., Sorooshian, S., 2005. Uncertainty assessment of hydrologic
637 model states and parameters: Sequential data assimilation using the particle filter. *Water*
638 *Resour. Res.* 41, n/a–n/a. doi:10.1029/2004WR003604
- 639 Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I - a
640 discussion of principles. *J. Hydrol.* 10, 282–290.
- 641 Neal, J., Schumann, G., Bates, P., Buytaert, W., Matgen, P., Pappenberger, F., 2009. A data
642 assimilation approach to discharge estimation from space 3649, 3641–3649. doi:10.1002/hyp

- 643 Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., 2011. Soil & Water Assessment Tool
644 Theoretical Documentation Version 2009.
- 645 NOAA, 2014. GFS Global Forecast System [WWW Document]. URL
646 <http://www.emc.ncep.noaa.gov/index.php?branch=GFS> (accessed 7.16.14).
- 647 Pagano, T.C., Wood, A.W., Ramos, M.-H., Cloke, H.L., Pappenberger, F., Clark, M.P., Cranston,
648 M., Kavetski, D., Mathevet, T., Sorooshian, S., Verkade, J.S., 2014. Challenges of Operational
649 River Forecasting. *J. Hydrometeorol.* 15, 1692–1707. doi:10.1175/JHM-D-13-0188.1
- 650 Pauwels, V.R.N., De Lannoy, G.J.M., 2009. Ensemble-based assimilation of discharge into rainfall-
651 runoff models: A comparison of approaches to mapping observational information to state
652 space. *Water Resour. Res.* 45, n/a–n/a. doi:10.1029/2008WR007590
- 653 Peterson, T.C., Vose, R.S., 1997. An overview of the global historical climatology network
654 temperature database. *Bull. Am. Meteorol. Soc.* 78, 2837–2849. doi:10.1175/1520-
655 0477(1997)078<2837:AOOTGH>2.0.CO;2
- 656 Rakovec, O., Weerts, a. H., Hazenberg, P., Torfs, P.J.J.F., Uijlenhoet, R., 2012. State updating of a
657 distributed hydrological model with Ensemble Kalman Filtering: effects of updating frequency
658 and observation network density on forecast accuracy. *Hydrol. Earth Syst. Sci.* 16, 3435–3449.
659 doi:10.5194/hess-16-3435-2012
- 660 Schellekens, J., Weerts, a. H., Moore, R.J., Pierce, C.E., Hildon, S., 2011. The use of MOGREPS
661 ensemble rainfall forecasts in operational flood forecasting systems across England and Wales.
662 *Adv. Geosci.* 29, 77–84. doi:10.5194/adgeo-29-77-2011
- 663 Seo, D.-J., Cajina, L., Corby, R., Howieson, T., 2009. Automatic state updating for operational
664 streamflow forecasting via variational data assimilation. *J. Hydrol.* 367, 255–275.
665 doi:10.1016/j.jhydrol.2009.01.019
- 666 Seo, D.-J., Koren, V., Cajina, N., 2003. Real-Time Variational Assimilation of Hydrologic and
667 Hydrometeorological Data into Operational Hydrologic Forecasting. *J. Hydrometeorol.* 4,
668 627–641.
- 669 Stisen, S., Sandholt, I., 2010. Evaluation of remote-sensing-based rainfall products through
670 predictive capability in hydrological runoff modelling. *Hydrol. Process.* 24, 879–891.
671 doi:10.1002/hyp.7529
- 672 Tang, Q., Gao, H., Lu, H., Lettenmaier, D.P., 2009. Remote sensing: hydrology. *Prog. Phys. Geogr.*
673 33, 490–509. doi:10.1177/0309133309346650
- 674 USGS, 2008. Global Land Cover Characteristics Data Base Version 2.0 [WWW Document]. URL
675 <http://edc2.usgs.gov/glcc/glcc.php> (accessed 7.16.14).
- 676 Weerts, A.H., El Serafy, G.Y.H., 2006. Particle filtering and ensemble Kalman filtering for state
677 updating with hydrological conceptual rainfall-runoff models. *Water Resour. Res.* 42, n/a–n/a.
678 doi:10.1029/2005WR004093

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

681

682 **Tables**

683 **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

684

685

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

686

687

688
689
690
691
692
693

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

694

695

696 **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%

697

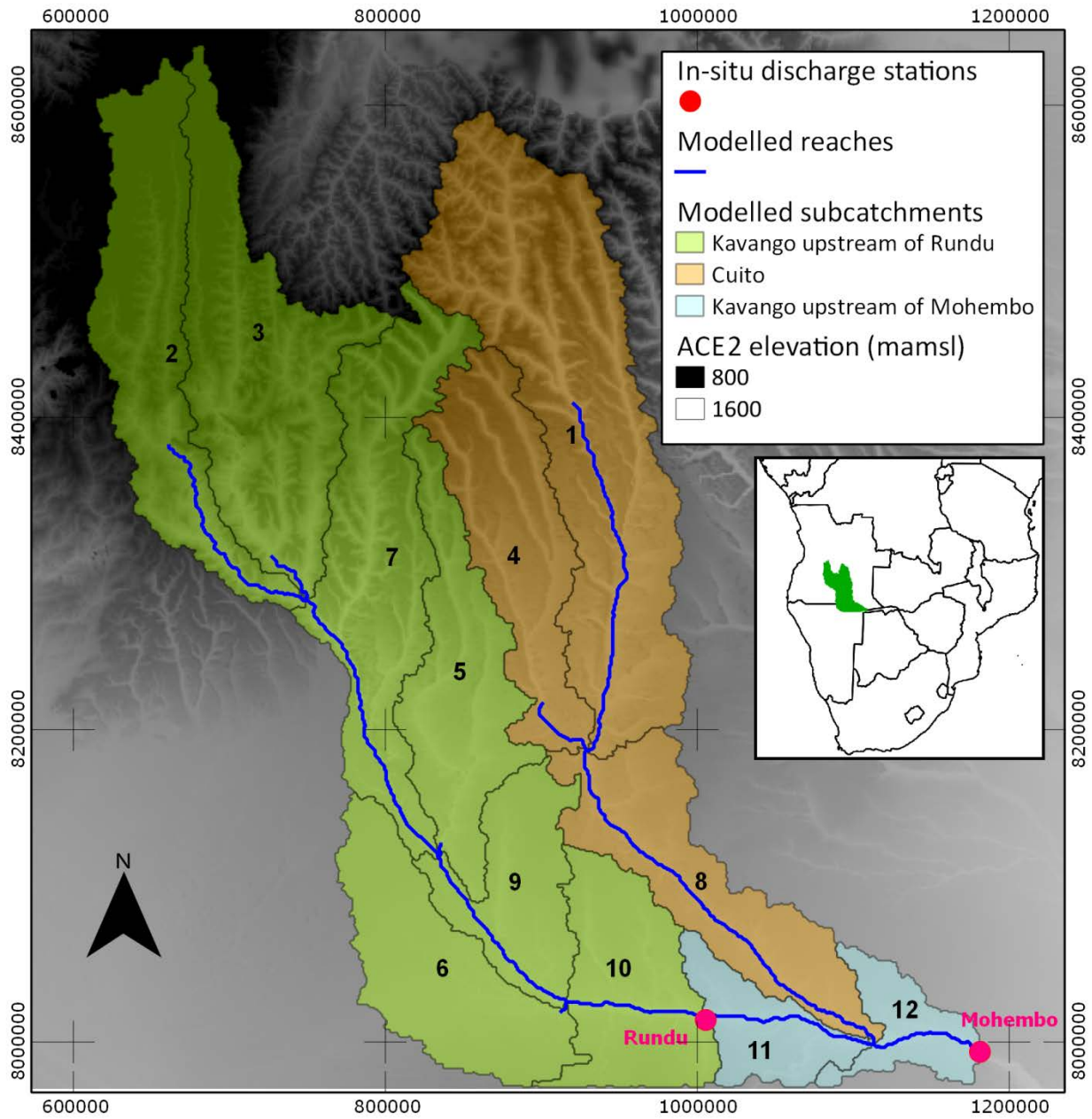
Period	In-situ station	Run	NSE (-)	RMSE (m ³ /s)	ME (m ³ /s)	Coverage (%)	Sharpness (m ³ /s)	Interval Skill Score (m ³ /s)	Mean of predicted observations (m ³ /s)	No. of predicted observations
Calibration Period (2005-2011)	Rundu	Open-Loop	0.73	105.6	-5.4	90.0	423.5	654.9	248.4	2440
		Assimilation	0.99	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	
	Moheumbo	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	
	Moheumbo	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)	Cove- rage (%)	Sharp- ness (m ³ /s)	Interval Skill Score (m ³ /s)	Persistence index (-)	CRPS (m ³ /s)	No. of predicted obser- vations
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

704

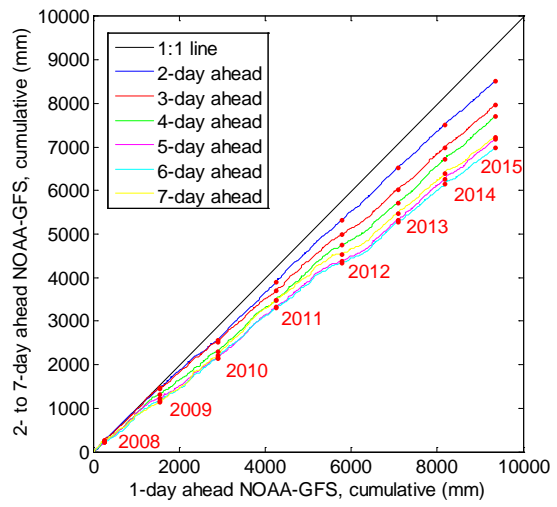
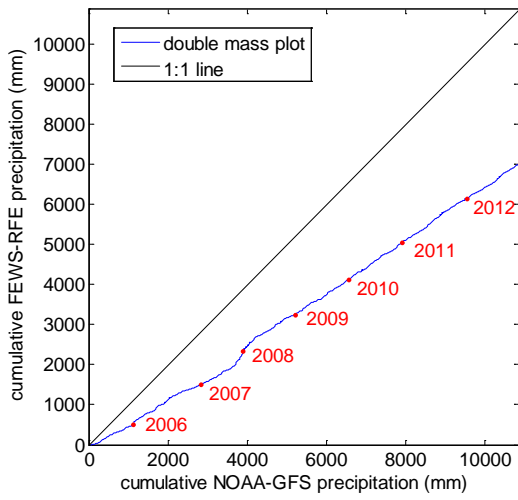
705 **Figures**



706

707 **Figure 1: Basemap for the Kavango River Basin with location of in-situ discharge stations. The coordinate system is UTM**
708 **33S, WGS84 datum. Inset map shows the location of the basin in Southern Africa.**

709



710

711

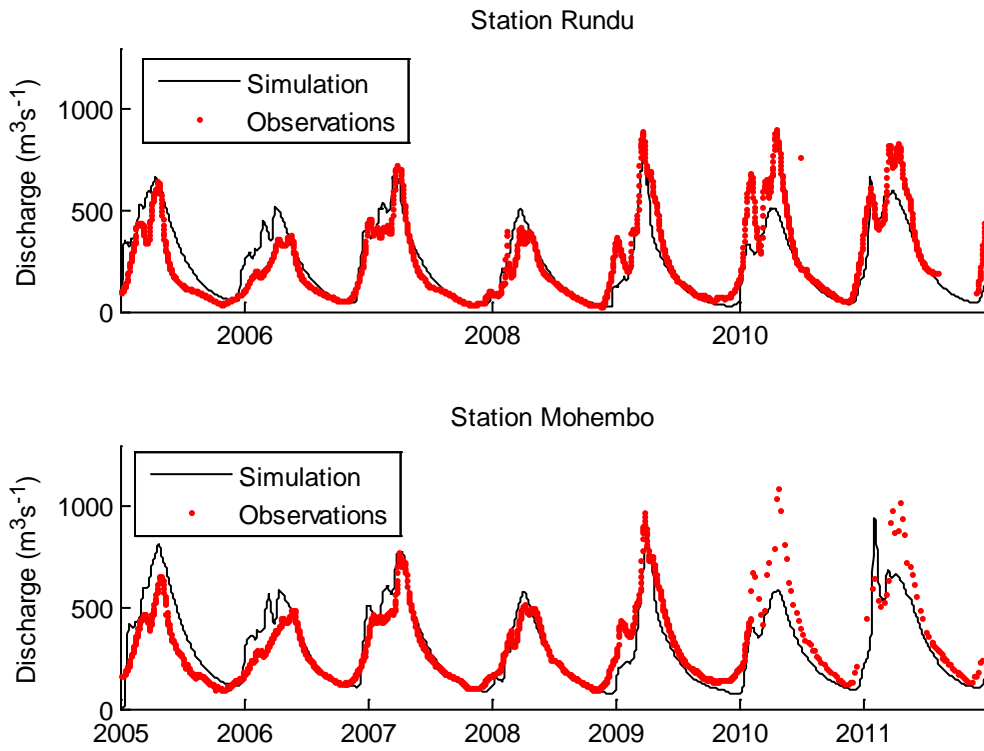
712

713

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.

714

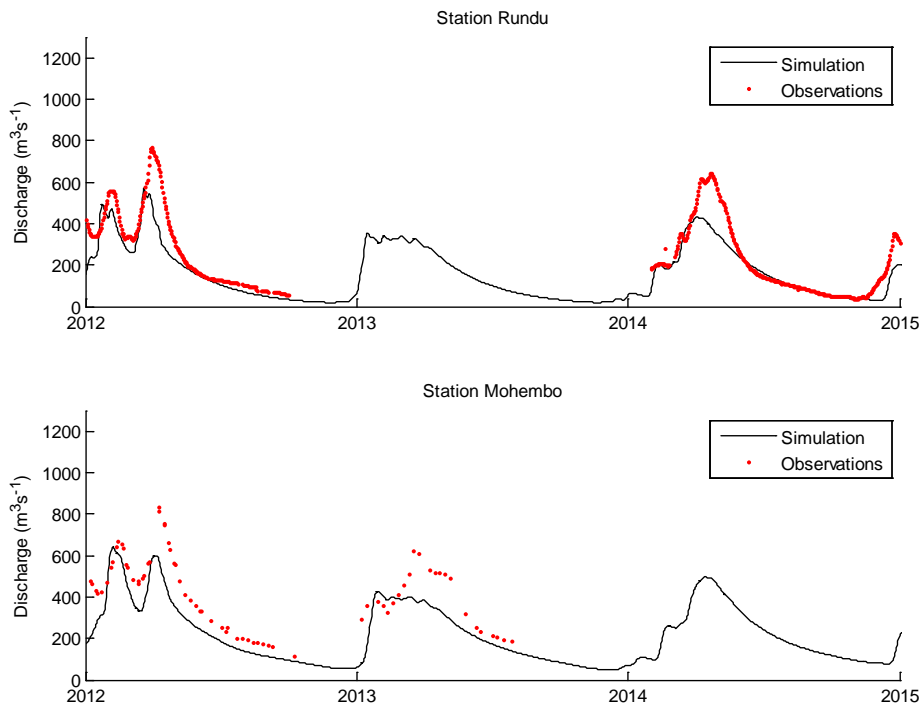
715



716

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

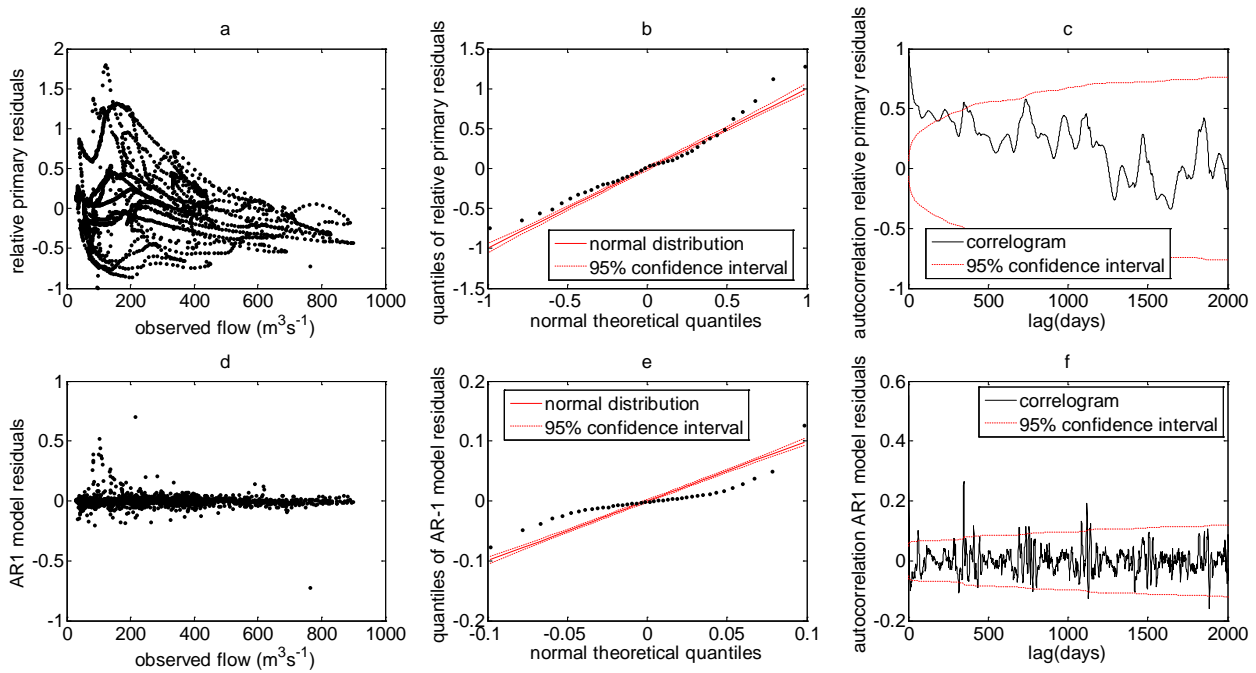
719



720

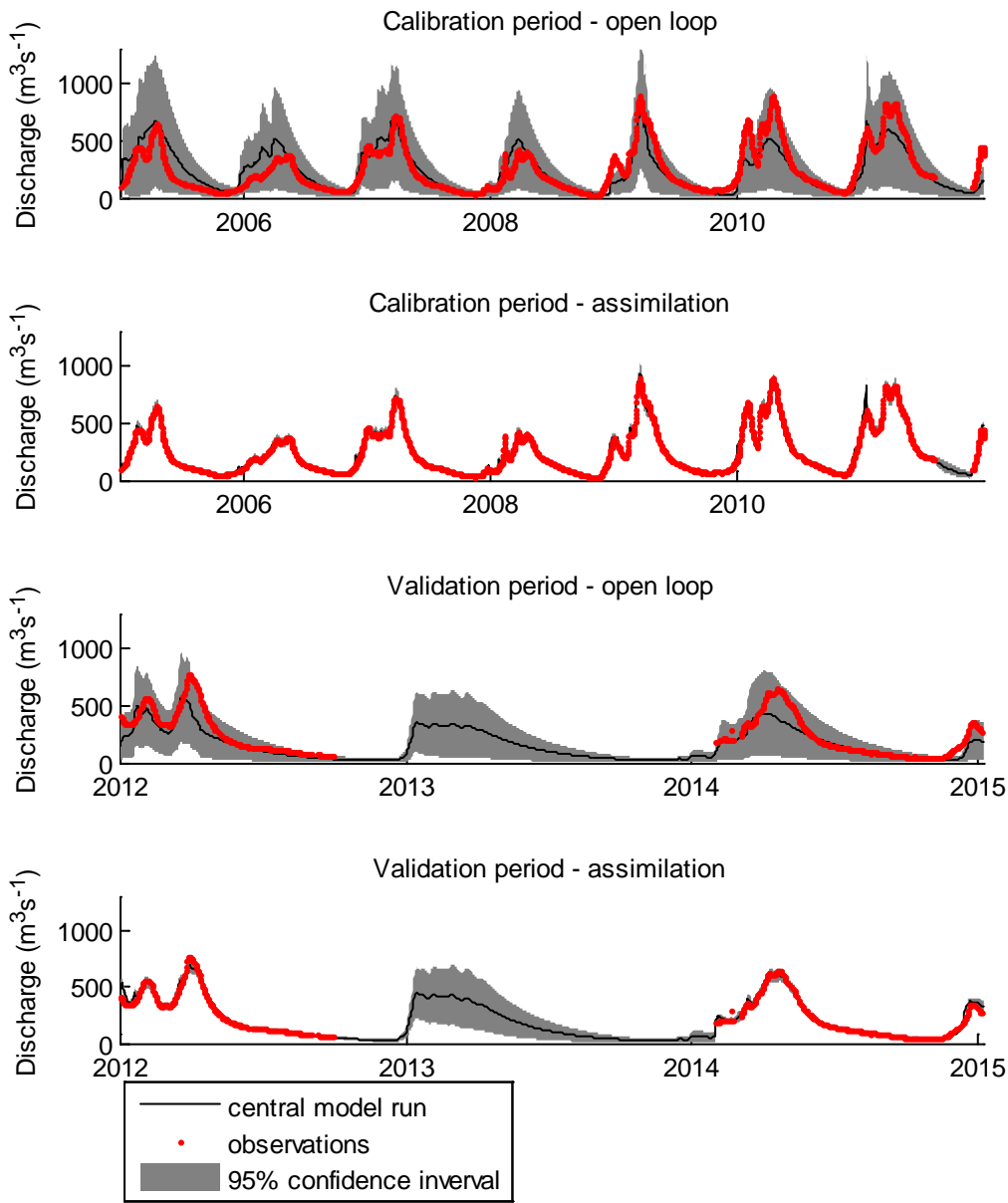
721 **Figure 4: Observed (red dots) and simulated (black lines) hydrographs for the validation period for Rundu (top) and**
 722 **Mohembo (bottom).**

723



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

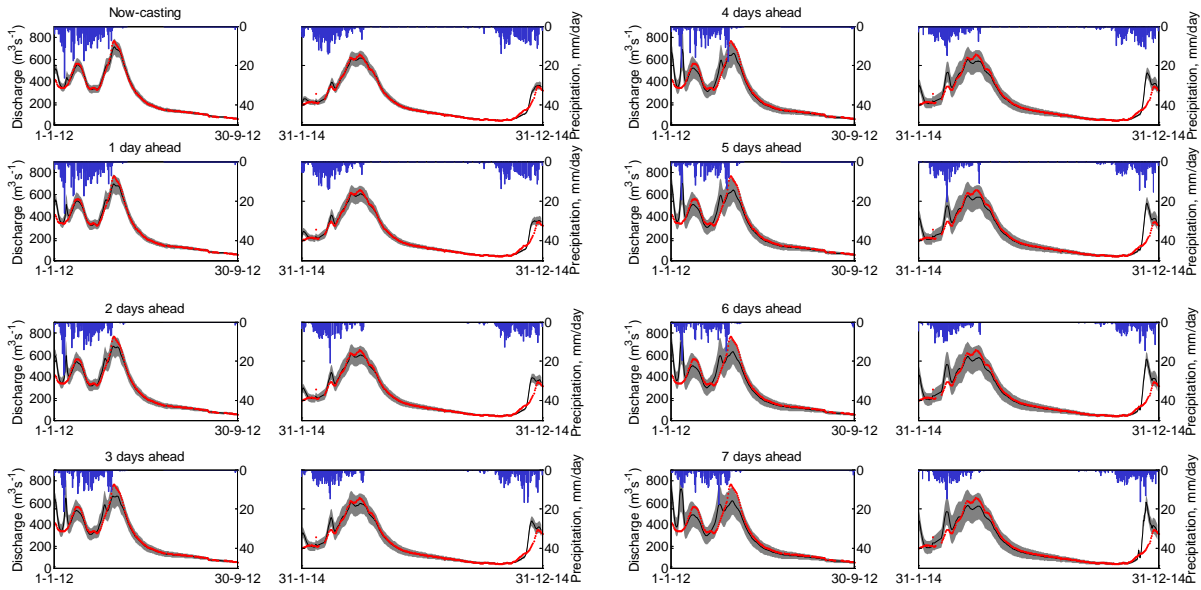
729



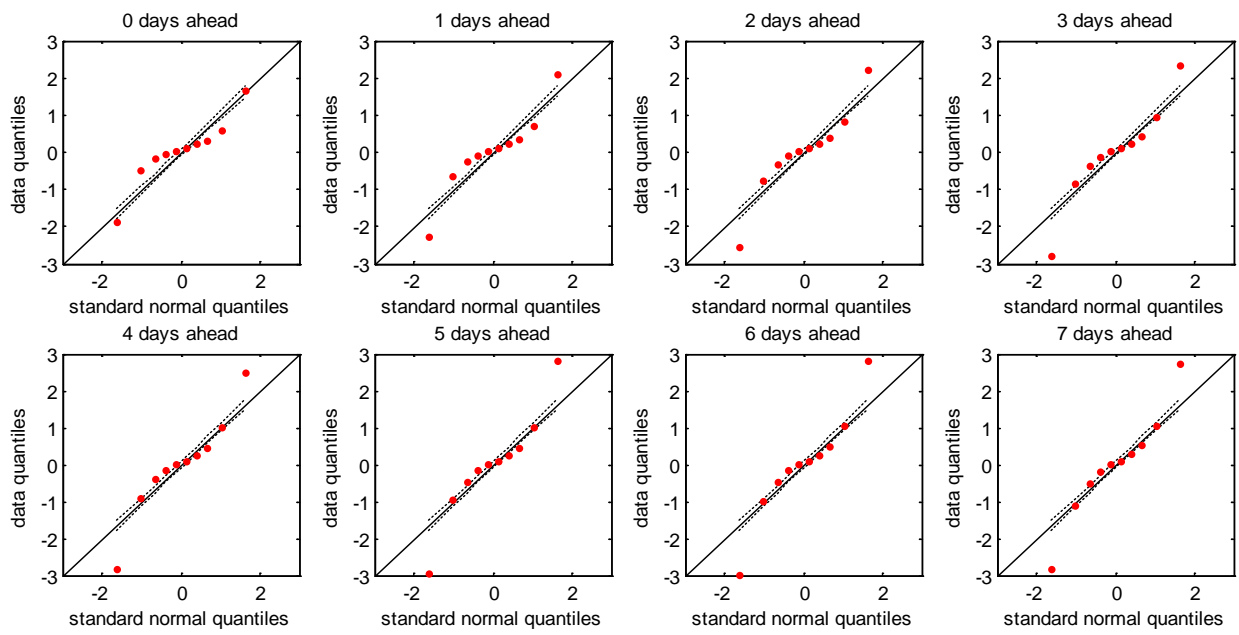
730

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

733



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



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

742