

Hydrological real-time modelling in the Zambezi river basin using satellite-based soil moisture and rainfall data

P. Meier, A. Frömel, and W. Kinzelbach

ETH Zürich, Institute of Environmental Engineering (IfU), Wolfgang-Pauli Strasse 15, 8093 Zürich, Switzerland

Received: 28 October 2010 – Published in Hydrol. Earth Syst. Sci. Discuss.: 10 November 2010

Revised: 11 March 2011 – Accepted: 18 March 2011 – Published: 23 March 2011

Abstract. Reliable real-time forecasts of the discharge can provide valuable information for the management of a river basin system. For the management of ecological releases even discharge forecasts with moderate accuracy can be beneficial. Sequential data assimilation using the Ensemble Kalman Filter provides a tool that is both efficient and robust for a real-time modelling framework. One key parameter in a hydrological system is the soil moisture, which recently can be characterized by satellite based measurements. A forecasting framework for the prediction of discharges is developed and applied to three different sub-basins of the Zambezi River Basin. The model is solely based on remote sensing data providing soil moisture and rainfall estimates. The soil moisture product used is based on the back-scattering intensity of a radar signal measured by a radar scatterometer. These soil moisture data correlate well with the measured discharge of the corresponding watershed if the data are shifted by a time lag which is dependent on the size and the dominant runoff process in the catchment. This time lag is the basis for the applicability of the soil moisture data for hydrological forecasts. The conceptual model developed is based on two storage compartments. The processes modeled include evaporation losses, infiltration and percolation. The application of this model in a real-time modelling framework yields good results in watersheds where soil storage is an important factor. The lead time of the forecast is dependent on the size and the retention capacity of the watershed. For the largest watershed a forecast over 40 days can be provided. However, the quality of the forecast increases significantly with decreasing prediction time. In a watershed with little soil storage and a quick response to rainfall events, the performance is relatively poor and the lead time is as short as 10 days only.

1 Introduction

The water resources in the Zambezi river basin are more and more developed. Feasibility studies for several new hydro-power plants are being carried out and new irrigation schemes are developed all over the river basin. While pressure on the resources is growing, long term forecasts of the discharge with a few weeks lead time can help optimize the operation of smaller reservoirs and water abstraction schemes for irrigation without neglecting the river-dependent ecosystems as an important water user.

Such a forecast of the inflows of the reservoir can be useful, especially if the management of a reservoir is targeted at the release of ecological flows. For these releases the timing is of great significance (Galat and Lipkin, 2000; Acreman and Ferguson, 2010). Operating a dam according to a strict rule curve without any information on future inflows leads to a very late flood pulse since the flood is attenuated until a prescribed water level is reached in the reservoir. With some information on the expected inflow ecological releases can mimic a more natural flow.

Kitanidis and Bras (1980a) stated that effective water management in a river basin system needs a reliable real-time forecast. This involves a continuous correction of the forecasts based on the prediction errors of earlier forecasts. The application of a model is accompanied by several sources of errors, such as model, input and parameter uncertainty. This leads to a deficient knowledge of the system states. Hence it is appropriate to use observations to update the states of the system (Kitanidis and Bras, 1980a,b). Sequential data assimilation techniques such as the Ensemble Kalman Filter (EnKF) provide a state of the art method to perform this update step (McLaughlin, 2002).

As soil moisture is a key parameter in land surface hydrology, the availability of area representative measurements offers a unique opportunity to improve hydrological modelling. The first satellite-derived global dataset on soil moisture was



Correspondence to: P. Meier
(meier@ifu.baug.ethz.ch)

presented by Wagner et al. (1999a). It was shown that runoff predictions were greatly improved when measured soil moisture, both from ground measurements and from remote sensing, were incorporated (Aubert et al., 2003; Crow and Ryu, 2009).

Recent studies have shown the usefulness of radar scatterometer derived soil moisture data for hydrological applications. Despite the generally coarse resolution, these data can be applied successfully for hydrological modelling since small scale spatial variability of the soil water content is averaged within the scatterometer footprint (Ceballos et al., 2005; Scipal et al., 2005).

The application of remotely sensed soil moisture data becomes more and more feasible. Several satellite missions have been launched, or will be launched in the near future, equipped with instruments to retrieve soil moisture information using microwave frequencies. These missions include the MetOp Advanced Scatterometer (ASCAT), the Soil Moisture and Ocean Salinity Mission (SMOS) and NASA's Soil Moisture Active-Passive instrument (SMAP) (Kerr et al., 2001; Naeimi et al., 2009; Piles et al., 2009). Wagner et al. (2007) showed the usefulness of high resolution soil moisture data from the Envisat Advanced Synthetic Aperture Radar (ASAR) instrument in hydrological modelling.

This study presents a prediction framework for river discharge based solely on remotely sensed data of soil moisture and rainfall, a simple conceptual model and data assimilation techniques. The performance of the prediction is evaluated in three different sub-basins of the Zambezi river basin. The availability of the input data in real-time allows the model to be operated in real-time, providing a prediction for discharge each time new input data are retrieved. When observation data are available the model state is updated using sequential data assimilation techniques (EnKF). The update step allows the model to be relatively simple.

The paper is organised as follows: after presenting the study area in Sect. 2, the data used for modelling (soil moisture, rainfall and discharge) are described in Sect. 3. The soil moisture – runoff model and its application in the real-time modelling framework is presented in Sect. 4. The performance of the model with and without real-time updating (deterministic mode and hindcast mode) is assessed (Sect. 5), and finally conclusions on the usefulness of such a prediction framework are drawn in Sect. 6.

2 Study area

This modelling approach is evaluated in three different sub-basins within the Zambezi River Basin (Fig. 1). The Zambezi river basin, the fourth largest river basin in Africa, lies in the semi-arid zone of southern Africa. Rainfall is strongly seasonal and occurs almost exclusively between October and March. The total amount of rainfall is on

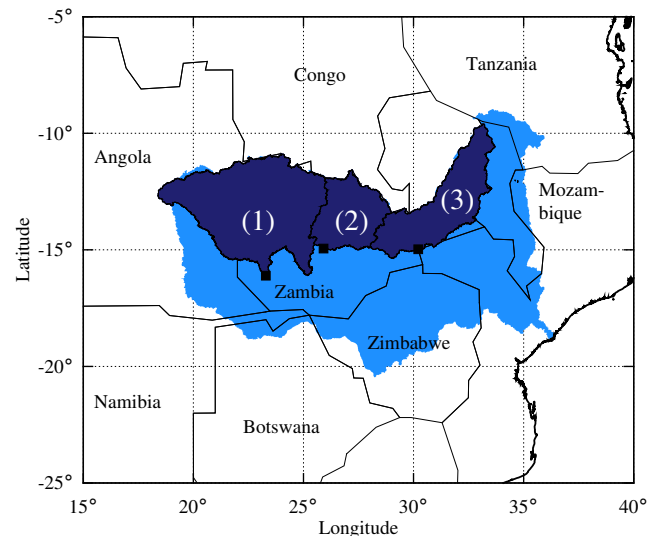


Fig. 1. Overview of the Zambezi River Basin (light blue) and the three watersheds where the model is applied. (1) Upper Zambezi; (2) Kafue River; (3) Luangwa River. The outlets of the watersheds are marked (■).

average around 1000 mm yr^{-1} , the potential evapotranspiration around 2000 mm yr^{-1} .

The three watersheds are (1) the Upper Zambezi upstream of the gauging station Senanga with an area of $281\,000 \text{ km}^2$, (2) the Kafue River where the discharge is measured at the Kafue Hook Bridge with an area of $95\,300 \text{ km}^2$ and (3) the Luangwa River which is gauged a few kilometers upstream of the confluence with the Zambezi River ($142\,070 \text{ km}^2$). These watersheds cover together more than one third of the whole Zambezi watershed and contribute more than one half of the total runoff at the mouth of the Zambezi, where the Upper Zambezi catchment contributes the largest amount ($850 \text{ m}^3 \text{ s}^{-1}$ mean annual discharge), the Kafue River discharges $300 \text{ m}^3 \text{ s}^{-1}$ and the Luangwa River $700 \text{ m}^3 \text{ s}^{-1}$.

The Upper Zambezi watershed is mainly characterized by gentle slopes and large floodplains along the course of the river. The largest floodplain, the Barotse Plains, spreads along the river for around 200 km, a maximal width of 50 km and an area of around 7500 km^2 . Similarly, the Kafue river basin features a floodplain, the Luangwa Swamps with an area of around 2500 km^2 . The main effect of the floodplains is the attenuation of the flow and the loss of water through increased evaporation (Vörösmarty and Moore, 1991). The Luangwa river valley is an extension of the East African rift valley characterized by steep slopes mainly in the upstream. The tributaries of the Luangwa drain the steep escarpment or the rift valley and therefore have a quick response to rainfall (Winsemius et al., 2009).

3 Data

3.1 Soil moisture

The vadose zone is one of the most important components of the global water cycle. Through the coupling of water and energy fluxes, soil moisture determines the local and global climate. However, the variability of soil moisture is very high in both space and time. Traditional measurement methods, such as Time Domain Reflectometry (TDR), are reasonably accurate, but they provide information on a very local scale. Monitoring large areas is nearly impossible. Typically, a correlation between such measurements can be observed if they are not more than a few tens of meters to a few hundred meters apart (Western et al., 2004). On larger scales the temporal variation seems to be very stable since it is mainly influenced by climatic conditions (Brocca et al., 2010). Therefore, remotely sensed soil moisture data provide a valuable dataset for hydrological monitoring on a larger scale.

There is a wide variety of techniques for measuring the soil water content through remote sensing. However, the data can only be retrieved by indirect measurements. Both active and passive methods rely on the measurement of radiation intensities in a certain range of wavelengths. For passive systems operating in the thermal infrared band the measurement target is the soil surface temperature (Verstraeten et al., 2006). The radiation intensity measured by systems operating in the microwave band is controlled by the dielectric constant of the top soil layer. Especially soil moisture products based on radar satellite imagery provide an attractive source for data since the influence of cloud cover and changing atmospheric conditions is minimal. Although they are governed by the same physical principles, generally three types of microwave techniques are distinguished: Passive radiometry and the two active methods using synthetic aperture radar (SAR) and radar scatterometer (Wagner et al., 2007). SAR systems, providing data with a high spatial resolution, show a good performance over bare soils. Despite the significantly lower spatial resolution, the multiple antenna configuration of scatterometers can facilitate the data processing to reduce the influence of vegetation on the signal (Baghdadi et al., 2008).

In this study the dataset of soil moisture derived from the radar scatterometer on board the two European remote-sensing satellites (ERS) is used (Wagner et al., 1999b). The ERS radar scatterometers are operating in the C-band at a frequency of 5.3 GHz using three sideways-looking antennas arranged at an angle of 45 degrees.

The measured back-scattering intensity is dependent on different properties of the surface, mostly on the surface roughness, the vegetation and the soil moisture. Generally wet soil results in higher back-scattering intensity than a dry soil. Since the arrangement of the antennas allows for the ruling out of the effects of vegetation, and the surface rough-

ness can be considered to be constant over time, the dry and the wet state of each pixel can be determined using a change detection algorithm (Wagner et al., 1999a).

3.2 Rainfall

Besides the soil moisture data which are described in the previous section, rainfall data are used as forcing data and measured discharge is applied for updating the model.

The rainfall dataset is provided by the Famine Early Warning Systems Network (FEWS NET) and can be downloaded free of charge from the internet. The data are available at a 10 day interval starting from July 1995. FEWS NET rainfall data incorporate various satellite-based rainfall estimates and data measured at gauging stations (Herman et al., 1997). Since the soil moisture data used are available only up to January 2002, the period where soil moisture data and rainfall data overlap is only little more than six years.

3.3 Discharge

Daily discharge data are available at the outlets of the three sub-basins. For the Kafue and the Luangwa sub-basin the data available cover the whole period where rainfall and soil moisture data are available simultaneously. The discharge of the Upper Zambezi sub-basin is measured from October 1996 only.

4 Methodology

4.1 Soil water column content

While the surface soil moisture can be directly derived from the scatterometer data, the soil column water content has to be estimated. This is due to the fact that the electromagnetic waves in the microwave bandwidth only penetrate the top few centimeters of the soil. Wagner et al. (1999b) proposed a method to calculate a so-called Soil Water Index (SWI) based on a simple two-layer moisture balance model. It computes a weighted average of the past measurements using an exponential filter of the form $\exp(-\frac{t}{T})$ and therefore acts as a low-pass filter. A value of 20 days was assigned to the parameter T globally. The dataset used in this study provides SWI data at a 10-daily time step. Scipal et al. (2005) concluded that even this low resolution soil moisture data can be applied in hydrological modelling.

The scatterometer based soil moisture data are strongly correlated to the occurrence of rainfall events but less correlated to the magnitude of these rainfall events (Fig. 2). In all three catchments the probability that a rainfall event has taken place grows for higher soil moisture in the same period, whereas the amount of rainfall is mainly correlated to soil moisture when the soil has not yet reached a certain degree of saturation (around values of 0.8).

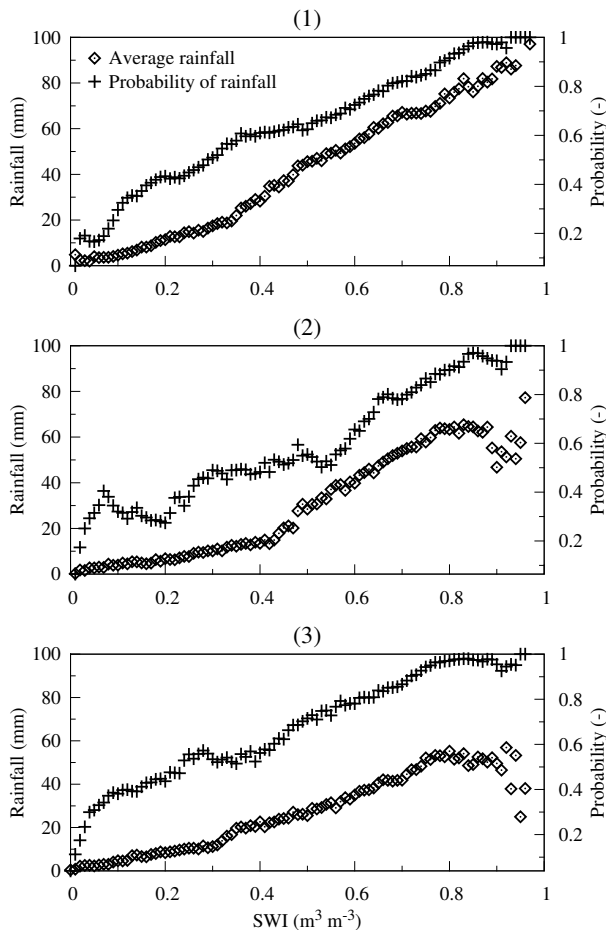


Fig. 2. The probability of rainfall event given a Soil Water Index (SWI) class and the average rainfall amount for the same classes for the three watersheds Upper Zambezi (1), Kafue River (2) and Luangwa River (3).

A similar effect can be observed if the correlation between the Basin Water Index (BWI) and the measured discharge is analyzed (Fig. 3). The BWI is calculated by averaging the SWI over the whole river basin (Scipal et al., 2005). The variation of the discharge is relatively small for low soil moisture values. If the values exceed 0.5 to 0.6 the variation suddenly increases. This indicates that the discharge is to some extent decoupled from the soil moisture as the soil approaches complete saturation. This decoupling is mainly caused by rainfall. Therefore modelling efforts which include rainfall data seem to be more realistic. It also shows that to obtain the best correlation the discharge has to be shifted by very different time lags. Not surprisingly, the largest watershed shows a long time lag of two months. For the second largest watershed (Luangwa River) the time lag has to be set to zero to obtain an optimal fit, indicating that the response time is significantly less than 10 days. The optimal time lag for the Ka-

fue River up to Kafue Hook Bridge is one month. These differences can be explained by the geological and geomorphic settings of the watersheds. The more gentle slopes of the Upper Zambezi and the Kafue River and the large wetlands retard the flow of the water. The water storage in wetlands cannot be tracked by the soil moisture product used. These wetlands are usually small compared to the area of the whole watershed. They therefore cause an additional retardation of the discharge which is mainly formed outside the wetlands.

4.2 Soil moisture – runoff model

In hydrological forecasting, fully distributed, physically based models provide the ability to account both for the heterogeneity of a watershed and physical changes of the system (e.g. induced by the construction of irrigation schemes or land use change). On the other hand, simple conceptual models can provide a satisfactory performance for forecasts. This can be an advantage, especially in regions with limited facilities for the measurement of relevant hydrological data.

The finding that the BWI correlates well with the observed discharge only for low values of BWI leads to the conclusion that additional data are needed to model the discharge. This is the reason why rainfall data are included in the model.

To model the discharge at the outlet of a basin a simple conceptual model was developed. The model consists of two compartments: a surface water storage and a subsurface water storage (Fig. 4). All input data of the model, the soil moisture and the rainfall, are averaged over the whole river basin. Hence, the spatial variability is not considered. The basin-averaged soil moisture is equivalent to the Basin Water Index (BWI) introduced by Scipal et al. (2005). The model we developed in this study is based on the following balance equations:

$$I_{GW} = k_1 AR(t)(1 - BWI(t)) \quad (1)$$

$$\frac{\Delta S_S(t)}{\Delta t} = k_1 AR(t) - I_{GW} - k_2 S_S(t-1) \quad (2)$$

$$\frac{\Delta S_{GW}(t)}{\Delta t} = \max(I_{GW} + k_3(BWI(t) - BWI(t-1)); 0) - k_4 S_{GW}(t-1) \quad (3)$$

where S_S and S_{GW} are the surface storage volume and the subsurface storage volume in m^3 , respectively, I_{GW} is the direct infiltration of rainfall to the subsurface storage in $m^3/10d$, R is the average rainfall in the river basin in $mm/10d$, A is the total area of the watershed in km^2 and k_i are the model parameters. The BWI is dimensionless and can take values between 0 and 1.

The storage compartments are considered to be single linear storages. Depending on the value of the BWI, a part of the rainfall is routed to the surface water storage whereas the remaining water volume is routed to the subsurface storage. If BWI is 0 all water is routed to the subsurface. If BWI is 1

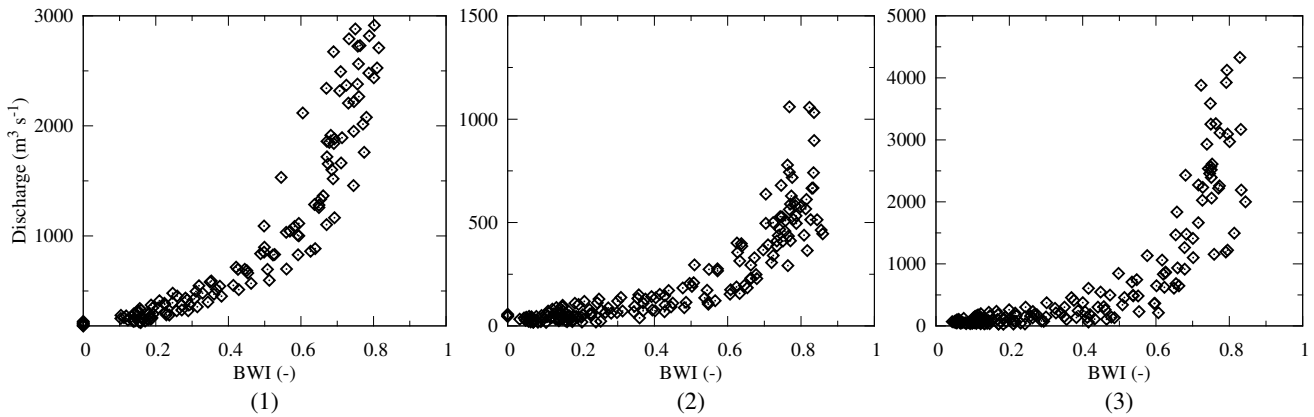


Fig. 3. Correlation between Basin Water Index (BWI) and discharge shifted by the time lag ($\Delta\tau$) which resulted in the best correlation. (1) Upper Zambezi: $\Delta\tau = 60$ d; (2) Kafue River: $\Delta\tau = 30$ d; (3) Luangwa River: $\Delta\tau = 0$ d.

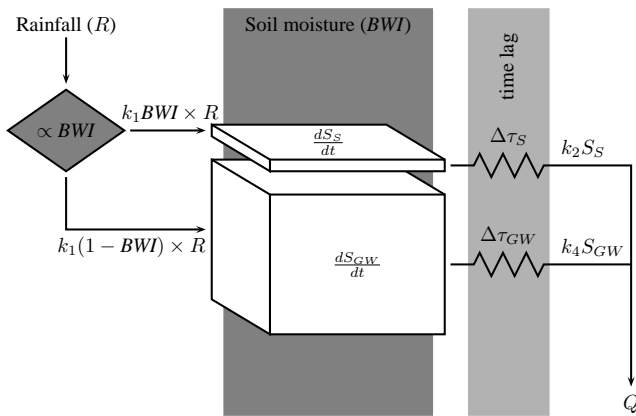


Fig. 4. Structure of the conceptual hydrological model.

all water is routed to the surface storage. The storage change in the subsurface is modeled through the measured change in soil moisture ($BWI(t) - BWI(t - 1)$). The sum of the rainfall routed to the subsurface and the measured change in soil moisture represent the recharge to the subsurface compartment which is not allowed to be negative in this model. For the surface runoff (Q_S) and the subsurface runoff (Q_{GW}) two different time lags $\Delta\tau_S$ and $\Delta\tau_{GW}$ are applied to calculate the total discharge according to Eq. (4).

$$Q(t) = Q_S(t - \Delta\tau_S) + Q_{GW}(t - \Delta\tau_{GW}) \quad (4)$$

with

$$Q_S(t) = k_2 S_S(t - 1) \quad \text{and} \quad Q_{GW}(t) = k_4 S_{GW}(t - 1).$$

According to Eqs. (1–4) the parameter k_1 is dimensionless, the parameters k_2 and k_4 have the unit s^{-1} and k_3 has the unit $m^3 s^{-1}$. A physical interpretation of the parameters assigns the parameter k_1 to the losses through evaporation and interception of rainfall before it enters a storage. Besides being

dependent on the infiltration properties of the soil it also depends on the average retention time in the catchment and is therefore correlated to the size of the watershed. For larger watersheds a lower value of k_1 is expected. The parameter k_3 relates the BWI to the total volume of water stored in the subsurface zone. The two parameters k_2 and k_4 are the rates at which the linear storage compartments are depleted. These model parameters are calibrated by running the model in deterministic mode using the Levenberg-Marquardt algorithm (Marquardt, 1963).

In a strict sense, parameter T which governs the estimation of the soil column water content (see Sect. 4.1) is a model parameter for itself. It drives the infiltration rate at which the measured water content at the surface flows downwards. A sensitivity analysis revealed that parameter T has only a marginal influence on the goodness of the fit between the observed and the modeled discharge.

The time lags ($\Delta\tau_S$ and $\Delta\tau_{GW}$) are mostly dependent on the size of the watershed. In this model the time lags are considered to be an integral multiple of the length of a single time step ($\Delta\tau = n\Delta t$, $n = 1, 2, 3, \dots$). Due to the discrete nature of the time lag, the parameter identification is done in two steps. For different pairs of $\Delta\tau_S$ and $\Delta\tau_{GW}$ the model parameters k_i are calibrated. The set of time lags with the minimal root mean square error (RMSE) between the observed and the computed flow together with the corresponding k_i are then chosen as the optimal parameter set. Since the input data used are available every ten days the time step Δt of the model was set to ten days.

A long time lag entails a long potential forecast period. Therefore, models using only soil moisture and rainfall as input have a longer lead time in larger watersheds.

4.3 Reference method

The regression model developed by Scipal et al. (2005) was used for comparison. This model uses a logarithmic

regression between soil moisture and discharge Eq. (5). It uses three parameters representing a baseflow (Q_0), a hydro-metric scaling factor χ_Q and a time lag $\Delta\tau$.

$$Q(t) = Q_0 + \chi_Q \ln\left(\frac{\text{BWI}_{\max}}{\text{BWI}_{\max} - \text{BWI}(t - \Delta\tau)}\right) \quad (5)$$

To assess the overall quality of the model presented in this article it is run in deterministic mode, without the data assimilation step. The simulated discharges are compared to the measured ones by calculating the root mean squared error (RMSE) and the Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970). Furthermore, they are compared to the discharges simulated with the reference model Eq. (5).

4.4 Real-time model

For effective real-time forecast of discharge the modeled states of the system need to be updated using observed discharges. This so-called data assimilation problem can be solved in different ways. In real-time applications a new assimilation problem is formulated at every time step. To solve this problem efficiently sequential assimilation techniques are considered superior (McLaughlin, 2002). Sequential assimilation algorithms, also known as filtering algorithms, are divided into two steps: first a propagation step, where the system states are propagated through time using a model and forcing data; second an update step, where the modeled states of the system are updated based on the difference between the model output and the real-world observation. To solve nonlinear filtering problems the Ensemble Kalman Filter (EnKF) has proven to be both efficient and robust (Evensen, 1994). EnKF has, along with standard batch calibration, the advantage of being able to incorporate a wide range of uncertainties. The uncertainties of forcing data, model parameters and model output are considered separately but can be incorporated in the same mathematical scheme (Thiemann et al., 2001).

The state variables which are updated in this study are the two storage volumes S_S and S_{GW} . As observation data the measured discharge is used. Observed discharge data are available on a daily basis. Since the temporal resolution of the model is 10 days, data assimilation is carried out in every time step.

For the real-time modelling framework an ensemble of randomly perturbed input and observation data are generated. The rainfall data ensemble is generated using the gamma distribution on the FEWS NET rainfall data. The gamma distribution only needs two parameters ($\Gamma(k, \theta)$). The expected value of a gamma distributed random variable X is defined as $E(X) = k\theta$ and the standard deviation as $\sigma_X = \sqrt{k}\theta$. The standard deviation is set to a fixed value ($\sigma_R = 50$ mm) which reflects the uncertainty of the rainfall data product (Herman et al., 1997). Using the measured rainfall amount as expected value of the perturbed rainfall for each time step, the two parameters k and θ can be calculated. Based on these

parameters the rainfall ensemble $\tilde{\mathbf{R}}_t$ at time t is generated according to Eq. (6).

$$\tilde{\mathbf{R}}_t = \begin{pmatrix} \tilde{R}_t^1 \\ \vdots \\ \tilde{R}_t^i \end{pmatrix}, \quad \text{with } \tilde{R}_t^i \sim \Gamma(k; \theta) = \Gamma\left(\frac{R_t^2}{\sigma_R^2}; \frac{\sigma_R^2}{R_t}\right), \quad (6)$$

where R_t is the measured rainfall at time t .

For the uncertainty of the BWI the standard deviation is set to 0.025. This is according to the standard error found by Ceballos et al. (2005) for the SWI. Whether this uncertainty can be translated directly to the BWI is unclear. Since the BWI is the arithmetic mean of all the SWI values in the watershed the uncertainty could be lower as well. A detailed analysis showed that this uncertainty has almost no influence on the quality of the forecast. The dominant sources of uncertainty in the real-time modelling framework are the rainfall data.

For the observed discharge data the variance of the added noise is proportional to their magnitude with a standard deviation of 0.05 times the measured value, as the absolute measurement error of discharge measurements is generally considered to be dependent on the discharge itself. Both the BWI and the discharge perturbations are considered to follow a Gaussian distribution.

To assess the possible accuracy of the forecast, the model is run in hindcast mode which includes the data assimilation step for the historic time series from the years 1995 to 2002.

5 Results and discussion

The parameters for the developed model are calibrated for all three watersheds. To analyze the performance of the model it was assessed both in a deterministic modelling mode and in hindcast modelling mode.

5.1 Deterministic model

The model parameters obtained by calibration in the deterministic mode are shown in Table 1. For the time lags one can see a similar dependency on the size and geomorphology of the different watersheds as it was already observed for the correlation analysis (Fig. 3). The Upper Zambezi catchment has by far the longest response time whereas the Luangwa river basin shows a relatively quick response.

The only model parameter that shows a distinct dependency on the area of a watershed is the parameter (k_3) that correlates the BWI to the total volume of water stored inside the subsurface storage. One can assume that the wetlands present in the Upper Zambezi and the Kafue watersheds have a huge impact on the water storage capacity.

The parameter k_1 mainly reflects the water losses in the watershed. The Upper Zambezi basin, being the largest watershed, shows the lowest value for k_1 . The Kafue River basin shows a very similar value of k_1 which suggests that the soil properties in the two basins are similar. However, due to the

Table 1. Estimated parameters for the three sub-basins and the 95% confidence interval for each parameter.

	Upper Zambezi	Kafue River	Luangwa River
$\Delta\tau_S$	40 d	20 d	10 d
$\Delta\tau_{GW}$	100 d	70 d	50 d
$k_1 (\times 10^{-5})$	4.22 ± 1.13	5.00 ± 0.99	10.44 ± 3.68
$k_2 (s^{-1})$	0.22 ± 0.09	0.29 ± 0.09	0.68 ± 0.46
$k_3 (\times 10^3 m^3 s^{-1})$	32.23 ± 6.41	5.61 ± 1.32	18.06 ± 5.58
$k_4 (s^{-1})$	0.15 ± 0.03	0.13 ± 0.03	0.35 ± 0.09

absence of detailed information on the soil properties in the area this statement cannot be verified. The influence of the size of the watershed on the parameter k_1 is marginal. While the area of the Luangwa basin is only one half of the size of the Upper Zambezi watershed the parameter value of k_1 is twice as big. This supports the conclusion that the water is drained quickly from the surface to the river and therefore losses are low.

While the parameters k_2 and k_4 show similar values for the Upper Zambezi basin and the Kafue river, they are much higher for the Luangwa river. This correlates well with the generally steeper slopes in the Luangwa basin where water flows faster.

The performance of the model running in the deterministic mode is illustrated in Fig. 5 and compared to the reference method. The modeled discharge agrees in general quite well with the measured flows. For all the applied models the root mean squared error (RMSE) and the Nash-Sutcliffe efficiency are calculated (Tables 2 and 3). The model developed gives better results than the reference method for the Upper Zambezi and the Kafue River basin where both the Nash-Sutcliffe efficiency and the RMSE are slightly higher or lower, respectively. It is clearly visible that the flood maxima are reproduced much better than in the reference method. Obviously the inclusion of rainfall information is most beneficial for the maxima as the correlation between BWI and rainfall deteriorates for large precipitation events. However, for the Luangwa river the model performance was not satisfactory when running in the deterministic mode. It even showed a slightly poorer performance than the reference method which also does not allow for adequate modelling of the situation.

Due to the parameter uncertainty the error band becomes large, especially in the wet season. This uncertainty is mostly attached to the model parameter k_1 since with a high rainfall amount a slight change in the parameter can greatly affect the amount of water which is routed to the system.

A drawback for the testing of the method is the short time period over which data are available. The model relies on soil moisture data, on rainfall data and on measured

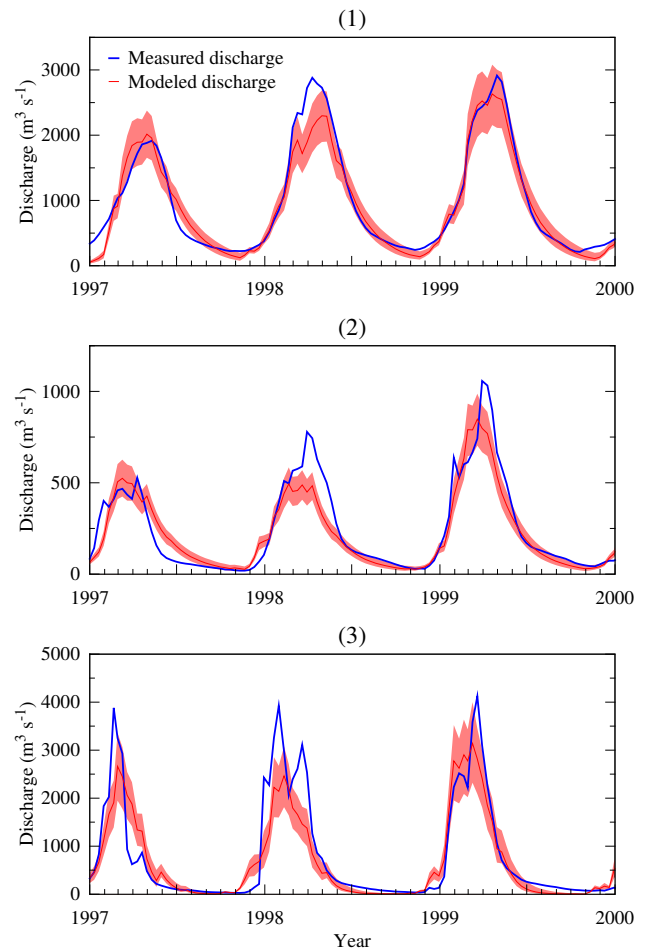


Fig. 5. The discharge simulated in the hindcast mode (red line) including the 95% confidence interval compared to the measured discharge (blue line) for the three watersheds Upper Zambezi (1), Kafue River (2) and Luangwa River (3).

discharge. The overlap of these three datasets dictates the longest continuous time span that can be modeled. Eventually it is only possible to test the model on a period of a bit more than six years, for the Upper Zambezi catchment even less. For this reason a validation of the model was not possible. The available data were used to obtain a stable calibration.

The application of the model provides a seamless integration of remote sensing products. With only four parameters and a simple conceptual formulation this model is applicable to a class of watersheds which comply with certain characteristics. All data used are developed to an operational standard. Therefore the user does not have to undertake extensive data processing. This model is especially suited for use in a real-time modelling framework.

Table 2. RMSE of the different forecast lead times up to the maximum possible lead time ($\Delta\tau_S$). For comparison the RMSE of the deterministic model run and the reference method are indicated. All values are given in $\text{m}^3 \text{s}^{-1}$.

	Upper Zambezi	Kafue River	Luangwa River
Lead time:			
Lag 4 (40 d)	281.5		
Lag 3 (30 d)	265.1		
Lag 2 (20 d)	238.8	70.5	
Lag 1 (10 d)	199.1	59.8	483.6
assimilated	131.4	46.5	412.6
Deterministic	269.3	99.5	513.7
Reference	285.0	103.8	502.6

Table 3. Nash-Sutcliffe efficiency of the different forecast lead times up to the maximum possible lead time ($\Delta\tau_S$). For comparison the Nash-Sutcliffe efficiency of the deterministic model run and the reference method are indicated.

	Upper Zambezi	Kafue River	Luangwa River
Lead time:			
Lag 4 (40 d)	0.85		
Lag 3 (30 d)	0.87		
Lag 2 (20 d)	0.90	0.84	
Lag 1 (10 d)	0.93	0.88	0.68
assimilated	0.98	0.96	0.80
Deterministic	0.90	0.82	0.74
Reference	0.88	0.80	0.75

5.2 Real-time model

For the successful application of the model for real-time prediction it does not necessarily have to be mechanistically correct but needs to reflect the correct tendency. When operating in the data assimilation mode the quality of the forecast is of interest. The length of the forecast period is defined by the shortest time lag ($\Delta\tau$) in the model. The ensemble of the forecast can be represented by the ensemble mean and the confidence interval. As time approaches the time of prediction for a certain timestep the error generally gets smaller (Fig. 6a, 6b and 6c). This statement is supported by the analysis of the RMSE and the Nash-Sutcliffe efficiency (Tables 2 and 3). While the prediction for the maximum forecast time shows the highest RMSE and lowest coefficient of efficiency, the model prediction gets significantly more accurate for shorter forecast periods. Again the absolute error of the prediction is much higher in the wet season, whereas the relative error is especially high during ascending and receding flows (Fig. 6a, 6b and 6c).

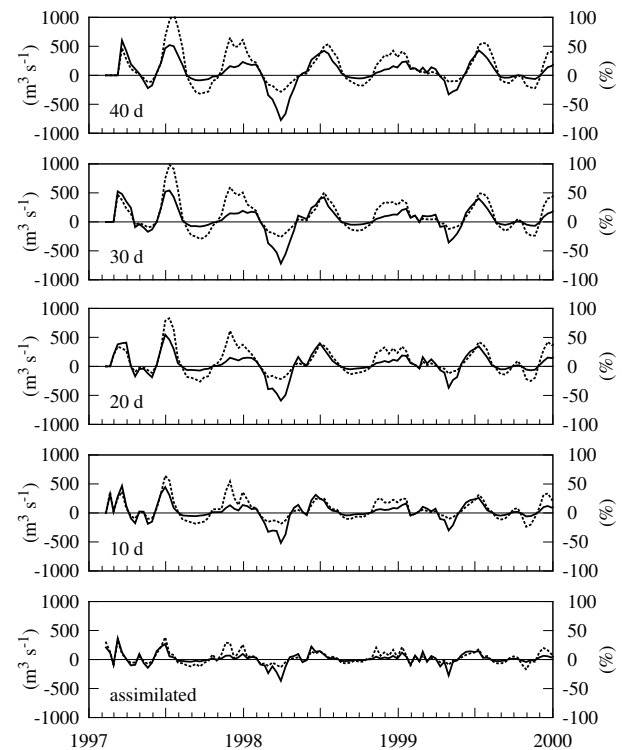


Fig. 6a. Absolute (solid line) and relative (dashed line) forecast error for the Upper Zambezi watershed for the different forecast periods and the assimilation step.

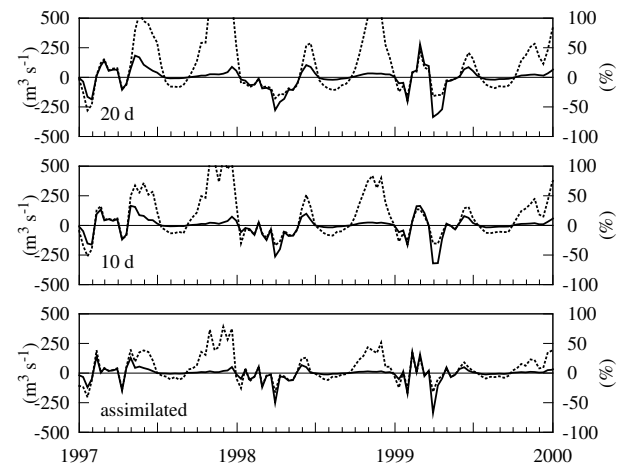


Fig. 6b. Absolute (solid line) and relative (dashed line) forecast error for the Kafue watershed for the different forecast periods and the assimilation step.

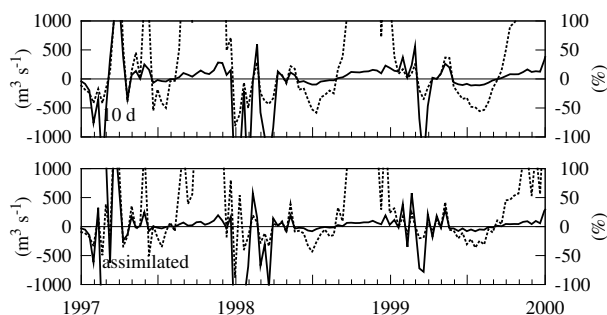


Fig. 6c. Absolute (solid line) and relative (dashed line) forecast error for the Luangwa watershed for the different forecast periods and the assimilation step.

These results show that the model presented is capable of providing useful discharge forecasts in semi-arid river basins. Yet this model can not be applied to every river since its model structure is not designed to reproduce the processes in watersheds with a relatively low storage volume and a quick response to rainfall events. This is the case in the Luangwa river basin where uncertainties become very large.

The SWI which was used to calculate the BWI assumes a uniform soil thickness everywhere. The actual thickness of the soil in a river basin does not have a big influence on the model results. The spatial variability of the soil thickness, however, has a huge influence on the results because certain areas with a relatively thin soil layer can suddenly dominate the behaviour of the system.

6 Conclusions

Hydrological modelling under conditions of data scarcity remains a big challenge. Recent developments offer promising opportunities to advance in this field. On one hand, real-time modelling techniques allow the assimilation of data in a model, updating the modeled system state every time observation data are available. On the other hand, techniques for extracting information on the hydrological cycle from remote sensing data have advanced in the past few years. While some years ago the satellite systems were designed to gather as many types of data as possible in order to provide the scientific community with data that could be exploited in several ways, nowadays satellite missions are designed for a specific purpose. Several satellite missions have been deployed recently, or will be launched in the near future, especially for the retrieval of soil moisture.

Radar scatterometer data were found especially promising for use in hydrological modelling where the soil moisture is one of the most important parameters. Since the radar signal penetrates only the top few centimeters of the soil, hence only giving information on the surface soil moisture, the water content in the soil column has to be modeled. The

simple two-layer model used to generate the SWI produces data which are appropriate to be used as input data for a conceptual model. Since rainfall is one important driver of soil moisture a conceptual model should also utilize rainfall data.

The prediction framework presented in this paper exploits the available data sets on rainfall and soil moisture. The relatively simple model consisting of two reservoirs, for the surface water and the subsurface water, and an infiltration process based on the soil moisture, shows a fairly good performance. Especially in watersheds where the storage of water in the soil is of high importance the model predictions are accurate. In the Luangwa river basin, which is dominated by steep slopes and quick runoff formation, the model's performance is not satisfactory.

Running the model in real-time with a data assimilation procedure provides long term forecasts which can be used for a wide variety of applications. To manage a river basin system such a forecast is beneficial since the discharge expected for the next few weeks can be quantified. Releases for power production, irrigation water demands or ecological flood releases can be planned based on this information.

If water management options for a period exceeding the forecast lead time have to be assessed the described model is not suitable because it is not physically based. Due to the relatively long time step flood forecasting is also not possible. If the quality of the input data is greatly improved flash flood forecasting could eventually be an option. More and more data from the newer satellite systems will be available in real-time. A higher temporal and/or a higher spatial resolution can greatly improve modelling efforts.

A higher spatial resolution of the data (e.g. as provided by Envisat ASAR) would allow a higher spatial resolution of the model. Since the BWI and the rainfall are averaged over the whole area, the runoff processes are also averaged. The spatial variability of the different runoff processes is completely neglected. If heavy convective rainfall is occurring in an area with high soil moisture a peak in the runoff should be observed. Therefore if the soil moisture data allow the model to divide the watershed into sub-basins a better prediction of peak discharges could be feasible. A higher temporal resolution of soil moisture data can allow models to account for the usually high temporal variability of the soil water content.

Further research should focus on the improvement of the quality of the data and the development of more sophisticated hydrological models tailored to use remotely sensed soil moisture data. The quantification of the relevant uncertainties demands attention as well. If the generally high uncertainties of the prediction presented in this study can be minimized such a forecast can be used for applications where low uncertainties are necessary.

Acknowledgements. The soil moisture data were kindly provided by the Institute of Photogrammetry and Remote Sensing at the Vienna University of Technology. The discharge data of the three rivers were provided by the Department of Water Affairs, Lusaka, Zambia. This research is funded by the Swiss National Science Foundation under project K-21K1-120266.

We would like to thank the anonymous reviewers for the time they invested in improving this article.

Edited by: F. Pappenberger

Reviewed by: three anonymous referees

References

- Acreman, M. C. and Ferguson, A. J. D.: Environmental flows and the European Water Framework Directive, *Freshwater Biol.*, 55, 32–48, 2010.
- Aubert, D., Loumagne, C., and Oudin, L.: Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall-runoff model, *J. Hydrol.*, 280, 145–161, doi:10.1016/S0022-1694(03)00229-4, 2003.
- Baghdadi, N., Cerdan, O., Zribi, M., Auzet, V., Darboux, F., El Hajj, M., and Kheir, R. B.: Operational performance of current synthetic aperture radar sensors in mapping soil surface characteristics in agricultural environments: application to hydrological and erosion modelling, *Hydrol. Process.*, 22, 9–20, doi:10.1002/hyp.6609, 2008.
- Brocca, L., Melone, F., Moramarco, T., and Morbidelli, R.: Spatial-temporal variability of soil moisture and its estimation across scales, *Water Resour. Res.*, 46, W02516, doi:10.1029/2009WR008016, 2010.
- Ceballos, A., Scipal, K., Wagner, W., and Martinez-Fernandez, J.: Validation of ERS scatterometer-derived soil moisture data in the central part of the Duero Basin, Spain, *Hydrol. Process.*, 19, 1549–1566, doi:10.1002/hyp.5585, 2005.
- Crow, W. T. and Ryu, D.: A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals, *Hydrol. Earth Syst. Sci.*, 13, 1–16, doi:10.5194/hess-13-1-2009, 2009.
- Evensen, G.: Sequential Data Assimilation With A Nonlinear Quasi-Geostrophic Model Using Monte-Carlo Methods To Forecast Error Statistics, *J. Geophys. Res.-Oceans*, 99, 10143–10162, 1994.
- Galat, D. and Lipkin, R.: Restoring ecological integrity of great rivers: historical hydrographs aid in defining reference conditions for the Missouri River, *Hydrobiologia*, 422, 29–48, 2000.
- Herman, A., Kumar, V. B., Arkin, P. A., and Kousky, J. V.: Objectively determined 10-day African rainfall estimates created for famine early warning systems, *Int. J. Remote Sens.*, 18, 2147–2159, 1997.
- Kerr, Y., Waldteufel, P., Wigneron, J.-P., Martinuzzi, J., Font, J., and Berger, M.: Soil moisture retrieval from space: the Soil Moisture and Ocean Salinity (SMOS) mission, *IEEE T. Geosci. Remote*, 39, 1729–1735, doi:10.1109/36.942551, 2001.
- Kitanidis, P. K. and Bras, R. L.: Real-Time Forecasting With a Conceptual Hydrologic Model 1. Analysis of Uncertainty, *Water Resour. Res.*, 16, 1025–1033, doi:10.1029/WR016i006p01025, 1980a.
- Kitanidis, P. K. and Bras, R. L.: Real-Time Forecasting With a Conceptual Hydrologic Model 2. Applications and Results, *Water Resour. Res.*, 16, 1034–1044, doi:10.1029/WR016i006p01034, 1980b.
- Marquardt, D.: An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Ind. Appl. Math.*, 11, 431–441, 1963.
- McLaughlin, D.: An integrated approach to hydrologic data assimilation: interpolation, smoothing, and filtering, *Adv. Water Resour.*, 25, 1275–1286, 2002.
- Naeimi, V., Bartalis, Z., and Wagner, W.: ASCAT Soil Moisture: An Assessment of the Data Quality and Consistency with the ERS Scatterometer Heritage, *J. Hydrometeorol.*, 10, 555–563, doi:10.1175/2008JHM1051.1, 2009.
- Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I – A discussion of principles, *J. Hydrol.*, 10, 282–290, 1970.
- Piles, M., Entekhabi, D., and Camps, A.: A change detection algorithm for retrieving high-resolution soil moisture from SMAP radar and radiometer observations, *IEEE T. Geosci. Remote*, 47, 4125–4131, doi:10.1109/TGRS.2009.2022088, 2009.
- Scipal, K., Scheffler, C., and Wagner, W.: Soil moisture-runoff relation at the catchment scale as observed with coarse resolution microwave remote sensing, *Hydrol. Earth Syst. Sci.*, 9, 173–183, doi:10.5194/hess-9-173-2005, 2005.
- Thiemann, M., Trosset, M., Gupta, H., and Sorooshian, S.: Bayesian Recursive Parameter Estimation For Hydrologic Models, *Water Resour. Res.*, 37, 2521–2535, doi:10.1029/2000WR900405, 2001.
- Verstraeten, W., Veroustraete, F., van der Sande, C., Grootaers, I., and Feyen, J.: Soil moisture retrieval using thermal inertia, determined with visible and thermal spaceborne data, validated for European forests, *Remote Sens. Environ.*, 101, 299–314, doi:10.1016/j.rse.2005.12.016, 2006.
- Vörösmarty, C. J. and Moore, B.: Modeling basin-scale hydrology in support of physical climate and global biogeochemical studies: An example using the Zambezi River, *Surv. Geophys.*, 12, 271–311, 1991.
- Wagner, W., Lemoine, G., Borgeaud, M., and Rott, H.: A study of vegetation cover effects on ERS scatterometer data, *IEEE T. Geosci. Remote*, 37, 938–948, doi:10.1109/36.752212, 1999a.
- Wagner, W., Lemoine, G., and Rott, H.: A method for estimating soil moisture from ERS scatterometer and soil data, *Remote Sens. Environ.*, 70, 191–207, 1999b.
- Wagner, W., Bloeschl, G., Pampaloni, P., Calvet, J.-C., Bizzarri, B., Wigneron, J.-P., and Kerr, Y.: Operational readiness of microwave remote sensing of soil moisture for hydrologic applications, *Nord. Hydrol.*, 38, 1–20, doi:10.2166/nh.2007.029, 2007.
- Western, A. W., Zhou, S.-L., Grayson, R. B., McMahon, T. A., Blöschl, G., and Wilson, D. J.: Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes, *J. Hydrol.*, 286, 113–134, doi:10.1016/j.jhydrol.2003.09.014, 2004.
- Winsemius, H. C., Schaeffli, B., Montanari, A., and Savenije, H. H. G.: On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information, *Water Resour. Res.*, 45, W12422, doi:10.1029/2009WR007706, 2009.