Real-time remote sensing driven river basin modelling using radar altimetry

S. J. Pereira-Cardenal¹, N. D. Riegels¹, P. A. M. Berry², R. G. Smith², A. Yakovlev³, T. U. Siegfried⁴, and P. Bauer-Gottwein¹

¹Department of Environmental Engineering, Technical University of Denmark, Miljøvej, Building 113, 2800 Kgs, Lyngby, Denmark
²Earth and Planetary Remote Sensing Laboratory, De Montfort University, The Gateway, Leicester, LE19BH, UK
³Hydrological Forecasting Department – UzHydromet, Tashkent, Uzbekistan
⁴The Water Center, The Earth Institute, Columbia University, New York, USA

Received: 20 October 2010 – Accepted: 21 October 2010 – Published: 25 October 2010
Correspondence to: P. Bauer-Gottwein (pbau@env.dtu.dk)

Published by Copernicus Publications on behalf of the European Geosciences Union.
Abstract

Many river basins have a weak in-situ hydrometeorological monitoring infrastructure. However, water resources practitioners depend on reliable hydrological models for management purposes. Remote sensing (RS) data have been recognized as an alternative to in-situ hydrometeorological data in remote and poorly monitored areas and are increasingly used to force, calibrate, and update hydrological models.

In this study, we evaluate the potential of informing a river basin model with real-time radar altimetry measurements over reservoirs. We present a lumped, conceptual, river basin water balance modelling approach based entirely on RS and reanalysis data: precipitation was obtained from the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA), temperature from the European Centre for Medium-Range Weather Forecast’s (ECMWF) Operational Surface Analysis dataset and reference evapotranspiration was derived from temperature data. The Ensemble Kalman Filter was used to assimilate radar altimetry (ERS2 and Envisat) measurements of reservoir water levels. The modelling approach was applied to the Syr Darya River Basin, a snowmelt-dominated basin with large topographical variability, several large reservoirs and scarce hydrometeorological data that is shared between 4 countries with conflicting water management interests.

The modelling approach was tested over a historical period for which in-situ reservoir water levels were available. Assimilation of radar altimetry data significantly improved the performance of the hydrological model. Without assimilation of radar altimetry data, model performance was limited, probably because of the size and complexity of the model domain, simplifications inherent in model design, and the uncertainty of RS and reanalysis data. Altimetry data assimilation reduced the mean error of the simulated reservoir water levels from 4.7 to 1.9 m, and overall model RMSE from 10.3 m to 6.7 m.

Because of its easy accessibility and immediate availability, radar altimetry lends itself to being used in real-time hydrological applications. As an impartial source of information about the hydrological system that can be updated in real time, the modelling
approach described here can provide useful medium-term hydrological forecasts to be used in water resources management.

1 Introduction

Hydrological models are constructed for two main purposes: to improve hydrological process understanding and to support practical decision making in water resources management. River basin management models typically operate at large scales and, given the complexity of most river basins, use semi-empirical lumped parameterizations of hydrological processes. Because the uncertainties inherent in such models are large, calibration and data assimilation techniques are essential to achieve satisfactory model performance. However, in-situ data availability is limited, particularly in the developing world, where many river basins are poorly gauged. Satellite-based data with high temporal resolution have the potential to fill critical information gaps in such ungauged or poorly gauged basins (e.g. Grayson et al., 2002; Lakshmi, 2004).

Remote sensing data can be used in hydrological models in two ways (Brunner et al., 2007): as input parameters (or forcing data) and as calibration data. The most popular remote sensing data sources for hydrological applications are multispectral imagery for the determination of actual evapotranspiration (e.g. Bastiaanssen et al., 1998; Jiang et al., 2001; Stisen et al., 2008b), active microwave sensors for mapping of soil moisture distribution (e.g. Parajka et al., 2006), total water storage change estimates from GRACE (e.g. Hinderer et al., 2006; Winsemius et al., 2006), and river and lake level variations from radar altimetry (e.g. Birkett, 2000; Alsdorf et al., 2001; Bjerklie et al., 2003; Calmant et al., 2008; Getirana et al., 2009) and interferometric SAR (e.g. Alsdorf et al., 2001; Wdowinski et al., 2004; Gondwe et al., 2010). Several previous studies have used remote sensing data in the context of river basin water balance modeling (e.g. Campo et al., 2006). Andersen et al. (2002) built a distributed hydrological model of the Senegal River Basin using precipitation derived from METEOSAT data and leaf area index (LAI) estimated from the normalized difference vegetation index (NDVI) from
NOAA AVHRR data. Stisen et al. (2008a) developed a distributed hydrological model of the same catchment using potential evapotranspiration (PET) estimated from global radiation, precipitation from satellite-derived cold cloud duration, and LAI calculated from NDVI. Boegh et al. (2004) used RS-data to derive PET and LAI as input to a distributed agro-hydrological model. Francois et al. (2003) used Synthetic Aperture Radar (SAR) estimates of soil moisture in a lumped rainfall-runoff model.

More recently, the potential of using remote sensing data in real time or near real time to update hydrological model state variables has been recognized. Data assimilation (DA) methods were first used in the fields of oceanography and meteorology, but have been used in hydrology since the 1990s (McLaughlin, 1995; Evensen, 2003). Several DA techniques are available, including the Particle Filter (Arulampalam et al., 2002) and the Reduced Rank Square Root Filter (Verlaan et al., 1997); The Ensemble Kalman Filter (Evensen, 2003) is used here because it has a simple conceptual formulation, it is easy to implement and is computationally efficient. Previous studies using data assimilation techniques on hydrological modeling included land surface models (e.g. Reichle et al., 2002), surface water models (e.g. Madsen et al., 2005) and groundwater models (e.g. Franssen et al., 2008). The value of DA for models used in water resources management is based on its ability to improve operational forecasts.

This study presents a semi-distributed river basin model of the Syr Darya River basin in Central Asia and analyses how operational model performance in real-time forecasting can be improved if the model is informed with real-time reservoir water levels based on radar altimetry. In situ data availability in the Syr Darya is extremely limited and model forcing variables are therefore exclusively based on remote sensing and reanalysis data. Precipitation is obtained from the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA; Huffman et al., 2007); daily temperature is obtained from the European Centre for Medium-Range Weather Forecast (ECMWF) Operational Surface Analysis dataset (Molteni et al., 1996); and reference evapotranspiration is derived from daily temperature using Hargreaves equation (Allen et al., 1998). The water level in a cascade of four reservoirs is simulated in the
model, and satellite radar altimetry data (Berry et al., 2005) are used to update the water level in the reservoirs using the Ensemble Kalman filter.

2 Study area

The Syr Darya River is located in the Central Asian republics of Kyrgyzstan, Uzbekistan, Tajikistan, and Kazakhstan and, along with the Amu Darya River, is one of two principal tributaries to the Aral Sea (Fig. 1). About 22 million people in the region depend on irrigated agriculture for their livelihoods, and 20% to 40% of GDP in the riparian countries is derived from agriculture, most of which is irrigated (Bucknall et al., 2003). Much of the region has an arid climate, with strongly seasonal precipitation and temperature patterns. The extensive development of irrigation in the basin is associated with a number of environmental problems including desiccation of the Aral Sea, which has lost up to 90% of its pre-1960 volume and has received international attention as an environmental disaster area (Micklin, 2007).

The Syr Darya River originates in the Tien Shan Mountains of Kyrgyzstan and is formed by the confluence of the Naryn and Karadarya rivers near the border of Kyrgyzstan and Uzbekistan. The population of the basin is approximately 20 million, with an area of about 400 000 km$^2$. Annual precipitation averages about 320 mm and ranges from 500–1500 mm in the mountain zones to 100–200 mm in desert regions near the Aral Sea (Schiemann et al., 2008). The bulk of runoff comes from melting snow and glaciers in the mountains of Kyrgyzstan. Because of the combined effects of snowmelt and glacial runoff, about 80% of runoff in the basin occurs between March and September. The onset of the snowmelt period shifts from early spring to early summer with increasing elevation, distributing snowmelt runoff over a period of several months. In the summer months, glacial ablation peaks and prolongs the period of peak runoff. Annual runoff averages about 39 km$^3$/year (approx. 96 mm/year), and approximately 90% of the river's mean annual flow is regulated by reservoirs (Savoskul et al., 2003).
The Syr Darya River was extensively developed for irrigation and hydropower during Soviet times, particularly after 1960, with the primary goal of producing cotton. Total irrigated area in the Aral Sea basin increased from 5 million hectares in 1965 to 7.9 million hectares in 2000 (Micklin, 2007). About 1.7 million hectares are currently irrigated directly from the Syr Darya River (Siegfried et al., 2007). Cotton is an important source of foreign exchange in Uzbekistan, and continued production through irrigated agriculture is a priority for the government (World Bank, 2004).

A significant change to the natural hydrological pattern of the basin occurred with the construction of the Toktogul Reservoir in 1974. Because the timing of the March–September natural runoff peak coincides with the irrigation season, substantial reservoir storage is not required to regulate seasonal runoff. However, the aggressive expansion of irrigation during Soviet times created a need for multi-year storage to store excess flows in wet years to supplement dry year flows. Toktogul Reservoir was constructed on the Naryn River (the principal tributary to the Syr Darya) to serve this purpose. The reservoir is the largest storage facility in the Aral Sea basin and has a total capacity of 19.5 km$^3$ (14 km$^3$ active storage). The construction of the reservoir was accompanied by the building of four smaller downstream reservoirs and power plants to maximize electricity generation from reservoir releases. The five facilities, commonly called the Naryn Cascade, have a combined generation capacity of 2870 MW (World Bank, 2004).

Toktogul Reservoir and the Naryn Cascade are at the heart of a dispute over management of the Syr Darya River that has existed since the downfall of the Soviet Union in 1991. In the Soviet system, the reservoir was operated to benefit irrigated agriculture and power was produced incidentally as flows were released to meet downstream demands. In 1992, the Central Asian riparian states agreed to continue Soviet water allocation policies and established the Interstate Commission for Water Coordination (ICWC) to oversee the allocation process. However, the new system immediately came under strain as the competing interests of the newly independent states emerged. Toktogul Reservoir came under the control of Kyrgyzstan, which is less dependent on
irrigated agriculture and lacks fossil fuel resources for energy generation. Kyrgyzstan had been supplied with fossil fuels under the Soviet system but found itself in the position of having to purchase energy supplies on world markets after 1991. Kyrgyzstan turned to hydropower for its own energy needs, which peak in winter because of heating demands, placing the country’s operational objectives in direct opposition to those of its downstream neighbors; Kyrgyzstan would prefer to store summer peak flows for winter power generation, while the downstream countries would like winter releases minimized to conserve water for the summer season (Biddison, 2002; World Bank, 2004; Siegfried et al., 2007). Increased winter releases also cause flooding, as many of the downstream irrigation works are not built to handle high flows and ice in the river bed reduces winter conveyance capacity (Biddison, 2002).

In these circumstances of mutual distrust between the up- and downstream countries, global datasets and assimilation of radar altimetry hold great promise for increasing transparency, reducing forecast uncertainty, and increasing the speed at which forecasts can be developed and updated. Because remotely-sensed data products are available to all, their increased use in the region has the potential to reduce distrust by providing a common base of information. The increasing availability of these products in real-time also has the potential to accelerate the forecasting process so that water allocation plans can be agreed upon earlier in the irrigation season.

3 Methods and data

3.1 River basin water balance modeling

The river basin water balance model was implemented as a combination of a rainfall-runoff model (NAM; Refsgaard et al., 1996; DHI, 2000) and a river network mass balance model (Mike Basin; DHI, 2009). Runoff in the subcatchments is simulated using NAM and is subsequently routed through the river network in Mike Basin. Irrigation agriculture districts are simulated as water users that abstract water from the Mike
Basin river network. The simulation is run in daily time steps from 1 January 2000 to 31 December 2007. A modeling flow chart is presented in Fig. 2.

The NAM model (Nedbør-Afstrømningsmodel, Danish for rainfall-runoff model) is a lumped conceptual modeling system consisting of mass balance equations that account for the water content in four different storages representing processes occurring in the land phase of the hydrological cycle: snow storage, surface storage, lower soil zone storage and groundwater storage (DHI, 2000). The minimum data requirements of the modeling system are precipitation, reference evapotranspiration and observed discharge. Daily mean temperature is also required if snowmelt contributes to runoff. A comparison of data requirements and performance of NAM with other lumped hydrological models is provided by Refsgaard et al. (1996). The four storages are typically modeled using a set of 17 parameters, about 10 of which are commonly used for model calibration.

Due to the limited amount of observed in-situ river discharge data, it was impossible to achieve a unique and stable calibration based on 10 free model parameters. Moreover, we observed that overland flow does not significantly contribute to the hydrological regime in the Syr Darya river basin. A more robust version of NAM was developed using only five free calibration parameters: two parameters describing surface and soil moisture storage and the others describing groundwater response times. The structure of this simplified version of NAM is shown in Fig. 3. Table 1 lists the parameters chosen to enforce this model structure. Because of its simplicity, the modeling approach is robust and appropriate given the general scarcity of observation data in the SDRB.

Precipitation falls as snow if the temperature is below 0 °C and as liquid precipitation (PL) otherwise. Each single catchment is divided into 10 separate elevation zones of equal area and the precipitation discrimination is done for each individual elevation zone separately, using temperature lapse rates based on Tsarev et al. (1994). Snow melt (SM) is modeled using a simple degree-day approach:
SM_{pot} = \begin{cases} M \times T, & T > 0 \\ 0, & \text{else} \end{cases} \quad (1)

SM = \min (SS, SM_{pot} \cdot \Delta t)

where \( \Delta t \) is the time step (days), \( SM_{pot} \) is the potential snowmelt (mm day\(^{-1}\)), \( T \) is the temperature in degrees Celsius, \( SS \) is the snow storage (mm), and \( M \) is a seasonally variable degree day factor (mm day\(^{-1}\) deg\(^{-1}\)) based on a parameterization proposed by Shenzis (1985) for the Central Asian Mountains. Snowmelt parameters used in the model are listed in Table 2. Snow melt is calculated for each elevation zone separately. Snow melt and liquid precipitation enter the surface storage (\( U \) in mm). The surface storage is depleted by evapotranspiration (ET). The ET rate is assumed to be equal to the reference ET rate. The surface storage has a maximum capacity (\( U_{\text{max}} \), in mm), which is a calibration parameter. If the surface storage is filled to maximum capacity, additional snow melt or liquid precipitation spills to the soil storage (\( L \), in mm). A balance equation is solved for the soil storage:

\[ \frac{dL}{dt} = SP - AET - PER \quad (2) \]

The symbol SP indicates spills from the surface storage to the soil storage. Water in the soil storage is depleted by actual soil ET (AET) and percolation (PER). Actual soil ET and percolation are calculated as functions of soil water storage:

\[ AET = (ET_{\text{ref}} - ET_U) \cdot \frac{L}{L_{\text{max}}} \quad (3) \]

\[ PER = SP \cdot \frac{L}{L_{\text{max}}} \quad (4) \]

where \( ET_{\text{ref}} \) is the reference ET (mm day\(^{-1}\)), \( ET_U \) is the ET drawn from the surface storage and \( L_{\text{max}} \) is the maximum soil storage (mm). \( L_{\text{max}} \) is a calibration parameter.
Percolation flows to two parallel groundwater reservoirs which are conceptualized as shallow and deep aquifers, respectively (GS, DG, in mm). A balance equation is solved for each of the two groundwater reservoirs:

\[
\frac{dG_S}{dt} = C_{l\text{ow}} \cdot \text{PER} - \text{BF}_{\text{shallow}}
\]

\[
\frac{dG_D}{dt} = (1 - C_{l\text{ow}}) \cdot \text{PER} - \text{BF}_{\text{deep}}
\]

Baseflow from the groundwater reservoirs to the river is calculated using a linear reservoir approach:

\[
\text{BF}_{\text{shallow}} = \frac{1}{C_{KBF}} \cdot \text{GS}_{\text{shallow}}
\]

\[
\text{BF}_{\text{deep}} = \frac{1}{C_{Klow}} \cdot \text{GS}_{\text{deep}}
\]

where \(C_{KBF}\) and \(C_{Klow}\) are the response times of the two linear reservoirs (days). The response times as well as the factor \(C_{l\text{ow}}\), which governs the partition of percolation between the shallow and deep aquifers, are calibration parameters. Rainfall-runoff processes are thus simulated using a very simple approach with five calibration parameters only: \(U_{\text{max}}, L_{\text{max}}, C_{KBF}, C_{Klow}\) and \(C_{l\text{ow}}\).

An automatic calibration module is available for NAM (Madsen, 2000). The module is based on the Shuffled Complex Evolution (SCE) algorithm, and it allows for the optimization of multiple objectives: (1) overall water balance; (2) overall RMSE, (3) peak flow RMSE and (4) low flow RMSE. The catchments were classified into calibration, validation, prediction and inactive catchments, based on the discharge data provided by the Operational Hydrological Forecasting Department (UzHydromet) in Tashkent, Uzbekistan. Catchments with continuous discharge records of 8 years or more were used for calibration, while those with 3–7 years of discontinuous data were used for validation. Prediction catchments are defined as those where no discharge data are available, but topography and land cover suggest considerable runoff. Areas where insignificant runoff is expected based on mean annual rainfall, slope and geology, were
made inactive. In total there were 9 calibration catchments, 9 validation catchments, 14 prediction catchments and 60 inactive catchments. Semi-automatic calibration was carried out for the calibration catchments.

The Mike Basin river network consists of nodes and reaches. The catchments de-water into the river network at catchment nodes. The water transfer from one node to the next is instantaneous, i.e. at every node a simple water balance equation is solved. The only nodes that can store water temporally are the reservoir nodes. Irrigation sites are introduced into the model as water demand nodes. Figure 1 shows the Mike Basin river network layout.

Information on the reservoirs was obtained from the Scientific Information Council of the Interstate Commission for Water Coordination in Central Asia (ICWC, 2009). The reservoirs in the Syr Darya River Basin are implemented as rule curve reservoirs. The reservoir water balance is calculated from inflow, outflow and losses. The level-area-volume curve is used to convert volume to water level. This information was provided by UzHydromet. Once the water level reaches the flood control level, all additional water is instantaneously routed downstream. Historical observed release time series are available from the ICWC (2009). The observed releases were prescribed as minimum downstream release time series for the various reservoirs over the historical simulation period. In real-time application mode of the model, these releases are replaced by planned/projected releases.

The irrigation areas in the Syr Darya River Basin are lumped into 6 major demand sites, following (Raskin et al., 1992). These are High Naryn, Fergana, Mid Syr, Chakir, Artur and Low Syr. Irrigation areas and crop distributions were taken from Raskin et al. (1992). Irrigation water demand was calculated using the FAO-56 methodology (Allen, 2000). Growing season time periods were estimated based on FAO-56. During the growing season, the soil water balance is calculated on a daily time step from precipitation, crop evapotranspiration and irrigation for each demand site. Precipitation is taken from the TMPA product (see below) and crop evapotranspiration is calculated using the FAO dual crop coefficient approach and reference ET. Irrigation abstractions
are calculated using the standard FAO-56 irrigation model. This model assumes that irrigation is triggered if the soil water content decreases below 50% of the readily available water. Soil water contents at field capacity and wilting point were uniformly set to 0.15 and 0.05 respectively. A total loss fraction of 0.3 was generally assumed for all demand sites.

### 3.2 Data assimilation

The Ensemble Kalman Filter (EnKF) has become a popular data assimilation technique in many environmental modeling applications because of its ease of implementation and its computational efficiency (Evensen, 2003). In the EnKF approach, the covariance matrix used in a traditional Kalman Filter is computed from an ensemble of model states. The mean of the ensemble is assumed to be the “truth” and the model error (or covariance) is represented by the covariance of the ensemble members. The ensemble members are then updated according to model and observation errors, as in a traditional Kalman Filter. Let $x_f$ be the $ns \times 1$ vector of forecasted model states. The model error covariance $P_f$ is

$$
P_f = (x_f - \bar{x}_f)(x_f - \bar{x}_f)^T \tag{7}
$$

where the overbar denotes an average over the ensemble. The model states of every member are then updated with the Kalman update equation

$$
x_a^i = x_f^i + K \left( y_i - Hx_f^i \right), \tag{8}
$$

where $x_a^i$ is the vector of updated model states for the $i$-th ensemble member, $H$ is an $no \times ns$ operator ($no$ is the number of observations) that transforms the states into observation space and $y$ is a $no \times 1$ vector that contains the observations for every state variable. The Kalman gain $K$ is defined by

$$
K = PH^T \left( HPH^T + R \right)^{-1} \tag{9}
$$

8358
where $R$ is the $n \times n$ error covariance matrix of the observations. A normally distributed, uncorrelated distribution is assumed for the observation errors.

We use the assimilation algorithm described in Verlaan (2003). The algorithm was adapted to run a set of coupled NAM – Mike Basin models automatically and to assimilate water level measurements for several reservoirs. The model state variables used in data assimilation are the water levels in the various reservoirs. In our implementation, we consider three sources of uncertainty, which dominate overall model uncertainty:

1. Uncertainty of the precipitation product used to force the rainfall-runoff model. The study area is large and has a complex topography. Comparisons between different precipitation products and the few available ground stations have shown significant deviations. Each ensemble member is therefore forced with a different precipitation input. The precipitation time series are generated using the relative error specified for our precipitation product (TRMM-3B42, see below).

2. The calibration parameters of the subcatchment rainfall-runoff models are highly uncertain. The ensemble members were run with individual random realizations of the 5 calibration parameters. All 5 parameters were assumed to be log-normally distributed around the calibration result and the sampling standard deviation was set to 0.6 log10 units for $U_{\text{max}}$ and $L_{\text{max}}$, 0.2 log10 units for CKBF and CKlow and 0.1 log10 units for Cqlow, where Cqlow is expressed in percent.

3. The irrigated areas and crop distributions at the 6 major demand sites are uncertain because the information from Raskin et al. (1992) is fairly old and probably does not match the present configuration of the irrigation districts. The ensemble members were therefore run with irrigation areas which were randomly perturbed by 20%.
3.3 Input and forcing datasets

All input and forcing datasets were obtained from remote sensing and reanalysis datasets. Table 3 provides an overview of the various data sources used in this study.

The digital elevation model (DEM) of the area was obtained from the Shuttle Radar Topography Mission (SRTM). The mission is described by Rabus et al. (2003), and an assessment of its results is provided by Rodriguez et al. (2006). The data with a 3 arc s (90 m) spatial resolution was resampled to 1 km spatial resolution. The 1 km DEM was used to delineate the river network and the subcatchments using automatic GIS routines.

The Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA; Huffman et al., 2007) was used as the data source for precipitation in the Syr Darya basin. The 3B42 research product was found suitable because of its temporal and spatial resolution (3 h and 0.25°, respectively) and the incorporation of surface observation data. The TMPA rainfall estimates have been validated in diverse regions, e.g. USA (Villarini et al., 2007), Argentina (Su et al., 2008) and Brazil (Collischonn et al., 2008). The 3B42 product comes as a 3-hourly product and includes pixel-based uncertainty estimates. The 3B42 product showed significantly lower precipitation amounts than observed at the available ground precipitation stations (Fig. 4a). Moreover, water balance calculations resulted in time-accumulated runoff coefficients as high as 0.9 for some subcatchments, which are clearly unrealistic. For the application of the model in real-time mode, the 3B42 product was compared to the TMPA real time product, 3B42-RT. For the period of comparison (October 2008 to December 2009), precipitation over the subcatchments of the Syr Darya river basin from 3B42-RT was 2.4 times higher than the precipitation calculated from 3B42 (Fig. 4b). In view of the water balance problems discussed above, we decided to adopt the overall precipitation amount from 3B42-RT and scaled the 3B42 product with a factor of 2.4.

Ten-day ground temperature observations were available from UzHydromet at five stations but could not be used to force the model because of the low temporal and
spatial resolution. ECMWF's operational surface analysis dataset, which includes 2-m temperature, was used instead (ECMWF, 2009). The data is available in near real-time and can thus be used in both historical and real-time mode. It has a temporal resolution of 6 h (00:00, 06:00, 12:00 and 18:00 UTC) and a spatial resolution of 0.5° up to 2006, and 0.25° thereafter. The temperature fields were averaged over daily periods, with a local time correction to the median longitude of the Syr Darya basin (70° E), i.e. UTC + 06:00. The pixel-wise daily mean temperature was then area-averaged over the catchments. The mean catchment elevation was used as the reference elevation when extrapolating the temperature to the different elevation zones in the catchment.

Input data for reference ET calculation based on the Penman-Monteith equation were not available. Reference ET was therefore computed from temperature using Hargreaves equation (Allen et al., 1998):

\[
ET_{\text{ref}} = 0.0023 \left( T_{\text{mean}} + 17.8 \right) \left( T_{\text{max}} - T_{\text{min}} \right)^{0.5} \cdot R_a
\]

where \( T_{\text{mean}} \) is defined as the daily average of \( T_{\text{max}} \) and \( T_{\text{min}} \) (not the average of all available temperature measurements) and \( R_a \) is the extraterrestrial radiation (converted to \( 10 \) mm day\(^{-1}\) using the latent heat of vaporization) for the corresponding Julian day and latitude. Hargreaves et al. (2003) present a comprehensive evaluation of the performance of Eq. (10). A temperature averaging period above 5 days is recommended; although some water resource studies (e.g. the IWMI World Climate Atlas) use 10-day temperature averages (Hargreaves et al., 2003). The \( ET_{\text{ref}} \) fields were calculated daily, then averaged over 10-day periods, and the resulting values area-averaged over the different catchments. Figure 5 presents a summary comparison between the various remote sensing forcing products and different in-situ control points.

### 3.4 Radar altimetry

Satellite radar altimetry was initially used in order to study the marine geoid and ocean dynamics (Rapley, 1990). However, over the past two decades different research groups have derived inland water heights from space-based radar altimetry.
(e.g. Cazenave et al., 1997; Berry et al., 2005; Cretaux et al., 2006). In this study, altimetry data re-tracked by the Earth and Planetary Remote Sensing Laboratory (EAPRS) over four large reservoirs was assimilated into the Mike Basin model in order to update the water levels of the reservoirs. The data used is derived from the ERS-2 and ENVISAT satellites, which cover 82° N to 82° S and have repeat cycles of 35 days. The altimetry data re-tracked by the EAPRS lab provide water level time series over a large number of inland water bodies in the Syr Darya River basin. In total, 39 usable ERS-2 targets and 37 usable ENVISAT targets were identified over rivers and lakes in the basin, but only those over the Toktogul, Chardara, Kayrakkum and Charvak Reservoirs (Fig. 1) were assimilated into the model. Frappart et al. (2006) report an accuracy of 0.25–0.53 m for the Radar Altimeter 2 (on board of ENVISAT) over lake targets in the Amazon basin. We found slightly lower precisions for the reservoirs in the Syr Darya, when comparing historical in-situ water levels and radar altimetry (Table 4). However, compared to the typical seasonal variation of the water levels in the reservoirs (between 7 and 50 m, see Table 4), the standard errors of the altimetry time series are small.

4 Results

Our rainfall-runoff modeling approach captures the dominant snowmelt process in the hydrological regime of the subcatchments: precipitation is accumulated during the winter and released throughout the melting season (Fig. 6). However, the calibration-validation process shows that model performance is very variable and that the models tend to underestimate runoff (Table 5). This is not surprising considering the size and complexity of the model domain, the uncertainties associated with remotely-sensed forcing data, and the simplicity of the modeling approach. If the model is run in “no-assimilation” mode and is not informed via assimilation of radar altimetry, its utility for water resources management applications is limited. However, model uncertainty can be significantly reduced through the assimilation of radar altimetry measurements of
reservoir water levels. Figure 7 compares the reservoir levels predicted by the assimilation scheme and by the corresponding no-assimilation model run for the historical simulation period. When altimetry measurements were available, their assimilation improved the results of the model considerably (Table 6), except at Kayrakkum Reservoir, which has a short residence time (Table 4) and discontinued altimetry measurements. The accuracy of the EnKF estimates is expected to improve proportionally to the square root of the ensemble size $n_e$ (Evensen, 2007), although in practical applications this is limited by the number of ensemble members that are computationally feasible to run. Initially, an ensemble size of 50 was chosen. Subsequent trials with larger ensembles resulted in insignificant improvements of the model performance. Figure 8 shows how, on average, model residuals increase over time following the assimilation of a radar altimetry datum for the four reservoirs. Generally, we observe an approximately linear increase of the model error as a function of time after assimilation. For Charvak reservoir, a period of moderate increase up to about day 40 after assimilation is followed by more pronounced increases after day 40. The orbit repeat cycle of Envisat is 35 days and assimilation of radar altimetry can thus keep the model error at a moderate level.

5 Discussion and conclusions

A modeling approach using only remotely-sensed and reanalysis data has been developed and applied to the Syr Darya River Basin. The ability of the river basin model to predict reservoir water levels in “no-assimilation” mode was limited. The generally low model performance can be due to inaccuracies in the RS input data, to the simplifications inherent in model structure (e.g. monthly snowmelt coefficient, lack of a glacial accumulation/ablation model), or to un-modeled dynamics in the hydrology of the basin (e.g. variation of irrigation water demand). Limited availability of in-situ discharge data for model calibration required that high resolution RS data be aggregated over very large areas (sometimes as large as 37 000 km$^2$). If more discharge stations were available, smaller subcatchments could be used and the high spatial resolution of the data...
products would have been better exploited. While these limitations are severe, they are fairly typical for the situation in many large, complex and poorly gauged river basins on the planet.

In such basins, we expect significant increases of model performance when assimilating real-time information based on remote sensing. We showed that assimilation of satellite altimetry measurements of reservoir water levels can reduce deviations between predicted and observed water levels. Even though the mean accuracy of the altimetry data was 0.86 m, it was sufficient to improve the performance of the model. Without such data, reservoir levels would diverge over time from the “true” state of the system. The ensemble Kalman filter implementation used in this study has reasonable computational requirements, with an ensemble size of 50 resulting in a good trade-off between model performance and computation time.

The modeling approach can be exploited to its full potential, if the modeling system is run in real time. A real time (RT) version of the model has been implemented and results are made available on the internet (http://tethys.eaprs.cse.dmu.ac.uk/RiverLake/info/river_modelling). The RT model uses the real time precipitation product (3B42-RT) instead of the research product (3B42). The 3B42-RT product does not incorporate gauge data, but becomes available ca. 6 h after observation (see Huffman et al., 2007; Huffman, 2008, 2009). The temperature data are available 7 h after observation. Updated reservoir levels are computed by the model with a total real-time delay of 2 days.

It is important to recognize that the assimilation of reservoir water levels violates the water balance in the system. In the state updating procedure (Eq. 8) water is simply added or abstracted from the reservoirs. Thus, the model is not suitable for long-term water resources scenario calculations, where mass balance has to be maintained. The application scenario for the tool is medium-term forecasting. The model is able to provide “best estimates” of reservoir levels (and thus water availability) with lead times of a few months. These best estimates are based on a hydrological model and the most recent available water levels. The model can also assimilate real-time water level data from in-situ stations. However, in-situ data typically become available at later times.
Moreover, remotely sensed water levels have the advantage of being accessible to all interested stakeholders and countries and can thus be considered as entirely impartial information.

The data assimilation approach described here could benefit the annual water allocation process in the Syr Darya basin by providing an efficient and transparent technology for updating hydrological forecasts in real time. The real-time capability could be used together with hydro-economic models of the basin (Cai et al., 2002) in real-time, adaptive water resources management. In the current management setup, the riparian states are supposed to agree on an allocation plan at the beginning of April; however, agreement is often delayed until well into the growing season, creating significant uncertainties for irrigation planning and increasing tensions between the riparian states. An obstacle to co-operation is the deteriorated state of the hydro-meteorological monitoring network that existed in Soviet times (Schar et al., 2004). In addition, national hydro-meteorological agencies now responsible for data collection are reluctant to share data and the annual process of data collection and forecasting can be delayed by poor communications infrastructure (Biddison, 2002). Because of all these issues, the real-time forecasting tool developed in this study has the potential to contribute to more efficient water resources management in the region.

In summary, radar altimetry data over inland water bodies is an innovative data source for hydrological applications. It can be used to update hydrological models in real time and can significantly enhance science-based decision support to water resources managers, particularly in poorly gauged river basins.

Acknowledgements. The TMPA data were provided by the NASA/Goddard Space Flight Center’s Laboratory for Atmospheres and PPS, which develop and compute the TMPA as a contribution to TRMM. The ECMWF temperature data was obtained through the Danish Meteorological Institute; the authors would like to thank Thomas Lorenzen for his support in providing this data. We thank the International Research School of Water Resources (fiva) for supporting T. Siegfried’s stay at the Technical University of Denmark.
This work is part of the River & Lake Project (Performance Improvement and Data Assimilation into Catchment Models for the Exploitation of River and Lake Products from Altimetry, ESRIN Contract No. 21092/07/I-LG), work package 700.

References


Francois, C., Quesney, A., and Ottlé, C.: Sequential assimilation of ERS-1 SAR data into a coupled land surface-hydrological model using an extended Kalman filter, J. Hydrometeorol.,


Table 1. NAM parameters chosen to enforce the rainfall-runoff model structure shown in Fig. 3.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Units</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CQOF</td>
<td>Overland flow runoff coefficient</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>CKIF</td>
<td>Time constant for interflow</td>
<td>hours</td>
<td>1e6</td>
</tr>
<tr>
<td>CK1,2</td>
<td>Time constants for routing overland flow</td>
<td>hours</td>
<td>10</td>
</tr>
<tr>
<td>TOF</td>
<td>Rootzone threshold value for overland flow</td>
<td>–</td>
<td>0.999</td>
</tr>
<tr>
<td>TIF</td>
<td>Rootzone threshold value for inter flow</td>
<td>–</td>
<td>0.999</td>
</tr>
<tr>
<td>TG</td>
<td>Rootzone threshold value for groundwater recharge</td>
<td>–</td>
<td>0</td>
</tr>
</tbody>
</table>

For a detailed description of these parameters and their interaction the reader is referred to DHI (2000).
Table 2. Snowmelt parameters used in the rainfall-runoff model (NAM).

<table>
<thead>
<tr>
<th>Lapse rate [deg 100 m⁻¹]</th>
<th>Degree day coefficient (M) [mm deg⁻¹ day⁻¹]</th>
</tr>
</thead>
<tbody>
<tr>
<td>wet</td>
<td>J</td>
</tr>
<tr>
<td>−0.7</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Table 3. Source and spatio-temporal resolution of the various datasets used in the model.

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Data Source</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Space</td>
</tr>
<tr>
<td>Remotely sensed and reanalysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precipitation</td>
<td>TMPA 3B42-RT</td>
<td>0.25°</td>
</tr>
<tr>
<td>Temperature</td>
<td>ECMWF</td>
<td>0.5°–0.25°</td>
</tr>
<tr>
<td>PET</td>
<td>Func. of Temp</td>
<td>0.5°–0.25°</td>
</tr>
<tr>
<td>Lake altimetry</td>
<td>ERS/ENVISAT</td>
<td>76 targets</td>
</tr>
<tr>
<td>DEM</td>
<td>SRTM</td>
<td>1000 × 1000 m</td>
</tr>
<tr>
<td>Observations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discharge</td>
<td>UzHydromet</td>
<td>18 stations</td>
</tr>
<tr>
<td>Reservoir release</td>
<td>ICWC</td>
<td>5 reservoirs</td>
</tr>
<tr>
<td>Comparison data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precipitation</td>
<td>UzHydromet</td>
<td>16 stations</td>
</tr>
<tr>
<td>Temperature</td>
<td>UzHydromet</td>
<td>5 stations</td>
</tr>
<tr>
<td>Reservoir levels</td>
<td>UzHydromet</td>
<td>4 reservoirs</td>
</tr>
</tbody>
</table>
Table 4. Altimetry targets over each reservoir. The altimetry error is reported here as the RMSE between the time series after the mean of each has been removed.

<table>
<thead>
<tr>
<th>Reservoir</th>
<th>Satellite</th>
<th>RMSE [m]</th>
<th>Level range [days]</th>
<th>Volume range [10^6 m^3]</th>
<th>Residence time [days]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chardara</td>
<td>ERS</td>
<td>0.63, 0.85</td>
<td>11.5</td>
<td>4700</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>ENVISAT</td>
<td>0.37, 0.48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toktogul</td>
<td>ERS</td>
<td>0.68</td>
<td>57.5</td>
<td>14 000</td>
<td>377.95</td>
</tr>
<tr>
<td></td>
<td>ENVISAT</td>
<td>0.89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charvak</td>
<td>ERS</td>
<td>1.81</td>
<td>55.4</td>
<td>1580</td>
<td>83.7</td>
</tr>
<tr>
<td></td>
<td>ENVISAT</td>
<td>1.42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kayrakkum</td>
<td>ERS</td>
<td>0.61</td>
<td>7.7</td>
<td>2600</td>
<td>46.1</td>
</tr>
</tbody>
</table>

Table 5. Results of the rainfall-runoff model calibration.

<table>
<thead>
<tr>
<th>Catchment ID</th>
<th>Station ID</th>
<th>Umax [mm]</th>
<th>Lmax [mm]</th>
<th>CKBF [h]</th>
<th>CKlow [h]</th>
<th>Cqlow [%]</th>
<th>$R^2$</th>
<th>WBE(^a) [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calibration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>16 055</td>
<td>10</td>
<td>100</td>
<td>480</td>
<td>8760</td>
<td>20</td>
<td>0.52</td>
<td>9</td>
</tr>
<tr>
<td>47</td>
<td>16 169</td>
<td>40</td>
<td>400</td>
<td>960</td>
<td>8760</td>
<td>30</td>
<td>0.62</td>
<td>3</td>
</tr>
<tr>
<td>55</td>
<td>16 176</td>
<td>40</td>
<td>400</td>
<td>720</td>
<td>8760</td>
<td>20</td>
<td>0.67</td>
<td>2</td>
</tr>
<tr>
<td>61</td>
<td>16 193</td>
<td>40</td>
<td>400</td>
<td>480</td>
<td>8760</td>
<td>10</td>
<td>0.69</td>
<td>97</td>
</tr>
<tr>
<td>145</td>
<td>16 198</td>
<td>10</td>
<td>100</td>
<td>960</td>
<td>8760</td>
<td>20</td>
<td>0.28</td>
<td>-25</td>
</tr>
<tr>
<td>62</td>
<td>16 202</td>
<td>40</td>
<td>400</td>
<td>480</td>
<td>8760</td>
<td>10</td>
<td>0.59</td>
<td>61</td>
</tr>
<tr>
<td>64</td>
<td>16 230</td>
<td>40</td>
<td>400</td>
<td>480</td>
<td>8760</td>
<td>10</td>
<td>0.74</td>
<td>34</td>
</tr>
<tr>
<td>148</td>
<td>16 279</td>
<td>10</td>
<td>100</td>
<td>960</td>
<td>8760</td>
<td>30</td>
<td>0.80</td>
<td>-13</td>
</tr>
<tr>
<td>147</td>
<td>16 290</td>
<td>10</td>
<td>100</td>
<td>960</td>
<td>8760</td>
<td>30</td>
<td>0.79</td>
<td>-12</td>
</tr>
<tr>
<td>Validation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>16 059</td>
<td>10</td>
<td>100</td>
<td>720</td>
<td>8760</td>
<td>20</td>
<td>0.45</td>
<td>27</td>
</tr>
<tr>
<td>12</td>
<td>16 121</td>
<td>10</td>
<td>100</td>
<td>720</td>
<td>8760</td>
<td>20</td>
<td>0.57</td>
<td>5</td>
</tr>
<tr>
<td>14</td>
<td>16 127</td>
<td>10</td>
<td>100</td>
<td>720</td>
<td>8760</td>
<td>20</td>
<td>0.51</td>
<td>-43</td>
</tr>
<tr>
<td>15</td>
<td>16 134</td>
<td>10</td>
<td>100</td>
<td>720</td>
<td>8760</td>
<td>20</td>
<td>0.43</td>
<td>-3</td>
</tr>
<tr>
<td>16</td>
<td>16 135</td>
<td>10</td>
<td>100</td>
<td>720</td>
<td>8760</td>
<td>20</td>
<td>0.42</td>
<td>-37</td>
</tr>
<tr>
<td>18</td>
<td>16 136</td>
<td>10</td>
<td>100</td>
<td>720</td>
<td>8760</td>
<td>20</td>
<td>0.66</td>
<td>-28</td>
</tr>
<tr>
<td>9</td>
<td>16 146</td>
<td>10</td>
<td>100</td>
<td>720</td>
<td>8760</td>
<td>20</td>
<td>0.58</td>
<td>-36</td>
</tr>
<tr>
<td>75</td>
<td>16 205</td>
<td>10</td>
<td>100</td>
<td>720</td>
<td>8760</td>
<td>20</td>
<td>0.44</td>
<td>55</td>
</tr>
<tr>
<td>49</td>
<td>16 510</td>
<td>10</td>
<td>100</td>
<td>720</td>
<td>8760</td>
<td>20</td>
<td>0</td>
<td>9</td>
</tr>
</tbody>
</table>

\(^a\) Water balance error.
### Table 6. Reservoir water level residuals with and without assimilation of radar altimetry data.

<table>
<thead>
<tr>
<th></th>
<th>Toktogul</th>
<th>Chardara</th>
<th>Kayrakkum</th>
<th>Charvak</th>
<th>Mean$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DA</td>
<td>mean(res)</td>
<td>−3.07</td>
<td>−0.46</td>
<td>3.08</td>
<td>−0.87</td>
</tr>
<tr>
<td></td>
<td>std(res)</td>
<td>5.33</td>
<td>1.15</td>
<td>5.26</td>
<td>15.11</td>
</tr>
<tr>
<td>No</td>
<td>mean(res)</td>
<td>−9.57</td>
<td>−3.18</td>
<td>3.62</td>
<td>−2.35</td>
</tr>
<tr>
<td>DA</td>
<td>std(res)</td>
<td>8.49</td>
<td>3.38</td>
<td>5.85</td>
<td>23.59</td>
</tr>
</tbody>
</table>

$^a$ Mean of absolute residuals across all reservoirs.
Fig. 1. Base map of the Syr Darya River Basin (SDRB).
Fig. 2. Block diagram of the modelling and assimilation approach, showing input data (left side), processing (center) and output data (right side). The input data in dark shading has been perturbed to create the ensembles. The processes on slanted shading only occur when there is altimetry data available.
Fig. 3. Structure of the rainfall-runoff model. Items on light shading are input data, those on dark shading relate to the calibration parameters.
Fig. 4. Comparison of precipitation datasets. (A) Mean observed annual precipitation at 16 stations and corresponding pixel from 3B42 research product. (B) Mean catchment precipitation from real time and research 3B42 products.
Fig. 5. 3B42 average monthly precipitation and ECMWF 10-day average temperatures against observed (UzHydromet) data at three locations (see Fig. 1 for locations).
Fig. 6. Rainfall-runoff modelling results in catchment 69.
Fig. 7. Reservoir level simulation for the historical period (in mamsl): with assimilation of radar altimetry (black), without assimilation of radar altimetry (red), and observed in-situ water levels (blue) for an ensemble size of 50. The shaded area indicates the 2.5 and 97.5% quantile range of the ensemble, and the crosses are altimetry measurements. For interpretation of the color reference in this figure, the reader is referred to the electronic version of this article.

8384
Fig. 8. Average absolute residuals of reservoir level after assimilation of altimetry data.