

Manuscript prepared for Hydrol. Earth Syst. Sci. Discuss.
with version 2.2 of the L^AT_EX class copernicus_discussions.cls.
Date: 12 March 2014

The suitability of remotely sensed soil moisture for improving operational flood forecasting

**N. Wanders¹, D. Karssenber¹, A. de Roo^{1,2}, S. M. de Jong¹, and
M. F. P. Bierkens^{1,3}**

¹Department of Physical Geography, Utrecht University, The Netherlands

²European Commission, Joint Research Centre, Ispra, Italy

³Deltares, Utrecht, The Netherlands

Correspondence to: N. Wanders
(n.wanders@uu.nl)

Abstract

We evaluate the added value of assimilated remotely sensed soil moisture for the European Flood Awareness System (EFAS) and its potential to improve the prediction of the timing and height of the flood peak and low flows. EFAS is an operational flood forecasting system for Europe and uses a distributed hydrological model (LISFLOOD) for flood predictions with lead times up to 10 days. For this study, satellite-derived soil moisture from ASCAT, AMSR-E and SMOS is assimilated into the LISFLOOD model for the Upper Danube basin and results are compared to assimilation of discharge observations only. To assimilate soil moisture and discharge data into the hydrological model, an Ensemble Kalman Filter (EnKF) is used. Information on the spatial (cross-) correlation of the errors in the satellite products, is included to ensure optimal performance of the EnKF. For the validation, additional discharge observations not used in the EnKF, are used as an independent validation dataset.

Our results show that the accuracy of flood forecasts is increased when more discharge observations are assimilated; the Mean Absolute Error (MAE) of the ensemble mean is reduced by 65%. The additional inclusion of satellite data results in a further increase of the performance: forecasts of base flows are better and the uncertainty in the overall discharge is reduced, shown by a 10% reduction in the MAE. In addition, floods are predicted with a higher accuracy and the Continuous Ranked Probability Score (CRPS) shows a performance increase of 5-10% on average, compared to assimilation of discharge only. When soil moisture data is used, the timing errors in the flood predictions are decreased especially for shorter lead times and imminent floods can be forecasted with more skill. The number of false flood alerts is reduced when more observational data is assimilated into the system. The added values of the satellite data is largest when these observations are assimilated in combination with distributed discharge observations. These results show the potential of remotely sensed soil moisture observations to improve near-real time flood forecasting in large catchments.

1 Introduction

Floods are extreme hydrological events caused by excessive water availability and may cause large economical, societal and natural damage. One example is the summer 2013 flood in central Europe producing historical high water levels in large parts of the Danube and Elbe catchments, causing a total estimated economic loss of 23 billion Euro (Aon Benfield, 2013). Due to their increasing impact on society, forecasting of these extreme events has become more important to increase preparedness and improve the response to and prevention of floods. This requires an increasing need to develop accurate and reliable flood forecasting systems. National forecasting systems have been developed in for example England (National Flood Forecasting System), Germany (Hochwasservorhersagezentral), Netherlands, Germany and Swiss (FEWS-Rhine & Meuse), Czech Republic (CHMI-IWSS), Sweden (SMHI) and most other countries in Europe. For transboundary river basins, national forecasting systems are often lacking skill and transboundary forecasting systems are preferred. To fulfil this need, the European Commission developed the European Flood Awareness System (EFAS) for flood forecasting up to a lead times of 10 days for the European continent (Thielen et al., 2009). Additionally, EFAS will contribute to understanding of flood events on a transboundary scale and will support international crisis management at the European level.

Flood forecasts are made for multiple basins, using distributed hydrological modelling. Systems like EFAS are highly dependent on the meteorological forcing provided as well as the pre-storm initial conditions of the catchment (Nester et al., 2012; Alfieri et al., 2013). To improve estimates of initial conditions data assimilation techniques have the potential to update incorrect model states with observational data to obtain the best possible estimate of the current status of the hydrological system. Discharge data is often used in these data assimilation frameworks, because it contains the integrated information of all other hydrological states (e.g. Vrugt et al., 2006; Clark et al., 2008; Rakovec et al., 2012). However, it is difficult to obtain these measurements in real-time in a way they can be used in EFAS. Observations might not be available in real-time, quality control can not be done in real-time or local data providers are unfortunately not willing to share the information. Measurements of hydrological states other than discharge

are rarely used for estimating the model's initial state while these may be of considerable value. In particular, measurements of the pre-storm soil moisture conditions could potentially improve flood forecasting systems, since initial soil moisture conditions are expected to have a large impact on the flood peaks during a storm event. The soil moisture content determines the amount of water which can still be stored in the unsaturated zone or percolate to the saturated zone and thereby influences the precipitation required to generate overland flow. However, field observations at continental scale are not available due to the limited number of observational networks and their low spatial support. Remotely sensed soil moisture retrievals from the microwave domain could potentially fill the need for soil moisture observations at the large spatial scales. Observations are globally available and revisit times per sensor are between 1 and 3 days depending on latitude. An additional advantage is that the data is available within 3 hours after observed and the satellites have a global coverage, while single discharge observations are only valid for the catchment scale.

Multiple studies have used remotely sensed soil moisture to improve discharge simulations in small catchments ($\leq 1000 \text{ km}^2$) and to correct for errors in pre-storm soil moisture conditions (Pauwels et al., 2001; Scipal et al., 2008; Brocca et al., 2010; Chen et al., 2011; Brocca et al., 2012; Matgen et al., 2012). These studies show that assimilation of these data improved the simulation of flood events and especially the height of the flood peak. For large-scale catchments, Draper et al. (2011) assimilated remotely sensed soil moisture from ASCAT over France to improve discharge simulations. It was concluded that the assimilation of soil moisture mainly corrected for biases in precipitation or incorrect model climatology. However, the potential to improve flood forecasts was not studied at the large-scale. The previously mentioned studies mainly focussed on the potential gain for flood forecasting, when only observations from a single sensor are assimilated. This potential can be increased by making use of soil moisture retrieved by multiple sensors, thereby increasing the quality and quantity of the observations. The value of combined assimilation of data from multiple sensors for operational flood forecasting at large-scale remains however unknown. Moreover, it is equally important to take into account that assimilation of remotely sensed soil moisture can lead to significant difference in the parametrization of the hydrological model (e.g. Santanello et al., 2007; Sutanudjaja et al., 2013;

Wanders et al., 2013) and this will also impact the potential gain from the assimilation of observations of other hydrological variables. Additionally the added value of the remotely sensed soil moisture compared to the assimilation of discharge observations has not been studied so far. Therefore, more research is required, especially in large-scale catchments using conjunctively

5 multi-sensor remotely sensed soil moisture observations and discharge data. The aim of this study is to determine the benefits of the assimilation of multi-sensor soil moisture observations in operational flood forecasting systems in large scale catchments. To achieve this aim, this research focuses on three main research questions: (i) Does the assimilation of remotely sensed soil moisture lead to increased forecasting skills in terms of forecast uncertainty

10 and forecast bias compared to assimilation of discharge observations? (ii) Does the assimilation of remotely sensed soil moisture increase the lead times at which floods can be accurately predicted? (iii) Is it possible to reduce the number of false flood alerts with the use of remotely sensed soil moisture? These research questions are answered using the EFAS model

15 setup, which enables a proper validation of the results in the context of a real operational system. Results of assimilating remotely sensed soil moisture are compared with assimilation of discharge data only. Also, the impact of the number of discharge observations is investigated, to properly validate in what situations the assimilation of remotely sensed soil moisture could improve flood forecasting. As a test-basin the Upper Danube catchment is selected which is one of the largest catchments in Europe containing a large number of locations with time series of

20 discharge. Satellite data from three microwave sensors (ASCAT, AMSR-E and SMOS) is used in the assimilation framework to increase the number of observations and the potential benefits of these observations for the flood prediction.

2 Material and Methods

2.1 Study area

25 The study area is the Upper Danube catchment upstream of Bratislava (catchment size $135 \cdot 10^3 \text{ km}^2$, Figure 1). The border of the Upper Danube is formed by the Alps in the South and the

catchment contains the northern part of Austria, the southern part of Germany, the South-eastern part of the Czech Republic and western Slovakia. Elevations range from 150 - 3150 *m* above sea level. In the catchment, daily discharge observations for 23 locations are available through the Global Runoff Data Centre (GRDC) which enable validation and assimilation (Figure 1).

5 With a split-sample approach discharge observations used for assimilation will not be used for validation to assure an independent validation of the improvements in the flood forecasting after the assimilation.

2.2 European Flood Awareness System

The European Flood Awareness System was developed in 2003 by the European Commission at the Joint Research Centre in Ispra and is being improved since¹. In 2012 EFAS became an operational service aiming to provide flood forecasts up to 10 days in advance over the European continent. At the core of the EFAS system is the hydrological model LISFLOOD which was originally developed by De Roo et al. (2000), later improved by Van Der Knijff et al. (2010) and running in the PCRaster modelling environment (Wesseling et al., 1996; Karszenberg et al., 15 2010). LISFLOOD was specifically developed for discharge simulations of large scale river basins. The model consists of a vegetation layer, four layers to simulate the unsaturated zone, two linear reservoirs to represent fast and slow responding groundwater systems and a channel network for discharge routing (Figure 2).

In this study, the original two layer representation of the unsaturated zone (De Roo et al., 2000; 20 Van Der Knijff et al., 2010) was replaced by a new unsaturated zone model component that uses four layers. This enables a more detailed representation of the soil moisture in the topsoil and results in modelled soil moisture that is directly comparable to the soil moisture observations retrieved from remotely sensed soil moisture. The layers have been added in the topsoil and possess a depth equal to the typical penetration depth of microwave sensors. The new model 25 set-up consists of unsaturated zone layers of 2 and 3 *cm* thick, respectively, the third layer represents the remaining part of the rooting depth (the topsoil, Figure 2). The root zone is sim-

¹www.efas.eu

ulated using the topsoil and evapotranspiration occurs from these layers. The evaporation for a particular layer is limited if soil moisture is below critical soil moisture conditions, in which case more water is extracted from the other soil moisture layers to compensate for the reduced evaporation. The abstraction per layer is linearly related to the total storage capacity of the layer. Thick layers will thus have a larger contribution to the evapotranspiration compared to thinner layers. When the entire root zone is below critical soil moisture conditions the evaporation is limited for the entire topsoil and actual evapotranspiration will be lower than potential evapotranspiration. Bare soil evapotranspiration occurs only from the first layer of 2 cm. Via capillary rise replenishment of the root zone can occur from the fourth unsaturated zone layer (the subsoil). The amount of capillary rise depends on the difference in hydraulic head between two layers and the average conductivity of the layers. The first layer will also largely impact the amount of surface runoff in the LISFLOOD model. The soil wetness of the first layer determines the infiltration capacity of the unsaturated zone and when the infiltration capacity is exceeded by rainfall or snowmelt this will generate overland flow. Sub-daily time steps are included to enable a stable performance of the soil moisture simulation, where the number of sub-daily time steps is dependent on the amount of potential infiltration and water storage in the unsaturated zone.

In order to use the best calibrated model for the study area, the hydrological model LISFLOOD was calibrated for the Upper Danube. For the calibration, soil moisture and discharge observations were used to calibrate the most sensitive model parameters. The parameters which were calibrated were related to the snow accumulation, infiltration and percolation through the unsaturated zone, the groundwater system and routing of discharge (Figure 2). A dual state and parameter Ensemble Kalman filter was used to calibrate LISFLOOD for the Upper Danube. A total of 300 members was used to estimate all parameters of the model over the period 2010-2011. The period was selected because satellite data from multiple sensors is available for this period. This resulted in calibrated parameters with distributions defined by 300 realizations of parameter sets, which could be used for hydrological simulations. The use of these parameter distributions allows to account for the uncertainty in the initial conditions and for different hydrological response to identical meteorological input. More detailed information on the prob-

abilistic model calibration set-up can be found in Wanders et al. (2013).

The meteorological forcing of EFAS consists of daily precipitation, daily potential evapotranspiration and the average daily temperature. EFAS uses meteorological forcing from the 51 members of the European Centre for Medium-Range Weather Forecasting Ensemble Prediction System (ECMWF-EPS). This results in 51 hydrological forecasts for every 12 hours at midday and midnight. The new set-up of EFAS which uses 300 realizations of parameter sets, differs from the original EFAS set-up which uses one parameter set. Additionally, the new set-up also uses a set of initial hydrological conditions which are forced with identical meteorological forcing. The original EFAS set-up only uses one parameter set and one initial hydrological condition for all meteorological forecasts. The EFAS setup used here will allow accounting for the uncertainty in the initial conditions which can be an important factor in flood forecasting. Throughout the manuscript the term EFAS will be used when talking about the entire forecasting system, i.e. the combination of meteorological forcing, hydrological model and resulting flood forecasts. The term LISFLOOD will be used when the focus is specifically on the data assimilation or the hydrological model.

2.3 Data

2.3.1 Satellite data

Remotely sensed soil moisture data from three satellites is used, namely SMOS, ASCAT and AMSR-E. SMOS is the first dedicated soil moisture satellite using fully polarized passive microwave signals at 1.41 GHz (L-band) observed at multiple angles (Kerr et al., 2012). The observation depth of SMOS is 5 *cm* with a spatial resolution of 35 – 50 *km* depending on the incident angle and the deviation from the satellite ground track. The revisit time of SMOS is within 1 – 3 days depending on the latitude. SMOS retrievals which are potentially contaminated with Radio Frequency Interference (RFI) have been removed. The observations from SMOS can be directly compared to the weighted average soil moisture content of the two top layers of LISFLOOD, together 5 *cm* thick.

AMSR-E is a multi-frequency passive microwave radiometer (6.9 GHz, C-band) and is a

widely used sensor for soil moisture retrievals. The spatial resolution of AMSR-E is between 36 and 54 *km* with an observation depth of 2 *cm* and a revisit time of 1 – 3 days. Several algorithms estimating surface soil moisture from AMSR-E observations exist (e.g. Njoku et al., 2003; Owe et al., 2008). One of the algorithms using exclusively satellite observations is the Land Parameter Retrieval Model (LPRM) which was used for this study. LPRM soil moisture products have been validated against in situ observations (e.g. Wagner et al., 2007; De Jeu et al., 2008; Draper et al., 2009), models (e.g. Loew et al., 2009; Crow et al., 2010; Bisselink et al., 2011) and other satellite products (e.g. Wagner et al., 2007; Dorigo et al., 2010). Observations from AMSR-E are compared to the first layer unsaturated zone layer of LISFLOOD.

Unlike SMOS and AMSR-E, ASCAT uses active microwave at a frequency of 5.3 GHz (C-band) to determine the soil moisture content (Wagner et al., 1999; Naeimi et al., 2009). ASCAT uses a change detection method (Naeimi et al., 2009) and data is provided relative to the soil moisture content of the wettest (field capacity) and driest (wilting point) soil moisture conditions measured (Wagner et al., 1999). The spatial resolution of ASCAT is around 25 *km*, the observation depth is 2 *cm* and the temporal resolution equals a revisit time of 1 – 3 days. As for AMSR-E, ASCAT observations are compared to the top layer of the unsaturated zone of the model simulations.

All satellite soil moisture products are used on an equal area Discrete Global Grid product (DGG). For the SMOS and ASCAT soil moisture product a DGG is available (Bartalis et al., 2006), while for the AMSR-E product a DGG is not available. Therefore, the AMSR-E data was projected on the DGG of SMOS using the nearest neighbour approach, because both satellites have roughly the same spatial resolution. The DGG of ASCAT uses equally spaced areas of 12.5 *km* while the other DGG uses a slightly lower resolution of 15 *km* between points.

Although the passive microwave satellite missions, SMOS and AMSR-E, give absolute soil moisture values in m^3m^{-3} , all satellite data was converted using a rescaling approach. The converted satellite values $\theta_{s,new}$ in m^3m^{-3} used for calibration are calculated by:

$$\theta_{s,new} = \frac{\theta_s - \theta_{s,5}}{\theta_{s,95} - \theta_{s,5}} (\theta_{FC} - \theta_{WP}) + \theta_{WP} \quad (1)$$

where θ_s are the observed satellite soil moisture values (-) at a DGG location, $\theta_{s,95}$ and $\theta_{s,5}$ are the 95th and 5th percentiles of satellite soil moisture values at the DGG location respectively (-), θ_{FC} and θ_{WP} are field capacity and wilting point of the modelled soil moisture values (m^3m^{-3}) at the DGG location. The average model values, θ_{FC} and θ_{WP} , are dependent on the soil texture and are averaged over the support unit of the satellite retrieval.

Frozen soils, snow accumulation and RFI hamper the soil moisture retrieval due to changes in the dielectric constant when water freezes. Therefore, retrievals done with 1) an air temperature below $4^\circ C$, 2) simulated snow accumulation and 3) the presence of RFI were not used in the calibration.

2.3.2 Discharge data

The Upper Danube catchment contains 23 locations where daily discharge observations are available (Figure 1). Time series of discharge are available from January 2000 until December 2011. Using a split sample approach the discharge of 7 stations was used for data assimilation into the forecasting system, while the other 16 stations were only used for validations of the forecasts. This approach is similar to the experimental set-up of Lee et al. (2012) and Rakovec et al. (2012), who used multiple interior discharge stations for validation and assimilation. Assimilation and validation stations are selected such that they are equally distributed over the catchment and are situated both in small rivers and the main Upper Danube river. This will allow to evaluate the impact of the data assimilation at different catchment sizes within the Upper Danube catchment.

2.4 Data assimilation

The Ensemble Kalman Filter (EnKF) is a Monte Carlo based approach which is highly suitable for data assimilation in high dimensional systems (Evensen, 1994; Burgers et al., 1998; Evensen, 2003, 2009), such as the LISFLOOD model. The EnKF is applied to update state variables of the hydrological model. The forward model is given by:

$$\Psi(t+1) = f(\Psi(t), F(t), p) \quad (2)$$

where f is the set of model equations, i.e. the model structure, representing the hydrological processes that lead to change in the system state over time, $\Psi(t)$ is the state of the model at time t , $F(t)$ the model forcing at time t (e.g. precipitation and evaporation) and p are the model parameters. The EnKF is applied on each daily time step using observations from remote sensing (when available, AMSR-E, SMOS and ASCAT) and discharge observations. If no observations of any kind are available no update will be performed. When only a limited number of observations are available these will be used to update the model. The general form of the EnKF (Evensen, 2003) is given by:

$$\Psi^a = \Psi^f + P^f H^T (H P^f H^T + R)^{-1} (Y - H \Psi^f) \quad (3)$$

where Ψ^a is the analysis of Ψ^f , the model forecast, P^f the error covariance matrix of the model, R is the measurement error covariance, H is the measurement operator which relates the model states Ψ to the satellite or discharge observations Y . The observations Y can be described by:

$$Y = H \Psi^t + \epsilon \quad (4)$$

where the true model state (Ψ^t) is transformed to the Y , using the measurement operator (H) and random noise ϵ with a zero mean and an error given by the measurement error covariance (R). The state error covariance matrix of the model prediction is directly calculated from the spread between the different ensemble members using:

$$P^f = \overline{(\Psi^f - \Psi^t)(\Psi^f - \Psi^t)^T} \quad (5)$$

where Ψ is the model state vector and the superscripts f and t represent the forecast and true state, respectively. Since the true state is not known it is assumed that:

$$P^f \approx P_e^f = \overline{(\Psi^f - \bar{\Psi}^f)(\Psi^f - \bar{\Psi}^f)^T} \quad (6)$$

where $\bar{\Psi}^f$ represents the ensemble average and it is assumed that the ensemble of model simulations is sufficient to represent the true state. The EnKF is implemented in the PCRaster modelling environment (Karssenberget al., 2010).

For the assimilation of the satellite data with the Ensemble Kalman Filter (EnKF), spatial

information on the measurement error covariance (R , Equation 3 and 4) is required. The structure of R is determined from estimates of Wanders et al. (2012) over Spain, obtained by using high resolution modelling of the unsaturated zone. From this study the local errors of each satellite product were determined as well as the spatial correlation of the errors of the satellites and the correlation between the errors of different sensors. This information can be used to simultaneously assimilate soil moisture observations from different sensors with the additional information about the error structure obtained from the observation error-covariance matrix. To avoid errors produced by downscaling of the satellite soil moisture, the average modelled soil moisture values are upscaled to the satellite resolution. Each individual satellite observation is then compared to the corresponding model average soil moisture at the same spatial support. The spatial support will differ for different sensors and hence it is important to correctly compare modelled soil moisture to observed soil moisture.

All observations are assimilated as daily averages, since this is the same temporal resolution as the meteorological forcing. The error covariance between the discharge observations is set to zero while the standard error for the discharge observations is assumed to be 30% of the discharge (e.g. Di Baldassarre and Montanari, 2009). It is assumed that the covariance between the satellite soil moisture observations and discharge observations equals zero.

2.5 Assimilation and ensemble hindcasting

In this study, observed satellite and discharge data for December 2010 - November 2011 are used in a hindcasting experiment for the Upper Danube. Only one year was selected to test the procedure since all satellite products are available for this time period with sufficient data quality. After the selected time period the AMSR-E sensor was shut down and before the selected period the quality of the SMOS observations was still below the potential maximum quality due to RFI contamination.

A data assimilation procedure was used to create a reanalysis time series of all state variables which are used as starting point for the hindcast (t_0). Model states are update with the observations and used to have a better estimate of initial condition at t_0 . Figure 3 provides a flowchart

that shows the full hindcasting procedure described below. The 300 parameter realizations from the probabilistic calibration were used to generate the reanalysis time series. As meteorological forcing for the analysis, observed time series of daily precipitation, daily potential evapotranspiration and the average daily temperature were used. Observations are interpolated between meteorological stations with an inverse distance interpolation. For every time step up till t_0 , observed state variables, remotely sensed soil moisture and/or discharge (depending on the scenario), are assimilated into the model. Assimilation is done on a daily time step, since information on the exact time of the discharge observations is largely unknown. Additionally, the model uses meteorological input with a temporal resolution of one day. Parameters are not updated in the assimilation. Thus, the same set of 300 parameter sets is used to generate the 300 ensemble members between analysis steps with the EnKF.

At t_0 , the start of the hindcast, the forward model (Equation 2) is used for the hindcasting of discharge and other state variables. After t_0 , the daily forcing from the ECMWF-EPS is used to drive the model simulation. The hindcast is evaluated based on the observed discharge for the hindcasting period. Like in EFAS, hindcasts are done at midday and midnight based on the latest simulations of the ECMWF-EPS leading to a total of 730 hindcasts. In the original forecasts from EFAS only one set of initial conditions is used, thereby neglecting the uncertainty in the initial conditions. In this experiment, 300 possible realizations of the initial conditions are available from the reanalysis. For each hindcast the 51 members of the ECMWF-EPS are used twice with random realizations from the 300 members of the reanalysis to create $n = 102$ realizations per hindcast. In this approach different meteorological forcing and initial conditions are used for each hindcast to have a better estimate of the forecast uncertainty. A four month simulation was performed using all 300 members in combination with all 51 meteorological forecasts. An analysis of the probability density functions of each hindcast showed that a total of 102 realizations showed no significant differences with a simulation consisting using all possible ($51 \times 300 = 15300$) realizations (for lead times up to 10 days). The significance was tested with a non-parametric Kolmogorov-Smirnoff test, which showed that distributions created with 102 and 15300 are identical ($p = 0.05$). In another set of runs, it was shown that using fixed initial conditions for the hydrological state leads to significantly different distributions.

The same holds for fixing the meteorological forecast for all 300 ensemble members which results in a significantly different probability density function compared to the run created with 15300 realizations. With this exercise we concluded that both the uncertainty in initial states as well as the forcing uncertainty need to be taken into account, but that it suffices to use a subset of the possible realizations to model this joint uncertainty. Hence, to reduce calculation times 102 realizations per hindcast were used in all scenarios. Calculation times for this new assimilation system are low. For a 10-day forecast with 102 members for the Upper Danube the required calculation time is 120 seconds on a 8-core machine with 2.26 GHz processors and 24 GB RAM.

2.6 Scenarios

The different scenarios used are given in Table 1 as well as the data used in the assimilation before the hindcasting was done. The parametrization was calibrated for the Upper Danube for the period 2010-2011 and was used to create analysis time series for each scenario. The calibration was based on the observations available for the reanalysis, so if both discharge and satellite data were available these were also used for the calibration of the hydrological model (Table 1).

Two additional scenarios have been included (bottom half of Table 1) to show the performance of the hindcasts in case of limited or no data availability. Both scenarios have been calibrated on discharge and use assimilation of satellite observations or no data.

2.7 Evaluation

The evaluation of each hindcast was done based on coefficient of variation (cv), Continuous Ranked Probability Score ($CRPS$, Hersbach (2000)), Mean Absolute Error (MAE), Brier Score (BS , Brier (1950)) and the number of false and true positive flood alerts. These scores were calculated for each lead time separately to evaluate the quality of the hindcast for different lead times.

To assess the spread of the ensemble of simulated discharges, the coefficient of variation was

determined with:

$$cv = \frac{1}{T} \sum_{t=1}^T \frac{\sigma_{Q_{mod}(t)}}{\overline{Q_{mod}(t)}} \quad (7)$$

where $\sigma_{Q_{mod}(t)}$ and $\overline{Q_{mod}(t)}$ (m^3d^{-1}) are the standard deviation and the mean of the ensemble of modelled discharge at time t , respectively, and T is the number of time step (days) in the reanalysis period.

The *CRPS* (Hersbach, 2000) was used to calculated whether the uncertainty of the forecast is correct and not over or underestimated. The *CRPS* is given by:

$$CRPS = \frac{1}{T} \sum_{i=1}^T \int_{x=-\infty}^{x=\infty} (F_i^f(x,t) - F_i^o(x,t))^2 dx \quad (8)$$

where $F_i^f(x,t)$ is the cumulative density function of the hindcast at time t , $F_i^o(x,t)$ is the cumulative density function of the observation at time t . $F_i^o(x,t)$ is given by a Heaviside function, with a step from 0 to 1 probability at the observed value. The *CRPS* is standardized by $\overline{Q_{obs}}$ for each validation location to enable a comparison between stations with a different magnitude of discharge.

To calculate if the hindcasts were biased the *MAE* was calculated using the ensemble mean of the forecast. The *MAE* is given by:

$$MAE = \frac{1}{T} \sum_{t=1}^T \frac{|\overline{Q_{mod}(t)} - Q_{obs}(t)|}{\overline{Q_{obs}}} \quad (9)$$

where $\overline{Q_{mod}(t)}$ and $Q_{obs}(t)$ (m^3d^{-1}) are the average hindcasted discharge and observed discharge at time t respectively and $\overline{Q_{obs}}$ is the average discharge over the evaluation period. *cv*, *CRPS* and *MAE* were used to evaluate the performance of each scenario and to determine the quality of each hindcasting scenario. Scores were standardized to enable a comparison between upstream and downstream stations without correcting for differences in discharge volumes. In

addition these scores were determined per lead time separately to enable a better comparison between the different scenarios and also to determine the flood forecasting performance of EFAS for different lead times.

To test the accuracy of the flood alerts (both timing and height of the flood peak), the Brier score is calculated for different flood thresholds and different lead times. The Brier score was calculated by:

$$BS = \frac{1}{T} \sum_{t=1}^T (f(t) - o(t)) \quad (10)$$

where $f(t)$ is the probability that discharge will exceed a certain threshold (calculated from the probability density function) and $o(t)$ is a binary value which is 0 if this threshold is not exceeded and 1 if it is exceeded. The Brier score can be calculated for different thresholds of discharge and different lead times. In this study we focussed on two threshold levels namely the 80th and 90th percentile of the discharge (Q_{80} , Q_{90}). Exceedance of these arbitrary levels will not necessarily cause flood situation, however to allow for evaluation of hindcast these high discharge events were used. Furthermore the number of false positives (flood forecast, no flood observed), missed (no flood forecasted, flood observed) and correctly forecasted (flood forecasted, flood observed) were calculated for each hindcasting scenario for the Q_{80} and Q_{90} . For Q_{80} and Q_{90} This resulted in 1035 and 2070 time steps with flooding Q_{90} and Q_{80} respectively divided over 10 to 18 flood events on average over the 16 validation locations.

3 Results

3.1 Reanalysis

To analyse the performance of the reanalysis the cv (Equation 7) is used to determine the uncertainty after the assimilation of the observations (Figure 4 and Table 2). In the Q_0 scenario, the model is not calibrated and no data is assimilated into the reanalysis to correct for incorrect model states. The uncertainty in the model simulation is large with a cv of 0.25. Uncertainty

even increases during extreme flood events, reducing the potential to use a model calibrated on expert knowledge without data assimilation for flood forecasting. The assimilation of three different satellite products ($Q0_{sat}$) results in a reduction of the cv of the discharge simulation to 0.136 compared to 0.25 for $Q0$ (Figure 4). This reduction is caused by the assimilation procedure which constrains the model to follow the observations and hence the spread between ensemble members is reduced. Soil moisture observations do not contain information on groundwater and routing processes, hence they impact the discharge simulation only indirect via surface runoff and percolation to the groundwater from the unsaturated zone. This results in the fact that the discharge simulations are not necessarily improved by assimilation of remotely sensed soil moisture observations.

Two scenarios were created where only discharge is assimilated into the model, namely $Q1$ and $Q7$. For $Q1$ only discharge from the outlet was used and for $Q7$, additional discharge observations (Figure 1) upstream were assimilated into the model. The assimilation of additional observation data reduces the cv to 0.08 for $Q1$ and to 0.04 for $Q7$, which is for both scenarios lower than for $Q0$ (Table 2). $Q1$ shows a small positive bias in the selected time period compared to the discharge observations. However, on average over the entire simulation a bias in $Q1$ is small.

Finally, two scenarios where both discharge and remotely sensed soil moisture observations are assimilated into the model ($Q1_{sat}$ and $Q7_{sat}$) were evaluated. In these scenarios the uncertainty is reduced compared to most other scenarios. However, peak discharge for $Q1_{sat}$ is overestimated, while baseflow simulations are better compared to $Q1$. Improved simulations are also observed with $Q7_{sat}$ compared to $Q7$ and the problem with overestimated peak discharge is gone $Q7_{sat}$ (Figure 4). An example time series is provided to show the impact of the satellite observations in the $Q7_{sat}$ scenario (Appendix Figure 10).

It must be mentioned that additional discharge data has a larger impact on the reduction of the uncertainty, than assimilation of remotely sensed soil moisture. Remotely sensed soil moisture enables a better simulation of the base flow compared to assimilation of discharge observation only. Although theoretical the reduction in uncertainty with the assimilation of remotely sensed soil moisture shows that this method has a high potential in sparsely gauged river basins to

correctly simulate discharges across the globe.

3.2 Hindcasting performance

The hindcast performance of each scenario was evaluated using the $CRPS$ (Equation 8) and the MAE (Equation 9). In general the uncertainty in the hindcast is reduced when more data is assimilated into the system leading to a better hindcast simulation (Figure 5). When more discharge data is assimilated, the uncertainty is more strongly reduced than with the assimilation of only remotely sensed soil moisture data (Figure 4 and 5). This is also confirmed by the $CRPS$ score for the different scenarios (Figure 6), where the decrease in $CRPS$ is strongest when more discharge data is used (Table 2). In general the $CRPS$ increases with increasing lead times for all scenarios with the exception of $Q1_{sat}$. Due to the larger spread for longer lead times (Figure 5) the $CRPS$ will increase, because forecasts with high uncertainty are penalized. The $CRPS$ for $Q1_{sat}$ is the highest indicating that this scenario has the lowest hindcasting skill of all scenarios (Figure 6 and Table 2). This is caused by the overestimation of most flood events, which results in a high $CRPS$. When more discharge data is assimilated ($Q0$ compared to $Q1$ and $Q7$) the $CRPS$ is reduced throughout the catchment for most locations including the outlet near Bratislava. When a combination of discharge data and satellite data is assimilated ($Q7_{sat}$), the quality of the hindcast is highest (Figure 5).

The MAE (Equation 9) is calculated for all scenarios for different lead times and locations (Figure 7). Compared to the scenario without assimilation of observations ($Q0$), only the scenarios where multiple discharge stations are assimilated ($Q7$ and $Q7_{sat}$) show an increase in performance. The best performance is generated by $Q7_{sat}$, which shows a low bias compared to the observed discharge. For $Q1_{sat}$ the MAE is relatively low, especially when compared to the $CRPS$. This is mainly caused by the accurate discharge simulation in base flow periods, resulting in a low MAE .

3.3 Flood hindcasting skill

The performance of each scenario was evaluated using the BS (Equation 10) and the number of false positive flood alerts. Due to the high spread within the ensemble the Q_0 in general has a low forecasting skill (Table 2). This is shown by the relative high BS (Figure 8) and the high number of false positive forecasts (Figure 9). Almost all flood events are correctly captured also for long lead times, which is caused by the overestimation of discharge in general (Figure 5). The overestimation of discharge also causes the high number of false positive flood forecasts, where around 90% of the exceedances of the threshold are incorrect and no flooding occurs. Compared to Q_0 the forecasting skill for $Q_{0_{sat}}$ is decreased, shown by an increasing BS and a higher number of false positives. The high number of false positives is the result of an even higher over estimation of the peak discharge in this scenario (Figure 5), which results in false flood alerts. The number of missed and correctly forecasted floods remains the same. The BS and the number of false positives for Q_1 and Q_7 is considerably lower than for Q_0 . Q_7 also has a better hindcast skill than the Q_1 caused by the increased number of observations used in the assimilation framework. The improved forecasting skill is also found in the BS for both $Q_{1_{sat}}$ and $Q_{7_{sat}}$ (Figure 8), which are for both scenarios lower than without the assimilation of remotely sensed soil moisture. For $Q_{1_{sat}}$ this is mainly caused by an increased performance in the upstream areas of the catchment, while $Q_{7_{sat}}$ shows an improved performance throughout the catchment. The number of false positive flood forecasts is reduced by 70% compared to the scenarios with only discharge assimilation, while the number of missed and correctly forecasted floods remains the same. This leads to the conclusion that even when the simulation of discharge throughout the catchment is used and discharge simulations are of a high quality, adding satellite data will lead to an improvement in the forecasting skills of the hydrological model.

3.4 Hindcasting performance with limited assimilation

Two additional scenarios have been evaluated were the model was calibrated on discharge observations alone and either remotely sensed soil moisture is assimilated ($Q_{7_{sat}DA}$) or no observations are assimilated ($Q_{7_{noDA}}$) in the reanalysis period (Table 2). The reanalysis for

$Q7_{noDA}$ shows the largest spread in the reanalysis (indicated by a large cv), while with the assimilation of remotely sensed soil moisture ($Q7_{satDA}$) this uncertainty is reduced. However, the uncertainties remain larger than for scenarios $Q7$ and $Q7_{sat}$ where in both cases discharge data has been assimilated.

5 The uncertainty in the hindcasting performance is reduced for $Q7_{noDA}$ compared to $Q7$. This indicates that the more accurate representation of the soil moisture will reduce the uncertainty in model simulations and hence hindcasts. For both $Q7_{noDA}$ and $Q7_{satDA}$ the MAE does not show a increased performance, indicating that the bias is not reduced compared to $Q7$.

10 As expected, the hindcast skills scores (MAE) are reduced when the satellite data is used in the assimilation compared to no assimilation. Also, compared to $Q7$ the hindcast skill for the extreme events is not increasing. However, compared to $Q7$ and $Q7_{noDA}$ the assimilation of satellite data will increase to hindcast skill for the less severe flood.

4 Conclusions

15 In this study we evaluated the added value of remotely sensed soil moisture in an operational flood forecasting system. The gain from assimilation of soil moisture observations is compared to assimilation of only discharge and the combination of discharge and soil moisture observations. The European Flood Awareness System (EFAS) was used for a hindcasting experiment in the Upper Danube. Hindcasts were made for a period of one year and the results compared
20 for six different scenarios.

The assimilation of remotely sensed soil moisture has an impact on the simulation of discharge as shown by other studies (e.g. Pauwels et al., 2001; Brocca et al., 2010, 2012; Draper et al., 2011). However, in this study we show that the impact is not only limited to small catchments with a spatial extent close to or smaller than the satellite resolution but also works for larger
25 catchments.

We show that the assimilation of remotely sensed soil moisture improves the flood forecasting especially when used in combination with assimilation of distributed discharge observations.

The uncertainty in the discharge simulations is reduced and biases in the simulation are reduced when satellite data is assimilated. In scenarios where only discharge from the outlet is used in combination with satellite observations, the peak discharges are generally overestimated. Although, this will result in a less accurate simulation of discharge it will not impact the quality of the forecasting of the flood events.

Floods are better predicted when soil moisture data is assimilated into EFAS in combination with discharge observations and the number of false alerts is reduced compared to scenarios when remotely sensed soil moisture observations are not used. Although the gain of using more discharge observations remains larger, soil moisture observations improve the quality of the flood alerts, both in terms of timing and in the exact height of the flood peak.

Two additional scenarios were studied, where only calibration of the hydrological model was used and no assimilation or assimilation of only satellite data. These scenarios were created to study the added value of the assimilation compared to only calibration of the hydrological model. We found that the *cv*, *CRPS*, *MAE* and *BS* are all reduced by the assimilation of remotely sensed soil moisture compared to no assimilation. However, assimilation of discharge reduces uncertainties more than assimilation of remotely sensed soil moisture. Simulations without data assimilation tend to have biases in the simulation and a larger ensemble spread than scenarios with data assimilation, while the reduced uncertainty resulting from assimilation will lead to a increased reliability of flood forecasts. These results show that the assimilation of soil moisture will result in an increased performance compared to not assimilating observations. This is important for ungauged basins, where satellite data is available and discharge observations are not available or not available in near-real time. Additionally these results show the added value of assimilation of observations into the EFAS system, compared to the current set-up.

In conclusion, we show that the uncertainty in the flood forecasts is reduced when discharge observations and satellite data are assimilated into the hydrological model of the EFAS system for the Upper Danube. The addition of remotely sensed soil moisture to existing discharge observations reduces the number of false positive flood alerts and thereby increases the reliability of the flood awareness system. Although the number of the data available via

satellite retrievals still remains a challenge in an operational system, the potential benefits could lead to a significant reduction in the false flood alerts, possibly also for other catchments. This will reduce the number of unnecessary precautions taken by the responsible governments and increase the confidence and willingness to act upon these flood alerts.

5

Acknowledgements. This research was funded by a grant from the user support program Space Research of NWO (contract number NWO GO-AO/30). JRC is acknowledged for being able to work with the Danube model set-up. GRDC and ECMWF are acknowledged for making publicly available their data. Jutta Thielen and two anonymous reviewers are acknowledged for their constructive reviews of the draft article.

10

References

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger, F.: GloFAS-global ensemble streamflow forecasting and flood early warning, *Hydrology and Earth System Sciences*, 17, 1161–1175, 2013.

15

Aon Benfield: Impact Forecasting - June 2013, Global Catastrophe Recap report, <http://www.claimsjournal.com/news/international/2013/07/12/232685.htm>, retrieved on 07-10-2013, 2013.

Bartalis, Z., Kidd, R., and Scipal, K.: Development and implementation of a Discrete Global Grid System for soil moisture retrieval using the MetOp ASCAT scatterometer, in: 1st EPS/MetOp RAO Workshop, ESA SP-618, Frascati, Italy, 15-17 May 2006, 2006.

20

Bisselink, B., van Meijgaard, E., Dolman, A. J., and De Jeu, R. A. M.: Initializing a regional climate model with satellite-derived soil moisture, *Journal of Geophysical Research*, 116, D02 121, doi:10.1029/2010JD014534, 2011.

Brier, G.: Verification of Forecasts Expressed in Terms of Probability, *Monthly Weather Review*, 78, 1–3, 1950.

25

Brocca, L., Melone, F., Moramarco, T., Wagner, W., Naeimi, V., Bartalis, Z., and Hasenauer, S.: Improving runoff prediction through the assimilation of the ASCAT soil moisture product, *Hydrology and Earth System Sciences*, 14, 1881–1893, doi:10.5194/hess-14-1881-2010, 2010.

Brocca, L., Moramarco, T., Melone, F., Wagner, W., Hasenauer, S., and Hahn, S.: Assimilation of

- Surface- and Root-Zone ASCAT Soil Moisture Products Into Rainfall Runoff Modeling, *Geoscience and Remote Sensing, IEEE Transactions on*, 50, 2542–2555, doi:10.1109/TGRS.2011.2177468, 2012.
- Burgers, G., Van Leeuwen, P., and Evensen, G.: Analysis Scheme in the Ensemble Kalman Filter, *Monthly Weather Review*, 126, 1719–1724, 1998.
- 5 Chen, F., Crow, W. T., Starks, P. J., and Moriasi, D. N.: Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture, *Advances in Water Resources*, 34, 526 – 536, doi: <http://dx.doi.org/10.1016/j.advwatres.2011.01.011>, 2011.
- Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., and Uddstrom, M. J.: Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model, *Advances in Water Resources*, 31, 1309 – 1324, doi:<http://dx.doi.org/10.1016/j.advwatres.2008.06.005>, 2008.
- 10 Crow, W. T., Miralles, D. G., and Cosh, M. H.: A Quasi-Global Evaluation System for Satellite-Based Surface Soil Moisture Retrievals, *IEEE Transactions on Geoscience and Remote Sensing*, 48, 2516–2527, 2010.
- 15 De Jeu, R. A. M., Wagner, W., Holmes, T. R. H., Dolman, A. J., Van de Giesen, N. C., and Friesen, J.: Global Soil Moisture Patterns Observed by Space Borne Microwave Radiometers and Scatterometers, *Surveys in Geophysics*, 28, 399–420, doi:doi10.1007/s10712-008-9044-0, 2008.
- De Roo, A., Wesseling, C., and Van Deursen, W.: Physically-based river basin modelling within a GIS: The LISFLOOD model., *Hydrological Processes*, 14, 1981–1992, 2000.
- 20 Di Baldassarre, G. and Montanari, A.: Uncertainty in river discharge observations: a quantitative analysis, *Hydrology and Earth System Sciences*, 13, 913–921, doi:10.5194/hess-13-913-2009, 2009.
- Dorigo, W. A., Scipal, K., Parinussa, R. M., Liu, Y. Y., Wagner, W., de Jeu, R. A. M., and Naeimi, V.: Error characterisation of global active and passive microwave soil moisture datasets, *Hydrology and Earth System Sciences*, 14, 2605–2616, doi:10.5194/hess-14-2605-2010, 2010.
- 25 Draper, C., Walker, J. P., Steinle, P., De Jeu, R. A. M., and Holmes, T. R. H.: An evaluation of AMSR-E derived soil moisture over Australia, *Remote Sensing of Environment*, 113, 703–710, 2009.
- Draper, C., Mahfouf, J.-F., Calvet, J.-C., Martin, E., and Wagner, W.: Assimilation of ASCAT near-surface soil moisture into the SIM hydrological model over France, *Hydrology and Earth System Sciences*, 15, 3829–3841, doi:10.5194/hess-15-3829-2011, 2011.
- 30 Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research: Oceans*, 99, 10 143–10 162, doi:10.1029/94JC00572, 1994.
- Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, *Ocean*

- Dynamics, 53, 343–367, doi:10.1007/s10236-003-0036-9, 2003.
- Evensen, G.: The ensemble Kalman filter for combined state and parameter estimation, *Control Systems, IEEE*, 29, 83 –104, doi:10.1109/MCS.2009.932223, 2009.
- Hersbach, H.: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems, *Weather and Forecasting*, 15, 559–570, doi:10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2, 2000.
- Karssenber, D., Schmitz, O., Salamon, P., de Jong, K., and Bierkens, M. F. P.: A software framework for construction of process-based stochastic spatio-temporal models and data assimilation, *Environmental Modelling & Software*, 25, 489 – 502, doi:10.1016/j.envsoft.2009.10.004, 2010.
- Kerr, Y., Waldteufel, P., Richaume, P., Wigneron, J., Ferrazzoli, P., Mahmoodi, A., Al Bitar, A., Cabot, F., Gruhier, C., Juglea, S., Leroux, D., Mialon, A., and Delwart, S.: The SMOS Soil Moisture Retrieval Algorithm, *Geoscience and Remote Sensing, IEEE Transactions on*, 50, 1384 –1403, doi:10.1109/TGRS.2012.2184548, 2012.
- Lee, H., Seo, D.-J., Liu, Y., Koren, V., McKee, P., and Corby, R.: Variational assimilation of streamflow into operational distributed hydrologic models: effect of spatiotemporal adjustment scale, *Hydrology and Earth System Sciences Discussions*, 9, 93–138, doi:10.5194/hessd-9-93-2012, 2012.
- Loew, A., Holmes, T. R. H., and De Jeu, R. A. M.: The European heat wave 2003: Early indicators from multisensoral microwave remote sensing?, *Journal of Geophysical Research*, 114, D05 103, doi:10.1029/2008JD010533, 2009.
- Matgen, P., Fenicia, F., Heitz, S., Plaza, D., de Keyser, R., Pauwels, V. R., Wagner, W., and Savenije, H.: Can ASCAT-derived soil wetness indices reduce predictive uncertainty in well-gauged areas? A comparison with in situ observed soil moisture in an assimilation application, *Advances in Water Resources*, 44, 49 – 65, doi:http://dx.doi.org/10.1016/j.advwatres.2012.03.022, 2012.
- Naeimi, V., Scipal, K., Bartalis, Z., Hasenauer, S., and Wagner, W.: An Improved Soil Moisture Retrieval Algorithm for ERS and METOP Scatterometer Observations, *IEEE Transactions on Geoscience and Remote Sensing*, 47, 1999–2013, doi:10.1109/TGRS.2008.2011617, 2009.
- Nester, T., Komma, J., Viglione, A., and Blschl, G.: Flood forecast errors and ensemble spread - A case study, *Water Resources Research*, 48, W10 502, doi:10.1029/2011WR011649, 2012.
- Njoku, E., Jackson, T. J., Lakshmi, V., Chan, T., and Nghiem, S.: Soil moisture retrieval from AMSR-E, *IEEE Transactions on Geoscience and Remote Sensing*, 41, 215 – 229, doi:10.1109/TGRS.2002.808243, 2003.
- Owe, M., De Jeu, R. A. M., and Holmes, T. R. H.: Multisensor historical climatology of satellite-derived global land surface moisture, *Journal of Geophysical Research*, 113, F01 002, doi:10.1029/

2007JF000769, 2008.

Pauwels, V. R., Hoeben, R., Verhoest, N. E., and Troch, F. P. D.: The importance of the spatial patterns of remotely sensed soil moisture in the improvement of discharge predictions for small-scale basins through data assimilation, *Journal of Hydrology*, 251, 88 – 102, doi:10.1016/S0022-1694(01)00440-1, 2001.

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

Santanello, J., Peters-Lidard, C. D., Garcia, M. E., Mocko, D. M., Tischler, M. A., Moran, M. S., and Thoma, D.: Using remotely-sensed estimates of soil moisture to infer soil texture and hydraulic properties across a semi-arid watershed, *Remote Sensing of Environment*, 110, 79 – 97, doi:10.1016/j.rse.2007.02.007, 2007.

Scipal, K., Drusch, M., and Wagner, W.: Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system, *Advances in Water Resources*, 31, 1101 – 1112, doi:10.1016/j.advwatres.2008.04.013, 2008.

Sutanudjaja, E. H., van Beek, L. P. H., de Jong, S. M., van Geer, F. C., and Bierkens, M. F. P.: Calibrating a large-extent high-resolution coupled groundwater-land surface model using soil moisture and discharge data, *Water Resources Research*, pp. n/a–n/a, doi:10.1002/2013WR013807, 2013.

Thielen, J., Bartholmes, J., Ramos, M., and de Roo, A.: The European Flood Alert System Part 1: Concept and development, *Hydrology and Earth System Sciences*, 13, 125–140, doi:10.5194/hess-13-125-2009, 2009.

Van Der Knijff, J. M., Younis, J., and De Roo, A. P. J.: LISFLOOD: a GIS based distributed model for river basin scale water balance and flood simulation, *International Journal of Geographical Information Science*, 24, 189–212, doi:10.1080/13658810802549154, 2010.

Vrugt, J., Gupta, H., Nuallin, B., and Bouten, W.: Real-time data assimilation for operational ensemble streamflow forecasting, *Journal of Hydrometeorology*, 7, 548–565, 2006.

Wagner, W., Lemoine, G., and Rott, H.: A Method for Estimating Soil Moisture from ERS Scatterometer and Soil Data, *Remote Sensing of Environment*, 70, 191 – 207, doi:10.1016/S0034-4257(99)00036-X, 1999.

Wagner, W., Blöschl, 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, *Nordic Hydrology*, 30, 1 – 20, doi:10.2166/nh.2007.029, 2007.

Table 1. Hindcasting scenarios for the EFAS system including abbreviations and assimilated data used to create a re-analysis time series from which hindcasts were initiated. The calibration indicates the data used by Wanders et al. (2013) to calibrate the hydrological model.

Scenario	Hindcast		Calibration
	Number of discharge stations	Satellite data	Data for calibration
$Q0$	0 stations	None	None, expert knowledge
$Q0_{sat}$	0 stations	All satellite data	Satellite data
$Q1$	1 stations	None	1 discharge observation
$Q1_{sat}$	1 stations	All satellite data	1 discharge station & satellite data
$Q7$	7 stations	None	7 discharge stations
$Q7_{sat}$	7 stations	All satellite data	7 discharge stations & satellite data
$Q7_{noDA}$	None	None	7 discharge stations
$Q7_{satDA}$	None	All satellite data	7 discharge stations

Wanders, N., Karssenber, D., Bierkens, M., Parinussa, R., de Jeu, R., van Dam, J., and de Jong, S.: Observation uncertainty of satellite soil moisture products determined with physically-based modeling, *Remote Sensing of Environment*, 127, 341 – 356, doi:10.1016/j.rse.2012.09.004, 2012.

5 Wanders, N., Bierkens, M., de Jong, S., de Roo, A., and Karssenber, D.: The benefits of using remotely sensed soil moisture in parameter identification of large-scale hydrological models, *Water Resources Research*, submitted, –, 2013.

Wesseling, C., Karssenber, D., Burrough, P., and Van Deursen, W.: Integrating dynamic environmental models in GIS: The development of a Dynamic Modelling language, *Transactions in GIS*, 1, 40–48, doi:10.1111/j.1467-9671.1996.tb00032.x, 1996.

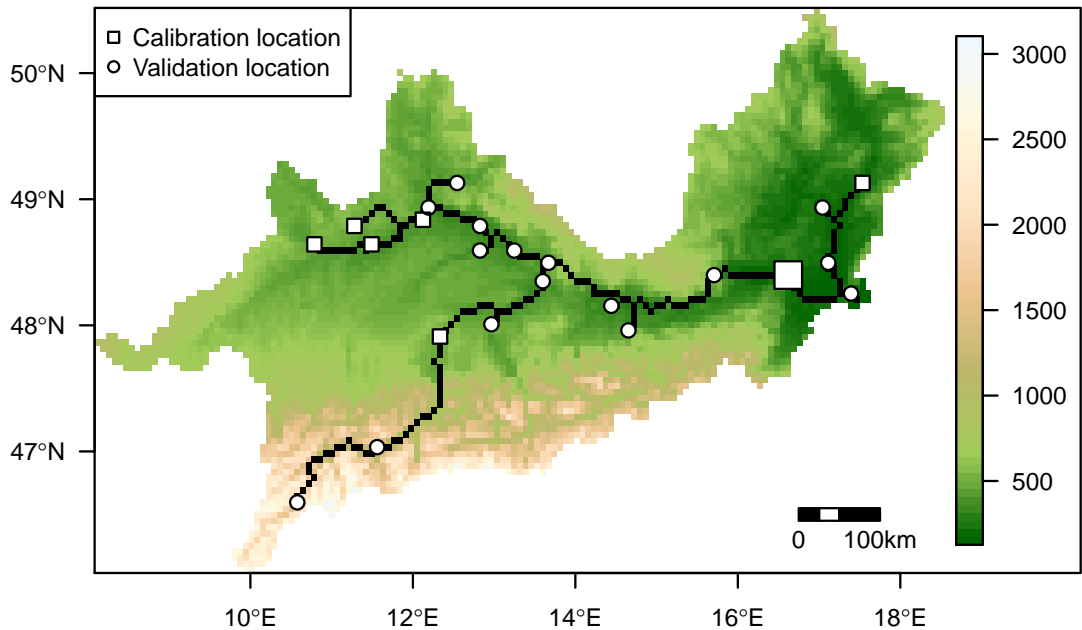


Fig. 1. Digital elevation map of the Upper Danube catchment, colours indicate elevation (m), indicated in black is the river network, square symbols indicate locations for calibration on discharge observations, circles indicate locations for validation on discharge observations. The large square near the outlet (right) is the location used for calibration if only one discharge time series is used (Q_1 and $Q_{1_{sat}}$).

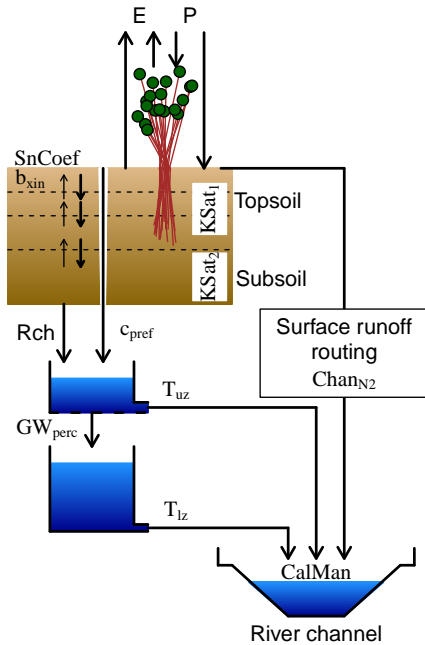


Fig. 2. LISFLOOD model set-up, with fluxes; precipitation (P), evaporation (E), recharge from the unsaturated zone to the groundwater (Rch). The calibration parameters of the model are; snowmelt coefficient ($SnCoef$), Xinanjiang shape parameter (b_{xin}), saturated conductivity of the topsoil ($KSat_1$), saturated conductivity of the subsoil ($KSat_2$), empirical shape parameter preferential macro-pore flow (c_{pref}), maximum percolation rate from upper to lower groundwater (GW_{perc}), reservoir constant upper groundwater (T_{uz}), reservoir constant lower groundwater (T_{lz}), surface runoff roughness coefficient ($Chan_{N2}$), channel Mannings roughness coefficient ($CalMan$).

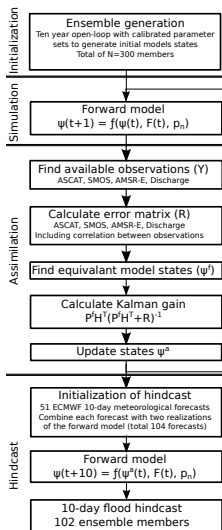


Fig. 3. Flowchart of the hindcasting procedure including initialization of the model and the construction of the reanalysis time series.

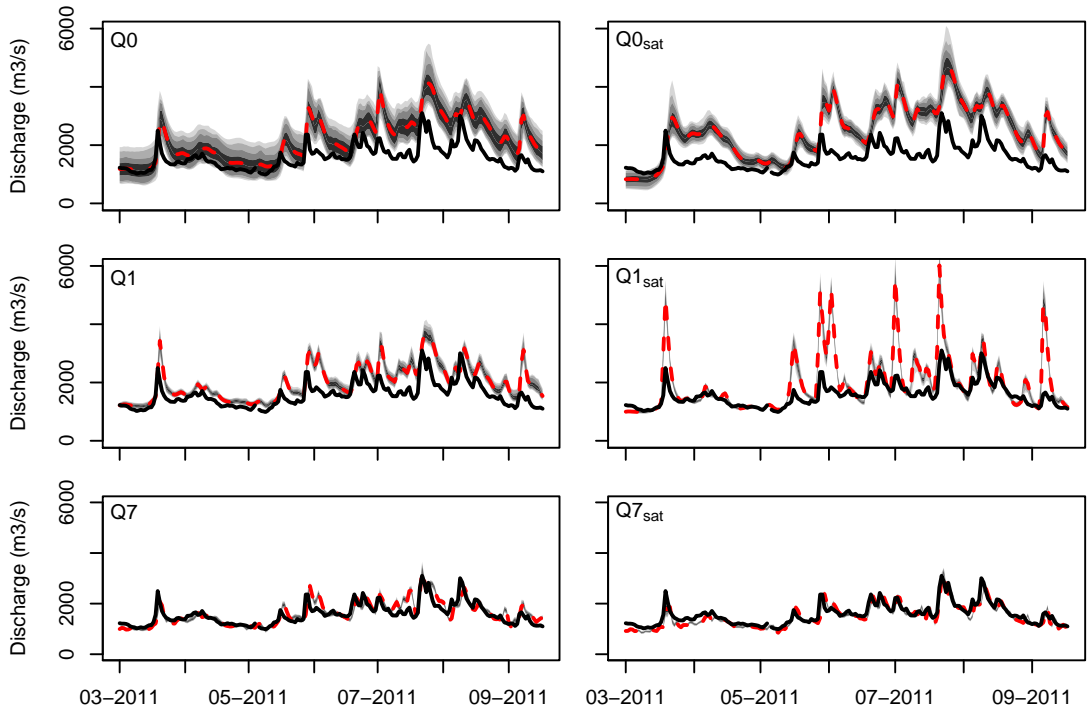


Fig. 4. Reanalysis time series of discharge at the outlet of the Upper Danube catchment (Figure 1) for part of the hindcasting period. In grey are all model realizations, the ensemble mean is given by the red line and the solid black line gives the observed discharge value. The different assimilation scenarios are indicated on the left; for explanation of scenarios see Table 1. The different assimilation scenarios are indicated in the top left corner of each plot.

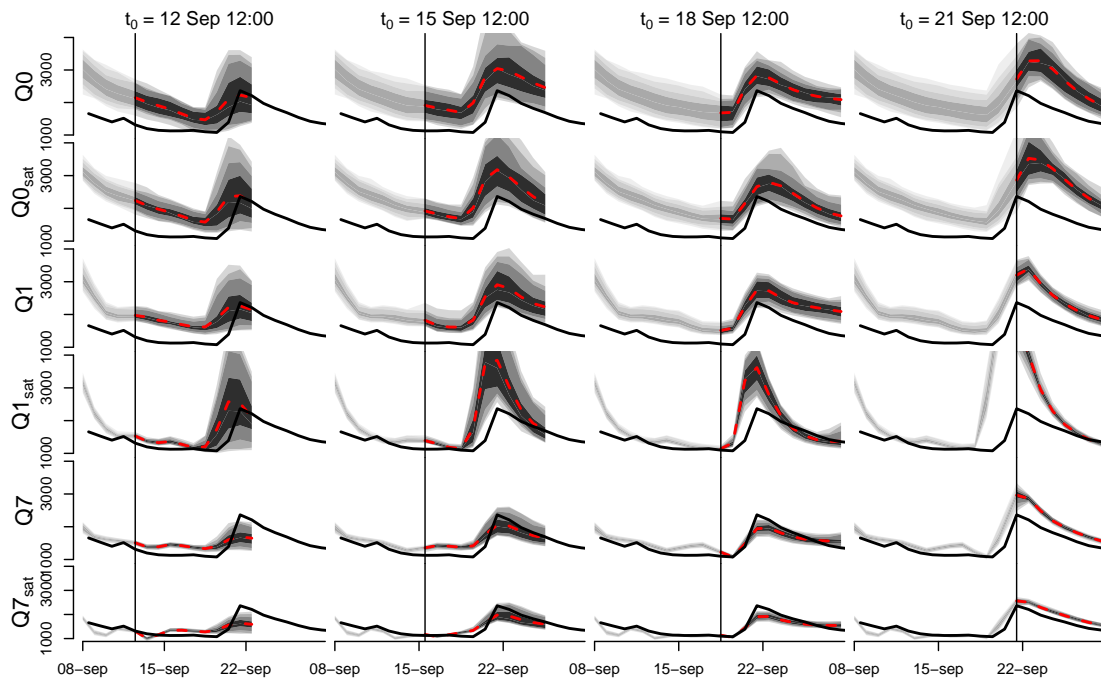


Fig. 5. Example forecast time series of discharge at the outlet of the Upper Danube catchment (Figure 1) for part of the hindcasting period. In grey are all model realizations, the ensemble mean is given by the red dashed line and solid black line gives the observed discharge value. Each column of figures gives the hindcast for a particular time, indicated by the vertical line.

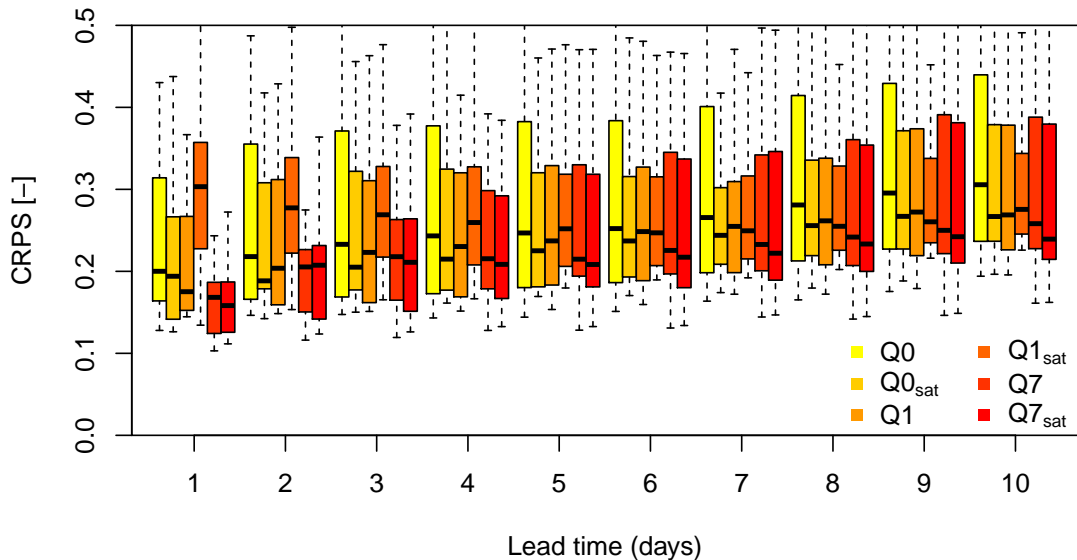


Fig. 6. Continuous Ranked Probability Scores (*CRPS*) for different forecasting times for the European Awareness Flood System (EFAS). Each boxes only contains the *CRPS* for the 16 validation locations for a period of 1 year with two forecasts per day.

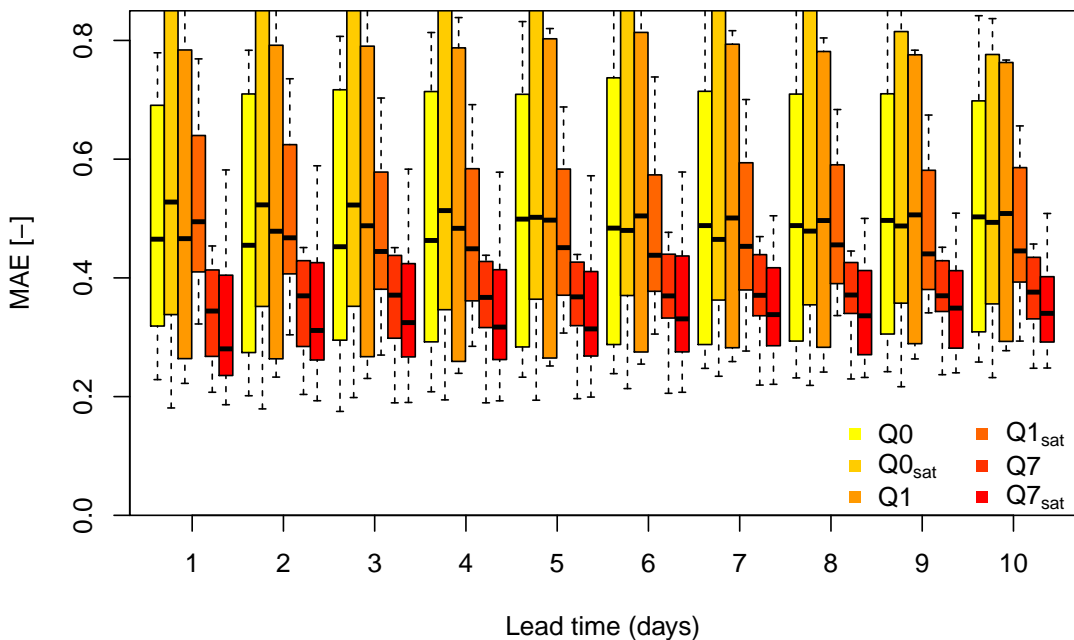


Fig. 7. Mean Absolute Error (MAE) for different forecasting times for the European Awareness Flood System (EFAS). MAE are standardized by dividing the MAE through the mean discharge. Each boxes only contains the MAE for the 16 validation locations for a period of 1 year with two forecasts per day.

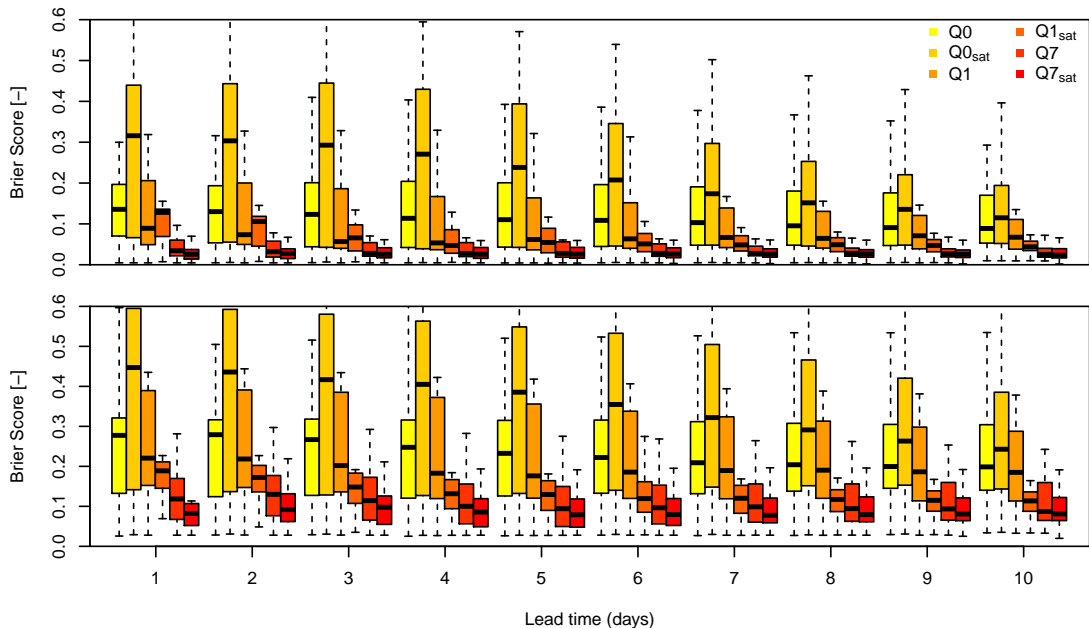


Fig. 8. Brier Score (BS) for different forecasting times for the European Flood Awareness System (EFAS) in the Upper Danube (Figure 1). Each boxes only contains the BS for the 16 validation locations for a period of 1 year with two forecasts per day. The Brier scores for the 90% threshold (top) and the 80% threshold (bottom) are given. A total of 1035 and 2070 time steps (90th and 80th percentile, respectively) with flooding were observed for the Upper Danube

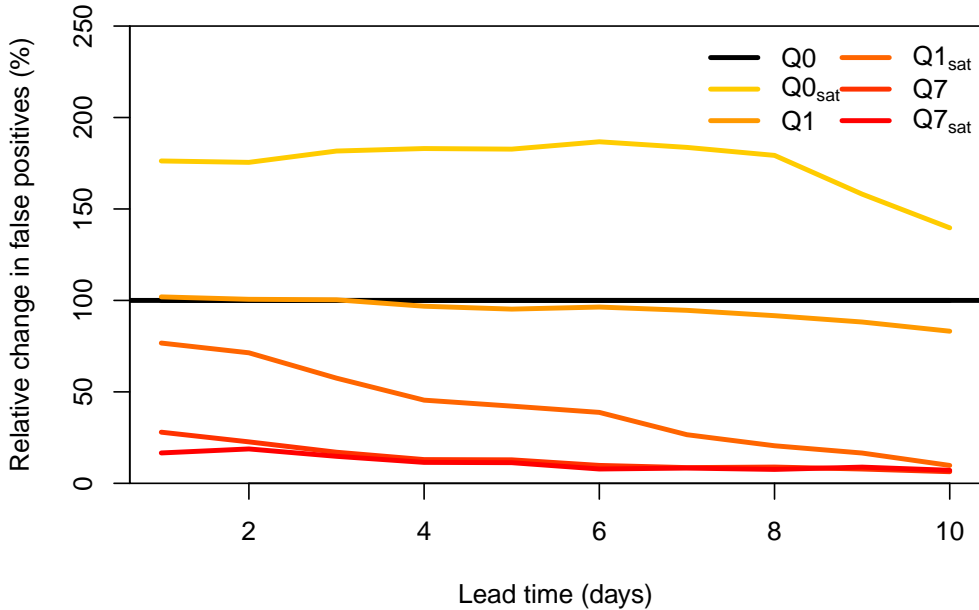


Fig. 9. Relative changes in false positive flood alerts for the 90th percentile threshold, compared to no assimilation scenario ($Q0$) for different forecasting times. A total of 1035 time steps with flooding were observed for the Upper Danube.

Table 2. Average skill scores for different hindcast scenarios for the EFAS system. Scores are averaged over different forecasting times and for different locations with discharge observation in the Upper Danube (Figure 1).

Scenario	cv	$CRPS$	MAE	$BS Q_{90}$	$BS Q_{80}$
$Q0$	0.272	0.328	0.620	0.130	0.257
$Q0_{sat}$	0.161	0.252	0.791	0.220	0.363
$Q1$	0.084	0.203	0.702	0.168	0.314
$Q1_{sat}$	0.075	0.306	0.508	0.084	0.177
$Q7$	0.049	0.186	0.382	0.038	0.166
$Q7_{sat}$	0.047	0.182	0.309	0.029	0.096
$Q7_{noDA}$	0.055	0.187	0.385	0.042	0.173
$Q7_{satDA}$	0.053	0.183	0.384	0.040	0.160

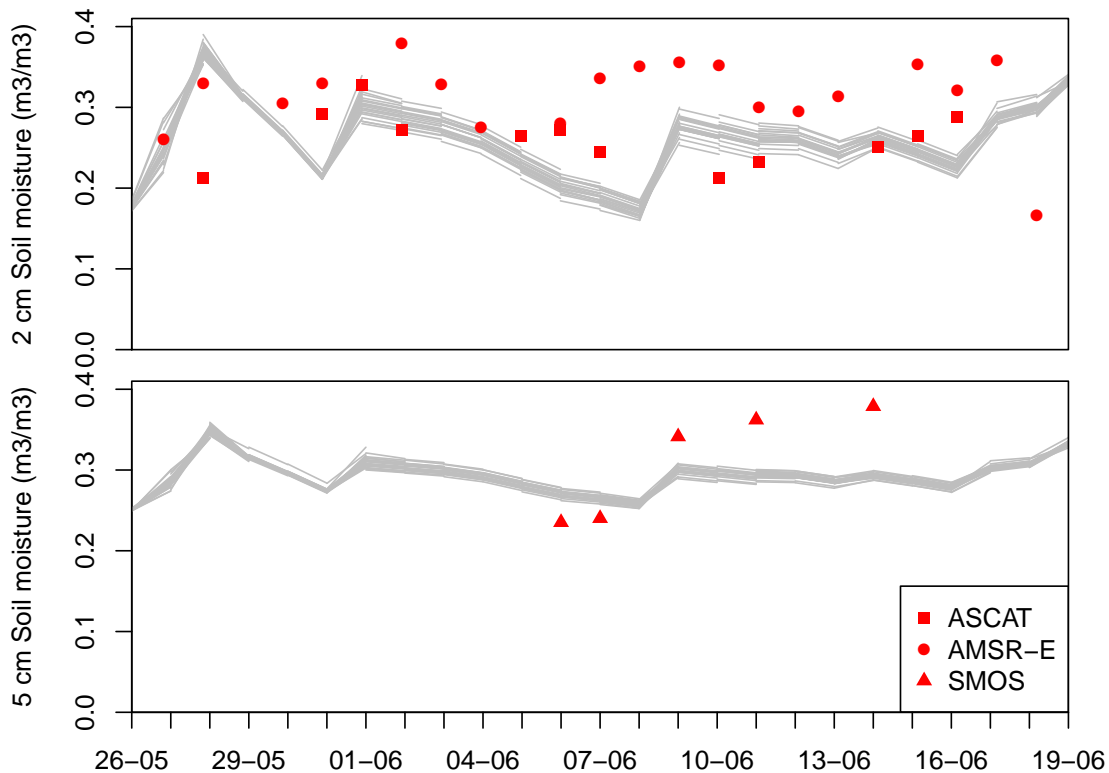


Fig. 10. Example time series (14°E , 48°N) of simulated soil moisture for $0-2\text{ cm}$ (upper panel) and the simulated soil moisture for $0-5\text{ cm}$ (lower panel) for the analysis period, $Q7_{sat}$ scenario. Observations of remotely sensed soil moisture are shown from three different sensors. Jumps in the time series at assimilation moments indicate the update of model states according to the observations.