Manuscript prepared for Hydrol. Earth Syst. Sci.
with version 5.0 of the LaTeX class copernicus.cls.
Date: 6 March 2015

Improving operational flood ensemble prediction by the assimilation of satellite soil moisture: comparison between lumped and semi-distributed schemes.

C. Alvarez-Garreton[1], **D. Ryu**[1], **A.W. Western**[1], **C.-H. Su**[1], **W.T. Crow**[2], **D.E. Robertson**[3], **and C. Leahy**[4]

[1]Department of Infrastructure Engineering, The University of Melbourne, Parkville, Victoria, Australia
[2]USDA-ARS Hydrology and Remote Sensing Laboratory, Beltsville, Maryland, United States
[3]CSIRO Land and Water, Australia
[4]Bureau of Meteorology, Melbourne, Victoria, Australia

Correspondence to: Camila Alvarez-Garreton
(calvarez@student.unimelb.edu.au)

Abstract. Assimilation of remotely sensed soil moisture data (SM-DA) to correct soil water stores of rainfall-runoff models has shown skill in improving streamflow prediction. In the case of large and sparsely monitored catchments, SM-DA is a particularly attractive tool. Within this context, we assimilate satellite soil moisture (SM) retrievals from the Advanced Microwave Scanning Radiometer (AMSR-E), the Advanced Scatterometer (ASCAT) and the Soil Moisture and Ocean Salinity (SMOS) instrument, using an Ensemble Kalman filter to improve operational flood prediction within a large (>40,000km^2) semi-arid catchment in Australia. We assess the importance of accounting for channel routing and the spatial distribution of forcing data by applying SM-DA to a lumped and a semi-distributed scheme of the probability distributed model (PDM). Our scheme also accounts for model error representation by explicitly correcting bias in soil moisture and streamflow in the ensemble generation process, and for seasonal biases and errors in the satellite data.

Before assimilation, the semi-distributed model provided a more accurate streamflow prediction (Nash-Sutcliffe efficiency, NSE=0.77) than the lumped model (NSE=0.67) at the catchment outlet. However, this did not ensure good performance at the "ungauged" inner catchments (two of them with NSE below 0.3). After SM-DA, the streamflow ensemble prediction at the outlet was improved in both the lumped and the semi-distributed schemes: the root mean square error of the ensemble was reduced by 22% and 24%, respectively; the false alarm ratio was reduced by 9% in both cases; the peak volume error was reduced by 58% and 1%, respectively; the ensemble skill was improved (evidenced by 12% and 13% reductions in the continuous ranked probability scores, respectively); and the ensemble reliability was increased in both cases (expressed by flatter rank histograms). SM-DA did not improve NSE.

Our findings imply that even when rainfall is the main driver of flooding in semi-arid catchments, adequately processed satellite SM can be used to reduce errors in the model soil moisture, which in turn provides better streamflow ensemble prediction. We demonstrate that SM-DA efficacy is enhanced when the spatial distribution in forcing data and routing processes are accounted for. At ungauged locations, SM-DA is effective at improving some characteristics of the streamflow ensemble prediction; however, the updated prediction is still poor since SM-DA does not address the systematic errors found in the model prior to assimilation.

1 Introduction

Floods have large costs to society, causing destruction of infrastructure and crops, erosion, and in the worst cases, injury and loss of life (Thielen et al., 2009). To reduce flood impacts on public safety and the economy, early and accurate alert systems are needed. These systems rely on hydrologic models, whose accuracy in turn is highly dependent on the quality of the data used to force and calibrate them. Therefore, in the case of sparsely monitored and ungauged catchments, flood prediction suffers from large uncertainties.

A plausible approach to reduce model uncertainties in the sparsely monitored catchments is to exploit remotely sensed hydro-meteorological observations to correct the states or parameters of the model in a data assimilation framework. Within this context, satellite soil moisture (SM) products are appealing given the vital role of SM in runoff generation. SM influences the partitioning of energy and water (rainfall, infiltration and evapotranspiration) between the land surface and the atmosphere (Western et al., 2002). Satellite SM observations provide global scale information and can be obtained in near real time at regular and reasonably frequent time intervals. This makes them valuable for improving the representation of catchment wetness. The accuracy of these observations has been assessed by a number of studies (Albergel et al., 2009; Draper et al., 2009; Albergel et al., 2010; Gruhier et al., 2010; Brocca et al., 2011; Albergel et al., 2012; Su et al., 2013). In general, they have shown promising performance with moderate correlation between satellite SM and ground data, but with significant bias at some locations.

In the last decade a large number of studies have explored satellite SM data assimilation (SM-DA) to correct the soil water states of models. These studies can be categorised into two main groups; the first, and larger group, has focused on the improvement of the SM predicted by the model (generally working with land surface models, e.g., Crow and van Loon, 2006; Crow and Reichle, 2008; Crow and Van den Berg, 2010; Reichle et al., 2008; Ryu et al., 2009). The second, and smaller group (where our study fits), has focused on the improvement of streamflow prediction in rainfall-runoff models (Francois et al., 2003; Brocca et al., 2010b, 2012; Alvarez-Garreton et al., 2013, 2014; Chen et al., 2014; Wanders et al., 2014).

Studies from the first group evaluate the prediction improvement of the same variable that is updated in the assimilation scheme (SM). Improvements in streamflow predictions investigated by studies in the second group are not exclusively influenced by better representation of SM. The potential improvement of streamflow predictions in the latter case is constrained by the particular runoff mechanisms operating within a catchment. Accordingly, even when a model structure and parametrisation are capable of representing the runoff mechanisms, improving streamflow prediction by reducing error in soil moisture depends on the error covariance between these two components. This error covariance (which in the model space will be defined by the representation of the different sources of uncertainty) may become marginal when the errors in streamflow come mainly from errors in rainfall input data (Crow and Ryu, 2009). This physical constraint is case specific and determines the potential skill of SM-DA for improving streamflow prediction. To understand and assess this skill, further studies focusing on the improvement of streamflow prediction are needed with different model characteristics, such as structure, parametrisation and performance before assimilation; and with different catchment characteristics, such as climate, scale, soils, geology, land cover and density of monitoring network. Among the latter, semi-arid catchments present distinct rainfall-runoff processes which have been rarely studied in SM-DA.

Here we address this gap by studying the Warrego River catchment in Australia, a large and sparsely monitored semi-arid basin. We set up the probability distributed model (PDM) within the catchment, and assimilate passive and active satellite SM products using an Ensemble Kalman filter (Evensen, 2003) for the purpose of improving operational flood prediction. We devise an operational SM-DA scheme to answer three main questions. 1) While rainfall is presumably the main driver of flood generation in semi-arid catchments, can we effectively improve streamflow prediction by correcting the antecedent soil water state of the model? 2) What is the impact of accounting for channel routing and the spatial distribution of forcing data on SM-DA performance? 3) What are the prospects for improving streamflow prediction within ungauged sub-catchments using satellite SM?.

A series of SM-DA experiments using a lumped version of PDM have already been undertaken in this study catchment by Alvarez-Garreton et al. (2014). They found that assimilating passive microwave satellite SM improved flood prediction, while highlighting specific limitations in their scheme. This paper expands on this previous result in a number of key ways. We improve the representation of model error by explicitly treating forcing, parameter and structural errors. We devise a more robust ensemble generation process by correcting biases in soil moisture and streamflow predictions. We incorporate additional satellite products and apply instrumental variable regression techniques for seasonal rescaling and observations error estimation. Furthermore, we employ a semi-distributed scheme to evaluate the advantages of accounting for channel routing and the spatial distribution of forcing data.

In this paper, Sect. 2 presents a description of the study catchment and the data used. Section 3 presents the methodology, including a description of the rainfall-runoff model, the EnKF formulation and the specific steps for setting up the SM-DA scheme. These include the error model estimation, estimation of profile SM based on the satellite surface data, the rescaling of satellite observations and observation error estimation. Section 4 presents the results and discussion. Section 5 summarises the main conclusions of the study.

2 Study area and data

The study area is the semi-arid Warrego catchment (42,870 km^2) located in Queensland, Australia (Fig.1). The catchment has an important flooding history, with at least three major floods within the last 15 years. The study area also features geographical and climatological conditions that enable satellite SM retrievals to have higher accuracy than in other areas. These conditions include the size of the catchment, the semi-arid climate and the low vegetation cover. Moreover, the ground-monitoring network within the catchment is sparse thus satellite data is likely to be more valuable than in well-instrumented catchments. The catchment has summer-dominated rainfall with mean monthly rainfall accumulation of 80 mm in January, and 20 mm in August. Mean maximum daily temperature in January is above 30°C and below 20°C in July. The runoff seasonality is characterised by peaks in summer months and minimum values in winter and spring. The mean annual precipitation over the catchment is 520 mm. Regarding the governing runoff mechanisms within the study catchment, Alvarez-Garreton et al. (2014) showed that streamflow has a negligible baseflow component and the surface runoff is generated only when a wetness threshold is exceeded. They concluded that soil moisture exerts an important control on the runoff generation mechanisms. In this work, the runoff mechanisms analysis is deepened by looking at model predictions (Sect. 3.1).

Daily rainfall data was computed from the Australian Water Availability Project (AWAP), which has a grid resolution of 0.05° (Jones et al., 2009). Hourly streamflow records were collected from the State of Queensland, Department of Natural Resources and Mines (*http://watermonitoring.dnrm.qld.gov.au*) (Fig.1). Daily discharge was calculated based on the daily AWAP time convention (9am-9am local time, UTC+10h). The flood classification for the study catchment (at the catchment outlet, N7) was provided by the Australian Bureau of Meteorology as river height threshold values, corresponding to minor, moderate and major floods. These threshold values expressed as streamflow (mm/day) are 0.06, 0.55 and 2.05, respectively and relate to flood impact rather than recurrence interval. The associated annual exceedance probability for the minor, moderate and major floods at N7 are 15.7%, 3.1% and 0.95%, respectively (calculated using the complete daily streamflow record period). Potential evapotranspiration was obtained from the Australian Data Archive for Meteorology database. Daily values were estimated by assuming a uniform daily distribution within a month.

Three satellite products were used here. The first was the Advanced Microwave Scanning Radiometer - Earth Observing System (AMS hereafter) version 5 VUA-NASA Land Parameter Retrieval Model Level 3 gridded product (Owe et al., 2008). AMS uses C- (6.9 GHz) and X-band (10.65 and 18.7 GHz) radiance observations to derive near-surface soil moisture (2 to 3 cm depth) using a land-surface radiative transfer model. The product used is in units of volumetric water content ($m^3\ m^{-3}$) and has a regular grid of 0.25°.

The second product was the TU-WIEN (Vienna University of Technology) ASCAT (ASC hereafter) data produced using the change-detection algorithm (Water Retrieval Package, version 5.4) (Naeimi et al., 2009). ASC transmits electromagnetic waves in C-band (5.3Gz) and measures the backscattered microwave signal. The change-detection algorithm assumes that land surface characteristics are relatively static over long time periods. Based on this, the differences between instantaneous backscatter coefficients and the historical highest and lowest values for a given incident angle, are related to changes in soil moisture (Wagner et al., 1999). The final SM estimate is provided in relative terms as the degree of saturation and has a nominal spatial resolution varying from 25 to 50 km.

The third satellite product was the Soil Moisture and Ocean Salinity satellite (SMO hereafter), version RE01 (Reprocessed 1-day global soil moisture product) SM provided by the Centre Aval de Traitement des Donnees. SMO uses L-band (1.4 GHz) detectors to measure microwave radiation emitted from depth of up to approximately 5 cm. Near-surface soil moisture is obtained in units of volumetric water content ($m^3\ m^{-3}$) at a spatial resolution of approximately 43 km by using the forward physical model inversion described by Kerr et al. (2012). The overpass times of the AMS, ASC and SMO satellites over the study catchment are 1.30am/pm, 10am/pm and 6am/pm local time (UTC+10h), respectively. Figure 2 summarises the period of record of the different datasets.

For each satellite dataset, a daily averaged SM was calculated for the complete catchment (or sub-catchment in the case of the semi-distributed scheme). The areal estimate of satellite SM over the catchment was given by averaging the values of ascending and descending satellite passes on days when more than 50% of the pixels had valid data. For the case of the passive sensors (AMS and SMO), we subtracted the long-term temporal mean of the ascending and descending datasets to remove the systematic bias between them (Brocca et al., 2011; Draper et al., 2009). Then, daily satellite SM was calculated as the average between the mean-removed ascending and descending passes (if both were available) or directly as the mean-removed available pass. For ASC retrievals, given the unbiased ascending and descending measurements, daily satellite SM was calculated from the actual ascending and descending values averaged over the catchment.

3 Methods

3.1 Lumped and semi-distributed model schemes

The probability distributed model (PDM) is a conceptual rainfall-runoff model that has been widely used in hydrologic research and applications (Moore, 2007), mainly over temperate and humid environments. The model was selected from amongst the set of models available within the flood forecasting system managed by the Australian Bureau of Meteorology. This selection was based on both the suitability of PDM to simulate ephemeral rivers (Moore and Bell, 2002) and preliminary analysis comparing PDM against other models such as the Sacramento soil moisture accounting model, which did not perform as well as PDM.

PDM is a parsimonious model where the runoff production is controlled by the absorption capacity of the soil (in-

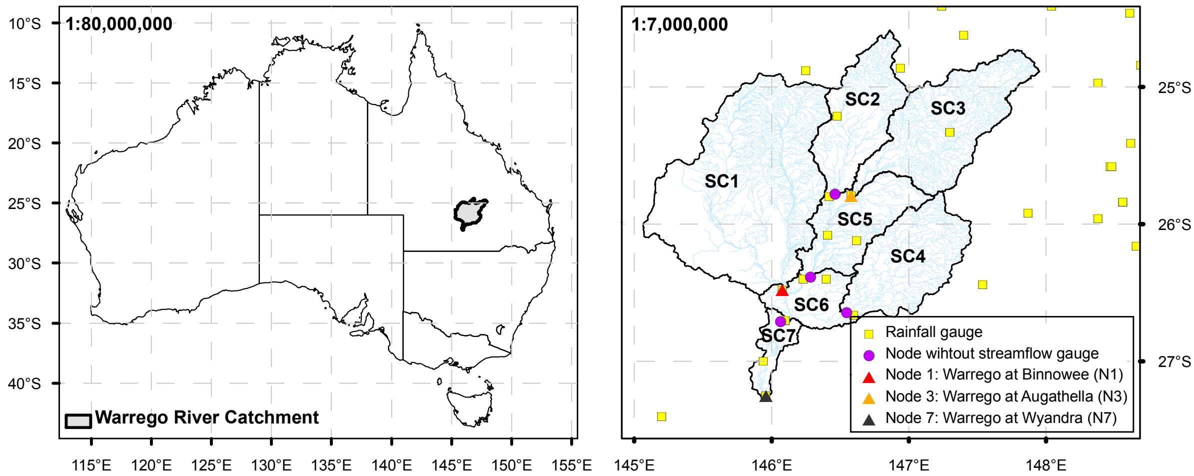

Fig. 1. The Warrego river basin located in Queensland, Australia (left panel). A close-up of the area is presented on the right panel. The lumped PDM scheme is set up over the entire catchment, while the semi-distributed scheme divides the total catchment in 7 sub-catchments (SC1 to SC7).

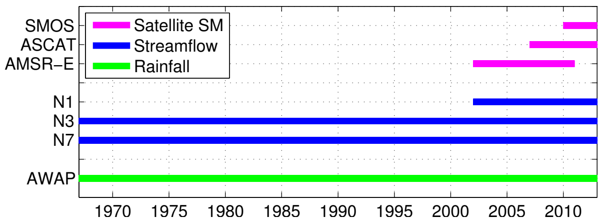

Fig. 2. Periods of record of the different datasets. The initial date of the plot was set as the beginning of the streamflow data record

cluding canopy and surface detention). This process is conceptualised by a store with a distribution of capacities across the catchment and the spatial distribution of these capacities is described by a probability distribution (Moore, 2007). The spatial variability of store capacities can be related to different soil depths, which was identified as the most dominant factor governing runoff variability in a semi-arid catchment (Jothityangkoon et al., 2001).

In the current formulation, the model treats soil moisture store (S_1 in Fig.3) over the entire catchment as a distributed variable with capacities (c) following a Pareto distribution function, $F(c)$. At a given time, the different stores receive water from rainfall and lose water by evaporation and groundwater recharge (drainage). The shallower stores with less capacity than a critical capacity, C^*, start to generate direct runoff while the rest accumulates the water as soil moisture. The proportion of the catchment that generates runoff can therefore be expressed in terms of the Pareto density function, $f(c)$, as

$$\mathrm{prob}\,(c \le C^*) = F\,(C^*) = \int_0^{C^*} f(c)dc. \qquad (1)$$

In this way, for a time t, the soil moisture over the entire catchment, $\boldsymbol{\theta}$ (water content of S_1), can be expressed as the summation of all the store capacities greater than $C^*(t)$:

$$\boldsymbol{\theta}(t) = \int_0^{C^*(t)} (1 - F(c))\,dc. \qquad (2)$$

Note that the critical capacity C^* varies in a time interval Δt based on the net rainfall rate during that time, P,

$$C^*(t + \Delta t) = C^*(t) + P\Delta t. \qquad (3)$$

Direct runoff is calculated based on Eq. 1 and routed through two cascade of reservoirs (S_{21} and S_{22} in Fig.3, with time constants k_1 and k_2, respectively). Subsurface runoff is estimated based on the drainage from S_1 and transformed into baseflow by using a storage reservoir (S_3 in Fig.3 with time constant k_b). These are then combined as total runoff, or streamflow. A detailed description of the model conceptualisation and the formulation of the different rainfall-runoff processes is presented in Moore (2007).

PDM was set up using both a lumped scheme and a semi-distributed scheme (see Fig.1). The semi-distributed scheme was configured with 7 sub-catchments (SC1 to SC7), each using the lumped version of PDM. The area and mean annual rainfall of each sub-catchment are summarised in Table 1. The river routing between upstream and downstream sub-catchments in the semi-distributed scheme was represented by a linear Muskingum method (Gill, 1978):

$$S = k_m\,(Ix + (1 - x)O)\,, \qquad (4)$$

where S is the storage within the routing reach, k_m is the storage time constant, I and O are the streamflow at the beginning and end of the reach, respectively, and x is a weighting factor parameter. The time constant parameters of the

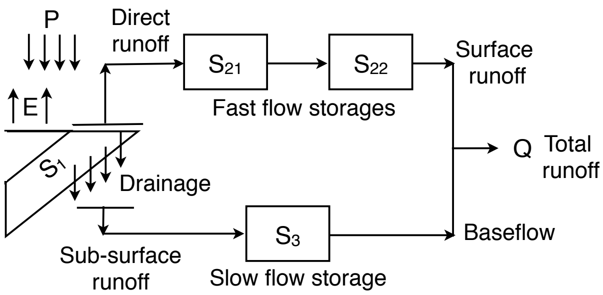

Fig. 3. The PDM scheme

storages S_{21}, S_{22} and S_3 (k_1, k_2 and k_b, respectively) were scaled by the area of each sub-catchment, and k_m from the Muskingum routing was scaled by the length of the river channel between corresponding nodes. The remaining model and routing parameters of the semi-distributed scheme were treated as homogeneous.

Table 1. Area and mean annual rainfall of the catchments used in the lumped and semi-distributed schemes.

Catchment	Area (km^2)	Mean annual rainfall (mm)
SC1	14,670	492
SC2	4,453	532
SC3	8,070	596
SC4	5,431	524
SC5	4,067	503
SC6	2,130	467
SC7	4,049	418
Total	42,870	512

The lumped and the semi-distributed models were calibrated by using a genetic algorithm (Chipperfield and Fleming, 1995) with an objective function based on the Nash-Sutcliffe model efficiency (NSE) (Nash and Sutcliffe, 1970). The models were calibrated for the period 01 January 1967 - 31 May 2003 and evaluation performed for the period 01 June 2003 - 02 March 2014. To make fair comparisons between the two model setups in a scenario where the inner catchments are ungauged, the semi-distributed scheme was calibrated using only the outlet gauge (N7 in Fig.1). The performance of the calibrated models was evaluated based on the NSE at the catchment outlet (N7, Fig.1) and at inner nodes N1 and N3, in the case of the semi-distributed scheme.

To analyse the runoff mechanisms simulated by the lumped and the semi-distributed schemes, we calculated the lag-correlation between rainfall and streamflow, and between antecedent SM and streamflow. This enables further understanding of the improvement in streamflow that can be expected by improving the simulated SM content through SM-DA.

3.2 EnKF formulation

The ensemble Kalman filter (EnKF) proposed by Evensen (2003) has been widely used in hydrologic applications given the nonlinear nature of runoff processes. In the EnKF, the error covariance between the model and observations is calculated from Monte Carlo-based ensemble realisations. In this way, the model and observation uncertainties are propagated and the streamflow prediction is treated as an ensemble of equally likely realisations. The uncertainty of the streamflow prediction can be derived from the ensemble, which provides valuable information for operational flood alert systems.

In a state-updating assimilation approach, the state ensemble is created by perturbing forcing data, parameters and/or states of the model with unbiased errors. As we will see in Sect. 3.3, an N-member ensemble of model soil moisture, $\boldsymbol{\theta} = \{\theta_1, \theta_2, ...\theta_N\}$, was generated by perturbing rainfall forcing data, the model parameter k_1, and $\boldsymbol{\theta}$. Then, the soil water error of member i at time t was estimated as

$$\theta_i^-(t)' = \theta_i^-(t) - \frac{1}{N}\sum_{i=1}^{N}\theta_i^-(t), \tag{5}$$

where the superscript "$-$" denotes the state prediction prior to the assimilation step. The error vector for time step t was defined as $\boldsymbol{\theta}^-(t)' = \{\theta_1^-(t)', \theta_2^-(t)', ..., \theta_N^-(t)'\}$ and the error covariance of the model state (P^-) was estimated at each time step as:

$$P^-(t) = \frac{1}{N-1}\boldsymbol{\theta}^-(t)' \cdot \left(\boldsymbol{\theta}^-(t)'\right)^T. \tag{6}$$

When a daily SM observation from AMS, ASC or SMO was available, each member of the background prediction $(\boldsymbol{\theta}^-)$ was updated. Before being assimilated, each of the three observation datasets was transformed to represent a profile SM and then rescaled to remove systematic differences between the model and the transformed observations (details in Sects. 3.5 and 3.6). We sequentially assimilated an N-member ensemble of the transformed and rescaled AMS, ASC and SMO (named $\boldsymbol{\theta}^{ams}$,$\boldsymbol{\theta}^{asc}$ and $\boldsymbol{\theta}^{smo}$, respectively) and updated each member of $\boldsymbol{\theta}^-$ with the following 3 steps:

1. If $\boldsymbol{\theta}^{ams}$ was available at time t,

   $$\theta_i^+(t) = \theta_i^-(t) + K_1(t)\cdot(\theta_i^{ams}(t) - H\theta_i^-(t)), \tag{7}$$

   where H is an operator that transforms the model state to the measurement space. Since the additive and multiplicative biases between the model predictions and the microwave retrievals were removed via rescaling in a separate step (see Section 3.6), H reduced to a unit matrix. The Kalman gain $K_1(t)$ was calculated as

   $$K_1(t) = \frac{P^-(t)H^T}{HP^-(t)H^T + R_1(t)}, \tag{8}$$

   where $R_1(t)$ is the error variance of $\boldsymbol{\theta}^{ams}$ estimated in the rescaling procedure (Sect. 3.6). If $\boldsymbol{\theta}^{ams}$ was not available, $\boldsymbol{\theta}^+(t) = \boldsymbol{\theta}^-(t)$.

2. If $\boldsymbol{\theta}^{asc}$ was available at time t, we updated the model soil moisture with

$$\theta_i^{++}(t) = \theta_i^{+}(t) + K_2(t)\cdot(\theta_i^{asc}(t) - H\theta_i^{+}(t)), \tag{9}$$

where $K_2(t)$ was calculated as

$$K_2(t) = \frac{P^{-}(t)H^T}{HP^{-}(t)H^T + R_2(t)}. \tag{10}$$

$R_2(t)$ is the error variance of $\boldsymbol{\theta}^{asc}$ and P^- is the model error covariance re-calculated by applying Eq.(6) to the updated soil moisture $\boldsymbol{\theta}^{+}(t)$. If $\boldsymbol{\theta}^{asc}$ was not available, $\boldsymbol{\theta}^{++}(t) = \boldsymbol{\theta}^{+}(t)$.

3. If $\boldsymbol{\theta}^{smo}$ was available at time t, we updated the model soil moisture with

$$\theta_i^{+++}(t) = \theta_i^{++}(t) + K_3(t)\cdot(\theta_i^{smo}(t) - H\theta_i^{++}(t)), \tag{11}$$

where $K_3(t)$ was calculated as

$$K_3(t) = \frac{P^{-}(t)H^T}{HP^{-}(t)H^T + R_3(t)}. \tag{12}$$

$R_3(t)$ is the error variance of $\boldsymbol{\theta}^{smo}$ and P^- is the model error covariance re-calculated by applying Eq.(6) to the updated soil moisture $\boldsymbol{\theta}^{++}(t)$. If $\boldsymbol{\theta}^{smo}$ was not available, $\boldsymbol{\theta}^{+++}(t) = \boldsymbol{\theta}^{++}(t)$.

In the case of the semi-distributed scheme, during the updating steps described above, each sub-catchment was treated independently and no spatial cross-correlation in the satellite measurements was considered. The order of the products assimilated in steps 1 to 3 was arbitrary; however, we checked that different orders did not significantly affect the SM-DA results.

3.3 Error model representation

The main sources of uncertainty in hydrologic models are the errors in the forcing data, the model structure and the incorrect specification of model parameters (Liu and Gupta, 2007). Generally, these errors are represented by adding unbiased synthetic noise to forcing variables, model state variables and/or model parameters.

The estimation of model errors is among the most crucial challenges in data assimilation, as it determines the value of the Kalman gain. In the case of a state updating SM-DA, the ability of the scheme to improve streamflow prediction will mainly depend on the covariance between the errors in SM states and modelled streamflow, which directly depends on the specific representation and estimation of the model errors.

To represent the forcing uncertainty, we adopted a multiplicative error model for the rainfall data (McMillan et al., 2011; Tian et al., 2013). In particular, we followed the scheme used in various SM-DA studies (e.g., Chen et al., 2011; Brocca et al., 2012; Alvarez-Garreton et al., 2014) and represented a spatially homogeneous rainfall error (ϵ_p) as

$$\epsilon_p \sim lnN(1,\sigma_p^2), \tag{13}$$

where σ_p is the standard deviation of the lognormal distribution. The above representation assumes a spatially homogeneous fraction of the error to the rainfall intensity, which could be an over simplification in a large area like the study catchment. However, it avoids the estimation of additional error parameters (e.g., spatial correlation parameter) in an already highly undetermined problem (see Sect. 3.4).

The parameter uncertainty was represented by perturbing the time constant parameter (k_1) for store S_{21}, a highly sensitive parameter of the model that directly affects the streamflow generation by influencing the water stored in both surface storages S_{21} and S_{22} (note that in the PDM formulation used, the time constant k_2 is calculated as a function of k_1). Given the lack of prior information about the structure of the parameter error (ϵ_k), we adopted a normally distributed multiplicative error with unit mean and standard deviation of σ_k, following previous SM-DA studies working with rainfall-runoff models (Brocca et al., 2010b, 2012).

Following the scheme used in most SM-DA experiments (e.g., Reichle et al., 2008; Crow and Van den Berg, 2010; Chen et al., 2011; Hain et al., 2012), the model structural error was represented by perturbing the SM prediction ($\boldsymbol{\theta}$) with a spatially homogeneous additive random error,

$$\epsilon_s \sim N(0,\sigma_s^2), \tag{14}$$

where σ_s is the standard deviation of the normal distribution.

The physical limits of SM (porosity as an upper bound and residual water content as a lower bound) are represented by the model through the storage capacity of S_1. When $\boldsymbol{\theta}$ approaches the limits of S_1, applying unbiased perturbation to $\boldsymbol{\theta}$ can lead to truncation bias in the background prediction. This can then result in mass balance errors and degrade the performance of the EnKF (Ryu et al., 2009). Moreover, the Kalman filter assumes unbiased state variables. This issue is of particular importance in arid regions like the study area, where the soil water content can be rapidly depleted by evapotranspiration and transmission losses, thus approaching the residual water content of the soil. To ensure that the state ensemble remained unbiased after perturbation we implemented the bias correction scheme proposed by Ryu et al. (2009).

The truncation bias correction consisted of running a single unperturbed model prediction (θ^{-0}) in parallel with the perturbed model prediction ($\theta_{i,}^{-}$). At each time step, the mean bias, $\delta(t)$, of the N-member ensemble prediction was calculated by subtracting $\theta^{-0}(t)$ from the ensemble mean, as follows (Ryu et al., 2009):

$$\delta(t) = \frac{1}{N}\sum_{i=1}^{N}\theta_i^{-}(t) - \theta^{-0}(t). \tag{15}$$

Then, a bias corrected ensemble of state variables, $\tilde{\theta}_i^-(t)$, was obtained by subtracting $\delta(t)$ from each member of the perturbed ensemble, $\theta_i^-(t)$.

Although the latter resulted in unbiased state ensembles, some important but subtle effects remain that arise from the highly non-linear nature of hydrologic model. These need to be guarded against in SM-DA. Representing model errors by adding unbiased perturbation to forcing, model parameters and/or model states can lead to a biased streamflow ensemble prediction (e.g., Ryu et al., 2009; Plaza et al., 2012), compared with the unperturbed model run. This biased streamflow ensemble prediction (open-loop hereafter) is degraded compared with the streamflow predicted by the unperturbed calibrated model. As a consequence, improvement of the open-loop after SM-DA will in part be due to the correction of bias introduced during the assimilation process itself.

To avoid overstating the SM-DA efficacy due to the above issue, we applied the bias correction scheme directly to the streamflow prediction (in both the open-loop and the assimilation runs). We used the unperturbed model run to estimate a mean bias in the streamflow (following Eq. 15, but using streamflow instead of soil moisture) and then corrected each ensemble member by subtracting this mean bias. This practical tool ensures that the streamflow ensemble mean maintains the performance skill of the unperturbed (calibrated) model run, thus avoiding artificial degradation of the unperturbed model run by bias. To our knowledge, this approach has not been applied in SM-DA previous studies.

3.4 Error model parameters calibration

To calibrate the error model parameters (σ_p, σ_k and σ_s), we evaluated the open-loop ensemble prediction (Q^{ol}) against the observed streamflow at the catchment outlet. In this study we used a maximum a posteriori (MAP) scheme, a Bayesian inference procedure detailed by Wang et al. (2009) that maximises the probability of observing historical events given the model and error parameters. In other words, it maximises the probability of having the streamflow observation within the open-loop streamflow.

Member i from the N-member open-loop can be expressed as

$$Q_i^{ol}(t) = Q^T(t) + \epsilon_m(t), \tag{16}$$

where Q^T is the (unknown) truth streamflow and ϵ_m is the error of the streamflow prediction and consists of forcing, parameter and states errors:

$$\epsilon_m(t) = f(\epsilon_p(t), \epsilon_k(t), \epsilon_s(t)). \tag{17}$$

The observed streamflow at N7 (Q_{obs}) can be expressed as a function of the same (unknown) truth and the streamflow observation error (ϵ_{obs}),

$$Q_{obs}(t) = Q^T(t) + \epsilon_{obs}(t). \tag{18}$$

Combining Eqs. 16 and 18, the model ensemble prediction of the observed streamflow ($\hat{Q}_{obs}$) is expressed as:

$$\hat{Q}_{obs}(t) = Q^{ol}(t) + \epsilon_m(t) + \epsilon_{obs}(t). \tag{19}$$

Following Li et al. (2014), ϵ_{obs} was assumed to be a serially independent multiplicative error following a normal distribution (mean 1 and standard deviation of 0.2). Then, the likelihood function (L) defining the probability of observing the historical streamflow data given the calibrated model parameters (x), and the error model parameters (σ_p, σ_k and σ_s), was expressed as

$$L(Q_{obs}|x, \sigma_p, \sigma_k, \sigma_s) = \Pi_{t=1}^{n} p(Q_{obs}(t)|\hat{Q}_{obs}(t)). \tag{20}$$

To maximise L, we applied a logarithm transformation to it and maximised the sum over time of the transformed function. The probability density function (p) at each time step was estimated by assuming that the ensemble prediction of the observed streamflow, $\hat{Q}_{obs}(t)$, follows a Gaussian distribution, with its mean and standard deviation computed using the ensemble members. The period used to calibrate the error model parameters was 01 January 1998 - 31 May 2003.

An important aspect to highlight about this error parameter calibration is that it is a highly underdetermined problem. Only one data set (streamflow at N7) is used to calibrate the error parameters, while there might be many combinations of error parameters that can generate similar streamflow ensemble (equifinality on the error parameters).

3.5 Profile soil moisture estimation

The aim of the stochastic assimilation detailed in Sect. 3.2 is to correct $\boldsymbol{\theta}$, which is a profile average SM representing a soil layer depth determined by calibration. By assuming a porosity of 0.46, (A-horizon information reported in McKenzie et al. (2000)), and the model S_1 storage capacity of 396 mm (420 mm) for the lumped (semi-distributed) scheme, this profile SM roughly represents the upper 1 m of the soil. On the other hand, the satellite SM observations represent only the few top centimetres of the soil column (see Sect. 2). To provide the model with information about more realistic dynamics of $\boldsymbol{\theta}$, we applied the exponential filter proposed by Wagner et al. (1999) to the satellite SM to estimate the soil wetness index (SWI) of the root-zone. SWI has been widely used to represent deeper layer SM based on satellite observations (e.g., Albergel et al., 2008; Brocca et al., 2009, 2010b, 2012; Ford et al., 2014; Qiu et al., 2014). SWI was recursively calculated as:

$$\mathrm{SWI}(t) = \mathrm{SWI}(t-1) + G(t)\left[\mathrm{SSM}(t) - \mathrm{SWI}(t-1)\right], \tag{21}$$

where $\mathrm{SSM}(t)$ is the satellite SM observation and $G(t)$ is a gain term varying between 0 and 1 as:

$$G(t) = \frac{G(t-1)}{G(t-1) + e^{-\left(\frac{t-(t-1)}{T}\right)}}. \tag{22}$$

T is a calibrated parameter that implicitly accounts for several physical parameters (Albergel et al., 2008). T was calibrated by maximising the correlation between SWI and the unperturbed model soil moisture (θ) during the first year of satellite data. This calibration period was selected to maximise the independent evaluation period (see Section 3.7); however, more representative values are likely to be obtained if a longer period was used for calibration. SWI was calculated independently for each of the AMS, ASC and SMO datasets (named $\mathrm{SWI}_{\mathrm{AMS}}$, $\mathrm{SWI}_{\mathrm{ASC}}$ and $\mathrm{SWI}_{\mathrm{SMO}}$, respectively) and then rescaled to remove systematic differences with the model prediction (Sect. 3.6).

3.6 Rescaling and observation error estimation

The systematic differences (e.g., biases) between θ and the SWI derived from each satellite product must be removed prior to applying a bias-blind data assimilation scheme (Dee and Da Silva, 1998). We applied instrumental variable (IV) regression to resolve the biases and estimate the measurement errors simultaneously (Su et al., 2014a). In three-data IV regression analysis, also known as triple collocation (TC) analysis (Stoffelen, 1998; Yilmaz and Crow, 2013), the model θ, the passive SWI and active SWI are used as the data triplet. As the sample size requirement for TC is stringent (Zwieback et al., 2012), a pragmatic threshold of 100 triplet sample was imposed (Scipal et al., 2008). During periods when only one satellite product was available (i.e., before ASC) or when the sample threshold for TC was not met, a two-data set IV regression using lagged variables (LV) was applied as a practical substitute (Su et al., 2014a). The LV analysis was performed on the model θ and a single satellite SWI, with the lagged variable coming from the model.

In most SM-DA experiments, the error in satellite SM has been treated as time-invariant (e.g., Reichle et al., 2008; Ryu et al., 2009; Crow and Van den Berg, 2010; Brocca et al., 2010b, 2012; Alvarez-Garreton et al., 2014); however, studies evaluating satellite SM products have shown an important temporal variability in the measurement errors (Loew and Schlenz, 2011; Su et al., 2014a). Since a data assimilation scheme explicitly updates the model prediction based on the relative weights of the model and the observation errors, assuming a constant observation error may lead to overcorrection of the model state if the actual error is higher, and vice versa.

Temporal characterisation of the observation error can be achieved by applying TC (or LV) to specific time windows of the observations and model predictions (for example, by grouping the triplets or doublets by month-of-the-year). There is however, a trade-off between the sampling window (which defines the temporal characterisation of the error) and the sample size (number of triplets in each subset). In an operational context this trade-off becomes more critical since only past observations are available. After analysing the temporal variability of the observation errors using the complete period of record (not shown here), we found that a 4-month sampling window can reproduce seasonality in errors while ensuring sufficient data samples for the TC and LV schemes. With this analysis we also assessed the suitability of using LV, which yielded similar results to TC although some positive bias in LV error variance estimates relative to TC was noted (not shown here).

Summarising, the procedure for rescaling and error estimation consists of:

1. From the start of the AMS dataset, we grouped LV triplets ($\mathrm{SWI}_{\mathrm{AMS}}(t)$, $\theta(t)$ and $\theta(t-1)$) into three subsets: Dec-Mar, Apr-Jul and Aug-Nov.

2. We applied LV and thus, estimated the observation error variance and rescaling factors for a given 4-month subset only when a minimum of 100 samples was reached (after one year of AMS dataset). After the first year of AMS, new seasonal triplets were added into the corresponding 4-month data pool (retaining all earlier triplets) and LV was applied to the updated subset.

3. When ASC was available, LV triplets ($\mathrm{SWI}_{\mathrm{ASC}}(t)$, $\theta(t)$ and $\theta(t-1)$) subsets were formed following step 1 criteria and LV was applied after the 4-month data pools had more than 100 samples, following step 2.

4. In parallel with step 3, TC triplets were formed using the two available satellite datasets ($\mathrm{SWI}_{\mathrm{AMS}}(t)$, $\mathrm{SWI}_{\mathrm{ASC}}(t)$ and $\boldsymbol{\theta}(t)$) and grouped into the 4-month subsets defined in step 1. TC was applied only when the 4-month data pools contained more than 100 samples (after approximately 3 years of ASC data).

5. Steps 3 and 4 were repeated when SMO was available. The triplets for TC in this case were given by $\mathrm{SWI}_{\mathrm{ASC}}(t)$, $\mathrm{SWI}_{\mathrm{SMO}}(t)$ and $\theta(t)$.

6. Once steps 1-5 were complete, a single time series of observations error variance and rescaling factors was constructed for each satellite-derived SWI by selecting TC results when available, and LV results if not. This criterion was adopted because LV is susceptible to bias due to auto-correlated errors in the model SM (Su et al., 2014a). The rescaled observations from AMS, ASC and SMO were named θ^{ams}, θ^{asc} and θ^{smo}, respectively.

3.7 Evaluation metrics

To evaluate the SM-DA results, we used six different metrics. Firstly, the normalised root mean squared difference (NRMSE) was calculated as the ratio of the root mean square error (RMSE) between the updated streamflow ensemble (Q^{up}) and the observed streamflow to the RMSE between the open-loop (ensemble streamflow prediction without assimilation, Q^{ol}) and the observed discharge:

$$\mathrm{NRMSE} = \frac{\frac{1}{N}\sum_{i=1}^{N}\sqrt{\sum_{t=1}^{T}\left(Q_i^{up}(t)-Q_{obs}(t)\right)^2}}{\frac{1}{N}\sum_{i=1}^{N}\sqrt{\sum_{t=1}^{T}\left(Q_i^{ol}(t)-Q_{obs}(t)\right)^2}}, \tag{23}$$

where $N = 1000$ is the number of ensemble members. The NRMSE provides information about both the spread of the ensemble and the performance the ensemble mean, which is considered as the best estimate of the ensemble prediction. Moreover, as it is calculated in linear streamflow space, it gives more weight to high flows.

To further evaluate the performance of the ensemble mean, we calculated the Nash Sutcliffe efficiency (NSE) for the entire evaluation period as follows (example for the open-loop case):

$$\mathrm{NSE_{ol}} = 1 - \frac{\sum_t \left(Q_{obs}(t) - \overline{Q^{ol}}(t)\right)^2}{\sum_t \left(Q_{obs}(t) - \overline{Q_{obs}}\right)^2}, \tag{24}$$

where $\overline{Q^{ol}}$ is the open-loop ensemble mean. Similarly, $\mathrm{NSE_{up}}$ was calculated by applying Eq.(24) to the updated ensemble mean ($\overline{Q^{up}}$).

We also estimated the probability of detection (POD) of daily flow rates (not flood events) exceeding minor, moderate and major floods, for the open-loop and the updated ensemble mean, as follows (example for the open-loop case):

$$\mathrm{POD_{ol}} = \frac{\#(\overline{Q^{ol}} >= Q_{obs}^{15.7\%} \ \&\ Q_{obs} >= Q_{obs}^{15.7\%})}{\#(Q_{obs} >= Q_{obs}^{15.7\%})}, \tag{25}$$

where the symbol $\#$ represents the number of times. $Q_{obs}^{15.7\%}$ is the observed streamflow corresponding to a minor flood classification. This corresponds to a flow (not flood) frequency of 15.7% (see Sect. 2). Similarly, $\mathrm{POD_{up}}$ was calculated by applying Eq.(25) to the updated ensemble mean ($\overline{Q^{up}}$). We estimated the false alarm ratio (FAR) for daily flows as (example for the open-loop case):

$$\mathrm{FAR_{ol}} = \frac{\#(\overline{Q^{ol}} >= Q_{obs}^{15.7\%} \ \&\ Q_{obs} < Q_{obs}^{15.7\%})}{\#(Q_{obs} < Q_{obs}^{15.7\%})}. \tag{26}$$

Similarly, $\mathrm{FAR_{up}}$ was calculated by applying Eq.(26) to the updated ensemble mean.

Finally, we calculated the aggregated peak volume error (PVE, in mm) of the ensemble mean, for days when the observed streamflow was above a minor flood classification (t^* days in Eq. 27). An example for the open-loop, PVE was calculated as

$$\mathrm{PVE_{ol}} = \sum_{t^*} \left(\overline{Q^{ol}}(t^*) - Q_{obs}(t^*)\right). \tag{27}$$

To evaluate the skill of the streamflow ensemble prediction before and after SM-DA, we calculated the continuos ranked probability score (CRPS; Robertson et al., 2013). CRPS is used as a measure of the ensemble errors. In the case of the deterministic unperturbed run, CRPS reduces to the mean absolute error. The reliability of the ensembles was also evaluated by inspecting the rank histograms of the ensemble following Anderson (1996). A reliable ensemble should have a uniform histogram while a u-shape (n-shape) histogram indicates that the ensemble spread is too small (large) (De Lannoy et al., 2006).

The evaluation period for the SM-DA was 01 June 2003 - 02 March 2014. This period is independent of all scheme component calibration periods (see Sects. 3.1, 3.4 and 3.5).

4 Results and discussion

4.1 Model calibration

The streamflow at the outlet of the study catchment (N7 in Fig.1) features long periods of zero-flow, a negligible baseflow component and sharp flow peaks after rainfall events, when the catchment has reached a threshold level of wetness (see observed streamflow in Fig.4).

The simulated streamflows from the lumped and the semi-distributed schemes are presented in Fig.4. To help visualisation of these time series, the calibration and evaluation periods were plotted separately. The evaluation period was further separated into two sub-periods, evaluation sub-period 1 (01 June 2003 - 30 April 2007), characterised by having only moderate and minor floods, and evaluation sub-period 2 (30 April 2007 - 02 March 2014), which had three major flooding events. The plots show that both the lumped and the semi-distributed models are generally able to capture the hydrologic behaviour of the catchment. As expected, the spatial distribution of forcing data and the channel routing accounted for by the semi-distributed scheme enhanced the overall performance of the model, with lower residual values through time (panels a.2, b.2 and c.2 in Fig.4) and consistently improved the simulation of peak flows.

Table 2 presents the evaluation statistics for the streamflow prediction in the calibration and evaluation periods, for both the catchment outlet and the inner catchments (notice that N1 does not have data in the calibration period). The different statistics in this table consistently show that, at the catchment outlet, the semi-distributed has consistently better performance than the lumped scheme in terms of RMSE, NSE, PEV and CRPS. Both schemes show better statistics in the evaluation period due to the higher flows over that period.

The good performance of the semi-distributed scheme at the catchment outlet was not reflected at the inner catchments. To explore the reasons for such bad performance, we separately calibrated the model parameters in those sub-catchments by using all the available N7, N1 and N3 observations. The results (not shown here) revealed that in this case, the model was able to adequately simulate streamflow in those sub-catchments (NSE in evaluation period of 0.78, 0.69 and 0.84 at N1, N3 and N7 nodes, respectively). Based on this, we argue that the problem of the poor model performance in the “ungauged” inner catchments is most likely due to sub-optimal parameter estimation (due to the limited information about catchment heterogeneity provided by the integrated catchment streamflow response) and unlikely to be due to errors in the input data or model structure.

To focus the analysis of the catchment runoff mechanisms on periods with flood events, the lag-correlation between the daily streamflow simulated at N7 and $\boldsymbol{\theta}$ (Fig.5), and between daily streamflow and the daily rainfall (Fig.6), was calculated for daily streamflow values greater than $Q_{obs}^{15.7\%}$, or minor

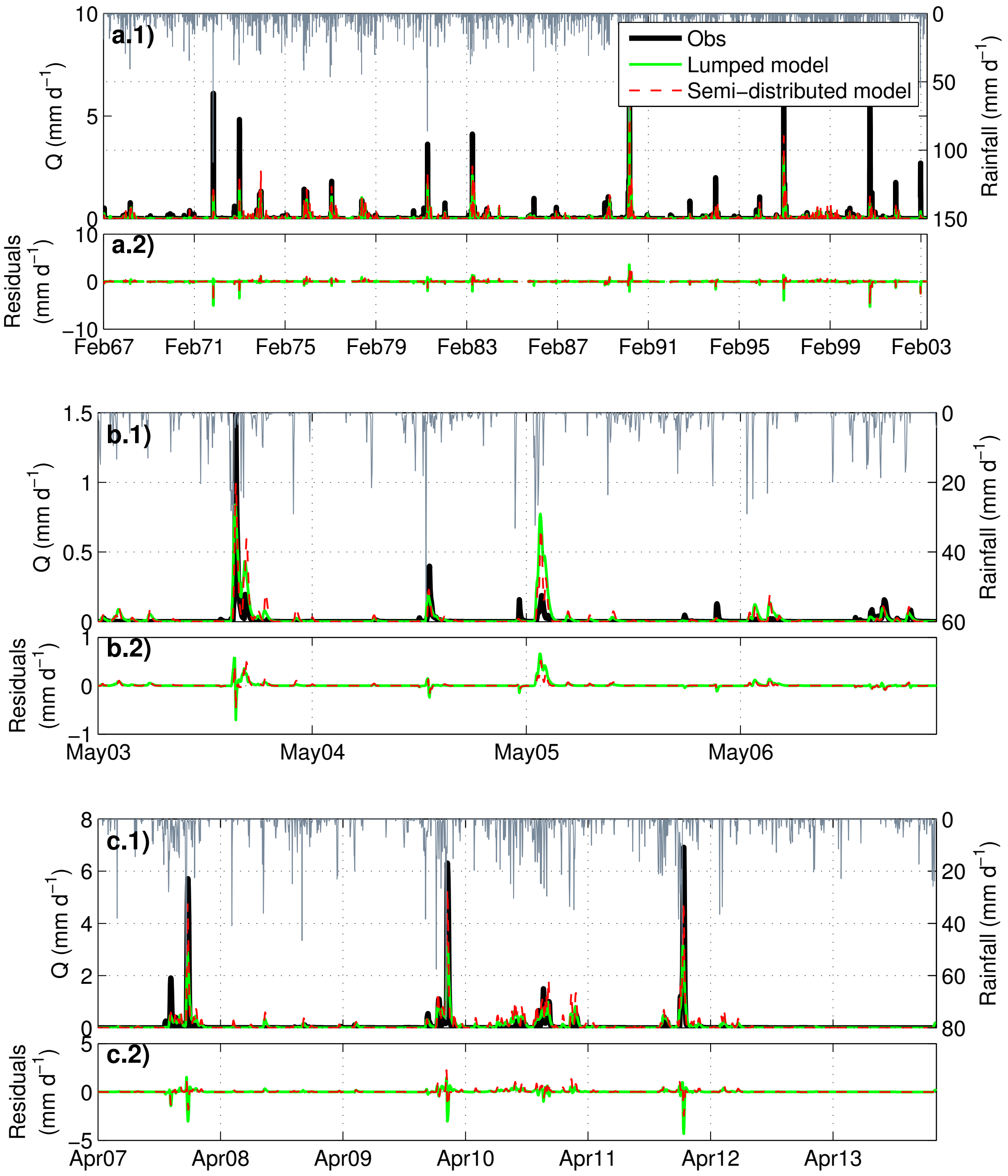

Fig. 4. Simulated and observed daily streamflow (Q) and model streamflow prediction residuals (simulated minus observed) at the catchment outlet (N7). (a.1) and (a.2) present the calibration period. (b.1) and (b.2) present evaluation sub-period 1, which has only moderate and minor flood events. (c.1) and (c.2) present evaluation sub-period 2, which has 3 major flood events. The daily rainfall plotted on the right axis correspond to the averaged rainfall over the entire catchment.

flood level. The lumped scheme indicates a stronger link between θ and streamflow than the semi-distributed scheme. This is represented by higher r values in panel a compared with panels b-h in Fig.5. Conversely the link between rainfall and streamflow is weaker in the lumped scheme (lower r values in panel a compared with panels b-h in Fig.6). These different representations of the catchment runoff response will have a direct impact on the skill of SM-DA to improve

Table 2. Model evaluation at the catchment outlet (N7) and at the inner catchments (N1 and N3), for calibration and evaluation periods. RMSE and PVE statistics are in units of mm.

Statistic	Lumped scheme	Semi-distributed scheme		
	(N7)	(N7)	(N1)	(N3)
$RMSE_{calib}$	0.19	0.18	-	0.3
$RMSE_{eval}$	0.21	0.18	0.53	0.46
NSE_{calib}	0.52	0.59	-	0.39
NSE_{eval}	0.67	0.77	0.28	-1.89
POD_{calib}	0.79	0.76	-	0.76
POD_{eval}	0.93	0.91	0.54	0.73
FAR_{calib}	0.09	0.10	-	0.15
FAR_{eval}	0.11	0.11	0.07	0.14
PVE_{calib}	-70.86	-39.99	-	168.23
PVE_{eval}	1.30	34.75	-100.53	115.52
$CRPS_{calib}$	0.29	0.28	-	0.58
$CRPS_{eval}$	0.56	0.33	0.92	0.49

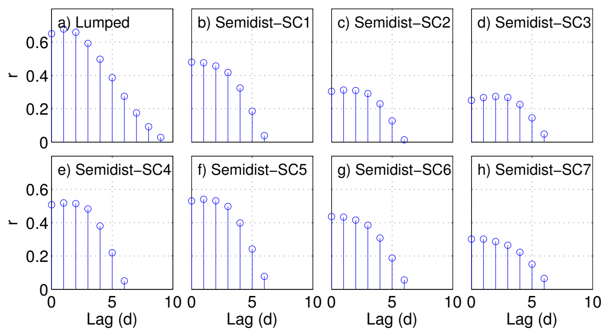

Fig. 5. Lag-correlation coefficient (r) between the simulated streamflow at N7 (mm d^{-1}), and θ (mm d^{-1}) from the lumped (a) and the semi-distributed (b)-(h) model schemes.

streamflow prediction. A strong relationship between θ and streamflow prediction suggests strong correlation between their errors, and therefore, greater potential improvement of streamflow resulting from an improved representation of θ.

If we assume that the semi-distributed scheme provides a better representation of runoff response within the entire catchment (based on its better model performance at the outlet), Figs. 5 and 6 also suggest that daily rainfall is the main control on runoff generation and thus has a stronger impact in the streamflow prediction than soil moisture. Figure 5 shows that flood prediction strongly depends on antecedent soil moisture for up to the preceding 3 days. The strong correlation found at lag-0 suggests that the real time SM correction given by the proposed SM-DA would be a good strategy to improve flood prediction.

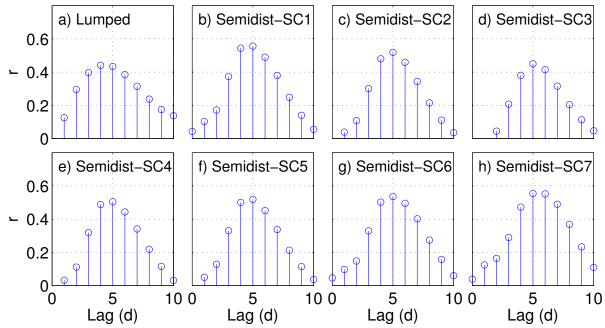

Fig. 6. Lag-correlation coefficient (r) between the simulated streamflow at N7 (mm d^{-1}), and the daily rainfall (mm d^{-1}) of the entire catchment (a) and the 7 sub-catchments (b)-(h).

4.2 Error model parameters and ensemble prediction

The calibrated error parameters for the lumped and the semi-distributed schemes are $\sigma_p = 1.286$ mm and 0.977 mm; $\sigma_s = 0.099$ and 0.03 and $\sigma_k = 0.084$ and 0.018, respectively. σ_s is expressed as a percentage of the total storage capacity (396 mm in the lumped scheme and 420 mm in the semi-distributed scheme) and σ_k is expressed as a percentage of the calibrated parameter k_1.

The rank histograms of the generated ensemble prediction (open-loop) are presented in Fig.7. The histograms at the catchment outlet (N7) are either n-shape or displaced to one side, for both the lumped and semi-distributed model schemes (Figs.7a and 7b, respectively). This suggests that the open-loop ensembles are slightly biased (with respect to the observed streamflow) and feature wider spread than an ideal ensemble. The width of the spread will be critical for the evaluation of SM-DA (Sect. 4.4) since any decrease of the spread would be considered as an improvement of the ensemble prediction.

The wider spread of the open-loop ensembles at the catchment outlet could be due to factors such as an over-prediction of error parameters by the MAP calibration algorithm, or the representation of the model error with time-constant error parameters. The latter becomes critical given the distinct behaviour of the intermittent streamflow response within the catchment, which could indicate distinct behaviour in the model errors as well.

The ensemble predictions at the inner nodes N1 and N3 (Figs.7c and 7d, respectively) feature high bias with respect to the observed streamflow (note that observations at N1 and N3 were not used to calibrate the error parameters). The large bias at these inner nodes result from the large errors in the calibrated model in SC1 and SC3 (see Sect. 4.1).

4.3 SWI estimation and rescaling

The satellite SM derived from AMS, ASC and SMO are presented in Fig.8a, for the lumped model. The satellite datasets

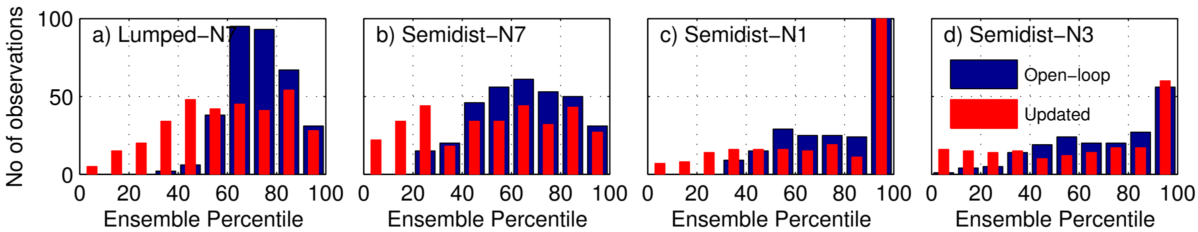

Fig. 7. Rank histograms of the open-loop and updated streamflow ensemble predictions. (a) presents the results from the lumped scheme at node N7. (b)-(d) present the results from the semi-distributed (semidist) scheme at nodes N7, N1 and N3.

feature significantly higher noise than the modelled θ. This can be explained by factors such as random errors in the satellite retrievals (Su et al., 2014b), and the rapid variation of water content in the surface layer of soil due to infiltration and evapotranspiration losses. Figure 8b presents the SWI derived from the satellite products, after seasonal rescaling (θ^{ams}, θ^{asc} and θ^{smo}). This plot shows better agreement between model and observations due to SWI filtering/transformation, even when the higher noise in the rescaled SWI time series is still present.

Figure 8c shows the seasonal observation error variance, and reveals a clear variation in the error with time. The variation of the seasonal error values is due to the alternative use of TC or LV and to the increasing sample size of each seasonal pool (see Section 3.6), which should reduce the uncertainties coming from finite sample size. One limitation of this procedure is its assumption that the errors vary seasonally without inter-annual variability. Since there are inter-annual cycles (wet and dry years), one may also expect the errors to vary with year. Ideally, moving-window estimation with windows smaller than 3 months should be considered, but that would cause greater sampling uncertainties for the TC and LV estimates. The inverse relationships between θ^{ams} and θ^{asc} error variances at some times could be due to the passive retrieval by AMS compared with the active ASC, among other factors.

A common error standard deviation value used in previous SM-DA studies is 3% m^3m^{-3} (e.g., Chen et al., 2011). This constant error, when transformed according to the soil moisture storage capacity of the model and the soil porosity (see Section 3.5) gives an error variance of 667 (750) mm^2 for the lumped (semi-distributed) scheme. As a simple comparison, these values are within the range of the error variance estimated through seasonal LV/TC; however, a comprehensive analysis of the impacts of accounting for seasonality in SM-DA is beyond the scope of this work.

Table 3 summarises the results of the SWI calibration and seasonal rescaling for the lumped model, showing the T parameter for each SWI and the correlation coefficient (r) between θ and the satellite SM before and after SWI transformation and rescaling (θ^{obs}). These results confirm the visual assessment of plots in Fig.8 by showing an important increase in the linear correlation coefficient with θ when satellite SM is transformed into SWI. The correlation is further increased after rescaling, which illustrates that there is clear benefit from performing seasonal bias correction. Note that applying a constant rescaling factor would have no impact on on the correlation between θ and θ^{obs}.

Table 3. Parameter T and correlation coefficient between model SM (θ) and satellite SM, before and after SWI transformation and rescaling. Results are presented for the entire catchment.

Dataset	T (days)	r between $\boldsymbol{\theta}$ and Satellite SM	SWI	θ^{obs}
AMS	3	0.65	0.74	0.94
ASC	11	0.77	0.92	0.97
SMO	40	0.46	0.79	0.93

The optimal T values (Table 3) are difficult to validate since there is no ground data to compare with and, given that it has been shown that they strongly depend on the physical processes of the study site (Ceballos et al., 2005), direct comparison with other studies cannot be made reliably. Indeed, previous studies have shown a wide range of optimal T values for soil depths ranging between 10 and 100 cm. As an example, in Fig.9 we have summarised the optimal T found in 5 different studies (Albergel et al., 2008; Brocca et al., 2009, 2010a; Ford et al., 2014; Wagner et al., 1999).

Previous studies have shown that optimal T value increases with layer depth (e.g., Brocca et al., 2010a). Results presented here show an increased T value for SMO, which would be inconsistent with L-band having a deeper penetration than AMS C-band (to limit the comparison within passive retrievals). We speculate that these differences might be due various factors, including the different retrievals methods (which have quite different assumptions pertaining to spatial heterogeneity) and the influence that radio-frequency interference noise. Moreover, to the best of our knowledge, the existing studies examining the dependence of T on the soil depths are usually based on a single satellite product against

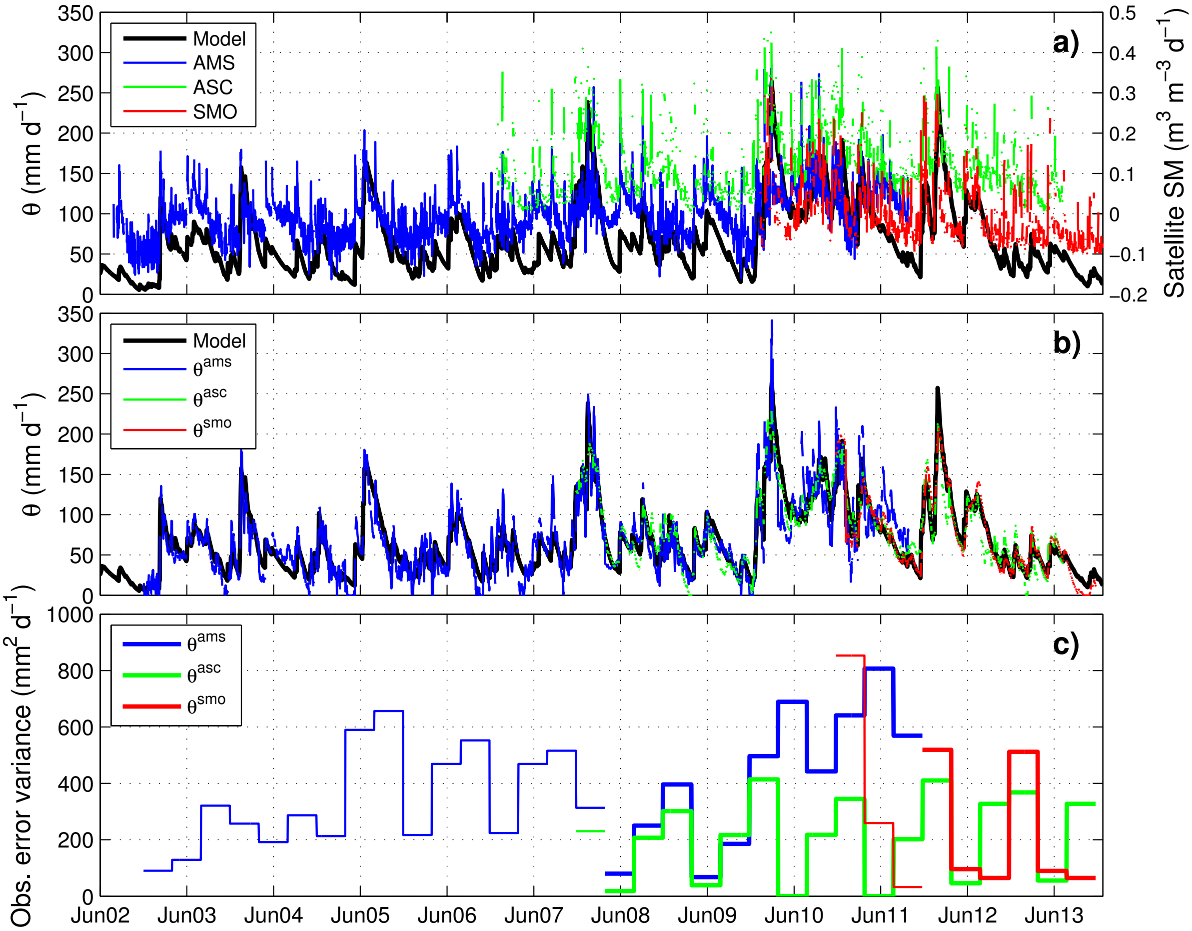

Fig. 8. (a) shows the model soil moisture on the left axis (θ) and the satellite soil moisture observations in the right axis. (b) shows the soil moisture on the model space, after the three satellite datasets were transformed into a soil wetness index (SWI) and then rescaled by using TC or LV (θ^{ams}, θ^{asc} and θ^{smo}). (c) shows the rescaled satellite SM observations error variance.

in situ measurements at variable depths. Hence it is difficult to compare our results against these studies due to the increased complexity due to different sensing and retrieval methods.

There are some key theoretical issues that should be considered when using SWI as a profile SM estimator. Firstly, the parameter T in Eq.(22) was estimated by maximising the correlation between SWI and θ, which could introduce cross-correlated errors between them. This would violate the IV regression assumption of no correlation between the errors among the triplets (Sect. 3.6). A way to overcome this issue, if data requirements are met, would be to estimate a profile SM independently of the rainfall-runoff model prediction, for example by using a physically-based model to transfer surface SM into deeper layers (e.g., Richards, 1931; Beven and Germann, 1982; Manfreda et al., 2014).

Secondly, the SWI formulation explicitly incorporates autocorrelation terms, which would result in autocorrelated errors in the observation, which violates an EnKF assumption: independence between observation and prediction errors. The autocorrelation in the observation error can be transferred to the updated $\boldsymbol{\theta}^{+}$ during the SM-DA updating step. In that case, the $\boldsymbol{\theta}^{-}$ background prediction error covariance at time $t+1$ would be correlated to the error of the rescaled SWI at time $t+1$. In contrast with the first issue

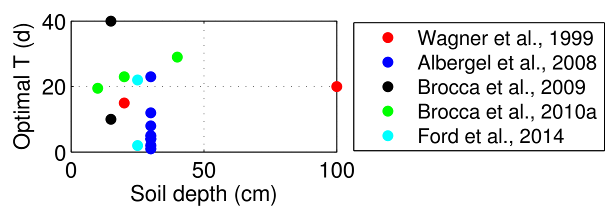

Fig. 9. Optimal T parameter against soil depth found in previous studies.

listed above, the violation of the EnKF assumption can not be avoided by replacing SWI with a physically-based model, since the latter would result in profile SM strongly correlated with previous states as well. Indeed, given the physical mechanisms of water flux in the unsaturated soil, this problem will be present whenever a profile SM estimated from satellite SM is used as an observation in an EnKF-based data assimilation framework. A way to overcome this could be to work with models that explicitly account for the water in the top few centimetres of soil and therefore can directly assimilate a (rescaled) satellite retrieval. However, the errors in satellite SM retrievals are probably already autocorrelated (Crow and Van den Berg, 2010).

Breaching some of the EnKF-based scheme and/or the IV-based rescaling assumptions could theoretically degrade the performance of the SM-DA scheme, when the variable analysed is soil moisture (Crow and Van den Berg, 2010; Reichle et al., 2008; Ryu et al., 2009). In this context, the performance of SM-DA with respect to the improvement in streamflow has been under-investigated. Alvarez-Garreton et al. (2013, 2014) show that in terms of streamflow prediction, SM-DA seems to be less sensitive to violation of these assumptions. Both the lower sensitivity and the apparent contradiction with previous studies analysing soil moisture prediction performance highlight the need for further studies focusing on SM-DA for the purposes of improving streamflow prediction from rainfall-runoff models.

4.4 Satellite soil moisture data assimilation

The ensemble predictions of streamflow and $\boldsymbol{\theta}$, before and after SM-DA, for both the lumped and the semi-distributed schemes at N7, are presented in Fig.10. The truncation bias correction (Sect. 3.3) was successful in creating an unbiased $\boldsymbol{\theta}$ ensemble when the unperturbed model approached the soil water storage bounds (Figs.10a.2 and 10b.2).

The rank histograms at N7, N1 and N3 are presented in Fig. 7. For all the evaluated nodes, the ensemble predictions are more reliable after SM-DA (flatter histograms compared with the open-loop). The consistent overestimation of the observed streamflow in the open-loop ensembles (diagonal histograms displaced towards the higher ensemble percentiles) is partially addressed by the SM-DA.

The evaluation statistics for the SM-DA are summarised in Table 4. The streamflow data of the inner catchments (N1 and N3) are used only for evaluation purposes in the semi-distributed scheme, therefore they are representative of "ungauged" inner catchments.

The NRMSE in Table 4 (all values below 1) demonstrates that the SM-DA was effective in reducing the streamflow prediction uncertainty (RMSE) across all gauged and ungauged catchments. The reductions in the RMSE ranged from 17 to 24% for the different evaluation nodes. The NRMSE combines precision improvement (i.e., reduction of ensemble spread) with prediction accuracy improvement (i.e., enhancement of ensemble mean performance) resulting from the SM-DA. Given that the ensemble open-loop spread was larger than an ideal ensemble (based on the n-shaped rank histograms in Fig.7), the reduction of the ensemble spread may be in part artificial.

Table 4. SM-DA evaluation statistics calculated at the catchment outlet (N7) and at the inner catchments (N1 and N3).

Statistic	Lumped scheme (N7)	Semi-distributed scheme (N7)	Semi-distributed scheme (N1)	Semi-distributed scheme (N3)
NRMSE	0.78	0.76	0.81	0.83
NSE_{ol}	0.67	0.77	0.28	-1.75
NSE_{up}	0.64	0.78	0.26	-1.39
POD_{ol}	0.96	0.92	0.56	0.69
POD_{up}	0.94	0.93	0.55	0.69
FP_{ol}	0.11	0.11	0.07	0.12
FP_{up}	0.10	0.10	0.06	0.11
PVE_{ol}	5.63	35.30	-96.87	56.42
PVE_{up}	-2.37	34.93	-109.66	40.71
$CRPS_{ol}$	0.32	0.26	0.74	0.20
$CRPS_{up}$	0.28	0.23	0.73	0.24

The performance of the ensemble mean was assessed by computing the NSE_{ol} and NSE_{up} (Table 4). At the catchment outlet, the NSE of the ensemble mean after SM-DA only improved for the semi-distributed scheme. At the ungauged catchments, SM-DA was effective at improving the performance of the ensemble mean only at N3, compared with the open-loop. However, the performance of the model in that catchment was still poor. This can be explained by the systematic errors present in the model for those catchments before assimilation, which were not addressed by the SM-DA.

The POD values at the catchment outlet (N7) show that before and after SM-DA, the model is consistently capable of detecting minor floods. Although this does not demonstrate an advantage of the SM-DA scheme proposed here, it does reflect the adequacy of the model ensemble prediction for simulating minor (and larger) floods. Consistently with previous results, the prediction of the semi-distributed model at the inner catchments is poorer in terms of detecting minor floods. The lower FAR values after SM-DA demonstrates the efficacy of the scheme in reducing the number of times the model predicted an unobserved minor flood, at both the gauged and the ungauged catchments.

The open-loop PVE was improved (lower PVE values) after SM-DA at N7 (for both the lumped and the semi-distributed schemes) and at N3. This was not the case however, for inner node N1, at which the PVE was higher after SM-DA, compared with the open-loop. When compared to the unperturbed model run (Table 2), the assimilation of satellite soil moisture improved the performance of the

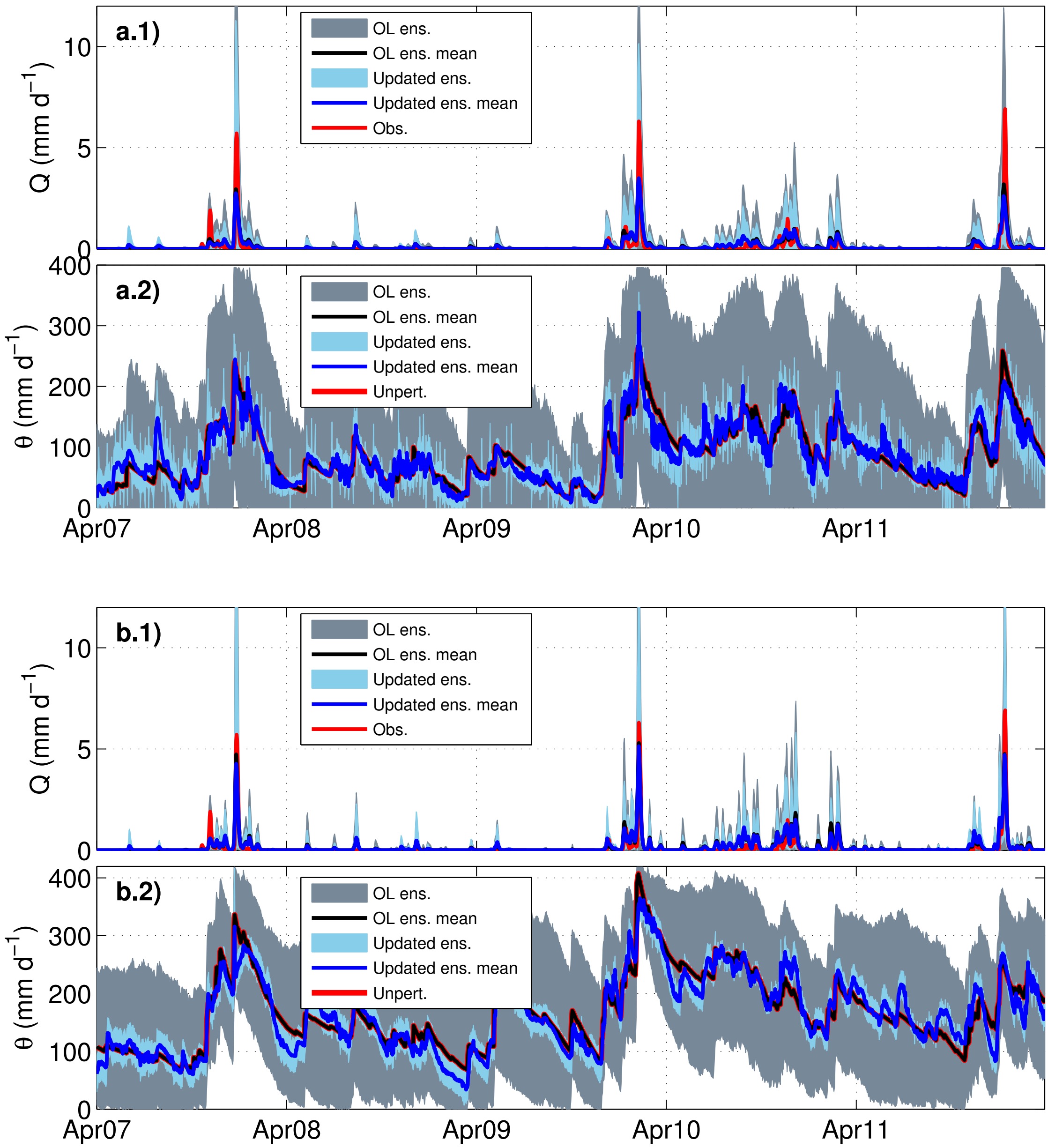

Fig. 10. Streamflow (Q in mm d^{-1}) and soil moisture ($\boldsymbol{\theta}$ in mm d^{-1}) ensemble prediction at the catchment outlet, before and after SM-DA for evaluation sub-period 2 (01 May 2007 - 02 March 2014), which had three major flooding events. (a.1) and (a.2) present the results for the lumped model. (b.1) and (b.2) present the results for the semi-distributed model.

model in terms of PVE at all the nodes and for both the lumped and semi-distributed schemes.

The skill of the ensembles after SM-DA was improved at the catchment outlet by 12% and 13% (expressed by a reduction in CRPS) for the lumped and semi-distributed scheme respectively, and by a 17% at N1. The skill of the updated ensemble was also consistently higher than the unperturbed model run (Table 2).

To summarise the efficacy of the SM-DA, we take into account the characteristics of the ensemble predictions (open-loop and updated) in terms of the their mean, skill and reliability. Overall, SM-DA was effective at improving streamflow ensemble predictions in the gauged and the ungauged catchments. By accounting for rainfall spatial distribution and routing process within the large study catchment, we improved the model performance at the outlet compared with a lumped homogeneous scheme.This led to greater improvements from the SM-DA for the semi-distributed model. The latter was achieved even though the relationship between θ and the streamflow prediction was weaker in the semi-distributed scheme (Fig.5). The proposed SM-DA scheme therefore, has the merits of improving streamflow ensemble predictions by correcting the SM state of the model, even when rainfall appears to be the main driver of the runoff mechanism (see Sect. 4.1).

5 Conclusions

This paper presents an evaluation of the assimilation of passive and active satellite soil moisture observations (SM-DA) into a conceptual rainfall-runoff model (PDM) for the purpose of reducing flood prediction uncertainty in a sparsely monitored catchment. We set up the experiments in the large semi-arid Warrego River Basin (>40,000 km^2) in south central Queensland, Australia. Within this context, we explore the advantages of accounting for the forcing data spatial distribution and the routing processes within the catchment.

The framework proposed here rigorously addressed the two main stages of a SM-DA scheme: model error representation and satellite data processing. We applied the different methods in the context of a sparsely monitored large catchment (i.e., limited data), under operational streamflow and flood forecasting scenarios (i.e., not future information is used in any of the presented methods).

The model error representation was the most critical step in the SM-DA scheme, since it determined the error covariance between observations and model state, and thus the potential efficacy of SM-DA. Moreover, the SM-DA evaluation was done against the open-loop ensemble prediction. We addressed key issues of the ensemble generation process by correcting truncation biases in soil moisture and streamflow predictions. This prevented an unintended degradation of the open-loop ensembles coming from perturbing a highly non-linear model. The open-loop ensembles at the catchment outlet provide key information about prediction uncertainty, which is required for assessing risks associated with water management decisions (Robertson et al., 2013). These ensembles showed a slight bias with respect to the observed streamflow and featured a wide spread. Further exploration of model error representation (sources of error and the structure of those errors) and error parameter estimation is required to improve the characteristics of the open-loop ensemble prediction.

In the satellite data processing, we highlighted that the use of an exponential filter to transfer surface information into deeper layers may potentially lead to violation of some of TC and EnKF assumptions (Sect. 4.3). Possible solutions to overcome this would be to use more physically-based methods to transfer satellite SM into deeper layers or to use a rainfall-runoff model that explicitly accounts for the surface soil layer that can directly assimilate a (rescaled) satellite SM product. However, both solutions are constrained by the ancillary data available for satisfactory implementation of a physically-based model. In the rescaling and error estimation procedure, we applied seasonal TC and LV to avoid error-in-variable biases. Applying these to correct biases in the SWI, showed improved agreement between observed and modelled SM. This seasonal approach is novel in the context of SM-DA and tends to lead to closer agreement between model and observations. Further investigation is required to assess the impacts and importance of accounting for seasonality in rescaling and error estimation.

The evaluation of the SM-DA results led to several insights. 1) The SM-DA was successful at improving the open-loop ensemble prediction at the catchment outlet, for both the lumped and the semi-distributed case. 2) Accounting for spatial distribution in the model forcing data and for the routing processes within the large study catchment improved the skill of the SM-DA at the catchment outlet. 3) The SM-DA was effective at improving streamflow prediction at the ungauged locations, compared with the open-loop. However, the updated prediction in those catchments was still poor, because the systematic errors before assimilation are not addressed by a SM-DA scheme.

This work provides new evidence of the efficacy of SM-DA in improving streamflow ensemble predictions within sparsely instrumented catchments. We demonstrate that SM-DA skill can be enhanced if the spatial distribution of forcing data and routing processes within the catchment are accounted for in large catchments. We show that SM-DA performance is directly related to the model quality before assimilation. Therefore we recommend that efforts should be focused on ensuring adequate models, while evaluating the trade-offs between more complex models and data availability.

Acknowledgements. The authors wish to thank one anonymous reviewer, Dr. Uwe Ehret and the Chief-Execute Editor Dr. Erwin Zehe for their constructive comments and suggestions on the earlier draft of the paper. This research was conducted with financial support from the Australian Research Council (ARC Linkage Project No. LP110200520) and the Australian Bureau of Meteorology. C. Alvarez-Garreton was supported by Becas Chile scholarship.

References

Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Piguet, B., Martin, E., et al.: From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations, Hydrol. Earth Syst. Sci., 12, 1323–1337, 2008.

Albergel, C., Rüdiger, C., Carrer, D., Calvet, J.-C., Fritz, N., Naeimi, V., Bartalis, Z., and Hasenauer, S.: An evaluation of ASCAT surface soil moisture products with in-situ observations in Southwestern France, Hydrol. Earth Syst. Sci., 13, 115–124, 2009.

Albergel, C., Calvet, J., De Rosnay, P., Balsamo, G., Wagner, W., Hasenauer, S., Naeimi, V., Martin, E., Bazile, E., Bouyssel, F., et al.: Cross-evaluation of modelled and remotely sensed surface soil moisture with in situ data in southwestern France, Hydrol. Earth Syst. Sci., 14, 2177–2191, 2010.

Albergel, C., de Rosnay, P., Gruhier, C., Muñoz-Sabater, J., Hasenauer, S., Isaksen, L., Kerr, Y., and Wagner, W.: Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations, Remote Sens. Environ., 118, 215–226, 2012.

Alvarez-Garreton, C., Ryu, D., Western, A. W., Crow, W. T., and Robertson, D. E.: Impact of observation error structure on satellite soil moisture assimilation into a rainfall-runoff model, in: MODSIM2013, 20th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, edited by Piantadosi, J., Anderssen, R., and Boland, J., pp. 3071–3077, 2013.

Alvarez-Garreton, C., Ryu, D., Western, A., Crow, W., and Robertson, D.: The impacts of assimilating satellite soil moisture into a rainfall–runoff model in a semi-arid catchment, J. Hydrol., 519, 2763–2774, 2014.

Anderson, J. L.: A method for producing and evaluating probabilistic forecasts from ensemble model integrations, J. Climate, 9, 1518–1530, 1996.

Beven, K. and Germann, P.: Macropores and water flow in soils, Water Resour. Res., 18, 1311–1325, 1982.

Brocca, L., Melone, F., Moramarco, T., and Morbidelli, R.: Antecedent wetness conditions based on ERS scatterometer data, J. Hydrol., 364, 73–87, 2009.

Brocca, L., Melone, F., Moramarco, T., Wagner, W., and Hasenauer, S.: ASCAT soil wetness index validation through in situ and modeled soil moisture data in central Italy, Remote Sens. Environ., 114, 2745–2755, 2010a.

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, Hydrol. Earth Syst. Sci., 14, 1881–1893, 2010b.

Brocca, L., Hasenauer, S., Lacava, T., Melone, F., Moramarco, T., Wagner, W., Dorigo, W., Matgen, P., Martínez-Fernández, J., Llorens, P., et al.: Soil moisture estimation through ASCAT and AMSR-E sensors: an intercomparison and validation study across Europe, Remote Sens. Environ., 115, 3390–3408, 2011.

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, IEEE T. Geosci. Remote, 50, 2542–2555, 2012.

Ceballos, A., Scipal, K., Wagner, W., and Martínez-Fernández, J.: Validation of ERS scatterometer-derived soil moisture data in the central part of the Duero Basin, Spain, Hydrol. Process., 19, 1549–1566, 2005.

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, Adv. Water Resour., 34, 526–536, 2011.

Chen, F., Crow, W. C., and Ryu, D.: Dual forcing and state correction via soil moisture assimilation for improved rainfall runoff modelling, J. Hydrometeorol., doi:10.1175/JHM-D-14-0002.1, 2014.

Chipperfield, A. and Fleming, P.: The MATLAB genetic algorithm toolbox, in: Applied Control Techniques Using MATLAB, IEE Colloquium on, pp. 10/1–10/4, 1995.

Crow, W. and Ryu, D.: A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals, Hydrol. Earth Syst. Sci., 13, 1–16, 2009.

Crow, W. T. and Reichle, R. H.: Comparison of adaptive filtering techniques for land surface data assimilation, Water Resour. Res., 44, W08 423, doi:10.1029/2008WR006 883, 2008.

Crow, W. T. and Van den Berg, M. J.: An improved approach for estimating observation and model error parameters in soil moisture data assimilation, Water Resour. Res., 46, W12 519, doi:10.1029/2010WR009 402, 2010.

Crow, W. T. and van Loon, E.: Impact of Incorrect Model Error Assumptions on the Sequential Assimilation of Remotely Sensed Surface Soil Moisture, J. Hydrometeorol., 7, 421–432, 2006.

De Lannoy, G. J., Houser, P. R., Pauwels, V., and Verhoest, N. E.: Assessment of model uncertainty for soil moisture through ensemble verification, J. Geophys. Res. - Atmos., 111, D10 101, doi:10.1029/2005JD006 367, 2006.

Dee, D. P. and Da Silva, A. M.: Data assimilation in the presence of forecast bias, Q. J. Roy. Meteor. Soc., 124, 269–295, 1998.

Draper, C. S., Walker, J. P., Steinle, P. J., de Jeu, R. A., and Holmes, T. R.: An evaluation of AMSR–E derived soil moisture over Australia, Remote Sens. Environ., 113, 703–710, 2009.

Evensen, G.: The ensemble Kalman filter: Theoretical formulation and practical implementation, Ocean Dynam., 53, 343–367, 2003.

Ford, T., Harris, E., and Quiring, S.: Estimating root zone soil moisture using near-surface observations from SMOS, Hydrol. Earth Syst. Sci., 18, 139–154, 2014.

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., 4, 473–487, 2003.

Gill, M. A.: Flood routing by the Muskingum method, J. Hydrol., 36, 353–363, 1978.

Gruhier, C., De Rosnay, P., Hasenauer, S., Holmes, T. R., De Jeu, R. A., Kerr, Y. H., Mougin, E., Njoku, E., Timouk, F., Wagner, W., et al.: Soil moisture active and passive microwave products: intercomparison and evaluation over a Sahelian site., Hydrol. Earth Syst. Sci., 2010.

Hain, C. R., Crow, W. T., Anderson, M. C., and Mecikalski, J. R.: An ensemble Kalman filter dual assimilation of thermal infrared and microwave satellite observations of soil moisture into the Noah land surface model, Water Resour. Res., 48, W11 517, doi:10.1029/2011WR011 268, 2012.

Jones, D. A., Wang, W., and Fawcett, R.: High-quality spatial climate data-sets for Australia, Australian Meteorological and Oceanographic Journal, 58, 233–248, 2009.

Jothityangkoon, C., Sivapalan, M., and Farmer, D.: Process controls of water balance variability in a large semi-arid catchment: downward approach to hydrological model development, J. Hydrol., 254, 174–198, 2001.

Kerr, Y. H., Waldteufel, P., Richaume, P., Wigneron, J.-P., Ferrazzoli, P., Mahmoodi, A., Al Bitar, A., Cabot, F., Gruhier, C., Juglea, S. E., et al.: The SMOS soil moisture retrieval algorithm, Geoscience and Remote Sensing, IEEE Transactions on, 50, 1384–1403, 2012.

Li, Y., Ryu, D., Western, A. W., Wang, Q., Robertson, D. E., and Crow, W. T.: An integrated error parameter estimation and lag-aware data assimilation scheme for real-time flood forecasting, J. Hydrol., 519, 2722–2736, 2014.

Liu, Y. Q. and Gupta, H. V.: Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, Water Resour. Res., 43, W07 401, doi:10.1029/2006WR005 756, 2007.

Loew, A. and Schlenz, F.: A dynamic approach for evaluating coarse scale satellite soil moisture products, Hydrol. Earth Syst. Sci., 15, 75–90, 2011.

Manfreda, S., Brocca, L., Moramarco, T., Melone, F., and Sheffield, J.: A physically based approach for the estimation of root-zone soil moisture from surface measurements, Hydrol. Earth Syst. Sci., 18, 1199–1212, 2014.

McKenzie, N. J., Jacquier, D., Ashton, L., and Cresswell, H.: Estimation of soil properties using the Atlas of Australian Soils, CSIRO Land and Water Canberra, 2000.

McMillan, H., Jackson, B., Clark, M., Kavetski, D., and Woods, R.: Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models, J. Hydrol., 400, 83–94, 2011.

Moore, R. and Bell, V.: Incorporation of groundwater losses and well level data in rainfall-runoff models illustrated using the PDM, Hydrology and Earth System Sciences Discussions, 6, 25–38, 2002.

Moore, R. J.: The PDM rainfall-runoff model, Hydrol. Earth Syst. Sci., 11, 483–499, 2007.

Naeimi, V., Scipal, K., Bartalis, Z., Hasenauer, S., and Wagner, W.: An improved soil moisture retrieval algorithm for ERS and METOP scatterometer observations, IEEE T. Geosci. Remote, 47, 1999–2013, 2009.

Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I: A discussion of principles, J. Hydrol., 10, 282–290, 1970.

Owe, M., de Jeu, R., and Holmes, T.: Multisensor historical climatology of satellite-derived global land surface moisture, J. Geophys. Res. - Earth, 113, F01 002, doi:10.1029/2007JF000 769, 2008.

Plaza, D., De Keyser, R., De Lannoy, G., Giustarini, L., Matgen, P., and Pauwels, V.: The importance of parameter resampling for soil moisture data assimilation into hydrologic models using the particle filter., Hydrol. Earth Syst. Sci., 16, 2012.

Qiu, J., Crow, W. T., Nearing, G. S., Mo, X., and Liu, S.: The impact of vertical measurement depth on the information content of soil moisture times series data, Geophys. Res. Lett., 41, 4997–5004, 2014.

Reichle, R. H., Crow, W. T., and Keppenne, C. L.: An adaptive ensemble Kalman filter for soil moisture data assimilation, Water Resour. Res., 44, W03 423, doi:10.1029/2007WR006 357, 2008.

Richards, L. A.: Capillary conduction of liquids through porous mediums, Physics, 1, 318–333, 1931.

Robertson, D. E., Shrestha, D. L., and Wang, Q. J.: Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting, Hydrol. Earth Syst. Sci., 17, 3587–3603, 2013.

Ryu, D., Crow, W. T., Zhan, X., and Jackson, T. J.: Correcting Unintended Perturbation Biases in Hydrologic Data Assimilation, J. Hydrometeorol., 10, 734–750, 2009.

Scipal, K., Holmes, T., De Jeu, R., Naeimi, V., and Wagner, W.: A possible solution for the problem of estimating the error structure of global soil moisture data sets, Geophys. Res. Lett., 35, L24 403, doi:10.1029/2008GL035 599, 2008.

Stoffelen, A.: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation, J. Geophys. Res. - Oceans, 103, 7755–7766, 1998.

Su, C., Ryu, D., Crow, W. T., and Western, A. W.: Beyond triple collocation: Applications to soil moisture monitoring, J. Geophys. Res. - Atmos., 119(11), 6416–6439, 2014a.

Su, C.-H., Ryu, D., Young, R. I., Western, A. W., and Wagner, W.: Inter-comparison of microwave satellite soil moisture retrievals over the Murrumbidgee Basin, southeast Australia, Remote Sens. Environ., 134, 1–11, 2013.

Su, C.-H., Ryu, D., Crow, W. T., and Western, A. W.: Stand-alone error characterisation of microwave satellite soil moisture using a Fourier method, Remote Sens. Environ., 154, 115–126, 2014b.

Thielen, J., Bartholmes, J., Ramos, M.-H., and De Roo, A.: The European Flood Alert System-Part 1: Concept and development., Hydrol. Earth Syst. Sci., 13, 2009.

Tian, Y., Huffman, G. J., Adler, R. F., Tang, L., Sapiano, M., Maggioni, V., and Wu, H.: Modeling errors in daily precipitation measurements: Additive or multiplicative?, Geophys. Res. Lett., 40, 2060–2065, 2013.

Wagner, W., Lemoine, G., and Rott, H.: A method for estimating soil moisture from ERS scatterometer and soil data, Remote Sens. Environ., 70, 191–207, 1999.

Wanders, N., Karssenberg, D., Roo, A. d., de Jong, S., and Bierkens, M.: The suitability of remotely sensed soil moisture for improving operational flood forecasting, Hydrol. Earth Syst. Sci., 18, 2343–2357, 2014.

Wang, Q., Robertson, D., and Chiew, F.: A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites, Water Resour. Res., 45, W05 407, doi:10.1029/2008WR007 355, 2009.

Western, A. W., Grayson, R. B., and Blöschl, G.: Scaling of soil moisture: A hydrologic perspective, Annual Review of Earth and Planetary Sciences, 30, 149–180, 2002.

Yilmaz, M. T. and Crow, W. T.: The optimality of potential rescaling approaches in land data assimilation, J. Hydrometeorol., 14, 650–660, 2013.

Zwieback, S., Scipal, K., Dorigo, W., and Wagner, W.: Structural and statistical properties of the collocation technique for error characterization, Nonlinear Porc. Geoph., 19, 69–80, 2012.