

# Streamflow data assimilation for soil moisture analysis

K. Warrach-Sagi and V. Wulfmeyer

Institute for Physics and Meteorology, University of Hohenheim, Stuttgart, Germany

Received: 21 May 2009 – Published in Geosci. Model Dev. Discuss.: 12 June 2009

Revised: 28 October 2009 – Accepted: 5 November 2009 – Published: 7 January 2010

**Abstract.** Streamflow depends on the soil moisture of a river catchment and can be measured with relatively high accuracy. The soil moisture in the root zone influences the latent heat flux and, hence, the quantity and spatial distribution of atmospheric water vapour and precipitation. As numerical weather forecast and climate models require a proper soil moisture initialization for their land surface models, we enhanced an Ensemble Kalman Filter to assimilate streamflow time series into the multi-layer land surface model TERRA-ML of the regional weather forecast model COSMO. The impact of streamflow assimilation was studied by an observing system simulation experiment in the Enz River catchment (located at the downwind side of the northern Black Forest in Germany). The results demonstrate a clear improvement of the soil moisture field in the catchment. We illustrate the potential of streamflow data assimilation for weather forecasting and discuss its spatial and temporal requirements for a corresponding, automated river gauging network.

precipitation is influenced by the interaction of the atmosphere with the land surface through the energy and water fluxes. Corresponding studies show the soil moisture influence on quantity and spatial distribution of precipitation (e.g. Schär et al., 1999; Hohenegger et al., 2008; Trier et al., 2004). Particularly, in summertime, continental QPF depends on the initialization of root zone soil moisture and other land surface states (Reichle et al., 2002; Hohenegger et al., 2009). Soil moisture not only depends on the weather but also on the local land surface characteristics (soil texture, vegetation, orography). But for this highly heterogeneous quantity, only scarce representative measurements are available at point locations (e.g. Bardossy and Lehmann, 1998; Grayson and Western, 1998). Multiple efforts to apply remote sensing to regions of scarce or shallow vegetation to obtain the skin layer soil moisture are currently under way (e.g. Crow and Wood, 2003; Dunne and Entekhabi, 2006; Drusch and Viterbo, 2007; Gao et al., 2007). So far, these techniques do not provide data for soil moisture estimates under dense vegetation and within the total soil profile. Hence, the knowledge of the soil moisture distribution is a key issue in NWP.

As the lower boundary of weather forecast and climate models, land surface models (LSM) calculate the coupled water and energy balance at each grid cell of the atmospheric model. On these scales ( $\geq 1 \text{ km}^2$ ), soil texture, topography and vegetation and, therefore, water and energy fluxes, soil moisture, runoff and soil temperature are highly heterogeneous (e.g. Kabat et al., 1997). This heterogeneity can neither be measured nor modelled explicitly at an acceptable cost. For each grid cell, the precipitation is balanced by the sum of evapotranspiration, runoff and soil moisture change. Evapotranspiration and soil moisture cannot be measured at this scale, over the large areas an atmospheric model is applied to (e.g. Beven, 2001; Pitman et al., 2004). Also seasonal to intra-seasonal climate simulations rely on a proper root zone soil moisture initialization (e.g. Conil et al., 2009; Seneviratne et al., 2006).

## 1 Introduction

Quantitative precipitation forecasting (QPF) is one of the most complex challenges in numerical weather prediction (NWP) (e.g. Rotach et al., 2009; Wulfmeyer et al., 2008). QPF failures can be due to errors in numerics, limited spatial resolution of the model, erroneous model physics, incorrect initial conditions and limited predictability. The skill of QPF, particularly on the mesoscale, is still strongly limited by uncertainties in initial conditions. Particularly, dynamics in complex terrain and the inhomogeneous distribution of water vapour are considered the most important unknowns in the initial fields. The water vapour field of the continental lower troposphere and, therefore, cloud formation and pre-



Correspondence to: K. Warrach-Sagi  
(warrach@uni-hohenheim.de)

Remotely sensed land surface data and air-temperature are currently assimilated to overcome errors in soil moisture and temperature simulation in NWP models (e.g. Hess, 2001; Seuffert et al., 2004; Crow and Wood, 2003; Gao et al., 2007).

Still unresolved problems are the soil moisture analysis in densely vegetated areas and in the root zone. Recently, various approaches of data assimilation were set up and analyzed to retrieve the root zone soil moisture at the regional scale in hydrological models. They mainly use Kalman Filter techniques and their modifications, which are outlined in detail, e.g. by Evensen (2006). Further, Evensen (2003) gives a detailed description and literature review of the Ensemble Kalman Filter (EnKF, Evensen, 1994). Walker et al. (2002) apply a modified Kalman Filter technique with a distributed hydrological model to retrieve the three-dimensional soil moisture from surface soil moisture measurements. This is a valuable approach in hydrology but due to the intense computational cost of a distributed hydrological model, not a tool currently suitable for NWP. Moradkhani et al. (2005) and Dunne and Entekhabi (2006), for example, use the Ensemble Kalman Smoother for root zone soil moisture analysis assimilating L-band radiobrightness temperatures in an area of the Southern Great Plains (USA) whose vegetation is mainly wheat and grasses (Drusch et al., 2001). At the German Weather Service (DWD), Hess (2001) implemented a method based on the EKF (Extended Kalman Filter) technique into the operational non-hydrostatic mesoscale weather forecast model COSMO (Doms et al., 2005) that adjusts the soil state to meet the observed atmospheric state. However, in his approach, the soil moisture and soil temperature do not necessarily match the reality, i.e., its usage is not consistent with the hydrologic interaction of the land surface and lower atmosphere. This is proven by Drusch and Viterbo (2007), who assimilated screen-level variables in ECMWF's Integrated Forecast System. If, due to the assimilation of screen level variables, the model's soil moisture and soil temperature are changed so that they may not reflect reality, this impacts other parameterizations and sub-models that rely on those variables, e.g. latent and sensible heat flux and runoff.

A data source that has only received attention in the past couple of years is streamflow from operational river gauging networks. Streamflow is a quantity that can be measured at relatively high accuracy (about >90%, LfU, 2002). If the runoff is transported to and within the river network, it can be compared to measured streamflow at gauging stations. Pauwels and De Lannoy (2006) published the application of a retrospective EnKF to assimilate streamflow data for soil moisture retrieval. Their synthetic tests show promising results for a 1000 km<sup>2</sup> catchment in Belgium and indicate improvements especially in case of precipitation underestimation. They apply it to the high resolution hydrological model TOPLATS. Komma et al. (2008) successfully applied the EnKF for soil moisture update in real-time flood forecasting

in a 622 km<sup>2</sup> catchment in Austria. However, they use a soil moisture model focusing on the hydrological model application, while in this study a land surface model for atmospheric models is applied. Clark et al. (2008) demonstrate that the standard implementation of EnKF is inappropriate and show the improved performance when streamflow is transformed into log space before applying EnKF with the distributed hydrological model TopNet. This is due to the large ranges in streamflow between peak flow and low flow, which can be 2 orders of magnitude or more.

Streamflow analyses allow for an evaluation of the model performance (e.g. Lohmann et al., 2004; Warrach-Sagi et al., 2008). In this study, we go a step further and study the potential of streamflow data assimilation for soil moisture analysis in a catchment, namely for initialisation of numerical weather prediction and climate models. We followed the most recent development in EnKF and applied it to the streamflow data assimilation for soil moisture initialization in a land surface model of the numerical weather prediction model COSMO.

In southern Germany, a network of automated river and precipitation gauges has been installed in the past couple of years by the federal services for flood monitoring. The federal state Baden-Württemberg has implemented a flood forecast centre, which is able to provide half-hourly updates of streamflow measurements at approximately 140 gauges at the rivers Rhein, Neckar, Donau, Main and their main tributaries. Similar warning systems are available in the federal state Bayern and Rheinland-Pfalz. Such automated networks provide a valuable source for operational streamflow data assimilation.

The square root algorithm for the EnKF (Evensen, 2004) is set up to assimilate streamflow data in TERRA-ML to analyse the soil-water content of the soil profile down to 2.43 m soil depth simulated by TERRA-ML. By means of an observing system simulation experiment (OSSE) in the Enz River catchment (Germany), we studied the potential of streamflow data assimilation and its spatial and temporal requirements for an automated river gauging network. The Enz catchment is on the downwind side of the Black Forest, i.e. QPF by the weather forecast model is often underestimating, making it a valuable test bed for streamflow data assimilation (Pauwels and De Lannoy, 2006). The study is carried out exemplarily with the land surface model TERRA-ML coupled to a river routing model (Warrach-Sagi et al., 2008). The multi-layer soil and vegetation model TERRA-ML serves as the lower boundary of the operational non-hydrostatic mesoscale weather forecast model COSMO (Doms et al., 2005). (COSMO is the acronym for the Consortium for Small-scale Modelling (<http://www.cosmo-model.org/>)). However, the data assimilation system can be set up for any land surface model that includes a river routing model to simulate streamflow.

## 2 Description of TERRA-ML and the river routing scheme

This study applies the stand alone version of TERRA-ML in the framework set up by Ament and Simmer (2006). The model configuration and parameters of TERRA-ML are taken from the German Weather Service's COSMO. In the framework, TERRA-ML is set up as if it is called by the COSMO, with the exception that the meteorology is read from a file instead of forecasted at the time step by the COSMO. This framework has the advantage that it allows the simulation of a gridded area (e.g. watershed) per time step mimicking a simulation with a weather forecast model. An important modification of TERRA-ML in this study is the parameterization of the hydraulic conductivity and diffusivity following Campbell (1974) instead of Rijtema (1969) due to the results of Graßelt et al. (2008) and Warrach-Sagi et al. (2008).

TERRA-ML and the river routing model are set up as described in detail by Warrach-Sagi et al. (2008), therefore, here only a summary is given. In COSMO, the model TERRA-ML has got 6 hydrological layers (layer depths from the surface: 0.01 m, 0.03 m, 0.09 m, 0.27 m, 0.81 m, and 2.43 m) and 8 thermal layers (layer depths from the surface: 0.01 m, 0.03 m, 0.09 m, 0.27 m, 0.81 m, 2.43 m, 7.29 m, and 21.87 m). The lower boundary condition is given by free drainage at 2.43 m depth and a constant climatological temperature below 7.29 m depth.

For model simulations, watersheds are divided into grid cells as in atmospheric models. For each grid cell, the one-dimensional vertical land surface model TERRA-ML is applied. The locally generated runoff of the LSM needs to be transported into and along the river system to compare it to streamflow measurements at gauging stations and to calculate the streamflow at various locations of the river. Based on the routing scheme described in detail by Lohmann et al. (1996, 2004) present a lumped optimized linear routing model, which Warrach-Sagi et al. (2008) coupled to TERRA-ML. The routing scheme describes the time runoff takes to reach the outlet of a grid cell and the water transport in the river network. It is assumed that water flows unidirectionally from grid cell to grid cell with eight possible directions through each side and corner of the grid cell.

## 3 The streamflow data assimilation system

The Ensemble Kalman Filter (EnKF) has been reviewed by many authors recently (e.g. Evensen, 2003, 2006; Pauwels and DeLannoy, 2006; Clark et al., 2008) and, therefore, here only a short description of its implementation for the streamflow data assimilation is given.

Both model results and observation, deviate from the true state. The goal of data assimilation is to find the best estimate of the state (e.g. soil moisture) from model simulations and

measurements (e.g. streamflow). One method is to estimate the mean state and the “maximum likelihood” including its covariance as uncertainty measure as it is provided e.g. by the EnKF.

Various algorithms solve the EnKF equations (see e.g. Evensen, 2006). For this study, we chose the square root algorithm for EnKF (<http://enkf.nersc.edu>) from Evensen (2004) due to the following aspects: it is stable, needs relatively little computing time, requires relatively little memory and it is straight-forward to implement. The following base line equations describe the EnKF as it is implemented:

$$\mathbf{x}_{e,n}^b = M(\mathbf{x}_{e,n-1}^a) \quad (1)$$

$$\mathbf{B}_{e,n} = \overline{(\mathbf{x}_{e,n}^b - \overline{\mathbf{x}_{e,n}^b}) \cdot (\mathbf{x}_{e,n}^b - \overline{\mathbf{x}_{e,n}^b})^T} \quad (2)$$

$$\mathbf{K}_{e,n} = \overline{(\mathbf{x}_{e,n}^b - \overline{\mathbf{x}_{e,n}^b}) \cdot (H(\mathbf{x}_{e,n}^b) - H(\overline{\mathbf{x}_{e,n}^b}))^T} \cdot \left( \mathbf{R}_{e,n}^{-1} + \overline{(H(\mathbf{x}_{e,n}^b) - H(\overline{\mathbf{x}_{e,n}^b})) \cdot (H(\mathbf{x}_{e,n}^b) - H(\overline{\mathbf{x}_{e,n}^b}))^T} \right)^{-1} \quad (3)$$

$$\mathbf{A}_{e,n} = \overline{(\mathbf{x}_{e,n}^a - \overline{\mathbf{x}_{e,n}^a}) \cdot (\mathbf{x}_{e,n}^a - \overline{\mathbf{x}_{e,n}^a})^T} = (\mathbf{I} - \mathbf{K}_{e,n} \mathbf{H}) \mathbf{B}_{e,n} \quad (4)$$

$$\mathbf{x}_{e,n}^a = \mathbf{x}_{e,n}^b + \mathbf{K}_{e,n} (\mathbf{y}_{e,n} - H(\mathbf{x}_{e,n}^b)) \quad (5)$$

Bold letters represent matrices,  $\mathbf{x}$  and  $\mathbf{y}$  are the vectors for the model state and observation.  $b$  is the background (i.e. initial state),  $a$  is the analysis,  $e$  is the ensemble member,  $n$  is the time step,  $T$  is the transpose and  $\mathbf{K}$  is the Kalman gain matrix.  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{R}$  are error covariance matrices of the analysis, background and observation,  $\mathbf{I}$  is the identity matrix,  $H$  is the observation operator (in this case, the river routing model), which transforms the variable from model space to observation space,  $\mathbf{H}$  is the tangent linear observation operator matrix of  $H$  and  $M$  is the model operator (in this case TERRA-ML). Though  $\mathbf{A}$  is not needed within the filtering process,  $\mathbf{A}$  is a valuable output for the application of the updated soil moisture fields as initial condition, for example, in weather prediction models running in a data assimilation mode. Furthermore,  $\mathbf{A}$  is critical information for the interpretation of the results. The last term of Eq. (4) demonstrates the relation of the EnKF to the Extended Kalman Filter.

Depending on the location within the catchment, the water needs more or less time to travel as streamflow through the river network. Water far away from the gauge arrives later than the runoff from grid cells close by. The travel time depends on the river itself and the form and orography of the catchment. This means that the streamflow measured at a gauge depends on the soil moisture distribution in the catchment for a time window from time  $t=0$  to  $t=m \cdot dt$ .  $m$  is the time step,  $dt$  is the time interval of one time step. This period of streamflow data needs to be assimilated. By this the EnKF becomes a “retrospective” EnKF, whose concept

is described by Pauwels and DeLannoy (2006). The time window depends on the catchment and is determined prior to the streamflow data assimilation. Then streamflow time series are assimilated depending on the catchments' time window to obtain the soil moisture (see, e.g., Sect. 4.2). Following Clark et al. (2008), the streamflow is transformed into log space before computing the error covariances since they demonstrated that this improves the filter performance.

Only the catchment's grid cells  $\gamma$  are part of the data assimilation system, i.e. each grid cell of the catchment gets an index  $h$  between 1 and  $\gamma$ . For example,  $\gamma$  grid cells belong to the catchment upstream of a gauge, i.e. the model state vector includes  $h*k*m$  soil moisture values.  $m$  the number of timesteps of length  $dt$  within the assimilation window,  $k$  is the index of the soil layers of each grid cell, the model has got  $k=\beta$  soil layers. The state vectors of the analysis and background  $x$  includes all soil moisture values  $\eta(h,k,n)$  and the simulated streamflow  $q(h,n)$  in each grid cell of the river network.  $n$  is the time step index of the streamflow assimilation window. The observation state vector  $y$  includes the observed streamflow timeseries  $Q$  from gauging station  $l$  at grid cell  $(i(l), j(l))$ .  $i$  and  $j$  are the indices of the grid cell in eastward and northward direction, to each catchment grid cell number  $h$  is defined by its model areas grid cell index  $i$  and  $j$ , i.e.  $Q(\gamma)=Q(i(l), j(l))$ . The assimilation window is from time  $t=0$  to  $t=m*dt$ ,  $dt$  is the length of the timesteps of the observed streamflow data,  $m$  the number of timesteps  $dt$  within the window. The equations for the state vectors are:

$$x = (\eta_{1,1,1}, \eta_{1,2,1}, \dots, \eta_{1,\beta,1}, \eta_{2,1,1}, \dots, \eta_{\gamma,\beta,m}, q_{1,1}, q_{2,1}, \dots, q_{\gamma,m}) \quad (6)$$

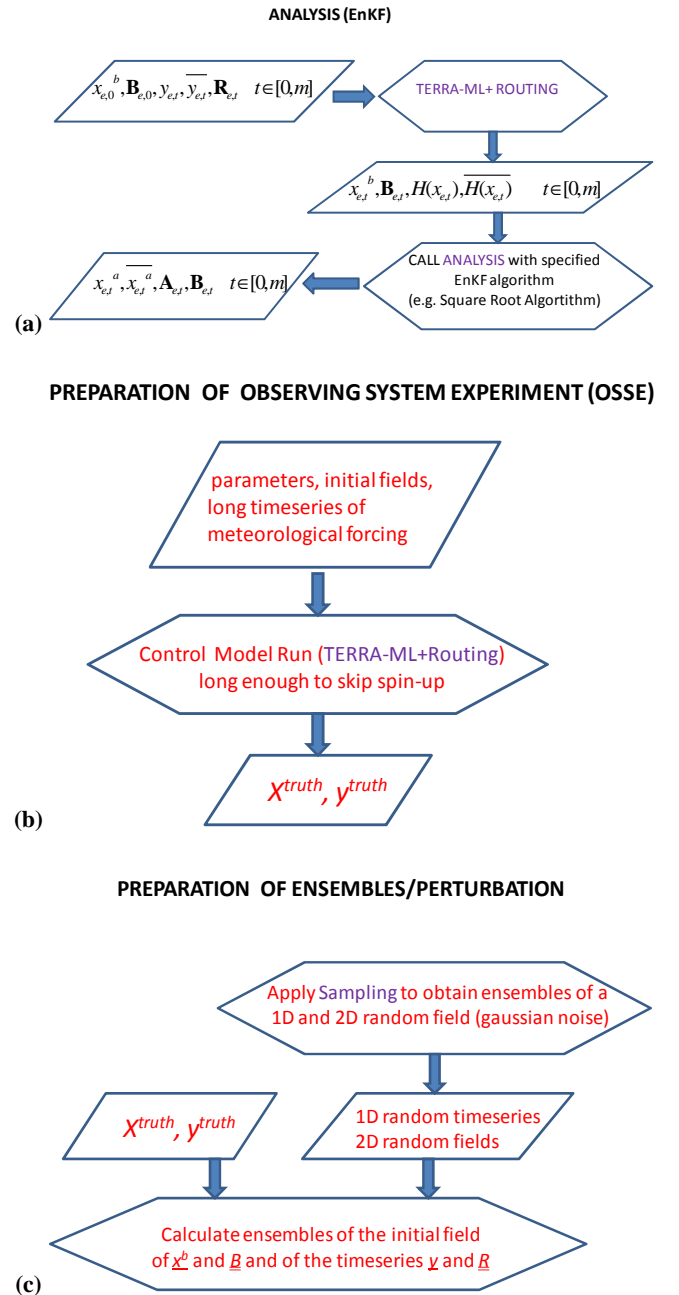
$$y = (Q_{l,1}, \dots, Q_{l,m}) \quad (7)$$

To illustrate the streamflow data assimilation system, Fig. 1a shows a flow chart for the analysis of the soil moisture at the initial timestep  $t=0$  as it is set up for this study.

## 4 The Observing System Experiment (OSSE)

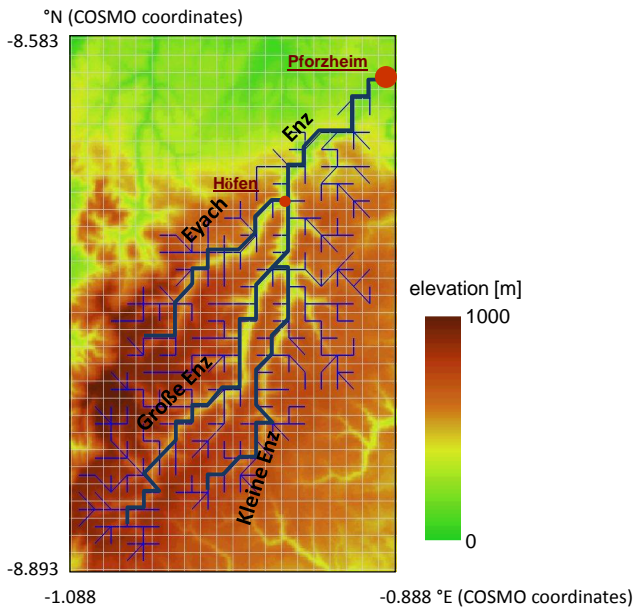
### 4.1 Study area: the Enz

The Black Forest is a mountain range that reaches from 47.5° S to 49° N at a width of approximately 50 km in Baden-Württemberg (Germany). Reaching from North to South, the Black Forest modifies significantly most frontal systems arriving from the Atlantic. In spite of its relatively low height (largest mountain Feldberg 1493 m a.m.s.l.), orographic lifting of unstable and moist air masses in this region results in the largest amount of precipitation in Germany except the northern front range of the Alps. In summer, the Black Forest is characterised by strong convection, thunderstorms, and the development of extreme precipitation events. The eastern Black Forest hosts about half of the tributaries of the Neckar, a major tributary of the Rhine. The western Black Forest drains directly to the Rhine. Most rivers



**Fig. 1.** Flow charts of (a) the streamflow data assimilation system for the soil moisture analysis of the initial time  $t=0$ . The applied model is TERRA\_ML with the river routing scheme as described by Warrach-Sagi et al. (2008), (b) the OSSE, and (c) the preparation of the perturbed soil moisture and streamflow observation data.

in Baden-Württemberg contain automated gauging stations from the flood forecast centre and streamflow data are available every 30 min.

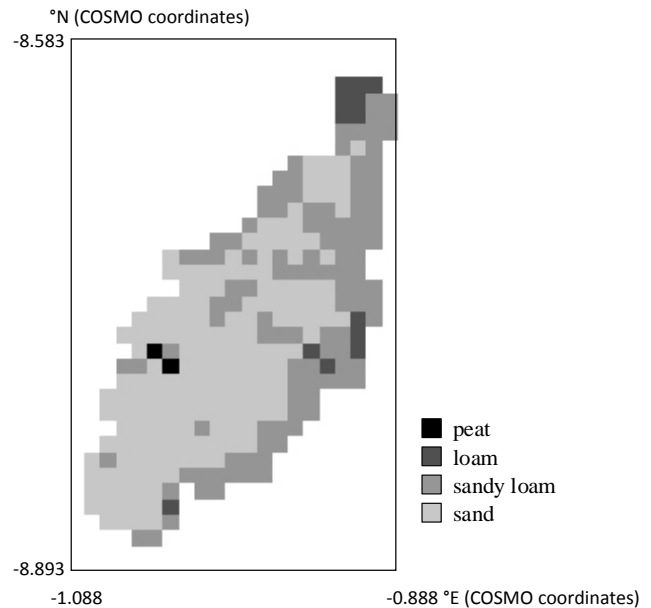


**Fig. 2.** The orography (based on the 90-m-orographic data from the SRTM) and the river network for the Enz catchment upstream of Pforzheim on the rotated spherical coordinate system of the COSMO on a grid resolution of  $0.01^\circ$  ( $\approx 1$  km).

The  $260 \text{ km}^2$  Enz catchment upstream of Pforzheim (upstream of the Nagold confluence) is on the downwind side of the Black Forest, i.e. precipitation is often underestimated by weather forecast models. Therefore, the Enz catchment (Fig. 2) was chosen for the streamflow data assimilation study. No water reservoirs interrupt the river system. Elevation of the catchment ranges between 350 and 930 m a.s.l. The catchment is characterised by forested (mixed deciduous and evergreen coniferous trees) upland areas and agriculturally used lowlands. Sandy and loamy soils dominate the upper Enz area (Fig. 3). Between 1997 and 2002 annual precipitation in the catchment ranged from 1088 to 1451 mm.

#### 4.2 Set-up of the OSSE

Warrach-Sagi et al. (2008) applied the coupled TERRA-ML-routing model to the Enz catchment upstream of Pforzheim and compared it to simulations of the flood forecast centre Baden-Württemberg and to observations. They showed that the model results and observations agree reasonably well. However, as is always the case, model results and observations both include errors and both differ from the true state. To assess the potential and requirements for streamflow data assimilation, an OSSE is set up, as illustrated in Fig. 1b. The results of the TERRA-ML-routing model for 1997 in the Enz river catchment (Warrach-Sagi et al., 2008) are assumed to be the “true” state, named “CONTROL” hereafter. The data assimilation experiment starts on 5 May 1997 with an ensemble of initial soil moisture fields in the catchment and

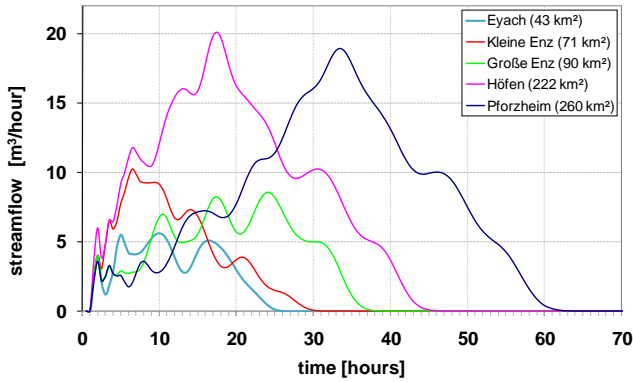


**Fig. 3.** Soil texture based on 1:200 000 soil map (BÜK 200) of the LGRB (Landesamt für Geologie, Rohstoffe und Bergbau) (Warrach-Sagi et al., 2008). In TERRA-ML the saturated soil moisture is  $0.364 \text{ m/m}$  for sand,  $0.445 \text{ m/m}$  for sandy loam,  $0.463 \text{ m/m}$  for loam and  $0.863 \text{ m/m}$  for peat.

an ensemble of streamflow at various locations in the river network. In this OSSE, the ensemble is limited to the perturbation of initial soil moisture fields rather than including additional ensembles of perturbed meteorological forcing. The reasons are twofold, firstly this allows for a better interpretation of the results and secondly the meteorological forcing is from measured station data, i.e. all forcing variables are consistent. Perturbing, for example, the temperature would mean to perturb the incoming radiation as well in a consistent manner and lead to quite a complex OSSE. The CONTROL streamflow serves as an “observation” which is assimilated for the soil moisture analysis. The analysis is then compared to the “true” state, i.e. the CONTROL soil moisture.

A flow duration check is carried out to obtain the assimilation time window for the whole basin at Pforzheim (upstream of Nagold confluence) and the sub basins *Große Enz* ( $90 \text{ km}^2$ ), *Kleine Enz* ( $71 \text{ km}^2$ ), *Eyach* ( $43 \text{ km}^2$ ) and upstream of Höfen ( $222 \text{ km}^2$ ), downstream of the confluence of the Eyach into the Enz (Fig. 2). For the flow duration check at the initial time step,  $0.002 \text{ kg/m}^2$  runoff are assumed for each grid cell. No more runoff is assumed afterwards. The routing model calculates the streamflow for each catchment (Fig. 4). Depending on the size and structure of the catchment, the time window until all water has left the catchment varies between 25 and 62 h. Experiments showed that in most cases an assimilation window of 90% of the time window lead to the best results in soil moisture distribution and catchments’ mean soil moisture. This is due to the following fact: The





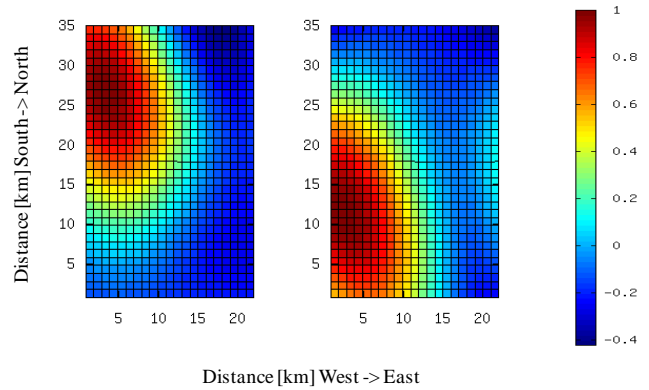
**Fig. 4.** For the flow duration check at the initial time step,  $0.002 \text{ kg/m}^2$  runoff are assumed for each grid cell in the Enz catchment. No more runoff is assumed afterwards. The river routing model calculates the streamflow for each sub-catchment.

closer the grid cell is to the gauging station, the shorter the part of the streamflow time series is responsible for its soil moisture. But since streamflow is an integrated quantity over the whole catchment, this is not separated in the EnKF. Less optimal results can be caused in those grid cells through the assimilation of this grid cell's "too long" assimilation windows. A denser gauging network would help to reduce these effects. This will be discussed in Sect. 4.4 in more detail.

### 4.3 Ensemble preparation

Figure 1c illustrates the ensemble preparation. For the OSSE, a period is chosen which does not include extreme events (such as flooding or drought or strong precipitation). A period in spring was chosen, when not only soil texture but also vegetation and weather control the soil moisture. Furthermore, in spring and summer soil moisture impacts the development of convection in the atmosphere. This study starts on 5 May 1997 (day 125). The initial soil moisture of the CONTROL simulation is perturbed applying the 2-D-pseudorandom sampling method and algorithm (<http://enkf.nersc.no>) of Evensen (2004) to obtain 100 ensemble members of initial soil moisture fields, which include no step-functions within the 2-D-area. (See Evensen, 2004, for more details on this approach.) The soil moisture of each grid cell is chosen to vary between +10% and -40% of the CONTROL soil moisture. This is to account for the typical underestimation of precipitation in NWP simulations in this area and to account for the fact that the precipitation might have been simulated in the wrong location within the catchment. The 2-D-pseudorandom fields vary up to  $d=\pm 1$  and examples are shown for 2 ensemble members in Fig. 5. According to the random number of each grid cell  $(i, j, k)$  of each ensemble member  $e$ , the soil moisture  $\eta$  in the grid cell is perturbed to

$$\eta_{i,j,k,e} = \eta_{i,j,k,c} \cdot d \cdot 0.1 \quad \forall \quad d > 0, \quad (8a)$$



**Fig. 5.** The initial soil moisture of the CONTROL simulation is perturbed applying the 2-D-pseudorandom sampling method and algorithm (<http://enkf.nersc.no>) of Evensen (2004) to obtain 100 ensemble members of initial soil moisture fields. The 2-D-pseudorandom fields vary up to  $d=\pm 1$  and examples are shown for 2 ensemble members.

$$\eta_{i,j,k,e} = \eta_{i,j,k,c} \cdot d \cdot 0.4 \quad \forall \quad d \leq 0, \quad (8b)$$

$i, j$  and  $k$  are the indices of the grid cell in eastward, northward and downward direction,  $c$  is the control state. Like in nature, soil moisture in the ensemble for each grid cell is always limited between saturation and air dryness point.

The background ensemble of the streamflow at Pforzheim simulated with TERRA-ML and the routing scheme from the initial ensemble of soil moisture fields shows that the ensemble does not converge during the first 200 h (Fig. 6a) even though the same atmospheric forcing is applied to each ensemble member. Most variability in streamflow between the ensemble members can be seen 30 and 70 h after the simulation started. The mean of the background ensemble streamflow is lower than the CONTROL streamflow (Fig. 6c).

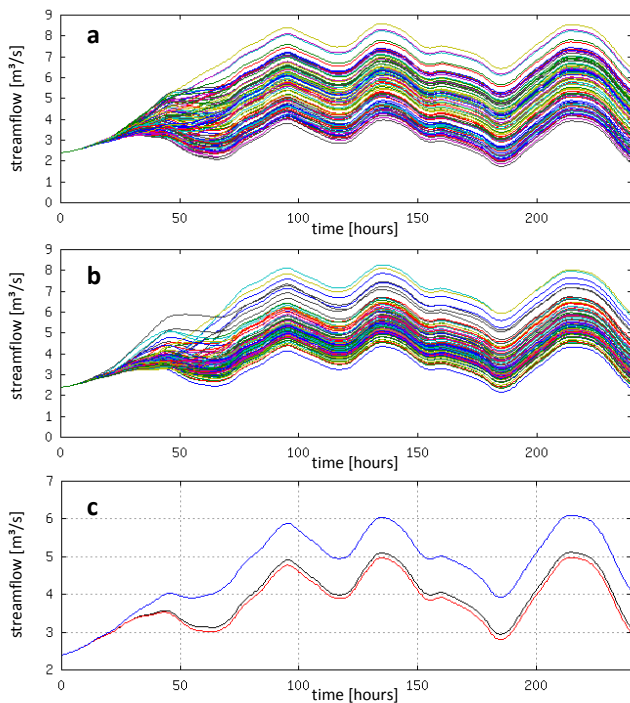
The CONTROL streamflow is perturbed by adding Gaussian noise. The 1-D-pseudorandom sampling method and algorithm (<http://enkf.nersc.no>) of Evensen (2004) to obtain 100 ensemble members is applied and streamflow perturbed by up to  $\pm 15\%$ , assuming that the error might be occasionally larger than the  $< 10\%$  assumed by LfU (2002).

### 4.4 Soil moisture analysis

TERRA-ML's soil column is 2.43 m deep (see Sect. 2). The soil-water content (SWC) of each grid cell depends on its soil depth and its soil moisture  $\eta$

$$\text{SWC}_{i,j} = \rho_w \cdot \sum_{k=1}^l (\eta_{i,j,k} \cdot (z(k) - z(k-1))), \quad (9)$$

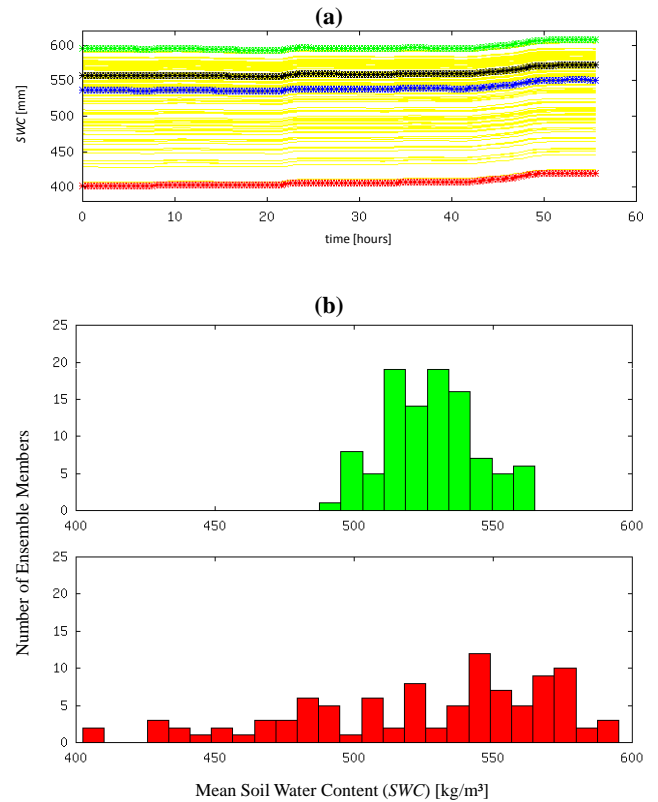
with the density of water  $\rho_w$ ,  $z(k)$  is the depth of the lower boundary of soil layer  $k$ ,  $l$  is the lowest soil layer. Streamflow data is assumed to be available at an half hourly time step like the observations made by the automated gauges in Baden-Württemberg.



**Fig. 6.** The streamflow at Pforzheim starting on 5 May 1997 with the CONTROL streamflow value: (a) simulated for all (background) ensemble members of initial soil moisture with TERRA-ML and river routing scheme, (b) ensemble of analysis members resulting from EnKF, and (c) streamflow from the CONTROL (blue) simulation, simulated with the initial mean soil moisture of the background ensemble (red) and simulated with the initial mean soil moisture analysis (black).

The analysis timeseries of the streamflow (Fig. 6b) show a narrower spread than the background (Fig. 6a). In both cases (background and analysis), most variability in streamflow between the ensemble members can be seen 30 and 70 h after the simulation started. Figure 7 shows the ensemble of the catchments' mean SWC upstream of Pforzheim. The timeseries of the SWC of the ensemble do not converge during the streamflow assimilation window (Fig. 7a) and the median SWC is not equal to the CONTROL SWC. Figure 7b shows the ensemble spread of the catchments' mean SWC at the initial time  $t=0$  for the background and the analysis for the catchment upstream of Pforzheim assimilating streamflow data from Pforzheim. The analysis ensemble has a lower spread and is closer to the CONTROL SWC. The ensemble mean SWC at  $t=0$  is  $525 \text{ kg/m}^2$  for the background,  $528 \text{ kg/m}^2$  for the analysis and  $557 \text{ kg/m}^2$  for the CONTROL.

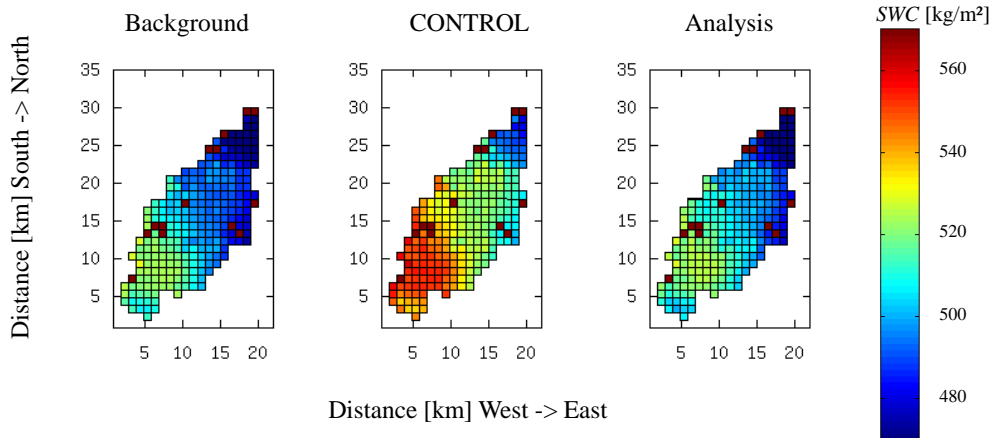
Figure 8 shows the spatial distribution of the SWC at time  $t=0$ . Note that single cells show larger SWC mainly due to different soil texture (peat and loam, see Fig. 3). While the ensemble mean of the background SWC is everywhere 5–6.5% lower than the CONTROL SWC, the ensemble mean



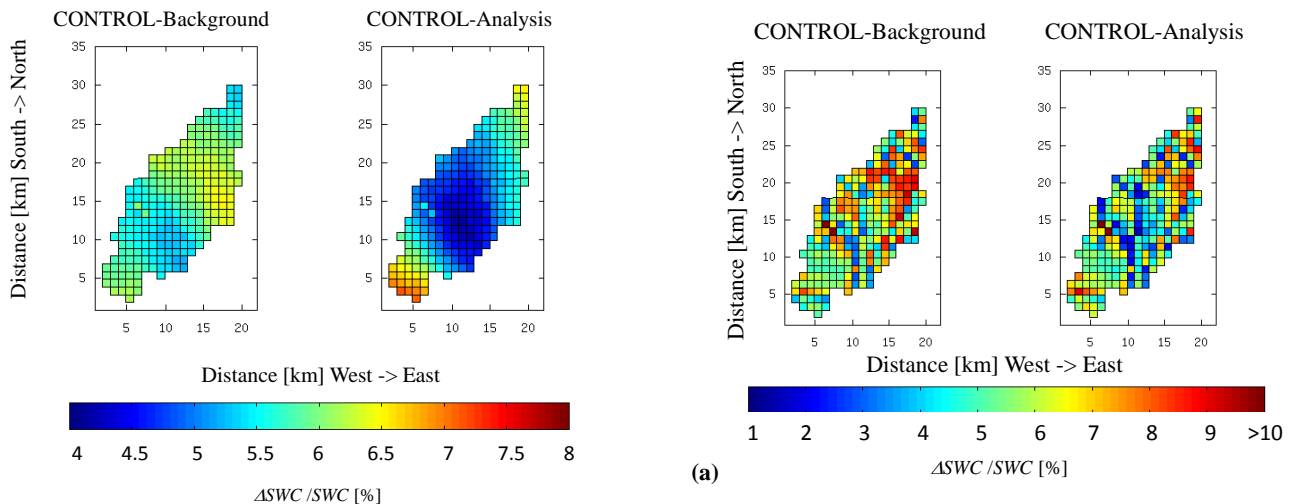
**Fig. 7.** TERRA-ML's soil column is 2.43 m deep. The soil-water content (SWC) of each grid cell depends on its soil depth (i.e. here 2.43 m) and its soil moisture (Eq. 9). SWC in the Enz catchment upstream of Pforzheim ( $260 \text{ km}^2$ ): (a) timeseries of SWC of the CONTROL (black \*) and of the ensemble members (yellow), their median SWC (blue \*), their minimum SWC (red \*) and their maximum SWC (green \*) during the assimilation window; (b) distribution of initial mean SWC ( $t=0$ ) between the ensemble members. CONTROL mean SWC is  $557 \text{ kg/m}^2$ . The analysis at  $t=0$  was obtained assimilating streamflow from the CONTROL model simulation from Pforzheim from  $t=0$  to  $t=56 \text{ h}$  with a 0.5 hourly timestep.

analysis SWC shows an improvement (Fig. 9). The analysis differs in more than half of the catchment by 4–4.5% from the CONTROL SWC. Only in a few upstream and downstream grid cells it is worse (7.5% upstream and 6.5% downstream) than the background SWC.

Figures 10 and 11 show the distribution of the differences in SWC for the soil layers of TERRA-ML. Note that TERRA-ML assumes the root depth at 0.8 m soil depth, i.e. layer 6 contains no roots and, therefore, does not contribute to the evapotranspiration in TERRA-ML. Most grid cells show an improvement of soil moisture through streamflow data assimilation in all soil layers, but it is lowest in the top 0.09 m and in the upstream grid cells. Strongest improvement is reached in a wide region in the middle of the catchment, this is most pronounced in the 4th layer (0.09–0.27 m depth).

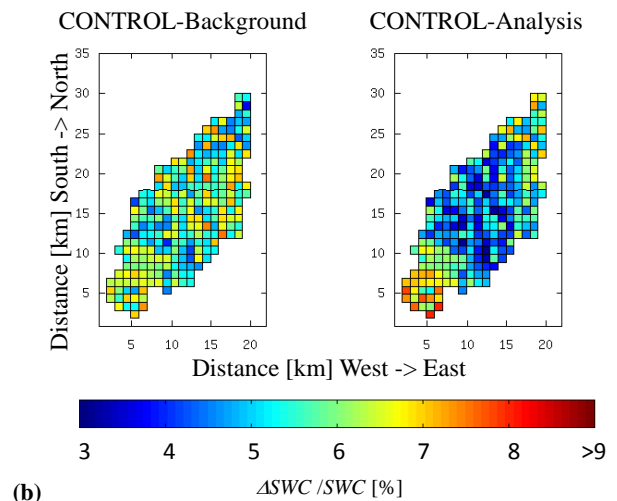


**Fig. 8.** TERRA-ML's soil column is 2.43 m deep. The soil water content (SWC) of each grid cell depends on its soil depth and its soil moisture (Eq. 9). The SWC is displayed for the initial time  $t=0$  (5 May 1997) for the background, CONTROL and analysis. The analysis at  $t=0$  was obtained assimilating streamflow from the CONTROL model simulation from Pforzheim from  $t=0$  to  $t=56$  h with a 0.5 hourly timestep.



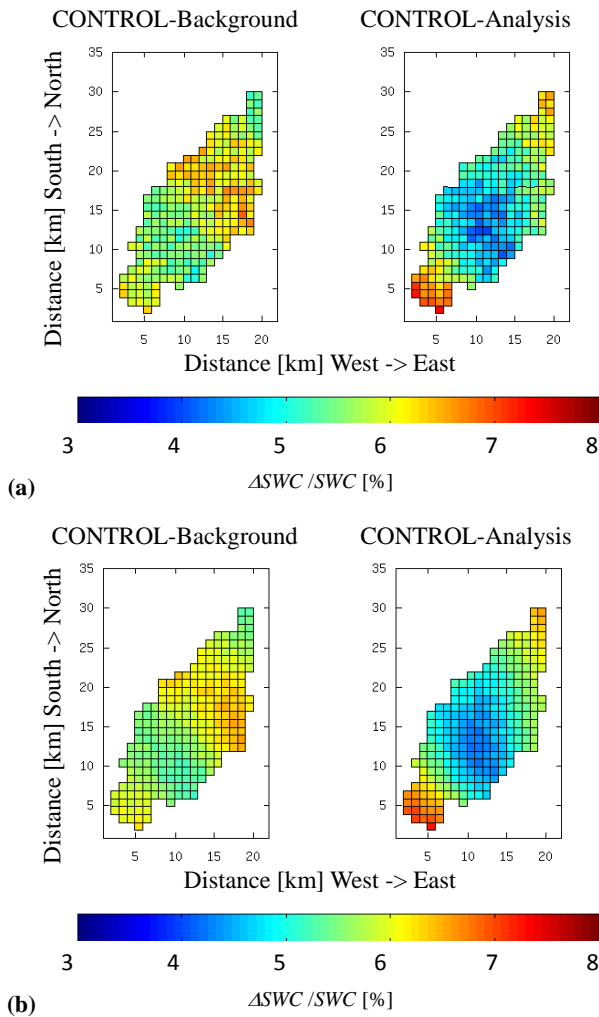
**Fig. 9.** The difference in soil-water content (SWC) of each grid cell relative to the CONTROL SWC of each grid cell for the initial time  $t=0$  (5 May 1997) for the background and analysis. The analysis at  $t=0$  was obtained assimilating streamflow from the CONTROL model simulation from Pforzheim from  $t=0$  to  $t=56$  h with a 0.5 hourly timestep.

The promising results from the 260 km<sup>2</sup> catchment led to a study about the potential impact of a denser network of gauges for the soil moisture analysis. Gauges were assumed to be at the outlet of the *Große Enz* (90 km<sup>2</sup>), the outlet of the *Kleine Enz* (71 km<sup>2</sup>) and the outlet of the *Eyach* (43 km<sup>2</sup>). Little impact was reached for the *Eyach*, but for all other catchments the SWC was improved. Figures 12, 13, and 14 show the results for the *Große Enz* catchment. Here, the impact of the streamflow data assimilation is much more pronounced, as can be seen from Figs. 11 and 13. The ensemble mean SWC at  $t=0$  is 535 kg/m<sup>2</sup> for the



**Fig. 10.** As Fig. 9, but (a) for the top 3 soil layers of TERRA-ML, i.e. 0–0.09 m depth of the soil, and (b) for the 4th soil layer of TERRA-ML, i.e. 0.09–0.27 m depth of the soil.



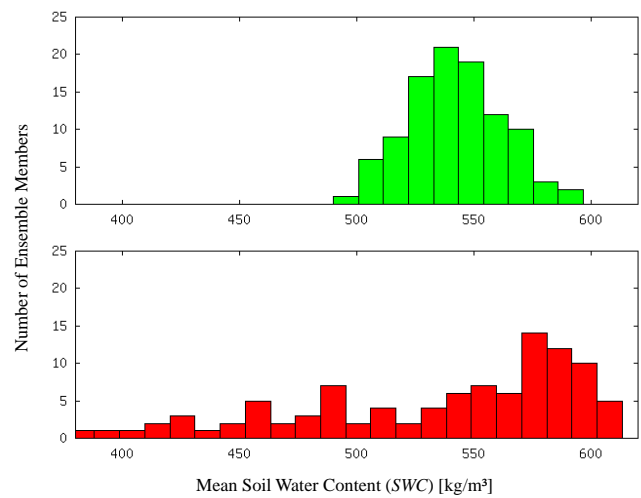


**Fig. 11.** As Fig. 9, but (a) for the 5th soil layer of TERRA-ML, i.e. 0.27–0.81 m depth of the soil, and (b) for the 6th soil layer of TERRA-ML, i.e. 0.81–2.43 m depth of the soil.

background, 542 kg/m<sup>2</sup> for the analysis and 568 kg/m<sup>2</sup> for the CONTROL. Figures 13 and 14 show that nowhere does the analysis lead to worse SWCs than the background.

Figure 15 shows the impact of assimilating streamflow from the CONTROL model simulation from the *Große Enz* outlet, *Kleine Enz* outlet and Eyach with a 0.5 hourly time step. Most areas show a positive impact of the data assimilation. Applying the mean of the soil moisture analysis of each layer and grid cell after the assimilation of *Große Enz*, *Kleine Enz*, Eyach and Pforzheim results in a slight (5%) improvement of the streamflow at Pforzheim (Fig. 6c) due to the improvement of the soil moisture.

All in all, the simulations show a gradient in the impact of the data assimilation. Close to the gauge location of the assimilated streamflow and at the furthest upstream grid cells, the data assimilation shows worse results than in the middle



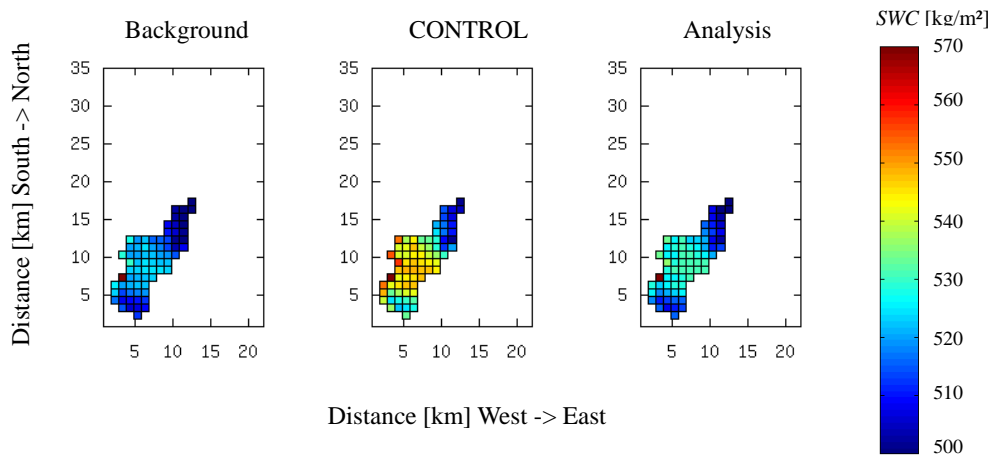
**Fig. 12.** TERRA-ML’s soil column is 2.43 m deep. The soil-water content (SWC) of each grid cell depends on its soil depth and its soil moisture (Eq. 9). Distribution of initial mean SWC ( $t=0$ ) in the *Große Enz* catchment (90 km<sup>2</sup>) between the ensemble members. CONTROL mean SWC is 568 kg/m<sup>2</sup>. The analysis at  $t=0$  was obtained assimilating streamflow from the CONTROL model simulation from Pforzheim from  $t=0$  to  $t=56$  h with a 0.5 hourly timestep.

areas. This is due to flow duration in the river network and the assimilation window. The grid cells close to the gauge would need shorter assimilation windows. However, the OSSE shows that the streamflow data assimilation has the potential to improve the soil moisture throughout the catchment and that a more dense gauging network would help to improve this even further.

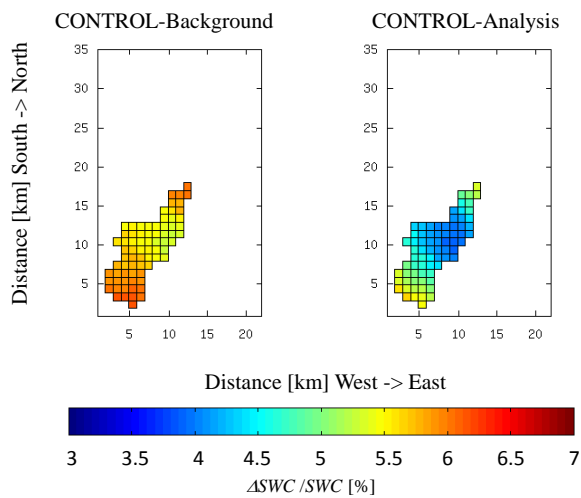
## 5 Conclusions

Numerical weather forecasting and climate modelling require an accurate soil moisture initialization for their land surface models. So far, the areal distribution of root zone soil moisture cannot be measured. Streamflow depends on the soil moisture of a river catchment and is measured at gauging stations of the rivers at relatively high accuracy.

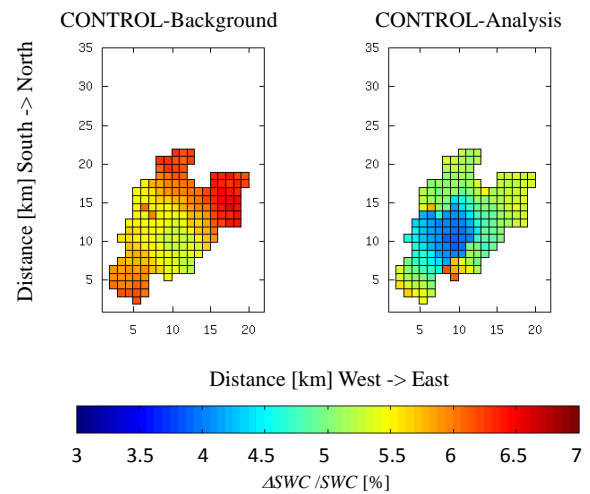
A retrospective EnKF was set up to assimilate streamflow into the multi-layer land surface model TERRA-ML of the regional weather forecast model COSMO. An OSSE was performed in the Enz River catchment located at the downwind side of the northern Black Forest (Germany). The results confirm the potential of streamflow data assimilation for improving soil moisture analyses. Further, we discussed the spatial and temporal requirements for an automated river gauging network. Half-hourly streamflow data is available from the automated gauges of the flood forecast centre of Baden-Württemberg (Germany) for approximately 140 gauges. Half-hourly resolution of streamflow data is



**Fig. 13.** TERRA-ML's soil column is 2.43 m deep. The soil-water content (SWC) of each grid cell depends on its soil depth and its soil moisture (Eq. 9). The SWC is displayed for the initial time  $t=0$  (5 May 1997) for the background, CONTROL and analysis. The analysis at  $t=0$  was obtained assimilating streamflow from the CONTROL model simulation from the *Große Enz* outlet from  $t=0$  to  $t=34$  h with a 0.5 hourly timestep. Note that the scaling is different from Fig. 8.



**Fig. 14.** The difference in soil-water content (SWC) of each grid cell relative to the CONTROL SWC of each grid cell for the initial time  $t=0$  (5 May 1997) for the background and analysis. The analysis was obtained assimilating streamflow from the CONTROL model simulation from *Große Enz* outlet from  $t=0$  to  $t=34$  h with a 0.5 hourly timestep.



**Fig. 15.** The difference in soil-water content (SWC) of each grid cell relative to the CONTROL SWC of each grid cell for the initial time  $t=0$  (5 May 1997) for the background and analysis. The analysis was obtained assimilating streamflow from the CONTROL model simulation from the *Große Enz* outlet, *Kleine Enz* outlet and *Eyach* outlet with a 0.5 hourly timestep.

sufficient for its assimilation for soil moisture analysis. In the upper Enz, an automated gauge is operational at Höfen. The OSSE shows that streamflow from this location can already improve SWC in the Enz catchment upstream of Höfen, but that a denser network would improve the SWC even more. Namely, at the outlets of smaller sub-catchments, like the *Große Enz*, this would be valuable, since the sub-catchments show a differently structured river network (Fig. 2) and flow duration (Fig. 4). Since the necessary assimilation window depends on the catchment size (e.g. 56 h for Pforzheim and

27 h for the *Kleine Enz*), a denser gauging network would shorten the assimilation time making it even more valuable for initialisation in numerical weather forecast models.

Warrach-Sagi et al. (2008) showed, for the study area, that the streamflow simulated with TERRA-ML underestimates the observation. This is due to model errors, land surface heterogeneity, spatial variability of meteorological conditions and errors in meteorological forcing data set, and soil and vegetation parameter uncertainty. Model errors may be, to a large extent, estimated applying TERRA-ML at

meteorological stations where also soil moisture, soil temperature and eddy correlation measurements are available. This, for example, is done during EVAGRIPS and published by Ament and Simmer (2006) and Johnsen et al. (2005). But the heterogeneity of the land surface and weather poses a large source of uncertainty. The usually underestimated streamflow will also increase the soil moisture when measured data is applied. Since the OSSE shows that the concept works, it is also expected that assimilating observed streamflow will cause a more realistic soil moisture pattern, namely during underestimated streamflow this will lead to larger soil moisture in the catchment. The catchment was chosen with care though, it is not dominated by deep groundwater flows, otherwise streamflow data assimilation should not be expected to be a suitable tool for soil moisture analysis in a land surface model like TERRA-ML.

Altogether the retrospective EnKF is a powerful method to assimilate streamflow data into a land surface model for root zone soil moisture analysis. The implementation of the square root algorithm for EnKF from Evensen (2004) is straight forward and can be used with any land surface model if a river routing model is attached.

An improved soil moisture is the first step in improving the simulation of the water fluxes. An immediate positive impact on the simulation of fluxes and atmospheric variables cannot be expected, but the optimization of initial fields is the first important step. With the improved soil moisture it will be possible to improve the parameterizations that are responsible for energy balance equations by means of re-analyses. Of course, the golden goal would be the assimilation of streamflow and other soil and atmospheric variables into a coupled atmosphere-land surface model system, e.g., COSMO-TERRA-ML or WRF-NOAH. This study is considered as a first step towards this direction and demonstrates that it is principally a possible path to follow.

*Acknowledgements.* We thank Dag Lohmann for the latest version of the river routing scheme and Felix Ament and Clemens Simmer for the stand-alone version of TERRA-ML. We thank the assistance from LUBW, namely Manfred Bremicker, Frank Eberspächer and Angela Sieber for providing the meteorological and hydrological data and the LGRB soil data in digital form for the project. Further, we highly appreciate the free access to the EnKF codes of Geir Evensen and its good documentation on <http://enkf.nersc.no>. Kirsten Warrach-Sagi thanks the German Research Foundation (DFG) for funding the project STREAMDATA within the Priority Program quantitative precipitation forecast (SPPI167).

Edited by: D. Lawrence

## References

- Ament, F. and Simmer, C.: Improved representation of land-surface heterogeneity in a non-hydrostatic numerical weather prediction model, *Bound.-Lay. Meteorol.*, 121(1), 153–174, 2006.
- Bárdossy, A. and Lehmann, W.: Spatial Distribution of Soil Moisture in a Small Catchment, Part 1: Geostatistical Analysis, *J. Hydrol.*, 206, 1–15, 1998.
- Beven, K. J.: *Rainfall-runoff modelling – the primer*, John Wiley and Sons Ltd, Chichester, 360 pp., 2001.
- Campbell, G. S.: A simple method for determining unsaturated conductivity from moisture retention data, *Soil Sci.*, 117(6), 311–314, 1974.
- 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, *Adv. Water Resour.*, 31, 1309–1324, 2008.
- Conil, S., Douville, H., and Tyteca, S.: Contribution of realistic soil moisture initial conditions to boreal summer climate predictability, *Clim. Dynam.*, 32(1), 75–93, 2009.
- Crow, W. T. and Wood, E. F.: The assimilation of remotely sensed soil brightness temperature imagery into a land-surface model using ensemble Kalman filtering: a case study based on ESTAR measurements during SGP97, *Adv. Water Resour.*, 26, 137–149, 2003.
- Doms, G., Förstner, J., Heise, E., Herzog, H.-J., Raschendorfer, M., Schrodin, R., Reinhardt, T., and Vogel, G.: A description of the non-hydrostatic regional model LM Part II: Physical parameterization, *Deutscher Wetterdienst, Offenbach*, 133 pp., 2005.
- Drusch, M., Wood, E. F., and Jackson, T.: Vegetative and atmospheric corrections for the soil moisture retrieval from passive microwave remote sensing data: Results from the Southern Great Plains Hydrology Experiment 1997, *J. Hydrometeorol.*, 2, 181–192, 2001.
- Drusch, M. and Viterbo, P.: Assimilation of Screen-Level Variables in ECMWF's Integrated Forecast System: A Study on the Impact on the Forecast Quality and Analyzed Soil Moisture, *Mon. Weather Rev.*, 135, 300–314, 2007.
- Dunne, S. and Entekhabi, D.: Land Surface State and Flux Estimation Using the Ensemble Kalman Smoother During the Southern Great Plains 1997 Field Experiment, *Water Resour. Res.*, 42(1), W01407, doi:10.1029/2005WR004334, 2006.
- Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, 99, 10143–10162, 1994.
- Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, *Ocean Dynam.*, 53, 343–367, 2003.
- Evensen, G.: Sampling strategies and square root analysis schemes for the EnKF, *Ocean Dynam.*, 54, 539–560, 2004.
- Evensen, G.: *Data Assimilation: The Ensemble Kalman Filter*, Springer, 280 pp., 2006.
- Gao, H., Wood, E. F., Drusch, M., and McCabe, M. F.: Copula-Derived Observation Operators for Assimilating TMI and AMSR-E Retrieved Soil Moisture into Land Surface Models, *J. Hydrometeorol.*, 8, 413–429, 2007.
- Graßelt, R., Warrach-Sagi, K., Schüttemeyer, D., Ament, F., and Simmer, C.: Influence of precipitation forcing on discharge simulation in the Sieg river catchment, *Meteorol. Z.*, 17, 763–773, 2008.
- Grayson, R. B. and Western, A. W.: Towards areal estimation of soil water content from point measurements: time and space stability of mean response, *J. Hydrol.*, 207, 68–82, 1998.

- Hess, R.: Assimilation of screen level observations by variational soil moisture analysis, *Meteorol. Atmos. Phys.*, 77, 145–154, 2001.
- Hohenegger, C., Brockhaus, P., and Schär, C.: Towards climate simulations at cloud-resolving scales, *Meteorol. Z.*, 17, 383–394, 2008.
- Hohenegger, C., Brockhaus, P., Bretherton, C. S., and Schär, C.: The Soil Moisture-Precipitation Feedback in Simulations with Explicit and Parameterized Convection, *J. Climate*, 22, 5003–5020, 2009.
- Johnsen, K.-P., Mengelkamp, H.-T., and Huneke, S.: Multi-objective calibration of the land surface scheme TERRA/LM using LITFASS-2003 data, *Hydrol. Earth Syst. Sci.*, 9, 586–596, 2005, <http://www.hydrol-earth-syst-sci.net/9/586/2005/>.
- Kabat, P., Hutjes, R. W. A., and Feddes, R. A.: The scaling characteristics of soil parameters: From plot scale heterogeneity to subgrid parameterization, *J. Hydrol.*, 190, 363–396, 1997.
- Komma, J., Blöschl, G., and Reszler, C.: Soil moisture updating by Ensemble Kalman Filtering in real-time flood forecasting, *J. Hydrol.*, 357, 228–242, 2008.
- LfU: Arbeitsanleitung Pegel- und Datendienst Baden-Württemberg – Durchflussermittlung mit Messflügeln, Landesanstalt für Umweltschutz Baden-Württemberg, Karlsruhe, Germany, 100 pp., 2002.
- Lohmann, D., Nolte-Holube, R., and Raschke, E.: A large-scale horizontal routing model to be coupled to land surface parameterization schemes, *Tellus A*, 48, 708–721, 1996.
- Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Cosgrove, B. A., Sheffield, J., Duan, Q., Luo, L., Higgins, W., Pinker, R. T., and Tarpley, J. D.: Streamflow and water balance intercomparisons of four land-surface models in the North American Land Data Assimilation System project, *J. Geophys. Res.*, 109, D07S91, doi:10.1029/2003JD003517, 2004.
- Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P.: Dual State-Parameter Estimation of Hydrological Models using Ensemble Kalman Filter, *Adv. Water Resour.*, 28, 135–147, 2005.
- Pauwels, V. R. N. and DeLannoy, G. J. M.: Improvement of modeled soil wetness conditions and turbulent fluxes through the assimilation of observed discharge, *J. Hydrometeorol.*, 7, 458–477, 2006.
- Pitman, A. J., Dolman, A. J., Kuijij, B., Valentini, R., and Baldocchi, D.: The Climate near the ground, in: *Vegetation, Water, Humans and the Climate: A new perspective on an interactive system*, edited by: Kabat, P., Claussen, M., Dirmeyer, P. A., Gashk, J. H. C., Bravo de Guenni, L., Meybeck, M., Pielke, R. A., Vorosmarty, C. J., Hutjes, R. W. A., and Lutkemaier, S., The IGBP Series, Springer, 9–19, 2004.
- Reichle, R. H., Mc Laughlin, D. B., and Entekhabi, D.: Hydrologic data assimilation with the Ensemble Kalman filter, *Mon. Weather Rev.*, 130, 103–114, 2002.
- Rijtema, P. E.: Soil moisture forecasting, Instituut voor Cultuurtechniek en Waterhuishouding, Wageningen, Technical Report Nota 513, 1969.
- Rotach, M. W., Arpagaus, M., Dorninger, M., Hegg, C., Frick, J., Montani, A., Ranzi, R., Bouttier, F., Buzzi, A., Frustaci, G., Mylne, K., Richard, E., Rossa, A., Schär, C., Staudinger, M., Volkert, H., Wulfmeyer, V., Bauer, H.-S., Ament, F., Denhard, M., Fundel, F., Germann, U., and Stoll, M.: MAP D-PHASE: Real-time Demonstration of Weather Forecast Quality in the Alpine Region, *B. Am. Meteorol. Soc.*, 90, 1321–1336, 2009.
- Schär, C., Lüthi, D., Beyerle, U., and Heise, E.: The soil-precipitation feedback: A process study with a regional climate model, *J. Climate*, 12, 722–741, 1999.
- Seneviratne, S. I., Lüthi, D., Litschi, M., and Schär, C.: Land-atmosphere coupling and climate change in Europe, *Nature*, 443, 205–209, 2006.
- Seuffert, G., Wilker, H., Viterbo, P., Drusch, M., and Mahfouf, J. F.: The Usage of Screen-Level Parameters and Microwave Brightness Temperature for Soil Moisture Analysis, *J. Hydrometeorol.*, 5, 516–531, 2004.
- Trier, S. B., Chen, F., and Manning, K. W.: A Study of Convection Initiation in a Mesoscale Model Using High-Resolution Land Surface Initial Conditions, *Mon. Weather Rev.*, 132, 2954–2976, 2004.
- Walker, J. P., Willgoose, G. R., and Kalma, J. D.: Three-dimensional soil moisture profile retrieval by assimilation of near-surface measurements: Simplified Kalman filter covariance forecasting and field application, *Water Resour. Res.*, 38, 1–18, 2002.
- Warrach-Sagi, K., Wulfmeyer, V., Grasselt, R., Ament, F., and Simmer, C.: Streamflow simulations reveal the impact of the soil parameterization, *Meteorol. Z.*, 17, 751–762, 2008.
- Wulfmeyer, V., Behrendt, A., Bauer, H.-S., Kottmeier, C., Corsmeier, U., Blyth, A., Craig, G., Schumann, U., Hagen, M., Crewell, S., Di Girolamo, P., Flamant, C., Miller, M., Montani, A., Mobbs, S., Richard, E., Rotach, M. W., Arpagaus, M., Russchenberg, H., Schlüssel, P., König, M., Gärtner, V., Steinacker, R., Dorninger, M., Turner, D. D., Weckwerth, T., Hense, A., and Simmer, C.: The Convective and Orographically-induced Precipitation Study: A Research and Development Project of the World Weather Research Program for improving quantitative precipitation forecasting in low-mountain regions, *B. Am. Meteorol. Soc.*, 89/10, 1477–1486, doi:10.1175/2008BAMS2367.1, 2008.