Implementation of Surface Soil Moisture Data Assimilation with Watershed Scale Distributed Hydrological Model

Eunjin Han[1], Venkatesh Merwade[2*] and Gary C. Heathman[3]

[1]Graduate Research Assistant, School of Civil Engineering, Purdue University, West Lafayette, IN. e-mail: ehan@purdue.edu

[2]Assistant Professor, School of Civil Engineering, Purdue University, West Lafayette, IN; *Corresponding author;* e-mail: vmerwade@purdue.edu

[3]Soil Scientist, USDA-ARS, National Soil Erosion Research Laboratory, West Lafayette, IN; email: Gary.Heathman@arcs.usda.gov

Abstract

This paper aims to investigate how surface soil moisture data assimilation affects each hydrologic process and how spatially varying inputs affect the potential capability of surface soil moisture assimilation at the watershed scale. The Ensemble Kalman Filter (EnKF) is coupled with a watershed scale, semi-distributed hydrologic model, the Soil and Water Assessment Tool (SWAT), to assimilate surface (5 cm) soil moisture. By intentionally setting inaccurate precipitation with open loop and EnKF scenarios in a synthetic experiment, the capability of surface soil moisture assimilation to compensate for the precipitation errors were examined. Results show that daily assimilation of surface soil moisture for each HRU improves model predictions especially reducing errors in surface and profile soil moisture estimation. Almost all hydrological processes associated with soil moisture are also improved with decreased Root Mean Square Error (RMSE) values through the EnKF scenario. The EnKF does not produce as much a significant improvement in streamflow predictions as compared to soil moisture estimates in the presence of large precipitation errors and the limitations of the infiltration-runoff model mechanism. Distributed errors of the soil water content also show the benefit of surface soil moisture assimilation and the influences of spatially varying inputs such as soil and landuse types. Thus, soil moisture update through data assimilation can be a supplementary way to overcome the errors created by inaccurate rainfall. Even though this synthetic study shows the potential of remotely sensed surface soil moisture measurements for applications of watershed scale water resources management, future studies are necessary that focus on the use of real-time observational data.

Key words: Soil Moisture; SWAT; Data assimilation; Ensemble Kalman Filter: Cedar Creek, Indiana

1. Introduction

One of the key variables in understanding land surface hydrologic processes is soil moisture because it controls the infiltration-runoff mechanism and energy exchange at the land-atmosphere boundary. Therefore, estimating soil moisture has been a long-standing research topic for various purposes in many areas: weather forecast in atmospheric science, flood or drought prediction in hydrology, water quality management in environmental science, irrigation operations in agricultural engineering and soil erosion in soil science (Walker et al., 2001).

Since the 1990s, remotely sensed soil moisture data has become much more available while overcoming limitations of traditional *in-situ* point measurements of soil moisture (Jackson et al., 1995; Njoku and Entekhabi, 1996; Wagner et al., 1999; Kerr et al., 2001; Njoku et al., 2003; Entekhabi et al., 2008). Remotely sensed data provide soil moisture estimates for the top soil layer (~5cm). However, information on the moisture condition in the root-zone and subsurface layers is more critical for understanding and simulating many hydrologic processes including evapotranspiration, surface runoff and subsurface flow. The need for profile soil moisture estimates has motivated researchers to integrate measured surface data and hydrologic models to obtain more accurate estimates of soil moisture content in the root zone through data assimilation techniques (Reichle et al., 2002b; Ni-Meister et al., 2006; Reichle et al., 2007; Sabater et al., 2007; Das et al., 2008; Draper et al., 2009). However, those surface soil moisture assimilation studies have been conducted at regional or global scales with land surface models in hydrometeorology for better initialization of soil moisture conditions. Even though great progress in surface soil moisture assimilation studies has been made in land surface atmospheric interactions, there is still a lack of research on utilizing remotely sensed soil moisture and data assimilation techniques for catchment scale water resource management problems (Troch et al., 2003).

The main objective of this study was to investigate the effect of surface soil moisture data assimilation on hydrological response at the catchment scale through a synthetic experiment. Especially, using intentionally limited rainfall input, we investigate how soil moisture update through surface soil moisture assimilation may compensate for errors in the hydrologic prediction due to the inaccurate rainfall. Simply focusing on the streamflow prediction overlooks the different contributions of each water balance component such as evapotranspiration, infiltration, surface runoff and lateral flow to the streamflow. Therefore, in this study, a physically-based catchment scale, continuous time, semi-distributed hydrologic

model, the Soil and Water Assessment Tool (SWAT), is used to determine and account for the sources of error in streamflow prediction. In addition, this study also aims to investigate the effects of spatially varying input such as landuse and soil type on data assimilation results.

In this study, one of the most popular data assimilation techniques, the Ensemble Kalman Filter (EnKF) is used to assimilate surface soil moisture observations into the model and a synthetic experiment is conducted assuming that uncertainties in the model and observations are known. Previous studies related to this study are summarized in section 1.1. Brief explanations about the EnKF and the SWAT model are described in section 2, followed by the illustration of how we conducted the synthetic assimilation experiments in section 3. Section 4 shows the results of the experiments with discussions.

1.1. Previous studies

This section briefly summarizes previous studies in the context of why surface soil moisture data assimilation at catchment scale is important, specifically for runoff prediction and agricultural applications.

Satellite-based surface soil moisture observations have received considerable attention in runoff or flood forecasts because antecedent soil moisture condition is a critical factor in rainfall-runoff modeling. The recently launched European Space Agency (ESA)'s Soil Moisture and Ocean Salinity **(**SMOS**)** mission (2009) and the upcoming NASA Soil Moisture Active/Passive (SMAP) mission (2014) are designed to better measure soil moisture on a global scale (Entekhabi et al., 2010; Kerr et al., 2010). One of the main application areas of SMAP is to improve flood forecasts using soil moisture measurements at high spatial and temporal resolutions (Entekhabi et al. 2010). Some previous studies have demonstrated the potential of using remotely sensed soil moisture to improve streamflow prediction through updating initial soil moisture conditions and finding correlation between soil moisture condition and runoff (Jacobs et al., 2003; Scipal et al., 2005; Weissling et al., 2007).

With regard to assimilating remotely sensed surface soil moisture into rainfall-runoff models, however, there are few studies to date. Pauwels et al. (2001) assimilated remotely surface soil moisture data into the TOPMODEL based Land-Atmosphere Transfer Scheme (TOPLATS) using the ‘nudging to individual observations’ and ‘statistical correction assimilation’ methods. They applied these methods for both the distributed and lumped versions of the model and concluded that improvement in the discharge prediction can be sufficiently obtained through assimilating the statistics of the remotely sensed soil moisture data into the lumped model.

Crow and Ryu (2009) adopted a smoothing framework for runoff prediction and used remotely sensed soil moisture to improve both pre-storm soil moisture conditions and external rainfall input. They showed that their smoothing framework improved streamflow prediction, especially for high flow events more than simply updating antecedent soil moisture conditions.

Microwave remote sensing soil moisture observations have high potential for agricultural applications. Some obvious examples are for crop yield forecasting, drought monitoring and early warning, insect and disease control, optimal fertilizer application, and operational decision-making for effective irrigation (Engman, 1991; Lakhankar *et al.*, 2009). However, very few studies exist on the application of remotely sensed soil moisture for agricultural operations. Jackson et al. (1987) tested the feasibility of using airborne microwave sensors to assess the preplanting soil moisture condition by creating a soil moisture map which represents overall soil moisture patterns and variations expected over large areas. Recently, soil moisture data from the Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E) was incorporated into a global agricultural model by Bolten et al (2009). Their preliminary results showed that surface soil moisture assimilation has promising potential to improve crop yield prediction capability of a global crop production decision support system.

Soil moisture information with high spatial and temporal resolution is also required for watershed or field scale agricultural applications. Timely and cost effective operational decisions for on-farm irrigation and trafficability should be based on field-specific information (Jackson *et al.*, 1987). Narasimhan et al. (2005) used the Soil and Water Assessment Tool (SWAT) to produce a long-term soil moisture dataset for the purpose of drought monitoring and crop yield prediction. They selected SWAT model because of its capability: 1) for simulating crop growth and land management and, 2) for incorporating all available spatial information for the watershed. SWAT model has also been successfully used to examine the temporal variability of soil moisture over longer time periods (e.g. 30 years) with the detailed land surface information (DeLiberty and Legates, 2003).

Based on the aforementioned previous studies, there is a strong need to estimate soil moisture content through assimilating remotely sensed soil moisture into a long-term, physically based distributed catchment scale hydrologic model. Most of the previous studies that explored data assimilation for runoff simulation used conceptual rainfall-runoff models (Aubert et al., 2003; Weerts and El Serafy, 2006; Crow and Ryu, 2009; van Delft et al., 2009) or lumped models (Jacobs et al. 2003) or for short-term periods with real measurements (Pauwels et al. 2001). From an agricultural aspect, soil moisture reserve between rainfall events

is also critical for scheduling water supply for crops. Therefore, it is desirable to apply data assimilation techniques to physically based and continuous time hydrological models to address various water resource problems at catchment scales.

Recently, SWAT has been used to assess the capability of the EnKF for catchment scale modeling (Xie and Zhang, 2010; Chen et al., 2011). Xie and Zhang (2010) explored combined state-parameter estimation using different types of measurements based on a synthetic experiment and demonstrated effective update of hydrological states and improved parameter estimation(CN_2) using the EnKF. Chen et al. (2011) conducted both synthetic and real data EnKF experiments. Their results showed improved update for soil moisture in the upper soil layers, but limited success for deeper soil layers and streamflow prediction due to the insufficient vertical coupling strength of SWAT.

2. Hydrologic Model and Data Assimilation

2.1 Soil and Water Assessment Tool (SWAT)

The Soil and Water Assessment Tool (SWAT) is categorized as a physically based basin-scale, continuous time and semi-distributed hydrologic model. Spatially distributed data related to soil, landuse, topography and weather are used as input to SWAT. The SWAT model has the capability to simulate plant growth, nutrients, pesticides and land management as well as hydrology on a daily time step and has been proven as an effective tool for assessing water resource and nonpoint-source pollution problems after being applied to watersheds of different scales and characteristics (Gassman *et al.*, 2007). In addition, it is considered as one of the promising models for long-term simulations in predominantly agricultural watersheds (Borah and Bera, 2003), similar to the study area, Upper Cedar Creek Watershed in Indiana, used in this paper. Considering the above factors, SWAT is used to accomplish the objectives of this study. The 2005 version of the SWAT model is used in this study, and detailed information of SWAT 2005 can be found in Neitsch et al. (2002).

In the SWAT model, a watershed is first subdivided into sub-basins based on topography, and each sub-watershed is further divided into hydrologic response units (HRU) based on soil and landuse characteristics. The soil profile can be subdivided into multiple layers for up to 2 meters. Hydrological processes in the SWAT including soil moisture accounting are based on the water balance equation in Eq. (1).

$$SW_t = SW_{t-1} + \sum_{i=1}^{t}\left(P_i - Q_{surf,i} - ET_i - Q_{loss,i} - Q_{gw,i}\right) \quad (1)$$

where SW_t is the soil water content above the wilting point at the end of day *t*. P_i is the amount of precipitation on day *i* and $Q_{surf,i}$, ET_i, $Q_{loss,i}$ and $Q_{gw,i}$ are the daily amounts of surface runoff, evapotranspiration, percolation into deep aquifer, and lateral subsurface flow, respectively. All components are estimated as mmH2O.

SWAT simulates surface runoff using either the modified SCS curve number (CN) method or Green Ampt Mein-Larson excess rainfall method (GAML) depending on the availability of daily or hourly precipitation data, respectively (Neitsch et al. 2002). In this study, the modified SCS CN method is used with daily precipitation data. The SCS CN method computes accumulated surface runoff (Q_{surf}) using the empirical relationship between rainfall (P_i), the initial abstraction (I_a) and retention parameter (S_i) as shown in Eq. (2). SWAT follows the typical assumption of $I_a = 0.2 \cdot S_i$

$$Q_{surf,i} = \frac{(P_i - I_a)^2}{(P_i - I_a + S_i)} = \frac{(P_i - 0.2S_i)^2}{(P_i + 0.8S_i)} \quad (2)$$

S_i is a function of the daily curve number(CN_i). Both S_i and CN_i vary spatially with the soil type, landuse management type, and slope.

$$S_i = 25.4\left(\frac{1000}{CN_i} - 10\right) \quad (3)$$

Since the daily curve number varies with the antecedent soil moisture condition, SCS defines three different curve numbers: CN_1-dry, CN_2-average moisture, and CN_3-wet. This classification, however, is too coarse to reflect antecedent soil moisture condition accurately. Thus SWAT adopted a new equation to compute S_i as a function of soil profile water content (SW_i).

$$S_i = S_{\max,i}\left(1 - \frac{SW_i}{[SW_i + \exp(w_1 - w_2 \cdot SW_i)]}\right) \quad (4)$$

where $S_{\max,i}$ is the maximum retention parameter that can be achieved on any given day. w_1 and w_2 are shape coefficients determined from the amount of water in the soil profile at field capacity and when fully saturated, and retention parameters for CN_1 and CN_3 conditions (Neitsch *et al.*, 2002).

USDA-SCS (USDA-SCS, 1972) states that the SCS curve number method is designed to estimate "direct runoff" which is composed of channel runoff, surface runoff and subsurface

flow, excluding base flow. However, SWAT 2005 uses the CN method to estimate "surface runoff" (Eq. (2)) and uses other equations to compute subsurface lateral flow and groundwater flow (Neitsch et al. 2002). Subsequently, the sum of the surface runoff, subsurface lateral flow and groundwater flow generates streamflow. Thus, according to Neitsch et al. (2002), this study uses the CN method in estimating only surface runoff (Eq. (2)).

The excess water available after initial abstractions and surface runoff infiltrates into the soil. The flow through each layer is simulated using a storage routing technique. Unsaturated flow between layers is indirectly modeled using the depth distribution of plant water uptake and soil water evaporation. Saturated flow is directly simulated and assumed that water is uniformly distributed within a given layer. When the soil water in the layer exceeds field capacity and the layer below is not saturated, downward flow occurs and its rate is governed by the saturated hydraulic conductivity. A kinematic storage routing technique is used to simulate lateral flow in the soil profile based on slope, hillslope length and saturated conductivity. Upward flow from a lower layer to an upper layer is simulated by the soil water to field capacity ratios of the two layers.

2.2 Ensemble Kalman Filter (EnKF)

In hydrology, data assimilation techniques have been used to improve model predictions by combining uncertain observations and imperfect information from hydrological processes represented in hydrological models (Walker and Houser, 2005). Among various data assimilation methods, the standard Kalman Filter is a sequential data assimilation method for linear dynamics minimizing the mean of the squared errors in state estimates. For nonlinear applications, the extended Kalman filter has been applied (Entekhabi et al., 1994; Walker and Houser, 2001; Draper et al., 2009), but this method requires high computational cost for the error covariance integration and is known as very unstable if the nonlinearities are severe (Miller et al., 1994; Reichle et al., 2002a). Evensen (1994) introduced the Ensemble Kalman Filter (EnKF) and proved that it could successfully handle strongly nonlinear systems with low computational cost unlike the extended Kalman Filter. The EnKF has gained popularity in hydrologic data assimilation and a number of previous studies have demonstrated its performance in improving hydrological predictions (Reichle et al., 2002a; Reichle et al., 2002b; Zhang et al., 2006; Clark et al., 2008; Komma et al., 2008; Xie and Zhang, 2010). In this study, the well-proven EnKF method is used to assimilate surface soil moisture observations into the SWAT model.

Hydrological processes such as infiltration, evapotranspiration and drainage to groundwater system influence soil water content in the root zone in a highly nonlinear manner. The state variable, soil moisture (θ_k) is a vector containing soil moisture values for each soil layer in a HRU and is described with a nonlinear model operator, $f_k(\cdot)$ at time step k.

$$\theta_{k+1} = f_k(\theta_k) + w_k \quad (5)$$

In the Eq. (5), w_k represents all uncertainties in the forcing data and model formulation due to the numerical approximation and imperfect knowledge of the hydrologic processes.

The surface soil moisture measurement (d_k) has error (v_k) due to the errors in the observational instruments or procedures. It is explained using the observation operator (H_k) and true state (θ_k) as follows:

$$d_k = H_k \theta_k + v_k \quad (6)$$

The EnKF is based on the concept of the statistical Monte Carlo method, where the ensemble of system states marches in state space and the mean of the ensemble is considered as the best estimate of the state. The initial ensemble of state vectors is generated with initial error vector (e_i) with N ensemble members ($i=1, ...,N$).

$$\theta_0^{i+} = \theta_0 + e_i \quad (7)$$

The EnKF approach assumes that the error terms, w, v and e, are white noise (uncorrelated in time) and have the Gaussian distribution with zero mean and covariance Q, R, and P respectively.

Once the initial ensemble is created, the ensemble of state vectors integrates forward in time through the nonlinear model operator (Eq. (8)) and is updated using the Kalman Gain (K_k) whenever the observations are available (Eq. (9)).

$$\theta_k^{i-} = f_{k-1}(\theta_{k-1}^{i+}) + w_{k-1}^i \quad (8)$$

$$\theta_k^{i+} = \theta_k^{i-} + K_k\left[d_k - H_k\theta_k^{i-} + v_k^i\right] \quad (9)$$

where superscripts '-' and '+' refer to the forecasted and updated state variables respectively. In the update step, the observation is perturbed with random errors (v_k^i), following Burgers et al. (1998). The Kalman gain (K_k) works as a weighting factor between uncertainties in model prediction and observations, and is determined by the state error covariance (P_k^-) and covariance matrix of the model predicted observation ($H_k P_k^- H_k^T$).

$$K_k = P_k^- H_k^T\left[H_k P_k^- H_k^T + R_k\right]^{-1} \quad (10)$$

Unlike the traditional Kalman Filter, the EnKF does not propagate the state error covariance

(P_k^-) explicitly. Instead, the EnKF simply estimates the error covariance using the forecasted ensemble of state and its mean (Eq. (11)), thus improving computational efficiency.

$$P_k^{-1} \approx \frac{1}{N-1}\sum_{i=1}^{N}\left[\left(\theta_k^{i-}-\theta_k^-\right)\left(\theta_k^{i-}-\theta_k^-\right)^T\right]^{-1} \qquad (11)$$

where $\theta_k^- = \frac{1}{N}\sum_{i=1}^{N}\theta_k^{i-}$

The average of the updated ensemble members is determined as a best estimate of the state variable.

3. Methodology

3.1 Study Area and Data for SWAT Simulation

The study area for this work is the Upper Cedar Creek Watershed (UCCW) which is located in the St. Joseph Watershed in northeastern Indiana (Fig. 1). The predominant landuse in the UCCW is agricultural, with major crops of corn and soybeans, and minor crops of winter wheat, oats, alfalfa, and pasture. The area receives approximately 94 cm of annual precipitation and has average daily temperatures ranging from -1 °C to 28 °C.

The United States Department of Agriculture, Agricultural Research Service (USDA-ARS) National Soil Erosion Research Laboratory (NSERL) has established an environmental monitoring network in UCCW as a part of the Source Water Protection Project and the Conservation Effect Assessment Project. The environmental monitoring network in UCCW has been operational since 2002. More details about the network can be found in Flanagan et al. (2008). The UCCW environmental monitoring network provides considerable meteorological data (10 minute rainfall, air temperature, solar radiation, wind speed and relative humidity) and every 10 minute soil moisture observations at four different depths (5 20, 40 and 60 cm). Recently, UCCW has been selected as one of several USA core sites to be used for calibration and validation of future SMAP data. Substantial *in-situ* data, and its importance for future remote sensing validation work makes UCCW an ideal test bed for this study.

In addition to the data from the monitoring network, daily precipitation and temperature data are available from nearby three National Climatic Data Center (NCDC) weather stations for the model simulation period, from 2003 to 2010. (Fig. 1).

For the SWAT simulation, sub-basins and stream network are delineated from the 30 meter Digital Elevation Model (DEM) from the National Elevation Dataset. ArcSWAT, an ArcGIS

interface for SWAT, is used for delineating the sub-basins and creating other input files for the model. By using a critical threshold area (CSA) of 500 ha, the UCCW is divided into 25 sub-basins. Further division of sub-basins into HRU units requires soil and land use information. Soil information is retrieved from the Soil Survey Geographic Database (SSURGO) available from the USDA Natural Resources Conservation Service. Major soil types in the UCCW and their areas are summarized in Table 1. Landuse information is obtained from the 2001 National Land Cover Data (NLCD 2001) produced by the Multi-resolution Land Characteristics Consortium. The NLCD 2001 data for the UCCW is reclassified to 15 different landcover classes for this study (Table 2). By using a 10% threshold for both landuse and soil, and 0% threshold for slope, a total of 209 HRUs are generated for the UCCW using ArcSWAT.

3.2 Experiment setup

3.2.1 Synthetic experiment

Data assimilation study requires reliable observed data and decent knowledge on the uncertainties in the model and observations. In the present study, a synthetic experiment is performed to better understand how soil moisture data assimilation affects various hydrologic processes in the watershed scale SWAT model. The synthetic experiment assumes that errors in the model and observations are known, which allows us to focus only on the effect of data assimilation. Thus, this synthetic experiment at watershed scale should contribute to the practical application of the forthcoming remotely sensed soil moisture data.

In this study, it is assumed that surface soil moisture observation for each HRU is available and the synthetic observed data is created by adding random observation errors with zero mean. Considering that the typical microwave remote sensing data have a penetrating depth up to 5cm, we also assume that the synthetically observed soil moisture is available for the top 5cm depth and the input SSURGO soil database is modified so that each HRU has the first two layers in the 5cm depth interval.

3.2.2 Scenarios setup

Our experiment includes three separate scenarios all run for the same time period: a true run, open loop and the EnKF. First the experiment starts with the implementation of the “true” state by running the model with all available rainfall data from the three NCDC stations and the seven NSERL rain gauges (Fig. 1). To represent our imperfect knowledge of the true hydrologic processes, the model is subsequently run with an intentionally poor set of initial conditions and with “limited” forcing data from only NCDC rain gauges. Hereafter, this is called the “open

loop" scenario. The third scenario, referred to as "EnKF" includes model integration with all the same input including rainfall and model parameters as the open loop, but updating soil moisture by assimilating daily (synthetic) observed surface soil moisture through the EnKF. By limiting precipitation input, which is the driving force of soil moisture and streamflow, while keeping other model parameters unchanged, the EnKF scenario enables the determination of the influence of the surface soil moisture assimilation on model predictions of profile soil water content and other hydrological responses.

Conventionally, precipitation data measured at point rain gauges have been used for hydrological modeling, despite its limitations in representing the spatial distribution and variability of actual precipitation. In reality, many watersheds rely on the precipitation data from the sparsely located rain gauges inside or around the watershed as our open loop scenario represents. Therefore, in this study, we investigate how the surface soil moisture assimilation may compensate for the errors from the inaccurate precipitation input by comparing the aforementioned three scenarios (true, open loop and EnKF).

SWAT assigns precipitation values at the subbasin level based on precipitation data measured at rain gauges. Each subbasin takes the precipitation value from the gage station that is closest to the centroid of the subbasin. Unfortunately, there is no NCDC rain gauge inside the UCCW. Therefore, when the precipitation data from the only three NCDC rain gauges are used (open loop and EnKF), SWAT assigns overestimated precipitation into the watershed (Fig. 2). The average precipitations of the true scenario and the open loop scenario (and EnKF) are 2.66 mm day^{-1} and 3.21 mm day^{-1}, respectively during the analysis period (from June 2008 to May 2009).

3.2.3 SWAT calibration

In order to create the true scenario, the SWAT model was calibrated using the daily streamflow measured at the outlet of the UCCW. The model ran for three years (April 2003 to April 2006) for warm-up and two years (May 2006 to May 2008) for calibration. Based on the sensitivity analysis results, 18 parameters were chosen for calibration including CN2, ESCO, SURLAG, TIMP, Ch_K2, SOL_AWC, BLAI, ALPHA_BF, SOL_Z, RCHRG_DP, SMTMP, EPCO, CH_N2, REVAPMN, SLOPE, SOL_ALB, CANMX, SOL_K. They were calibrated using the autocalibration tool, Parasol (Parameter Solutions) method, implemented in the ArcSWAT. After the calibration, the model was validated for one year (June 2008 to May 2009). The coefficient of determination (R^2) and the Nash and Sutcliffe model efficiency coefficient (E_{NS}) are 0.46 and 0.44 for calibration period and 0.42 and 0.37 for validation period, respectively with the daily streamflow predictions. Although these statistical metrics seem to

indicate marginal model performance, they are within an acceptable range at the daily time step (Moriasi et al. 2007). This validation period will be used for our data assimilation experiment. For the true scenario, the model runs from 2003 to 2009 with the calibrated parameters, and the simulation results only from the last one year (June 2008 to May 2009) are used for analysis. For the open loop and EnKF scenarios, the model runs from January 2008 to May 2009 with an intentionally short warm-up period for poor initial conditions and the same calibrated parameters as the true scenario. Again the last one year simulation results (June 2008 to May 2009) are used for analysis.

3.2.4 Evaluation metrics

Evaluation of the surface soil assimilation through the EnKF is first conducted by comparing time series soil water storage and other representative hydrological responses obtained from the true, open loop and EnKF scenarios. Root mean square error (RMSE), mean bias error (MBE) and correlation coefficient (R) are used to compare the results quantitatively.

$$RMSE = \sqrt{\frac{\sum (T - S)^2}{n}}$$

$$MBE = \frac{\sum (T - S)}{n}$$

$$R = \frac{\sum \left(T - \bar{T}\right)\left(S - \bar{S}\right)}{\sqrt{\sum \left(T - \bar{T}\right)^2} \sqrt{\sum \left(S - \bar{S}\right)^2}}$$

where T and S are true and simulated values respectively, and n is the number of data (days) $\bar{T}$ and $\bar{S}$ are averages of true and simulated values respectively.

In addition, distributed soil moisture maps are shown to better visualize and illustrate the spatial differences in the results and to determine the effect of spatially varying input.

3.3 Implementation of the EnKF into SWAT

3.3.1 Ensemble simulation with SWAT

The SWAT model runs on a daily time step. This study assumes that the daily soil moisture observations are available, and therefore soil moisture condition is reinitialized for each day through the analysis (update) step of the EnKF. The framework of hydrologic processes in the SWAT and additional routines due to the EnKF are illustrated in Fig. 3. The default SWAT routines in Fig. 3 are excerpted from Neitsch et al. (2002). Each HRU runs independently and one dimensional EnKF scheme is applied to each HRU separately.

Ensemble simulation starts on day 153 (June 1) of 2008 and 100 initial ensemble members are created by adding random noise to the initial condition (soil moisture simulation results from day 152). SWAT represents soil water content as a concept of soil water storage in mmH_2O (vs $m^3\ m^{-3}$) and therefore the range of soil moisture values for each layer varies with the thickness of the soil layer. The initial perturbation noise is assigned to the normalized soil moisture values (mmH_2O per 1cm). The initial perturbation noise has a Gaussian distribution with zero mean and 0.001 mmH_2O of standard deviation, and the magnitude of the standard deviations are scaled by the thickness of each layer according to Ryu et al. (2009). Note that forcing variables (e.g. precipitation) and model parameters are not perturbed explicitly to generate ensemble members in this study. Adding model error (w_k in the Eq. (5)) takes into account all uncertainties raised from forcing variables and parameter estimation, as well as model physics.

Each of the ensemble members of soil moisture goes through separate hydrological processes and generates different subsequent variables such as Leaf Area Index and evapotranspiration. All subsequent variables from different ensemble members are also averaged after the ensemble forecasting step and the average values are considered as the best estimations of the day.

3.3.2 Specification of model and observation errors

At the end of each day, the predicted ensemble soil moisture values are updated through the EnKF analysis step. In the analysis step, model errors (w_k in the Eq. (5)) with zero mean Gaussian distribution are added. The standard deviations of the model errors are determined by the current soil moisture content from the true scenario using a scaling factor of 0.01. Observations are also perturbed with the observation errors (v_k in the Eq. (9)) which have zero mean Gaussian distribution and standard deviation of 0.01.

Procedures for soil moisture perturbation by adding system variance (model error) are described as follows:

1) Compute a weight (wl_j) for each layer (j) in order to take account for the different thicknesses of each layer (Ryu et al., 2009).
2) Compute model noise (w_k in the Eq. (5)) by applying a scaling factor (0.01) and a weight (wl_j) for each layer.

$$w_j = wl_j \cdot \left(0.01 \cdot \theta_{true,j}\right) \cdot r_j$$

where $\theta_{true,j}$ is soil moisture estimation from the true scenario and r_j is a random number from

a Gaussian distribution with zero mean and variance one for $j=1,\ldots,n$ where n is the number of soil layers.

For soil moisture perturbation, a constant scaling factor is applied to all layers every daily time step. Vertical correlation between perturbations is also an important factor in determining how surface soil moisture assimilation affects deeper layers and other dependent variables. Chen et al. (2011) showed the impact of error coupling with the SWAT model by comparing the results from zero vertical error correlation and perfect vertical error correlation. In the present study, perfect vertical error correlation was applied.

Several past studies have attempted to specify observation and model error statistics. Clark et al. (2008) and Xie and Zhang (2010) used the scaling factor approach for straightforward estimation of model and observation error. Reichle et al. (2002b) assigned temporally constant standard deviations of errors. The scaling factor approach used in this study allows standard deviation of the model errors to vary with time, which is a better representation of the real uncertainties than the time invariant standard deviation of errors. Since our synthetic experiment uses the same model parameters as the true scenario, the source of model error is mainly from errors in the precipitation. Therefore, in this study, a simple but more operational approach of using the scaling factor is adopted for model error estimation. This approach overestimates the real errors when the soil moisture content is high, but is advantageous to test the robustness of the EnKF (Xie and Zhang, 2010). The disadvantage of this approach, however, is that it may enhance the nonlinear impact of the saturation (or wilting point) threshold by inflating (or decreasing) the standard deviation of soil moisture predictions. In spite of convenient application of the scaling factor approach, more sophisticated approaches such as an adaptive filtering seem to be desirable for future real data assimilation studies.

Time invariant standard deviation of observation errors is adopted in this study because errors in remotely sensed soil moisture come from vegetation cover, surface roughness, soil properties, radio frequency interference (RFI), and retrieval algorithms, all of which are not proportional to the surface soil moisture condition (Schmugge et al., 2002; Entekhabi et al., 2010). In addition, if the same scaling factor approach is applied to the observation error, it will assign unreasonably high weight on the observation accuracy when the soil is very dry because SWAT defines soil water content excluding the amount of water held at wilting point, with the minimum soil water content in the SWAT being zero.

3.3.3 Additional steps in analysis procedure

The bounded nature of soil moisture between wilting point and saturated soil water content

makes the application of the EnKF more complicated. Reichle et al. (2002a) mentioned that in an operational perspective, the violation of its inherent assumption, Gaussian distribution, would have the greatest impact when the soil is very dry and there is high skewed forecast error. Crow and Wood (2003) showed that the skewed model ensembles and a non-Gaussian error structure negatively impacted the EnKF's performances. Because of this unique characteristic of soil moisture, some additional steps are added after the analysis step. When the soil is very dry or wet (close to the wilting point or saturated condition), the best estimates of the soil moisture (average of ensemble) after the analysis step may exceed the actual physical limits of the soil moisture. If the average of the ensemble of the soil moisture becomes negative, the best estimate of the ensemble is adjusted to 10^{-6} mmH_2O. In the case of an ensemble average higher than the maximum soil moisture limit, saturated soil moisture content minus 10^{-6} mmH_2O replaces the best estimate.

The concept of a simple bias correction method adopted by Ryu et al. (2009) is implemented in this study to take account of the effect of the bounded range of soil moisture. Mean bias is computed using the unperturbed soil moisture prediction and the ensemble is corrected by subtracting the mean bias from the perturbed soil moisture ensemble. After this adjustment, all ensemble members exceeding minimum or maximum soil water content are replaced by 10^{-6} mmH_2O (almost wilting point) or saturated water content. This boundary truncation might shift the average of ensemble again. Therefore, these bias corrections and boundary truncations are repeated 10 times to reduce the remaining bias. Boundary truncation and the simple bias correction approach are applied at each update time step for all state variables (soil moisture vector). A simple boundary truncation might cause the mean of the state variable to be shifted. For example, for a wet soil moisture condition, the simple boundary truncation may shift the mean of soil moisture higher than the actual mean from the analysis step. Therefore, this repetition of boundary truncation and bias correction may be advantageous in generating less mass balance error by decreasing the biases (Ryu et al. 2009).

4. Results and Discussion

4.1 Effect of surface soil moisture assimilation on hydrologic processes

This section describes how surface soil moisture data assimilation affects subsequent hydrological processes. First, in this study, inaccurate precipitation is the main source of error in soil moisture predictions as well as other hydrological processes. Fig. 4 shows the daily simulation results of surface (a) and profile (b) soil moisture in the watershed which is the sum

of area-weighted soil water content from all HRUs. Soil moisture prediction errors are significant especially during the winter time (from day 353 of the first year to day 67 of the second year) due to the cumulatively overestimated precipitation shown in Fig. 2b. There exist distinct discrepancies in soil moisture predictions between the true and the open loop scenario. Soil moisture update through the EnKF draws the inaccurate soil moisture prediction in the open loop scenario closer to the true state. The correlation coefficient increases from 0.585 to 0.747 and from 0.906 to 0.942 for surface and profile soil moisture respectively (Table 3). The RMSE and MBE also decrease in both surface and profile soil moisture predictions with the EnKF. The improved results with the EnKF support the previous studies (Das et al. 2008; Draper et al. 2009; Reichle et al. 2002b; Sabater et al. 2007) further demonstrating the potential of current and forthcoming remotely sensed surface soil moisture data to improve the profile soil moisture estimations for land surface hydrology through data assimilation. The results of this study also show that the errors in soil moisture estimation due to inaccurate precipitation can be partially compensated by accommodating surface soil moisture observations.

Temporal variations in the RMSE of the subsequent variables are shown in Fig. 5 where daily RMSE is based on the errors of all HRUs. It is apparent that reduced errors in surface soil moisture prediction (Fig. 5a) are clearly identified for almost every time step. In spite of the prompt decrease of errors in surface soil moisture prediction, the magnitude of improvements in the profile soil water content varies with time as shown in Fig. 5b with reduction in errors with the EnKF being distinct in winter (from day 353 to 67). Limited success in updating profile soil water content for some periods may be caused by the weak model vertical coupling of soil moisture in SWAT subsurface physics (Chen et al., 2011). Another reason is that non-linearities in model physics and the bounded nature of soil moisture result in suboptimal update with the EnKF by violating the Gaussian assumption. When optimality of ensemble perturbation is checked as in Reichle et al. (2002a), non-symmetric ensemble distribution is found with the soil moisture condition close to saturation or wilting points (not shown). In addition, consistently lower forecast and analysis error variances compared with actual error were found, resulting in estimates less than optimal.

The improved soil moisture prediction affects other subsequent hydrological responses. Table 3 shows that the EnKF scenario reduces errors in other hydrological variables compared to the open loop scenario even though the correlation coefficient for some variables may be slightly lower (e.g., SHALLST, QDAY, RCHRG, GW_Q and GWSEEP).

Similar to the profile soil moisture in Fig. 5b, SHALLST (depth of water in shallow aquifer)

in Fig. 5c and GW_Q (groundwater contribution to streamflow) in Fig. 5f are not noticeably affected except during winter conditions. Little differences in RMSE are observed in the prediction of QDAY (surface runoff) in Fig. 5d and LATQ (lateral flow) in Fig. 5e. Since accurate prediction of surface runoff (QDAY) and groundwater contribution (GW_Q) are critical to streamflow prediction, it is expected from these results that surface soil moisture assimilation may not improve streamflow prediction significantly. Further discussion of these finding is provided in section 4.2. Application of EnKF reduced the RMSE in evapotranspiration prediction throughout the experiment period (Fig. 5g). This is because evaporation occurs mainly on the soil surface and the improved surface soil moisture condition contributes to greater accuracy in simulating evaporation. Finally, the curve number for a day (CNDAY) is computed based on the profile soil water content. Therefore, the trend of error reduction in the curve number prediction (Fig. 5h) is very similar to the trend of the profile soil moisture (Fig. 5b).

4.2 Streamflow prediction

The answer to the question "Can surface soil moisture data assimilation improve streamflow prediction?" depends on the accuracy of precipitation input and antecedent soil moisture condition. In order to answer to this question, four simple and different cases can be considered accounting for only antecedent soil moisture conditions and the accuracy of precipitation data, and assuming that the EnKF always improves soil moisture predictions close to the true values.

1) Model predicted antecedent soil moisture is less than the true soil moisture ($\theta_{predicted} < \theta_{true} \approx \theta_{EnKF}$) and current precipitation is overestimated.
2) Model predicted antecedent soil moisture is greater than the true soil moisture ($\theta_{predicted} > \theta_{true} \approx \theta_{EnKF}$) and current precipitation is overestimated.
3) Model predicted antecedent soil moisture is less than the true soil moisture ($\theta_{predicted} < \theta_{true} \approx \theta_{EnKF}$) and current precipitation is underestimated.
4) Model predicted antecedent soil moisture is greater than the true soil moisture ($\theta_{predicted} > \theta_{true} \approx \theta_{EnKF}$) and current precipitation is underestimated.

Cases 2 and 3 would provide improvement in streamflow prediction by updating soil moisture through the EnKF. In Case 2, the overestimated antecedent soil moisture condition will aggregate the error from the overestimated precipitation and therefore, updated (corrected to lower) soil moisture with the EnKF will generate less error than the open loop. The same principle can be applied to Case 3 where improved runoff prediction with the EnKF is expected

compared to the open loop. However, Cases 1 and 4 will generate more errors with the EnKF than with the open loop. In Case 1, the underestimated antecedent soil moisture condition counterbalances the error in the overestimated precipitation. Therefore, improved runoff prediction occurs with the open loop rather than with the EnKF in this case. The same principle is applied to Case 4. Therefore, a hypothesis, "Assimilating surface soil moisture observation into a hydrologic model will improve streamflow (or surface runoff) prediction" is valid only if we have accurate precipitation information. Crow and Ryu (2009) pointed out this limitation of updating solely the antecedent soil moisture condition and designed an assimilation system that simultaneously updates soil moisture and corrects rainfall input by taking into account soil moisture observations.

Case 1 occurs in the results of this study during the summer, especially around day 259. The EnKF improves soil moisture prediction (Fig. 4b and Fig. 5b). However, because of inaccurately overestimated precipitation, improved soil moisture with the EnKF (higher soil moisture prediction than the open loop results) results in overestimated runoff and streamflow (Fig. 5d and Fig. 6). The lower model performance using the EnKF is primarily due to inaccurate precipitation. Therefore, in this case, surface soil moisture assimilation does not appear to improve streamflow prediction.

Results during winter time represent Case 2 where both soil moisture and precipitation are overestimated. Application of the EnKF improves streamflow prediction during winter time (Fig. 6, day 42 to 87). This is because the open loop continues overestimating soil moisture during winter, but the EnKF improves the soil moisture prediction (lowers the overestimated soil moisture in Fig. 4). Even though the improvement in runoff prediction (QDAY) due to the improved soil moisture with the EnKF is not significant (Fig. 5d), the improved groundwater contribution (GW_Q) also contributes to the improved (reduced) streamflow in Fig. 5f.

In this study, streamflow prediction does not show significant improvement even after applying surface soil moisture assimilation. The primary reason for this is that most of the errors in streamflow prediction are due to inaccurate precipitation input (Fig. 6, Fig. 7 and Table 4). Second, SWAT model physics does not have sufficient vertical coupling strength to constrain soil water content in deeper layers or in the root zone using the surface soil moisture observations (Chen et al., 2011). Since surface runoff generation depends on profile soil water content, failure to significantly improve estimates of profile soil water content impede successful surface runoff (or streamflow) prediction. Therefore, improvement in streamflow prediction is not expected when the improvement of profile soil water content with the EnKF

is of little or marginal consequence as shown in Fig. 5b. Lastly, the unsuccessful improvement in streamflow prediction may be attributed to the limited sensitivity of surface runoff (or streamflow) prediction to the change in soil moisture with the CN method implemented in SWAT. Surface runoff is the main contributor to streamflow especially with high rainfall intensity. Therefore, if the updated soil water content with the EnKF is not reflected properly to the surface runoff estimation, improved streamflow prediction cannot be expected.

In order to test model sensitivity in these regards, the relationship between surface runoff generation and different soil moisture conditions for various curve numbers is illustrated in Fig. 8 based on Eq. (2), (3) and (4). The soil type, GnB2, of which water storage at field capacity and saturation are 101 mm and 201 mm respectively for a soil depth of 1246 mm, is used to create Fig.8. In the case of high curve numbers (CN_2=93), a 20 mm change in soil moisture from 200 to 180 mm results in reducing surface runoff by only 2 mm. This decrease in runoff prediction will result in a 4.57 m^3 sec^{-1} streamflow decrease, which is not sufficient to overcome the significant streamflow overestimation during the winter in Fig. 6.

As Fig. 8 shows, the sensitivity of surface runoff to soil water content varies with the curve number, which is a function of soil type and landuse. This study also shows that the degree of improvement in surface runoff prediction with the EnKF is highly related to soil type and landuse. The areas which have low runoff error with the EnKF consist of soil types BoB, Hw, SrB2, StC3 and Se (Table 1). Hydrologic soil groups of those soil types are A or B (Table 1). Furthermore, the combinations of those soil types and landuse (HAY and FRSD) provide the smallest errors because of their hydrologic soil group (A or B) and low CN_2 (35 ~60). On the contrary, the main three soil types (BaB2, GnB2 and Pe) which cover 67% of the area in the UCCW produce high errors in surface runoff prediction regardless of landuse type. Their CN_2 is relatively high (77 ~ 89). For most of the areas in UCCW the EnKF, therefore, cannot decrease the error in the runoff prediction because surface runoff is not very sensitive to the soil moisture change with high CN_2 values. That is, the slope of the graph is small for high CN_2 values in Fig. 8. Especially when soil moisture is high (close to saturation, 200mm in case of GnB2), the slope is very small. During the winter time, highest soil moisture estimation errors exist and the EnKF improves profile soil moisture up to 20 mm in Fig. 4b, but in this period, the soil moisture condition is very wet, so its improvement is difficult to reflect in surface runoff.

In this study, surface soil moisture assimilation was minimally successful in improving streamflow predictions. However, if we consider only streamflow results, traditional

calibration methods can improve streamflow simulation considerably, even in the presence of inaccurate precipitation. In SWAT, streamflow prediction has a higher sensitivity to changes in certain parameters rather than changing soil moisture conditions. Parameter adjustments through the calibration can change streamflow prediction effectively by increasing the E_{NS}. However, conventional calibration methods do not directly take into account the uncertainties in precipitation or model structure, or the possibility of model parameters that change temporally. Data assimilation techniques usually focus on updating state variables (soil moisture in this study) on the premise that model parameters are pre-specified. Therefore, new frameworks which can address both the limitations of conventional parameter calibrations and data assimilation are required for future advances in hydrologic modeling. Moradkhani et al. (2005), for example, showed the possibility of estimating both model states and parameters simultaneously using a dual state-parameter estimation method based on the EnKF.

4.3 CN method vs. Green Ampt method

Even though the CN method has been widely used in various hydrologic models, it has limitations. Garen and Moore (2005) stresses some issues with the broad use of the CN method. Since the CN method was developed as an event model for the prediction of flood streamflow conditions, they argue that it is questionable to apply the CN method for a continuous model and daily flow of ordinary magnitude. In addition, they also mention that daily time step might not be appropriate to simulate the infiltration excess mechanism which is designed for hourly (or subhourly) time steps.

SWAT 2005 provides another option to estimate surface runoff, the Green Ampt Mein-Larson excess rainfall method (GAML). This method is based on the Green & Ampt infiltration method and requires subdaily precipitation input. The CN method has been more widely used than the GAML method because of the difficulties in obtaining subdaily precipitation data and uncertain benefit of using the GAML compared to the CN method. While Kannan et al. (2007) and King et al. (1999) found no significant advantages of using the GAML instead of the CN method with SWAT, Jeong et al. (2010) showed that the GAML outperforms the CN method and suggested that the quality of subdaily precipitation data and the size of study area influence the results. In addition, Jeong et al. (2010) also showed that even the subhourly simulation performs poorly under low or medium flows.

In this study, a simple experiment is conducted to test if the GAML method has a greater potential than the CN method for improving runoff (and eventually streamflow) prediction with

surface soil moisture data assimilation. As mentioned in section 4.2, sensitivity of surface runoff generation to the soil moisture change determines how successfully the soil moisture update through the EnKF will contribute to the improvement of the surface runoff prediction. For the two main soil types, GnB2 and SrB2, surface runoff was computed by using the CN method and the GAML method for one day (June 21, 2008) when precipitation was 23mm day^{-1}. For the GAML method, 10 minute precipitation data were used. The amount of water held in soil profile at field capacity is 101.5 and 240 mmH_2O for GnB2 and SrB2, respectively. Water content at saturation is 210 and 400 mmH_2O for GnB2 and SrB2, respectively.

Fig. 9 shows the different runoff-soil moisture relationships from the CN and GAML methods. The CN method maintains a smooth relationship between runoff and soil moisture, however, the GAML method produces drastic changes when soil moisture is near field capacity (101.5 and 240 mmH_2O for GnB2 and SrB2 respectively). This is because the GAML code implemented in SWAT2005 replaces any soil moisture value that is greater than field capacity soil moisture with the field capacity value in computing infiltration rate. Therefore, under higher soil moisture conditions, the GAML method produces much less surface runoff than CN method, which is appropriate in reducing the overestimated streamflow. However, the slope of the graph for GAML is much smaller than the slope of the CN method above field capacity. Therefore, the impact of the improved soil moisture with the EnKF might be difficult to see with the GAML method when the soil moisture is above field capacity. However, soil moisture less than the field capacity has a high sensitivity to changes in soil moisture. The two different soil types show different shapes of the relationship because of their different field capacity and saturated water content values. Therefore, effectiveness of the soil moisture assimilation with the GAML depends on the soil type and soil moisture condition. In using either the CN method or GAML method, it is very difficult and somewhat complicated to determine precisely how changes in soil moisture affect streamflow prediction.

Successful improvements of streamflow prediction through soil moisture data assimilation can be expected only if the model is based on correct linkage between soil moisture conditions and surface runoff generation. Even though SWAT 2005 adopts a modified CN method which accounts for antecedent soil moisture condition continuously, its effectiveness is not thoroughly proven. In addition, application of the CN method has been questioned by many studies and its modification for runoff simulation has been introduced (Michel et al., 2005; Mishra and Singh, 2006; Chung et al., 2010; Sahu et al., 2010), specifically with SWAT (Kannan et al., 2008; Kim and Lee, 2008; Wang et al., 2008; White et al., 2010). In addition, various factors such as

quality of rainfall data, rainfall intensity, soil type, landuse and identification of critical source area make streamflow prediction more complicated. Therefore, more careful selection and development of runoff simulation procedures effectively taking account for soil moisture variations should be required to enhance streamflow prediction with future soil moisture assimilation studies.

4. 4 Spatial variation in soil moisture prediction

In this section, the impact of spatially varying input, specifically landuse and soil type, on surface soil moisture assimilation is illustrated. To visualize spatial distribution of output, an HRU map is created by overlaying landuse and the SSURGO soil map generated from ArcSWAT. Since a single slope is defined in the initial SWAT setup, the slope is not taken into account in creating the HRU map. Then, HRU level outputs are assigned to each HRU to show spatially distributed results.

Precipitation as a forcing variable is the most important factor for successful soil moisture estimation. Fig. 10 shows time-averaged RMSE of precipitation. Time-averaged RMSE results of surface and profile soil moisture prediction are shown in Fig. 11 and 12 respectively. The RMSE distribution of the open loop and the EnKF, especially for profile soil moisture distribution, coincides with the distribution of precipitation in general. That is to say, higher precipitation errors within an area result in greater errors in soil moisture estimation.

The EnKF reduces errors in surface and profile soil moisture compared to the results of the open loop. The average RMSE errors in surface soil moisture estimates in Fig. 11 are 5.05 and 3.43 mmH_2O for the open loop and the EnKF, respectively. For profile soil moisture estimation, the open loop and the EnKF have 22.77 and 19.77 mmH_2O of average RMSE, respectively in the Fig. 12.

Within a subbasin where constant precipitation is assigned, types of landuse and soil determine the magnitude of errors in the soil moisture estimation. One distinct example is shown in subbasin 2 in Fig. 11a. Even though the subbasin has same amount of precipitation throughout the area, some of the areas (light blue in the Fig. 11a) have highly underestimated surface soil moisture (much higher RMSE) than others. The areas consist of HRU 11(FRSD, Hw), 18(AGRR, Hw) and 21(WETF, GnB2). The soil type Hw is classified as hydrologic soil group A (Table 1) which has a high infiltration rate.

Fig. 13 compares the results of different combinations of landuse and soil types. Fig. 13b is the reference because it consists of the major landuse (AGRR) and soil type (GnB2).

Comparing Fig. 13b and 13c explains the effect of different soil types on soil moisture estimation. Although landuse type and precipitation are same, Fig. 13c which has more infiltration rate because of the soil type Hw, exaggerates both the underestimated and overestimated errors compared to the soil type GnB2 in Fig. 13b.

Types of landuse also affect the soil moisture variations. Fig. 13a and 13d show that landuse types, forest (FRSD) and wetland (WETF), produce greater errors than agricultural area. Proportion of precipitation intercepted by canopy is large in forest area. Therefore, the actual precipitation that reaches the ground (after canopy interception) in this area may be much smaller than other landuse areas, which will exaggerate errors in soil moisture estimation. Further explanation regarding the errors due to different landuse types follows in section 4.5.

4.5 Spatiotemporal variation in soil moisture prediction

Aforementioned results show significant effects of errors in precipitation on hydrologic responses. In this section, the effect of precipitation is excluded by selecting a drydown period when no precipitation is received throughout the watershed. Therefore, the effectiveness of the EnKF can be further evaluated.

Fig. 14c and 14d show the antecedent surface soil moisture condition in terms of deviations (errors) from the true condition on day 258 before the precipitation events on day 259 shown in Fig. 14a and 14b. The western areas of the UCCW have overestimated soil moisture (negative error) and eastern areas have underestimated states (positive error) on day 258. On day 259 in Fig. 14e and 14f, highly overestimated precipitation in subbasins 1, 3, 4, 5, 6 and 7 do not make much differences in soil moisture status because soil moisture is already overestimated (close to saturated condition) and the overestimated precipitation becomes surface runoff instead of infiltrating into the soil and increasing soil moisture. However, slightly overestimated precipitation in subbasins 2, 8, 14, 15, 18, 19, 20, 21 and 24 infiltrates into soil and decreases the errors in the previously underestimated soil moisture.

After the precipitation on day 259, there is no precipitation at all throughout the watershed until day 263. While errors in surface soil moisture estimation in the open loop do not change significantly day by day, the EnKF results show noticeably decreasing errors in one or two days. In other words, errors in the EnKF results become close to zero sooner than the open loop results. In regard to the profile soil moisture variations, the EnKF produces better results (smaller errors in general) than the open loop results even though the magnitudes of improvements are not as much as the ones with the surface soil moisture (not shown).

Interestingly, some areas in the open loop give soil moisture results that remain highly underestimated (high positive errors with blue color) even after the precipitation on day 259 (Fig. 14g, i and k). Those areas correspond to certain distributions of landuse type, such as forest (FRSD) and hay (HAY). A large portion of initial precipitation is intercepted by canopy in this area and thus, actual precipitation generating surface runoff and infiltration on the ground becomes very small. Table 5 shows the amount of initial precipitation that is intercepted by the leaves of plants; less than 20% of initial precipitation reaches the ground surface in the forest area in the case of the open loop and the EnKF. The amount of canopy interception is computed based on the values of leaf area index, amount of water held in canopy storage and potential leaf area index. The difference in the results between the open loop and the EnKF arise from the differences in the simulation results of those variables. In SWAT, those variables related to crop growth are, therefore, very important in simulating surface runoff and subsurface flow as well as soil moisture because actual precipitation being a forcing variable is affected greatly by the canopy interception.

Contrary to the open loop, surface soil moisture assimilation through the EnKF prevents this long-lasting underestimation for certain landuse types (Fig. 14h, j and l). The results of profile soil moisture estimation in the open loop scenario also show the same problem, high underestimation with a specific landuse type (not shown). These long-lasting underestimated (dry) soil moisture conditions will affect simulation of crop growth and other hydrologic processes such as plant water uptake and subsurface flow in the long term. The errors due to specific landuse types can be minimized to a certain degree by optimizing parameters through calibration. However, considering that most calibrations for a distributed model are achieved by focusing on the streamflow and overall water balance, it is difficult to overcome this issue for specific soil types. However, the soil moisture update through the EnKF enables us to overcome this problem and to prevent further accumulation of errors in hydrologic predictions by improving soil moisture estimates.

5. Conclusion

In this study, a synthetic experiment was conducted to investigate how surface soil moisture assimilation affects hydrological processes in the Upper Cedar Creek Watershed using the SWAT hydrologic model and the Ensemble Kalman Filter. This study compared three scenarios: 1) a true scenario with no errors in model, precipitation and soil moisture observations; 2) an open loop scenario with limited precipitation information; and 3) the EnKF scenario with the

same imperfect information as the open loop but assimilating observed surface (~5cm) soil moisture every day using the EnKF data assimilation technique.

Soil moisture update through the EnKF improved surface and profile soil moisture estimations compared to the open loop. In addition, the EnkF improved predictions of other subsequent hydrological variables with reduced errors even though the magnitude of the improvements varied according to different variables and rainfall accuracy. The most significant improvements in the prediction of those variables were found during the winter months due to large errors in cumulatively overestimated precipitation.

The capability of surface soil moisture assimilation for improving streamflow prediction was constrained by the accuracy of precipitation and characteristics of the rainfall-runoff mechanism in SWAT. Highly overestimated or underestimated streamflow peaks corresponded to the inaccurate rainfall events. Updated soil moisture conditions after applying the EnKF does not always lead to improvements in surface runoff (or streamflow) predictions. This is because in using the curve number method, sensitivity of surface runoff generation to changes in soil moisture is affected by various factors such as soil type, landuse type, rainfall intensity, CN_2 and antecedent soil moisture conditions. A simple experiment in this study showed that both the CN method and the Green-Ampt method in the SWAT have some limitations in reflecting soil moisture updates into the surface runoff routine. Therefore, it is essential to have accurate precipitation input to achieve improved streamflow predictions through surface soil moisture assimilation using SWAT. This fact was the motivation behind several previous studies for exploring methods to improve precipitation input using soil moisture retrievals (Crow and Bolten, 2007; Crow et al., 2009; Crow et al., 2011). Second, greater effort is needed in developing better excess rainfall simulation algorithms to overcome the limitations of the current CN method so that continuous model processes under various rainfall intensities are taken into account.

Distributed RMSE maps from the surface and profile soil moisture predictions illustrated the effects of different landuse and soil types on soil moisture estimation. The EnKF results in much less errors in soil moisture estimation than the open loop scenario throughout the watershed. However, depending on the soil and landuse type, the magnitude of reduced errors with the EnKF varied. When we removed the effect of inaccurate precipitation and examined a certain drydown period, the EnKF scenario performed much better than the open loop scenario. The EnKF returned the soil moisture condition close to the true state more quickly than the open loop scenario reducing the prediction errors that resulted from inaccurate or

limited precipitation data. In addition, the EnKF was shown to overcome the problem of consistently underestimated soil moisture in some areas compared with the open loop due to a certain soil type and landuse.

The synthetic data modeling experiment conducted in this study assumed that surface soil moisture observations for each HRU in the SWAT model were available and that model/observational errors were known. However, for real world applications, further studies are required to answer the following questions. 1) How can we best use remotely sensed soil moisture observations in coarse resolution in time and space for a watershed scale hydrologic model? This question will lead to more studies on developing temporal and spatial downscaling methods (Kaheil et al., 2008; Merlin et al., 2010). 2) How can we determine the uncertainties in precipitation, models and observations? Various approaches have been presented in previous studies with the development of data assimilation techniques, some of which being mentioned in this paper. Although this study demonstrates the potential of remotely sensed surface soil moisture measurements and data assimilation for applications of watershed scale water resources management, future studies using actual observed data are necessary to effectively transfer the science to practical applications.

Acknowledgements

The authors thank Dr. Rao Govindaraju, Dr. Wade Crow and anonymous reviewers for all their constructive comments and suggestions for further developing and improving the quality of this manuscript.

Reference

Aubert, D., Loumagne, C. and Oudin, L., 2003. Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall-runoff model. Journal of Hydrology, 280(1-4): 145-161.

Bolten, J., Crow, W., Zhan, X., Reynolds, C. and Jackson, T., 2009. Assimilation of a satellite-based soil moisture product into a two-layer water balance model for a global crop production decision support system. In: S.K. Pard (Editor), Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications. Springer-Verlang, London, United Kingdom, pp. 449-464.

Borah, D.K. and Bera, M., 2003. SWAT model background and application reviews, 2003 ASAE Annual International Meeting, Las Vegas, Nevada.

Burgers, G., van Leeuwen, P.J. and Evensen, G., 1998. Analysis scheme in the Ensemble Kalman Filter. Monthly Weather Review, 126: 1719.

Chen, F., Crow, W.T., Starks, P.J. and Moriasi, D.N., 2011. Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture. Advances in Water Resources, 34(4): 526-536.

Chung, W.H., Wang, I.T. and Wang, R.Y., 2010. Theory-based SCS-CN method and its applications. Journal of Hydrologic Engineering, 15(12): 1045-1058.

Clark, M.P. et al., 2008. Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. Advances in Water Resources, 31: 1309-1324.

Crow, W.T. and Bolten, J.D., 2007. Estimating precipitation errors using spaceborne surface soil moisture retrievals. Geophysical Research Letters, 34(8).

Crow, W.T., Huffman, G.J., Bindlish, R. and Jackson, T.J., 2009. Improving Satellite-Based Rainfall Accumulation Estimates Using Spaceborne Surface Soil Moisture Retrievals. Journal of Hydrometeorology, 10(1): 199-212.

Crow, W.T. and Ryu, D., 2009. A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals. Hydrology and Earth System Sciences, 13(1): 1-16.

Crow, W.T., van den Berg, M.J., Huffman, G.J. and Pellarin, T., 2011. Correcting rainfall using satellite-based surface soil moisture retrievals: The Soil Moisture Analysis Rainfall Tool (SMART). Water Resources Research, 47.

Crow, W.T. and Wood, E.F., 2003. 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. Advances in Water Resources, 26(2): 137-149.

Das, N.N., Mohanty, B.P., Cosh, M.H. and Jackson, T.J., 2008. Modeling and assimilation of root zone soil moisture using remote sensing observations in Walnut Gulch Watershed during SMEX04. Remote Sensing of Environment, 112(2): 415-429.

DeLiberty, T.L. and Legates, D.R., 2003. Interannual and seasonal variability of modelled soil moisture in Oklahoma. International Journal of Climatology, 23(9): 1057-1086.

Draper, C.S., Mahfouf, J.F. and Walker, J.P., 2009. An EKF assimilation of AMSR-E soil moisture into the ISBA land surface scheme. Journal of Geophysical Research - Atmospheres, 114.

Engman, E.T., 1991. Applications of microwave remote sensing of soil moisture for water resources and agriculture. Remote Sensing of Environment, 35(2-3): 213-226.

Entekhabi, D., Nakamura, H. and Njoku, E.G., 1994. Solving the inverse problems for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations. IEEE Transactions on Geoscience and Remote Sensing, 32(2): 438-448.

Entekhabi, D. et al., 2008. The Soil Moisture Active/Passive Mission (SMAP), Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International, pp. III - 1-III - 4.
Entekhabi, D. et al., 2010. The Soil Moisture Active Passive (SMAP) Mission. Proceedings of the Ieee, 98(5): 704-716.
Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research - Oceans, 99: 10143-10162.
Flanagan, D., Huang, C.H., Pappas, E., Smith, D. and Heathman, G., 2008. Assessing conservation effects on water quality in the St. Joseph River Watershed, Proceedings of the AgroEnviron 2008 (Sixth International Symposium AgroEnviron), Antalya, Turkey, pp. 12.
Garen, D.C. and Moore, D.S., 2005. Curve number hydrology in water quality modeling: uses, abuses, and future directions. Journal of the American Water Resources Association, 41: 377-388.
Gassman, P.W., Reyes, M.R., Green, C.H. and Arnold, J.G., 2007. The Soil and Water Assessment Tool: Historical development, applications, and future research directions Transactions of the ASABE 50(4): 1211-1250.
Jackson, T.J., Hawley, M.E. and O'Neill, P.E., 1987. Preplanting soil moisture using passive microwave sensors. JAWRA Journal of the American Water Resources Association, 23: 11-19.
Jackson, T.J., Vine, D.M.L., Swifi, C.T., Schmugge, T.J. and Schiebe, F.R., 1995. Large area mapping of soil moisture using the ESTAR passive microwave radiometer in Washita'92. Remote Sensing of Environment, 53: 27-37.
Jacobs, J.M., Myers, D.A. and Whitfield, B.M., 2003. Improved rainfall/runoff estimates using remotelyl sensed soil moisture. Journal of the American Water Resources Association, 39: 313-324.
Jeong, J. et al., 2010. Development and integration of sub-hourly rainfall–runoff modeling capability within a watershed model. Water Resources Management, 24(15): 4505-4527.
Kaheil, Y.H., Gill, M.K., McKee, M., Bastidas, L.A. and Rosero, E., 2008. Downscaling and Assimilation of Surface Soil Moisture Using Ground Truth Measurements. Geoscience and Remote Sensing, IEEE Transactions on, 46(5): 1375-1384.
Kannan, N., Santhi, C., Williams, J.R. and Arnold, J.G., 2008. Development of a continuous soil moisture accounting procedure for curve number methodology and its behaviour with different evapotranspiration methods. Hydrological Processes, 22(13): 2114-2121.
Kannan, N., White, S.M., Worrall, F. and Whelan, M.J., 2007. Sensitivity analysis and identification of the best evapotranspiration and runoff options for hydrological modelling in SWAT-2000. Journal of Hydrology, 332(3-4): 456-466.
Kerr, Y.H. et al., 2010. The SMOS Mission: New Tool for Monitoring Key Elements of the Global Water Cycle. Proceedings of the Ieee, 98(5): 666-687.
Kerr, Y.H. et al., 2001. Soil moisture retrieval from space: the Soil Moisture and Ocean Salinity (SMOS) mission. Geoscience and Remote Sensing, IEEE Transactions on, 39(8): 1729-1735.
Kim, N.W. and Lee, J., 2008. Temporally weighted average curve number method for daily runoff simulation. Hydrological Processes, 22(25): 4936-4948.
King, Kevin W. , Arnold, J.G. and Bingner, R.L., 1999. Comparison of Green-Ampt and Curve Number methods on Goodwin Creek watershed using SWAT Transactions of the ASABE, 42(4): 919-926
Komma, J., Blőschl, G. and Reszler, C., 2008. Soil moisture updating by Ensemble Kalman Filtering in real-time flood forecasting. Journal of Hydrology, 357: 228-242.
Lakhankar, T., Karakauer, N. and Khanbilvardi, R., 2009. Applications of microwave remote sensing of soil moisture for agricultural applications. International Journal of Terraspace

Science and Engineering, 2(1): 81-91.
Merlin, O., Al Bitar, A., Walker, J.P. and Kerr, Y., 2010. An improved algorithm for disaggregating microwave-derived soil moisture based on red, near-infrared and thermal-infrared data. Remote Sensing of Environment, 114(10): 2305-2316.
Michel, C., Andr?ssian, V. and Perrin, C., 2005. Soil Conservation Service Curve Number method: How to mend a wrong soil moisture accounting procedure? Water Resources Research, 41(2): W02011.
Miller, R.N., Ghil, M. and Gauthiez, F., 1994. Advanced data assimilation in strongly nonlinear dynamical systems. Journal of the Atmospheric Sciences, 51(8): 1037-1056.
Mishra, S.K. and Singh, V.P., 2006. A relook at NEH-4 curve number data and antecedent moisture condition criteria. Hydrological Processes, 20: 2755-2768.
Moradkhani, H., Sorooshian, S., Gupta, H.V. and Houser, P.R., 2005. Dual state-parameter estimation of hydrological models using ensemble Kalman filter, Advances in Water Resources, pp. 135-147.
Narasimhan, B., Srinivasan, R., Arnold, J.G. and Di Luzio, M., 2005. Estimation of long-term soil moisture using a distributed parameter hydrologic model and verification using remotely sensed data. Transactions of the ASAE, 48(3): 1101-1113.
Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R. and King, K.W., 2002. Soil and water assessment tool. Theoretical documentation: Version 2005. TWRI TR-191. Texas Water Resources Institute, College Station, TX.
Ni-Meister, W., Houser, P.R. and Walker, J.P., 2006. Soil moisture initialization for climate prediction: Assimilation of scanning multifrequency microwave radiometer soil moisture data into a land surface model. J. Geophys. Res., 111(D20): D20102.
Njoku, E.G. and Entekhabi, D., 1996. Passive microwave remote sensing of soil moisture. Journal of Hydrology, 184: 101-129.
Njoku, E.G., Jackson, T.J., Lakshmi, V., Chan, T.K. and Nghiem, S.V., 2003. Soil moisture retrieval from AMSR-E. IEEE Transactions on Geoscience and Remote Sensing, 41(2): 215-229.
Pauwels, V.R.N., Hoeben, R., Verhoest, N.E.C. and De Troch, F.i.P., 2001. The importance of the spatial patterns of remotely sensed soil moisture in the improvement of discharge predictions for small-scale basins through data assimilation. Journal of Hydrology, 251(1-2): 88-102.
Reichle, R.H. et al., 2007. Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). Journal of Geophysical Research - Atmospheres, 112.
Reichle, R.H., McLaughlin, D.B. and Entekhabi, D., 2002a. Hydrologic data assimilation with the Ensemble Kalman Filter. Monthly Weather Review, 130: 103-114.
Reichle, R.H., Walker, J.P., Koster, R.D. and Houser, P.R., 2002b. Extended versus Ensemble Kalman Filtering for land data assimilation. Journal of Hydrometeorology, 3: 728-740.
Ryu, D., Crow, W.T., Zhan, X. and Jackson, T.J., 2009. Correcting unintended perturbation biases in hydrologic data assimilation. Journal of Hydrometeorology, 10(3): 734-750.
Sabater, J.M., Jarlan, L., Calvet, J.-C., Bouyssel, F. and De Rosnay, P., 2007. From near-surface to root-zone soil moisture using different assimilation techniques. Journal of Hydrometeorology, 8: 194-206.
Sahu, R.K., Mishra, S.K. and Eldho, T.I., 2010. An improved AMC-coupled runoff curve number model. Hydrological Processes, 24: 2834-2839.
Schmugge, T.J., Kustas, W.P., Ritchie, J.C., Jackson, T.J. and Rango, A., 2002. Remote sensing in hydrology. Advances in water resources, 25: 1367-1385.

Scipal, K., Scheffler, C. and Wagner, W., 2005. Soil moisture-runoff relation at the catchment scale as observed with coarse resolution microwave remote sensing. Hydrology and Earth System Sciences, 9(3): 173-183.
Troch, P.A., Paniconi, C. and McLaughlin, D., 2003. Catchment-scale hydrologic modeling and data assimilation. Advances in Water Resources, 26: 131-135.
USDA-SCS, 1972. SCS National Engineering Handbook, Section 4, Hydrology. Chapter 10, Estimation of Direct Runoff from Storm Rainfall. Hydrology. U.S. Department of Agriculture, Soil Conservation Service, Washington, D.C. pp. 10.1-10.24.
van Delft, G., El Serafy, G.Y. and Heemink, A.W., 2009. The ensemble particle filter (EnPF) in rainfall-runoff models. Stochastic Environmental Research and Risk Assessment, 23(8): 1203-1211.
Wagner, W., Lemoine, G. and Rott, H., 1999. A method for estimating soil moisture from ERS scatterometer and soil data. Remote Sensing of Environment, 70(2): 191-207.
Walker, J.P. and Houser, P.R., 2001. A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations. Journal of Geophysical Research-Atmospheres, 106(D11): 11761-11774.
Walker, J.P. and Houser, P.R., 2005. Hydrologic data assimilation. In: U. Aswathanarayana (Editor), Advances in Water Science Methodologies Taylor & Francis, A.A. Balkema, The Netherlands, pp. 230.
Walker, J.P., Willgoose, G.R. and Kalma, J.D., 2001. One-dimensional soil moisture profile retrieval by assimilation of near-surface observations: a comparison of retrieval algorithms. Advances in Water Resources, 24: 631-650.
Wang, X., Shang, S., Yang, W. and Melesse, A.M., 2008. Simulation of an agricultural watershed using an improved Curve Number method in SWAT. Transactions of the ASABE, 51(4): 1323-1339.
Weerts, A.H. and El Serafy, G.Y.H., 2006. Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. Water Resources Research, 42(9): 17.
Weissling, B.P., Xie, H. and Murray, K.E., 2007. A multitemporal remote sensing approach to parsimonious streamflow modeling in a southcentral Texas watershed, USA. Hydrology and Earth System Sciences Discussions 4(1): 1-33.
White, E.D. et al., 2010. Development and application of a physically based landscape water balance in the SWAT model. Hydrological Processes, 25(6): 915-925.
Xie, X. and Zhang, D., 2010. Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter. Advances in Water Resources, 33(6): 678-690.
Zhang, S., Li, H., Zhang, W., Qiu, C. and Li, X., 2006. Estimating the soil moisture profile by assimilating near-surface observations with the ensemble Kalman filter (EnKF). Advances in Atmospheric Sciences, 22(6): 936-945.

List of Figures

Figure 1. Study area: Upper Cedar Creek Watershed in Indiana, USA.

Figure 2. Comparison of two different precipitation inputs (a) difference in precipitation rate (true precipitation from all available rain gauges subtracted by precipitation only from the NCDC rain gauges) (b) cumulative total precipitation.

Figure 3. Schematic of surface soil moisture data assimilation with the SWAT.

Figure 4. Watershed average soil water content prediction: (a) surface (~5cm) soil water content, (b) profile soil water content. Precipitation difference = true precipitation – open loop precipitation.

Figure 5. Daily RMSE of: (a) surface (~5cm) soil water content, (b) profile soil water content, (c) depth of water in shallow aquifer (SHALLST), (d) surface runoff loading to main channel for day in HRU (QDAY), (e) amount of water in lateral flow in HRU for the day (LATQ), (f) groundwater contribution to streamflow from HRU for the day (GW_Q), (g) Evapotranspiration (ET), (h) curve number for current day (CNDAY).

Figure 6. Streamflow prediction.

Figure 7. Area-weighted sum of daily precipitation.

Figure 8. Relationship between surface runoff and soil water condition with various CN2 number in Curve Number method when rainfall is 23 mm/day.

Figure 9. Comparison of CN method and Green Ampt method for the relationship between surface runoff and soil water condition: (a) soil type: GnB (b) soil type: SrB.

Figure10. Time averaged RMSE of precipitation. The numbers indicate subbasin number.

Figure11. Time averaged RMSE of surface (~5cm) soil moisture from: (a) open loop, (b) EnKF. The numbers indicate subbasin number.

Figure12. Time averaged RMSE of profile soil moisture from: (a) open loop, (b) EnKF. The numbers indicate subbasin number.

Figure13. Surface soil moisture estimation error of open loop scenario (error = true – open loop): (a) HRU 11 (FRSD, Hw), (b) HRU 17 (AGRR, GnB2), (c) HRU 18 (AGRR, Hw), (d) HRU 21 (WETF, GnB2).

Figure14. Precipitation ((a) and (b)) and surface soil moisture error variation during drydown period((c) ~ (l)).

List of Tables.

Table 1. Soil types in UCCW.
Table 2. Landuse in UCCW.
Table 3. Statistical results of the simulated watershed average soil water content.
Table 4. Errors in precipitation and outflow of the watershed during simulation period.
Table 5. Effect of canopy interception (DAY259)

***Abbreviations**

ALPHA_BF	Alpha factor for groundwater recession curve [days]
RCHRG_DP	Deep aquifer percolation fraction
REVAPMN	Threshold water depth in the shallow aquifer for "revap" [mm]
CANMX	Maximum canopy storage
CN2	Initial SCS CN II value
SOL_K	Saturated hydraulic conductivity [mm hr^{-1}]
SOL_Z	Depth to bottom of soil layer [mm]
SOL_AWC	Available water capacity of soil layer [mmH2O $mmsoil^{-1}$]
SOL_ALB	Albedo when soil is moist
SLOPE	Average slope steepness [m m^{-1}]
ESCO	Soil evaporation compensation factor
EPCO	Plant water uptake compensation factor
SURLAG	Surface runoff lag time [days]
SMTMP	Snow melt base temperature [◦C]
TIMP	Snow pack temperature lag factor
CH_N2	Manning's "n" value for the main channel
CH_K2	Effective hydraulic conductivity in main channel alluvium [mm hr^{-1}]
BLAI	Maximum (potential) leaf area index
SOL_SW	Amount of water stored in the soil profile on any given day [mm H_2O]
SHALLST	Depth of water in shallow aquifer [mm H_2O]
DEEPST	Depth of water in deep aquifer [mm H_2O]
INFLPCP	Amount of precipitation that infiltrates into soil [mm H_2O]
QDAY	Surface runoff loading to main channel for day in HRU [mm H_2O]
LATQ	Amount of water in lateral flow in HRU for the day [mm H_2O]
RCHRG	Amount of water recharging both aquifers on current day in HRU [mm H_2O]
GW_Q	Groundwater contribution to streamflow from HRU on current day [mm H_2O]
GWSEEP	Amount of water recharging deep aquifer on current day [mm H_2O]
CNDAY	Curve number for current day, HRU and at current soil moisture
ET	Actual amount of evapotranspiration that occurs on day in HRU [mm H_2O]

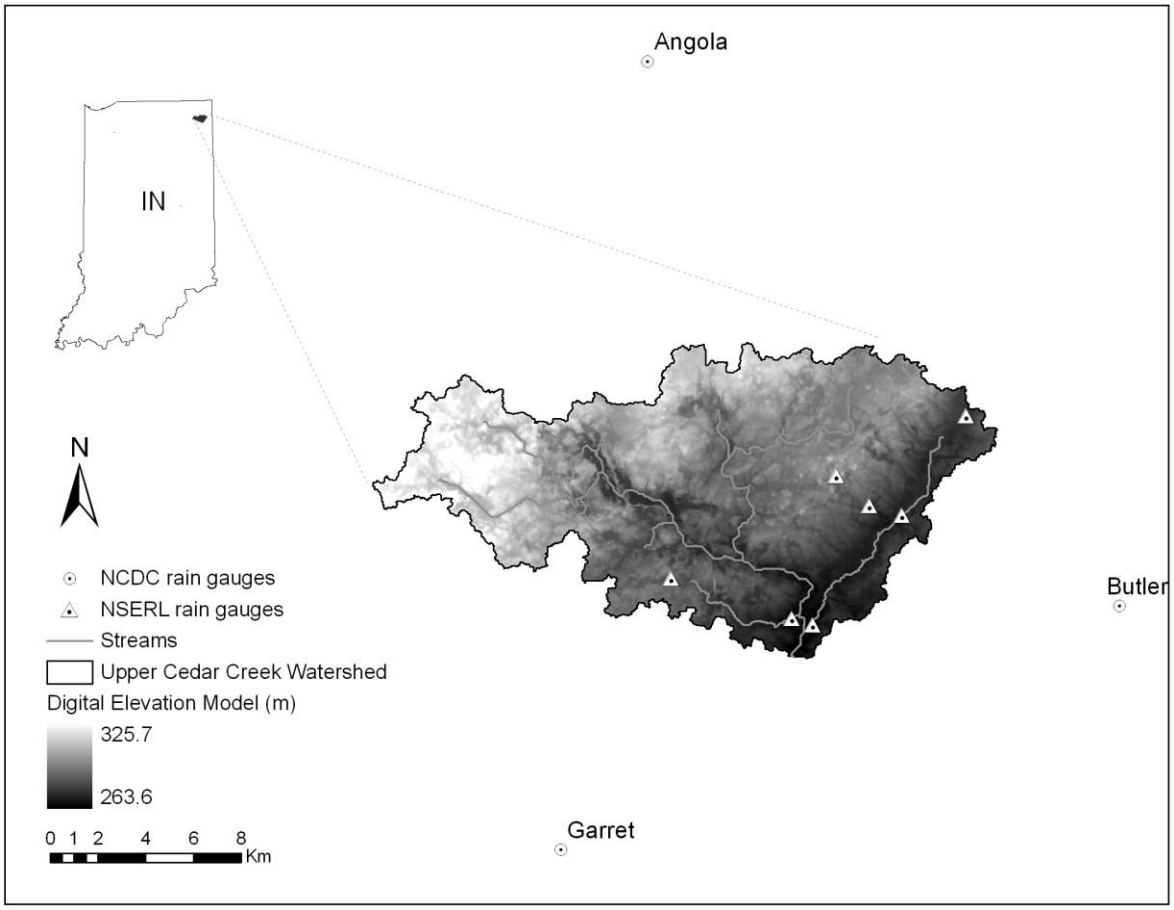
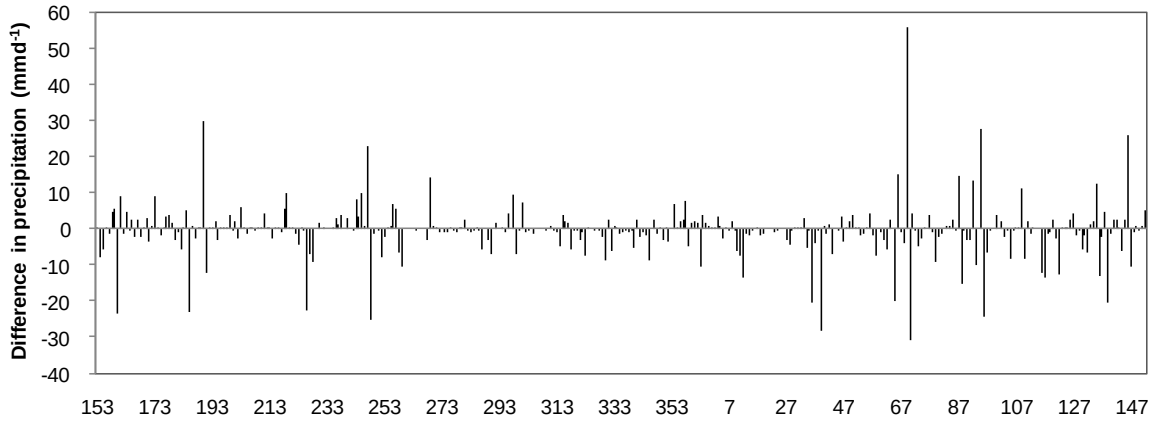


Figure 1. Study area: Upper Cedar Creek Watershed in Indiana, USA.

(a)



(b)

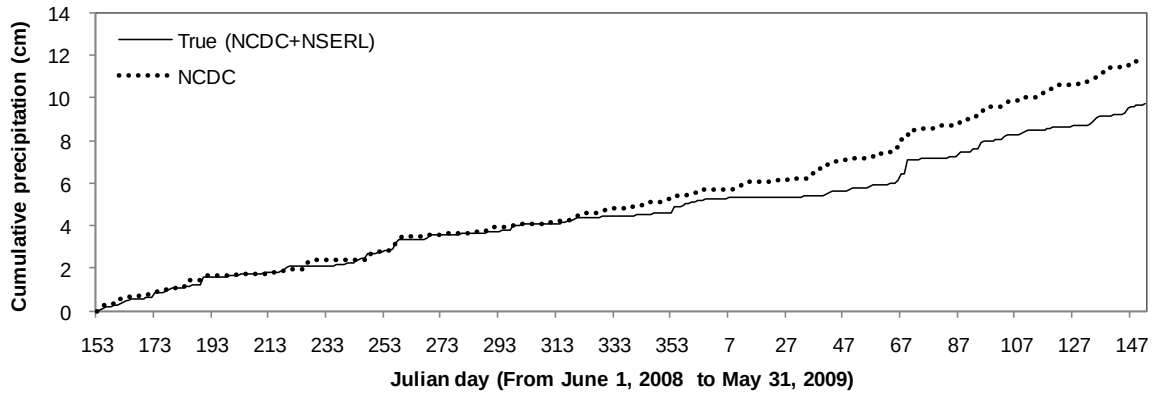


Figure 2. Comparison of two different precipitation inputs: (a) difference in precipitation rate (true precipitation from all available rain gauges subtracted by precipitation only from the NCDC rain gauges) (b) cumulative total precipitation.

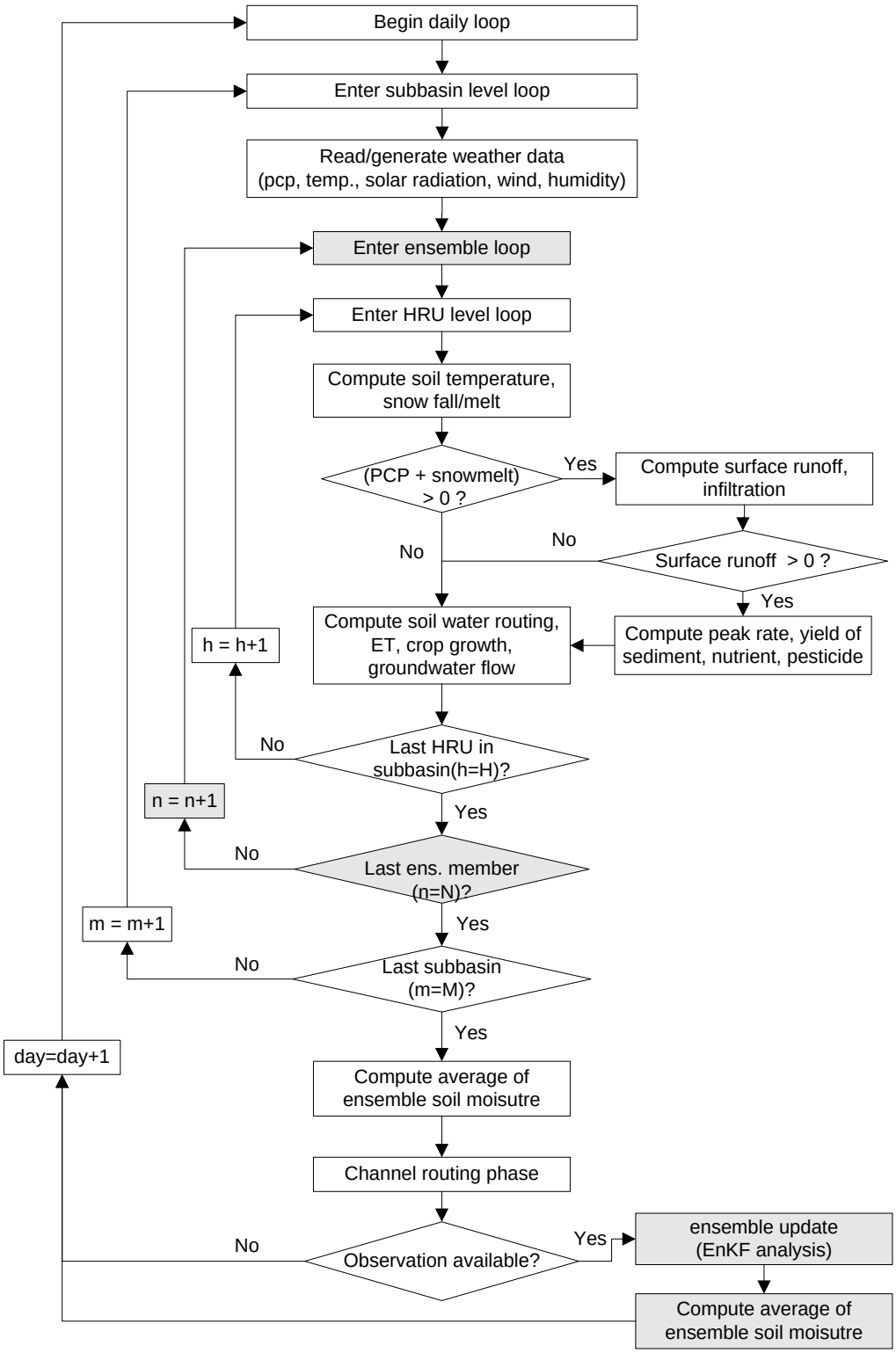


Figure 3. Schematic of surface soil moisture data assimilation with the SWAT. Additional routines due to the EnKF are shown in gray color. H, N and M refer the number of HRUs in a subbasin, total ensemble size (100) and total subbasin number (25) in the UCCW respectively.

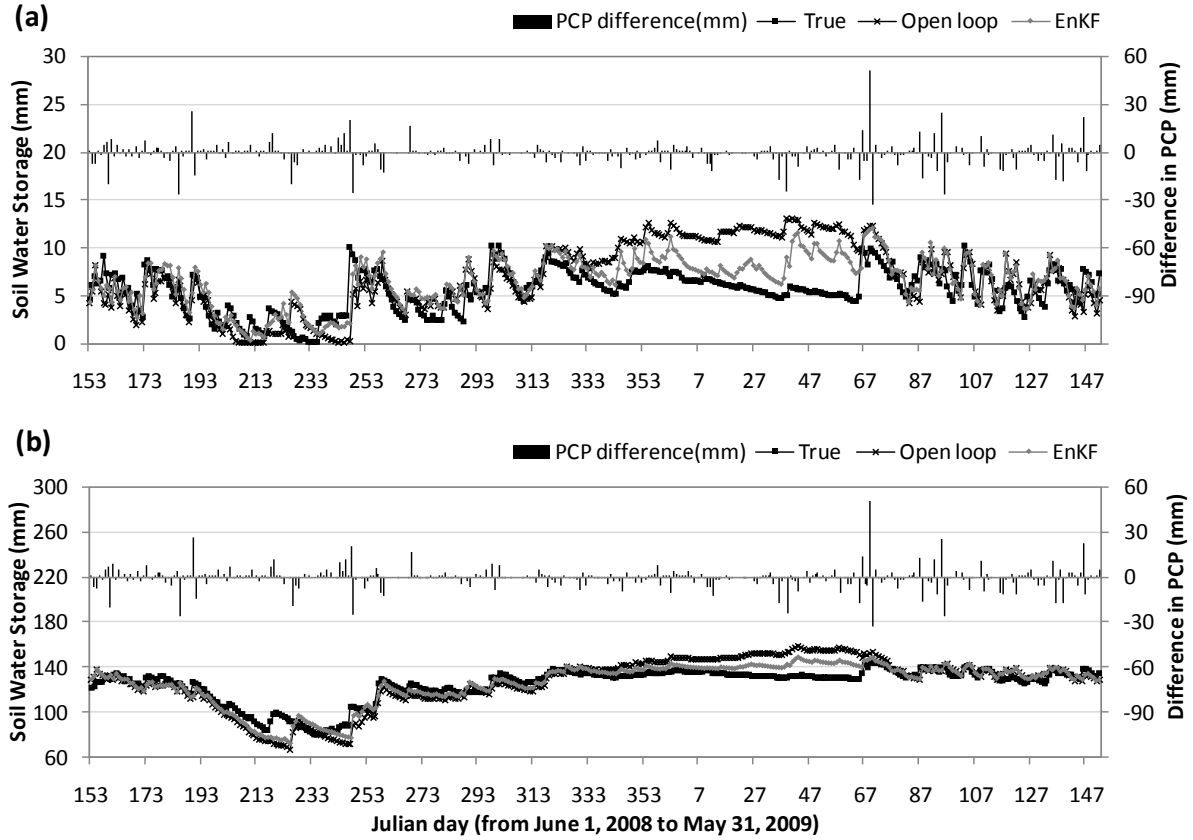
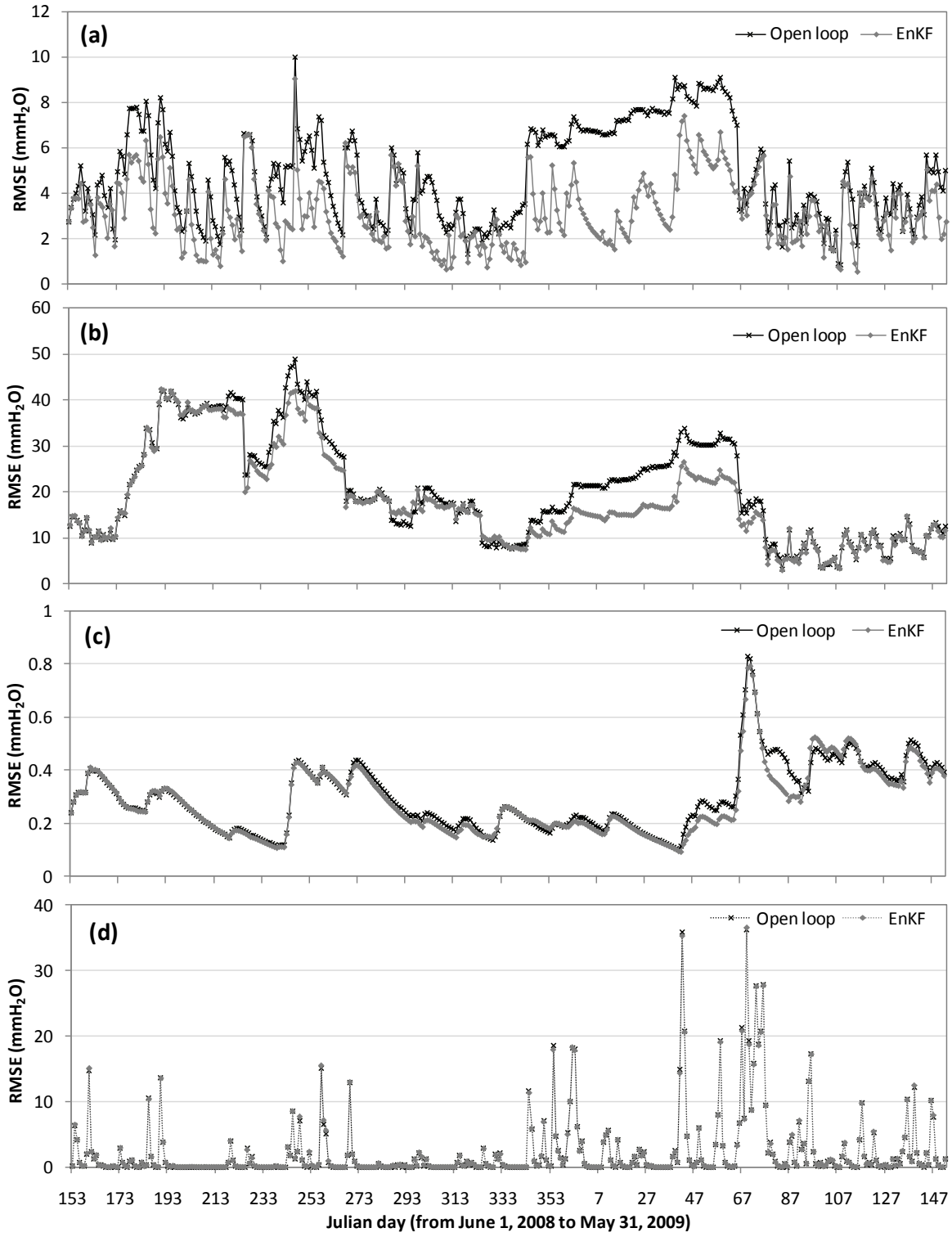


Figure 4. Watershed average soil water content prediction: (a) surface (~5cm) soil water content, (b) profile soil water content. Precipitation difference = true precipitation – open loop precipitation.


(a)
Open loop
EnKF
RMSE (mmH2O)
(b)
(c)
(d)
153 173 193 213 233 253 273 293 313 333 353 7 27 47 67 87 107 127 147
Julian day (from June 1, 2008 to May 31, 2009)

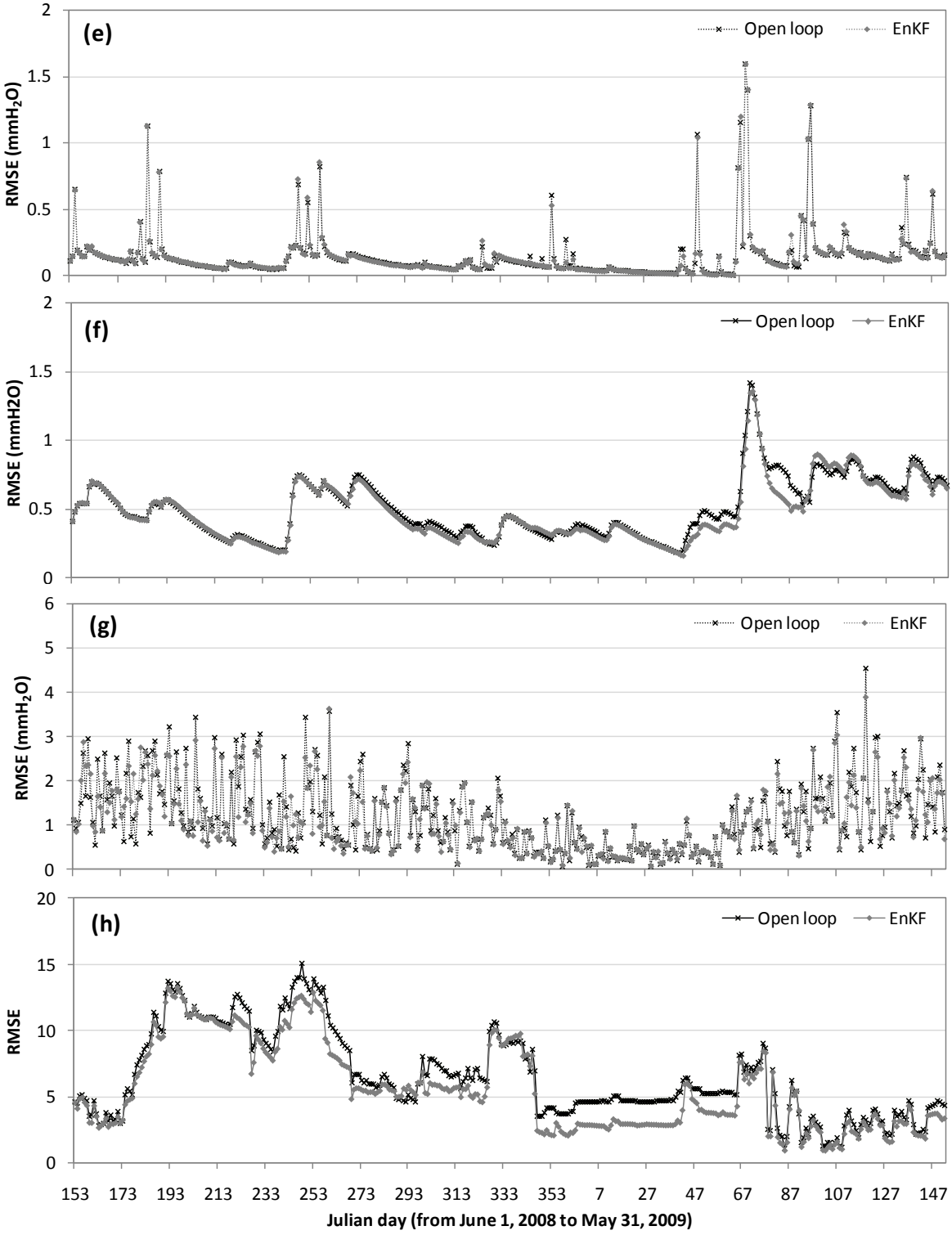


Figure 5. Daily RMSE of: (a) surface (~5cm) soil water content, (b) profile soil water content, (c) depth of water in shallow aquifer (SHALLST), (d) surface runoff loading to main channel for day in HRU (QDAY), (e) amount of water in lateral flow in HRU for the day (LATQ), (f) groundwater contribution to streamflow from HRU for the day (GW_Q), (g) Evapotranspiration (ET), (h) curve number for current day (CNDAY).

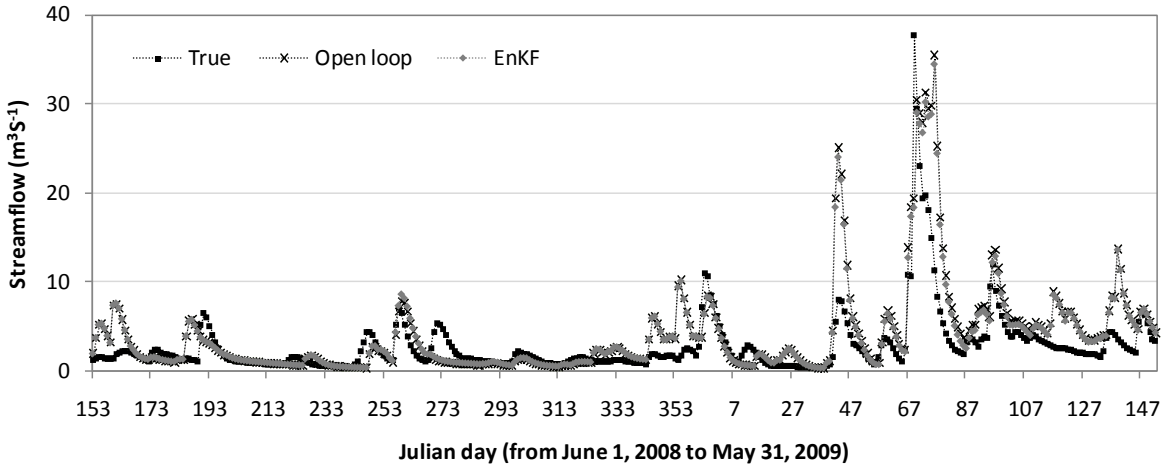


Figure 6. Streamflow prediction.

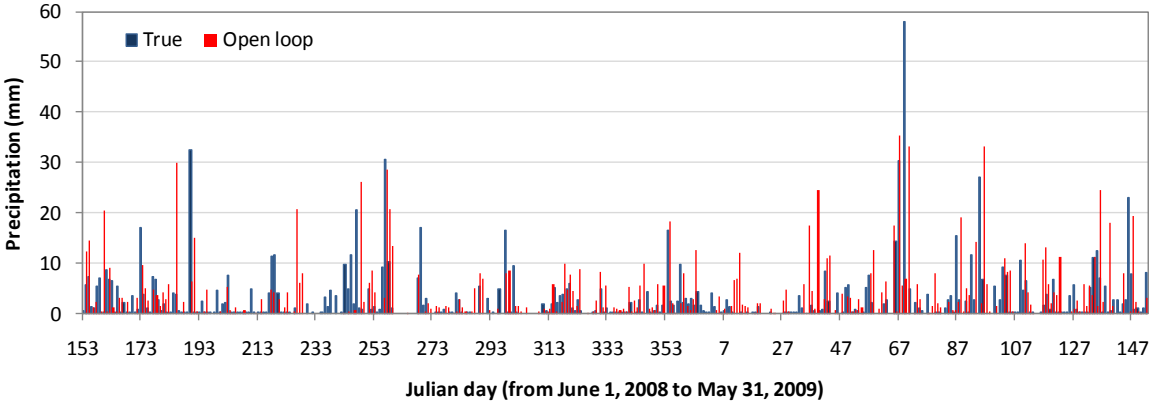


Figure 7. Area-weighted sum of daily precipitation.

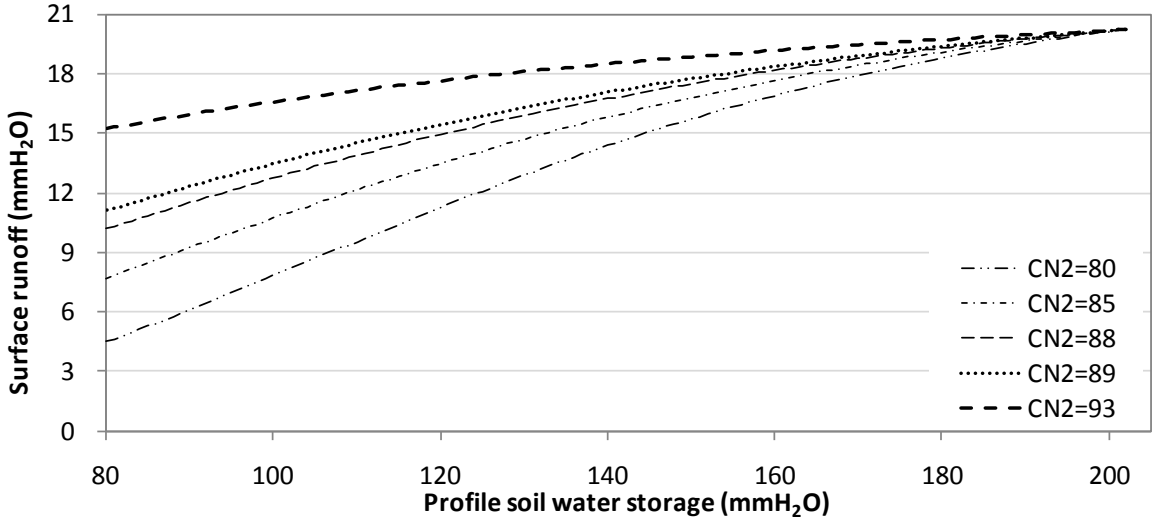


Figure 8. Relationship between surface runoff and soil water condition with various CN_2 numbers in Curve Number method when rainfall is 23 mm/day.

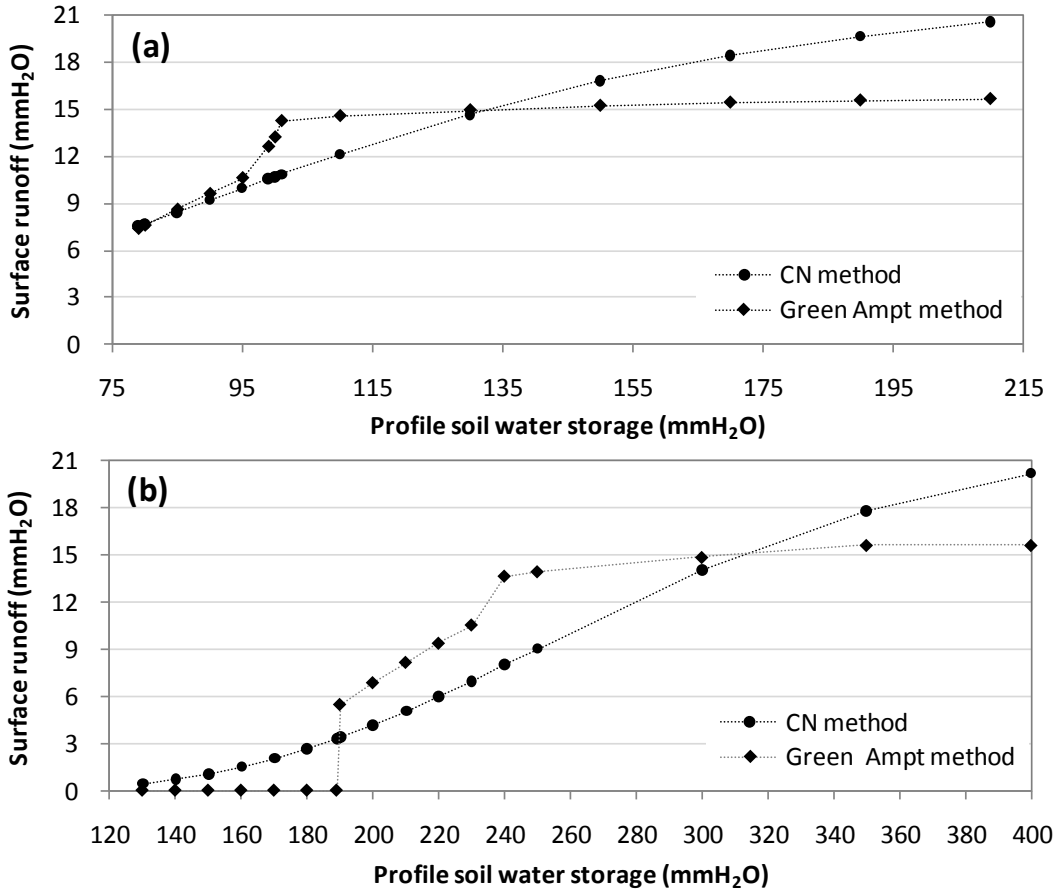


Figure 9. Comparison of CN method and Green Ampt method for the relationship between surface runoff and soil water condition: (a) soil type: GnB (b) soil type: SrB.

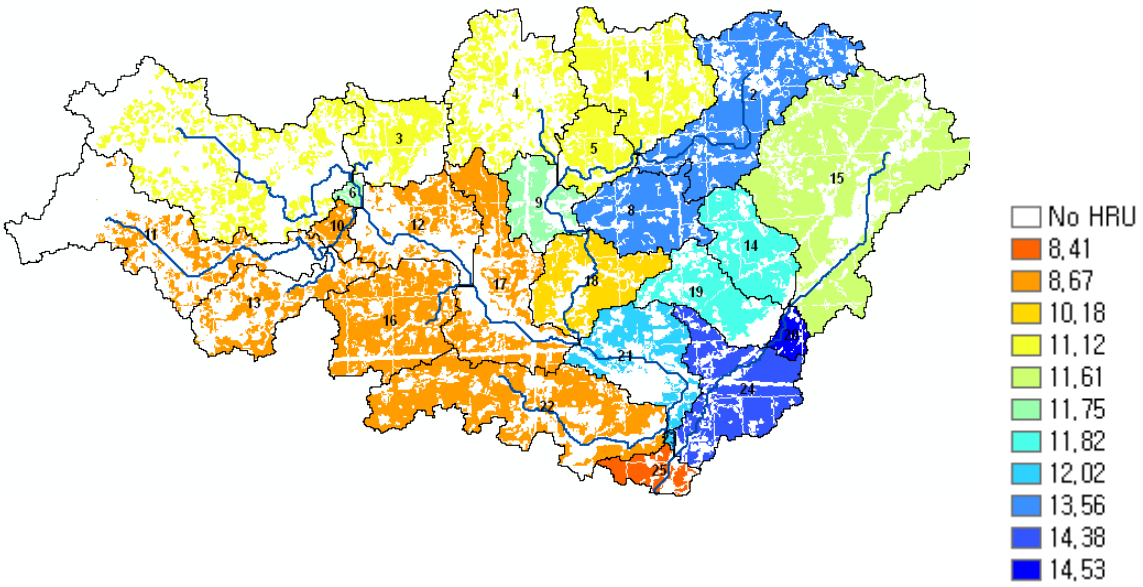


Figure10. Time averaged RMSE of precipitation. The numbers indicate subbasin number. (unit: mm)

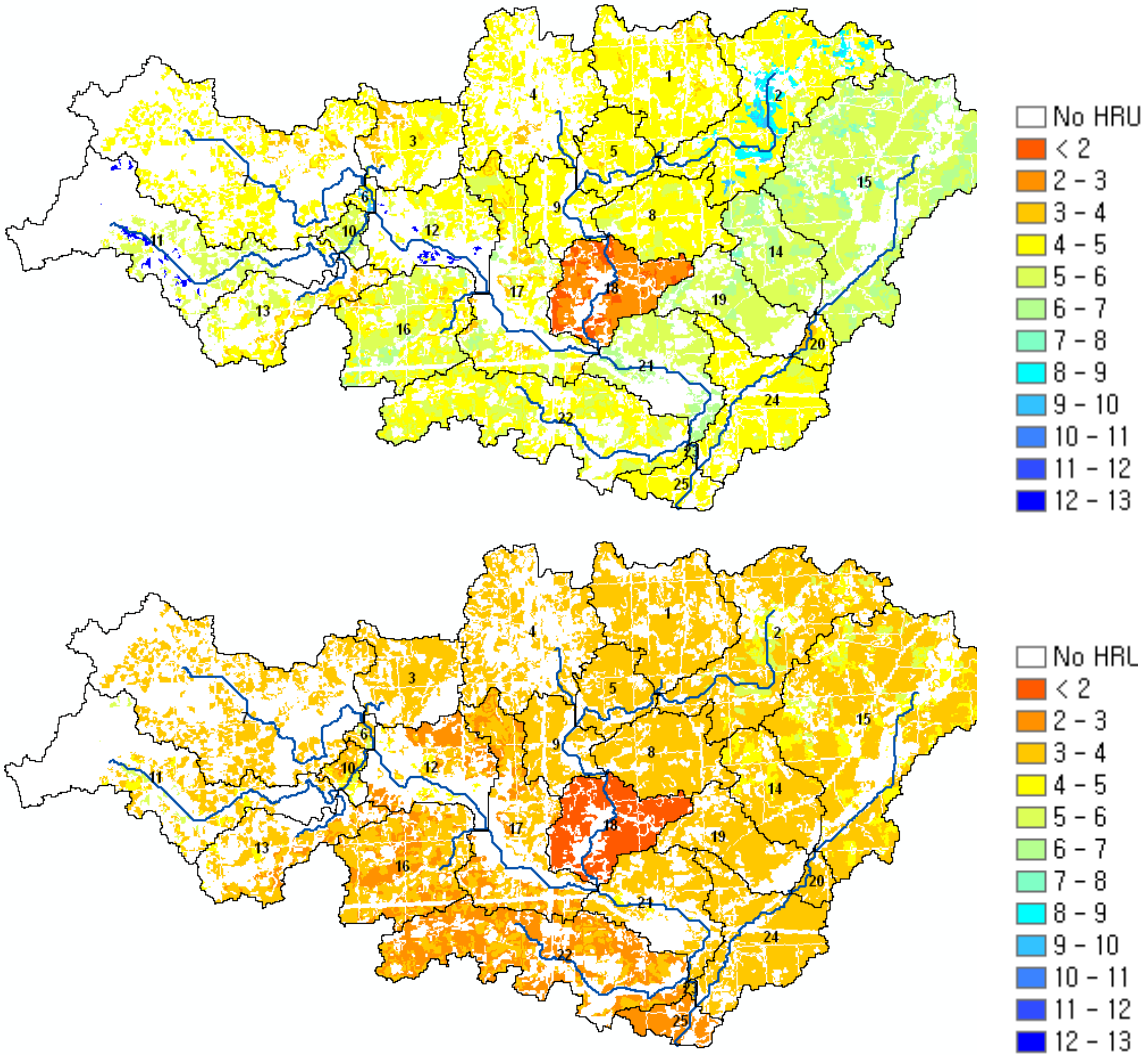


Figure11. Time averaged RMSE of surface (~5cm) soil moisture from: (a) open loop, (b) EnKF. The numbers on the map indicate subbasin number. (unit: mmH_2O)

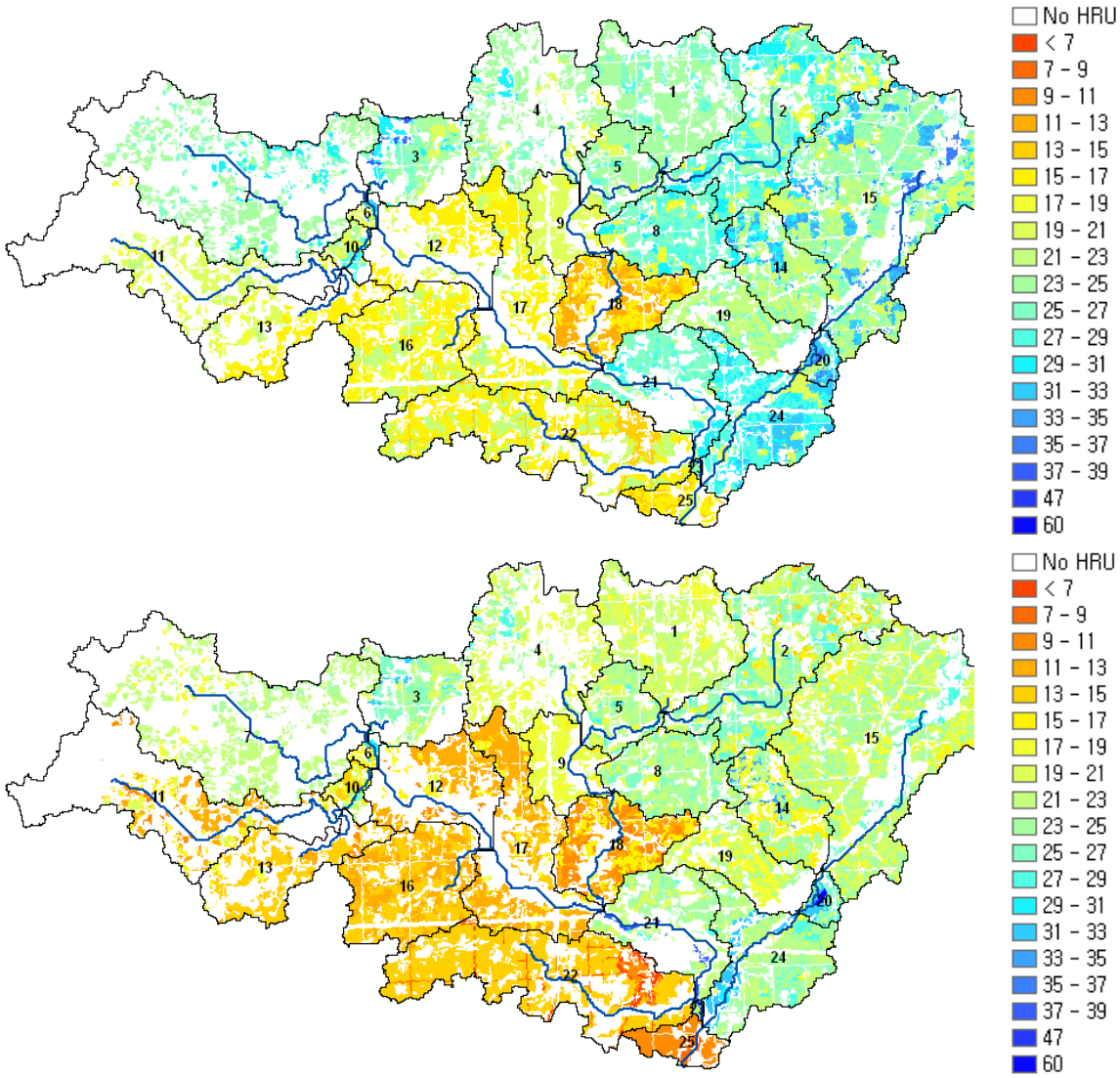


Figure12. Time averaged RMSE of profile soil moisture from: (a) open loop, (b) EnKF. The numbers on the map indicate subbasin number. (unit: mmH_2O)

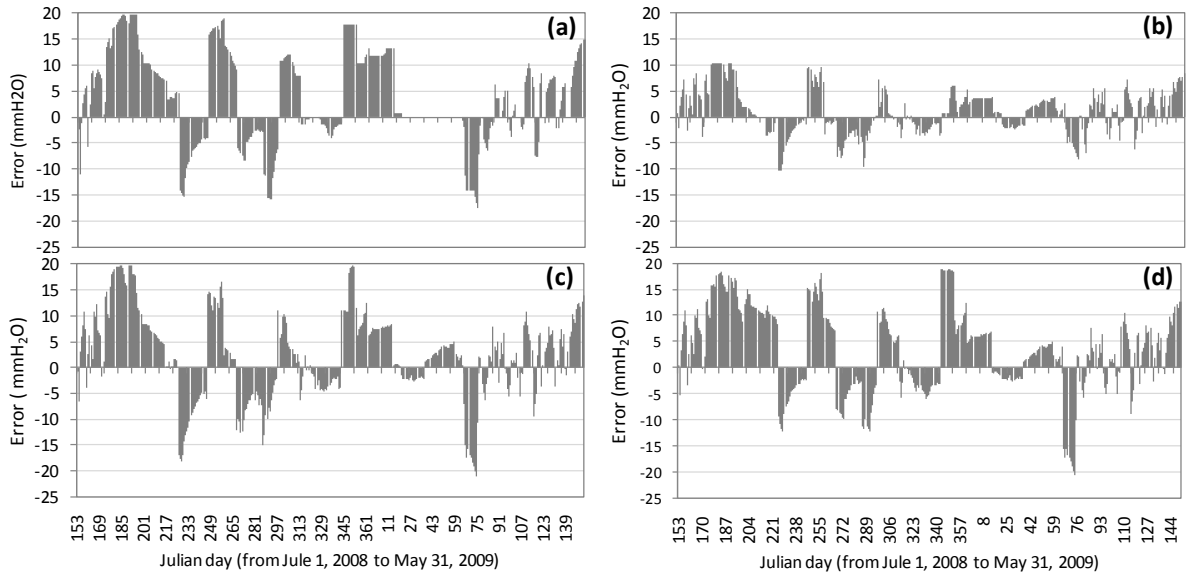
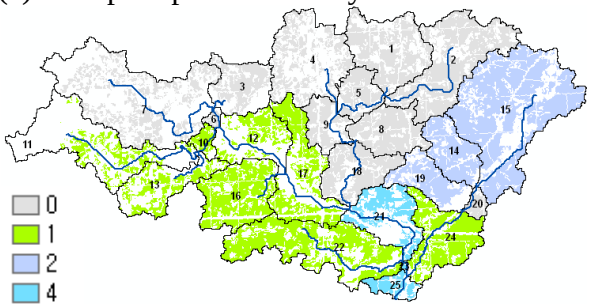
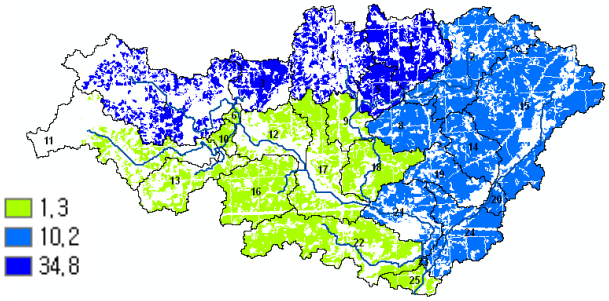


Figure13. Surface soil moisture estimation error of open loop scenario (error = true – open loop): (a) HRU 11 (FRSD, Hw), (b) HRU 17 (AGRR, GnB2), (c) HRU 18 (AGRR, Hw), (d) HRU 21 (WETF, GnB2).

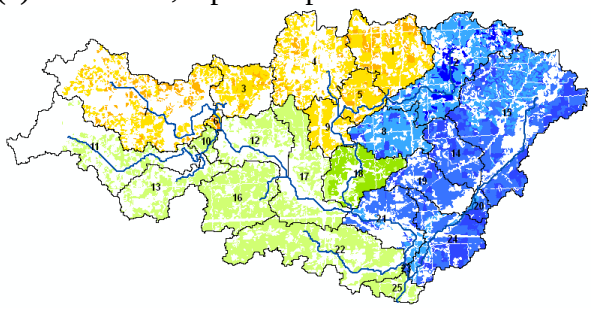
(a) True precipitation on day 259



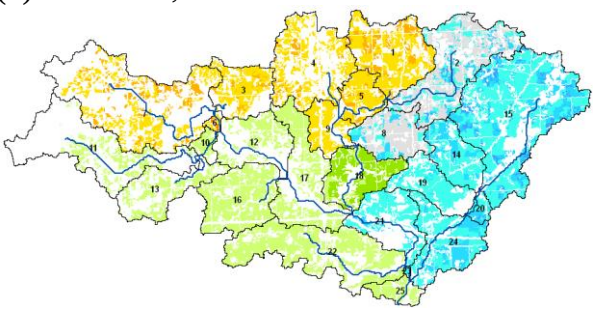
(b) Open loop precipitation on day 259



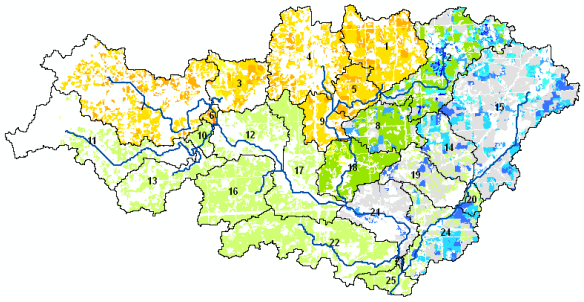
(c) DAY=258, Open loop



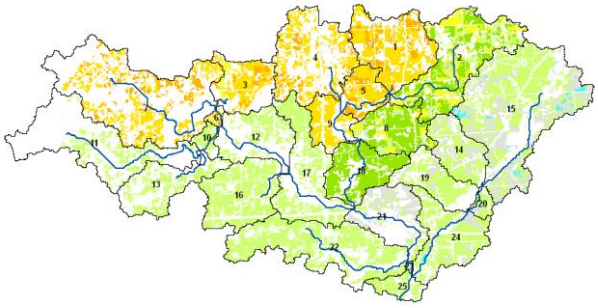
(d)D AY=258, EnKF



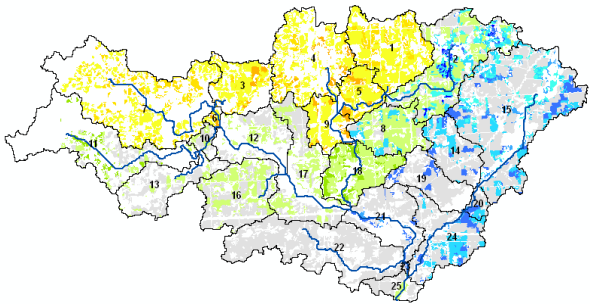
(e) DAY=259, Open loop (after precipitation)



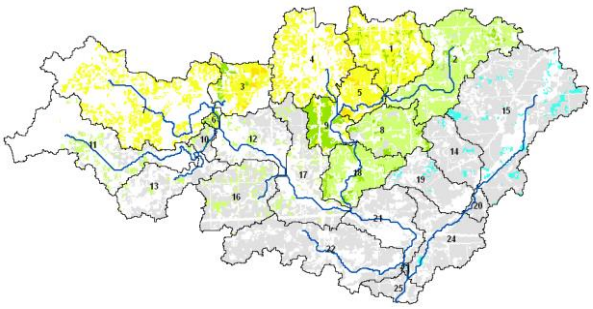
(f) DAY=259, EnKF (after precipitation)



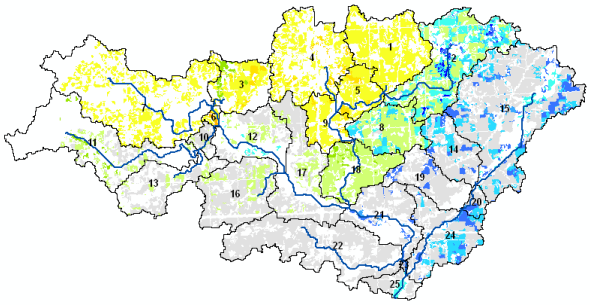
(g) DAY=260, Open loop



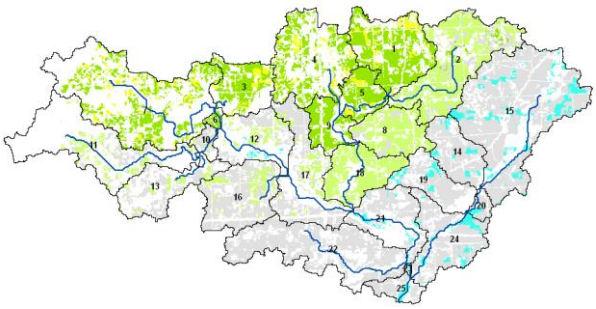
(h) DAY=260, EnKF



(i) DAY=261, Open loop



(j) DAY=261, EnKF



(k) DAY=262, Open loop

(l) DAY=262, EnKF

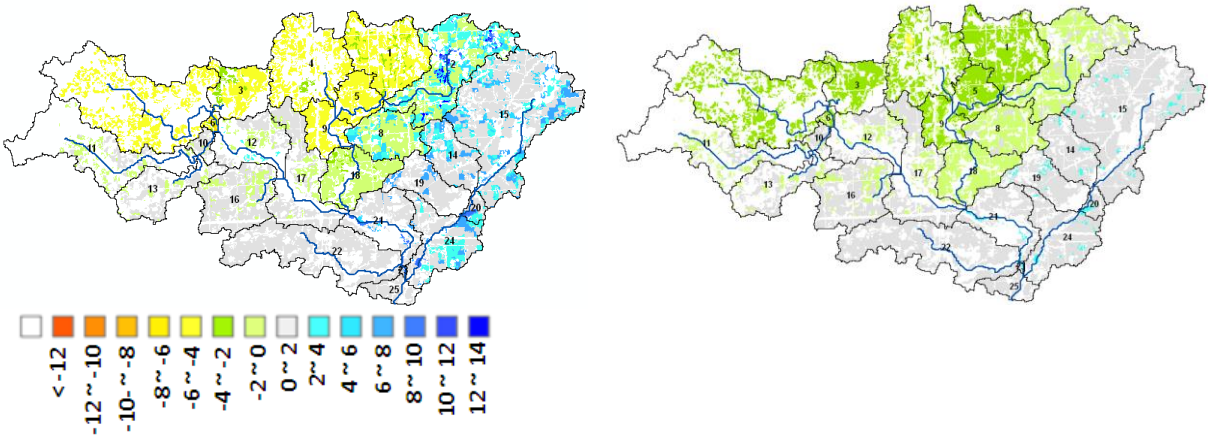


Figure14. Precipitation ((a) and (b)) and surface soil moisture error variation during drydown period((c) ~ (l)) (error = true – open loop or EnKF). (unit: mmH_2O)

Table 1. Soil types in UCCW.

Soil type		Area		Hydrologic soil group
		[km^2]	Proportion (%)	
GnB2	Glynwood loam	52.69	26.67	C
BaB2	Blount silt loam	43.89	22.21	D
Pe	Pewamo silty clay	35.59	18.01	C
SrB2	Strawn loam	30.71	15.55	B
RaB	Rawson sandy loam	7.43	3.76	C
MrC3	Morley silty clay loam	5.38	2.72	D
StC3	Strawn clay loam	4.83	2.44	B
Hw	Houghton muck	4.59	2.32	A
Re	Rensselaer loam	4.14	2.09	B
BoB	Boyer sandy loam	3.84	1.94	A
Se	Sebewa sandy loam	1.47	0.75	B
All others		3.01	1.54	

Table 2. Landuse in UCCW.

Landuse classification		Area	
		[km^2]	Proportion (%)
AGRR	Agricultural	113.67	57.54
HAY	Hay	30.75	15.56
FRSD	Forest, deciduous	23.59	11.94
URLD	Residential, low intensity	8.38	4.24
WETF	Wetlands-forested	7.99	4.04
URMD	Residential, medium intensity	4.03	2.04
RNGB	Range, brush	3.82	1.94
RNGE	Range, grasses	2.89	1.46
WATR	Water	0.91	0.46
URHD	Residential, high intensity	0.76	0.38
WETN	Wetlands, non-forested	0.41	0.21
FRSE	Forest, evergreen	0.16	0.08
UIDU	Industrial	0.10	0.05
FRST	Forest, mixed	0.08	0.04
SWRN	Barren land	0.02	0.01
	Sum	197.57	99.99

Table 3. Statistical results of the simulated watershed average soil water content.

	R		**RMSE**		**MBE**	
	Open loop	EnKF	Open loop	EnKF	Open loop	EnKF
SOL_SW (5cm)	0.585	0.747	3.214	1.931	-1.275	-0.879
SOL_SW	0.906	0.942	11.176	6.832	-1.533	-0.317
SHALLST	0.929	0.923	0.127	0.097	-0.054	-0.041
DEEPST	0.986	0.995	2.444	2.438	2.444	2.438
INFLPCP	0.339	0.339	3.572	3.572	-0.152	-0.171
QDAY	0.458	0.455	3.059	3.025	-0.452	-0.442
LATQ	0.844	0.860	0.035	0.034	-0.013	-0.014
RCHRG	0.927	0.921	0.219	0.168	-0.092	-0.071
GW_Q	0.929	0.923	0.217	0.165	-0.092	-0.071
GWSEEP	0.930	0.924	0.001	0.001	0.000	0.000
CNDAY	0.976	0.977	2.125	1.617	0.641	0.358
ET	0.696	0.736	0.852	0.849	-0.111	-0.200

Table 4. Errors in precipitation and outflow of the watershed during simulation period.

	True	Open loop	EnKF
Input precipitation (mm year^{-1})	937.02	1175.37	1175.37
Outflow (mm year^{-1})	438.47 (47%)	641.94 (55%)	630.85 (54%)
Error in precipitation (mm year^{-1})	-	-238.36	-238.36
Error in streamflow (mm year^{-1})	-	-203.46	-192.37

Table 5. Effect of canopy interception (DAY259)

HRU(Landuse)	Initial precipitation		Precipitation after canopy interception		
	True	Open loop/EnKF	True	Open loop	EnKF
118,119 (FRSD)	2.4	10.2	0	1.83	1.83
115,116,117 (AGRR)			1.989	10.151	10.08