Streamflow and Soil Moisture Assimilation in the SWAT model Using the Extended Kalman Filter

by

Leqiang Sun

Dissertation submitted to the

Faculty of Graduate and Postdoctoral Studies

In partial fulfillment of the requirements

for the Ph.D. degree in

Civil Engineering

Department of Civil Engineering

Faculty of Engineering

University of Ottawa

Abstract

Numerical models often fail to accurately simulate and forecast a hydrological state in operation due to its inherent uncertainties. Data Assimilation (DA) is a promising technology that uses real-time observations to modify a model's parameters and internal variables to make it more representative of the actual state of the system it describes. In this thesis, hydrological DA is first reviewed from the perspective of its objective, scope, applications and the challenges it faces. Special attention is then given to nonlinear Kalman filters such as the Extended Kalman Filter (EKF). Based on a review of the existing studies, it is found that the potential of EKF has not been fully exploited.

The Soil and Water Assessment Tool (SWAT) is a semi-distributed rainfall-runoff model that is widely used in agricultural water management and flood forecasting. However, studies of hydrological DA that are based on distributed models are relatively rare because hydrological DA is still in its infancy, with many issues to be resolved, and linear statistical models and lumped rainfall-runoff models are often used for the sake of simplicity. This study aims to fill this gap by assimilating streamflow and surface soil moisture observations into the SWAT model to improve its state simulation and *forecasting* capability. Unless specifically defined, all '*forecasts*' in Italic font are based on the assumption of a perfect knowledge of the meteorological forecast. EKF is chosen as the DA method for its solid theoretical basis and parsimonious implementation procedures.

Given the large number of parameters and storage variables in SWAT, only the watershed scale variables are included in the state vector, and the Hydrological Response Unit (HRU) scale variables are updated with the *a posteriori/a priori* ratio of their watershed scale counterparts. The Jacobian matrix is calculated numerically by perturbing the state variables. Two case studies are carried out with real observation data in order to verify the effectiveness of EKF assimilation. The upstream section of the Senegal River (above Bakel station) in western Africa is chosen for the streamflow assimilation, and the USDA ARS Little Washita experimental watershed is chosen to examine surface soil moisture assimilation. In the case of streamflow assimilation, a spinoff study is conducted to compare EKF state-parameter assimilation with a linear autoregressive (AR) output assimilation to improve SWAT's flood *forecasting* capability.

The influence of precipitation forecast uncertainty on the effectiveness of EKF assimilation is discussed in the context of surface soil moisture assimilation.

In streamflow assimilation, EKF was found to be effective mostly in the wet season due to the weak connection between runoff, soil moisture and the curve number (CN2) in dry seasons. Both soil moisture and CN2 were significantly updated in the wet season despite having opposite update patterns. The flood *forecast* is moderately improved for up to seven days, especially in the flood period by applying the EKF subsequent open loop (EKFsOL) scheme. The *forecast* is further improved with a newly designed quasi-error update scheme.

Comparison between EKF and AR output assimilation in flood *forecasting* reveals that while both methods can improve *forecast* accuracy, their performance is influenced by the hydrological regime of the particular year. EKF outperformed the AR model in dry years, while AR outperformed the EKF in wet years. Compared to AR, EKF is more robust and less sensitive to the length of the *forecast* lead time. A combined EKF-AR method provides satisfying results in both dry and wet years.

The assimilation of surface soil moisture is proved effective in improving the full profile soil moisture and streamflow estimate. The setting of state and observation vector has a great impact on the assimilation results. The state vector with streamflow and all-layer soil moisture outperforms other, more complicated state vectors, including those augmented with intermediate variables and model parameters. The joint assimilation of surface soil moisture and streamflow observation provides a much better estimate of soil moisture compared to assimilating the streamflow only. The updated SWAT model is sufficiently robust to issue improved *forecasts* of soil moisture and streamflow after the assimilation is 'unplugged'. The error quantification is found to be critical to the performance of EKF assimilation. Nevertheless, the application of an adaptive EKF shows no advantages over using the trial and error method in determining time-invariant model errors.

The robustness of EKF assimilation is further verified by explicitly perturbing the precipitation 'forecast' in the EKF subsequent *forecasts*. The open loop model without previous EKF update is more vulnerable to erroneous precipitation estimates. Compared to streamflow *forecasting*, soil moisture *forecasting* is found to be more resilient to erroneous precipitation input.

Acknowledgements

First and foremost, I would like to express my appreciation and special thanks to my academic advisors Dr. Ioan Nistor and Dr. Ousmane Seidou. Dr. Nistor is one of the most intelligent and humorous professors I have ever worked with. He is a caring person who is always ready to offer help, and he has been a tremendous mentor to me, not only academically but also in career planning. Dr. Seidou guided me throughout the whole process of my PhD study and provided me with invaluable suggestions and advice. He is truly a humble and talented person, and the knowledge and experience I gained by working with him will benefit me for the rest of my life. I am greatly honored to have been a student of these two exceptional professors.

I would like to kindly thank Dr. Colin Rennie, Dr. Abhijit Sarkar, Dr. Majid Mohammadian, and Dr. Tad Murty for serving on various committees during my PhD studies. Thank you all for the time you made for me from your tight schedules and for the suggestions you provided. Many thanks to Dr. Marie-Amélie Boucher from Université du Québec à Chicoutimi for her extensive comments and suggestions to improve this thesis. Additional thanks to Dr. Kalifa Goïta from the University of Sherbrooke and his student Youssef Jallouli for their discussions about remote-sensing soil moisture derivation. Thanks also to Dr. Kailei Liu from Hohai University, China, who also offered help in reviewing parts of this thesis.

Special thanks to my good friend Andre Purdy for proofreading my thesis and for the linguistic assistance he offered. Also, thanks to Mr. Chris Buchanan and Mr. Robert Hart for the proofreading he did in the revision of this thesis.

Last but not least, the greatest appreciation to my family back in China. I am deeply grateful to my mother and father, who are the best parents ever, for the selfless sacrifices they made when I was studying overseas. This thesis would not have been possible without their continuous support and unconditional encouragement.

This study has been funded by the uOttawa-CSC four-year joint scholarship.

Table of Contents

List of Figures

List of Tables

ACRONYMS

3D VAR	3-Dimensional Variational
4D VAR	3-Dimensional Variational
ACF	Autocorrelation function
AEnKF	Asynchronous EnKF
AMC	Antecedent moisture condition
AMSR-E	Advanced Microwave Scanning Radiometer - Earth Observing System
ANN	Artificial neural network
AR	Autoregressive model
ARMA	Autoregressive–moving-average model
ARS	Agricultural Research Service (USDA)
ASCAT	Advanced Scatterometer
ATLAS	Auto-Tuned Land Data Assimilation System
CFSR	Climate Forecast System Reanalysis
CN	Curve number
DA	Data assimilation
DAO	Data assimilation Office (NASA)
DEM	Digital elevation model
ECMWF	European Center for Medium-range Weather Forecasting
EDA	Evolutionary data assimilation
EKF	Extended Kalman Filter
EKFsOL	EKF subsequent open loop
EnKF	Ensemble Kalman Filter
EnKS	Ensemble Kalman Smoother
FAO	Food and Agriculture Organization (UN)
fvGCM	Finite-volume general circulation model
GA	Genetic algorithm
GLCC	Global Land Cover Characterization
HBV	Hydrologiska Byråns Vattenbalansavdelning model
HRU	Hydrological response unit
ISCCP	International Satellite Cloud Climatology Project
ISURF	Integrated Sensitivity and Uncertainty analysis Framework
KF	Kalman Filter
MA	Moving average model
MEs	Moment equations
MISP	Mutually Interactive State-Parameter
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NCEP	National Centers for Environmental Prediction
NSC	Nash–Sutcliffe Coefficient

NWP	Numerical Weather Prediction
OL	Open Loop
PBIAS	Percent bias
PF	Particle Filter
PSO	Particle Swarm Optimization
QPFs	Quantitative precipitation forecasts
RR	Rainfall-Runoff
RMSE	Root-mean-square error
SCA	Snow covered area
SCE	Snow cover extent
SCS	Soil Conservation Service
SEEK	Singular Evolutive Extended Kalman filter
SEKF	Simplified EKF
SODA	Simultaneous Optimization and Data Assimilation method
SWAT	Soil and Water Assessment Tool model
SWE	Snow Water Equivalent
UKF	Unscented Kalman Filter
USDA	United States Department of Agriculture
USGS	U.S. Geological Survey
VIC	Variable Infiltration Capacity model

List of Symbols

$CN1$	curve number with dry soil moisture
$CN2$	curve number with average soil moisture
$CN3$	curve number with wet soil moisture
d_k	index weighted coefficient
$depth$	the depth of the layer (mm)
ET_i	the daily evapotranspiration (mm)
$\widetilde{H}_k$	the new observation function
H_k	observation transition function
$\dot{H}_k$	the linearized observation function
I	Identity matrix
J_M	Jacobian matrix
k	time step
K_k	Kalman gain
L	the slope length for lateral subsurface flow (m)
$latlyr$	lateral subsurface flow in layer (mm)
$lyrtile$	the drainage tile flow in soil layer on the current day (mm)
$\widetilde{M}_k$	the augmented model operator
M_k	model transition function
$\dot{M}_k$	the linearized model function
N	the time series length (days)
p_{i+t}	the initial precipitation forecast
p'_{i+t}	the perturbed precipitation forecast
P_i	the daily percolation (mm)
P_k^a	a posteriori state estimate error matrix
P_k^f	a priori state estimate error matrix
pG	the percolation to the deep aquifer (mm)
Q	the daily surface runoff (mm)
$\overline{Q^o}$	the mean of the observed streamflow series (m^3/s)
$\tilde{Q}_k$	the new model error covariance matrix
Q_i	the daily runoff (mm)
Q_i^o	the streamflow observation of the i^{th} day (m^3/s)
Q_i^s	the model prediction of the i^{th} day (m^3/s)
Q_k	model error covariance matrix
q_k	model error mean
QR_i	the daily return flow (mm)
R	the daily rainfall (mm)
R_i	the daily precipitation
R_k	observation error covariance matrix
r_k	observation error mean

Rc	the recharge from the soil profile (mm)
rev	the root uptake from the shallow aquifer (mm)
Rf	the return flow (mm)
S	the retention parameter
S_k	the covariance matrix of the parameter errors
$sepday$	the percolation from soil layer (mm)
$slope$	the average slope steepness (m/m)
sol_fc	the amount of water held in the soil layer at field capacity (mm)
sol_k	the saturated hydraulic conductivity of soil layer (mm/hour)
sol_st	the amount of water stored in the soil layer on the current day (mm)
sol_ul	the amount of water held in the soil layer at saturation (mm)
$ST1$	surface soil layer (depth of 0-100mm)
$ST2$	subsurface soil layer (depth of 100-400mm)
$ST3$	deeper soil layer (depth 400-500mm).
SW	the initial soil water content (mm)
sw_{excess}	the soil water excessing the field capacity (mm)
SW_t	the soil water content of day t (mm)
u_k	forcing term
v_k	the innovation
V_{sa}	the shallow aquifer storage (mm)
W_{sa}	the water use from the shallow aquifer (mm)
x_{k-1}^a	a posteriori state vector
$x_k^a(i)$	the ith element of the a posteriori state vector
x_k^f	a priori state
$x_k^f(i)$	the ith element of the a priori state.
y_k^o	observation vector
Z_k	the augmented state vector
α	scaling factor
γ_{i+t}	the perturbation factor
ε_k	observation error
$\tilde{\eta}_k$	the augmented model error
η_k	model error
θ_k	parameters in augmented state vector
σ	error variance (covariance)

Chapter 1 Introduction

1.1 Motivation

Rainfall-runoff models can be classified as either lumped models or distributed models, according to their spatial description of the watershed. The simulation of hydrological processes in lumped models is averaged over the watershed. Distributed models divide the watershed into multiple calculation units on which the basic hydrological processes are simulated before they are lumped with continuous equations. Distributed models are able to consider the heterogeneity of topography, soil properties and meteorological inputs within the watershed, but they usually feature more a complicated model structure and parameters compared to lumped models. In addition, the parameters in distributed models usually have physical representations, and therefore it is possible to verify the simulation with *in situ* or remote-sensing observations.

The Soil and Water Assessment Tool (SWAT) model divides the watershed into Hydrological Response Units (HRUs) with homogeneous physical properties such as topography, land use or soil type, while the simulation of the hydrological processes on HRUs are based on an empirical Soil Conservation Service (SCS) method instead of a deterministic physical method, and therefore it is often described as a 'semi-distributed' model (Arnold et al., 1998; Beven et al., 1995; Liang et al., 1994). Semi-distributed hydrological models blur the boundary between conceptual lumped models and physically distributed models (Corzo et al., 2009). To avoid confusion, the 'semi-distributed' SWAT is still considered as a 'distributed model' in this study.

Despite its seeming advantages, a complex distributed hydrological model does not guarantee a better simulation or forecast result than a simple lumped model (Bergström, 1991). This is likely due to the various additional uncertainties involved in the operation of distributed models, which arise from initial condition, inputs, model structure, parameterization, etc. (Liu and Gupta, 2007).

Data assimilation (DA) is a method of continuously correcting model simulations with the most recent measurements. Originally used in areas like atmospheric and oceanographic sciences (Dee, 1995; Derber and Rosati, 1989; Ghil and Malanotte-Rizzoli, 1991), DA has drawn more and more attention from hydrologists for its powerful role in real-time correction of the initial

condition and parameters of hydrological models (Robinson and Lermusiaux, 2000). The difference between a regular hydrological model and a model with a 'real-time' correction function is shown in Figure 1-1. In the updated model, measurements are collected at the same time as the model is being run. The feedback information from the comparison between the measurements and model outputs is calculated by the 'updating procedure' (which is a DA algorithm in this case) and utilized to update the initial state and/or parameters of the forecast model.

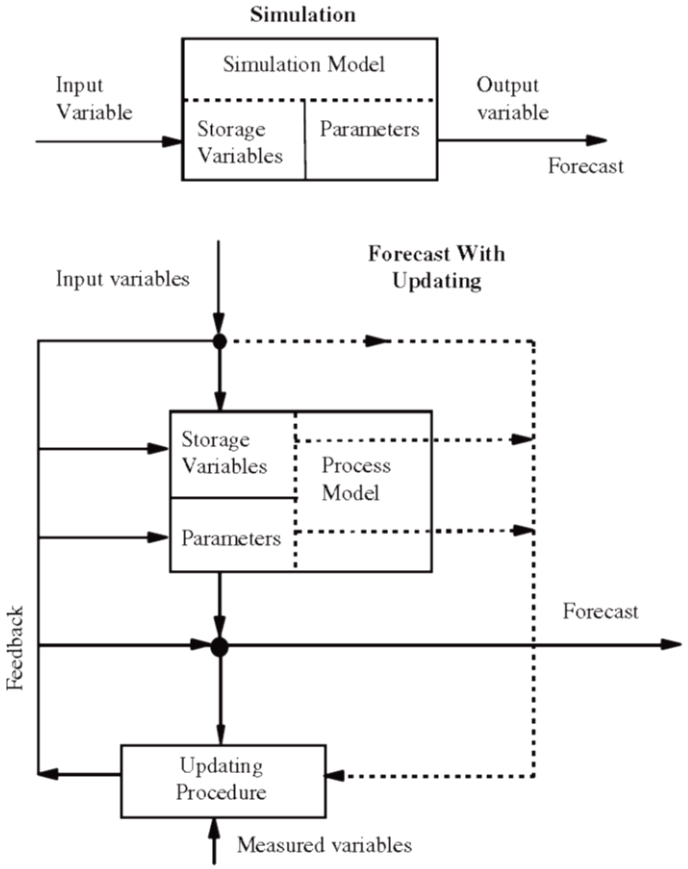


Figure 1-1 Schematic Diagram of 'real-time' updating procedures (WMO, 1992)

Many approaches have been developed in hydrological DA (Houser et al., 2005; Walker and Houser, 2005), and the Kalman Filter is a widely used sequential method in hydrological DA mostly due to its simple implementation procedures. The conventional Kalman Filter is only applicable in a linear system in which both dynamic and observation transition models are linear (Kalman, 1960), and the Extended Kalman Filter (EKF) is a straightforward derivation of the

conventional Kalman Filter designed to cope with a nonlinear system. For a nonlinear system with lower complexity, it is arguably the *de facto* standard method to expand the Kalman Filter for nonlinear systems due to its good performance in practice (Rabier, 2003). However, the application of EKF in distributed or semi-distributed rainfall runoff models is rare so far due to the following:

- the instability of EKF in complex nonlinear models (Evensen, 1994)
- the very large computational cost if a fully distributed state vector is used
- difficulty in linearizing rainfall runoff models, which encourages alternative methods such as the Ensemble Kalman Filter (EnKF).

The purpose of this study is to explore the possibility of updating the SWAT model using EKF, thereby improving its capability of issuing *forecasts* of hydrological states. Two types of observation, streamflow at watershed outlet and *in situ* surface soil moisture, are assimilated into SWAT. Streamflow is chosen because it is usually the only available observation in practice. The surface soil moisture is not currently as available as streamflow; however, this might change with the development of remote-sensing technologies (Brocca et al., 2011). Meanwhile, in the foreseeable future, the algorithm developed with *in situ* observation data will provide a ready-to-use tool in the application of remote-sensing data (Chen et al., 2011). This study also aims to compare EKF state-parameter assimilation with output assimilation, in which the output error, instead of the model itself, is updated in the *forecast*. Another topic discussed in this study is the robustness of EKF assimilation with erroneous precipitation input in the *forecast* after SWAT is updated with EKF.

In order to focus on the validation of EKF assimilation, it is presumed that the meteorological input at the *forecast* stage after using EKF can be replaced by observation or be predefined for research purposes. Hence the streamflow and soil moisture '*forecast*' of SWAT is actually a 'perfect forecast' (Thiboult and Anctil, 2015) instead of a real forecast.

The flood forecast is critical, especially to larger rivers, given the catastrophic consequences if an accurate forecast cannot be obtained in advance. Therefore, the upstream section of the Senegal River in western Africa is chosen for the streamflow assimilation. The assimilation of *in situ* surface soil moisture is expected to serve as a pilot project for the assimilation of remote-sensing soil moisture data. It is crucial to assure the quality and representativeness of the *in situ*

observations within the watershed. The Little Washita experimental watershed has a dense soil moisture observation network run by the USDA ARS office, making it the perfect candidate for this study.

1.2 Objectives

The primary objective of this study is to seek a parsimonious DA procedure to improve the simulation and *forecasting* of a distributed (semi-distributed) rainfall-runoff model. The ideal method is expected to balance model performance, generalizability and computational cost. To be specific, this study is devoted to the coupling of the SWAT model with EKF DA in the simulation and *forecasting* of streamflow and soil moisture. This objective is reached from the following aspects:

- Development of a framework for the SWAT-EKF DA system
- The assimilation of streamflow observations at the watershed outlet to improve the flood *forecasting* capability of SWAT
- The assimilation of *in situ* surface soil moisture in improving the simulation and *forecasting* of soil moisture in deeper layers of the soil profile as well as streamflow at the watershed outlet
- Comparing the effectiveness of EKF state-parameter assimilation with output assimilation in flood *forecasting*
- Discussing the robustness of EKF assimilation regarding the influence of erroneous precipitation input in the *forecasting* of streamflow and soil moisture

1.3 Novelty

1) The application of EKF in a distributed hydrological model. EKF has rarely been used in hydrological data assimilation, much less in the case of a distributed hydrological model such as SWAT. The coupling of EKF with the SWAT model is challenging mostly because of the 'distributed' feature of SWAT and the difficulty in linearizing the model. The solution to these issues is key to applying EKF in SWAT, and many ideas have been developed in this study to tackle these issues.

2) The use of the ratio method in the update of state. Instead of stacking all distributed variables in the state vector, only the watershed scale variables are involved in the online EKF calculation in order to obtain an *a posteriori/a priori* ratio, and then the HRU scale variables are updated with this ratio offline. The ratio method mitigates the computational burden of EKF

exponentially by reducing the degree of freedom of the state vector. It also avoids the quantification of the spatial error correlation between HRU variables, which is critical to the generalizability of this method to operational applications.

3) The numerical linearization of SWAT. An analytical solution to the Jacobian matrix is impossible, as the state-space equations based on SWAT are implicit. A numerical solution with the original Fortran version of SWAT with perturbed state variables would influence the online operation of SWAT, as such operation is sequential. In this study, the Matlab version of SWAT 2009 is applied to facilitate the offline Jacobian matrix calculation without influencing the online running of SWAT.

In addition to the above innovations in the implementation of EKF, this study also makes contributions to the development of Kalman-type hydrological DA. These contributions include but are not limited to:

- Assessing the performance of the assimilated model in *forecast* mode. This has been ignored in most previous hydrological DA studies, as they generally focus on the experimental stage of DA.
- Developing a hybrid method that combines state-parameter assimilation with output assimilation. This new method combines the robustness of the former and the short-term effectiveness of the latter.
- Quantified analysis of the impact of precipitation forecast error on the assimilated model

1.4 Thesis Overview

The rest of the thesis is organized as follows:

Chapter 2 reviews the literature on Kalman-type hydrological DA. The theory, scope and challenges of Kalman filters are introduced, and special attention is given to the comparison of EKF with EnKF. This chapter is based on the following publication:

Leqiang Sun, Ousmane Seidou, Ioan Nisto, Kailei Liu. A Review of Kalman type Hydrological Data Assimilation (2015). Hydrological Science Journal, doi:10.1080/02626667.2015.1127376.

Chapter 3 is the streamflow assimilation study in which the implementation procedures of coupling EKF with SWAT have been described in detail. This chapter is based on the following publication:

Leqiang Sun, Ioan Nistor, Ousmane Seidou. Streamflow Data Assimilation in SWAT Model Using Extended Kalman Filter (2015). Journal of Hydrology, 531 (2015). 671-684. doi:10.1016/j.jhydrol.2015.10.060.

Chapter 4 is the surface soil moisture assimilation study in which the factors that influence the performance of EKF assimilation, such as the setting of state and observation vector and model error quantification, have been discussed. This chapter is based on the following publication:

Leqiang Sun, Ousmane Seidou ; Ioan Nistor , et al. *The assimilation of soil moisture and streamflow with the Extended Kalman Filter in SWAT model (2016). Submitted to Journal of Hydrology.*

Chapter 5 compares the EKF state-parameter assimilation with AR output assimilation in flood *forecasting* with SWAT. This chapter is based on the following publication:

Leqiang Sun, Ousmane Seidou ; Ioan Nistor. Towards streamflow forecasts improvement of the deterministic hydrological model: state assimilation versus output assimilation (2015). Submitted to ASCE Journal of Hydrologic Engineering.

Chapter 6 discusses the impact of erroneous precipitation forecasting on streamflow and soil moisture *forecasts* issued with the assimilated SWAT model.

Chapter 7 summarizes the conclusions and provides recommendations for future work.

Chapter 2 Literature Review

There is great potential in Data Assimilation (DA) for the purposes of uncertainty identification and real-time correction of hydrological models. This chapter reviews the latest developments in Kalman filters, particularly nonlinear Kalman Filters such as the Extended Kalman Filter (EKF) used in hydrological DA. Hydrological DA targets, methodologies and applicability are examined. The recent applications of EKF and EnKF in hydrological DA are summarized and assessed. Lastly, this review highlights the existing challenges in the implementation of EKF and EnKF, especially error determination and joint parameter estimation.

2.1 Introduction

In the past few decades, hydrological models have significantly benefited from improvements in computational capacity and the availability of multi-source measurement data to simulate and forecast the hydrological processes. However, due to the increasing involvement of the uncertainties stemming from initial conditions and inputs and outputs, operational hydrological models will remain inherently imprecise.

Data assimilation (DA) is a procedure developed to optimally merge information from model simulations and independent observations with appropriate modeling (Liu et al. 2012b). DA could provide optimized initial conditions, updated parameters and even improved structures for dynamic models. Originally used in areas like atmospheric and oceanographic science (Dee 1995, Derber and Rosati, 1989), DA has drawn more and more attention from hydrologists for its performance in the real-time correction of hydrological models (Robinson Lermusiaux, 2000).

Walker and Houser (2005) and Reichle (2008) summarized the basic hydrological DA methods; Moradkhani (2008) reviewed the remote-sensing measurement techniques and their applications in data assimilation; Liu and Gupta (2007) discussed the role of DA in addressing the uncertainties in hydrological models and discussed the application of the hydrological DA in operational scenarios (Liu et al., 2012b); and Montzka et al. (2012) examined the joint assimilation of observational data precedents from different spatial scales and different data types. Reviews that focus on a specific method are relatively rare compared to reviews with a wider scope (e.g., Van Leeuwen, 2009).

The Kalman filter is a classic sequential method that has been widely used in hydrological DA for more than two decades (Evensen, 1994; McLaughlin, 1995). Compared to other methods such as particle filters and variational methods, the Kalman filter is easier to implement and can produce comparable or better results with a lower computation demand (Weerts and El Serafy, 2006; Abaza et al., 2014b). Furthermore, the Kalman filter is very flexible in coupling with hydrological models and has more derivative variants than any other method. Among these variants, the Extended Kalman Filter (EKF) and the Ensemble Kalman Filter (EnKF) were both developed to extend the application of the Kalman filter to nonlinear systems; also, they are the two main descendants of the linear Kalman filter. EKF applies a straightforward Taylor extension scheme to linearize a nonlinear system. In spite of its advantages, EKF has endured a poor reputation for being unstable when applied to complex nonlinear hydrological models. EnKF avoids direct linearization by statistically analyzing the ensemble members, and although it increases the computational cost, EnKF is one of the most widely used hydrological DA methods. Despite some comparison case studies (Reichle et al., 2002b; El Serafy and Mynett, 2004; Dumedah and Coulibaly, 2012), there lacks a critical review that specifically focuses on the Kalman filter. The objective of this review is to fill this gap by assessing the latest developments and challenges of Kalman-type hydrological DA, especially EKF and EnKF. However, this review is not intended to judge which method is better, but rather regards them as different solutions to the same problem. This is not only because they are both derived from the linear Kalman filter, but also because they face some similar issues when applied in hydrological DA.

In the rest of this chapter, hydrological DA, including its targets and methods, is introduced in Section 2. Kalman filter (including EKF and EnKF) theories and state of the art applications in hydrological DA are discussed in Section 3. Section 4 discusses the issues regarding the implementation of Kalman filters in hydrological models, and the summary and conclusions are given in Section 5.

2.2 Hydrological Data Assimilation

2.2.1 Hydrological DA targets

DA was first used in the 1950s for numerical weather forecast models; however, hydrologists did not pay attention to it until the 1990s (Evensen, 1994; McLaughlin, 1995). Proper use of DA

may help to handle the uncertainties from model inputs, initialization and propagation of the states, model structures, and even model parameters (Vrugt et al., 2006; Liu and Gupta, 2007; He et al., 2012). Meanwhile, global DA may improve regional field estimations by achieving more accurate external boundary condition estimations (Robinson and Lermusiaux, 2000), and local DA could also lead to improved global estimations (Clark et al., 2008).

The development of remote sensing has promoted the application of DA in hydrological models. The remotely-sensed hydrological data that currently have the potential to be applied in hydrological models includes (Walker and Houser 2005, Houser et al. 2005, Xu et al. 2014):

1) overland parameters (e.g., topography, land cover, albedo)
2) forcing inputs (e.g., precipitation, humidity, and temperature)
3) states (e.g., soil moisture, snow cover)
4) fluxes (e.g., carbon flux)

The overland parameters are usually regarded as 'static' in models, even though this is not true in long-term forecasts. Forcing inputs have the potential to replace traditional ground-based observations with the development of remote-sensing instruments and more accurate retrieval algorithms. Neither overland parameters nor forcing inputs are the main interests of DA for hydrological modeling at this stage. The states and fluxes provide validation to the intermediate processes of the models. Such validation was not available before, and hence it drew extra attention.

Snow cover has a high albedo and thermal properties, as well as a medium-term water storage capacity; therefore, the assimilation of snow observations could improve hydrological predictions (Walker et al., 2003). Andreadis and Lettenmaier (2006) used EnKF to assimilate remotely-sensed MODIS SCE snow observation data into the VIC hydrological model in order to update snow water equivalent (SWE) estimates. From these estimates, a simple snow-depletion curve scheme from SNOTEL SWE data and MODIS imagery (Moradkhani, 2008) was used to form the observation operator. Clark et al. (2006) assimilated the observations of a snow-covered area (SCA) to update the hydrological model, and they found that the assimilation of SCA information results in minor improvements in the accuracy of streamflow simulations near the end of the snowmelt season. Both of them found that the snow cover update works better during the snowmelt season than the snow accumulation season. Further research with

regard to snow cover assimilation can be found in the literature (Parajka and Blöschl, 2008; Rodell and Houser, 2004; Sheffield et al., 2003; Pan et al., 2003; Kumar et al., 2008).

Galantowicz et al. (1999) demonstrated Kalman filter retrieval of soil moisture profiles and temperatures from L-band radio brightness observations, while Crow and Wood (2003) applied EnKF to assimilate the airborne measurements of surface brightness temperature into a TOPMODEL-based Land–Atmosphere Transfer Scheme (TOPLATS). Due to the short memory the land surface skin temperature holds, it is suggested to combine other state variables with longer memories, such as deeper soil temperature or moisture, to obtain longer-term DA effectiveness (Walker et al., 2003). The surface temperature products are also sensitive to terrain, vegetation and cloud contamination; hence, multiple sets of products and better retrieval algorithms are required for continuous DA operation (Huang et al., 2008).

Soil moisture is the key variable in runoff generation process, and it is logical to assimilate soil moisture because it is a continuous storage indication as well as a natural bridge to transform the hydrological model into state space equations. The correct estimation of antecedent soil moisture content is critical to streamflow simulation when the soil is neither too dry nor over-saturated (Reichle et al., 2002a). Soil moisture assimilation has been proved beneficial to streamflow predictions (Chen et al., 2014; Wanders et al., 2014b). The applicability of soil moisture requires the ground-based measurements to be dense enough (Chen et al., 2011); otherwise, spatial interpolation errors have to be considered. Satellite remote-sensing can provide an economical and continuous observation of soil moisture.

Parrens et al. (2014) assimilated *in situ* soil moisture observations into a soil model at a local scale, while Reichle et al. (2002a) assimilated L-band microwave brightness temperature into a Land Surface Model to estimate near-surface soil moisture. Crow and Ryu (2009) used remotely-sensed soil moisture retrievals to correct soil moisture state and satellite rainfall products, and Alvarez-Garreton et al. (2014) assimilated the surface soil moisture and the soil wetness index derived from the passive microwave AMSR-E. Using EnKF, Brocca et al. (2012) assimilated the surface and root-zone soil moisture products derived from the active microwave ASCAT into a rainfall runoff model. Recently, soil moisture assimilation has also been used to assist in parameter identification for hydrological models (Wanders et al., 2014a; Tran et al., 2014).

Despite the existence of some root-zone soil moisture byproducts from surface-soil moisture observations (Wagner et al., 1999; Das et al., 2008), one of the drawbacks of remotely sensing soil moisture is that it can only monitor surface or near-surface moisture (Moradkhani, 2008; Han et al., 2012). Preliminary experiments showed that surface soil moisture assimilation has a minimal effect on the simulation of deep layer soil moisture (Chen et al., 2011), although the latter has a more significant impact on runoff simulations (Brocca et al., 2012; Han et al., 2012). Houser et al. (1998) argued that remotely-sensed soil moisture must be supplemented by *in situ* surface and root zone observations in order to specify error correlation, calibrate parameters, and validate the model-calculated fields. Draper et al. (2011) found that it might be more effective to address the cause of model bias instead of relying on soil moisture assimilation to correct it. Some research has combined soil moisture assimilation with the correction of other rainfall runoff model forcings, such as precipitation, in order to improve streamflow predictions (Chen et al., 2014; Massari et al., 2014). The application of soil moisture remote-sensing may also involve issues like rescaling (Kaheil et al., 2008; Sahoo et al., 2013), error evaluation (Doubková et al., 2012; Alvarez-Garreton et al., 2013), and radiative transfer modeling (Verhoef and Bach 2003; Reichle, 2008), among others. A more detailed explanation of the limitations of soil moisture assimilation can be found in the literature (Vereecken et al., 2015).

Streamflow is the most commonly used prognostic observation variable, and is sometimes the only one available (Clark et al., 2008; Abaza et al., 2014a; Trudel et al., 2014, Samuel et al., 2014; Randrianasolo et al., 2014). Great efforts have been made to improve streamflow forecasting using output assimilation/error assimilation over the past two decades (Anctil et al., 2003; Yu & Chen, 2005; Broersen, 2007; Sene, 2008). The output assimilation/error assimilation methods treat streamflow forecasts as a pure model output and update it by adding errors calculated with another independent procedure or model. Such procedures/models could either be nonlinear, like Artificial Neural Networks (ANNs) (Anctil et al., 2003), or linear, such as the Autoregressive–Moving-Average (ARMA) model (Broersen, 2007; Chen et al., 2015). Output/error assimilation of streamflow is relatively simple to implement, as there is no feedback to the original rainfall runoff model from the modification of model outputs.

Besides being the direct assimilation goal (Liu et al., 2012a), streamflow is also the primary observation in assimilating other state variables and parameters (Coustau et al., 2013). In

hydrological DA, streamflow is often treated as a diagnostic variable, and is therefore not updated directly (Clark et al., 2008). In the case of the nonlinear measurement operator, it is "probably" the only choice to augment the state vector with the streamflow (Evensen, 2003). By doing so, the nonlinear measurement operator would be reduced to a linear matrix; otherwise, one would have to linearize the measurement operator. Pauwels and De Lannoy (2009) attempted to linearize the nonlinear discharge-watershed storage relationship (observation operator) in a 'brutal' way. Linearization was undertaken within both a simple time series model and a conceptual model HBV, and the results indicated that to obtain a better assimilation result, direct linearization should be bypassed.

Streamflow assimilation is different from other state variables because it involves the issue of routing. When runoff is assimilated into a hydrological model at a given time step, not only does the state of the watershed at a given location (e.g., outlet) need to be updated and propagated forward, but also the state at a number of different locations at multiple previous time steps (Pauwels and De Lannoy, 2006). This is especially the case in large and distributed hydrological models. Many authors actually choose not to get into the details of complicated channel networks in order to reduce the complexity of the DA implementation (Weerts and El Serafy, 2006; Clark et al., 2008).

Despite the unique challenges faced by the assimilation of these observations, they share some common issues in their implementation, most notably in the quantification of observation errors and the determination of observation operators to connect them with the model output (Andreadis and Lettenmaier 2006). It is also worth pointing out that the assimilation of one variable does not necessarily improve the estimation of the other variables. Trudel (2014) reported that assimilation of the streamflow at the outlet of a watershed would distort the estimate of soil moisture. The combination of multiple observation variables seems to be superior to single-variable assimilation (Xie et al., 2008; Trudel et al., 2014).

The early 21st century saw the development of various global and regional land assimilation systems (Cosgrove et al., 2003; Rodell et al., 2004; Mitchell et al., 2004; Kumar et al., 2008). These systems enable the incorporation of multi-source observations, multi-models and multi-assimilation schemes in creating optimal land surface state outputs. Due to possible water and energy balance issues, a major concern in land assimilation systems is that if they should be

coupled to the atmospheric models (Walker et al., 2003). Betts (2003) compared the water budget of ECMWF 40-year reanalysis and NASA DAO fvGCM with the hydrological balance of the VIC model and the radiative fluxes with the basin averages derived from ISCCP. They found that the runoff from both atmospheric models was significantly underestimated compared to the VIC runoff simulation, which is consistent with the observed streamflow. Large errors are also observed for radiation fluxes and surface temperature between the ECMWF 40, fvGCM and ISCCP data. Pan (2006) suggested a constrained EnKF to maintain the benefit of DA without violating the water balance principle. Boulet (2000) and Bøgh (2004) used a simple water and energy balance model that allows the direct application of remote-sensing data.

2.2.2 DA methods

DA methods can be divided into different categories based on different standards (Rakovec et al., 2015). According to the dimension they focus on, these approaches can be classified as objective analysis methods and time-dependent methods (Wang and Kou 2009).

2.2.2.1. Objective analysis

Objective analysis aims to minimize the error between the observation field and background field (usually the output of the numerical models) by 'fusing' the new observations into the background with spatial dimensions. Typical objective analysis methods include successive correction (Cressman, 1959; Barnes, 1964), optimal correction (Gandin, 1963; Lorenc, 1981), statistical bias correction (Piani et al., 2010), Newtonian nudging (Houser et al., 1998; Paniconi et al., 2003), variational methods (Reichle et al., 2001; Seo et al., 2003; Seo et al., 2009; Navon, 2009), etc. Detailed descriptions about the theory and development of the above methods can be found in the literature (Navon, 2009).

Since objective analysis methods mostly take 'snapshots' of the background at a given time (McLaughlin, 1995), the temporal dimension is usually not well incorporated with the spatial dimensions. Even though some of them do consider the temporal dimension, such as 4D VAR and Newton nudging, time-variant objective methods are most simply viewed as a dynamic extension of the time-invariant version of the objective methods (McLaughlin, 2002).

Among the various objective analysis methods, the successive correction and nudging methods fail to consider errors in the observations, while 3D VAR and 4D VAR ignore the uncertainties

in the models. Objective analysis methods usually involve huge computations (Clark et al., 2008). Houser et al. (2005) suggested that the adjoints should be calculated as the model is developed, but this is by no means an easy task for distributed hydrological models.

2.2.2.2. Time dependent methods

Time-dependent methods use a probabilistic framework and estimate the system state sequentially by propagating information forward in time (Bertino et al., 2003). The strength of the time-dependent methods is in time-series analysis. Time-dependent methods work on a fixed but moving time window, and only the most recent observations that fall into this window are incorporated into the final estimation results. Typical time-dependent methods that are frequently used in hydrological DA include the linear Kalman filter (Kalman 1960) and its variants, such as the Extended Kalman Filter (EKF) (Puente and Bras, 1987), the Ensemble Kalman filter (EnKF) (Evensen, 1994), the Unscented Kalman Filter (UKF) (Wan and Van Der Merwe, 2000), the Particle Filter (Pham, 2001; Moradkhani et al., 2005a; Weerts and El Serafy, 2006), and the H-infinity filter (Moradkhani et al., 2005a; Wang and Cai, 2008; L ü et al., 2010), among others. There are also some alternative approaches to solving specific issues in hydrological DA, most notably in the application of a genetic algorithm (GA) in the estimation of pixel-based soil hydraulic parameters for hydroclimatic modeling (Ines and Mohanty 2008) and evolutionary-based assimilation in streamflow simulations in ungauged watersheds (Dumedah and Coulibaly 2012). The spatial variation in the relationship between the background field and observation field is of less concern in time-dependent methods. Meanwhile, time-dependent methods are capable of handling more uncertainties (Moradkhani, 2008) and are less complex to implement compared to some objective analysis methods that require an inverse or a joint model (Bertino et al., 2003).

Time-dependent methods can be well explained with the Wiener-Kolmogorov estimation theory (Wiener, 1949; Kolmogorov et al., 1941). Suppose $s(t)$ is the original signal and $\widehat{s(t)}$ is the estimated signal; the estimation error is then defined as:

$$e(t) = s(t+\alpha) - \widehat{s(t)} \tag{2-1}$$

Where α is the delay of the estimation. In other words, the error is the difference between the estimated signal and the true signal shifted by α. Depending on the value of α, the estimation problem can be described as shown in Figure 2-1, i.e.:

i) if α>0 , the estimation is a prediction problem (error is reduced when $\widehat{s(t)}$ is similar to a later value of *s(t)*);

ii) if α=0 , the estimation is a filtering problem (error is reduced when $\widehat{s(t)}$ is similar to *s(t)*) ;

iii) if α<0 , the estimation is a smoothing problem (error is reduced when $\widehat{s(t)}$ is similar to an earlier value of *s(t)*).

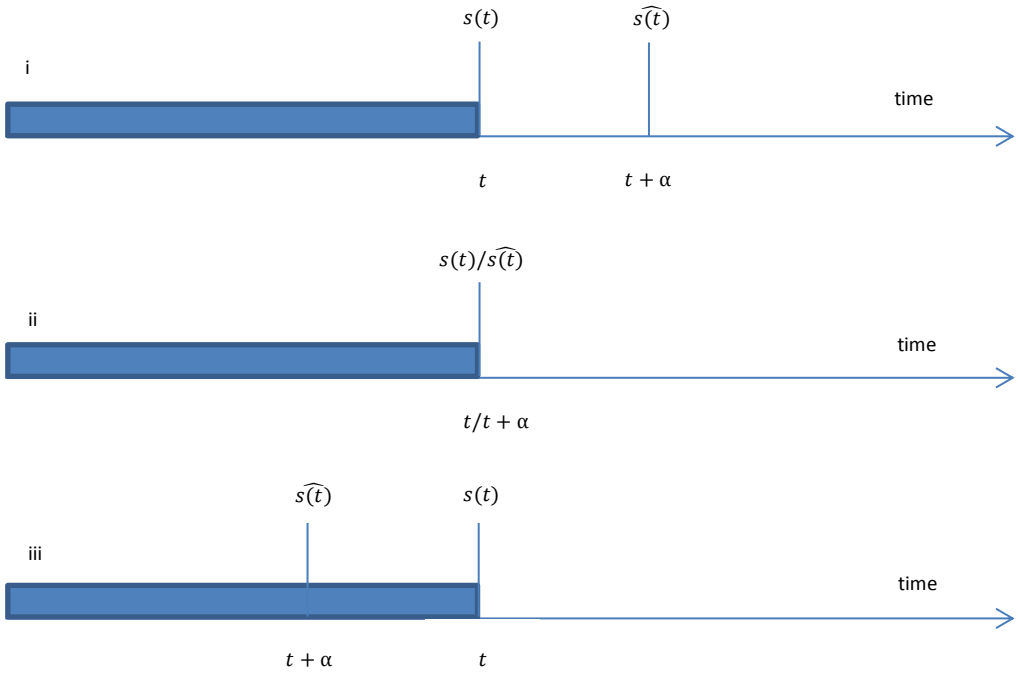


Figure 2-1 Types of estimation problem: i) prediction; ii) filtering; iii) smoothing

In a broad sense, both smoothing and filtering are time-dependent DA methods, and both combine the advantages of measurements and model outputs. In a narrower sense, only filtering problems count as time-dependent DA methods, as DA only deals with 'real-time' measurements instead of historical ones. In the case that the information in the time series is only propagated forward without any backward loop or window, the time-dependent methods are also termed "sequential" methods. The prediction problems are often used as the validation of the smoothing and filtering problems. The traditional batch calibration of hydrological

models can be categorized as a smoothing problem (McLaughlin, 2002). Filtering technologies usually play an important role to improve the prediction after the calibration; hence, they are widely used in the real-time modules of operational models (Divac et al., 2009; McMillan et al., 2013).

2.2.2.3 Applicability of DA methods to hydrological models

Many approaches have been applied in hydrological DA (Houser et al., 2005, Walker and Houser, 2005). The most commonly used methods include the variational method, particle filters and Kalman filters. Although traditionally dominant in numerical weather forecasts, variational methods have not been widely used in hydrological DA. Variational methods such as 3D VAR assume that forecast error statistics are isotropic and largely homogeneous with little variation over time (Houtekamer and Mitchell, 1998). Consideration of the time dimension would overwhelmingly increase the computational burden. It is also complicated to develop an adjoint model for distributed hydrological models (Clark et al., 2008). Successful cases using variational methods in hydrological DA are mostly based on simpler lumped models (Seo et al., 2009, Seo et al., 2003, Abaza et al., 2014b). Abaza et al. (2014b) compared the classic variational method with EnKF, and they found that the latter is more stable in streamflow assimilation.

Particle filters are also widely used in hydrology DA (Weerts and El Serafy, 2006, Pham, 2001, Moradkhani et al., 2005a). One of the major advantages of particle filters is that the system does not need to be Gaussian (Liu et al., 2012b). Particle filters are also better at handling model nonlinearities compared to other sequential methods (Moradkhani et al., 2005a). The fact that particle filters use full prior density means that this method is more computationally intensive (Weerts and El Serafy, 2006). The operational applications of particle filters in distributed hydrological models are also limited due to the setup of the particle numbers (Liu et al., 2012b).

Despite the development of DA methods in ‘interpolation’, ‘smoothing’ or ‘filtering’, not many of them are actually extensively validated in ‘forecast’ mode, not to mention in operational hydrological forecasts (Liu et al., 2012b). One of the problems that hydrological DA faces is the short ‘efficient period’. The updating of model structures and storages generally has a major impact only on forecasts with shorter lead times (El Serafy and Mynett, 2004, Knight and Shamseldin, 2006). A possible reason for this is that DA deals with uncertainties from different

data sources in a statistical way, rather than in a physical or mechanical way (McLaughlin, 2002). It somehow 'detours' the imperfect model structure and parameters in hydrologic models by adding more weight to the model uncertainties, thereby emphasizing observations in the final results.

Numerical Weather Prediction (NWP) is an initial problem that is very sensitive to state variation. Hydrological modeling, on the other hand, is a process problem that relies on model structures and parameters. For this reason, DA might not work as well in hydrology as in NWP. However, because of the 'conceptual' nature of most hydrological models, DA does have the potential to improve model forecasts, either by updating the initial condition or modifying the model parameters.

A potential development direction of hydrological DA is the combination of objective analysis methods and sequential methods. 4D VAR expands the strength of 3D VAR by considering the temporal evolution of variables within a fixed time window. However, depending on the model's structure, it could be expensive to calculate the gradients of the cost function (Bin and Ying, 2005). Methods that combine the merits of both objective analysis and sequential methods, such as 4D VAR and EnKF, are generally regarded as the most promising DA approach in Numerical Weather Prediction (Lorenc, 2003, Kalnay et al., 2007). However, as far as the authors are aware, such a DA method still does not exist in operational hydrological simulation and forecasting.

2.3 EKF and EnKF

The linear Kalman filter is a classic sequential method. Together with its multiple nonlinear variants such as EKF and EnKF, the Kalman filters have become a very promising method cluster used in hydrological DA.

2.3.1 Linear Kalman Filter

A Kalman filter is an optimal estimator that recursively couples the most recent measurements into the linear model to update the model's state output (Kalman, 1960). Under the assumption that the linear system is a stochastic process with Gaussian noises, the Kalman filter produces the best estimation with a minimum mean square error.

The Kalman filter works on a stochastic system in the form of state space equations (Hamilton, 1994):

$$x_{k+1} = M_k x_k + B_k u_k + \eta_k \tag{2-2}$$

$$\mathrm{y}_k^o = H_k x_k + \varepsilon_k \tag{2-3}$$

Equation (2-2) is the model function that propagates state x from step k to step $k+1$. x_k is the *a posteriori* state vector and x_{k+1} is the *a priori* state. M_k is the model function (also known as the dynamic function or dynamic operator). B_k is a linear matrix to convert the dimension of residual vector u_k to state vector x_k. u_k is the 'forcing' term of the model in the form of linear residuals. η_k is the model estimation error (covariance matrix=Q_k). Sometimes, in order to simplify a problem, u_k is regarded as part of M_k and its estimation error is included in η_k, and in this case B_k=0.

Equation (2-3) is an observation function that relates the state vector x_k to the observation y_k^o, with observation error ε_k (covariance matrix=R_k). To guarantee the linearity of the system, the observation operator H_k should also be linear. Normally, the number of observation variables is limited, while state variable combinations could be infinite. It is likely that the observation vector dimension (degree of freedom) is much smaller than that of the state vector.

A typical implementation of the Kalman filter is described below (Dr écourt, 2003). At time step 0, set up the initial estimation error covariance matrix P_0^a of the initial state x_0^a. The model propagation without an explicit form of u_k is given by:

$$x_k^f = M_{k-1} x_{k-1}^a \tag{2-4}$$

$$P_k^f = M_{k-1} P_{k-1}^a M_{k-1}^T + Q_{k-1} \tag{2-5}$$

The superscript 'a' represents the *a posteriori* estimation, and the superscript 'f' means the a priori estimation.

The state is then updated by:

$$x_k^a = x_k^f + K_k \cdot (y_k^o - H_k x_k^f) \tag{2-6}$$

where $y_k^o - H_k x_k^f$ is the innovation and the Kalman gain K_k is calculated with:

$$K_k = P_k^f H_k^T [R_k + H_k P_k^f H_k^T]^{-1} \quad (2\text{-}7)$$

The estimation error is finally updated by:

$$P_k^a = (I - K_k H_k) P_k^f \quad (2\text{-}8)$$

With the updated estimation error, the filter can restart from equation (2-4) and (2-5) to begin another loop.

Equation (2-6) shows that the updated state is a linear combination of the observation and the model estimate (Drécourt, 2003). The gain will change between 0 and H^{-1} depending on the uncertainty comparison between the model outputs and the observations. For example, if the dynamic model is 100% accurate (which is unrealistic), then the new observation would not be considered at all in the updated state, as the gain is 0, and vice versa.

The linear Kalman filter tends to be more commonly used in stochastic models (Bolzern et al. 1980; Bergman and Delleur 1985; Szöllősi-Nagy and Mekis 1988) and channel routing problems (Huang, 1999), which are easier to linearize (Wu et al., 2008; Fan, 1991; Georgakakos and Bras, 1982; Sun et al., 2013). However, there are a few successful cases where the linear Kalman filter was applied in complex rainfall runoff models. Liu et al. (2011) and Lee and Singh (1999) coupled a Tank model with a linear Kalman filter to estimate the model's parameter and outflow respectively. Kim et al. (2005) applied the Kalman filter in a one dimension physically based distributed CDRMV3 model with the storage of whole watershed as state and outlet discharge as measurement, while a discharge-storage (Q-S) curve was used as the observation equation, in addition to a Monte Carlo ensemble being used as the dynamic equation. The drawback of such an operation is that it could be difficult to find the linear relationship between discharge and storage, even at a local scale.

2.3.2 Nonlinear Kalman Filter

The Kalman filter is only applicable in a linear system where both the model function M_k and the observation function H_k are linear. In the case of a nonlinear system (Drécourt, 2003):

$$x_k^f = M_k(x_{k-1}^a, \eta_k) + u_k \quad (2\text{-}9)$$

$$y_k^o = H_k(x_k^f, \varepsilon_k) \quad (2\text{-}10)$$

Where M_k is the model function and H_k is the observation function, the linear Kalman filter is not applicable even if only one of them is nonlinear.

A hydrological system is a highly complicated nonlinear system with a large degree of freedom. In cases like stochastic black-box stream flow prediction models, a linear Kalman filter might apply. However, it is not feasible to count on a linear Kalman filter in the DA of distributed/semi-distributed hydrological models.

To cope with nonlinear problems, many different versions of modified Kalman filters have been developed (Shamir et al., 2010; Muluye, 2011; Dumedah and Coulibaly, 2012; de Rosnay, et al, 2013; Chen, et al., 2013; Gharamti and Hoteit, 2014; Rakovec et al., 2015). A large proportion of these new methods is based on EKF and EnKF.

2.3.2.1 EKF

2.3.2.1.1 Theory of EKF

A prerequisite to applying EKF is that the nonlinear system should be continuously derivable. Under this condition, a Taylor extension is applied at the estimated point. It is therefore possible to obtain the converted linear dynamic matrix by expanding the nonlinear functions at the estimated point. Apply this to both model function and observation function to get equation (2-11) and equation (2-12) (Walker and Houser, 2005):

$$\dot{M}_k = \frac{\partial M}{\partial x_k}|_{(x_{k-1}^a, 0)} \tag{2-11}$$

$$\dot{H}_k = \frac{\partial H}{\partial x_k}|_{(x_k^f, 0)} \tag{2-12}$$

The linearized $\dot{M}_k$ and $\dot{H}_k$ are then used in the calculation of error covariance propagation; all subsequent steps are the same as in the linear Kalman filter (Lewis et al., 2008). The EKF may be applied for the latest observation as many times as needed to reduce the linearization errors. However, such recursive operations do not guarantee a better estimation than a one-time EKF, and the computational burden is significantly increased (Puente and Bras, 1987).

In many cases, nonlinear systems do not have explicit analytical solutions, and the derivatives can only be calculated numerically. As the state is a vector with multiple variables, the linearized matrix can be expressed as a Jacobian matrix of the partial differential functions.

EKF keeps the first order of the Taylor expansion, while it ignores the higher-order terms. Theoretically, a high-order EKF that keeps more than the first-order terms is more precise, as more information is kept (Tanizaki, 1996, Simon, 2006); however, it is still biased because higher-orders terms are still ignored (Tanizaki, 1996). In reality, there are few cases supporting the superiority of higher-order EKF over first-order EKF (Sadeghi and Moshiri, 2007; Ermolaev and Volynsky, 2014). This is especially true when the analytical solutions of the system are unavailable, because the simultaneous numerical solution of the Jacobian and Hessian matrixes may introduce more errors (Roth and Gustafsson, 2011; Vittaldev et al., 2012).

2.3.2.1.2 Application of EKF

EKF has been widely used in operational soil moisture analysis by various weather agencies (Hess, 2001; de Rosnay et al., 2013; Fairbairn et al., 2014). De Rosnay et al. (2013) described the operational implementation of simplified EKF (SEKF) (Draper et al., 2009; Mahfouf et al. 2009), in which EKF has been simplified by setting constant background errors in the ECMWF land surface analysis system. SEKF linearizes the observation operator with finite differences by adding individual perturbations to each element of the model state vector. Such a feature significantly increases computational cost with the increase of dynamic model resolution. To solve this problem, one should either modify the model structure or calculate the Jacobian matrix in an off-line land surface model simulation (Mahfouf et al., 2009). The comparison between SEKF and EnKF shows that SEKF provides performance similar to EnKF, but does not surpass it (Fairbairn et al., 2014).

Ge (1984) applied EKF on a 3-layer Storage Excess runoff model. Depending on the existence of precipitation and the fulfillment of each soil layer, they set ten different runoff generation scenarios. To avoid defining a threshold function, each scenario was assigned a state space function. By doing so, it was found that only two state space functions needed to be linearized. A Taylor Extension was then used with these two functions analytically. The constants from the truncation of the Taylor extension are treated as the linear-driven terms of the model function and the linear residuals of the observation function. Neither is involved in the state error

propagation explicitly, but they are integral components of the model and observation functions and cannot be neglected.

Kitanidis and Bras (1980a) linearized the Sacramento model by using statistical linearization and a 'describing function' technique. Instead of linearizing the nonlinear formulas directly, a group of simple linear functions that produce similar outputs with the same inputs as the nonlinear formulas is applied. The focus is thus shifted to the determination of the coefficients of the linear functions using statistical methods. This scheme is subjective because the candidate linear functions are not unique, depending on the selection of criteria or the evaluation functions to assess the fitness of the linear functions. Another challenge for this scheme is that it assumes that the inputs follow a constrained Gaussian distribution, and most importantly, that they should fall into a very narrow range for the linear function replacement. The computational burden of statistical linearization is also a big issue.

Instead of implementing analytical linearization, Walker and Houser (2001) numerically derived the dynamic transition matrix in the propagation of the estimation error covariance. The state vector contains three intermediate soil moisture parameters, and the forecast equations of state were linearized with a first-order Taylor series extension. The remote-sensing surface soil moisture is chosen as the observation, which is related to the state vector through a complex nonlinear equation that is also linearized with a Taylor extension. The numerical solution significantly increases computational cost, but not as much as statistical linearization (Kitanidis and Bras, 1980b). Therefore, it is important to provide a state vector of a manageable size. The problem with the numerical solution is in determining the perturbation of the independent variables: small perturbations may lead to numerical problems, while large perturbations will cause a much greater loss of accuracy, and could also risk hitting the nonlinear threshold (Reichle et al., 2002b).

2.3.2.1.3 Assessment of EKF

The robustness issue of EKF has been well discussed in hydrological DA, and much of this concerns the divergence problem (Ljung, 1979). The divergence of EKF can be caused by a number of reasons (Fitzgerald, 1971). The first concerns the inappropriate estimation of model and observation error covariance matrix, which will be discussed extensively in 2.4.2. Secondly, divergence can also be caused by incorrect estimation of the initial estimation error P. Since the

estimation of P at time k relies on P at time k-1, a large error in initial P could recursively propagate until there is a divergence of the EKF (Tao et al., 2005). The third reason concerns the truncation errors arising from the Taylor expansion: although first-order local linearization is adequate to account for differential nonlinearity in some models (Reichle et al., 2002b), the existence of such errors makes it harder for the model errors to meet the whiteness assumption. The last issue is with the errors from the numerical calculation of the Jacobian matrix, as the perturbation scale of the independent variables involves subjective trials.

Many efforts have been made to alleviate the divergence of EKF, and adaptive algorithms (Jwo and Wang, 2007) are widely used to prevent the divergence caused by inappropriate error quantifications. Ljung (1979) introduced an innovation model by adding a parameterized Kalman gain term to the original EKF state space equations; he argued that the main reason for the divergence was the lack of coupling between Kalman gain K and model parameters. Tao et al. (2005) rescaled the error covariance matrix by multiplying them by a factor; once divergence is detected, the Kalman gain is frozen until stability is restored.

Compared to other nonlinear methods, the implementation of EKF is direct and straightforward. Due to its outstanding performance in practice, it is arguably the de facto standard method to expand the Kalman filter to nonlinear systems (Rabier, 2003). However, it is not clear if EKF can outperform other DA methods in nonlinear distributed hydrological models even though strategies such as low rank singular evolutive extended Kalman filter (SEEK) (Pham et al., 1998) can reduce the computational requirements of EKF.

2.3.2.2 EnKF

2.3.2.2.1 Theory of EnKF

The Ensemble Kalman Filter (EnKF) is a Monte Carlo approach for nonlinear filtering problems (Evensen, 1994). There are two types of EnKF (Kalnay et al., 2007): perturbed observation EnKF (Evensen 1994, Burgers et al. 1998) and square root EnKF (Anderson, 2001, Bishop et al., 2001, Whitaker and Hamill, 2002). Perturbed observations cause additional sampling errors (Kalnay et al., 2007), but this type of EnKF can handle nonlinearities better than square root EnKF (Lawson and Hansen, 2004).

EnKF is based on the approximation of the conditional probability densities of the state errors covariance by a large finite number of randomly generated model trajectories. EnKF does not need any derivation of the model operator or observation operator, as is performed in EKF. Instead, it generates a set of realizations (the ensemble) and propagates them through the model operator independently. Then, it derives the *a priori* state error covariance through statistical analysis of the ensemble (Evensen, 1994).

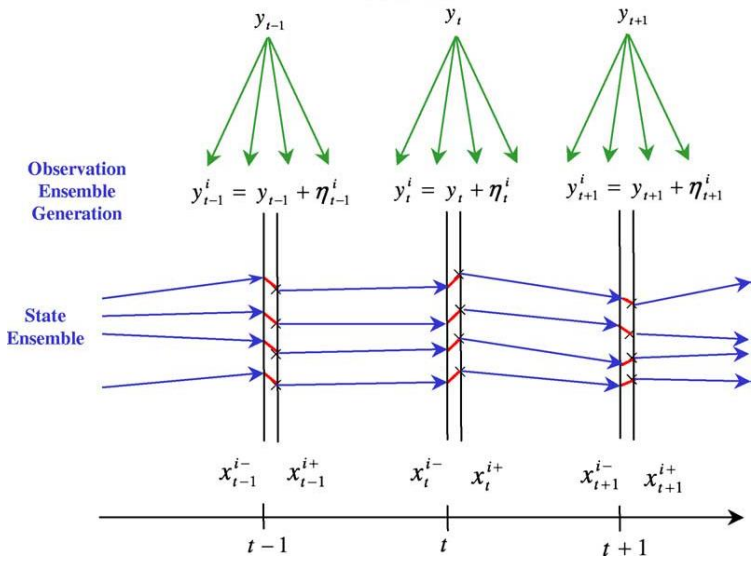


Figure 2-2 Schematic of EnKF with perturbed observations (Moradkhani et al. 2005b)

The general idea of perturbed observation EnKF is demonstrated in Figure 2-2. In the forecast step, the *a priori* state ensemble is created by adding random perturbations to the best estimate of the initial state. Then, the *a priori* error covariance matrices are approximated from the state and model output ensemble error matrices (Gillijns et al., 2006). In the analysis step, the state ensemble are integrated forward in parallel, based on the original model. The observation is perturbed by adding Gaussian noises of measurement errors (Burgers et al., 1998). Then, the *a posteriori* error covariance matrix can be estimated from the ensemble error matrix of the *a posteriori* states. Meanwhile, the Kalman gain is calculated from the forecast error covariance matrices (Reichle et al., 2002b; Moradkhani et al., 2005b).

2.3.2.2.2 Application of EnKF

Although the propagation of state error covariance does not require linearization of the model, the parallel computation of the ensemble means that EnKF is an expensive method to use. Nevertheless, EnKF is a widespread approach in hydrological DA.

Weerts and El Serafy (2006) compared EnKF with two particle filters for flood forecasting with the conceptual rainfall-runoff model HBV-96, and found that EnKF outperformed both particle filters. Dumedah and Coulibaly (2013) compared Evolutionary Data Assimilation (EDA) with methods based on the integration of Pareto-optimality into both an EnKF and a PF (ParetoPF). Rafieeinasab et al. (2014) compared EnKF with a maximum likelihood ensemble filter, which is an ensemble extension of variational assimilation. It is argued that the former is less sensitive to observation, model errors, and parameter uncertainties, and it performs reasonably well with a smaller ensemble size. Borup et al. (2015) discussed the utilization of EnKF in the assimilation of observations that are 'out-of-range', defining this method as 'partial EnKF', and the positive results indicated that EnKF is versatile enough to take advantage of the imperfect but still valuable observation information instead of simply discarding it.

Madsen et al. (2003) presented a procedure to assimilate observed water levels and fluxes in the MIKE 11 Flood Forecasting system using EnKF. Borup (2014) demonstrated the application of EnKF in assimilating water level and flow observations into distributed urban drainage models. Liu et al. (2015) used EnKF to assimilate AMSR-E snow depth into a land surface model for stream flow predictions.

Xie and Zhang (2010) implemented synthetic simulation experiments for the application of EnKF with a SWAT model. Sensitivity analysis with regard to error specification, initial realization, and ensemble size were also demonstrated in their thesis. The application of EnKF in SWAT has also been extensively studied in other works (Han et al., 2012; Lei et al., 2014).

Clark et al. (2008) attempted to apply EnKF to assimilate streamflow observations into a distributed hydrological model; however, the researchers found it unsuccessful due to the inappropriate selection of state variables. Reichle et al. (2002a) assimilated remote-sensing soil moisture into a land surface model with a very large state vector and a small ensemble number, and argued that a large ensemble number is needed to obtain a robust error variance estimation.

Shi et al. (2014) found that the EnKF assimilation of multivariate observations applies strong constraints on parameter estimation in a physically-based land surface hydrological model. Abaza et al. (2014a) introduced hyper-parameter perturbation factors by comparing the hydrological ensemble prediction system (H-EPS) spread to its mean forecast error for the perturbation of the system in EnKF. This research outlined the importance of input replicate generation in the implementation of EnKF. Panzeri et al. (2015) described the implementation of stochastic moment equations (MEs) based on EnKF in groundwater flow simulation, and the researchers found it more efficient than the traditional EnKF.

2.3.2.2.3 Assessment of EnKF

Compared to EKF, EnKF is less likely to diverge because the Taylor expansion and Jacobian matrix calculation are avoided. Meanwhile, EnKF accounts for more model errors, as there is no truncation of higher-order terms as in the Taylor expansion (Reichle et al., 2002b). Considering the computational burden of linearization, it is unrealistic to put all the representative variables from the calculation unit of the distributed model into the state vector (Gillijns et al., 2006). Therefore, it is difficult to consider both the horizontal correlations in the model and the measurement errors in EKF. However, this is not an issue in EnKF (Reichle et al., 2002b).

Computational efficiency is likely the primary concern before one applies EnKF. A major drawback of EnKF is that it assumes that the prior PDFs of the model states are Gaussian; hence, the posterior states are only determined by the first two moments of the prior density (Weerts and El Serafy, 2006). Theoretically, the larger the ensemble, the more accurate the estimation of the mean and variance of prior density (Madsen et al., 2003; Rasmussen et al., 2015). It is argued that the EnKF estimate would not converge to an optimum unless the ensemble size is large, especially in the case of high-dimensional problems (Reichle et al., 2002a, Reichle et al., 2002b, Crow and Wood, 2003). One of the preconditions to consider EnKF superior to EKF is that the ensemble size should reach a certain number of members so that the major errors are statistical noises, rather than a closure problem or unbounded error variance growth encountered by EKF (Evensen, 1997). In reality, the parallel running of distributed hydrological models is a huge challenge with regard to computational capacity. Although the determination of a reasonable ensemble size is regarded as a case by case problem (Xie and Zhang, 2010), more

efforts are expected in deriving efficient algorithms to trade off between ensemble size and balanced forecast analyses, as well as representative error statistics (Mitchell et al., 2002).

2.4 Implementation of Kalman Filters

2.4.1 Filter element selection

An incomplete list of recent hydrological DA work has been made in order to summarize their features, including the DA methods chosen, hydrological models, state, parameters (if any), and observations (see Table 2-1). Depending on the assimilation purpose and the available observations, the selection of state variables to assimilate can be very flexible. A state variable can either be intermediate (e.g., soil moisture content, snow water equivalent) or prognostic (e.g., runoff/discharge). Soil moisture is by far the most popular state variable because of its crucial role in describing the model with state space equations. Nowadays, there are many remote-sensing products that can provide temporally and spatially continuous soil moisture observations (Moradkhani, 2008). On the other hand, streamflow forecasting is one of the ultimate goals of hydrological modeling, and with the easy availability of streamflow observations, these are also a widely used state variable.

Table 2-1 Literature summary of states and parameter(s) selection

Literature	Method	Model	State	Parameter(s)	Observations
(Reichle et al. 2002a)	EnKF	(Reichle et al. 2001)	soil moisture and soil temperature	-	L-band microwave radio brightness
(Madsen et al. 2003)	EnKF	MIKE 11	water level and flux	-	water level and flux
(Moradkhani et al. 2005b)	EnKF	HyMOD	water storage	Rq, Rs, α, bexp, Cmax	stream flow
(Vrugt et al. 2006)	EnKF	SAC-SMA	Six soil zone reservoir variables, three routing variables	14 time invariant parameters	stream flow
(Weerts and El Serafy 2006)	EnKF, Particle filter	HBV96	soil moisture, snow water, interception storage, upper and	-	discharge

			lower zone, etc.		
(Clark et al. 2008)	EnKF	Topnet	soil storage, aquifer storage, surface storage, stream flow prediction	-	stream flow
(Xie and Zhang 2010)	EnKF	SWAT	runoff, water storage. evapotranspiration, etc.	CN2	stream flow, soil water content, evapotranspiration
(Dechant and Moradkhani 2011)	EnKF, Particle filter	Snow 17, SAC-SMA	snow water equivalent (SWE)	All model parameters	microwave radiance
(Chen et al. 2011)	EnKF	SWAT	soil moisture	-	surface soil moisture
(Cammalleri and Ciraolo 2012)	EnKF	SVAT	soil moisture	All model parameters	surface temperature, latent heat flux, soil water content
(Han et al. 2012)	EnKF	SWAT	soil moisture	-	surface soil moisture
(Trudel et al. 2014)	EnKF	CATHY	pressure head, incoming and outgoing discharge	-	stream flow
(Samuel et al. 2014)	EnKF	SAC-SMA	soil moisture content, SWE, etc	10 time variant parameters	stream flow, soil moisture
(Kitanidis and Bras 1980b)	EKF	NWS	soil water content, additional impervious storage, channel storage, routing coefficient	-	stream flow
(Puente and Bras 1987)	EKF	Conceptual lumped model	water equivalent mass in cloud, soil water content, channel reservoir storage, etc.	-	discharge, precipitation
(Francois et al. 2003)	EKF	GRKAL	soil moisture	-	surface soil moisture
(Aubert et al.	EKF	GR4J	water level in soil reservoir and water	-	soil moisture, stream

2003)			level in routing reservoir		flow
(Kumar and Kaleita 2003)	EKF	LSM	soil temperature	-	near-surface soil temperature
(Jonsdottir et al. 2006)	EKF, maximum likelihood	A stochastic model similar to HBV and NAM	temperature, snow, upper and lower water content	All model parameters	discharge
(Lü et al. 2011)	EKF, PSO	1-D Richards' Equation	soil moisture content	Ks,θr, θs,α,n	surface soil moisture

In most cases, the model parameters are assumed to be time-invariant once they are calibrated. However, it is possible to treat some parameters as time-variant, and update them together with the state variables (Samuel et al., 2014), even though this is still not as widespread as it should be. The selection of the parameters is less flexible than that of the state variables because it should coordinate with the state variable update. Besides that, the parameters selected should be sensitive enough to reflect the update (Xie and Zhang, 2013). In the case of a model being simple enough, it is possible to include all the parameters in an assimilation; however, this would present a huge challenge for complex models. He et al. (2012) developed an Integrated Sensitivity and Uncertainty analysis Framework (ISURF) to screen and identify sensitive model parameters and assess the uncertainty structure of model parameters in EnKF.

2.4.2 Determination of errors

Another key issue in the implementation of the Kalman filter is the determination of model and observation errors. The model errors are more difficult to describe as compared to observation errors, because the latter can sometimes be predefined based on measurements, sampling methods or empirical formulas (McMillan et al., 2010). Overestimation of model errors reduces confidence in the model, and thus the filter would rely overly on observations; on the other hand, underestimation of model errors exaggerates the accuracy of the model and also wastes the information from the new observations (Kitanidis and Bras, 1980b). The improper selection of model errors may lead to unacceptable results or a divergence of the Kalman filter. Puente and

Bras (1987) argued that proper error quantification of the model is even more important than the selection of DA methods.

Clark et al. (2008) quantified the errors in streamflow measurements as a fixed proportion of the discharge observations, while the model errors were generated by perturbing the precipitation forcing and model states (e.g., soil moisture and aquifer storage) with varying temporal perturbations. However, the more general method to realize the time variant estimation of model and observation errors is by employing adaptive filtering. If properly used, adaptive filtering can partly mitigate the influence of the inaccurate (if not incorrect) setup of the initial model and observation errors. Adaptive filters are divided into four categories: Bayesian, maximum likelihood, correlation, and covariance matching (Mehra, 1972).

Bayesian methods aim to obtain recursive equations for the *a posteriori* probability density of states and errors given the new observations are available. The Bayesian approach is considered the theoretical basis for all other approaches (Sarkka and Nummenmaa 2009). Examples of Bayesian approaches include state augmentation methods (Wan and Van Der Merwe, 2000) and interacting multiple model methods (Rong Li and Bar-Shalom, 1994). To improve the operational flood forecasting, Li et al. (2014) used a maximum a posteriori method constrained by prior information drawn from flow-gauging data to estimate model and observation errors in the EnKS assimilation of discharge.

The Sage-Husa method (Sage and Husa, 1969; Sage and Wakefield, 1972), which is generally used in discrete linear systems, is a maximum posterior probability method that recursively estimates the mean and covariance of the unknown system and observation noises. The Sage-Husa method is based on the assumption that the system is stationary and that therefore, the model and observation error covariance will eventually converge to constants. In the case of non-stationary systems, Duan et al. (2003) proposed rescaling the model errors with a newly defined time-varying factor to avoid the possible divergence of the filter.

EKF and EnKF are both suboptimal filters, and this is mainly due to the representativeness of errors (Reichle et al., 2002b; Mitchell et al., 2002). For suboptimal filters with non-white innovations, Mehra (1970) suggested an auto-correlation function of the innovation process to obtain asymptotically unbiased and consistent estimates of model and observation errors.

However, this method requires the number of the unknown elements in the model error to be no greater than the state vector size by observation vector size; in addition, the structure of the model error should also be predefined. This method does not guarantee the uniqueness of the estimated model error under the given conditions (Odelson et al., 2006). Thus, the representativeness of the sample correlation function is questionable, and minimum covariance may not be reached due to the possible correlation between the observations (Neethling and Young, 1974).

The basic idea behind the covariance matching method is to make the residuals consistent with their theoretical covariances (Mehra, 1972). This method is particularly popular in atmospheric data assimilations (Dee, 1995). Reichle et al. (2008) modified the covariance matching approach and used it in the surface soil moisture assimilation. In a nutshell, their modified method assumes that the internal diagnostics of the assimilation system are consistent with the values expected from the input parameters, and as a result, adaptiveness could be realized by adjusting the 'adaptive tuning factor', which is the ratio of the true input error variances and their initial values. Like most other covariance matching methods (Mehra, 1970), this method works better for the identification of observation errors than for model errors. It is also believed that covariance matching techniques give biased estimates of the true covariances (Odelson et al., 2006).

In general, adaptive update methods work well in linear systems with Gaussian errors under stationary conditions (Reichle et al., 2008). However, some of them exploit the 'optimality' of the Kalman filter, especially in the assumption of whiteness in the innovation sequences (Kitanidis and Bras, 1980b). In the case of non-stationary systems or non-white errors, computation becomes extremely tedious and complicated. Crow and Reichle (2008) compared four adaptive filtering schemes in land data assimilation and subsequently found that they all had the problem of low convergence. It has also been argued that the current self-adjustment or adaptive methods are feasible only when system dimensionality is reduced (McLaughlin, 2002). To seek an economical DA approach that requires a minimum assumption to describe the unknown model and observational error characteristics, Vrugt et al. (2005) and Crow and Yilmaz (2014) used the Auto-Tuned Land Data Assimilation System (ATLAS) for the integration of two remote-sensing soil moisture products into a water balance model. They

combined four separate adaptive filtering solutions (Triple Collocation, Innovation, Merged, and Red Modeling Error) to estimate the observation and model errors. The application of ATLAS on a simple forecast model leads to improved surface soil moisture analysis, although its power with more complex models is yet to be investigated (Crow and Yilmaz, 2014).

2.4.3 Parameter estimation

Due to current limits in the knowledge of hydrological processes, it is impossible to predetermine all the parameters in hydrological models. The most common practice first involves the initialization of the parameters, which are then adjusted with batch calibration until certain criteria are met. Once determined, the calibrated parameters are assumed to be consistent in future simulations and are rarely updated again.

The Kalman filter is designed to update the state of a system for a better initialization of the forecast. However, it is not easy to find the state vector with a significant representativeness to update the overall estimation. Moreover, a pure state update neglects the propagation of the uncertainties in model parameters (L ü et al., 2011). Thus, in order to simultaneously update the model parameters with the state, it is desirable to account for both the uncertainties in parameters as well as the state estimation (Liu and Gupta, 2007). There are three main schemes to estimate the model parameters: state augmentation, dual state-parameter estimation, and hybrid solutions.

2.4.3.1 State augmentation

State augmentation is realized by adding the parameters to the original state vector; thereafter, both the original dynamic model and observation model are modified (Hendricks Franssen and Kinzelbach, 2008). It is rare that the relationship between state variables and model parameters are linear, and therefore, the linear Kalman filter is normally inapplicable when the state vector is augmented. In the case of augmented EnKF, Hendricks Franssen and Kinzelbach (2008) argued that the state and parameters can be jointly updated with either an iterative or non-iterative approach. In the case of EKF, one can normally implement the Taylor expansion and Jacobian matrix methods as usual. However, it is worth pointing out that the parameters are usually independent from the prognostic state variables, which means that the corresponding derivatives in the Jacobian matrix are zeros. On the other hand, the prognostic state variables are not independent of the parameters.

Based on Equation (2-9) and Equation (2-10), supposing Z_k is the augmented new 'state vector' that contains both the original state vector X_k and parameters θ_k, the augmentation process can be demonstrated as follows (Gharamti and Hoteit, 2014; Wang and Wang, 1985):

$$Z_k = \begin{bmatrix} X_k \\ \theta_k \end{bmatrix} = \begin{bmatrix} M_k(X_{k-1}, \theta_{k-1}) + \eta_k \\ \theta_{k-1} \end{bmatrix} = \widetilde{M}_k(Z_k) + \tilde{\eta}_k \tag{2-13}$$

Where $\widetilde{M}_k$ is the new model operator and $\tilde{\eta}_k$ is the new model error, assuming the parameters are constants, the new model error can be expressed as:

$$\tilde{\eta}_k = \begin{bmatrix} \eta_k \\ 0 \end{bmatrix} \sim N(0, \tilde{Q}_k) \tag{2-14}$$

where

$$\tilde{Q}_k = \begin{bmatrix} Q_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{2-15}$$

where Q_k is the original model error covariance and $\mathbf{0}$ is zero matrix.

Meanwhile, the observation operator is also transformed to:

$$\widetilde{H}_k = [H_k \ \ \mathbf{0}] \tag{2-16}$$

where H_k is the original observation function.

Equations (2-14) and (2-15) are based on the assumption that the parameters are constants. In the case of varying parameters, the following equations should be used (Wang et al., 2009):

$$\tilde{\eta}_k = \begin{bmatrix} \eta_k \\ \varepsilon_k \end{bmatrix} \sim N(0, \tilde{Q}_k) \tag{2-17}$$

where ε_k is the parameter error. Assuming ε_k is independent from η_k, then

$$\tilde{Q}_k = \begin{bmatrix} Q_k & \mathbf{0} \\ \mathbf{0} & S_k \end{bmatrix} \tag{2-18}$$

where S_k is the covariance matrix of the parameter errors.

Augmentation of parameters could lead to more complex numerical solutions to the Jacobian matrix. When numerical techniques such as forward, backward or central Euler schemes are utilized (Chang and Latif, 2009), EKF might be very sensitive to the chosen step size of the parameters. EKF is also sensitive to truncation errors and round-off errors of the parameters, which therefore makes it more difficult to choose the optimal step size (Gharamti and Hoteit, 2014). Although EnKF is more capable of handling state vectors with a large degree of freedom, the increase in the degree of freedom caused by parameter augmentation, together with the spurious long-range correlations between state variables and parameters, may contribute to false parameter estimation and offset the benefits of parameter updates (Reichle and Koster, 2003; Xie and Zhang, 2013). Reichle and Koster (2003) applied localization (Keppenne and Rienecker, 2002) to constrain the correlations of the state vector elements. Xie and Zhang (2013) introduced a partitioned parameter update scheme. This new scheme is believed superior to the regular state augmentation (Xie and Zhang, 2013; Rakovec et al., 2015). Two major drawbacks arise in this method: first, the repeated running of the model not only significantly increases computational demand, but may also result in the accumulation of errors; and second, it has been found that the update order of the parameter segments may influence the assimilation results (Xie and Zhang, 2013).

2.4.3.2 Dual state-parameter filter

The dual state-parameter filter method dates back to the development of the Mutually Interactive State-Parameter (MISP) method (Todini, 1978). In this method, two Kalman Filters are run simultaneously: one for the state update and one for the model parameter update. For each time step, the two filters are run alternatively once a new observation is available: the update of state benefits from the update of the parameters through the update of the model function, and the update of the parameters benefits from the update of state from the update of the innovation (Bergman and Delleur, 1985; Wang and Wang 1985).

The equations of a dual state-parameter EKF with a nonlinear model function and a nonlinear observation function are provided, based on the following assumptions (Haykin et al., 2001):

First, the parameters and parameter-estimate errors are propagated:

$$\theta_k^f = \theta_{k-1}^a \tag{2-19}$$

$$P(\theta)_k^f = \lambda P(\theta)_{k-1}^a \tag{2-20}$$

where θ_k^f is the a priori parameter estimate, θ_{k-1}^a is the a posteriori parameter estimate, $P(\theta)_k^f$ is the a priori estimate error of the parameters, $P(\theta)_{k-1}^a$ is the a posteriori estimate error of the parameters, and $\lambda \in (0,1]$ is a 'forgetting factor' (Haykin et al., 2001, Zhong and Brown, 2009).

Then, after running the state filter:

$$x_k^f = M_k(x_{k-1}^a, u_k, \theta_k^f) \tag{2-21}$$

$$P(x)_k^f = \dot{M}_k P(x)_{k-1}^a \dot{M}_k^T + Q_k \tag{2-22}$$

$$K(x)_k = P(x)_k^f \dot{H}_k^T [\dot{H}_k P(x)_k^f \dot{H}_k^T + R_k]^{-1} \tag{2-23}$$

$$x_k^a = x_k^f + K(x)_k \cdot (y_k^o - \dot{H}_k x_k^f) \tag{2-24}$$

$$P(x)_k^a = (I - K_k \dot{H}_k) P(x)_k^f \tag{2-25}$$

where $P(x)_k^f$ is the a priori estimate error of state, $P(x)_{k-1}^a$ is the a posteriori estimate error of state, and $K(x)_k$ is the Kalman gain for the state. $\dot{M}_k$ and $\dot{H}_k$ are the linearized model and observation functions calculated with Equation (2-11) and Equation (2-12). The rest of the variables have the same meanings as in Equation (2-4) to Equation (2-8). A major difference between the dual state-parameter filters and the state-only filter is that the model function in the former is updated with the parameters.

Define:

$$v_k = (y_k^o - \dot{H}_k x_k^f) \tag{2-26}$$

$$C_k = \frac{\partial v_k}{\partial \theta}|_{\theta=\theta_k^f} = -\frac{\partial (\dot{H}_k x_k^f)}{\partial \theta}|_{\theta=\theta_k^f} = -x_k^f \cdot \frac{\partial \dot{H}_k}{\partial \theta}|_{\theta=\theta_k^f} - \dot{H}_k \cdot \frac{\partial x_k^f}{\partial \theta}|_{\theta=\theta_k^f} \tag{2-27}$$

The parameters can be updated:

$$K(\theta)_k = P(\theta)_k^f C_k^T [C_k P(\theta)_k^f C_k^T + R_k]^{-1} \tag{2-28}$$

$$\theta_k^a = \theta_k^f + K(\theta)_k \cdot v_k \tag{2-29}$$

The application of the dual state-parameter method is less common compared to state augmentation. Moradkhani et al. (2005b) described the application of dual state-parameter filters in EnKF, and in that case, the Kalman gain of the state is calculated with the cross-covariance of the state ensemble and prediction ensemble, instead of using Equation (2-23). The Kalman gain of the parameters is calculated through the cross-covariance of the parameter ensemble and prediction ensemble instead of using Equation (2-28) (Moradkhani, 2005b).

2.4.3.3 Hybrid solutions

The last category of the parameter update is named 'hybrid solutions', because it involves a third-party scheme in addition to Kalman filters. L ü et al. (2011) coupled EKF with the optimal parameter estimation, which is realized with a particle swarm optimization to obtain the dual state-parameter estimation of root zone soil moisture within the Richards equation. This coupling was found to improve the simulation results. Vrugt et al. (2005, 2006) presented the Simultaneous Optimization and Data Assimilation (SODA) method, which uses EnKF to recursively update model states conditioned on an assumed parameter set within the inner loop, while additionally estimating time-invariant values in an outer global optimization loop using the shuffled complex evolution metropolis stochastic-ensemble optimization approach. It is proved that SODA not only improves the estimates of model parameters and state variables, but also creates reliable model prediction uncertainty bounds and a time series of state and output innovation.

2.4.4 Routing problem

EKF and EnKF are designed to propagate forward, with the model's forecasted state predictions being updated at the same time observations are obtained (Crow and Ryu, 2009). In reality, the gap between the continuous hydrological states and a discrete observation usually makes this assumption hard to satisfy. For example, when an hourly rainfall-runoff model is used in DA, the discrete discharge 'observation' at the watershed outlet is not only controlled by the watershed water storage 'state' at the current hour, but also that of the last few hours considering the lag time that the water within the watershed took to travel to the outlet (Pauwels and De Lannoy, 2009). Neglecting the lag time may cause serious consequences if a distributed model is used in a large watershed and the underground water flow is taken into account. This is

an issue that all sequential methods face when they are used in hydrological DA. However, it is discussed here in the context of EnKF because of its extensive application in such problems.

Weerts and El Serafy (2006) avoided the routing problem by shifting the state forward for a period, during which time most of the rainfall leaves the watershed to reach the outlet. In other words, the future observation is 'borrowed' to be used in the current assimilation. As simple as this is, this method neglects the baseflow component in the channel; also, the determination of the 'shift time' could be subjective depending on the size of the watershed, as it is obviously different from the time of concentration. Pauwels and De Lannoy (2006, 2009) suggested a 'retrospective' EnKF in which the current state vector is augmented with a few past states. The number of augmented historical states is determined by the time of concentration. Assuming that the assimilation of each time step provides the initial condition for the next step, the EnKF is reapplied forward from the start of the concentration through the current time step. Unfortunately, Pauwels and De Lannoy's (2009) results do not seem solid enough to support the efficiency of such methods, due to the accumulation of model errors after each retrospective run. Meanwhile, in order to improve the initial condition, McMillan et al. (2013) suggested a recursive EnKF that iteratively updates past and present model states with lags up to the concentration time of the catchment, hence the streamflow forecast. This method is more stable compared to EnKF, although the computational cost could be an issue if the model is complicated and the watershed is large. The Ensemble Kalman Smoother (EnKS) uses the EnKF analysis as a first estimate, and propagates the updating backward in time, and the state covariance matrix is extended to past time steps (Dunne and Entekhabi, 2005, Evensen and Van Leeuwen 2000). Li et al. (2013, 2014) found EnKS superior to EnKF only when the soil moisture was updated without routing storage being considered. EnKS is considered superior to EnKF when they are verified in the hypothetical forecasting mode. Sakov et al. (2010) introduced Asynchronous EnKF (AEnKF), which is roughly equivalent to 4D VAR without a tangent linear or adjoint model, but does however require the forward integration of the model (Rakovec et al., 2015). AEnKF is believed to be more suitable for large-scale systems compared to EnKS because it does not require multiple ensemble updates. Rakovec et al. (2015) demonstrated the application of AEnKF by augmenting the state vector with past forecast observations, and argued that from operational aspects, AEnKF can be considered an effective method for model state updating.

It is worth pointing out that the prerequisite to use the above methods is that the routing scheme used in the model is the unit hydrograph instead of the storage routing scheme (i.e., linear reservoirs). The lag time is not an issue when the storage routing scheme is applied (Vrugt et al., 2006, Moradkhani et al., 2005a).

To conclude, the existing schemes to deal with routing problem mostly involve: 1) augmenting the state vector with historical observations; and 2) retrospective runs of EnKF. The second method may slightly violate the concept of 'filtering' , but nevertheless it seems it has been well accepted to cope with the routing issue.

2.5 Summary and Conclusions

The latest developments in Kalman-type hydrological DA are discussed in this chapter, and major attention is paid to the implementation and assessment of EKF and EnKF.

EKF features direct linearization of a system by applying a Taylor expansion and solving the Jacobian matrix. The robustness issue of EKF implementation is well discussed, and much of this concerns the divergence problem. Many factors can cause divergence in EKF: the incorrect determination of the model and observation error, improper initialization, truncation error of the Taylor expansion, and numerical calculation errors of the Jacobian matrix, among others. EnKF is a Monte Carlo method based on the Gaussian distribution assumption of the ensemble. Despite many scholars regarding EnKF as a solution to the computation issues EKF faces, the computational efficiency of EnKF itself is a primary concern. The determination of a reasonable ensemble size is regarded as a case to case problem, although more efforts are expected to balance the ensemble size and forecast accuracy, as well as error representativeness.

It is critical to determine the model and observation error in a Kalman filter. Adaptive filtering is one of the possible solutions to this issue. Four categories of adaptive filter are summarized: Bayesian, maximum likelihood, correlation, and covariance matching. Despite adaptive filters having already been applied to stationary linear systems with Gaussian errors, their applicability with regard to non-stationary systems or systems with non-white errors is still questioned. It is also believed that adaptive methods only work well when the model dimensions are significantly reduced.

Joint assimilation of state and parameters has been adopted in many hydrological DA studies. There are three major schemes to realize this: state augmentation, dual filter and hybrid methods. In the foreseeable future, it is unlikely that the joint estimation of state and parameters would completely replace batch calibration. Nevertheless, such an operation may help to extend the 'efficient period', during which the forecast can still benefit from the previous assimilation, considering that the short influence period is one of the obstacles that limits the application of Kalman filter DA in operational hydrological forecasting (Muluye, 2011; Agboma and Lye, 2014; Grassi and de Magistris, 2014).

2.6 Challenges

Despite achievements in Kalman-type hydrological DA, some challenges remain. These include, but are not limited to:

1) The quantification of model and observation errors. Even though adaptive filters may alleviate the influence of improper initial error quantifications, it is not clear as to what extent they may help. Most of the current research still tends to quantify model and observation errors using trial and error or empirical equations without further updates to adaptive filters.

2) The assimilation of model parameters through Kalman filters is still in its infancy. The 'update' of the parameters may contradict the hypothesis that the parameters are 'calibrated' constants, and there is a lack of specific comparison studies between batch calibration and Kalman filter assimilation updating for the determination of parameters.

3) The application of Kalman hydrological DA in large-scale watersheds. The main difference between large and small watersheds is the temporal mismatch of outlet streamflow observations and the updates of watershed and channel water storage. The updates of watershed and channel water storage may not be verified with the outlet streamflow observations right away, considering the significant travel time for the water to reach the outlet in a large watershed.

4) The unknown true state. Most of the EnKF applications utilize synthetic simulation results to represent the true state of the system (some even use them as quasi-observations) (Crow and Wood, 2003). In the forecasts, both EKF and EnKF applications eventually compare their assimilation results with the observations, of which the historical parts have already been assimilated. This indicates that the observations are consciously or unconsciously presumed as

identical to the true state and are free of errors (Ahsan and O'Connor, 1994). The identification of the ‘true state’ is also a key issue to advance Kalman hydrological DA from the design and experiment stages to the operational stage.

5) The handling of nonlinearity. EKF and EnKF are two of the most recognized nonlinear variants of the linear Kalman filter, yet they are both suboptimal and far from perfect. Other methods are needed to balance assimilation accuracy and computational cost as well as implementation simplicity.

6) The coupling of DA with data-driven models (Solomatine and Ostfeld, 2008). The traditional process-based rainfall runoff models place greater emphasis on model structure and parameters, and this hinders the coupling of Kalman hydrological DA at least in two ways: in the failure to create an interface to merge multiple source observations (e.g., remote-sensing products), and in the models being less sensitive to the initial condition compared to input and parameters.

Chapter 3 Streamflow Assimilation

In this chapter, EKF is coupled with a distributed process-based hydrological model – the Soil and Water Assessment Tools (SWAT) model. The streamflow observations are assimilated into SWAT to jointly update the state and parameters in order to improve streamflow simulation, and the Jacobian matrix is calculated numerically in the linearization of the SWAT model. A newly-developed ratio update strategy is used to circumvent the computational demands of EKF in a distributed model. Twin experiments are conducted to test the power of EKF, and the benefit of using EKF prior to issuing one- to seven-day streamflow *forecasts* was also assessed. Supposing a perfect knowledge of the future climate, two different schemes were used to issue the *forecasts*: EKF-subsequent open loop, and quasi-error updating.

This chapter is organized as follows:

- Section 1 introduces the application of EKF in hydrological DA.
- In Section 2, the development of the Senegal River SWAT model is introduced.
- The EKF and SWAT coupling procedures are explained in Section 3.
- The experiment setups are described in Section 4.
- The DA results, including the streamflow *forecast* verifications, are analyzed in Section 5.
- Lastly, the conclusions and recommendations are provided in Section 6.

3.1 Introduction

Data Assimilation (DA) has been widely used in the uncertainty analysis and state update of hydrological models for more than two decades (Walker and Houser, 2005a; Houser et al., 2012; Liu et al., 2012c). DA reconciles the observations and the model outputs based on real-time analysis of their errors. Although the availability of multi-source observation products boosted hydrological DA, conventional, easy-to-get observations such as streamflow data can also be assimilated into the models (Clark et al., 2008; Abaza et al., 2014a; Xie et al., 2014).

Despite the growing use of distributed models, hydrological simulation is primarily a temporal problem, and hence sequential methods are largely preferred over objective analysis in hydrological DA (Mclaughlin, 1995; Rafieeinasab et al., 2014). Sequential DA has been successfully applied to both lumped models (Abaza et al., 2014b) and distributed models (Clark et al., 2008; Salamon and Feyen, 2009; Trudel et al., 2014).

EnKF has been the preferred sequential method in most nonlinear hydrological DA. It handles large-size state vectors better than EKF, and also, no linearization of the model and observation operators is required during assimilation. However, EnKF assumes that the ensemble sample follows the same Gaussian distribution as the entire state space (Harlim and Hunt, 2007; Le Gland et al., 2009), and this may cause degeneration in EnKF if the size of the state space is much larger than the ensemble size (Papadakis et al., 2010), while increases in the ensemble size involve more computational cost.

EKF linearizes nonlinear systems by using a Taylor expansion, and applies the linearized functions in the propagation of the estimation error covariance. Due to robustness issues (Evensen, 1994) and the complexity of linearization operations in hydrological models, most of the trials with EKF hydrological DA were done with lower-dimensional lumped conceptual models with simpler model structures (Francois et al., 2003; Shamir et al., 2010; Muluye, 2011).

The Soil and Water Assessment Tool (SWAT) is a popular distributed conceptual hydrological model (Arnold et al., 1998a). Xie and Zhang (2010) argued that the application of EnKF in the SWAT model could provide acceptable model parameters and satisfactory state estimations; Han et al. (2012) coupled EnKF with SWAT to assimilate the surface soil moisture to compensate for precipitation errors; and Lei et al. (2014b) assimilated the soil moisture using the Ensemble Kalman Smoother (EnKS). However, despite these achievements with regard to EnKF or similar methods, EKF has never been used in DA with a SWAT model.

3.2 EKF and SWAT Model

3.2.1 The Extended Kalman Filter (EKF)

The Kalman Filter is an optimal estimator that recursively couples the most recent measurements into the linear model in order to update the model states (Kalman, 1960). Under the assumption that the linear system is a stochastic process with Gaussian noises, the Kalman filter produces optimal estimation with a minimum mean-square error. The Kalman filter is only applicable in linear systems where both the dynamic and observation functions are linear. EKF was developed to deal with a nonlinear system, which can be represented by the state-space functions:

$$x_k^f = M_k(x_{k-1}^a, u_k, \eta_k) \tag{3-1}$$

$$y_k^o = H_k(x_k^f, \varepsilon_k) \tag{3-2}$$

where M_k is the dynamic model function and H is the observation function. Equation (3-1) propagates the state from step *k-1* to step *k*, x_{k-1}^a is the a posteriori state vector that was updated at the last step, and x_k^f is the a priori state to be updated at current step. u_k is the 'forcing' term which is normally not updated, and η_k is the model estimation error (covariance matrix=*Q*). Equation (3-2) is the observation function that relates the state vector x_k^f to the observation y_k^o with observation error ε_k (covariance matrix=*R*).

A prerequisite to applying EKF is that the nonlinear system $M(x)$ is must be continuously derivable. Under this condition, applying the Taylor extension at the estimated point *a*:

$$M(x) = M(a) + \frac{M'(a)}{1!}(x-a) + \frac{M''(a)}{2!}(x-a)^2 + \cdots + \frac{M^{(n)}(a)}{n!}(x-a)^n + R_n(x) \tag{3-3}$$

Truncating the second- and higher-order terms as well as the error residual $R_n(x)$, Equation (3-3) becomes:

$$M(x) = M(a) + \frac{M'(a)}{1!}(x-a) \tag{3-4}$$

or

$$M(x) = M'(a)x + M(a) - M'(a)a \tag{3-5}$$

$M(x)$ now only contains the first-order term and the constants.

It is therefore possible to obtain the converted linear dynamic matrix by expanding the nonlinear functions at the estimated point. Apply this to both model function and observation function to get Equation (3-6) and Equation (3-7):

$$M'_k = \frac{\partial M_k}{\partial x_k}\Big|_{(x_{k-1}^a, u_k, 0)} \tag{3-6}$$

$$H'_k = \frac{\partial H_k}{\partial x_k}\Big|_{(x_k^f, 0)} \tag{3-7}$$

As the state is a vector with multiple variables, the linearized matrix can be expressed as a Jacobian matrix of partial differential functions as Equation (3-8):

$$J_M=\begin{bmatrix} \frac{\alpha M_1}{\alpha x_1} & \frac{\alpha M_1}{\alpha x_2} & .. & \frac{\alpha M_1}{\alpha x_n} \\ \frac{\alpha M_2}{\alpha x_1} & .. & .. & \frac{\alpha M_2}{\alpha x_n} \\ \vdots & .. & .. & \vdots \\ \frac{\alpha M_n}{\alpha x_1} & \frac{\alpha M_n}{\alpha x_2} & .. & \frac{\alpha M_n}{\alpha x_n} \end{bmatrix} \tag{3-8}$$

It is difficult to obtain an analytical solution for Equation (3-8) in most cases, and a numerical solution is often suggested.

The linearized model and observation functions are then used in the error covariance propagation, and all the rest of the steps are the same as in the implementation of the regular Kalman Filter, as follows.

In the first step of EKF, set up the initial estimation error covariance matrix P_0^a and the initial state x_0^a. The state propagation model without an explicit form of u_k is given by:

$$x_k^f = M_k' x_{k-1}^a \tag{3-9}$$

in which M_k' is the linearized model function.

The estimation error is given by:

$$P_k^f = M_k' P_{k-1}^a {M_k'}^T + Q_{k-1} \tag{3-10}$$

The superscript 'a' represents the *a posteriori* estimation, and superscript 'f' means the *a priori* estimation. Q_{k-1} is the model error covariance.

The state is then updated via:

$$x_k^a = x_k^f + K_k \cdot v_k \tag{3-11}$$

where $v_k = y_k^o - H'_k x_k^f$ is the innovation and the Kalman gain K_k is given by:

$$K_k = P_k^f H'^T_k [R_k + H'_k P_k^f H'^T_k]^{-1} \tag{3-12}$$

where H'_k is the linearized observation function and R_k is the observation error covariance.

Lastly, the estimation error is updated as:

$$P_k^a = (I - K_k H'_k) P_k^f \tag{3-13}$$

With the updated estimation error, the filter can restart from Equations (3-9) and (3-10) to begin another loop.

Equation (3-11) shows that the updated state is a linear combination of the observations and the model estimation. The gain *K* changes between **0** and *1/H'*$_k$ depending upon the uncertainties comparison between the model outputs and the observations. *K* is positively correlated to *P* and *Q*, and negatively correlated to *R*. The EKF may be applied at the latest observation iteratively as many times as needed to reduce the linearization errors (Puente and Bras, 1987).

3.2.2 The SWAT model

The SWAT model is a distributed conceptual hydrological model based on the Soil Conservation Service (SCS) runoff generation theory (Srinivasan et al., 1998; Arnold et al., 1998a; Jayakrishnan et al., 2005). The calculation unit of SWAT is the Hydrological Response Unit (HRU), which features a unique combination of three watershed characteristics: soil type, land use type, and slope.

The hydrology component of SWAT is based on the water balance equation (Arnold et al., 1998a):

$$SW_t = SW + \sum_{i=1}^{t}(R_i - Q_i - ET_i - P_i - QR_i) \tag{3-14}$$

where SW_t (mm) is the soil water content of day *t*, SW (mm) is the initial soil water content, R_i (mm) is the daily precipitation, Q_i (mm) is the daily runoff, ET_i (mm) is the daily evapo-transpiration, P_i (mm) is the daily percolation, and QR_i (mm) is the daily return flow.

The generated runoff consists of three parts: surface runoff, lateral runoff, and groundwater runoff. The surface runoff volume is calculated with the SCS curve number equation:

$$\begin{aligned} Q &= \frac{(R-0.2S)^2}{R+0.8S}, && R > 0.2S \\ Q &= 0, && R \le 0.2S \end{aligned} \tag{3-15}$$

where Q is the daily surface runoff (mm), R is the daily rainfall (mm) and S is the retention parameter determined by curve number (CN) with the SCS equation (Saleh et al., 2009):

$$S = 254(\frac{100}{CN} - 1) \tag{3-16}$$

The curve number is a function of soil permeability, land use, and antecedent soil water conditions. The curve number calculated on condition of average antecedent soil water is termed CN2. The value of CN2 normally varies between 30 and 100, and the curve numbers from dry antecedent soil water condition CN1 and moist antecedent soil water conditions CN3 are calculated based on CN2. CN2 is regarded as one of the most sensitive parameters for streamflow simulation in SWAT, and the calibration of CN2 is critical to the success of SWAT.

The newest SWAT version provides two methods to calculate the retention parameters: the traditional antecedent soil moisture content-based method, and the accumulated plant ET-based method. In this thesis, we use the traditional soil moisture content-based method in which S is tuned according to the equation (Arnold et al., 1998a):

$$S = S_1(1 - \frac{FFC}{FFC + \exp[w1 - w2 * FFC]}) \tag{3-17}$$

where S_1 is the theoretical value calculated with CN1, and FFC is the fraction of field capacity, which is a function of soil water content SW (mm) and field capacity water content FC (mm). w_1 and w_2 are shape parameters.

The actual daily curve number is then calculated with:

$$dailyCN = \frac{25400}{(S + 254)} \tag{3-18}$$

The lateral flow is calculated with Sloan's kinetic storage model (Sloan and Moore, 1984), and the groundwater flow is calculated from the variation of the shallow aquifer storage (Arnold et al., 1998a):

$$Vsa_i = Vsa_{i-1} + Rc - revap - rf - percGW - WUsa \tag{3-19}$$

where *Vsa* (mm) is the shallow aquifer storage, *Rc* (mm) is the recharge from the soil profile, *revap* (mm) is the root uptake from the shallow aquifer, *rf* (mm) is the return flow, *percGW* (mm) is the percolation to the deep aquifer, and *WUsa* (mm) is the water use from the shallow aquifer.

3.2.3 SWAT model setup

The SWAT model was applied to the upstream Senegal River watershed controlled by the Bakel station (area=420,546 km^2) (see Figure 3-1). The climate of this area is the Sahel (Sudan) type, with distinct wet and dry seasons, and the annual precipitation is mostly concentrated in the wet season (DIA, 2007).

The streamflow at Bakel Station is characterized by long dry spells with unpredictably large seasonal floods. Once flooded, the downstream floodplain may suffer from inundation for weeks to months (Fraval et al., 2002). The Manantali Dam on the Bafing tributary has reduced the peak flood discharge since 1988, while the reduced peak flow has attracted more people to the floodplain for flood-recession agriculture, and this increased the risk of hazardous floods in the future (Sandholt et al., 2003; Dumas et al., 2010).

The southeastern part of the watershed provides the bulk of the runoff in the Senegal River, yet it lacks a dense rain gauge station network. Given that streamflow observations are available at three interior stations (Gourbassi, Manantali and Oualia) on the southern tributaries, these stations are treated as point inputs to the SWAT model, and therefore their control areas, which cover the southeast and the Manatali Dam, are subtracted from the watershed.

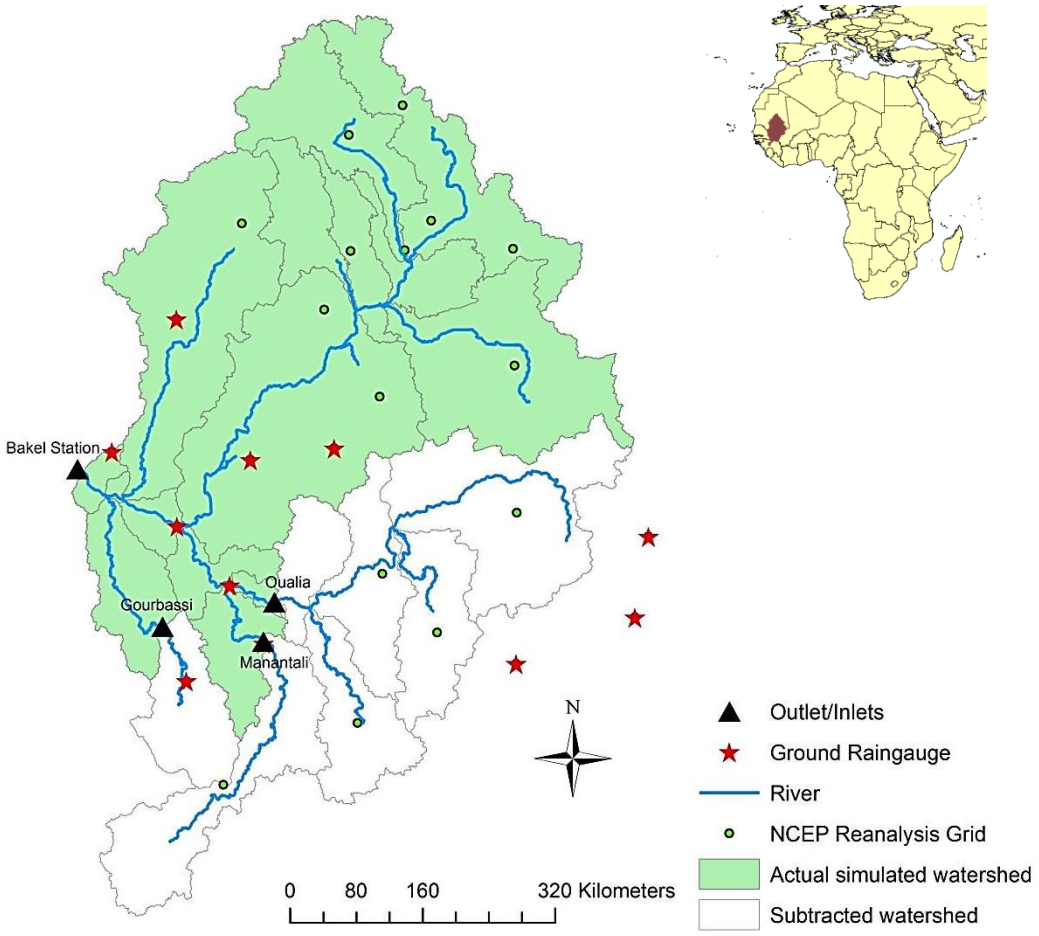


Figure 3-1 Senegal Watershed above Bakel Station

The land-use data is from the USGS Global Land Cover Characterization (GLCC) database, and the soil type data is from the FAO/UNESCO Digital Soil Map of the World. They can both be downloaded from the third-party website www.waterbase.org. The meteorological data, such as temperature, relative humidity, wind speed, and solar radiation, are obtained from the database of the National Centers for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (CFSR). Considering the importance of the precipitation data in streamflow simulations, and the well-known uncertainties in precipitation reanalysis products, we try to use the ground precipitation observations as much as possible, even though rain gauge stations are scarce and unevenly distributed. After a quality control check, only eight stations inside the

watershed were chosen. These stations are clustered in the downstream area (see Figure 3-1). To make up for the uneven distribution of the chosen rainfall stations, four additional southeast stations that are outside, but close to, the watershed are added. For those sub-basins with no ground rainfall stations, the precipitation data from the NCEP grid closest to the center-point of the sub-basin are adopted. The hybrid precipitation network is then interpolated with the Thiessen method to assign each sub-basin a representative area-weighted precipitation. These representative precipitation values are then assigned to a 'pseudo' station located at the center point of each sub-basin. These 'pseudo' stations are used as direct precipitation inputs to the SWAT model.

Table 3-1 SWAT parameter calibration results (1990-1998)

Parameter	Unit	Initial Lower Limit	Initial Upper Limit	Calibrated Lower Limit	Calibrated Upper Limit	Fitted Value
r__CN2.mgt	/	-0.3	0.3	-0.03	0.04	0
v__ALPHA_BF.gw	days	0	1	0.44	0.81	0.62
v__GW_DELAY.gw	days	20	450	381.63	532.98	457.3
v__GWQMN.gw	mm	0	500	514.29	670.91	592.6
v__GW_REVAP.gw	/	0	0.2	0.03	0.07	0.05
r__SOL_AWC(1).sol	/	-0.1	0.5	-0.02	0.29	0.14
r__SOL_K(1).sol	mm/hr	-0.2	0.8	0.16	0.36	0.26
r__SOL_BD(1).sol	g/cm^3	-0.2	0.8	0.94	1.32	1.13
r__RCHRG_DP.gw	/	-0.2	0.2	0.13	0.23	0.18
r__SOL_Z(1).sol	mm	-0.2	1	0.31	0.56	0.44
v__REVAPMN.gw	mm	0	100	7.34	49.31	28.32
v__ALPHA_BNK.rte	/	0	1	0.47	0.82	0.65

Twelve variables are selected to be calibrated based on the results of the built-in sensitivity analysis of ArcSWAT and other literature (White and Chaubey, 2005; Schuol and Abbaspour, 2006; Schuol et al., 2008; Arnold et al., 2012). The calibration benchmark is the daily streamflow observations at the Bakel Station between 1990 and 1998. After doing a baseflow separation of the streamflow observations, manual calibration was carried out to satisfy the water balance of both the surface runoff and total water generation. Then, more delicate calibrations were conducted with SWAT-CUP (Abbaspour, 2008). Manual calibration is especially useful for larger estimation errors. The calibration results are shown in Table 3-1 and

the sensitivity analyses are shown in Figure 3-2, in which the prefix ‘r’ means ‘relative’ and ‘v’ means ‘replace’; the suffix ‘.mgt’ means ‘management’, ’.gw’ means ‘groundwater’, ’.sol’ means ‘soil’, and ‘.rte’ means ‘channel parameters’. In Figure 3-1, smaller P-values or larger absolute t-values indicate more sensitivity.

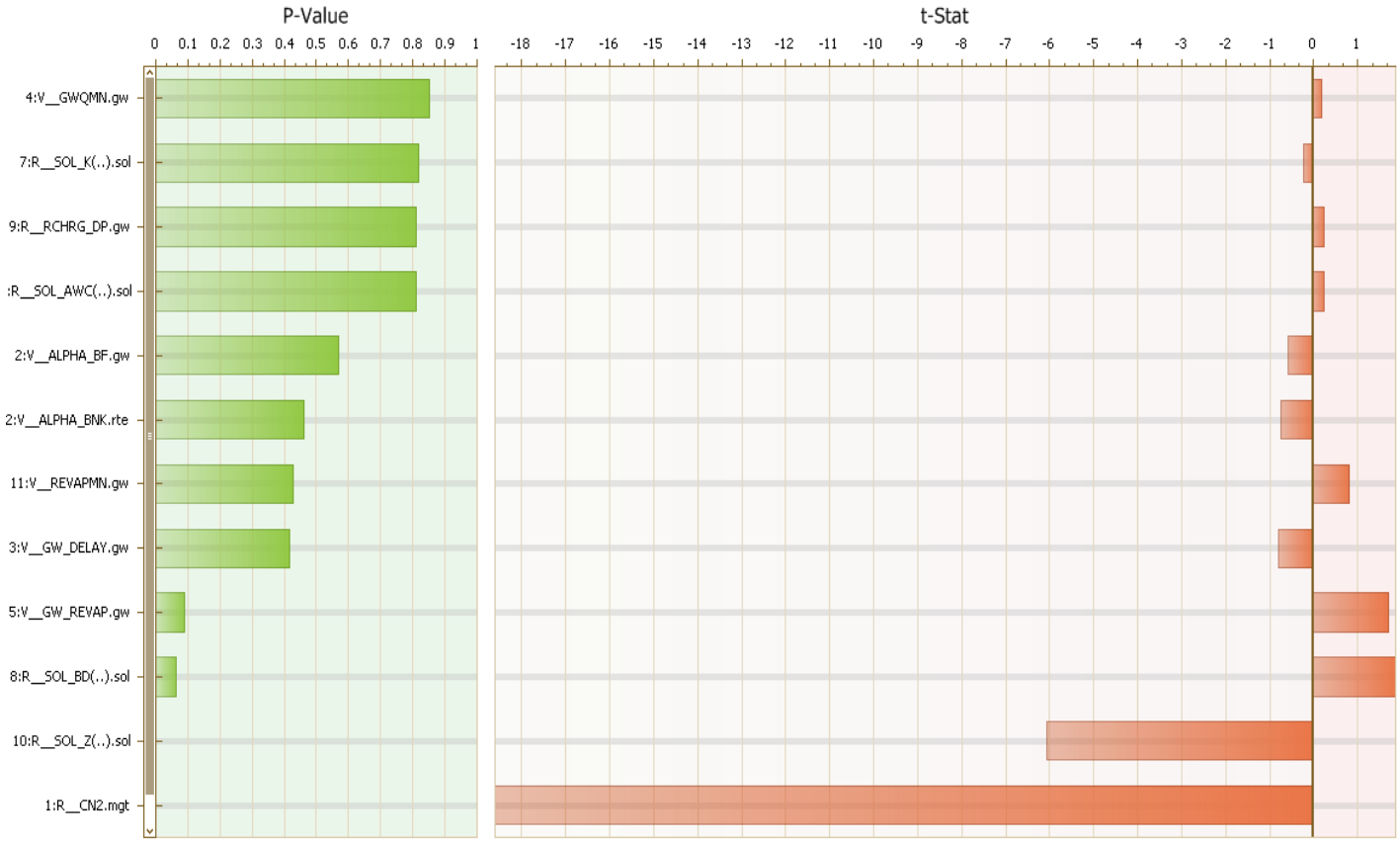


Figure 3-2 Sensitivity analysis of the calibrated parameters

3.3 EKF Coupling in SWAT

3.3.1 State selection and ratio update method

The EKF must be implemented on the basis of the state space equations, and it is common in hydrological DA to select the water storage variables in the state vector (Clark et al., 2008; Samuel et al., 2014). The SWAT model considers the runoff with three major components: surface runoff, lateral runoff and groundwater runoff. The surface runoff and lateral runoff are mainly controlled by the soil moisture content, and the groundwater runoff is controlled mainly by the shallow and deep aquifer water storage. In this study, the initial values of the shallow and deep aquifer water depths were found to be insensitive to the overall streamflow at the outlet, and hence neither of them is included in the state vector. Some studies have used water content variables such as channel water storage (Kitanidis and Bras, 1980b), snow water equivalent (Dechant and Moradkhani, 2011), and atmospheric water mass (Puente and Bras, 1987) in the streamflow assimilation. In this thesis, the atmospheric water mass was not considered, as this is a variable that SWAT could not simulate; nor was the channel water storage considered, as the reach storage correlations are too complicated in the river network. Therefore, soil moisture is selected as the only water storage variable in the state vector.

The only available observation for this research is streamflow. Although it is straightforward to build a nonlinear relationship between streamflow and watershed water content, the numerical solution to the Jacobian matrix of the observation function would be difficult to obtain. To solve this problem, we add the SWAT-simulated streamflow in the state vector, even though it is a prognostic variable (Clark et al., 2008). This way we obtain a simple linear observation operator with an element of 1 with regard to the streamflow observation, and zeros for the rest (Evensen, 1994).

In addition to state variable updates, it is possible to update the model parameters with DA (Moradkhani et al., 2005a). In EKF the parameters can be added into the state vector but treated as regular state variables in the calculations. Updating parameters could be more effective than updating state variables due to their crucial roles in conceptual models. The SWAT model has numerous parameters, among which CN2 is one of the most commonly tuned, as it controls the surface runoff generation process (Santhi et al., 2001; Zhang et al., 2009; Xie and Zhang, 2010). To maintain the stability of the EKF and reduce the calculation burden, we only update CN2 in

this thesis. Given the special form of the observation operator mentioned previously, the augmented CN2 can be treated like another regular time-variant state variable, and no extra modification to the filter is needed.

One of the issues in the implementation of EKF in distributed models is horizontal correlation of the distributed variables. This issue can be avoided in EnKF by simply stacking all the distributed variables in the state vector, but this brings heavy computational burdens in EKF. In this work, only the watershed average value of a given parameter is included in the state vector. The update at the HRU level is completed using a 'correction ratio', as:

$$ratio(i) = \frac{x_k^a(i)}{x_k^f(i)} \tag{3-20}$$

where $x_k^a(i)$ is the i^{th} element of the *a posteriori* state vector and $x_k^f(i)$ is the i^{th} element of the *a priori* state vector, both state vectors are at the watershed scale. Note that the simulated streamflow is not a distributed variable and is therefore not directly updated using this method. Although this method ignores the spatial correlation of the distributed variables, it dramatically reduces the computational burden of EKF assimilation in distributed models.

Since both the numerical solution of the Jacobian matrix and the ratio method are purely mathematical, there is a chance that the updated HRU scale state variables may violate their physical thresholds. To avoid this, both the ratios and the updated HRU scale variables are constrained within certain limits (Wang et al., 2009). The ratios are set to between 0.5 and 1.5; the soil moistures are set to between 0 and 100 mm; and the CN2 should be between 30 and 100.

3.3.2 Numerical calculation of the Jacobian matrix

The Jacobian matrix is calculated numerically using Equation (3-21). In Equation (3-21), Q(k) is the streamflow simulation, sol_sw(k) is the soil water content simulation, and CN2(k) is the curve number. Notice that in each differential equation, the denominator is the perturbation of the a posteriori estimate of the last time step, while the numerator is the change in the a prior estimate of the current time step, although Walker and Houser (2005) argued that if the a posteriori estimate is unknown, the denominators can be replaced by the perturbation of the a priori estimate. Since all the elements are partial differential equations, each column should be computed independently. As a prognostic variable, the change of the stream flow Q of the last

time step would not influence the model simulation of the current time step. Hence, for the first column, we always have:

$$\frac{\partial Q^f(k)}{\partial Q^a(k-1)} = 1, \frac{\partial sol\_sw^f(k)}{\partial Q^a(k-1)} = \frac{\partial QCN2^f(k)}{\partial Q^a(k-1)} = 0.$$

$$J(k) = \begin{bmatrix} \frac{\partial Q^f(k)}{\partial Q^a(k-1)} & \frac{\partial Q^f(k)}{\partial sol\_sw^a(k-1)} & \frac{\partial Q^f(k)}{\partial CN2^a(k-1)} \\ \frac{\partial sol\_sw^f(k)}{\partial Q^a(k-1)} & \frac{\partial sol\_sw^f(k)}{\partial sol\_sw^a(k-1)} & \frac{\partial sol\_sw^f(k)}{\partial CN2^a(k-1)} \\ \frac{\partial CN2^f(k)}{\partial Q^a(k-1)} & \frac{\partial CN2^f(k)}{\partial sol\_sw^a(k-1)} & \frac{\partial CN2^f(k)}{\partial CN2^a(k-1)} \end{bmatrix} \quad (3\text{-}21)$$

The perturbation scale is important to the calculation of the Jacobian matrix. Too small a perturbation of the denominator may either be unable to cause any change in the numerator or lead to divergence, whilst too large a perturbation may violate the local linearity assumption. Using the trial and error method, both soil moisture and CN2 are perturbed by -11% of their values.

Ideally, the perturbation scale should be time-variant. For example, a -11% change of the soil moisture may be large enough to hit the nonlinear threshold when the antecedent moisture condition (AMC) is average, while it might be too small to cause any change in the runoff generation if the AMC is too dry or too saturated. Nevertheless, as an experimental project, it is set to be a time-invariant value in this research for simplicity.

3.3.3 Determination of errors

The model errors have various sources such as input parameters, model structure, boundary and initial conditions, etc. (Liu and Gupta, 2007). Underestimation or overestimation of the model error may cause the assimilation results to overly rely on the observations or the model output (Kitanidis and Bras, 1980b). Improper quantification of the system noise could lead to unacceptable results or divergence of the Kalman filter (Puente and Bras, 1987). The observation errors are usually caused by the measurement methods, equipment and operation handling. For example, streamflow observation errors can be traced back to both the measurements of the stages and the stage-discharge curve.

In this thesis the trial and error method (Puente and Bras, 1987) is used to quantify the model error covariance matrix *Q*. First, we assume that the true state of the calendar date within a one-year period is equal to the average model outputs for the same calendar date in the previous few years. *Q* is then calculated with the model error time series for each variable in the state vector, where the model error is defined as the difference between the model output state and the true state calculated with the model outputs between 1990 and 1999, as shown in Equation (3-22):

$$Q = E((\boldsymbol{x} - \bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}})^T) \tag{3-22}$$

Where $\boldsymbol{x}$ is the state vector, $\bar{\boldsymbol{x}}$ is the true state vector calculated with the average model outputs from 1990 to 1999 for the same calendar date as $\boldsymbol{x}$. Notice equation (3-22) does not apply to the calculation of the error covariance between CN2 and other state variables as well as its own variance, because CN2 is regarded as a constant in SWAT model. In this case the CN2 related covariances and variance are set manually.

The previous two steps are only used to determine the initial scales of *Q*. In order to obtain an optimal *Q* one has to adjust the orders of the elements of *Q* based on the predefined performance criteria. In this thesis the optimal *Q* is obtained by observing the *forecast* improvements in the EKF-subsequent open loop runs.

The observation error covariance matrix *R* is the stream flow variance simply because it is the only variable in the observation vector. Although *R* can also be calculated via trial and error, here we use an empirical method that assumes that the observation error is proportional to the scale of the streamflow observation (Clark et al., 2008), hence:

$$R_k = (\alpha y_k)^2 \tag{3-23}$$

where y_k is the observed streamflow, and α is an arbitrary scaling factor. In this study, we set $\alpha = 0.1$.

3.3.4 Implementation procedures

The flowchart of the implementation procedures of SWAT coupled with EKF is shown in Figure 3-2. The Fortran version of SWAT 2009 is transferred into a Matlab environment. The Matlab version of SWAT has two major advantages over the Fortran version: it is more convenient to handle the modification of the state by saving all the variables in a temporary

database, and the calculation of the Jacobian matrix is easier. As shown in Figure 3-3, the first-day simulation is used as the warm-up run. Starting from the second day, EKF is triggered as long as there are observations. During EKF, we first run a regular 'simulate' to generate the *a priori* estimate. The SWAT model is linearized in parallel with the generation of the background field, and with the streamflow observation and the linearized SWAT, the conventional Kalman filter is executed to obtain the a posteriori estimate of state. Then, the watershed scale ratios are calculated to update the HRU scale states and parameters. An entire loop is completed at this point, and another daily simulation of SWAT would start the next loop and repeat this process over and over again.

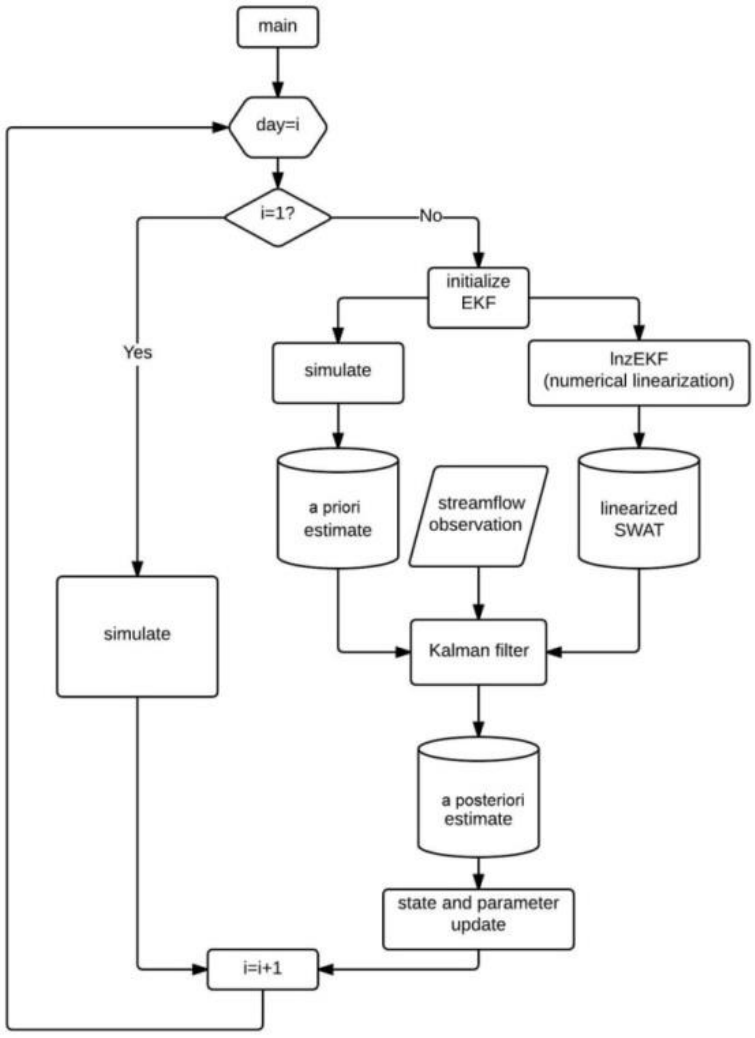


Figure 3-3 The EKF coupled SWAT implementation flowchart

3.4 Experiment Setup

3.4.1 Twin experiments

As a case study, the twin experiments were carried out from January 1st, 1999 to December 30th, 1999. The year 1999 is chosen because of the exceptional precipitation in the wet season of this year that caused severe flooding at the Bakel Station (Dia, 2007), and SWAT was found unable to simulate the flooding despite that it generally performed well in the previous calibration period.

To set up the twin experiments, we first run the SWAT model without EKF assimilation in order to provide a reference 'open loop', and then rerun the model using the same setup as the open loop but assimilate the observed daily streamflow with EKF. The behaviours of EKF are analyzed by comparing the state and parameters in the second run with those from the open loop run.

The twin experiments are different from the widely used synthetic experiments in EnKF assimilations (Liu et al., 2012c). The assimilation run in EnKF synthetic experiments involves artificial perturbations on the observations, model inputs and states in the ensemble simulations. The true state is represented either by an unperturbed model realization (Crow and Ryu, 2009b) or a random perturbed ensemble member (Kumar et al., 2009). By contrast, the assimilation run in the EKF twin experiments is a single realization of SWAT without any perturbation. More importantly, no 'true' state is predefined because it is never known from the operational perspective unless the model and observation errors are perfectly quantified.

The Kalman Filter degenerates to a pure prediction problem if the observation time series and 'true' signals are identical (i.e., the true state is measurable). In such cases, the application of the Kalman filter is redundant (Ahsan and Oconnor, 1994). The comparisons between the 'open loop' results and the EKF assimilated results only reveal the behaviors or effectiveness of EKF, which is the basis to evaluate its potentials in improving the *forecast*. Nevertheless, for design purposes the EKF estimates may serve as the pseudo true state in the verification of the *forecast*, based on the assumption that the model and observation errors are properly determined.

3.4.2 Forecast verification

One of the ultimate goals of DA is to provide optimized initial conditions for the forecasts. Technically, the *a priori* estimation of EKF could be regarded as a one-day-ahead forecast. It is reasonable to verify whether the *forecast* with a longer lead time is improved or not after using EKF. The influence of DA on the streamflow forecast depends on the specific model, catchment, DA methodology, etc. (Komma et al., 2008; Randrianasolo et al., 2014).

It is worth pointing out that the '*forecasts*' referred to in this study are 'perfect forecasts'. They are calculated based on the assumption that the model inputs (such as weather information) can be predefined or perfectly known. Biased inputs would cause extra errors in the *forecast*; however, this issue is not a focus of this thesis.

Here we tested *forecasts* with a lead time of one to seven days, and used two schemes to evaluate the influence of the EKF on the streamflow *forecast*. The first scheme is called 'EKF-subsequent Open Loop' (EKFsOL). As the name indicates, this scheme runs open loop simulations online for seven days after the EKF. Figure 3-4 demonstrates the algorithm for the EKFsOL. Each color represents a calculation loop, and the arrow in each box shows the computation flow direction: Day0 ~ Day6 are sequential, and after Day7 the computation will return back to the EKF. Day0 is the sequential EKF DA series running on a daily basis without lead time (although technically the a priori estimate can be regarded as a one-day *forecast*). The EKF flow is paused after the current day EKF is completed, the model status is saved in a temporary state database, the open loop SWAT is then triggered for the next seven days, and the *forecast* result for each lead day is stored in column Day1 ~ Day7. At the end of the seventh-day *forecast*, Day7, the model is switched back to the EKF flow by restoring the model back to the temporary state database saved earlier. Repeating this process, we get the complete time series of the *forecast* for each lead day.

Date	Day0	Day1	Day2	Day3	Day4	Day5	Day6	Day7
27/07/1999	EKF↘							
28/07/1999	EKF↘	OL↘						
29/07/1999	EKF↘	OL↘	OL↘					
30/07/1999	EKF↘	OL↘	OL↘	OL↘				
31/07/1999	EKF↘	OL↘	OL↘	OL↘	OL↘			
01/08/1999	EKF↘	OL↘	OL↘	OL↘	OL↘	OL↘		

02/08/1999	EKF↘	OL↘	OL↘	OL↘	OL↘	OL↘	OL↘	
03/08/1999	EKF↘	OL↘	OL↘	OL↘	OL↘	OL↘	OL↘	OL↖
04/08/1999	EKF↘	OL↘	OL↘	OL↘	OL↘	OL↘	OL↘	OL↖
……	……	……	……	……	……	……	……	……

Figure 3-4 Demonstration of the EKFsOL forecast evaluation scheme

Besides the straightforward EKFsOL, the t days lead streamflow *forecast* results are also estimated with a newly defined variable:

$$Q\_fct(iie+t) = Q\_ol(iie+t) + K(iie,1) * (obs(iie) - Q\_ol(iie+t)) \,, t = 1,2,..,7 \quad (3\text{-}24)$$

where iie is the day EKF is implemented, $Q\_fct(iie+t)$ is the t days lead-time streamflow *forecast* after day iie, $Q\_ol(iie+t)$ is the open loop result, $K(iie,1)$ is the streamflow Kalman gain for day iie, and $obs(iie)$ is the streamflow observation for day iie. Equation (3-24) is a modified Kalman state update equation, with the major difference being that in Equation (3-24) the observation and Kalman gain are fixed to the value of day iie because the future observations and Kalman gains are unknown. Through such modifications, this new variable takes full advantage of the information in the latest Kalman gain, observations and future model outputs.

Equation (3-24) is regarded as a 'quasi' error update method. Like the regular error update methods, it only uses model outputs and historical observations without interfering with the model inputs and states. It is 'quasi' because in a standard error update scheme the errors (the difference between the observations and the model output for the same day) are usually estimated independently rather than relying on external parameters such as the Kalman gain in Equation (3-24).

3.4.3 Performance criteria

Three statistical indicators are used to assess the streamflow *forecast* accuracy, as follows:

Nash–Sutcliffe coefficient:

$$NSC = 1 - \frac{\sum_{i=1}^{N}(Q_i^o - Q_i^s)^2}{\sum_{i=1}^{N}(Q_i^o - \overline{Q^o})^2} \quad (3\text{-}25)$$

Percent bias:

$$PBIAS = 100 * \frac{\sum_{i=1}^{N}(Q_i^o - Q_i^s)}{\sum_{i=1}^{N} Q_i^o} \quad (3\text{-}26)$$

Root mean square error (m^3/s):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Q_i^o - Q_i^s)^2}{N}} \quad (3\text{-}27)$$

where N is the time series length (days), Q_i^o (m^3/s) is the true streamflow of the i^{th} day, $\overline{Q^o}$(m^3/s) is the mean of the true streamflow series, and Q_i^s(m^3/s) is the *forecast* streamflow for the i^{th} day.

3.5 Results and Discussion

3.5.1 State and parameter updates

The results of the twin experiments are critical to assess the effectiveness of EKF. The comparison between the EKF a posteriori streamflow estimate and the open loop simulation at Bakel Station for 1999 is shown in Figure 3-5. The SWAT model constantly overestimates the dry season streamflow from January to July, and significantly underestimates the wet season flood from August to November. Table 3-2 summarizes the statistical indicators of the open loop (OL) and EKF streamflow estimates for the full year and part of the wet season that covers the majority of the flood hydrograph. Note that for pure comparison purposes, Q_i^o in the evaluation equations is replaced with the streamflow observations. The NSC for the open loop estimate in the chosen wet period is only 0.1.

The assimilated streamflow is very similar to the observations. The grey dotted line in Figure 3-5 shows the value of the Kalman gain K for the streamflow. During the dry season K almost equals 1.0, which means that the model output is only slightly taken into consideration by the EKF DA. This should not be surprising given that the majority of the dry season streamflow is baseflow from ground water runoff; the state vector only contains soil moisture content and CN2, which are more critical to the surface and lateral runoff. For this reason, in the next few sections we will only assess the streamflow *forecasts* for late July to late September.

The watershed soil moisture assimilation results are shown in Figure 3-6, and the watershed precipitation is also displayed in Figure 3-6. Compared to streamflow, watershed soil moisture

is very sensitive to precipitation. The precipitation hence has great potential in the validation of the soil moisture estimate, considering the difficulties of directly measuring the watershed soil moisture. Similar to the case with streamflow, the influence of the assimilation on the watershed soil moisture is significant in the wet season but not in the dry season. This is likely because the surface runoff occurrence threshold is never reached in the dry season, and therefore the streamflow does not respond to changes in soil moisture content.

Correspondingly, the soil moisture will not be updated from the new information from streamflow observations. For the wet season, EKF first moderately increases the soil moisture during the rising flood period but then significantly reduces it in the recession period, despite the fact that the streamflow increases throughout the whole flood period. Such a conflict is not uncommon in multiple-variable hydrological DA (Samuel et al., 2014; Trudel et al., 2014). Samuel et al. (2014) found that assimilating streamflow alone would distort the soil moisture estimate, and vice versa. A possible solution for this issue is the combined assimilation of both variables.

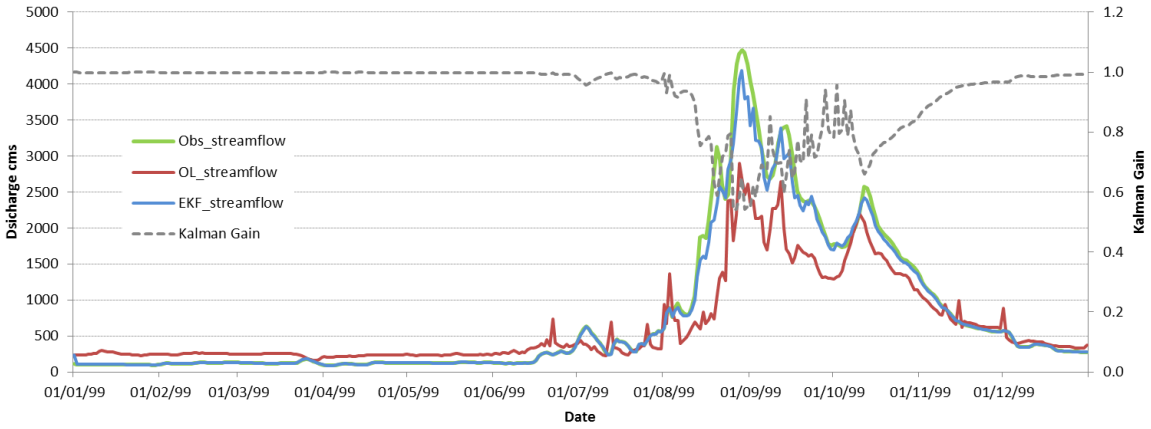


Figure 3-5 Bakel Station 1999 streamflow EKF assimilation results compared to observations (Obs) and Open Loop (OL)

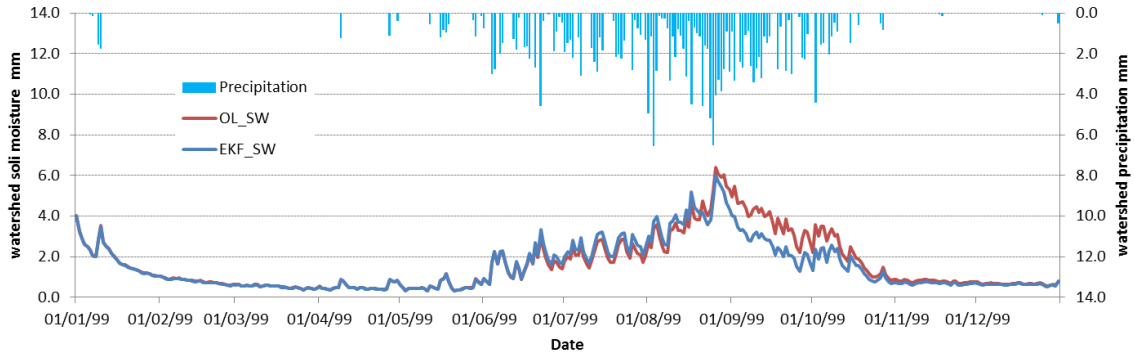


Figure 3-6 Watershed average soil moisture (mm) EKF assimilation results compared to Open Loop (OL)

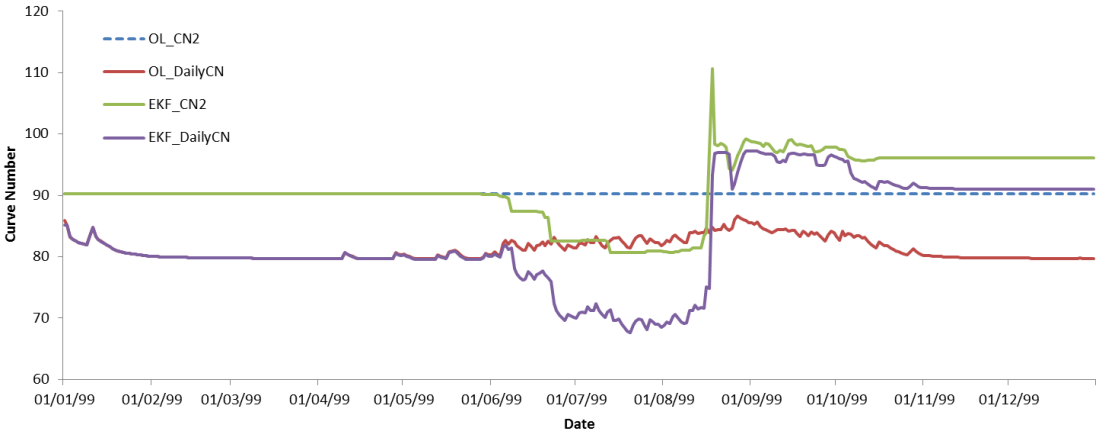


Figure 3-7 Watershed average soil CN2 and dailyCN assimilation results compared to Open Loop (OL)

Figure 3-7 displays the updates of CN2 and dailyCN, which is the curve number that is adjusted with the soil moisture content based on CN2. The dailyCN is the curve number directly used in the calculation of the retention parameters. In the open loop model, the model parameters are presumed as constants in future simulations once they have been calibrated. In the case of the SWAT model, unless CN2 is redefined in the crop operation files, it remains constant, and dailyCN only varies with the soil moisture content. The validation of the traditional batch calibration is based on the assumption that the hydrological system is a steady system without abrupt changes. However, as shown in Figure 3-5, the calibrated model does not perform well for exceptional or extreme flood simulations.

Again, the CN2 and dailyCN are also only updated in the wet season. Both of them steadily decline from early June to early August. Note that in the same period the streamflow Kalman

gain remains close to 1.0 and the soil moisture is slightly higher. There is then a sharp increase, and further increases until early October. Then there is a slight drop to a steady but high level. In the same period, the streamflow Kalman gain fluctuates dramatically and the soil moisture sees a major reduction, as shown in Figure 3-6. The CN2 and dailyCN share a similar update pattern, but the latter has a larger update magnitude, as it also reflects the update of the soil moisture despite that there is an obvious spike in the increase of CN2. This spike sets the value of CN2 to 111, but it immediately falls back to below 100. It is also noted that the reduced soil moisture between early August and mid-October shrinks the difference between the CN2 and dailyCN.

It is worth pointing out that the update patterns of CN2 and dailyCN vary with the quantification of Q and R, although the details are not shown here. Furthermore, both of them reach a steady level after the flood season. However, it is too ambitious to extrapolate the new steady level as a long term 'constant', as the updated parameters are the result of a strict filtering procedure without a long-range *forecast* capability.

Figure 3-8 displays the watershed soil moisture and CN2 update ratios from July 27th to September 30th, 1999. The change in the ratios correctly reflects the soil moisture and CN2 assimilation results analyzed earlier. As shown in Figure 3-8, it is obvious that the soil moisture and CN2 have opposite update patterns.

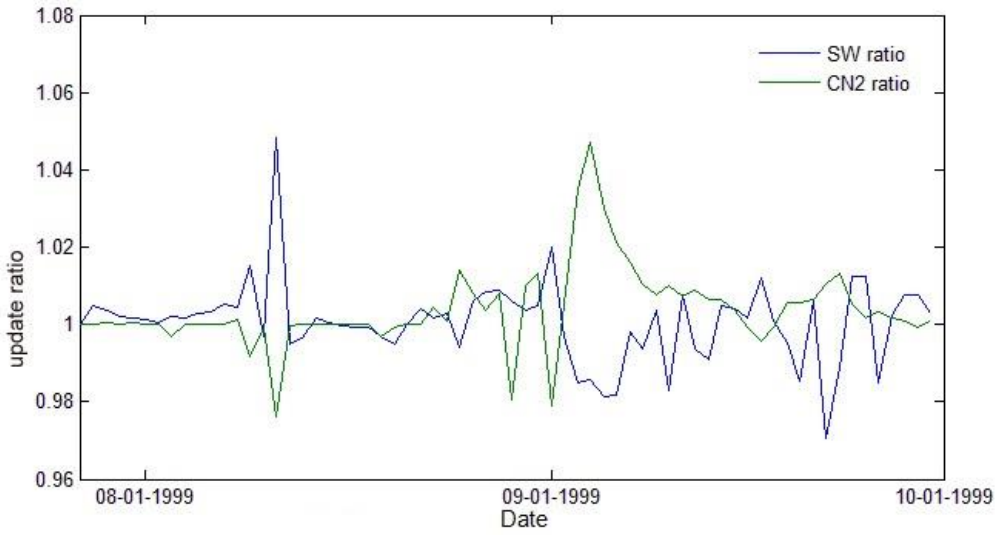


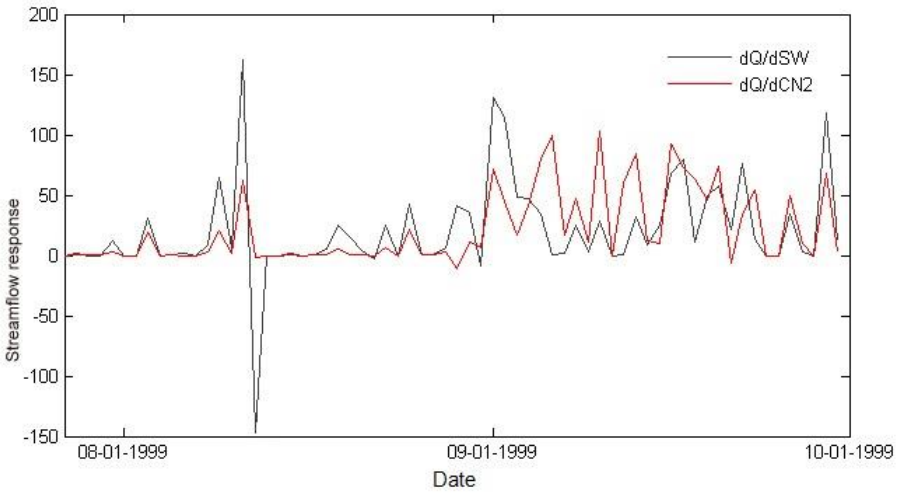
Figure 3-8 The soil moisture update ratio (SW-ratio) and the CN2 update ratio (CN2-ratio)

Table 3-2 Statistical indicators of Open Loop (OL) and EKF streamflow estimates

	January 1-December 30		July 27-September 30	
	OL	EKF	OL	EKF
NSC	0.77	0.99	0.10	0.93
PBIAS	16.59	4.34	37.95	7.78
RMSE/mm	484.99	121.35	1078.75	298.16

Although we constrained the magnitude of the ratios to between 0.5 and 1.5, it seems these boundaries have rarely been reached.

(a)



(b)

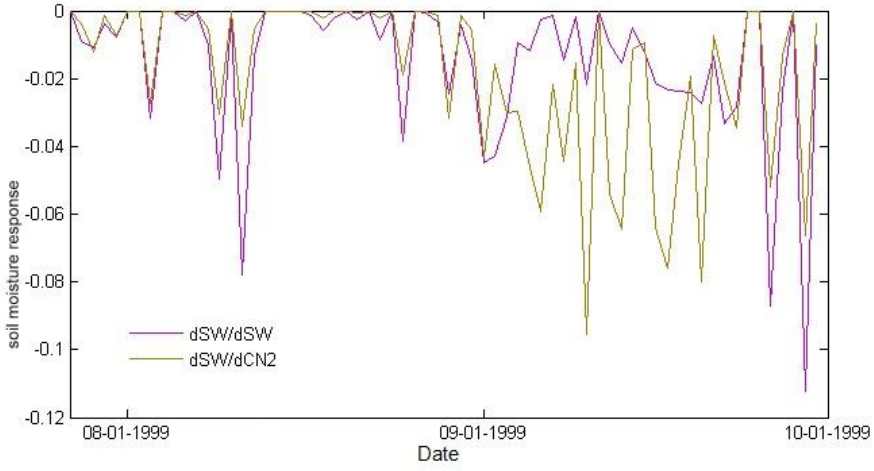


Figure 3-9 (a) Streamflow response to the antecedent soil moisture and CN2, and (b) soil moisture response to the antecedent soil moisture and CN2.

Since we presume that the model error and observation error are constant throughout the simulation, the estimate error is updated mainly due to the propagation of the model function. In this case, it is the linearized SWAT model in the form of a numerical Jacobian matrix with a perturbation magnitude of -11%. The proper calculation of the Jacobian matrix depends on the sensitivity of the state variables to the perturbations. Figure 3-9 (a) shows the streamflow response to the perturbation of the antecedent soil moisture and CN2, and Figure 3-9 (b) shows the soil moisture response to the perturbation of the antecedent soil moisture and CN2. Both are only carried out from July 27th to September 30th. In general, the streamflow shows a positive feedback while the soil moisture shows negative feedback to the negative soil moisture (SW) and CN2 perturbations. The opposite responses of streamflow and soil moisture may explain why the streamflow is increased while the soil moisture is reduced in the flood recession period.

3.5.2 Forecast verification

Table 3-3 provides the statistical indicators for the open loop (OL), EKF (Day0) and EKFsOL *forecasts* for the one- to seven-day leads for the period July 27th to September 30th. Figure 3-10 shows the EKFsOL *forecasts* for (a) one day, (b) two days, (c) three days, and (d) seven days. The bars in Figure 3-10 represent the differences between the *forecast* and the true state, which is replaced with the EKF assimilation results from the previous section. The indicators show that the EKFsOL *forecast* carries on the benefit of EKF for up to 7 days. The *forecast* accuracy decreases with the increase of the lead time. The average decrease rate for NSC is 0.03/day, and the average increase rate for PBIAS and RMSE is 0.54/day and 24.36/day respectively.

As shown in Figure 3-10, a prominent feature of the EKFsOL *forecasts* is the dramatic fluctuation. They frequently swing between the observations and the open loop simulation. Despite occasionally surpassing the observations, the EKFsOL results fall mostly between the observations and open loop simulation. The main difference between the different lead *forecasts* is the ability to capture the peak flow at the end of August. While Day1 and Day2 moderately improve the underestimation of the peak flow compared to open loop, Day7 worsens the underestimation. In Figure 3-10 (a) it is also noted that EKFsOL performs better for the flood events in the rising period than in the recession period, and similar features can also be observed in the pure EKF assimilation results. Such a performance discrepancy among different flood

phases was also observed by Komma et al. (2008). Although it is not clear what causes such a discrepancy, a more accurate flood rise *forecast* is more valuable than a flood recession *forecast*.

Table 3-4 shows the statistical indicators for the quasi-error update forecasts for one- to seven-day leads. Figure 3-11 displays the quasi-error forecast for the lead of one day (a), two days (b), three days (c) and seven days (d). Compared to EKFsOL, this method provides much better indicators. However, with the increase in lead time, the NSC decreases at a rate of 0.06/day, the PBIAS increases at 1.08/day, and the RMSE increases at 69.67 m^3/s /day. These dramatic deterioration rates indicate that this method is less robust compared to the EKFsOL forecast.

The effectiveness of the quasi-error update forecast largely depends on the Kalman gain of streamflow, which partitions the forecast into future model outputs and historical observations. If K equals 1.0, the forecast completely ignores the model output and shifts the historical observations forward. The closer K approaches to 1.0, the more such a 'shift forward' can be observed. The lower K values around the peak flow in the assimilation mode reduce the peak flow and the surrounding high flows when they are shifted forward in the forecast mode. For this reason, the quasi-error method shows better performance for the recession period than for the rising period of the floods.

Figure 3-12 and Figure 3-13 compare the mean, interquartile range, spread and outliers of the forecast errors for these two schemes. As shown in Figure 3-13, the quasi-error update scheme has very good performance for shorter lead times; however, the errors spread faster than the EKFsOL and have an obvious decreasing trend. It is reasonable to speculate that EKFsOL is more stable and that with longer forecasts it may outperform the quasi-error update forecasts.

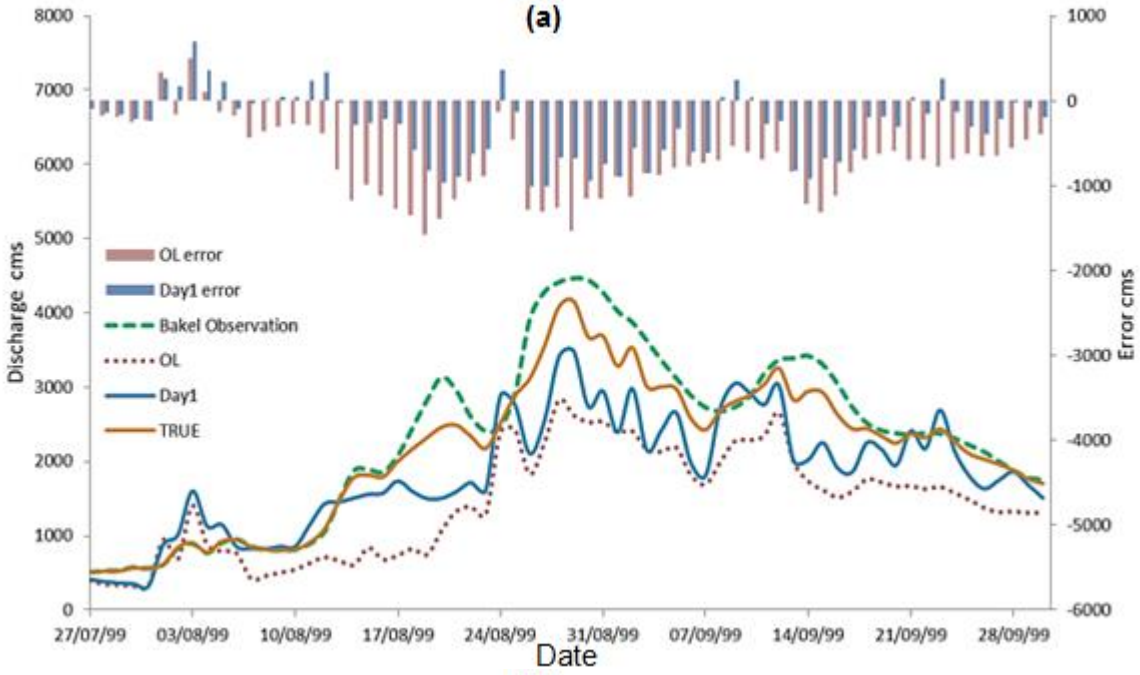
Table 3-3 Statistical indicators for OL and EKFsOL forecasts for one- to seven-day lead times

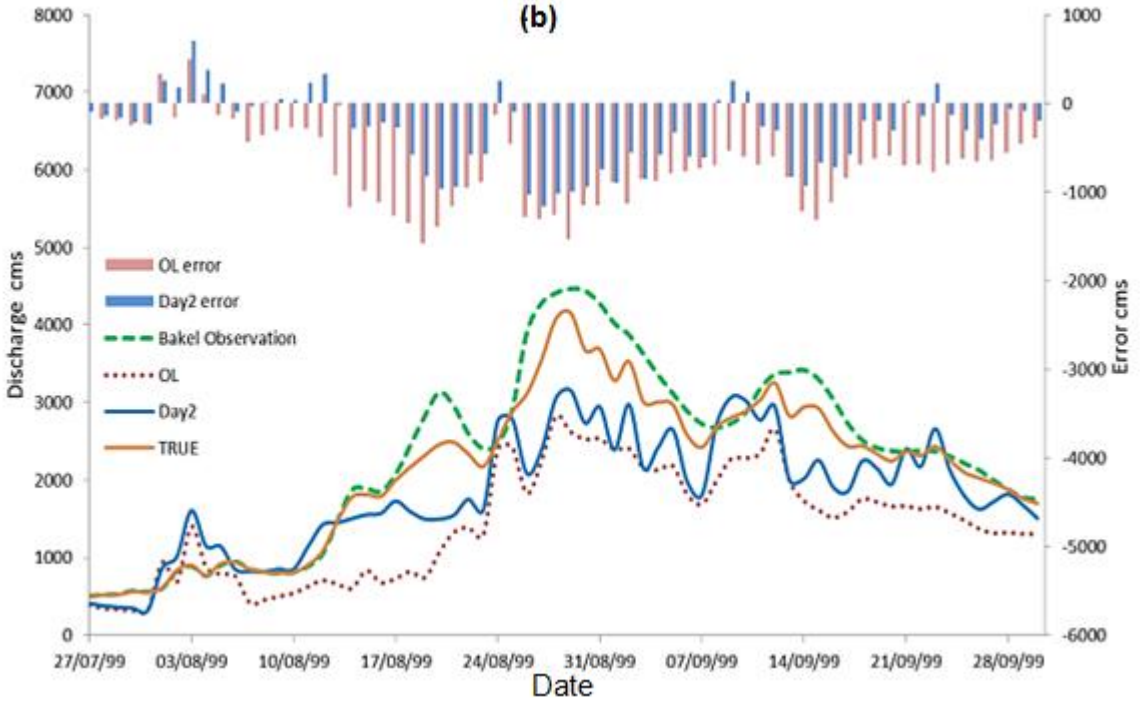
	OL	DAY1	DAY2	DAY3	DAY4	DAY5	DAY6	DAY7
NSC	0.28	0.74	0.72	0.68	0.63	0.59	0.57	0.53
PBIAS	32.71	13.56	14.27	14.98	15.59	16.26	16.80	17.35
RMSE/mm	824.31	491.65	515.47	552.60	590.56	619.63	636.57	662.16

Table 3-4 Statistic indicators for quasi-error forecasts for one- to seven-day lead times

	DAY1	DAY2	DAY3	DAY4	DAY5	DAY6	DAY7
NSC	0.98	0.93	0.85	0.76	0.70	0.64	0.58
PBIAS	0.57	1.66	2.97	4.16	5.35	6.71	8.15

RMSE/mm	143.04	264.56	380.91	471.66	535.55	585.41	630.71





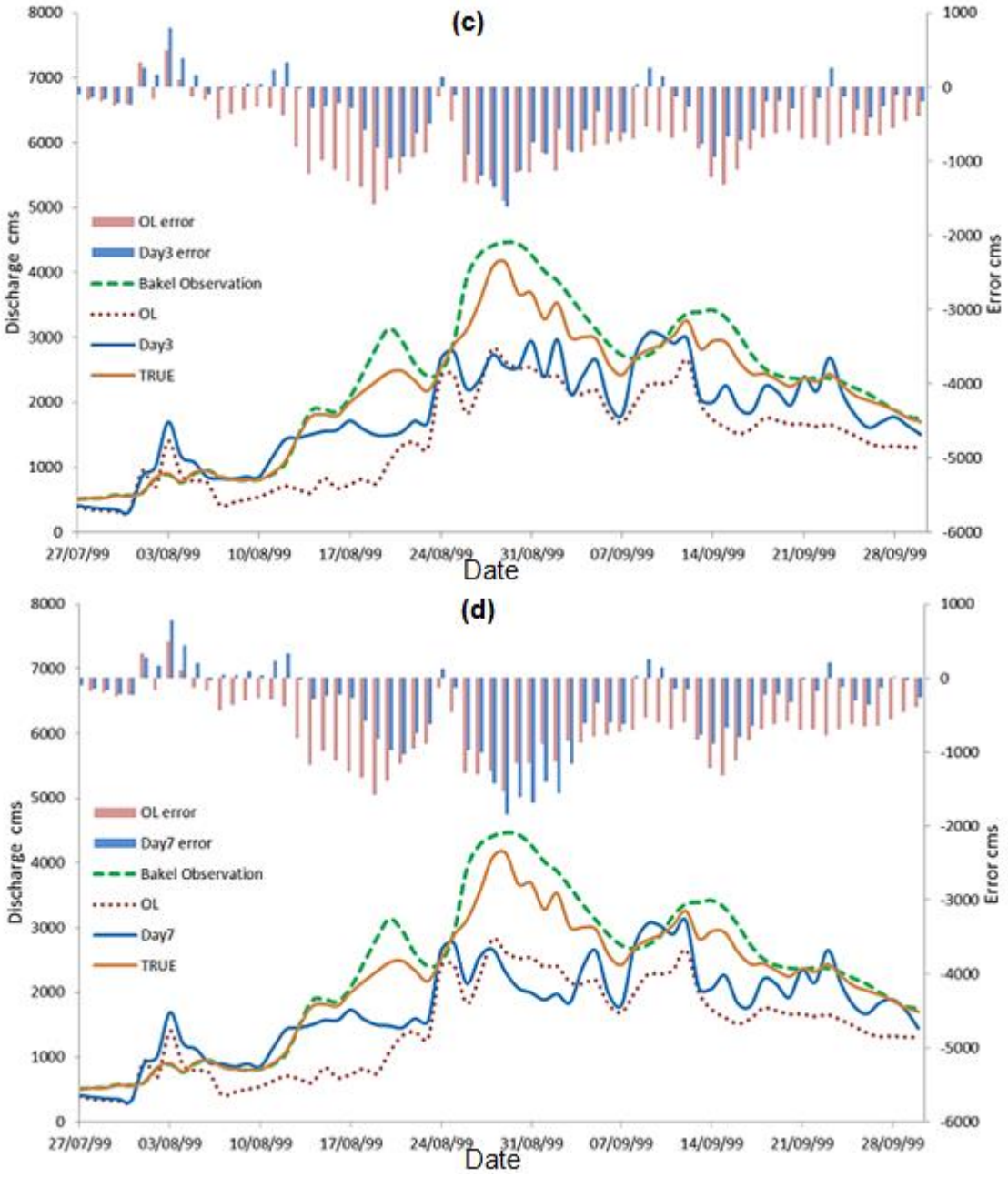
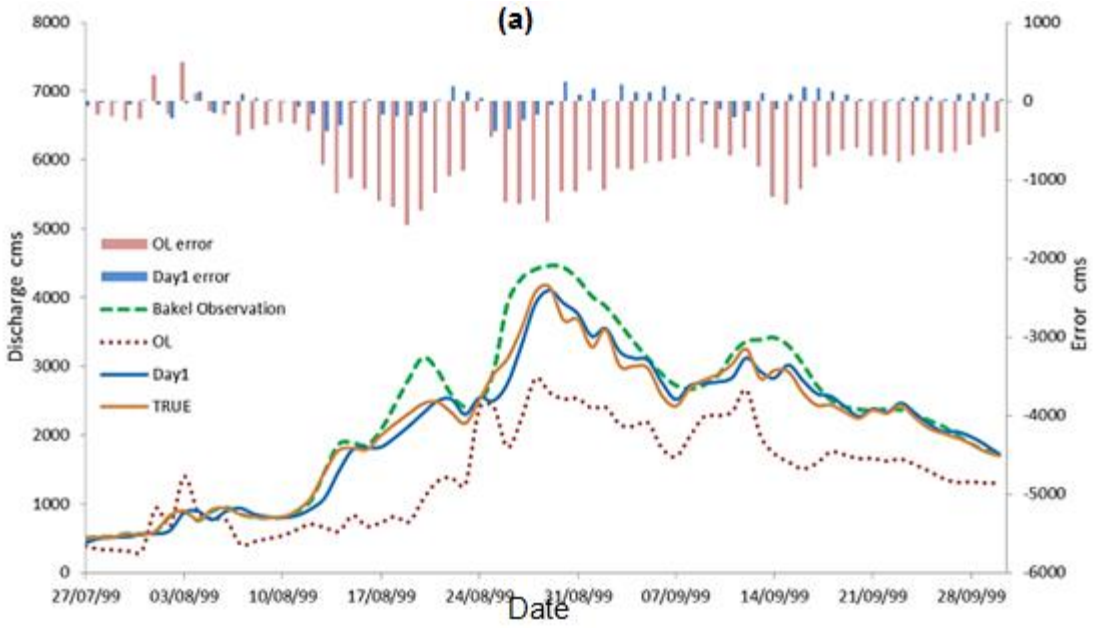
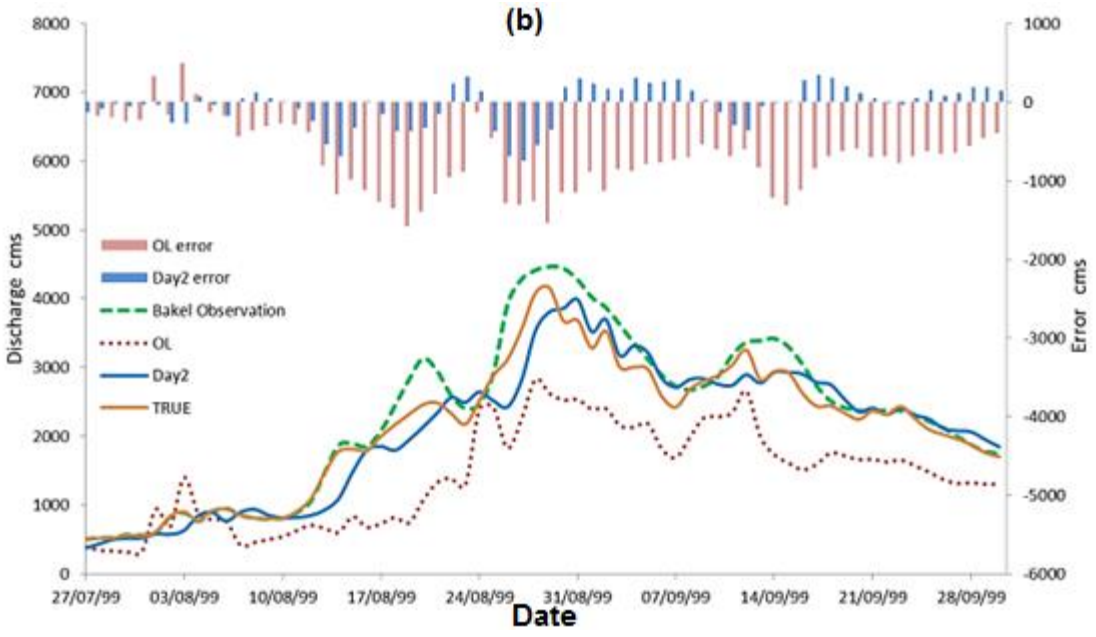


Figure 3-10 **The EKFsOL forecast for Day1 (a), Day2 (b), Day3 (c) and Day7 (d)**


(a)
Discharge cms
Error cms
OL error
Day1 error
Bakel Observation
OL
Day1
TRUE
Date
27/07/99
03/08/99
10/08/99
17/08/99
24/08/99
31/08/99
07/09/99
14/09/99
21/09/99
28/09/99


(b)
Discharge cms
Error cms
OL error
Day2 error
Bakel Observation
OL
Day2
TRUE
Date
27/07/99
03/08/99
10/08/99
17/08/99
24/08/99
31/08/99
07/09/99
14/09/99
21/09/99
28/09/99

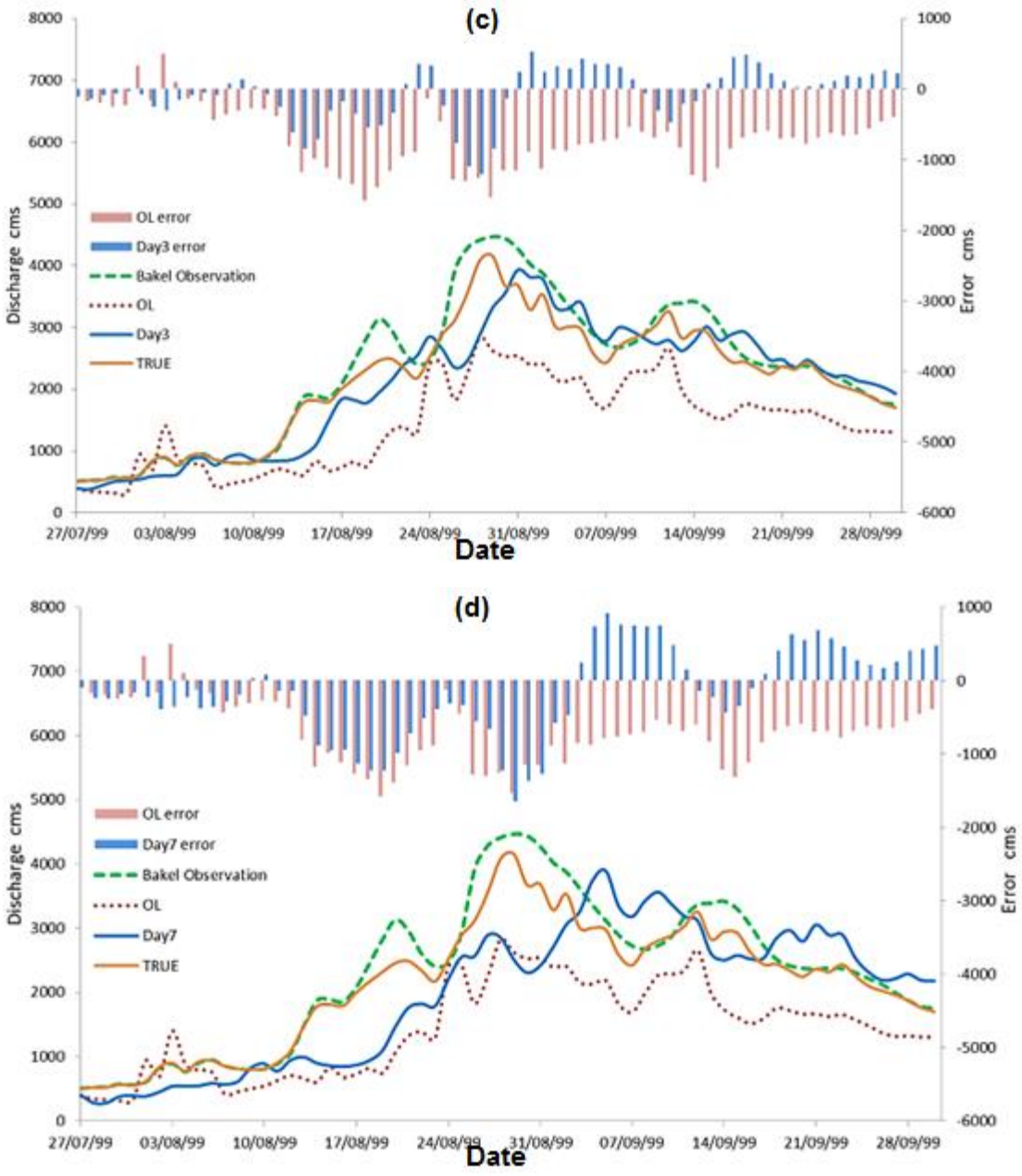


Figure 3-11 The quasi-error forecasts for Day1 (a), Day2 (b), Day3 (c) and Day7 (d)

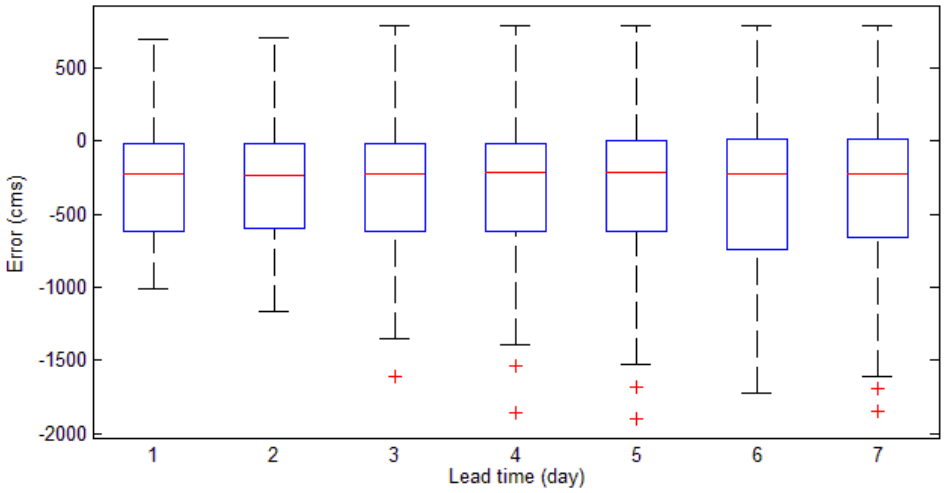


Figure 3-12 Boxplot of the EKFsOL forecast errors for different lead times

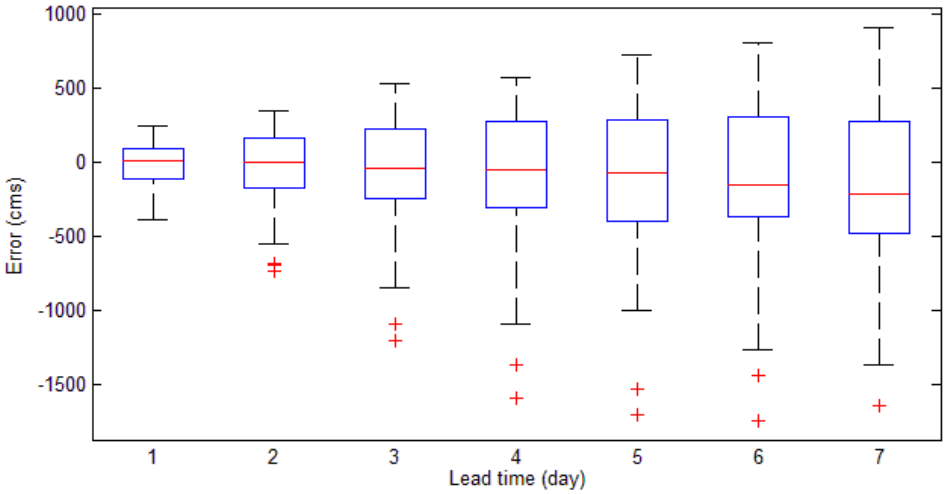


Figure 3-13 Boxplot of the quasi-error update forecast errors for different lead times

3.6 Conclusions and Recommendations

The EKF was coupled with SWAT by assimilating the daily streamflow observations at the outlet of a large-scale watershed. Twin experiments were adopted to assess the effectiveness of EKF assimilation by evaluating the state and parameter updates, and more importantly, the possible improvement in the streamflow forecasting. The forecasts are all 'perfect forecasts' based on the assumption of perfect knowledge of the meteorological forecast.

The SWAT model is translated from Fortran to Matlab to allow temporary state matrix storage, and the streamflow output of SWAT is included in the state vector to avoid linearization of the observation equation. SWAT is treated as a 'black-box model' in the numerical solution of the Jacobian matrix, in which each element is calculated with a partial differential equation.

In order to control the size of the state vector, only the watershed scale state variables are assimilated directly. The HRU scale variables are then updated with the *a posteriori/a priori* correction ratio of their watershed scale counterparts. This method balances the computational burden and the update requirements of the distributed variables, despite the fact that their horizontal correlations are neglected.

EKF has significantly updated both the soil moisture and CN2 in the wet season, but not in the dry season. In the wet season the soil moisture first increases and then declines, while the CN2 first declines and then increases. Such opposition is also found between the streamflow response and soil moisture response to the same perturbation on the antecedent soil moisture and CN2 in the calculation of the Jacobian matrix for model linearization.

In a case study of a high flood forecast, the effectiveness of EKF is verified with both the EKFsOL and quasi-error update forecast schemes for up to seven days. The quasi-error update forecast generally outperforms EKFsOL, but has less robustness. The longer the lead time, the less capable EKFsOL is of capturing the peak flow. EKFsOL performs better in the rising period of the flood events than in the recession period, while the quasi-error forecast does the opposite.

In conclusion, EKF is a promising assimilation method for DA with the SWAT model. This thesis only assimilates the outlet streamflow due to limited observational resources, and in future works it is desirable to include more observations such as watershed soil moisture. Due to the opposite update patterns of the streamflow and soil moisture content, it would be interesting to explore this problem with the water balance analysis. Further research is also expected to focus on the quantification of the model and observation errors as well as evaluation of the state vector with different parameters, etc.

Chapter 4 Soil Moisture Assimilation

In this chapter, EKF is applied to assimilate real *in situ* surface soil moisture into the SWAT model using the same coupling procedure as described in Chapter 3. It is anticipated that this will improve the simulation and 'perfect forecast' (or *forecast*) of deeper-layer soil moisture and streamflow. Soil moisture assimilation is compared to streamflow assimilation and the combined assimilation of both soil moisture and streamflow. Considering the key role of the surface soil moisture in the movement and redistribution of moisture in the full soil profile, both the direct-insert method and open loop simulations are used as references in assessing EKF performance for data assimilation. The selection of the state and observation vectors is discussed in this chapter. This chapter also highlights the importance of quantification of the model error, especially the error covariance between different state variables.

The rest of this chapter is organized as follows: Section 1 introduces the assimilation of soil moisture in hydrological models; Section 2 is a brief review of the methodology of EKF and the SWAT model; Section 3 introduces the study area and data; Section 4 presents the EKF-SWAT coupling scheme, the setting of assimilation scenarios and the error quantification strategy; the results and discussion are provided in Section 5; and the conclusions are provided in Section 6.

4.1 Introduction

Data Assimilation (DA) is a method that tries to estimate the true state of a system by optimally merging observations with dynamic model simulations based on analysis of their uncertainties. Hydrological DA is particularly valuable nowadays with the availability of more observations of the hydrological processes. Among these observations, surface soil moisture and streamflow are two of the most frequently used in hydrological DA. The Kalman filter and its variants are specific DA methods that have been widely accepted as efficient ways to improve the initial condition and parameterization of rainfall-runoff (RR) models (Sun et al., 2015b).

Soil moisture is among the key variables that control the rainfall runoff process, and it has been well studied in hydrological DA (Feddes et al., 1988; Houser et al., 1998; Vrugt et al., 2001; Simunek et al., 2005; Arnold et al., 2012). The correct estimation of antecedent soil moisture content is not only critical for streamflow simulation (Chen et al., 2014; Tayfur et al., 2014; Wanders et al., 2014b), but also to the parameter identification for hydrological models (Shi et al., 2014; Tran et al., 2014; Wanders et al., 2014a; Chen et al., 2015). It is logical to assimilate

soil moisture because it is a continuous water storage state bridging the gap between the hydrological model and state space equations.

Traditionally, soil moisture observations have been obtained with an in situ measurement network (Parrens et al., 2014; Shi et al., 2014). Such networks should be dense enough to estimate the watershed soil moisture; otherwise, the spatial interpolation error would have to be considered (Chen et al., 2011; Al-Yaari et al., 2014). Nowadays, satellite remote-sensing can provide continuous and economical spatial-temporal observation of soil moisture (Xu et al., 2014). There exist many different types of remote-sensing soil moisture products. The majority of them are surface or near-surface soil products (Moradkhani, 2008), and root-zone products are only available occasionally (Wagner et al., 1999; Das et al., 2008). However, the subsurface and root-zone soil moisture have a more significant impact on runoff generation (Brocca et al., 2012; Han et al., 2012; Lei et al., 2014). The widespread application of remote-sensing soil moisture estimates in operational hydrological simulations is thus far limited due to issues like rescaling (Kaheil et al., 2008; Sahoo et al., 2013), error evaluation (Doubkov á et al., 2012; Draper et al., 2013), retrieval algorithms (Verhoef and Bach, 2003; Vereecken et al., 2015), etc. Houser et al. (1998) argued that in hydrological applications, remotely-sensed soil moisture must be supplemented by *in situ* surface and root-zone observations. Therefore, *in situ* surface soil moisture observations will be used in this study.

As opposed to soil moisture products, streamflow measurements are often more available, and hence have been frequently used in hydrological DA (Clark et al., 2008; Abaza et al., 2014a; Randrianasolo et al., 2014; Samuel et al., 2014; Trudel et al., 2014; Mazzoleni et al.,2015). Streamflow is often treated as a diagnostic variable which is not updated directly within the state vector (Liu et al., 2012), and augmenting the state vector with streamflow is seen by some researchers as the only way to obtain a linear measurement operator (Evensen, 2003; Pauwels and De Lannoy, 2009). Although many studies have involved assimilating distributed streamflow observations within a watershed, their results are mixed (Clark et al., 2008; Trudel et al., 2014; Mazzoleni et al., 2015).

In hydrological DA, the assimilation of one variable does not necessarily improve the estimation of the other variables. Sun et al. (2015a) found that assimilating streamflow observations does not simultaneously update the estimates of streamflow and soil moisture in the same way.

Trudel (2014) reported that the assimilation of streamflow distorts estimates of soil moisture. The combined assimilation of multiple observation variables is found to be superior to single-variable assimilation (Crow and Van Loon, 2006; Trudel et al., 2014; Lopez et al., 2015).

The Extended Kalman Filter (EKF) is one of the quintessential nonlinear Kalman filters. EKF linearizes the model by using a Taylor expansion (Dr écourt, 2003; Kumar and Kaleita, 2003; L ü et al., 2011; de Rosnay et al., 2013). Given the difficulty in obtaining explicit solutions of the derivatives of complex hydrological models, numerical methods are often preferred in the calculation of the Jacobian matrix (Walker and Houser, 2001). As the linearization procedure is the only difference between EKF and the linear Kalman filter, the implementation of EKF is usually more straightforward than other types of nonlinear Kalman filter.

The synthetic twin experiment is probably the most widely used strategy in Kalman hydrological DA to verify the effectiveness of assimilation (Crow and Wood, 2003; Lei et al., 2014; Shi et al., 2014). In this method the model and observation errors are artificial (Crow and Van Loon, 2006), and the artificial errors added to the inputs, parameters and model state may deteriorate the model if they are biased (Ryu et al., 2009; Alvarez-Garreton et al., 2015). In order to generalize the conclusions from synthetic twin experiments to real data assimilation, it is vital to correctly quantify the model and observation errors (Moradkhani et al., 2005; Vrugt et al., 2005; Liu and Gupta, 2007; Noh et al., 2014). Puente and Bras (1987) argued that correct error quantification is even more important than the selection of the DA method itself. Due to the challenges in addressing the structure, source, and types as well as magnitude of the model and observation errors (Crow and Van Loon, 2006; Draper et al., 2011; Noh et al., 2014), so far the application of soil moisture and streamflow assimilation in operational forecast settings has been rare (Komma et al., 2008; Chen et al., 2011; Alvarez-Garreton et al., 2015).

In this study, *in situ* surface soil moisture and streamflow observations are assimilated into the Soil and Water Assessment Tool (SWAT) with EKF. Similar to the work of Sun et al. (2015a), a ratio method was suggested to update the states on the Hydrological Response Unit (HRU) scale. A case study is carried out in the Little Washita River Experimental Watershed run by the USDA ARS in southwestern Oklahoma (Elliott et al., 1993). Despite the availability of soil moisture observations in the deeper subsurface layers, only the surface soil moisture is assimilated, and different configurations of state and observation vectors are studied with

numerical experiments. Instead of using synthetic experiments, this thesis assimilates real streamflow and surface soil moisture observation data to generate and verify short-term *forecasts* of streamflow and full-profile soil moisture. The sensitivity of EKF performance to model error is also investigated.

4.2 Methodology

4.2.1 The Extended Kalman Filter

The Kalman filter is an optimal estimator that recursively couples measurements into the linear model with a minimum estimated error covariance (Kalman, 1960). To apply the Kalman filter, transition models must be based on a well-defined linear system. This poses a huge challenge to the application of Kalman filter in hydrological data assimilation, as most rainfall runoff models are highly nonlinear. The Extended Kalman Filter (EKF) is the direct heir of the conventional linear Kalman filter, as it linearizes the nonlinear transition models while still following the algorithm of the linear filter (Walker and Houser, 2005).

Consider a nonlinear system in the form of state-space functions:

$$x_k^f = M_k(x_{k-1}^a, u_k, \eta_k) \tag{4-1}$$

$$y_k^o = H_k(x_k^f, \varepsilon_k) \tag{4-2}$$

where M_k is the model transition function and $H_k$is the observation transition function. Equation (4-1) propagates the state from step *k-1* to step *k*. x_{k-1}^a is the *a posteriori* state vector and x_k^f is the a priori state. u_k is the forcing term, which is the input of the model in this study, and η_k is the model error. Equation (4-2) is the observation function that relates the state vector x_k^f to the observation y_k^o with observation error ε_k.

Assuming η_k and ε_k are jointly Gaussian and independent, we have $E(\eta_k) = \mathbf{0}$ and $E(\varepsilon_k) = \mathbf{0}$. Suppose the state vector $x_k = (x_k^1, x_k^2, \dots, x_k^m)$ and the observation vector $y_k^o = (y_k^{o1}, y_k^{o2}, \dots, y_k^{on})$, the model error covariance matrix can be defined as a symmetric positive semi-definite matrix:

$$Q_k = E(\eta_k {\eta_k}^T) = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{m1} \\ \vdots & \ddots & \vdots \\ \sigma_{1m} & \cdots & \sigma_m^2 \end{bmatrix} \tag{4-3}$$

The diagonal elements of this matrix represent the error variances of each member in the state vector. The off-diagonal elements are the error covariance between different members of the state vector. Q_k degenerates into a diagonal matrix if the errors in state vector members are independent from each other.

The observation error covariance matrix can be defined as another symmetric positive semi-definite matrix:

$$R_k = E(\varepsilon_k {\varepsilon_k}^T) = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{n1} \\ \vdots & \ddots & \vdots \\ \sigma_{1n} & \cdots & \sigma_n^2 \end{bmatrix} \tag{4-4}$$

The quantification of Q_k and R_k is critical, yet there lacks a systemic scheme especially for the determination of Q_k. More discussion about this topic can be found in the relevant sections in this thesis.

The implementation procedure of EKF can be found in section 3.2.1.

4.2.2 Adaptive EKF

In the case of optimal filtering, the model and observation error covariance can be updated recursively with adaptive filtering based on the assumption that the innovation sequences are white noises with a mean of zero (Mehra, 1970; Wood and Szöllösi-Nagy, 1978). However, such an assumption is hard to meet in EKF due to the truncation of the higher-order errors in the Taylor expansion. In this case, the model and observation errors follow a new Gaussian distribution (Meng et al., 2000), $\eta_k \sim N(q_k, Q_k)$, $\varepsilon_k \sim N(r_k, R_k)$, in which q_k and r_k are the new means of the model and observation errors. The adaptive EKF used here is a modified version of the classic Bayes estimation algorithm (Sage and Husa, 1969; Meng et al., 2000).

In the modified adaptive EKF, the basic procedures are the same as for conventional EKF, except that equation (3-9) is rewritten as:

$$x_{k+1}^f = M_{k+1}(x_k^a) + q_k \tag{4-5}$$

The new innovation will be:

$$v_{k+1} = y_{k+1}^o - {H'}_{k+1} x_{k+1}^f - r_k \tag{4-6}$$

and $q_{k+1}, Q_{k+1}, r_{k+1}$ and R_{k+1} are updated with the following equations:

$$q_{k+1} = (1 - d_k)q_k + d_k[x^a_{k+1} - M_{k+1}(x^a_k)] \quad (4\text{-}7)$$

$$Q_{k+1} = (1 - d_k)Q_k + d_k[P^a_{k+1} + K_{k+1} \cdot v_{k+1} \cdot v^T_{k+1} \cdot K^T_{k+1} - M'_{k+1} P^a_k {M'_{k+1}}^T] \quad (4\text{-}8)$$

$$r_{k+1} = (1 - d_k)r_k + d_k[y^o_{k+1} - H_{k+1}(x^f_{k+1})] \quad (4\text{-}9)$$

$$R_{k+1} = (1 - d_k)R_k + d_k[v_{k+1} \cdot v^T_{k+1} - H'_{k+1} P^f_{k+1} {H'_{k+1}}^T] \quad (4\text{-}10)$$

In the above equations, d_k is an index-weighted coefficient defined as (He et al., 2011):

$$d_k = \frac{1-\gamma}{1-\gamma^{k+1}} \quad (4\text{-}11)$$

where γ is the forgetting factor (Meng et al., 2000) with $0 < \gamma < 1$.

4.2.3 The Soil and Water Assessment Tool model

The Soil and Water Assessment Tool (SWAT) model is a distributed basin scale conceptual hydrological/water quality model (Arnold et al., 1998; Srinivasan et al., 1998; Jayakrishnan et al., 2005). The SWAT model is a highly comprehensive model to simulate hydrology, chemicals, sediments, crop growing, agriculture management, etc. within a watershed. The SWAT model has three major components: (1) Sub-basin, (2) Reservoir Routing, and (3) Channel Routing (Spruill et al., 2000). The land area in a sub-basin is divided into hydrological response units (HRUs) based a unique combination of soil type, land use type, and slope. The sub-basin may also contain a tributary channel and a main channel. Users are allowed to prepare input at the level of the watershed, sub-basin, or HRU as well as to add point sources. The model calibration can also be implemented at any of these levels. For the channel routing, SWAT provides two options: the variable storage method and the Muskingum method.

To drive a SWAT model, the following basic inputs are required: DEM, weather input (precipitation, temperature, solar radiation, wind speed, humidity), soil map, and land-use map. Streamflow observations are needed for the purposes of calibration and validation. All these inputs can be processed with ArcSWAT, which is an ArcGIS-ArcView extension and graphical user input interface for SWAT.

The hydrology component of SWAT is based on the water balance equation (Arnold et al., 1998):

$$SW_t = SW + \sum_{i=1}^{t}(R_i - Q_i - ET_i - P_i - QR_i) \quad (4\text{-}12)$$

where SW_t (mm) is the soil water content for day t, SW (mm) is the initial soil water content, R_i (mm) is the daily precipitation, Q_i (mm) is the daily runoff, ET_i (mm) is the daily evapo-transpiration, P_{i} (mm) is the daily percolation, and QR_i (mm) is the daily return flow.

The generated runoff consists of three parts: surface runoff, lateral runoff, and groundwater runoff. The surface runoff volume can be calculated with either the SCS curve number method or the Green-Ampt method. In this study, the SCS curve number method is used:

$$\begin{aligned} Q &= \frac{(R-0.2S)^2}{R+0.8S}, & R > 0.2S \\ Q &= 0, & R \leq 0.2S \end{aligned} \quad (4\text{-}13)$$

where Q (mm) is the daily surface runoff, R (mm) is the daily rainfall, and S is the retention parameter determined by the curve number (CN) with the SCS equation (Saleh et al., 2009):

$$S = 254(\frac{100}{CN} - 1) \quad (4\text{-}14)$$

The curve number is a function of soil permeability, land use and antecedent soil water conditions. The curve number calculated on a condition of average antecedent soil water conditions is termed CN2, and the value of CN2 normally varies between 30 and 100. The curve numbers from dry antecedent soil water condition CN1 and moist antecedent soil water conditions CN3 are calculated based on CN2. CN2 is one of the most sensitive parameters for streamflow simulation in SWAT.

The SWAT model presumes that percolation and lateral subsurface flow from a soil layer will not happen unless the field capacity is exceeded. The lateral flow in each subsurface soil layer is calculated with Sloan's kinetic storage model (Sloan and Moore, 1984):

$$latlyr = \alpha * \frac{2*depth*sw_{excess}}{sol_{ul}-sol_{fc}} * sol_k * slope * \frac{0.024}{L}, \quad latlyr < sw_{excess} \quad (4\text{-}15)$$

$$latlyr = sw_{excess}, \quad latlyr > sw_{excess} \quad (4\text{-}16)$$

where $latlyr$ (mm) is the lateral subsurface flow in a layer; α is an scaling factor; $depth$ (mm) is the depth of the layer; sw_{excess} (mm) is the soil water exceeding the field capacity; sol_{ul} (mm) is the amount of water held in the soil layer at saturation; sol_{fc} (mm) is the amount of water held in the soil layer at field capacity; sol_k (mm/hour) is the saturated hydraulic conductivity of the soil layer; $slope$ (m/m) is the average slope steepness, and L (m) is the slope length for lateral subsurface flow.

The soil moisture in each subsurface layer is updated with:

$$sol_{st} = sol_{st} - sepday - latlyr - lyrtile \tag{4-17}$$

where sol_{st} (mm) is the amount of water stored in the soil layer on the current day; $sepday$ (mm) is the percolation from soil layer; and $lyrtile$ (mm) is the drainage tile flow in soil layer on the current day.

The groundwater flow is calculated from the variation of the shallow aquifer storage (Arnold et al., 1998):

$$V_{sa} = V_{sa} + Rc - rev - rf - pG - W_{sa} \tag{4-18}$$

where V_{sa} (mm) is the shallow aquifer storage; Rc (mm) is the recharge from the soil profile; rev (mm) is the root uptake from the shallow aquifer; rf(mm) is the return flow; pG (mm) is the percolation to the deep aquifer; and W_{sa} (mm) is the water use from the shallow aquifer.

The original SWAT model is written in Fortran format, but in this study SWAT 2009 is translated into Matlab for the convenience of state storage in the calculation of the Jacobian matrix.

4.3 Study Area and Data

The study area is the USDA ARS Little Washita River Experimental Watershed located in southwestern Oklahoma (Figure 4-1). The elevation of the Little Washita watershed varies from 327 m to 486 m. The mean annual precipitation is 760 mm and the mean annual temperature is 16 degrees Celsius. A large volume of *in situ* measurement data is available for this 611 km^2 watershed, dating back to 1936 (Allen and Naney, 1991; Elliott et al., 1993). Currently, the Agricultural Research Service (ARS) monitors the environmental conditions of the Little

Washita watershed with a 20-station network called the Little Washita Micronet (http://ars.mesonet.org/).

The Little Washita watershed has been extensively studied in soil moisture estimation and streamflow simulations (Jackson et al., 1995; Van Liew and Garbrecht, 2003; Cosh et al., 2006; Li et al., 2012; Steiner et al., 2014; Yan et al., 2015). Van Liew and Garbrecht (2003) developed a SWAT model within the Little Washita watershed and confirmed that SWAT is capable of providing adequate simulations for hydrologic investigations there. Li et al. (2012) assimilated the raw AMSR-E soil moisture retrievals with EnKF and found that such assimilation could improve the root-zone soil moisture estimate.

The soil moisture data used in this study are from the Little Washita Micronet network between February 1 and December 31, 2009. The soil water content is measured every 15 minutes at 20 stations at the depths of 5cm, 25cm and 40cm below the ground (black crossings in Figure 4-1), although in this study we only use the daily average values (24-hr summary data).

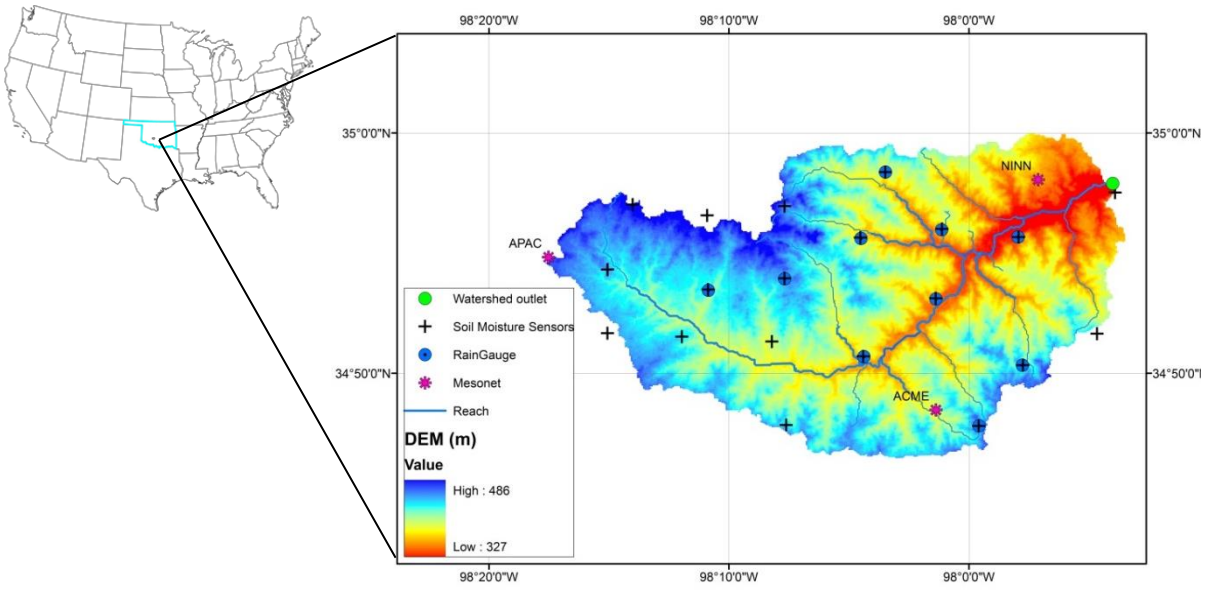


Figure 4-1 The Little Washita Micronet network

The land-use data is from the USGS Global Land Cover Characterization (GLCC) database. The summary of the land-use divisions is presented in Table 4-1. The soil type data are from the FAO/UNESCO Digital Soil Map of the World. According to the FAO soil classification, there

are only two major soil types found in this watershed: Duane soils (62.74%) and Lyons soils (37.26%). There exist more detailed classifications of both land-use and soil type in this watershed (Allen and Naney, 1991), but the GLCC and FAO data are chosen for the sake of simplicity.

One of the challenges in assimilating *in situ* soil moisture data into the SWAT model is that the soil water content measured at the sensor depths does not represent the soil water content in the theoretical soil layers used in the model. In this study, this problem is solved by re-dividing the soil profile. Assuming the soil water content measured at the depth of the sensor is the average soil water content of a 'virtual' layer centered at the sensor depth, three virtual layers are added to the soil profile: ST1 (0-100 mm) for the 50 mm sensor, ST2 (100 mm-400 mm) for the 250 mm sensor, and ST3 (400 mm-500 mm) for the 450 mm sensor (Figure 4-2). The Duane soil (Figure 4-2, left) has three layers whose bottom depths are 106.1 mm, 965.0 mm and 1320.8 mm. Since the depth of the original first layer is very close to that of the virtual first layer, the two are combined as one layer, and therefore the new profile would have five layers. The Lyons soil (Figure 4-2, right) has three layers whose bottom depths are 228.6 mm, 914.4 mm and 1524.0 mm, and the new soil profile would have six layers, as shown in Figure 4-2. In Figure 4-2, the left column represents the Duane soil, and the right column represents the Lyons soil. The green line indicates the bottom of the representative layer for the 50 mm sensor (ST1), the red line indicates the bottom of the representative layer for the 250 mm sensor (ST2), and the blue line indicates the bottom of the representative layer for the 400 mm sensor (ST3).

Most of the meteorological data to drive the SWAT model are from three Oklahoma Mesonet stations (NINN, ACME, and APAC), which are located in the northeastern, southern, and western areas of the watershed respectively. Given the importance of precipitation spatial distribution in SWAT simulations, the precipitation data are provided by ten Little Washita Micronet network stations inside the watershed, instead of the three Mesonet stations. The SWAT model is calibrated with the daily streamflow observations at the site of USGS 07327550 (Little Washita River, east of Ninnekah) from January 1st, 2007 to December 31st, 2008 and verified with the observations from January 1st, 2009 to December 31st, 2009.

Table 4-1 Land-use summary for the Little Washita watershed (GLCC)

Land-use	Description	%Wat.Area
SAVA	Savanna	37.65
FODB	Deciduous Broadleaf Forest	3.29
MIGS	Mixed Grassland/shrubland	56.61
SHRB	Shrubland	2.44

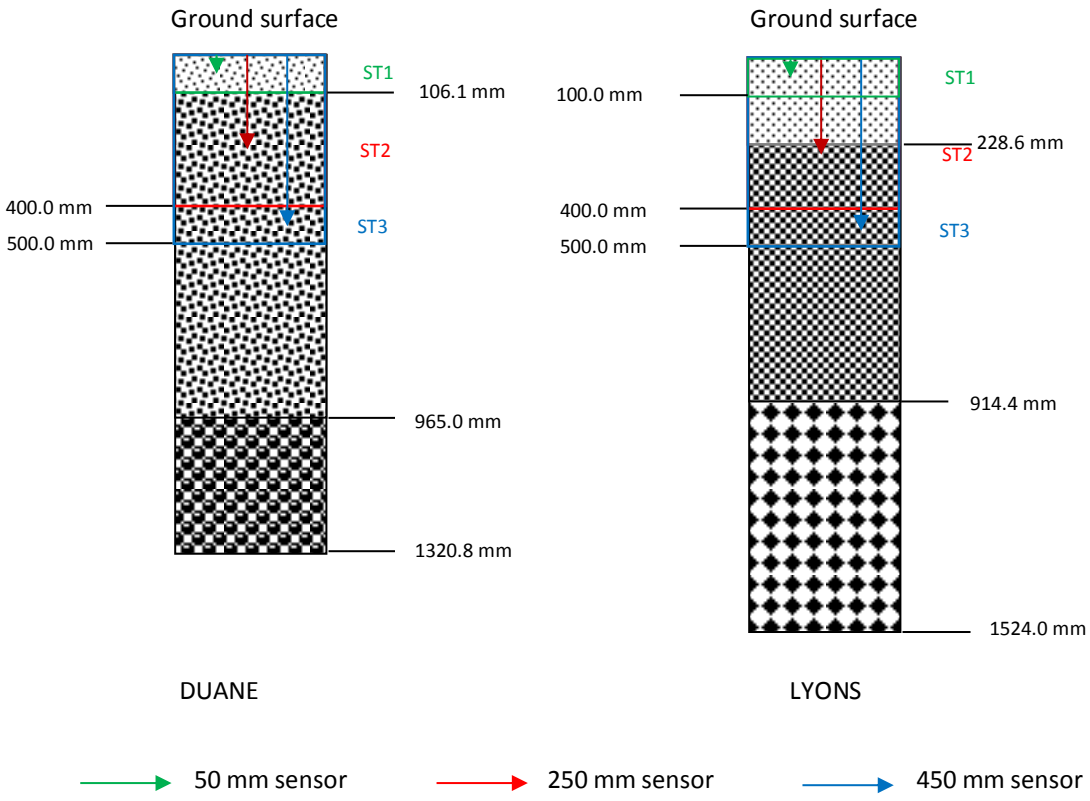


Figure 4-2 Schematic sketch of soil layer re-divisions

4.4 Implementation of EKF

4.4.1 The ratio method

Since the Jacobian matrix is solved numerically by perturbing each element in the state vector, the computational burden largely depends on the size of the state vector. On the other hand, it is challenging to determine the spatial correlations between different HRU scale variables, and it is

therefore not practical to include all state variables at the HRU scale in the state vector. Instead, a ratio method that only updates the state variables at the watershed scale is applied (Sun et al., 2015a). In this method, the update of the variable or parameter at the HRU scale is accomplished using a 'correction ratio' defined as:

$$ratio(i) = \frac{x_k^a(i)}{x_k^f(i)} \tag{4-19}$$

where $x_k^a(i)$ is the i^{th} element of the a posteriori state vector and $x_k^f(i)$ is the i^{th} element of the a priori state vector. Once the watershed scale ratio is calculated, the a priori state variable at the HRU scale will be updated by multiplying the same ratio.

4.4.2 State and observation vector setup

A state variable can either be intermediate, which serves as the 'go-between' from input to output but has no corresponding observations, or prognostic, which can be verified by observations. In addition, it is possible to augment the state vector with the model parameters in order to enhance the benefit of state updates (Gharamti and Hoteit, 2014).

In order to simplify the observation transition function, the state vector selection should also consider the availability of observations. In this study, two types of observation are used: the average daily surface soil moisture over the watershed, and the daily streamflow observations at the watershed outlet. These are either assimilated individually or together into the SWAT model. The estimates and *forecasts* of streamflow and deeper layer soil moisture will be examined to verify whether any improvement has been brought by assimilating these observations.

As shown in Table 4-2, seven experiments were designed to select the state and observation vectors. Experiments titled A to E are used to select the state vector x1with a fixed observation vector *y1* (surface soil moisture). Experiments F and G then use the selected state vectors (from A to E) to compare other observation vectors (streamflow *y2* and combined assimilation *y3*).

In Experiment A, only model-simulated streamflow and watershed average surface soil moisture in ST1 is considered.

In Experiment B, the state vector consists of streamflow and the average soil moisture of three layers: surface (ST1), subsurface (ST2) and deeper layer (ST3). For both soil types, ST1 refers

to the soil moisture content in surface layer 1, with a thickness of 100 mm. ST2 refers to layer 2 for DUANE, but layer 2 plus layer 3 for LYONS. In both cases, it has a thickness of 300 mm. ST3 refers to layer 3 for DUANE, but layer 4 plus layer 5 for LYONS, and in both cases, ST3 has a thickness of 100 mm.

Experiment C is an extension of Experiment A with two more variables added to the state vector: the amount of surface runoff lag over one day (*surf_bs*, mm) and the amount of lateral flow lag over one day (bss, mm). Since the SWAT simulation is done on a daily basis, the motivation for adding these two variables is to examine the impact of lagging water on the estimates of streamflow and soil moisture content. CN2 is arguably the most critical parameter in controlling water redistribution between runoff and soil moisture, and it is also added in the state vector.

Experiment D and Experiment E are both extensions of Experiment B. The only difference between Experiment D and Experiment E is that the former does not include CN2, while the latter does. Notice that among all the candidate state variables and parameters, all but streamflow are distributed, and therefore only their watershed average values are used.

Table 4-2 State vector design

Experiment	Streamflow	ST1	ST2	ST3	surf_bs	bss	CN2
A	*x1*	*x1, y1*					
B	***x1***	***x1, y1***	***x1***	***x1***			
C	*x1*	*x1, y1*			*x1*	*x1*	*x1*
D	*x1*	*x1, y1*	*x1*	*x1*	*x1*	*x1*	
E	*x1*	*x1, y1*	*x1*	*x1*	*x1*	*x1*	*x1*
F	***x2, y2***	***x2***	***x2***	***x2***			
G	***x3, y3***	***x3, y3***	***x3***	***x3***			

Note:

x1 - state vector of surface soil moisture assimilation

x2 - state vector of streamflow assimilation

x3 - state vector of combined surface soil moisture and streamflow assimilation

y1 - observation vector of surface soil moisture assimilation

y2 - observation vector of streamflow assimilation

y3 - observation vector of combined surface soil moisture and streamflow assimilation

4.4.3 Error quantification

The conventional synthetic twin experiments assume that the model and observation errors are known, and hence they are only used to test the effectiveness of the DA method rather than their applicability. In operational hydrological DA, however, error quantification, especially model error quantification, is a big issue as there lacks a systematic solution. This is mainly due to the difficulty in obtaining the source, property and magnitude of modeling errors in land data assimilation (Crow and Van Loon, 2006). Therefore, the trial and error method is the general rule in such studies.

However, the trial and error method has a few drawbacks. First, it is subjective. The success of this method depends on the experience and skills of the modellers. In this study, the traversal algorithm (Chang et al., 1977) is applied to first search for the optimal order of the errors, then to quantify the errors more accurately. Second, it is a very expensive method. Assuming that none of the errors' variance or covariance is known in advance, each of them in the error covariance matrix should be determined individually. Therefore, the computational cost increases exponentially with increases in size of the state vector. This affirms the value of controlling the state vector size in this study by using ratio method. Thirdly, the objective function may converge to a local optimal point instead of the global optimal point. In order to reduce the computational burden, it is necessary to manually set the upper and lower boundaries for each element in the error covariance matrix. However, this risks leaving the global optimal point out of the range that the trial and error method is working on. This problem can be solved by combining more intensive computation with guidance from experts.

Based on the results of the initial trial and error method, it is found that the assimilation is more sensitive to the error covariance between different state variables than to the error variances of the same state variable. In other words, it is more effective to adjust the values of the non-diagonal elements than the diagonal elements of the model error covariance matrix.

A study by Chen et al. (2011) showed that vertical soil water coupling is insufficiently represented in the SWAT model and that this may influence the effectiveness of the assimilation of surface soil moisture in improving deeper layer soil moisture estimation. This highlights the value of error covariance determination of soil moisture content in different layers. Meanwhile, the quantification of the error covariance between streamflow and soil moisture is also found to be critical when their observations are assimilated to improve the model simulations of each other (Alvarez-Garreton et al., 2015). Considering the computational burden, a major focus is given to the quantification of these two groups of error covariance: streamflow and the soil moisture content of each layer, and soil moisture between different layers. For the sake of simplicity, the error covariance related to the intermediate state variable *surf_bs*, *bss* and the model parameter CN2 are determined randomly within reasonable scopes.

The observation errors are assumed to be directly proportional to the magnitude of the observations. They are calculated by multiplying the observations with a time-invariant factor. Meanwhile, we assume that the streamflow observation errors are independent of the soil moisture observation errors, and hence the error covariances between these two are zero in the observation error covariance matrix.

4.4.4 Assessment

In simulation scenarios, assessment of DA effectiveness depends on the selection of the observation variables to be assimilated. For example, when the surface soil moisture is chosen to be assimilated, the model simulations of the soil moisture of two deeper layers and the streamflow on the same day can be assessed, but not the new surface moisture results, as it is not a SWAT simulation but rather the result of the 'artificial' configuration of EKF. For the same reason, when the streamflow is assimilated, the soil moisture of all three layers on the same day can be assessed, but not the streamflow result itself. In *forecast* scenarios, it is desirable that the forecast be improved through the assimilation of historical observations. The short-term *forecasts* for the following seven days are evaluated in order to verify the benefit of the assimilation. To issue the *forecasts*, SWAT will be assimilated with the observation available at time t, then run in 'open-loop' (i.e., without further assimilation) for times t+1, t+2, …, t+7. In this case, the *forecast* values of all variables in the state vector (including the assimilated variable itself) can be used in the evaluation. Two statistic indicators, the Nash–

Sutcliffe coefficient (NSC) and the root mean square error (RMSE), are used in the evaluation of model performance.

4.5 Results and Discussion

4.5.1 Soil assimilation only

In the first case, the surface soil moisture observations for the ST1 layer (0-100 mm, sensor depth of 5 cm) from February 1st to December 31st, 2009 are assimilated into SWAT with EKF to improve the estimate of soil moisture content in the ST2 layer (100-400 mm deep, sensor depth of 25 cm), ST3 layer (400-500 mm deep, sensor depth of 45 cm), and streamflow at the watershed outlet. The model errors are assumed as time-invariant and quantified by trial and error, and the observation error is assumed to be 10% of the observation itself. Table 4-3 and Table 4-4 compare the results of EKF assimilation with those from the open-loop SWAT model (no assimilation) and direct-insert method, in which the surface soil moisture estimate is replaced by observation online instead of being assimilated with EKF. The results of EKF Experiment B, as described in Table 4-2, are displayed in Figure 4-3~Figure 4-5. In these figures, 'obs' represents 'observation', 'OL' represents 'open-loop', and 'ins' represents 'direct-insert'.

Based on NSC and RMSE, it is clear that the assimilation of surface soil moisture can significantly improve the estimates of the soil moisture in deeper layers as well as the streamflow at the watershed outlet. Without assimilation, SWAT severely overestimates both soil moisture (see Figure 4-3 and Figure 4-4) and streamflow (see Figure 4-5). Replacing the model estimate of the surface soil moisture with the observation only marginally improves the model estimate of soil moisture and streamflow compared to updating it with EKF.

Table 4-3 NSC of streamflow, ST2 and ST3 for different experiments

	open_loop	direct_insert	EKF				
			A	B	C	D	E
Streamflow	-0.14	0.28	0.42	0.41	0.42	0.36	0.31
ST2	-7.26	-6.70	-4.71	-1.51	-4.90	-1.54	-1.02
ST3	-0.67	-0.01	0.37	0.60	0.36	0.58	0.50

Table 4-4 RMSE of streamflow, ST2 and ST3 for different experiments

	open_loop	direct_insert	EKF

			A	B	C	D	E
streamflow (m^3/s)	2.54	2.02	1.81	1.82	1.80	1.90	1.97
ST2 (mm)	34.12	32.94	28.45	18.87	28.93	18.98	16.93
ST3 (mm)	5.52	4.30	3.38	2.69	3.40	2.76	3.02

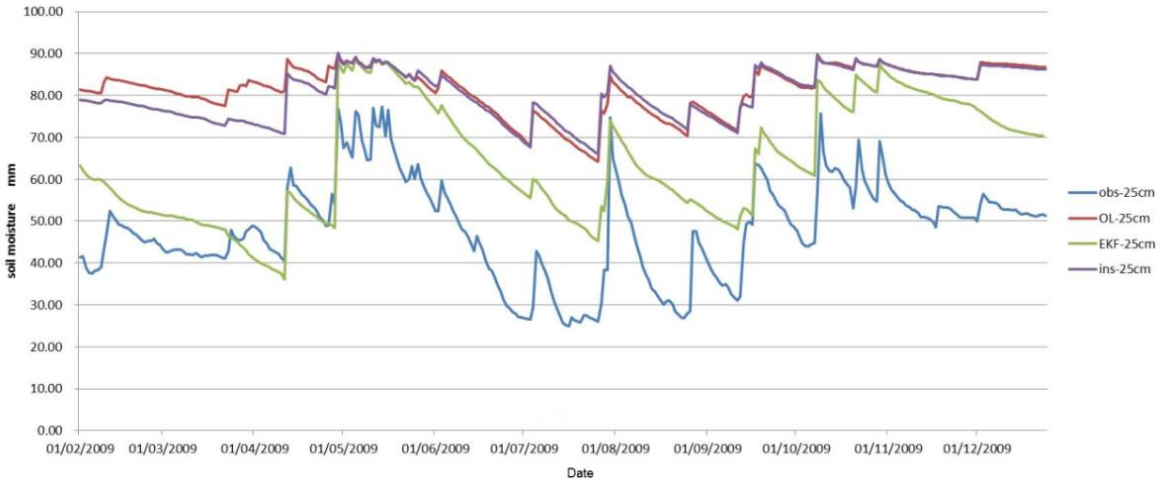


Figure 4-3 Soil moisture content of ST2 layer estimated by surface soil moisture assimilation (EKF Experiment B)

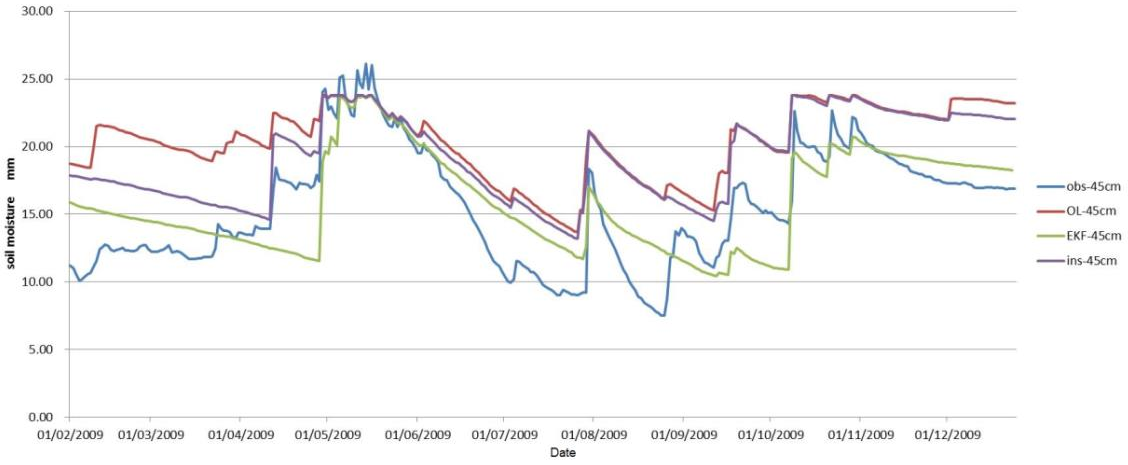


Figure 4-4 Soil moisture content of ST3 layer estimated by surface soil moisture assimilation (EKF Experiment B)

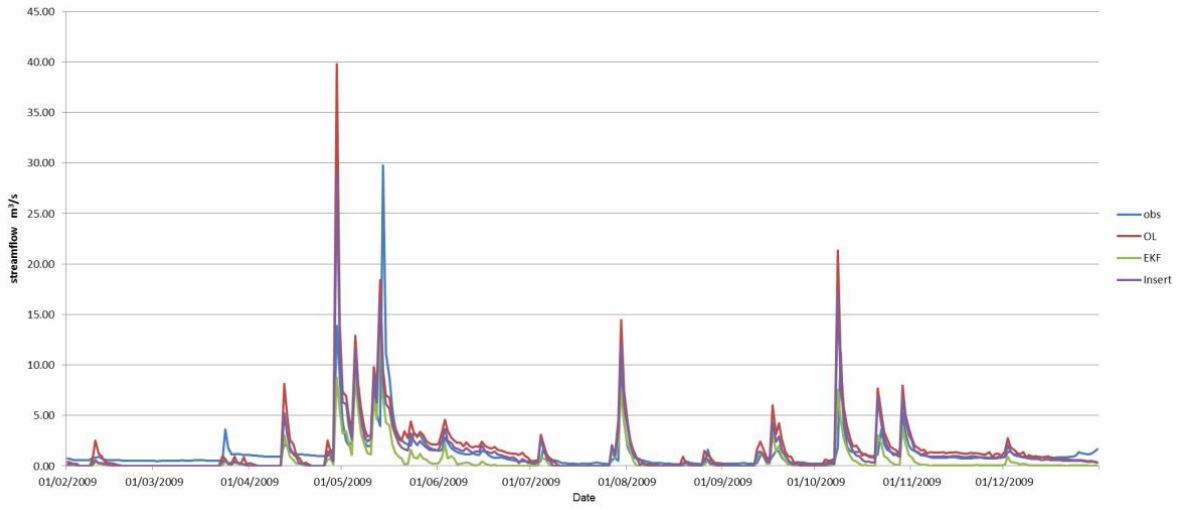


Figure 4-5 Streamflow estimated by surface soil moisture assimilation (EKF Experiment B)

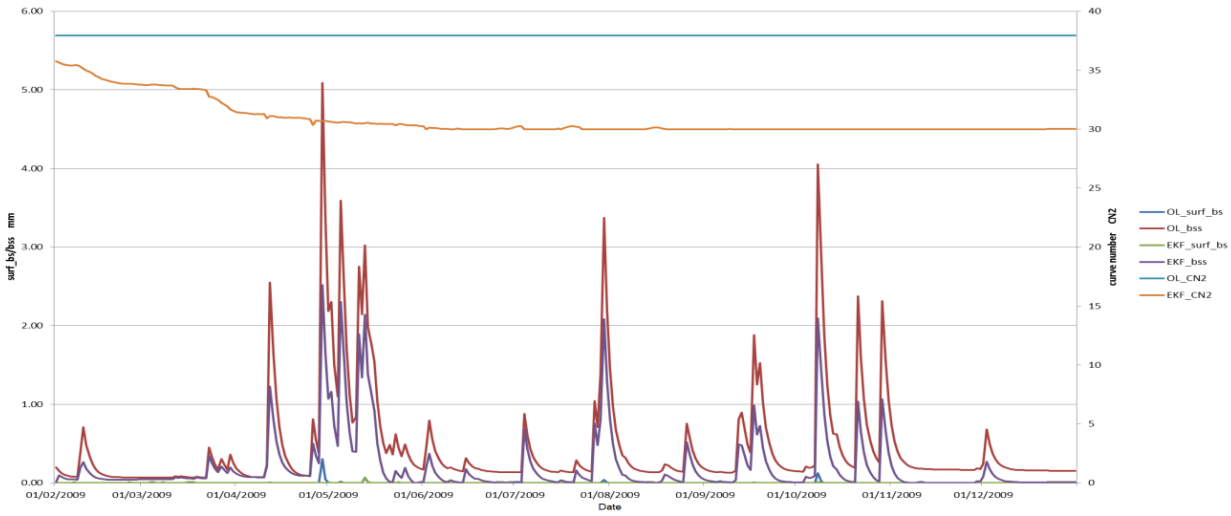


Figure 4-6 The update of intermediate state variables and model parameter CN2 (EKF Experiment E)

It is found that Experiment B significantly outperforms Experiment A in improving the estimates of soil moisture in ST2 and ST3. However, adding intermediate state variables and model parameters to the state vector fails to improve the performance of either Experiment A or Experiment B, and this is especially the case for Experiment C, in which ST2 and ST3 are absent but the state vector is augmented with the lagged surface runoff (*surf_bs*) and lagged lateral flow (*bss*) as well as the curve number CN2.

All experiments have similar accuracy in streamflow estimates. The streamflow improvement is mainly in the reduction of the overestimated peak flow, indicating the great potential of updating soil moisture in improving the *forecasting* of a flood event.

The assimilation results of Experiment E (not shown here) are highly similar to those of Experiment B despite the intermediate state variables and the curve number CN2 being significantly updated in Experiment E (see Figure 4-6). The overall results of Experiment E are not as good as those of Experiment B, and this is likely due to the correction of the overestimates of soil moisture and streamflow in open-loop SWAT being overly enhanced by the update of CN2. First, during the calibration of SWAT, CN2 plays a key role, as it is the most sensitive parameter to the streamflow output of the model. Considering the small size of the watershed, the streamflow at the outlet is very low at most times of year, and hence the model is less sensitive to other parameters such as *sol_k* (saturated hydraulic conductivity), which controls the movement of lateral flow. Second, the error covariance between the model parameters and other state variables might not be determined properly. The error covariances between model parameters and other state variables are more complex to quantify compared to those between regular state variables, as the true values and properties of these parameters are never known, and they have already been calibrated as time-invariant values.

4.5.2 Streamflow assimilation only

In the second case, the streamflow at the outlet is the only observation being assimilated. This new experiment is termed Experiment F (see Table 4-2). Since it has been proved that the augmented state vector does not necessarily produce better results, Experiment F has the same state vector as Experiment B. The model error covariance matrix remains the same as in the first case, and the observation error is also assumed to be 10% of the streamflow observation.

The soil moisture estimation results from the streamflow assimilation with EKF Experiment F are shown in Figure 4-7~Figure 4-9. The statistical indicators NSC and RMSE are listed in Table 4-5. Note that the direct-insert method is not applied in this case because replacing the outlet streamflow would not influence the watershed soil moisture.

As shown in Figure 4-7~Figure 4-9, the update of soil moisture is limited when only the streamflow is assimilated. This is especially true for the surface layer ST1, for which the RMSE

is only reduced from 8.42m^3/s to 7.87m^3/s, and the deeper layer ST3, for which the RMSE is only reduced from 5.52m^3/s to 5.03m^3/s. The update of the subsurface layer ST2 soil moisture is more substantial, where the RMSE is reduced by almost 30% and the NSC is increased from -7.26 to -2.97. Figure 4-8 shows that the overestimation in the open-loop model has been significantly corrected via the EKF assimilation, but such a correction seems very unstable (e.g., the abrupt changes in April and October). This is not surprising considering the complex nonlinear relationship between streamflow and the soil moisture content in each layer, especially those layers with thinner depths (ST1 and ST3) because they are more sensitive to the perturbation of the streamflow in solving the Jacobian matrix. The more positive feedback of the thicker subsurface layer to streamflow assimilation implies the potential to improve the overall soil moisture estimate by better soil layer depth reconfiguration and more robust numerical solutions to the Jacobian matrix.

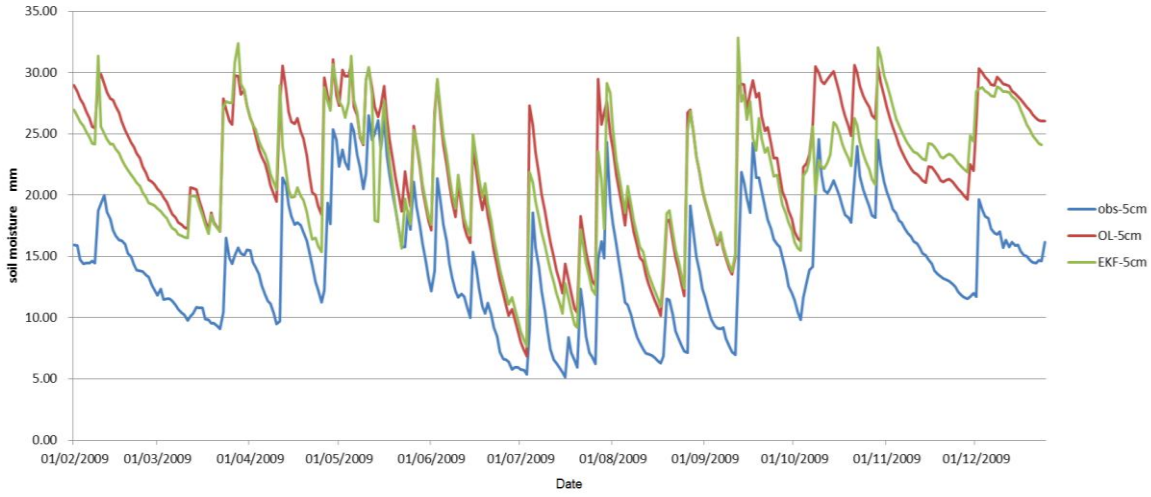


Figure 4-7 Soil moisture content of ST1 layer estimated by streamflow assimilation (EKF Experiment F)

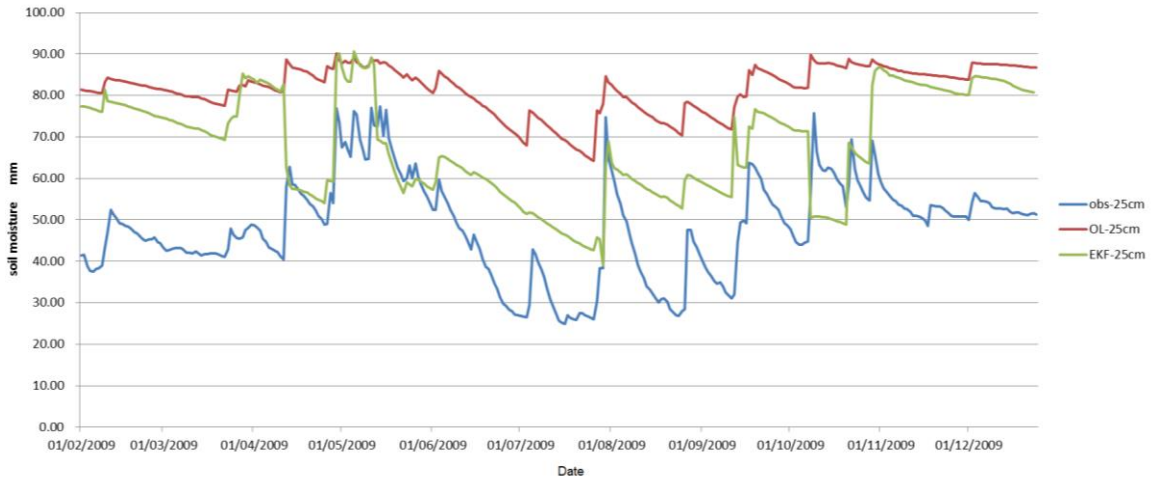


Figure 4-8 Soil moisture content of ST2 layer estimated by streamflow assimilation (EKF Experiment F)

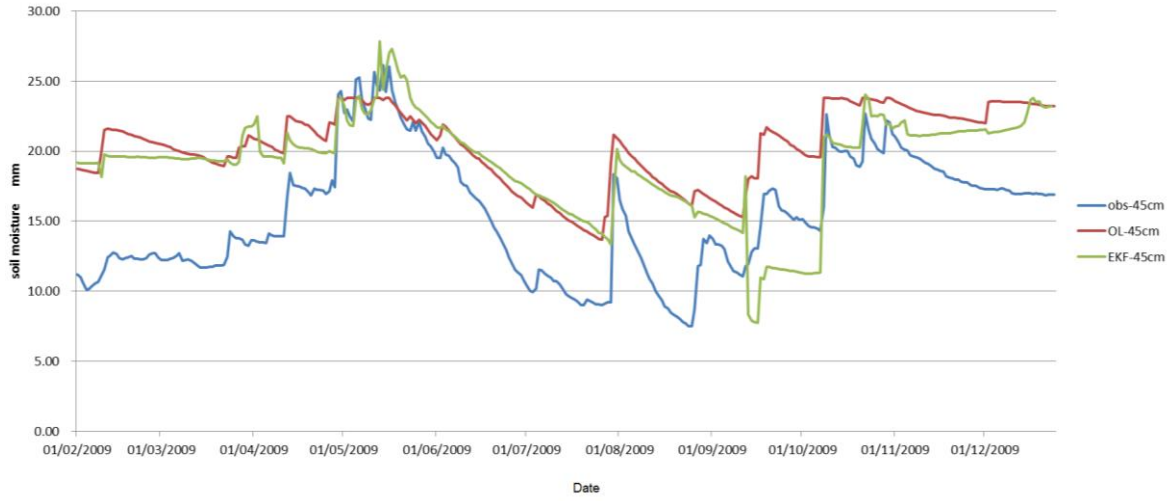


Figure 4-9 Soil moisture content of ST3 layer estimated by streamflow assimilation (EKF Experiment F)

Table 4-5 NSC and RMSE for three soil layers (ST1, ST2, ST3) in streamflow assimilation

	OL-ST1	OL-ST2	OL-ST3	EKF-ST1	EKF-ST2	EKF-ST3
NSC	-1.91	-7.26	-0.67	-1.54	-2.97	-0.39
RMSE(mm)	8.42	34.12	5.52	7.87	23.64	5.03

4.5.3 Combined assimilation

In the third case, both the surface soil moisture and the outlet streamflow are jointly assimilated with EKF. This experiment is termed Experiment G (see Table 4-2). For simplicity, Experiment G also has the same state vector setting as Experiment B, and the model errors are the same as

in the first two cases. For the observation error matrix, the diagonal elements are the variance of the presumed observation error (10% of the surface soil moisture and streamflow), and the off-diagonal elements are zero, based on the assumption that the streamflow observation errors are independent of the surface soil moisture observation errors.

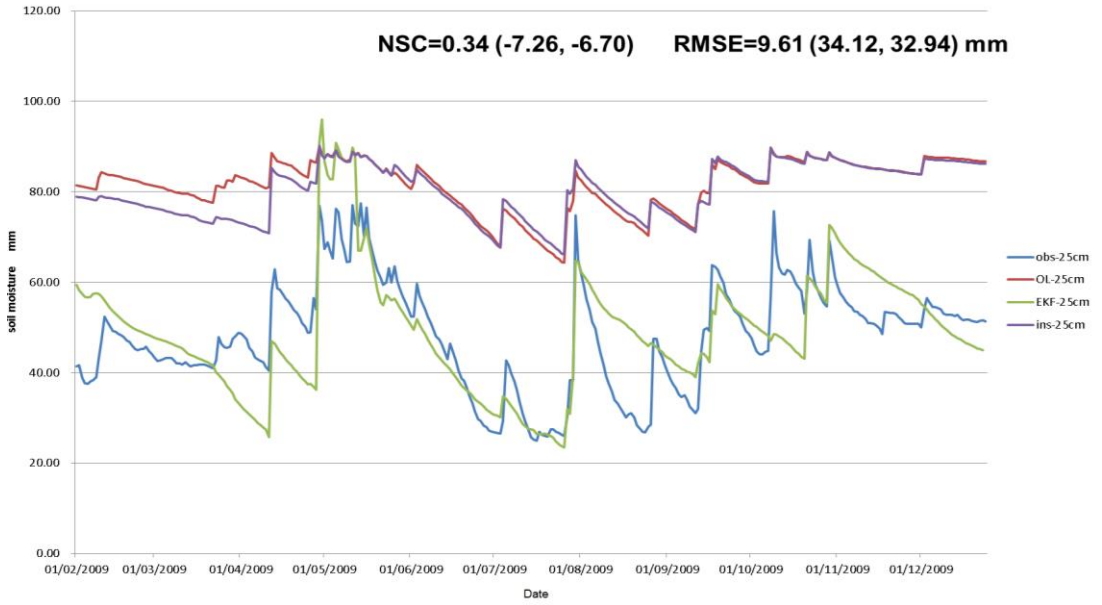


Figure 4-10 Soil moisture content of ST2 layer estimated by combined assimilation (EKF Experiment G)

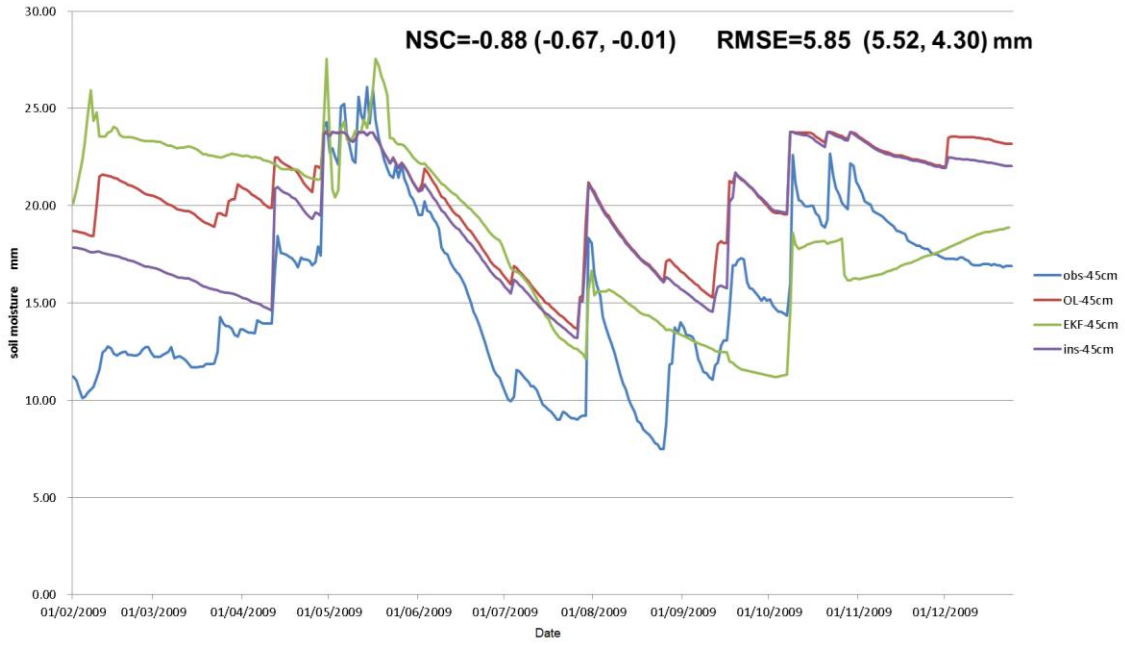


Figure 4-11 Soil moisture content of ST3 layer estimated by combined assimilation (EKF Experiment G)

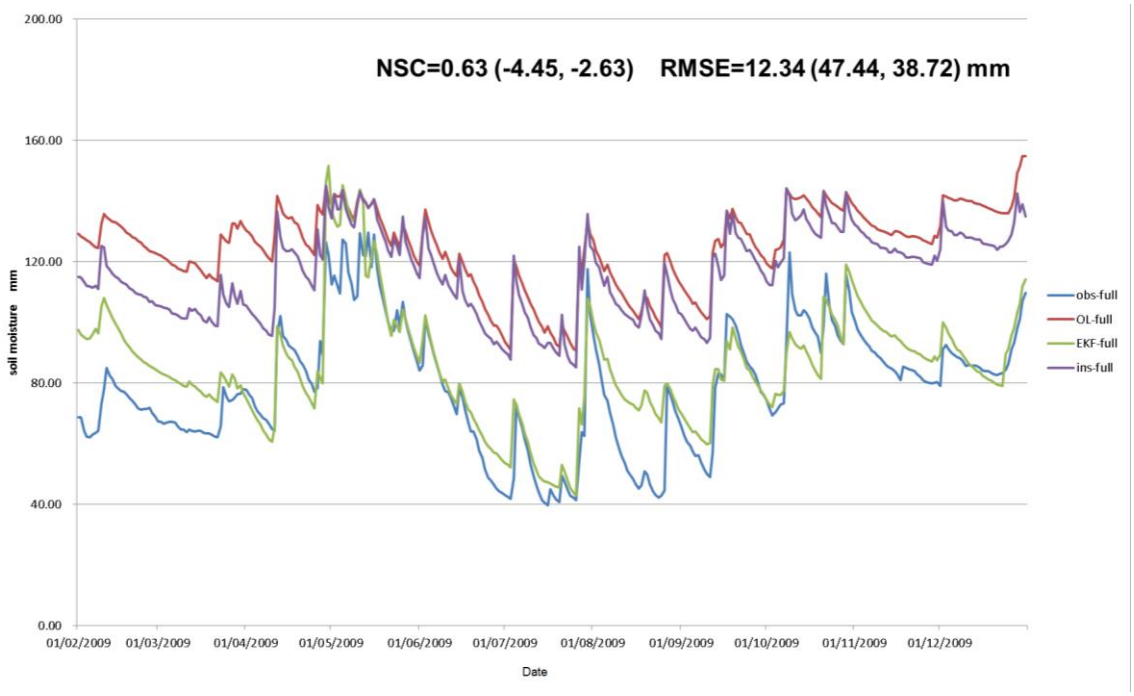


Figure 4-12 Soil moisture content of full profile estimated by combined assimilation (EKF Experiment G)

The assimilation results for the soil moisture in layers ST2 and ST3 and the full soil profile (ST1+ST2+ST3) are displayed in Figure 4-10~Figure 4-12. Notice that the 'full soil profile' does not represent the whole root zone but just the portion from the ground surface to the bottom of ST3, as shown in Figure 4-2. The NSC and RMSE are displayed inside the figures, and the values in the brackets indicate the results of the open-loop model and direct-insert method (applied to surface soil moisture observation only) respectively.

Compared to ST1 and ST3, the correct estimate of soil moisture in ST2 is particularly important, as it is the main body of the full soil profile (300mm out of 500mm). As shown in Figure 4-10, with combined assimilation, the NSC of ST2 soil moisture has been improved to 0.34 from -7.26 in the open-loop model and -6.70 using the direct-insert method, and the RMSE has been reduced by 72% to 9.61mm from 34.12mm in open-loop. Despite the slightly distorted deeper soil moisture estimate in ST3 (Figure 4-11), the combined assimilation improves the soil moisture estimate of the full soil profile significantly, as shown in Figure 4-12. The NSC has been improved to 0.63 from -4.45 in open-loop and the RMSE has been reduced to 12.34mm from 47.44mm in open-loop. The purple line in Figure 4-12 shows the full soil profile estimate of replacing the surface soil moisture with *in situ* observations. It is found that the improvement

of total soil moisture estimate is only minimal with the direct-insert method. This highlights the value of considering the vertical soil moisture interaction between different layers, which is realized by quantifying the error covariance. Meanwhile, it proves that the improvement in the soil moisture estimate of the full soil profile mostly contributes to improvement in the soil moisture estimate in subsurface layer ST2. It is noted that the improvement in the soil moisture estimate of the full soil profile is more significant than that of any single layer (excluding ST1, which is the direct assimilation target layer). This probably results from the fact that EKF does not update the soil moisture estimate in ST2 and ST3 with the same pattern. For example, in April 2009, although the ST2 soil moisture is underestimated, the underestimation has been offset by the overestimation of ST3 soil moisture in the same period, leading to a better estimate of the overall soil moisture.

4.5.4 Forecast mode

In practice, modellers expect that updates of state from current DA will benefit *forecasts* in the future. In this thesis, the one- to seven-day forecasts of full profile soil moisture and streamflow issued with the state-updated SWAT model are examined. These *forecasts* are called 'perfect forecasts' based on the assumption that the future climate input is perfectly known. For the case of the open-loop model, the *forecasts* for all lead times are the same because they all have the same, non-updated initial condition and climate input. However, in the case of EKF assimilation, despite EKF not being involved in the *forecast* period, better *forecasting* is still expected from the updated initial state before a *forecast* is issued.

The performance of the one- to seven-day *forecasts* of both soil moisture and streamflow are shown in Figure 4-13 for NSC and Figure 4-14 for RMSE. These figures only show the results of Experiment B for the three cases discussed above: Case 1 represents only the surface soil moisture assimilation; Case 2 represents only the streamflow assimilation; and Case 3 represents the combined assimilation. 'sm' means the full profile soil moisture, and 'flow' means streamflow.

For the open-loop model, the sm NSC remains -4.45 and the sm RMSE remains 47.71 mm for all *forecast* leads (not shown in the figure). As shown in Figure 4-13 and Figure 4-14, the soil moisture *forecasts* have been improved for all cases, and the combined assimilation outperforms the other two cases. The improvement wanes for all cases, however, with increases in the

forecast lead. For example, in the case of combined assimilation, the one-day *forecast* NSC is 0.57 while the seven-day *forecast* NSC drops to only 0.23.

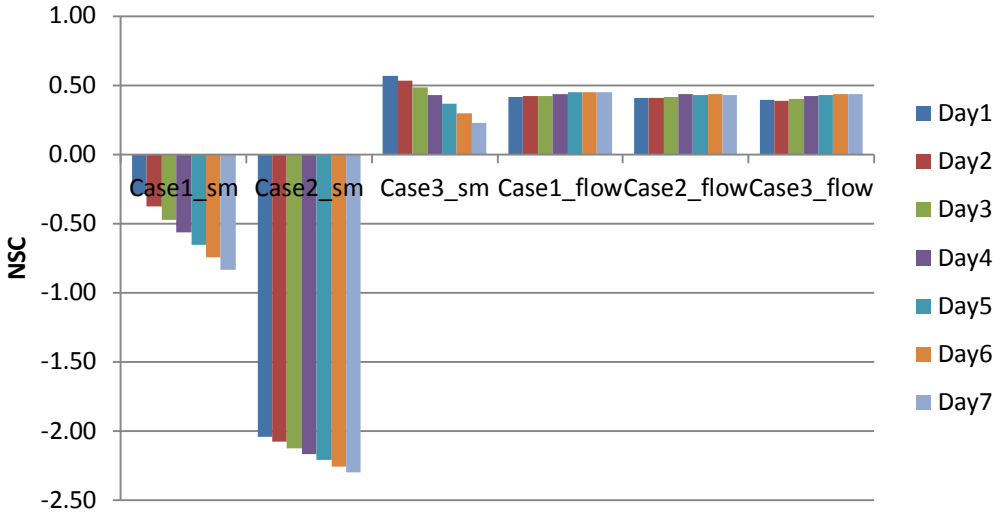


Figure 4-13 NSC for one- to seven-day lead forecasts (soil moisture and streamflow)

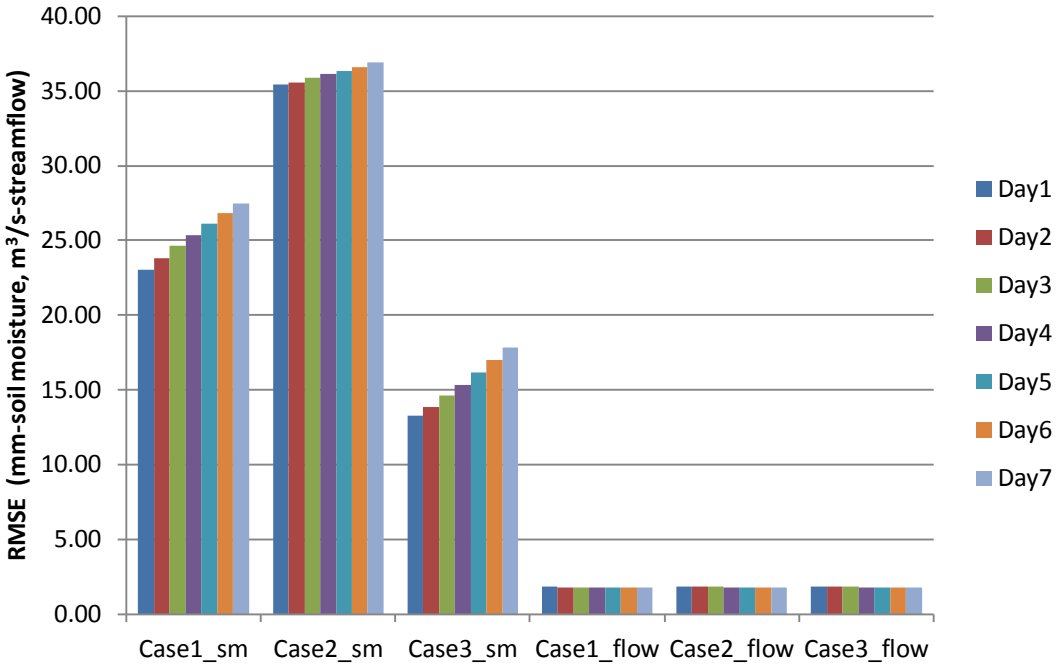


Figure 4-14 RMSE for one- to seven-day lead forecasts (soil moisture *mm* and streamflow m^3/s)

For the open-loop model, the flow NSC remains -0.14 and the flow RMSE remains 2.53 m^3/s for all *forecast* leads (not shown in the figure). Improved streamflow *forecasts* are noted for all cases. Unlike in the sm *forecast*, it is hard to tell which case provides the best streamflow *forecast*, but there is no sign of a wane in improvement with increases in *forecast* lead. In some cases, longer lead times even result in slightly better *forecasts.*

4.5.5 Model error sensitivity

An example of the sensitivity analysis of the model error quantification is shown in Figure 4-15 and Figure 4-16. Figure 4-15 shows the joint influence of error covariance between streamflow and soil moisture content in the ST2 layer, which is marked as cov (flow, ST2), and that between streamflow and soil moisture in ST3, which is marked as cov (flow, ST3), on the estimate of soil moisture in ST2 layer in Experiment G. The precondition of this analysis is that all other error variance and covariance are predefined. As shown in Figure 4-15, with both cov (flow, ST2) and cov (flow, ST3) varying between -0.01 and 0.01, NSC varies from -3.35 to 0.35 and RMSE varies from 9.59 mm to 24.85 mm. The accuracy of the ST2 soil moisture estimate is strongly related to cov (flow, ST2), while only moderately related to cov (flow, ST3). In general, lower cov (flow, ST2) and higher cov (flow, ST3) lead to a more accurate estimate of ST2 layer soil moisture. Hence, in the study, cov (flow, ST2) is set as -0.01 and cov (flow, ST3) is chosen as 0.006.

Figure 4-16 shows the joint influence of error covariance between streamflow and surface soil moisture cov (flow, ST1) and that between streamflow and soil moisture in ST3 cov (flow, ST3) on the estimate of soil moisture in ST3 layer in Experiment G. Specifically, cov (flow, ST2) is set as -0.01. It is found that the estimate of the soil moisture in ST3 is not sensitive to cov (flow, ST1), but sensitive to cov (flow, ST3). As cov (flow, ST3) changes from -0.01 to 0.01, NSC first increases, then decreases, and RMSE first decreases, then increases. A local optimal point (possibly the global optimal point) is found at (0.2, 0.3) where NSC=0.63 and RMSE=2.61 mm. Meanwhile, within this interval, the worst NSC is merely -3.18 (compared to -0.67 in open-loop) and the worst RMSE is as high as 8.70 mm (compared to 5.52 mm in open-loop). In other words, the model simulations can be distorted instead of improved if the error covariance is not quantified properly.

The adaptive EFK theory is usually based on the assumption that the EKF is optimal or only moderately sub-optimal, so that the model errors can be updated online with Bayesian theory (Sage and Husa, 1969; Mehra, 1970). In other words, the model errors are time-variant. In this study, a numerical experiment is carried out using the adaptive EKF method (Meng et al., 2000) to update both the model and observation errors online. The initial errors are pre-determined using the trial and error method. Unfortunately, it is found that the adaptive EKF actually worsens the model performance instead of improving it. The major problem is that the error could not converge (results not presented). This is likely due to the violation of the Gaussian assumption of the model errors during the numerical solution to the Jacobian matrix in a highly nonlinear system. Further study on the potentials of adaptive filtering in EKF is desired.

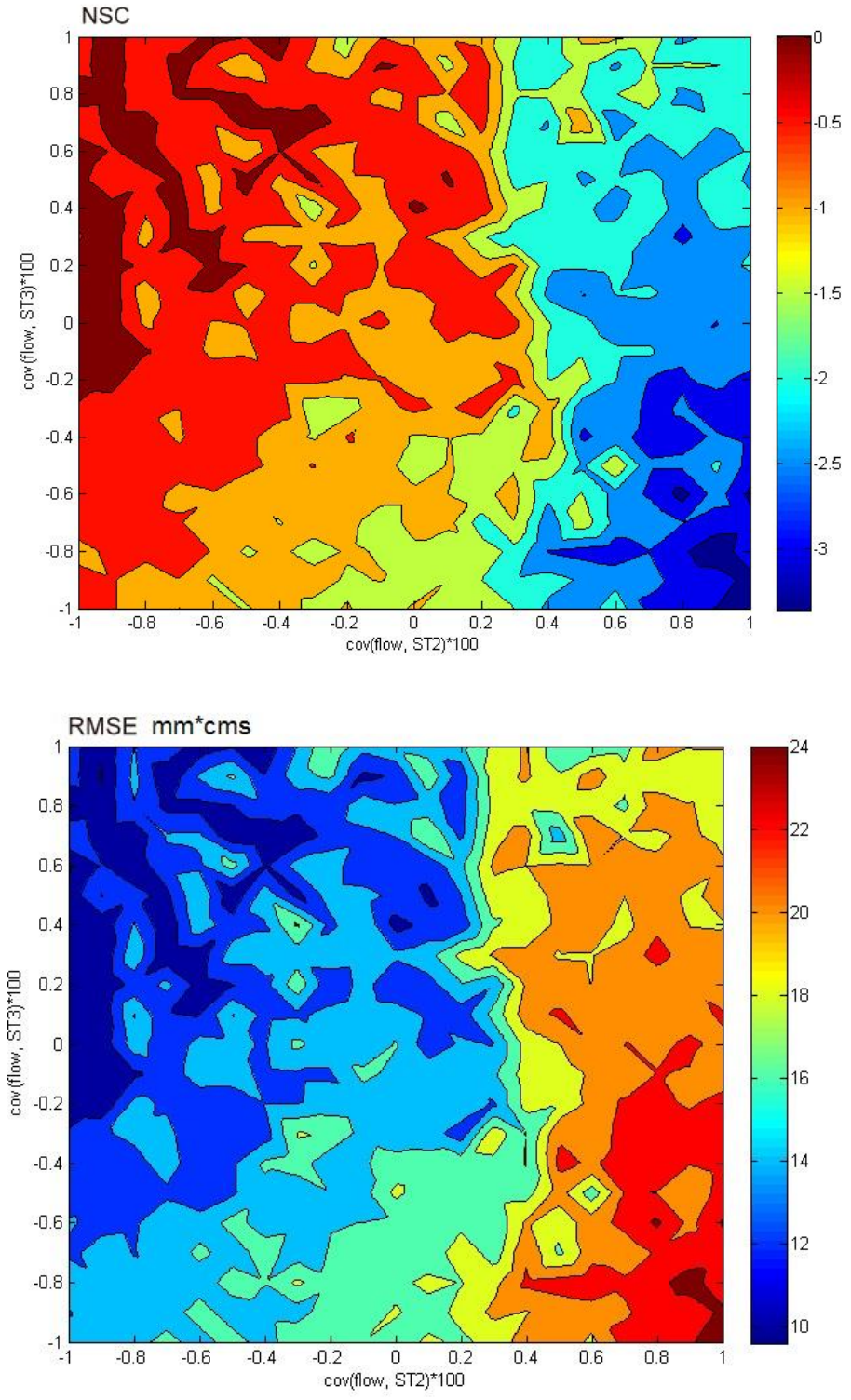


Figure 4-15 The joint influence of cov (flow, ST2) and cov (flow, ST3) on the simulation of ST2 layer soil moisture

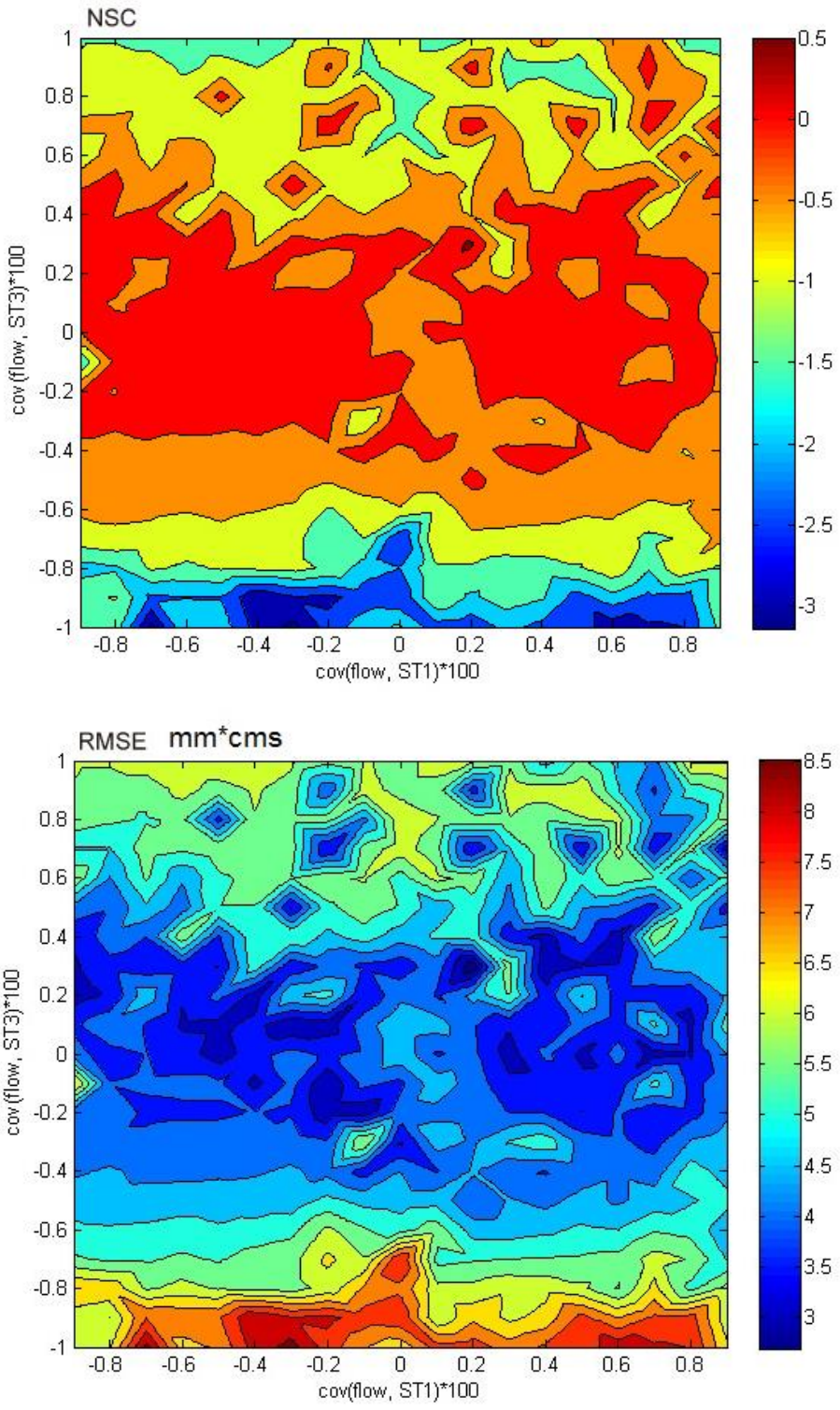


Figure 4-16 The joint influence of cov (flow, ST1) and cov (flow, ST3) on the simulation of ST3 layer soil moisture

4.6 Conclusions

In this study, EKF is applied in the assimilation of real-life surface soil moisture and watershed outlet streamflow in a SWAT model in order to improve the estimates of subsurface and deeper-layer soil moisture as well as streamflow output. The ratio method is used to simplify the state vector. EKF assimilations with different combinations of state variables and observation variables are evaluated based on their capability to improve the estimates of other variables of interest. The model outputs from the assimilated model are compared with those from the open-loop model and the model updated with the 'direct-insert' method. The robustness of the assimilated model is further verified in 'perfect forecast' mode. This thesis has also discussed the sensitivity of model error quantification, especially the error covariance between different state variables.

The EKF assimilation is proved effective in improving the soil moisture estimates and streamflow simulation. However, the effectiveness is influenced by the setting of both the state vector and the observations to be assimilated. The case of state vector with streamflow and soil moisture in all layers outperforms other cases with more complicated state vectors, including those augmented with intermediate variables and model parameters. This shows that the update of model parameters does not necessarily improve the model performance. However, improvement might be expected if the model is been calibrated sufficiently and the error covariances between the model parameters and other state variables are quantified properly. The combined assimilation of both surface soil moisture and streamflow significantly improves the estimates of soil moisture in the full soil profile, while assimilating streamflow only provides very limited improvement in soil moisture estimates.

The updated SWAT model is robust enough to issue better short-term 'perfect forecasts' of soil moisture and streamflow. The soil moisture *forecast* benefits most from the combined assimilation, then from surface soil moisture assimilation, and least from streamflow assimilation. Such benefits tend to wane with increases in the *forecast* lead. The streamflow *forecast* seems sensitive neither to the type of assimilated observations nor to increases in the *forecast* lead. It seems that flood events, instead of a long series of daily streamflows, would benefit more from EKF assimilation.

The performance of the EKF assimilation is directly related to the accuracy of the error quantification. The examples in the case of combined assimilation show that estimates of soil moisture in a specific layer are more influenced by the error covariance between streamflow and the soil moisture in this layer, while not sensitive to the error covariance between streamflow and soil moisture in other layers. The application of adaptive EKF shows no advantages over using the trial and error method to determine time-invariant model errors.

Chapter 5 Comparison of State-parameter Assimilation and Output Assimilation

This chapter compares the EKF state-parameter assimilation with output assimilation in improving streamflow *forecasts* using the SWAT model. Again, the *forecasts* referred to in this chapter are all 'perfect forecasts' with meteorological forecasts replaced with observations. The state-parameter assimilation is performed by updating the stored water content and curve number with EKF, and the output assimilation is carried out by updating the model output errors with autoregressive (AR) models. The performance of the two data assimilation techniques is compared for a dry year and a wet year. A hybrid method that combines both state-parameter assimilation and output assimilation was also developed.

The chapter is organized as follows: Section 1 reviews the development of state-parameter assimilation and output assimilation; Section 2 introduces the methodology; Section 3 introduces the SWAT model and the research area; Section 4 explains the experimental design details, including the implementation of state and parameter assimilation with EKF and the output assimilation with the AR model; the results and analysis are presented in Section 5; and the conclusions are given in Section 6.

5.1 Introduction

Streamflow *forecasting* can be performed with either stochastic or physically-based models. Stochastic models are data-driven and usually simple to implement, but they lack a connection with the real world (Todini, 1988; Todini, 2007). Physically-based models have stronger physical rationality, but do not necessarily generate more accurate forecasts than stochastic models (Zealand et al., 1999; Anctil et al., 2004) because of the additional uncertainties in inputs, parameterization, and model structure (Beven, 1989).

A method that involves both deterministic and stochastic components could benefit from the robustness of the former and the adaptability of the latter (Todini, 2007). Examples of such approaches include probabilistic forecasting (Krzysztofowicz, 1999; Krzysztofowicz, 2001), combined forecasting (Jeong and Kim, 2009; Chen et al., 2015), and ensemble forecasting (Cloke and Pappenberger, 2009; Velázquez et al., 2011). In these approaches, the deterministic models are typically applied within a statistical framework (Ajami et al., 2007; Georgakakos et al., 2004). Challenges in such approaches include computational burden and outputs that are less

intuitive for decision-makers when they are used in an operational setting (Ajami et al., 2007; Cloke and Pappenberger, 2009; De Kleermaeker and Verkade, 2013).

Data assimilation (DA) is a method that combines the strengths of the deterministic model and statistical estimation methods by integrating the most recent observations into model outputs (Kitanidis and Bras, 1980b; Vrugt et al., 2006). DA is easy to implement, as it normally only requires one model and minimum historical observations. In a broad sense, DA includes input assimilation (Chen et al., 2014; Massari et al., 2014), state and parameter assimilation (Vrugt et al., 2005; Hendricks Franssen and Kinzelbach, 2008; Lü et al., 2011; Moradkhani et al., 2005b), and output assimilation (Anctil et al., 2003; Yu and Chen, 2005; Broersen, 2007; Sene, 2008).

State and parameter DA has experienced rapid development since the 1990s. This type of DA can be divided into objective methods and sequential methods. Compared to objective methods, sequential methods are capable of handling all sources of uncertainty (Moradkhani, 2008) and are less complex to implement. Sequential methods commonly used in hydrologic DA include Particle Filter (Pham, 2001; Moradkhani et al., 2005a; Weerts and El Serafy, 2006), H-infinity filter (Wang and Cai, 2008; Lü et al., 2010), and the linear Kalman Filter (Kalman, 1960) and its heirs, like the Extended Kalman Filter (EKF) (Puente and Bras, 1987), Ensemble Kalman Filter (EnKF) (Evensen, 1994), Unscented Kalman Filter (UKF) (Wan and Van Der Merwe, 2000), etc.

Theoretically, the Extended Kalman Filter (EKF) is the standard method to expand the linear Kalman filter to nonlinear systems. If done properly, the implementation of EKF is neat and straightforward. However, due to the complexity of the linearization of the nonlinear hydrological model, this method has been rarely used in hydrological DA, especially in the case of distributed models (Kitanidis and Bras, 1980b; Walker and Houser, 2001).

Output assimilation (also known as error update or error assimilation) is fundamentally different from state and parameter assimilation. In a typical output assimilation, a stochastic model is calibrated using the historical model output error, and the improved forecast is calculated by adding the predicted error to the initial model forecast. Examples of stochastic models used in output assimilation include Auto-regressive Moving Average (ARMA) models (Angela, 1982; Broersen, 2007; Wu et al., 2012; Chen et al., 2015), Artificial Neural Networks (ANNs) (Anctil

et al., 2003; Jain and Srinivasulu, 2004; Khac‐Tien Nguyen and Hock‐Chye Chua, 2012; Jeevaragagam and Simonovic, 2013), the 'dual pass' method (Pagano et al., 2011), hybrid methods that combine ARMA or ANNs with Kalman filters (Bidwell and Griffiths, 1994; Muluye, 2011), etc. Xiong and O'connor (2002) compared the AR model with other more complicated models and found that the latter do not necessarily outperform the standard AR model.

Output assimilation methods are independent of the structure and parameters of deterministic models, and hence they are usually more efficient to apply. The deterministic models do not receive any feedback from the error update, but only generate more error samples into the training pool if the update is carried out online (Tingsanchali and Gautam, 2000). Compared to state-parameter assimilation methods, output assimilation methods are incapable of handling uncertainties in the observations, and neither can they improve the deterministic models' structure and parameters. For this reason, output assimilation is usually found to be more effective than deterministic models for shorter-range forecasting (Angela, 1982).

Madsen and Skotner (2005) described a combined state and output assimilation scheme in real-time flood forecasting with the MIKE 11 model. The state update procedure was used in the hydrodynamic module, and error forecasting was implemented with a second-order autoregressive model. Bauer-Gottwein et al. (2015) applied the standard linear Kalman filter in the state update of the Muskingum routing model in the Soil and Water Assessment Tool (SWAT) model. Both of these two studies only focused on the hydrodynamic module of the deterministic models. There is a growing consensus in the literature (Berg and Mulroy, 2006; Berthet et al., 2009; Brocca et al., 2009; Javelle et al., 2010; McMillan et al., 2013) that updates of the initial conditions of the watershed, especially the antecedent moisture content (AMC), have great potential in the enhancement of streamflow forecasting. Hybrid assimilation of the rainfall runoff module may improve longer-range forecasting, but to the authors' knowledge, such research is rare (Bauer-Gottwein et al., 2015).

5.2 Extended Kalman Filter and ARMA Model

5.2.1 Extended Kalman Filter

The introduction of EKF theory can be found in section 3.2.1.

5.2.2 ARMA models

It is acknowledged that a stationary stochastic process can be approximated by a suitable auto-regressive moving average (ARMA) model (Bidwell and Griffiths, 1994; Choi, 2012). Given a time series x, an ARMA model of orders p and q can be described by:

$$x_t + a_1 x_{t-1} + \cdots + a_p x_{t-p} = \varepsilon_t + b_1 \varepsilon_{t-1} + \cdots + b_q \varepsilon_{t-q} \quad (5\text{-}1)$$

where $a_1, \ldots, a_p$, $b_1, \ldots, b_q$ are parameters, $\varepsilon_t, \ldots, \varepsilon_{t-q}$ are white noise. The ARMA model degenerates into an AR model if q=0 and an MA model if p=0. The determination procedure of the orders and parameters of an ARMA model is well established by the Box-Jenkins method (Box and Jenkins, 1976). Although higher orders generally lead to more accurate forecast results, lower orders are found to be sufficient and more applicable for short-term forecasting (Broersen and Weerts, 2005; Wu et al., 2012).

In order to obtain superior forecast accuracy, the following aspects should be considered: data selection, principal components, cross-validation, and optimal combination of independent variables (Garen, 1992). In more advanced cases, more complicated methods such as step-wise regression can be used to select and evaluate the order of importance of the predictors (Berger and Entekhabi, 2001; Ssegane et al., 2012). The AR model normally only involves the time series of one variable. In this thesis, the performance of multi-variable AR models is compared with that of single-variable models. Although multi-variable models violate the definition of an AR model, such examination may help to determine an optimal model that is more effective and robust.

5.3 The SWAT Model

5.3.1 Model structure

The introduction of the SWAT model can be found in section 3.2.2.

5.3.2 Study area

The SWAT model is developed for the watershed of the Bakel Hydrometric Station on the Senegal River in Senegal (see Figure 5-1). The watershed has an area of 420,546 km^2. The streamflow at Bakel Station is characterized by eight months of low flow (November to June) but unpredictably large floods in the rainy season, from July to October (Rimkus, 2005). Given

that streamflow observations are available at three interior stations (Gourbassi, Manantali and Oualia) on the southern tributaries (Table 5-1), these stations (marked 'Inlet' in Figure 5-1) are treated as point inputs for the SWAT model. By such implementation, the SWAT model becomes a nonlinear input-output model, and the state update is only applied to the blue-shaded area in Figure 5-1. For the sake of simplicity, in this thesis the errors in the streamflow observations of these three interior stations are not considered explicitly in the implementation of EKF.

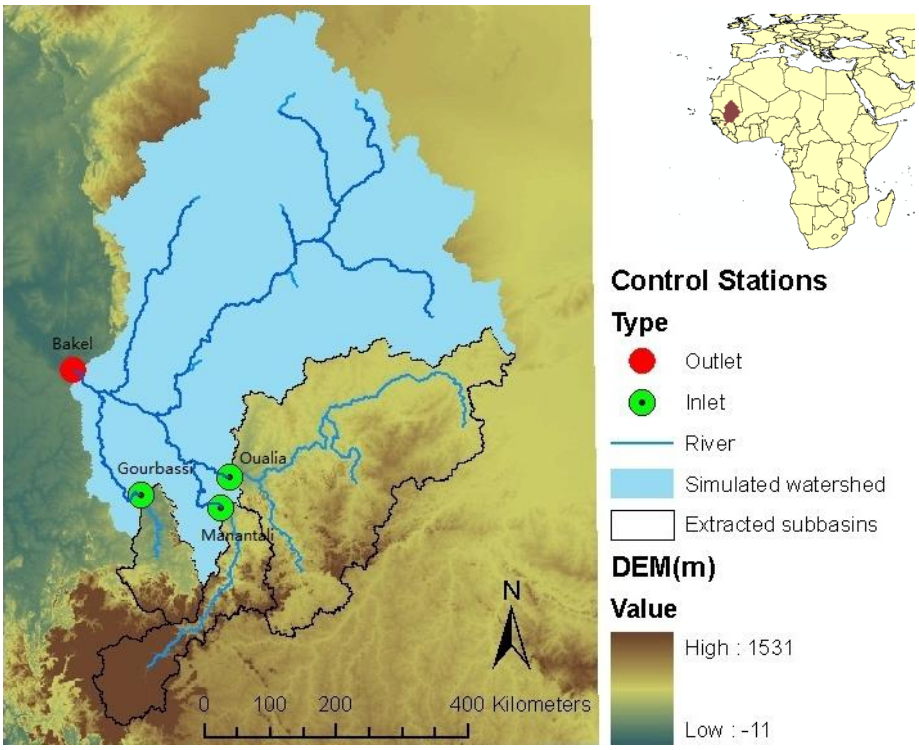


Figure 5-1 **Senegal River Watershed above Bakel Station (The extracted sub-basins are subtracted from the watershed, and only the blue-shaded watershed area is simulated)**

Table 5-1 Statistics of the streamflow observations (1988-2006) for Bakel and the three interior stations

Stations	Mean (m^3/s)	Standard Error (m^3/s)	Cs	Cv
Oualia	77	41	1.03	0.54
Gourbassi	76	31	0.65	0.41
Manantali	214	77	1.22	0.36
Bakel	446	167	1.12	0.37

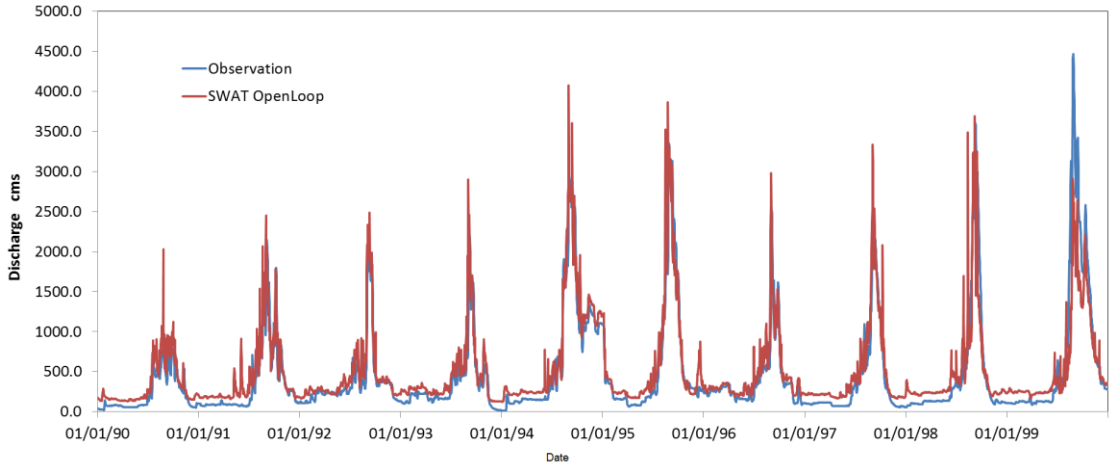


Figure 5-2 Bakel Station streamflow observations vs. SWAT model output (1990-1999)

The SWAT model is calibrated using the data from 1990 to 1998 and validated with data from 1999 (see Figure 5-2). The streamflow predictions for one- to seven-day leads in 1992 and 1999 are evaluated. 1992 was a relatively dry year for which SWAT performs well. 1999 was a particularly wet year, and the performance of the SWAT simulation was lower than average. The comparison of the model's performance in these two representative years reveals the effectiveness of the assimilation methods in different hydrological regimes. Nevertheless, the results should be interpreted objectively, and more case studies are required for the generalizability of these conclusions.

5.4 EKF State Assimilation and ARMA Output Assimilation

5.4.1 EKF state assimilation and forecast assessment

5.4.1.1 Ratio update method

Hydrological DA allows for a lot of flexibility in the selection of state variables, and the water storage variables are the ones most commonly considered (Clark et al., 2008; Samuel et al., 2014). SWAT is usually operated with a daily time step; however, the water generated from HRUs may take more than one day to reach the channel. To handle this issue, SWAT divides the surface, lateral and ground water runoff into two classifications: water that reaches the channel in the current day, and lagged water, which takes more than one day to reach the channel. In this study, three intermediate water storage variables in the state vector are chosen

based on trial and error: soil water content of the full soil profile, lagged surface runoff, and lagged lateral runoff.

Besides the water storage variables, it is often beneficial to update the model parameters in the hydrological DA (Moradkhani et al., 2005a; Lü et al., 2011). In EKF, the parameters can be augmented in the state vector. In this thesis, only CN2 is selected because it is not only critical to the runoff simulation but is also found to be the most sensitive parameter affecting the simulation results.

The streamflow simulated by SWAT is also placed in the state vector. This way, the linear observation operator would be a vector whose designation is 1 with regard to streamflow, and zeroes for the rest (Evensen, 1994). Hence, the final state vector contains the model output of streamflow, the three intermediate water storage variables: soil moisture, lagged surface runoff and lagged lateral runoff as well as the curve number CN2. The observation operator is simply a linear vector of [1, 0, 0, 0, 0].

One of the issues in the implementation of EKF in distributed models is the horizontal correlation of the distributed variables. Stacking all the distributed variables in the state vector would lead to an exceptionally large dimension of the state vector. In this study, the ratio method described in the previous chapters has been used to update the HRU state. Compared to 'adding' or 'replacing' methods, the ratio method maintains the maximum spatial heterogeneity of the original distributed variables and parameters. Note that the simulated streamflow is not a distributed variable and hence is not updated using this method.

5.4.1.2 Error Quantification

The true state is a trade-off between model output and observations based on the comparison between their errors. Hence, the determination of the true state is essentially the quantification of the model errors and the observation errors. The model errors have various sources, such as inputs, parameters, model structure, boundary and initial conditions, etc. (Liu and Gupta, 2007). Underestimation or overestimation of model errors may cause the assimilation results to rely overly on the observations or the model output (Kitanidis and Bras, 1980a). Observation errors are usually caused by the measurement method and/or equipment error. For example,

streamflow observation errors are not only caused by the determination of the stage-discharge curve, but may also be due to errors during the stage measurements.

In this thesis, the model error covariance matrix Q is determined by the trial and error method, following the work of Puente and Bras (1987). The observation error covariance matrix R is the stream flow variance calculated with Equation 3-23..

Despite the importance of error quantification, the trial and error method is still the mainstream method, and therefore the error quantification process is often regarded as the 'calibration' of the DA model. It is worth pointing out that observation errors and/or imperfect initialization of the model might be offset by cumulative updates of the errors (Mehra, 1972).

5.4.2 Output assimilation

Output assimilation is based on the assumption that the errors in the model output are additive. Figure 5-3 shows the sample autocorrelation function (ACF) of both streamflow observation and the SWAT output errors for 1990-1999, and shows that both have a strong autocorrelation for up to 20 lags. The error ACF is less robust than the observation ACF, indicating that the ARMA model, which also includes observations as input in addition to pure errors, might produce better predictions.

For simplicity, two two-order AR models are suggested:

$$e(i+t) = a_1(i)e(i) + a_2(i)e(i-1) \tag{5-2}$$

$$e(i+t) = a_1(i)e(i) + a_2(i)e(i-1) + b_1(i)Q\_obs(i) + b_2(i)Q\_obs(i-1) \tag{5-3}$$

where $e(i+t)$ is the predicted error of t days ahead of day i, $a_1(i)$, $a_2(i)$, $b_1(i)$, and $b_1(i)$ are the time-variable coefficients calculated with the least linear square method, $e(i), e(i-1)$ are the errors of the SWAT model output at day i and $i-1$, $Q\_obs(i)$ and $Q\_obs(i-1)$ are the streamflow observations at day i and day $i-1$ respectively.

With the predicted error, the predicted streamflow can be calculated with:

$$Q(i+t) = Q_{ol}(i+t) + e(i+t) \tag{5-4}$$

where $Q_{ol}(i+t)$ is the SWAT output of day $i+t$, assuming that the climate prediction is reliable.

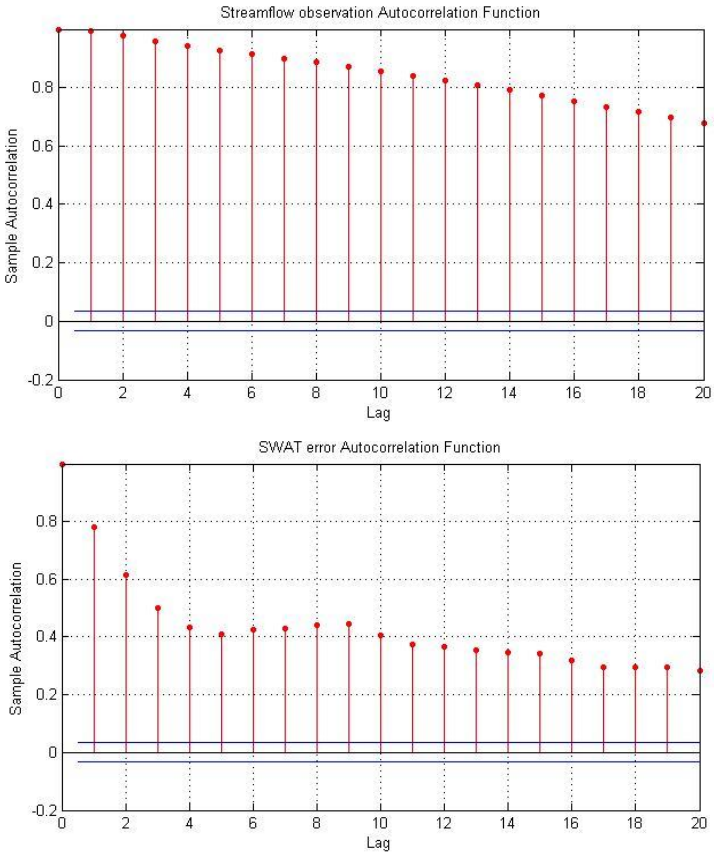


Figure 5-3 Autocorrelation functions of streamflow (upper) and streamflow errors (lower) at Bakel Station, 1990-1999. The blue lines indicate the 0.05 significance level for the ACF.

5.4.3 Experimental Design

Four numerical experiments were designed in order to evaluate the performance of different assimilation schemes. As shown in Table 5-2, Experiment A is simply the open-loop SWAT model without any assimilation. It is therefore used as the reference experiment. Experiment B applies the AR models as output assimilation based on the output of Experiment A. Depending on the selection of AR inputs, Experiment B is further divided into two sub-experiments: in Experiment B-I, the inputs of the AR models are the errors between the model output and the observation for the previous two days (see Equation (5-2)), and in Experiment B-II, the inputs are the errors as well as the observations of the previous two days (see Equation (5-3). The outputs of the two sub-experiments are both calculated with Equation (5-4). In Experiment C,

EKF is used for the state and parameter assimilation. EKF is first applied to SWAT continuously up to the date the *forecast* is issued. Before that, SWAT is run in open-loop mode to issue *forecasts* Q_{ol}^{EKF}. In Experiment D, EKF and the AR model are used simultaneously, but the error prediction model here is not based on Experiment A but on Experiment C. This means that the model errors in both the training and validation stages refer to the differences between the observations and the outputs of the SWAT model for which the parameters have been updated using EKF. In *forecasting* mode, the errors are simply added to Q_{ol}^{EKF}. The AR model used to simulate the errors for 1992 (1999) is trained with the 1990-1991 (1990-1998) observations. RMSE and NSC are used to assess the accuracy of the model predictions.

Table 5-2 Experimental Settings

Experiment	Description	Assimilation type
A	SWAT open-loop Q_{ol}	-
B (B-I, B-II)	AR error update with Q_{ol}	Output assimilation
C	EKF subsequent SWAT open-loop Q_{ol}^{EKF}	State and parameter assimilation
D	AR error update with Q_{ol}^{EKF}	Hybrid

5.5 Results and Analysis

5.5.1 State-parameter assimilation

5.5.1.1 Assimilation results

The comparison between EKF assimilation and the open-loop model output at Bakel Station for 1992 and 1999 is shown in Figure 5-4 and Figure 5-5. For 1992, SWAT in open-loop mode (Experiment A, red line) mostly overestimates the observed streamflow (green line). For 1999, SWAT in open-loop mode overestimates the dry season streamflow but significantly underestimates the wet season flood. The EKF assimilated streamflow (blue line) is similar in observations versus the model output for both years. This indicates that there is a higher confidence in the observations than in the model output.

The grey dashed lines in Figure 5-4 and Figure 5-5 show the values of the Kalman gain *K* for streamflow. With high *K* values, the model output is less considered by the EKF assimilation for the dry season. *K* drops to lower values only in major flood periods. One of the possible reasons for this is that the majority of the streamflow in the dry season is baseflow originating from ground water, while the state vector only contains water storage variables and parameters that

are critical to surface and lateral runoff. In addition, given that the model errors are set to be constant while the observation errors are proportional to the scale of the streamflow, the larger observation errors give the model outputs more confidence, leading to lower *K* values. Considering the different performances of EKF before and after July, the streamflow *forecasts* from July to December are discussed separately as 'wet season' streamflow in the following sections (unless the term 'wet season' is redefined).

The soil moisture assimilation results for 1992 and 1999 at the watershed scale are shown in Figure 5-6 and Figure 5-7 respectively. In both years, soil moisture increases significantly in the wet season. Despite some fluctuations, applying EKF increases the soil moisture estimation for the wet season of 1992. In contrast, applying EKF significantly reduces the soil moisture estimation for the wet season of 1999, especially during the peak flood period. Note that the SWAT model overestimates the flood in 1992, while it underestimates the flood in 1999.

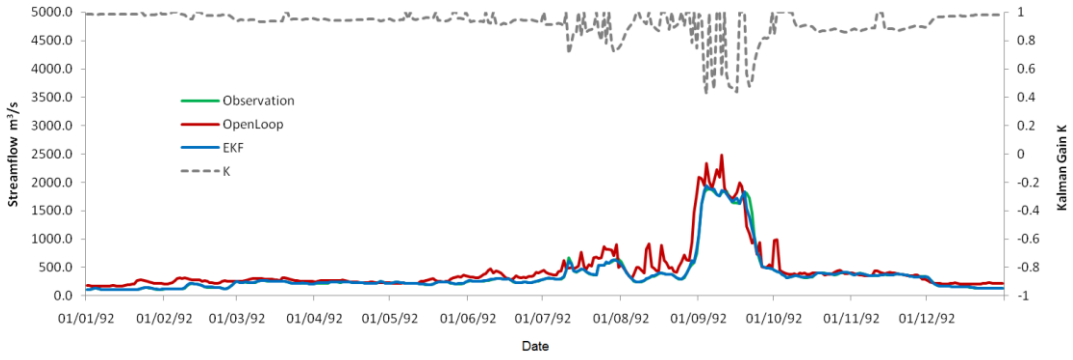


Figure 5-4 1992 streamflow EKF assimilation results compared to observations (Obs) and open-loop (OL)

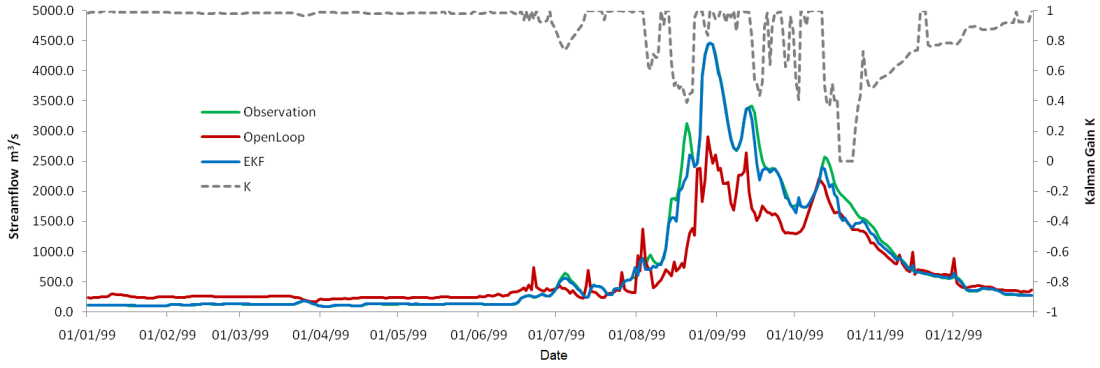


Figure 5-5 1999 streamflow EKF assimilation results compared to observations (Obs) and open-loop (OL)

Because of the scarcity of *in situ* watershed soil moisture measurements, it is challenging to verify the soil moisture assimilation results of EKF. The different update patterns for 1992 and 1999 reveal that EKF assimilation is influenced more by the accumulation of precipitation than the intensity of a single precipitation event. Compared to streamflow, the soil moisture is more sensitive to the watershed precipitation. This implies that the simulation results of specific soil moisture models that use precipitation as inputs may provide more accurate soil moisture estimates (Hopmans et al., 2002; Montzka et al., 2013).

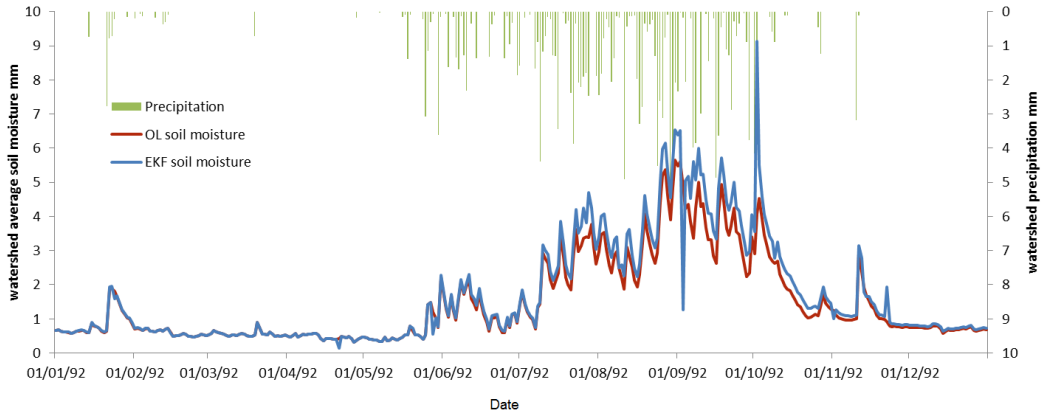


Figure 5-6 Watershed average soil moisture (mm) EKF assimilation results compared to open-loop (OL) for 1992

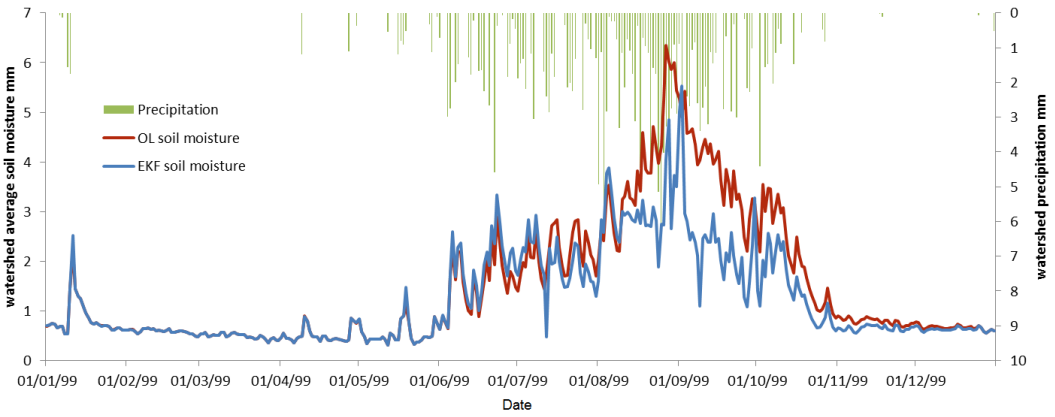


Figure 5-7 Watershed average soil moisture (mm) EKF assimilation results compared to open-loop (OL) for 1999

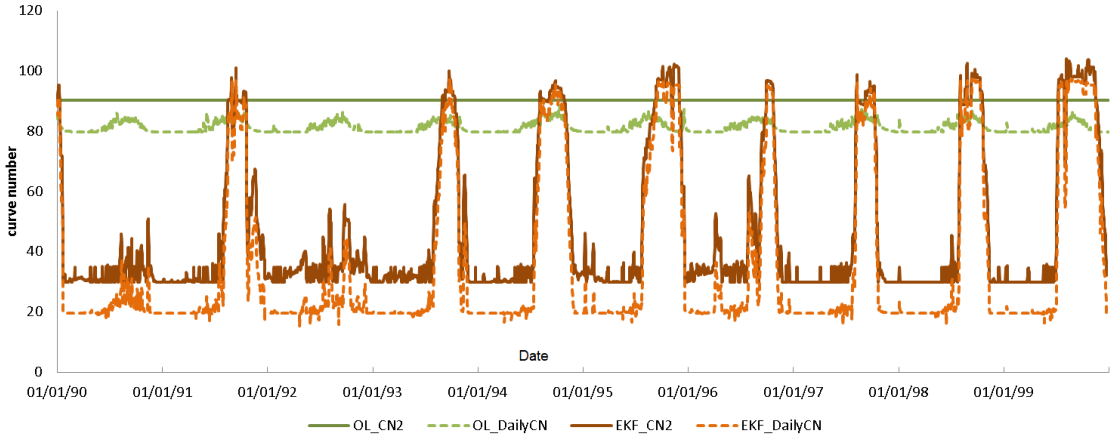


Figure 5-8 CN2 and dailyCN updates for 1992

Figure 5-8 displays the updates for CN2 and dailyCN from 1990 through 1999. DailyCN is the soil moisture adjusted curve number used in the final surface runoff calculation. Both CN2 and dailyCN are directly proportional to the surface runoff generation (i.e., a larger curve number leads to larger runoff, and subsequently increased streamflow). In SWAT, CN2 remains constant unless it is redefined in future crop operation processes, while dailyCN varies with the soil moisture content. The traditional batch calibration technologies assume that the calibrated model is robust enough (at least for the validation period). However, as shown in Figure 5-3, the calibrated model does not perform well for the exceptional flood in 1999.

The EKF updated curve number shows dramatically different patterns from those in the open-loop case. Both CN2 and dailyCN are greatly reduced in the dry season, while they are significantly increased in the flood season. CN2 and dailyCN are reduced during the flood period of 1992; however, they are increased during the flood period of 1999. They both correctly reflect the streamflow update results in Figure 5-4 and Figure 5-5. DailyCN sees more fluctuations than CN2 because the soil moisture is also adjusted by EKF. It is interesting to observe that, although the curve numbers were actually not used in the dry season because of the absence of surface runoff, EKF still 'drags down' the curve numbers to reduce the streamflow overestimation of the SWAT open-loop.

5.5.1.2 Streamflow forecast improvement

In order to verify the improvement in SWAT streamflow *forecasts* due to EKF assimilation, the one- to seven-day SWAT streamflow *forecasts* after using EKF are evaluated. A full

hydrograph is obtained for each lead because EKF is applied continuously. To avoid confusion, the open-loop *forecasts* issued after the application of EKF up to the date the *forecast* is issued are termed 'EKF subsequent open-loop' (EKFsOL). Figure 5-9 and Figure 5-10 show the hydrographs of these EKFsOL results for 1992 and 1999 respectively, and Table 5-3 and Table 5-4 are the NSC and RMSE for the wet season flood simulation for those two years, also respectively. Day1~ Day 7 represents the EKFsOL *forecasts* for the one- to seven-day leads.

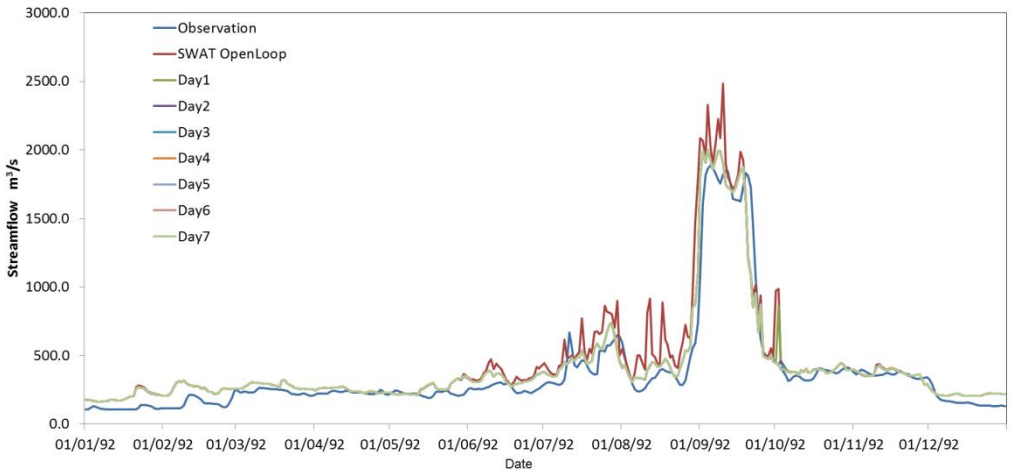


Figure 5-9 EKF subsequent open-loop (EKFsOL) streamflow *forecast* (1992)

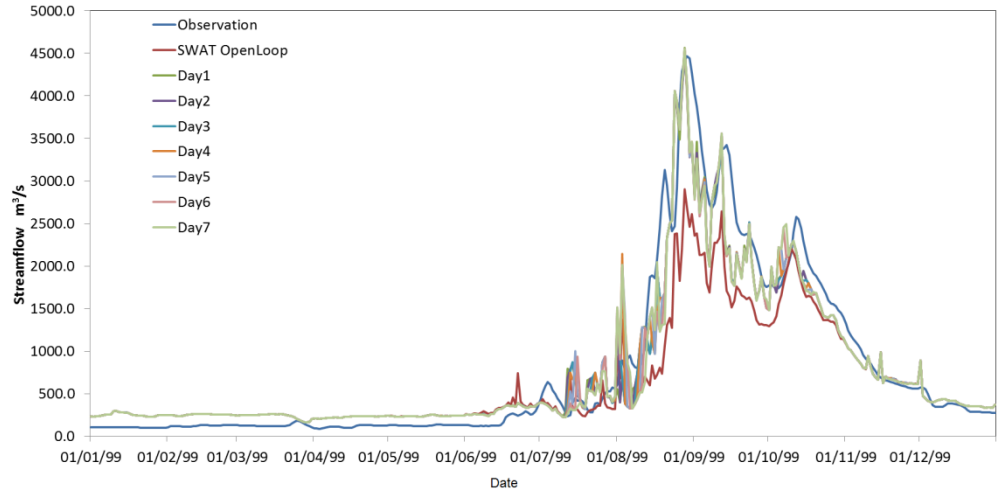


Figure 5-10 EKF subsequent open-loop (EKFsOL) streamflow *forecast* (1999)

The EKFsOL is better than the pure SWAT open-loop (OL) in both years. In 1992, the erroneous fluctuations of OL are mostly eliminated, and the flood is less overestimated in EKFsOL. In the wet season, NSC is improved from 0.77 to 0.91, and RMSE is reduced from

223 m^3/s to less than 140 m^3/s. In 1999, the EKFsOL results are much closer to the observations than are the underestimated results of the OL simulation. For the wet season, NSC is improved from 0.62 to more than 0.84, and RMSE is reduced from 682 m^3/s to less than 440 m^3/s.

Table 5-3 EKF subsequent open-loop (EKFsOL) *forecast* statistics (1992)

	OL	DAY1	DAY2	DAY3	DAY4	DAY5	DAY6	DAY7
NSC	0.77	0.91	0.92	0.92	0.92	0.92	0.93	0.92
RMSE(m^3/s)	223.08	138.60	134.87	134.83	134.83	134.88	134.93	134.96

Table 5-4 EKF subsequent open-loop (EKFsOL) *forecast* statistics (1999)

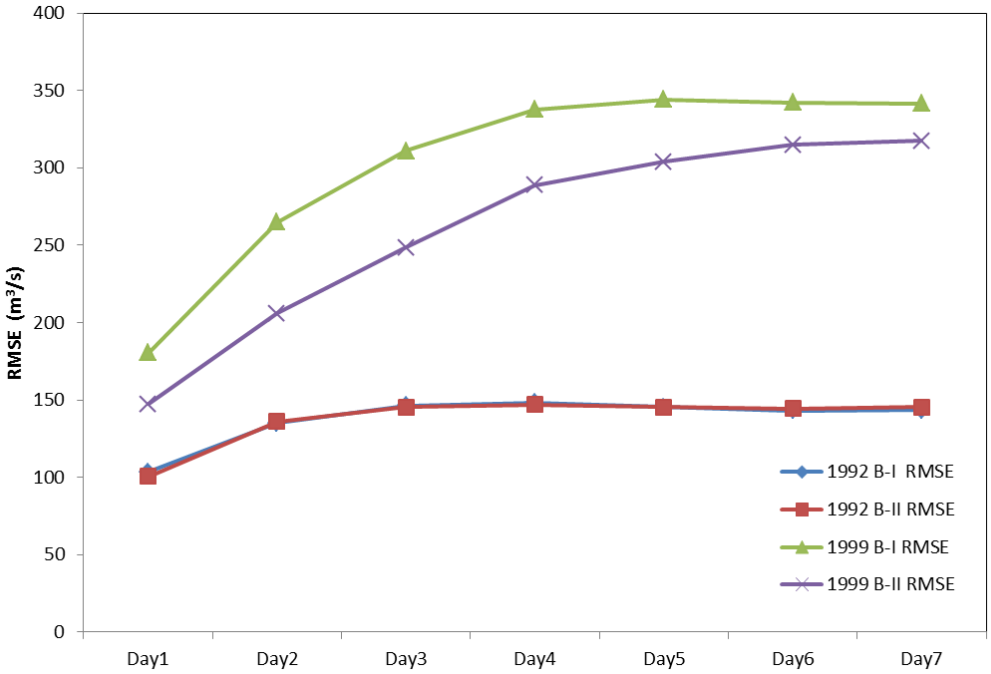
	OL	DAY1	DAY2	DAY3	DAY4	DAY5	DAY6	DAY7
NSC	0.62	0.86	0.85	0.85	0.85	0.84	0.84	0.84
RMSE(m3/s)	681.56	407.46	419.32	420.57	432.02	435.37	437.00	439.21

One interesting finding is that the benefit of EKF seems to be not very sensitive to the length of the lead time. This is especially the case for the wet season of 1992, in which the streamflow is relatively low and SWAT performs reasonably well. EKFsOL in general performs better for shorter leads than longer leads. Again, such a trend is not as significant as one might expect, indicating that the state and parameter updates with EKF are robust.

5.5.2 Output assimilation results

Figure 5-11 compares the wet season performance of the two sub-experiments of Experiment B. For inputs, the single-variable input Experiment B-I only uses the errors from the previous two days, while the multi-variable input Experiment B-II also uses streamflow observations. The accuracy of both sub-experiments decreases rapidly for the first four days as the lead gets longer. No obvious differences between these two sub-experiments were observed for 1992, while B-II significantly outperforms B-I for 1999. The better performance of the combined input experiments agrees with the results from other research (Anctil et al., 2003; Jeevaragagam and Simonovic, 2013).

a)



b)

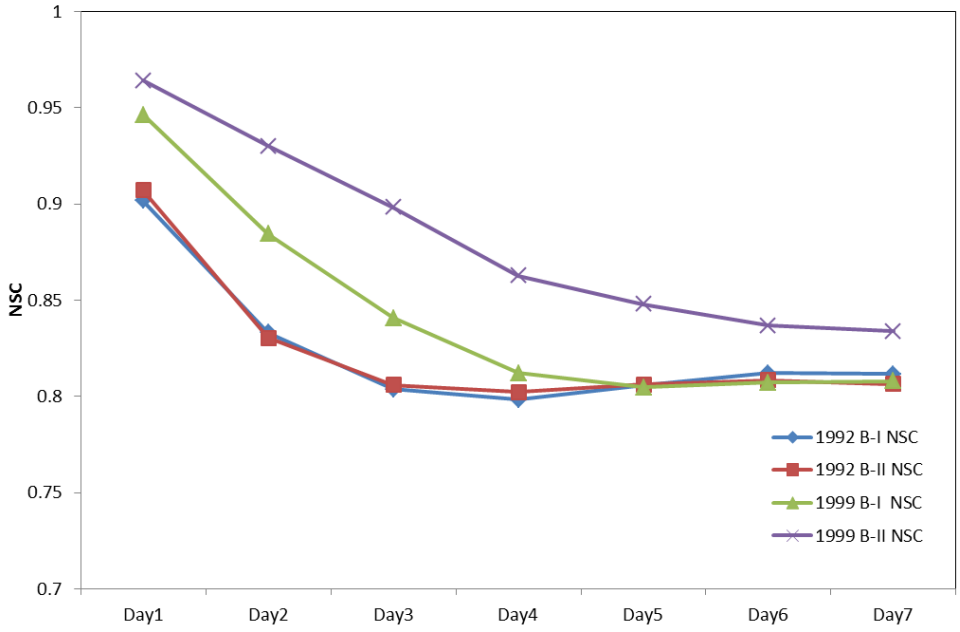


Figure 5-11 Wet season *forecast* performance for the two sub-experiments of Experiment B: a) RMSE; b) NSC

Figure 5-12 and Figure 5-13 show the *forecasts* of sub-experiment B-II, whose RMSE and NSC values are presented in Table 5-5. A major difference between sub-experiment B-II and

Experiment C is that the former predicts the streamflow of the dry season better than the wet season, while the latter does not improve dry season prediction. The dry season streamflow is more stable, which is easier for the AR model to handle. EKF only updates the states and parameters related to surface and lateral runoff, which in the dry season contributes little to the channel streamflow.

For 1992, the AR model provides accurate result for one-day *forecasts*, but the accuracy decreases rapidly for the two-day and longer *forecasts*. As shown in Figure 5-12, the *forecasts* for longer leads have more fluctuations than those from the observations. More importantly, the flood peak is often underestimated. The 1999 *forecasts* are generally very accurate compared to the SWAT open-loop simulations, despite that they are also sensitive to lead time.

Experiment D is designed to combine EKF with the AR model: instead of assimilating the errors between the SWAT open-loop simulation and the observations, as Experiment B does, the errors between the EKFsOL and the observations are assimilated into the model outputs. The idea behind this experiment is that the 'filtered' model errors might be more suitable for the AR model. Note that the EKFsOL have seven time series, but only the one-day *forecast* series is used in Experiment D because it is the most accurate among the seven.

The *forecasts* of Experiment D for 1992 and 1999 are shown in Figure 5-14 and Figure 5-15 respectively, and their RMSE and NSC results are presented in Table 5-6. As shown in Table 5-6, Experiment D significantly outperforms Experiment B for 1992. For example, the one-day wet season *forecast* RMSE and NSC for Experiment B are 100.64 m^3/s and 0.91 respectively, but improve to 67.59 m^3/s and 0.96 for Experiment D; the seven-day *forecast* RMSE and NSC results for Experiment B are 145.29 m^3/s and 0.81 respectively, but improve to 102.68 m^3/s and 0.90 for Experiment D. The *forecasts* for different leads are more clustered, as shown in Figure 5-14, indicating more stability and less sensitivity to lead length. For 1999, a notable improvement compared to Experiment B can only be seen for leads longer than four days. Comparing Figure 5-15 and Figure 5-13, it is also seen that Experiment D produces more clustered *forecasts* than Experiment B.

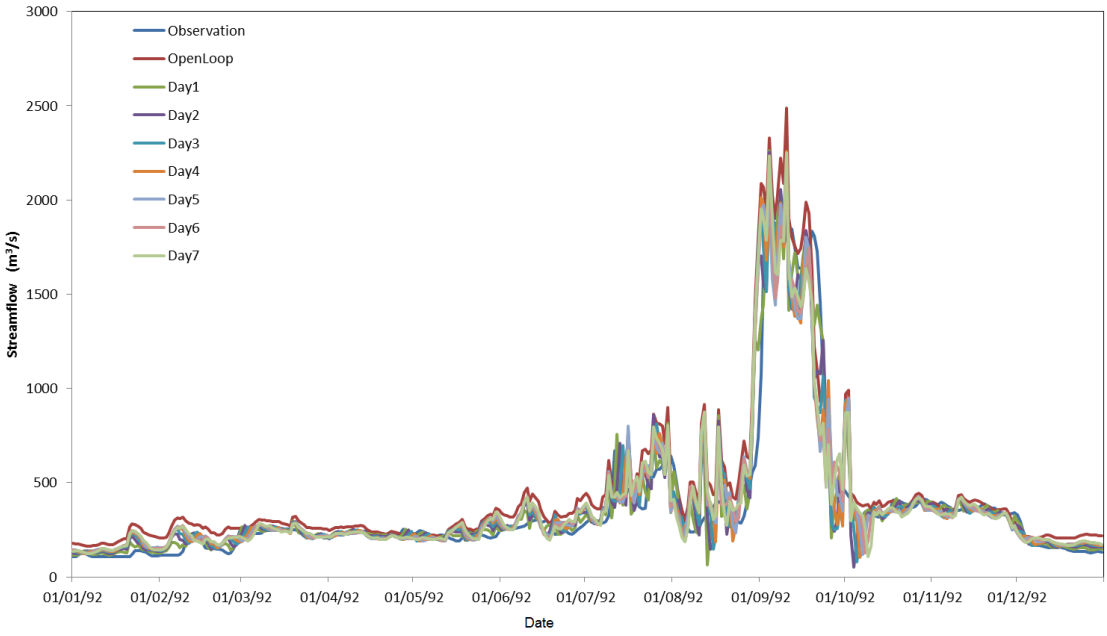


Figure 5-12 The streamflow *forecasts* for different leads of the multi-variable input AR model (Experiment B-II) (1992, full year)

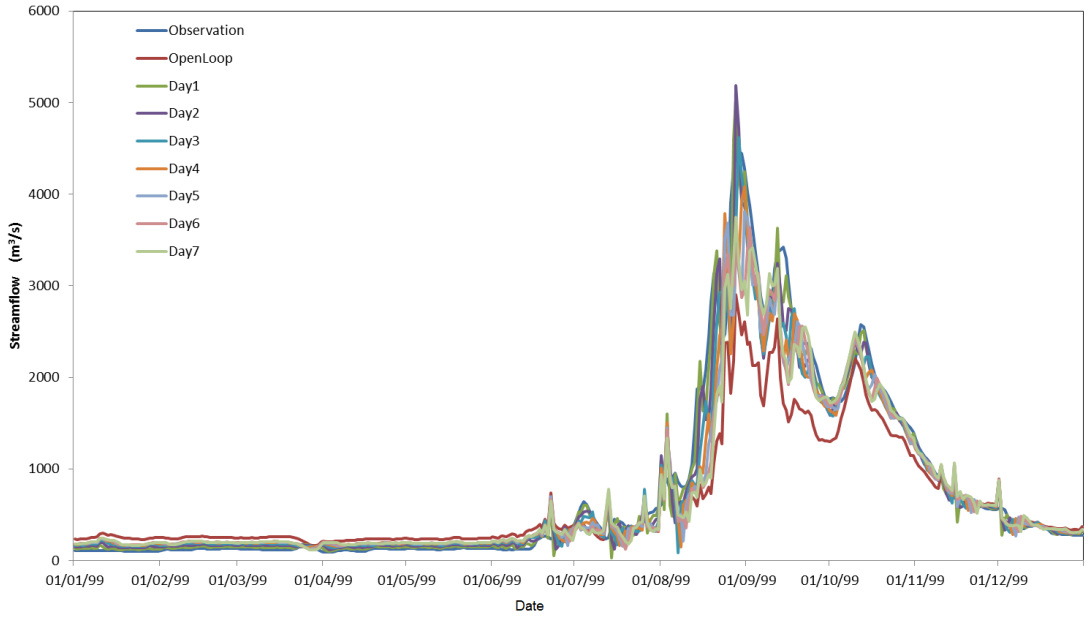


Figure 5-13 The streamflow *forecasts* for different leads of the multi-variable input AR model (Experiment B-II) (1999, full year)

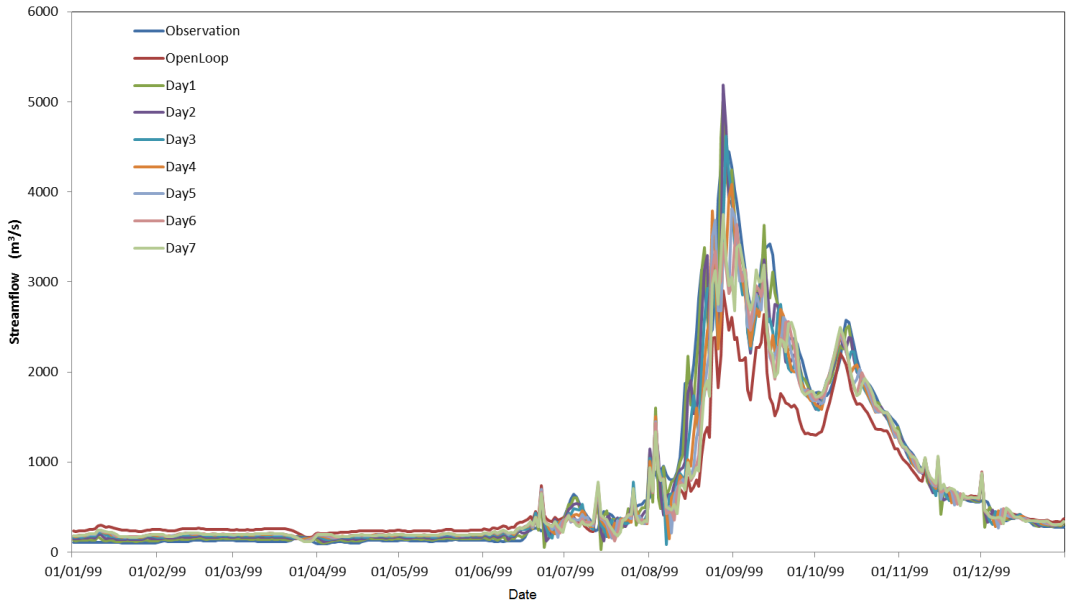


Figure 5-14 The streamflow *forecasts* for different leads of the hybrid model (Experiment D) (1992, full year)

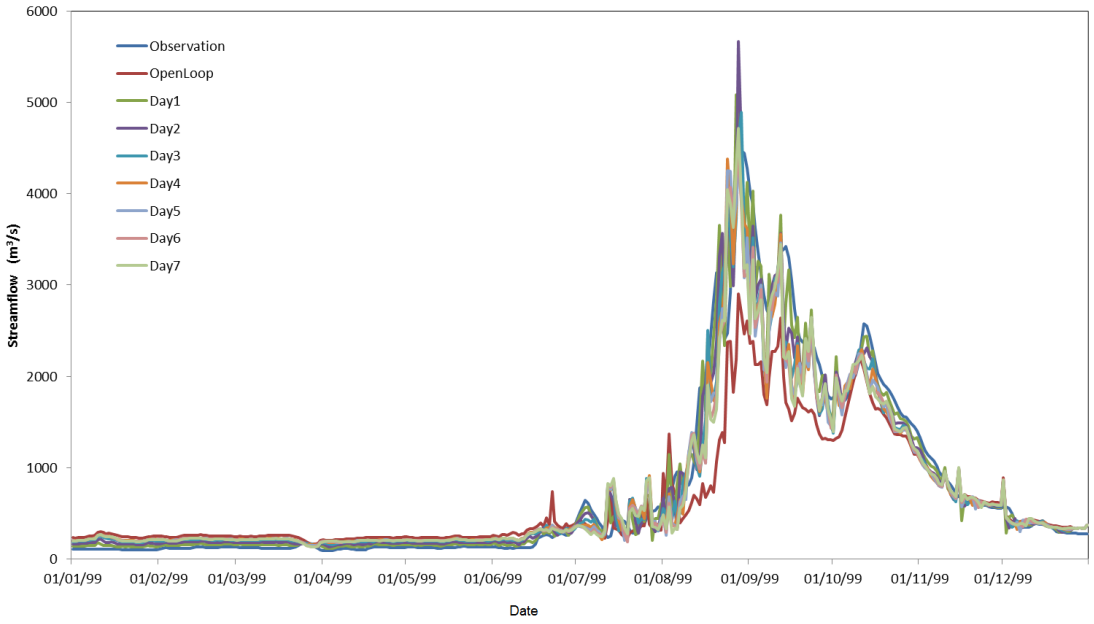


Figure 5-15 The streamflow *forecasts* for different leads of the hybrid model (Experiment D) (1999, full year)

Furthermore, Experiment D provides more accurate forecasts for the flood rise limb and flood peak discharge from early August to early September, while Experiment B captures more features of the flood recession period after mid-September. These are valuable findings if these methods are to be applied in an operational flood *forecasting* and warning system.

Table 5-5 The *forecast* statistics of the multi-variable input AR model (Experiment B-II)

1992		Day1	Day2	Day3	Day4	Day5	Day6	Day7
Full year	RMSE(m^3/s)	101.93	138.41	148.73	150.72	149.66	148.98	149.50
	NSC	0.92	0.86	0.84	0.83	0.84	0.84	0.84
Wet season	RMSE(m^3/s)	100.64	136.02	145.49	146.86	145.41	144.59	145.29
	NSC	0.91	0.83	0.81	0.80	0.81	0.81	0.81
1999		**Day1**	**Day2**	**Day3**	**Day4**	**Day5**	**Day6**	**Day7**
Full year	RMSE(m^3/s)	150.23	209.82	253.44	293.92	309.46	320.30	323.10
	NSC	0.98	0.96	0.94	0.91	0.91	0.90	0.90
Wet season	RMSE(m^3/s)	147.40	206.06	248.51	288.62	304.00	314.71	317.60
	NSC	0.96	0.93	0.90	0.86	0.85	0.84	0.83

Table 5-6 The *forecast* statistics of the hybrid model (Experiment D)

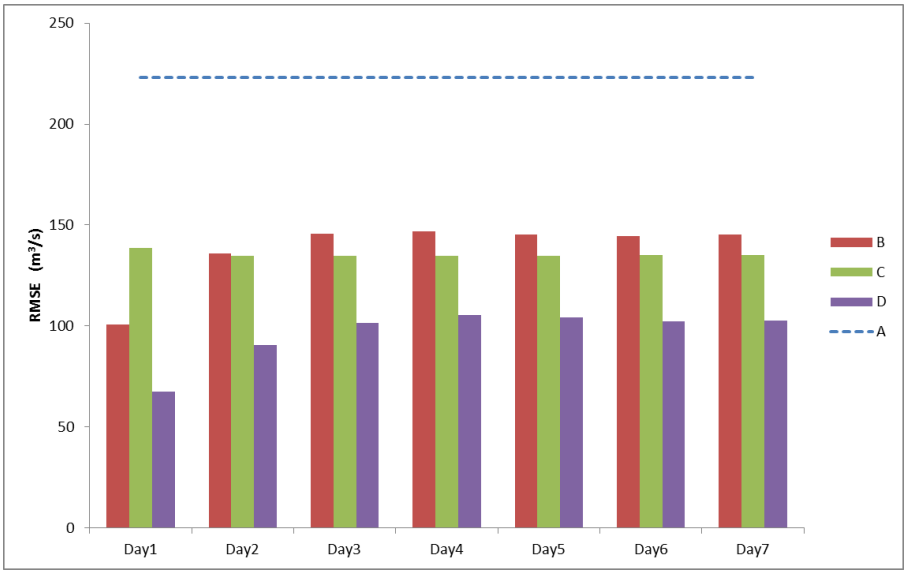
1992		Day1	Day2	Day3	Day4	Day5	Day6	Day7
Full year	RMSE(m^3/s)	69.47	94.19	106.47	111.04	110.18	108.44	108.64
	NSC	0.96	0.93	0.92	0.91	0.91	0.91	0.91
Wet season	RMSE(m^3/s)	67.59	90.39	101.39	105.30	104.25	102.46	102.68
	NSC	0.96	0.93	0.91	0.90	0.90	0.90	0.90
1999		**Day1**	**Day2**	**Day3**	**Day4**	**Day5**	**Day6**	**Day7**
Full year	RMSE(m^3/s)	190.86	244.42	274.72	298.70	303.91	307.49	311.58
	NSC	0.96	0.94	0.93	0.91	0.91	0.91	0.90
Wet season	RMSE(m^3/s)	189.46	240.91	268.55	291.23	296.78	300.48	304.96
	NSC	0.94	0.90	0.88	0.86	0.86	0.85	0.85

5.5.3 Comparison of experiments

The comparison of the wet season *forecasts* of all the experiments is summarized in Figure 5-16. Since the EKF assimilation is not effective for dry season streamflow forecasting, only the wet season results are shown here. Also, the NSC results are not presented because while they correspond to the RMSE results, the patterns in the latter are more obvious.

All three assimilation experiments, B, C and D, significantly improve upon the Experiment A *forecasts* (the open-loop SWAT simulation without any assimilation). However, the "best" assimilation scheme is difficult to determine because the experiments perform differently in different experimental years.

a)



b)

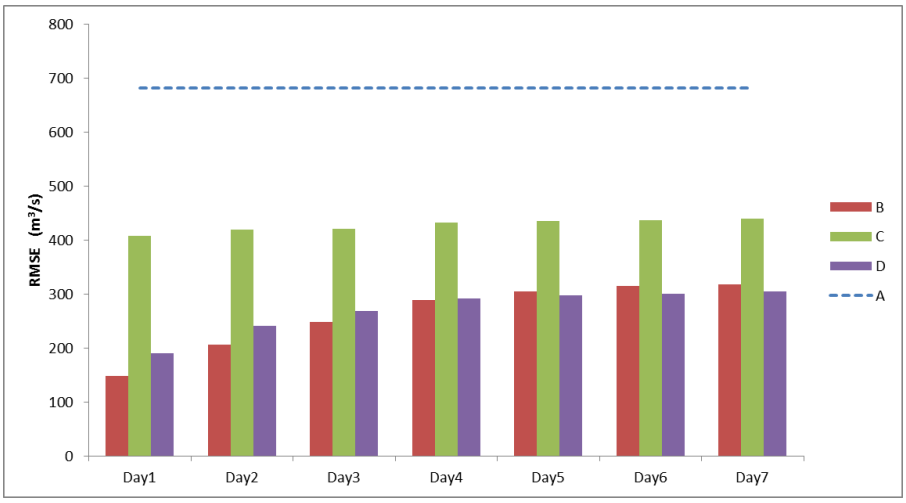


Figure 5-16 Wet season RMSE for all four experiments: a) 1992 wet season RMSE; b) 1999 wet season RMSE

In 1992, although the one-day *forecasting* of Experiment B is more accurate than that of Experiment C, the latter has a marginal advantage over the former for *forecast* times longer than one day. Meanwhile, Experiment D significantly outperforms Experiment B and Experiment C for all *forecast* leads. All experiments except Experiment C are sensitive to lead time. For 1999, Experiment B and Experiment D share similar patterns, while the former has a lower RMSE for shorter leads and the latter a lower RMSE for longer leads. What Figure 5-16 does not show is, as was discussed previously, that the hybrid model Experiment D provides more accurate

forecasts for flood rise and peak periods, while the pure output assimilation B provides more satisfactory *forecasts* for the flood recession period. Both Experiment B and Experiment D outperform Experiment C for all *forecast* leads, especially for shorter leads.

5.6 Conclusions

This study compares the potential improvement in streamflow *forecasts* issued with the SWAT model by using state-parameter assimilation and output assimilation. EKF is used to update the water storage variables and CN2 in the state-parameter assimilation, and AR models are applied to update the estimate errors in the output assimilation. It is found that both methods can improve *forecast* accuracy; however their performance is influenced by the hydrological regime of the year.

The state-parameter assimilation outperforms the output assimilation in the dry year, while the latter outperforms the former in the wet year. The flood rise limb and peak discharge are more accurately *forecast* by state-parameter assimilation, while the recession limb is captured better by output assimilation. Compared to output assimilation, the state-parameter assimilation results are more robust and less sensitive to the length of the lead.

The state-parameter assimilation is capable of updating watershed state variables such as soil moisture and model parameters in addition to streamflow. However, due to the selection of the state vector in this study, such updates are only valid in the wet season, when surface and lateral runoff occurs. In contrast, the output assimilation handles the dry season *forecast* better than the wet season *forecast*.

The traditional state and output hybrid methods mostly focus on the hydrodynamic module of a hydrological model. In this study we combined output assimilation with state-parameter assimilation, which works in the rainfall-runoff part of SWAT. This new hybrid method provides the most accurate *forecasts* in the dry year, and *forecasts* that are similar to state-parameter assimilation *forecasts* but better than output assimilation *forecasts* in the wet year.

Chapter 6 Impact of Uncertainties in Precipitation Forecasts

In the previous chapters, 'perfect forecasts' are generated based on the assumption that the future climate is perfectly known. In this chapter, these *forecasts* will be re-assessed with probabilistic climate prediction. Random errors were added to the actual precipitation, and the results were used to replace the observed precipitation in the previous numerical experiments. Instead of 'perfect forecasts' or forecasts, these forecasts are named 'perturbed *forecasts*'. The combined assimilation of streamflow and surface soil moisture in Experiment G in Chapter 4 is used in this study.

6.1 Methodology

The probabilistic forecasts of daily precipitation can be obtained either with statistical methods (Barnston, 1994; Sittichok et al., 2014; Toth et al., 2000) or directly from Numerical Weather Prediction (NWP) models (Brown et al., 2012; Ebert et al., 2003; Rutherford, 1972). It is critical to evaluate the influence of erroneous precipitation forecasts to hydrological simulation and forecasts of other state variables such streamflow and soil moisture.

For simplicity, of the three types of precipitation forecast error (displacement error, amplitude error and residual error) (Hoffman et al., 1995), only the residual errors are considered in this study. This is reasonable since it is a short-term forecast problem within a small temporal scale (11 months). Under this assumption, an economical precipitation multiplier model is used to quantify the precipitation errors (Vrugt et al., 2008; Salamon and Feyen, 2009; McMillan et al., 2011).

In this method, the precipitation forecast of *t* days after day *i* is perturbed as:

$$p'_{i+t} = p_{i+t} + \gamma_{i+t} p_{i+t} \tag{6-1}$$

where p_{i+t} is the initial precipitation forecast, which is replaced by ground observation in this study; γ_{i+t} is the perturbation factor, which is assumed to follow a Gaussian distribution:

$$\gamma \sim N(0, \sigma) \tag{6-2}$$

where σ is the error variance of γ.

The multiplier method is very flexible and maintains the credibility of the original forecast and the adjustment is 'somehow scale dependent' (Vrugt et al., 2008). Unlike other methods that add increments to the original forecast, the multiplier method may avoid the issue of negative precipitation values.

6.2 Results and Discussion

In order to assess the influence of uncertain precipitation input on the *forecasting* of soil moisture and streamflow, the one- to seven- day *forecasts* with the precipitation perturbed with errors of different scales are studied. Considering the characteristics of the precipitation used, these *forecasts* will be referred as 'perturbed *forecasts*'. The scale of errors is adjusted by quantifying the error variance of the perturbation factor σ from 0.1 to 0.5. The perturbed precipitation is first used in the open-loop models, and then in the updated model with combined assimilation (with state vector as in Experiment G in Chapter 4) in order to simulate soil moisture and watershed outlet streamflow.

Two examples of the perturbation results are displayed in Figure 6-1. In Figure 6-1, 'pcp_obs' is the observed daily precipitation, 'P2' is the perturbed precipitation when $\sigma = 0.2$ and 'P4' is the perturbed precipitation when $\sigma = 0.4$. 'E0.2' represents the generated perturbation factor γ with an error variance of 0.2, and 'E0.4' represents the generated perturbation factor γ with an error variance of 0.4. As shown in Figure 6-1, the perturbed precipitation depends on both the magnitude of the 'true' precipitation (i.e., the observed precipitation) and the perturbation factor γ. A larger perturbation factor γ leads to more discrepancy between the perturbed precipitation and the real precipitation.

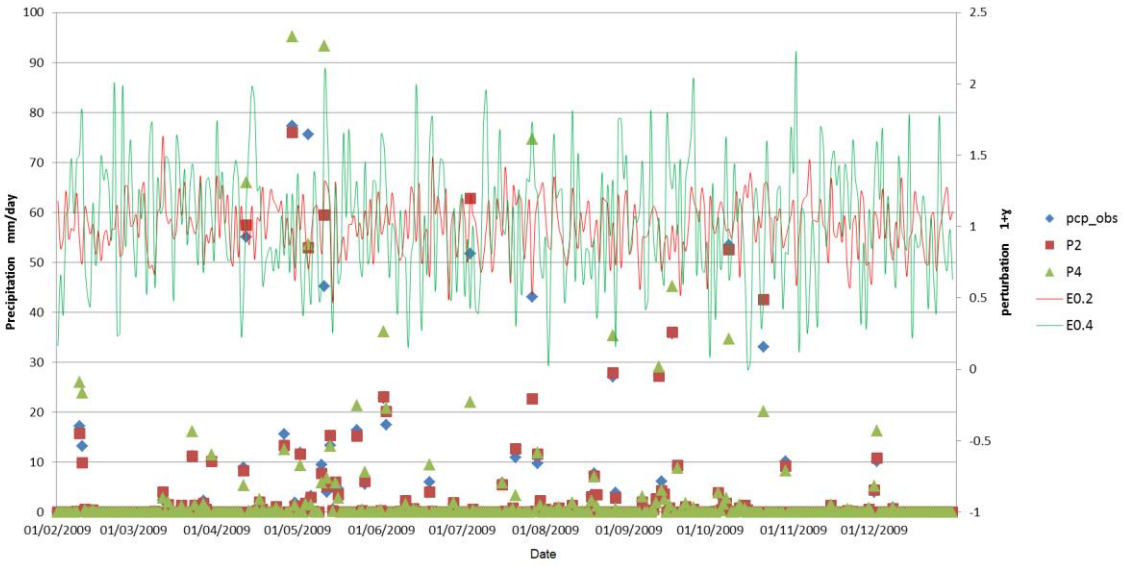


Figure 6-1 The precipitation perturbation results for $\sigma = 0.2$ and $\sigma = 0.4$

Table 6-1 Influence of uncertain precipitation on open-loop SWAT simulation

σ	0	0.1	0.2	0.3	0.4	0.5
RMSE_sm /mm	47.41	47.54	48.22	48.98	48.78	49.66
NSC_sm	-4.45	-4.48	-4.64	-4.82	-4.77	-4.99
RMSE_flow /cms	2.53	2.47	2.38	2.97	2.98	3.27
NSC_flow	-0.14	-0.09	-0.01	-0.57	-0.58	-0.91

The influence of the erroneous precipitation on open-loop SWAT simulations is displayed in Table 6-1. The column with $\sigma = 0$ represents the case in which the precipitation is not perturbed at all. As shown in this table, the perturbed *forecasts* of both the full profile soil moisture (sm) and streamflow (flow) are generally less accurate with larger perturbation errors. The only exception is the streamflow perturbed *forecast* with lower precipitation perturbations. The streamflow perturbed *forecasts* of $\sigma = 0.1$ and $\sigma = 0.2$ are actually slightly better than the case of no perturbation. This is possibly because the perturbed precipitation happens to offset the influence of insufficient calibration of the model on the streamflow simulation.

The accumulated NSC and RMSE for soil moisture and streamflow perturbed *forecasts* are displayed in Figure 6-2~Figure 6-5. In these figures, 'case3_sm' represents the soil moisture perturbed *forecast* from the combined assimilation of surface soil moisture and watershed streamflow, but with unperturbed precipitation input, 'case3_flow' represents the streamflow perturbed *forecast* from combined assimilation with unperturbed precipitation input. 'PE_sm1', 'PE_sm2',…, 'PE_sm5' means the soil moisture perturbed *forecast* from the combined assimilated SWAT model with perturbed precipitation whose perturbation error covariance $\sigma = 0.1, 0.2, \ldots, 0.5$. 'PE_flow1', 'PE_flow2',…, 'PE_flow5' means the streamflow perturbed *forecast* from the combined assimilated SWAT model with perturbed precipitation whose perturbation error covariance $\sigma = 0.1, 0.2, \ldots, 0.5$.

As discussed in the previous chapters, the perturbed *forecasts* with the open-loop SWAT model for all leads have the same accuracy for the same inputs. In this study, it is assumed that the scale of the errors in the precipitation for all leads is the same. This assumption is necessary to maintain the whiteness of the residual errors in operation, although it may not be satisfied as the QPFs tend to be more accurate for shorter perturbed *forecast* leads. For this reason, the open-loop simulation results with perturbed precipitation, if presented in Figure 6-2~Figure 6-5, would be a constant value across all leads that sums up the statistical indicators for the different perturbation error scales in Table 6-1. For example, in Figure 6-1, the accumulated NSC of soil moisture from the open-loop SWAT with perturbed precipitation would be -23.70, which is far below the majority of the accumulated NSC from the assimilated models. Hence, they are normally not presented in these figures for display purposes.

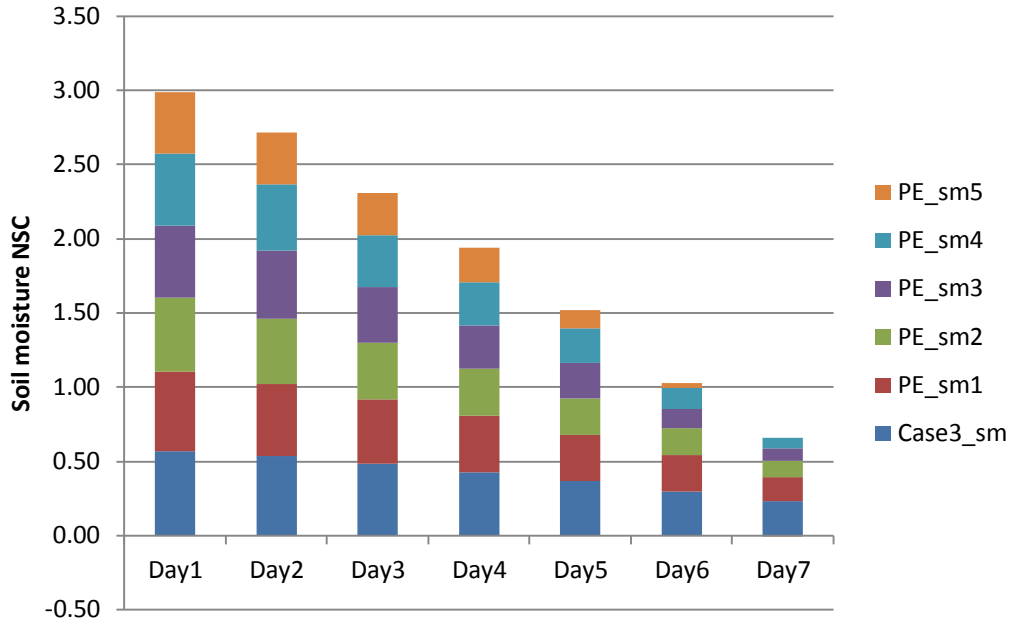


Figure 6-2 The accumulated NSC of soil moisture perturbed *forecasts* for different leads and perturbation scales

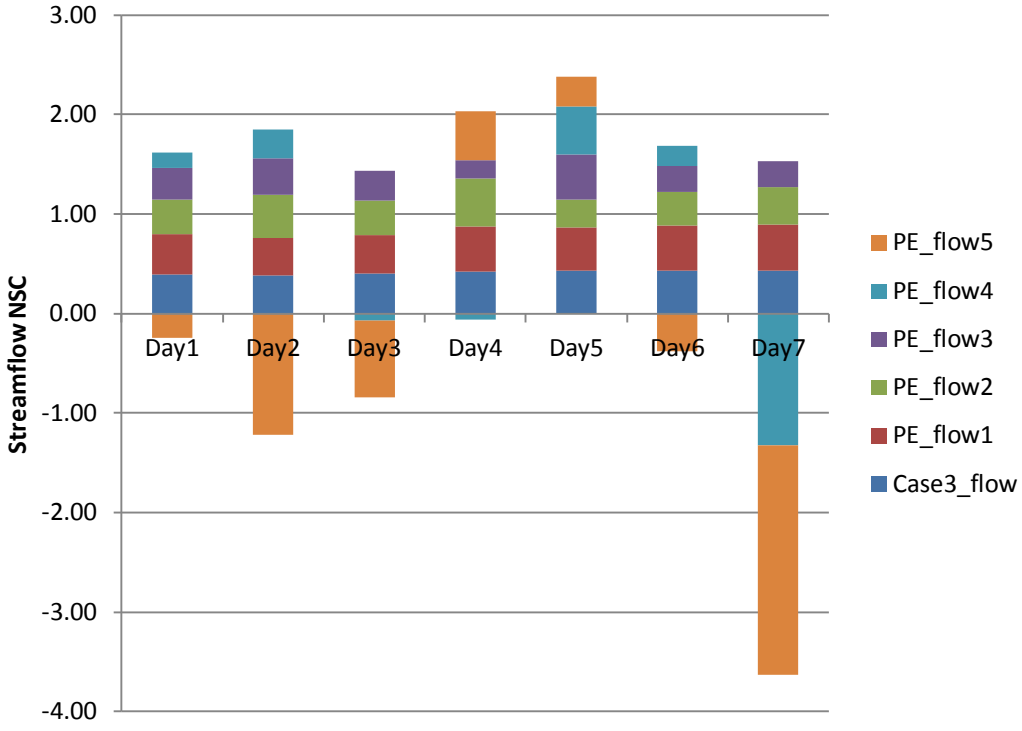


Figure 6-3 The accumulated NSC of streamflow perturbed *forecasts* for different leads and perturbation scales

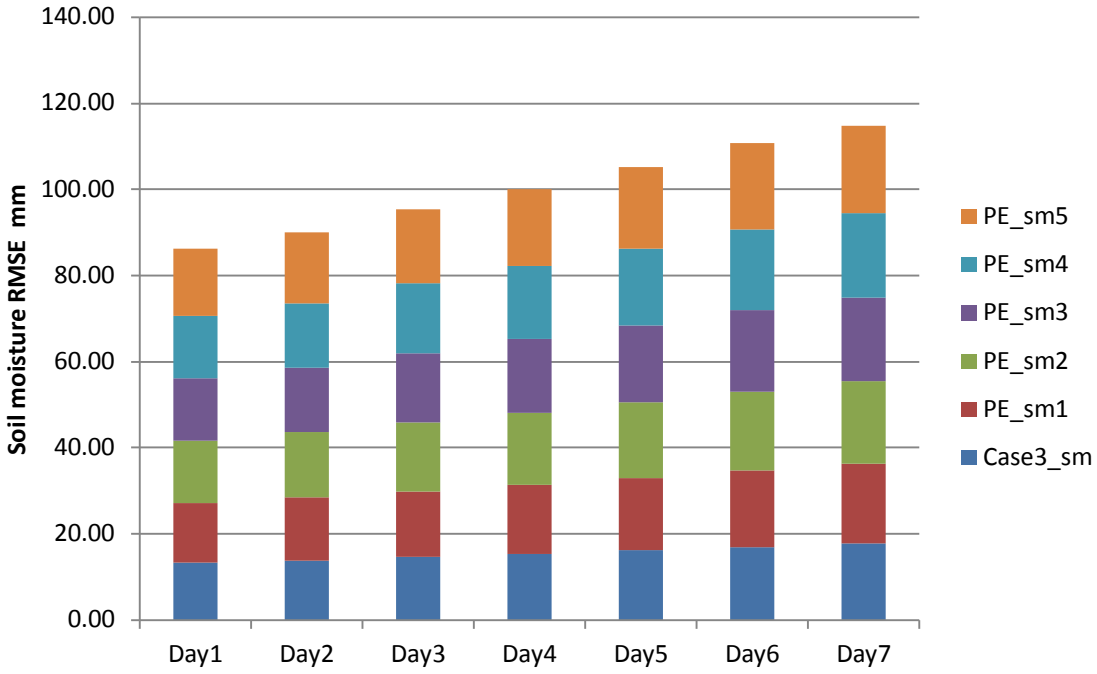


Figure 6-4 The accumulated RMSE of soil moisture perturbed *forecasts* for different leads and perturbation scales

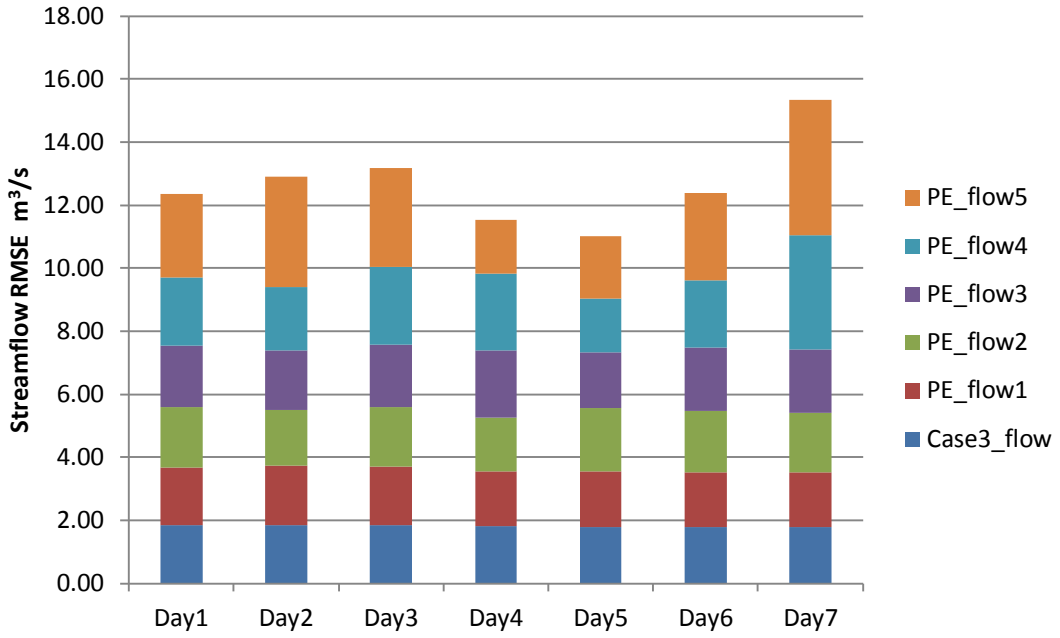


Figure 6-5 The accumulated RMSE of streamflow perturbed *forecasts* for different leads and perturbation scales

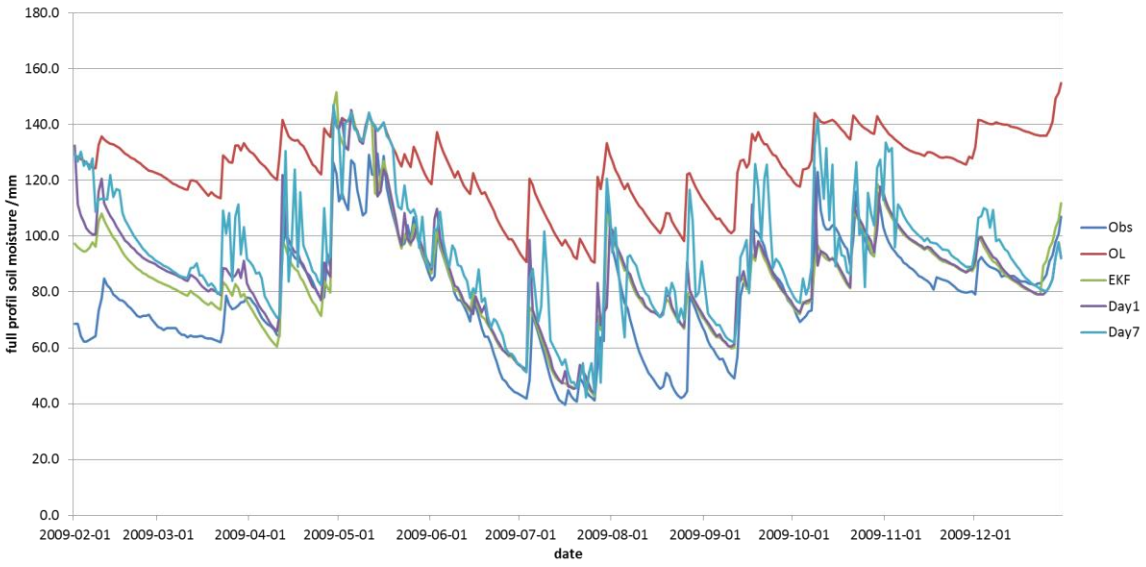


Figure 6-6 Soil moisture perturbed *forecasts* with perturbed precipitation ($\sigma = 0.5$)

The soil moisture perturbed *forecast* with perturbed precipitation of $\sigma = \mathbf{0.5}$ is displayed in Figure 6-6. 'obs', 'OL' and 'E5' are the observed soil moisture, open-loop simulated soil moisture respectively. E5_day1 is the one-day perturbed *forecast* of the assimilated model with perturbed precipitation, and E5_day7 is the seven-day perturbed *forecast* with perturbed precipitation.

For soil moisture perturbed *forecasts*, the benefit of EKF assimilation carries on for all seven leads, even in the case of maximum perturbed precipitation. However, erroneous precipitation tends to degrade the assimilation benefit. The perturbed *forecast* accuracy is inversely proportional to the scale of the error added to the precipitation. For a one-day perturbed *forecast*, the NSC of the assimilated model with 'perfect' precipitation is 0.57; with a precipitation perturbation error covariance of 0.1 it becomes 0.53; and with a perturbation error covariance of 0.5 it decreases to 0.41. The RMSE for the 'perfect' precipitation and perturbed precipitation with σ of 0.1 and 0.5 is 13.29mm, 13.87mm and 15.53mm respectively.

The accuracy of perturbed *forecasts* also decreases with increases in the length of the lead, and decreases faster with larger precipitation errors. The NSC of the seven-day perturbed *forecast* with 'perfect' precipitation is 0.23; with a precipitation perturbation error covariance of 0.1 it is reduced to 0.16; and with a perturbation error covariance of 0.5 it is worsened to 0.00. It can be noted that this is not the worst case, as the NSC for the open-loop model without assimilation is

well below zero. The RMSE for the 'perfect' precipitation and perturbed precipitation with $\boldsymbol{\sigma}$ of 0.1 and 0.5 are 17.80mm, 18.55mm and 20.34mm respectively. The decrease rate of NSC for perfect precipitation is 0.04/day for 'perfect' precipitation, 0.05/day for an error covariance of 0.1, and 0.06/day for an error covariance of 0.5.

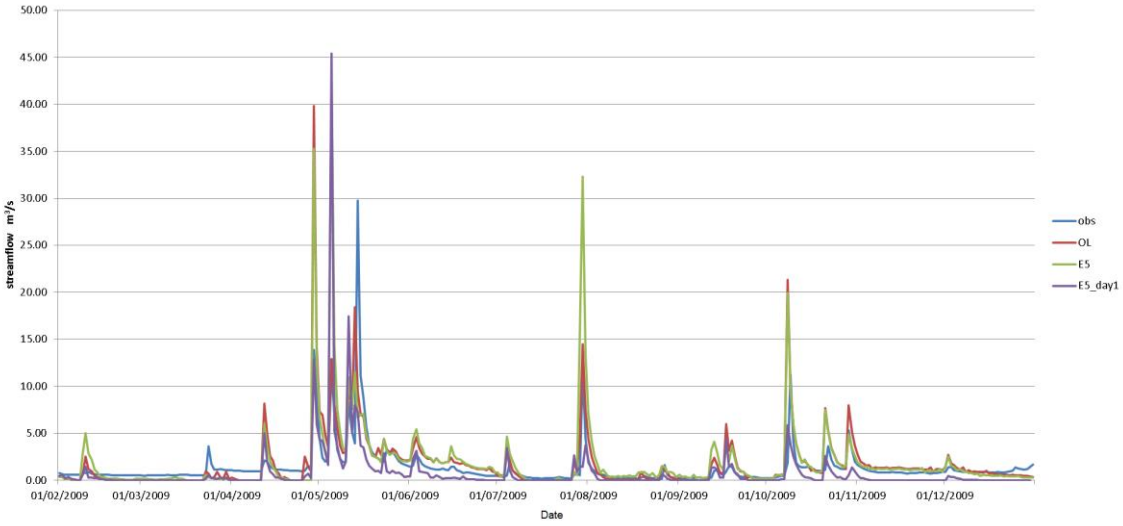


Figure 6-7 Streamflow perturbed *forecast* with perturbed precipitation ($\sigma = 0.5$)

Compared to soil moisture perturbed *forecasting*, streamflow perturbed *forecasting* is more significantly influenced by uncertainties in precipitation. Larger errors ($\sigma = 0.4, 0.5$) added to precipitation can cause significant errors in streamflow perturbed *forecasts*, which cannot be corrected by assimilating surface soil moisture and watershed output streamflow. However, similar to the case of 'perfect' precipitation, streamflow perturbed *forecasts* with perturbed precipitation are not sensitive to longer perturbed *forecast* leads. In other words, perturbed *forecast* accuracy does not necessarily decrease with longer leads. For example, with a perturbation of $\sigma = 0.5$, the NSC for a one-day perturbed *forecast* is -0.25 but increases to 0.45 for a four-day perturbed *forecast*, and drops to -2.31 for a seven-day perturbed *forecast*. This implies that for longer leads, streamflow perturbed *forecasts* are very unstable with perturbed precipitation, either with or without data assimilation.

Figure 6-7 shows the streamflow perturbed *forecast* with perturbed precipitation of $\sigma = 0.5$. 'E5' is the perturbed *forecast* from open-loop SWAT with perturbed precipitation, and 'E5_day1' is the one-day perturbed *forecast* from the assimilated model with perturbed precipitation.

Although the assimilated model manages to mitigate the overestimations of open-loop for some flood events (e.g., 01/08/2009, 15/10/2009), it creates new overestimates (e.g., 08/05/2009). Therefore, it fails to improve on the streamflow perturbed *forecasts* of open-loop SWAT with perturbed precipitation.

6.3 Conclusions

In this chapter, the uncertainties in the quantitative precipitation forecasts (QPFs) are considered explicitly in the perturbed *forecast* mode of EKF assimilation. The precipitation forecasts are perturbed with a multiplier model whose parameter σ is extensively assessed in the soil moisture and streamflow perturbed *forecast* experiments with a combined assimilation of surface soil moisture and watershed streamflow.

The EKF assimilated model issues better perturbed *forecasts* of soil moisture and streamflow than open-loop models, despite that the perturbed precipitation generally worsens the perturbed *forecasts* in both open-loop and assimilated models. However, soil moisture and streamflow react differently to such perturbations. It is found that the soil moisture perturbed *forecasts* can still be improved with combined assimilation, even with the maximum perturbation for the longest *forecast* lead. Perturbed *forecasts* are less accurate with longer leads, even more so than the case of unperturbed precipitation. Compared to soil moisture, a streamflow perturbed *forecast* is more vulnerable to precipitation perturbation. The combined assimilation will not necessarily improve the streamflow perturbed *forecast* of the model when the precipitation perturbation is very large and the *forecast* lead is very long.

Chapter 7 Conclusions and Recommendations

7.1 Summary and Conclusions

In this study, data assimilation of streamflow and soil moisture is implemented in a semi-distributed model (SWAT: Soil and Water Assessment Tool) with EKF. Based on the assumption of a predefined meteorological forecast, EKF state-parameter assimilation was assessed and compared with AR output assimilation in a 'perfect' flood *forecasting* scenario. The robustness of EKF assimilation is examined by perturbing the precipitation forecast to generate perturbed *forecasts* of soil moisture and streamflow.

A ratio method was developed and used to reduce the computational burden by updating the HRU-scale variables offline with the a posteriori/a priori correction ratio of their watershed scale counterparts, which are then updated with EKF online. SWAT is treated as a 'black-box model' in the numerical solution of the Jacobian matrix in which each element is calculated with a partial differential equation.

In the streamflow assimilation experiments, the measured streamflow at the Bakel Station is assimilated into a SWAT model of the Senegal watershed in western Africa. Twin experiments were adopted to assess the effectiveness of EKF in updating the model state and parameters and in improving streamflow *forecasts*.

EKF led to significant updates in both soil moisture and CN2 for the wet season but not for the dry season. In the wet season, soil moisture first increases and then declines, while the CN2 first declines and then increases. Such an opposition is also found in the responses of streamflow and soil moisture to the perturbation of soil moisture.

In a case study of a flood *forecast*, the effectiveness of EKF is verified with both the EKFsOL and quasi-error update *forecast* schemes for up to seven days. The quasi-error update *forecast* generally outperforms EKFsOL, but has less robustness. The longer the lead, the less capable EKFsOL is in capturing the peak flow. EKFsOL performs better in the flood rise period, while the quasi-error *forecast* performs better in the recession period.

In the soil moisture assimilation experiments, observed *in situ* surface soil moisture was assimilated in an attempt to improve the simulation and *forecasting* of soil moisture in deeper

layers, as well as the streamflow. The Little Washita Experimental Watershed was chosen as the case study area. EKF assimilation schemes with different combinations of state and observation variables were evaluated based on their capability to improve the estimate of other variables of interest. The outputs from the assimilated model are compared to those from the open-loop and direct-insert models. The robustness of the assimilated model is further verified in the *forecast* mode. This study also discussed the sensitivity of model error quantification, especially the error covariance between different state variables.

The EKF assimilation of the surface soil moisture is proved to be effective in improving soil moisture and streamflow simulation. However, its effectiveness is influenced by the structure of the state and observation vectors. The state vector with soil moisture in all layers plus streamflow outperforms other, more complicated state vectors, including those augmented with intermediate variables and model parameters. The combined assimilation of surface soil moisture and streamflow significantly improves the estimate of the soil moisture in the full soil profile, while the assimilation of streamflow alone only provides very limited improvement in the soil moisture simulation.

After being updated with EKF, the SWAT model is robust enough to result in better short-term *forecasts* of soil moisture and streamflow. Soil moisture *forecasts* benefit most from the combined assimilation and least so from streamflow assimilation. Such benefit tends to wane with increases of the *forecast* lead. The streamflow *forecast* appears sensitive neither to the type of assimilated observations nor to the increase of the *forecast* lead.

The performance of EKF assimilation is directly related to the accuracy of the model error quantification. The estimate of the soil moisture in a specific layer is more influenced by the error covariance between streamflow and the soil moisture in this layer, while it is not sensitive to the error covariance between streamflow and soil moisture in other layers. The application of an adaptive EKF shows no advantages over using the trial and error method.

As a spin-off study, EKF state-parameter assimilation and AR output assimilation are compared. The EKF is used to update the water storage variables and CN2 in the state-parameter assimilation, and the AR models are applied to update the estimate errors in the output assimilation. It is found that both methods can improve *forecast* accuracy, but their performance

is influenced by the hydrological regime of the year. The state-parameter assimilation outperforms the output assimilation in a dry year, while the output assimilation outperforms the state-parameter assimilation in a wet year. The flood rise limb and peak discharge is more accurately *forecast* by state-parameter assimilation, while the recession limb is captured better by output assimilation. Compared to output assimilation, the state-parameter assimilation results are more robust and less sensitive to the length of the lead.

State-parameter assimilation is capable of updating watershed state variables such as soil moisture and model parameters in addition to streamflow. With the selected state vector in this study, such updates are only valid in the wet season, when surface and lateral runoff occurs. In contrast, the output assimilation handles the dry season *forecast* better than that of the wet season.

The output assimilation is combined with the state-parameter assimilation, which works on the rainfall-runoff part of SWAT. This new hybrid method provides the most accurate *forecasts* in the dry year and *forecasts* that are similar to state-parameter assimilation but better than output assimilation in the wet year.

One of the novelties of this study is the validation of EKF assimilation in the *forecasts* of streamflow and soil moisture. The assumption behind this method is that the future climate is perfectly known. If not, it is important to examine whether the EKF assimilation can still improve *forecasts* of soil moisture and streamflow. In the last part of this thesis the errors in the quantitative precipitation forecasts (QPFs) are considered explicitly when they are used as inputs for the SWAT model. The precipitation forecasts are perturbed with a multiplier model which is extensively assessed with the combined assimilation of surface soil moisture and streamflow.

Perturbed precipitation generally worsens the *forecasting* of soil moisture and streamflow in both open-loop and assimilated models. However soil moisture and streamflow reacts differently to such perturbations. It is found that perturbed *forecasts* of soil moisture can still be improved, even with a maximum perturbation error and longest lead time. Perturbed *forecasts* of streamflow are more vulnerable to the erroneous precipitation forecasts.

In conclusion, EKF is a promising DA method to improve the simulation and *forecasting* ability of a distributed model such as SWAT. The performance of EKF assimilation is influenced by the setting of both state and observation vectors as well as the quantification of model errors. EKF-assimilated SWAT could result in improved streamflow and soil moisture *forecasts*, even with inaccurate precipitation inputs. The combination of state-parameter assimilation and output assimilation has great potential in improving streamflow *forecasting*.

7.2 Recommendations

Although this study has successfully coupled EKF with SWAT, some issues remain, and further work is needed in several areas:

1. Quantification of model and observation errors (Q and R): In this study, the model error is determined using the trial and error method, and the observation error is determined with an empirical method. Ideally, the model error should be quantified during the development of the rainfall runoff model, and the observation error should be quantified in the process of taking measurements. Adaptive filtering seems to be an alternative; however, a case study conducted in this thesis shows that the adaptive filtering is not superior to the conventional trial and error method.
2. Comparison of EKF and EnKF: As pointed out in the literature review, EnKF is probably the most popular hydrological DA method despite a potentially large computational burden and the difficulty of determining the ensemble size. It would be interesting to compare EKF with EnKF when they are both used in a distributed model such as SWAT.
3. Assimilation of remote sensing soil moisture: With the method based on assimilating *in situ* surface soil moisture, it is desirable to apply it to other watersheds without *in situ* soil moisture. Remote-sensing satellites provide more accessible soil moisture products, but widespread application of them is still to come.
4. The update of precipitation input: The update of precipitation by assimilating other observations such as soil moisture is attractive, because rainfall runoff models are usually more sensitive to input than it is to the initial state.

References

Abaza, M., Anctil, F., Fortin, V., Turcotte, R., 2014a. Sequential streamflow assimilation for short-term hydrological ensemble forecasting. Journal of Hydrology, 519: 2692-2706.

Abaza, M., Garneau, C., Anctil, F., 2014b. Comparison of Sequential and Variational Streamflow Assimilation Techniques for Short-Term Hydrological Forecasting. Journal of Hydrologic Engineering, 20(2): 04014042.

Abbaspour, K.C., 2008. SWAT-CUP2: SWAT Calibration and Uncertainty Programs–A User Manual. Duebendorf, Switzerland: Swiss Federal Institute of Aquatic Science and Technology (Eawag), Department of Systems Analysis, Integrated Assessment, and Modeling (SIAM).

Agboma, C. and Lye, L. 2014. Hydrologic Memory Patterns Assessment over a Drought-Prone Canadian Prairies Catchment. Journal of Hydrologic Engineering, 20(7), 04014084.

Ahsan, M., Oconnor, K.M., 1994. A Reappraisal of the Kalman Filtering Technique, as Applied in River Flow Forecasting. Journal of Hydrology, 161(1-4): 197-226. DOI:10.1016/0022-1694(94)90129-5

Ajami, N. K., Q. Duan, and S. Sorooshian, 2007. An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, Water Resources Research, 43(1).

Allen, P.B., Naney, J.W., 1991. Hydrology of the Little Washita river watershed, Oklahoma: Data and analyses. ARS-US Department of Agriculture, Agricultural Research Service (USA).

Alvarez-Garreton, C., Ryu, D., et al., 2013. Impact of observation error structure on satellite soil moisture assimilation into a rainfall–runoff model. In: MODSIM, 20th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, pp. 3071–3077.

Alvarez-Garreton, C., et al. 2014. The impacts of assimilating satellite soil moisture into a rainfall–runoff model in a semi-arid catchment. Journal of hydrology, 519, 2763-2774.

Alvarez-Garreton, C. et al., 2015. Improving operational flood ensemble prediction by the assimilation of satellite soil moisture: comparison between lumped and semi-distributed schemes. Hydrology and Earth System Sciences, 19(4): 1659-1676.

Al-Yaari, A. et al., 2014. Global-scale evaluation of two satellite-based passive microwave soil moisture datasets (SMOS and AMSR-E) with respect to Land Data Assimilation System estimates. Remote Sensing of Environment, 149: 181-195.

Anctil, F., Perrin, C. and Andreassian, V., 2003. ANN output updating of lumped conceptual rainfall/runoff forecasting models. Journal of the American Water Resources Association,39(5), 1269–1279..

Anctil, F., C. Perrin, and V. Andréassian, 2004. Impact of the length of observed records on the performance of ANN and of conceptual parsimonious rainfall-runoff forecasting models, Environmental Modelling & Software, 19(4), 357-368.

Anderson, J. L. 2001. An ensemble adjustment Kalman filter for data assimilation. Monthly Weather Review, 129(12), 2884-2903.

Andreadis, K. M. and Lettenmaier, D. P. 2006. Assimilating remotely sensed snow observations into a macroscale hydrology model. Advances in Water Resources, 29(6), 872-886.

Angela, L., 1982. Combination of a conceptual model and an autoregressive error model for improving short time forecasting, Nordic Hydrology, 13(4), 233-246.

Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams, 1998. Large area hydrologic modeling and assessment part I: Model development1, edited, pp. 73-89.

Arnold, J.G. et al., 2012. Swat: Model Use, Calibration, and Validation. Transactions of the Asabe, 55(4): 1491-1508.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment part I: Model development1, pp. 73-89. DOI:10.1111/j.1752-1688.1998.tb05961.x

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), 145-161.

Barnes, S. L. 1964. A technique for maximizing details in numerical weather map analysis. Journal of Applied Meteorology, 3(4), 396-409.

Barnston, A.G., 1994. Linear statistical short-term climate predictive skill in the Northern Hemisphere. Journal of Climate, 7(10): 1513-1564.

Bauer-Gottwein, P., I. Jensen, R. Guzinski, G. Bredtoft, S. Hansen, and C. Michailovsky, 2015. Operational river discharge forecasting in poorly gauged basins: the Kavango River basin case study, Hydrology and Earth System Sciences, 19(3), 1469-1485.

Berg, A. A., and K. A. Mulroy, 2006. Streamflow predictability in the Saskatchewan/Nelson River basin given macroscale estimates of the initial soil moisture status, Hydrological sciences journal, 51(4), 642-654.

Berger, K. P., and D. Entekhabi, 2001. Basin hydrologic response relations to distributed physiographic descriptors and climate, Journal of Hydrology, 247(3), 169-182.

Bergman, M. and Delleur, J. 1985. Kalman filter estimation and prediction of daily stream flows : I ,Review, algorithm and simulation experiments. Water Resources Bulletin, 21(5), 827-832.

Bergström, S., 1991. Principles and confidence in hydrological modelling. Nord. Hydrol., 22(2): 123-136.

Berthet, L., V. Andréassian, C. Perrin, and P. Javelle, 2009. How crucial is it to account for the antecedent moisture conditions in flood forecasting? Comparison of event-based and continuous approaches on 178 catchments, Hydrology and Earth System Sciences Discussions(13), p. 819 - p. 831.

Bertino, L., Evensen, G. and Wackernagel, H. 2003. Sequential data assimilation techniques in oceanography. International Statistical Review, 71(2), 223-241.

Betts, A. K., et al. 2003. Intercomparison of water and energy budgets for five mississippi subbasins

between ecmwf reanalysis (era - 40) and nasa data assimilation office fvgcm for 1990–1999. Journal of Geophysical Research: Atmospheres (1984–2012), 108(D16).

Beven, K., 1989. Changing ideas in hydrology—the case of physically-based models, Journal of hydrology, 105(1), 157-172.

Beven, K. et al., 1995. Topmodel. Computer models of watershed hydrology.: 627-668.

Bidwell, V., and G. Griffiths,1994. Adaptive flood forecasting: an application to the Waimakariri River, Journal of hydrology. New Zealand, 32(2), 1-15.

Bin, W. and Ying, Z. 2005. A new data assimilation approach. Acta Meteorologica Sinica, 63(5):694-701.

Bishop, C. H., Etherton, B. J. and Majumdar, S. J. 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. Monthly Weather Review, 129(3), 420-436.

Bøgh, E., et al. 2004. Incorporating remote sensing data in physically based distributed agro-hydrological modelling. Journal of hydrology, 287(1), 279-299.

Bolzern, P., Ferrario, M. and Fronza, G. 1980. Adaptive real-time forecast of river flow-rates from rainfall data. Journal of hydrology, 47(3), 251-267.

Borup, M., 2014. Real Time Updating in Distributed Urban Rainfall Runoff Modelling. Thesis (PhD). Technical University of Denmark.

Borup, M., et al. 2015. A partial ensemble Kalman filtering approach to enable use of range limited observations. Stochastic Environmental Research and Risk Assessment, 29(1), 119-129.

Boulet, G., et al. 2000. A simple water and energy balance model designed for regionalization and remote sensing data utilization. Agricultural and Forest Meteorology, 105(1), 117-132.

Box, G. E., and G. M. Jenkins, 1976. Time series analysis: Forecasting and control, in Holden-Day series in time series analysis, edited, Holden-Day.

Brocca, L. et al., 2011. Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe. Remote Sensing of Environment, 115(12): 3390-3408.

Brocca, L. et al., 2012. Assimilation of surface-and root-zone ASCAT soil moisture products into rainfall–runoff modeling. Geoscience and Remote Sensing, IEEE Transactions on, 50(7): 2542-2555.

Brocca, L., F. Melone, T. Moramarco, and V. Singh, 2009. Assimilation of observed soil moisture data in storm rainfall-runoff modeling, Journal of Hydrologic Engineering, 14(2), 153-165.

Broersen, P. M. 2007. Error correction of rainfall-runoff models with the ARMAsel program. Instrumentation and Measurement, IEEE Transactions on, 56(6), 2212-2219.

Broersen, P. M., and A. H. Weerts, 2005. Automatic error correction of rainfall-runoff models in flood forecasting systems, paper presented at Instrumentation and Measurement Technology Conference, 2005. IMTC 2005. Proceedings of the IEEE, IEEE.

Brown, J.D., Seo, D.-J., Du, J., 2012. Verification of Precipitation Forecasts from NCEP's Short-Range

Ensemble Forecast (SREF) System with Reference to Ensemble Streamflow Prediction Using Lumped Hydrologic Models. Journal of Hydrometeorology, 13(3): 808-836.

Burgers, G., Jan van Leeuwen, P. and Evensen, G. 1998. Analysis scheme in the ensemble Kalman filter. Monthly Weather Review, 126(6), 1719-1724.

Cammalleri, C. and Ciraolo, G. 2012. State and parameter update in a coupled energy/hydrologic balance model using ensemble Kalman filtering. Journal of hydrology, 416, 171-181.

Chang, R.L. and Pavlidis, T., 1977. Fuzzy decision tree algorithms. IEEE Transactions on systems, Man, and cybernetics, 1(7), pp.28-35.

Chang, S.-Y. and Latif, S. M. I. 2009. Extended Kalman filtering to improve the accuracy of a subsurface contaminant transport model. Journal of Environmental Engineering, 136(5), 466-474.

Chen, F., Crow, W. T. and Ryu, D. 2014. Dual forcing and state correction via soil moisture assimilation for improved rainfall–runoff modeling. Journal of hydrometeorology, 15(5), 1832-1848.

Chen, F., Crow, W.T., Starks, P.J., 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.

Chen, H., et al. 2013. Hydrological data assimilation with the Ensemble Square-Root-Filter: Use of streamflow observations to update model states for real-time flash flood forecasting. Advances in Water Resources, 59, 209-220.

Chen, L., et al. 2015. Real-time error correction method combined with combination flood forecasting technique for improving the accuracy of flood forecasting. Journal of hydrology, 521, 157-169.

Chen, L., Y. Zhang, J. Zhou, V. P. Singh, S. Guo, and J. Zhang, 2015. Real-time error correction method combined with combination flood forecasting technique for improving the accuracy of flood forecasting, Journal of Hydrology, 521, 157-169.

Chen, W., Huang, C., Shen, H., Li, X., 2015. Comparison of ensemble-based state and parameter estimation methods for soil moisture data assimilation. Advances in Water Resources.

Choi, B., 2012. ARMA model identification, Springer Science & Business Media.

Clark, M. P., et al. 2006. Assimilation of snow covered area information into hydrologic and land-surface models. Advances in Water Resources, 29(8), 1209-1221.

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(10): 1309-1324. DOI:10.1016/j.advwatres.2008.06.005

Cloke, H., and F. Pappenberger, 2009. Ensemble flood forecasting: a review, Journal of Hydrology, 375(3), 613-626.

Corzo, G. et al., 2009. Combining semi-distributed process-based and data-driven models in flow simulation: a case study of the Meuse river basin. Hydrology and Earth System Sciences, 13(9): 1619-1634.

Cosgrove, B. A., et al. 2003. Real‐time and retrospective forcing in the North American Land Data

Assimilation System (NLDAS) project. Journal of Geophysical Research: Atmospheres (1984–2012), 108(D22).

Cosh, M.H., Jackson, T.J., Starks, P., Heathman, G., 2006. Temporal stability of surface soil moisture in the Little Washita River watershed and its applications in satellite soil moisture product validation. Journal of Hydrology, 323(1): 168-177.

Coustau, M., et al. 2013. Benefits and limitations of data assimilation for discharge forecasting using an event-based rainfall–runoff model. Natural Hazards and Earth System Science, 13(3), 583-596.

Cressman, G. P. 1959. An operational objective analysis system. Monthly Weather Review, 87(10), 367-374.

Crow, W.T., 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.

Crow, W.T., Van Loon, E., 2006. Impact of incorrect model error assumptions on the sequential assimilation of remotely sensed surface soil moisture. Journal of hydrometeorology, 7(3): 421-432.

Crow, W. and Reichle, R. 2008. Comparison of adaptive filtering techniques for land surface data assimilation. Water resources research, 44, W08423, doi:10.1029/2008WR006883.

Crow, W. 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. and Yilmaz, M. T. 2014. The Auto‐Tuned Land Data Assimilation System (ATLAS). Water resources research, 50(1), 371-385.

Cuo, L., Pagano, T.C., Wang, Q., 2011. A review of quantitative precipitation forecasts and their use in short-to medium-range streamflow forecasting. Journal of hydrometeorology, 12(5): 713-728.

Das, N., et al. 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.

Das, N., Mohanty, B., Cosh, M., Jackson, T., 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.

De Kleermaeker, S., and J. Verkade, 2013. A decision support system for use of probability forecasts, paper presented at ISCRAM 2013: Proceedings of the 10th International Conference on Information Systems for Crisis Response and Management, Baden-Baden, Germany, 12-15 May 2013, ISCRAM.

de Rosnay, P. et al., 2013. A simplified Extended Kalman Filter for the global operational soil moisture analysis at ECMWF. Quarterly Journal of the Royal Meteorological Society, 139(674): 1199-1213.

Dechant, C., Moradkhani, H., 2011. Radiance data assimilation for operational snow and streamflow

forecasting. Advances in Water Resources, 34(3): 351-364.DOI:10.1016/j.advwatres.2010.12.009

DeChant, C., Moradkhani, H., 2014a. Hydrologic Prediction and Uncertainty Quantification. 387-414. CRC press, Taylor & Francis Group.

DeChant C.M., and H. Moradkhan, 2014b. Toward a Reliable Prediction of Seasonal Forecast Uncertainty: Addressing Model and Initial Condition Uncertainty with Ensemble Data Assimilation and Sequential Bayesian Combination, Journal of Hydrology, special issue on Ensemble Forecasting and data assimilation, 519, 2967-2977, DOI: 10.1016/j.jhydrol.2014.05.045.

Dee, D. P. 1995. On-line estimation of error covariance parameters for atmospheric data assimilation. Monthly Weather Review, 123(4), 1128-1145.

Derber, J., Rosati, A., 1989. A global oceanic data assimilation system. Journal of Physical Oceanography, 19(9): 1333-1347.

DIA, A.M., 2007. Adapting to climate variability in the Senegal River basin in West Africa LERG,University of Dakar, Senegal.

Dinh Tuan Pham, 2001. Stochastic Methods for Sequential Data Assimilation in Strongly Nonlinear Systems. Mon. Wea. Rev., 129, 1194–1207. DOI: http://dx.doi.org/10.1175/1520-0493(2001)129<1194:SMFSDA>2.0.CO;2

Divac, D., et al. 2009. A procedure for state updating of SWAT-based distributed hydrological model for operational runoff forecasting. Journal of the Serbian Society for Computational Mechanics, 3(1), 298-326.

Doubková, M., Van Dijk, A.I., Sabel, D., Wagner, W., Blöschl, G., 2012. Evaluation of the predicted error of the soil moisture retrieval from C-band SAR by comparison against modelled soil moisture estimates over Australia. Remote sensing of environment, 120: 188-196.

Draper, C. et al., 2013. Estimating root mean square errors in remotely sensed soil moisture over continental scale domains. Remote Sensing of Environment, 137: 288-298.

Draper, C., et al. 2011. Assimilation of ASCAT near-surface soil moisture into the SIM hydrological model over France. Hydrology and Earth System Sciences, 15(12), 3829-3841.

Draper, C., Mahfouf, J. F. and Walker, J. 2009. An EKF assimilation of AMSR‐E soil moisture into the ISBA land surface scheme. Journal of Geophysical Research: Atmospheres (1984–2012), 114(D20).

Draper, C., Mahfouf, J.-F., Calvet, J.-C., Martin, E., Wagner, W., 2011. Assimilation of ASCAT near-surface soil moisture into the SIM hydrological model over France. Hydrology and Earth System Sciences, 15(12): 3829-3841.

Drécourt, J.-P. 2003. Kalman filtering in hydrological modeling, Technical Report. Hørsholm, Denmark, DAIHM.

Duan, Z., Han, C. and Dang, H. 2003. An adaptive Kalman Filter with dynamic rescaling of process

noise. Proceedings of the Sixth Inter-national Conference of Information Fusion, vol.2, 1310-1315.

Dumas, D. et al., 2010. Large Dams and Uncertainties: The Case of the Senegal River (West Africa). Soc Natur Resour, 23(11): 1108-1122. DOI:10.1080/08941920903278137

Dumedah, G. and Coulibaly, P. 2012. Integration of evolutionary based assimilation into Kalman-type methods for streamflow simulations in ungauged watersheds. Journal of hydrology, 475, 428-440.

Dunne, S. and Entekhabi, D. 2005. An ensemble‐based reanalysis approach to land data assimilation. Water resources research, 41(2).

Ebert, E.E., Damrath, U., Wergen, W., Baldwin, M.E., 2003. The WGNE assessment of short-term quantitative precipitation forecasts. Bulletin of the American Meteorological Society, 84(4): 481-492.

El Serafy, G. and Mynett, A., 2004. Comparison of EKF and EnKF in Sobek River: Case study Maxau-Ijssel. ed. Proceedings of the 6th International Conference on Hydroinformatics. World Scientific, Singapore, ISBN, 981-238.

Elliott, R., Schiebe, F., Crawford, K., Peter, K., Puckett, W., 1993. A unique data capability for natural resources studies, International Winter Meeting of the American Society of Agricultural Engineering, Chicago, pp. 14-17.

Ermolaev, P. and Volynsky, M., 2014. The second order extended Kalman filter and Markov nonlinear filter for data processing in interferometric systems. ed. Journal of Physics: Conference Series, 012015.

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 (1978–2012), 99(C5), 10143-10162.

Evensen, G. 1997. Advanced data assimilation for strongly nonlinear dynamics. Monthly Weather Review, 125(6), 1342-1354.

Evensen, G. and Van Leeuwen, P. J. 2000. An ensemble Kalman smoother for nonlinear dynamics. Monthly Weather Review, 128(6), 1852-1867.

Evensen, G. 2003. The ensemble Kalman filter: Theoretical formulation and practical implementation. Ocean dynamics, 53(4), 343-367.

Fairbairn, D., et al. 2014. Comparing the SEKF with the DEnKF on a land surface model [online]. Available from: http://www.meteo.fr/cic/meetings/2014/CNA/presentations/3_2_Fairbairn.pdf. (Accessed March 31,3015)

Fan, Y. C. 1991. Prediction of stage-discharge in river using Kalman filter. Journal of Hydrodynamics, 3(1), 51-55.

Feddes, R.A., Kabat, P., Van Bakel, P., Bronswijk, J., Halbertsma, J., 1988. Modelling soil water dynamics in the unsaturated zone—state of the art. Journal of Hydrology, 100(1): 69-111.

Fitzgerald, R. J. 1971. Divergence of the Kalman filter. Automatic Control, IEEE Transactions on, 16(6), 736-747.

Francois, C., Quesney, A., Ottle, C., 2003. Sequential assimilation of ERS-1 SAR data into a coupled land surface-hydrological model using an extended Kalman Filter. Journal of Hydrometeorology, 4(2): 473-487. DOI:10.1175/1525-7541(2003)4<473:Saoesd>2.0.Co;2

Fraval, P. et al., 2002. The quest for integrated and sustainable water management in the Senegal River Valley, 5th Inter-Regional Conference on Environment and Water ENVIROWATER, Ouagadougou.

Galantowicz, J. F., Entekhabi, D. and Njoku, E. G. 1999. Tests of sequential data assimilation for retrieving profile soil moisture and temperature from observed L-band radiobrightness. Geoscience and Remote Sensing, IEEE Transactions on, 37(4), 1860-1870.

Gandin, L. S. 1965. Objective analysis of meteorological fields (Vol. 242). R. Hardin (Ed.). Jerusalem: Israel program for scientific translations.

Garen, D. C. ,1992. Improved techniques in regression-based streamflow volume forecasting, Journal of Water Resources Planning and Management, 118(6), 654-670.

Ge, G. 1984. The Kalman Filtering Algorithm for the Model of Runoff Formation at the Natural Storage Point. Journal of Chendu Univeristy of Technology, 1(4), 69-78.

Georgakakos, K. P. and Bras, R. L. 1982. Real‐time, statistically linearized, adaptive flood routing. Water resources research, 18(3), 513-524.

Georgakakos, K. P., D.-J. Seo, H. Gupta, J. Schaake, and M. B. Butts, 2004. Towards the characterization of streamflow simulation uncertainty through multimodel ensembles, Journal of Hydrology, 298(1), 222-241.

Georgakakos, K.P., Hudlow, M.D., 1984. Quantitative precipitation forecast techniques for use in hydrologic forecasting. Bulletin of the American Meteorological Society, 65(11): 1186-1200.

Gharamti, M., Hoteit, I., 2014. Complex step-based low-rank extended Kalman filtering for state-parameter estimation in subsurface transport models. Journal of Hydrology, 509: 588-600.

Ghil, M., Malanotte-Rizzoli, P., 1991. Data assimilation in meteorology and oceanography. Advances in geophysics, 33: 141-266.

Gift Dumedah and Paulin Coulibaly.2013. Integration of an evolutionary algorithm into the ensemble Kalman filter and the particle filter for hydrologic data assimilation. Journal of Hydroinformatics, Vol 16 No 1 pp 74–94. doi:10.2166/hydro.2013.088.

Gillijns, S., et al. 2006. "What is the ensemble Kalman filter and how well does it work?", Proc. Amer. Control Conf., pp.4448 -4453.

Grassi, S. and de Magistris, P. S. 2014. When long memory meets the Kalman filter: A comparative study. Computational Statistics & Data Analysis, 76, 301-319.

Hamilton, J. D., 1994. Time series analysis. Princeton, NJ: Princeton University press.

Han, E.J., Merwade, V., Heathman, G.C., 2012. Implementation of surface soil moisture data

assimilation with watershed scale distributed hydrological model. Journal of Hydrology, 416: 98-117. DOI:10.1016/j.jhydrol.2011.11.039

Hansen, M., R. DeFries, J.R.G. Townshend, and R. Sohlberg, 1998. UMD Global Land Cover Classification, 1 Kilometer, 1.0, Department of Geography, University of Maryland, College Park, Maryland, 1981-1994.

Harlim, J., Hunt, B.R., 2007. A non‐Gaussian Ensemble Filter for Assimilating Infrequent Noisy Observations. Tellus A, 59(2): 225-237. DOI:10.1111/j.1600-0870.2007.00225.x

Haykin, S. S.2001. Kalman filtering and neural networks. New York, NY: John Wiley & Sons Ltd.

He, H., Xiong, R., Zhang, X., Sun, F., Fan, J., 2011. State-of-charge estimation of the lithium-ion battery using an adaptive extended Kalman filter based on an improved Thevenin model. Vehicular Technology, IEEE Transactions on, 60(4): 1461-1469.

He, M., et al. 2012. An integrated uncertainty and ensemble-based data assimilation approach for improved operational streamflow predictions. Hydrology and Earth System Sciences, 16(3), 815-831.

Hendricks Franssen, H. and Kinzelbach, W. 2008. Real‐time groundwater flow modeling with the ensemble Kalman filter: Joint estimation of states and parameters and the filter inbreeding problem. Water resources research, 44(9).

Hess, R. 2001. Assimilation of screen-level observations by variational soil moisture analysis. Meteorology and Atmospheric Physics, 77(1-4), 145-154.

Hoffman, R.N., Liu, Z., Louis, J.-F., Grassoti, C., 1995. Distortion representation of forecast errors. Monthly Weather Review, 123(9): 2758-2770.

Hopmans, J. W., D. R. Nielsen, and K. L. Bristow, 2002. How useful are small-scale soil hydraulic property measurements for large-scale vadose zone modeling?, GEOPHYSICAL MONOGRAPH-AMERICAN GEOPHYSICAL UNION, 129, 247-258.

Houser, P. R., De Lannoy, G. J. and Walker, J. P. 2005. Hydrologic Data Assimilation. In Advances in Water Science Methodologies, Aswathanarayana, A. (ed.), AA Balkema, The Netherlands, pp. 230.

Houser, P. R., et al. 1998. Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. Water resources research, 34(12), 3405-3420.

Houser, P.R., De Lannoy, G.J., Walker, J.P., 2012. Hydrologic Data Assimilation In: Approaches to Managing Disaster - Assessing Hazards, Emergencies and Disaster Impacts. Tiefenbacher J (Ed.), InTech,. DOI:10.5772/31246.

Houtekamer, P. L. and Mitchell, H. L. 1998. Data assimilation using an ensemble Kalman filter technique. Monthly Weather Review, 126(3), 796-811.

Huang, C., Li, X. and Lu, L. 2008. Retrieving soil temperature profile by assimilating MODIS LST products with ensemble Kalman filter. Remote Sensing of Environment, 112(4), 1320-1336.

Huang, W.-C. 1999. Kalman Filter Effective to Hydrologic Routing? Journal of Marine Science and

Technology, 7(1), 65-71.

Ines, A. V. and Mohanty, B. P. 2008. Near‐surface soil moisture assimilation for quantifying effective soil hydraulic properties using genetic algorithm: 1. Conceptual modeling. Water resources research, 44(6).

Jackson, T.J., Le Vine, D.M., Swift, C.T., Schmugge, T.J., Schiebe, F.R., 1995. Large area mapping of soil moisture using the ESTAR passive microwave radiometer in Washita'92. Remote Sensing of Environment, 54(1): 27-37.

Jacobs, J. M., D. A. Myers, and B. M. Whitfield, 2003. IMPROVED RAINFALL/RUNOFF ESTIMATES USING REMOTELY SENSED SOIL MOISTURE1, JAWRA Journal of the American Water Resources Association, 39(2), 313-324.

Jain, A., and S. Srinivasulu, 2004. Development of effective and efficient rainfall‐runoff models using integration of deterministic, real‐coded genetic algorithms and artificial neural network techniques, Water Resources Research, 40(4), W04302.

Javelle, P., C. Fouchier, P. Arnaud, and J. Lavabre, 2010. Flash flood warning at ungauged locations using radar rainfall and antecedent soil moisture estimations, Journal of Hydrology, 394(1–2), 267-274.

Jayakrishnan, R., Srinivasan, R., Santhi, C., Arnold, J.G., 2005. Advances in the application of the SWAT model for water resources management. Hydrological Processes, 19(3): 749-762. DOI:10.1002/Hyp.5624

Jeevaragagam, P., and S. P. Simonovic, 2013. IMPROVEMENT OF STREAMFLOW SIMULATION FOR GAUGED SITE OF HYDROLOGICAL MODEL, Malaysian Journal of Civil Engineering, 25(2), 239-253.

Jeong, D. I., and Y.-O. Kim, 2009. Combining single-value streamflow forecasts–A review and guidelines for selecting techniques, Journal of Hydrology, 377(3), 284-299.

Jonsdottir, H., Madsen, H. and Palsson, O. P. 2006. Parameter estimation in stochastic rainfall-runoff models. Journal of hydrology, 326(1), 379-393.

Jwo, D.-J. and Wang, S.-H. 2007. Adaptive fuzzy strong tracking extended Kalman filtering for GPS navigation. Sensors Journal, IEEE, 7(5), 778-789.

Kaheil, Y.H., Gill, M.K., McKee, M., Bastidas, L.A., 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.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. Journal of Fluids Engineering, 82(1): 35-45. DOI:10.1115/1.3662552

Kalnay, E., et al. 2007. 4‐D‐Var or ensemble Kalman filter? Tellus A, 59(5), 758-773.

Keppenne, C. L. and Rienecker, M. M. 2002. Initial testing of a massively parallel ensemble Kalman filter with the Poseidon isopycnal ocean general circulation model. Monthly Weather Review, 130(12), 2951-2965.

Khac‐Tien Nguyen, P., and L. Hock‐Chye Chua, 2012. The data‐driven approach as an operational real‐time flood forecasting model, Hydrological Processes, 26(19), 2878-2893.

Kim, S., Tachikawa, Y. and Takara, K., 2005. Real-time prediction algorithm with a distributed hydrological model using Kalman filter. Annual journal of hydraulic engineering, JSCE, (49), 163-168.

Kitanidis, P. K. and Bras, R. L. 1980a. Real-time forecasting with a conceptual hydrologic model 1. Analysis of Uncertainty. WATER RESOURCES RES., 16(6), 1034-1044.

Kitanidis, P. K. and Bras, R. L. 1980b. Real‐time forecasting with a conceptual hydrologic model: 1. Analysis of Uncertainty. Water resources research, 16(6), 1025-1033.

Knight, D. and Shamseldin, A., 2006. River basin modelling for flood risk mitigation. D. W. Knight and A. Y. Shamseldin, eds., A. A. Balkema, Brookfield, VT, 301–334..

Kolmogorov, A. N., Doyle, W. and Selin, I. 1941. Interpolation and extrapolation of stationary random sequences. Izv. Akad. Nauk. SSSR Ser. Mat., 1941,5, 3–14. (in Russian) Translation in A. N. Shiryagev (Ed.), Selected works of A. N. Kolmogorov (Vol. 23, pp. 272–280). Norwell, MA: Kluwer.

Komma, J., Blöschl, G., Reszler, C., 2008. Soil moisture updating by Ensemble Kalman Filtering in real-time flood forecasting. Journal of Hydrology, 357(3): 228-242.

Krzysztofowicz, R., 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model, Water Resources Research, 35(9), 2739-2750.

Krzysztofowicz, R., 2001. The case for probabilistic forecasting in hydrology, Journal of Hydrology, 249(1), 2-9.

Kumar, P., Kaleita, A.L., 2003. Assimilation of near-surface temperature using extended Kalman filter. Advances in Water Resources, 26(1): 79-93.

Kumar, S. V., et al. 2008. A land surface data assimilation framework using the land information system: Description and applications. Advances in Water Resources, 31(11), 1419-1432.

Kumar, S.V., Reichle, R.H., Koster, R.D., Crow, W.T., Peters-Lidard, C.D., 2009. Role of Subsurface Physics in the Assimilation of Surface Soil Moisture Observations. Journal of Hydrometeorology, 10(6): 1534-1547. DOI:10.1175/2009jhm1134.1

Lawson, W. G. and Hansen, J. A. 2004. Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth. Monthly Weather Review, 132(8), 1966-1981.

Le Gland, F., Monbet, V., Tran, V.-D., 2009. Large sample asymptotics for the ensemble Kalman Filter. Oxford Handbook of Nonlinear Filtering. ed D Crisan and B Rozovskii. Oxford: Oxford University Press,pp 598–631.

Lee, Y. and Singh, V. 1999. Tank model using Kalman filter. Journal of Hydrologic Engineering, 4(4), 344-349.

Lei, F.N., Huang, C.L., Shen, H.F., Li, X., 2014. Improving the estimation of hydrological states in the

SWAT model via the ensemble Kalman smoother: Synthetic experiments for the Heihe River Basin in northwest China. Advances in Water Resources, 67: 32-45. DOI:10.1016/j.advwatres.2014.02.008

Leisenring, M., and Moradkhani, H., 2012. Analyzing the uncertainty of suspended sediment load prediction using sequential data assimilation. Journal of Hydrology, 468-469:268-282.

Lewis, F. L., Xie, L. and Popa, D., 2008. Optimal and robust estimation: with an introduction to stochastic control theory. CRC.

Li, B., Toll, D., Zhan, X., Cosgrove, B., 2012. Improving estimated soil moisture fields through assimilation of AMSR-E soil moisture retrievals with an ensemble Kalman filter and a mass conservation constraint. Hydrology and Earth System Sciences, 16(1): 105-119.

Li, Y., et al. 2013. Assimilation of stream discharge for flood forecasting: The benefits of accounting for routing time lags. Water resources research, 49(4), 1887-1900.

Li, Y., et al. 2014. An integrated error parameter estimation and lag-aware data assimilation scheme for real-time flood forecasting. Journal of hydrology, 519, 2722-2736.

Liang, X., Lettenmaier, D.P., Wood, E.F., Burges, S.J., 1994. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. Journal of Geophysical Research: Atmospheres (1984–2012), 99(D7): 14415-14428.

Liu, J., DOAN, C. D. and Liong, S.-Y. 2011. Conceptual Rainfall-Runoff Model with Kalman Filter for Parameter and Outflow Updating. Advances in Geosciences (a 6-volume Set)-Volume 23: Hydrological Science (hs), 23, 133.

Liu, S., et al. 2012a. Improved regional hydrologic modelling by assimilation of streamflow data into a regional hydrologic model. Environmental Modelling & Software, 31, 141-149.

Liu, Y., and H. V. Gupta, 2007. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, Water Resources Research, 43(7), W07401.

Liu, Y. et al., 2012b. Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. Hydrology and Earth System Sciences, 16(10): 3863-3887. DOI:DOI 10.5194/hess-16-3863-2012

Liu, Y., et al. 2015. Blending satellite‐based snow depth products with in situ observations for streamflow predictions in the Upper Colorado River Basin. Water resources research. 51(2), 1182-1202.

Ljung, L. 1979. Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems. Automatic Control, IEEE Transactions on, 24(1), 36-50.

Lopez, P.L. et al., 2015. Improved large-scale hydrological modelling through the assimilation of streamflow and downscaled satellite soil moisture observations.

Lorenc, A. 1981. A global three-dimensional multivariate statistical interpolation scheme. Monthly Weather Review, 109(4), 701-721.

Lorenc, A. C. 2003. The potential of the ensemble Kalman filter for NWP—a comparison with 4D‐Var.

Quarterly Journal of the Royal Meteorological Society, 129(595), 3183-3203.

Lü, H., et al. 2010. Using a H∞ filter assimilation procedure to estimate root zone soil water content. Hydrological processes, 24(25), 3648-3660.

Lü, H., et al. 2011. Dual state-parameter estimation of root zone soil moisture by optimal parameter estimation and extended Kalman filter data assimilation. Advances in Water Resources, 34(3), 395-406.

Madsen, H., et al. 2003. Data assimilation in the MIKE 11 Flood Forecasting system using Kalman filtering. International Association of Hydrological Sciences, Publication, (281), 75-81.

Madsen, H., and C. Skotner, 2005. Adaptive state updating in real-time river flow forecasting—A combined filtering and error forecasting procedure, Journal of Hydrology, 308(1), 302-312.

Mahfouf, J. F., et al. 2009. A comparison of two off‐line soil analysis schemes for assimilation of screen level observations. Journal of Geophysical Research: Atmospheres (1984–2012), 114(D8).

Massari, C., L. Brocca, T. Moramarco, Y. Tramblay, and J.-F. D. Lescot, 2014. Potential of soil moisture observations in flood modelling: Estimating initial conditions and correcting rainfall, Advances in Water Resources, 74, 44-53.

Mazzoleni, M., Alfonso, L., Chacon-Hurtado, J., Solomatine, D., 2015. Assimilating uncertain, dynamic and intermittent streamflow observations in hydrological models. Advances in Water Resources, 83: 323-339.

McLaughlin, D. 1995. Recent developments in hydrologic data assimilation. Reviews of Geophysics, 33(S2), 977-984.

McLaughlin, D. 2002. An integrated approach to hydrologic data assimilation: interpolation, smoothing, and filtering. Advances in Water Resources, 25(8), 1275-1286.

McMillan, H., et al. 2010. Impacts of uncertain river flow data on rainfall‐runoff model calibration and discharge predictions. Hydrological processes, 24(10), 1270-1284.

McMillan, H., Jackson, B., Clark, M., Kavetski, D., Woods, R., 2011. Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models. Journal of Hydrology, 400(1): 83-94.

McMillan, H., et al. 2013. Operational hydrological data assimilation with the recursive ensemble Kalman filter. Hydrology and Earth System Sciences, 17(1), 21-38.

Mehra, R. K. 1970. On the identification of variances and adaptive Kalman filtering. Automatic Control, IEEE Transactions on, 15(2), 175-184.

Mehra, R. K. 1972. Approaches to adaptive filtering. Automatic Control, IEEE Transactions on, 17(5), 693-698.

Meng, Q.-h., Sun, Y.-c., Cao, Z.-l., 2000. Adaptive extended Kalman filter (AEKF)-based mobile robot localization using sonar. Robotica, 18(05): 459-473.

Mitchell, H. L., Houtekamer, P. and Pellerin, G. 2002. Ensemble Size, Balance, and Model-Error

Representation in an Ensemble Kalman Filter*. Monthly Weather Review, 130(11), 2791-2808.

Mitchell, K. E., et al. 2004. The multi‐institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. Journal of Geophysical Research: Atmospheres (1984–2012), 109(D7).

Montzka, C., et al. 2012. Multivariate and multiscale data assimilation in terrestrial systems: A review. Sensors, 12(12), 16291-16333.

Montzka, C., J. P. Grant, H. Moradkhani, H.-J. H. Franssen, L. Weihermüller, M. Drusch, and H. Vereecken, 2013. Estimation of radiative transfer parameters from L-band passive microwave brightness temperatures using advanced data assimilation, Vadose Zone Journal, 12(3).

Moradkhani, H., et al. 2005a. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. Water resources research, 41(5). DOI: 10.1029/2004WR003604.

Moradkhani, H., et al. 2005b. Dual state–parameter estimation of hydrological models using ensemble Kalman filter. Advances in Water Resources, 28(2), 135-147.

Moradkhani, H., 2008. Hydrologic remote sensing and land surface data assimilation. Sensors, 8(5): 2986-3004.

Moradkhani, H., C.M. DeChant and S. Sorooshian, 2012. Evolution of Ensemble Data Assimilation for Uncertainty Quantification using the Particle Filter-Markov Chain Monte Carlo Method, Water Resources Research,48,W12520, doi:10.1029/2012WR012144.

Muluye, G.Y., 2011. Improving long-range hydrological forecasts with extended Kalman Filters. Hydrolog Sci J, 56(7): 1118-1128. DOI:10.1080/02626667.2011.608068

Navon, I. M. 2009. Data assimilation for numerical weather prediction: a review. In Data assimilation for atmospheric, oceanic and hydrologic applications. Springer Berlin Heidelberg. pp. 21-65.

Neethling, C. and Young, P. 1974. Comments on" Identification of optimum filter steady-state gain for systems with unknown noise covariances". Automatic Control, IEEE Transactions on, 19(5), 623-625.

Noh, S.J., Rakovec, O., Weerts, A.H., Tachikawa, Y., 2014. On noise specification in data assimilation schemes for improved flood forecasting using distributed hydrological models. Journal of Hydrology, 519: 2707-2721.

Odelson, B. J., Rajamani, M. R. and Rawlings, J. B. 2006. A new autocovariance least-squares method for estimating noise covariances. Automatica, 42(2), 303-308.

Pagano, T., Q. Wang, P. Hapuarachchi, and D. Robertson, 2011. A dual-pass error-correction technique for forecasting streamflow, Journal of Hydrology, 405(3), 367-381.

Pan, M., et al. 2003. Snow process modeling in the North American Land Data Assimilation System (NLDAS): 2. Evaluation of model simulated snow water equivalent. Journal of Geophysical Research: Atmospheres (1984–2012), 108(D22).

Pan, M. and Wood, E. F. 2006. Data assimilation for estimating the terrestrial water budget using a constrained ensemble Kalman filter. Journal of hydrometeorology, 7(3), 534-547.

Paniconi, C., et al. 2003. Newtonian nudging for a Richards equation-based distributed hydrological model. Advances in Water Resources, 26(2), 161-178.

Panzeri, M., et al. 2015. EnKF coupled with groundwater flow moment equations applied to Lauswiesen aquifer, Germany. Journal of hydrology, 521, 205-216.

Papadakis, N., Memin, E., Cuzol, A., Gengembre, N., 2010. Data assimilation with the weighted ensemble Kalman Filter. Tellus Series a-Dynamic Meteorology and Oceanography, 62(5): 673-697. DOI:10.1111/j.1600-0870.2010.00461.x

Parajka, J. and Blöschl, G. 2008. The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models. Journal of hydrology, 358(3), 240-258.

Parrens, M., et al. 2014. Assimilation of surface soil moisture into a multilayer soil model: design and evaluation at local scale. Hydrology and Earth System Sciences, 18(2), 673-689.

Pauwels, V. R. and De Lannoy, G. J. 2006. Improvement of modeled soil wetness conditions and turbulent fluxes through the assimilation of observed discharge. Journal of hydrometeorology, 7(3), 458-477.

Pauwels, V., De Lannoy, G.J., 2009. Ensemble‐based assimilation of discharge into rainfall‐runoff models: A comparison of approaches to mapping observational information to state space. Water resources research, 45(8).

Pham, D. T., Verron, J. and Christine Roubaud, M. 1998. A singular evolutive extended Kalman filter for data assimilation in oceanography. Journal of Marine systems, 16(3), 323-340.

Pham, D. T. 2001. Stochastic methods for sequential data assimilation in strongly nonlinear systems. Monthly Weather Review, 129(5), 1194-1207.

Piani, C., Haerter, J. and Coppola, E. 2010. Statistical bias correction for daily precipitation in regional climate models over Europe. Theoretical and Applied Climatology, 99(1-2), 187-192.

Puente, C. E. and Bras, R. L. 1987. Application of nonlinear filtering in the real time forecasting of river flows. Water resources research, 23(4), 675-682.

Rabier, F., Liu,Z., 2003. Variational assimilation:theory and overview: theory and overview. Proc. Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean. Reading, United Kingdom, ECMWF, 2003,29–43..

Rafieeinasab, A., et al. 2014. Comparative evaluation of maximum likelihood ensemble filter and ensemble Kalman filter for real-time assimilation of streamflow data into operational hydrologic models. Journal of hydrology, 519, 2663-2675.

Rafieeinasab, A., Seo, D.J., Lee, H., Kim, S., 2014. Comparative evaluation of maximum likelihood ensemble filter and ensemble Kalman Filter for real-time assimilation of streamflow data into operational hydrologic models. Journal of Hydrology, 519: 2663-2675. DOI:10.1016/j.jhydrol.2014.06.052

Rakovec, O., et al. 2015. Operational aspects of asynchronous filtering for hydrological forecasting. Hydrology and Earth System Sciences Discussions, 12(3), 3169-3203.

Randrianasolo, A., Thirel, G., Ramos, M.H., Martin, E., 2014. Impact of streamflow data assimilation and length of the verification period on the quality of short-term ensemble hydrologic forecasts. Journal of Hydrology, 519: 2676-2691. DOI:10.1016/j.jhydrol.2014.09.032

Rasmussen, J., et al. 2015. Data assimilation in integrated hydrological modeling using ensemble Kalman filtering: evaluating the effect of ensemble size and localization on filter performance. Hydrology and Earth System Sciences Discussions, 12(2), 2267-2304.

Reichle, R. H., McLaughlin, D. B. and Entekhabi, D. 2001. Variational data assimilation of microwave radiobrightness observations for land surface hydrology applications. Geoscience and Remote Sensing, IEEE Transactions on, 39(8), 1708-1718.

Reichle, R. H., McLaughlin, D. B. and Entekhabi, D. 2002a. Hydrologic data assimilation with the ensemble Kalman filter. Monthly Weather Review, 130(1), 103-114.

Reichle, R. H., et al. 2002b. Extended versus ensemble Kalman filtering for land data assimilation. Journal of hydrometeorology, 3(6), 728-740.

Reichle, R. H. and Koster, R. D. 2003. Assessing the impact of horizontal error correlations in background fields on soil moisture estimation. Journal of hydrometeorology, 4(6), 1229-1242.

Reichle, R. H. 2008. Data assimilation methods in the Earth sciences. Advances in Water Resources, 31(11), 1411-1418.

Rimkus, S., 2005. Performance measuring in the case of the Senegal River watershed management, ETH Zürich: Seminar on Science and Politics of the International Freshwater Management, Working Paper, Zürich.

Robinson, A. R. and Lermusiaux, P. F. 2000. Overview of data assimilation. Harvard reports in physical/interdisciplinary ocean science, 62, 1-13.

Rodell, M. and Houser, P. 2004. Updating a land surface model with MODIS-derived snow cover. Journal of hydrometeorology, 5(6), 1064-1075.

Rong Li, X. and Bar-Shalom, Y. 1994. A recursive multiple model approach to noise identification. Aerospace and Electronic Systems, IEEE Transactions on, 30(3), 671-684.

Roth, M. and Gustafsson, F. 2011. An efficient implementation of the second order extended Kalman filter. ed. Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on, , 1-6.

Rutherford, I.D., 1972. Data assimilation by statistical interpolation of forecast error fields. Journal of the Atmospheric Sciences, 29(5): 809-815.

Ryu, D., Crow, W.T., Zhan, X., Jackson, T.J., 2009. Correcting unintended perturbation biases in hydrologic data assimilation. Journal of hydrometeorology, 10(3): 734-750.

Sadeghi, B., Moshiri, B. 2007. " Second order EKF and Unscented Kalman Filter fusion for tracking maneuvering targets," IEEE International Conference on Information Reuse and Integration, pp.

514-519, 13-15 August 2007.

Sage, A. P. and Husa, G. W., 1969. Adaptive filtering with unknown prior statistics. ed. Proceedings of joint automatic control conference, 760-769.

Sahoo, A.K., De Lannoy, G.J., Reichle, R.H., Houser, P.R., 2013. Assimilation and downscaling of satellite observed soil moisture over the Little River Experimental Watershed in Georgia, USA. Advances in Water Resources, 52: 19-33.

Sakov, P., Evensen, G. and Bertino, L. 2010. Asynchronous data assimilation with the EnKF. Tellus A, 62(1), 24-29.

Salamon, P., Feyen, L., 2009. Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter. Journal of Hydrology, 376(3-4): 428-442. DOI:10.1016/j.jhydrol.2009.07.051

Saleh, D.K., Kratzer, C.R., Green, C.H., Evans, D.G., 2009. Using the Soil and Water Assessment Tool (SWAT) to Simulate Runoff in Mustang Creek Basin, California. U.S. Geological Survey Scientific Investigations Report 2009–5031, 28 p., Reston,VA.

Samuel, J., Coulibaly, P., Dumedah, G., Moradkhani, H., 2014. Assessing model state and forecasts variation in hydrologic data assimilation. Journal of Hydrology, 513: 127-141. DOI:10.1016/j.jhydrol.2014.03.048

Sandholt, I. et al., 2003. Remote sensing techniques for flood monitoring in the Senegal River Valley. Geografisk Tidsskrift-Danish Journal of Geography, 103(1): 71-81. DOI:10.1080/00167223.2003.10649481

Santhi, C. et al., 2001. Validation of the SWAT model on a large river basin with point and nonpoint sources. JAWRA Journal of the American Water Resources Association, 37(5): 1169–1188. DOI:10.1111/j.1752-1688.2001.tb03630.x

Sarkka, S. and Nummenmaa, A. 2009. Recursive noise adaptive Kalman filtering by variational Bayesian approximations. Automatic Control, IEEE Transactions on, 54(3), 596-600.

Schuol, J., Abbaspour, K., 2006. Calibration and uncertainty issues of a hydrological model (SWAT) applied to West Africa. Advances in geosciences, 9(9): 137-143.

Schuol, J., Abbaspour, K.C., Srinivasan, R., Yang, H., 2008. Estimation of freshwater availability in the West African sub-continent using the SWAT hydrologic model. Journal of Hydrology, 352(1-2): 30-49. DOI:10.1016/j.jhydrol.2007.12.025

Seo, D.-J., Koren, V. and Cajina, N. 2003. Real-time variational assimilation of hydrologic and hydrometeorological data into operational hydrologic forecasting. Journal of hydrometeorology, 4(3), 627-641.

Sene, K., 2008. Flood warning, forecasting and emergency response. Springer Science & Business Media.

Seo, D.-J., et al. 2009. Automatic state updating for operational streamflow forecasting via variational data assimilation. Journal of hydrology, 367(3), 255-275.

Shamir, E., Lee, B.J., Bae, D.H., Georgakakos, K.P., 2010. Flood Forecasting in Regulated Basins Using the Ensemble Extended Kalman Filter with the Storage Function Method. Journal of Hydrologic Engineering, 15(12): 1030-1044. DOI:10.1061/(Asce)He.1943-5584.0000282

Sheffield, J., et al. 2003. Snow process modeling in the North American Land Data Assimilation System (NLDAS): 1. Evaluation of model‐simulated snow cover extent. Journal of Geophysical Research: Atmospheres (1984–2012), 108(D22).

Shi, Y., Davis, K.J., Zhang, F., Duffy, C.J., Yu, X., 2014. Parameter estimation of a physically based land surface hydrologic model using the ensemble Kalman filter: A synthetic experiment. Water Resources Research, 50(1): 706-724.

Simon, D. 2006. Using nonlinear Kalman filtering to estimate signals. Embedded Systems Design, 19(7), 38.

Simunek, J., Van Genuchten, M.T., Sejna, M., 2005. The HYDRUS-1D software package for simulating the movement of water, heat, and multiple solutes in variably saturated media, version 3.0, HYDRUS software series 1. Department of Environmental Sciences, University of California Riverside, Riverside Edition.

Sittichok, K. et al., 2014. Statistical seasonal rainfall and streamflow forecasting for the Sirba watershed, West Africa, using sea surface temperatures. Hydrological Sciences Journal,1-11. DOI:10.1080/02626667.2014.944526

Sloan, P.G., Moore, I.D., 1984. Modeling Subsurface Stormflow on Steeply Sloping Forested Watersheds. Water Resources Research, 20(12): 1815-1822. DOI:10.1029/Wr020i012p01815

Solomatine, D. P., and A. Ostfeld. 2008. Data-driven modelling: Some past experiences and new approaches, J. Hydroinformatics, 10(1), 3–22. doi:10.2166/hydro.2008.015.

Spruill, C., Workman, S., Taraba, J., 2000. Simulation of daily and monthly stream discharge from small watersheds using the SWAT model. Transactions of the ASAE, 43(6): 1431-1439.

Srinivasan, R., Ramanarayanan, T.S., Arnold, J.G., Bednarz, S.T., 1998. Large area hydrologic modeling and assessment part II: Model application. JAWRA Journal of the American Water Resources Association, 34(1): 91-101. DOI:10.1111/j.1752-1688.1998.tb05962.x

Ssegane, H., E. Tollner, Y. Mohamoud, T. Rasmussen, and J. Dowd, 2012. Advances in variable selection methods I: Causal selection methods versus stepwise regression and principal component analysis on data of known and unknown functional relationships, Journal of Hydrology, 438-439, 16-25.

Steiner, J.L. et al., 2014. Long-term environmental research: The Upper Washita River experimental watersheds, Oklahoma, USA. Journal of environmental quality, 43(4): 1227-1238.

Sun, L., et al., 2013. Prediction of Daily Discharge at Bakel (Senegal) using Multiple Linear Regression, Kalman Filer and Artificial Neural Networks. CSCE 2013, 3rd Specialty Conference on Disaster Prevention and Mitigation. Montreal,Canada.

Sun, L., Nistor, I., Seidou, O., 2015a. Streamflow data assimilation in SWAT model using Extended

Kalman Filter. Journal of Hydrology, 531, Part 3: 671-684. DOI:http://dx.doi.org/10.1016/j.jhydrol.2015.10.060

Sun, L., Seidou, O., Nistor, I., Liu, K., 2015b. Review of The Kalman Type Hydrological Data Assimilation. Hydrological Science Journal In Press. DOI:10.1080/02626667.2015.1127376

Szöllősi-Nagy, A. and Mekis, E. 1988. Comparative analysis of three recursive real-time river flow forecasting models: deterministic, stochastic, and coupled deterministic-stochastic. Stochastic Hydrology and Hydraulics, 2(1), 17-33.

Tanizaki, H., 1996. Nonlinear filters: estimation and applications. New York: Springer Science & Business Media.

Tao, W., et al. 2005. A Method of Preventing Divergence of EKF. JOURNAL OF PROJECTILES, ROCKETS, MISSILES AND GUIDANCE, 25(2), 484-486.

Tayfur, G., Zucco, G., Brocca, L., Moramarco, T., 2014. Coupling soil moisture and precipitation observations for predicting hourly runoff at small catchment scale. Journal of Hydrology, 510: 363-371.

Tingsanchali, T., and M. R. Gautam, 2000. Application of tank, NAM, ARMA and neural network models to flood forecasting, Hydrological Processes, 14(14), 2473-2487.

Todini, 1978. E. Todini Mutually interactive state-parameter (MISP) estimation, C.L. Chiu (Ed.), Proc. AGU Chapman Conf. on Application of Kalman Filter to Hydrology, Hydraulics and Water Resources, Department of Civil Engineering, University of Pittsburgh, Pittsburgh (1978), pp. 135–151

Todini, E. ,1988. Rainfall-runoff modeling—Past, present and future, Journal of Hydrology, 100(1), 341-352.

Todini, E., 2007. Hydrological catchment modelling: past, present and future, Hydrology and Earth System Sciences, 11(1), 468-482.

Toth, E., Brath, A., Montanari, A., 2000. Comparison of short-term rainfall prediction models for real-time flood forecasting. Journal of Hydrology, 239(1): 132-147.

Tran, A. P., et al. 2014. Joint estimation of soil moisture profile and hydraulic parameters by ground‐penetrating radar data assimilation with maximum likelihood ensemble filter. Water resources research, 50(4), 3131-3146.

Tran, A.P., Vanclooster, M., Zupanski, M., Lambot, S., 2014. Joint estimation of soil moisture profile and hydraulic parameters by ground‐penetrating radar data assimilation with maximum likelihood ensemble filter. Water Resources Research, 50(4): 3131-3146.

Trudel, M., Leconte, R. and Paniconi, C. 2014. Analysis of the hydrological response of a distributed physically-based model using post-assimilation (EnKF) diagnostics of streamflow and in situ soil moisture observations. Journal of hydrology, 514, 192-201.

Van Leeuwen, P. J. 2009. Particle filtering in geophysical systems. Monthly Weather Review, 137(12), 4089-4114.

Van Liew, M.W., Garbrecht, J., 2003. Hydrologic simulation of the Little Washita River experimental watershed using SWAT. Journal of the American Water Resources Association, 39(2): 413-426.

Velázquez, J., F. Anctil, M. Ramos, and C. Perrin, 2011. Can a multi-model approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures, Advances in Geosciences, 29(29), 33-42.

Vereecken, H. et al., 2015. Soil hydrology: Recent methodological advances, challenges, and perspectives. Water Resources Research.

Verhoef, W., Bach, H., 2003. Remote sensing data assimilation using coupled radiative transfer models. Physics and Chemistry of the Earth, Parts A/B/C, 28(1): 3-13.

Vittaldev, V., Russell, R. P., Arora, N., & Gaylor, D. 2012. Second-order Kalman filters using multi-complex step derivatives. In Proceedings of the AAS/AIAA Space Flight Mechanics Meeting 2012, Kauai, Hawaii.

Vrugt, J., Wijk, M.v., Hopmans, J.W., Šimunek, J., 2001. One - , two - , and three - dimensional root water uptake functions for transient modeling. Water Resources Research, 37(10): 2457-2470.

Vrugt, J. A., et al. 2005. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. Water resources research, 41(1).

Vrugt, J. A., et al. 2006. Real-time data assimilation for operational ensemble streamflow forecasting. Journal of hydrometeorology, 7(3), 548-565.

Vrugt, J.A., Ter Braak, C.J., Clark, M.P., Hyman, J.M., Robinson, B.A., 2008. Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. Water Resources Research, 44(12). W00B09.

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, 106(11), 11.

Walker, J. P., Houser, P. R. and Reichle, R. H. 2003. New technologies require advances in hydrologic data assimilation. EOS, Transactions American Geophysical Union, 84(49), 545-551.

Walker, J. P. and Houser, P. R. 2005. Hydrologic data assimilation. Advances in water science methodologies. Londres: Taylor & Francis, ed, 1, 25-48.

Wan, E. A., and R. Van Der Merwe, 2000. The unscented Kalman filter for nonlinear estimation, paper presented at Adaptive Systems for Signal Processing, Communications, and Control Symposium. AS-SPCC. The IEEE 2000, IEEE.

Wanders, N., et al. 2014a. The benefits of using remotely sensed soil moisture in parameter identification of large - scale hydrological models. Water resources research, 50(8), 6874-6891.

Wanders, N., et al. 2014b. The suitability of remotely sensed soil moisture for improving operational flood forecasting. Hydrology and Earth System Sciences, 18(6), 2343-2357.

Wang, D. and Cai, X. 2008. Robust data assimilation in hydrological modeling–A comparison of Kalman and H-infinity filters. Advances in Water Resources, 31(3), 455-472.

Wang, D.B., Chen, Y.G., Cai, X.M., 2009. State and parameter estimation of hydrologic models using the constrained ensemble Kalman Filter. Water Resources Research, 45(11). DOI:10.1029/2008wr007401

Wang, W. and Kou, X. 2009. Methods for hydrological data assimilation and advances of assimilating remotely sensed data into rainfall-runoff models. Journal of Hohai University: Natural Sciences, 37(5), 556-562.

Wang, Y. and Wang, S. 1985. The application of adaptive data assimilation in hydrological forecast. Hydrology, 1, 56-58.

Weerts, A. H., and G. Y. H. El Serafy, 2006. Particle filtering and ensemble Kalman Filtering for state updating with hydrological conceptual rainfall-runoff models, Water Resour. Res., 42, W09403, doi:10.1029/2005WR004093.

Whitaker, J. S. and Hamill, T. M. 2002. Ensemble data assimilation without perturbed observations. Monthly Weather Review, 130(7), 1913-1924.

White, K.L., Chaubey, I., 2005. Sensitivity analysis, calibration, and validations for a multisite and multivariable swat model. JAWRA Journal of the American Water Resources Association, 41(5): 1077-1089. DOI:10.1111/j.1752-1688.2005.tb03786.x

Wiener, N., 1949. Extrapolation, interpolation, and smoothing of stationary time series. MIT press Cambridge, MA.

Wishner, R., Tabaczynski, J., Athans, M., 1969. A comparison of three non-linear filters. Automatica, 5(4): 487-496. DOI:10.1016/0005-1098(69)90110-1

WMO, 1992. Simulated real-time intercomparison of hydrological models. Oper Hydrol.Rep.38,WMO No.779,Geneva.

Wood, E.F., Szöllösi‐Nagy, A., 1978. An adaptive algorithm for analyzing short‐term structural and parameter changes in hydrologic prediction models. Water Resources Research, 14(4): 577-581.

Wu, S.-J., H.-C. Lien, C.-H. Chang, and J.-C. Shen, 2012. Real-time correction of water stage forecast during rainstorm events using combination of forecast errors, Stochastic Environmental Research and Risk Assessment, 26(4), 519-531.

Wu, X.-l., et al. 2008. Kalman filtering correction in real-time forecasting with hydrodynamic model. Journal of Hydrodynamics, Ser. B, 20(3), 391-397.

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.

Xie, X. and Zhang, D. 2013. A partitioned update scheme for state‐parameter estimation of distributed hydrologic models based on the ensemble Kalman filter. Water resources research, 49(11), 7350-7365.

Xie, X., Meng, S., Liang, S., Yao, Y., 2014. Improving streamflow predictions at ungauged locations

with real-time updating: application of an EnKF-based state-parameter estimation strategy. Hydrology and Earth System Sciences, 18(10): 3923-3936. DOI:10.5194/hess-18-3923-2014

Xiong, L., and K. M. O'connor, 2002. Comparison of four updating models for real-time river flow forecasting, Hydrological Sciences Journal, 47(4), 621-639.

Xu, X. et al., 2014. Assimilation of SMOS soil moisture in the MESH model with the ensemble Kalman filter, Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International. IEEE, pp. 3766-3769.

Xu, X., Li, J. and Tolson, B. A. 2014. Progress in integrating remote sensing data and hydrologic modeling. Progress in Physical Geography, 0309133314536583.

Yan H., DeChant C. M., and Moradkhani H. 2015. Improving Soil Moisture Profile Prediction With the Particle Filter-Markov Chain Monte Carlo Method. IEEE Transactions on Geoscience and Remote Sensing, 1-14, DOI 10.1109/TGRS.2015.2432067.

Yu, P.-S. and Chen, S.-T. 2005. Updating Real-Time Flood Forecasting Using a Fuzzy Rule-Based Model/Mise à Jour de Prévision de Crue en Temps Réel Grâce à un Modèle à Base de Règles Floues. Hydrological Sciences Journal, 50(2), 265-278.

Zealand, C. M., D. H. Burn, and S. P. Simonovic, 1999. Short term streamflow forecasting using artificial neural networks, Journal of hydrology, 214(1), 32-48.

Zhang, X.S., Srinivasan, R., Bosch, D., 2009. Calibration and uncertainty analysis of the SWAT model using Genetic Algorithms and Bayesian Model Averaging. Journal of Hydrology, 374(3-4): 307-317. DOI:10.1016/j.jhydrol.2009.06.023

Zhong, J. and Brown, M. 2009. Joint State and Parameter Estimation for Biochemical Dynamic Pathways with Iterative Extended Kalman Filter: Comparison with Dual State and Parameter Estimation. The Open Automation and Control Systems Journal, 2(1), 69-77.