3-21-2016

Hydrologic Modeling in Dynamic Catchments: A Data Assimilation Approach

Sahani Darshika Pathiraja
University of New South Wales

Ashish Sharma
University of New South Wales

Lucy Marshall
University of New South Wales

Hamid Moradkhani
Portland State University, hamidm@pdx.edu

Citation Details

Pathiraja, S., L. Marshall, A. Sharma, and H. Moradkhani (2016), Hydrologic modeling in dynamic catchments: A data assimilation approach, Water Resour. Res., 52, 3350–3372, doi:10.1002/2015WR017192.

AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2015WR017192

Hydrologic modeling in dynamic catchments: A data assimilation approach

S. Pathiraja[1], L. Marshall[1], A. Sharma[1], and H. Moradkhani[2]

[1]School of Civil and Environmental Engineering, University of New South Wales, Sydney, New South Wales, Australia, [2]Department of Civil and Environmental Engineering, Portland State University, Portland, Oregon, USA

Key Points:
- A data assimilation based method for modeling dynamic catchments is investigated
- Prior parameters are critical in determining performance of data assimilation
- Improved approach for time varying parameter estimation presented

Supporting Information:
- Supporting Information S1

Correspondence to:
S. Pathiraja,
s.pathiraja@unsw.edu.au

Citation:
Pathiraja, S., L. Marshall, A. Sharma, and H. Moradkhani (2016), Hydrologic modeling in dynamic catchments: A data assimilation approach, *Water Resour. Res.*, *52*, 3350–3372, doi:10.1002/2015WR017192.

Received 6 MAR 2015
Accepted 16 MAR 2016
Accepted article online 21 MAR 2016
Published online 4 MAY 2016

Abstract The transferability of conceptual hydrologic models in time is often limited by both their structural deficiencies and adopted parameterizations. Adopting a stationary set of model parameters ignores biases introduced by the data used to derive them, as well as any future changes to catchment conditions. Although time invariance of model parameters is one of the hallmarks of a high quality hydrologic model, very few (if any) models can achieve this due to their inherent limitations. It is therefore proposed to consider parameters as potentially time varying quantities, which can evolve according to signals in hydrologic observations. In this paper, we investigate the potential for Data Assimilation (DA) to detect known temporal patterns in model parameters from streamflow observations. It is shown that the success of the DA algorithm is strongly dependent on the method used to generate background (or prior) parameter ensembles (also referred to as the parameter evolution model). A range of traditional parameter evolution techniques are considered and found to be problematic when multiple parameters with complex time variations are estimated simultaneously. Two alternative methods are proposed, the first is a Multilayer approach that uses the EnKF to estimate hyperparameters of the temporal structure, based on apriori knowledge of the form of nonstationarity. The second is a Locally Linear approach that uses local linear estimation and requires no assumptions of the form of parameter nonstationarity. Both are shown to provide superior results in a range of synthetic case studies, when compared to traditional parameter evolution techniques.

1. Introduction

All hydrologic models are simplified representations of the complex spatially and temporally varying rainfall-runoff conversion processes. Consequently, significant emphasis is often placed on the model parameters to account for any unresolved physical processes, scale factors and unknown physiographic attributes [*Gupta et al*., 1998; *Bardossy*, 2007]. It is customary to batch calibrate a hydrologic model against a particular data period, with the end goal of deriving a globally optimum parameter vector [e.g., *Duan et al*., 1994] or stationary distribution of parameters [e.g., *Beven and Freer*, 2001]. This approach is problematic for a number of reasons. Firstly, the optimal parameter set(s) are no longer suitable if the catchment conditions change. Catchments are fundamentally dynamic in nature, often undergoing changes due to natural processes (e.g., bushfire, changes in surface soil cover due to erosion) or human intervention (e.g., urbanization, diversions etc.). The effect of catchment dynamics on hydrologic variables such as streamflow has been widely documented (see for instance *Kuczera* [1987] and *Scott and Van Wyk* [1990] in regards to changes in the runoff regime after bushfires, and *Siriwardena et al*. [2006] in relation to afforestation/deforestation). Second, model parameters often display a strong dependence on the calibration data used to derive them, due largely to the empirical nature of many conceptual hydrologic models [*Ebtehaj et al*., 2010]. In particular, several studies have noted the dependence between model parameters and the dominant climatic regime of the calibration period, even when the catchment itself has remained relatively stationary [*Sorooshian et al*., 1983; *Choi and Beven*, 2007; *Wu and Johnston*, 2007]. *Coron et al*. [2012] also noted the potential for seasonal variations in parameters, due to changes in the dominant runoff generation mechanisms between seasons in some catchments. Lastly, the adoption of a time invariant optimal parameter distribution does not necessarily guarantee consistent model performance even within the calibration period. The chosen objective function strongly influences the calibration parameters [*Thyer et al*., 2009; *Madsen*, 2003; *Wang et al*., 2010; *Pathiraja et al*., 2012], with emphasis often placed on a particular streamflow characteristic (e.g., peak flows). It is near impossible for a time invariant set of model parameters to accurately simulate all

characteristics of the time series of system states (such as peak flow, base flows, volume of runoff, soil moisture storage fluctuations etc.) [*Moussa and Chahinian*, 2009; *Efstratiadis and Koutsoyiannis*, 2010; *Westerberg et al*., 2011]. Multiobjective criterion type approaches have shown promise in defining parameter spaces which provide improved simulations over a range of characteristics, but trade-offs are almost always required [*Madsen*, 2003; *Efstratiadis and Koutsoyiannis*, 2010].

One possible approach to improve the applicability of hydrologic models over long time scales is to allow the model parameters to evolve with time. Time varying parameter systems have been investigated in fundamental systems theory over the last few decades [e.g., *Richards*, 1983; *Schwartz and Ozbay*, 1990; *Mohammadpour and Scherer*, 2012], the most common being Linear Time Variant systems where the state space transition matrix is time varying [e.g., *Khalil*, 1996]. A well-known example of a time varying parameter system is the variation in aerodynamic coefficients during take-off, cruising and landing for high speed aircraft [e.g., *Tomas-Rodriguez and Banks*, 2010]. In hydrologic applications, model parameterizations would adjust to varying climatic regimes and/or changes in catchment conditions based on signals in the input and observed data under such a framework. It can be argued that time variations in model parameters are an artefact of missing system processes. However, it is not always tractable to identify the "correct" system equations to model such missing processes. Allowing the model parameters to vary provides a mechanism with which to adjust the model structure to reflect such missing processes, when they become evident in the observations. Only a few studies have examined strategies for diagnosing parameter nonstationarity in hydrologic systems. The most common approach in the literature involves split-sample calibration, whereby the available data set is subdivided and calibration undertaken separately for each period [e.g., *Thirel et al*., 2015; *Merz et al*., 2011; *de Vos et al*., 2010; *Gharari et al*., 2013; *Seibert and McDonnell*, 2010]. *Westra et al*. [2014] examined the matter from a different perspective, by assuming one or more parameters vary in time as a function of selected covariates. Their experiment on the Scott Creek catchment showed an improvement in model predictions when the soil storage parameter in the GR4J model was made to vary based on seasonal, annual and long term trends. *Jeremiah et al*. [2013] and *Marshall et al*. [2006] modeled parameter variations as a function of the catchment state in a Hierarchical Mixtures of Experts (HME) framework.

Another possible approach, which has rarely been explored, is to extract parameter variation signals from hydrologic observations using Data Assimilation. Data Assimilation (DA) offers a framework to detect the potentially time varying nature of model parameters, by adjusting them in real time as observations become available [*Liu and Gupta*, 2007; *Evensen*, 2006]. It allows for the rapid quantification of both model parameters and prognostic variables in real time using uncertain observations [*Evensen*, 2006; *Liu et al*., 2012; *Gelb*, 1974]. Prior (a.k.a. background) estimates of a system variable (usually model simulations) are combined with observation(s), both of which are assumed to contain random errors, to produce an updated estimate with lower error variance than the background or observation [*Evensen*, 2006]. The widely used Kalman Filter (KF) [*Kalman*, 1960] provides optimal updates for linear dynamical systems with normally distributed variables and errors [*Kalman*, 1960; *Evensen*, 2006]. However, most real world applications do not satisfy these assumptions. A plethora of DA algorithms have been developed to accommodate less stringent conditions, including Kalman Filter variants (e.g., the Extended Kalman Filter [*Jazwinski*, 1970], Ensemble Kalman Filter [*Evensen*, 2006], and the Unscented Kalman Filter [*Wan and Van Der Merwe*, 2000]) and Particle Filters (e.g., Particle Filter-SIR [*Moradkhani et al*., 2005a], Regularised Particle Filter [*Musso et al*., 2001], and the Particle Filter-MCMC [*Moradkhani et al*., 2012]). Although such methods have been widely applied in the geosciences, they are suboptimal and there is no clear consensus on the most appropriate methods for use in complex nonlinear systems with nonnormal distributions. The Ensemble Kalman Filter has been used frequently in hydrologic applications with considerable success, despite its inability to adequately accommodate nonnormal distributions [see e.g., *Komma et al*., 2008; *Reichle et al*., 2002; *Weerts and El Serafy*, 2006]. Data Assimilation algorithms in general have predominantly been used in hydrology for state updating, for instance for characterizing soil moisture [*Houser et al*., 1998; *Walker et al*., 2001; *Matgen* et al., 2012], rainfall runoff modeling [*Aubert et al*., 2003; *Weerts and El Serafy*, 2006] and flood forecasting [*Vrugt et al*., 2006; *Noh et al*., 2013; *Li et al*., 2013] (for a detailed review, refer to *Liu and Gupta* [2007]). This has in recent years been extended to include parameter estimation within the DA framework, although it has mainly been limited to deriving stationary parameter distributions [e.g., *Maneta and Howitt*, 2014; *Moradkhani et al*., 2005b; *Simon and Bertino*, 2012; *Yang and Delsole*, 2009; *Aksoy et al*., 2006; *Xie et al*., 2014; *Annan et al*., 2005]. One of the few applications of DA for time varying parameter estimation in

hydrology is the work of *Smith et al.* [2008]. They examined the time evolution of parameter distributions from particle filtering as a way of diagnosing model structural inadequacy. Rapid parameter fluctuations in time were noted, although it is unclear to what extent this is driven by nonuniqueness of parameter solutions and the stochastic nature of the DA algorithm. *Vrugt et al.* [2013] and *Salamon and Feyen* [2009] also applied Particle Filter based algorithms to examine the time evolution of model parameters. However, none of these studies addressed the question of whether a structured time variation in a parameter can be detected using DA (e.g., in *Vrugt et al.* [2013] the true parameters were assumed constant in the synthetic study using HyMOD). *Lin and Beck* [2007] used a recursive parameter evolution algorithm to estimate two state variables and two parameters (one time varying, one time invariant) in a biomass-substrate model. The algorithm was shown to successfully replicate the temporal pattern of the parameters in the synthetic case study. However its effectiveness remains to be seen when the dimensionality of time varying parameters is increased and the number of observation variables is reduced.

In this paper, we examine the potential of Data Assimilation for estimating time varying hydrologic model parameters. It is demonstrated that the parameter evolution model, which is used to generate background parameter ensembles, is critical in determining the effectiveness of the filter in this regard. A number of commonly used parameter evolution models are applied within an EnKF framework. They are used to detect time variations in parameters of the Probability Distributed Model (PDM) [*Moore*, 2007], a lumped conceptual hydrologic model. Parameter estimation is undertaken for multiple scenarios with synthetic data, so as to allow for a robust assessment of algorithm performance. The aim here is to investigate the efficacy of existing parameter evolution models for a range of parameter temporal patterns. Two alternative parameter evolution models are proposed which are specifically suited to time varying parameter applications. The first is a Multilayered approach that uses the EnKF to estimate hyperparameters of the temporal structure, based on apriori knowledge of the form of nonstationarity. The second is a Locally Linear approach that uses local linear estimation and requires no assumptions of the form of parameter nonstationarity.

The remainder of this paper is structured as follows. In section 2, we describe the Joint State-Parameter Estimation framework, including the Dual EnKF algorithm of *Moradkhani et al.* [2005b] and the various parameter evolution models examined in the study. Section 3 provides details of the experimental setup of the synthetic study, and the known time variations imposed on model parameters for the various case studies. Results from all parameter evolution models considered are provided in section 4, along with a discussion of their efficacy in various scenarios. Finally, we conclude with a summary of the main outcomes of the study as well as proposed future work.

2. Joint State Parameter Estimation with the EnKF

The Ensemble Kalman Filter (EnKF) is a Monte Carlo application of the well-known Kalman Filter [*Evensen*, 2006]. In the EnKF, distributions of the system variables are replaced with random samples or an ensemble. This means that unlike the original Kalman Filter [*Kalman*, 1960] or the Extended Kalman Filter [*Jazwinski*, 1970], the model error covariance matrix can be approximated by the sample covariance, instead of being specified apriori and explicitly propagated forward in time [*Evensen*, 2006]. The sample covariance of the system variables provide an estimate of the error covariance, assuming unbiasedness. The main limitation of the EnKF is its suboptimality for nonlinear dynamics with nonnormally distributed errors [*Evensen*, 2006]. However several studies have demonstrated the usefulness of the EnKF in a variety of highly nonlinear systems [see e.g., *Komma et al.*, 2008; *Reichle et al.*, 2002; *Weerts and El Serafy*, 2006]. For further details on the EnKF, refer to *Evensen* [2006].

The EnKF and its variants have been successfully applied for estimating both system state variables and time invariant model parameters [e.g., *Maneta and Howitt*, 2014; *Moradkhani et al.*, 2005b; *Simon and Bertino*, 2012; *Yang and Delsole*, 2009; *Aksoy et al.*, 2006; *Samuel et al.*, 2014]. Simultaneous estimation of states and parameters in a DA framework is traditionally undertaken by one of two methods: (1) augmenting the state vector with parameter variables so that updating of states and variables occurs concurrently [e.g., *Smith et al.*, 2013; *Franssen and Kinzelbach*, 2008]; or (2) Updating parameters and states separately through two sequential filters, also known as Dual State – Parameter estimation [*Moradkhani et al.*, 2005b; *Lü et al.*, 2011; *Leisenring and Moradkhani*, 2012]. The parameter update equation in the state augmentation

and dual updating approaches are equivalent when applied with Kalman filtering methods (refer Appendix A). However, unlike the augmented approach, dual filtering allows biases in the model simulations of system states to be removed (at least partially) prior to state updating. This is important, as the successful implementation of any DA algorithm is reliant on unbiased priors [*Dee*, 2005]. We therefore adopt the dual state – parameter estimation approach in this study.

A general framework for undertaking dual State - Parameter estimation using the EnKF was described in *Moradkhani et al*. [2005b], although this focused on estimating a stationary distribution of parameters. Here a methodology is presented for the estimation of time varying parameters in a dual updating framework. The algorithm is summarized below.

2.1. Dual EnKF

Suppose a dynamical system at any given time t can be described by a vector of states $\boldsymbol{x}_t$ and a vector of associated model parameters $\boldsymbol{\theta}_t$. The distributions of system variables at time t are represented by an ensemble of states $\{\boldsymbol{x}_t^i\}_{i=1:n}$ and parameters $\{\boldsymbol{\theta}_t^i\}_{i=1:n}$ each with n members. A plus (+) superscript indicates an updated ensemble whereas a minus (-) superscript indicates the background or prior ensemble. Hereafter, we use the term "prior" to refer to the background or model simulated ensemble and "update" to refer to the analysis or updated ensemble after assimilation. The Dual EnKF procedure is undertaken as follows:

1. *Generate a prior parameter ensemble*. At any given time t, begin by generating a prior parameter ensemble:

$$\boldsymbol{\theta}_{t+1}^{i-} = g\left(\boldsymbol{\theta}_t^{i+},\ z\right) \quad for\ i=1:n \tag{1}$$

   where g is a parameter evolution model to be chosen by the user, with input parameters z. Parameter evolution is discussed in more detail in section 2.2.

2. *Incorporate observation and forcing uncertainty*. A set of perturbed observations $\{\boldsymbol{y}_{t+1}^i\}_{i=1:n}$ is generated using the measurement error characteristics. Here we adopt the method of *Houtekamer and Mitchell* [1998] and *Burgers et al*. [1998] for incorporating observation uncertainty into the analysis ensemble. Stochastic perturbations are applied to the observation vector to produce an ensemble of observations using the assumed observation error statistics:

$$\boldsymbol{y}_{t+1}^i = \boldsymbol{y}_{t+1}^o + \epsilon_{t+1}^i \quad for\ i=1:n \tag{2}$$

$$\epsilon_{t+1}^i \sim N\left(0,\ \boldsymbol{\Sigma}_{t+1}^{y^o y^o}\right) \tag{3}$$

   where $\boldsymbol{y}_{t+1}^o$ denotes the raw observation and $\boldsymbol{\Sigma}_{t+1}^{y^o y^o}$ denotes the observation error covariance matrix. A similar approach can be adopted to account for forcing error (refer *Moradkhani et al*. [2005b]) to generate a set of perturbed forcings $\{\boldsymbol{u}_{t+1}^i\}_{i=1:n}$. Zero forcing error is assumed in this study, as the focus is on the rainfall to runoff conversion processes and parameter adjustments to compensate for forcing error are undesired.

3. *Generate simulated observations using the prior parameters*. The updated parameter ensemble is determined by first calculating simulated observations using the best available estimate of model states and prior parameters:

$$\hat{\boldsymbol{x}}_{t+1}^i = f\left(\boldsymbol{x}_t^{i+},\ \boldsymbol{\theta}_{t+1}^{i-},\ \boldsymbol{u}_{t+1}^i\right) \quad for\ i=1:n \tag{4}$$

$$\hat{\boldsymbol{y}}_{t+1}^i = h\left(\hat{\boldsymbol{x}}_{t+1}^i,\ \boldsymbol{\theta}_{t+1}^{i-}\right) \quad for\ i=1:n \tag{5}$$

   where f is the set of model equations, h is an observation operator that converts model states to observed variables, $\hat{\boldsymbol{y}}_{t+1}^i$ indicates the ith ensemble member of the model simulation of the observed variable using the prior parameters, and $\boldsymbol{u}_{t+1}^i$ is a vector of forcing inputs (e.g., rainfall and evapotranspiration).

4. *Perform the Kalman update of the parameters*. The prior parameter ensemble is then updated using the Kalman update equation and the covariance between parameters and simulated observations:

$$\boldsymbol{\theta}_{t+1}^{i+} = \boldsymbol{\theta}_{t+1}^{i-} + \mathbf{K}_{t+1}^{\theta}\left(\boldsymbol{y}_{t+1}^i - \hat{\boldsymbol{y}}_{t+1}^i\right) \quad for\ i=1:n \tag{6}$$

$$\mathbf{K}_{t+1}^{\theta} = \boldsymbol{\Sigma}_{t+1}^{\theta\hat{y}}\left[\boldsymbol{\Sigma}_{t+1}^{\hat{y}\hat{y}} + \boldsymbol{\Sigma}_{t+1}^{y^o y^o}\right]^{-1} \tag{7}$$

   where $\boldsymbol{\Sigma}_{t+1}^{\theta\hat{y}}$ is a matrix of the cross covariance between parameters $\{\boldsymbol{\theta}_{t+1}^{i-}\}_{i=1:n}$ and simulated observed variables $\{\hat{\boldsymbol{y}}_{t+1}^i\}_{i=1:n}$ and $\boldsymbol{\Sigma}_{t+1}^{\hat{y}\hat{y}}$ is the covariance matrix of the simulated observations.

5. *Generate simulated observations using the updated parameters*. The prior state ensemble $\{\boldsymbol{x}_{t+1}^{i-}\}_{i=1:n}$ is then generated using the updated model parameter ensemble $\{\boldsymbol{\theta}_{t+1}^{i+}\}_{i=1:n}$ and the model equations:

$$\boldsymbol{x}_{t+1}^{i-} = f\left(\boldsymbol{x}_t^{i+},\ \boldsymbol{\theta}_{t+1}^{i+},\ \boldsymbol{u}_{t+1}^{i}\right) \quad for\ i=1:n \tag{8}$$

$$\tilde{\boldsymbol{y}}_{t+1}^{i} = h\left(\boldsymbol{x}_{t+1}^{i-},\ \boldsymbol{\theta}_{t+1}^{i+}\right) \quad for\ i=1:n \tag{9}$$

where $\tilde{\boldsymbol{y}}_{t+1}^{i}$ indicates the simulated observed variable using the updated parameters.

6. *Perform the Kalman update of the states*. The Kalman equation for correlated measurement and process noise is used to update states. This is because the standard Kalman equation (equations (6) and (7)) assumes that the errors in the prior and observation are independent. This assumption is no longer valid in the dual updating approach, as the simulated states have been generated from parameters that were updated using observations from the same time step. Therefore, updating is undertaken taking into account the potential correlation between observation and prior noise:

$$\boldsymbol{x}_{t+1}^{i+} = \boldsymbol{x}_{t+1}^{i-} + \mathbf{K}_{t+1}^{x}\left(\boldsymbol{y}_{t+1}^{i} - \tilde{\boldsymbol{y}}_{t+1}^{i}\right) \quad for\ i=1:n \tag{10}$$

$$\mathbf{K}_{t+1}^{x} = \left[\boldsymbol{\Sigma}_{t+1}^{x\tilde{y}} + \boldsymbol{\Sigma}_{t+1}^{\varepsilon_x y^o}\right]\left[\boldsymbol{\Sigma}_{t+1}^{\tilde{y}\tilde{y}} + \boldsymbol{\Sigma}_{t+1}^{\varepsilon_{\tilde{y}} y^o} + \left(\boldsymbol{\Sigma}_{t+1}^{\varepsilon_{\tilde{y}} y^o}\right)^{\mathbf{T}} + \boldsymbol{\Sigma}_{t+1}^{y^o y^o}\right]^{-1} \tag{11}$$

$$\varepsilon_{x}{}_{t+1}^{i} = \boldsymbol{x}_{t+1}^{i-} - \hat{\boldsymbol{x}}_{t+1}^{i};\ \ \varepsilon_{\tilde{y}}{}_{t+1}^{i} = \tilde{\boldsymbol{y}}_{t+1}^{i} - \hat{\boldsymbol{y}}_{t+1}^{i} \tag{12}$$

where $\boldsymbol{\Sigma}_{t+1}^{x\tilde{y}}$ is a matrix of the cross covariance between states $\{\boldsymbol{x}_{t+1}^{i-}\}_{i=1:n}$ and simulated observed variables $\{\tilde{\boldsymbol{y}}_{t+1}^{i}\}_{i=1:n}$ from Step 5; $\boldsymbol{\Sigma}_{t+1}^{\varepsilon_x y^o}$ represents the covariance between $\{\varepsilon_{x}{}_{t+1}^{i}\}_{i=1:n}$ and the observations; $\boldsymbol{\Sigma}_{t+1}^{\varepsilon_{\tilde{y}} y^o}$ represents the covariance between the $\{\varepsilon_{\tilde{y}}{}_{t+1}^{i}\}_{i=1:n}$ and the observations; and $()^{\boldsymbol{T}}$ represents the transpose operator. Note that only the Kalman gain matrix is adjusted, the form of the Kalman update equation is the same as the standard Kalman equation. For further details on equations (10)–(12), refer to Appendix B.

2.2. Existing Parameter Evolution Models

The parameter evolution model (equation (1)) is required to generate a suitable prior parameter ensemble at each time. In the case of time invariant parameters, a logical choice for the prior parameter ensemble, $\{\boldsymbol{\theta}_{t+1}^{i-}\}_{i=1:n}$ would be the updated parameter ensemble from the previous time step, $\{\boldsymbol{\theta}_{t}^{i+}\}_{i=1:n}$. However, this approach can result in a steadily reducing sample variance, since the variance of an update is always smaller than its background. If the background ensemble variance is small compared to the observation error variances, the updated ensemble will be dominated by the model simulations, which are seen to have high confidence because of their small error variance. This slow loss of sensitivity to observations as a result of reduced background variance is known as filter divergence [*Anderson*, 2007]. A common method of dealing with this issue is to introduce stochastic noise into the background parameter ensemble [*Gordon et al*., 1993]. Here we examine the performance of three existing parameter perturbation techniques when applied to estimating nonstationary parameters. These are:

1. *Standard Kernel Smoother (SKS)* [*West*, 1993]. In this approach, individual ensemble members are drawn from a truncated multivariate normal distribution with heteroscedastic covariance:

$$\boldsymbol{\theta}_{t+1}^{i-} \sim TMVN\left(\boldsymbol{\theta}_{t}^{i+},\ s^2\boldsymbol{\Sigma}_{t}^{\theta}\right) \tag{13}$$

where $\boldsymbol{\Sigma}_{t}^{\theta}$ is the covariance matrix of the updated parameter ensemble at time t and s^2 is a smoothing parameter which must be tuned by the user. The resulting prior parameter distribution, $p\left(\boldsymbol{\theta}_{t+1}^{-}\right)$ is a mixture of normals with variance greater than the previous updated ensemble, $\{\boldsymbol{\theta}_{t}^{i+}\}_{i=1:n}$ [*Liu and West*, 2001]. For this study, the SKS has been modified so that parameters are instead sampled from independent univarate truncated normal distributions, since parameters are forced to vary independently in the synthetic case study (refer section 3). The bounded nature of the PDM model parameters necessitates the use of a truncated normal distribution.

2. *Kernel Smoother with location shrinkage (KSLS)* [*Liu and West*, 2001]. This technique has been used in a number of studies [see e.g., *Maneta and Howitt*, 2014; *Moradkhani et al*., 2005b; *Xie et al*., 2014]. As stated above and in *Liu and West* [2001], the Standard Kernel Smoother produces a prior ensemble with increased variance compared to the updated ensemble from the previous time step. *Liu and West* [2001]

argued that this increased variance would result in a loss of information over time. They suggested that the mean of the Standard Kernel Smoother should be adjusted according to equation (14), so that the posterior variance from the previous time step would be maintained. For further details of the kernel smoother with location shrinkage, refer to [*Liu and West*, 2001].

$$\boldsymbol{\theta}_{t+1}^{i-} \sim TMVN\left(a\boldsymbol{\theta}_t^{i+} + (1-a)\overline{\boldsymbol{\theta}}_t^{+},\ h^2\boldsymbol{\Sigma}_t^{\theta}\right) \quad \text{for } i=1::n \tag{14}$$

where $h^2=1-a^2$, $a=\frac{3\delta-1}{2\delta}$, $\delta \in \left[\frac{1}{5},\ 1\right]$ $\overline{\boldsymbol{\theta}}_t^{+}$ is the updated parameter ensemble mean vector at time t, $\boldsymbol{\Sigma}_t^{\theta}$ is the covariance matrix of the updated parameter ensemble at time t and δ is a scalar smoothing parameter, typically between 0.95 and 0.99.

3. *Homoscedastic Kernel Smoother (HKS)* [e.g., *Smith et al.*, 2008]. This method is similar to the Standard Kernel Smoother, with the important distinction that the covariance of the sampling distribution is fixed in time. The parameters are evolved in time using independent truncated normal distributions with standard deviations fixed in time and equal to a fraction of the feasible parameter range:

$$\boldsymbol{\theta}_{t+1}^{i-} \sim TMVN\left(\boldsymbol{\theta}_t^{i+},\ \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right) \tag{15}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \begin{pmatrix} (\boldsymbol{v}.\Delta_{\theta_1})^2 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & (\boldsymbol{v}.\Delta_{\theta_m})^2 \end{pmatrix}$

Δ_{θ_j} is the width of the feasible parameter range for the jth parameter variable and v is a user defined fraction of the feasible parameter range (between 0 and 1, *Smith et al.* [2008] adopted 0.2 in their study.)

2.3. A Multilayer EnKF Approach

Here we propose an alternative parameter evolution approach that utilizes apriori knowledge of the form of parameter nonstationarity. Just as prior states (ie. the model simulations) are determined from model equations, improved prior parameters could be derived by evolving parameters through a persistence model instead of an arbitrary random walk. In this approach, the functional form of parameter variation with time is assumed to be known and incorporated into the EnKF. For instance, suppose a parameter θ varies linearly with time, its functional form would be $\theta(t) = \theta(t-1) + a$ where a is a time invariant hyperparameter to be estimated. The dual layer filter described in section 2.1 is extended to three sequential filters, so that the hyperparameters that describe the time evolution of parameters are estimated through the first level filter. This is followed by parameter and state estimation in the second and third level filters respectively. This Multilayer method (ML) is undertaken as follows:

1. Apply the Standard Kernel Smoothing approach (SKS) to generate the background hyperparameter ensemble, $\left\{\boldsymbol{\psi}_{t+1}^{i-}\right\}_{i=1:n}$ which is assumed time invariant (refer equation (16)). Note we propose using the SKS because of the increased susceptibility of the KSLS to filter divergence, which can produce biased ensemble means following convergence.

$$\boldsymbol{\psi}_{t+1}^{i-} \sim TMVN\left(\boldsymbol{\psi}_t^{i+},\ s^2\boldsymbol{\Sigma}_t^{\psi}\right) \quad \text{for } i=1:n \tag{16}$$

2. Estimate parameters $\breve{\boldsymbol{\theta}}_{t+1}^{i}$ using the background hyperparameters $\boldsymbol{\psi}_{t+1}^{i-}$ and assumed parameter persistence equations l:

$$\breve{\boldsymbol{\theta}}_{t+1}^{i} = l\left(\boldsymbol{\theta}_t^{i+}, \boldsymbol{\psi}_{t+1}^{i-}\right) \quad \text{for } i=1:n \tag{17}$$

3. Calculate model states $\breve{\boldsymbol{x}}_{t+1}^{i}$ using the model equations f, parameters $\breve{\boldsymbol{\theta}}_{t+1}^{i}$ and forcing data $\boldsymbol{u}_{t+1}^{i}$. Determine simulated observations $\breve{\boldsymbol{y}}_{t+1}^{i}$ using the observation operator h, and model states $\breve{\boldsymbol{x}}_{t+1}^{i}$ and parameters $\breve{\boldsymbol{\theta}}_{t+1}^{i}$:

$$\breve{\boldsymbol{x}}_{t+1}^{i} = f\left(\boldsymbol{x}_t^{i+}, \breve{\boldsymbol{\theta}}_{t+1}^{i}, \boldsymbol{u}_{t+1}^{i}\right) \quad \text{for } i=1:n \tag{18}$$

$$\breve{\boldsymbol{y}}_{t+1}^{i} = h\left(\boldsymbol{x}_{t+1}^{i},\ \breve{\boldsymbol{\theta}}_{t+1}^{i}\right) \quad \text{for } i=1:n \tag{19}$$

4. Update the hyperparameters $\boldsymbol{\psi}_{t+1}^{i-}$ using the perturbed observations $\boldsymbol{y}_{t+1}^{i}$ (as discussed in section 2.1) and the simulated observations $\breve{\boldsymbol{y}}_{t+1}^{i}$ in the standard Kalman update equation:

$$\psi_{t+1}^{i+}=\psi_{t+1}^{i-}+K_{t+1}^{\psi}\left(y_{t+1}^{i}-\breve{y}_{t+1}^{i}\right) \quad \text{for } i=1:n \tag{20}$$

$$K_{t+1}^{\psi}=\Sigma_{t+1}^{\psi\breve{y}}\left[\Sigma_{t+1}^{\breve{y}\breve{y}}+\Sigma_{t+1}^{y^o}\right]^{-1} \tag{21}$$

where $\Sigma_{t+1}^{\psi\breve{y}}$ is a matrix of the cross covariance between hyperparameters $\left\{\psi_{t+1}^{i-}\right\}_{i=1:n}$ and simulated observed variables $\left\{\breve{y}_{t+1}^{i}\right\}_{i=1:n}$; and $\Sigma_{t+1}^{\breve{y}\breve{y}}$ is the covariance matrix of the simulated observations.

5. Generate the background parameters θ_{t+1}^{i-} using the updated hyperparameters ψ_{t+1}^{i+} and the parameter persistence equations I:

$$\theta_{t+1}^{i-}=I\left(\theta_{t}^{i+},\ \psi_{t+1}^{i+}\right) \quad \text{for } i=1:n \tag{22}$$

6. Generate model simulations of the observed variables $\left\{\hat{y}_{t+1}^{i}\right\}_{i=1:n}$ using the background parameter ensemble, $\left\{\theta_{t+1}^{i-}\right\}_{i=1:n}$ as per Step 2 in section 2.1.
7. Update the model parameters using the Kalman update equation for correlated process and measurement noise (equations (23) and (24)). Equations (23) and (24) are used in place of equations (6) and (7) as the errors in the simulated observed variable and observations are no longer independent.

$$\theta_{t+1}^{i+}=\theta_{t+1}^{i-}+K_{t+1}^{\theta}\left(y_{t+1}^{i}-\hat{y}_{t+1}^{i}\right) \quad \text{for } i=1:n \tag{23}$$

$$K_{t+1}^{\theta}=\left[\Sigma_{t+1}^{\theta\hat{y}}+\Sigma_{t+1}^{\varepsilon_\theta y^o}\right]\left[\Sigma_{t+1}^{\hat{y}\hat{y}}+\Sigma_{t+1}^{\varepsilon_{\hat{y}} y^o}+\left(\Sigma_{t+1}^{\varepsilon_{\hat{y}} y^o}\right)^{T}+\Sigma_{t+1}^{y^o y^o}\right]^{-1} \tag{24}$$

$$\varepsilon_{\theta}{}_{t+1}^{i}=\theta_{t+1}^{i-}-\breve{\theta}_{t+1}^{i};\ \ \varepsilon_{\hat{y}}{}_{t+1}^{i}=\hat{y}_{t+1}^{i}-\breve{y}_{t+1}^{i} \tag{25}$$

where $\Sigma_{t+1}^{\theta\hat{y}}$ is a matrix of the cross covariance between parameters $\left\{\theta_{t+1}^{i-}\right\}_{i=1:n}$ and simulated observed variables $\left\{\hat{y}_{t+1}^{i}\right\}_{i=1:n}$ from Step M6; $\Sigma_{t+1}^{\varepsilon_\theta y^o}$ represents the covariance between $\left\{\varepsilon_{\theta}{}_{t+1}^{i}\right\}_{i=1:n}$ and the observations; $\Sigma_{t+1}^{\varepsilon_{\hat{y}} y^o}$ represents the covariance between $\left\{\varepsilon_{\hat{y}}{}_{t+1}^{i}\right\}_{i=1:n}$ and the observations; and $()^{T}$ represents the transpose operator.

8. Update states as per Steps 5 and 6 in section 2.1.

2.4. A Locally Linear Approach

We also propose a second parameter evolution model, which unlike the Multilayer EnKF, requires no assumptions about the form of parameter nonstationarity. This method involves linearly extrapolating the ensemble mean of the updated parameters from the previous two time steps to propose a background ensemble, as demonstrated in Figure 1. If the change in updated parameters over a single time step exceeds a predefined threshold value, the gradient of the linear fit is determined based on the preceding updated parameters. This minimizes both overfitting and avoids parameter drift due to isolated large updates. Parameter values and the linearized rate of change of parameters at time $t = 0$ are required to initialize the algorithm.

The locally linear approach (LL) is undertaken as follows:

1. Apply the SKS to generate an ensemble of background parameters for any given time $t+1$, $\left\{\theta_{\boldsymbol{int},\ t+1}^{\boldsymbol{i}}\right\}_{i=1:n}$:

$$\theta_{\boldsymbol{int},\ t+1}^{\boldsymbol{i}}\sim TMVN\left(\theta_{\boldsymbol{t}}^{\boldsymbol{i}+},\ s^2\Sigma_{t}^{\theta}\right) \quad \text{for } i=1:n \tag{26}$$

where s^2 is a smoothing parameter which must be tuned by the user. For this study, the parameters are instead sampled from independent univarate truncated normal distributions, since parameters are forced to vary independently in the synthetic case study (refer section 3).

2. Calculate the gradient of the linear function fitted to the updated parameter ensemble mean from time *t-1* and time t. Repeat for time *t-2* and time *t-1*:

$$m_t=\frac{\left(\overline{\theta_{\boldsymbol{t}}^{+}}-\overline{\theta_{\boldsymbol{t-1}}^{+}}\right)}{\Delta_t} \ \text{ and } \ m_{t-1}=\frac{\left(\overline{\theta_{\boldsymbol{t-1}}^{+}}-\overline{\theta_{\boldsymbol{t-2}}^{+}}\right)}{\Delta_{t-1}} \tag{27}$$

where $\Delta_t=t-(t-1)$ and $\Delta_{t-1}=(t-1)-(t-2)$.

3. Define the gradient for linearly extrapolating the updated parameter ensemble mean at time $t+1$:

$$m_{\boldsymbol{t+1}}=\begin{cases} m_t, & |m_t|\ \le\ m_{\max} \\ m_{t-1}, & |m_t|\ >\ m_{\max} \end{cases} \tag{28}$$

where m_{max} is a user defined positive which indicates the maximum allowable jump over a single time step.

4. Shift the parameter ensemble from Step 1, $\theta^i_{int,\ t+1}$, so that $\overline{\theta^-_{t+1}} = \overline{\theta^+_t} + \boldsymbol{m}_{t+1}.\ \Delta_{t+1}$, that is, calculate:

$$\theta^{i-}_{t+1} = \theta^i_{int,\ t+1} + \left(\overline{\theta^+_t} + \boldsymbol{m}_{t+1}.\ \Delta_{t+1} - \overline{\theta_{int,t+1}}\right) \quad for\ i=1:n \tag{29}$$

3. Synthetic Case Studies With the Probability Distributed Model

The ability of the Dual EnKF to estimate nonstationary model parameters was investigated through an application to the lumped conceptual hydrologic model, PDM [*Moore*, 2007]. The PDM consists of three buckets representing the main storage and routing components of the catchment, ie. soil moisture, surface runoff routing and groundwater flow. In this version, soil storage capacity is assumed to be Pareto distributed with shape parameter b and maximum point soil storage depth c_{max}. Excess rainfall is partitioned into a surface runoff store and groundwater store through the splitting parameter α. Outflow from these two linear routing tanks is controlled by parameters k_s (for the surface runoff store) and k_b (for the groundwater store). A schematic of the model is shown in Figure 2.

State and parameter estimation was undertaken by assimilating streamflow observations into the PDM at a daily time step for approximately 3 years. In order to properly assess algorithm performance, the true system states and model parameters must be known apriori. Synthetic daily rainfall, potential evapotranspiration (PET) and model parameters were therefore specified for use in the PDM to develop the "true" system variables. Catchment averaged rainfall derived from the 5km x 5km gridded daily rainfall SILO database [*Jeffrey et al*., 2001] for the Snowy Creek Catchment in Victoria, Australia was used as the basis for synthetic rainfall. Extended periods of no rainfall were replaced with rainfall events from wet years, to minimize periods of zero observability associated with dry conditions, and to avoid the need for likelihood functions designed to operate for such ephemeral catchments [*Smith et al*., 2010]. Zero (or small) streamflow values contain no information about the model parameters, meaning that extended dry periods create additional challenges for the time varying parameter estimation problem. Daily PET data were developed by uniformly disaggregating monthly average PET data from evapotranspiration maps published by the Australian Bureau of Meteorology as part of their Climate Atlas Series (see http://www.bom.gov.au/climate/how/newproducts/IDCetatlas.shtml). Mean annual PET for the synthetic time series is approximately 1000 mm/yr whilst mean annual rainfall is 1800 mm/yr. Multiple time varying parameter scenarios were developed in order to assess algorithm performance for a range of situations, as detailed in Table 1. These include parameter combinations with a strong signal in the streamflow (Scenarios 1 and 2) and a weaker signal in the streamflow (Scenarios 3 and 4). Signal strength is defined by the magnitude of runoff residuals when parameters are assumed constant and equal to their initial values (refer supporting information Figure S1). We considered a sinusoidal variation in the model parameters, following the work of *Paik et al. [*2005], *Coron et al*. [2012] and *Ye et al*. [1997] who noted the potential for seasonal variations in hydrologic model parameters; and also *He et al*. [2011] where snow melt factors are treated as sinusoidally varying. Linear trends in model parameters corresponding to a steady gradual change in land cover (e.g., afforestation/deforestation or urbanization) [*Croke et al*., 2004] were also considered. Finally, parameter sets with different forms of nonstationarity (Scenario 3) and differing rates of change (Scenario 2) were also devised to assess the ability to detect temporally correlated parameter fluctuations.

3.1. Specification of EnKF components

We now describe the specification of the components of the EnKF, such as initialization and error characterization. In order to characterize the observation uncertainty, perturbed observation ensembles were generated following the method of *Houtekamer and Mitchell* [1998] and *Burgers et al*. [1998], using known error variances. A heteroscedastic error variance was considered for generating the synthetic observations, and defined as a proportion of the true streamflow, so that larger flows have greater errors than low flows. In this study, d was chosen to be 0.1, and the error characteristics assumed known. A truncated normal distribution was utilized to ensure positive streamflow observations:

$$q^i_{obs}(t) = q_{obs}(t) + \varepsilon^i_q \ \ where\ \varepsilon^i_q \ \sim\ TN(0,\ d\ \mathrm{x}\ q_{true}(t)) \quad for\ i=1:n \tag{30}$$

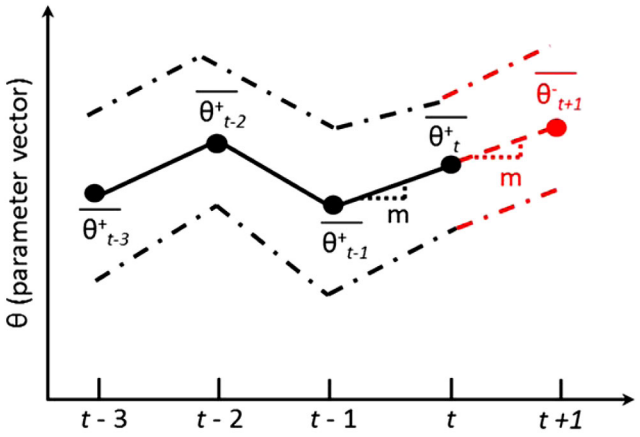

Figure 1. Schematic of the Locally Linear Approach for proposing background parameters. The circles indicate the parameter ensemble mean. The mean of the background parameters for the current time ($\overline{\theta_{t+1}^{-}}$) is estimated by linearly extrapolating from the updated parameter ensemble mean from the previous time step. The gradient (m) is calculated from the updated parameters from the previous two time steps (t and $t-1$). A background ensemble for time $t+1$ is generated using the Standard Kernel Smoother, and the entire ensemble shifted so that its mean is equal to $\overline{\theta_{t+1}^{-}}$.

Initial parameter ensembles were generated by sampling from a truncated normal distribution with mean equal to the true parameter value at $t = 0$ (see Table 2). Perfect initial conditions were adopted as the purpose of the study is to assess the performance of the filter in detecting parameter variations over time, rather than to correct for initial condition error. For applications involving real data, this would reflect a situation where model parameters have been well-calibrated over a stationary period and used to initialize the algorithm. Hyperparameters in the Multilayer approach are assumed time-invariant, therefore initial hyperparameter ensembles were developed by sampling from a uniform distribution with bounds determined based on the feasible parameter range and initial parameters. All hyperparameters were updated simultaneously. Linearly varying parameters have only one hyperparameter, a ($\theta_t = \theta_{t-1} + a$), whilst sinusoidally varying parameters have four hyperparameters, b, c, d and e ($\theta_t = b\sin\left(\frac{2\pi}{c}t + d\right) + e$). Initial states were also generated In order to minimize sampling error, an ensemble size of 100 members was adopted, based on the findings of *Moradkhani et al.* [2005b] and *Aksoy et al.* [2006]. Due to the stochastic-dynamic nature of the algorithm, ensemble statistics were calculated over 50 realizations of the dual filtering process. For the sake of comparison, the same set of perturbed observations and initial ensembles were used when evaluating the different parameter evolution models.

The difficulties in estimating time varying parameters with increasing dimensionality have been noted in previous studies [*Yang and Delsole*, 2009]. In order to examine the effects of increased dimensionality on algorithm performance, dual filtering for the scenarios described in Table 1 was undertaken for two case studies. The first involves updating only the routing parameters and states of the PDM (the surface runoff routing coefficient k_s, surface storage state S_s, groundwater routing coefficient k_b, groundwater storage state S_g, and the excess runoff splitting parameter α) in the EnKF, with the soil storage state and parameters assumed known. The second involves estimating all model states and parameters within the Dual EnKF framework.

All the parameter evolution models evaluated are parametric in that they require the specification of a tuning parameter (ie. s^2 for the SKS, Multilayer and Locally Linear methods; v for the HKS; and δ for the KSLS). An objective method for specifying such tuning parameters is through Variance Variable Multipliers [*Leisenring and Moradkhani*, 2012]. This approach considers the tuning parameters as dynamic variables, which are adjusted in time based on the absolute bias and the 95th percentile uncertainty bound. The absolute bias refers to the absolute difference between forecast and observed streamflow from previous time steps, whilst the 95th percentile uncertainty bound (ub_t) refers to the distance between the ensemble mean and 5th or 95th percentile of the forecast:

$$ub_t = \begin{cases} \overline{\hat{y}_t} - \hat{y}_t^{5} & \text{if } y_t < \overline{\hat{y}_t} \\ \hat{y}_t^{95} - \overline{\hat{y}_t} & \text{if } y_t > \overline{\hat{y}_t} \end{cases} \tag{31}$$

where $\overline{\hat{y}_t}$ = ensemble mean forecast streamflow at time t, $\hat{y}_t^{5}$ = 5th percentile forecast streamflow at time t, $\hat{y}_t^{95}$ = 95th percentile forecast streamflow at time t.

An initial tuning parameter value is supplied and increased at any given time if the absolute bias is greater than the 95th percentile uncertainty bound (indicating the ensemble variance is too small because the

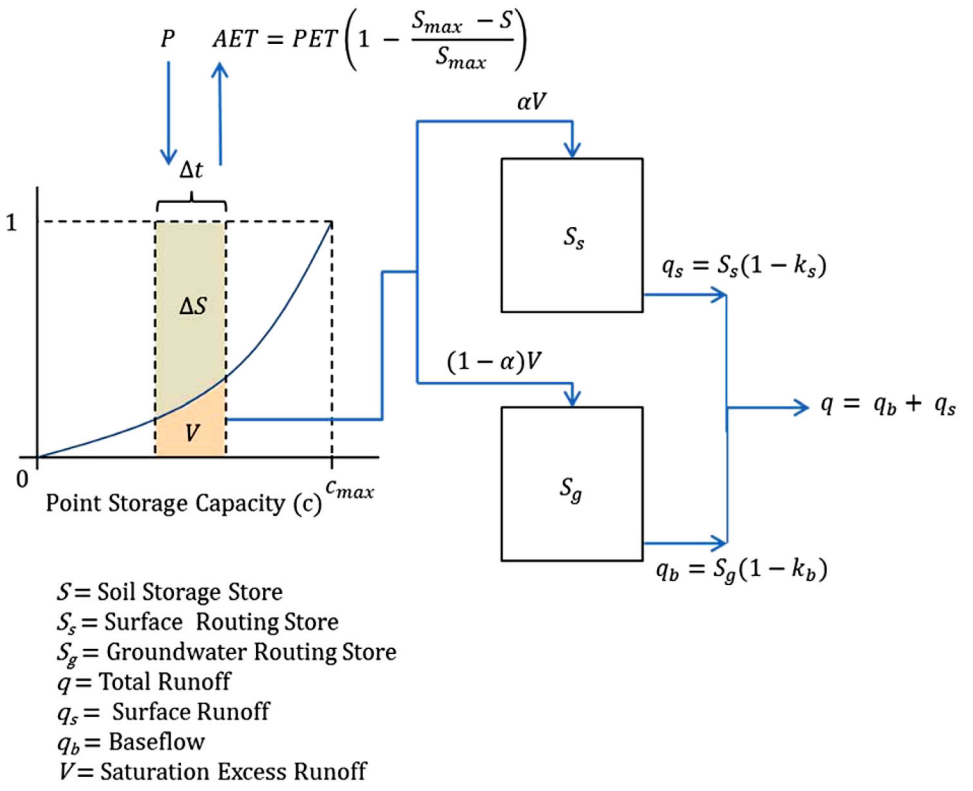

Figure 2. PDM Schematic (adapted from *Moore* [2007]).

observation lies outside the forecast ensemble) and decreased if the absolute bias is less than the 95th percentile uncertainty bound (indicating the ensemble variance is too large). For further details on Variable Variance Multipliers, refer to *Leisenring and Moradkhani* [2012]. Initial tuning parameter values of 0.03 for the SKS and LL methods and 0.002 for the ML method were adopted within the Variable Variance Multiplier framework. The remaining methods (KSLS and HKS) are not amenable to the VVM approach, hence, trial experiments with a range of tuning parameters were undertaken. For the KSLS, a range of values for δ between 0.95 and 0.99 (as recommended in *Liu and West* [2001]) were evaluated and found to produce similar parameter ensembles trajectories. In the HKS, $v = 0.2$ was tested following *Smith et al.* [2008], however it was found that smaller values were more suitable (refer section 4).

4. Results and Discussion

Clear differences between the estimation performance of each of the parameter evolution models were seen when detecting temporal patterns in parameters. The results demonstrate that the parameter evolution model, i.e., the mechanism for defining the background (or prior) parameter ensemble, has a significant impact on the detection of temporal patterns in parameters. A summary of the parameter estimation performance for the various scenarios is provided in Figure 3. The Nash Sutcliffe Efficiency (NSE) of the parameter ensemble mean and the percentage of time the parameter ensemble mean is within acceptable limits (denoted PTW here) are provided. Both statistics are provided as averages over the estimated parameters. Acceptable limits for each parameter are defined for time t as $\theta_t^* + \sigma d_p$ where θ_t^* is the true parameter value at time t, d_p is the feasible parameter range (Table 2) and σ is a fraction varying between 5% and 10%. Results are provided for each scenario for Case Study 1 (estimating routing parameters only) in the top quadrant and for Case Study 2 (estimating full model parameters) in the bottom quadrant. Average NSE values closer to 1 and Average PTW scores closer to 100% indicate better performance.

When calculating the average Nash Sutcliffe Efficiency over all the model parameters, the Multilayer approach gave higher average NSE values in comparison to the other parameter evolution models, regardless of the

Table 1. True Parameters and Impact on Streamflow for the Various Scenarios

	Scenario 1—High Impact on Q, Similar Parameter Variation	Scenario 2—High Impact on Q, Different Rates of Change	Scenario 3—Low Impact on Q, Different Persistence Types	Scenario 4—Low Impact on Q, Similar Parameter Variation
b	$b(t+1)=b(t)-2\times10^{-4}$	As per Scenario 1	As per Scenario 1	As per Scenario 1
c_{max}	$c_{max}(t+1)=c_{max}(t)+6.08\times10^{-2}$	As per Scenario 1	As per Scenario 1	As per Scenario 1
k_s	$k_s(t)=\ 0.1\sin\left(\frac{2\pi t}{300}+0\right)+0.5$	$k_s(t)=\ 0.1\sin\left(\frac{2\pi t}{200}-0.2\pi\right)+0.5$	As per Scenario 1	As per Scenario 1
k_b	$k_b(t)=\ 0.05\sin\left(\frac{2\pi t}{300}+0\right)+0.85$	$k_b(t)=\ 0.05\sin\left(\frac{2\pi t}{150}+0.4\pi\right)+0.85$	$k_b(t+1)=k_b(t)-1\times10^{-4}$	As per Scenario 1
α	$\alpha(t)=\ 0.15\sin\left(\frac{2\pi t}{300}+\pi\right)+0.7$	$\alpha(t)=\ 0.15\sin\left(\frac{2\pi t}{300}+0.2\pi\right)+0.7$	$\alpha(t+1)=\alpha(t)-2\times10^{-4}$	$\alpha(t)=\ 0.15\sin\left(\frac{2\pi t}{300}+0\right)+0.7$

parameter scenario or the case study investigated (Figure 3, left). The Locally Linear method provides good quality estimates for all scenarios involving the updating of routing parameters only (average NSE is 0.55). Both methods yield superior parameter estimation performance compared to the existing parameter evolution models (SKS, KSLS and HKS). However, the performance of the Locally Linear method degrades when the parameter dimensionality is increased, producing results similar to the SKS in the full model case study. We now examine the various parameter evolution models and test scenarios in further detail.

4.1. Estimation of Routing Parameters

We begin by considering the first case study, where the soil storage parameters are assumed known and only the routing parameters of the PDM are estimated for Scenarios 1–4. The results demonstrate the potential of both the Multilayer approach and Locally Linear method even when the information content in observations is low (Scenarios 3 and 4) and when model parameters vary at different rates with different structures (Scenario 2 and 3 respectively). Across Scenarios 1–4, the Multilayer approach and the Locally Linear Method provide the highest NSE and PTW scores (refer Figure 3). The Locally Linear Method, which is less restrictive in terms of its assumptions, provides parameter estimates with comparable quality to the Multilayer approach for all scenarios in this case study. For the scenario with the most complex combination of time varying parameters (Scenario 2), the long term temporal structure is mostly well represented by the Locally Linear method (see Figure 4), with the Multilayer approach demonstrating the most superior performance. Both the Multilayer and Locally Linear methods were also found to correctly identify stationary parameters when applied to scenarios with a combination of time varying and time invariant parameters.

The existing parameter evolution models (KSLS, SKS and HKS) have varying degrees of success in the routing parameter estimation case study. The KSLS parameter evolution model performs poorly for all scenarios, and is unsuitable for estimating parameters which vary over lengthy time scales. Representative results from the KSLS for Scenario 1 (ie. parameter variations have a strong impact on streamflow) demonstrate that the shrinkage of kernel locations encourages filter divergence (Figure 5) That is, the ensemble variance is continuously reduced over time, such that the filter becomes insensitive to observations. Background parameters are subsequently seen as having high confidence and are no longer updated, leading to convergence to a stationary distribution. This result is to some extent expected, given that the KSLS has been designed specifically for the detection of time invariant parameters [*Liu and West*, 2001].

The predecessor to the KSLS, the SKS, is more suited to hydrologic parameter estimation than the KSLS where parameter changes occur over long time scales (see Figure 4). Unlike the KSLS, the SKS is less susceptible to filter divergence because of its in-built variance inflation. The background parameter ensemble $\{\boldsymbol{\theta}_{t+1}^{i-}\}_{i=1:n}$ from the SKS has larger variance than the updated ensemble from the previous time step $\{\boldsymbol{\theta}_{t}^{i+}\}_{i=1:n}$, unlike in the KSLS where the shrinkage of locations retains the variance of $\{\boldsymbol{\theta}_{t}^{i+}\}_{i=1:n}$ [*Liu and*

Table 2. Sampling Distributions of Initial Parameters and Their Feasible Ranges[a]

				Feasible Range	
Parameter	Description	Units	Initial Sampling Distribution	Lower Bound	Upper Bound
b	Pareto-distributed soil storage shape parameter		$N(b(t=0), 0.001)$	0.10	1.50
c_{max}	Maximum point soil storage depth	(mm)	$N(c_{max}(t=0), 10)$	150	350
k_s	Surface Runoff Routing Coefficient		$N(k_s(t=0), 0.008)$	0.00	0.69
k_b	Groundwater Routing Coefficient		$N(k_b(t=0), 0.001)$	0.70	0.99
α	Excess Runoff Splitting Parameter		$N(\alpha(t=0), 0.001)$	0.50	0.99

[a]For all parameters, the mean of the initial sampling distribution is equal to the true initial parameter.

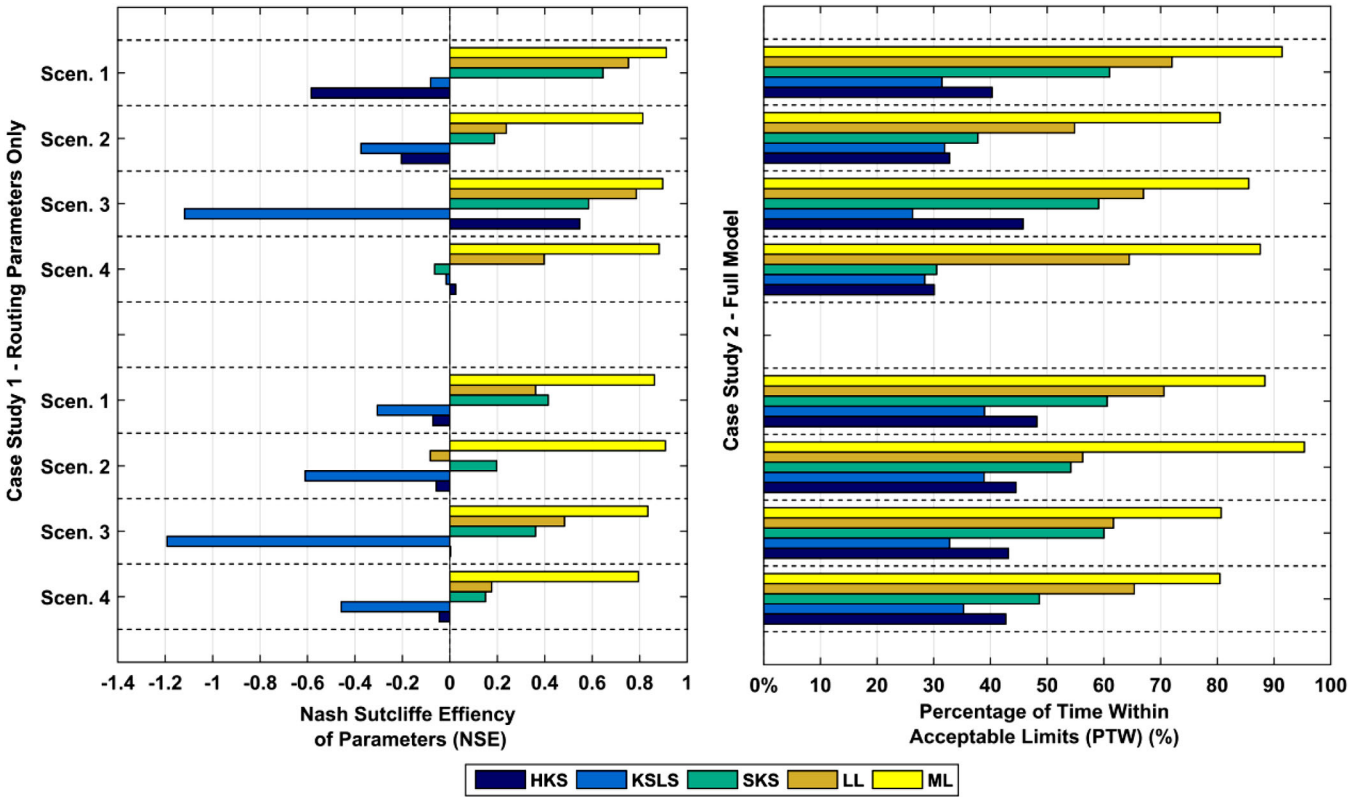

Figure 3. Summary performance metrics from using a suite of parameter evolution models in an EnKF framework for estimating time varying parameters. A range of time varying parameters were examined (Scenarios 1–4, refer Table 1) in two separate Case Studies 1) Updating routing parameters only and 2) Updating all model parameters. The NSE of the parameter ensemble mean (averaged over all parameters) is shown in the left panel. The right panel shows the percentage of time the ensemble mean is within acceptable limits (averaged over all parameters). Acceptable limits for each parameter are defined for time $\boldsymbol{t}$ as $\theta_t^* + \sigma \boldsymbol{d_p}$ where θ_t^* is the true parameter value at time $\boldsymbol{t}$, $\boldsymbol{d_p}$ is the feasible parameter range (Table 2) and σ is a fraction varying between 5% and 10%. HKS = Homoscedastic Kernel Smoother, KSLS = Kernel Smoother with Location Shrinkage, SKS= Standard Kernel Smoother, LL = Locally Linear, ML = Multilayer.

West, 2001]. Whilst the SKS performs reasonably well for Scenario 1 where parameter variations strongly impact streamflow response (on average, parameters are within their acceptable limits approximately 60% of the time), its estimation of α and k_b in particular for Scenario 2 and 4 is poor (parameters within acceptable limits for only 30–40% of the time). Scenario 2 in particular shows that even if the parameter variation signal in streamflow is fairly strong, the SKS struggles when multiple parameters vary at different rates (refer Figure 4).

Similar to the KSLS, parameter trajectories from the HKS fail to replicate the true parameter variations in all scenarios. Figure 5 shows the ensemble mean trajectories for the 3 year simulation period with $v = 0.2$ (as adopted in *Smith et al.* [2008]) for a representative case, Scenario 1. Rapid temporal fluctuations in the ensemble mean (as much as 0.2 over a time step) as well as the width of the ensemble are indicative of overdispersion due to the perturbation process. Smaller values of v led to reduced short-term parameter fluctuations, but were still unable to detect long-term patterns (refer Figure 5b shown for $v = 0.005$). Although the HKS is fairly similar to the SKS in its structure, it would seem that the use of a time invariant variance for parameter resampling leads to either overdispersion or underdispersion at certain times, degrading the overall performance of the filter.

4.2. Estimation of All Model Parameters

We now consider the case where all the parameters of the PDM are estimated simultaneously by the filter, that is, the Pareto distribution shape parameter b and maximum point soil storage depth c_{max} are also estimated. As was the case with the first case study, the Multilayer approach produces high quality estimates for all parameters, across all scenarios, with average NSE and PTW values in excess of 0.85 and 85% respectively. The linear trend in the soil storage parameters b and c_{max} was captured well by the SKS, Multilayer

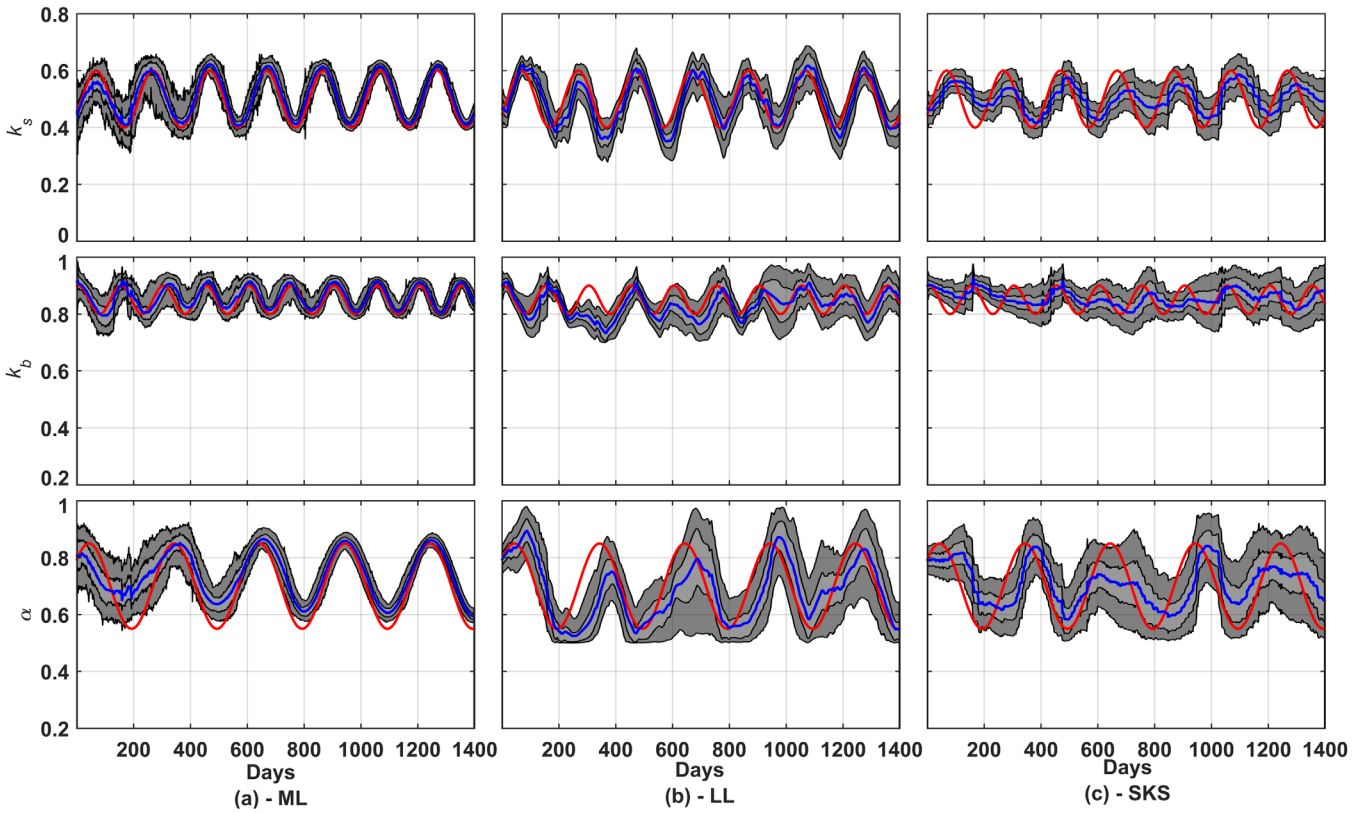

Figure 4. Updated parameter ensemble for Case Study 1 (Routing parameter updating only), Scenario 2. The darker grey areas indicate the middle 90% of the ensemble, whilst the lighter grey areas indicate the middle 50% of the ensemble. The blue line indicates the ensemble mean whilst the red line indicates the synthetic true parameter. Results are provided for (a) the Multilayer Method (ML), (b) the Locally Linear Method (LL), and (c) the Standard Kernel Smoother (SKS).

and Locally Linear methods in all scenarios (refer Figure 6). For both the HKS and the KSLS, inclusion of the soil storage parameters produced similar poor results to the case where only the routing parameters were estimated. However, for the remaining parameter evolution models, estimation of the routing parameters (α, k_s and k_b) was degraded in comparison to the case where b and c_{max} were assumed known. Inclusion of the soil storage parameters only slightly degraded the quality of routing parameter estimates (in terms of capturing the overall trend) for Scenario 3, where k_b and α both vary linearly in time (see Figure 6). However in terms of parameter bias, the NSE reduced significantly for the routing parameters (in particular, for the Locally Linear method, see Table 3), and the sinusoidal temporal structure in Scenarios 1, 2 and 4 was less pronounced in the ensemble mean (see e.g., Figure 7). This reduction in estimation quality is attributed to the sensitivity of these parameters to errors in the excess runoff, particularly α and k_b. Biases in the soil storage state accumulate over time and are not fully compensated for in the state updating step due to small errors in the soil storage parameters. This leads to errors in the excess runoff which degrade estimates of the less observable routing parameters, k_b and α. This is particularly problematic when k_b and α have complex time varying structures (e.g., high frequency sinusoidal), as the impacts of parameter fluctuations on the streamflow may be drowned out by errors in the excess runoff. Despite the reduced parameter estimation quality, the Locally Linear method still leads to superior estimation of streamflow compared to the Standard Kernel Smoother across the scenarios.

The full model parameter case study demonstrates the efficacy of the Multilayer approach, particularly in situations where several parameters are time varying and there is reduced information content in the observations. However, it is limited to situations where the form of nonstationarity can be defined apriori and the hyperparameters which govern the nonstationarity are time invariant. The Locally Linear approach is a suitable alternative which has modest apriori knowledge requirements (initial parameter value and initial rate of change). It has been shown to provide parameter estimates of comparable quality to the Multilayer approach in low dimensional systems, even when the impact of parameter variation on the observations is

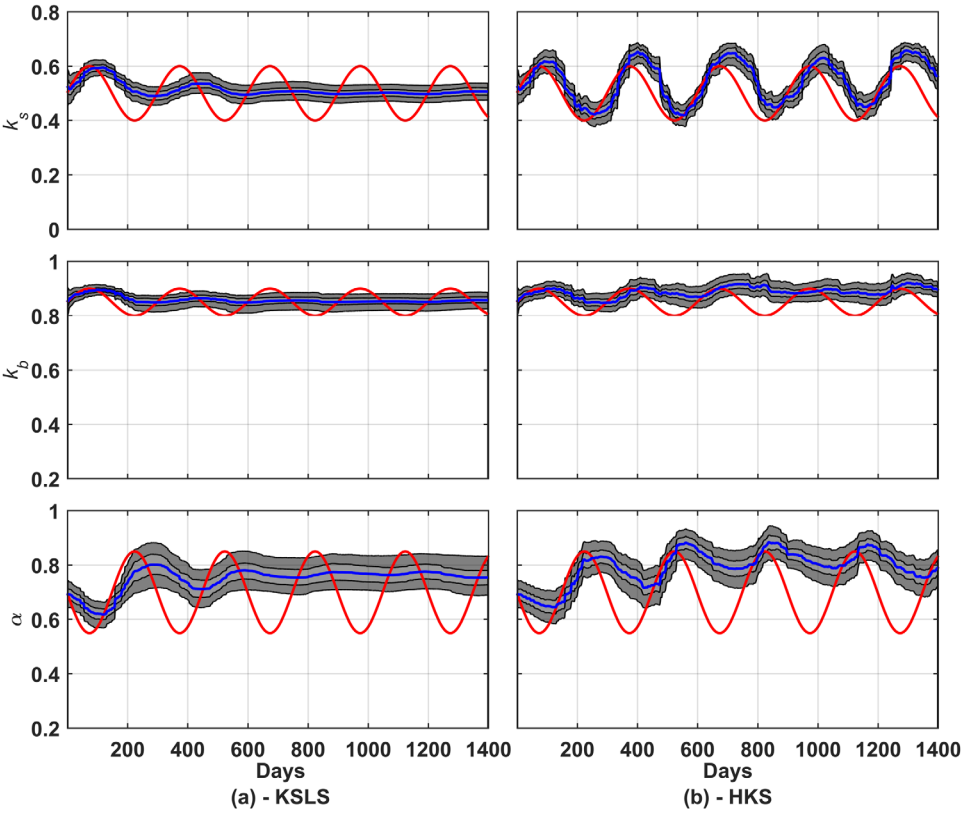

Figure 5. Updated parameter ensemble for Case Study 1 (Routing parameter updating only), Scenario 1. The darker grey areas indicate the middle 90% of the ensemble, whilst the lighter grey areas indicate the middle 50% of the ensemble. The blue line indicates the ensemble mean whilst the red line indicates the synthetic true parameter. Results are shown for (a) the Kernel Smoother with Location Shrinkage and (b) the Homoscedastic Kernel Smoother.

weak. However, the Local Linear approach is less resistant to periods of poor observability, and was shown to perform poorly when the dimensionality and nonlinearity was increased. It remains to be seen whether similar dimensionality issues arise when the Local Linear method is applied to other model structures. Issues of observability and identifiability in parameter estimation using DA will be investigated in more detail in a follow up paper.

Table 3. Percentage Change in NSE of the Routing Parameters When All Model Parameters are Estimated, Compared to When Only the Routing Parameters are Estimated[a]

		Percentage change in NSE $\left(\frac{NSE_{full\ model} - NSE_{routing\ only}}{\lvert NSE_{routing\ only}\rvert}\right)$		
Scenario	Parameter	Locally Linear Method	Standard Kernel Smoother	Multilayer Method
Scenario 1	k_s	−28%	−43%	0%
	k_b	−252%	−42%	−15%
	α	−42%	−86%	−4%
Scenario 2	k_s	−70%	−81%	2%
	k_b	−485%	−552%	12%
	α	−134%	−58%	21%
Scenario 3	k_s	−19%	−85%	−4%
	k_b	−115%	−81%	−15%
	α	−42%	−40%	−15%
Scenario 4	k_s	−17%	−62%	−9%
	k_b	−938%	−5%	−29%
	α	−114%	5%	−18%
Average over all parameters and scenarios:		−188%	−94%	−6%

[a]Negative values indicate a reduction in the NSE score when all model parameters are estimated simultaneously. The quality of the parameter estimates is almost always degraded when the dimensionality is increased.

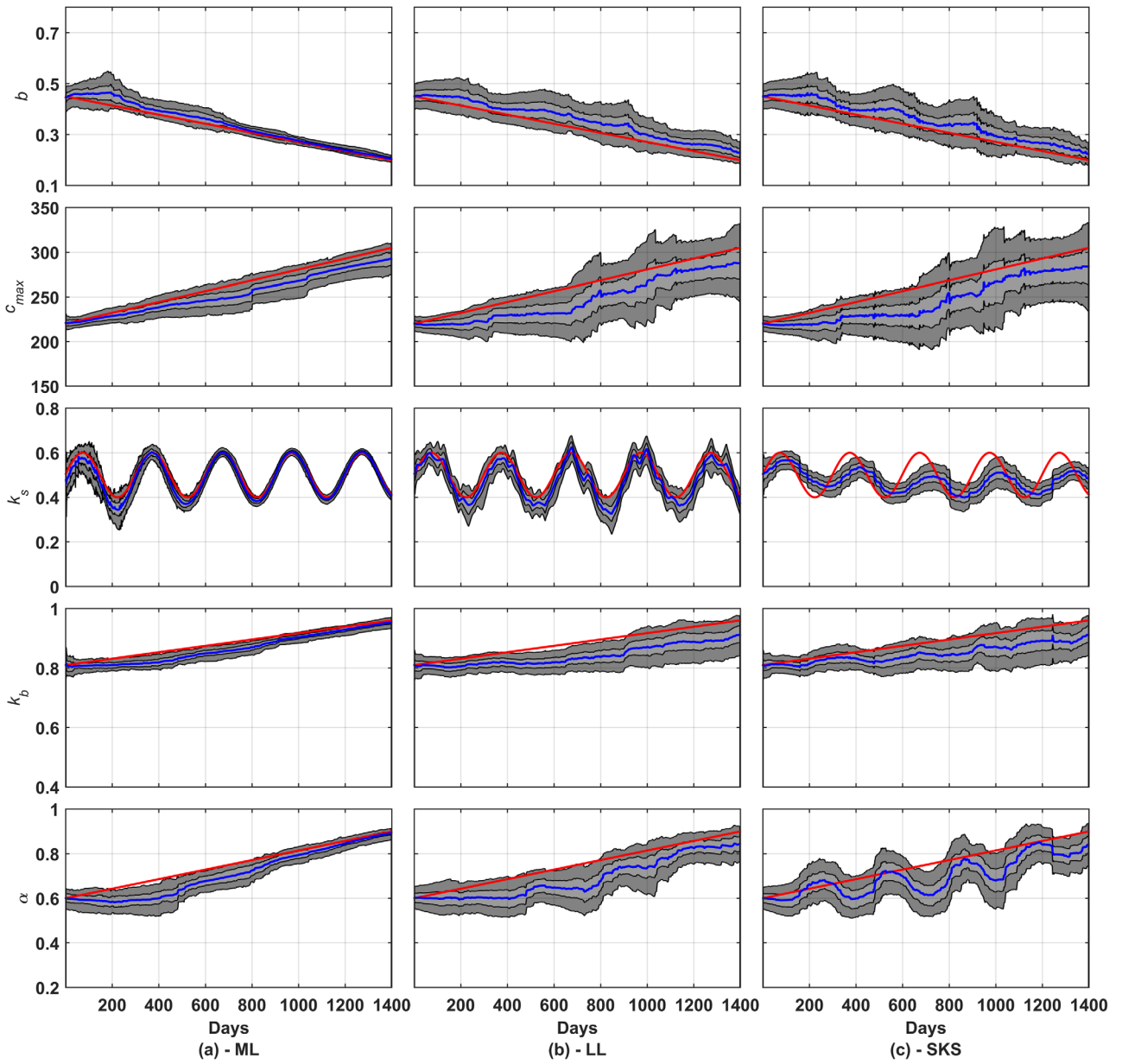

Figure 6. Updated parameter ensemble for Case Study 2 (Full model parameter updating), Scenario 3. The darker grey areas indicate the middle 90% of the ensemble, whilst the lighter grey areas indicate the middle 50% of the ensemble. The blue line indicates the ensemble mean whilst the red line indicates the synthetic true parameter. Results are provided for (a) the Multilayer Method (ML), (b) the Locally Linear Method (LL), and (c) the Standard Kernel Smoother (SKS).

4.3. Robustness of the Multilayer Method

Additional testing was undertaken to determine the robustness of the Multilayer method to errors in the assumed parameter dynamics. Any given parameter θ was assumed to follow a more general structure:

$$\theta(\mathrm{t}) = a\sin\left(\frac{2\pi t}{b} + c\right) + d + et \tag{32}$$

where a, b, c, d and e are the 5 hyperparameters updated in the first stage of the filter for each model parameter and t = time. The algorithm was found to be fairly robust to changes in assumptions about the temporal structure of parameters, with only one parameter experiencing a reduction in performance due to poor identifiability. For the first case study, results were largely similar to the case where there are no ambiguities in the dynamical structure of parameters (refer supporting information Figure S2 for

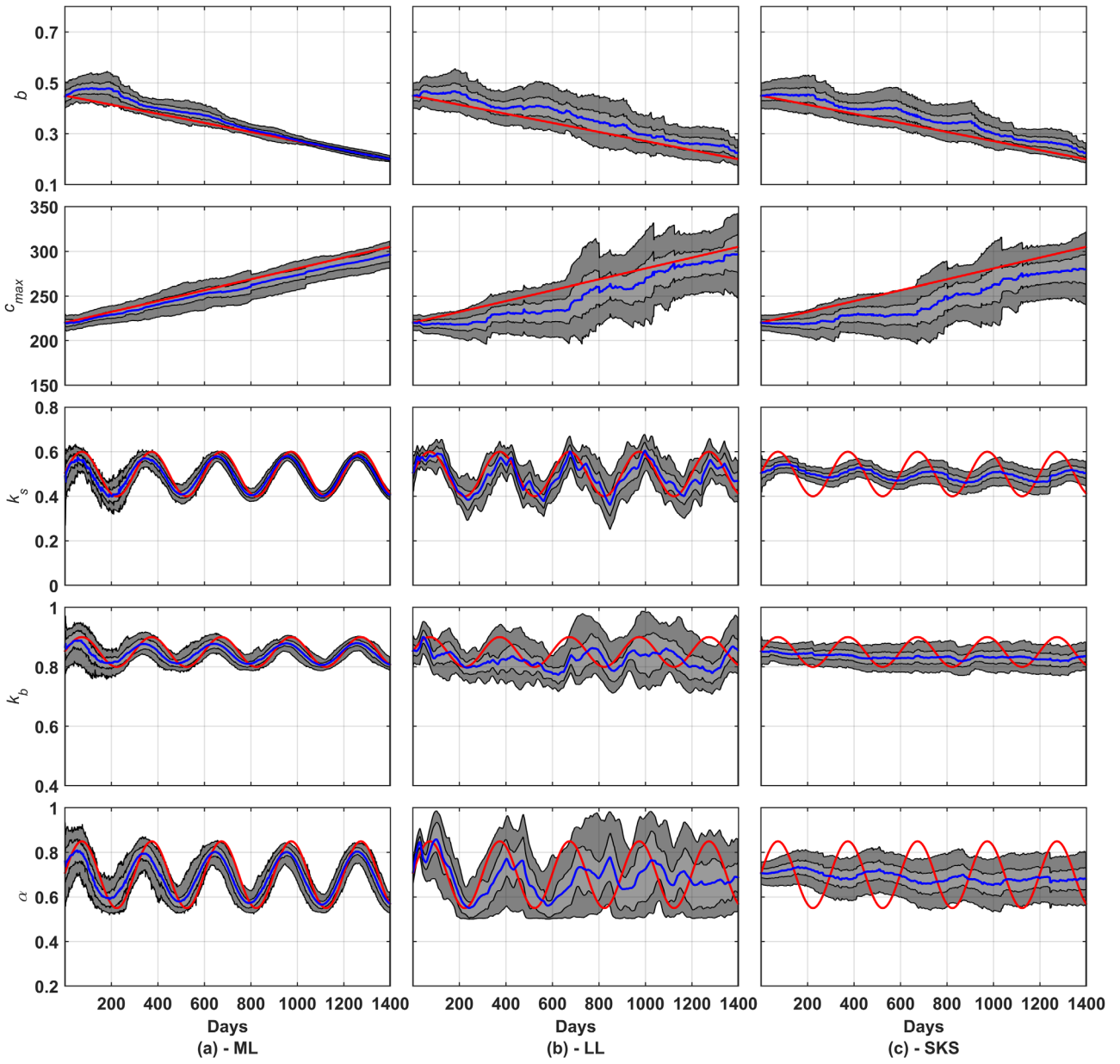

Figure 7. Updated parameter ensemble for Case Study 2 (Full model parameter updating), Scenario 4. The darker grey areas indicate the middle 90% of the ensemble, whilst the lighter grey areas indicate the middle 50% of the ensemble. The blue line indicates the ensemble mean whilst the red line indicates the synthetic true parameter. Results are provided for (a) the Multilayer Method (ML), (b) the Locally Linear Method (LL), and (c) the Standard Kernel Smoother (SKS).

representative results). Only a minor degradation in performance was seen for Case Study 2. All parameters except for c_{max} were well represented, with performance similar to that of the perfect parameter structure case (refer supporting information Figure S3). The reduction in c_{max} performance is attributed to the weak correlation between its hyperparameters and the observations. This means that introducing further unknowns in temporal structure reduces the ability of the filter to detect changes in this parameter. Ultimately, the method was found to be fairly robust to incorrect assumptions in parameter temporal structure, but its performance in this case may reduce as dimensionality and nonlinearity increase and parameter identifiability reduces.

4.4. Improving the Locally Linear Approach—Local Linear Regression?

The Locally Linear approach demonstrated strong potential for the estimation of time varying parameters in the first case study, with minimal apriori knowledge required. The method generates background

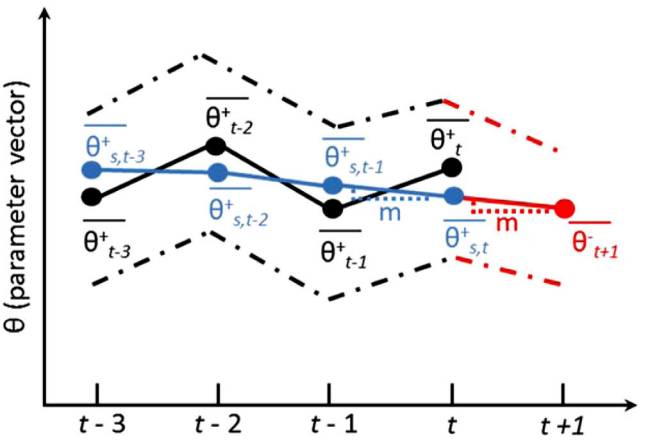

Figure 8. Schematic of the Locally Linear Approach combined with local linear regression for proposing background parameters. Initially, local linear regression is undertaken on the updated parameter ensemble means from previous time steps (refer to the blue line). The mean of the background parameters for the current time $\{\theta^{-}_{t+1}\}$ is then estimated by linear extrapolation from the regressed values.

parameters at any given time by using the rate of change of the updated parameters from the previous two time steps. A natural question might be whether the Locally Linear approach could be improved by considering a longer historical time window of updated parameters than simply the previous two time steps to propose the background. This was examined by applying Local Linear Regression (hereafter LLR) to the full time history of updated parameters, thereby producing a temporally smoothed time series. The commonly used Epanechnikov kernel [*Epanechnikov*, 1969] was adopted as the weighting function and a number of bandwidth sizes examined. Background parameters were then proposed in a similar fashion to the Locally Linear approach, except that the smoothed parameters from the previous two time steps were used for extrapolation (refer Figure 8).

Incorporating LLR into the Locally Linear approach produced temporally smoother parameter trajectories than when it was not used. However, its ability to replicate the temporal structure of parameters was significantly reduced compared to the nonsmoothed Local Linear approach.

Figure 9 shows the parameter ensemble when combining the Locally Linear approach with LLR for Case Study 1 - Scenario 1. Bandwidth values of 10, 20 and 5 days were adopted for all parameters based on trial runs. It can be seen that the filter produces a delayed response near the crests and troughs where the parameters experience the most rapid change in gradient, (see for instance near days 800–1000 for k_b) .

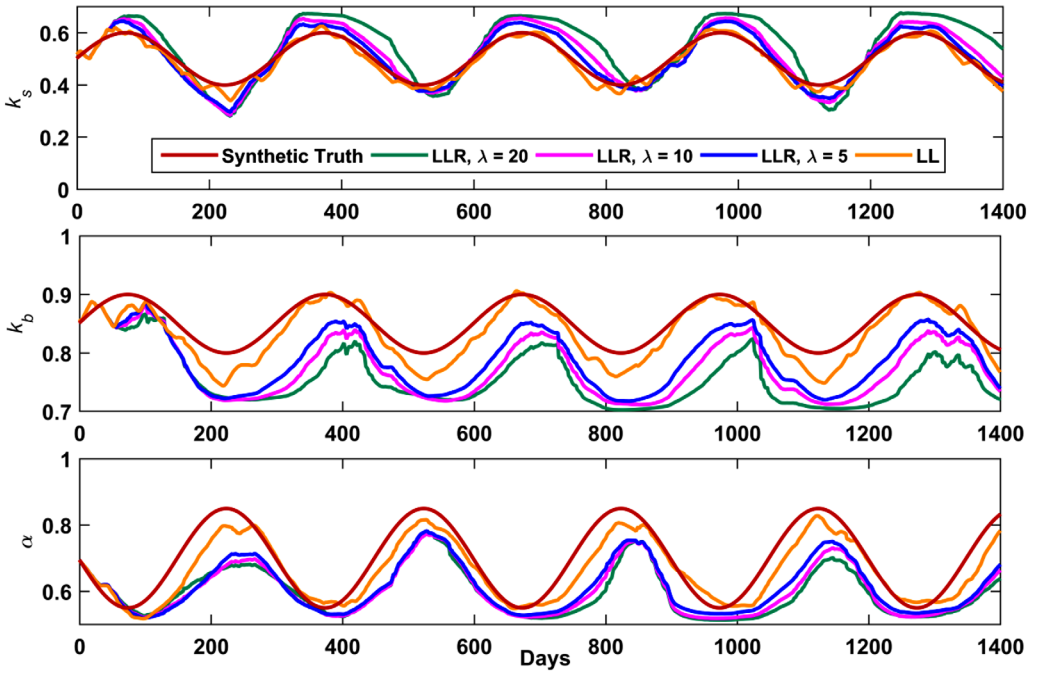

Figure 9. Parameter ensemble mean trajectories for the Locally Linear Method with Local Linear Regression (LLR) and the Locally Linear Method (LL) for Case Study 1 (Routing parameter updating only), Scenario 1.

This behavior occurs due to the LLR smoothing out regions of parameter updates of duration less than the bandwidth, which may indicate a critical change in the parameter. This is demonstrated here where even a relatively small bandwidth of 5 days is used.

5. Conclusions

The accuracy of hydrologic predictions from conceptual rainfall runoff models is strongly dependent on the calibration parameters used to derive them. Consideration of a unique parameter set or stationary distribution of parameters is problematic in that 1) it ignores potential changes in the catchment that may have occurred outside the calibration period (e.g., changing land use); 2) they have a tendency to be highly dependent on the dominant climatic regime of the calibration period [*Sorooshian et al.*, 1983; *Choi and Beven*, 2007; *Wu and Johnston*, 2007]; and 3) it is difficult to adequately simulate all features of hydrologic variables (such as peak flows, low flows, volume of runoff) due to model structural deficiencies [*Moussa and Chahinian*, 2009; *Efstratiadis and Koutsoyiannis*, 2010; *Westerberg et al.*, 2011]. A framework which relaxes the need for a time invariant parameter set has the potential to improve the transferability of hydrologic models in time, particularly for nonstationary catchments. This paper examined the potential for Data Assimilation (DA) to detect time variations in model parameters from hydrologic observations, namely streamflow. It serves as an important first step in examining whether DA can be used to detect a structured time variation in multiple model parameters simultaneously. Joint state-parameter estimation with the Ensemble Kalman Filter was undertaken for a range of synthetic experiments using the conceptual rainfall runoff model, the Probability Distributed Model (PDM). It was demonstrated that the success of the DA algorithm is strongly dependent on the mechanism for generating the background (or prior) parameter ensembles, an issue which has not been investigated in previous studies [*Smith et al.*, 2008; *Vrugt et al.*, 2013]. Existing techniques such as the Kernel Smoother with Location Shrinkage [*Liu and West*, 2001] and the Standard Kernel Smoother [*West*, 1993] are shown to be poorly suited for time varying parameter estimation from streamflow observations, particularly when multiple parameters vary in a complex, uncorrelated fashion. Two alternative parameter evolution models are proposed which are specifically suited to time varying parameter applications. The first is a Multilayer approach which uses the EnKF to estimate hyperparameters of the temporal structure, based on apriori knowledge of the form of nonstationarity. The second is a Locally Linear approach that uses local linear estimation to propose temporal changes and requires no assumptions of the form of parameter nonstationarity. The results demonstrate that both methods more accurately capture temporal variations in parameters compared to the traditional parameter evolution models. This indicates that incorporating information about temporal structure into the background, in addition to ensuring sufficient variance, can improve the detection of time varying parameters. Expert knowledge of how catchment nonstationary affects model parameters (e.g., whether parameters will vary linearly or cyclically with time to reflect changes in catchment conditions) will significantly improve filter performance when changing processes affect several model parameters. This is demonstrated by the superior performance of the Multilayer method compared to other approaches examined. The method was shown to be fairly robust to inaccurate assumptions of parameter dynamics, meaning that reasonable performance can still be obtained by considering all the likely modes of parameter variability together. This however has the potential to degrade the detection of less identifiable parameters. The Locally Linear method requires only knowledge of the initial parameters and their likely rate of change. It was shown to provide parameter estimates of comparable quality to the Multilayer approach in low dimensional systems, indicating that it can be a suitable alternative where changes in catchment conditions lead to time variations in only a few model parameters. Small increases in dimensionality and nonlinearity have the potential to degrade the performance of the Locally Linear method, indicating a limit to which several time varying parameters can be estimated from streamflow observations alone, without the addition of further external information (as in the case of the Multilayer method). The optimum number of parameters that could be satisfactorily estimated by the Locally Linear method would be model dependent, and also depend on the signal strength in the observed variables. An application of the proposed methods to a real data case study will be investigated in a follow up publication. This will allow assessment of the extent to which time varying parameters can compensate for missing processes and the resulting improvement in streamflow prediction. Future work will also examine the performance of the proposed parameter evolution models with other DA algorithms such as the Ensemble Kalman Smoother (EnKS) and Particle Filter, along with a range of model structures.

Appendix A

Here we demonstrate that the parameter update equation in the state augmentation approach is equivalent to the dual update procedure when applied with Kalman filtering methods. For the purposes of demonstration, we will assume our system is one dimensional (the same logic applies for higher dimesions):

$$\boldsymbol{x}_{t+1} = x_{t+1};\ \boldsymbol{y}_{t+1} = y_{t+1};\ \boldsymbol{\theta}_{t+1} = \theta_{t+1}$$

We start with the state augmentation approach for parameter updating. The augmented state vector is defined as:

$$\boldsymbol{z}_{t+1} = \begin{bmatrix} x_{t+1} \\ \theta_{t+1} \end{bmatrix} \tag{A1}$$

The augmented state is updated using the Kalman update equation:

$$\boldsymbol{z}_{t+1}^{i+} = \boldsymbol{z}_{t+1}^{i-} + \mathbf{K}_{t+1}\left[y_{t+1}^{i} - \hat{y}_{t+1}^{i}\right] \quad for\ i = 1:n \tag{A2}$$

$$\mathbf{K}_{t+1} = \boldsymbol{\Sigma}_{t+1}^{z\hat{y}}\left[\Sigma_{t+1}^{\hat{y}\hat{y}} + \Sigma_{t+1}^{y^o y^o}\right]^{-1} \tag{A3}$$

That is,

$$\begin{bmatrix} x_{t+1}^{i+} \\ \theta_{t+1}^{i+} \end{bmatrix} = \begin{bmatrix} x_{t+1}^{i-} \\ \theta_{t+1}^{i-} \end{bmatrix} + \frac{1}{\left(\Sigma_{t+1}^{\hat{y}\hat{y}} + \Sigma_{t+1}^{y^o y^o}\right)} \begin{bmatrix} cov\left(\left\{x_{t+1}^{i-}\right\}_{i=1:n}, \left\{\hat{y}_{t+1}^{i-}\right\}_{i=1:n}\right) \\ cov\left(\left\{\theta_{t+1}^{i-}\right\}_{i=1:n}, \left\{\hat{y}_{t+1}^{i-}\right\}_{i=1:n}\right) \end{bmatrix} \left[y_{t+1}^{i} - \hat{y}_{t+1}^{i}\right] \quad for\ i = 1:n \tag{A4}$$

Since

$$\boldsymbol{\Sigma}_{t+1}^{z\hat{y}} = cov\left(\boldsymbol{z}_{t+1}^{-},\ \hat{y}_{t+1}\right) = cov\left(\left\{\begin{bmatrix} x_{t+1}^{i-} \\ \theta_{t+1}^{i-} \end{bmatrix}\right\}_{i=1:n},\ \left\{\hat{y}_{t+1}^{i-}\right\}_{i=1:n}\right) = \begin{bmatrix} cov\left(\left\{x_{t+1}^{i-}\right\}_{i=1:n}, \left\{\hat{y}_{t+1}^{i-}\right\}_{i=1:n}\right) \\ cov\left(\left\{\theta_{t+1}^{i-}\right\}_{i=1:n}, \left\{\hat{y}_{t+1}^{i-}\right\}_{i=1:n}\right) \end{bmatrix} \tag{A5}$$

Now extract the parameter update equation (i.e., second element of vector):

$$\theta_{t+1}^{i+} = \theta_{t+1}^{i-} + \frac{1}{\left(\Sigma_{t+1}^{\hat{y}\hat{y}} + \Sigma_{t+1}^{y^o y^o}\right)} cov\left(\left\{\theta_{t+1}^{i-}\right\}_{i=1:n}, \left\{\hat{y}_{t+1}^{i-}\right\}_{i=1:n}\right)\left[y_{t+1}^{i} - \hat{y}_{t+1}^{i}\right] \quad for\ i = 1:n \tag{A6}$$

$$= \theta_{t+1}^{i-} + \Sigma_{t+1}^{\theta\hat{y}}\left[\Sigma_{t+1}^{\hat{y}\hat{y}} + \Sigma_{t+1}^{y^o y^o}\right]^{-1} y_{t+1}^{i} - \hat{y}_{t+1}^{i}\right] \quad for\ i = 1:n \tag{A7}$$

This is the equivalent to the parameter update equation presented in the Dual approach (refer equations (6) and (7)).

Appendix B

Here we provide further details on the state updating step in the dual state-parameter estimation framework (i.e., equations (10–12)). As noted earlier, the observations have been utilized to generate the background model states (by way of the updated model parameters). There is therefore potentially some correlation between the model simulation noise (also referred to as process noise) and the observation (or measurement) noise. In such cases, the Kalman update equation remains unchanged, but the Kalman gain matrix now includes terms accounting for the covariance between process and measurement noise [e.g., *Simon*, 2006]:

$$\mathbf{K} = \left[\mathbf{P}_t\mathbf{H}^{\mathsf{T}} + \mathbf{C}_t\right]\left[\mathbf{H}\mathbf{P}_t\mathbf{H}^{\mathsf{T}} + \mathbf{H}\mathbf{C} + \mathbf{C}_t^{\mathsf{T}}\mathbf{H}^{\mathsf{T}} + \boldsymbol{R}_t\right]^{-1} \tag{B1}$$

where $\boldsymbol{x}_{t+1} = \mathbf{F}\boldsymbol{x}_t + \mathbf{G}\boldsymbol{u}_{t+1} + \boldsymbol{\omega}_t$, $\boldsymbol{y}_t = \mathbf{H}\boldsymbol{x}_t + \upsilon_t$, $cov(\omega_t,\ \omega_t) = \mathbf{P}_t =$ background or process error covariance, $cov(\upsilon_t, \upsilon_t) = \mathbf{R}_t =$ observation error covariance, $cov(\omega_t, \upsilon_t) = \mathbf{C}_t$ = covariance between background and observation errors, $\mathbf{H}$ = linearized observation operator, $\mathbf{F}$ and $\mathbf{G}$ represent linearized model dynamics.

In an EnKF framework, error covariances are estimated by the sample covariances of the monte carlo ensembles (assuming unbiasedness). For instance, in standard state updating using the EnKF, the following estimate is used:

$$\begin{aligned}\mathbf{P}_t\mathbf{H}_t^{\mathbf{T}} &= cov(\omega_t, \omega_t)\mathbf{H}_t^{\mathbf{T}} = cov(\omega_t, \mathbf{H}\omega_t)\\ &= \mathbf{\Sigma}_t^{xy}\end{aligned} \tag{B2}$$

= covariance between model simulated states and simulated observed variables.

The same approach can therefore be used to estimate $\mathbf{C}_t$, with $\mathbf{C}_t$ given by the cross covariance between the simulated model states using the updated parameters and the observations. This is problematic for a number of reasons. Firstly, consider the two ensembles of model states available prior to state updating:

$$\hat{\boldsymbol{x}}_{t+1}^{i} = f\left(\boldsymbol{x}_t^{i+},\ \boldsymbol{\theta}_{t+1}^{i-},\ \boldsymbol{u}_{t+1}\right) \quad for\ i=1:n \tag{B3}$$

$$\boldsymbol{x}_{t+1}^{i-} = f\left(\boldsymbol{x}_t^{i+},\ \boldsymbol{\theta}_{t+1}^{i+},\ \boldsymbol{u}_{t+1}\right) \quad for\ i=1:n \tag{B4}$$

where $\left\{\boldsymbol{x}_{t+1}^{i-}\right\}_{i=1:n}$ represents the model simulated states using the prior parameters and $\left\{\hat{\boldsymbol{x}}_{t+1}^{i}\right\}_{i=1:n}$ represents the model simulated states using the updated parameters. To simplify the notation, we define the following:

$$\begin{aligned}X &= \left\{\boldsymbol{x}_{t+1}^{i-}\right\}_{i=1:n}\\ \hat{X} &= \left\{\hat{\boldsymbol{x}}_{t+1}^{i}\right\}_{i=1:n}\\ Y &= \left\{\boldsymbol{y}_{t+1}^{i}\right\}_{i=1:n}\end{aligned}$$

X is potentially correlated with the Y (the observation errors at time $t+1$), whilst $\hat{X}$ is not, as the observations have not been used in any capacity to generate these states. Now X and $\hat{X}$ are highly correlated with each other. Due to sampling issues, it is possible that $corr\left(\hat{X},\ Y\right)$ is nonzero, and since X and $\hat{X}$ are highly correlated, this nonzero correlation will inflate the estimate of $\mathbf{C}_{t+1}$. A more appropriate estimate for $\mathbf{C}_{t+1}$ is hence given by $\mathbf{\Sigma}_{t+1}^{\varepsilon_x y^o}$:

$$\mathbf{C}_{t+1} \equiv \mathbf{\Sigma}_{t+1}^{\varepsilon_x y^o} = cov\left(X - \hat{X},\ Y\right) = cov\left(\left\{\varepsilon_{x}{}_{t+1}^{i}\right\}_{i=1:n},\ \left\{\boldsymbol{y}_{t+1}^{i}\right\}_{i=1:n}\right) \tag{B5}$$

where $\varepsilon_{x}{}_{t+1}^{i} = \boldsymbol{x}_{t+1}^{i-} - \hat{\boldsymbol{x}}_{t+1}^{i}$

To see why this is the case, note that equation (B5) can be written as:

$$\mathbf{\Sigma}_{t+1}^{\varepsilon_x y^o} = cov\left(X - \hat{X},\ Y\right) = cov(X, Y) - cov\left(\hat{X}, Y\right) \tag{B6}$$

Equation (B6) shows that any spurious correlations between $\hat{X}$ and Y are removed from $cov(X, Y)$, and that $\mathbf{\Sigma}_{t+1}^{\varepsilon_x y^o}$ is equivalent to $cov(X, Y)$ if $corr\left(\hat{X},\ Y\right)$ is indeed zero (ie. $\mathbf{C}_{t+1} = cov(X, Y)$ if $corr\left(\hat{X},\ Y\right) = 0$). It was found from experiments in this study that such spurious correlations have the potential to overestimate $\mathbf{C}_{t+1}$, leading to degraded parameter and state estimates.

Now the term $\mathbf{HC}_{t+1}$ is simply the covariance between the process noise in observation space and the observations. Using the estimate of $\mathbf{C}_{t+1}$ from equation (B5), we have:

$$\begin{aligned}\mathbf{\Sigma}_{t+1}^{\varepsilon_y y^o} \equiv \mathbf{HC}_{t+1} &= \mathbf{H}cov\left(X - \hat{X},\ Y\right) = cov\left(\mathbf{H}X - \mathbf{H}\hat{X},\ Y\right)\\ &= cov\left(\left\{\varepsilon_{y}{}_{t+1}^{i}\right\}_{i=1:n},\ \left\{\boldsymbol{y}_{t+1}^{i}\right\}_{i=1:n}\right)\end{aligned} \tag{B7}$$

where $\varepsilon_{y}{}_{t+1}^{i} = \tilde{\boldsymbol{y}}_{t+1}^{i} - \hat{\boldsymbol{y}}_{t+1}^{i}$.

Acknowledgments
This study was funded by the Australian Research Council as part of the Discovery Project DP140102394. Marshall is additionally supported through a Future Fellowship FT120100269. Data utilized in this study can be made available from the authors upon request.

References

Aksoy, A., F. Zhang, and J. Nielsen-Gammon (2006), Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model, *Mon. Weather Rev.*, *134*, 2951–2970.

Anderson, J. L. (2007), An adaptive covariance inflation error correction algorithm for ensemble filters, *Tellus, Ser. A*, *59*(2), 210–224, doi: 10.1111/j.1600-0870.2006.00216.x.

Annan, J., J. Hargreaves, N. Edwards, and R. Marsh(2005), Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter, *Ocean Modell.*, *8*(1-2), 135–154, doi:10.1016/j.ocemod.2003.12.004.

Aubert, D., C. Loumagne, and L. Oudin(2003), Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall–runoff model, *J. Hydrol.*, *280*(1-4), 145–161, doi:10.1016/S0022-1694(03)00229-4.

Bardossy, A. (2007), Calibration of hydrological model parameters for ungauged catchments, *Hydrol. Earth Syst. Sci.*, *11*, 703–710.
Beven, K., and J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *J. Hydrol.*, *249*, 11–29.
Burgers, G., P. J. van Leeuwen, and G. Evensen (1998), On the analysis scheme in the ensemble Kalman filter, *Mon. Weather Rev.*, *126*, 1719–1724.
Choi, H. T., and K. Beven(2007), Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework, *J. Hydrol.*, *332*, 316–336.
Coron, L., V. Andréassian, C. Perrin, J. Lerat, J. Vaze, M. Bourqui, and F. Hendrickx (2012), Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments. *Water Resour. Res.*, *48*, W05552, doi:10.1029/2011WR011721.
Croke, B. F.W., W. S. Merritt, and A. J. Jakeman (2004), A dynamic model for predicting hydrologic response to land cover changes in gauged and ungauged catchments, *J. Hydrol.*, *291*(1-2), 115–131, doi:10.1016/j.jhydrol.2003.12.012.
Dee, D. P. (2005), Bias and data assimilation, *Q. J. R. Meteorol. Soc.*, *131*(613), 3323–3343, doi:10.1256/qj.05.137.
de Vos, N. J., T. H. M. Rientjes and H. V. Gupta (2010), Diagnostic evaluation of conceptual rainfall-runoff models using temporal clustering, *Hydrol. Processes*, *24*, 2840–2850, doi:10.1002/hyp.7698.
Duan, Q., S. Sorooshian, and V. K. Gupta (1994), Optimal use of the SCE-UA global optimization method for calibrating watershed models, *J. Hydrol.*, *158*, 265–284.
Ebtehaj, M., H. Moradkhani, and H. V. Gupta (2010), Improving robustness of hydrologic parameter estimation by the use of moving block bootstrap resampling, *Water Resour. Res.*, *46*, W07515, doi:10.1029/2009WR007981.
Efstratiadis, A., and D. Koutsoyiannis(2010), One decade of multi-objective calibration approaches in hydrological modelling: A review, *Hydrol. Sci. J.*, *55*(1), 58–78, doi:10.1080/02626660903526292.
Epanechnikov, V. A. (1969), Non-parametric estimation of a multivariate probability density, *Theory Probab. Appl.*, *14*(1), 153–158, doi: 10.1137/1114019.
Evensen, G. (2006), *Data Assimilation: The Ensemble Kalman Filter*, Springer, N. Y.
Franssen, H. H. J., and W. Kinzelbach(2008), Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem, *Water Resour. Res.*, *44*, W09408, doi:10.1029/2007WR006505.
Gelb, A. (1974), *Applied Optimal Estimation*, MIT Press, Cambridge, U. K.
Gharari, S., M. Hrachowitz, F. Fenicia, and H. H. G. Savenije (2013), An approach to identify time consistent model parameters: Sub-period calibration, *Hydrol. Earth Syst. Sci.*, *17*(1), 149–161, doi:10.5194/hess-17-149-2013.
Gordon, N. J., D. J. Salmond, and A. F. M. Smith (1993), Novel approach to non-linear/non-Gaussian Bayesian state estimation, *IEE Proc., Part F*, *140*, 107–113.
Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, *34*(4), 751–763, doi:10.1029/97WR03495.
He, M., T. S. Hogue, K. J. Franz, S. A. Margulis, and J. A. Vrugt (2011), Characterizing parameter sensitivity and uncertainty for a snow model across hydroclimatic regimes, *Adv. Water Resour.*, *34*(1), 114–127, doi:10.1016/j.advwatres.2010.10.002.
Houser, P. R., W. J. Shuttleworth, J. S. Famiglietti, H. V. Gupta, K. H. Syed, and D. C. Goodrich (1998), Integration of soil moisture remote sensing and hydrological modeling using data assimilation, *Water Resour. Res.*, *34*(12), 3405–3420.
Houtekamer, P. L., and H. L. Mitchell (1998), Data assimilation using an ensemble Kalman filter technique, *Mon. Weather Rev.*, *126*, 796–811.
Jazwinski, A. H. (1970), *Stochastic Processes and Filtering Theory*, Elsevier, N. Y.
Jeffrey, S. J., J. O. Carter, K. B. Moodie, and A. R. Beswick (2001), Using spatial interpolation to construct a comprehensive archive of Australian climate data, *Environ. Modell. Software*, *16*(4), 309–330.
Jeremiah, E., L. Marshall, S. A. Sisson, and A. Sharma(2013), Specifying a hierarchical mixture of experts for hydrologic modeling: Gating function variable selection, *Water Resour. Res.*, *49*, 2926–2939, doi:10.1002/wrcr.20150.
Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, *J. Basic Eng.*, *82* (1960), 35–45.
Khalil, H. K. (1996), *Nonlinear Systems*, 143 pp., Simon & Schuster, Upper Saddle River, N. J.
Komma, J., G. Blöschl, and C. Reszler (2008), Soil moisture updating by Ensemble Kalman Filtering in real-time flood forecasting, *J. Hydrol.*, *357*(3-4), 228–242, doi:10.1016/j.jhydrol.2008.05.020.
Kuczera, G. (1987), Prediction of water yield reductions following a bushfire in ash-mixed species eucalyptus forest, *J. Hydrol.*, *94*, 215–236.
Leisenring, M., and H. Moradkhani (2012), Analysing the uncertainty of suspended sediment load prediction using sequential data assimilation, *J. Hydrol.*, *468–469*, 268–282.
Li, Y., D. Ryu, A. W. Western, and Q. J. Wang (2013), Assimilation of stream discharge for flood forecasting: The benefits of accounting for routing time lags, *Water Resour. Res.*, *49*, 1887–1900, doi:10.1002/wrcr.20169.
Lin, Z., and M. B. Beck (2007), On the identification of model structure in hydrological and environmental systems, *Water Resour. Res.*, *43*, W02402, doi:10.1029/2005WR004796.
Liu, J., and M. West (2001), Combined parameter and state estimation in simulation based filtering, in *Sequential Monte Carlo Methods in Practice*, edited by A. Doucet, N. de Freitas, and N. Gordon, pp. 197–223, Springer, N. Y.
Liu, Y., and H. V. Gupta (2007), Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, *Water Resour. Res.*, *43*,W07401, doi:10.1029/2006WR005756.
Liu, Y., A. H. Weerts, M. Clark, H.-J. Hendricks Franssen, S. Kumar, H. Moradkhani, and P. Restrepo (2012), Advancing data assimilation in operational hydrologic forecasting: Progresses, challenges, and emerging opportunities, *Hydrol. Earth Syst. Sci. Discuss.*, *16*, 3863–3887. doi:10.5194/hessd-9-3415-2012.
Lü, H., Z. Yu, Y. Zhu, S. Drake, Z. Hao, and E. A. Sudicky (2011), Dual state-parameter estimation of root zone soil moisture by optimal parameter estimation and extended Kalman filter data assimilation, *Adv. Water Resour.*, *34*(3), 395–406, doi:10.1016/j.advwatres.2010.12.005
Madsen, H. (2003), Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives, *Adv. Water Resour.*, *26*(2), 205–216, doi:10.1016/S0309-1708(02)00092-1.
Maneta, M. P., and R. Howitt(2014), Stochastic calibration and learning in nonstationary hydroeconomic models, *Water Resour. Res.*, *50*, 3976–3993, doi:10.1002/2013WR015196.
Marshall, L., A. Sharma, and D. J. Nott (2006), Modelling the catchment via mixtures: Issues of model specification and validation, *Water Resour. Res.*, *42*, W11409, doi:10.1029/2005WR004613.
Matgen, P., F. Fenicia, S. Heitz, D. Plaza, R. de Keyser, V. R. N. Pauwels, and H. Savenije (2012), Can ASCAT-derived soil wetness indices reduce predictive uncertainty in well-gauged areas? A comparison with in situ observed soil moisture in an assimilation application, *Adv. Water Resour.*, *44*, 49–65, doi:10.1016/j.advwatres.2012.03.022.

Merz, R., J. Parajka, and G. Blöschl (2011), Time stability of catchment model parameters: Implications for climate impact analyses, *Water Resour. Res.*, *47*, W02531, doi:10.1029/2010WR009505.
Mohammadpour, J., and C. W. Scherer (2012), *Control of Linear Parameter Varying Systems With Applications*, 3 pp., Springer, N. Y.
Moore, R. J. (2007), The PDM rainfall-runoff model, *Hydrol. Earth Syst. Sci.*, *11*(1), 483–499, doi:10.5194/hess-11-483-2007.
Moradkhani, H., K.-L. Hsu, H. Gupta, and S. Sorooshian(2005a), Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, *Water Resour. Res.*, *41*, W05012, doi:10.1029/2004WR003604.
Moradkhani, H., S. Sorooshian, H. V. Gupta, and P. R. Houser (2005b), Dual state–parameter estimation of hydrological models using ensemble Kalman filter, *Adv. Water Resour.*, *28*(2), 135–147, doi:10.1016/j.advwatres.2004.09.002.
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 Resour. Res.*, *48*, W12520, doi:10.1029/2012WR012144.
Moussa, R., and N. Chahinian(2009), Comparison of different multi-objective calibration criteria using a conceptual rainfall-runoff model of flood events, *Hydrol. Earth Syst. Sci.*, *13*, 519–535.
Musso, C., N. Oudjane, and F. LeGland (2001), Improving regularised particle filters, in *Sequential Monte Carlo Methods in Practice*, edited by A. Doucet, J. F. G. de Freitas, and N. J. Gordon, pp. 247–271, Springer, N. Y.
Noh, S. J. I. N., Y. Tachikawa, M. Shiiba, and S. Kim (2013), Sequential data assimilation for streamflow forecasting using a distributed hydrologic model: Particle filtering and ensemble Kalman filtering, in *Floods: From Risk to Opportunity, IAHS Publ.*, vol. 357, 341–349.
Paik, K., J. H. Kim, H. S. Kim, and D. R. Lee (2005), A conceptual rainfall-runoff model considering seasonal variation, *Hydrol. Processes*, *19*(19), 3837–3850, doi:10.1002/hyp.5984.
Pathiraja, S., S. Westra, and A. Sharma (2012), Why continuous simulation? The role of antecedent moisture in design flood estimation, *Water Resour. Res.*, *48*, W06534, doi:10.1029/2011WR010997.
Reichle, R. H., D. B. McLaughlin, and D. Entekhabi (2002), Hydrologic Data Assimilation with the Ensemble Kalman Filter, *Mon. Weather Rev.*, *130*(1), 103–114, doi:10.1175/1520-0493(2002)130 < 0103:HDAWTE>2.0.CO;2.
Richards, J. A. (1983), *Analysis of Periodically Time-Varying Systems*, Springer, N. Y.
Salamon, P., and L. Feyen (2009), Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter, *J. Hydrol.*, *376*(3-4), 428–442, doi:10.1016/j.jhydrol.2009.07.051.
Samuel, J., P. Coulibaly, G. Dumedah, and H. Moradkhani (2014), Assessing model state variation in hydrologic data assimilation, *J. Hydrol.*, *513*, 127–141, doi:10.1016/j.jhydrol.2014.03.048.
Schwartz, C. A., and H. Ozbay (1990), An identification procedure for linear continuous time systems with jump parameters, in *Proceedings of the International Symposium MTNS-89 Volume 1 on Realization and Modelling in Systems Theory*, edited M. A. Kaashoek, J. H. van Schuppen, and A. C. M. Ran, pp. 471-480, Birkhaüser Boston, Berlin.
Scott, D. F., and D. B. Van Wyk (1990), The effects of wildfire on soil wettability and hydrological behaviour of an afforested catchment, *J. Hydrol.*, *121*, 239–256.
Seibert, J., and J. J. McDonnell (2010), Land-cover impacts on streamflow: A change-detection modelling approach that incorporates parameter uncertainty, *Hydrol. Sci. J.*, *55*(3), 316–332, doi:10.1080/02626661003683264.
Simon, D. (2006), *Optimal State Estimation, Kalman, H Infinity, and Non-linear Approaches*, 1st ed. Wiley-Interscience.
Simon, E., and L. Bertino (2012), Gaussian anamorphosis extension of the DEnKF for combined state parameter estimation: Application to a 1D ocean ecosystem model, *J. Mar. Syst.*, *89*(1), 1–18, doi:10.1016/j.jmarsys.2011.07.007.
Siriwardena, L., B. L. Finlayson, and T. A. McMahon (2006), The impact of land use change on catchment hydrology in large catchments: The Comet River, Central Queensland, Australia, *J. Hydrol.*, *326*(1-4), 199–214, doi:10.1016/j.jhydrol.2005.10.030.
Smith, P. J., K. J. Beven, and J. A. Tawn (2008), Detection of structural inadequacy in process-based hydrological models: A particle-filtering approach, *Water Resour. Res.*, *44*, W01410, doi:10.1029/2006WR005205.
Smith, P. J., G. D. Thornhill, S. L. Dance, A. S. Lawless, D. C. Mason, and N. K. Nichols (2013), Data assimilation for state and parameter estimation: Application to morphodynamic modelling, *Q. J. R. Meteorol. Soc.*, *139*(671), 314–327, doi:10.1002/qj.1944.
Smith, T., A. Sharma, L. Marshall, R. Mehrotra, and S. Sisson (2010), Development of a formal likelihood function for improved Bayesian inference of ephemeral catchments, *Water Resour. Res.*, *46*, W12551, doi:10.1029/2010WR009514.
Sorooshian, S., V. K. Gupta, and J. F. Fulton (1983), Evaluation of maximum likelihood parameter estimation techniques for and length on model credibility, *Water Resour. Res.*, *19*(1), 251–259.
Thirel, G., V. Andréassian, C. Perrin, J. N. Audouy, L. Berthet, P. Edwards, and J. Vaze (2015), Hydrology under change: An evaluation protocol to investigate how hydrological models deal with changing catchments, *Hydrol. Sci. J.*, *60*(7–8), 1184–1199. doi:10.1080/02626667.2014.967248.
Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and R. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modelling: A case study using Bayesian total error analysis, *Water Resour. Res.*, *45*, W00B14, doi:10.1029/2008WR006825.
Tomas-Rodriguez, M., and S. P. Banks (2010), *Linear, Time-varying Approximations to Nonlinear Dynamical Systems: With Applications in Control and Optimization, Lect. Notes Control Inf. Sci.*, Springer-Verlag, Berlin Heidelberg, pp. 75–76.
Vrugt, J. A., H. V. Gupta, B. O. Nuallain, and W. Bouten (2006), Real-time data assimilation for operational ensemble streamflow forecasting, *J. Hydrometeorol.*, *7*, 548–565.
Vrugt, J. A., C. J. F. ter Braak, C. G. H. Diks, and G. Schoups (2013), Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications, *Adv. Water Resour.*, *51*, 457–478, doi:10.1016/j.advwatres.2012.04.002.
Walker, J. P., G. R. Willgoose, and J. D. Kalma (2001), One-dimensional soil moisture profile retrieval by assimilation of near-surface observations: A comparison of retrieval algorithms, *Adv. Water Resour.*, *24*(6), 631–650.
Wan, E. A., and R. Van Der Merwe (2000), The Unscented Kalman Filter for Nonlinear Estimation, in *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000, AS-SPCC*, IEEE, Lake Louise, Alberta, Canada, pp. 153–158.
Wang, Y. C., P. S. Yu, and T. C. Yang (2010), Comparison of genetic algorithms and shuffled complex evolution approach for calibrating distributed rainfall-runoff model, *Hydrol. Processes*, *24*(8), 1015–1026.
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.
West, M. (1993), Mixture Models, Monte Carlo, Bayesian updating and dynamic models, in *Proceedings of the 24th Symposium on the Interface on Computing Science and Statistics*, pp. 325-333, Interface Found. of North Am., Fairfax Station, Va.
Westerberg, I. K., J. L. Guerrero, P. M. Younger, K. J. Beven, J. Seibert, S. Halldin, and C.-Y. Xu (2011), Calibration of hydrological models using flow-duration curves, *Hydrol. Earth Syst. Sci.*, *15*(7), 2205–2227, doi:10.5194/hess-15-2205-2011.

Westra, S., M. Thyer, M. Leonard, D. Kavetski, and M. Lambert (2014), A strategy for diagnosing and interpreting hydrological model nonstationarity, *Water Resour. Res.*, *50*, 5090–5113, doi:10.1002/2013WR014719.

Wu, K., and C. A. Johnston (2007), Hydrologic response to climate variability in a Great Lakes Watershed: A case study with the SWAT model, *J. Hydrol.*, *337*, 187–199.

Xie, X., S. Meng, S. Liang, and Y. Yao (2014), Improving streamflow predictions at ungauged locations with real-time updating: Application of an EnKF-based state-parameter estimation strategy, *Hydrol. Earth Syst. Sci.*, *18*(10), 3923–3936, doi:10.5194/hess-18-3923-2014.

Yang, X., and T. Delsole (2009), Using the ensemble Kalman filter to estimate multiplicative model parameters, *Tellus, Ser. A*, *61*(5), 601–609, doi:10.1111/j.1600-0870.2009.00407.x.

Ye, W., B. C. Bates, N. R. Viney, M. Sivapalan, and A. J. Jakeman (1997), Performance of conceptual rainfall-runoff models in low-yielding ephemeral catchments, *Water Resour. Res.*, *33*(1), 153–166.