frontiers

Dual Ensemble Kalman Filter Applied To Raster Parameters in a Hydrologic Model

William Cook [1,*], **Prof. Jesse Johnson** [2] **and Prof. Marko Maneta** [1,2]

[1] *Laboratory X, Institute X, Department X, Organization X, City X , State XX (only USA, Canada and Australia), Country X*

[2] *Laboratory X, Institute X, Department X, Organization X, City X , State XX (only USA, Canada and Australia), Country X*

Correspondence*:
Corresponding Author
william.cook@umontana.umt.edu

ABSTRACT

Abstract goes here.

Keywords: hydro, model, hydro, model, hydro, model, kalman, filter

1 INTRODUCTION

This research is part of the **daWUAP** project led by Marko Maneta. The **daWUAP** team has created a hydrological rainfall-runoff model that predicts streamflows across the state of Montana. **daWUAPhydroengine** is informed by a variety of smaller models, including a groundwater model, snow water equivalient (swe) model, and agricultural component model. Despite this complexity, **daWUAPhydroengine** is designed to be a quick and efficient model.

This paper focuses on the calibration of **daWUAPhydroengine** via a Dual State Parameter Estimation Ensemble Kalman Filter. Uniquely, this filter operates on multiple high dimensional parameters represented as raster images. These raster images are both ingested and outputted by the **daWUAPhydroengine** and were previously set at a constant value. The filter aims to both calibrate and distribute these variables geospatially.

Hydrological models generate various states that are generally determined by a set of time-invariant parameters. In practice these parameters are calibrated continuously and are tweaked as soon as new observations become available. Parameter calibration techniques have recently been the subject of research Xie and Zhang (2010) Sorooshian et al. (1993), although many of the earliest calibration methods are unable to account for all sources of error Evensen (1994). More recently, sequential parameter estimation through techniques such as the Ensemble Kalman Filter have been developed that can 1) simultaneously calibrate parameter and state estimates and 2) take all sources of uncertainty into account Evensen (2003).

The Kalman Filter is an efficient model for continuous parameter estimation because it is a sequential data assimilation algorithm, only needing the previous timestep's state estimate, parameter estimate, and co-variance matrices to update the current timestep's state estimate, parameter estimate, and co-variance matrices. Although the original Kalman Filter Kalman (1960) works well for linear problems, more advanced techniques must be used for non-linear dynamics. The extended Kalman Filter Jazwinski (1970) works for mildly non-linear systems but does not function optimally on heavily non-linear systems Miller et al. (1994). The Unscented Kalman Filter Julier and Uhlmann (1997) is an all-around improvement on the Extended Kalman Filter that allows for the filtering of highly non-linear systems. The Ensemble Kalman Filter Evensen (1994), a predecessor to the Unscented Kalman Filter, filters non-linear systems by generating an 'ensemble' of model instances and adding unique noise to each model's forcing data. The main advantage of this ensemble based approach is the substitution of the original Kalman Filter's error covarience matrix with an ensemble covariance matrix, which allows for the efficient computation of the covariance of high dimensional state vectors. The Dual State Ensemble Kalman Filter Moradkhani et al. (2005) extends the Ensemble Kalman Filter to also include static parameter estimation. This paper examines the Dual State Ensemble Kalman Filter's application to high dimensional geospatially distributed raster data.

Research has been done on applying dual state parameter estimation ensemble Kalman filtering to hydrologic models Moradkhani et al. (2005). Further research into the calibration of hydrologic parameters, particularly high dimensional geospatial parameters, will help both inform future hydrologic models and allow for examination of the effectiveness of different techniques that distribute geospatial parameters across a landscape.

The optimization of the parameters must be done in a way that is reasonably efficient. Since models such as **daWUAPhydroengine** ingest and output dense raster data, for example, many simpler calibration implementations will become unwieldy because of large co-variance matrices. Furthermore, since **daWUAPhydroengine** is designed to be a quick and efficient model, it is preferable that any calibration component is similarly efficient.

The main contribution of this paper is a new 'moving window' model for kernel smoothing. The 'moving window' technique allows the parameter perturbation technique be informed by a set number of previous timesteps. This paper also covers the case application of the Dual Ensemble Kalman Filter framework to the **daWUAP** hydrologic engine, which necessitates the calibration of 4 large raster parameters. Finally, the code used for this paper is modular and should be usable with any Bayesian model.

Section 2 covers the theory behind the Dual State Ensemble Kalman Filter. Section 3 discusses the 'moving window' improvement and its implementation within the DSKE parameter sampling methods. Section 4 covers the application of the Dual State Ensemble Kalman Filter to **daWUAPhydroengine**.

2 THE DUAL ENSEMBLE KALMAN FILTER

According to Jazwinski Jazwinski (1970) any discrete nonlinear stochastic-dynamic model can be defined as:

$$x_{t+1} = f(x_t, u_t, \theta_t) + \varepsilon_t \tag{1}$$

where x_t is an n dimensional vector representing the state variables of the model at time step t, u_t is a vector of forcing data (e.g temperature or precipitation) at time step t, and θ_t is a vector of model parameters which may or may not change per time step (e.g *soil beta* or *DDF*). The non-linear function f takes these variables as inputs. The noise variable ε_t accounts for both model structural error and for any uncertainty in the forcing data.

A state's observation vector z_t can be defined as

$$z_t = h(x_t, \theta_t) + \delta_t \tag{2}$$

Where the x_t vector represents the true state, θ_t represents the true parameters, $h(.)$ is a function that determines the relationship between observation and state vectors, and δ_t represents observation error. δ_t is Gaussian and independent of ε_t.

The Dual State Ensemble Kalman Filter can be split into three subsections: The prediction phase, the parameter correction phase, and the state correction phase.

2.1 Prediction Phase

In a Dual Ensemble Kalman filter, each ensemble member *i* is represented by a stochastic model similar to (1). The modified equation is as follows:

$$x^{i-}_{t+1} = f(x^{i+}_t, u^i_t, \theta^{i-}_t) + \omega_t, \quad i = 1, ..., n \tag{3}$$

Where n is the total number of ensembles. The $-/+$ superscripts denote corrected $(+)$ and uncorrected $(-)$ values. Note that θ^{i-}_t's t superscript does not necessarily denote that θ is time variant but rather indicates that parameter values change as they are filtered over time. The noise term ω_t accounts for model error and will hereafter be excluded from the state equation.

Errors in the forcing data are accounted for through the perturbation the forcing data vector u_t with random noise ζ^i_t to generate a unique variable u^i_t for each ensemble. ζ^i_t is drawn from a normal distribution with a covarience matrix Q^i_t.

$$u^i_{t+1} = u_t + \zeta^i_t, \quad \zeta^i_t \sim N(0, Q^i_t) \tag{4}$$

To generate the priori parameters θ^{i-}_{t+1} an evolution of the parameters similar to the evolution of the state variables must be implemented. To accomplish this the kernel smoothing technique developed by WestWest (1993) is used. Legacy implementations of parameter evolution added a small perturbation sampled from $N(0, \Sigma^\theta_t)$, where Σ^θ_t represents the covariance matrix of θ at timestep t. This legacy method of evolution resulted in overly disposed parameter samples and the loss of continuity between two consecutive points in timeChen et al. (2008). Kernel smoothing has been used effectively to solve this problem in previous Dual Ensemble Kalman filter implementations Moradkhani et al. (2005) and similar models Chen et al. (2008).

86

$$\theta^{i-}_{t+1} = a\theta^{i+}_t + (1-a)\bar{\theta}^+_t + \tau^i_t \tag{5}$$

$$\tau^i_t = N(0, h^2 V_t) \tag{6}$$

Where $\bar{\theta}^+_t$ is the mean of the parameters with respect to the ensembles, $V_t = var(\theta^{i+}_t)$, a is a shrinkage factor between (0,1) of the kernel location, and h is a smoothing factor. h is defined by $\sqrt{1 - a1/2}$, while a is generally between (.45,.49). Note that h and a tend to vary per model and optimal values for these parameters are generally found via experimentation Moradkhani et al. (2005) Anderson et al. (1999) Annan et al. (2005) Chen et al. (2008).

2.2 Parameter Correction Phase

In an Ensemble Kalman Filter, observations are perturbed to reflect model error. Therefore, the variable z^i_{t+1} is defined as follows:

$$z^i_{t+1} = z_{t+1} + \eta^i_{t+1}, \quad \eta^i_{t+1} = N(0, R_{t+1}) \tag{7}$$

Where z_{t+1} is an observation vector defined by (2) and η^i_{t+1} is a random perturbation drawn from a normal distribution with covarience matrix R_{t+1}. A set of state predictions that can be related to the observations are generated by running the priori state vector through the function $h(.)$:

$$\hat{y}^i_{t+1} = h(x^{i-}_{t+1}, \theta^{i-}_{t+1}) \tag{8}$$

The parameter update equation is similar to the update equation of the linear Kalman filter ($\hat{x}^+_t = \hat{x}^-_t + K_t(z_t - H\hat{x}_t)$) Notably, parameters are corrected in lieu of the states:

$$\theta^{i+}_{t+1} = \theta^{i-}_{t+1} + K^\theta_{t+1}(z^i_{t+1} - \hat{y}^i_{t+1}) \tag{9}$$

To facilitate this, K^θ_{t+1} is defined as

$$K^\theta_{t+1} = \frac{\Sigma^{\theta,\hat{y}}_{t+1}}{\Sigma^{\hat{y},\hat{y}}_{t+1} + R_{t+1}} \tag{10}$$

where $\Sigma^{\theta,\hat{y}}_{t+1}$ is the cross covariance of θ_{t+1} and $\hat{y}_{t+1}$, $\Sigma^{\hat{y},\hat{y}}_{t+1}$ is the covarience of $\hat{y}_{t+1}$, and R_{t+1} is the observation error matrix from (7).

2.3 State Correction Phase

After θ^{i+}_{t+1} has been calculated the model is run again (3) with the θ^{i+}_{t+1} replacing θ^{i-}_{t+1}.

$$x^{i-}_{t+1} = f(x^{i+}_t, u^i_t, \theta^{i+}_t), \quad i = 1, ..., n \tag{11}$$

After a new state vector is generated it is re-run through (8) with the new parameter vector:

$$\hat{y}^i_{t+1} = h(x^{i-}_{t+1}, \theta^{i+}_{t+1}) \tag{12}$$

The corrected state vector is then run through the state update equation

$$x^{i+}_{t+1} = x^{i-}_{t+1} + K^x_{t+1}(z^i_{t+1} - \hat{y}^i_{t+1}) \tag{13}$$

$$K^x_{t+1} = \frac{\Sigma^{x,\hat{y}}_{t+1}}{\Sigma^{\hat{y},\hat{y}}_{t+1} + R_{t+1}} \tag{14}$$

where $\Sigma_{t+1}^{x,\hat{y}}$ is the cross covariance of x_{t+1} and $\hat{y}_{t+1}$.

REFERENCES

Anderson, J. L., Anderson, S. L., Anderson, J. L., and Anderson, S. L. (1999). A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts. *Monthly Weather Review* 127, 2741–2758. doi:10.1175/1520-0493(1999)127⟨2741:AMCIOT⟩2.0.CO;2

Annan, J. D., Lunt, D. J., Hargreaves, J. C., and Valdes, P. J. (2005). Parameter estimation in an atmospheric GCM using the Ensemble Kalman Filter. *Nonlinear Processes in Geophysics* 12, 363–371. doi:10.5194/npg-12-363-2005

Chen, M., Liu, S., Tieszen, L., and Hollinger, D. (2008). An improved state-parameter analysis of ecosystem models using data assimilation. *Ecological Modelling* 219, 317–326. doi:10.1016/J.ECOLMODEL.2008.07.013

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research* 99, 10143. doi:10.1029/94JC00572

Evensen, G. (2003). The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics* 53, 343–367. doi:10.1007/s10236-003-0036-9

Jazwinski, A. H. (1970). Stochastic Processes and Filtering. *Mathematics in Science and Enginerring*

Julier, S. J. and Uhlmann, J. K. (1997). New extension of the Kalman filter to nonlinear systems (International Society for Optics and Photonics), vol. 3068, 182. doi:10.1117/12.280797

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering* 82, 35. doi:10.1115/1.3662552

Miller, R. N., Ghil, M., and Gauthiez, F. (1994). Advanced Data Assimilation in Strongly Nonlinear Dynamical Systems. *Journal of the Atmospheric Sciences* 51, 1037–1056. doi:10.1175/1520-0469(1994)051⟨1037:ADAISN⟩2.0.CO;2

Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R. (2005). Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources* 28, 135–147. doi:10.1016/j.advwatres.2004.09.002

Sorooshian, S., Duan, Q., and Gupta, V. K. (1993). Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model. *Water Resources Research* 29, 1185–1194. doi:10.1029/92WR02617

West, M. (1993). Mixture models, Monte Carlo, Bayesian updating, and dynamic models. *Computing Science and Statistics* doi:10.1002/ange.200802345

Xie, X. and Zhang, D. (2010). Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter. *Advances in Water Resources* 33, 678–690. doi:10.1016/J.ADVWATRES.2010.03.012

FIGURE CAPTIONS

Figure 1. Enter the caption for your figure here. Repeat as necessary for each of your figures

Figure 2. This is a figure with sub figures, **(A)** is one logo, **(B)** is a different logo.