

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

William Cook<sup>1,\*</sup>, Prof. Jesse Johnson<sup>2</sup> and Prof. Marko Maneta<sup>1,2</sup>

<sup>1</sup>Laboratory X, Institute X, Department X, Organization X, City X, State XX (only USA, Canada and Australia), Country X

<sup>2</sup>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.umn.edu

## 2 ABSTRACT

3 Abstract goes here.

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

## 1 INTRODUCTION

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

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

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

22 The Kalman Filter is an efficient model for continuous parameter estimation because it is a sequential  
23 data assimilation algorithm, only needing the previous timestep's state estimate, parameter estimate, and  
24 co-variance matrices to update the current timestep's state estimate, parameter estimate, and co-variance  
25 matrices. Although the original Kalman Filter [Kalman \(1960\)](#) works well for linear problems, more  
26 advanced techniques must be used for non-linear dynamics. The extended Kalman Filter [Jazwinski \(1970\)](#)  
27 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 covariance 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:

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

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  $\theta_t$  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  $\varepsilon_t$  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

$$z_t = h(x_t, \theta_t) + \delta_t \quad (2)$$

Where the  $x_t$  vector represents the true state,  $\theta_t$  represents the true parameters,  $h(\cdot)$  is a function that determines the relationship between observation and state vectors, and  $\delta_t$  represents observation error.  $\delta_t$  is Gaussian and independent of  $\varepsilon_t$ .

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:

$$x_{t+1}^{i-} = f(x_t^{i+}, u_t^i, \theta_t^{i-}) + \omega_t, \quad i = 1, \dots, n \quad (3)$$

Where  $n$  is the total number of ensembles. The  $-/+$  superscripts denote corrected (+) and uncorrected (−) values. Note that  $\theta_t^{i-}$ 's  $t$  superscript does not necessarily denote that  $\theta$  is time variant but rather indicates that parameter values change as they are filtered over time. The noise term  $\omega_t$  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  $\zeta_t^i$  to generate a unique variable  $u_t^i$  for each ensemble.  $\zeta_t^i$  is drawn from a normal distribution with a covariance matrix  $Q_t^i$ .

$$u_{t+1}^i = u_t + \zeta_t^i, \quad \zeta_t^i \sim N(0, Q_t^i) \quad (4)$$

To generate the priori parameters  $\theta_{t+1}^{i-}$  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 West (1993) is used. Legacy implementations of parameter evolution added a small perturbation sampled from  $N(0, \Sigma_t^\theta)$ , where  $\Sigma_t^\theta$  represents the covariance matrix of  $\theta$  at timestep  $t$ . This legacy method of evolution resulted in overly disposed parameter samples and the loss of continuity between two consecutive points in time (Chen 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)).

$$\theta_{t+1}^{i-} = a\theta_t^{i+} + (1-a)\bar{\theta}_t^+ + \tau_t^i \quad (5)$$

$$\tau_t^i = N(0, h^2 V_t) \quad (6)$$

Where  $\bar{\theta}_t^+$  is the mean of the parameters with respect to the ensembles,  $V_t = \text{var}(\theta_t^{i+})$ ,  $a$  is a shrinkage factor between (0,1) of the kernel location, and  $h$  is a smoothing factor.  $h$  is defined by  $\sqrt{1-a}/2$ , 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)).

## 92 2.2 Parameter Correction Phase

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

$$z_{t+1}^i = z_{t+1} + \eta_{t+1}^i, \quad \eta_{t+1}^i = N(0, R_{t+1}) \quad (7)$$

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

$$\hat{y}_{t+1}^i = h(x_{t+1}^{i-}, \theta_{t+1}^{i-}) \quad (8)$$

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

$$\theta_{t+1}^{i+} = \theta_{t+1}^{i-} + K_{t+1}^\theta(z_{t+1}^i - \hat{y}_{t+1}^i) \quad (9)$$

100 To facilitate this,  $K_{t+1}^\theta$  is defined as

$$K_{t+1}^\theta = \frac{\Sigma_{t+1}^{\theta, \hat{y}}}{\Sigma_{t+1}^{\hat{y}, \hat{y}} + R_{t+1}} \quad (10)$$

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

## 103 2.3 State Correction Phase

104 After  $\theta_{t+1}^{i+}$  has been calculated the model is run again (3) with the  $\theta_{t+1}^{i+}$  replacing  $\theta_{t+1}^{i-}$ .

$$x_{t+1}^{i-} = f(x_t^{i+}, u_t^i, \theta_t^{i+}), \quad i = 1, \dots, n \quad (11)$$

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

$$\hat{y}_{t+1}^i = h(x_{t+1}^{i-}, \theta_{t+1}^{i+}) \quad (12)$$

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

$$x_{t+1}^{i+} = x_{t+1}^{i-} + K_{t+1}^x(z_{t+1}^i - \hat{y}_{t+1}^i) \quad (13)$$

$$K_{t+1}^x = \frac{\Sigma_{t+1}^{x, \hat{y}}}{\Sigma_{t+1}^{\hat{y}, \hat{y}} + R_{t+1}} \quad (14)$$

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

## 108 REFERENCES

- 109 Anderson, J. L., Anderson, S. L., Anderson, J. L., and Anderson, S. L. (1999). A Monte Carlo  
 110 Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts.  
 111 *Monthly Weather Review* 127, 2741–2758. doi:10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2  
 112 Annan, J. D., Lunt, D. J., Hargreaves, J. C., and Valdes, P. J. (2005). Parameter estimation in an  
 113 atmospheric GCM using the Ensemble Kalman Filter. *Nonlinear Processes in Geophysics* 12, 363–371.  
 114 doi:10.5194/npg-12-363-2005  
 115 Chen, M., Liu, S., Tieszen, L., and Hollinger, D. (2008). An improved state-parameter analysis of ecosystem  
 116 models using data assimilation. *Ecological Modelling* 219, 317–326. doi:10.1016/J.ECOLMODEL.  
 117 2008.07.013  
 118 Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte  
 119 Carlo methods to forecast error statistics. *Journal of Geophysical Research* 99, 10143. doi:10.1029/  
 120 94JC00572  
 121 Evensen, G. (2003). The Ensemble Kalman Filter: Theoretical formulation and practical implementation.  
 122 *Ocean Dynamics* 53, 343–367. doi:10.1007/s10236-003-0036-9  
 123 Jazwinski, A. H. (1970). Stochastic Processes and Filtering. *Mathematics in Science and Engineering*  
 124 Julier, S. J. and Uhlmann, J. K. (1997). New extension of the Kalman filter to nonlinear systems  
 125 (International Society for Optics and Photonics), vol. 3068, 182. doi:10.1117/12.280797  
 126 Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic*  
 127 *Engineering* 82, 35. doi:10.1115/1.3662552  
 128 Miller, R. N., Ghil, M., and Gauthiez, F. (1994). Advanced Data Assimilation in Strongly Nonlinear  
 129 Dynamical Systems. *Journal of the Atmospheric Sciences* 51, 1037–1056. doi:10.1175/1520-0469(1994)  
 130 051<1037:ADAISN>2.0.CO;2  
 131 Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R. (2005). Dual state-parameter estimation  
 132 of hydrological models using ensemble Kalman filter. *Advances in Water Resources* 28, 135–147.  
 133 doi:10.1016/j.advwatres.2004.09.002  
 134 Sorooshian, S., Duan, Q., and Gupta, V. K. (1993). Calibration of rainfall-runoff models: Application of  
 135 global optimization to the Sacramento Soil Moisture Accounting Model. *Water Resources Research* 29,  
 136 1185–1194. doi:10.1029/92WR02617  
 137 West, M. (1993). Mixture models, Monte Carlo, Bayesian updating, and dynamic models. *Computing*  
 138 *Science and Statistics* doi:10.1002/ange.200802345  
 139 Xie, X. and Zhang, D. (2010). Data assimilation for distributed hydrological catchment modeling via  
 140 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.