**LITERATURE REVIEWS**

**Paper 1 : Data Assimilation Methods in the Earth Science** (Reichle, 2008)

**Why data assimilation:**

~ To merge measurements of any type, including remote sensing observations, with estimates from geophysical models.

~ to combine the complementary information from measurements and models of the Earth System into an optimal estimate of the geophysical fields of interest.

**Why data assimilation of RS observations is a good idea:**

1. Coverage
2. Observability
3. Resolutions
4. Data volume and redundancy
5. Additional information from models

**Methods:**

~ Initial conditions of parameters are crucial to reduce accumulating error.

~ Theory of data assimilation in ES rest on mathematical theory of ESTIMATION THEORY.

~ Data assimilation rely on linear theory and assume Gaussian error distribution.

~ Simplistic method may include replacing the model estimate with observation (***direct insertion***)- snow cover observation.

1. **A simple data assimilation system**

~ Considering a scalar model variable ***m*** with uncertainty

**~** And a corresponding scalar observation ***o*** with uncertainty

~ The goal is to find the least-squares estimates X-bar of the true state X based.

~ Determine the objective function J as:

Text, letter

Description automatically generated

1. **The Kalman Filter**

~ Share the static update with some of the variational techniques.

~ Explicitly compute the error covariances through additional matrix equation that propagates error information from one update time to the next, subject to possibly uncertain dynamics.

~ The error propagation is prohibitively expensive for large scale applications in the traditional Kalman Filter and its non-linear variant, the Extended Kalman Filter.

~ Can be derived from objective function

~ Successful for soil moisture data assimilation

~ The reduced-rank approximations such as the Ensemble Kalman filter (EnKF) are designed to reduce the number of degrees of freedom to a manageable level.

**Paper 2: Multi-Scale Hydrological Evaluation of the National Water Model Streamflow Data Assimilation** (Seo et al., 2021)

~ Basically, the incorporation of streamflow observations into NWM modeling framework and updates model-simulated values using the observed ones.

~ Reduce many sources of errors and uncertainties in meteorological inputs.

~ Performs the evaluation of the performance of streamflow DA realized in the NWM.

~ The DA scheme used is called ***nudging***, as it propagates downstream only.

~ Nudging consists of direct insertion (observed value replaces the model value without considering the uncertainty).

~ Nudge at the assimilation location is the difference between observed and model streamflow (i.e., model error) with limited temporal interpolation.

~ Alternative approach will include spatial smoothing with assigned weights

~ 70 stations were used (spatial/routed DA)

~ NWM consists of LSM and water routing elements (diffusive wave surface routing, saturated subsurface flow routing, and Muskingum-Cunge channel routing)

**Method**

Qd(t) = *C*1[Qu(t-1) + Nd(t-1)] + *C*2[Qu(t) + Nd(t-1)] + *C*3[Qd(t-1) + Nd(t-1)] + ()

Where,

Q = streamflow discharge

*t* = current time

*t* – 1 = previous time

*d* = downstream

*u* = upstream

*C*1, *C*2, *C*3 = Coefficients calculated using routing parameters

ql = lateral flow

D = wedge storage contribution from lateral flow.

Nd(t-1) = nudge

**DA Evaluation**

Text, letter

Description automatically generated

**Paper 3: Data Assimilation for streamflow forecasting: State-Parameter Assimilation versus Output Assimilation.** (Sun et al., 2017)

**Overview:**

* Compares two unique data assimilation techniques: (1) state-parameter assimilation (2) output assimilation in improving streamflow forecasting.
* Extended Kalman Filter (EKF) were used to perform the state Parameter assimilation.
* Autoregressive (AR) models were used by updating the model output errors.
* EKF were used is performed by updating the stored water content and soil curve number.

**Extended Kalman Filter and ARMA model**

**Extended Kalman Filter:** Optimal estimator that recursively couples observations into a linear model to update the model states.

* Kalman filter can only be applied to linear systems where both the dynamic and observation functions are linear.
* EKF deal with nonlinear system, which can be represented by the state-space functions.

= (1)

= (2)

where,

*Mk* = dynamic model function; = forcing term; = model error; = Observation,

*Hk* = observation function; = a posterior state vector; = a priori state; = error; = covariance matrix of ; = covariance matrix of

* EKF requires the non-linear system to be continuously derivable.
* Uses the first order derivative of the Taylor extension at an estimated point to represent the nonlinear equation

**ARMA Models:** Approximates stationary stochastic process by a suitable autoregressive moving average (ARMA) model. Expressed as:

Xt + a1xt-1 + … + apxt-p = 𝜀t + b1𝜀t-1 + … + bq𝜀t-q  (3)

where a1,……… ap, and b1,….. bq are parameters; and εt,……εt−q are white noise

**Part 4: Novel approach to nonlinear/non-Gaussian Bayesian state estimation.** (Gordon et al., 1993)

Overview

* Uses the Bootstrap filter technique
* Proposed for implementing recursive Bayesian filters.
* The required density of the state vector is represented as a set of random samples
* These samples are updated and propagated by the algorithm.
* The study showed that the bootstrap filter is greatly superior to the standard extended Kalman Filter.

Text

Description automatically generated with medium confidence

**Summary Literature Review**

**Functional Algorithm of the Ensemble Kalman Filter**

The procedure of running two or more related but different analytical models and then synthesizing the results into a single score or spread to improve the accuracy of predictive analytics and data mining applications is known as Ensemble modeling (Koshin et al., 2020). This study uses the ensemble Kalman Filter (EnKF), which is very similar to the unscented Kalman Filter (UKF) but rather uses the Monte Carlo method to choose a large number of sigma points (Reichle et al., 2002). EnKF starts by randomly generating many points distributed about the filter’s initial state, this distribution is proportional to the filter’s covariance. EnKF uses an ensemble of many state vectors that are randomly sampled around the estimate by adding perturbations at each update for predicting the next step (Shen & Tang, 2015). When the filter is initialized a large number of points are drawn from the initial state and covariance, the algorithm then proceeds to the prediction step where the sigma points are passed through the state transition function, and then perturbed by adding a bit of noise to account for the process noise (Yamanaka et al., 2019). At the update step the sigma points are translated into measurement space by passing them through the measurement function, they are perturbed by a small amount to account for the measurement noise (Roth et al., 2017). Figure 1 shows a covariance ellipse of two standard deviations illustrating how the points from a 2 by 2 array matrix of covariance P with mean. The general equations for implementing the EnKF is expressed mathematically as:

(1)

(2)

(3)

(4)

(5)

(6)

(7)

(8)

Where, **x** is the state mean, **P** is the covariance, ***N*** is the number of sigma points, is the set of sigma points, **K** is the Kalman gain of the update step, while **R** is the measurement noise matrix. Readers are directed to Roger (2015) for more information on the working mechanism and implementations of the ensemble Kalman filter.

|  |
| --- |
| Chart, scatter chart  Description automatically generated |

Figure 1. covariance ellipse plots representing two standard deviations

**USGS Data Retrieval**

The study adapted a python library/package (dataretrieval) for obtaining USGS streamflow data with emphasis on instantaneous flow (iv) directly from web service of the National Water Information Service (NWIS). The developed scripts get inputs such as start and end dates, site id, and the service which can vary and includes options such as instantaneous flow (iv), daily values (dv), statistics (stat), site info (site), discharge peaks (peaks), discharge measurements (measurements), and water quality samples (qwdata) (Jem, 2018). Peculiar to this study, we pulled stations only within the HUC01 and subset gauges in our study domain (sub-regional camel basins) for testing and implementations. This subsetting was done using ArcGIS tools and an additional R-package for building and/or subsetting Hydrofabric. The study developed two python scripts and one Jupyter notebook to run the scripts. The first script (named “usgs”) served as the USGS data retrieval script with output as the instantaneous flow. The second script “usgs\_bmi” contain the BMI wrapper that sets the data retrieval output to be callable within the NEXTGEN framework, while the notebook is the framework that runs both scripts.

**Study Area**

The study basin (Figure 1) is located at southeast of Connecticut, and made up of 7 catchments, 3 USGS gauges, and 5 nexuses. Five criteria were used to properly select an appropriate sub-regional basin for testing and running the developed data assimilation in this study. Firstly, the study ensured the selected basin contained multiple USGS gauges and nexus for efficient data assimilation and validation. Secondly, the selected USGS gauges for data assimilation were not more than 1km away from the nexus with a nexus available downstream and an outlet USGS gauge for validation. Thirdly, the catchment sizes were considerably small enough to reduce uncertainties in lumped models, the smallest and largest catchments measured 6.35 km2 and 21.95 km2 respectively, while the total estimated basin area were 84.18 km2. Fourthly, the basin itself is a camel basin with known natural flow. Finally, our USGS gauges is close to a Next Generation Weather Radar (NEXRAD) site to ensure efficient model performance and further calibration.

Map

Description automatically generated

Map

Description automatically generated

Figure 1. study basin located in southeast Connecticut, showing the 7 catchments, 3 USGS gauges and 5 nexuses. One of the nexuses overlap with the southernmost USGS gauge. The nexuses are in green, USGS gauges in red, and the flowpaths in blue.

**REFERENCES**

Gordon, N. J., Salmond, D. J., & Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F Radar and Signal Processing*, *140*(2), 107. https://doi.org/10.1049/ip-f-2.1993.0015

Jem, S. (2018). *dataretrieval: Download hydrologic data* [Python]. USGS-python. https://github.com/USGS-python/dataretrieval (Original work published 2018)

Koshin, D., Sato, K., Miyazaki, K., & Watanabe, S. (2020). An ensemble Kalman filter data assimilation system for the whole neutral atmosphere. *Geoscientific Model Development*, *13*(7), 3145–3177. https://doi.org/10.5194/gmd-13-3145-2020

Reichle, R. H. (2008). Data assimilation methods in the Earth sciences. *Advances in Water Resources*, *31*(11), 1411–1418. https://doi.org/10.1016/j.advwatres.2008.01.001

Reichle, R. H., McLaughlin, D. B., & Entekhabi, D. (2002). Hydrologic Data Assimilation with the Ensemble Kalman Filter. *Monthly Weather Review*, *130*(1), 103–114. https://doi.org/10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2

Roth, M., Hendeby, G., Fritsche, C., & Gustafsson, F. (2017). The Ensemble Kalman filter: A signal processing perspective. *EURASIP Journal on Advances in Signal Processing*, *2017*(1), 56. https://doi.org/10.1186/s13634-017-0492-x

Seo, B.-C., Krajewski, W. F., & Quintero, F. (2021). Multi-Scale Hydrologic Evaluation of the National Water Model Streamflow Data Assimilation. *JAWRA Journal of the American Water Resources Association*, *57*(6), 875–884. https://doi.org/10.1111/1752-1688.12955

Shen, Z., & Tang, Y. (2015). A modified ensemble Kalman particle filter for non-Gaussian systems with nonlinear measurement functions. *Journal of Advances in Modeling Earth Systems*, *7*(1), 50–66. https://doi.org/10.1002/2014MS000373

Sun, L., Seidou, O., & Nistor, I. (2017). Data Assimilation for Streamflow Forecasting: State–Parameter Assimilation versus Output Assimilation. *Journal of Hydrologic Engineering*, *22*(3), 04016060. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001475

Yamanaka, A., Maeda, Y., & Sasaki, K. (2019). Ensemble Kalman filter-based data assimilation for three-dimensional multi-phase-field model: Estimation of anisotropic grain boundary properties. *Materials & Design*, *165*, 107577. https://doi.org/10.1016/j.matdes.2018.107577