EVALUATION OF MAXIMUM LIKELIHOOD ENSEMBLE FILTER FOR REAL-TIME ASSIMILATION OF STREAMFLOW DATA IN OPERATIONAL STREAMFLOW FORECASTING

Arezoo Rafieei Nasab¹, Dong-Jun Seo¹, Haksu Lee²,³, Sunghee Kim⁴

¹Dept. of Civil Eng., The Univ. of Texas at Arlington, Arlington, TX
²NOAA/NWS, Office of Climate, Water and Weather Services, Silver Spring, MD
³Len Technologies, Oak Hill, VA
⁴Consultant, Bethesda, MD
In this presentation

- Motivation
- Methodology
  - EnKF, MLEF
- Problem formulation
- Error modeling
  - Model, observation errors
- Comparative evaluation of EnKF and MLEF
  - Under homoscedastic, heteroscedastic errors
  - Sensitivity analysis
- Conclusions and future research recommendations
Motivation

• Streamflow is the most widely available, high information-content hydrologic data for inference of soil moisture states of the basin
  • Assimilating streamflow data, however, involves highly nonlinear observation equations
    • Ensemble Kalman filter (EnKF)
      • Relative simple and easy to implement
      • Optimal only if the observation equation is linear
    • Maximum likelihood ensemble filter (MLEF)
      • Ensemble extension of variational assimilation (VAR)
      • Can handle nonlinear observation equations
      • No need for adjoint code
Ensemble Kalman filter

Problem: Nonlinear observation operation

Solution?: Augment the state vector $x$ with $H(x)$
(still assuming Gaussian pdf for new model vector)
Maximum likelihood ensemble filter

Use square-root forecast (prior) error covariance

\[ P_f^{1/2} = [p_1^f \quad p_2^f \quad \cdots \quad p_{N,s}^f] \]

\[ p_i^f = M(x + p_i^a) - M(x) \]

Minimize cost function in the ensemble subspace

\[ J = \frac{1}{2} [x - x^f]^T P_f^{-1} [x - x^f] + \frac{1}{2} [y_{obs} - H(x)]^T R^{-1} [y_{obs} - H(x)] \]

Similar to VAR, but:

- Non-differentiable iterative minimization with superior (Hessian) preconditioning
- Reduced –rank solution in ensemble subspace
- Estimate of analysis uncertainty available

From Zupanski (2005)
MLEF – Extension for model error

- Zupanski (2005) does not consider model error

- To account for model uncertainty, augment the square root of the forecast covariance matrix

\[
P_f(k) = M_{k-1,k} P_a(k-1) M_{k-1,k}^T + Q(k-1)
\]

\[
= \{M_{k-1,k} P_a(k-1)\}^{1/2} \{M_{k-1,k} P_a(k-1)\}^{T/2} + Q^{1/2}(k-1)Q^{T/2}(k-1)
\]

\[
P_f^{T/2}(k) = [\{M_{k-1,k} P_a(k-1)\}^{1/2} \ Q^{1/2}(k-1)]
\]
Fixed-lag smoother formulation

Beginning of the assimilation window
End of the assimilation window = prediction time

Assimilation window ~ length of unit hydrograph
Assimilation cycle ~ nominally once per hour

k-L-1  k-L  k-L+1  ...  k-1  k  k+1  k+2

Time (hrs)

All or part of the precipitation, PE, and streamflow data valid within the assimilation window is assimilated.
Fixed-lag smoother formulation (cont.)

1) Prescribe the initial background model states and their covariance

3) Solve for the initial model states, biases for precipitation and PE utilizing all available data within the current assimilation window

4) Integrate the model to the end of the assimilation window to obtain the updated IC’s valid at the current prediction time, k

2) Propagate the model states and their uncertainty an hour forward

Time (hrs)
Comparatively evaluation of EnKF and MLEF – Homoscedastic errors

- Study basin
  - MTPT2 in WGRFC (435 km², time-to-peak of 17 hrs)

- Nominal parameter settings
  - Streamflow obs error variance: 0.01 (cms)²
  - Precipitation obs error variance: 10 (mm/hr)²
  - Potential evaporation obs error variance = 1 (mm/hr)²
  - Runoff obs error variance = 0.1 (mm/hr)²
  - Model error (as a fraction of soil water bucket size) = 0.05 (i.e. 5%)
  - Ensemble size = 30
  - Number of streamflow data used in DA = 1 (valid at the end of the assimilation window)

- The same settings are used for both MLEF and EnKF
All y-axis units are in mm
mtpt2_00580_2004051424

FLOW (CMS)

TIME ELAPSED (HRS)

- OBSERVED
- SIMULATED
- MLEF: CONTROL
- MLEF: ENSEMBLE MEAN
- EnKF: ENSEMBLE MEAN
- MLEF: ENSEMBLE MEMBERS
- EnKF: ENSEMBLE MEMBERS
All y-axis units are in mm
mtpt2_00253_1997031820

FLOW (CMS)

TIME ELAPSED (HRS)

- OBSERVED
- SIMULATED
- MLEF: CONTROL
- MLEF: ENSEMBLE MEAN
- EnKF: ENSEMBLE MEAN
- MLEF: ENSEMBLE MEMBERS
- EnKF: ENSEMBLE MEMBERS
All y-axis units are in mm
Sensitivity analysis

- Model error (frac)
  - fraction of soil water bucket size: 0, 0.05, 5, 25

- Streamflow obs error variance (\( \sigma^2_q \))
  - 0.01, 0.1, 1 and 10 (cms)^2

- Ensemble size (ns)
  - 9, 30, 50 and 100 members

- Number of streamflow obs assimilated per cycle (nf)
  - 1, 2, 4 and 8 obs within the assimilation window

- Obs error variances in precipitation, PE and runoff
Streamflow Prediction

frac = 0

frac = 0.05

frac = 5

frac = 25
Streamflow Prediction

\[ \sigma^2_q = 0.01 \text{ (cms)} \]

\[ \sigma^2_q = 0.1 \text{ (cms)} \]

\[ \sigma^2_q = 1 \text{ (cms)} \]

\[ \sigma^2_q = 10 \text{ (cms)} \]
Heteroscedastic error modeling

Error variance in model runoff

\[
Q(t) = \int_0^t \{I(\tau) + w(\tau)\} \times u(t - \tau) d\tau
\]

\[
\sigma^2_{eq} = \sigma^2_w \int_0^t \int_0^t u(t - \tau) \times u(t - s) ds d\tau
\]

\[
\sigma^2_{eq} = \left(\frac{\sqrt{Q_{obs} + 135.5 - 11.5537}}{0.0714}\right)^2 \text{ (mm/hr)}^2
\]

Error variance in obs

\[
\sigma^2_q = (C_q \times Q_{obs} + \text{additive})^2 \text{ (cms)}^2
\]

\[
\sigma^2_p = (C_p \times P_{obs} + \text{additive})^2 \text{ (mm/hr)}^2
\]

\[
\sigma^2_e = 1 \text{ (mm/hr)}^2
\]
Streamflow Prediction

**CP = 0.05 - CQ = 0.003**

- Simulated
- MLP: Control
- MLP: Ensemble Mean
- ENF: Ensemble Mean
- WO DA: Ensemble Mean

**Lead Time (HRS)**

**CP = 0.15 - CQ = 0.003**

- Simulated
- MLP: Control
- MLP: Ensemble Mean
- ENF: Ensemble Mean
- WO DA: Ensemble Mean

**Lead Time (HRS)**

**CP = 0.25 - CQ = 0.003**

- Simulated
- MLP: Control
- MLP: Ensemble Mean
- ENF: Ensemble Mean
- WO DA: Ensemble Mean

**Lead Time (HRS)**

**CP = 0.05 - CQ = 0.03**

- Simulated
- MLP: Control
- MLP: Ensemble Mean
- ENF: Ensemble Mean
- WO DA: Ensemble Mean

**Lead Time (HRS)**

**CP = 0.15 - CQ = 0.03**

- Simulated
- MLP: Control
- MLP: Ensemble Mean
- ENF: Ensemble Mean
- WO DA: Ensemble Mean

**Lead Time (HRS)**

**CP = 0.25 - CQ = 0.03**

- Simulated
- MLP: Control
- MLP: Ensemble Mean
- ENF: Ensemble Mean
- WO DA: Ensemble Mean

**Lead Time (HRS)**

**CP = 0.05 - CQ = 0.3**

- Simulated
- MLP: Control
- MLP: Ensemble Mean
- ENF: Ensemble Mean
- WO DA: Ensemble Mean

**Lead Time (HRS)**

**CP = 0.15 - CQ = 0.3**

- Simulated
- MLP: Control
- MLP: Ensemble Mean
- ENF: Ensemble Mean
- WO DA: Ensemble Mean

**Lead Time (HRS)**

**CP = 0.25 - CQ = 0.3**

- Simulated
- MLP: Control
- MLP: Ensemble Mean
- ENF: Ensemble Mean
- WO DA: Ensemble Mean

**Lead Time (HRS)**
Conclusions & future research recommendations

- MLEF generally improves streamflow prediction over EnKF; the improvement is:
  - very significant at short lead times
  - consistent over varying conditions of observational and model errors and ensemble size
- At large lead times, EnKF tends to perform slightly better than MLEF
  - suggests possible overfitting by MLEF
- Performance of MLEF is much less sensitive to error modeling and ensemble size than that of EnKF
  - important consideration for operational applications
- While the streamflow results appear similar, the soil moisture results are quite different between MLEF and EnKF
  - reflects large differences in their solutions
Conclusions & future research recommendations (cont.)

- Approximate gradient evaluation in MLEF is not always successful (compared to the adjoint-based)
  - May result in temporal discontinuity streamflow and soil moisture results
  - Improved error modeling may alleviate the problem
- Need to test on larger-dimensional problems with varying degree of under-determinedness
  - Assess computational requirements as well
- Assess the quality of analysis (i.e. updated) ensembles via rigorous ensemble verification for both streamflow and soil moisture
- Compare with iterative EnKF
THANK YOU

Questions?

For more info, please contact arezoo.rafieeinasab@mavs.uta.edu.