COMPARATIVE EVALUATION OF ENKF AND MLEF FOR ASSIMILATION OF STREAMFLOW DATA INTO NWS OPERATIONAL HYDROLOGIC MODELS

Arezoo Rafieei Nasab[1], Dong-Jun Seo[1], Haksu Lee[2,3], Sunghee Kim[1]

[1]Dept. of Civil Eng., The Univ. of Texas at Arlington, Arlington, TX

[2]NOAA/NWS, Office of Climate, Water and Weather Services, Silver Spring, MD

[3]Len Technologies, Oak Hill, VA

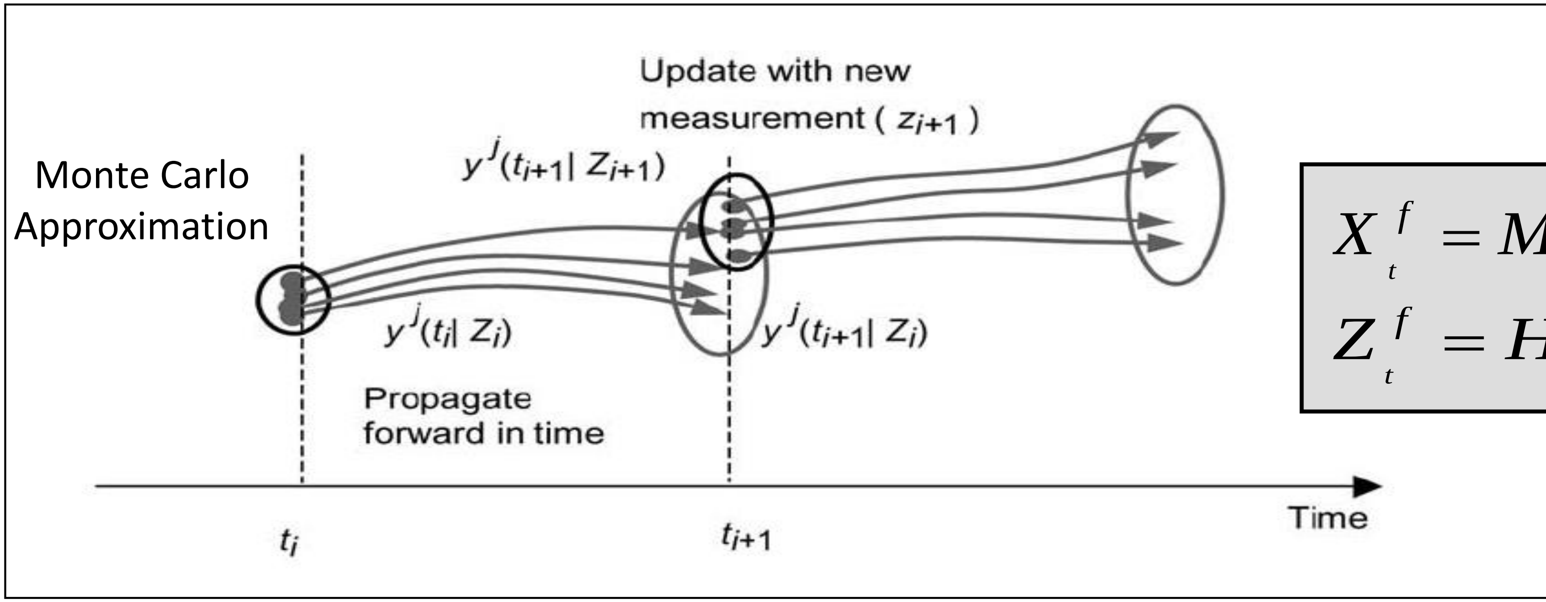
In this presentation

- Motivation
- Methodology
  - EnKF, MLEF
- Problem formulation
- Comparative evaluation of EnKF and MLEF
  - Homoscedastic versus heteroscedastic error modeling
  - 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



$$X_t^f = M_t(X_{t-1})$$
$$Z_t^f = H_t . X_t^f$$

Recursive updating of each ensemble trace

Problem: Nonlinear observation operation

$$Y_n = \begin{bmatrix} X_t^f \\ Z_t^f \end{bmatrix}$$

Solution?: Augment the state vector *x* with *H(x)*

The obs eq is linear only in appearance, still assumes linear response of streamflow to soil moisture near the solution

$$H = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Maximum likelihood ensemble filter

Use square-root forecast error covariance

$$\boldsymbol{P}_f^{1/2} = [\boldsymbol{p}_1^f \quad \boldsymbol{p}_2^f \quad \cdots \quad \boldsymbol{p}_{N_s}^f]$$

$$\boldsymbol{p}_i^f = \mathcal{M}(\boldsymbol{x} + \boldsymbol{p}_i^a) - \mathcal{M}(\boldsymbol{x})$$

Ensemble size

$$\boldsymbol{p}_i^f = \begin{pmatrix} p_{1,i}^f \\ p_{2,i}^f \\ \vdots \\ p_{S,i}^f \end{pmatrix}$$

State-space dimension

Minimize cost function (in ensemble subspace)

$$J = \frac{1}{2}[\boldsymbol{x} - \boldsymbol{x}^f]^T \boldsymbol{P}_f^{-1}[\boldsymbol{x} - \boldsymbol{x}^f] + \frac{1}{2}[\boldsymbol{y}_{obs} - \mathcal{H}(\boldsymbol{x})]^T \boldsymbol{R}^{-1}[\boldsymbol{y}_{obs} - \mathcal{H}(\boldsymbol{x})]$$

Similar to VAR, but:

- Uses non-differentiable iterative minimization with superior (Hessian) preconditioning
- Provides reduced–rank solution in ensemble subspace
- Estimates analysis uncertainty

From Zupanski (2005)

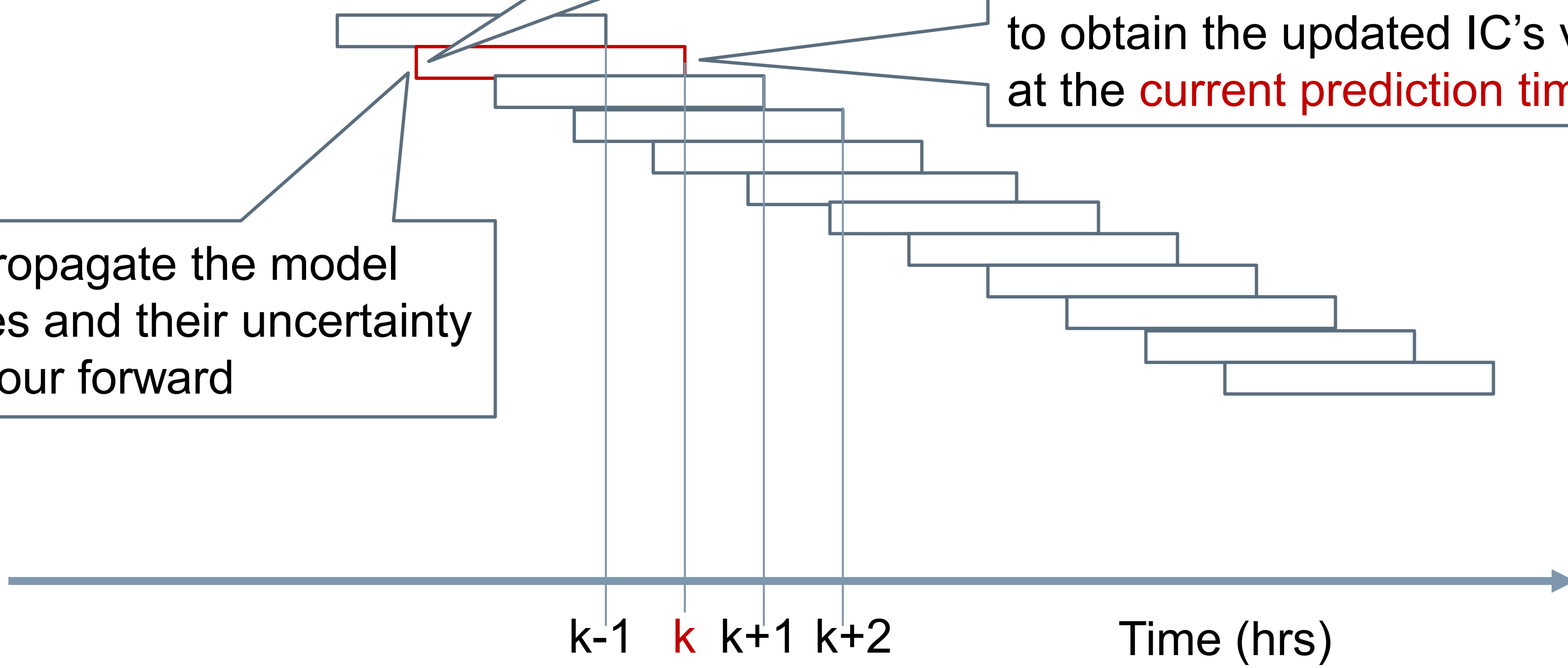
Fixed-lag smoother formulation

1) Prescribe the initial background model states and their covariance

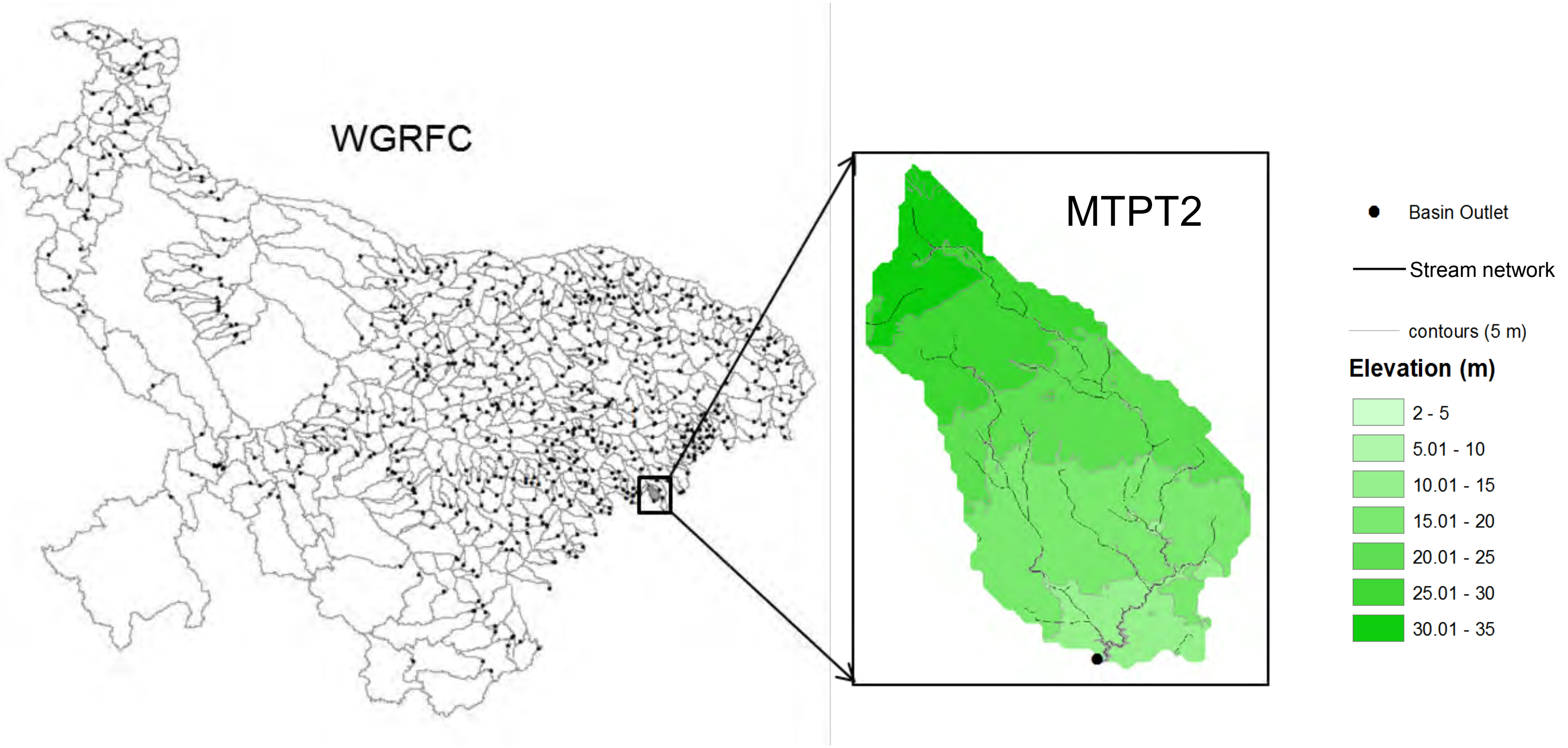
3) Solve for the initial model states, biases for precipitation and PE utilizing all available data within the current assimilation window

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



Study area



MTPT2 in WGRFC (435 km^2, time-to-peak ~17 hrs)

Summary of parameter settings

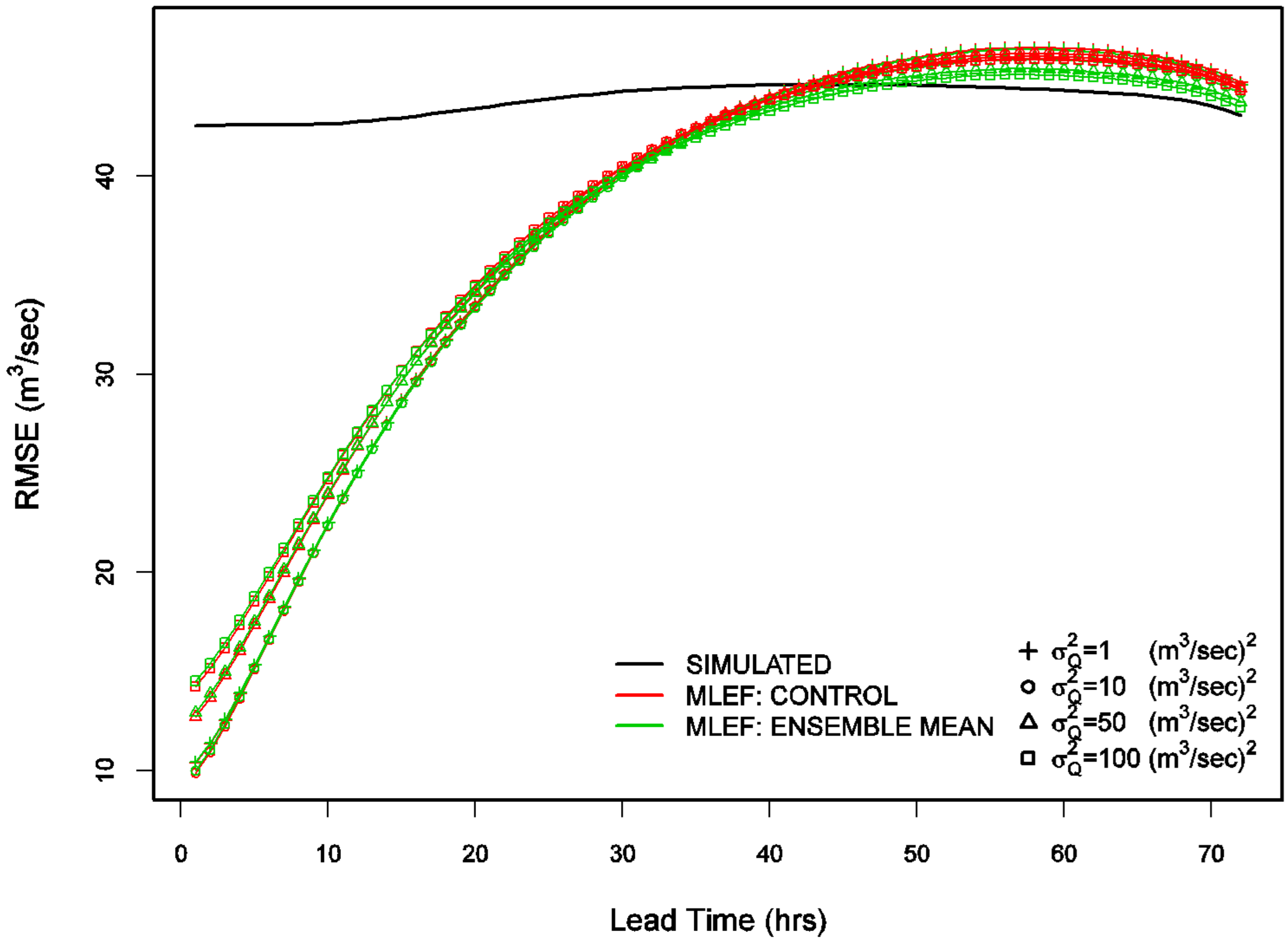
	Experiment	Stream-flow error variance $(m^3/s)^2$	MAP $(mm/hr)^2$	Additive error to TCI $(mm/hr)^2$	Fractional dynamical model error	Ensemble size	No. of Stream-flow data used/cycle
1	Streamflow error variance	**1, 10, 50, 100**	10	1	0.03	30	1
1	Additive error variance to TCI	10	10	**0.01, 0.1, 1, 10**	0.03	30	1
2	Homoscedastic error modeling	10	10	1	0.03	30	1
2	Heteroscedastic error modeling	**C_Q=0.03, 0.3**	**C_P=0.15, 0.25**	**Function of Z_Q**	0.03	30	1
3	Fractional dynamical model error	10	10	1	**0, 0.025, 0.075, 0.1**	30	1
3	Ensemble size	10	10	1	0.025	**5, 9, 30, 50**	1
3	No. of streamflow data used per cycle	10	10	1	0.025	30	**1, 2, 4, 8**

MAPE error variance = 1 $(mm/hr)^2$

Sensitivity to streamflow observation error variance

- The smaller the error variance, the closer the fit through the observed streamflow
- Smaller RMSE at short lead times but at some expense of larger RMSE at large lead times
- The DA-aided simulation has slightly larger RMSE than DA-less simulation at large lead times

uncertainty modeling needs improvement



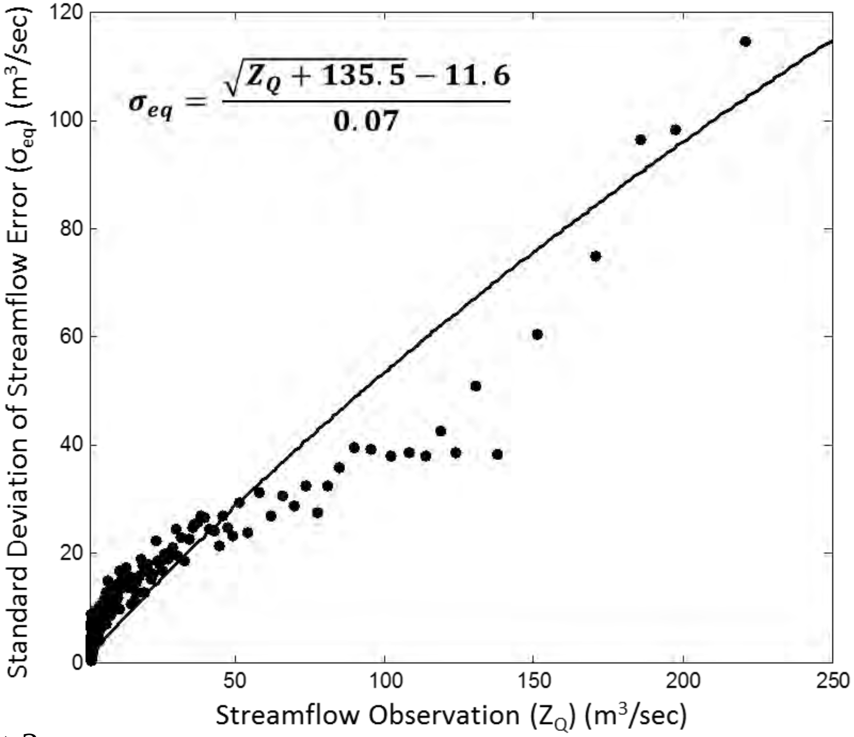
Heteroscedastic error modeling

Error variance in model runoff

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

$$\sigma^2{}_{eq} = \sigma^2{}_w \int_0^t \int_0^t u(t-\tau) \times u(t-s)dsd\tau$$

$$\sigma^2{}_{eq} = \left(\frac{\sqrt{Q_{obs} + 135.5} - 11.6}{0.07}\right)^2 \quad \text{(cms)}^2$$



Error variance in obs

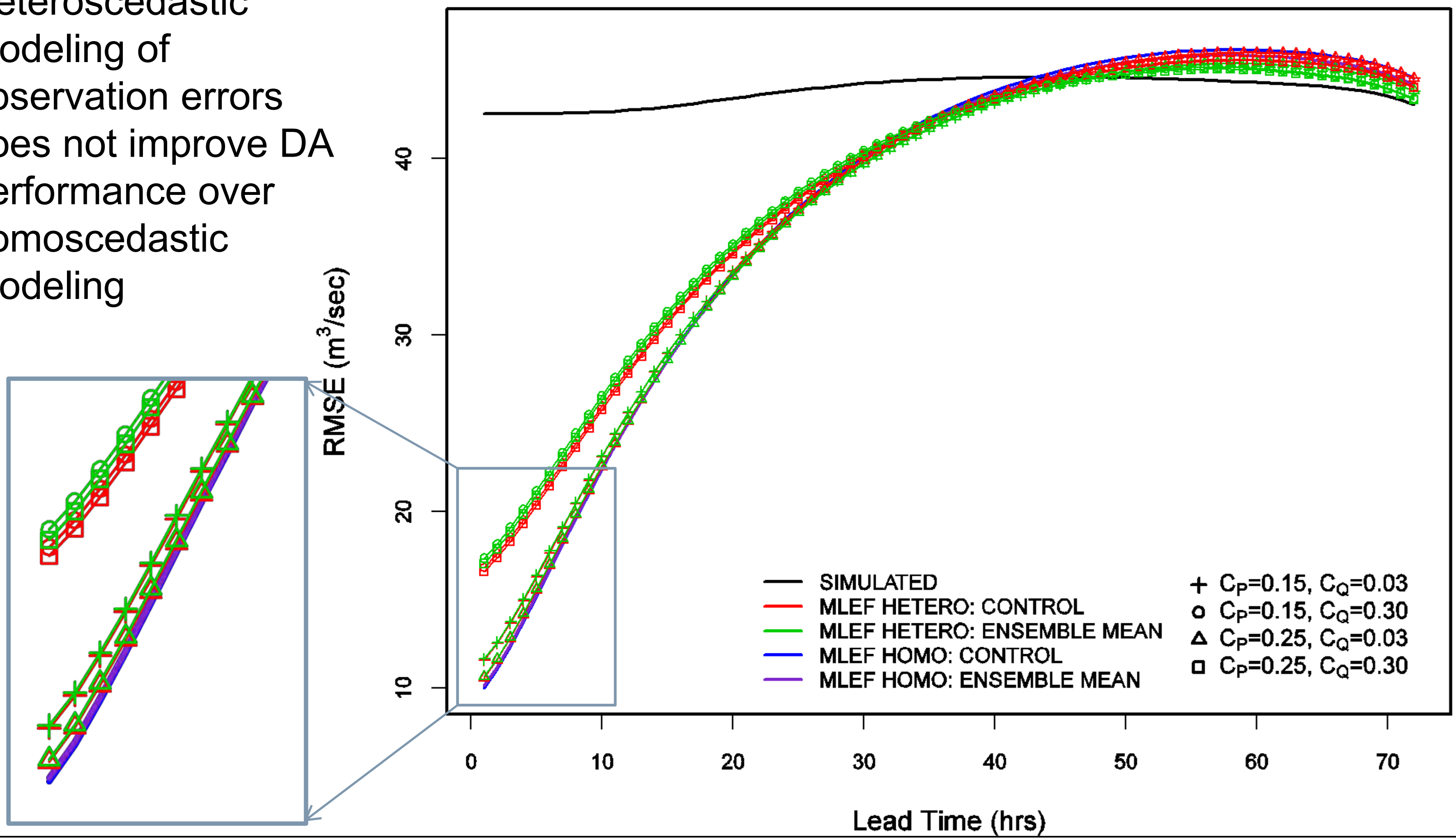
$$\sigma^2{}_q = (\boxed{C_q} * Q_{obs} + additive\ )^2 \quad \text{(cms)}^2$$

$$\sigma^2{}_P = (\boxed{C_p} * P_{obs} + additive)2 \quad \text{(mm/hr)}^2$$

$$\sigma^2{}_e = 1 \ \text{(mm/hr)}^2$$

Heteroscedastic error modeling

- Heteroscedastic modeling of observation errors does not improve DA performance over homoscedastic modeling



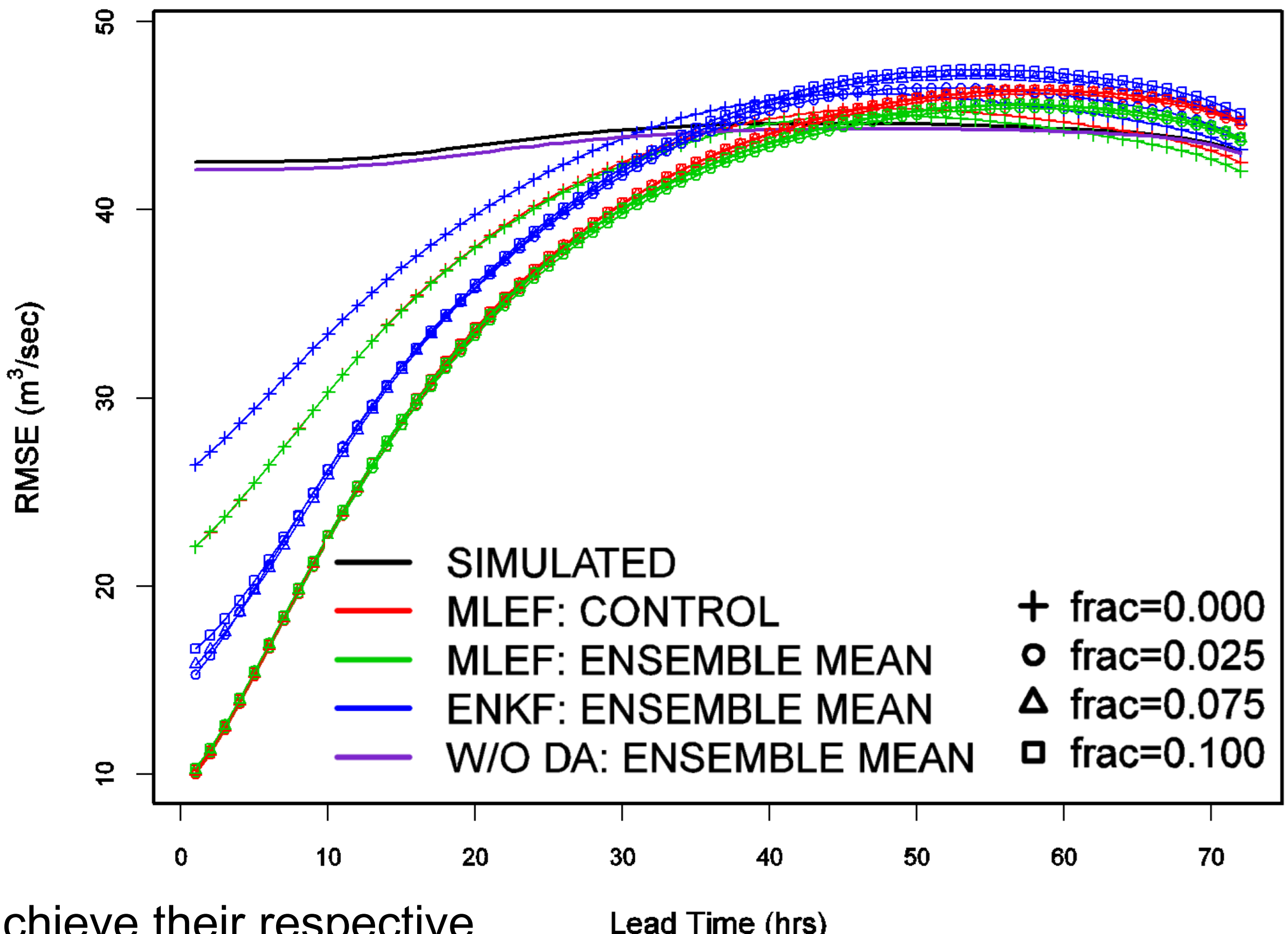
Summary of parameter settings

	Experiment	Streamflow error variance $(m^3/s)^2$	MAP $(mm/hr)^2$	Additive error to TCI $(mm/hr)^2$	Fractional dynamical model error	Ensemble size	No. of Streamflow data used/cycle
1	Streamflow error variance	**1, 10, 50, 100**	10	1	0.03	30	1
	Additive error variance to TCI	10	10	**0.01, 0.1, 1, 10**	0.03	30	1
2	Homoscedastic error modeling	10	10	1	0.03	30	1
	Heteroscedastic error modeling	**C_Q=0.03, 0.3**	**C_P=0.15, 0.25**	**Function of Z_Q**	0.03	30	1
3	Fractional dynamical model error	10	10	1	**0, 0.025, 0.075, 0.1**	30	1
	Ensemble size	10	10	1	0.025	**5, 9, 30, 50**	1
	No. of streamflow data used per cycle	10	10	1	0.025	30	**1, 2, 4, 8**

MAPE error variance = 1 $(mm/hr)^2$

Model error

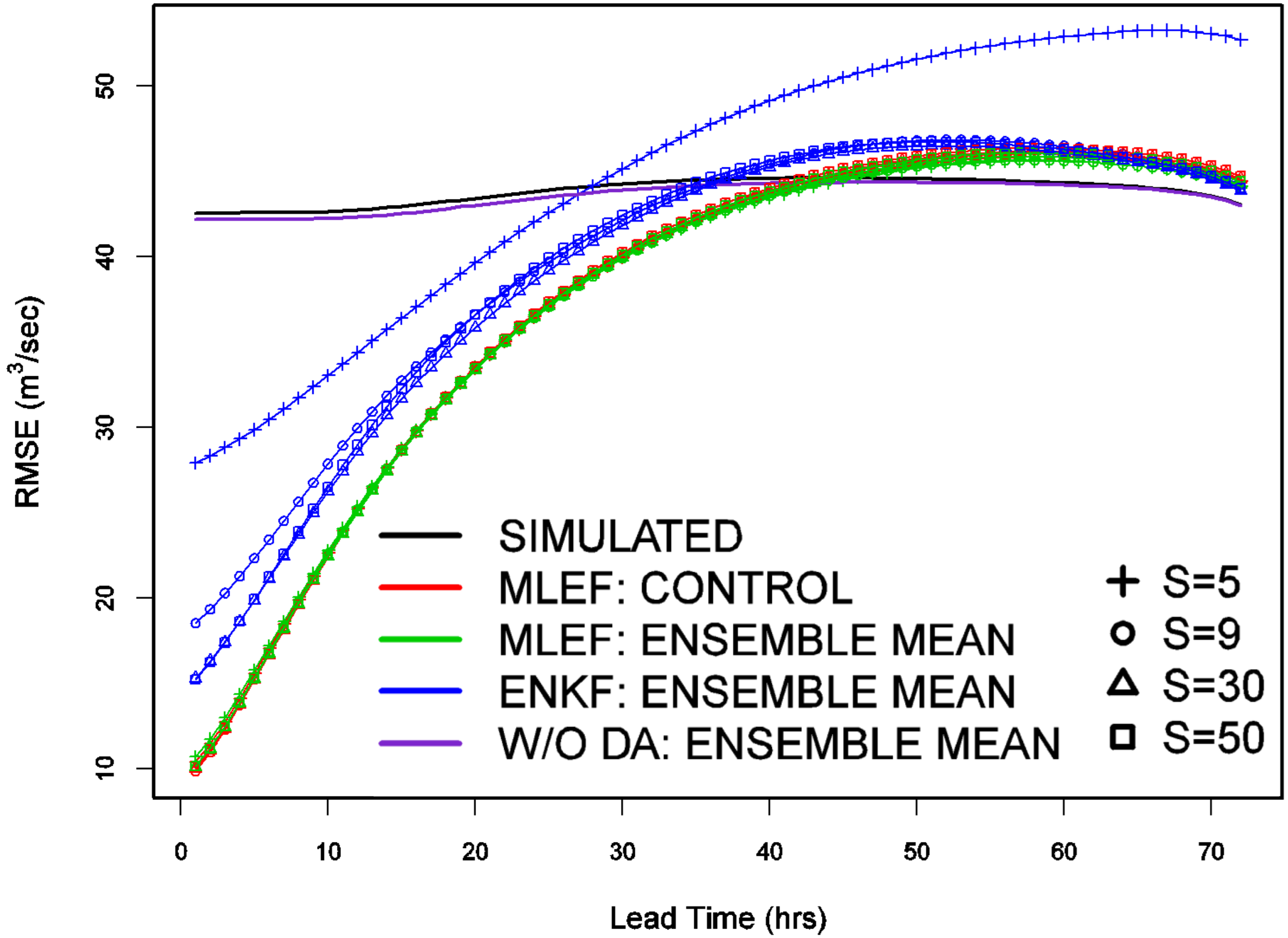
- MLEF
  - fraction of soil water bucket size
- EnKF
  - fraction of soil water content
- Accounting for model errors in soil moisture dynamics improves the performance of DA significantly at short lead times



- Both MLEF and EnKF achieve their respective best with a fraction of 0.025.

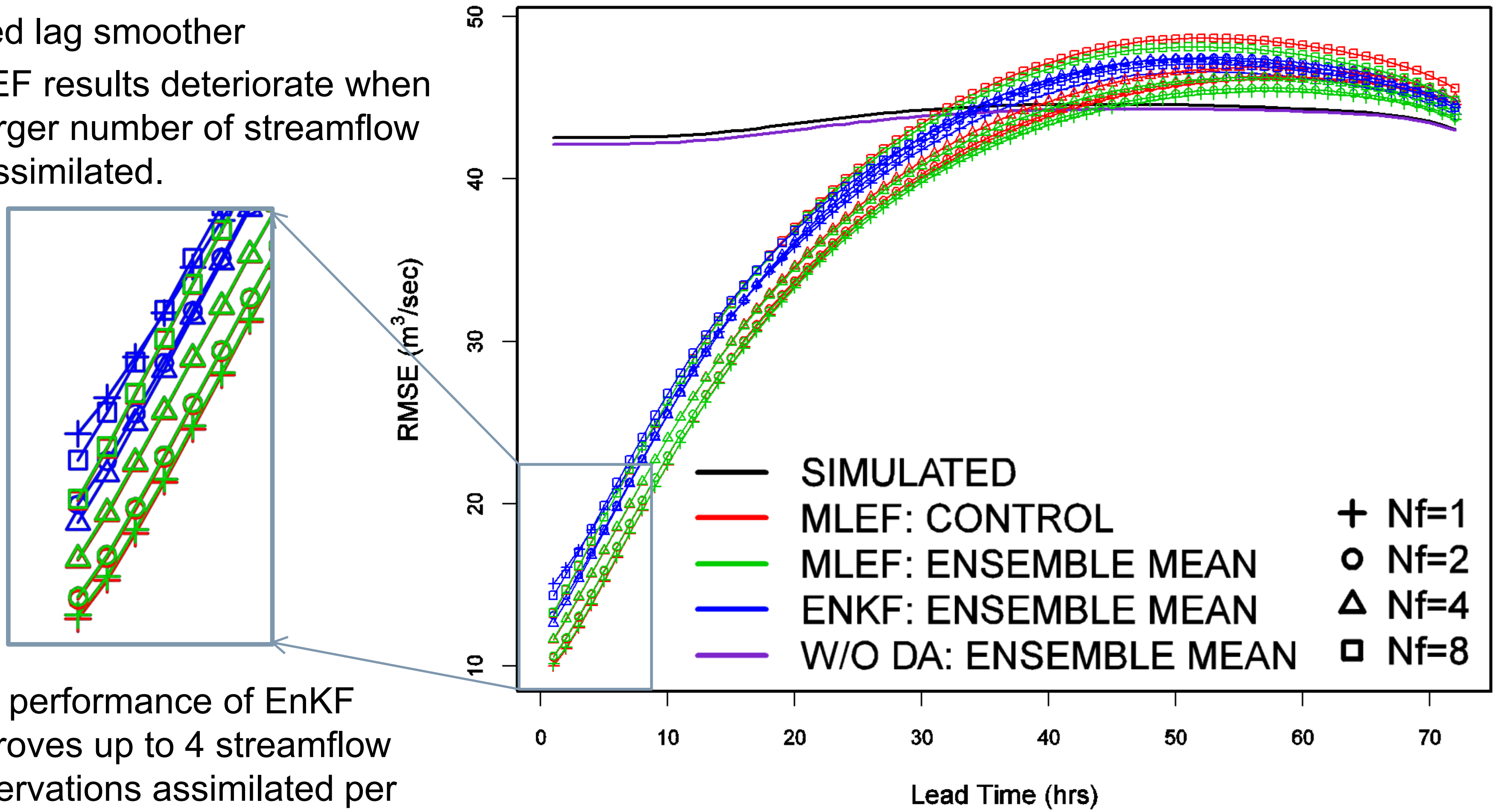
Ensemble size

- MLEF is not very sensitive to ensemble size
- The EnKF solution generally improves with increasing ensemble size but does not come close to the MLEF solution even with 50 members
- The CPU time for MLEF is considerably smaller than that for EnKF



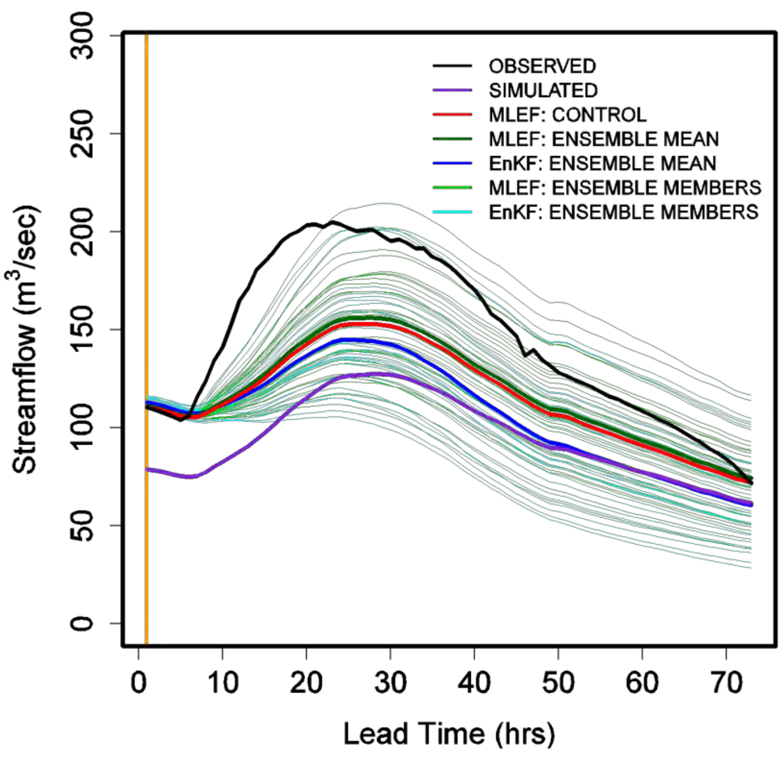
Number of streamflow obs assimilated per cycle

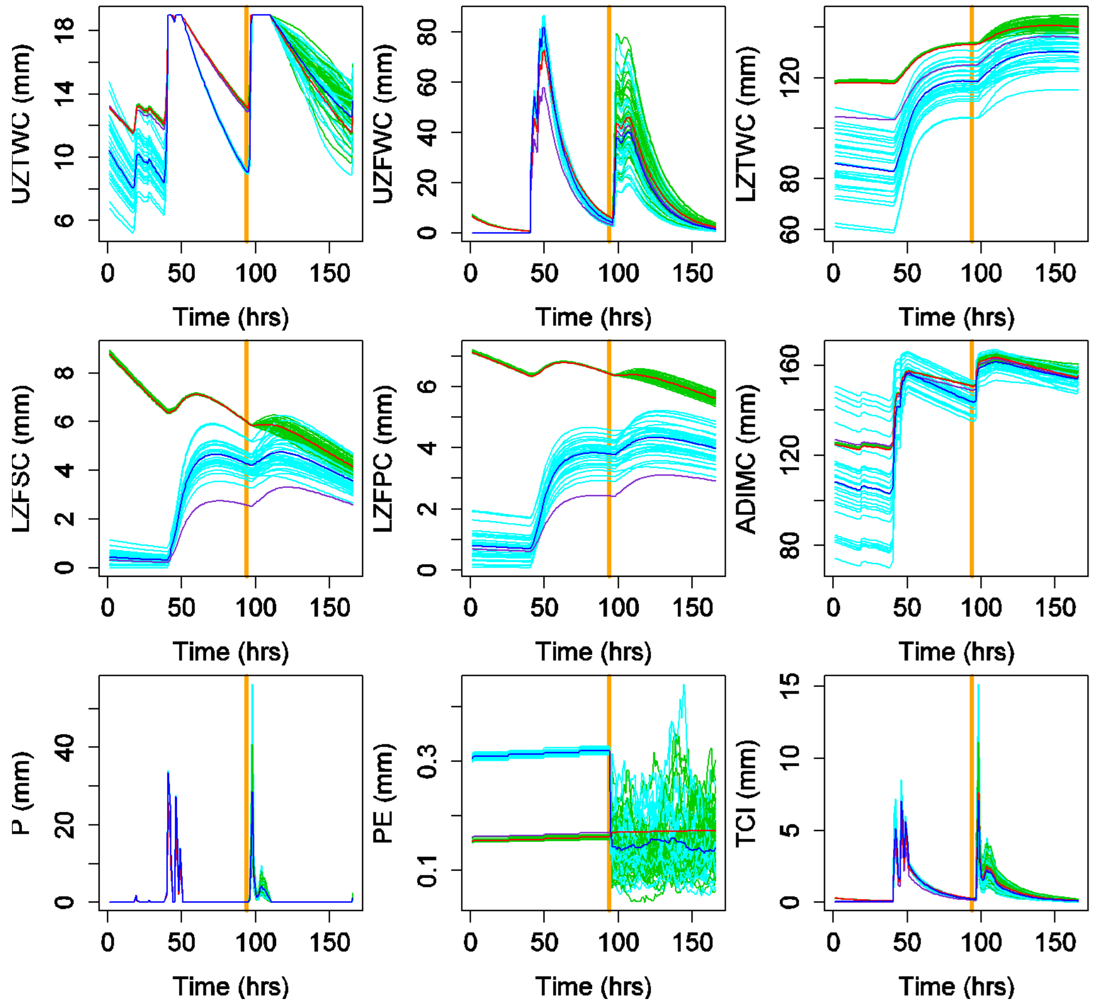
- Fixed lag smoother
- MLEF results deteriorate when a larger number of streamflow is assimilated.



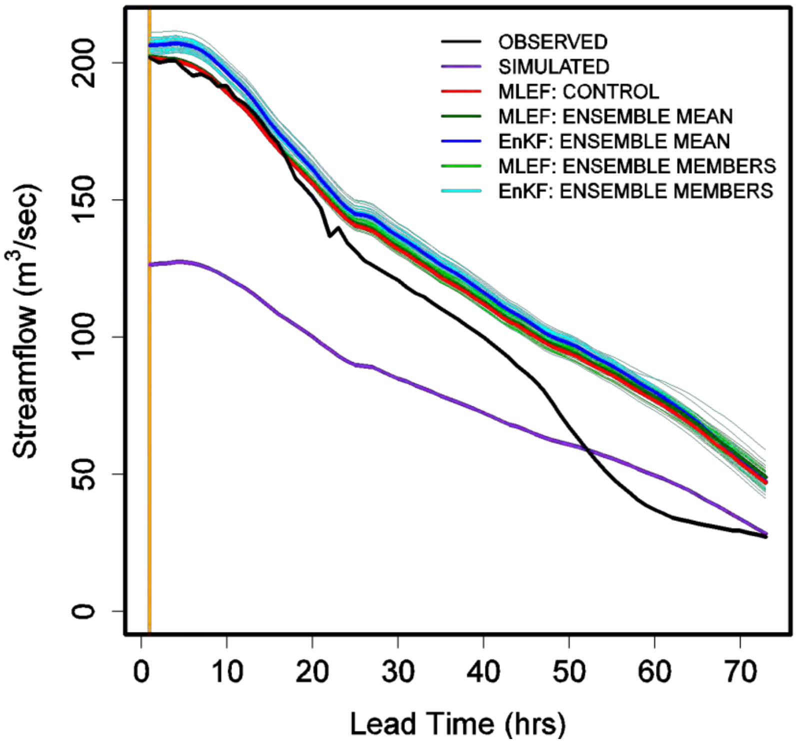
- The performance of EnKF improves up to 4 streamflow observations assimilated per cycle and then decreases

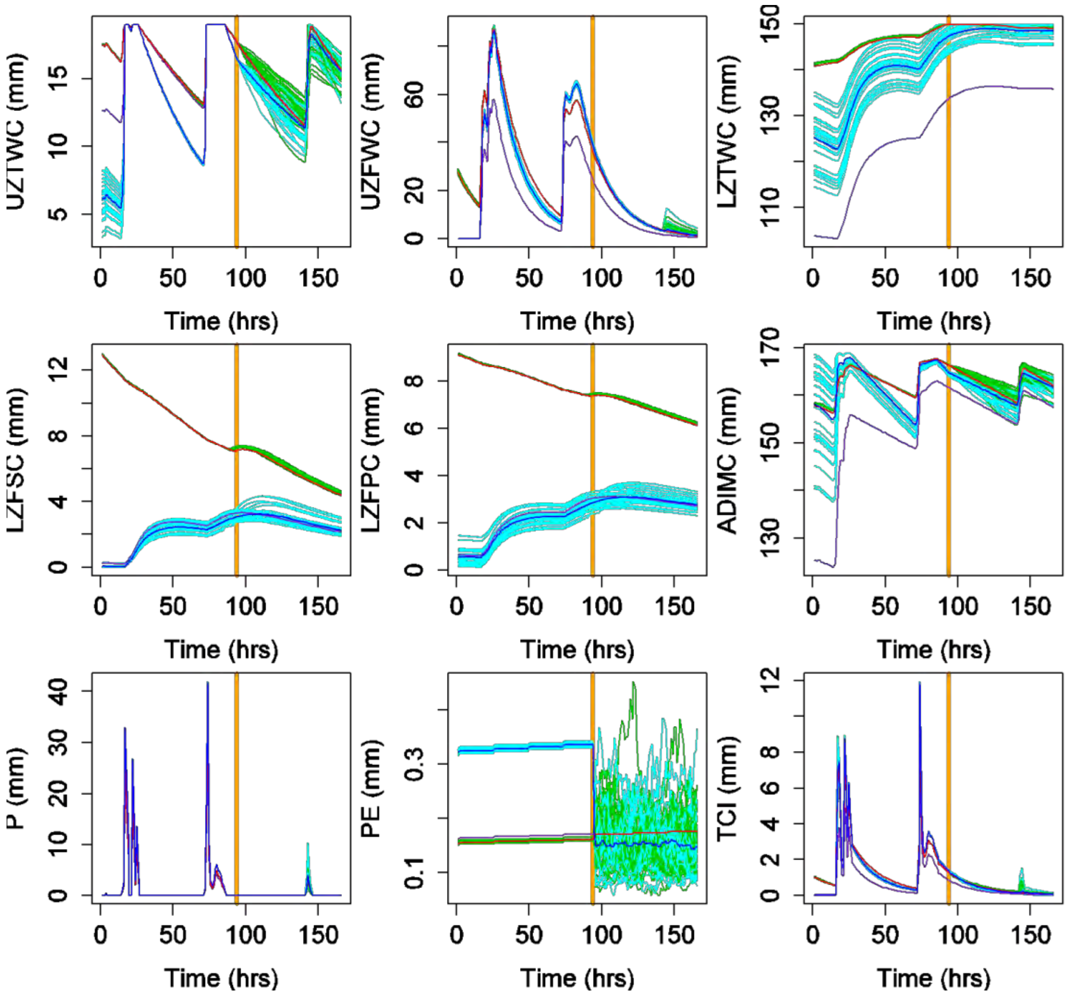
Example results



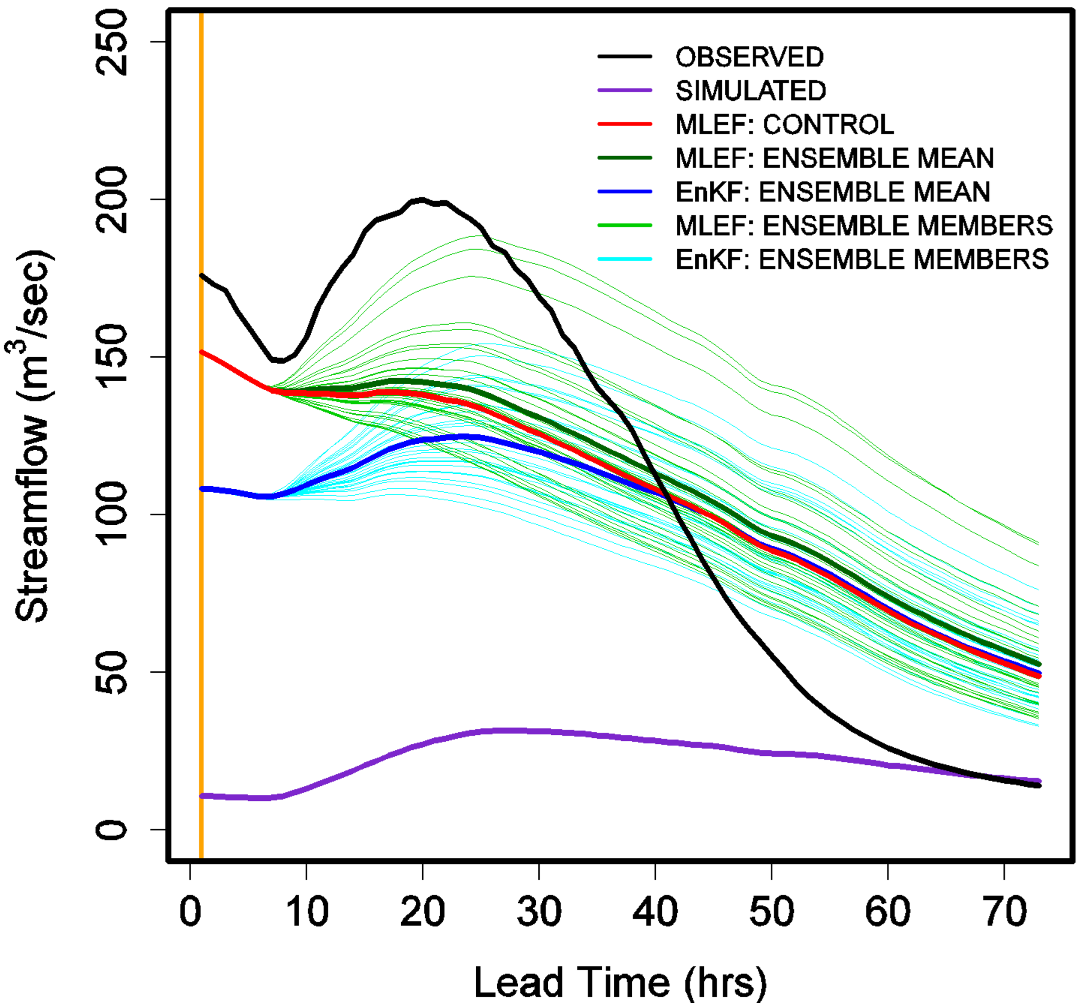


Example results (cont.)





An example of significantly different performance between MLEF and EnKF



Conclusions & future research recommendations

- MLEF generally improves streamflow prediction over EnKF
  - very significant at short lead times
- 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
- Computational requirements for MLEF is smaller than those for EnKF
- While the streamflow results appear similar, the soil moisture results are quite different between MLEF and EnKF
  - Reflects possible under-determinedness of the problem

Conclusions & future research recommendations (cont.)

- Approximate gradient evaluation in MLEF is not always successful (compared to the adjoint-based)
  - May result in temporal discontinuity in streamflow and soil moisture results

- Need to test on larger-dimensional problems with varying degree of under-determinedness (will be covered by Sunghee Kim on Session 8: Real-world Applications of Data Assimilation in Operational Hydrology)

- Assess the quality of analysis (i.e. updated) ensembles via rigorous ensemble verification for both streamflow and soil moisture

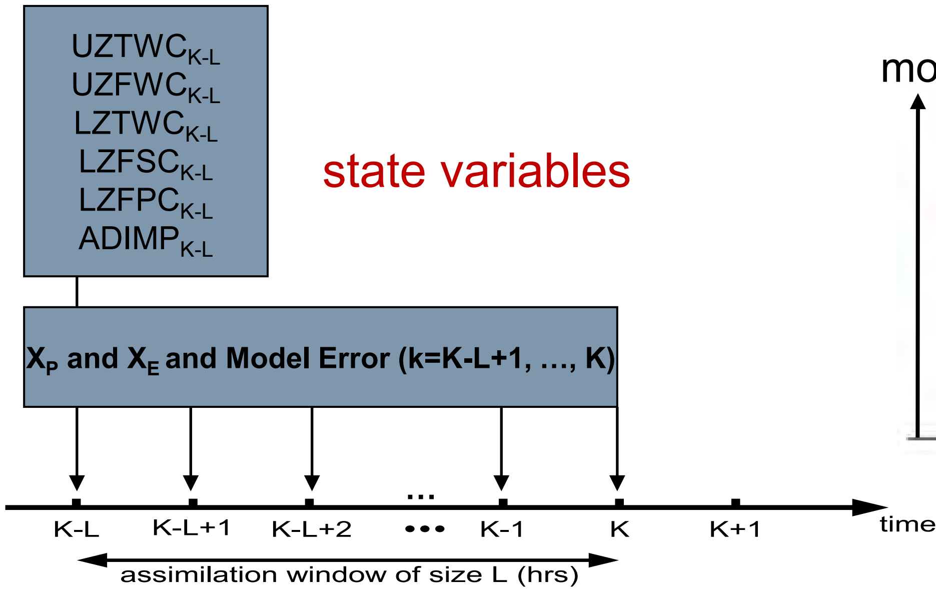
THANK YOU

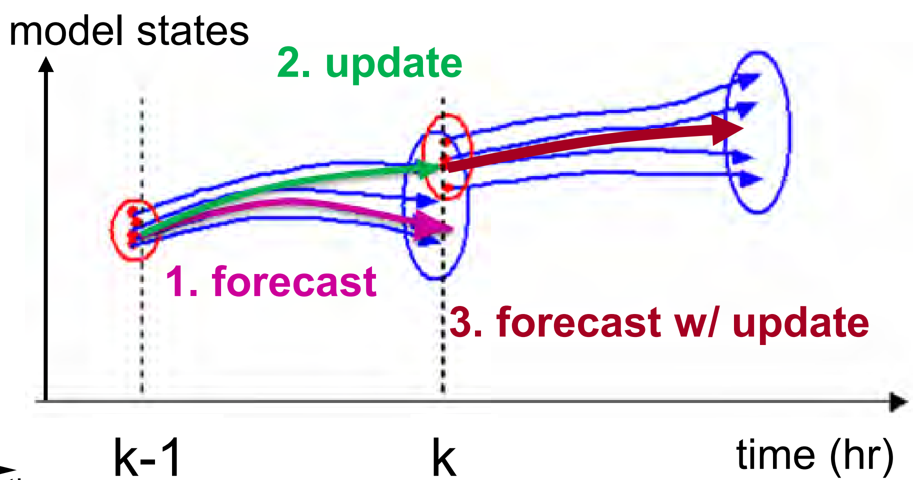
Questions?

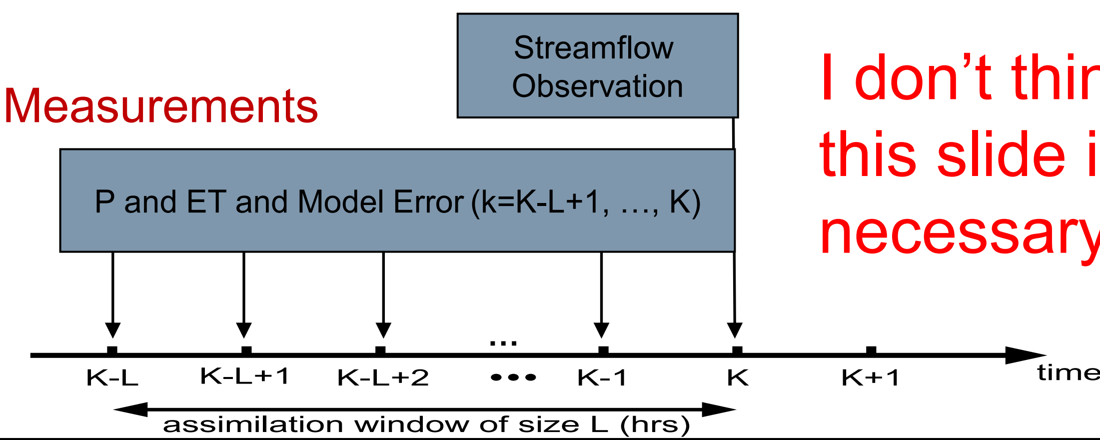
For more info, please contact arezoo.rafieeinasab@mavs.uta.edu.

Formulation of assimilation problem

Lumped SAC - Unit Hydrograph (1-hr timestep)







I don't think this slide is necessary