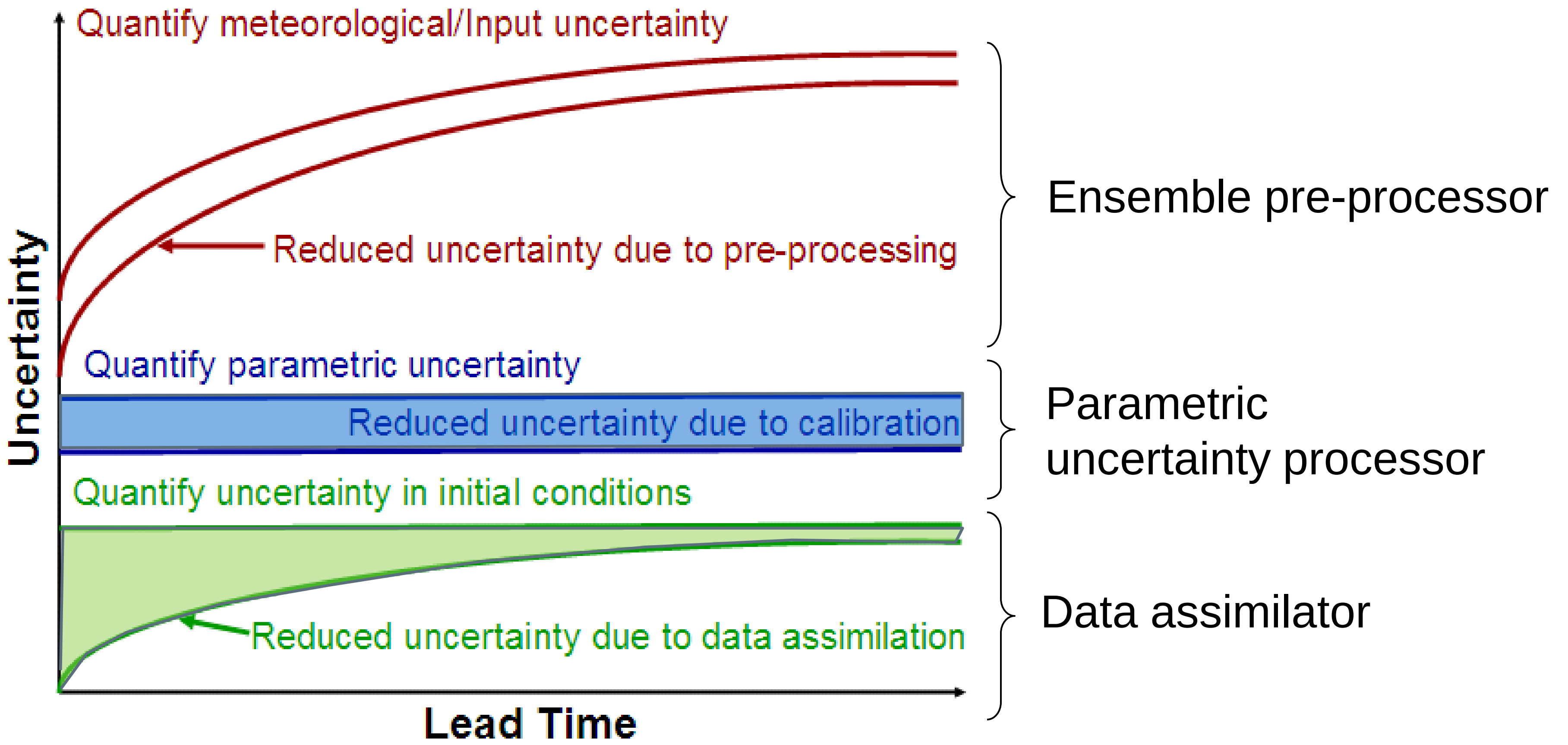
DATA ASSIMILATION IN ENSEMBLE WATER FORECASTING - CHALLENGES AND OPPORTUNITIES

D.-J. Seo[1], Haksu Lee[2,3], Yuqiong Liu[4,5], Victor Koren[2,9], Arezoo Rafieei Nasab[1], Hamideh Riazi[1] Sunghee Kim[1], Changmin Shin[6], Ridwan Siddique[1,8], Yu Zhang[2], Dongsoo Kim[7], Beomgeun Kim[1]

[1]The University of Texas at Arlington, Arlington, TX, USA
[2]NWS Office of Hydrologic Development, Silver Spring, MD, USA
[3]LenTech, Reston, VA, USA
[4]NASA Goddard Space Flight Center, Greenbelt, MD, USA
[5]University of Maryland, College Park, MD, USA
[6]National Institute of Environmental Research, Incheon, Korea
[7]National Climatic Data Center, Asheville, NC, USA
[8]Now at Pennsylvania State University, State College, PA, USA
[9]Retired

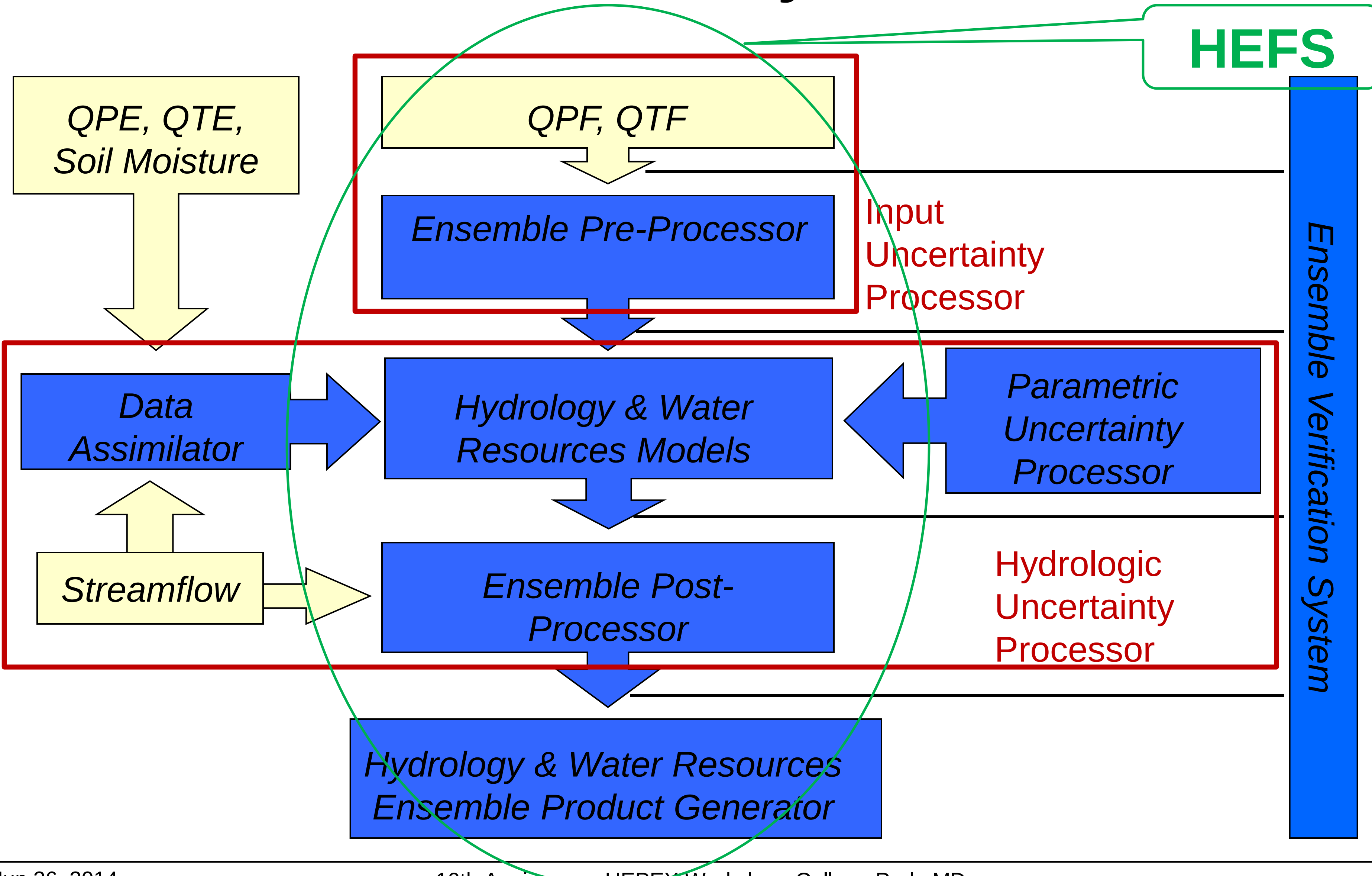
Uncertainties in Hydrologic Forecast



Structural uncertainty residual uncertainty] Ensemble post –processor, multimodel ensemble

Flow regulations: A large challenge

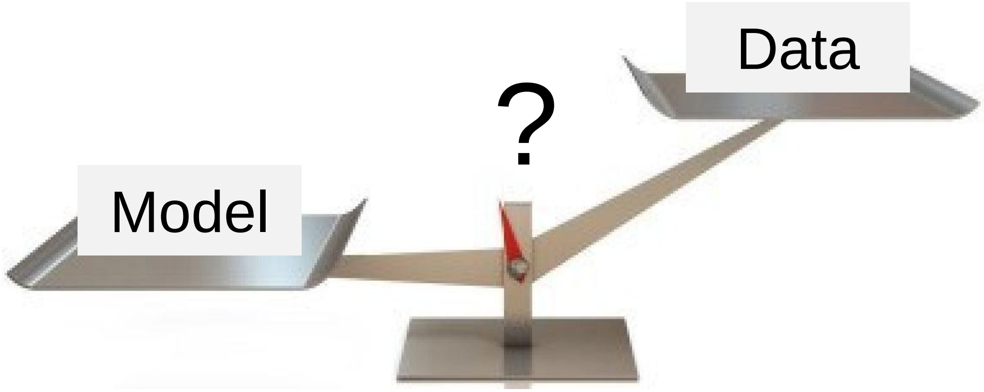
Elements of Hydrologic Ensemble Prediction System



Data assimilation (DA)

- ***"All models are wrong, but some are useful." (George E. P. Box)***
- ***"Models are to be used but not to be believed." (Henry Theil)***

- ***Most observations are useful, but some are wrong.***
- ***Observations are to be believed but not always to be used.***



History of DA in NWS

- Kitanidis, P. K., and R. L. Bras, 1978. Real time forecasting of river flows. Ralph M. Parsons Lab, Dept of Civil Eng, MIT, TR235.
- Sittner, W. T., and K. M. Krouse, 1979. Improvement of Hydrologic simulation by utilizing observed discharge as an indirect input. NOAA Tech. Memo. NWS HYDRO-38, Silver Spring, MD.
- Carroll, 1979. A procedure to incorporate snow course data in the NWSRFS. Proceedings, Modeling of Snow Cover Runoff, Hanover, NH, 351-358.
- Peck, E. L., E. R. Johnson, K. M. Krouse, T. R. Carroll, and J. C. Schaake, Jr, 1980. Hydrological update techniques used by the US National Weather Service, Proceedings of the Oxford Symposium, IAHS Publ. No. 129.
- Day, G., 1990. A methodology for updating a conceptual snow model with snow measurements, NOAA Technical Report NWS 43, Dept of Commerce, Silver Spring, MD.
- Koren, V. and J. Schaake, 1993. Nile Technical Note #147: Updating Algorithm and Program.
- Georgakakos, K. P. and J. A. Sperflage, 1995. Hydrologic Forecast System – HFS: A user's manual. HRC Tech. Note 1, Hydrologic Research Center, San Diego, CA, 17pp.
- McManamon, A., R. K. Hartman, and R. Hills, 1995. Implementation of the Snow Estimation and Updating System (SEUS) in the Clearwater River Basin, Idaho, Proceedings: 63rd Annual Western Snow Conference, Sparks, NV, 56-65.
- Snow Update System user manual, Version 2.00.02 Updated 2003-09-26, Riverside Technology, Inc.

NWSRFS Operations in use (as of October 2007)

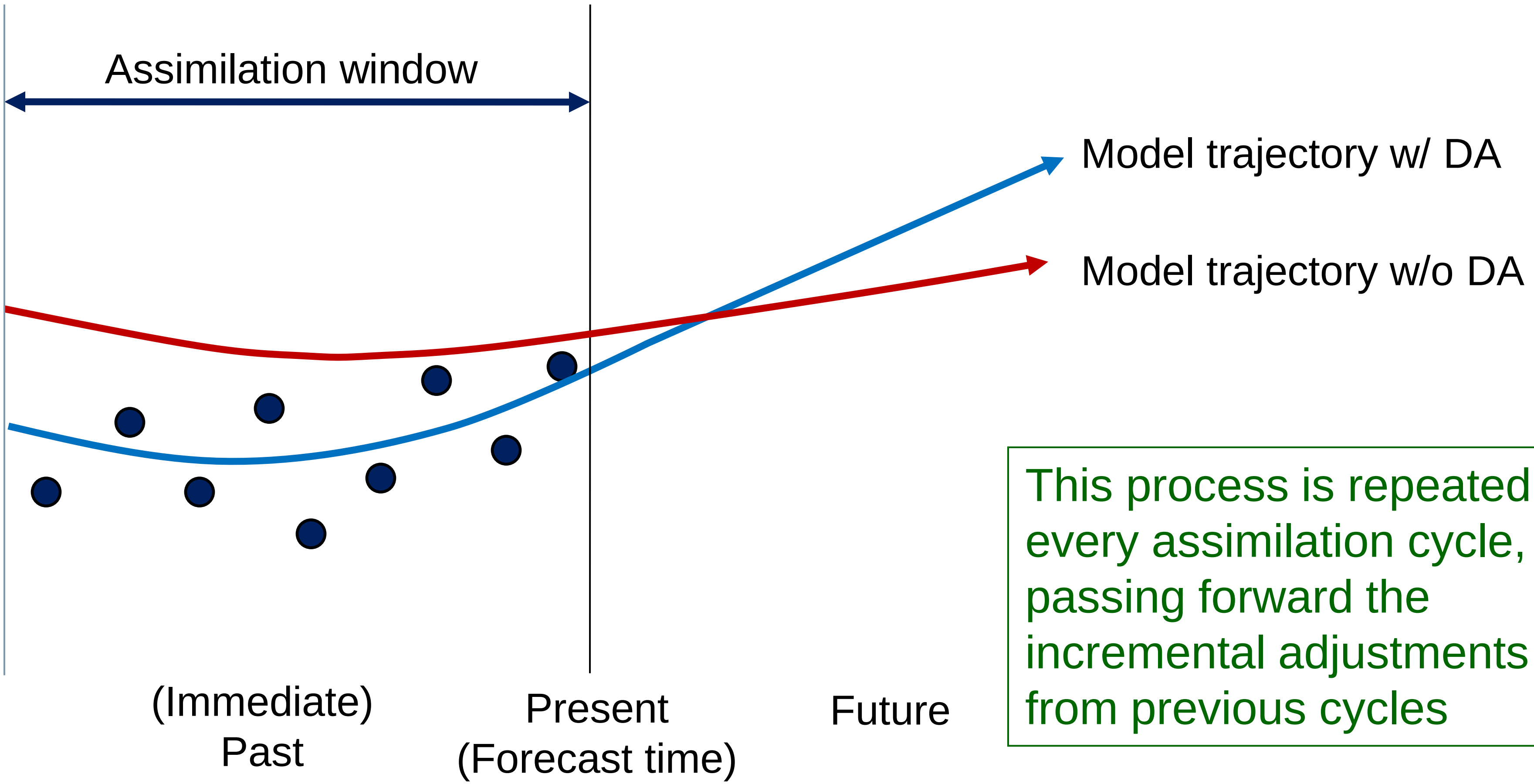
Operation	#RFCs
ADD/SUB	13
ADJUST-H	1
ADJUST-Q	13
ADJUST-T	2
API-CONT	1
API-HFD	1
API-MKC	1
BASEFLOW	4
CHANGE-T	13
CHANLOSS	11
CLEAR-TS	13
CONS_USE	3
DELTA-TS	3
DWOPER	3
FFG	10
FLDWAV	6
GLACIER	1
LAG/K	13
LAY-COEF	1
LIST-FTW	2
LOOKUP	11
LOOKUP3	6
MEAN-Q	13
MERGE-TS	12
MULT/DIV	4
MUSKROUT	3
NOMSNG	9
PLOT-TS	8
PLOT-TUL	13
RES-SNGL	11
RSNWELEV	4
SAC-SMA	12
SARROUTE	1
RES-J	7
SET-TS	9
SNOW-17	10
SSARRESV	1
SS-SAC	1
STAGE-Q	13
STAGEREV	1
TATUM	2
TIDEREV	2
UNIT-HG	13
WEIGH-TS	10

Recent prototype DA development in NWS

- **State updating for lumped hydrologic model**
  - Sacramento, unit hydrograph
  - 2DVAR
    - Seo et al. (2003, 2008)
    - Implemented in the Site Specific Hydrologic Prediction (SSHP) system
  - Maximum likelihood ensemble filter (MLEF, Zupanski 2005)
- **State updating for distributed hydrologic model**
  - gridded Sacramento, kinematic-wave routing
  - 4DVAR
    - Lee et al. (2011, 2012)
- **Parameter updating for hydrologic routing model**
  - 3-parameter Muskingum routing (O'Donnell 1985)
  - 1DVAR
    - Lee et al. (2011)

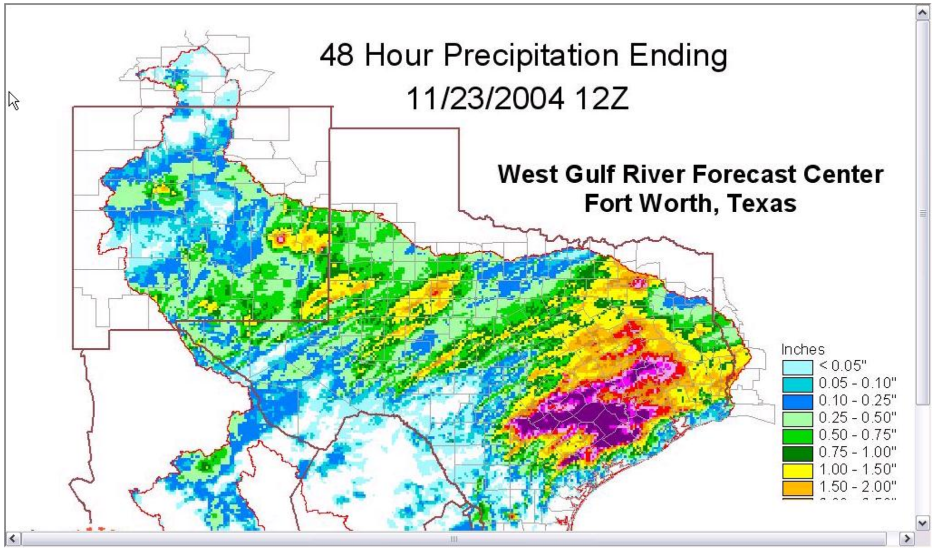
Variational assimilation (VAR)

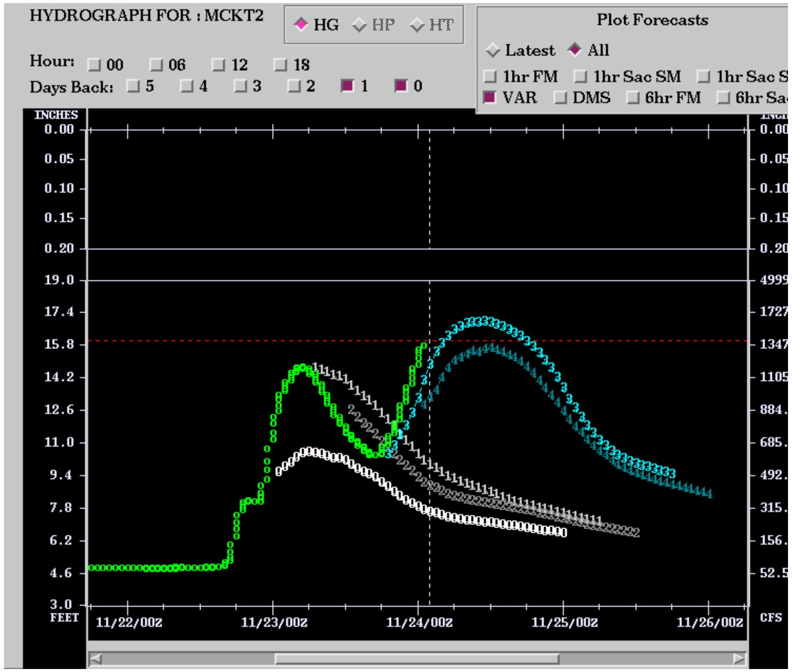
Successive (in time) “batch-by-batch” least-squares curve fitting

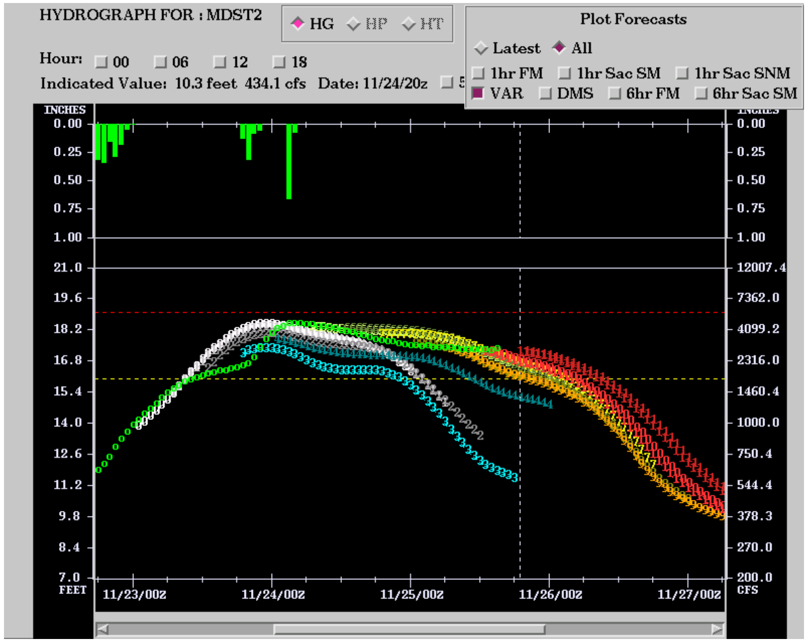


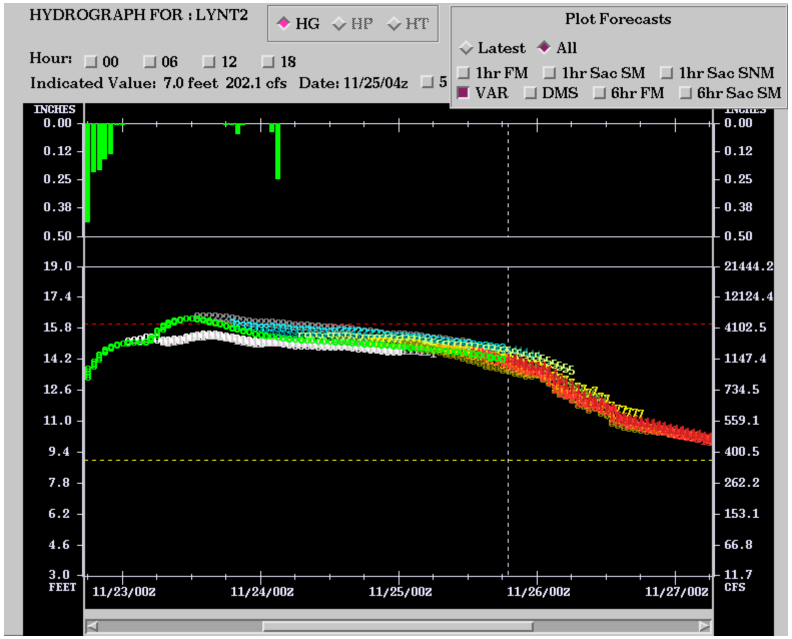
This process is repeated every assimilation cycle, passing forward the incremental adjustments from previous cycles

2DVAR-aided forecast as time-lagged ensembles

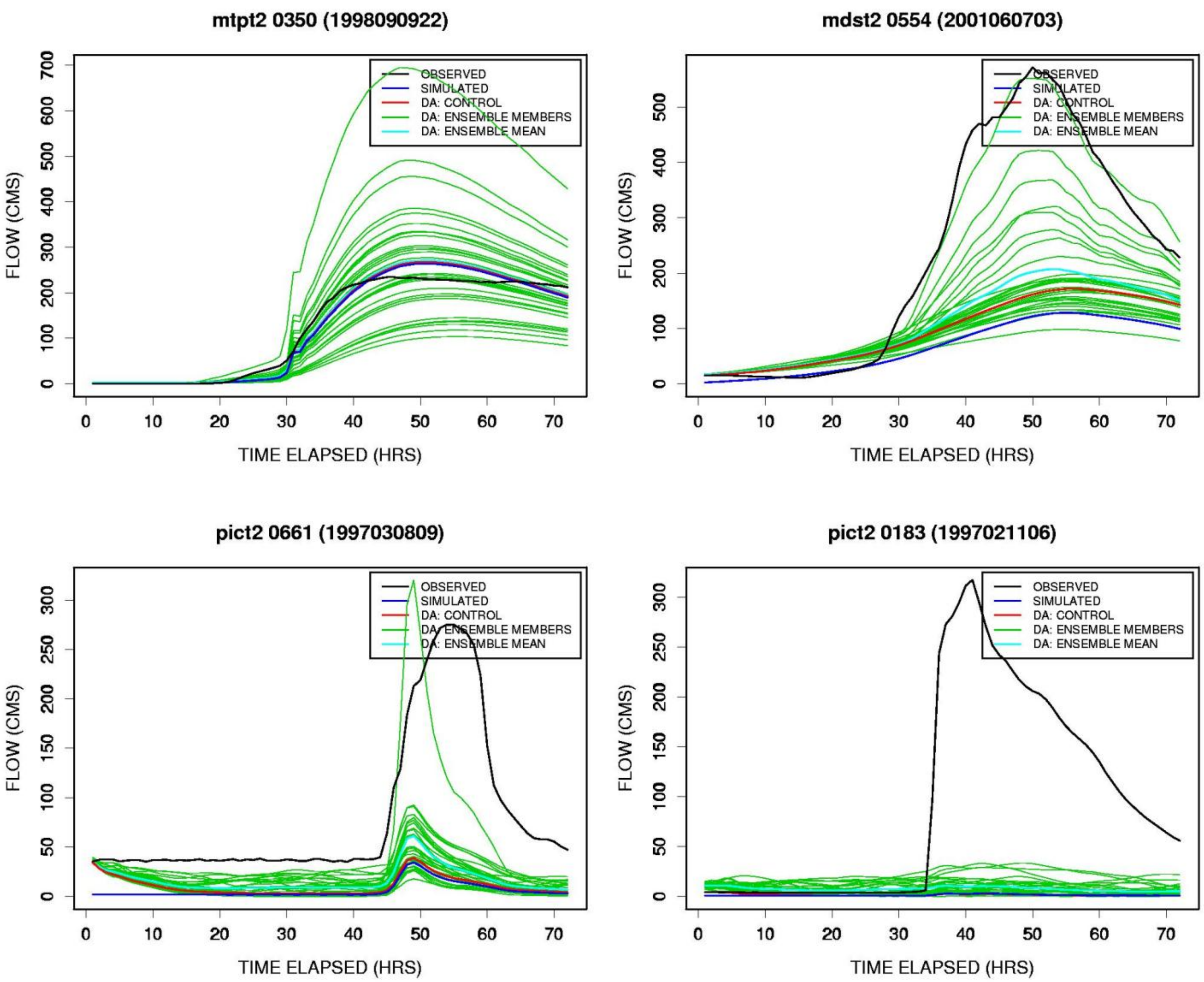






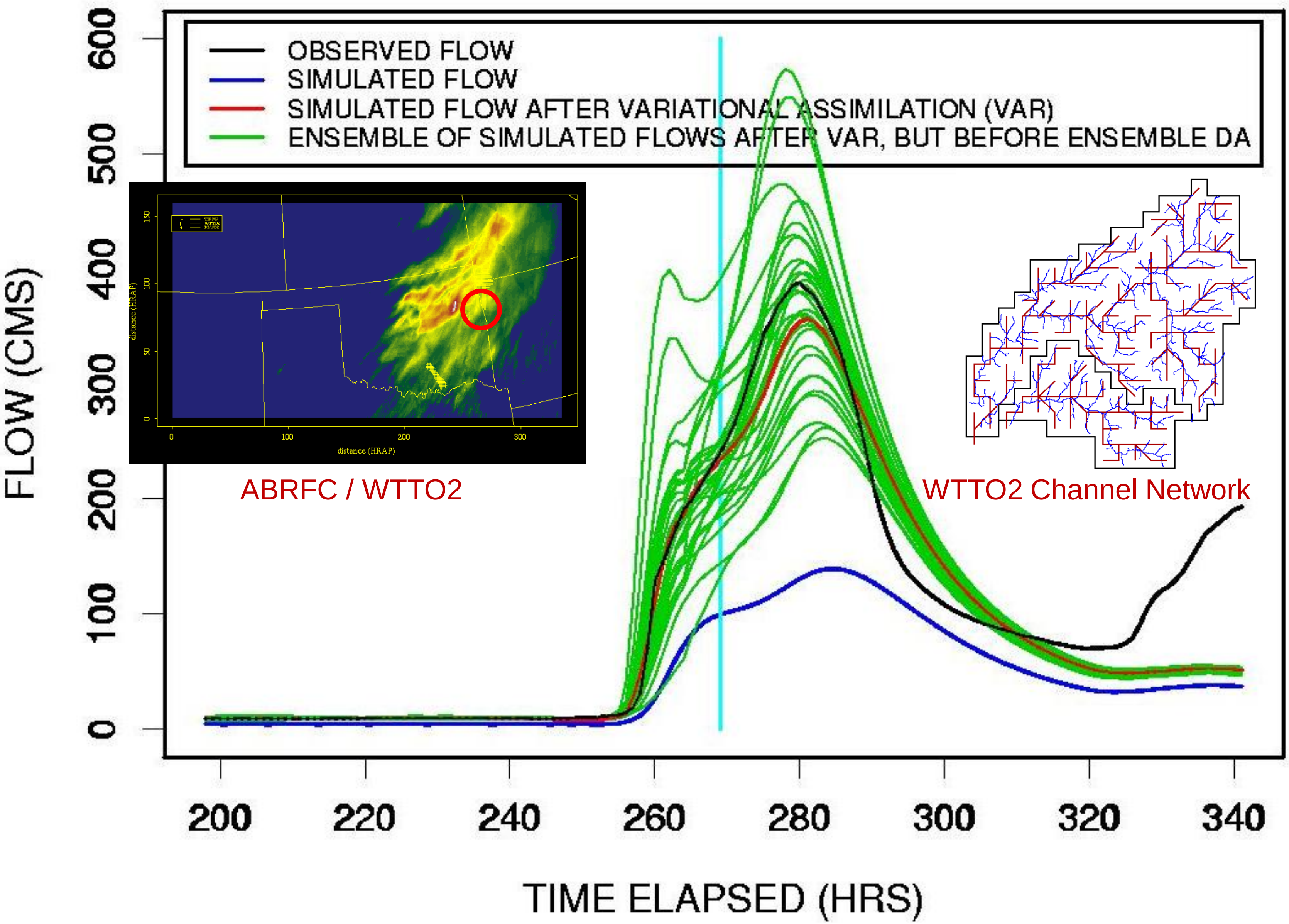


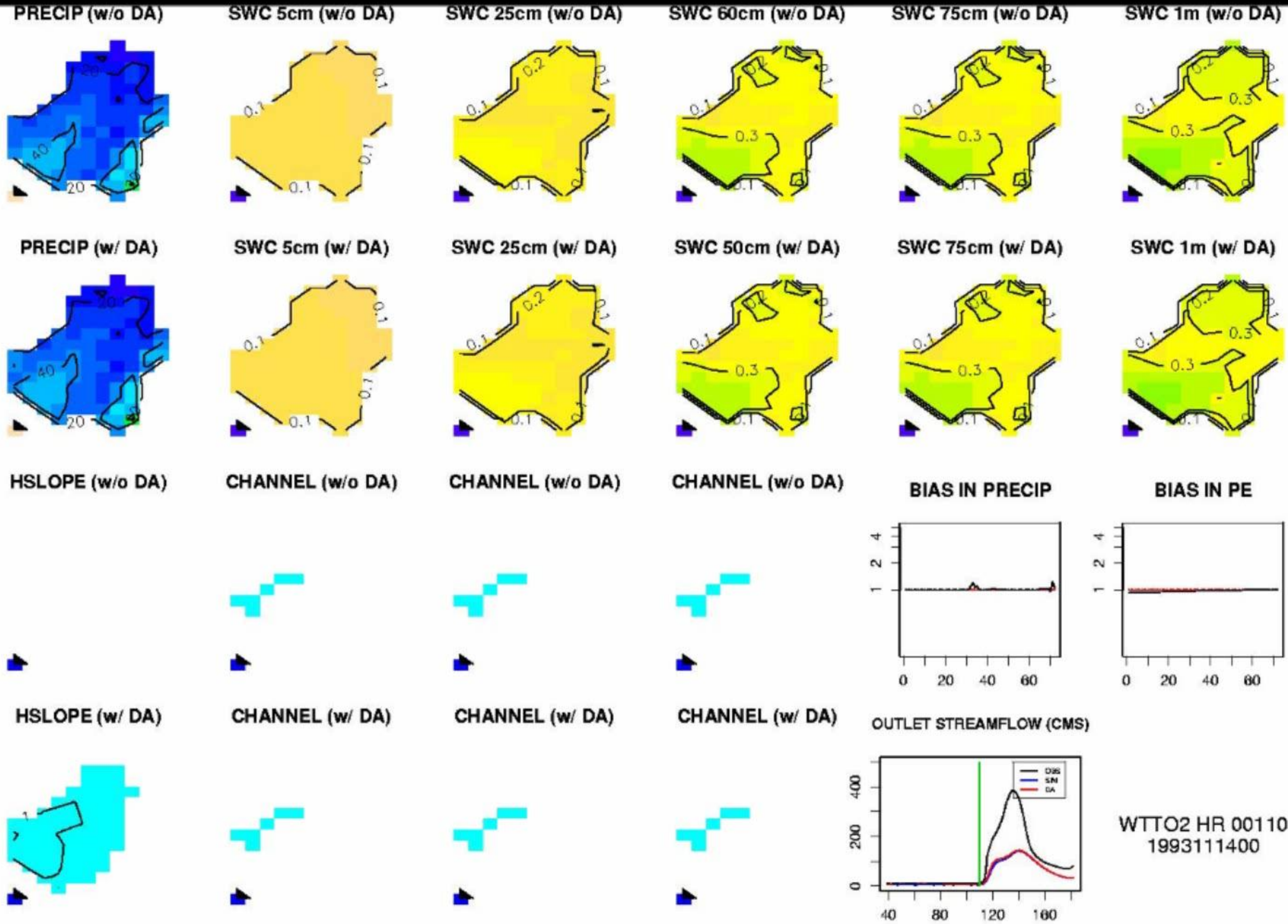
Ensemble filter for lumped SAC-UHG for assimilation of streamflow, precipitation and potential evaporation (PE)



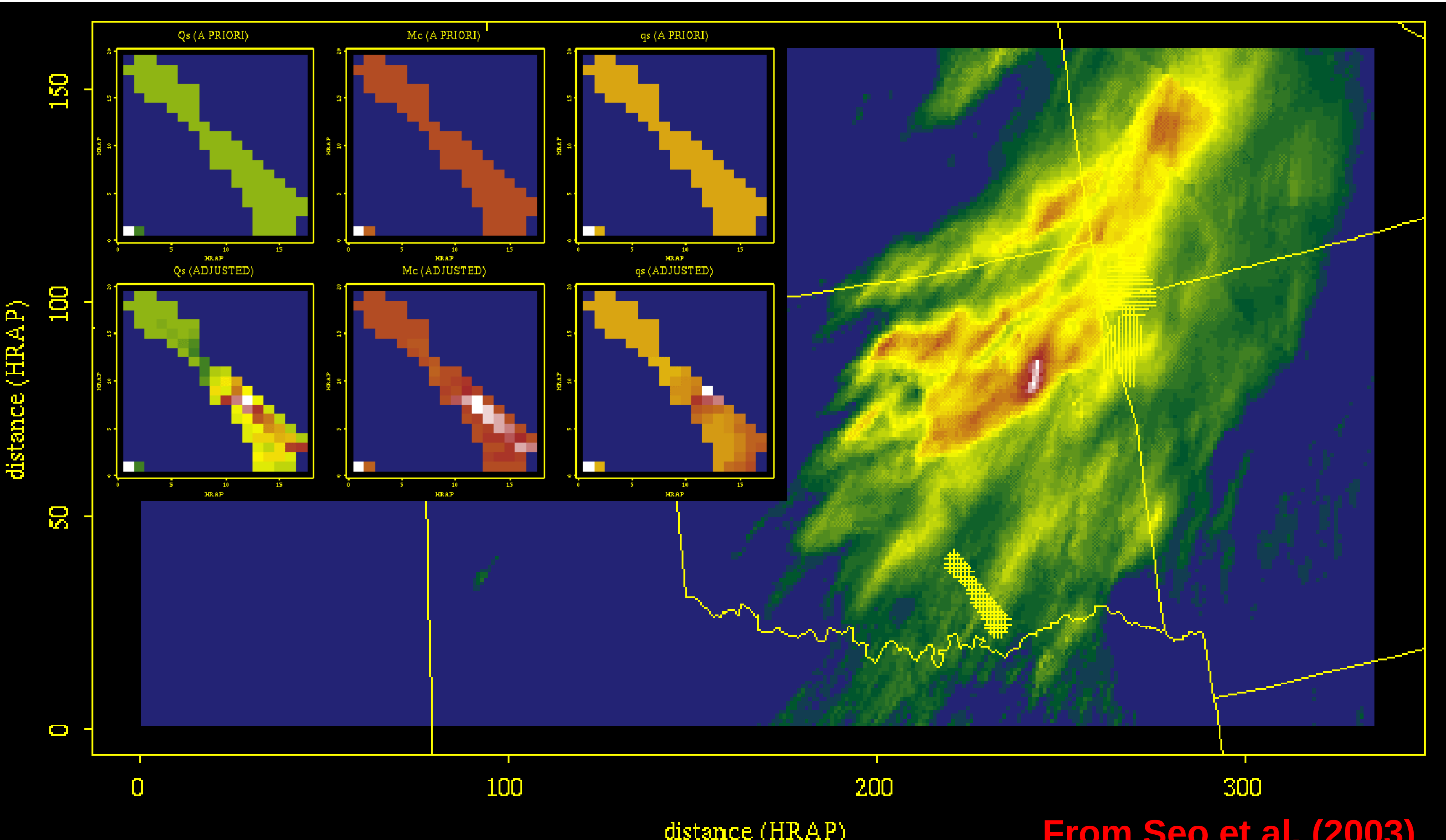
A prototype ensemble filter, a variant of maximum likelihood ensemble filter (Zupanski 2005), for lumped SAC-UHG shows potential (upper plots) and need for improvement (lower plots), including accounting of phase errors and improved error modeling.

ILLUSTRATION OF DATA ASSIMILATION WITH DISTRIBUTED MODEL

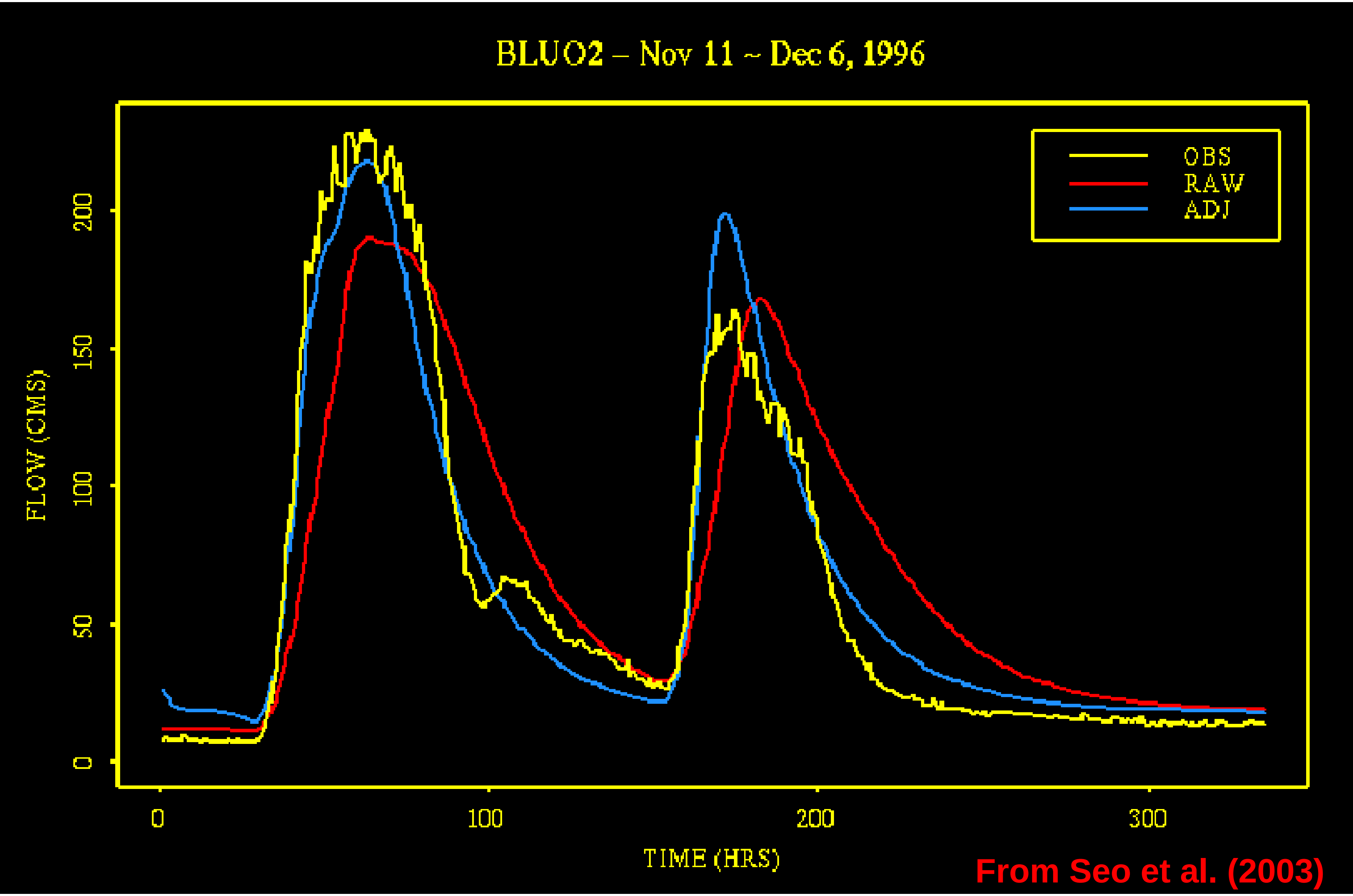



PRECIP (w/o DA)
SWC 5cm (w/o DA)
SWC 25cm (w/o DA)
SWC 60cm (w/o DA)
SWC 75cm (w/o DA)
SWC 1m (w/o DA)
PRECIP (w/ DA)
SWC 5cm (w/ DA)
SWC 25cm (w/ DA)
SWC 50cm (w/ DA)
SWC 75cm (w/ DA)
SWC 1m (w/ DA)
HSLOPE (w/o DA)
CHANNEL (w/o DA)
CHANNEL (w/o DA)
CHANNEL (w/o DA)
BIAS IN PRECIP
BIAS IN PE
HSLOPE (w/ DA)
CHANNEL (w/ DA)
CHANNEL (w/ DA)
CHANNEL (w/ DA)
OUTLET STREAMFLOW (CMS)
OBS
SIM
DA
WTTO2 HR 00110
1993111400

Parameter estimation/optimization of distributed hydrologic routing model

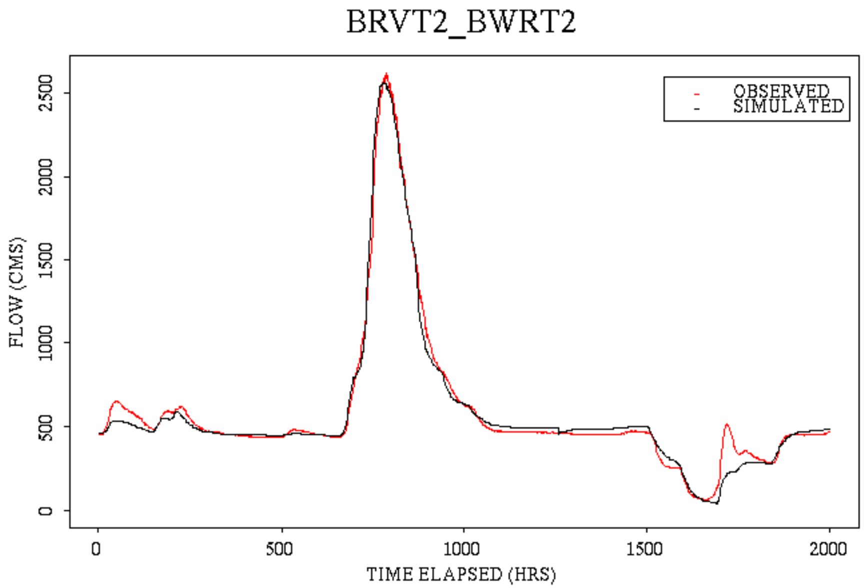


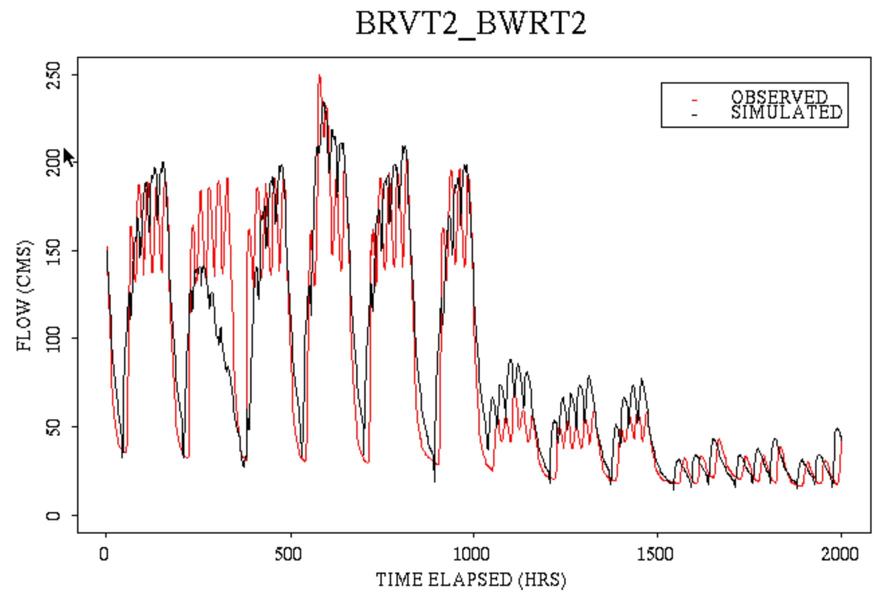
From Seo et al. (2003)

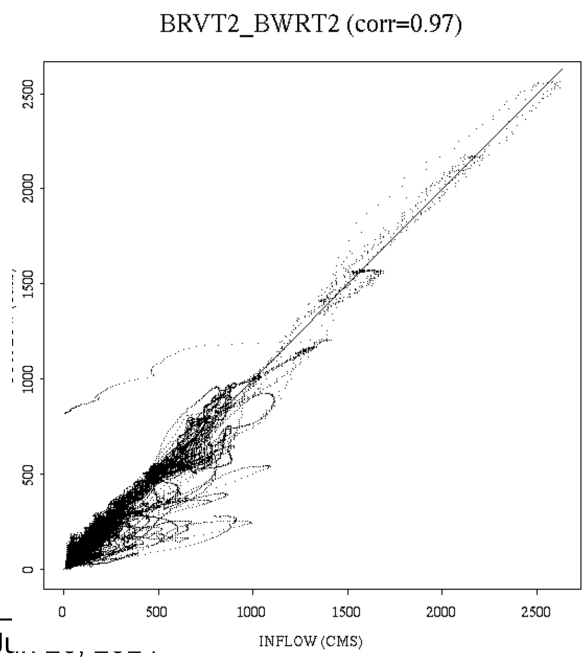


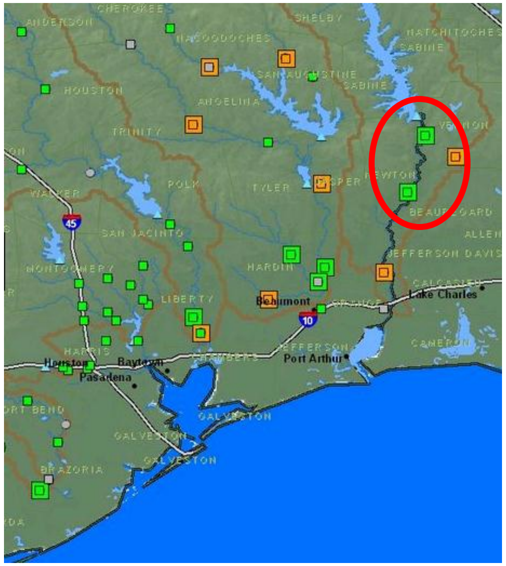
BLUO2 – Nov 11 ~ Dec 6, 1996
OBS
RAW
ADJ
FLOW (CMS)
0
50
100
150
200
0
100
200
300
TIME (HRS)
From Seo et al. (2003)

Toward ensemble DA for hydrologic routing – parameter estimation for variable 3-parameter Muskingum routing (O'Donnell 1985)









Many NWS forecast points require routing of upstream forecasts to the downstream locations. Very often, such routing is subject to flow regulations (upper-right) and sources and sinks that are not very well accounted for (lower-left). Real-time DA can help improve forecast accuracy and quantify uncertainty.

2DVAR - Lessons learned

- Use the same, operational models (soil moisture accounting, snow, routing, etc.)
  - Model physics and parameters must be the same and completely transferable
- Allow forecaster control
  - To reflect any prior or additional information that the forecaster may have
  - Restart (warm or cold) may be necessary if the model deviates from the real world
- Provide, and effectively present, model-dynamical information that explains the DA results
  - Displays of with- and without-DA results over multiple time periods for pattern identification
- Clearly demonstrate the value of DA through objective comparative verification
  - In the context of the end-to-end forecast process
  - Relative to the current operational practice
- Training

Community Hydrologic Prediction System (CHPS)

Flexible, open modeling architecture *linking* program elements

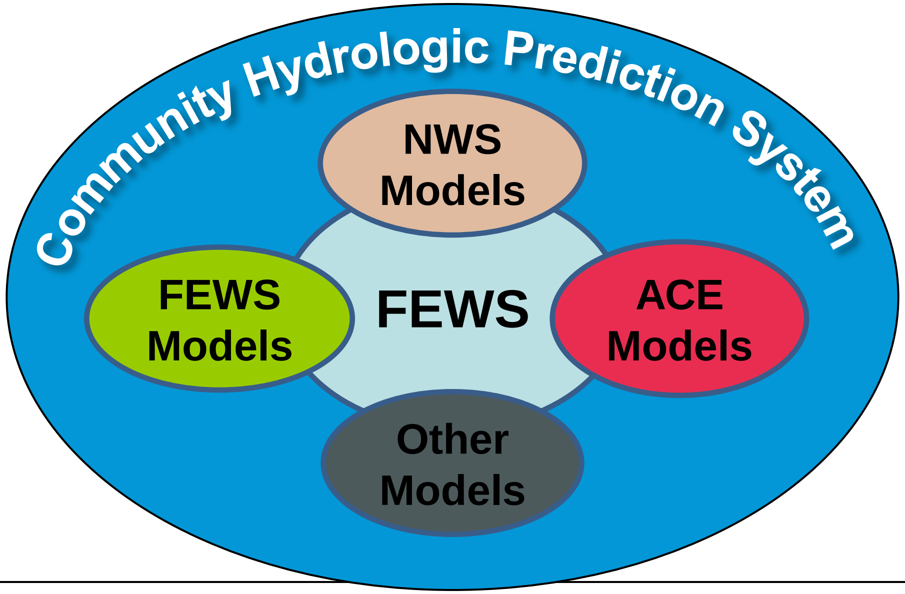
- Modular software to enhance collaboration and accelerate R2O
- Extension of the Flood Early Warning System (FEWS) architecture:
  - Incorporates NWS models with models from FEWS, U.S. Army Corps of Engineers (ACE), and academia

Implementation Status:

✓ AWIPS-II compatible prototype hardware and software for all RFCs

✓ Conducting parallel operations at 4 RFCs, remaining by early 2011

✓ Retire legacy system in early 2012

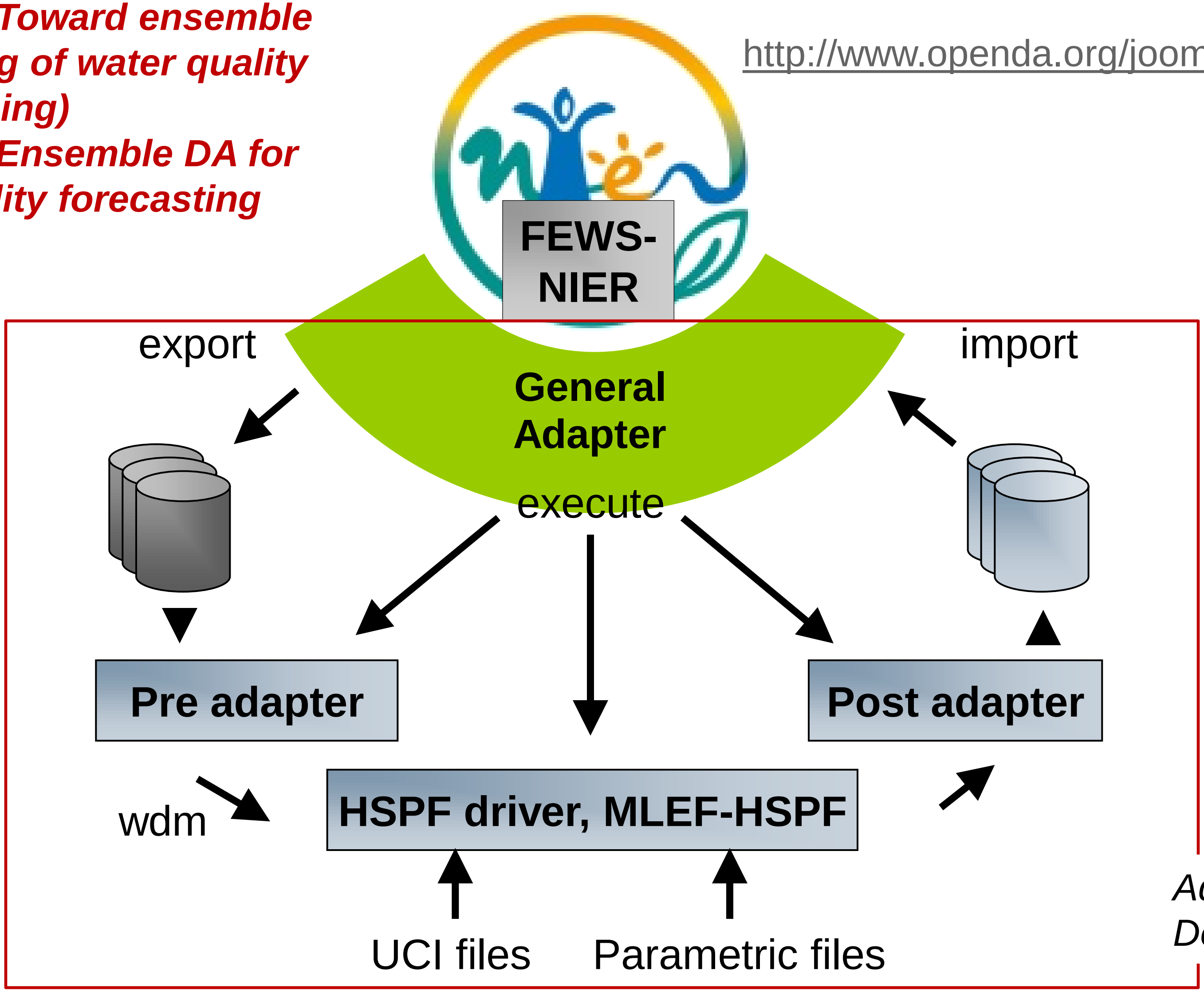


From Carter (2010)

Implementation of MLEF-HSPF in FEWS-NIER as a new model

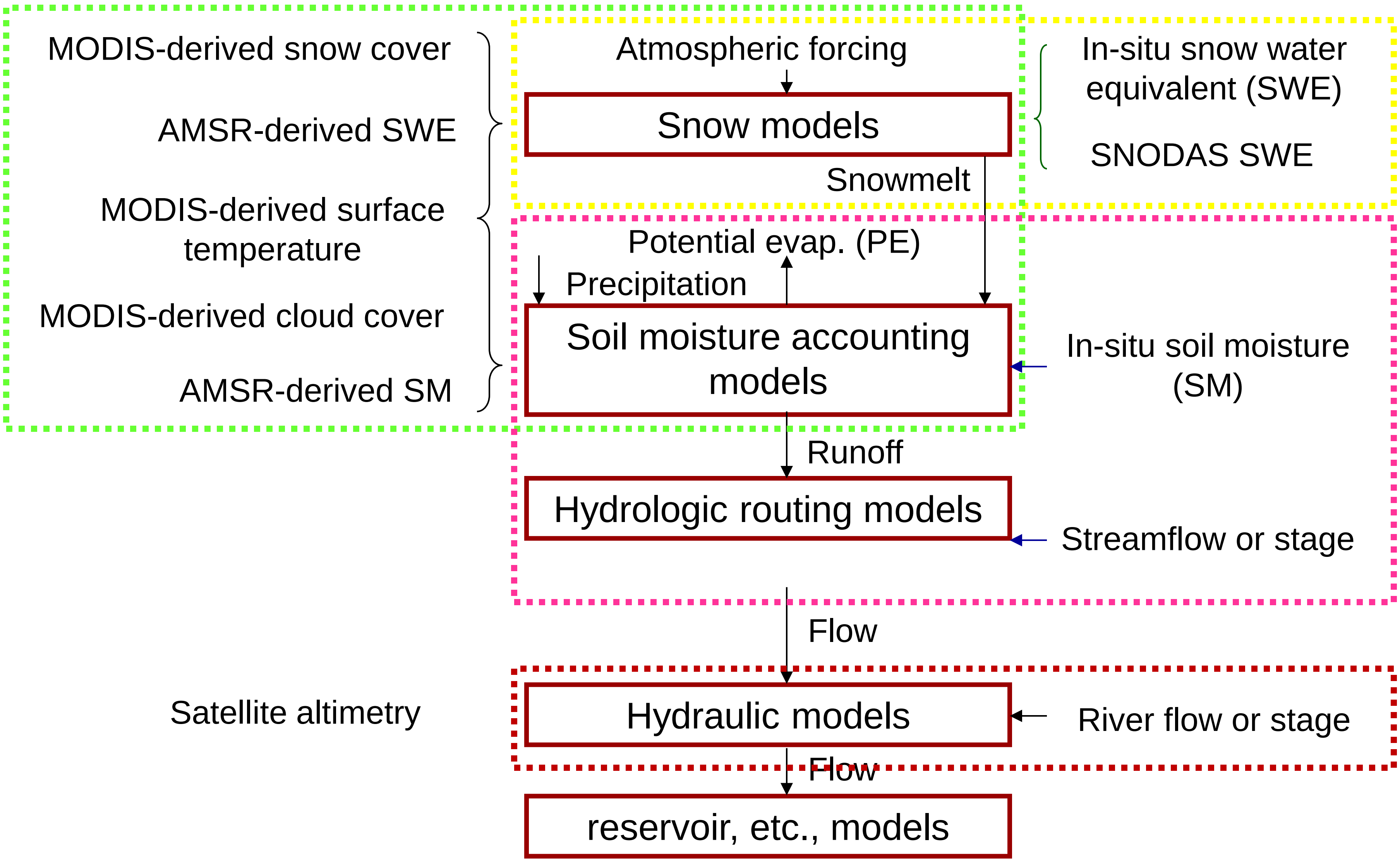
- ***Kim et al. Toward ensemble forecasting of water quality (this morning)***
- ***Kim et al. Ensemble DA for water quality forecasting (poster)***

http://www.openda.org/joomla/index.php



Adapted from Deltares (2012)

Operational hydrologic data assimilation - Strategy



Adapted from OHD Strategic Science Plan 2010

DA strategy for operational hydrologic forecasting

- Decompose $Z = f(X, V)$

- To illustrate, decompose:

$$\begin{bmatrix} Z_1 \\ Z_2 \\ . \\ . \\ Z_n \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & . & . & H_{1m} \\ H_{21} & H_{22} & . & . & H_{2m} \\ . & & & & \\ . & & & & \\ H_{n1} & H_{n2} & . & . & H_{nm} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ X_m \end{bmatrix} + \begin{bmatrix} V_1 \\ V_2 \\ . \\ . \\ V_n \end{bmatrix}$$

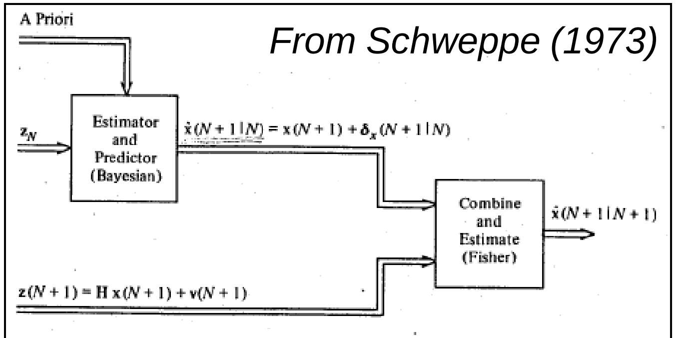


Figure 6.2.1 Basis of argument.

From Schweppe (1973)

into smaller ones such that:

- the suboptimal solutions from the decomposed problems are close to the optimal solution for the full-blown problem
- the resulting DA process is forecaster-controllable

Questions

- Uncertain error statistics
  - Nonlinear, heteroscedastic, flow- and scale-dependent
- Underdetermined systems
  - Paucity of observations
  - Rank deficiency a large issue
- Nonlinear observations
  - Streamflow for soil moisture
- Minimization criteria
  - Need for DA is for out-of-the-ordinary/extreme events
    - Climate change, urbanization

The "PQR" problem (from the 1st HEPEX DA Workshop in Delft, Nov, 2010)

$$\hat{\mathbf{x}}(N+1|N+1) = \boldsymbol{\Sigma}(N+1|N+1) \times \{\mathbf{H}'(N+1)\mathbf{R}^{-1}(N+1)\mathbf{z}(N+1) + \boldsymbol{\Sigma}^{-1}(N+1|N)\underline{\boldsymbol{\Phi}(N)\hat{\mathbf{x}}(N|N)}\}$$

$$\boldsymbol{\Sigma}(N+1|N) = \boldsymbol{\Phi}(N)\boldsymbol{\Sigma}(N|N)\boldsymbol{\Phi}'(N) + \mathbf{G}(N)\mathbf{Q}(N)\mathbf{G}'(N)$$

$$\boldsymbol{\Sigma}(N+1|N+1) = [\mathbf{H}'(N+1)\mathbf{R}^{-1}(N+1)\mathbf{H}(N+1) + \boldsymbol{\Sigma}^{-1}(N+1|N)]^{-1}$$

$$\boldsymbol{\Sigma}(0|0) = \boldsymbol{\psi}$$

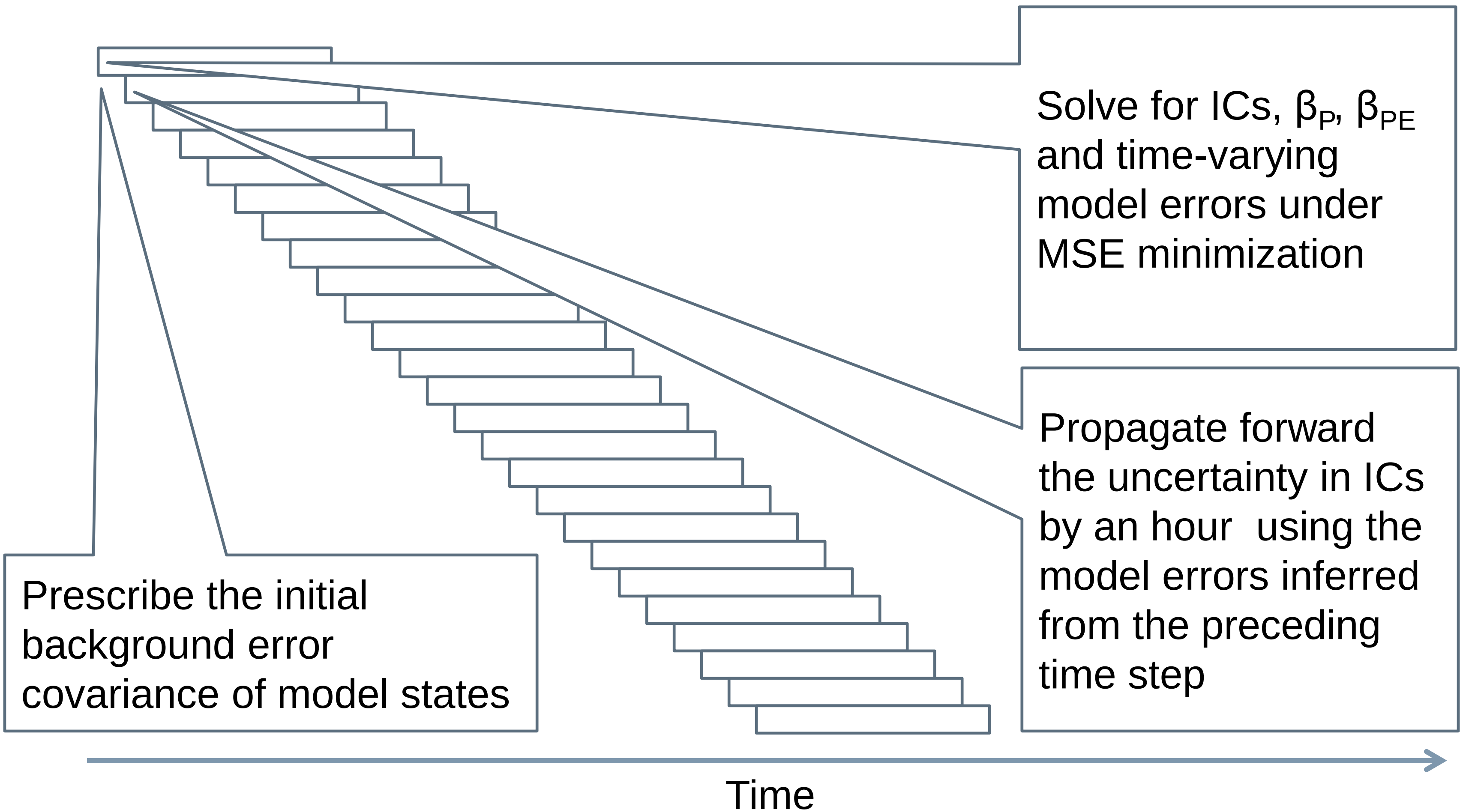
$$\hat{\mathbf{x}}(0|0) = \mathbf{0}$$

From Schweppe (1973)

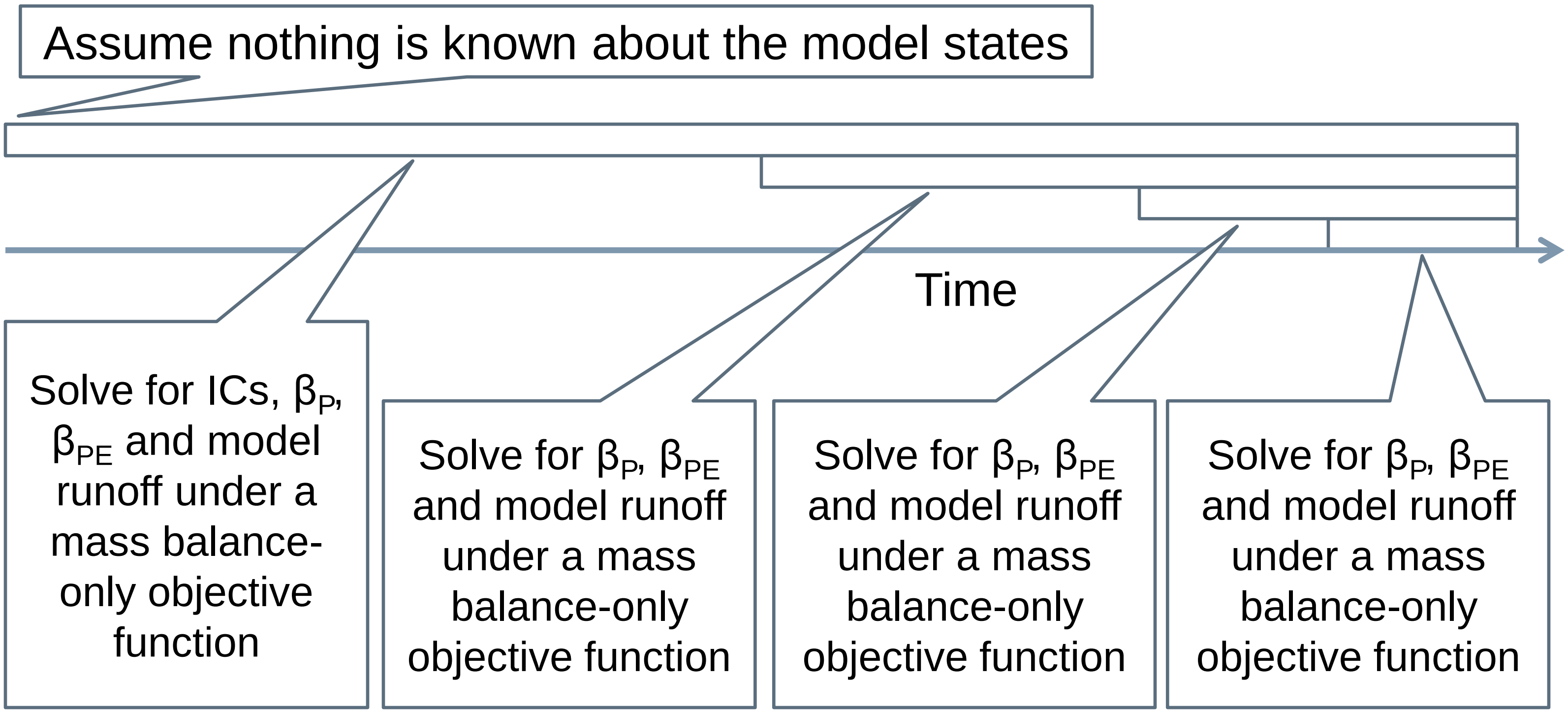
Proposed approach

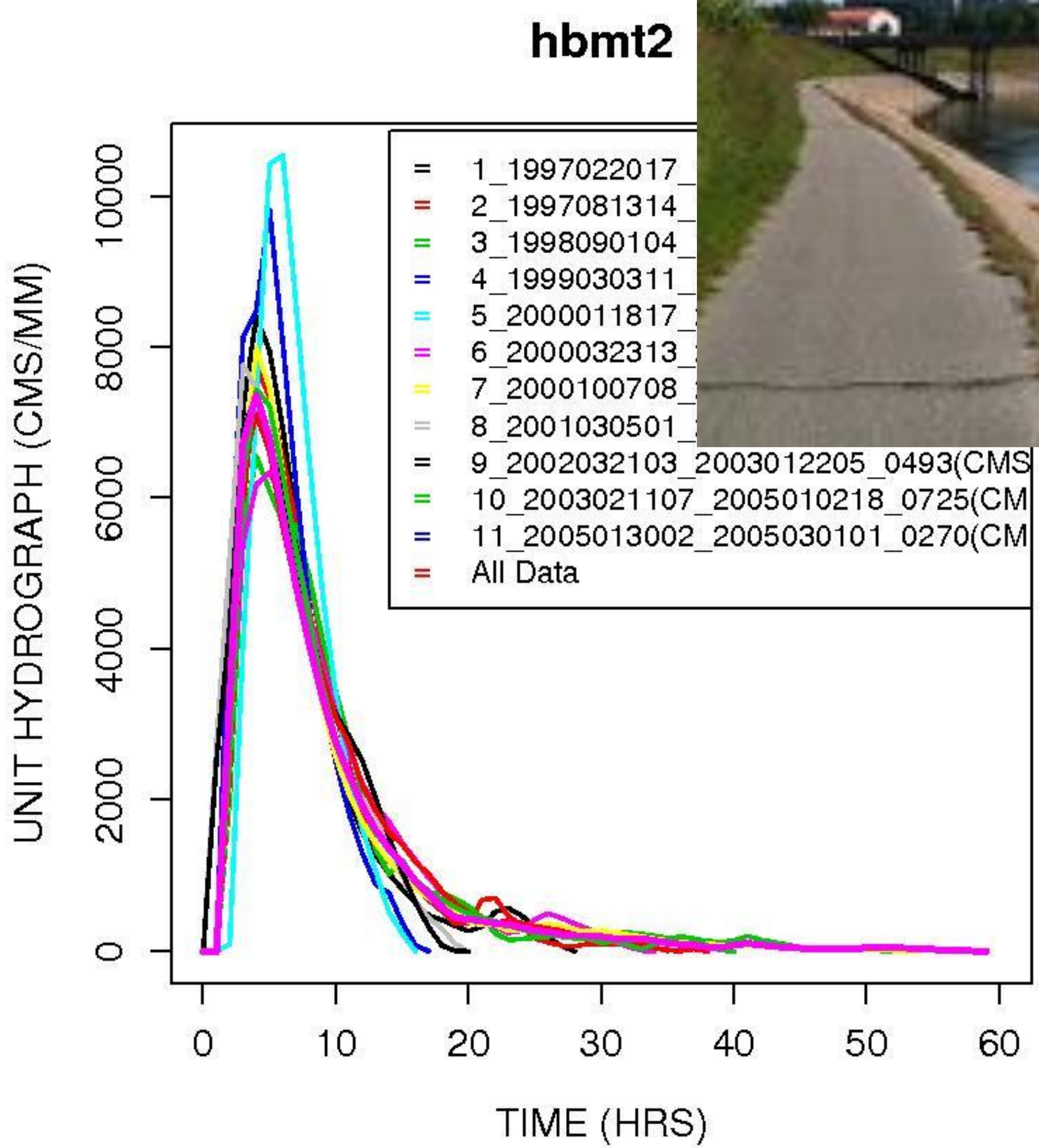
- Multi-Scale Bias Correction (MSBC)
  - The hydrologic processes (and hence model errors) are multiscale in nature due to different residence times at work.
  - Due to paucity of hydrologic observations, the DA problems are likely to be underdetermined.
- Adaptive Error Modeling (AEM)
  - Rather than modeling process-specific errors in soil moisture and routing dynamics, model the aggregate errors in runoff simulation based on observed streamflow for
    - parsimony
    - adaptive accounting of heteroscedasticity and timing errors.

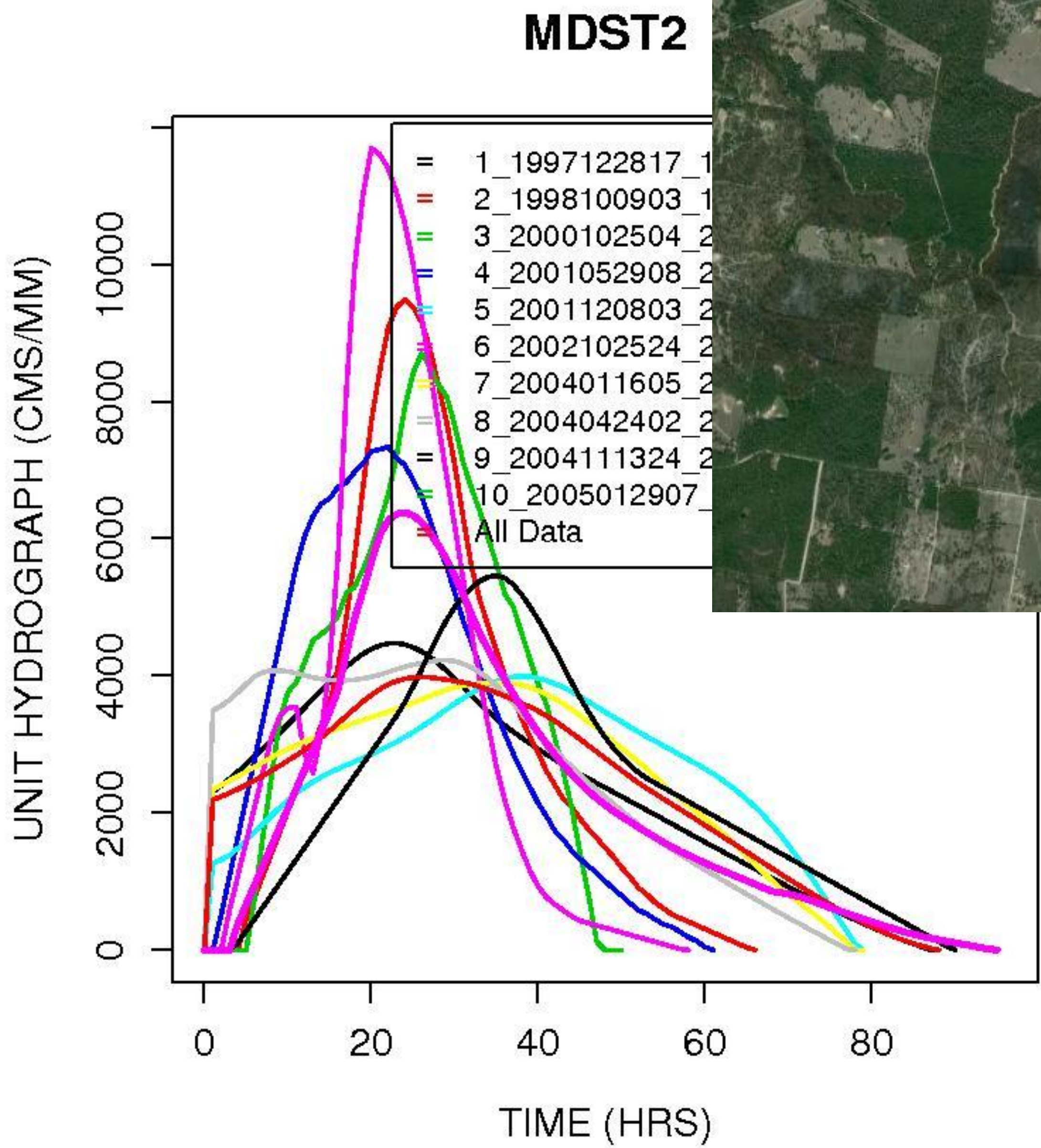
2DVAR with forward propagation of IC

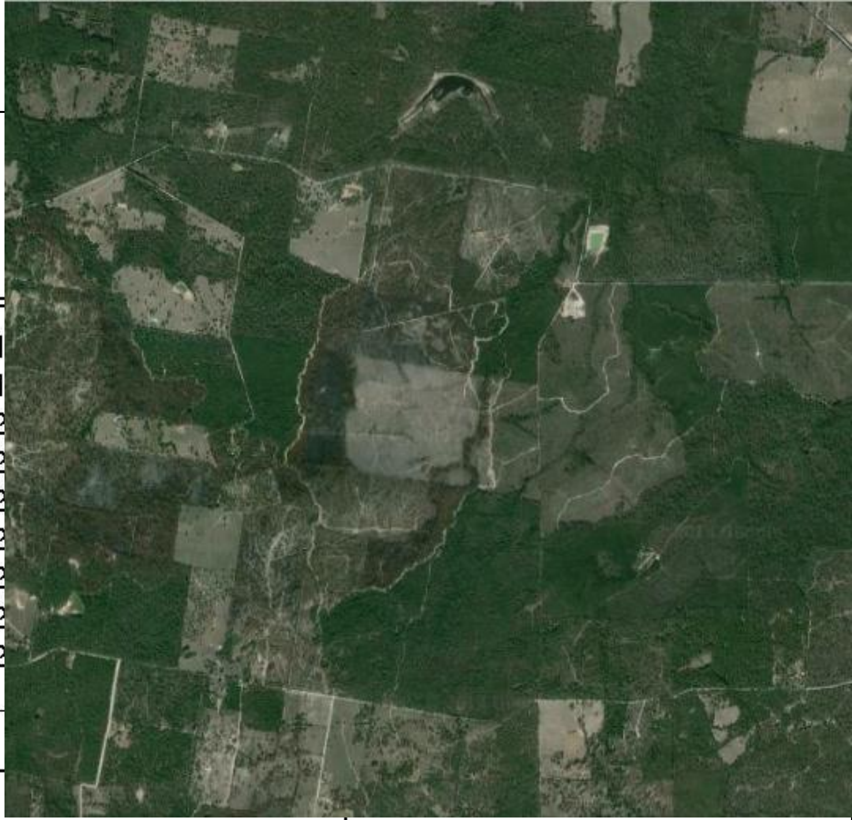


Multi-Scale Bias Correction (MSBC)

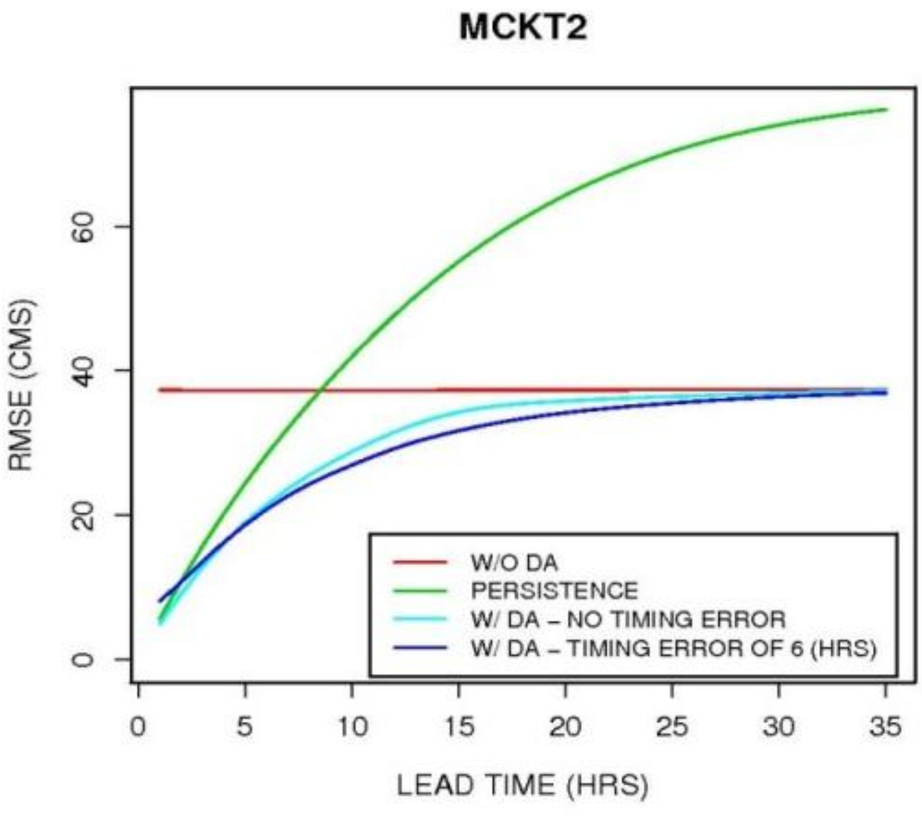


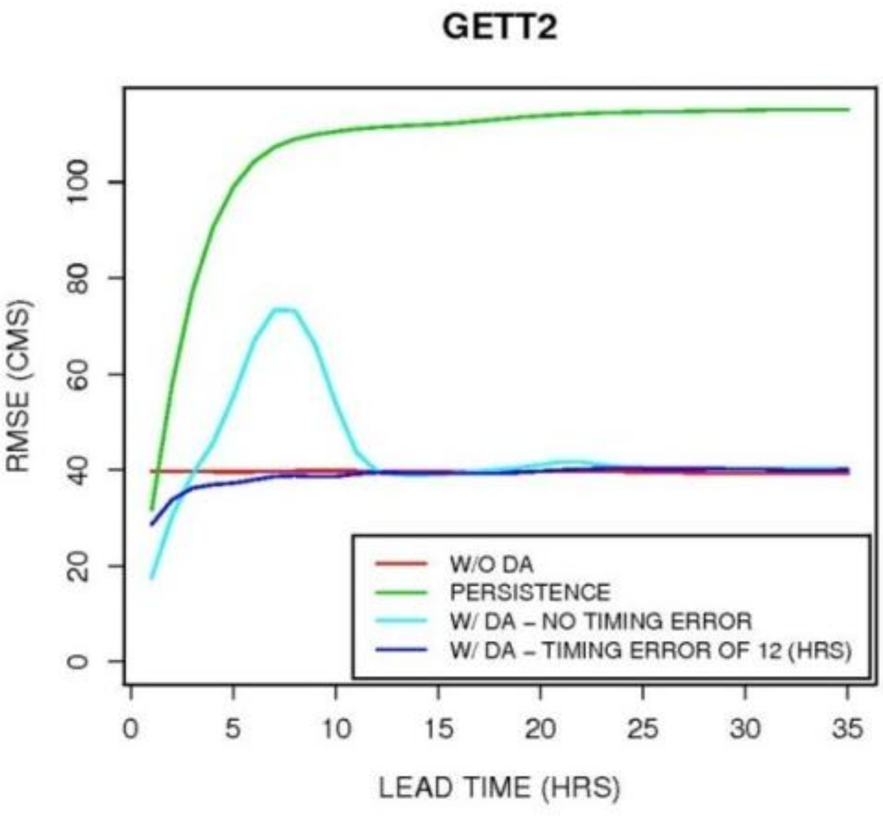

hbmt2
UNIT HYDROGRAPH (CMS/MM)
10000
8000
6000
4000
2000
0
1_1997022017_
2_1997081314_
3_1998090104_
4_1999030311_
5_2000011817_
6_2000032313_
7_2000100708_
8_2001030501_
9_2002032103_2003012205_0493(CMS
10_2003021107_2005010218_0725(CM
11_2005013002_2005030101_0270(CM
All Data
0
10
20
30
40
50
60
TIME (HRS)


MDST2
1_1997122817_1
2_1998100903_1
3_2000102504_2
4_2001052908_2
5_2001120803_2
6_2002102524_2
7_2004011605_2
8_2004042402_2
9_2004111324_2
10_2005012907_
All Data
UNIT HYDROGRAPH (CMS/MM)
0
2000
4000
6000
8000
10000
0
20
40
60
80
TIME (HRS)



Effect of timing errors in updating of soil water states

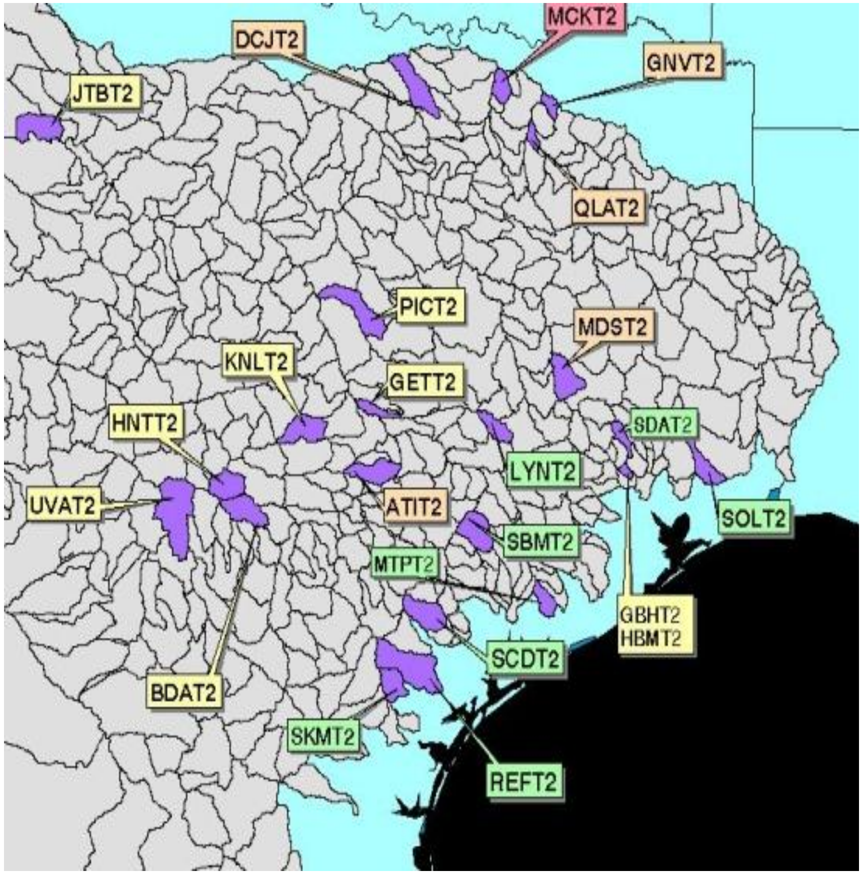




T_p=13 (hrs)

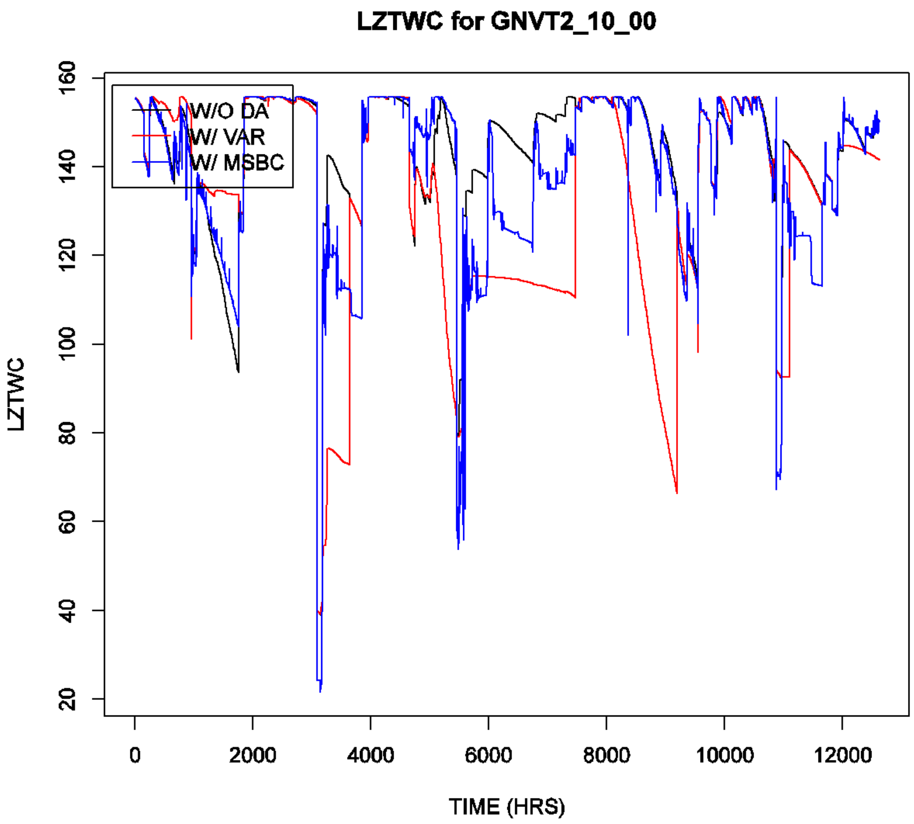
T_p=10 (hrs)

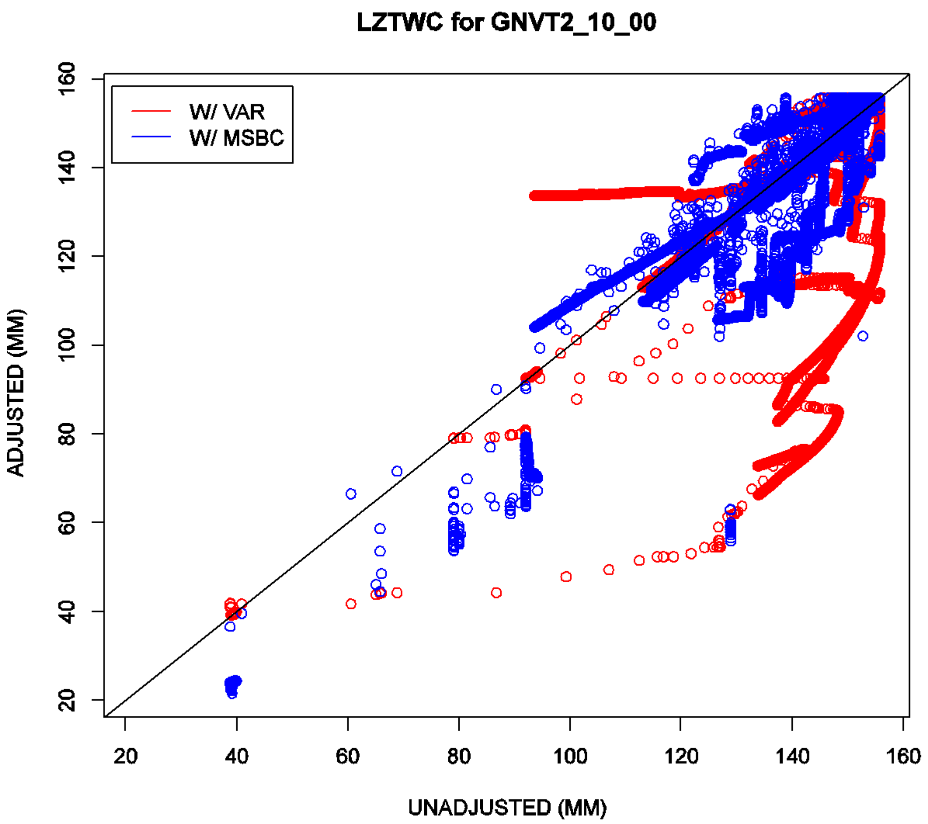
MSBC vs. 2DVAR



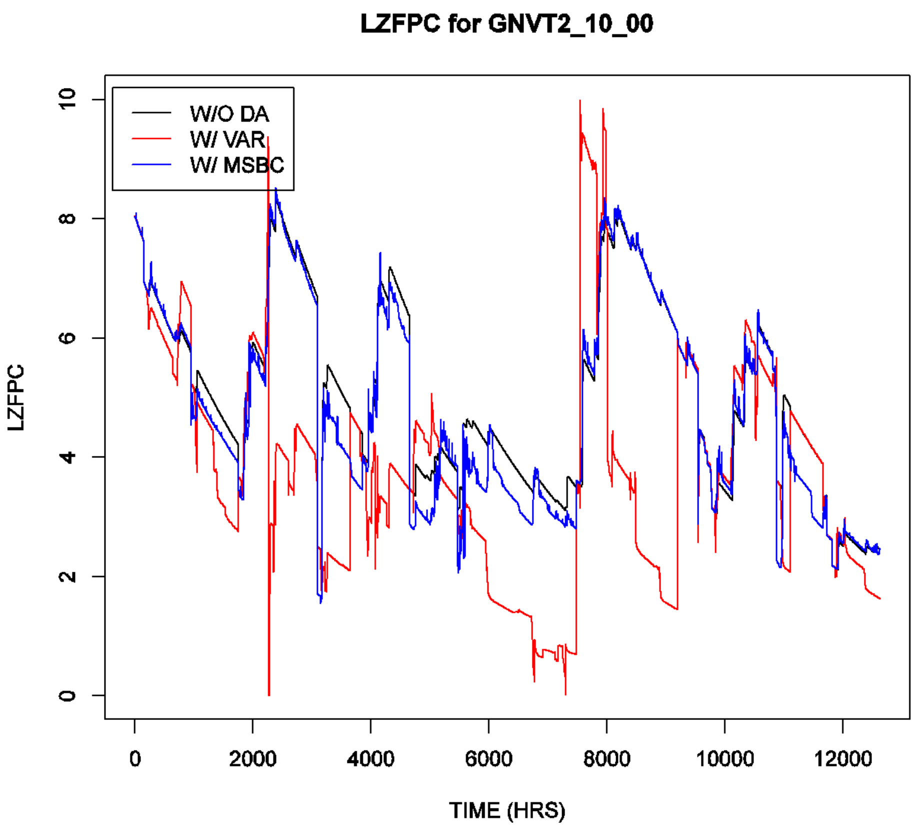
2DVAR better	3 basins
MSBC better	10 basins
Comparable	10 basins

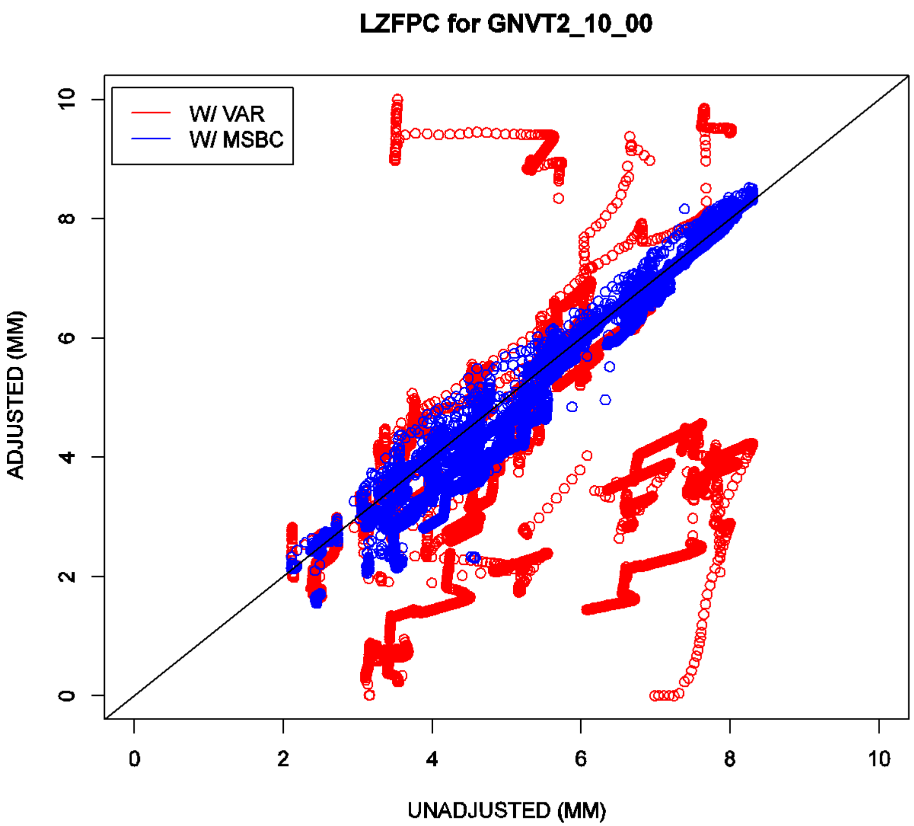
MSBC-updated model states stay much closer to the base (i.e. un-updated) model states





MSBC-updated model states stay much closer to the base model states (cont.)





Dealing with nonlinear observations

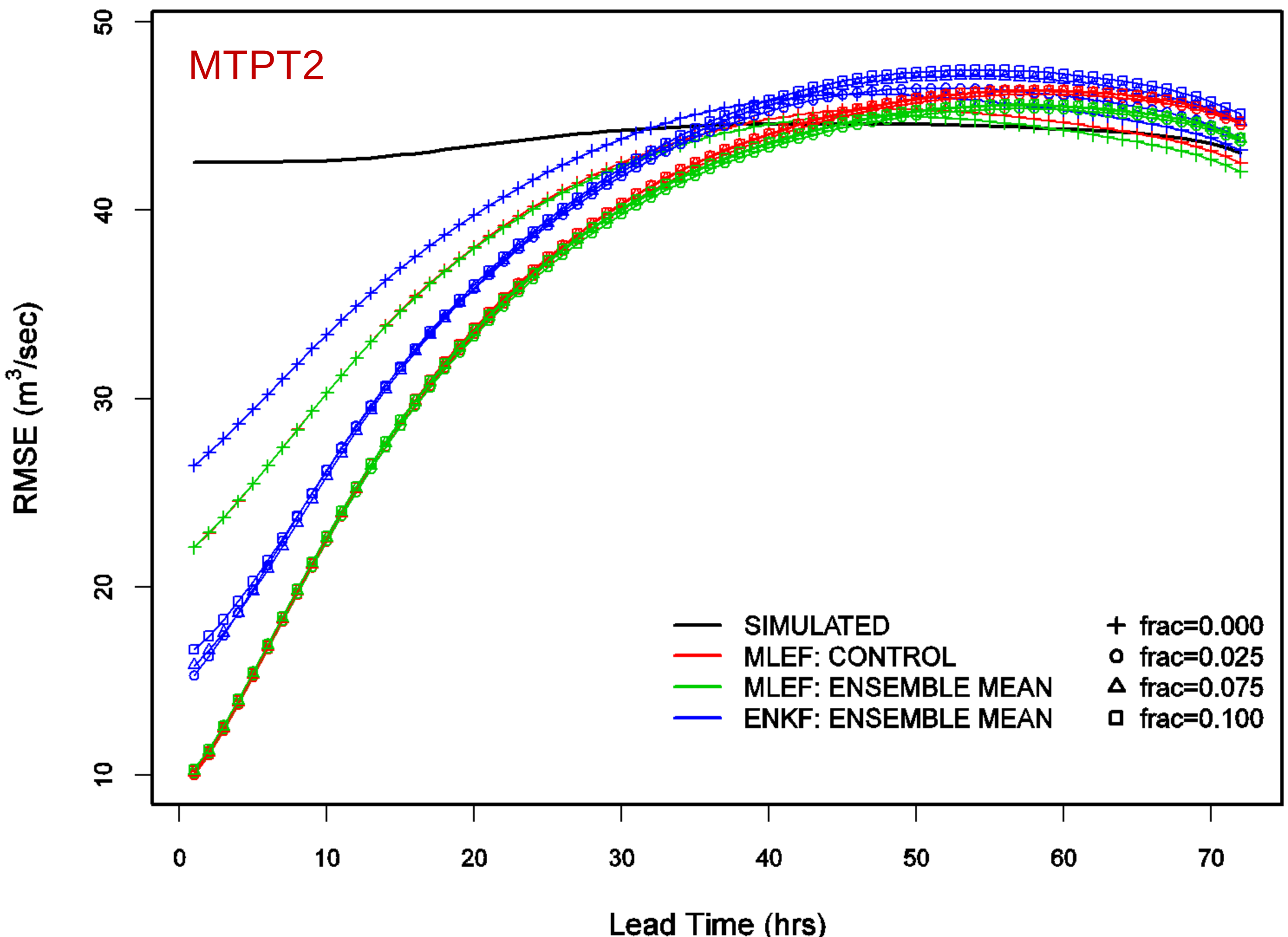
$$\mathbf{x}(n+1) = \mathbf{\Phi}(n)\mathbf{x}(n) + \mathbf{G}(n)\mathbf{w}(n)$$

$$\mathbf{z}(n) = \mathbf{H}(n)\mathbf{x}(n) + \mathbf{v}(n), \qquad n = 1, \ldots$$

$\mathbf{x}(n)$	state, a K_1 vector
$\mathbf{z}(n)$	observation, a K_2 vector
$\mathbf{v}(n)$	white observation uncertainty, a K_2 vector
$\mathbf{w}(n)$	white system driving uncertainty, a K_3 vector
$\mathbf{x}(0)$	initial condition which may be uncertain
n	time

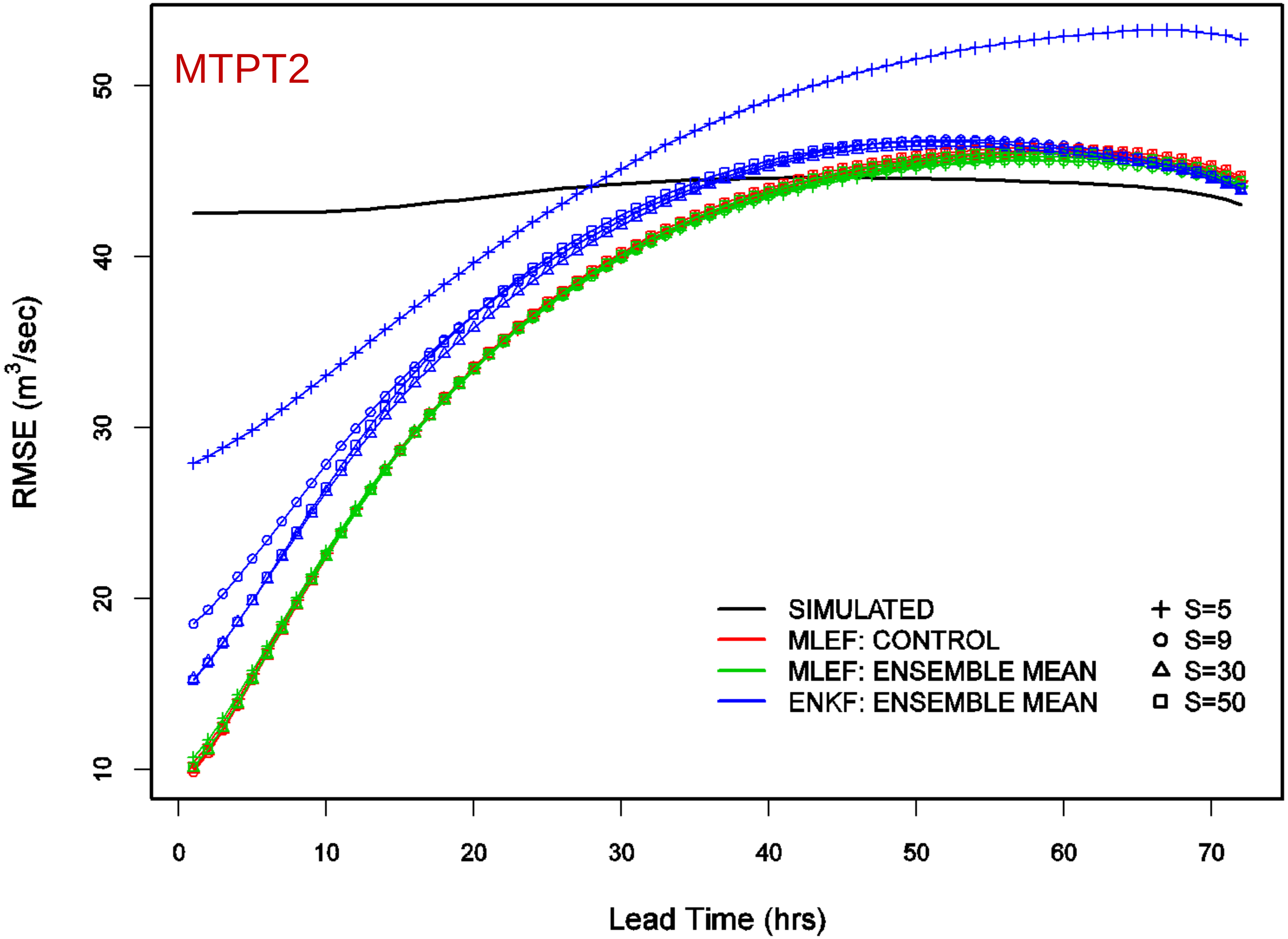
From Schweppe (1973)

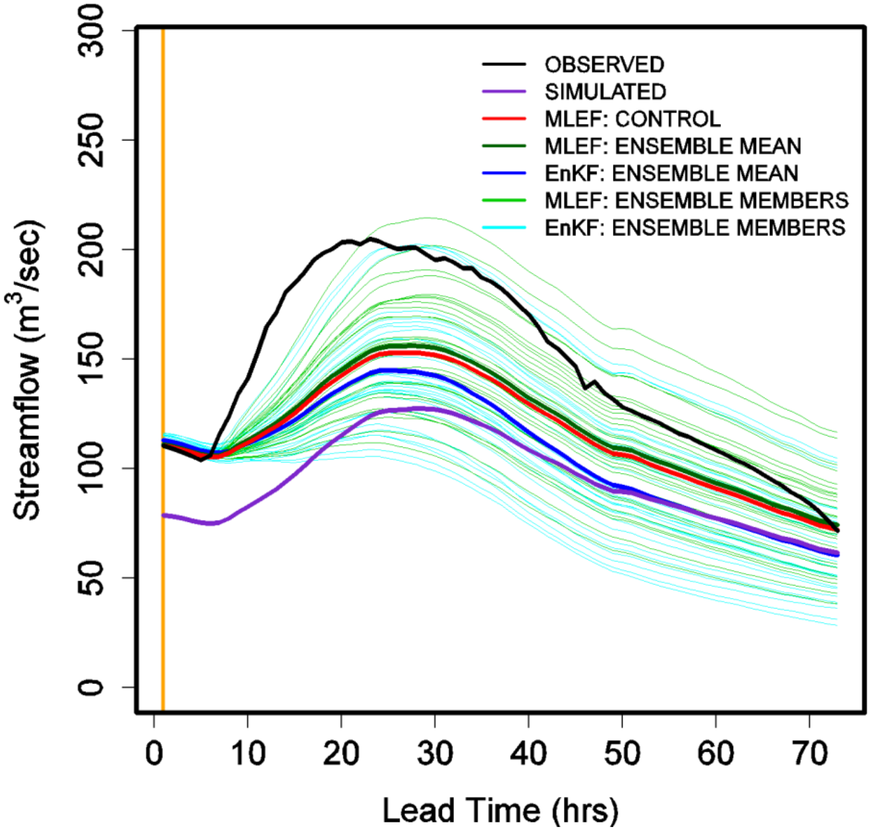
EnKF (Evensen 1994) vs. MLEF (Zupanski 2005)



From Rafieei Nasab et al. (2014)

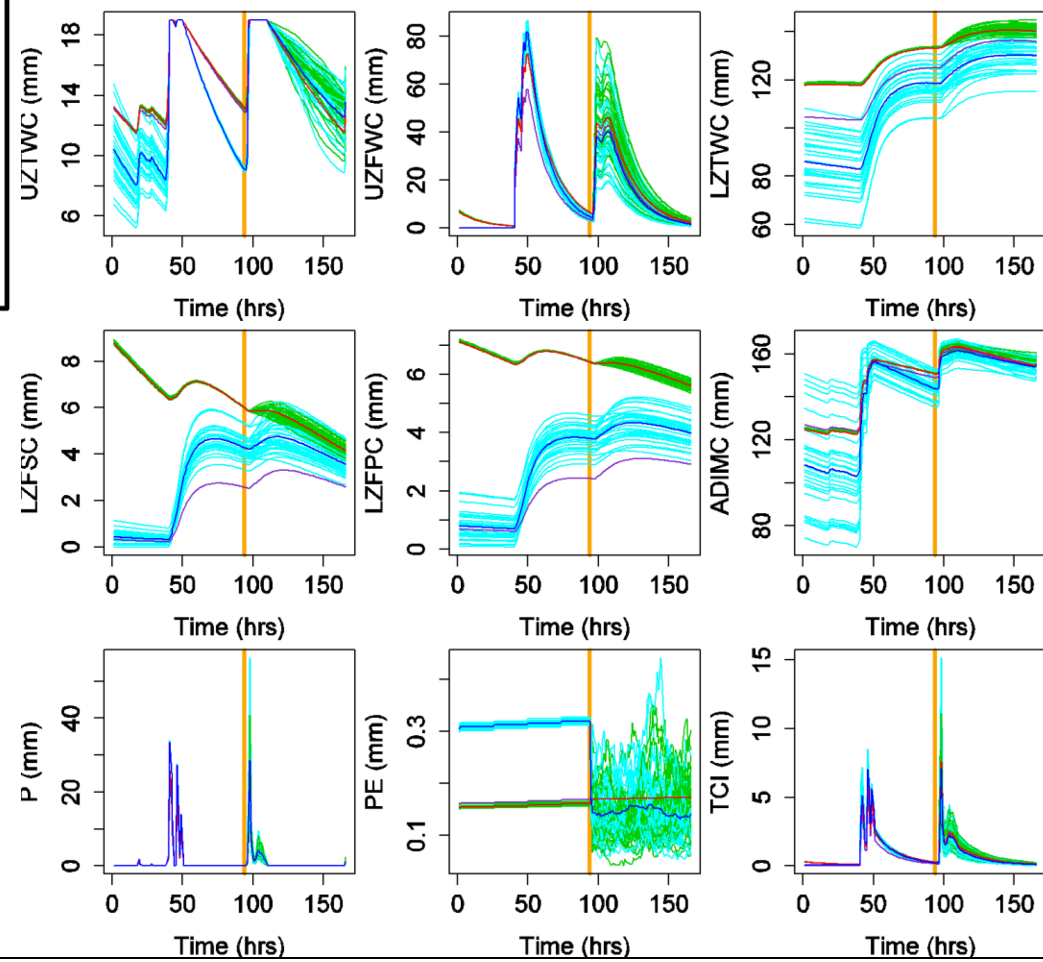
Sensitivity to ensemble size

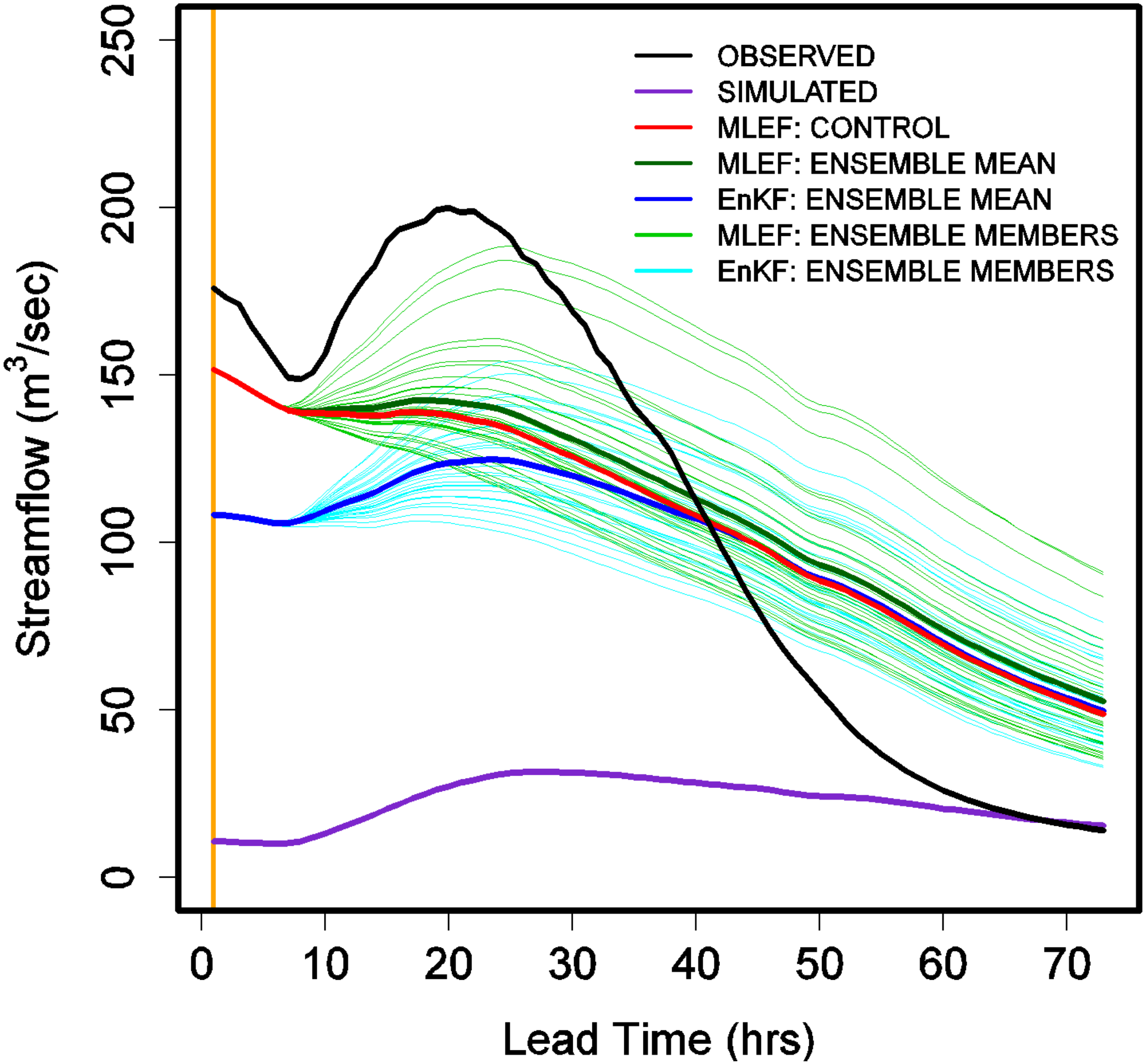




EnKF and MLEF streamflow ensembles are often similar

EnKF and MLEF solutions for soil moisture, however, are quite different





EnKF solution may be very poor.

Does minimization of mean square error suffice?

Choose the $\mathbf{x}(N)$ and $\mathbf{w}(n)$, $n = 0, \ldots, N-1$, which minimize

$$J[\mathbf{x}(N), \mathbf{w}(0) \cdots \mathbf{w}(N-1)]$$
$$= \sum_{n=1}^{N} [\mathbf{z}(n) - \mathbf{H}(n)\mathbf{x}(n)]'\mathbf{R}^{-1}(n)[\mathbf{z}(n) - \mathbf{H}(n)\mathbf{x}(n)] \qquad (6.5.1)$$
$$+ \sum_{n=0}^{N-1} \mathbf{w}'(n)\mathbf{Q}^{-1}(n)\mathbf{w}(n) + \mathbf{x}'(0)\boldsymbol{\psi}^{-1}\mathbf{x}(0)$$

subject to the constraint that

$$\mathbf{x}(n+1) = \mathbf{\Phi}(n)\mathbf{x}(n) + \mathbf{G}(n)\mathbf{w}(n)$$

where $\mathbf{R}(n)$, $\mathbf{Q}(n)$, and $\boldsymbol{\psi}$ are positive definite matrices chosen by engineering judgement.

Let $\hat{\mathbf{x}}(N|N)$ denote the resulting value of $\mathbf{x}(N)$. If one actually performs the minimization using Lagrange multipliers, the resulting equations for $\hat{\mathbf{x}}(N|N)$ are the same as those of Sections 6.2 and 6.3.

From Schweppe (1973)

Motivation for adding penalty for Type-II CB

- For accurate estimation/prediction of large amounts, reducing conditional bias (CB), in particular Type-II CB, is just as important as minimizing unconditional error variance

- Type-I CB $E[X \mid \hat{X}] - \hat{X}$ (analogous to reliability)

- Type-II CB $E[\hat{X} \mid X] - X$ (analogous to discrimination)

  - Climatological estimates are conditionally unbiased in the Type-I sense but conditionally biased in the Type-II sense
  - Perfect estimates are conditionally unbiased both in the Type-I and Type-II sense

- The focus here is on Type-II CB

Fisher solution to optimal linear estimation

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v}$$

$\mathbf{x}$: completely unknown

$$E\{\mathbf{v}\} = 0$$

$$E\{\mathbf{v}\mathbf{v}'\} = \mathbf{R}$$

$$\Sigma = [H^T R^{-1} H]^{-1}$$

$$\hat{X} = [H^T R^{-1} H]^{-1} H^T R^{-1} Z$$

From Schweppe (1973)

"Fisher-like" solution for minimizing J=$\sum_{EV}$+α$\sum_{CB}$, α≥0

$$\Sigma = \mathrm{B}[\hat{H}^T \Lambda^{-1} H]^{-1}$$

$$X^* = [\hat{H}^T \Lambda^{-1} H]^{-1} \hat{H}^T \Lambda^{-1} Z$$

where

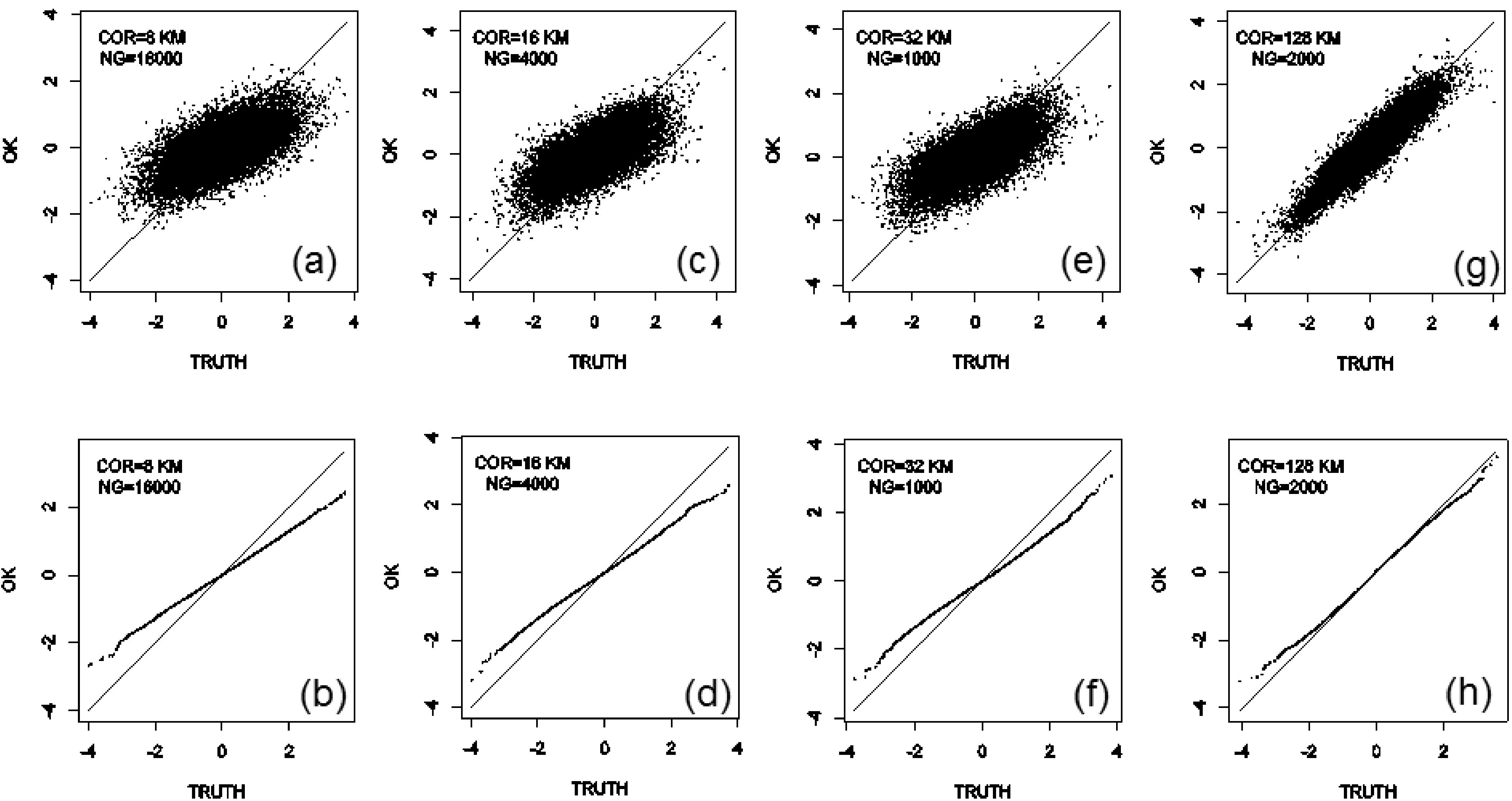
$$\hat{H}^T = H^T + \alpha \Psi_{XX}^{-1} \Psi_{XZ}$$

$$\Lambda = R + \alpha(1-\alpha)\Psi_{ZX} \Psi_{XX}^{-1} \Psi_{XZ} - \alpha H \Psi_{XZ} - \alpha \Psi_{ZX} H^T$$

$$\mathrm{B} = \alpha \Psi_{XX} \hat{H}^T \Lambda^{-1} \hat{H} + (1+\alpha) I$$

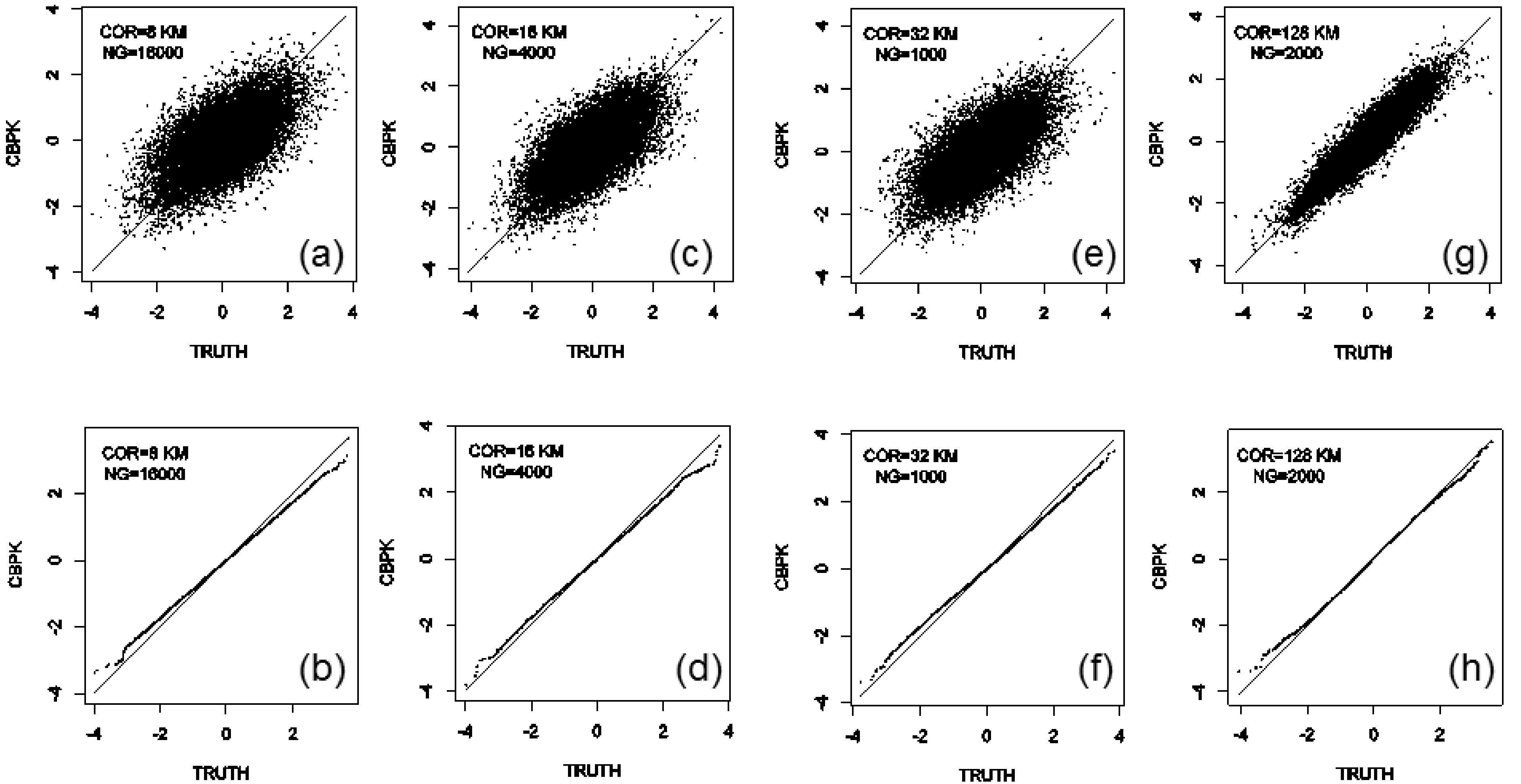
From Seo (2013), Seo et al. (2014)

OK/KF estimates vs. truth

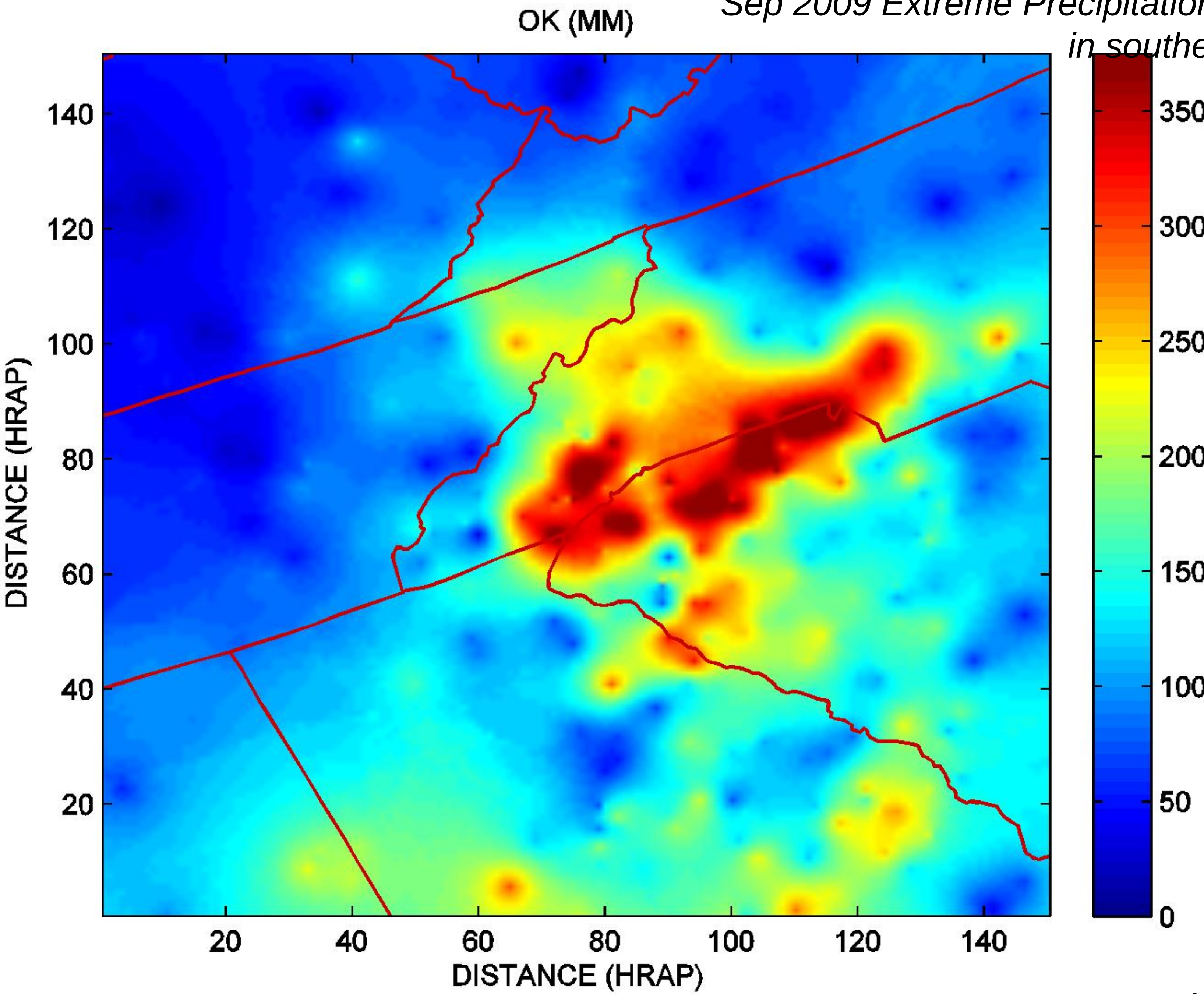


From Seo (2013)

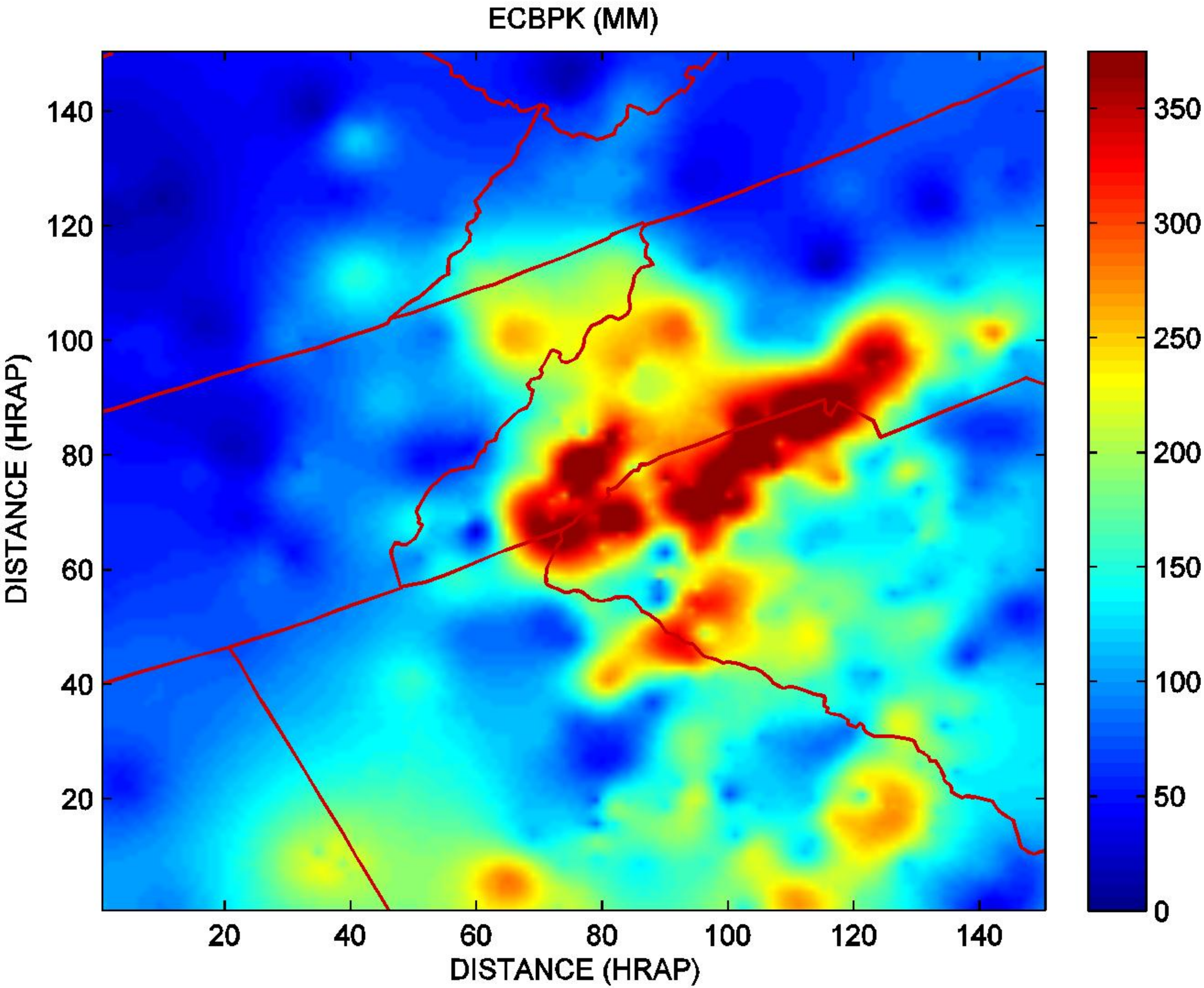
CBPK/CBPKF estimates vs. truth

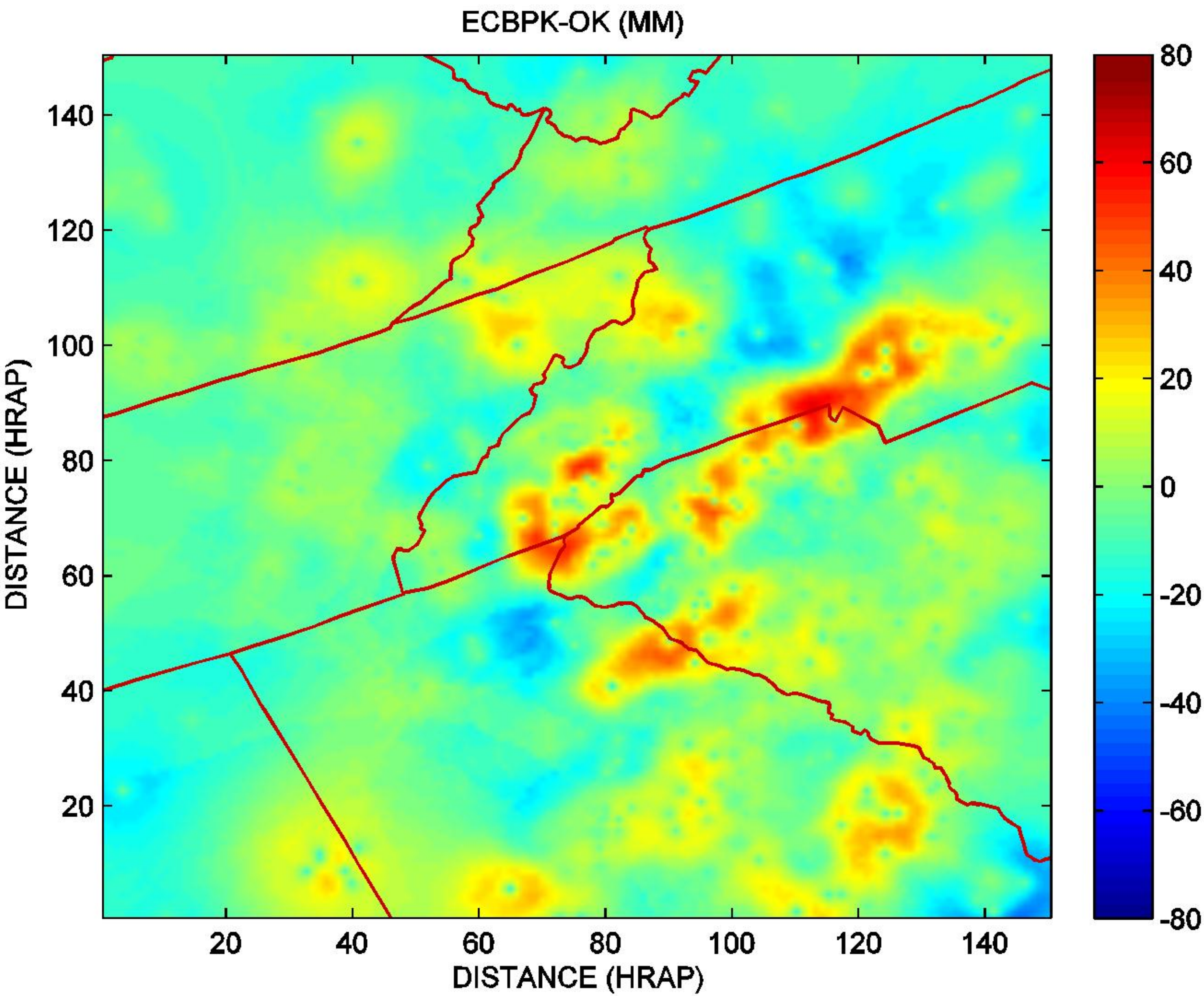


Sep 2009 Extreme Precipitation Event
in southeast US

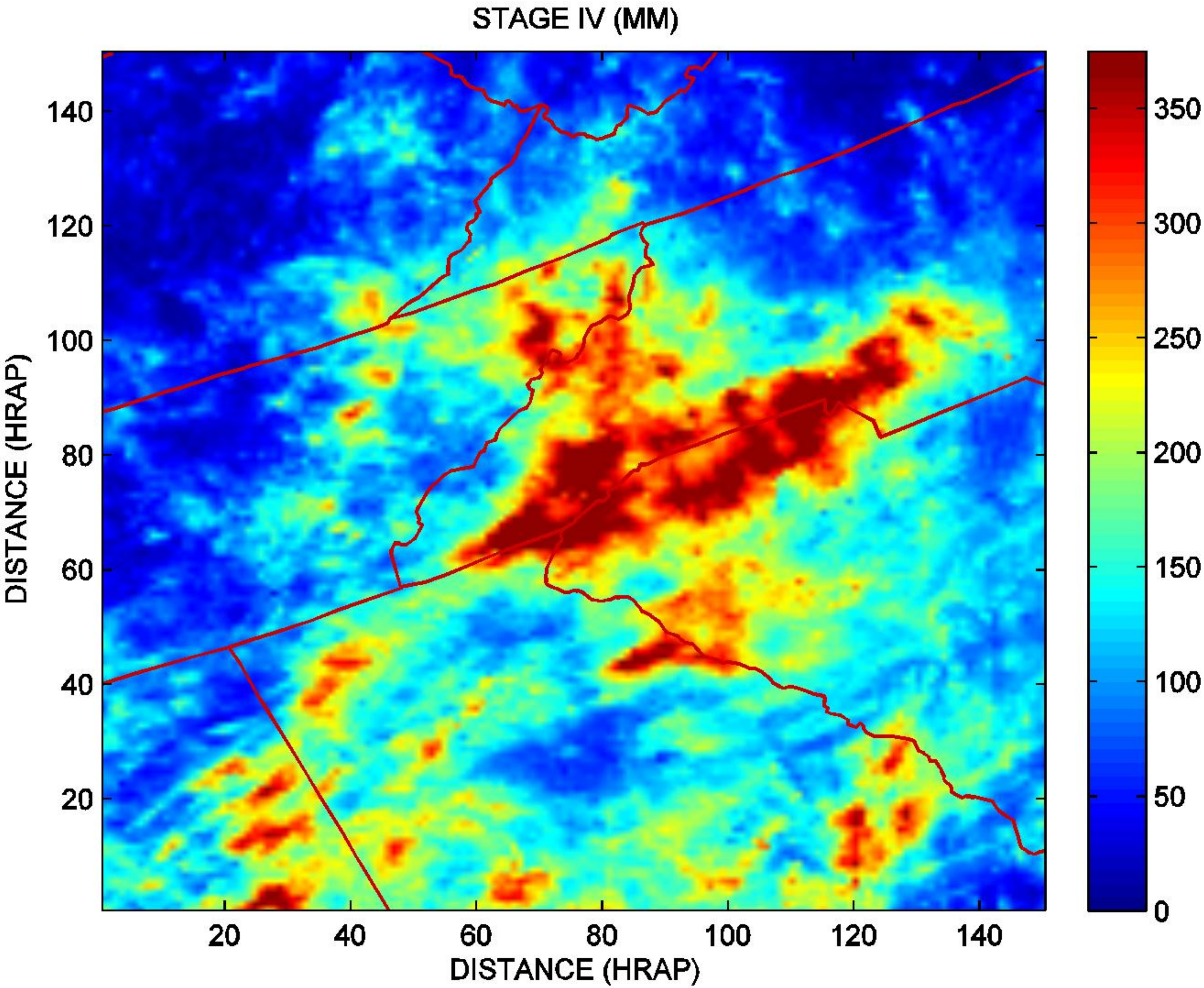


From Seo et al. (2014)

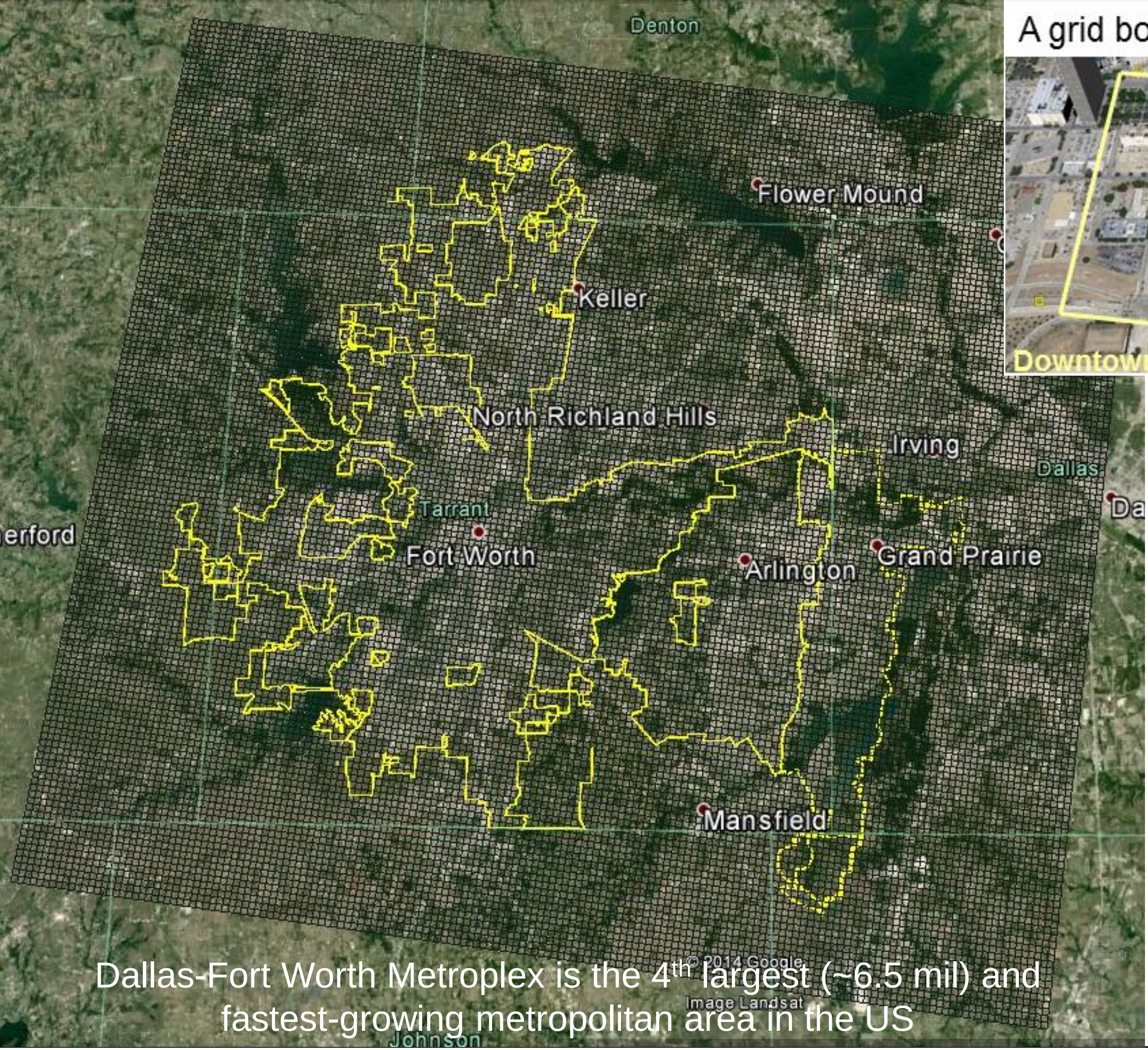

ECBPK (MM)
DISTANCE (HRAP)
DISTANCE (HRAP)
20
40
60
80
100
120
140
0
50
100
150
200
250
300
350

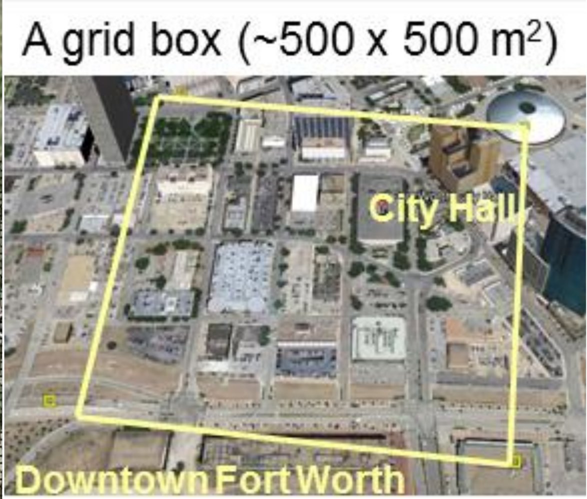


ECBPK-OK (MM)
DISTANCE (HRAP)
DISTANCE (HRAP)
20
40
60
80
100
120
140
80
60
40
20
0
-20
-40
-60
-80



STAGE IV (MM)
DISTANCE (HRAP)
DISTANCE (HRAP)
20
40
60
80
100
120
140
0
50
100
150
200
250
300
350

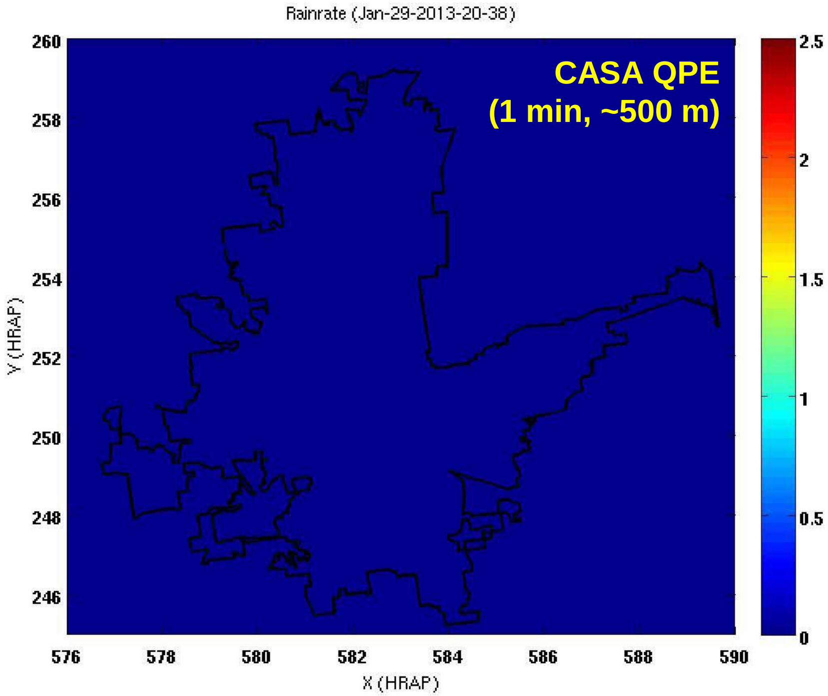


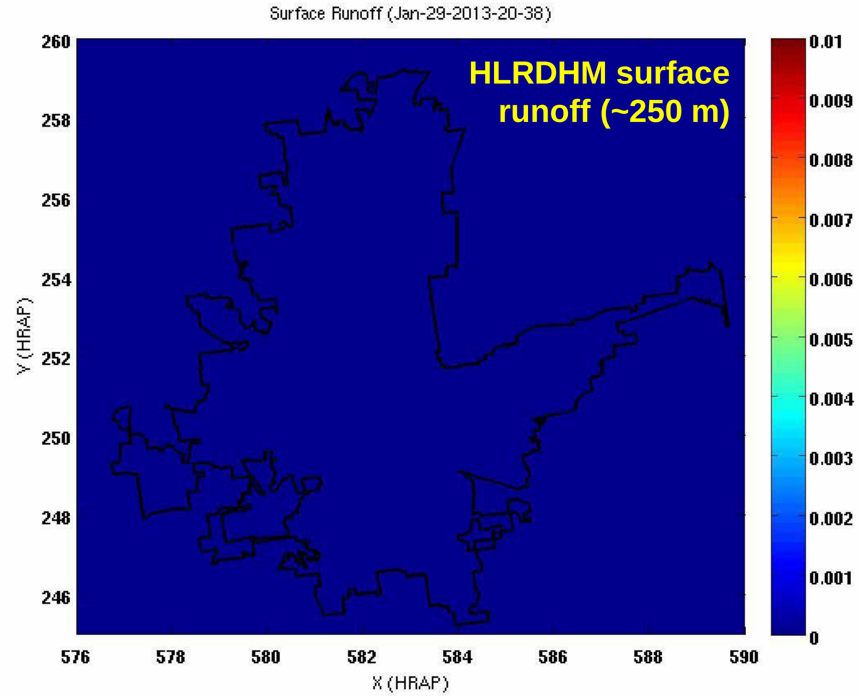


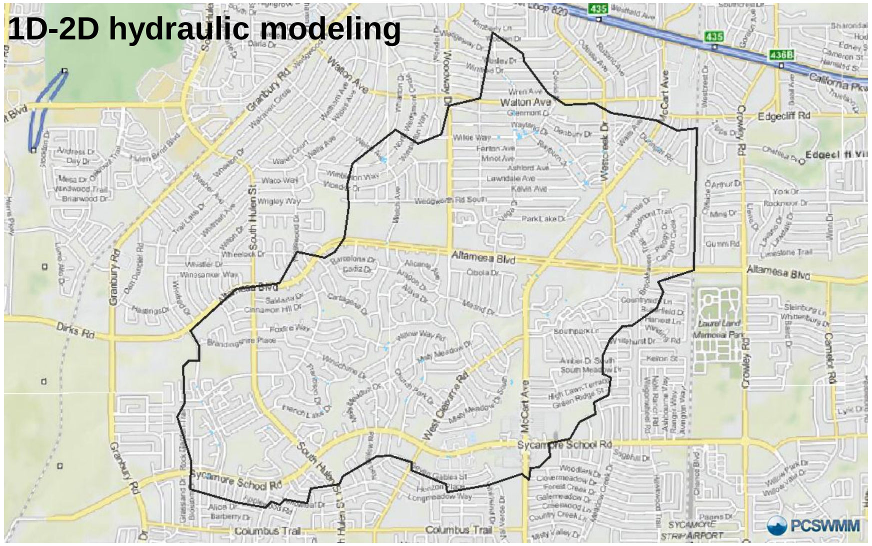
Dallas-Fort Worth Metroplex is the 4th largest (~6.5 mil) and fastest-growing metropolitan area in the US

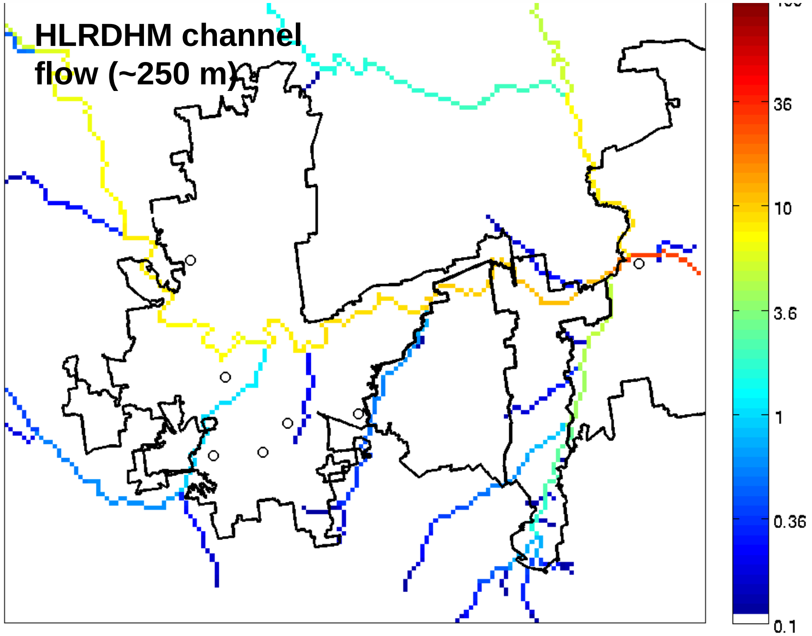
The HLRDHM domain encompassing Fort Worth, Arlington and Grand Prairie.

Overlaid is the 500x500 m^2 CASA QPE grid.

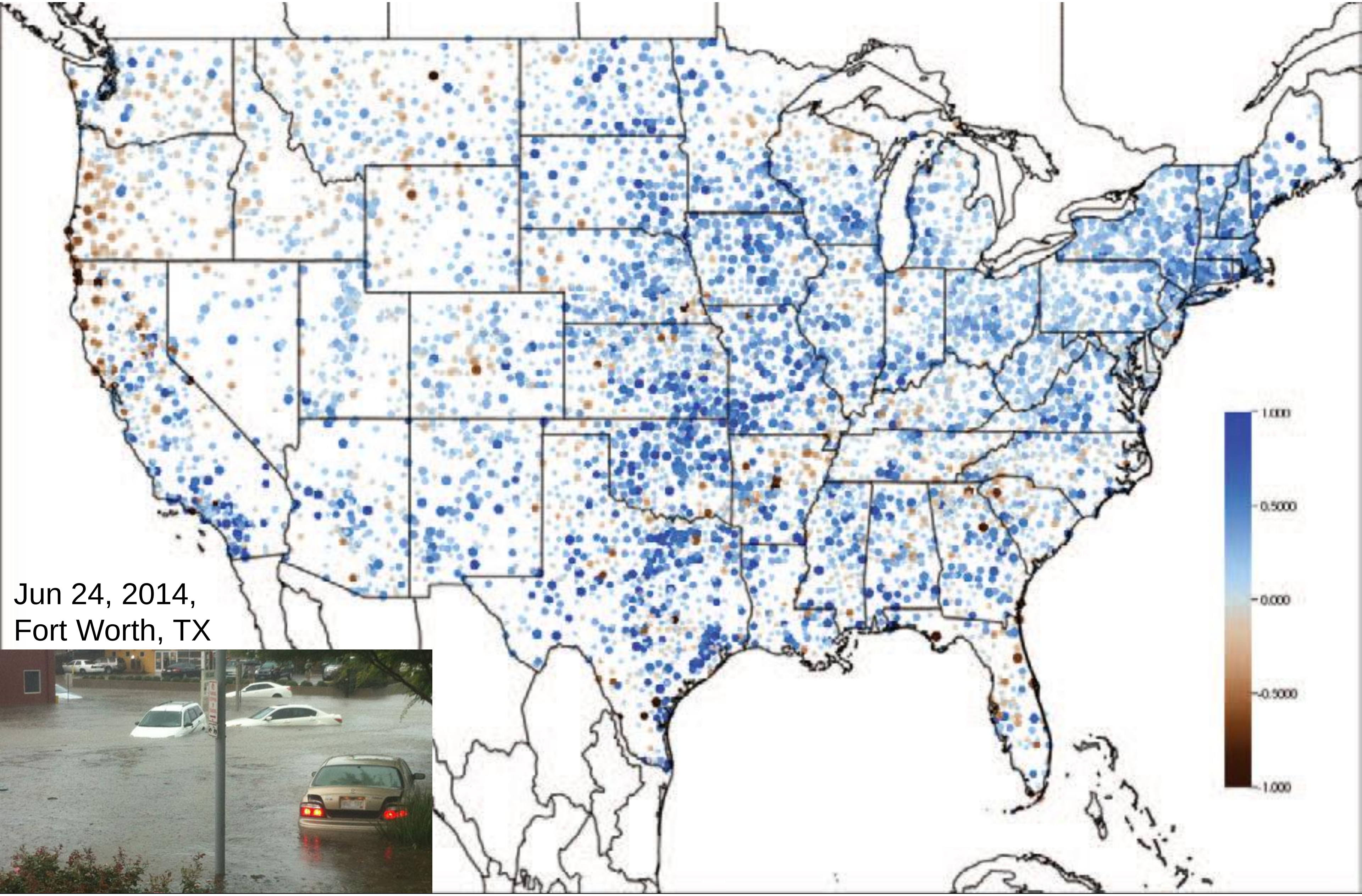

Rainrate (Jan-29-2013-20-38)
CASA QPE
(1 min, ~500 m)
Y (HRAP)
X (HRAP)
260
258
256
254
252
250
248
246
576
578
580
582
584
586
588
590
2.5
2
1.5
1
0.5
0


Surface Runoff (Jan-29-2013-20-38)
HLRDHM surface
runoff (~250 m)
Y (HRAP)
X (HRAP)
260
258
256
254
252
250
248
246
576
578
580
582
584
586
588
590
0.01
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0


1D-2D hydraulic modeling
PCSWMM

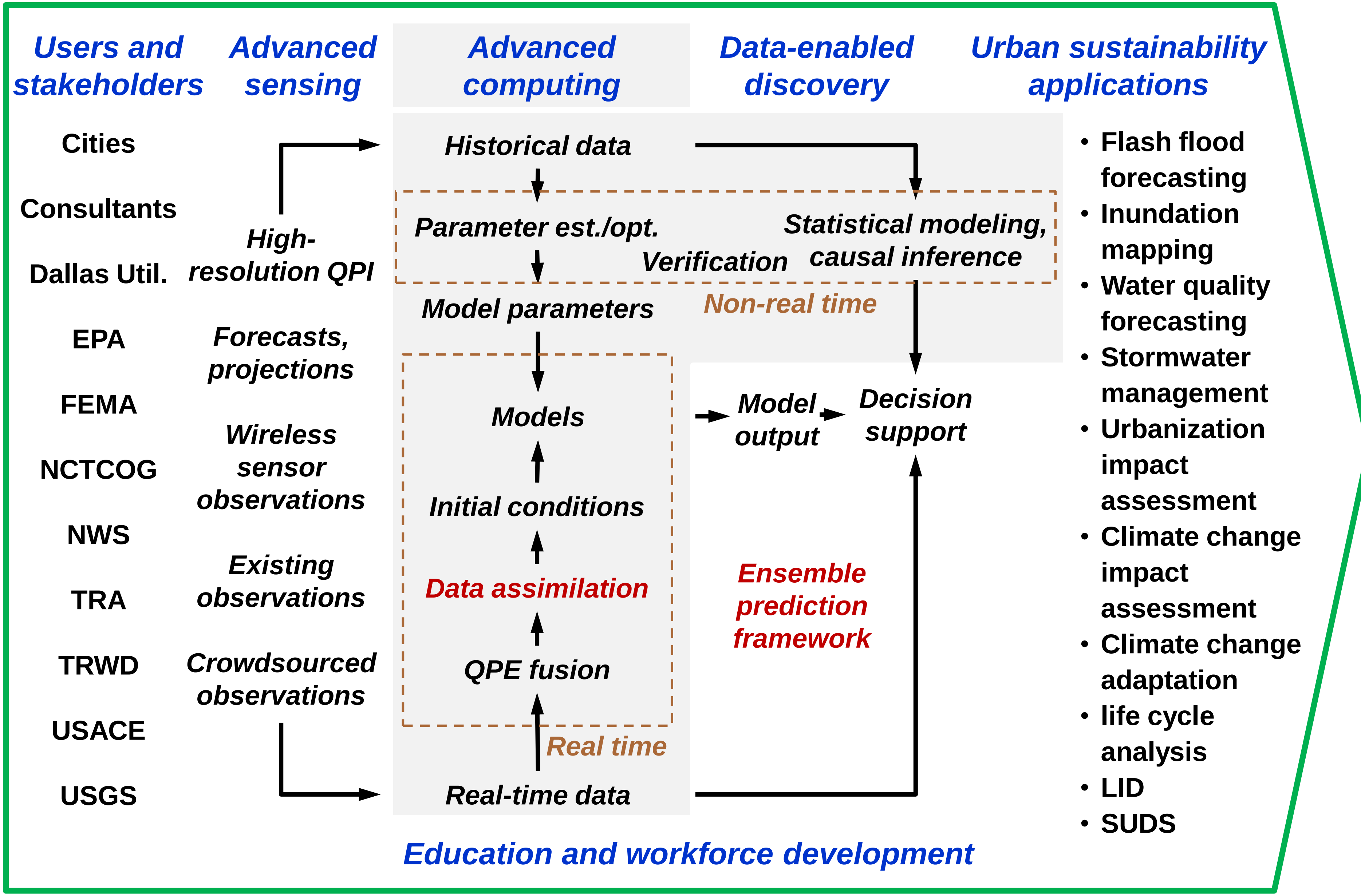

HLRDHM channel
flow (~250 m)
36
10
3.6
1
0.36
0.1

Changes in observed 20-yr return value of the daily accumulated precipitation (in.) from 1948 to 2010 (Kunkel et al. 2013)

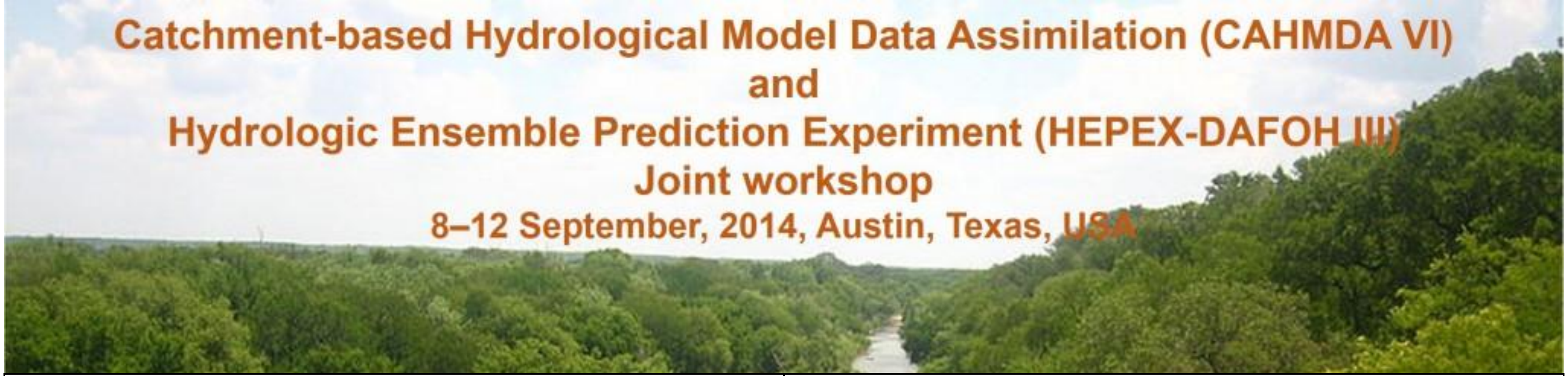


Jun 24, 2014,
Fort Worth, TX

Ensemble prediction for urban applications – Dallas-Fort Worth Metroplex example



CAHMDA-DAFOH http://www.jsg.utexas.edu/ciess/cahmda-vi-hepex-dafoh-iii/



Abstract submission Deadline	June 30, 2014 (extended)
Early Registration Deadline	June 30, 2014
Training Course	Sep 06-07, 2014
Joint CAHMDA-VI and HEPEX-DAFOH III Workshop	Sep 08-12, 2014
Excursions and Activities	Sep 13, 2014

THANK YOU

For more information, contact:

djseo@uta.edu