IMPROVING REAL-TIME WATER QUALITY FORECASTING VIA ENSEMBLE DATA ASSIMILATION

Sunghee Kim[1], Hamideh Riazi[2], Changmin Shin[1,2], D.-J. Seo[1]

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

[2]Water Quality Control Center, National Institute of Environmental Research, Incheon, Korea

Outline

- Background
- Why data assimilation (DA) for water quality forecasting?
- What is DA?
  - Introduction
  - Illustration
- Evaluation
  - Hindcasting experiment
  - some results
- Conclusions and ongoing work
  - A plea/plug

Overview of Operational WQ Forecast in NIER/Korea

The procedure of 7-day WQ forecast

Weather data

- *Weather observation data*
- *Weather forecasting data (UM-Regional / UM-Global)*

Hydrologic data

- *Flow & stage monitoring data*
- *Dam water release data & plan*

Water quality data

- *Manual WQ monitoring data*
- *Automatic WQ monitoring data*
- *Tele-monitoring system (TMS) data*

Data Preprocessing

Watershed modeling (HSPF)

- Update ICs & BCs
- *HSPF model run*
- Predict Flow & WQ
- Validate the prediction

River WQ modeling (EFDC)

- Update ICs & BCs
- *EFDC model run*
- Predict WQ
- Validate the prediction

Report forecasting results

FEWS-NIER

- Built on the Flood Early Warning System (FEWS) developed by Deltares
- Uses:
  - Hydrologic Simulation Program-FORTRAN (HSPF) for watershed modeling
  - Environmental Fluid Dynamics Code (EFDC) for river modeling

NWS's new operational hydrologic forecasting system, the Community Hydrologic Prediction System (CHPS), is also built on the FEWS.

Why data assimilation (DA)?

- There are many sources of error in the end-to-end water quality forecast process
- Reducing uncertainty in the initial conditions (IC) is one of the most cost-effective ways to improve forecast accuracy
  - Current practice - Manual updating of IC's based on real-time observations
  - Need - Automatic updating using DA

HSPF state variables

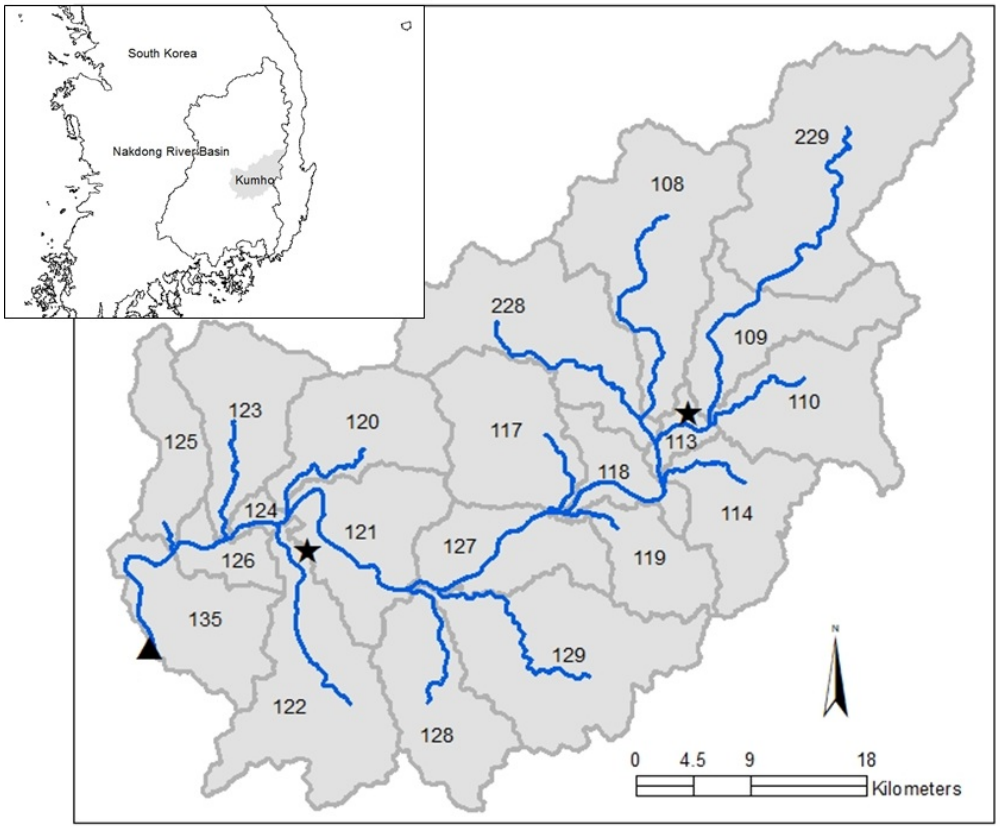
Module	Variable name	Observation data	Definition
PERLND	CEPS	No	interception storage
	SURS	No	surface (overland flow) storage
	UZS	No	upper zone storage
	IFWS	No	interflow storage
	LZS	No	lower zone storage
	AGWS	No	active groundwater storage
	GWVS	No	index to groundwater slope
	SQO-NH_4	No	storage of NH_4 on the surface
	SQO-NO_3	No	storage of NO_3 on the surface
	SQO- PO_4	No	storage of PO_4 on the surface
	SQO- BOD	No	storage of BOD on the surface
IMPLND	RETS	No	retention storage
	SURS	No	surface (overland flow) storage
	SQO-NH_4	No	storage of NH_4 on the impervious surface
	SQO-NO_3	No	storage of NO_3 on the impervious surface
	SQO- PO_4	No	storage of PO_4 on the impervious surface
	SQO- BOD	No	storage of BOD on the impervious surface
RCHRES	VOL	cal.	volume of water in the RCHRES at end of interval
	TW	Yes	water temperature
	DOX	Yes	dissolved oxygen concentration
	BOD	Yes	biochemical oxygen demand concentration
	NO_3	Yes	dissolved concentration of NO_3
	TAM	Yes	dissolved concentration of TAM (incl. NH_3, NH_4)
	PO_4	Yes	Dissolved concentration of PO_4
	PHYTO	cal.	phytoplankton concentration (as biomass)
	TN	Yes	Total nigrogen
	TP	Yes	Total phosphorus
	TOC	Yes	Total organic carbon

- ***Most state variables in water quality models are not observed in reality***
- ***Very large uncertainties exist in model ICs***

Plus precipitation and PE(potential evapotranspiration)

Nakdong River Basin (23,817 km^2)

Kumho Sub-Basin (2,000 km^2)



31 model segments →
292 variables to adjust

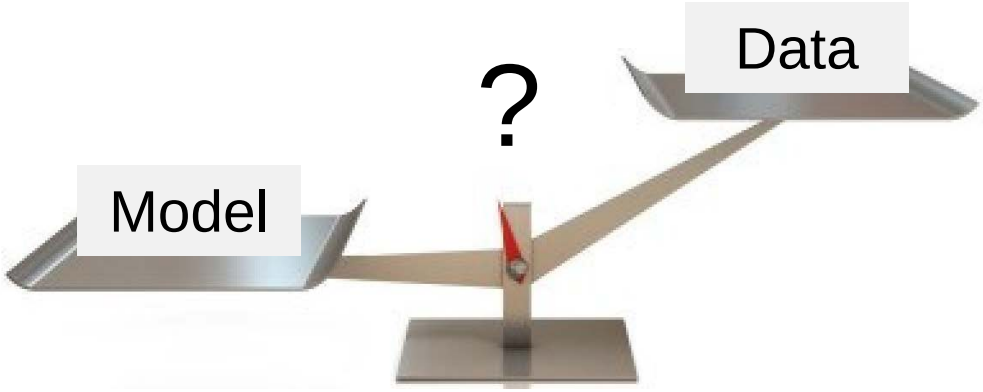
Data assimilation (DA)

- Rooted in optimal estimation/control theory
- Often referred to as state updating in hydrology (1970's ~)
- Popularized by highly successful applications in oceanography and weather forecasting (1980's ~)

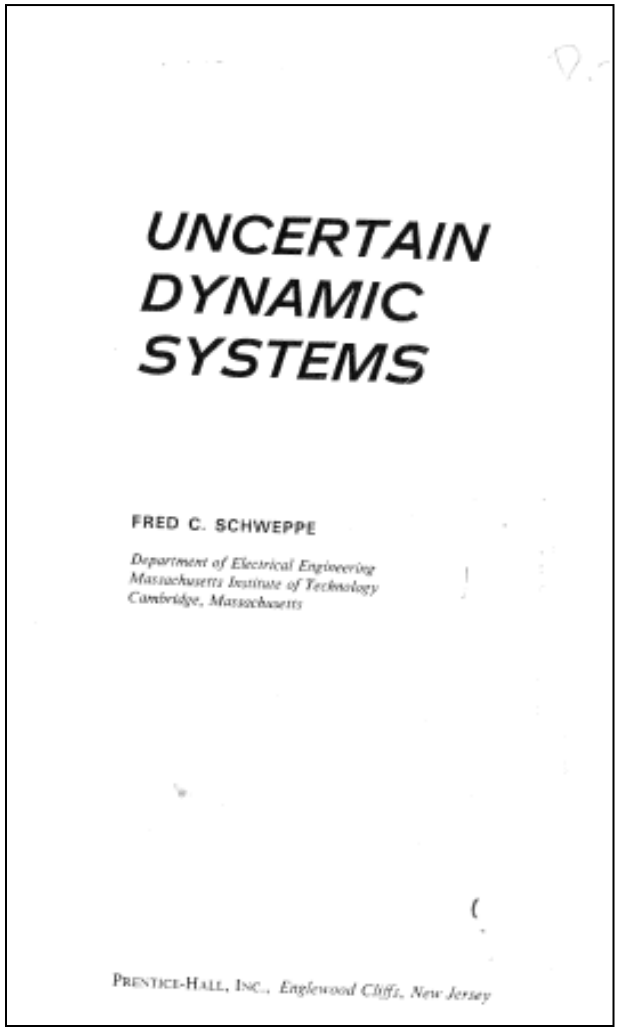
DA (cont.)

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

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



DA theory



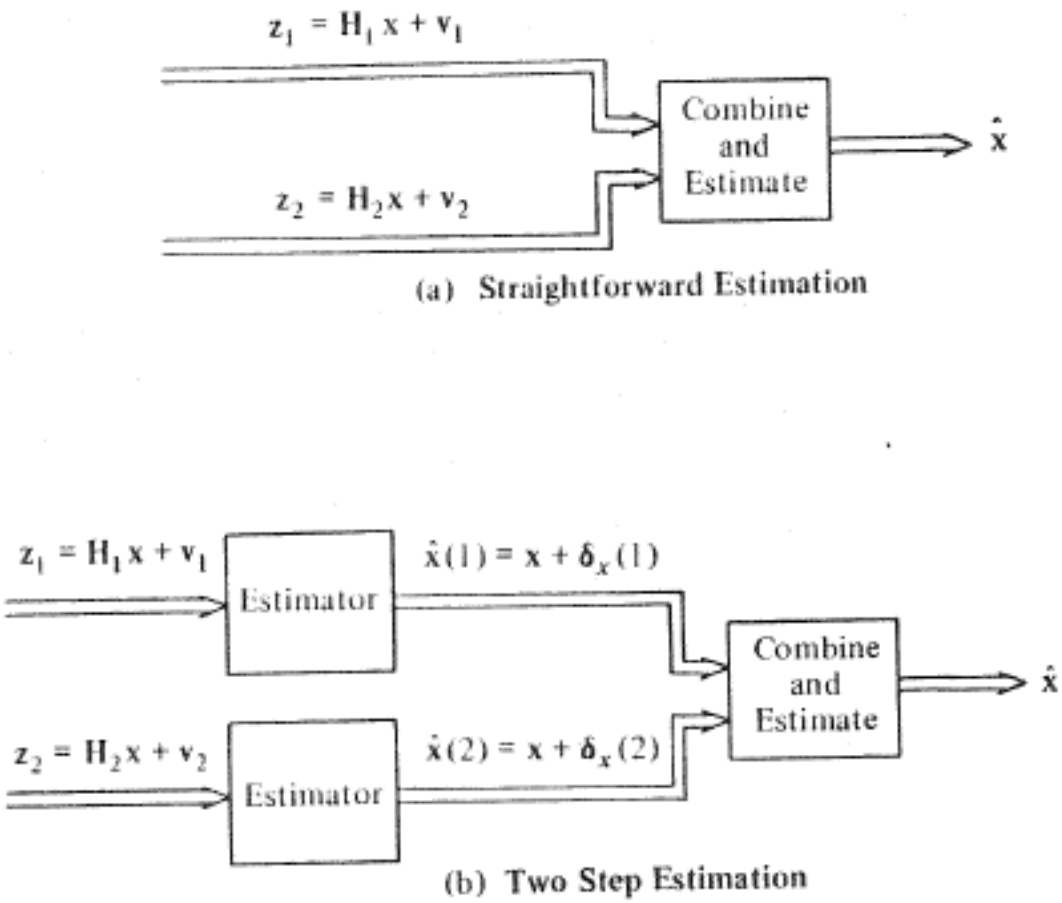
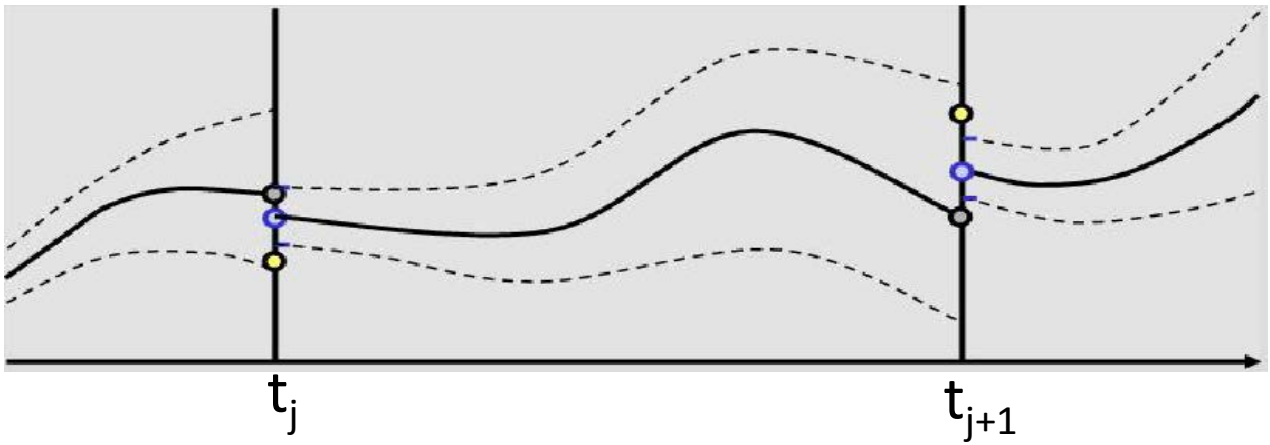


Figure 5.7.1 A dichotomy.

Kalman filtering (Kalman 1960)



- Model prediction
- New measurement
- Analysis from DA
- Model error bounds

Prediction:

$$\hat{X}_t = f_I\left(I_t \mid \theta, X_{t-1}\right)$$
$$\hat{O}_t = f_o\left(\hat{X}_t \mid \phi\right)$$

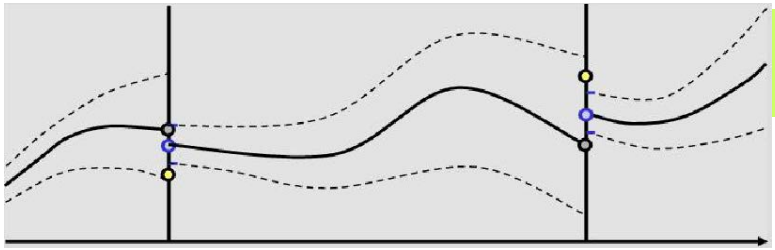
(Bayesian) Updating:

$$X_t^a = \hat{X}_t + K_t(O_t - \hat{O}_t)$$

Kalman Gain

Prediction

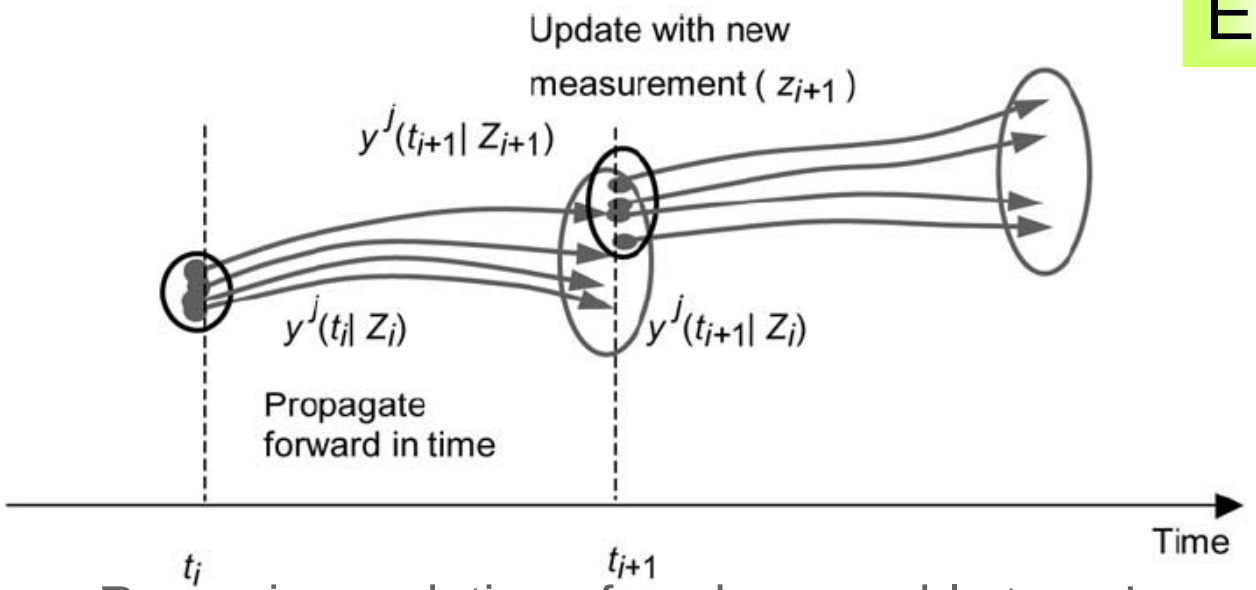
Ensemble Kalman filtering (Evensen 1994)



KF

Monte Carlo Approximation

EnKF



Recursive updating of each ensemble trace!

DA (cont.)

- Methodology for obtaining a "best" estimate of the state of the system by optimally combining model predictions and observations.

$$x_k = M_{k,k-1}(x_{k-1}) + w_{k-1}$$

$$y_k = H(x_k) + \varepsilon_k$$

x: Model state vector

M: Dynamical model

w: Model error

y: Observations vector

H: Observation operator

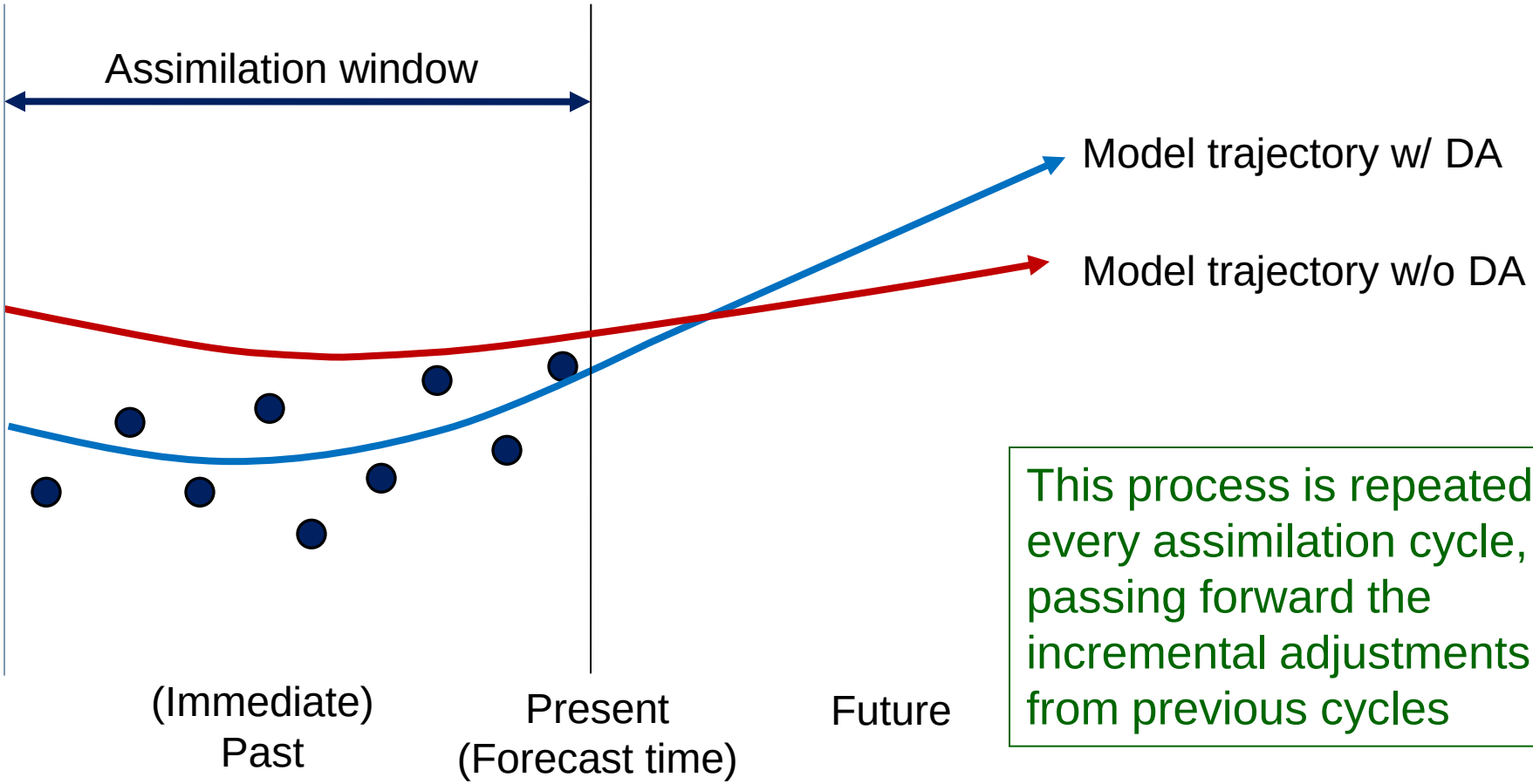
ε: Observation error

Minimize

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

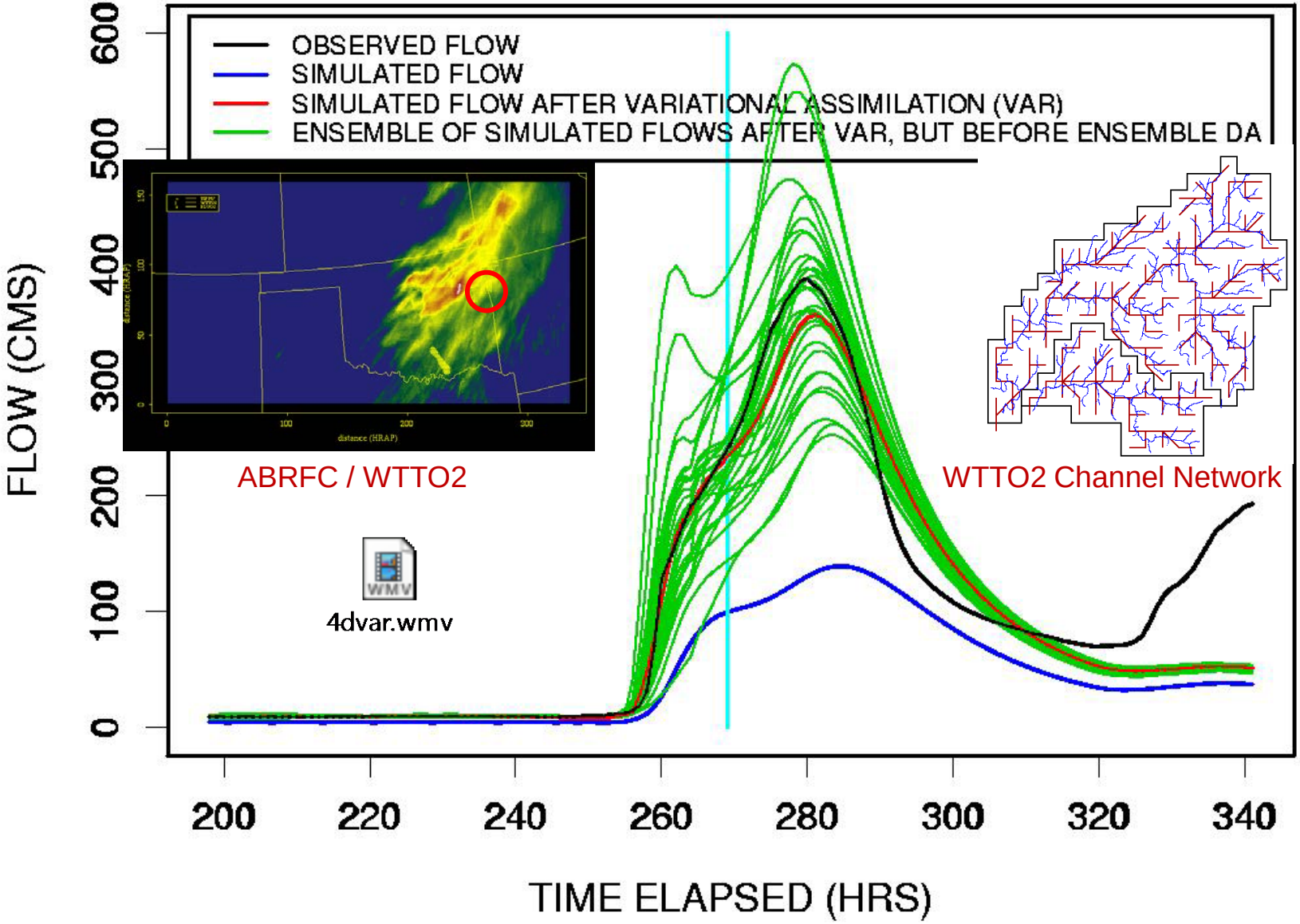
Variational assimilation (VAR)

Analogous to successive (in time) “batch-by-batch” curve fitting via least-squares regression



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

Illustration of DA using distributed hydrologic model



DA (cont.)

- Different techniques available
  - Linear vs. nonlinear model dynamics
  - Linear vs. nonlinear observation equation
  - Control vs. ensemble
  - Variational assimilation (VAR), ensemble Kalman filter (EnKF), etc.

Ensemble DA techniques

Ensemble Kalman filter (EnKF)

- Evensen (1994)
- Can handle nonlinear model dynamics
- Cannot (well) handle nonlinear observation equations
- Algorithmically (relatively) simple

Maximum likelihood ensemble filter (MLEF)

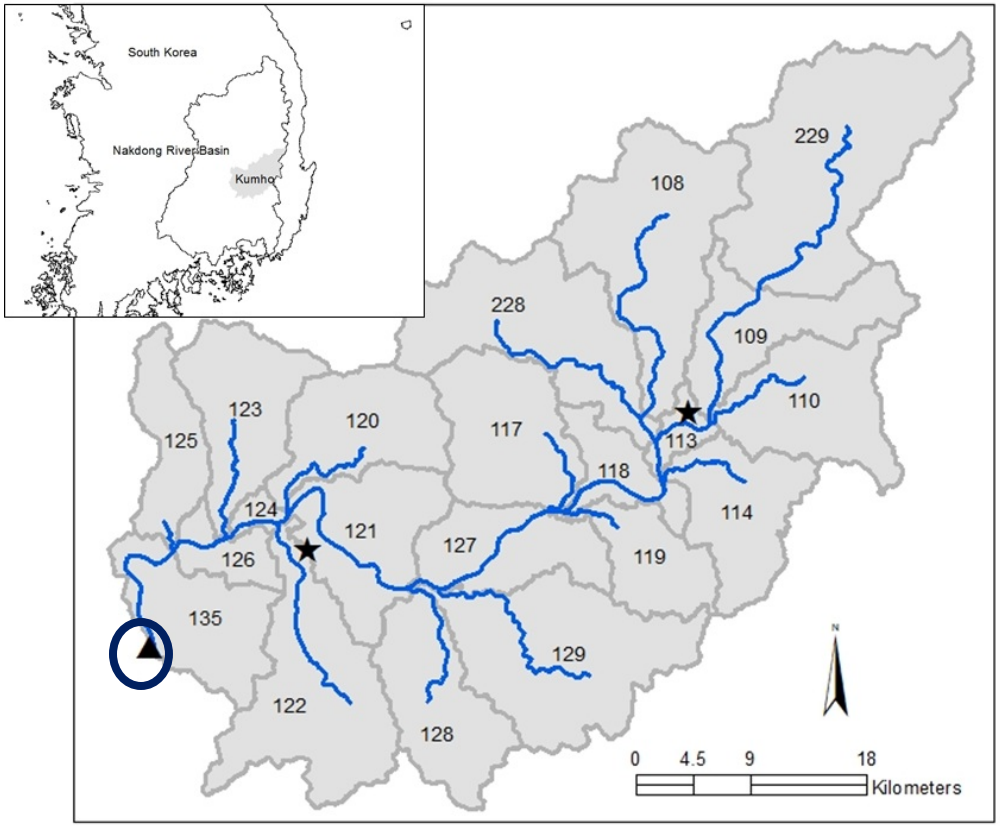
- Zupanski (2005)
- Can handle nonlinear model dynamics
- Can handle nonlinear observation equations
- Algorithmically more involved

DA hindcast experiment

- 2008~2009 for the Kumho Catchment
- DA method: MLEF (Zupanski, 2005), fixed-lag smoother formulation
- Assimilation cycle: once per day
- Assimilation window: 7 days
- Lead time: 72 hours
- Control variables: 292 in total
  - 28 state variables, 2 multiplicative adjustment factors (for precipitation and PE)
  - 31 model segments
- Observed forcing used instead of predicted forcing
  - Precipitation, potential evapotranspiration (based on air temperature, wind, solar radiation)
- Assimilate streamflow (QIN), TW, NH_4 ,NO_3,PO_4,CHL-a, TN,TP, TOC, BOD, DO (available ~ once a week)

Nakdong River Basin (23,817 km^2)

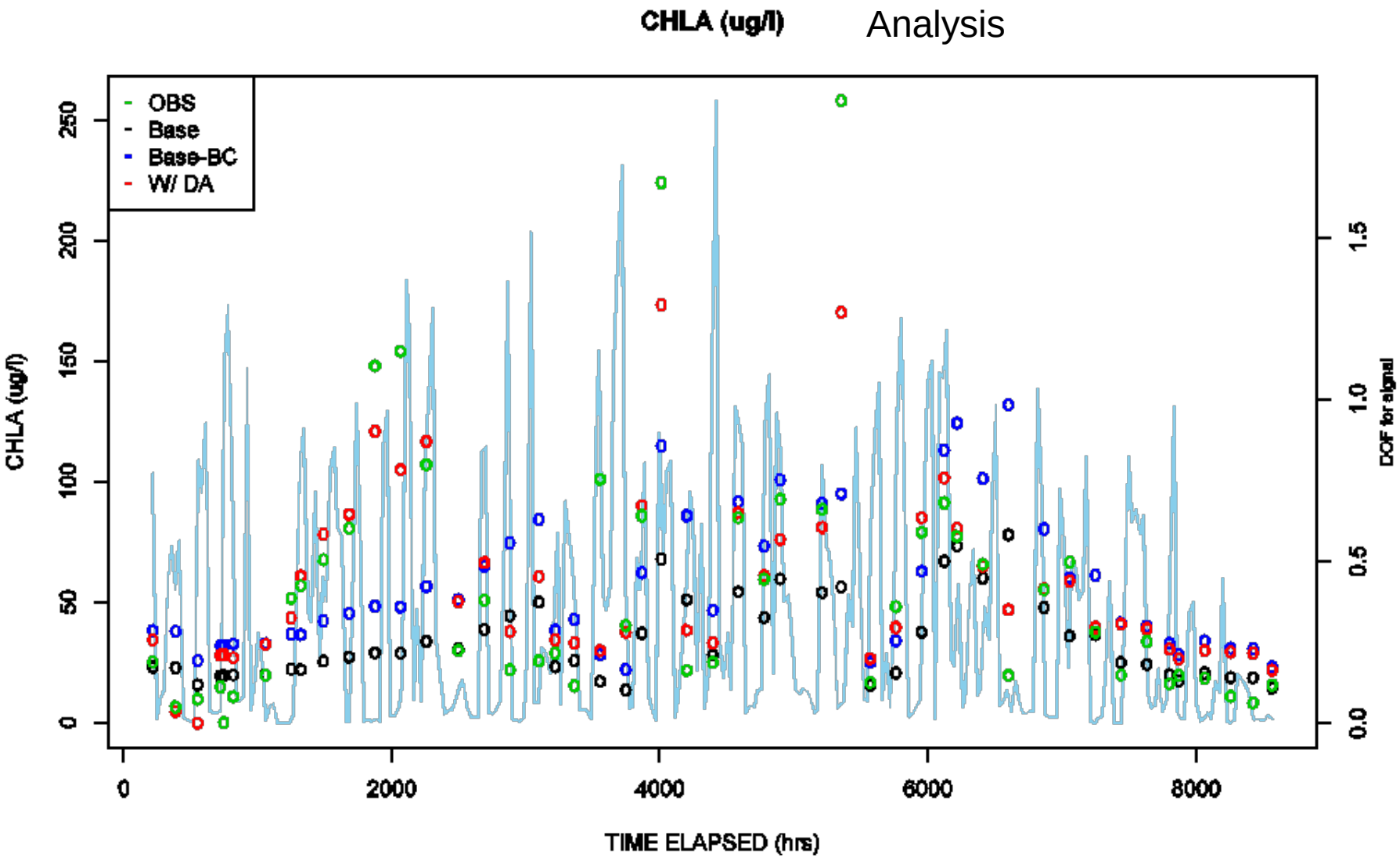
Kumho Sub-basin (2,000 km^2)



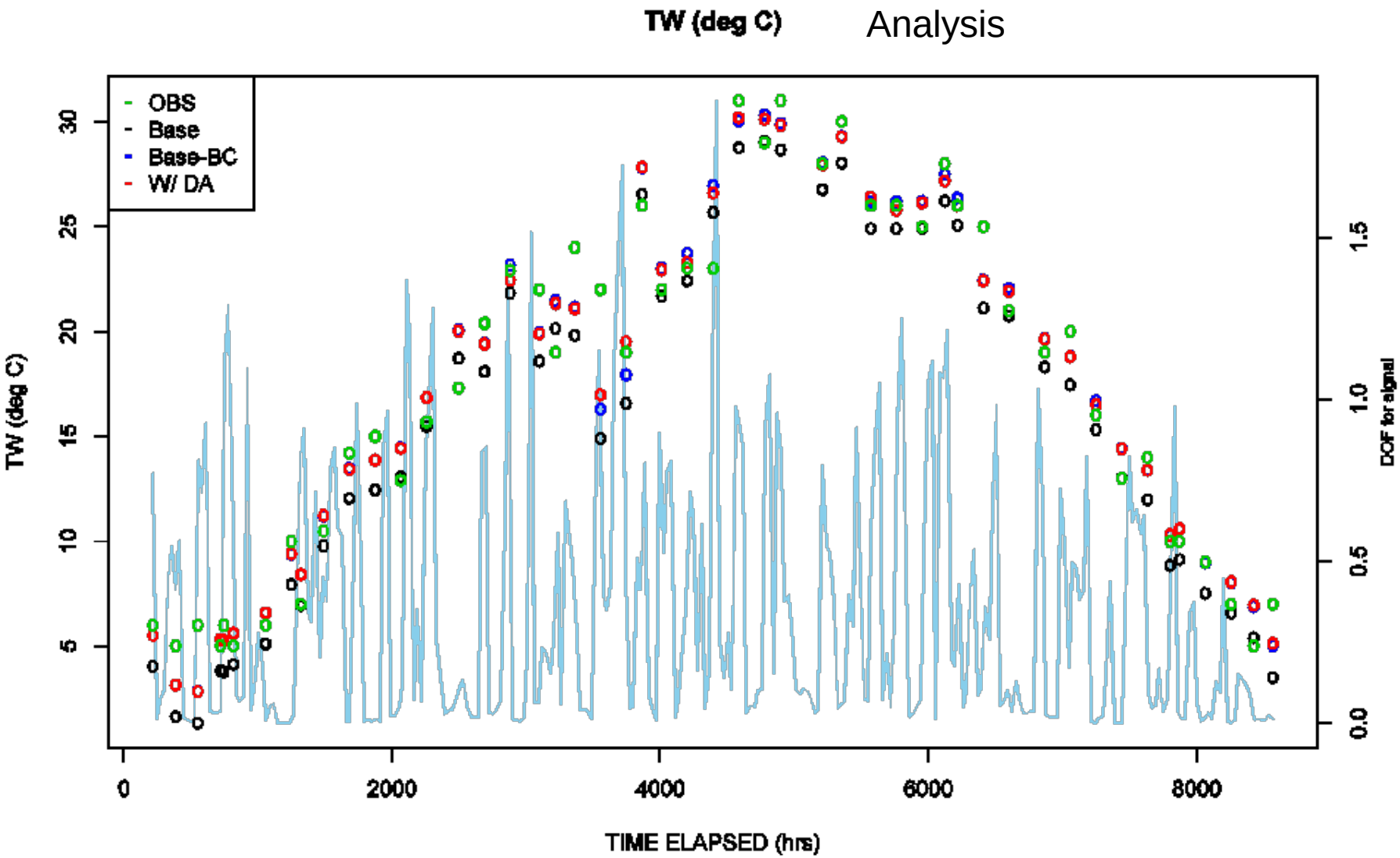
Evaluation of MLEF-HSPF: Example results

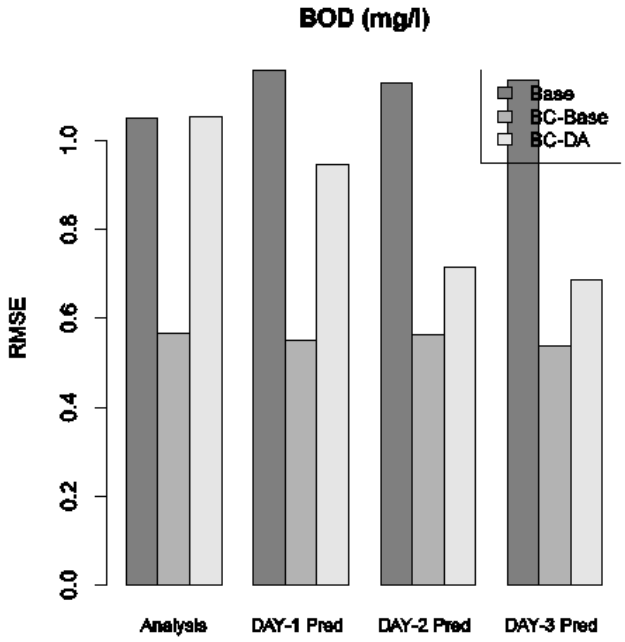
- For Kumho, 2008
- Reach 135
- Base: HSPF simulation without bias correction or DA
- BC-Base: HSPF simulation with bias correction (as post processing)
- BC-DA: HSPF simulation with bias correction (as part of the observation equation) and DA

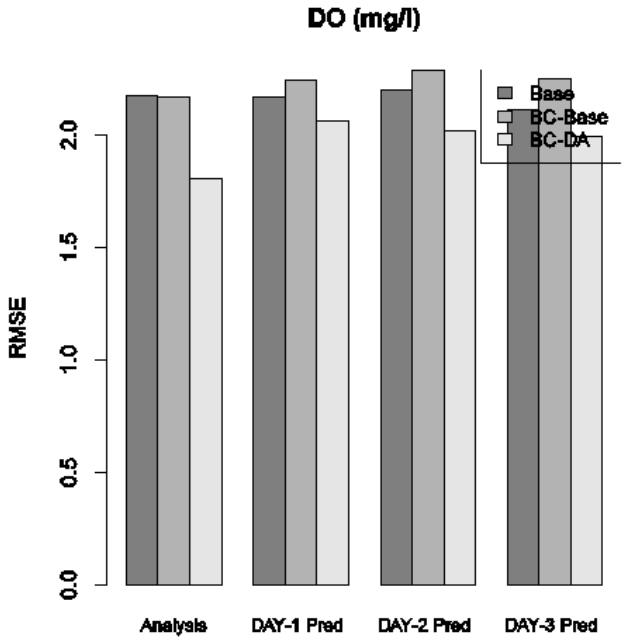
Analysis

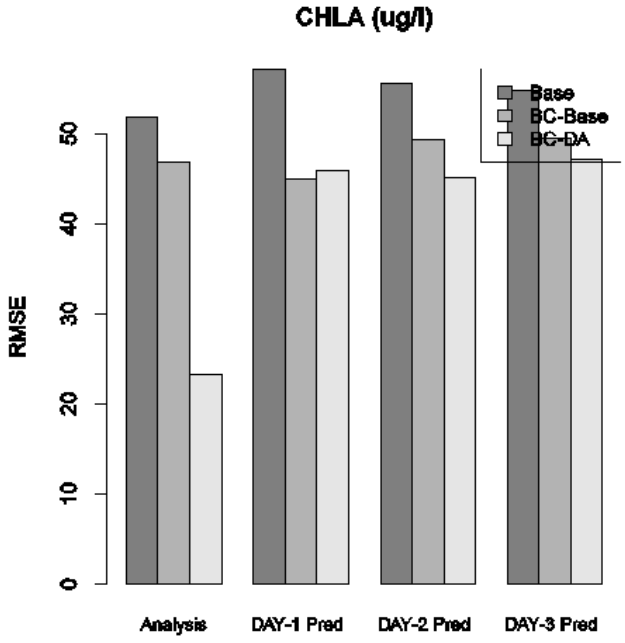

CHLA (ug/l)
OBS
Base
Base-BC
W/ DA
CHLA (ug/l)
0
50
100
150
200
250
DOF for signal
0.0
0.5
1.0
1.5
0
2000
4000
6000
8000
TIME ELAPSED (hrs)

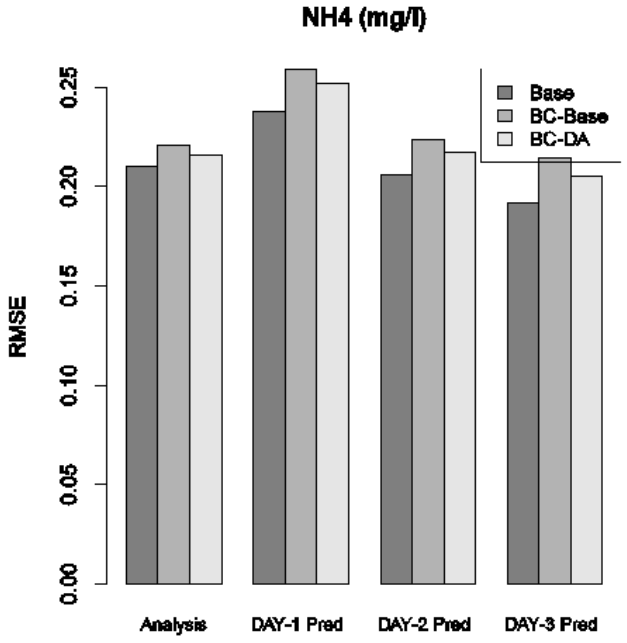
Analysis

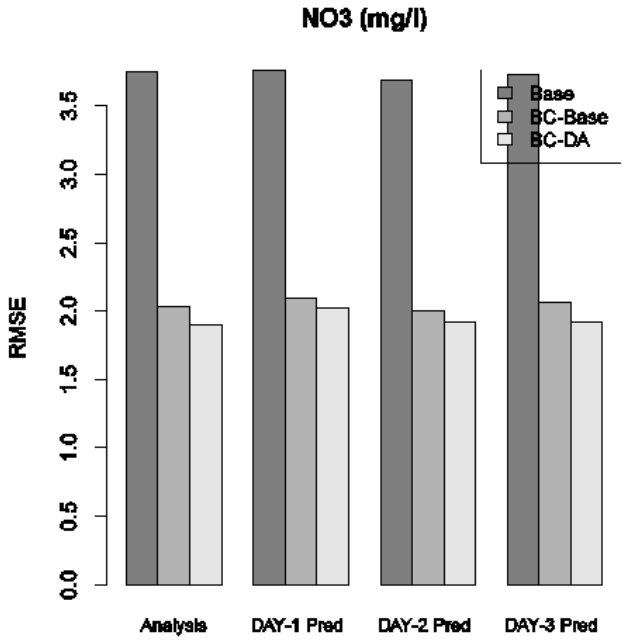

TW (deg C)
OBS
Base
Base-BC
W/ DA
TW (deg C)
DOF for signal
TIME ELAPSED (hrs)
0
2000
4000
6000
8000
5
10
15
20
25
30
0.0
0.5
1.0
1.5

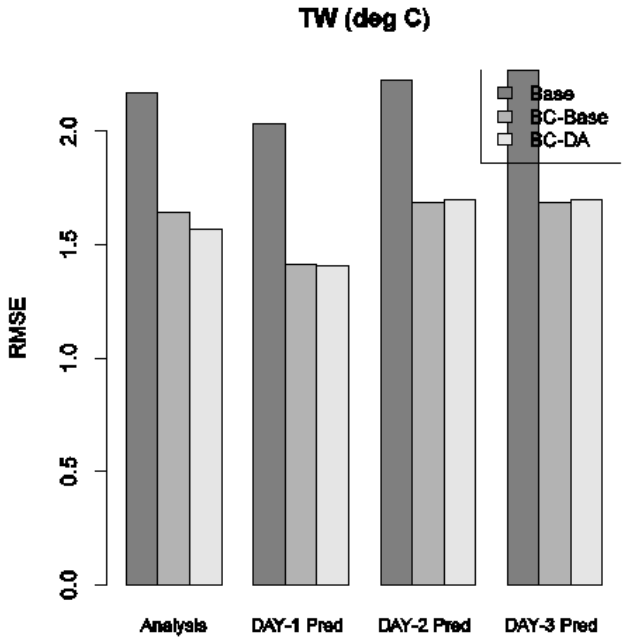

BOD (mg/l)
RMSE
0.0
0.2
0.4
0.6
0.8
1.0
Base
BC-Base
BC-DA
Analysis
DAY-1 Pred
DAY-2 Pred
DAY-3 Pred

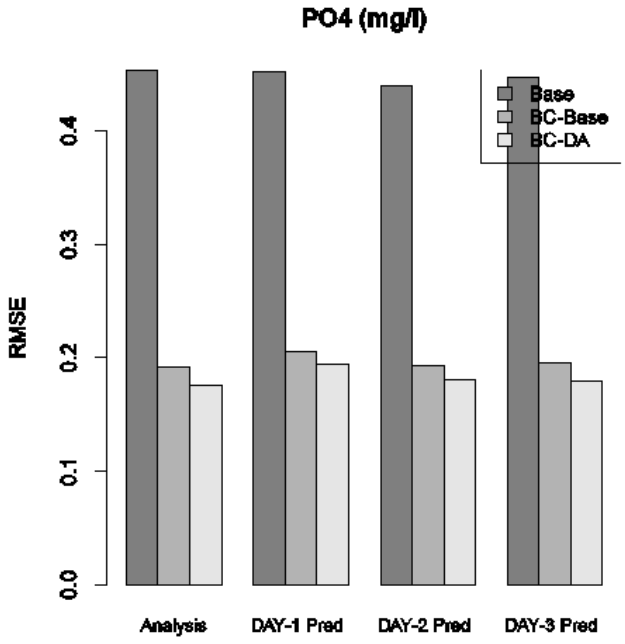

DO (mg/l)
RMSE
0.0
0.5
1.0
1.5
2.0
Base
BC-Base
BC-DA
Analysis
DAY-1 Pred
DAY-2 Pred
DAY-3 Pred

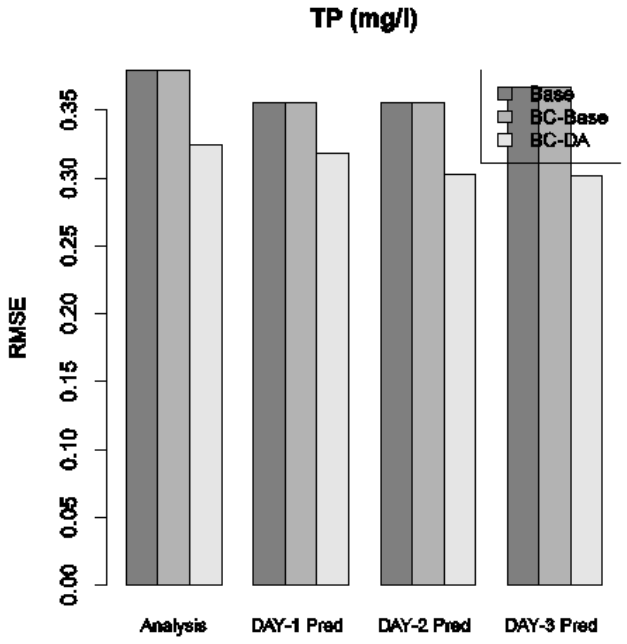

CHLA (ug/l)
RMSE
0
10
20
30
40
50
Base
BC-Base
BC-DA
Analysis
DAY-1 Pred
DAY-2 Pred
DAY-3 Pred

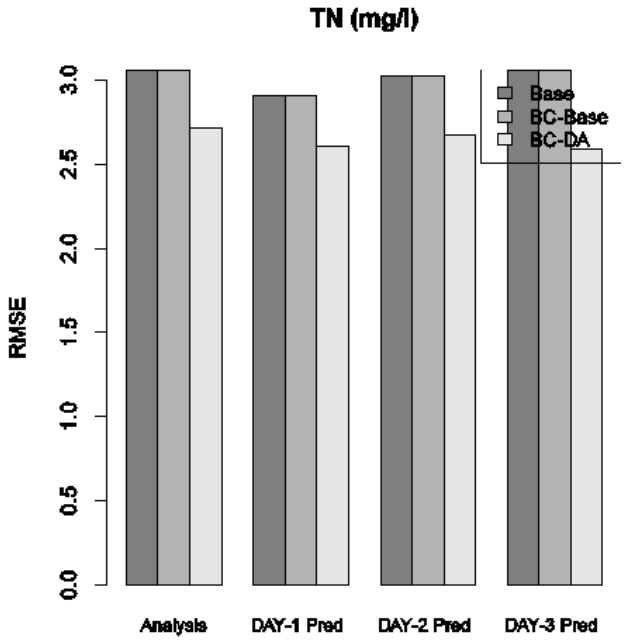

NH4 (mg/l)
RMSE
0.00
0.05
0.10
0.15
0.20
0.25
Base
BC-Base
BC-DA
Analysis
DAY-1 Pred
DAY-2 Pred
DAY-3 Pred

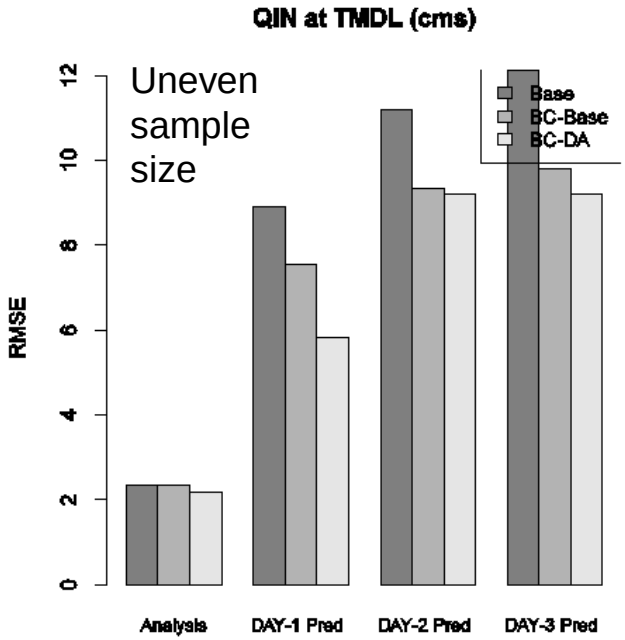

NO3 (mg/l)
RMSE
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Base
BC-Base
BC-DA
Analysis
DAY-1 Pred
DAY-2 Pred
DAY-3 Pred

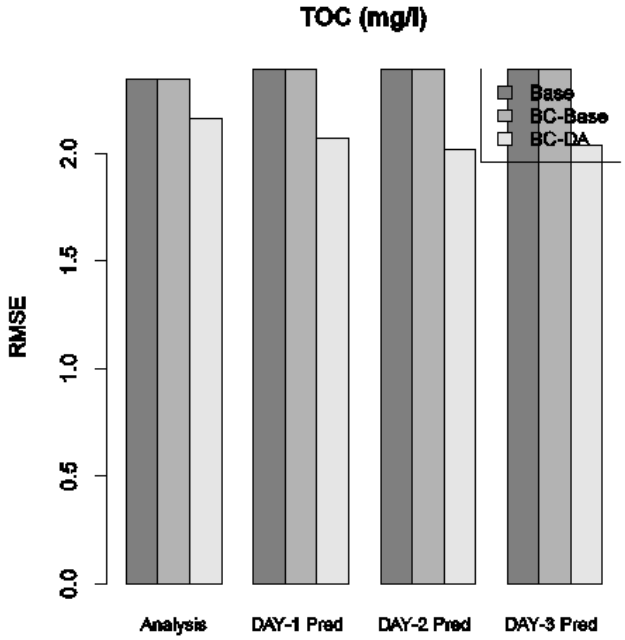

TW (deg C)
RMSE
0.0
0.5
1.0
1.5
2.0
Base
BC-Base
BC-DA
Analysis
DAY-1 Pred
DAY-2 Pred
DAY-3 Pred


PO4 (mg/l)
RMSE
0.0
0.1
0.2
0.3
0.4
Base
BC-Base
BC-DA
Analysis
DAY-1 Pred
DAY-2 Pred
DAY-3 Pred


TP (mg/l)
RMSE
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
Base
BC-Base
BC-DA
Analysis
DAY-1 Pred
DAY-2 Pred
DAY-3 Pred







- Significant to substantial improvement in accuracy in ~72 hr prediction for most variables
  - No improvement for NH_4
- Details to be reported in the Journal of Hydrology Special Issue on Ensemble Prediction and DA

Conclusions and ongoing work

- Real-time water quality forecasting is an increasingly important tool in water resources and hazard management
- For actionable decision making, accurate short-term water quality forecasts are essential
- Because most state variables in water quality models are not observed in reality, very large uncertainties exist in model ICs
- DA adds significant to substantial skill over DA-less (base) prediction by reducing the IC uncertainties
  - With the exception of NH_4

Conclusions and ongoing work (cont.)

- MLEF handles nonlinear observation equations very well
- Accounting for model (conditional) biases in the observation equation is important
- Ongoing work
  - Systematic sensitivity analysis
    - Improve performance
    - Assess possible under-determinedness of the inverse problem
  - Multi-catchment evaluation
  - Implementation in FEWS-NIER
- Ensemble DA $\in$ ensemble forecasting

THANK YOU

Q/A, Discussion

For more information, contact:

djseo@uta.edu

Acknowledgements

- Funding source
  - Water Quality Control Center, National Institute of Environmental Research (NIER), Incheon, Korea

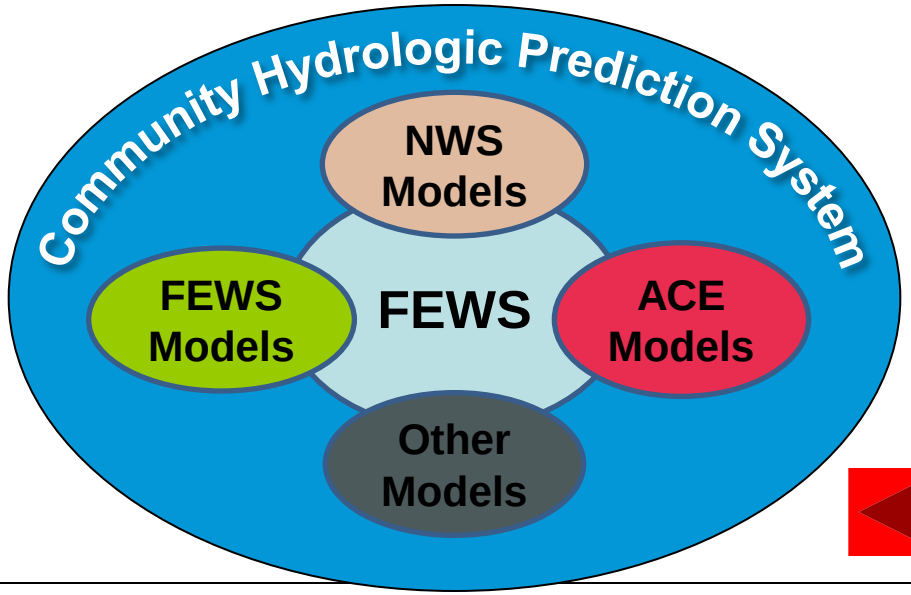
Community Hydrologic Prediction System (CHPS)

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

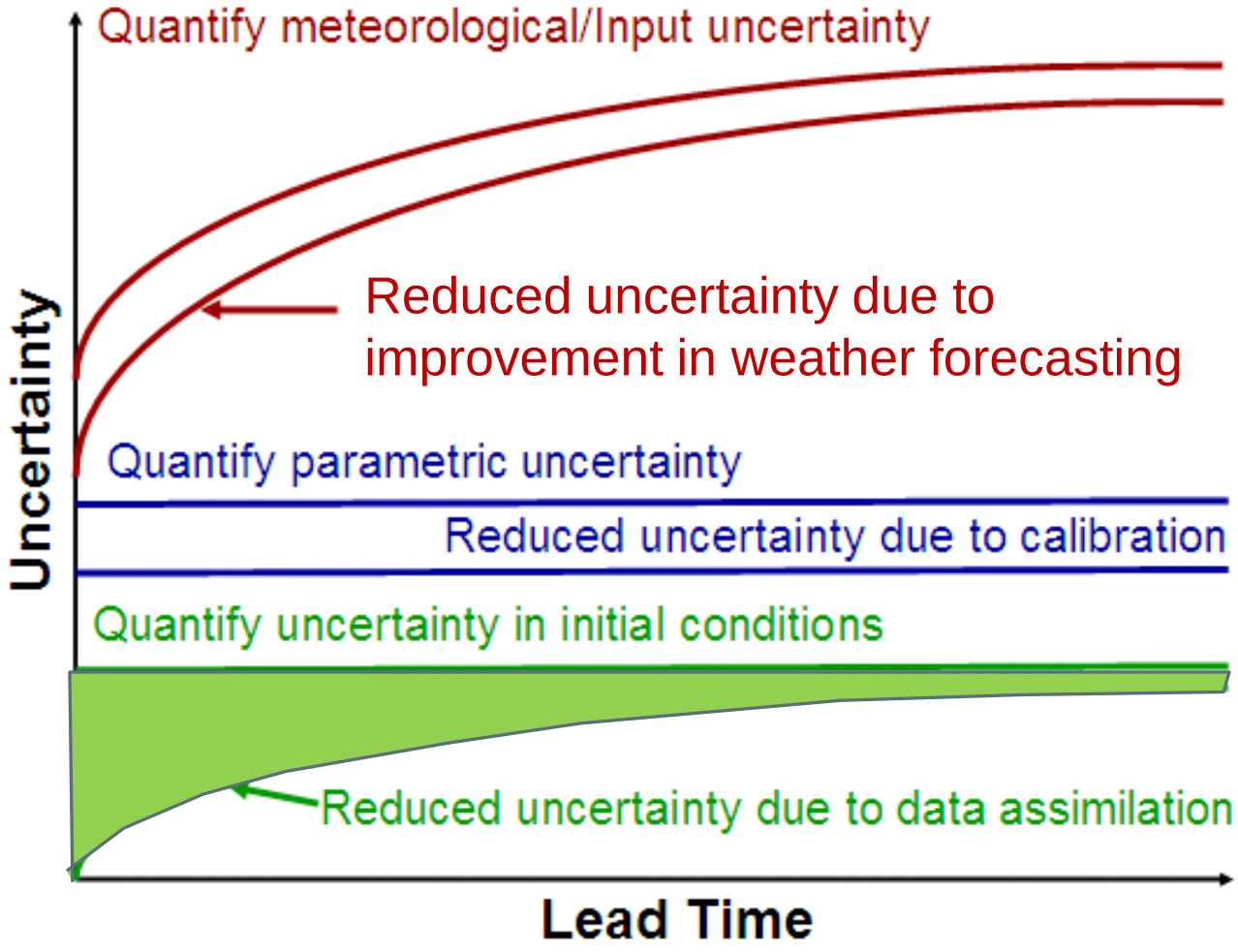
Implementation Status:

- ✓ 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)

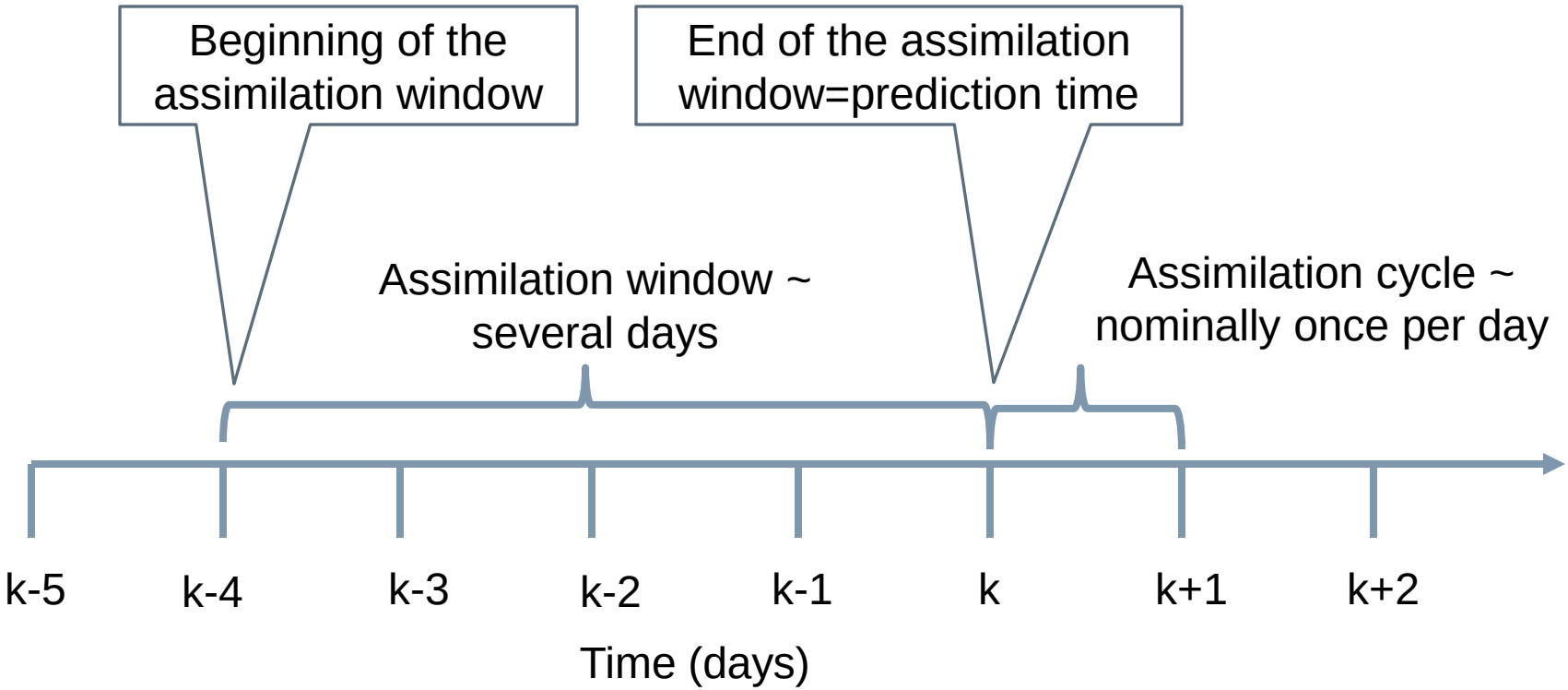
Uncertainties in real-time water quality forecasting



Structural uncertainty,

Flow regulations

Fixed-lag smoother formulation



All or part of the precipitation, PE, flow, stage, water quality data valid within the assimilation window is assimilated.

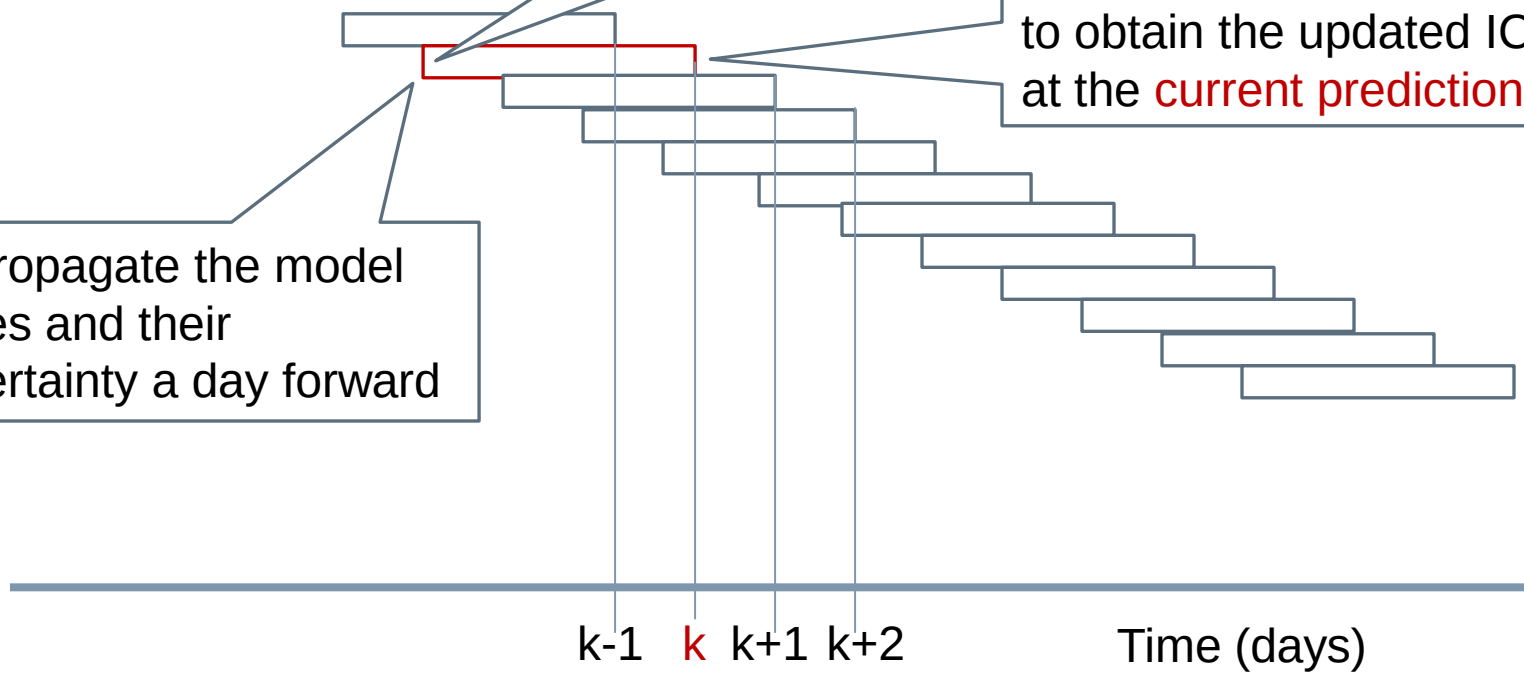
Fixed-lag smoother formulation (cont.)

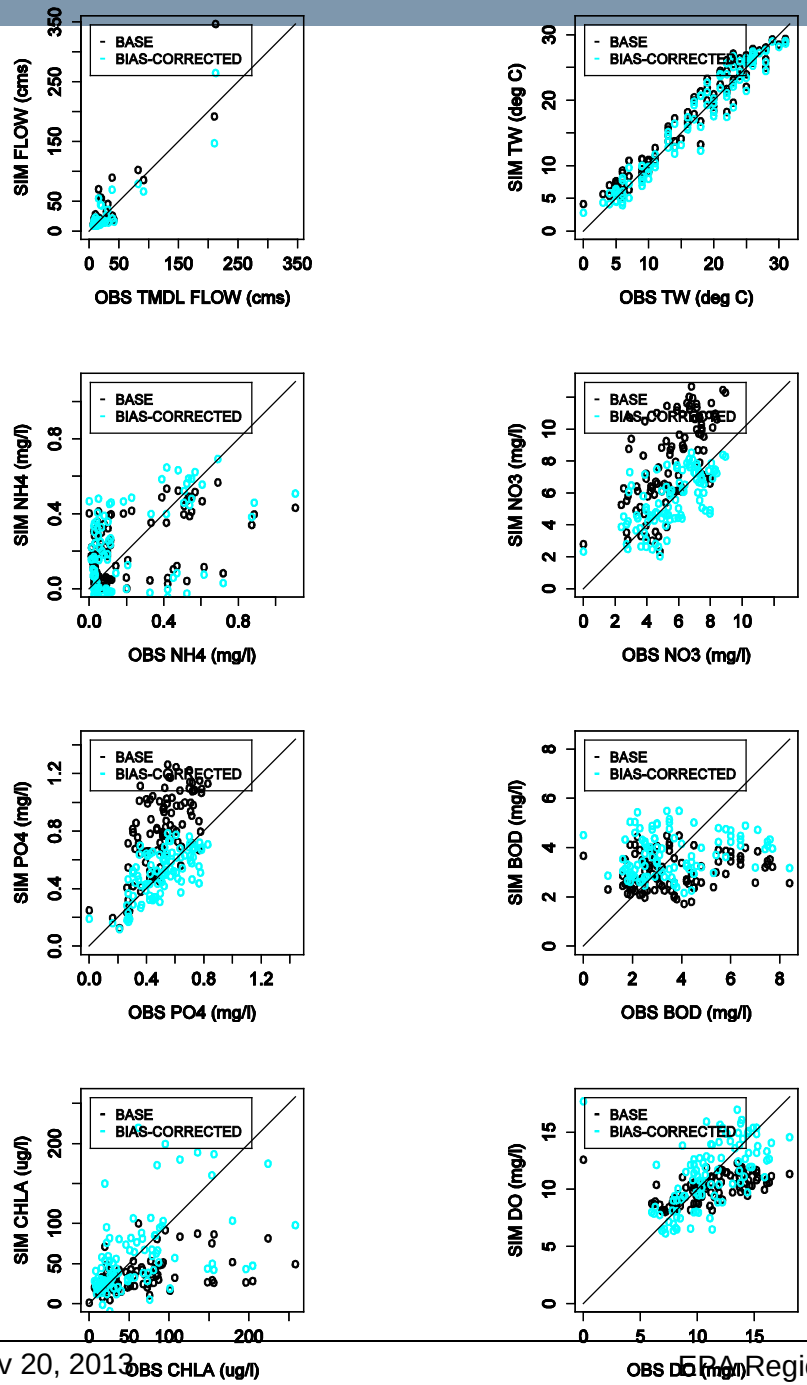
1) Prescribe the initial background model states and their covariance

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 a day forward





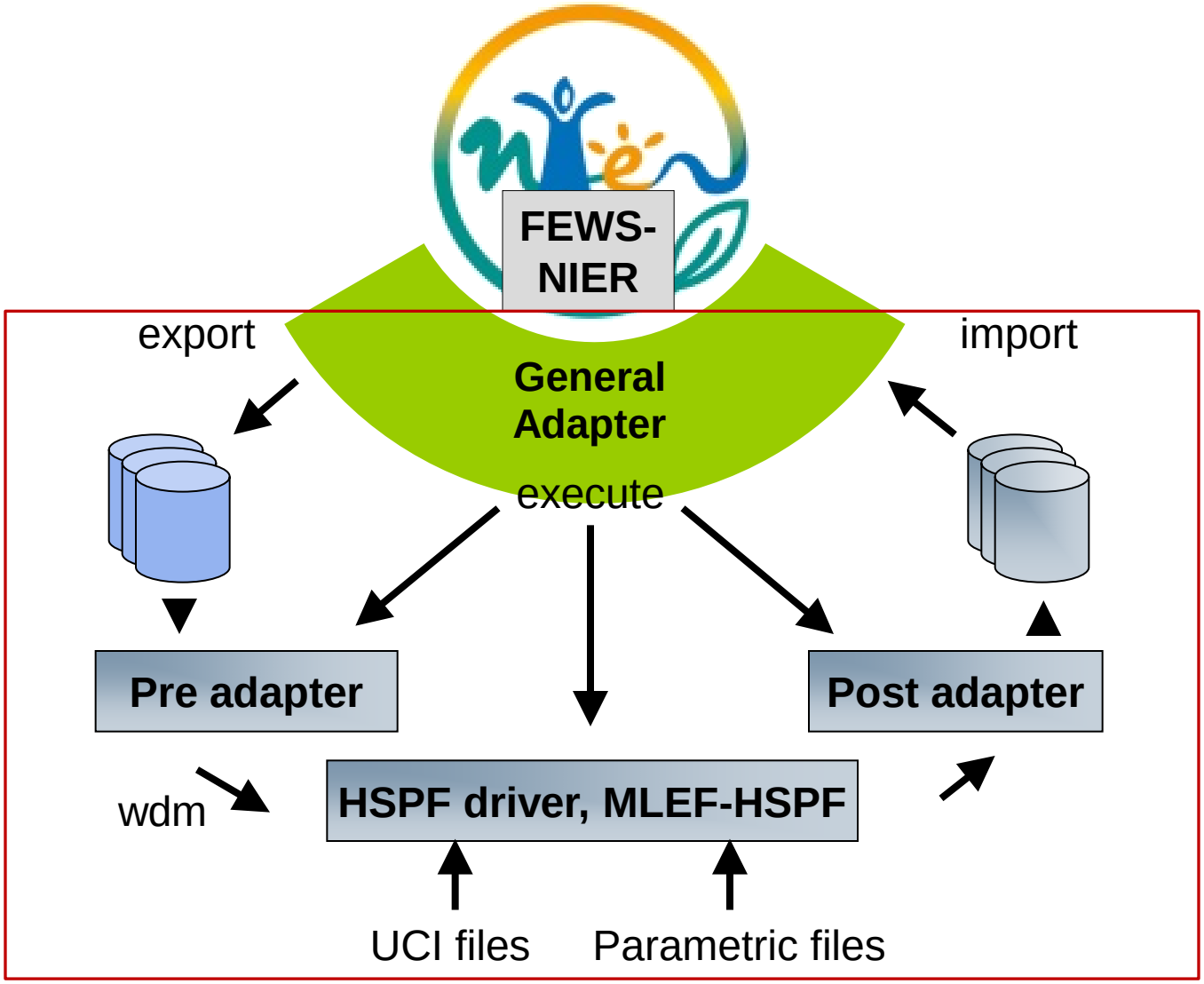
Bias correction

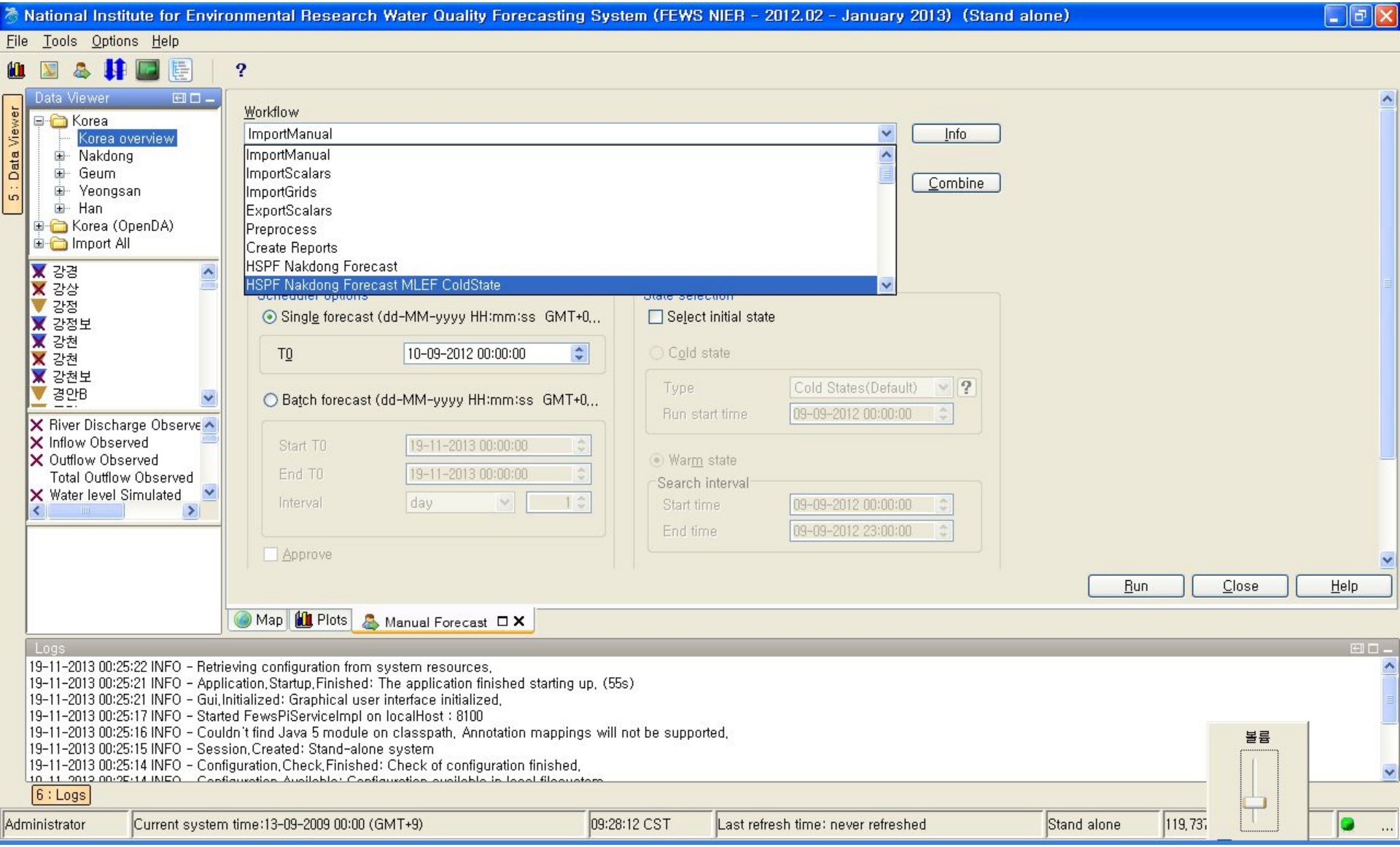
- Model simulation is usually conditionally biased
- Develop linear regression that removes conditional bias (Seo 2013)
- Apply bias correction as a post processor to model simulation (BC-Base)
- Incorporate bias correction in the observation equation of DA (BC-DA)

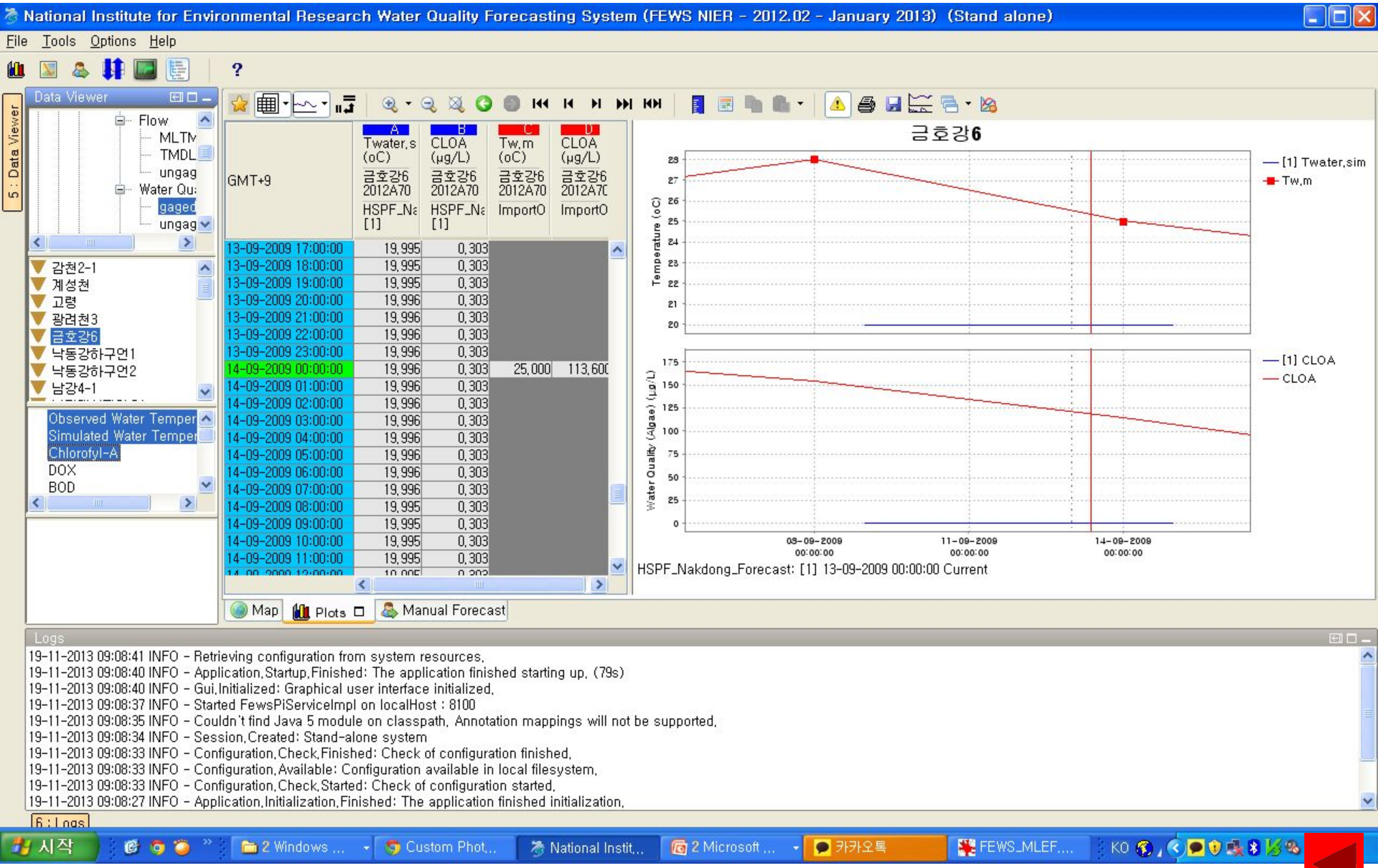
Evaluation of DA performance for selected catchments in hindcast mode

- Candidate catchments
  - **Kumho**, Andongdam, Banbyunchun, **Naeseongchun**, **Younggang**, Whichun, **Gamchun**, Hwang River, Nam River
- Evaluation
  - Compare HSPF prediction skills with and without DA at EFDC boundary condition locations

Implementation of MLEF-HSPF in FEWS-NIER




National Institute for Environmental Research Water Quality Forecasting System (FEWS NIER - 2012.02 - January 2013) (Stand alone)
File Tools Options Help
Data Viewer
5 : Data Viewer
Korea
Korea overview
Nakdong
Geum
Yeongsan
Han
Korea (OpenDA)
Import All
강경
강상
강정
강정보
강천
강천
강천보
경안B
River Discharge Observe
Inflow Observed
Outflow Observed
Total Outflow Observed
Water level Simulated
Workflow
ImportManual
ImportManual
ImportScalars
ImportGrids
ExportScalars
Preprocess
Create Reports
HSPF Nakdong Forecast
HSPF Nakdong Forecast MLEF ColdState
Info
Combine
Single forecast (dd-MM-yyyy HH:mm:ss GMT+0...
T0
10-09-2012 00:00:00
Batch forecast (dd-MM-yyyy HH:mm:ss GMT+0...
Start T0
19-11-2013 00:00:00
End T0
19-11-2013 00:00:00
Interval
day
1
Approve
Select initial state
Cold state
Type
Cold States(Default)
Run start time
09-09-2012 00:00:00
Warm state
Search interval
Start time
09-09-2012 00:00:00
End time
09-09-2012 23:00:00
Run
Close
Help
Map
Plots
Manual Forecast
Logs
19-11-2013 00:25:22 INFO - Retrieving configuration from system resources.
19-11-2013 00:25:21 INFO - Application.Startup.Finished: The application finished starting up. (55s)
19-11-2013 00:25:21 INFO - Gui.Initialized: Graphical user interface initialized.
19-11-2013 00:25:17 INFO - Started FewsPiServiceImpl on localHost : 8100
19-11-2013 00:25:16 INFO - Couldn't find Java 5 module on classpath. Annotation mappings will not be supported.
19-11-2013 00:25:15 INFO - Session.Created: Stand-alone system
19-11-2013 00:25:14 INFO - Configuration.Check.Finished: Check of configuration finished.
6 : Logs
볼륨
Administrator
Current system time:13-09-2009 00:00 (GMT+9)
09:28:12 CST
Last refresh time: never refreshed
Stand alone


National Institute for Environmental Research Water Quality Forecasting System (FEWS NIER - 2012.02 - January 2013) (Stand alone)
File Tools Options Help
Data Viewer
Flow
MLTM
TMDL
ungag
Water Qu
gaged
ungag
감천2-1
계성천
고령
광려천3
금호강6
낙동강하구언1
낙동강하구언2
남강4-1
Observed Water Temper
Simulated Water Temper
Chlorofyl-A
DOX
BOD
GMT+9
A Twater,s (oC) 금호강6 2012A70 HSPF_Na [1]
B CLOA (μg/L) 금호강6 2012A70 HSPF_Na [1]
C Tw,m (oC) 금호강6 2012A70 ImportO
D CLOA (μg/L) 금호강6 2012A7(ImportO
13-09-2009 17:00:00 19,995 0,303
13-09-2009 18:00:00 19,995 0,303
13-09-2009 19:00:00 19,995 0,303
13-09-2009 20:00:00 19,996 0,303
13-09-2009 21:00:00 19,996 0,303
13-09-2009 22:00:00 19,996 0,303
13-09-2009 23:00:00 19,996 0,303
14-09-2009 00:00:00 19,996 0,303 25,000 113,60(
14-09-2009 01:00:00 19,996 0,303
14-09-2009 02:00:00 19,996 0,303
14-09-2009 03:00:00 19,996 0,303
14-09-2009 04:00:00 19,996 0,303
14-09-2009 05:00:00 19,996 0,303
14-09-2009 06:00:00 19,996 0,303
14-09-2009 07:00:00 19,996 0,303
14-09-2009 08:00:00 19,995 0,303
14-09-2009 09:00:00 19,995 0,303
14-09-2009 10:00:00 19,995 0,303
14-09-2009 11:00:00 19,995 0,303
금호강6
Temperature (oC)
[1] Twater,sim
Tw,m
Water Quality (Algae) (μg/L)
[1] CLOA
CLOA
09-09-2009 00:00:00
11-09-2009 00:00:00
14-09-2009 00:00:00
HSPF_Nakdong_Forecast: [1] 13-09-2009 00:00:00 Current
Map
Plots
Manual Forecast
Logs
19-11-2013 09:08:41 INFO - Retrieving configuration from system resources.
19-11-2013 09:08:40 INFO - Application.Startup.Finished: The application finished starting up. (79s)
19-11-2013 09:08:40 INFO - Gui.Initialized: Graphical user interface initialized.
19-11-2013 09:08:37 INFO - Started FewsPiServiceImpl on localHost : 8100
19-11-2013 09:08:35 INFO - Couldn't find Java 5 module on classpath. Annotation mappings will not be supported.
19-11-2013 09:08:34 INFO - Session.Created: Stand-alone system
19-11-2013 09:08:33 INFO - Configuration.Check.Finished: Check of configuration finished.
19-11-2013 09:08:33 INFO - Configuration.Available: Configuration available in local filesystem.
19-11-2013 09:08:33 INFO - Configuration.Check.Started: Check of configuration started.
19-11-2013 09:08:27 INFO - Application.Initialization.Finished: The application finished initialization.
시작
2 Windows ...
Custom Phot...
National Instit...
2 Microsoft ...
카카오톡
FEWS_MLEF...
KO

Need for ensemble forecasting and verification

In 2006, National Research Council recommended that NWS produce **uncertainty-quantified products, expand verification** and make information easily available to all users in near real time.

- **Forecasters** get objective guidance for **level of confidence** in forecasts
- **End users** decide whether to take action based on their **risk tolerance**

THE SCIENCE OF NOAA'S OPERATIONAL HYDROLOGIC ENSEMBLE FORECAST SERVICE

BY JULIE DEMARGNE, LIMIN WU, SATISH K. REGONDA, JAMES D. BROWN, HAKSU LEE, MINXUE HE, DONG-JUN SEO, ROBERT HARTMAN, HENRY D. HERR, MARK FRESCH, JOHN SCHAAKE, AND YUEJIAN ZHU

HEFS extends hydrologic ensemble services from 6-hour to year-ahead forecasts and includes additional weather and climate information as well as improved quantification of major uncertainties.

- To appear in the Jan 2014 issue of the Bulletin of the American Meteorological Society (BAMS)
- Available online at http://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-12-00081.1