Ensemble data assimilation for operational water quality forecasting

Sunghee Kim[1], Hamideh Riazi[1], Dong-Jun Seo[1], Changmin Shin[2]
[1]Department of Civil Engineering, The University of Texas at Arlington, Arlington, TX, USA
[2]Water Quality Control Center, National Institute of Environmental Research, Inchon, Korea

Abstract

Due to large dimensionality of the state vector and sparsity of observations, the initial conditions (IC) of watershed water quality models are subject to large uncertainty. To improve forecast accuracy and to quantify uncertainty in the ICs, an ensemble data assimilation (DA) procedure has been developed for the Hydrologic Simulation Program – Fortran (HSPF). To utilize all available hydrologic and water quality observations, it is important that the DA technique be able to handle strong nonlinearity in hydrologic and biochemical observation equations in addition to nonlinear model dynamics. The procedure, which is being implemented for operational use at the Water Quality Control Center of the National Institute of Environmental Research in Korea, uses maximum likelihood ensemble filter (MLEF) which combines strengths of variational assimilation (VAR) and ensemble Kalman filter (EnKF). The observations assimilated are water temperature, dissolved oxygen (DO), biochemical oxygen demand (BOD), ammonium (NH_4), nitrate (NO_3), phosphate (PO_4), chlorophyll-a (CHL-a) and streamflow. Verification of the control results has been carried out for the Kumho Catchment in the Nakdong River Basin in Korea (Kim et al. 2014). Verification of the ensemble results is currently under way using the Ensemble Verification System (EVS, Brown et al. 2010).

Objective

- Many sources of error exist in the end-to-end water quality forecast process
- Reducing uncertainty in the initial conditions (IC) is a cost-effective way of improving forecast accuracy
  - Current practice: Manual updating of ICs based on real-time observations
  - Need: Automatic updating using DA
- Issues:
  - Nonlinear hydrologic and biochemical observation equations
  - A large number of state variables, sparse observations

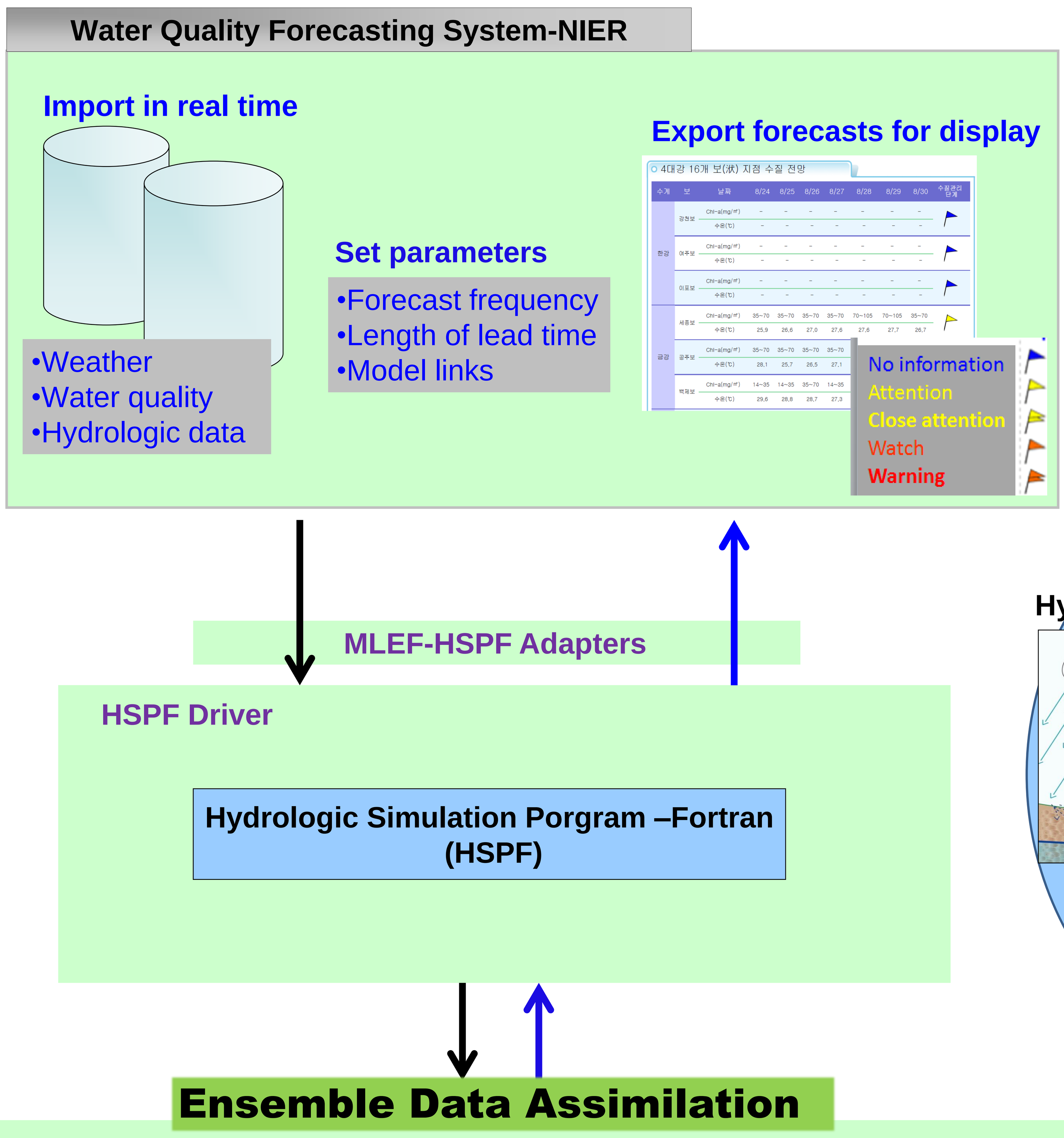
(e.g. 333 state variables using 28 HSPF state variables for 31 model segments for the Kumho Catchment in the Nakdong River Basin)

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)
	ORP	No	Dead organic refractory phosphorus
	ORN	No	Dead organic refractory nitrogen
	ORC	No	Dead organic refractory carbon

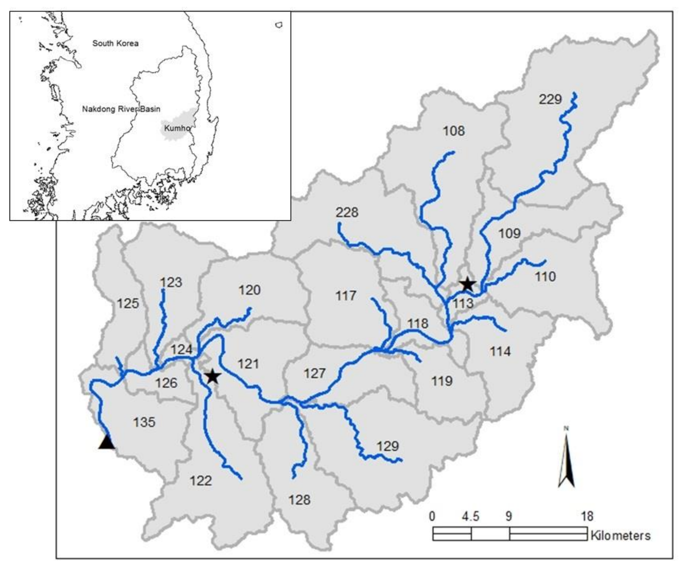
HSPF state variables

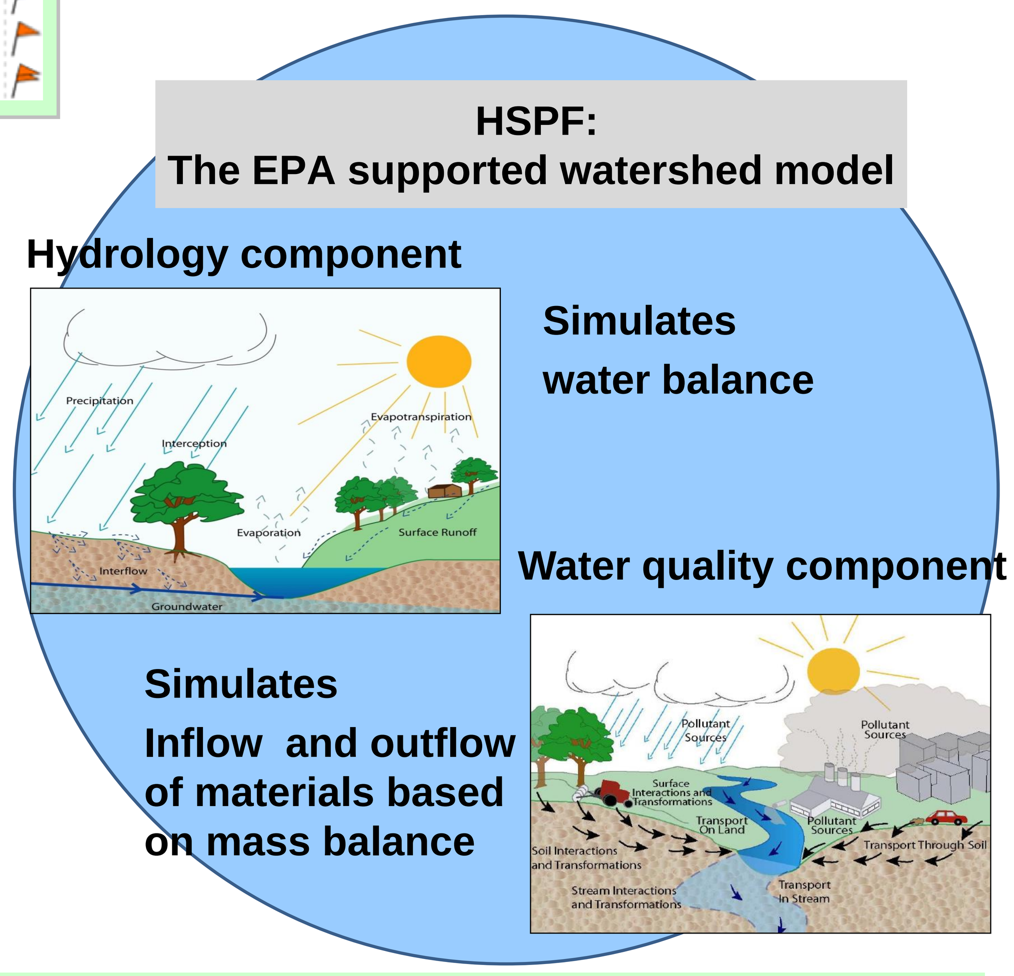
Most HSPF state variables are not observed

Operational Water Quality forecasting



Study Area (2110 km²)





Maximum Likelihood Ensemble Filter (MLEF)

- Zupanski (2005)
- Ensemble extension of variational assimilation, but, no adjoint necessary
- Can handle nonlinear model dynamics
- Can handle nonlinear observation equation
- Algorithmically somewhat complex
- Minimizes cost function in ensemble space

Square root forecast error covariance

$$P_f^{1/2} = [p_f^1 \ldots p_f^S]$$

$$p_f^i = M(x + p_a^i) - M(x)$$

Square root analysis error covariance

$$P_a^{1/2} = [p_a^1 \ldots p_a^S] = P_f^{1/2}(I + C)^{-1/2}$$

$$C = \begin{bmatrix} z_1^T z_1 \ldots z_1^T z_S \\ \cdot \\ \cdot \\ z_S^T z_1 \ldots z_S^T z_S \end{bmatrix}$$

$$z^i \approx R^{-1/2}H(x + p_f^i) - R^{-1/2}H(x)$$

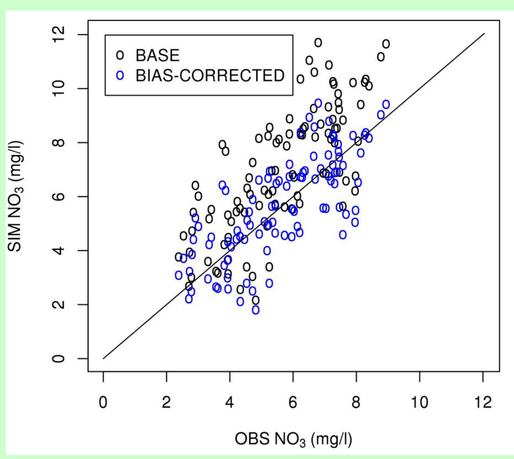
$$p_f^i = \begin{pmatrix} p_f^1 \\ p_f^2 \\ \vdots \\ p_f^N \end{pmatrix} \qquad p_a^i = \begin{pmatrix} p_a^1 \\ p_a^2 \\ \vdots \\ p_a^N \end{pmatrix}$$

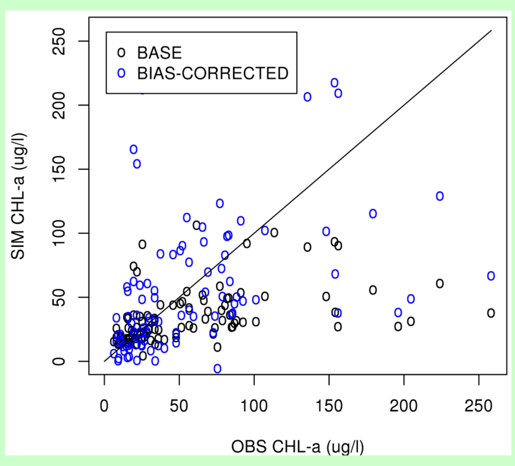
S :Ensemble size $i = 1, \ldots, S$
N :Size of control vector in physical space
M( ): Dynamical model
H( ): Nonlinear observation operator
R : Observation error covariance

Bias Correction

Apply conditional bias-penalized linear regression to correct bias (Seo 2013)
Apply bias correction as a post processor to model prediction (BC-Base)
Incorporate bias correction in the observation equation of DA (BC-DA)

$$J = E_{Z,Y}[(Z - Y)^2] + \alpha E_Y[\{E_Z[Z \mid Y] - Y\}^2]$$

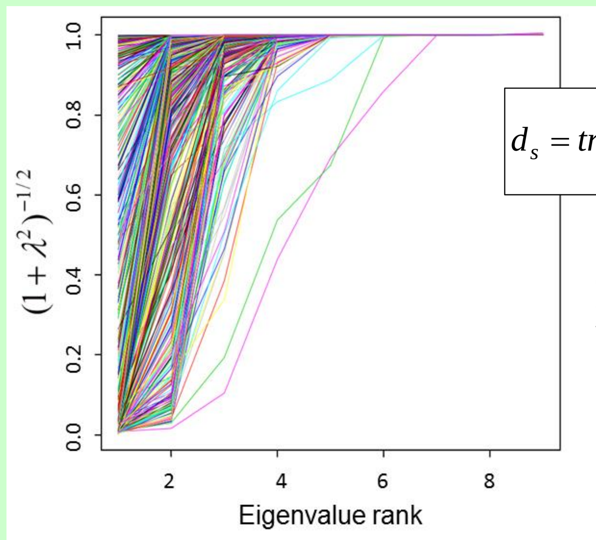




- NO_3 (left): the bias correction procedure reduces systematic first-order conditional bias effectively
- CHL-a (right): heteroscedastic errors still exist

DA Parameters

Assimilation cycle	Once per day
Assimilation windows	7 days
Lead time	72 hours
Control variables	333
Ensemble Size	9



$$d_s = tr[\{I + C(x)\}^{-1}C(x)] = \sum_i \frac{\lambda_i^2}{1+\lambda_i^2}$$

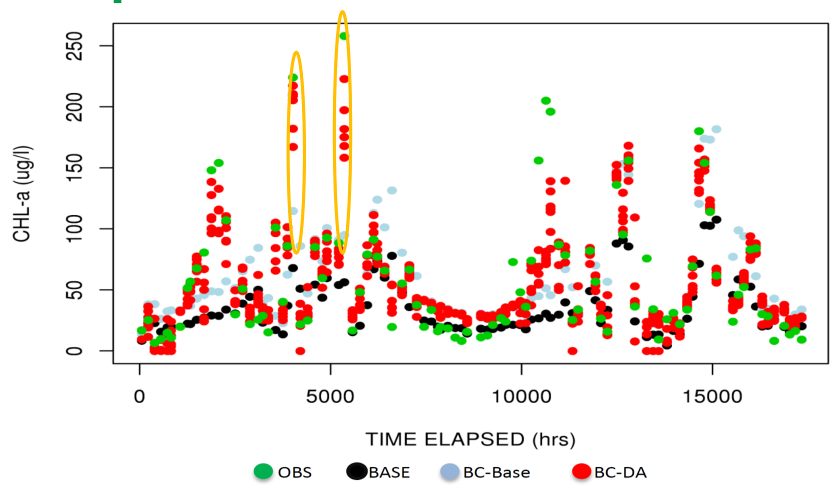
d_s: Degrees of freedom for signal
λ_i^2 : the i-th eigenvalue of C(x)

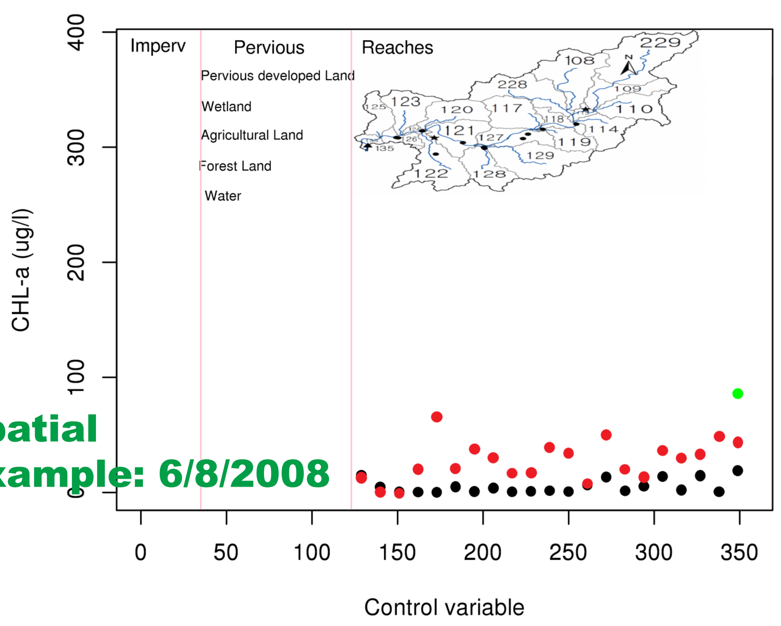
Eigenvalue spectra of $[I + C(x_{opt})]^{-1/2}$ for all assimilation cycles in 2008 (above) indicates that about 7 ensemble members capture all information in all assimilation cycles

Results (NO_3, CHL-a as examples)

Analysis

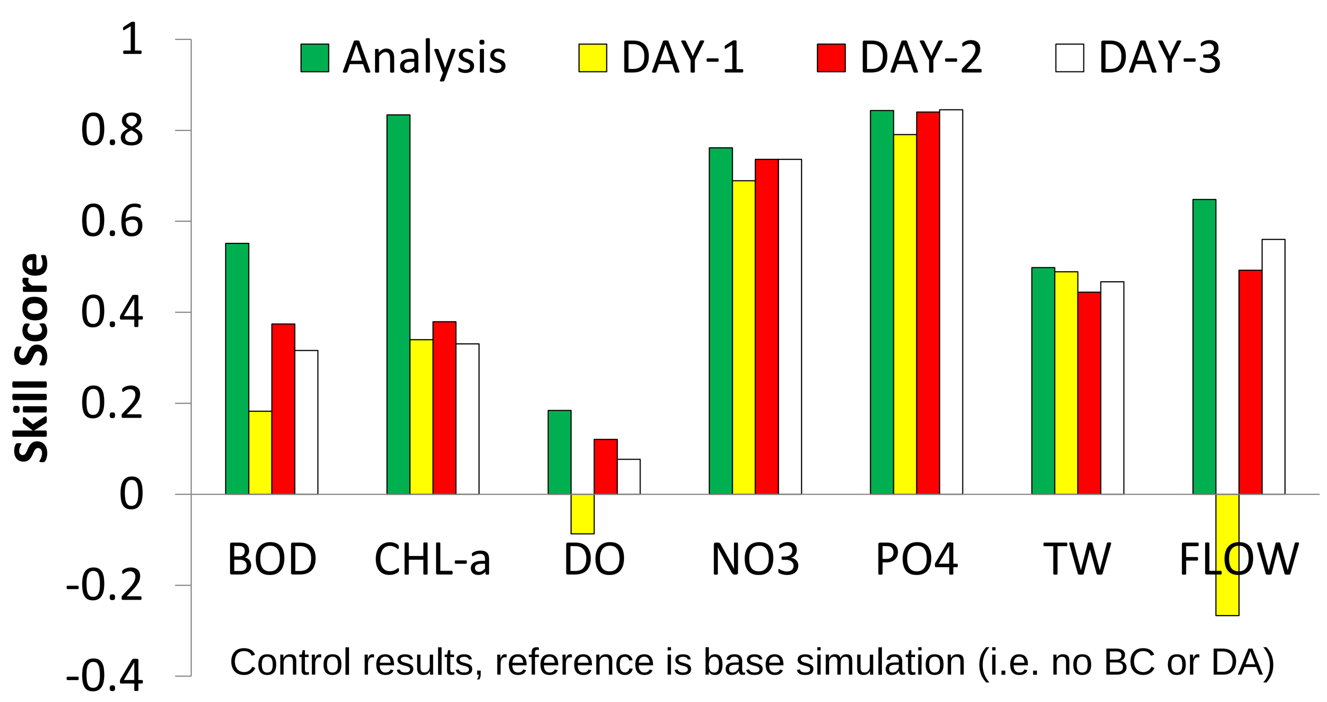
Temporal





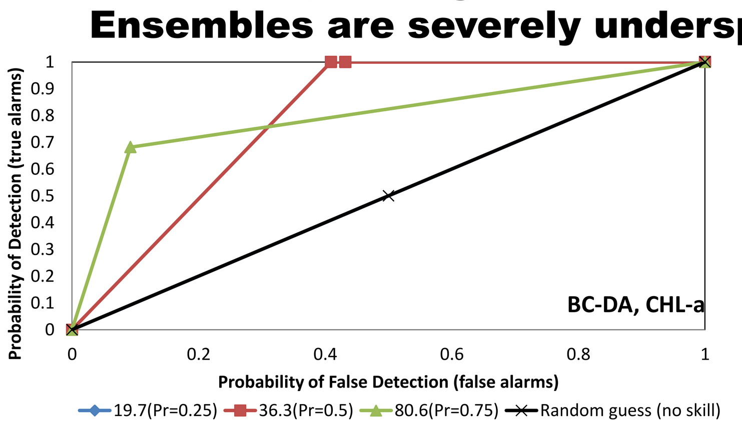
Spatial Example: 6/8/2008

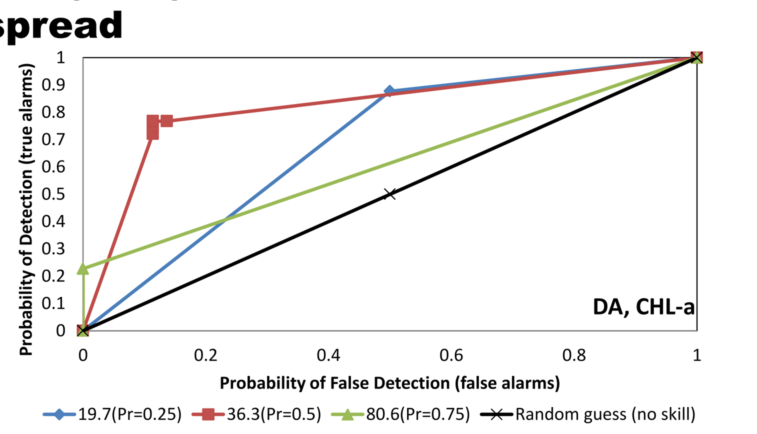
Forecast Verification



Control results, reference is base simulation (i.e. no BC or DA)

Relative Operating Characteristic (ROC) at lead time of 72 hr – Ensembles are severely underspread





Conclusions

- The DA procedure adds substantial predictive skill for all observed variables except DO. Reduction in root mean square error (RMSE) ranges from 11 to 60% for Day-1 through 3 predictions. The reduction is the largest for NO_3 and PO_4 at about 47 and 59%, respectively, owing largely to the bias correction component of DA. The second largest reduction is for TW at about 25%.
- Correction of model bias as part of the observation equation is important.
- MLEF handles nonlinear observation equations very well.

Ongoing work

- Address underspread
- Additional evaluation using multiple catchments
- Comparison of DA performance using observations from interior monitoring stations

References

Brown J.D., Demargne J., Seo D-J., and Liu Y. (2010) The Ensemble Verification System (EVS): a software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations. *Environmental Modelling and Software*, 25(7), 854-872.
Kim, S., D.-J. Seo, H. Riazi and C. Shin, Improving water quality forecasting via data assimilation – Application of maximum likelihood ensemble filter to HSPF, submitted to Journal of Hydrology.
Seo, D-J. 2013. Conditional bias-penalized kriging, Stochastic Environmental Research and Risk Assessment, DOI 10.1007/s00477-012-0567-z.
Zupanski, M. 2005. Maximum likelihood ensemble filter: Theoretical aspects. Mon. Weather Rev. 133, 1710–1720.

Acknowledgements

The work reported herein is supported by the National Institute of Environmental Research, the Republic of Korea, under the Agreement of the Cooperative Study between Geosystem Research Corporation, Korea, and The University of Texas at Arlington. This support is gratefully acknowledged.`