DATA ASSIMILATION IN ENSEMBLE WATER FORECASTING - CHALLENGES AND OPPORTUNITIES

D.-J. Seo¹, Haksu Lee²,³, Yuqiong Liu⁴,⁵, Victor Koren²,⁹, Arezoo Rafieei Nasab¹, Hamideh Riazi¹ Sunghee Kim¹, Changmin Shin⁶, Ridwan Siddique¹,⁸, Yu Zhang², Dongsoo Kim⁷, Beomgeun Kim¹

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

- Quantify meteorological/Input uncertainty
- Quantify parametric uncertainty
- Quantify uncertainty in initial conditions

Reduced uncertainty due to:
- Pre-processing
- Calibration
- Data assimilation

Lead Time

Structural uncertainty
Residual uncertainty

Flow regulations: A large challenge

Ensemble pre-processor
Parametric uncertainty processor
Data assimilator
Ensemble post-processor, multimodel ensemble
Elements of Hydrologic Ensemble Prediction System

QPE, QTE, Soil Moisture

QPF, QTF

Ensemble Pre-Processor

Hydrology & Water Resources Models

Ensemble Post-Processor

Hydrology & Water Resources Models

Hydrology & Water Resources Ensemble Product Generator

Data Assimilator

Streamflow

Input Uncertainty Processor

Parametric Uncertainty Processor

Hydrologic Uncertainty Processor

HEFS

Ensemble Verification 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

- Snow Update System user manual, Version 2.00.02 Updated 2003-09-26, Riverside Technology, Inc.
# NWSRFS Operations in use (as of October 2007)

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

48 Hour Precipitation Ending
11/23/2004 12Z
West Gulf River Forecast Center
Fort Worth, Texas

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

- OBSERVED FLOW
- SIMULATED FLOW
- SIMULATED FLOW AFTER VARIATIONAL ASSIMILATION (VAR)
- ENSEMBLE OF SIMULATED FLOWS AFTER VAR, BUT BEFORE ENSEMBLE DA

FLOW (CMS)

ABRFC / WTTO2

WTTO2 Channel Network

TIME ELAPSED (HRS)
Parameter estimation/optimization of distributed hydrologic routing model

From Seo et al. (2003)
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.opendata.org/joomla/index.php

Adapted from Deltares (2012)
Operational hydrologic data assimilation - Strategy

MODIS-derived snow cover
AMSR-derived SWE
MODIS-derived surface temperature
MODIS-derived cloud cover
AMSR-derived SM

Atmospheric forcing
Snow models
Snowmelt
Potential evap. (PE)
Precipitation
Soil moisture accounting models
Runoff
Hydrologic routing models

In-situ snow water equivalent (SWE)
SNODAS SWE
In-situ soil moisture (SM)
Streamflow or stage

Satellite altimetry

Flow
Hydraulic models
River flow or stage
Flow
reservoir, etc., models

Adapted from OHD Strategic Science Plan 2010
DA strategy for operational hydrologic forecasting

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

- To illustrate, decompose:

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

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

From Schweppe (1973)
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)

\[ \mathcal{f}(N + 1 | N + 1) = \Sigma(N + 1 | N + 1) \]
\[ \times \{ H'(N + 1)R^{-1}(N + 1)z(N + 1) \]
\[ + \Sigma^{-1}(N + 1 | N)\Phi(N)\mathcal{f}(N | N) \} \]
\[ \Sigma(N + 1 | N) = \Phi(N)\Sigma(N | N)\Phi'(N) + G(N)Q(N)G'(N) \]
\[ \Sigma(N + 1 | N + 1) = [H'(N + 1)R^{-1}(N + 1)H(N + 1) \]
\[ + \Sigma^{-1}(N + 1 | N)]^{-1} \]
\[ \Sigma(0 | 0) = \psi \]
\[ \mathcal{f}(0 | 0) = 0 \]

From Schwepppe (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

Prescribe the initial background error covariance of model states

Solve for ICs, $\beta_p$, $\beta_{PE}$ and time-varying model errors under MSE minimization

Propagate forward the uncertainty in ICs by an hour using the model errors inferred from the preceding time step

Time
Multi-Scale Bias Correction (MSBC)

Assume nothing is known about the model states

- Solve for ICs, $\beta_P$, $\beta_{PE}$ and model runoff under a mass balance-only objective function
- Solve for $\beta_P$, $\beta_{PE}$ and model runoff under a mass balance-only objective function
- Solve for $\beta_P$, $\beta_{PE}$ and model runoff under a mass balance-only objective function
- Solve for $\beta_P$, $\beta_{PE}$ and model runoff under a mass balance-only objective function

Time
UNIT HYDROGRAPH (CMS/MM)

TIME (HRS)

hbmt2

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
Effect of timing errors in updating of soil water states

\[ \text{\(T_p=13\) (hrs)} \]

\[ \text{\(T_p=10\) (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

\[ x(n + 1) = \Phi(n)x(n) + G(n)w(n) \]
\[ z(n) = H(n)x(n) + v(n), \quad n = 1, \ldots \]

- \( x(n) \): state, a \( K_1 \) vector
- \( z(n) \): observation, a \( K_2 \) vector
- \( v(n) \): white observation uncertainty, a \( K_2 \) vector
- \( w(n) \): white system driving uncertainty, a \( K_3 \) vector
- \( x(0) \): 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

MTPT2

RMSE (m³/sec)

Lead Time (hrs)

SIMULATED
MLEF: CONTROL
MLEF: ENSEMBLE MEAN
ENKF: ENSEMBLE MEAN

S=5
S=9
S=30
S=50
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 $x(N)$ and $w(n)$, $n = 0, \ldots, N - 1$, which minimize

$$J[x(N), w(0) \cdots w(N - 1)]$$

$$= \sum_{n=1}^{N} [z(n) - H(n)x(n)]'R^{-1}(n)[z(n) - H(n)x(n)]$$

$$+ \sum_{n=0}^{N-1} w'(n)Q^{-1}(n)w(n) + x'(0)\psi^{-1}x(0)$$

subject to the constraint that

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

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

Let $\hat{x}(N|N)$ denote the resulting value of $x(N)$. If one actually performs the minimization using Lagrange multipliers, the resulting equations for $\hat{x}(N|N)$ 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 \[ E[X | \hat{X}] - \hat{X} \] (analogous to reliability)

- Type-II CB \[ E[\hat{X} | X] - X \] (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

\[ z = Hx + \nu \]

\[ x: \text{completely unknown} \]

\[ E\{\nu\} = 0 \]

\[ E\{\nu\nu'\} = R \]

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

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

*From Schweppe (1973)*
“Fisher-like” solution for minimizing
\[ J = \sum_{EV} + \alpha \sum_{CB}, \ \alpha \geq 0 \]

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

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

where

\[ \hat{H}^T = H^T + \alpha \Psi_{xx}^{-1} \Psi_{xz} \]

\[ \Lambda = R + \alpha (1 - \alpha) \Psi_{xx}^{-1} \Psi_{xz} - \alpha H \Psi_{xz} - \alpha \Psi_{zx} H^T \]

\[ B = \alpha \Psi_{xx} \hat{H}^T \Lambda^{-1} \hat{H} + (1 + \alpha) I \]

*From Seo (2013), Seo et al. (2014)*
OK/KF estimates vs. truth

From Seo (2013)
CBPK/CBPKF estimates vs. truth

(a) COR=8 KM NG=18000
(b) COR=8 KM NG=18000
(c) COR=16 KM NG=4000
(d) COR=16 KM NG=4000
(e) COR=32 KM NG=1000
(f) COR=32 KM NG=1000
(g) COR=128 KM NG=2000
(h) COR=128 KM NG=2000
From Seo et al. (2014)
Improving estimation of heavy-to-extreme precipitation
Improving estimation of heavy-to-extreme precipitation
The HLRDHM domain encompassing Fort Worth, Arlington and Grand Prairie.

Overlaid is the 500x500 m² CASA QPE grid.

Dallas-Fort Worth Metroplex is the 4th largest (~6.5 mil) and fastest-growing metropolitan area in the US.
Hydraulic modeling using radar rainfall data

CASA QPE (1 min, ~500 m)

HLRDHM surface runoff (~250 m)

1D-2D hydraulic modeling

HLRDHM channel flow (~250 m)
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

**Users and stakeholders**
- Cities
- Consultants
- Dallas Util.
- EPA
- FEMA
- NCTCOG
- NWS
- TRA
- TRWD
- USACE
- USGS

**Advanced sensing**
- High-resolution QPI
- Forecasts, projections
- Wireless sensor observations
- Existing observations
- Crowdsourced observations

**Advanced computing**
- Historical data
- Parameter est./opt.
- Models
- Initial conditions
- Data assimilation
- QPE fusion
- Real-time data

**Data-enabled discovery**
- Statistical modeling, causal inference
- Verification
- Non-real time

**Urban sustainability applications**
- Ensemble prediction framework

**Decision support**
- Model output

**Education and workforce development**

- Real-time data

**Ensemble prediction framework**
- Flash flood forecasting
- Inundation mapping
- Water quality forecasting
- Stormwater management
- Urbanization impact assessment
- Climate change impact assessment
- Climate change adaptation
- Life cycle analysis
- LID
- SUDS
CAHMDA-DAFOH

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

<table>
<thead>
<tr>
<th>Event</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract submission Deadline</td>
<td>June 30, 2014 (extended)</td>
</tr>
<tr>
<td>Early Registration Deadline</td>
<td>June 30, 2014</td>
</tr>
<tr>
<td>Training Course</td>
<td>Sep 06-07, 2014</td>
</tr>
<tr>
<td>Joint CAHMDA-VI and HEPEX-DAFOH III Workshop</td>
<td>Sep 08-12, 2014</td>
</tr>
<tr>
<td>Excursions and Activities</td>
<td>Sep 13, 2014</td>
</tr>
</tbody>
</table>

*Catchment-based Hydrological Model Data Assimilation (CAHMDA VI) and Hydrologic Ensemble Prediction Experiment (HEPEX-DAFOH III) Joint workshop 8–12 September, 2014, Austin, Texas, USA*
THANK YOU

For more information, contact:
djseo@uta.edu