A NEW (OLD?) APPROACH IN DATA ASSIMILATION FOR OPERATIONAL HYDROLOGIC FORECASTING

D.-J. Seo\textsuperscript{1}, Arezoo Rafieei Nasab\textsuperscript{1}, Haksu Lee\textsuperscript{2,3}, Sunghee Kim\textsuperscript{4} and Victor Koren\textsuperscript{5}

\textsuperscript{1}Dept. Of Civil Engineering, The University of Texas at Arlington, Arlington, TX, USA
\textsuperscript{2}NOAA/NWS/Office of Climate, Water, and Weather Services, Silver Spring, MD, USA
\textsuperscript{3}University Corporation for Atmospheric Research, Boulder, CO, USA
\textsuperscript{4}Consultant, Bethesda, MD, USA
\textsuperscript{5}NOAA/NWS/Office of Hydrologic Development, Silver Spring, MD, USA
Motivation

• All modern estimation theory-based data assimilation (DA) techniques assume that the statistical properties of the errors in the model and in the data are perfectly known.

• In hydrologic reality, however, the above assumption is rarely, if ever, met.

• The errors associated with rainfall-runoff processes are nonlinear and flow- and scale-dependent, and hence are very difficult to model accurately, particularly given the general paucity of hydrologic observations.
The “PQR” problem (from the 1st Workshop in Delft, Nov, 2010)

\[ \mathbf{x}(N+1 | N) = \mathbf{H}'(N+1)\mathbf{R}^{-1}(N+1)\mathbf{z}(N+1) \\
+ \mathbf{e}^{-1}(N+1 | N)\mathbf{H}(N+1) \mathbf{z}(N+1)\mathbf{H}'(N+1) \mathbf{R}^{-1}(N+1)\mathbf{z}(N+1) \\
+ \mathbf{e}^{-1}(N+1 | N)\mathbf{H}(N+1) \mathbf{z}(N+1)\mathbf{H}'(N+1) \mathbf{R}^{-1}(N+1)\mathbf{z}(N+1) \\
\]

From Schweppe (1973)
Motivation (cont.)

• Compared to the atmospheric processes, the hydrologic processes for surface and groundwater flow operate over a much wider range of time scales:
  • Water in the atmosphere ~ 8 days
  • Water in rivers ~ 17 days
  • Groundwater ~ days to millennia
• The multiscale nature of hydrologic processes makes inaccurate modeling of uncertainty rather unforgiving.
  • Difficult to recover from a wrong turn
Motivation (cont.)

• If the statistical properties of the errors involved are perfectly known, and if the inverse problem is well-posed, the existing sequential DA techniques, or batch DA techniques with forward propagation of the updated initial conditions (IC), may be expected to account for time scale-dependent dynamics well.

• In reality, the above conditions are rarely met.

• The DA solutions based on inaccurate modeling of scale-dependent error statistics may not be representative of the model dynamics even after a long warm-up period if they are associated with long-memory processes.
Sacramento Soil Moisture Components

SAC-SMA Model

- Precipitation
  - Pervious
  - Impervious

Evaporation

Upper Zone

Lower Zone

Impervious and Direct Runoff

Surface Runoff

Interflow

Supplemental Baseflow

Primary Baseflow
Prescribe the initial background error covariance of model states

Solve for the initial model states, biases for precipitation and PE, and time-varying model errors under the MSE objective function

Propagate the uncertainty in the initial model states an hour forward using the model errors inferred from the preceding time step

VAR with forward propagation of initial condition
SAC state dynamics with and without updating by VAR
(Aug 29 - Sep 10, 2001)
Each trace is a 36-hr DA-aided forecast w/ clairvoyant QPF
SAC state dynamics with and without updating by VAR (Nov 11-21, 1998)
Each trace is a 36-hr DA-aided forecast w/ clairvoyant QPF.
Proposed new (old?) approach

- Relax unattainable assumptions such as perfectly known statistical properties of the model errors as much as possible.
- Separate scale-disparate dynamics
  - Soil moisture accounting
  - Routing

- Divide the DA problem into multiscale bias correction (MSBC) and adaptive error modeling (AEM).
Proposed approach (cont.)

• Decomposes the inverse problem into two parts with disparate time scales: a first-order bias estimation problem and a higher-order error modeling problem.
  • The former operates at time scales where the random errors may average out, whereas the latter operates at a much smaller scale.

• The resulting inverse problems are likely to be of lower dimensionality and hence less likely to be underdetermined.
• Allows different DA techniques for highly nonlinear soil moisture accounting (SMA) and mildly nonlinear routing.
Proposed approach (cont.)

- 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.
Multi-Scale Bias Correction (MSBC)

Assume nothing is known about the model states

Solve for the initial model states and multiplicative biases for precipitation, PE and model runoff under a mass balance-only objective function

Solve for multiplicative biases for precipitation, PE and model runoff under a mass balance-only objective function

Solve for multiplicative biases for precipitation, PE and model runoff under a mass balance-only objective function

Solve for multiplicative biases for precipitation, PE and model runoff under a mass balance-only objective function
For each scale employed in MSBC:

\[
\begin{align*}
\text{Minimize} \quad & J_{k}^{\text{MSBC}} = \frac{1}{2}[U^{T}Z_{q} - U^{T}H_{qq}^{\text{SMA}}(X_{p}, X_{e}, X_{r})]^{T} \sigma_{ss}^{-2} \\
& \quad + \frac{1}{2}[Z_{p} - H_{pp}X_{p}]^{T} R_{pp}^{-1}[Z_{p} - H_{pp}X_{p}] \\
& \quad + \frac{1}{2}[Z_{e} - H_{ee}X_{e}]^{T} R_{ee}^{-1}[Z_{e} - H_{ee}X_{e}] + \frac{1}{2}[Z_{r} - H_{rr}X_{r}]^{T} R_{rr}^{-1}[Z_{r} - H_{rr}X_{r}] \\
\text{subject to} \quad & X_{s,j} = F(X_{s,j-1}, X_{p}, X_{e}, X_{r}), \quad j = k - l + 1, \ldots, k \\
& X_{s,j}^{\text{min}} \leq X_{s,j} \leq X_{s,j}^{\text{max}}, \quad j = k - l + 1, \ldots, k
\end{align*}
\]

\(X_{p}\) = (multiplicative) bias in observed precipitation
\(X_{e}\) = (multiplicative) bias in observed potential evaporation
\(X_{r}\) = (additive) bias in modeled runoff
Adaptive Error Modeling (AEM)

Minimize \( J_k^{AEM} = \frac{1}{2} [Z_q - H_{qq}^{route}(X_w)]^T R_q^{-1} [Z_q - H_{qq}^{route}(X_w)] + \frac{1}{2} X_w^T R_{ww}^{-1} X_w \)

subject to \( TCI_j + x_{w,j} \geq 0, \quad j = k - l + 1, \ldots, k \)

\( X_w = \) (additive) time-varying error in modeled runoff
\( TCI_j = \) total channel inflow (i.e. runoff) at time step \( j \)

Timing error accounted for in \( H_{qq}^{route}(X_w) \)
## Accounting for timing errors

### Table 1: Geomorphological and hydroclimatological characteristics of the study basins in WGRFC

<table>
<thead>
<tr>
<th>No.</th>
<th>Basin ID</th>
<th>USGS ID</th>
<th>Time-to-peak, h</th>
<th>Area, km²</th>
<th>Average annual discharge, m³/s</th>
<th>Average annual precipitation, mm</th>
<th>Impervious area, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ATIT2</td>
<td>081590000</td>
<td>9</td>
<td>844</td>
<td>2.45</td>
<td>684</td>
<td>5.8</td>
</tr>
<tr>
<td>2</td>
<td>DCJT2</td>
<td>080535000</td>
<td>6</td>
<td>1039</td>
<td>2.38</td>
<td>684</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>GBHT2</td>
<td>080760000</td>
<td>7</td>
<td>137</td>
<td>2.58</td>
<td>640</td>
<td>28.8</td>
</tr>
<tr>
<td>4</td>
<td>GETT2</td>
<td>081049000</td>
<td>10</td>
<td>334</td>
<td>3.59</td>
<td>1175</td>
<td>1.3</td>
</tr>
<tr>
<td>5</td>
<td>GNVT2</td>
<td>080172000</td>
<td>16</td>
<td>212</td>
<td>1.47</td>
<td>605</td>
<td>7.6</td>
</tr>
<tr>
<td>6</td>
<td>HBMT2</td>
<td>080750000</td>
<td>4</td>
<td>246</td>
<td>2.28</td>
<td>982</td>
<td>44.1</td>
</tr>
<tr>
<td>7</td>
<td>HNTT2</td>
<td>081655000</td>
<td>3</td>
<td>769</td>
<td>8.69</td>
<td>999</td>
<td>–</td>
</tr>
<tr>
<td>8</td>
<td>JTBT2</td>
<td>080796000</td>
<td>8</td>
<td>945</td>
<td>1.93</td>
<td>640</td>
<td>–</td>
</tr>
<tr>
<td>9</td>
<td>KNLT2</td>
<td>081520000</td>
<td>7</td>
<td>904</td>
<td>1.28</td>
<td>491</td>
<td>0.3</td>
</tr>
<tr>
<td>10</td>
<td>LYNT2</td>
<td>081101000</td>
<td>19</td>
<td>508</td>
<td>2.08</td>
<td>868</td>
<td>2.1</td>
</tr>
<tr>
<td>11</td>
<td>MCKT2</td>
<td>080589000</td>
<td>13</td>
<td>427</td>
<td>1.63</td>
<td>806</td>
<td>1.6</td>
</tr>
<tr>
<td>12</td>
<td>MDST2</td>
<td>080658000</td>
<td>24</td>
<td>870</td>
<td>4.05</td>
<td>859</td>
<td>10.7</td>
</tr>
<tr>
<td>13</td>
<td>MTPT2</td>
<td>081626000</td>
<td>25</td>
<td>435</td>
<td>4.90</td>
<td>877</td>
<td>5.4</td>
</tr>
<tr>
<td>14</td>
<td>PICT2</td>
<td>081010000</td>
<td>6</td>
<td>1178</td>
<td>4.69</td>
<td>763</td>
<td>0.5</td>
</tr>
<tr>
<td>15</td>
<td>QLAT2</td>
<td>080173000</td>
<td>10</td>
<td>197</td>
<td>4.15</td>
<td>701</td>
<td>5.7</td>
</tr>
<tr>
<td>16</td>
<td>REFT2</td>
<td>081895000</td>
<td>36</td>
<td>1787</td>
<td>2.41</td>
<td>929</td>
<td>1.3</td>
</tr>
<tr>
<td>17</td>
<td>SBMT2</td>
<td>081643000</td>
<td>14</td>
<td>896</td>
<td>4.79</td>
<td>763</td>
<td>1.4</td>
</tr>
<tr>
<td>18</td>
<td>SCDT2</td>
<td>081769000</td>
<td>18</td>
<td>932</td>
<td>5.08</td>
<td>763</td>
<td>0.9</td>
</tr>
<tr>
<td>19</td>
<td>SOLT2</td>
<td>080417000</td>
<td>82</td>
<td>1746</td>
<td>2.82</td>
<td>763</td>
<td>–</td>
</tr>
<tr>
<td>20</td>
<td>UVAT2</td>
<td>081900000</td>
<td>2</td>
<td>1981</td>
<td>13.1</td>
<td>1315</td>
<td>4.8</td>
</tr>
</tbody>
</table>
MDST2

UNIT HYDROGRAPH (CMS/MM)

TIME (HRS)

23
Effect of timing errors in updating of soil water states

From Seo et al. (2010)
MSBC - Initial results

• For 23 headwater basins in Texas
• Compared with VAR
  • Fix-lag smoother formulation with continuous forward propagation of IC
• Adjusted for timing errors for the smallest assimilation window
MSBC - Initial results (cont.)

- How to assess performance?
  - Only outlet streamflow observations are available
  - Streamflow prediction error vs. lead time
    - The smaller, the better.
  - Departure of updated soil water states from the base model states
    - The smaller, the better.
  - Why? DA assumes that the model is realistic; in the absence of any other information, the model dynamics is to be trusted
In both VAR and MSBC, the IC’s are forward-propagated.
MSBC vs. VAR (cont.)
MSBC vs. VAR (cont.)
MSBC vs. VAR (cont.)
MSBC vs. VAR (cont.)

![Graphs showing Comparison]

- **JBT2_02_00**
  - **VAR**
  - **MSBC**
  - **NO UPDATING**
  - **PERSISTENCE**

- **KNLT2_02_18**
  - **VAR**
  - **MSBC**
  - **NO UPDATING**
  - **PERSISTENCE**

**RMSE (CMS)**

**LEAD TIME (HRS)**
MSBC vs. VAR (cont.)

![Graphs showing comparison of MSBC vs. VAR](image)

- **LYNT2_02_06**
  - VAR
  - MSBC
  - NO UPDATING
  - PERSISTENCE

- **MCKT2_02_06**
  - VAR
  - MSBC
  - NO UPDATING
  - PERSISTENCE

Parameters:
- **LEAD TIME (HRS)**
- **RMSE (CMS)**
MSBC vs. VAR (cont.)

**MDST2_02_00**

![Chart 1: MDST2_02_00](chart1.png)

- **VARIABLES:** VAR, MSBC, NO UPDATING, PERSISTENCE
- **Y-Axis:** RMSE (CMS)
- **X-Axis:** LEAD TIME (HRS)

**MTPT2_02_00**

![Chart 2: MTPT2_02_00](chart2.png)

- **VARIABLES:** VAR, MSBC, NO UPDATING, PERSISTENCE
- **Y-Axis:** RMSE (CMS)
- **X-Axis:** LEAD TIME (HRS)
MSBC vs. VAR (cont.)

PICT2_02_00

QLAT2_02_00

- VAR
- MSBC
- NO UPDATING
- PERSISTENCE

RMSE (CMS)

LEAD TIME (HRS)
MSBC vs. VAR (cont.)
MSBC vs. VAR (cont.)

- SKMT2_02_00
- SOLT2_02_00

![Graphs showing RMSE vs. lead time for different models.]

- VAR
- MSBC
- NO UPDATING
- PERSISTENCE

LEAD TIME (HRS)

RMSE (CMS)

0 5 10 15 20 25 30 35
MSBC vs. VAR (cont.)

RMSE vs. lead time

<table>
<thead>
<tr>
<th>VAR better</th>
<th>3 basins</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSBC better</td>
<td>10 basins</td>
</tr>
<tr>
<td>Comparable</td>
<td>10 basins</td>
</tr>
</tbody>
</table>

UVAT2_02_06
MSBC generally improves over VAR in outlet streamflow prediction
MSBC-updated model states are much closer to the base model states (i.e. reduces under-determinedness)
MSBC-updated model states are much closer to the base model states (cont.)
Time-averaged Euclidean norm of normalized SAC states with and without DA

High-flow events only
Time-averaged Euclidean norm

\[ \text{TAEN} = \frac{1}{T} \sqrt{x_{1,t=1}^2 + x_{1,t=2}^2 + \cdots + x_{1,t=T}^2 + \cdots + x_{6,t=1}^2 + x_{6,t=2}^2 + \cdots + x_{6,t=T}^2} \]

where

\[ x_{1,t=k} = \frac{UZTW_{C_{t=k}^{\text{noDA}}} - UZTW_{C_{t=k}^{DA}}}{UZTW_{M}} \]

\[ x_{2,t=k} = \frac{UZF_{W_{C_{t=k}^{\text{noDA}}} - UZF_{W_{C_{t=k}^{DA}}}}}{UZF_{W_{M}}} \]

\[ x_{3,t=k} = \frac{LZTW_{C_{t=k}^{\text{noDA}}} - LZTW_{C_{t=k}^{DA}}}{LZTW_{M}} \]

\[ x_{4,t=k} = \frac{LZFS_{C_{t=k}^{\text{noDA}}} - LZFS_{C_{t=k}^{DA}}}{LZFS_{M}} \]

\[ x_{5,t=k} = \frac{LZFP_{C_{t=k}^{\text{noDA}}} - LZFP_{C_{t=k}^{DA}}}{LZFP_{M}} \]

\[ x_{6,t=k} = \frac{ADIM_{C_{t=k}^{\text{noDA}}} - ADIM_{C_{t=k}^{DA}}}{UZTW_{M} + LZTW_{M}} \]
Time-averaged Euclidean norm of normalized SAC states with and without DA (cont.)

High-flow events only

The MSBC-updated states are much closer to the base model states than the VAR-updated for both upper- and lower-zone soil water.
Water balance components

High-flow events only

Runoff (mm)

Precipitation (mm)

- Q(obs)
- Q(noDA)
- Q(var)
- Q(msbc)

- P(obs)
- P(noDA)
- P(var)
- P(msbc)

Sep 10-12, 2012
2nd Int. DA Workshop, Incheon
Water balance components (cont.)

High-flow events only

Potential Evap. (mm)

Actual Evap. (mm)

Water balance components (cont.)

High-flow events only

Potential Evap. (mm)

Actual Evap. (mm)
Budyko curve

Ea: Actual evaporation (mm)
P: Precipitation (mm)
Ep: Potential evaporation (mm)

Because Ea observations are unavailable, “observed” Ea/P (blue dots) is calculated from the runoff ratio.
Possible criticism

• Why not model uncertainty propagation (as best as one can) to provide (hopefully) informative prior for filtering?
  • This runs into the same “PQR” problem
• How to select time scales objectively?
  • We are exploring wavelet analysis and principal component analysis of runoff error time series for this
    • Identify scales at which model errors are dominant
• Added computational requirements
  • This is an issue but not a crippling one
  • Explore smart telescoping of the assimilation window
In closing

- In hydrologic reality, the “PQR” assumption is rarely, if ever, met.
  - The errors associated with rainfall-runoff processes are nonlinear and flow- and scale-dependent, and hence are extremely difficult to model accurately.

- Relax unattainable assumptions such as perfectly known statistical properties of the model errors as much as possible.
  - Decomposes the inverse problem into two parts with disparate time scales: a first-order bias estimation problem (multiscale bias correction) and a higher-order error modeling problem (adaptive error modeling).

- MSBC shows promising results.
  - Work is ongoing for objective selection of scales and computationally economical implementation for real-time applications.
  - Plan to apply to large systems such as the Nakdong River Basin in Korea and the Trinity River Basin in TX.
Nakdong River Basin (23,817 km$^2$)
Trinity River Basin (40,380 km²)
THANK YOU

For more information, contact:
djseo@uta.edu