Data Assimilator (DA) for Hydrology Laboratory’s Research
Distributed Hydrologic Model (HL-RDHM)

Haksu Lee\textsuperscript{1,2}, Dong-Jun Seo\textsuperscript{1,2}, Victor Koren\textsuperscript{1}

\textsuperscript{1}NOAA/NWS/Office of Hydrologic Development
\textsuperscript{2}University Corporation for Atmospheric Research
Acknowledgments

- NOAA Climate Program Office/Climate Prediction Program for the Americas (CPPA)
  - Core Project: Pedro Restrepo, John Schaake, Ken Mitchell
  - “Incorporating knowledge of observational uncertainties in streamflow forecasting applications in the Western US”: Andrew Slater, Martyn Clark
- NASA
  - “Improving NOAA/NWS River Forecast Center Decision Support with NASA Satellite and Land Information System Products”: Pedro Restrepo, Ashutosh Limaye, Christa Peters-Lidard
- AHPS/XEFS
- Hydrologic Ensemble Prediction Group - Limin Wu, Julie Demargne, James Brown, Satish Regonda, Yuqiong Liu
- Hydrology Group - Ziya Zhang
- Hydraulics Group - Fekadu Moreda
Predicting Floods to Droughts In Your Neighborhood

River Services
(600 miles per forecast point)

Water Resource Services
(6 square mile forecast basins)

River Conditions

From Carter (2006)

Soil Conditions

OHD Seminar, May 29, 2008
Objective of the project

• Develop a prototype data assimilator (DA) for distributed hydrologic models in HL-RDHM for more accurate, high-resolution analysis and prediction of streamflow and soil moisture
  – by reducing uncertainty in the model initial conditions (i.e. model soil moisture)
Outline of the presentation

• Models used
• Technique used
  – What is 4DVAR?
  – How does 4DVAR work?
• Questions investigated
• Approach
  – Synthetic experiments
  – Real-world experiment
• Conclusions
• Next steps
Models used

  - Gridded (~4x4 km²) soil moisture accounting models (SAC, API)
  - Gridded snow ablation model (SNOW-17)
  - Kinematic-wave routing

- The prototype DA assimilates the following data into gridded SAC-kinematic wave routing models (Seo et al. 2003b, Lee et al. a,b):
  - Streamflow (at outlet and interior locations)
  - Gridded precipitation
  - Potential evaporation (PE)
  - In-situ soil moisture
SAC-HT allows translation of SAC states to soil moisture, and hence assimilation of soil moisture data into SAC
Technique used

• 4-dimensional variational assimilation, or 4DVAR
  – Arguably the most advanced data assimilation (DA) technique used in operational weather forecasting today
  – More amenable to forecaster control than ensemble Kalman filter/smooth (Evensen 1994, Evensen and van Leeuwen 2000)
  – Amenable to ensemble DA via maximum likelihood ensemble filter (MLEF) (Zupanski 2005)
What does 4DVAR do?

- Given all available data, the model(s) and the prescribed uncertainties for them, adjust the selected variables (e.g. the model states) such that the model results best fit the data
  - Under user-prescribed *criterion* (usually minimization of mean square errors)
  - Necessarily *model-dynamically consistent*
  - Not unlike what a human forecaster may do
  - As in any curve fitting, subject to over-fitting (too large a degree of freedom) and under-fitting (too small a degree of freedom)
What does 4DVAR do? (cont.)

Forecast time-
Forecast time-(n-1)
Forecast time-(n-2)

Assimilation window
Forecasting period

OBS
SIM
DA

Flow (CMS)

0 50 100 150 200 250 300

1600 1620 1640 1660 1680 1700 1720

ELAPSED TIME (HRS)
How does 4DVAR work?

Adjust model states, and observed precipitation and PE so that the model-simulated flow is sufficiently close to the observed data.

\[
\text{Minimize} \quad J_k = \frac{1}{2} \left[ Z_q - H_{qq} (X_{s,k-l}, X_p, X_e) \right]^T R^{-1}_{qq} \left[ Z_q - H_{qq} (X_{s,k-l}, X_p, X_e) \right] + \frac{1}{2} \left[ Z_\theta - H_{\theta \theta} (X_{s,k-l}, X_p, X_e) \right]^T R^{-1}_{\theta \theta} \left[ Z_\theta - H_{\theta \theta} (X_{s,k-l}, X_p, X_e) \right] \\
+ \frac{1}{2} \left[ Z_p - H_{pp} X_p \right]^T R^{-1}_{pp} \left[ Z_p - H_{pp} X_p \right] + \frac{1}{2} \left[ Z_e - H_{ee} X_e \right]^T R^{-1}_{ee} \left[ Z_e - H_{ee} X_e \right] \\
+ \frac{1}{2} \left[ Z_b - H_{ab} X_{s,k-l} \right]^T R^{-1}_{bb} \left[ Z_b - H_{ab} X_{s,k-l} \right]
\]

subject to \( X_{s,j} = F(X_{s,j-1}, X_{p,j}, X_{e,j}), \quad j = k - l + 1, \ldots, k \)

\[
X_{s,i}^{\min} \leq X_{s,k,i} \leq X_{s,i}^{\max}, \quad j = k - l, \ldots, k; \quad i = 1, \ldots, 6
\]

How good is my streamflow data?

How good is my soil moisture data?

How good is my precipitation data?

How good is my potential evaporation (PE) data?

What do I know about the initial soil moisture states?

Whatever adjustments I am making must abide by the model dynamics.

The adjustments must be within physically realistic bounds.

OHD Seminar, May 29, 2008
4DVAR

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)  ABRFC/WTTO2  CHANNEL (w/o DA)  BIAS IN PRECIP  BIAS IN PE

HSLOPE (w/ DA)  CHANNEL (w/ DA)  OUTLET STREAMFLOW (CMS)

WTTO2 HR 001993
1993111711

OHD Seminar, May 29, 2008
Questions investigated

• What is the value of assimilating streamflow (outlet, interior) data for improved accuracy in monitoring (analysis) and prediction of streamflow and soil moisture?
  – According to uncertainty in the initial soil moisture conditions

• What is the value of assimilating in-situ soil moisture data?

• What is the value of assimilating gridded precipitation data
Approach

• Carry out **synthetic** and **real-world** experiments
• Why synthetic experiments?
  – In reality, truth is unknown and many uncertainties complicate understanding and interpretation
• In synthetic experiments:
  – **Truth is known**
    ✅ Easier to evaluate DA performance
  – Can **separate different uncertainty sources**
    ✅ Initial condition uncertainty (ICU)
    ✅ Precipitation uncertainty (PU)
    ✅ Other Observational uncertainty
    ✅ Model structural and parametric uncertainty
  – More likely to **gain and advance understanding** on hydrologic DA with distributed models
Synthetic experiments

• Methodology
  – Assume “true” initial soil moisture states (IC), streamflow (Q) and soil moisture (S) observations, and observed precipitation (P)

  – Perturb with low, medium and high levels of noise
    • IC, Q, S (Experiment 1, Lee et al.a)
    • IC, Q, S, P (Experiment 2, Lee et al.b)

  – Assimilate the observations via 4DVAR
  – Repeat above 2 steps to generate ensembles
  – Assess the quality of posterior ensembles
  – Monte-Carlo type of 4DVAR

\(\text{a,bin preparation}\)

Mar 12, 2008
Study basin - Eldon (795 km²)

- Soil moisture site
- Stream gauge

- SW: Westville
- Qc: Christie
- Qd: Dutch
- Qe: Eldon

- 64.7 km²
- 105.1 km²

- Contours: 50 [m]
  - Elevation [m]:
    - 106 - 174
    - 175 - 243
    - 244 - 312
    - 313 - 381
    - 382 - 449
    - 450 - 518
    - 519 - 587
    - 588 - 656
    - 657 - 725

Mar 12, 2008
Case studied
2000/6/22/18Z - 2000/6/24/6Z

Elapsed time (hrs)
Streamflow (m³/s)

Assimilation window

Precip [mm/36hr]
- 100 - 102.4
- 102.4 - 104.8
- 104.8 - 107.2
- 107.2 - 109.6
- 109.6 - 112
- 112 - 114.4
- 114.4 - 116.7
- 116.7 - 119.1
- 119.1 - 121.5
- 121.5 - 123.9
- 123.9 - 126.3
- 126.3 - 128.7
- 128.7 - 131.1
- 131.1 - 133.5
- 133.5 - 135.9
- 135.9 - 138.3
- 138.3 - 140.7
- 140.7 - 143
- 143 - 145.4
- 145.4 - 147.8
Synthetic Experiment I: Sensitivity of DA to initial condition uncertainty (ICU) and observational uncertainties
Synthetic Experiment I: Results

RMSE for Q for assimilation period

Average RMSE of streamflow simulation over all grid boxes (cms)

ICUL (initial condition uncertainty level):
\[ \text{ICUL} = \frac{\text{std}}{X_{\text{max}}} \]

SS (Skill Score)
\[ SS = 1 - \frac{\text{MSE}_{\text{DA}}}{\text{MSE}} \]

SS = 1 perfect
SS > 0 skillful
SS ≤ 0 no skill
Synthetic Experiment I: Results (cont.)

Average RMSE of soil moisture simulation over all grid boxes

RMSE for SWC25 for assimilation period

- w/o DA
- w/ $Q_E$ (uncertain Q)
- w/ $Q_E$
- w/ $Q_{ECD}$
- w/ $Q_{ECD} S_W$

OHD Seminar, May 29, 2008
Streamflow results for assimilating accurate streamflow & soil moisture obs under uncertain IC

Eldon
Assumed truth
Before DA
After DA

Christie

Dutch

w/o DA
w/ Q
w/ Q
w/ Q
w/ Q

Eldon
Christie
Dutch

Eldon
Christie
Dutch

Eldon
Christie
Dutch

Eldon
Christie
Dutch

OHD Seminar, May 29, 2008
Synthetic Experiment II: Sensitivity of DA to precipitation uncertainty (PU)
Precipitation Uncertainty Model

\[ P_k(u) = B_k \, O_k(u) + \sigma \, Z_k(u) \]

\[ \ln B_k = a_1 \ln B_{k-1} + W_k \] (Smith and Krajewski 1991)

where

- \( P_k(u) \): perturbed rainfall at location \( u \) at hour \( k \) (mm)
- \( O_k(u) \): reference rainfall at hour \( k \) (mm)
- \( B_k \): mean field bias at hour \( k \)
- \( Z_k(u) \): spatially-correlated standard normal random noise

\[ 2\sigma/O_k(u) = \begin{cases} 1-0.02 \, O_k(u) & \text{if } O_k(u) \leq 25.4 \, (mm) \\ 0.5 & \text{if } O_k(u) > 25.4 \, (mm) \end{cases} \] (Carpenter and Georgakakos 2006)

where

- \( \sigma \): rainfall amount-dependent standard deviation of the noise
Impact of additionally assimilating precipitation on streamflow prediction

Eldon

Christie

Dutch

w/o DA

w/ $Q_{ECD} S_W$

w/ $Q_{ECD} S_W P$

Assumed truth
Before DA
After DA
Total precipitation over the assimilation window [mm/hr] (input)

Total streamflow at the outlet over the assimilation window [mm/hr] (output)
Water balance!!!

Soil moisture

precipitation

streamflow

w/o da

w/ $Q_{ECD} \cdot S_W$

w/ $Q_{ECD} \cdot S_W \cdot P$

OHD Seminar, May 29, 2008
Impact of additionally assimilating precipitation to streamflow and soil moisture simulation

- Large precipitation uncertainty
- Medium streamflow observation uncertainty
- Medium soil moisture observation uncertainty
Impact of mean field bias ($B'=B?\alpha$) and noise ($\sigma'=a_3\sigma$) to streamflow simulation via DA

Outlet streamflow, in-situ soil moisture, gridded precipitation assimilated

- Medium initial condition uncertainty
- Medium streamflow observation uncertainty
- Medium soil moisture observation uncertainty

Mar 12, 2008
Real-World Experiment
Real-World Experiment: Questions

• The models are never perfect
  – Structural errors
  – Parametric errors

• Soil moisture is seldom observed directly, and never at the model grid scale

• How to account for these uncertainties?

• How do these uncertainties impact DA?
Experiment Design

• Setup
  − Assimilation window: 36 hrs
  − Error variance for precipitation: sample variance
  − Error variance for streamflow: sensitivity analysis
  − Error variance for soil moisture: data analysis & model simulation

• Data
  − Streamflow: 1997 – 2000
  − Precipitation: ABRFC-produced operational multisensor QPE

Acknowledgment: We would like to thank the Oklahoma Climatological Survey for allowing the use of the Oklahoma Mesonet soil moisture data.
Uncertainties associated with in-situ soil moisture obs (OK Mesonet)

- **Device error:**
  - Soil moisture sensor error (CSI 229-L) (e1)
  - Numerical precision error (e2)
  - Device limit to measure extreme values (e3)

- **Scaling** (e4)
  - pt to HRAP scale error estimated by cdf matching technique
  - bias correction is done by cdf matching

- **Spatial variability** (e5)

- **Overall error variances** (=e1+e2+e3+e4+e5)
Estimated standard deviation for soil moisture error

Standard Deviation for Soil Moisture Error

≤ 0.05 m³/m³ (Walker and Houser, 2004) is useful for data assimilation
Streamflow time series for the biggest event in yr 2000:
lead time = 0 hr

Eldon

Christie

Dutch

obs
w/o DA
Qe
Qe Qc Qd
Qe P
Qe Qc Qd P
Qe Sw
Qe Qc Qd Sw
Qe P Sw
Qe Qc Qd P Sw

OHD Seminar, May 29, 2008
Hourly Soil Moisture at Westville for yr 2000

~ 2.5 months

$S_W$ at 25cm
rmse vs. lead time for streamflow for yr 2000

Eldon

Christie

Dutch

w/o DA
Qe
Qe Qc Qd
Qe P
Qe Qc Qd P
Qe Sw
Qe Qc Qd Sw
Qe P Sw
Qe Qc Qd P Sw

Lead time

OHD Seminar, May 29, 2008
Event 1 - rising limb

Event 1 - recession limb

Event 2 - rising limb

Event 2 - recession limb

Assimilation results
Prediction results

OHD Seminar, May 29, 2008 37
Conclusions

- Assimilating streamflow, in-situ soil moisture and QPE data in real time has large potential value for high-resolution analysis and prediction of streamflow and soil moisture.

- However, its potency is sensitive to the uncertainty in the initial soil moisture conditions, the quality of observations and the goodness of the models (and their parameters) used.

- It is seen that:
  - If the initial conditions are highly uncertain, soil moisture observations have larger positive impact than streamflow observations.
  - If the initial conditions are less uncertain, accurate streamflow observations have larger positive impact than soil moisture observations.
Conclusions (cont.)

• **Assimilating QPE**, in addition to streamflow and soil moisture observations, *improves water balance calculations*
  – If precipitation uncertainty is not properly accounted for in DA, streamflow balance may be improved, but only at the expense of deteriorated soil moisture balance

• If there are **large uncertainties in QPE** and in the initial conditions, assimilating **soil moisture observations** has **large positive impact** on analysis and prediction of soil moisture and streamflow

• **Assimilating streamflow observations** at both the outlet and interior locations generally *improves streamflow prediction* at those locations

• **Assimilating soil moisture observations** have **large positive impact** on model soil moisture states on cold starts
Upshot of all this

- A prototype DA has been developed that is capable of assimilating streamflow, in-situ soil moisture and gridded QPE into SAC and kinematic wave routing models of HL-RDHM
- Results thus far are encouraging, and points out salient observational, scientific and practical issues to be addressed
- Gained much understanding on how the major sources of uncertainty impact the performance of DA and what the next steps are toward improving operational worthiness
- The immediate next step is to simplify the current prototype to avoid “overfitting” and reduce computational burden (ongoing – should also help forecaster control of the DA), and to evaluate performance for multiple basins (ongoing)
- The new prototype to be considered for integration with HL-RDHM in the CHPS/FEWS/XEFS environment
Next Steps

• **Simplify** the current prototype
  – Avoid overfitting, reduce amount of computation
• Further **assess model errors** and their impact
• **Better understand** in-situ soil moisture measurement (HMT/Robert Zamora)
• Assimilate **satellite-derived soil moisture** data (w/ NCEP/EMC)
  – Into SAC-HT via LIS
  – Assimilate satellite-aided model soil moisture fields into the prototype DA
• **Develop 4DVAR into ensemble DA** using, e.g., maximum likelihood ensemble filter
Thank you

Q&A, discussion
Appendix
Uncertainty model for initial SAC states

\[ X_{k}(0:k) = X_{\text{max}} \left[ \exp(\eta_k) - 1 \right] + X_{\text{true}} \]

where \( \eta_k = -0.5 \ln \left[ 1 + (a_{IC}X_{\text{max}})^2 \right] + \varepsilon_k \left[ \ln \left( 1 + (a_{IC}X_{\text{max}})^2 \right) \right]^{1/2} \),

\( \varepsilon_k \sim k\text{-th spatially correlated } N(0,1) \text{ random deviate} \)
Uncertainty model for in-situ soil moisture obs

\[ Z_S(t,k) = Z_S(t) + a_S \, w(t,k) \]

Where \( w(t,k) \) is the \( k \)-th temporally correlated \( N(0,1) \) random deviate.

\[ \begin{align*}
&\text{Generated soil moisture obs (}a_S=0.03) \\
&\leq 0.05\ m^3/m^3 \text{ (Walker and Houser, 2004) is useful for data assimilation}
\end{align*} \]
Uncertainty model for streamflow obs

\[ Q(t:k) = Q(t) + a_q Q(t) \cdot w(t:k) \]

where \( w(t) \) is \( k \)-th temporally correlated \( N(0,1) \) random deviate
Impact of additionally assimilating precipitation to streamflow simulation

- Large precipitation uncertainty
- Medium streamflow observation uncertainty
- Medium soil moisture observation uncertainty
Impact of additionally assimilating precipitation to soil moisture simulation at 25-cm depth

- Large precipitation uncertainty
- Medium streamflow observation uncertainty
- Medium soil moisture observation uncertainty
How additionally assimilating precipitation may reduce PU
Large perturbations to mean-field bias (median=3)

- Medium initial condition uncertainty
- Medium streamflow observation uncertainty
Vision for Ensemble & DA

- Improved accuracy
- Reliable uncertainty estimates
- Benefit-cost effectiveness maximized

The Big Picture:
- Precipitation, temperature, humidity, surface pressure, wind, insolation
- Soil moisture, snowpack
- Streamflow, stage
- Reservoir level, temperature
- Flow, stage, velocity, temperature
- Sea level, wind, pressure, temperature

- Global coupled ocean-land surface-atm. model
- Regional coupled land surface-atm. model
- Uncoupled land surface

Hydrologic Ensemble Post-Processor

Hydrology and Water Resources Models

Parametric Ensemble Processor

Atmospheric Ensemble Pre-Processor

Hydrology and Water Resources Ensemble Product Generator

OHD Seminar, May 29, 2008