Towards operationalizing ensemble DA in hydrologic forecasting

Albrecht Weerts
Ensemble Kalman Filter Rhine assimilation water levels 14 locations

Operational since 1-1-2008 with 48 members

See also Werner et al., 2014
Results EnKF at Lobith over 2 year hindcast (2006/2007)

(a) RMSE in forecast level (m)

- Lobith (with EnKF)
- Lobith (simulated)

(b) Water Level (m)

- Observed
- Simulated

Januari 2007
Different forecasting paradigms

**North America**
- Primarily Manual process
  - Forcings QC in-house
  - Interactive Forecasting
  - Involved
  - Low-frequency (daily)
  - 6 hr timesteps
  - Ensembles climatology based

**Europe**
- Primarily Automated process
  - Forcings QC out-of-house
  - Automated Forecasting
  - Observant
  - High frequency (hourly)
  - Hourly time steps
  - Ensembles NWP based

From Gijsbers et al. 2011
Some conclusions from the workshop in 2010

- Role forecaster!
  - different for each organisation / institutional setup

- Manual vs Automated DA approach
  - Europe mainly automated (deterministic)
  - USA often manual => Runtime MODIFICATIONS
    > Using observed streamflow to infer appropriate adjustments to:
      - Model states
      - Timing and magnitude of forcings

- Changes in budget/priorities or other outside factors
pro’s / con’s automated DA

Essential in “fast response” situations where there is no time for manual analysis;
Consistent DA is more reproducible than subjective forecaster specified adjustments.

Updates to improve performance in one time domain (e.g. short) may degrade performance in another (e.g. long);
Improved simulation does not guarantee improved forecast performance;
Forecasters don’t (won’t) trust what they can’t see or understand;
May reduce the need for the forecaster to understand what the models are doing (Robotic Operations).
Operational forecasters perform *manual* data assimilation by leveraging their expertise.

Automatic DA is critical for very short lead times where manual analysis is not feasible.

DA techniques should be engineered to assist the forecaster in *understanding and applying* the most appropriate adjustments suggested by the data (guidance).

Development, implementation and institutionalization of forecaster-supervised automatic DA in operational hydrology will require:

- Collective efforts from both the research and the operational communities
- Cross-cutting community planning and action to capitalize on recent advances and to seize the window of opportunity
So where are we?

• Joint paper in HESS 2012 Liu et al. 2012 (cited 41 times scholar.google.com) as result of workshop in Delft

• Main challenges for operational hydrologic DA:
  • theoretic or mathematical challenges;
  • characterization of uncertainties;
  • Integrating newly emerging observations such as remote sensing data;
  • real-time control of water resources systems and hydraulic structures;
  • community-based approach to hydrologic DA
    • verification
    • developing community-based generic modeling/DA tools
Rakovec et al. (2012a) presented a spatially distributed hourly ensemble rainfall generator which was used in Rakovec et al. (2012b) to study effect of update frequency, number and location of streamflow gauges.
Conclusions:

- Best results in terms of the RMSE were achieved using all observations, which includes all six discharge gauges.
- Given the travel time of the catchment, an updating frequency of 12 h seems to be the most appropriate.
- Most sensitivity in routing stores.
In Noh et al (2014) the rainfall generator of Rakovec et al. (2014) was used in combination with the lagged particle filter (Noh et al, 2011) to study the effect of noise on the DA results on 3 Catchments in Japan.

Conclusions: results show that better specification of input noise requires smaller noise on state variables to get similar performance.
Operational aspects of asynchronous filtering

Asynchronous Ensemble Kalman Filter (Sakov et al., 2010) updates model at the analysis step using past observations over a time window:

The Asynchronous EnKF is particularly attractive from a forecasting perspective as more observations can be used with **hardly any extra additional computational time!**

Rakovec et al. 2014 (submitted to WRR)
**EnKF** updates model states at time $k$ as:

$$X_k^+ = X_k^- + K_k(y_k - H_kX_k^-),$$

where $X_k^+$ is the new updated (posterior) model state matrix, $X_k^-$ is the forecasted (prior) model state matrix. $K_k$ is the Kalman gain (a weighting factor of the errors in model $H_kX_k^-$ and observations $y_k$).

**Asynchronous EnKF** is a simple modification of the EnKF, in which $X_k$ is augmented with the past forecasted observations from $W$ previous time steps $H_kX_k$:

$$\tilde{X}_k = \begin{pmatrix} X_k \\ H_{k-1}X_{k-1} \\ H_{k-2}X_{k-2} \\ \vdots \\ H_{k-W}X_{k-W} \end{pmatrix}$$
- 8 largest flood peaks observed since 1998
- Model noise: perturbation of soil moisture reservoir with spatio-temporally correlated error model (36 members)
- Sensitivity of the AEnKF to the assimilated time window: $W = 0h, W = 5h, W = 11h$
- Four partitioned state updating schemes for model states being updated (thus included in the model analysis).

<table>
<thead>
<tr>
<th>name</th>
<th>Q</th>
<th>H</th>
<th>SM</th>
<th>UZ</th>
<th>LZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>no update</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>noSM</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>HQ</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Model states: discharge (Q), water level (H), soil moisture (SM), upper zone (UZ), and lower zone (LZ). Snow and interception storages not shown.
Conceptualization into a grid

- 1 km² resolution
- 8 model states
- Lumped routing substituted by KW model
Results

![Graph showing RMSE vs. lead time with different update scenarios and augmentation values.]

Update:
- no update
- all
- noSM
- HQ

Augmentation W:
- 0
- 5
- 11
Validation

Update:
- no update
- all
- noSM
- HQ

Augmentation W:
- 0
- 5
- 11
Update:
- no update
- all
- noSM
- HQ

Augmentation W:
- 0
- 11
Conclusions AEnKF

• The AEnKF: an effective method for model state updating taking into account more (e.g. all) observations at hardly any additional computational burden;
• Partitioned update scheme: reducing the number of model states using AEnKF can lead to better forecasts of discharge;
• Largest improvements in the forecast accuracy using RMSE were observed for the scenario, when the soil moisture was left out from the analysis (because SM obs. not available);
• Updating only routing states leads to very good performance at time of update, but it deteriorates sharply at longer lead times;
• Keeping the quick catchment response storage (upper zone; UZ) in the model analysis is important especially for longer lead times;
Delft-FEWS provides an open shell system for managing forecasting processes and/or handling time series data.

Delft-FEWS incorporates a wide range of general data handling utilities, while providing an open interface to any external (forecasting) model.

The modular and highly configurable nature of Delft-FEWS allows it to be used effectively for data storage and retrieval tasks, simple forecasting systems and in highly complex systems utilising a full range of modelling techniques.

Delft-FEWS can either be deployed in a stand-alone, manually driven environment, or in a fully automated distributed client-server environment.
Models coupled

Delft FEWS
Operational Forecasting Platform

<table>
<thead>
<tr>
<th>Model</th>
<th>Type</th>
<th>Supplier/Owner</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISIS</td>
<td>Hydrodynamics</td>
<td>HR/Halcrow</td>
<td>UK</td>
</tr>
<tr>
<td>PDM</td>
<td>Rainfall-Runoff</td>
<td>CEH</td>
<td>UK</td>
</tr>
<tr>
<td>TCM</td>
<td>Rainfall-Runoff</td>
<td>CEH</td>
<td>UK</td>
</tr>
<tr>
<td>KW</td>
<td>Routing (kinematic wave)</td>
<td>CEH</td>
<td>UK</td>
</tr>
<tr>
<td>PACK</td>
<td>SnowMelt</td>
<td>CEH</td>
<td>UK</td>
</tr>
<tr>
<td>ARMA</td>
<td>Error Correction</td>
<td>CEH</td>
<td>UK</td>
</tr>
<tr>
<td>PRTF</td>
<td>Event Based RR</td>
<td>PlanB</td>
<td>UK</td>
</tr>
<tr>
<td>TRITON</td>
<td>Surge propagation/Overtopping</td>
<td>PlanB</td>
<td>UK</td>
</tr>
<tr>
<td>STP</td>
<td>Transfer functions</td>
<td>EA</td>
<td>UK</td>
</tr>
<tr>
<td>DODO</td>
<td>Routing (layered Muskingum)</td>
<td>EA</td>
<td>UK</td>
</tr>
<tr>
<td>MCRM</td>
<td>Rainfall-Runoff</td>
<td>EA</td>
<td>UK</td>
</tr>
<tr>
<td>Modflow99/VKD</td>
<td>3D groundwater</td>
<td>Deltas/Adam Taylor</td>
<td>Netherlands/UK</td>
</tr>
<tr>
<td>Mike11</td>
<td>Hydrodynamics</td>
<td>DHI</td>
<td>Denmark</td>
</tr>
<tr>
<td>NAM</td>
<td>Rainfall-Runoff</td>
<td>DHI</td>
<td>Denmark</td>
</tr>
<tr>
<td>LISFLOOD</td>
<td>Distributed Rainfall-Runoff</td>
<td>JRC</td>
<td>Italy</td>
</tr>
<tr>
<td>TOPKAPI</td>
<td>Rainfall-Runoff</td>
<td>Univ. of Bologna</td>
<td>Italy</td>
</tr>
<tr>
<td>HBV</td>
<td>Rainfall-Runoff (inc snowmelt)</td>
<td>SHMI</td>
<td>Sweden</td>
</tr>
<tr>
<td>Vito</td>
<td>Distributed Rainfall-Runoff</td>
<td>Vieux &amp; Associates</td>
<td>USA</td>
</tr>
<tr>
<td>SWMM</td>
<td>Urban Rainfall-Runoff</td>
<td>USGS</td>
<td>USA</td>
</tr>
<tr>
<td>HEC-RAS</td>
<td>Hydrodynamics</td>
<td>USACE</td>
<td>USA</td>
</tr>
<tr>
<td>Snow17</td>
<td>SnowMelt</td>
<td>NWS</td>
<td>USA</td>
</tr>
<tr>
<td>SACS/M</td>
<td>Rainfall-Runoff</td>
<td>NWS</td>
<td>USA</td>
</tr>
<tr>
<td>Unit/H</td>
<td>Unit Hydrograph</td>
<td>NWS</td>
<td>USA</td>
</tr>
<tr>
<td>PRMS</td>
<td>Rainfall-Runoff</td>
<td>Univ. of Karlsruhe</td>
<td>Germany</td>
</tr>
<tr>
<td>SyHP</td>
<td>Hydrodynamics</td>
<td>BIG</td>
<td>Germany</td>
</tr>
<tr>
<td>SOBEK</td>
<td>Hydrodynamics, Water Quality, RR</td>
<td>Deltas</td>
<td>Netherlands</td>
</tr>
<tr>
<td>SOBEK-2d</td>
<td>Linked 1d/2d inundation modelling</td>
<td>Deltas</td>
<td>Netherlands</td>
</tr>
<tr>
<td>Sacramento</td>
<td>Rainfall-Runoff</td>
<td>Deltas</td>
<td>Netherlands</td>
</tr>
<tr>
<td>RIBASIM</td>
<td>Water distribution + Reservoir</td>
<td>Deltas</td>
<td>Netherlands</td>
</tr>
<tr>
<td>REW</td>
<td>Distributed Rainfall-Runoff</td>
<td>Deltas</td>
<td>Netherlands</td>
</tr>
<tr>
<td>DELFT3D</td>
<td>2/3D Hydrodynamics/Water quality</td>
<td>Deltas</td>
<td>Netherlands</td>
</tr>
<tr>
<td>TWAM</td>
<td>2D Hydrodynamics</td>
<td>PlanB</td>
<td>UK</td>
</tr>
</tbody>
</table>
What is OpenDA

OpenDA is an open source toolbox for data assimilation and parameter calibration in a generic modeling context.

It encompasses:

• An architecture for applying (stochastic) data assimilation algorithms to deterministic models
• A set of interfaces that define interactions between components
• A library of data assimilation and calibration methods:
  • ensemble KF, ensemble square root KF, 3DVar, …
  • Dud, Simplex, Powell, Conjugate Gradient, …
What is OpenDA

- Open source (LGPL)
- Written in Java / C / Fortran
- Current version: OpenDA 2.1
- Available for Windows, Linux & Mac
- Website: [www.openda.org](http://www.openda.org) with downloads, documentation, support
- The OpenDA Association:
Application in of OpenDA in Delft-FEWS
Coupling with models; without openDA

1. input
2. simulation
3. output

Pre adapter

Post adapter

import

export

execute

Delft-FEWS

simulation model

General Adapter
Coupling with models; via openDA blackbox

1. input
2. simulation
3. output
Coupling with models; via openDA dll (in memory)

Delft-FEWS

export
General Adapter
import

openDAinFEWS

openDA
OpenDA model wrapper

Pre adapter
Post adapter

simulation model

Ensemble-KalmanFilter
KalmanFilter-SteadyState
3 examples Delft-FEWS-OpenDA-Model

OpenStreams
Distributed Hydrologic Model
(Belgium – Ourthe catchment)

EFDC 2D Water Quality Model
(Korea – Han River)
Example OpenStreams (distributed hydrological model)

- Goal: More accurate flood forecasts
- Current Status and setup:
  - HBV-96 OpenStreams model available for the Ourthe (Belgium)
  - Python model OpenDA wrapped via Java Embedded Python (JEP)
  - Perturbations applied to precipitation fields (temporal and spatial correlated if needed)
  - Assimilation of measured discharges at outlet (measurements at intermediate locations is also possible or used as verification)

- Future:
  - Ready for operational testing in Dutch operational system RWsOS Rivieren, FEWS Taiwan or any other Delft-FEWS flood and water management forecasting system;
  - Next step using satellite observations (e.g. soil moisture, snow, flooded areas etc)
Example OpenStreams (distributed hydrological model)

updated (EnKF)

forecasted

open loop

measurements
Example EFDC (Environment Fluid Dynamics Code)

- Goal: Improved accuracy Water Quality Forecasts
- Current Status and setup:
  - EFDC 2D-model available for Han River (Korea)
  - Model OpenDA wrapped in memory (.dll/.so)
  - Pertubations applied to radiation, WQ inflows river (temporal correlated)
  - Assimilation of measured WQ parameters (PO$_4$ or algae) at various locations along the river

- Future:
  - Ready for operational testing in FEWS NIER or any other FEWS / EFDC users;
  - Possible use satellite measurements algae blooms, water temperature;
Example EFDC: Implementation in an operational forecasting system

Observation ■, Forecast ◢, Analysis ◙ and Ensemble statistics

Real-time data assimilation of PO$_4$

여주보
Together with Riverside, NCAR, Deltares USA investigate the feasibility of guided (ensemble based) DA for operational forecasting in the US Tennessee River basin.

Main goal: Guide the manual modification process and realize considerable time gains.
• Delft-FEWS is wrapped (via piwebservice) in OpenDA to enable parallel execution of FEWS workflows (and the chained models within these workflows, including SACSMA, UNITHG and LAGK) in a DA framework;

• Multiple DA algorithms configured for testing
  • EnKF, AEnKF, PF, (DEnKF, EnSR,…), etc

• Verification results for a variety of basins will be presented at AGU 2014;
Concluding remarks

• A lot of research has been conducted the last couple of years on the topic of Hydrologic Ensemble DA;

• Operational ensemble based (guided) DA is becoming feasible and the obstacles for usage are only institutional (e.g. resources, forecasts process, other priorities, etc);

• Community-based generic modeling/DA tools are needed and a useful tool also to cross the bridge between research and operations;