Hydro-DART: Ensemble Streamflow Data Assimilation using WRF-Hydro and DART

Application to Hurricane Florence

Moha Gharamti
https://dart.ucar.edu/
gharamti@ucar.edu

Date: Nov. 20, 2020

National Center for Atmospheric Research
Data Assimilation Research Section (DAReS) - TDD - CISL
Motivation
Hurricane Florence

- Tropical wave \(\leadsto\) tropical storm \(\leadsto\) Category 4 Hurricane
- Landfall on Sep. 14 (Carolinas) with winds up to 150 mph
- Catastrophic damages to coastal communities [\$25 billion]
- Flooding magnitude greatly exceeded the levels observed due to Hurricane Matthew in 2016
Hurricane Florence

- Tropical wave $\rightarrow$ tropical storm $\rightarrow$ Category 4 Hurricane
- Landfall on Sep. 14 (Carolinas) with winds up to 150 mph
- Catastrophic damages to coastal communities [$25$ billion]
- Flooding magnitude *greatly exceeded* the levels observed due to Hurricane Matthew in 2016
The goal is to interface the Data Assimilation Research Testbed (DART; Anderson, 2003) with WRF-Hydro (NOAA’s NWM; Gochis, 2020) to enhance flood prediction during Hurricane Florence.
The goal is to interface the Data Assimilation Research Testbed (DART; Anderson, 2003) with WRF-Hydro (NOAA’s NWM; Gochis, 2020) to enhance flood prediction during Hurricane Florence.

- Regional subdomain of the NWM CONUS
- NWM channel network based on NHDPlus v.2
- ~ 67K reaches
- Hourly streamflow assimilation
- 107 USGS gauges
- EAKF: 80 members
THE COUPLED HYDROLOGIC-ASSIMILATION FRAMEWORK
The Hydrologic Model

Channel + Bucket Configuration:

- **Streamflow Model:** Muskingum-Cunge hydrograph routing
- **Groundwater Bucket Model:** Mitigate baseflow deficiencies
The Hydrologic Model

Channel + Bucket Configuration:

- **Streamflow Model**: Muskingum-Cunge hydrograph routing
- **Groundwater Bucket Model**: Mitigate baseflow deficiencies

![Diagram of the Hydrologic Model](image)
The Hydrologic Model

Channel + Bucket Configuration:

- **Streamflow Model**: Muskingum-Cunge hydrograph routing
- **Groundwater Bucket Model**: Mitigate baseflow deficiencies

Full model run from 2010-10-01 to 2018-07-01

Deterministic NWM model chain from forcing through aggregation

Beyond 2010-07-01: NWM operational analysis
The Hydrologic Model

Channel + Bucket Configuration:

- **Streamflow Model:** Muskingum-Cunge hydrograph routing
- **Groundwater Bucket Model:** Mitigate baseflow deficiencies

Full model run from 2010-10-01 to 2018-07-01

Deterministic NWM model chain from forcing through aggregation

One-way runoff fluxes used as input forcing to the channel+bucket sub-model

- **Streamflow Data Assimilation System**
  - **DART**
  - **USGS streamflow observations**
  - **Streamflow (cms)**
  - **Bucket head (m)**

DA period: 2018-09-07 → 2018-10-08
Forcing and Ensemble Uncertainty

- Apply Gaussian perturbations to the boundary fluxes to the streamflow and bucket models every hourly forecast step.

- To create realistic model variability, we follow a "multi-physics" approach ([Berner et al., 2011](#)) and perturb the channel parameters:

  1. top width, $T$
  2. bottom width, $B$
  3. side slope, $m$
  4. Manning’s N, $n$
  5. width of compound channel, $T_{cc}$
  6. Manning’s N of compound channel, $n_{cc}$

Sampling uniformly under some physical constraints!
Apply Gaussian perturbations to the boundary fluxes to the streamflow and bucket models every hourly forecast step.

To create realistic model variability, we follow a "multi-physics" approach (Berner et al., 2011) and perturb the channel parameters:

1. top width, $T$
2. bottom width, $B$
3. side slope, $m$
4. Manning’s N, $n$
5. width of compound channel, $T_{cc}$
6. Manning’s N of compound channel, $n_{cc}$

Sampling uniformly under some physical constraints!
Serial DA scheme: process observations one after the other


How to mitigate typical filtering issues?

i. Sampling Errors: due to limited ensemble size

\[ x_{j,k}^{a(i)} = x_{j,k}^{f(i)} + \alpha \Delta x_{j}^{(i)}; \quad j, k, i : \{\text{space, time, ensemble}\} \]

→ Along-The-Stream (ATS) Localization \([0 < \alpha < 1]\)

ii. Model Biases: e.g., physical parameters, boundary conditions, ...  

\[ x_{j}^{f|a(i)} = \sqrt{\lambda} \left( x_{j}^{f|a(i)} - \bar{x}_{j}^{f|a} \right) + \bar{x}_{j}^{f|a}; \quad f|a : \{\text{forecast or analysis}\} \]

→ Spatially and Temporally Varying Adaptive Inflation \([\sqrt{\lambda} \geq 1]\)
Serial DA scheme: process observations one after the other


How to mitigate typical filtering issues?

i. **Sampling Errors:** due to limited ensemble size

\[ x_{j,k}^{a(i)} = x_{j,k}^{f(i)} + \alpha \Delta x_{j}^{(i)} ; \quad j, k, i : \{\text{space, time, ensemble} \} \]

→ Along-The-Stream (ATS) Localization \[0 < \alpha < 1\]

ii. **Model Biases:** e.g., physical parameters, boundary conditions, ...

\[ x_{j}^{f|a(i)} = \sqrt{\lambda} \left( x_{j}^{f|a(i)} - \bar{x}_{j}^{f|a} \right) + \bar{x}_{j}^{f|a} ; \quad f|a : \{\text{forecast or analysis} \} \]

→ Spatially and Temporally Varying Adaptive Inflation \[\sqrt{\lambda} \geq 1\]
DART: The Data Assimilation Research Testbed

- Serial DA scheme: process observations one after the other

### How to mitigate typical filtering issues?

#### i. Sampling Errors:

Due to limited ensemble size

\[ x_{j,k}^{a(i)} = x_{j,k}^{f(i)} + \alpha \Delta x_{j}^{(i)}; \quad j, k, i : \{ \text{space, time, ensemble} \} \]

→ Along-The-Stream (ATS) Localization \[0 < \alpha < 1\]

#### ii. Model Biases:

E.g., physical parameters, boundary conditions, ...

\[ x_{j}^{f|a(i)} = \sqrt{\lambda} \left( x_{j}^{f|a(i)} - \bar{x}_{j}^{f|a} \right) + \bar{x}_{j}^{f|a}; \quad f|a : \{ \text{forecast or analysis} \} \]

→ Spatially and Temporally Varying Adaptive Inflation \[\sqrt{\lambda} \geq 1\]
DART: The Data Assimilation Research Testbed

- Serial DA scheme: process observations one after the other

How to mitigate typical filtering issues?

i. **Sampling Errors**: due to limited ensemble size

\[
x^{a(i)}_{j,k} = x^{f(i)}_{j,k} + \alpha \Delta x^{(i)}_j; \quad j, k, i : \{\text{space, time, ensemble}\}
\]

→ Along-The-Stream (ATS) Localization \[0 < \alpha < 1\]

ii. **Model Biases**: e.g., physical parameters, boundary conditions, ...

\[
x^{f|a(i)}_j = \sqrt{\lambda} \left( x^{f|a(i)}_j - \bar{x}^{f|a}_j \right) + \bar{x}^{f|a}_j; \quad f|a : \{\text{forecast or analysis}\}
\]

→ Spatially and Temporally Varying Adaptive Inflation \[\sqrt{\lambda} \geq 1\]
Serial DA scheme: process observations one after the other


### How to mitigate typical filtering issues?

**i. Sampling Errors:** due to limited ensemble size

\[
x_{j,k}^{a(i)} = x_{j,k}^{f(i)} + \alpha \Delta x_{j}^{(i)}; \quad j, k, i : \{\text{space, time, ensemble}\}
\]

→ Along-The-Stream (ATS) Localization \([0 < \alpha < 1]\)

**ii. Model Biases:** e.g., physical parameters, boundary conditions, ...

\[
x_{j}^{f|a(i)} = \sqrt{\lambda} \left( x_{j}^{f|a(i)} - \bar{x}_{j}^{f|a} \right) + \bar{x}_{j}^{f|a}; \quad f|a : \{\text{forecast or analysis}\}
\]

→ Spatially and Temporally Varying Adaptive Inflation \([\sqrt{\lambda} \geq 1]\)
Along-The-Stream (ATS) Localization

ATS localization aims to mitigate not only spurious correlations, due to limited ensemble size, but also physically incorrect correlations between unconnected state variables in the river network.
Along-The-Stream (ATS) Localization

ATS localization aims to mitigate not only spurious correlations, due to limited ensemble size, but also physically incorrect correlations between unconnected state variables in the river network.
Along-The-Stream (ATS) Localization

ATS localization aims to mitigate not only spurious correlations, due to limited ensemble size, but also physically incorrect correlations between unconnected state variables in the river network

- □ best performance using 100 km
- □ larger radii give rise to spurious correlations and smaller ones limit the amount of useful information
- □ G-C outperforms other correlation functions
## ATS vs Regular Localization

<table>
<thead>
<tr>
<th>Tar River at Tarboro (NWIS 02083500)</th>
<th>ATS</th>
<th>Reg 20</th>
<th>Reg 10</th>
<th>Reg 5</th>
<th>Reg 2</th>
<th>Reg 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior RMSE</td>
<td>5.58</td>
<td>18.54</td>
<td>8.86</td>
<td>33.46</td>
<td>41.61</td>
<td>34.32</td>
</tr>
<tr>
<td>Posterior RMSE</td>
<td>4.93</td>
<td>17.82</td>
<td>6.75</td>
<td>25.11</td>
<td>33.66</td>
<td>26.41</td>
</tr>
<tr>
<td>Prior Bias</td>
<td>-1.13</td>
<td>-11.65</td>
<td>-1.71</td>
<td>-20.24</td>
<td>-18.09</td>
<td>-11.07</td>
</tr>
<tr>
<td>Posterior Bias</td>
<td>-0.85</td>
<td>-11.41</td>
<td>-0.74</td>
<td>-20.37</td>
<td>-17.16</td>
<td>-10.01</td>
</tr>
<tr>
<td>Prior Spread</td>
<td>1.20</td>
<td>3.29</td>
<td>2.80</td>
<td>10.90</td>
<td>10.84</td>
<td>9.54</td>
</tr>
<tr>
<td>Posterior Spread</td>
<td>1.55</td>
<td>3.00</td>
<td>2.27</td>
<td>6.28</td>
<td>6.43</td>
<td>5.17</td>
</tr>
</tbody>
</table>

- Performance using ATS localization is significantly better (~ 40%)
- Using ATS, one can increase the effective localization radius
- Regular localization with large radii fails (correlating physically unrelated variables)
Adaptive Covariance Inflation

The algorithm is adaptive in time, based on Bayes’ theorem, and results in spatially varying fields (El Gharamti, 2018):

\[ p \left( \lambda | d^f | a \right) \approx p \left( d^f | a | \lambda \right) \cdot p(\lambda) \]
The algorithm is adaptive in time, based on Bayes’ theorem, and results in spatially varying fields (El Gharamti, 2018):

\[
p\left(\lambda|d^f|a\right) \approx p\left(d^f|a|\lambda\right) \cdot p(\lambda)
\]
Adaptive Covariance Inflation

The algorithm is adaptive in time, based on Bayes’ theorem, and results in spatially varying fields (El Gharamti, 2018):

\[
p(\lambda | d_f | a) \approx p(d_f | a | \lambda) \cdot p(\lambda)
\]

Prior pdf

Inverse Gamma

Posterior pdf

Likelihood

Time-Avg. Streamflow Prior Inflation

Time-Avg. Bucket Prior Inflation
Adaptive Covariance Inflation

The algorithm is adaptive in time, based on Bayes’ theorem, and results in spatially varying fields (El Gharamti, 2018):

\[ p\left(\lambda | d_f^a\right) \approx p\left(d_f^a | \lambda\right) \cdot p(\lambda) \]

- **Posterior pdf**
- **Prior pdf**
- **Inverse Gamma Likelihood**

Large inflation in densely observed areas
After landfall, the model’s streamflow prediction (Open Loop) is significantly smaller than the posterior along Pee Dee River in South Carolina.
After landfall, the model’s streamflow prediction (Open Loop) is significantly smaller than the posterior along Pee-Dee River in South Carolina.
After landfall, the model’s streamflow prediction (Open Loop) is significantly smaller than the posterior along Pee-Dee River in South Carolina.
After landfall, the model’s streamflow prediction (Open Loop) is significantly smaller than the posterior along Pee-Dee River in South Carolina.
After landfall, the model’s streamflow prediction (Open Loop) is significantly smaller than the posterior along Pee-Dee River in South Carolina.
After landfall, the model’s streamflow prediction (Open Loop) is significantly smaller than the posterior along Pee-Dee River in South Carolina.
After landfall, the model’s streamflow prediction (Open Loop) is significantly smaller than the posterior along Pee-Dee River in South Carolina.
After landfall, the model’s streamflow prediction (Open Loop) is significantly smaller than the posterior along Pee-Dee River in South Carolina.

A sizable increase in prior inflation to counter the bias in the modeled streamflow!
The rank histogram for the open loop is heavily skewed to the right indicating that the gauge data is larger than the ensemble.
Bias Mitigation

The rank histogram for the open loop is heavily skewed to the right indicating that the gauge data is larger than the ensemble.

The probability of the observation to fall outside the open loop ensemble is greater than 50%.

The observed discharge is statistically indistinguishable from the prior ensemble.
NOAA’s National Water Model configuration of the WRF-Hydro framework is coupled to the Data Assimilation Research Testbed (DART) to improve ensemble streamflow forecasts under extreme rainfall conditions during Hurricane Florence in Sep. 2018.

To address sampling errors, Along-The-Stream (ATS) Localization is proposed. The algorithm provides improved information propagation in the stream network.

Adaptive Inflation is extremely useful and is able to serve as a vigorous bias correction scheme which varies both spatially and temporally.
NOAA’s National Water Model configuration of the WRF-Hydro framework is coupled to the Data Assimilation Research Testbed (DART) to improve ensemble streamflow forecasts under extreme rainfall conditions during Hurricane Florence in Sep. 2018.

To address sampling errors, Along-The-Stream (ATS) Localization is proposed. The algorithm provides improved information propagation in the stream network.

Adaptive Inflation is extremely useful and is able to serve as a vigorous bias correction scheme which varies both spatially and temporally.
NOAA’s National Water Model configuration of the WRF-Hydro framework is coupled to the Data Assimilation Research Testbed (DART) to improve ensemble streamflow forecasts under extreme rainfall conditions during Hurricane Florence in Sep. 2018.

To address sampling errors, Along-The-Stream (ATS) Localization is proposed. The algorithm provides improved information propagation in the stream network.

Adaptive Inflation is extremely useful and is able to serve as a vigorous bias correction scheme which varies both spatially and temporally.
NOAA’s National Water Model configuration of the WRF-Hydro framework is coupled to the Data Assimilation Research Testbed (DART) to improve ensemble streamflow forecasts under extreme rainfall conditions during Hurricane Florence in Sep. 2018.

To address sampling errors, Along-The-Stream (ATS) Localization is proposed. The algorithm provides improved information propagation in the stream network.

Adaptive Inflation is extremely useful and is able to serve as a vigorous bias correction scheme which varies both spatially and temporally.

https://dart.ucar.edu/

Thank You!