Data Assimilation in Hydrology and Streamflow Forecasting

Hurricane Florence Flooding 2018

Moha Gharamti

https://dart.ucar.edu/

gharamti@ucar.edu

Special thanks to: James Mccreight, Tim Hoar, Seong Jin Noh, Arezoo RafieeiNasab, Ben Johnson, Nancy Collins

EnKF Workshop 2021

National Center for Atmospheric Research

Data Assimilation Research Section (DAReS) - TDD - CISL

CISL Computational & Information Systems Lab

NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

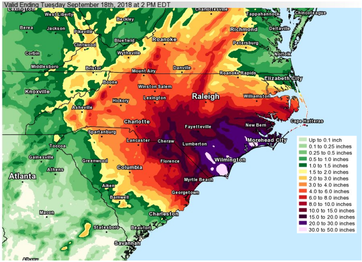
Overview

Motivation

1. Why Streamflow Forecasting?

Hurricane Florence (2018):

- Tropical wave ⇝ tropical storm ⇝ **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 (2016) and Floyd (1999) **combined!**



Rainfall estimates from Hurricane Florence (*Source: NWS*)

Hurricane Florence eye during landfall (*Source: NWS*)

1. Why Streamflow Forecasting?

- Predicting major floods during extreme rainfall events is crucial
  1. Save lives ($\sim$ 50 people died due to Florence Flooding)
  2. Limit damages (via advance warnings)
  3. Protect infrastructure, socio-economic impacts, ...

Flooded city of New Bern, NC

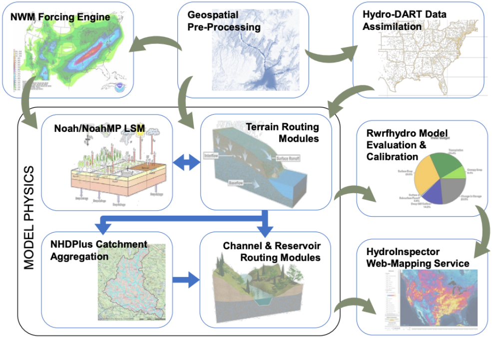
HydroDART: Streamflow Prediction Framework

2.1 The Model: WRF-Hydro

WRF-Hydro [Gochis et al., 2020]: Weather Research and Forecasting hydrological modeling system. Research compartment of the **National Water Model (NWM)**

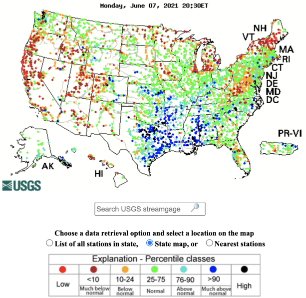
A community-based system, providing:

- Prediction of major water cycle components such as precipitation, soil moisture, snowpack, groundwater, streamflow, inundation
- Reliable streamflow prediction across scales
- A robust framework for land-atmosphere coupling studies



Streamflow (in cfs) simulation over CONUS for the 2019-2020 water year (*Source: NOAA, NWC, NWS*)

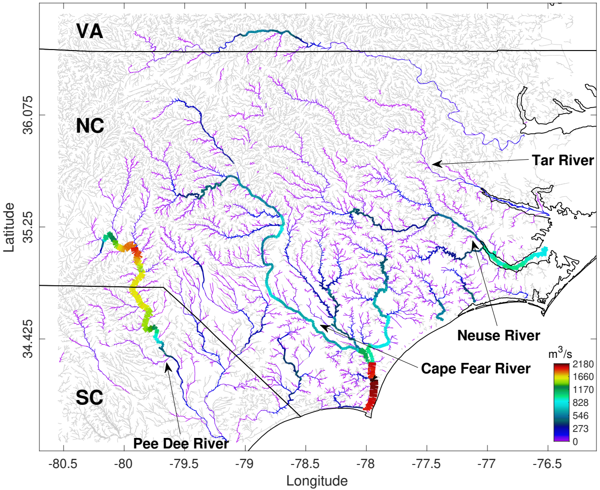
USGS operates one of the largest stream-gauging enterprises in the world (more than 11,000 gauges)



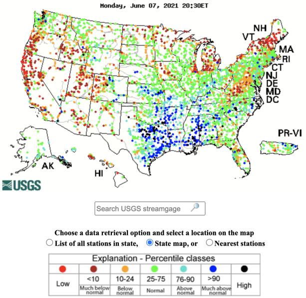
Source: https://waterwatch.usgs.gov/

2.2 The Data

USGS operates one of the largest stream-gauging enterprises in the world (more than 11,000 gauges)



Streamflow map of the Florence domain simulated using WRF-Hydro during the flooding event

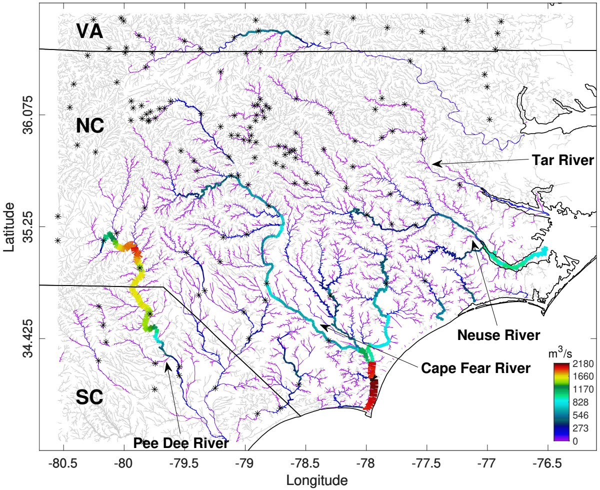


Source: https://waterwatch.usgs.gov/

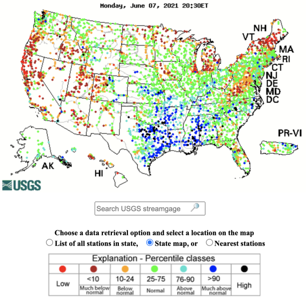
- Regional subdomain of the NWM CONUS
- NWM channel network
- ~ 67K reaches

2.2 The Data

USGS operates one of the largest stream-gauging enterprises in the world (more than 11,000 gauges)



Streamflow map of the Florence domain simulated using WRF-Hydro during the flooding event

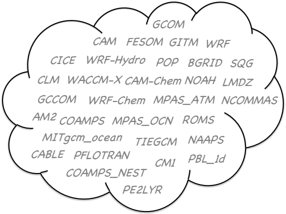


Source: https://waterwatch.usgs.gov/

- Regional subdomain of the NWM CONUS
- NWM channel network
- ~ 67K reaches
- Hourly streamflow assimilation
- 107 USGS gauges

2.3 DA Tool: DART

- A community facility for ensemble DA [Anderson et al., 2008; BAMS], developed and maintained by the Data Assimilation Research Section (DAReS) in CISL at NCAR
  - *Framework:*
    - Flexible, portable, well-tested, extensible, **free!**
    - Source code distributed on GitHub: NCAR/DART
    - Models: Toy to HUGE, including CESM
    - Observations: Real, synthetic, novel
  - *Research:*
    - Theory based, widely applicable techniques
    - Nonlinear filters, nonGaussian approaches
    - Adaptive inflation, Localization, ...
  - *Teaching:* Extensive tutorial materials and exercises



- ~ 50 UCAR member universities & more than 100 other sites
- Collaborations with external partners

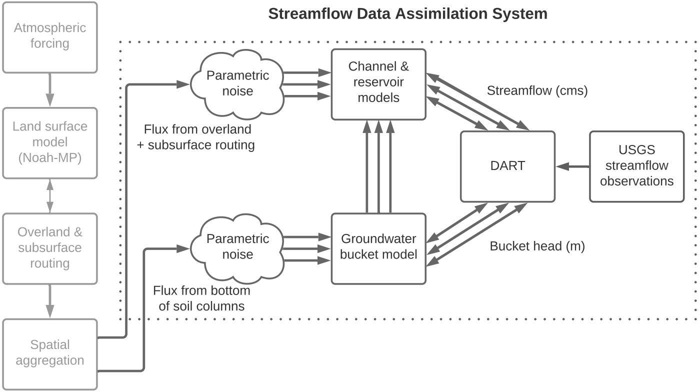
https://dart.ucar.edu/
https://docs.dart.ucar.edu/

2.4 Interfacing WRF-Hydro and DART

$$\mathbf{x} = \begin{bmatrix} x1 : \text{Streamflow} \\ x2 : \text{Bucket} \end{bmatrix}; \qquad \mathbf{y} = \begin{bmatrix} \mathbf{I}, & \mathbf{O} \end{bmatrix} \mathbf{x} + \varepsilon; \qquad N_e = 80$$

2.4 Interfacing WRF-Hydro and DART

$$\mathbf{x} = \begin{bmatrix} x1 : \text{Streamflow} \\ x2 : \text{Bucket} \end{bmatrix}; \qquad \mathbf{y} = \begin{bmatrix} \mathbf{I}, & \mathbf{O} \end{bmatrix} \mathbf{x} + \varepsilon; \qquad N_e = 80$$

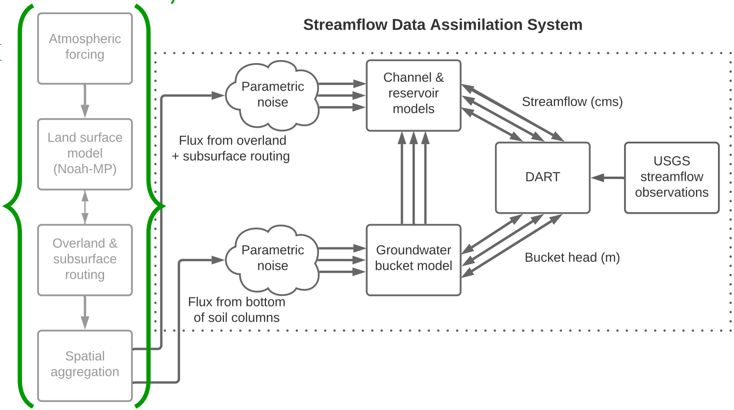


2.4 Interfacing WRF-Hydro and DART

$$\mathbf{x} = \begin{bmatrix} x1 : \text{Streamflow} \\ x2 : \text{Bucket} \end{bmatrix}; \qquad \mathbf{y} = \begin{bmatrix} \mathbf{I}, & \mathbf{O} \end{bmatrix} \mathbf{x} + \varepsilon; \qquad N_e = 80$$

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

Deterministic NWM model chain from forcing through aggregation



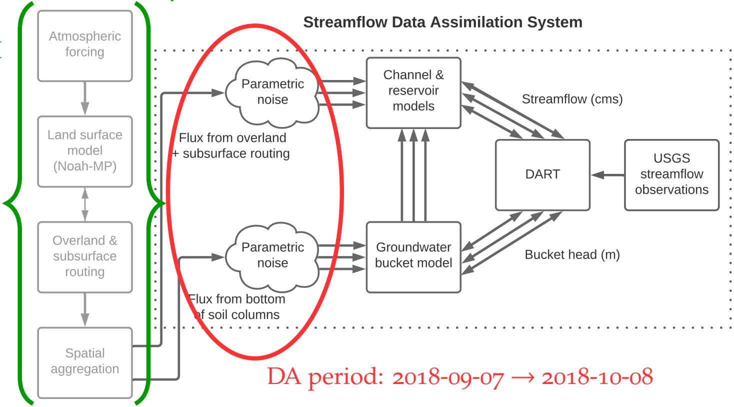
2.4 Interfacing WRF-Hydro and DART

$$\mathbf{x} = \begin{bmatrix} x1 : \text{Streamflow} \\ x2 : \text{Bucket} \end{bmatrix}; \qquad \mathbf{y} = \begin{bmatrix} \mathbf{I}, & \mathbf{O} \end{bmatrix} \mathbf{x} + \varepsilon; \qquad N_e = 80$$

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



DA period: 2018-09-07 → 2018-10-08

2.4 Interfacing WRF-Hydro and DART

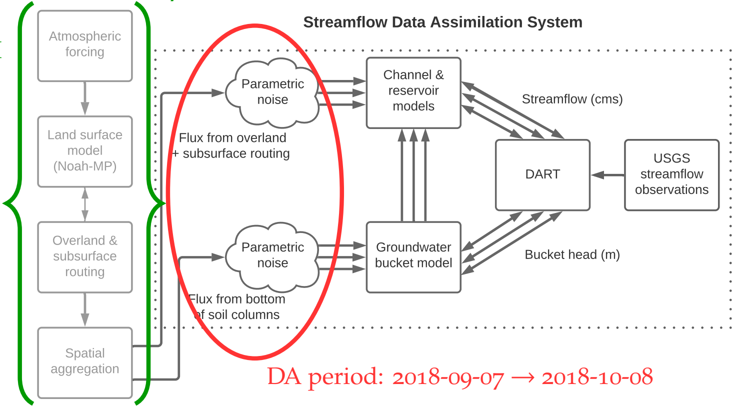
$$\mathbf{x} = \begin{bmatrix} x1 : \text{Streamflow} \\ x2 : \text{Bucket} \end{bmatrix}; \qquad \mathbf{y} = \begin{bmatrix} \mathbf{I}, & \mathbf{O} \end{bmatrix} \mathbf{x} + \varepsilon; \qquad N_e = 80$$

https://github.com/NCAR/wrf_hydro_nwm_public
https://github.com/NCAR/wrf_hydro_dart
https://github.com/NCAR/DART

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



DA period: 2018-09-07 → 2018-10-08

DA Enhancements & Results

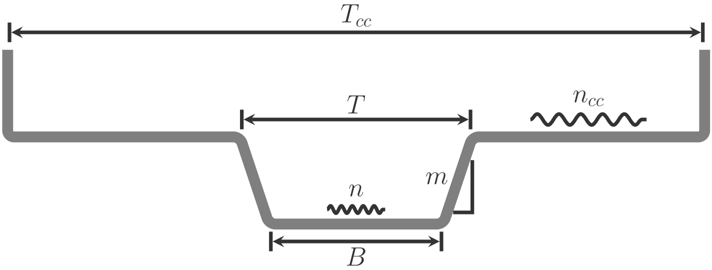
3.1 Forcing and Ensemble Uncertainty

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

3.1 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-configuration" approach 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!



3.2.1 Along-The-Stream (ATS) Localization

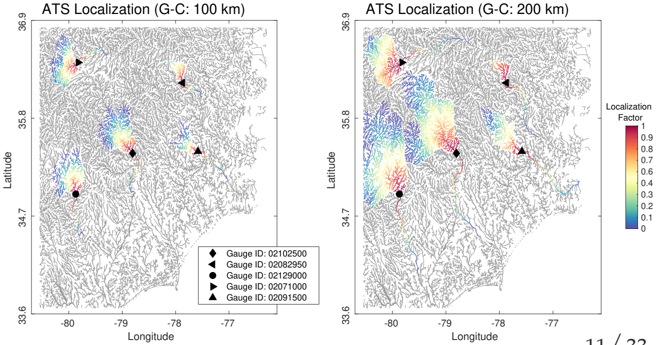
$$\mathbf{x}_{j,k}^{a(i)} = \mathbf{x}_{j,k}^{f(i)} + \alpha \Delta x_j^{(i)} \qquad 0 < \alpha < 1 \qquad \text{(Localization Factor)}$$

- Small ensemble sizes produce imperfect sample covariances [Houtekamer and Mitchell, 2001; MWR], yielding spurious correlations

3.2.1 Along-The-Stream (ATS) Localization

$$\mathbf{x}_{j,k}^{a(i)} = \mathbf{x}_{j,k}^{f(i)} + \alpha \Delta x_j^{(i)} \qquad 0 < \alpha < 1 \qquad \text{(Localization Factor)}$$

- Small ensemble sizes produce imperfect sample covariances [Houtekamer and Mitchell, 2001; MWR], yielding spurious correlations
- ATS localization [El Gharamti et al., 2020; HESS] 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**

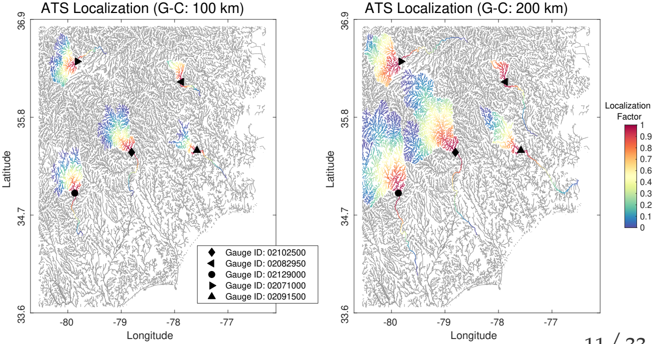


3.2.1 Along-The-Stream (ATS) Localization

$$\mathbf{x}_{j,k}^{a(i)} = \mathbf{x}_{j,k}^{f(i)} + \alpha \Delta x_j^{(i)} \qquad 0 < \alpha < 1 \qquad \text{(Localization Factor)}$$

Functionality/Characteristics:

1. Flow of information only travels downstream (tree-like shapes)
2. Total number of close reaches depend on the size of the basin
3. Observations in different catchments do not have common close reaches

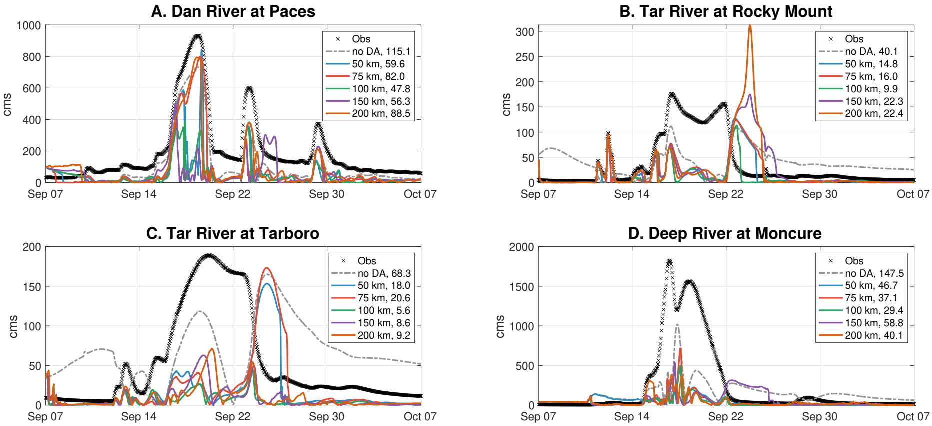


3.2.2 Does regular localization even work?

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

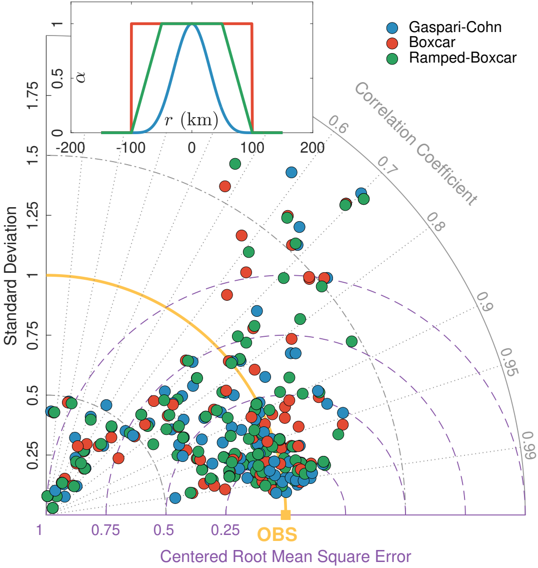
- Performance using ATS localization is significantly better ($\sim$ 40%)
- Using ATS, one can increase the effective localization radius
- Regular localization with large radii fails (correlating physically unrelated variables)

3.2.3 Tuning ATS Localization; [i] Radius



- Test with different localization radii: 50, 75, 100, 150, 200 km
- Larger radii degrade the accuracy (giving rise to spurious correlations)
- Smaller radii limit the amount of useful information
- Best performance with 100 km

3.2.4 Tuning ATS Localization; [ii] Correlation Function



- Averaging over all gauges, the correlation coefficient was: Gaspari-Cohn (**0.83**), Boxcar (**0.77**) and Ramped-Boxcar (**0.79**)
- Gaspari-Cohn outperforms other functions

3.3.1 Dealing with Variance Underestimation

- Variance underestimation often happens in ensemble-based systems due to sampling errors and model biases

3.3.1 Dealing with Variance Underestimation

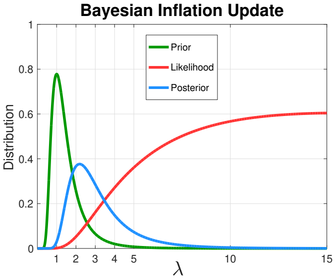
- Variance underestimation often happens in ensemble-based systems due to sampling errors and model biases
- Spatially and Temporally Varying Adaptive Covariance Inflation [Anderson 2009; Tellus], [El Gharamti 2018; El Gharamti et al. 2019; MWR]:
  1. Assume inflation factor, λ to be a random variable
  2. Use the data to estimate λ at every point in the domain

3.3.1 Dealing with Variance Underestimation

- Variance underestimation often happens in ensemble-based systems due to sampling errors and model biases
- Spatially and Temporally Varying Adaptive Covariance Inflation [Anderson 2009; Tellus], [El Gharamti 2018; El Gharamti et al. 2019; MWR]:
  1. Assume inflation factor, λ to be a random variable
  2. Use the data to estimate λ at every point in the domain
  3. Bayes' rule: $p(\lambda|d) \approx p(\lambda) \cdot p(d|\lambda)$

     Prior $p(\lambda)$; an Inverse Gamma pdf
     Likelihood $p(d|\lambda)$; a Gaussian function
     - $d = |y^o - \overline{\mathbf{x}}_j^f|$ is the innovation
     - Derosiers Innovation statistics: $\mathbb{E}(d) = 0; \quad \mathbb{E}(d^2) = \sigma_o^2 + \lambda\sigma_f^2$

     Posterior $p(\lambda|d)$
  4. Can inflate prior or posterior covariance!

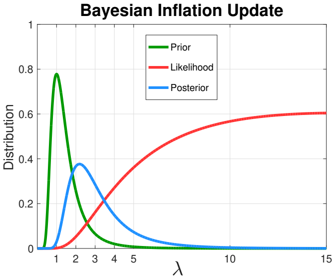


3.3.1 Dealing with Variance Underestimation

- Variance underestimation often happens in ensemble-based systems due to sampling errors and model biases
- Spatially and Temporally Varying Adaptive Covariance Inflation [Anderson 2009; Tellus], [El Gharamti 2018; El Gharamti et al. 2019; MWR]:
  1. Assume inflation factor, λ to be a random variable
  2. Use the data to estimate λ at every point in the domain
  3. Bayes' rule: $p(\lambda|d) \approx p(\lambda) \cdot p(d|\lambda)$

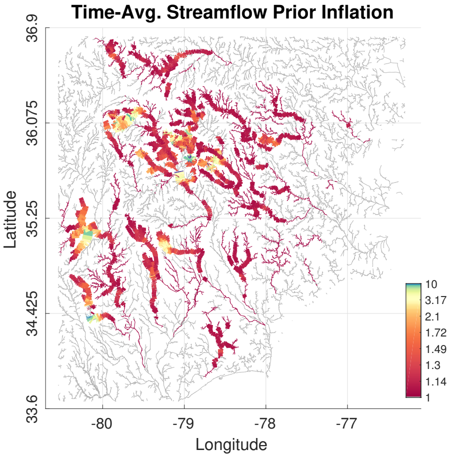
     Prior $p(\lambda)$; an Inverse Gamma pdf
     Likelihood $p(d|\lambda)$; a Gaussian function
     - $d = |y^o - \overline{\mathbf{x}}_j^f|$ is the innovation
     - Derosiers Innovation statistics:
       $\mathbb{E}(d) = 0; \quad \mathbb{E}(d^2) = \sigma_o^2 + \lambda\sigma_f^2$

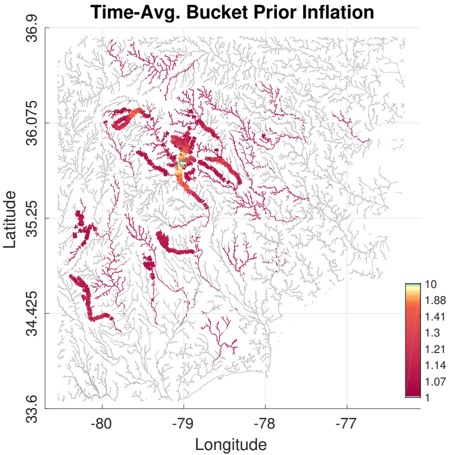
     Posterior $p(\lambda|d)$
  4. Can inflate prior or posterior covariance!



- Other forms [Raanes et al. 2019; QJRMS], [Tandeo et al. 2020; MWR]

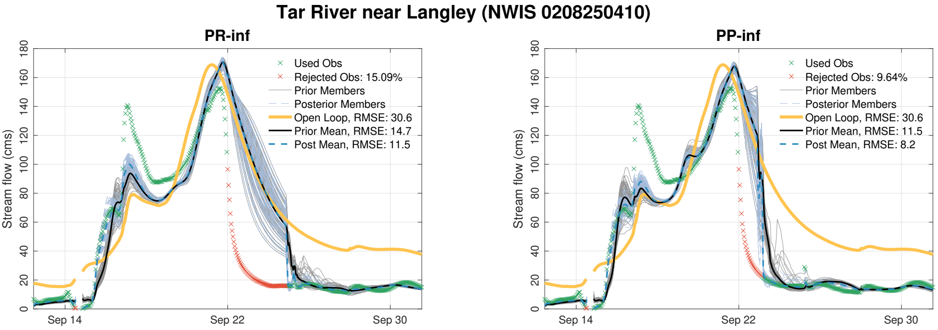
3.3.2 Inflation on the River Network





- Inflation follows tree-like shapes thanks to ATS localization
- Larger inflation in densely observed regions

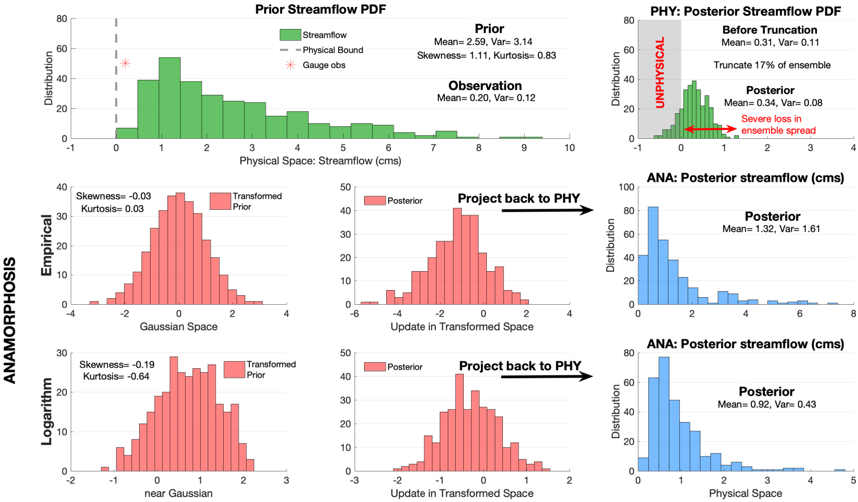
3.3.3 Combining Prior and Posterior Inflation



- Posterior inflation **does not** yield satisfactory results
- Adding posterior inflation on top of prior inflation helps improve accuracy
- Falling limb of hydrograph (PP-inf) better fits the data. Recession happens almost 2 days earlier (rejects less data)
- May argue that posterior inflation could be resolving other regression issues such as sampling noise and nonGaussianity

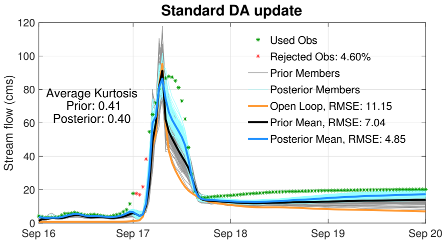
3.4 Gaussian Anamorphosis

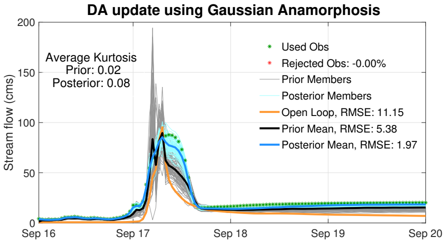
Streamflow is a positive quantity. Make sure the DA framework produces physically meaningful updates using GA [Simon and Bertino 2009; OS]



3.4 Gaussian Anamorphosis

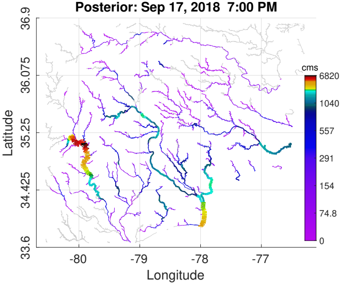
Streamflow is a positive quantity. Make sure the DA framework produces physically meaningful updates using GA [Simon and Bertino 2009; OS]

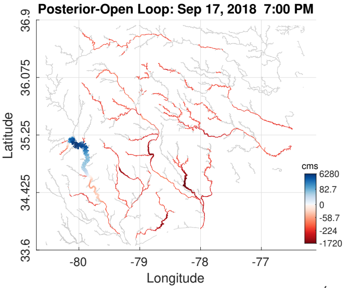




- Observation rejection is improved with GA
- Better fit to the observations on Sep. 17$^{\text{th}}$
- Higher order moments are almost completely eliminated using GA

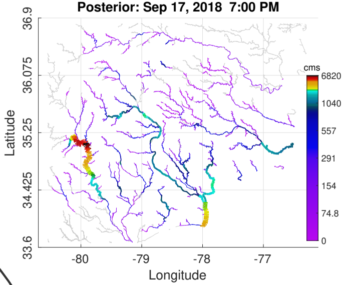
After landfall, the model's streamflow prediction (Open Loop) is significantly smaller than the posterior along Pee-Dee River in South Carolina

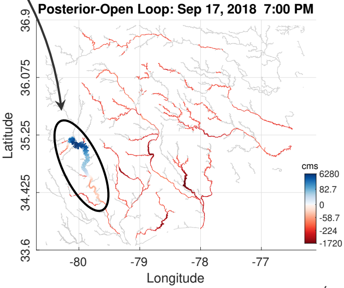




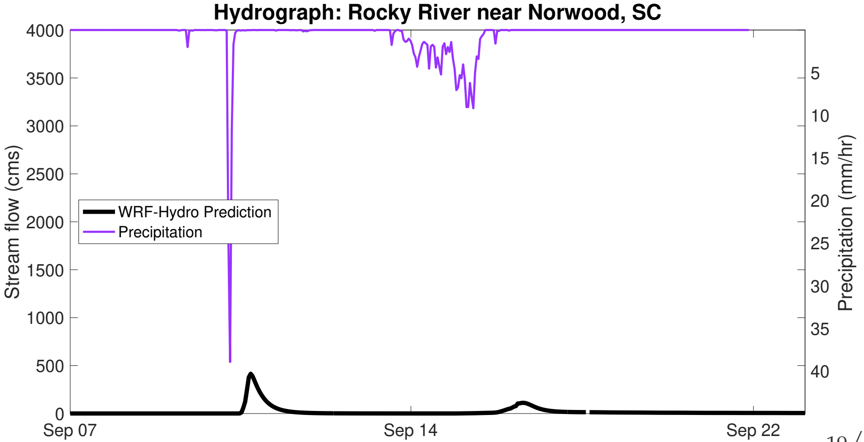
3.5 Bias Mitigation

After landfall, the model's streamflow prediction (Open Loop) is significantly smaller than the posterior along Pee-Dee River in South Carolina



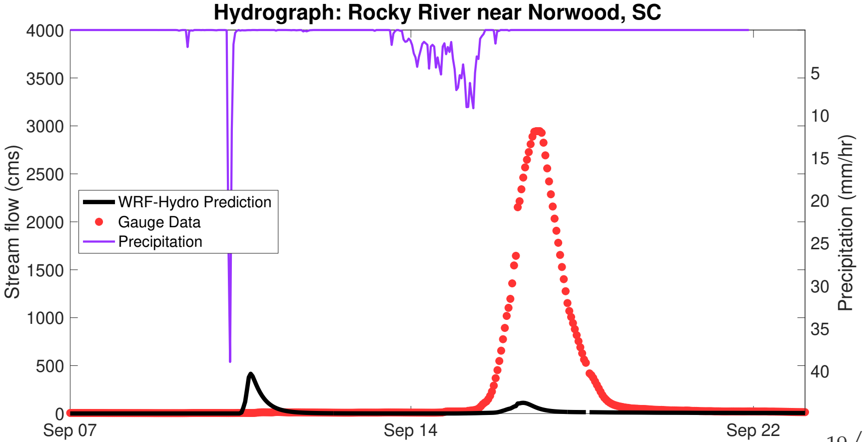


After landfall, the model's streamflow prediction (Open Loop) is significantly smaller than the posterior along Pee-Dee River in South Carolina



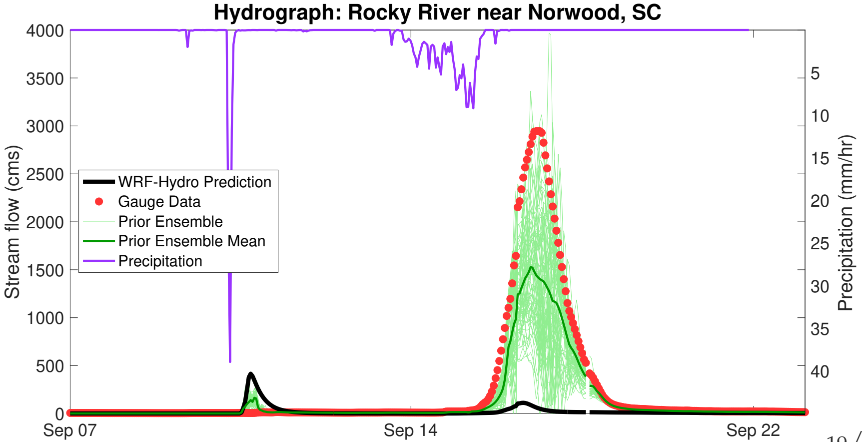
3.5 Bias Mitigation

After landfall, the model's streamflow prediction (Open Loop) is significantly smaller than the posterior along Pee-Dee River in South Carolina



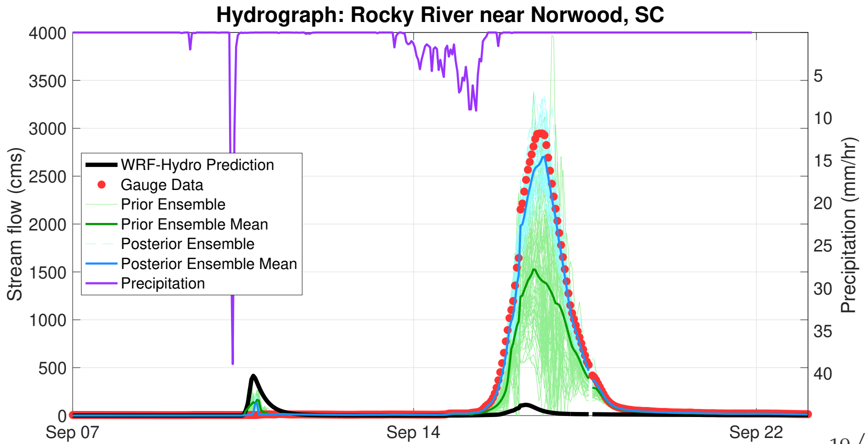
3.5 Bias Mitigation

After landfall, the model's streamflow prediction (Open Loop) is significantly smaller than the posterior along Pee-Dee River in South Carolina

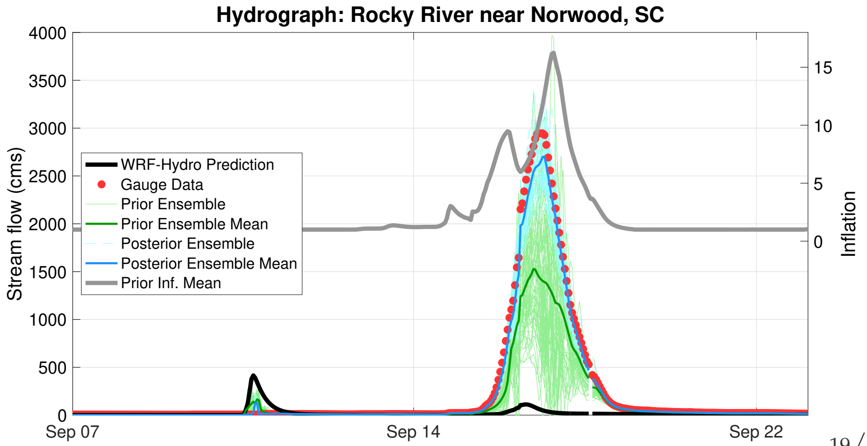


3.5 Bias Mitigation

After landfall, the model's streamflow prediction (Open Loop) is significantly smaller than the posterior along Pee-Dee River in South Carolina

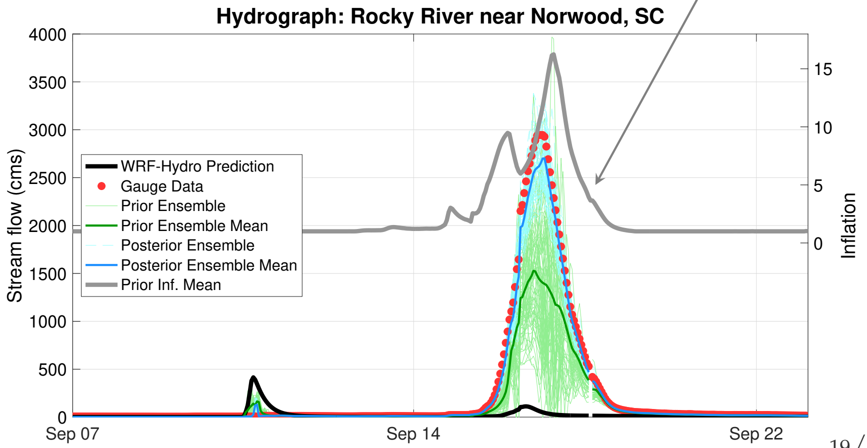


After landfall, the model's streamflow prediction (Open Loop) is significantly smaller than the posterior along Pee-Dee River in South Carolina



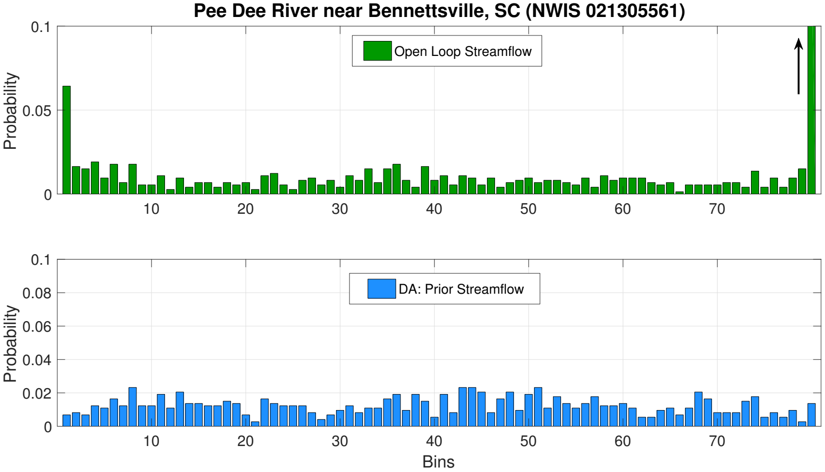
After landfall, the model's streamflow prediction (Open Loop) is significantly smaller than the posterior along Pee-Dee River in South Carolina

A sizable increase in prior inflation to counter the bias in the modeled streamflow!



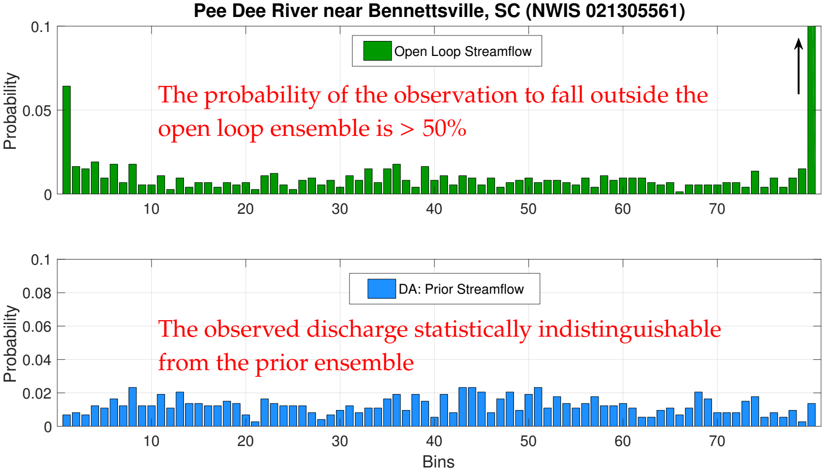
3.5 Bias Mitigation

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



3.5 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



Conclusion

4.1 Summary

- HydroDART: an advanced streamflow DA tool
  1. Provides hourly skillful streamflow estimates
  2. Enhanced ensemble uncertainty assessment
  3. Introduces Along-The-Stream localization
  4. Supports a variety of DA algorithms: e.g., Adaptive Inflation
  5. Supports *parameter (model + hyper) estimation*

4.1 Summary

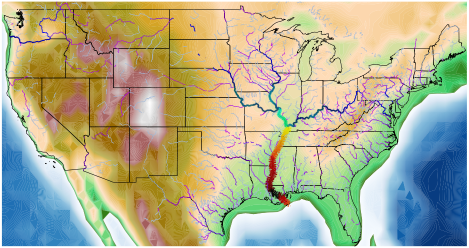
- HydroDART: an advanced streamflow DA tool
  1. Provides hourly skillful streamflow estimates
  2. Enhanced ensemble uncertainty assessment
  3. Introduces Along-The-Stream localization
  4. Supports a variety of DA algorithms: e.g., Adaptive Inflation
  5. Supports *parameter (model + hyper) estimation*

- This study made use of HydroDART for Hurricane Florence flooding events (2018) in the Carolinas
  1. Explored effectiveness of ATS localization (as compared to regular localization methods)
  2. Tested inflation scheme in highly nonlinear scenarios. Closer look at prior and posterior inflation strategies

4.2 Future Research Directions

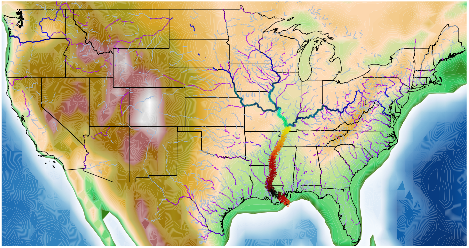
- Full CONUS streamflow reanalysis for the past 30 years:
  $\longrightarrow$ Explore hybrid EnKF-OI approaches:

  Adaptive: [El Gharamti 2021; MWR]
  Analogs: [Grooms 2021; QJRMS]



4.2 Future Research Directions

- Full CONUS streamflow reanalysis for the past 30 years:
  → Explore hybrid EnKF-OI approaches:
  Adaptive: [El Gharamti 2021; MWR]
  Analogs: [Grooms 2021; QJRMS]



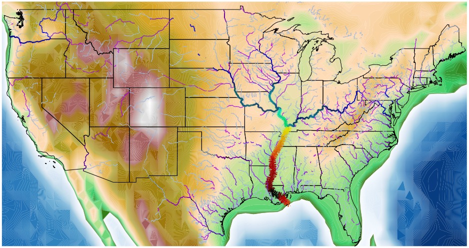
- A collaborative project with USGS; 2 main goals:
  1. Assimilate gauge temperature data (investigate effects on streamflow)
  2. Placement of gauges (OSSE studies)

4.2 Future Research Directions

- Full CONUS streamflow reanalysis for the past 30 years: → Explore hybrid EnKF-OI approaches:

  Adaptive: [El Gharamti 2021; MWR]

  Analogs: [Grooms 2021; QJRMS]



- A collaborative project with USGS; 2 main goals:
  1. Assimilate gauge temperature data (investigate effects on streamflow)
  2. Placement of gauges (OSSE studies)

- Coupling the LSM with WRF-Hydro:
  1. Assimilate soil moisture & streamflow; weak vs strong coupling
  2. Assimilate snow data (thickness, SWE, ...)