Improving Flood Prediction with Cutting-Edge Data Assimilation

Collaborators: Special thanks to Arezoo Rafieeinasab (HAP/RAL), James McCreight (CPAESS/USGS), Tim Hoar (Retired), the entire DAReS team (CISL)

Moha Gharamti

DAReS-TDD-CISL, NSF NCAR

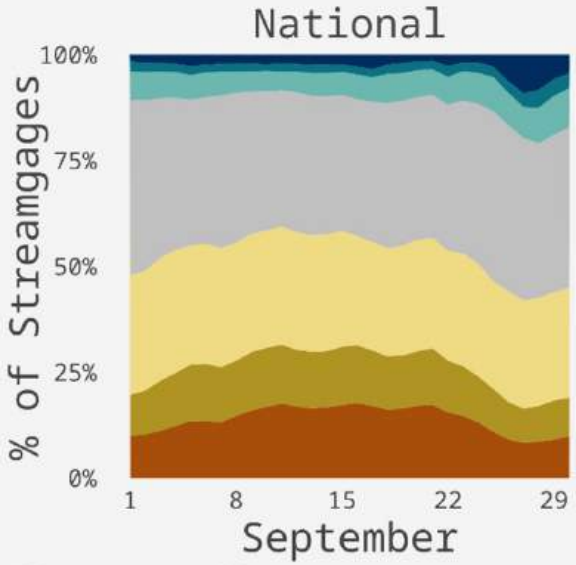
January 21, 2025

This material is based upon work supported by the NSF National Center for Atmospheric Research, which is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement No. 1852977.

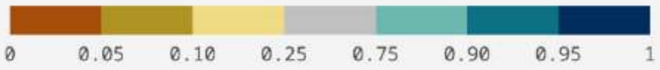
Outline

1. **Flood Prediction: Motivation**
2. **Hydro-DART: The Hydrologic Prediction System**
    1. WRF-Hydro and Data
    2. The Data Assimilation Research Testbed (DART)
    3. Flooding Applications
3. **Novel DA Tools and Algorithms**
    1. The EnKF
    2. Along-The-Stream Localization
    3. Adaptive Covariance Inflation
    4. Hybrid Ensemble-Variational Scheme
4. **Current Activities**
5. **Future Plans**

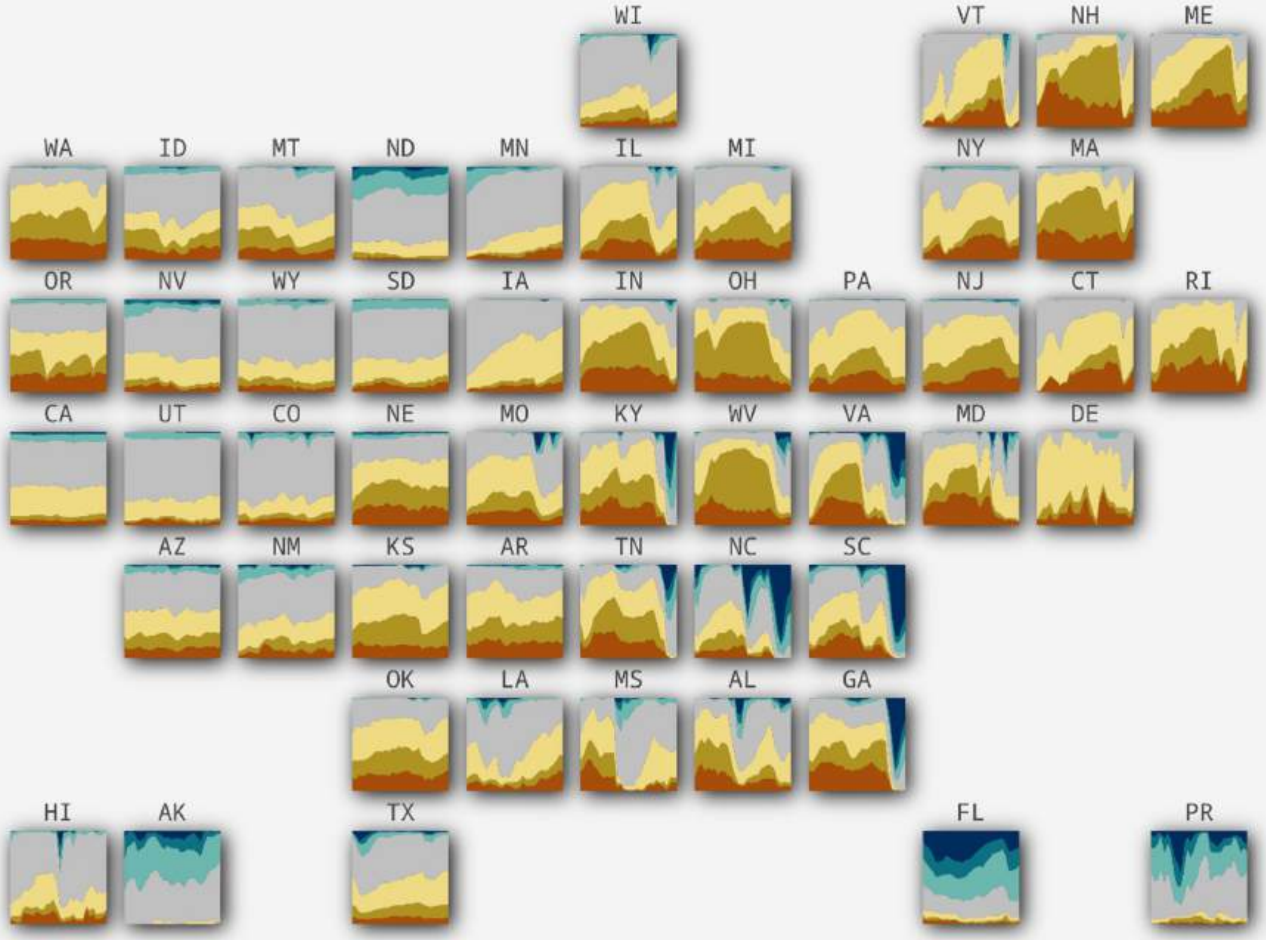
September 2024 STREAMFLOW



Flow percentile at USGS streamgages relative to the historic record.



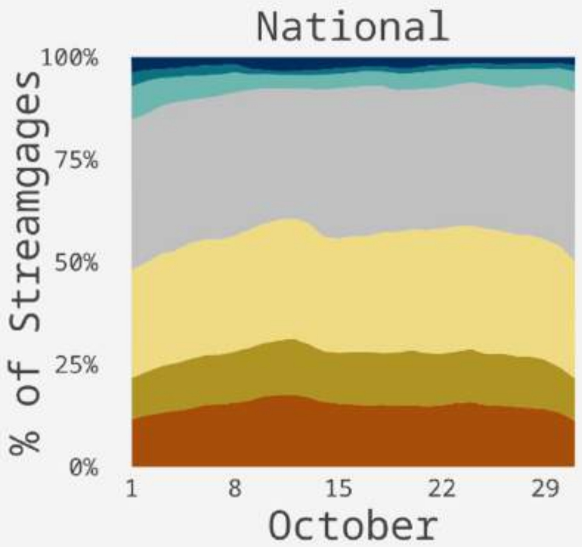
USGS
science for a changing world



Data: USGS Water Data for the Nation

NSF | NCAR

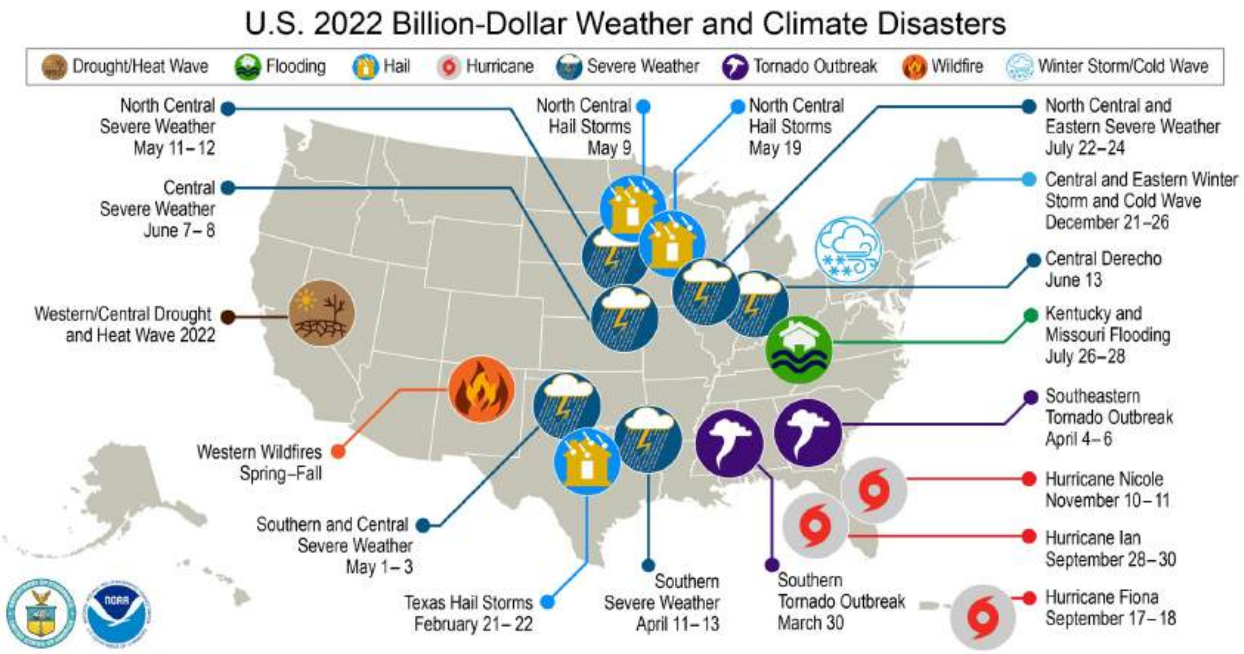
October 2024
STREAMFLOW



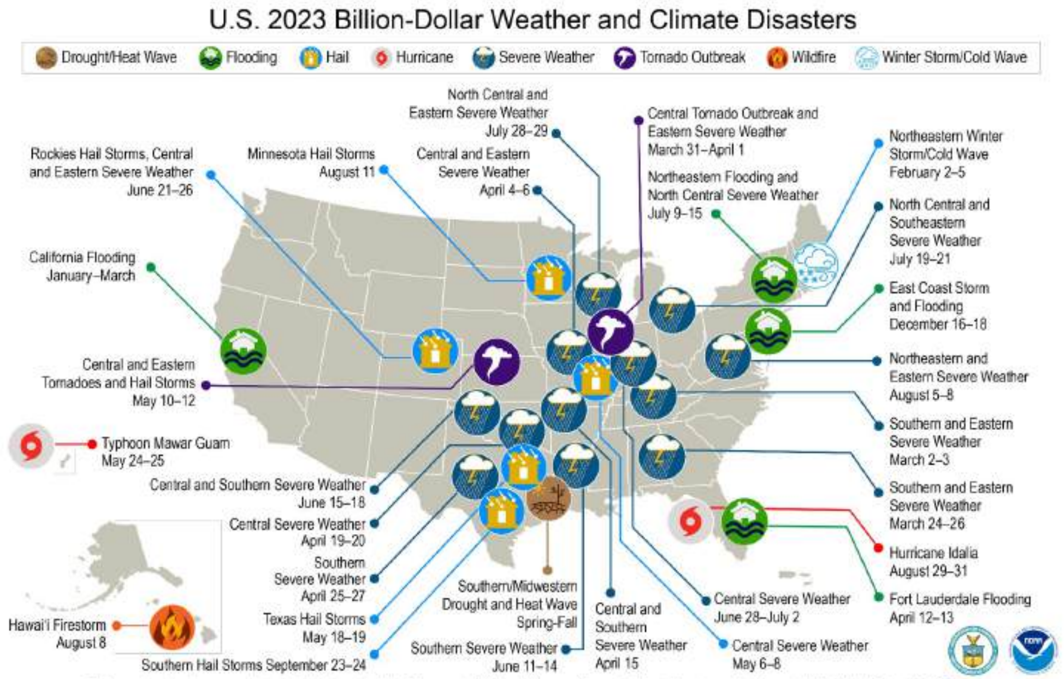
Flow percentile at USGS streamgages relative to the historic record.

USGS science for a changing world

Data: USGS Water Data for the Nation

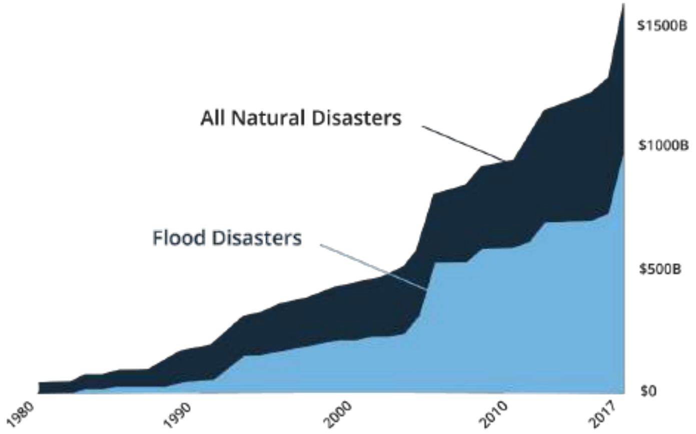


This map denotes the approximate location for each of the ***18 separate billion-dollar weather and climate disasters that impacted the United States in 2022.***



This map denotes the approximate location for each of the ***28 separate billion-dollar weather and climate disasters that impacted the United States in 2023.***

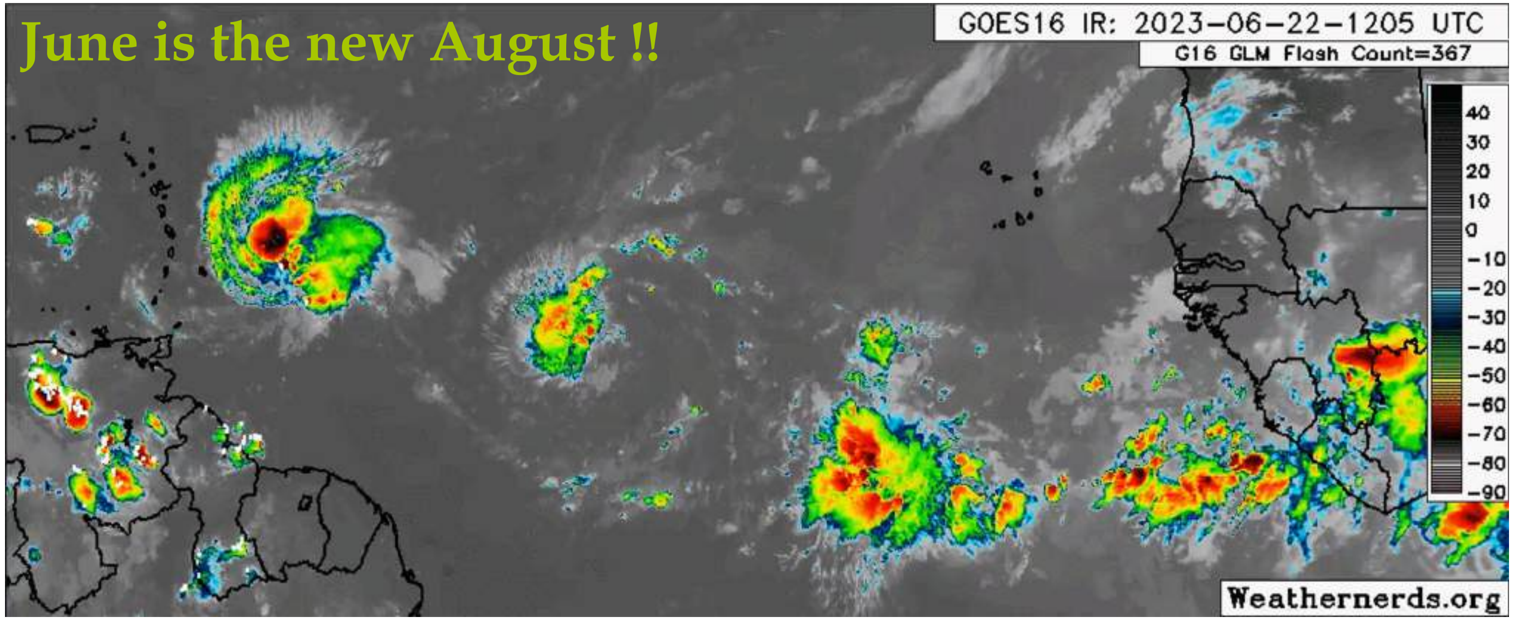
- ❑ Extreme rainfall, hurricanes, and the associated flooding events are consistent recurring disasters
- ❑ **$1 Trillion** since 1980 and **$850B** since 2000
- ❑ For the majority of Americans, ⅔ of their wealth is tied up in their home. 1 flood can wipe out a lifetime of savings!



1. Motivation: Hurricanes and Flooding

- ❑ Subtropical or tropical cyclogenesis in the Atlantic Ocean
- ❑ Saffir-Simpson Scale for hurricane wind speeds
  - ❑ Major: 3-4-5; up to 160 mph
- ❑ **Helene:** Category 4; 140 mph winds; +30 inches of rain; ~200 fatalities

Hurricane Helene: Flood damage near Swannanoa River on Oct 3, 2024 [Getty Images]



Tropical storms west of Africa and in the Atlantic Ocean in June, 2023

- ❑ ~2.5 trillion gallons of rain a day
- ❑ Torrential rain leads to freshwater (inland) flooding
- ❑ Catastrophic damages to infrastructure

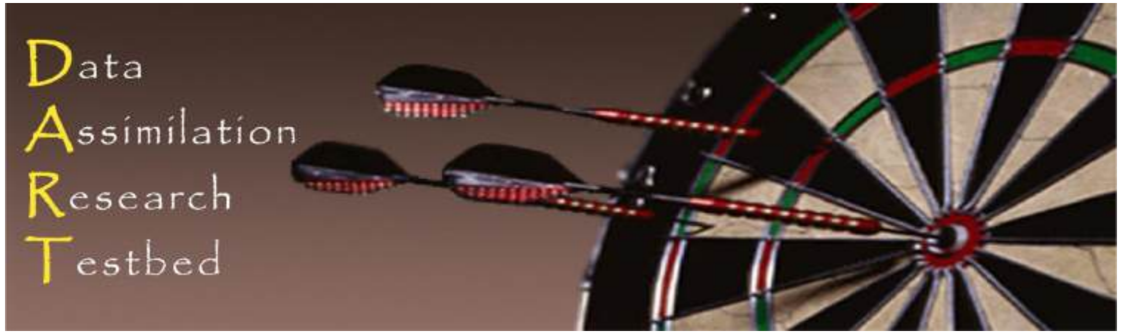
How well can we predict these flooding events?

2. Hydro-DART: The Hydrologic Prediction System

Hydro-DART: A hydrologic ensemble prediction system that integrates NSF NCAR's Weather Research and Forecasting Hydrological Model (WRF-Hydro) with NSF NCAR's Data Assimilation Research Testbed (DART)

- ❑ Hourly and sub-hourly streamflow assimilation
- ❑ A highly robust prediction framework
- ❑ State-of-the-art data assimilation (DA) tools
- ❑ .. and many more including snow DA

WRF-Hydro®



NSF NCAR

2. Hydro-DART: The Hydrologic Prediction System

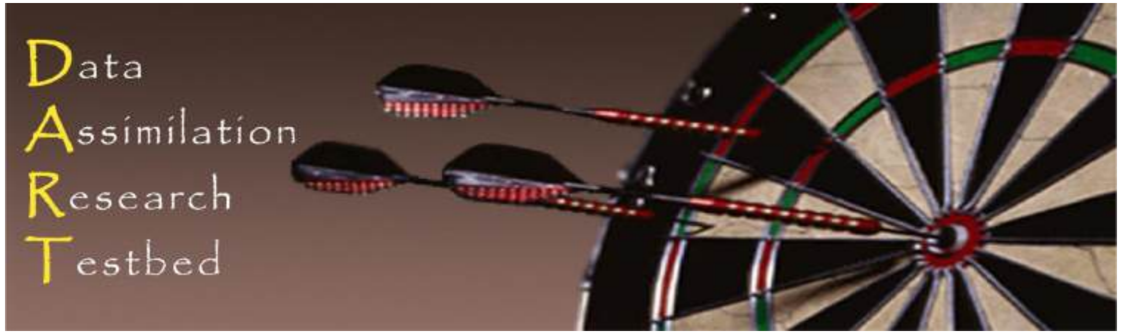
Hydro-DART: A hydrologic ensemble prediction system that integrates NSF NCAR's Weather Research and Forecasting Hydrological Model (WRF-Hydro) with NSF NCAR's Data Assimilation Research Testbed (DART)

- ❑ Hourly and sub-hourly streamflow assimilation
- ❑ A highly robust prediction framework
- ❑ State-of-the-art data assimilation (DA) tools
- ❑ .. and many more including snow DA

Python Wrapper **Configuration** **Diagnostics**

YAML

WRF-Hydro®



2. Hydro-DART: The Hydrologic Prediction System

Hydro-DART: A hydrologic ensemble prediction system that integrates NSF NCAR's Weather Research and Forecasting Hydrological Model (WRF-Hydro) with NSF NCAR's Data Assimilation Research Testbed (DART)

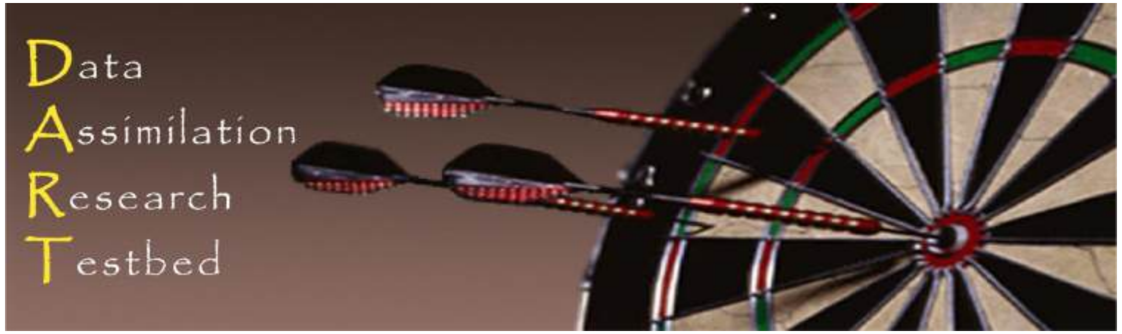
- ❑ Hourly and sub-hourly streamflow assimilation
- ❑ A highly robust prediction framework
- ❑ State-of-the-art data assimilation (DA) tools
- ❑ .. and many more including snow DA

Github
NCAR/DART/tree/main/models/wrf_hydro

Python Wrapper | **Configuration** | **Diagnostics**

YAML

WRF-Hydro®



2.1 The Hydrologic Model: WRF-Hydro

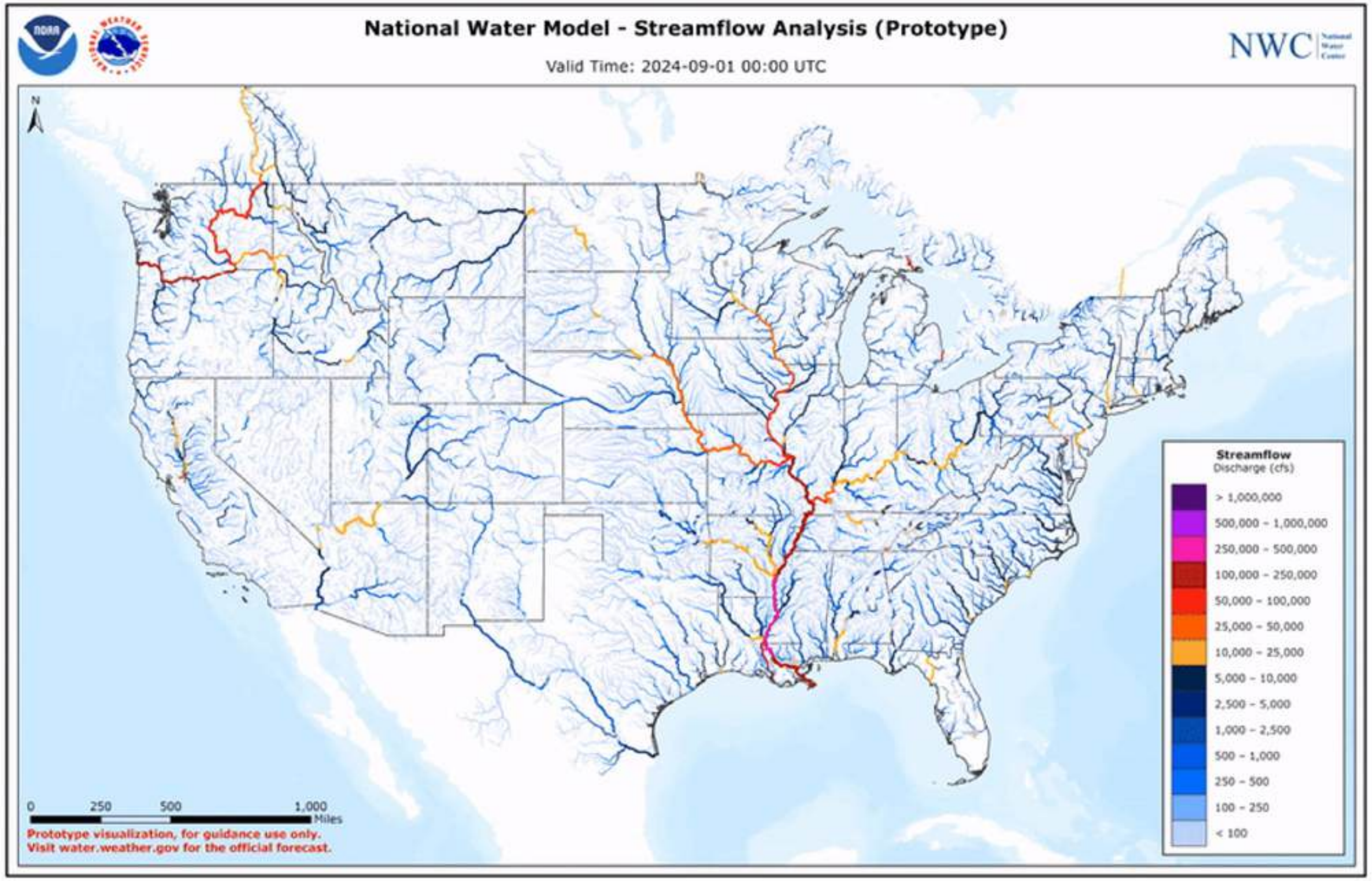
WRF-Hydro: NSF NCAR Weather Research and Forecasting model (WRF) hydrological modeling system. Research component of the National Water Model (NWM)

Community-based system [Gochis et al., 2020]:

- ❑ Major water cycle components
- ❑ Reliable streamflow predictions
- ❑ Land-atmosphere coupling

Open Source

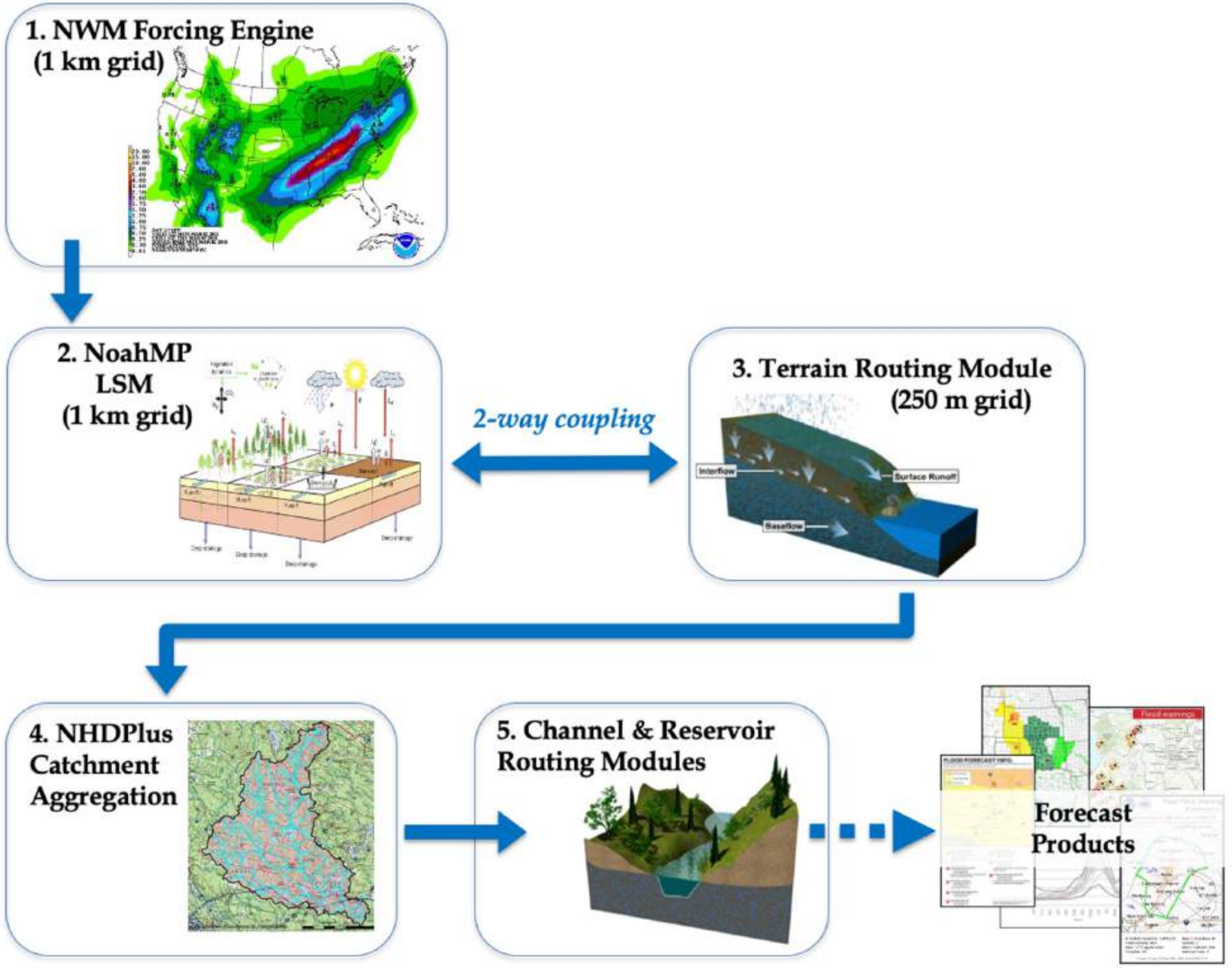
Github: `NCAR/wrf_hydro_nwm_public`
Web: `ral.ucar.edu/projects/wrf_hydro`
Docs: `wrf-hydro.readthedocs.io`



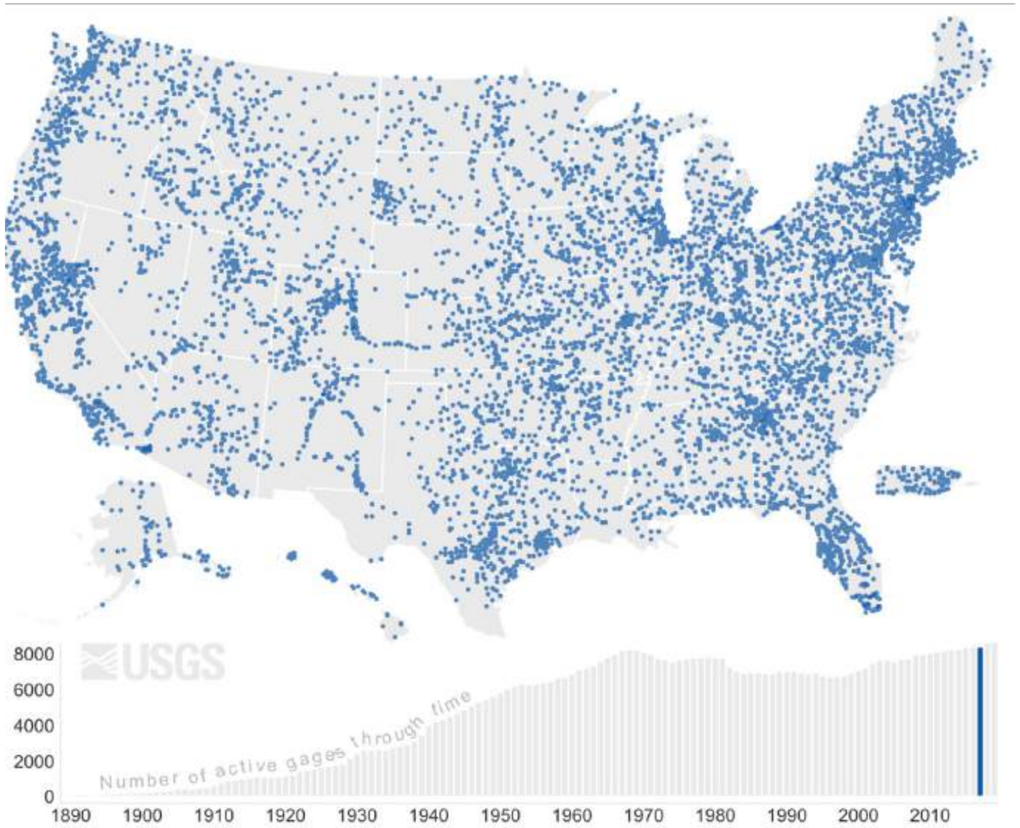
Streamflow in cubic feet per second (cfs) as simulated by NOAA's NWM (v3.0) for the month of September, 2024

2.1 The Hydrologic Model: WRF-Hydro

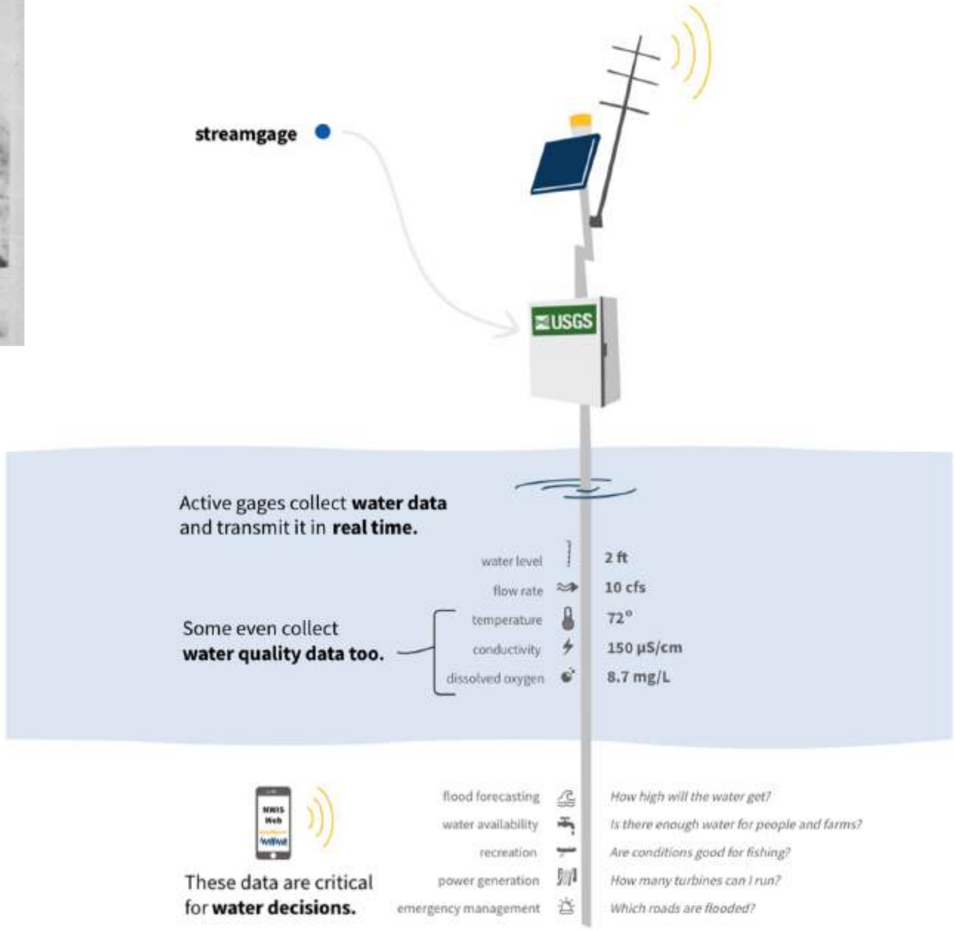
1. Atmospheric driver
2. Noah-MP LSM
   a. 1 km spatial resolution
   b. 1 hr temporal resolution
3. Terrain routing
   a. Finer 250 m resolution
   b. Resolve local topographic features (e.g., depressions)
4. Geospatial framework: USGS National Hydrography Dataset (NHD)
   a. Medium-resolution
   b. Streams and catchments
5. **Channel + subsurface routing**



2.2 Hydro-DART: Streamflow Data



1889: Rio Grande at Embudo, New Mexico



- ❑ USGS operates a network of more than 9000 stream gauges nationwide
- ❑ Hourly (+sub-hourly) assimilation of streamflow data

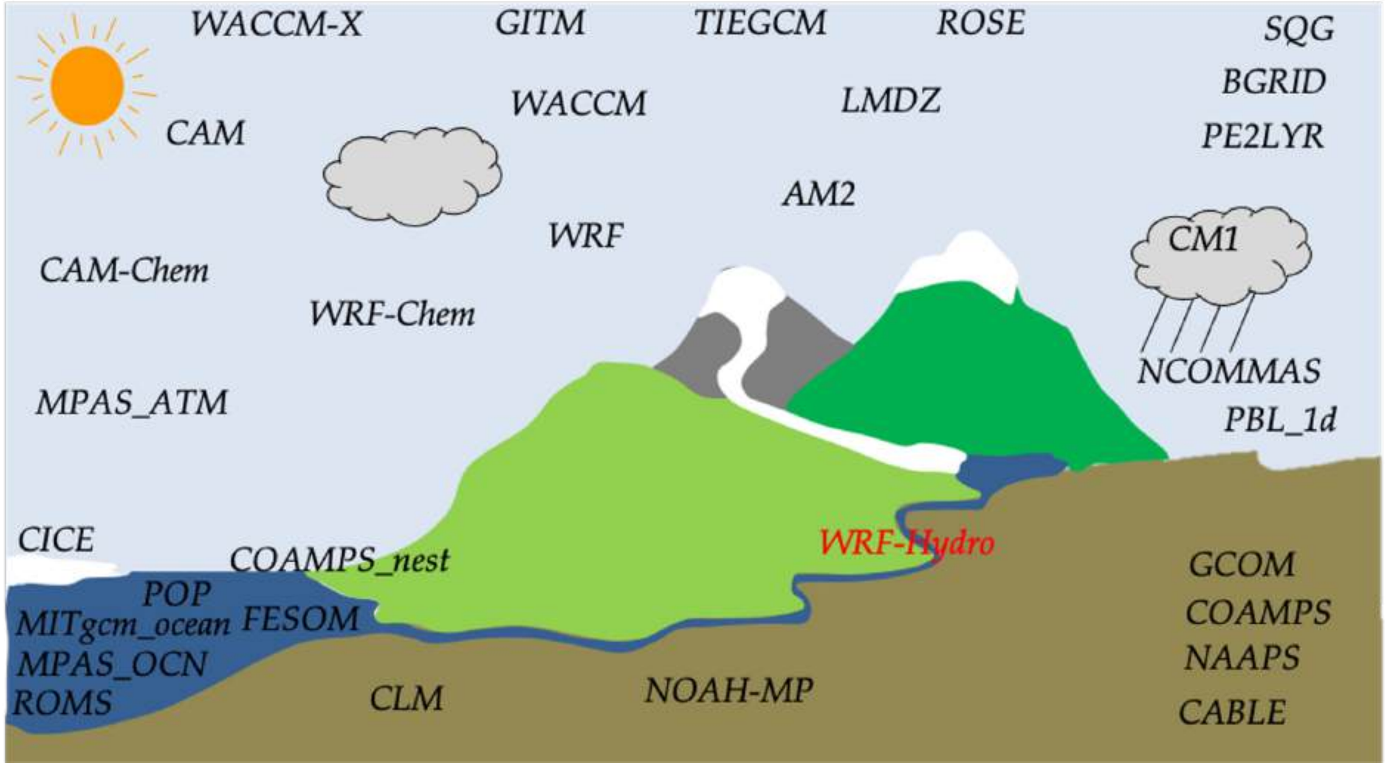
2.3 The Data Assimilation Research Testbed

DART: A community facility for **ensemble** Data Assimilation, developed and maintained by DAReS in CISL

Github: NCAR/DART
Web: dart.ucar.edu
Docs: docs.dart.ucar.edu

What can you do with DART:

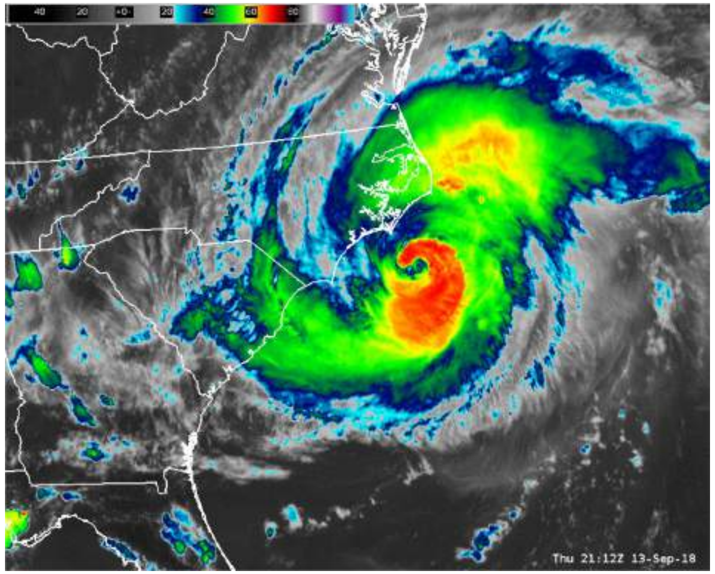
- ❑ Ensemble forecasting/reanalysis
- ❑ Model improvement and predictability
- ❑ Sensitivity analysis
- ❑ OSE, OSSE + DA algorithms
- ❑ Observation design/evaluation



Anderson, J., T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, A. Arellano, 2009: The Data Assimilation Research Testbed: A Community Facility. Bull. Amer. Meteor. Soc. ***90****, 1283-1296*

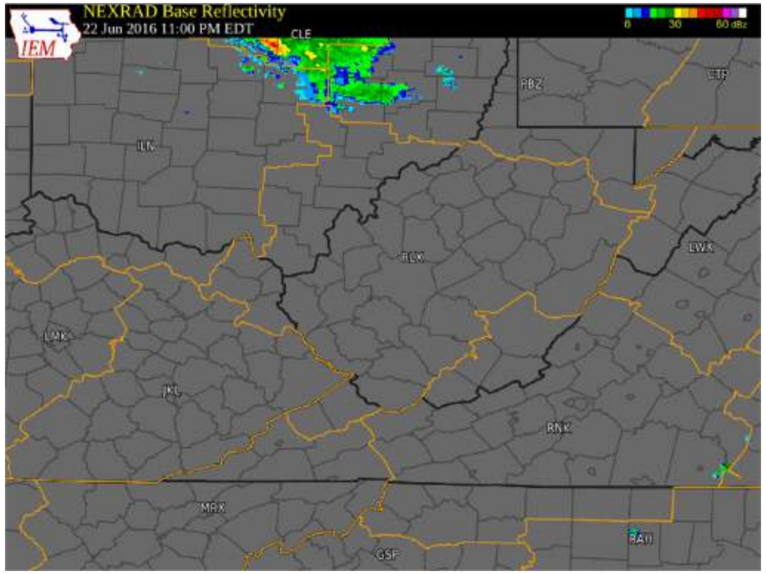
2.4 Flooding Scenarios

Hurricane Florence, NC (2018)



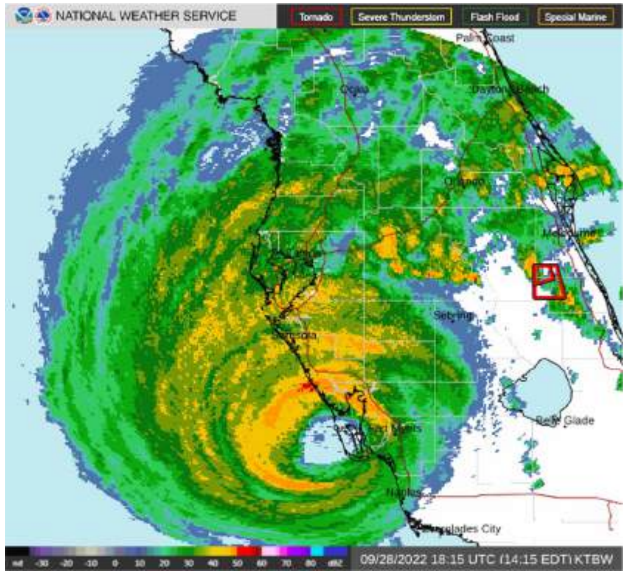
- Category 4 hurricane
- Landfall: Sep. 14 (Carolinas)
- Winds up to 150 mph
- Damages: $25 billion
- ~50 people died
- Flooding magnitude exceeded Matthew ('16) and Floyd ('99) **combined**

Flash Flood, WV (2016)



- Several thunderstorms
- Flash flooding, June 2016
- Damages: $1.2 billion
- ~23 people died
- 8-10" of rainfall over a 12-hr period
- Occurrence probability: **1/1000**

Hurricane Ian, FL (2022)



- Category 4 hurricane
- Landfall: Sep. 28
- US record: 5th strongest
- Damages: **> $112 billion**
- ~150 people died
- Precipitation exceeded **20 inches**

3. Novel Data Assimilation Algorithms and Tools

Bayes Theorem:

$$p(x|y) \propto p(x) \cdot p(y|x)$$

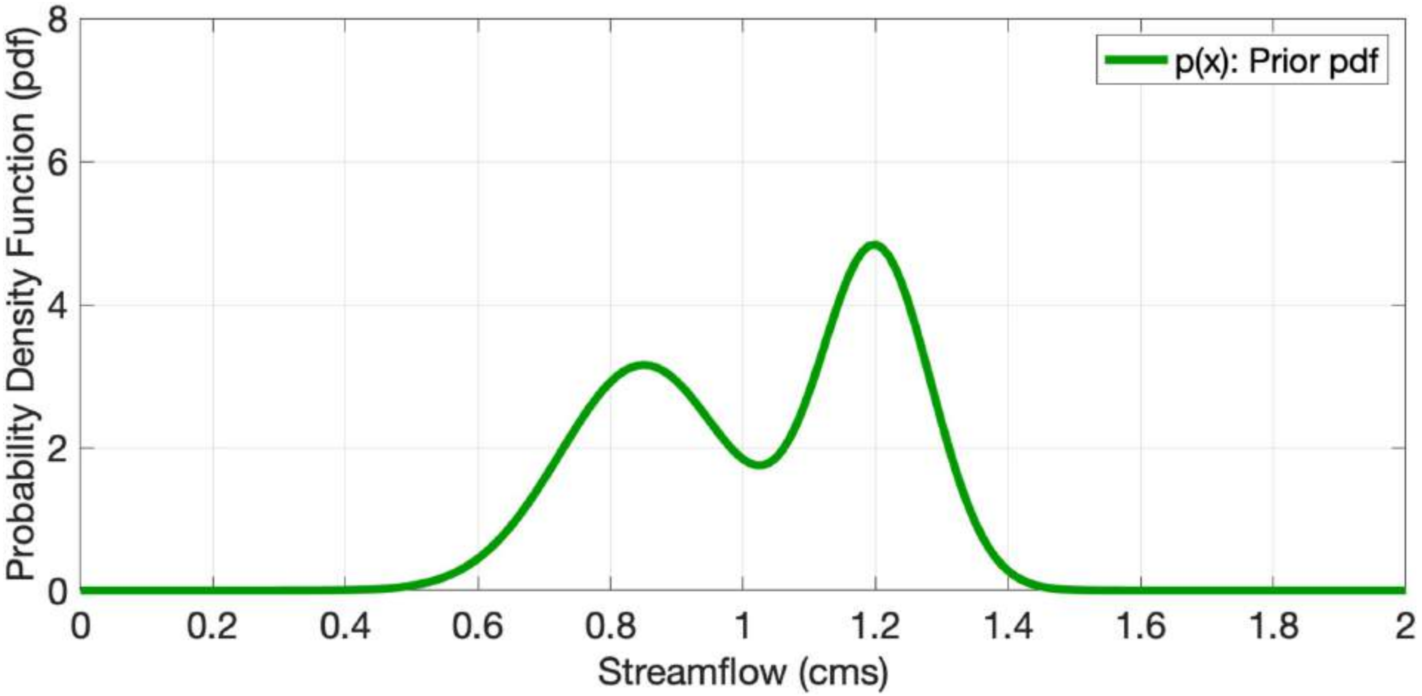
- **x**: model state variables such as streamflow
- **y**: observations i.e., gauge flow data
- **p(x)**: prior (forecast) distribution
- **p(y|x)**: observation likelihood function
- **p(x|y)**: posterior (analysis) distribution

3. Novel Data Assimilation Algorithms and Tools

Bayes Theorem:

$$p(x|y) \propto p(x) \cdot p(y|x)$$

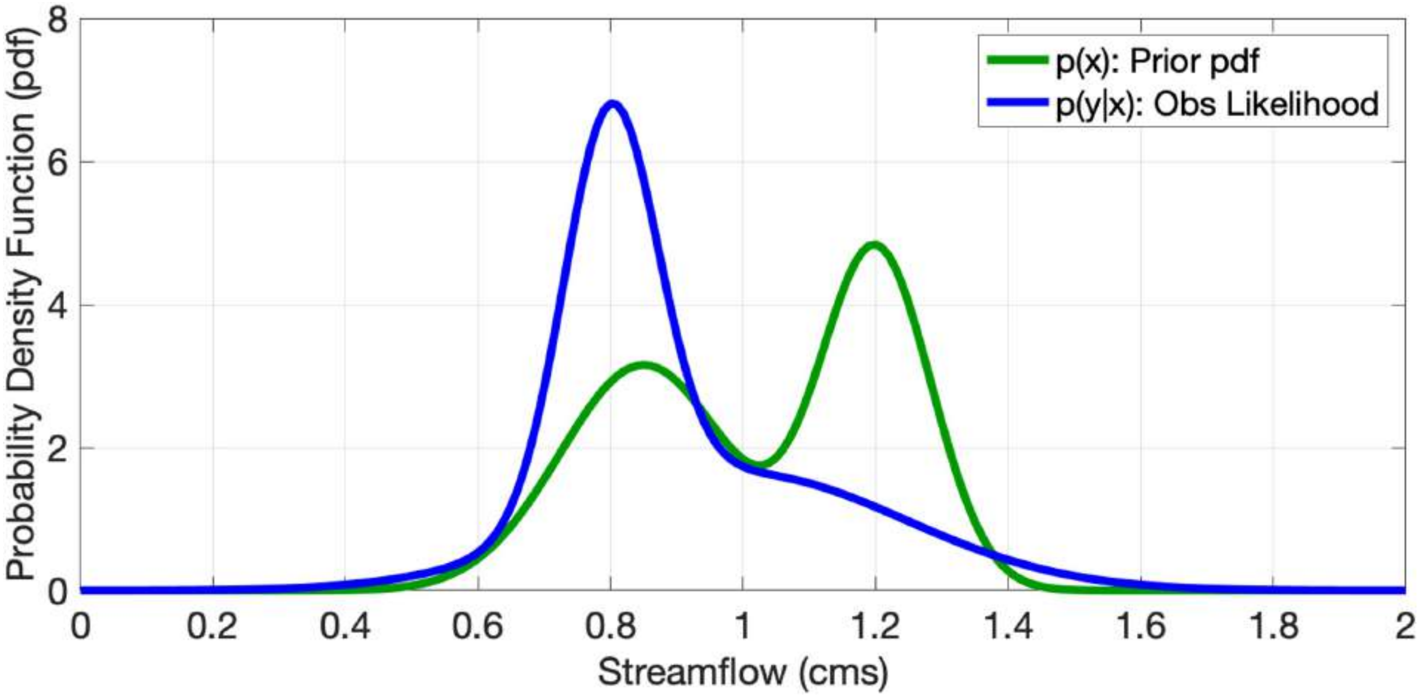
- **x**: model state variables such as streamflow
- **y**: observations i.e., gauge flow data
- **p(x)**: prior (forecast) distribution
- **p(y|x)**: observation likelihood function
- **p(x|y)**: posterior (analysis) distribution



Bayes Theorem:

$$p(x|y) \propto p(x) \cdot p(y|x)$$

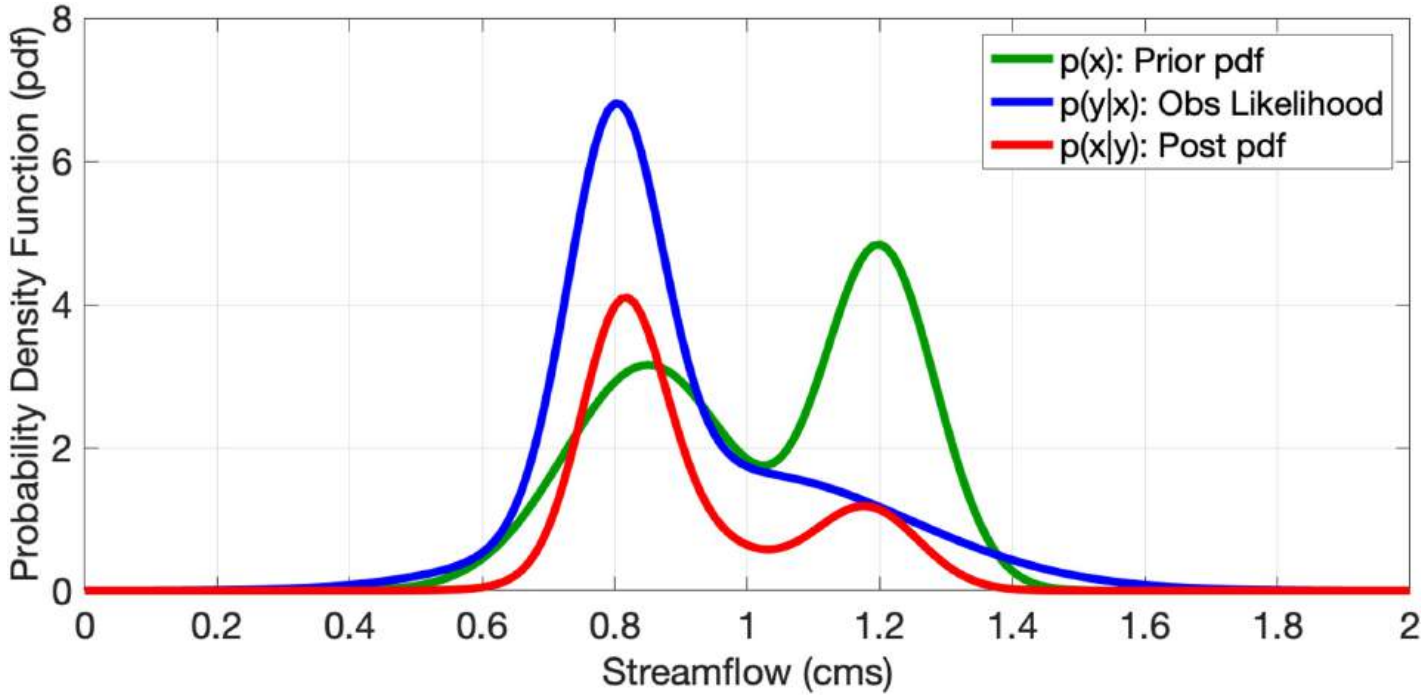
- **x**: model state variables such as streamflow
- **y**: observations i.e., gauge flow data
- p(x): prior (forecast) distribution
- p(y | x): observation likelihood function
- p(x | y): posterior (analysis) distribution



Bayes Theorem:

$$p(x|y) \propto p(x) \cdot p(y|x)$$

- **x**: model state variables such as streamflow
- **y**: observations i.e., gauge flow data
- p(x): prior (forecast) distribution
- p(y | x): observation likelihood function
- p(x | y): posterior (analysis) distribution



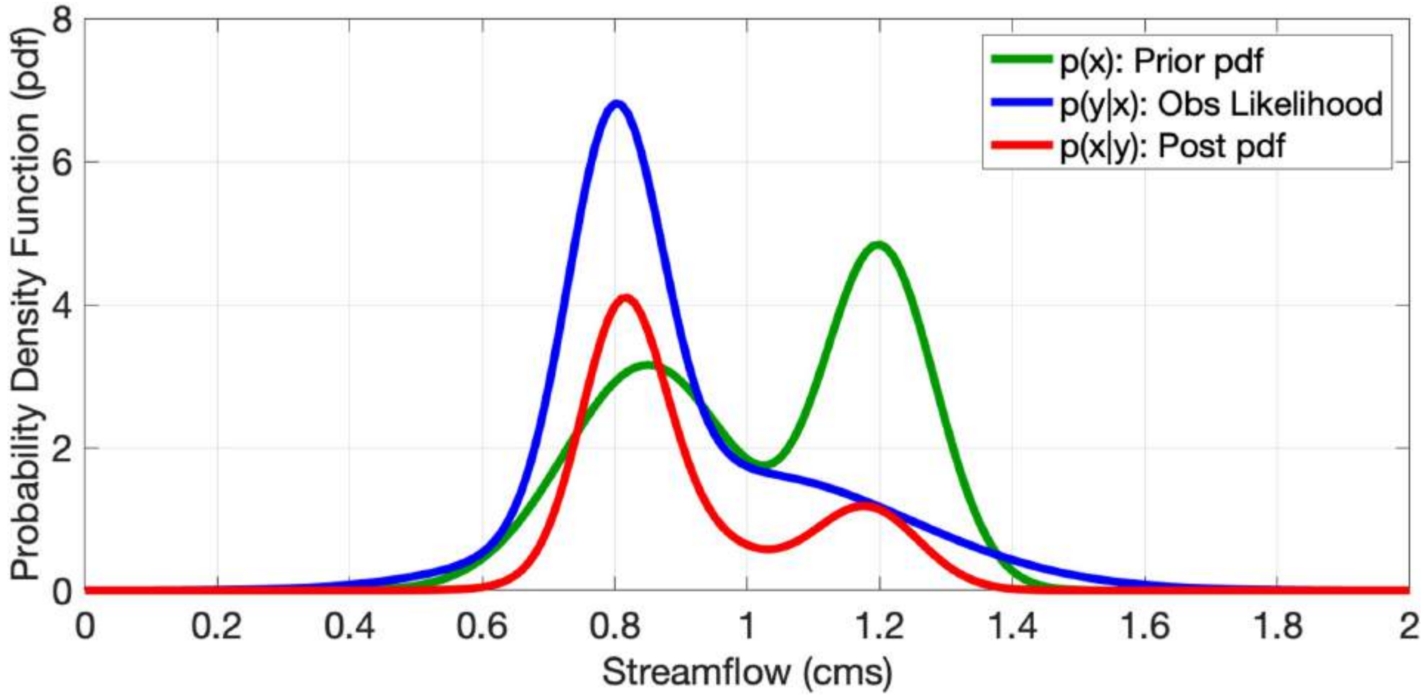
3. Novel Data Assimilation Algorithms and Tools

Bayes Theorem:

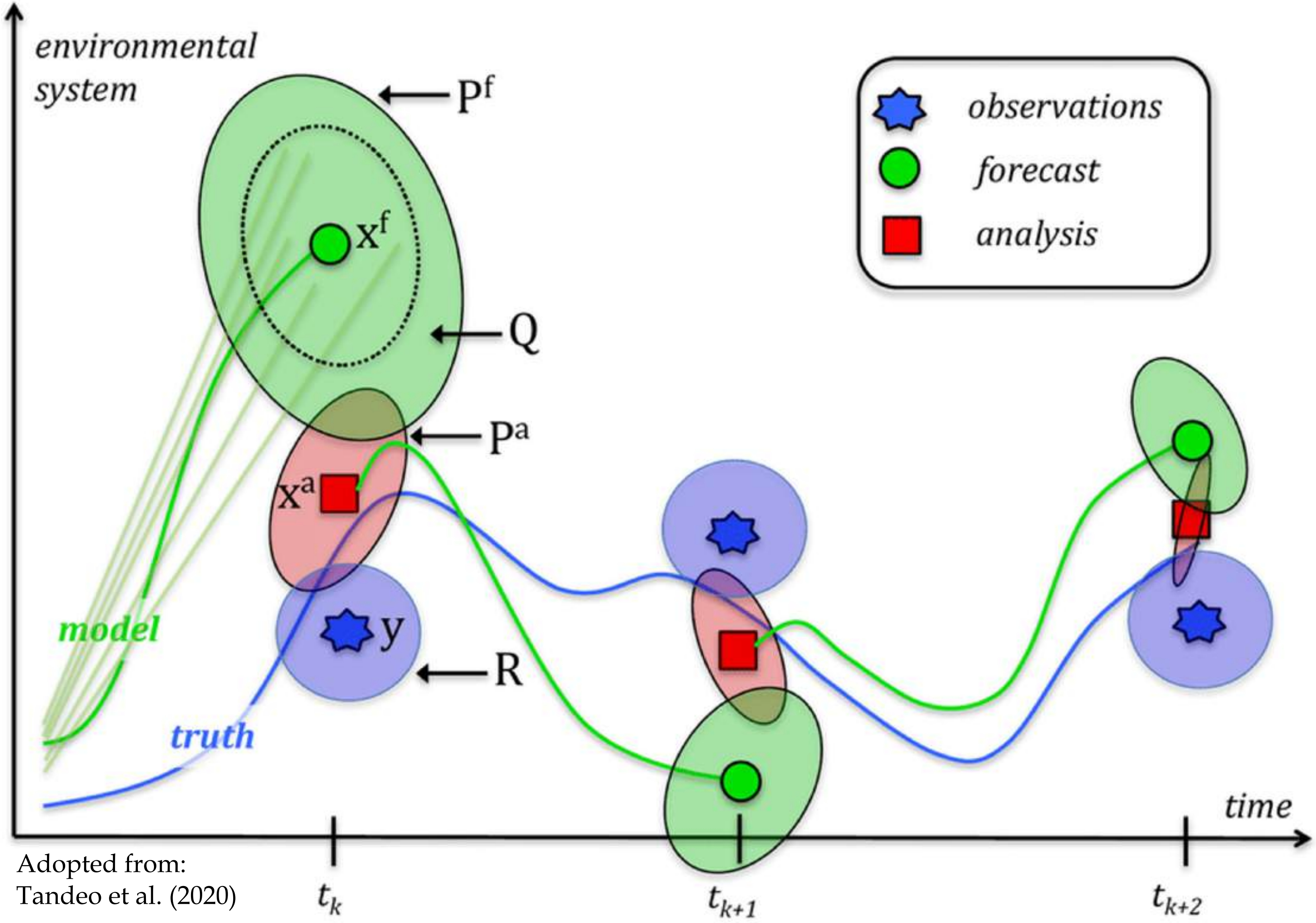
$$p(x|y) \propto p(x) \cdot p(y|x)$$

- **x**: model state variables such as streamflow
- **y**: observations i.e., gauge flow data
- p(x): prior (forecast) distribution
- p(y | x): observation likelihood function
- p(x | y): posterior (analysis) distribution

Solve the Bayesian problem sequentially using the **ensemble Kalman filter**!



3.1 DA Algorithms: The Ensemble Kalman Filter (EnKF)



Adopted from:
Tandeo et al. (2020)

- Use an ensemble (a set of model state realizations) to estimate the pdf
- Gaussian approximation
- Recursive Algorithm
- DA cycle:
  - Propagation (or forecast)
  - Update (or analysis)

3.1 DA Algorithms: The EnKF

$$p(x) \sim \mathcal{N}\left[\overline{\mathbf{x}}^f, \mathbf{P}^f\right]$$

$$\overline{\mathbf{x}}^f = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i^f$$

Prior Ensemble Mean

$$\mathbf{P}^f = \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathbf{x}_i^f - \overline{\mathbf{x}}^f\right)\left(\mathbf{x}_i^f - \overline{\mathbf{x}}^f\right)^T$$

Prior background covariance

$$p(x|y) \sim \mathcal{N}\left[\overline{\mathbf{x}}^a, \mathbf{P}^a\right]$$

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(\mathbf{y}^o - \mathbf{H}\overline{\mathbf{x}}^f\right)$$

Posterior Ensemble Mean

3.1 DA Algorithms: The EnKF

$$p(x) \sim \mathcal{N}\left[\overline{\mathbf{x}}^f, \mathbf{P}^f\right]$$

$$\overline{\mathbf{x}}^f = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i^f$$

Prior Ensemble Mean

$$\mathbf{P}^f = \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathbf{x}_i^f - \overline{\mathbf{x}}^f\right)\left(\mathbf{x}_i^f - \overline{\mathbf{x}}^f\right)^T$$

Prior background covariance

$$p(x|y) \sim \mathcal{N}\left[\overline{\mathbf{x}}^a, \mathbf{P}^a\right]$$

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(\mathbf{y}^o - \mathbf{H}\overline{\mathbf{x}}^f\right)$$

Posterior Ensemble Mean

3.1 DA Algorithms: The EnKF

$$p(x) \sim \mathcal{N}\left[\overline{\mathbf{x}}^f, \mathbf{P}^f\right]$$

$$\overline{\mathbf{x}}^f = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i^f$$

Prior Ensemble Mean

$$\mathbf{P}^f = \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathbf{x}_i^f - \overline{\mathbf{x}}^f\right)\left(\mathbf{x}_i^f - \overline{\mathbf{x}}^f\right)^T$$

Prior background covariance

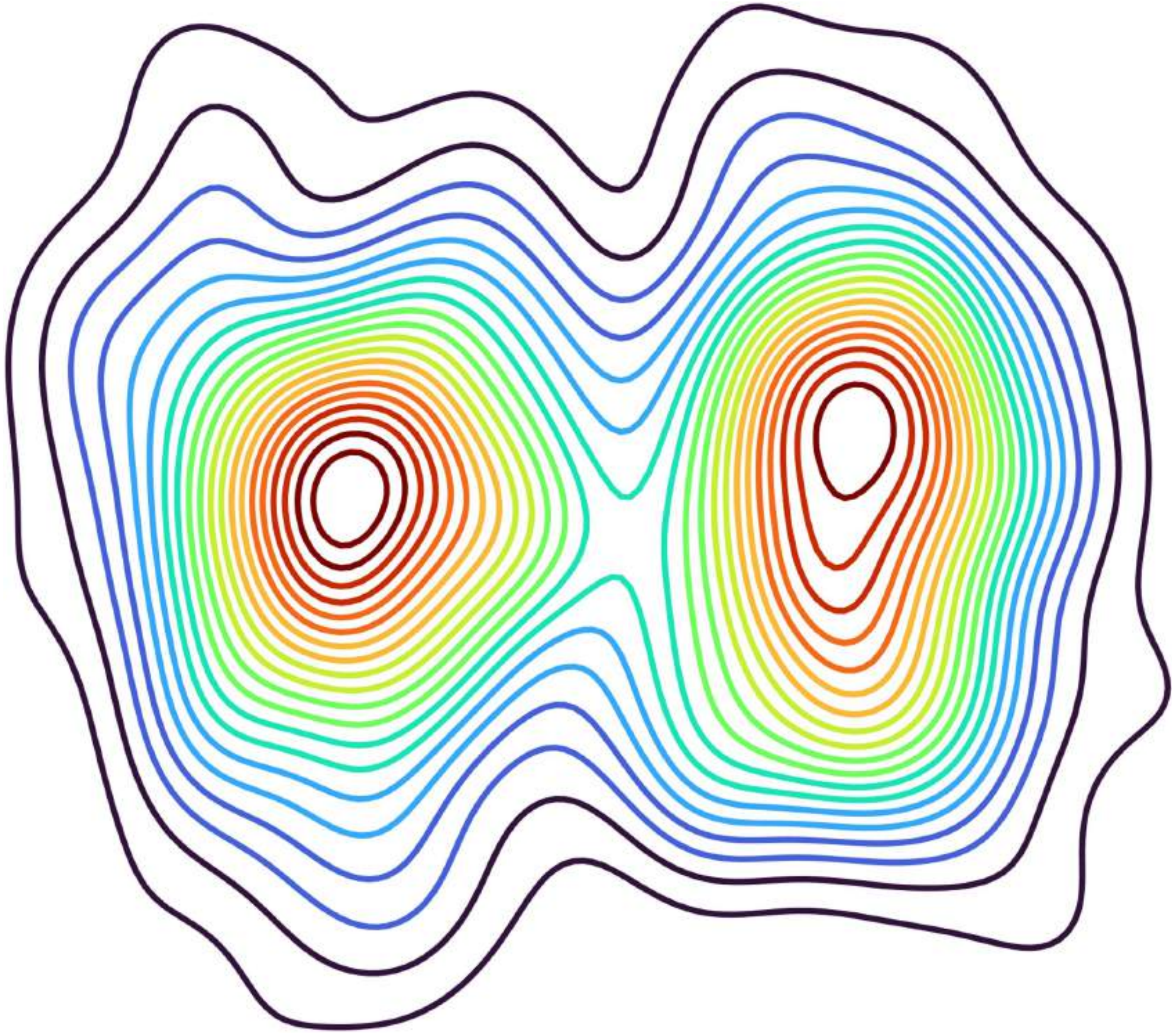
Ensemble Size

$$p(x|y) \sim \mathcal{N}\left[\overline{\mathbf{x}}^a, \mathbf{P}^a\right]$$

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(\mathbf{y}^o - \mathbf{H}\overline{\mathbf{x}}^f\right)$$

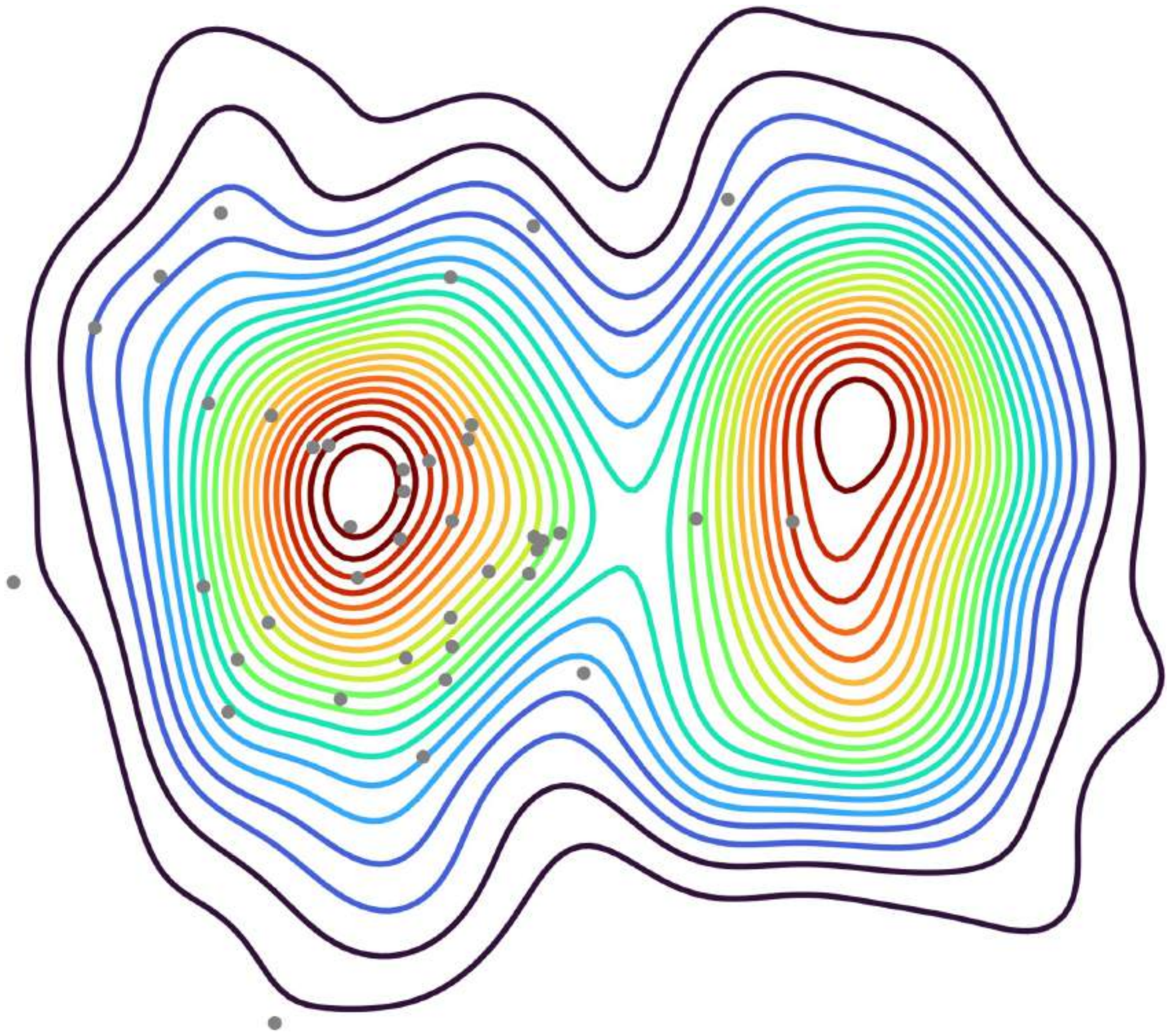
Posterior Ensemble Mean

3.1 DA Algorithms: EnKF Sampling



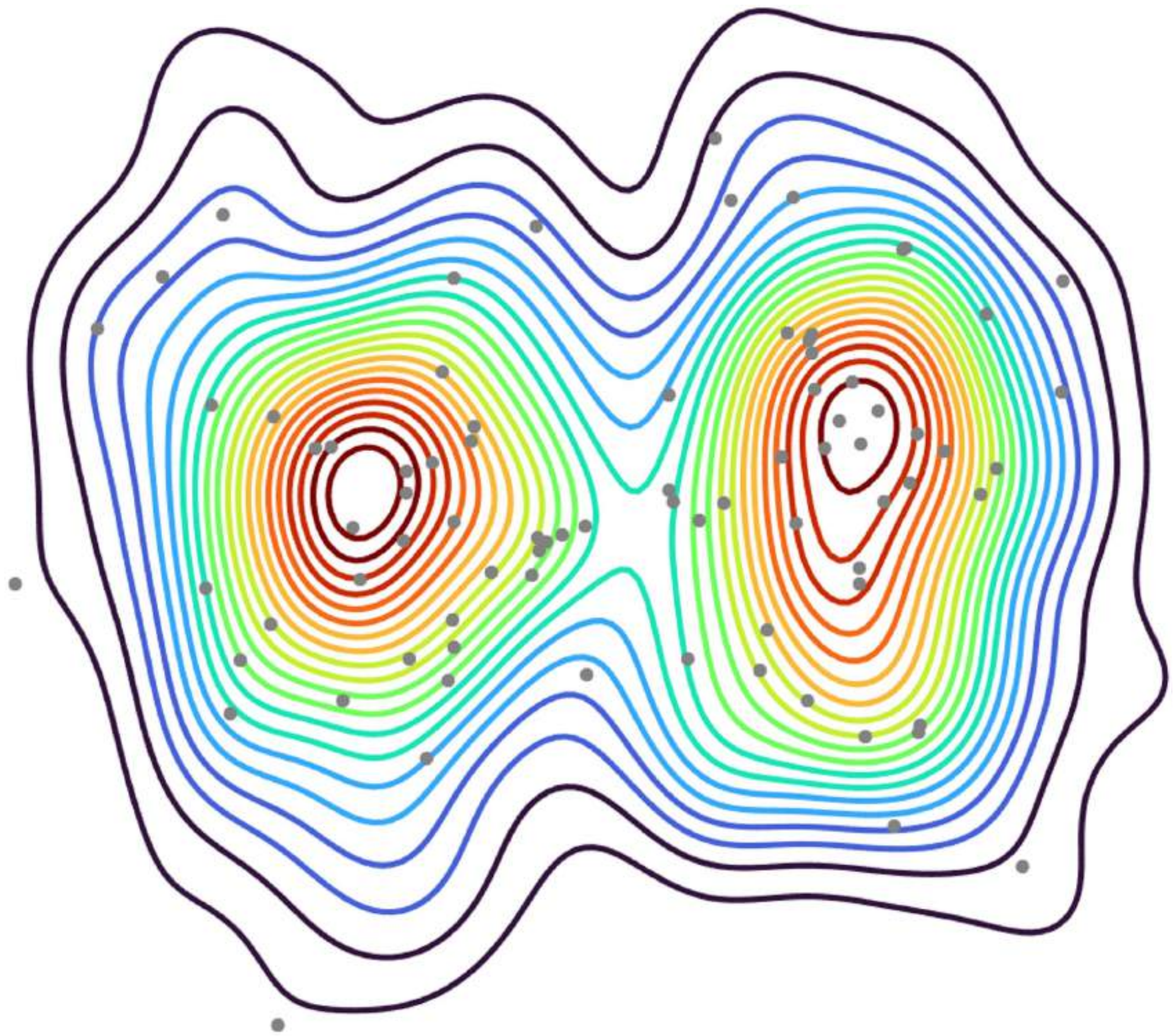
- Draw samples from a known bi-modal distribution
- Do these samples provide a good representation of the entire pdf?
  - 40 members
  - 80 members
  - 400 members
  - …

3.1 DA Algorithms: EnKF Sampling



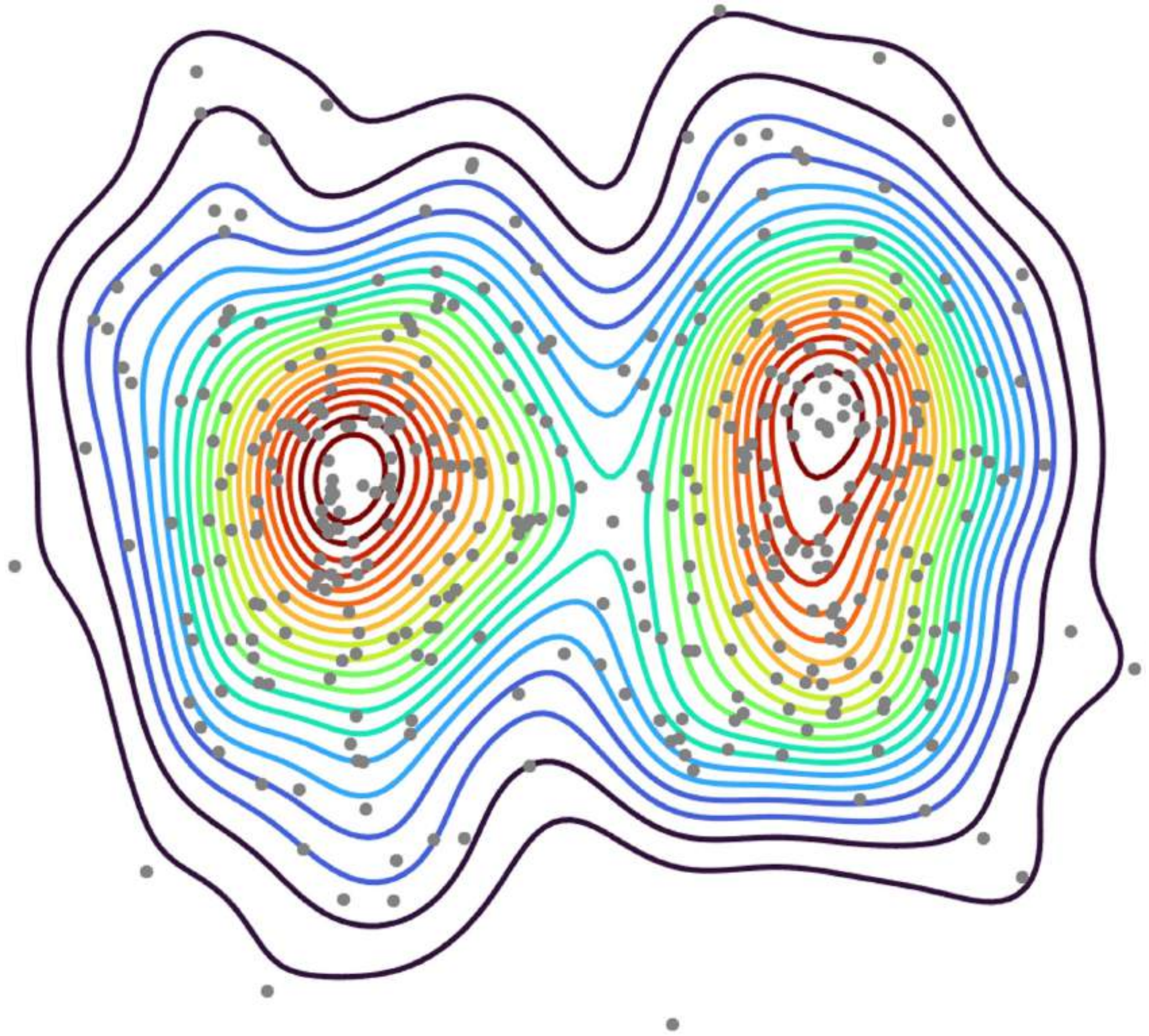
- Draw samples from a known bi-modal distribution
- Do these samples provide a good representation of the entire pdf?
  - 40 members
  - 80 members
  - 400 members
  - …

3.1 DA Algorithms: EnKF Sampling



- Draw samples from a known bi-modal distribution
- Do these samples provide a good representation of the entire pdf?
  - 40 members
  - 80 members
  - 400 members
  - …

3.1 DA Algorithms: EnKF Sampling

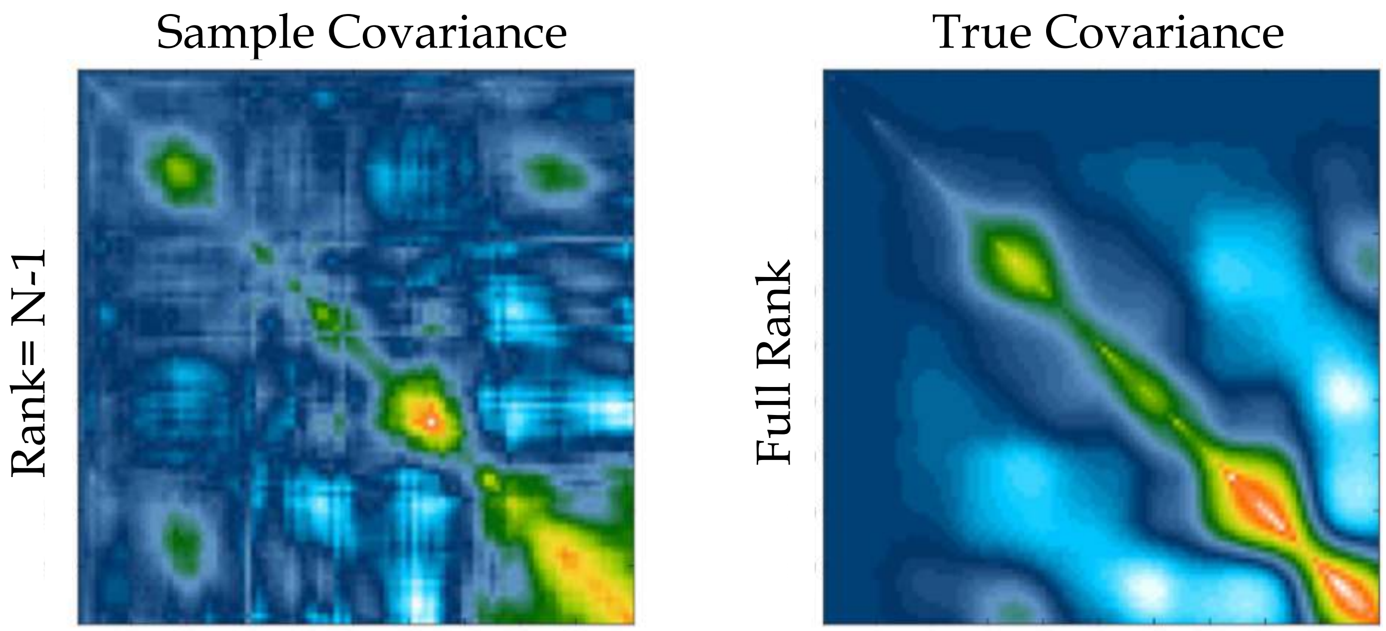


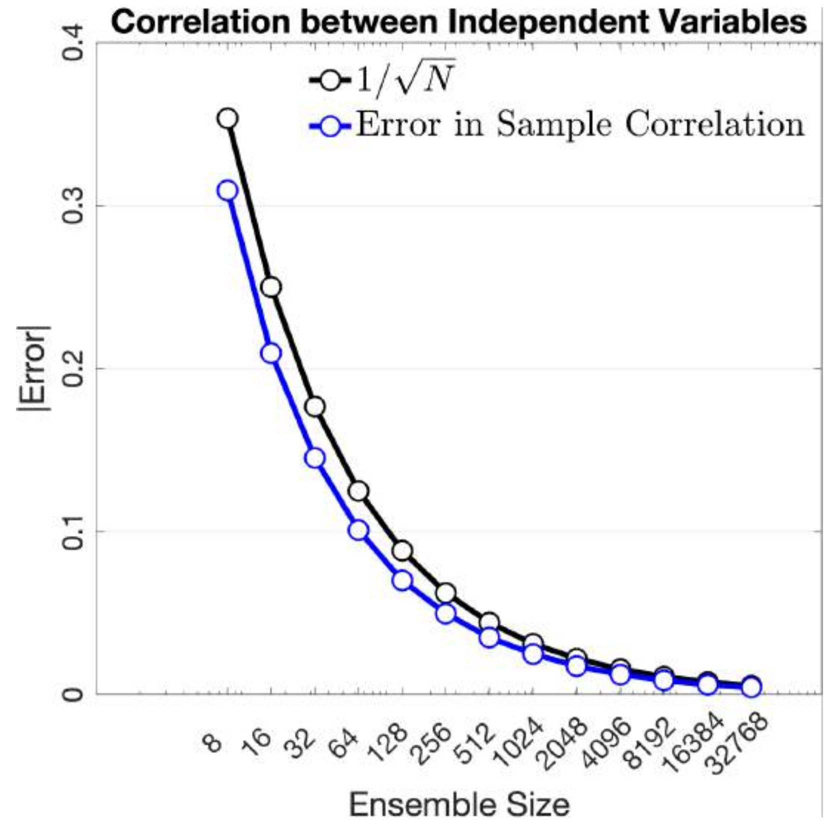
- Draw samples from a known bi-modal distribution
- Do these samples provide a good representation of the entire pdf?
  - 40 members
  - 80 members
  - 400 members
  - …

3.1 DA Algorithms: EnKF Limitations

➔ **Geophysical Models: Can't afford running large ensemble sizes!**

- *Sampling errors:* deteriorate $\mathbf{P}^f$
  - True variance is underestimated
  - Rank deficiency; N << model size
  - Noisy and spurious correlations





3.1 DA Algorithms: EnKF Limitations

- **Can't afford running large ensemble sizes!**
  - *Sampling errors:* deteriorate $\mathbf{P}^f$
    - True variance is underestimated
    - Rank deficiency; N << model size
    - Noisy and spurious correlations

- **Model errors**
  - *Ensemble collapse*
    - Uncertain parameters
    - Forcing (precipitation) errors

- **Physical Variability**
  - *Reduced ensemble spread*
    - Low-flow conditions
    - Simplified dynamics

3.1 DA Algorithms: EnKF Limitations

- **Can't afford running large ensemble sizes!**
  - *Sampling errors:* deteriorate $\mathbf{P}^f$
    - True variance is underestimated
    - Rank deficiency; N << model size
    - Noisy and spurious correlations

$$\mathbf{P}^f = \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathbf{x}_i^f - \overline{\mathbf{x}}^f\right)\left(\mathbf{x}_i^f - \overline{\mathbf{x}}^f\right)^T$$

- **Model errors**
  - *Ensemble collapse*
    - Uncertain parameters
    - Forcing (precipitation) errors

- Issues degrade the estimate of the sample background covariance
- Noisy, inaccurate, and underestimated background covariance leads to suboptimal updates

$$\mathbf{P}^f \rightarrow \mathbf{P}^{\text{true}}$$

- **Physical Variability**
  - *Reduced ensemble spread*
    - Low-flow conditions
    - Simplified dynamics

3.2 Physical Uncertainty: Parameter Perturbations

➔ Increase the physical variability of the hydrologic model, especially during low-flow periods

3.2 Physical Uncertainty: Parameter Perturbations

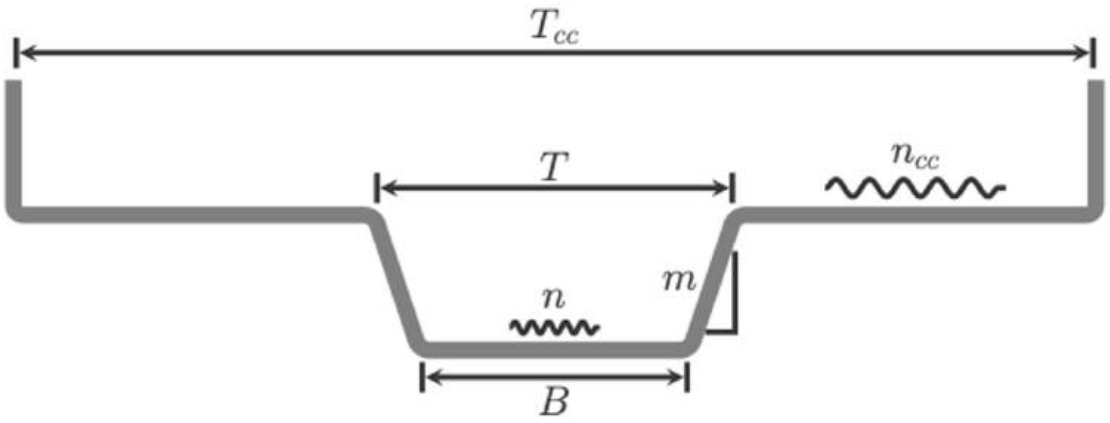
➔ Increase the physical variability of the hydrologic model, especially during low-flow periods

Multi-Configuration Ensemble:

Perturb uncertain channel parameters to create realistic variability

- Top width, T
- Bottom width, B
- Side slope, m
- Manning's N, n
- Width of compound channel, Tcc
- Manning's N of compound channel, ncc

Sampling uniformly under some physical constraints



3.2 Physical Uncertainty: Parameter Perturbations

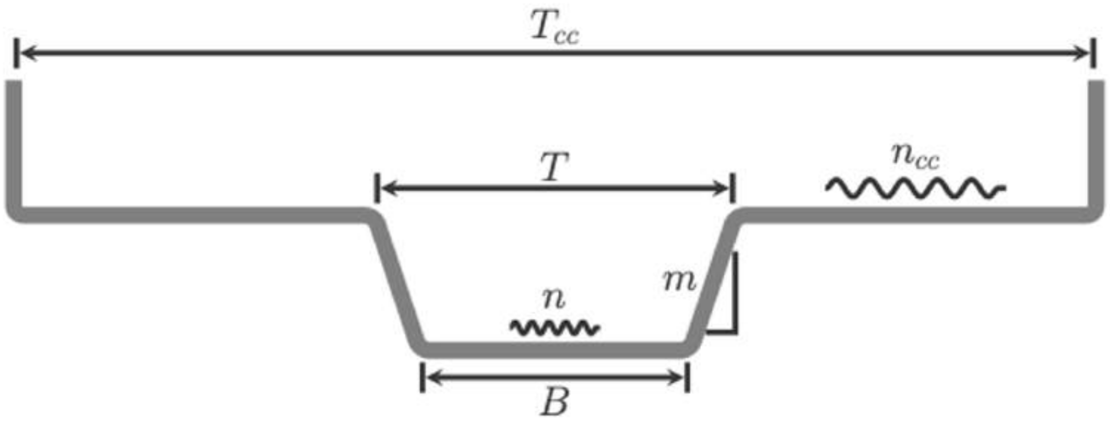
➔ Increase the physical variability of the hydrologic model, especially during low-flow periods

Multi-Configuration Ensemble:

Perturb uncertain channel parameters to create realistic variability

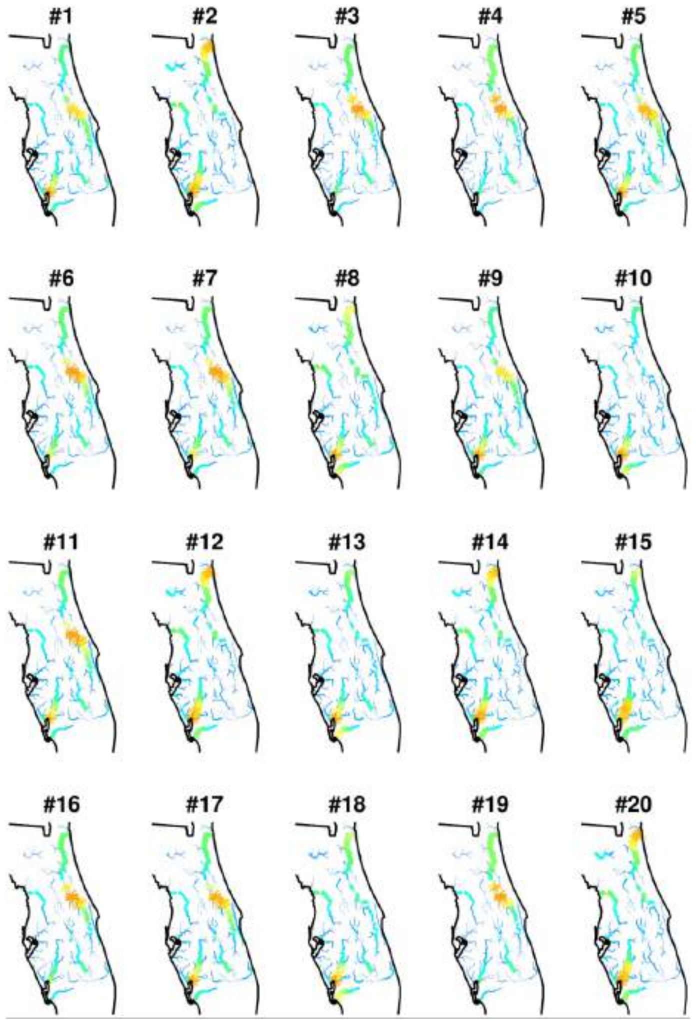
- Top width, T
- Bottom width, B
- Side slope, m
- Manning's N, n
- Width of compound channel, Tcc
- Manning's N of compound channel, ncc

Sampling uniformly under some physical constraints



Forcing Ensemble:

Generate ensemble perturbations to the boundary fluxes: surface and groundwater bucket

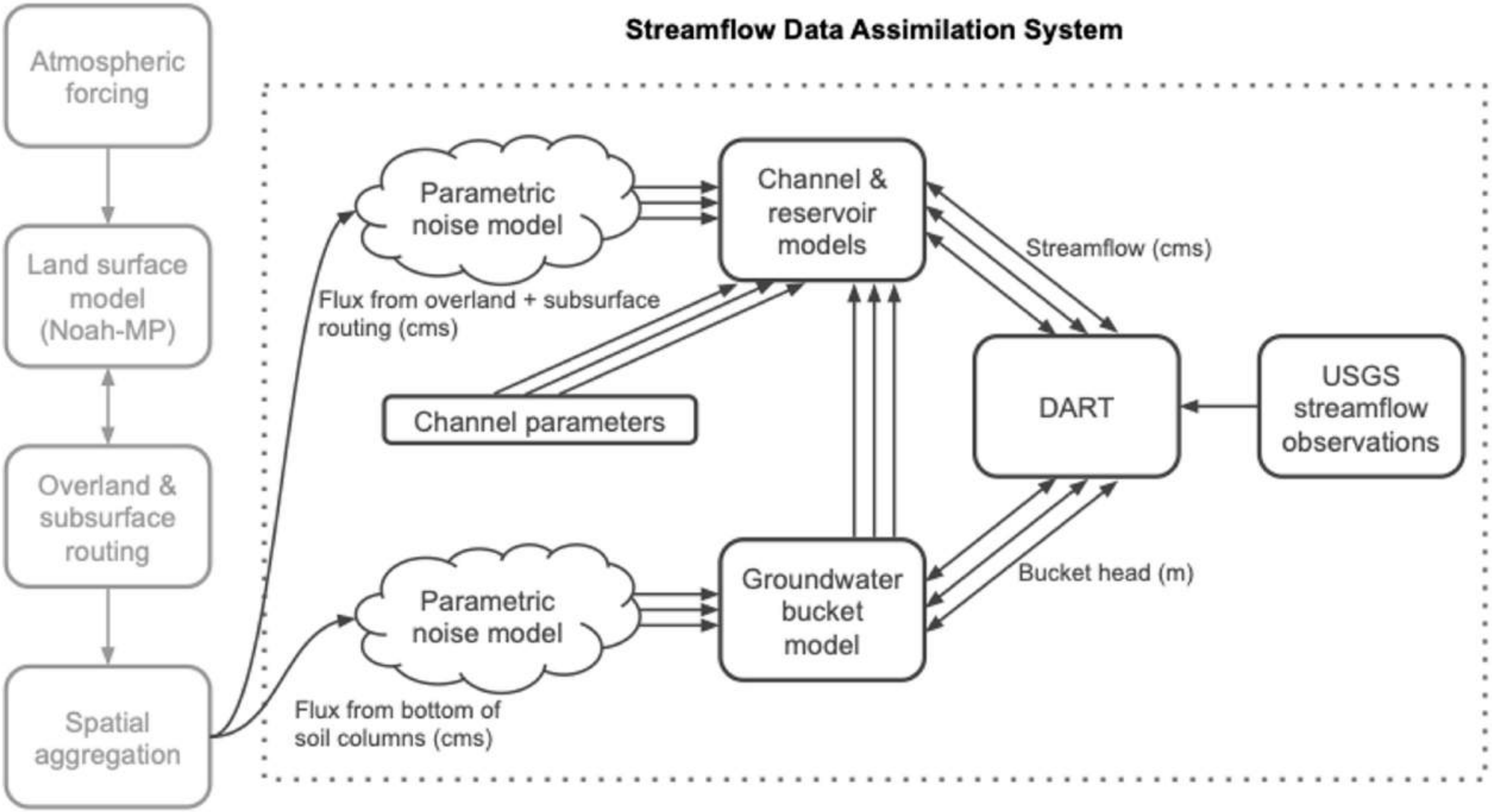


Streamflow multi-configuration ensemble realizations for the Ian flooding domain

3.2 Physical Uncertainty: Parameter Perturbations

Added variability due to forcing and parameters perturbations

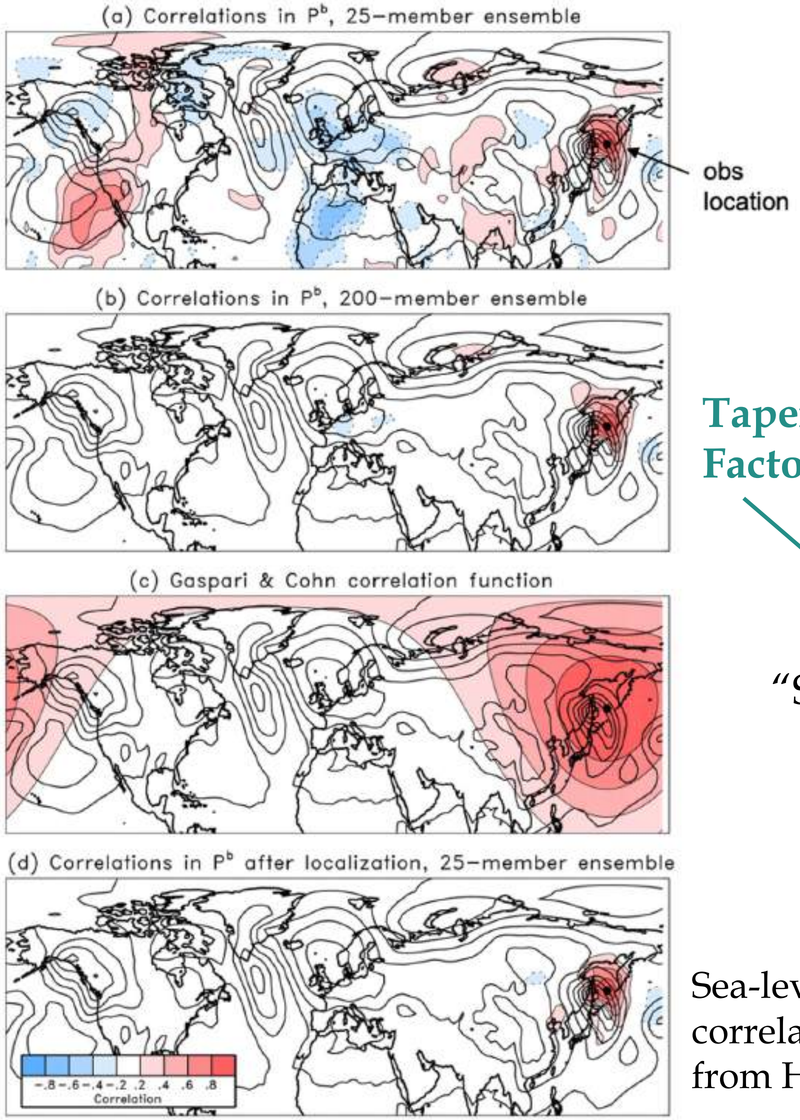
$\widehat{\mathbf{P}}^f$



Ensembles are denoted by groups of three arrows (N >> 3). The spatially distributed streamflow and bucket head states comprise the "state" vector passed to DART for updating by USGS streamflow observations

3.3 Along-The-Stream Localization

- **Localization: A way to tackle sampling errors**
  - deal with spurious, long-distance correlation
  - improves the rank of the covariance
  - Euclidean distance-based
  - widely used in NWP



Taper: Localization Factor [0,1]

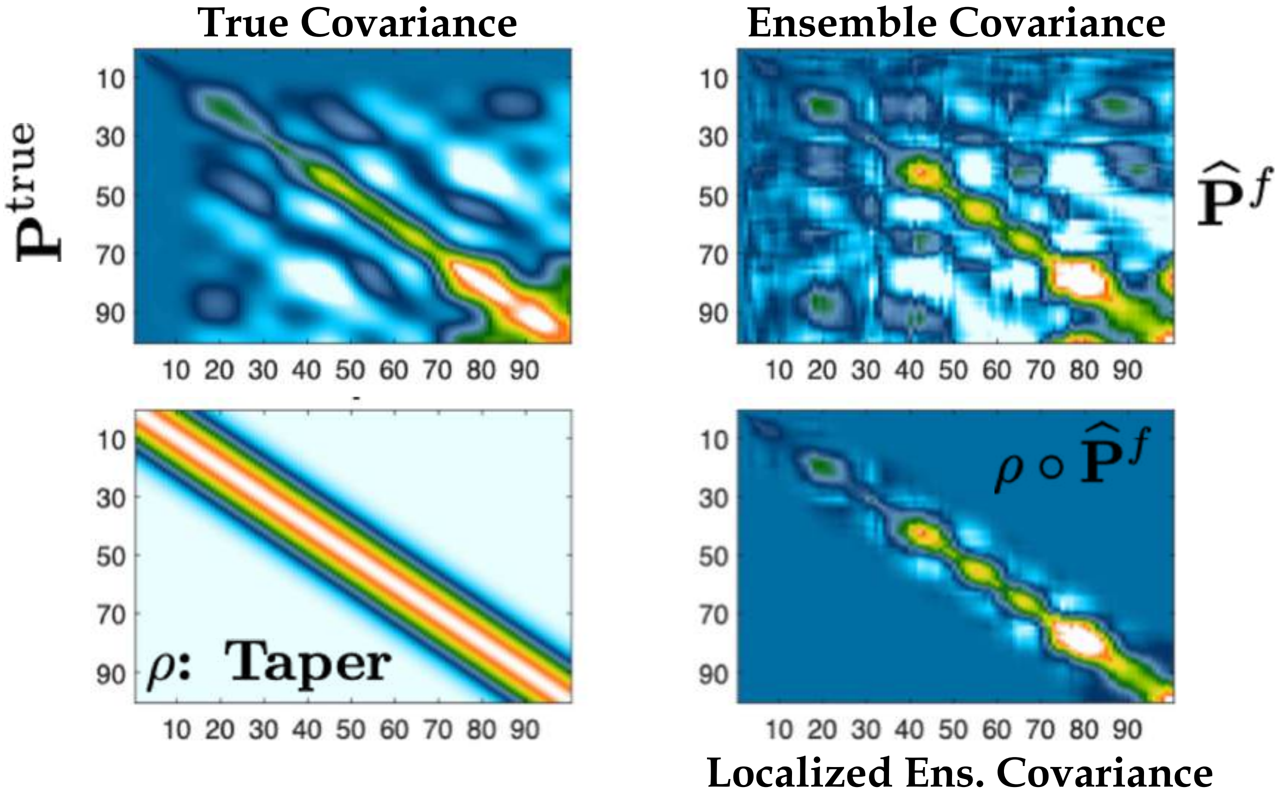
$$\rho \circ \widehat{\mathbf{P}}^f$$

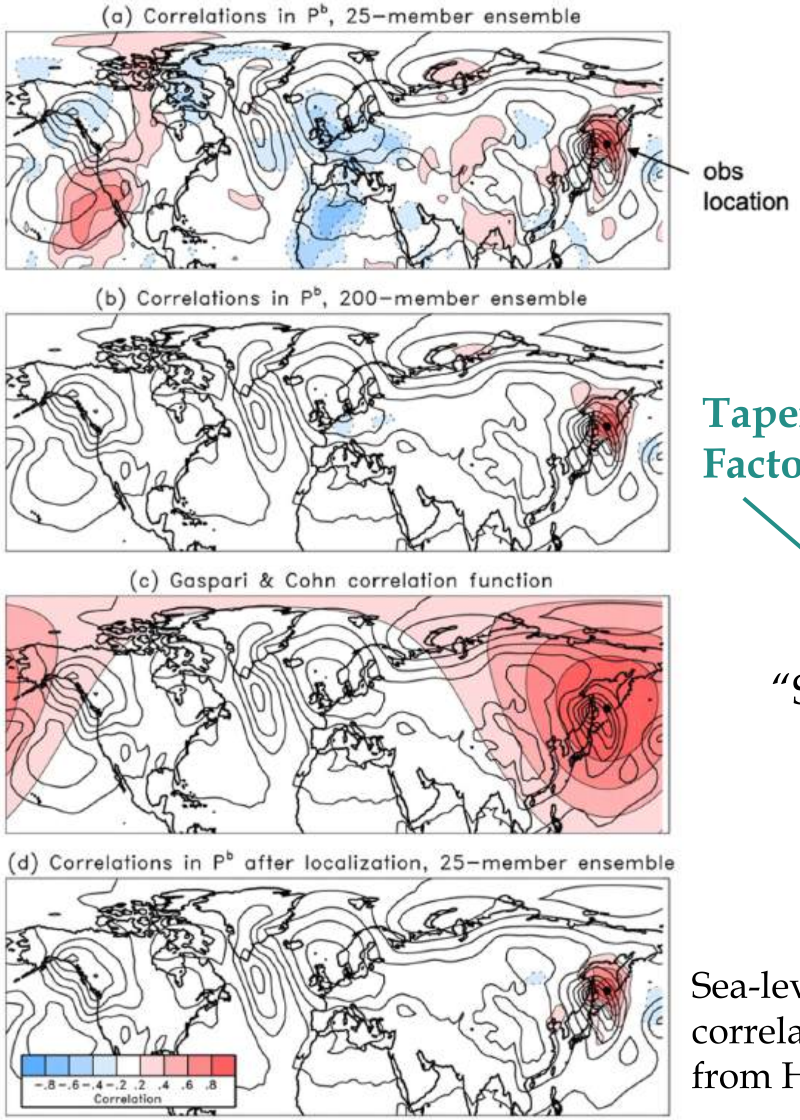
"Schur Product"

Sea-level pressure correlations. Adopted from Hamill (2003)

3.3 Along-The-Stream Localization

- Localization: A way to tackle sampling errors
  - deal with spurious, long-distance correlation
  - improves the rank of the covariance
  - Euclidean distance-based
  - widely used in NWP





Taper: Localization Factor [0,1]

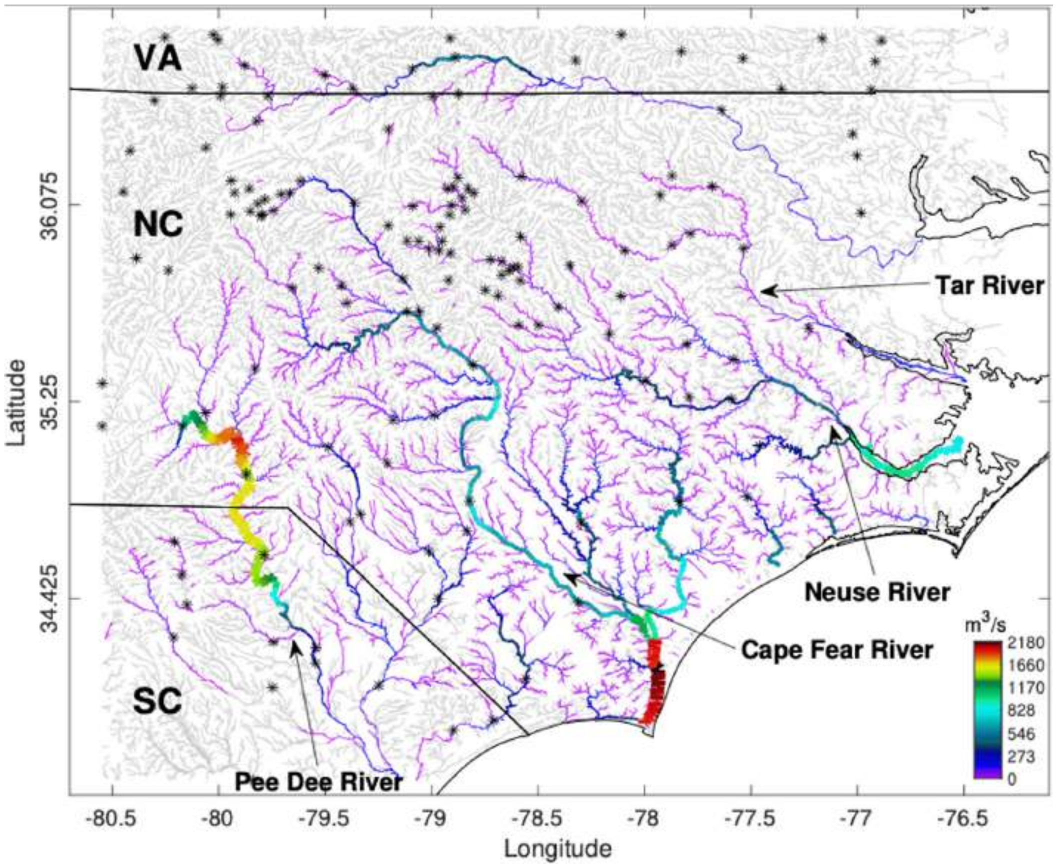
$\rho \circ \widehat{\mathbf{P}}^f$

"Schur Product"

Sea-level pressure correlations. Adopted from Hamill (2003)

3.3 Along-The-Stream Localization

- Given a stream network (unstructured grid), **How to localize the impact of the gauge to neighboring streams?**



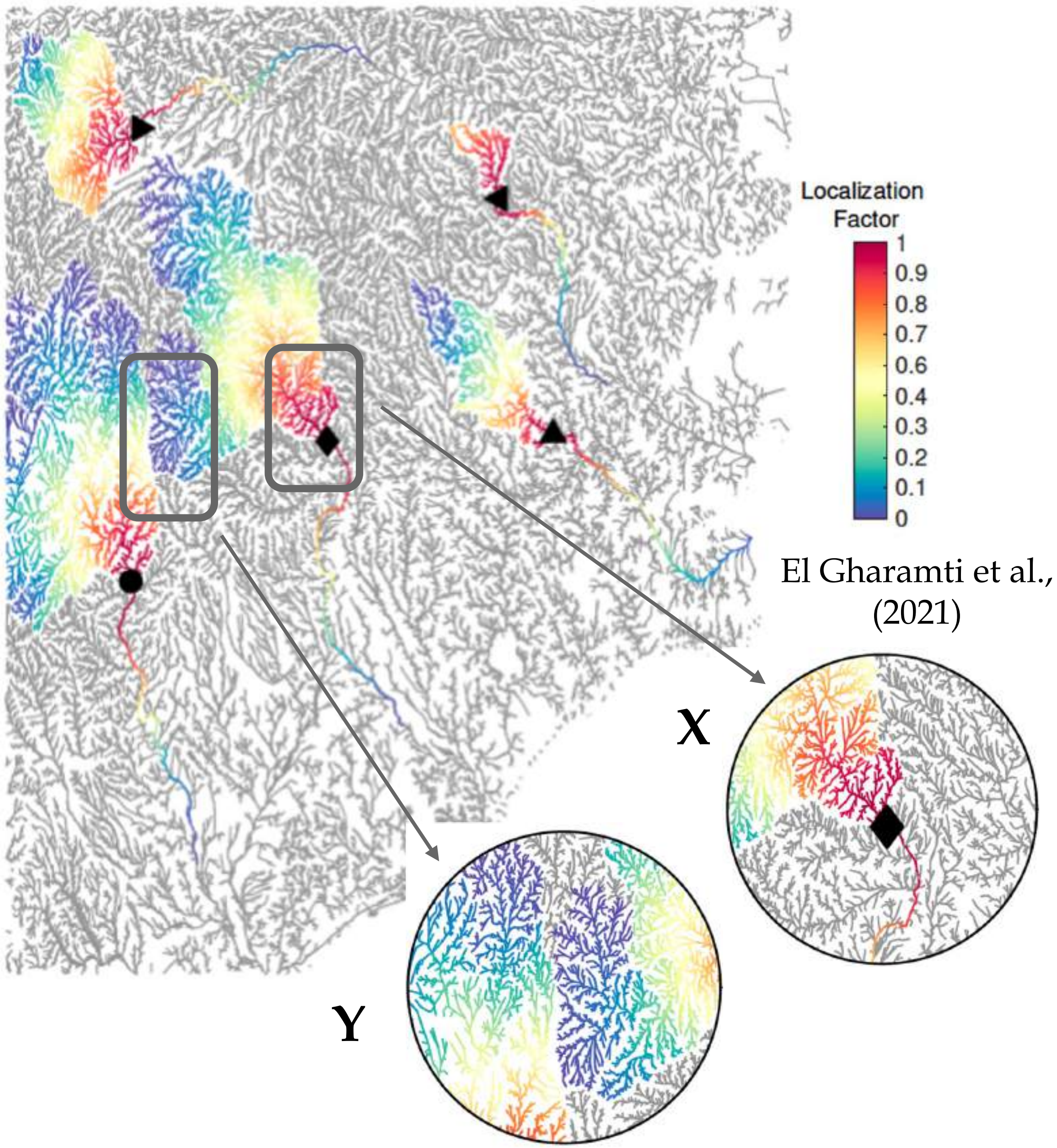
3.3 Along-The-Stream Localization

- ❑ Given a stream network (unstructured grid), **How to localize the impact of the gauge to neighboring streams?**

- ❑ **Along-The-Stream (ATS) Localization**
  - ❑ A topological localization strategy
  - ❑ Adheres to the stream network
  - ❑ Improves information propagation

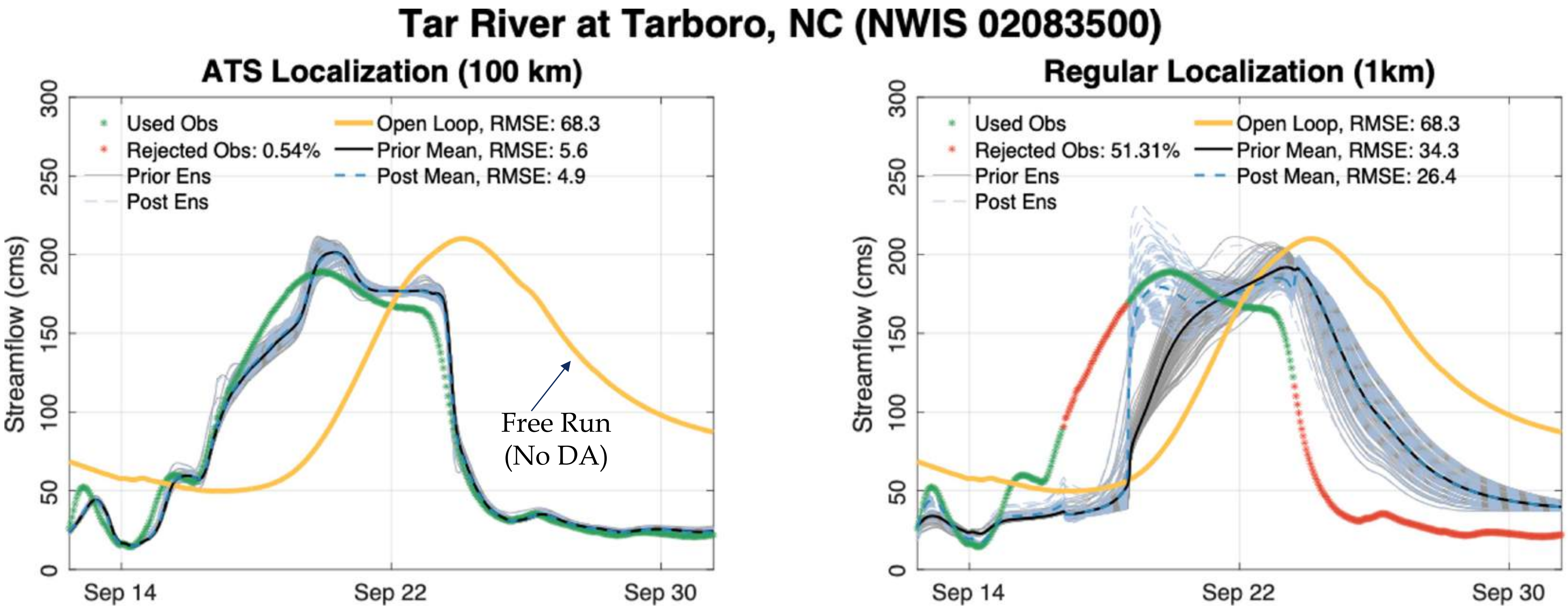
- ★ **Functionality:**
  - ○ Two reaches could be physically close but unrelated if they belong to different catchments
  - ○ **ATS localization** mitigates not only spurious correlations but also physically incorrect ones between unconnected state variables

X: Tree-like shapes: Downstream from a gauge, information flows only downstream
Y: Observations in **different catchments do not have common close reaches**

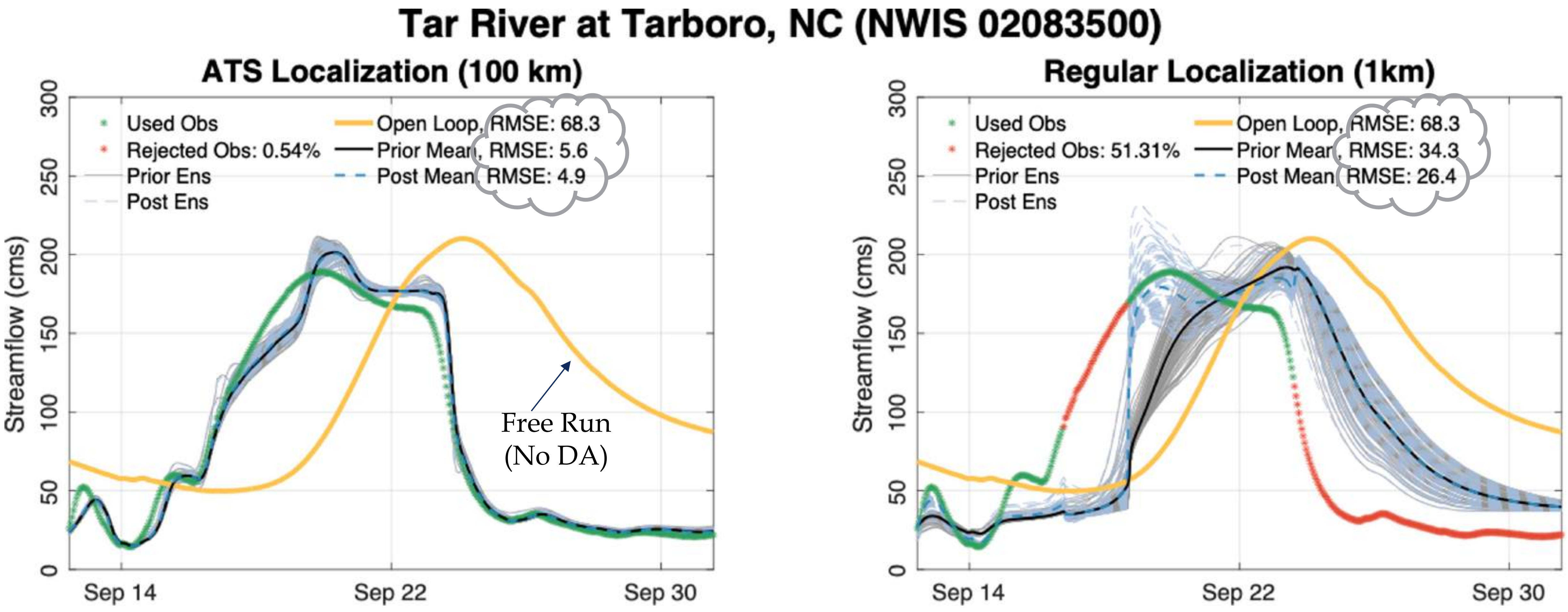


El Gharamti et al., (2021)

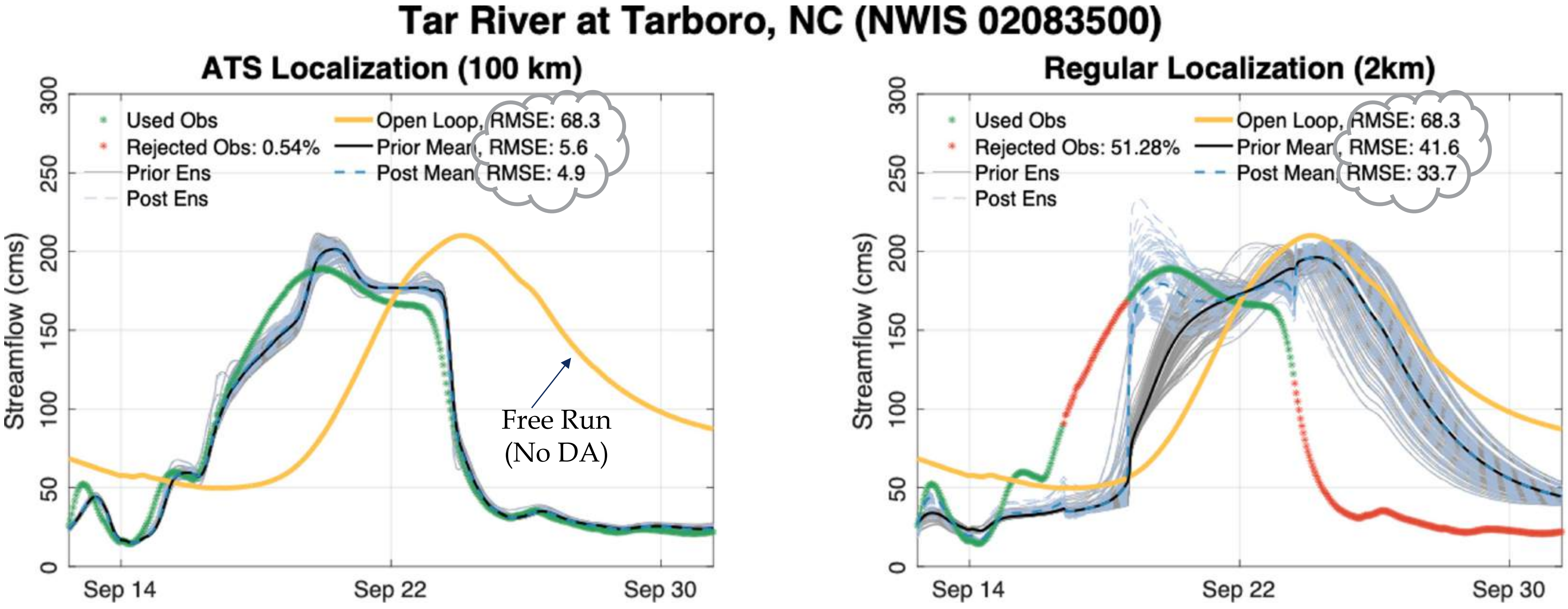
3.3.1 ATS Localization: Performance



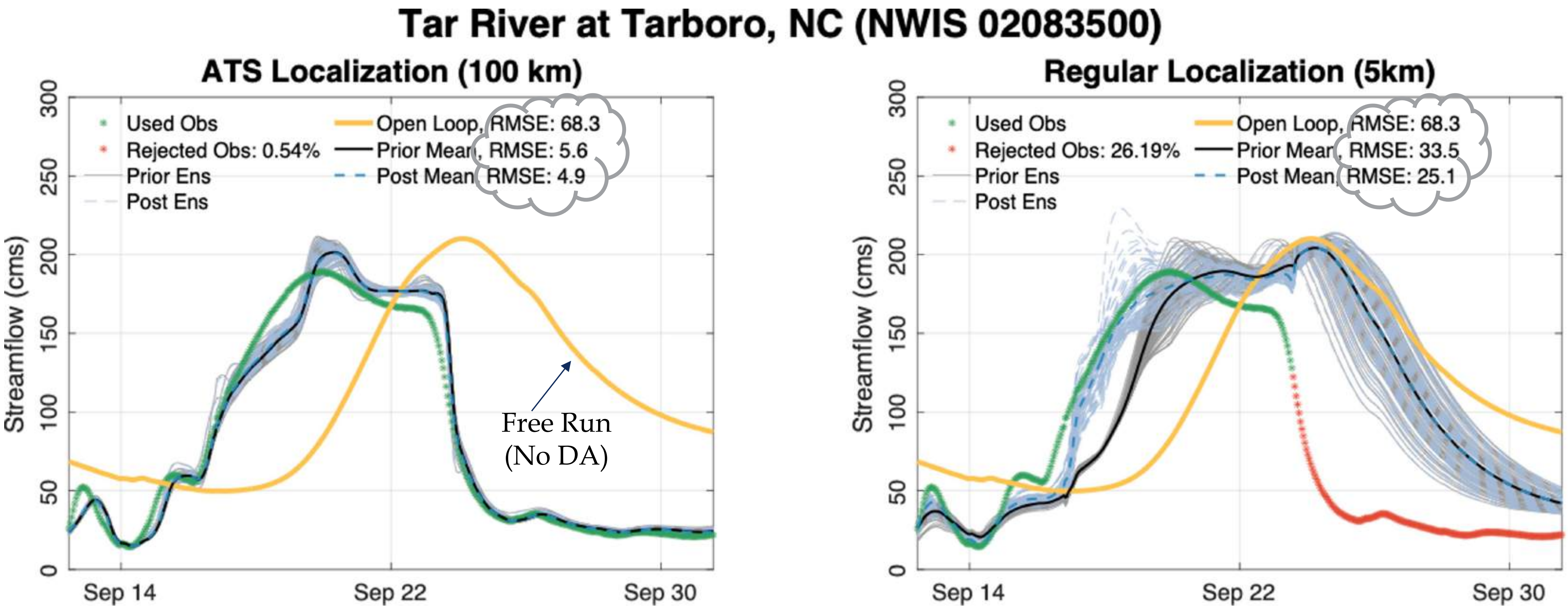
3.3.1 ATS Localization: Performance



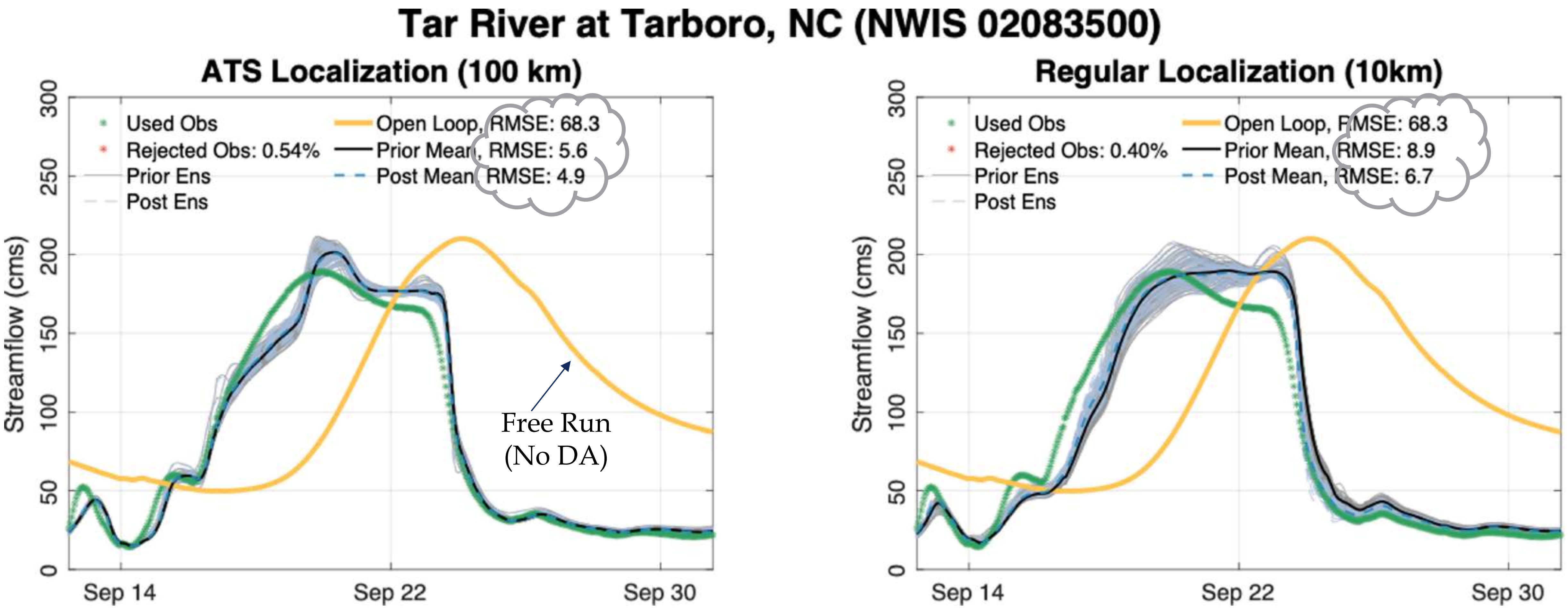
3.3.1 ATS Localization: Performance



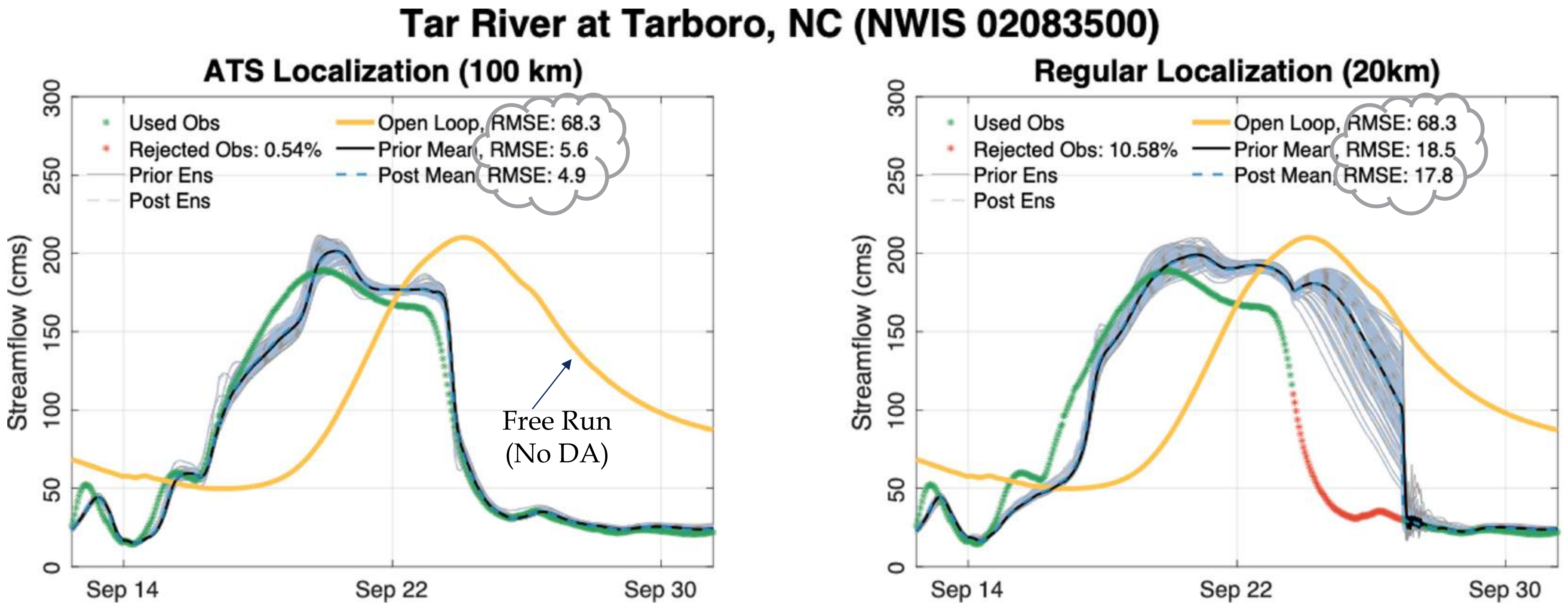
3.3.1 ATS Localization: Performance



3.3.1 ATS Localization: Performance

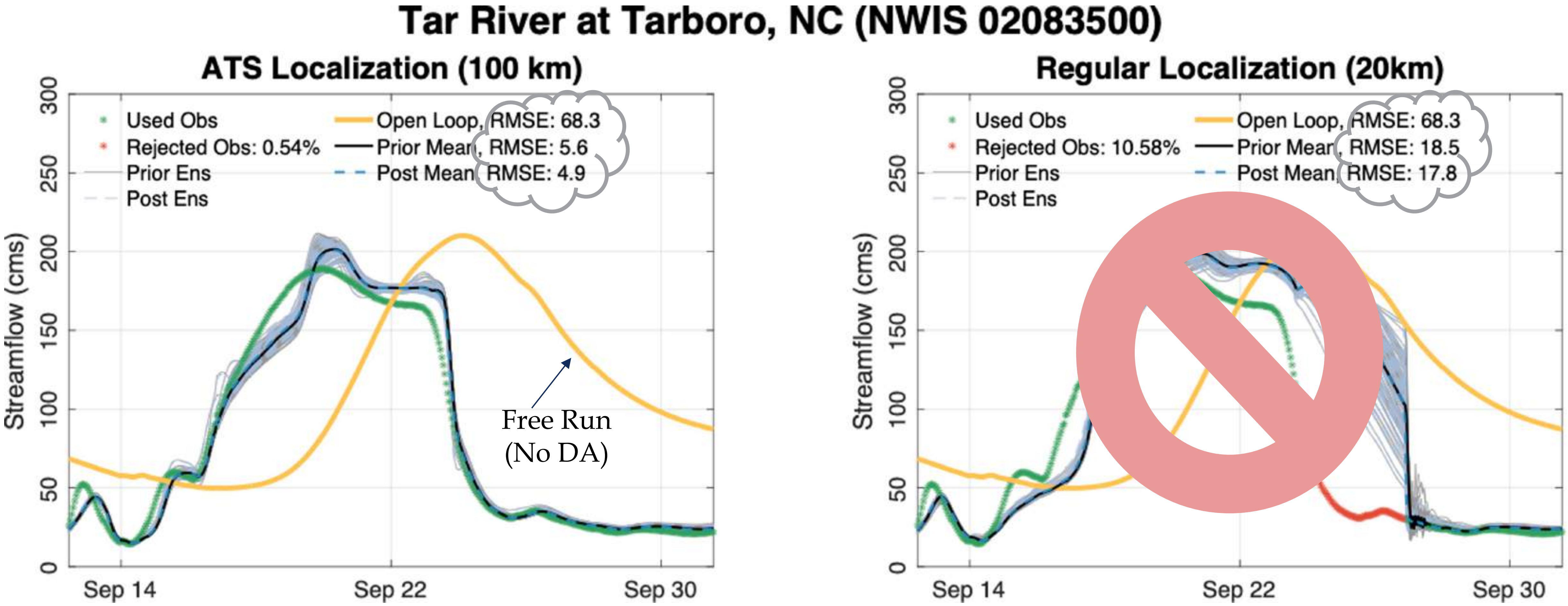


3.3.1 ATS Localization: Performance



ATS enhances the prediction accuracy **(up to 40%)**

3.3.1 ATS Localization: Performance



ATS enhances the prediction accuracy **(up to 40%)**

Regular localization fails with large distances

3.4 Adaptive Covariance Inflation

- Tackle underestimated variance through **Inflation**
  - Inflation increases the variance of the ensemble (mean remains the same)
  - It is equivalent to scaling the covariance

- Counteracts underestimated variance and can be utilized to mitigate model biases
- **Quite effective**; can be applied to the prior and/or posterior ensemble

$$\widetilde{\mathbf{x}}_i^f \leftarrow \sqrt{\lambda}\left[\mathbf{x}_i^f - \overline{\mathbf{x}}^f\right] + \overline{\mathbf{x}}^f$$

Inflation Factor > 1

$$\lambda\left(\rho \circ \widehat{\mathbf{P}}^f\right)$$

3.4 Adaptive Covariance Inflation

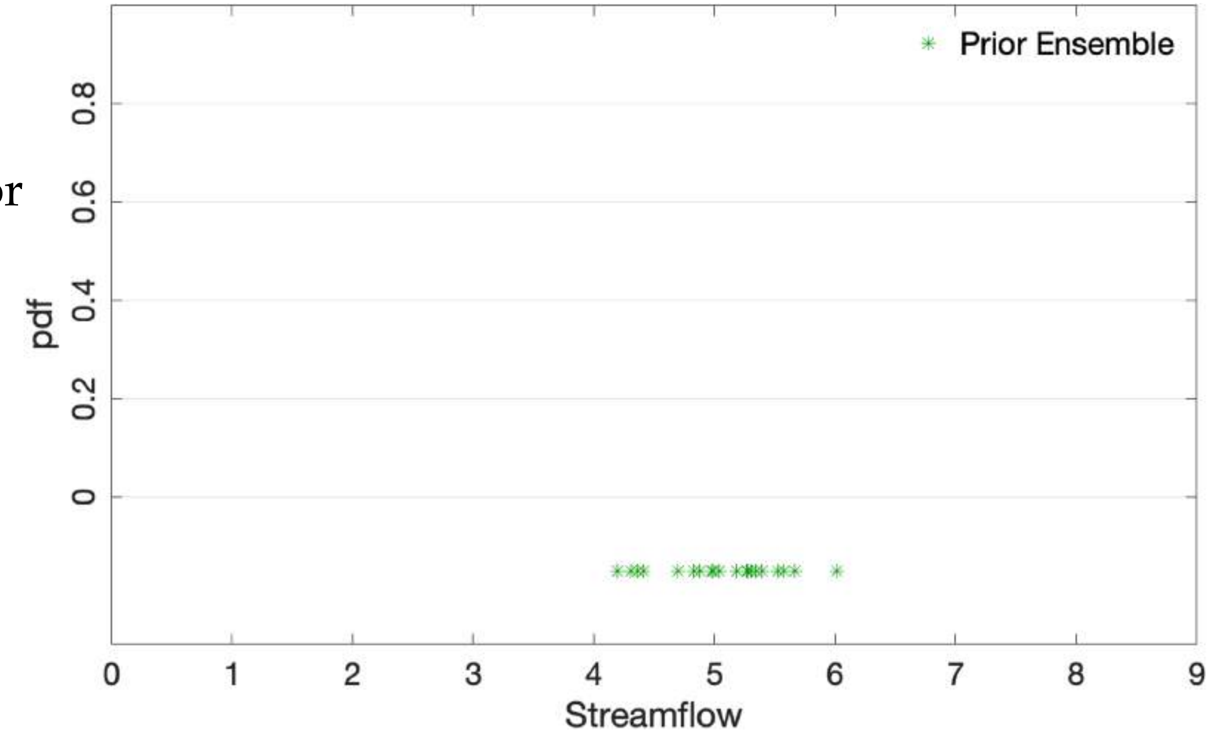
- Tackle underestimated variance through **Inflation**
  - Inflation increases the variance of the ensemble (mean remains the same)
  - It is equivalent to scaling the covariance

- Counteracts underestimated variance and can be utilized to mitigate model biases
- **Quite effective**; can be applied to the prior and/or posterior ensemble

$$\widetilde{\mathbf{x}}_i^f \leftarrow \sqrt{\lambda}\left[\mathbf{x}_i^f - \overline{\mathbf{x}}^f\right] + \overline{\mathbf{x}}^f$$

Inflation Factor > 1

$$\lambda\left(\rho \circ \widehat{\mathbf{P}}^f\right)$$



3.4 Adaptive Covariance Inflation

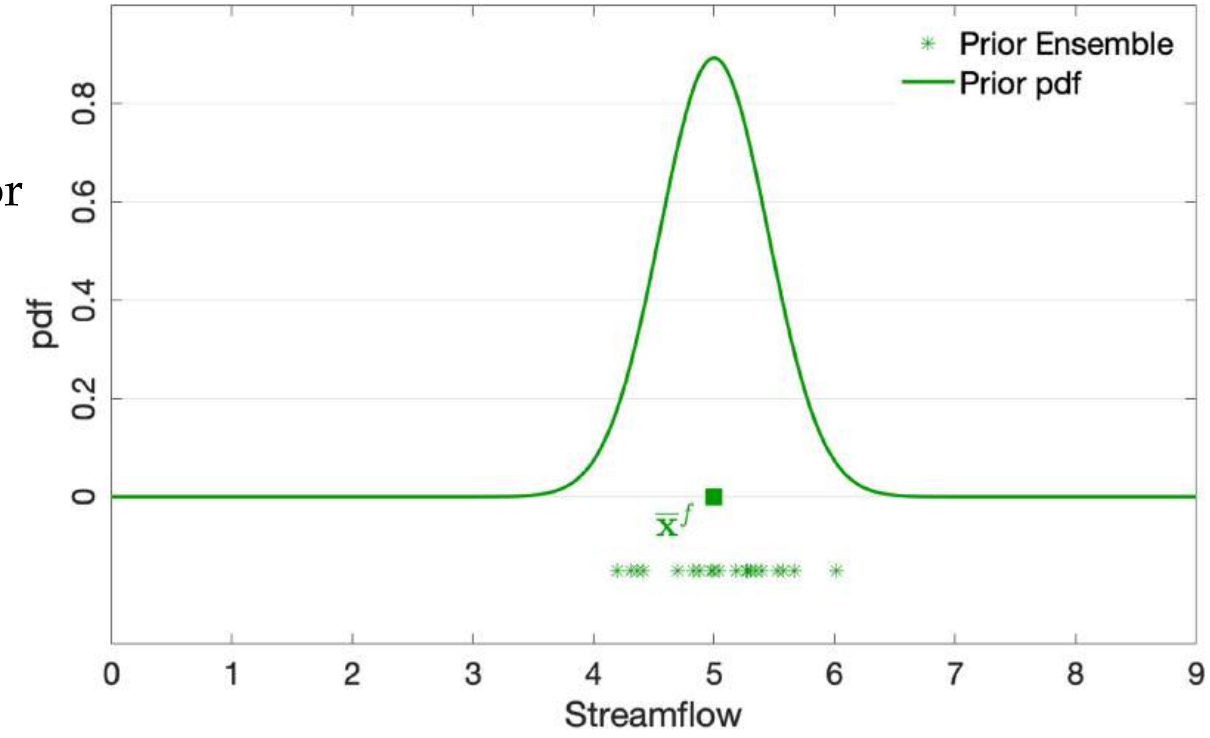
- Tackle underestimated variance through **Inflation**
  - Inflation increases the variance of the ensemble (mean remains the same)
  - It is equivalent to scaling the covariance

- Counteracts underestimated variance and can be utilized to mitigate model biases
- **Quite effective**; can be applied to the prior and/or posterior ensemble

$$\widetilde{\mathbf{x}}_i^f \leftarrow \sqrt{\lambda}\left[\mathbf{x}_i^f - \overline{\mathbf{x}}^f\right] + \overline{\mathbf{x}}^f$$

Inflation Factor > 1

$$\lambda\left(\rho \circ \widehat{\mathbf{P}}^f\right)$$



3.4 Adaptive Covariance Inflation

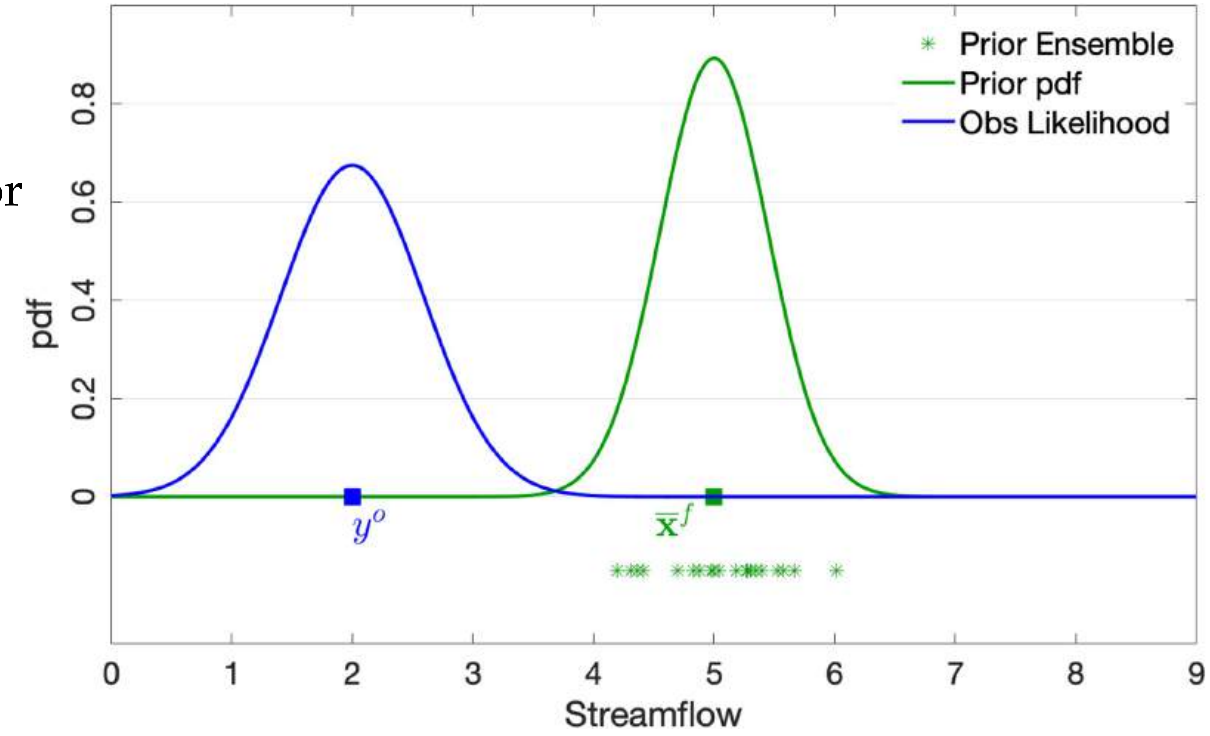
- Tackle underestimated variance through **Inflation**
  - Inflation increases the variance of the ensemble (mean remains the same)
  - It is equivalent to scaling the covariance

- Counteracts underestimated variance and can be utilized to mitigate model biases
- **Quite effective**; can be applied to the prior and/or posterior ensemble

$$\widetilde{\mathbf{x}}_i^f \leftarrow \sqrt{\lambda}\left[\mathbf{x}_i^f - \overline{\mathbf{x}}^f\right] + \overline{\mathbf{x}}^f$$

Inflation Factor > 1

$$\lambda\left(\rho \circ \widehat{\mathbf{P}}^f\right)$$



3.4 Adaptive Covariance Inflation

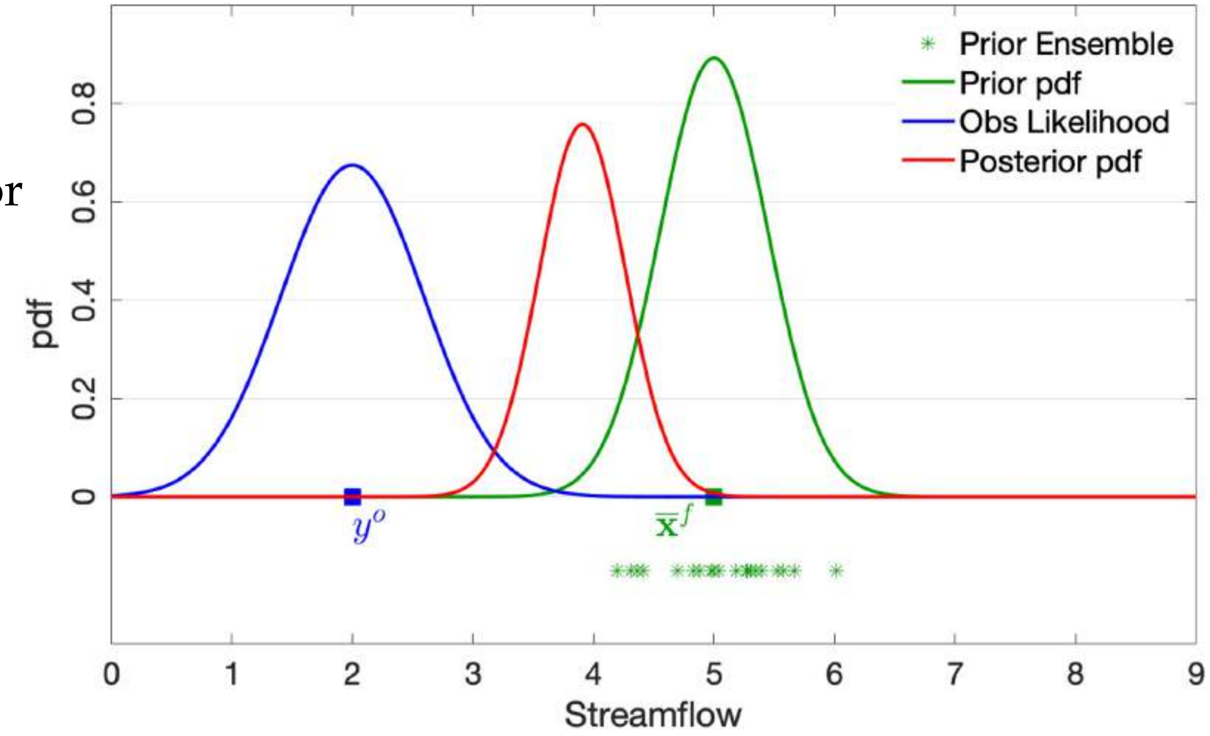
- Tackle underestimated variance through **Inflation**
  - Inflation increases the variance of the ensemble (mean remains the same)
  - It is equivalent to scaling the covariance

- Counteracts underestimated variance and can be utilized to mitigate model biases
- **Quite effective**; can be applied to the prior and/or posterior ensemble

$$\tilde{\mathbf{x}}_i^f \leftarrow \sqrt{\lambda}\left[\mathbf{x}_i^f - \overline{\mathbf{x}}^f\right] + \overline{\mathbf{x}}^f$$

Inflation Factor > 1

$$\lambda\left(\rho \circ \widehat{\mathbf{P}}^f\right)$$



3.4 Adaptive Covariance Inflation

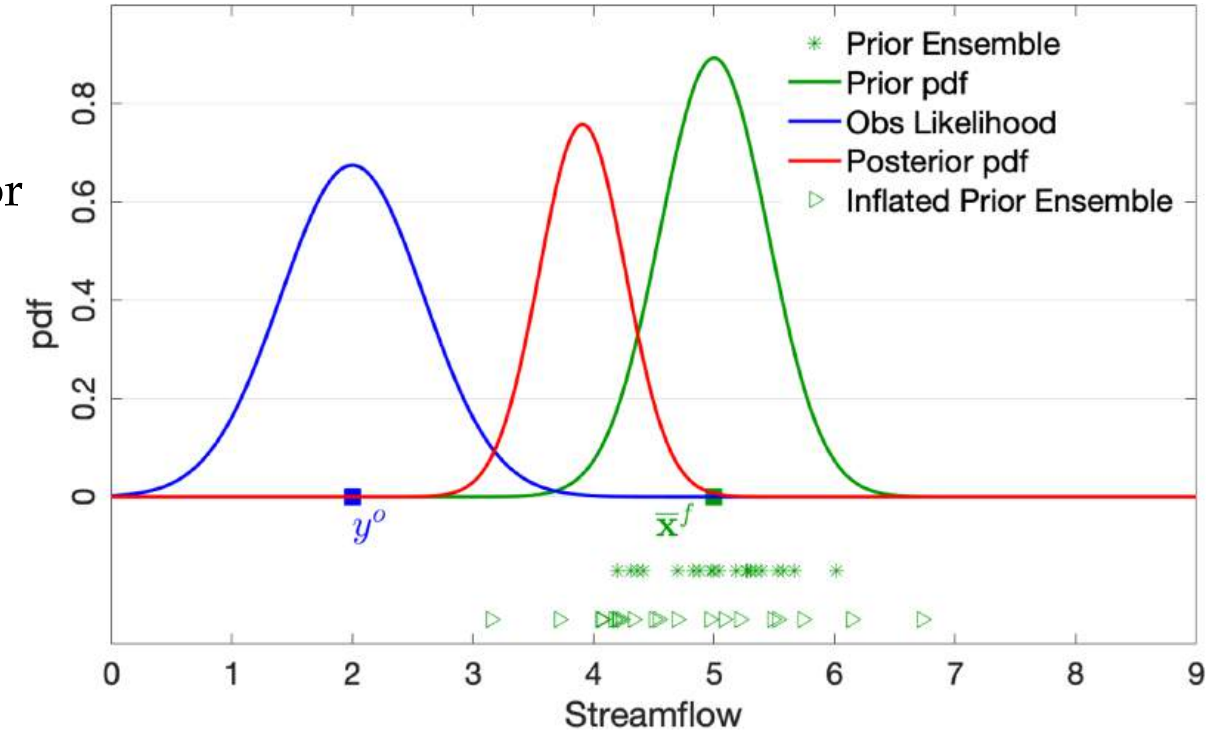
- Tackle underestimated variance through **Inflation**
  - Inflation increases the variance of the ensemble (mean remains the same)
  - It is equivalent to scaling the covariance

- Counteracts underestimated variance and can be utilized to mitigate model biases
- **Quite effective**; can be applied to the prior and/or posterior ensemble

$$\widetilde{\mathbf{x}}_i^f \leftarrow \sqrt{\lambda}\left[\mathbf{x}_i^f - \overline{\mathbf{x}}^f\right] + \overline{\mathbf{x}}^f$$

Inflation Factor > 1

$$\lambda\left(\rho \circ \widehat{\mathbf{P}}^f\right)$$



3.4 Adaptive Covariance Inflation

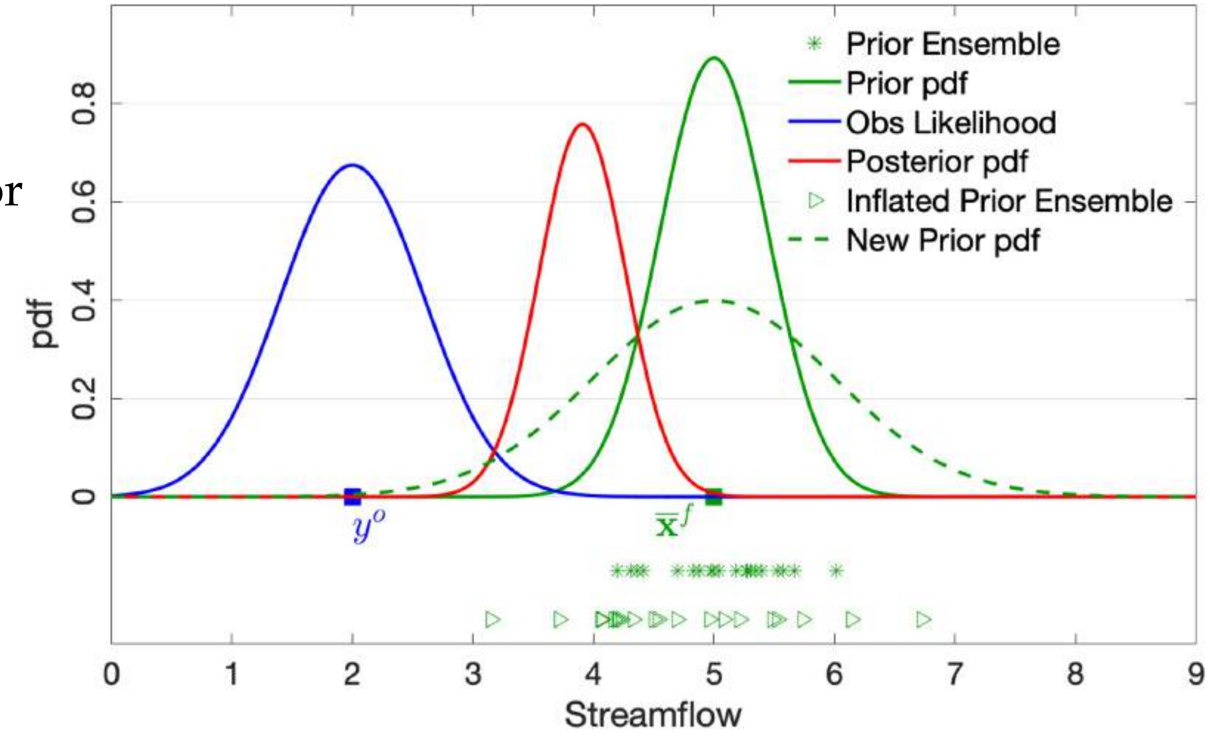
- ❑ Tackle underestimated variance through **Inflation**
  - ❑ Inflation increases the variance of the ensemble (mean remains the same)
  - ❑ It is equivalent to scaling the covariance

- ❑ Counteracts underestimated variance and can be utilized to mitigate model biases
- ❑ **Quite effective**; can be applied to the prior and/or posterior ensemble

$$\widetilde{\mathbf{x}}_i^f \leftarrow \sqrt{\lambda}\left[\mathbf{x}_i^f - \overline{\mathbf{x}}^f\right] + \overline{\mathbf{x}}^f$$

Inflation Factor > 1

$$\lambda\left(\rho \circ \widehat{\mathbf{P}}^f\right)$$



3.4 Adaptive Covariance Inflation

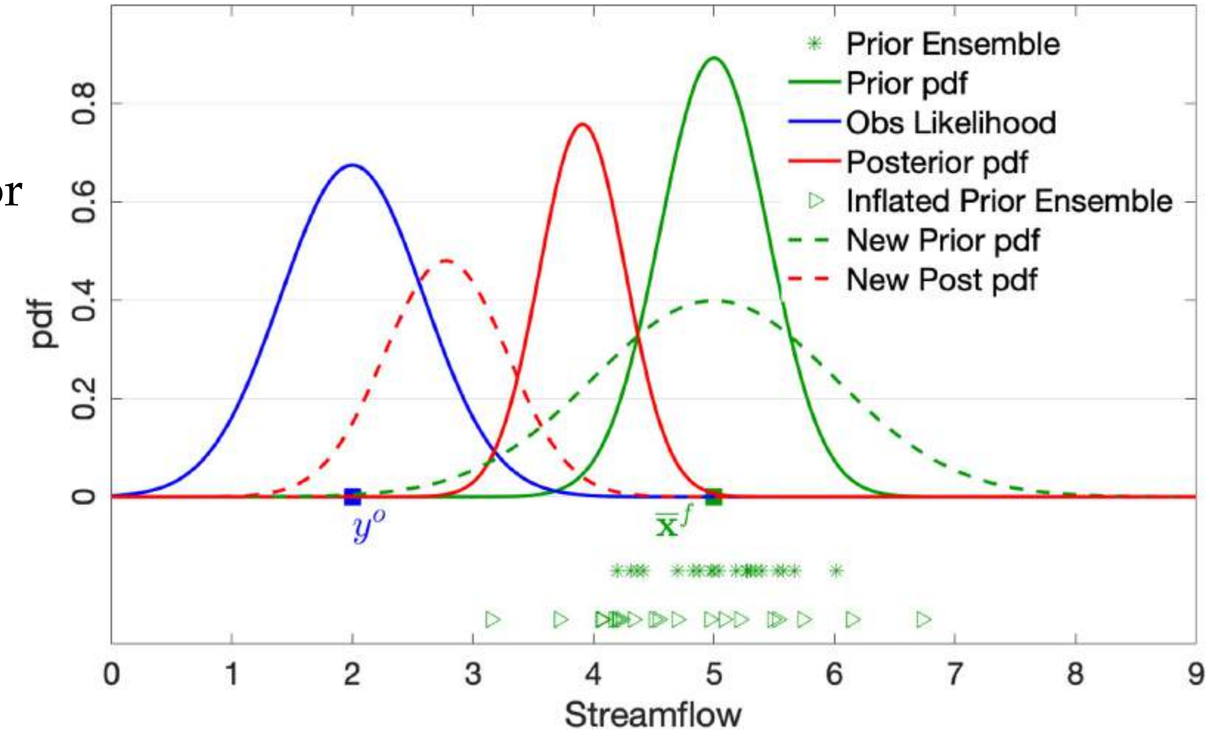
- Tackle underestimated variance through **Inflation**
  - Inflation increases the variance of the ensemble (mean remains the same)
  - It is equivalent to scaling the covariance

- Counteracts underestimated variance and can be utilized to mitigate model biases
- **Quite effective**; can be applied to the prior and/or posterior ensemble

$$\tilde{\mathbf{x}}_i^f \leftarrow \sqrt{\lambda}\left[\mathbf{x}_i^f - \overline{\mathbf{x}}^f\right] + \overline{\mathbf{x}}^f$$

Inflation Factor > 1

$$\lambda\left(\rho \circ \widehat{\mathbf{P}}^f\right)$$



3.4 Adaptive Covariance Inflation

- Tackle underestimated variance through **Inflation**
  - Inflation increases the variance of the ensemble (mean remains the same)
  - It is equivalent to scaling the covariance

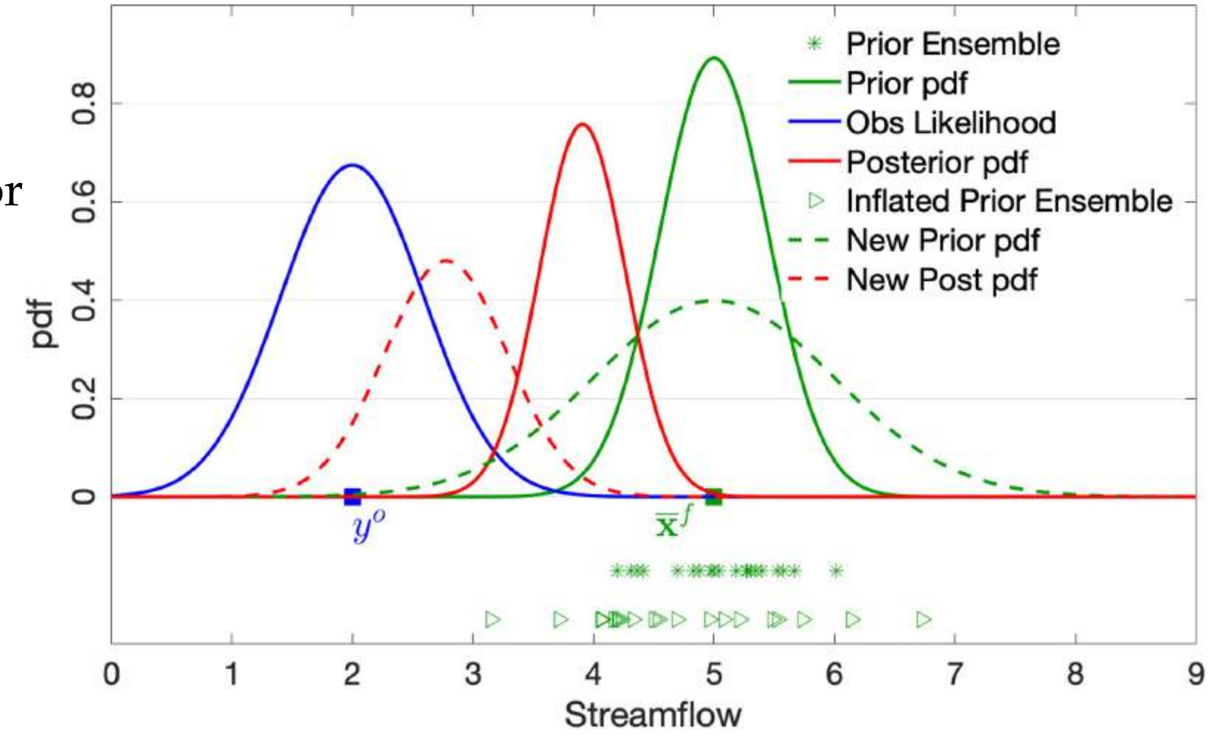
- Counteracts underestimated variance and can be utilized to mitigate model biases
- **Quite effective**; can be applied to the prior and/or posterior ensemble

$$\widetilde{\mathbf{x}}_i^f \leftarrow \sqrt{\lambda}\left[\mathbf{x}_i^f - \overline{\mathbf{x}}^f\right] + \overline{\mathbf{x}}^f$$

Inflation Factor > 1

$$\lambda\left(\rho \circ \widehat{\mathbf{P}}^f\right)$$

What is an appropriate inflation value? How to choose it?



3.4 Adaptive Covariance Inflation

Spatially and Temporally Varying Adaptive Scheme

[MWR: El Gharamti, 2018]

- ❑ Inflation factor is assumed to be a random variable
- ❑ Use the available streamflow data to estimate it

$$p(\lambda|y) \propto p(\lambda) \cdot p(y|\lambda)$$

3.4 Adaptive Covariance Inflation

Spatially and Temporally Varying Adaptive Scheme

[MWR: El Gharamti, 2018]

- ❑ Inflation factor is assumed to be a random variable
- ❑ Use the available streamflow data to estimate it

$$p(\lambda|y) \propto p(\lambda) \cdot p(y|\lambda)$$

FEBRUARY 2018 GHARAMTI 623

Enhanced Adaptive Inflation Algorithm for Ensemble Filters

MOHAMAD EL GHARAMTI

National Center for Atmospheric Research, Boulder, Colorado

(Manuscript received 27 June 2017, in final form 2 January 2018)

ABSTRACT

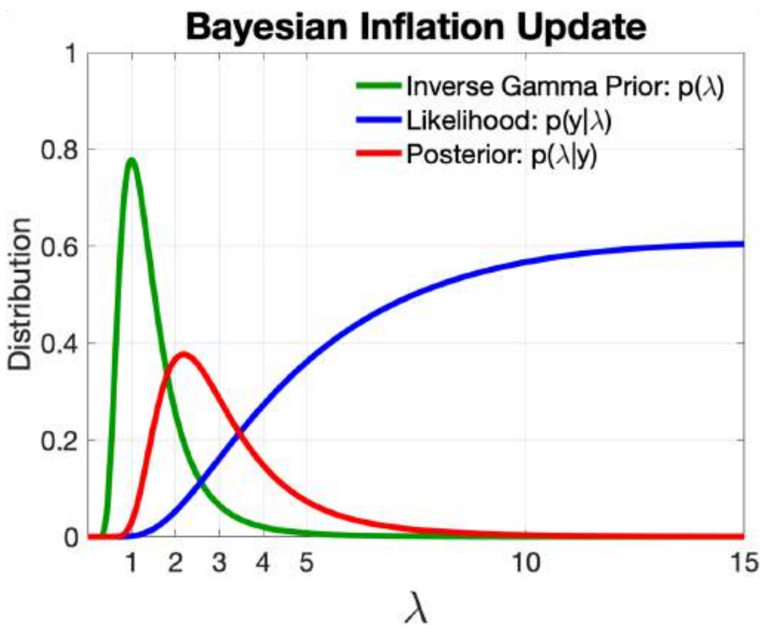
Spatially and temporally varying adaptive inflation algorithms have been developed to combat the loss of variance during the forecast due to various model and sampling errors. The adaptive Bayesian scheme of Anderson uses available observations to update the Gaussian inflation distribution assigned for every state variable. The likelihood function of the inflation is computed using model-minus-data innovation statistics. A number of enhancements for this inflation scheme are proposed. To prevent excessive deflation, an inverse gamma distribution for the prior inflation is considered. A non-Gaussian distribution offers a flexible framework for the inflation variance to evolve during the update. The innovations are assumed random variables, and a correction term is added to the mode of the likelihood distribution such that the observed inflation is slightly larger. This modification improves the stability of the adaptive scheme by limiting the occurrence of negative and physically intolerable inflations. The enhanced scheme is compared to the original one in twin experiments using the Lorenz-63 model, the Lorenz-96 model, and an idealized, high-dimensional atmospheric model. Results show that the proposed enhancements are capable of generating accurate and consistent state estimates. Allowing moderate deflation is shown to be useful.

1. Introduction

The ensemble Kalman filter (EnKF) is an efficient estimation tool that has been used extensively in many geophysical applications during the last two decades (Evensen 2009). It operates sequentially by combining information from a prior state of a system and its associated variance with noisy observations to produce an analyzed posterior state with an updated uncertainty (Evensen 2003). Inflation has been introduced into stochastic (e.g., EnKF) and deterministic (e.g., Anderson 2001; Hoteit et al. 2002; Whitaker and Hamill 2002; Hunt et al. 2007; Sakov and Oke 2008) ensemble filters as a way to counteract the underestimation of the true variance. Large portions of the variance are often lost during the forecast

Systematic observational and representativeness errors within the analysis step are also expected to lead to underestimates of the ensemble variance (Furrer and Bengtsson 2007).

Studies within the data assimilation (DA) literature have proposed a number of techniques that deal with all kinds of system errors. The majority of these techniques, apart from localization, which removes spurious correlations from the prior covariance, can be seen as a form of inflation. These can be split into four distinct categories. The first category is background covariance inflation, which itself is divided into additive and multiplicative inflation. Additive inflation (Mitchell and Houtekamer 2000) adds a random perturbation drawn from a specific



3.4 Adaptive Covariance Inflation

Spatially and Temporally Varying Adaptive Scheme

[MWR: El Gharamti, 2018]

- ❑ Inflation factor is assumed to be a random variable
- ❑ Use the available streamflow data to estimate it

$$p(\lambda|y) \propto p(\lambda) \cdot p(y|\lambda)$$

FEBRUARY 2018 GHARAMTI 623

Enhanced Adaptive Inflation Algorithm for Ensemble Filters

MOHAMAD EL GHARAMTI

National Center for Atmospheric Research, Boulder, Colorado

(Manuscript received 27 June 2017, in final form 2 January 2018)

ABSTRACT

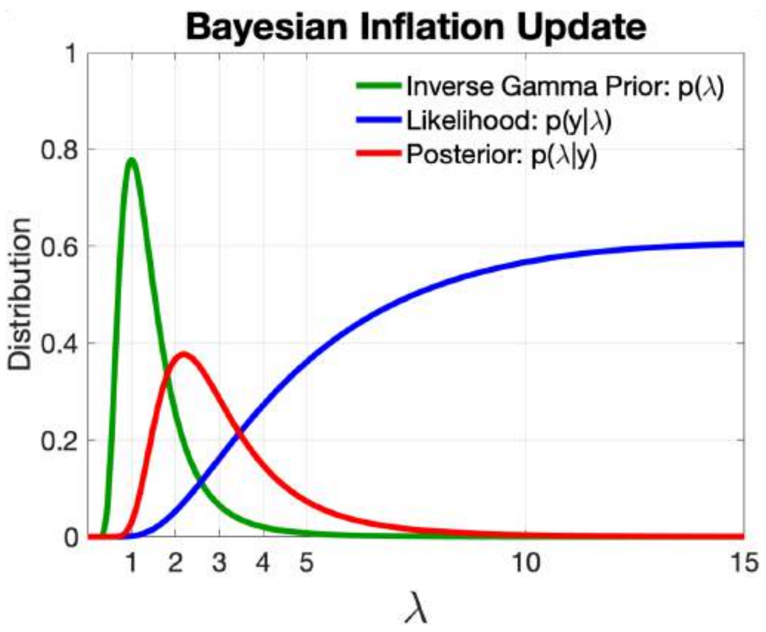
Spatially and temporally varying adaptive inflation algorithms have been developed to combat the loss of variance during the forecast due to various model and sampling errors. The adaptive Bayesian scheme of Anderson uses available observations to update the Gaussian inflation distribution assigned for every state variable. The likelihood function of the inflation is computed using model-minus-data innovation statistics. A number of enhancements for this inflation scheme are proposed. To prevent excessive deflation, an inverse gamma distribution for the prior inflation is considered. A non-Gaussian distribution offers a flexible framework for the inflation variance to evolve during the update. The innovations are assumed random variables, and a correction term is added to the mode of the likelihood distribution such that the observed inflation is slightly larger. This modification improves the stability of the adaptive scheme by limiting the occurrence of negative and physically intolerable inflations. The enhanced scheme is compared to the original one in twin experiments using the Lorenz-63 model, the Lorenz-96 model, and an idealized, high-dimensional atmospheric model. Results show that the proposed enhancements are capable of generating accurate and consistent state estimates. Allowing moderate deflation is shown to be useful.

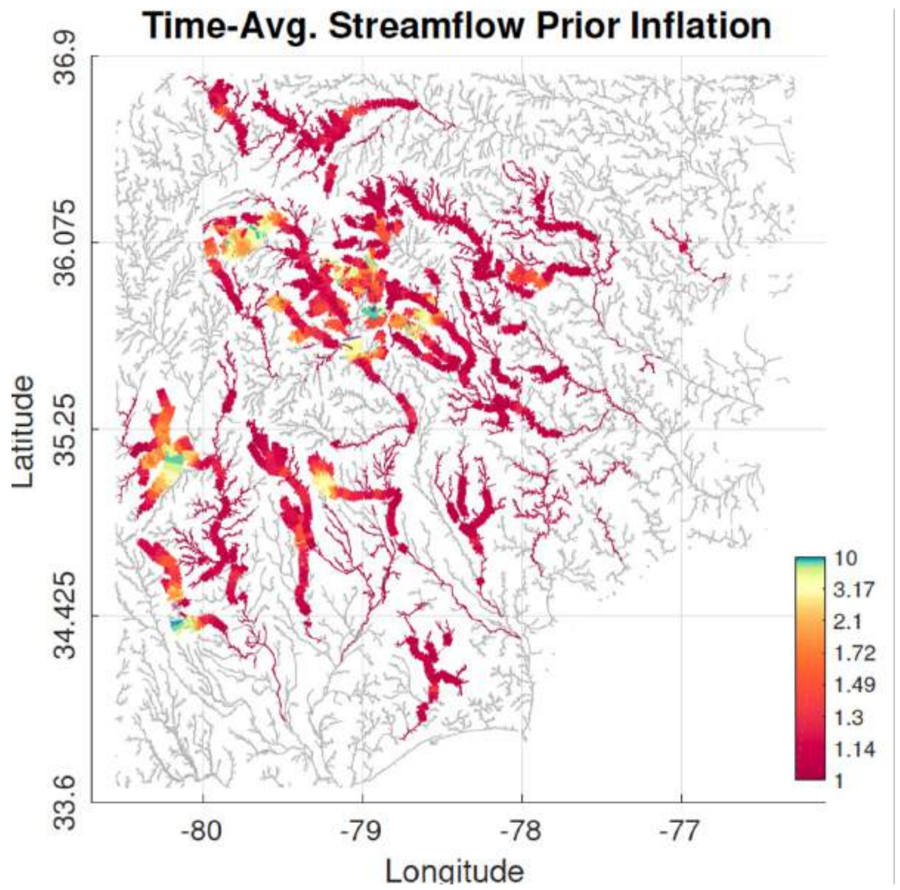
1. Introduction

The ensemble Kalman filter (EnKF) is an efficient estimation tool that has been used extensively in many geophysical applications during the last two decades (Evensen 2009). It operates sequentially by combining information from a prior state of a system and its associated variance with noisy observations to produce an analyzed posterior state with an updated uncertainty (Evensen 2003). Inflation has been introduced into stochastic (e.g., EnKF) and deterministic (e.g., Anderson 2001; Hoteit et al. 2002; Whitaker and Hamill 2002; Hunt et al. 2007; Sakov and Oke 2008) ensemble filters as a way to counteract the underestimation of the true variance. Large portions of the variance are often lost during the forecast

Systematic observational and representativeness errors within the analysis step are also expected to lead to underestimates of the ensemble variance (Furrer and Bengtsson 2007).

Studies within the data assimilation (DA) literature have proposed a number of techniques that deal with all kinds of system errors. The majority of these techniques, apart from localization, which removes spurious correlations from the prior covariance, can be seen as a form of inflation. These can be split into four distinct categories. The first category is background covariance inflation, which itself is divided into additive and multiplicative inflation. Additive inflation (Mitchell and Houtekamer 2000) adds a random perturbation drawn from a specific





3.4.1 Adaptive Inflation: Algorithm

$$\textbf{Innovations:}\ d = y - x_b = \varepsilon_o + \varepsilon_b,$$

$$\mathbb{E}\left[d^2\right] = \mathbb{E}(\varepsilon_o^2) + \mathbb{E}(\varepsilon_b^2) + 2\mathbb{E}\left(\varepsilon_o\varepsilon_b\right),$$

$$= \sigma_o^2 + \lambda_o\widehat{\sigma}_b^2$$

$$\approx \left[\frac{1}{N}\sum_{i=1}^{N} d_i\right]^2 + \mathbb{V}(d) = \sigma_o^2 + \sigma_b^2 + \frac{N-1}{N}\widehat{\sigma}_b^2,$$

$$\Rightarrow \lambda_o^* = \frac{\mathbb{E}\left[d^2\right] - \sigma_o^2}{\widehat{\sigma}_b^2} + \frac{1}{N} = \lambda_o + \frac{1}{N},$$

$$\lambda_o^k = \left[\gamma(\lambda_b^k - 1) + 1\right]^2, \quad \gamma = \kappa|r|, \quad k = 1, 2, \dots, N_x,$$

$$\theta^2 = \left\{\left[\gamma(\lambda^k - 1) + 1\right]^2 - \frac{1}{N}\right\}\widehat{\sigma}_b^2 + \sigma_o^2,$$

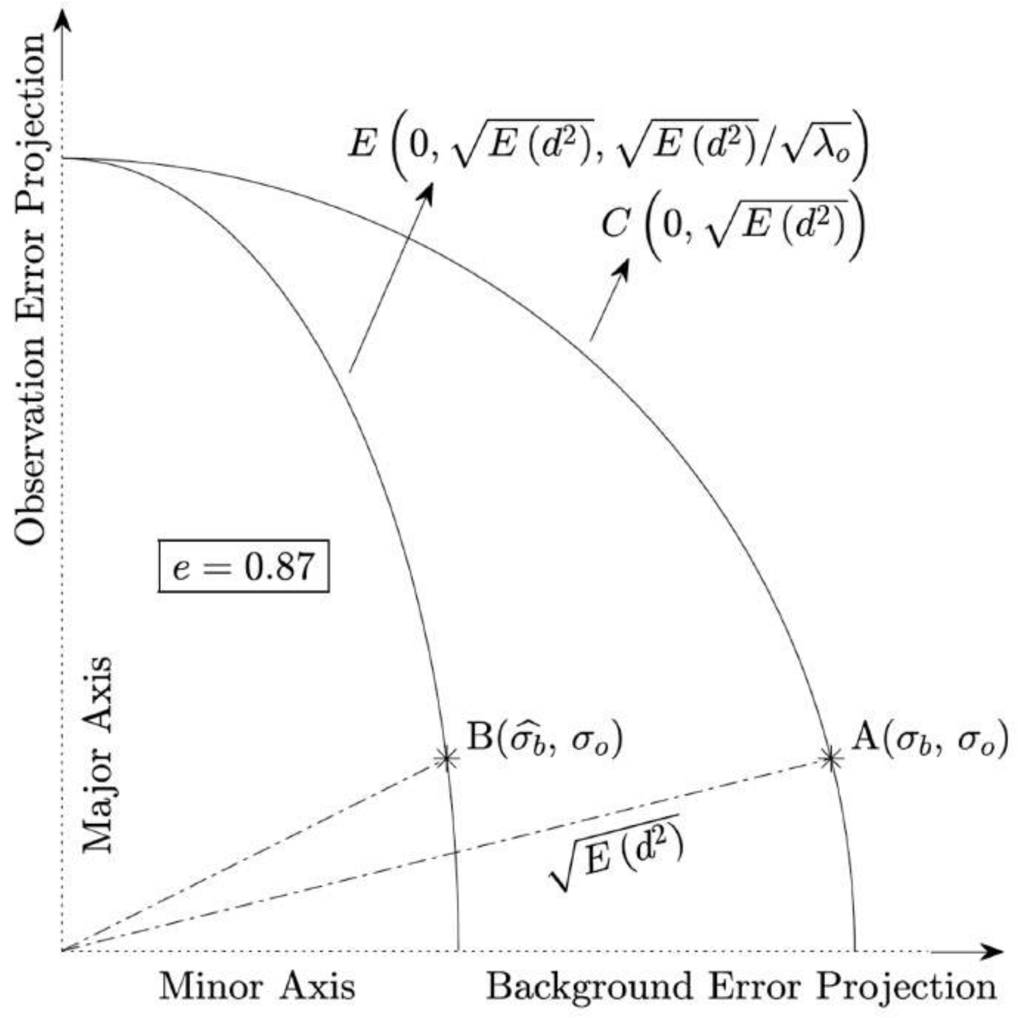
$$\textbf{Likelihood:}\ p\left(\lambda^k|d\right) \propto \frac{1}{2\pi\sigma_{\lambda_b,k}\theta}\exp\left[-\frac{\left(\lambda^k - \lambda_b^k\right)^2}{2\sigma_{\lambda_b,k}^2} - \frac{d^2}{2\theta^2}\right]$$

$$\textbf{Prior:}\ p(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{-\alpha-1}\exp\left[-\frac{\beta}{\lambda}\right]$$

$$\textbf{Posterior:}\ p(d|\lambda) = \frac{\beta^\alpha\lambda^{-\alpha-1}}{\sqrt{2\pi}\theta\Gamma(\alpha)}\exp\left[-\frac{d^2}{2\theta^2} - \frac{\beta}{\lambda}\right],$$

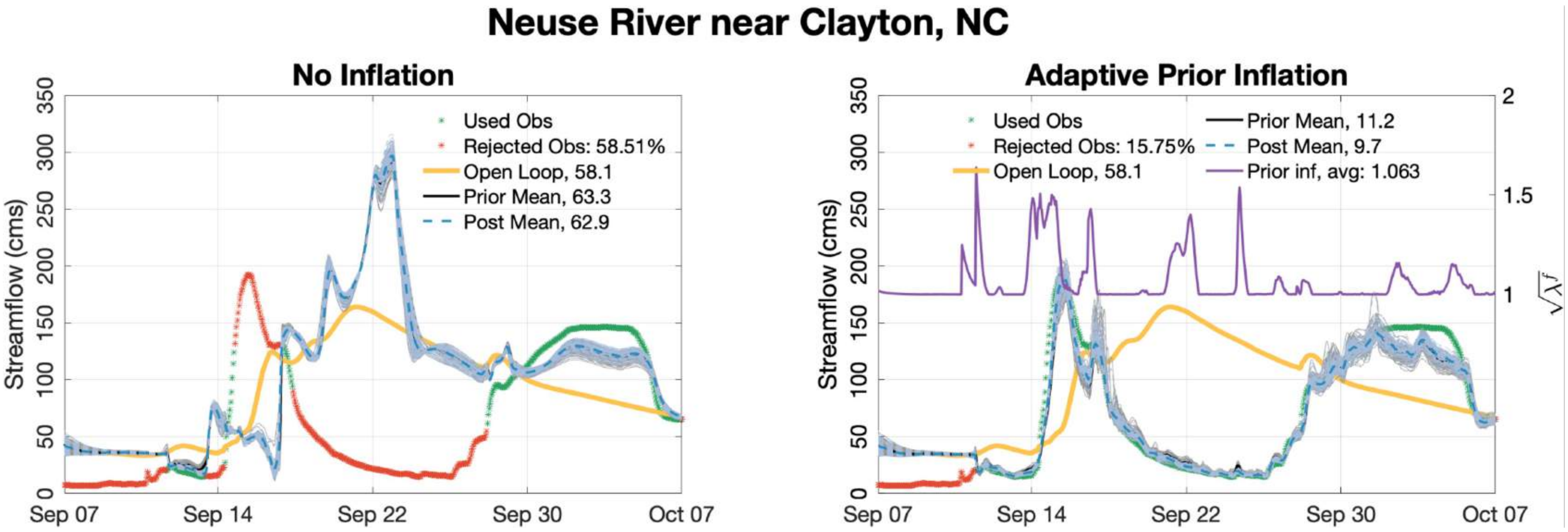
$$\cong p(d|\lambda_b) + \left.\frac{\partial p(d|\lambda)}{\partial\lambda}\right|_{\lambda_b}(\lambda - \lambda_b) + \mathcal{O}\left(\lambda - \lambda_b\right)^2,$$

$$\Rightarrow \left(1 - \frac{\lambda_b}{\beta}\right)\lambda^2 + \left(\frac{\bar{\ell}}{\ell'} - 2\lambda_b\right)\lambda + \left(\lambda_b^2 - \frac{\bar{\ell}}{\ell'}\lambda_b\right) = 0$$

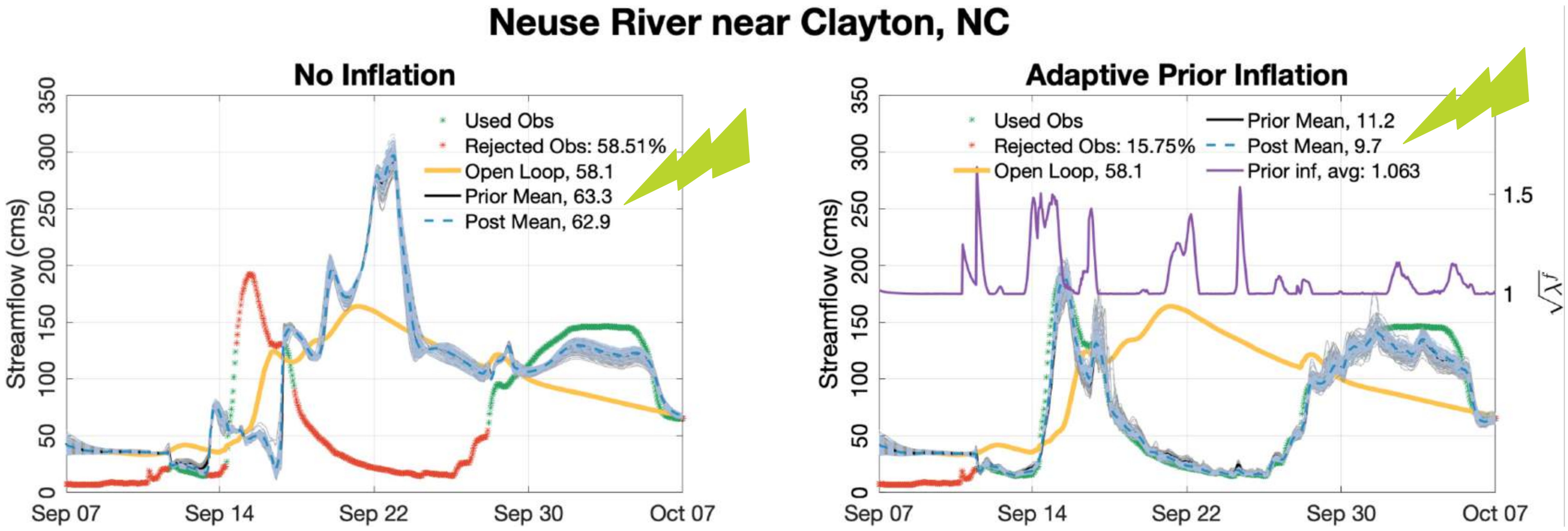


Geometrical interpretation of the adaptive inflation scheme

3.4.2 Inflation: Accurate Streamflow Predictions

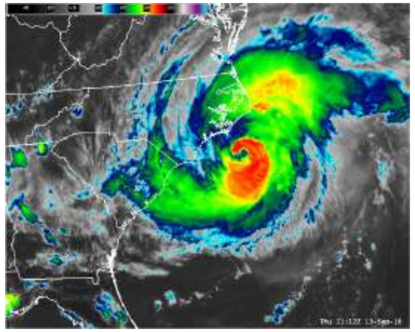


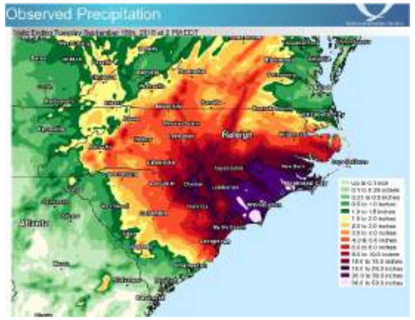
3.4.2 Inflation: Accurate Streamflow Predictions

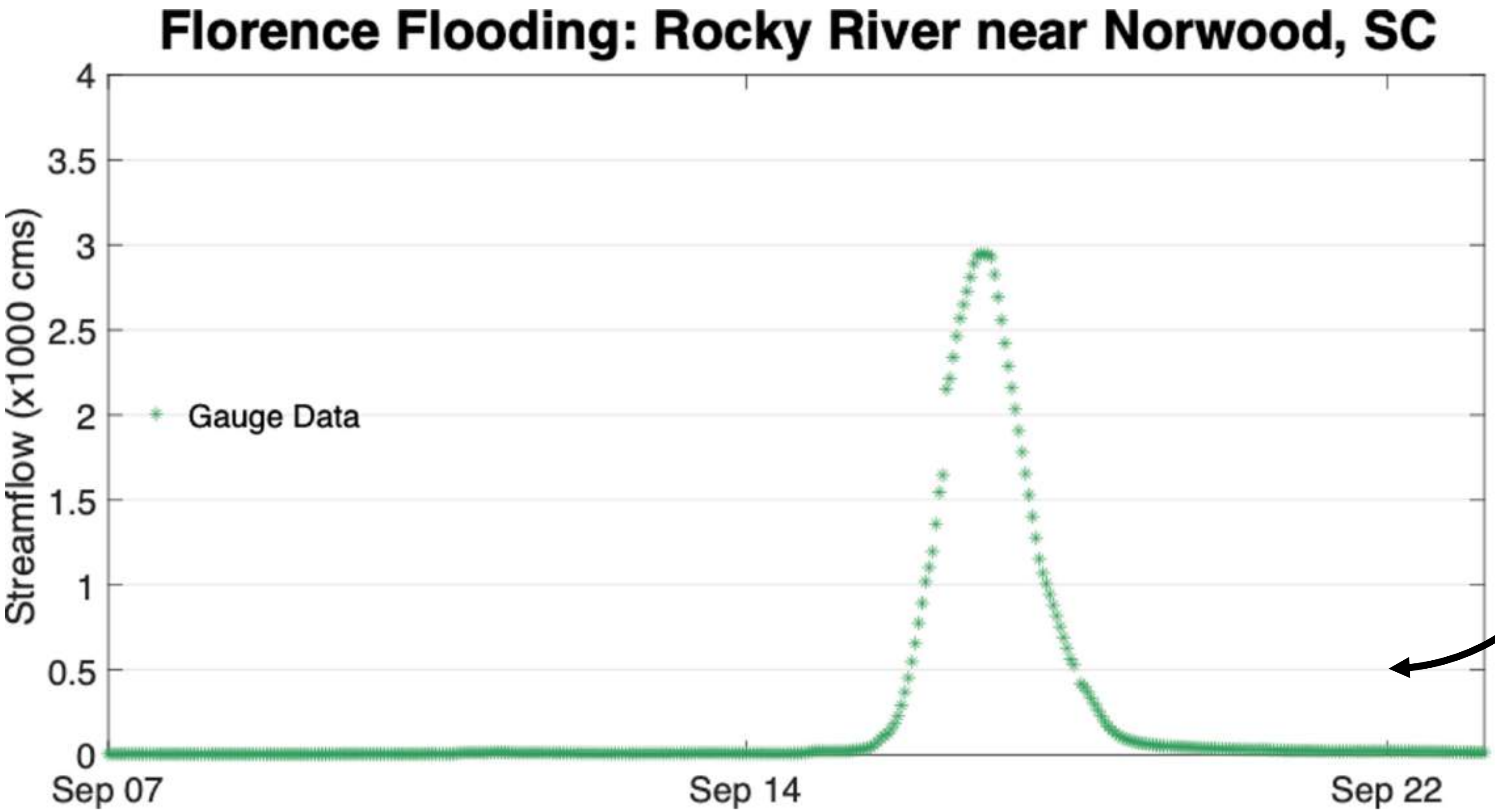


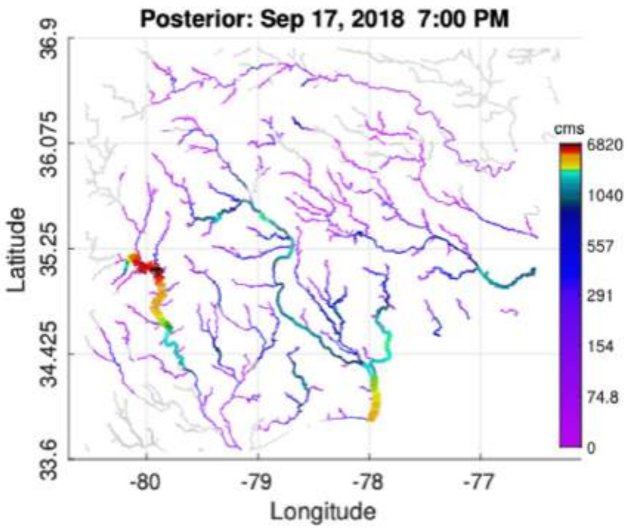
~80% Accuracy

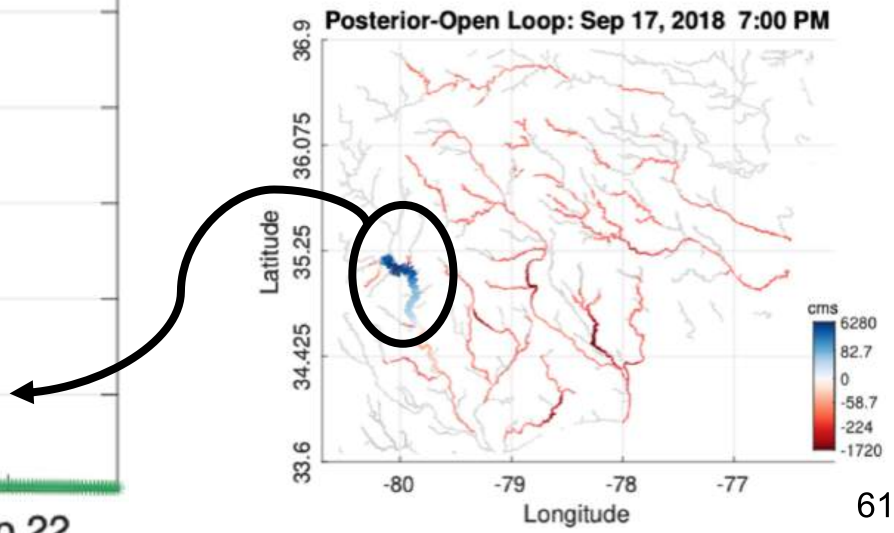
3.4.3 Inflation: Bias Correction Tool



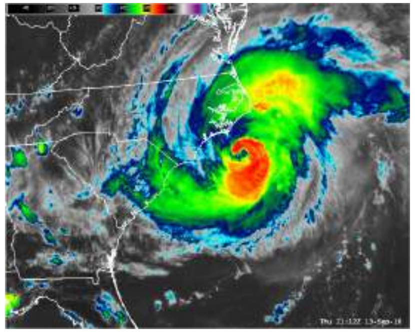


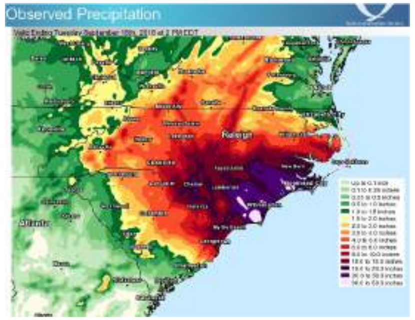


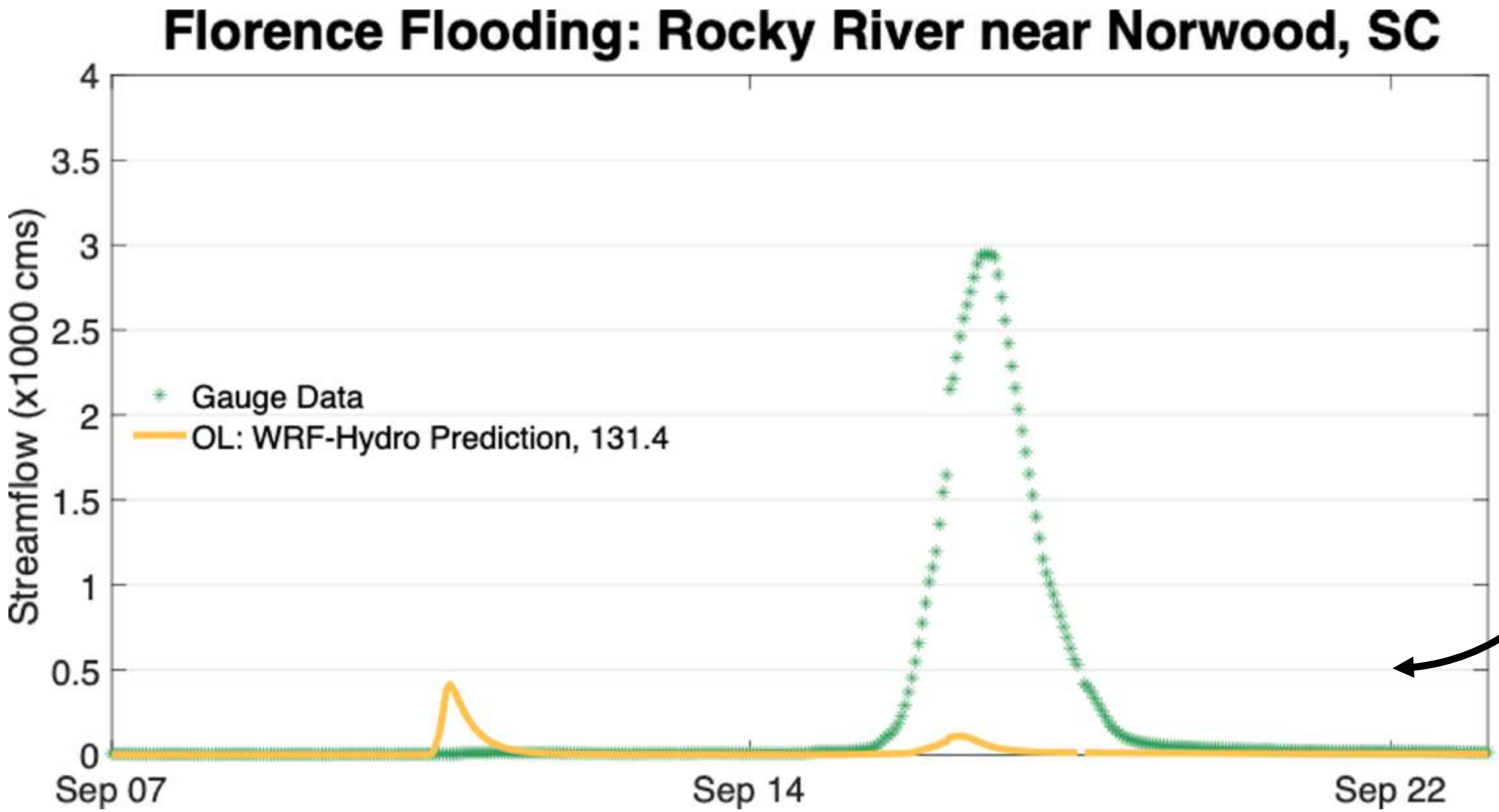


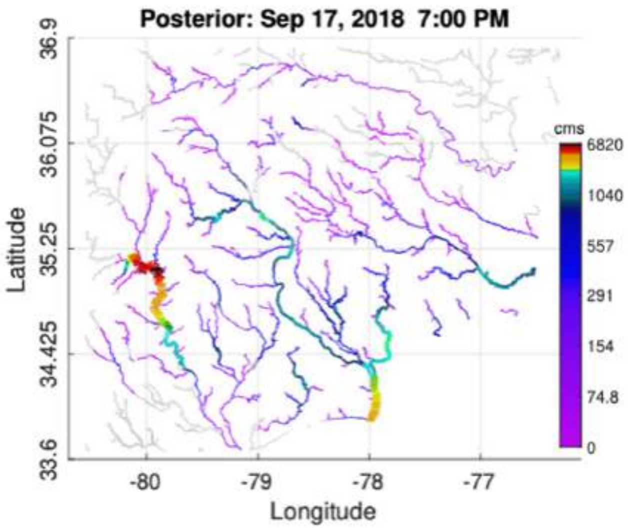


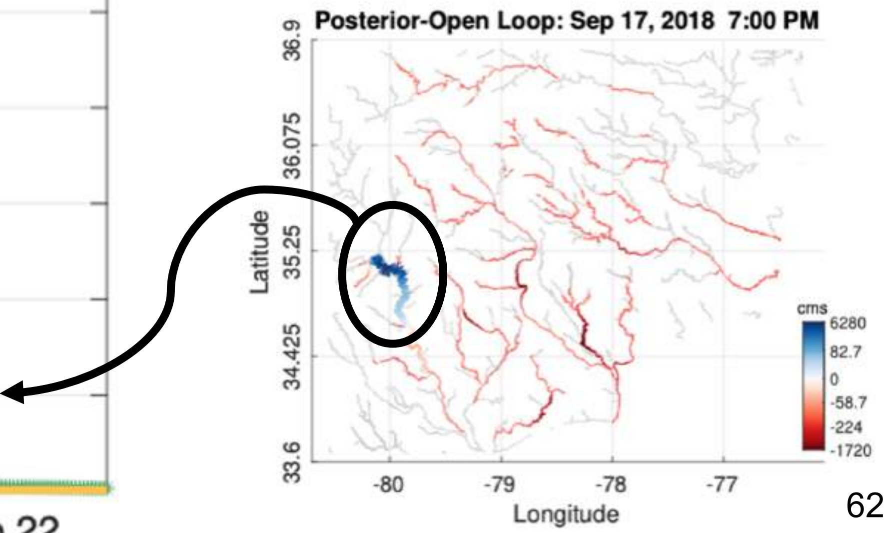
3.4.3 Inflation: Bias Correction Tool



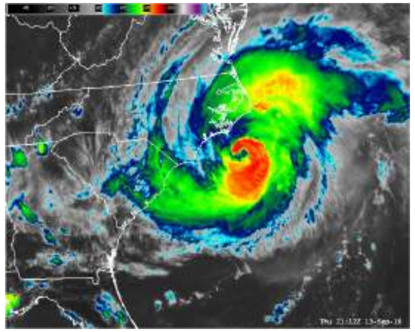


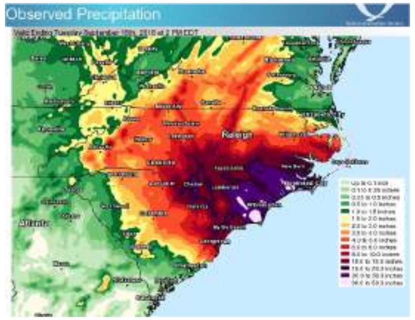


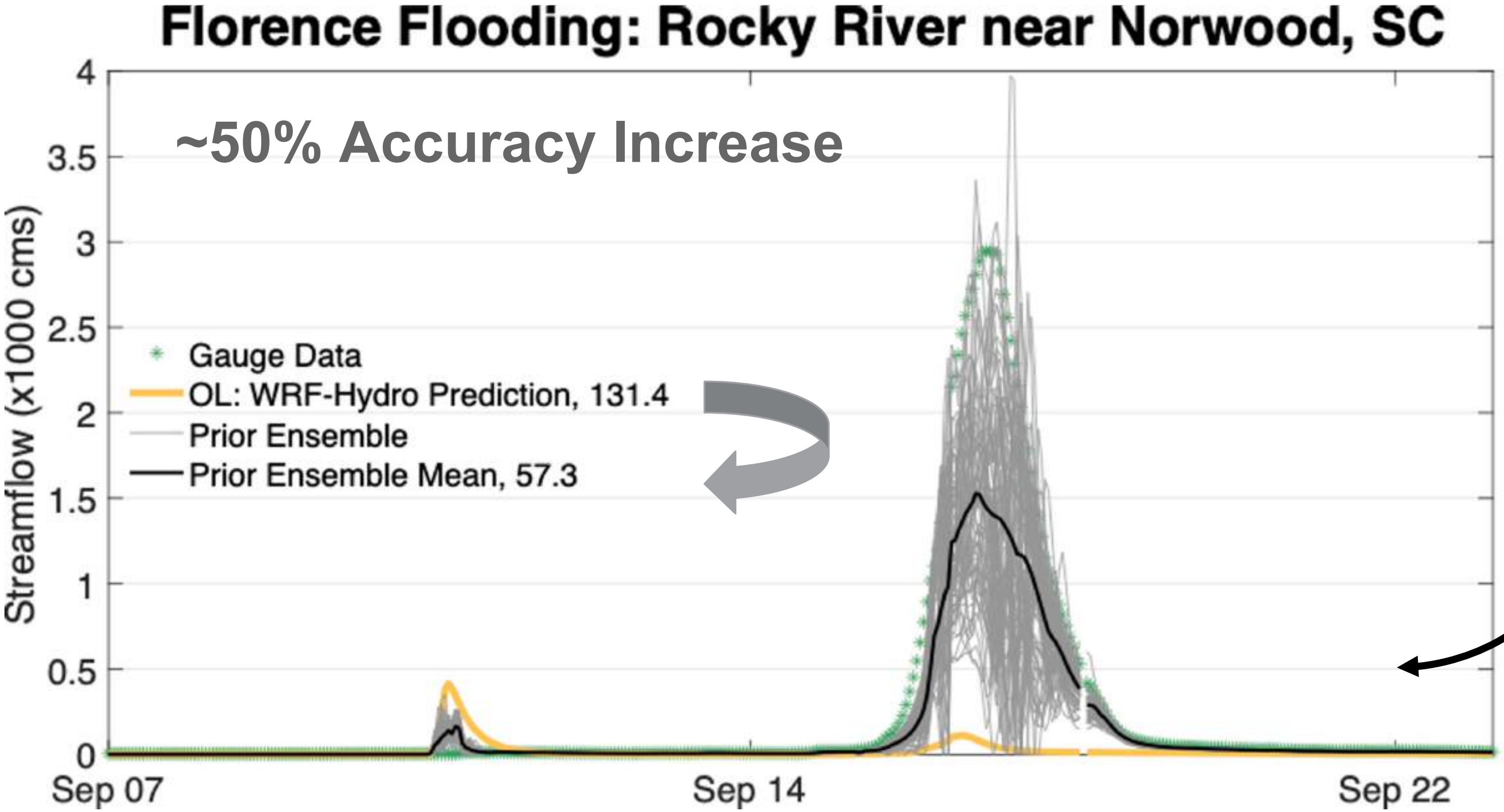


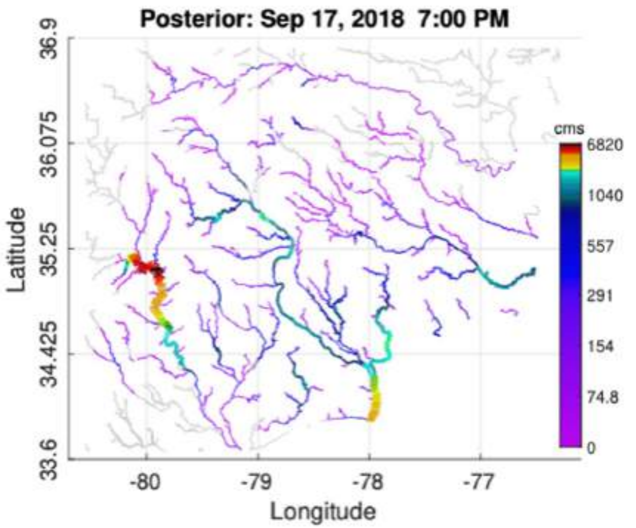


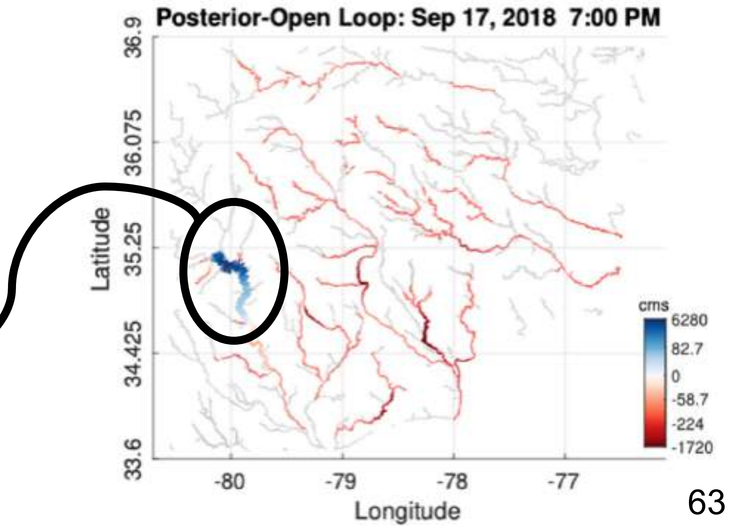
3.4.3 Inflation: Bias Correction Tool



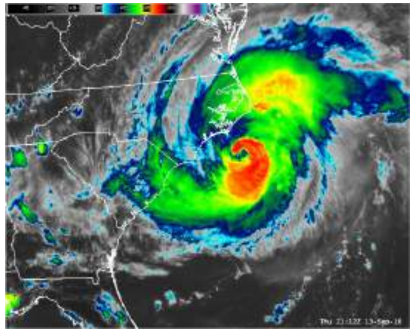


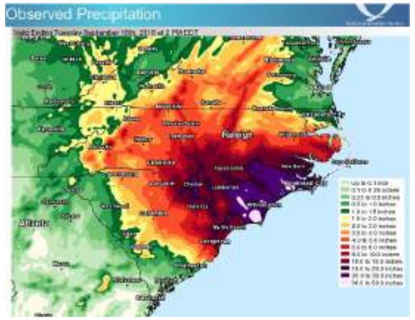


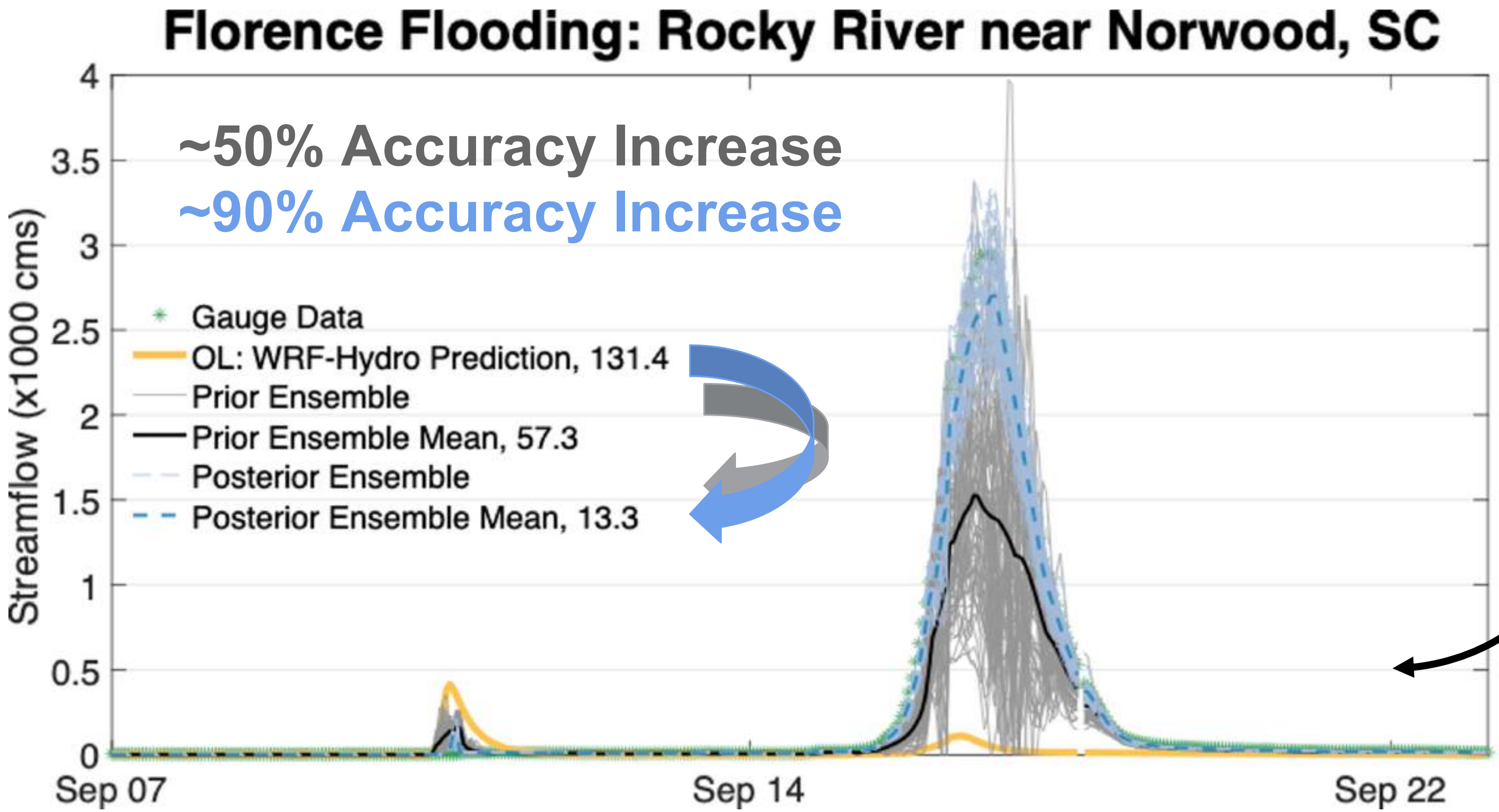


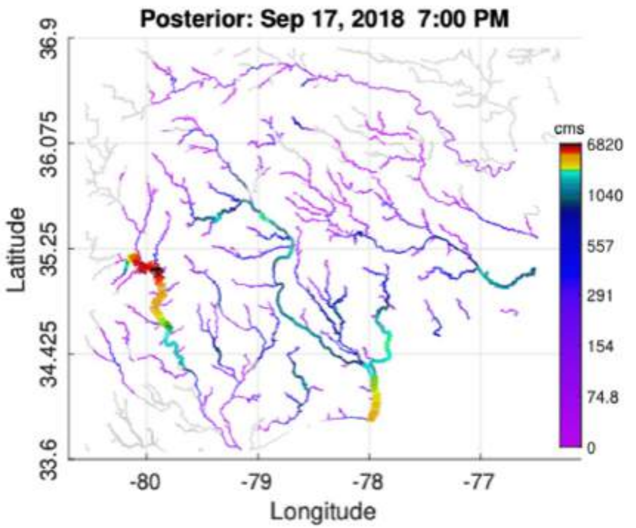


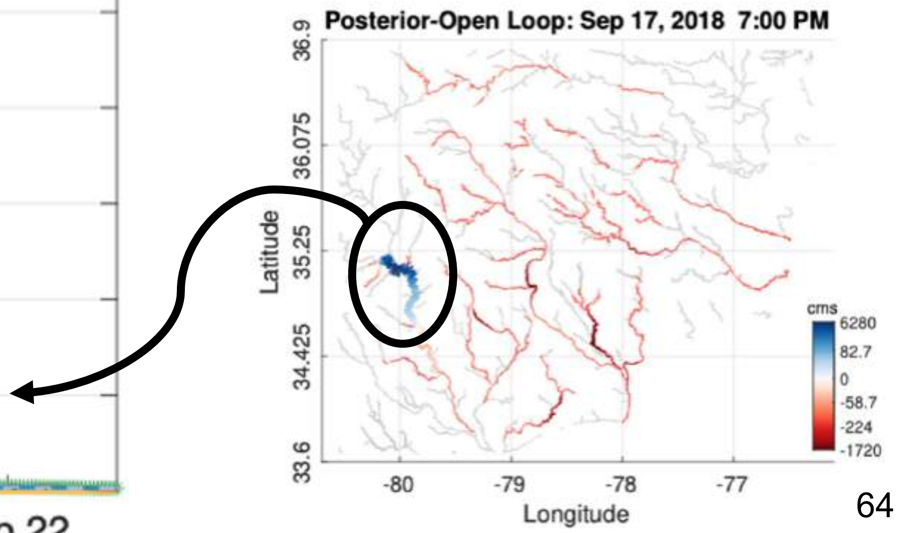
3.4.3 Inflation: Bias Correction Tool



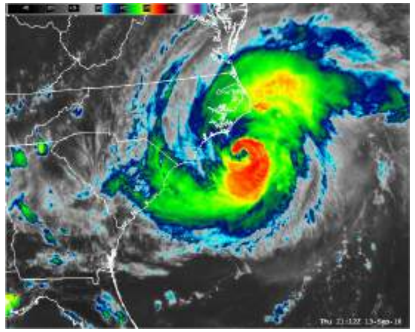


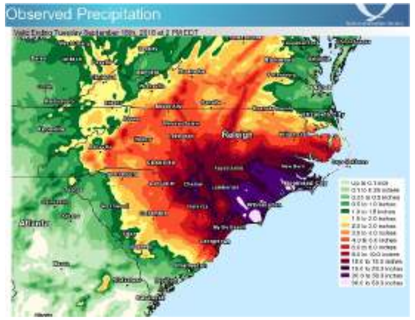


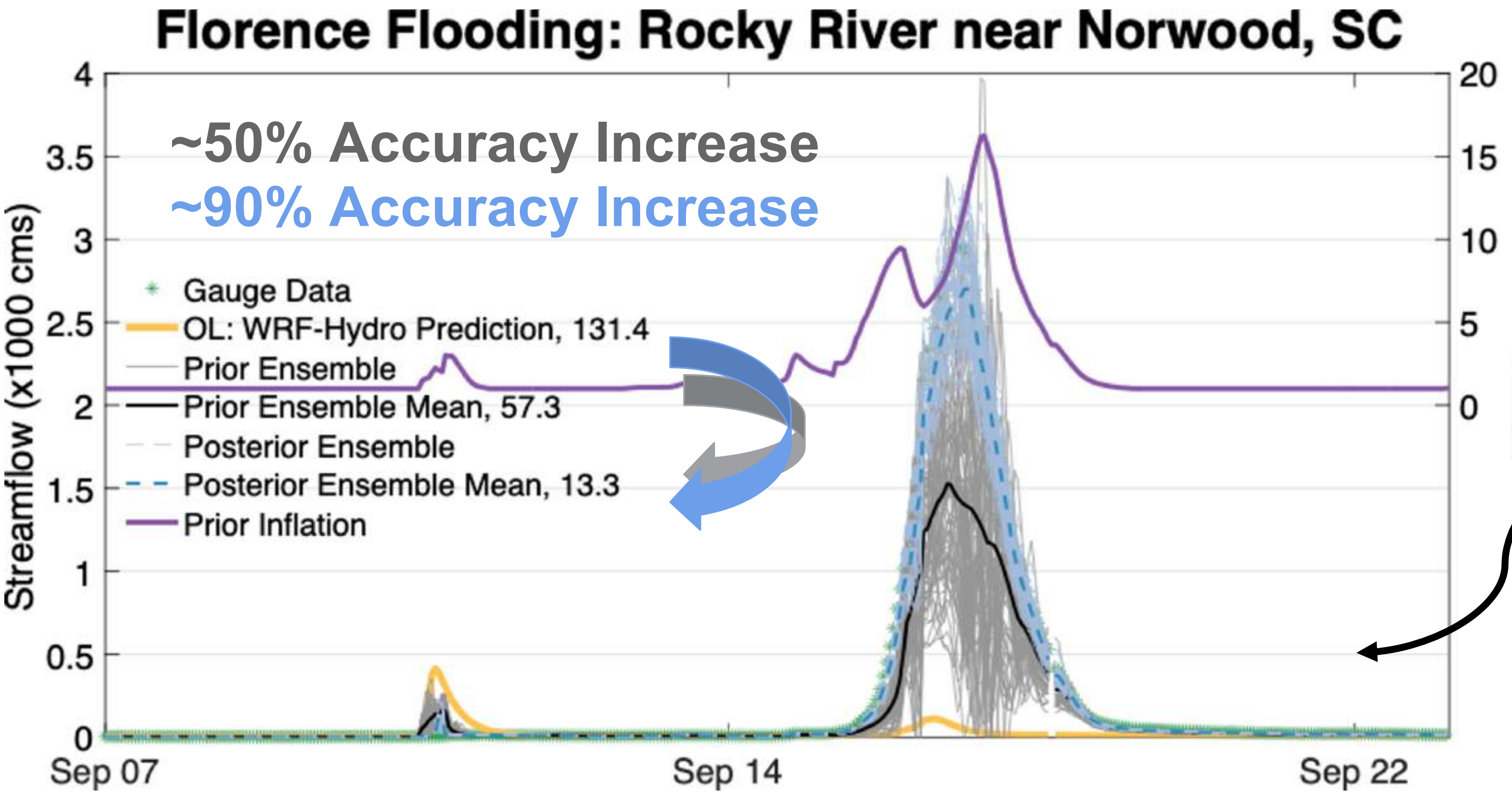


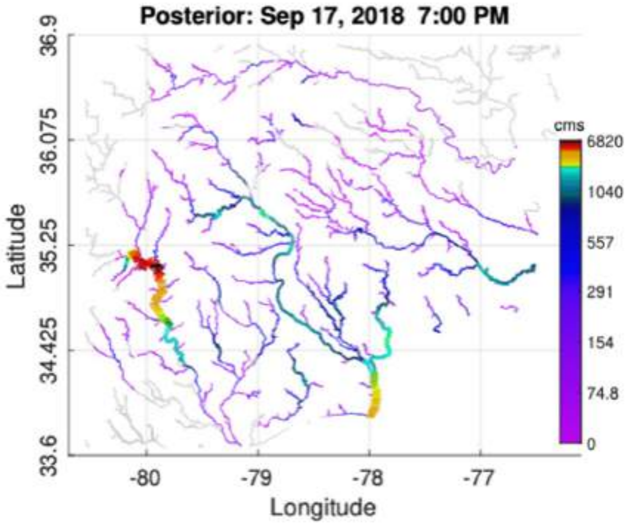


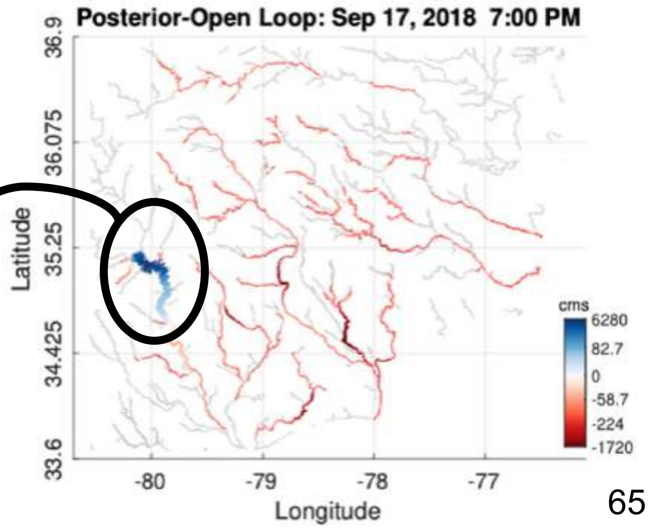
3.4.3 Inflation: Bias Correction Tool











3.4.4 Inflation: Hydro-DART Assessment

Hydrol. Earth Syst. Sci., 25, 5315–5336, 2021
https://doi.org/10.5194/hess-25-5315-2021

Hydrology and Earth System Sciences
Open Access
EGU

Ensemble streamflow data assimilation using WRF-Hydro and DART: novel localization and inflation techniques applied to Hurricane Florence flooding

Mohamad El Gharamti[1], **James L. McCreight**[2], **Seong Jin Noh**[3], **Timothy J. Hoar**[1], **Arezoo RafieeiNasab**[2], and **Benjamin K. Johnson**[1]

[1]NCAR, Computational and Information Systems Laboratory (CISL), Boulder CO, USA
[2]NCAR, Research Application Laboratory (RAL), Boulder CO, USA
[3]Civil Engineering, Kumoh National Institute of Technology, Gumi, South Korea

Correspondence: Mohamad El Gharamti (gharamti@ucar.edu)

Received: 4 December 2020 – Discussion started: 25 March 2021
Revised: 22 July 2021 – Accepted: 27 August 2021 – Published: 29 September 2021

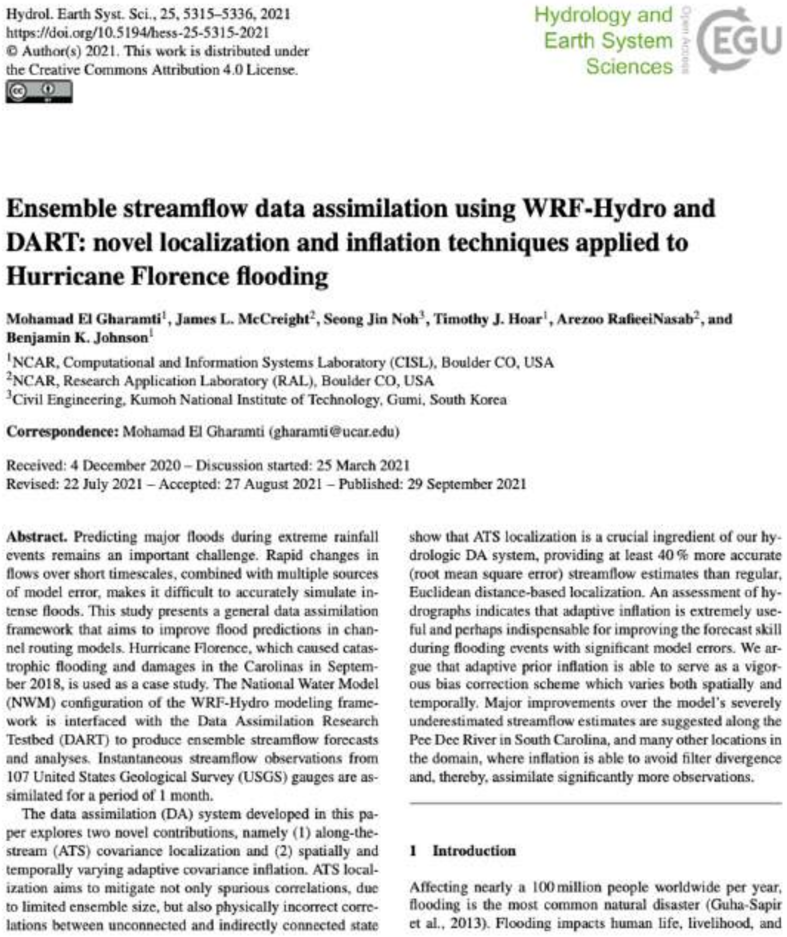
Abstract. Predicting major floods during extreme rainfall events remains an important challenge. Rapid changes in flows over short timescales, combined with multiple sources of model error, makes it difficult to accurately simulate intense floods. This study presents a general data assimilation framework that aims to improve flood predictions in channel routing models. Hurricane Florence, which caused catastrophic flooding and damages in the Carolinas in September 2018, is used as a case study. The National Water Model (NWM) configuration of the WRF-Hydro modeling framework is interfaced with the Data Assimilation Research Testbed (DART) to produce ensemble streamflow forecasts and analyses. Instantaneous streamflow observations from 107 United States Geological Survey (USGS) gauges are assimilated for a period of 1 month.

The data assimilation (DA) system developed in this paper explores two novel contributions, namely (1) along-the-stream (ATS) covariance localization and (2) spatially and temporally varying adaptive covariance inflation. ATS localization aims to mitigate not only spurious correlations, due to limited ensemble size, but also physically incorrect correlations between unconnected and indirectly connected state show that ATS localization is a crucial ingredient of our hydrologic DA system, providing at least 40 % more accurate (root mean square error) streamflow estimates than regular, Euclidean distance-based localization. An assessment of hydrographs indicates that adaptive inflation is extremely useful and perhaps indispensable for improving the forecast skill during flooding events with significant model errors. We argue that adaptive prior inflation is able to serve as a vigorous bias correction scheme which varies both spatially and temporally. Major improvements over the model's severely underestimated streamflow estimates are suggested along the Pee Dee River in South Carolina, and many other locations in the domain, where inflation is able to avoid filter divergence and, thereby, assimilate significantly more observations.

1 Introduction

Affecting nearly a 100 million people worldwide per year, flooding is the most common natural disaster (Guha-Sapir et al., 2013). Flooding impacts human life, livelihood, and

3.4.4 Inflation: Hydro-DART Assessment



Hydrol. Earth Syst. Sci., 25, 5315–5336, 2021
https://doi.org/10.5194/hess-25-5315-2021

Hydrology and Earth System Sciences
Open Access
EGU

Ensemble streamflow data assimilation using WRF-Hydro and DART: novel localization and inflation techniques applied to Hurricane Florence flooding

Mohamad El Gharamti[1], James L. McCreight[2], Seong Jin Noh[3], Timothy J. Hoar[1], Arezoo RafieeiNasab[2], and Benjamin K. Johnson[1]

[1]NCAR, Computational and Information Systems Laboratory (CISL), Boulder CO, USA
[2]NCAR, Research Application Laboratory (RAL), Boulder CO, USA
[3]Civil Engineering, Kumoh National Institute of Technology, Gumi, South Korea

Correspondence: Mohamad El Gharamti (gharamti@ucar.edu)

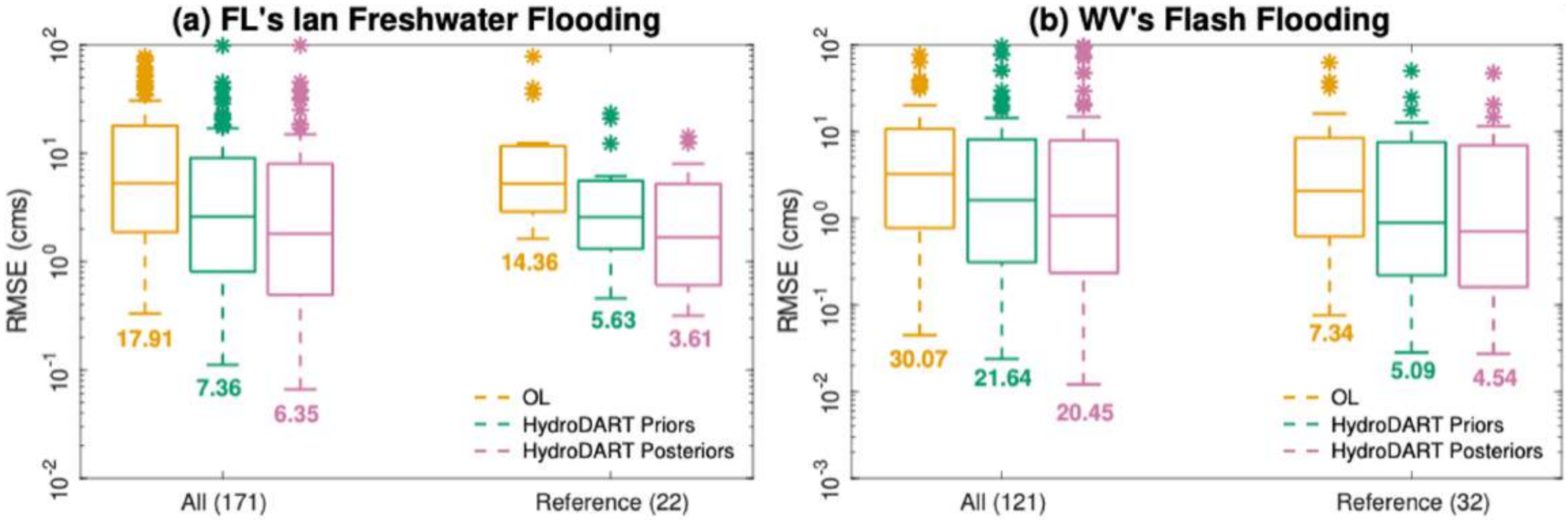
Received: 4 December 2020 – Discussion started: 25 March 2021
Revised: 22 July 2021 – Accepted: 27 August 2021 – Published: 29 September 2021

Abstract. Predicting major floods during extreme rainfall events remains an important challenge. Rapid changes in flows over short timescales, combined with multiple sources of model error, makes it difficult to accurately simulate intense floods. This study presents a general data assimilation framework that aims to improve flood predictions in channel routing models. Hurricane Florence, which caused catastrophic flooding and damages in the Carolinas in September 2018, is used as a case study. The National Water Model (NWM) configuration of the WRF-Hydro modeling framework is interfaced with the Data Assimilation Research Testbed (DART) to produce ensemble streamflow forecasts and analyses. Instantaneous streamflow observations from 107 United States Geological Survey (USGS) gauges are assimilated for a period of 1 month.

The data assimilation (DA) system developed in this paper explores two novel contributions, namely (1) along-the-stream (ATS) covariance localization and (2) spatially and temporally varying adaptive covariance inflation. ATS localization aims to mitigate not only spurious correlations, due to limited ensemble size, but also physically incorrect correlations between unconnected and indirectly connected state show that ATS localization is a crucial ingredient of our hydrologic DA system, providing at least 40 % more accurate (root mean square error) streamflow estimates than regular, Euclidean distance-based localization. An assessment of hydrographs indicates that adaptive inflation is extremely useful and perhaps indispensable for improving the forecast skill during flooding events with significant model errors. We argue that adaptive prior inflation is able to serve as a vigorous bias correction scheme which varies both spatially and temporally. Major improvements over the model's severely underestimated streamflow estimates are suggested along the Pee Dee River in South Carolina, and many other locations in the domain, where inflation is able to avoid filter divergence and, thereby, assimilate significantly more observations.

1 Introduction

Affecting nearly a 100 million people worldwide per year, flooding is the most common natural disaster (Guha-Sapir et al., 2013). Flooding impacts human life, livelihood, and



Hydro-DART estimates consistently outperformed the model's performance for FL's Ian flooding (>50%) and WV's flash flood (~30%)

3.5 Hybrid Ensemble-Variational Scheme

$$\alpha \left[\lambda \left(\rho \circ \widehat{\mathbf{P}}^f \right) \right] + (1 - \alpha)\mathbf{B} \approx \mathbf{P}^{\text{true}}$$

Hybrid Weighting Coefficient; [0, 1]

Hybrid EnKF-OI ⟶ Linearly combine:

- ❑ The flow-dependent ensemble covariance
- ❑ A static background covariance; often used in OI and 3/4D-VAR systems

$$\Rightarrow \begin{cases} \textbf{EnKF: } \lambda \left(\rho \circ \widehat{\mathbf{P}}^f \right) & \alpha = 1 \\ \textbf{EnOI: B} & \alpha = 0 \\ \textbf{Hybrid Form} & 0 < \alpha < 1 \end{cases}$$

- ❑ **Hybridization**: A better estimate of the true covariance!
- ❑ Can tackle **sampling errors, model biases** and **computational issues** all simultaneously

- ❑ **Adaptive scheme** for the weight [MWR: El Gharamti, 2020]

3.5 Hybrid Ensemble-Variational Scheme

$$\alpha\left[\lambda\left(\rho\circ\widehat{\mathbf{P}}^{f}\right)\right]+(1-\alpha)\mathbf{B}\approx\mathbf{P}^{\text{true}}$$

Hybrid Weighting Coefficient; [0, 1]

Hybrid EnKF-OI ⟶ Linearly combine:

- ❑ The flow-dependent ensemble covariance
- ❑ A static background covariance; often used in OI and 3/4D-VAR systems

$$\Rightarrow\begin{cases}\textbf{EnKF: }\lambda\left(\rho\circ\widehat{\mathbf{P}}^{f}\right) & \alpha=1\\ \textbf{EnOI: B} & \alpha=0\\ \textbf{Hybrid Form} & 0<\alpha<1\end{cases}$$

- ❑ **Hybridization**: A better estimate of the true covariance!
- ❑ Can tackle **sampling errors, model biases** and **computational issues** all simultaneously

- ❑ **Adaptive scheme** for the weight [MWR: El Gharamti, 2020]

Hybrid Ensemble–Variational Filter: A Spatially and Temporally Varying Adaptive Algorithm to Estimate Relative Weighting

MOHAMAD EL GHARAMTI[a]

[a] *National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 30 March 2020, in final form 8 September 2020)

ABSTRACT: Model errors and sampling errors produce inaccurate sample covariances that limit the performance of ensemble Kalman filters. Linearly hybridizing the flow-dependent ensemble-based covariance with a time-invariant background covariance matrix gives a better estimate of the true error covariance. Previous studies have shown this, both in theory and in practice. How to choose the weight for each covariance remains an open question especially in the presence of model biases. This study assumes the weighting coefficient to be a random variable and then introduces a Bayesian scheme to estimate it using the available data. The scheme takes into account the discrepancy between the ensemble mean and the observations, the ensemble variance, the static background variance, and the uncertainties in the observations. The proposed algorithm is first derived for a spatially constant weight and then this assumption is relaxed by estimating a unique scalar weight for each state variable. Using twin experiments with the 40-variable Lorenz 96 system, it is shown that the proposed scheme is able to produce quality forecasts even in the presence of severe sampling errors. The adaptive algorithm allows the hybrid filter to switch between an EnKF and a simple EnOI depending on the statistics of the ensemble. In the presence of model errors, the adaptive scheme demonstrates additional improvements compared with standard enhancements alone, such as inflation and localization. Finally, the potential of the spatially varying variant to accommodate challenging sparse observation networks is demonstrated. The computational efficiency and storage of the proposed scheme, which remain an obstacle, are discussed.

KEYWORDS: Bayesian methods; Filtering techniques; Inverse methods; Kalman filters; Ensembles; Numerical weather prediction/forecasting

1. Introduction

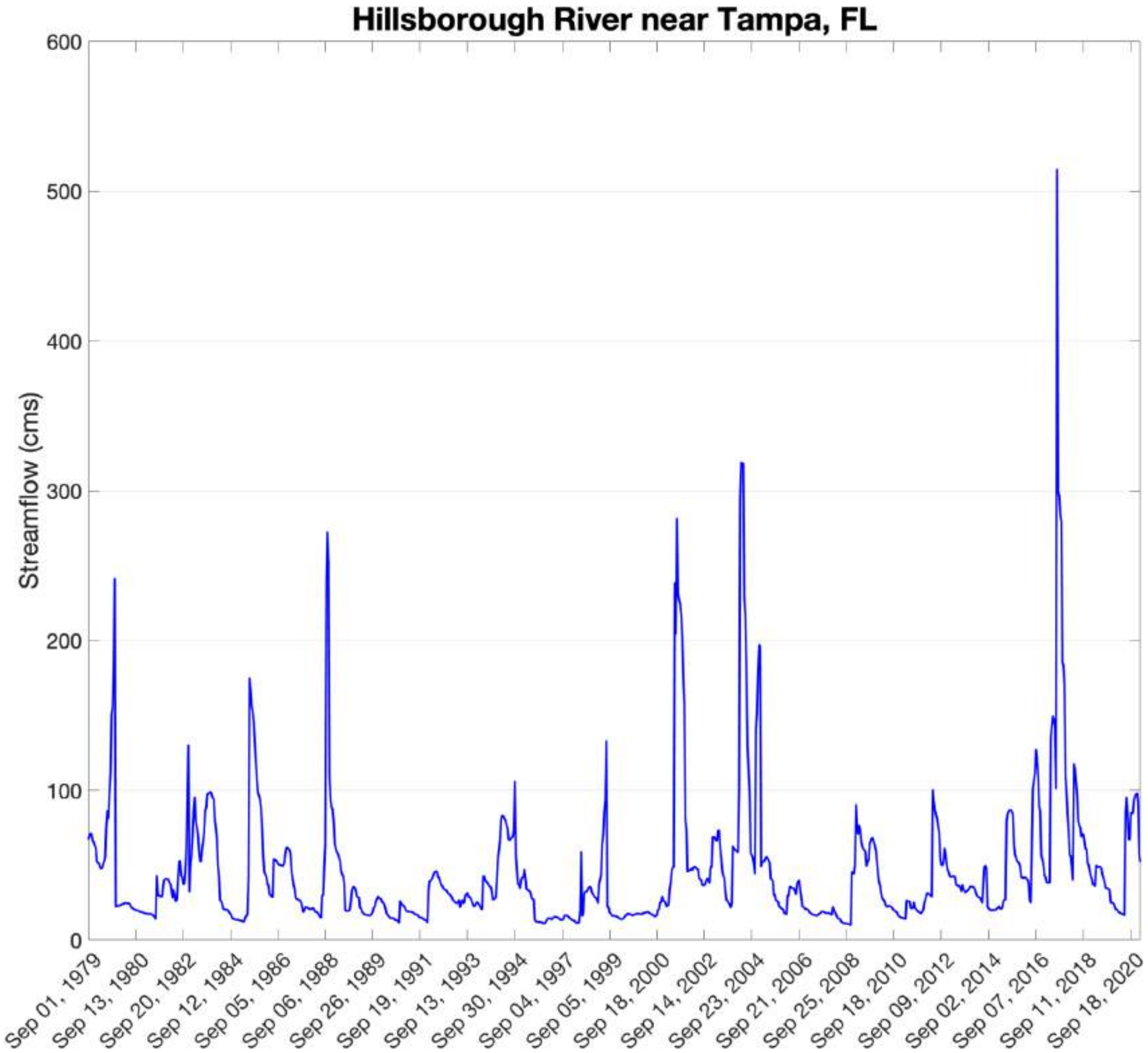
For the past ~2 decades, the pursuit of improving the performance of ensemble Kalman filters (EnKFs) has been the focus of many research studies across various Earth system applications including numerical weather prediction (NWP) and oceanography (Carrassi et al. 2018, and references therein). The EnKF (Evensen 2003) approximates the background error covariance matrix with the sample covariance from an ensemble of model states. For unbiased forecasts and

(e.g., Houtekamer and Mitchell 1998; Bishop and Hodyss 2009a,b; Anderson 2012; Lei et al. 2016) and the use of multiphysics ensembles (e.g., Skamarock et al. 2008; Meng and Zhang 2008; Berner et al. 2011). Inflation increases the spread of the ensemble around the ensemble mean without changing the rank of the covariance. The rank may change if the inflation is designed to vary in space (El Gharamti et al. 2019). Localization, on the other hand, tackles sampling error by reducing spurious correlations in the error covariance usually resulting in a full-rank matrix. The multiphysics, often referred

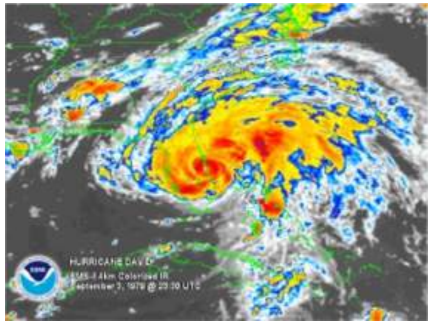
3.5.1 Hybrid Scheme: Climatology

- Here, we estimate the static background using the model's climatology
- 42-year retrospective WRF-Hydro model simulation: 1979-2020
- **B** is approximated using a large climatological ensemble (1000)

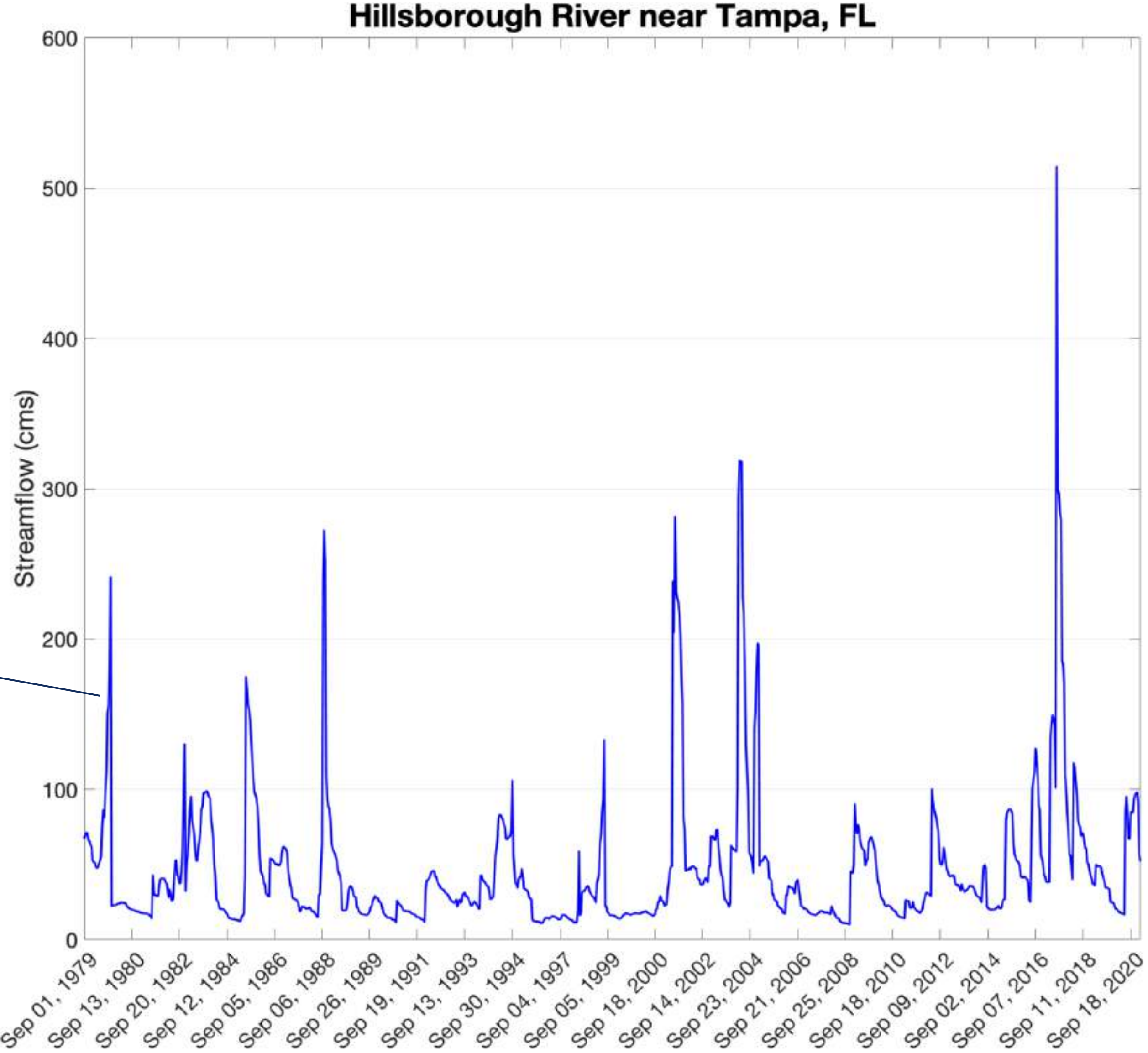
3.5.1 Hybrid Scheme: Climatology



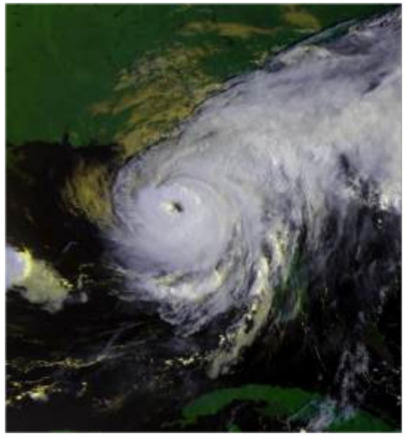
3.5.1 Hybrid Scheme: Climatology

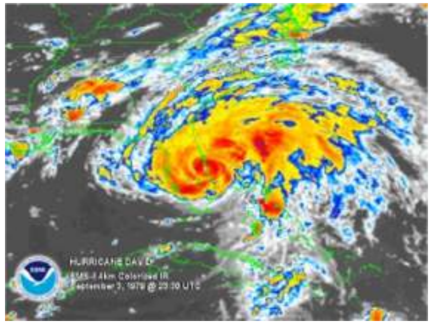


Hurricane David (1979)

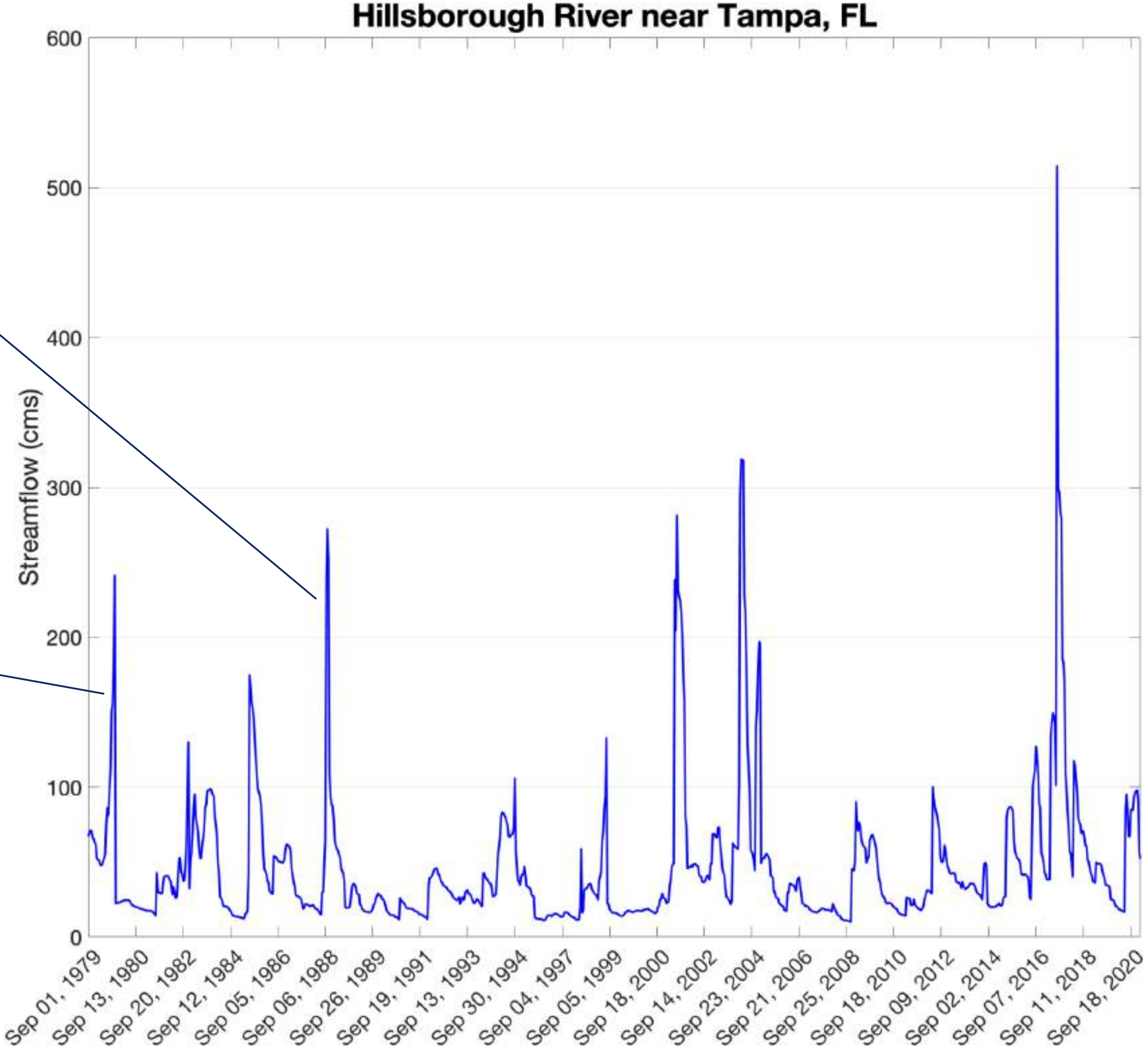


Hurricane Elena (1985)



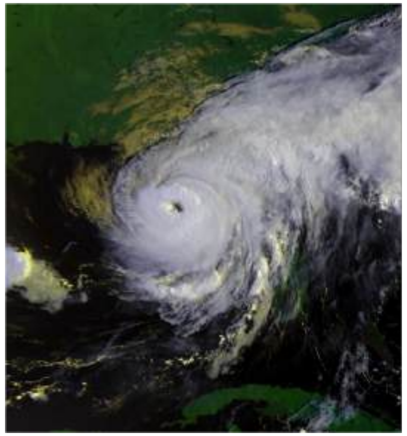


Hurricane David (1979)

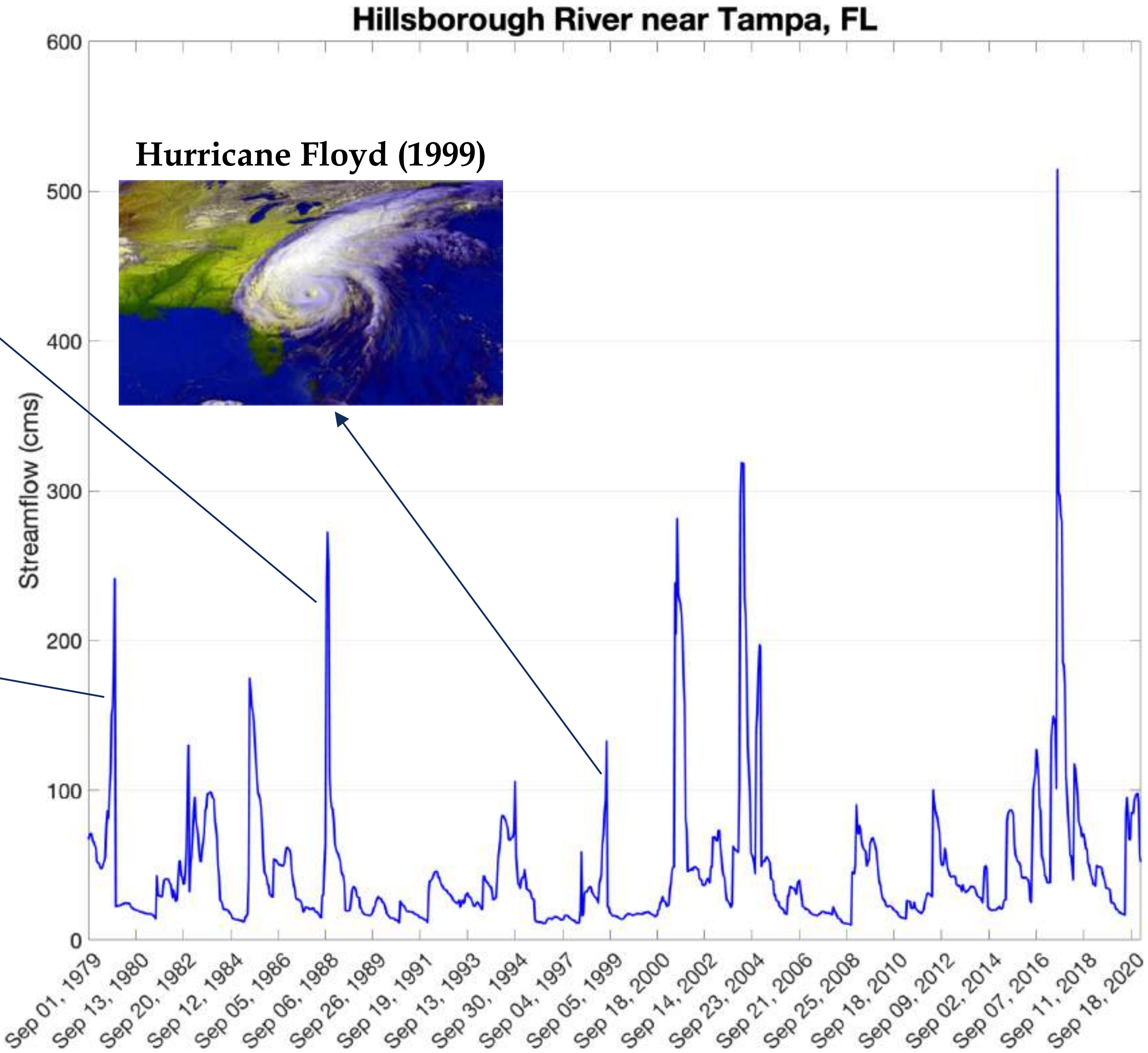


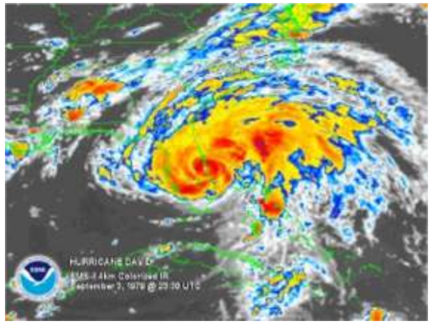
3.5.1 Hybrid Scheme: Climatology

Hurricane Elena (1985)



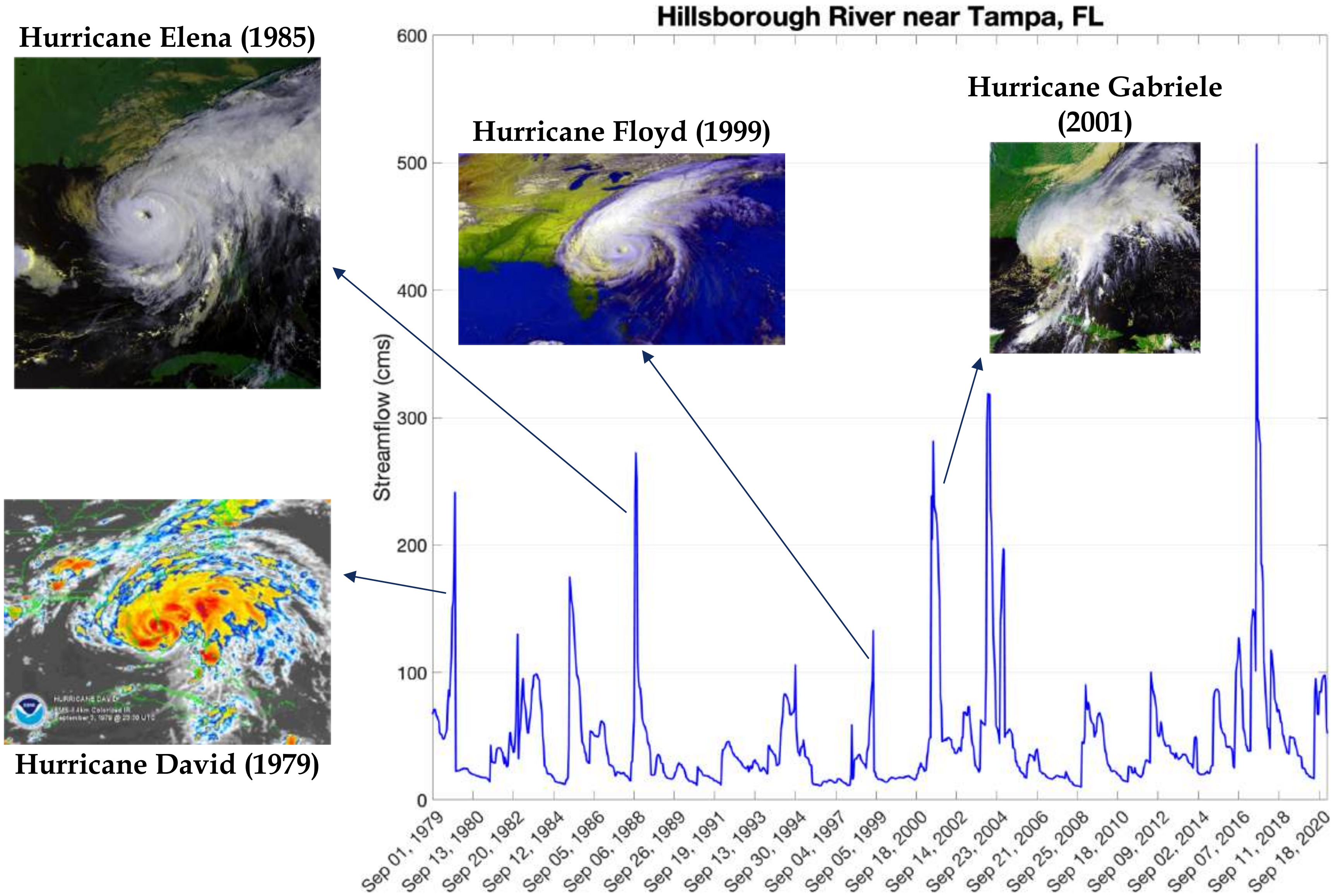
Hurricane Floyd (1999)



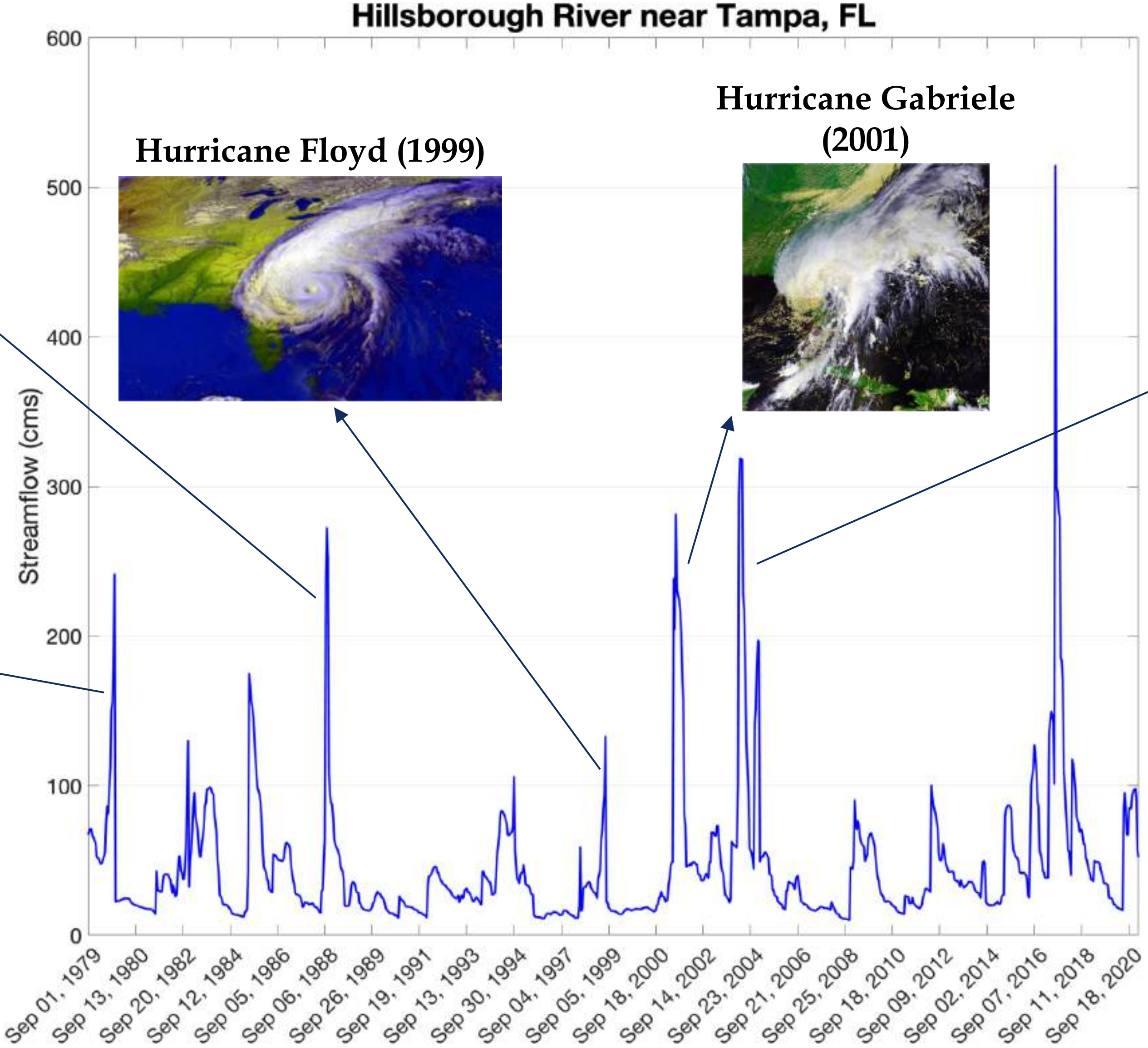


Hurricane David (1979)

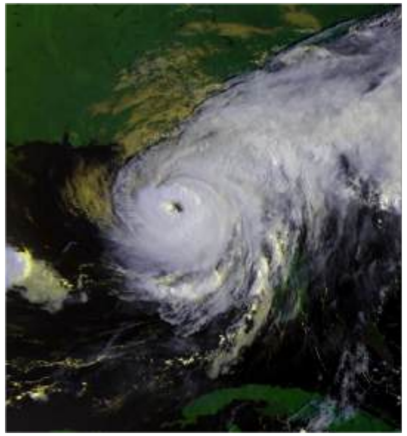
3.5.1 Hybrid Scheme: Climatology



3.5.1 Hybrid Scheme: Climatology



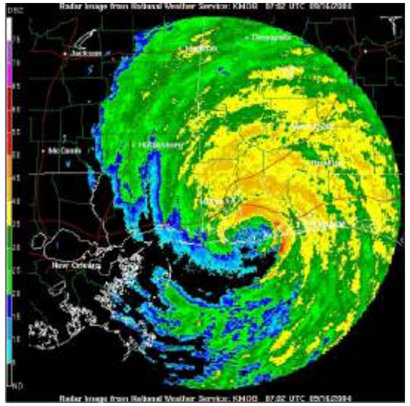
Hurricane Elena (1985)

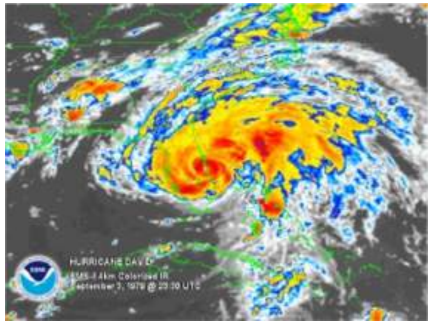


Hurricane Floyd (1999)

Hurricane Gabriele (2001)

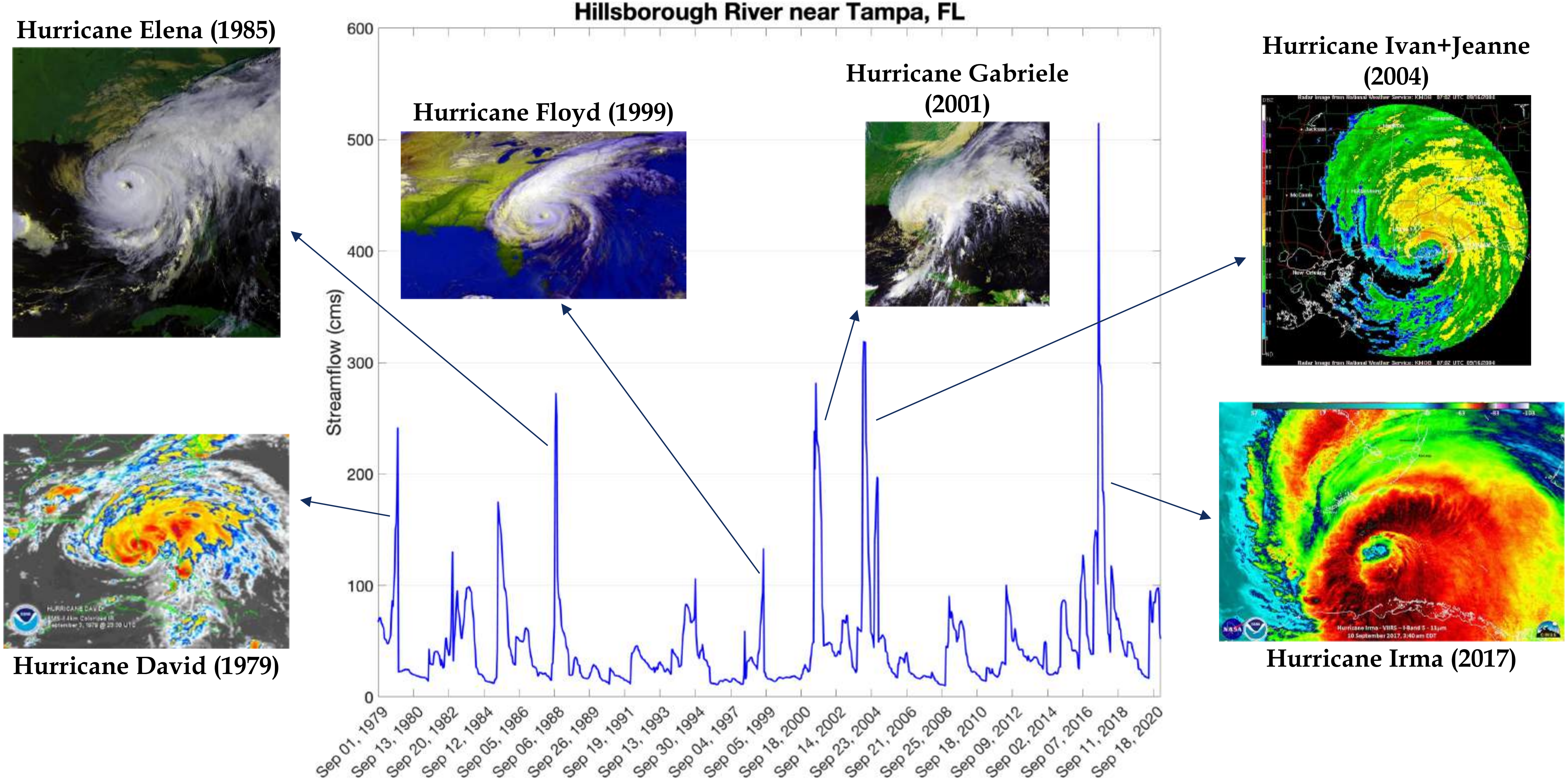
Hurricane Ivan+Jeanne (2004)





Hurricane David (1979)

3.5.1 Hybrid Scheme: Climatology



Hurricane Elena (1985)

Hurricane David (1979)

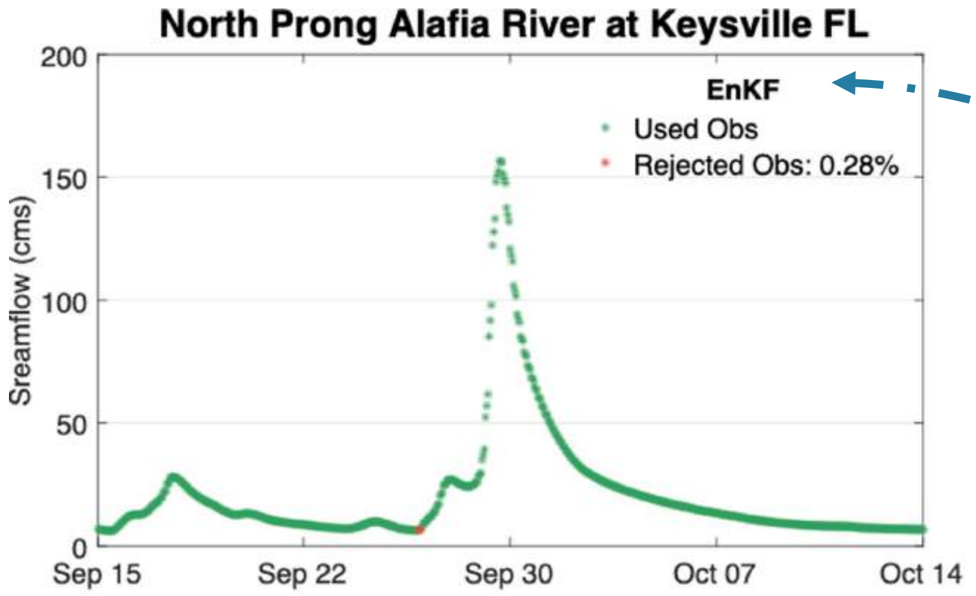
Hurricane Floyd (1999)

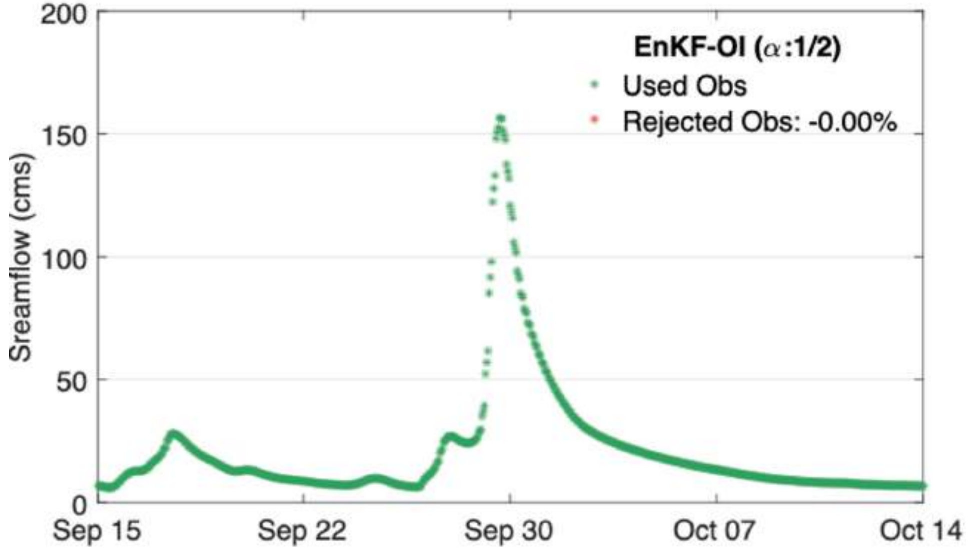
Hurricane Gabriele (2001)

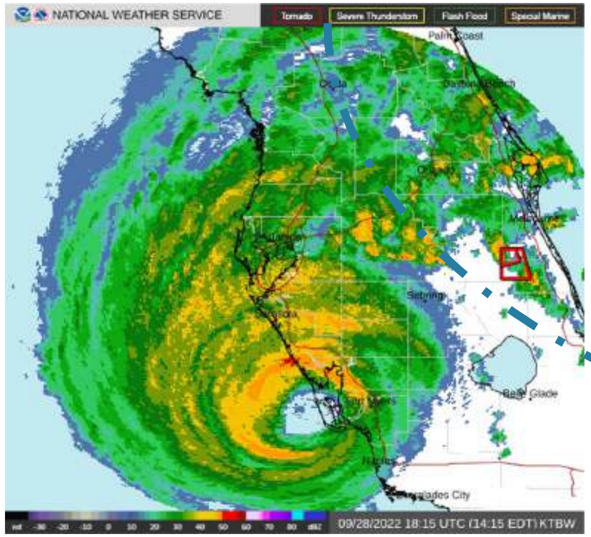
Hurricane Ivan+Jeanne (2004)

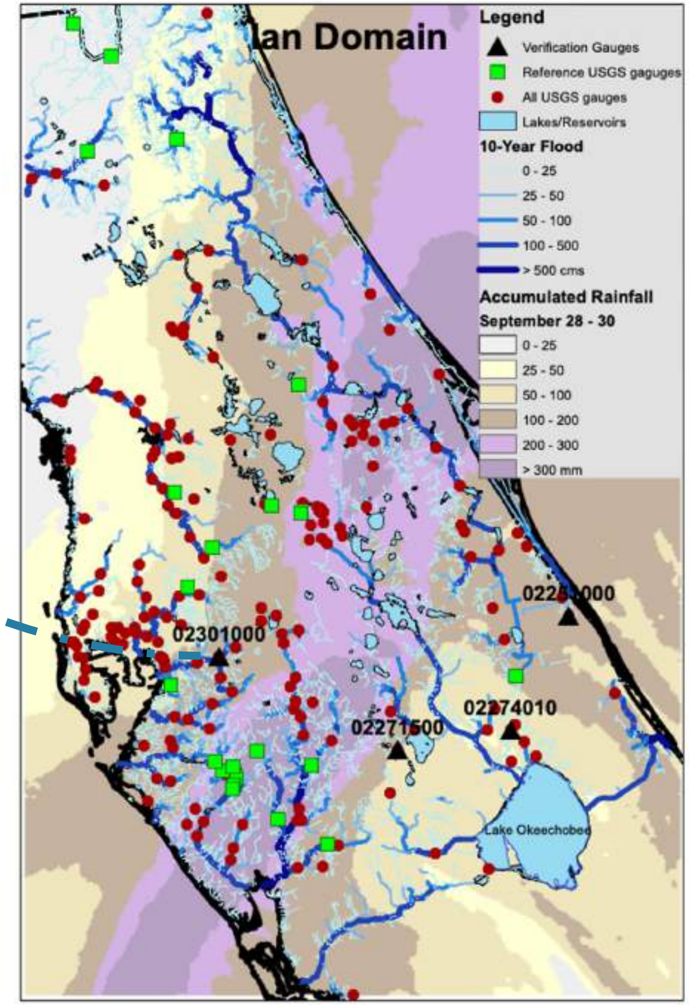
Hurricane Irma (2017)

3.5.2 Hybrid Scheme: Accuracy vs the EnKF

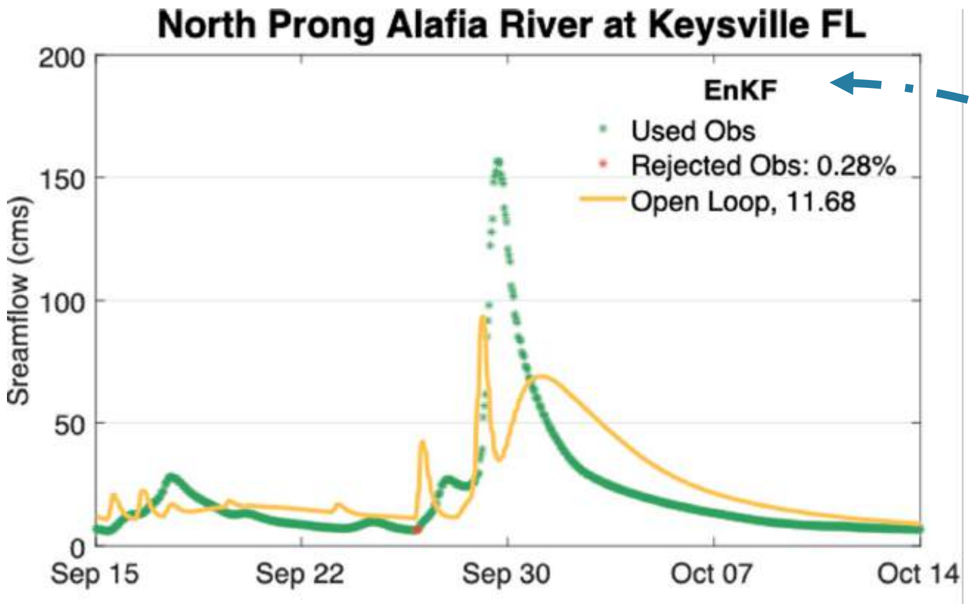


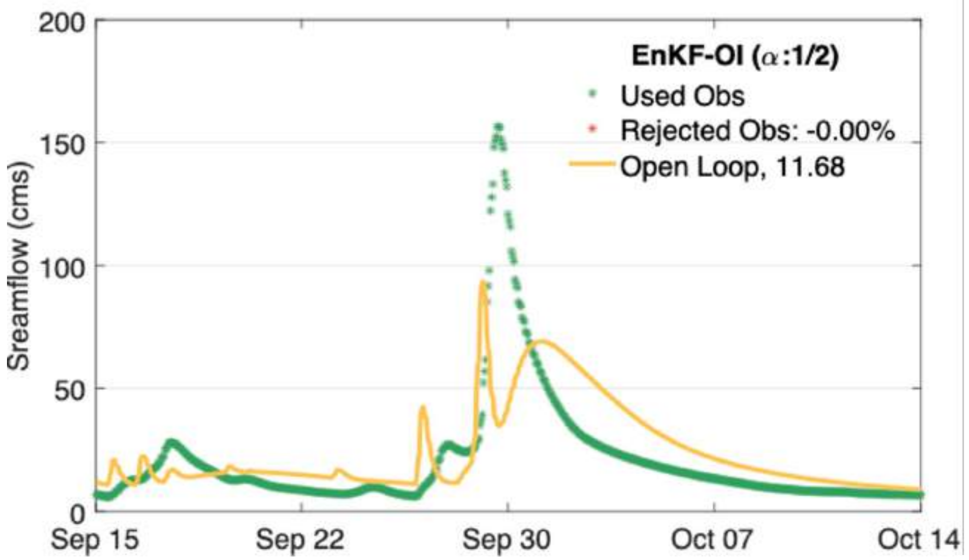


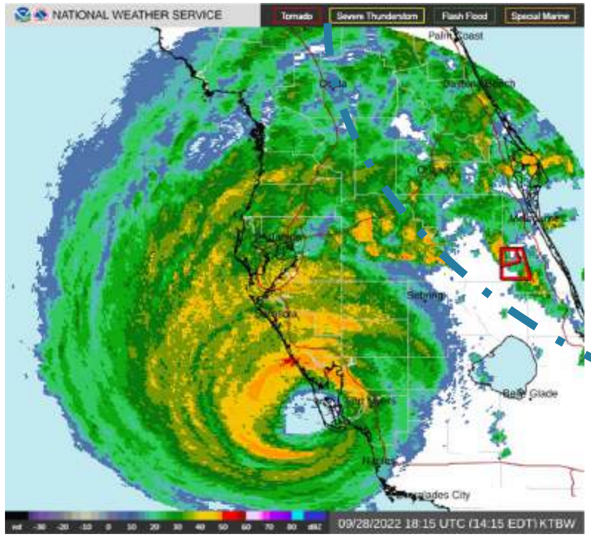


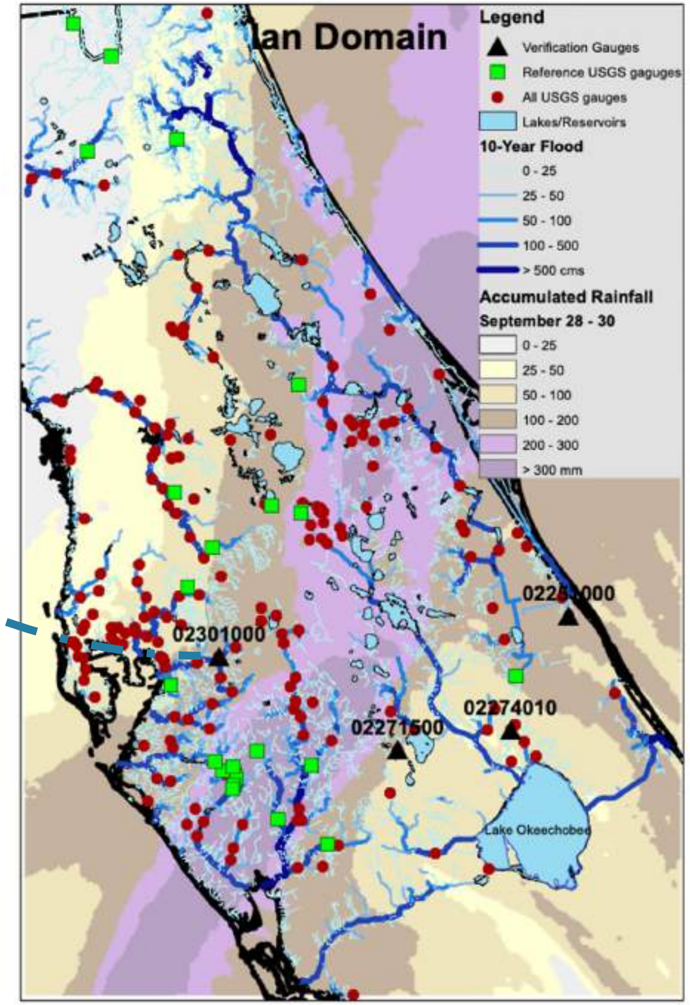


3.5.2 Hybrid Scheme: Accuracy vs the EnKF

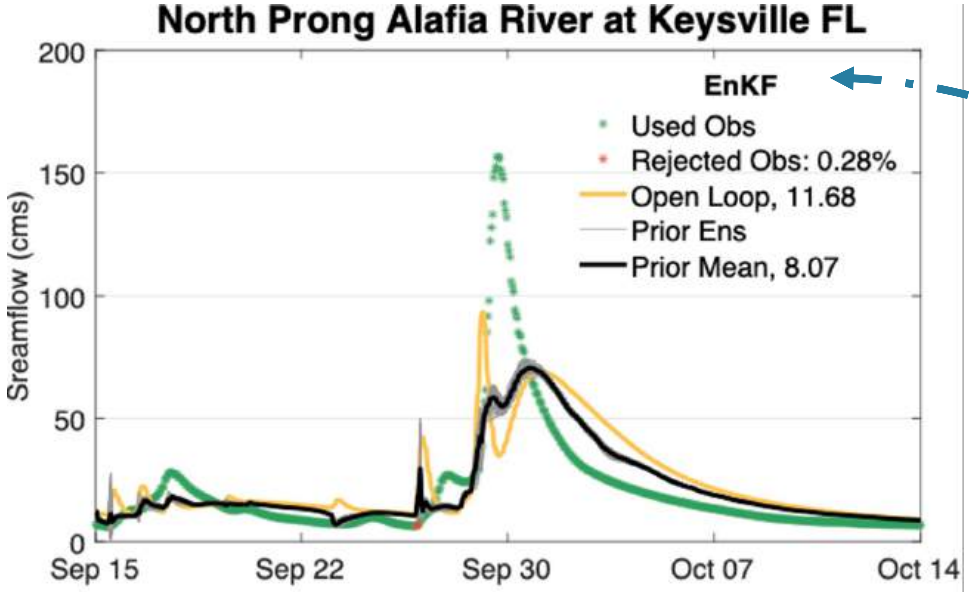


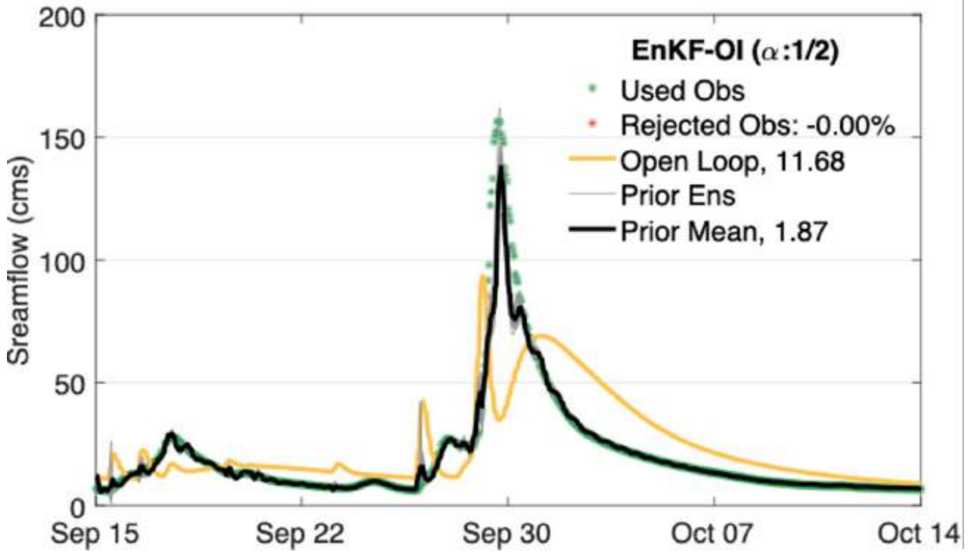


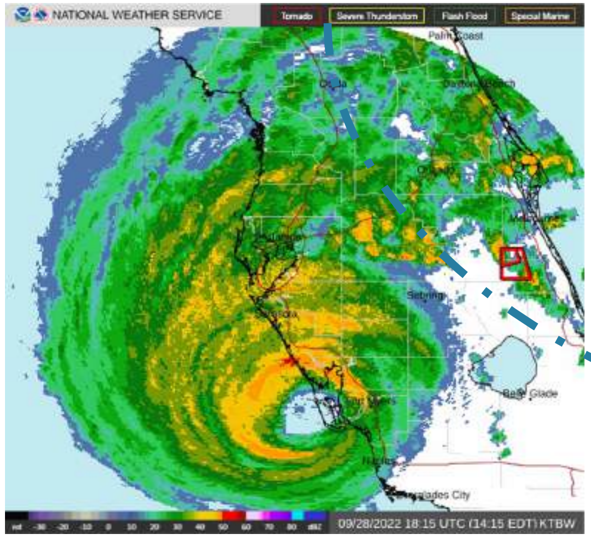


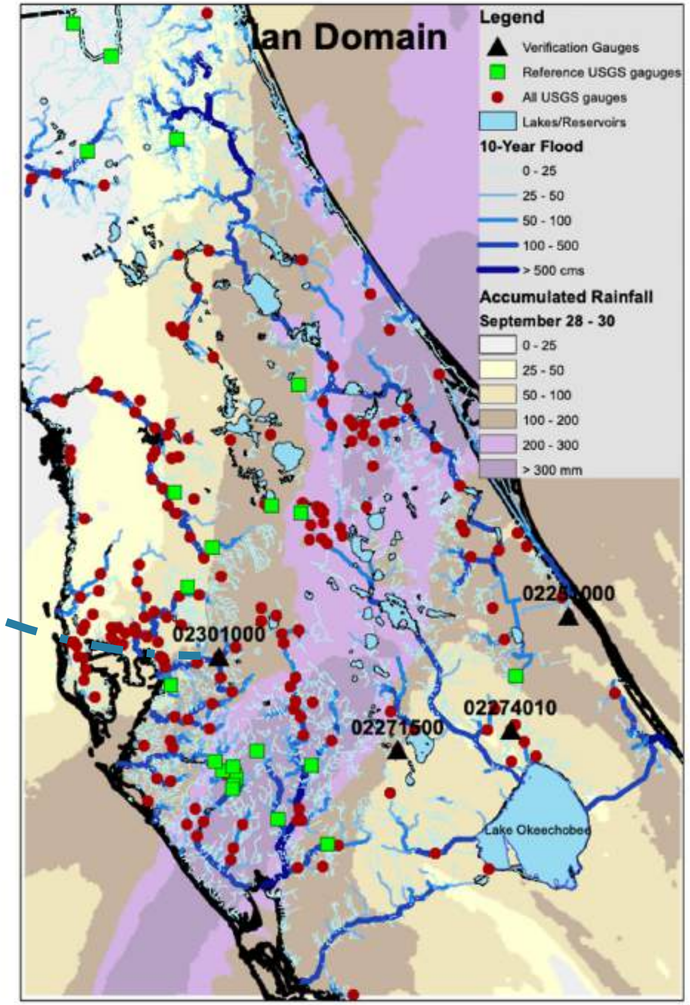


3.5.2 Hybrid Scheme: Accuracy vs the EnKF

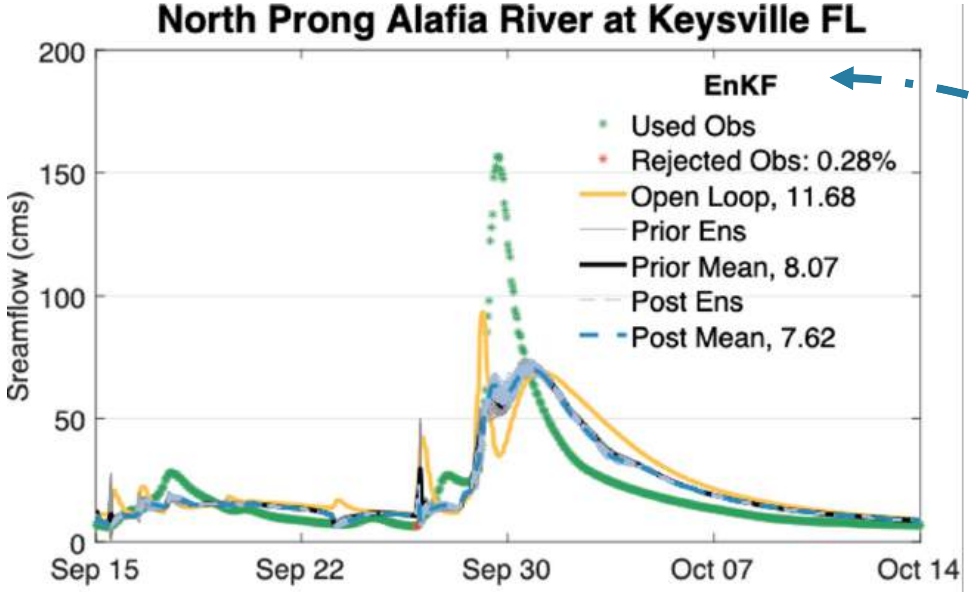


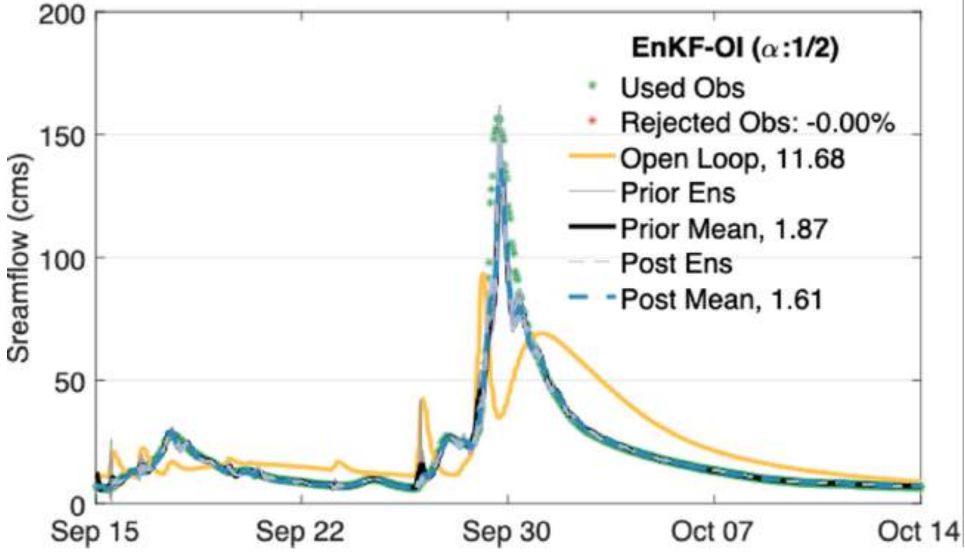


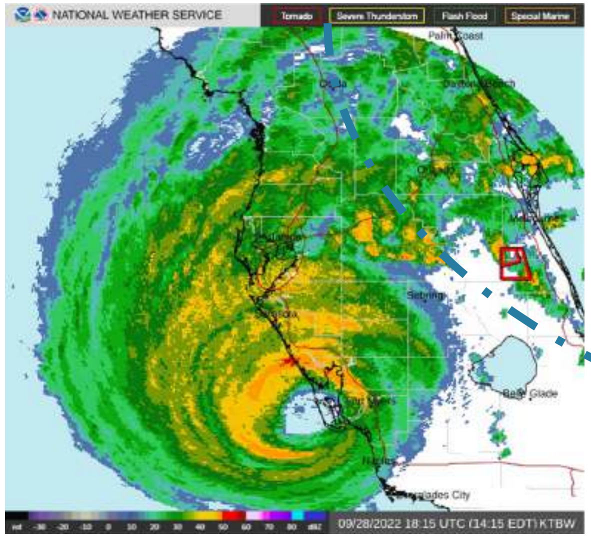


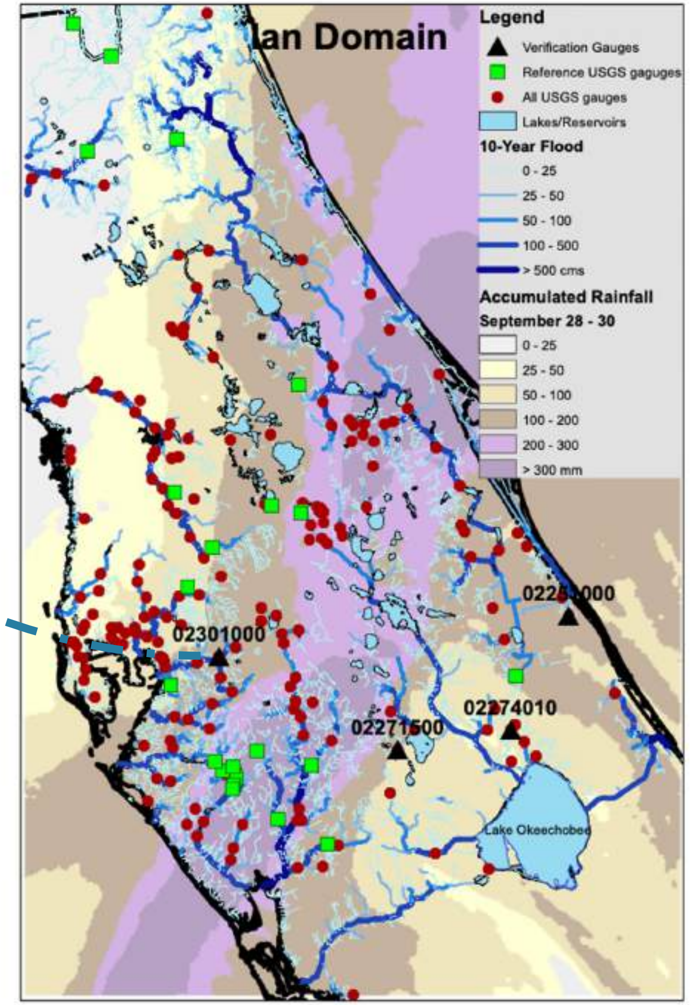


3.5.2 Hybrid Scheme: Accuracy vs the EnKF

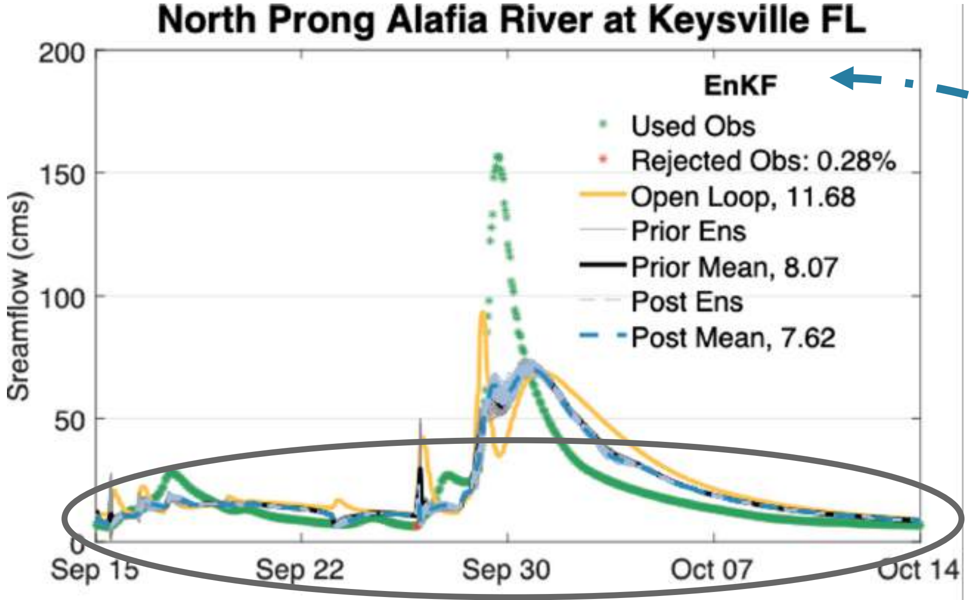


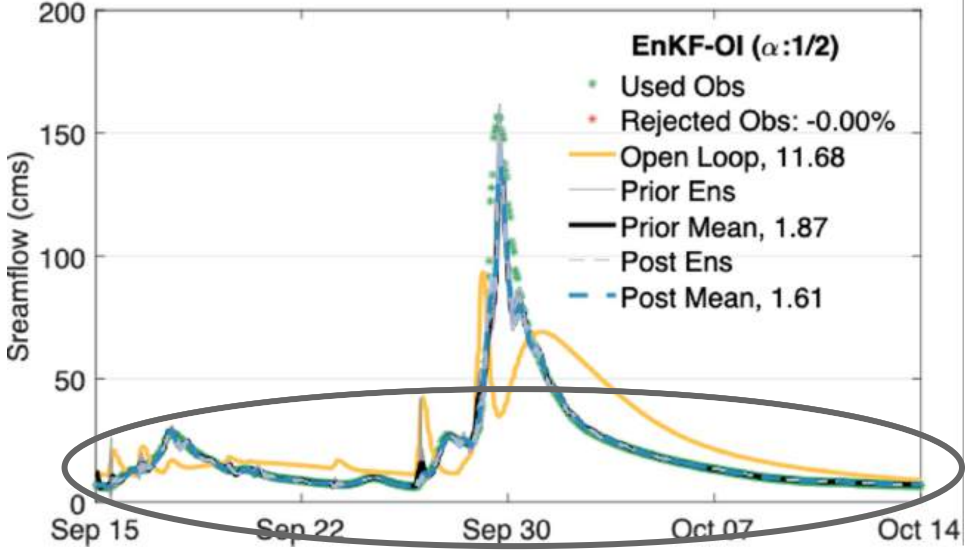


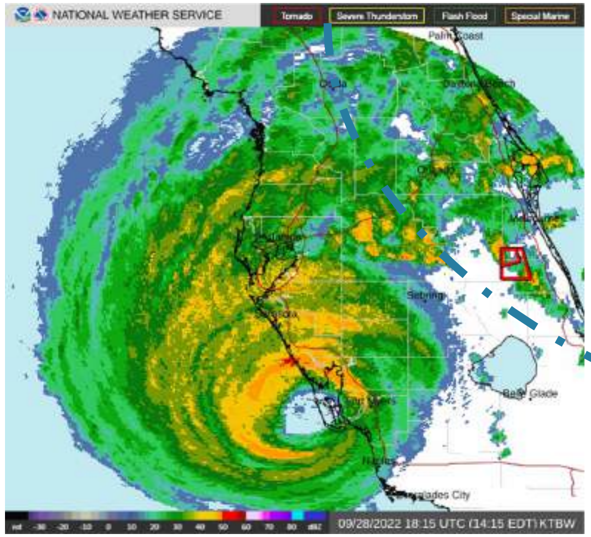


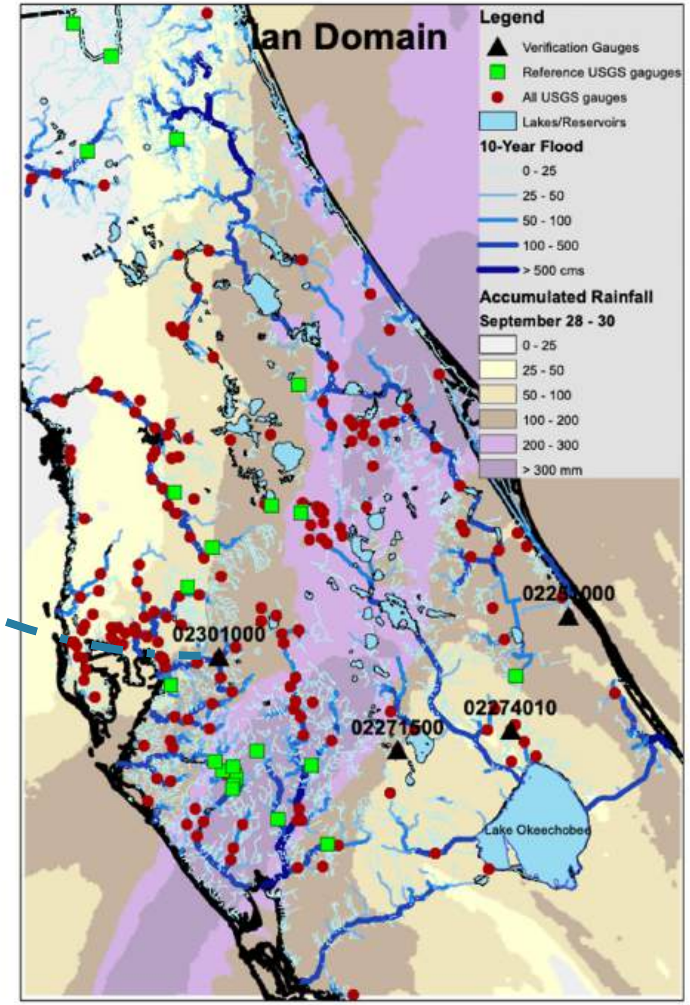


3.5.2 Hybrid Scheme: Accuracy vs the EnKF

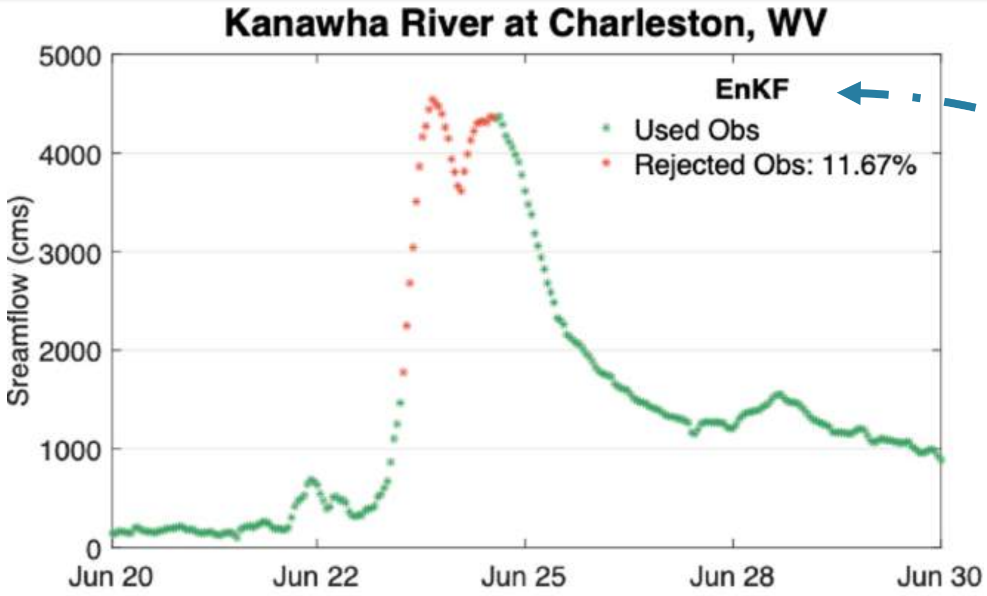


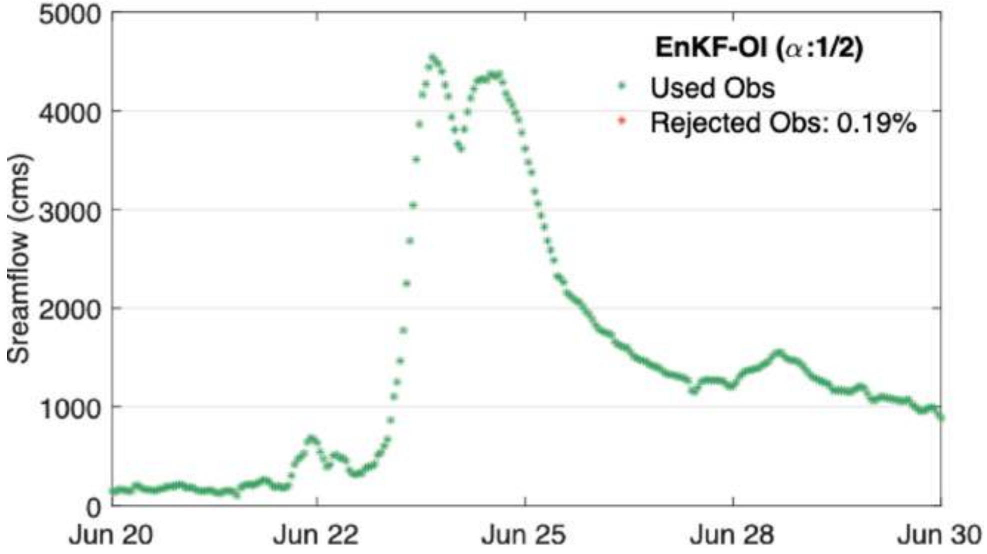


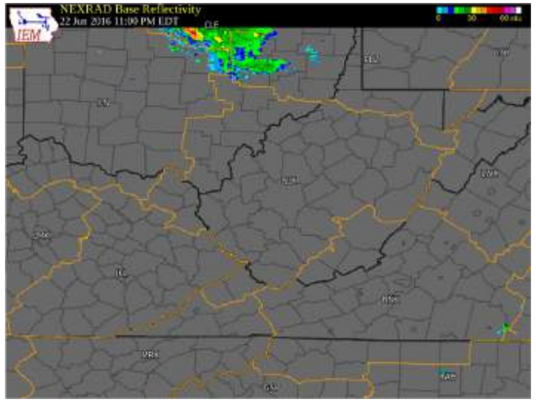


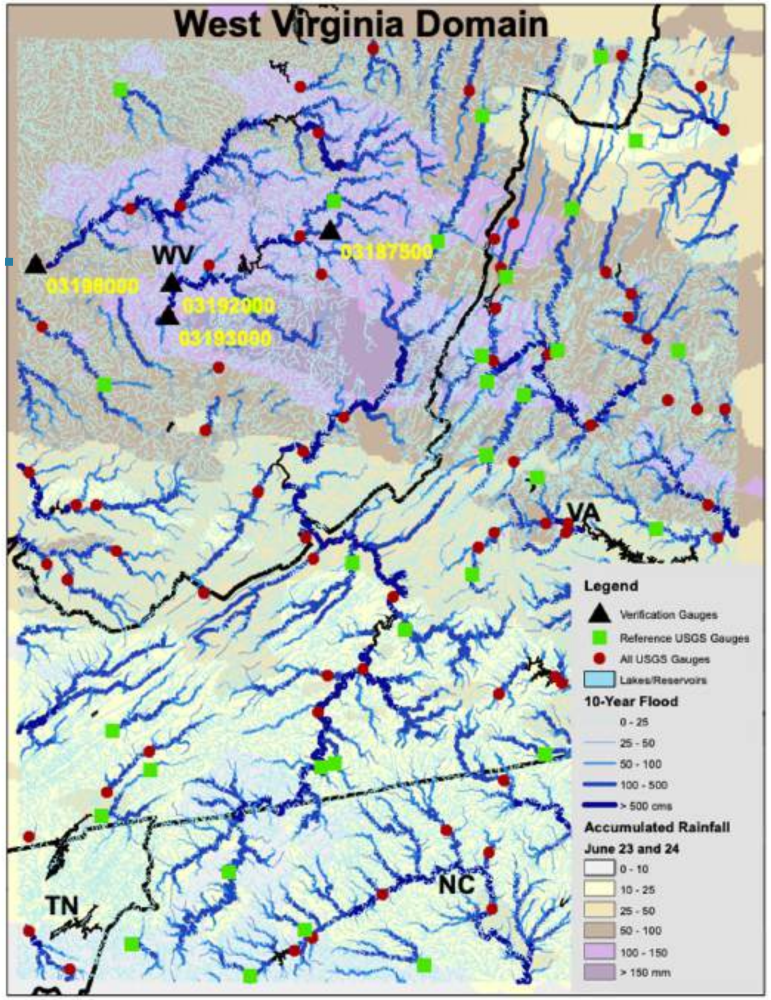


3.5.2 Hybrid Scheme: Accuracy vs the EnKF

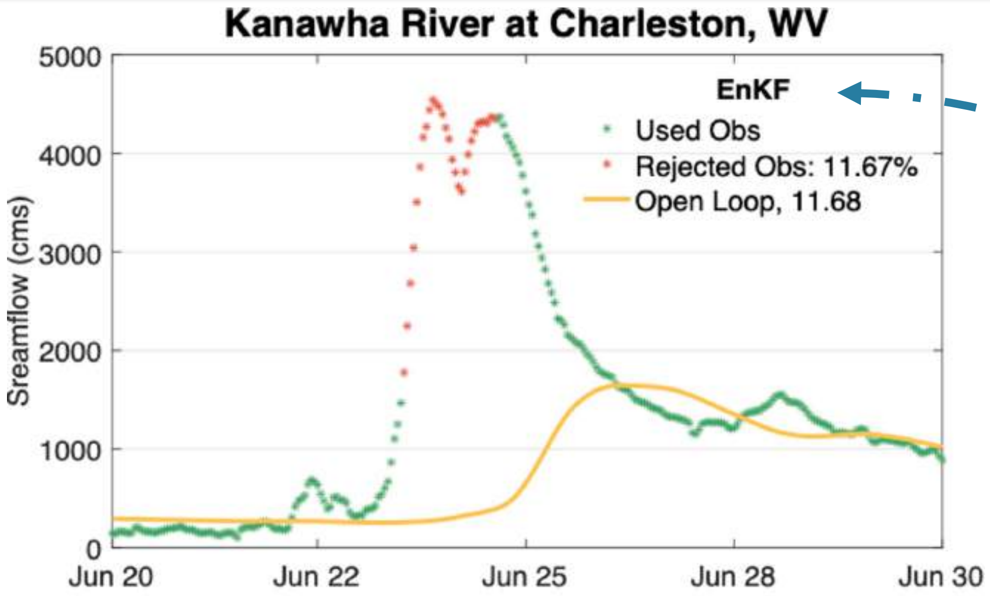


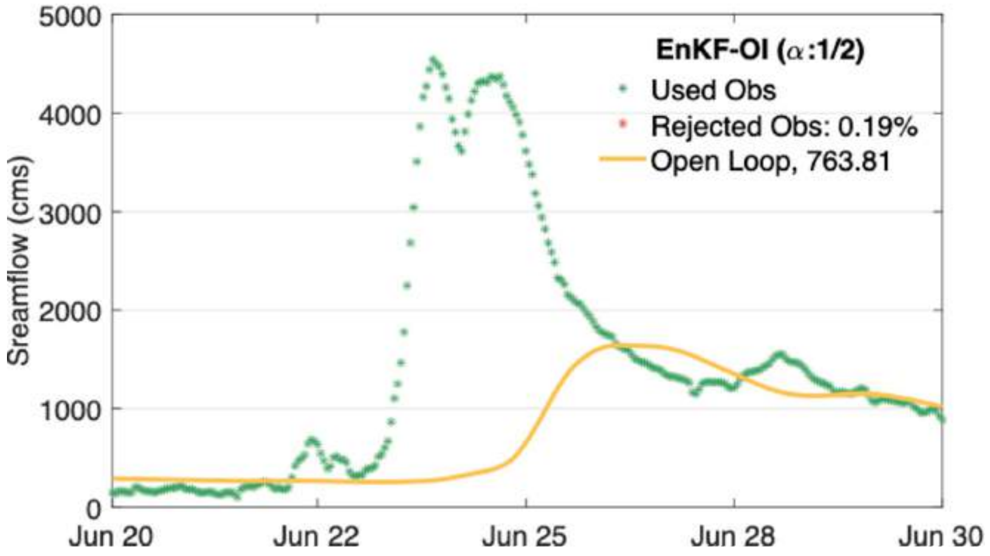


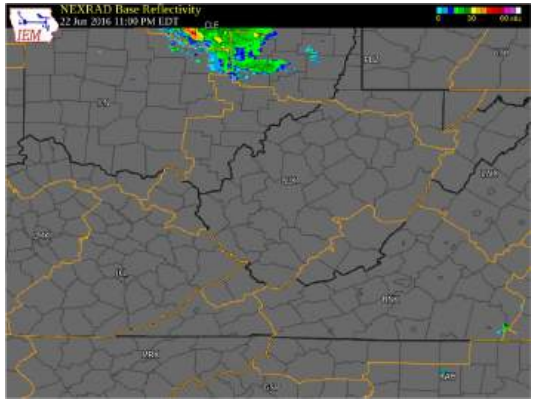


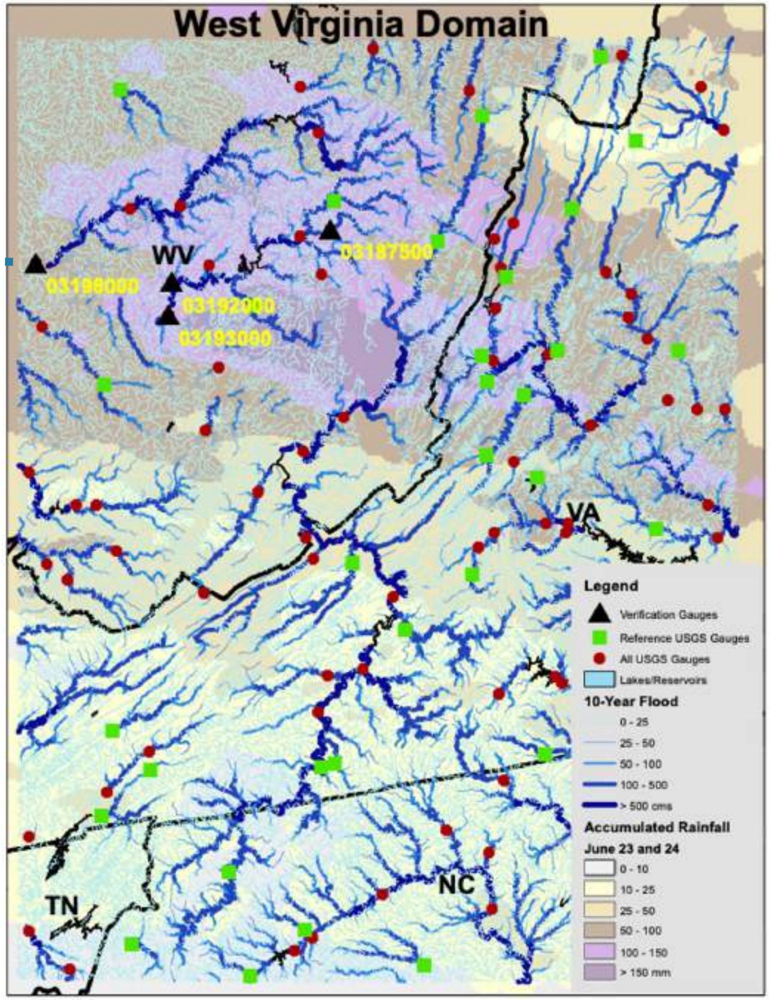


3.5.2 Hybrid Scheme: Accuracy vs the EnKF

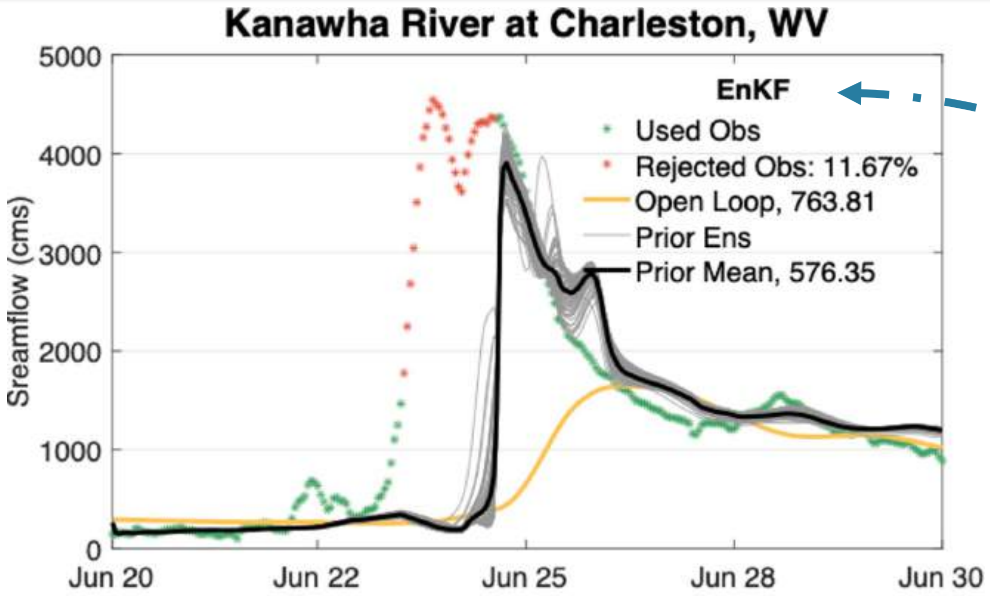


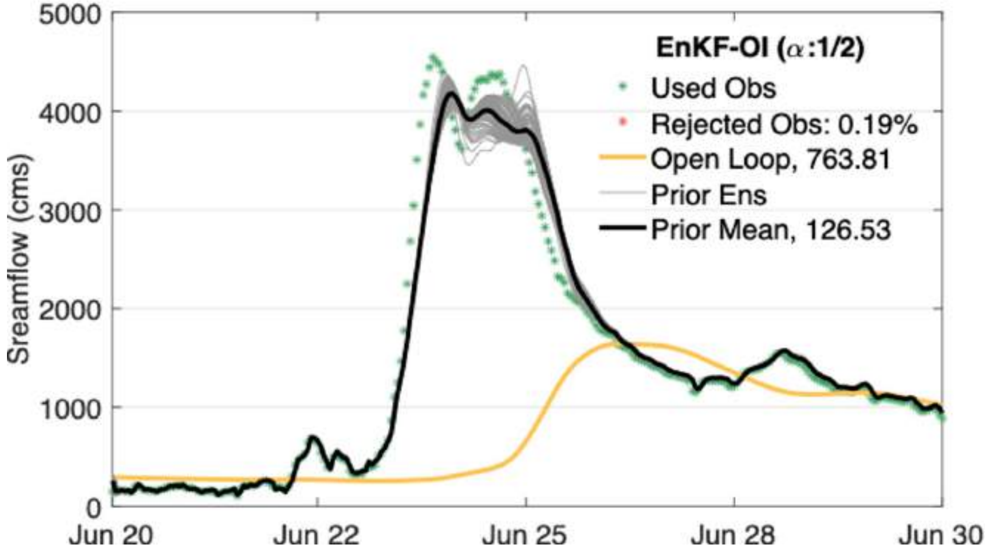


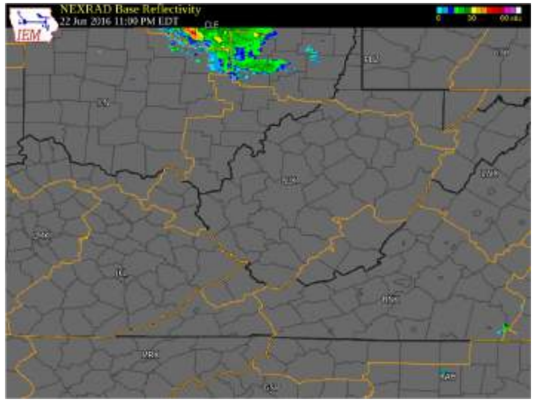


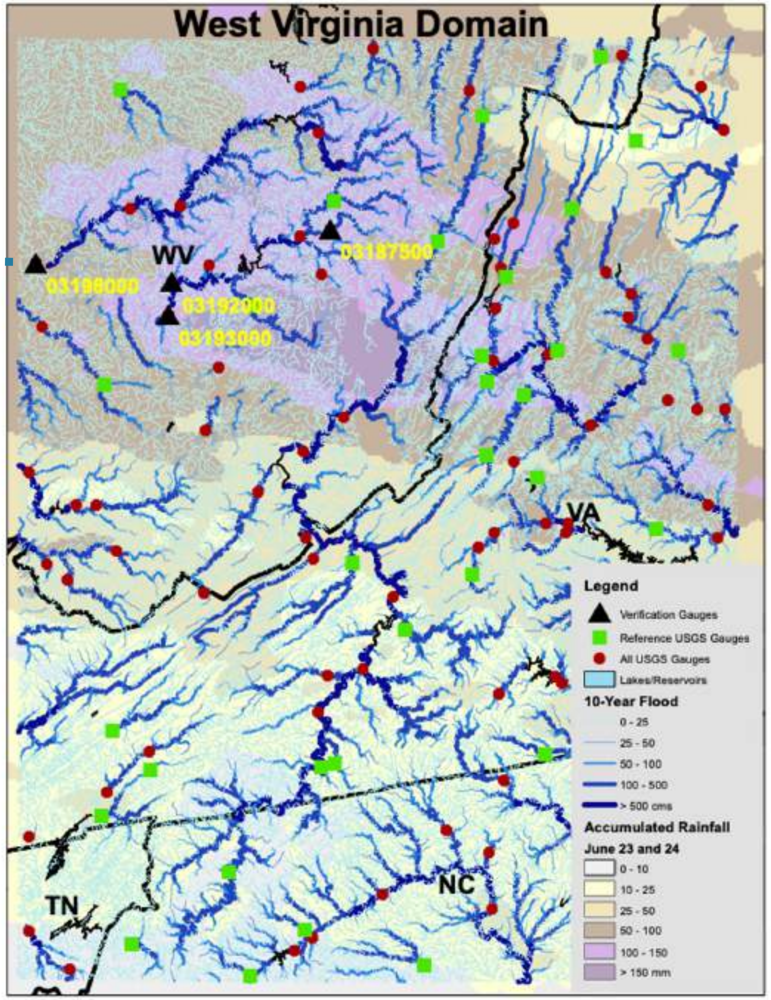


3.5.2 Hybrid Scheme: Accuracy vs the EnKF

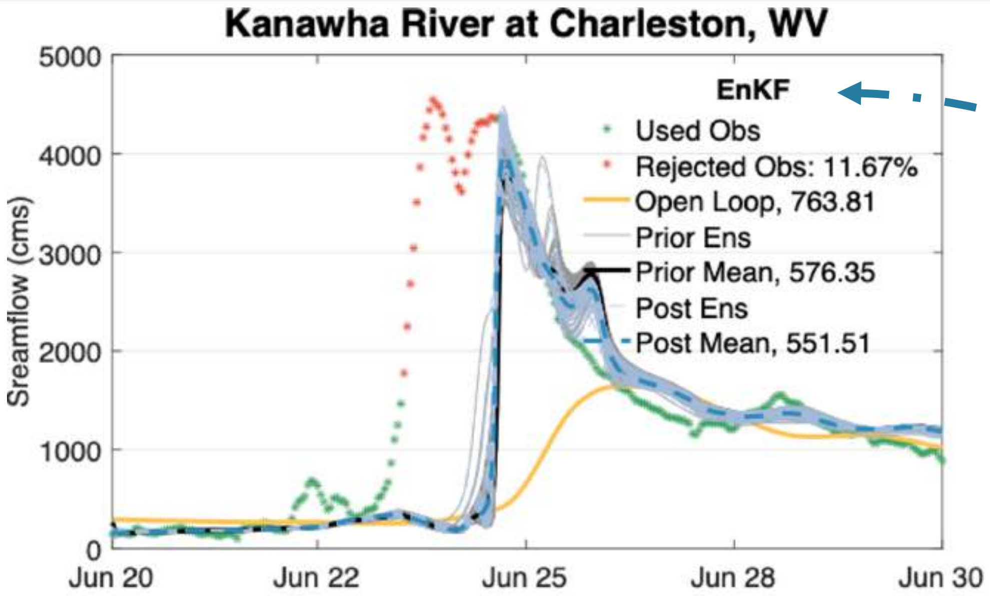


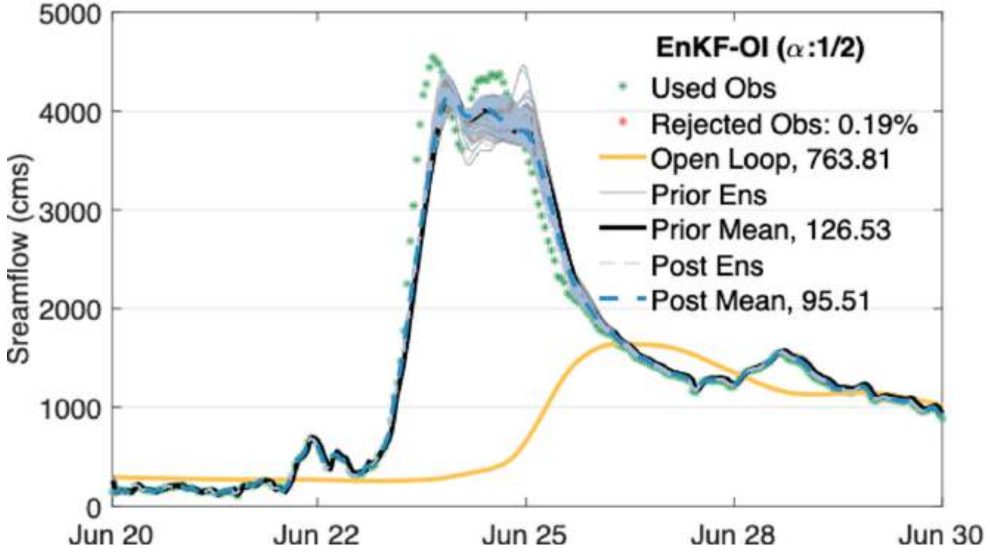


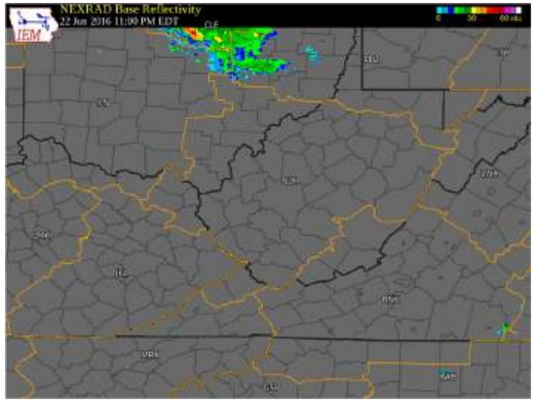


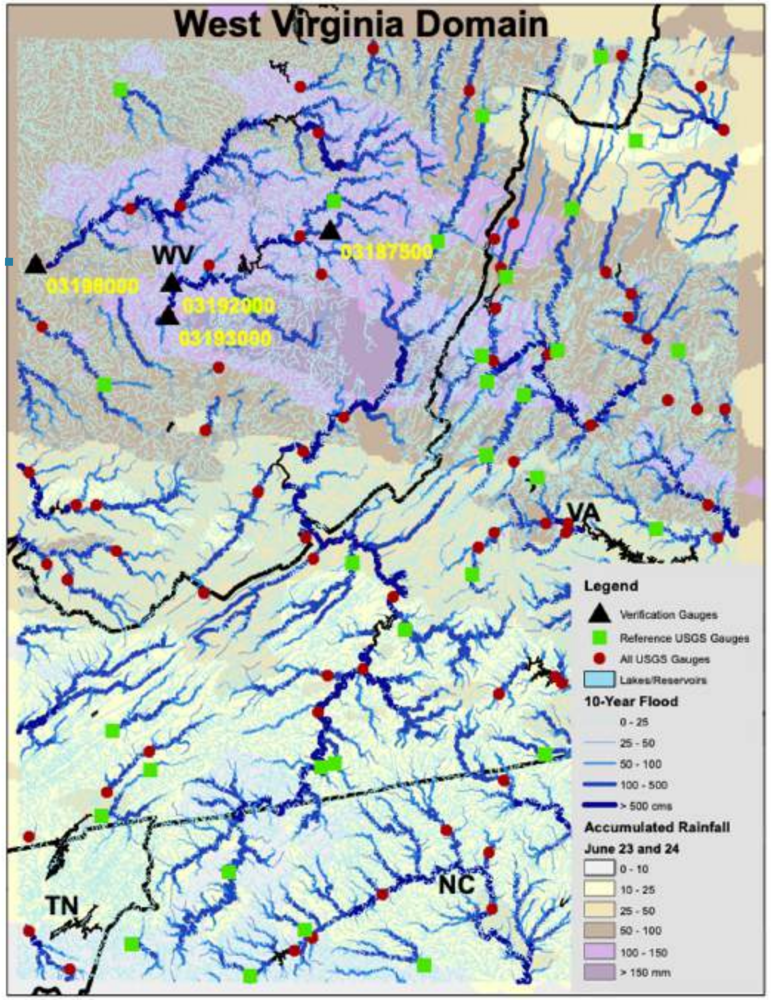


3.5.2 Hybrid Scheme: Accuracy vs the EnKF

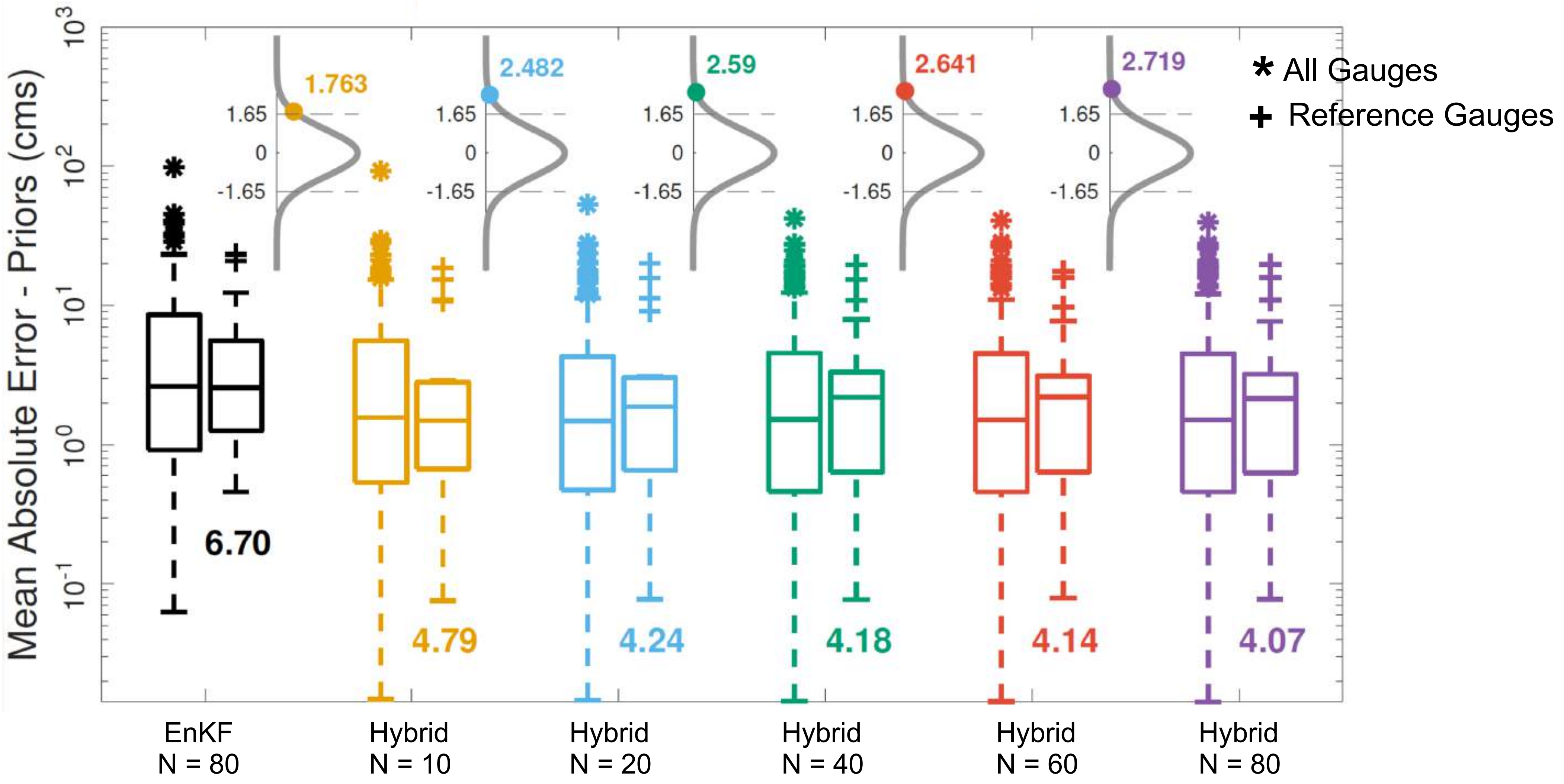






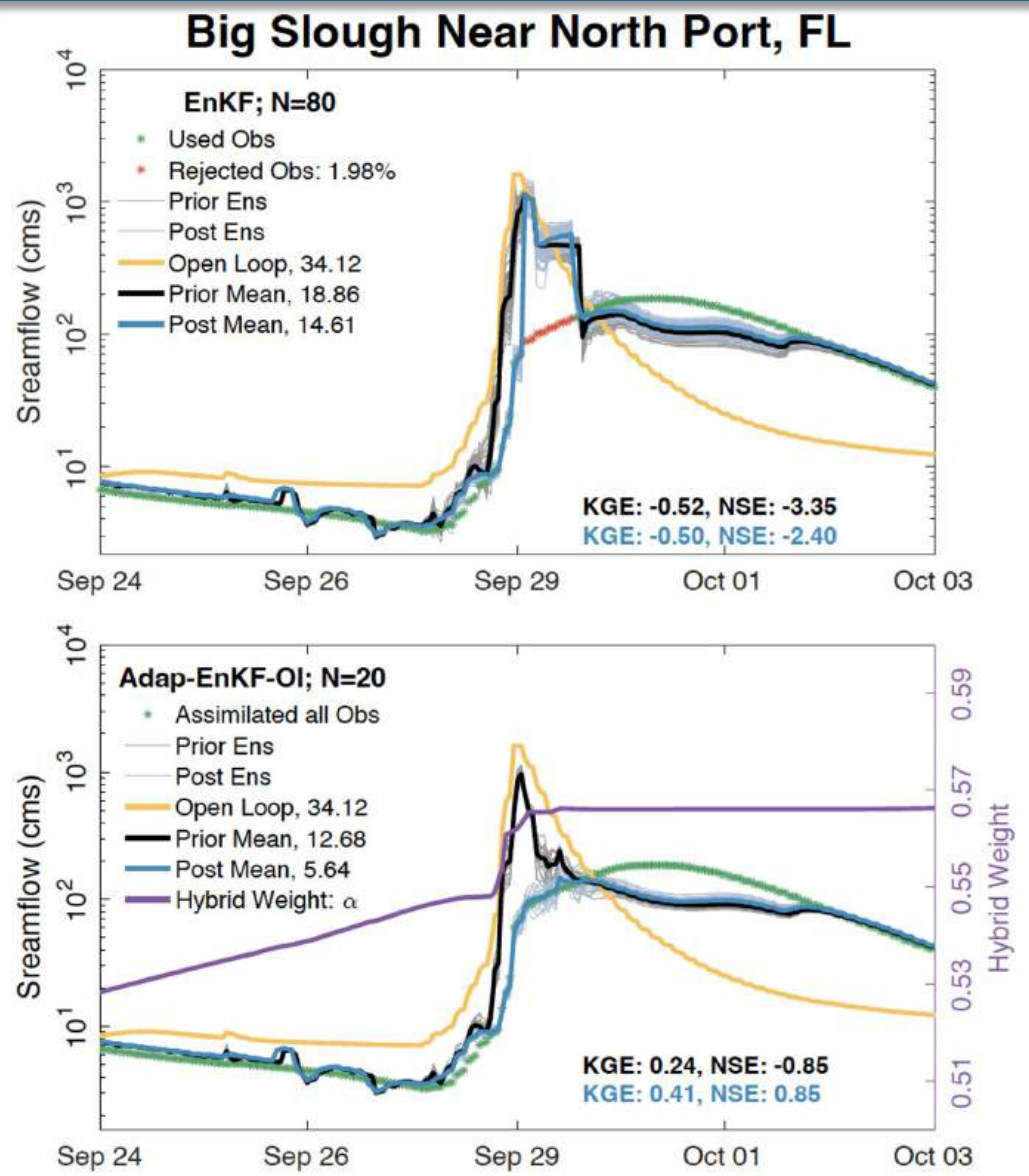


3.5.3 Hybrid Scheme: Computational Efficiency

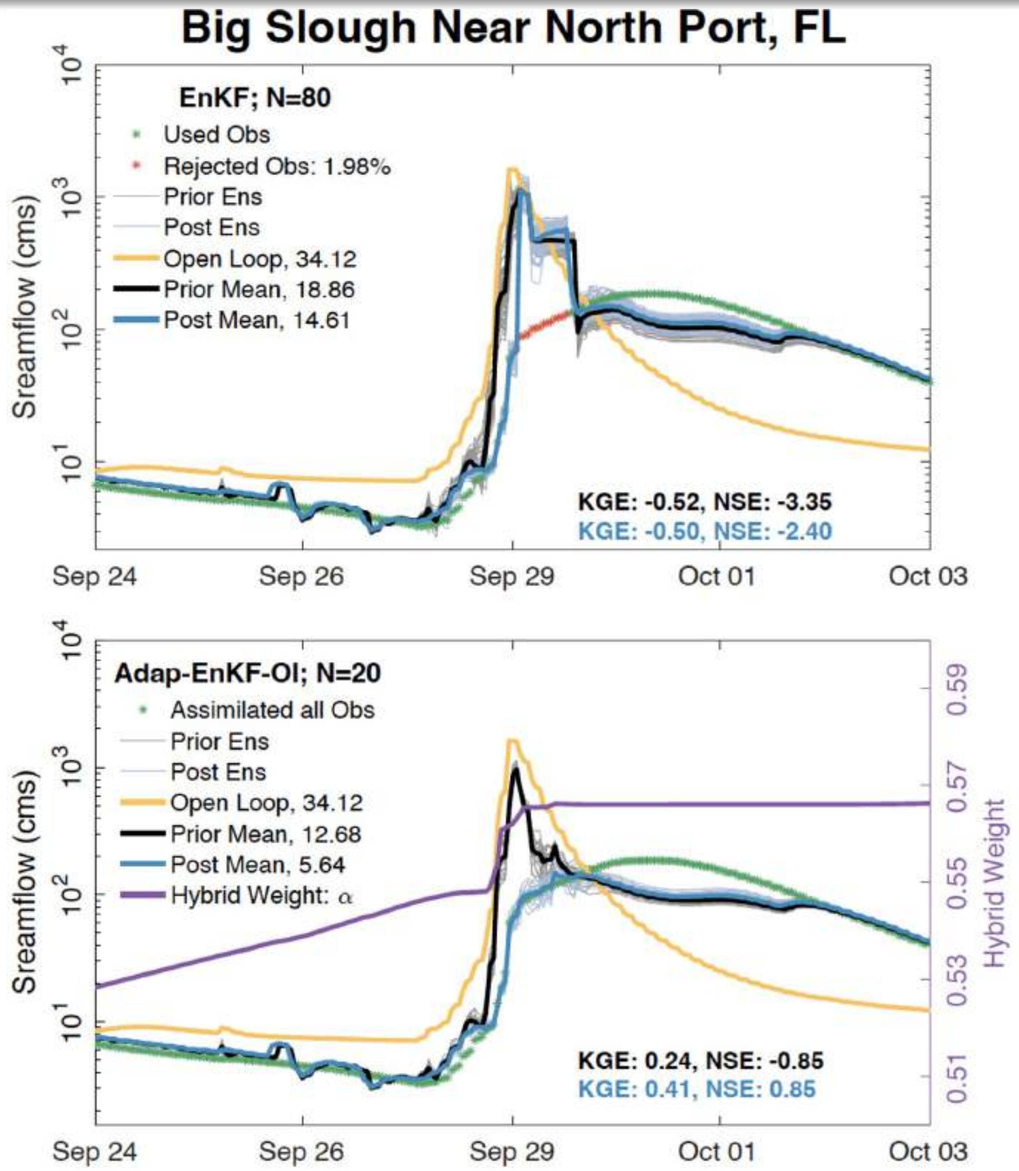


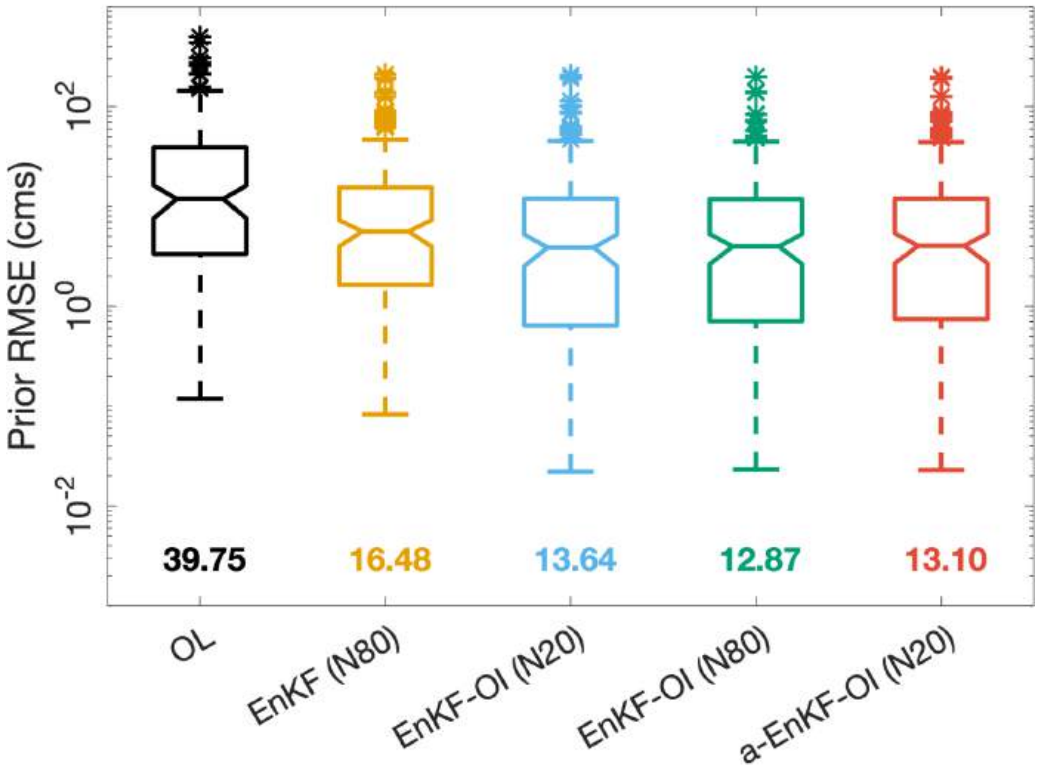
The hybrid scheme is able to provide skillful predictions using **only a quarter** of the computational resources!

3.5.4 Hybrid Scheme: Adaptive Form

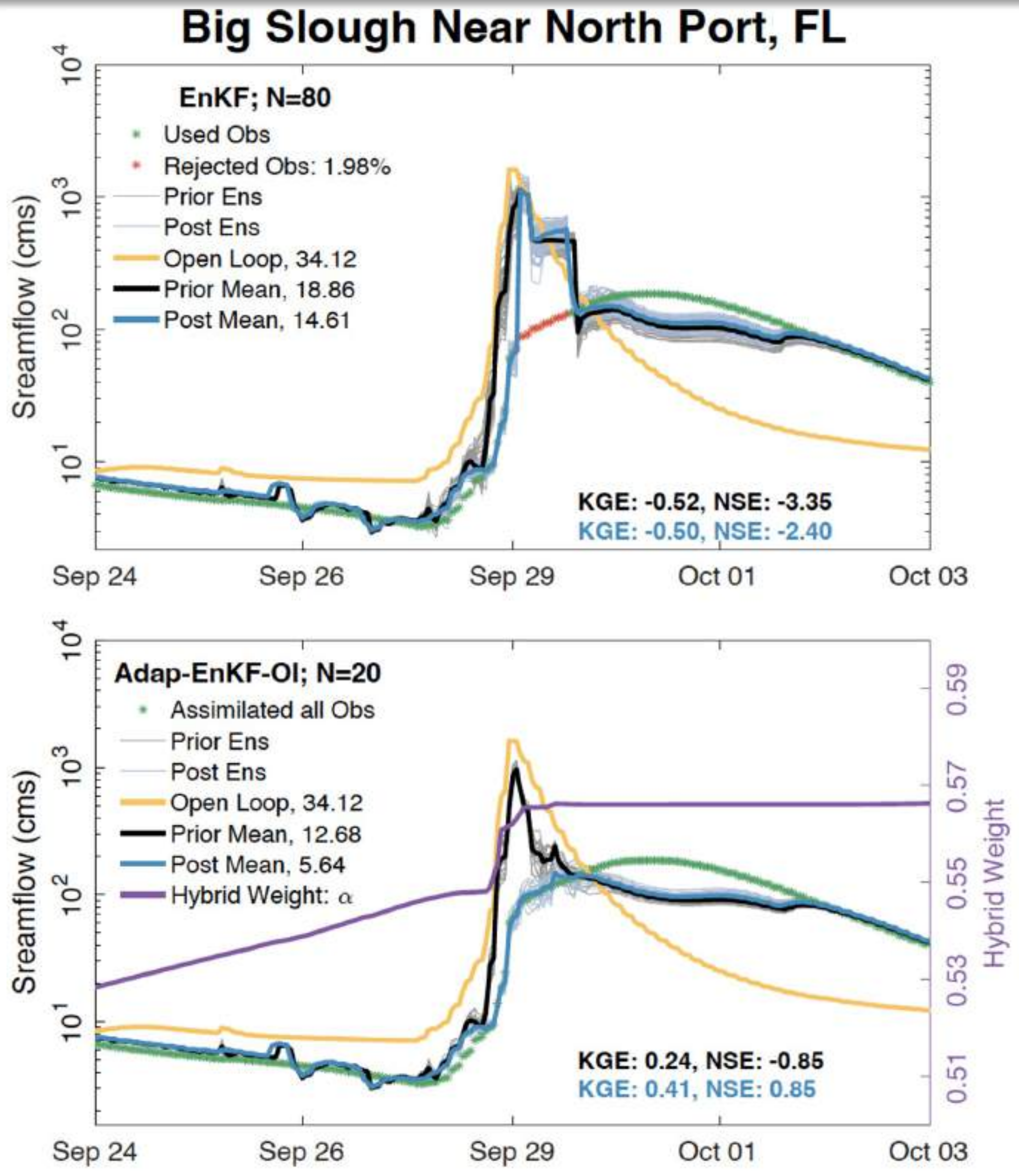


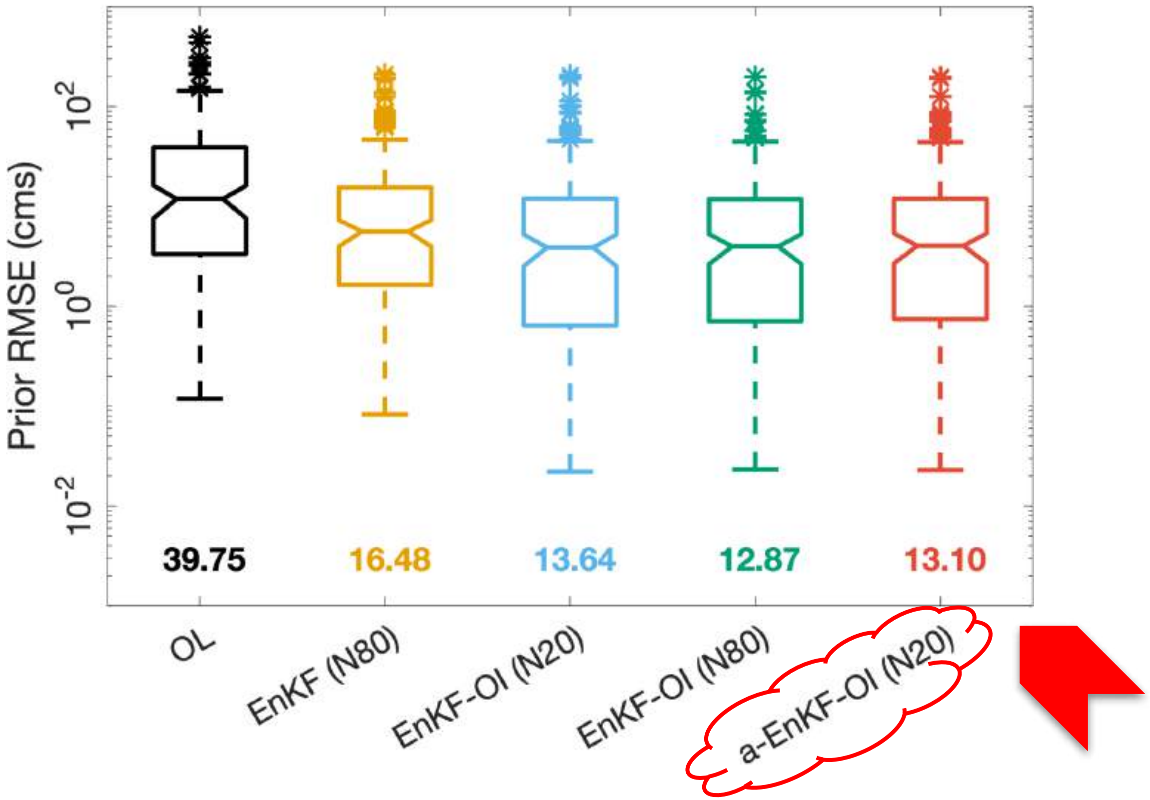
3.5.4 Hybrid Scheme: Adaptive Form





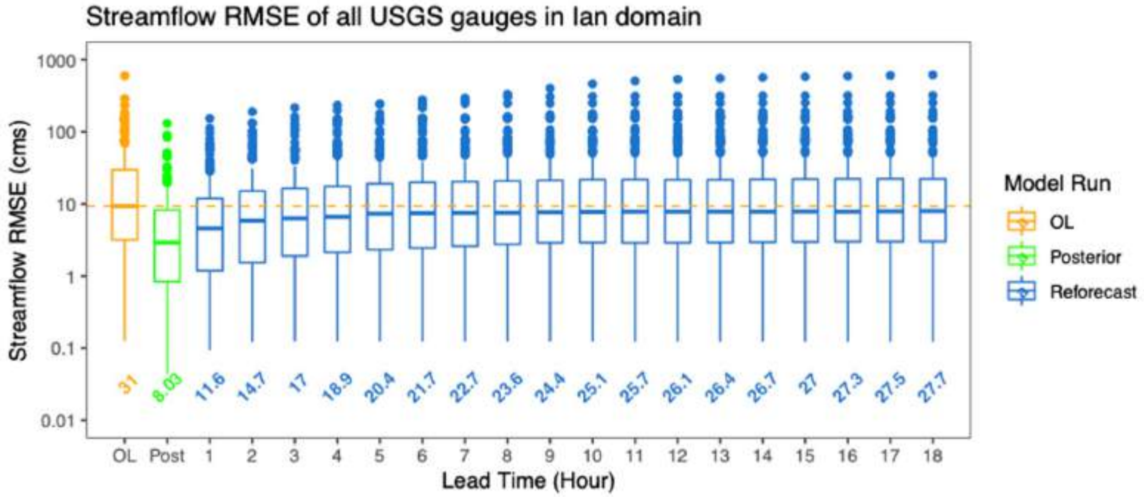
3.5.4 Hybrid Scheme: Adaptive Form

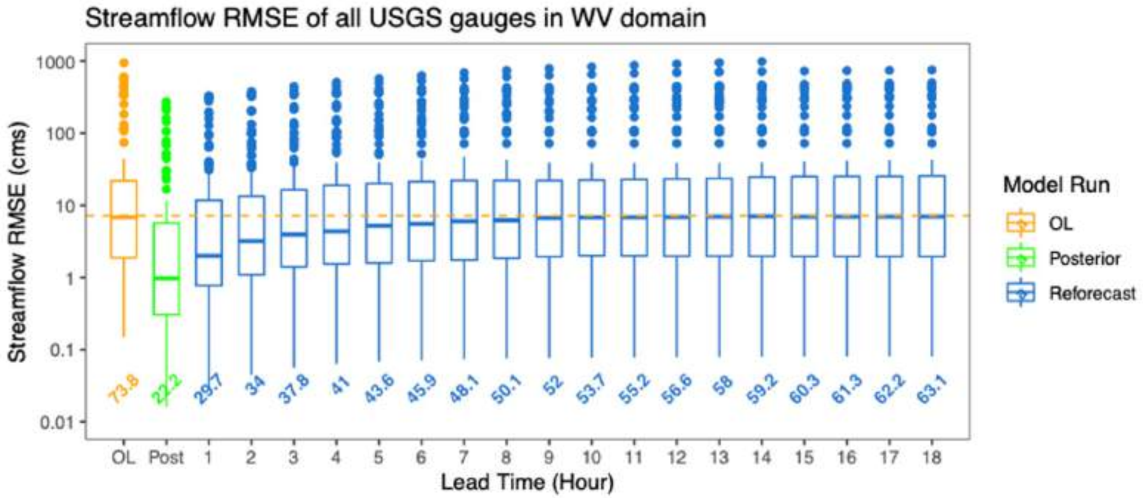




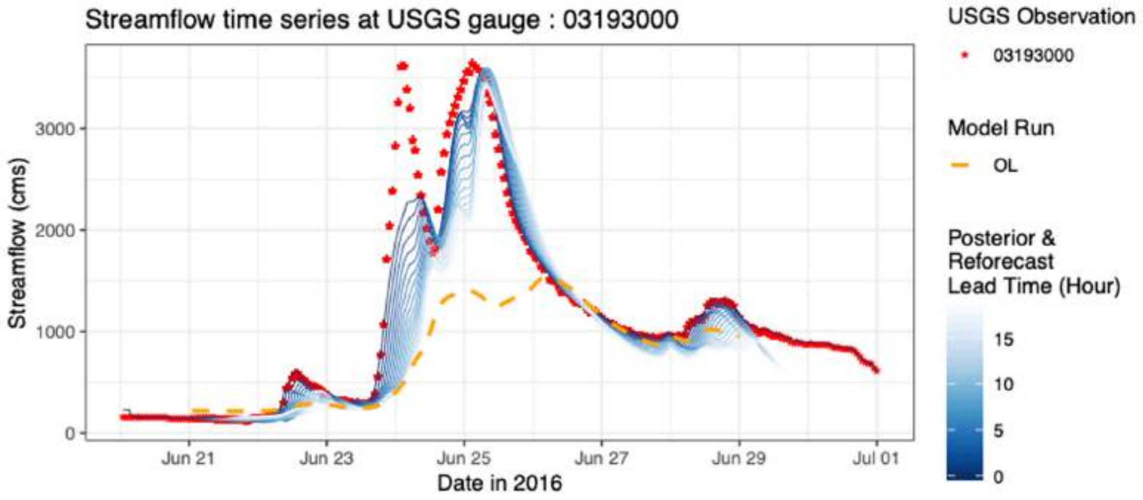
Using **only 20 flow-dependent ensemble members**, the adaptive hybrid EnKF-OI scheme provides **robust performance**, **high quality streamflow estimates** and **minimal use of computational resources**

3.5.5 Hybrid Scheme: Forecast Assessment





Early Warnings: Up to 18 hours ahead of flood peaks!



3.5.5 Hybrid Scheme: Forecast Assessment

❑ For more detailed information, refer to our recent work [El Gharamti et al., 2024]

Hydrol. Earth Syst. Sci., 28, 3133–3159, 2024
https://doi.org/10.5194/hess-28-3133-2024

Hydrology and Earth System Sciences
Open Access
EGU

Leveraging a novel hybrid ensemble and optimal interpolation approach for enhanced streamflow and flood prediction

Mohamad El Gharamti[1], **Arezoo Rafieeinasab**[2], **and James L. McCreight**[3]
[1]NCAR, Computational and Information Systems Laboratory (CISL), Boulder, CO, USA
[2]NCAR, Research Applications Laboratory (RAL), Boulder, CO, USA
[3]UCAR, CPAESS Cooperative Programs for the Advancement of Earth System Science, Boulder, CO, USA

Correspondence: Mohamad El Gharamti (gharamti@ucar.edu)

Received: 13 November 2023 – Discussion started: 25 January 2024
Revised: 15 May 2024 – Accepted: 28 May 2024 – Published: 19 July 2024

Abstract. In the face of escalating instances of inland and flash flooding spurred by intense rainfall and hurricanes, the accurate prediction of rapid streamflow variations has become imperative. Traditional data assimilation methods face challenges during extreme rainfall events due to numerous sources of error, including structural and parametric model uncertainties, forcing biases, and noisy observations. This study introduces a cutting-edge hybrid ensemble and optimal interpolation data assimilation scheme tailored to precisely and efficiently estimate streamflow during such critical events. Our hybrid scheme uses an ensemble-based framework, integrating the flow-dependent background streamflow covariance with a climatological error covariance derived from historical model simulations. The dynamic interplay (weight) between the static background covariance and the evolving ensemble is adaptively computed both spatially and temporally. By coupling the National Water Model (NWM) configuration of the WRF-Hydro modeling system hybrid data assimilation system propels streamflow forecasts up to 18 h in advance of flood peaks, marking a substantial advancement in flood prediction capabilities.

1 Introduction

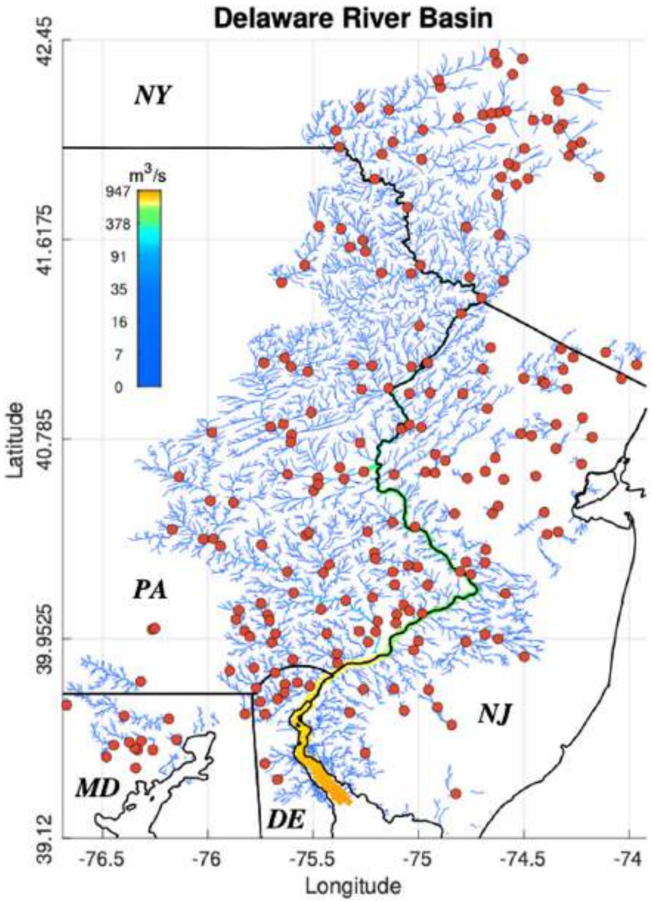
Flooding can stem from various causes, including prolonged rainfall events like tropical storms or hurricanes, as well as intense rainfall over short periods or complications such as debris and ice jams. When examining events causing at least a billion dollars in damage, river and urban flooding alone accounts for 7.4 % of US natural disasters from 1980 to 2023. Tropical cyclones top the list, contributing to 52 % of the damage (Smith, 2020).

Tropical storms and hurricanes are characterized by destructive winds, storm surge, and catastrophic flooding. Hur-

4.1 Current Activities: Optimal Observation Design

USGS **Next Generation Water Observing System (NGWOS)**:

- streamflow quantitative information
- evapotranspiration, snowpack, soil moisture, ..
- water use, water quality constituents
- connections between surface and groundwater

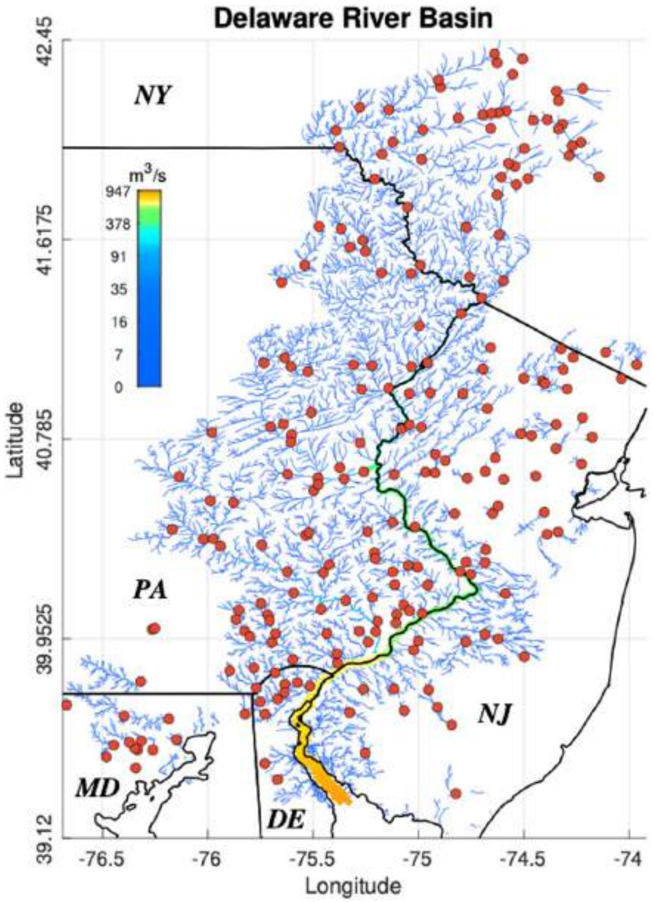


Where to place the new gauges?

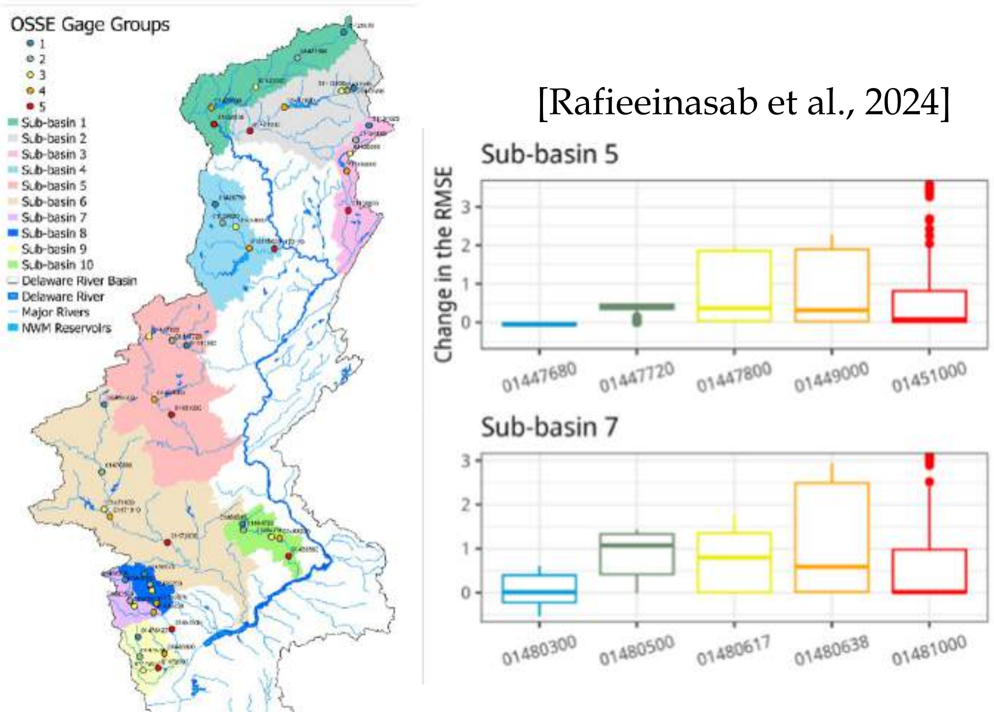
4.1 Current Activities: Optimal Observation Design

USGS **Next Generation Water Observing System (NGWOS)**:

- streamflow quantitative information
- evapotranspiration, snowpack, soil moisture, ..
- water use, water quality constituents
- connections between surface and groundwater



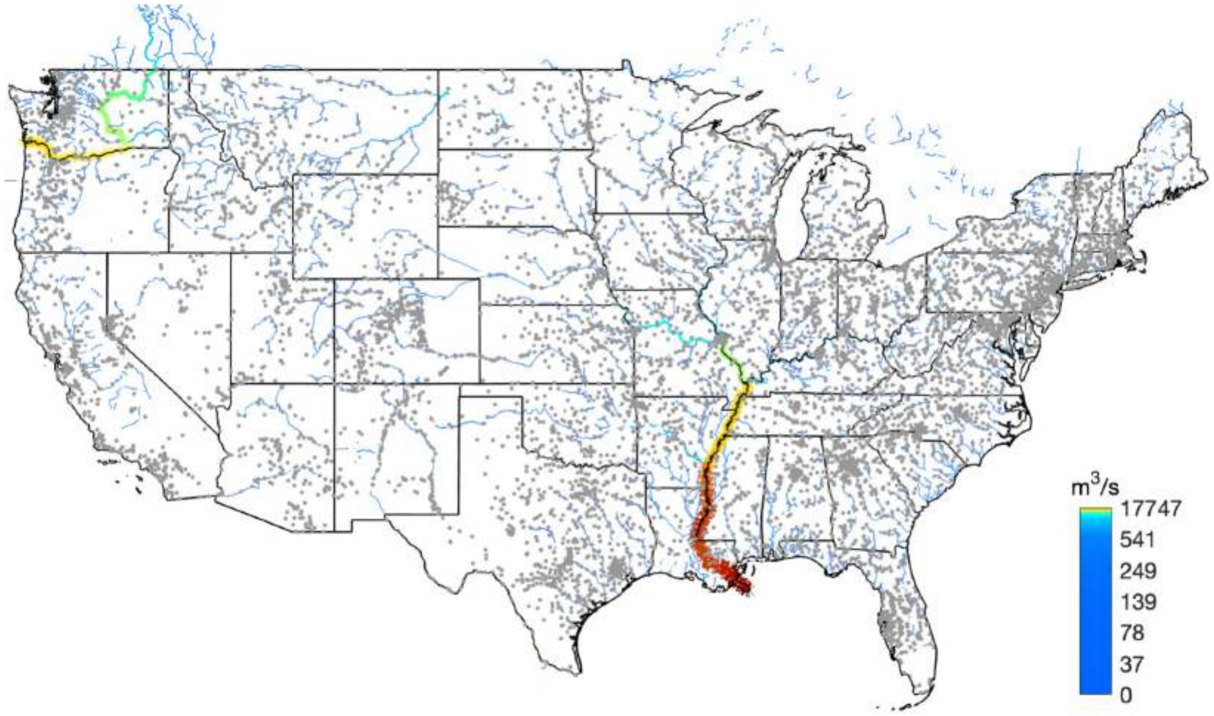
Where to place the new gauges?

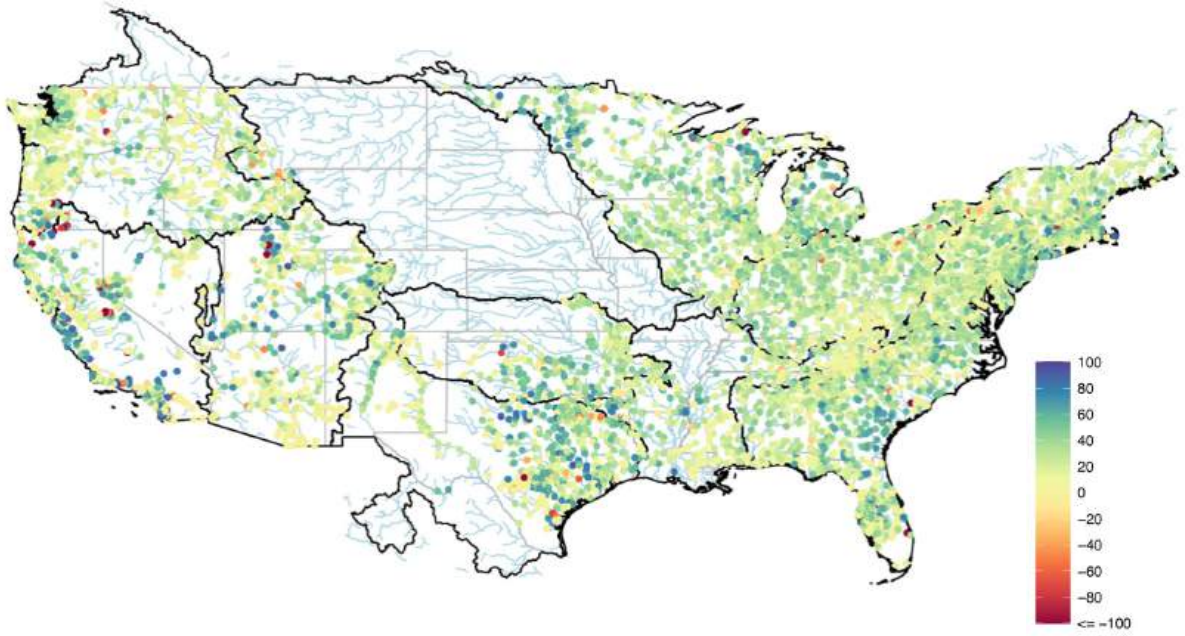


[Rafieeinasab et al., 2024]

1. Observing System Simulation Experiments (OSSEs)
2. Assimilation Impact: Upstream vs Downstream gauges
3. *Next Steps:* Optimize for accuracy, uncertainty, …

4.2 Current Activities: CONUS-wide Predictions





- **Hydro-DART** experiments across CONUS
  - ATS Localization sensitivity runs
  - Hybrid: **B** ⟶ Climatology tuning
  - Non-Gaussian filtering
  - A full reanalysis

Preliminary Results: Percentage of RMSE reduction, by River Forecast Centers (RFC), using Hydro-DART as compared to the Open Loop

5.1 Future Plans: Coupled DA with Hydro-DART

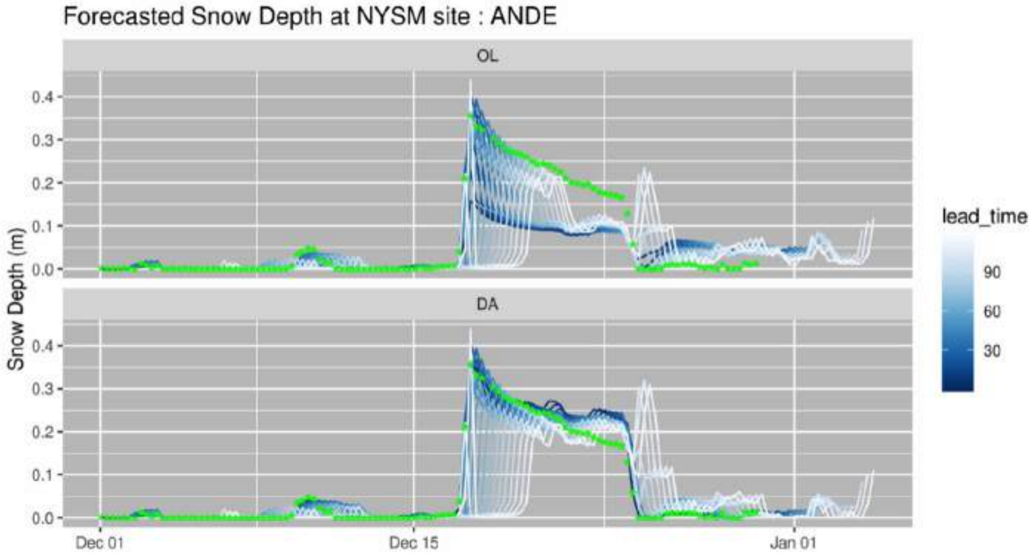
Enhance prediction skill of the coupled land-hydro system:

- ❑ On top of streamflow, we would like to integrate the land surface model, Noah-MP and conduct **weakly and strongly coupled DA**
- ❑ Study the impact of assimilating soil moisture, snow depth, .. on streamflow and vice versa

5.1 Future Plans: Coupled DA with Hydro-DART

Enhance prediction skill of the coupled land-hydro system:

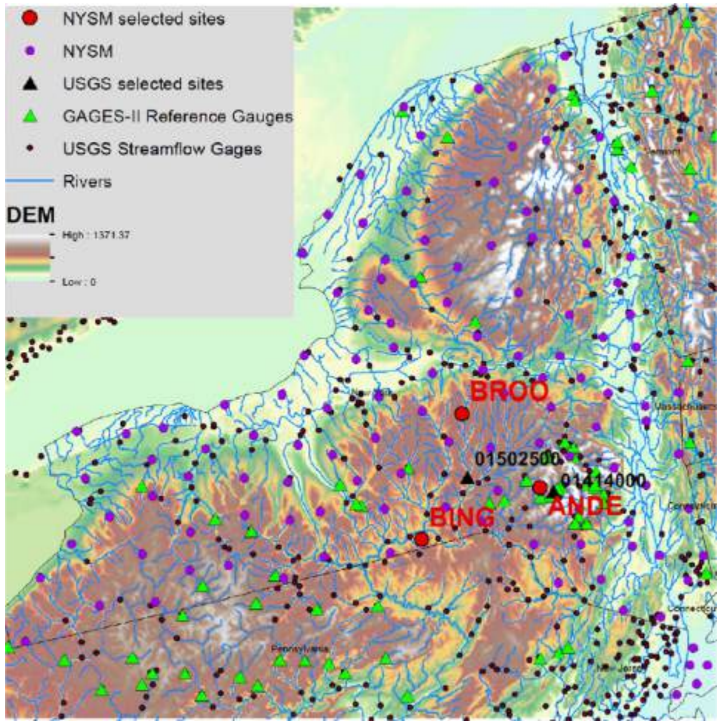
- ❑ On top of streamflow, we would like to integrate the land surface model, Noah-MP and conduct **weakly and strongly coupled DA**
- ❑ Study the impact of assimilating soil moisture, snow depth, .. on streamflow and vice versa



Time-series of snow depth (up to lead time of 120 hours) for the OL and the DA runs at the NYSM site ANDE [Rafieeinasab et al., 2023]

Initial snow DA work:

- ➔ Snow depth measurements from NYSM are assimilated every hour (Dec. 2020)
- ➔ Update streamflow, soil moisture content, accumulated melt, soil ice content, …



Domain area. Purple circles: Location of New York State Mesonet (NYSM) gauges

5.2 Future Plans: pywatershed+DART

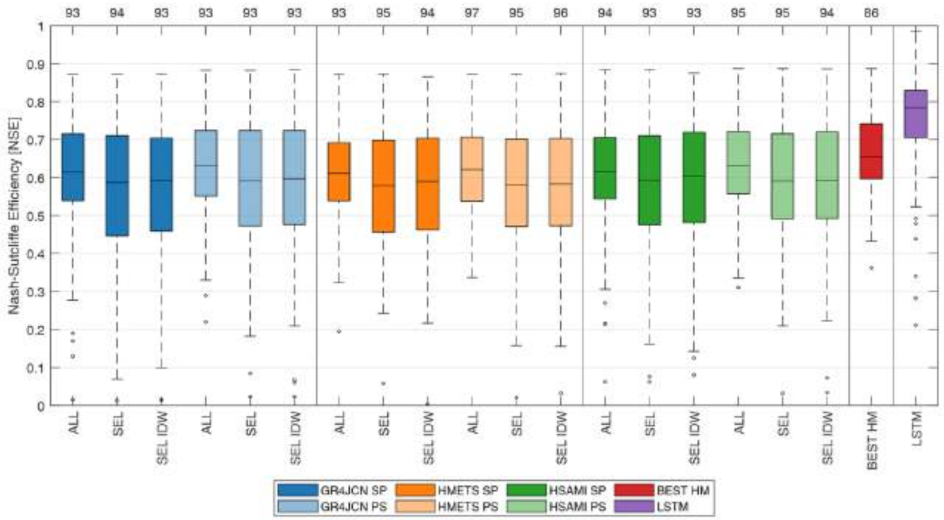
Pywatershed is Python package for simulating hydrologic processes motivated by the need to modernize important, legacy hydrologic models at the USGS, particularly the Precipitation-Runoff Modeling System (PRMS)

Github: `EC-USGS/pywatershed`
Docs: `pywatershed.readthedocs.io`

- ❑ **Goal**: Build an interface between USGS pywatershed hydrological model and DART
- ❑ Interface to mimic Hydro-DART tools
- ❑ Kick-off in early 2025 (very soon)

5.3 Future Plans: ML+Hydro-DART

- **Ungauged Basins**: A basin with insufficient hydrological data (quantity and quality) to draw meaningful predictions
- **PUB: Prediction in Ungauged Basins** was the decadal problem of the International Association of Hydrological Sciences (IAHS) from 2003–2012



Performance of different process-based and LSTM hydrological models. Adopted from Arsenault et al., 2023.

- **Our Idea:** Train a **Generative AI model** over regions where observational data is abundant, and then use it over regions where the observational data is sparse or does not exist and generate *pseudo data*
- *Pseudo Streamflow Observations* can then be effectively used for model calibration and assimilation purposes, enhancing its forecasting skills (flooding and drought)

Hydro-DART on Github: **NCAR/DART/tree/main/models/wrf_hydro**

E-mail: gharamti@ucar.edu
DART: dart@ucar.edu

THANK YOU QUESTIONS?!