Ensemble streamflow data assimilation with NOAA's National Water Model: Novel methods and evaluation applied to hurricane Florence hindcasts.

James McCreight, Moha Gharamti, Seongjin Noh, Tim Hoar, Arezoo RafieeiNasab, Erin Towler

National Center for Atmospheric Research (USA), U. Texas Arlington (USA), Kumoh National Institute of Technology (South Korea)

Exec Summary: TLDR:

1. Nudging is worse than the open-loop ensemble.
2. Ensemble data assimilation improves forecasts but updating aquifer storage is tricky.
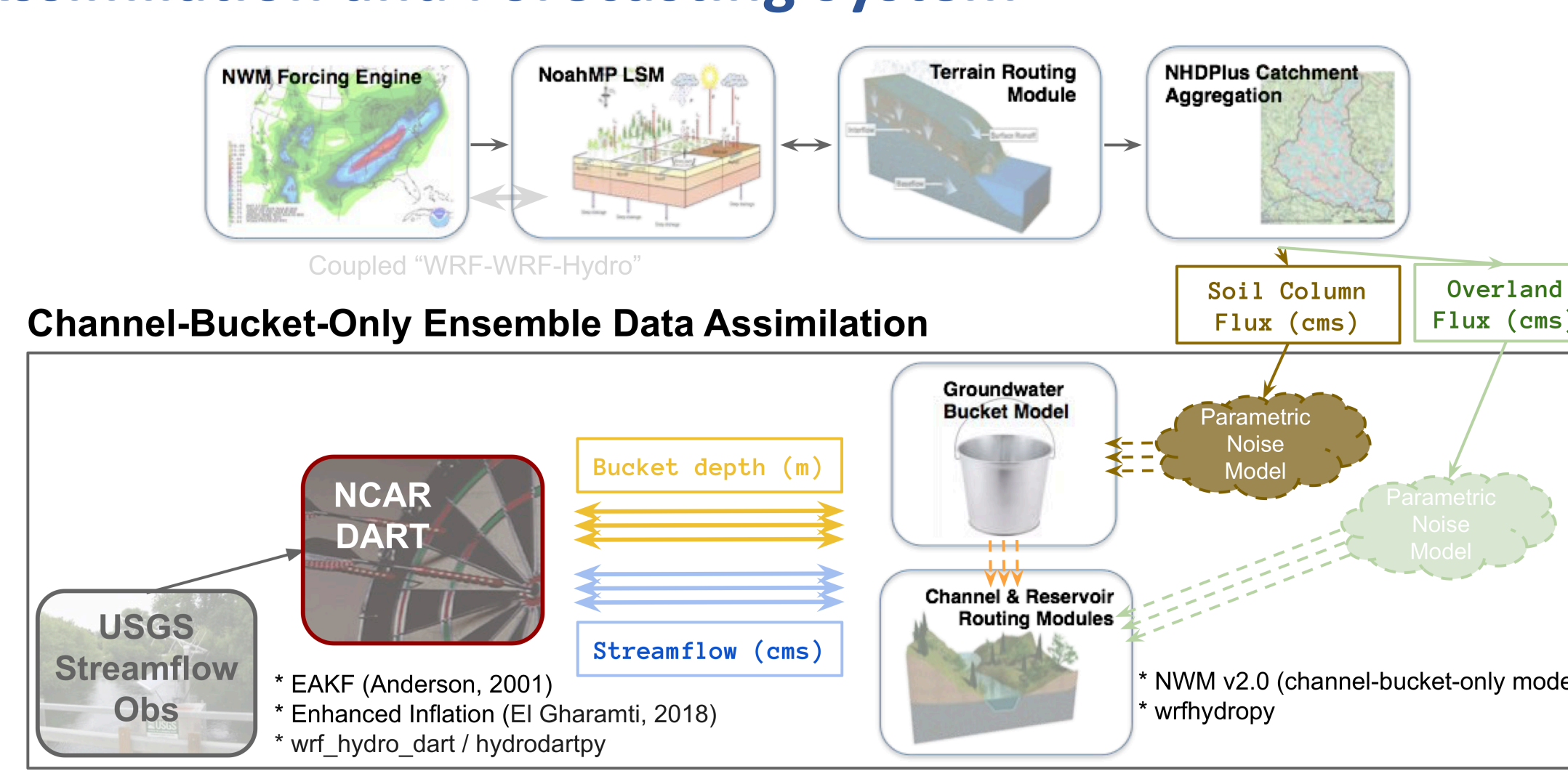3. Forcing errors dominate forecasts: improvements from initial conditions fade after ~10 hours.
4. Novel wavelet approach identifies "events" and their timing errors.

Ensemble Streamflow Assimilation and Forecasting System

US National Water Model:
- NOAA Office of Water Prediction
- Operational on continental US
- Forced by NWP products from NOAA National Weather Service.
- 1km land model
- 250m overland and subsurface hydrologic routing
- Streams and lakes on vector network (~1.5 km)
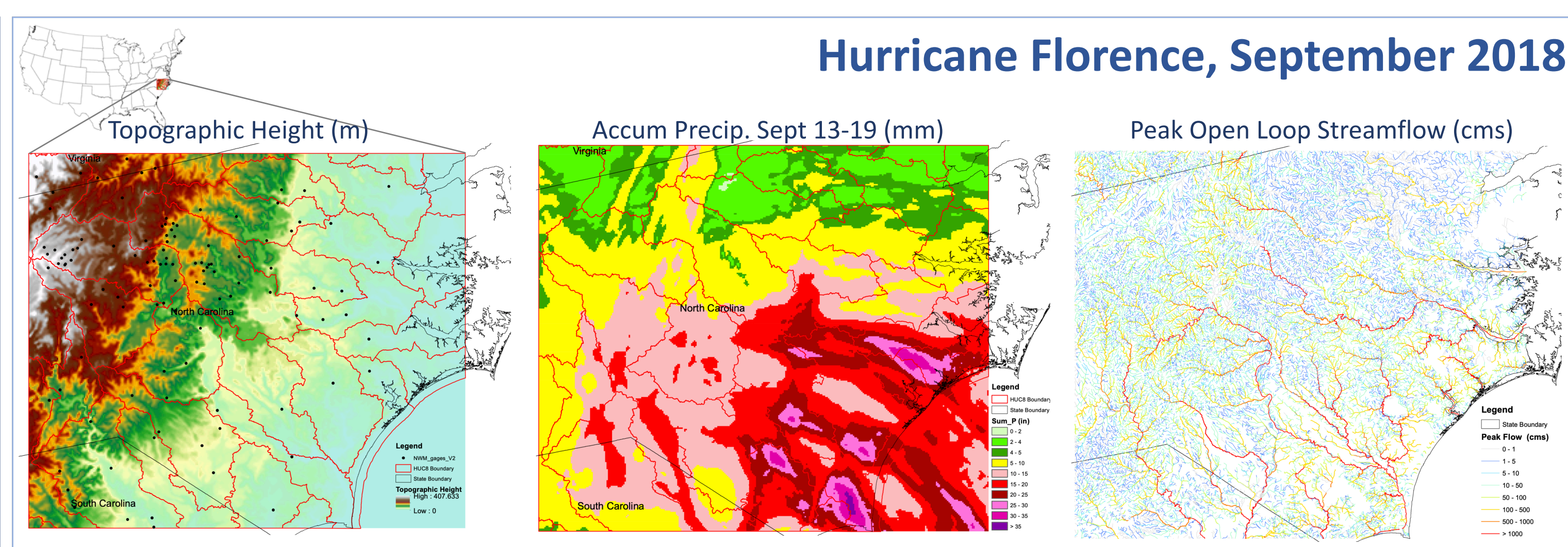- *Streamflow nudging data assimilation: USGS observations*
- Aquifer "bucket" model

Channel-Lake-Bucket Submodel:
- Forced by fluxes from the NWM.
- Computational expense ~1/30

NCAR's Data Assimilation Research Testbed (DART):
- Ensemble assimilation is performed within NCAR's DART.
- Ensemble Adjustment (deterministic) Kalman Filter (Anderson, 2001)
- Spatially- and temporally- varying adaptive inflation (Gharamti, 2018)

Hurricane Florence, September 2018

Topographic Height (m) | Accum. Precip. Sept 13-19 (mm) | Peak Open Loop Streamflow (cms)

- Record-breaking rainfall totals (> 30 in = 76 cm)
- Storm surge
- 105 USGS streamflow gages in the domain
- *Real-time flows not viable for assimilation experiments, using post-corrected flows*
- Widespread flooding
- Realtime NWM AnA forcings used for both analysis and forecasts (*hindcast experiments*)

Methods and Design

Uncertainty:
1. Flux perturbations (depicted above)
2. Channel parameters (6 including geometry and mannings N)
3. Bucket parameters (2)

State Updating:
1. Streamflow
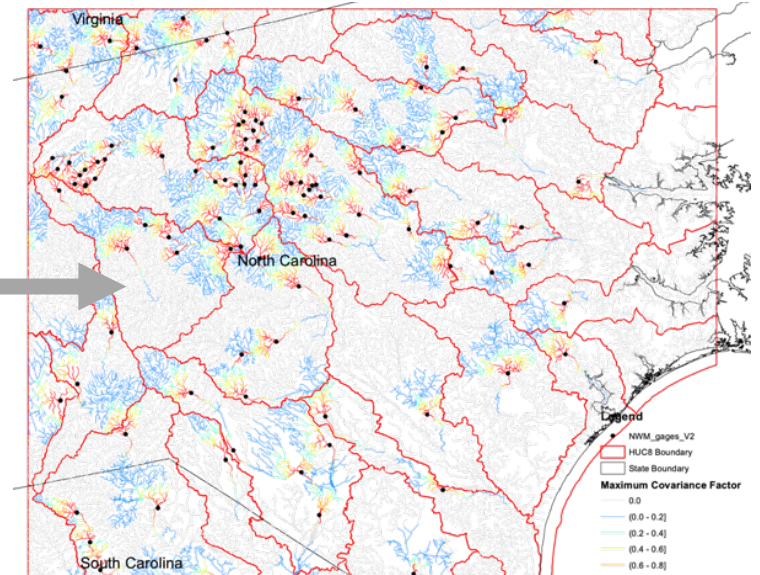2. Streamflow & bucket depth

Experiments:
- **open-loop-multiphysics**: channel & bucket param uncertainty
- **nudging**: deterministic, flux perturb (all ensembles have perturbed fluxes)
- **nobucket-channel-multiphysics**: No bucket updating, channel param uncertainty
- **nobucket-multiphysics**: No bucket updating, channel & bucket param uncertainty
- **bucket-multiphysics**: Bucket updating, channel & bucket param uncertainty

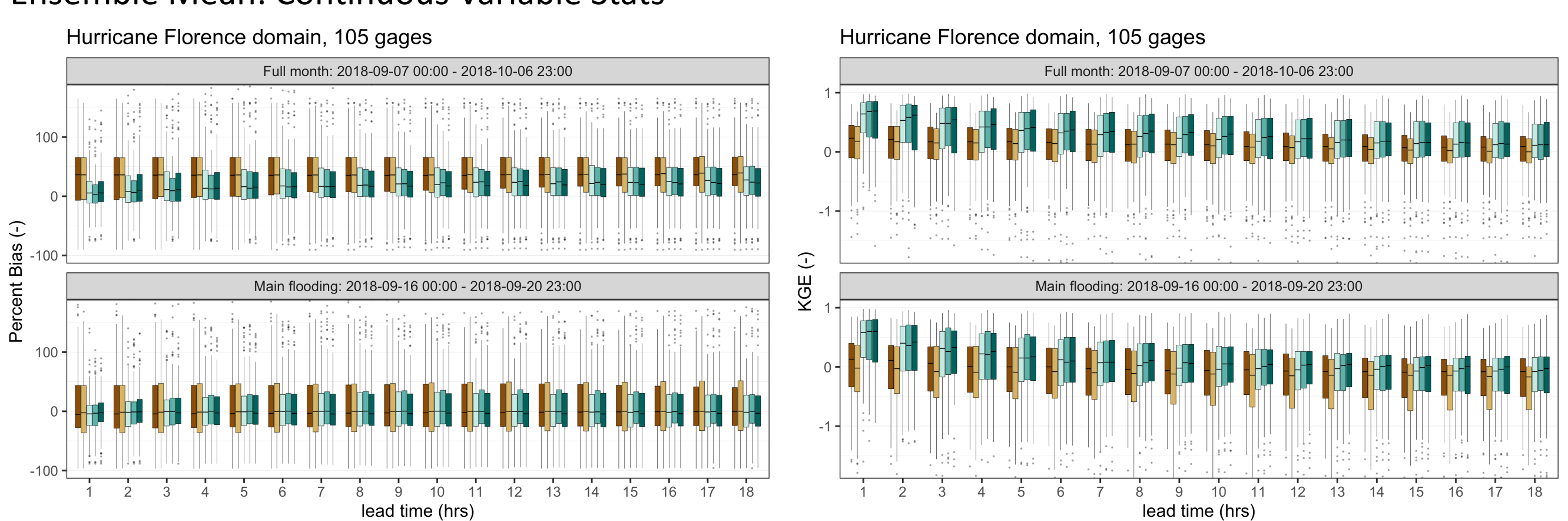**Observation Error Variance**: 20%

Localization: Along-stream metric (as-the-fish-swims) localization strategy

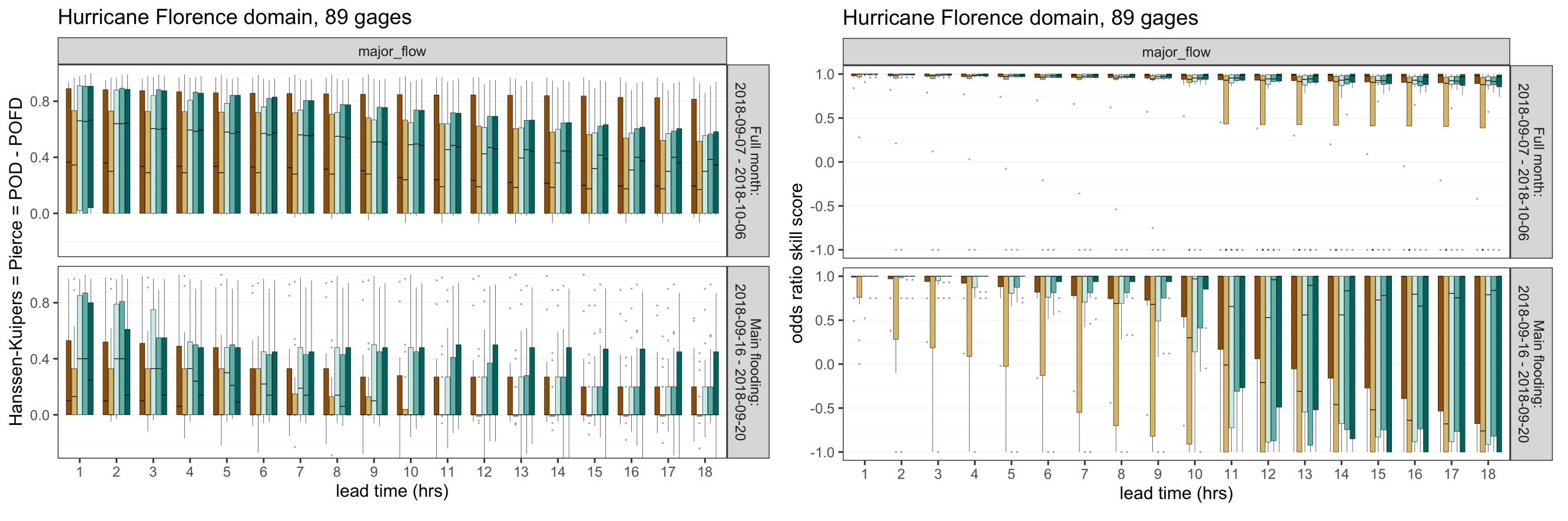**Gaussian Anamorphosis**: technique for the non-gaussian updated states…. Still in development.

Forecast Evaluation

Ensemble Mean: Continuous Variable Stats

Ensemble Mean: Dichotomous Variable (Contingency) Stats

Nudging is worse than the open-loop ensemble:
- Ensemble mean evaluation:
  - Full period: generally similar distributions to open-loop at all lead times
  - Flooding period: generally worse distributions than open-loop
    - Hansen-Kuipers = POD – POFD: IQR negative for large lead times
    - ORSS: Nudging struggles against random chance to discriminate major floods
- Ensemble distribution evaluation: Time-lagged ensembles
  - "Incomplete" until verifying: Evaluate as a function of min lead time
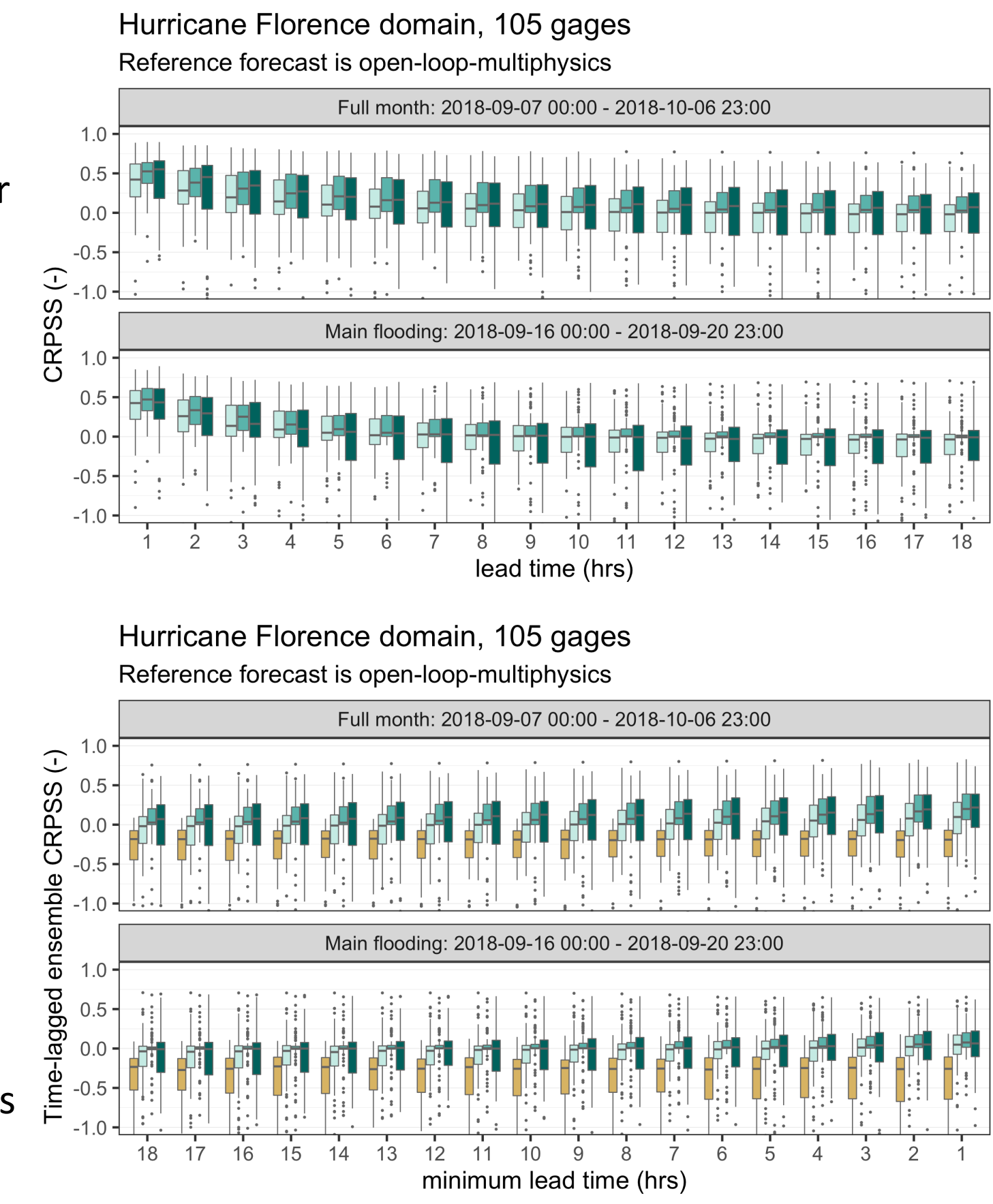 - Time-lagged ensemble CRPSS: Much worse than open-loop

Forcing errors dominate forecasts:
- Ensemble mean evaluation:
  - For all stats, the distributions of the assimilation experiments converge towards the open-loop distribution within about ~10 hours
  - The presented stats are the exception to this rule: convergence is slower
- Ensemble distribution evaluation:
  - Two of the experiments converge to the reference (open-loop) around 9h
  - **Coupled atmosphere – hydro data assimilation to fix "forcing" errors**

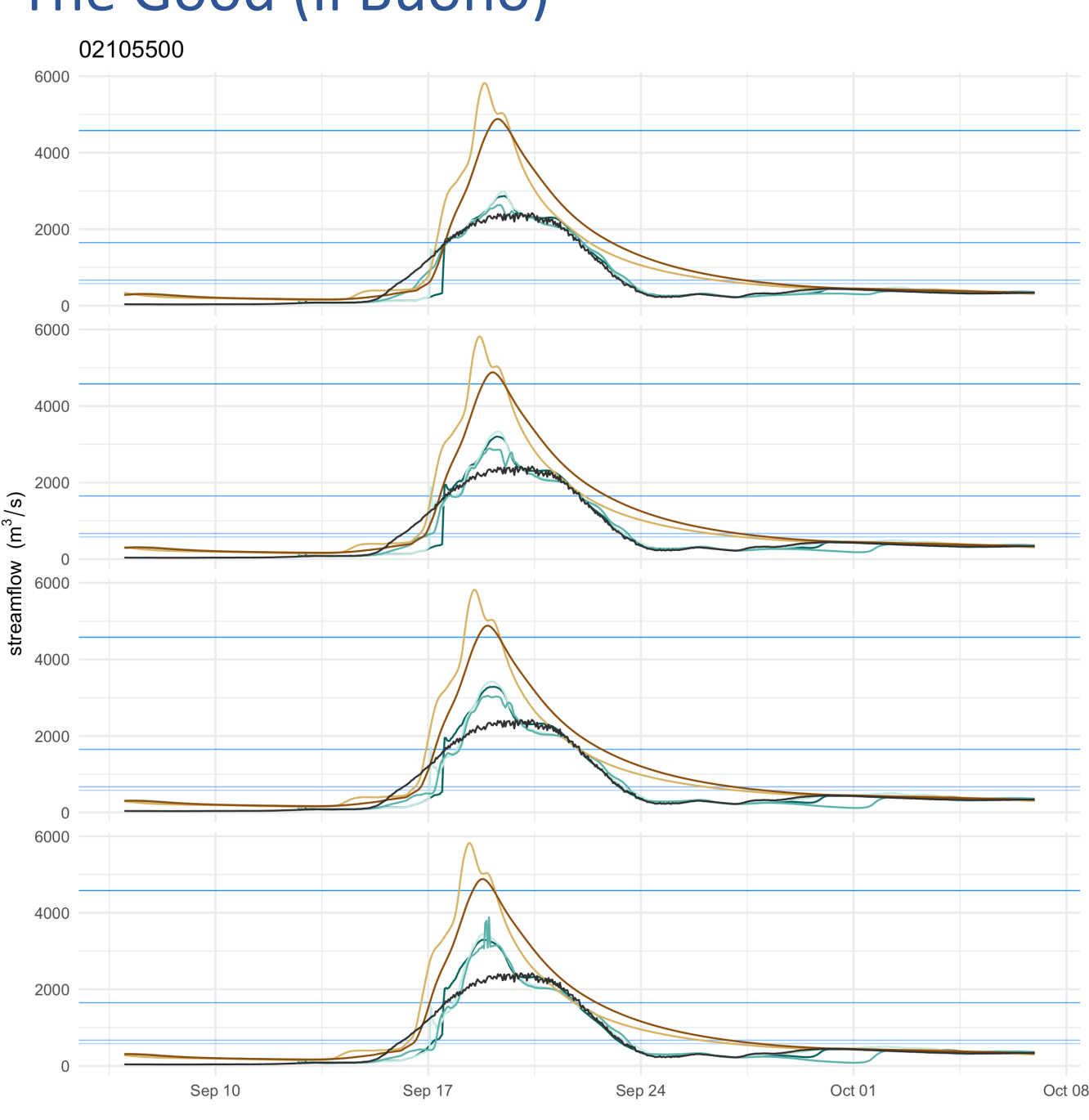### Ensemble DA improves forecasts, updating bucket is tricky:
- Updating the bucket (aquifer) level has pros and cons
- Ensemble mean evaluation:
  - Generally bucket-multiphysics has similar or slightly better distributions as the other assimilation experiments
- Ensemble distribution evaluation:
  - Bucket-multiphysics has wider distributions with better and worse CRPSSs
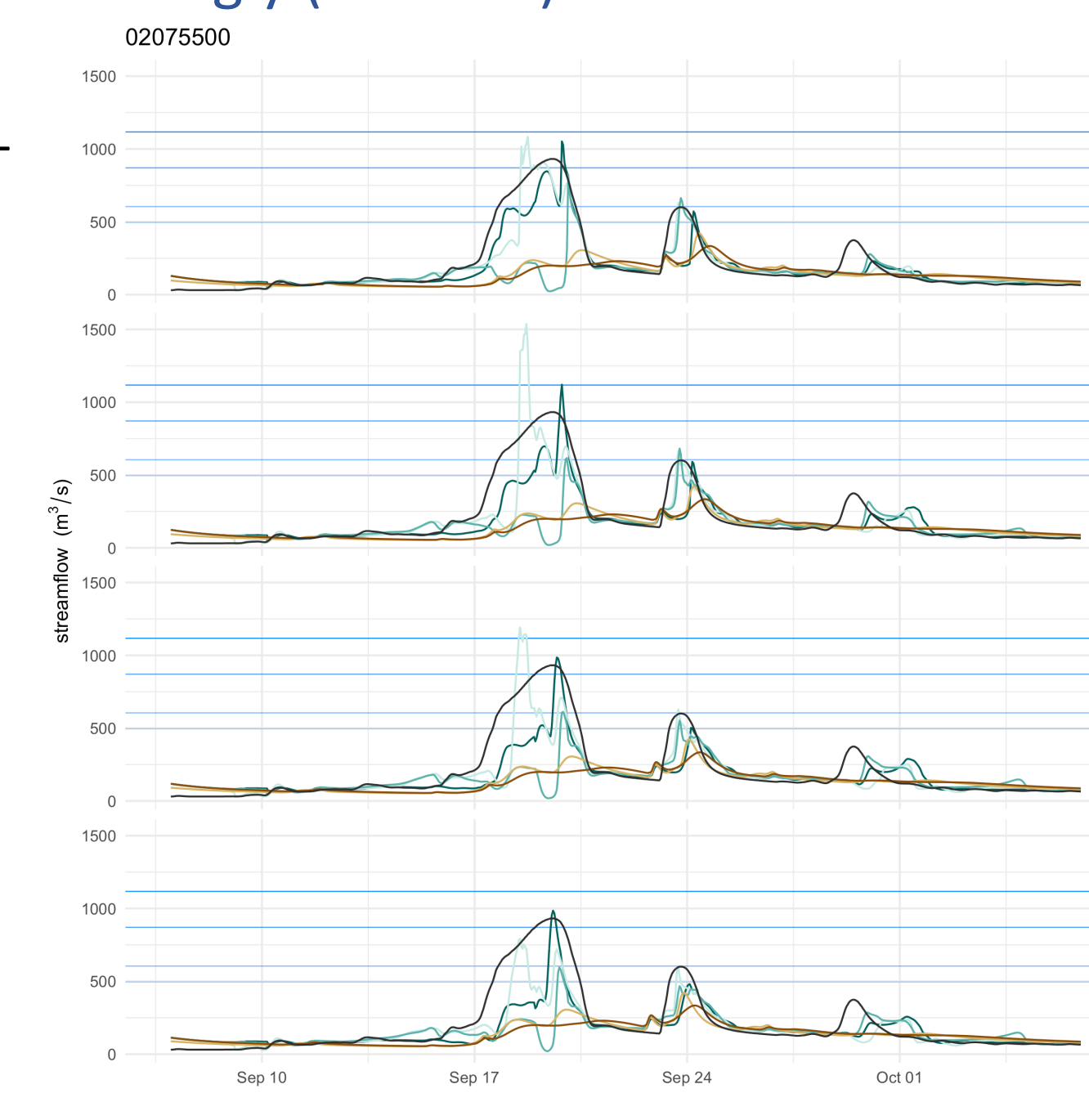
Ensemble Distribution: Probabilistic Variable Stats

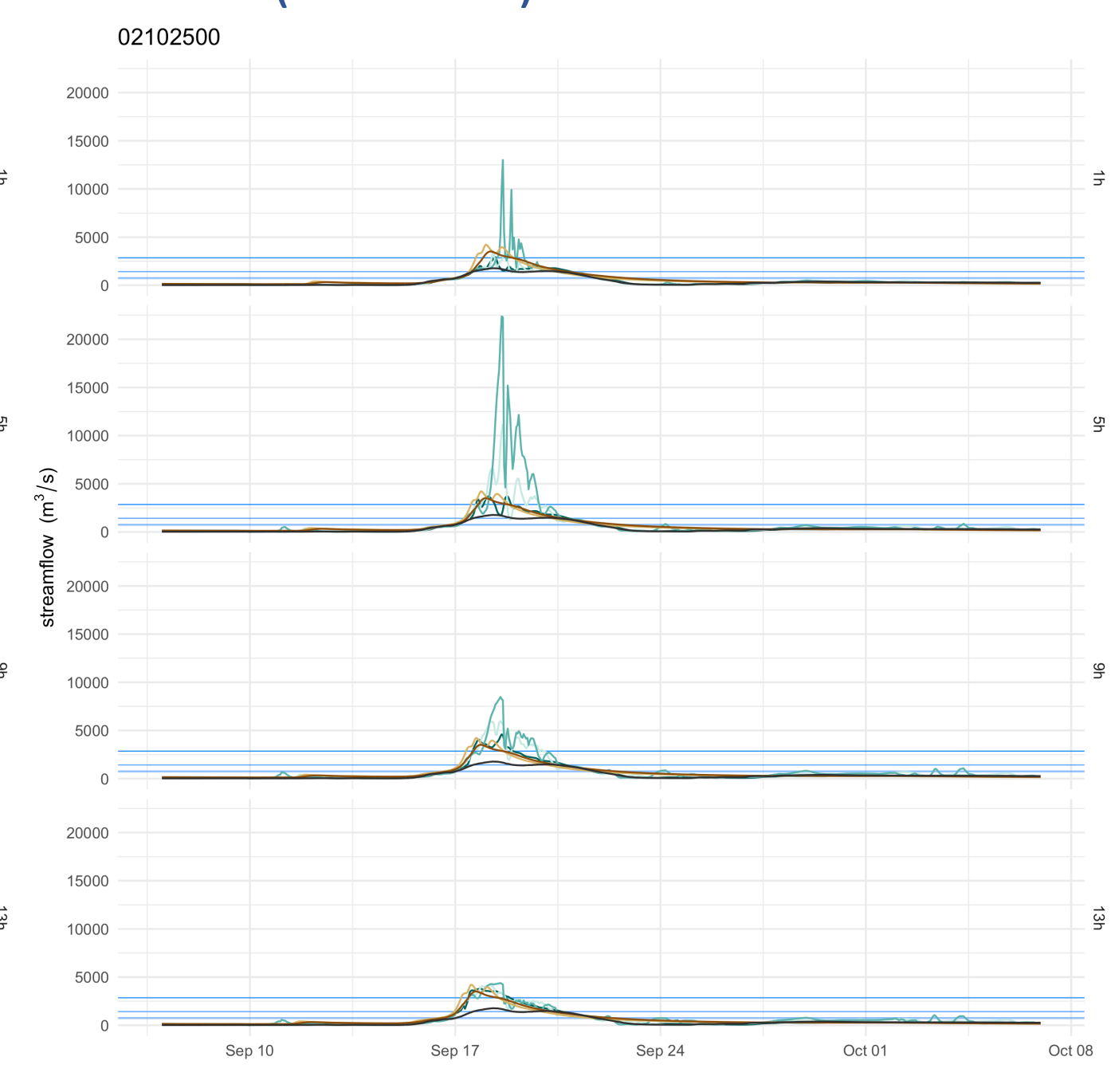## Evaluation Illustrated: Example Ensemble Mean Forecast Hydrographs

The Good (Il Buono)

experiment
- bucket-multiphysics
- nobucket-channel-multiphysics
- nobucket-multiphysics
- nudging
- open-loop-multiphysics
- observed

flow threshold
- major
- moderate
- minor
- action

The Good:
- Strong bias correction (high bias reduction)
  - Even at long lead times, not converging to OL
  - Are upstream corrections to thank?
- Assimilation eliminates a "major" false alarm
- Descending limb is much more accurate
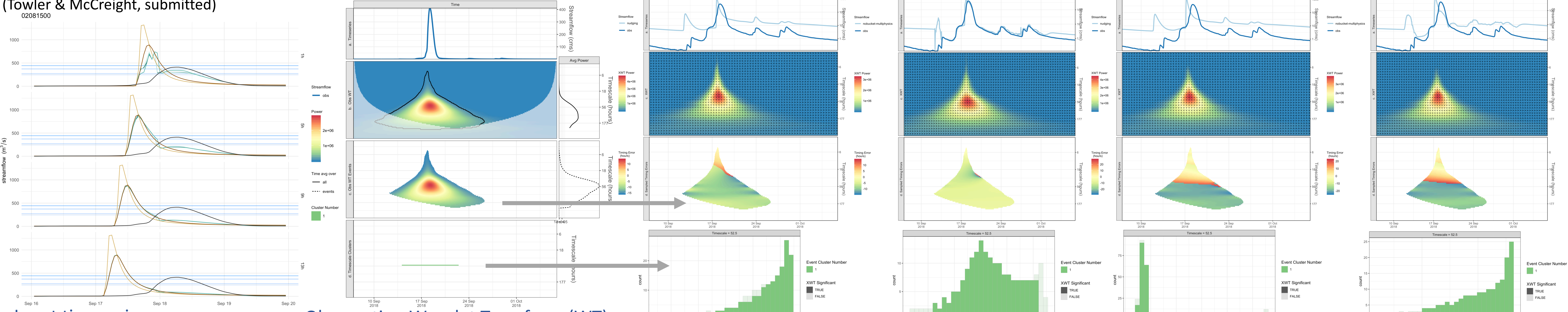- General bias improvements outside flooding

The Ugly:
- Bias correction (low bias correction)
- Open loop and nudging register "miss"
- bucket-multiphysics: much better during first event, worse during second event
- nobucket-channel-multiphysics: False alarms for first event at medium lead times
- nobucket-multiphysics: unrealistic oscillatory behavior, worse misses than nudging and open-loop. **Smoother?**

The Bad:
- False alarms galore!
- Highly unrealistic spikes reflected in the mean are under investigation

The Ugly (Il Brutto)

The Bad (Il Cattivo)

Novel Wavelet Event Detection and Timing Error Quantification
(Towler & McCreight, submitted)

Input timeseries:
- Simple event example
- Observed
- Modeled:
  - 1 and 13 hour forecasts
  - Nudging & nobucket-multiphysics

Observation Wavelet Transform (WT):
- Significance defines events (b -> c)
- Events have dimensions: time, period
- Peaks in event spectrum: "characteristic" periods of observed events = ~53hr
- Timescale event clusters (c -> d)

Cross-Wavelet Transforms (XWT):
- XWT phase (panel b) gives timing error on observed events (panel c)
- XWT significance (not shown) defines timing error significance
- Timing error stats (bottom) calculated on events' characteristic periods defined by the observation WT and its event spectrum

Timing Error Distributions:
- Stats can be taken on the significant timing errors
- 1 hr: nudging is ~11hr early, nobucket-multiphysics is ~5hr early
- 13 hr: nudging is ~23hr early, nobucket-multiphysics is ~20.5hr early
- Timing errors are converging with lead time (as are forecasts).