Ensemble Streamflow Forecasting with the Coupled GFS-Noah Modeling System

Dingchen Hou, Kenneth Mitchell, Zoltan Toth,*
*Dag Lohmann** and Helin Wei**

Environmental Modeling Center/NCEP/NOAA, USA
5200 Auth Road Camp Springs, MD 20746
* EMC/NCEP/NOAA and SAIC
** Risk Management Solution Ltd. UK

Acknowledgement:
Dongjun Seo, Pedro Restrepo and John Schaake, OHD/NOAA
George Gayno, Yuejian Zhu, Jesse Meng, Bo Cui and Youlong Xia,
EMC/NCEP/NOAA

3rd HEPEX Workshop
Stresa, Italy June 27-29, 2007

OUTLINE

- **Introduction**
- **Configuration and Experimental Design**
- **Case Studies**
- **Statistical Evaluation of the Results**
  - Temporal Correlation
  - Continuous Ranked Probability Score (CRPS)
- **Conclusions and Discussions**

Introduction

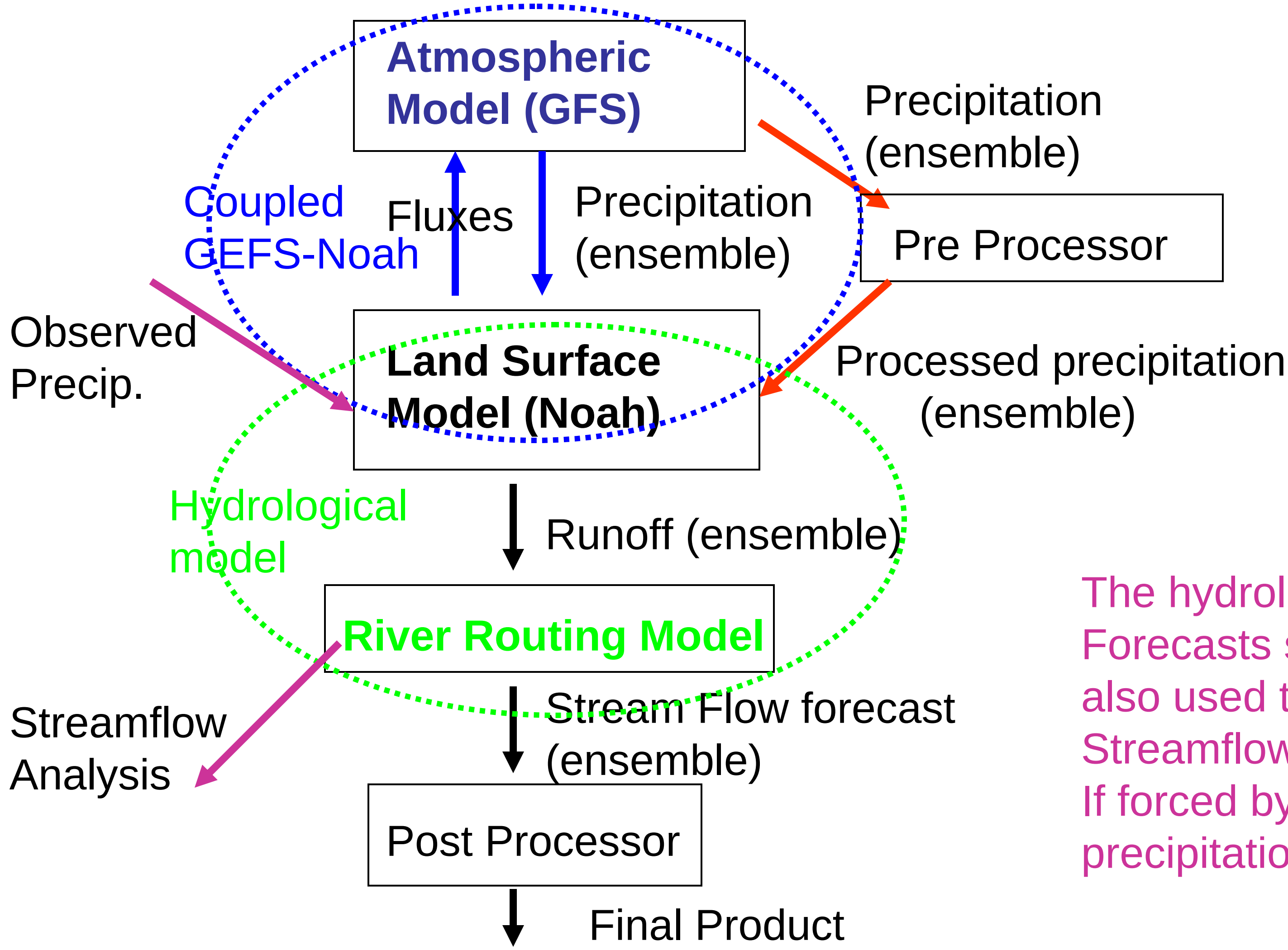
Ensemble Streamflow Forecast: Two Possible Approaches

A) (Proposed approach)

Use the NWP precip. Forecast
Retain the Ensemble members
Retain as much precip. info as possible

B) (Traditional approach)

Pre-processing of NWP precip. forecast
Regenerate ensemble members
Retain less precip. forecast info.



The hydrological Forecasts system is also used to generate Streamflow Analysis, If forced by observed precipitation.

Introduction: Purpose and Strategy

Purpose:

- Demonstrate feasibility of **gridded** river flow forecast in operational ensemble forecast systems (e.g. GEFS).
- Test the quality of the forcing to the hydrological model from the coupled GFS-Noah ensemble forecasting system and identify simple online procedure to improve it.
- Establish suitable configuration for the air-land-river coupled system which can be used with any river routing model.
- Develop suitable strategy to account for uncertainties.
- Test suitable methods for calibrating the products.

General Strategy:

- Focusing on natural (uncontrolled) flow forecast to support water management decisions (e.g., Georgakakos et al, 2006);
- Using NLDAS streamflow simulations as analysis, which is from estimated real precipitation and matches the observations well;
- Keeping global domain in mind with domestic and international users, while CONUS domain being used in this study.
- Developing river flow forecast capacity as a component of the ESMF system;
- Generating hind cast data set for post processing.

Configuration of the River Routing Model

Configuration of the River Routing Model

- **River Routing Model**: linear program, distributed approach, same as used in NLDAS (Lohmann et al., 1998, 2004).
- **CONUS domain**, 1/8 degree grid size (same as NLDAS).
- **River Flow Direction Mask**: A D8 model, river stream in each grid point is discharged to 1 of the eight main directions (Lohmann, et al, 2004).
- **Initial Condition**: NLDAS streamflow analysis.
- **Forcing**: Runoff from global ensemble forecasts (GEFS, control and 10 perturbed members) and the high resolution control forecast (GFS), interpolated to NLDAS grid and 1 hour intervals.
  - Downscaling not considered yet
- **Uncertainty** considered in river routing:
  - in forcing, included partially
  - In hydrological model, ignored but systematic model errors can be corrected via post processing
- **Evaluation**: Using NLDAS streamflow analysis as the verification. Observation may be used in follow up study.
  - Natural flow is compared
  - Uncertainty associated with the meteorological forcing is isolated
  - Consistent with the focus of the present study

Forecast Example (initiated April 1st, 2006) Stream Flow

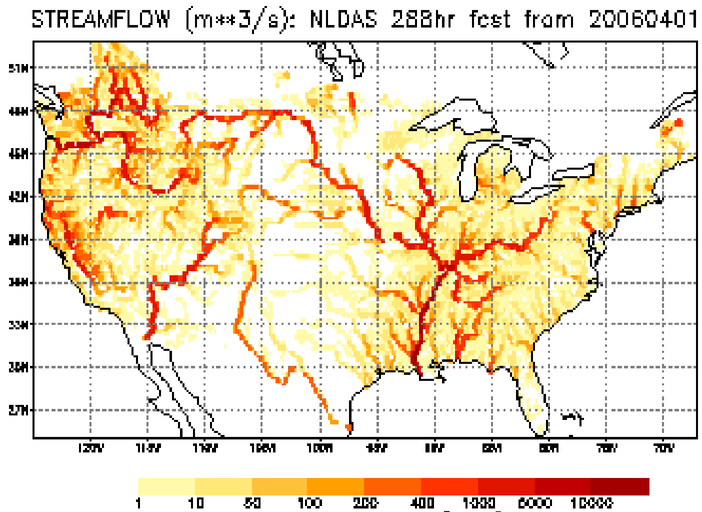
Stream Flow, Analysis and Ensemble Mean Forecast
Error of Ensemble Mean and Ensemble Spread
Forecast Starting at 00Z, April 1st, 2006. lead time **12 days**

Ensemble mean is similar to the Analysis;

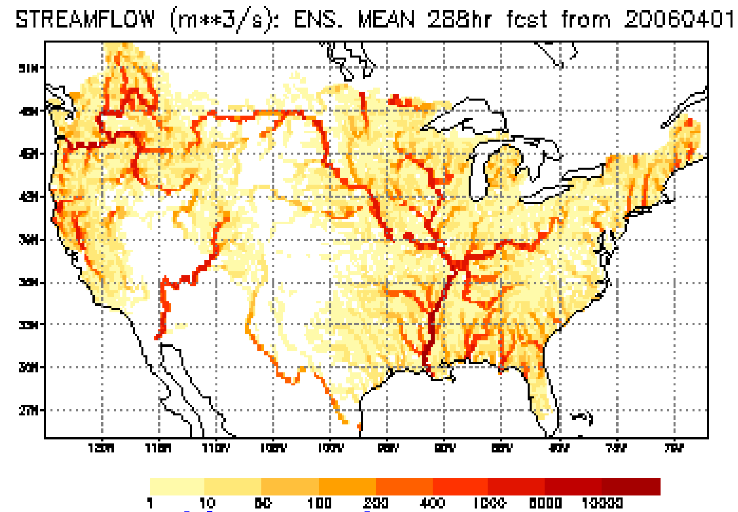
Geographic distribution of positive and negative errors.

Note the scales for error and spread is 1/10 of that for analysis and Ensemble mean.

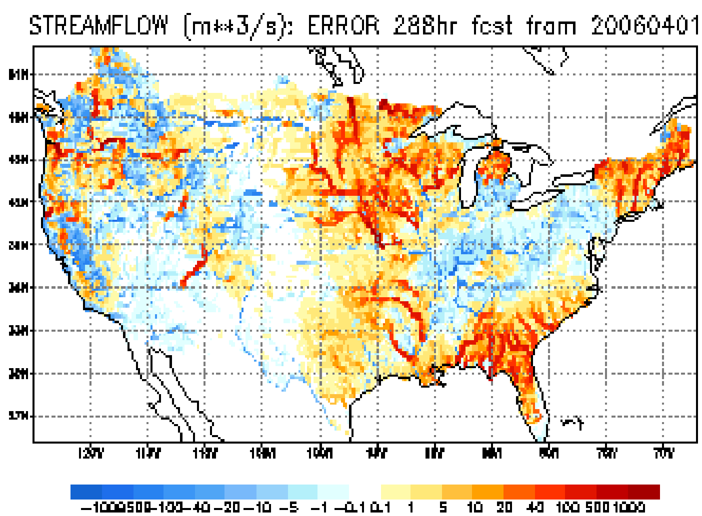
Analysis (NLDAS)



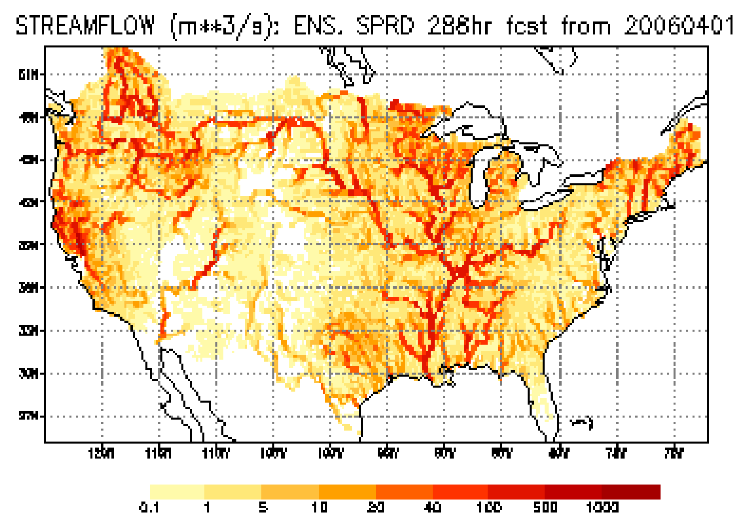
Ensemble Mean



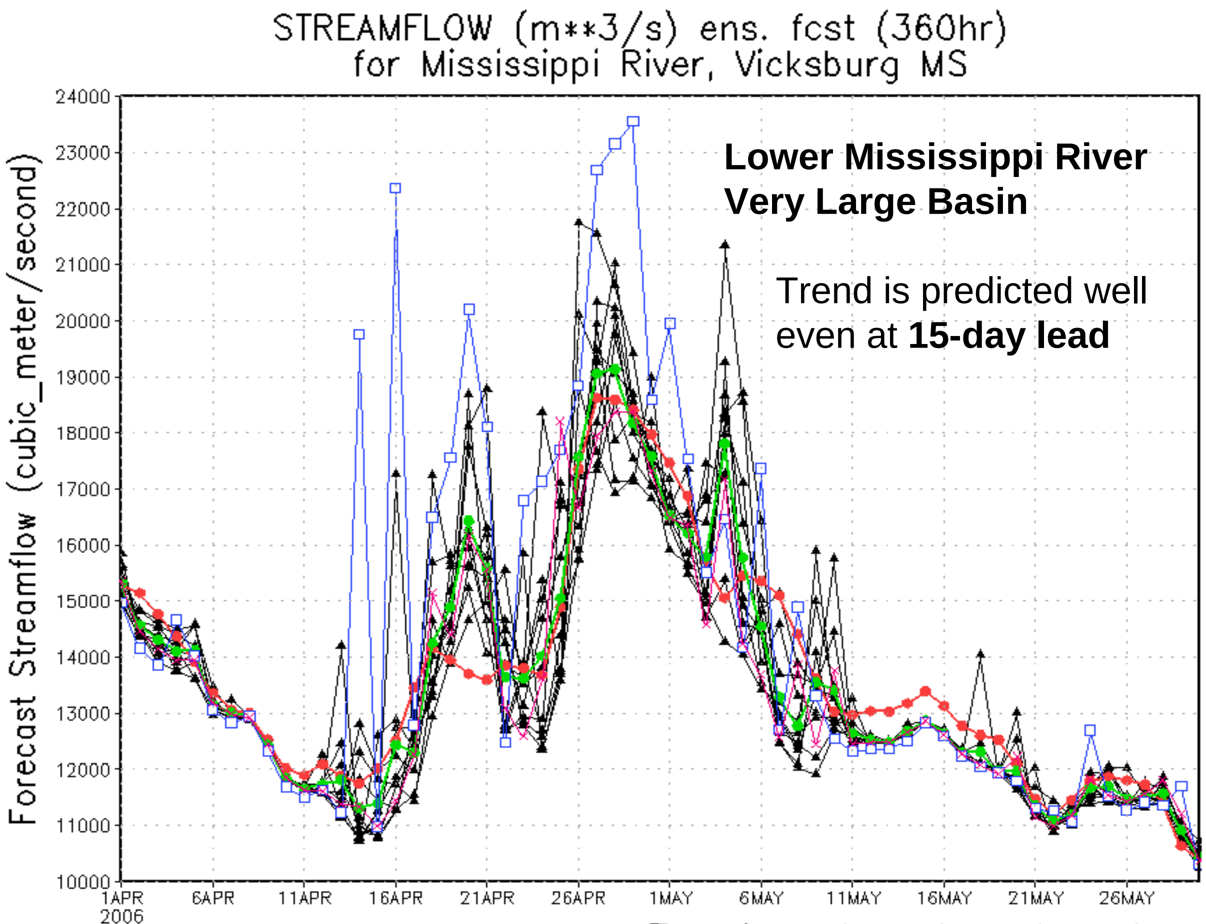
Error=Ens. Mean - analysis



Ensemble Spread



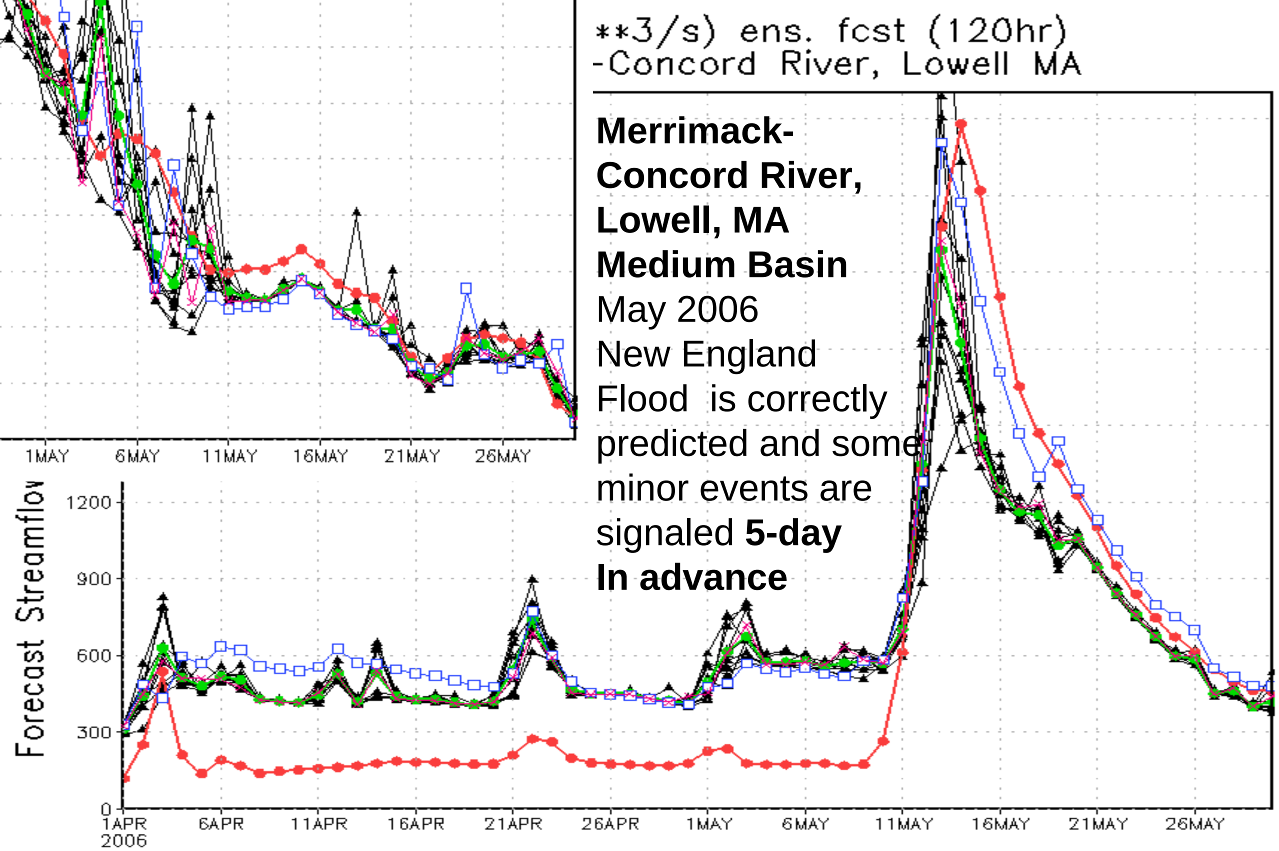
Time Series of Forecasts and Analysis



Lower Mississippi River
Very Large Basin

Trend is predicted well even at **15-day lead**

Positive Correlation between Forecasts and Analysis for all Lead times

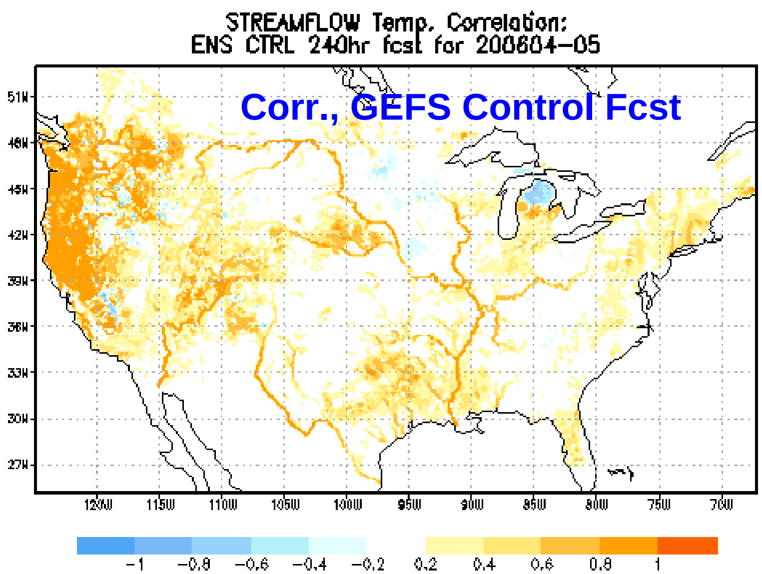


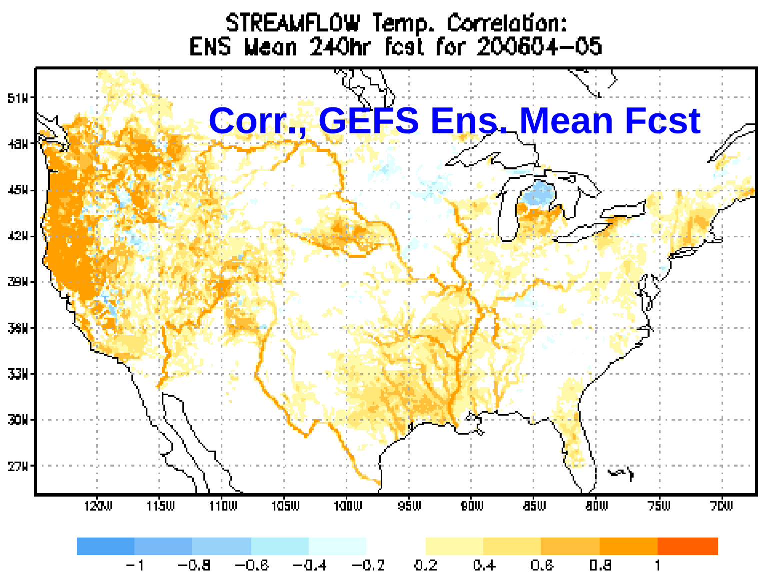
Merrimack-Concord River, Lowell, MA
Medium Basin
May 2006 New England Flood is correctly predicted and some minor events are signaled **5-day In advance**

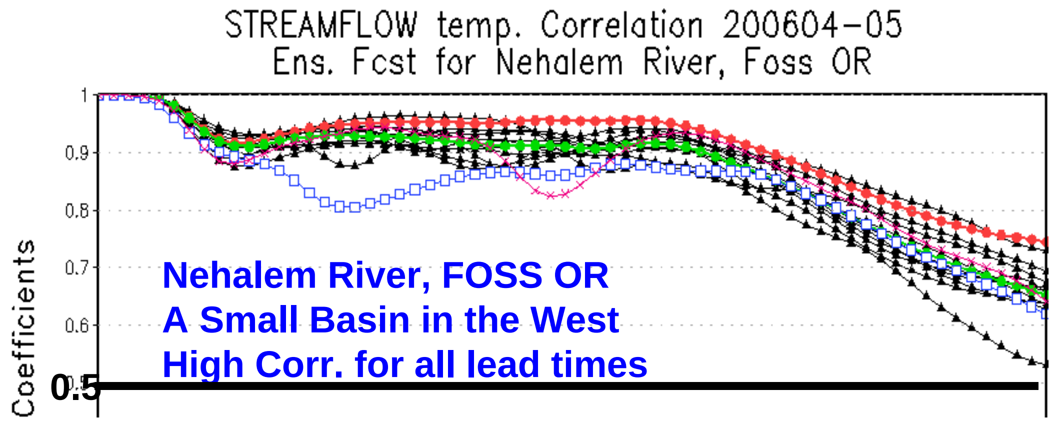
----- GEFS members
----- GEFS ens. mean
----- GEFS control
----- GFS high resolution
----- NLDAS Analysis

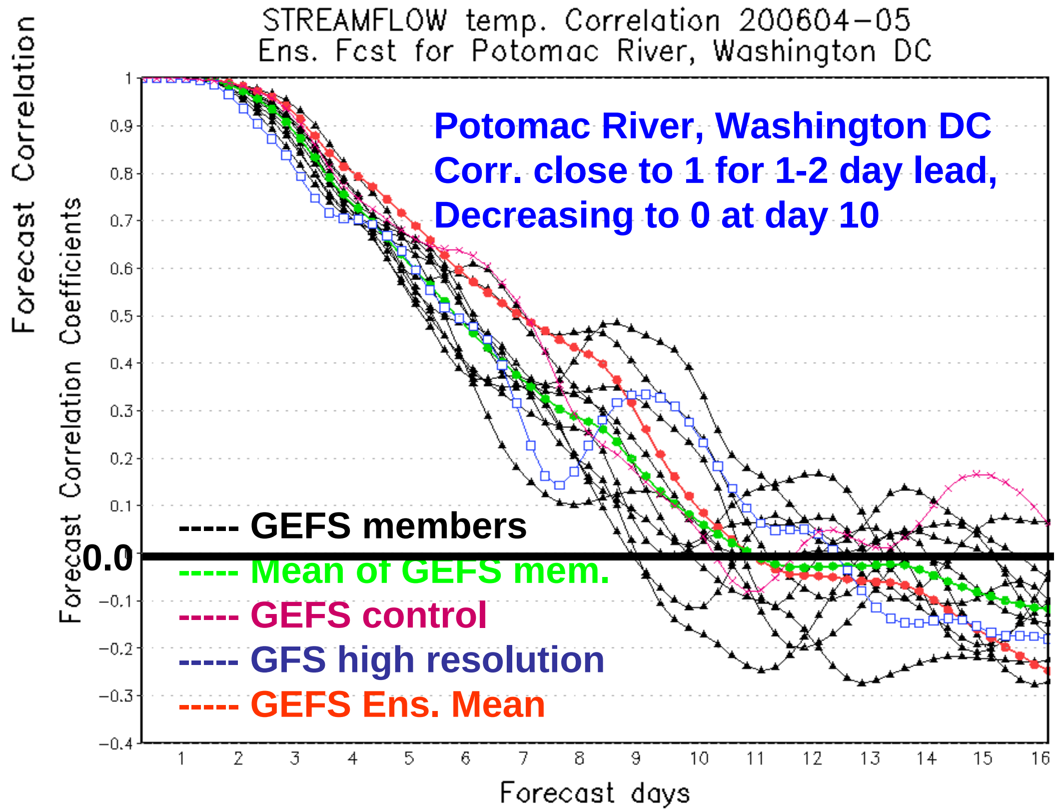
Temporal Correlation

Between Forecasts and Analysis





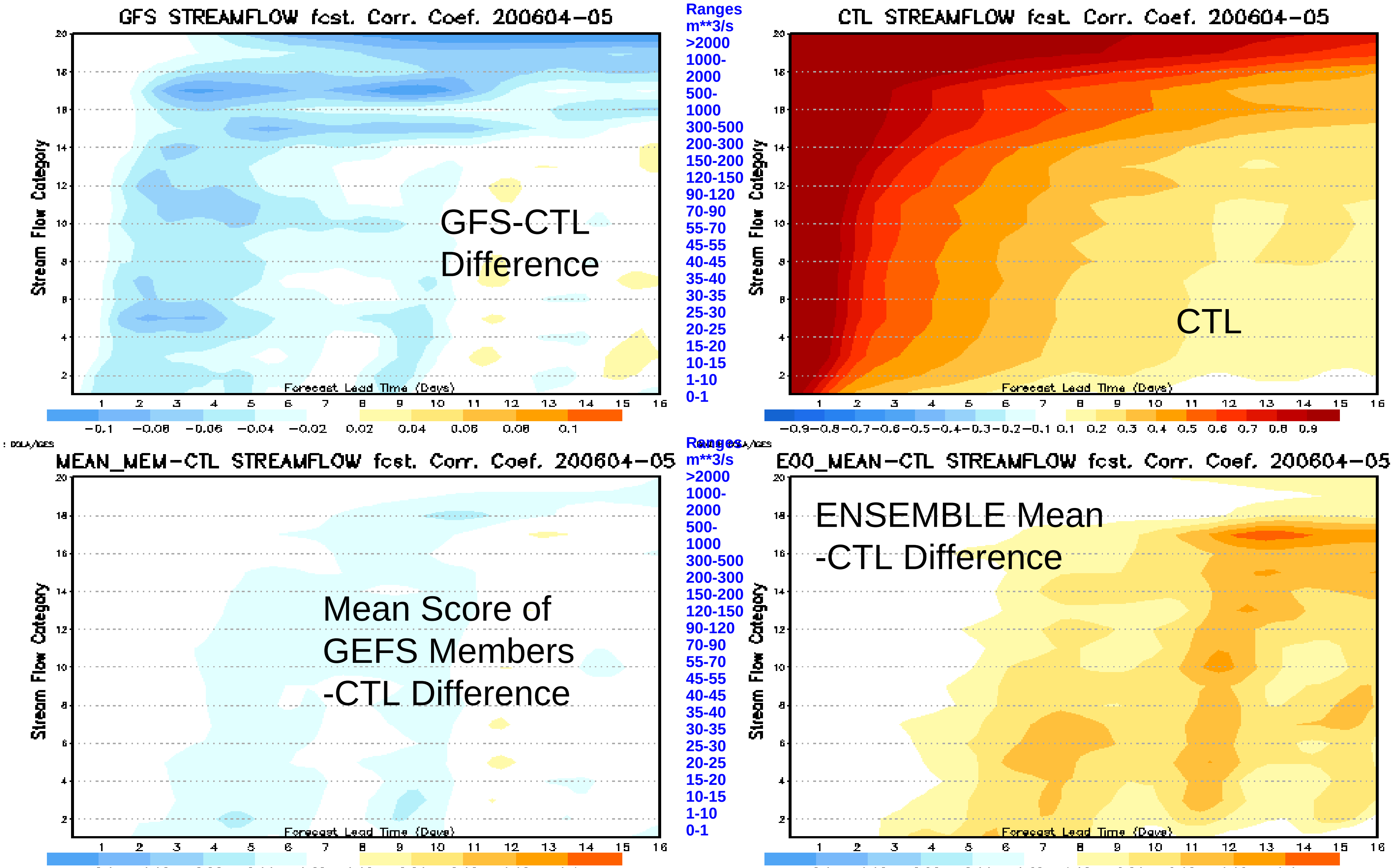




Correlation Coefficient as Function of Lead Time and Mean Flow

The high resolution GFS forecast has lower correlation, especially for day 2-5 over small basins and for week 2 forecast over largest basins.

Major Improvement due to ensemble approach.



CRPSS

- **Continuous Ranked Probability Score (CRPS)**
  - The integral of the Brier scores at all possible threshold values for a continuous predictand (Hersbach 2000; Toth et al. 2003)
  - Averaged over the test data
  - Reduces to **Mean Absolute Error (MAE)** for a single value (deterministic) forecast.

- **CRPS** is calculated for
  - GFS high resolution (single) forecast
  - GEFS control (single) forecast
  - GEFS 10-member mean (deterministic-style) forecast
  - Probabilistic forecast based on GEFS 10 member ensemble

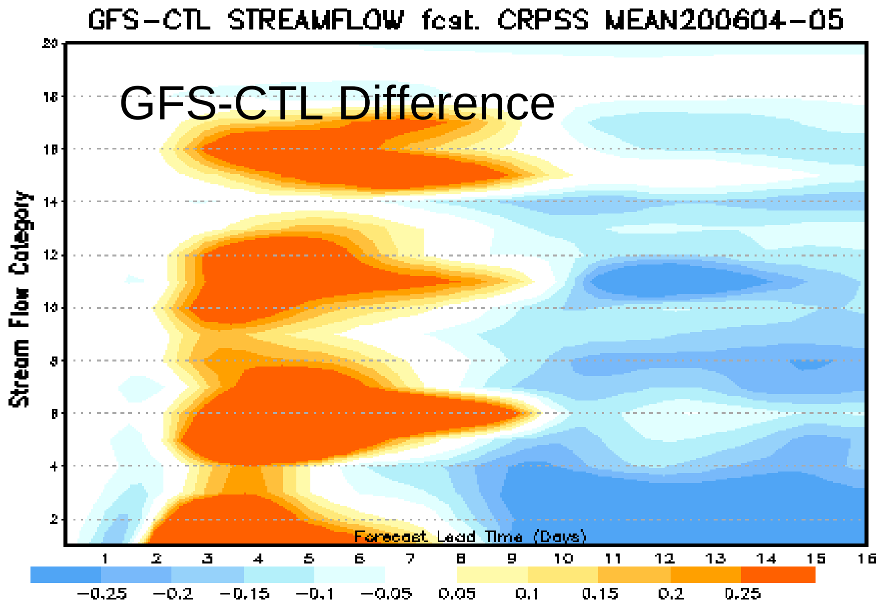
- **Continuous Ranked Probability Skill Score CRPSS=1-CRPS/CRPS_ref**
  - Reference forecast: persistent forecast (forecast=initial)
    - Not the best choice. Generating forecast without precip. Forcing is an alternative
  - CRPSS is less or equal to 1.0
  - <=0, no skill compared with reference forecast
  - >0, some skill over the reference forecast

CRPSS as a Function of Lead Time and Mean Flow, Raw Forecasts

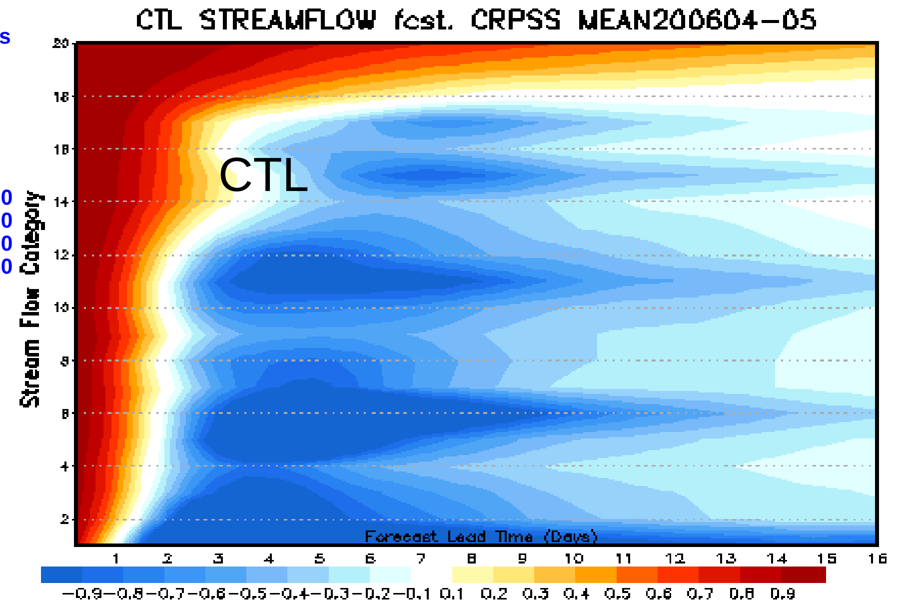
Slight Improvement due to ensemble approach

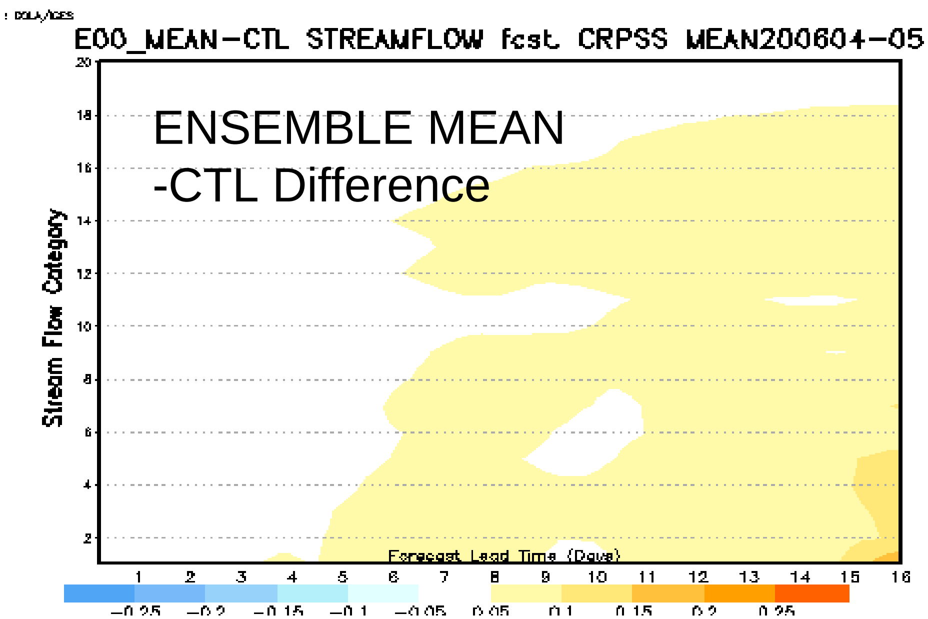
Major Improvement due to probabilistic forecast

High resolution GFS is superior for 2-8 day lead

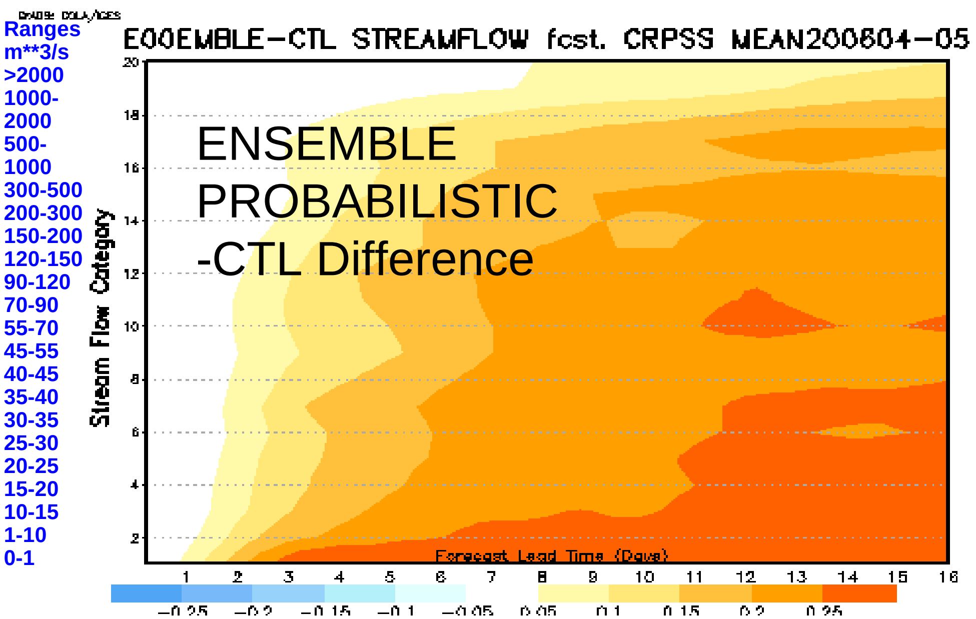


Ranges
m**3/s
>2000
1000-2000
500-1000
300-500
200-300
150-200
120-150
90-120
70-90
55-70
45-55
40-45
35-40
30-35
25-30
20-25
15-20
10-15
1-10
0-1





Ranges
m**3/s
>2000
1000-2000
500-1000
300-500
200-300
150-200
120-150
90-120
70-90
55-70
45-55
40-45
35-40
30-35
25-30
20-25
15-20
10-15
1-10
0-1

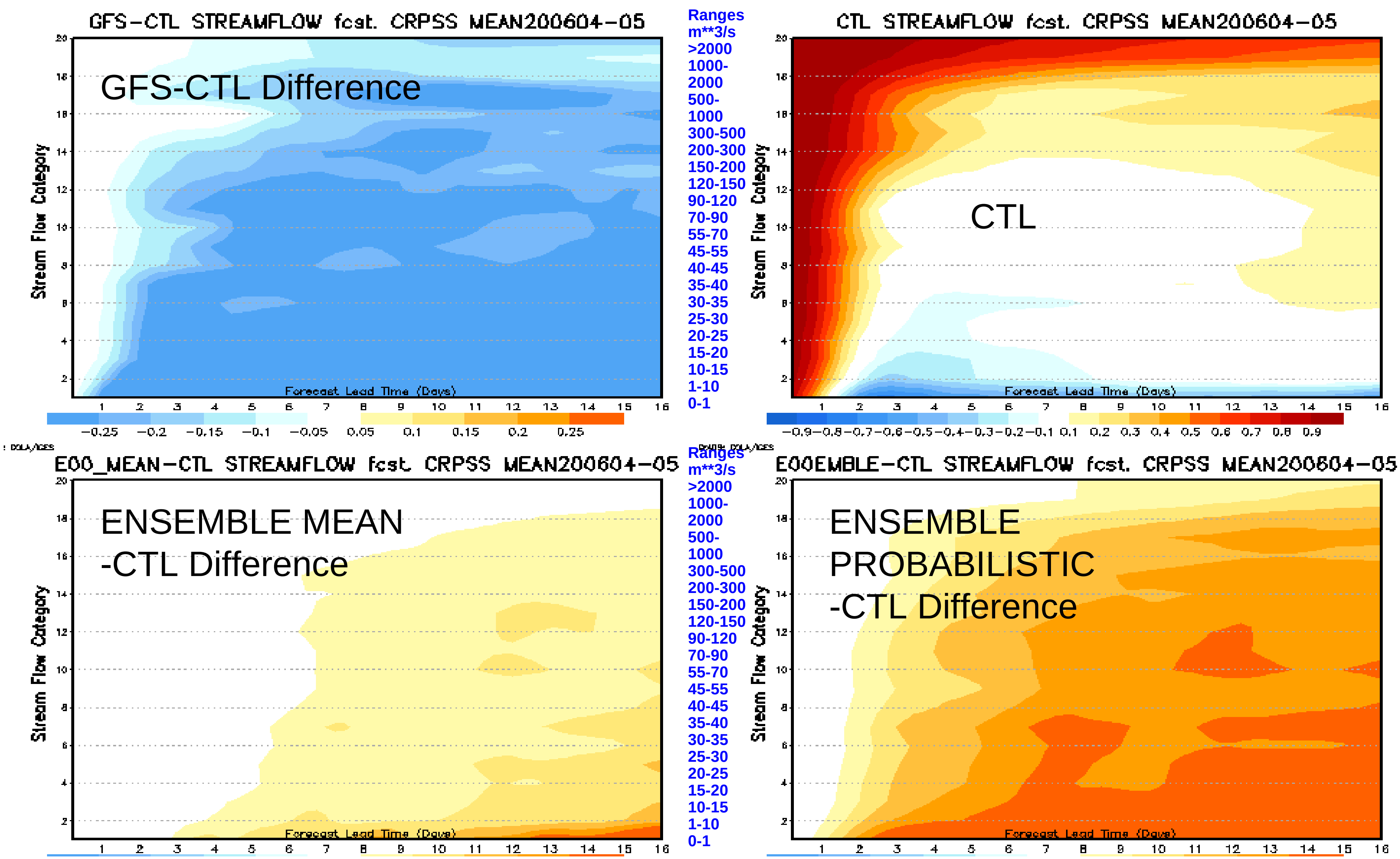


CRPSS as Function of Lead Time and Mean Flow, After Bias-reduction

(Using dependent training data set, not a practical bias correction)

Slight/major Improvement due to ensemble approach/probabilistic forecast

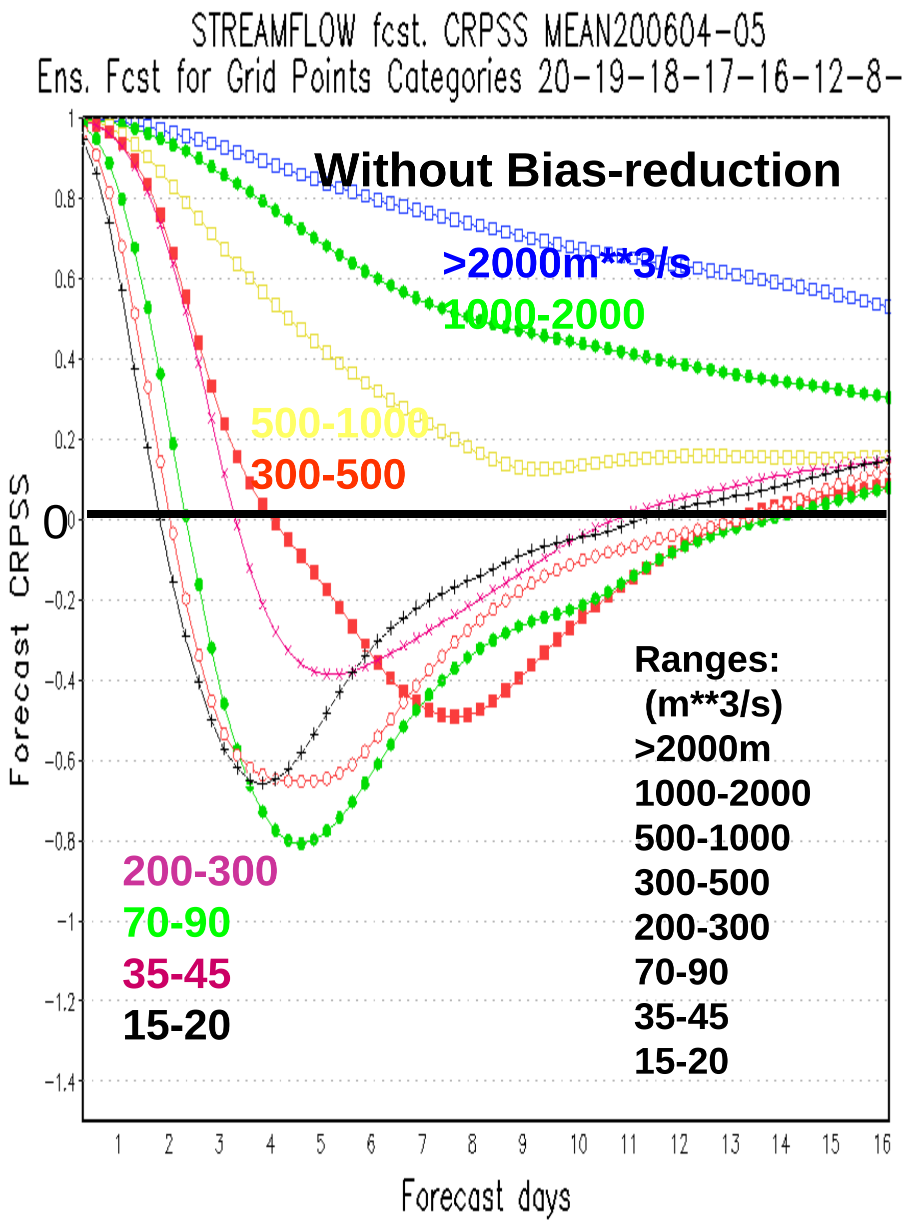
High resolution GFS is not as good as the ensemble control

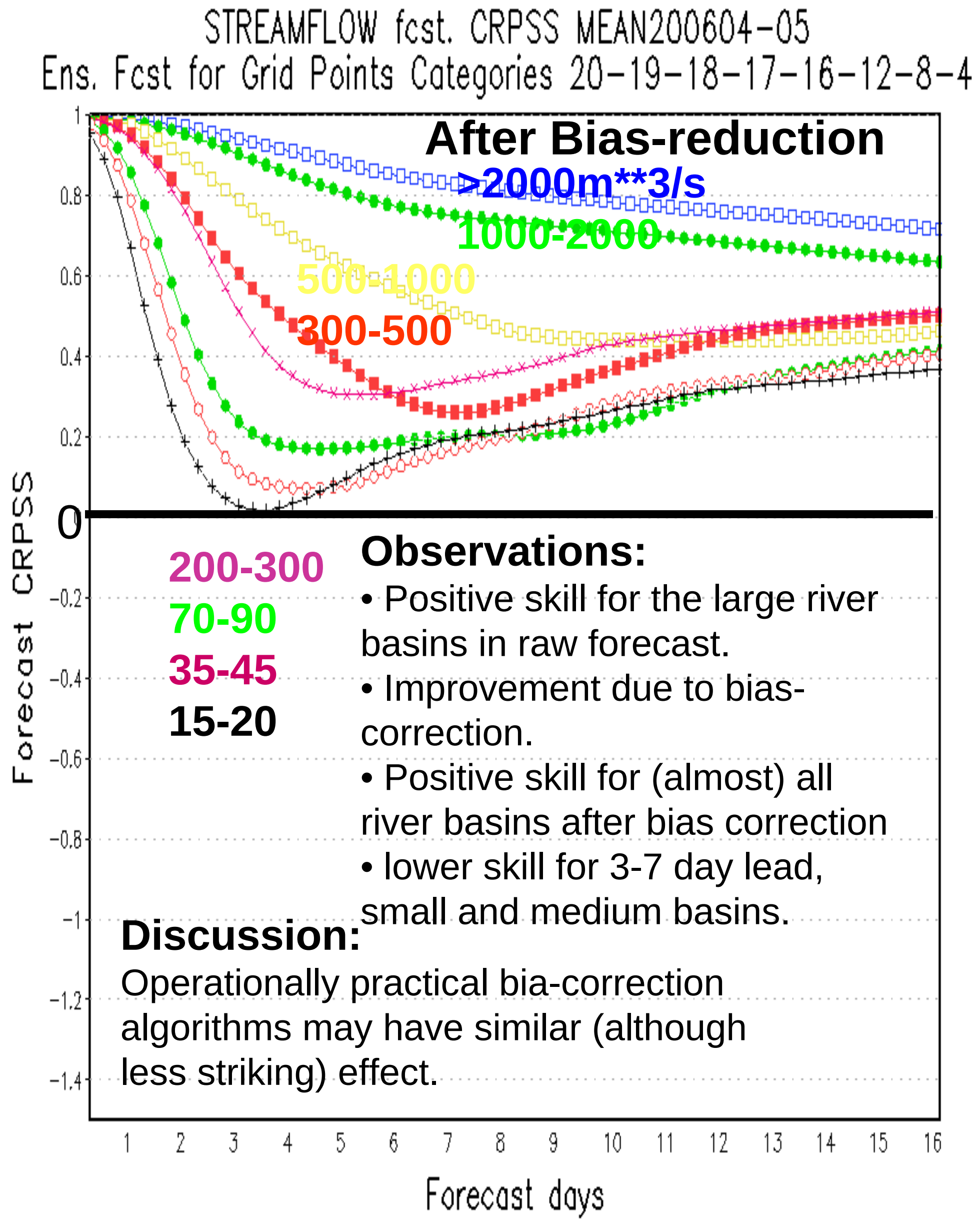


Effect of Bias Correction

CRPSS of The Ensemble Based Probabilistic Forecast

(Averaged over selected ranges of mean Stream Flow)





CRPSS

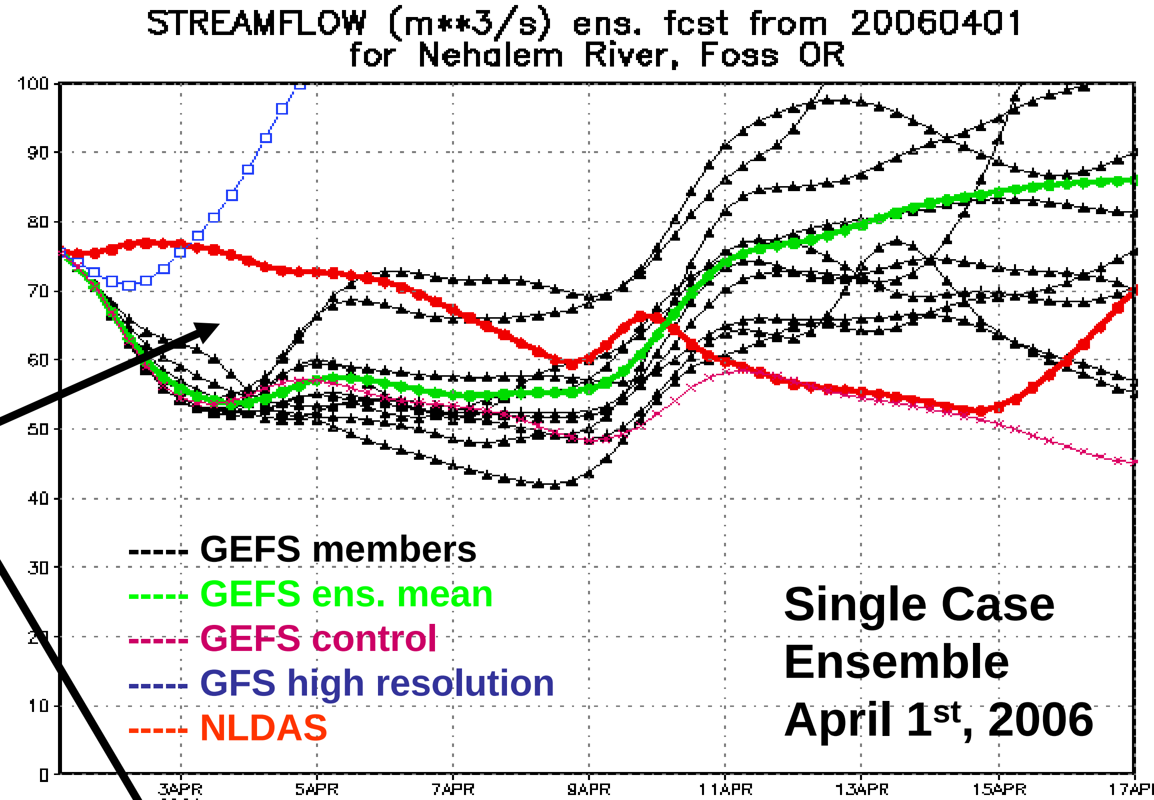
Lack of skill for small and medium basins with 3-7 days lead, even after bias correction

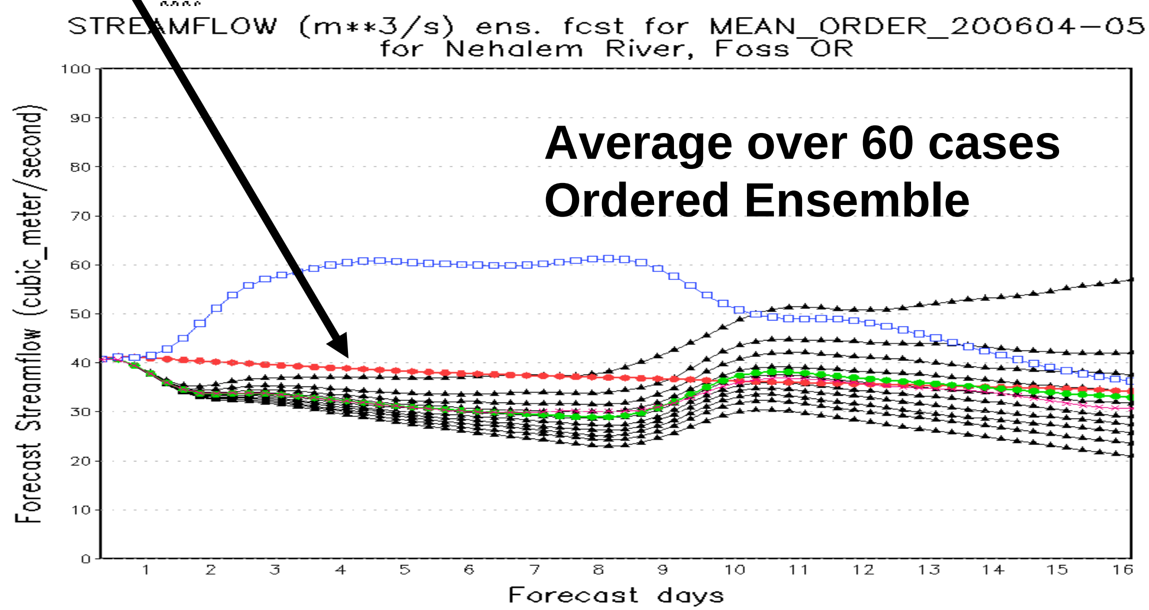
Possible explanation:
Bias and insufficient spread in the streamflow forecast due to deficiencies in the forcing (precipitation and/or runoff forecast) generated by the GEFS system

- Bias
- Insufficient spread on grid and subgrid scales.
- Spatial and temporal resolution

Possible Solutions:

- Downscaling of precipitation/runoff;
- Bias correction of precipitation/runoff.





Conclusions and Discussions

1. Distributed river routing system (coupled GEFS, NOAH and the Lohmann River Flow model) generates reasonable gridded river flow forecast.
   – The coupled GFS-Noah system provides reasonable forcing to the river routing model

2. The ensemble approach, especially the ensemble-based probabilistic forecast, improves the forecast skill significantly.
   – Ensemble spread is comparable to the forecast error in first moment

3. Large basin forecasts are more skillful with higher correlation and positive CRPSS for all lead times up to 16 days.
   – GEFS provides reasonable forcing

4. Medium/small basin forecasts, especially for short to medium lead time, suffer from underdispersion (insufficient spread).
   – Downscaling of hydro-meteorological forcing is needed.
   – Forecast can be improved and calibrated through bias correction.

5. For the small and medium basins at lead time of 2-7 days, the high resolution GFS forecast is superior to the lower resolution runs in that it has smaller bias, but this is balanced by lower forecast-analysis correlation.
   – The GEFS ensemble, with suitable post processing, can outperform higher resolution single forecast

Further Development Plan

Evaluation

- Using actual USGS streamflow observations at unregulated basins. (Ohio River, in corporation with Ohio RFC)
- Corporation initiatives from other RFCs welcome

Configuration

- Expand to global domain (at 0.5 degree resolution)

Improvement of the Forcing (precipitation/runoff)

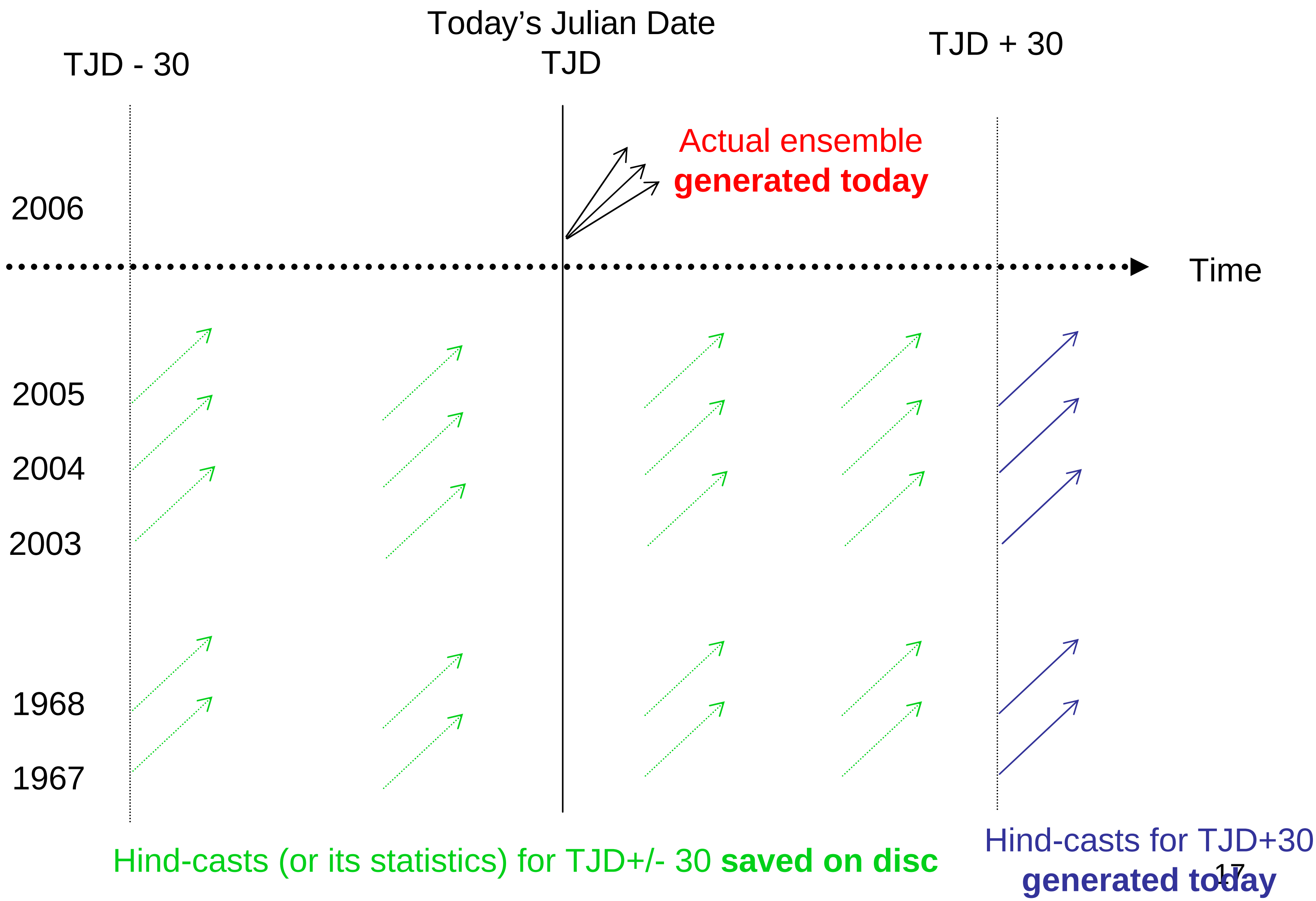
- Bias correction
- Downscaling

Calibration of the Product

(post-processing of streamflow forecast)

- Bias correction to the streamflow output for better product.
- Generate a hind-cast data set for a better estimate of bias.

REAL-TIME GENERATION OF HIND-CAST DATASET?



Thank You!

Background

Introduction: Background

- River routing experiment in analysis mode of the North American Land Data Assimilation System (NLDAS) project (Lohmann et al, 2004) revealed potential benefit of river flow forecast in NWP.

- Coupling of Atmospheric and Land Surface components of NWP systems (Mitchell et al, 2004) facilitates **gridded** stream flow forecast in NWP.

- Existence of uncertainty in initial conditions, model structure and forcing needs to be considered with an ensemble approach.

Configuration and Design of Current Experiment

(Approach A: Two Way coupling)

Experimental Design

- **Period:** April 1st to May 30th, 2006
- **Forecast Cycle:** 00Z
- **Forecast Length:** 384 hours (16days)
- **Domain:** CONUS

Configuration of the NCEP Global Ensemble Forecasting System (GEFS) (operational before May 31st 2006**)**

- **Model:** Two way coupled GFS-Noah
- **Ensemble Size:** 10 Members
- **Ensemble Generation:** Breeding
- **Resolution**: T126L28 for ensemble members and control forecast
  T382L64 (0-180h) and T190L64 (180-384)
  for GFS high resolution forecast (GFS)
- **Output:** Runoff
  1.0 deg. by 1.0 deg. grid, every 6h for ensemble members and control
  0.5 deg. By 0.5 deg. grid, every 6h for GFS high resolution forecast

Forecast Example **(initiated April 1st, 2006)** Stream Flow

Forced by GFS,GEFS Forecast and NLDAS Product

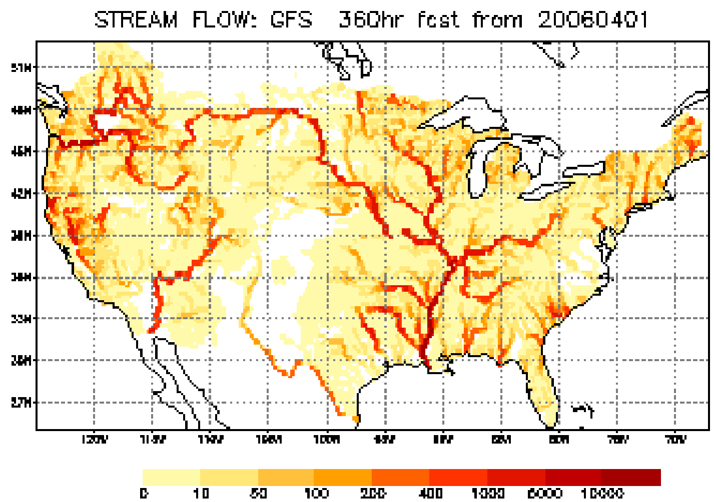
Forecast Starting at 00Z, April 1st, 2006. Lead time **15 days**

Single control forecasts similar to each other;

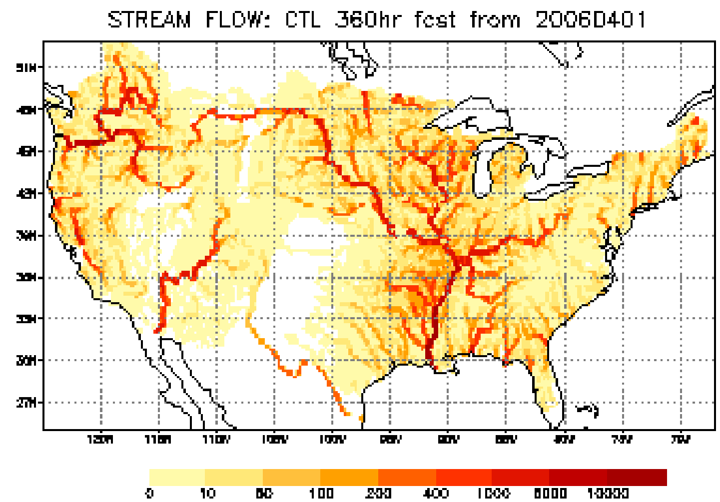
Ensemble mean is similar to the analysis.

This suggests the ensemble mean has its value in stream flow forecast.

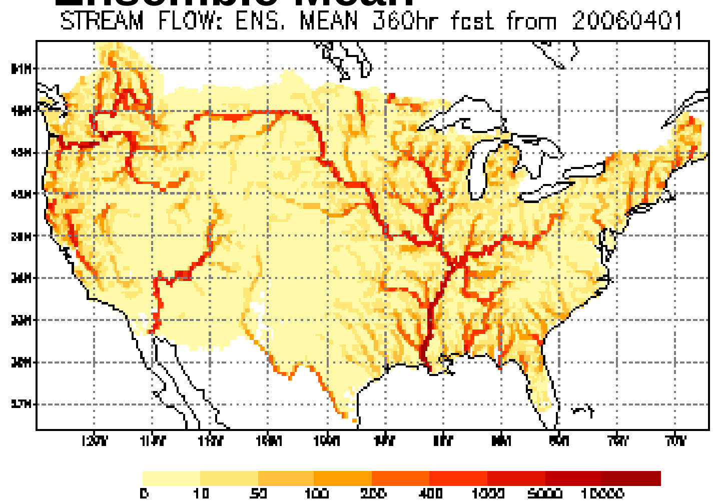
GFS High Res.



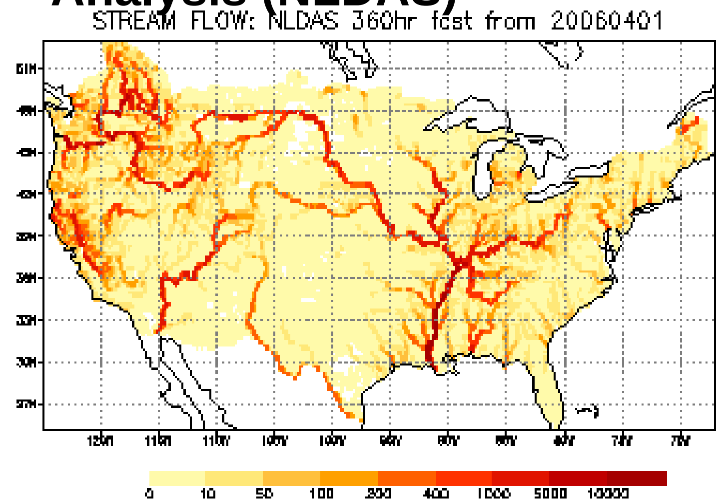
Ensemble (low res.) Control



Ensemble Mean



Analysis (NLDAS)



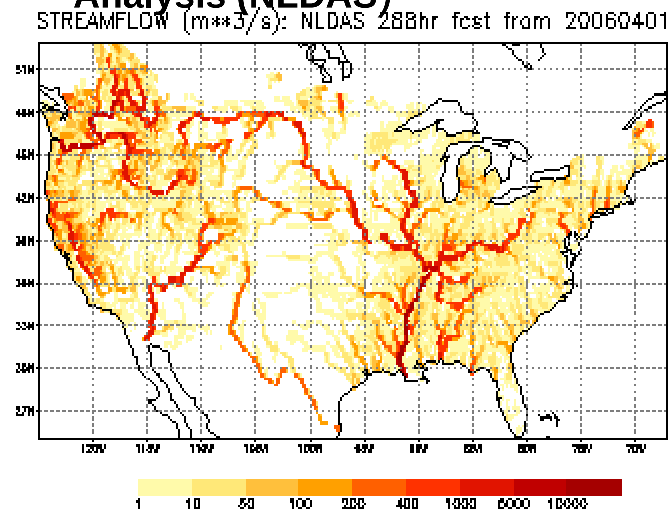
Forecast Example (initiated April 1st, 2006) Stream Flow

Stream Flow, Analysis and Ensemble Mean Forecast
Absolute Error of Ensemble Mean and Ensemble Spread
Forecast Starting at 00Z, April 1st, 2006. lead time **12 days**

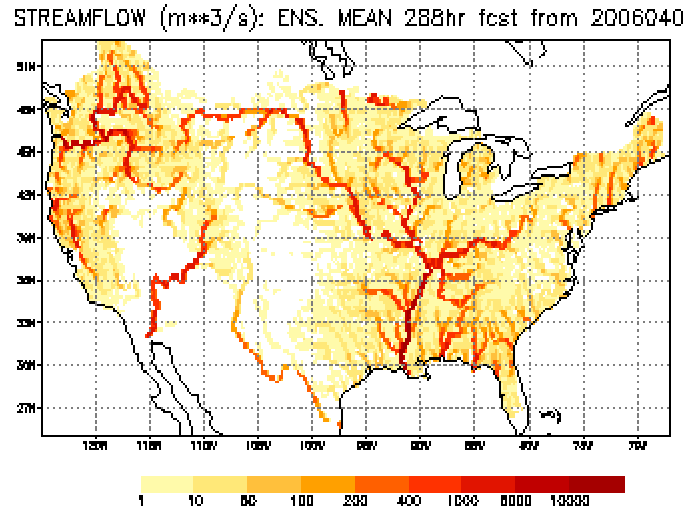
Same as previous slide except for the error, where absolute value is plotted to compared with the spread.

Spread is comparable to error, but the value is smaller, especially in the West.

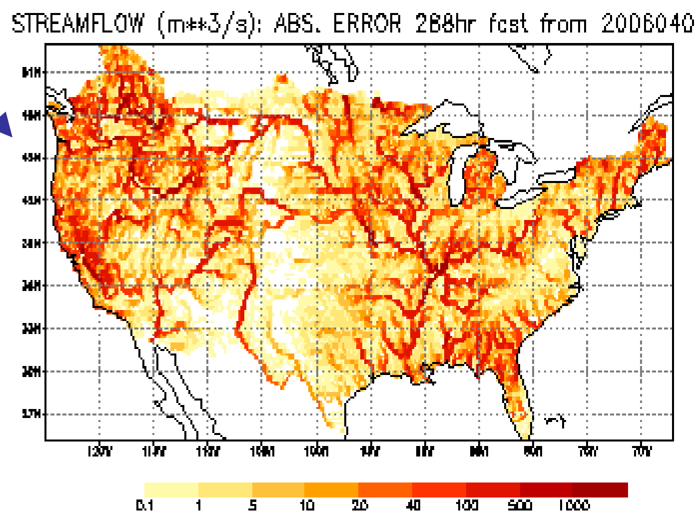
Analysis (NLDAS)



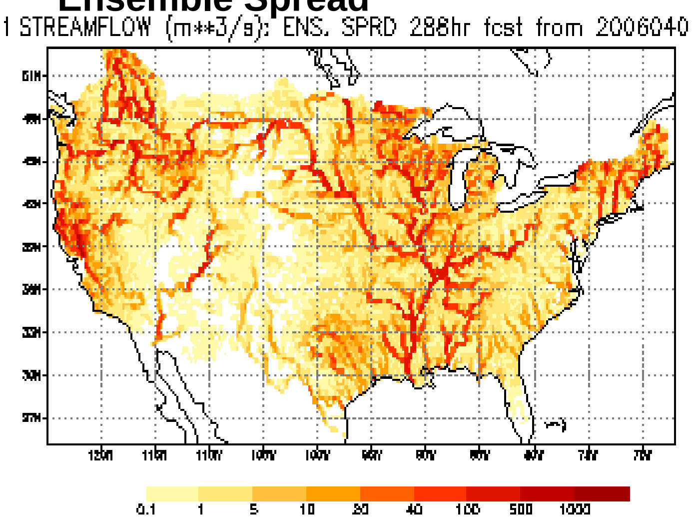
Ensemble Mean



Absolute Error



Ensemble Spread



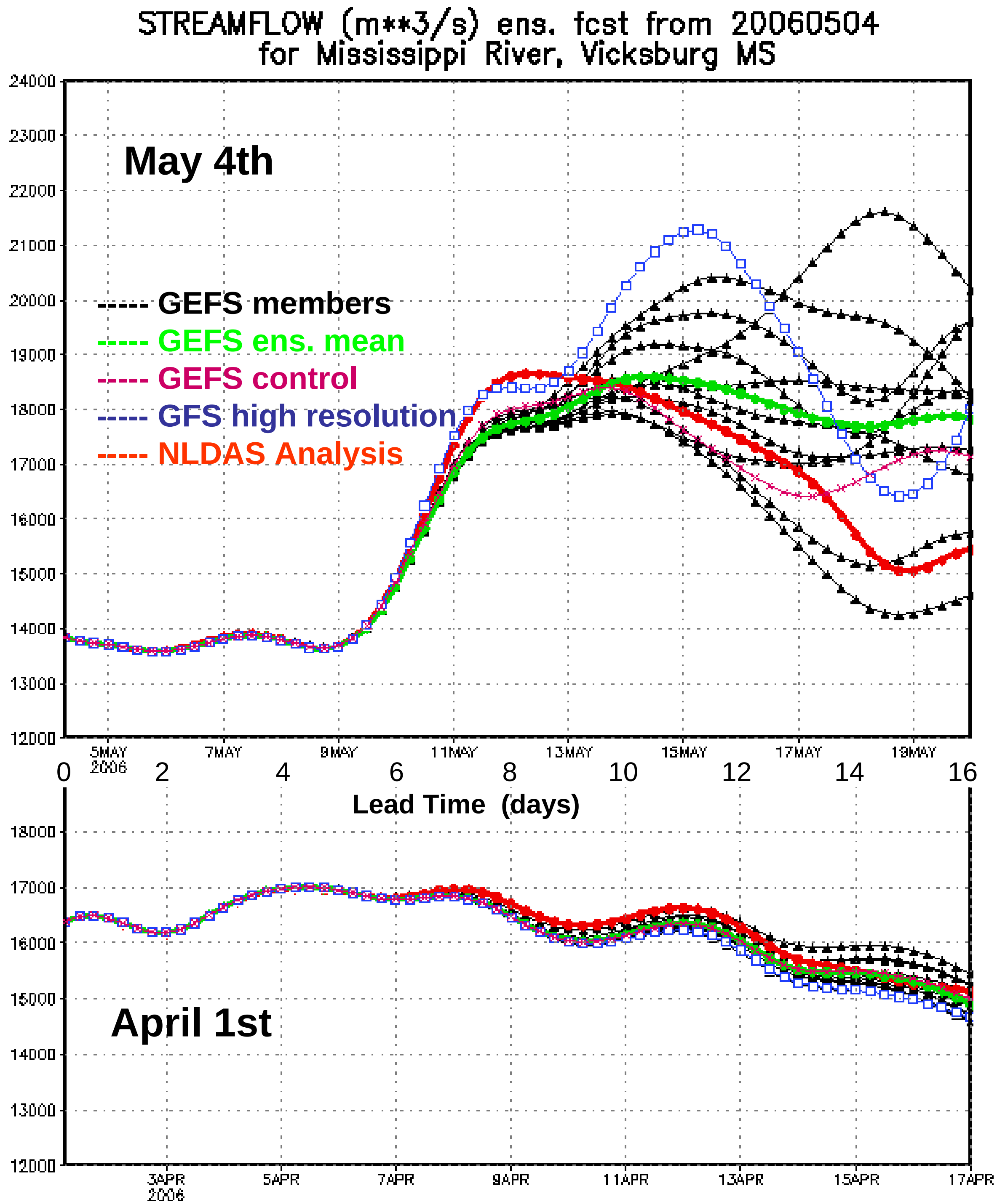
Forecast Examples

Mississippi, River
Vicksburg, MS
The Large Basin

May 4th case
A major mid-range event well predicted; Significant spread in extended range

April 1st case
Without a major event,
all simulations are similar and spread is small.

Trend and events picked up.
Short lead time dominated
by initial condition, showing little spread.
Spread Increases with time.

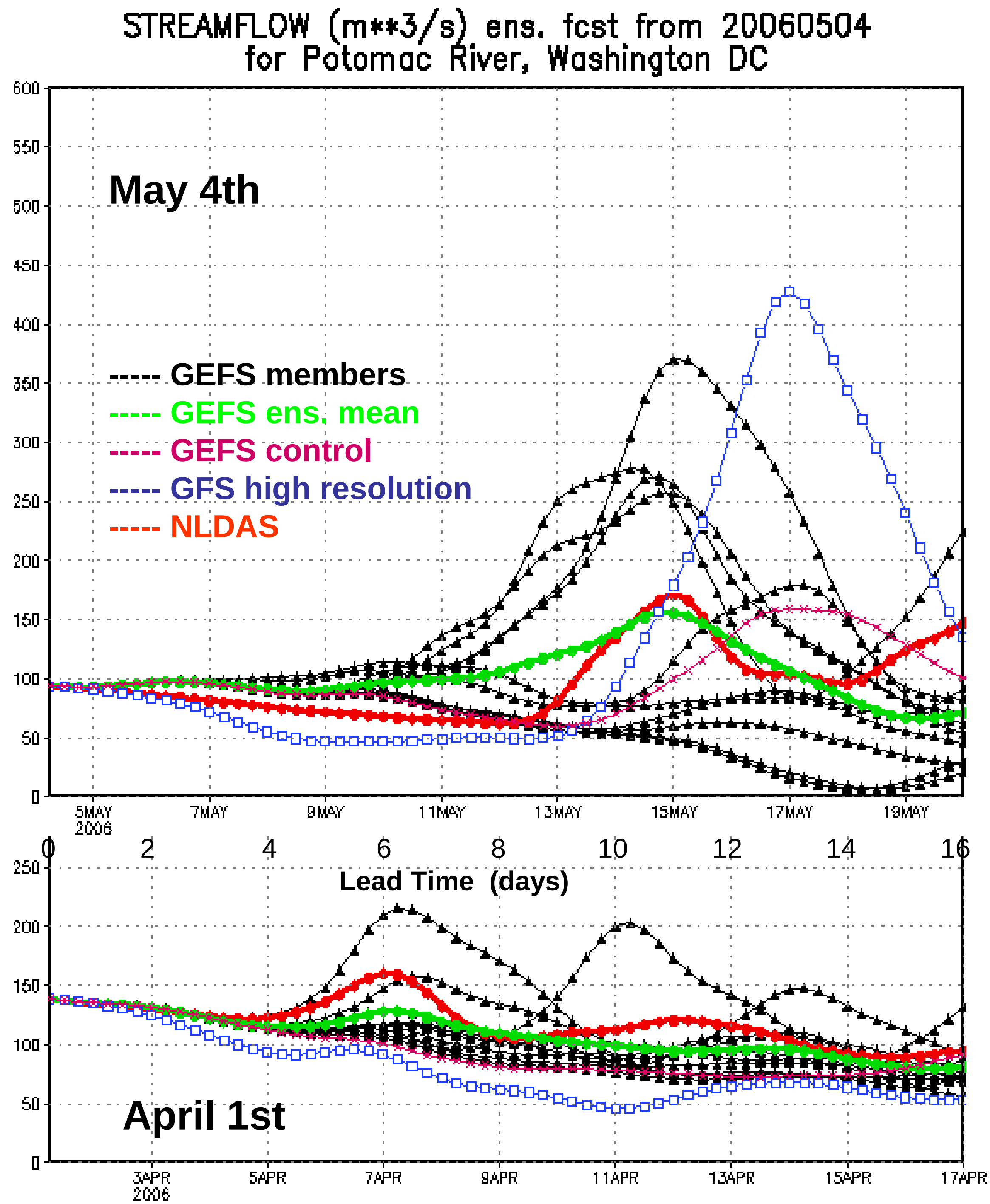


Potomac River
A Medium Sized Basin

In both cases

Single forecasts are insufficient.

Non-linear evolution of ensemble members help to improve forecast and catch major flood events.

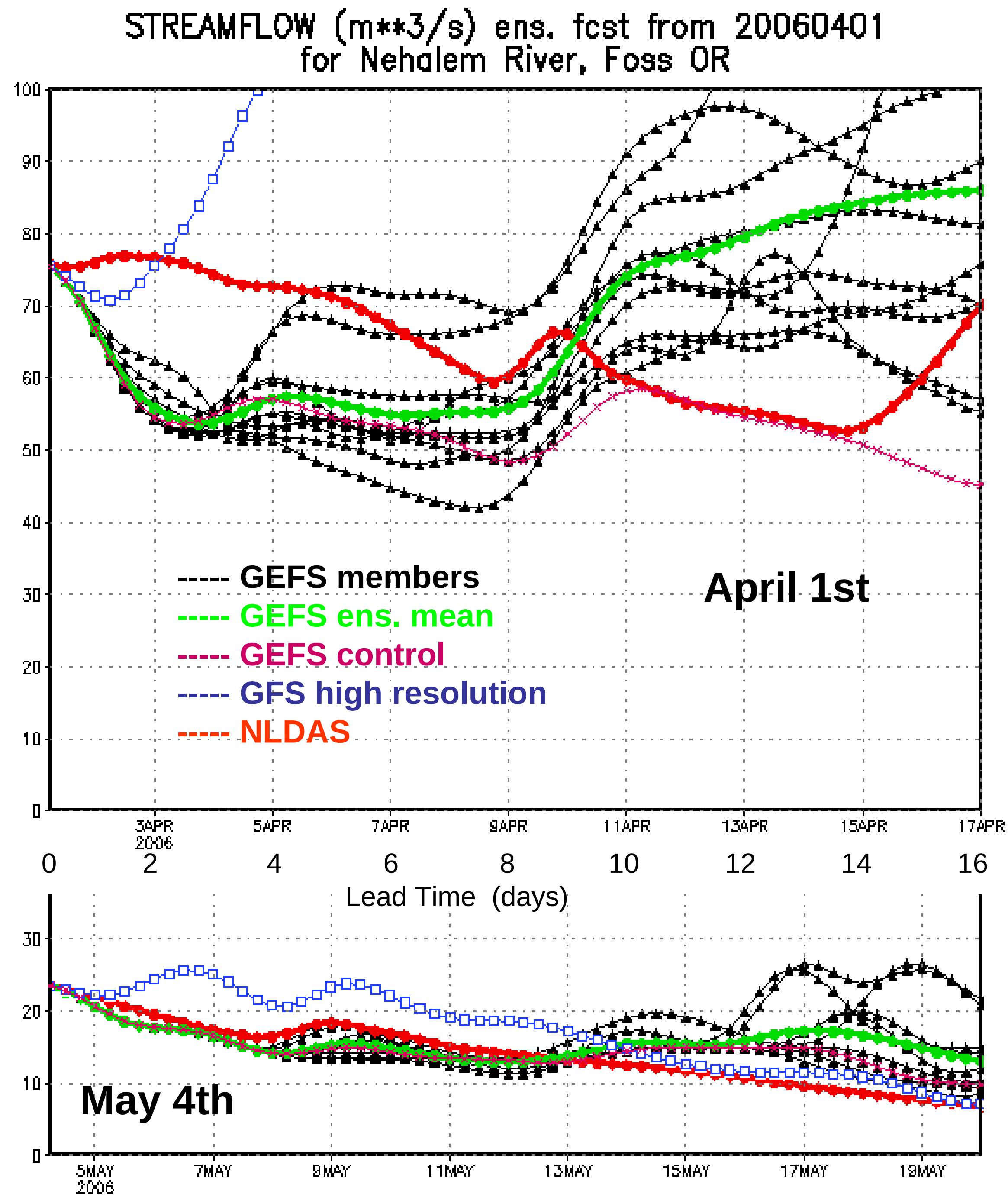


Nehalem River, FOSS OR
A Small Basin

A challenge for the models. April 1st, large forecast discrepancy from day 1 despite significant spread

Possible causes of the problem in the short range forecast

- Lack of spread in precip. fcst. on grid and subgrid scale.
- Spatial and temporal resolution of the runoff.
- Bias of precipitation (and runoff) forecast



Merrimack-Concord River
Lowell, MA
A Medium Sized Basin

Major Problem

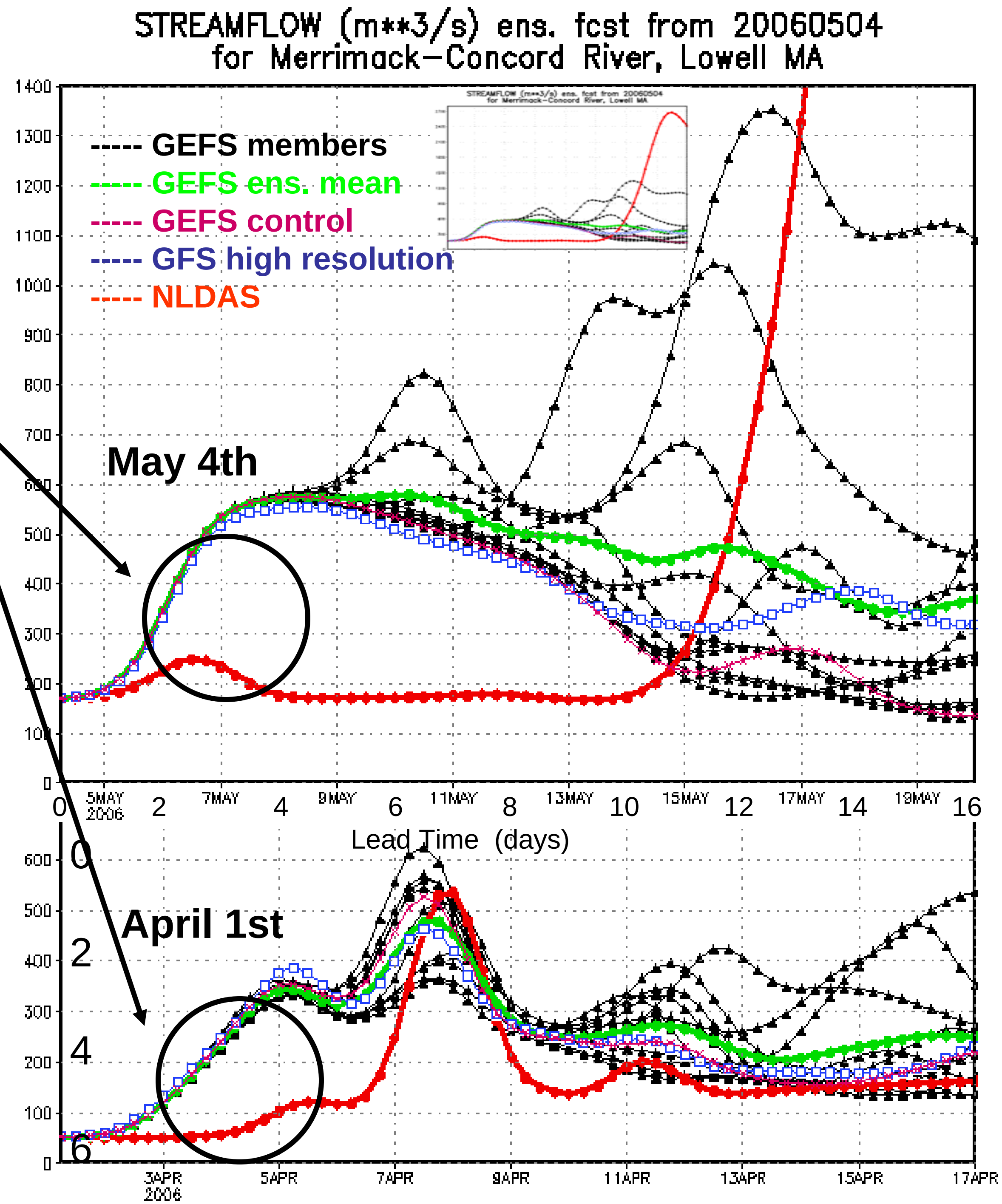
Underdispersive ensemble in grid and subgrid scale precipitation.

Mid-May Flood Event

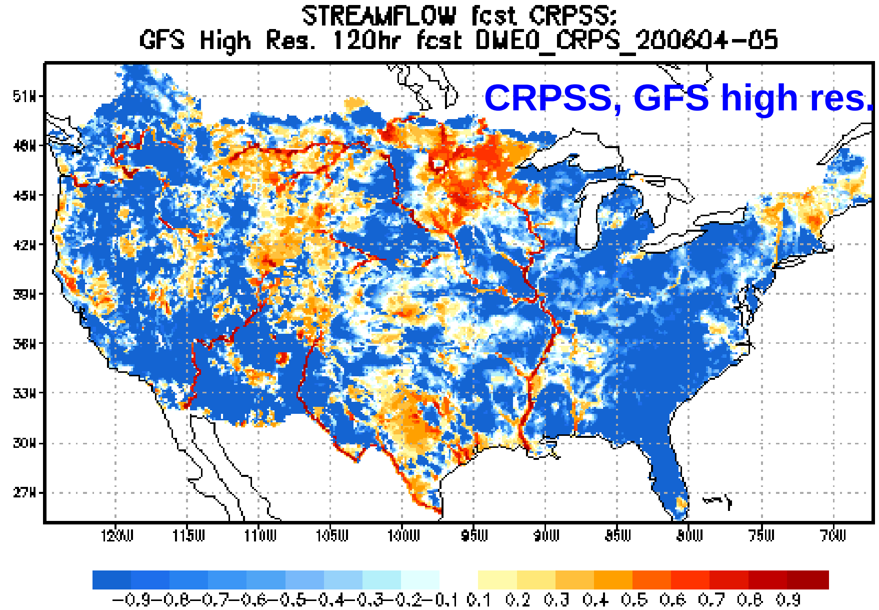
Compared with the Early-April event, the Mid-May event is harder for the model to simulate. Nevertheless, the ensemble shows some skill indicating a major event with 10+ day lead, various amplitude and timing.

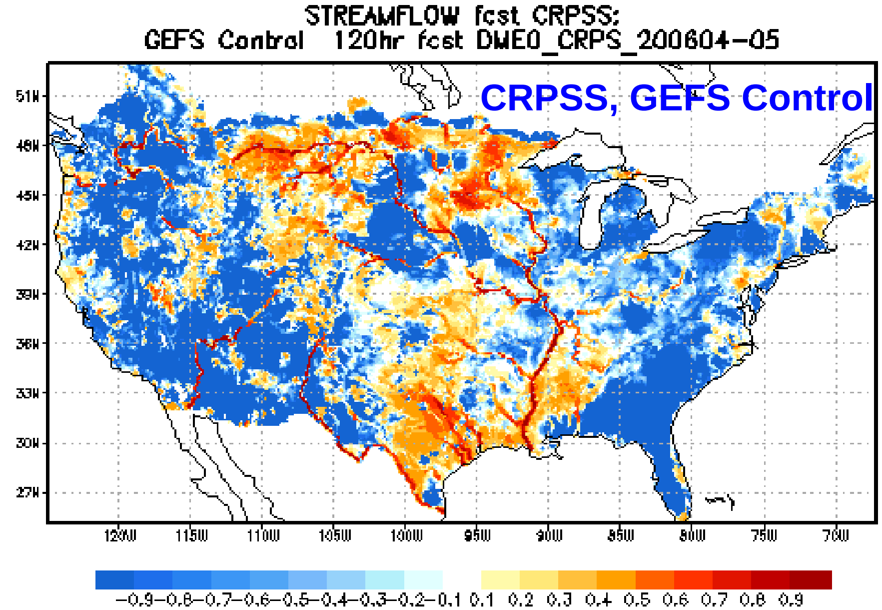
Early April

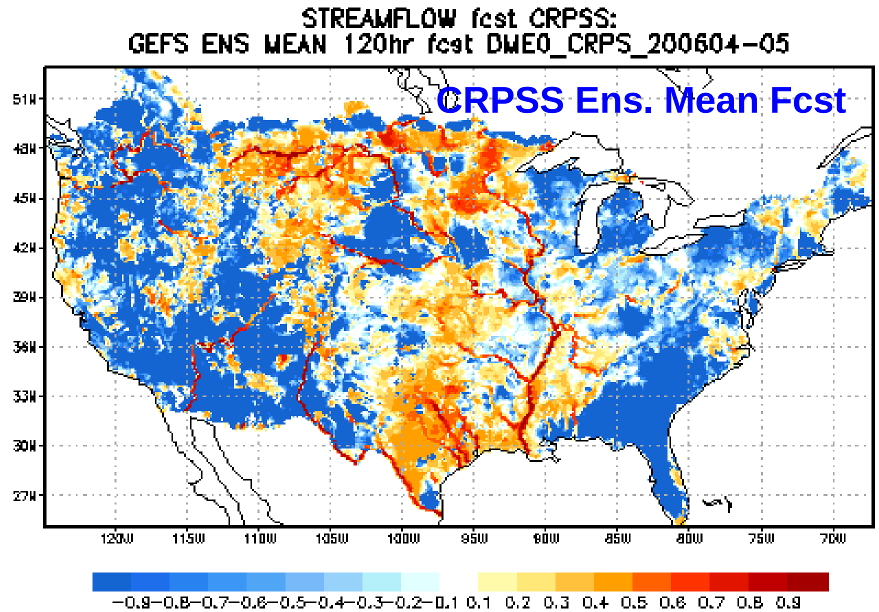
Major event forecast despite short range over-forecast

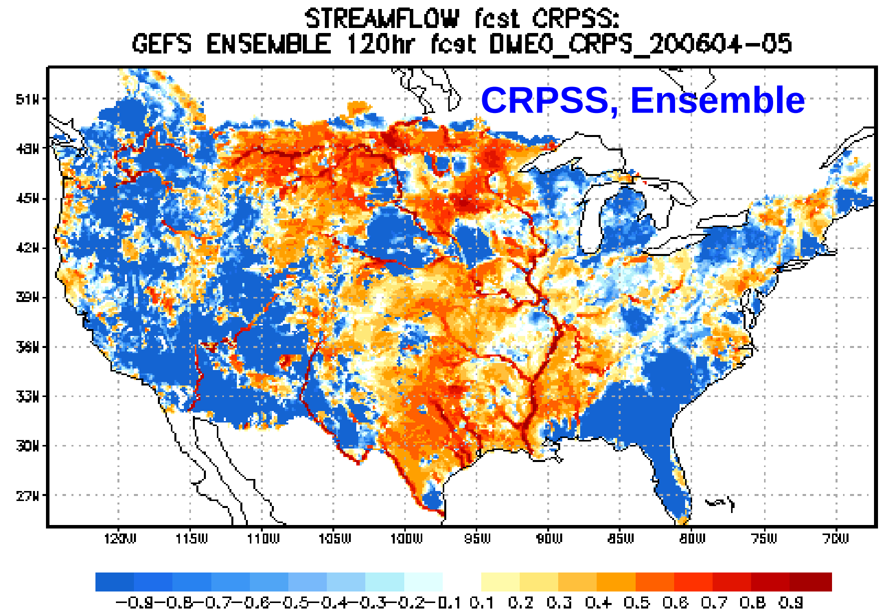


CRPSS of Various Forecasts (lead time 120h)



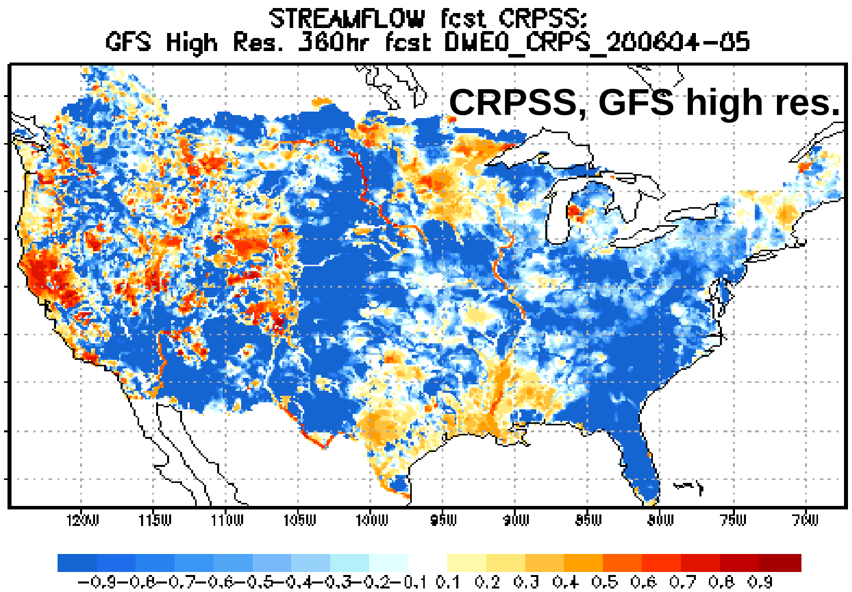


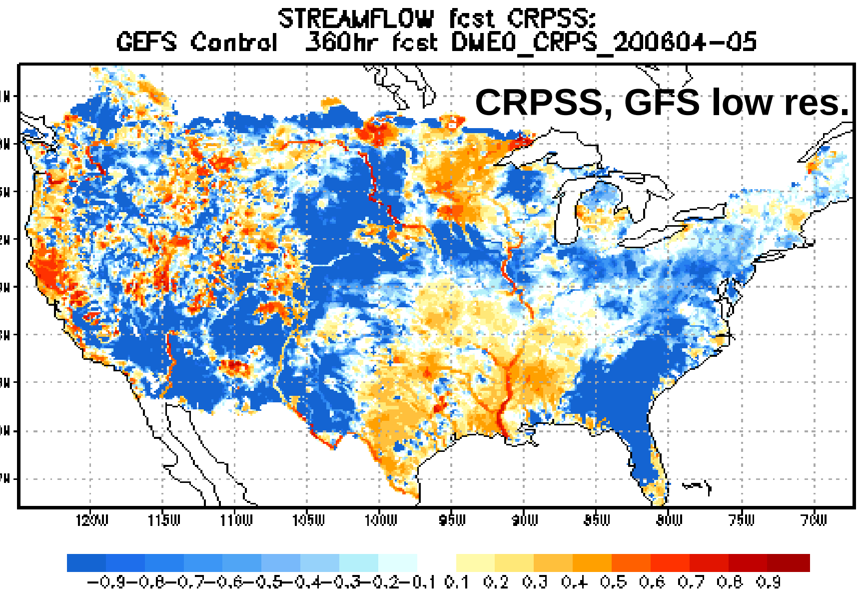


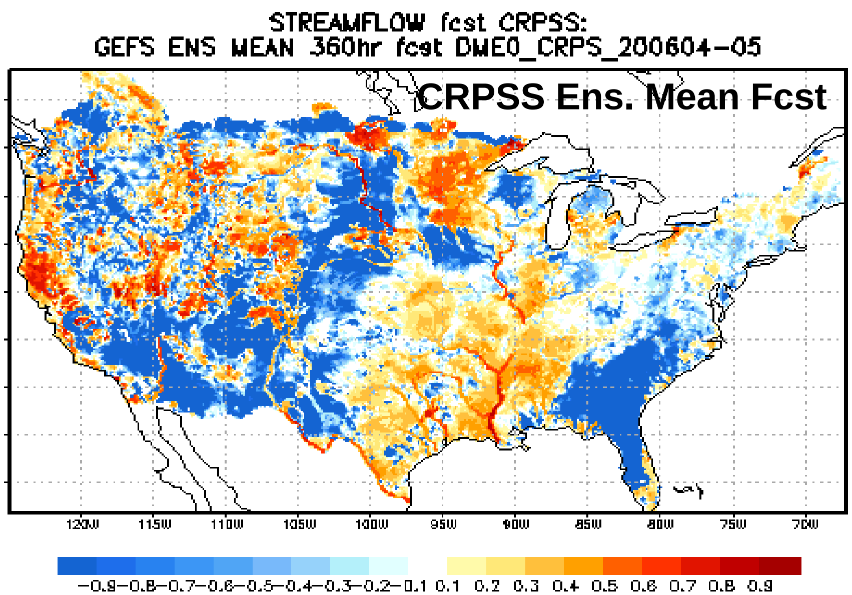


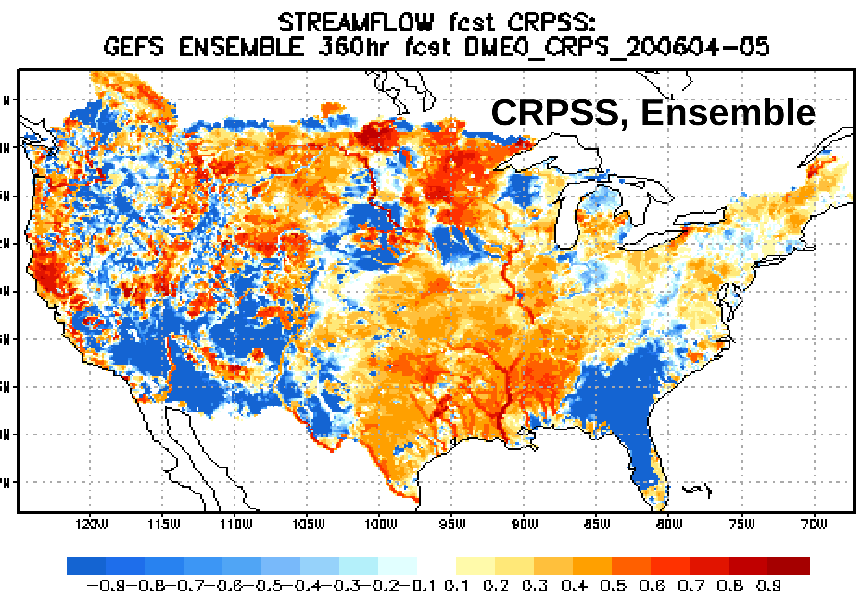
CRPSS of Various Forecasts (lead time 360h)

(After Bias-correction with dependent training period)



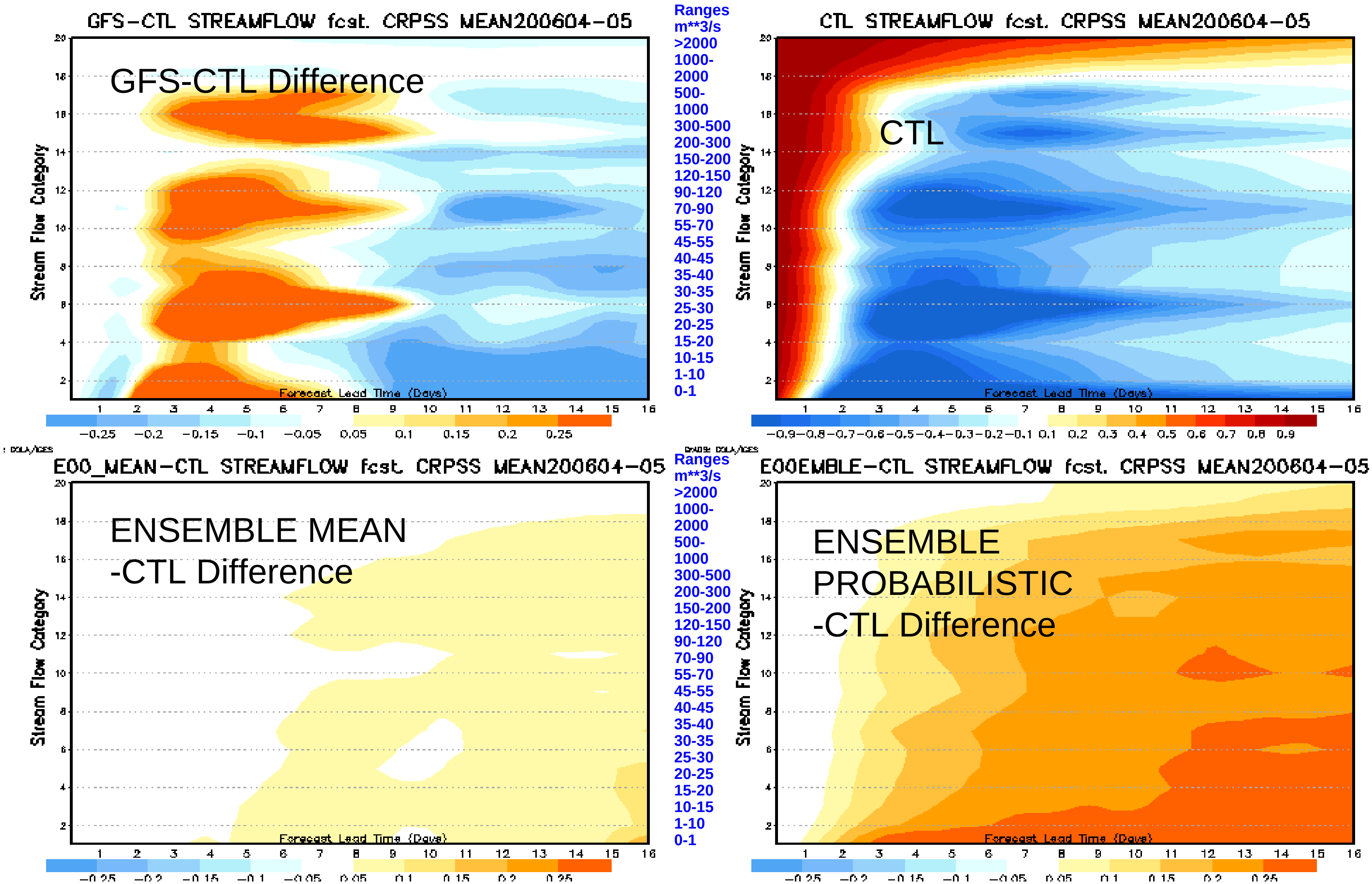






CRPSS as a Function of Lead Time and Mean Flow, Raw Forecasts

Slight Improvement due to ensemble approach
Major Improvement due to probabilistic forecast
High resolution GFS is superior for 2-8 day lead



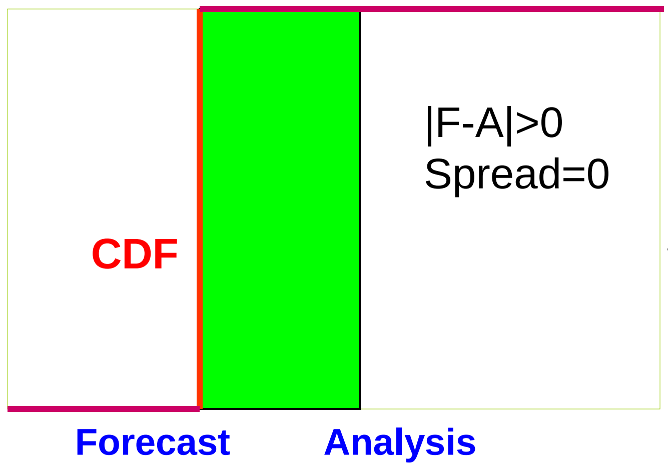
How CRPS Reflects Errors in 1^{st} (position) and 2^{nd} (dispersion) Moments?

In the situation where 1^{st} moment error exists (|Fmean-A|>0), CRPS is minimized if

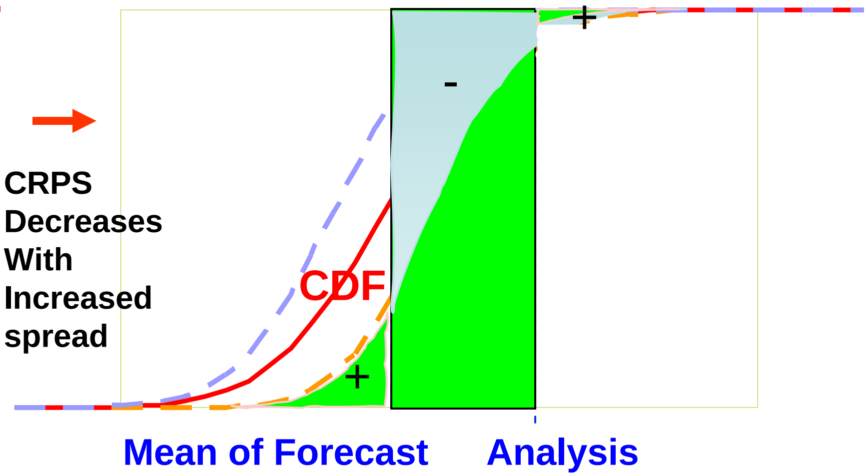
Spread ~ |Fmean-A|

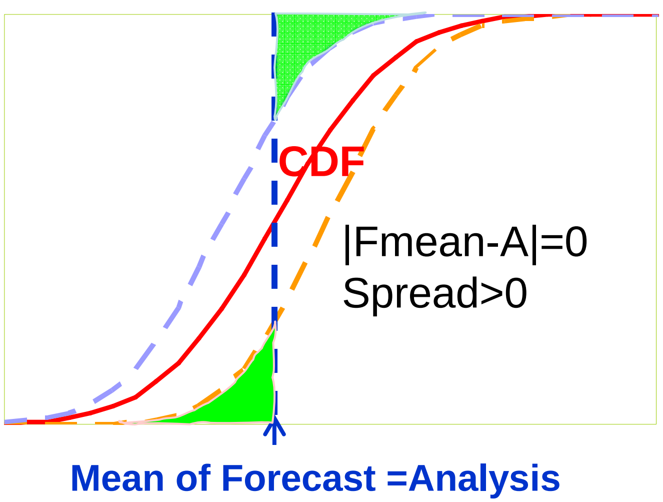
(an idealized ensemble).

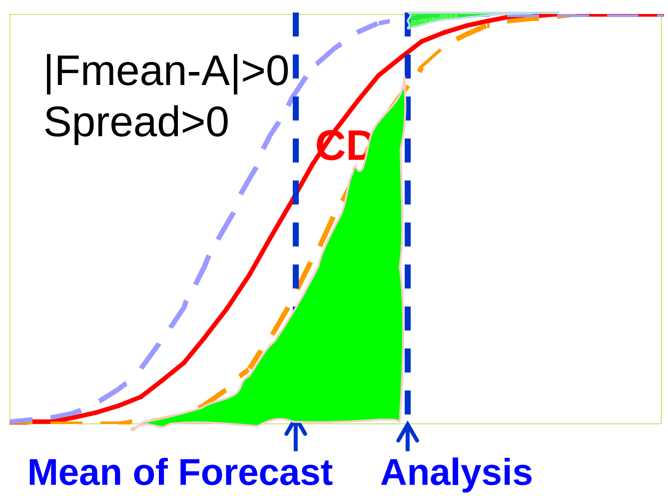
CRPS is smaller if (1) the analysis is closer to the mean of the forecast pdf and (2) spread is smaller (CRPS=0 for a perfect deterministic forecast).



CRPS Decreases With Increased spread







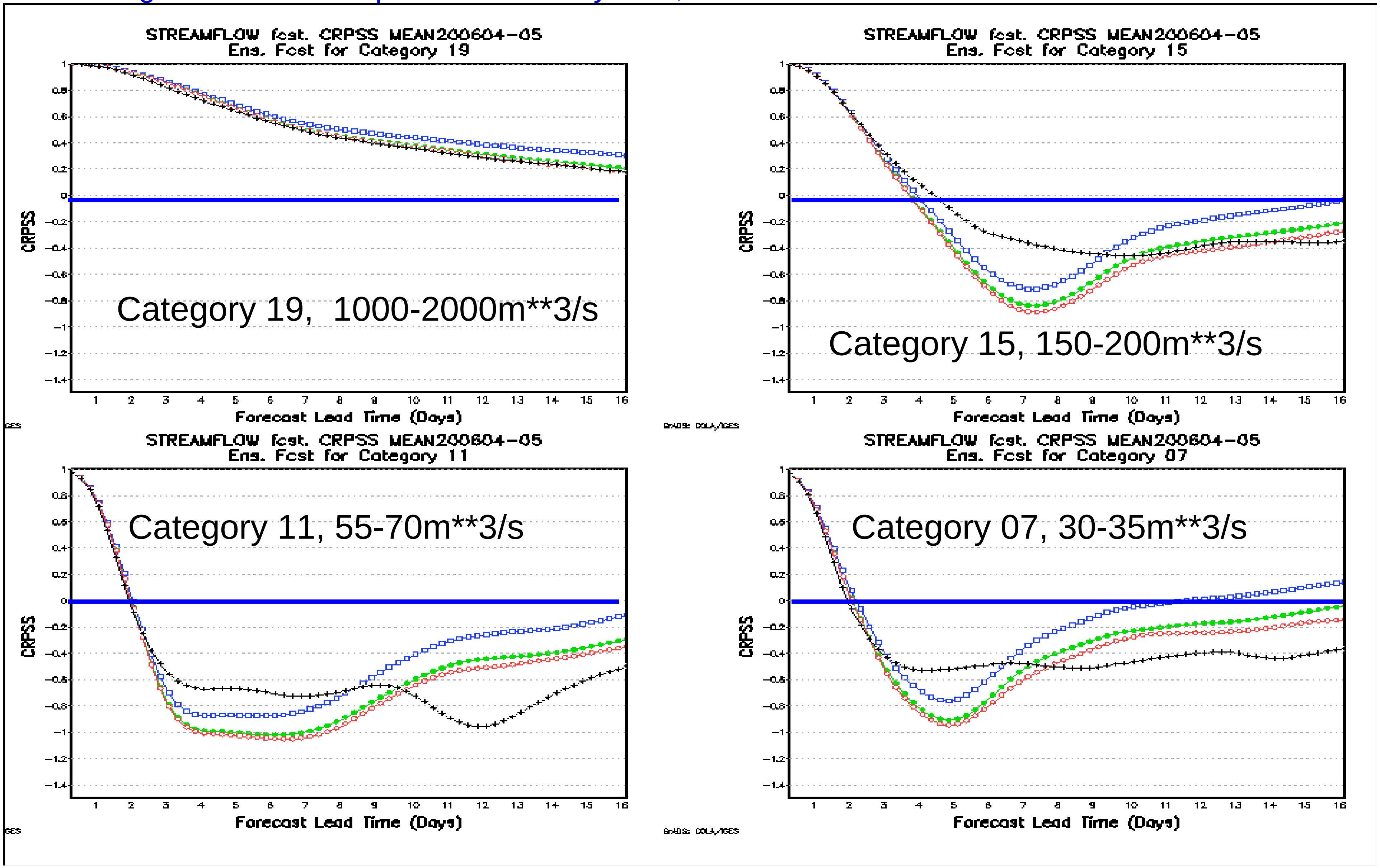
CRPS can be reduced by bias correction (adjustment of the first moment) and/or spread inflation (adjustment of the second moment)

Category-mean of CRPSS (Probabilistic based on GEFS)

Slight Improvement due to ensemble approach
Major Improvement due to probabilistic forecast

High res. GFS is superior for 2-8 day lead, small and medium basins

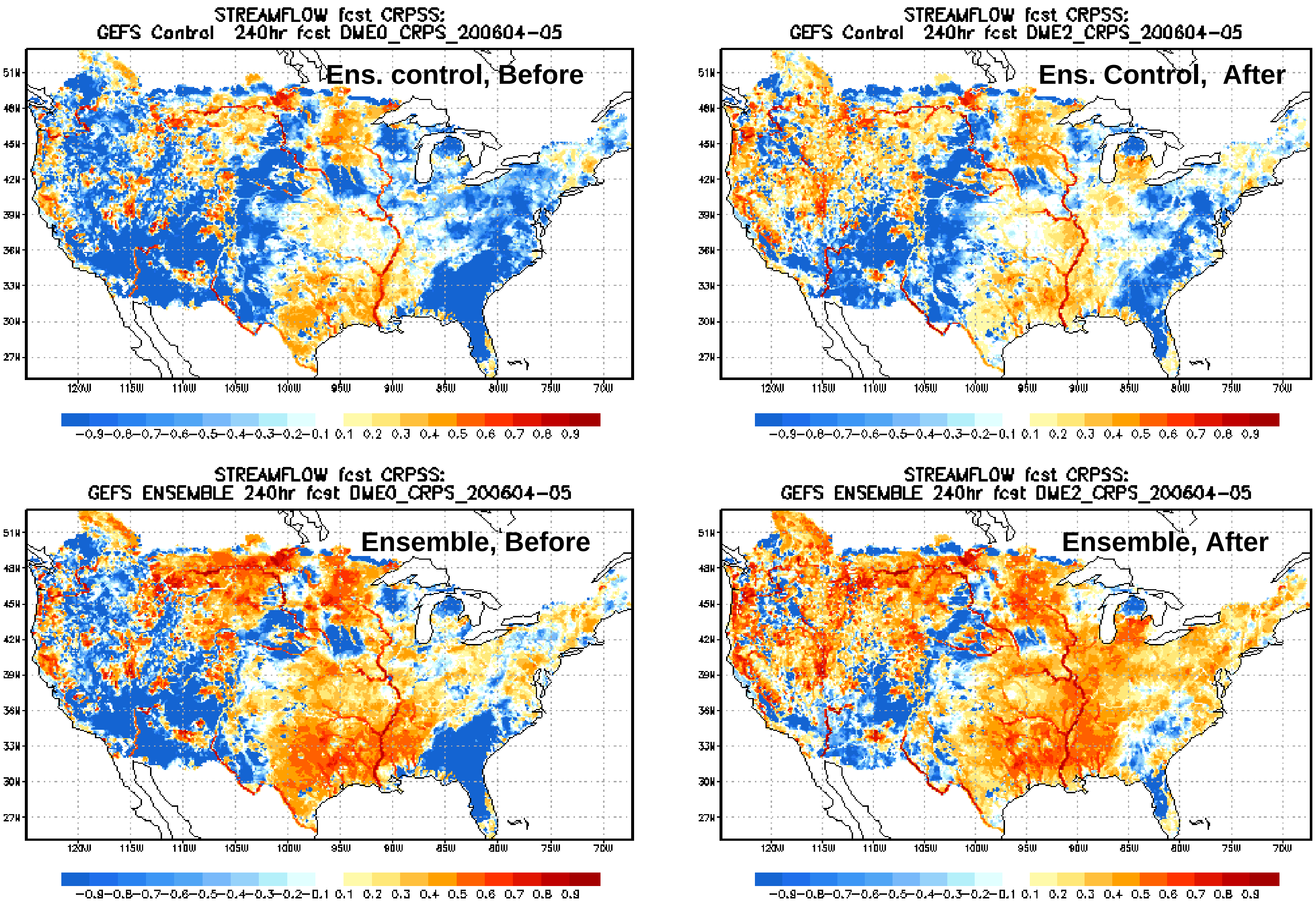
--- GFS
--- CTL
--- ENS. MEAN
--- ENSEMBLE



Effect of Bias Correction

CRPSS of Ensemble Control and Ensemble

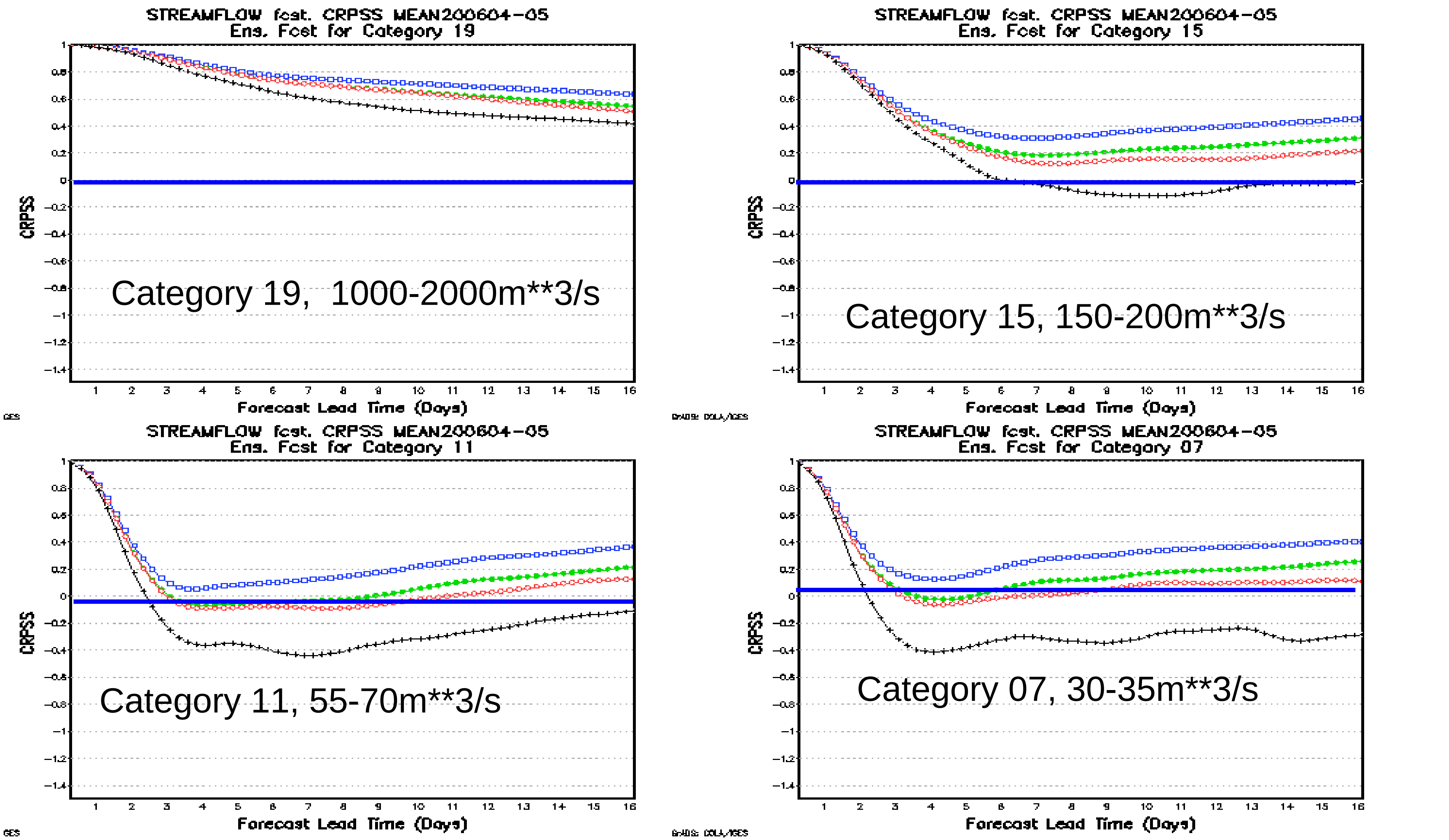
(Lead Time 240h; before and after bias-correction)



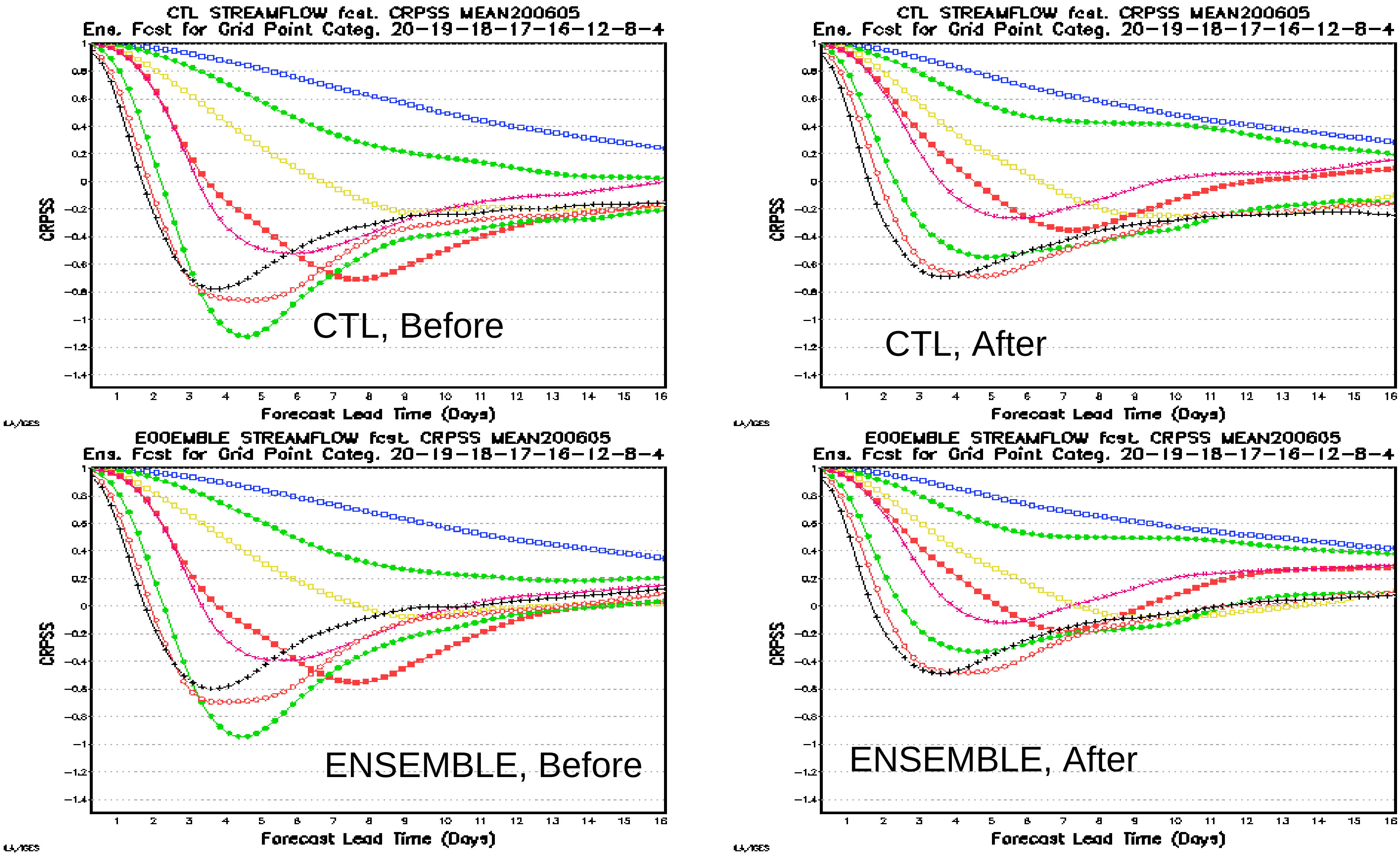
Category-mean of CRPSS, After Bias Correction (Probabilistic Forecast based on GEFS)

High res. GFS is NOT superior

--- GFS
--- CTL
--- ENS. MEAN
--- ENSEMBLE



Bias Correction with Independent Training Data Set
(Training: April; Evaluation: May)



MOTIVATION FOR ATM / LAND / HYDRO ENSEMBLE EXPS

- Purpose of seminar
  - Share initial results
  - Seek advice and collaboration

- Main goal of experiments
  - Evaluate quality of meteorological forcing (precipitation)

- Approach
  - Work with a land surface & river routing model that is readily available
    - Focus is not on particular land/hydro models used, that's secondary
  - Study quality of river flow forecasts to learn about shortcomings in meteorological forcing (ensemble)

- Outcomes
  - Use results to adjust priorities for THORPEX and related work on improving ensemble forcing for hydrological applications
  - Explore possibility of distributed atmospheric/land surface / hydro ensemble forecasting
    - Is there any promise with available simple models and approaches used?
    - Work collaboratively to further explore this avenue with better models, techniques, etc

XEFS PLANS

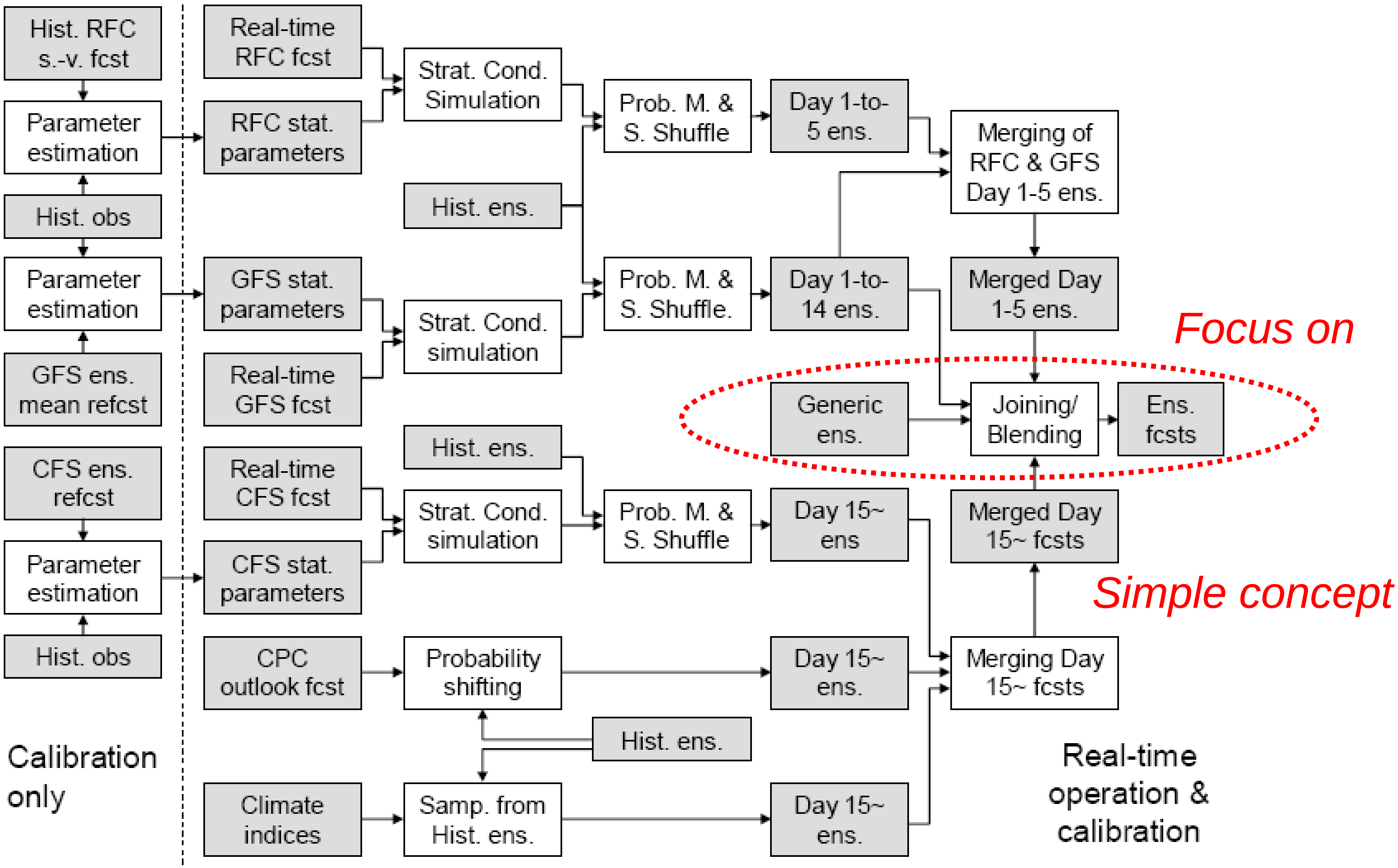


Figure 2.2 EPP3 science algorithm suite workflow diagram.

From "The Experimental ensemble Forecast System (XEFS) Design and Gap Analysis", report of the XEFS Design and Gap Analysis Team, NOAA/NWS

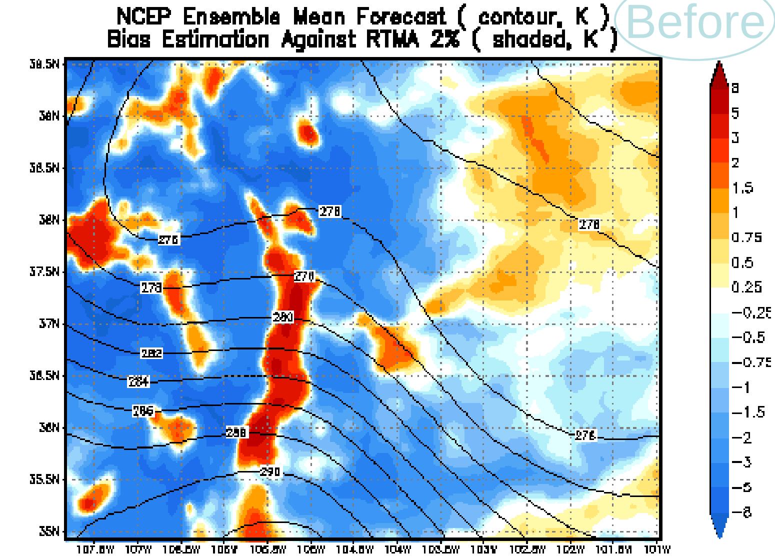
PROBABILISTIC NUMERICAL GUIDANCE FOR HIGH IMPACT EVENTS

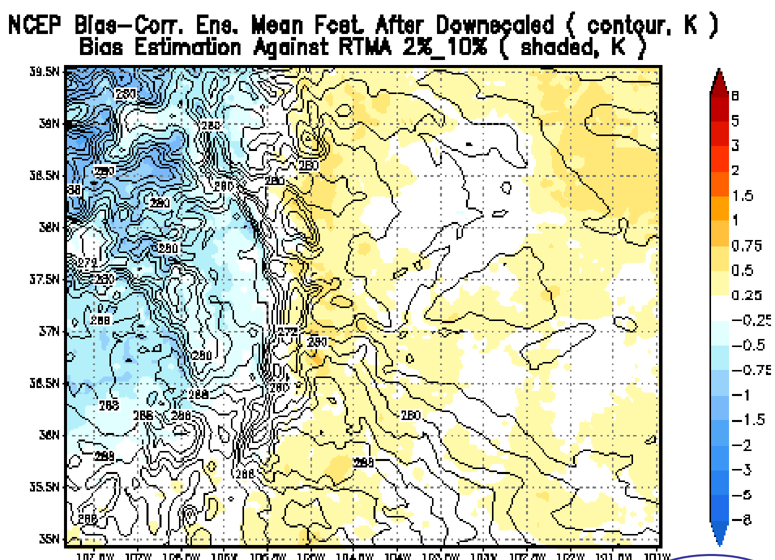
- Mini-POP
  - Developed under EMP & STI (THORPEX)
- Goal
  - Bias corrected & downscaled ensemble forecasts for wide variety of users
    - NCEP Service Centers, WFOs modify numerical first guess, keep ensemble format
    - Generate any and all products from primary bias-corrected / downscaled ensemble
- Flagship
  - North American Ensemble Forecast System
    - Joint NCEP / Canadian ensemble
      - Bias correction of first moment for 35 quasi-normal variables
      - Combination of two ensembles
      - Climate anomalies for ~20 variables
- Future plan includes
  - Bias correction of all model variables on model grids
    - Unified Bayesian approach
      - All time scales (SREF, NAEFS, CFS)
      - All variables, including precipitation
    - Hind-casts as needed generated in real time
      - Allows frequent model updates
  - Downscaling to NDFD (or similar) grid, using RTMA analysis
    - Preliminary example for 2m temperature (10m winds also available)
      - More advanced downscaling approaches to be explored
        - » Capture case dependent information on fine scales
        - » Add stochastic perturbations to represent uncertainty on NDFD scales

UNDER TESTING - Ensemble Mean Forecast bias & RMS error before & after bias correction & downscaling

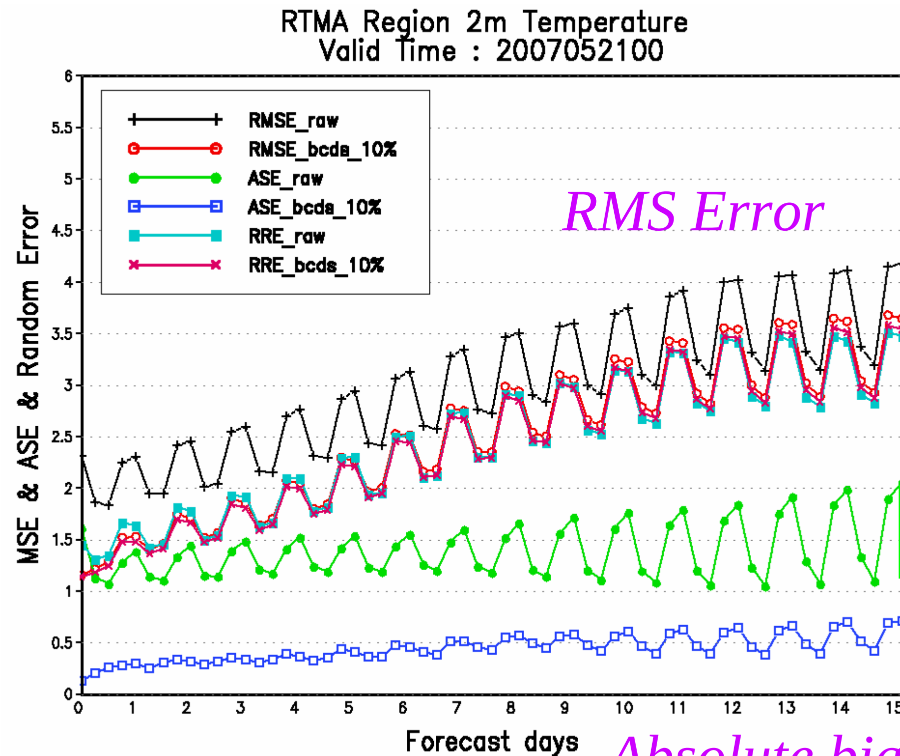
24hr

Before





After



RMS Error

Before

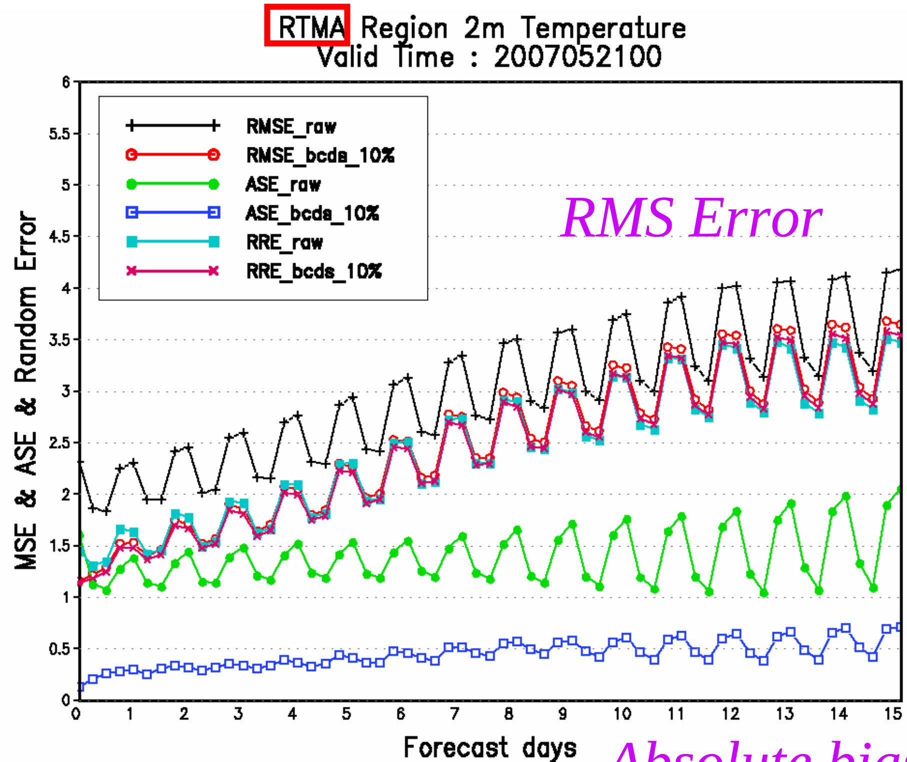
After

Before

After

Absolute bias error

QUALITATIVE COMPARISON OF ADAPTIVE BIAS CORRECTION & DOWNSCALING METHODS WITH EXISTING APPROACHES



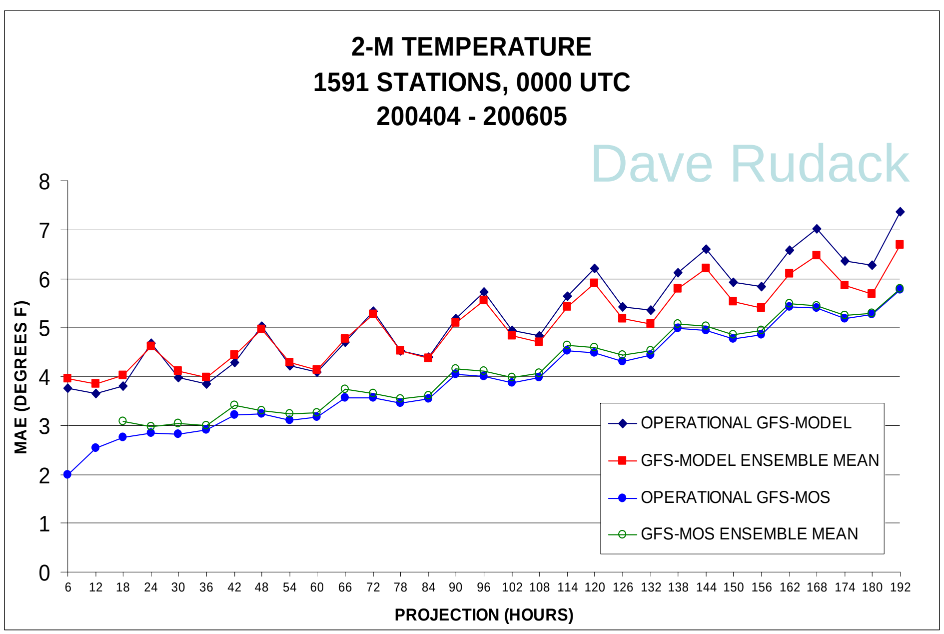
Before

After

Before

After

Absolute bias error



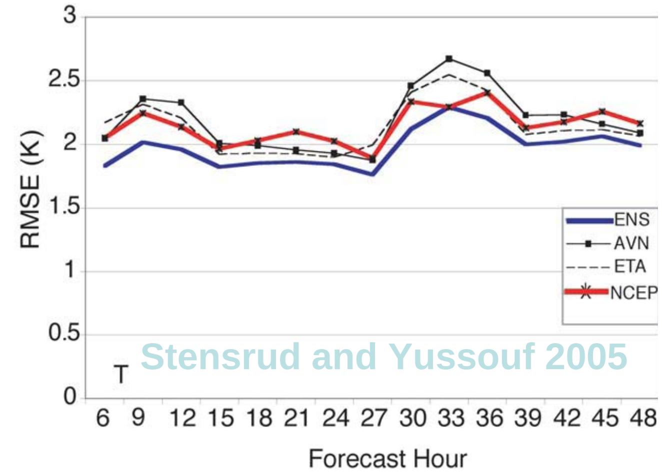


FIG. 5. Values of root-mean-square error (K) plotted as a function of forecast hour for (top) 2-m temperature from the full 31 member BCE (blue), the NCEP-only BCE (red), and the AVN (solid black line) and Eta (dashed line) MOS. Results are calculated at 1258 station locations for both the ensemble and AVN and Eta MOS data (after Stensrud and Yussouf 2005).

REAL-TIME GENERATION OF HIND-CAST DATASET?