Ensemble Verification Update

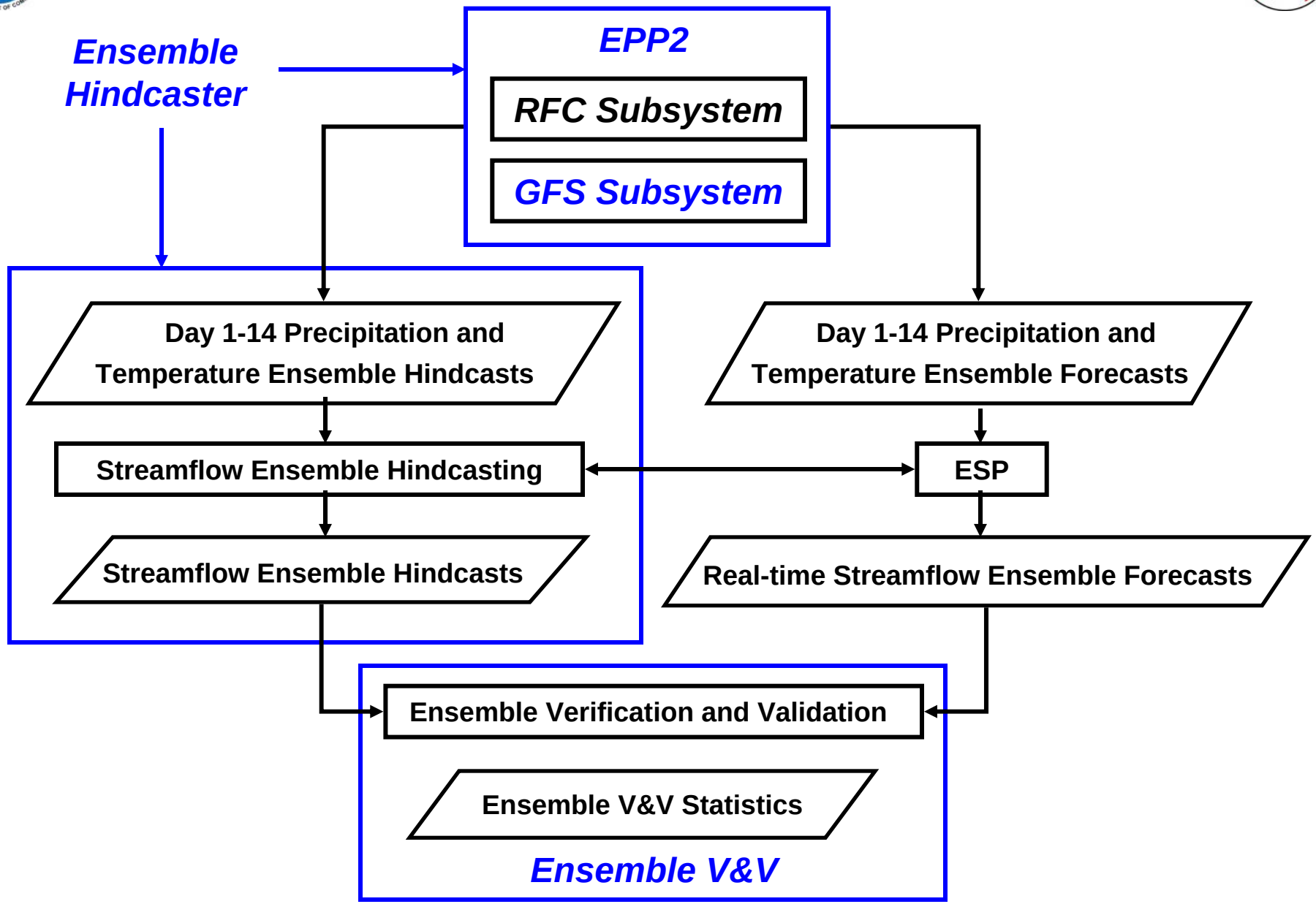
D.-J. Seo[1,2], Julie Demargne[1,2], Shuzheng Cong[1,3],
James Brown[1,2], Limin Wu[1,4], Hank Herr[1,5]

[1]Hydrologic Ensemble Prediction Group
Hydrologic Science and Modeling Branch
[5]Hydrologic Software Engineering Branch

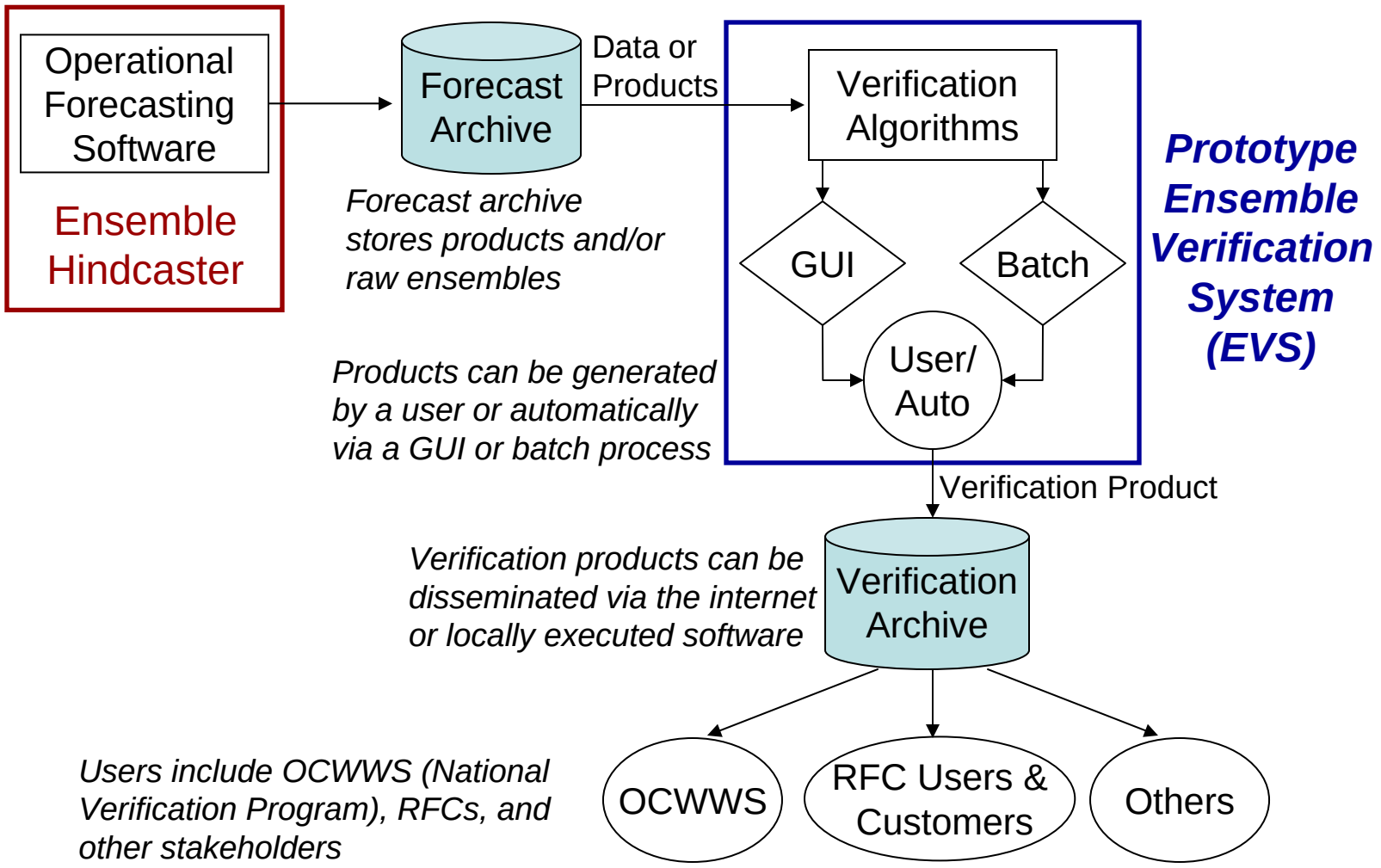
Hydrology Laboratory
Office of Hydrologic Development
NOAA/National Weather Service

[2]University Corporation for Atmospheric Research
[3]Retired
[4]RS Information Systems, Inc.

FY06 Ensemble Projects and their Relationships



Vision for Operational Ensemble Verification System

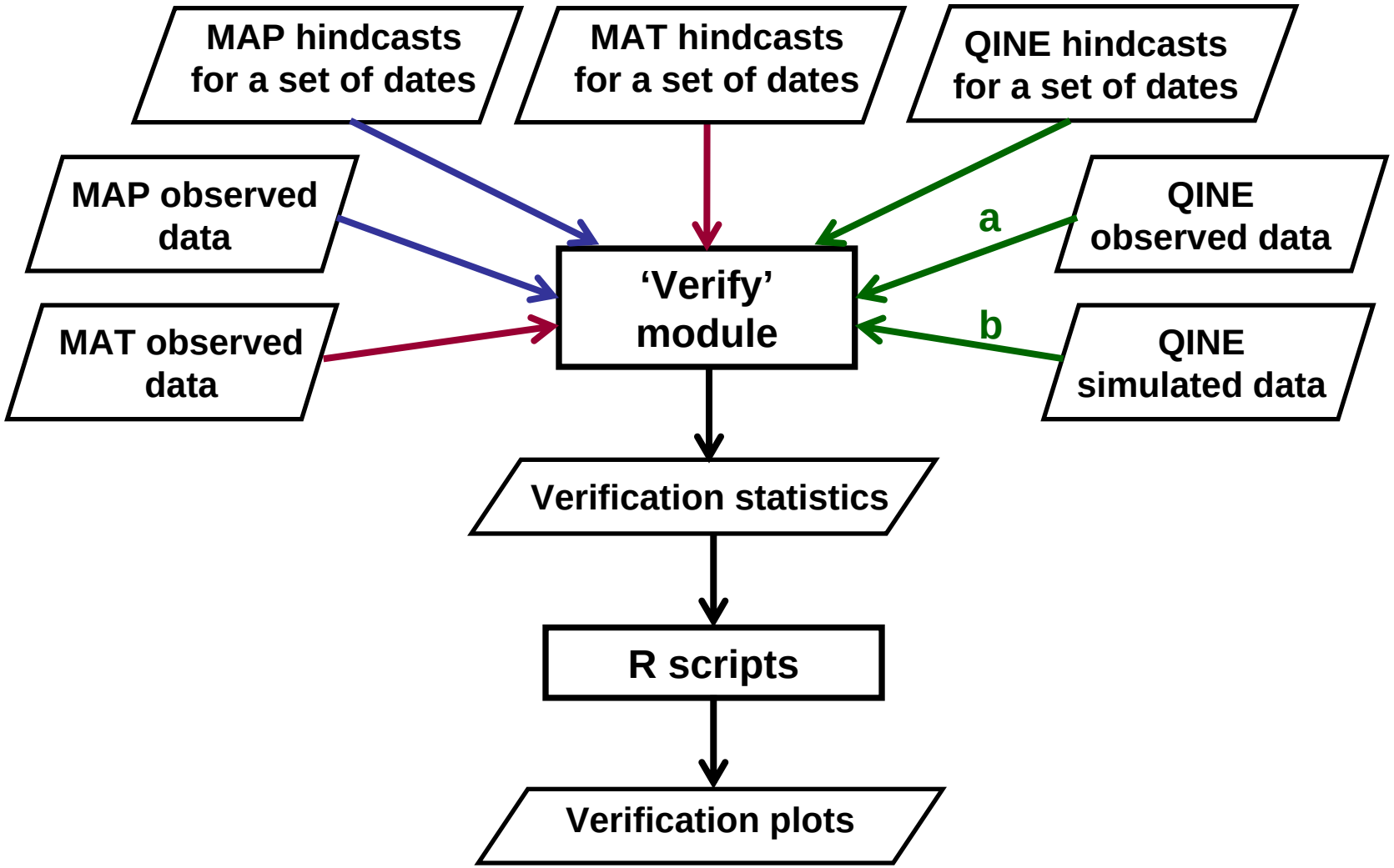


FY06 Activities

- Improved general robustness of the verification algorithms in the prototype Ensemble Verification System (EVS)
- Developed and implemented new capabilities, including:
  - Separation of input and hydrologic uncertainties in ensemble streamflow verification
  - Aggregation of ensemble verification results over multiple basins to reduce sampling uncertainties
- Enhanced the R scripts for graphical display of various verification statistics

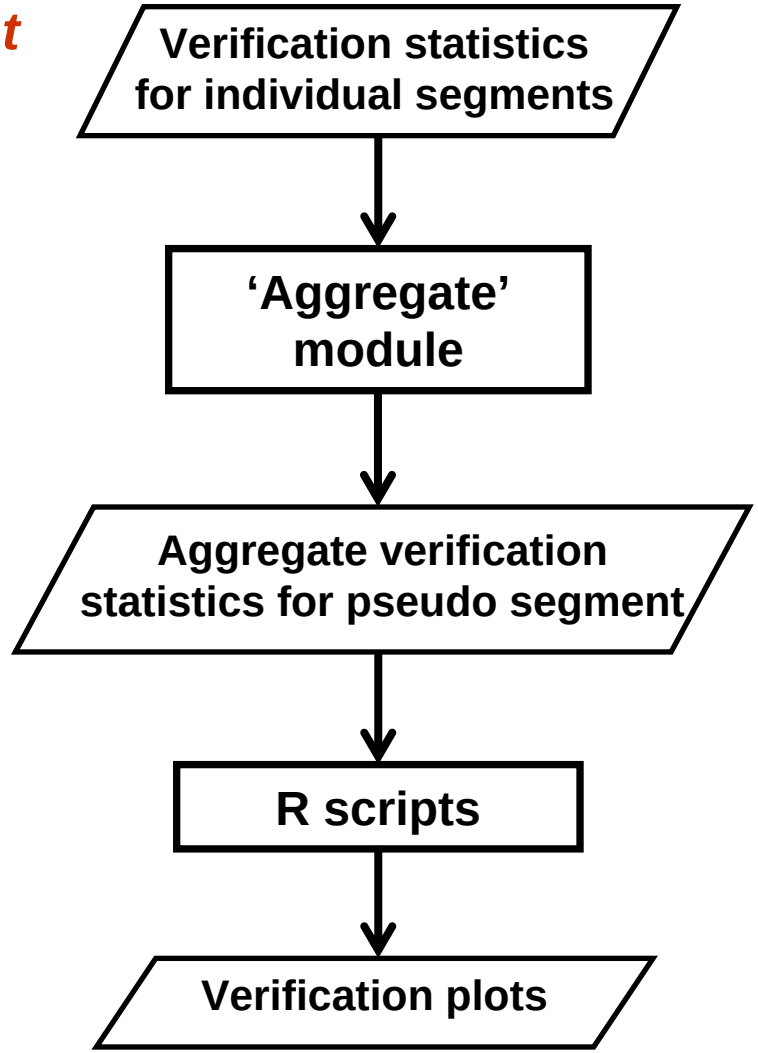
Ensemble Verification: Data & Processes

For one segment



Ensemble Verification: Data & Processes (cont.)

For a group of segment



‘Verify’ Module

- Pair observations and ensemble forecasts (with same or different time systems) (optional)
- Compute verification statistics for deterministic and probabilistic forecasts
  - Scatter plots
  - Deterministic verification of ensemble mean, comparison with persistence forecast: Ratio, Mean Error, Root Mean Square Error, Correlation Coefficient
  - Brier Score (BS) and its decomposition, Brier Skill Score (BSS)
  - Rank Probability Score (RPS), Rank Probability Skill Score (RPSS)
  - Reliability Diagram (# of bins to be defined)
  - Relative Operating Characteristics (# of points to be defined)

‘Aggregate’ Module

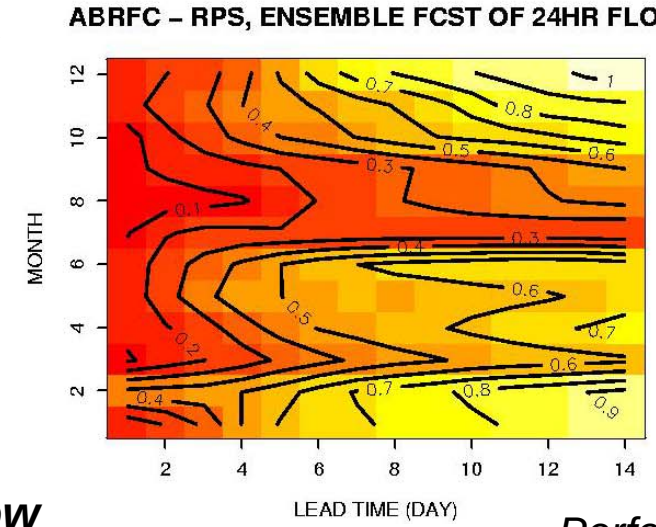
- Compute aggregate verification statistics as the weighted average of individual basin statistics (given number of observed events) to generate input files for R scripts:
  - Brier Score (BS) and its decomposition, Brier Skill Score (BSS)
  - Rank Probability Score (RPS), Rank Probability Skill Score (RPSS)
  - Reliability Diagram (# of bins to be defined)
  - Relative Operating Characteristics (# of points to be defined)

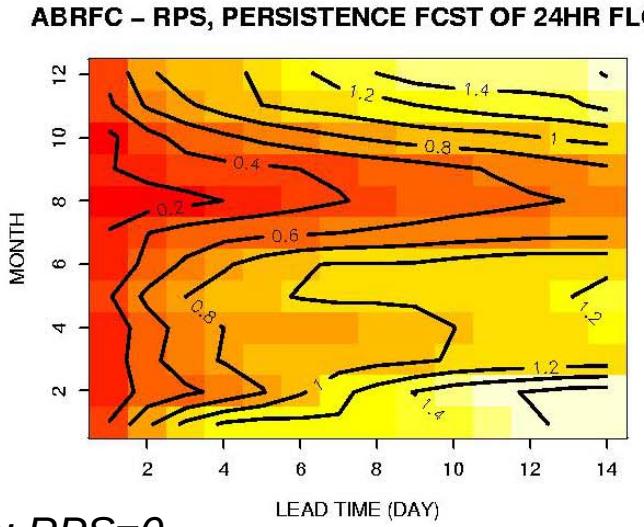
R Scripts

- A set of R scripts to be run for each verification statistic:
  - Scatter plots (3 temporal scales)
  - Deterministic verification of ensemble means, comparison with persistence forecast: Ratio, Mean Error, Root Mean Square Error, Correlation Coefficient (monthly scale)
  - Brier Score (BS) and its decomposition, Brier Skill Score (BSS) (annual scale)
  - Rank Probability Score (RPS), Rank Probability Skill Score (RPSS) (monthly scale)
  - Reliability Diagram (3 temporal scales)
  - Relative Operating Characteristics (3 temporal scales)

EVS Output: RPS & RPSS

Rank Probability Score measures overall forecast performance

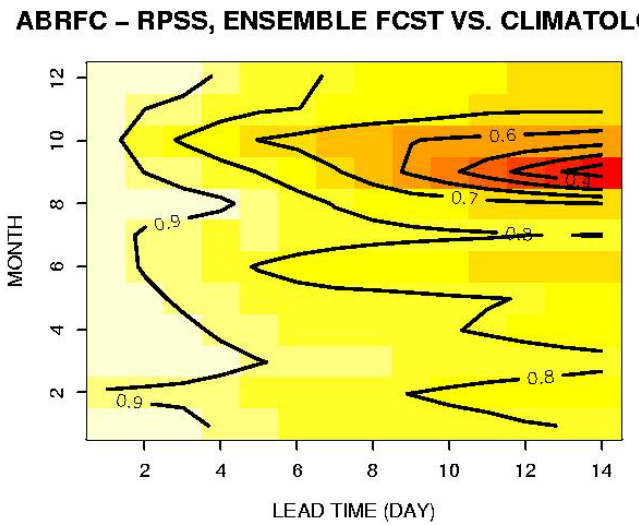


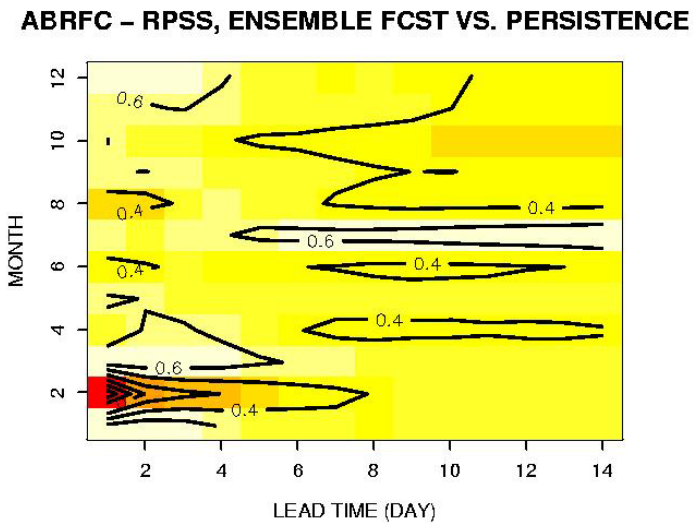


24-hr flow

Perfect score: RPS=0

Rank Probability Skill Score measures improvement over reference forecast





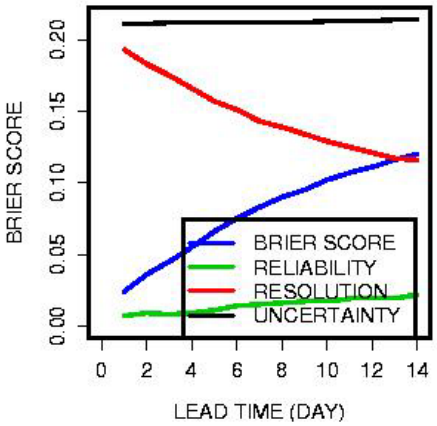
Perfect score: RPSS=1

EVS Output: BS & BSS

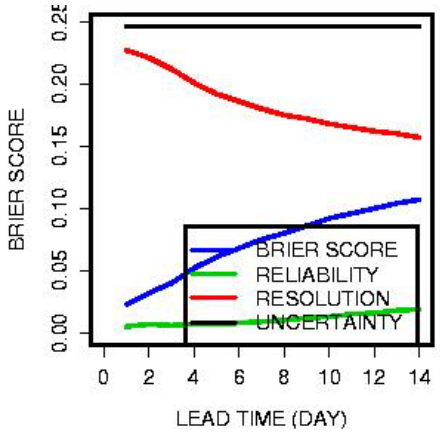
BS = Reliability – Resolution + Uncertainty

Brier Score measures mean squared probability error

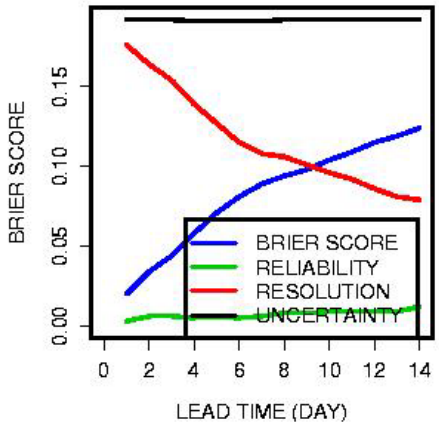
Flow - 25th Per.



Flow - 50th Per.



Flow - 75th Per.

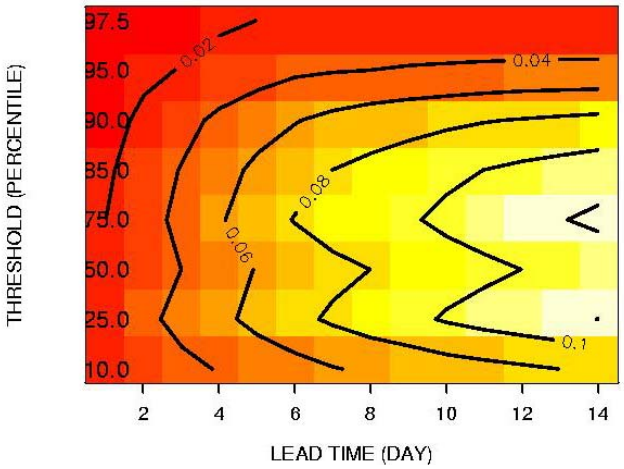


24-hr flow

Brier Score

Perfect score: BS=0

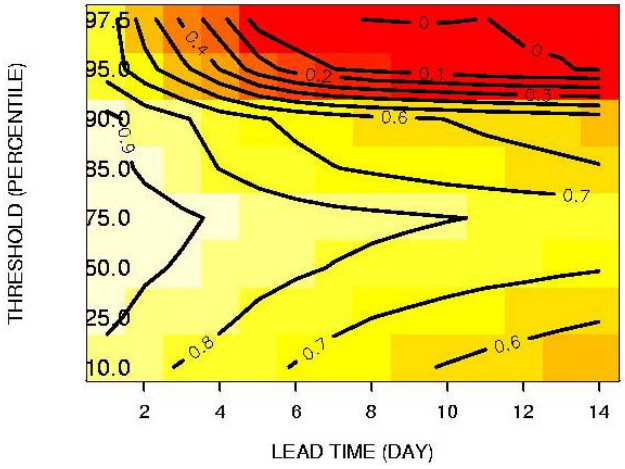
ABRFC – BS, ENSEMBLE FCST OF 24HR FLOW



Brier Skill Score

Perfect score: BSS=1

ABRFC – BSS, ENSEMBLE FCST VS. CLIMATOLOGY



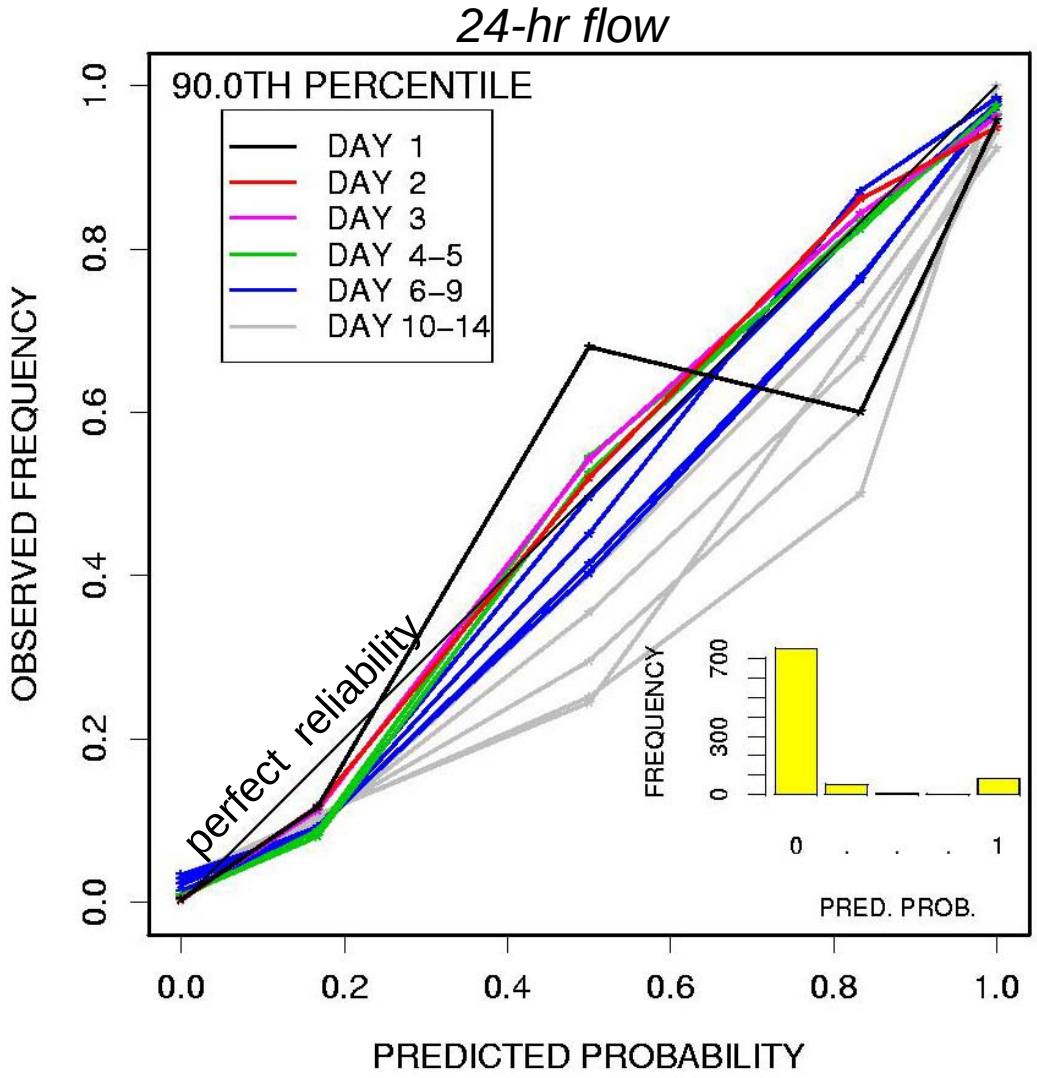
EVS Output: Reliability Diagram

- **Reliability Diagram** measures agreement between forecast probability and mean observed frequency

Deviation from diagonal gives conditional bias

Below diagonal: over-forecasting
Above diagonal: under-forecasting

Range of forecast probabilities is divided into K bins (+ 0 and 1 bins). Histogram gives sample size in each bin for lead day 1.



EVS Output: ROC

- **Relative Operating Characteristic** measures resolution (ability of forecast to discriminate between events & non-events)

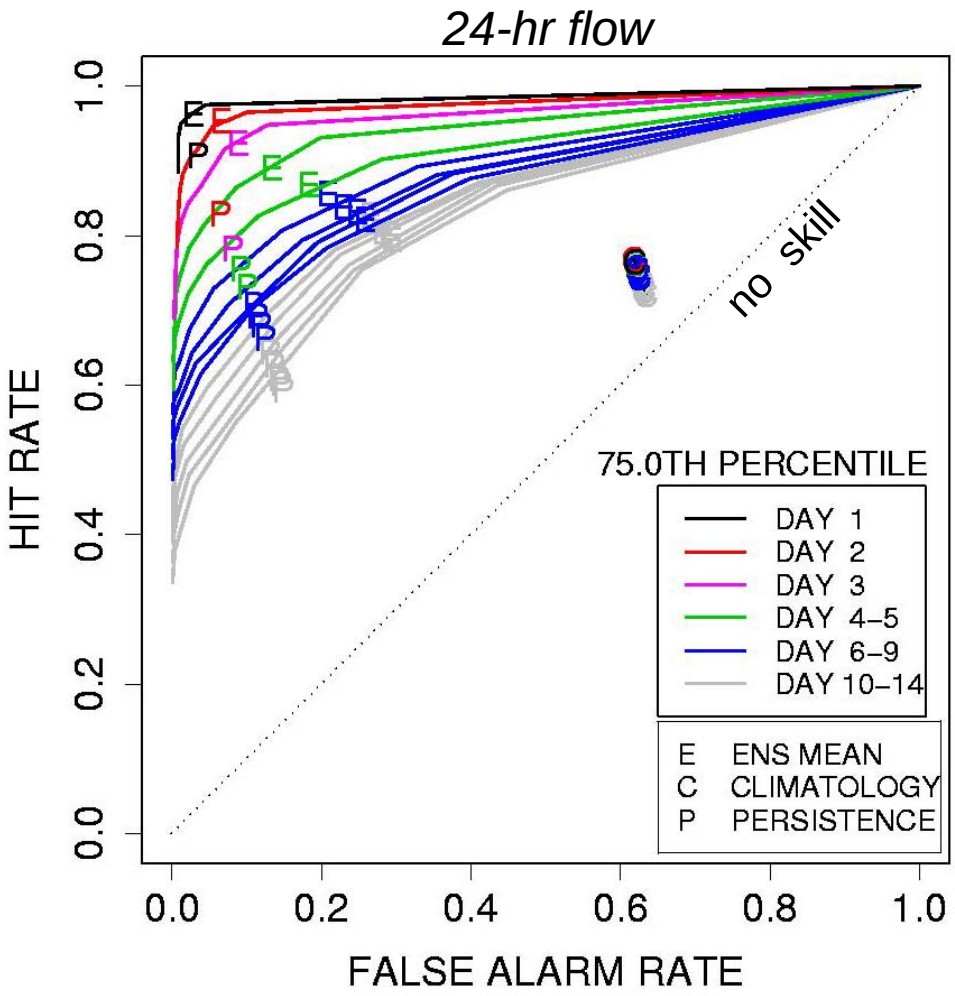
$$\text{Hit Rate (HR)} = \frac{\text{Hits}}{\text{Hits + Misses}}$$

Perfect score: HR = 1

$$\text{False Alarm Rate (FAR)} = \frac{\text{False alarms}}{\text{False alarms + Correct rejections}}$$

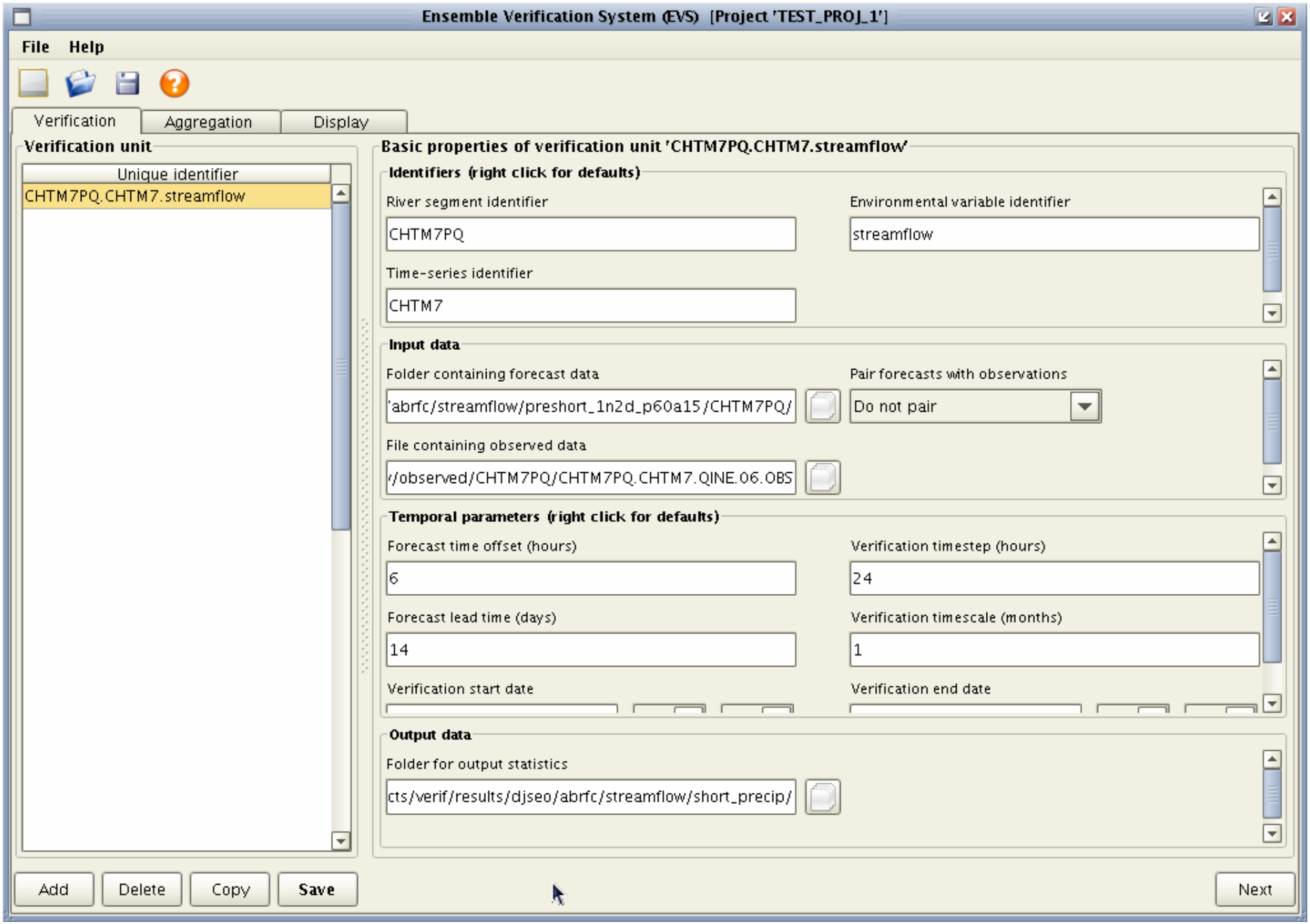
Perfect score: FAR = 0

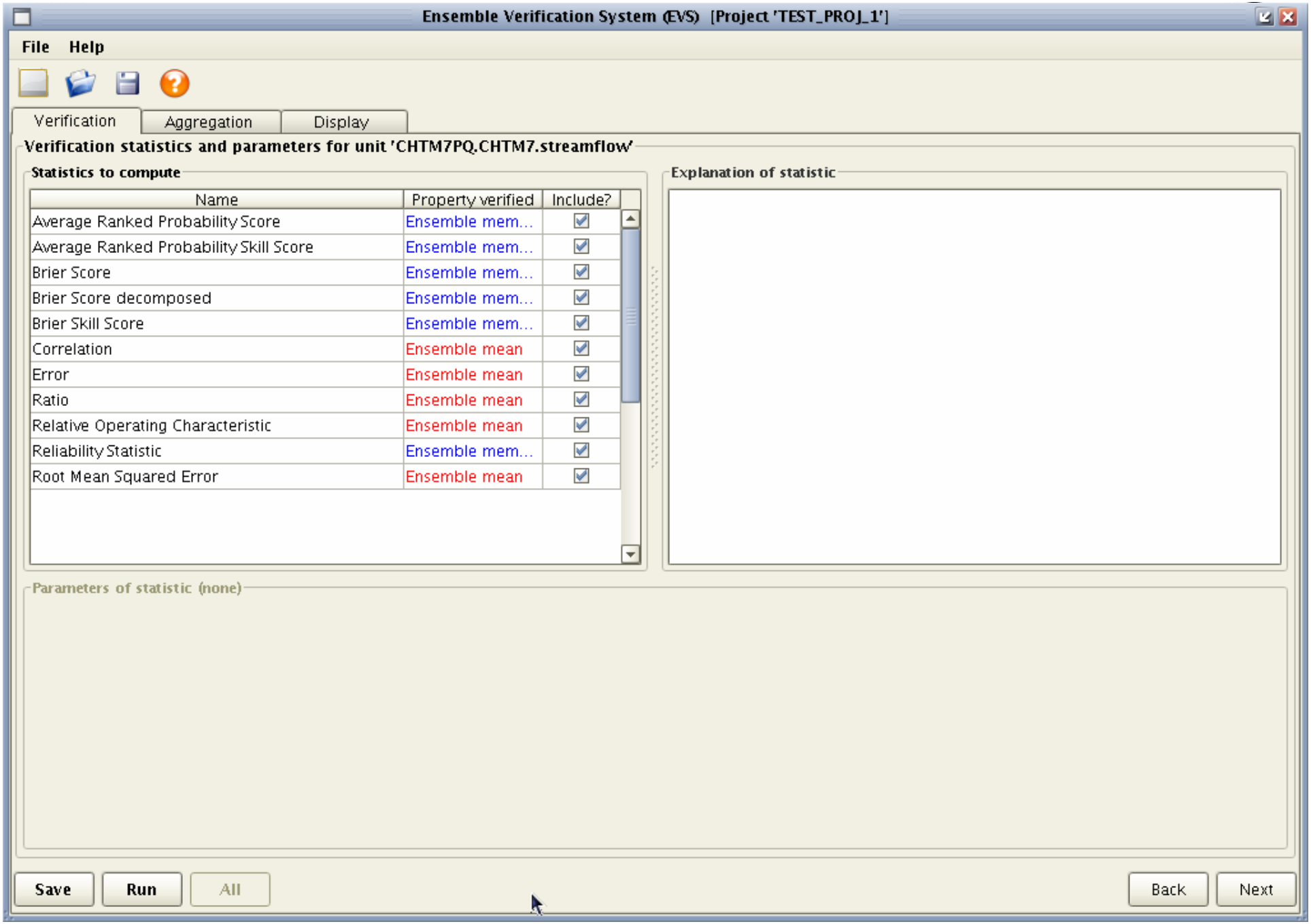
ROC with 10 probability thresholds & ROC for Ensemble Mean, Monthly Climatology Mean, & Persistence forecasts

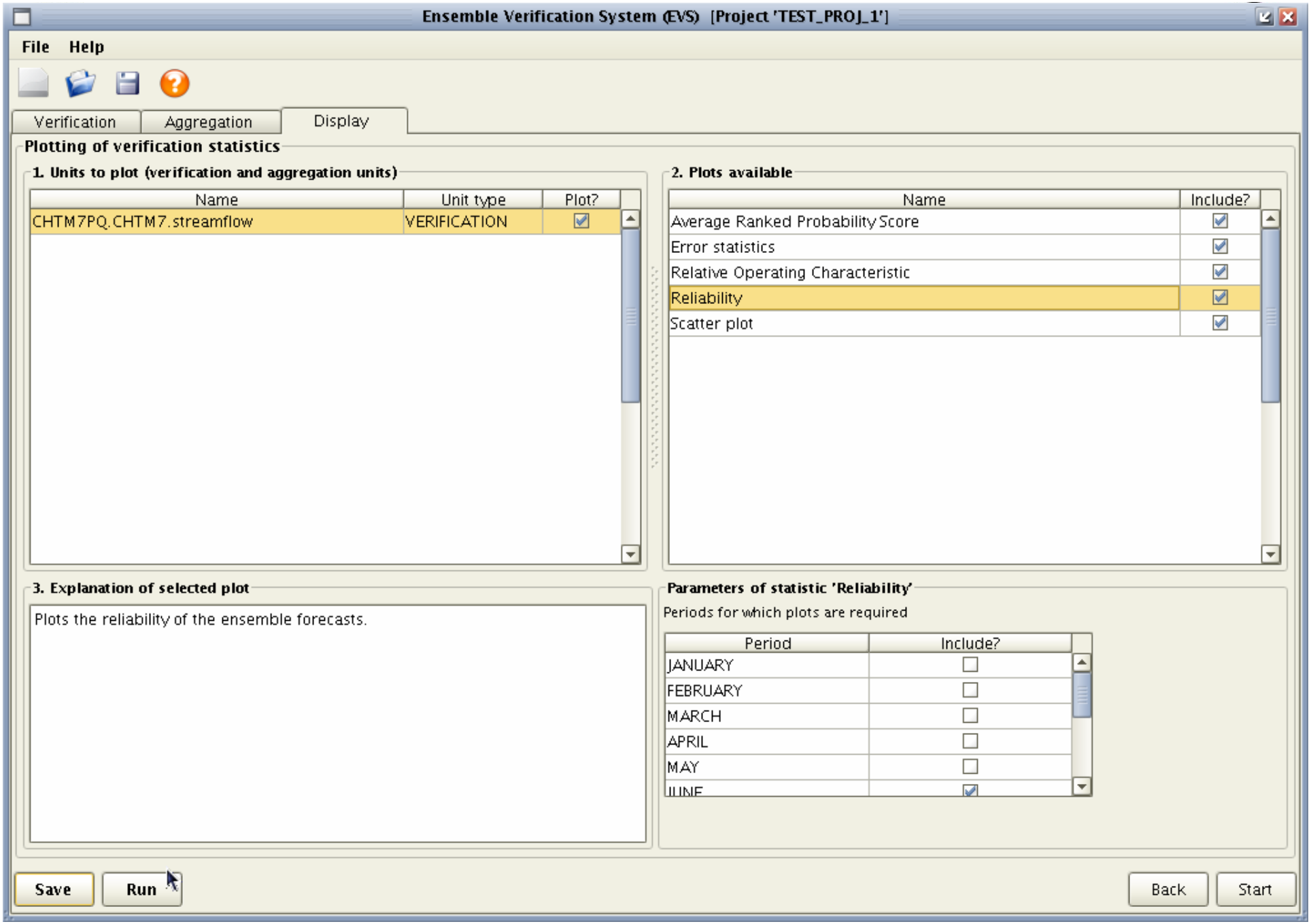


FY07 Activities

- Developed a prototype user interface for ‘verify’ and ‘aggregate’ modules and R scripts
  - Verify, aggregate or display
  - Choose verification statistics to calculate
  - Choose plots to generate


Ensemble Verification System (EVS) [Project 'TEST_PROJ_1']
File Help
Verification Aggregation Display
Verification unit
Unique identifier
CHTM7PQ.CHTM7.streamflow
Basic properties of verification unit 'CHTM7PQ.CHTM7.streamflow'
Identifiers (right click for defaults)
River segment identifier
CHTM7PQ
Environmental variable identifier
streamflow
Time-series identifier
CHTM7
Input data
Folder containing forecast data
'abrfc/streamflow/preshort_1n2d_p60a15/CHTM7PQ/
Pair forecasts with observations
Do not pair
File containing observed data
/observed/CHTM7PQ/CHTM7PQ.CHTM7.QINE.06.OBS
Temporal parameters (right click for defaults)
Forecast time offset (hours)
6
Verification timestep (hours)
24
Forecast lead time (days)
14
Verification timescale (months)
1
Verification start date
Verification end date
Output data
Folder for output statistics
cts/verif/results/djseo/abrfc/streamflow/short_precip/
Add
Delete
Copy
Save
Next

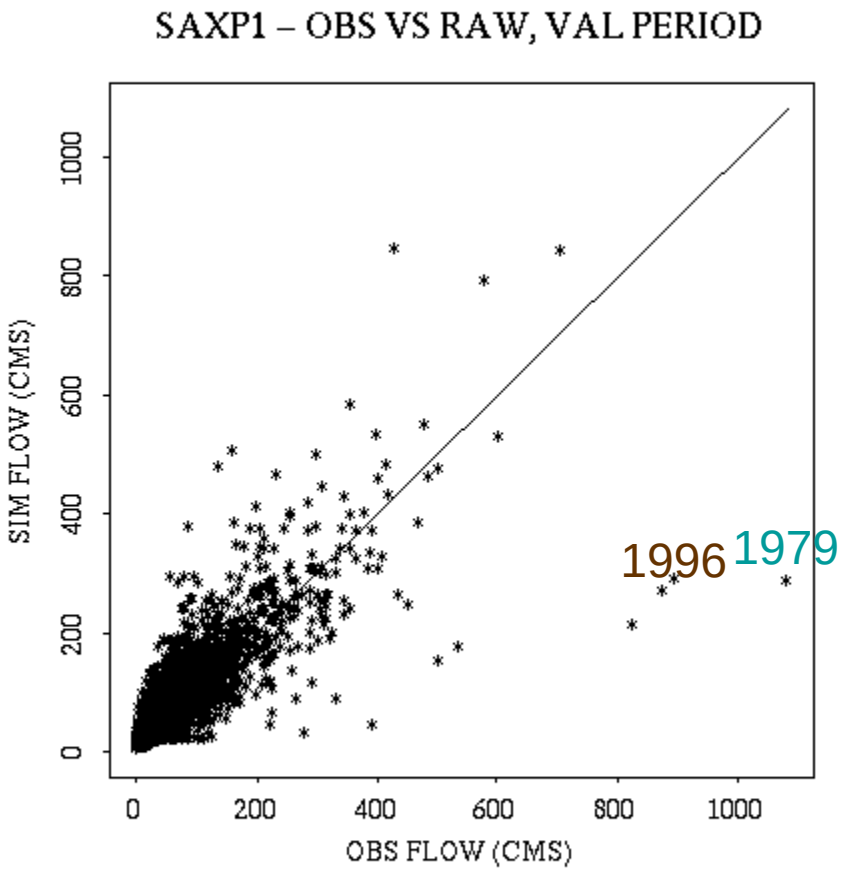

Ensemble Verification System (EVS) [Project 'TEST_PROJ_1']
File Help
Verification Aggregation Display
Verification statistics and parameters for unit 'CHTM7PQ.CHTM7.streamflow'
Statistics to compute
Name | Property verified | Include?
Average Ranked Probability Score | Ensemble mem... | ☑
Average Ranked Probability Skill Score | Ensemble mem... | ☑
Brier Score | Ensemble mem... | ☑
Brier Score decomposed | Ensemble mem... | ☑
Brier Skill Score | Ensemble mem... | ☑
Correlation | Ensemble mean | ☑
Error | Ensemble mean | ☑
Ratio | Ensemble mean | ☑
Relative Operating Characteristic | Ensemble mean | ☑
Reliability Statistic | Ensemble mem... | ☑
Root Mean Squared Error | Ensemble mean | ☑
Explanation of statistic
Parameters of statistic (none)
Save
Run
All
Back
Next

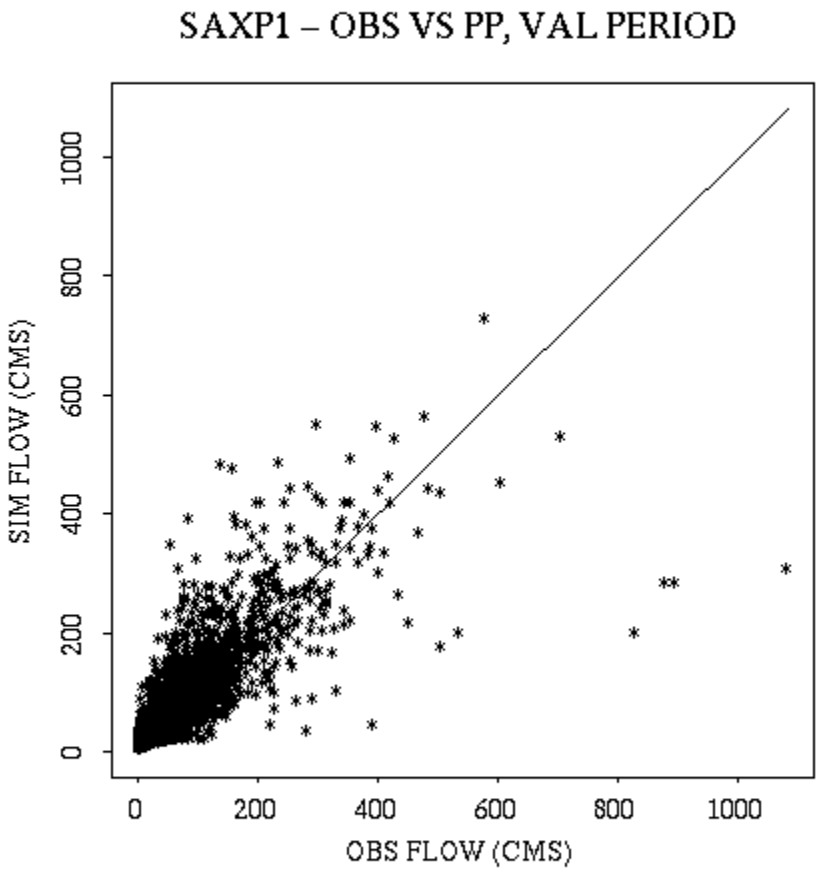

Ensemble Verification System (EVS) [Project 'TEST_PROJ_1']
File
Help
Verification
Aggregation
Display
Plotting of verification statistics
1. Units to plot (verification and aggregation units)
Name
Unit type
Plot?
CHTM7PQ.CHTM7.streamflow
VERIFICATION
2. Plots available
Name
Include?
Average Ranked Probability Score
Error statistics
Relative Operating Characteristic
Reliability
Scatter plot
3. Explanation of selected plot
Plots the reliability of the ensemble forecasts.
Parameters of statistic 'Reliability'
Periods for which plots are required
Period
Include?
JANUARY
FEBRUARY
MARCH
APRIL
MAY
JUNE
Save
Run
Back
Start

FY07 Activities (cont.)

- AHPS funding expected for:
  - Add metrics from the Verification Plan (and Confidence Intervals) and enhance display capabilities in EVS, release and support experimental EVS
  - Develop error analysis, enhance ensemble hindcaster, release and support experimental hindcaster
  - Evaluate existing archiving capabilities

Developing new ways to visualize goodness of ensemble forecasts and new/different performance measures for hydrology that are easy to understand and to communicate

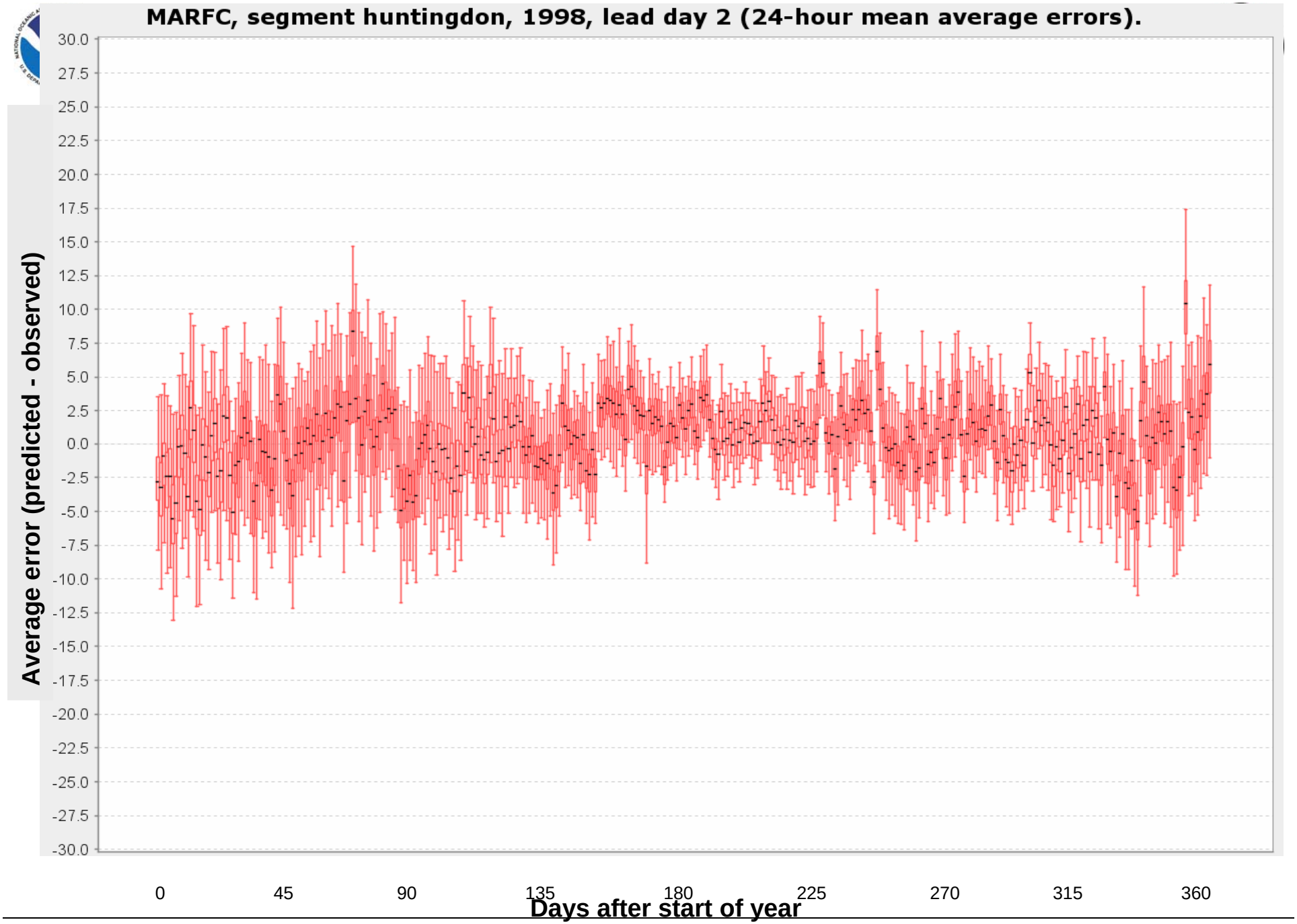


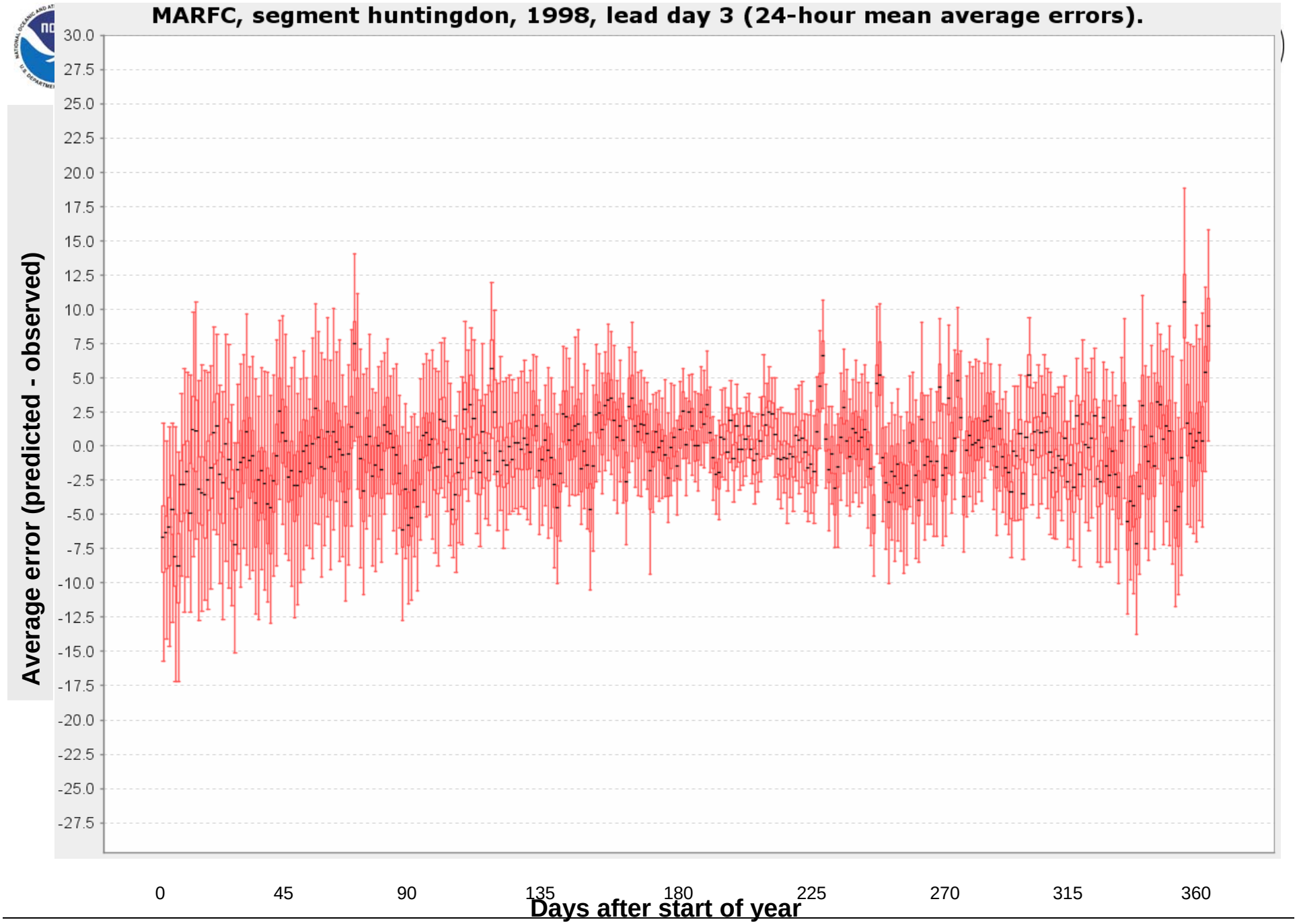


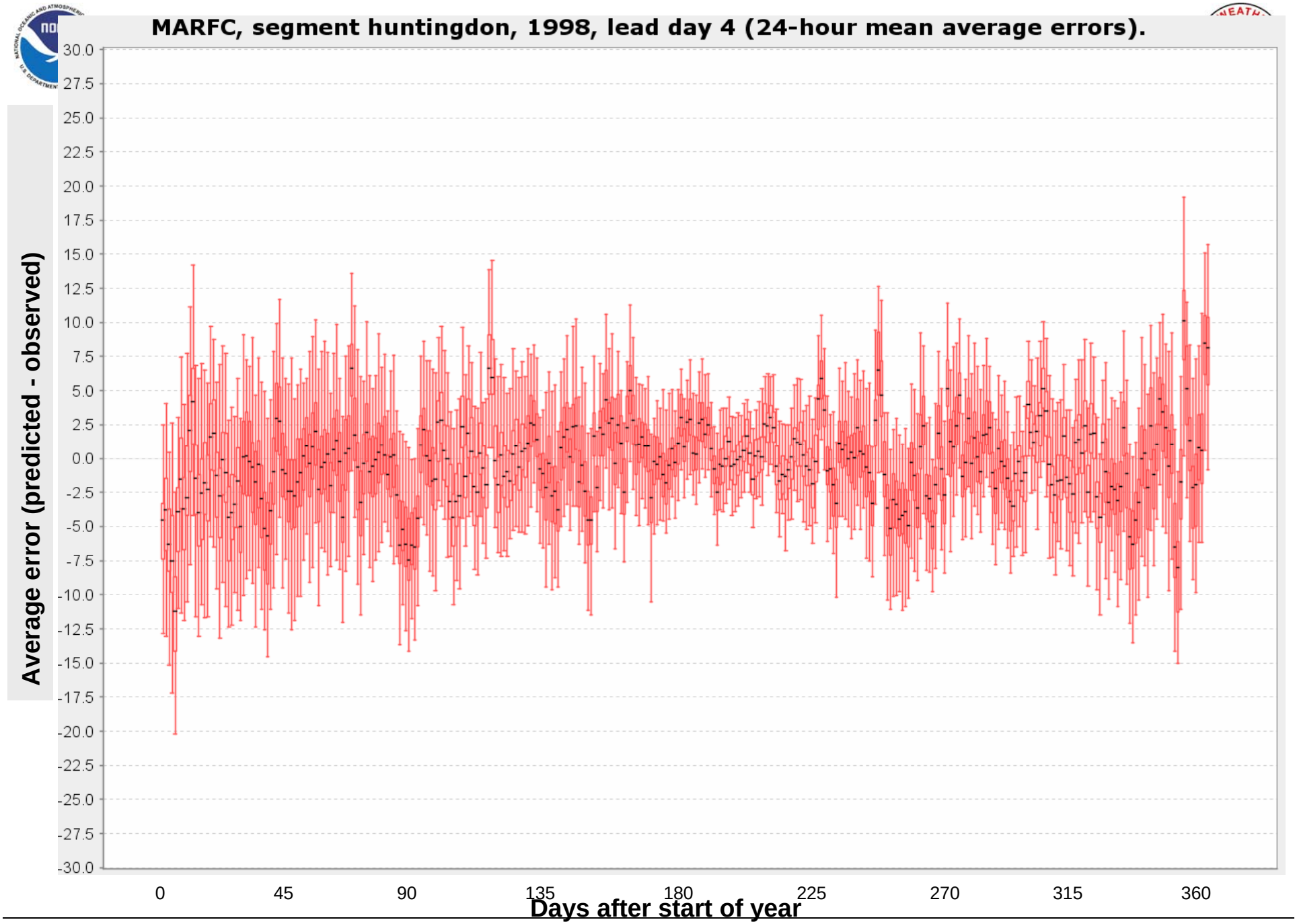
w/ rain-on-snow events included

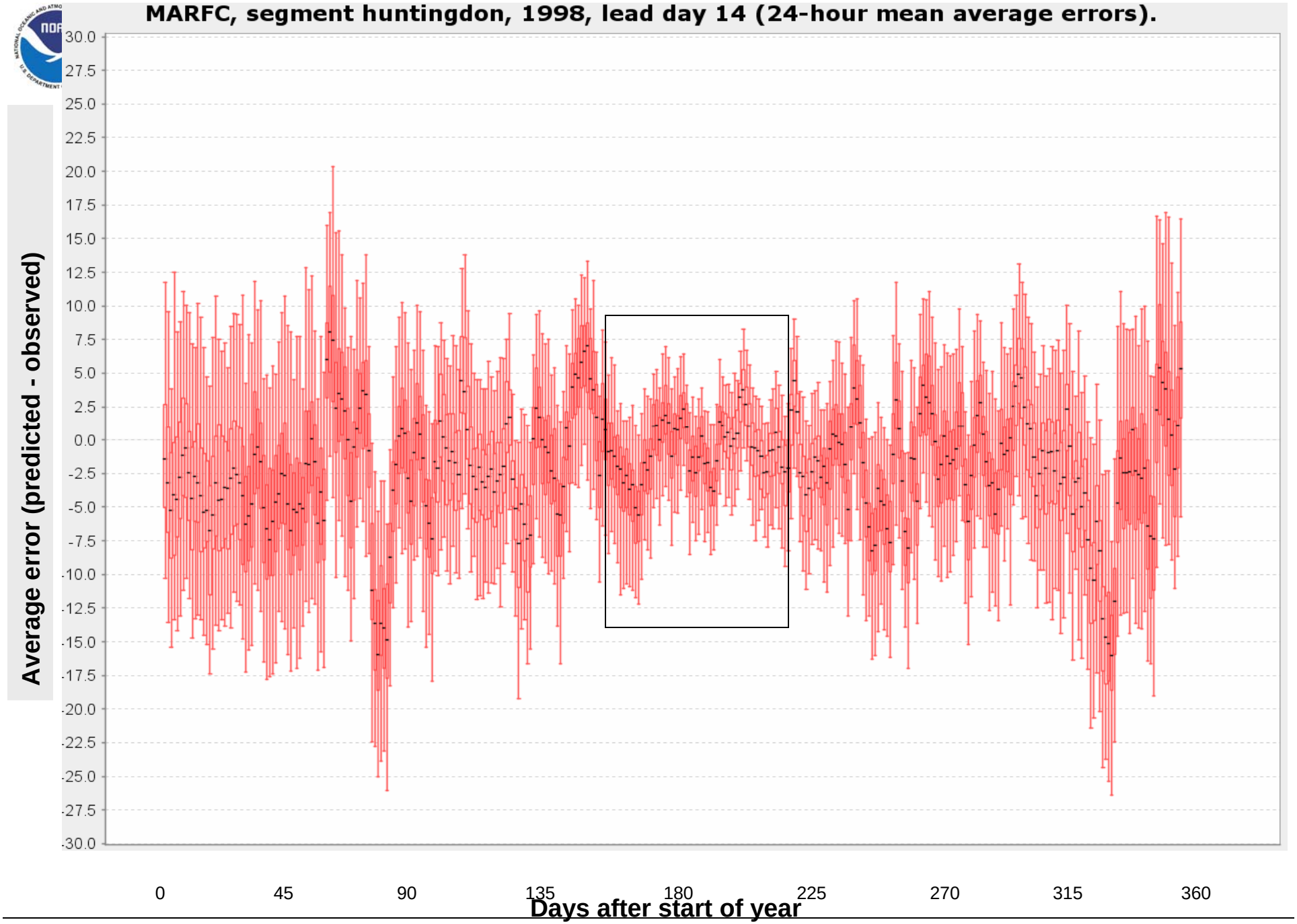
Temperature


MARFC, segment huntingdon, 1998, lead day 1 (24-hour mean average errors).
Average error (predicted - observed)
30.0
27.5
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-2.5
-5.0
-7.5
-10.0
-12.5
-15.0
-17.5
-20.0
-22.5
-25.0
-27.5
0
45
90
135
180
225
270
315
360
Days after start of year

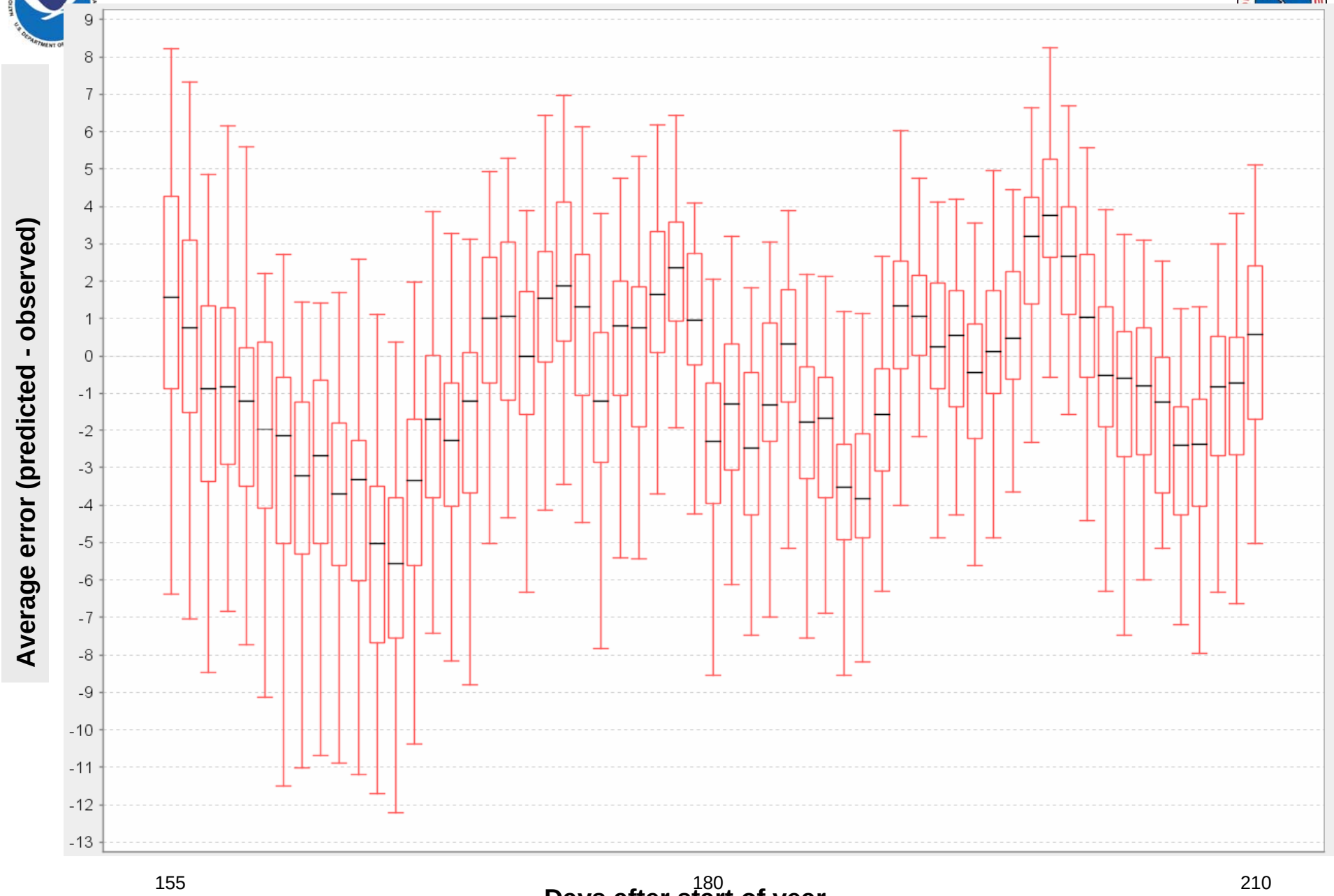

MARFC, segment huntingdon, 1998, lead day 2 (24-hour mean average errors).
Average error (predicted - observed)
30.0
27.5
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-2.5
-5.0
-7.5
-10.0
-12.5
-15.0
-17.5
-20.0
-22.5
-25.0
-27.5
-30.0
0
45
90
135
180
225
270
315
360
Days after start of year


MARFC, segment huntingdon, 1998, lead day 3 (24-hour mean average errors).
Average error (predicted - observed)
30.0
27.5
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-2.5
-5.0
-7.5
-10.0
-12.5
-15.0
-17.5
-20.0
-22.5
-25.0
-27.5
0
45
90
135
180
225
270
315
360
Days after start of year

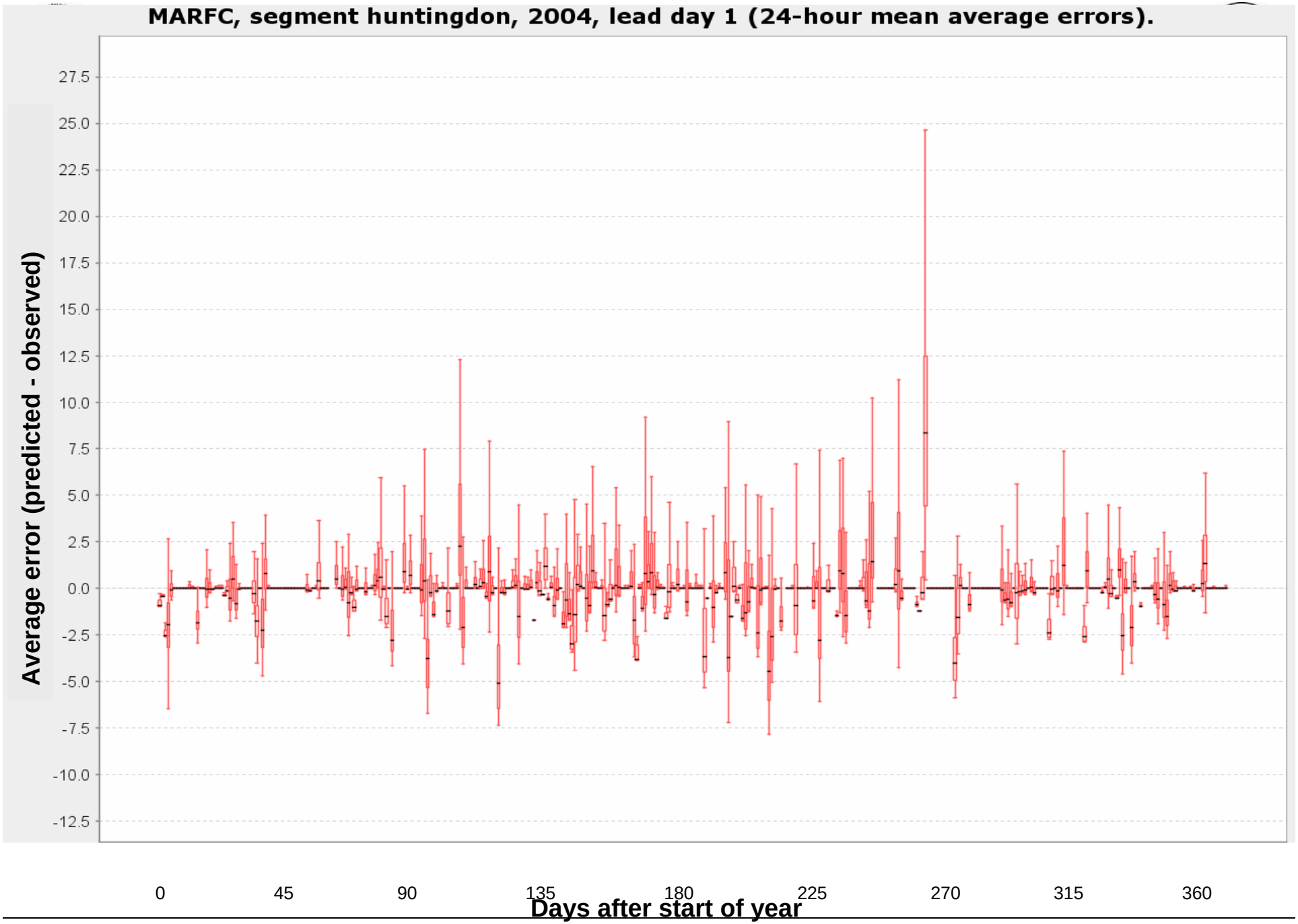

MARFC, segment huntingdon, 1998, lead day 4 (24-hour mean average errors).
Average error (predicted - observed)
30.0
27.5
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-2.5
-5.0
-7.5
-10.0
-12.5
-15.0
-17.5
-20.0
-22.5
-25.0
-27.5
-30.0
0
45
90
135
180
225
270
315
360
Days after start of year

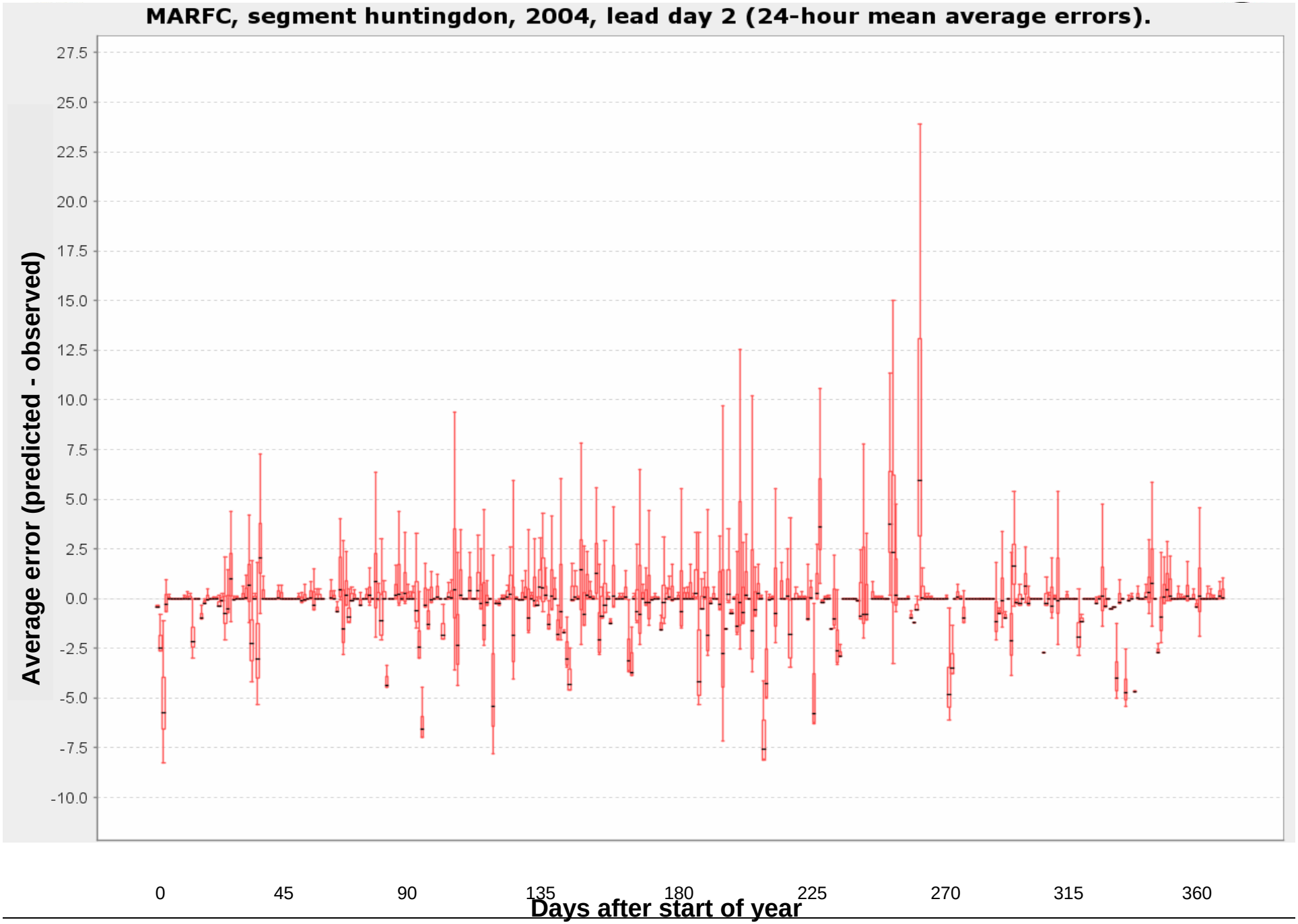

MARFC, segment huntingdon, 1998, lead day 14 (24-hour mean average errors).
Average error (predicted - observed)
30.0
27.5
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-2.5
-5.0
-7.5
-10.0
-12.5
-15.0
-17.5
-20.0
-22.5
-25.0
-27.5
-30.0
0
45
90
135
180
225
270
315
360
Days after start of year

MARFC, huntingdon, 1998, lead day 14, zoom on July/August



Precipitation


MARFC, segment huntingdon, 2004, lead day 1 (24-hour mean average errors).
Average error (predicted - observed)
27.5
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-2.5
-5.0
-7.5
-10.0
-12.5
0
45
90
135
180
225
270
315
360
Days after start of year


MARFC, segment huntingdon, 2004, lead day 2 (24-hour mean average errors).
Average error (predicted - observed)
27.5
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-2.5
-5.0
-7.5
-10.0
0
45
90
135
180
225
270
315
360
Days after start of year

Thank you