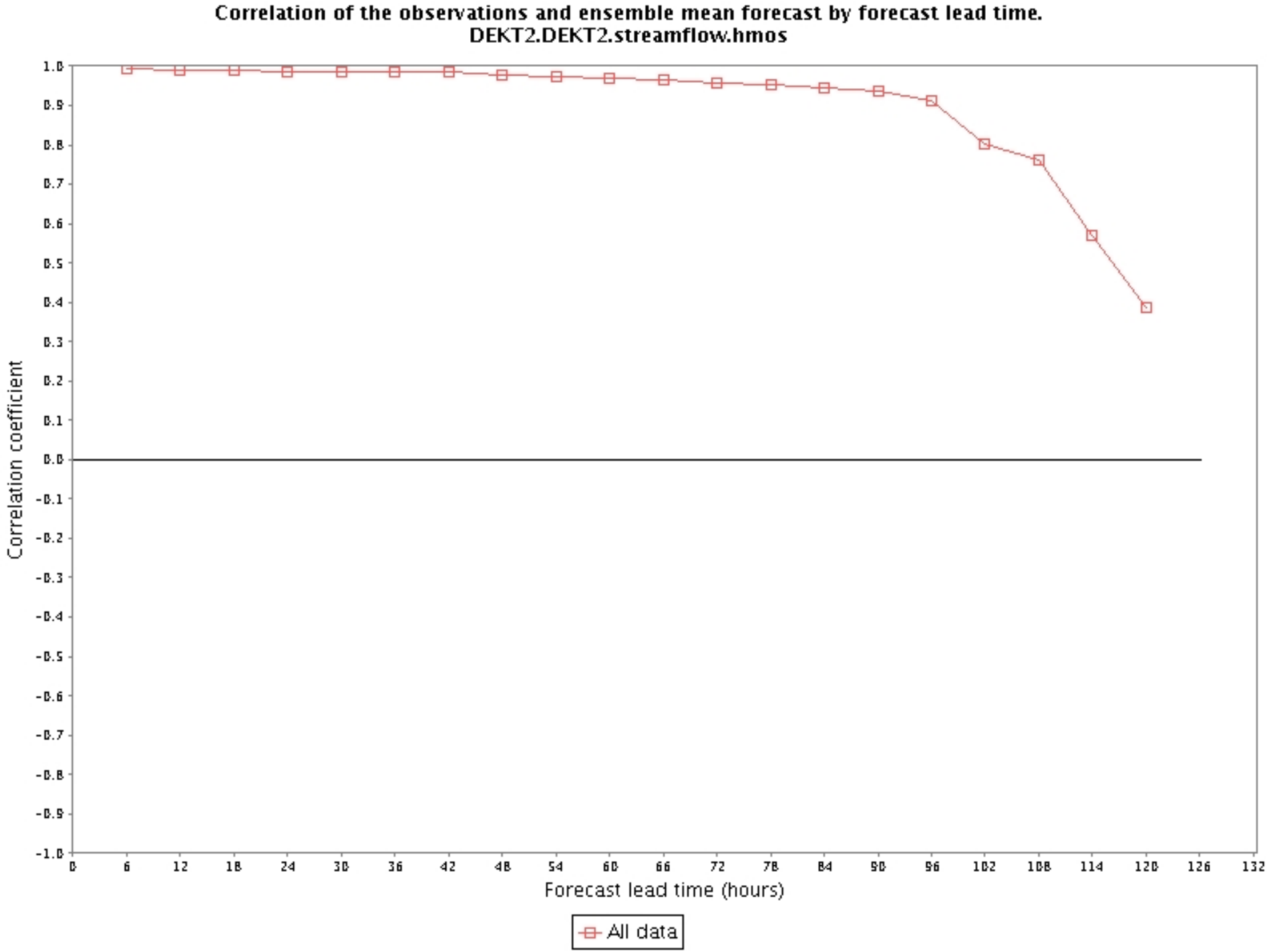
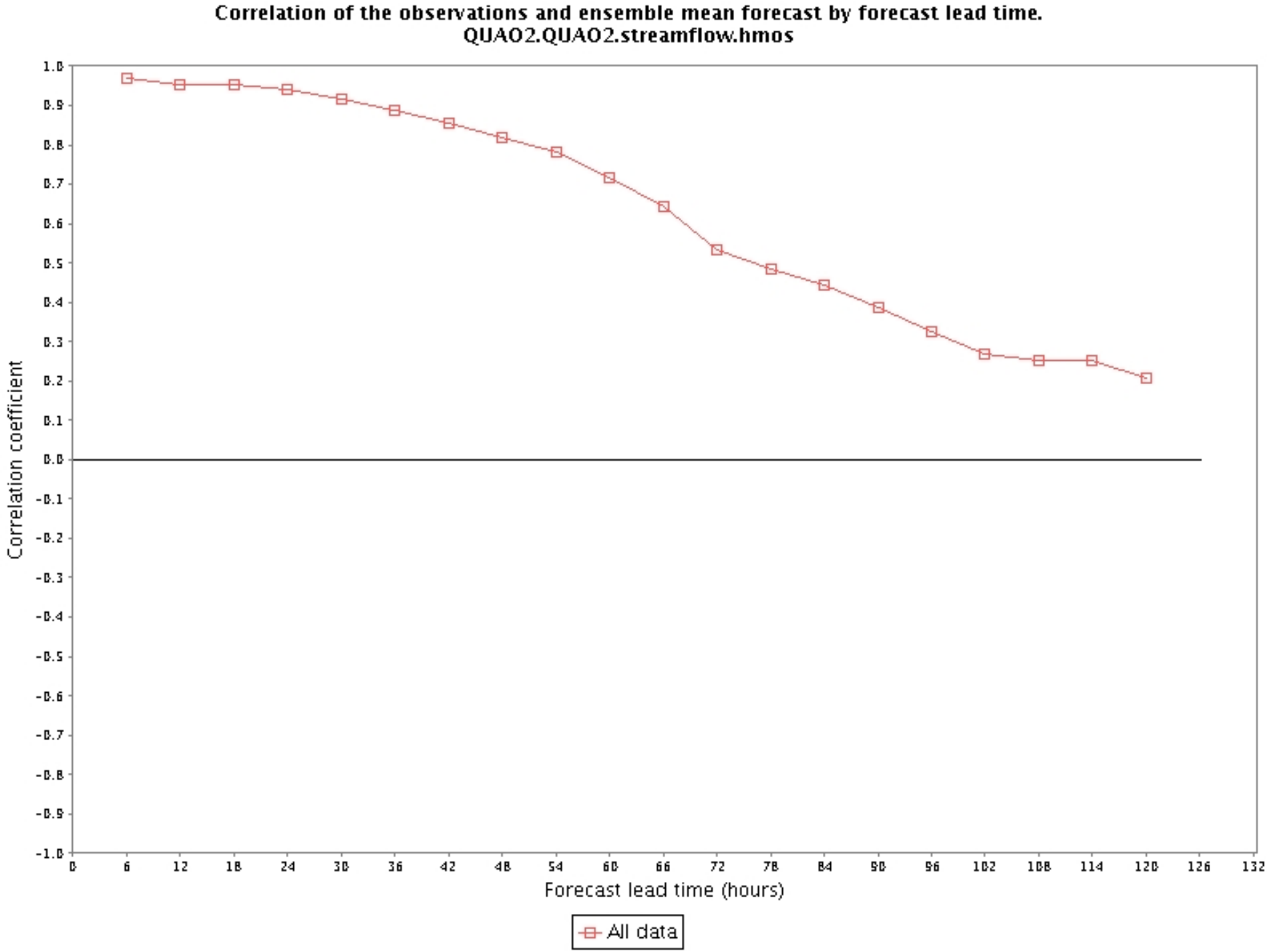
HMOS EVS results

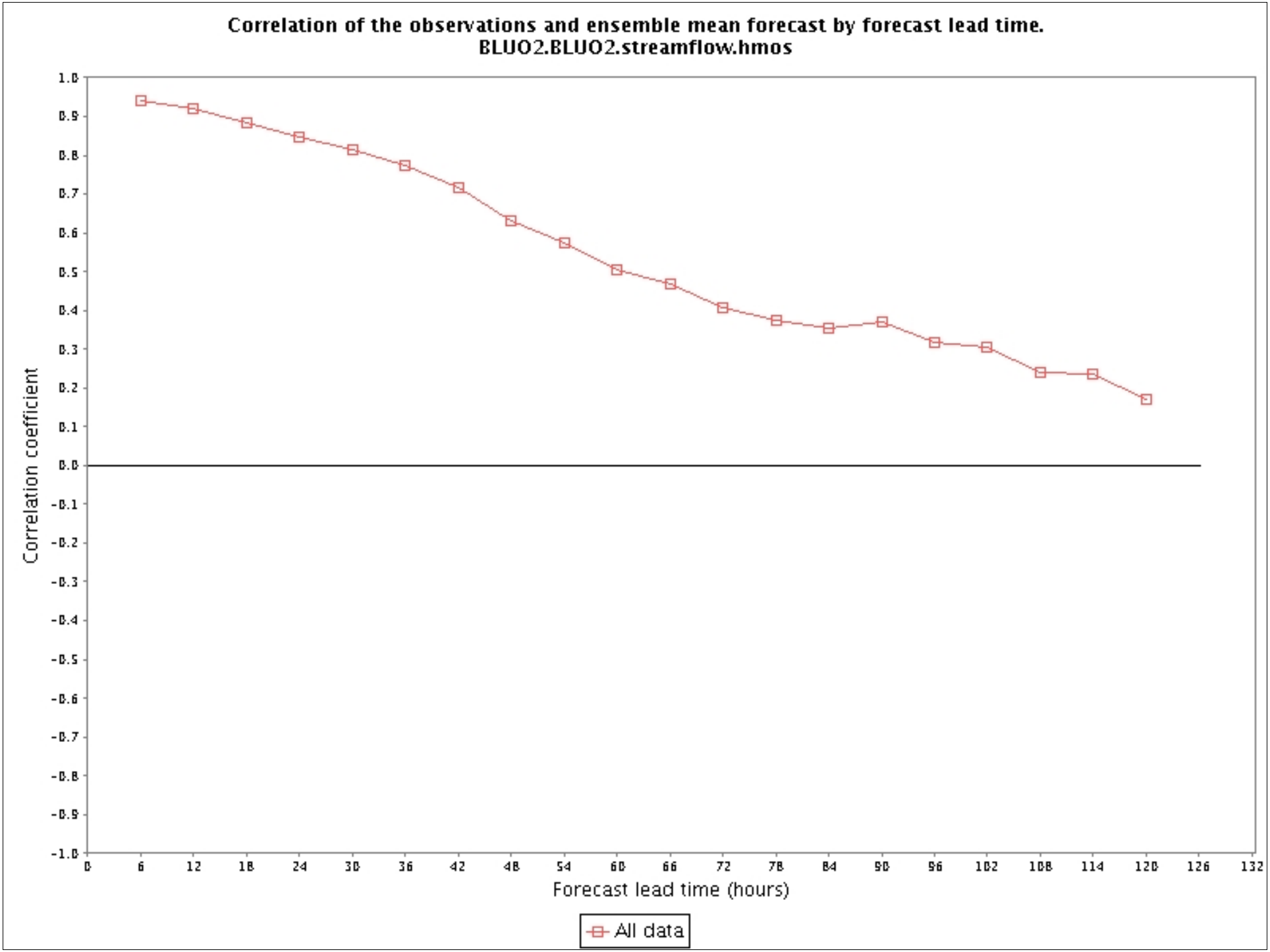
Bill Lawrence

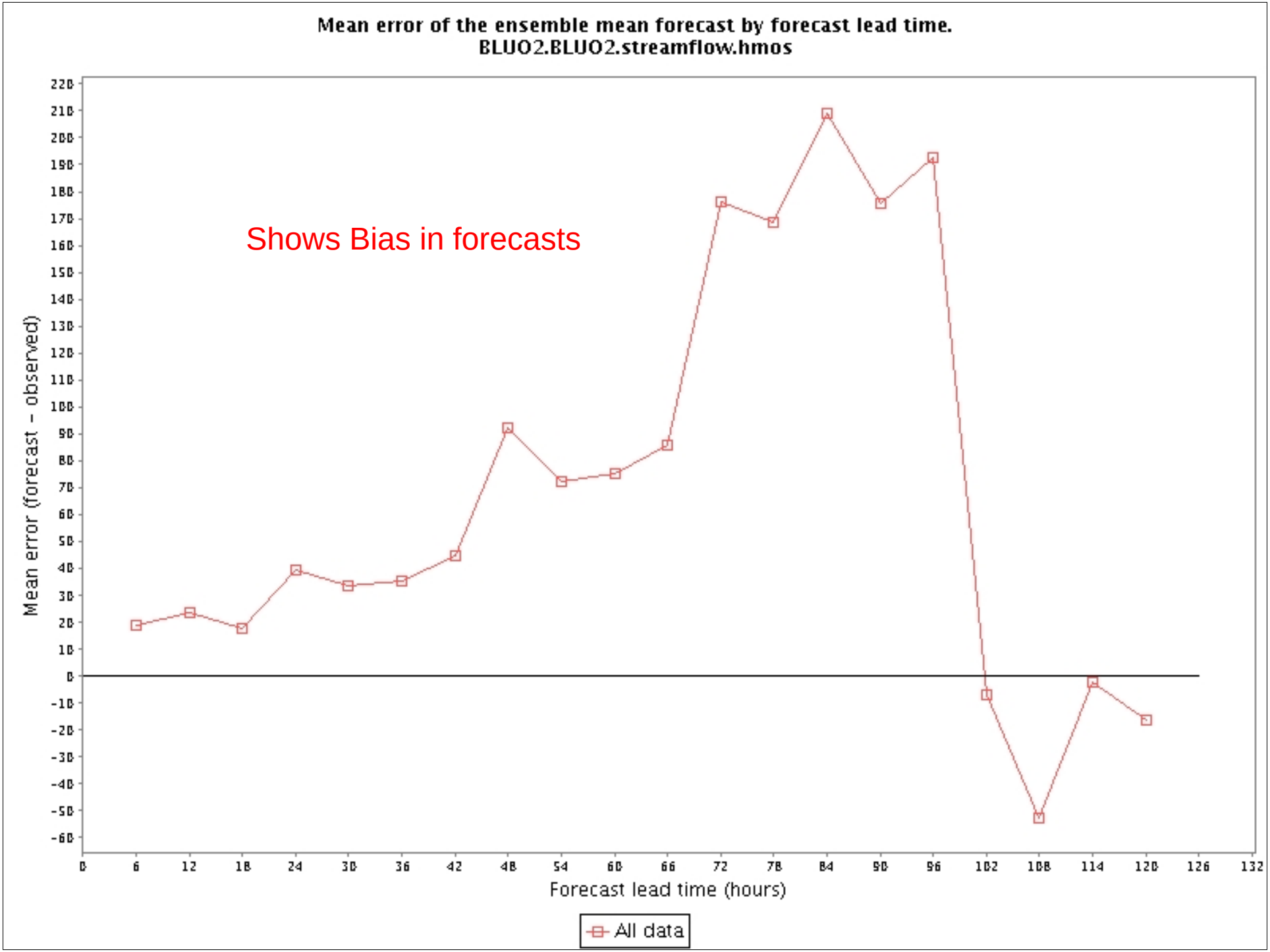
NWS Verification Team Meeting
11/10/2008

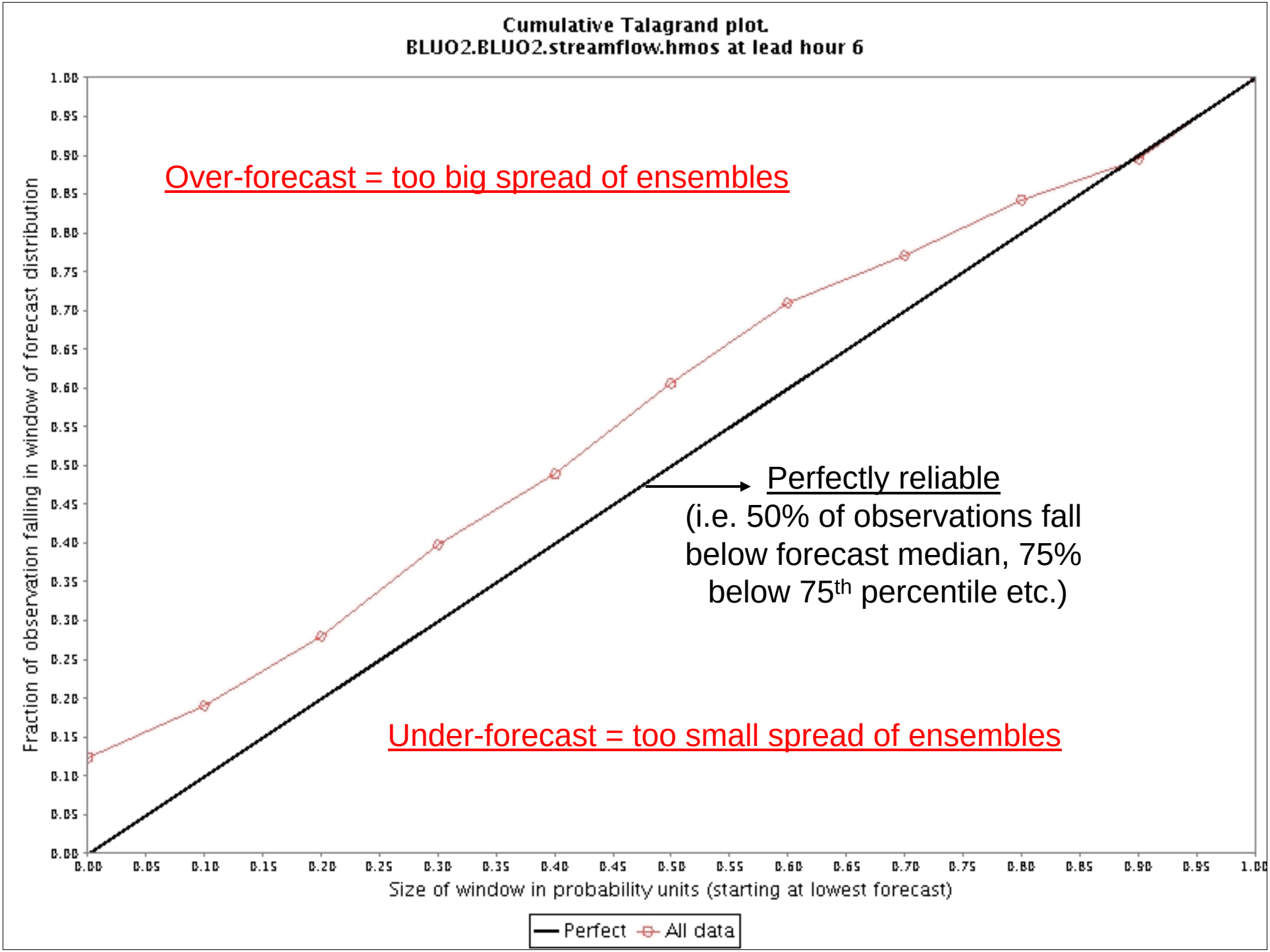
- HMOS run retrospectively for 11+ year period; Feb 97 to Oct 08
- 3 Points studied.
- BLUO2 – Headwater, smallest overall basin
- QUAO2 – Mid-side basin, not a headwater
- DEKT2 – Larger total basin, somewhat regulated.
- EVS used to study resulting .CS files

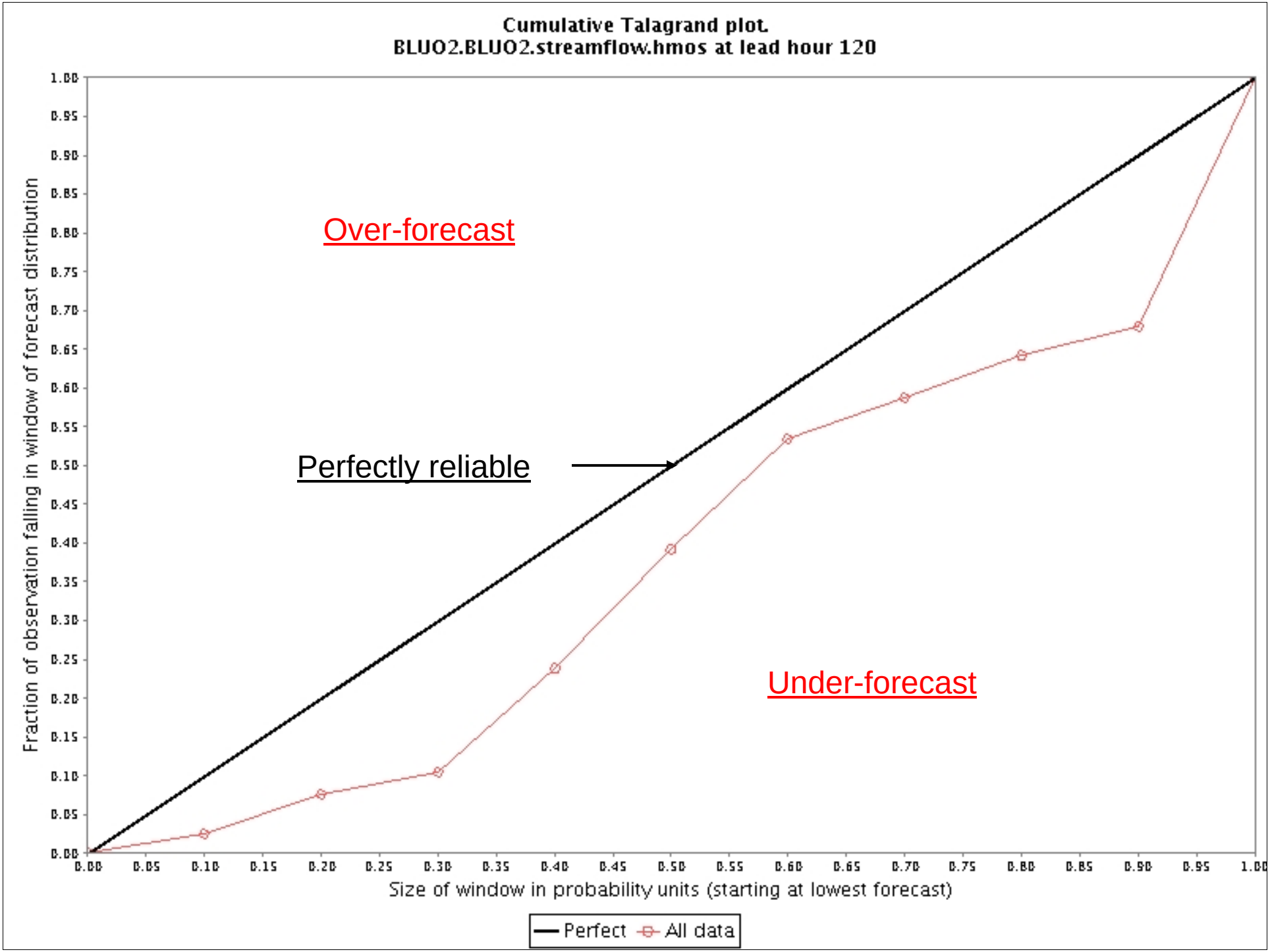

Correlation of the observations and ensemble mean forecast by forecast lead time.
DEKT2.DEKT2.streamflow.hmos
Correlation coefficient
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
-0.9
-1.0
0
6
12
18
24
30
36
42
48
54
60
66
72
78
84
90
96
102
108
114
120
126
132
Forecast lead time (hours)
All data

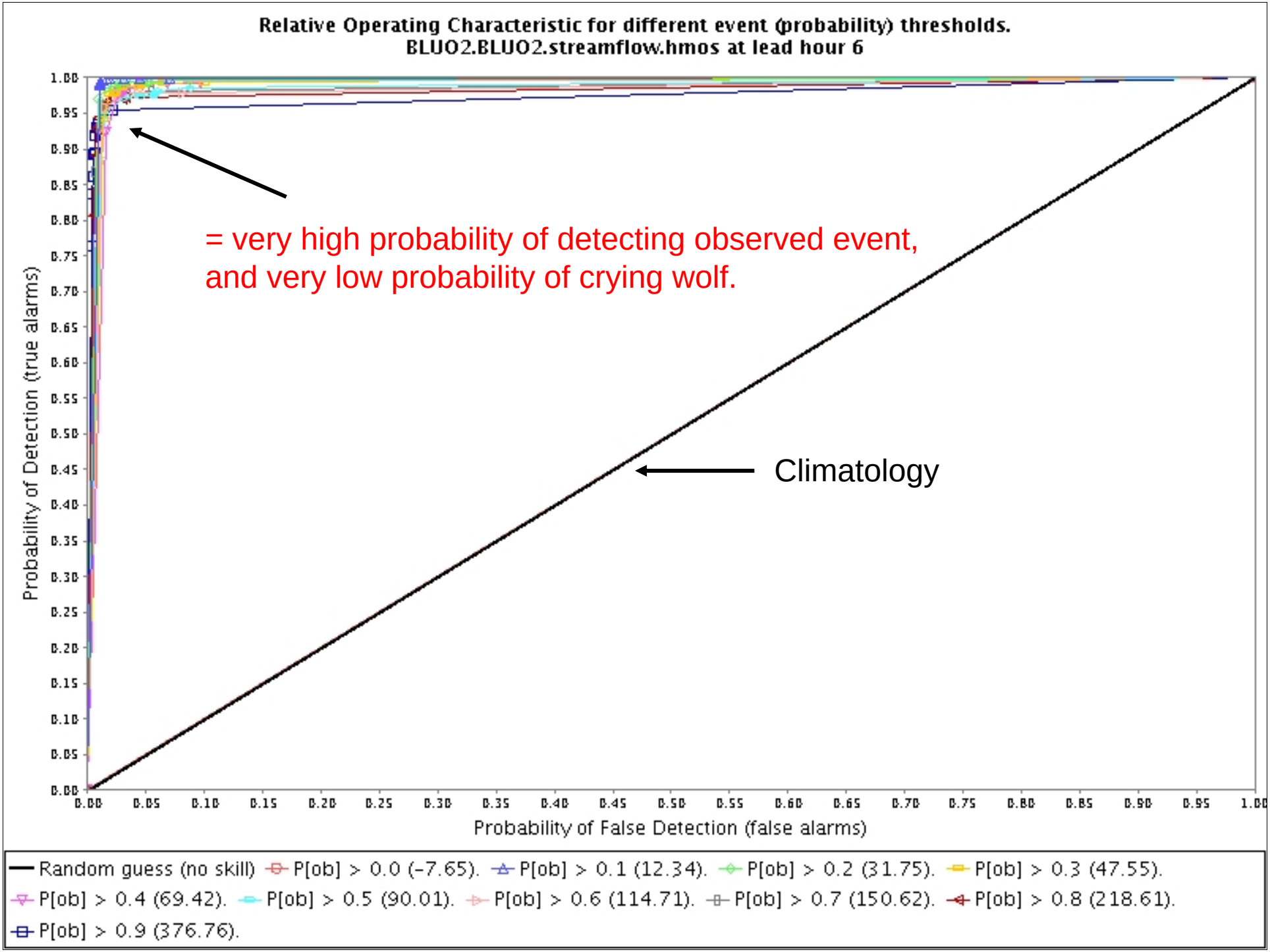

Correlation of the observations and ensemble mean forecast by forecast lead time.
QUAO2.QUAO2.streamflow.hmos
Correlation coefficient
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
-0.9
-1.0
0
6
12
18
24
30
36
42
48
54
60
66
72
78
84
90
96
102
108
114
120
126
132
Forecast lead time (hours)
All data

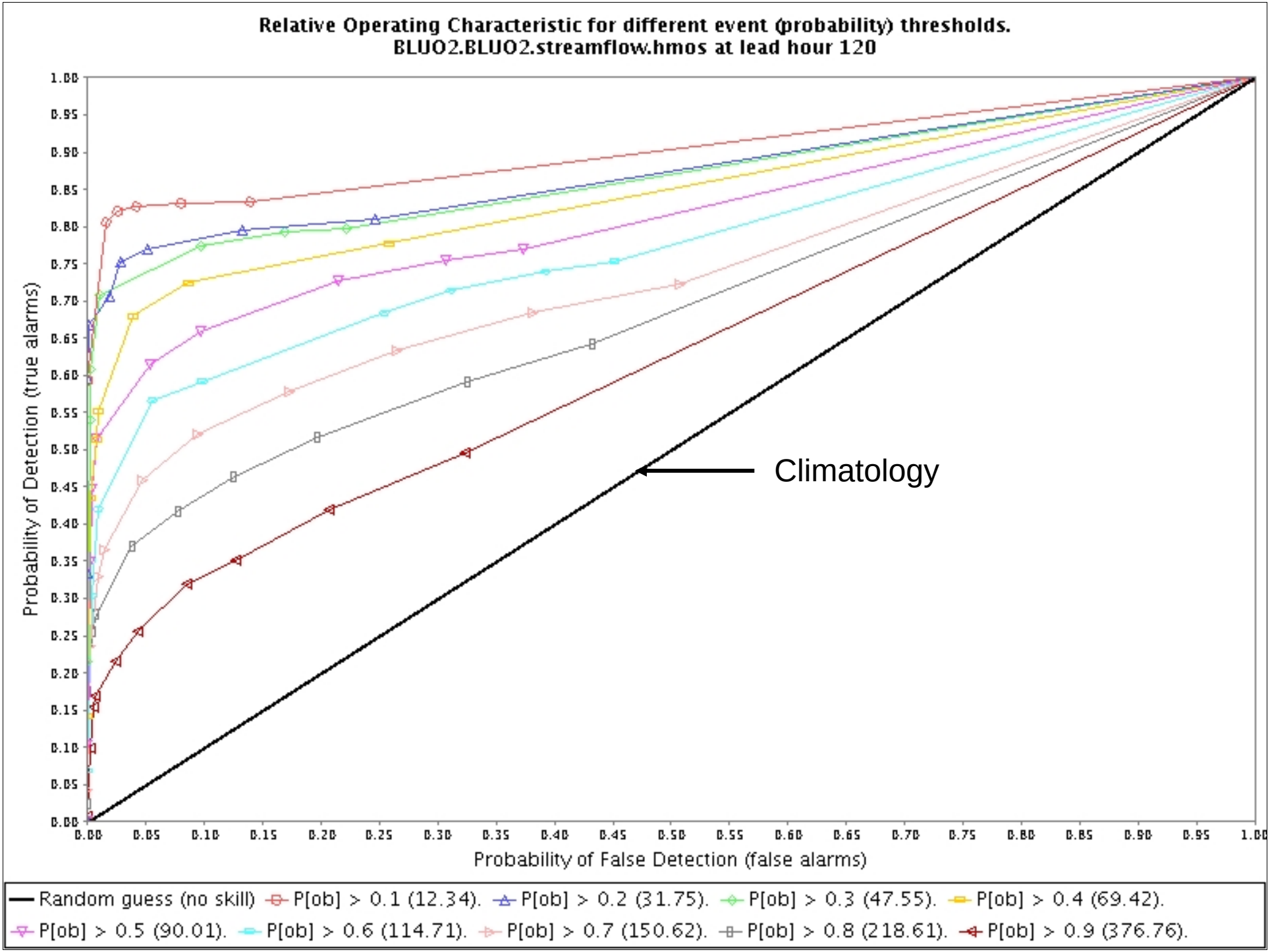

Correlation of the observations and ensemble mean forecast by forecast lead time.
BLUO2.BLUO2.streamflow.hmos
Correlation coefficient
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
-0.9
-1.0
0
6
12
18
24
30
36
42
48
54
60
66
72
78
84
90
96
102
108
114
120
126
132
Forecast lead time (hours)
All data

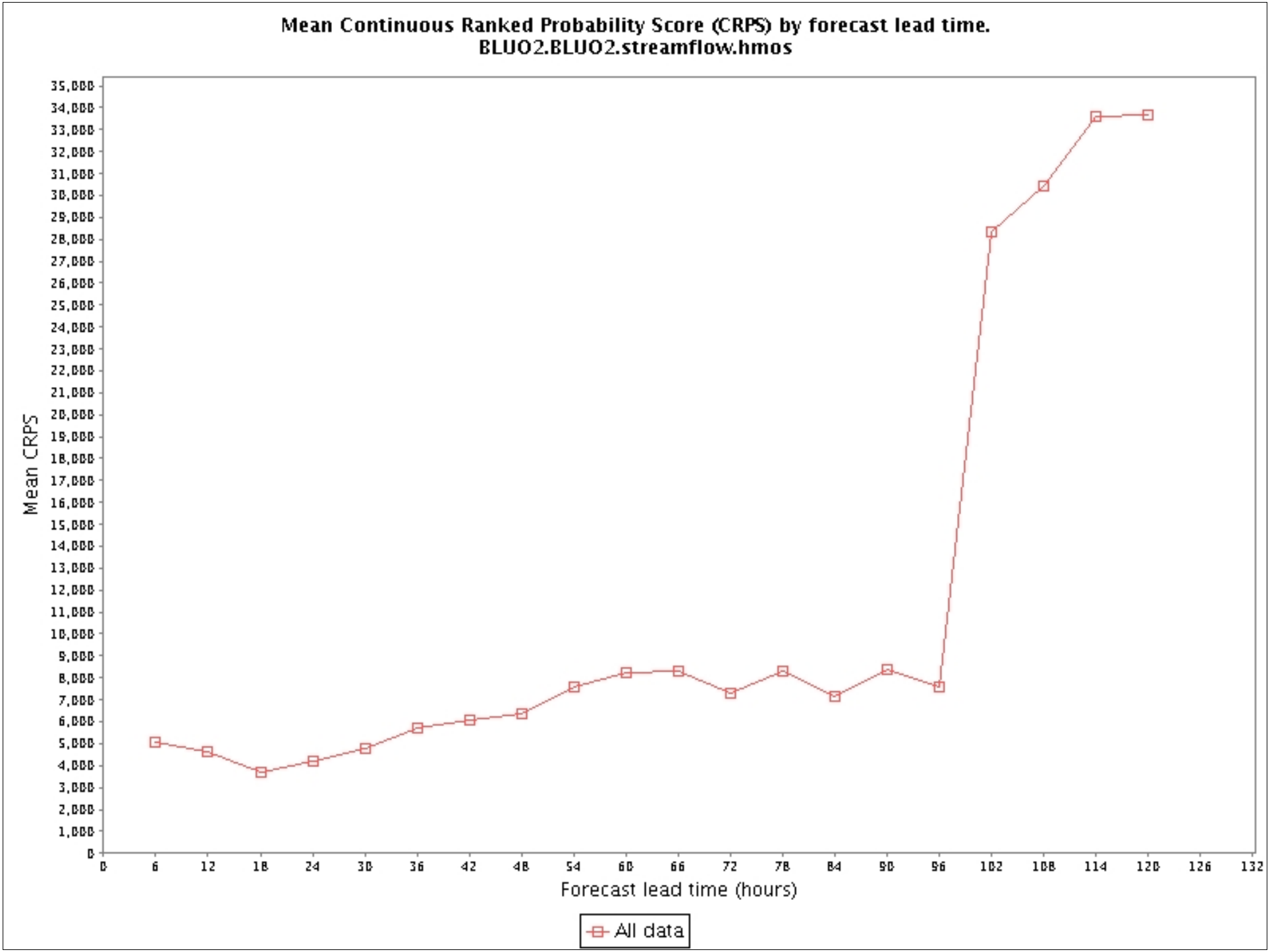

Mean error of the ensemble mean forecast by forecast lead time.
BLUO2.BLUO2.streamflow.hmos
Shows Bias in forecasts
Mean error (forecast - observed)
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
-20
-30
-40
-50
-60
0
6
12
18
24
30
36
42
48
54
60
66
72
78
84
90
96
102
108
114
120
126
132
Forecast lead time (hours)
All data


Cumulative Talagrand plot.
BLUO2.BLUO2.streamflow.hmos at lead hour 6
Fraction of observation falling in window of forecast distribution
Size of window in probability units (starting at lowest forecast)
Perfect
All data
Over-forecast = too big spread of ensembles
Perfectly reliable
(i.e. 50% of observations fall below forecast median, 75% below 75th percentile etc.)
Under-forecast = too small spread of ensembles


Cumulative Talagrand plot.
BLUO2.BLUO2.streamflow.hmos at lead hour 120
Fraction of observation falling in window of forecast distribution
Size of window in probability units (starting at lowest forecast)
Over-forecast
Perfectly reliable
Under-forecast
Perfect
All data


Relative Operating Characteristic for different event (probability) thresholds.
BLUO2.BLUO2.streamflow.hmos at lead hour 6
= very high probability of detecting observed event,
and very low probability of crying wolf.
Climatology
Probability of Detection (true alarms)
Probability of False Detection (false alarms)
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00
Random guess (no skill)
P[ob] > 0.0 (-7.65).
P[ob] > 0.1 (12.34).
P[ob] > 0.2 (31.75).
P[ob] > 0.3 (47.55).
P[ob] > 0.4 (69.42).
P[ob] > 0.5 (90.01).
P[ob] > 0.6 (114.71).
P[ob] > 0.7 (150.62).
P[ob] > 0.8 (218.61).
P[ob] > 0.9 (376.76).


Relative Operating Characteristic for different event (probability) thresholds.
BLUO2.BLUO2.streamflow.hmos at lead hour 120
Probability of Detection (true alarms)
Probability of False Detection (false alarms)
Climatology
Random guess (no skill)
P[ob] > 0.1 (12.34).
P[ob] > 0.2 (31.75).
P[ob] > 0.3 (47.55).
P[ob] > 0.4 (69.42).
P[ob] > 0.5 (90.01).
P[ob] > 0.6 (114.71).
P[ob] > 0.7 (150.62).
P[ob] > 0.8 (218.61).
P[ob] > 0.9 (376.76).


Mean Continuous Ranked Probability Score (CRPS) by forecast lead time.
BLUO2.BLUO2.streamflow.hmos
Mean CRPS
35,000
34,000
33,000
32,000
31,000
30,000
29,000
28,000
27,000
26,000
25,000
24,000
23,000
22,000
21,000
20,000
19,000
18,000
17,000
16,000
15,000
14,000
13,000
12,000
11,000
10,000
9,000
8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000
0
0 6 12 18 24 30 36 42 48 54 60 66 72 78 84 90 96 102 108 114 120 126 132
Forecast lead time (hours)
All data

Conclusions/Results

Preliminary results appear to show that HMOS provides discriminate fairly well, as evident on ROC Charts.

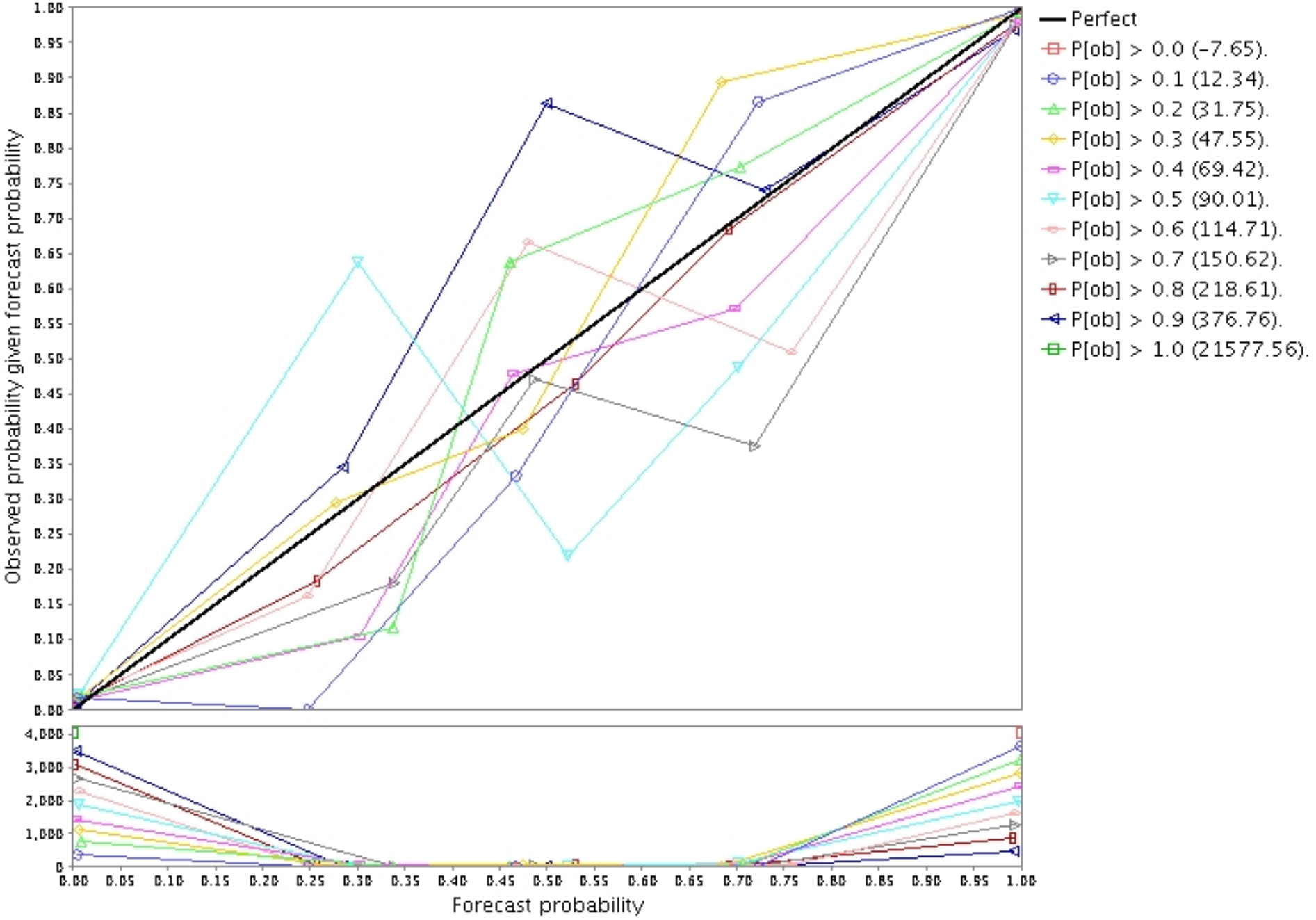
- Preliminary results appear to show that HMOS provides fairly reliable forecasts by viewing the Cumulative Talagrand Charts
- Both reliability and discrimination tends to suffer the further out in time the forecast is for.
- EVS gives metrics for both ensembles (probabilistic) and mean fcsts (deterministic)

Conclusions/Results

- Results appear better at larger basins, where local effects tend to be dampened by routed water, etc.
- A bit disturbed about the mean error value (that shows bias), as they appear quite high.
- Something odd occurring in day 5 results, not quite sure what is causing these odd results.
- Ensemble metrics for HMOS are much better than original short term ensemble method used by ABRFC, MARFC and CNRFC.

Additional Slides

Reliability diagram for various event thresholds (upper) and sample counts (lower).
BLUO2.BLUO2.streamflow.hmos at lead hour 6



Reliability diagram for various event thresholds (upper) and sample counts (lower).
BLUO2.BLUO2.streamflow.hmos at lead hour 120