



4th International Verification Methods Workshop, Helsinki, 8th June 2009

The Ensemble Verification System (EVS): a software tool for verifying ensemble forecasts at discrete locations

James Brown, Dong-Jun Seo, Julie
Demargne, Yuqiong Liu

james.d.brown@noaa.gov



1. What is the EVS and why is it needed?

What is the EVS?

Java tool for diagnostic verification

- Diagnose biases to improve models ('offline')
- Distinct from real-time verification ('online')

Types of forecasts/observations handled

- RFC ens. forecasts are mainly time-series
- Continuous numeric variables
- Issued at discrete locations (or areas)
- Any forecast lead time and frequency



Why is the EVS needed?

Had to be user-friendly (operations)

- 13 NWS RFCs, 4 use ensemble forecasts
- They need a tool to test/justify ensembles

Had to be flexible, but user-oriented

- Flexible conditioning (e.g. flow when $< 0^{\circ}\text{C}$)
- Metrics of varying detail (user-focused help)

Had to be simple to add new metrics

- Object-oriented and well-documented



2. How does the EVS work?



How does the EVS work?

Two methods of execution (on any OS)

- GUI and command line. GUI is structured....

3. Verification (at specific locations)

- Specify locations, data sources, metrics etc.

5. Aggregation (across locations)

- Locations with certain common properties

3. Output (graphical and numerical)

File Help



Verification Aggregation Output

Three stages (tabbed panes)

Verification metrics to compute

Metrics to compute

Name	Property verified	Include?
Correlation coefficient	Ensemble mean	<input checked="" type="checkbox"/>
Mean error	Ensemble mean	<input checked="" type="checkbox"/>
Root mean squared error	Ensemble mean	<input checked="" type="checkbox"/>
Brier score	Ensemble distribution	<input checked="" type="checkbox"/>
Mean continuous ranked probability score	Ensemble distribution	<input checked="" type="checkbox"/>
Mean error of probability diagram	Ensemble distribution	<input checked="" type="checkbox"/>
Mean capture rate diagram	Ensemble distribution	<input checked="" type="checkbox"/>
Modified box plot pooled by lead time	Ensemble distribution	<input checked="" type="checkbox"/>
Modified box plot per lead time by observed value	Ensemble distribution	<input checked="" type="checkbox"/>
Relative operating characteristic	Ensemble distribution	<input checked="" type="checkbox"/>
Relative operating characteristic score	Ensemble distribution	<input checked="" type="checkbox"/>

Explanation of metric 'Mean continuous ranked probability score'

MEAN CONTINUOUS RANKED PROBABILITY SCORE (CRPS)

The CRPS summarizes the quality of a continuous probability forecast with a single number (a score). It measures the integrated squared difference between the cumulative distribution function (cdf) of a forecast, $F_Y(y)$ and the corresponding cdf of the observation, $\mathbf{1}\{\}$:

$$CRPS(x, F_Y) = \int_{-\infty}^{\infty} (F_Y(y) - \mathbf{1}(y \geq x))^2 dy$$

where $\mathbf{1}\{\}$ is a step function that reaches probability 1.0 for values greater than or equal to the observation, and has probability 0.0 otherwise. In practice, the CRPS is averaged across a number, n , of paired forecasts and observations, which leads to the mean CRPS:

$$\overline{CRPS} = 1/n \sum^n CRPS(x, F_Y)$$

Parameters of metric 'Mean continuous ranked probability score'

Edit thresholds [optional]

Threshold values
All data
0.0
0.05
0.1
...

Add

Delete

Metrics

Details of selected metric.

Basic params. of selected metric

Navigation

More

Save

Run

All

Back

Next



3. Example application



Ensemble flow forecasts

NWS-ESP forecasts at “NFDC1”

- North Fork of the American River in CA (dam)
- Daily mean inflows forecast from 1979-2002
- 14 forecast lead days, 50 ensemble members

Summary of verification results

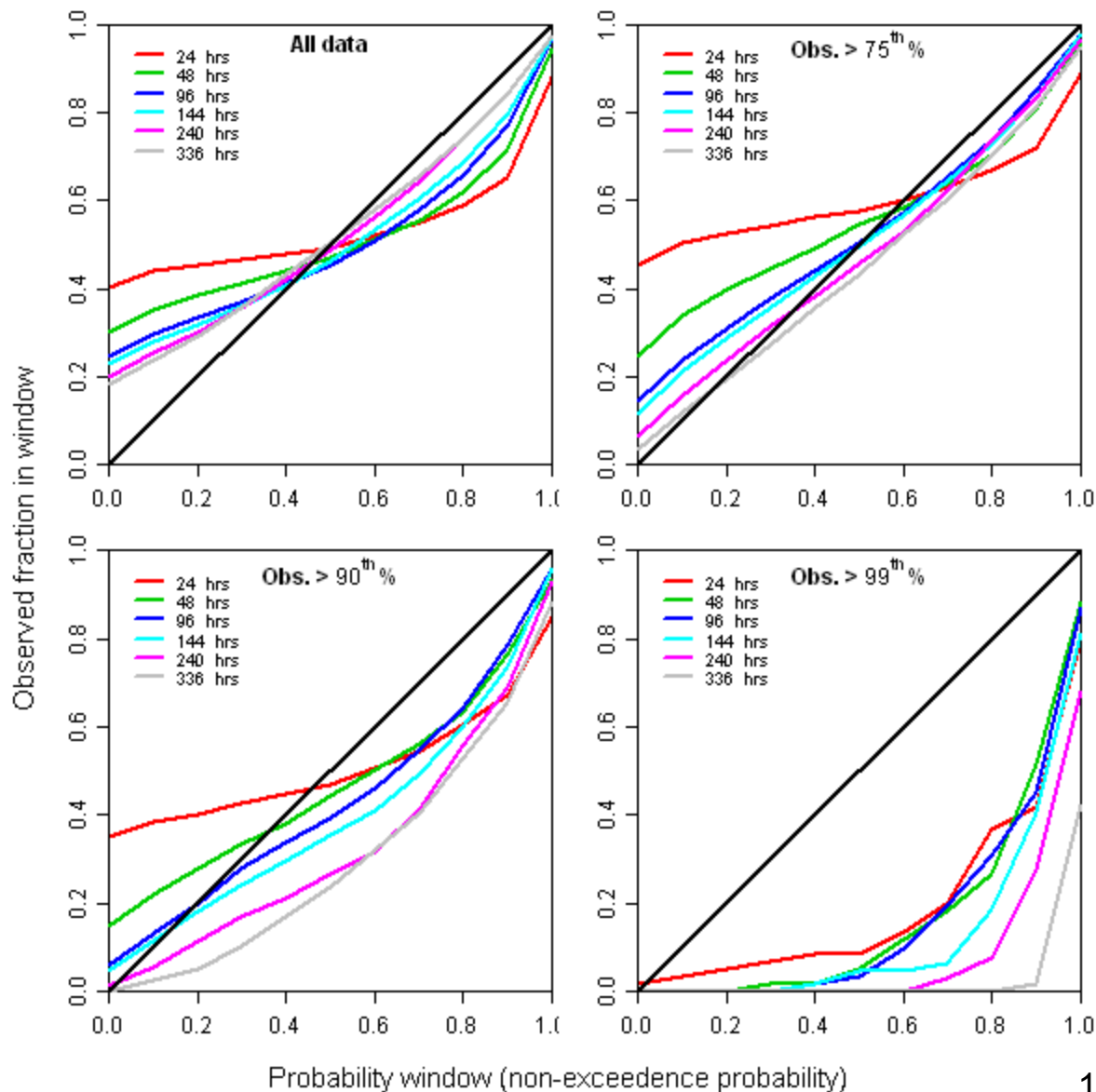
- Reliable/discriminatory for moderate flow
- But: conditional biases effect low/high flows
- Illustrated with two ‘unconventional’ metrics

Spread-bias plots

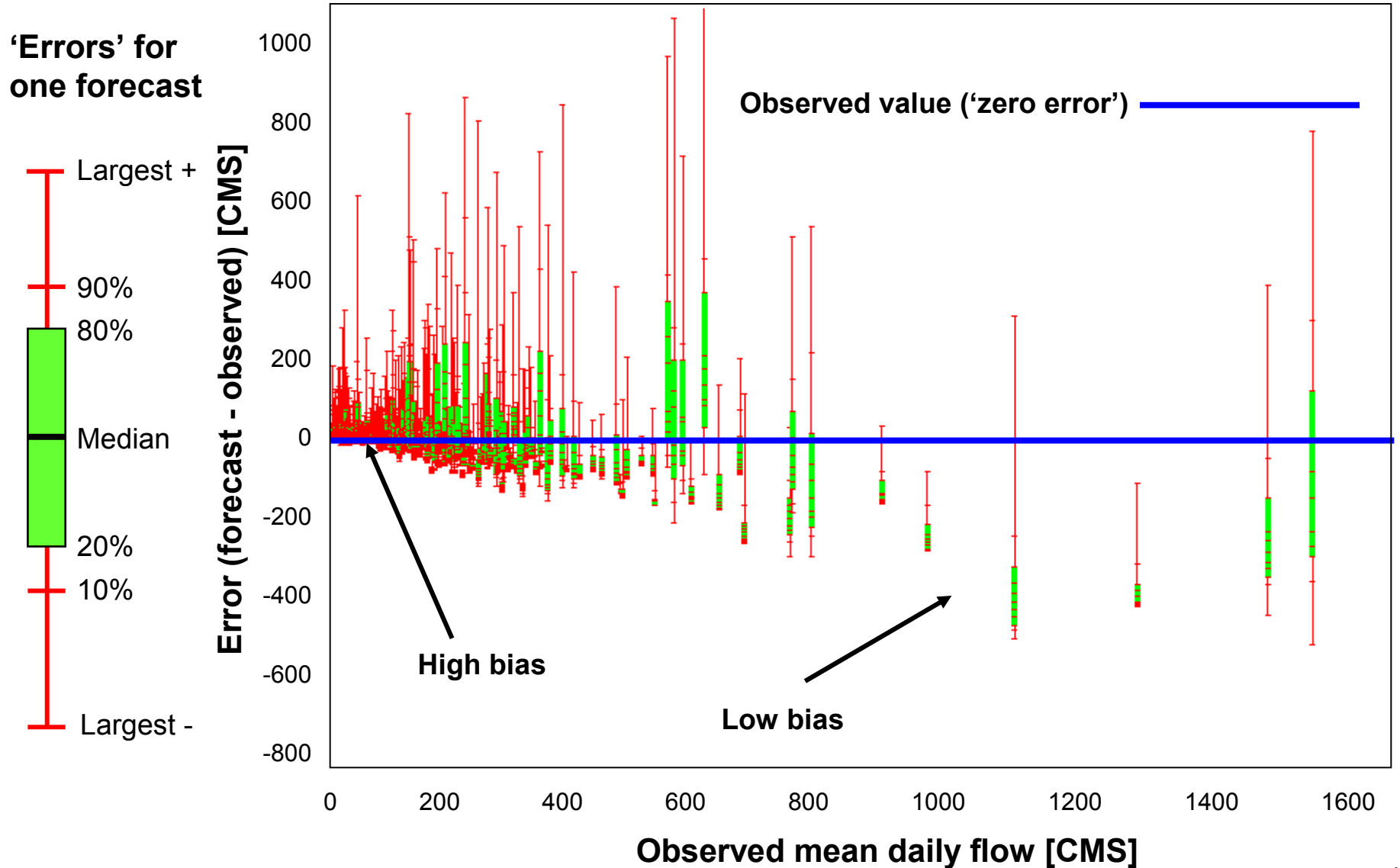
- **Measures reliability**
- **Similar to Talagrand**
- **Define interval of constant prob. W_Y , on support of forecast,**

$$W_Y = [a, b | a, b \in [0, 1]]$$
- **Count fraction of times obs., x^o , falls in W_Y over n pairs**

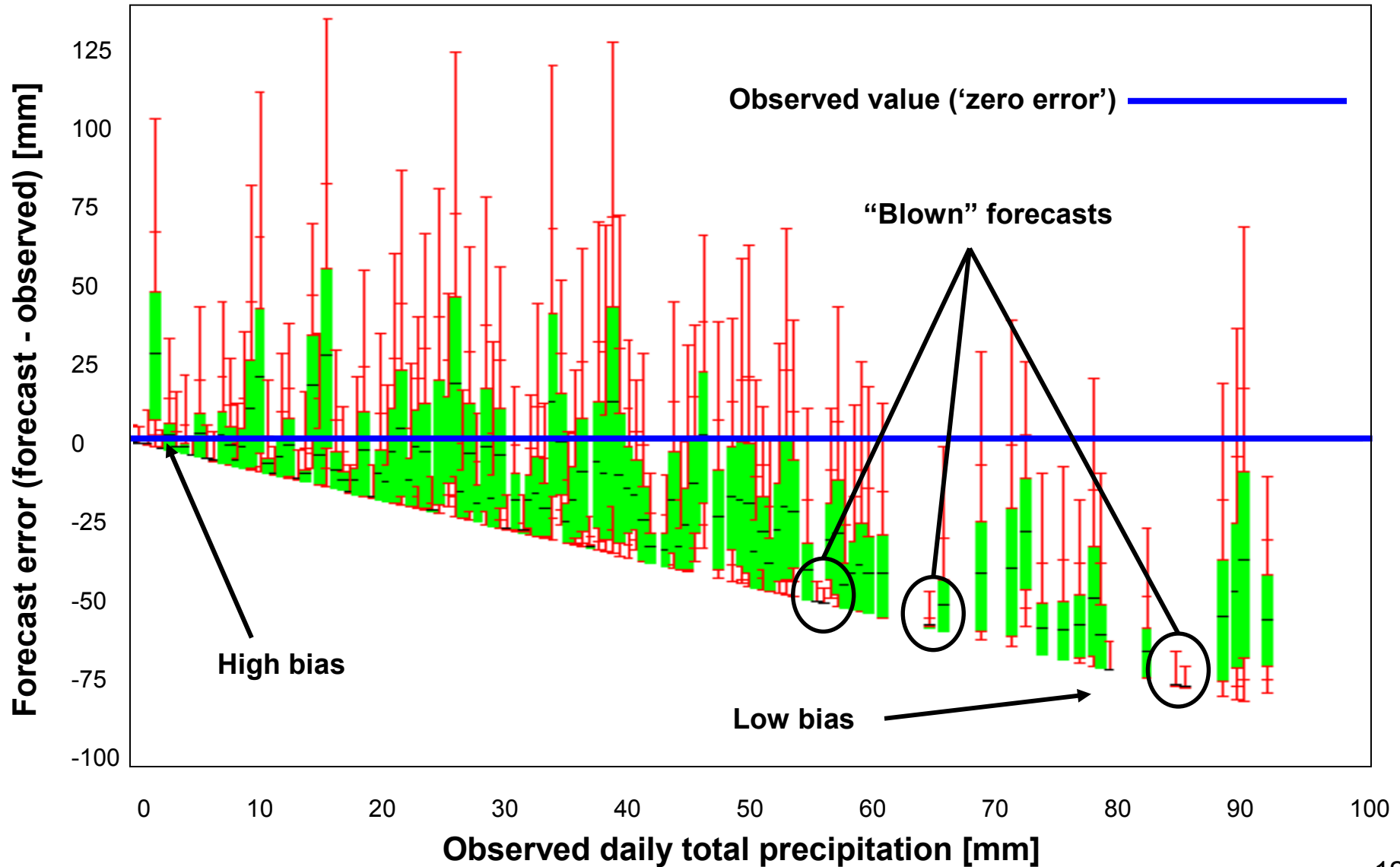
$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}\{F_{Y_i}(x_i^o) \in W_Y\}$$
- **Should be equal to width of W_Y**
- **Repeat for all W_Y ; plot**



Box plot of flow errors (day 1)



Precipitation (day 1, NFDC1)



Next steps

To make EVS widely used (beyond NWS)

- Paper for EMS. Use in HEPEX Ver. Testbed

Current research (two examples)

- 1) Operational hydrology driven by extremes
 - Visualization of “raw” errors (e.g. Bradley)
 - Quantification of sampling uncertainties
- 7) Error source analysis (where to improve?)
 - Hydrograph timing errors (X-wavelets)



Additional slides

File Help



Properties of selected location

Verification Aggregation Output

Verification unit

Unique identifier
nfdc1hlf.Precipitation
nfdc1huf.Precipitation

Basic properties of verification unit 'nfdc1hlf.Precipitation'

Identifiers (right click for defaults)

Location identifier:

Additional identifier [optional]:

Environmental variable identifier:

Input data

Files or folder containing forecast data:

Time zone of forecasts:

File containing observed data:

Time zone of observations:

Verification window

Start of verification period (in forecast time system):

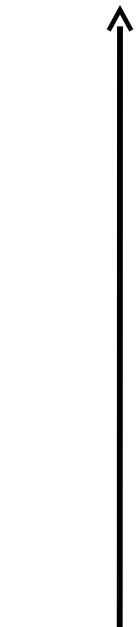
Forecast lead period:

End of verification period (in forecast time system):

Aggregation period [optional]:

Output data

Folder for output statistics:



Locations



Data sources



Verification parameters



Output data

Add Delete Copy Save

File Help



Common properties of discrete locations

Verification Aggregation Output

Candidate aggregation unit(s)

Unique identifier

nfdc1

Properties of 'nfdc1'

Parameter values

Aggregation unit identifier

nfdc1

Environmental variable identifier

Precipitation

Forecast lead period

14 DAY

Aggregation of lead period

24 HOUR

Start of verification period (YYYY/MM/DD)

1979

1

1

End of verification period (YYYY/MM/DD)

1996

12

31

Verification units to include in aggregation

Available units (specify S to weigh by sample size)

Unique identifier	Weight	Include?
nfdc1 hlf.Precipitation	0.5	<input checked="" type="checkbox"/>
nfdc1 huf.Precipitation	0.5	<input checked="" type="checkbox"/>

Aggregation units

Verification units
(discrete locations)

Output data

Folder for aggregated statistics

C:\Documents and Settings\James Brown\Desktop\EVS_demo_05_



Output data location

Delete

Save

Run

Back

Next

File Help



Verification Aggregation Output

Output of results

Units with results (verification and aggregation units)

Name	Unit type
nfdc1.hlf.Precipitation	VERIFICATION
nfdc1.huf.Precipitation	VERIFICATION
nfdc1	AGGREGATION

Verification / Aggregation units

Metrics for selected unit

Products for selected unit

Product	Include?
Brier score	<input type="checkbox"/>
Correlation coefficient	<input checked="" type="checkbox"/>
Mean capture rate diagram	<input type="checkbox"/>
Mean continuous ranked probability score	<input type="checkbox"/>
Mean error	<input type="checkbox"/>
Mean error of probability diagram	<input type="checkbox"/>
Modified box plot per lead time by observed value	<input type="checkbox"/>
Modified box plot pooled by lead time	<input type="checkbox"/>
Relative operating characteristic	<input type="checkbox"/>

Forecast lead times for selected product

Lead time (hours)	Include?
24.0	<input checked="" type="checkbox"/>
48.0	<input checked="" type="checkbox"/>
72.0	<input checked="" type="checkbox"/>
96.0	<input type="checkbox"/>
120.0	<input type="checkbox"/>
144.0	<input type="checkbox"/>
168.0	<input type="checkbox"/>
192.0	<input type="checkbox"/>
216.0	<input type="checkbox"/>

Lead times available

Output options

Output options

Write Display

Graphical output

Write graphical output

Output format:

JPEG file (*.jpg)

Options for selected output format:

Image width (pixels): 800

Image height (pixels): 600

Numerical output

Write numeric

Output format:

XML file (*.xml)

Run

Back

Start

Metrics available

Metric name	Quality attribute tested	Discrete events?	Detail
Mean error	Ensemble mean (deterministic)	No	Lowest
RMSE	Ensemble mean (deterministic)	No	Lowest
Correlation coefficient	Ensemble mean (deterministic)	No	Lowest
Brier Score	Lumped error score	Yes	Low
Brier Skill Score	Lumped error score vs. climatology	Yes	Low
Mean CRPS	Lumped error score	No	Low
Mean CRPS reliability	Lumped reliability score	No	Low
Mean CRPS resolution	Lumped resolution score	No	Low
CRPSS	Lumped error score vs. climatology	No	Low
ROC score	Lumped discrimination score	Yes	Low
Mean error in prob. plot	Reliability (unconditional bias)	No	Low
Spread-bias diagram	Reliability (conditional bias)	No	High
Reliability diagram	Reliability (conditional bias)	Yes	High
ROC diagram	Discrimination	Yes	High
Modified box plots	Error visualization	No	Highest