

THIS LOG CONTAINS A LIST OF CHANGES BETWEEN SUCCESSIVE VERSIONS OF THE EVS

FROM EVS 1.0 BETA (build date 10/12/07) to EVS 1.0 (build date 05/09/08).

* Feature upgrades and modifications not related to Graphical User's Interface) *

- Allowed real-valued thresholds for all metrics.
- Allowed probability thresholds for all metrics, not just Brier, ROC, and Reliability.
- Included the option for thresholding with a closed interval (i.e. a "between" condition).
- Supported the use of symmetric windows around the forecast median in the Spread-Bias plot.
- Included sharpness (sample-count) plot in the Reliability diagram.
- Changed the definition of probability thresholds in Spread-Bias plot, Mean Capture Rate diagram and Box Plots. Previously, these thresholds referred to plotting positions (i.e. plot resolution) and NOT thresholds of the observed distribution. They now refer to thresholds of the observed distribution for consistency with all other metrics. Plotting positions are now determined with a 'points count' parameter. For example, a point count of 10 for the Spread-Bias plot will construct a plot comprising 10 points.
- Added a new 'points count' parameter for the Spread-Bias plot, Mean Capture Rate diagram, Box Plots and Reliability diagram, which allows the resolution of those diagrams to be altered.
- Included the option to change the default temporal aggregation function from the mean over a specified period to one of several other functions, including the minimum, maximum, and total (i.e. accumulation).
- Included units in the plots that comprise real units (mean error, RMSE, Mean Capture Rate diagram, box plots) once those units have been added to the observed and forecast support for a verification unit.
- Included an option to animate a sequence of verification graphics at different lead times.
- Included writing of sample counts to an XML file when writing other numerical results.
- Included writing of conditional pairs to XML format as well as the original pairs.
- Included an option to ignore global value conditions on a per-metric basis. For example, if a condition was applied to consider only those pairs whose ensemble mean temperatures exceeded freezing, this condition could be ignored on a per-metric basis.
- Enabled backwards compatibility with old project files (i.e. projects with old options will run as before).
- Enhanced and updated documentation.
- Improved memory management for Aggregation units.

* Feature upgrades and modifications related to Graphical User's Interface *

- Removed the table containing reference forecasts, which are not yet enabled.
- Improved the labeling of various options (e.g. 'time zones' rather than 'time systems').

FROM EVS 1.0 (build date 05/09/08) to EVS 2.0 (build date 10/12/09).

* Feature upgrades and modifications not related to Graphical User's Interface *

- Added multiplication factor in support dialog to allow simple changes between measurement units (more complex operations, such as a change in temperature units, are not yet supported).
- Implemented reading of PI-XML observations
- Implemented reading of PI-XML forecasts
- Implemented reading of ASCII observations
- Implemented reading of ASCII forecasts
- Changed representation of forecast lead times from integer hours to double-precision float hours to allow verification of forecasts with lead times in fractional hours, thereby extending the EVS to arbitrary forecast lead times.
- Rewrote the online documentation and updated the mathematical formulas for all of the verification metrics.
- Implemented an R script for each metric in the EVS to read in the XML output and produce high quality plots in EPS format for scientific papers.
- Modified calculation of the mean CRPS to account for the relative position of the observation between the two adjacent ensemble members.
- Added ROC score to the available metrics and included a plot by forecast lead time (same as with other scores). The calculation is based on Mason and Graham (2000).
- Added a sample size metric and associated plot to compute the basic sample size information by forecast lead time and threshold. This may be used for exploratory purposes before computing other verification metrics. In future, we may add further metrics for data exploration (of the observed and forecast data rather than the verification pairs).
- Added modified box plot by size of observed value to GUI (previously via the command line only).
- Modified the spatial aggregation routine to compute the expected (mean) value of each metric across a set of Verification Units rather than pooling paired data.
- Included ability to perform a weighted spatial aggregation. The weight is uniform by default and must sum to 1. A non-uniform weight is also permissible and a weight of "S" is used to weight by the sample size at the first lead time (i.e. maintaining constant weights across lead times). If a

verification metric is not available for a given lead time or threshold the weights are automatically rescaled to sum to 1, maintaining the correct relative weighting of the available metrics.

- Relaxed constraints on spatial aggregation to allow aggregation for Verification Units with different start and end dates.
- Improved the efficiency of file reading for external file formats to ensure that only data within the specified start and end dates and forecast lead times are fully read (otherwise only the file headers are read to check this information).
- Implemented backwards compatibility for the above features so that they do not prevent running of old EVS projects. However, the aggregation routine and CRPS update has been swapped without the option of backwards compatibility. Thus, old projects with spatial aggregation will produce different results in the EVS 2.0. The CRPS update was a bug fix, voiding the need for backwards compatibility.
- Updated the algorithm for CRPS to the method described in Hersbach, H., 2000: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. There are small numerical differences between the old and new algorithms. Also, the procedure described in Hersbach assumes that a constant number of ensemble members is available, whereas the previous method for computing CRPS had no such constraint. Thus, differences will be seen when comparing numbers between systems for which some forecasts comprise null ensemble members.
- Added the decomposition of the CRPS into reliability, resolution and uncertainty, as described in Hersbach (2000).
- Added the Brier Skill Score (BSS) for an arbitrary reference forecast selected by the user.
- Added the Continuous Ranked Probability Skill Score (CRPSS) for an arbitrary reference forecast selected by the user.

* Feature upgrades and modifications related to Graphical User's Interface *

- Removed Time-Series ID and renamed River Segment ID to Location ID.
- Moved basic output options from pop-up window to main Output window.
- Implemented enhanced error dialog with improved error messages.
- Implemented enhanced progress monitor to monitor and display progress of paired-file reading (and included a pair count in the paired file to enable this).
- Updated the aggregation window to include a weighting input in the table of verification metrics. Also renamed some features in this window.
- Added the option to select an arbitrary reference forecast for a skill metric.
- Added the option to show skill score decompositions in a tabbed pane (similar to metrics with one plot per lead time), which may be animated.
- Improved display of zero error line in plots (extended continuously).
- Improved auto-scaling of axes in plots.