

Model Evaluation Tools (MET) verification statistics visualization

P. Oldenburg, J. Halley Gotway, T. Jensen - National Center for Atmospheric Research (NCAR) / Research Applications Laboratory (RAL) & Developmental Testbed Center (DTC)



Abstract

The Model Evaluation Tools (MET) was developed by the Developmental Testbed Center (DTC) to provide the numerical prediction community with common software incorporating the latest advances in forecast verification. Recently, the DTC developed a system for storing and visualizing MET verification statistics by integrating database and plotting software. The primary focus of this development was to provide a configurable tool, called METViewer, for use across several different DTC testing and evaluation projects. Evaluation activities that the DTC performed over the past year, for which METViewer was utilized, include the Quasi-Normal Scale Elimination (QNSE) Sensitivity Test, the Hydrometeorology Testbed Collaboration (HMT) and the Hazardous Weather Testbed Collaboration (HWT). It is anticipated that METViewer will be available to the community in the future.

METViewer ingests MET output into a standard SQL database, which facilitates searching and sorting of the verification statistics. Using a user-supplied plot XML specification as input, METViewer gathers and aggregates statistics from the database and generates one or more plots. The three primary types of statistics handled are continuous, categorical and object based. Additional calculations, such as confidence intervals, bootstrap resampling and observation frequencies, are performed during the aggregation and plotting process. The primary focus, to date, has been to produce monthly and seasonal aggregations of statistics plotted over lead time, valid date and precipitation threshold, along with vertical profile plots, for a number of atmospheric variables, including: temperature, moisture and winds.

System Design

METViewer consists of three interworking components, all implemented in Java. These components interact with a SQL database and the R statistical package to manage data and generate plots, respectively. All three components interact using XML.

Database Loader:

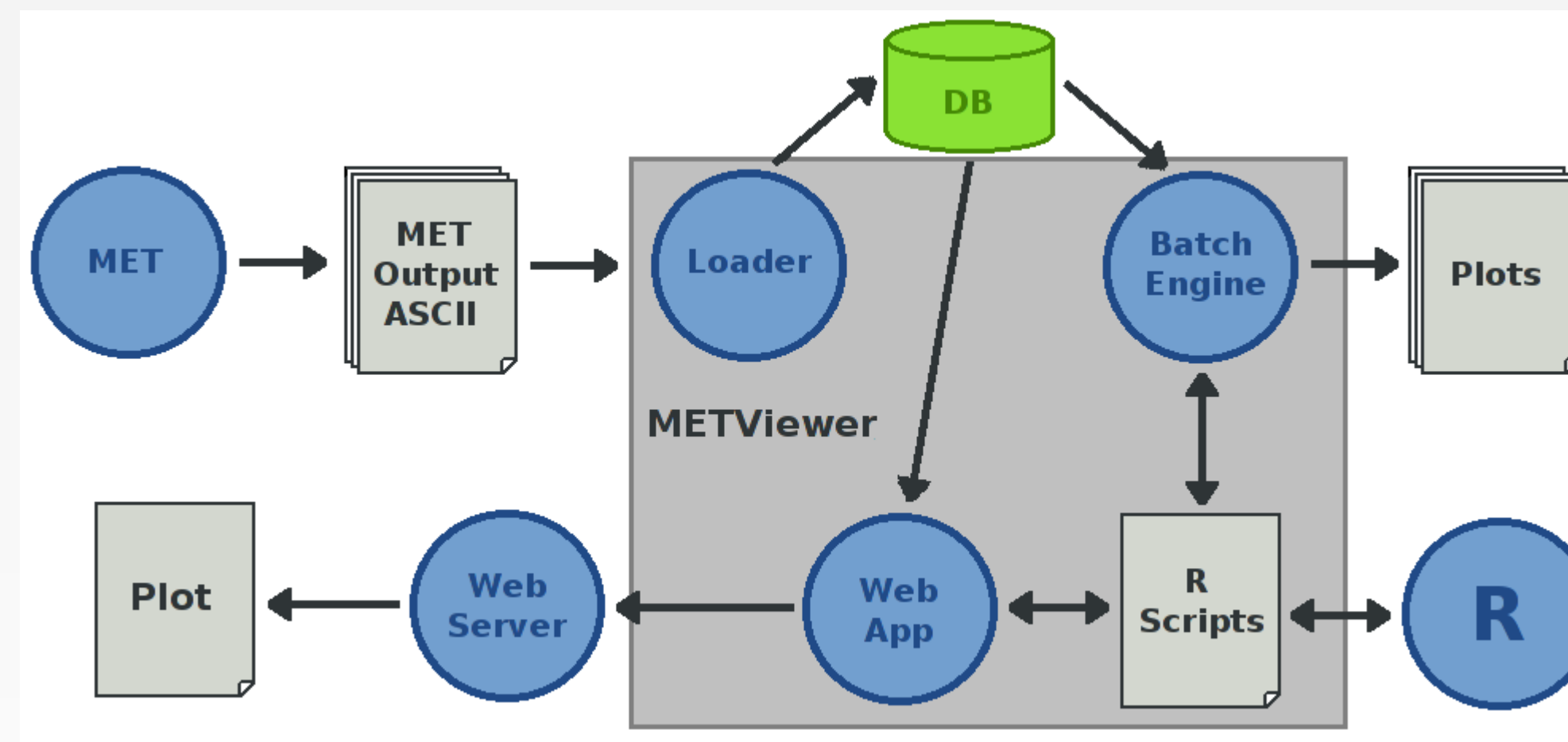
Reads an XML job specification and loads one or more MET output ASCII files into a SQL database.

Web Application:

Uses an Apache Tomcat JSP web interface to collect plot information and build an XML plot job specification. Submits the XML to the Batch Engine to generate a plot and present the plot.

Batch Engine:

Reads an XML plot job specification from the command line or Web Application, pulls verification statistics from the database and creates one or more plots using R.



A typical METViewer usage pattern is to design a plot using the web application, and then expand the resulting XML plot specification to generate a large number of plots. Verification experiments typically involve a number of cases, models and statistics which are compared to one another.

Plot Specification

In METViewer, the user specifies several criteria that define the characteristics of the plot. These criteria are passed to METViewer in an XML document with a defined format.

Plot Type:

- Series Plot: displays one or more series of statistics with configurable confidence intervals
- Box Plot: displays boxes for each series member, using the Tukey paradigm
- Bar Plot: series values appear as bars
- Rank Histogram: for ensembles, displays a histogram showing the frequencies of the observation rank among the ensemble members

Independent Variable:

Lead Time, Valid Time, Initialization Time, Forecast Threshold, Pressure Level, etc.

Dependent Variable(s):

Combination of forecast variables (e.g. precip, temp) and verification statistics. Several types of verification statistics are supported.

- Continuous: RMSE, ME, etc.
- Categorical: PODY, GSS, etc.
- Object Based: Centroid Distance, Area Ratio, etc.

Case Information:

Fixed sets of values that apply to the plot data (e.g. date ranges, observation types)

Series Information:

Specifies what the various plot series represent (e.g. models, thresholds, neighborhood sizes) and information about confidence intervals

Aggregation Information:

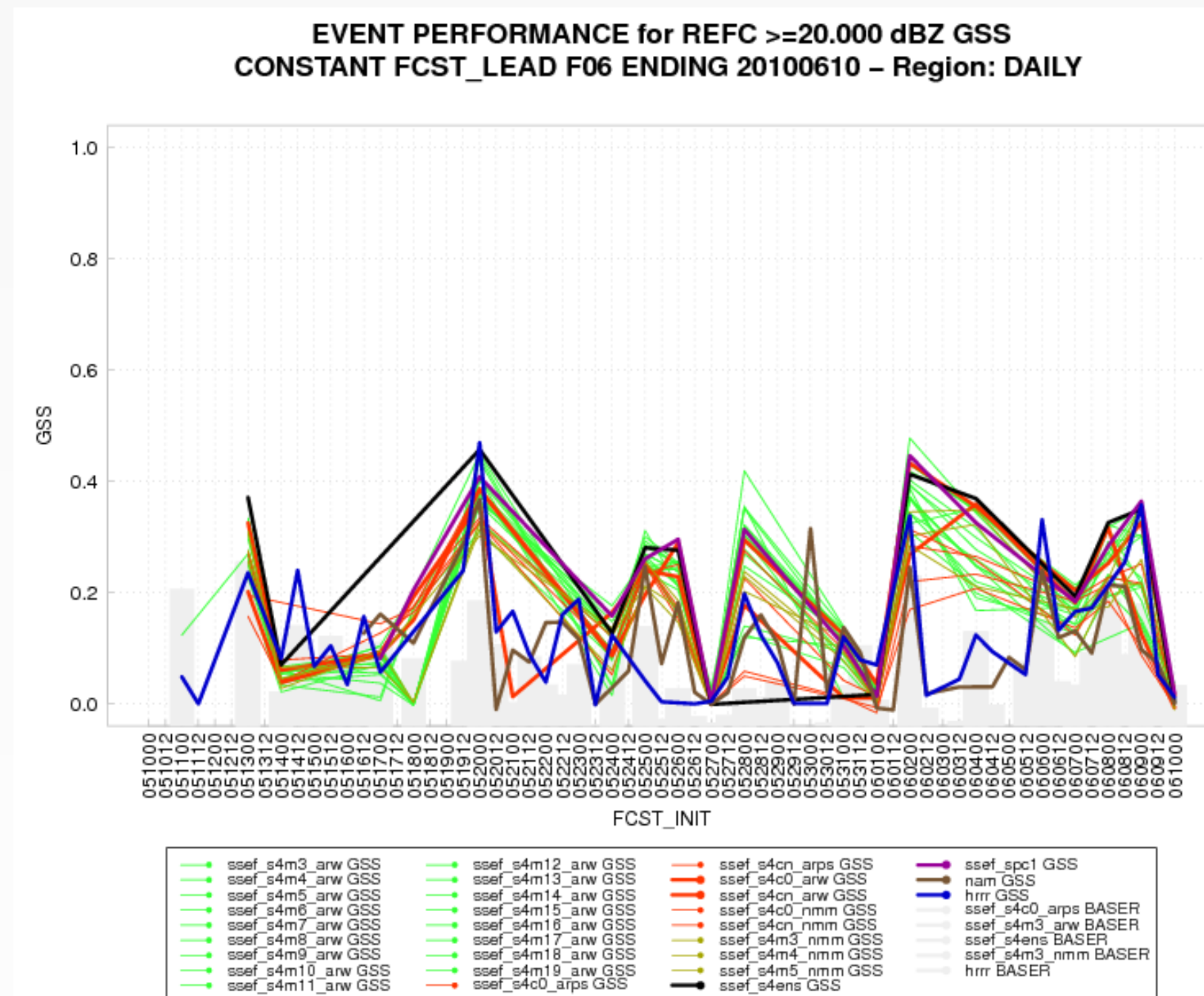
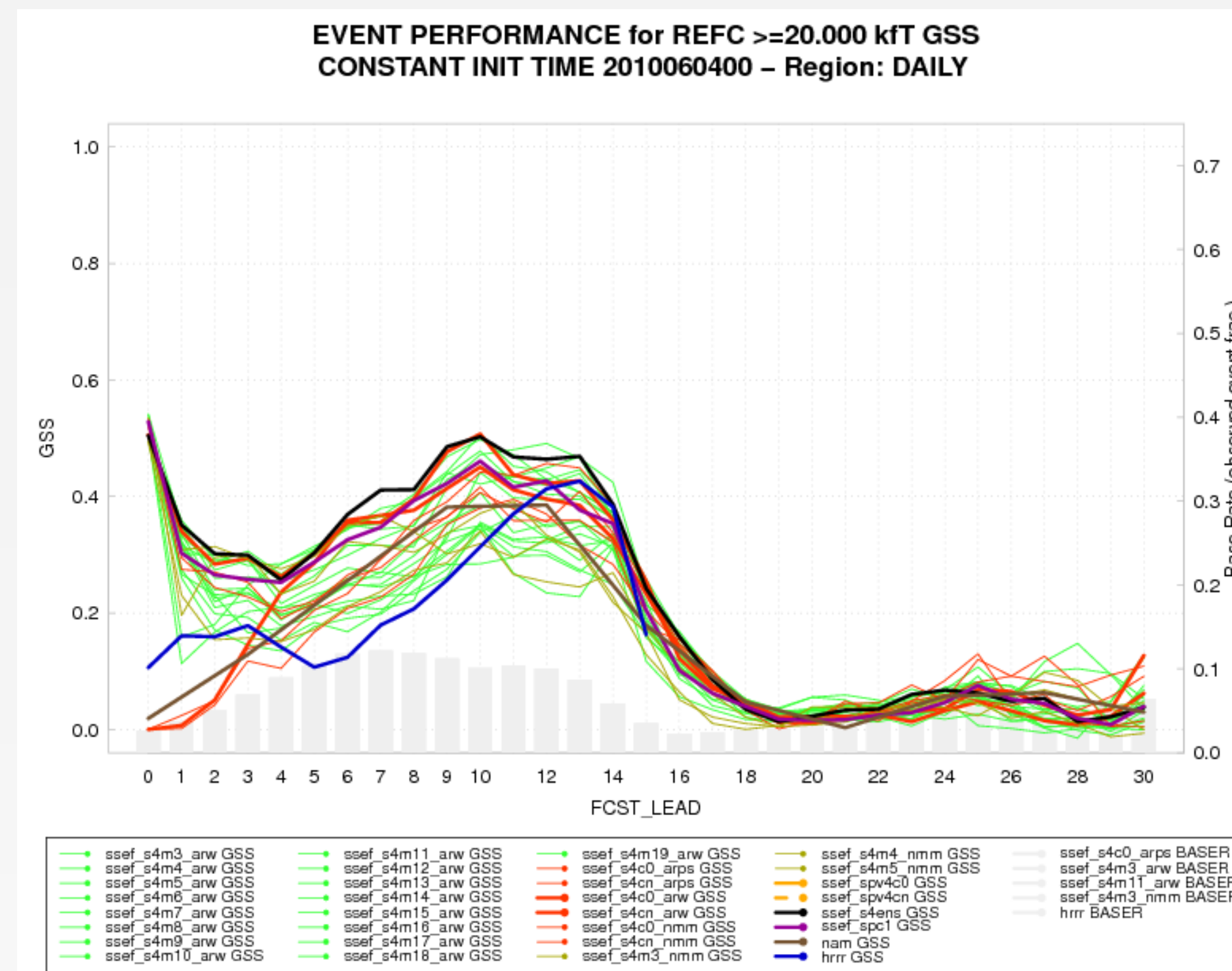
Controls the way that contingency tables or partial sums are combined to calculate stats

Plot Formatting:

- Geometry: Sizes, locations, orientations
- Colors: Series lines or boxes, legend, fonts and lines
- Plot Bounds: Automatically calculated or custom, linear or log scale

Time Series Plots

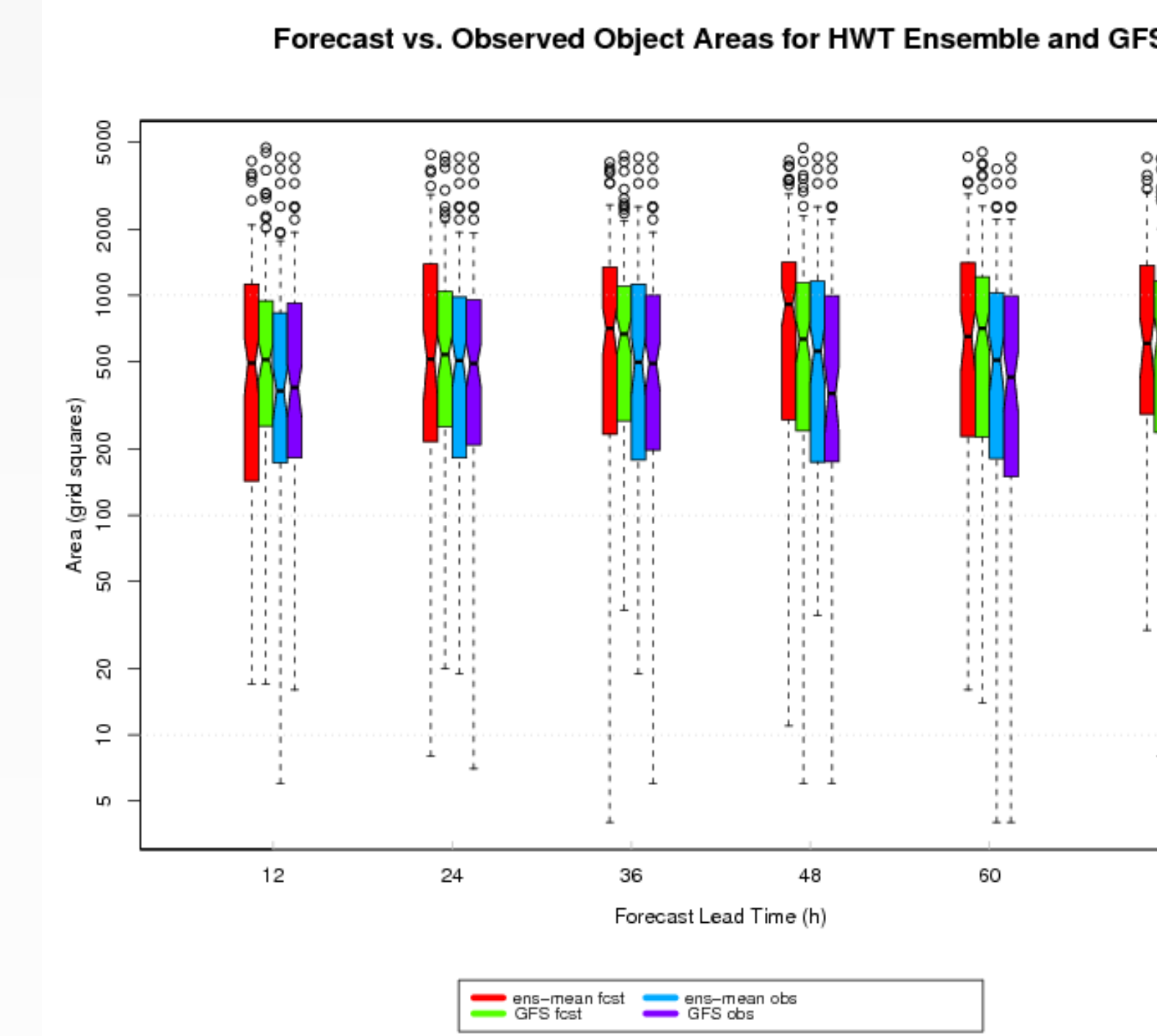
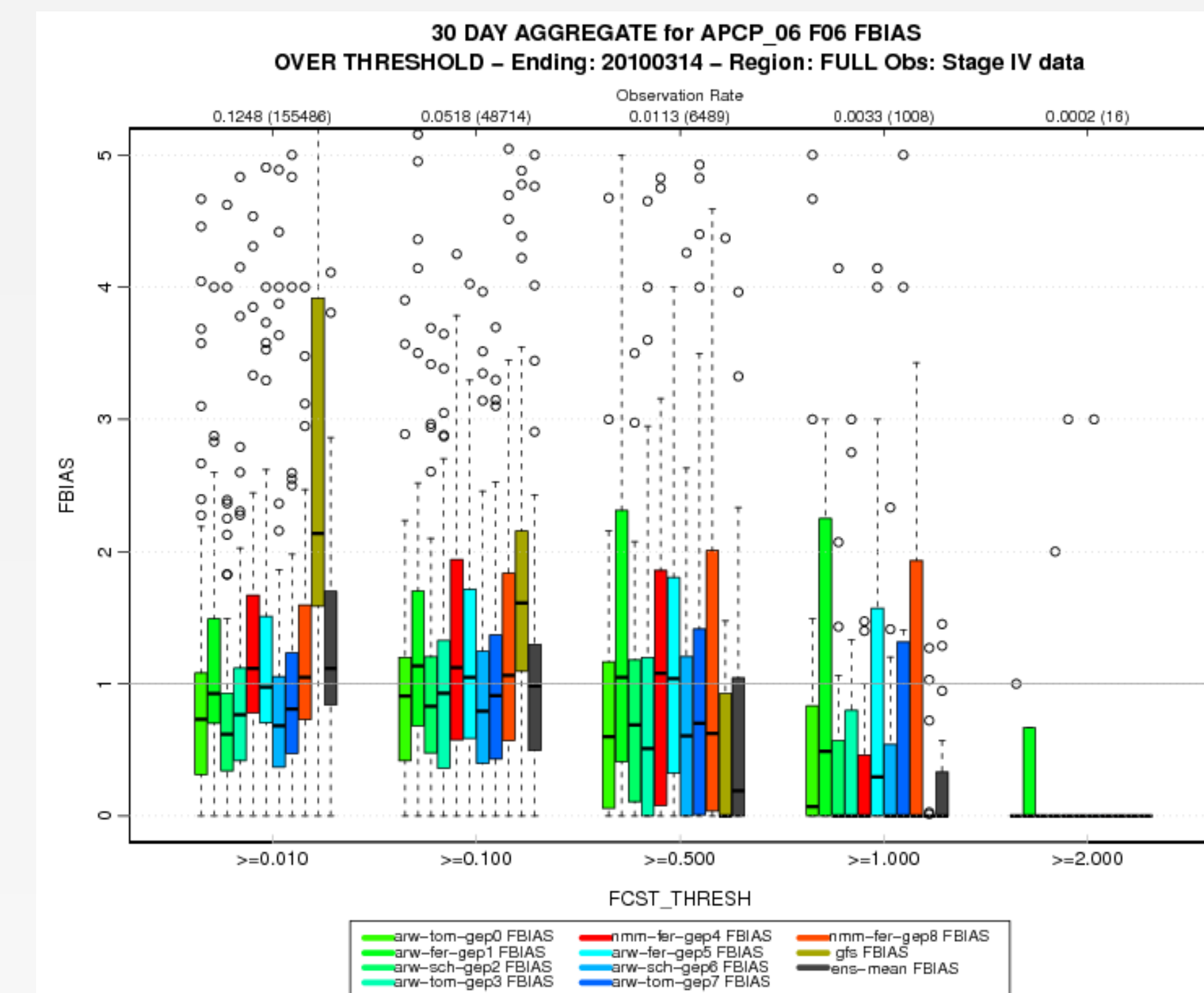
Time series plots use a measure of time as an independent variable, whether lead time (number of hours) or valid or initialization times. In the former case, all data on the plot is either from a fixed initialization time or a fixed valid time. For a lead series plot, forecast performance typically declines as the number of hours increases, reflecting increasing uncertainty.



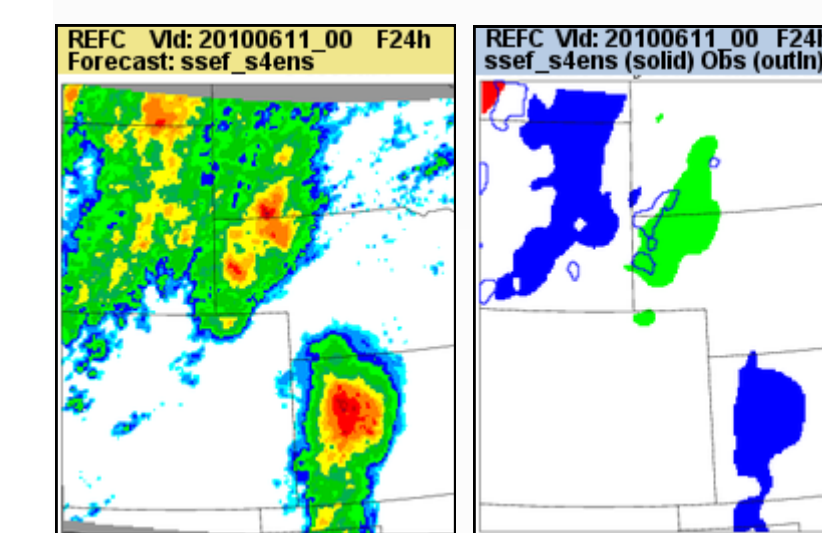
For initialization or valid time series plots, the x-axis values are dates. In this case the date format is MMDDHH, showing the initialization time of each forecast. To a certain degree, forecast performance can depend on observed activity, which is shown using gray bars on the right hand y-axis scale.

Box Plots

This box plot shows a representation of the distributions of frequency bias (FBIAS) values for each model, showing median, quantiles and outliers. Forecast thresholds can be used as the independent variable for categorical verification statistics in METViewer. This plot shows that the number of events decreases to nearly zero as threshold increases and the FBIAS score degrades from a perfect score of 1.

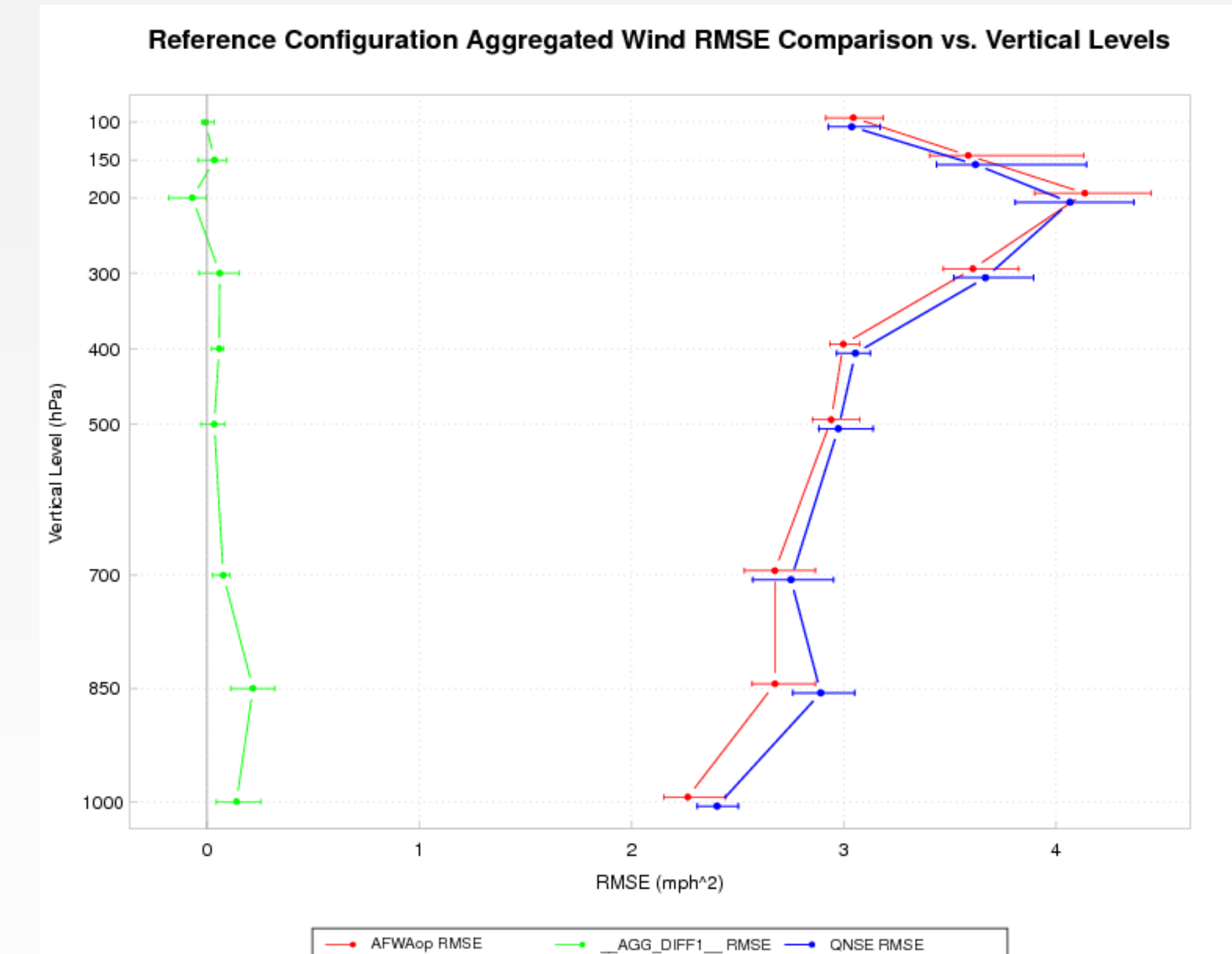


MET calculates object based statistics which can be plotted in several different ways. A few of the objects whose areas are measured can be seen in the "postage stamp" plots below.

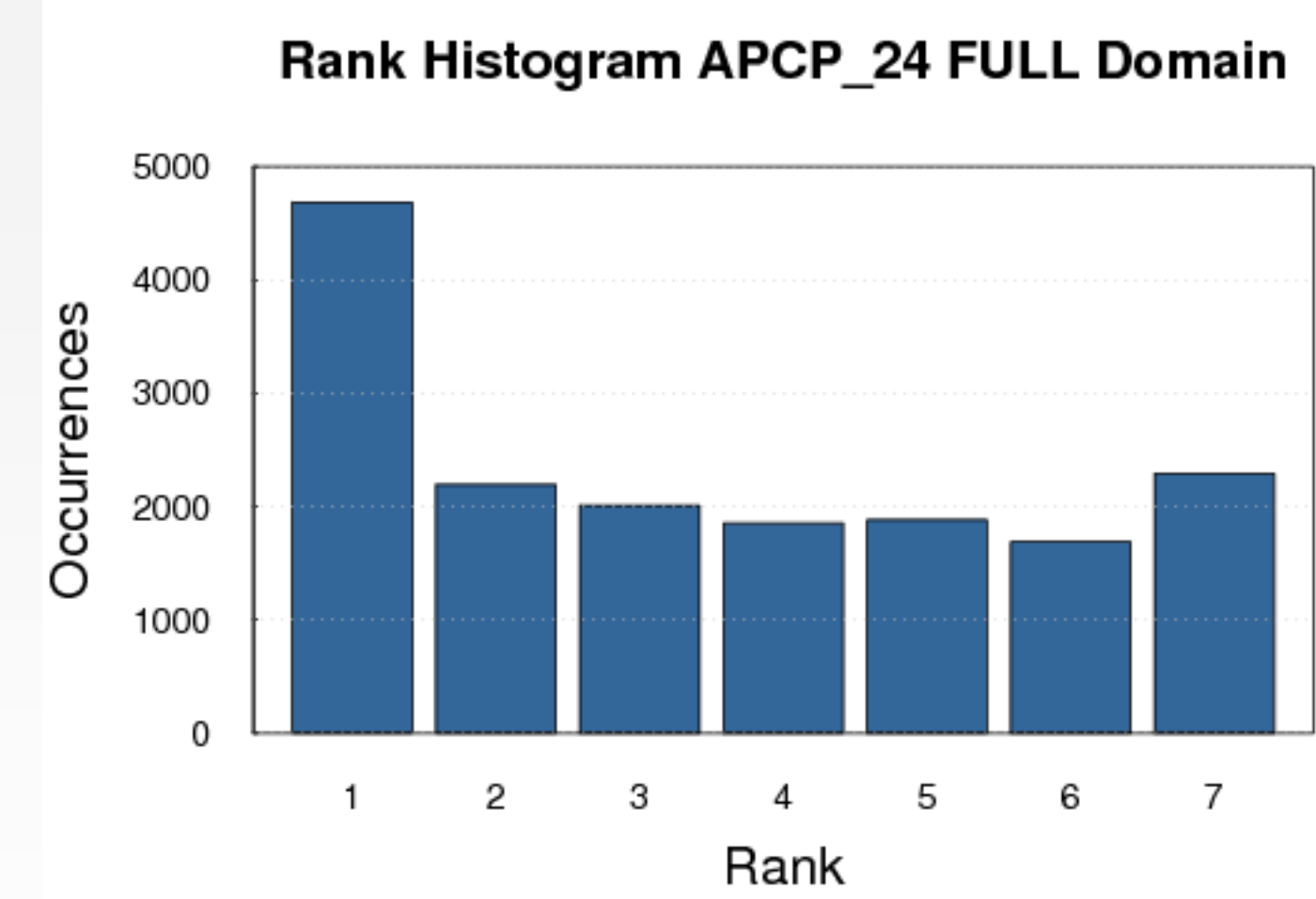


Vertical Levels Plots

Some observations, such as rawinsonde, provide vertical atmospheric measurements. When these observations are used to verify forecast data, statistics can be calculated at several vertical levels. This plot shows a comparison between two models and the pairwise difference between them. Bootstrapping is used to calculate confidence intervals on the difference to find statistically significant differences in performance.



Rank Histograms



Observation rank is used to verify ensembles and analyze the spread of the ensemble members. A rank histogram plot shows an aggregation of observation ranks over a large number of cases. Different rank histogram shapes indicate different types of performance including bias, under-dispersiveness and over-dispersiveness. In this plot, the ensemble is biased high, perhaps because the ensemble members frequently forecast a small amount of precipitation when there is none observed.

References
 Davis, C.A., B.G. Brown, and R.G. Bullock, 2006a: Object-based verification of precipitation forecasts, Part I: Methodology and application to mesoscale rain areas. Monthly Weather Review, 134, 1772-1784.
 Hamill, T.M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. Mon. Wea. Rev., 129, 5507-5560.
 Wilks, D., 2006: Statistical methods in the atmospheric sciences. Elsevier, San Diego.