

MET with Ensembles

John Halley Gotway
Tressa Fowler
NCAR/RAL/JNTP
and
Developmental Testbed Center (DTC)
Project Manager: Tara Jensen
MET Team: Randy Bullock, Tatiana Burek, Julie Prestopnik,
Minna Win, Howard Soh, Jim Frimel, Hank Fisher
Tina Kalb, Dan Adriaansen, Barb Brown

Outline of Talk

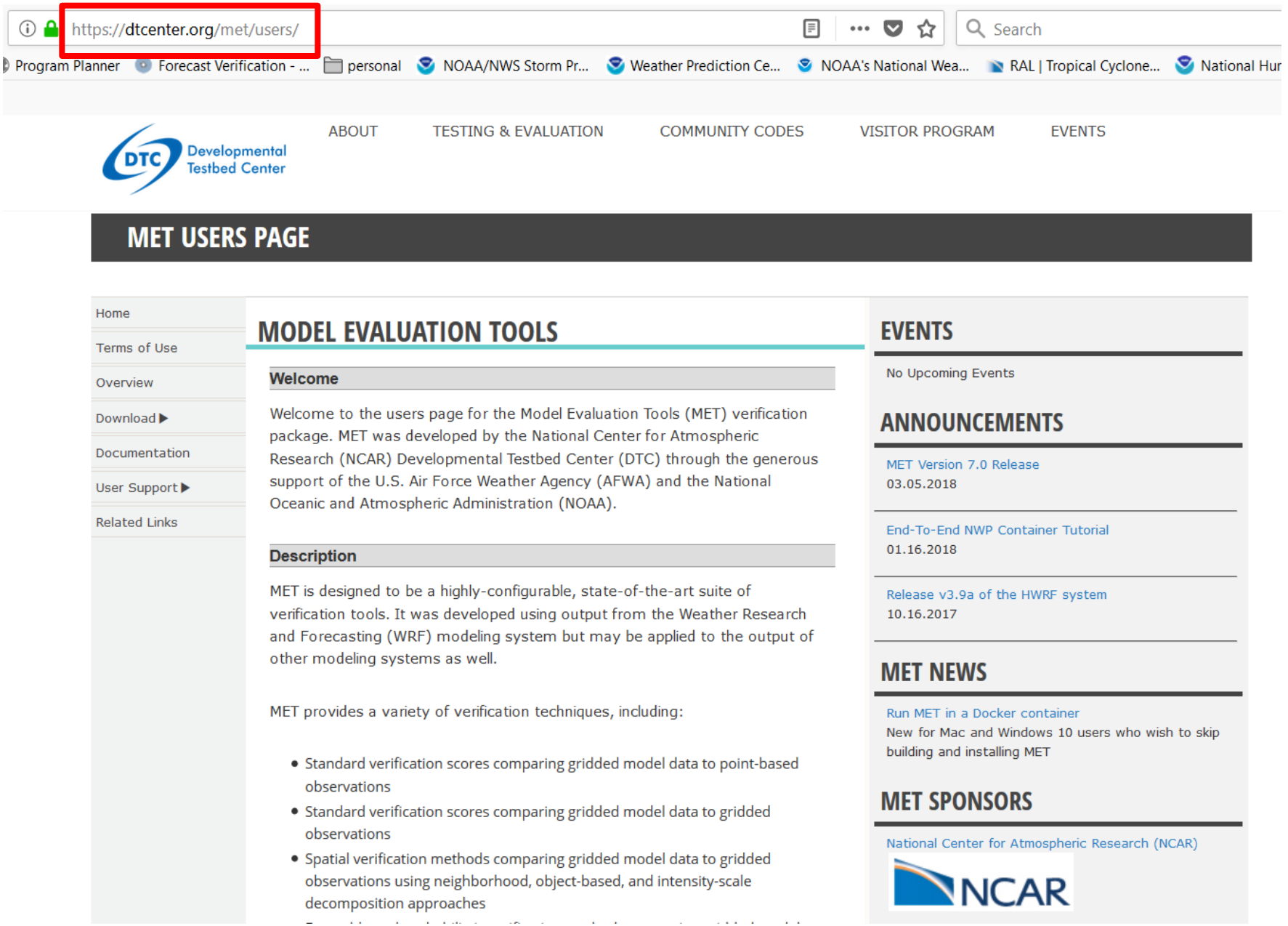
- **Overview of MET software**
- **METViewer Database and Display**
- **MET+ Automation**
- **Ensemble Verification by Tressa Fowler**
- **Ensemble-Stat Tool**
- **met-7.0 (current) and met-7.1 (summer)**
- **Docker containers**

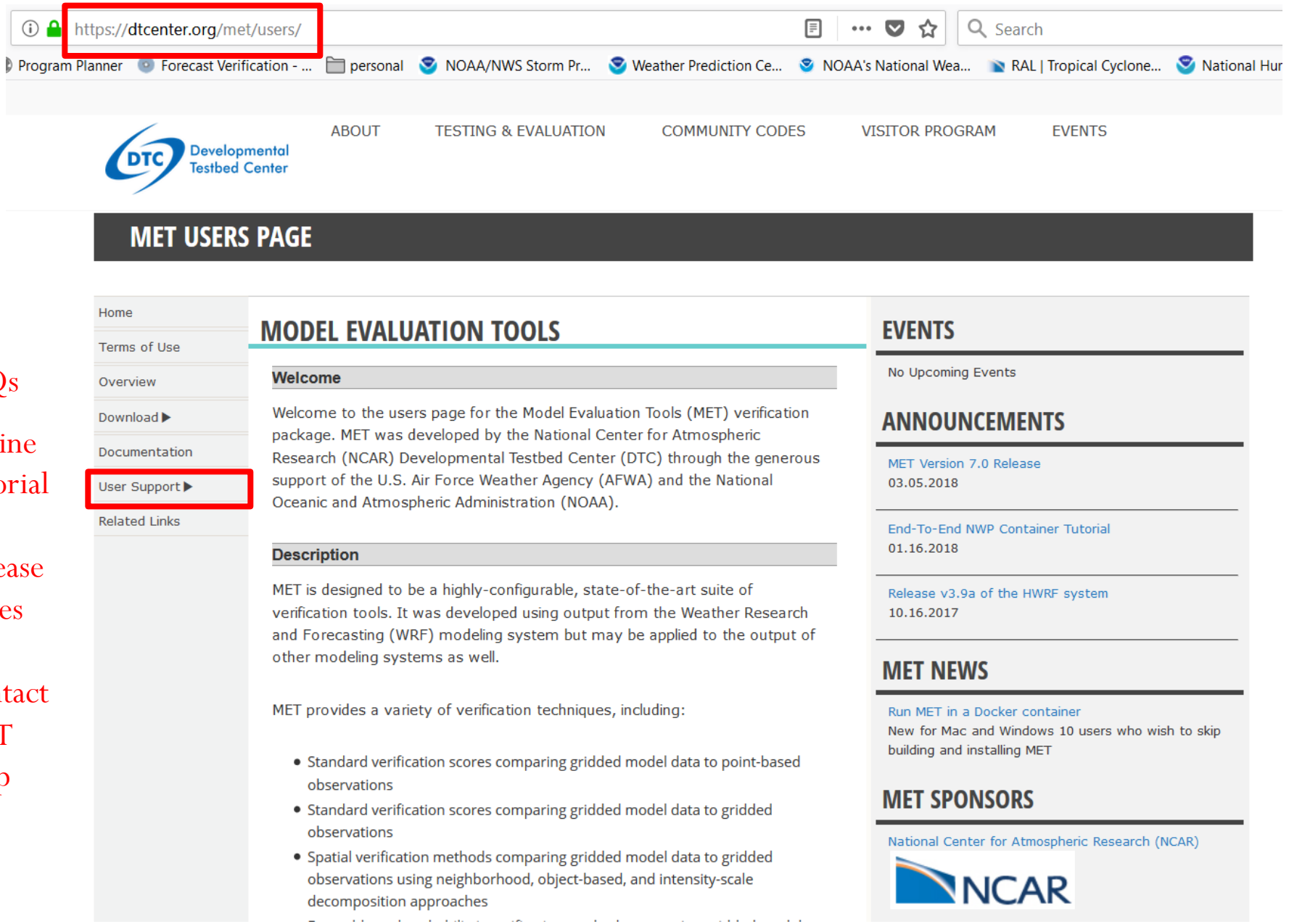
MET Package

First MET Release:
January 2008

- **MET is community code supported by DTC**
  - Feely available and open source
  - 3600+ registered users
  - 133+ countries, 30% from USA
  - Universities, Government, Private Companies, Non-Profits
- Download MET release and compile locally.
  - Register and download: **www.dtcenter.org/met/users**
  - C++ with calls to some Fortran libraries
  - Linux with GNU, PGI, or Intel compilers
- Support
  - Online tutorial and in-person tutorials given yearly
  - **met_help@ucar.edu** help desk
  - 250+ support tickets in past year

A Decade of Community Verification


https://dtcenter.org/met/users/
Search
Program Planner
Forecast Verification - ...
personal
NOAA/NWS Storm Pr...
Weather Prediction Ce...
NOAA's National Wea...
RAL | Tropical Cyclone...
National Hur
DTC Developmental Testbed Center
ABOUT
TESTING & EVALUATION
COMMUNITY CODES
VISITOR PROGRAM
EVENTS
MET USERS PAGE
Home
Terms of Use
Overview
Download ▶
Documentation
User Support ▶
Related Links
MODEL EVALUATION TOOLS
Welcome
Welcome to the users page for the Model Evaluation Tools (MET) verification package. MET was developed by the National Center for Atmospheric Research (NCAR) Developmental Testbed Center (DTC) through the generous support of the U.S. Air Force Weather Agency (AFWA) and the National Oceanic and Atmospheric Administration (NOAA).
Description
MET is designed to be a highly-configurable, state-of-the-art suite of verification tools. It was developed using output from the Weather Research and Forecasting (WRF) modeling system but may be applied to the output of other modeling systems as well.
MET provides a variety of verification techniques, including:
• Standard verification scores comparing gridded model data to point-based observations
• Standard verification scores comparing gridded model data to gridded observations
• Spatial verification methods comparing gridded model data to gridded observations using neighborhood, object-based, and intensity-scale decomposition approaches
EVENTS
No Upcoming Events
ANNOUNCEMENTS
MET Version 7.0 Release
03.05.2018
End-To-End NWP Container Tutorial
01.16.2018
Release v3.9a of the HWRF system
10.16.2017
MET NEWS
Run MET in a Docker container
New for Mac and Windows 10 users who wish to skip building and installing MET
MET SPONSORS
National Center for Atmospheric Research (NCAR)
NCAR

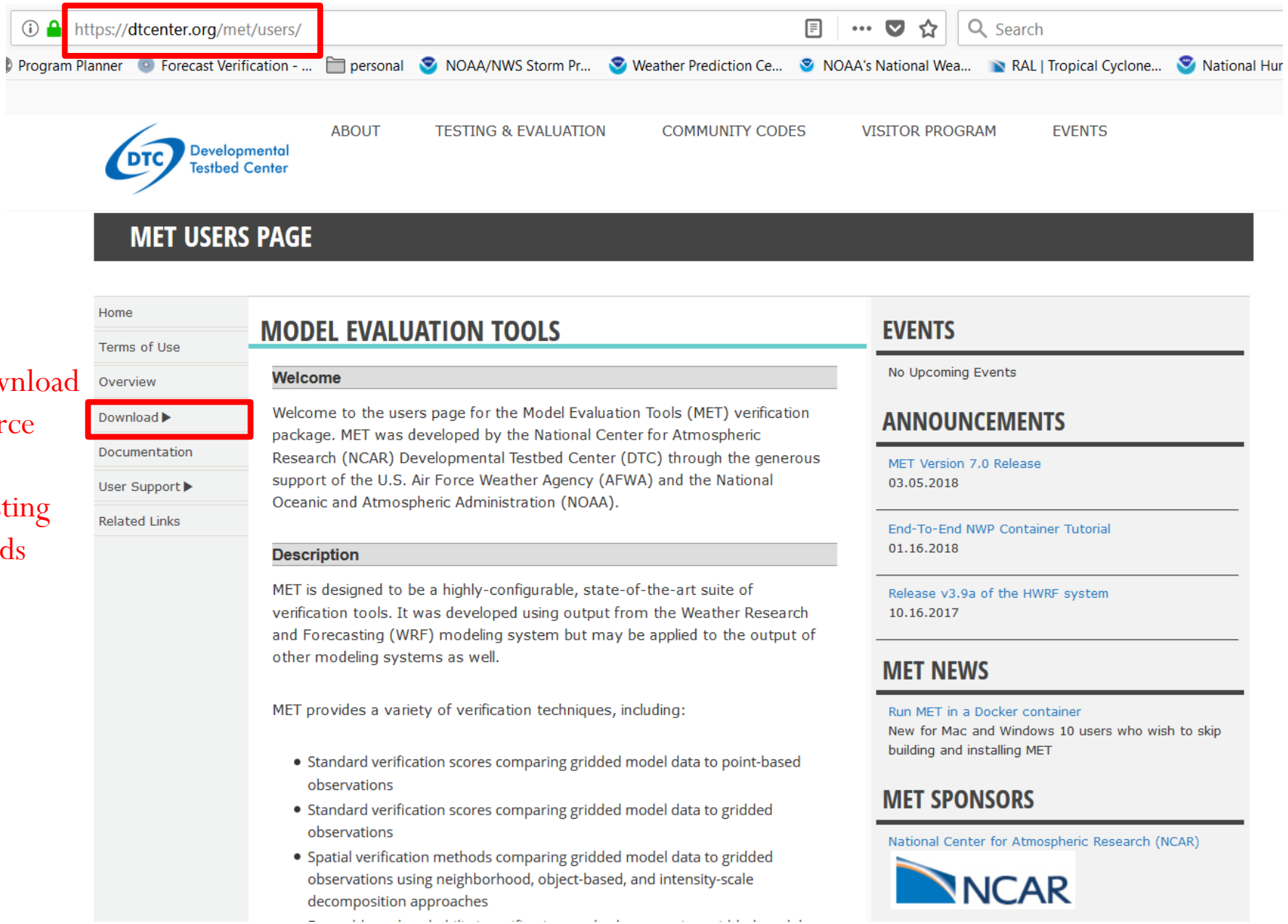


FAQs

Online Tutorial

Release Notes

Contact MET Help



Download Source

Existing Builds

METV7.0 EXISTING MET BUILDS

METv7.0 Existing MET Builds

- NCAR RAL machines
  - MET BUILD: /usr/local/met
    PATCH DATE:
- NCAR machine cheyenne
  - PATCH DATE:
    MODULES:
    - module use /glade/p/ral/jnt/MET/MET_releases/cheyenne/modulef
    - module load met/7.0
- NOAA machine theia
  - PATCH DATE:
    MODULES:
    - module use /contrib/modulefiles
    - module load met/7.0

NCAR-RAL Machines

Cheyenne

Theia

Gaea

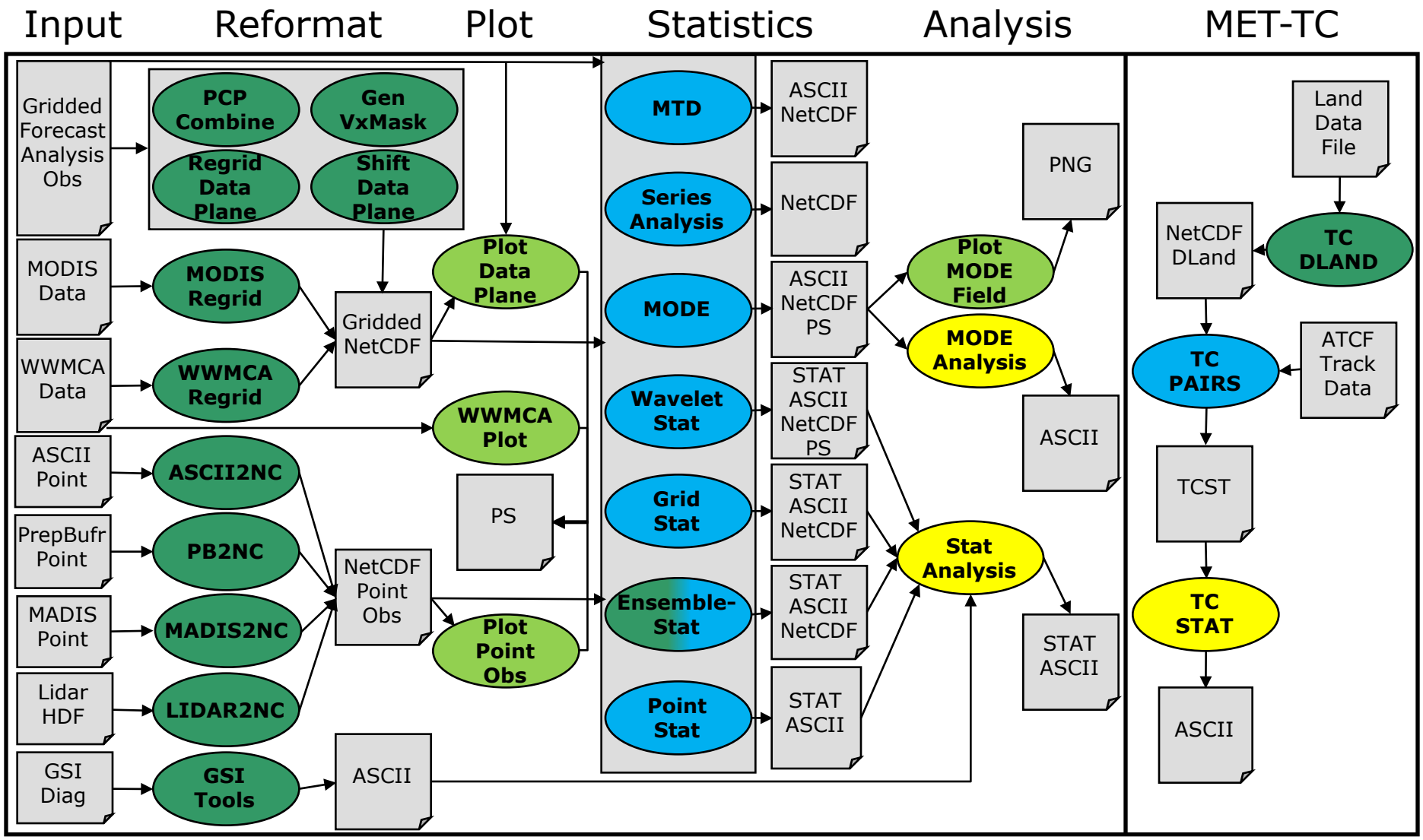
Jet

WCOSS Machines

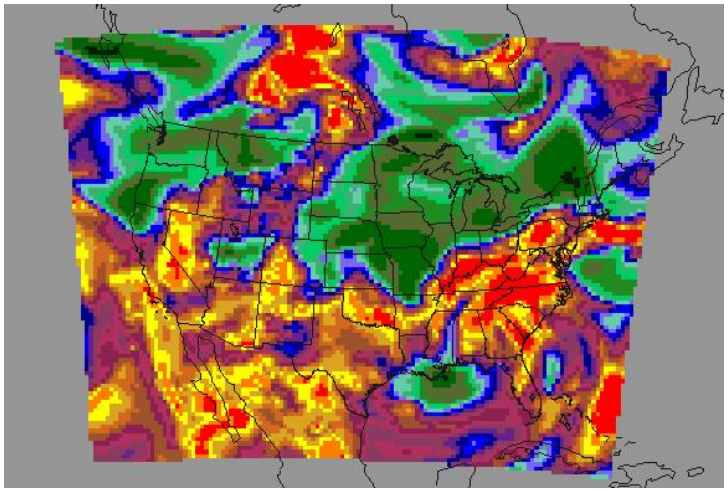
Docker Container:

https://github.com/NCAR/container-dtc-met

MET Overview v7.0

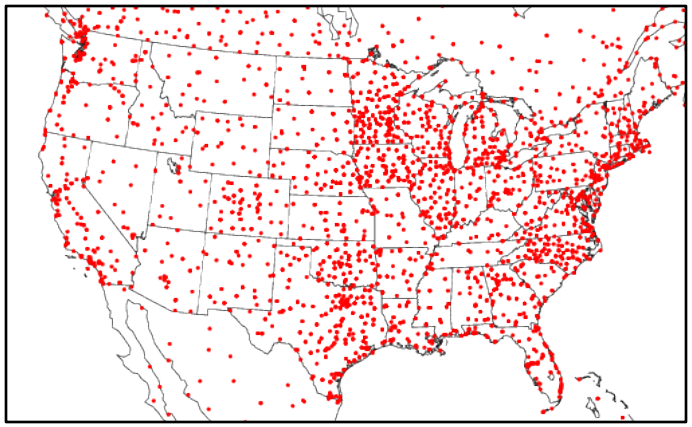


Point-Stat: Overview

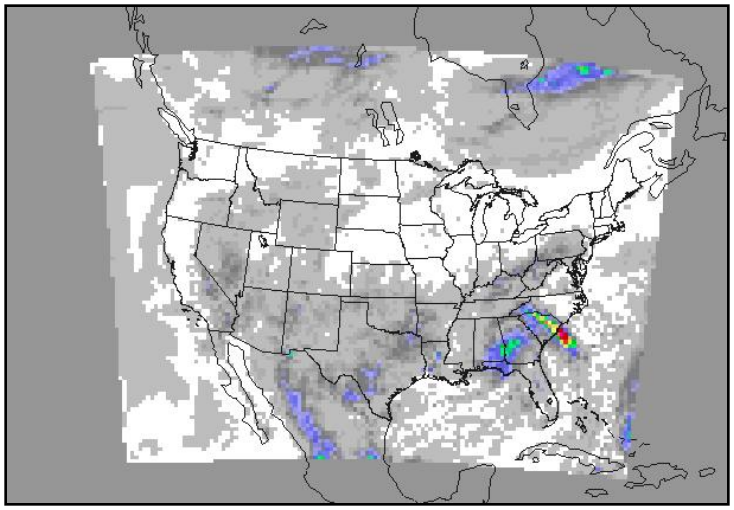


- Compare **gridded forecasts** to **point observations.**
- Accumulate matched pairs over a defined area at a single point in time.
- Verify one or more variables/levels.
- Analysis tool provided to aggregate through time.

- Verification methods:
  - Continuous statistics for raw fields.
  - Single and Multi-Category counts and statistics for thresholded fields.
  - Parametric and non-parametric confidence intervals for statistics.
  - Compute partial sums for raw fields and/or the raw matched pair values.
  - Methods for probabilistic forecasts.
  - HiRA spatial verification method.

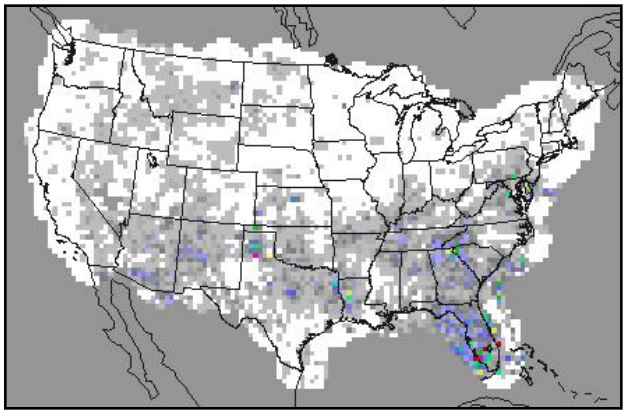


Grid-Stat: Overview



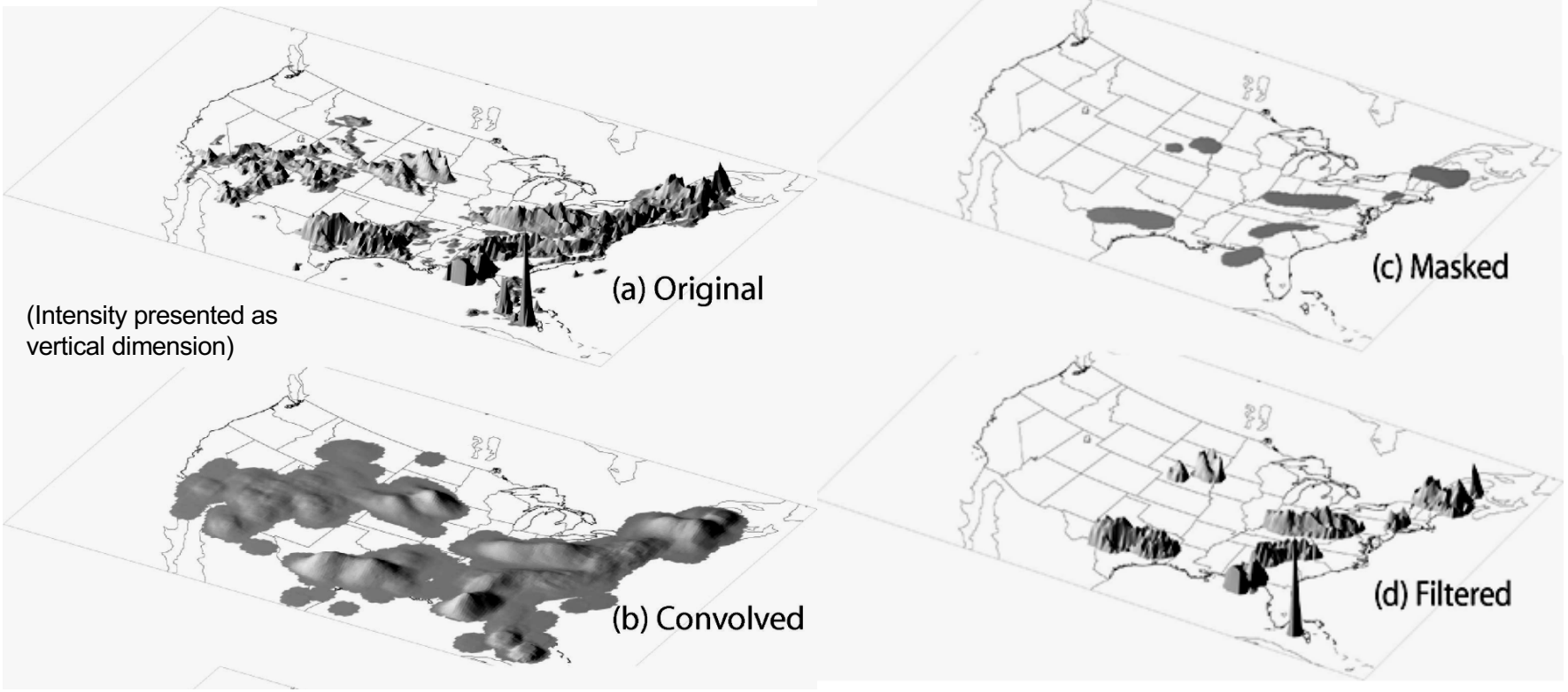
- Compare **gridded forecasts** to **gridded observations** on the same grid.
- Accumulate matched pairs over a defined area at a single point in time.
- Verify one or more variables/levels.
- Analysis tool provided to aggregate through time.

- Verification methods:
  - Continuous statistics for raw fields.
  - Single and Multi-Category counts and statistics for thresholded fields.
  - Parametric and non-parametric confidence intervals for statistics.
  - Compute partial sums for raw fields.
  - Methods for probabilistic forecasts.
  - Continuous statistics and categorical counts/statistics using neighborhood verification method and gradients.



MODE: Object Identification

Source: Davis 2006

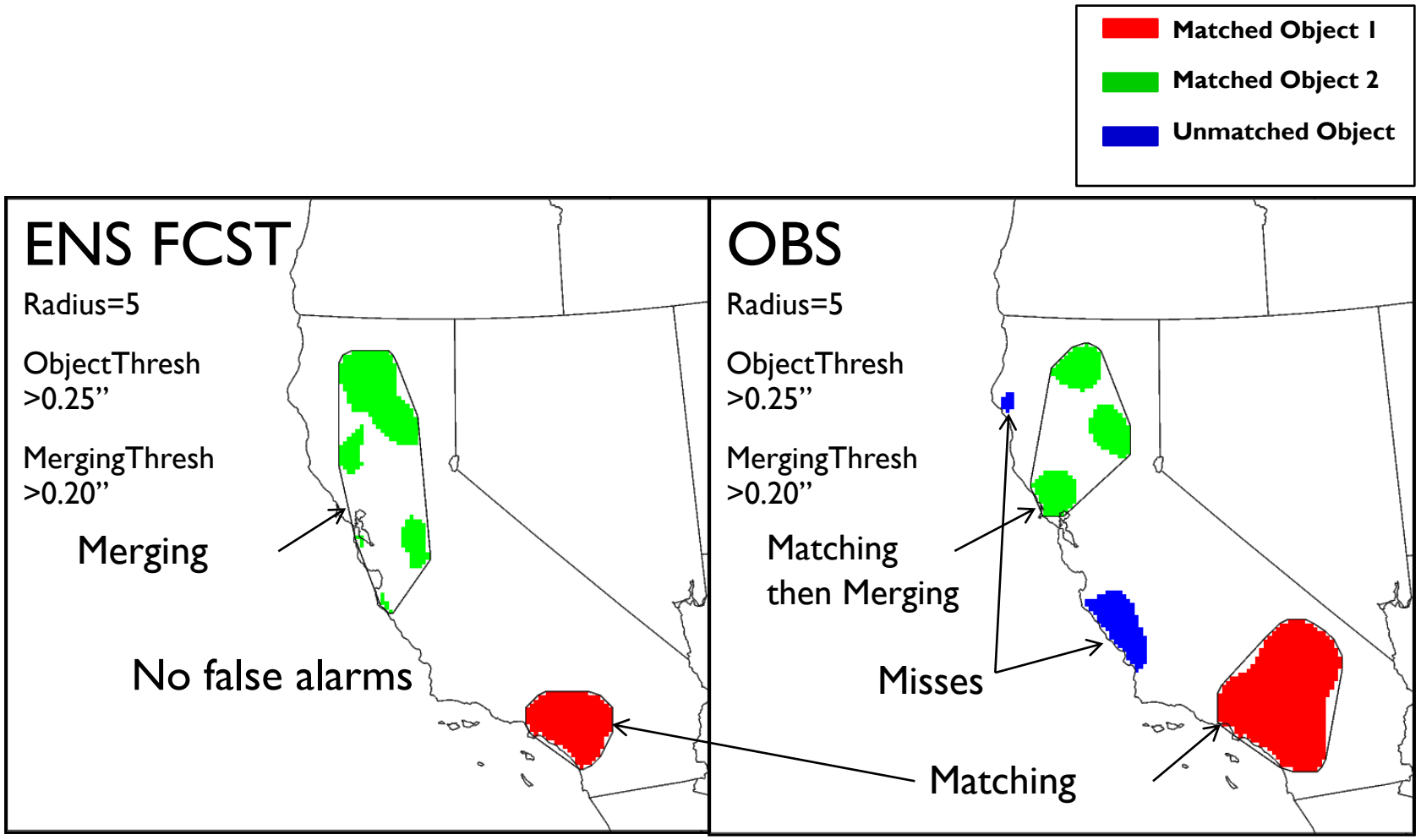


Smoothing radius
(in grid squares)

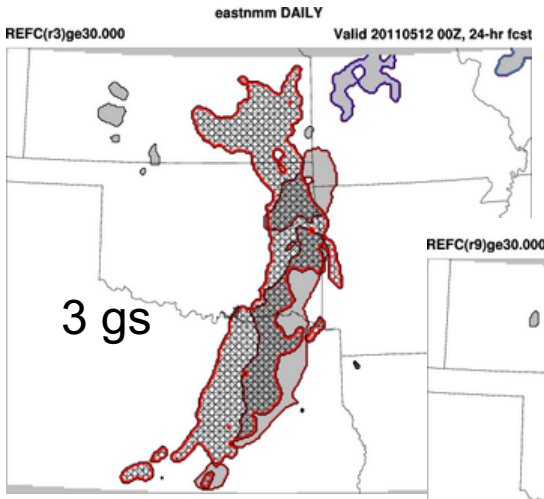
Accumulation threshold
(in mm)

*** User-defined parameters in configuration file**

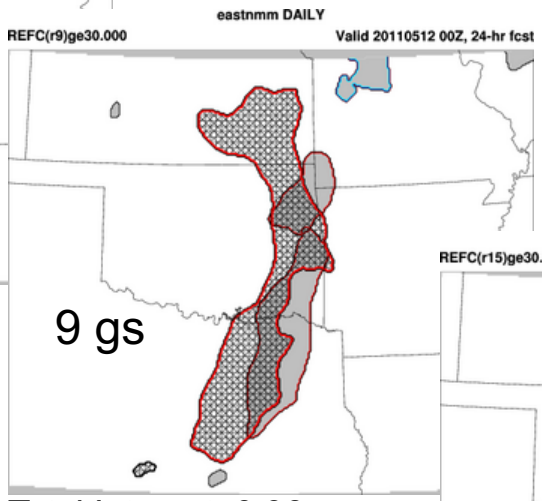
MODE: Precip Example



MODE: REFC > 30 Example
Impact of Smoothing Radius

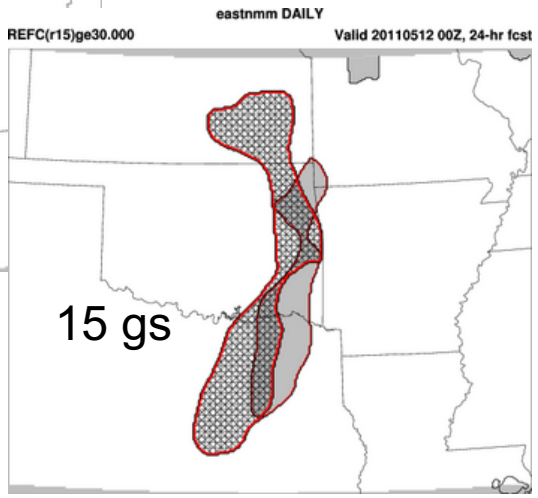


Total Interest: 0.96
Area Ratio: 0.53
Centroid Distance: 95km
P90 Intensity Ratio: 1.00



Total Interest: 0.96
Area Ratio: 0.57
Centroid Distance: 94km
P90 Intensity Ratio: 1.02

Convolution Radius Increases

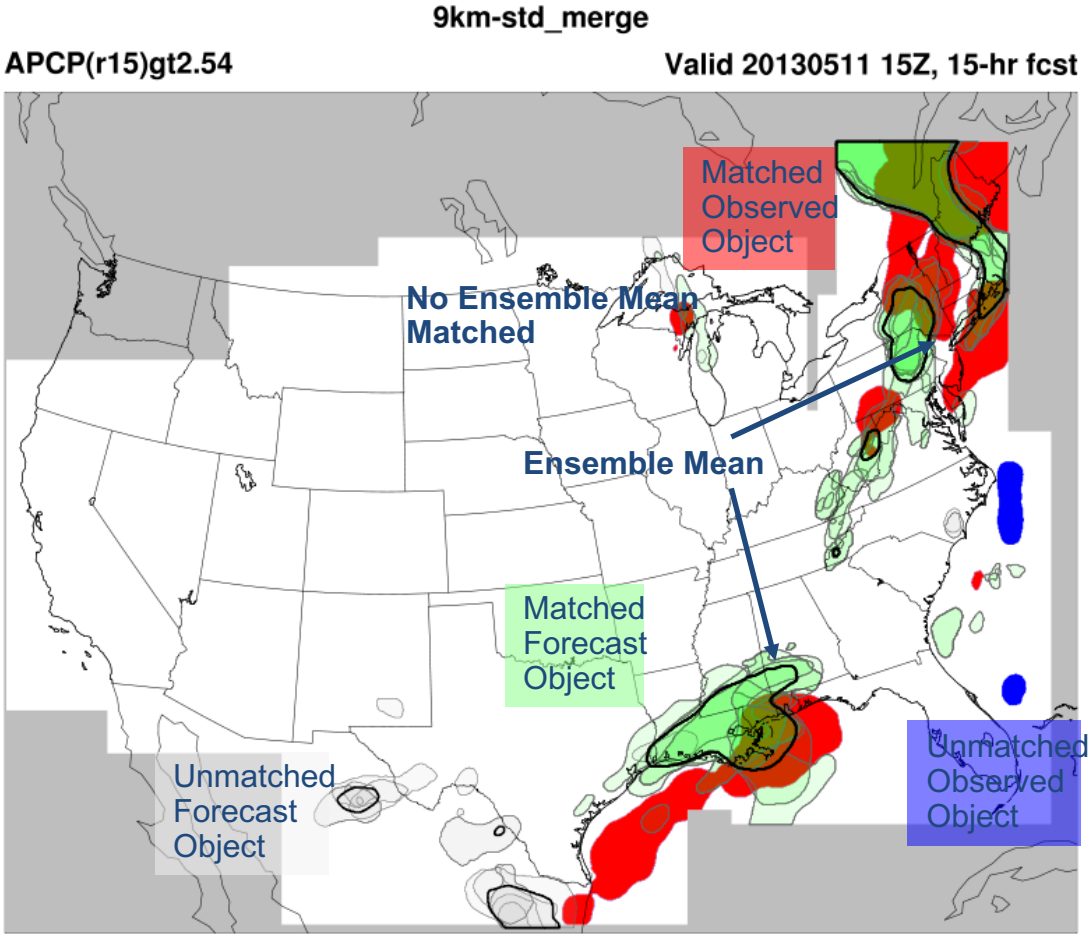


Total Interest: 0.96
Area Ratio: 0.53
Centroid Distance: 92km
P90 Intensity Ratio: 1.04

MODE: Ensemble Precip Example

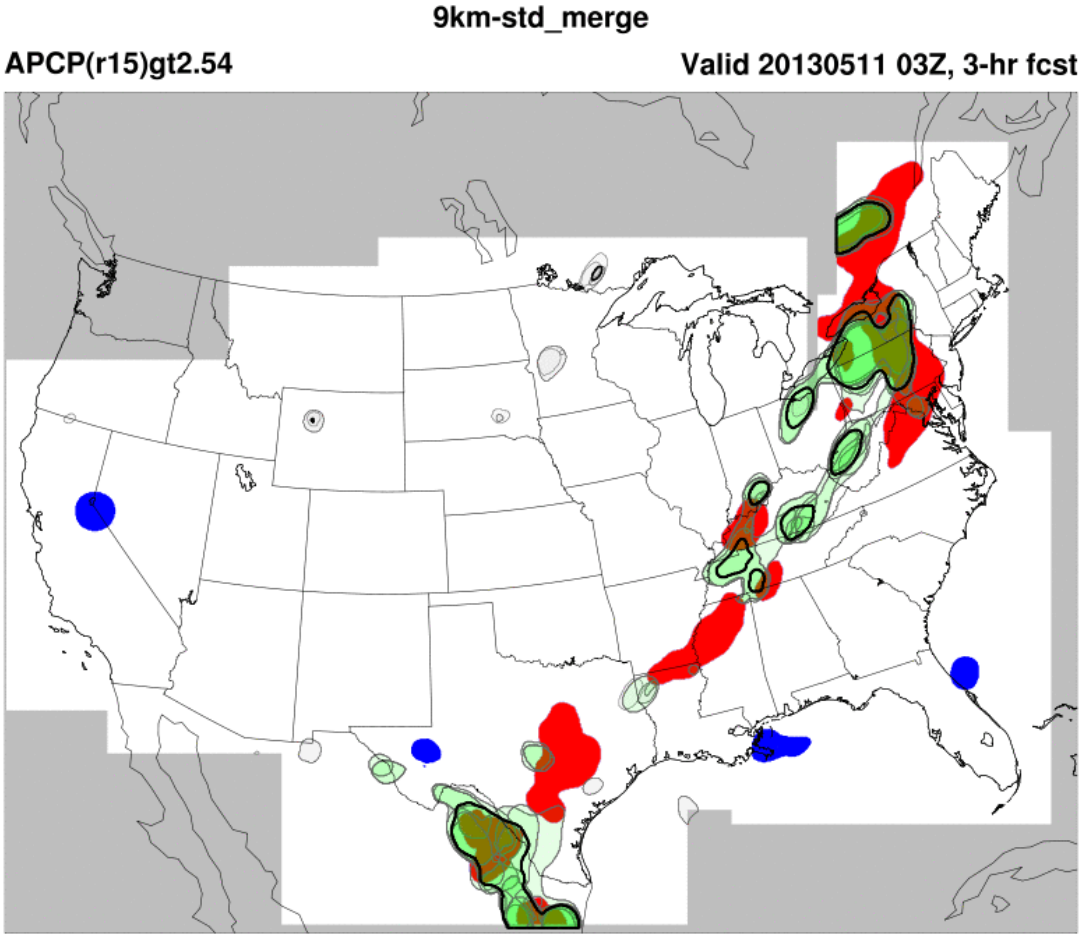
Example May 11, 2013

DTC SREF Tests – ARW Members



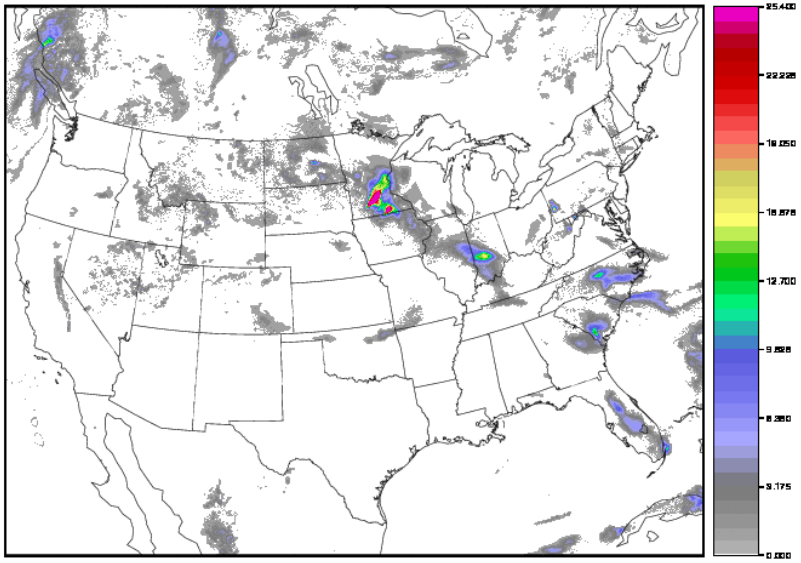
MODE: Ensemble Precip Example

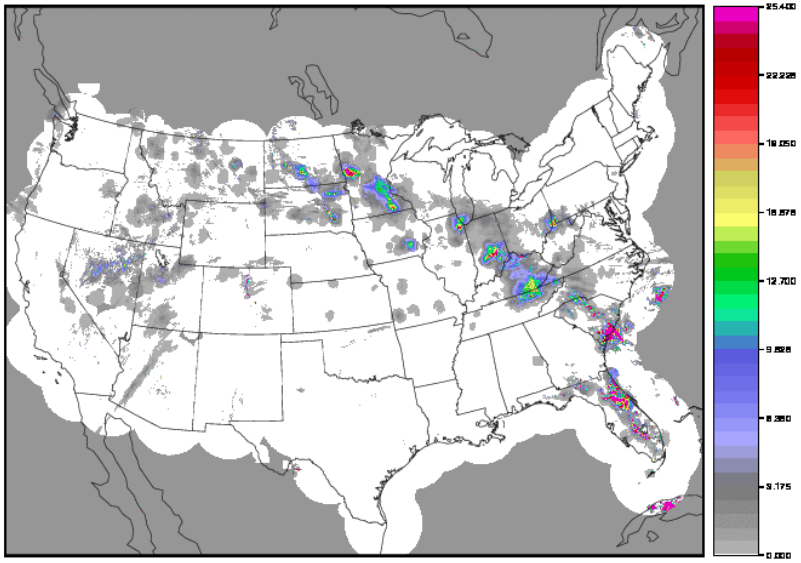
Spread increases With Time



Series-Analysis: Example

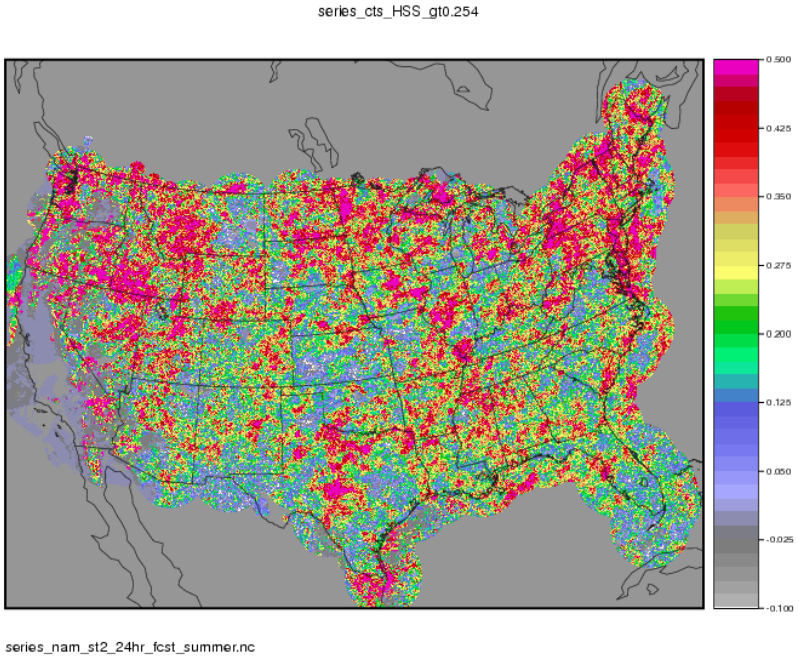
- Define series as:
  - Same field from multiple files.
  - Different fields from the same file.
  - Example: 24hr NAM fcst of 3hr APCP vs StageII



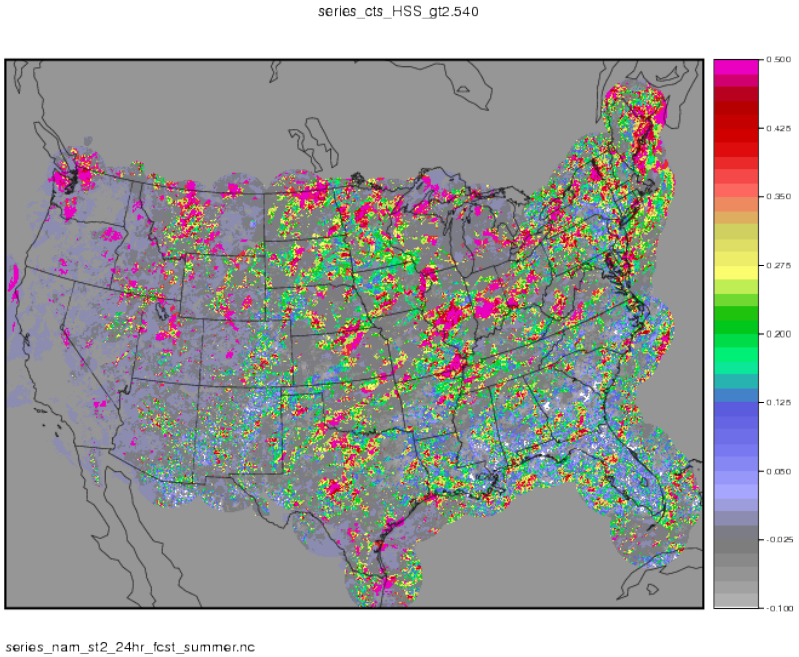


Series-Analysis: Example Writes Gridded Output Statistics

3hr APCP > 0.254 mm
(0.01 in)



3hr APCP > 2.54 mm
(0.1 in)



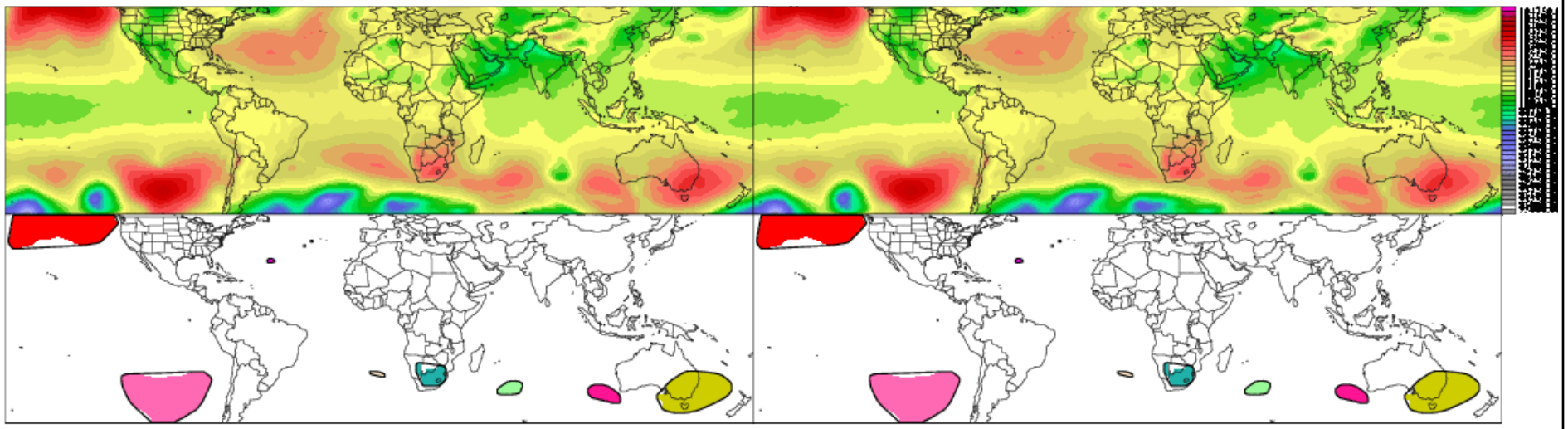
MODE Time-Domain

GFS vs Analysis

f000 to f240 every 6 hours

Objects >= 1025mb

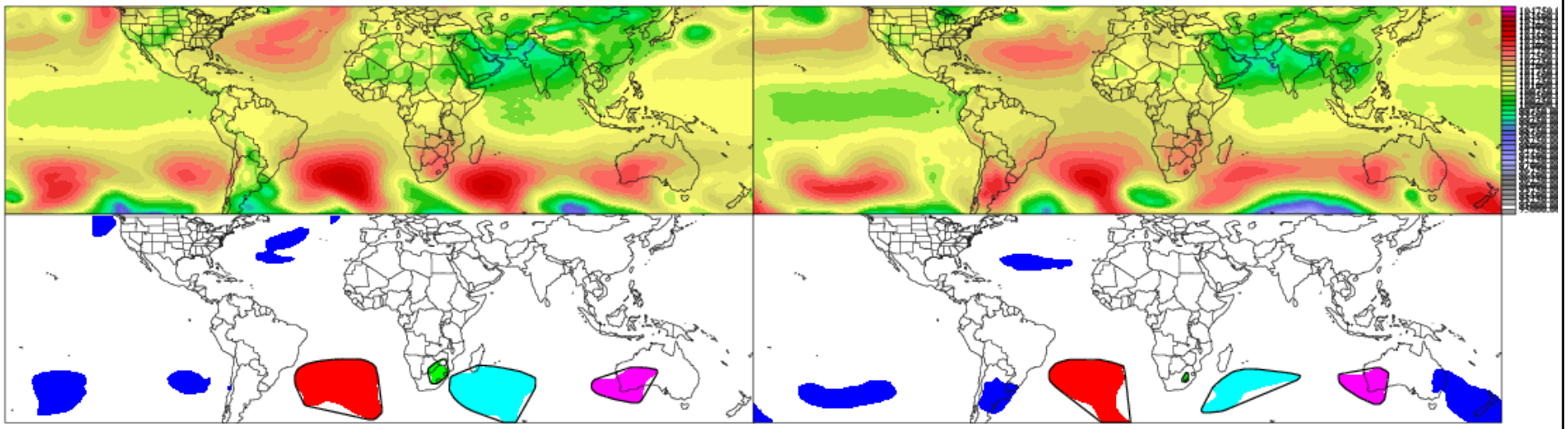
Forecast **Analysis**



MODE Time-Domain

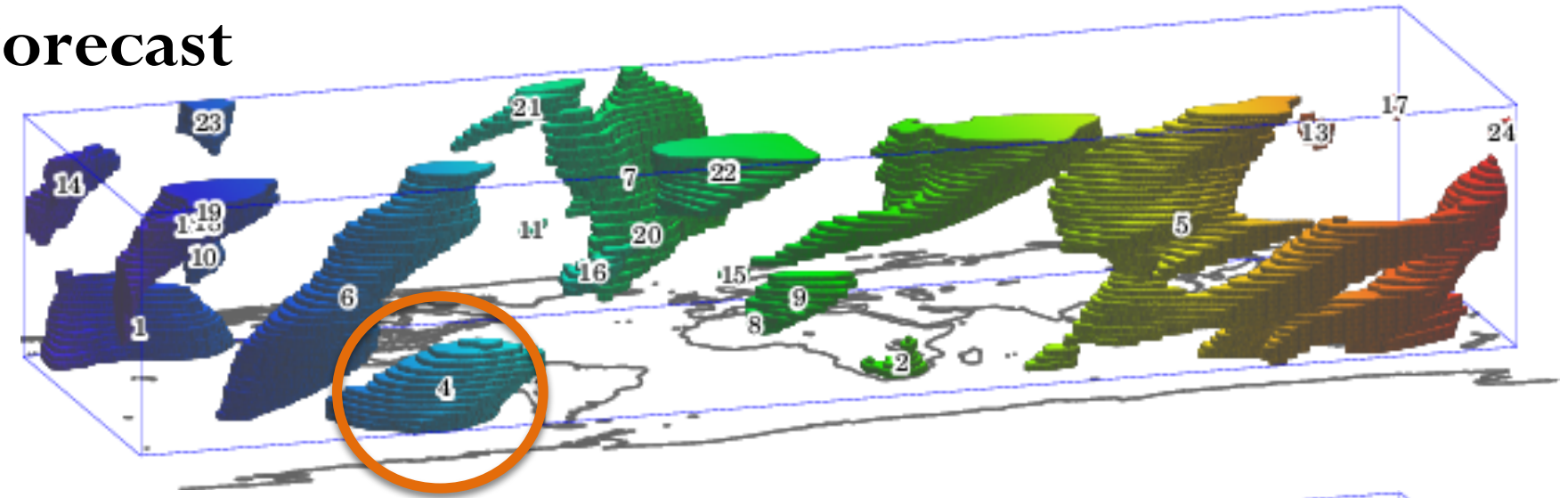
f240	Pair 1	Pair 2	Pair 3	Pair 4
Centroid Distance	4.62	10.31	10.12	11.22
Fcst Area	13983	1097	12146	5132
Obs Area	11584	133	6713	4225
Intersection	10857	133	5723	3610

Forecast **Analysis**

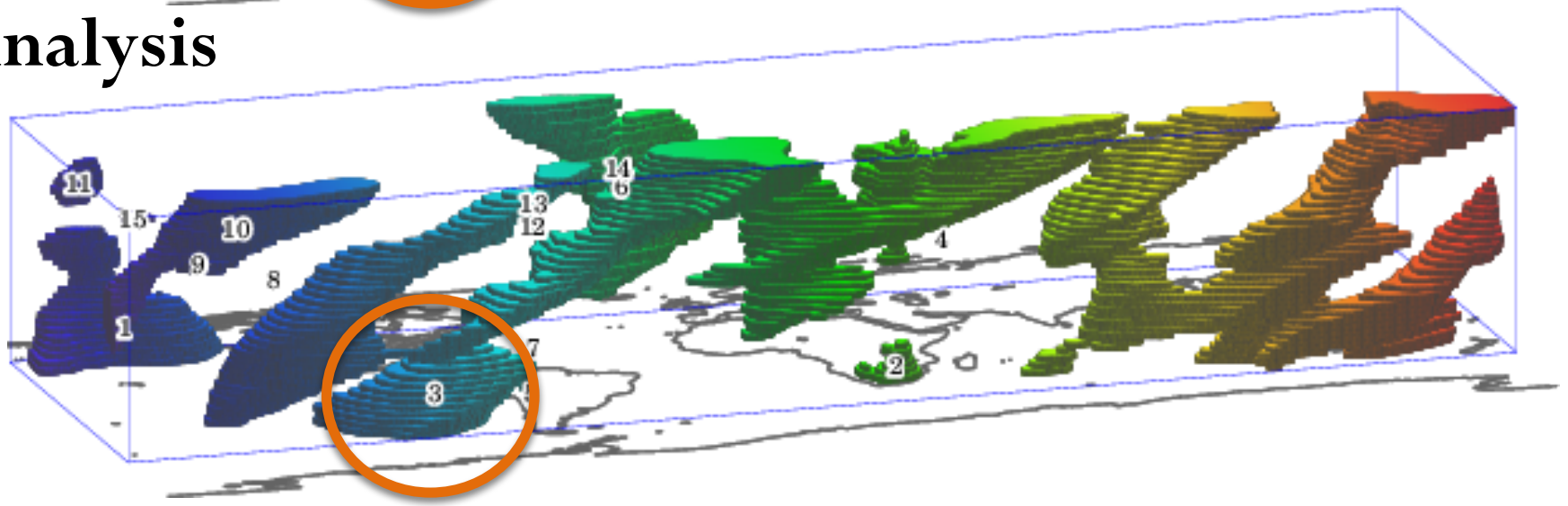


MODE Time-Domain

Forecast

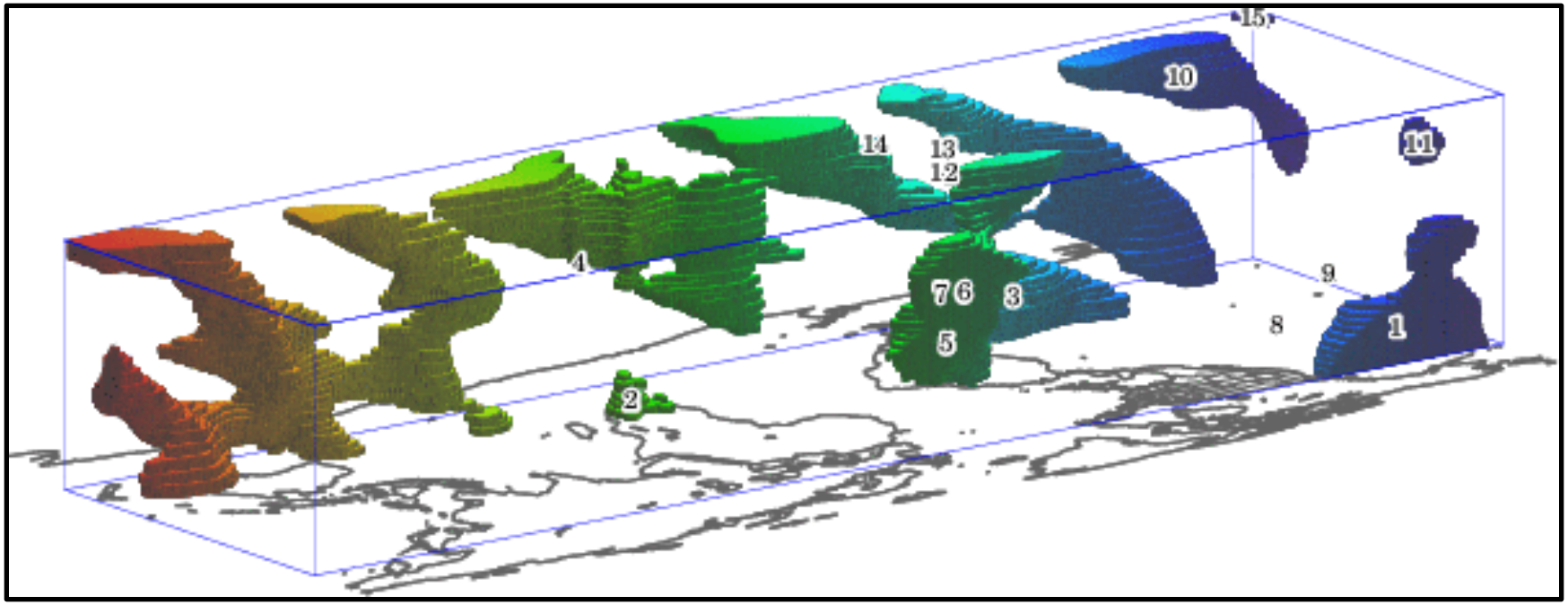


Analysis



MODE Time-Domain

f000 – f240	Max Inten	Volume	Centroid(x,y,t)	Velocity
Fcst Object 4	103927	111493	336, 57, 4.19	2.85
Analysis Object 3	103914	113692	335, 59, 4.27	2.79



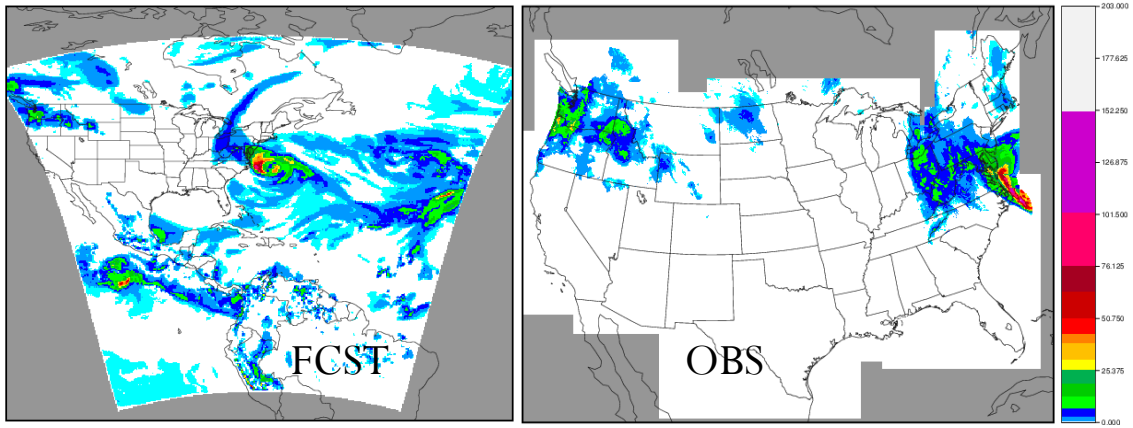
MET Config Entries

```
//
// Output model name to be written
//
model = "WRF";
//
// Output description to be written
// May be set separately in each "obs.field" entry
//
desc = "NA";
//
// Output observation type to be written
//
obtype = "ANALYS";
//
// Verification grid
//
regrid = {
  to_grid    = OBS;
  method     = BUDGET;
  width      = 2;
  vld_thresh = 0.5;
  shape      = SQUARE;
}
```

```
//
// May be set separately for each task
//
censor_thresh    = [ <0 ];
censor_val       = [ 0 ];
//
// Fields to be verified
//
fcst = {
  field = [
    {
      name      = "APCP";
      level     = [ "A03" ];
      cat_thresh = [ >0.0, >=5.0 ];
    }
  ];
}
obs = fcst;
```

Automated Regridding in MET

Config file: grid to verify on: FCST, OBS, or USER DEFINED



Old method: Regrid outside MET

Regrid to FCST or OBS - requires at least 1 more file

Regrid to USER DEFINED - requires 2 more files

Impact #1 – Decreased complexity & storage requirements

Automated regridding could save **0.5 to 7.5 GB per operational cycle**

Equates to **60 GB – 1 TB per month** of storage

Impact #2 – Less complexity for using climatologies

Climatologies may not be on same grid as forecasts.

See Impact #1

MTD

Series Analysis

MODE

Wavelet Stat

Grid Stat

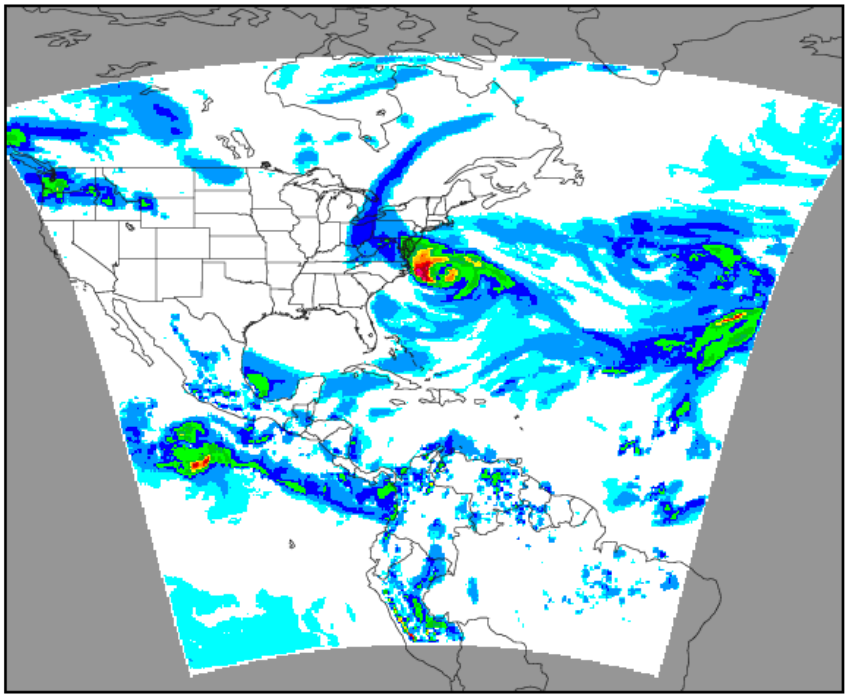
Ensemble-Stat

Point Stat

Automated Regridding

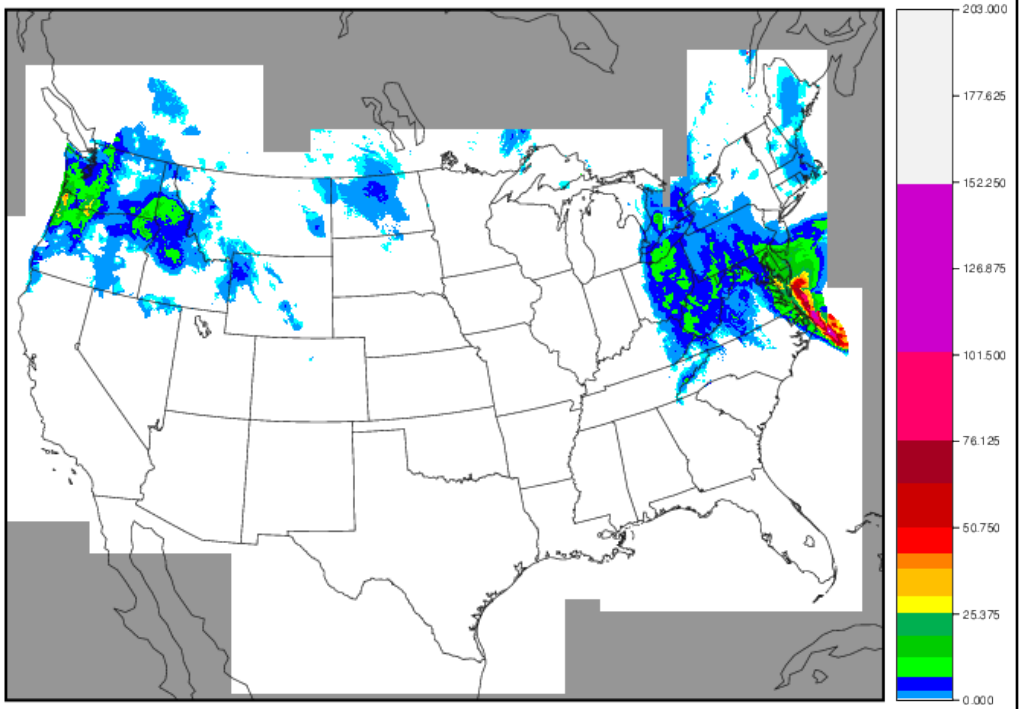
0.25° HWRF

Sandy 6-hour APCP



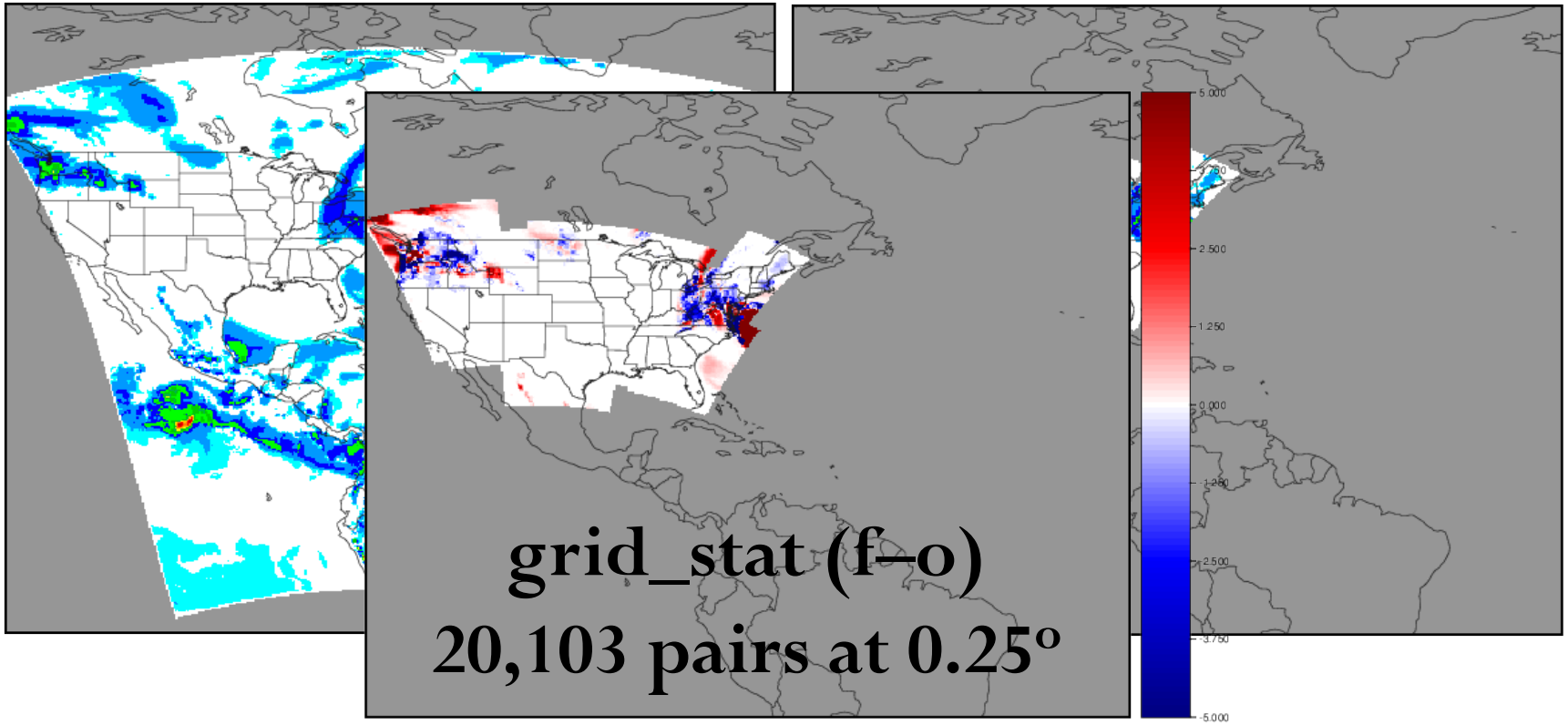
4km StageIV

6-hour APCP



Automated Regridding

to_grid = FCST;

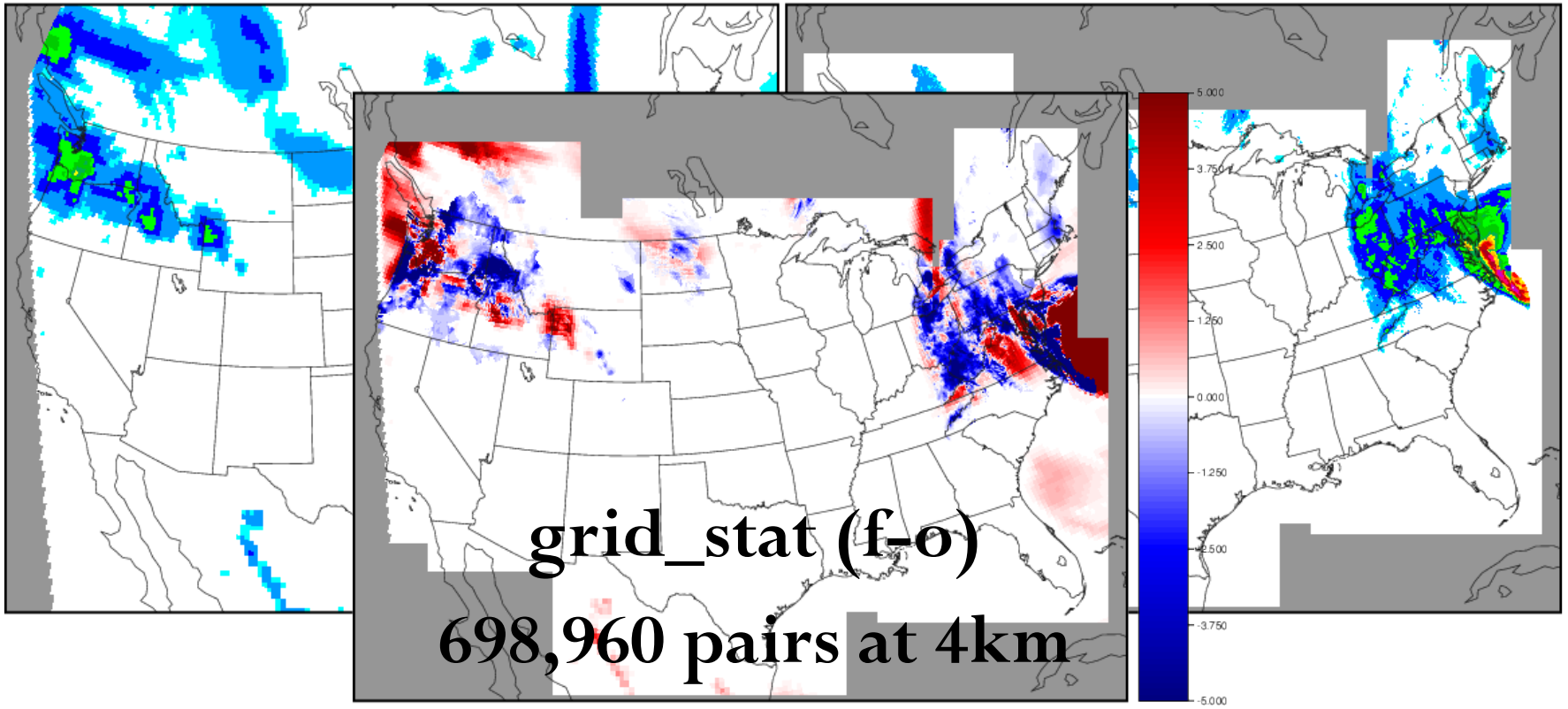


grid_stat (f–o)

20,103 pairs at 0.25°

Automated Regridding

to_grid = OBS;

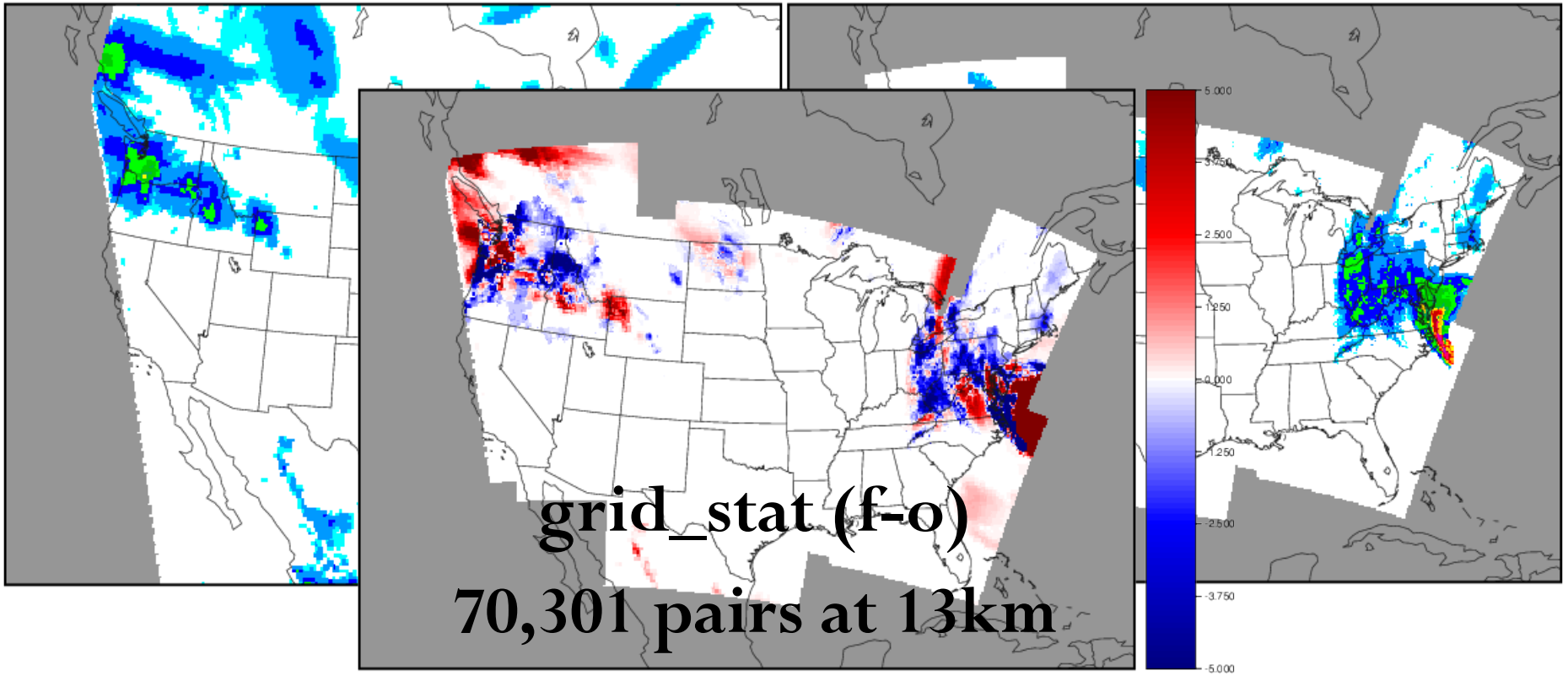


grid_stat (f-o)

698,960 pairs at 4km

Automated Regridding

to_grid = "G130";



grid_stat (f-o)

70,301 pairs at 13km

- Named grid, gridded data file, or grid definition string

Output from Point and Grid-Stat

```
output_flag = {
   fho      = STAT;
   ctc      = STAT;
   cts      = STAT;
   mctc     = STAT;
   mcts     = STAT;
   cnt      = STAT;
   sl1l2    = STAT;
   sal1l2   = STAT;
   vl1l2    = STAT;
   val1l2   = STAT;
   vcnt     = STAT;
   pct      = NONE;
   pstd     = NONE;
   pjc      = NONE;
   prc      = NONE;
   eclv     = NONE;
   nbrctc   = STAT; (*GS)
   nbrcts   = STAT; (*GS)
   nbrcnt   = STAT; (*GS)
   grad     = STAT; (*GS)
   mpr      = STAT (*PS)
}
```

CATEGORICAL

CONTINUOUS

VECTOR WINDS

PROBABILISTIC

NEIGHBORHOOD

MATCHED PAIRS

CTC Output Line Type

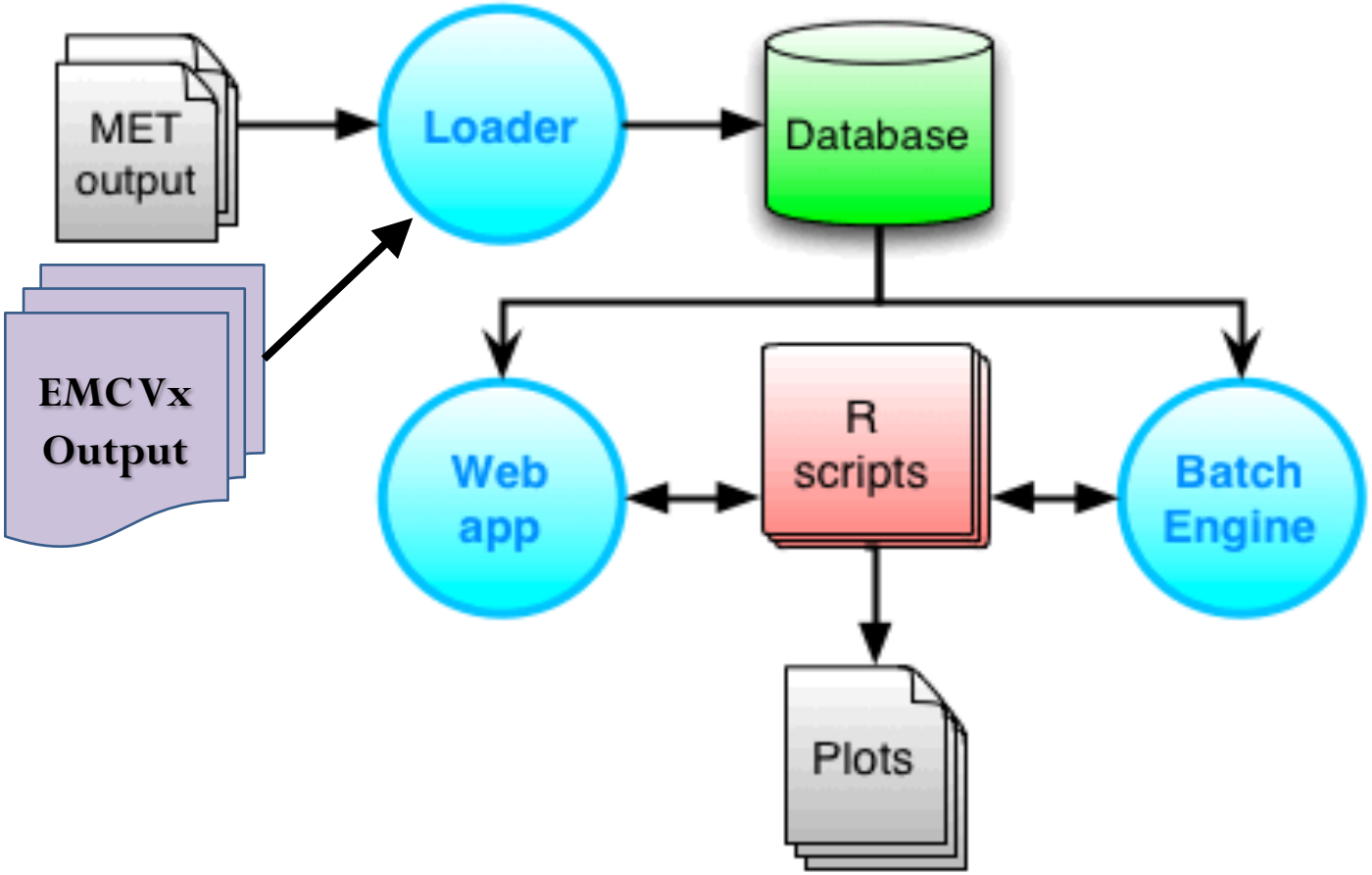
VERSION	V7.0
MODEL	WRF
DESC	NA
FCST_LEAD	360000
FCST_VALID_BEG	20070331_120000
FCST_VALID_END	20070331_120000
OBS_LEAD	000000
OBS_VALID_BEG	20070331_103000
OBS_VALID_END	20070331_133000
FCST_VAR	TMP
FCST_LEV	Z2
OBS_VAR	TMP
OBS_LEV	Z2
OBTYPE	ADPSFC

VX_MASK	FULL
INTERP_MTHD	UW_MEAN
INTERP_PNTS	1
FCST_THRESH	>273.000
OBS_THRESH	>273.000
COV_THRESH	NA
ALPHA	NA
LINE_TYPE	CTC
TOTAL	**4003**
FY_OY (hits)	**3111**
FY_ON (f.a.)	**78**
FN_OY (miss)	**215**
FN_ON (c.n.)	**599**

METViewer

METViewer components

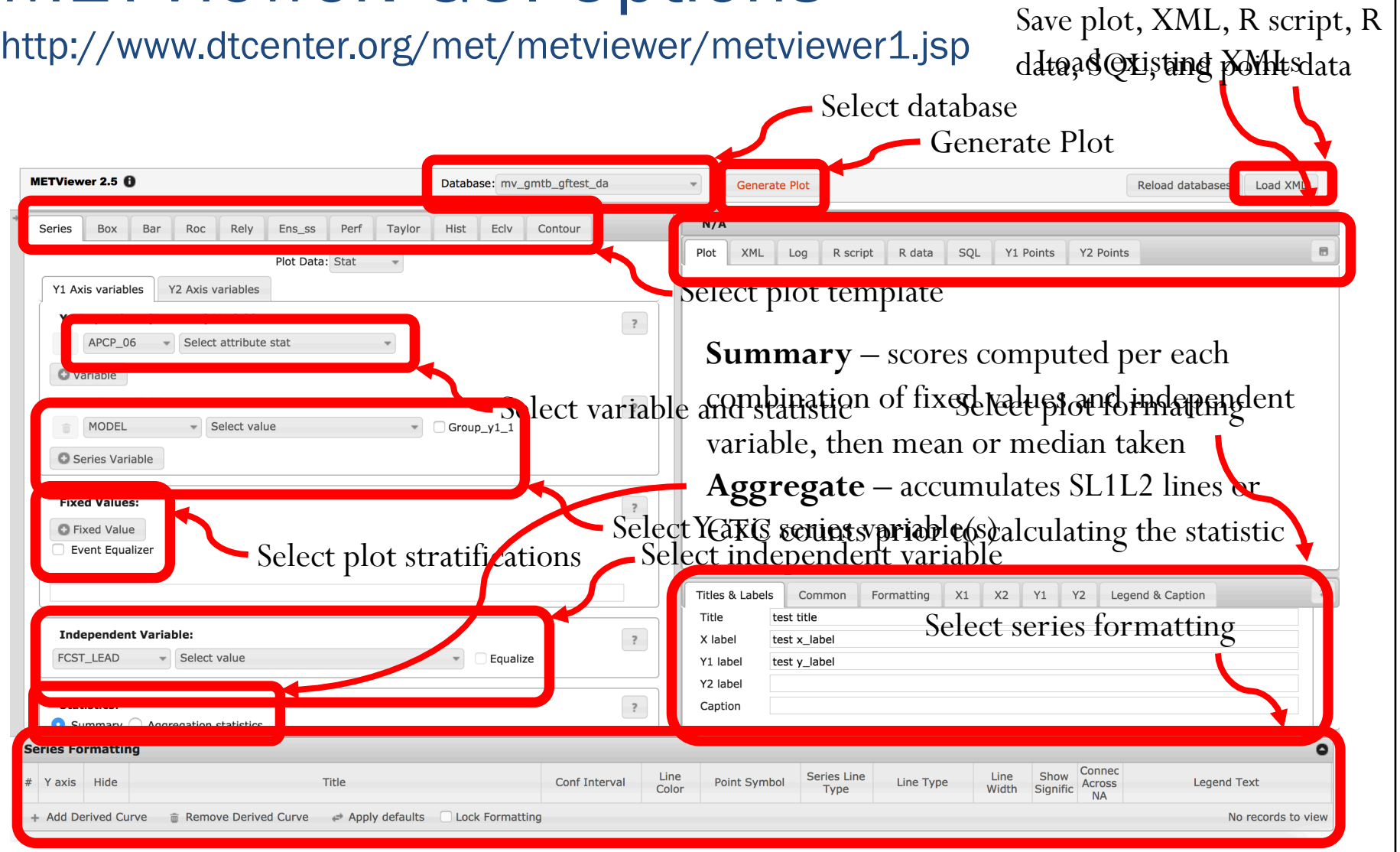
Packages: Java, Apache/Tomcat, MySQL, R statistics



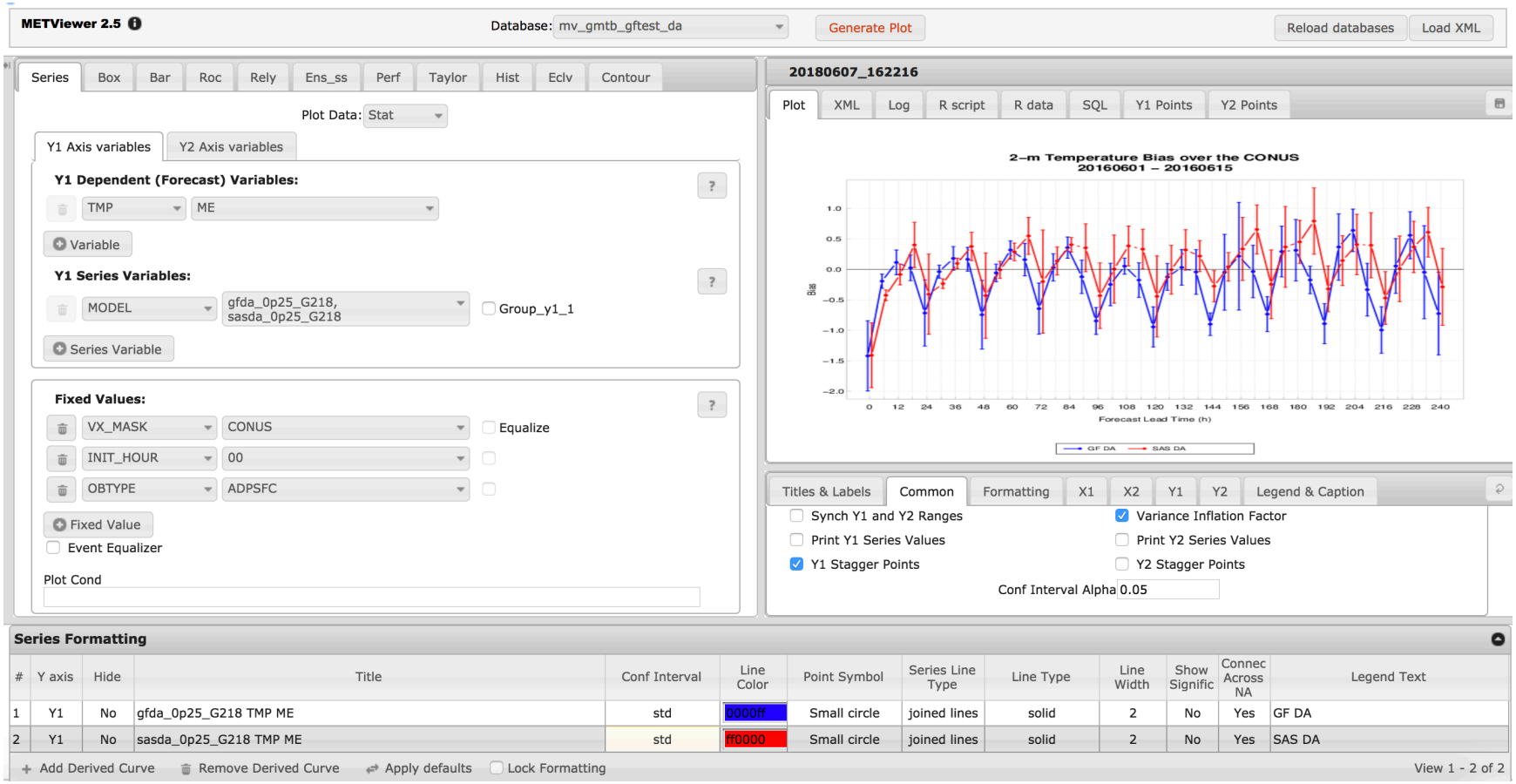
Database and Display analysis tool

METViewer: GUI Options

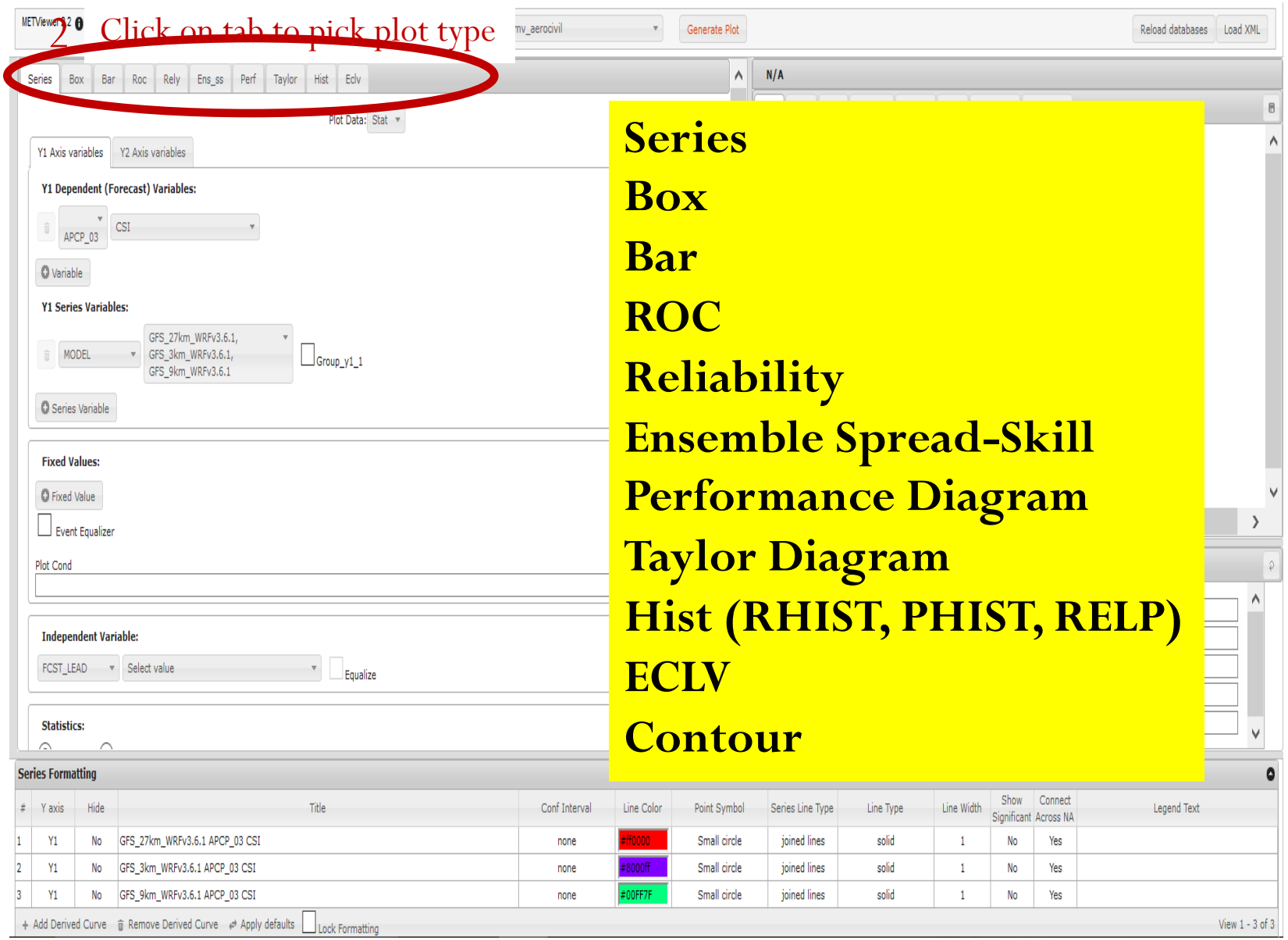
http://www.dtcenter.org/met/metviewer/metviewer1.jsp



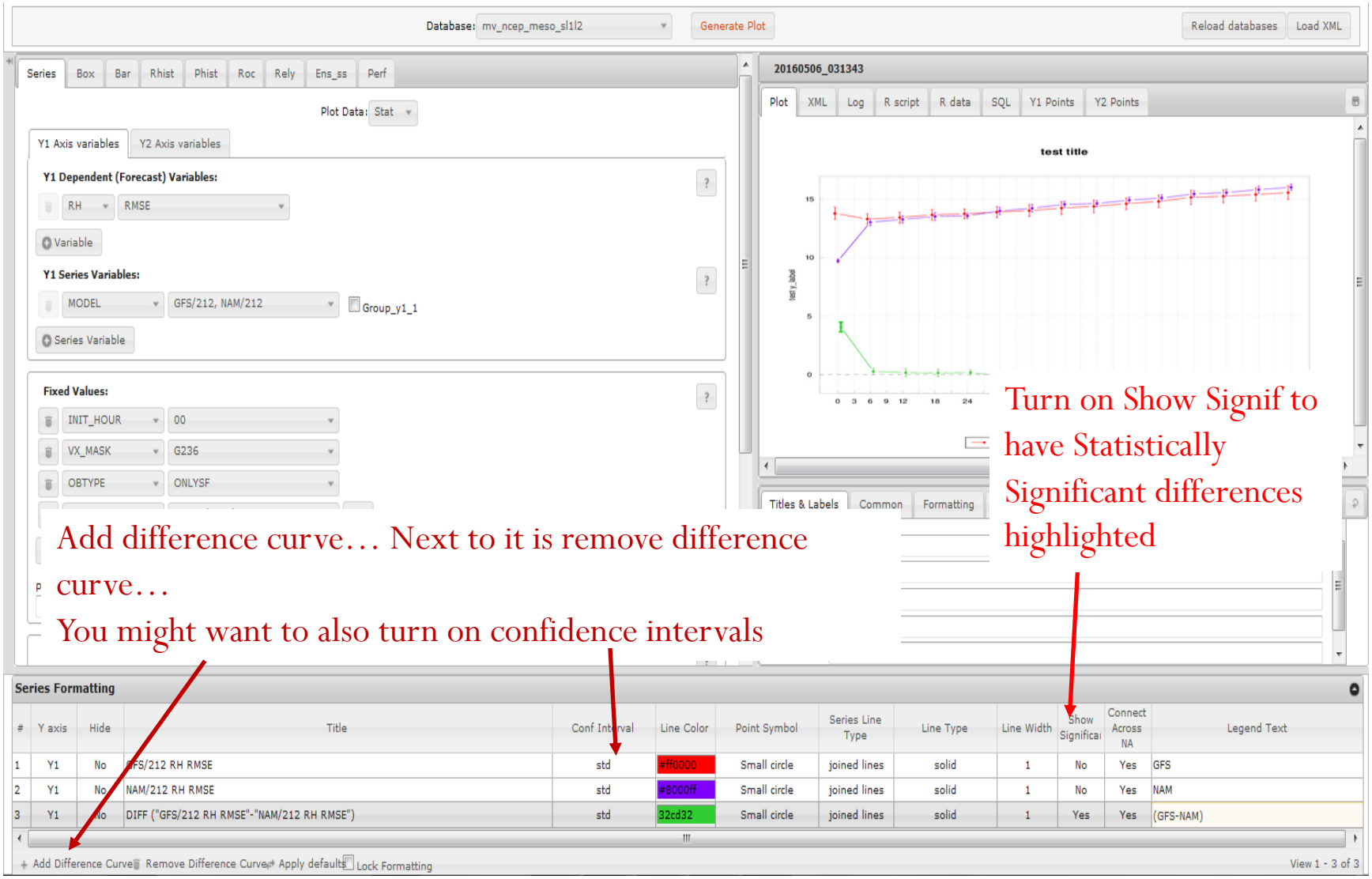
METViewer: GUI Plot



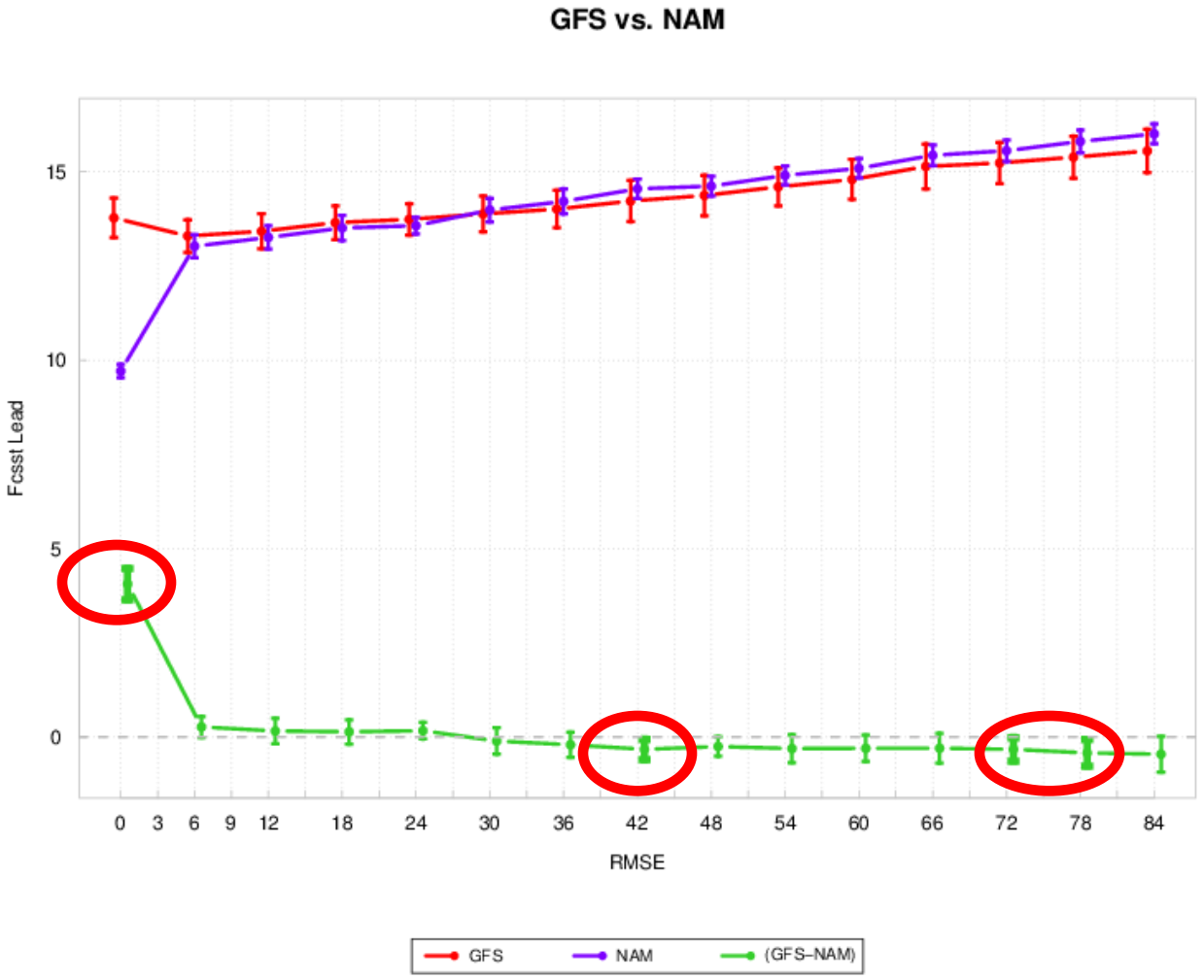
METViewer: Plot Templates



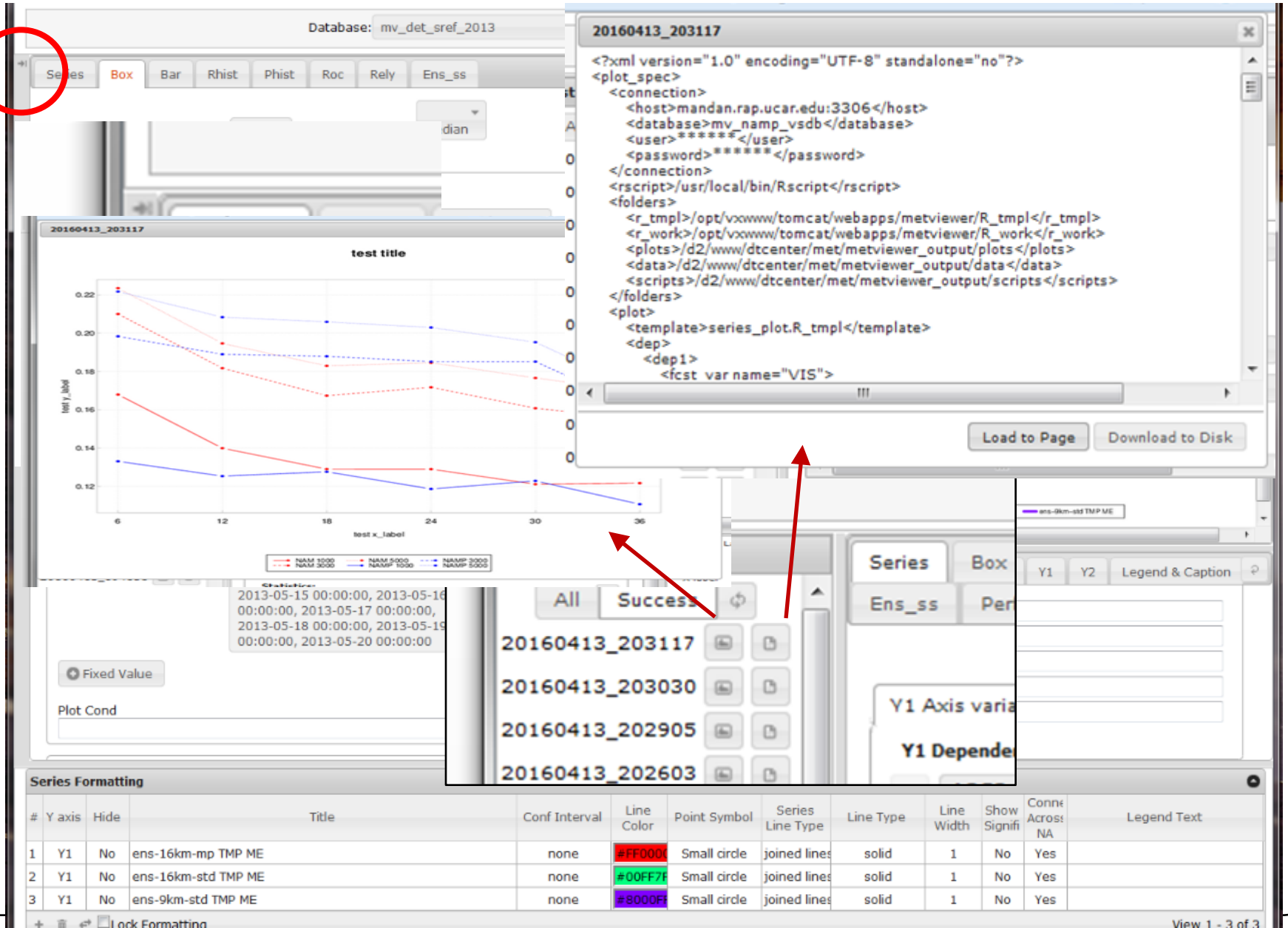
METViewer: Statistically Significant Differences



Statistically Significant Plot

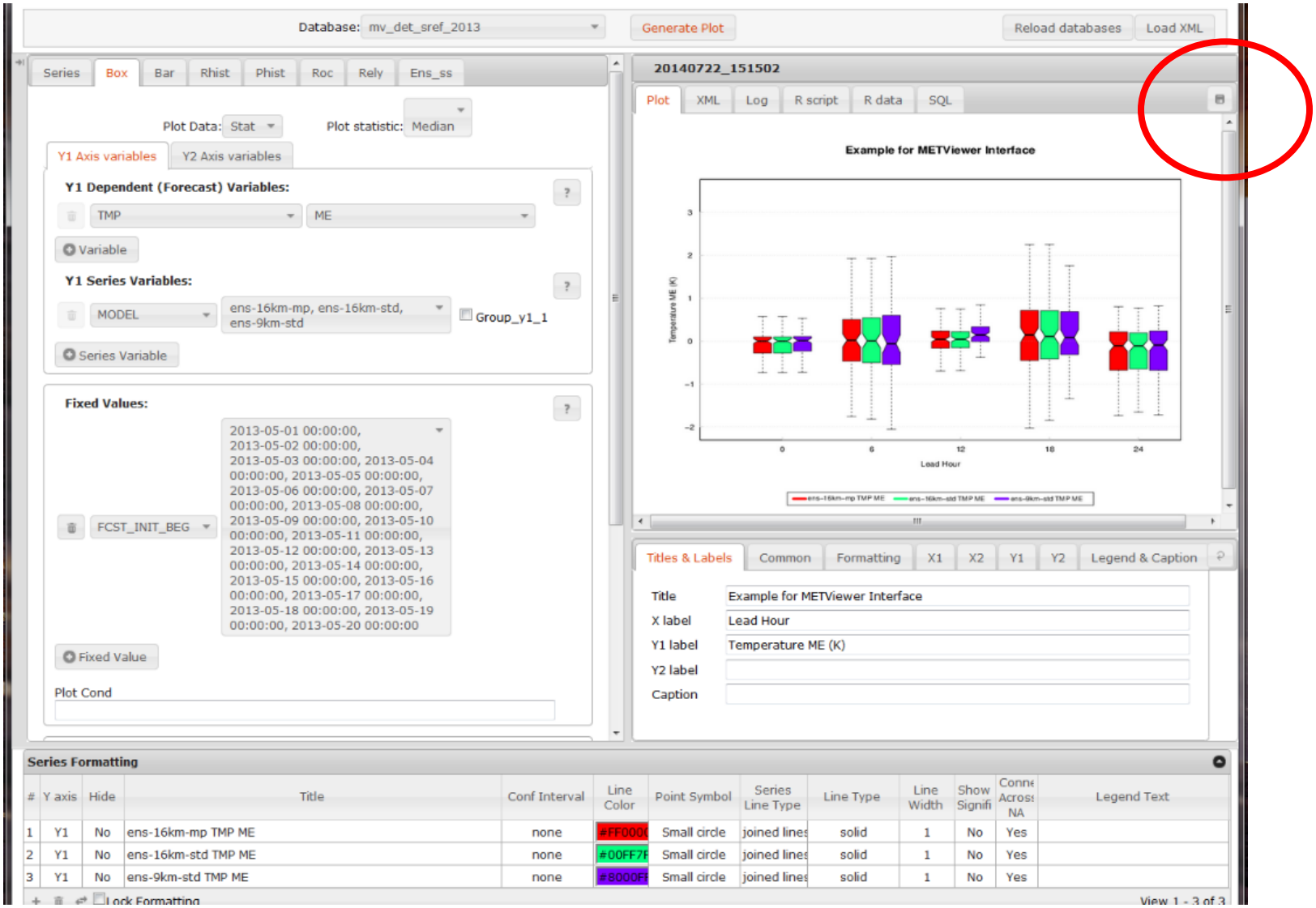


METViewer History



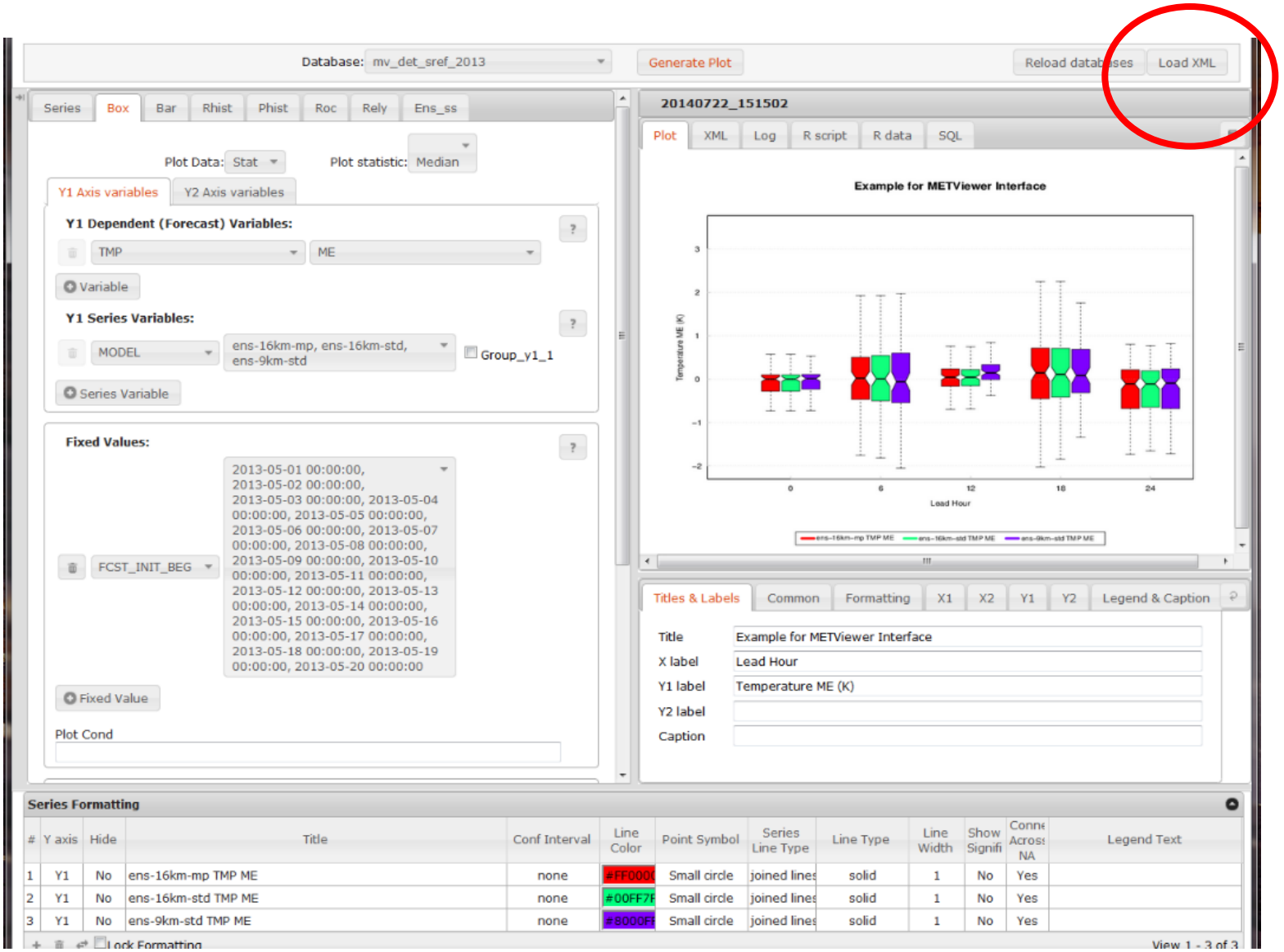
METViewer

Save Plots, XML,
Data, Rscrips, etc
**Based on which tab is selected



METViewer

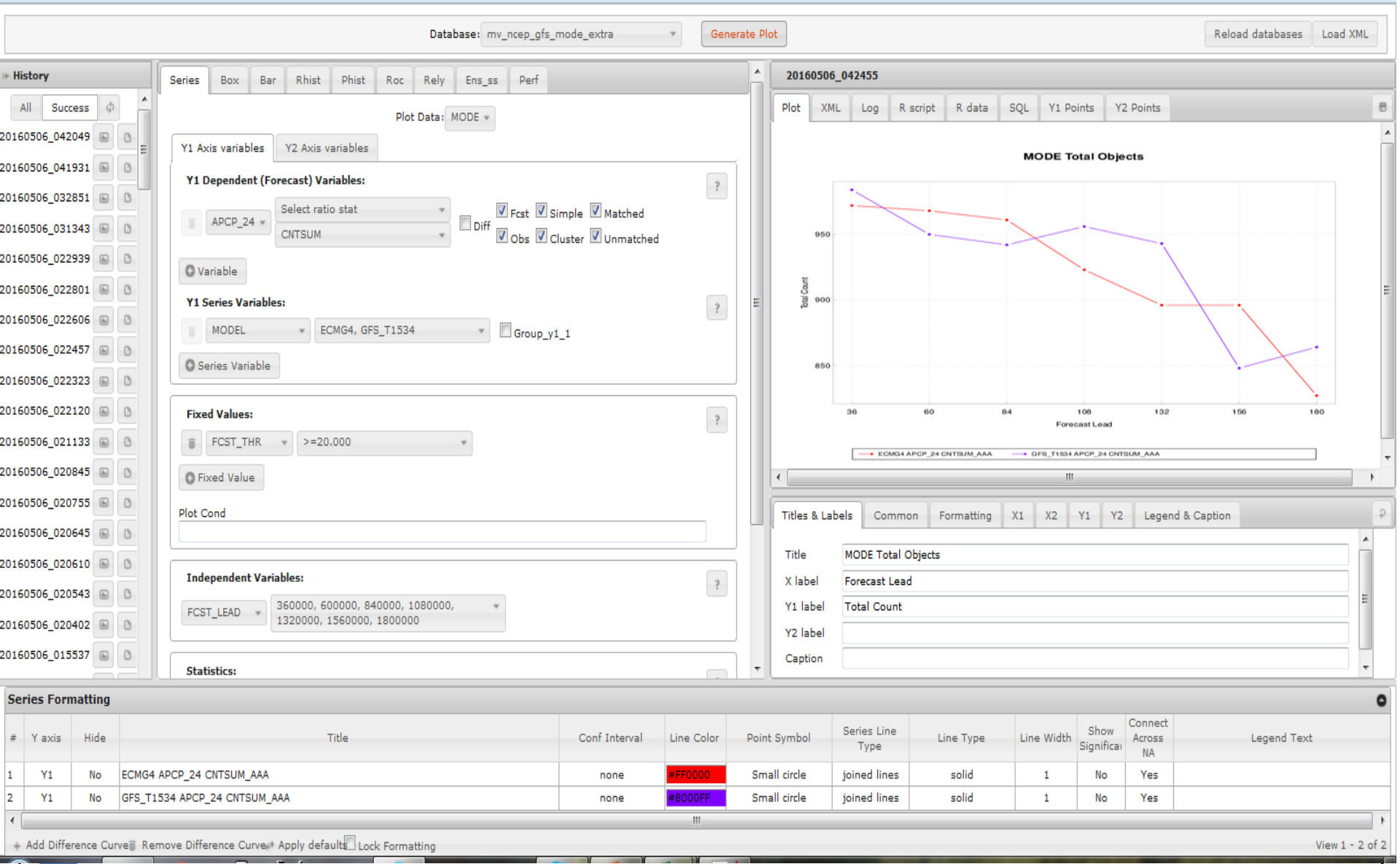
Upload XML scripts from your system



METViewer


Database: mv_det_sref_2013
Generate Plot
Reload databases
Load XML
Series
Box
Bar
Plot Data: Stat
Y1 Axis variables
Y2 Axis variables
Y1 Dependent (Forecast) Variables
TMP
ME
Variable
Y1 Series Variables:
MODEL
ens-16km-mp, ens-16km-std, ens-9km-std
Group_y1_1
Series Variable
Fixed Values:
FCST_INIT_BEG
Fixed Value
Plot Cond
Plot Data: Choose between Stat, MODE, and MTD
SQL
for METViewer Interface
Temperature ME (K)
Lead Hour
Titles & Labels
Common
Formatting
X1
X2
Y1
Y2
Legend & Caption
Title
Example for METViewer Interface
X label
Lead Hour
Y1 label
Temperature ME (K)
Y2 label
Caption
Series Formatting
| Y axis | Hide | Title | Conf Interval | Line Color | Point Symbol | Series Line Type | Line Type | Line Width | Show Signifi | Conne Across NA | Legend Text
1 | Y1 | No | ens-16km-mp TMP ME | none | Small circle | joined lines | solid | 1 | No | Yes
2 | Y1 | No | ens-16km-std TMP ME | none | Small circle | joined lines | solid | 1 | No | Yes
3 | Y1 | No | ens-9km-std TMP ME | none | Small circle | joined lines | solid | 1 | No | Yes
Lock Formatting
View 1 - 3 of 3

MODE Interface



METViewer Scorecard

Specify Statistics

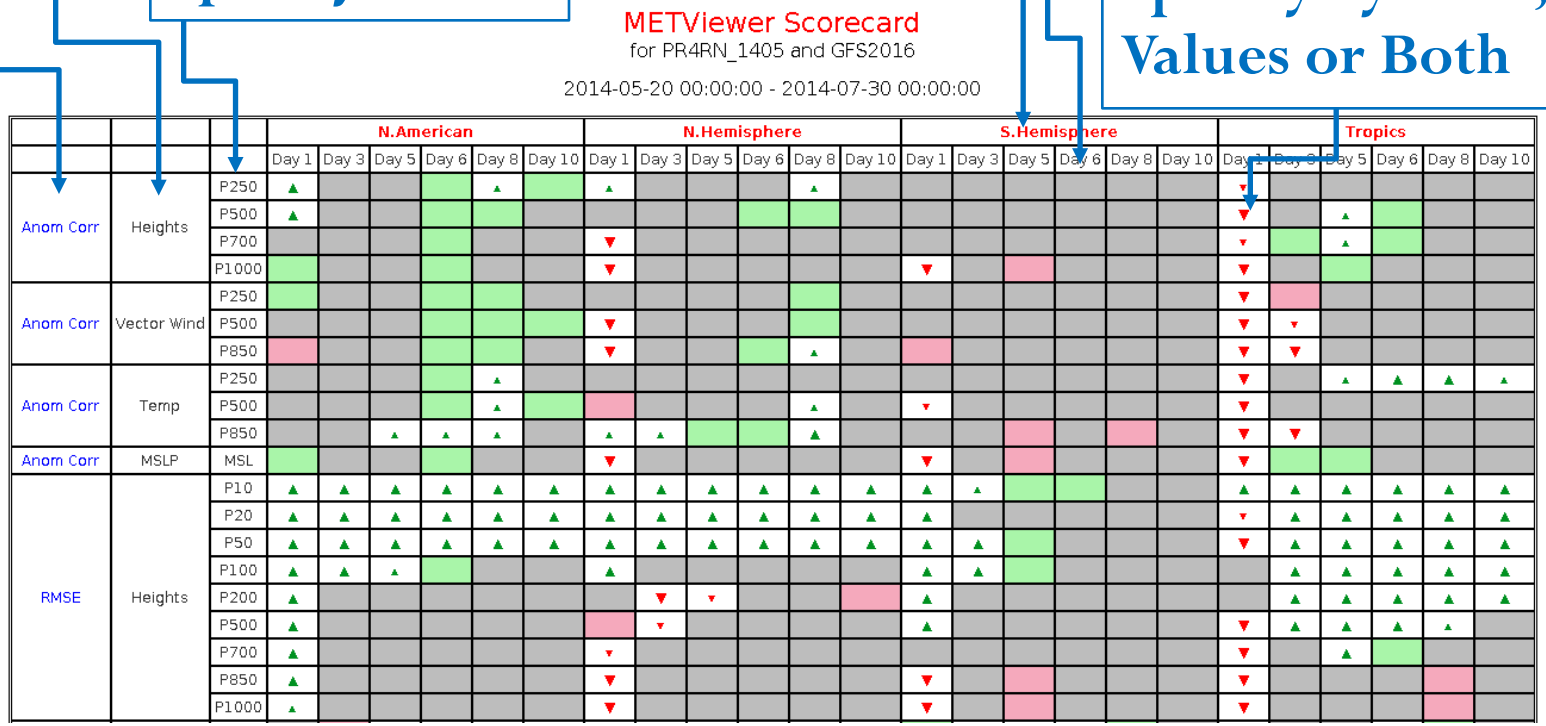
Specify Variables

Specify Levels

Specify Regions

Specify Aggregations

Specify Symbol, Values or Both



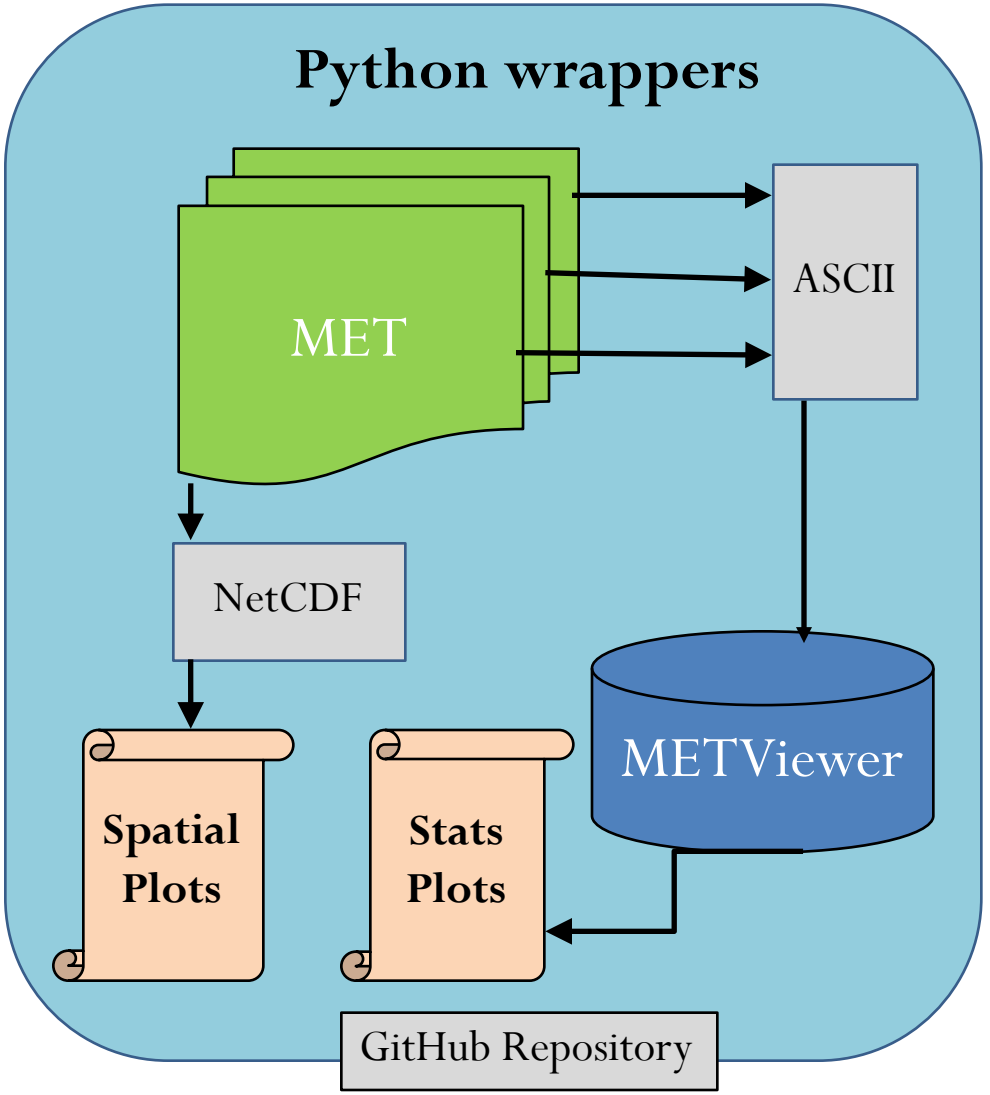
Submit to batch engine of METViewer

MET+

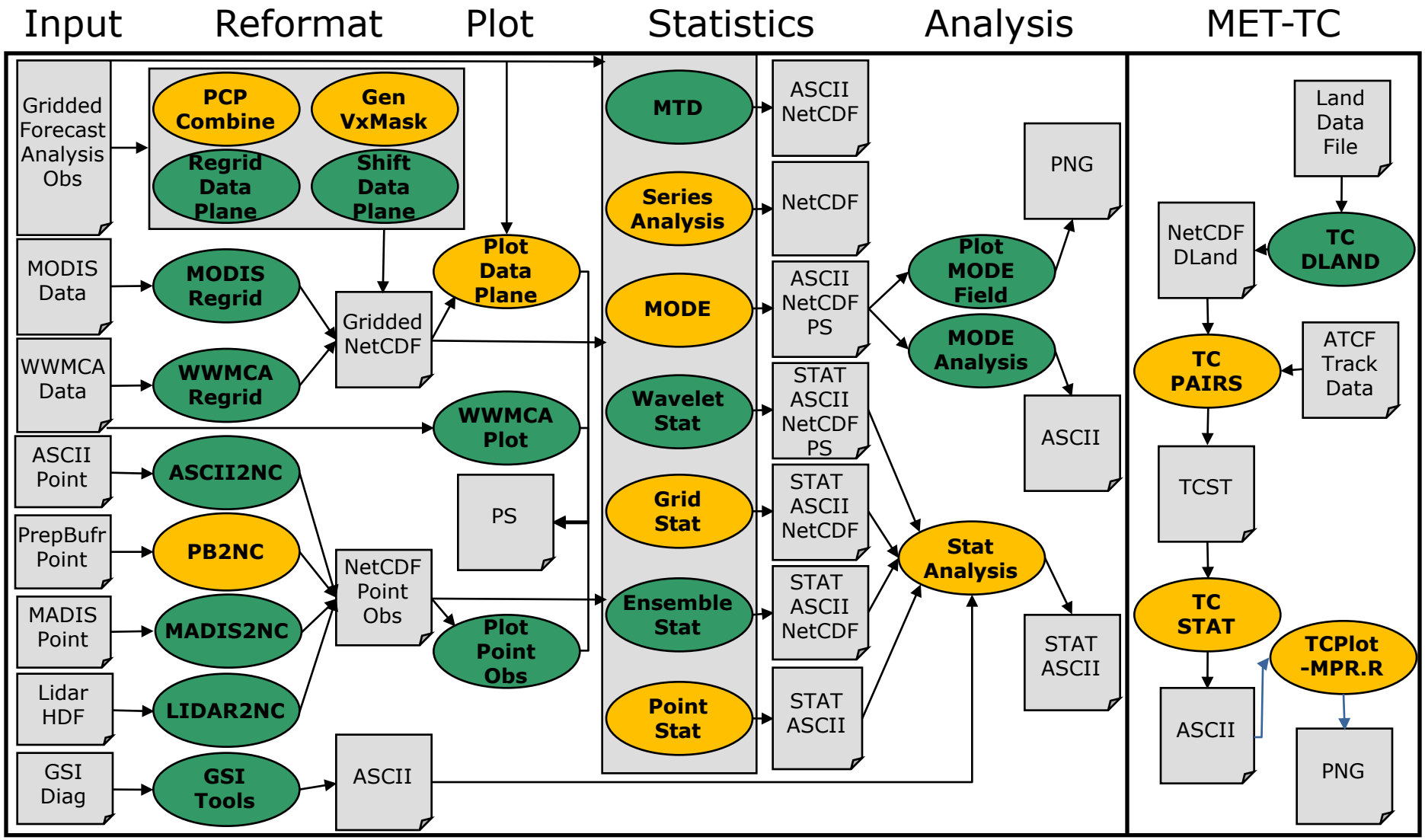
MET+ Unified Package

- Python wrappers around MET and METViewer:
- Simple to set-up and run
- Automated plotting of 2D fields and statistics

Initial system - Global deterministic with plans to generalize across scales when possible to quickly spin-up Ensembles, High Resolution & Global Components



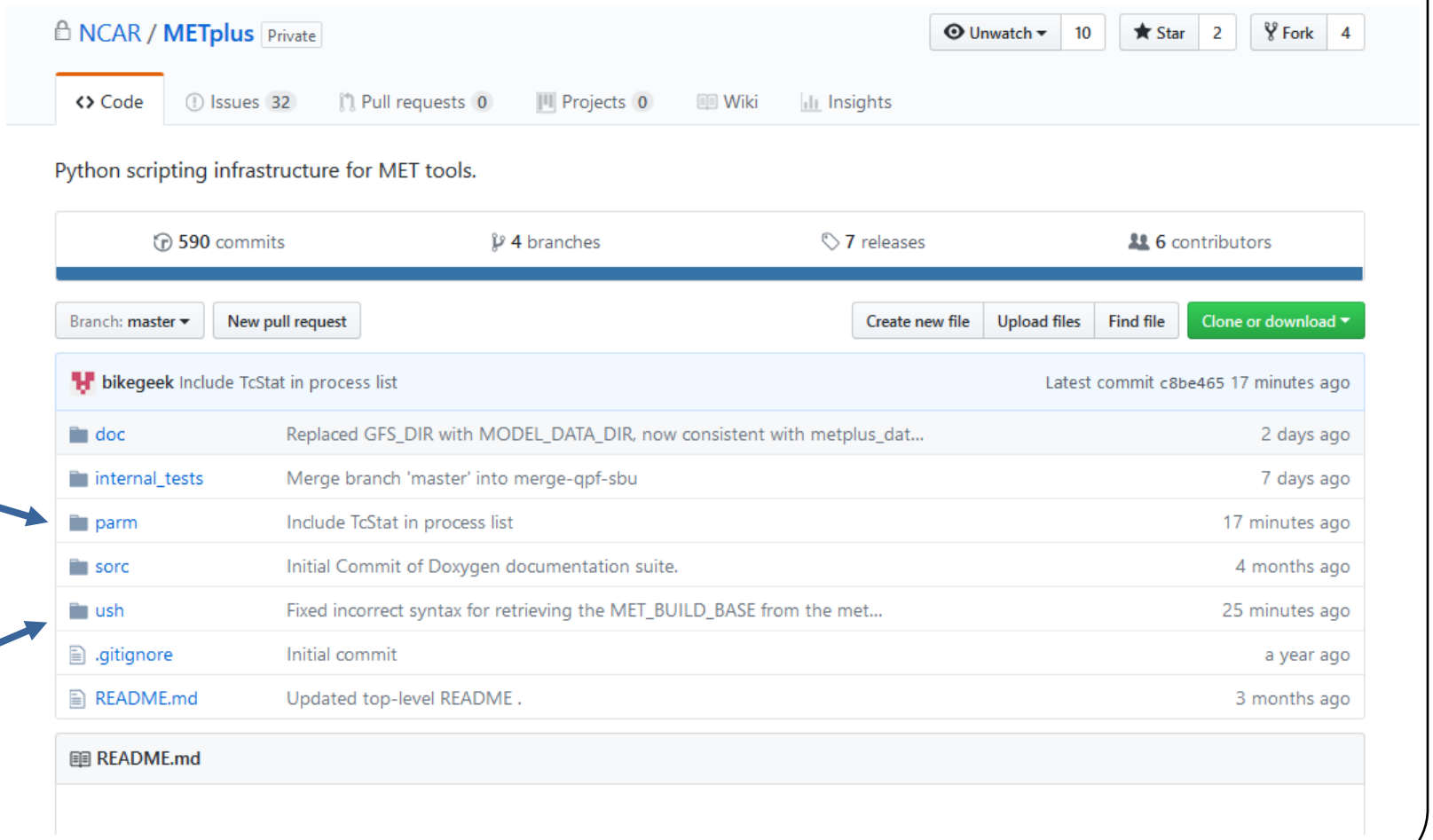
Currently Wrapped by a Use Case



Copyright 2017, University Corporation for Atmospheric Research, all rights reserved

What does wrapped by Python mean?

At https://github.com/NCAR/METplus/



Control File and Config

Python Scripts

What does wrapped by Python mean?

METplus/parm/use_cases/feature_relative

feature_relative.conf

```
#
#     LISTS AND SETTINGS
#

#     Processes to run in master script (master_met_plus.py)

PROCESS_LIST = ["run_tc_pairs.py", "extract_tiles.py", "series_by_lead.py"]

#
#     NOTE: "TOTAL" is a REQUIRED cnt statistic used by the series analysis scripts
#

STAT_LIST = ["TOTAL", "FBAR", "OBAR", "ME", "MAE", "RMSE", "BCMSE", "E50", "EIQR", "MAD"]

#     Dates must be in YYYYMMDD format
#     INIT_HOUR_INC is the increment in integer format
#     INIT_HOUR_END should be a string in HH or HHH format

INIT_DATE_BEG = "20141201"
INIT_DATE_END = "20150331"
INIT_HOUR_INC =  6
INIT_HOUR_END = "18"

#     Used by extract_tiles.py to define the records of interest from the grib2 file

VAR_LIST = ["HGT/P500", "PRMSL/Z0", "TMP/Z2", "PWAT/L0", "HGT/P250", "TMP/P850", "TMP/P500", "UGRD/P250", "VGRD/P250" ]
EXTRACT_TILES_VAR_LIST = []

#     Used for performing series analysis based on lead time
```

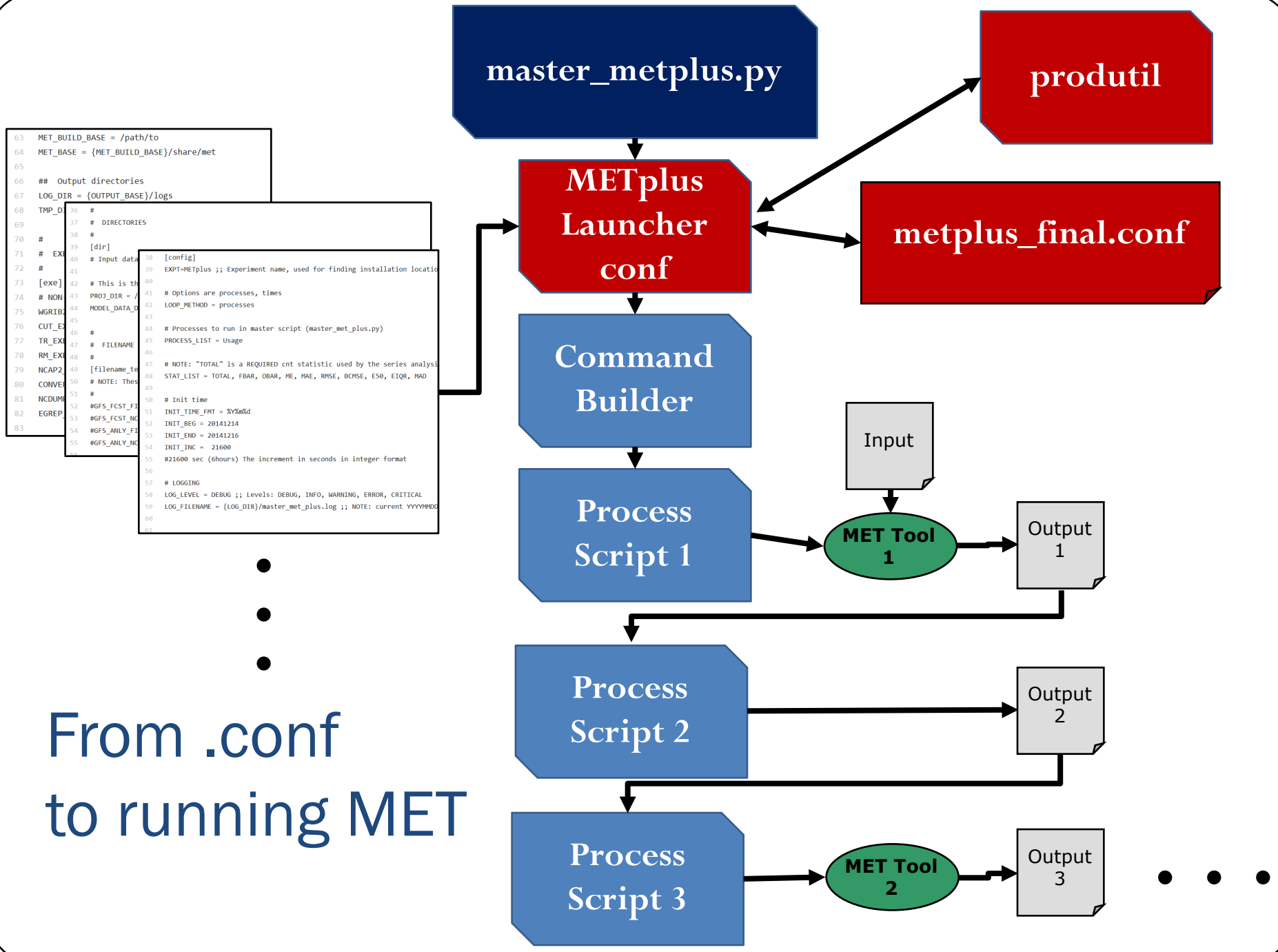
What does wrapped by Python mean?

At https://github.com/NCAR/METplus/

In Configs: Environment variables passed in from Constants File

Series_Analysis_Config

```
30
31  ////////////////////////////////////////////////////////////////////////////////
32
33  cat_thresh  = [ NA ];
34  cnt_thresh  = [ NA ];
35  cnt_logic   = UNION;
36
37  //
38  // Forecast and observation fields to be verified
39  //
40  fcst = {
41
42     field = [
43       {
44         name  = "${NAME}";
45         level = [ "${LEVEL}" ];
46       }
47     ];
48
49  }
50  obs = fcst;
51
52  ////////////////////////////////////////////////////////////////////////////////
53
54  //
55  // Climatology mean data
```

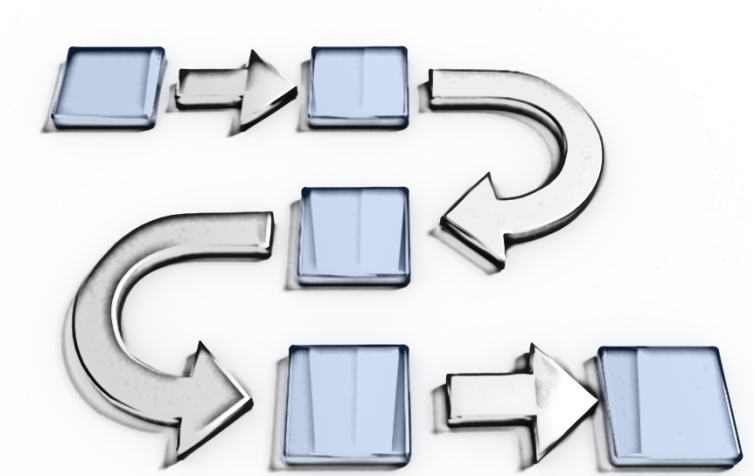

master_metplus.py
produtil
METplus Launcher conf
metplus_final.conf
Command Builder
Process Script 1
Process Script 2
Process Script 3
Input
MET Tool 1
MET Tool 2
Output 1
Output 2
Output 3
From .conf
to running MET

MET+ Coding Standards

- NCEP Coding Standards
- NCO WCOSS Implementation Standards for directory structure and script naming conventions (http://www.nco.ncep.noaa.gov/idsb/implementation_standards/)
- pep8 for Code Style
- Doxygen and Python docstrings for documentation

Aligning with NCEP Workflow

- Discussing and planning to collaborate on new dynamical core (FV3) workflow developers.
- Using PRODUTIL package for logging and constants file parsin plus likely other features
- Plan to use Rocoto workflow management (NOAA tool) for dev environments and make autonomous to also use with ecFlow or Cylc for operations



MET+ Beta - Prerequisites

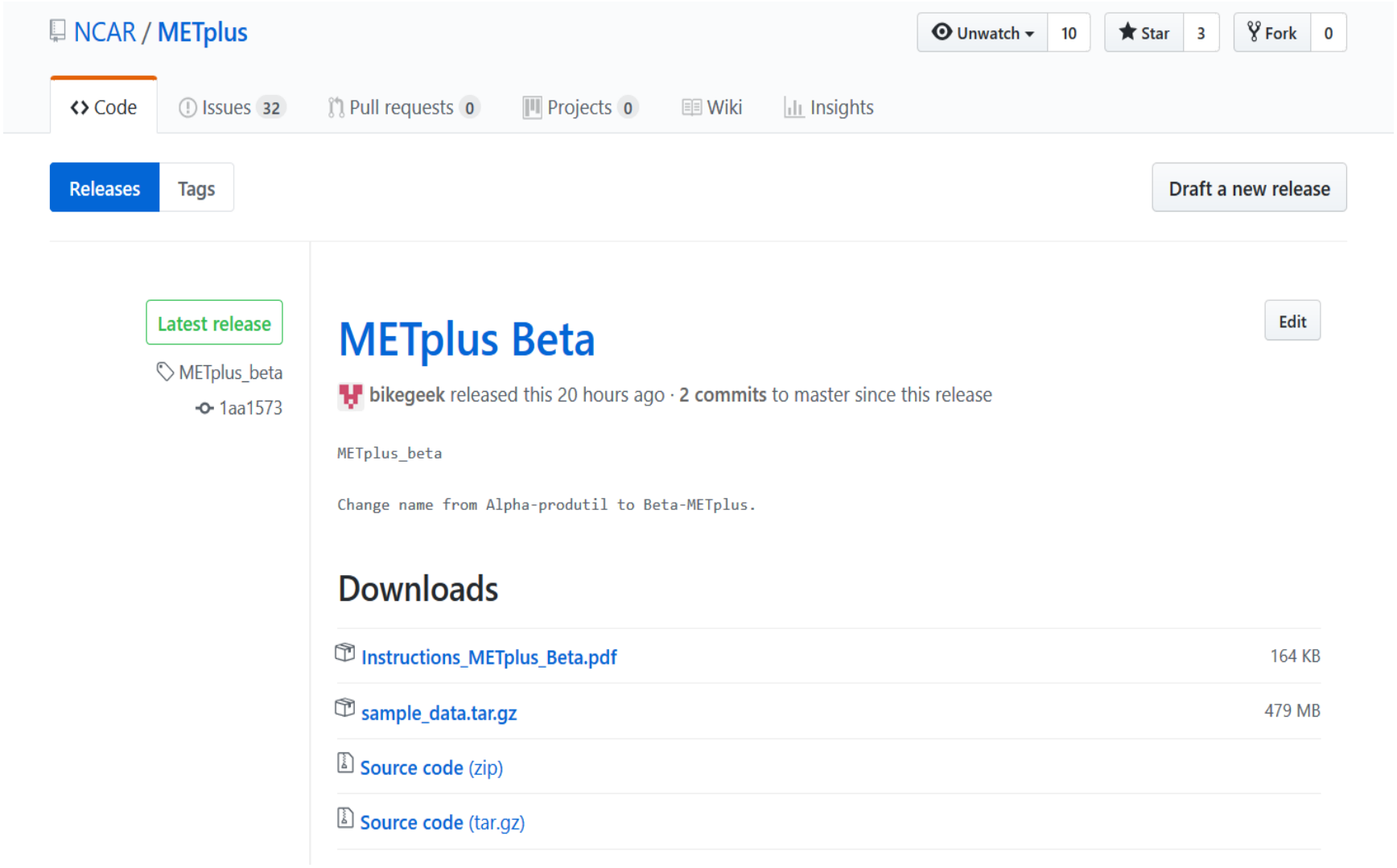
- Python 2.7 ***When we started this was specified by NCO*
- R version 3.25 ** *Only if you are using PlotTCMpr.R tool*
- nco (netCDF operators)
- MET version 6.0 or later installed
  ** *Tool is designed to sit on-top of MET and should be version insensitive after METv6.0*
- Basic familiarity with MET

- **User:** Access the public release at:
  https://github.com/NCAR/METplus/releases

-OR-

- Use install on Theia or WCOSS
  ** *Only on Gyre right now, will populate on Surge, Tide and Luna as access is available*
- **Developer:** Need a github account
  https://github.com/NCAR/METPlus/ then proceed like a User

Grabbing the Release



Operational Directory Structure

- doc/ - Doxygen documentation
- internal_tests/ - developer tests
- parm/ - where configs live
- README.md - general README
- sorc/ - executables
- ush/ - python scripts

3 Use Cases

- Track and Intensity
  - Using MET-TC to pair up ATCF track files
  - PlotTCMPR.R to compute track and intensity errors and plot
- Feature Relative
  - Use MET-TC to pair up ATCF track files
  - Extract 30deg by 30deg tiles from GFS Forecast and Analysis files for comparison
  - Use Series-Analysis to compute statistics for the stack of tiles over CONUS
  - Use Plot-Data-Plane to generate quick look plots
- QPF
  - Use Pcp-Combine to accumulate 1-hr QPE into 3-hr accumulation
  - Use Grid-Stat to compute Categorical statistics

Ensemble Verification Metrics

Tressa Fowler

Acknowledgments: Laurie Wilson, Barbara Brown, Matt Pocernich, Tom Hamill, CAWCR

Questions to ask before beginning?

- How were the ensembles constructed?
  - Poor man's ensemble (distinct members)
  - Multi-physics (distinct members)
  - Random perturbation of initial conditions (anonymous members)
- How are your forecasts used?
  - Improved point forecast (ensemble mean)
  - Probability of an event
  - Full distribution

Approaches to evaluating ensemble forecasts

- As individual members
  - Use methods for continuous or categorical forecasts
- As probability forecasts
  - Create probabilities by applying thresholds or statistical post-processing
- As a full distribution
  - Use individual members or fit a distributions through post-processing

Approaches to evaluating ensemble forecasts

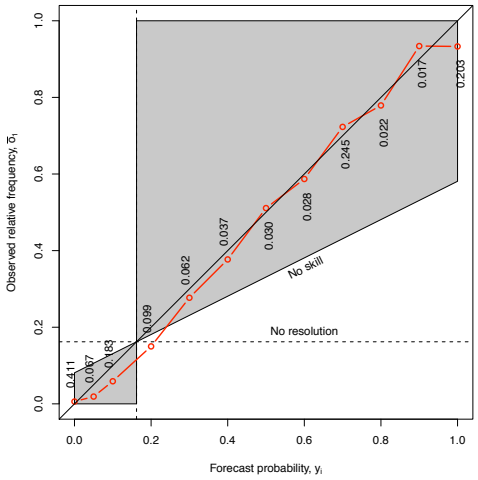
- As individual members
- As probability forecasts
  - Brier Score
  - Reliability diagram
  - ROC diagram
  - Economic Cost Loss Value
- As a full distribution
  - Rank Histograms
  - Spread / Skill comparison
  - Ranked Probability Score / Continuous RPS
  - Ignorance (log p) score

Verifying a probabilistic forecast

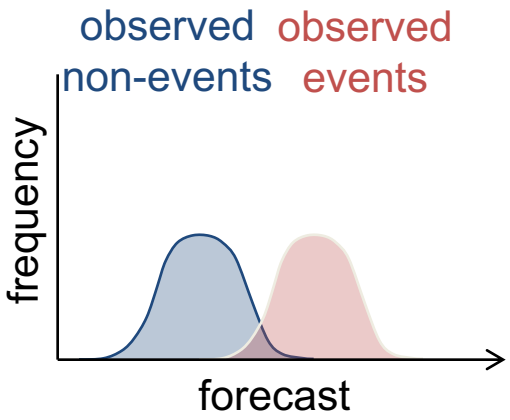
- You cannot verify a probabilistic forecast with a single observation.
- The more data you have for verification, (as with other statistics) the more certain you are.
  - *Evaluation of probability forecasts generally requires larger sample sizes than other types of forecasts*
- Rare events (low probability) require more data to evaluate.
- These comments refer to probabilistic forecasts developed by methods other than ensembles as well.

Properties of a perfect probabilistic forecast of a binary event.

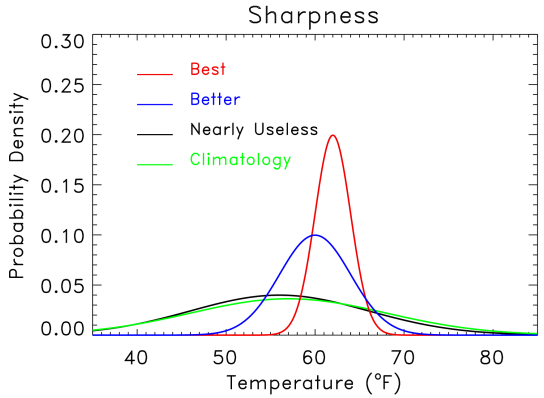
Reliability



Resolution



Sharpness



The Brier Score

- Mean square error of a probability forecast

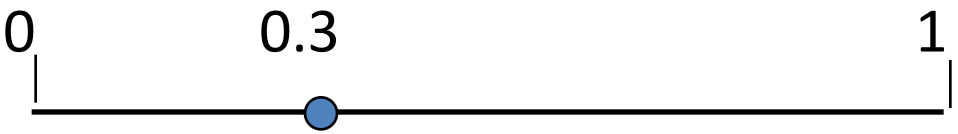
$$BS = \frac{1}{n}\sum_{i=1}^{n}(f_i - x_i)^2$$

where n is the number of forecasts

f_i is the forecast prob on occasion i

x_i is the observation (0 or 1) on occasion i

- Weights larger errors more than smaller ones



Components of the Brier Score

- **Reliability**

  Measures how well the conditional relative frequency of events matches the forecast

$$\frac{1}{n}\sum_{i=1}^{I} N_i(f_i - \bar{x}_i)^2$$

- **Resolution**

  Measures how well the forecasts distinguish situations with different frequencies of occurrence

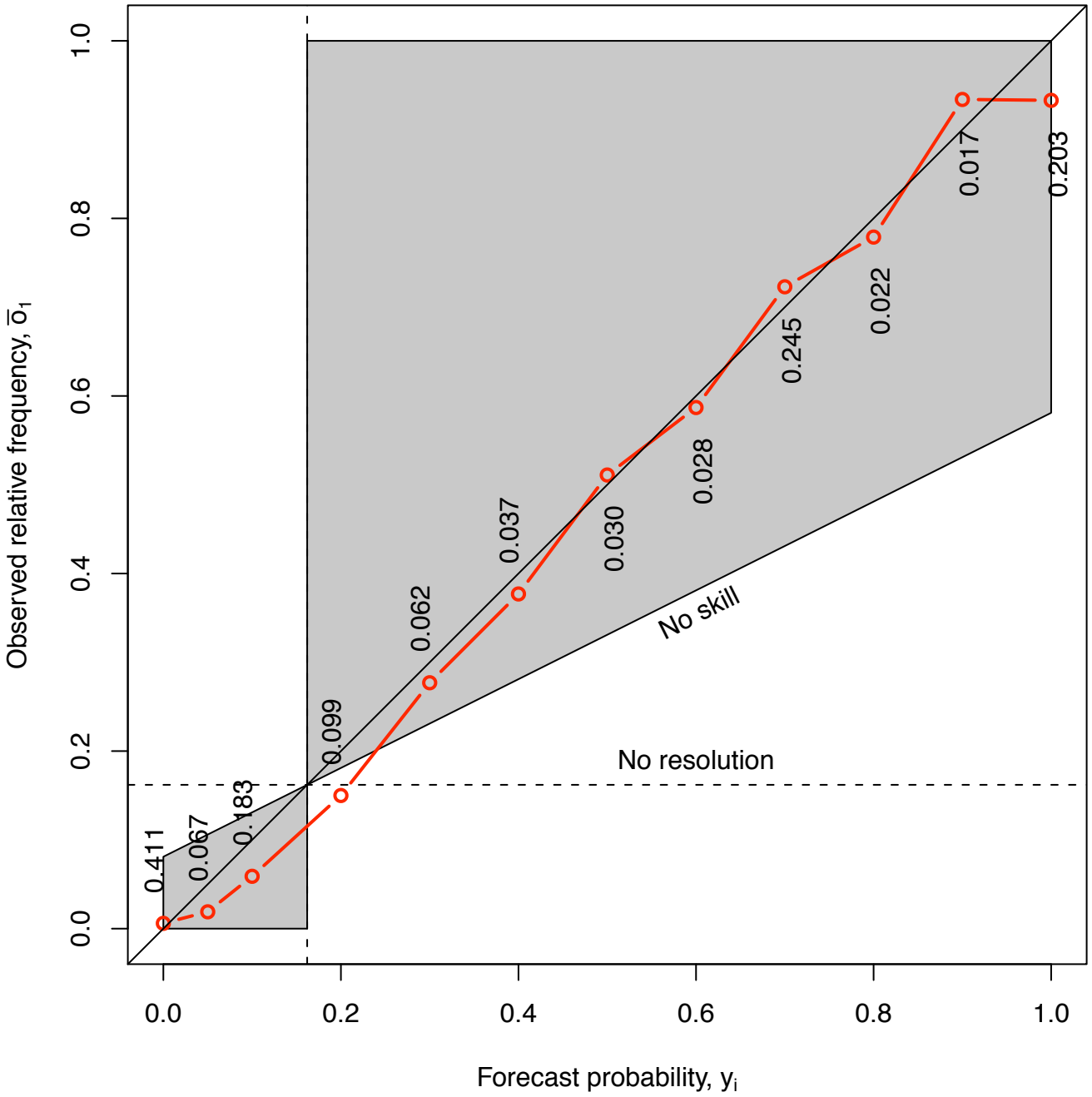
$$\frac{1}{n}\sum_{i=1}^{I} N_i(\bar{x}_i - \bar{x})^2$$

- **Uncertainty**

  Measures the variability in the observations (i.e., the difficulty of the forecast situations)

$$\bar{x}(1-\bar{x})$$

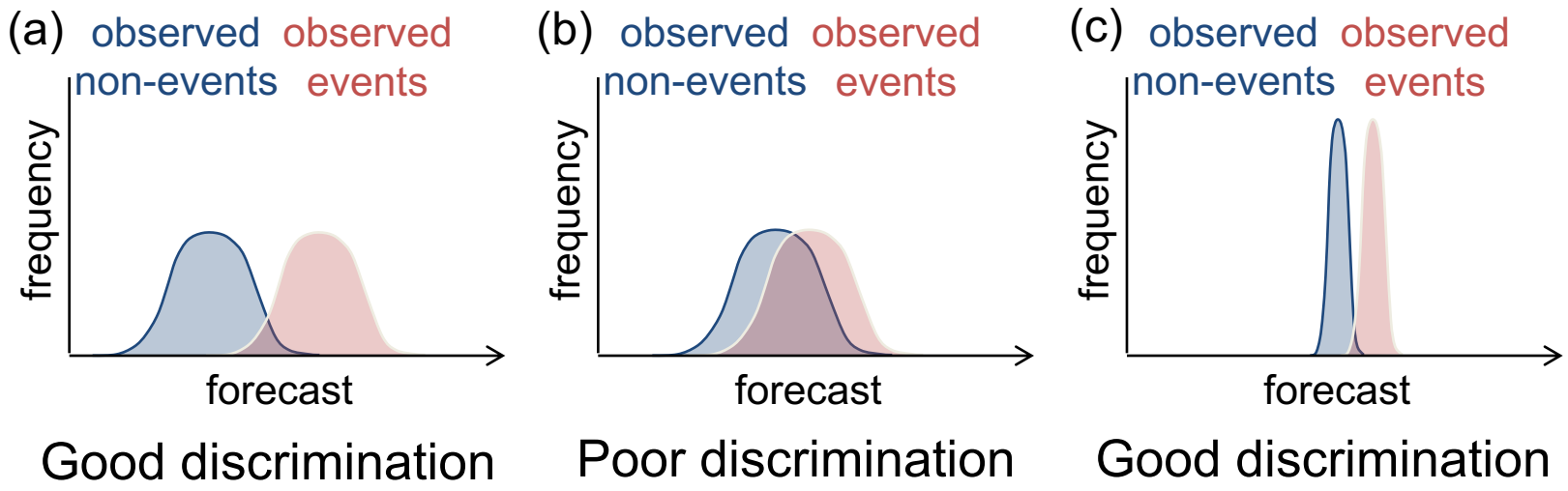
Looking at Brier Score *components* is critical to understand forecast performance



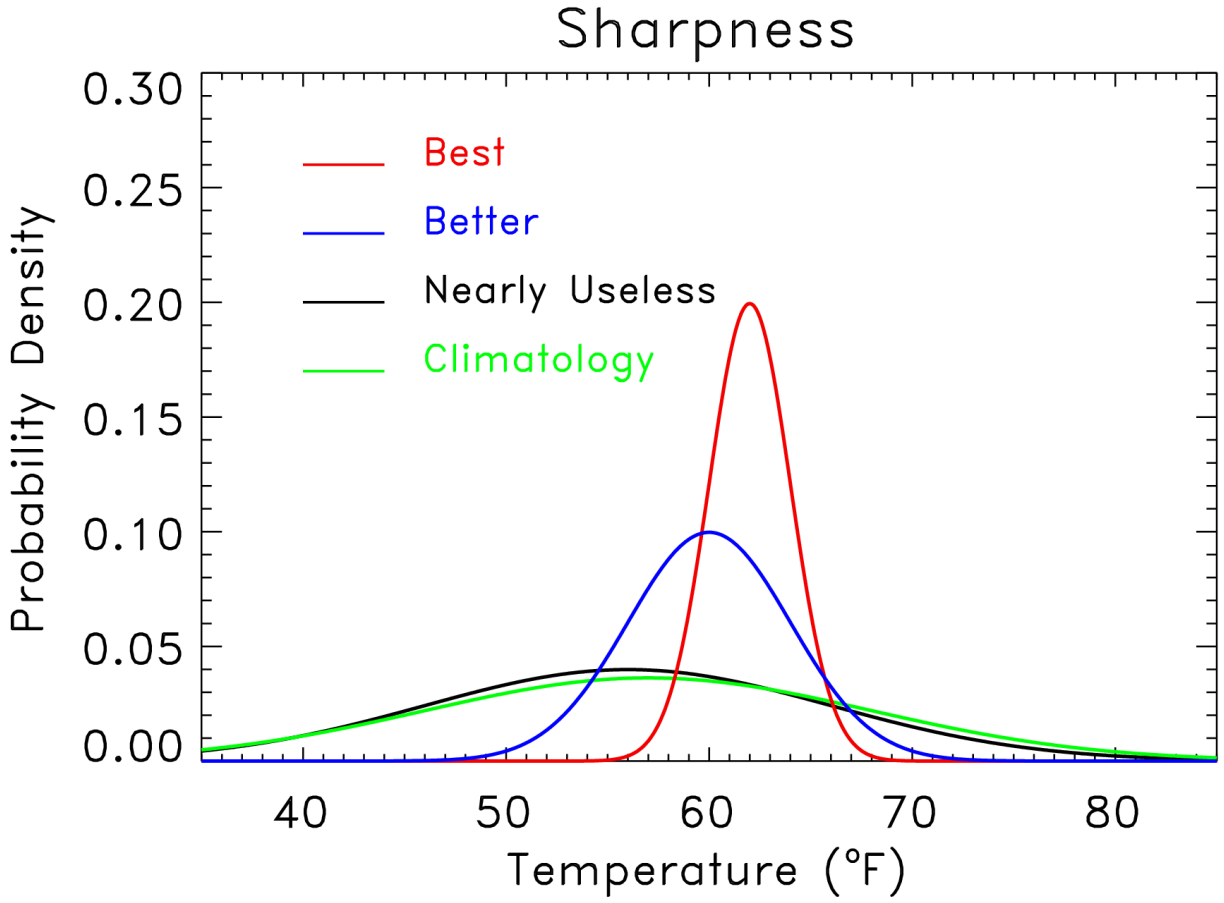
Attribute (Reliability) diagram shows reliability, resolution, skill

Discrimination

- *Discrimination*: The ability of the forecast system to clearly distinguish situations leading to the occurrence of an event of interest from those leading to the non-occurrence of the event.
- Depends on:
  - Separation of means of conditional distributions
  - Variance within conditional distributions



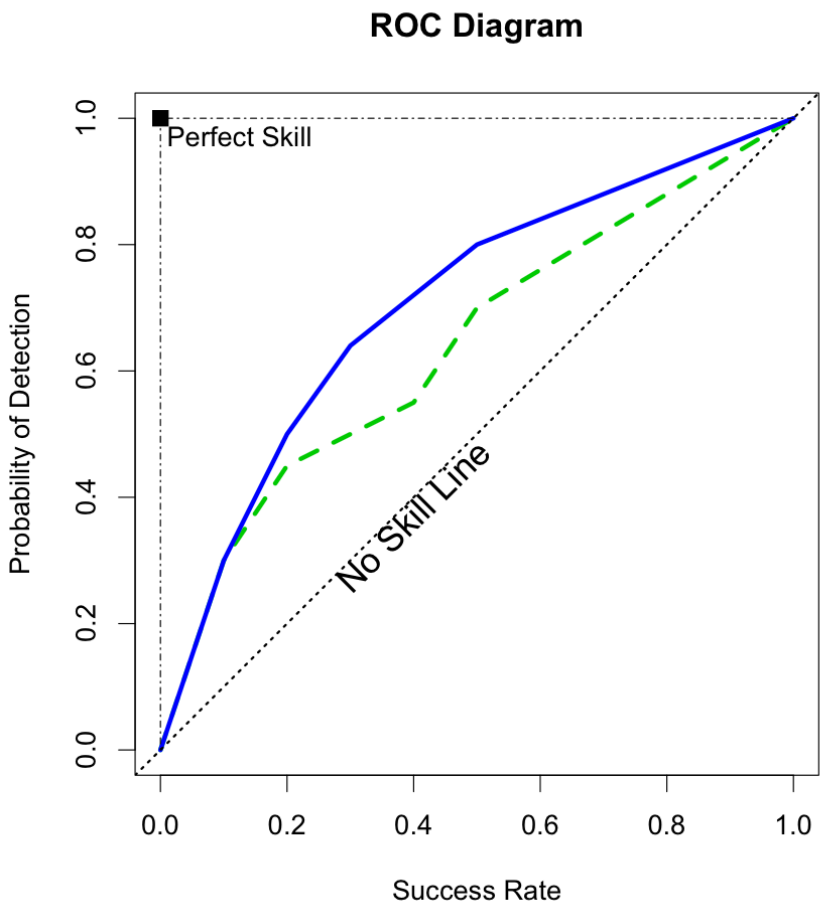
Sharpness also important



"**Sharpness**" measures the specificity of the probabilistic forecast. Given two reliable forecast systems, the one producing the sharper forecasts is preferable.

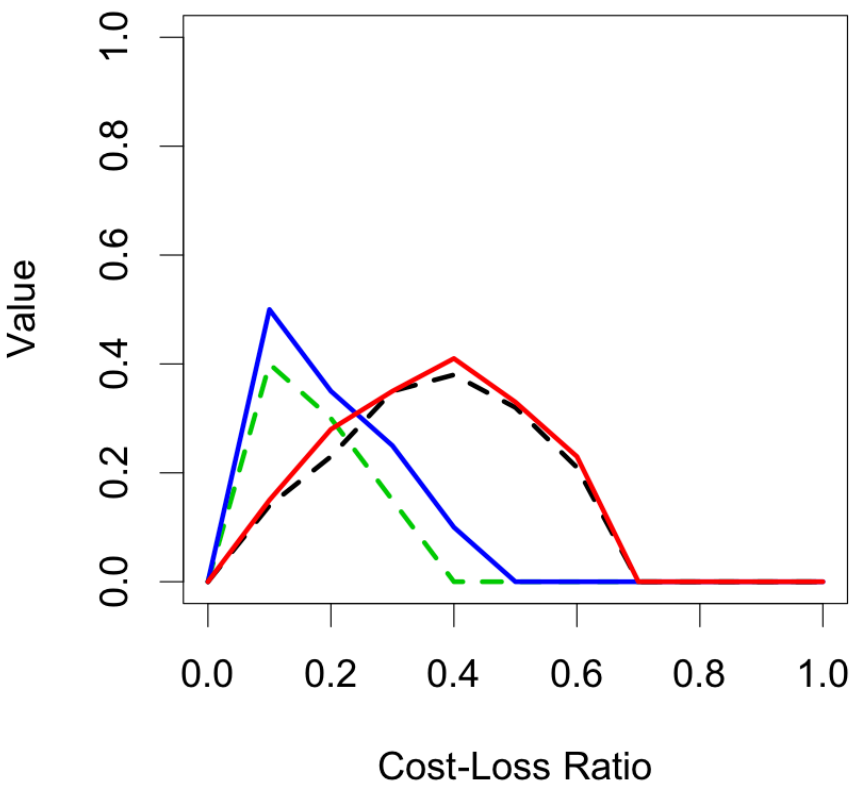
But: don't want sharp if not reliable. Implies unrealistic confidence.

Receiver Operating Characteristic



- Plot POD vs. (1-FAR) for all probability thresholds.
- Upper left corner is perfect.
- Facilitates comparisons even when probabilities are uncalibrated (e.g. conditionally biased).
- Area under ROC also used as a metric.

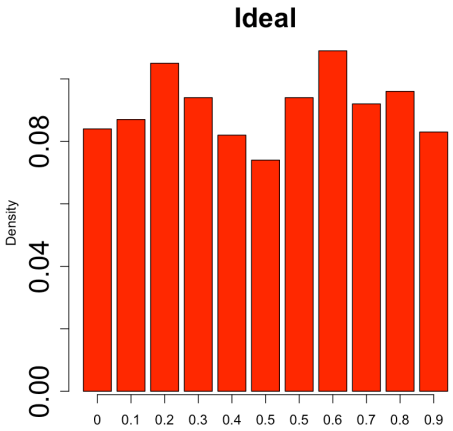
Economic Cost Loss Value Plot

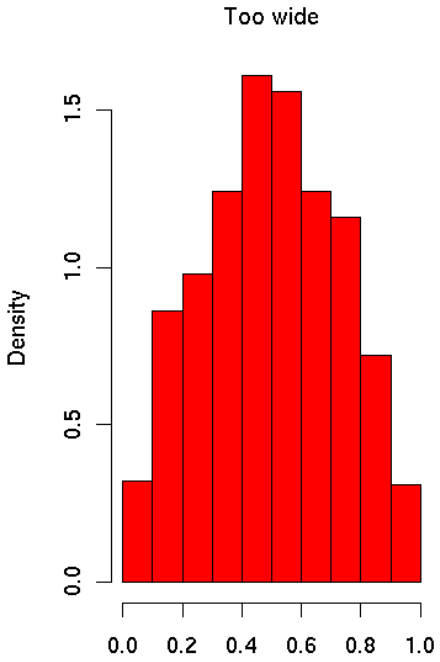


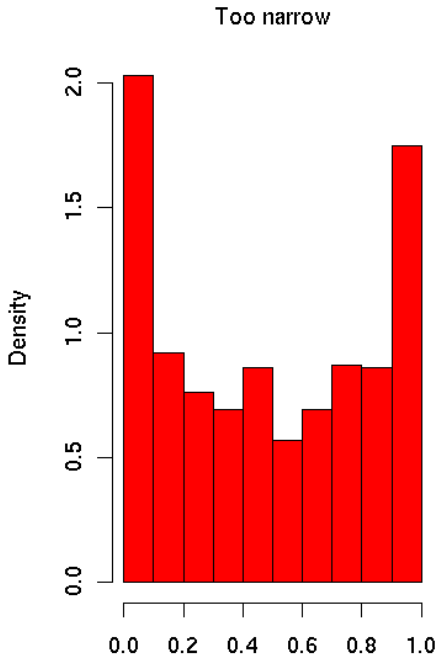
The envelope of relative value curves on the ECLV plot represents the *potential* value since all decision thresholds are possible for probability forecasts.

	Event	**No Event**
Action	Cost	Cost
No Action	Loss	0

Evaluating ensembles – Rank Histograms





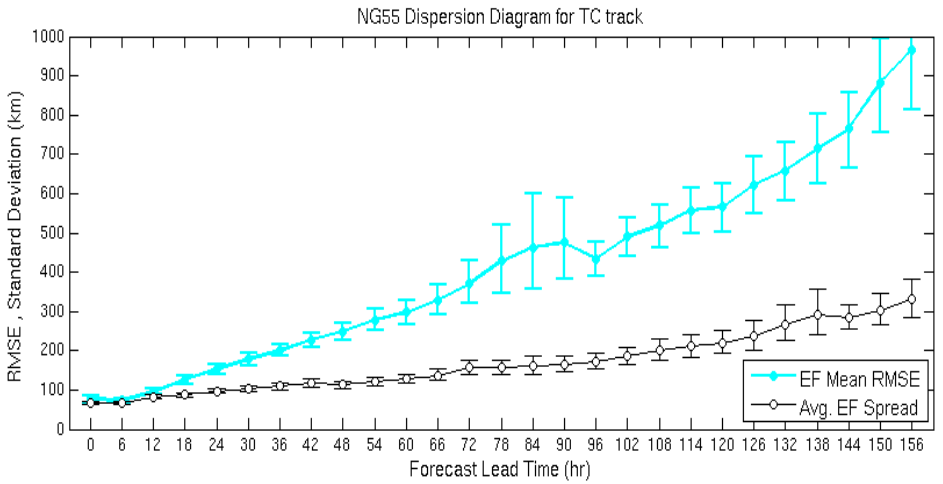


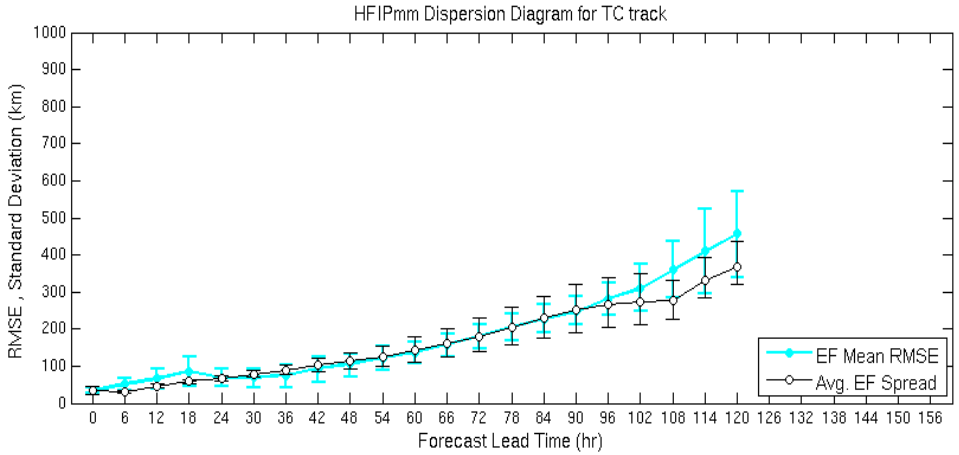
- For a perfect ensemble, the observation comes from the same distribution as the ensemble.
- Thus, ranking the observation among the ensemble member should result in a flat (uniform) distribution.

Evaluating ensembles – Spread Skill diagram

- For a perfect ensemble, the spread of the ensemble represents the error in the forecast.
- We can compare spread (std. dev.) with skill (RMSE) to see if they are close.

Spread-skill





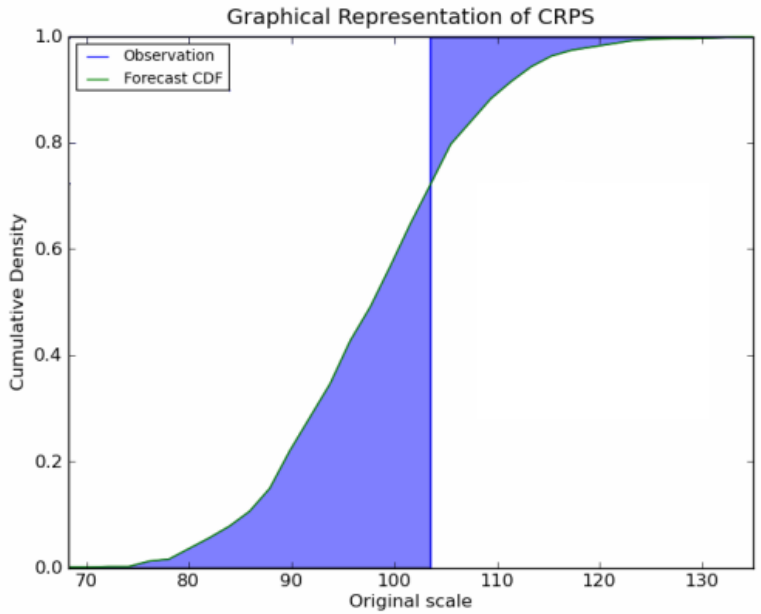
Continuous and discrete rank probability scores

- Measures of accuracy for
  - Multiple category forecasts (e.g., precipitation type)

    Rank Probability Score (RPS)
  - Continuous distributions (e.g., ensemble distribution)

    Continuous Ranked Probability Score (CRPS)

Brier Score, RPS, and CRPS are all MSE-type scores, but with different formats of input data.



Evaluating ensembles – Ignorance (log p) score

The information in a forecast = the reduction of uncertainty due to the forecasts.

$$\text{IGN} = -\log p(a_j)$$

where $p(a_j)$ is the probability of event a_j.

In MET, these are averaged.

Measure	Attribute evaluated	Comments
Probability forecasts		
Brier score	Accuracy	Based on squared error
Resolution	Resolution (resolving different categories)	Compares forecast category climatologies to overall climatology
Reliability	Calibration	
Skill score	Skill	Skill involves *comparison* of forecasts
Sharpness measure	Sharpness	Only considers distribution of forecasts
Relative Operating Characteristic (ROC)	Discrimination	Ignores calibration
C/L Value	Value	Ignores calibration
Ensemble distribution		
Rank histogram	Calibration	Can be misleading
Spread-skill	Calibration	Difficult to achieve
CRPS	Accuracy	Squared difference between forecast and observed distributions Analogous to MAE in limit
log p score (IGN)	Accuracy	Local score, rewards for correct category; infinite if observed category has 0 density

Useful references

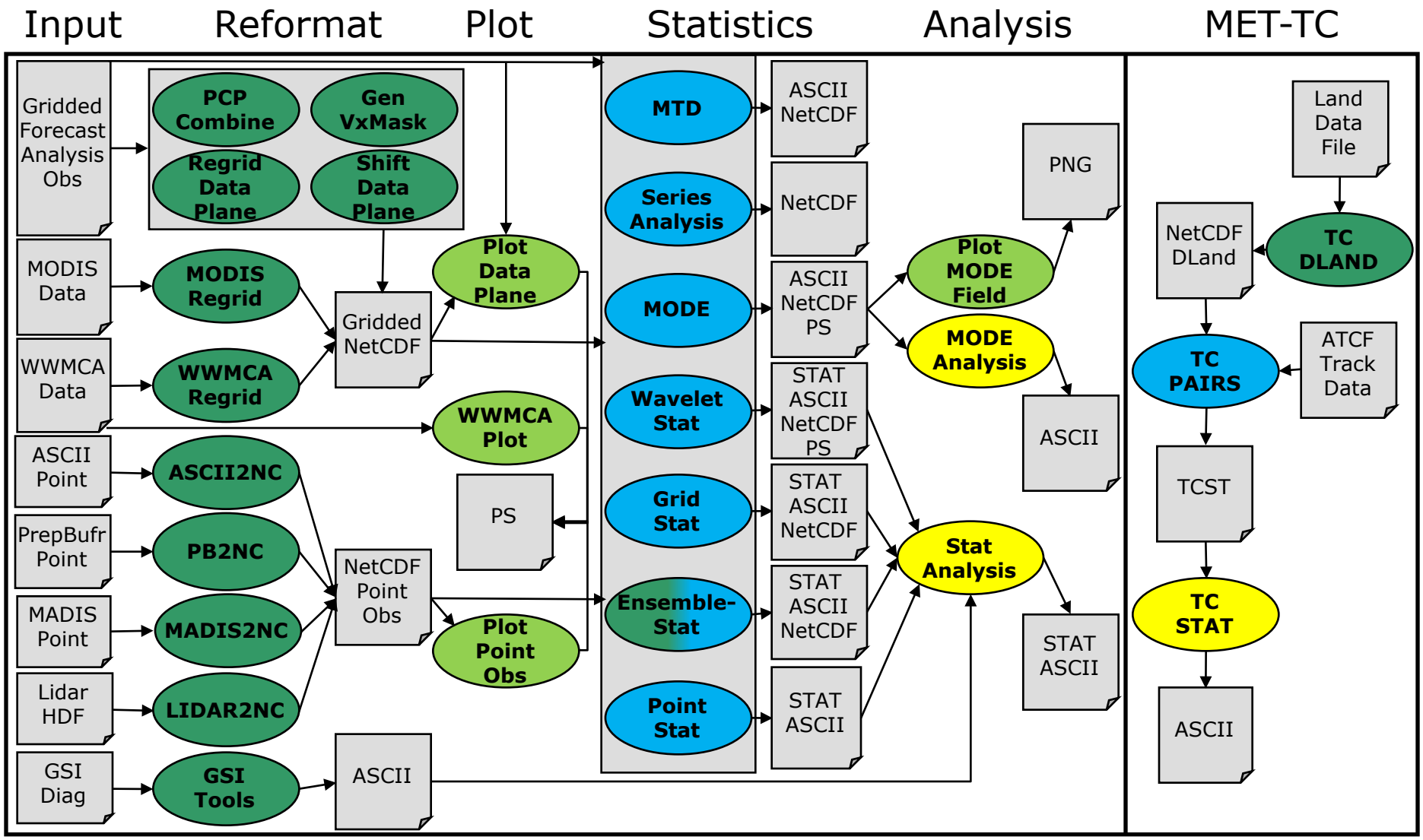
- **Good overall references** for forecast verification:
  - (1) Wilks, D.S., 2011: *Statistical Methods in the Atmospheric Sciences (3rd Ed).* Elsevier, 704 pp.
  - (2) WMO Verification working group forecast verification web page, http://www.cawcr.gov.au/projects/verification/
  - (3) Jolliffe and Stephenson Book: Jolliffe, I.T., and D.B. Stephenson, 2012: *Forecast Verification. A Practitioner's Guide in Atmospheric Science*., 2nd Edition, Wiley and Sons Ltd.
- **Verification tutorial – Eumetcal** (http://www.eumetcal.org/-learning-modules-)
- **Rank histograms**: Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev*., **129**, 550-560.
- **Spread-skill relationships**: Whitaker, J.S., and A. F. Loughe, 1998: The relationship between ensemble spread and ensemble mean skill. *Mon. Wea. Rev*., **126**, 3292-3302.
- **Brier score, continuous ranked probability score, reliability diagrams**: Wilks text again.
- **Relative operating characteristic**: Harvey, L. O., Jr, and others, 1992: The application of signal detection theory to weather forecasting behavior. *Mon. Wea. Rev*., **120**, 863-883.
- **Economic value diagrams**:
  - (1)Richardson, D. S., 2000: Skill and relative economic value of the ECMWF ensemble prediction system. *Quart. J. Royal Meteor. Soc*., **126**, 649-667.
  - (2) Zhu, Y, and others, 2002: The economic value of ensemble-based weather forecasts. *Bull. Amer. Meteor. Soc*., **83**, 73-83.
- **Overestimating skill**: Hamill, T. M., and J. Juras, 2006: Measuring forecast skill: is it real skill or is it the varying climatology? *Quart. J. Royal Meteor. Soc.*, Jan 2007 issue. http://tinyurl.com/kxtct

Verifying Ensembles & Probability Forecasts with MET

- Ensemble-Stat Tool
  - Ensemble Mean Fields
  - Probability Fields
  - Ensemble Continuous Statistics
  - Rank Histograms, PIT Histograms, Relative Position
  - Spread-Skill Calculation

- Point-Stat and Grid-Stat Tool
  - Brier Score + Decomposition
  - Reliability Diagrams
  - Receiver Operating Characteristic Diagram + Area Under the Curve
  - Joint/Conditional factorization table

MET Overview v7.0



Ensemble-Stat Capabilities

Reads:

- Gridded ensemble member files
- ***Gridded AND point*** observations files

Calculates:

- Ensemble Mean, Standard Deviations, Mean $\pm$ 1 SD fields
- Ensemble Min, Max, and Range fields
- Ensemble Valid Data Count field
- Ensemble Relative Frequency by threshold fields (i.e. probability)
- Ensemble Continuous Statistics (in met-7.1)
- Rank, Relative Position, and PIT Histograms (if Obs Field Provided)
- Ensemble Spread-Skill (if Obs Field Provided)
- Application of observation error (in met-7.1)

Writes:

- Ensemble products in a NetCDF file
- Stat file with Continuous Ensemble Statistics, Rank Histograms, PIT Histograms, RELP Histograms, Spread-Skill partial sums, and Point Observation Ranks
- Gridded field of Observation Ranks in a NetCDF file

Ensemble Stat Tool: Usage

Usage: ensemble_stat

n_ens ens_file_1 \ ... ens_file_n | ens_file_list	Number of Ensemble members followed by list of ensemble member names OR ens_file_list (the name of an ASCII file with names of members)
config_file	Config file name
[-grid_obs file]	Name of gridded or point observed file – Required if Rank Histograms desired (optional)
[-point_obs file]	
[-ssvar_mean file]	Specify an ensemble mean model data file for use in calculating ensemble spread-skill (optional)
[-obs_valid_beg time]	YYYYMMDD[_HH[MMSS]] format to set the beginning and end of the matching observation time window (optional)
[-obs_valid_end time]	
[-outdir path]	Set output directory (optional)
[-log file]	Outputs log messages to the specified file (optional)
[-v level]	Set level of verbosity (optional)

Ensemble-Stat: Configuration

- Many configurable parameters – only set a few:
  - ens_thresh - All members must be available; vld_thresh – all data in grid must be valid
  - 24hr Accumulated Precip (APCP)
  - Composite Reflectivity (REFC)
  - N-S component of Wind (UGRD)
  - Thresholds used for Ensemble Relative Freq (i.e. probability)
  - GRIB1_ptv = 129; Use GRIB Table 129 instead of Table 2

```
//
// Ensemble product fields to be processed
// (i.e. mean, min, max, stdev fields)
//
ens = {
   ens_thresh = 1.0;
   vld_thresh = 1.0;

   field = [
      {
         name       = "APCP";
         level      = [ "A24" ];
         cat_thresh = [ >0.0, >=10.0 ];
      },
      {
         name       = "REFC";
         level      = [ "L0" ];
         cat_thresh = [ >=35.0 ];
         GRIB1_ptv  = 129;
      },
      {
         name       = "UGRD";
         level      = [ "Z10" ];
         cat_thresh = [ >=5.0 ];
      },
   ];
}
```

Ensemble-Stat: Configuration

- Many configurable parameters – only set a few:
  - Use ADPSFC message type for point obs
  - Use 24hr precip for gridded obs field
  - Bin size for spread-skill calcuation is 0.1 mm
  - Output rank histogram, obs rank, and ssvar calculations
  - Output all ensemble products like mean, stdev, etc…

```
//
// Forecast and observation fields to be
// verified (i.e. RHIST, PHIST, SSVAR)
//
fcst = {

   message_type = [ "ADPSFC" ];

   field = [
      {
         name                = "APCP";
         level               = [ "A24" ];
         ens_ssvar_bin_size = 0.1;
      }
   ];
}
obs = fcst;
```

```
output_flag = {
   ecnt  = BOTH;
   rhist = BOTH;
   phist = BOTH;
   orank = BOTH;
   ssvar = BOTH;
   relp  = BOTH;
};
```

```
ensemble_flag = {
   mean       = TRUE;
   stdev      = TRUE;
   minus      = TRUE;
   plus       = TRUE;
   min        = TRUE;
   max        = TRUE;
   range      = TRUE;
   vld_count  = TRUE;
   frequency  = TRUE;
   rank       = TRUE;
   weight     = FALSE;
};
```

Ensemble-Stat Tool: Run

```
ensemble_stat \
  6 sample_fcst/2009123112/*gep*/d01_2009123112_02400.grib \
  config/EnsembleStatConfig \
  -grid_obs sample_obs/ST4/ST4.2010010112.24h \
  -point_obs out/ascii2nc/precip24_2010010112.nc \
  -outdir out/ensemble_stat -v 2
```

NOTE:
You can pass in a path with wildcards to pull out the files you would like to process or you can pass in an ASCII filename that contains a list of ensemble members

Gridded and Obs fields are included for use in calculating Rank Histogram, PIT Histogram, and Spread-Skill

Ensemble-Stat Tool: Run

```
*** Running Ensemble-Stat on APCP using GRIB forecasts, point observations, and gridded observations ***
DEBUG 1: Default Config File: /d3/projects/MET/MET_releases/met-6.0/data/config/EnsembleStatConfig_default
DEBUG 1: User Config File: config/EnsembleStatConfig
GSL_RNG_TYPE=mt19937
GSL_RNG_SEED=1
DEBUG 1: Ensemble Files[6]:
DEBUG 1:    ../data/sample_fcst/2009123112/arw-fer-gep1/d01_2009123112_02400.grib
DEBUG 1:    ../data/sample_fcst/2009123112/arw-fer-gep5/d01_2009123112_02400.grib
DEBUG 1:    ../data/sample_fcst/2009123112/arw-sch-gep2/d01_2009123112_02400.grib
DEBUG 1:    ../data/sample_fcst/2009123112/arw-sch-gep6/d01_2009123112_02400.grib
DEBUG 1:    ../data/sample_fcst/2009123112/arw-tom-gep3/d01_2009123112_02400.grib
DEBUG 1:    ../data/sample_fcst/2009123112/arw-tom-gep7/d01_2009123112_02400.grib
DEBUG 1: Gridded Observation Files[1]:
DEBUG 1:    ../data/sample_obs/ST4/ST4.2010010112.24h
DEBUG 1: Point Observation Files[1]:
DEBUG 1:    ../out/ascii2nc/precip24_2010010112.nc
DEBUG 2:
DEBUG 2: ----------------------------------------------------------------------------
DEBUG 2:
DEBUG 2: Processing ensemble field: APCP/A24
DEBUG 2:
DEBUG 2: ----------------------------------------------------------------------------
...
Processing gridded verification APCP_24/A24 versus APCP_24/A24, for observation type MC_PCP, over region FULL, for interpolation method
UW_MEAN(1), using 15480 pairs.
---------------------------------------------------------------------------
DEBUG 1: Output file: out/ensemble_stat/ensemble_stat_20100101_120000V.stat
DEBUG 1: Output file: out/ensemble_stat/ensemble_stat_20100101_120000V_rhist.txt
DEBUG 1: Output file: out/ensemble_stat/ensemble_stat_20100101_120000V_phist.txt
DEBUG 1: Output file: out/ensemble_stat/ensemble_stat_20100101_120000V_orank.txt
DEBUG 1: Output file: out/ensemble_stat/ensemble_stat_20100101_120000V_ssvar.txt
DEBUG 1: Output file: out/ensemble_stat/ensemble_stat_20100101_120000V_ens.nc
DEBUG 1: Output file: out/ensemble_stat/ensemble_stat_20100101_120000V_orank.nc
```

Ensemble-Stat Tool: Output

Output from out/ensemble_stat/ensemble_stat_20100101_120000V_rhist.txt

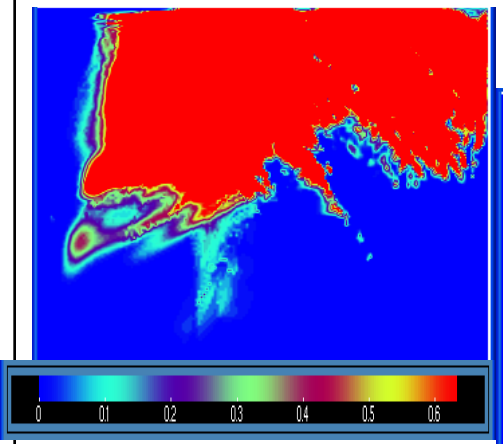
VERSION MODEL FCST_LEAD FCST_VALID_BEG FCST_VALID_END OBS_LEAD OBS_VALID_BEG OBS_VALID_END FCST_VAR FCST_LEV OBS_VAR OBS_LEV OBTYPE VX_MASK INTERP_MTHD INTERP_PNTS FCST_THRESH OBS_THRESH COV_THRESH ALPHA LINE_TYPE TOTAL CRPS IGN N_RANK RANK_1 RANK_2 RANK_3 RANK_4 RANK_5 RANK_6 RANK_7

V6.0 WRF 240000 20100101_120000 20100101_120000 000000 20100101_103000 20100101_133000 APCP_24 A24 APCP_24 A24 ADPSFC FULL UW_MEAN 1 NA NA NA NA

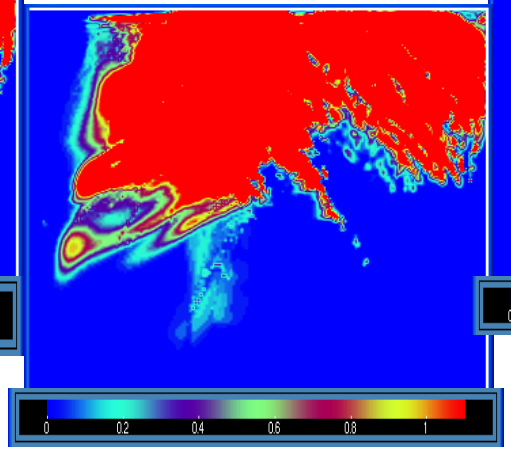
RHIST 1125 8.21904 6.53721 7 261 160 138 141 149 111 165

CRPS IGN RANK HIST

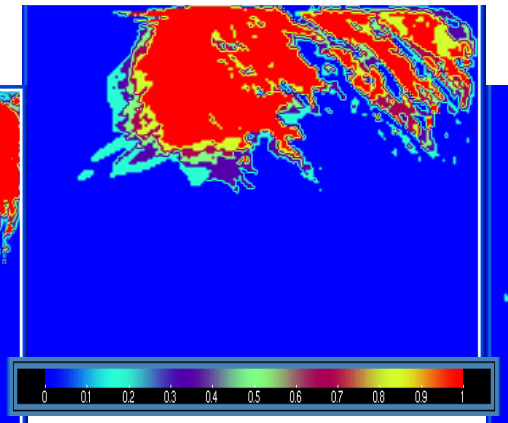
Output from out/ensemble_stat/ensemble_stat_20100101_120000V_ens.nc (using ncview)



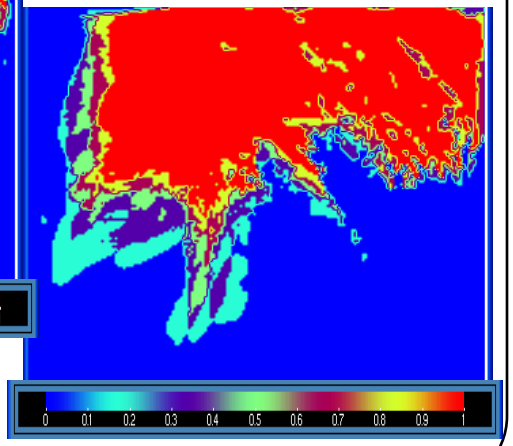
Ensemble Mean



Ensemble StdDev

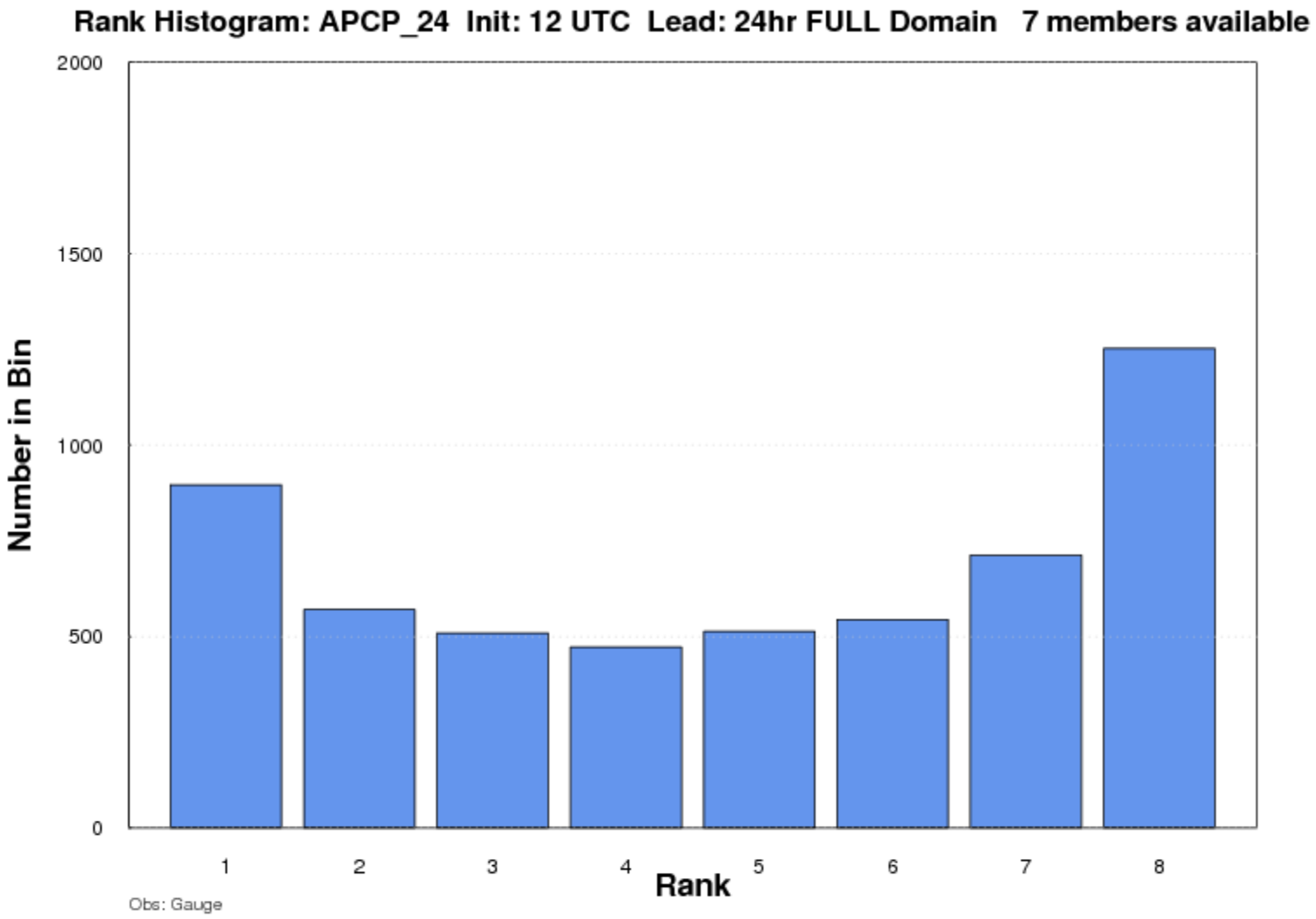


Prob > 5 mm



Prob > 0 mm

Rank Histogram



RELP Line Type

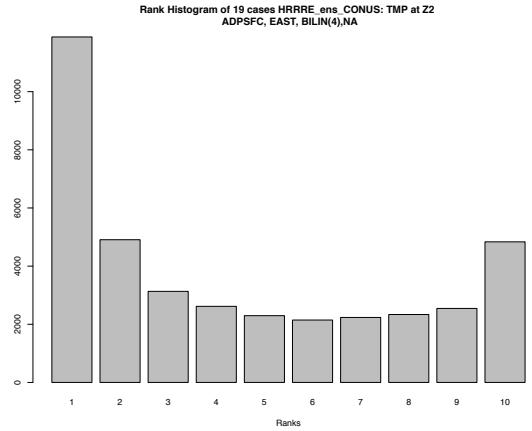
- Added support to Ensemble-Stat for new relative position (**RELP**) line type.
  - Same as VSDB RELP line type.
  - Similar to the Rank Histogram line type, but no ranking is done.
  - N-th histogram bar indicates how often ensemble member N was closest (in absolute value) to the observation.
  - Ties are assigned equally to all tied members.

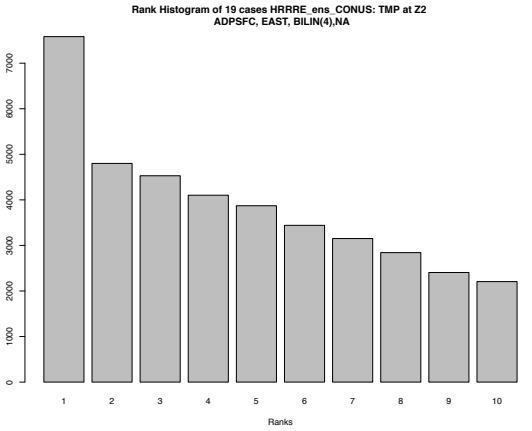
Ensemble-Stat Tool: Updates

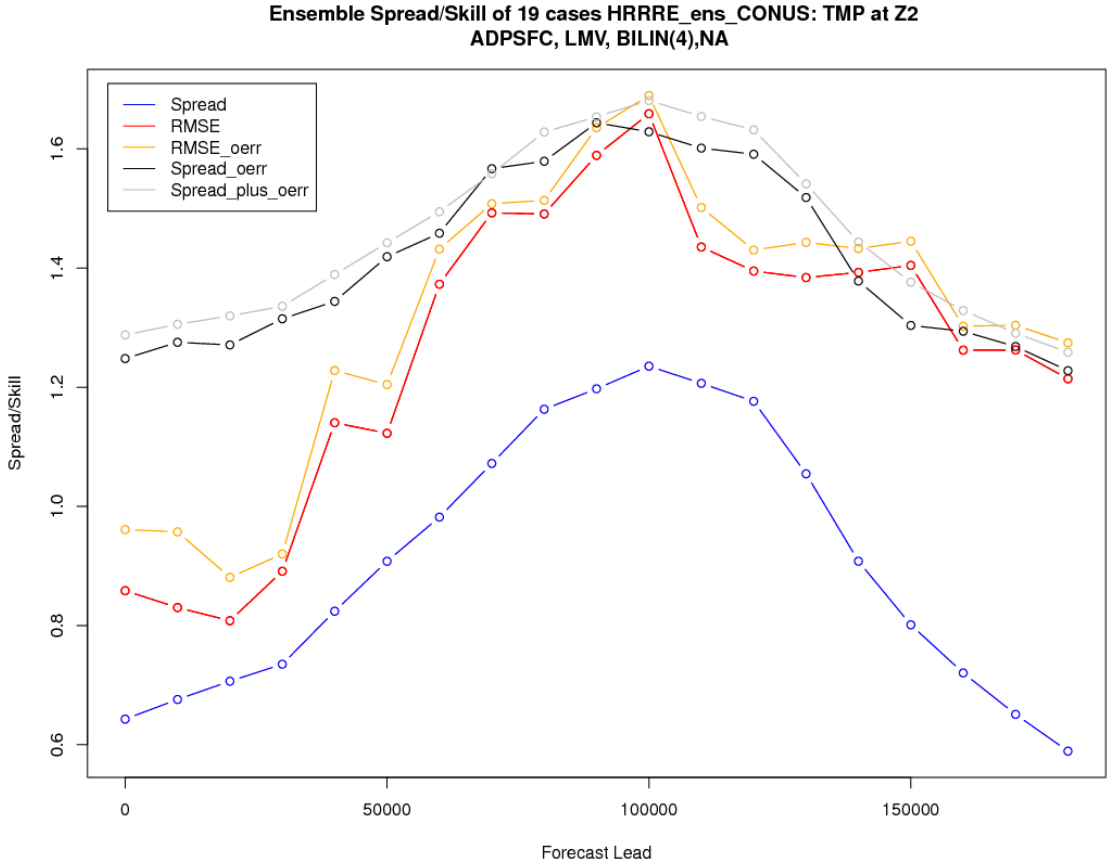
- **Current met-7.0 release:**
  - Config file option to exclude ties (i.e. all ensemble members and observation are identical) … skip_const = TRUE;
  - Config file option to filter pairs by the observation values … obs_thresh = [>0]; // e.g. only observations of APCP > 0
  - Added BSS_SMPL statistic to the PSTD line type, making MET consistent with METViewer.
- **Upcoming met-7.1 release:**
  - Read gridded data from user-supplied Python script.
  - Read GOES-16 AOD datasets.
  - Add new continuous ensemble statistics (ECNT) line type.
  - Apply user-configurable observation error in ensemble verification.

Ensemble-Stat: Observation Error

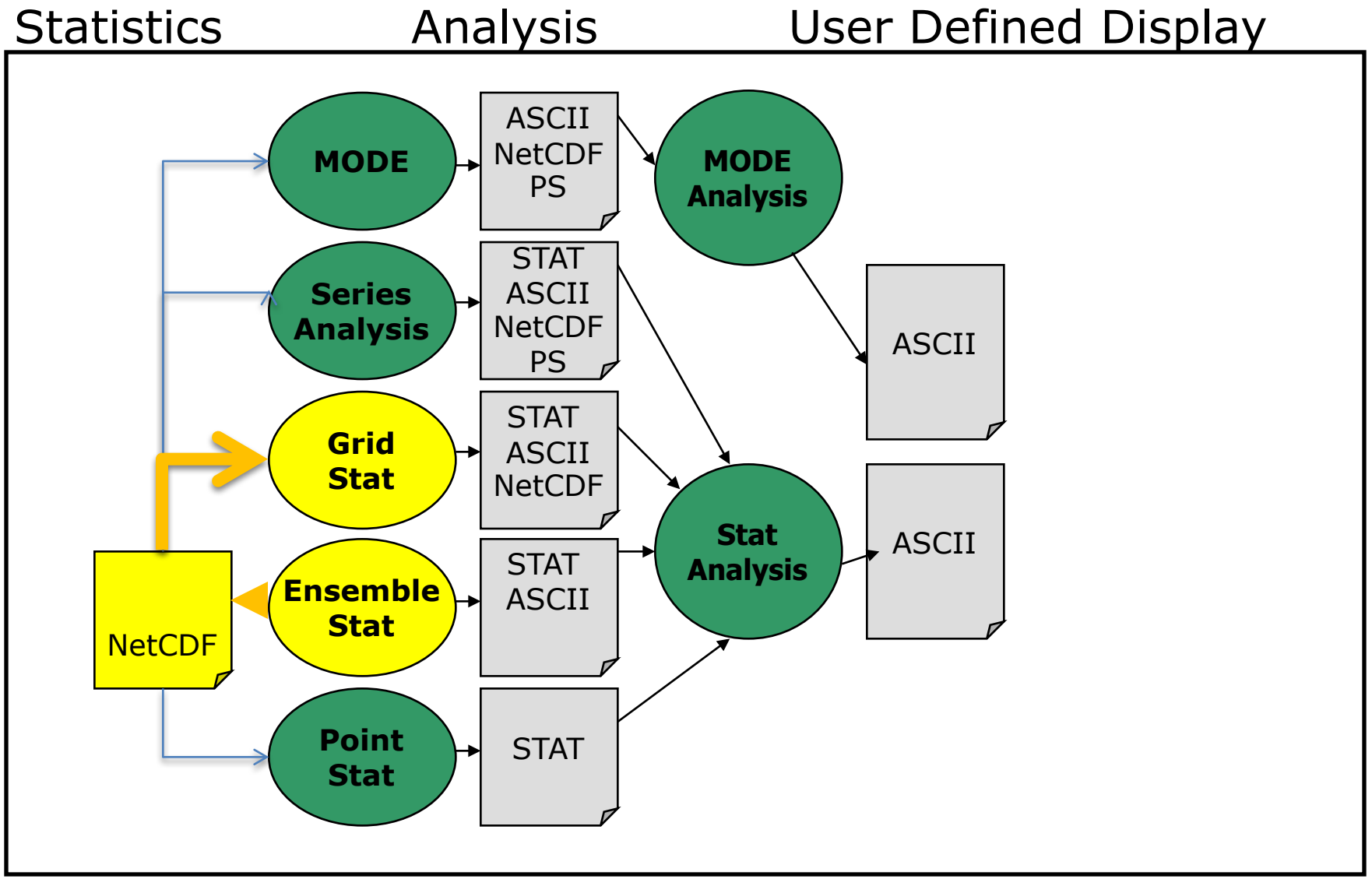
- User has fine control of observation error assumptions.
- Observation error perturbations are applied to the ***ensemble member values*** … Counter-intuitive but correct







Probability fields



Verifying Probabilities

- Probabilistic verification method tools:
  - Grid-Stat, Point-Stat, and Stat-Analysis
- Define Nx2 contingency table using:
  - Multiple forecast probability thresholds
  - One observation threshold

Forecast	Observation		Total
	o = 1 (e.g., "Yes")	o = 0 (e.g., "No")	
p_1 = midpoint of (0 and threshold1)	n_{11}	n_{10}	$n_{1\cdot} = n_{11} + n_{10}$
p_2 = midpoint of (threshold1 and threshold2)	n_{21}	n_{20}	$n_{2\cdot} = n_{21} + n_{20}$
⋮	⋮	⋮	⋮
p_j = midpoint of (threshold$i$ and 1)	n_{i1}	n_{i0}	$n_j = n_{j1} + n_{j0}$
Total	$n_{\cdot 1} = \Sigma n_{i1}$	$n_{\cdot 0} = \Sigma n_{i0}$	$T = \Sigma n_i$

Example:

- FCST: Probability of precip

[0.00, 0.25, 0.50, 0.75, 1.00]

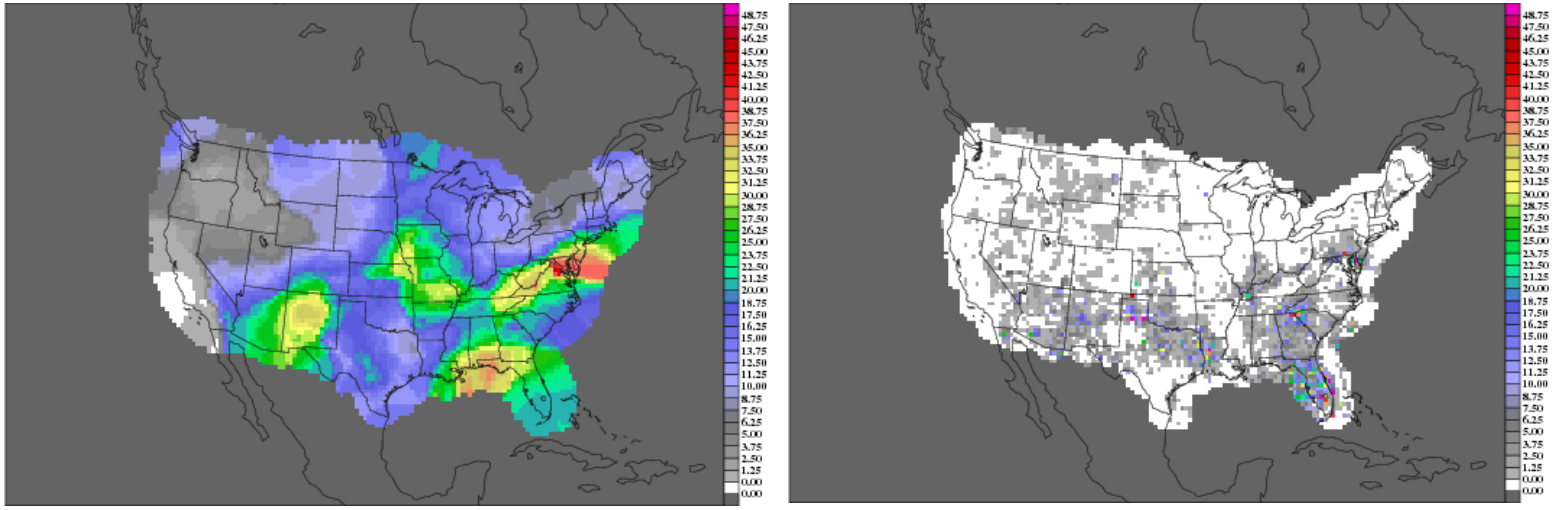
==0.25

- OBS: Accumulated precip

> 0.00

Verifying Probabilities: Example

- Verify probability of precip with total precip:



- Configuration file settings:

```
fcst = {
 field = [
  {
   name          = "POP";
   level         = [ "Z0" ];
   //cat_thresh  = [ >=0.0, >=0.25, >=0.50, >=0.75, >=1.00 ];
   cat_thresh    = [ ==0.25 ];
   prob          = TRUE;
  }
 ];
};
```

```
obs = {
 field = [
  {
   name        = "APCP";
   level       = [ "A12" ];
   cat_thresh  = [ >0.0 ];
  }
 ];
};
```

Grid-Stat: Probability Config

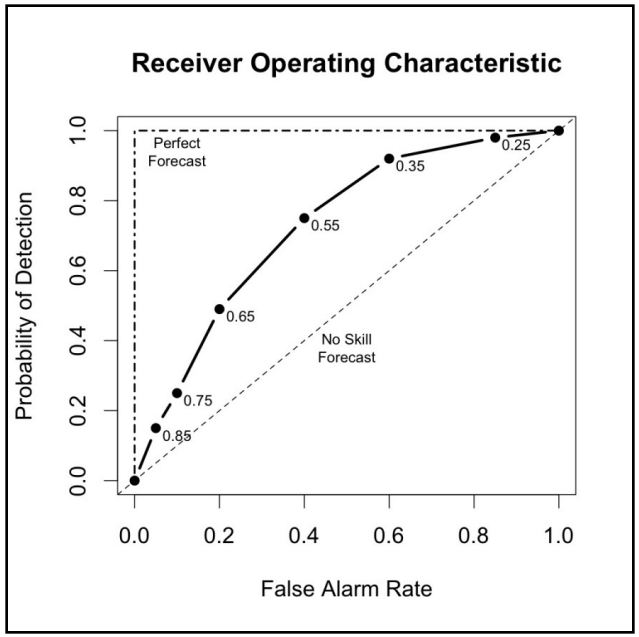
- Many configurable parameters – only set a few:
  - APCP_24… is name of ens mean in netcdf file
  - prob = True important
  - cat_thresh used for reliability and roc curves
  - Use 24hr Accumulation in grib file threshold at >10 mm
  - Generate probabilistic statistics

```
fcst = {
  wind_thresh = [ NA ];

  field = [
    {
      name       = "APCP_24_A24_ENS_FREQ_ge10.000";
      level      = [ "(*,*)" ];
      prob       = TRUE;
      cat_thresh = [ >=0.0, >=0.1, >=0.2, >=0.3,
  >=0.4, >=0.5, >=0.6, >=0.8, >=1.0 ];
      //cat_thresh = [ ==0.1 ];
      }
  ];

};
obs = {

  wind_thresh = [ NA ];

  field = [
    {
      name       = "APCP";
      level      = [ "A24" ];
      cat_thresh = [ >10.000 ];
    }
  ];

};
```

```
output_flag = {
   fho     = NONE;
   ctc     = NONE;
   cts     = NONE;
   mctc    = NONE;
   mcts    = NONE;
   cnt     = NONE;
   sl1l2   = NONE;
   vl1l2   = NONE;
   val1l2  = NONE;
   pct     = BOTH;
   pstd    = BOTH;
   pjc     = BOTH;
   prc     = BOTH;
   eclv    = BOTH;
   nbrctc  = NONE;
   nbrcts  = NONE;
   nbrcnt  = NONE;
   grad    = NONE;
};
```

Grid-Stat: Probability Run

- Output written to .stat file and, if desired, to individual text files:
  - PCT – Probability Contingency Table Counts
  - PSTD – Probability Contingency Table Scores
    - Brier Score, Reliability, Resolution, Uncertainty, Area Under ROC
  - PJC – Joint/Continuous Statistics of Probabilistic Variables
    - Calibration, Refinement, Likelihood, Base Rate, Reliability points
  - PRC – ROC Curve Points for Probabilistic Variables

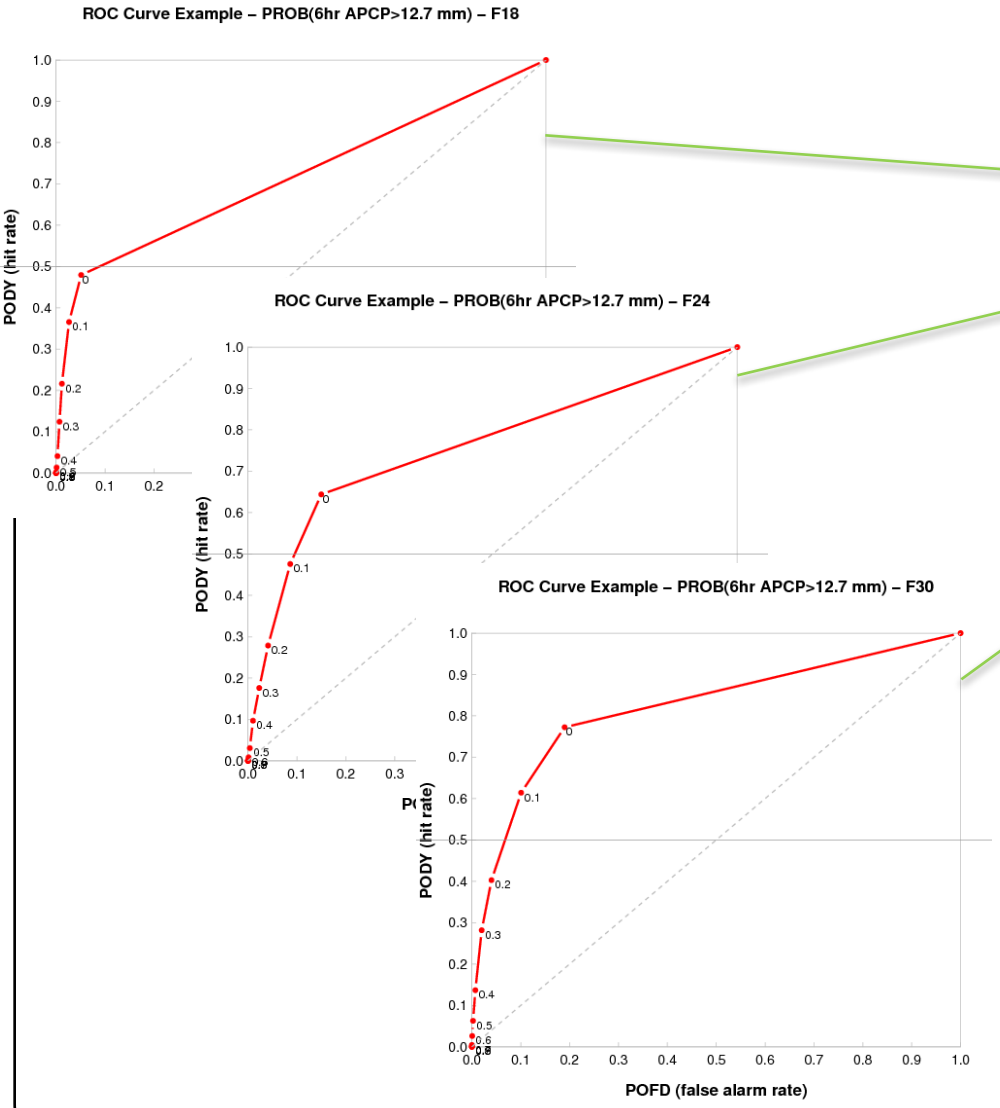


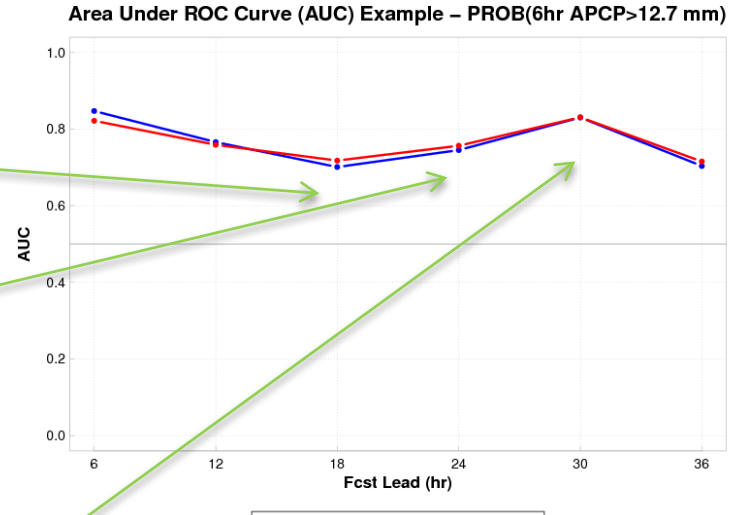
ECLV Line Type

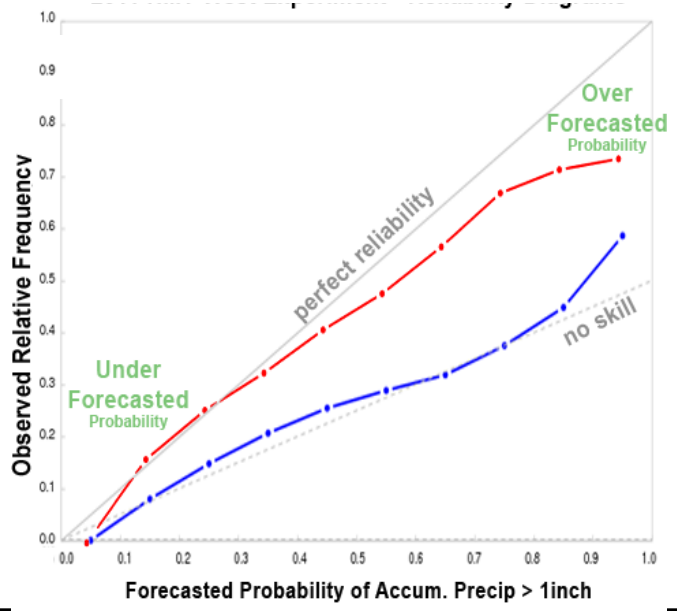
- met-6.1 added support to Grid-Stat and Point-Stat for computation of the Economic Cost-Loss Value (**ECLV**) line type.
- Equivalent to the VSDB ECON line type, except…
  - ECON is only generated when evaluating ensemble probabilities.
  - ECLV from 2x2 contingency table yield a single curve.
  - ECLV from Nx2 probabilistic contingency table yields N curves.
- Configuration file option to specify the cost/loss ratios to be evaluated:
  - eclv_points = 0.05; // equal spacing
  - eclv_points = [0.05, 0.10, 0.15, 0.20, 0.25, 0.50, 0.75]; // non-equal

```
//   - The "eclv_points" entry specifies the economic cost/loss ratio points
//   to be evaluated. For each cost/loss ratio specified, the relative value
//   will be computed and written to the ECLV output line. This entry may
//   either be specified as an array of numbers between 0 and 1 or as a single
//   number. For an array, each array entry will be evaluated. For a single
//   number, all evenly spaced points between 0 and 1 will be evaluated, where
//   eclv_points defines the spacing. Cost/loss values are omitted for
//   ratios of 0.0 and 1.0 since they are undefined.
```

Grid-Stat: Probability Examples

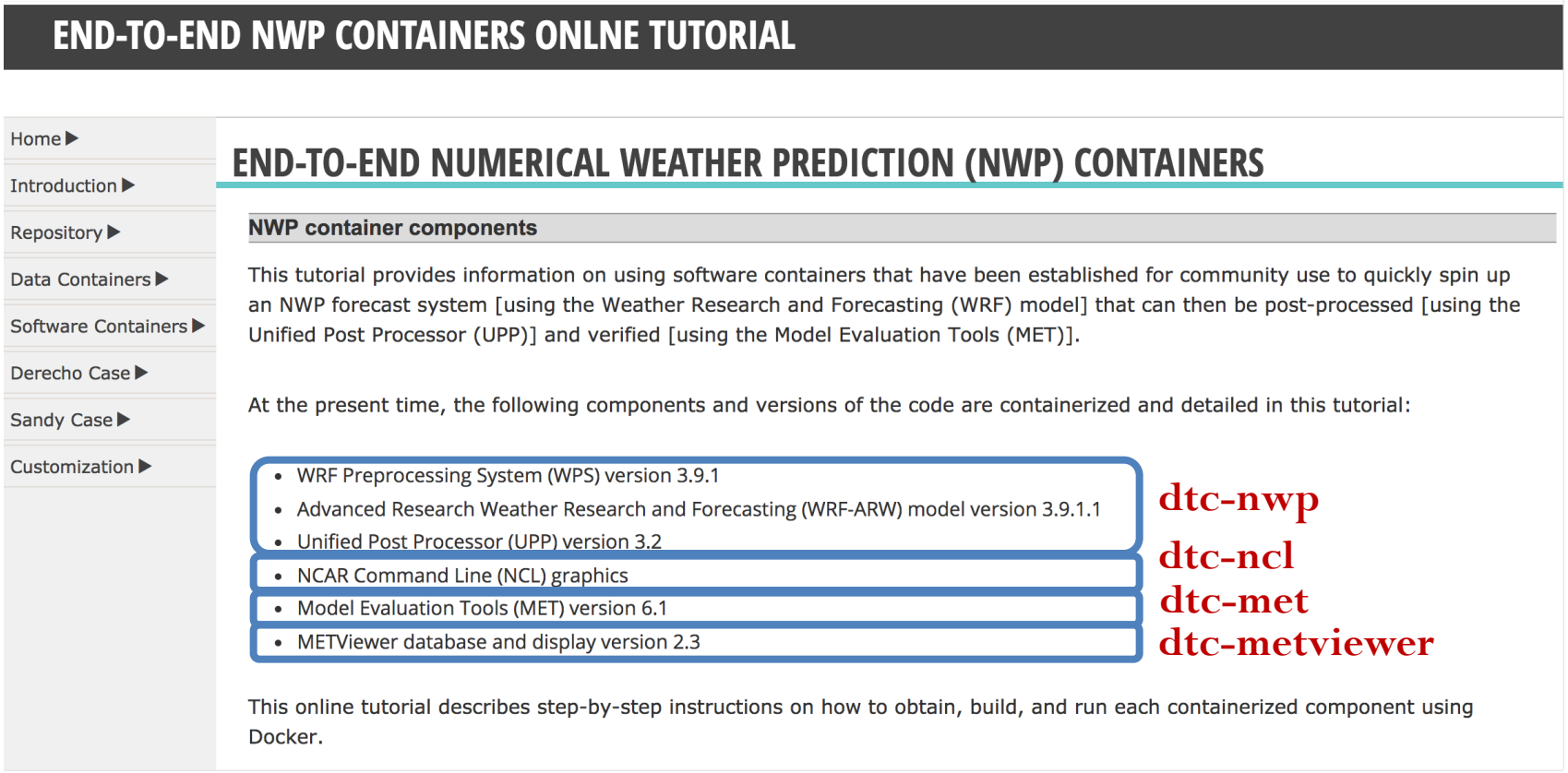






Container NWP Tutorial

https://dtcenter.org/met/docker-nwp/tutorial/container_nwp_tutorial/index.php



Thanks and Questions?

Contacts:

John Halley Gotway (Software) – johnhg@ucar.edu

Tressa Fowler (Statistician) – tressa@ucar.edu

Tara Jensen (Project Manager) – jensen@ucar.edu

MET Helpdesk: ***met_help@ucar.edu***

http://www.dtcenter.org/met/users/support/met_help.php

Support for this project provided by NGGPS Federally Funded Opportunities and made available to the community through the Developmental Testbed Center (DTC) which is funded by NOAA, Air Force and NSF