The Model Evaluation Tools (MET) Tutorial

Tressa L. Fowler
John Halley Gotway
Randy Bullock

June 2009

Today you get just a taste of MET

1.5 hour
MET tutorial

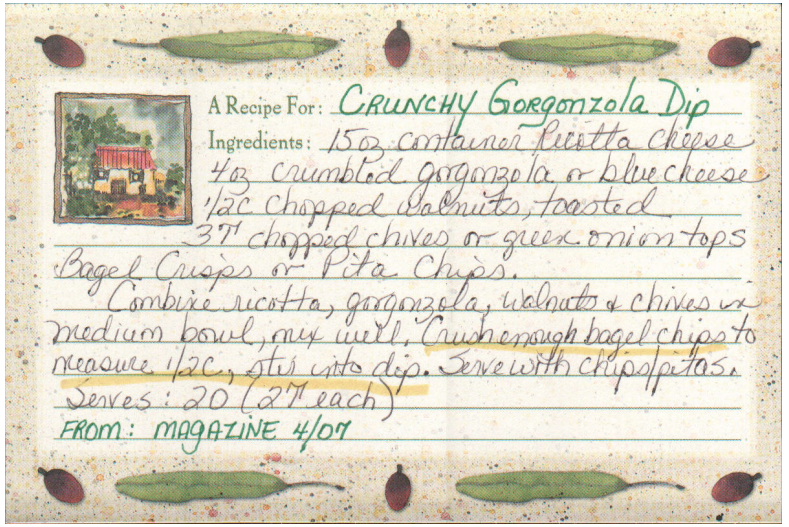
1.5 day MET tutorial July 23-24, 2009

Overview

- Existing MET tools (Tressa)

- Recent MET enhancements (John)

- Imminent MET tools (Tressa and Randy)



MET Online Tutorial

http://www.dtcenter.org/met/users/support/online_tutorial/METv2.0/index.php

Walks you through MET tools command line by command line.

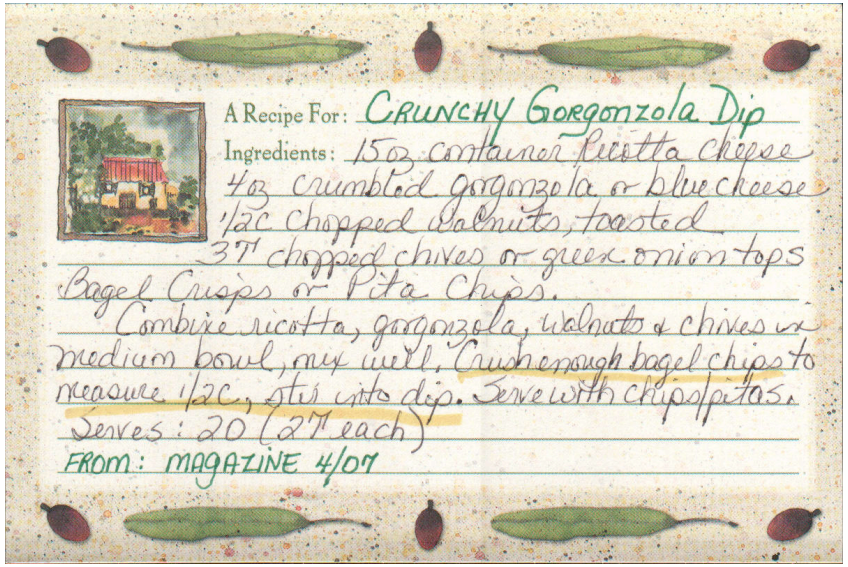
Existing MET Tools

- Data preprocessing
  - Convert ascii data to netCDF (ascii2nc tool)
  - Convert prepBUFR to netCDF (pb2nc tool)
  - Accumulate precipitation over time (pcp_combine)
- Individual forecast verification
  - Verify forecast with point observations (Point-Stat)
  - Verify forecast with gridded observations (Grid-Stat)
    - Neighborhood methods
  - Verify forecast objects with observed objects (MODE)
- Cumulative analysis (Stat-Analysis and MODE-Analysis Tools)

New MET Tools

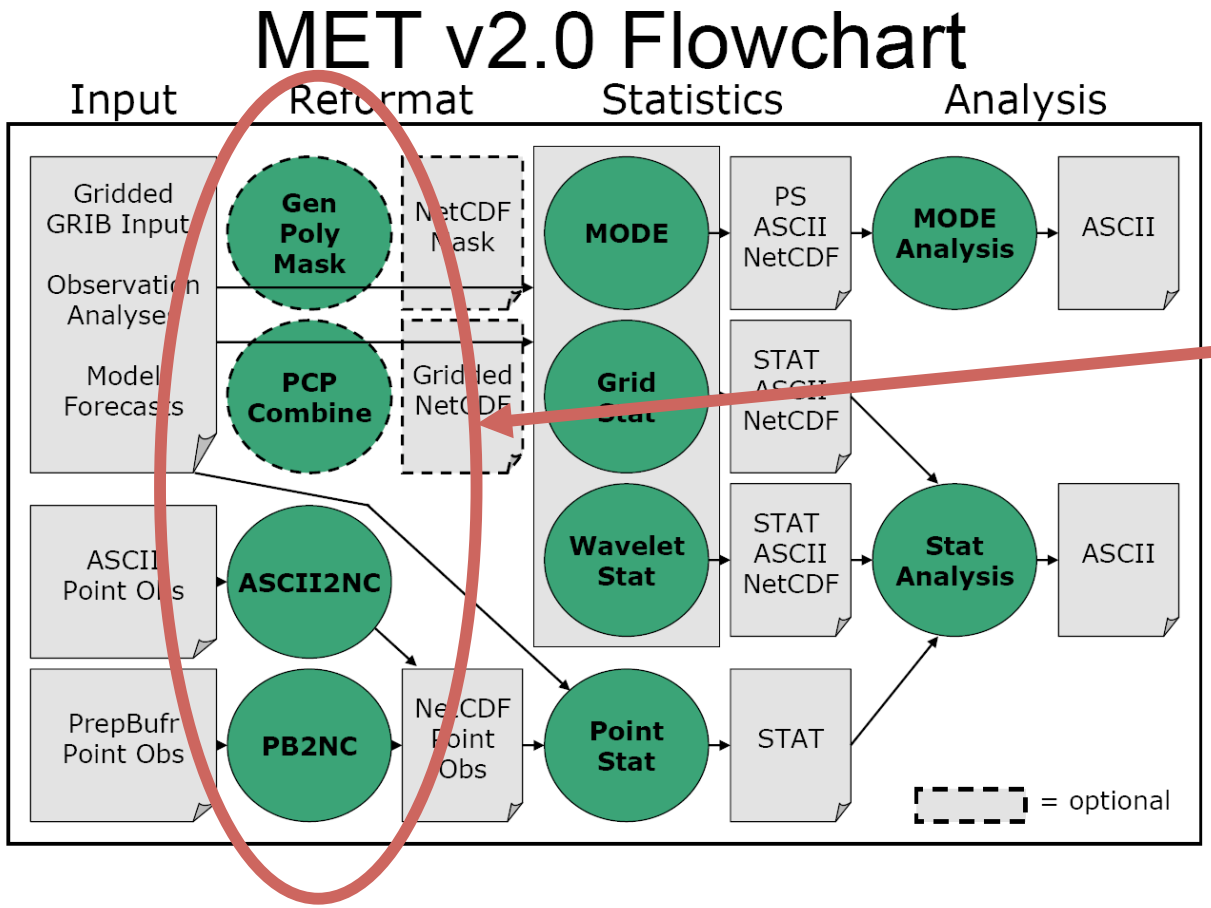
- Generate Polyline Masking Region (GenPolyMask tool)
- Probabilistic forecast verification (Point-Stat and Grid-Stat)
- Wind forecast verification (Stat-Analysis).
- Intensity Scale (aka Wavelet) tool

Imminent MET Tools



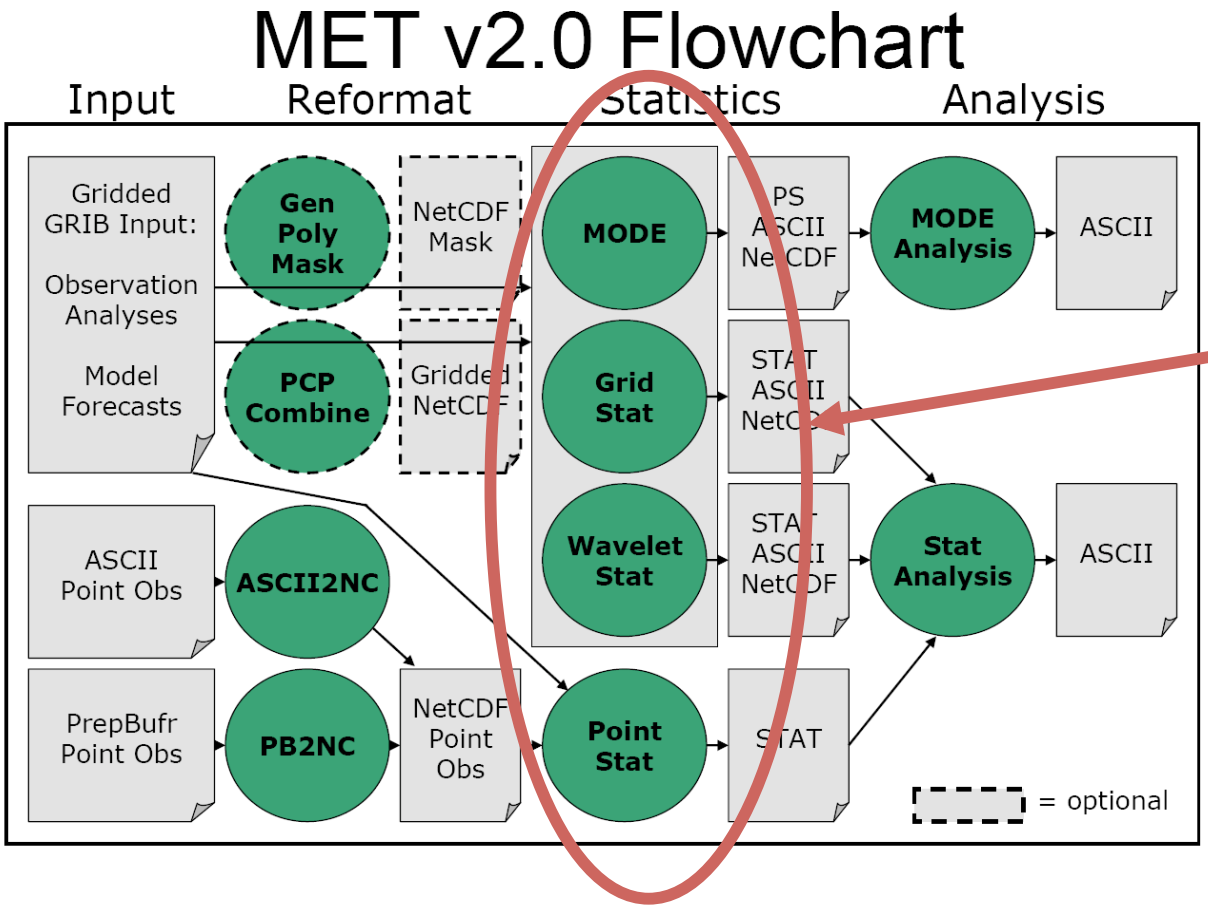
- MODE time domain
- Ensemble forecast verification
- Satellite data ingest
- Cloud verification

MET is…



[Data Preprocessing tools]: Place data in the format(s) expected by the statistics tools

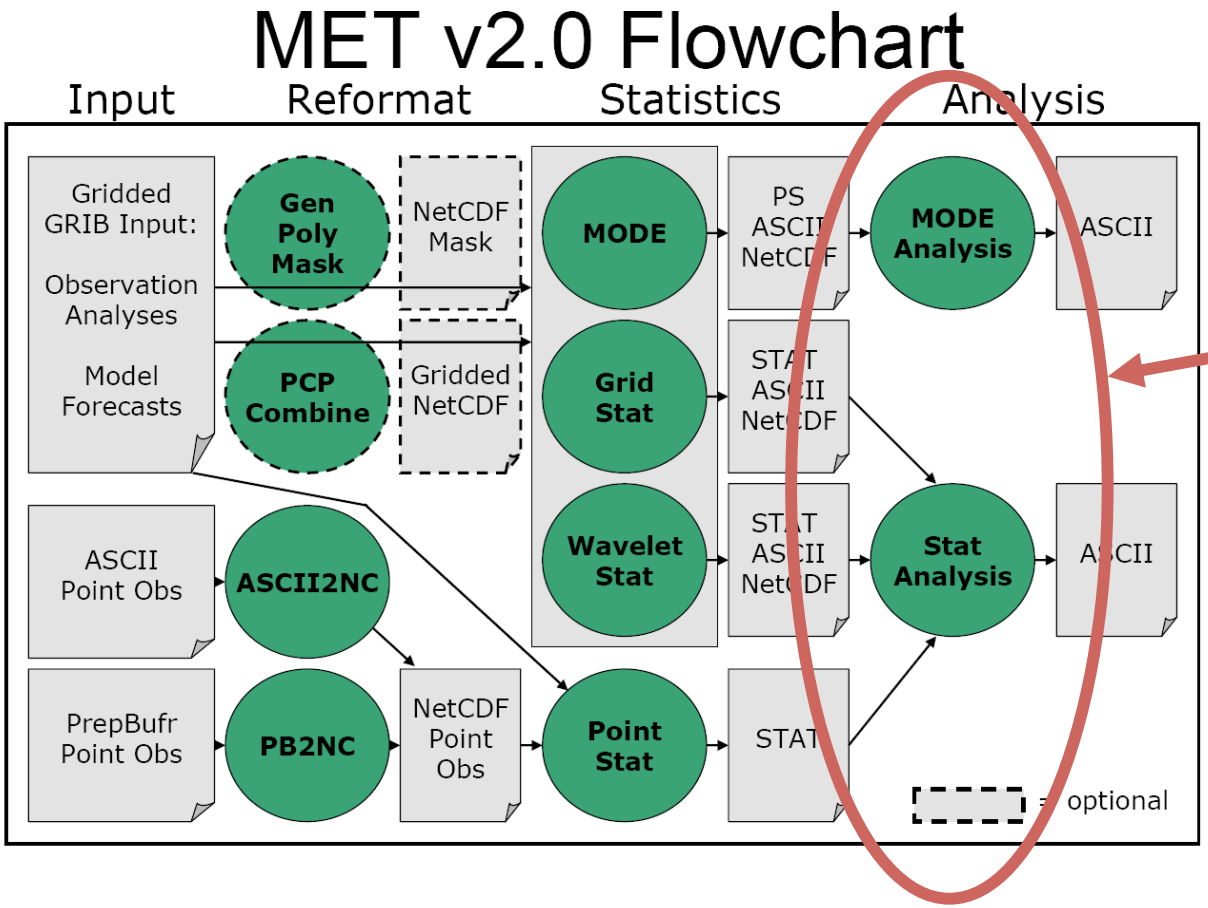
MET is…



Individual Forecast verification tools

- Traditional methods
  - Gridded obs
  - Point obs
  - Confidence intervals
- Spatial methods
  - Object-based
  - Neighborhood
  - Wavelet

MET is…



(Cumulative) Analysis tools

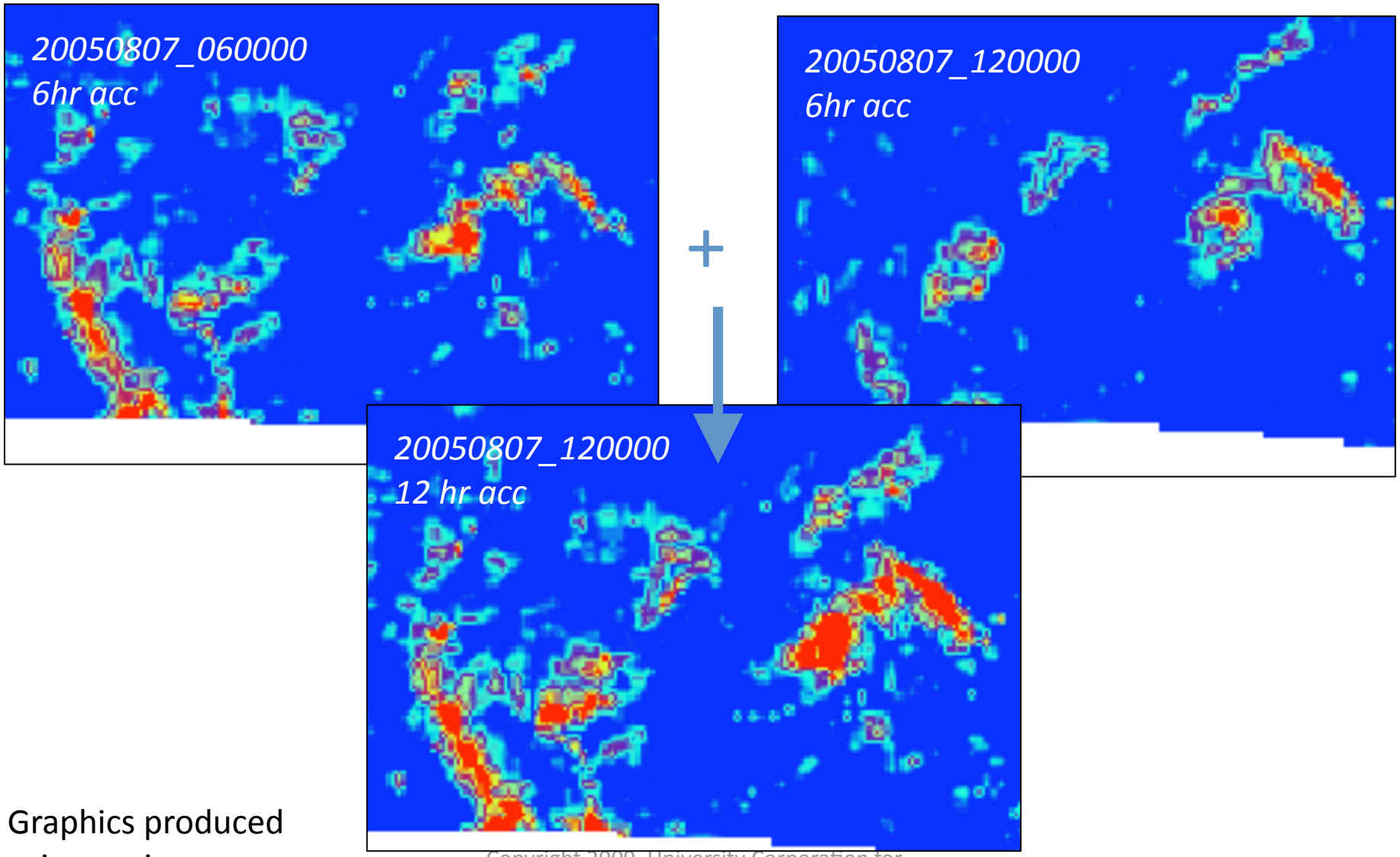
- Summarize statistics across cases
- Stratify according to various criteria (e.g., lead time)

PCP-Combine Tool

- Functionality:
  - Sum precipitation across multiple files.
  - Add precipitation in 2 files (i.e. NMM output).
  - Subtract precipitation in 2 files (i.e. ARW output).

- Data formats:
  - Reads GRIB.
  - Writes gridded NetCDF as input to stats tools.

PCP-Combine Example



Graphics produced using ncview

PB2NC Tool

- Functionality:
  - Filter and reformat PREPBUFR point observations into intermediate NetCDF format.
  - Configuration file specifies:
    - Observation types, variables, locations, elevations, quality marks, and times to retain or derive for use in Point-Stat.
- Data formats:
  - Reads PREPBUFR using NCEP's BUFRLIB.
  - Writes point NetCDF as input to Point-Stat.
- ~~CWORDSH utility for FORTRAN blocking~~

ASCII2NC Tool

- Functionality:
  - Reformat ASCII point observations into intermediate NetCDF format.
  - One input ASCII format supported (10 columns):
    - Message_Type, Station_ID, Valid_Time
    - Lat(Deg North), Lon(Deg East), Elevation(msl)
    - Grib_Code, Level, Height(msl), Observation_Value
  - No configuration file.

- Data formats:
  - Reads ASCII.
  - Writes point NetCDF as input to Point-Stat.
  - Support additional ASCII formats based on user input.

MET Statistics modules (Point and Grid Stat): Traditional verification measures

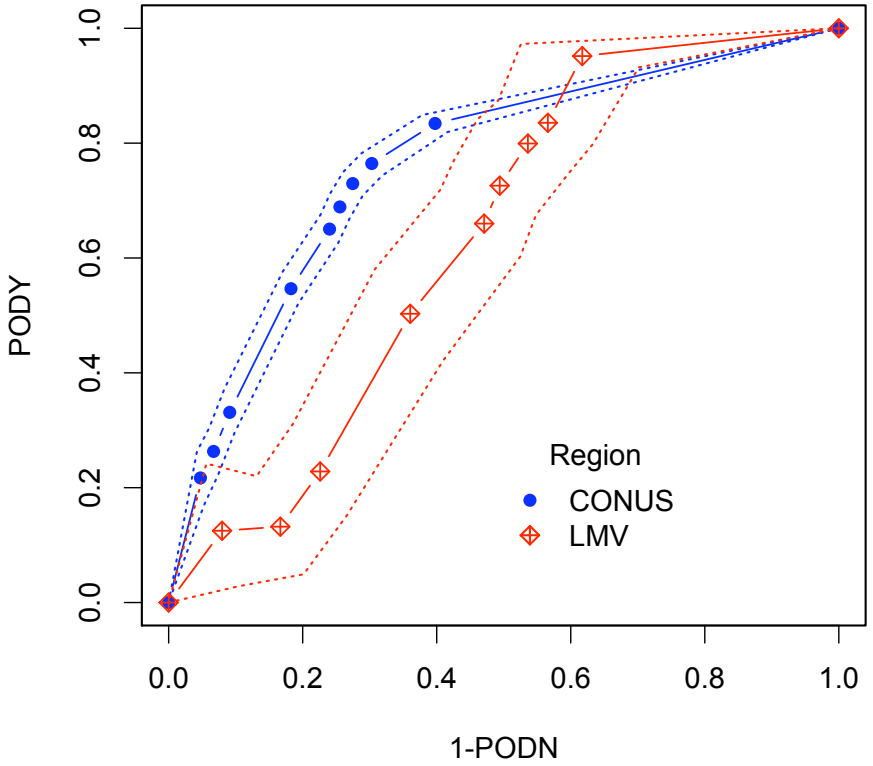
- Gridded and point verification
  - Multiple interpolation and matching options
- Statistics
  - **Continuous** - RMSE, BCRMSE, Bias, Correlation, etc.
  - **Categorical** - POD, FAR, CSI, GSS, Odds Ratio, etc.
  - **Probabilistic** - Brier Score, Reliability, ROC, etc. in v2.0

Matching approaches:

MET allows users to select the number of forecast grid points to match to a point observations and the statistic to use to summarize the forecasts.

MET Statistics modules (Point and Grid Stat): Confidence Intervals (CIs)

- MET provides two CI approaches
  - Normal
  - Bootstrap
- CIs are critical for appropriate and meaningful interpretation of verification results
  - Ex: *Regional comparisons*



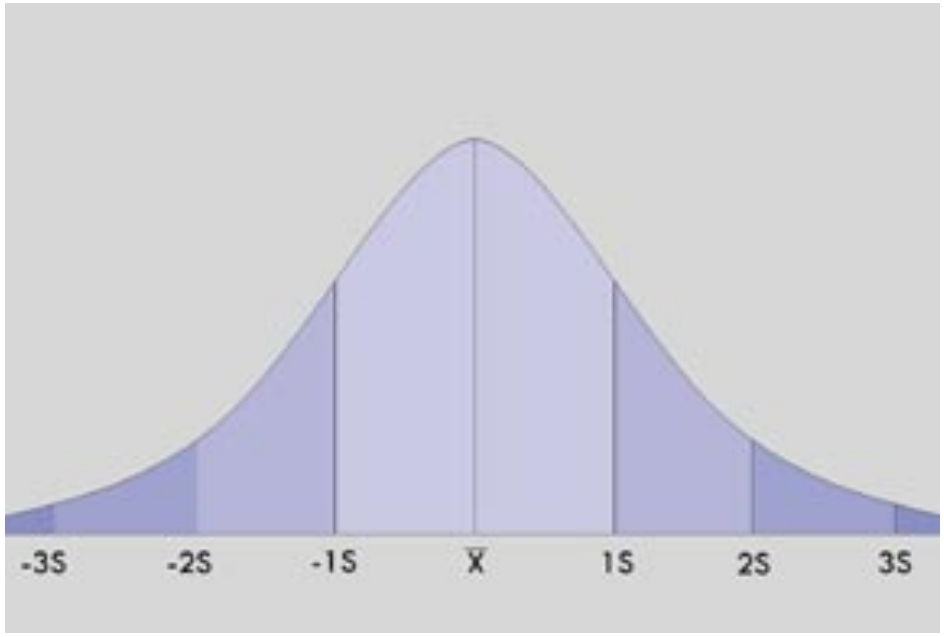
Accounting for Uncertainty

- Observational
- Model
  - Model parameters
  - Physics
  - Verification scores
- Sampling
  - Verification statistic is a realization of a random process.
  - What if the experiment were re-run under identical conditions?

Confidence Intervals (CI's)

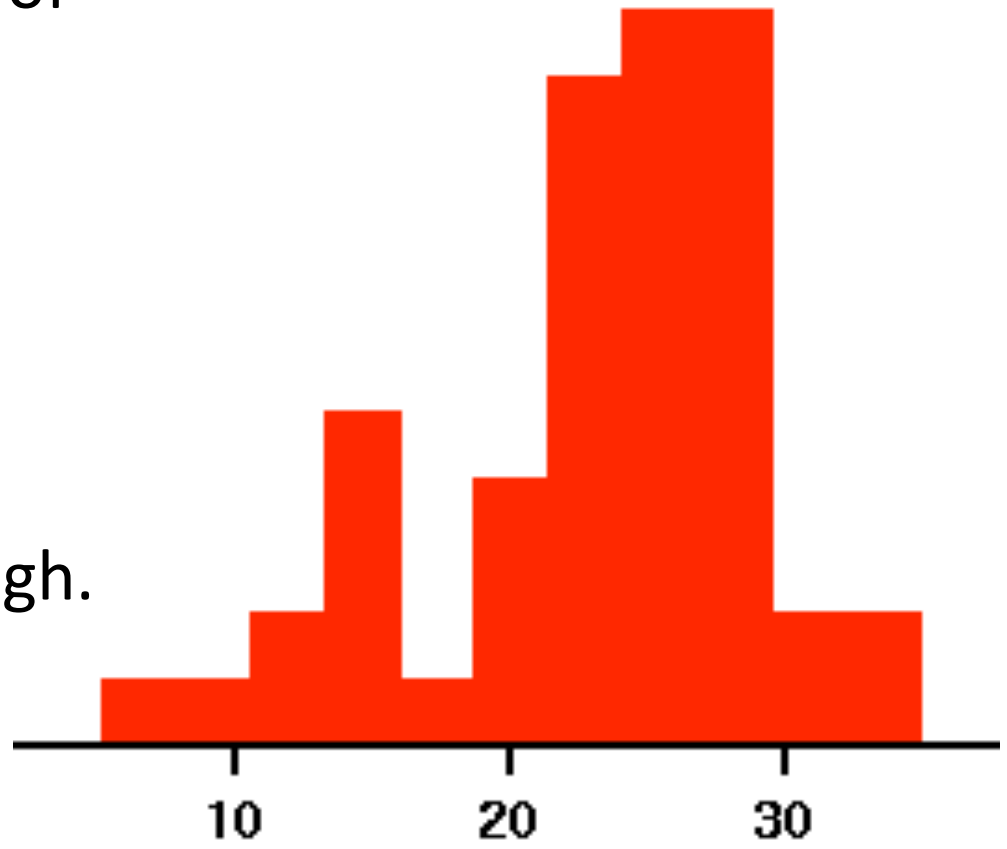
- Parametric
  - Assume the observed sample is a realization from a known *population* distribution with possibly unknown parameters (e.g., normal).
  - Normal approximation CI's are most common.
  - Quick and easy.

$$\hat{\theta} \pm z_{\alpha/2} s\hat{e}(\theta)$$



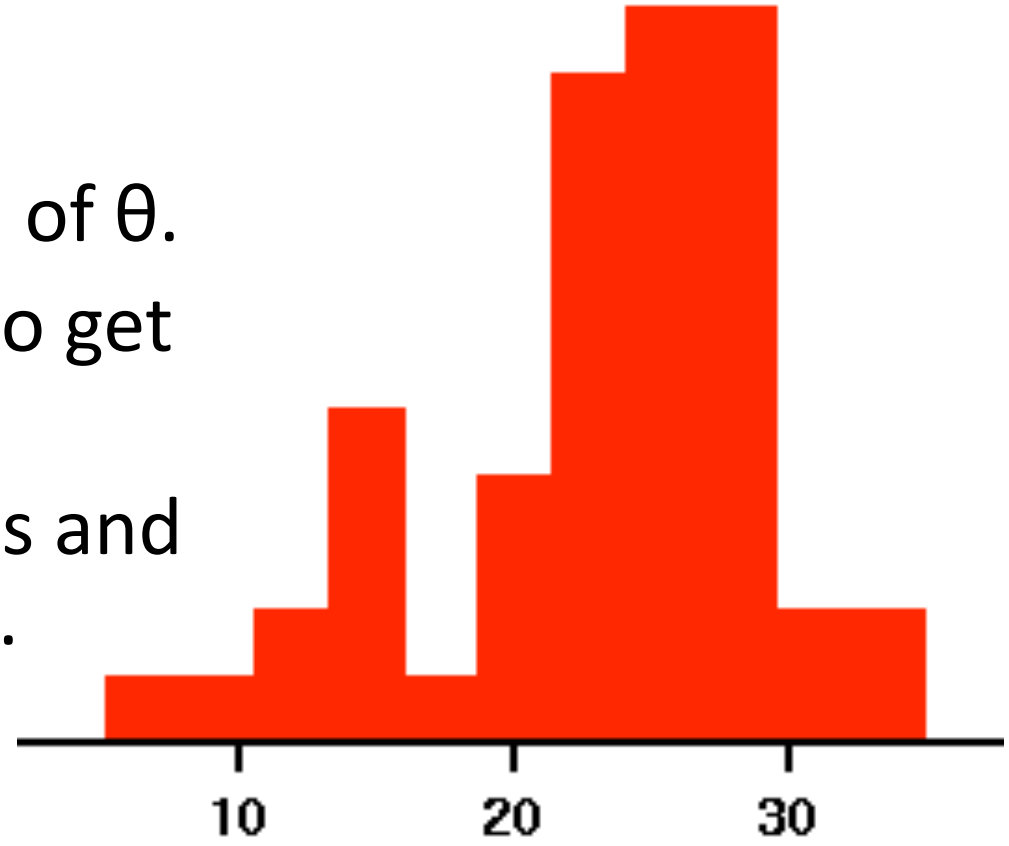
Confidence Intervals (CI's)

- Nonparametric
  - Assume the distribution of the observed sample is representative of the *population* distribution.
  - Bootstrap CI's are most common.
  - Can be computationally intensive, but easy enough.



Bootstrap Confidence Intervals (CI's)

- Resample from data *with replacement*.
- Calculate statistic θ
- Repeat to get empirical distribution (histogram) of θ.
- Count in on both ends to get CI (percentile method)
- Do BCa to adjust for bias and skewness in resampling.



MET Statistics modules: Spatial verification approaches

- Meaningful evaluations of spatially-coherent fields (e.g., precipitation)

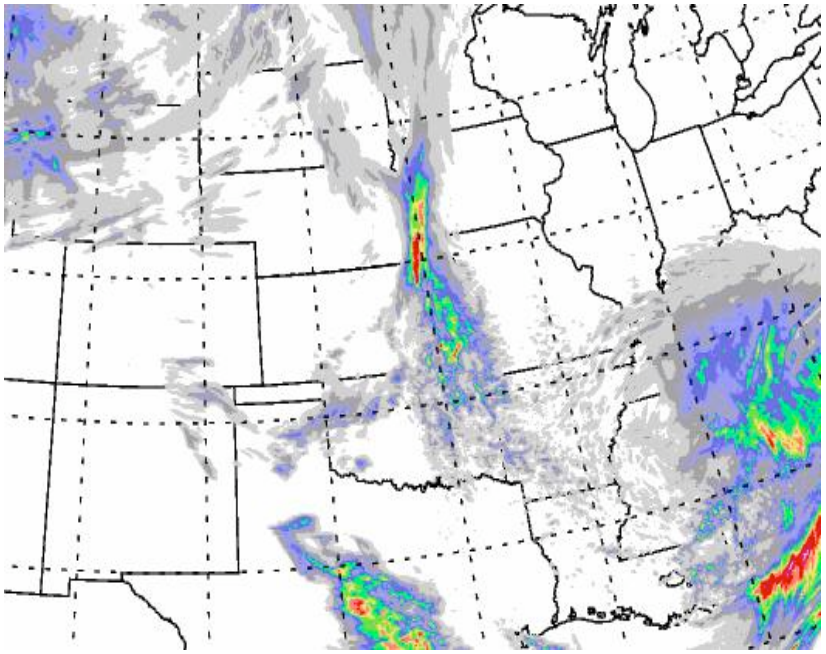
  Examples

  - *What* is wrong with the forecast?
  - At what scales does the forecast perform well?
  - How does the forecast perform on attributes of interest to users?

- Methods included in MET
  - Object-based: Method for Object-based Diagnostic Evaluation (MODE)
  - Neighborhood; Example: Fractional Skill Score (FSS in Grid Stat)
  - Scale-separation: Casati's Intensity-Scale measure (Wavelet Tool)

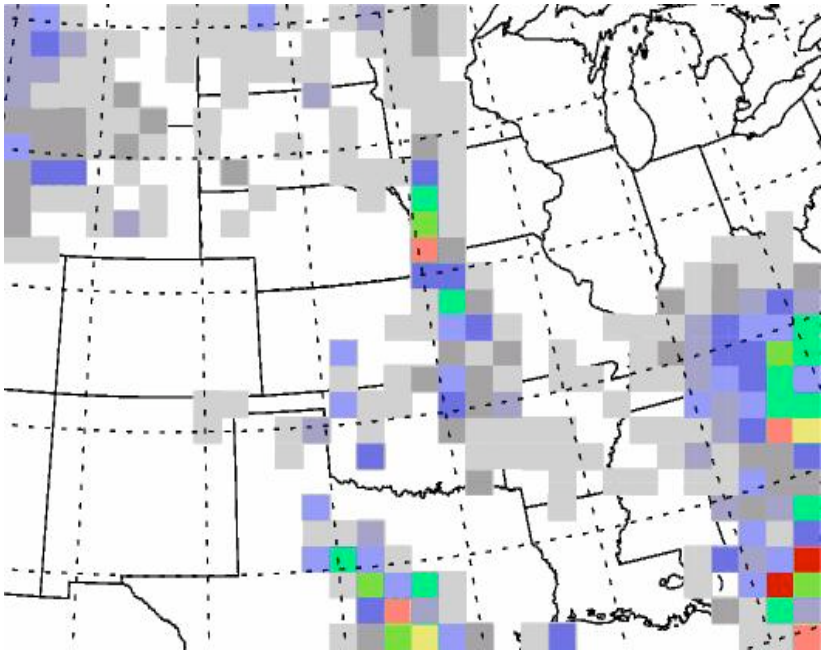
Neighborhood verification methods (Grid-Stat Tool)

- Also called “fuzzy” verification
- Upscaling
  - Put observations and/or forecast on coarser grid
  - Calculate traditional metrics
- Provide information about scales where the forecasts have skill



Neighborhood verification methods

- Also called “fuzzy” verification
- Upscaling
  - Put observations and/or forecast on coarser grid
  - Calculate traditional metrics
- Provide information about scales where the forecasts have skill

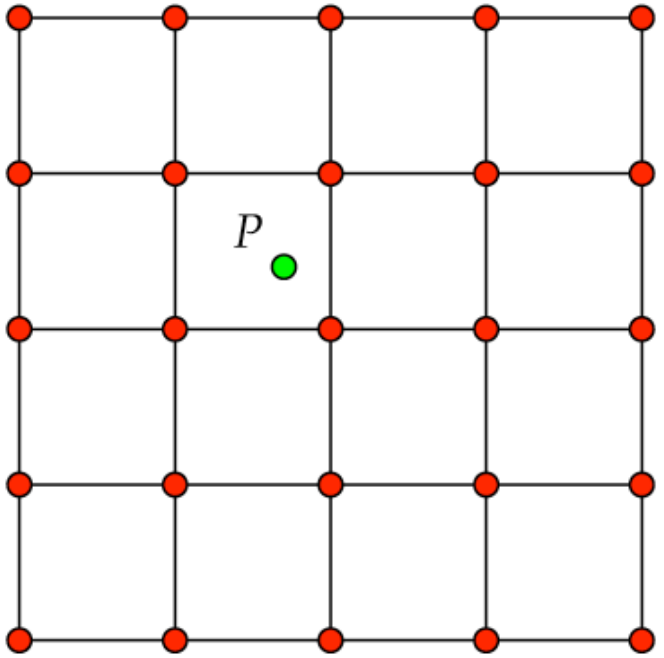


Interpolation

Need to Choose:

(1) Method

(2) Width



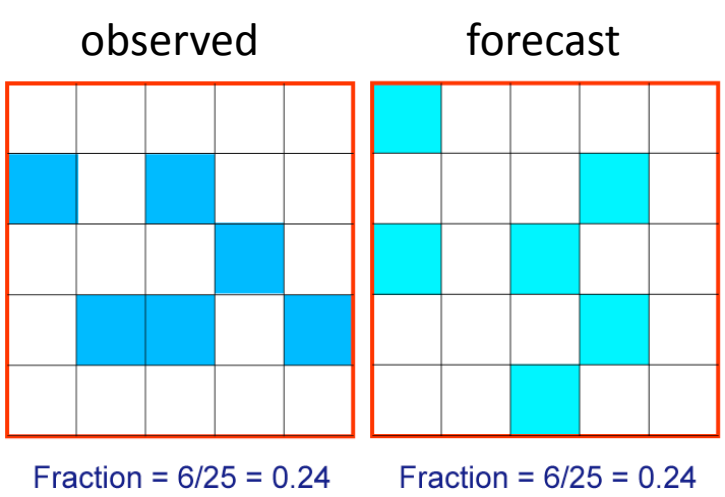
Interpolation Methods

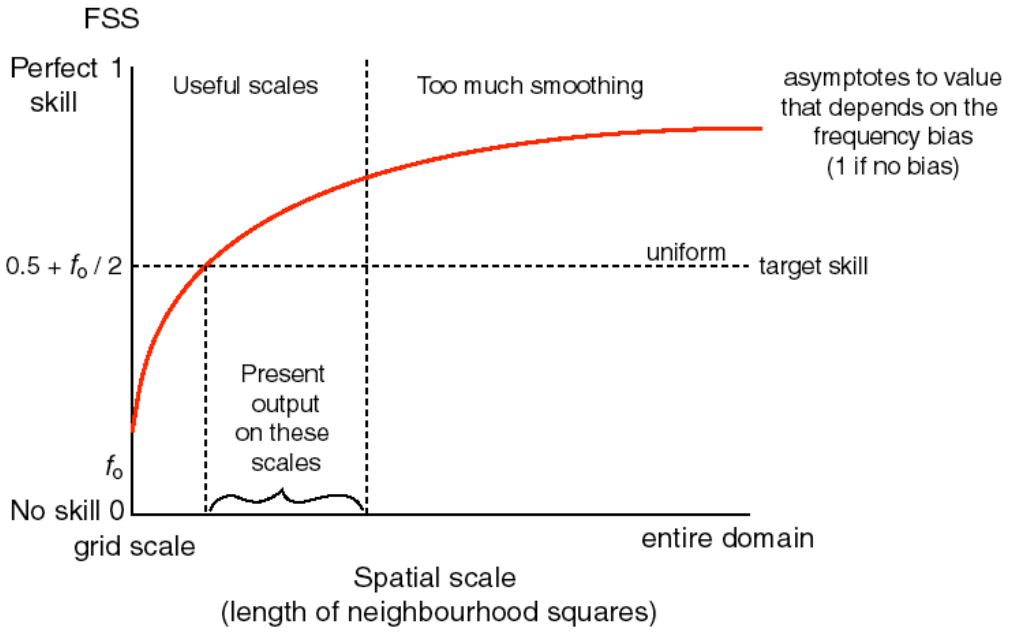
	Min	Max	Median	UW Mean	DW Mean	Nearest Nbr	Least Squares
Point Stat	✓	✓	✓	✓	✓	✓	✓
Grid Stat	✓	✓	✓	✓	N/A	N/A	N/A

For Grid Stat, these are smoothing methods.

Neighborhood verification methods

Example: Fractional skill score (Roberts and Lean, MWR, 2008)





From Mittermaier 2008

Ebert (2008; Met Applications) describes the neighborhood methods in MET

Stat and MODE Analysis Tools

Used to :

- Filter
- Summarize
- Aggregate

results over many times, leads, thresholds, domains, etc.

Stat Analysis Tool: Run aggregate

"-job aggregate -dump_row out/aggr_ctc_job.stat -level P850-750"

Point Stat Output *(i.e. point_stat_out.stat)*

```
V2.0     WRF     … ADPUPA G212 … TMP
    P850-750 … >278.00 CTC
    401      192      11      24      174
    UW_MEAN 1
```

```
V2.0     WRF     … ADPSFC G212 … TMP
    P850-750 … >278.00 CTC
    167      25      23      0      119
    UW_MEAN  1
```

(NOTE: header modified to show only pertinent info)

		OBS		
FCST		Y	N	
	Y	192	11	*203*
	N	24	174	*198*
		216	*185*	**401**

		OBS		
FCST		Y	N	
	Y	25	23	*48*
	N	0	119	*119*
		25	*142*	**167**

Stat Analysis Tool: Run aggr

Stat Analysis Output *(i.e. stat_analysis.out)*

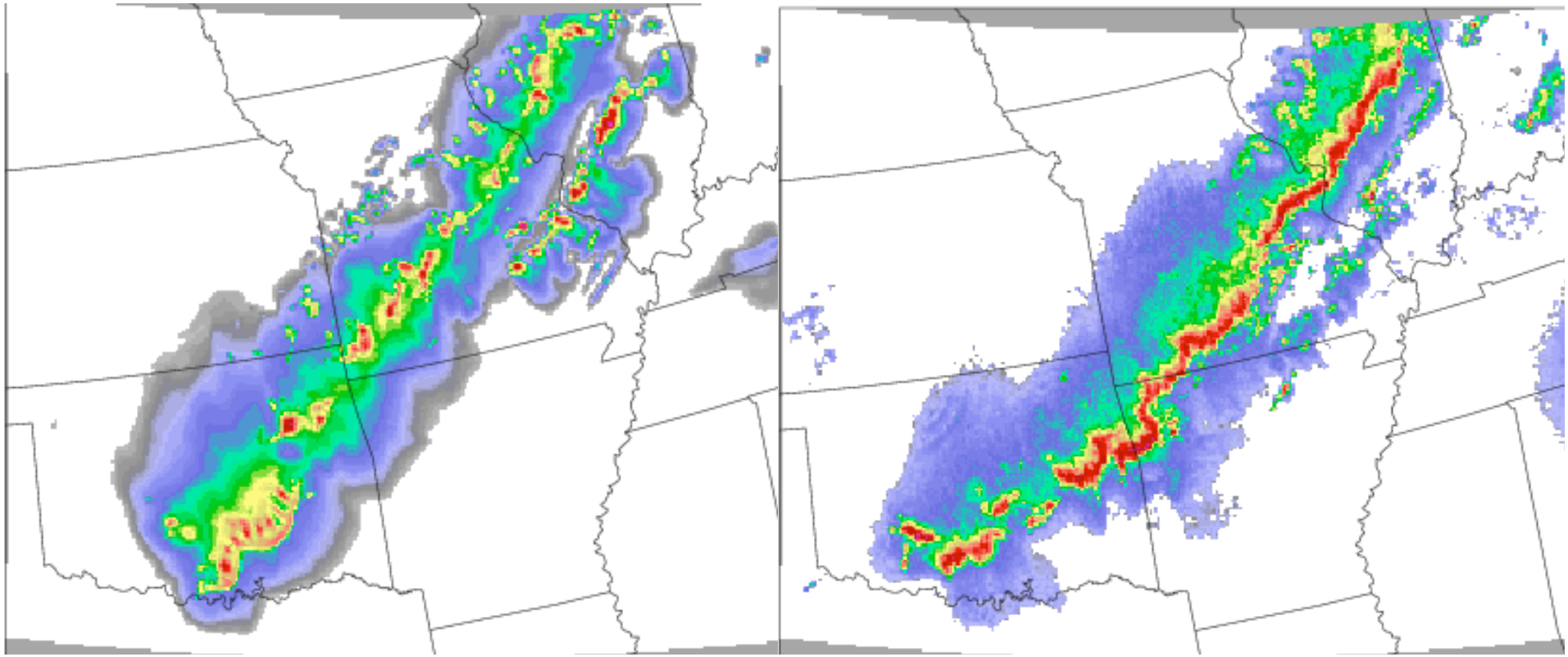
```
JOB_LIST: -job aggregate
-vx_mask G212 -line_type CTC
   -fcst_thresh >278.000 -var TMP
   -level P850-750 -dump_row out/
   aggr_ctc_job.stat
COL_NAME:        TOTAL
   FY_OY            FY_ON
   FN_OY          FN_ON
   INTERP_MTHD      INTERP_PNTS
CTC:             568                217
   34               24          293
```

		OBS		
FCST		Y	N	
	Y	217	34	251
	N	24	293	317
		241	327	568

MODE Example

Forecast

Observation



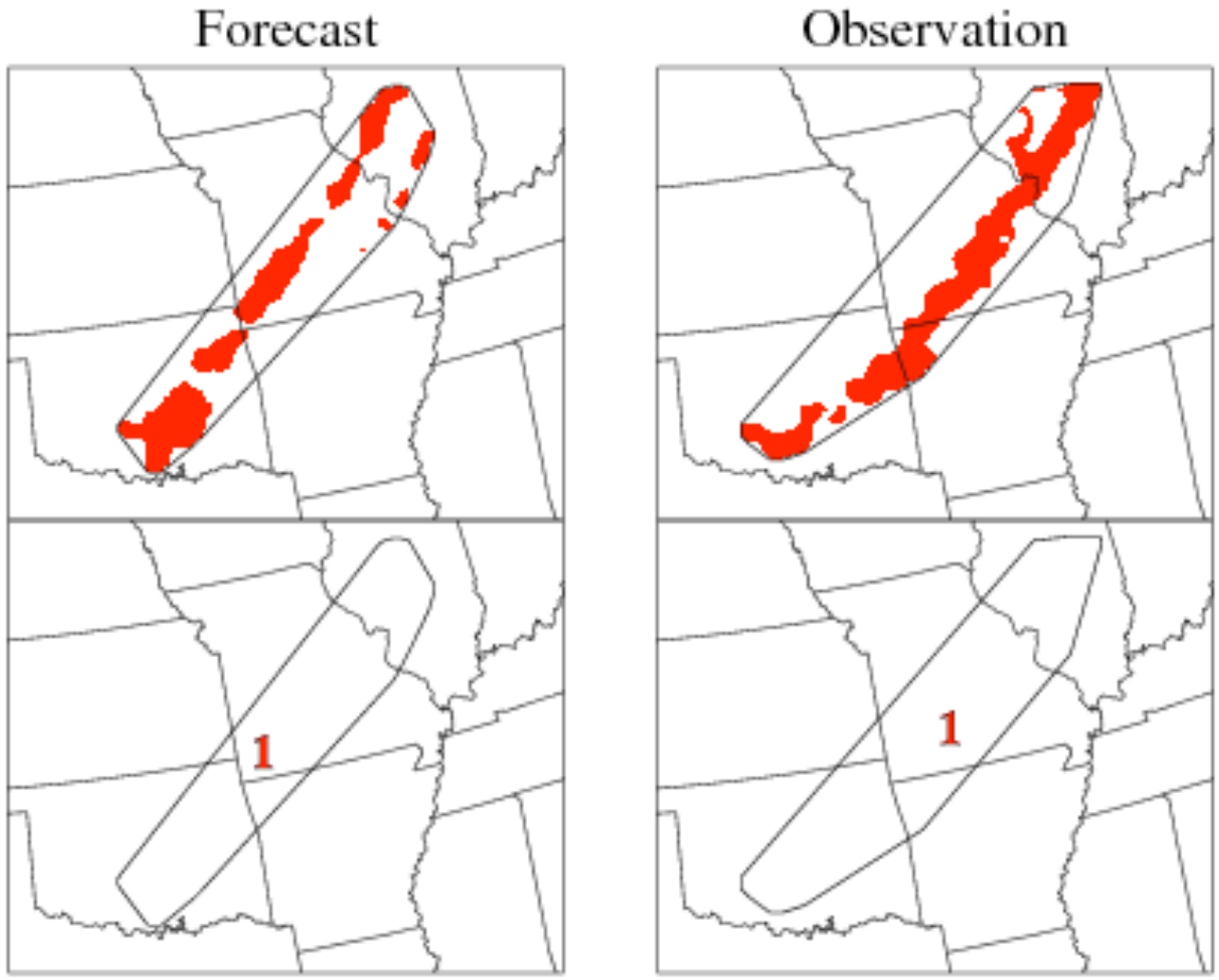
30 DBz threshold

Smooth

Threshold

Merge

Compare

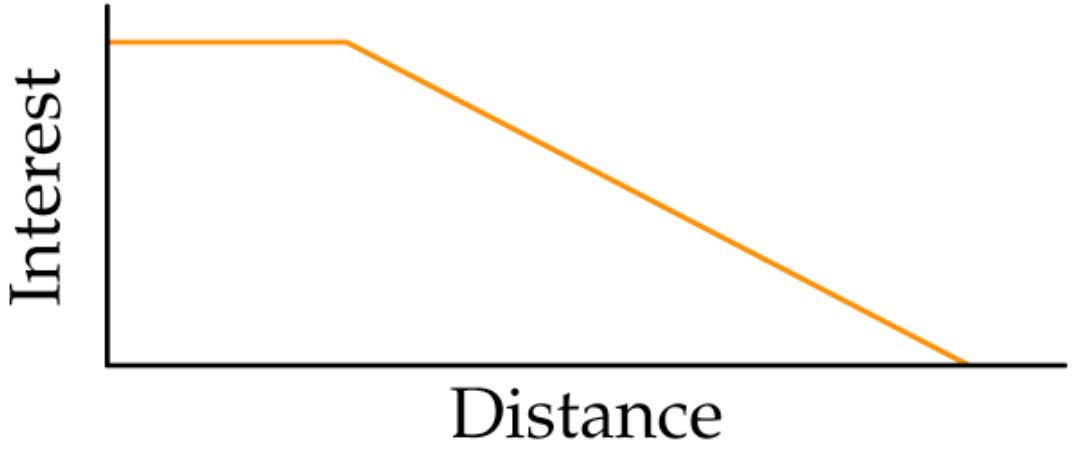


CLUS PAIR	CEN DIST	ANG DIFF	FCST AREA	OBS AREA	INTER AREA	UNION AREA	SYM DIFF	FCST INT50	OBS INT50	FCST INT90	OBS INT90	TOT INTR
1	24.31	2.65	3973	5109	963	8119	7156	34.00	35.95	45.00	50.63	0.9653

Interest Maps

Map attributes to interest values.

Example: Centroid Distance



All interest maps can be changed in the config file.

And now John will cover the enhancements to MET for version 2.0.

RECENT ENHANCEMENTS TO THE MODEL EVALUATION TOOLS (MET)

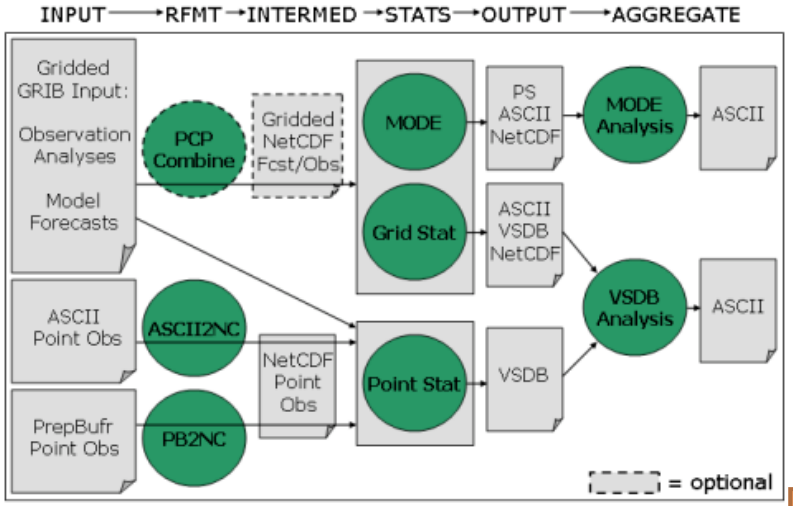
26 June 2009

Release History

- METv0.9: Beta release – July, 2007
- METv1.0: First official release – January, 2008
- METv1.1: Incremental upgrades – July, 2008
- METv2.0: Current release – April, 2009
  - About 500 registered users from 66 countries
  - 50/50 University/Non-University users
  - On-line tutorial updated for METv2.0.
  - Hands-on tutorial offered with the WRF-Tutorial
    - Previous – February, 2009
    - Upcoming – July, 2009

METv1.1 vs METv2.0

METv1.1 Flowchart



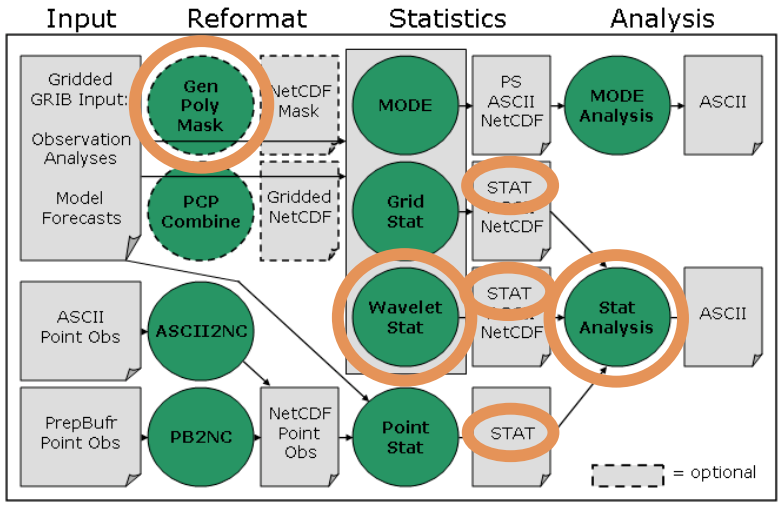
Visible Changes:

- Gen-Poly-Mask Tool
- Wavelet-Stat Tool
- VSDB to STAT Format

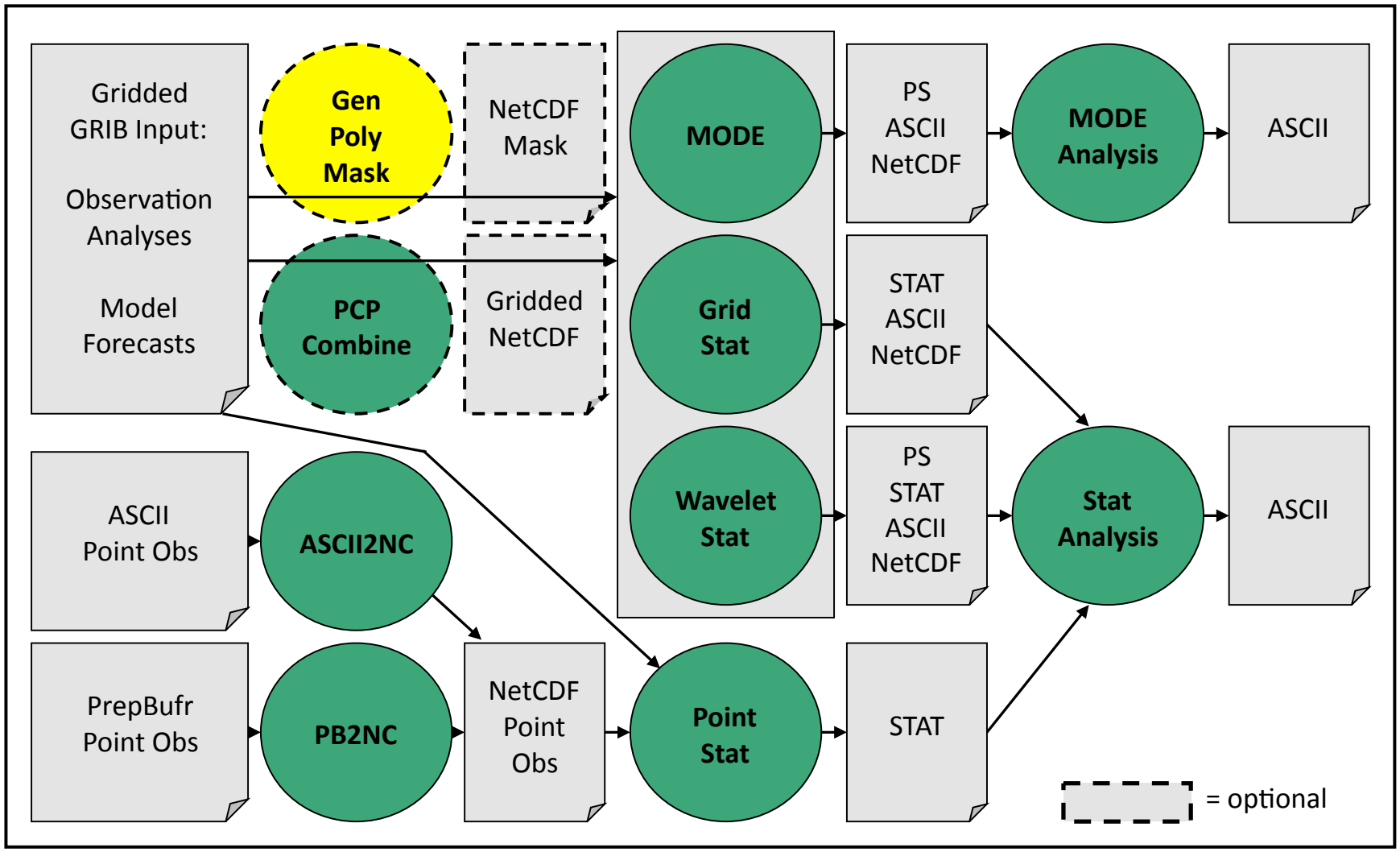
Internal Changes:

- Verifying Probabilities
- Comparing Different Fields
- Verifying Winds
- Internal Fortran-Blocking

METv2.0 Flowchart

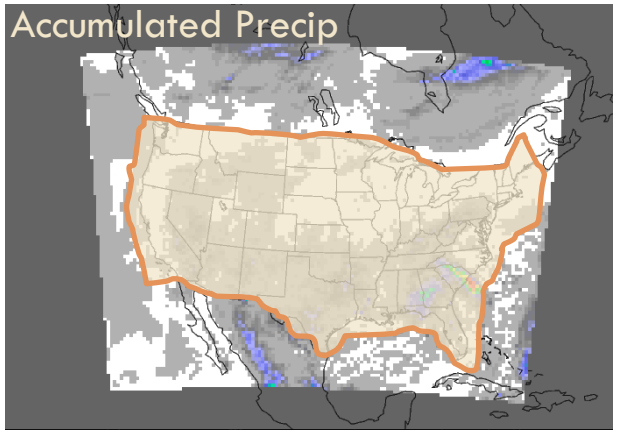


Gen-Poly-Mask Tool

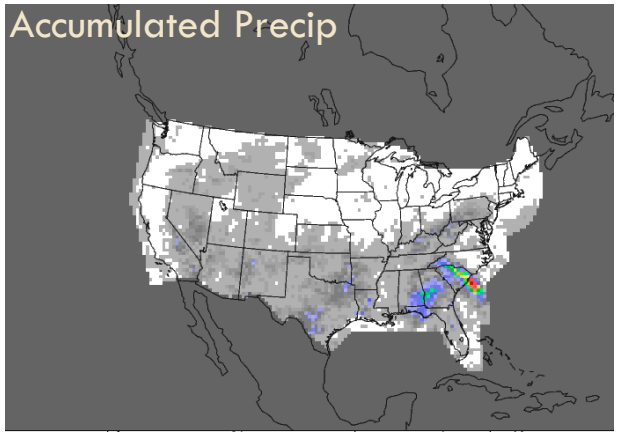


Gen-Poly-Mask Tool

- Inputs
  - GRIB file defining domain
  - ASCII Polyline verification region (Lat/Lon)
- Ouput
  - NetCDF file with masking bitmap



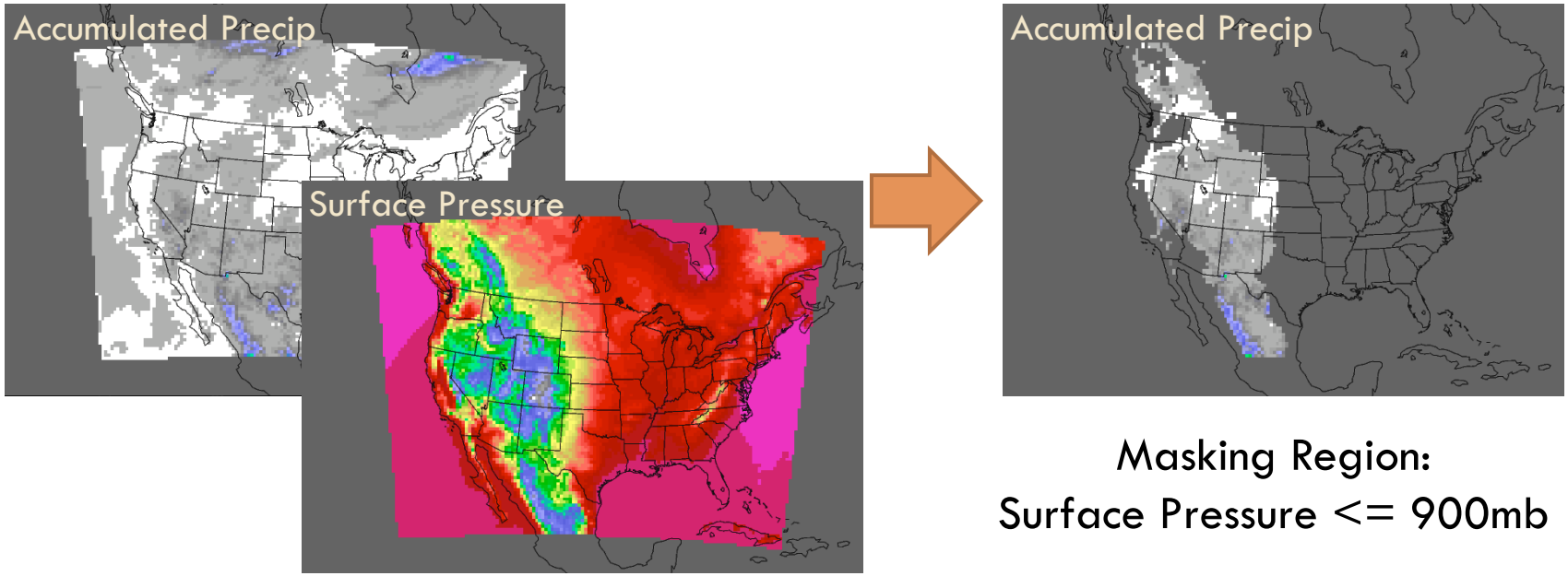

```
CONUS
31.1931 -120.4211
31.2291 -120.4976
31.2650 -120.5741
31.3009 -120.6123
31.3369 -120.6506
31.3728 -120.6888
31.4087 -120.6888
31.4447 -120.7270
...
```



Define once, apply many times

Data Masking

- Choose a data field and threshold to define the masking region.
  - Same grid as data to be verified.

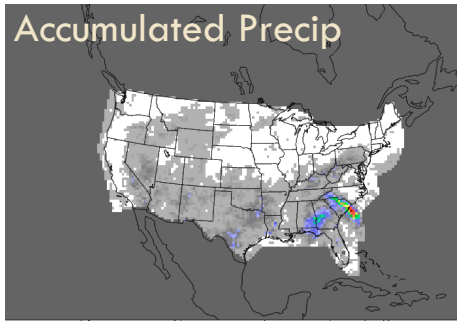


Masking Region:
Surface Pressure <= 900mb

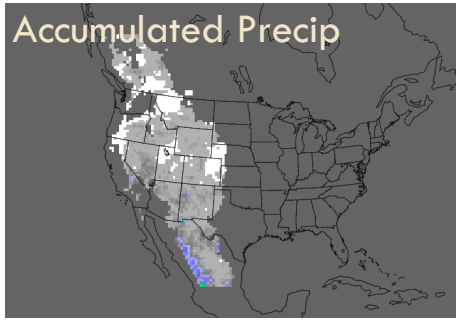
Masking Options

- Masking for Grid-Stat, Point-Stat, and MODE:

1. Ouput of Gen-Poly-Mask:

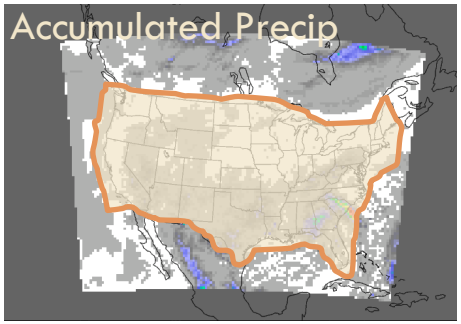


2. Gridded data field and threshold:



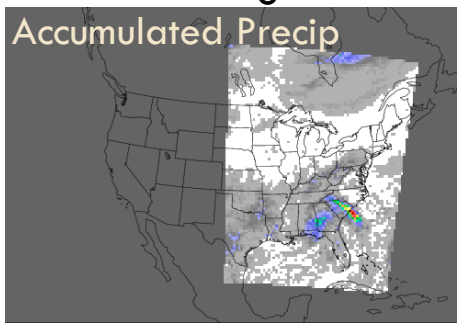
Surface Pressure <= 900mb

3. Lat/Lon Polyline file:




```
CONUS
31.1931 -120.4211
31.2291 -120.4976
31.2650 -120.5741
31.3009 -120.6123
31.3369 -120.6506
31.3728 -120.6888
31.4087 -120.6888
31.4447 -120.7270
...
```

4. Pre-defined grid:

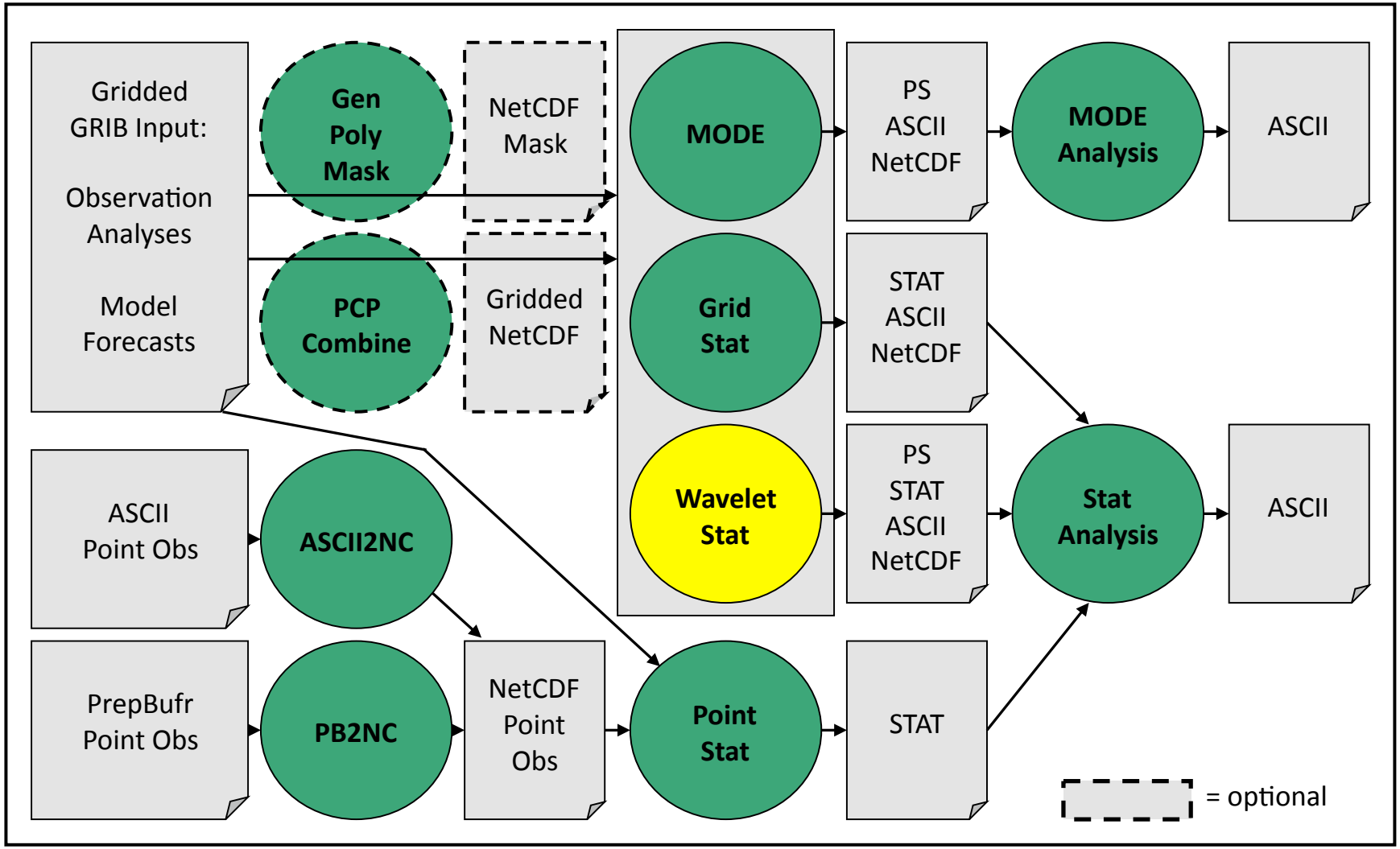


Grid = "DTC166"

NCEP Grids:

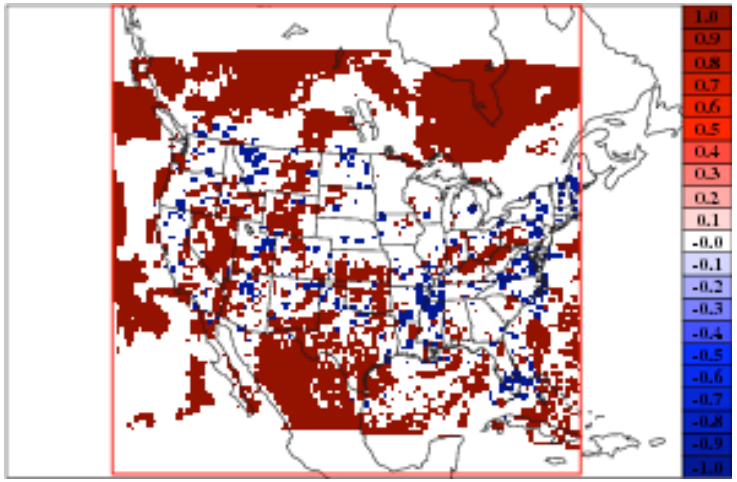
- 83 of them
- Named "GNNN"
- New custom grids require code change

Wavelet-Stat Tool

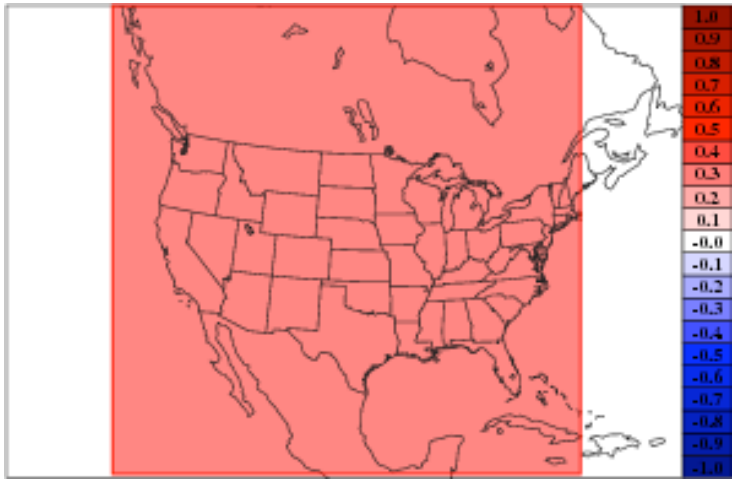


Wavelet-Stat Tool: Overview

- Implements Intensity-Scale verification technique, Casati et al. (2004)
- Evaluate skill as a function of intensity and spatial scale of the error.
- Method:
  - Threshold raw forecast and observation to create binary images.
  - Decompose binary thresholded fields using wavelets (Haar as default).
  - For each scale, compute the Mean Squared Error (MSE) and Intensity Skill Score (ISS).
  - At what spatial scale is this forecast skillful?

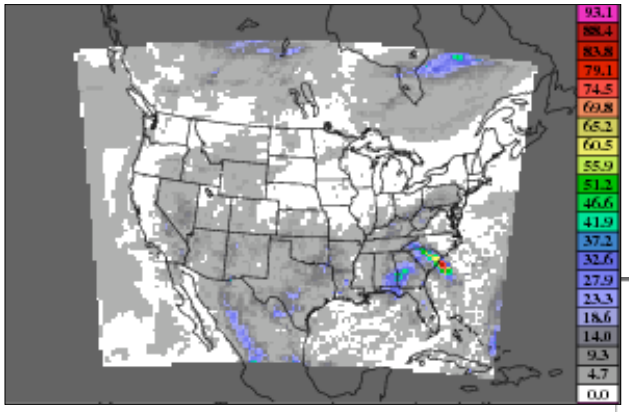


Difference (F-O) for precip > 0 mm

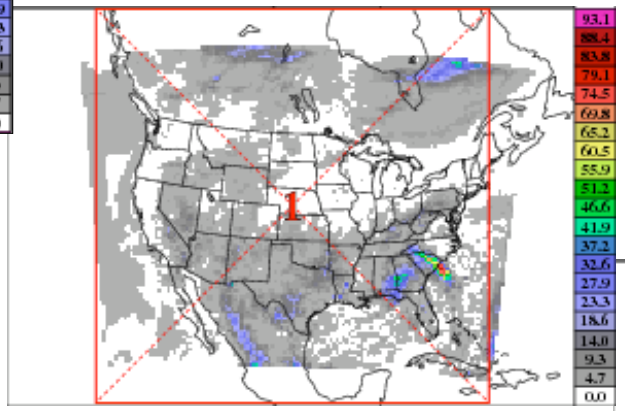


Wavelet decomposition difference

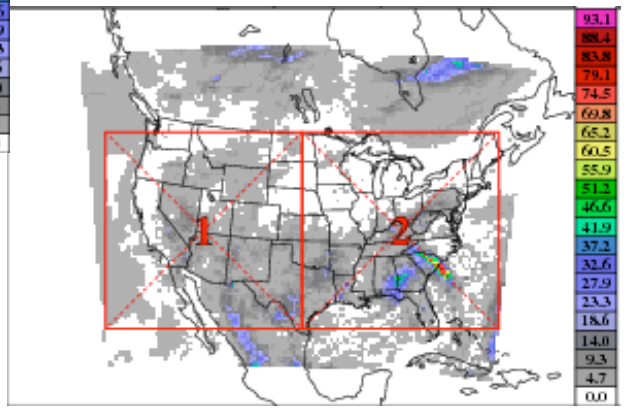
Wavelet-Stat Tool: Configure



- Handling missing data:
  - Set to zero for precipitation.
  - Set to mean of field for continuous variables.



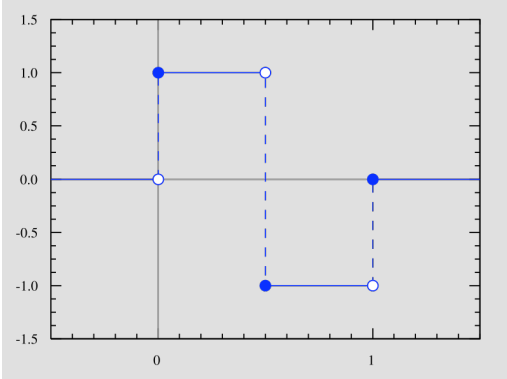
- 2^n x 2^n Grid
- Tiling options:
  - Automatic tile selection
  - User-defined tile(s)
  - Pad to nearest 2^n x 2^n

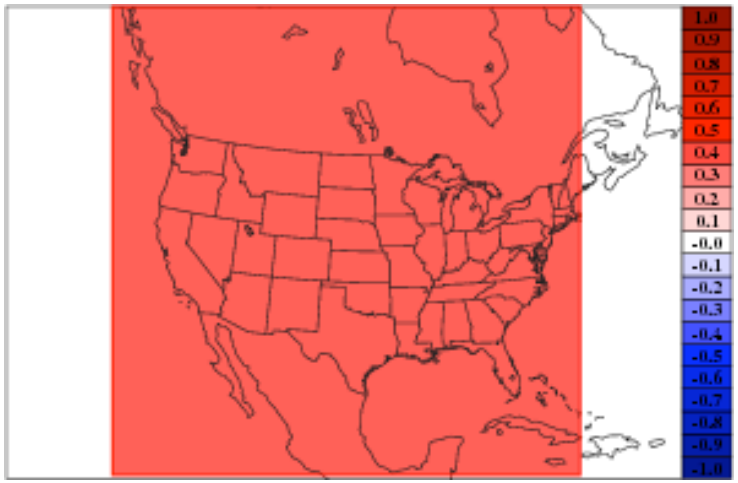


Wavelet-Stat Tool: Wavelets

- Haar, centered
  - Used in Casati et al. (2004)
  - Default configuration
  - Discontinuous data
  - 1 member
- Daubechies, centered
  - 9 members
- B-spline, centered
  - 11 members

Haar Wavelet





Daubechies (10) decomposition

Wavelet-Stat Tool: Output

1. ASCII STAT file
   - ISC (Intensity Skill-Score) line for each tile/threshold/scale
     - Header columns
     - Mean-Squared Error (MSE) and Intensity Skill Score (ISC)
     - Fcst&Obs Energy Squared (FENERGY2, OENERGY2)
     - Base Rate (BASER) and Frequency Bias (FBIAS)
2. NetCDF file
   - For each tile/threshold/scale
     - Forecast, Observation, and Difference fields
3. PostScript summary plot
   - Difference field image for each tile/threshold/scale

Model Name:

Init Time:
Valid Time:
Lead Time:
Accum Time:

Tile Method:
Tile Count:
Tile Dim:
Tile Corner:

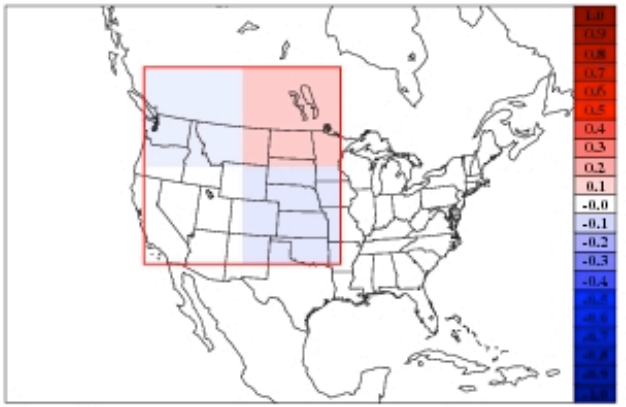
Mask Missing:
Wavelet(k):

W W W

Frequency B Frequency B Frequency B
Base Rate: Base Rate: Base Rate:
Mean-Square Mean-Square Mean-Square

Wavelet-Stat: APCP/A24, Tile 1, >0.100, Scale 6

Difference (F-0)



Frequency Bias:	1.82519	Intensity Skill Score:	0.92589
Base Rate:	0.28491	Fcst Energy Squared (%):	0.01550 (2.98)
Mean-Squared Error (%):	0.00539 (1.42)	Obs Energy Squared (%):	0.02233 (7.84)

W W W

Frequency B Frequency B Frequency B
Base Rate: Base Rate: Base Rate:
Mean-Square Mean-Square Mean-Square

Wavelet-Stat: APCP/A24, Tile 1, >0.100, Scale 7

Difference (F-0)

Frequency Bias:	1.82519	Intensity Skill Score:	0.23925
Base Rate:	0.28491	Fcst Energy Squared (%):	0.27042 (52.00)
Mean-Squared Error (%):	0.05528 (14.60)	Obs Energy Squared (%):	0.08117 (28.49)

Wavelet-Stat Tool: Summary

- Decomposes error by spatial scale.
- Options for selecting:
  - Field and thresholds
  - Wavelet type and shape
  - 2^n x 2^n tile(s) definition
    - Keep tiles fixed for multiple cases in time.
- Added support to STAT-Analysis tool to aggregate ISC data through time.

Verifying Probabilities

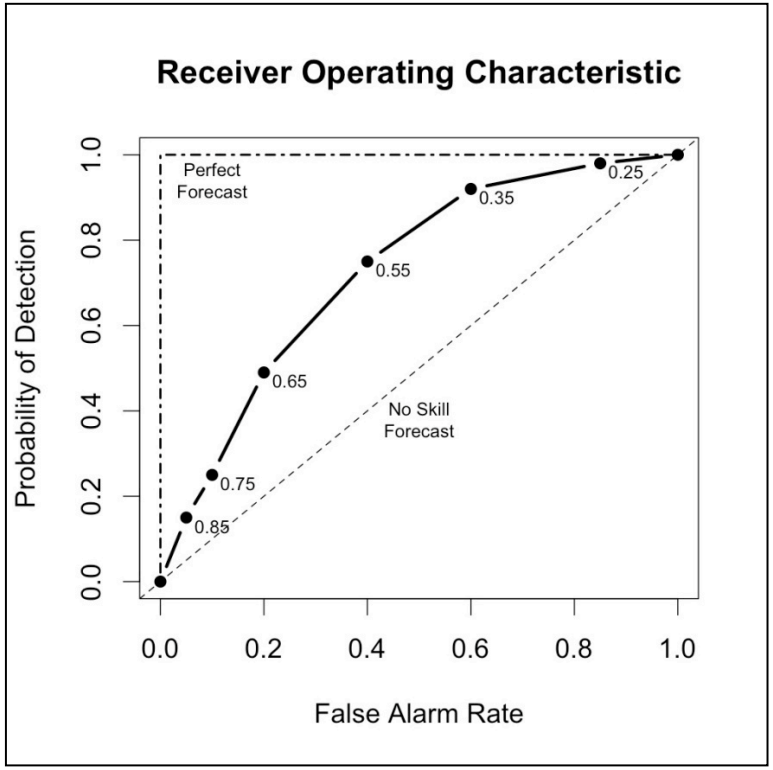
- Probabilistic verification methods added for Grid-Stat, Point-Stat, and Stat-Analysis.
- Define Nx2 contingency table using:
  - Multiple forecast probability thresholds
  - One observation threshold

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:
  - Probability of precip
    [0.00, 0.25, 0.50, 0.75, 1.00]
  - Accumulated precip
    > 0.00

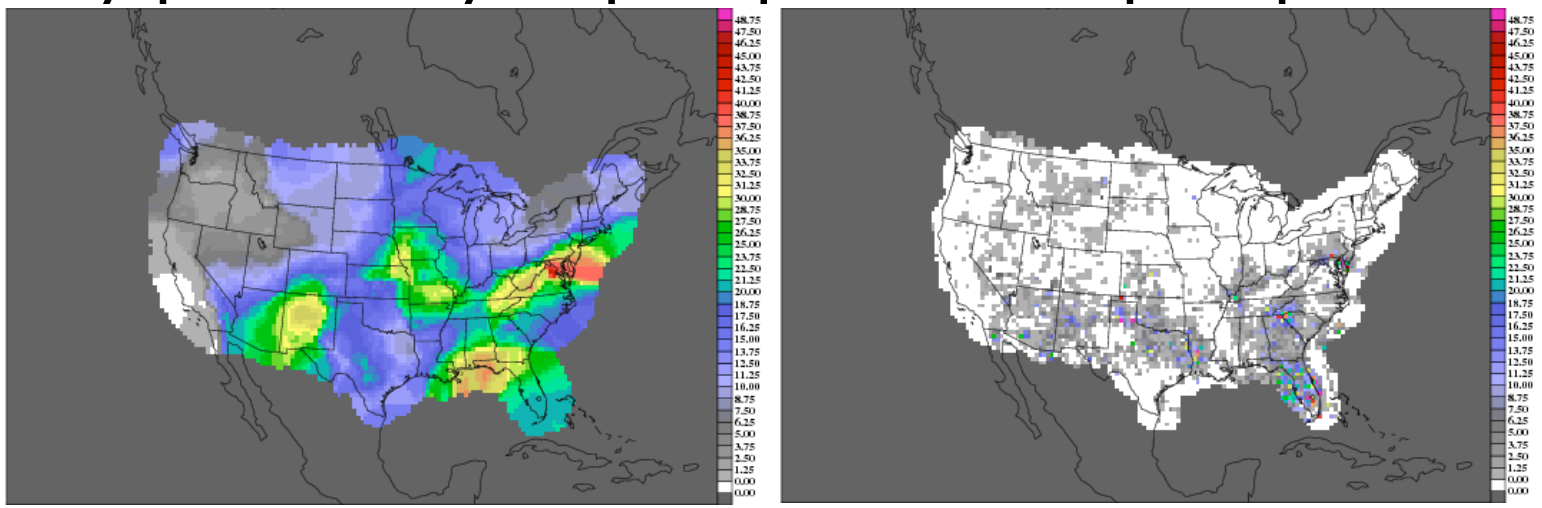
Verifying Probabilities: Output

- Statistical Output (Line Type):
  - Nx2 Table Counts (PCT)
  - Joint/Conditional factorization table with calibration, refinement, likelihood, and base rate by threshold (PJC)
  - Receiver Operating Characteristic (ROC) plot points by threshold (PRC)
  - Reliability, resolution, uncertainty, area under ROC Curve, and Brier Score (PSTD)



Verifying Probabilities: Example

- Verify probability of precip with total precip:



- Configuration file settings:
  - fcst_field[] = ["POP/Z0/PROB"];
  - obs_field[] = ["APCP/A12"];
  - fcst_thresh[] = ["ge0.00 ge0.25 ge0.50 ge0.75 ge1.00"];
  - obs_thresh[] = ["gt0.00"];

Comparing Different Fields

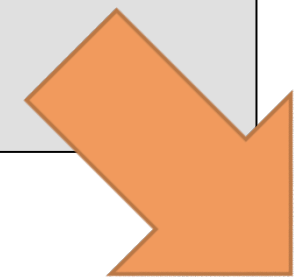
- For probabilities, compare two different fields.
- Generalize MET tools to compare any two fields.
- User must interpret results.
- Example: Total precip vs. convective precip
  - Configuration file settings:
    - fcst_field[] = ["APCP/A24"];
    - obs_field[] = ["ACPCP/A24"];
    - fcst_thresh[] = ["gt0.0 ge20.0"];
    - obs_thresh[] = []; (leave blank to use fcst setting)

VSDB to STAT

METv1.1

VSDB File format:

- 11 Line Types
- 10 common header columns
  - Times, var, level



METv2.0

STAT File format:

- 15 Line Types
- 21 common header columns
  - Fcst times, vars, levels
  - Obs times, vars, levels

Post-processing scripts/tools may need to be modified.

Verifying Winds

- Verify u, v, and speed, but not wind direction.
- Incremental support for verification of winds:
  - Enhancements for Point-Stat and Grid-Stat:
    - Add wind speed thresholds to determine which U/V pairs are included in the vector partial sums (VL1L2).
  - Enhancements for Stat-Analysis:
    - Support new job to aggregate one or more vector partial sum lines and compute statistics for the wind direction errors.
      - Mean forecast and observation wind directions, mean error (F-O), and mean absolute error

Wind Direction: Example

Point-Stat: VL1L2 Lines

VX_MASK	THRESH	LINE_TYPE	TOTAL	UFBAR	VFBAR	UOBAR	VOBAR	UVFOBAR	UVFFBAR	UVOOBAR
DTC_165	>=1.000	VL1L2	653	1.91117	0.07900	1.40658	-0.06126	13.01039	18.12575	20.31649
DTC_165	>=3.000	VL1L2	279	3.13561	-0.35096	2.87061	-0.30072	26.50472	30.03257	38.25362
DTC_165	>=5.000	VL1L2	96	5.21268	-2.74580	5.47813	-2.01667	49.90791	51.10427	70.78802
DTC_166	>=1.000	VL1L2	2431	-1.62742	0.25391	-1.23402	-0.04393	18.48309	29.70179	21.89615
DTC_166	>=3.000	VL1L2	1610	-1.84581	0.16061	-1.47491	-0.11217	24.45214	36.67400	29.36032
DTC_166	>=5.000	VL1L2	520	-0.93518	-0.45435	-0.25923	-0.49558	37.21821	52.51917	47.26483

Stat-Analysis: aggregate_stat jobs

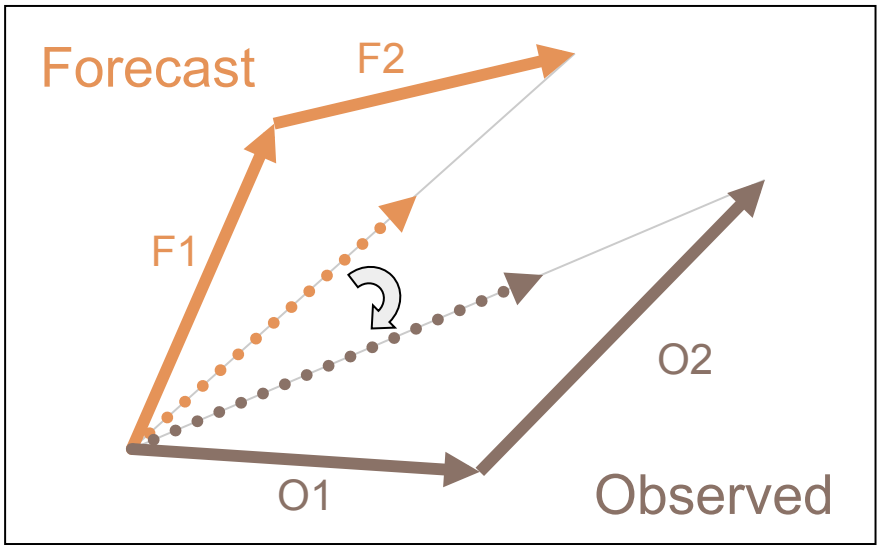
```
JOB_LIST:      -job aggregate_stat -fcst_thresh >=1.000 -line_type VL1L2 -out_line_type WDIR
     COL_NAME: TOTAL FBAR      OBAR     ME        MAE
ROW_MEAN_WDIR: 2     183.25038 0.22749  -3.02289  7.88372
    AGGR_WDIR: 3084  103.87238 85.96574 -17.90663 NA
```

```
JOB_LIST:      -job aggregate_stat -fcst_thresh >=3.000 -line_type VL1L2 -out_line_type WDIR
     COL_NAME: TOTAL FBAR     OBAR     ME        MAE
ROW_MEAN_WDIR: 2     5.67967  0.81565  -4.86402  4.86402
    AGGR_WDIR: 1889  94.38140 80.45939 -13.92200 NA
```

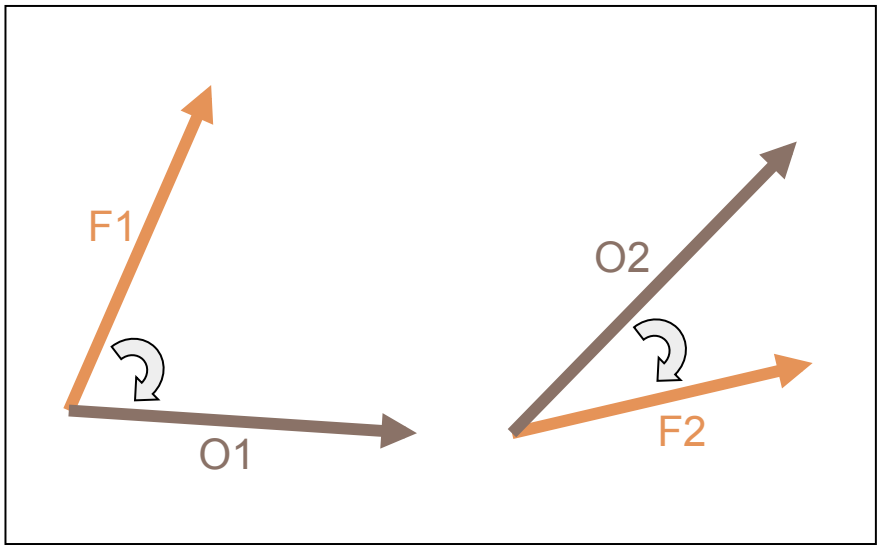
```
JOB_LIST:      -job aggregate_stat -fcst_thresh >=5.000 -line_type VL1L2 -out_line_type WDIR
     COL_NAME: TOTAL FBAR      OBAR      ME        MAE
ROW_MEAN_WDIR: 2     0.93288   338.91179 -22.02109 22.02109
    AGGR_WDIR: 616   358.38152 319.08761 -39.29391 NA
```

Wind Direction: Output

- AGGR_WDIR
  1. Aggregate VL1L2 partial sums lines
  2. Derive wind directions and errors

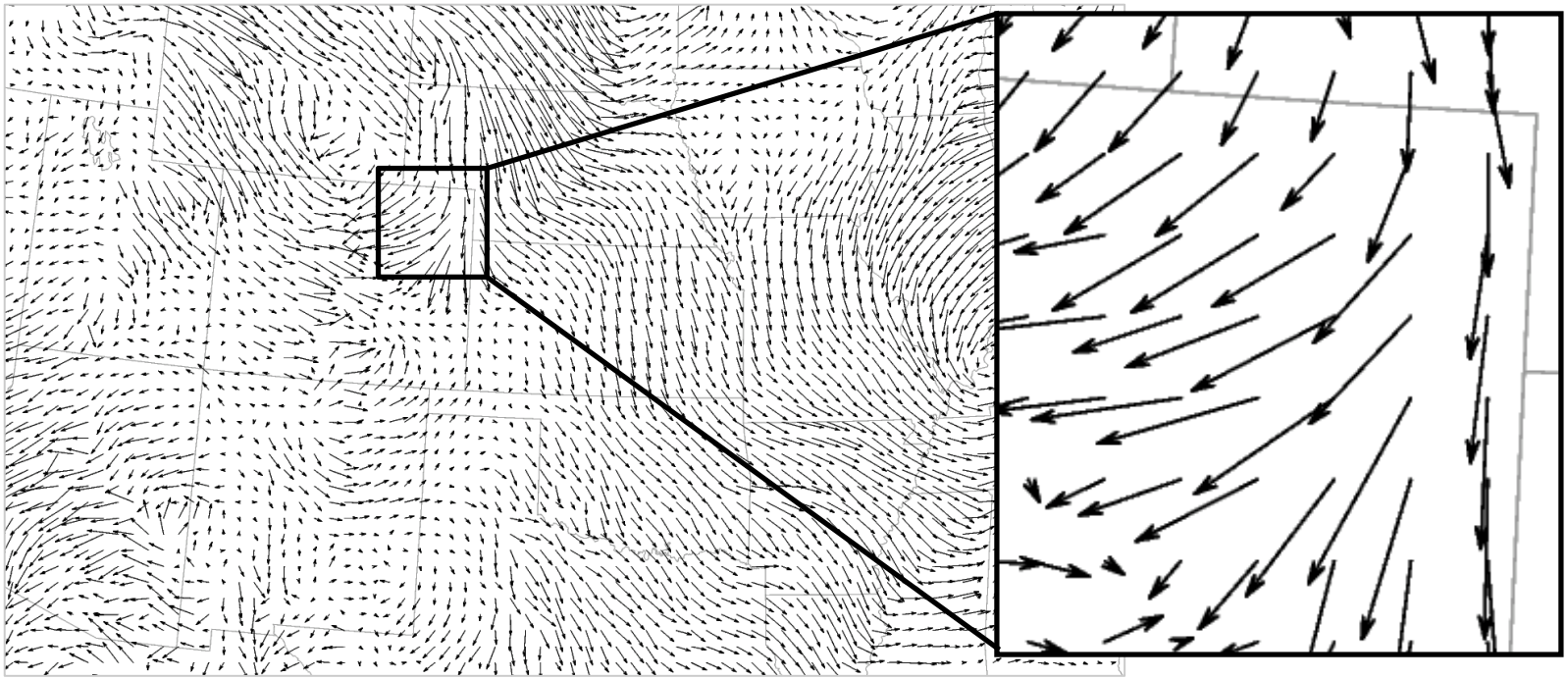


- ROW_MEAN_WDIR
  1. Derive wind directions and errors for each VL1L2 line
  2. Compute mean of errors



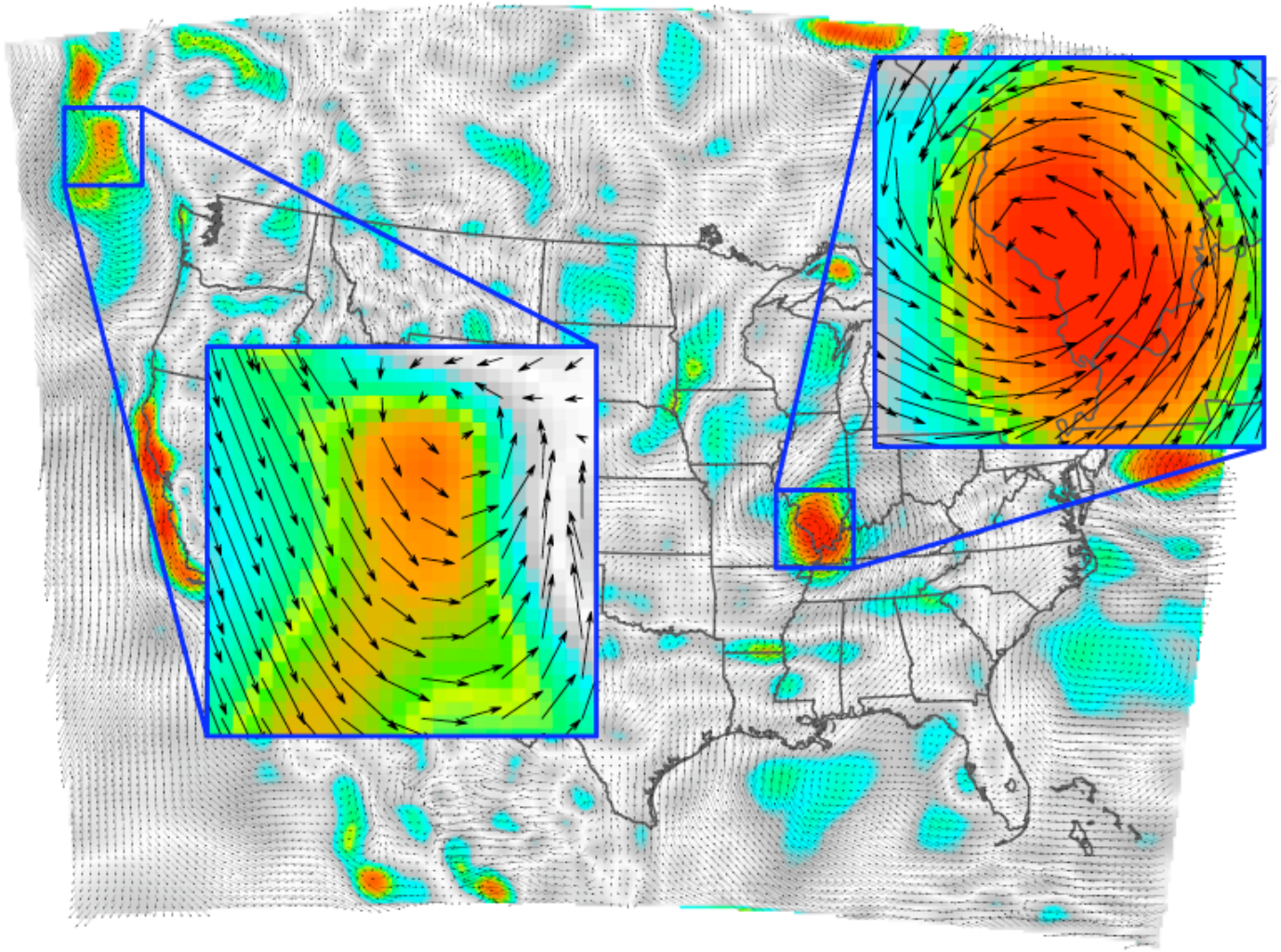
Wind Direction: Suggestions

- When aggregating, wind directions can cancel.
  1. Verify over regions with unimodal wind direction.
  2. Verify u and v components separately.



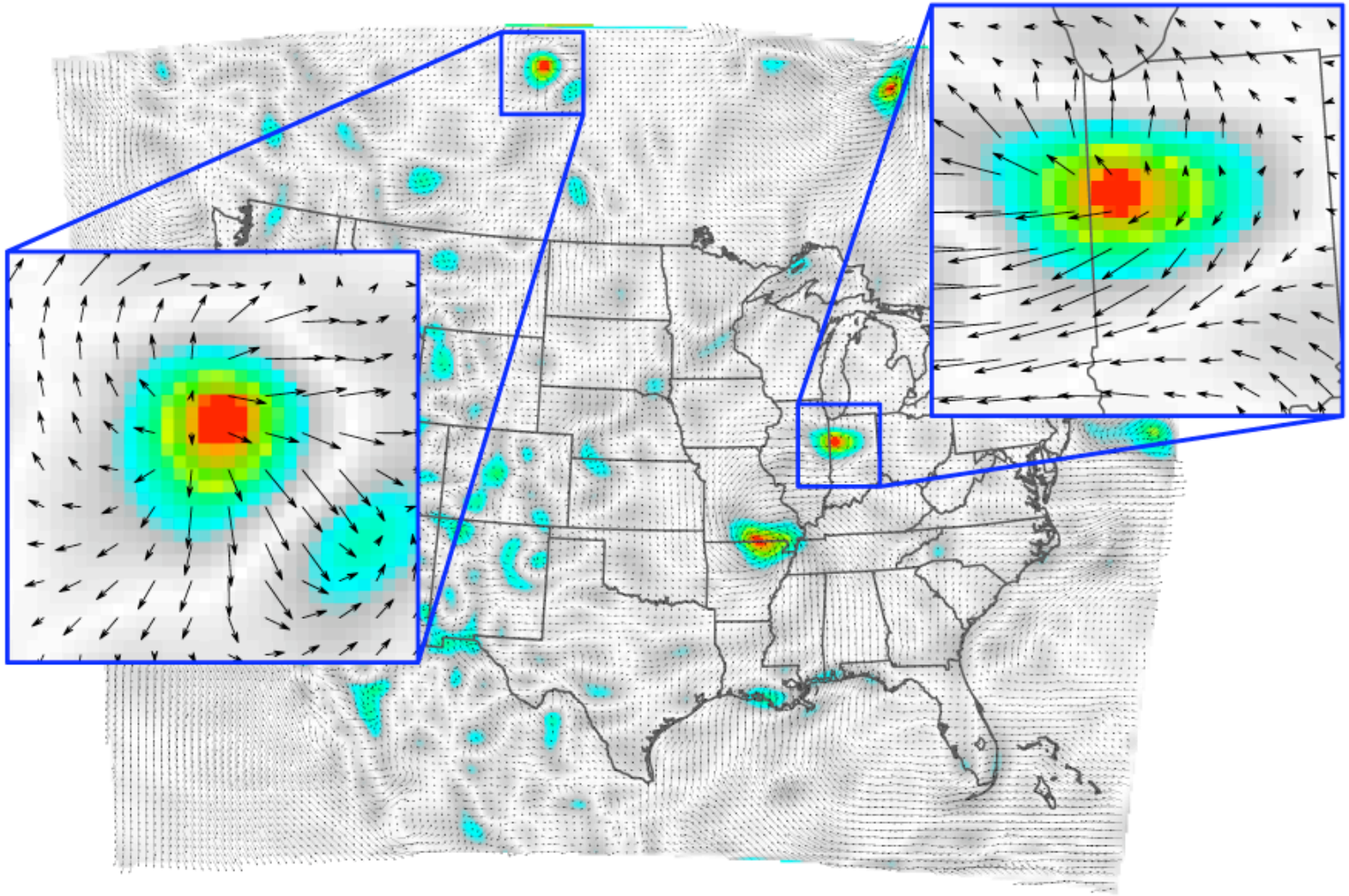
Verifying Winds: MODE

Divergence	$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}$
Curl	$\frac{\partial u}{\partial y} - \frac{\partial v}{\partial x}$
Speed	$\sqrt{u^2 + v^2}$



Curl

Valid Jul 12, 2005 12h Lead 00h

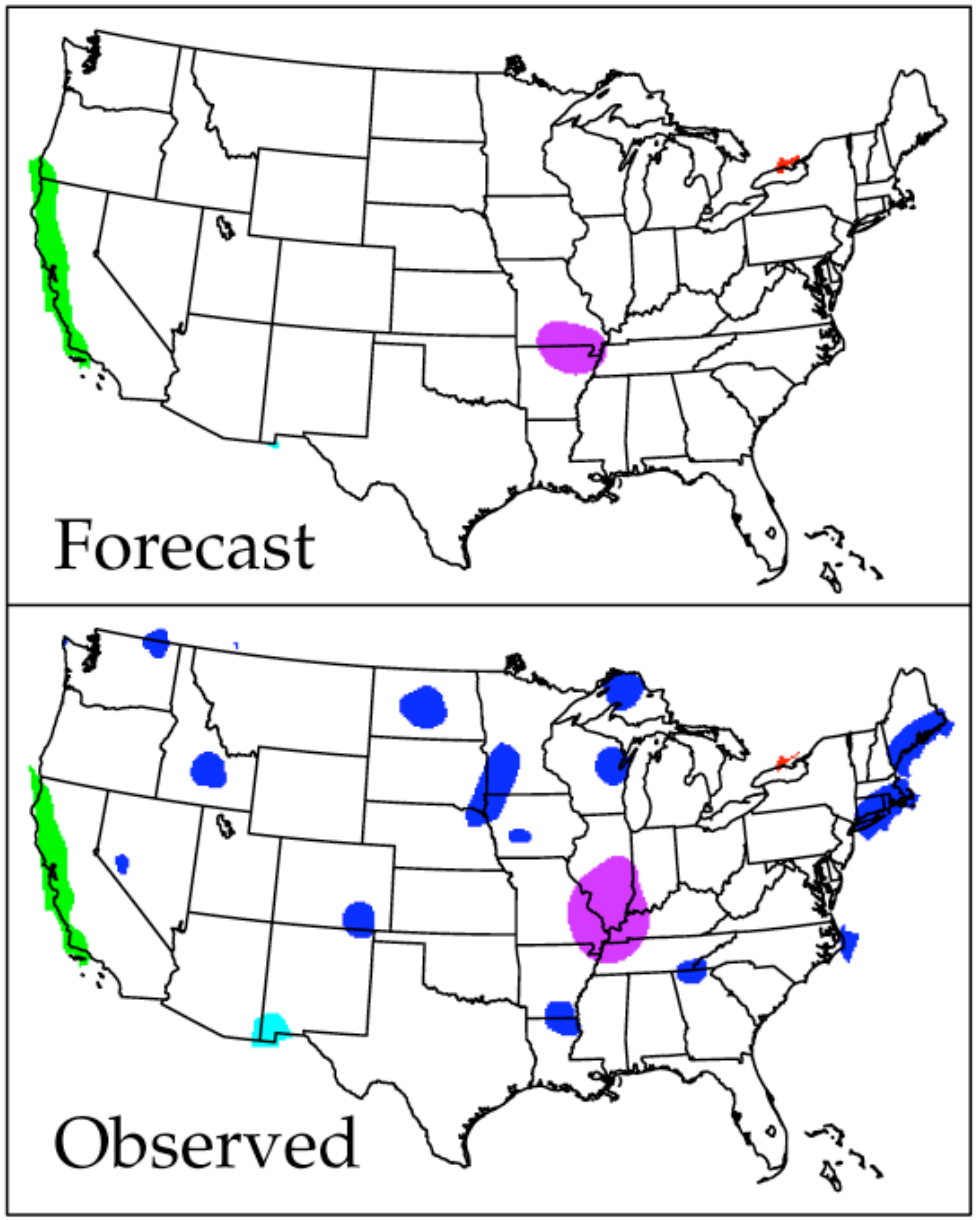


Divergence

Valid Jul 12, 2005 12h Lead 36h

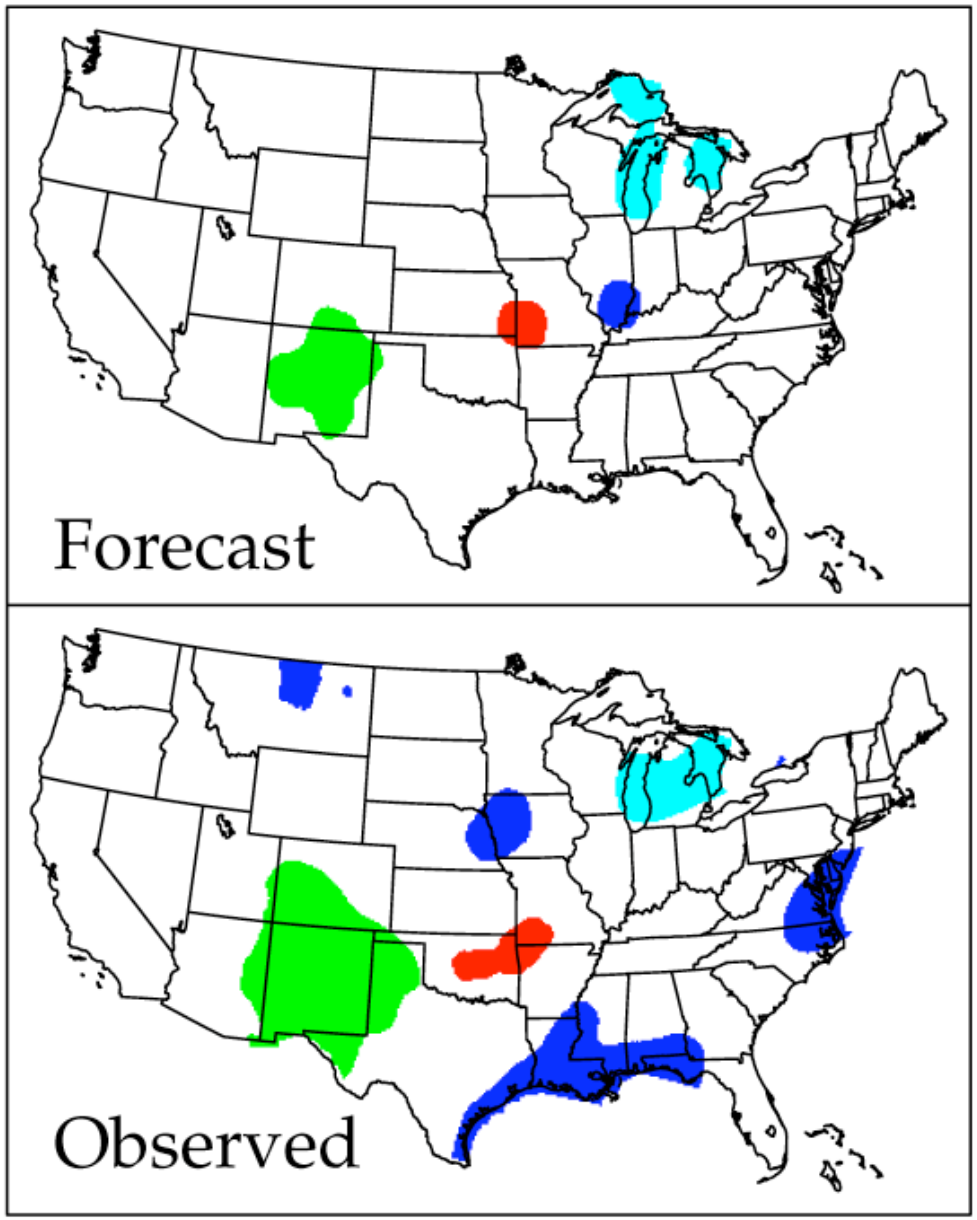
Curl Example

Jul 12, 2005



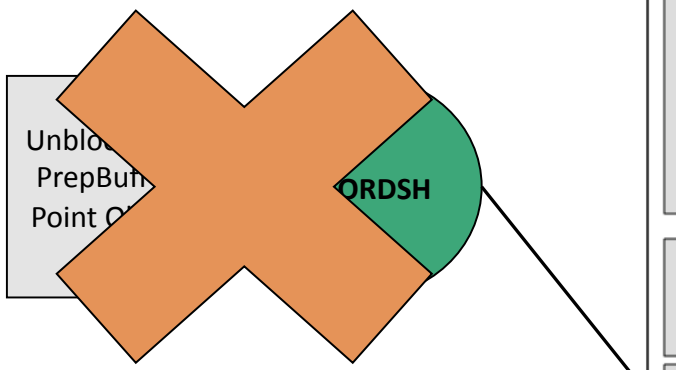
Speed Example

Feb 16, 2006

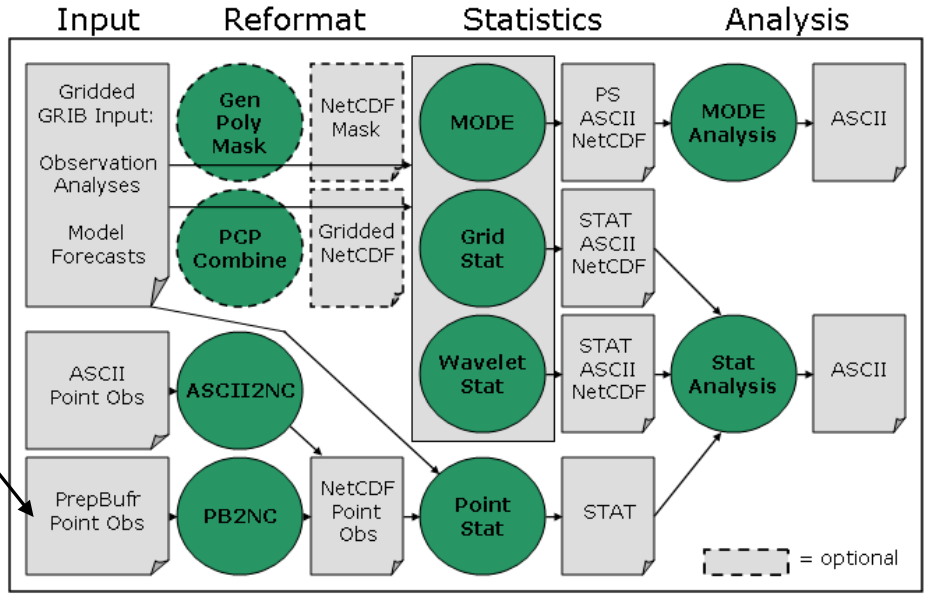


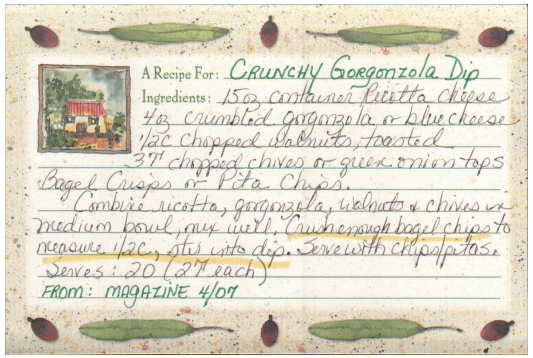
Fortran-Blocking

- No need to run the cwordsh utility on PrepBufr files
- Fortran-blocking performed within PB2NC tool

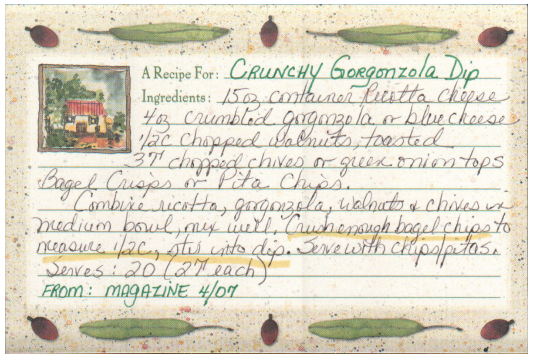


METv2.0 Flowchart





Future Work



- Major releases of MET once per year.
- Continued research and development of forecast evaluation methods and tools:
  - Verification of ensembles
  - Cloud verification
  - Use of satellite data (HDF5/NetCDF4)
  - Database/Display system for MET output (Example)
  - MODE time domain (DEMO)
- Sample plotting scripts on MET website (R code)
  - Please contribute your plotting scripts!

2009 Spring Experiment 9-panel Loop Page

http://www.dtcenter.org/plots/hwt/includes/loop_9panel_00z_20km30dBZ.php?starthr=00&date=20090514

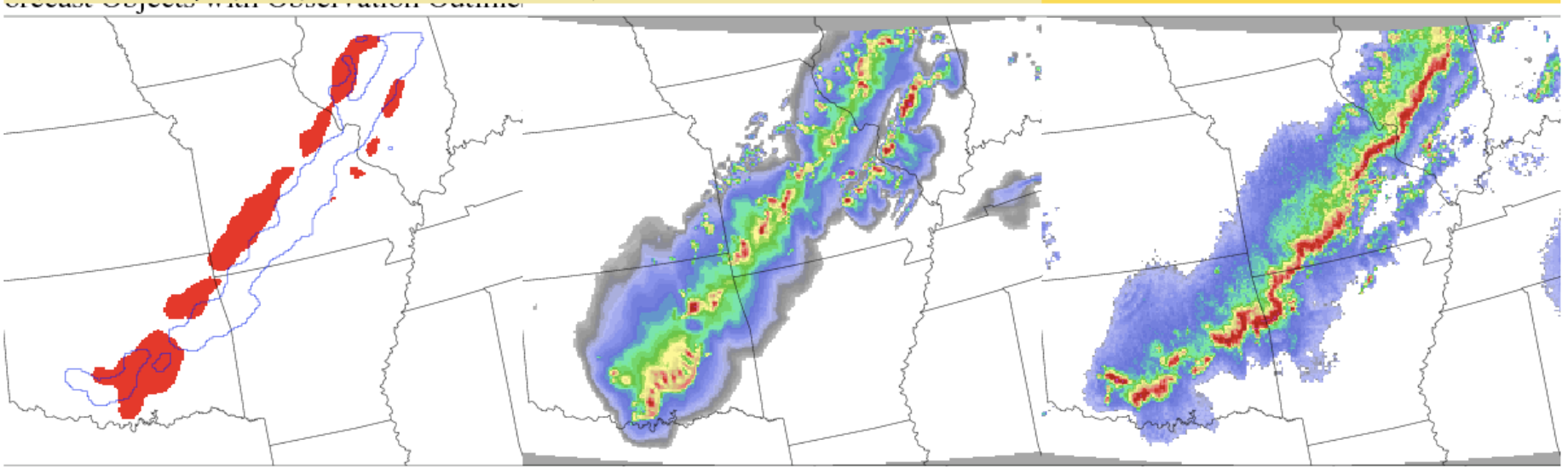
Gmail Comcast.net Sign In Google Maps Apple News (1836) Popular NOAA Wthr Library Broomfield

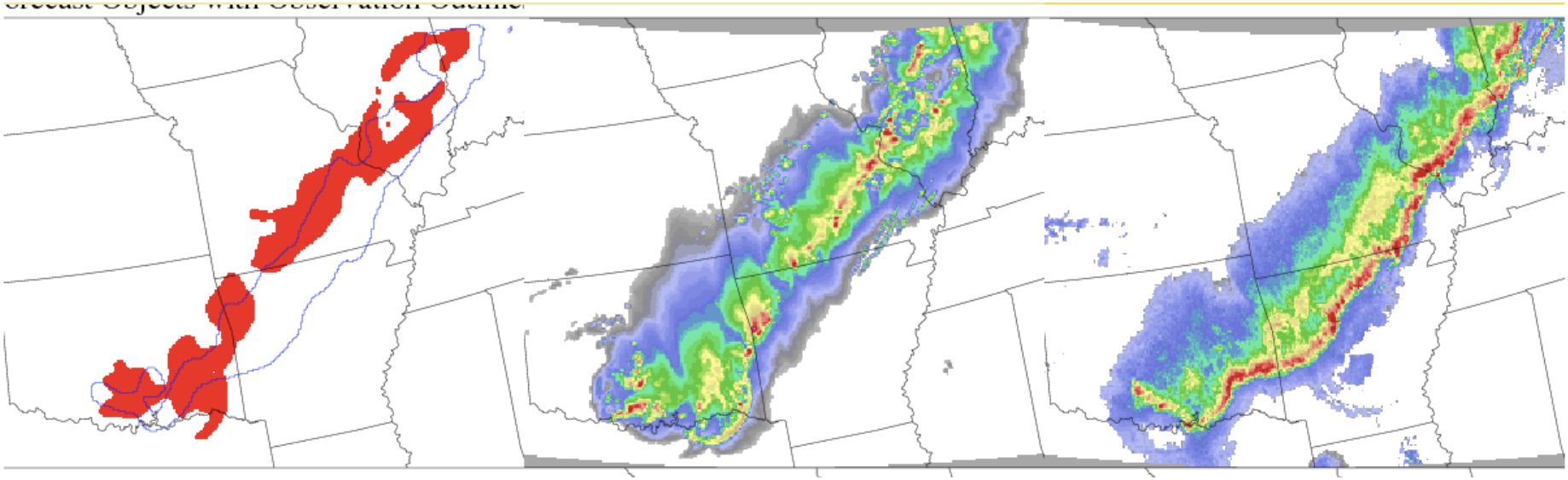
NSSL/SPC Spring Experiment 2009: Model Comparison Webpage

DTC MODE Model Comparison - 20090514

Frame 6 of 13 20090514/05Z Start Stop < > 10 img/sec Dwell Rock

capsc0 Fcst and Obs Objects (solid/line) REFC Valid: 20090514_0500 | capsc0 Fcst Field REFC Valid: 20090514_0500 | Obs Field REFC Valid: 20090514_0500





Further Details

- For more detail on the METv2.0 changes:
  - MET User's Guide
    - www.dtcenter.org/met/users/docs/overview.php
  - README within the MET release

Thank You

For more information:

http://www.dtcenter.org/met/users/

Questions for you

- What types (formats) of data do you use for verification?
- What is your biggest verification need?
- Do you use WRF ARW or NMM?