The Model Evaluation Tools (MET) for Objective Evaluation of Forecasts and Observational Datasets

Tara Jensen

Tressa Fowler

NCAR/RAL /JNT

and

Developmental Testbed Center (DTC)

Thank you to ECMWF and the NRL teams for Forecast Graphics

Thank You

MET
Model Evaluation Tools

Support for MET is provided by

the Developmental Testbed Center (DTC),

NOAA

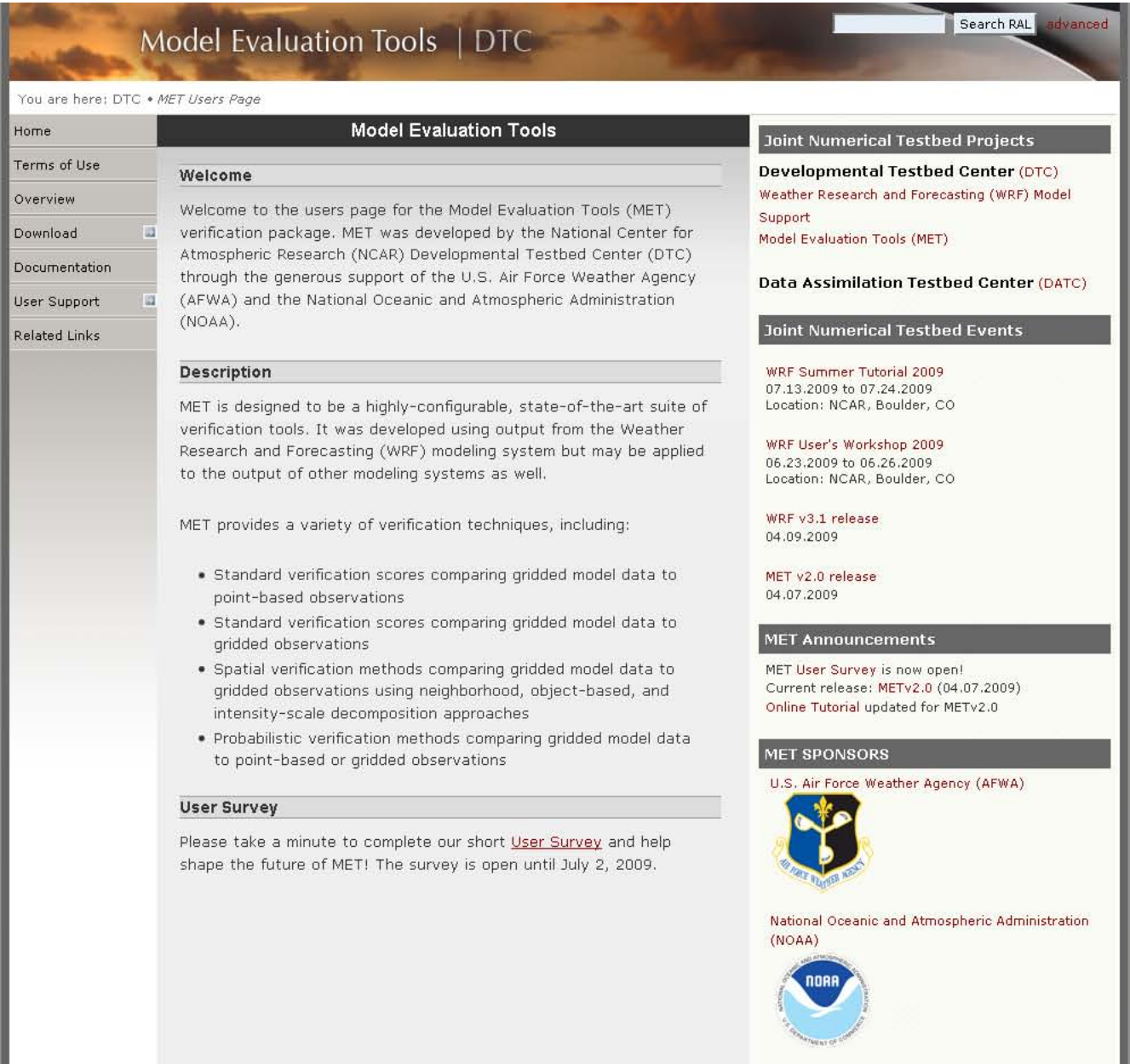
and

the Air Force Weather Agency (AFWA).

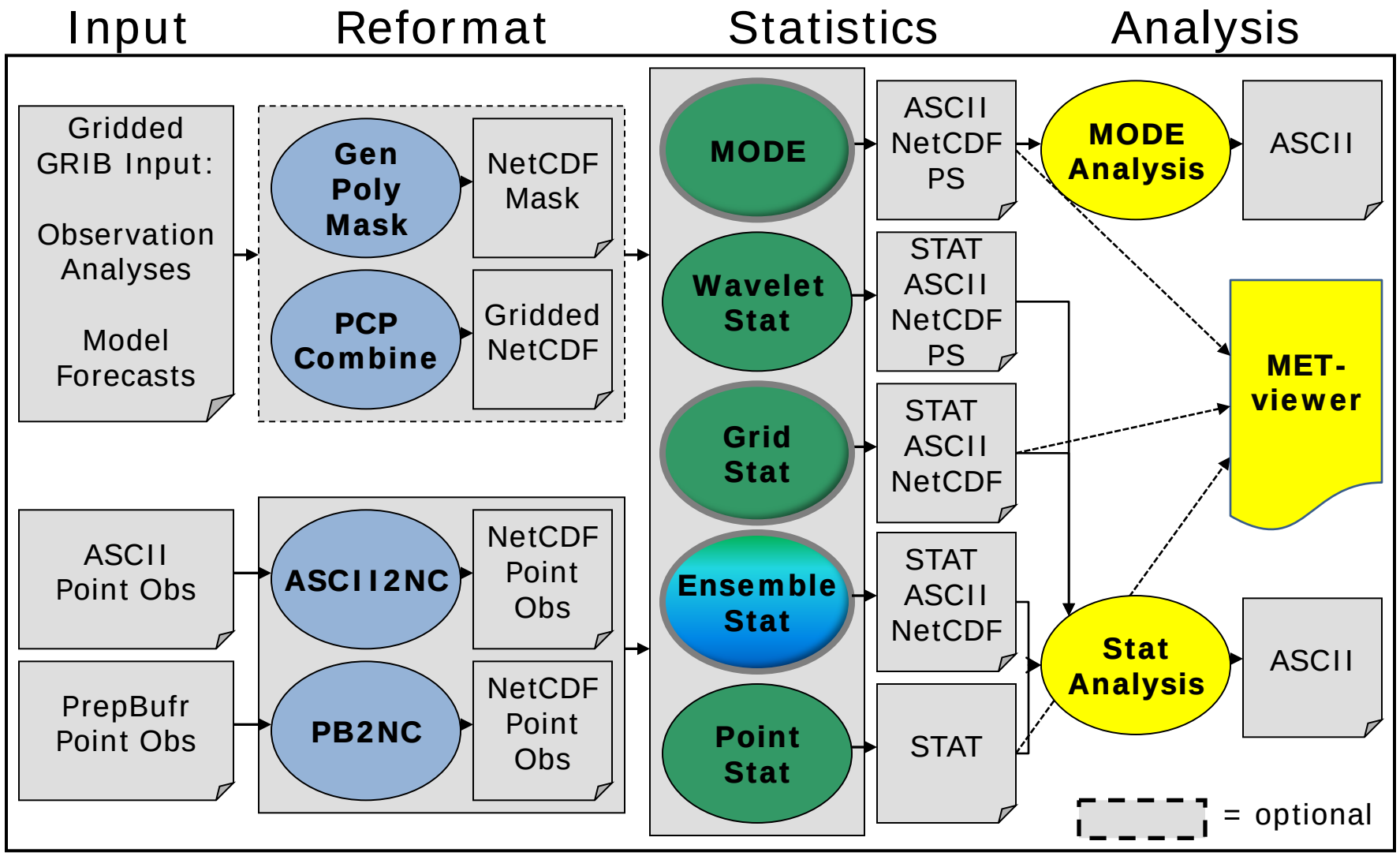
MET Package

- MET is community code supported by DTC that is free to download – registration required
- Download MET release and compile locally.
  - Register and download: **www.dtcenter.org/met/users**
- Language:
  - Primarily in C++ with calls to a Fortran library
- Supported Platforms and Compilers:
  1. Linux with GNU compilers
  2. Linux with Portland Group (PGI) compilers
  3. Linux with Intel compilers
  4. IBM machines with IBM compilers

www.dtcenter.org/met/users



MET v3.0 Tools



Data	MET Tool
Gridded Forecasts Gridded Observations	Grid Stat (traditional or neighborhood) Ensemble Stat Wavelet Stat MODE
Gridded Forecasts Point Observations	Point Stat Ensemble Stat

Typical MET Verification Process

- Determine Gridded Forecast Source
  - Get data into GRIB1 format
- Determine Observation Source
  - Point: Run through MET pre-processing to put into MET netCDF
  - Gridded: Put on same grid as forecast
- Set up config file for Statistical Tool
  - Typically tell it what:
    - Variable (based on GRIB ID and Level or Accumulation Interval)
    - Thresholds
    - What type of output you'd like – Categorical, Continuous, Probabilistic
- Run Statistical Tool
- Run Analysis Tool to filter data and aggregate then Plot OR
- Run METViewer to view filter, aggregate and plot thru UI
- Analyze

50%

10%

20%

20%

Traditional Statistics (Deterministic Forecasts)

Point Stat and Grid Stat

Point Stat: Overview

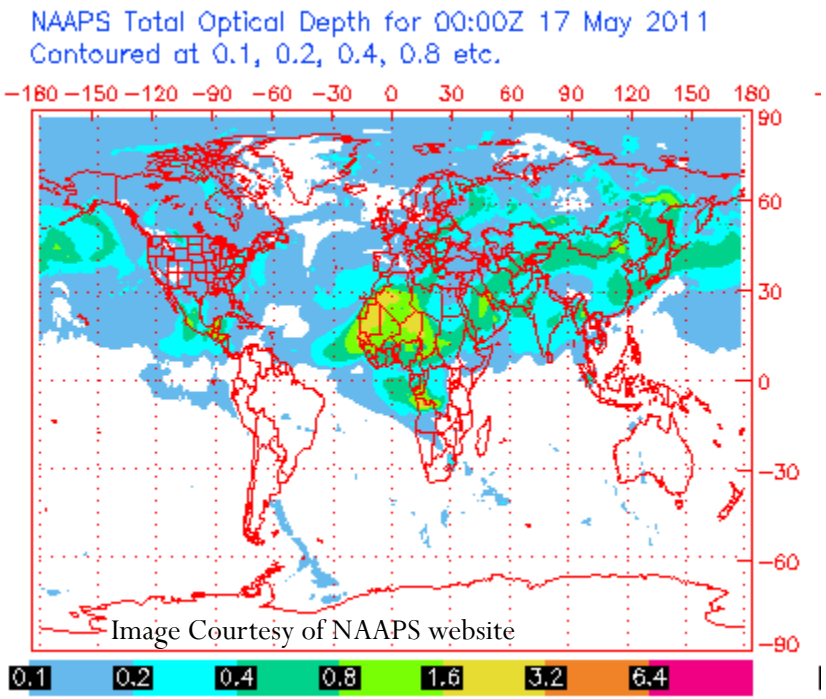
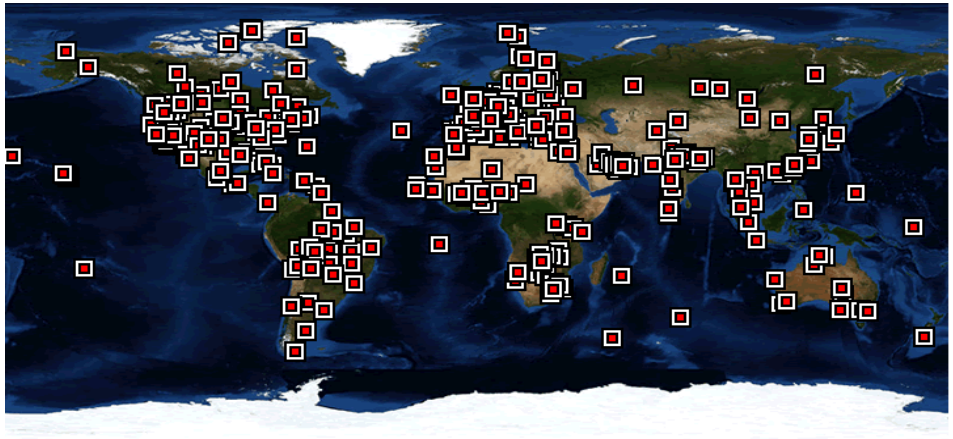


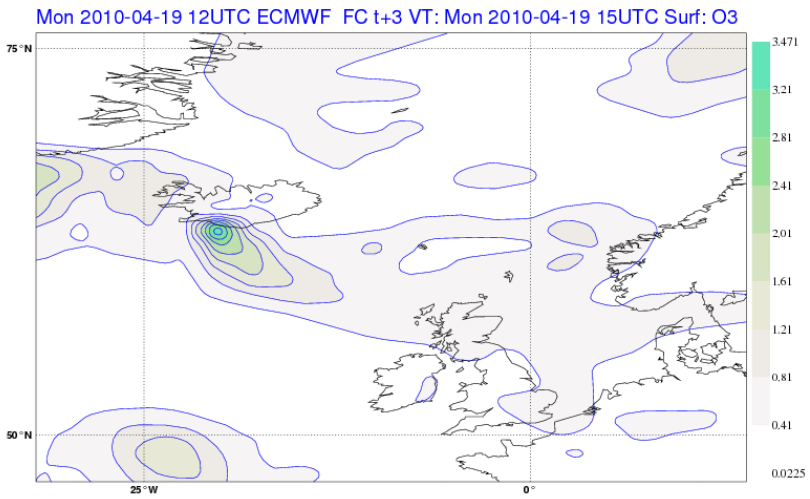
Image Courtesy of NAAPS website

- 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.

- 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.


Aeronet Sites – Image Courtesy of Aeronet website

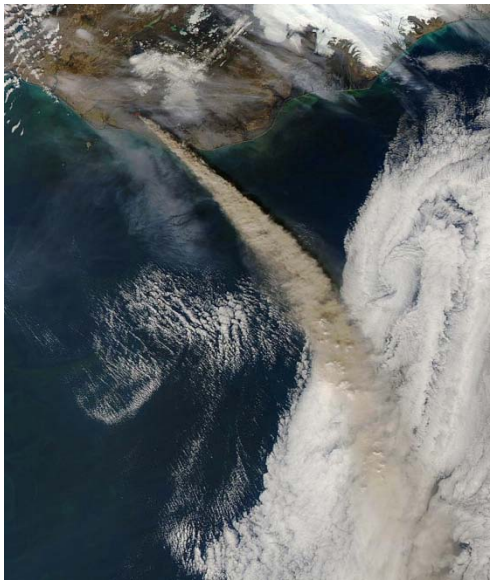
Grid Stat: Overview



ECMWF Forecast courtesy of Angela Benedetti

- 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.



Aqua (NASA MODIS/AFP/Getty Images)
2010 May 8

Details on Categorical and Continuous Statistics

Appendix C of MET Documentation

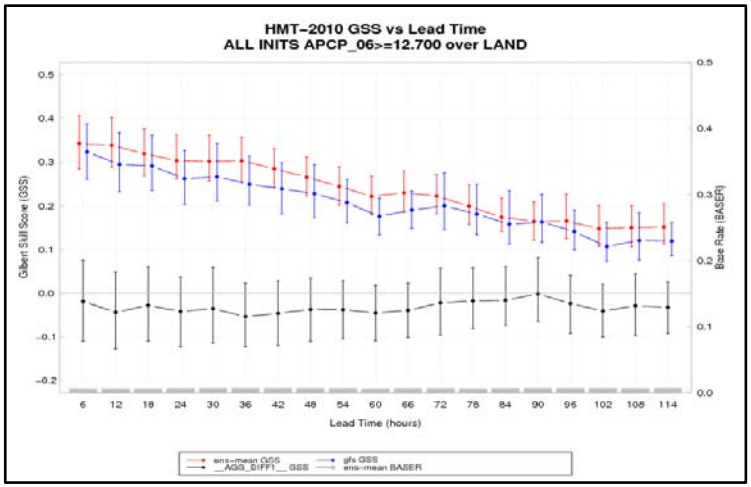
- http://www.dtcenter.org/met/users/docs/overview.php

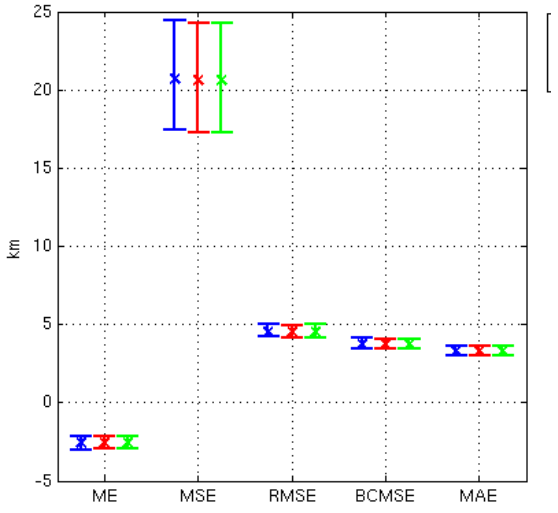
Joint Group on Forecast Verification Page

- http://www.cawcr.gov.au/projects/verification/

Confidence Intervals in MET

- Normal Approximation
  - Calculated for all statistics for which this is appropriate
- Bootstrapping
  - Can be turned on in config file
  - Number of repetitions are user defined
- Limits output to ascii output

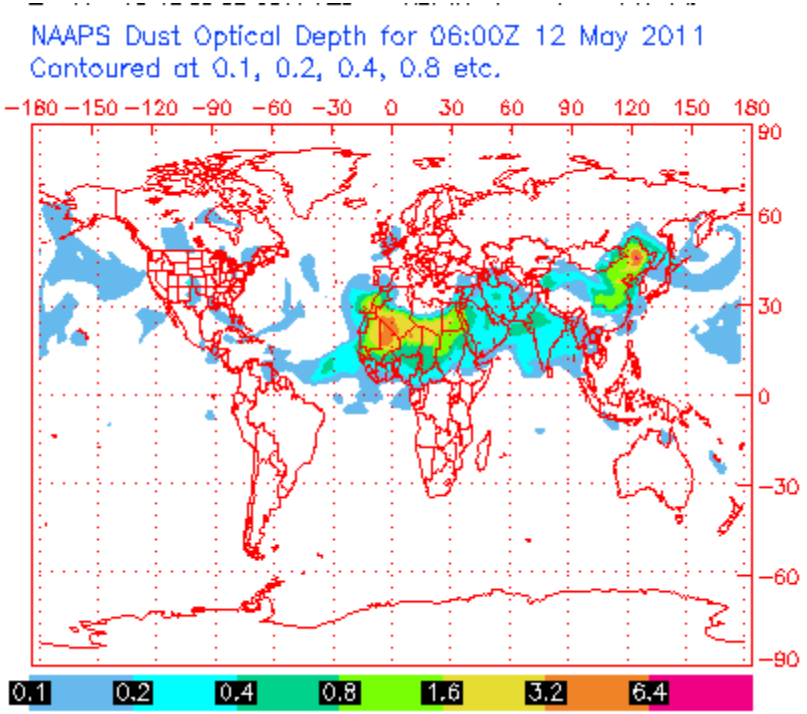




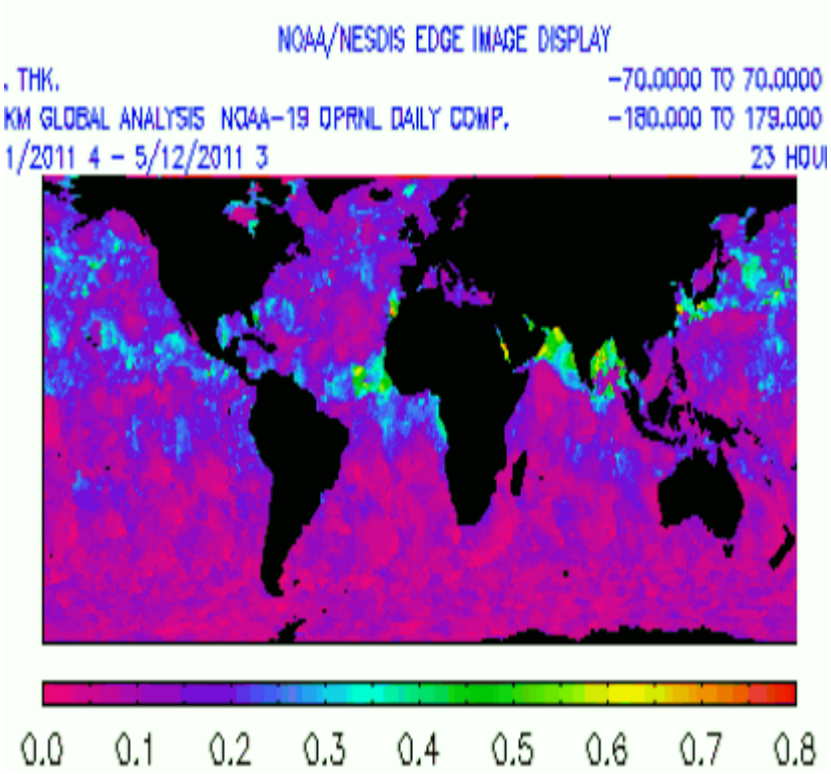
Object Based Evaluation

Method for Object Based Diagnostic Evaluation (MODE)

Typical situation



Forecast



Observation

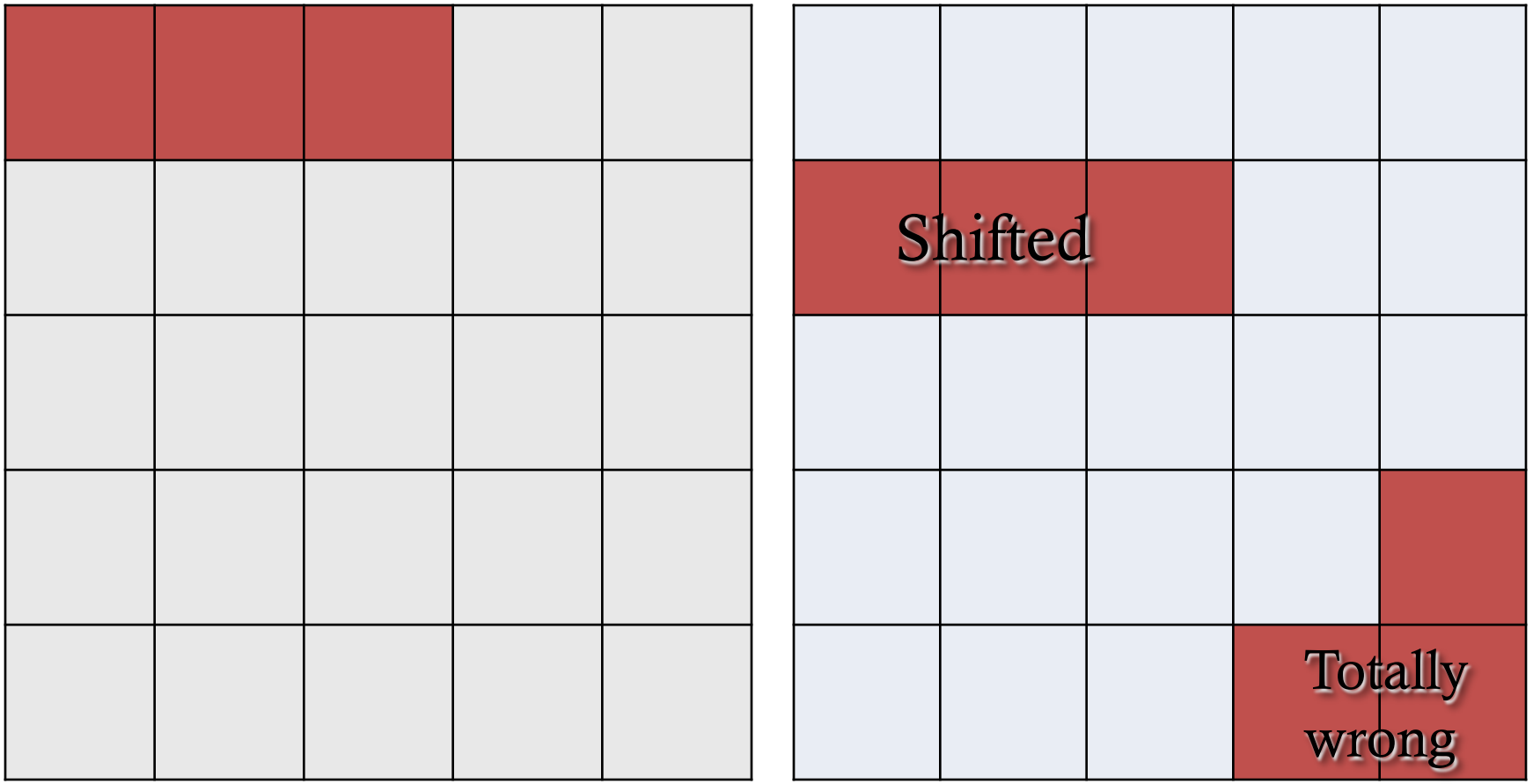
Object verification
is more like what
humans
do.

Objects recognize
the spatial relationship
between points.

Simple example

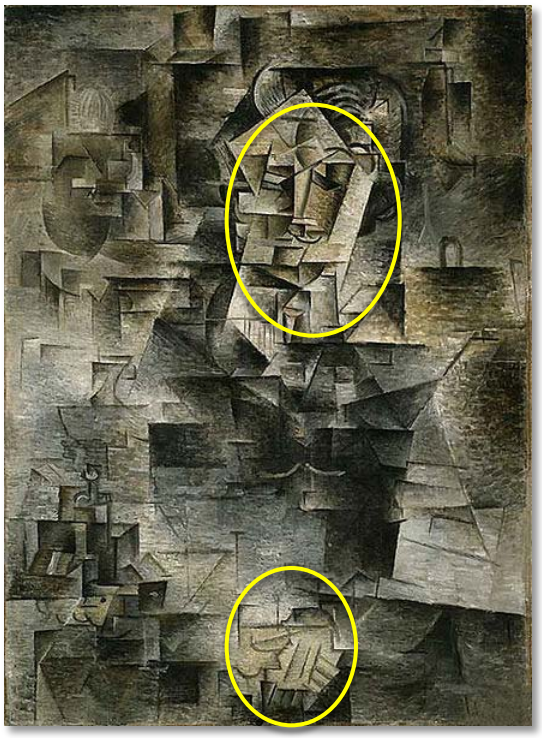
Observed

Forecast

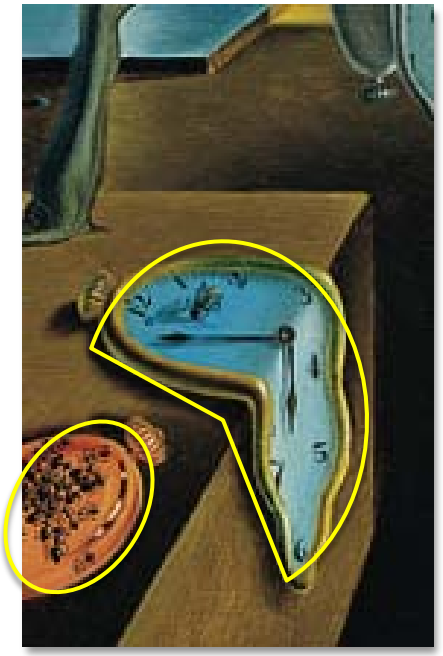


REAL - observed

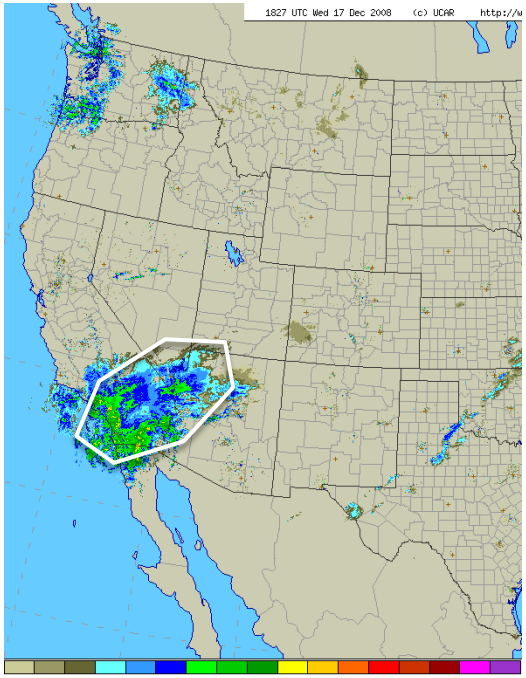
Forecast 1 –
Distorted view of reality



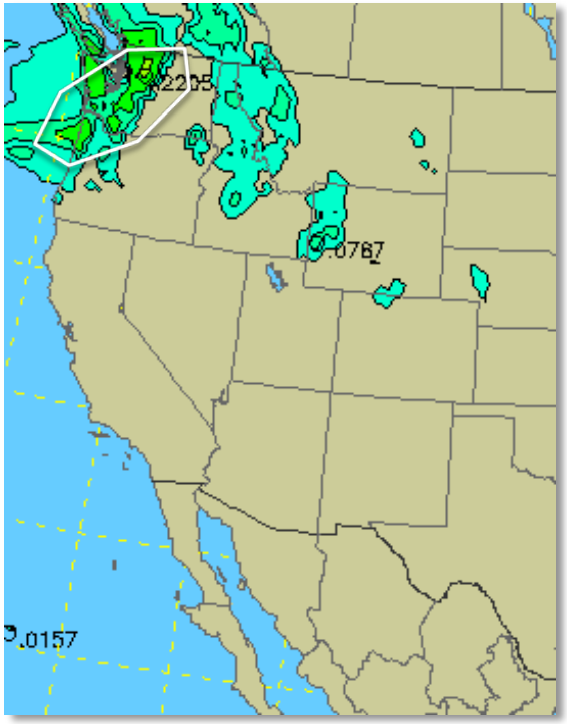
Forecast 2 –
Another distorted
view of reality



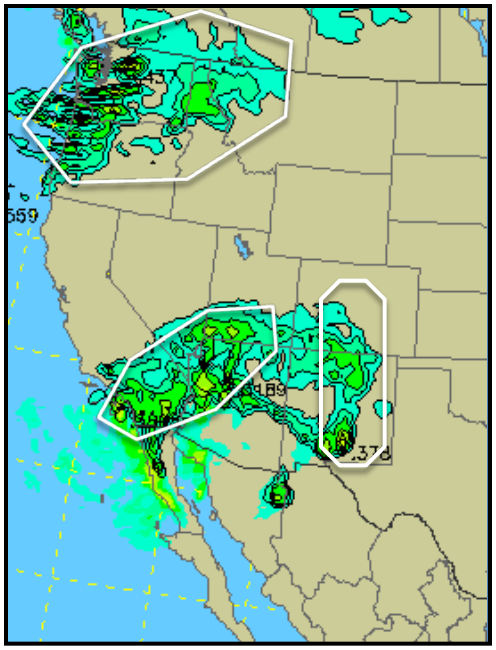
REAL - observed



Forecast 1 –
Distorted view of reality



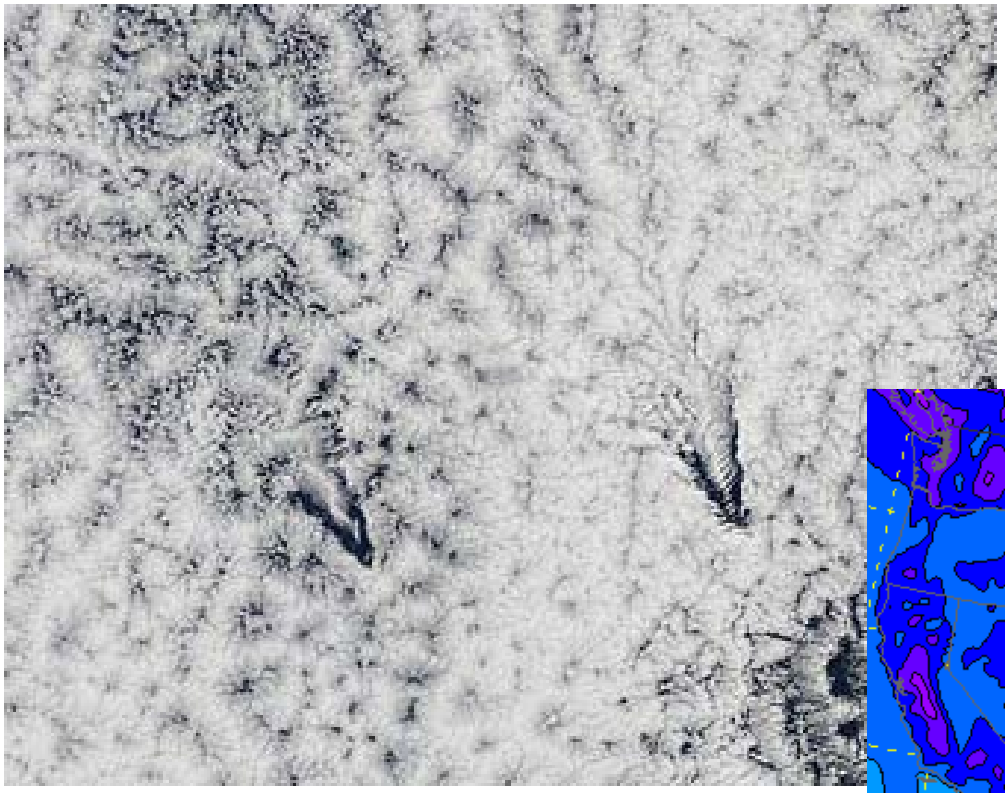
Forecast 2 –
Another distorted
view of reality



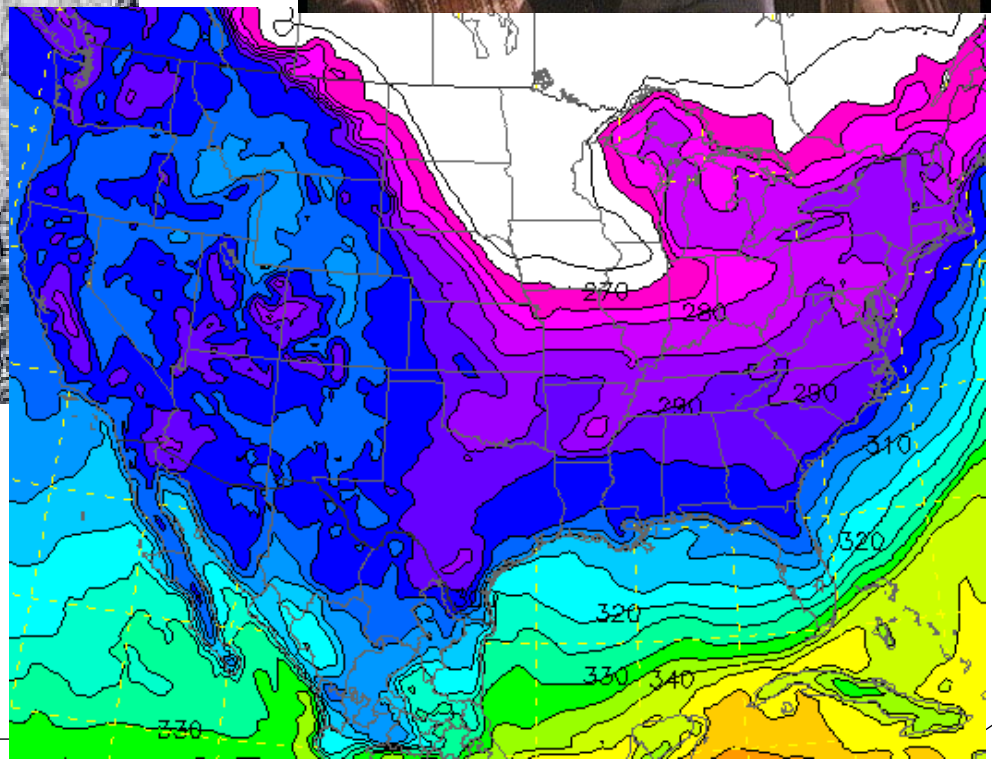
Comparing objects can tell you things about your forecast like . . .

This:	Instead of this:
30% Too Big (area ratio=1.3)	POD = 0.35
Shifted west 1 km (centroid distance = 1km)	FAR = 0.7235
Rotated 15° (angle diff = 15%)	CSI = 0.1587
Peak Rain 1/2" too much (diff in 90th percentile of intensities = 0.5)	

Verifying with objects doesn't always make sense . . .

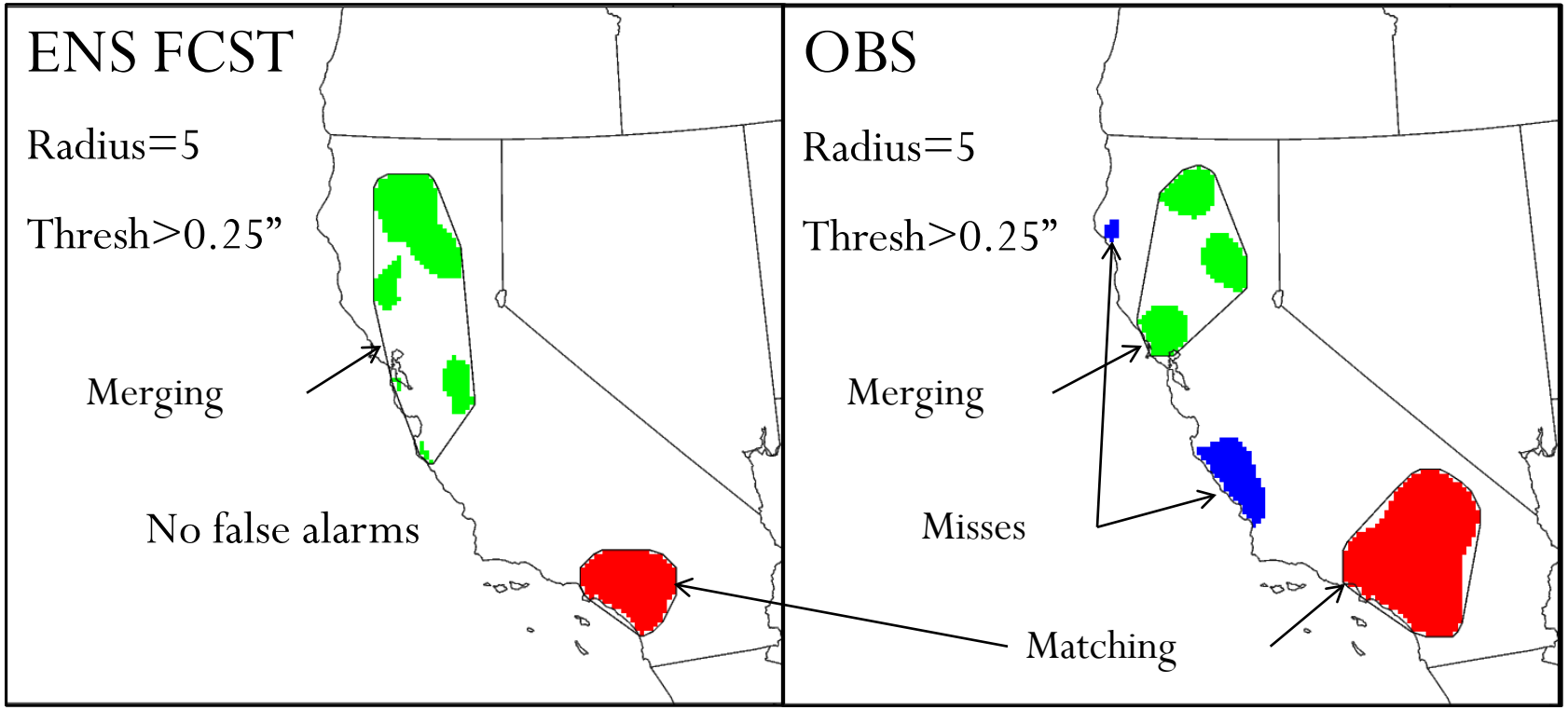


NASA/MODIS - **May 9, 2011 – Ship-wave-shape wave clouds induced by Juan Fernandez Islands, off Chile**

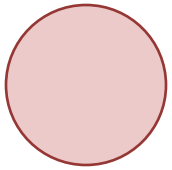


Object Oriented Method: MODE

From HMT: 18hr forecast valid at 2010-02-06 18Z

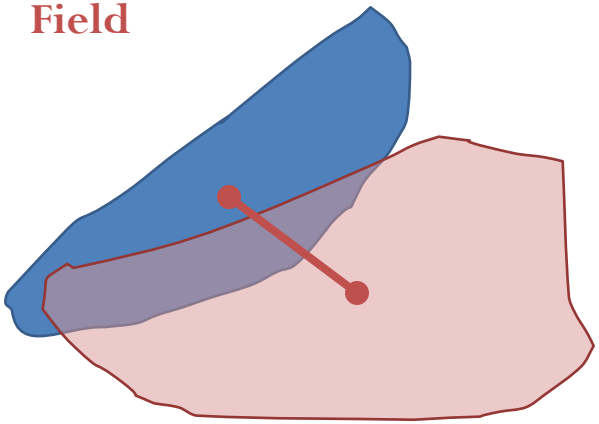


Use of Pair Attributes defined by MODE

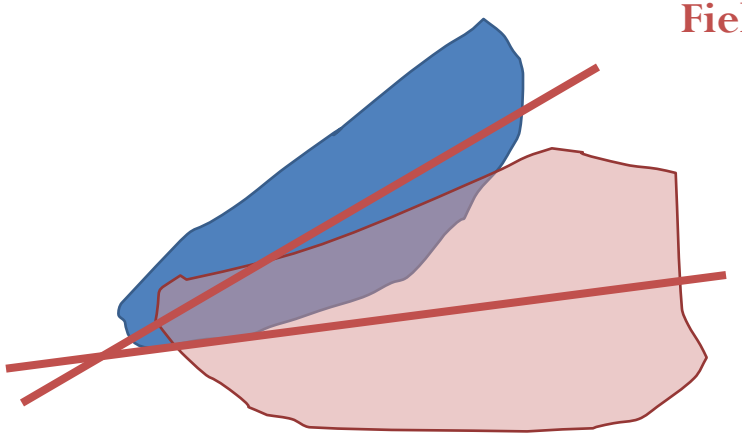


Forecast Field

Observed Field

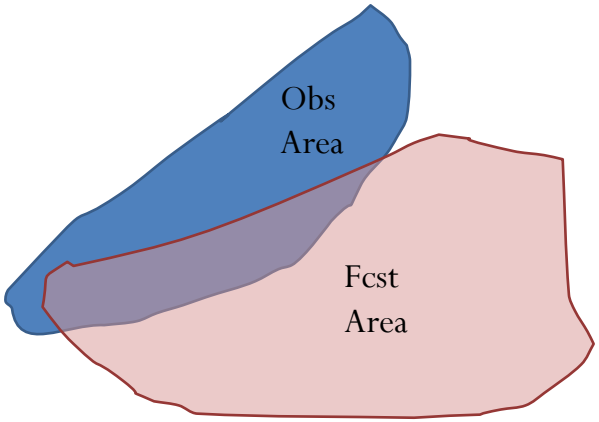


Centroid Distance: Provides a quantitative sense of spatial displacement of forecast.
Small is good



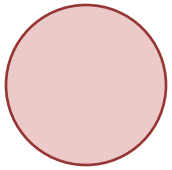
Axis Angle: For non-circular objects – gives measure of orientation errors.
Small is good

Area Ratio =
Fcst Area
Obs Area



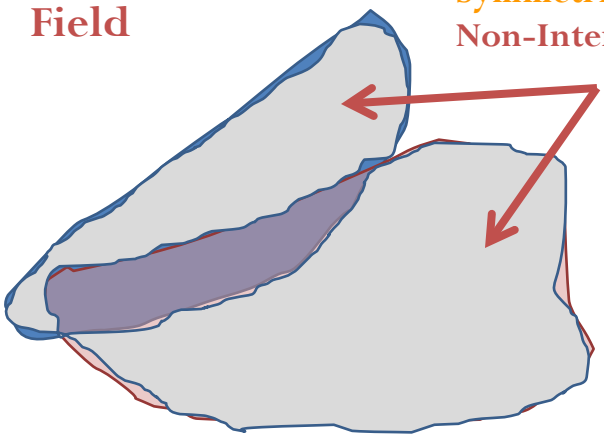
Area Ratio: Provides an objective measure of whether there is an over- or under-prediction of areal extent of forecast.
Close to 1 is good

Use of Pair Attributes defined by MODE



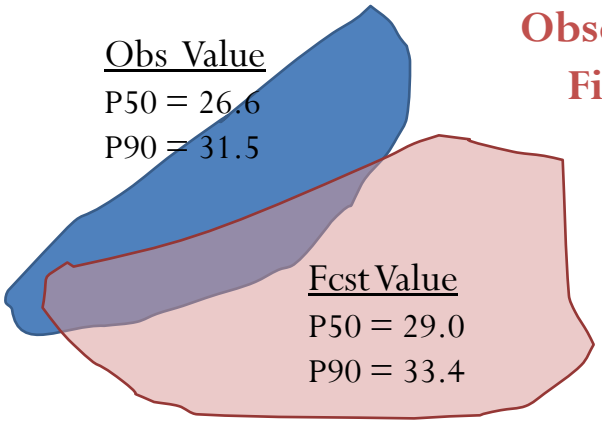
Forecast Field

Symmetric Difference: Non-Intersecting Area

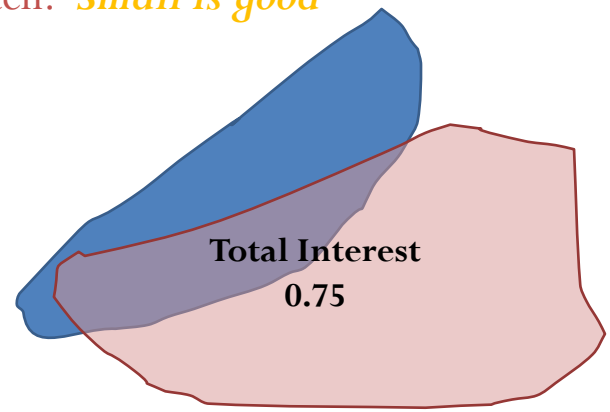


Symmetric Diff: May be a good summary statistic for how well Forecast and Observed objects match. *Small is good*

Observed Field

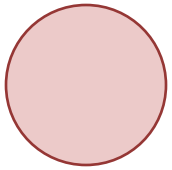


P50 | P90 Int: Provides objective measures of Median (50th percentile) and near-Peak (90th percentile) intensities found in objects. *Ratio close To 1 is good*



Total Interest: Summary statistic derived from fuzzy logic engine with user-defined Interest Maps for all these attributes plus some others. *Close to 1 is good*

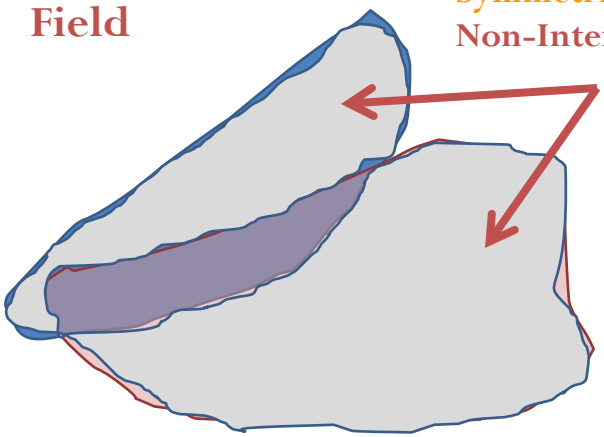
Use of Pair Attributes defined by MODE



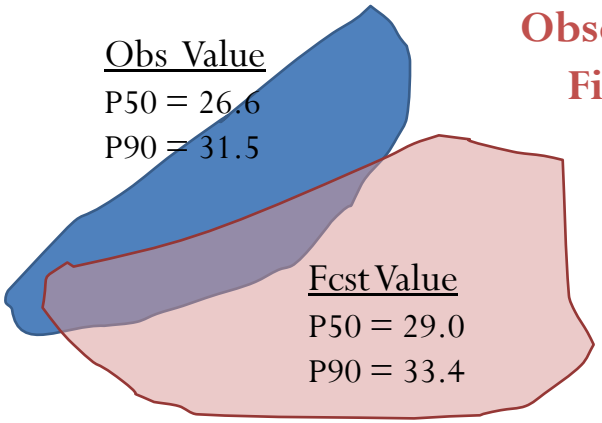
Forecast Field

Observed Field

Symmetric Difference: Non-Intersecting Area



Symmetric Diff: May be a good summary statistic for how well Forecast and Observed objects match. *Small is good*



P50 | P90 Int: Provides objective measures of Median (50th percentile) and near-Peak (90th percentile) intensities found in objects. *Ratio close To 1 is good*

Angle_diff & **Sym_diff** less so **Total Int.** higher

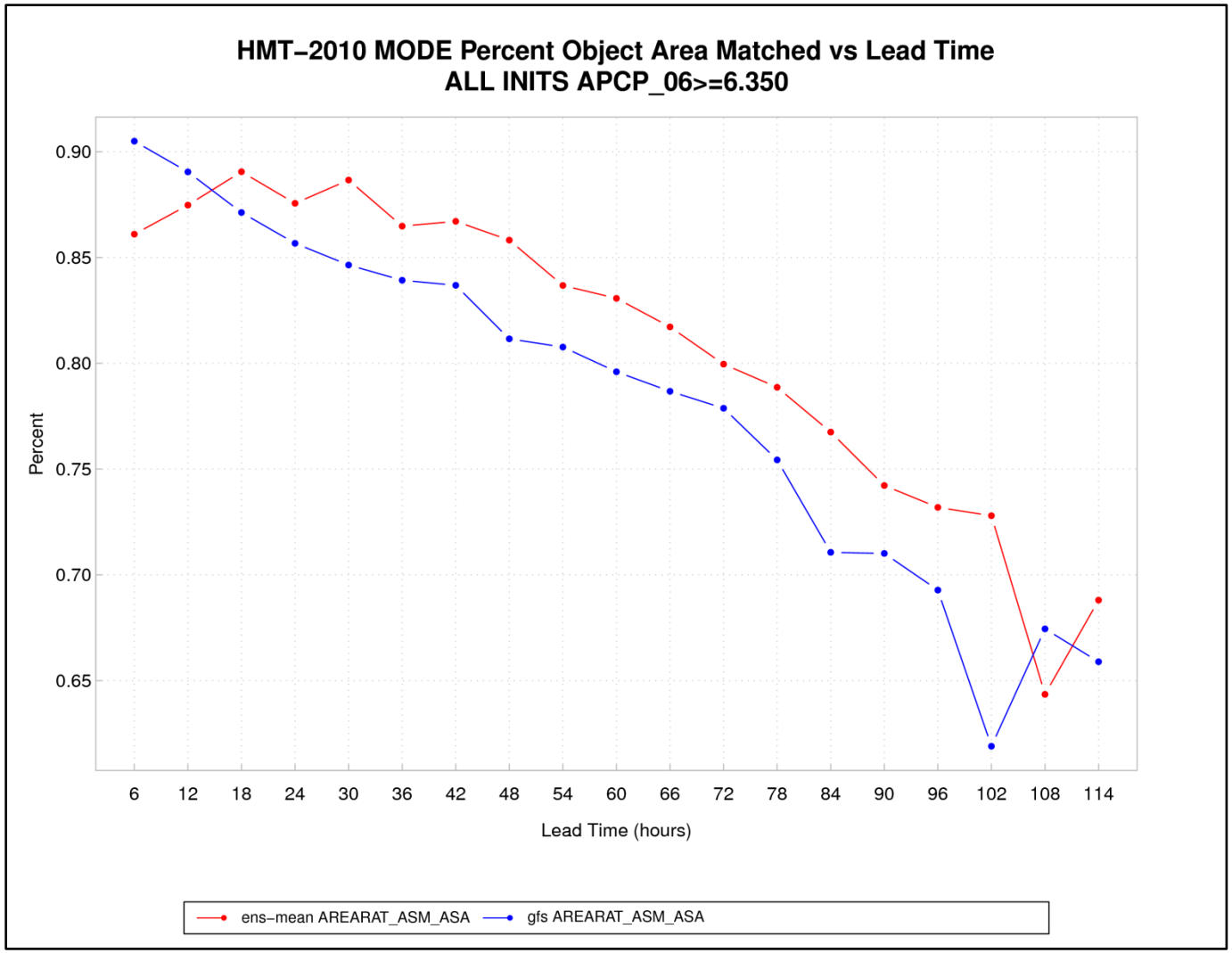
Total Interest: Summary statistic derived from fuzzy logic engine with user-defined Interest Maps for all these attributes plus some others. *Close to 1 is good*

MODE: Percent Area Matched

06hr
accumulations

- Ensemble Mean matches a higher percentage of total object area than GFS

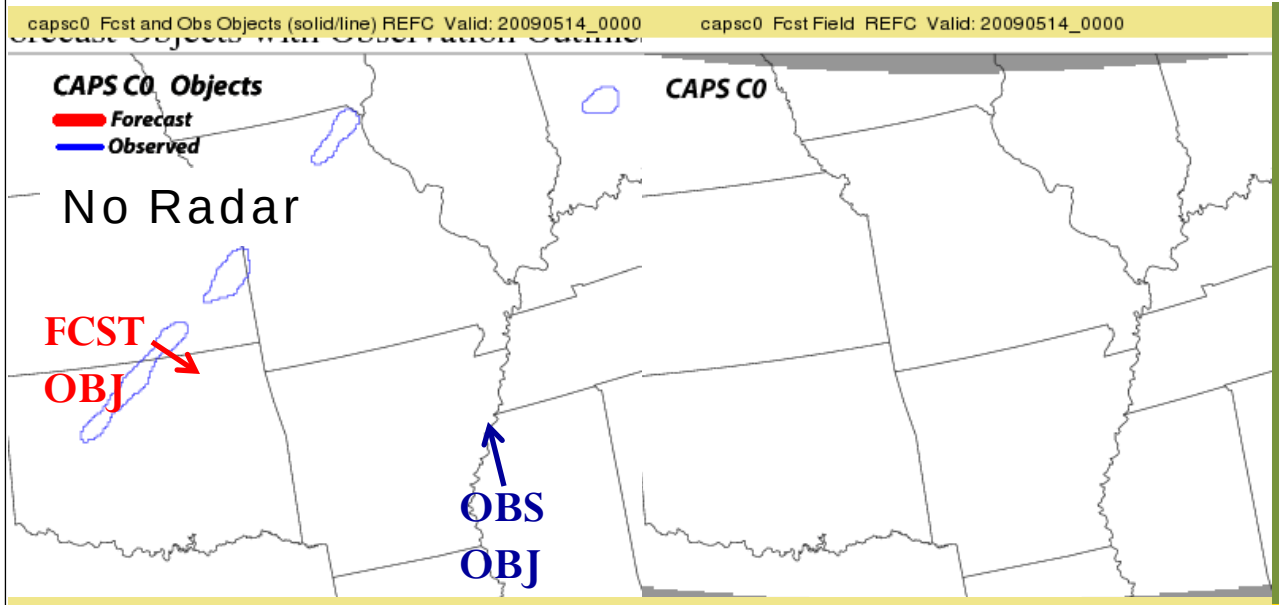
From HMT



MODE has been used to evaluate

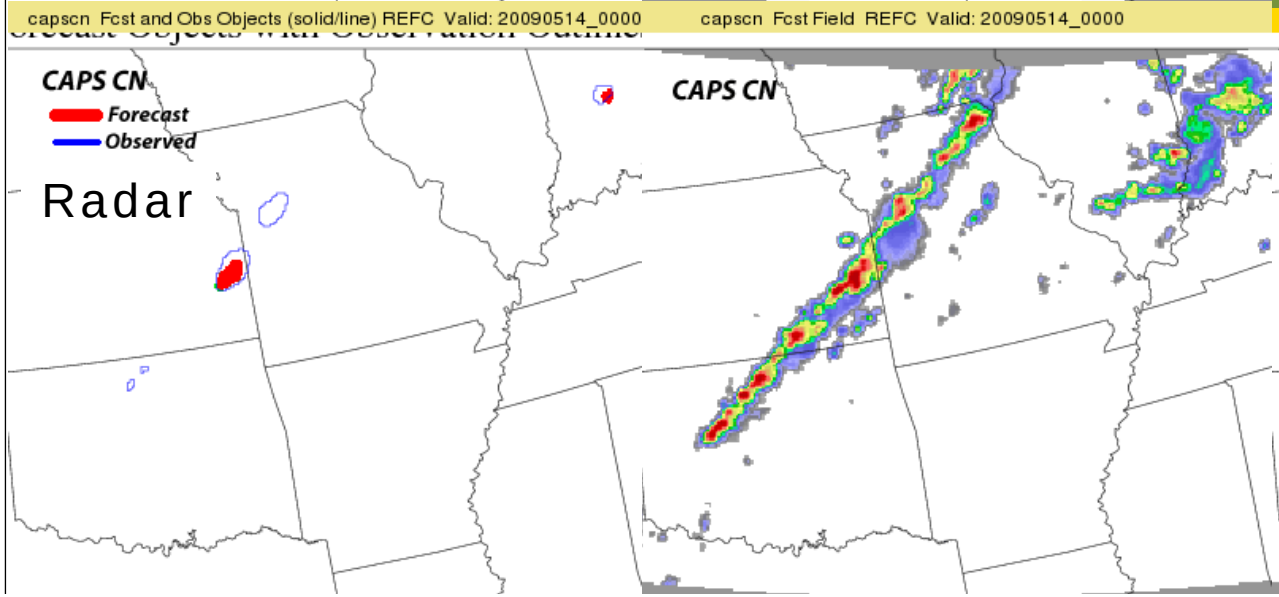
- Precipitation
  - Standard Accumulation Intervals
  - Probability Fields
- Reflectivity
  - Composite
  - Radar Echo Top
- Precipitable Water / IWV
- A-Train 2-D vertical curtain of satellite fields
- World-Wide Merged Cloud Analysis (WWMCA)

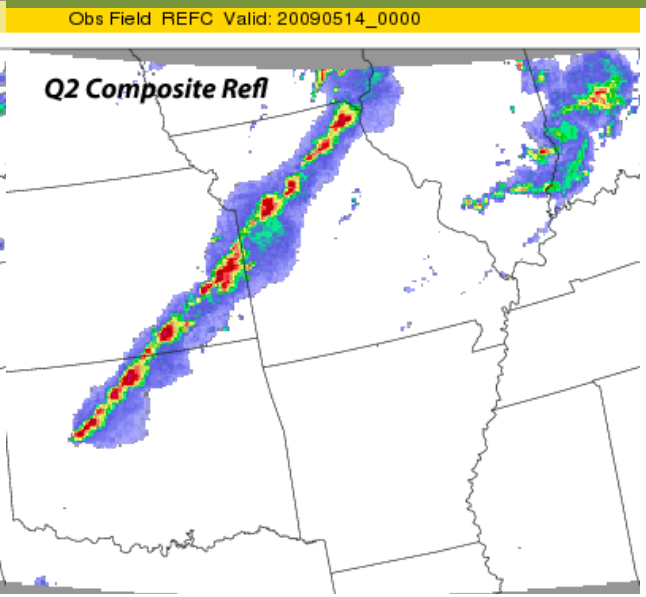
14 May 2009 Init: 00 UTC Spatial Thresh: 30dBZ



High Impact Display allowed HWT to identify phase and convective mode errors

Still working on how to quantify objectively





Objects

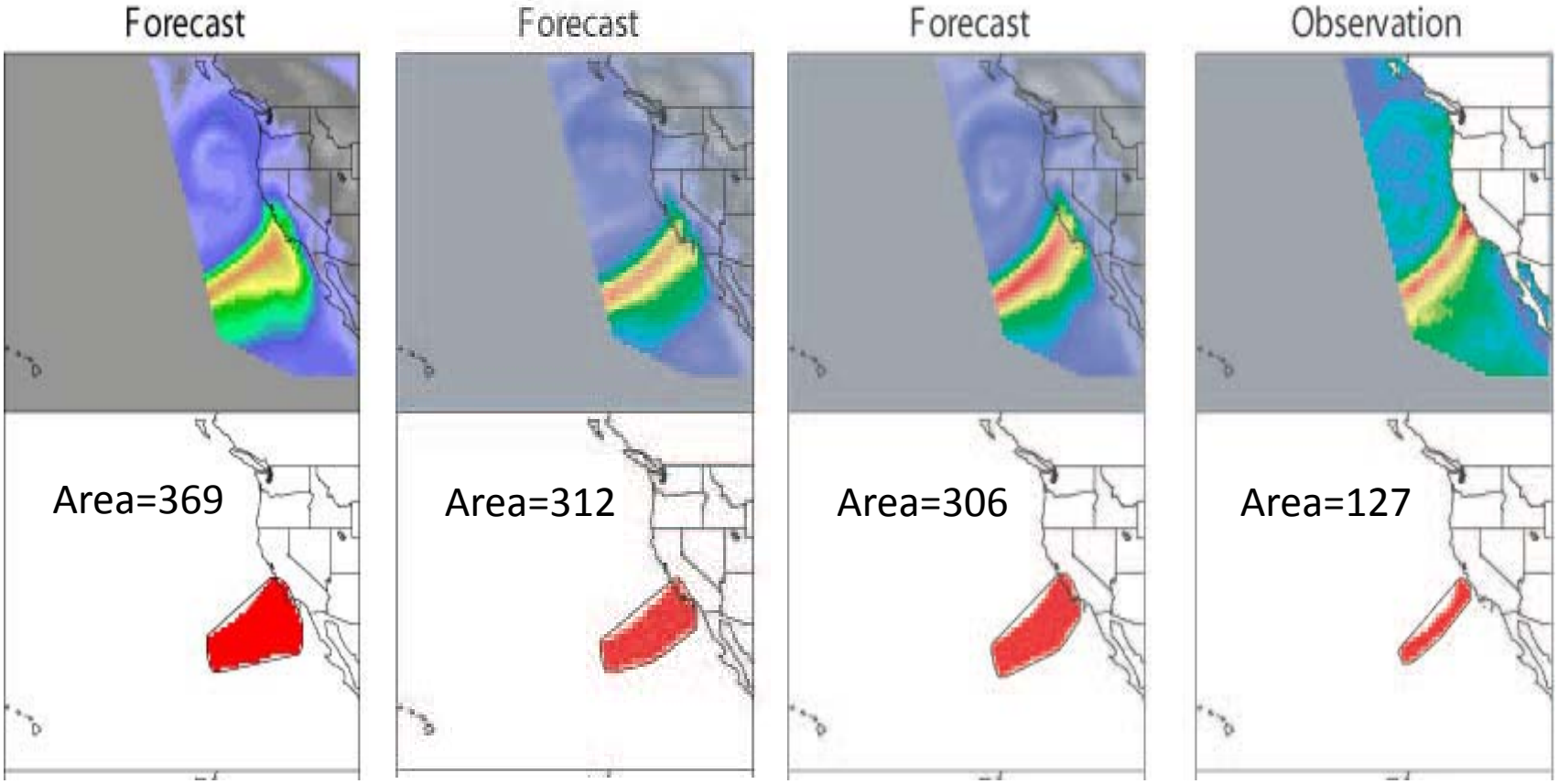
Forecast Field

Observed Field

Example: PWAT / IWV

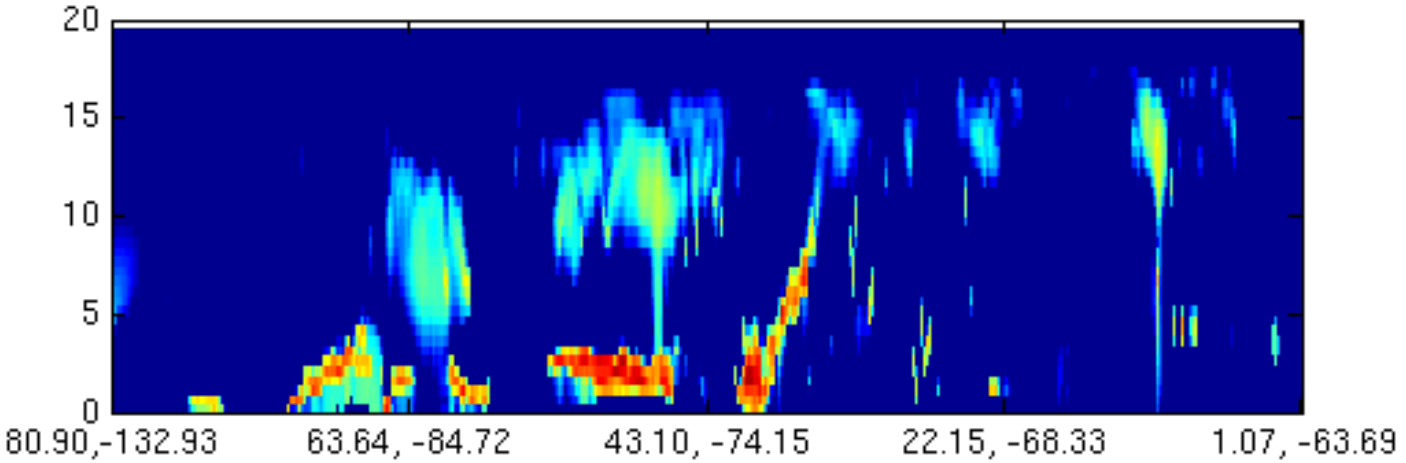
GFS Precipitable Water

SSMI Integrated Water Vapor

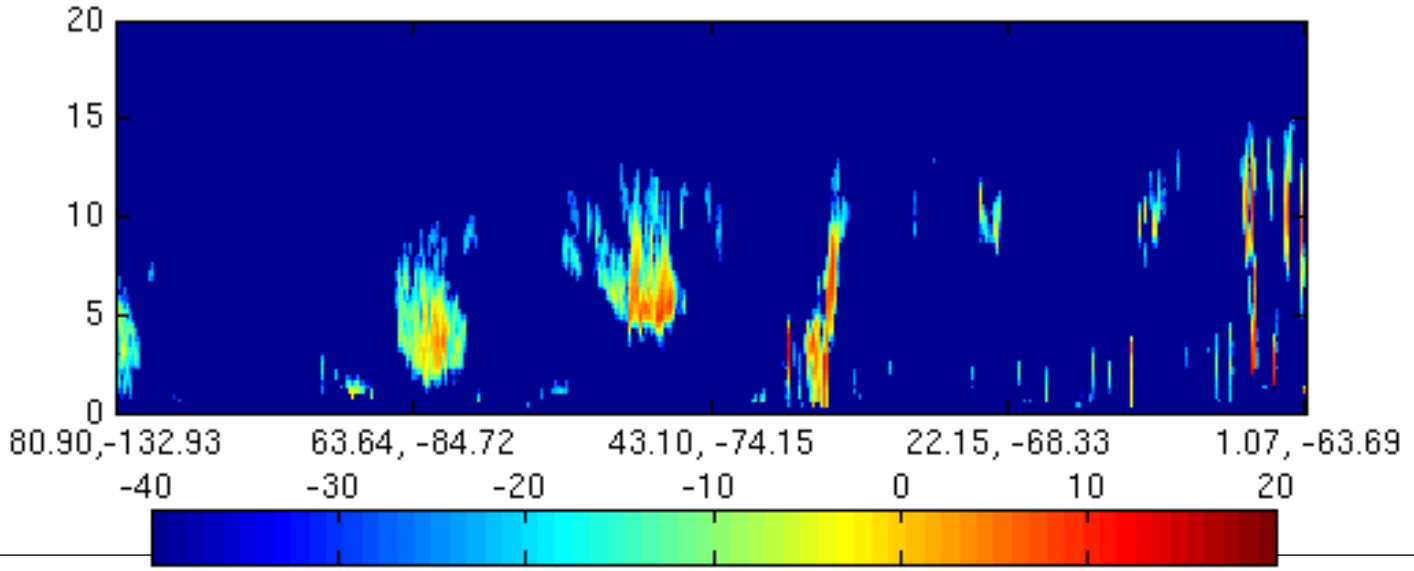


CloudSat/NWP Comparison: Object Based: Reflectivity

Model



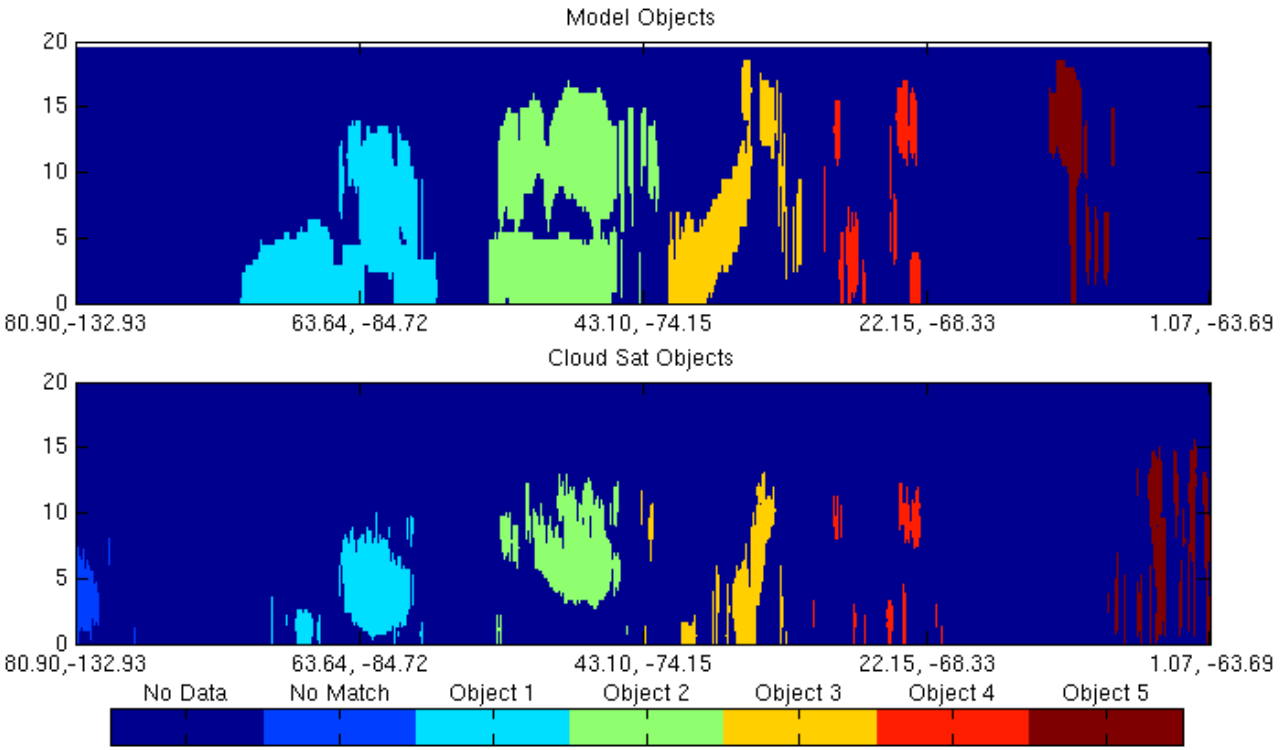
CloudSat



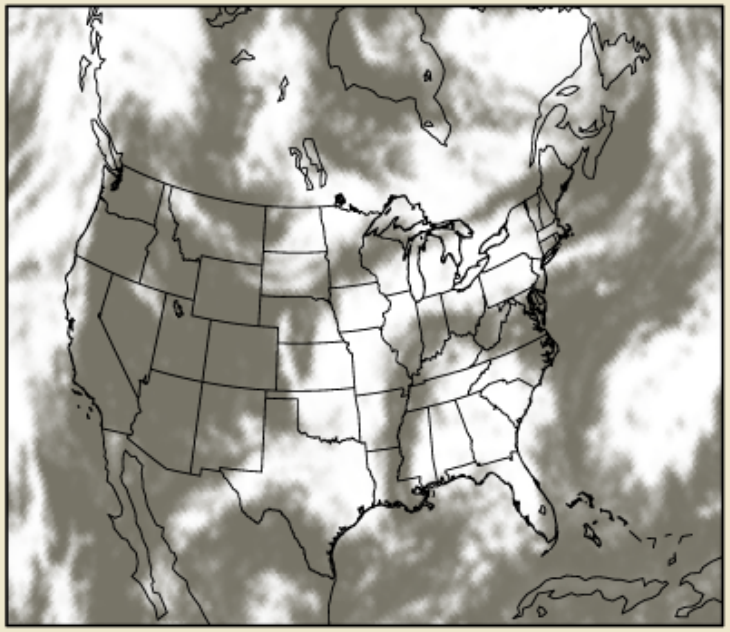
CloudSat/NWP Comparison: Object Based: Objects

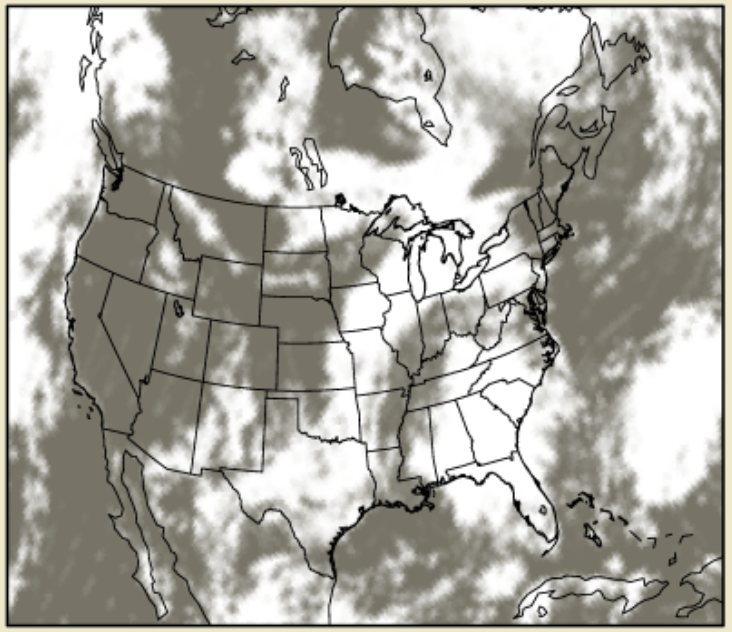
Model

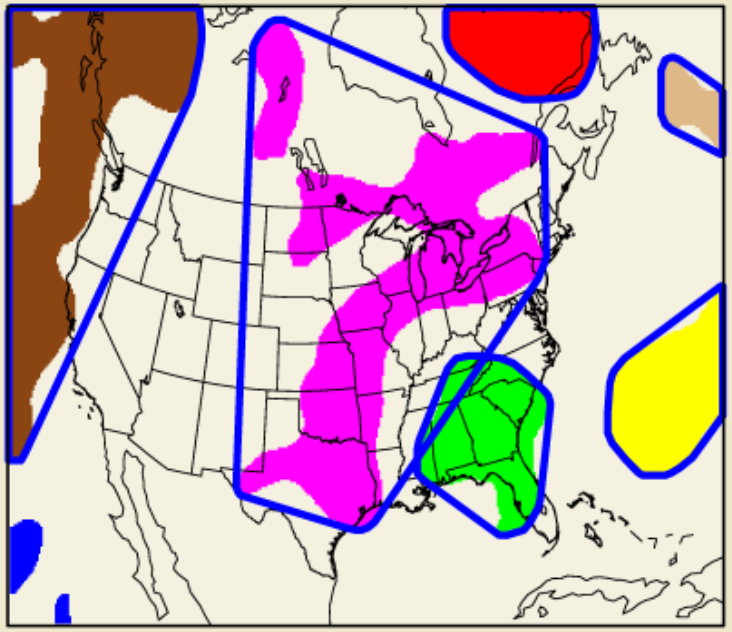
CloudSat

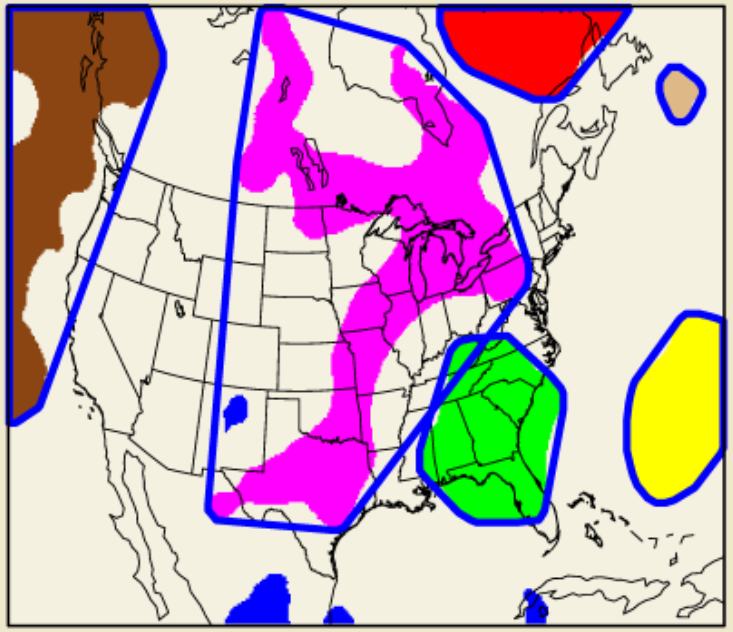


Forecast









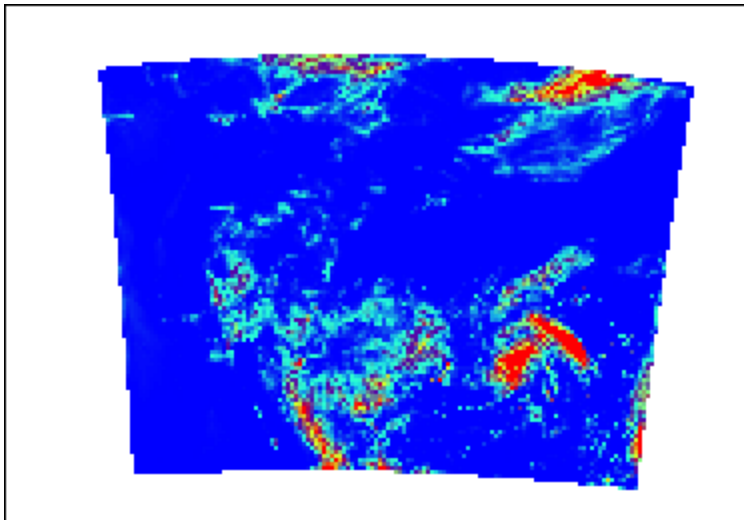
Observed

Scale Decomposition

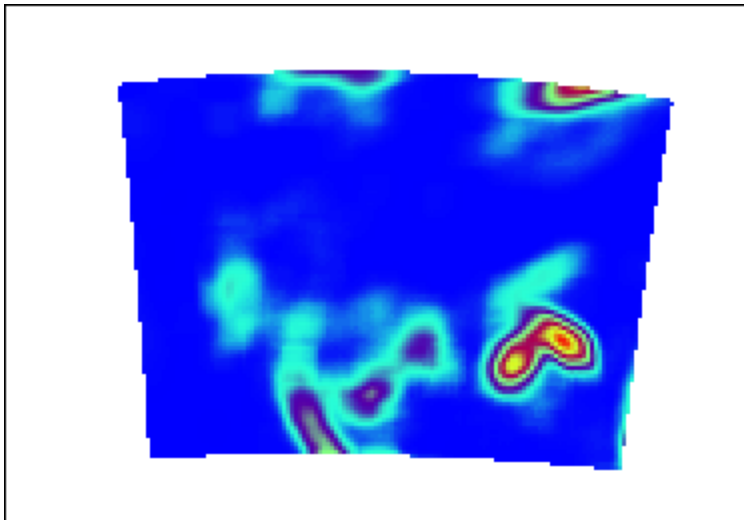
Grid Stat and Wavelet Stat

Neighborhood Methods - Smoothing

- Smoothing Filters in MET
  - Minimum, Maximum, Median, Mean



original

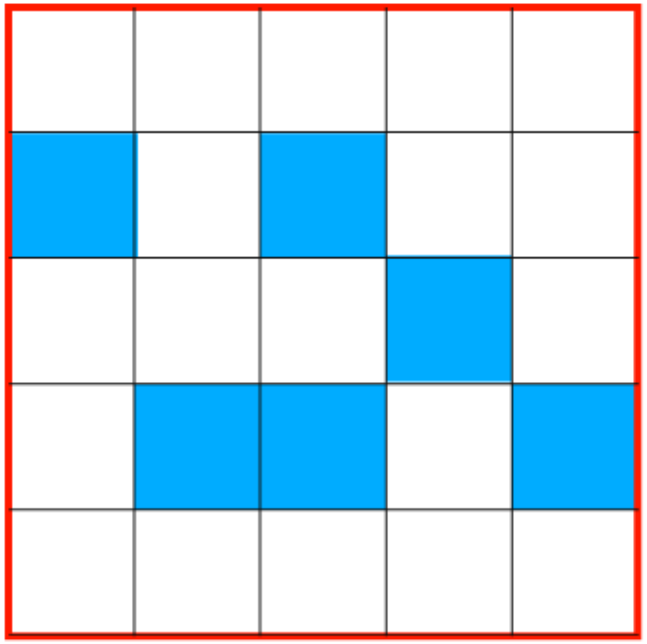


mean

Neighborhood Methods: Fractional coverage of events

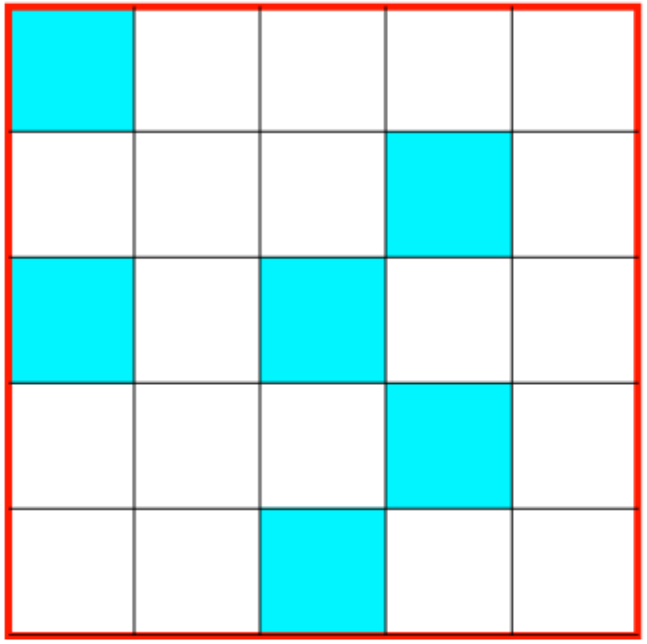
Intensity threshold exceeded where squares are blue

observed



Fraction = 6/25 = 0.24

forecast



Fraction = 6/25 = 0.24

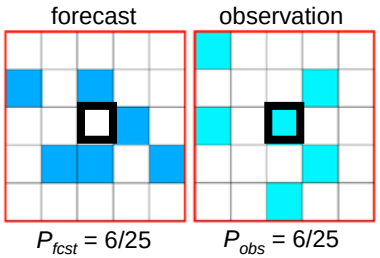
P is the fractional event frequency within the neighborhood.

This is calculated for all *n* grid points in the domain

slide from Mittermaier

Neighborhood Methods

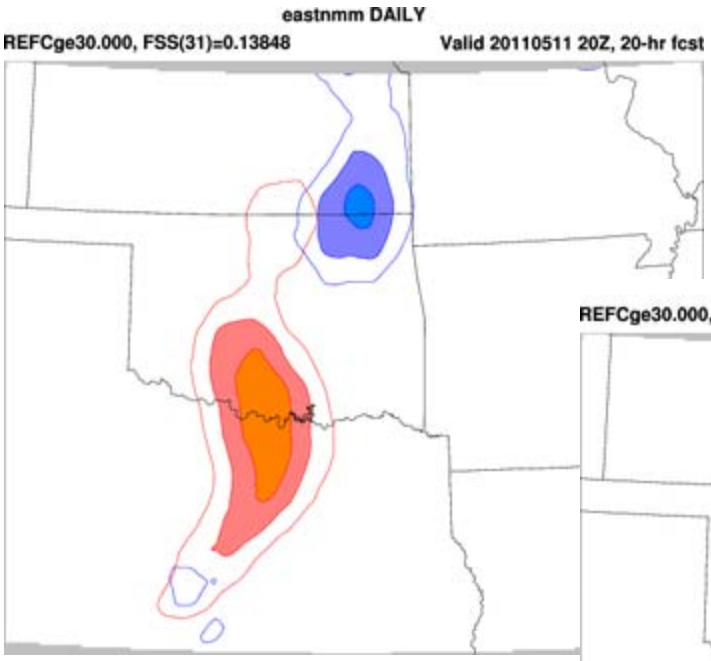
n = grid points in the domain



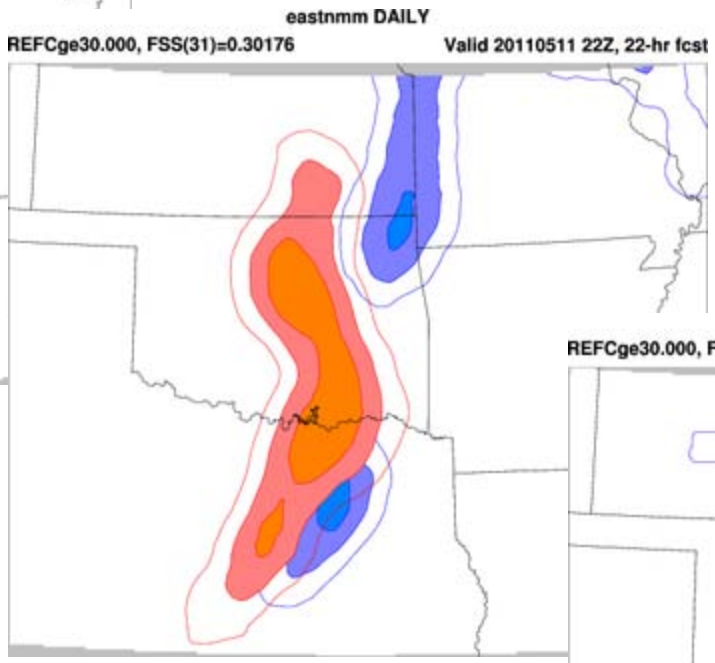
$$1 - \frac{\text{Fractions Brier Score}}{\text{(worst possible) Fractions Brier Score}}$$

Fractions Skill Score (FSS) of Roberts and Lean (2008)

Example – REFC > 30 dBZ



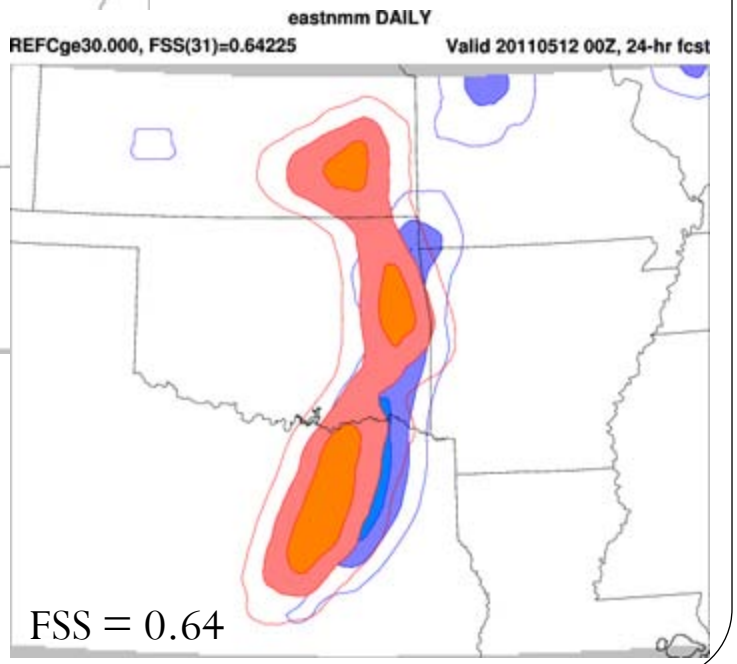
FSS = 0.14



FSS = 0.30

Time

Overlap

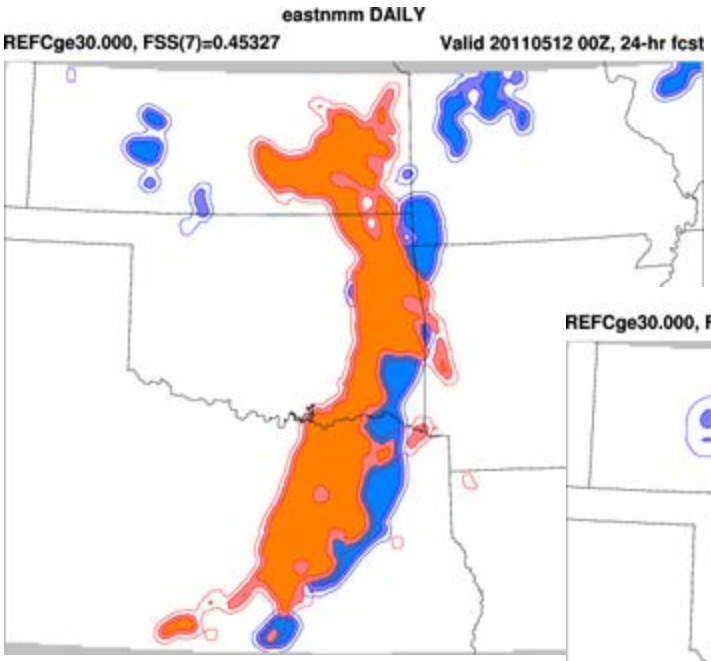


FSS = 0.64

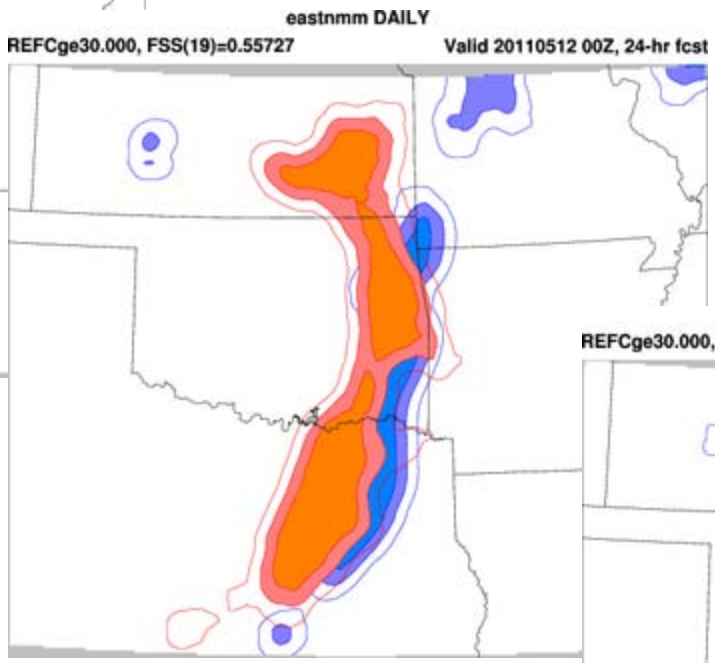
Forecast Fractional Coverage

Observed Fractional Coverage

Example – REFC > 30 dBZ

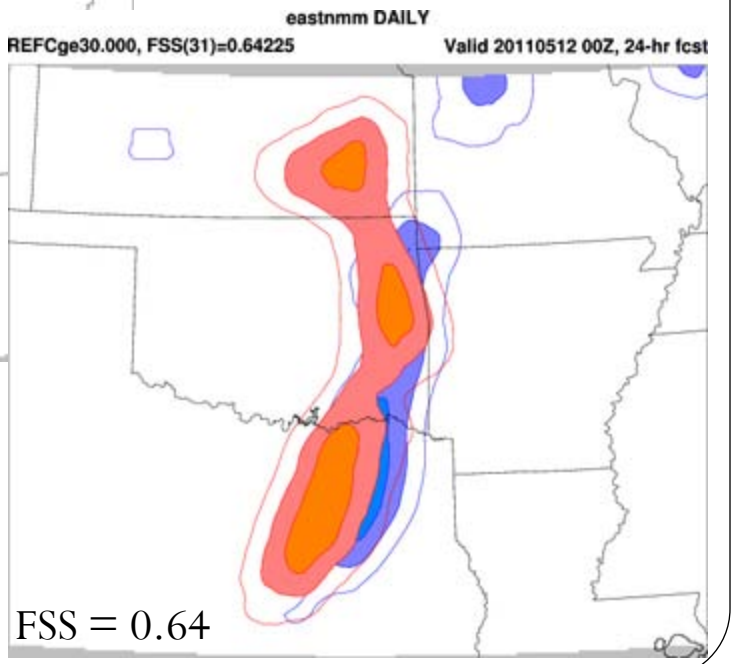


FSS = 0.45



FSS = 0.56

Neighborhood Size



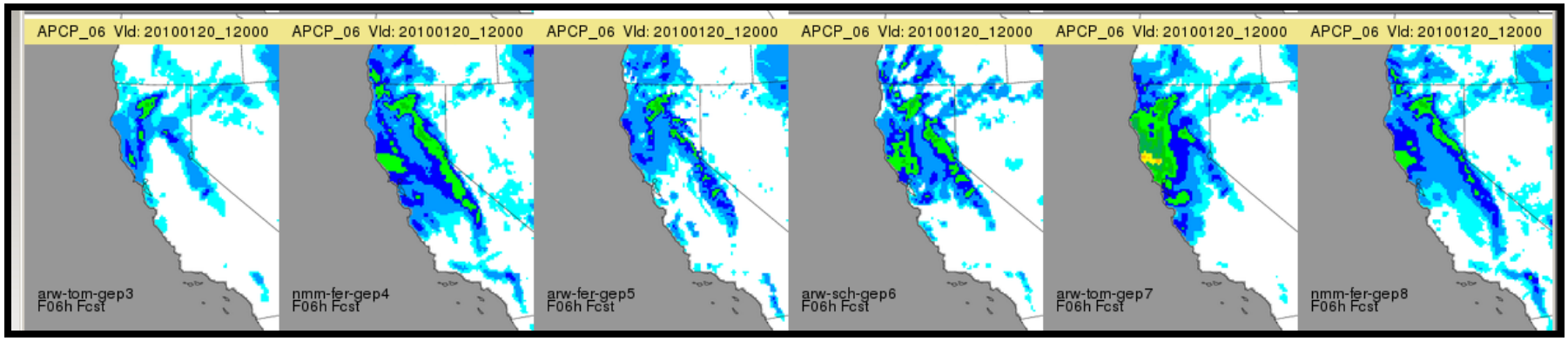
FSS = 0.64

Increase in FSS with increasing neighborhood size is a characteristic of method and can be attributed to the smoothing that occurs

Ensembles

Ensemble Stat, Point Stat, Grid Stat and MODE

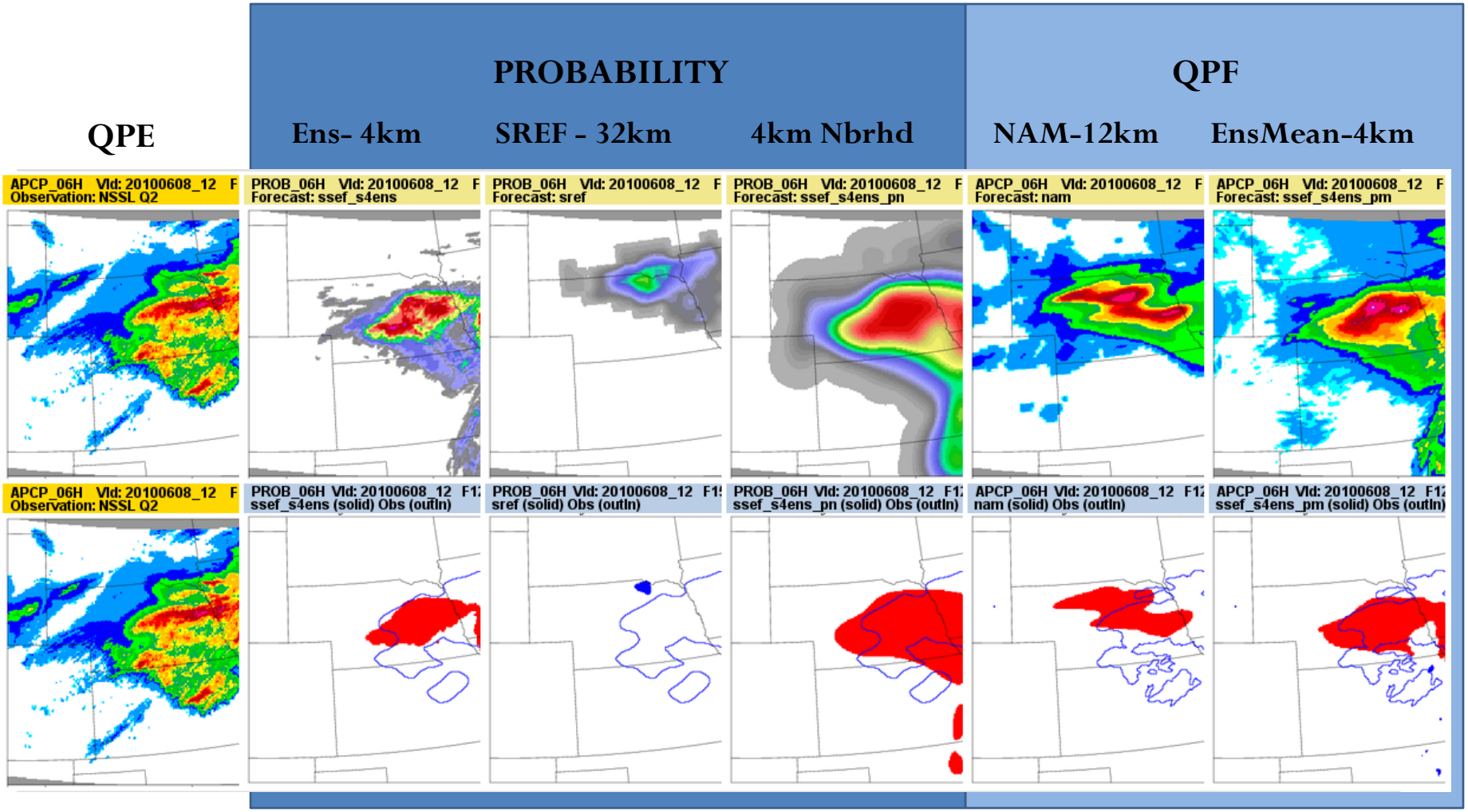
Ensemble Stat: Overview



- Ensemble Stat primarily a pre-processing tool
  - Mean, Spread, Mean $\pm$ 1 STDev, Min, Max, Range, Number of valid members
- Calculates rank histograms and outputs:
  - Bins and Counts
  - Matched Pair ranks

MODE on Probabilistic Data

Probabilistic Fields (PQPF) and QPF Products

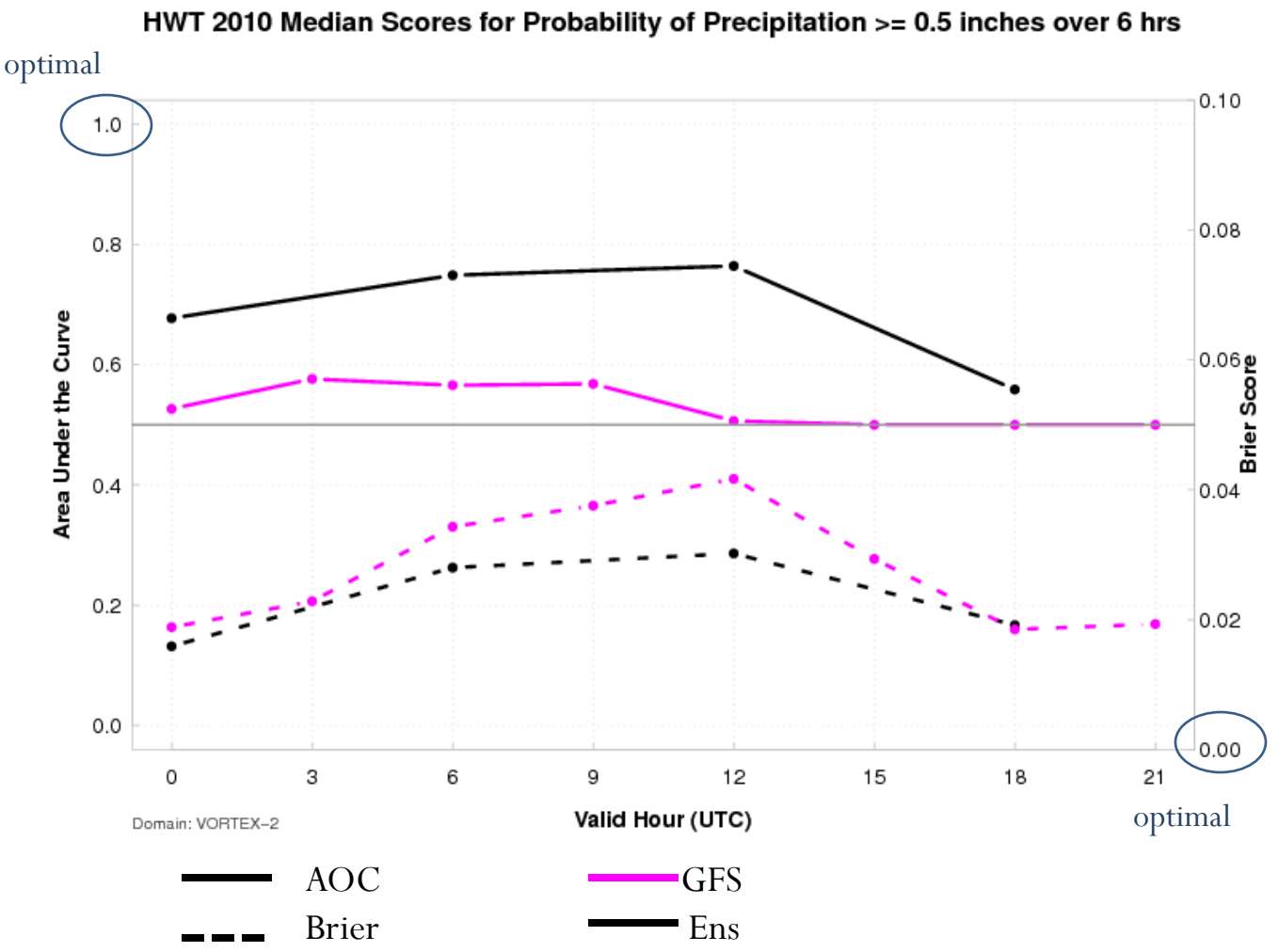


Probabilistic Metrics in Grid-Stat

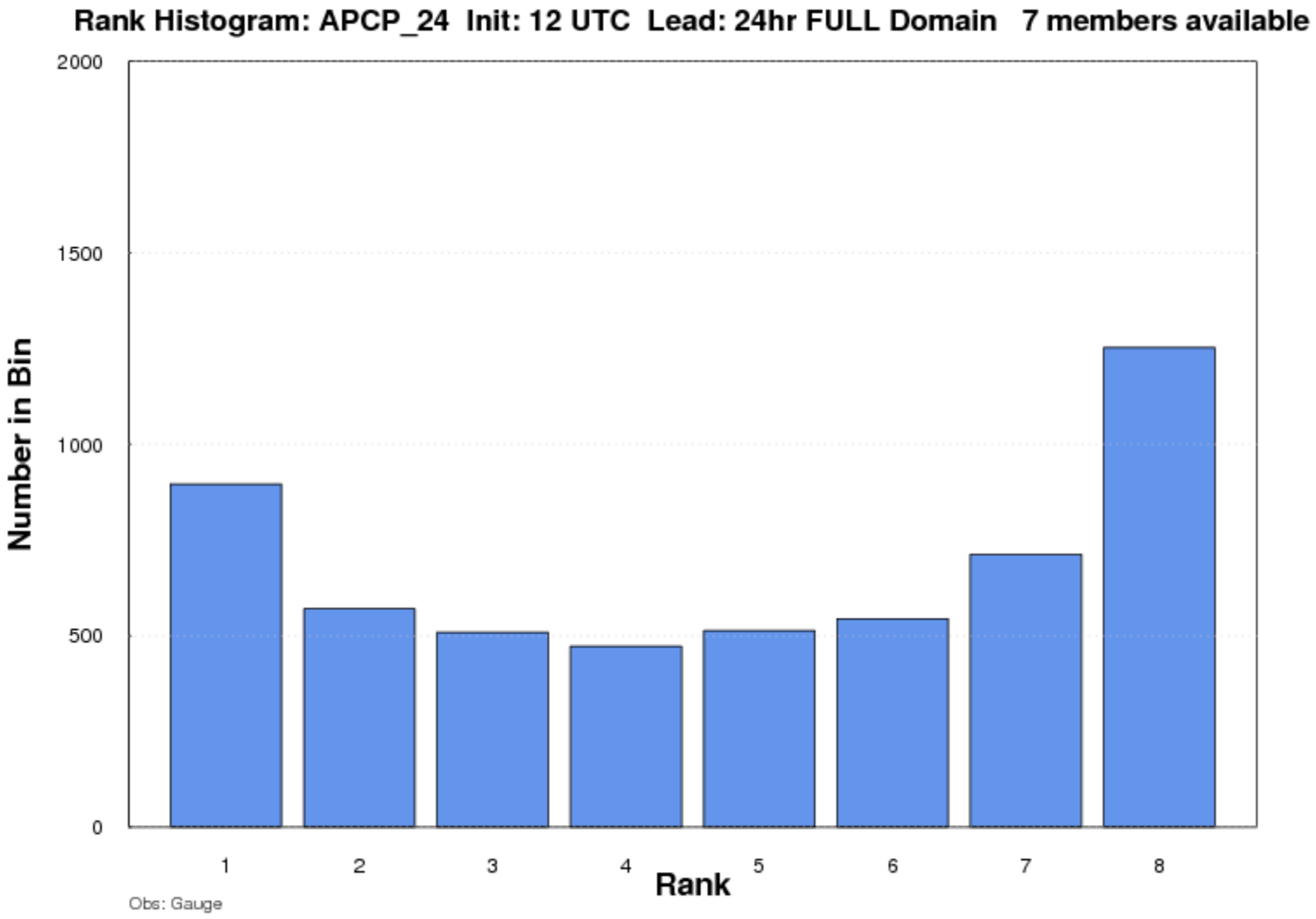
Mesoscale vs. Convection Allowing Ensemble

06hr
accumulations

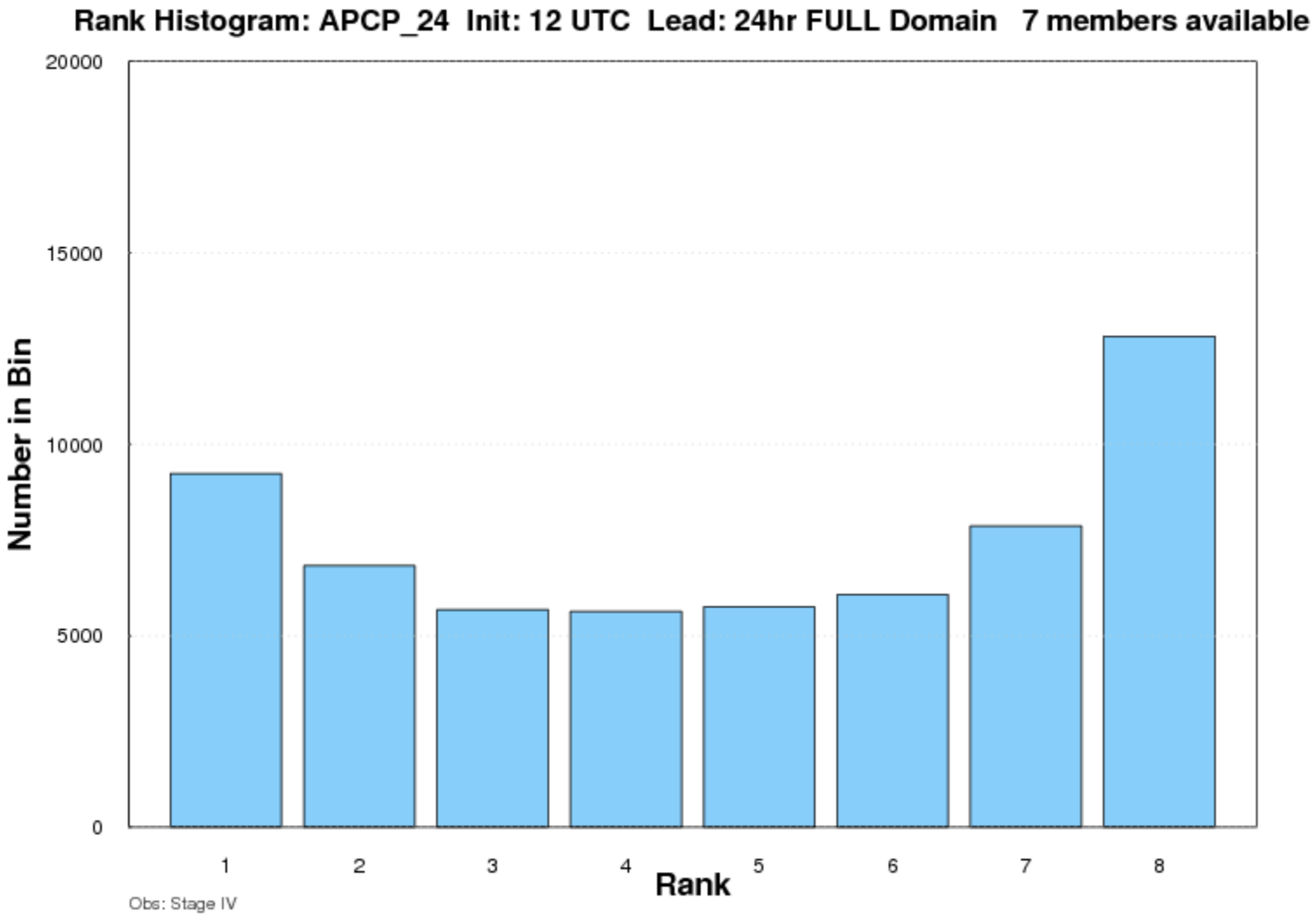
- **SREF PQPF** - minimal skill
- **HWT Ens PQPF** - nominal skill at 0.5”
- **HWT Neighbhd PQPF** – not show has slightly higher skill



HMT Ensemble for Extreme Cool Season Precip over Northern CA using Gauge Data



HMT Ensemble for Extreme Cool Season Precip over Northern CA using Stage IV QPE



MET tutorial

June 27 – 28

NCAR Foothills Lab

Registration and Information:

http://www.dtcenter.org/events/workshops11/met_tutorial.php

Thank Yous and Further Information

DTC would like to thank you for your interest and the assistance of all of our collaborators…

JNT: http://www.ral.ucar.edu/jnt
DTC: http://www.dtcenter.org
MET: http://www.dtcenter.org/met

Email: tressa@ucar.edu
jensen@ucar.edu

Support for the Developmental Testbed Center (DTC), is provided by NOAA, AFWA NCAR and NSF