



# 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

*ICAP Workshop – Boulder, Colorado – 12 May 2011*



# Thank You



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](http://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

You are here: DTC • [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
- Probabilistic verification methods comparing gridded model data to point-based or gridded observations

#### User Survey

Please take a minute to complete our short [User Survey](#) and help shape the future of MET! The survey is open until July 2, 2009.

#### Joint Numerical Testbed Projects

##### Developmental Testbed Center (DTC)

[Weather Research and Forecasting \(WRF\) Model Support](#)  
[Model Evaluation Tools \(MET\)](#)

##### Data Assimilation Testbed Center (DATC)

#### Joint Numerical Testbed Events

[WRF Summer Tutorial 2009](#)  
07.13.2009 to 07.24.2009  
Location: NCAR, Boulder, CO

[WRF User's Workshop 2009](#)  
06.23.2009 to 06.26.2009  
Location: NCAR, Boulder, CO

[WRF v3.1 release](#)  
04.09.2009

[MET v2.0 release](#)  
04.07.2009

#### MET Announcements

[MET User Survey is now open!](#)  
Current release: [METv2.0](#) (04.07.2009)  
[Online Tutorial updated for METv2.0](#)

#### MET SPONSORS

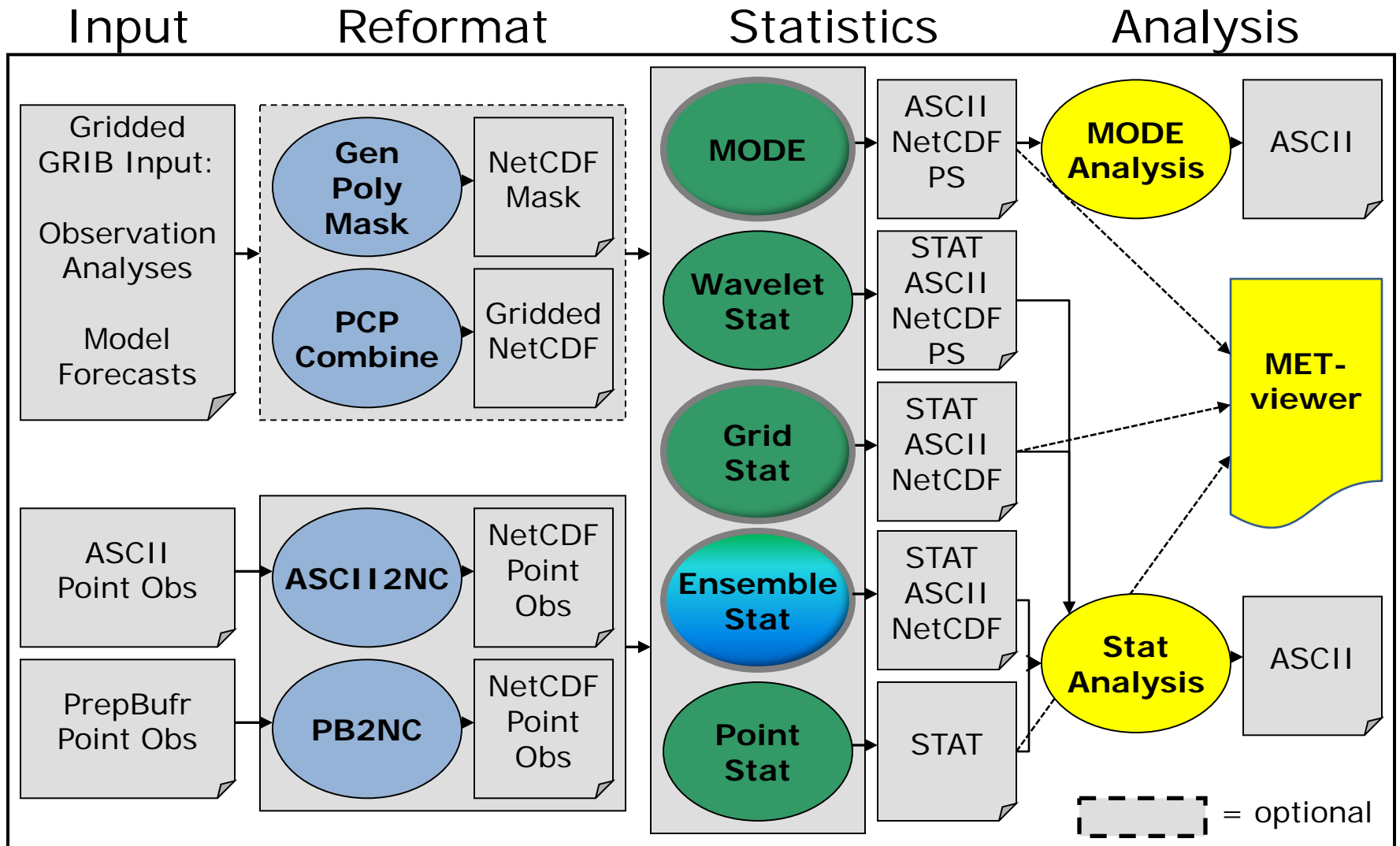
U.S. Air Force Weather Agency (AFWA)



National Oceanic and Atmospheric Administration (NOAA)



# MET v3.0 Tools



<b>Data</b>	<b>MET Tool</b>
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)

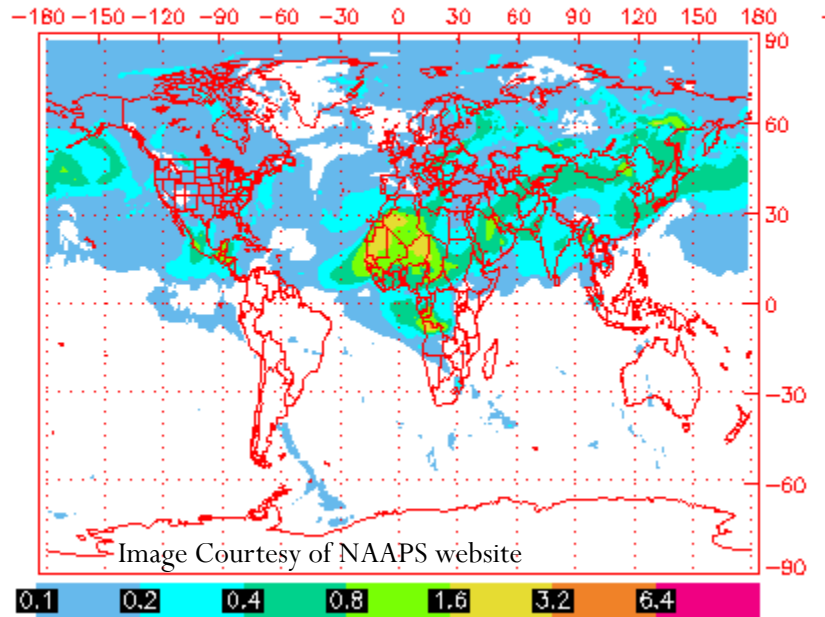
---

Point Stat and Grid Stat



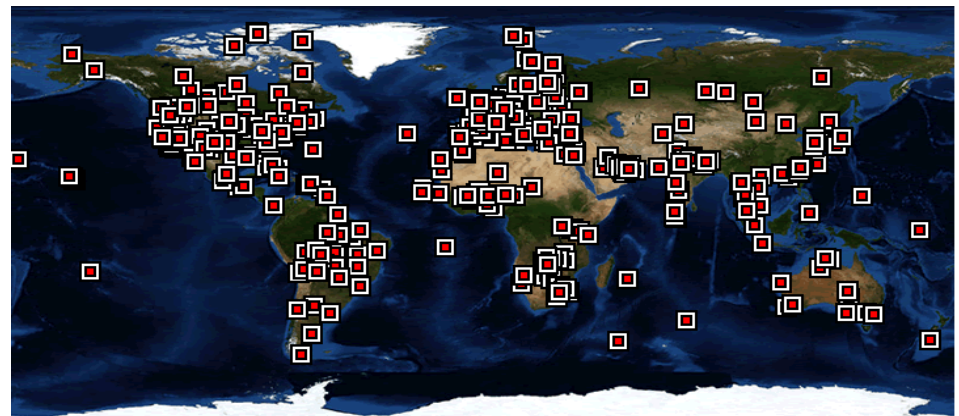
# Point Stat: Overview

NAAPS Total Optical Depth for 00:00Z 17 May 2011  
Contoured at 0.1, 0.2, 0.4, 0.8 etc.



- 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.

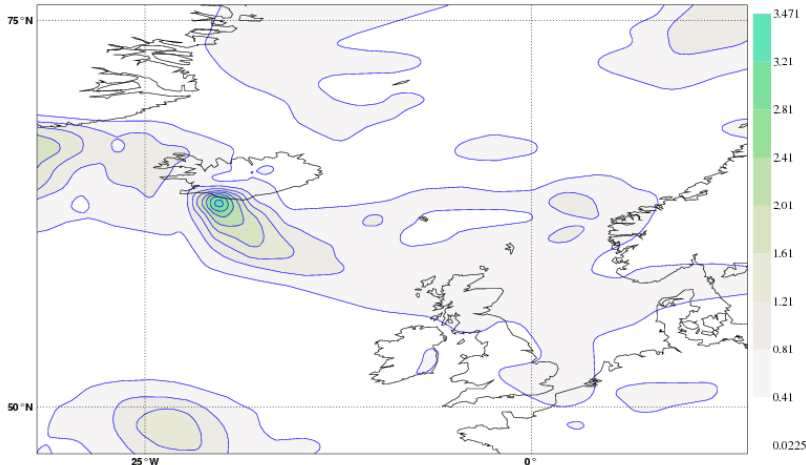


Aeronet Sites – Image Courtesy of Aeronet website

# Grid Stat: Overview

- 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.

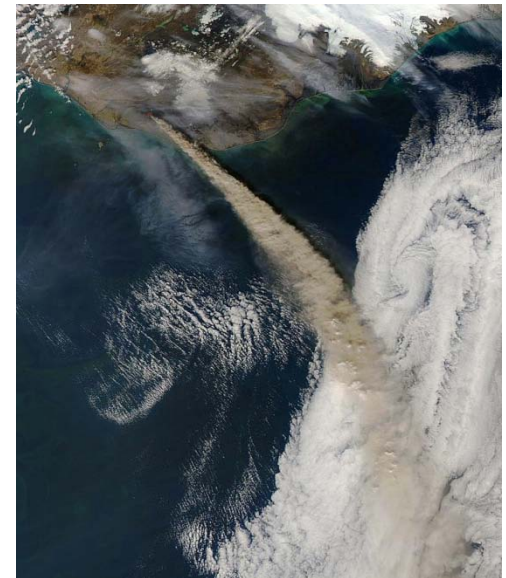
Mon 2010-04-19 12UTC ECMWF FC t+3 VT: Mon 2010-04-19 15UTC Surf: O3



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.

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

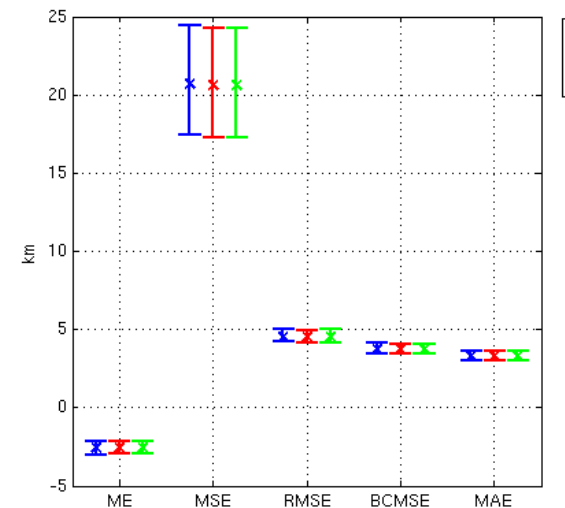
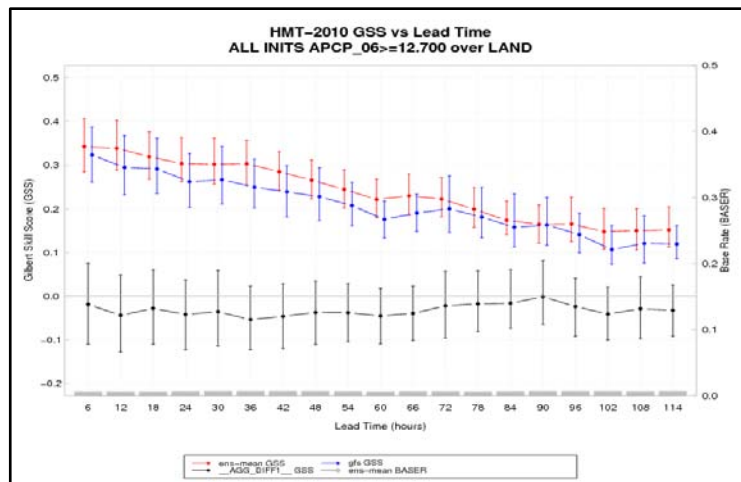
- <http://www.dtcenter.org/met/users/docs/overview.php>

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



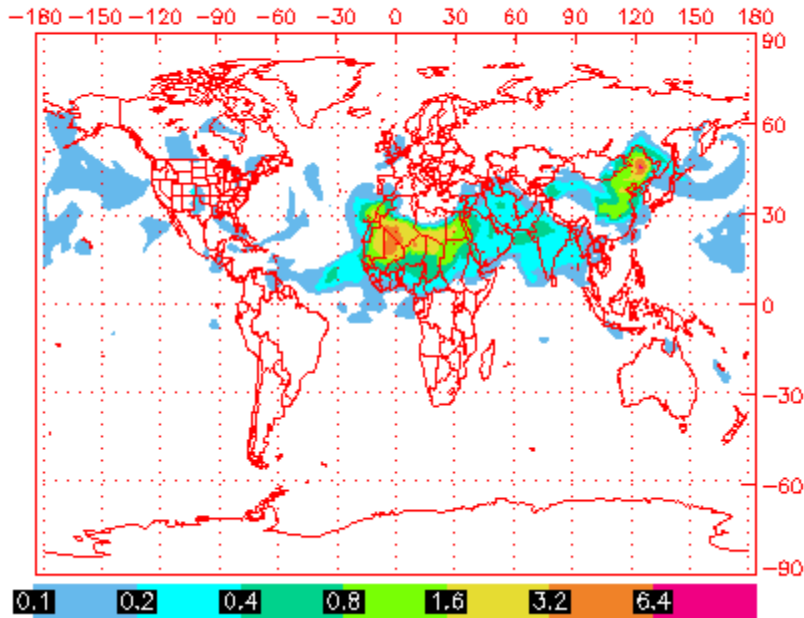
# Object Based Evaluation

---

Method for Object Based Diagnostic Evaluation (MODE)

# Typical situation

NAAPS Dust Optical Depth for 06:00Z 12 May 2011  
Contoured at 0.1, 0.2, 0.4, 0.8 etc.



Forecast

NOAA/NESDIS EDGE IMAGE DISPLAY

. THK.

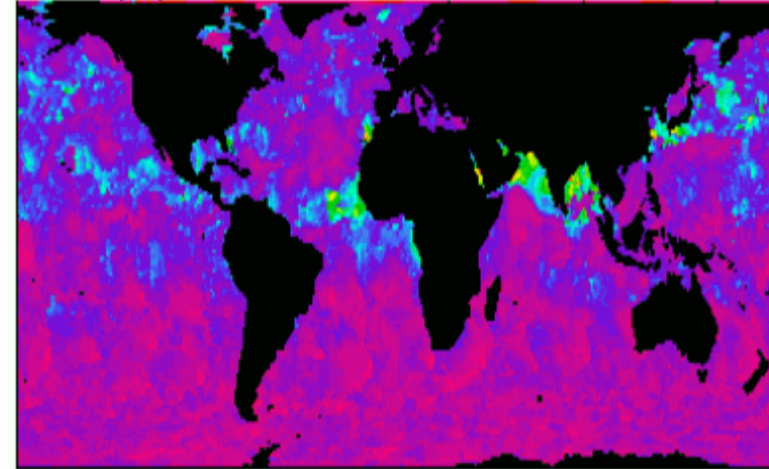
KM GLOBAL ANALYSIS NOAA-19 OPRNL DAILY COMP.

1/2011 4 - 5/12/2011 3

-70.0000 TO 70.0000

-180.000 TO 179.000

23 HOUR



0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8

Observation

Object verification  
is more like what  
humans  
do.

Objects recognize  
the spatial relationship  
between points.

# Simple example

Observed

Red	Red	Red	Light Gray	Light Gray
Light Gray	Light Gray	Light Gray	Light Gray	Light Gray
Light Gray	Light Gray	Light Gray	Light Gray	Light Gray
Light Gray	Light Gray	Light Gray	Light Gray	Light Gray
Light Gray	Light Gray	Light Gray	Light Gray	Light Gray

Forecast

Light Gray	Light Gray	Light Gray	Light Gray	Light Gray
Red	Red	Red	Light Gray	Light Gray
Light Gray	Light Gray	Light Gray	Light Gray	Light Gray
Light Gray	Light Gray	Light Gray	Light Gray	Red
Light Gray	Light Gray	Light Gray	Red	Red

Shifted

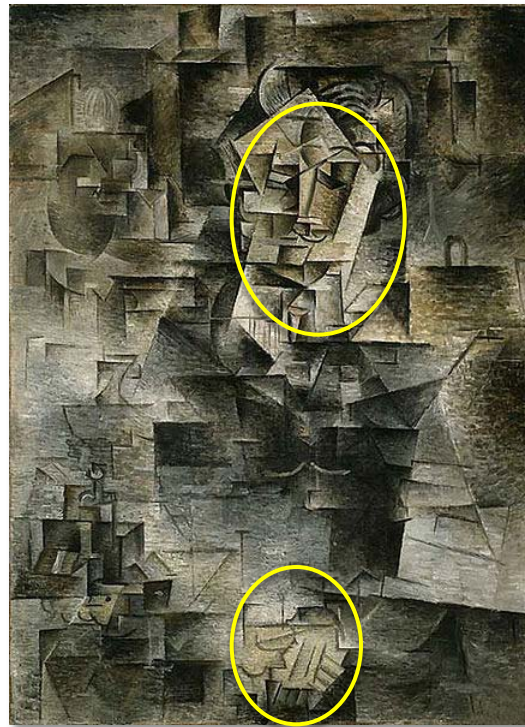
Totally  
wrong



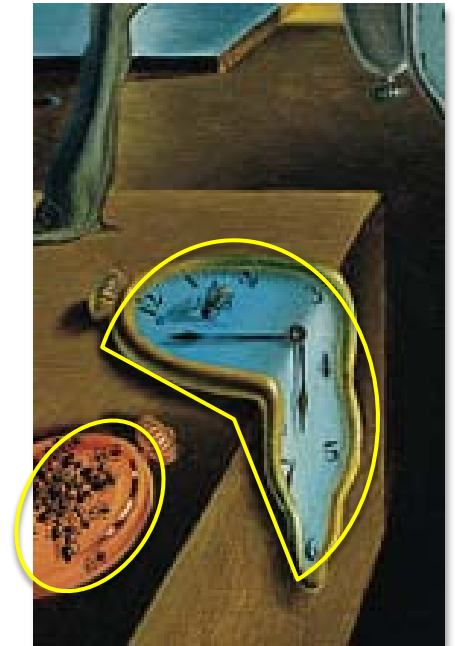
REAL - observed



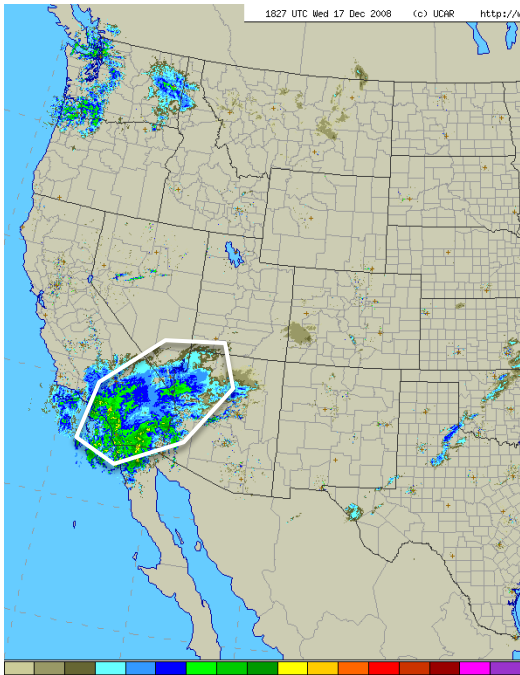
Forecast 1 –  
Distorted view of reality



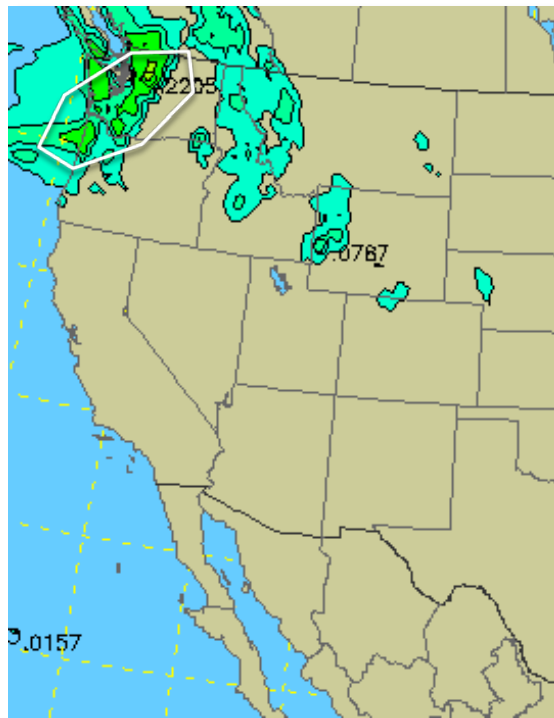
Forecast 2 –  
Another distorted  
view of reality



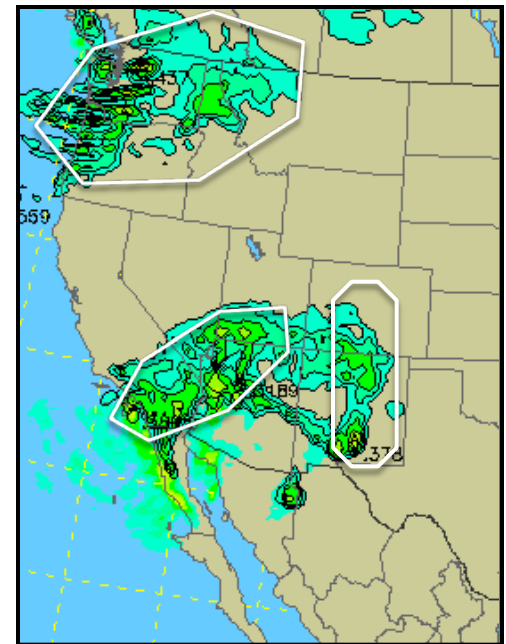
# REAL - observed



# Forecast 1 – Distorted view of reality



# Forecast 2 – Another distorted view of reality



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

---

This:

30% Too Big  
(area ratio=1.3)

Shifted west 1 km  
(centroid distance = 1km)

Rotated 15°  
(angle diff = 15%)

Peak Rain 1/2" too much  
(diff in 90<sup>th</sup> percentile of intensities = 0.5)

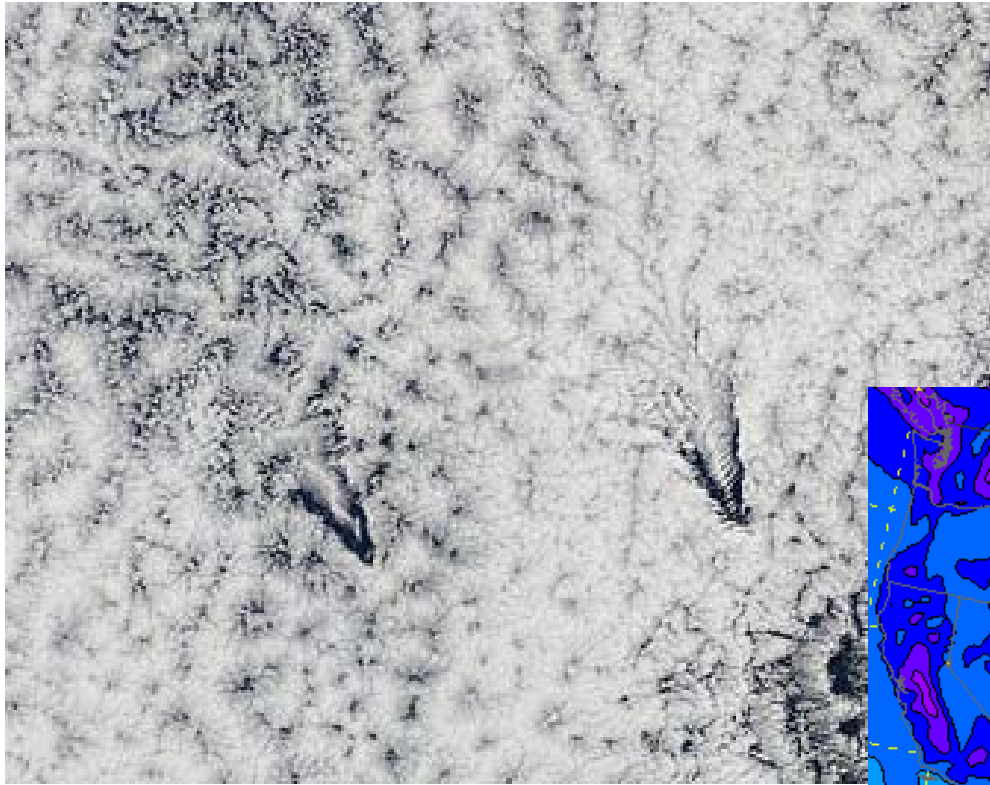
Instead of this:

POD = 0.35

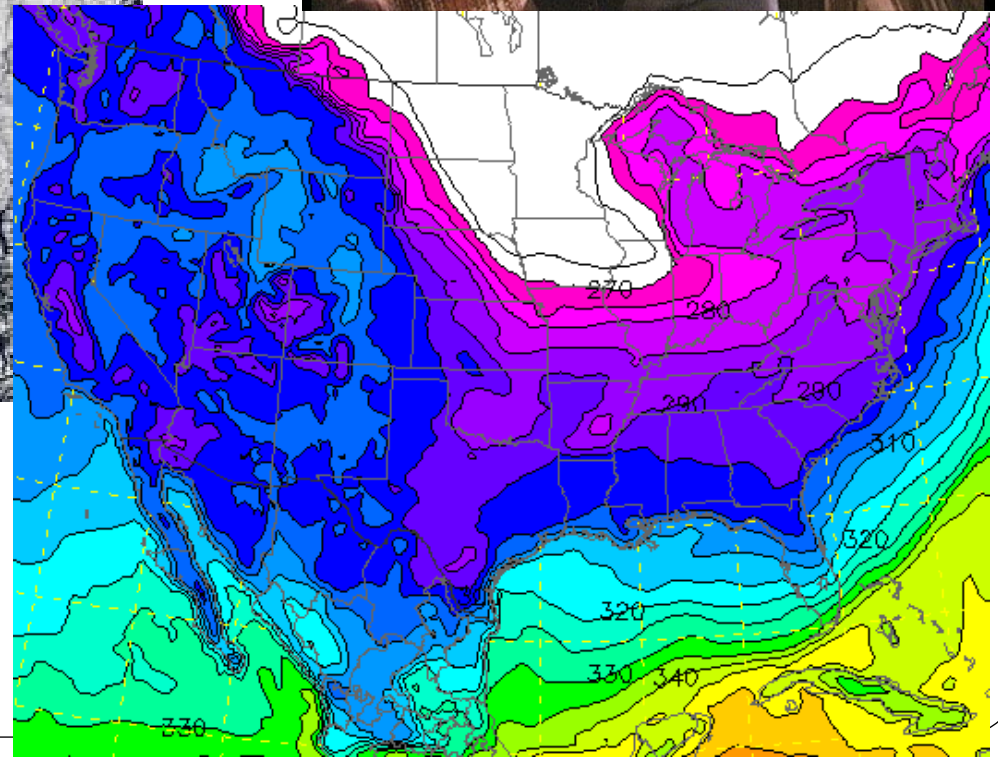
FAR = 0.7235

CSI = 0.1587

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

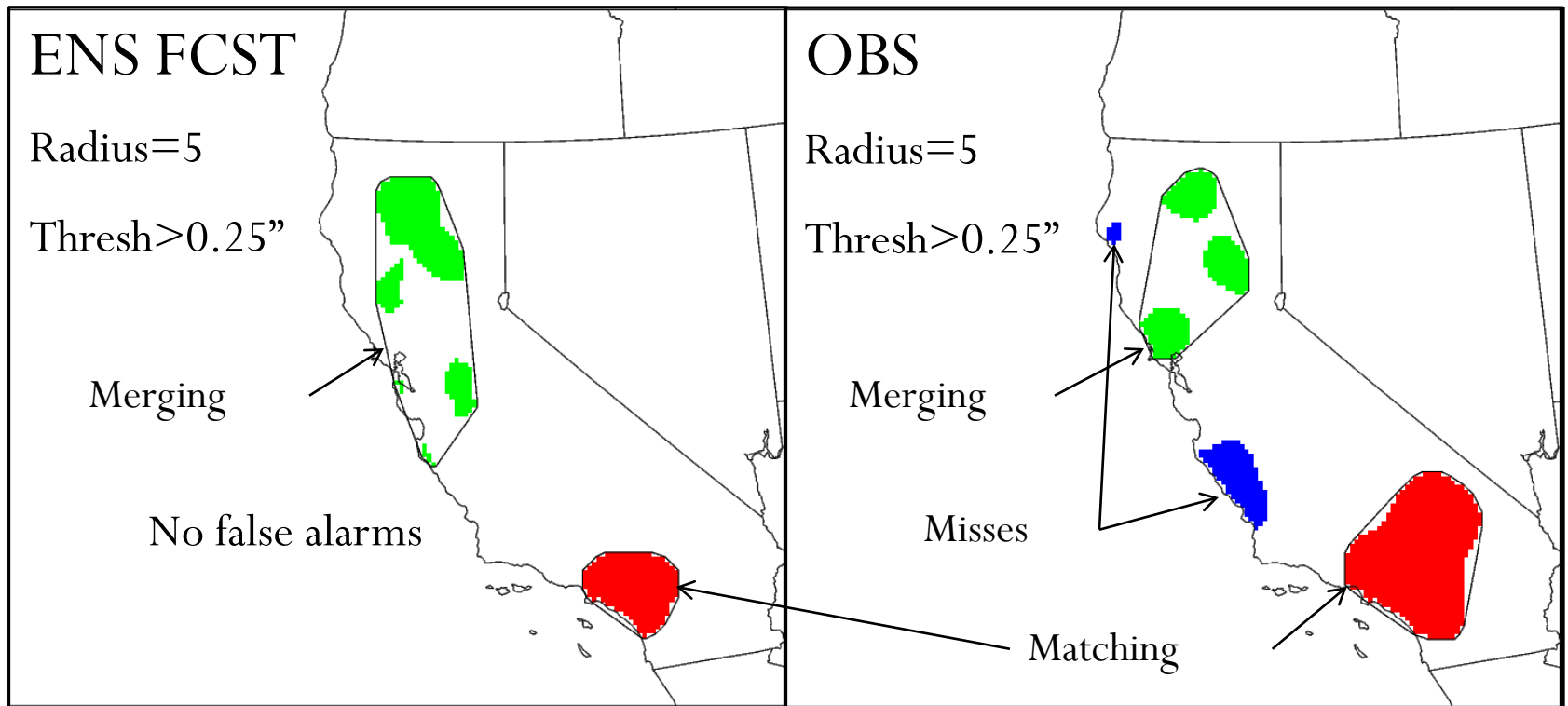


NASA/MODIS - May 9, 2011 - Ship-wave-shape wave clouds induced by Juan Fernandez Islands, off Chile

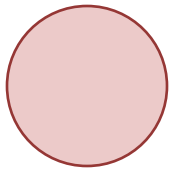


# Object Oriented Method: MODE

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



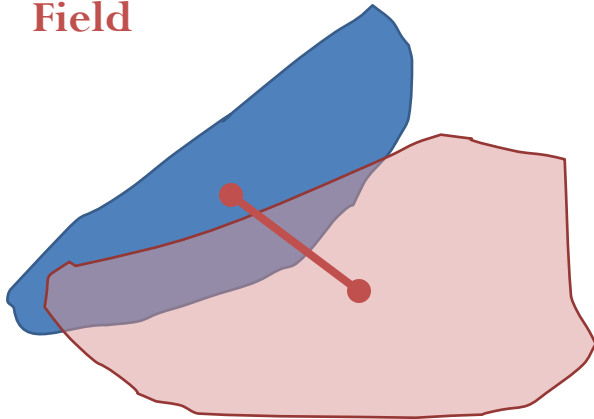
# Use of Pair Attributes defined by MODE



Forecast  
Field

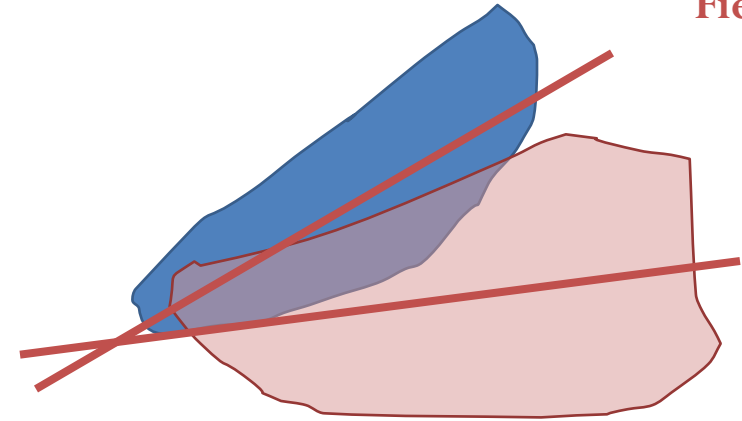


Observed  
Field



**Centroid Distance:** Provides a quantitative sense of spatial displacement of forecast.

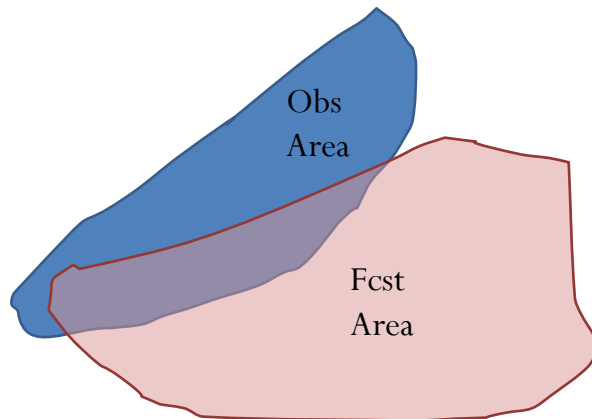
*Small is good*



**Axis Angle:** For non-circular objects – gives measure of orientation errors.

*Small is good*

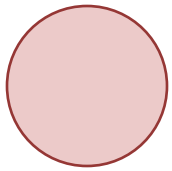
**Area Ratio =**  
$$\frac{\text{Fcst Area}}{\text{Obs Area}}$$



**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

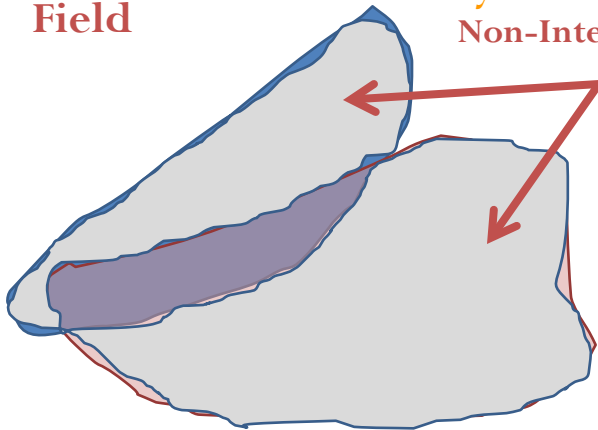


**Forecast  
Field**



**Observed  
Field**

**Symmetric Difference:  
Non-Intersecting Area**

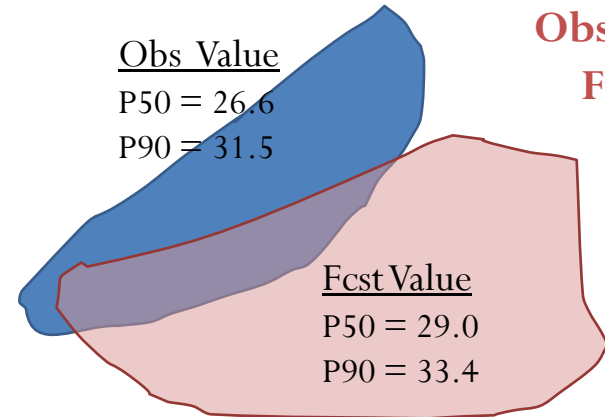


**Symmetric Diff:** May be a good summary statistic for how well Forecast and Observed objects match. *Small is good*

Obs Value

P50 = 26.6

P90 = 31.5



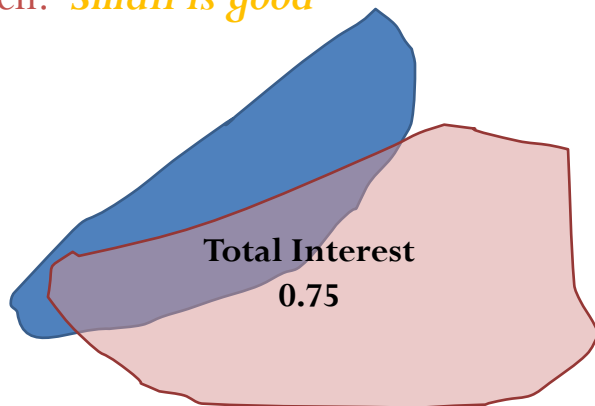
Fcst Value

P50 = 29.0

P90 = 33.4

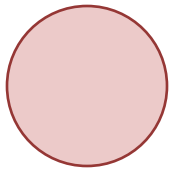
**P50 | P90 Int:** Provides objective measures of Median (50<sup>th</sup> percentile) and near-Peak (90<sup>th</sup> percentile) intensities found in objects.

*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*



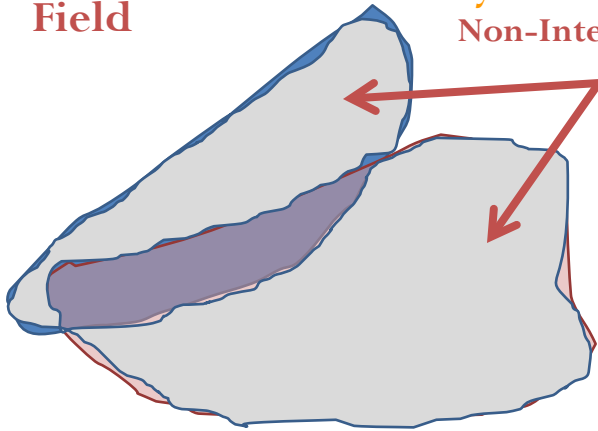
# Use of Pair Attributes defined by MODE



**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*

Obs Value

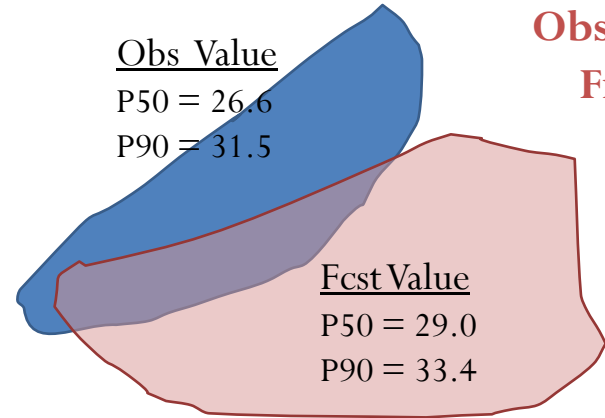
P50 = 26.6

P90 = 31.5

Fcst Value

P50 = 29.0

P90 = 33.4



**P50 | P90 Int:** Provides objective measures of Median (50<sup>th</sup> percentile) and near-Peak (90<sup>th</sup> 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*

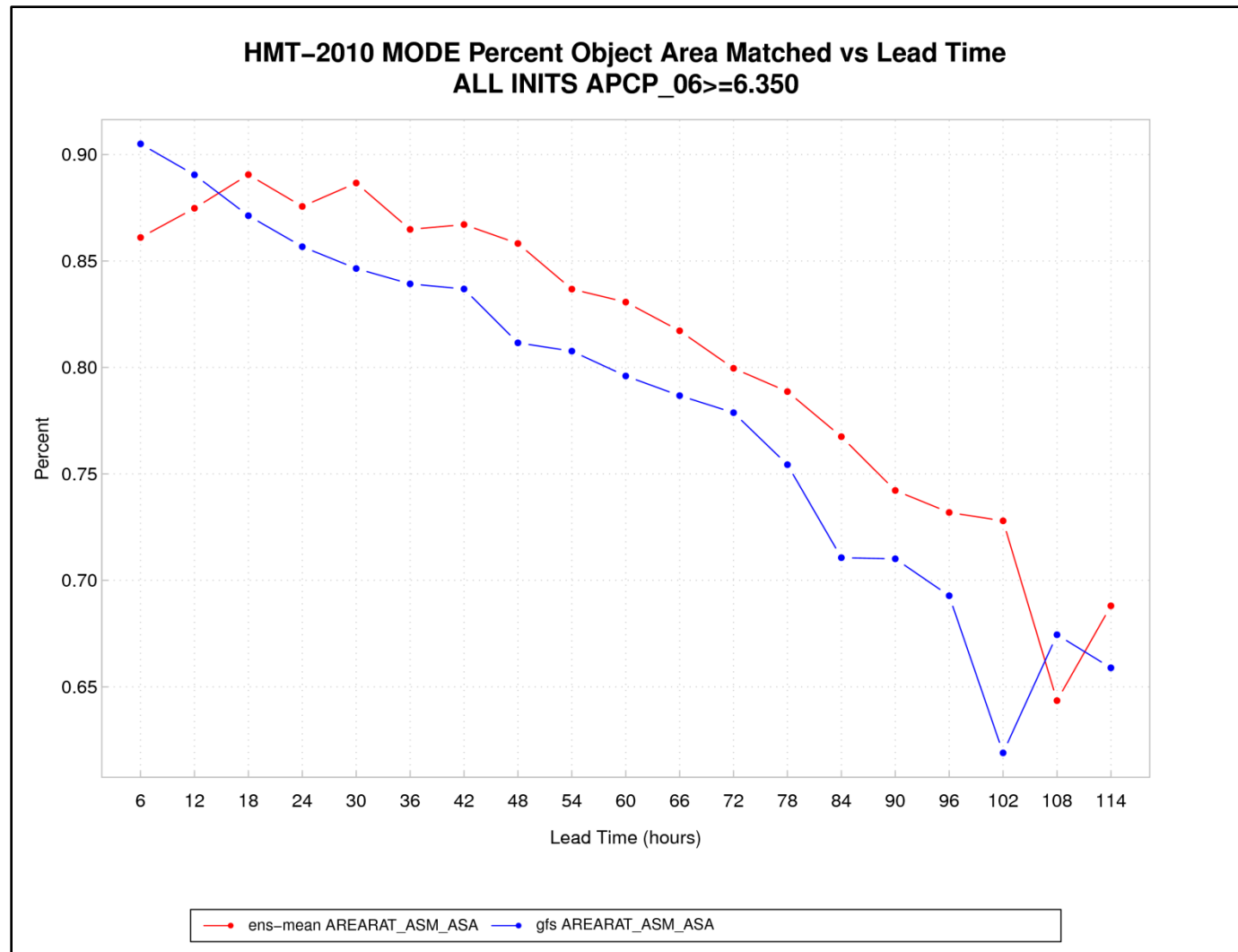


# MODE: Percent Area Matched

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)

capsc0 Fcst and Obs Objects (solid/line) REFC Valid: 20090514\_0000

capsc0 FcstField REFC Valid: 20090514\_0000

**CAPS C0 Objects**

— Forecast  
— Observed

**CAPS C0**

No Radar

FCST  
OBJ

OBS  
OBJ

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

Still working on how to quantify objectively

capscn Fcst and Obs Objects (solid/line) REFC Valid: 20090514\_0000

capscn FcstField REFC Valid: 20090514\_0000

Obs Field REFC Valid: 20090514\_0000

**CAPS CN**

— Forecast  
— Observed

**CAPS CN**

Radar

**Q2 Composite Refl**

Objects

Forecast  
Field

Observed  
Field

# Example: PWAT / IWV

GFS Precipitable Water

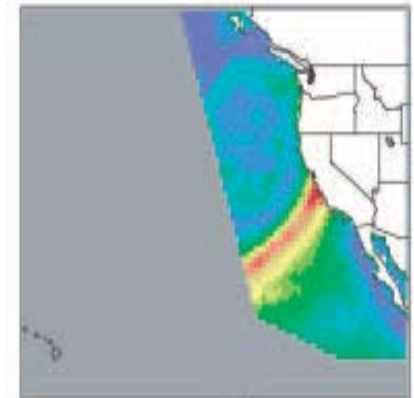
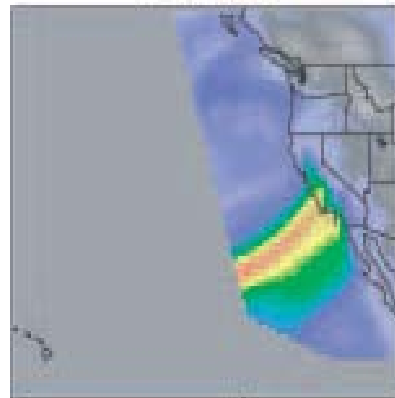
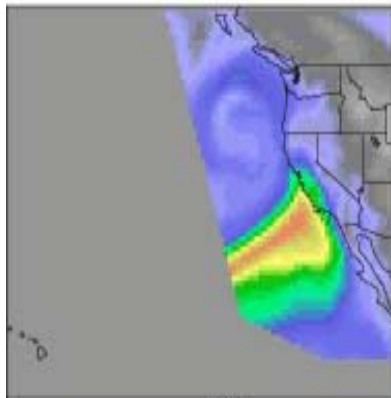
SSMI Integrated  
Water Vapor

Forecast

Forecast

Forecast

Observation



Area=369

Area=312

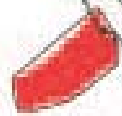
Area=306

Area=127

72 hr

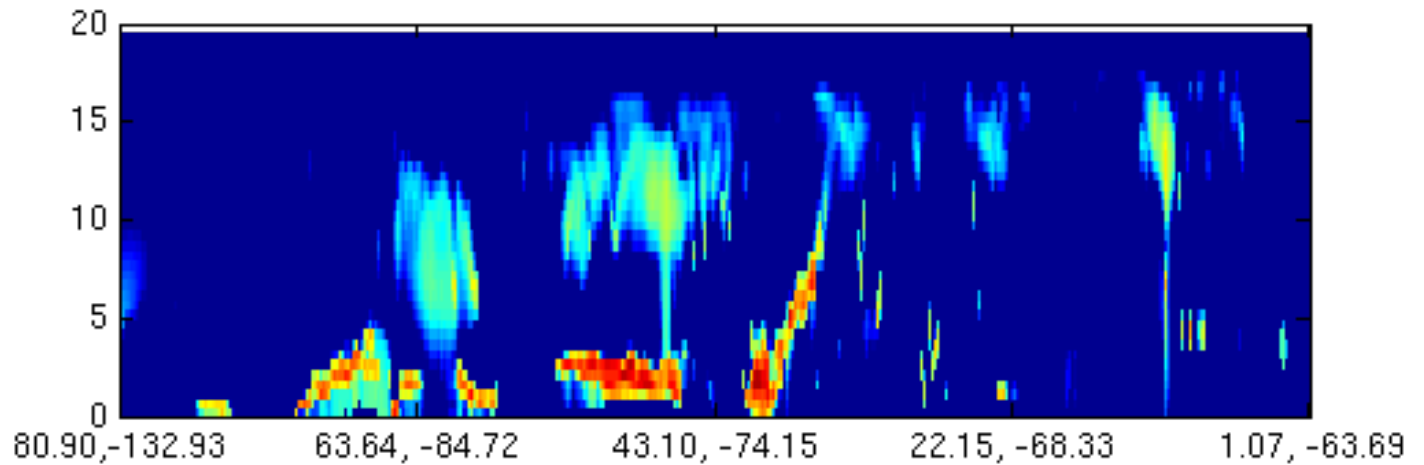
48 hr

24 hr

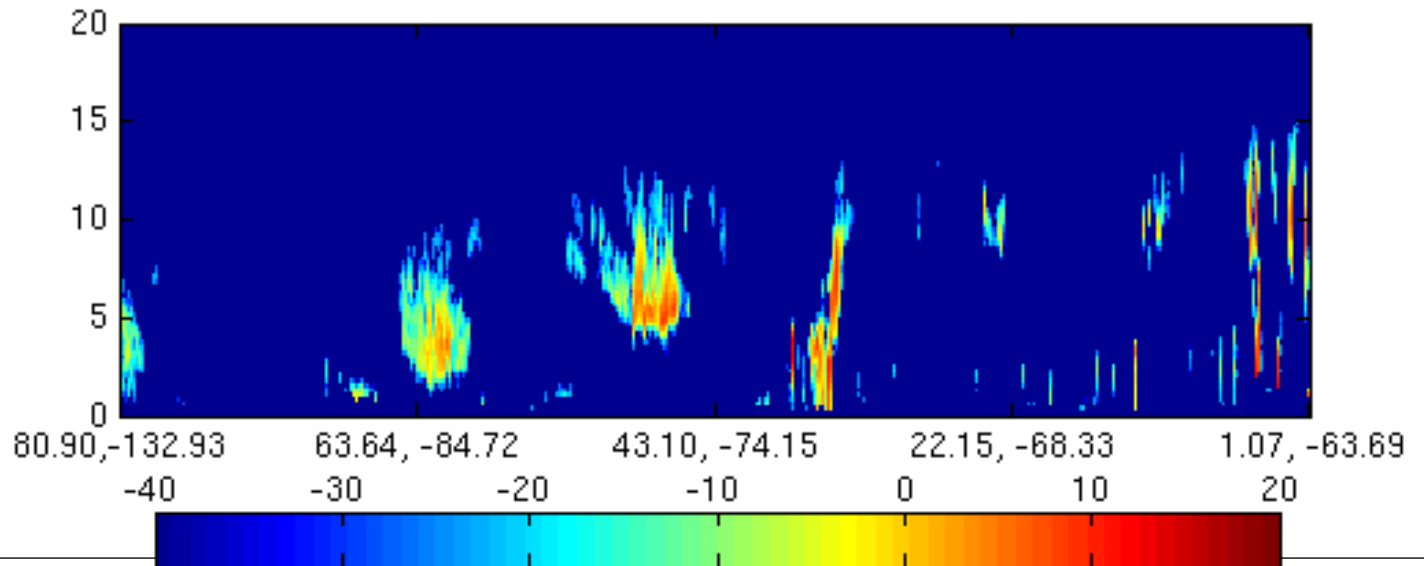


# CloudSat/NWP Comparison: Object Based: Reflectivity

Model

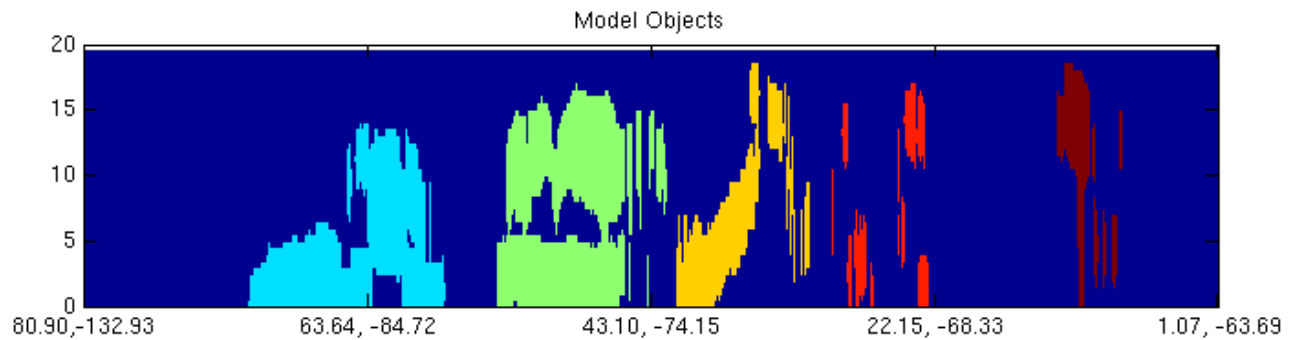


CloudSat

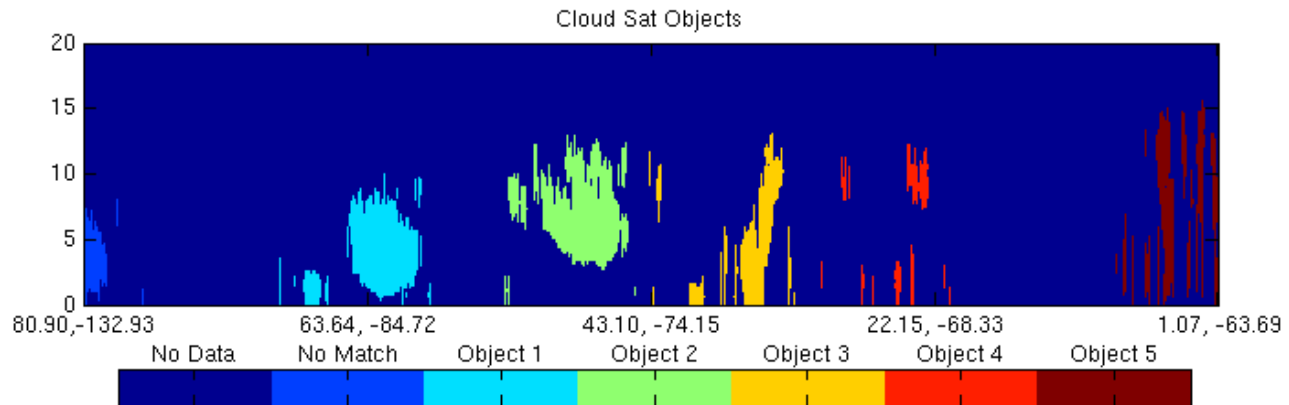


# CloudSat/NWP Comparison: Object Based: Objects

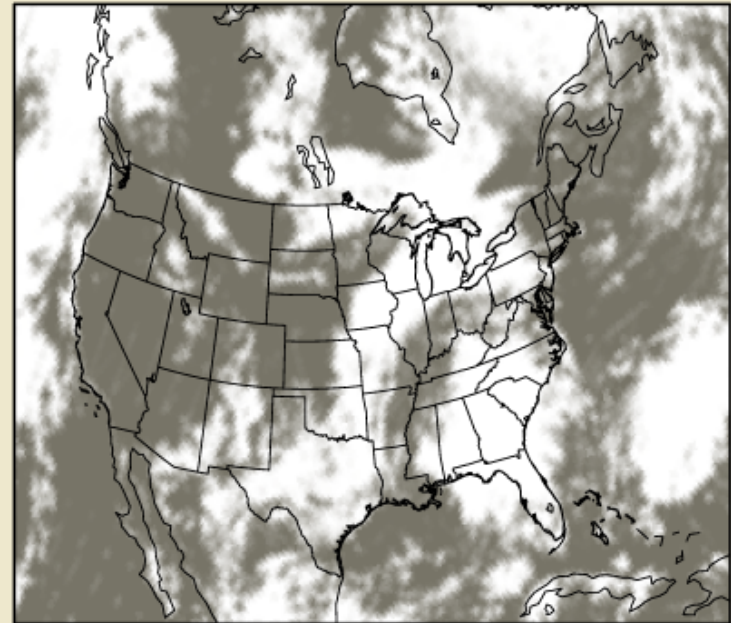
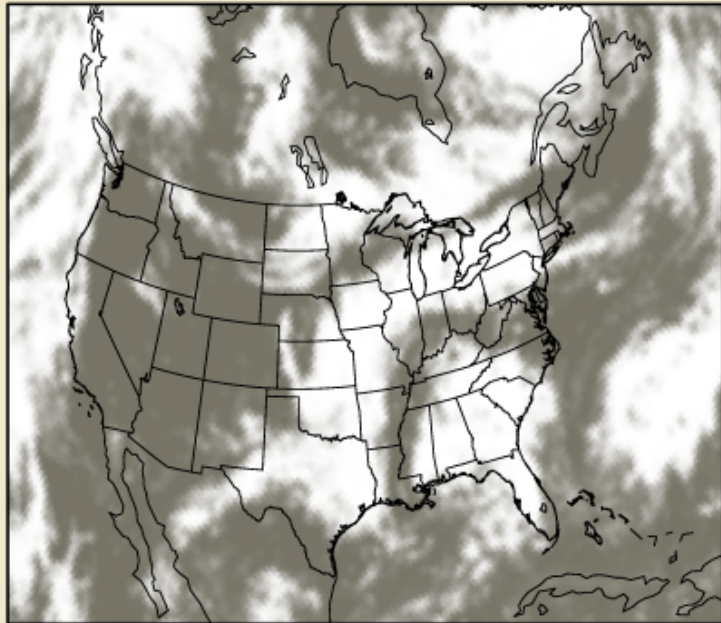
Model



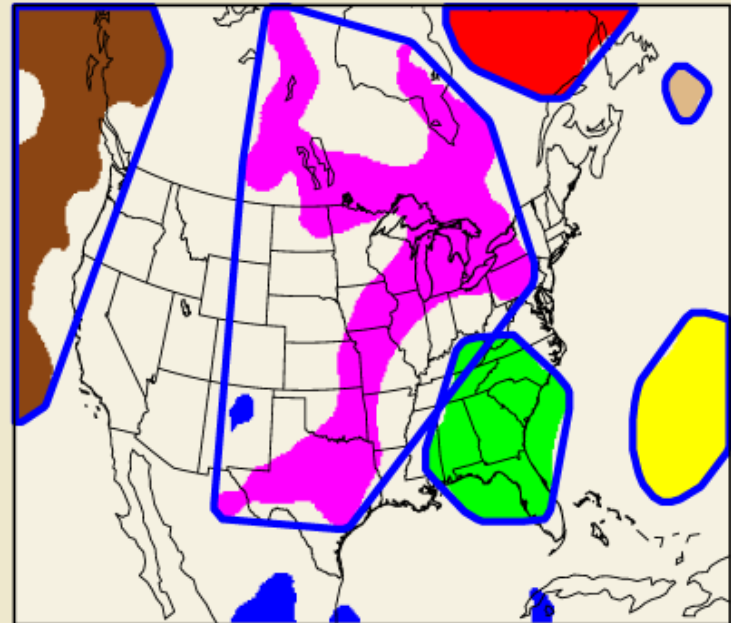
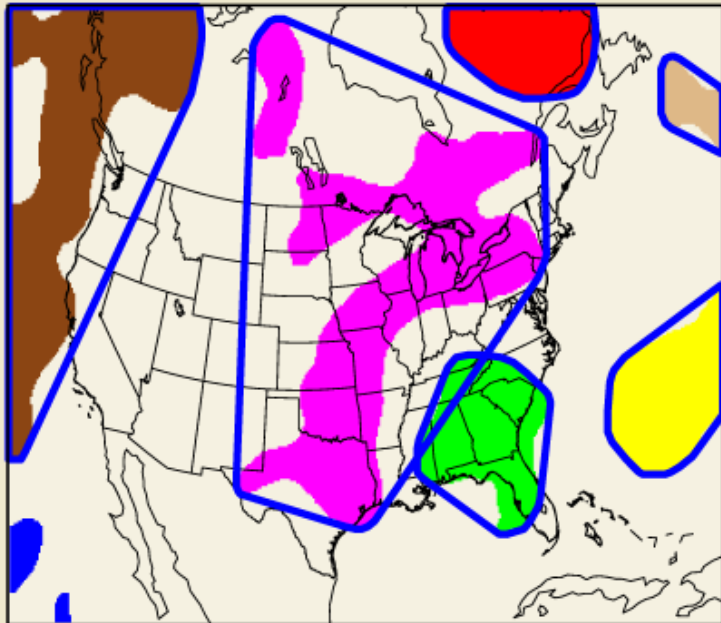
CloudSat



F  
o  
r  
e  
c  
a  
s  
t



O  
b  
s  
e  
r  
v  
e  
d



# Scale Decomposition

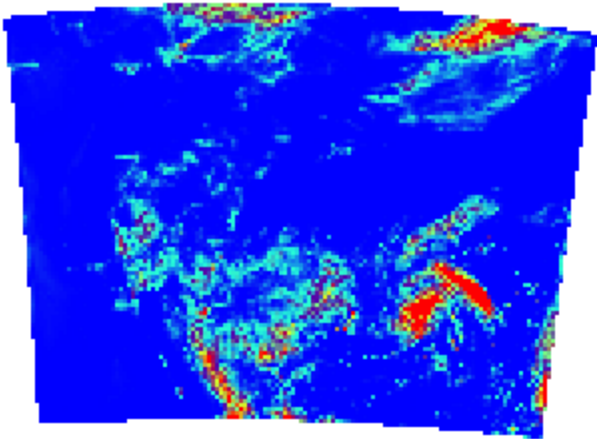
---

Grid Stat and Wavelet Stat

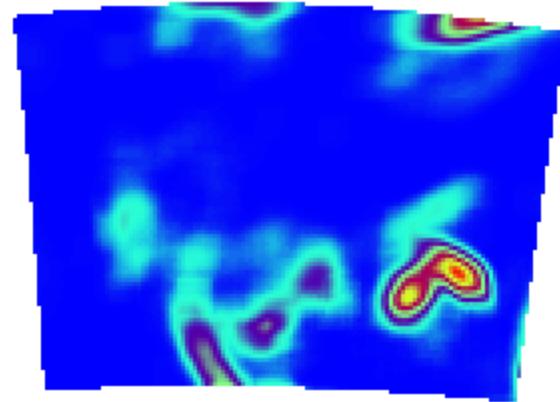


# Neighborhood Methods - Smoothing

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



original

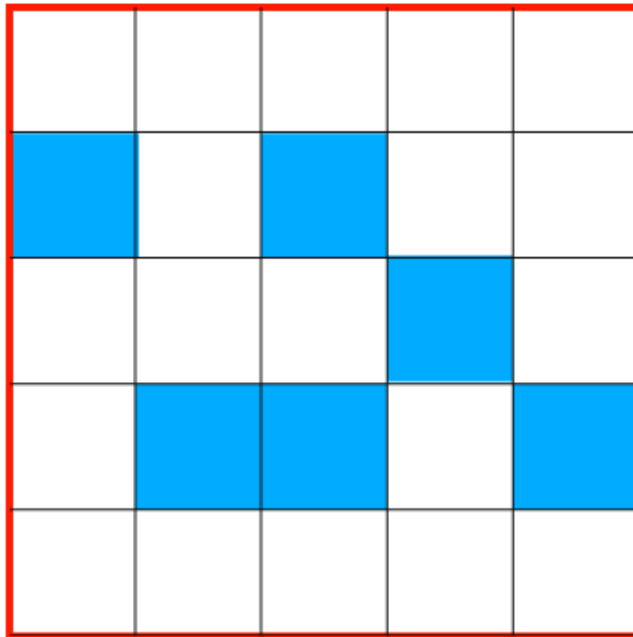


: mean

# Neighborhood Methods: Fractional coverage of events

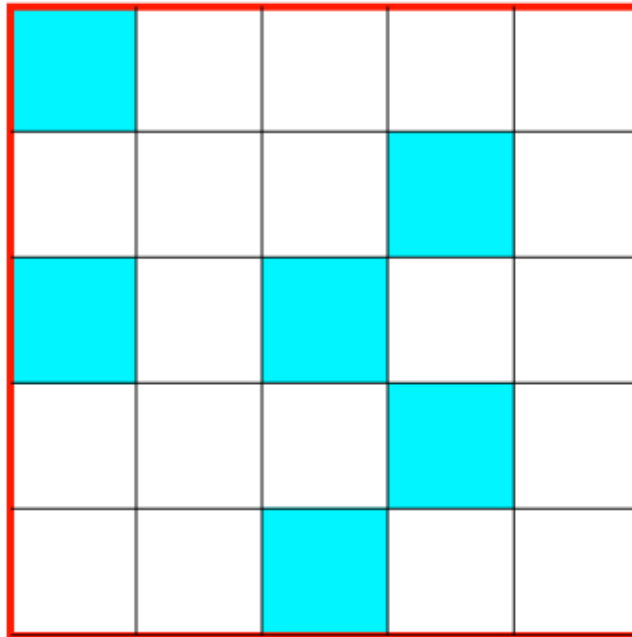
Intensity threshold exceeded where squares are blue

observed



Fraction =  $6/25 = 0.24$

forecast



Fraction =  $6/25 = 0.24$

$P$  is the fractional event frequency  
within the neighborhood.

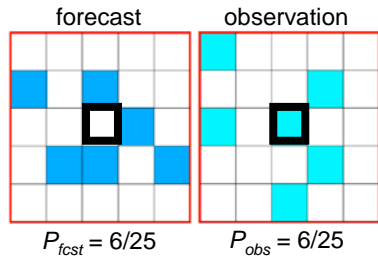
This is calculated for all  $n$  grid points in the domain

copyright 2010 UCAR, all rights reserved

slide from Mittermaier

# Neighborhood Methods

$n =$  grid points in the domain



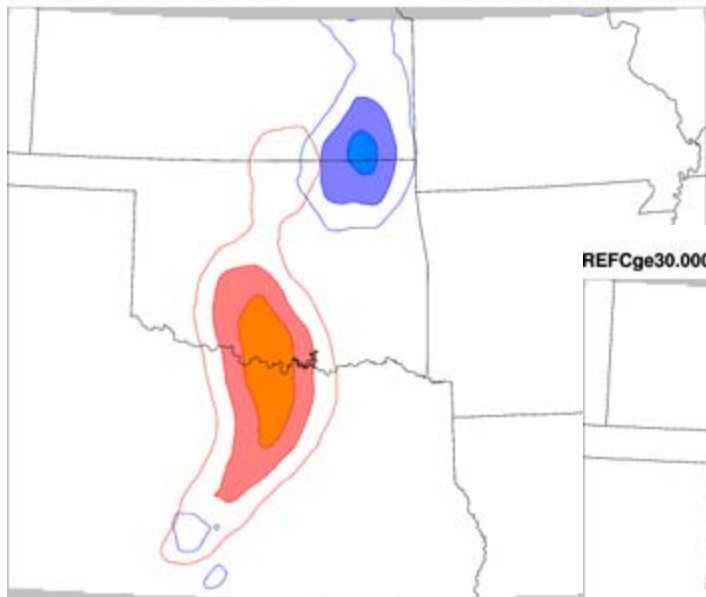
$$1 - \frac{\text{Fractions Brier Score}}{\text{Fractions Brier Score}}$$

(worst possible)

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

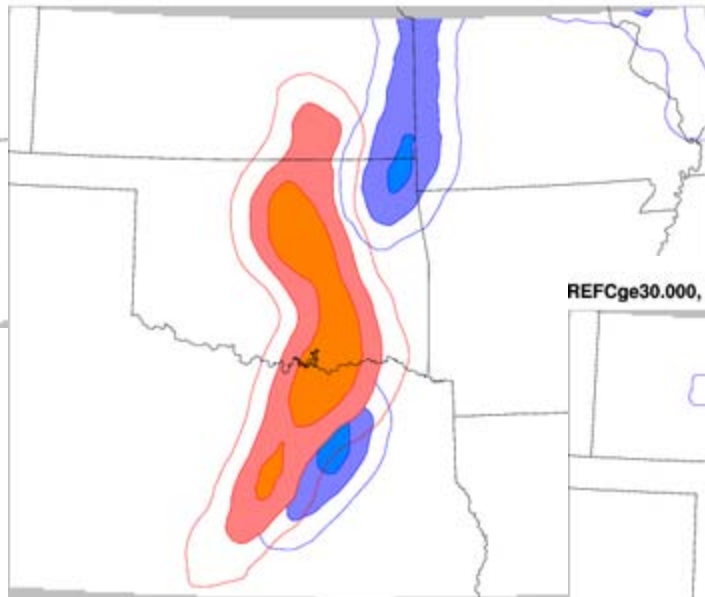
# Example – REFC > 30 dBZ

eastnmm DAILY  
REFCge30.000, FSS(31)=0.13848 Valid 20110511 20Z, 20-hr fcst



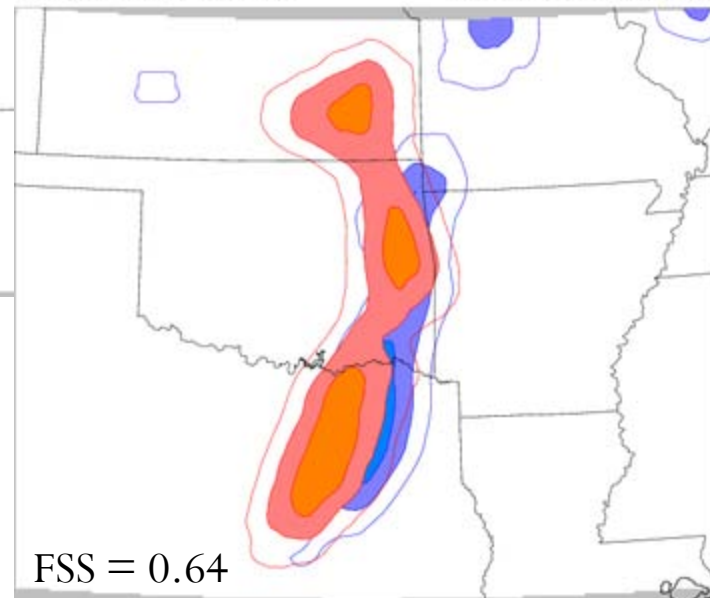
FSS = 0.14

eastnmm DAILY  
REFCge30.000, FSS(31)=0.30176 Valid 20110511 22Z, 22-hr fcst



FSS = 0.30

eastnmm DAILY  
REFCge30.000, FSS(31)=0.64225 Valid 20110512 00Z, 24-hr fcst



FSS = 0.64

Time

Overlap



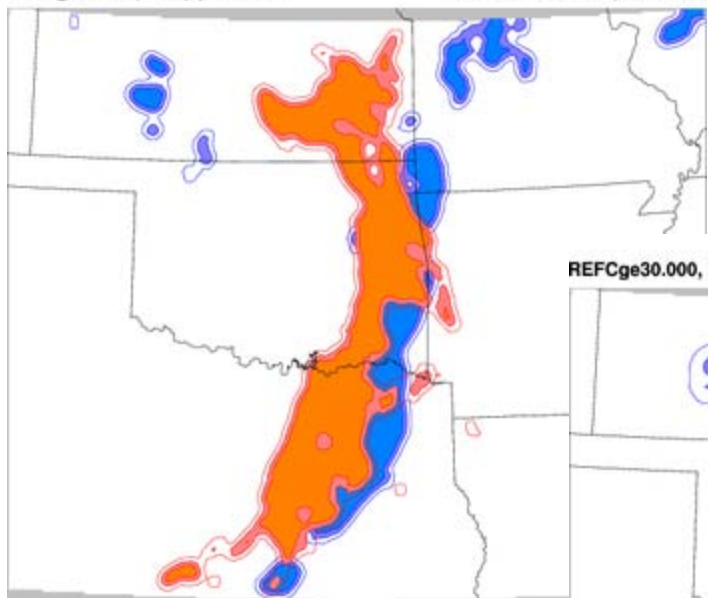
Forecast Fractional Coverage



Observed Fractional Coverage

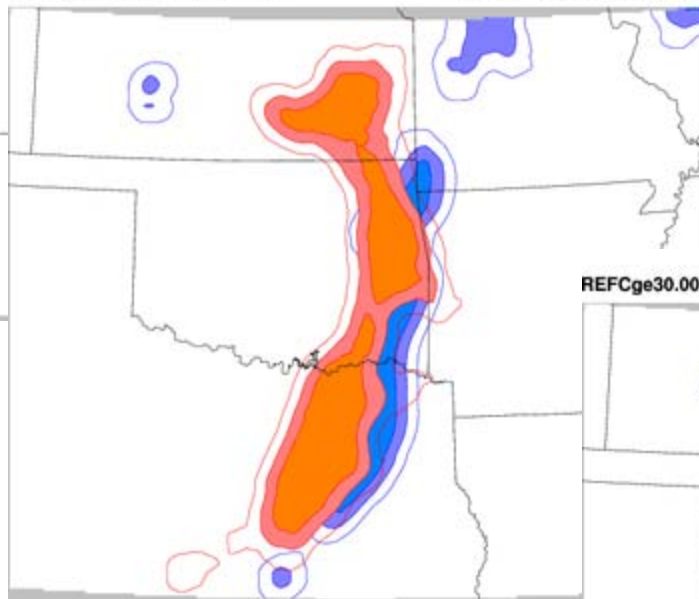
# Example – REFC > 30 dBZ

eastnmm DAILY  
REFCge30.000, FSS(7)=0.45327 Valid 20110512 00Z, 24-hr fcst



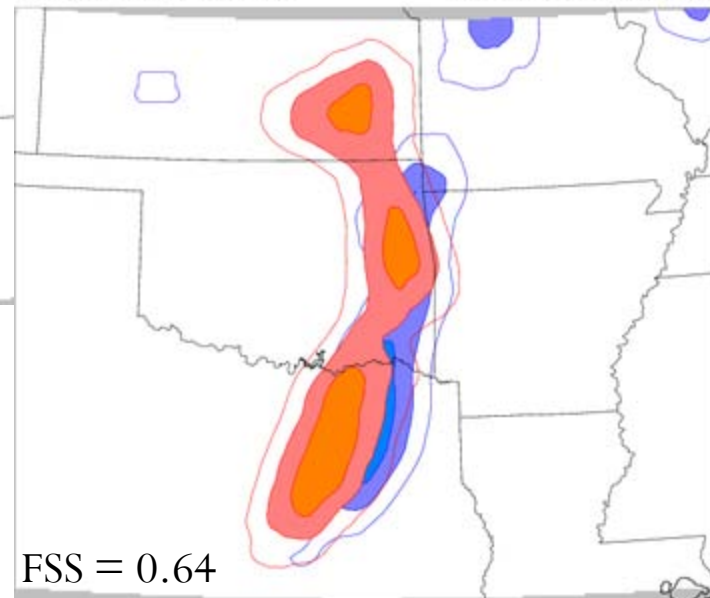
FSS = 0.45

eastnmm DAILY  
REFCge30.000, FSS(19)=0.55727 Valid 20110512 00Z, 24-hr fcst



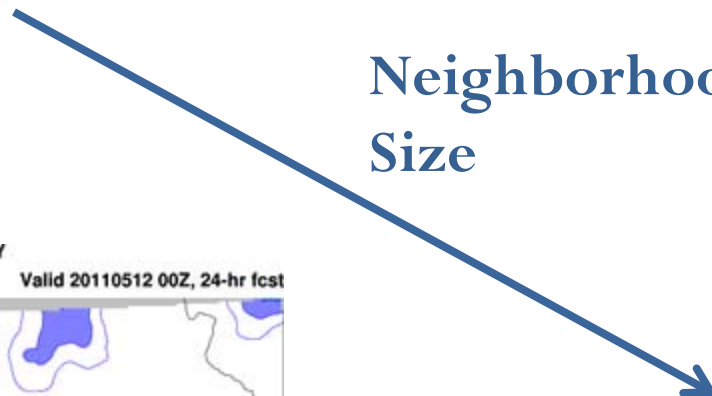
FSS = 0.56

eastnmm DAILY  
REFCge30.000, FSS(31)=0.64225 Valid 20110512 00Z, 24-hr fcst



FSS = 0.64

Neighborhood  
Size



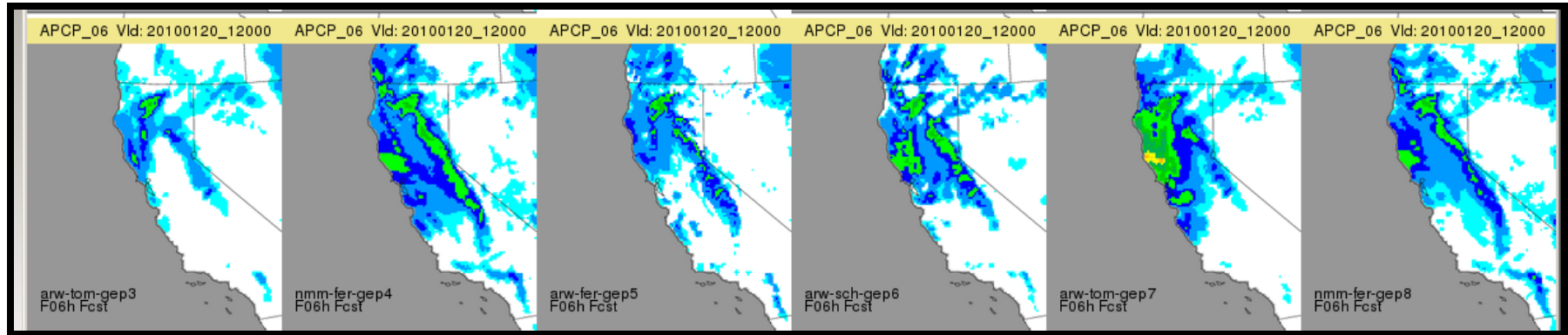
Increase in FSS with increasing neighborhood size is a characteristic of method and can be attributed to the smoothing that occurs

# Ensembles

---

Ensemble Stat, Point Stat, Grid Stat and MODE

# Ensemble Stat: Overview

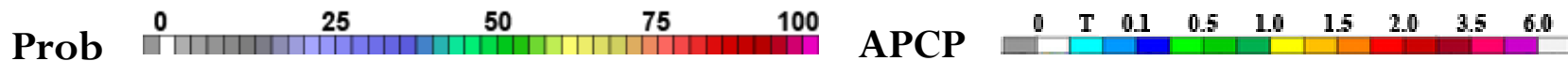
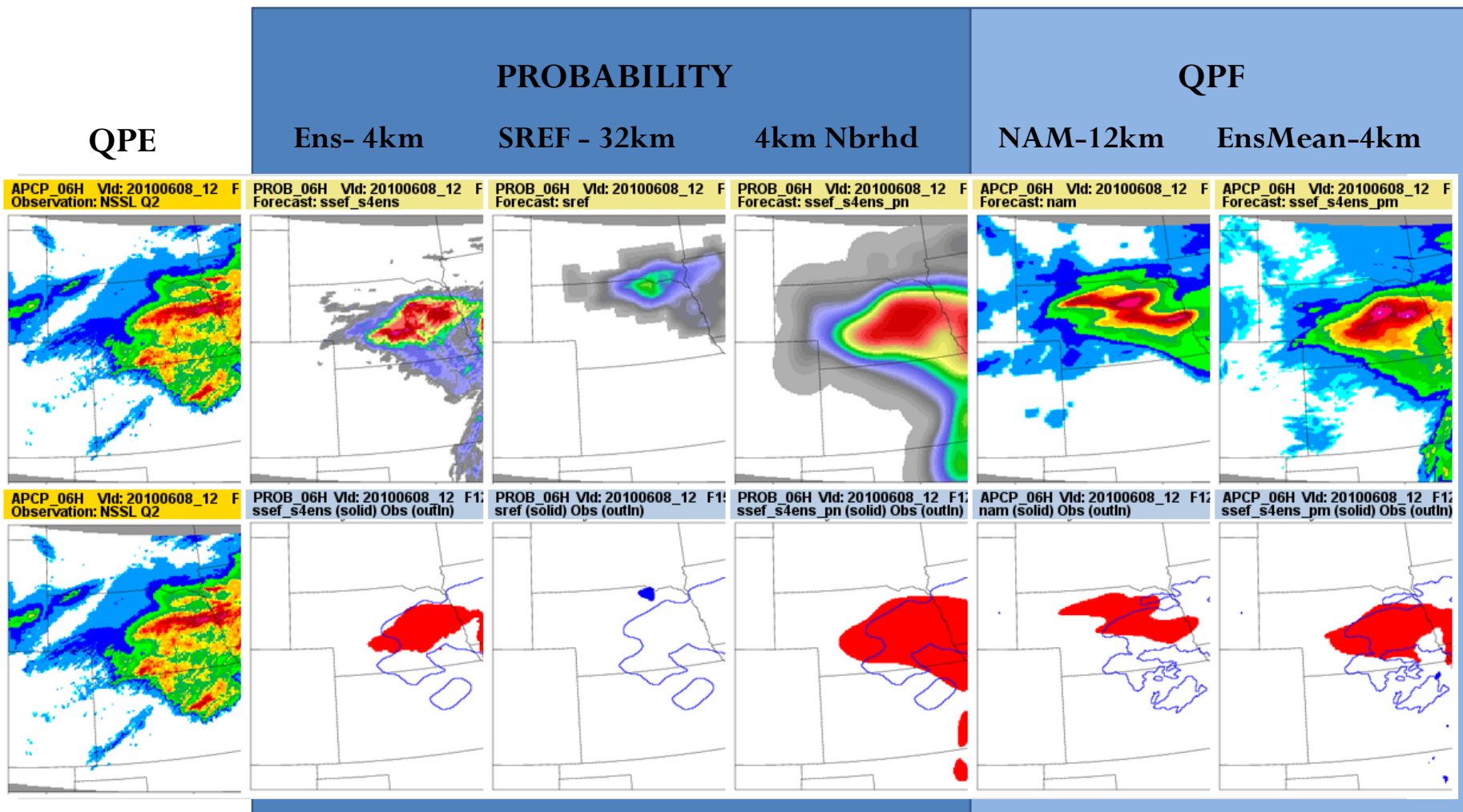


- Ensemble Stat primarily a pre-processing tool
  - Mean, Spread, Mean  $\pm$  1 STDev, Min, Max, Range, Number of valid members
- Calculates rank histograms and outputs:
  - Bins and Counts
  - Matched Pair ranks

# MODE on Probabilistic Data



# Probabilistic Fields (PQPF) and QPF Products



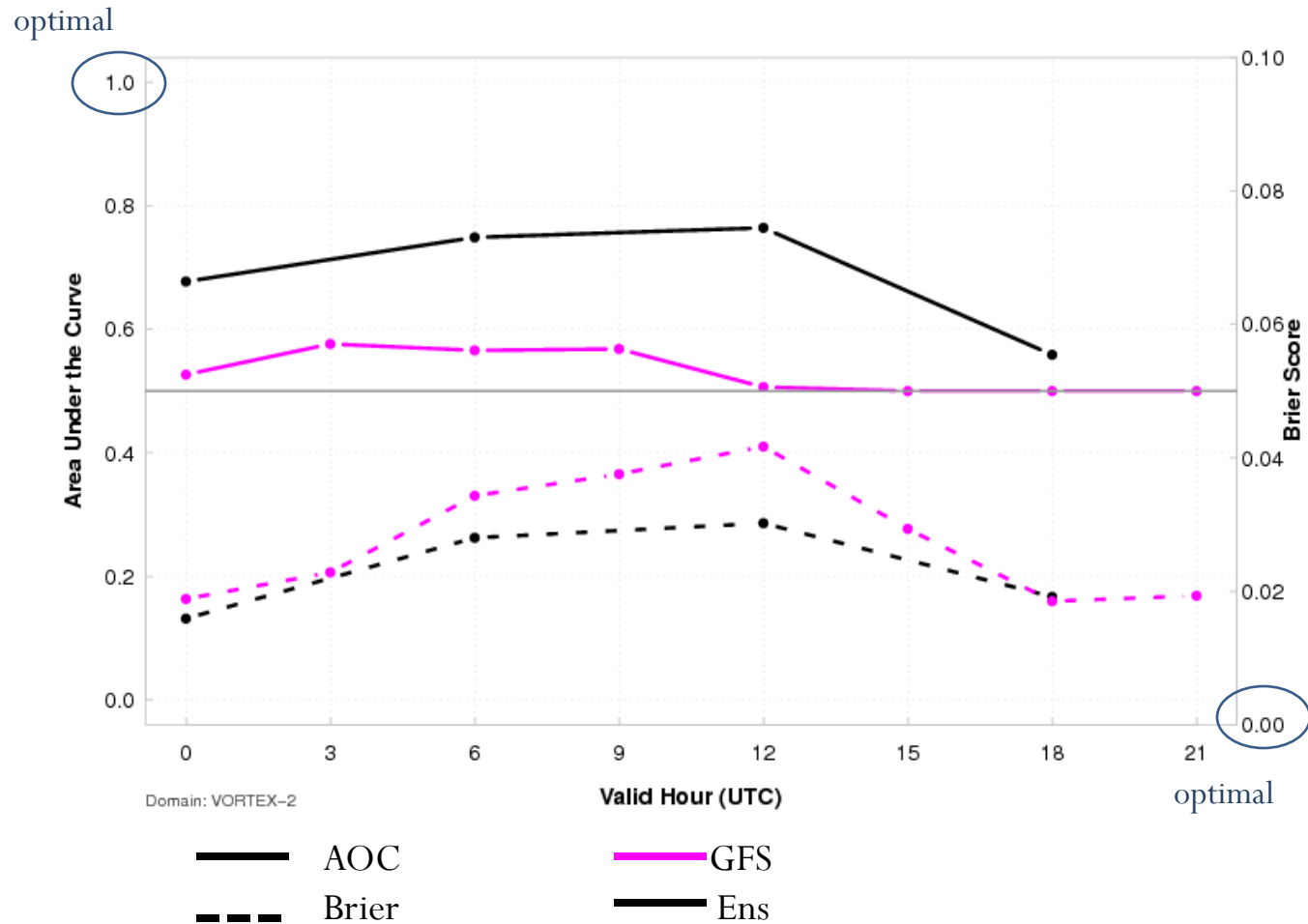
# Probabilistic Metrics in Grid-Stat

# Mesoscale vs. Convection Allowing Ensemble

06hr  
accumulations

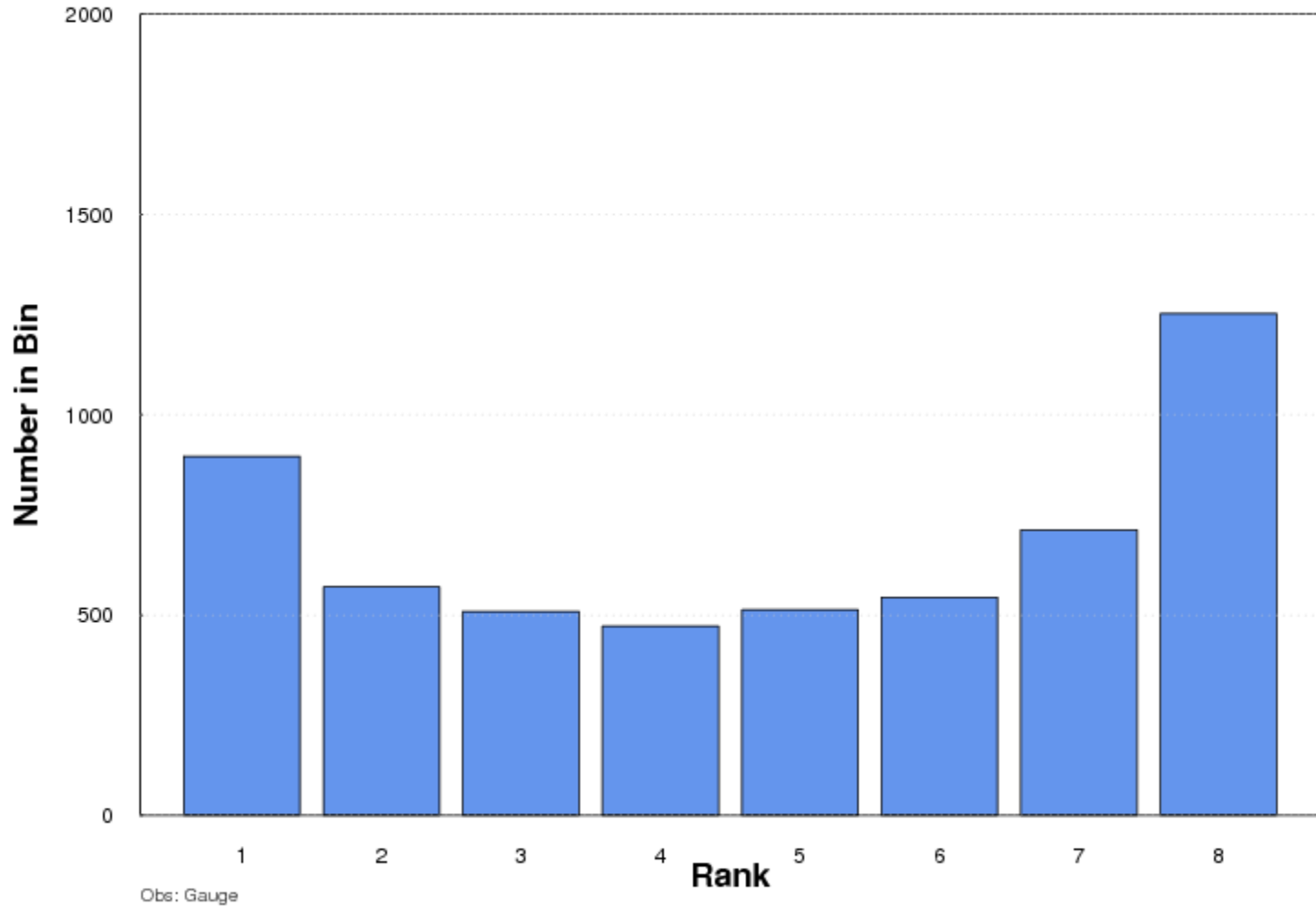
- **SREF PQPF** - minimal skill
- **HWT Ens PQPF** - nominal skill at 0.5"
- **HWT Neighbhd PQPF** – not show has slightly higher skill

HWT 2010 Median Scores for Probability of Precipitation  $\geq 0.5$  inches over 6 hrs



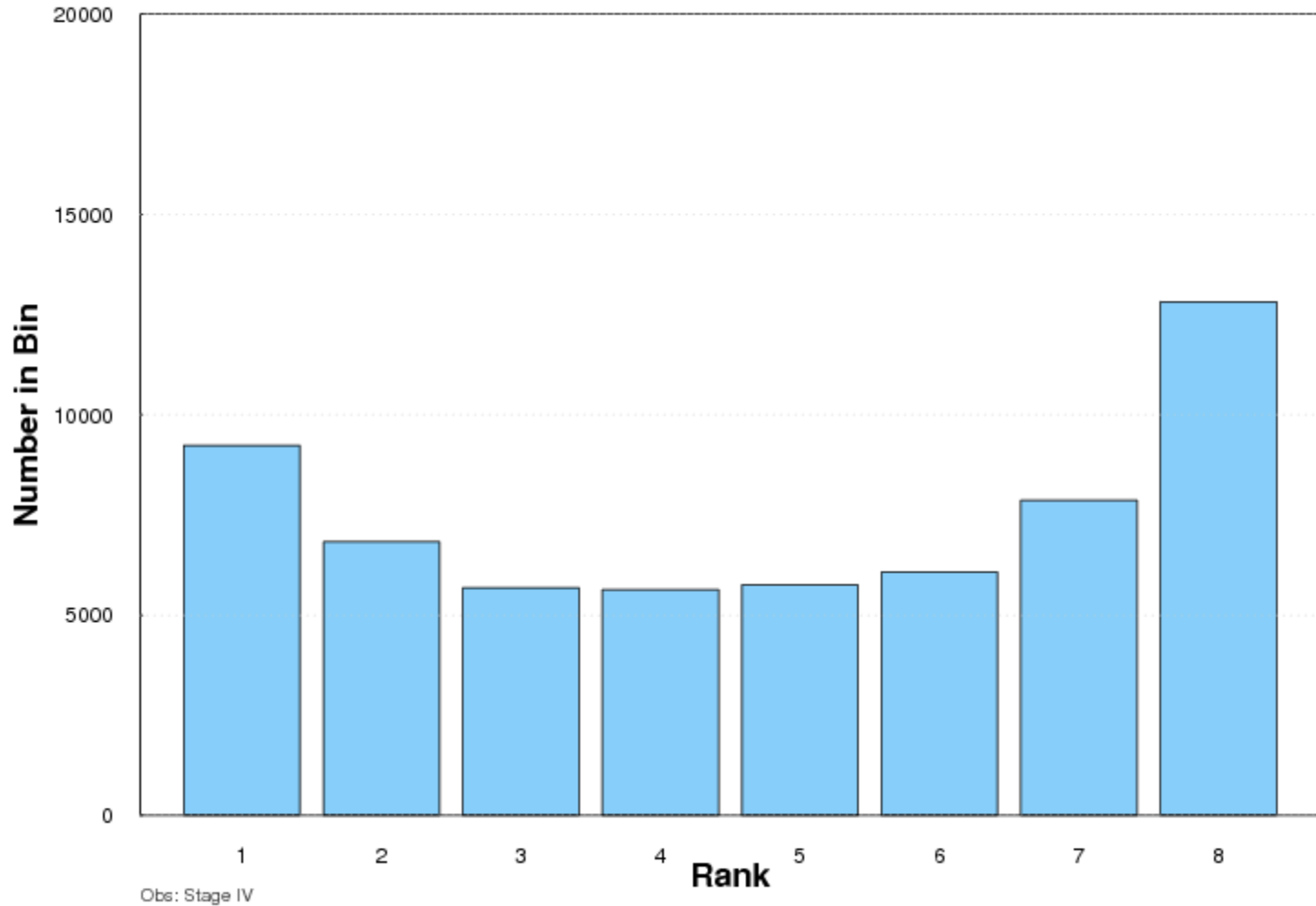
# HMT Ensemble for Extreme Cool Season Precip over Northern CA using Gauge Data

Rank Histogram: APCP\_24 Init: 12 UTC Lead: 24hr FULL Domain 7 members available



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

Rank Histogram: APCP\_24 Init: 12 UTC Lead: 24hr FULL Domain 7 members available



# MET tutorial

June 27 – 28

NCAR Foothills Lab

Registration and Information:

[http://www.dtcenter.org/events/workshops11/met\\_tutorial.php](http://www.dtcenter.org/events/workshops11/met_tutorial.php)

# Thank You 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](mailto:tressa@ucar.edu)  
[jensen@ucar.edu](mailto:jensen@ucar.edu)

Support for the Developmental Testbed Center (DTC),



is provided by  
NOAA, AFWA  
NCAR and NSF

