Methods for Evaluation of Cloud Predictions

Barbara Brown, Tara Jensen, John Halley Gotway,
Kathryn Newman, Eric Gilleland, Tressa Fowler,
and Randy Bullock

7th International Verification Methods Workshop
Berlin, Germany
10 May 2017

Motivation and Goals

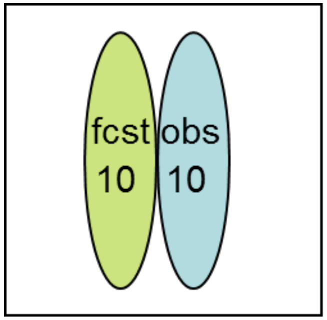
- Motivation
  - Clouds have important impacts on activities of the US Air Force and are a prime focus of the 557th Weather Wing
  - Skill of cloud forecasts impact decision making (e.g., uncertainty in cloud cover predictions can change operational decisions)
- Goals
  - *Long-term*: Create a meaningful cloud verification "index" for AF applications
  - *Short-term*: Identify useful components of such an index

Approach

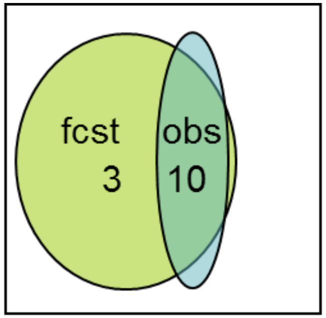
1. **Standard methods** based on traditional metrics (continuous, categorical)
2. Investigate **object-based and distance metrics** to provide forecast quality information that
   - Provides diagnostic, user-relevant information
   - Includes methods not subject to “hazards” of traditional verification (e.g., entanglement of spatial displacement with other errors)

Initial focus on CONUS, fractional coverage (TCA = Total Cloud Amount)

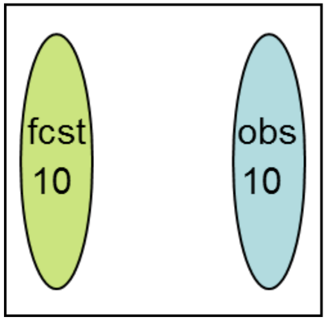
Secondary: Global forecasts



Hi res forecast
RMS ~ 4.7
POD=0, FAR=1
TS=0



Low res forecast
RMS ~ 2.7
POD~1, FAR~0.7
TS~0.3



Verification Questions

- Which methods provide useful information about the performance of cloud forecasts?
- Do spatial methods have a role to play in evaluation of clouds?
- Would distance metrics be a useful addition to the cloud verification toolbox?

Conclusions First…

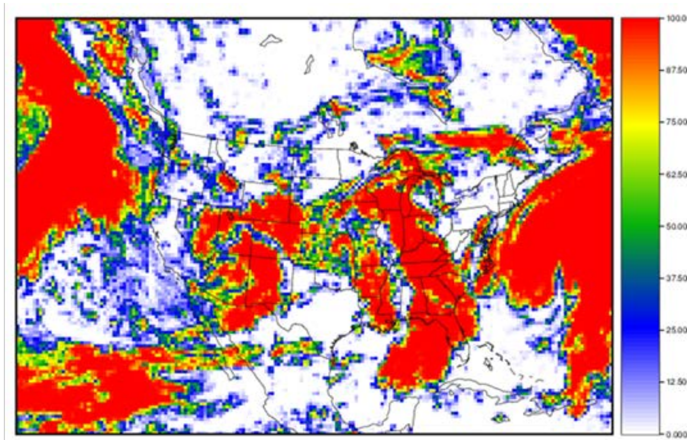
- ~~**Continuous methods** (RMSE, MAE, etc.) do not provide much useful information regarding TCA performance – primarily due to discontinuous nature of clouds~~
  - ~~Edges~~
  - ~~Tendency of products toward 0 or 100% values~~
- **Point observations are less useful overall than satellite-based analyses due to limited availability globally**
- **Categorical methods** (POD, FAR, etc.) are more useful for answering relevant questions about cloud occurrence
  - Especially when presented in a diagnostic multivariate form
- **Object-based methods** have promise of providing useful information – when configured appropriately
- **Distance metrics** can provide interesting diagnostic information – but need to be explored more

Observations, Analyses, and Forecasts

NCAR

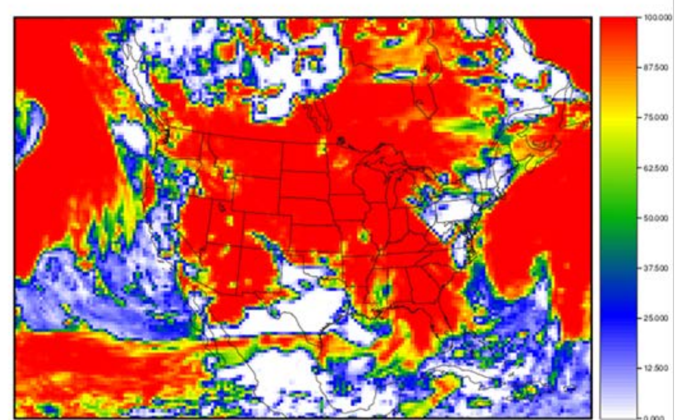
- **“Observations” and Analyses**
  - WWMCA (gridded World-Wide Merged Cloud Analysis)
  - WWMCA-R (WWMCA updated in post-analysis with all obs available)
- **Forecasts**
  - 2 global models (72 h)
    - **GALWEM** (AF implementation of UK Unified Model)
    - **GFS** (NCEP Global Forecast System)
  - **DCF** (Diagnostic Cloud Forecast)
    - Bias-corrected GALWEM and GFS
  - **ADVCLD**: Advection (persistence) model (9 h)
- Sample data for 4 seasons (1 week each)
- NCEP grid 212 (polar stereographic; 40 km)
- Model Evaluation Tools (MET) and Spatial-Vx R package used for all analyses

WWMCA



grid_stat_REGRID_G212_ADVCLD_vs_WWMCA_000000L_20160108_030000V_pairs.nc

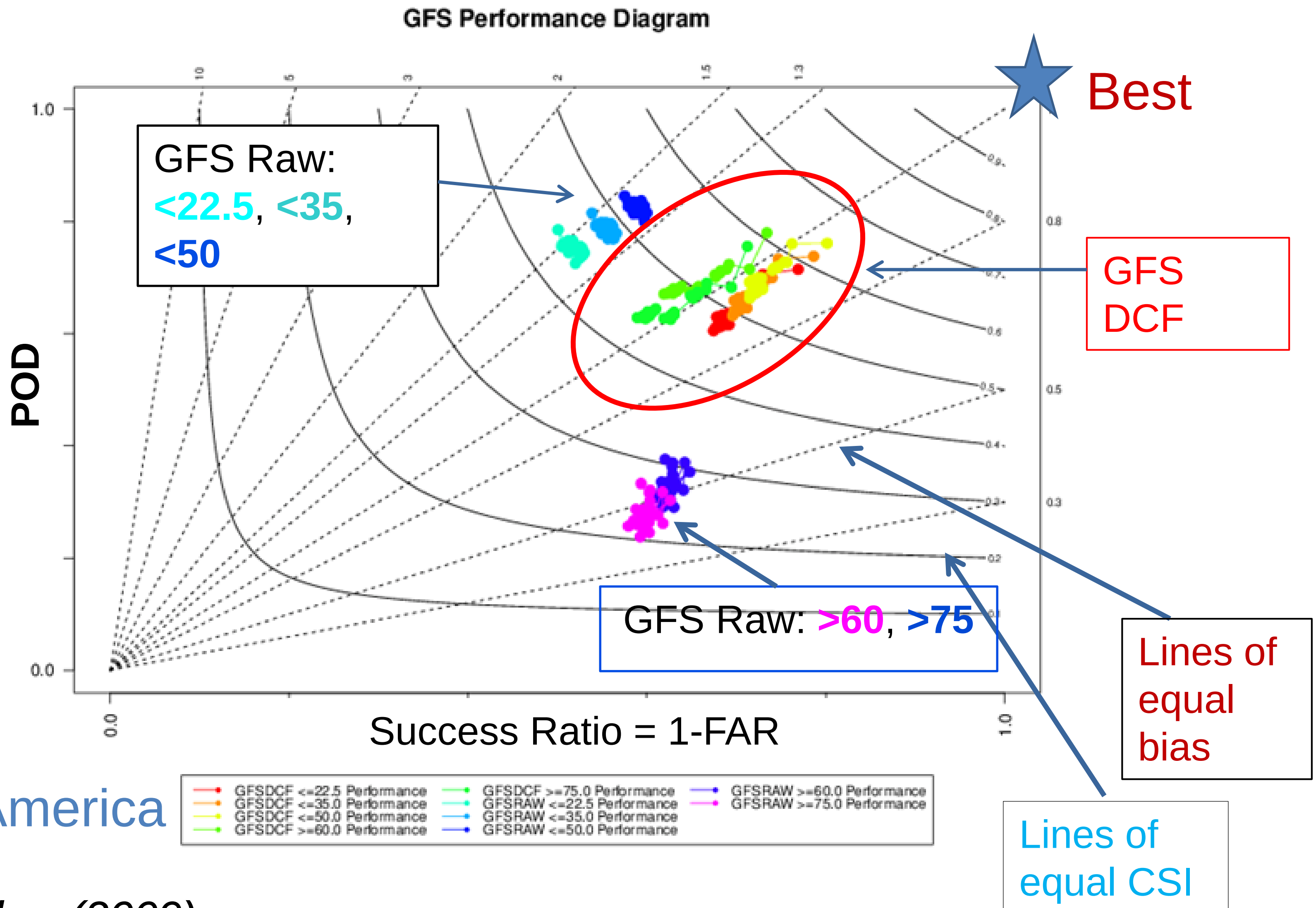
GALWEM



grid_stat_REGRID_G212_ADVCLD_vs_GALWEM_000000L_20160108_030000V_pairs.nc

Gridded comparisons: Categorical statistics

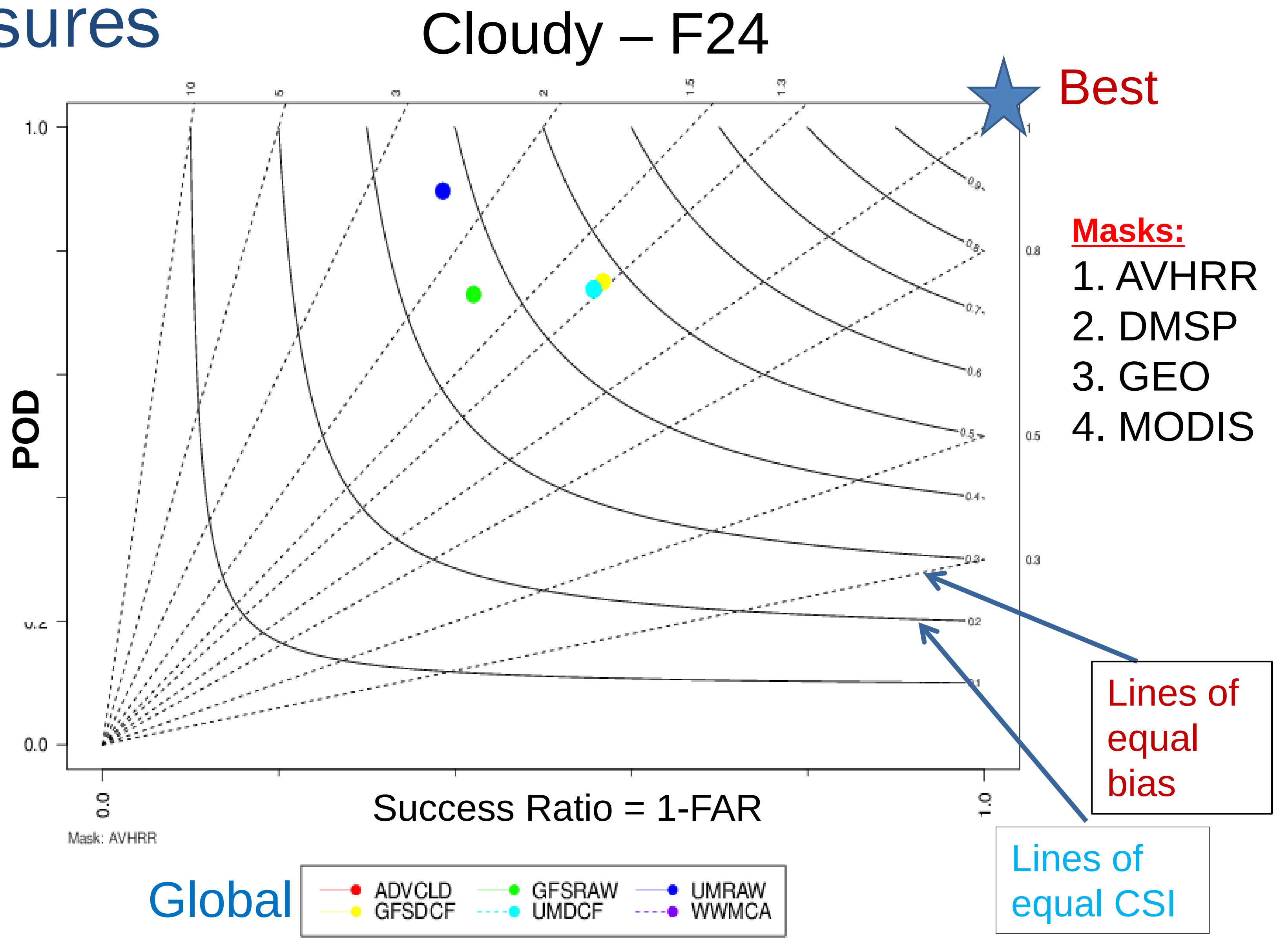
Performance Diagrams using WWMCA-R as the verification grid



N. America

After Roebber (2009)

Performance Diagram: Multiple Categorical Measures



Models:
GFSDCF
GFSRAW
UMDCF
UMRAW

Analysis:
World Wide Merged Cloud Analysis (WWMCA) -reanalysis

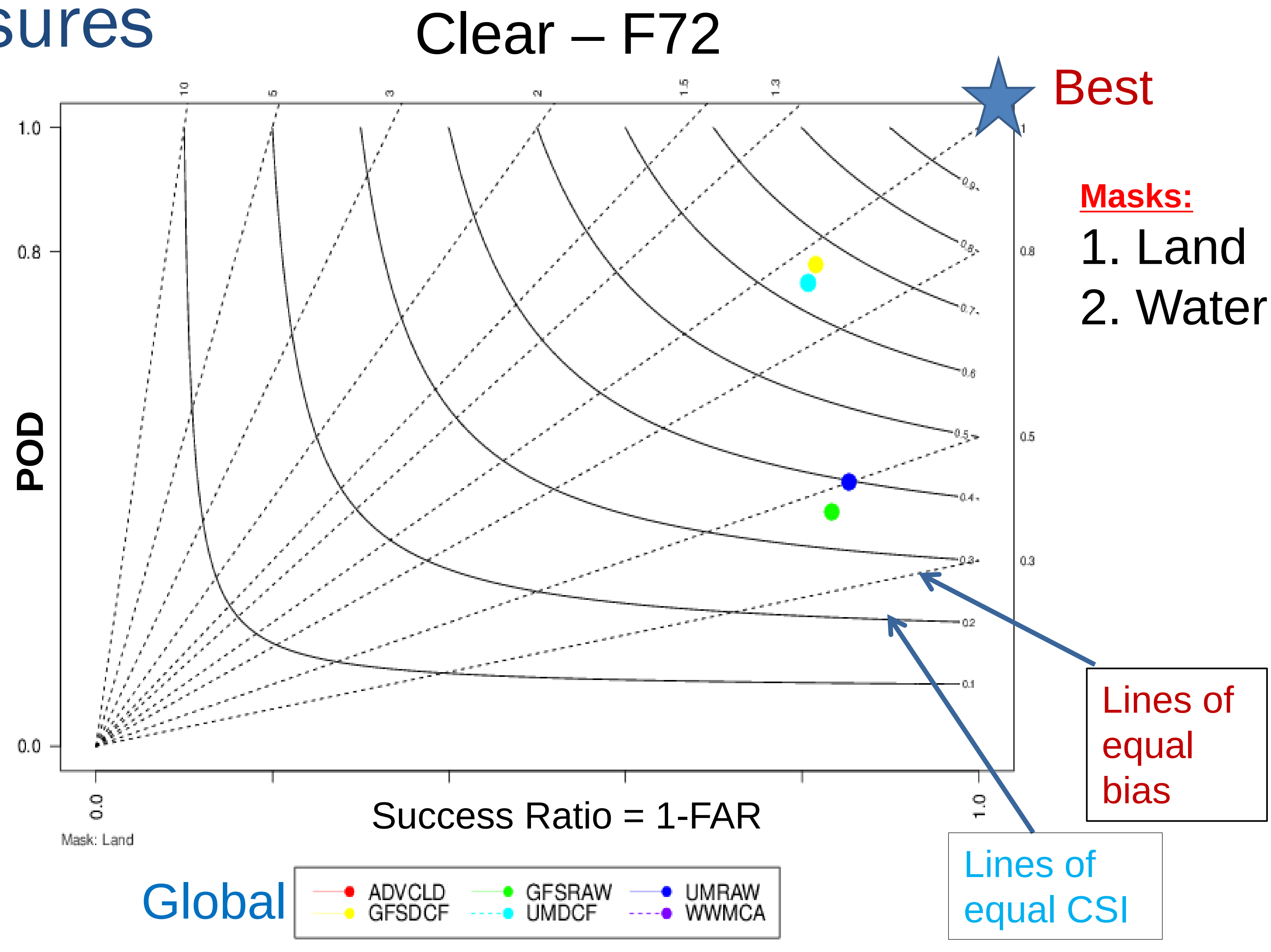
Masks:
1. AVHRR
2. DMSP
3. GEO
4. MODIS

Global

Performance Diagram: Multiple Categorical Measures

Models:
GFSDCF
GFSRAW
UMDCF
UMRAW

Analysis:
World Wide Merged Cloud Analysis (WWMCA) -reanalysis

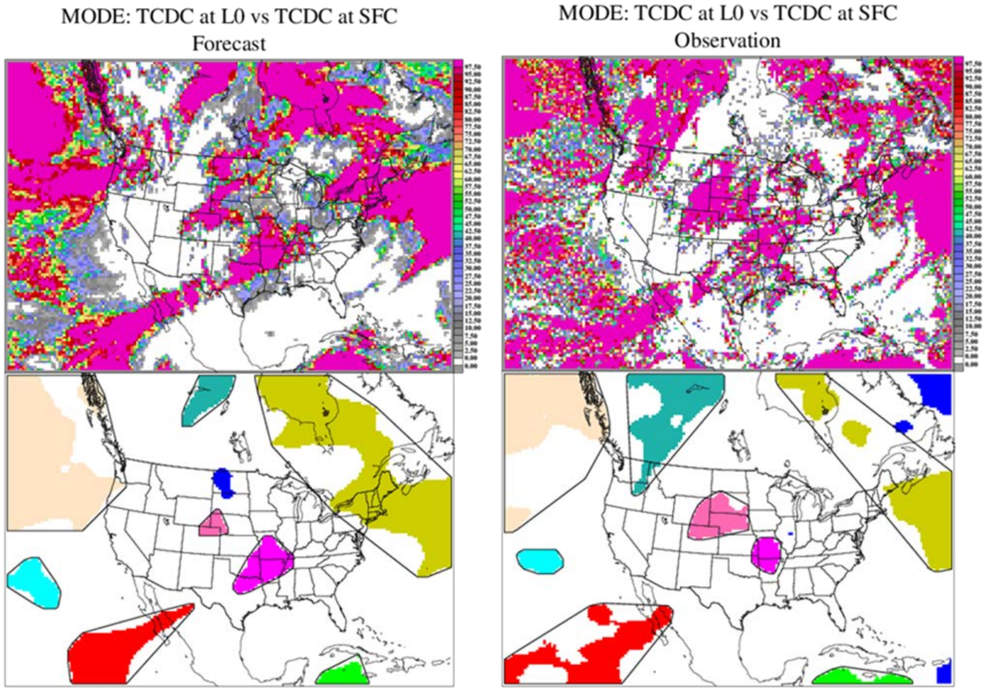


Masks:
1. Land
2. Water

Application of MODE

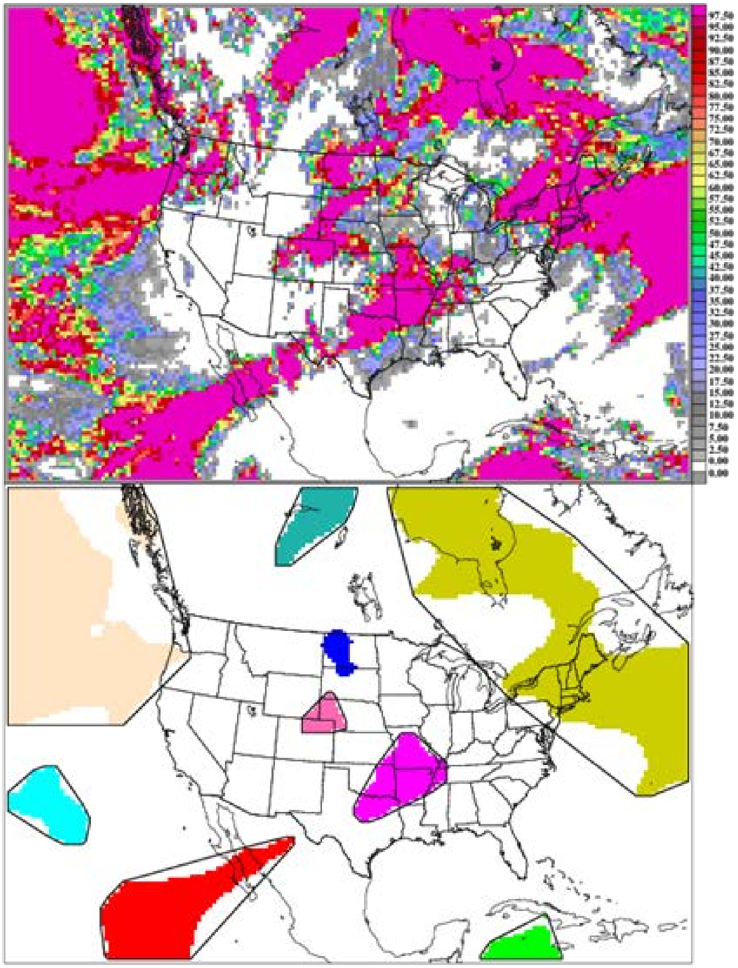
MODE (Method for Object-based Diagnostic Evaluation) process:

- Identify ***relevant*** features in obs and forecast fields
- Use fuzzy logic engine to ***match clusters*** of forecasts and observed features
- ***Summarize characteristics*** of objects and differences between pairs of objects

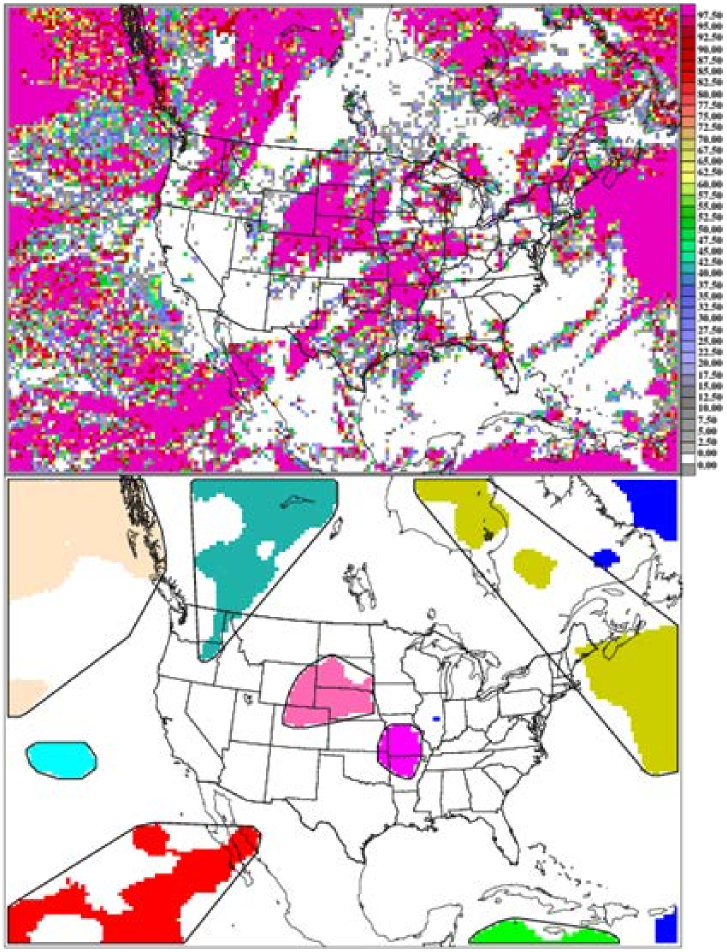


MODE Object-Based Approach

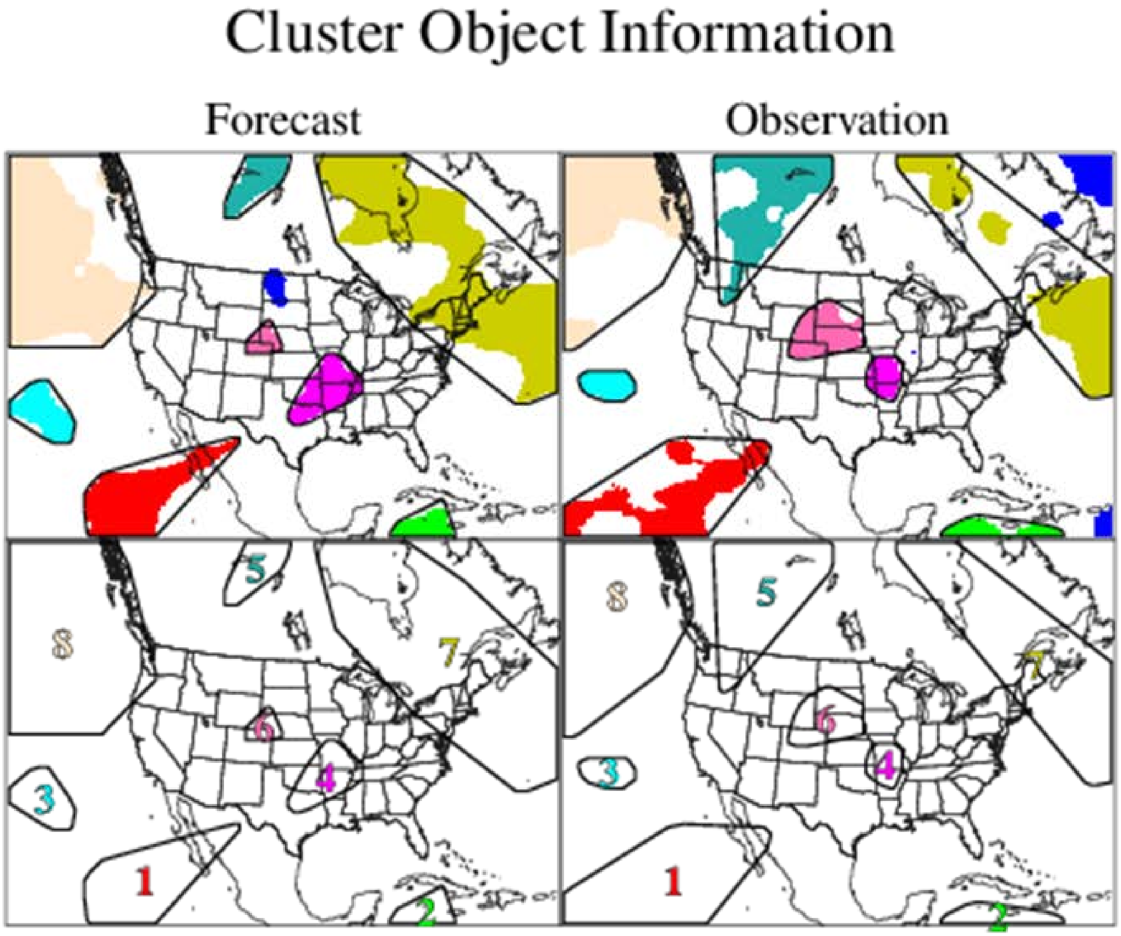
GALWEM



WWMCA



11 November 2015; Cloudy Threshold (TCA > 75)

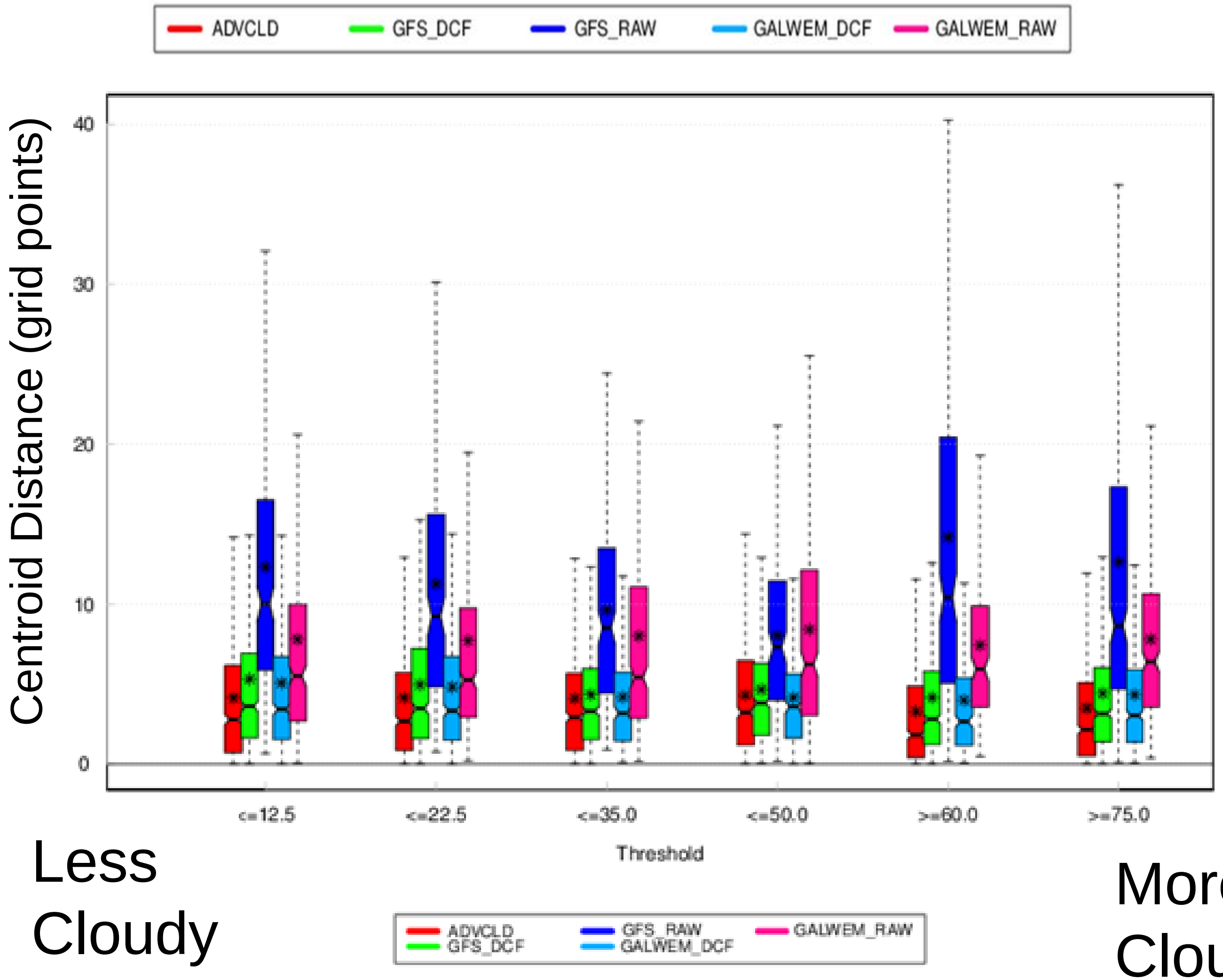


- Some displacement of all clusters
- Large area differences, for some objects

… Etc.

CLUS PAIR	CEN DIST	ANG DIFF	FCST AREA	OBS AREA	INTER AREA	UNION AREA	SYMM DIFF	FCST INT 50	OBS INT 50	FCST INT 90	OBS INT 90	TOT INTR
1	8.53	10.08	689	816	504	1001	497	100.00	100.00	100.00	100.00	1.0000
2	6.18	10.69	131	138	87	182	95	100.00	100.00	100.00	100.00	1.0000
3	9.80	35.64	247	145	33	359	326	89.00	100.00	100.00	100.00	0.9411
4	4.69	51.94	299	130	121	308	187	100.00	100.00	100.00	100.00	0.9158
5	16.56	13.02	229	829	196	862	666	100.00	100.00	100.00	100.00	0.9018
6	3.47	19.33	81	305	81	305	224	100.00	100.00	100.00	100.00	0.8958
7	11.74	2.27	2366	1049	1001	2414	1413	100.00	100.00	100.00	100.00	0.9407
8	15.77	38.71	1921	1157	773	2305	1532	100.00	100.00	100.00	100.00	0.9607

Example MODE summary result: Centroid Distance



Less Cloudy

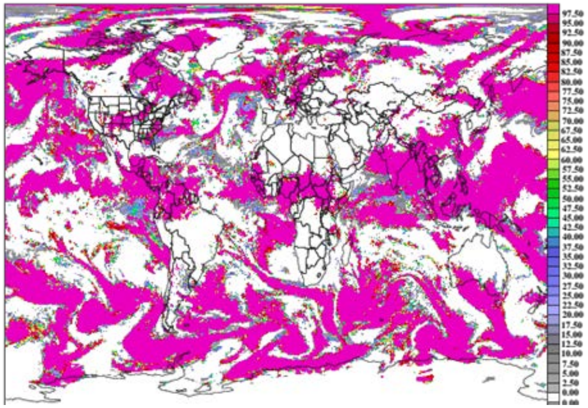
More Cloudy

Global MODE

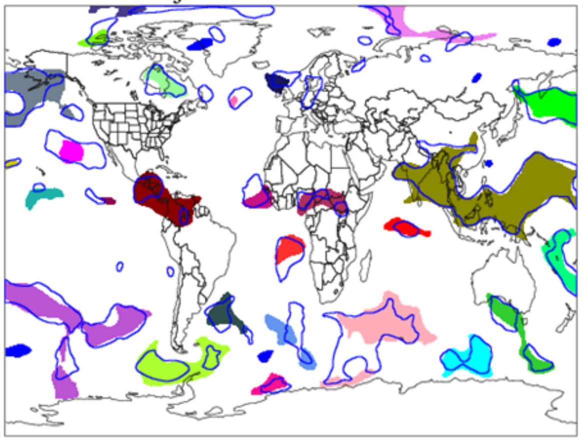
Adjustments for Global application of MODE:

- Larger convolution radius
- Changes in weights and interest values for centroid distance and area ratio for matching

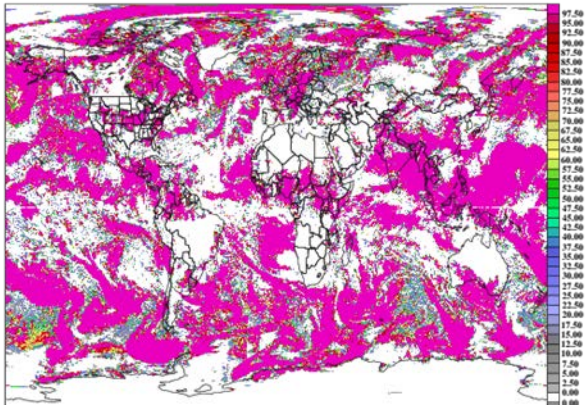
Cloudy



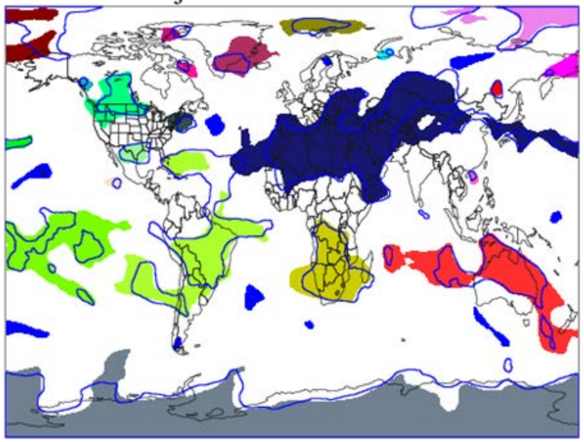
Forecast Objects with Observation Outlines



Clear

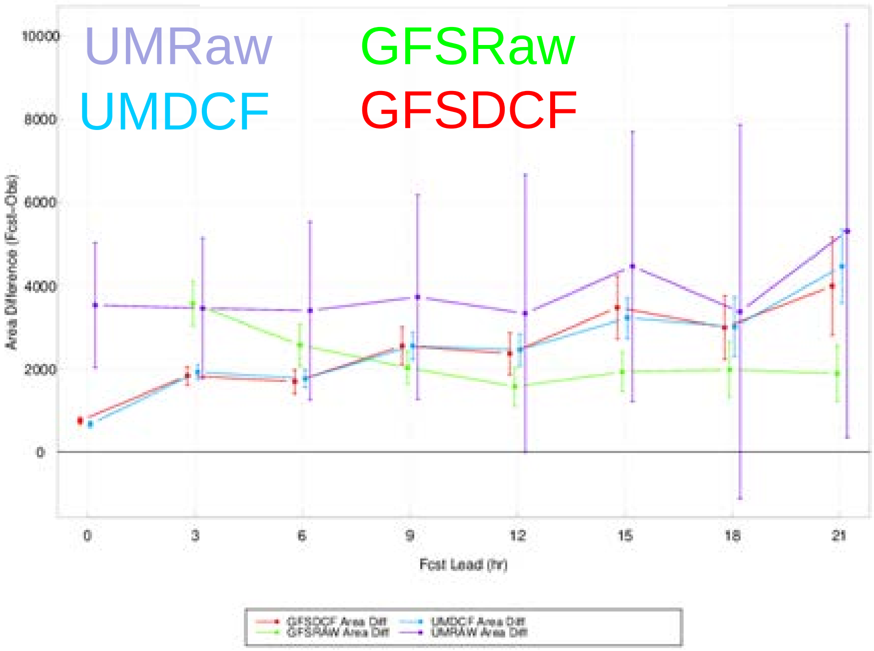


Forecast Objects with Observation Outlines

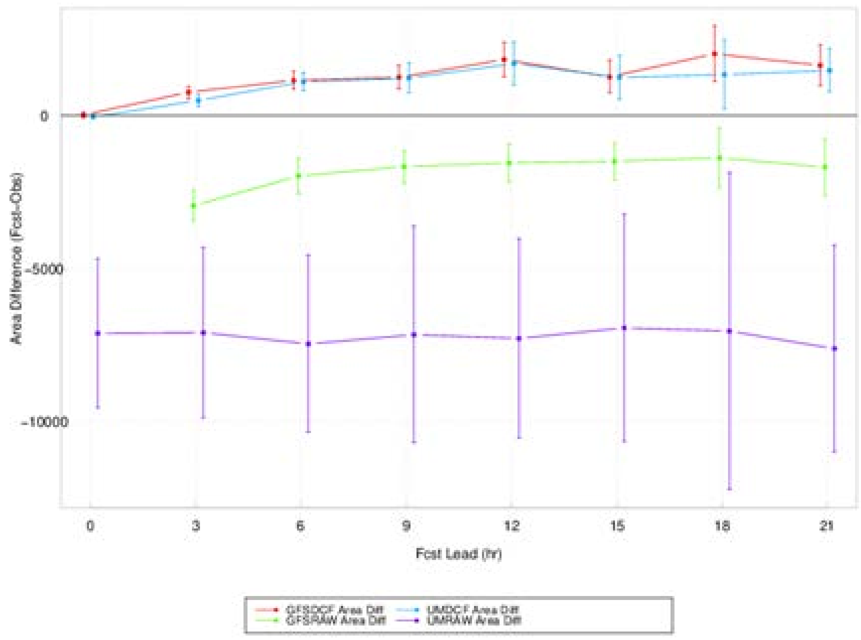


Global MODE Cluster Areas

Cloudy



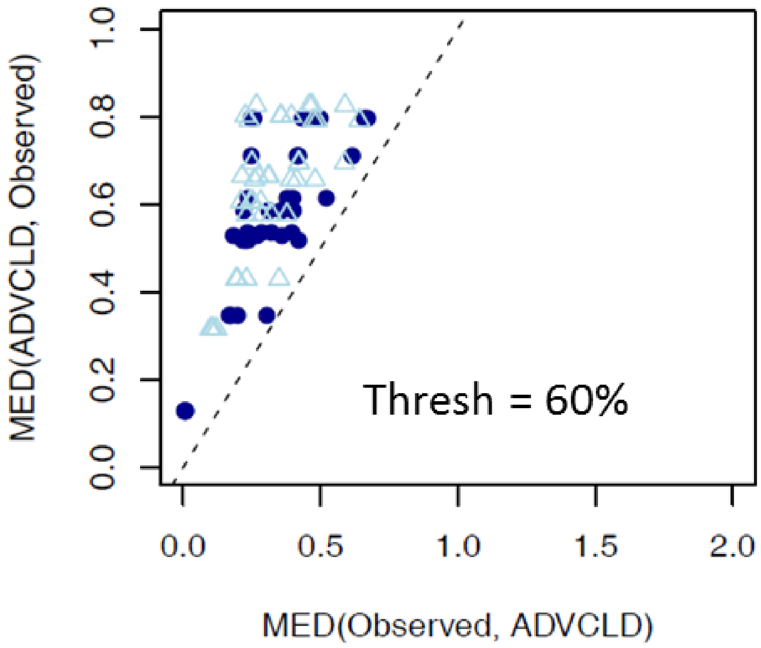
Clear

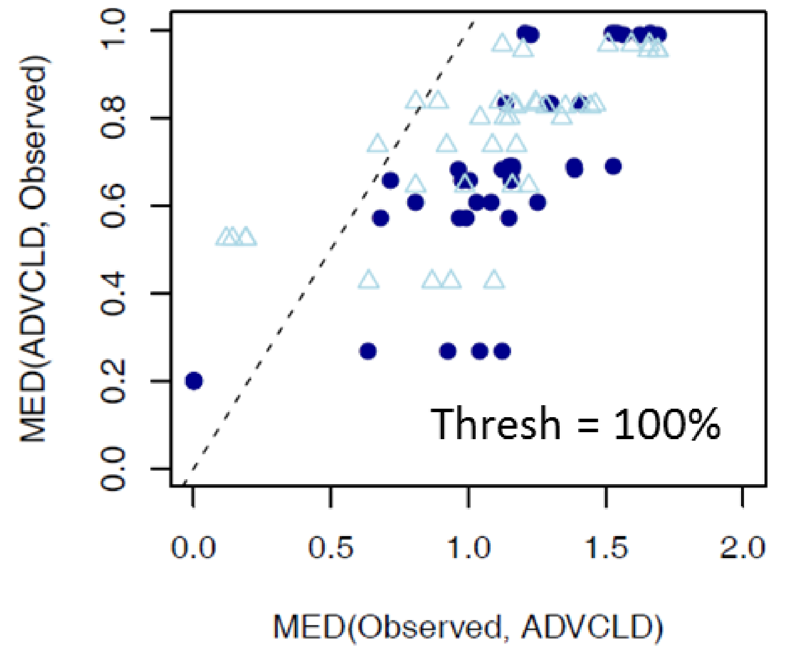


No Pairwise significant differences for Cloudy Cluster Areas

All Pairwise differences for Raw models significant for Clear Cluster Areas

Mean Error Distance





Examine average error distance from all obs points to the nearest forecast point [**MED(forecast, obs)**], and from all forecast points to the nearest obs point [**MED(obs, forecast)**]

- Above diagonal: Misses
- Below diagonal: False alarms

Other promising approaches:

- ***Hausdorff and Baddeley Delta metrics***
- ***Image warping***
- ***Geometric measures***

Gilleland 2017 (WAF)

Conclusions

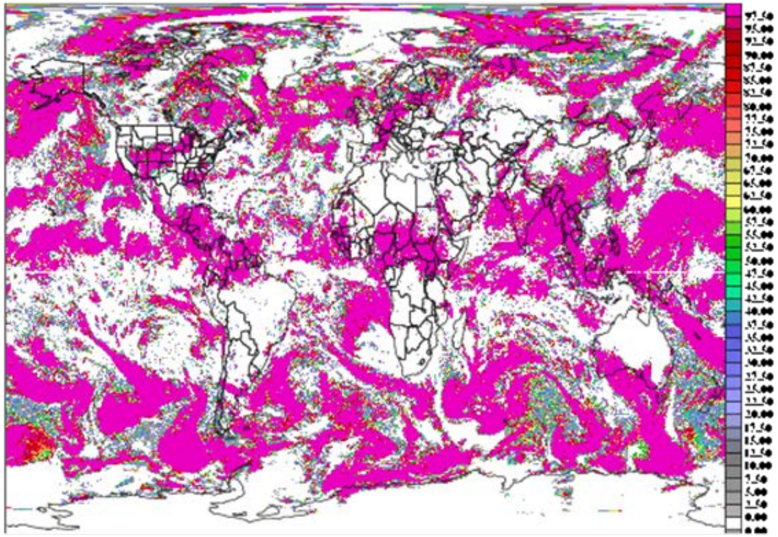
- Categorical methods are the most useful "traditional" approach for evaluating TCA
  - ***Diagnostic plots*** (box plots, performance diagrams) aid in interpretation of results

- Spatial and distance metrics have many benefits and are promising approaches
  - ***MODE configurations*** *depend greatly on scale of evaluation (e.g., global vs. regional)*

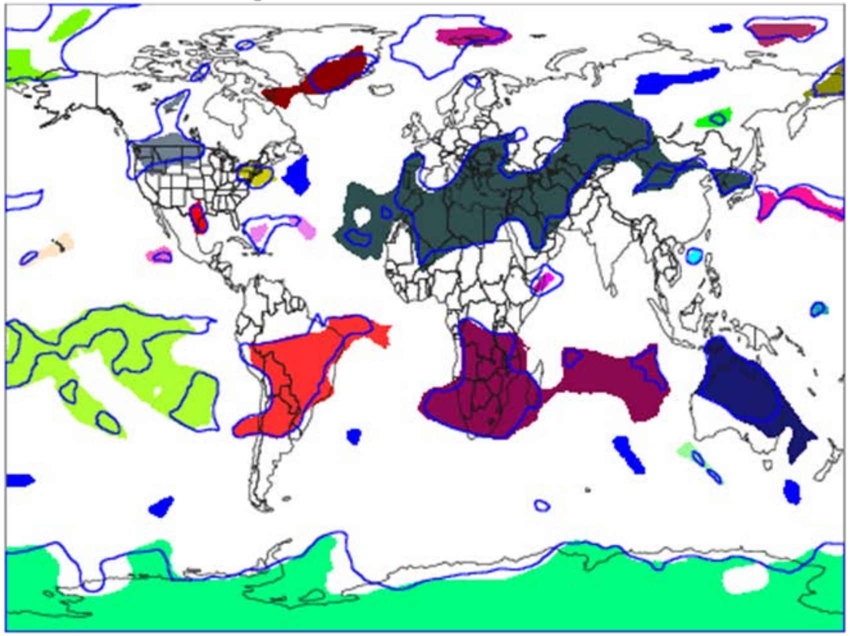
- On a global scale, MODE is especially useful for ***evaluation of non-cloudy areas***

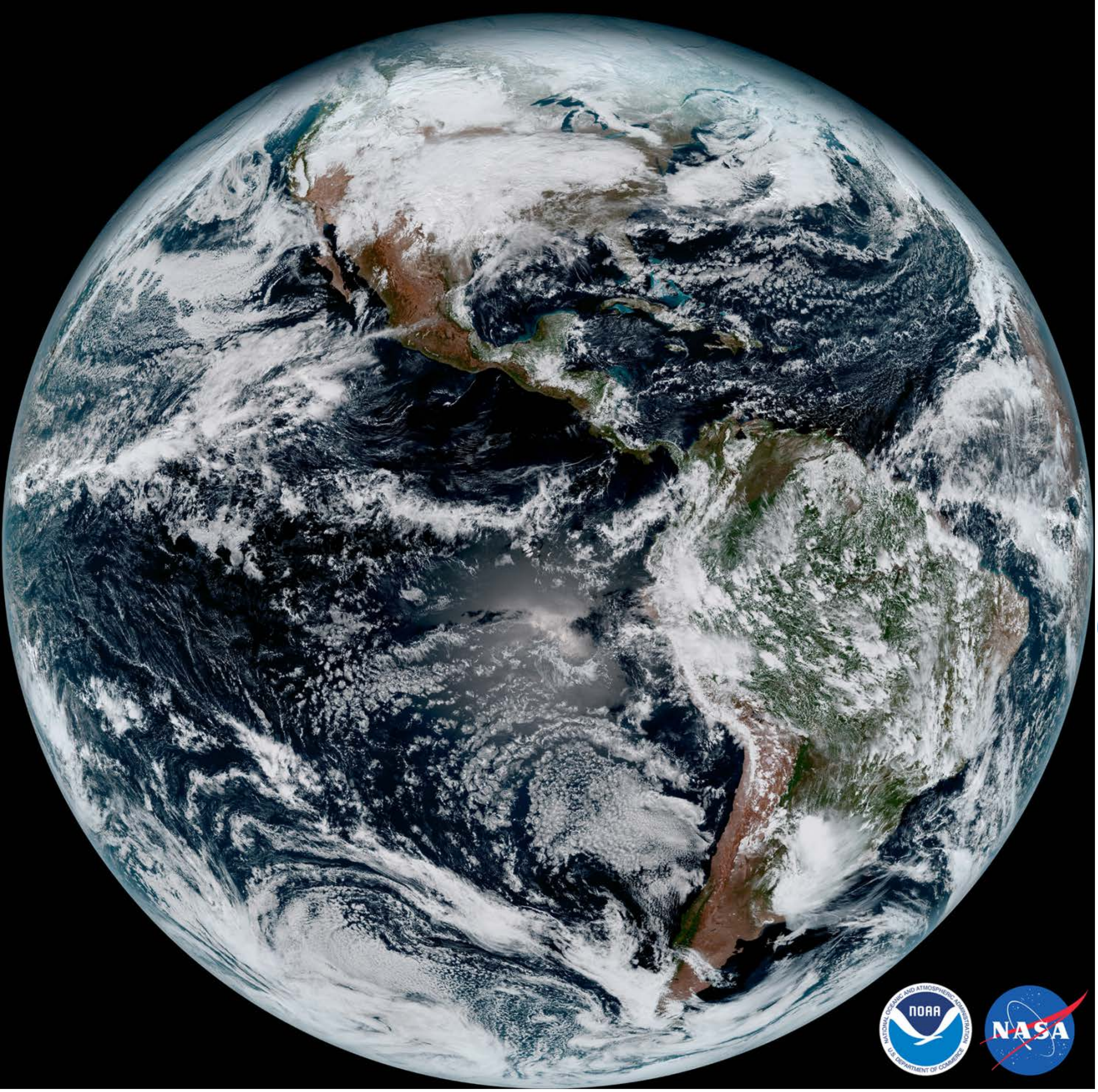
MODE: TCDC at L0 vs TCDC at SFC

Observation



Forecast Objects with Observation Outlines





Thank You

DTC
Developmental Testbed Center

NCAR