Model Verification using satellite observations and object-based verification methods

Sarah M. Griffin *

Jason A. Otkin *

Christopher M. Rozoff ❖

Justin M. Sieglaff *

Lee M. Cronce *

Curtis R. Alexander ★

Tara L. Jensen ^

Jamie K. Wolff ^

* Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin–Madison, Madison, WI

❖Research Applications Laboratory, National Center for Atmospheric Research, Boulder, CO

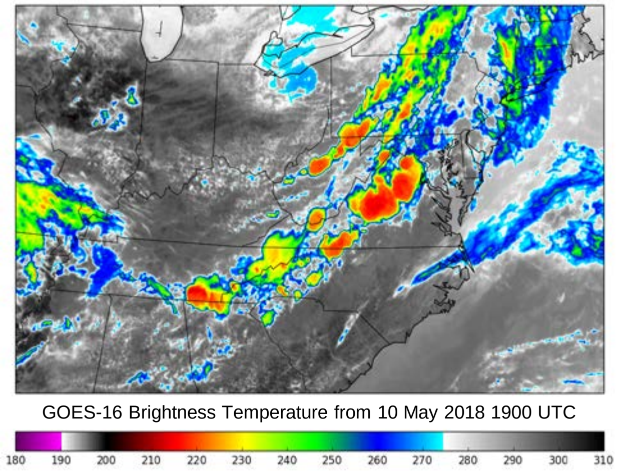
★ NOAA/Earth System Research Laboratory, Boulder, CO

^ Research Applications Laboratory, National Center for Atmospheric Research, and Developmental Testbed Center, Boulder, CO

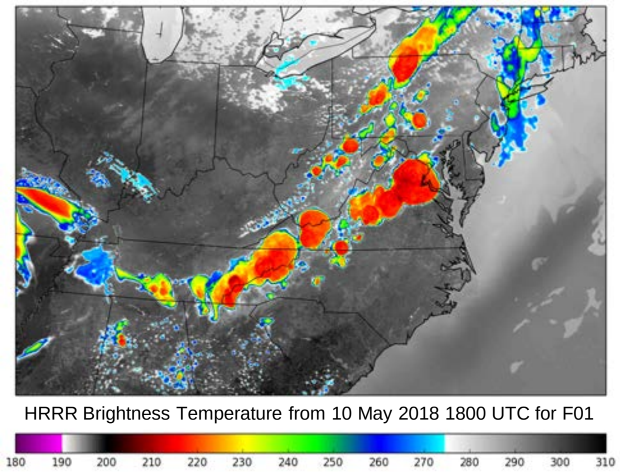
Motivation

Many traditional metrics for assessing forecast accuracy are point-based metrics.

What would you define as the most accurate forecast?

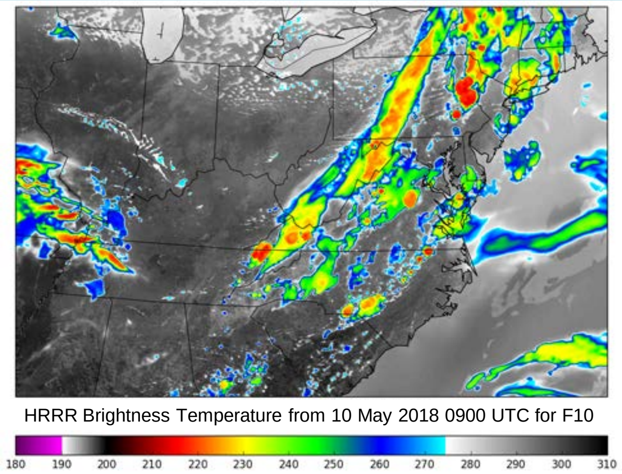


GOES-16 Brightness Temperature from 10 May 2018 1900 UTC



HRRR Brightness Temperature from 10 May 2018 1800 UTC for F01

Mean Absolute Error = 9.18 K
Mean Bias Error = 6.71 K



HRRR Brightness Temperature from 10 May 2018 0900 UTC for F10

Mean Absolute Error = 13.08 K
Mean Bias Error = 1.69 K

How can we assess how accurately a model is predicting cloud features when traditional metrics can be biased by non-cloud features or displacement?

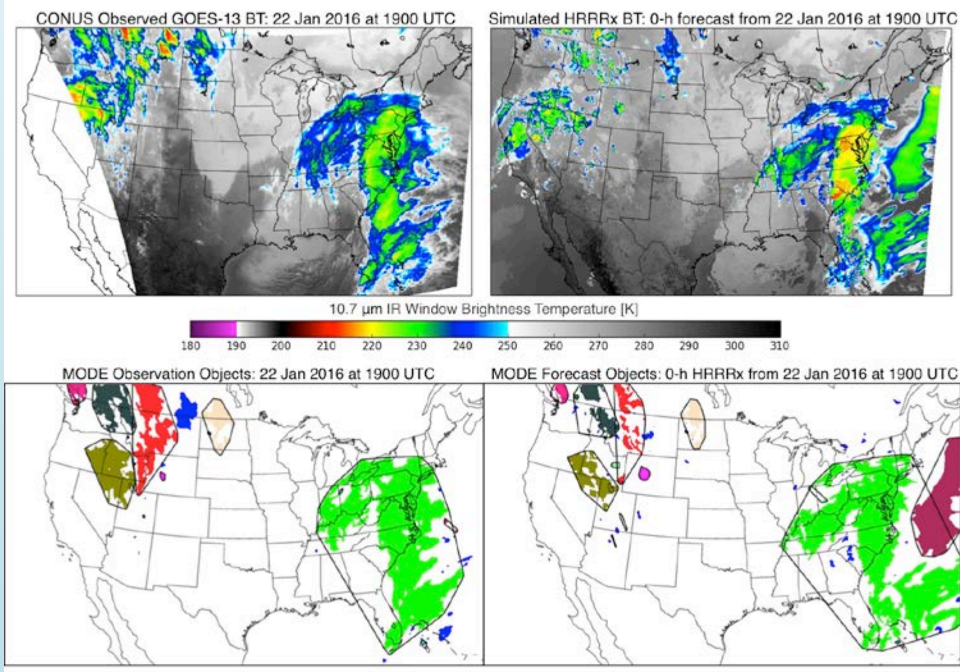
Object-Based Tracking

Program called method for object-based diagnostic evaluation (**MODE**)

MODE identifies and matches objects in two different fields.

1) Smooth and threshold the forecast and observed Brightness Temperature (BT) fields.
   - BTs colder than a certain threshold
   - Objects of a minimum size.
2) Calculate various object attributes for each observed and forecast cloud object.
3) Match forecast and observed cloud objects using a fuzzy-logic algorithm and calculate attributes of paired objects, such as intersection area and distance.
4) Output attributes for individual objects and matched object pairs for assessment.

MODE



Circled objects represent clusters. One or more objects can be in a cluster.
Useful if one or more object from one field matches up with a single object in the other.

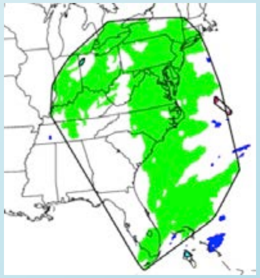
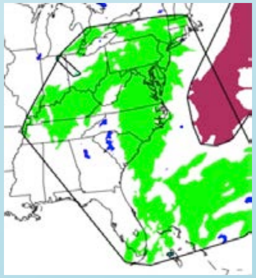
MODE: How well do objects match?

MODE measures the correspondence between two objects by computing the **interest value** between matched objects.

Observation Forecast

Interest Value = 1.00

Observation Forecast

Interest Value = 0.9287

Interest Value is a weighted average of attribute interest values.

Attribute	Weight	Description
Centroid Boundary Distance	4 (25%)	Distance between objects' centers of mass
Boundary Distance	3 (18.75%)	Minimum distance between the objects
Convex Hull Distance	1 (6.25%)	Minimum distance between the polygons surrounding the objects
Angle Difference	1 (6.25)	Orientation-angle difference
Area Ratio	4 (25%)	Ratio of the forecast and observation objects' areas (lowest value)
Intersection Area Ratio	3 (18.75%)	Ratio of the observation (forecast) object to the objects' intersection area (lowest value)

MODE Composite Score (MCS)

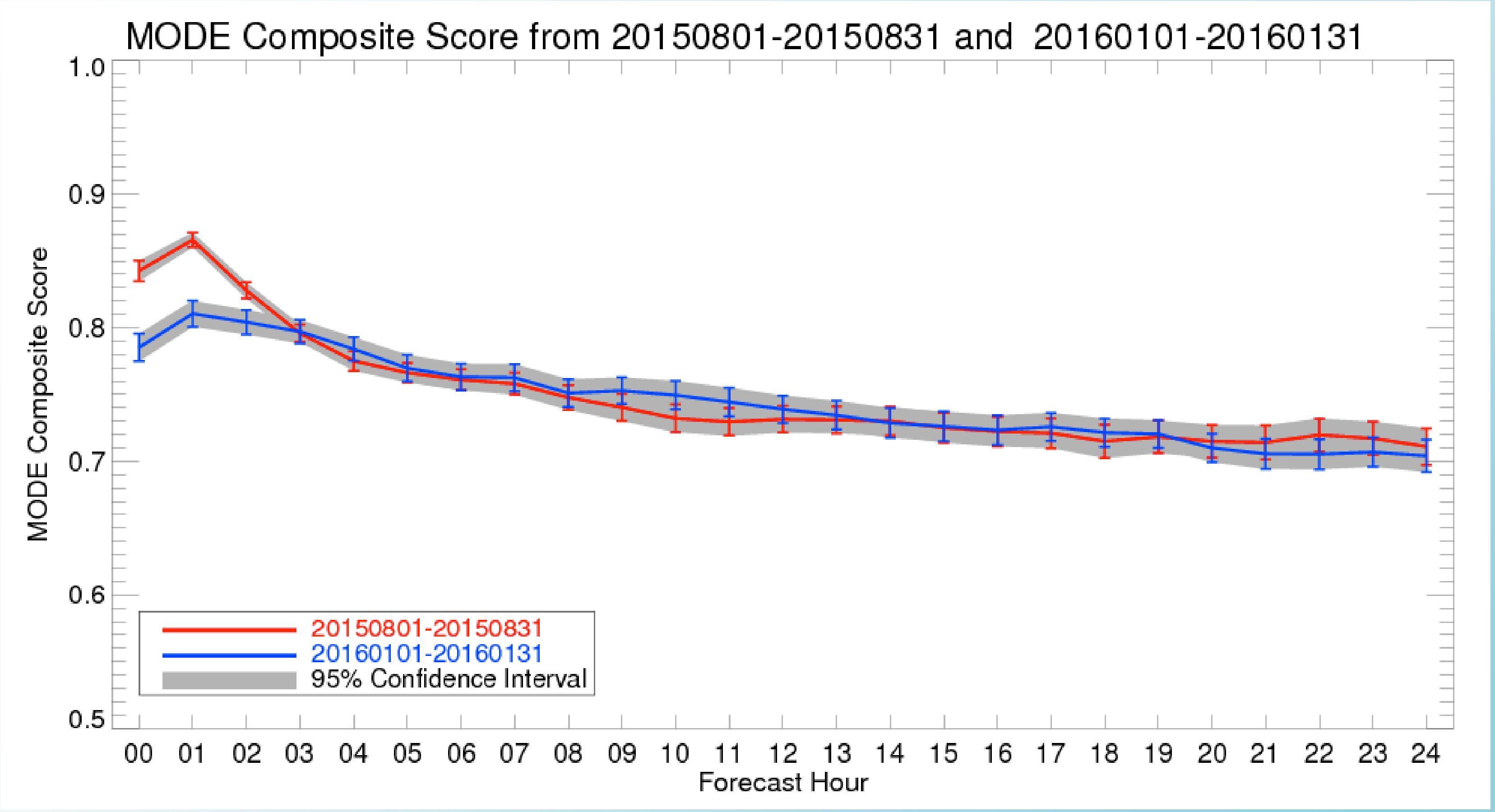
Combines wealth of information about objects into a single value

$$MCS = \sum_{i=1}^{C} \frac{Area_{\text{Observed Cluster}}(i)}{\text{Total Area}} * \text{Interest Value } (i) + \sum_{j=1}^{O} \frac{Area_{\text{Observed Object}}(j)}{\text{Total Area}} * \text{Interest Value } (j)$$

- Area-weighted metric that uses the MODE interest values for each object
- Area weighting used so that more weight is given to larger cloud objects
- Total Area = observation area + unmatched forecast area
- Area of each object or cluster multiplied by its interest value and then summed over all identified objects and clustered
- MODE Composite Score ranges from 0 to 1.
  - Zero indicates no skill
  - One indicates perfect skill

Results

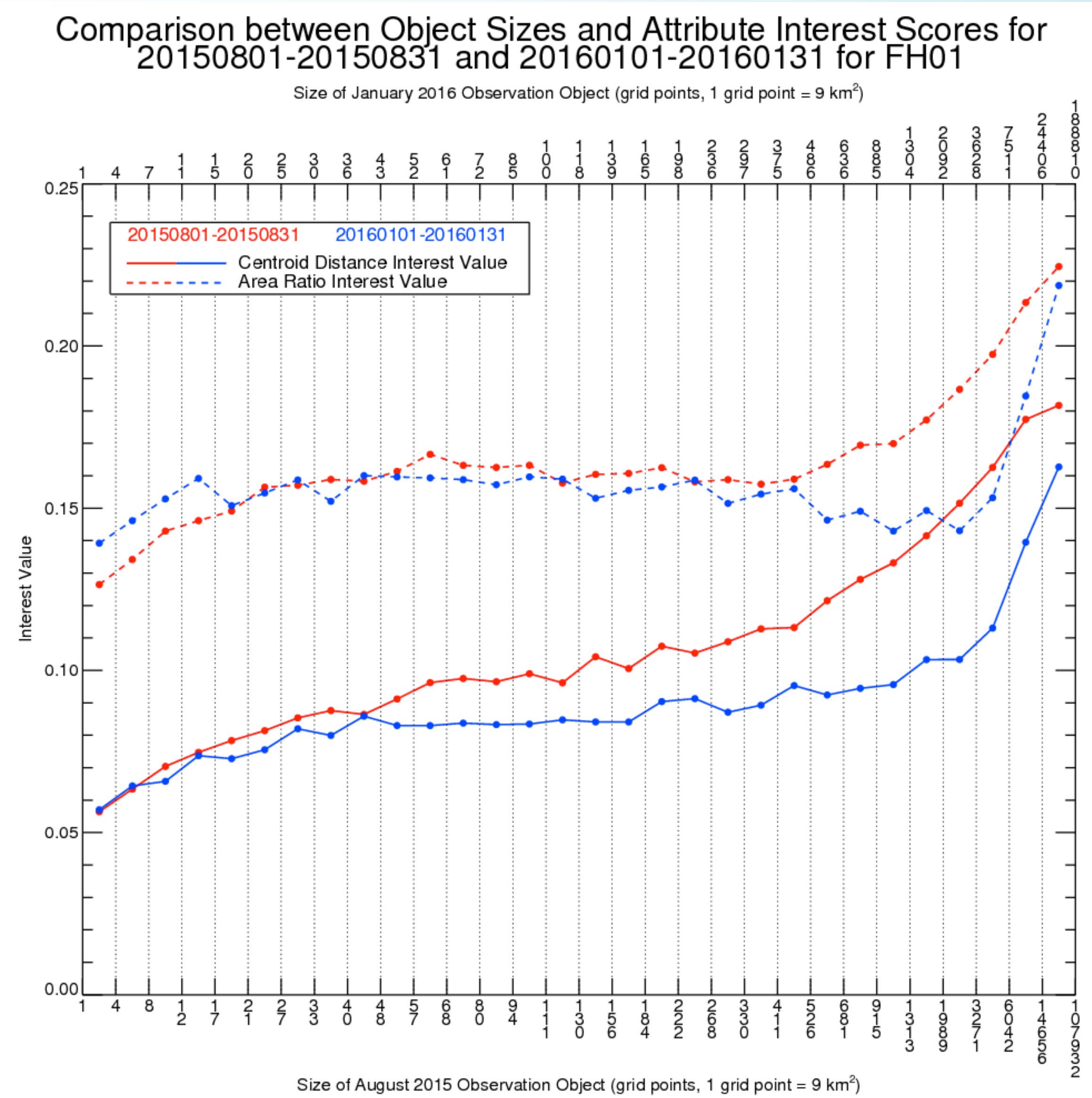
Compared experimental HRRR simulated BT to GOES observed BT for August 1-30, 2015 and January 1-31, 2016.



Why are the forecasts from August more accurate than the January forecasts for early hours?

Why is the 1-h forecast more accurate than the 0-1 forecast?

August more skillful than January



Centroid distance (solid line):

Higher for August than January for most objects.

The centers of objects are closer.

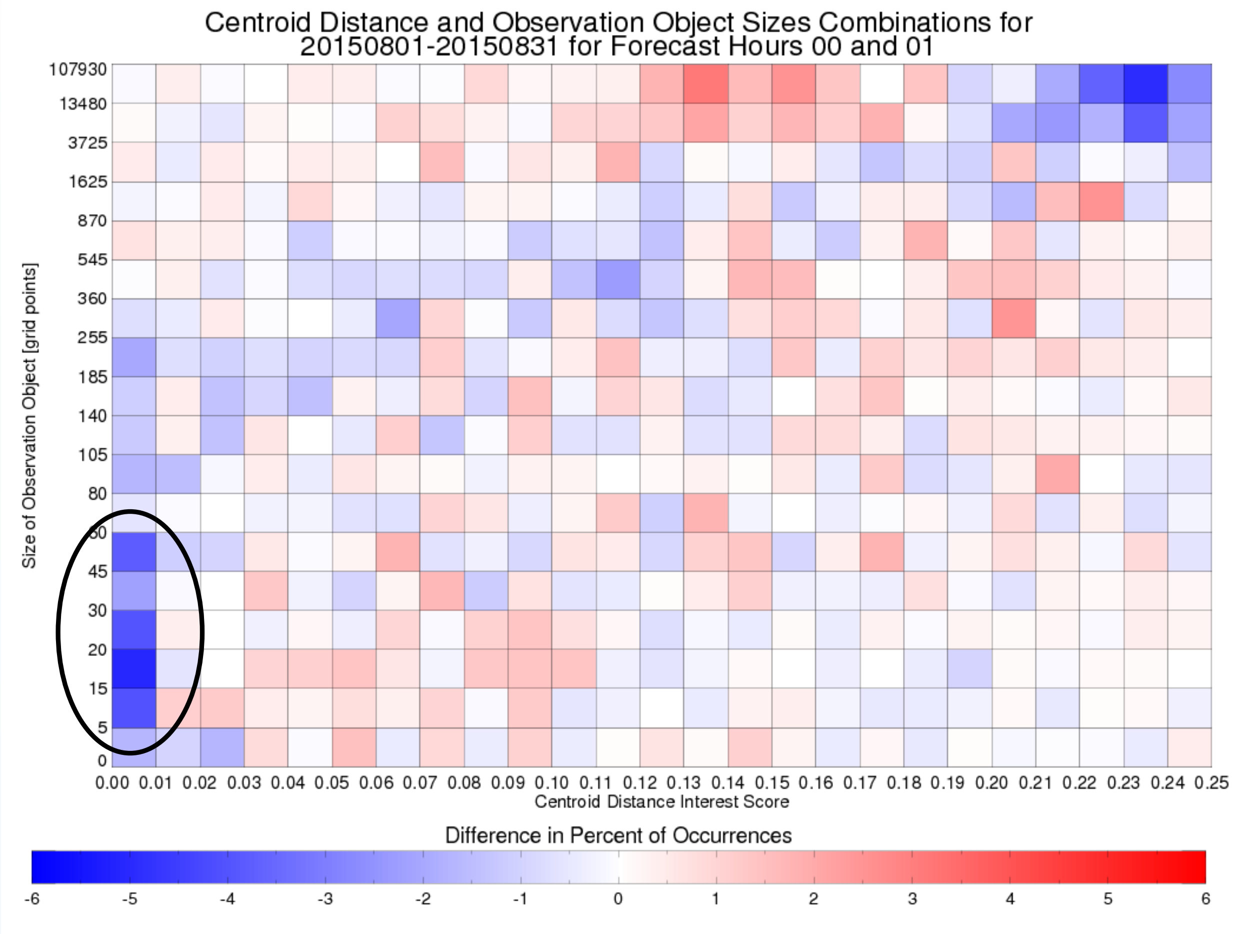
Area ratio (dashed line):

Similar for August than January except for larger objects.

The centers of larger objects are closer in August.

1-h forecast more skillful than 0-h analysis

$$\text{Difference in Percent of Occurrences} = \frac{\text{Number(score, size)}_{FH\,00}}{\text{Total(size)}_{FH\,00}} - \frac{\text{Number(score, size)}_{FH\,01}}{\text{Total(size)}_{FH\,01}}$$



Greater Occurrence For Forecast Hour 1

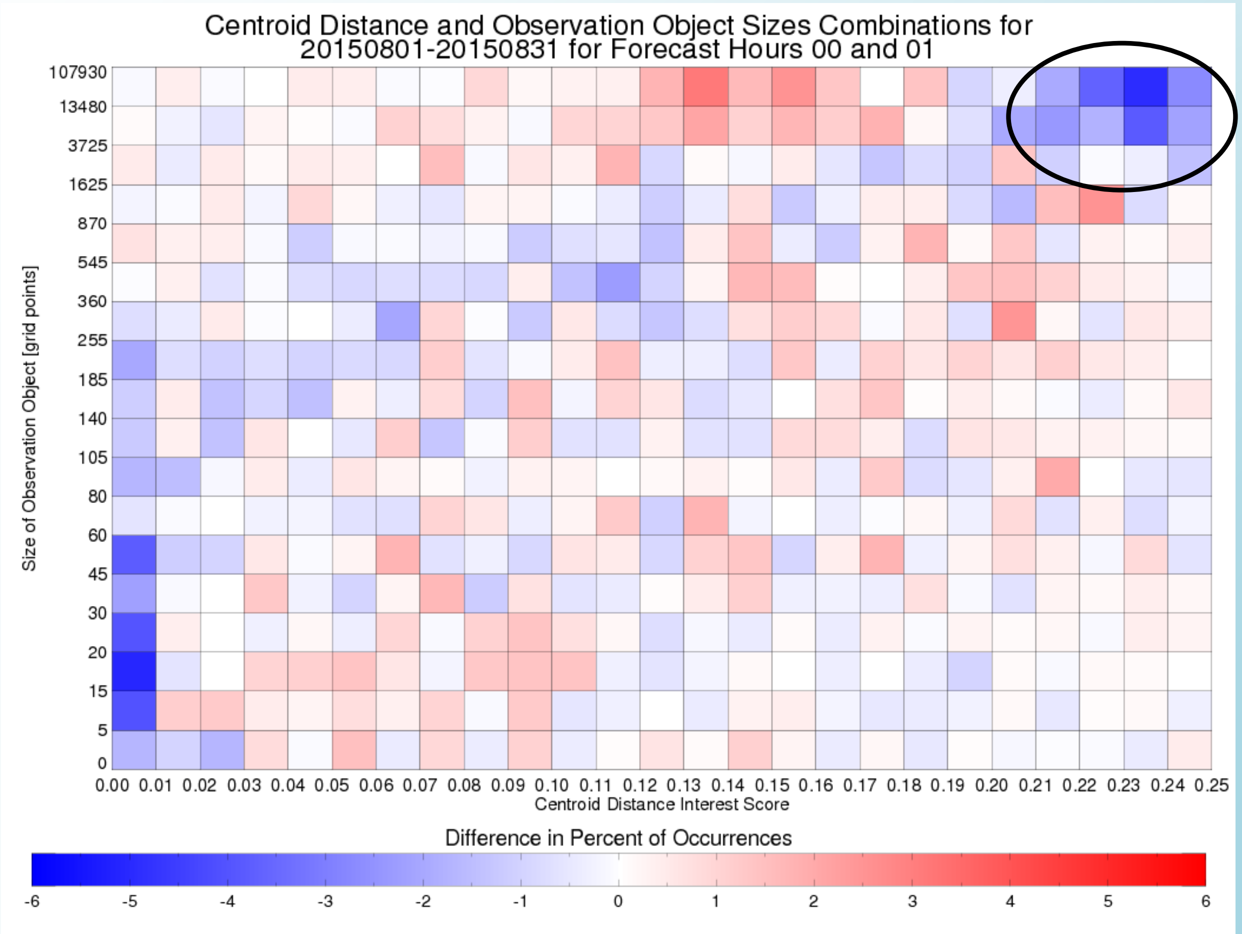
Greater Occurrence For Forecast Hour 0

Centroid distance: distance between the center points of objects

- For small objects, more forecast hour 1 objects have low interest scores

1-h forecast more skillful than 0-h analysis

$$\text{Difference in Percent of Occurrences} = \frac{\text{Number(score, size)}_{FH\ 00}}{\text{Total(size)}_{FH\ 00}} - \frac{\text{Number(score, size)}_{FH\ 01}}{\text{Total(size)}_{FH\ 01}}$$



Greater Occurrence For Forecast Hour 1

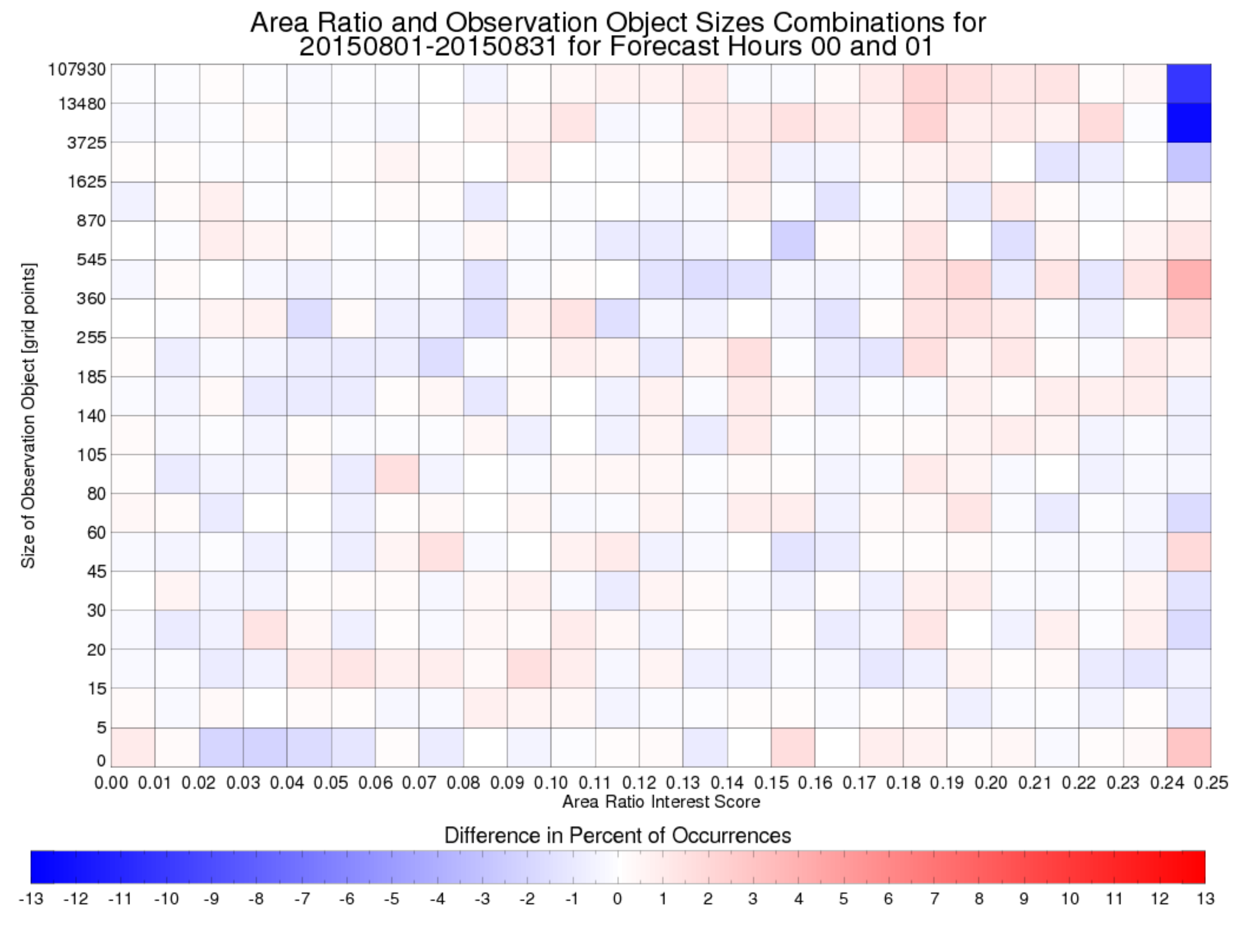
Greater Occurrence For Forecast Hour 0

Centroid distance: distance between the center points of objects

- For small objects, more forecast hour 1 objects have low interest scores
- For larger objects, more forecast hour 1 objects have larger interest scores
- Improvement in larger objects has a greater impact on the MODE Composite Score

1-h forecast more skillful than 0-h analysis

$$\text{Difference in Percent of Occurrences} = \frac{\text{Number(score, size)}_{FH\,00}}{\text{Total(size)}_{FH\,00}} - \frac{\text{Number(score, size)}_{FH\,01}}{\text{Total(size)}_{FH\,01}}$$



Greater Occurrence For Forecast Hour 1

Greater Occurrence For Forecast Hour 0

Area ratio: ratio of the number of observed to forecast grid points

- Small differences for most interest scores and object sizes
- For larger objects, higher percentage of forecast hour 1 object matches have larger interest scores
- Improvement in large cloud objects has a greater impact on the MODE Composite Score

Conclusions

We described object-based statistics package called MODE which was used to assess the accuracy of experimental HRRR cloud cover (using brightness temperatures as a proxy).

MODE is highly configurable, allowing users to decide the threshold and minimum size of objects as well as what is most important when defining object matches.

We've developed a metric (MODE Composite Score) to distill all of the information from MODE into a single value.

Forecast accuracy can all be assessed using object or attribute interest scores, which can better explain **WHY** one forecast is more accurate than the other.