QPF Verification Comparison between the GFS and NAM Operational Models

Jamie Wolff, Barbara Brown, John Halley Gotway, Michelle Harrold, Louisa Nance, Paul Oldenburg and Zach Trabold

National Center for Atmospheric Research (NCAR)

Collaborator: Ying Lin of NCEP/EMC

12th Annual WRF Users Workshop

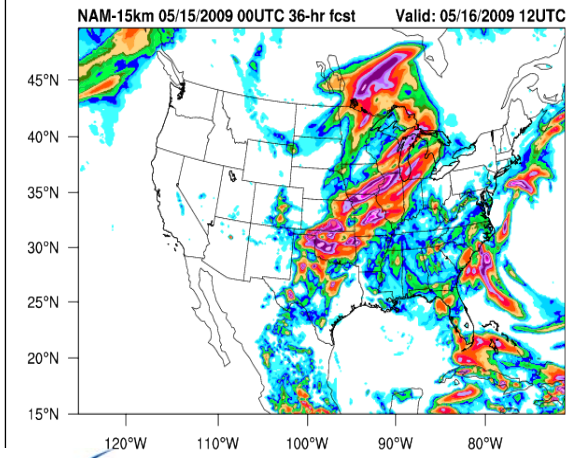
June 23, 2011

Overview

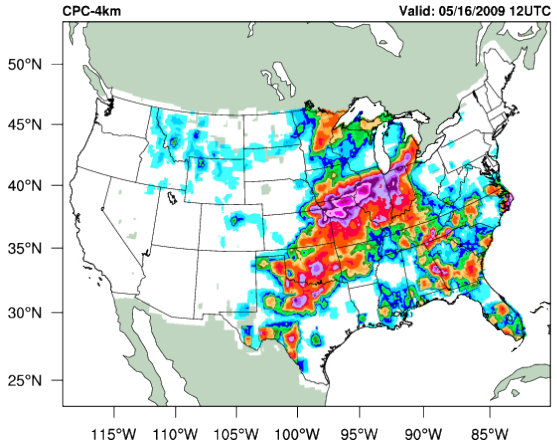
- Goal: Assess the performance of the Global Forecast System (GFS) and North American Mesoscale (NAM) operational models, which differ significantly in horizontal resolution
  - Secondary goal: Demonstrate the utility of, and the attributes available from, new spatial verification techniques

24-h Accumulated Total Precipitation (in)

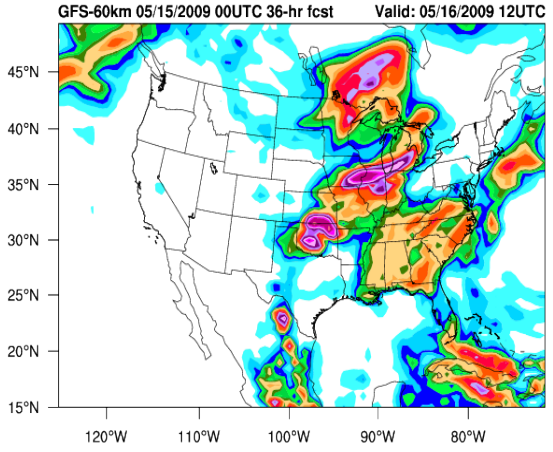
NAM (higher resolution)

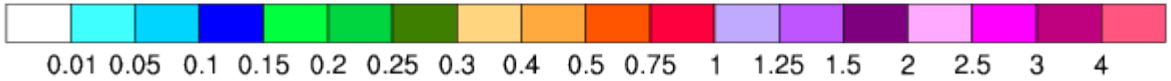


Observations



GFS (coarser resolution)



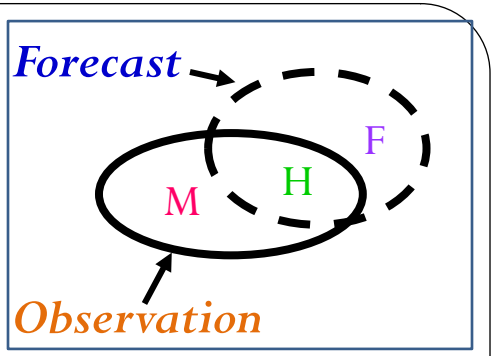


Experiment Design

- Native forecast datasets
  - GFS: global Gaussian grid (half-degree resolution)
  - NAM: E-grid domain (~12-km resolution)
- Native observation datasets
  - NCEP Stage II analyses: 3-h observed precip accum (4-km resolution)
  - NCEP/CPC analyses: 24-h observed precip accum (1/8-degree resolution)
- Test Period: 18 Dec 2008 – 15 Dec 2009
- Retrospective forecasts: 00 UTC daily initializations out to 84 h (with output available every 3 h)
- Common grid: 4-km, 15-km, 60-km CONUS
- Verification: Model Evaluation Tools v3.0

Verification domain

Model Verification



- Traditional Verification Metrics
  - **Gilbert Skill Score (GSS)**: Fraction of obs and/or fcst events that were correctly predicted

$$\frac{\#\text{Hits} - \#\text{Hits}_{\text{rand}}}{\#\text{Hits} + \#\text{Misses} + \#\text{False Alarm} - \#\text{Hits}_{\text{rand}}}$$

$$\#\text{Hits}_{\text{rand}} = \frac{(\text{Total Fcst Area})(\text{Total Obs Area})}{\text{Total Area}}$$

Range: -0.33 to 1. Perfect: 1

  - **Frequency Bias**: Ratio of the frequency of forecast events to the frequency of observed events

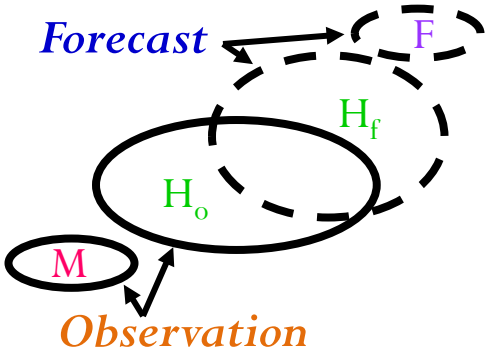
$$\frac{\#\text{Hits} + \#\text{False Alarm}}{\#\text{Hits} + \#\text{Misses}} \quad \text{-or-} \quad \frac{\text{Total Fcst Area}}{\text{Total Obs Area}}$$

Range: 0 to ∞. Perfect: 1 (Under-forecast<1, Over-forecast>1)

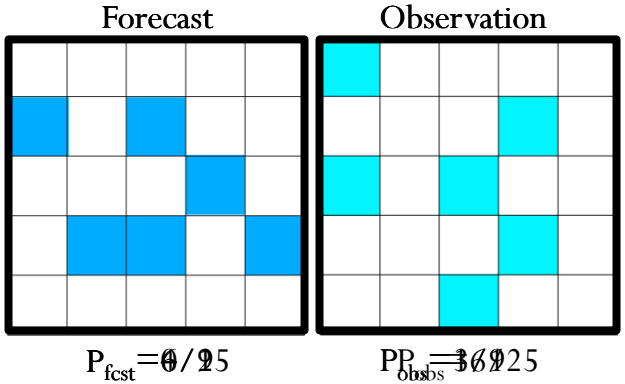
- Computed confidence intervals (CIs) at the 99% level, using a bootstrapping technique
- Identified statistically significant (SS) differences between scores

Model Verification, Cont.

- Spatial Verification Techniques
  - **Method for Object-based Diagnostic Evaluation (MODE)**: *Identify*, *merge* and *match* objects in forecast and observed fields
    - Example attributes:
      - centroid distance, boundary distance, angle difference, area ratio, percent coverage, intersection area ratio, etc.



  - **Fractional Skill Score (FSS)**: Obtain a measure of how forecast skill varies with spatial scale

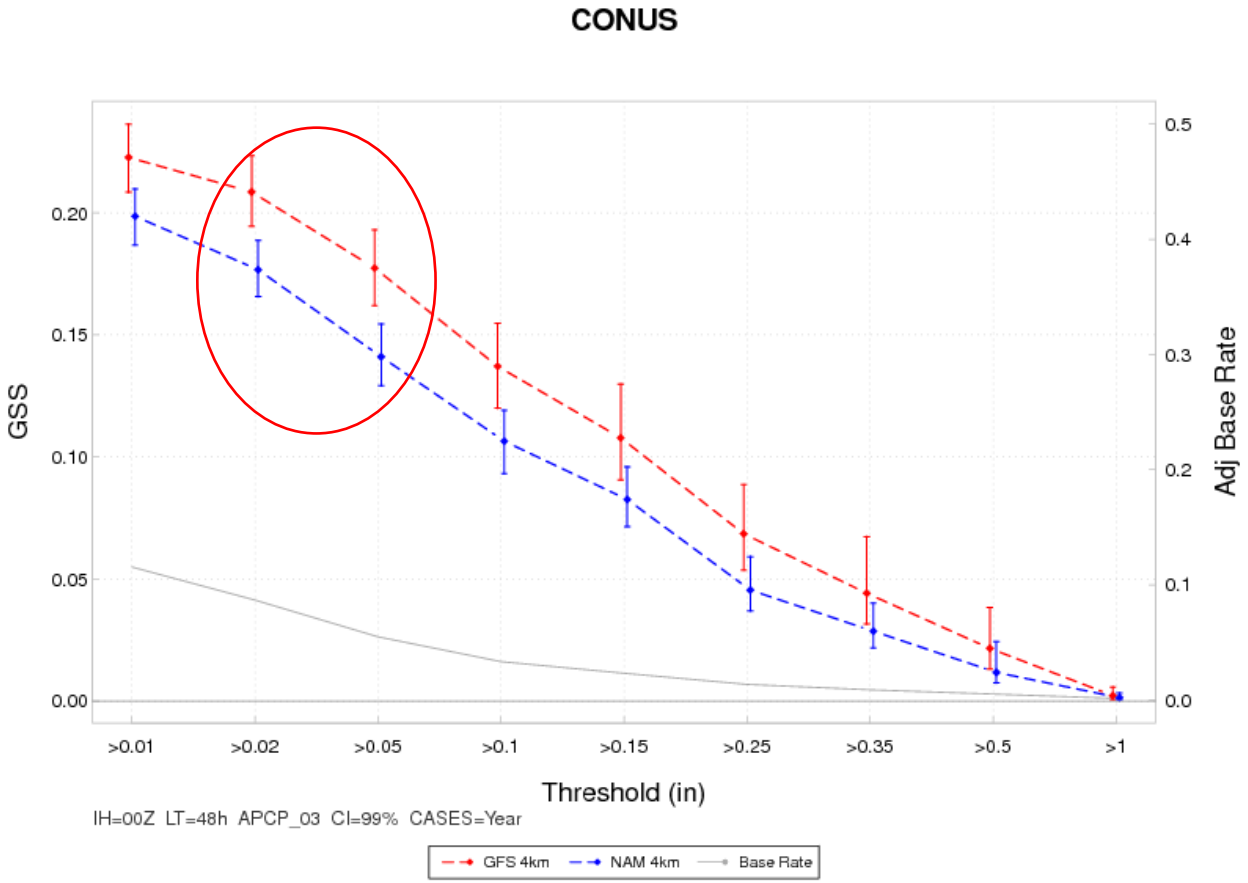


Courtesy: Mittermaier

Verification Results

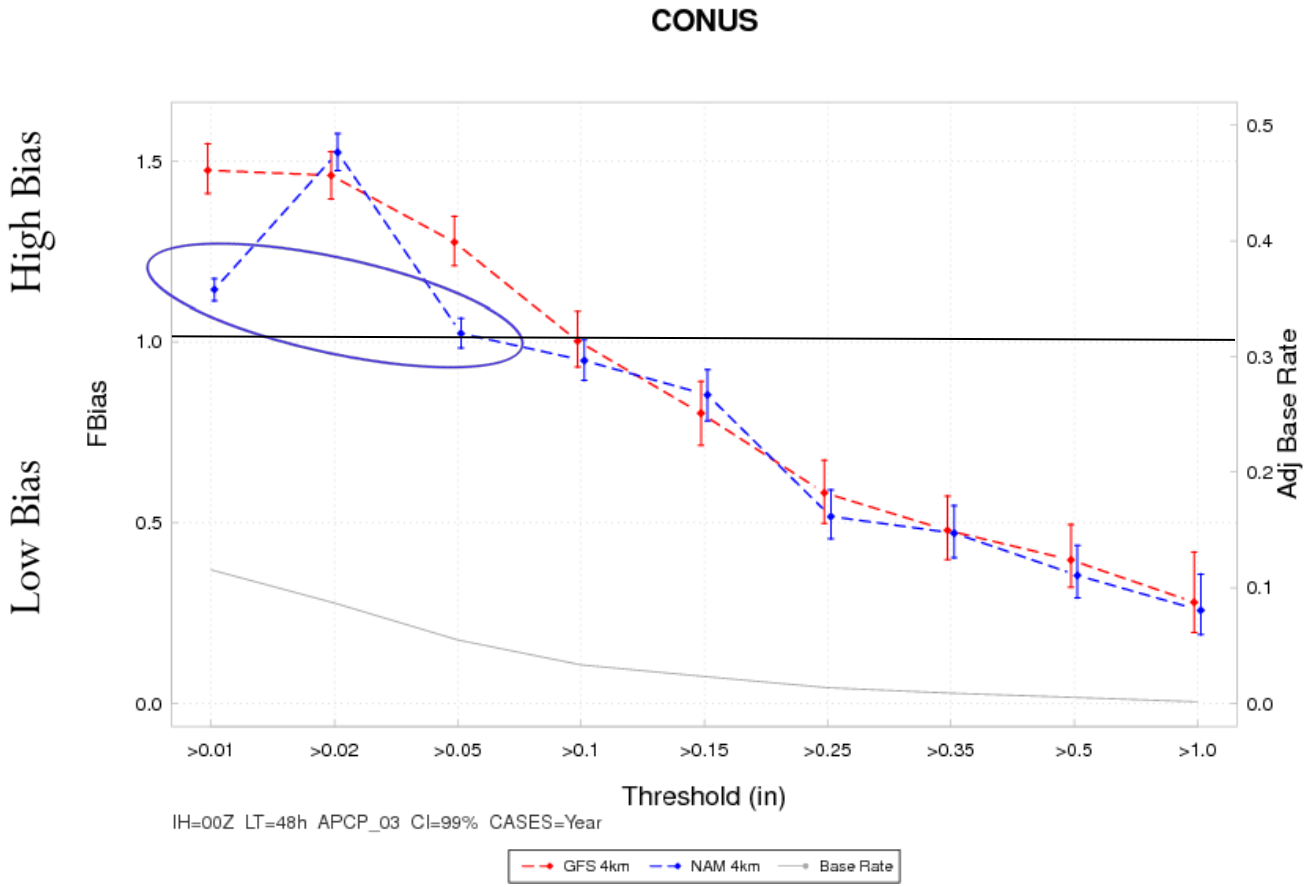
Traditional Verification:
Gilbert Skill Score (GSS)

Higher skill



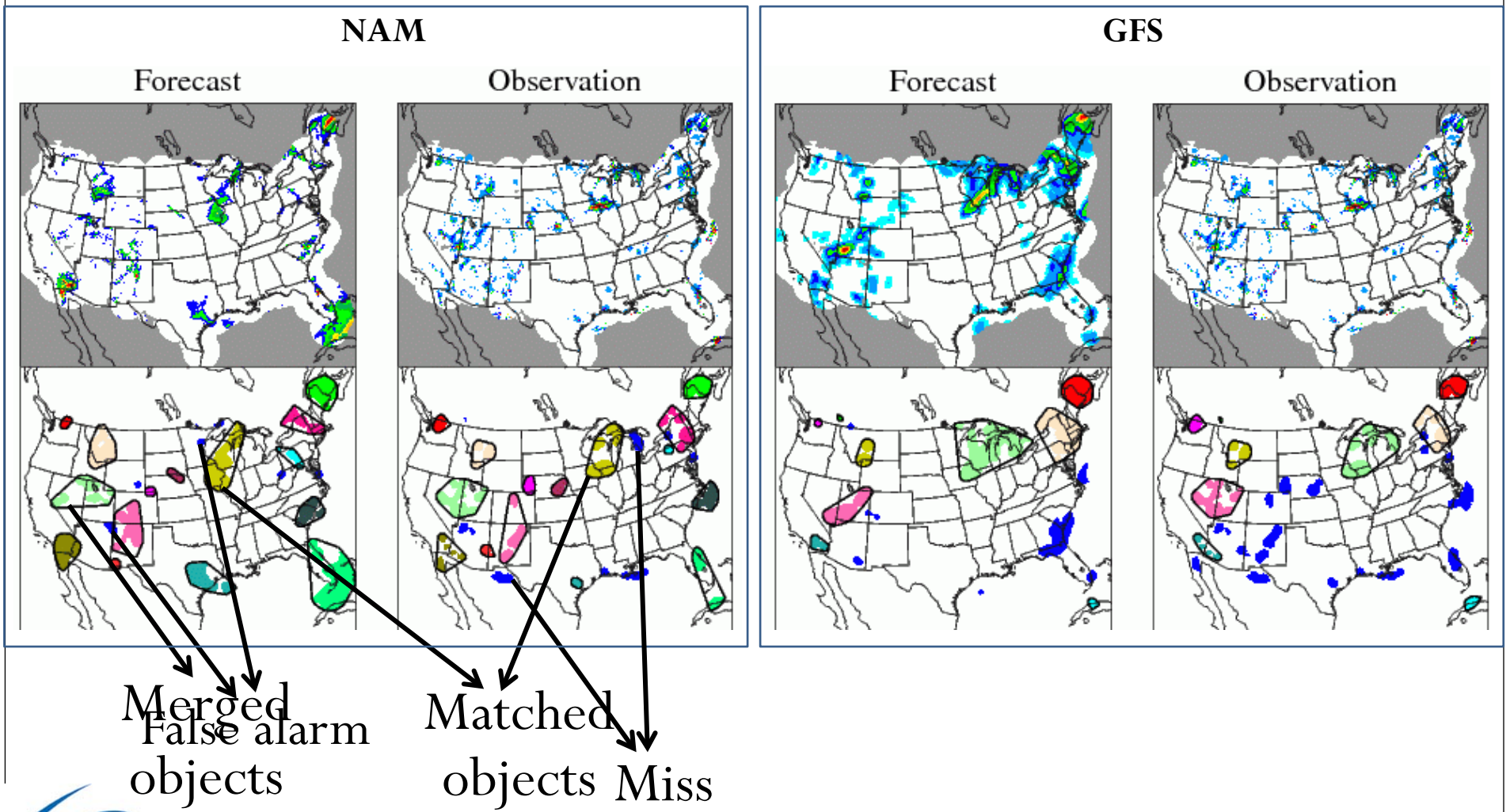
- With two exceptions, the scores are not statistically different when measuring performance based on this traditional metric

Traditional Verification:
Frequency Bias

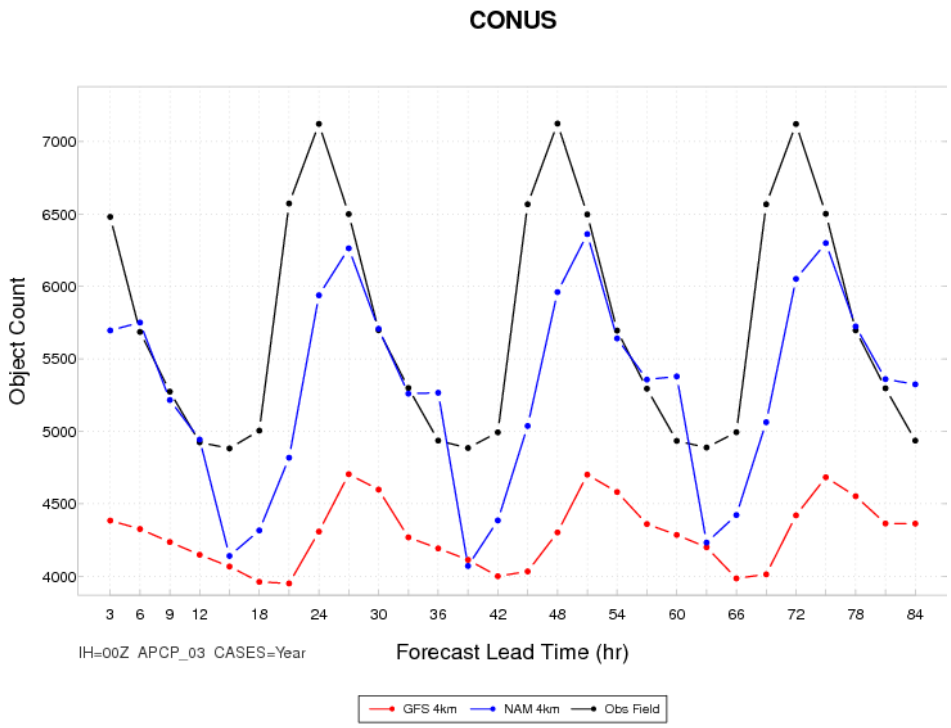


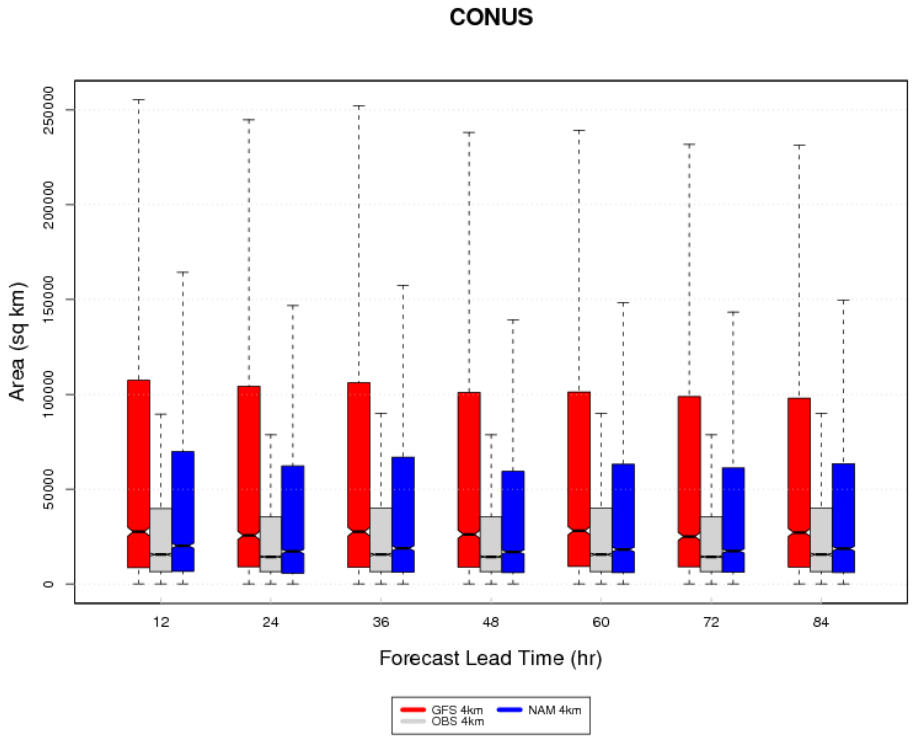
- Again, with two exceptions, the scores are not statistically different when measuring performance based on this traditional metric

Spatial Verification:
Method for Objection-based Diagnostic Evaluation (MODE)



MODE:
Object Counts and Areas

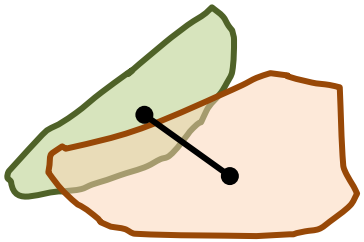


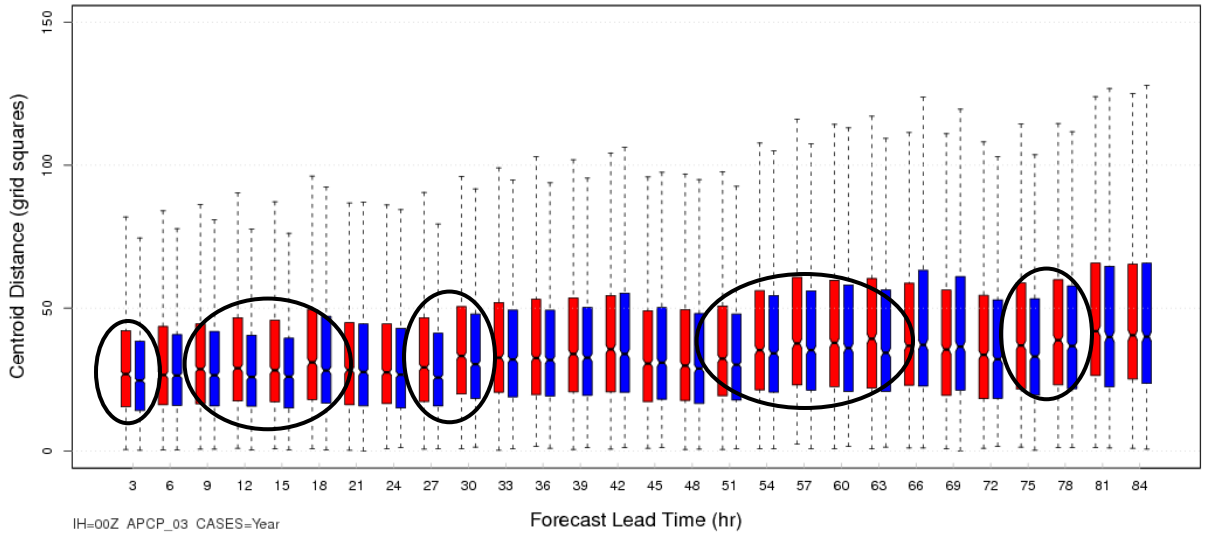


- Counts and size distribution for objects defined within the **NAM4** forecast are more consistent with the obs field than the **GFS4** forecast

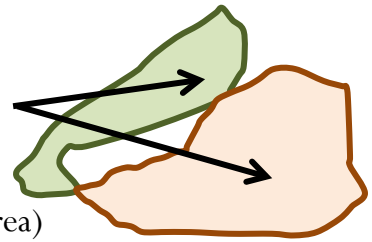
MODE Attributes (matched objects):

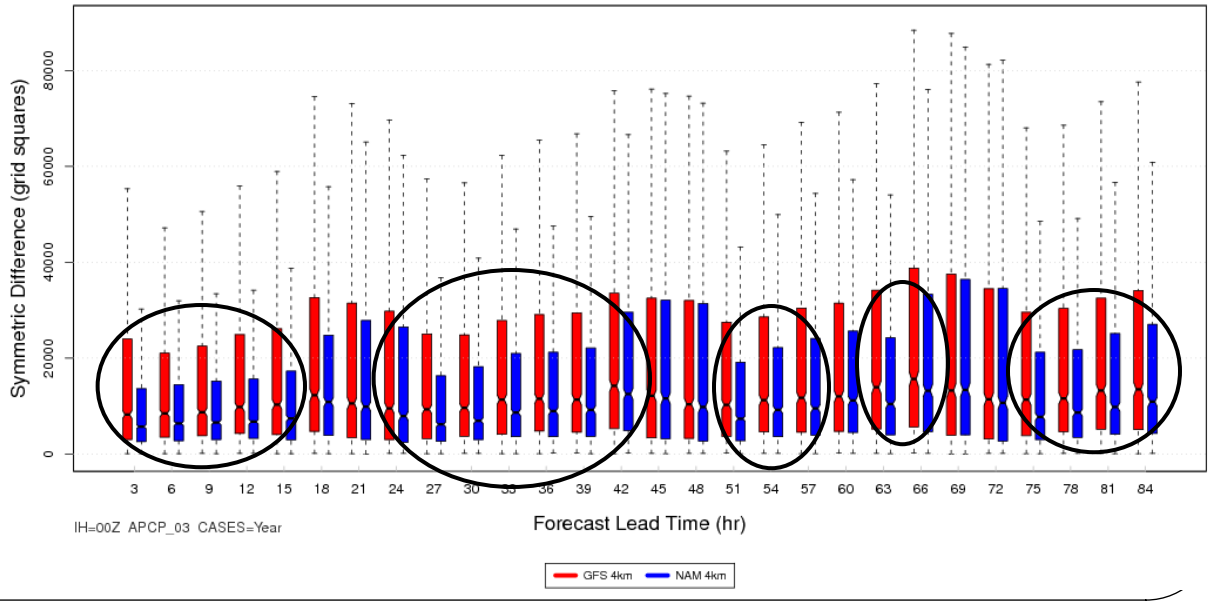
Example:
Centroid
Distance



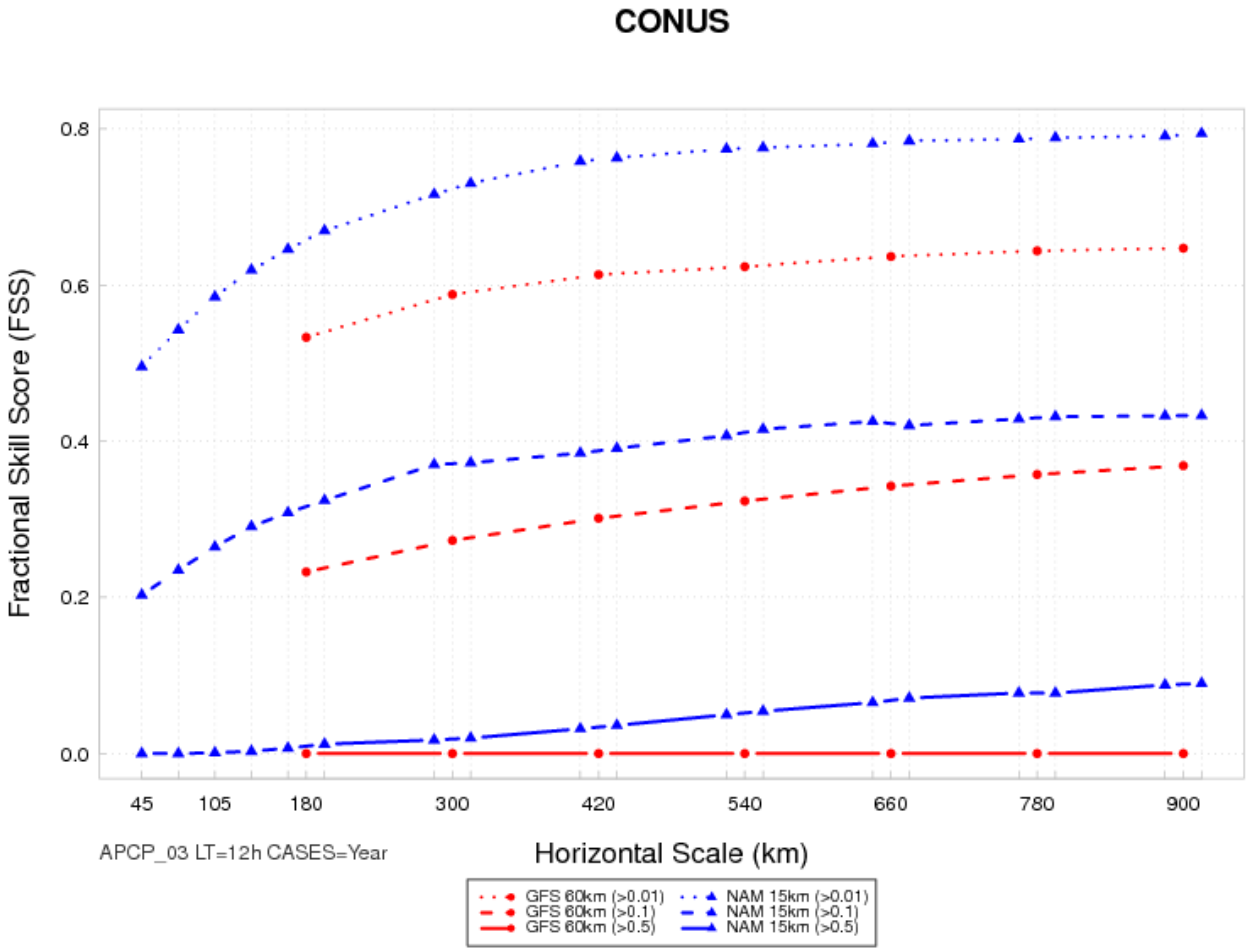


Example:
Symmetric
Difference
(non-intersecting area)



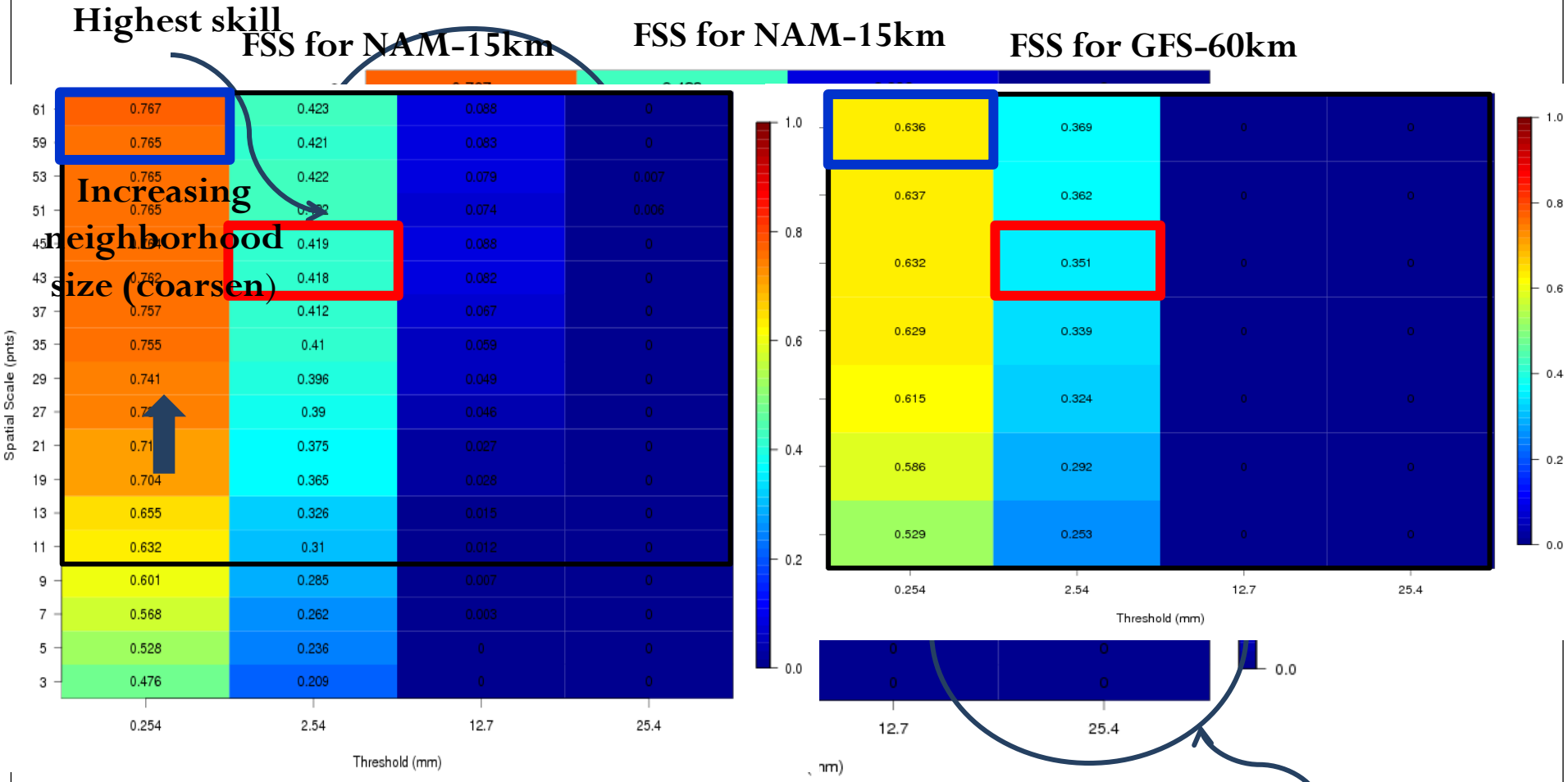


Spatial Verification:
Fractional Skill Score



- **NAM15** is consistently higher than the **GFS60** across all thresholds (12h lead time shown)

FSS: Quilt Plot



- FSS for **NAM15** forecast is consistently higher than the **GFS60** forecast regardless of spatial scale or threshold

Summary

- Even though, subjectively, the higher-resolution models can provide added benefit, traditional verification metrics show no notable, consistent improvement in scores
- Advanced spatial verification techniques can provide useful information on forecast skill for high-resolution models
  - MODE
    - NAM objects (counts and area) more closely reproduce those of the observation field
    - Example attributes of matched objects favor the NAM
  - FSS
    - NAM has consistently higher skill than the GFS at comparable spatial scales
- For more information, see:

http://verif.rap.ucar.edu/eval/gfs_nam_pcp/

Acknowledgements: *The DTC is funded by the National Oceanic and Atmospheric Administration, the Air Force Weather Agency, and the National Center for Atmospheric Research (NCAR). NCAR is sponsored by the National Science Foundation.*