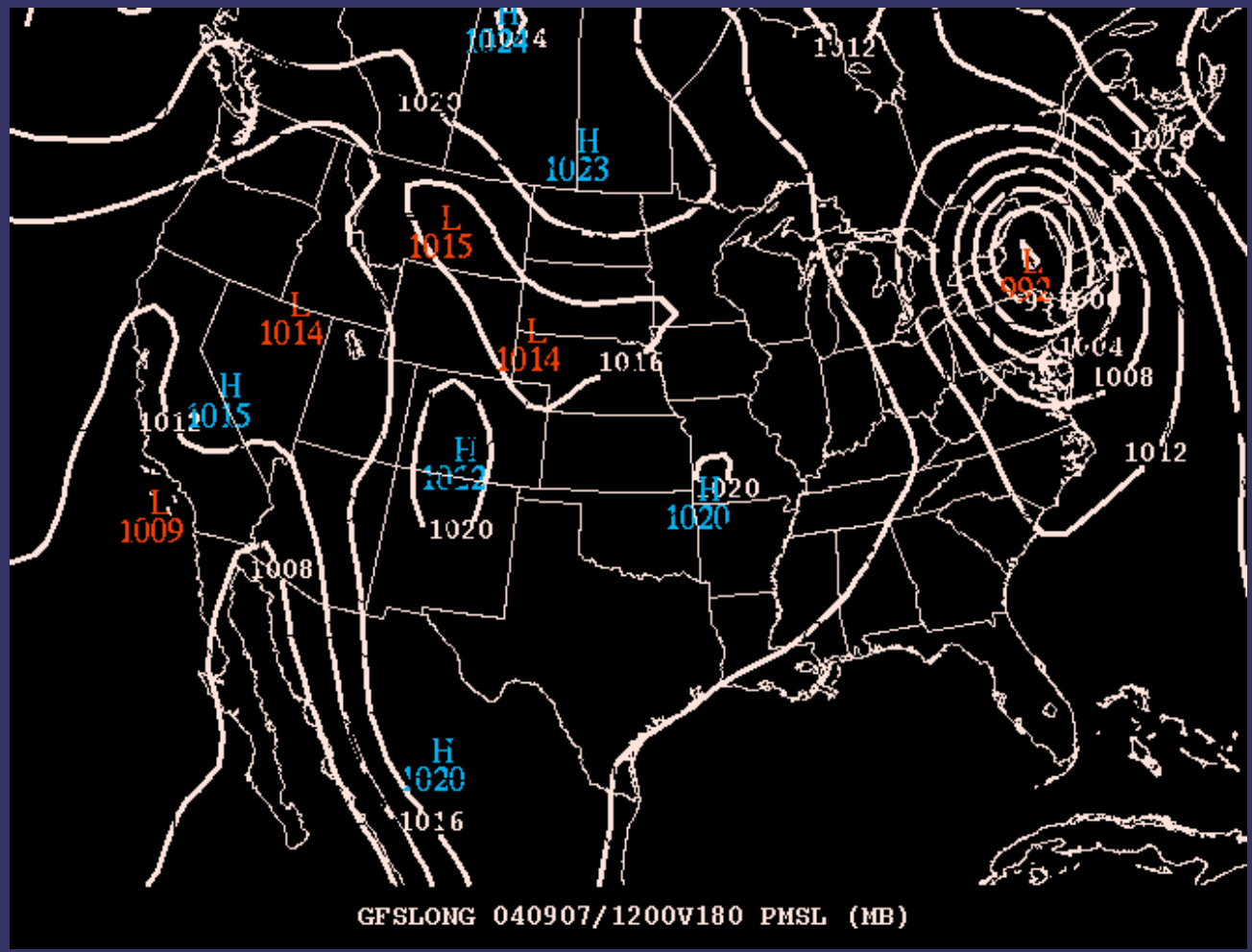
The design and evaluation of a measure of forecast consistency for the Collaborative Convective Forecast Product

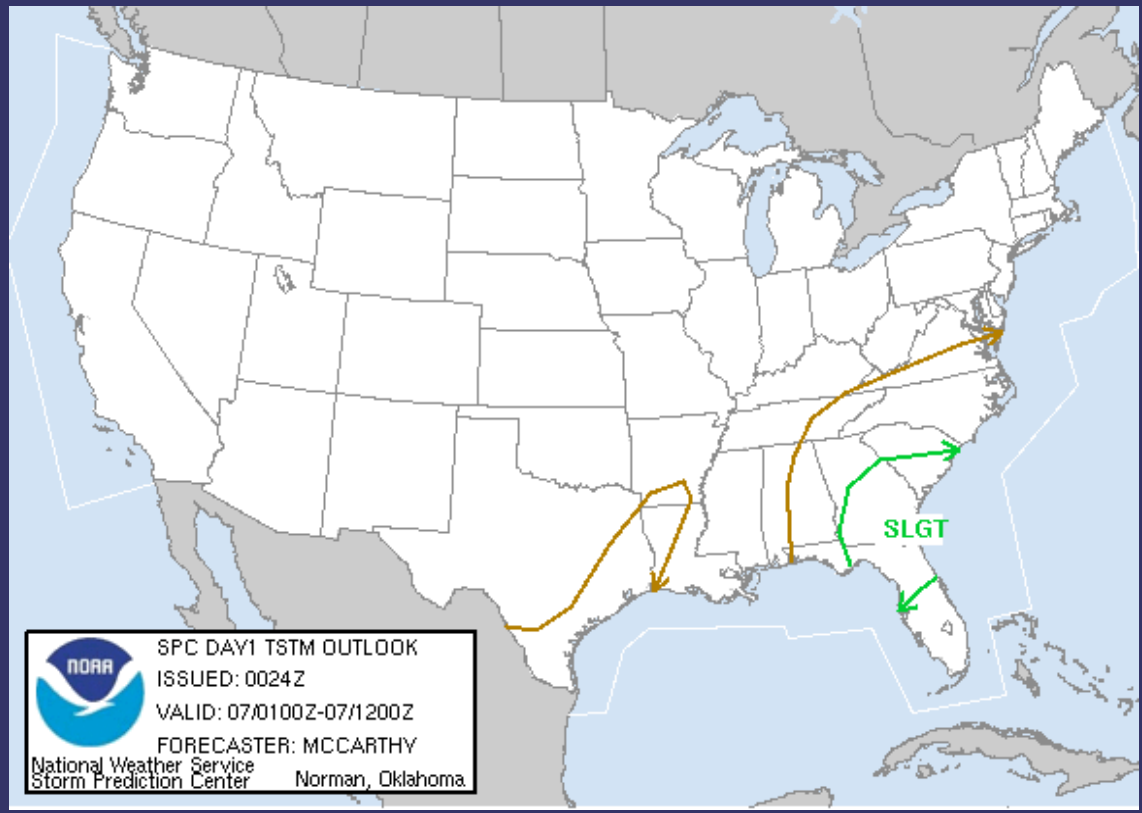
Mike Kay
mike.kay@noaa.gov
CIRES/NOAA/FSL

How does one make decisions based on these forecasts?!

Imagine if you were a utility company in North Carolina and needed to decide if you should bring in extra crews to help with storm damage repairs. How would you deal with this set of forecasts?



Does it appear to be easier to make decisions based upon this set of forecasts?



Forecast Consistency

Refers to the similarity (or lack thereof) between a series of forecasts with different lead times but identical valid times

Often referred to by operational meteorologists as d(prog)/dt referring to the change in a series of numerical model forecasts as lead time decreases

For example, if a model has been too cold recently then forecasters may attempt to adjust for this in the most recent forecast, by assuming these trends are continuing

Similarly, if a model trends towards moving a cold front through an area more quickly forecasters are likely to assume that the trend of increased speed for frontal passage is correct

Forecast Consistency (cont.)

Forecast consistency is potentially a perilous rule of thumb to be using owing to human ignorance (Gilovich 1993)

Hamill (2003) looked at d(prog)/dt for a long series of NWP forecasts from a frozen model and found it to be a poor forecast tool:

Extrapolation of forecast trends showed little utility

The amount of disagreement between the forecasts did not serve as an indicator of accuracy

What makes this work different than Hamill (2003)?

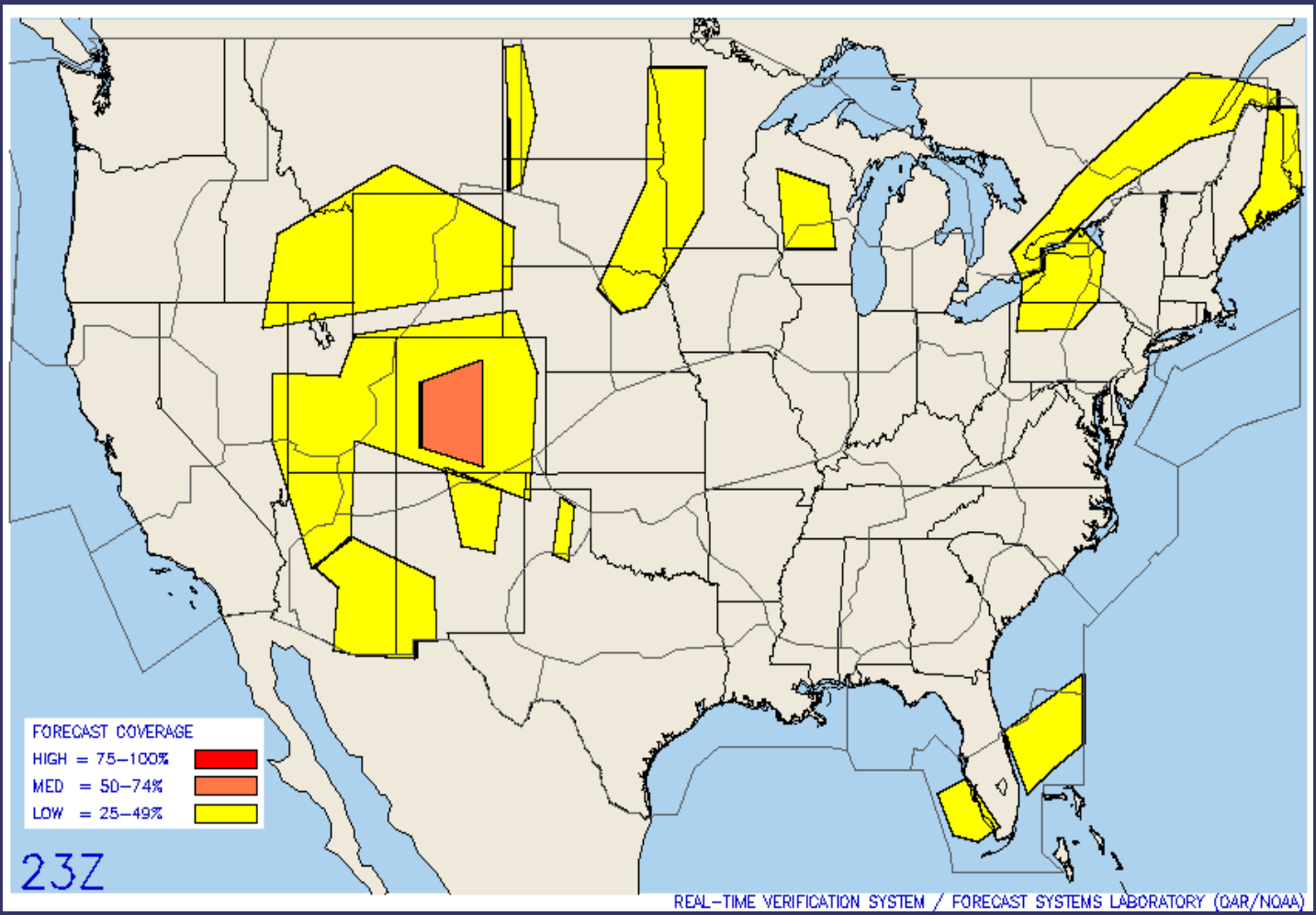
A quantitative look at consistency of human-generated forecasts has never been done

The forecasts we're looking at have some element of consistency built into the forecast process. If our measure proves valuable, and with a different result than Hamill, it may be suitable for inclusion in the forecast process for reducing data overload issues

NHC discussion on Ivan:

EXCEPT THE GFDL WHICH HAS SHIFTED A LITTLE TO THE RIGHT. THE OFFICIAL FORECAST TRACK IS SHIFTED ABOUT 60 N MI LEFT OF THE PREVIOUS ADVISORY AT 72 HOURS AND IS STILL TO THE RIGHT OF ALL GUIDANCE EXCEPT FOR THE GFS. *IF I DID NOT HAVE A PREVIOUS FORECAST TO MAINTAIN SOME CONTINUITY WITH...I WOULD HAVE SHIFTED THE TRACK EVEN FURTHER TO THE LEFT.*

Forecast Information



Collaborative Convective Forecast Product

Used for strategic planning of air traffic routes in the U.S.

Issued numerous times each day with 2-, 4-, and 6-hour leadtimes at each issuance

Used 1171 forecasts from 1 Mar 2003 to 25 Oct 2003

Definition of consistency

We want a measure of that captures what a forecaster would view as consistency:

1. Forecast areas have similar locations and orientations

2. Attributes of those areas (e.g., forecast confidence) similar

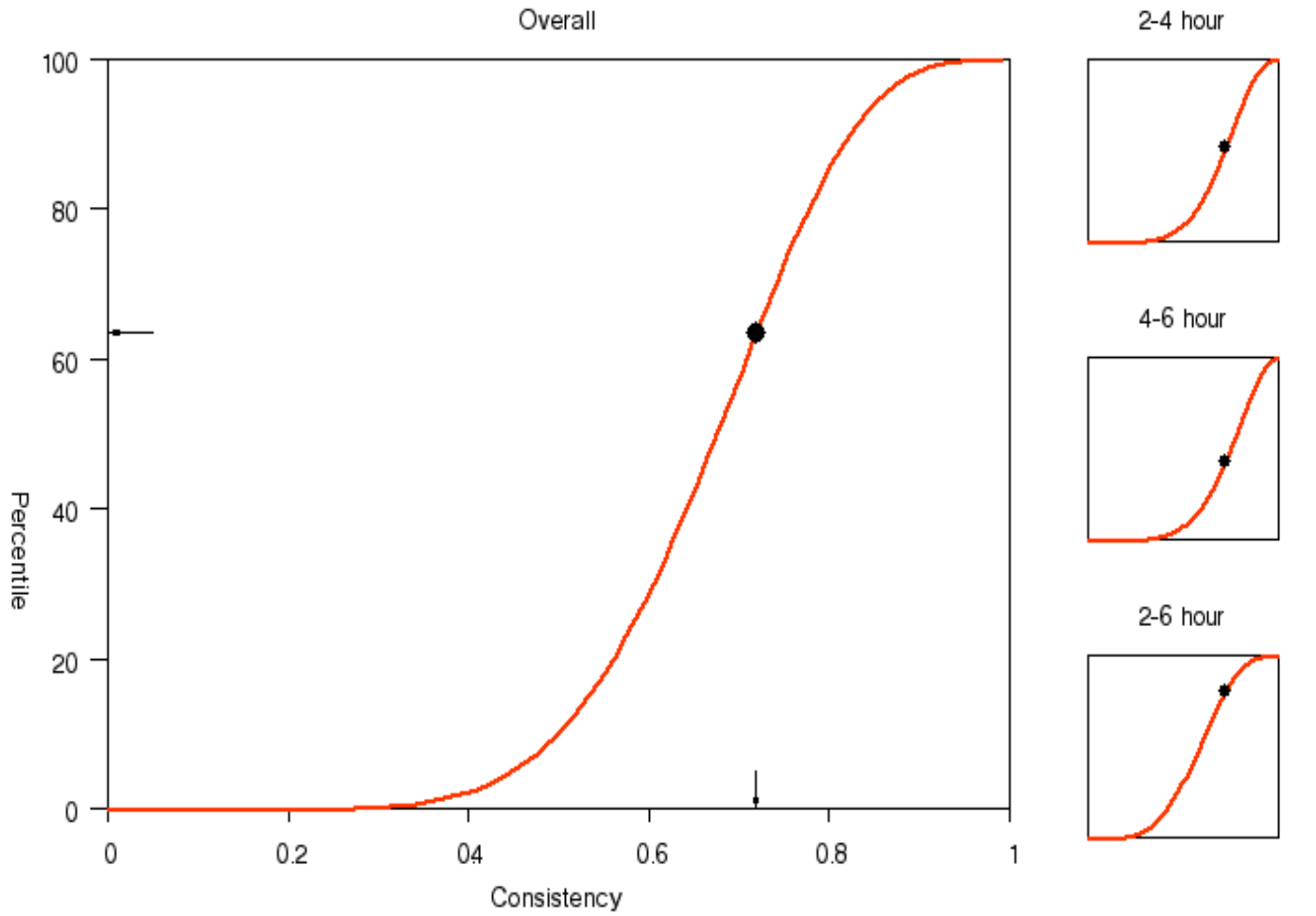
$$C = \frac{cor_{2\_4} + cor_{2\_6} + cor_{4\_6}}{3}$$

Two modifications:

We allow forecast areas to move around (to maximize correlation) up to 200km

Score is rescaled to [0,1]

How do you present the score to users?



Use the cumulative distribution function

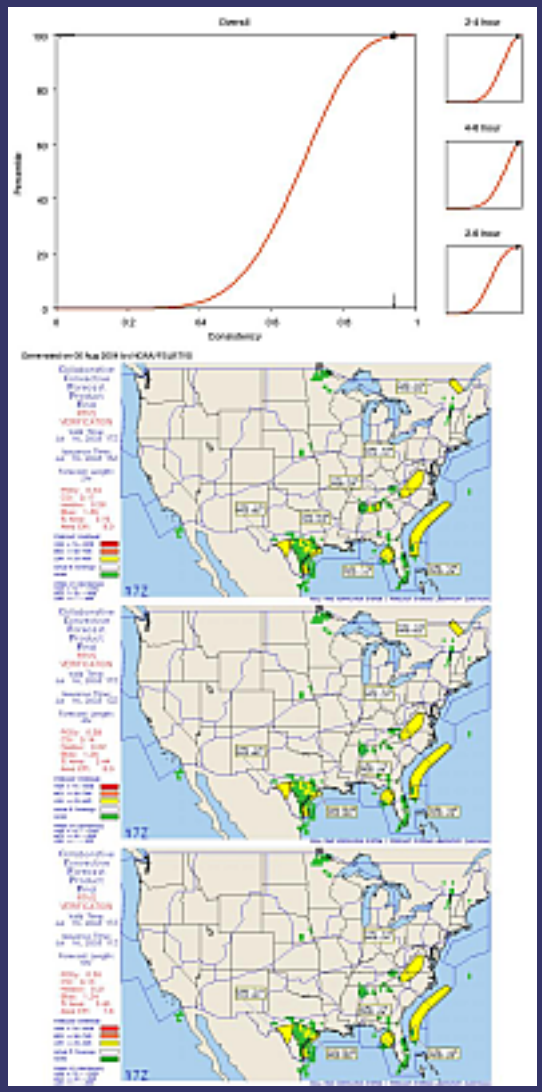
Presents the score relative to all other scores

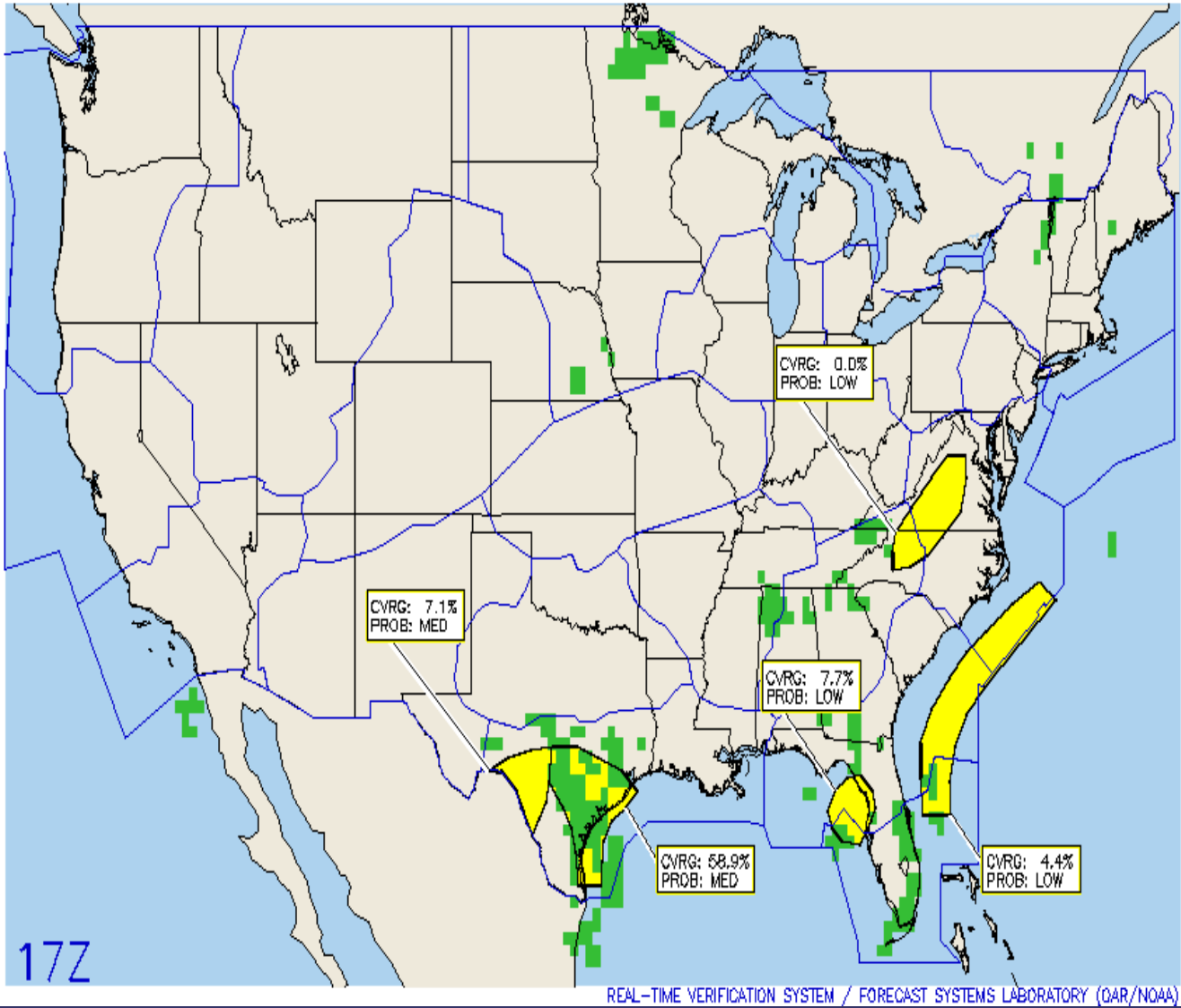
Easy to interpret the diagram

Allows for rapid understanding

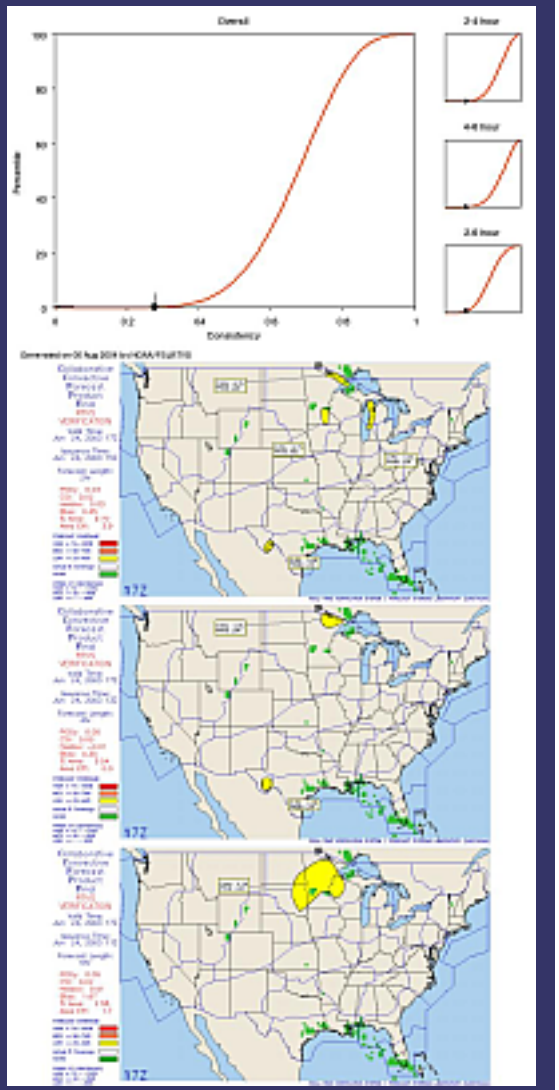
Combine with the actual forecasts to give complete information

Forecaster display





Poor consistency



Collaborative Convective Forecast Product Final
RTVS VERIFICATION
Valid Time: Jun 24, 2003 17Z
Issuance Time: Jun 24, 2003 11Z
Forecast Length: 6hr

PODy: 0.05
CSI: 0.02
Heidke: 0.01
Bias: 1.67
% Area: 2.58
Area Eff: 1.7

FORECAST COVERAGE
HIGH = 74–100%
MED = 50–74%
LOW = 25–49%
Actual % Coverage
NCWD

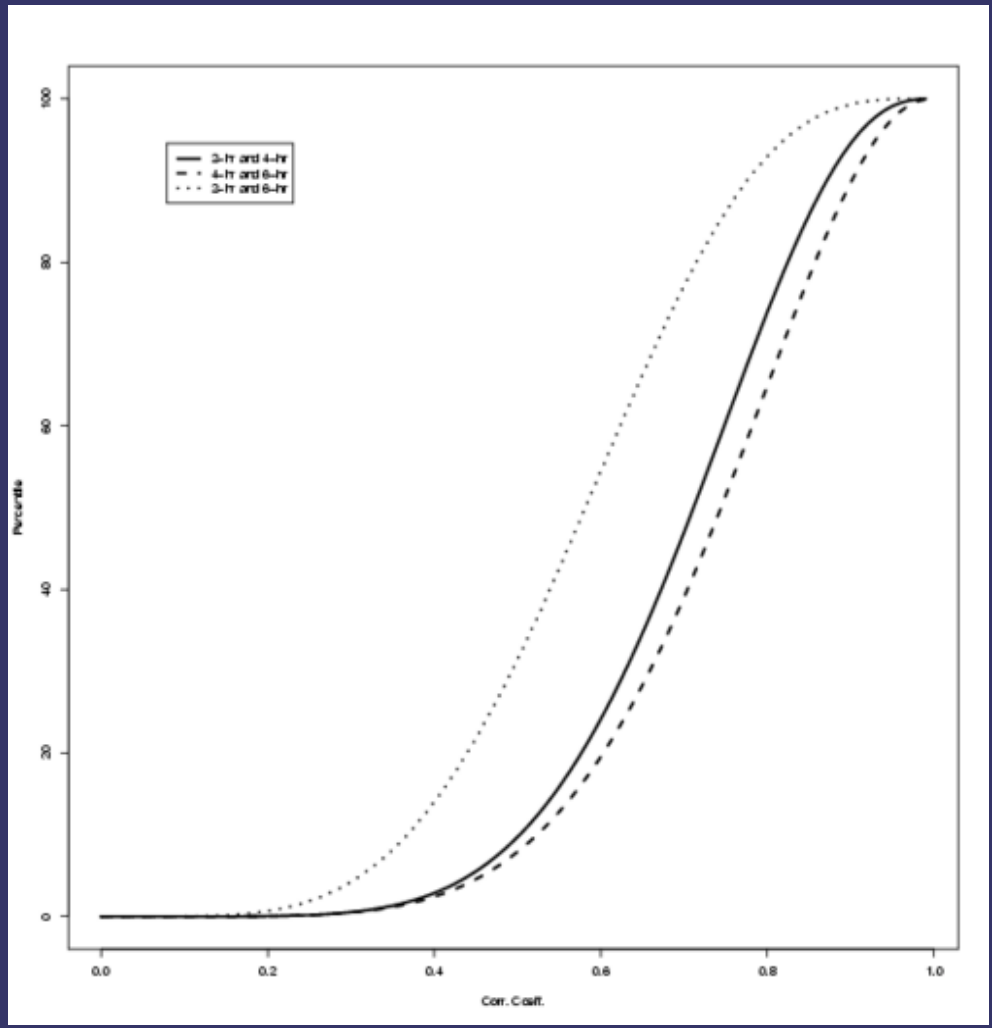
PROB OF CONFIDENCE:
HIGH = 70 – 100%
MED = 40 – 69%
LOW = 1 – 39%

CVRG: 2.7%
PROB: LOW

17Z

REAL-TIME VERIFICATION SYSTEM / FORECAST SYSTEMS LABORATORY (OAR/NOAA)

What can we learn from consistency?

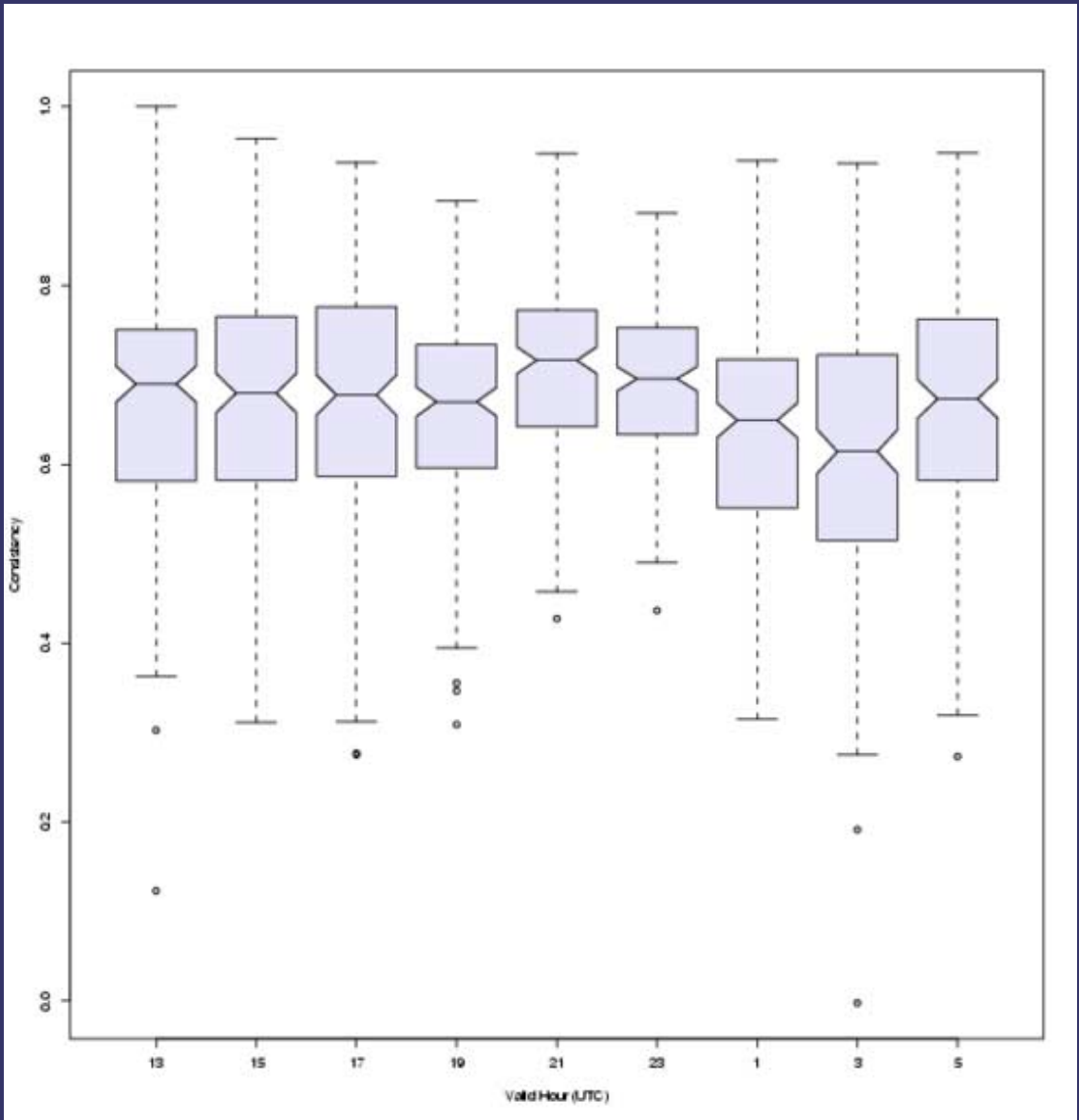


The 2-4 hr and 4-6 hr correlations are more similar than the 2-6 hr correlation

This is a Good Thing

If the distributions were identical there'd be no reason to issue the short-range forecasts and one could issue the 6-hr only!

How does consistency change during the day?

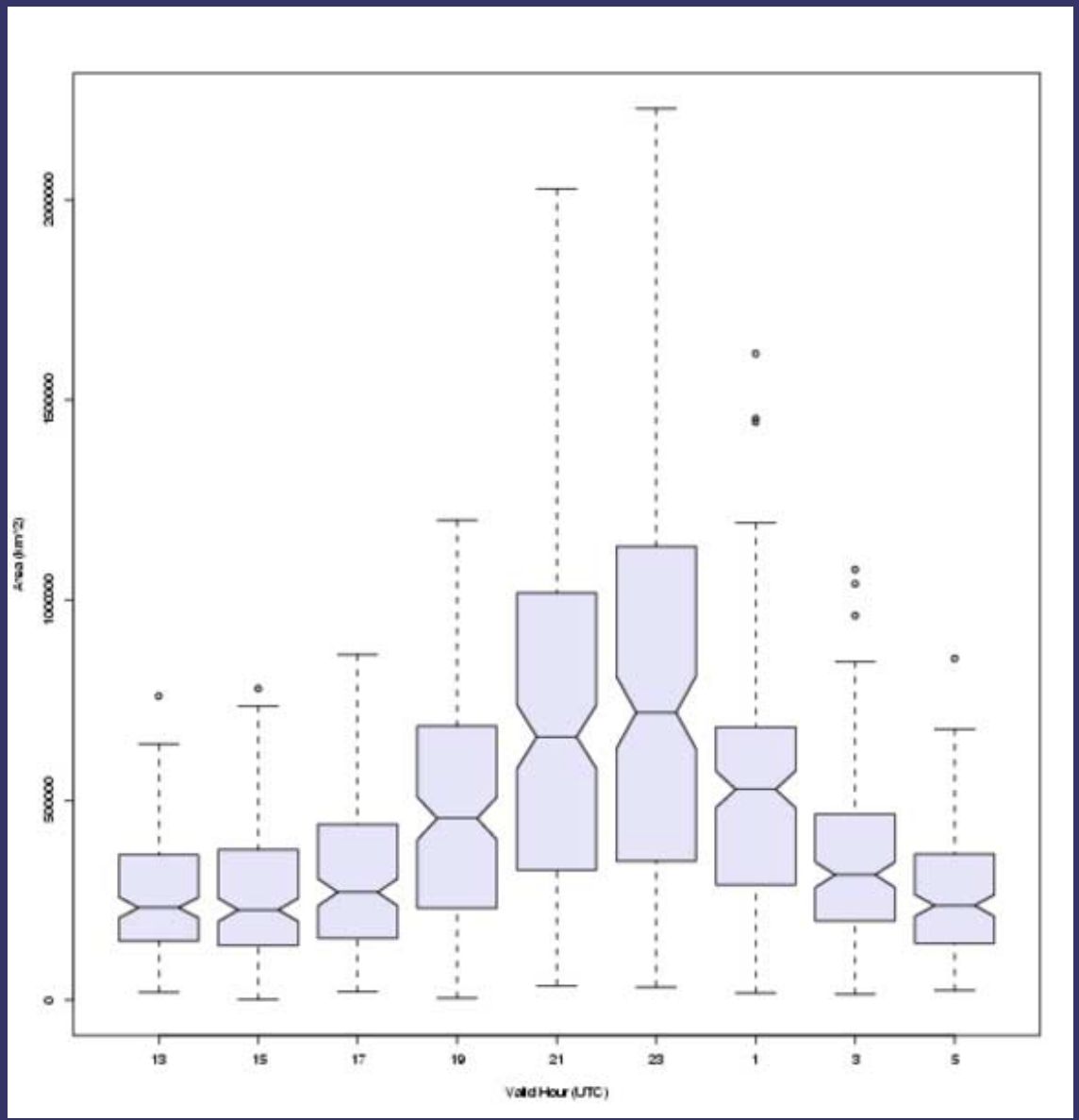


CCFP most critical in the morning *before* convective initiation

Clearly tied to diurnal cycle of convection

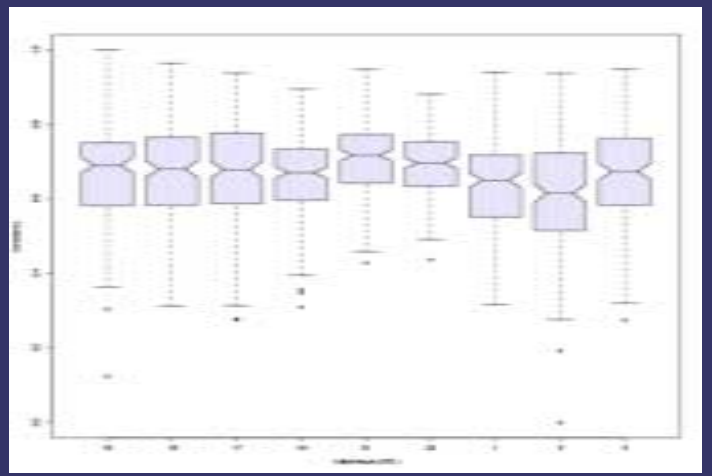
Consistency falls off markedly in the early evening – convective decay not well understood

Afternoon forecasts are large

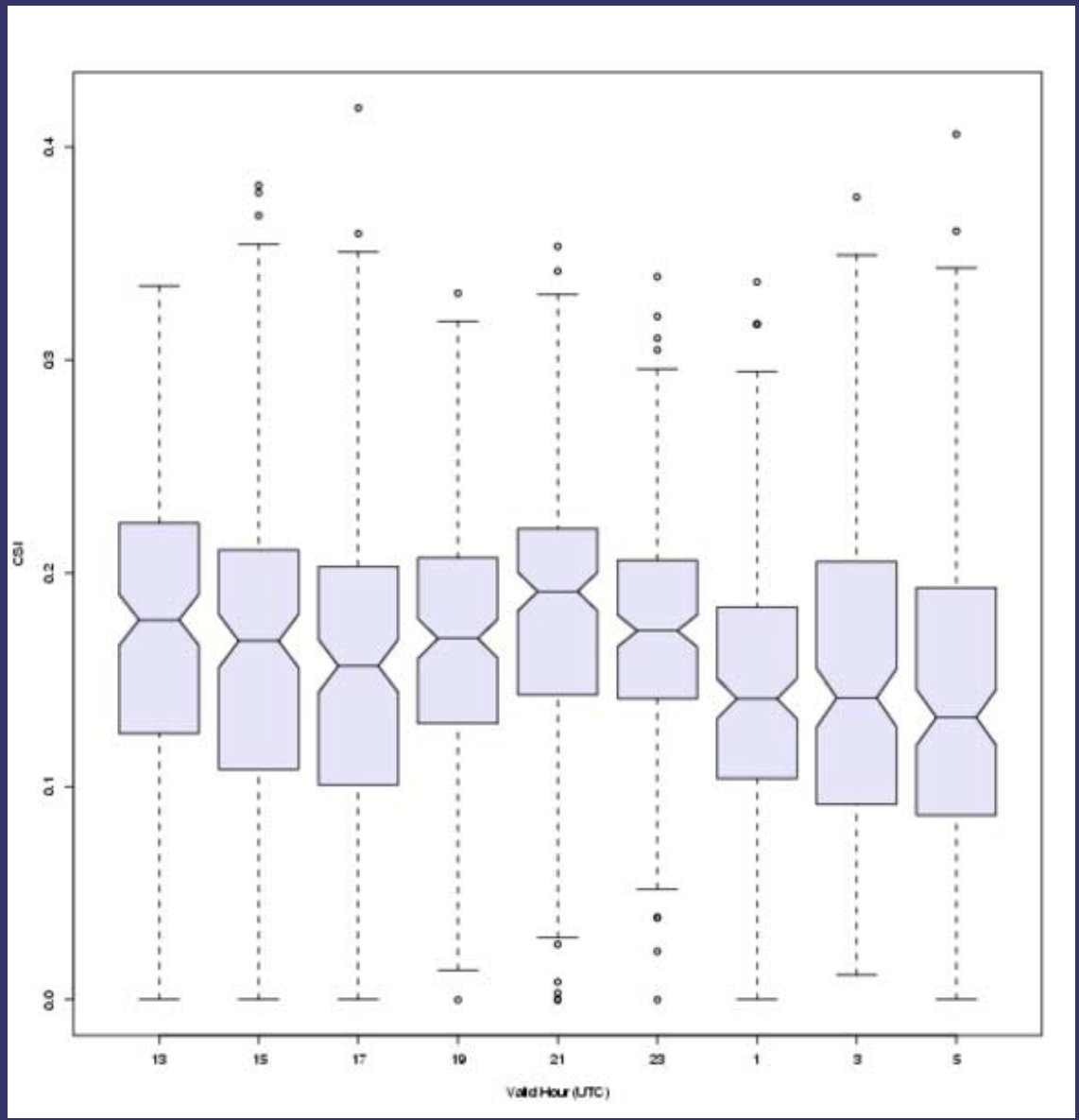


Diurnal trend is very pronounced

Consistency



Diurnal variation in skill

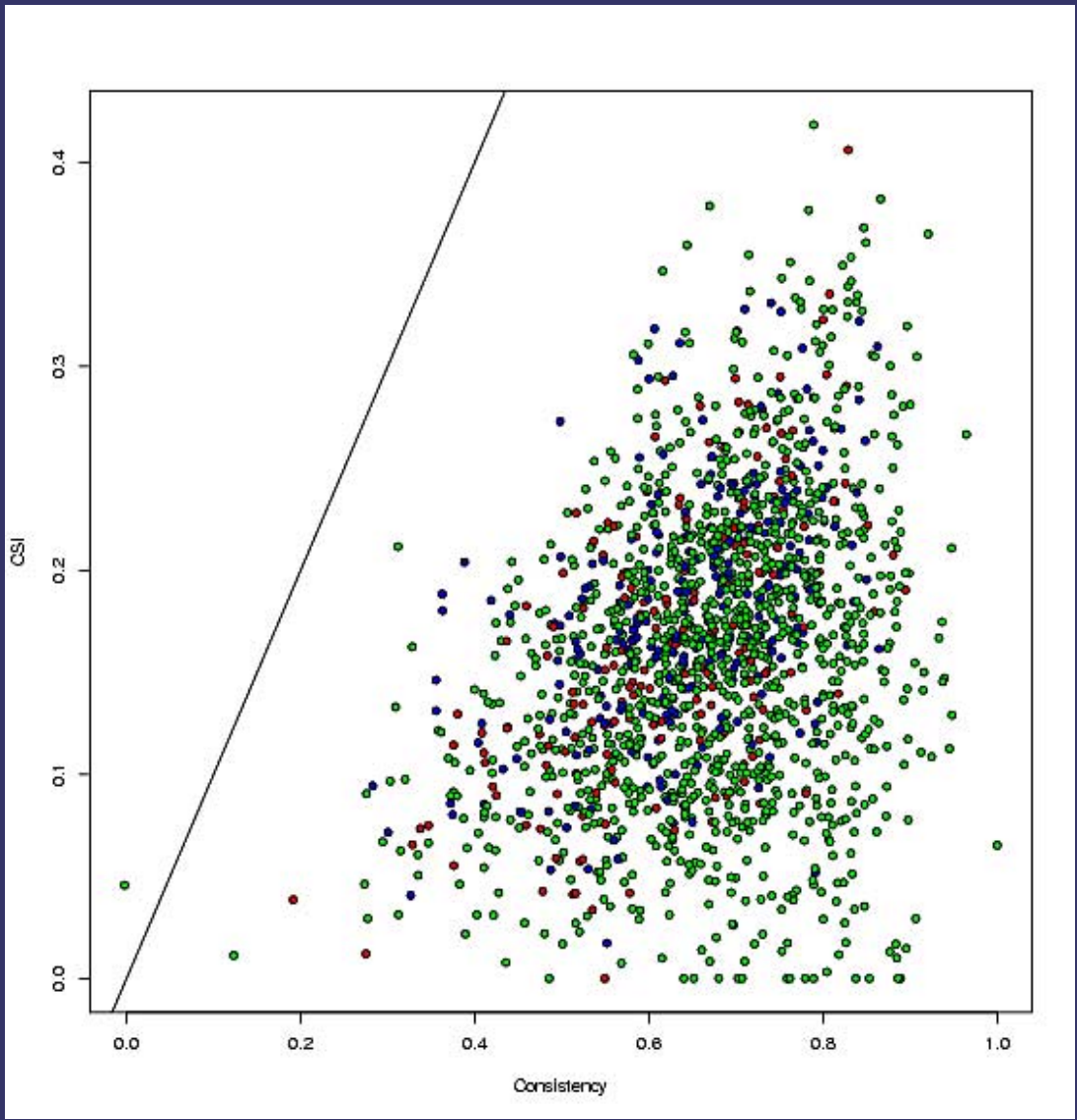


Average the CSI from each component forecast

Increasing skill during the day followed by rapid decline again in the evening

Unclear why there's good skill at 13Z

Can consistency be used as a proxy for accuracy?



This is how real users think

r=0.31

Stratification by areal trend does not indicate utility either (e.g., fcsts trending smaller or larger (indicating inconsistency) don't have higher CSI)

Conclusions

Consistency was quantitatively defined for a series of human-generated forecasts

Consistency can be used as a tool to illustrate aspects of forecasts that might otherwise be overlooked

Useful for forecast evaluation but not necessarily verification. (Useful as a forecast attribute however.)

Similar results to those found by Hamill (2003) for NWP forecasts indicate that consistency should not be considered a proxy for accuracy!

Low consistency rather than high consistency may be of most importance to users as it indicates times when they should be most alert