Quality Assessment Techniques for Evaluating Convective Weather Products used for Air Traffic Management

Steve Lack, Geary Layne, Mike Kay, Sean Madine, Missy Petty, and Jennifer Mahoney

Forecast Impact and Quality Assessment Section, NOAA/ESRL/GSD

Outline

- Goal
- QA Assessment Framework
- Diagnostic Evaluation Methods
- Translation Evaluation Methods
- Supplemental Evaluation
- Consistency Methodology
- Stratifications
- Conclusions

Goal

Independently evaluate the utility of various convective weather forecasts using a meteorological assessment framework in the context of the end-user

*There is no one-size-fits-all model to an evaluation (single score). Results using different methodologies must be put into context while telling a consistent, coherent story suitable for the decision maker at scales they are concerned with.

QA Assessment Framework

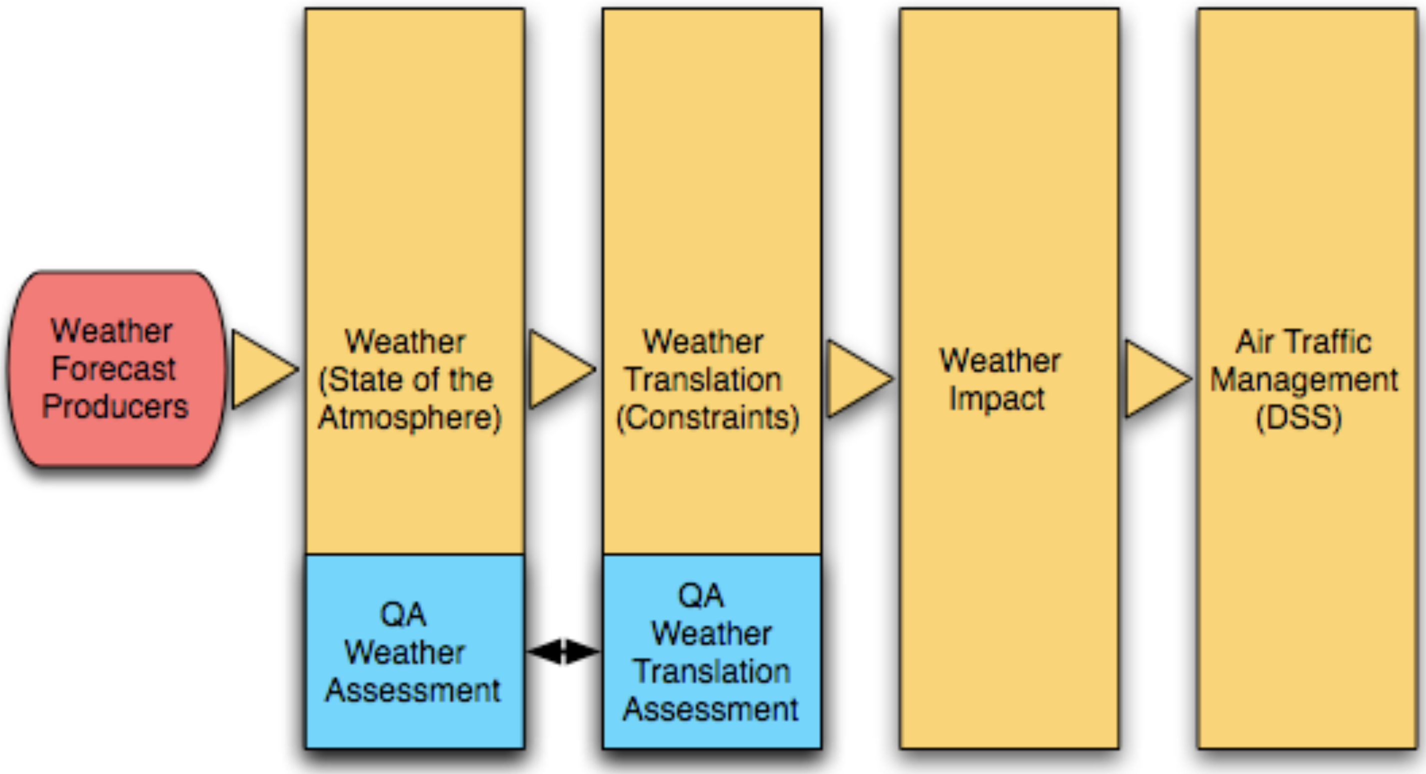
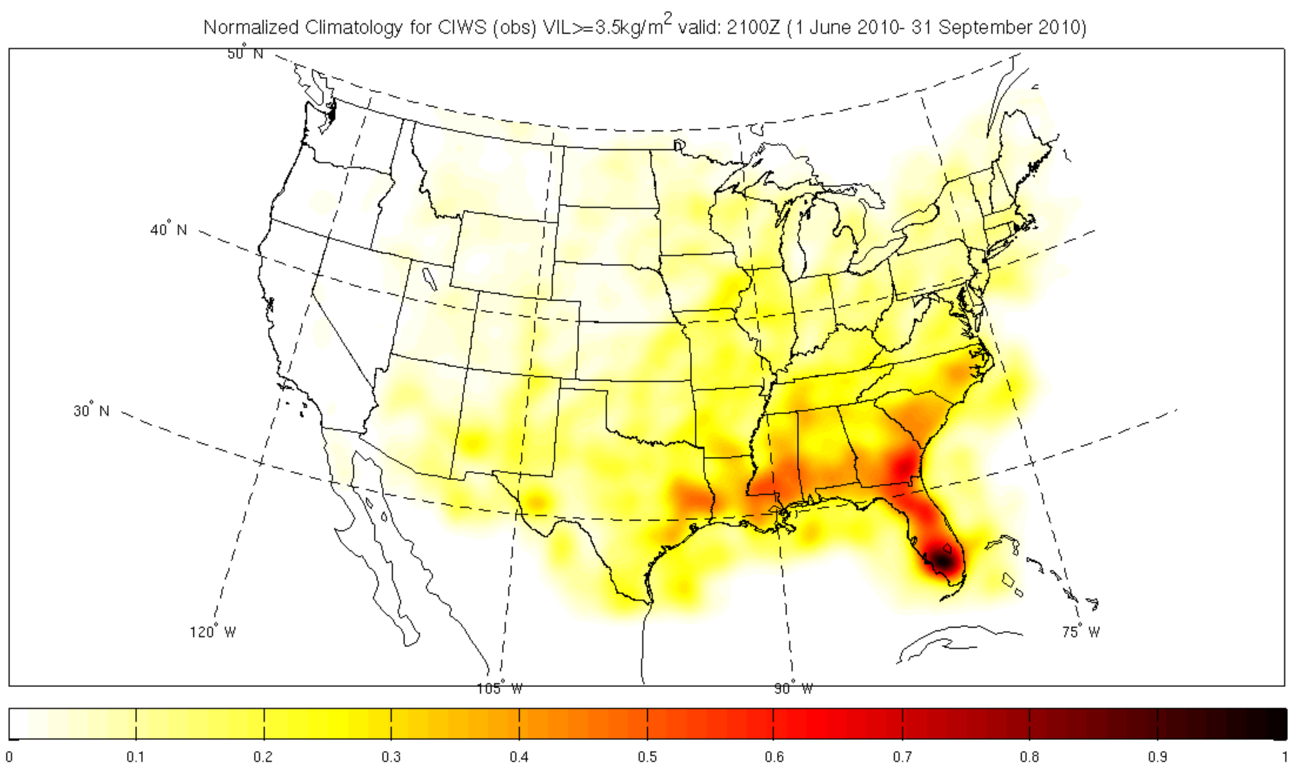


Diagram adapted from JPDO ATM Weather Integration Plan Draft v1.5 (June 30, 2010). QA Assessment addresses the quality of forecasts in the context of both Weather and Weather Translation

Diagnostic Evaluation Methods

- Climatological Analysis
  - Seasonal Trends
  - Diagnostic Images (Temporal and Spatial)
- Product Diagnostics (no inter-comparisons)
  - Upscaled categorical scores for deterministic forecasts
    - POD, FAR, CSI, Bias, etc.
  - Reliability diagrams for probabilistic forecasts
  - Categorical analysis
    - Distributions of areal coverage, echo top analysis

Normalized CIWS climatology Summer 2010



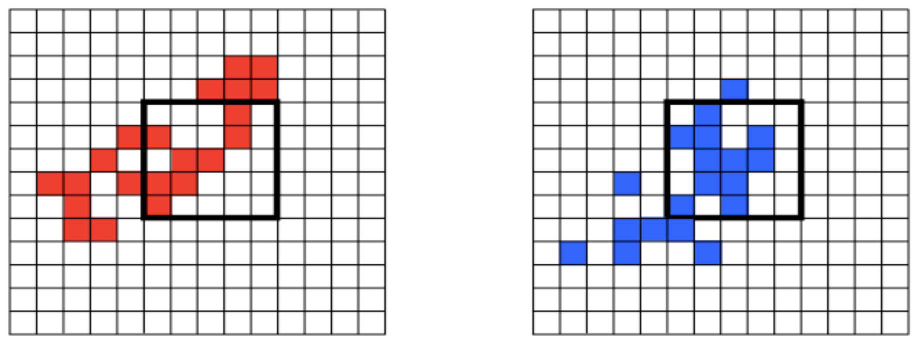
Examining climatology diagnostic images shows the user where convection occurs most of the time on a broad scale at select valid times for both the observations and forecasts.

Translation Evaluation Methods

- Translate forecasts and observations to common weather picture
  - Skill can be assessed at multiple scales
  - Treat all forecasts the same (probabilistic)
- Metric-based translation
  - Fractions Skill Score (FSS)
    - Apply baselines for comparison (climo and uniform)
- Permeability-based translation
  - Flow Constraint Index (FCI)
    - Mincut Bottleneck Approach

Example: Fractions skill score
(Roberts and Lean 2005)

Compares fractional coverage in forecast with fractional coverage in observations



observation   forecast

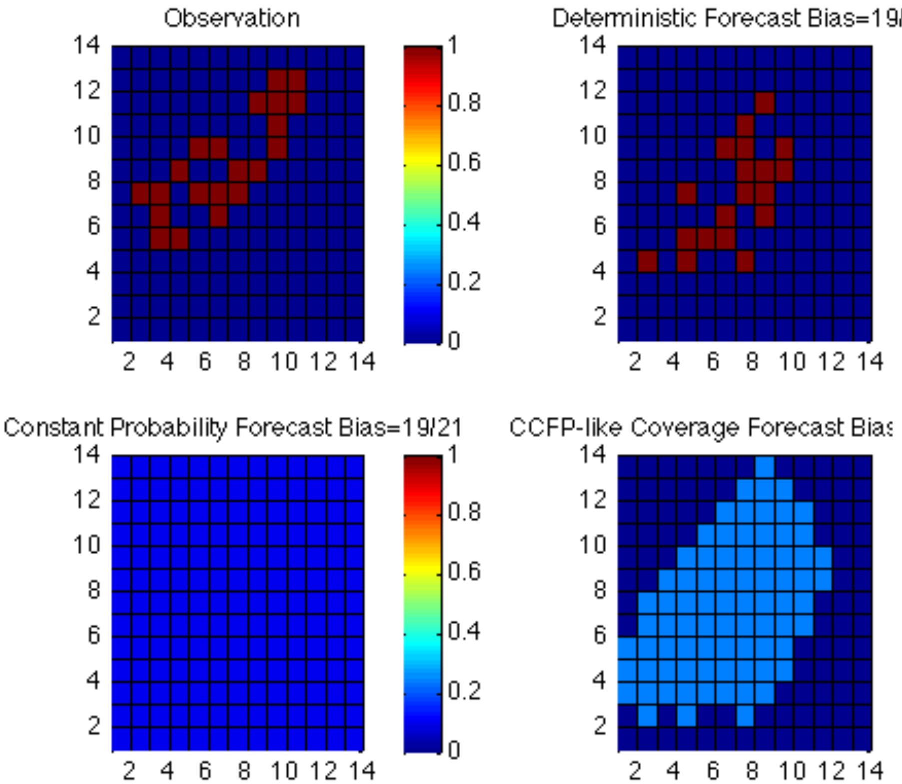
$$FSS = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}(P_{fcst} - P_{obs})^2}{\frac{1}{N}\sum_{i=1}^{N}P_{fcst}^{\;2} + \frac{1}{N}\sum_{i=1}^{N}P_{obs}^{\;2}}$$

Useful to see at what spatial scales a forecast starts to have skill

Australian Government
Bureau of Meteorology

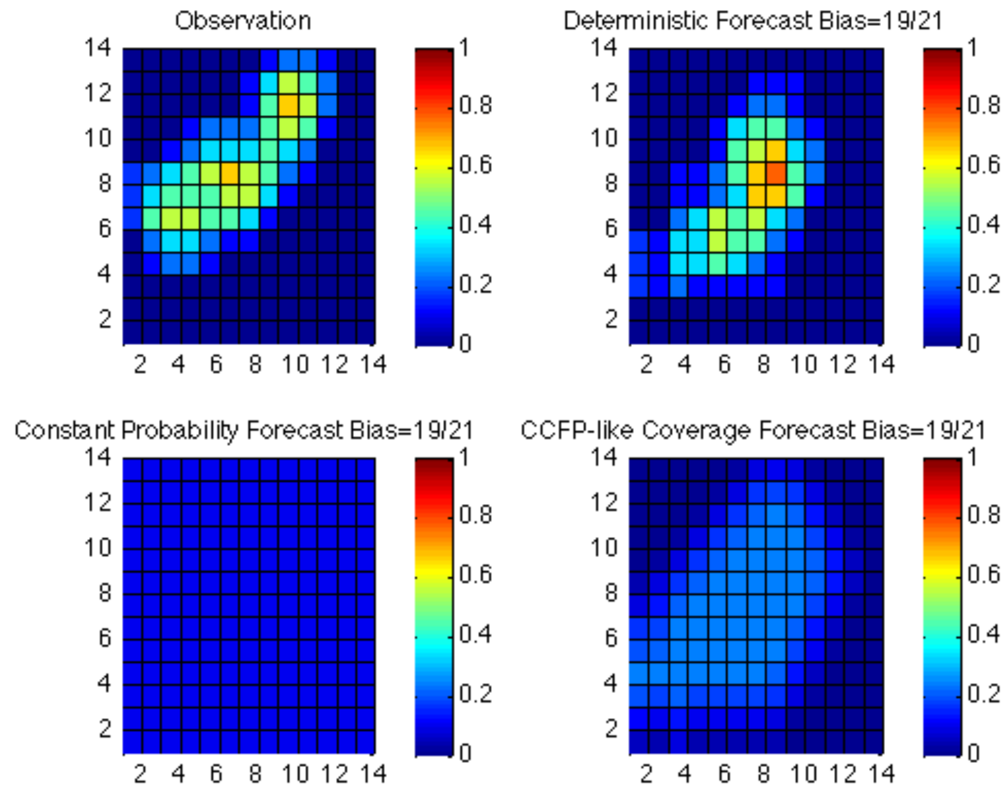
8

From Ebert, 2nd QPF Conference, Boulder, CO, 5-8 June 2006

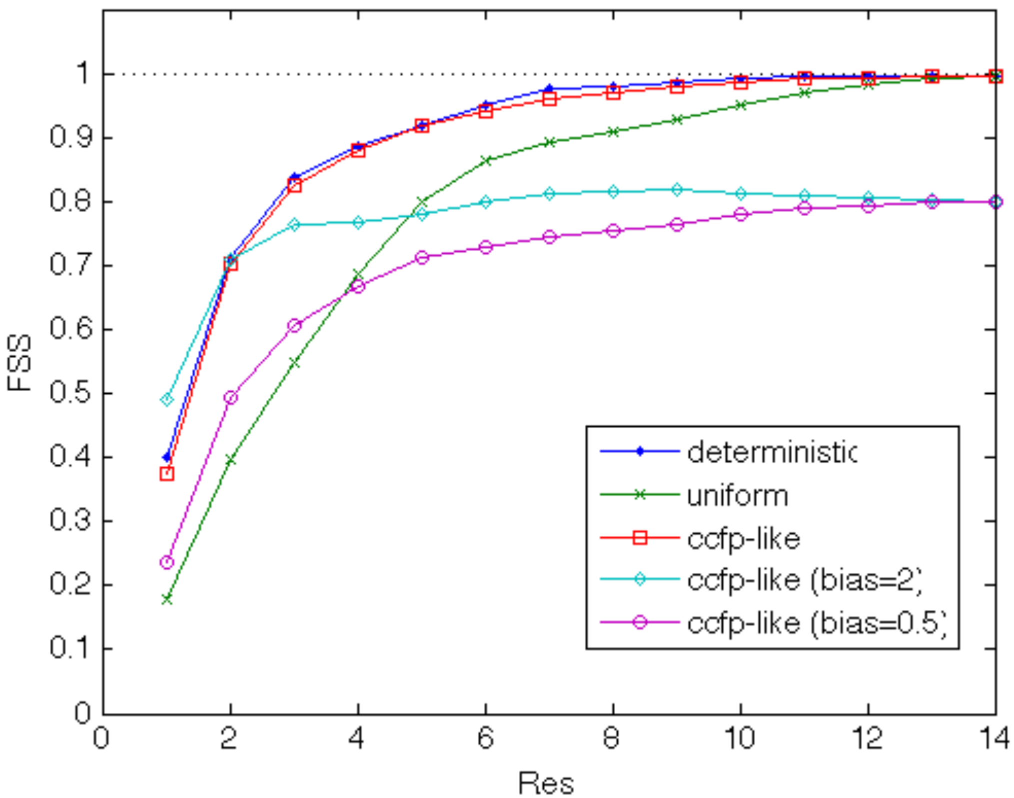


Native Sample Image

Fractionalized 3x3 Grid



FSS Sample Skill Image



The uniform field with bias close to 1 will score highly as the neighborhood approaches the size of the forecast domain. High or low bias will hurt forecasts as resolution increases.

Mincut Bottleneck as Applied (Flow Constraint Index FCI)

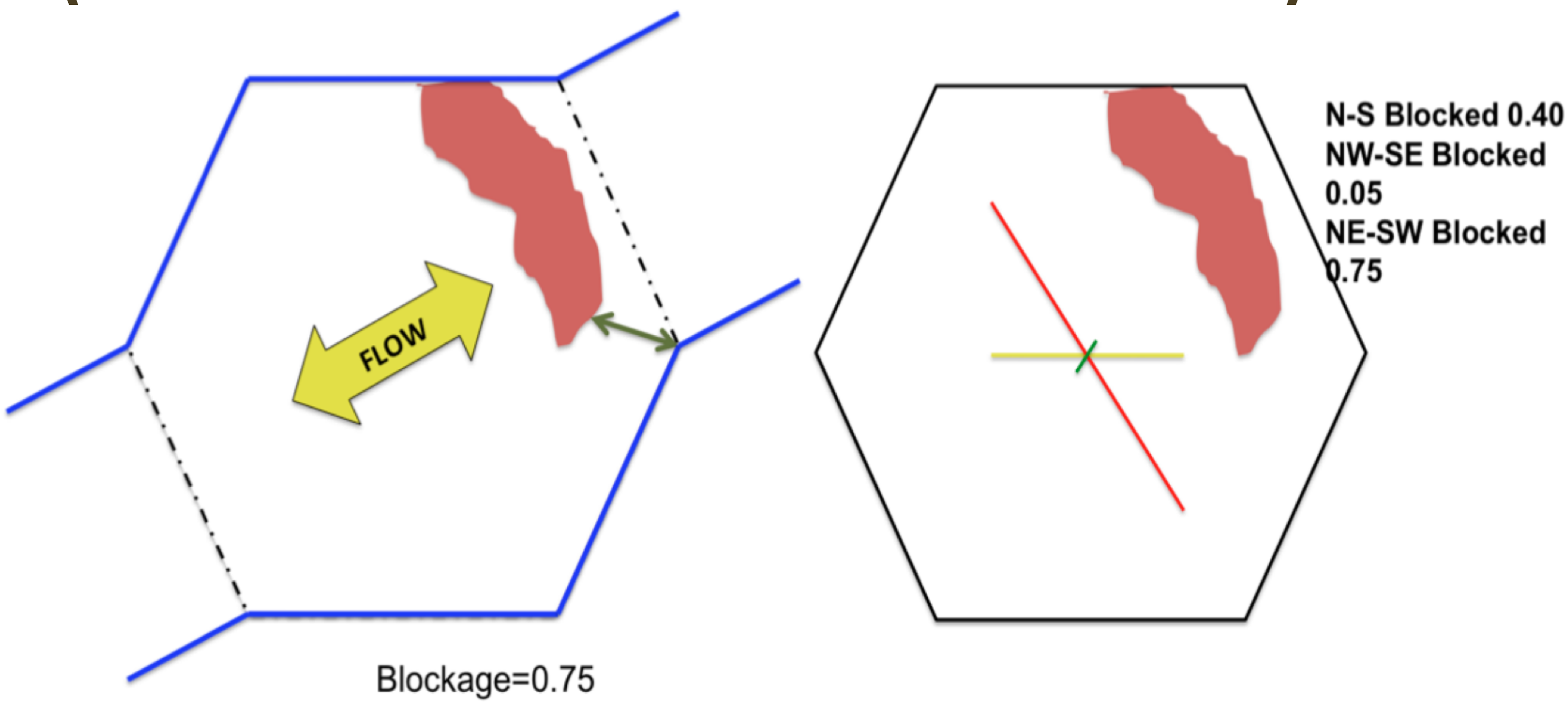
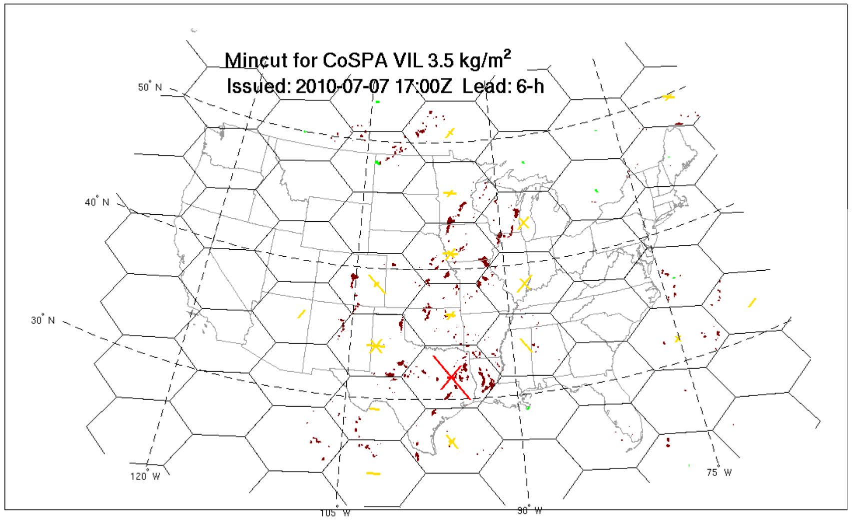
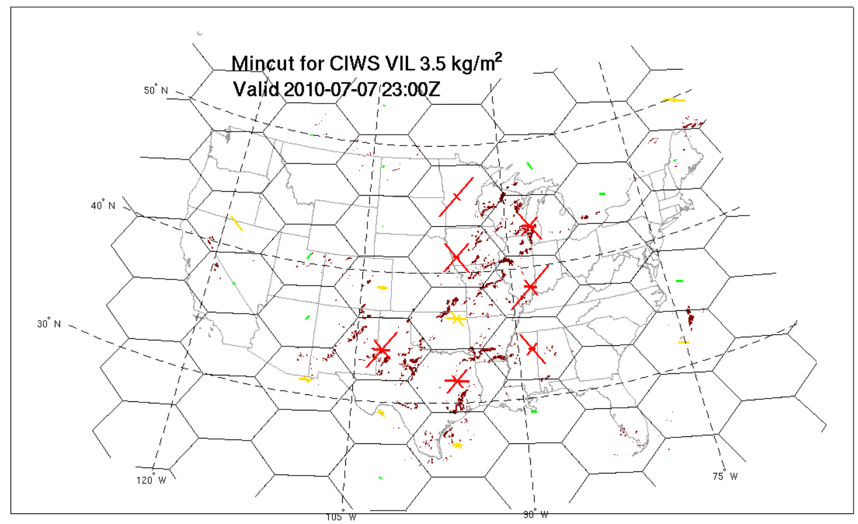
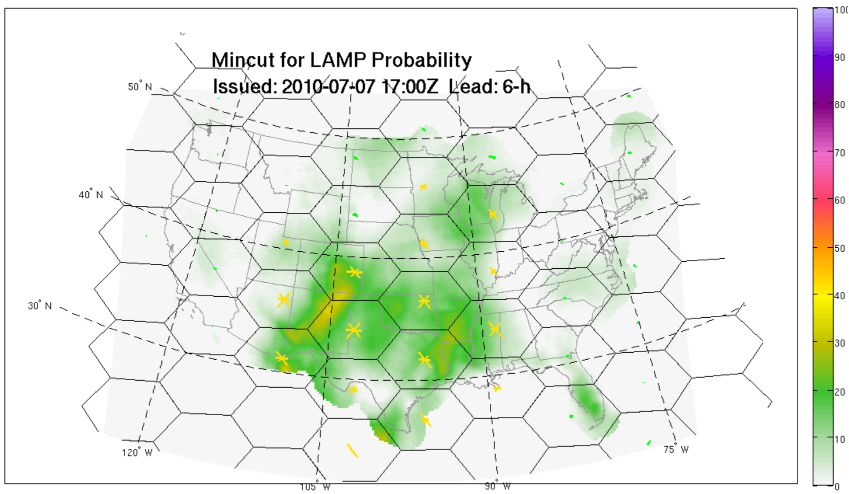


Illustration of the Mincut Bottleneck technique for a hexagonal grid; note that hazards can be defined by weather attributes or weather avoidance objects, and can be deterministic or probabilistic. Useful to assess the permeability of airspace. Can include jetway corridors over high altitude sectors or ARTCCs instead of hexagons.

Mincut Example: 2010-07-07
Valid 23Z, forecasts lead-time: 6-h



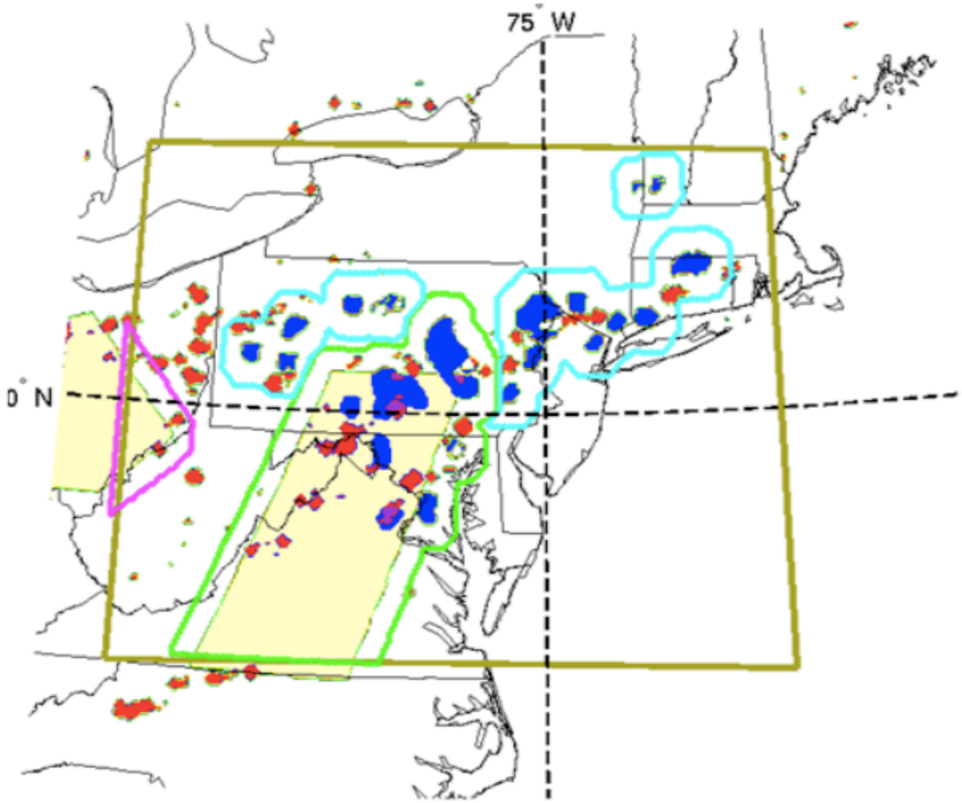




Mincut (FCI) Yellow Threshold (0.15) Scores in Vx Domain:

CoSPA Hits: 13; Misses: 2;
False Alarms: 2; CSI: 0.76
LAMP Hits: 12; Misses: 3;
False Alarms: 3; CSI: 0.67

Supplemental Evaluation



Example of domain decomposition for use in the supplemental evaluation. CCFP is shown in beige, a supplemental forecast is shown in blue and observations are in red. Forecast agreement of convection is outlined in green, disagreement in magenta and cyan.

Consistency Methodology

- Important for end-user for **confidence** to commit to decisions
- Uses FCI (Mincut-Bottleneck) translation field
- Consistency metric uses the Correspondence Ratio (CR) (Stensrud and Wandishin 2000)

$$CR = \frac{I}{U} = \frac{\sum_{i=1}^{n} f_{1,i} \bigcap f_{2,i} \bigcap .. \bigcap f_{m,i}}{\sum_{i=1}^{n} f_{1,i} \bigcup f_{2,i} \bigcup .. \bigcup f_{m,i}}$$

(m is the mth forecast and i is the ith point with n being the total number of points)

- Comparison of a single forecast product across the same valid time
- Comparison of multiple forecasts as same lead-time
- Other consistency scores will be assessed

Stratifications

- Primary Stratifications
  - Hazardous convection which is defined at VIP-level 3
  - Pre-initiation issuances corresponding to strategic planning telecons
    - Sensitive times for strategic planning
  - Examine skill in traffic-sensitive areas (NE core)
- Additional Stratifications
  - Post-initiation issuances
  - SE US and Gulf Coast regions (ATL, JAX, MEM, IAH, MIA)
  - Western CONUS (PHX, DEN, SLC)
  - Multiple VIP-level thresholds and Echo Top thresholds

Conclusions

- There is no single score that can define the value of the product, it must be looked at it in context using end-user guidelines
- The methodologies applied should provide a consistent and coherent picture of the value of the product
- Scores should incorporate confidence intervals, especially when comparing different sample sizes
- Forecasts are being produced and modified rapidly and there is a need to baseline results (LAMP, CoSPA)
  - Consistent verification methodology
  - New versions of forecasts run against old versions over the same time periods
  - Characterization of the benefits of the new version (look at individual moving pieces…small changes can have big consequences)

This research is in response to requirements and funding provided by the Federal Aviation Administration. The views expressed are those of the authors and do not necessarily represent the official policy and position of the U.S. Government.