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

Steven A. Lack[1,2], Geary J. Layne[1,2], Michael P. Kay[1,2], and Sean Madine[1,3], Melissa A. Petty[1,3], and Jennifer L. Mahoney[1]

NOAA Earth System Research Laboratory (ESRL), 325 Broadway, Boulder, Colorado 80305[1]

Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado at Boulder, UCB 216, Boulder, Colorado 80309[2]

Cooperative Institute for Research in the Atmosphere (CIRA), Colorado State University, Ft. Collins, CO 80523[3]

1. Introduction

There is no one-size-fits-all model (i.e. a single score) to assessing the skill of a forecast. Attempting to summarize skill in a single score may not give an end user sufficient information required to determine the usefulness of a particular forecast. Additionally, using a narrow set of meteorological verification contexts and techniques can yield misleading information on how the products may benefit an end user (e.g. an operational air traffic planner). To this end, results using different methodologies must be put into context while telling a consistent, coherent story suitable for the decision maker. When the products being evaluated have fundamentally different characteristics (e.g. probabilistic vs. deterministic), forecasts should be translated to a common field using a methodology that is consistent with how the product is interpreted by the user. New forecasts being considered for operations must be compared and evaluated with respect to current operational standards; often these candidate forecasts are considered to be supplemental products to the operational standard as they are being introduced to the user community. This paper presents a framework that encompasses techniques that can aid a decision maker in determining the utility of the candidate forecast. This framework includes but is not limited to a metric-based translation, a permeability-based translation, and a supplemental analysis with respect to the current operational standard. These transformation processes and the benefits and issues with their implementation will be presented in the following sections. Section 2 will discuss the metric-based translation using the Fractions Skill Score (FSS) approach. Section 3 will discuss the impact-based translation using the Flow Constraint Index (FCI), which is based on the Mincut Bottleneck translation technique. Section 4 will examine the techniques used to assess a product as a supplement to an operational standard forecast.

The work herein was driven by the Quality Assessment Product Development Team (QA PDT) task to determine the value of advanced convective forecasts as they pertain to aviation planning, specifically strategic Traffic Flow Management (TFM), for the 2010 convective season (Madine et al., 2011). The current operational convective forecast, the Collaborative Convective Forecast Product (CCFP), was used to define the baseline of performance against which experimental products were compared. CoSPA (Wolfson et al., 2008) and the Localized Aviation MOS Program (Ghirardelli, 2005) /Collaborative Convective Forecast Product Hybrid (LAMP CCFP Hybrid, LCH), along with CoSPA's parent model the High Resolution Rapid Refresh (HRRR; Weygandt et al., 2010), were the focus of the study.

Corresponding author address: Steven A. Lack, 325 Broadway R/GSD5, Boulder, CO 80305, steven.a.lack@noaa.gov

2. Fractions Skill Score

The Fractions Skill Score (FSS) developed by Roberts and Lean (2008) is a neighborhood-based verification approach commonly used in assessing the skill of numerical weather prediction (NWP) models at various resolutions. The FSS compares the percent coverage in the forecast to the percent coverage of the observation for a given neighborhood about a reference pixel for all pixels in the forecast field, and is given is given in equation (1). The FSS has a valid range between 0 (worst) and 1 (best) and is similar in interpretation to the critical success index (CSI).

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

The FSS can be applied to any type of forecast (dichotomous or probabilistic). There is no need to threshold probabilistic forecasts as all forecasts are treated probabilistically in this methodology. Dichotomous forecasts are thresholded at an operationally meaningful threshold; the probabilistic forecasts are assumed to be forecasting probabilities of events defined by this threshold. The dichotomous forecasts are simply assigned probability 1 for an event meeting or exceeding a threshold and 0 otherwise. Treating the forecasts and the observations the same makes this metric powerful for forecast comparisons.

Figure 1 shows an example of how percent coverage is calculated in a domain. A 5x5 neighborhood is created around the center pixel for this example. The observation at the center pixel is assigned a value of 0.32 (P_{obs}) and the forecast at the center pixel is assigned a value of 0.44 (P_{fcst}). This procedure is repeated for all pixels in the native domain and equation 1 is applied to the results for the calculation of the FSS using a 5x5 neighborhood.

The example in Figure 1 is further expanded upon with the inclusion of a constant probability forecast and CCFP-like forecasts. The constant probability forecast, also known as the uniform forecast, can be viewed as a useful baseline of skill when displaying plots of the FSS. The uniform forecast is simply the base rate of the observation applied at every pixel in the forecast domain. Figure 2 shows the observation and deterministic forecast as a binary image along with a forecast of constant probability and a CCFP-like forecast. In this example, both the uniform forecast and CCFP-like forecast are created so that the bias of each new forecast is consistent with the original deterministic forecast (bias=19/21; 0.905). The CCFP-like forecast is simply a dilation of the deterministic field so that the two forecasts represent a forecast over approximately the same region of the domain. Figure 3 shows the fractionalized grid for a 3x3 neighborhood for each of the forecasts in Figure 2.

It can be shown that as the neighborhood of interest approaches the size of the entire study domain, the FSS is a function of the forecast bias given by (2). This is demonstrated in Figure 4, which includes the FSS curves for the three forecasts in Figure 2, with two additional CCFP-like forecasts with bias=2 and bias=0.5.

$$\lim FSS_{neighborhood \to Domain} = 1 - \frac{(bias_* - 1)^2}{(bias_*^2 + 1)}$$

$$\text{where } bias_* = {}^{1}\!/_{bias} \text{ if } bias < 1 \qquad (2)$$

As shown in Figure 4, the FSS curves for the deterministic, uniform, and CCFP-like forecasts converge to the same value (~0.995) following equation 2 as they all have the same overall bias. The CCFP-like forecast with bias=2 and bias=0.5 also converge to the same solution; however, the high bias forecast is favored at high resolutions. It may be possible for a forecast to give a false indication of skill with the FSS at high resolutions by having spatially accurate forecasts with biases slightly larger than 1. The uniform forecast has relatively poor skill until the full domain is approached (no sharpness).

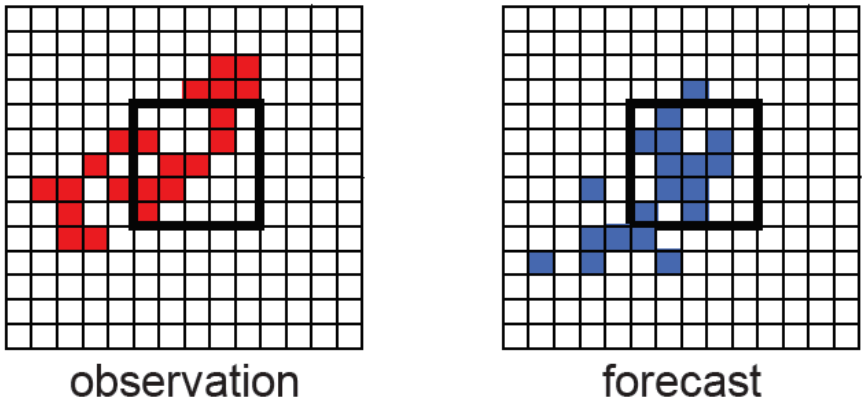

Figure 1. Example of 5x5 neighborhoods about the center pixel for input into the FSS. (Borrowed from E. Ebert)

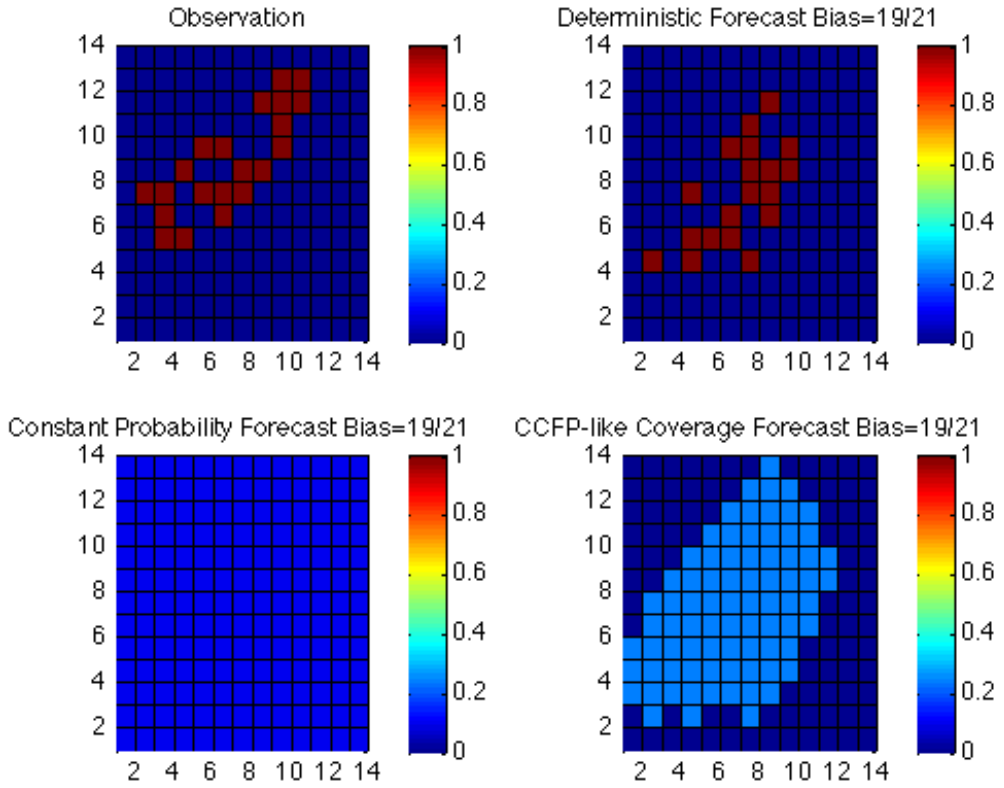

Figure 2. Observation field (top left) with a deterministic forecast (top right), a constant probability forecast (uniform; bottom left) and a CCFP-like forecast (bottom right) for the illustration of the FSS. All forecasts have the same bias (fractional coverage).

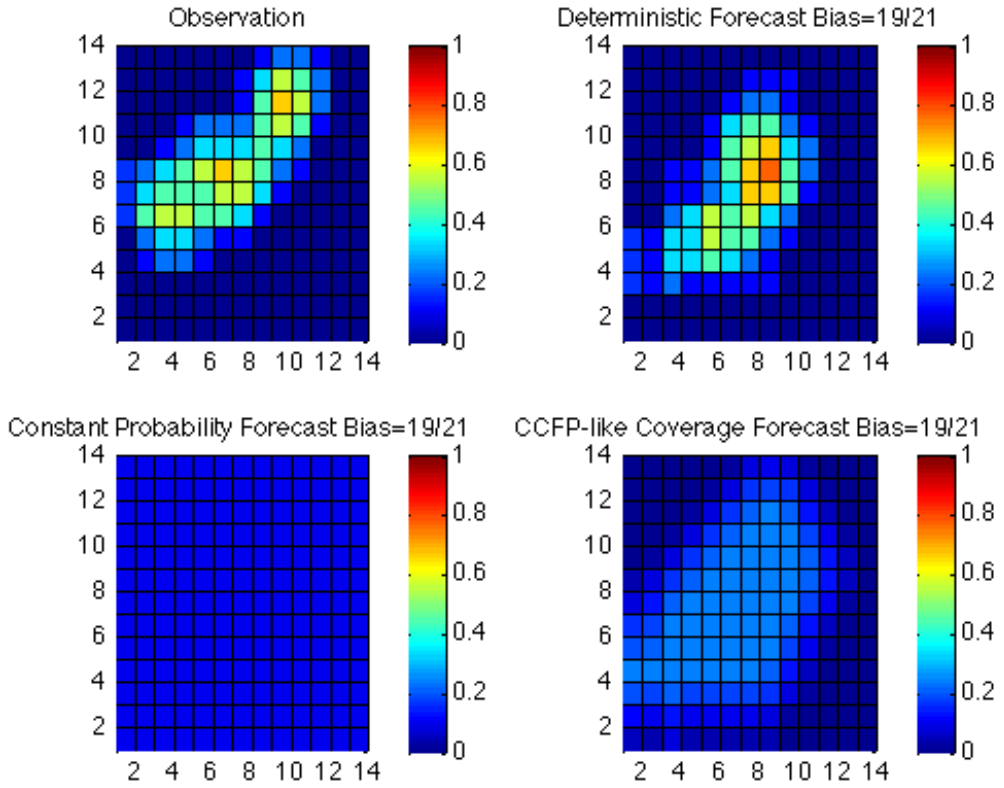

Figure 3. Fractionalized grid for a 3x3 neighborhood for the observation field (top right) with the deterministic forecast (top right), the constant probability forecast (uniform; bottom left) and the CCFP-like forecast (bottom right).

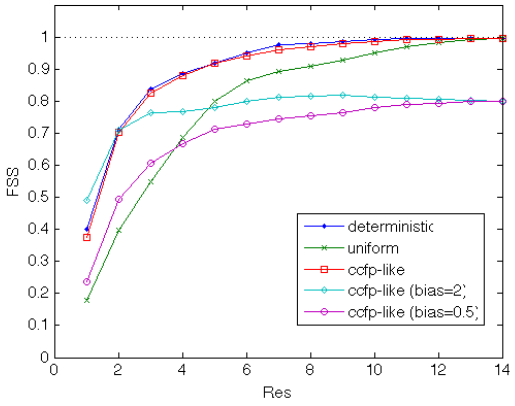

Figure 4. FSS as a function of resolution for the five different idealized forecasts.

3. Flow Constraint Index

The Flow Constraint Index is a measure of the hypothetical reduction in air space capacity due to the presence of convection. This measure can provide insight into the meteorological structure and resolution to which air traffic flow may be most sensitive. The methodology follows closely with Layne and Lack (2010) which has been adapted from Krozel et al. (2004).. The strengths of this methodology are 1) the forecast and observation are translated to facilitate comparisons in an aviation operational environment, and 2) like the FSS this metric allows a direct comparison of dichotomous and probabilistic forecasts.

The Mincut-Bottleneck approach estimates permeability by calculating the

minimum distance across a specified geometry from source to sink (perpendicular to the corridor of flow) using obstructions (convective objects) as nodes. Permeability is defined as the ratio of the minimum distance found with convection present to the minimum distance across the corridor. Flow reduction is defined as one minus this ratio (3). This is referred to as the Flow Constraint Index (FCI).

$$\mathrm{FCI} = 1 - \frac{Mincut_{convection}}{Mincut_{corridor}} \qquad (3)$$

A simplified example using a regular grid for a deterministic forecast or observation is shown in Figure 6. The method can be visualized as path lengths (PL) through probabilistic and deterministic forecasts or observations as shown in Figure 7. The minimum paths across the corridor are given in parentheses and shown with arrows; the FCI (flow reduction) is also given.

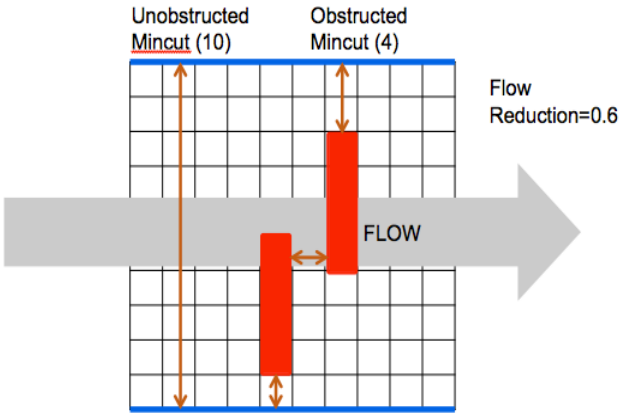

Figure 6. Simplified 10x10 grid showing an example of the Mincut-Bottleneck technique. The deterministic convective objects are shown in red, obstructing the implied flow in gray for the given corridor shown in blue. The minimum paths across the corridor are given in parentheses and shown with arrows.

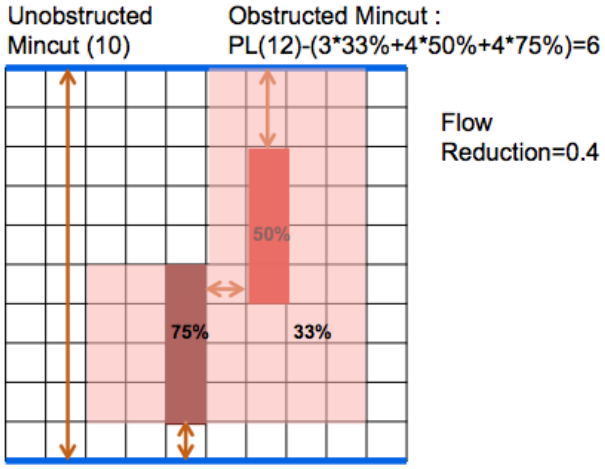

Figure 7. Simplified 10x10 grid showing an example of the Mincut-Bottleneck technique for a probabilistic field. Shades of red show varied probabilistic convective fields between the corridor bounds shown in blue. The minimum paths across the corridor are given in parentheses and shown with arrows.

For the work herein, the Mincut-Bottleneck approach estimates permeability in a hexagonal grid. For this study two hexagon heights were chosen, 75-nm and 300-nm to approximate the size of the average super high sector and that of Air Route Traffic Control Centers (ARTCCs), respectively. These hexagon grids are overlaid onto the CONUS to intersect the forecast domain. This allows useful comparisons of different candidate forecasts in an operational framework at scales of interest to strategic decision-making. For a given hexagonal grid three corridors are defined for permeability calculations. Figure 8 (left) shows one of the three corridors of interest with the hexagon highlighted by the blue lines. In this example, the convective node (in red) is represented by an observation at the hazardous convective threshold for aviation (VIP-level 3); however, any meaningful threshold can be applied. The green arrow represents the minimum

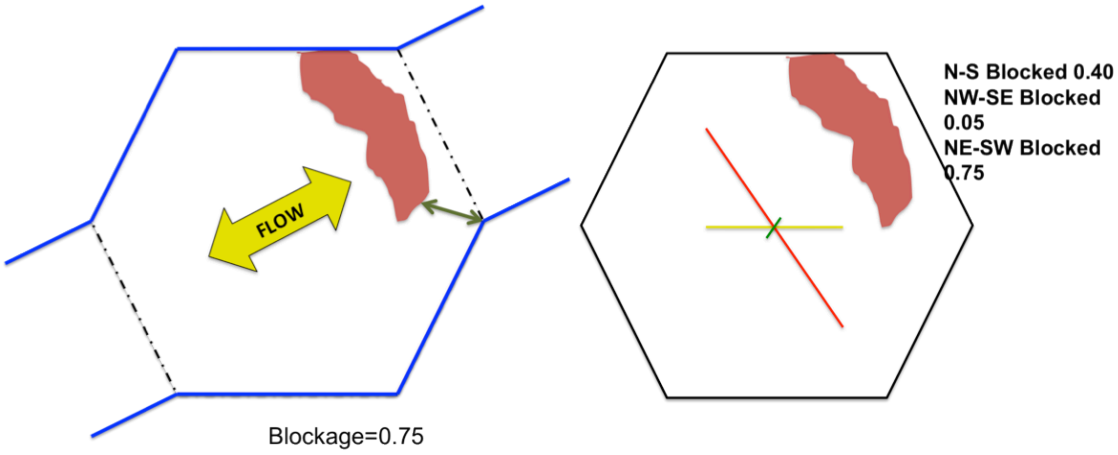

Figure 8. An example of the Mincut-Bottleneck technique for a given hexagon given a convective object (red). The mincut is shown by a green arrow for one corridor (blue) in the given hexagon that results in a flow reduction for the given flow direction (yellow) (left). The resulting FCI for all three corridors in the hexagon is shown (right).

distance from the convective node to the edge of the corridor. This is repeated for all three corridors in the hexagon and the result is shown in Figure 8 (right). The lines in Figure 8 (right) indicate the FCI of the hexagon for each corridor. The lines run perpendicular to the flow direction of their associated corridor, and the length of the line is proportional to the magnitude of the corridor's FCI. In this example, the NE to SW corridor is most impacted by the convective object, while the NW to SE route is the least impacted.

A deterministic forecast (e.g. CoSPA) is strictly defined as probability=0 for a pixel not reaching the threshold and probability=1 for all pixels meeting or exceeding the defined threshold. It is important to note that raw probabilities in the forecast are utilized when the FCI translation is applied (i.e. no thresholds are applied to probabilistic fields). Once the translation is completed, standard categorical skill scores can be calculated for a series of FCI values. This is accomplished by setting up the standard 2x2 contingency table for a given FCI threshold, including: hits, misses, false detections, and correct negatives. A primary analysis metric using the FCI (blockage) in the hexagons is the critical success index (CSI). A resultant plot would show CSI values plotted against increasingly impactful events (FCI values) for each forecast product of interest.

4. Supplemental Relationships

The assessment of a forecast that may be used in operations often goes through a process to evaluate the skill of the candidate forecast as it supplements the operational standard. Following are two approaches to assess supplemental relationships. The first approach involves the use of sub-domains based on the operational standard product, the CCFP. The forecast domain is partitioned into sub-domains of forecast agreement and disagreement, where forecast skill is then evaluated. The frequency of occurrence of each of the sub-domains is equally important as the skill of the product within each sub-domain. For example, the sub-domain defined by the presence of both the

supplemental forecast and the operation forecast may occur 80% of the time and on average cover 20% of the forecast domain.

The second approach for evaluating supplemental relationships uses clustering techniques to examine the candidate forecasts at CCFP-like scales. This allows for the examination of skill at CCFP-like granularity to see the benefit that high-resolution forecasts may potentially add in terms of sharpness or confidence. For instance, if there is considerable agreement between a high-resolution forecast when it is transformed into a CCFP-like product and the operational forecast, the end user may have more confidence in trusting both forecasts. This work closely follows Lack et al. (2010a).

The transformation of a deterministic observation or forecast field into CCFP-like scales is accomplished Using Fast Fourier Transform (FFT) band passes to convert spatial intensity to frequency space following Lack et al. (2010b). An example of this clustering technique is shown in Figure 9 for radar reflectivity over Texas. For each cluster identified, the percent coverage of observations is calculated for those clusters exceeding the minimum size criterion for CCFP (3000 sq mi). From the obtained percent coverage for the identified cluster, a CCFP coverage category is assigned based on historical distributions of observed coverage for sparse coverage/low confidence, sparse coverage/high confidence and medium coverage and above areas. An example clustering of the CIWS analysis field is shown in Figure 10. This example depicts areas of strong frequency signals above the VIP 2 threshold. With this rule in place the convection in N Alabama and W Tennessee is not identified as a CCFP-like polygon due to its isolated nature and large separation between convective objects.

5. Stratifications

In addition to having meaningful metrics that tell a consistent story, appropriate stratifications should be made based on both meteorologically significant regimes and user-specific evaluation goals. Climatological examinations help set up stratifications in meteorological contexts, such to identify locations of interest where dynamic forcing may be different over the study period. Knowledge of air traffic patterns and decision-making criteria aid in user-specific evaluation goals. It also serves to identify sensitive times where having good forecast performance is crucial, onset and cessation. Finally, it is important to put results in the context of where the heaviest air traffic is located, i.e., the NE US, as this is what often drives the needs of the end user.

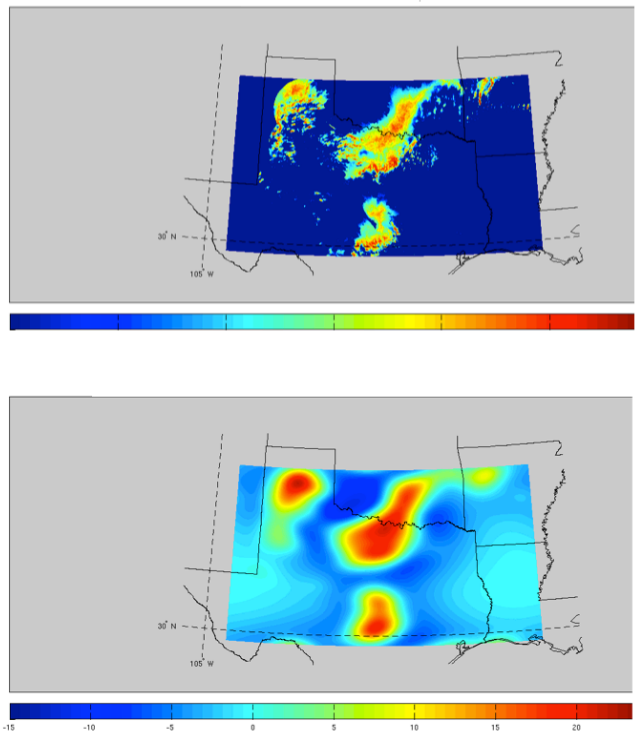

Figure 9. Reflectivity composite in N TX (top) FFT clustering example for a particular band pass (bottom). If thresholded at yellow 4 CCFP-sized features are evident at this particular pass.

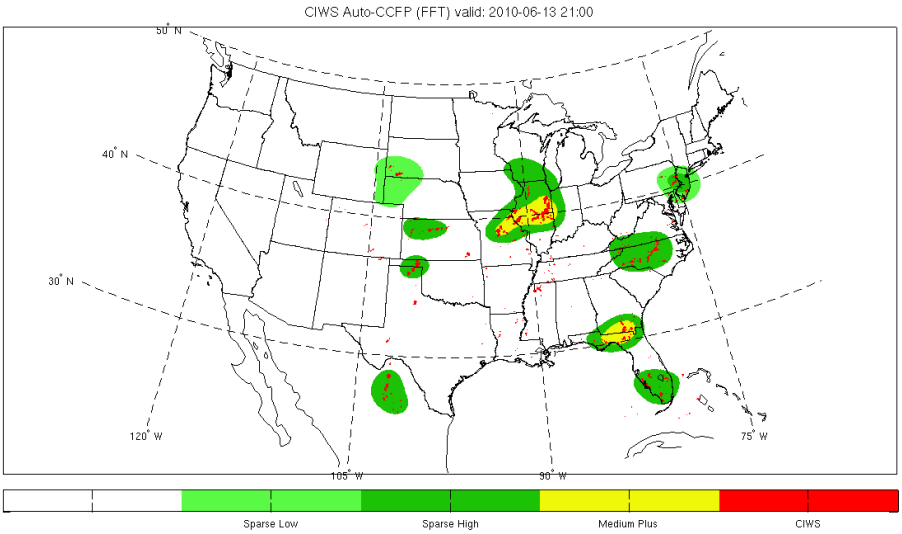

Figure 10. An example of a scaled CIWS field valid at 21Z on 13 June 2010, light green represents sparse/low coverage, dark green sparse/high coverage, yellow medium coverage or higher, and red is actual CIWS VIL at 3.5 kg m^{-2} or greater.

6. Conclusions

Although examining forecasts by their definition and computing statistics in a common framework is beneficial, comparisons are difficult to make when forecasts of different types are under consideration. In the end, all forecasts being considered for use in a particular operational setting must be evaluated using appropriate transformations to understand their relevance in the user-specific decision making process. Translations should be capable of resolving skill at various scales and meteorological impact thresholds to relate to the decision maker. The FSS, FCI, and conversion to CCFP-like scales are useful translations to employ for this purpose.

7. References

Ghirardelli, J., 2005: An overview of the redeveloped Localized Area MOS Program (LAMP) for short-range forecasting. *Preprints, 21st Conference on Weather Analysis and Forecasting/17th Conference on Numerical Weather Prediction, Washington, D.C*.

Krozel, J., Penny, S., Prete, J., and Mitchell, J.S.B., 2004: Comparison of Algorithms for Synthesizing Weather Avoidance Routes in Transition Airspace, *AIAA Guidance, Navigation, and Control Conf., Providence, RI.*

Lack, S.A., G.J. Layne, M.P. Kay, S. Madine, and J. Mahoney, 2010a: Evaluating supplemental relationships between convective forecast products for aviation. 14th Conference on Aviation,

Range, and Aerospace Meteorology (ARAM). Atlanta, GA.

Lack, S. A., G. L. Limpert, and N. I. Fox, 2010b: An object-oriented multi-scale verification scheme. *Weather and Forecasting*, **25**, 79-92.

Layne, G.J. and S.A. Lack, 2010: Methods for estimating air traffic capacity reductions due to convective weather for verification. 14th Conference on Aviation, Range, and Aerospace Meteorology (ARAM). Atlanta, GA.

Madine, S., S.A. Lack, G.J. Layne, and J.L. Mahoney, 2011: Quality Assessment of CoSPA. 42 pp. Prepared for the FAA AWRP.

Roberts and Lean, 2008: Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. *Mon. Wea. Rev.*, **136**, 78–97.

Weygandt, S., T. Smirnova, C. Alexander, S. Benjamin, K. Brundage, B. Jamison, and S. Sahm, 2010: The High Resolution Rapid Refresh (HRRR): Recent enhancements and evaluation during the 2009 convective season. *Preprints, 14th Conference on Aviation, Range and Aerospace Meteorology (ARAM), Atlanta, GA*.

Wolfson, M. M., Dupree, W. J., Rasmussen, R., Steiner, M., Benjamin, S., Weygandt, S., 2008: Consolidated Storm Prediction for Aviation (CoSPA)**,** *Preprints, 13th conference on Aviation, Range, and Aerospace Meteorology (ARAM).*

Acknowledgements

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