2.6 Evaluating supplemental relationships between convective forecast products for aviation

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

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

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

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

1. Introduction

Current air traffic operations involving convective weather require making strategic planning decisions (2-h to 6-h lead times) at 11, 13, and 15 UTC planning points. Air traffic operations use weather information at these planning points to issue airspace flow programs (AFP), ground delay programs (GDP), and other route advisories. As the national airspace (NAS) becomes increasingly complex, additional forecasts that attempt to provide more structural information are being evaluated to supplement the current operational baseline, the Collaborative Convective Forecast Product (CCFP). Forecasts that provide deterministic structure (e.g. simulated radar reflectivity) and probabilistic information are potential supplements to CCFP. The Forecast Verification Section (FVS) within NOAA/ESRL serves as an independent assessment team for the evaluation of weather products in this framework (including but not limited to convective weather) for operational use within the Federal Aviation Administration (FAA).

Diagnostic studies should be performed to fully understand the strengths and weaknesses of the candidate supplement. These diagnostic studies are intended to give a baseline of skill and to provide potential areas where the product could add information to CCFP to enhance the planning process. These diagnostic studies include stratifying skill by geographic regions, issuance and lead time. Performance is also assessed at a variety of scales from the product's native resolution through CCFP-like resolutions.

Once the background diagnostic study of the product is complete, the supplemental study begins in the context of the candidate forecast's strengths and weaknesses. This provides some insight for the development of a potential concept of use (ConUse) for using the CCFP in conjunction with the supplemental forecast. Three major questions are addressed when evaluating the supplemental relationship to CCFP.

- Does the forecast provide finer-scale structural information inside of a CCFP polygon?
- Does the forecast provide confidence to the planner when used in conjunction with the CCFP polygon?
- Does the product have skill beyond the 6-h lead provided by CCFP?

Evaluating a product in a supplemental fashion is primarily accomplished through the use of a four-quadrant joint domain decomposition analysis between the operational product (CCFP) and the candidate supplemental forecast. Two of the quadrants indicate forecast agreement (both forecasts indicate the occurrence of an event and both forecasts indicate the absence of an event) and two of the

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

quadrants represent disagreement (one forecast indicates an event and the other indicates the absence of the event). The quantification of these quadrants and the methodology employed vary depending on the type of forecast product (deterministic vs. probabilistic for example) and the desired questions being answered.

This paper will detail some of the techniques used for diagnostic studies of forecasts in section 2. Section 3 will detail how the 3 major questions for supplemental relationships are evaluated. Section 4 will include some future directions of evaluations.

2. Diagnostic Study

The diagnostic study performed is focused on understanding the baseline skill of a product and comparing that skill to that of the current operational standard CCFP. Diagnostics include determining and understanding potential systematic strengths and weaknesses of forecasts considered to supplement CCFP or eventually replace CCFP in the decision-making process. Basic verification practices are normally accomplished during this stage such as examining the bias of the product as a function of lead time and geographic domain. Computing standard skill scores such as the critical success index (CSI), probability of detection (POD), and false alarm ratio (FAR) as a function of resolution are also useful by tying back to producer-derived statistics to provide a common ground of comparison (Wilks, 1995). In the instance for a probabilistic forecast, reliability is often computed with various stratifications. Relating these basic skill scores to CCFP skill can highlight some potential issues in using the supplemental forecast in instances of lesser skill by CCFP. Traditionally, during the evaluation of convective forecasts the National Convective Weather Diagnostic (NCWD) field is used as the observation field (Megenhardt et al. 2004). NCWD is considered to be hazardous for aviation at VIP-level 3 and above. Although NCWD is the primary observation throughout many of the studies performed herein, other observations, such as the Corridor Integrated Weather System (CIWS) analysis field (Evans and Ducot, 2006), also are heavily used for consistency in all evaluations. This section will detail some of the techniques used.

Examining climatological temporal smears of observations and comparing the selected forecast at a specific valid time for a set of lead times can provide insight into systematic flaws in the domain of a specific forecast. These smears are generated over a specified period of study such as a convective season. The smears are created by averaging the occurrence of convection (observed or forecasted) at each grid box for the set of days used in the study. A Gaussian smoothing operator is then applied to the domain at a scale which can be much larger than an individual grid box which allows for systematic signals on larger scales to be seen effectively. An example of a climatological smear for hazardous convection compared to CCFP polygon issuance for July and August 2009 is shown in Figure 1. These smears are useful coarse-scale diagnostics to see if a forecast is giving you first-order skill at forecast correct locations and structures with corresponding relative frequencies.

Although smears may provide insightful qualitative views of coarse-scale forecast performance, it is necessary to quantify this skill. Skill of the candidate supplemental forecast should be on par with the skill a CCFP polygon provides at the scale of typical CCFP polygons. This requires an appropriate scaling technique to transform a candidate probabilistic or dichotomous forecast into CCFP-like information.

Probabilistic forecasts are generally the easier of the two forecast types to handle. To get information on CCFP-like resolution for a probabilistic forecast one just has to find the probability threshold that yields similar size to that of a CCFP polygon. This

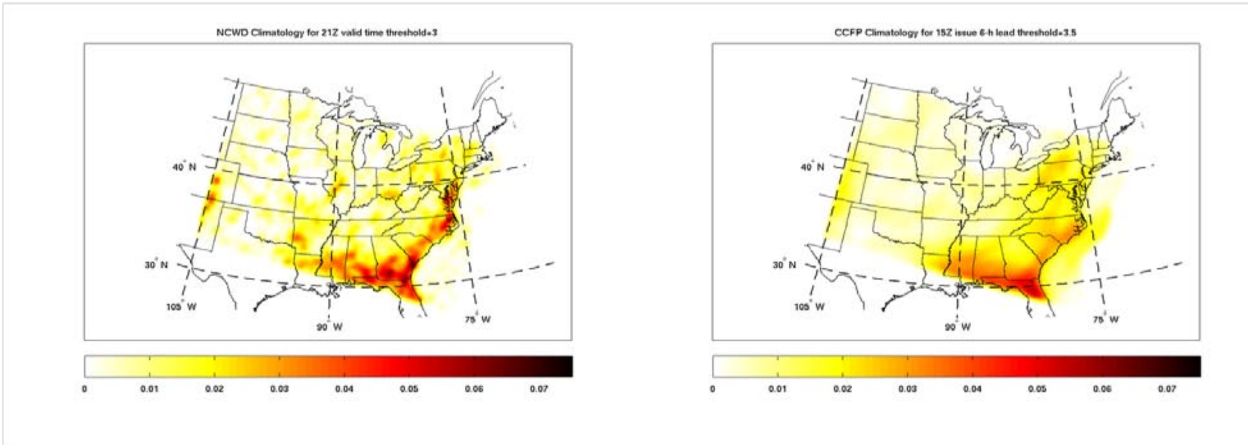

Figure 1. Hazardous NCWD climatology for July and August 2009 valid at 21Z compared with CCFP polygon climatology issued at 15Z with a 6-h lead.

is achieved by bias correcting the forecast using CCFP as the target. These thresholds can be issue and lead-time dependent and vary seasonally.

Changing a deterministic observation or forecast field into CCFP-like scales requires the use of a clustering methodology. Two passes are used in scaling the observation or forecast field plus an additional smoothing operation. The first pass clusters observations that obtain observed coverage based on historical distributions of actual CCFP coverage inside a sparse coverage (Kay et al. 2007), low confidence polygon while exceeding the minimum size criterion for CCFP (3000 sq mi). The second pass clusters higher density observations consistent with CCFP nested medium or high coverage polygons. The smoothing is accomplished by a series of erosion and dilation operations. An example clustering of the hazardous NCWD field is shown in Figure 2.

From these scaled images, skill scores (such as CSI and bias) can be calculated for the low coverage regions which would match to the scale of CCFP polygon issuance in the domain. This provides a useful baseline diagnostic comparison between the operational standard and the new forecast product being considered for traffic flow management (TFM).

From a traffic management perspective, knowledge of air traffic density plays a key role in determining when to go forward with using a product in operations. Geographical stratifications need to be made in the best interest of TFM. For instance, the busiest hubs are located in the northeast US, and therefore forecast products must perform well in this region. Air traffic flow is sensitive to the structure of convection in the high traffic NE and therefore a candidate forecast must exhibit skill in providing structural information. This requires a forecast to have some degree of sharpness; the forecast must go out on a limb to forecast extreme events. When CCFP does not contain nested or medium coverage polygons the forecast is not considered sharp; this is typical of CCFP before the 15Z planning telecon. For example, a sharp forecast will go out on a limb and forecast extreme events; ideally at the same frequency at which the extreme events occurs in observations on similar temporal and spatial scales. A forecast based on climatology, although reliable over long time periods, does not exhibit sharpness as it forecasts smooth probabilities with little variation from day to day. Understanding the underlying model to the particular forecast of interest is a good first step in discovering sharpness issues. Sharpness may be measured in terms of ARTCC or sector permeability described by Layne and Lack (2010).

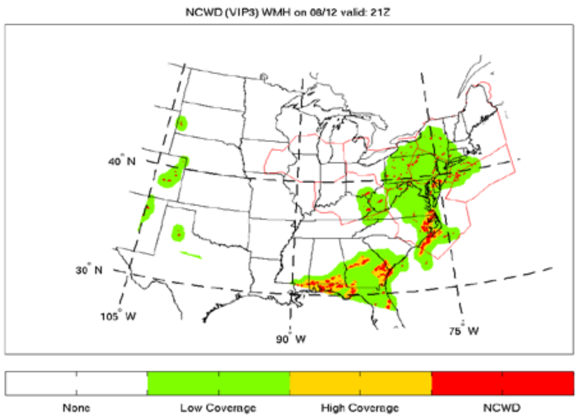

Figure 2. An example of a scaled NCWD field valid at 21Z on 12 August 2009, green represents lower coverage, yellow high coverage, and red is actual NCWD VIP-level 3 and greater.

Important features of a candidate forecast are discovered by running some simple diagnostic tests including the climatological smears, skill at CCFP-like scales, skill at different resolutions, standard bias trends as a function of lead time, and sharpness tests. The climatological smears may indicate systematic problems in certain geographical regions. Assessing skill at CCFP-like scales can indicate whether or not a forecast has similar qualities to CCFP at large scale, which is important when considering if a forecast agrees or disagrees with CCFP. In addition, examining a forecast at different thresholds may reveal if the product has skill but needs calibration. The methods above illustrate a sample of the approaches used to provide diagnostic feedback that could provide useful background information on how the product could perform as a successful supplement to CCFP.

3. Supplemental Study

This section will explore techniques outlining a forecast's usefulness in adding structure, confidence, or value beyond CCFP's maximum 6-h lead time.

Regardless if the forecast is deterministic or probabilistic, assessing the usefulness of the product beyond the 6-h lead can be accomplished by examining the skill using the scaling technique outlined in the previous section. It is important to look at the trend of the skill leading up to the 6-h lead time and beyond. If the forecast's skill at CCFP-like scales is giving comparable information at the 6-h lead to CCFP and does not decay significantly beyond this period the product may give the planner CCFP-like information beyond the 6-h lead. Information on finer structural skill beyond 6-h can be assessed using methods outlined below.

The methods for assessing structure and confidence are slightly different when working with probabilistic versus deterministic forecasts. As the methodology used for the candidate forecasts vary, slightly different answers are derived for the specific supplemental relationship questions which can be used as input to ConUse formulation. All of these methods use a joint domain decomposition approach to understand the forecast quality. First, the occurrence of agreement and disagreement are quantified. In these regions of agreement and disagreement several methods are used to determine a forecast's usefulness in a supplemental role. These approaches will be outlined in the following subsections.

3.1 Joint Domain Decomposition

The four quadrants that make up the joint domain distribution include the agreement of the forecasts of an event, the agreement of a non-event, and the two quadrants of forecast disagreement. The common forecast domain is decomposed into the four quadrants making up a contingency table, an example of domain decomposition is shown in Figure 3. A clustering algorithm is used to find general regions of forecasting agreement and disagreement. The example shows a supplemental deterministic forecast (blue), along with CCFP polygons (yellow), and

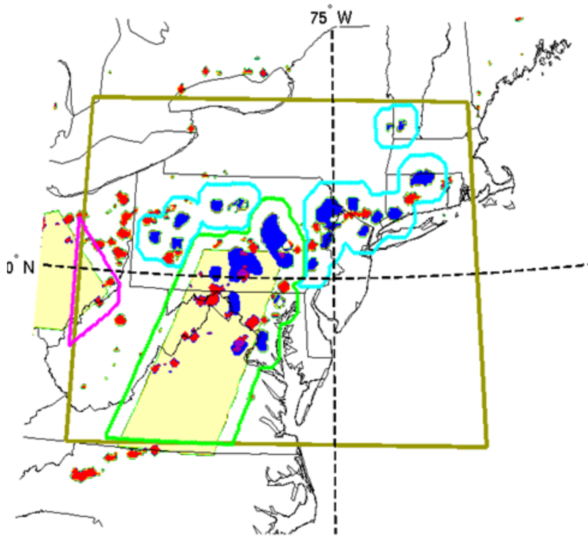

Figure 3. Example of domain decomposition for use in the evaluation. CCFP is shown in yellow, a supplemental forecast is shown in blue and observations are shown in red. Forecast agreement of convection is outlined in green, disagreement in magenta and cyan. The rest of the domain (brown outline) is considered agreement of no forecasted convection.

corresponding NCWD truth field (red). Forecast agreement is shown by the regions outlined in green, while existence of CCFP with no supplemental forecast is shown in magenta, and the supplemental forecast alone is shown outlined in cyan. The rest of the domain is considered agreement of no forecast of significant convection.

The quadrants are aggregated over the season and summarized by the average areal percent of agreement and the average frequency of occurrence. This evaluation allows the user to understand how often the primary CCFP forecast and the supplemental forecast agree and disagree. In regions of forecast agreement the supplemental forecast should add structure or confidence to the CCFP polygon while in regions of disagreement the supplemental forecast should add skill by resembling actual nearby observed convection.

Within the decomposed domains of interest several approaches can be leveraged to determine supplemental forecast skill as it pertains to the primary forecast of interest. These methods vary depending on the type of supplemental forecast presented, deterministic versus probabilistic. Probabilistic forecasts may be evaluated by looking at reliability, resolution, and sharpness within the sub-regions. In the case of deterministic forecasts, object-oriented approaches may be used to evaluate the structure of the convection in the sub-regions. Example methodologies of forecast evaluation for the two types of forecasts will be explored in the following sub-sections.

3.2 Probabilistic Approach

When dealing with a probabilistic forecast, the probability field must add confidence and/or structure to the CCFP polygon by issuing hotter than normal probabilities in regions where convection is likely to occur. Confidence is measured by examining the coverage of convection inside a CCFP polygon compared with the mean, median, or maximum probabilities given by the supplemental forecast's probability field. If the trend of higher convective coverage in the observation field (NCWD) occurs inside a CCFP polygon with higher probabilities in the supplemental forecast field the supplement is said to add confidence to the planning process. This analysis can be repeated for any CCFP polygon type for further stratifications. It can also be repeated outside of CCFP polygons to see what value is added when the probability field is disjoint with respect to the CCFP polygons.

A probabilistic field adds structure to a CCFP polygon if the probabilistic field is sharp enough to provide finer scale structure inside the given polygon. This is evaluated by trimming the CCFP polygon by increasing probability bins given by the supplemental forecast product. The desired goal is to have higher probabilities contain more observed convective coverage than lower probabilities within the CCFP polygon. This acts to discriminate regions where convection is likely to occur within the broader CCFP polygon. An illustration of a probabilistic forecast that adds structure is shown in Figure 4, where the maroon objects are hazardous convection, the orange outline is the CCFP polygon, and the probabilities range from green to yellow. In Figure 4, it is clear that more convection is occurring in the 30% probability bin adding structural information to the southern extent of the CCFP polygon.

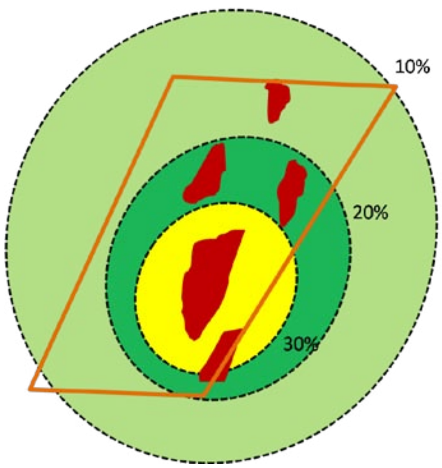

Figure 4. An example of a probabilistic field (green to yellow) adding structural information to a CCFP polygon (orange) with hazardous convection (maroon) overlaid.

3.3 Deterministic Approaches

In terms of confidence, the probabilistic method described in the previous section can be applied to a deterministic supplemental forecast. In this case, convective coverage for both the observation and supplemental forecast can be calculated when confined within any selected stratification of CCFP polygon type. If the trend of observed convection to forecasted convection in the supplemental product is similar, the forecast adds confidence to the CCFP polygon. Although such a confidence study is important, the structural information added from a supplemental deterministic forecast is more intriguing. Structural supplements from a deterministic forecast can be measured in several ways with each method answering slightly different questions.

One method for assessing the additional structural information provided by the deterministic forecast combined with a measure of confidence is by utilizing the scaling approach shown in Figure 2. The yellow regions in Figure 2 represent dense NCWD coverage (or forecast coverage); these regions may cause more problems for aviation than the broader green regions (CCFP-like scale convection). Taking the yellow regions from the deterministic supplemental forecast and clipping them to CCFP polygons can give a broader sense of confidence combined with finer scale structure, especially when stratifications based on CCFP polygon type are made. It is known that operational planners tend to weigh information based on sparse coverage, low confidence CCFP polygons lower than high confidence CCFP polygons. However, if a deterministic forecast indicates areas of dense coverage in these polygons it may add structural information and confidence to the planner that the CCFP polygon warrants attention if this trend is verified over a season.

Additional structural information can be assessed by examining the distribution of objects observed and forecasted within a CCFP polygon including size, shape, and orientation with or without applying the scaling (i.e. yellow areas). This allows for the comparison of observed convective mode and forecasted convective mode. If the deterministic forecast and observation agree on convective mode within a CCFP polygon even if the objects are not collocated this provides significant structural information to the planner in terms of potential permeability of the airspace. More precise measures of skill can be addressed by using object-oriented verification practices. The Procrustes verification approach (Lack et al. 2010) has been used to assess attributes such as displacement of convective objects.

4. Future directions

Near future effort for convective forecast verification will include echo tops both as a forecast input and observation. This hinges on the understanding of the concept of use for the echo top field in the strategic planning process from the air traffic decision point of view. The en route flow problem is highly sensitive to accurate echo top information and clearly needs to be made a focus in a future studies.

Although temporal resolution has been addressed in various studies performed by the Forecast Verification Section, more effort is underway to communicate results of added temporal resolution to the planning process. One aspect of increased temporal resolution is the ability to determine onset and cessation of significant convective events more accurately than the 2-h intervals associated with CCFP. A product with higher temporal resolution should add more reliable information than the 2-h lead time interval CCFP. Additionally a product that updates more frequently than CCFP must have some level of consistency. A forecast that changes its story from one issue to the next can only add confusion to a planner. These additional temporal resolution studies will be prominent in future evaluations when applicable.

5. References

Evans, J. E., Ducot, E. R., Corridor Integrated Weather System, MIT Lincoln Laboratory Journal, Volume 16, Number 1, 2006.

Kay, M.P., S. Madine, J.L. Mahoney, and J.E. Hart, 2007: 2007 Convective Forecast Scientific Evaluation. Prepared for the FAA Systems Operations Group.

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

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

Megenhardt, D. L., C. Mueller, S. Trier, D. Ahijevych, and N. Rehak, 2004: NCWF-2 Probabilistic Forecasts. *Preprints, 11th AMS Conference on Aviation, Range, and Aerospace Meteorology*.

Wilks, 1995: Statistical Methods in the Atmospheric Sciences, Academic Press.

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.