AN OPERATIONAL PERSPECTIVE FOR EVALUATING CONVECTIVE NOWCASTS 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

ABSTRACT

As the United States is moving toward automated air traffic decision tools (NextGen), evaluating forecasts for potential operational use is key. Air traffic management, current or future, needs convective nowcasts (0-6 h) for information on how to best route traffic between aviation centers, sectors, and jetways. Current and future operational air traffic management needs focus on how to best supplement (or replace) the current operation standard, the coarse Collaborative Convective Forecast Product (CCFP). Finer scale products (i.e. simulated radar reflectivity from models) have to be evaluated for additional information, namely value added by increasing structural information and increasing temporal resolution. Structure can be quantified by examining bias behavior of this fine scale forecast or by providing information on the convective objects or clusters of objects within, and outside of, the broad-scale operational standard forecast. Additional structural information comes from evaluating the porosity of sectors when overlaid with convective objects for the assessment of potential reductions in air traffic capacity. Value added by increasing the temporal resolution of forecasts can be evaluated by a planning point evaluation of all forecasts covering a valid range of time. In this evaluation, all lead times from a forecast are assessed from forecast initial time until a point beyond the valid time of the final lead using all observations present in that valid range. A common thread to all of these evaluations is being able to stratify all days from a study period into different degrees of air traffic impact which is made possible from a normalized air traffic impact score. Other important stratifications involve the delineation of significant and non-significant convection. This note will give some detail to the approaches outlined above.

1. INTRODUCTION

The aviation forecasting community, when evaluating their own products, often find ways to increase standard categorical skill scores (CSI, Bias, POD, etc.) when presenting results. In addition, they may find ways to tune their models to score the best for specific high impact events while allowing lower impact events to score poorly. In the operational realm, forecasts must be reliable for more than a few select cases and exhibit skill beyond information obtained from categorical skill scores. As the United States is moving toward automated air traffic decision tools (Next Generation Air Transportation System—NextGen), evaluating forecasts for potential operational use is of primary importance. Air traffic management, current or future, needs reliable and accurate convective forecast information from 0 to 8 h to be used to route air traffic between convective weather and through aviation centers, sectors, and jetways.

The Federal Aviation Administration (FAA) has two main branches interested in weather and air traffic management (TFM). The FAA Systems Operations Group is the body that governs current air traffic management for the National Airspace System (NAS), while the FAA Aviation Weather Research Program (AWRP) is concerned with supporting the weather needs of the future Next Generation Air Transportation System (NextGen). Therefore, utilization of convective weather information by the two groups differs, thus affecting how the forecasts are assessed for quality. This use of the convective weather information is expected to converge as NextGen's initial operating capability (IOC) date approaches (2013). In order to achieve this goal, blending of new forecast products with current NAS management will become a necessity.

This paper outlines approaches used by the Forecast Verification Section (FVS) within NOAA/ESRL/GSD which serves as an independent evaluator of convective forecast products for both FAA AWRP and FAA Systems Operations. Section 2 will highlight some of the basic verification mechanics and metrics that are used for all of our convective evaluations. Section 3 will highlight some of the specific concerns for the current state of air traffic management, while Section 4 will highlight newer techniques to evaluate NextGen applications of convective weather forecasts. Some future efforts will also be outlined.

2. VERIFICATION MECHANICS

Currently, the standard operational aviation products often drive the construction of the evaluation mechanics. For example, the Collaborative Convective Forecast Product (CCFP) issued by the Aviation Weather Center (AWC) is considered the current operational convective forecast standard. The

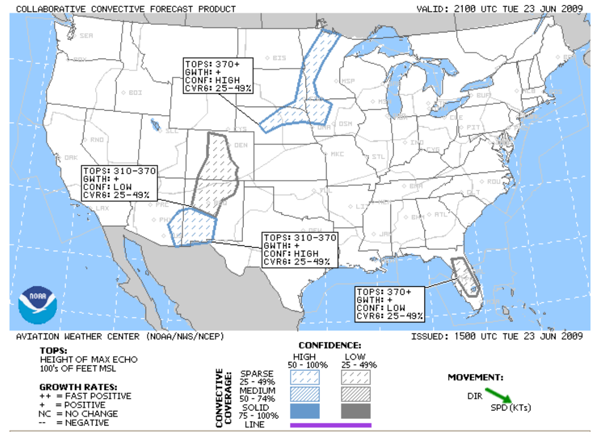

Figure 1. Example CCFP forecast from 23 June 2009 issued at 1500 UTC valid at 2100 UTC.

evaluation of newer forecast products must exhibit value in some way relative to the CCFP. An example CCFP forecast is shown in Figure 1. CCFP forecasts are categorical forecast polygons with attributes pertaining to hazardous convective weather coverage (defined by the FAA to be radar echoes of Video Integrator and Processor (VIP) level 3 or greater; ~40 dBZ), forecast confidence, growth of the hazardous weather, and potential maximum echo top values.

2.1 VERIFICATION STRATIFICATIONS

The definition of CCFP drives the basic verification mechanics and data stratifications used to evaluate both CCFP and alternative convective forecasts. Stratifications are then separated into two major categories during the evaluation: convective intensity and forecast resolution. Additionally, situational stratifications are implemented based on the estimated impact of convection on the flow and management of air traffic during the day.

The convective intensity of primary concern is easily adapted to the definition of hazardous convection set by the FAA (see FAA Advisory Circular and FAA Aeronautical Information Manual). Traditionally, the National Convective Weather Diagnostic (NCWD) has been used as the observed verification field with a VIP

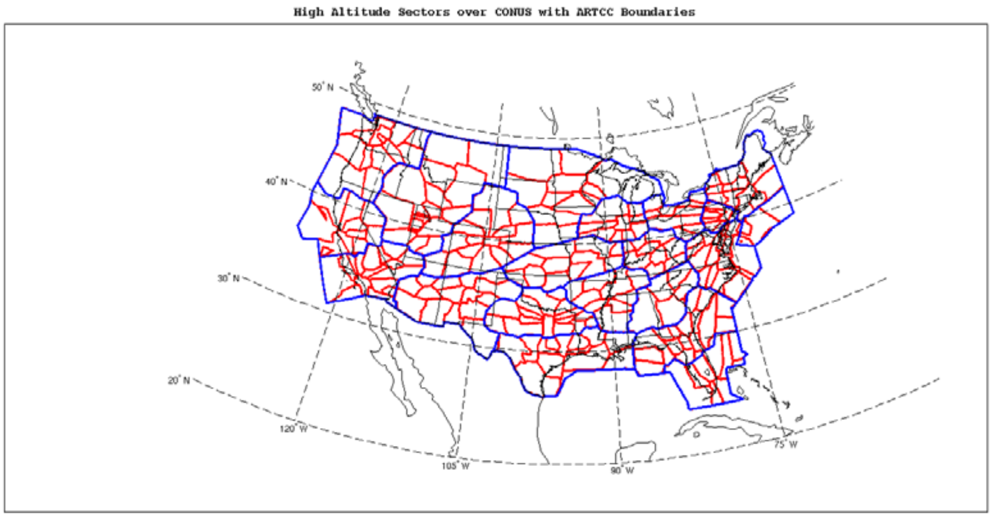

Figure 2. En route (ultra high altitude) sectors (red) over CONUS with ARTCC boundaries blue).

level 3 threshold applied (hazardous convection). NCWD (described by Megenhardt et al. 2004) considers radar vertically integrated liquid (VIL) from the WSR-88D (NEXRAD) Level III data and lightning data from a variety of sources displayed on a 4-km grid. Radar echoes below 15 kft are filtered out of this product. Other products such as composite reflectivity are also considered as candidates for the observation field in some studies. While VIP level 3 is clearly of concern to operational traffic management, it is of use to examine lower (and higher) intensities. Lower intensity evaluations allow for important feedback to the convective forecast developers on issues such as calibration with respect to intensity levels.

Resolution stratifications are also useful in diagnosing skill of convective forecasts for potential use in air traffic planning. Altering the resolution allows examination of the forecast scale most relevant to FAA operations. As CCFP does not have a defined resolution, convective coverage on high altitude air traffic sectors provides a useful resolution stratification that is directly applicable to FAA operations. This also provides a consistent framework for evaluating CCFP and other convective forecasts in the probabilistic realm (ensemble) or the fine-scale deterministic realm (simulated radar reflectivity). An en route sector grid can be seen in Figure 2.

Selecting days on which to evaluate the performance of the collection of convective forecasts aids in the understanding of convective performance for operations. A normalized air traffic index which was evolved from Callaham et al. (2001: similar to Klein et al., 2008) is computed daily for the entire convective season based on NCWD coverage, position, and historical scheduled flight traffic. This metric allows for the discrimination between high and low impact weather days based on air traffic and convective weather coverage. Historically, organized convection along a strong cold front situated to impact the NE US has the highest aviation impact. Combining this traffic index with convective coverage across the CONUS can lead to further stratifications such as days where a lot of convection was present but did not impact the NE US. A collection of scores

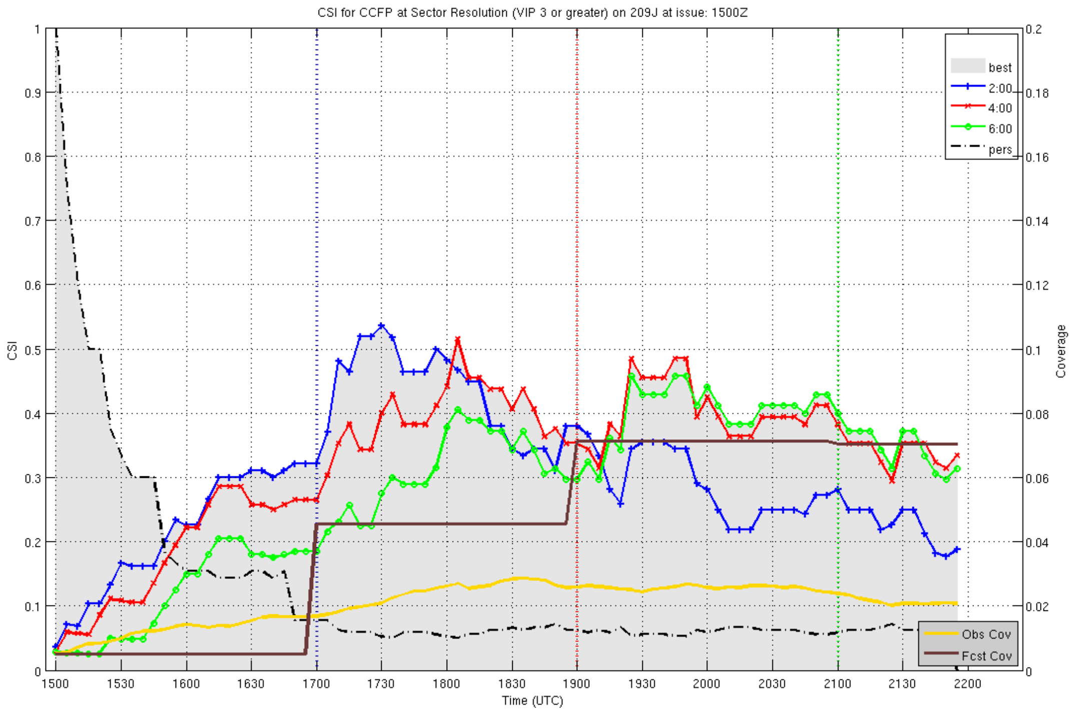

Figure 3. Planning point time-series plot of sector-based CSI for CCFP from 27 July 2008 issued at 1500Z. All CCFP lead times are scored during the valid period, 2-h lead (blue), 4-h lead (red), 6-h lead (green). The gray background represents the best skill at any NCWD 5-min observation. Black dashed line is persistence using the first NCWD 5-min observation. The yellow line is the average NCWD sector coverage, and the brown line is the average forecast sector coverage.

can be calculated from the possible stratifications listed above, including: convective intensity, potential air traffic impact, and resolution. These range from standard categorical scores (POD, CSI) in the deterministic forecast realm, to continuous scores (Brier, RMSE) in the probabilistic realm. Additionally, dichotomous and continuous scores can be calculated when forecasts are converted into sector coverage. For a detailed description of meteorological skill scores see Wilks (1995).

2.2 DISPLAYING RESULTS

From the statistics gathered from the many stratifications listed above, new displays have been developed for the effective communication of the results. One such display is the planning point time-series plot. The idea behind this plot is to show the quality of all forecast lead times for one issue time against every observation from the issuance time until an hour or two beyond the last forecast lead times valid period for any skill score. This allows the end-user to view the skill of the product in an operational framework for strategic planning or automated planning. A sample planning point time-series plot for CCFP is shown in Figure 3. The skill score of interest in Figure 3 is CSI based on en route sector coverage with 5% as the threshold for an impacted sectors (Kay et al., 2007). NCWD at VIP 3 or greater is used as the observation. It can be inferred using this skill score from one issue time that CCFP's 2-h lead is the best forecast available beginning at 1600 UTC (persistence is best

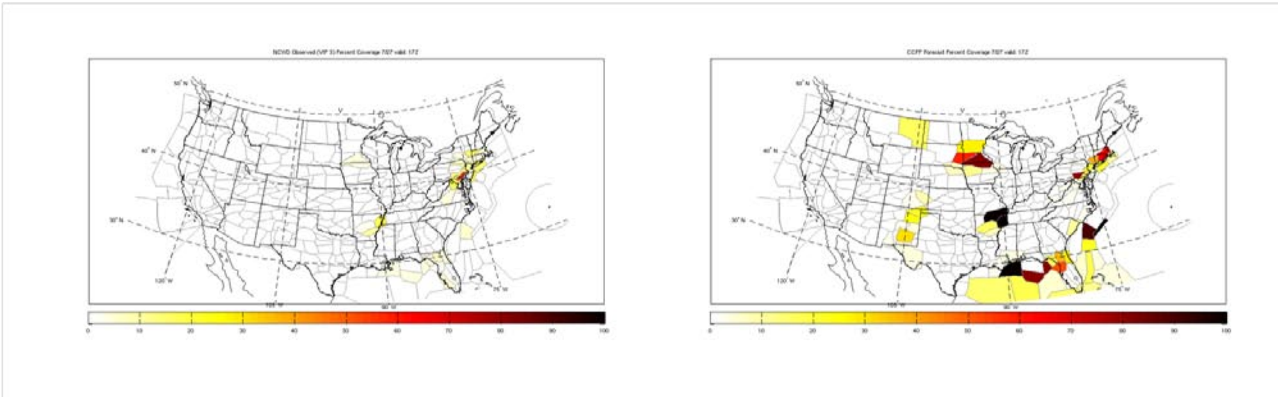

Figure 4. Example of NCWD sector percent coverage (left) versus CCFP sector percent coverage (right) valid at 1700 UTC on 27 July 2008. Warmer colors indicate higher percent coverage.

for the first hour) until the 4-h lead becomes the best around 1800 UTC. Both the 4-h lead and 6-h lead have comparable CSI scores of around 0.4 beyond 1930 UTC. These techniques help diagnose forecasting problems at specific ATM planning points. For example, this shows the shortcoming in CCFP in forecasting from the 0-1 hour time frame. Persistence is the best forecast out to approximately an hour out. Although this was a sample from one day and one issuance the scores may be aggregated over an entire convective season with confidence intervals applied.

A further diagnostic visualization of skill, especially in the case of CCFP, is the depiction of sector coverage over the CONUS (Figure 4). This can give the end-user some depiction of structure over the NAS in the observation (NCWD) field and the forecast field at a given forecast valid time. Both of these sample diagnostic plots of skill will be available in the Network Enabled Verification Service (NEVS) provided by NOAA (Madine et al., 2009).

3. CURRENT TFM OPERATIONS

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 NAS becomes increasingly complex, additional forecasts that may provide more structural information are being evaluated to supplement the current operational baseline (CCFP). Forecasts that provide deterministic structure (simulated radar reflectivity) and probabilistic information are potential supplements to CCFP. One such forecast being evaluated is the Localized Aviation MOS Program (LAMP) thunderstorm probability field (Charba and Liang, 2005).

Evaluating a product in a supplemental fashion is primarily accomplished through the use of a four-quadrant joint probability distribution (JPD) analysis between the operational product and the candidate supplemental forecast (Table 1). 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 quadrants represent disagreement (one forecast indicates an event and the other indicates the absence of the event).

The JPD approach allows for both the validation and creation of a concept of use when decisions are made based on more than one forecast. The JPD indicates how often the forecasts agree and disagree over the length of the study period (normally a convective season). Further, in each of the JPD quadrants, skill can be characterized based on the observation of interest. For

Forecast Region Present	CCFP Yes	CCFP No
Supplement Yes	Agreement of event(char. skill when products agree on event)	Supplement indicates event when CCFP does not (char. skill of supp.)
Supplement No	CCFP indicates an event (char. skill of CCFP when no supp. forecast is present)	Agreement of no event (char. the observation field if present)

Table 1. The schematic of a JPD for CCFP and a supplemental product; agreement regions are highlighted in green, disagreement in yellow.

example, it may be that the forecasts agree the majority of time during the convective season, however, during the small percentage of time that they disagree one should trust the supplement product for TFM decisions.

It is generally understood that CCFP does not contain structural information on fine scales. The forecast parameters of CCFP only indicate broad coverage ranges, which typically have a high bias. For example, a sparse CCFP polygon indicates convective coverage amounts between 25-49%, rather than indicating an organized pattern of convective coverage, which is more useful to operational decision-makers.

When evaluating a forecast for supplemental usage with CCFP, the secondary forecast should provide more information to the decision-maker than CCFP alone. This is mainly accomplished by examining the structure added by the supplemental product to CCFP when the forecasts are in agreement. Structure of the supplemental forecast is measured using a few techniques. If the supplemental product is a simulated reflectivity-like product, the structure can be characterized by comparing distributions of forecast convective objects and observed convective objects. This gives a good approximation of how well the supplemental forecast identifies the type of convection within a CCFP polygon. Standard skill scores can also be used as a proxy for the skill of the location of the convection in the deterministic realm. If the supplemental forecast is probabilistic, a reliability approach can be used to measure structure. When the supplemental probabilistic forecast agrees with the CCFP polygon, the reliability of the supplement in the specific CCFP polygon type can be calculated based upon the observation occurrence. For example, if high probabilities of the supplemental forecast are collocated within a CCFP polygon one would expect most of the convection to occur in this highlighted area. The reliability and distributions of probabilities within the CCFP polygons can yield whether or not the probabilistic forecast is truly adding benefit to the CCFP forecast.

4. NEXTGEN APPLICATIONS

In the future, NextGen applications will involve making real-time routing decisions based on convective weather in an automated framework. The candidate forecasts tend to be of high temporal and spatial resolution. The NextGen framework also assumes that air traffic management will move away from a network of set jetways and move to flexible jetways or even free flight. As the future of airspace usage is relatively unknown, verifying potential forecasts for usage in the context of the NextGen framework is difficult.

The underlying theme of evaluating forecasts for air traffic flow planning is the reliability of the forecast in identifying convective structure, location, and intensity. A solid linear convective system is significantly challenging for air traffic management, whereas a linear convective system with gaps between convective cores is slightly more manageable. In our current sector-based framework, skill has been assessed by examining convective coverage of the observation and the forecast within a high altitude sector. (Note: A convective coverage of 5% within a sector is a general marker for when the flow of air traffic begins to become constrained within at sector.) In some cases this may lead to a misrepresentation of skill when looking for information on the accuracy of convective structure. For instance, a sector may have its observation coverage exceeding 5% when the forecast coverage is below 5%; however, the observation field may contain many isolated cells whereas the forecast is an organized line of convection perpendicular to main jetways in the sector. In the current dichotomous framework, this would yield a miss although the sector from the observation perspective may only have a slight reduction in capacity whereas the forecast would make the sector impassable. It is therefore necessary to develop a toolset to measure the accuracy of convective structure for NextGen applications.

Two methods for estimating sector capacity reduction due to the presence and orientation of convection have been developed, Euclidean Distance and Mincut Bottleneck. Both methods utilize the identification of significant convection overlaid onto a sector and the estimation of the direction of traffic flow by bounding the sector to an ellipse to identify the length and orientation of the major axis.

The Euclidean Distance approach estimates capacity by calculating the distance to the nearest non-zero valued pixel of convection and the edge of the sector. A buffer is then applied as an avoidance zone from the sector's boundary and an avoidance zone from convection. The area, major axis length and orientation, as well as the maximum distance found in the buffered Euclidean distance field is then calculated and compared to that of the sector with no convection present. If the major axis of the Euclidean distance field does not approximate the length of the sector with no convection present, there are no open lanes to get air traffic to one end of the sector to the other and the sector is said to be completely closed. If the field does meet the length of the sector with no convection present, a ratio of the areas of the buffered regions and the maximum Euclidean distances are used as estimates of sector capacity.

The Mincut Bottleneck approach estimates capacity by calculating the minimum distance across the sector from a source and sink node (perpendicular to the major axis of the sector) using convective objects as nodes. The minimum distance found from the forecast and observation for the particular sector is then compared to the sector without convection to get its estimate of capacity. The convective objects may be dilated to estimate an air traffic avoidance field. The Mincut Bottleneck methodology for sector capacity reduction estimates comes from proposed air space management for NextGen applications (Krozel et al., 2004).

From the capacity reduction estimates calculated from both of the above estimates, dichotomous CSI scores can be made from quartiles of estimated reduction. Continuous scores can also be used as an aggregate measure of skill such as the Brier score and root mean squared error (RMSE). Graphical comparisons can also be made similar to those created for percent coverage (Figure 6). Figure 6 shows the comparison of the NCWD observation field capacity reduction estimate to a simulated reflectivity forecast valid at 1700 UTC on 27 July 2008, a relatively high impact day. The observation field shows considerable capacity reduction on sectors going in and out of the NE US, while the forecast field only hints at some capacity reduction in the NE.

The methods estimating sector capacity reduction currently use sector-based geometry. It is assumed that the representation of air traffic flow between major airports will remain close to current operations in the near future (IOC timeframe). This may be replaced with icosahedral (or similar) geometry as flow associated with NextGen is realized.

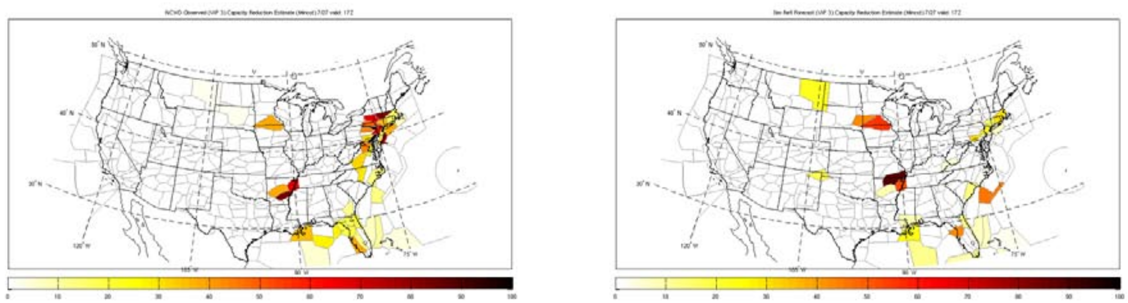

Figure 6. Sector capacity reduction estimates using the Mincut Bottleneck scheme for NCWD (left) and a simulated reflectivity forecast (right) on 27 July 2008 valid at 1700UTC. Warmer colors indicate a larger capacity reduction.

5. FUTURE WORK

The methods outlined throughout this paper will be adapted into the real-time verification framework, NEVS, being developed by NOAA/ESRL/GSD. From this verification service, an end-user (automated FAA decision tool or meteorological analyst) will be able to aggregate statistics over convective seasons and make useful stratifications as outlined earlier. Use of this tool will allow a more insightful analysis of the impact of convective weather on current and future air traffic management.

An effort currently ongoing is to incorporate convective echo tops, as convective forecast and analysis fields now contain such information. Differentiating between intense thunderstorms at altitudes that may affect en route traffic is yet another key to NextGen planning. For example, a low-topped mesoscale convective system (MCS) moving through the Midwest may be of low en route impact between the East and West Coasts of the U.S. but may cause limited terminal impact. This situation will end up having a lower operation cost than a severe weather outbreak in the same geographic region due to significant reroutes around higher topped convective storms. Being able to differentiate between the two weather features improve aviation traffic flow management decisions. In the near future, both the Euclidean Distance and Mincut Bottleneck will be extended to handle probabilistic forecasts. Both methods will also be compared with capacity measures from the Aircraft Situation Display (ASD) database for accuracy.

Work is ongoing in the creation of a meaningful suite of deterministic simulated radar reflectivity CCFP realizations based on historical observed convection within different size, shape, and attribute CCFP polygons. This will help answer the question of CCFP defining structure and may give a stochastic fine-scale solution as an alternative or supplement to high resolution model-based solutions.

6. REFERENCES

Callaham, M.B., J.S. DeArmon, A.M. Cooper, J.H. Goodfriend, D. Moch-Mooney, and G.H. Solomos: 2001, Assessing NAS Performance: Normalizing for the Effects of Weather. *4th USA/Europe Air Traffic Management R&D Symposium*, Santa Fe.

Charba, J.P., and F. Liang, 2005: Automated two hour thunderstorm guidance forecasts. *Preprints, Conference on Meteorological Applications of Lightning Data*. San Diego.

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.

Klein, A., R. Jehlen, S. Kavoussi, D. Hickman, D. Simenauer, M. Phaneuf, and T. MacPhail, 2008: Quantification of Predicted Impact of Weather on Air Traffic and Comparison with Actual Impact, *13th Conference on Aviation, Range and Aerospace Meteorology*.

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.

Madine, S., N. Matheson, M. Petty, D. Schaffer, and J. Mahoney 2009: The Network-Enabled Verification Service (NEVS): Providing verification of weather forecast products in NextGen. *Preprints* 25th Conference on International Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, Phoenix, AZ.

Megenhardt, D. L., C. Mueller, S. Trier, D. Ahijevych, and N. Rehak, 2004: NCWF-2 Probabilistic Forecasts. *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.