89th AMS Annual Meeting
ARAM Special Symposium on Weather - Air Traffic
Phoenix, AZ / 11-15 January 2009

P1.1

THE 2008 CoSPA FORECAST DEMONSTRATION*
(COLLABORATIVE STORM PREDICTION FOR AVIATION)•

W. Dupree†, D. Morse, M. Chan, X. Tao, H. Iskenderian, C. Reiche, and M. Wolfson
MIT Lincoln Laboratory, Lexington, MA

J. Pinto, J. K. Williams, D. Albo, S. Dettling, and M. Steiner
NCAR Research Applications Laboratory, Boulder, CO

S. Benjamin and S. Weygandt
NOAA ESRL Global Systems Division, Boulder, CO

1. INTRODUCTION

Air traffic congestion caused by convective weather in the US has become a serious national problem. Several studies have shown that there is a critical need for timely, reliable and high quality forecasts of precipitation and echo tops with forecast time horizons of up to 12 hours in order to predict airspace capacity (Robinson et al. 2008, Evans et al. 2006 and FAA REDAC Report 2007). Yet, there are currently several forecast systems available to strategic planners across the National Airspace System (NAS) that are not fully meeting Air Traffic Management (ATM) needs. Furthermore, the use of many forecasting systems increases the potential for conflicting information in the planning process, which can cause situational awareness problems between operational facilities.

One of the goals of the Next Generation Air Transportation System (NextGen) is to consolidate these redundant and sometimes conflicting forecast systems into a Single Authoritative Source (SAS) for aviation uses. The FAA initiated an effort to begin consolidating these systems in 2006, which led to the establishment of a collaboration between MIT Lincoln Laboratory (MIT LL), the National Center for Atmospheric Research (NCAR) Research Applications Laboratory (RAL), the NOAA Earth Systems Research Laboratory (ESRL) Global Systems Division (GSD) and NASA, called the Consolidated Storm Prediction for Aviation (CoSPA; Wolfson et al. 2008). The on-going collaboration is structured to leverage the expertise and technologies of each laboratory to build a CoSPA forecast capability that not only exceeds all current operational forecast capabilities and skill, but that provides enough resolution and skill to meet the demands of the envisioned NextGen decision support technology. The current CoSPA prototype for 0-6 hour forecasts is planned for operation as part of the NextGen Initial Operational Capability (IOC) in 2013. CoSPA is funded under the FAA's Aviation Weather Research Program (AWRP).

The first CoSPA research prototype demonstration was conducted during the summer of 2008. Technologies from the Corridor Integrated Weather System (CIWS; Evans and Ducot 2006), National Convective Weather Forecast (NCWF; Megenhardt et al. 2004), and NOAA's Rapid Update Cycle (RUC; Benjamin et al. 2004) and High Resolution Rapid Refresh (HRRR; Benjamin et al. 2009) models were consolidated along with new technologies into a single high-resolution forecast and display system.

Historically, forecasts based on heuristics and extrapolation have performed well in the 0-2 hour window, whereas forecasts based on Numerical Weather Prediction (NWP) models have shown better performance than heuristics past 3-4 hours (Figure 1). One of the goals of CoSPA is to optimally blend heuristics and NWP models into a unified set of aviation-specific storm forecast products with the best overall performance possible.

The CoSPA prototype demonstration began in July 2008 with 2-6 hr forecasts of Vertically-Integrated Liquid water (VIL) that seamlessly matched with the 0-2 hr VIL forecasts available in CIWS. These real-time forecasts have been made available to the research team and FAA

*This work was sponsored by the Federal Aviation Administration under Air Force Contract No. FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.
•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 or position of the FAA.
†*Corresponding author address*: William Dupree, MIT Lincoln Laboratory, 244 Wood Street, Lexington, MA 02420-9185; e-mail: dupree@ll.mit.edu

management only through a web-based interface[1]. This paper discusses the system infrastructure, the forecast display, the forecast technology and performance of the 2-6 hr VIL forecast.

Our early assessment based on the 2008 demonstration is that CoSPA is showing tremendous promise for greatly improving strategic storm forecasts for the NAS. Early user feedback during CoSPA briefings suggested that the 6 hr forecast time horizon **be extended to 8 hours** to better meet their planning functions, and that **forecasts of Echo Tops** must also be included.

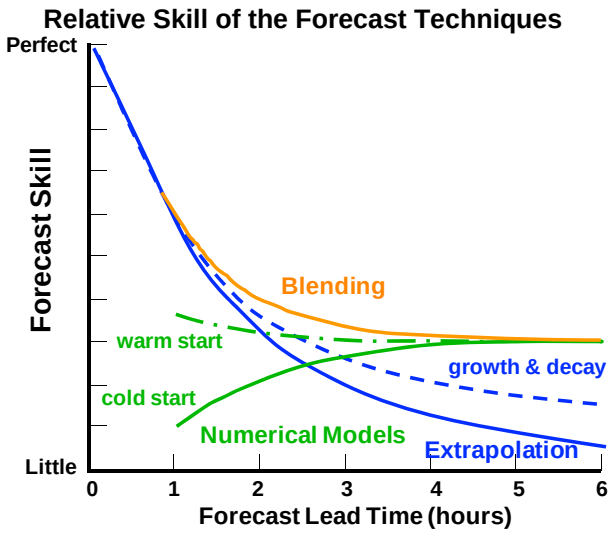

Figure 1: Depicted is a notional view of forecast skill as a function of lead time. Short-term heuristic forecasts based on tracking and trending tend to perform better in the 0-3 hour time range, however NWP forecasts show better performance at the longer lead times of 3-6 hours. We seek to optimize the performance by blending heuristic and NWP forecasts.

2. SYSTEM ARCHITECTURE

In line with the concepts of a virtual distributed system as envisioned by NextGen (NextGen ConOps, 2007), the CoSPA system was designed as a distributed set of processing nodes that are linked together by a network. The NextGen Network Enabled Weather (NNEW) working group is currently exploring a number of data formats and web services that could be used to exchange data across this distributed system. We are using the common gridded data format NetCDF4 for CoSPA as this format will likely be adopted by NNEW and NextGen System Wide Information Management (SWIM) system. However, use of some native data formats are still necessary until the NNEW technology matures. Additionally, CoSPA currently uses File Transfer Protocol (FTP) and Local Data Manager (LDM) servers as an interim solution for data transport until the NNEW web services have matured to a point that they are reliable enough to handle the immense data flow required by a prototype high-resolution operational real-time system.

The network and data flow diagrams are provided in Figures 2a and 2b. Many sources of sensor and meteorological data are ingested by MIT LL, NCAR and NOAA/GSD for the heuristic and NWP models. We are utilizing the CIWS 0-2 hr forecast data streams as a starting point for the 2-6 hr CoSPA forecasts. A set of CoSPA-specific modules that extend the heuristic forecast out to 6 hours (soon to be 8 hours) are built on top of the CIWS platform.

Once the MIT LL extrapolation and HRRR model data become available, they are ingested into the blending algorithm described in Section 4.3, which is run at NCAR. Upon completion, the blended forecast data are sent back to MIT LL where the data are post-processed for display on the CoSPA website.

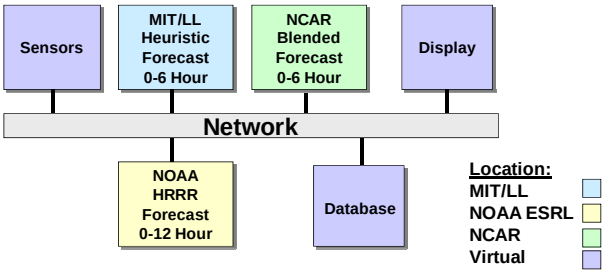

Figure 2a: Depicted is the general network architecture of the CoSPA system. Processing is distributed among the three laboratories and data is moved using LDM, and FTP in the common netCDF4 data format wherever possible. In the future these distributed nodes can be plugged directly into NNEW SOA architecture.

3. COSPA WEB DISPLAY

3.1 Website Overview

The CoSPA website was developed primarily as a tool for the researchers to monitor and evaluate the performance of the forecast. The website consists of three primary sections: 1) a Situational Display (SD); 2) a playback data archive; and 3) a forecast comparison and analysis tool.

[1] CoSPA Website: http://cospa.wx.ll.mit.edu

The CoSPA web situation display was built onto the CIWS web situation display. This has the advantage of a high degree of code reuse, cost reduction, and benefits from years of user feedback from demonstrations and operational use in CIWS. Additionally, once the forecast is ready for prototype operations, this web platform can be used for real-time application if desired.

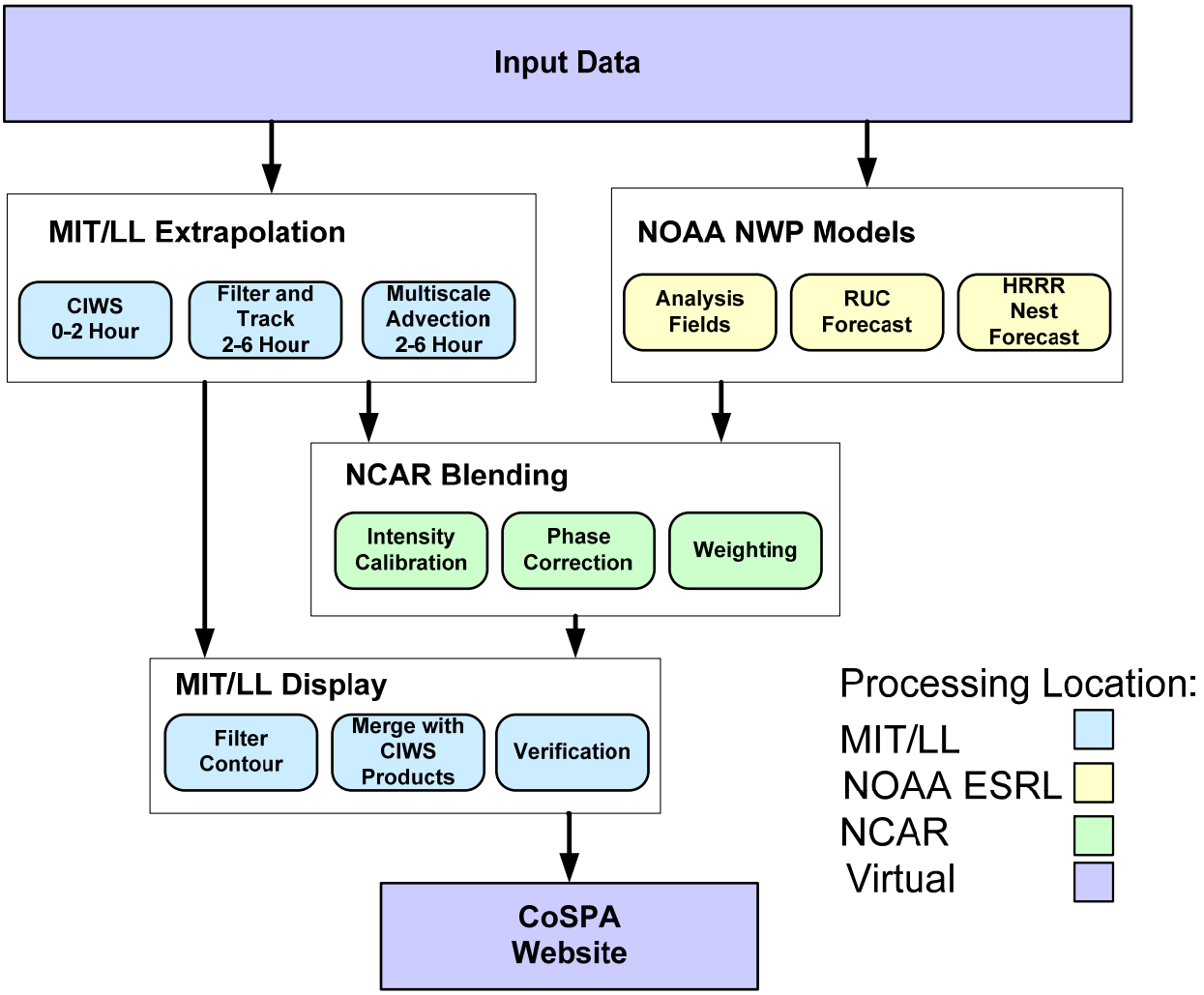

Figure 2b: Shown are the combined functions and dataflow for the CoSPA system. The arrows represent dataflow between processes that are either local or remote.

3.2 Web Situation Display and Products

The web SD has two primary modes. In the first mode it can be used to show the current weather situation and in the second mode it can provide forecasts of VIL via static forecast images, animation loops and contour overlays. VIL is displayed in 6 aviation-specific color levels sometimes referred to as VIP[2] levels (see Table 1 for equivalence to VIL in kg/m^2). In this paper we will refer to VIP levels as VIL levels.

[2] VIP stands for Video Integrated Processor and derives from an early radar weather display used in aviation.

The SD functionality is very robust. Agile panning and zooming are available via mouse control. An operator can zoom the SD to the terminal level with pixel resolution of 1 km (0.5nm) and continuously zoom out to full CONUS level with 7 km resolution (4 nm). The display can run in an animation loop mode with 6 hours of past weather that transitions into a forecast with 6 hours of future weather. The user can also adjust the frame rate frequency continuously from 15 min to 3 hours (Figure 3a) if desired. Weather information is available over the CONUS from the current time to 6 hours prior. From 0-2 hours, the CONUS CIWS forecast is displayed; as the

display transitions to the CoSPA 2-6 hour forecast the domain becomes restricted to the Northeast Corridor (Figure 3b), which is the current limit of the HRRR computational domain. Previous forecasts can be evaluated by activating the "Verification" contours product in which forecasted contours of VIL level 3 valid at the current time are overlaid on the current VIL image (Figure 3c).

The "Short Trends" product shows storm growth and decay areas that have been detected over the past 15 to 18 minutes. Recent storm growth is shown in orange areas with a black cross hatch, and recent storm decay is shown in navy blue areas (Figure 3d). The "Storm Motion" product displays black vectors estimating storm speed and direction and storm extrapolated positions are displayed in aqua-colored markings predicting where the edge of level 3+ VIL is projected to be in 20 minutes (Figure 3d). The "Lightning" product displays recent cloud-to-ground lightning strikes using real-time data from the National Lightning Detection Network (NLDN).

The "Echo Tops" product displays a radar measurement of current cloud top height at a 1 km resolution. This can be useful for determining whether or not aircraft can fly over storms that may appear to be not traversable based on VIL intensity alone. In addition, Echo Tops Tags can be enabled to show cloud top height in units of hundreds of feet (Figure 3e).

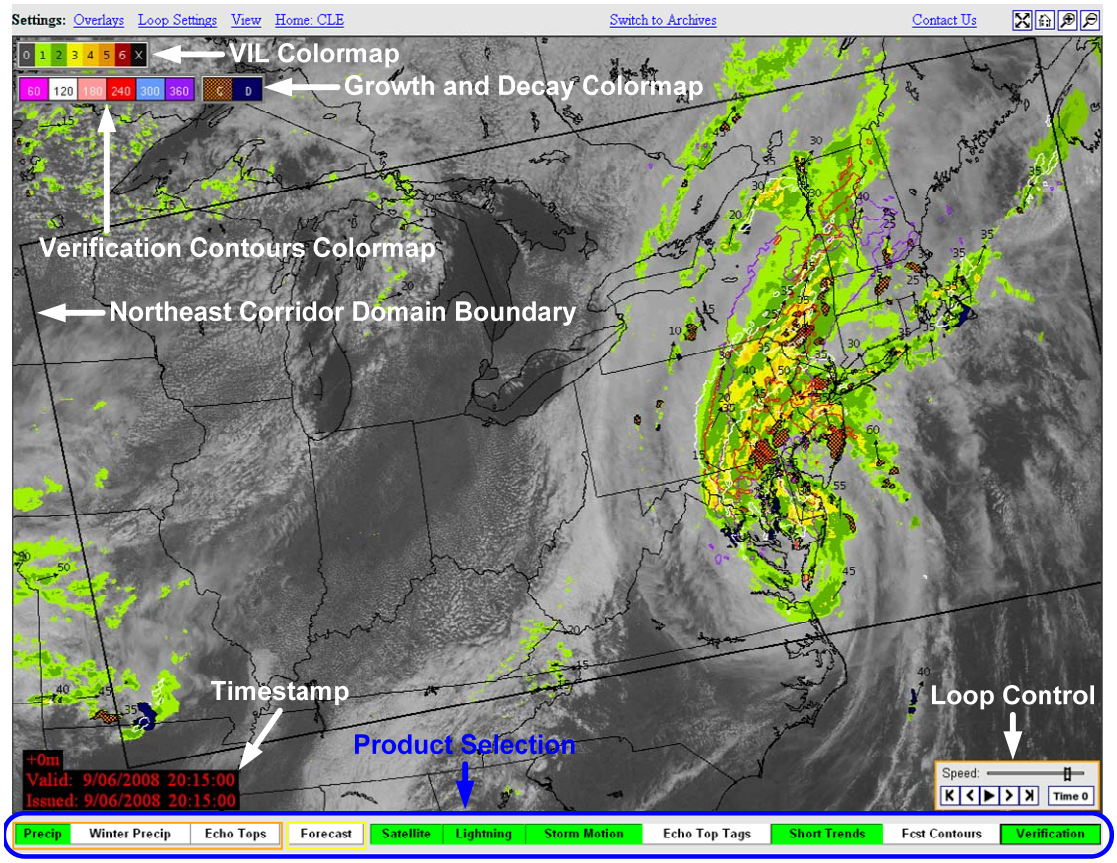

Figure 3a. The web situation display for CoSPA is shown for 2015 UTC on September 6, 2008 with Hurricane Hanna moving up the east coast. The display can run as an animation loop with 6 hours of past weather that transitions into a forecast out to 6 hours. The tabs at the bottom show the various products that can be displayed. These products were developed specifically for ATM use through years of user interaction.

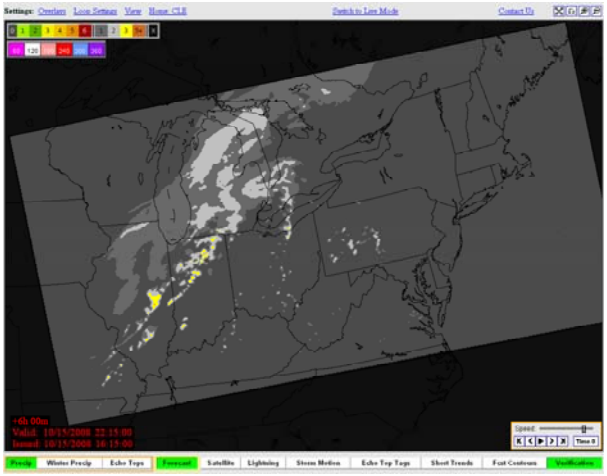

Figure 3b. The web situation display for CoSPA is shown with the 6 hour VIL forecast at 1615 UTC on October 15, 2008. 2-6 hour forecasts are currently limited to this Northeast Corridor domain.

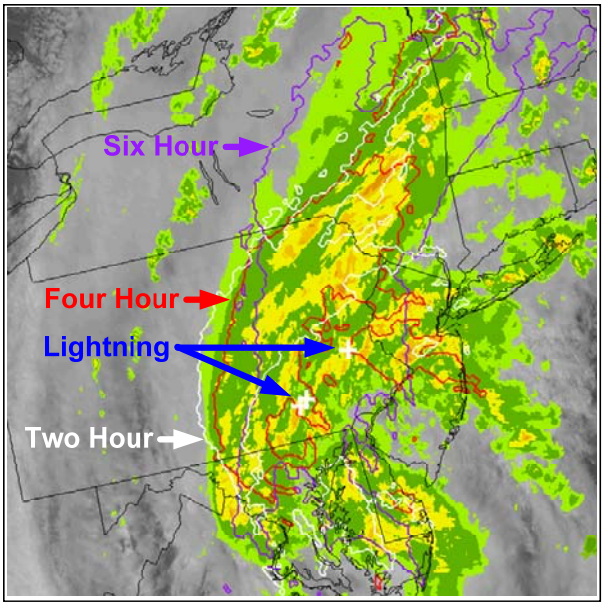

Figure 3c. A zoomed in section of Figure 3a is shown with the Verification and Lightning products. Lightning strikes are marked by white crosses. Verification contours for past forecasts are shown in white, red, and purple for the 2, 4, and 6 hour previous forecasts respectively.

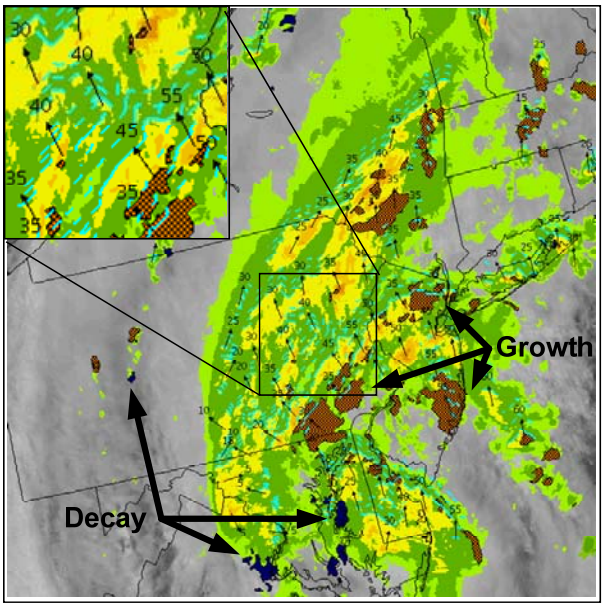

Figure 3d. This figure shows short term growth and decay trends (Short Trends) and Storm Extrapolated positions (SEP). The SEPs are shown in the aqua-colored lines and indicate where the edge of level 3+ VIL is moving (see zoomed in box).

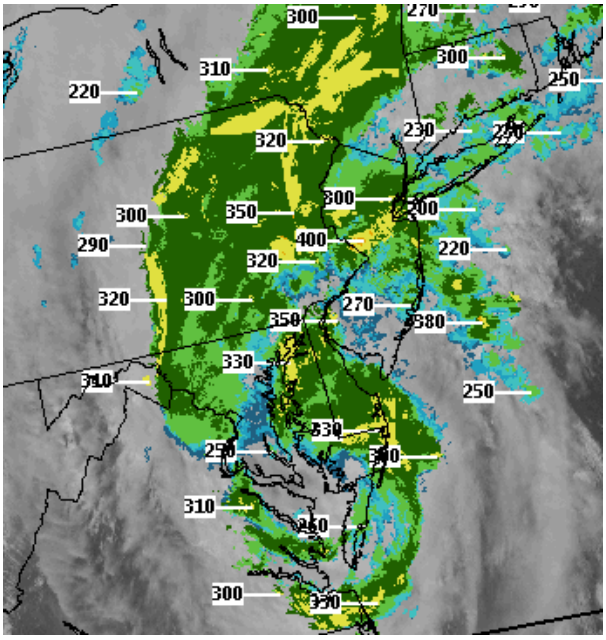

Figure 3e: The web situation display for CoSPA is shown with the Echo Tops Mosaic and Echo Tops Tags at 2015 UTC on September 6, 2008 in units of hundreds of kilofeet.

Table 1: Showing equivalence of VIP levels to VIL in units of kg/m^2 and 8 bit encoded values called digital VIL. Column 4 shows the colors for past and current VIL, and column 5 shows the colors for forecasted VIL.

VIP or VIL Levels	VIL (digital VIL)	VIL (kg/m^2)	Past and Current VIL	Forecast VIL
1	16	0.15	Level 1	Level 1
2	74	0.76	Level 2	Level 2
3	133	3.47	Level 3	Level 3/4
4	160	6.92	Level 4	Level 3/4
5	181	12.0	Level 5	Level 5/6
6	219	31.6	Level 6	Level 5/6

3.3 Verification Tools

A number of tools are available on the website for CoSPA forecast verification and analysis. The Verification tab on the web situation display can be applied at the current time to show how well past forecasts performed. As discussed above, contours can be displayed to depict the location of the level 3+ VIL, and are available from 1 hour to 6 hours in intervals of 1 hour. Figure 3c shows the 2, 4, and 6-hour verification contours.

A link to a forecast analysis tool is provided for comparison and analysis of various forecast products. It can be used to compare forecasts such as the CoSPA Blended VIL, MIT LL Extrapolated VIL, and the Collaborative Convective Forecast Product (CCFP).

4. FORECAST TECHNOLOGY

This section discusses the three main components of the CoSPA forecast: the heuristic extrapolation forecast, the HRRR numerical model and the blending algorithm. For the 0-2 hour forecast, CIWS technology is used; for a review of CIWS see Dupree et al. (2005, 2006) and Wolfson and Clark (2006).

4.1 Extrapolation Forecast

Tracking storms and producing forecasts on multiple scales remains an area of active research (Bellon and Zawadzki, 1994, Wolfson et al., 1999, Seed and Keenan, 2001, Lakshmanan et al., 2003 and Dupree et al., 2002,2005). All of these studies apply some type of scale classifier to separate features and use a tracking method on those features, usually either cross-correlation tracking or mean-squared error tracking. A key result is that all studies show large-scale features are more predictable than small-scale features, and large-scale features can be extrapolated to longer time horizons with greater accuracy than small-scale features. Additionally, the longer the forecast time horizon, the larger the minimum scale at which meaningful motion data may be extracted. For the 0-2 hour forecast used in CIWS, we used two scales: the envelope scale indicative of line storm, supercell, or stratiform weather motions, and the single cell scale.

The CoSPA extrapolation technique was designed to adapt and improve upon the advection techniques used in CIWS. The 0-2 hour forecast uses a simple Eulerian-like extrapolation technique which works well for short-term motion; however, to obtain forecasts out to 6 hours, larger scales are necessary. Additionally, the Eulerian techniques do not capture rotational motion very well, and therefore it is desirable to separate or decompose short-term rotations of the smaller scale storm features from larger-scale translational motion of the entire storm system. Hurricanes provide a good example of the separate scales of motion: smaller short-lived convective elements typically rotate around the center of the hurricane in bands, while the hurricane as a whole typically translates more or less in a straight line over much longer (~6-12 hour) time scales. Here we present a new extrapolation technique that combines these scales and uses the motion decomposition to produce a seamless 0-6 hour forecast.

The motion prediction consists of three fundamental steps: 1) filtering and tracking, 2) interpolation of motion fields, and 3) advection of the weather. First, for the filter and track step, the motion of a storm system must be determined and distilled into motion vector fields at several somewhat independent scales. To create the raw motion vectors from the observed data, the input precipitation (VIL) images are filtered with a set of mean filters followed by cross correlation on each output time series. Three scales are used for the extrapolation; these are the cell, envelope and synoptic scales shown in Figure 4a. Two of the three motion scales are created in the CIWS system: the cell scale, a 13 km diameter circular mean filter with a 6 minute correlation, and the envelope scale, a 13x69 km filter with an 18 minute correlation time. A new scale, created for CoSPA, applies for longer time horizons: the synoptic scale, a 101x201 km filter with a 45 minute correlation time. For the interpolation step, each set of raw motion vectors is interpolated to

create a smooth vector map for each scale. A data quality editing routine is applied, which compares the updated vectors to a reference field and determines a deviation weight for each vector. Vectors that don't agree with the reference field get no or very low weights, whereas vectors that are consistent with the reference get higher or full weights. The vectors are then interpolated based on the assigned weights. For the synoptic scale, model winds are used as proxies for motion for areas without radar data and a variational technique is used to blend the observed motion with the model winds. A weighted mean of several pressure levels of the RUC winds are used for this model wind estimate. The interpolated vector fields as well as the VIL growth and decay image are then passed into the advection routine.

The advection process uses two steps to move the separate scales. First, a pseudo-Lagrangian[3] advection is applied to the small scale motions (cell and envelope), and second, an Eulerian advection step is applied to the synoptic scale. For the first step, the synoptic motion is subtracted from the cell and envelope scales, and the resulting field is applied in a pseudo-Lagrangian sense to the VIL image. The method works as follows: a pixel is advected with a 15-minute time step, and then placed at a new location. The pixel is then picked up and advected with the motion of its new location. The pixel therefore should approximately follow the streamline of the small scale (rotational) motion field. The cell vectors are used out to the 30-minute time horizon, and then the envelope vectors are used out to a 90-minute time horizon. Because the cell- and envelope-scale motions don't apply well at the longer time scales, the cell and envelope scale vectors are reduced over time and do not persist beyond a 90-minute time horizon. Once the final time horizon is reached, the Eulerian step is applied, in which the rotated pixels are advected using the synoptic-scale motion vectors.

A schematic of this process is shown in Figure 4b, with the smaller arrows representing the Lagrangian steps and the large arrow representing the Eulerian step. The two steps combine to move the weather to its final location.

An example of the extrapolation forecast is shown in Figure 4c. The 27 July 2008 line storm was well established at 19 UTC, and the forecast shows the line moving to the east. The extrapolation forecast shows some deformation at longer forecast times, but shows largely reasonable motion and shape of the line storm. Growth and decay trends are used in the 0-2 hour forecast; however, no growth or decay is included in the 2-6 hour extrapolation forecast. This is remedied by blending the extrapolation with the model-based HRRR forecasts, as described later.

4.2 *High Resolution Rapid Refresh (HRRR)*

An experimental version of the Weather Research and Forecasting (WRF) model called the High Resolution Rapid Refresh (HRRR) model (Benjamin et al. 2009) is being run at NOAA's ESRL GSD laboratory. The HRRR model is a 3-km resolution model that is nested inside an experimental version of the 13-km Rapid Update Cycle (RUC) model that assimilates three-dimensional radar reflectivity data with an effective method based on a diabatic Digital Filter Initialization (DFI) technique. The HRRR model benefits from the RUC radar data assimilation through the lateral boundaries throughout the forecast as well as in improved initial conditions. In addition, the high resolution of the HRRR obviates the need for convective parameterization, further reducing uncertainty of the forecast and allowing the model to produce realistic convective structures vital for improved forecast fidelity.

The HRRR model updates once an hour and generates forecasts out to 12 hours. VIL forecasts have been made available at a special 15 minute time horizon frequency for the CoSPA forecast system in order to take advantage of the blending technology.

For the 2008 summer demonstration, HRRR was run in the Northeast corridor domain as shown in Figure 5. The HRRR has shown remarkable skill at depicting storm organization and evolution. In particular, the HRRR typically provides clear guidance in the distinction between scattered and organized convection, which is critical information for aviation planning.

[3] Here we refer to the method as pseudo-Lagrangian because no mathematical Lagrangian operators are utilized in the algorithm.

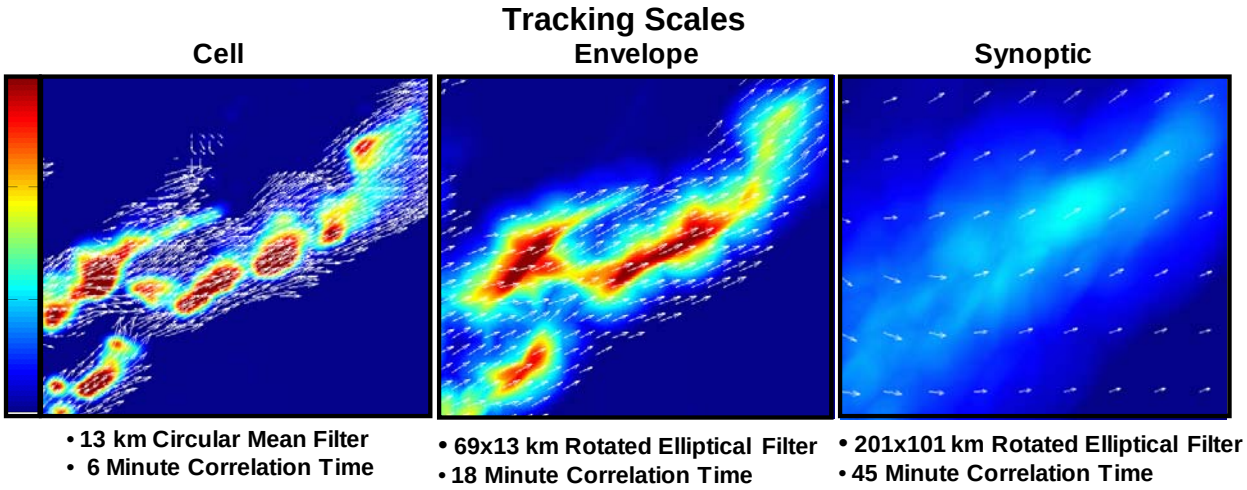

Figure 4a: Three scales used to create raw motion vectors from smallest to largest: cell, envelope and synoptic. Motion vectors are in white, background is spatially filtered interest image. VIL is passed through the filters, then correlated with previous images to calculate the best motion vector.

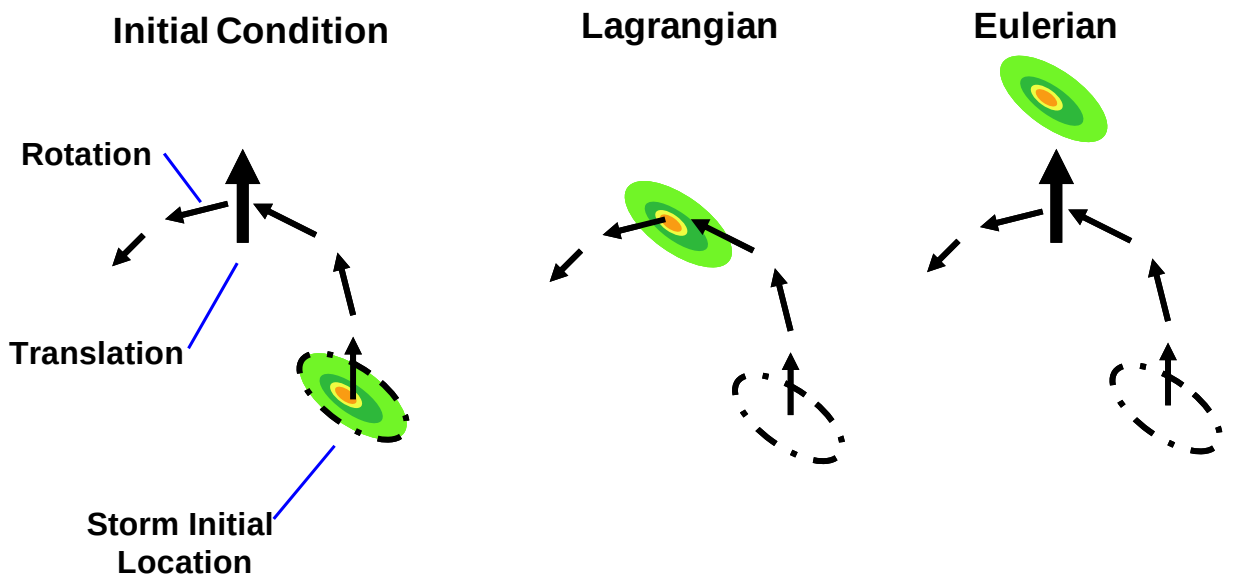

Figure 4b: Schematic depicting multiscale advection technique. Initial image is advected in many short steps using successive small scale vectors, then in one large step using large scale vectors. Small arrows represent small scale vectors used in the small scale step, the large arrow represents vectors used in large scale step.

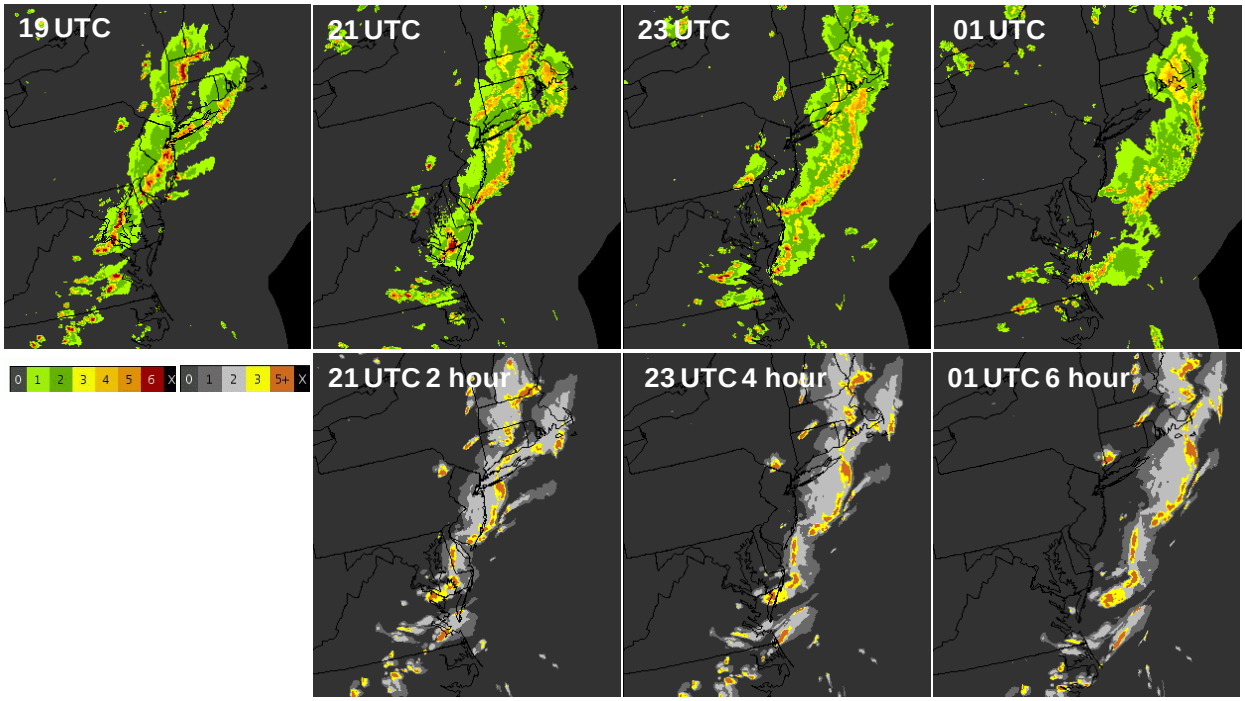

Figure 4c: Example of extrapolation forecast for 27 July 2008 at 19 UTC. The upper left hand panel shows initial weather, the remaining upper panels show truth for 2, 4, and 6 hours and the bottom panels (from left to right) show the forecast for two, four and six hour lead times.

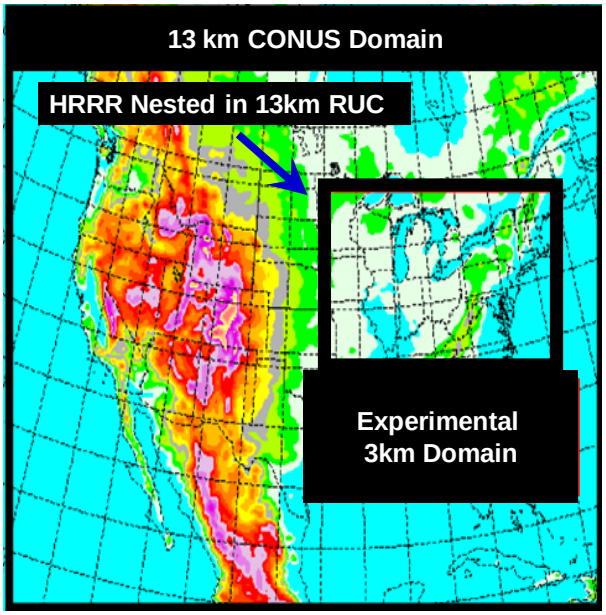

Figure 5: HRRR model nested in the Rapid Refresh (RUC) model. Depicted is the Experimental Northeast domain over which the HRRR was run during 2008.

4.3 *Blending*

The blending algorithm (Figure 6a) has been designed to combine extrapolation and model forecasts of VIL to produce a seamless, rapidly-updating 0-6 hour forecast of weather intensity. This is done through a calibration of model data, a phase correction to remove location errors in the model and statistically-based weighted averaging. In CoSPA, heuristic extrapolation forecasts of VIL from MIT LL are blended with VIL forecasts from the HRRR model.

The model VIL field (also called the Rain Water Path; RWP) is obtained by vertically integrating the snow, rain and graupel water mixing ratios over the depth of the atmosphere. The model VIL field is then calibrated by performing a frequency matching procedure so that the distribution of modeled VIL values matches the observed distribution. In this case, the observed VIL is reported as an 8 bit integer (i.e., 0-254) and is called digital VIL (see Table 1). Thus, an additional step converts the modeled VIL field (units of g m^{-2}) to a calibrated 8-bit integer. An example of the frequency matching is shown in Figure 6b for a given forecast. The resulting conversion equation is determined via a least

squares fit and is found to vary as a function of the model forecast generation time. The calibrated 8-bit model VIL ($digVIL_m$) may then be determined empirically following:

$$digVIL_m = b_1 + b_2 VIL_m$$

if $VIL_m \leq threshold$

$$digVIL_m = a_1 + a_2 \log(a_3 (VIL_m - 2.0)^{\gamma})$$

if $VIL_m > threshold$

where VIL_m is the modeled VIL field (g m^{-2}), the coefficients ($b1$, $b2$, a_1, a_2, and a_3) are allowed to vary as a function of forecast generation time and γ is a constant.

Spatial offsets between $digVIL_m$ and the observed $digVIL_o$ are then reduced using an Eulerian phase correction. Phase errors in $digVIL_m$ are determined using a variational echo tracker that compares the current radar mosaic image of VIL with the modeled VIL from the most recent forecast valid at the observation time. Due to the time it takes to run and transmit the model forecast, the VIL analysis is typically compared with a three hour forecast. Since the comparison is made at a single time, the resulting correction is called an Eulerian phase correction. When multiple times are compared, a phase error tendency can be calculated and used to correct the model – this is called a Lagrangian phase correction – the magnitude of which may vary with time. The phase error vectors are determined by minimizing a cost function that has two constraints: distance error and relative smoothness – see equations 9-11 in Germann and Zawadski (2002). These vectors are then used to correct the model VIL field at each model output forecast lead time. An example is shown in Figure 6c. In this example, subtle shifts (generally to the south) in $digVIL_m$ reduce the area covered by lower thresholds and generally reduce edge-to-edge offsets (e.g., see increased overlap of $digVIL_m$ with observations within the cyan circle).

So far only the Eulerian phase correction has been implemented in the real-time system, owing to its simplicity and computational efficiency; however, later releases of the blending system will include the Lagrangian option. It is noted that the Eulerian phase correction does not perform well in situations where the phase errors change rapidly with forecast lead time.

Time-varying weights are determined from relative performance of the phase corrected model and the extrapolation forecasts (Figure 6d). The performance is determined by looking at a combination of Bias and CSI scores. Generally, the model is given more weight at the longer lead times, with equal weighting at about 4 hours. However, the model weights can vary as a function of the time of day, the model is given more weight (at the earlier lead times) during the period of most rapid storm initiation and growth over the CONUS (i.e., 10-15 UTC) as this period of rapid change is difficult to handle through observation-based approaches.

Timing of the real-time data feeds is a critical aspect of the system. A new forecast is generated every 15 min with forecast output frequency of 15 min out to 6 hours. The forecast blending hinges on the latency of the model forecast, which is typically 2-3 hours old by the time it is available (this latency will be reduced for next year's demo). The phase correction procedure then matches the current radar mosaic data with the appropriate forecast lead time and performs the phase correction. At the same time, the final blending step is constantly "watching" for new forecasts to combine using the weighted averaging.

The resulting final forecast is a combination of heuristic extrapolation and high resolution model forecasts of VIL. The average skill scores (CSI, bias) obtained for a single day (27 July 2008) shown in Figure 6d illustrate the relative performance of each of the components of CoSPA for VIL levels 2 (minimal en route impacts) and 3 (en route aviation impacts possible). Note how the "cross-over" point decreases as the threshold increases. That is, the model skill increases relative to extrapolation/heuristics as a function of storm intensity. Also note that the phase correction and blending algorithm adds more skill at the lower thresholds. In addition, it is critical to keep biases less then 1.5 to minimize false alarms.

The final resulting forecast aims to optimally combine extrapolation and heuristics with high resolution NWP output. This is demonstrated in Figure 6e in which rapid storm growth is well captured by the blending algorithm with small storms accurately depicted to grow into large aviation-impacting clusters along the East coast corridor in 6 hours. Forecasts that accurately depict storm evolution and morphology are critical for making well-informed decisions related to routing air traffic across the NAS.

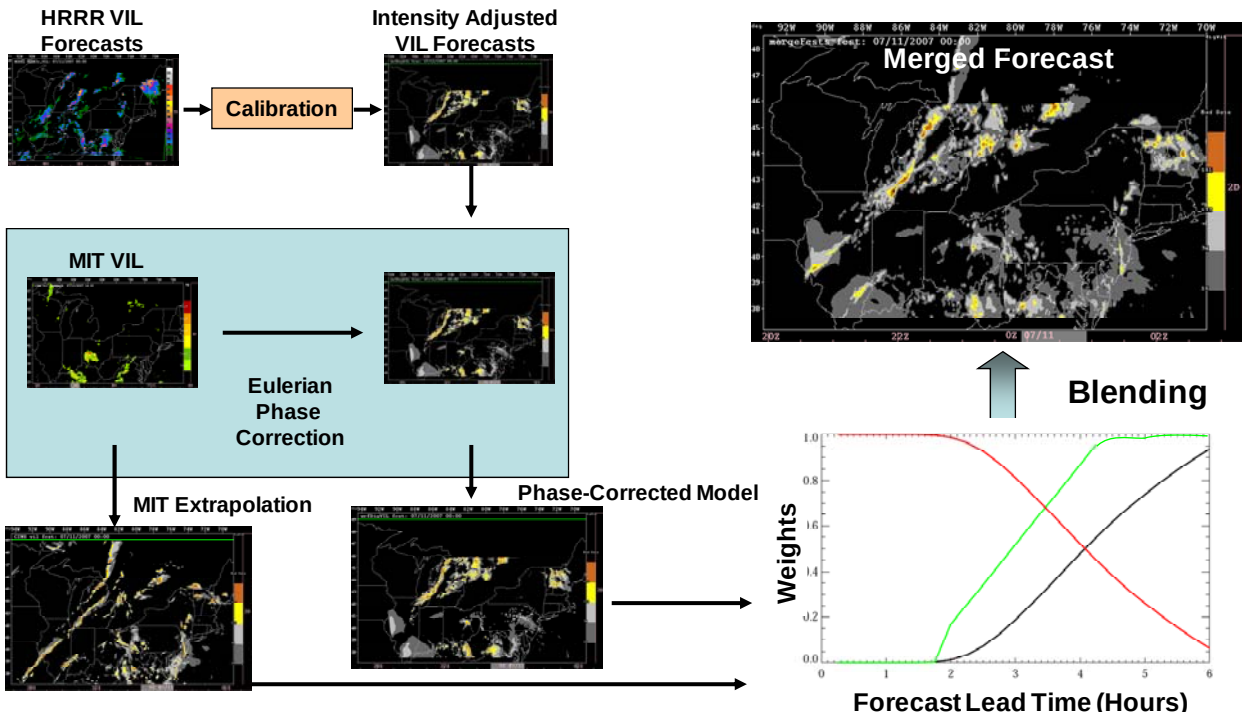

Figure 6a: Schematic flowchart of current blending algorithm used in CoSPA. The lower-left panel gives the weight as a function of forecast lead time. The red and black curves give the standard weights for extrapolation and model respectively, while the green curve depicts the weights used between 15-20 UTC.

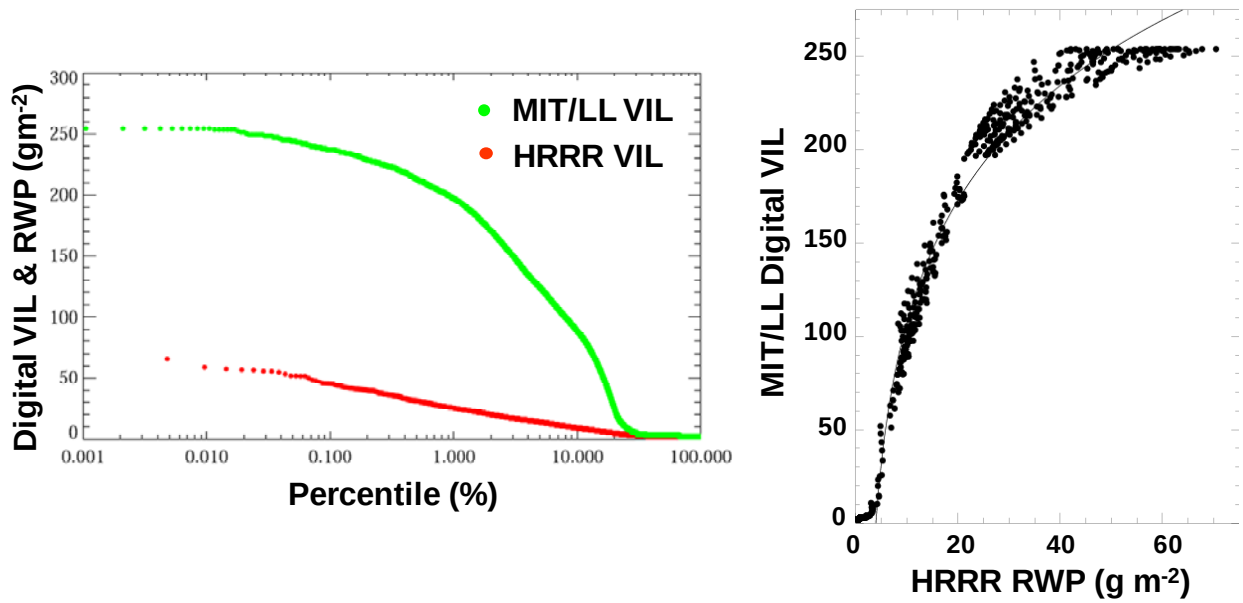

Figure 6b: Left panel shows a comparison of the frequency distributions of MIT LL radar-digital VIL and HRRR model VIL (called RWP in the figure) (red). The right panel shows the results of frequency binning between the HRRR model VIL (RWP) and MIT LL radar-digital VIL for an entire forecast (2-12 hours). The line is the empirical fit to the data.

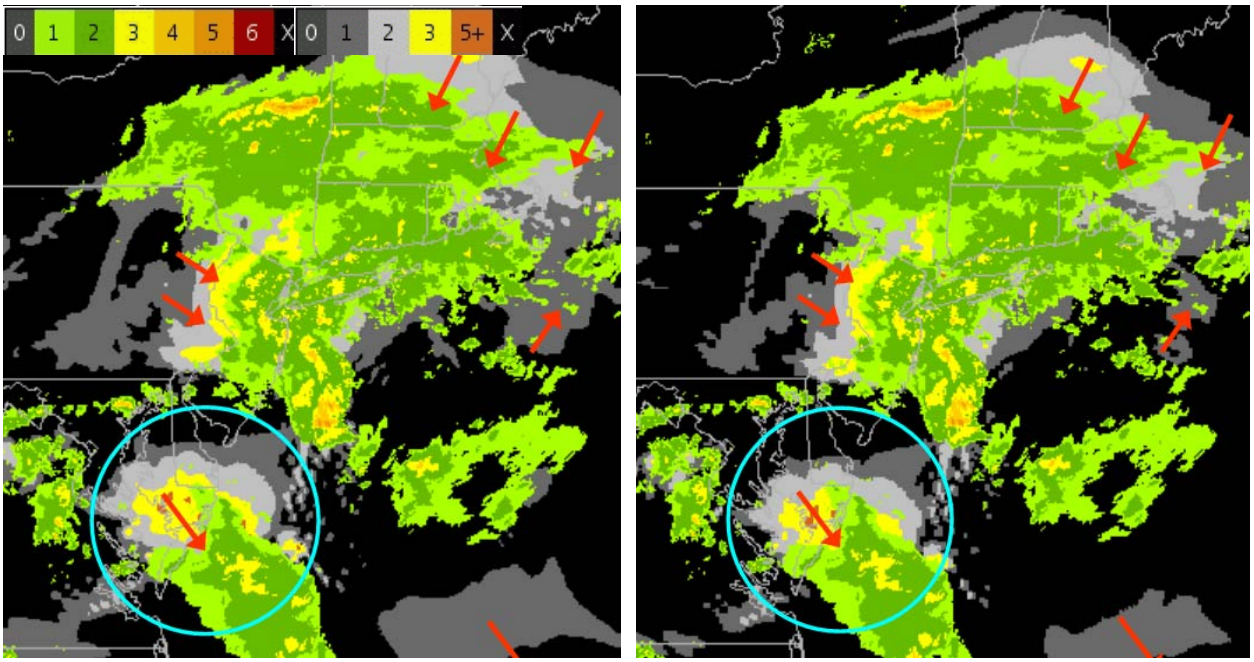

Figure 6c: Images depicting the modeled 4 hour forecast of VIL field (gray shades) and observed VIL (colors) before (left) and after (right) Eulerian phase correction. The arrows represent the direction of the shift. The cyan circle denotes area where storms where shifted 50 km to the southeast (closer to reality).

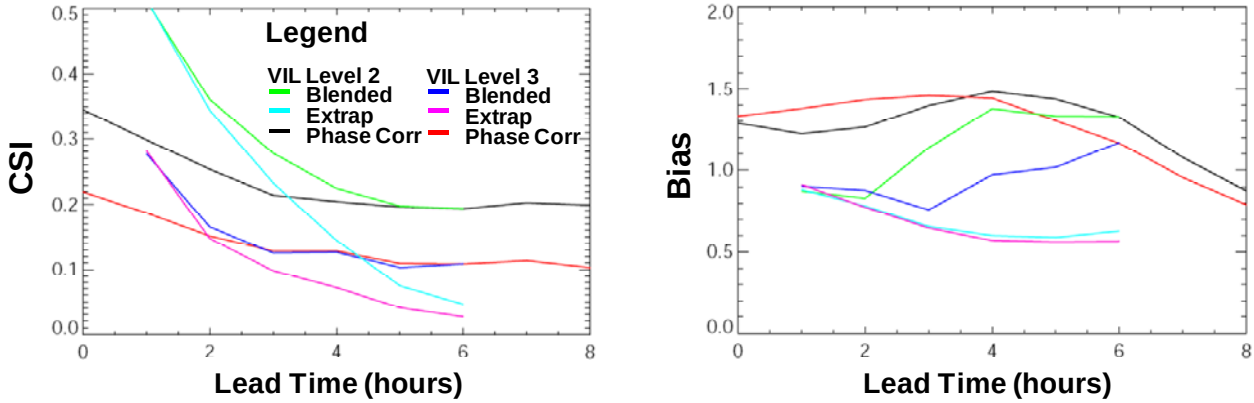

Figure 6d: Forecast skill scores (left - CSI and right – Bias) as a function of lead time for two VIL thresholds (VIL level 2 and VIL level 3) for phase-corrected model (black, red) and extrapolation (cyan, magenta), blended forecasts (green, blue) for forecasts issued between 14 and 22 UTC 27 July 2008.

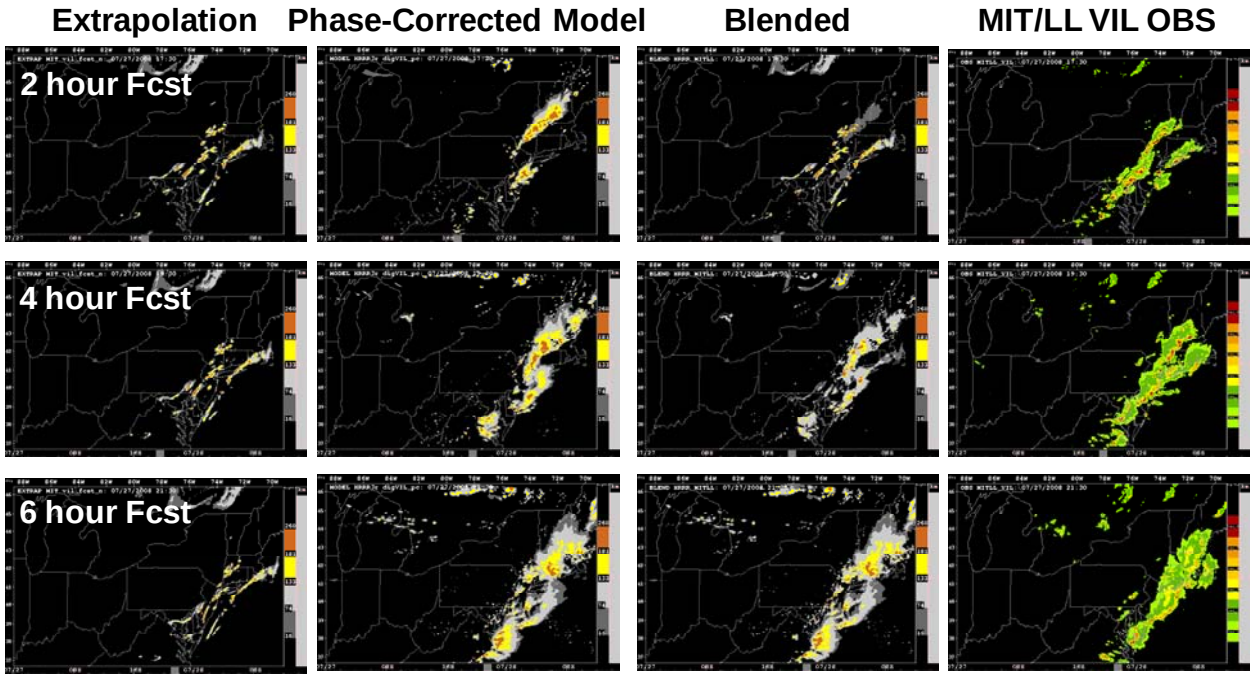

Figure 6e: Example of blending algorithm showing inputs at 2, 4, 6 hour lead times, blended forecast, and VIL observations at forecast valid times 1730, 1930 and 2130 UTC for forecasts issued at 1530 UTC on 27 July 2008.

5. VERIFICATION AND PERFORMANCE

The performance of the CoSPA forecast was monitored throughout the summer at both MIT LL and NCAR through a variety of methods. The methods include: 1) Real-time verification code, 2) on site assessment by meteorologists, and 3) post performance analysis using playback and analysis tools.

We ran real-time verification modules to generate forecast statistics and verification map products. One useful map product was a binary forecast score image that geographically depicts hits, misses and false alarms for various intensity thresholds levels by comparing the VIL truth and VIL forecasts at forecast valid times. An example of the forecast score images is depicted in Figure 7a and provides a qualitative assessment of the skill. Daily time series of forecast statistics were also computed. These included Bias, Probability of Detection (POD), Critical Success Index (CSI), and Probability of False Alarm (PFA; Wilks 1995).

On site assessment consisted of monitoring CoSPA products for data dropouts, problem forecasts and potentially operationally-significant events. Verification contours can be displayed on a current VIL image, outlining the location of significant weather from the past 2, 4, and 6 hour forecasts, valid at the current time. In addition, the Configurable Interactive Data Display (CIDD; http://www.rap.ucar.edu/CIDD/user_manual/CIDD_manual.html) was used by the staff monitors to analyze the forecast system, including primitive predictors.

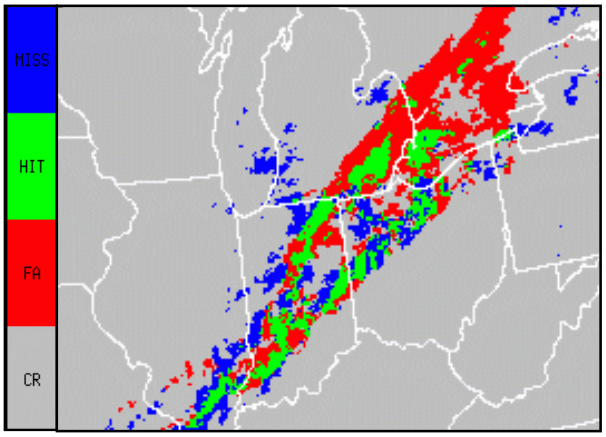

Figure 7a: Example of a forecast score image created by thresholding the VIL 4 hour forecast and truth at VIL level 2 and superimposing the images. Forecast hits are shown in green, misses in blue, false alarms in red, and correct rejections in grey.

For post analysis events, an archive was added to the web page to allow case days to be reviewed. In addition, a web tool called the CoSPA Forecast Analysis Tool was developed to

compare CoSPA's blended forecast with other forecast products.

Using all of these methods, it has become clear that CoSPA has forecast skill and performs well in most situations. An example 4-hour forecast of a large-scale line is shown in Figure 7b. Note the large scale character and location of the line storm was well-resolved. This example also shows that this line is for the most part permeable, and planes could fly through gaps in most areas along the line, except perhaps a small section in Southern Illinois. Figure 7c shows an example from 8 August 2008 in which convection grew up during a 6 hour period. The 6-hour forecast issued at 11Z shows a nice example of large-scale, widespread initiation. The location of the large-scale instability region is evident, as is the scattered airmass mode of the convection. Even though the exact locations of the storms are not perfect, the forecast clearly shows that air traffic could move through this weather, since there are many open flow paths.

Storm location, structure, and scale have emerged as having an important role in CoSPA performance. The forecast is more robust for larger scale convection, such as a line or bow echo, but the growth mode of smaller scale weather can also be well-forecasted out to 6 or 8 hours. Under-forecasting of VIL intensity has been observed to occur in the 3-4 hour blended forecasts, whereas the 6 hour forecast appears to over-forecast the intensity of convection. In addition, weather that enters the HRRR domain from the western edge has been shown to be forecasted more poorly than in other regions of the domain; this is likely due to boundary effects in the HRRR. This problem can likely be alleviated by expanding the computation domain in the HRRR. The CoSPA forecasts continue to be monitored, analyzed, and enhanced to address these issues.

6. FUTURE WORK

6.1 Traffic Flow Metrics Scoring

A new route blockage algorithm has been developed and has potential use as a scoring metric in CoSPA (Matthews et al., 2009). The new algorithm applies a filter to both VIL and Echo Tops to remove non-route-impacting weather, according to a new convective weather avoidance model. The algorithm can be run on both observed and forecasted storms, and hence can be scored as a forecast of route blockage. Work is currently being conducted to test and improve the performance of this scoring metric on the CoSPA forecasts, with introduction into the system planned for 2009.

6.2 Convective Initiation

Convective initiation (CI) remains a nowcasting challenge at all forecast time horizons. At short forecast time horizons (under 1 hour), precursors to CI can be identified in infrared imagery from the Geostationary Operational Environmental Satellites (GOES) prior to the existence of radar echoes. An example of a satellite-based CI precursor is the rate of cloud-top cooling, which has been shown by Roberts and Rutledge (2003) to be an indicator of convective initiation.

15-16 October 2008

Truth **4 Hour Blended Forecast**

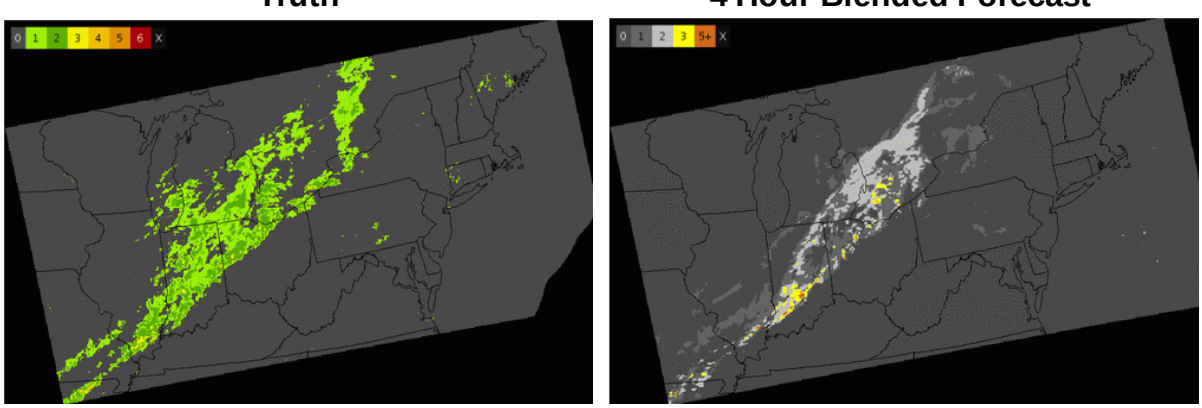

Figure 7b: Comparison of the forecast truth at 01 UTC 16 October 2008 (left panel) with the 4 hour blended forecast issued at 21 UTC on 15 October 2008 (right panel) viewed using the CoSPA Forecast Analysis Tool available on the CoSPA website. Note that the large scale character and location of the line storm was resolved.

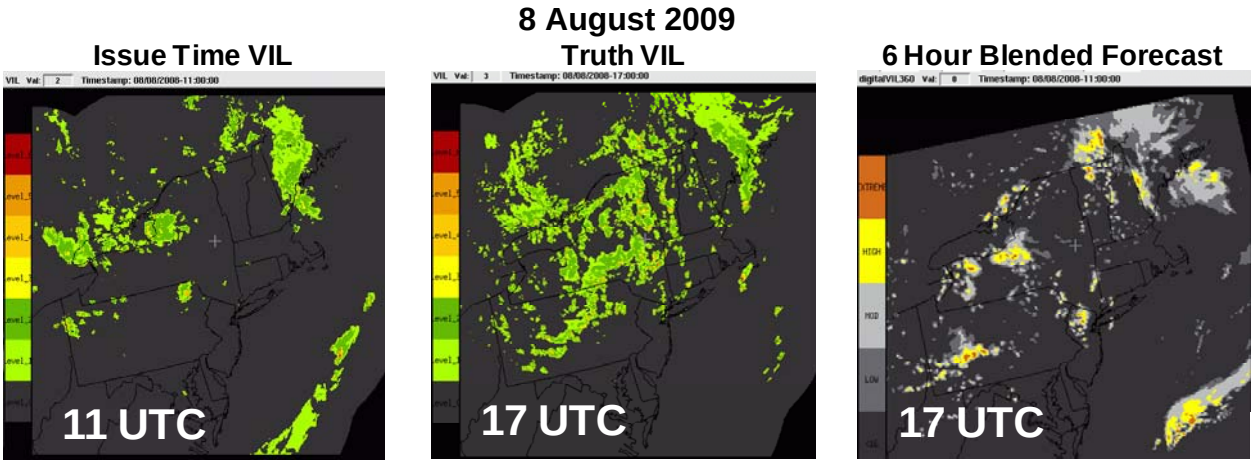

Figure 7c: Comparison of initial and observed VIL with the 6 hour blended forecast issued at 11 UTC on 8 August 2008. Note that the forecast predicted the growth of the scattered airmass storms in this case.

Using cloud-top cooling rate and other satellite-based indicators to forecast CI is the basis of the Satellite Convection Analysis and Tracking System, known as SATCAST (Mecikalski and Bedka 2006). The SATCAST system uses a cloud mask component, an atmospheric motion vector component, and a nowcasting component to create eight satellite-derived CI indicators based on tracking and trending of cloud properties in multiple infrared channels. The eight indicators are combined into a single CI nowcast field with values from zero to eight, where pixels with higher values indicate a higher confidence in CI.

SATCAST development was funded by the NASA Advanced Satellite Aviation-weather Products (ASAP) program. The NASA ASAP program provides satellite-derived meteorological products and expertise to the FAA weather research community. The SATCAST system is currently being transitioned to and tested in the CoSPA environment to evaluate its benefit to convective forecasts for aviation. As part of this transition and evaluation, algorithms have been developed that utilize the SATCAST CI indicators in CoSPA. These algorithms use a combination of atmospheric variables (including stability and lower-tropospheric winds) and image processing to create enhanced CI interest in regions identified by SATCAST to be favorable for CI. The CI interest is then combined with other CoSPA interest fields to create a forecast.

Figure 8 provides an example of a CoSPA 90-minute forecast with and without SATCAST CI. SATCAST CI pixels with high scores (red pixels in Figure 8a) identify potential areas of CI in western Minnesota. Two of these areas (circled in Figure 8a) are in a region of atmospheric instability as determined from the CoSPA stability mask (not shown) and are therefore highly favored for CI. CI interest is generated in these two areas and is used in the CoSPA forecast. Although there are no radar echoes over western Minnesota at the time the forecast is made, storms subsequently developed over western Minnesota, and 90 minutes later these storms moved over central Minnesota (Figure 8b). The 90-minute CoSPA VIL forecast without SATCAST CI (Figure 8c) shows no evidence of these initiating storms, whereas the CoSPA forecast that utilizes the information provided by SATCAST initiates these storms correctly (Figure 8d). Future work will continue to improve the use of the SATCAST CI indicators in CoSPA with the goal of improving short-term CI in convective forecasts.

6.3 Random Forest

Methods for enhancing the fusion of observation and NWP model data have begun to show promise as an approach to improving CoSPA's short-range forecasts. These methods may help developers address several challenges. Since incorporating a new data source into the fuzzy logic forecast engine is a time-intensive process, it would be helpful to first objectively identify the potential contribution of new candidate predictor fields, taking into account that there may be complex dependencies and some predictors may be important only in certain scenarios. It would also be desirable to create a forecast engine performance benchmark to provide information on a minimum level of expected skill based on a given set of input fields. Finally, NextGen applications require CoSPA to provide

estimates of forecast uncertainty and probabilistic forecasts.

Early results suggest that a non-linear statistical analysis technique called *random forests* (Breiman 2001) holds promise for addressing these challenges (Williams et al. 2008a, b). A Random Forest (RF) is a set of decision trees, created via an automated "training" process, that collectively relate a vector of predictor values (e.g., model fields and observation features at a map pixel) to a targeted output quantity (e.g., whether there will be a storm there one hour later). A representative training data set consisting of vectors of predictor values with associated true outcomes is used to construct the RF; each decision tree is trained on a random subset of the data, and at each node the best split is chosen from a random subset of the predictor fields. By constraining their power in this way, a set of weakly-correlated decision trees is formed that may serve as an "ensemble of experts" by "voting" on the correct classification of a novel data point. Moreover, during training, the RF produces objective estimates of variable importance that may be very useful for selecting a minimal, skillful set of predictors or comparing the value of a new candidate predictor with others.

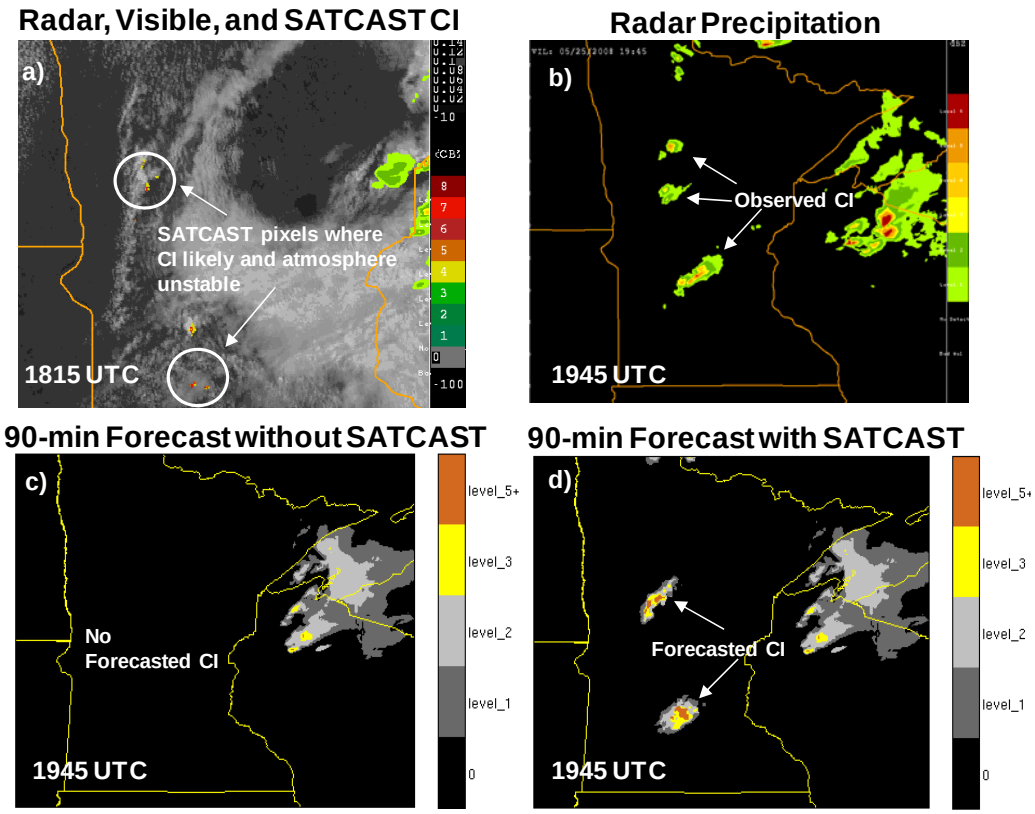

Figure 8: Example of using SATCAST CI products in the CoSPA convective forecast system for 25 May 2008. a) SATCAST CI nowcast scores (≥4 shown), visible satellite, and VIL over MN at 1815 UTC. Circled pixels indicate highly favored regions of CI where there are high SATCAST CI nowcast scores and the CoSPA stability mask (not shown) indicates that the atmosphere is unstable. Note that there is no VIL in these regions at this time. b) The observed VIL 90 minutes later shows that CI has occurred in western MN and the storms have moved over central MN. c) The VIL forecast without SATCAST CI does not depict the newly-developed storms over central MN, whereas the forecast with the SATCAST CI (d) captures the convective initiation.

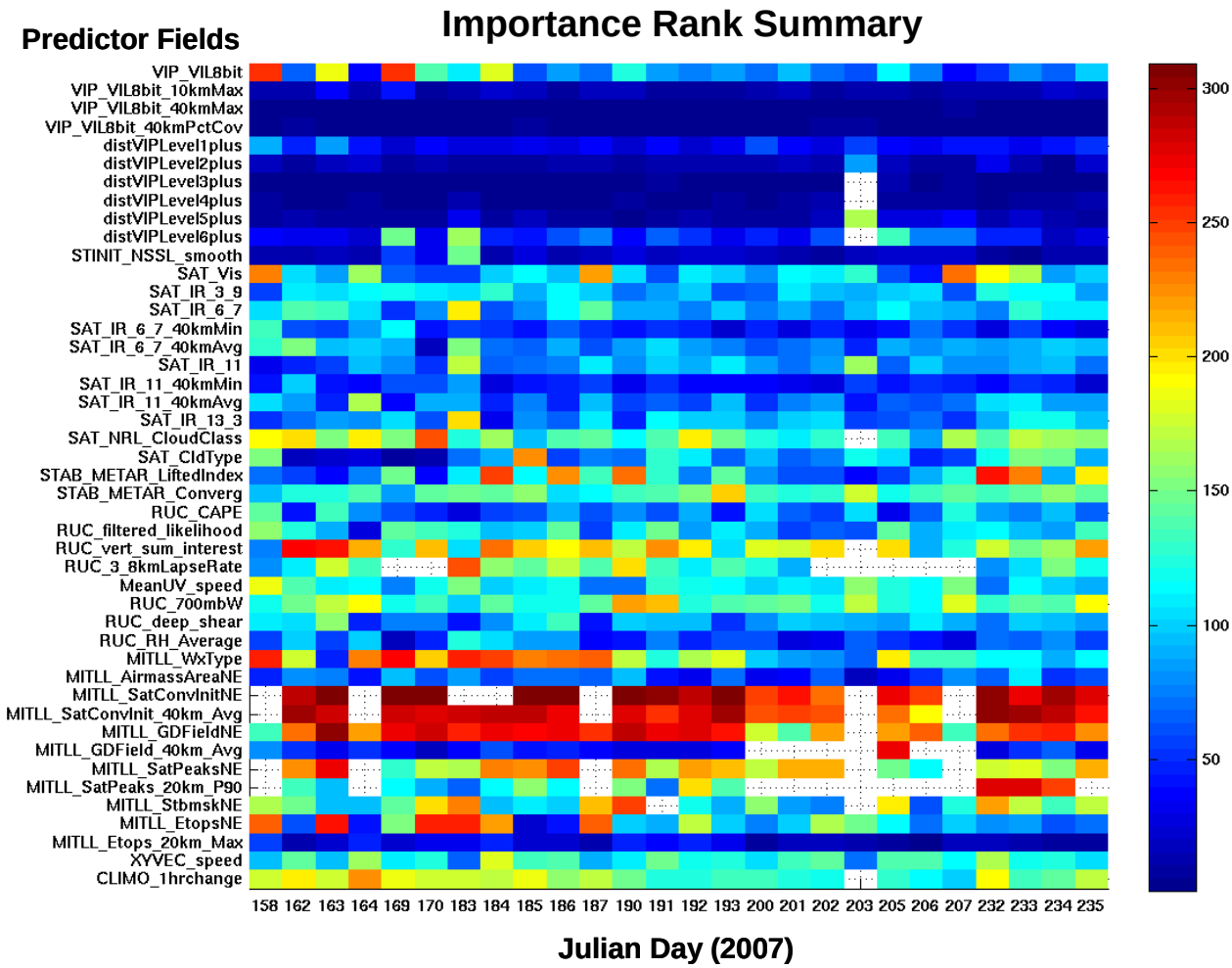

Figure 9a: Color-scaled summary plot of the RF importance ranks for a subset of 45 candidate predictor fields (y-axis) in predicting whether VIL ≥ 3 at a pixel one hour later as a function different days (x-axis). Lower ranks (blue colors) represent greater importance. Variation may be due to differences in the synoptic regime, or to small sample sizes for some of the days.

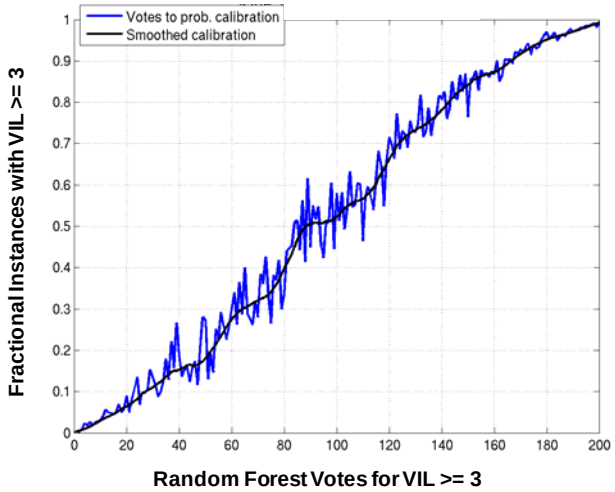

Figure 9b: Reliability diagram (blue line) and smoothed calibration curve (black line) for an RF trained to predict VIL ≥ Level 3 (equivalent VIL ≥ 3.5 kg/m^2) at a pixel, evaluated on an independent test set.

Initial experiments have focused on one-hour predictions of storm intensity (VIL level) based on a large number of candidate predictor fields: radar-based VIL, composite reflectivity, echo tops, and accumulated precipitation; satellite radiances and derived cooling rate, cloud type and land use fields; RUC fields including relative humidity, CAPE, CIN, and winds; METAR-derived fields including convergence, lifted index, and relative humidity; CIWS feature detection fields including airmass area detection, weather type, boundaries, growth and decay, etc.; storm climatology data; distances to storm intensity contours; and local disc statics from selected data fields designed to capture information at various scales. The RF technique was used to rank the importance of

these candidate predictor fields, over 300 in all, for predicting various VIL levels under a variety of conditions (e.g., isolated initiation). A sample result in Figure 9a shows how the ranks of a subset of 45 fields vary for 26 days selected from the summer of 2007. Figure 9b, which depicts the observed frequency of VIL ≥ level 3 (equivalent VIL ≥ 3.5 kg m^{-2}) as a function of the number of RF votes, which demonstrates that the RF empirical model has considerable predictive skill. A set of RFs trained for various storm thresholds may be used to form a probability distribution over intensity levels, or to derive a deterministic prediction and uncertainty estimate. Thus, the RF technique may assist in identifying predictors, creating an empirical model that can be used as a benchmark, and providing a data fusion method for addressing forecast uncertainty.

7. SUMMARY

A research demonstration of 0-6 hour forecasts of VIL has been running since 17 July 2008. The forecasts are a blend between extrapolation and NWP forecasts and show promising skill at predicting aviation-specific content including storm mode and permeability structure. For the summer 2009 we plan to add blended forecasts of Echo Tops, extend both the VIL and Echo Tops forecasts out to 8 hours in the Northeast corridor, and provide these forecasts to key members of the operational ATC community. Beyond 2010 we plan to provide CONUS coverage with companion forecast error estimates for probabilistic use of the forecast information.

8. REFERENCES

Bellon, A. and I. Zawadzki, 1994: Forecasting of hourly accumulations of precipitation by optimal extrapolation of radar maps, *J. of Hydrol.*, **157**, 211-233.

Benjamin, S. G., T. G. Smirnova, S. S. Weygandt, M. Hu, S. R. Sahm, B. D. Jamison, M. M. Wolfson, and J. O. Pinto, 2009: The HRRR 3-km storm-resolving, radar-initialized, hourly updated forecasts for air traffic management. *AMS Aviation, Range and Aerospace Meteorology Special Symposium on Weather-Air Traffic Management Integration*, Phoenix AZ

Benjamin, S. G., G. A. Grell, J. M. Brown, T. G. Sminova, and R. Bleck, 2004: Mesoscale weather prediction with the RUC hybrid isentropic-terrain-following coordinate model. *Mon. Wea. Rev.*, **132**, 473-494.

Breiman, L., 2001: Random forests. *Machine Learning*, **45**, 5-32.

Dupree, W.J., M. Robinson, R. DeLaura, R. A. P. Bieringer, 2006: Echo Tops Forecast Generation and Evaluation of Air Traffic Flow Management Needs in the National Airspace System, *AMS 12th Conference on Aviation, Range, and Aerospace Meteorology*, Atlanta, GA.

Dupree, W.J., M.M. Wolfson, R.J. Johnson Jr., R.A. Boldi, E.B. Mann, K. Theriault Calden, C.A. Wilson, P.E. Bieringer, B.D. Martin, and H. Iskenderian, 2005: FAA Tactical Weather Forecasting in the United Stares National Airspace, *Proceedings from the World Weather Research Symposium on Nowcasting and Very Short Term Forecasts*. Toulouse, France.

Dupree, W.J., R.J. Johnson, M.M. Wolfson, K.E. Theriault, B.E. Forman, R.A. Boldi, and C. A. Wilson, 2002: Forecasting Convective Weather Using Multiscale Detectors and Weather Classification – Enhancements to the MIT Lincoln Laboratory Terminal Convective Weather Forecast. *AMS 10th Conference on Aviation, Range, and Aerospace Meteorology*, Portland, Oregon, 132-135.

Evans, J. E. and E. R. Ducot, 2006: Corridor Integrated Weather System. *Lincoln Laboratory Journal*, **16**, 59-80.

Evans, J., M. Weber and W. Moser, 2006: Integrating Advanced Weather Forecast Technologies into Air Traffic Management Decision Support, *MIT Lincoln Laboratory Journal*, v. 16, n. 1, pp. 81-96 (available for download at http://www.ll.mit.edu/mission/aviation/publications/publications.html)

FAA REDAC, 2007: "Weather-Air Traffic Management Integration Final Report," Weather – ATM Integration Working Group (WAIWG) of the National Airspace System Operations Subcommittee, Federal Aviation Administration (FAA) Research, Engineering and Development Advisory Committee (REDAC). 3 October 2007 (to be available at http://research.faa.gov/redac/)

Germann, U. and I. Zawadski, 2002: Scale-dependence of the predictability of precipitation from continental radar images. Part I: Description and Methodology. *Mon. Wea. Rev.*, **130**, 2859-2873.

Lakshmanan, V., R. Rabin, and V. Debrunner, 2003: Multiscale Storm Identification and Forecast, *J. Atm. Res.*, 367-380.

Matthews, M, M. Wolfson, R. DeLaura, J. Evans and C. Reiche, 2009: Measuring the uncertainty of weather forecast specific to air traffic management operations. *AMS Aviation, Range and Aerospace Meteorology Special Symposium on Weather-Air Traffic Management Integration*, Phoenix AZ

Mecikalski, J.R. and K.M. Bedka, 2006: Forecasting convective initiation by monitoring evolution of moving cumulus in daytime GOES imagery. *Mon. Wea. Rev.*, **134**, 49-78.

Megenhardt, D. L., C. Mueller, S. Trier, D. Ahijevych, and N. Rehak, 2004: NCWF-2 Probabilistic Forecasts. *AMS Eleventh Conf. Aviat. Range Aerospace Meteorol.*, paper 5.2.

NextGen ConOps, Joint Planning and Development Office, 2007: Concept of Operations of the Next Generation Air Transportation System, http://www.faa.gov/about/office_org/headquarters_offices/ato/publications/nextgenplan/resources/view/NextGen_v2.0.pdf

Roberts, R. D. and S. Rutledge, 2003: Nowcasting storm initiation and growth using GOES-8 and WSR-88D data. *Wea. Forecasting*, **18**, 562-584.

Robinson, M., W. Moser, and J. Evans, 2008: Measuring the Utilization of Available Aviation System Capacity in Convective Weather. *AMS 13th Conference on Aviation, Range, and Aerospace Meteorology (ARAM)*, New Orleans, LA.

Seed, A. W. and T. Keenan, 2001: A Dynamic Spatial Scaling Approach to Advection Forecasting. *30th International Conference on Radar Meteorology*, 19-24 July 2001, 492-494.

Wilks, D. S. 1995: *Statistical Methods in Atmospheric Sciences*, Academic Press, 467 pp.

Williams, J. K., D. A. Ahijevych, C. J. Kessinger, T. R. Saxen, M. Steiner and S. Dettling, 2008a: A machine-learning approach to finding weather regimes and skillful predictor combinations for short-term storm forecasting. *AMS 6th Conference on Artificial Intelligence Applications to Environmental Science and 13th Conference on Aviation, Range and Aerospace Meteorology*, paper J1.4.

Williams, J. K., D. Ahijevych, S. Dettling and M. Steiner, 2008b: Combining observations and model data for short-term storm forecasting. In W. Feltz and J. Murray, Eds., Remote Sensing Applications for Aviation Weather Hazard Detection and Decision Support. *Proceedings of SPIE*, 7088, paper 708805.

Wolfson, M.M., W. J. Dupree, R. Rasmussen, M. Steiner, S. Benjamin, S. Weygandt, 2008: Consolidated Storm Prediction for Aviation (CoSPA), *AMS 13th Conference on Aviation, Range, and Aerospace Meteorology*, New Orleans, LA, 2008.

Wolfson, M. M. and D. Clark, 2006: Advanced Aviation Weather Forecasts, *Lincoln Laboratory Journal*, Vol. 16, Number 1. 31-58.

Wolfson, M.M., B.E. Forman, R.G. Hallowell, and M.P. Moore, 1999: The Growth and Decay Storm Tracker, *AMS 8th Conference on Aviation, Range, and Aerospace Meteorology*, Dallas, TX, 58-62.