AutoNowcaster Pilot Evaluation Study

Prepared by:

Steve Lack[1,2], Matthew Wandishin[1,2], Missy Petty[1,3], and Jennifer Mahoney[1]

12 January 2011

Affiliations:
1 – National Oceanic and Atmospheric Administration, Earth System Research Laboratory, Global Systems Division (NOAA/ESRL/GSD), Forecast Impact and Quality Assessment Section
2- Cooperative Institute for Research in Environmental Sciences (CIRES) and NOAA/ESRL/GSD
3 – Cooperative Institute for Research in the Atmosphere (CIRA) and NOAA/ESRL/GSD

Corresponding Author:
J.L. Mahoney (NOAA/ESRL/GSD, 325 Broadway, Boulder, CO 80303;
Jennifer.Mahoney@noaa.gov)

Table of Contents

Executive Summary

This study summarizes the quality of the AutoNowcaster (ANC) forecast product, developed by the National Center for Atmospheric Research (NCAR), which was performed at the request of the National Weather Service NextGen Program. The ANC is a product being considered for use in aviation decision making.

The AutoNowcaster forecast product was evaluated with a focus on three main areas: the extrapolation component, comparison of the fully automated and human-in-the-loop (HITL) initiation forecast, and for its utility in an aviation-specific capacity, with specific focus on the Dallas Fort Worth Air Route Traffic Control Center (ZFW ARTCC) and adjacent areas. Data were gathered during the demonstration project for the period 19 April – 23 May 2010.

It must be emphasized that due to small sample sizes from which the results are drawn, all conclusions **must** be interpreted as indicators of forecast quality rather than definitive conclusions.

Initial indications of ANC quality from the study include:

- Pertaining to the extrapolation component, the ANC extrapolation forecasts require calibration. In order to eliminate the large underforecast bias the ANC extrapolation needed to be treated as a forecast of 43 dBZ or greater.
- When the ANC was treated as an extrapolation forecast of 43 dBZ echoes (thereby removing the forecast bias), the quality of the ANC relative to CIWS improved.
- Compared to the fully automated initiation forecast, human interaction (choosing the ANC regime and adding boundaries) improves the forecast quality by reducing the area of initiation 1 forecasts, while increasing both the number and size of embedded initiation 2 and initiation 3 area. The skill of the forecasts with human interaction remains comparable with those that were automated.
- About 10% of observed initiation is captured within a forecast region, 25% is within 30 miles, and 75% is within 120 miles. When an initiation forecast contains an initiation 2 or 3 area, the median distance is reduced by half, while the proportion of perfect forecasts is nearly doubled. In absolute terms, the human contribution is small (i.e., there is little difference in the continuous initiation potential fields between the HITL and the fully automated ANC forecasts); substantial differences occur when the field is discretized into initiation 1, 2 and 3 categories.
- Limiting the evaluation to the four low-altitude sectors containing the DFW arrival cornerposts, the increased skill possessed by forecasts with areas of higher embedded potential is once more apparent. The hit rate for

occurrence of convection within a sector is nearly double when an initiation 2 area is present when compared with an initiation 1 area only. This result holds for both the fully automated and HITL forecasts.

- Human input increases forecast association, measured as the correlation between time series of forecast and observed sector coverage, though the correlations for both the fully automated and HITL forecasts are very low.
- The nudging tool was used sparingly (~10% of all forecasts), but the nudged forecasts did improve slightly upon the HITL forecasts. The correlations are substantially higher, though still not strong, when the sector coverage time series are converted to binary series of the presence or absence of convection.

Further analysis of the ANC should continue over a longer period of time and for different meteorological regimes.

1. Introduction

An objective assessment of the AutoNowcaster, developed by the National Center for Atmospheric Research, was conducted by the Forecast Impact and Quality Assessment Section (FIQAS) of the Earth System Research Laboratory of NOAA at the request of the National Weather Service (NWS). In this study, the quality of the ANC for 3 main questions are examined: 1) does the ANC provide an increase in forecast quality when compared to the operational baseline represented by the Corridor Integrated Weather System (CIWS), 2) how accurate is the convective initiation field, and does the forecast quality improve when boundaries and regime changes are introduced into the forecast, and 3) does the ANC provide helpful information to the aviation planner at the Center Weather Service Unit (CWSU) Air Route Traffic Control Center (ARTCC) in Dallas Fort Worth (FWD).

The assessment of ANC was performed during the period from 19 April-23 May 2010 and facilitated by the NWS and the Meteorological Development Laboratory (MDL). During this period, forecaster representatives (FRs) were assigned to produce convective forecasts using the ANC by interacting with the forecast boundaries and forecast regimes. All forecasts produced by the ANC during this period were collected and analyzed. In addition, detailed analyses of the boundary and regime modifications as well as the role of the forecaster in the ANC forecast process was assessed.

In addition to the objective assessment, a subjective assessment was also performed, including an assessment of the forecast process and use of the ANC tool for generating forecasts. Only the results from the objective assessment are provided in this report. In addition, this objective assessment provided the opportunity to develop new verification techniques that were used specifically for the convective initiation assessment.

Note: It is imperative for the reader to remember that the results presented here are to be considered initial indicators of forecast accuracy only. Further assessments are needed before definitive conclusions can be determined.

Section 2 provides a summary of the data used in the study. The approach is provided in Section 3, the verification methodologies summarized in Section 4, the results listed in Section 5, and conclusions provided in Section 6. Section 7 summarizes an alternative convective initiation detection scheme, while references are listed in Section 8.

2. Data

Data for the study included the ANC and the Corridor Integrated Weather System (CIWS) forecasts and radar data as the 'truth' field. Over 5000 ANC forecasts were issued during the period from 19 April-23 May 2010.

2.1 ANC

The ANC is unique among short-term forecasts (also known as nowcasts). It provides both a pseudo-reflectivity convective nowcast that is extrapolated from the existing convection as well as a field indicating the likelihood of convective initiation (CI). The product is issued approximately every 5 minutes with each issue containing a forecast valid 1-h from issuance. The forecasts combine extrapolated radar information with a fuzzy logic algorithm. The fuzzy logic algorithm incorporates a variety of meteorological information such as radar, satellite, soundings, surface stations, and numerical weather prediction model output with conceptual models to produce a short-term forecast of convective initiation, growth, and dissipation. The ANC provides an extrapolation forecast as well as a convective initiation potential field. The extrapolation field provides the location of thunderstorms at reflectivity values of 35 dBZ or greater. The convective initiation potential consists of a smoothed initiation field generated from overlapping areas of convective instability and forcing. Threshold values are applied to the CI field resulting in three increasing levels of initiation potential. As the level of CI potential increases from values of 1 to 3, the likelihood that the convection will occur is also expected to increase. Each CI potential level is evaluated. More information on the ANC system can be found in Mueller et al. (2003).

The ANC is produced over a domain covering most of TX along with portions of the neighboring states (Fig.1, blue box). However, to limit the influence of the boundary on the forecast quality, the ANC and the CIWS forecasts are assessed in the smaller domain identified by the red box in Fig. 1. The red box was derived by applying a 40-km buffer to the area inside the larger domain.

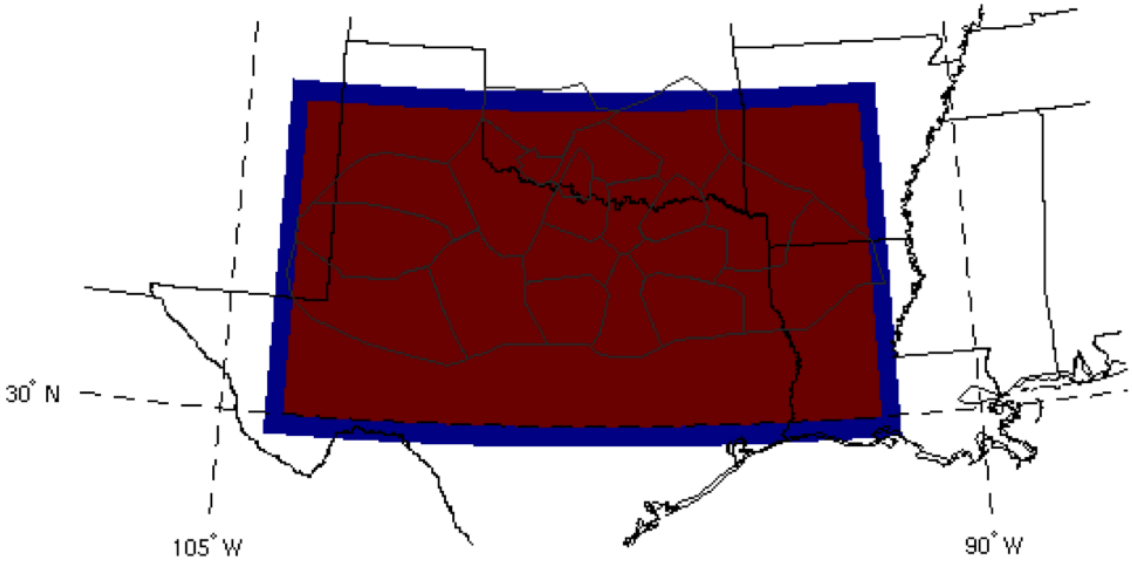

Fig. 1. *The ANC verification domain,40-km buffer applied shown in red. The full forecast domain shown in blue. ZFW sectors shown in gray.*

2.2 CIWS Forecast

The Corridor Integrated Weather System (CIWS) developed by Massachusetts Institute of Technology Lincoln Laboratory (Evans and Ducot 2006) provides a 0-2 h forecasts of vertically integrated liquid (VIL) and echo tops over the continental United States that include cell growth and decay. Unlike ANC, CIWS does not forecast the initiation of new thunderstorms. Since the CIWS forecast is currently the operational standard, the 1-h CIWS forecasts were used in this study to baseline the quality of the ANC 1-h forecast.

2.3 Radar

The radar data is used to evaluate the quality of the ANC and CIWS forecasts. The radar data used in this study is derived from the several Weather Surveillance Radar-1988 Doppler Radars. The data files used to evaluate the ANC intersect the ANC forecast domain and are available at roughly six-minute intervals. The time between successive radar files typically vary by a few minutes depending on the radar-scanning strategy. The radar data is also an input to the ANC. However, the radar files used for verification of the ANC are not the same files used to build the forecast, thus providing an independent dataset for verification.

3. Approach

The quality of the ANC was assessed in an environment that was similar to how it might be used in NWS operations. Thus, a 'forecaster-intensive' period was defined over a 5 week period (19 April–23 May 2010) where forecaster representatives (FR) were expected to produce hourly convective forecasts using the ANC within the Dallas/Ft Worth domain (identified in Fig. 1). Four meteorologists from MDL and 1 forecaster from the NWS Headquarters were assigned to this activity. The FR's were required to produce convective forecasts using the ANC and, when needed, adjust the automated forecasts with human-generated boundaries and convective regime modifications to produce more refined convective forecasts. These adjustments were provided in one of the 3 forms: a default mixed-regime with no HITL, HITL with regime changes and boundary inputs, and HITL with regime changes, boundary inputs and nudging. One FR was dedicated to this forecast activity per week. Each FR was trained on the ANC system by MDL staff prior to the start of the evaluation.

The quality of ANC is assessed from 3 main perspectives:

- The meteorological accuracy of the ANC extrapolation forecast of reflectivity values greater than 35 dBZ as compared to the operational baseline, which for this study is CIWS.

- The meteorological accuracy of the convective initiation field for 3 variants of the ANC forecast: the fully automated mixed-regime; human-modified regimes and boundary additions; and human-modified regime, boundary modifications, and use of a polygon nesting tool to alter the entire initiation potential field. (Although the forecast variation in which the human used the polygon-nudging tool will be included in the analysis, this tool was employed in only 10% of all forecasts, primarily during late April.)

- The quality of the forecast in the context of CWSU ARTCC operations, where the forecast coverage within the low-altitude sectors containing the four en-route air traffic arrival corner posts is evaluated.

Overall forecast quality was assessed and specific case studies were analyzed.

4. Methodology

4.1 Forecast Matching Technique

The forecast matching technique is used when the 1-h ANC extrapolation forecasts are compared to the 1-h CIWS forecasts. To perform the matching, the forecast time interval, map projection, and data values are matched spatially and temporally as summarized below.

The CIWS forecast is produced every 5 min at consistent 5-min intervals. The ANC forecast is issued roughly every 5-min, but the interval at which it is provided often varies by a few minutes. Therefore, in order to time match the ANC and CIWS forecasts, the CIWS forecasts, at 15 minute intervals, are paired with the nearest-in-time ANC extrapolation forecasts, with a maximum allowable time gap between the two forecasts of 2.5 min. The 15-min interval selected for evaluation was sufficient to capture changes in the forecast while also alleviating storage constraints that occurred by storing the large 5-min CIWS files. .

The ANC, CIWS, and the radar observation grids each possess different native map projections. The ANC is at roughly a 2-km resolution while CIWS and the radar data have a 1-km resolution. To facilitate the comparisons, the CIWS and radar data are mapped onto the ANC grid using a maximum composite mapping, wherein each grid cell is assigned the value of the highest reflectivity within the cell volume.

CIWS produces forecasts of vertically integrated liquid (VIL), while the ANC forecasts reflectivity. Therefore, the CIWS VIL values are mapped to dBZ values using a logarithmic fit-lookup table provided by Troxel and Engholm (1990) to match them to the ANC and radar observation values. In addition, multiple reflectivity thresholds are applied to the CIWS forecasts and radar observations to calibrate the forecasts (through bias correction) during the comparison. Traditionally, a bias correction is obtained by adjusting the forecast threshold until the forecast area matches the observed area. Because the ANC field does not contain values below 35 dBZ, it was necessary to adjust the threshold of the radar observations to match the forecast area.

4.2 Convective Initiation Technique

Evaluation of the ANC convective initiation potential field presents unique challenges. The goal of this section of the study is to answer the following two-part question: Is there a correspondence between the ANC convective initiation forecast and observed convective initiation? If so, how close are the observed cells to the forecast areas?

Traditional verification metrics, however, are insufficient in this context as they usually require some final form of the forecast and observation field for comparison. While the ANC is one of the few forecasts that distinguish growing and decaying convection from newly initiated convection, observations clearly do not make this distinction. The challenge in evaluating the ANC forecast in the context of CI is in establishing an objective identification of initiation in the observation.

The approach to classify CI is described for observations; however, this approach is applied to the ANC forecasts of sufficient temporal and spatial resolution. At the highest spatial and temporal resolution of the observations (which is determined by radar scan frequency) the classification of CI is accomplished using object-oriented verification approaches. In general, the object-oriented verification scheme identifies objects in the observation field and compares it to that of a forecast with similarly defined objects. For the context of initiation, the forecast is replaced with the observation field from the previous time step. The verification scheme then matches like objects and assigns a penalty based on minimizing the cost function of all possible matches. In the end, some objects in the current time step may not match any objects from the previous time step. These unmatched objects are considered initiation if certain requirements are met. Namely, the new object must be located a minimum distance from existing objects (to distinguish initiation from growth on the edge of a system) and its intensity must surpass a minimum threshold. Other requirements are included for quality control; for example, the new object must exceed a minimum size to ensure that a pre-existing storm suddenly appearing after a radar outage is not labeled as initiation.

4.3 Procrustes Object-oriented Technique

The Procrustes technique (Lack et al., 2010) is the object-oriented verification approach used to identify CI in this study. Figure 2 gives an example of how the Procrustes technique can identify initiation. Using the Procrustes technique, two different intensity thresholds are used to detect initiation in the observation, 25-dBZ and 30-dBZ. This means that if an echo that was 24 dBZ (29 dBZ) in the previous timestep grew to 25 dBZ or above (30 dBZ or above), the CI detection scheme would register a new initiation object. Using both thresholds, that is, requiring echoes to grow from less than 25 dBZ to 30 dBZ or greater in one time step, allows for the examination of significant convection.

Once initiation objects are identified in the observation field, the accuracy of the initiation forecast is determined by the minimum edge-to-edge distance between forecast and observed initiation objects. A window is applied around the forecast valid time to examine the temporal sensitivity of the CI potential. Finally, in addition to the comparison between forecasts of any convection and only significant convection, forecast objects are evaluated as a function of the maximum CI potential

within the object.

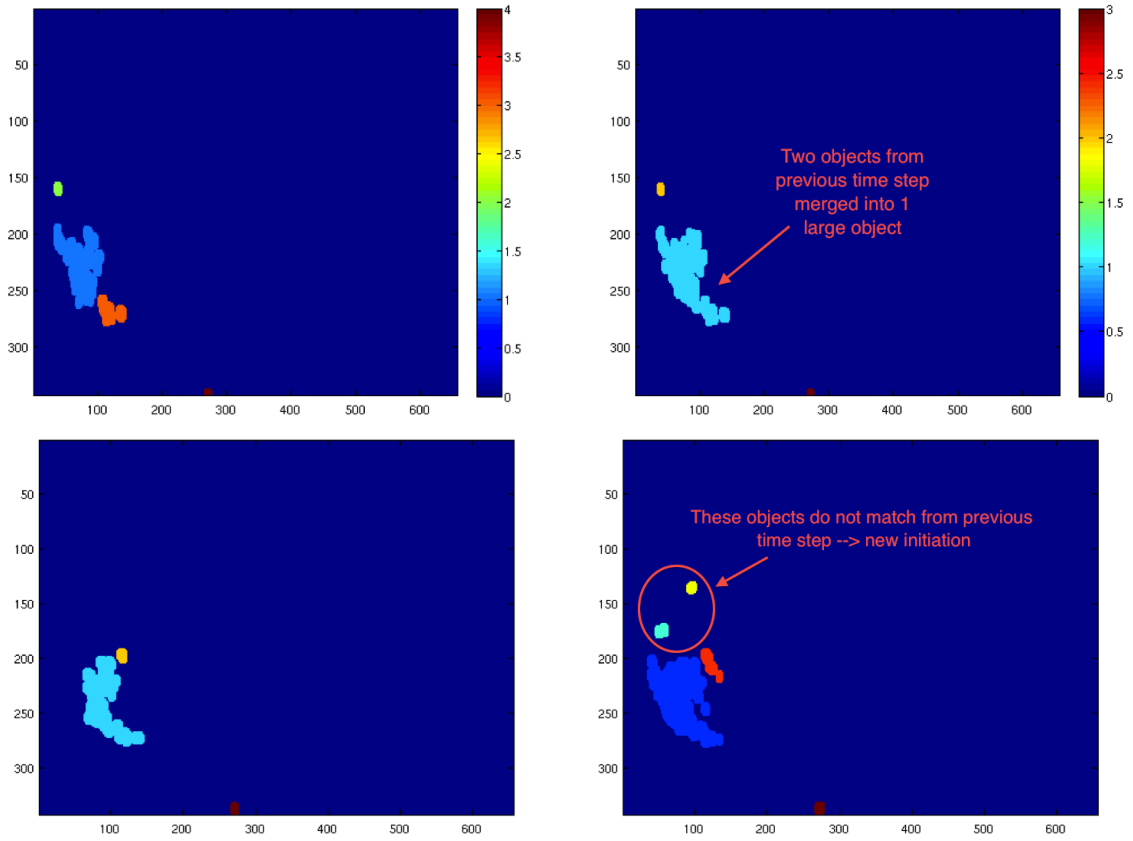

Fig. 2. *Example of consecutive time stamps with no initiation cells present (top) and consecutive images with initiation (bottom). No initiation was detected in the top image although merging present. Two new cells initiated in the bottom image with considerable growth of the existing convection.*

4.4 Operational Context

A primary aviation user of the ANC forecasts is the Dallas/Fort Worth CWSU forecaster, providing support for Dallas/Fort Worth (ZFW) air traffic planning. ZFW planners continuously balance the air traffic flow through aviation sectors with the amount or coverage of convective weather occurring in the sector. Higher levels of convective coverage in sectors impact the arrival rates at airports and cause airport delays and flight cancellations. Therefore, accurate forecasts of convective initiation at 1-h leads and indications of convective 'density' or coverage in critical air traffic sectors would be useful to the air traffic planner.

Therefore, the objective of this assessment is to measure the trends in sector coverage above 25 dBZ with the presence of an initiation potential field. The results are stratified by CI threshold. In addition, the quality of the forecast and the forecaster confidence for embedded CI regions was also examined. The ANC analysis is targeted on the arrival corner posts in ZFW shown in Fig. 3 and includes: Glen Rose (JEN), Cedar Creek (CQY), Bowie (UKW), and Bonham (BYP).

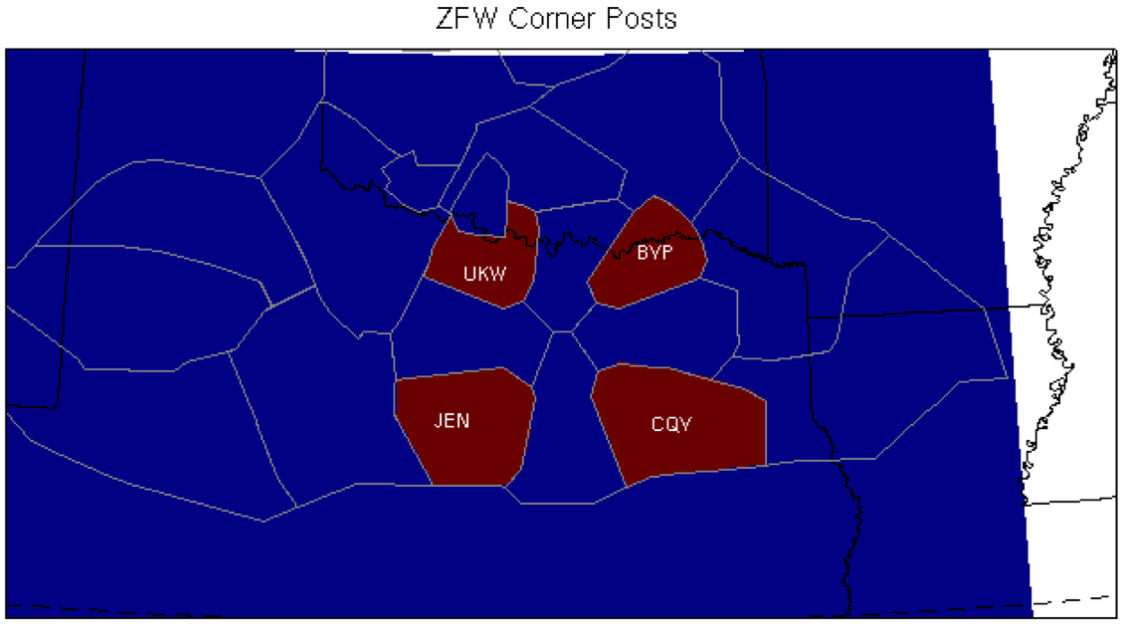

Fig. 3. *Arrival corner posts in ZFW serve as the basis for an aviation centric evaluation.*

5. Results

5.1 ANC and CIWS Extrapolation Comparison

Figure 4 shows box-and-whisker plots for bias for the ANC (right) and CIWS (left). A box-and-whisker plot represents a distribution of values: the box denotes the interquartile range (IQR; i.e., 25th to 75th percentiles), while the notches represent roughly a 95% confidence interval for differences between the medians of the two distributions. The whiskers extend 1.5 times the length of the box, with all data points further than this from the median drawn as circles. A log base 2 scale is used to illustrate the degree of bias difference between the two products (e.g., an overforecast bias of 2 is equidistant from the perfect bias (=1) line as an underforecast bias of 1/2).

The results indicate that the ANC and CIWS extrapolation forecasts possess a substantial bias: the ANC tends to significantly underforecast convection at its defined level of 35 dBZ, while the CIWS overforecasts convection at this threshold. In addition, as shown by the distances of the median from the 1.0 line, it appears that the ANC requires more 'tuning' to arrive at a bias of 1.0 than does CIWS product. When the forecasts are bias corrected by adjusting the observation threshold to above 40 dBZ (i.e., a less common event, and thus more difficult forecast), the bias gap between the two products is noticeably reduced.

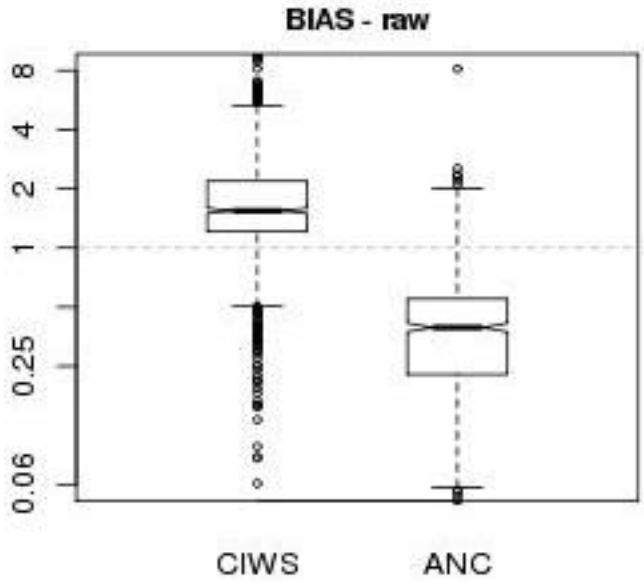

Fig. 4. *Boxplot of Bias for the CIWS product (left) and the ANC extrapolation forecast (right) at 35 dBZ for the full study period on a $\log_2$ scale. Notches show an estimate of statistical significance.*

Figures 5 shows the Critical Success Index (CSI) for the products at the 35-dBZ threshold with a corresponding cumulative distribution function (CDF). The CSI is a measure of accuracy that rewards correct forecasts (hits) while punishing both false alarms and missed events. The CDF presents all of the forecasts for which there was no observed convection anywhere within the domain, arranged by the size of the forecasted convective areas. These cases are not included in the CSI calculations because they can result in division by zero. The results in Fig. 5a indicate that the skill for the ANC at the 35-dBZ threshold is lower than the skill of CIWS at the 35-dBZ threshold. However, the ANC is less likely to have extrapolated convection in cases in which no convection is observed as shown in Fig. 5b (the red dots begin with a value of roughly 0.97, indicating that for 97% of all non-convective events the ANC correctly forecast no convection). Figure 5b also shows that for both ANC and CIWS, forecast objects are small when no observed convection is present. That is, these false alarms are unlikely to draw undue attention from forecasters.

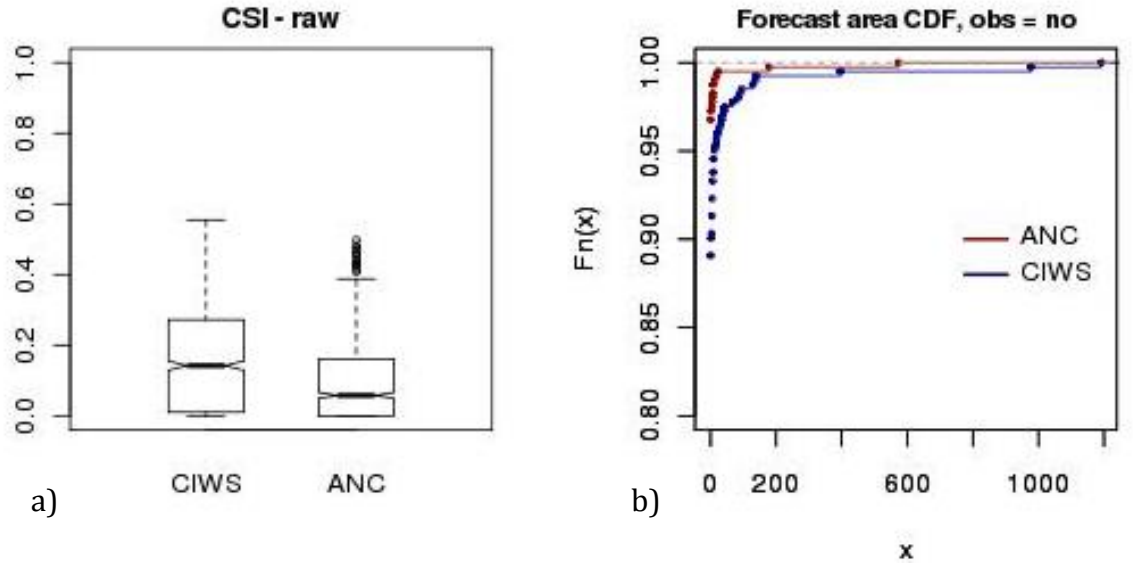

Fig. 5 a,b. *CSI (a); observation and forecast thresholds for CIWS (right) and ANC (left) are 35-dBZ; corresponding CDF (b)) for ANC (red) and CIWS (blue).*

As previously stated, correcting for the forecast biases provide a more meaningful and direct comparison of forecast quality between the two forecasts. Following the bias-correction method outlined in Section 4.1, the forecast and observation thresholds were adjusted in three ways: 1) observation and forecast thresholds for CIWS and ANC are 35 dBZ (Figs. 4,5), 2) observation threshold is 30 dBZ and CIWS threshold is 35 dBZ; observation threshold is 43 dBZ and ANC threshold is 35 dBZ (Figs. 6,7), and 3) observation threshold of 43 dBZ and CIWS threshold of 46 dBZ, observation threshold of 43 dBZ and ANC threshold of 35 dBZ (Fig. 8). Adjustment 2 achieves bias correction but at the expense of much different event thresholds and thus event frequencies. Adjustment 3 equalizes both the bias and event frequency, allowing for a direct comparison between the two forecast products.

The results shown in Fig. 6 indicate that when the bias correction is applied according to adjustment 2, the ANC retains a larger interquartile range than the CIWS product reflecting a greater variability in the forecast coverage relative to the observed coverage. The CSI values for CIWS are nearly double those of the ANC (Fig. 7a); this result is not unexpected as the ANC is forecasting a rarer event, an inherently more difficult task. The increase in frequency of total false alarms (i.e., convection is forecast but not observed anywhere in the domain) for the ANC forecasts relative to the unadjusted forecasts (from roughly 3% to roughly 7%) is expected from the adjustment to the higher, less common observed threshold. The even larger increase in total false alarms for the CIWS forecasts (from roughly 11% to 20%) is exactly opposite of what would be expected. Time constraints did not permit an investigation into the cause of this result. When the bias is corrected as in

Adjustment 3 (Fig. 8), so that the two systems are forecasting the same events, the skill of the two forecasts is more similar than for the other cases. However, both forecasts show modest accuracy, as indicated by the median of the box plots, of around 0.1 CSI (Fig. 8a). Slight improvement in skill for the CIWS over the ANC is noted. At this higher threshold, the number of total false alarms is now similar for both forecast products, and the forecast coverage of those false alarms is reduced (Fig. 8b).

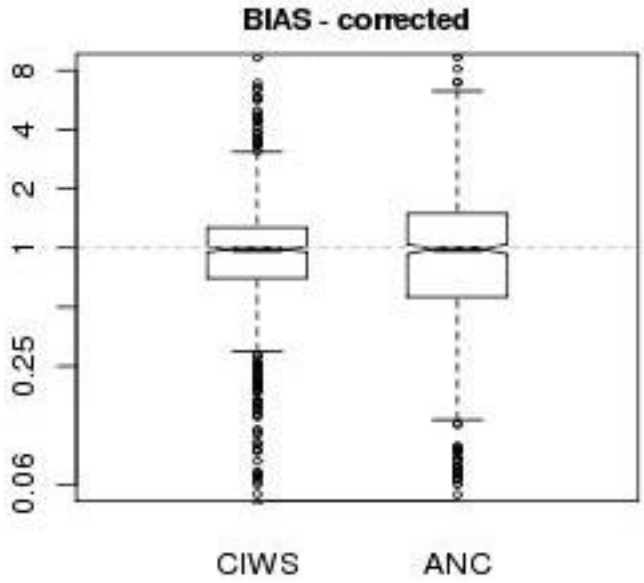

Fig. 6. *Boxplot of Bias after bias correction is applied. Observation and CIWS threshold is 30-dBZ; observation and ANC threshold is 43 dBZ (rarer event).*

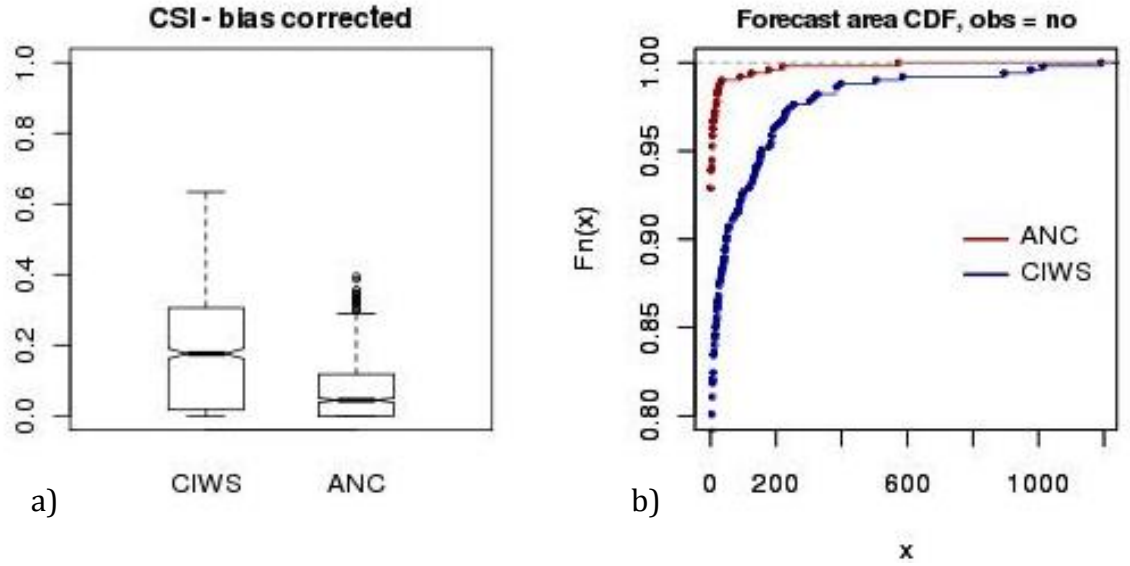

Fig. 7 a,b. *CSI (a); observation threshold of 30 dBZ and CIWS threshold of 35-dBZ; observation threshold of 43 dBZ and ANC threshold of 35-dBZ; corresponding CDF (b) for ANC (red) and CIWS (blue).*

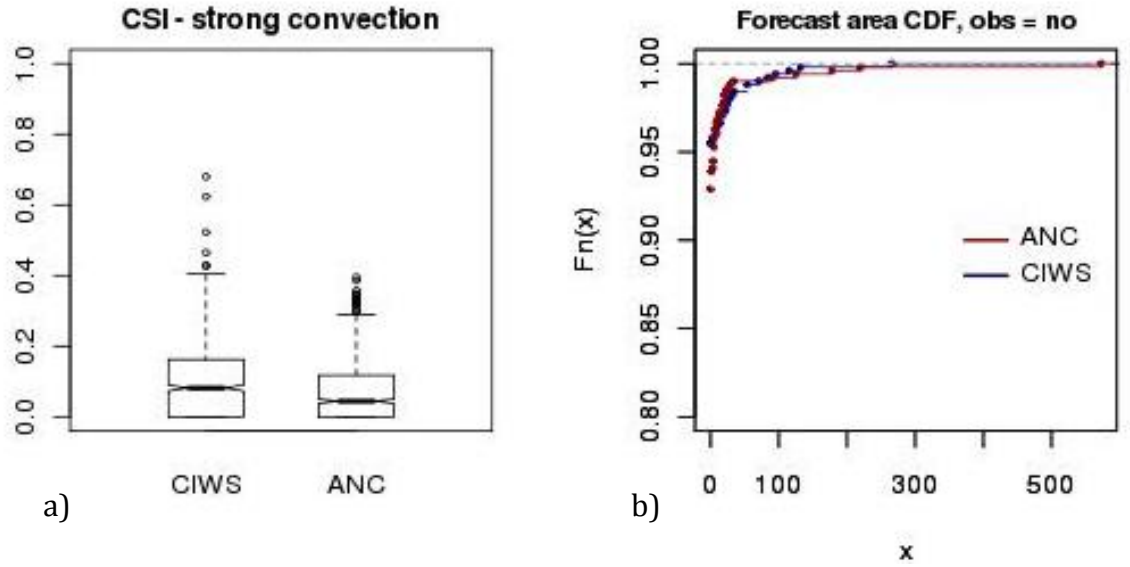

Fig. 8 a,b. *CS(a);, observation threshold of 43 dBZ and CIWS threshold of 46-dBZ, observation threshold of 43-dBZ and ANC threshold of 35 dBZ, corresponding CDF (b) for ANC (red) and CIWS (blue).*

5.2 Convective Initiation Investigation

The methodology used to evaluate the convective initiation potential field is outlined in Section 4.2. The fully automated ANC product in the mixed-regime mode with no HITL is compared to the ANC product with the HILT and without polygon nudging applied. Results are presented for all convective initiation (reflectivity increases to above 25 dBZ in successive radar scans) and for significant initiation only (reflectivity increases from less than 25 dBZ to 30 dBZ or greater in successive radar scans).

Figure 9 shows a time series plot of the areas of convective initiation potential for the fully automated ANC product running in mixed-regime mode without HITL (top) as compared to the ANC with HITL (bottom). Results in Fig. 9, and confirmed by Table 1, indicate that forecast areas from the fully automated ANC forecast are refined when HITL is applied. For instance, convective initiation 1 regions are substantially larger in the automated version without HITL than in the HITL version. In addition, the initiation 2 regions (areas with a higher likelihood for convection initiation) are larger in the ANC with the HITL than the ANC without the HITL, possibly suggesting that the human is adding confidence to regions of higher likelihood for convective initiation.

While not identical the time series (Fig. 9) of areal coverage of CI potential from the fully automated and HITL forecasts are very similar, possibly due to the dependencies on the background RUC field and similar instances where no initiation field is present but initiation cells are detected. In both forecasts, there seem to be few instances of initiation detection without an initiation potential field present somewhere in the domain (i.e., there are few domain-wide misses).

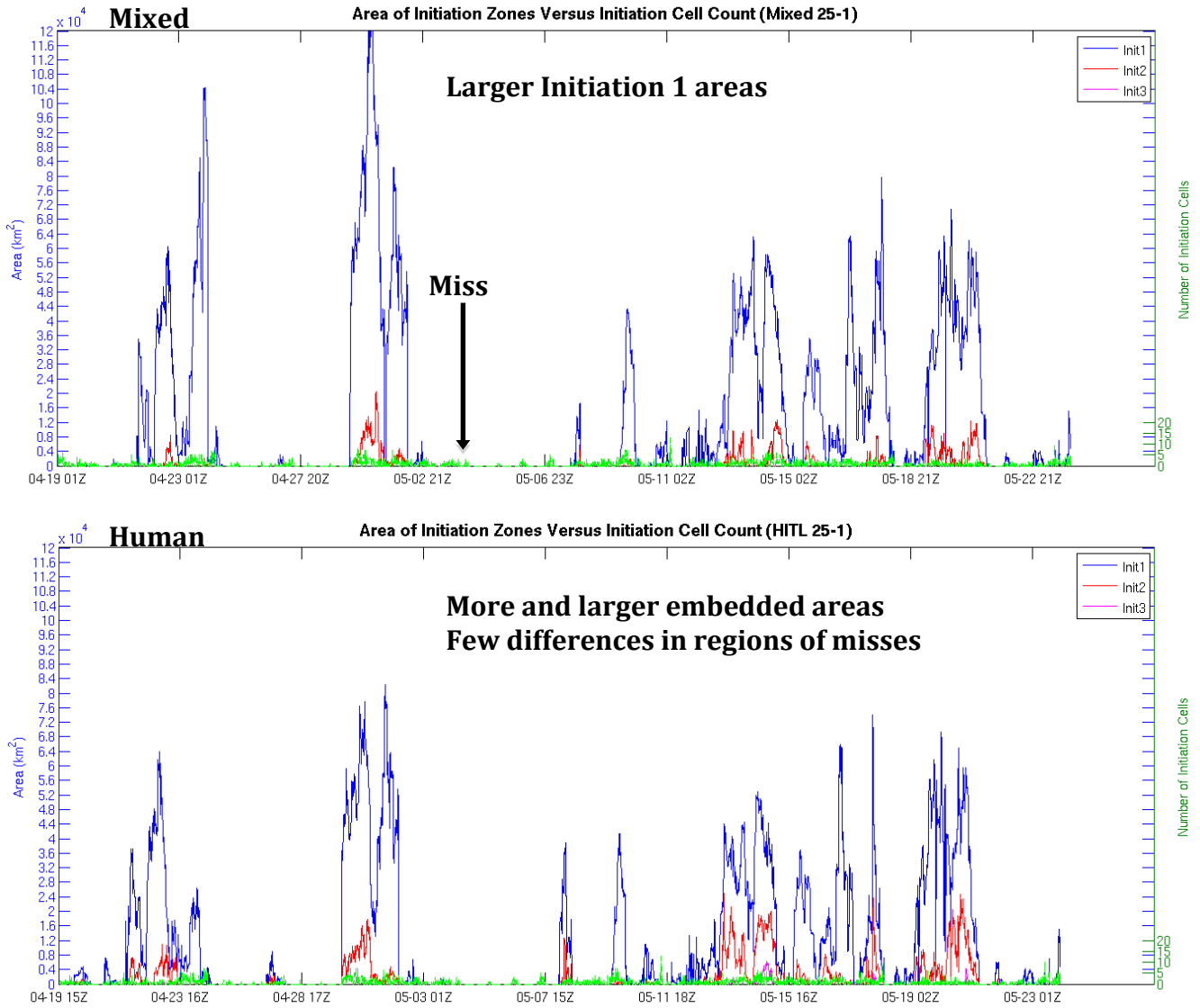

Fig. 9. *Areas of initiation potential fields for the fully automated ANC product running in mixed-regime mode (above) compared to the HITL ANC product (below) for the study period. Initiation 1 blue, initiation 2 red, and initiation 3 magenta. The green line is number of initiation cells detected in the domain.*

Table 1. Number and average size (km^2) of initiation objects by initiation threshold.

		Mixed	Human
Init_1	count	2383	2591
	area	25840	20570
Init_2	count	1144	1618
	area	3060	4984
Init_3	count	0	267
	area	0	1480

To further assess the accuracy of the CI potential field, edge-to-edge distance from observed initiation object to observed forecast initiation potential object is examined.

Figure 10 shows the distribution of observed distances away from any initiation object throughout the forecast period for the three variations of the ANC forecast which include: HITL no nudging, mixed-regime automated, and HITL nudging allowed. The distances were computed for any initiation (i.e., 25 dBZ) and significant initiation (i.e., 30 dBZ). If an observed initiation cell is coincident with a forecast of initiation, the distance is set to 0 and is considered perfect. In Fig. 10 it can be seen that significant initiation cells (represented in the diagram on the right) tend to be closer to the initiation potential regions. The differences in distance away from initiation forecast are not significant between HITL forecasts and fully automated ANC forecasts without HITL. The results were computed for a variety of time windows and initiation detection thresholds and were shown to be insensitive to either. The lack of sensitivity of the time window may come from the use of the RUC and its 1-h updates (low temporal resolution) as opposed to the ANC 5-min update cycle.

When the location of CI forecast areas are compared to the location of the observations for any initiation, 25% of the observed initiation cells fall within 30 miles of a forecast convective initiation region. When higher initiation forecast regions are present within an area of initiation 1, the distance to an observed initiation cell is reduced by nearly half. Higher intensity regions are typically larger and appear to be more prevalent in HITL ANC forecasts when compared to the fully automated version with no HITL.

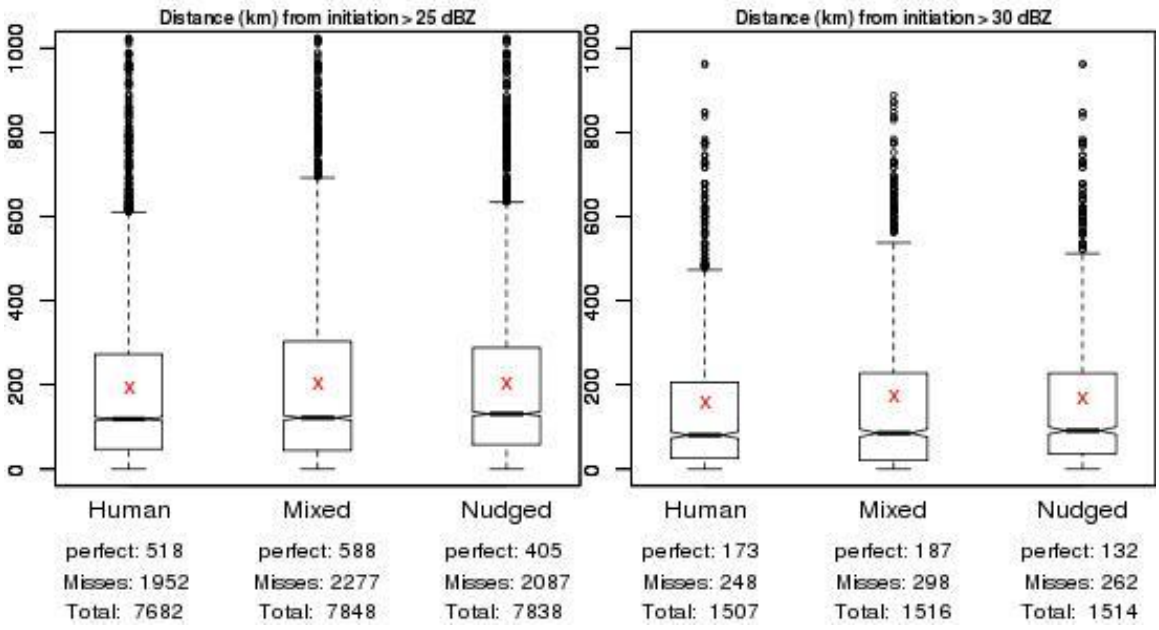

Fig. 10. *Comparison of the any initiation (left) and significant initiation (right) using a +/- 5 minute time window. Boxplots of the distance away (km) from any initiation forecast object for the three variations of ANC forecasts are shown, HITL no nudging (left), mixed-regime automated (center), HITL nudging allowed (right). Total initiation cells detected, total misses, and perfect forecasts are noted in the text.*

Figure 11 shows the sensitivity of the distance between forecast and observed initiation objects to the presence of nested ANC convective initiation objects (objects where a higher confidence area is embedded in an init 1 area). It can be noted that the distance decreases significantly when convective initiation nests exist in the forecast initiation potential field. This appears to show that a user may gain confidence in a forecast by focusing on initiation zones with increasing nests of CI likelihood. Evidence also comes from the fact that initiation cells are always detected in the observation field somewhere in the domain if a nested potential field is present in the forecast.

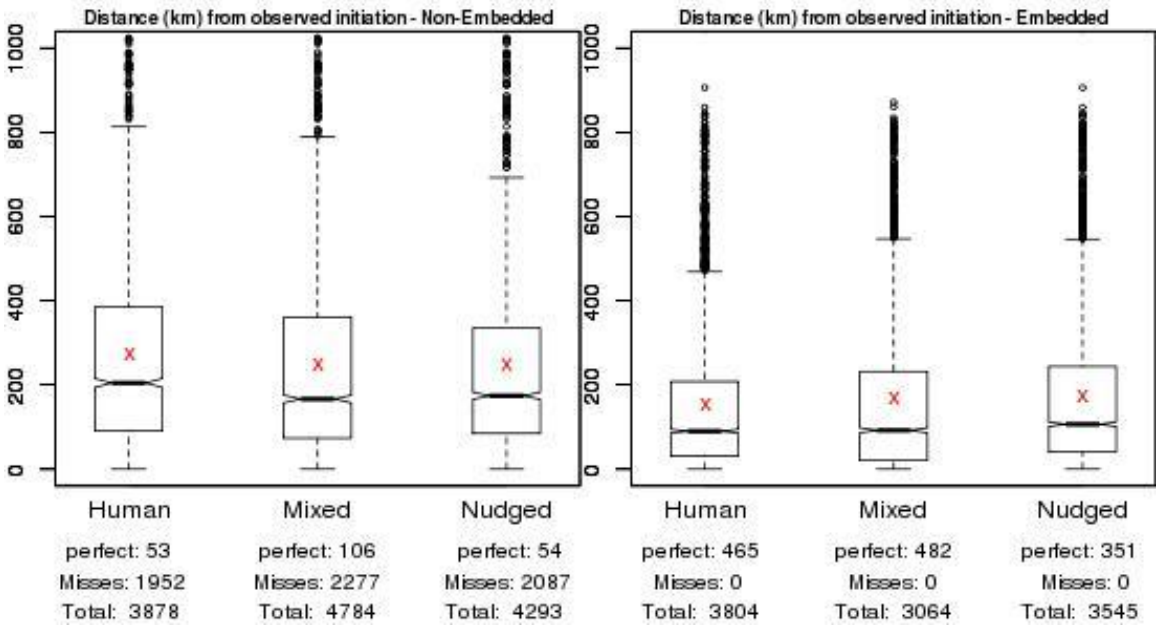

Fig. 11. *As in Fig. 10, but for forecasts with initiation potential 1 field only (left) and embedded areas of higher initiation potential (right).*

5.3 Application to Aviation Operations

In the four corner-post sectors examined, a higher confidence area embedded within an initiation 1 area does signal an, often substantially, greater likelihood of convective initiation within the sector. Examination of time series of forecast and observed sector coverage indicates that the HITL appears to be adding value in terms of the temporal correlation between forecasts and observations over the fully automated ANC setting. Remember that these results provide only initial indicators of forecast performance. Further investigation is needed.

Within each of the four ZFW low-altitude sectors containing a DFW-arrival corner post (Fig.3), the sector coverage of observed reflectivity greater than 25 dBZ is stratified according to the degree of confidence provided by the ANC forecasts. For the fully automated, mixed-regime forecasts (Fig. 12), the hit rate when an initiation 2 area is present is about double the hit rate obtained when only the initiation 1 threshold is reached. (The hit rate is defined here to mean percentage of all forecasts for which there was also convection observed *somewhere in the sector*, i.e., forecast and observation need not overlap.) The greater average—both mean (red 'X') and median (thick horizontal line)—sector coverages for the initiation 2 forecasts compared to the initiation 1 forecasts are the result of fewer null, or 'no', forecasts of higher likelihood.

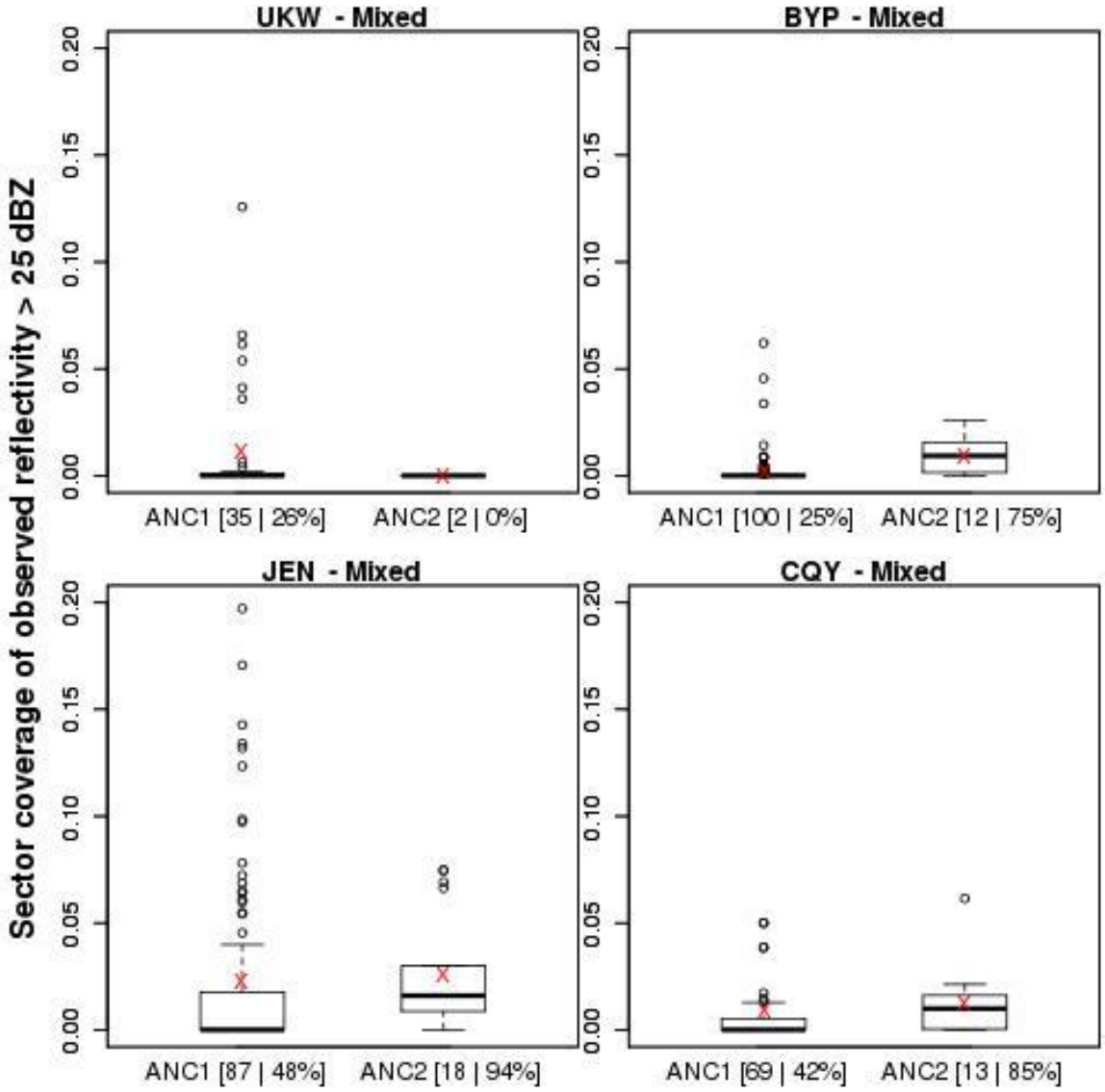

Fig. 12. *Box plots of observed sector coverage of greater than 25-dBZ reflectivity for the ZFW low-altitude sectors containing the UKW (upper-left panel), JEN (lower-left), BYP (upper-right), and CZY (lower-right) corner posts, stratified by the highest-confidence forecast within the sector for the fully automated ANC forecasts. The numbers in each panel denote the number of forecasts of each confidence level and the hit rate corresponding to those forecasts.*

The HITL forecasts (Fig. 13) follow a similar pattern to the fully automated forecasts, namely a substantial increase in the hit rate for initiation 2 forecasts when compared to forecasts of initiation 1. That is, the greater confidence (i.e., higher potential) of the forecast is substantiated by a greater likelihood of occurrence. As with the fully automated forecasts, there is no clear pattern of greater coverage density for the initiation 2 forecasts over the initiation 1 forecasts. Similarly, comparing the hit rates for the HITL and the fully automated forecasts yields a

mixed picture; neither forecast clearly dominates the other. As seen in Section 3.2, however, adding the human to the forecast process increases the prevalence of initiation 2 forecasts.

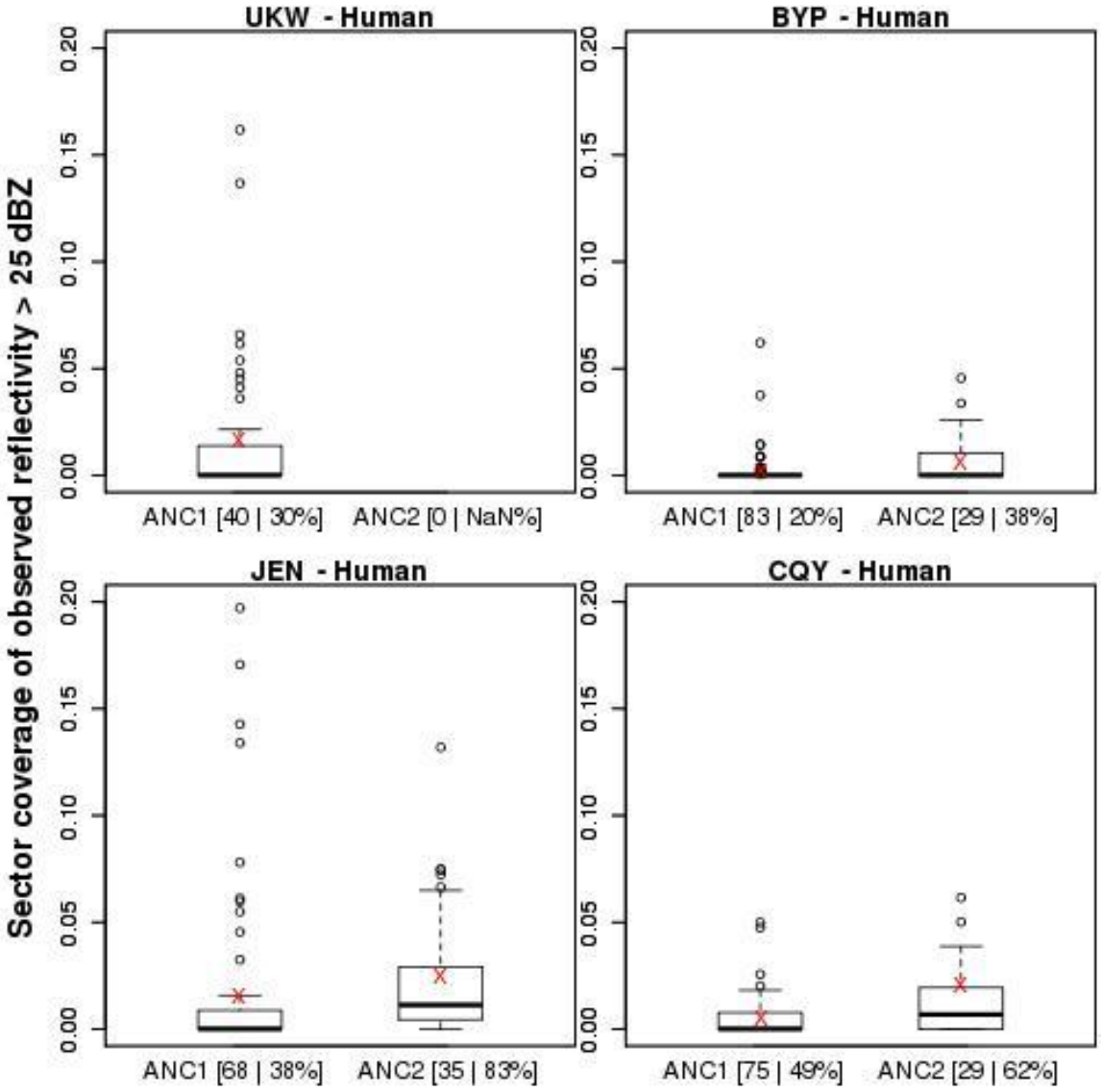

Fig. 13. *As in Fig. 12, but for the HITL forecasts.*

The forecast sector coverage also can be examined as a time series through the duration of the study and correlated with the time series of observed sector coverage (Fig. 14). The time series highlight the fact that, while this study consists of about 5000 separate forecasts, there are very few meteorological events being considered. In addition, recall that when considering initiation over the whole domain (Fig. 9) there are a few missed events, but no false alarms. Narrowing focus on the low-altitude sectors (Fig. 14), one sees not only missed events (e.g., 3 May) but also false alarms (e.g., 30 April). Since there were no false alarms over the whole domain these low-altitude sector false alarms must be caused by location errors by the ANC, i.e., locating the initiation in the wrong sectors. Correlations are low for all regions, but the HITL doubles the success of the fully automated forecasts. This pattern is consistent whether one considers total sector coverage, sector coverage of initiation 1 only, of initiation 1 only when initiation 2 is also present, or simply presence or absence of initiation (Fig. 15). The low-altitude sector containing the UKW corner post is an outlier in terms of having lower overall correlations and is the only sector in which the fully automated product outperforms the HITL. From the time series plot (Fig. 14) it can be seen that initiation is forecast much less often in this region, likely amplifying the impact of small differences between the forecasts. The nudged forecasts perform similarly to the HITL forecasts (Fig. 15), which is to be expected since the nudging was used in only about a tenth of the forecasts. Most of the time, then, the nudged forecast is identical to the HITL forecast.

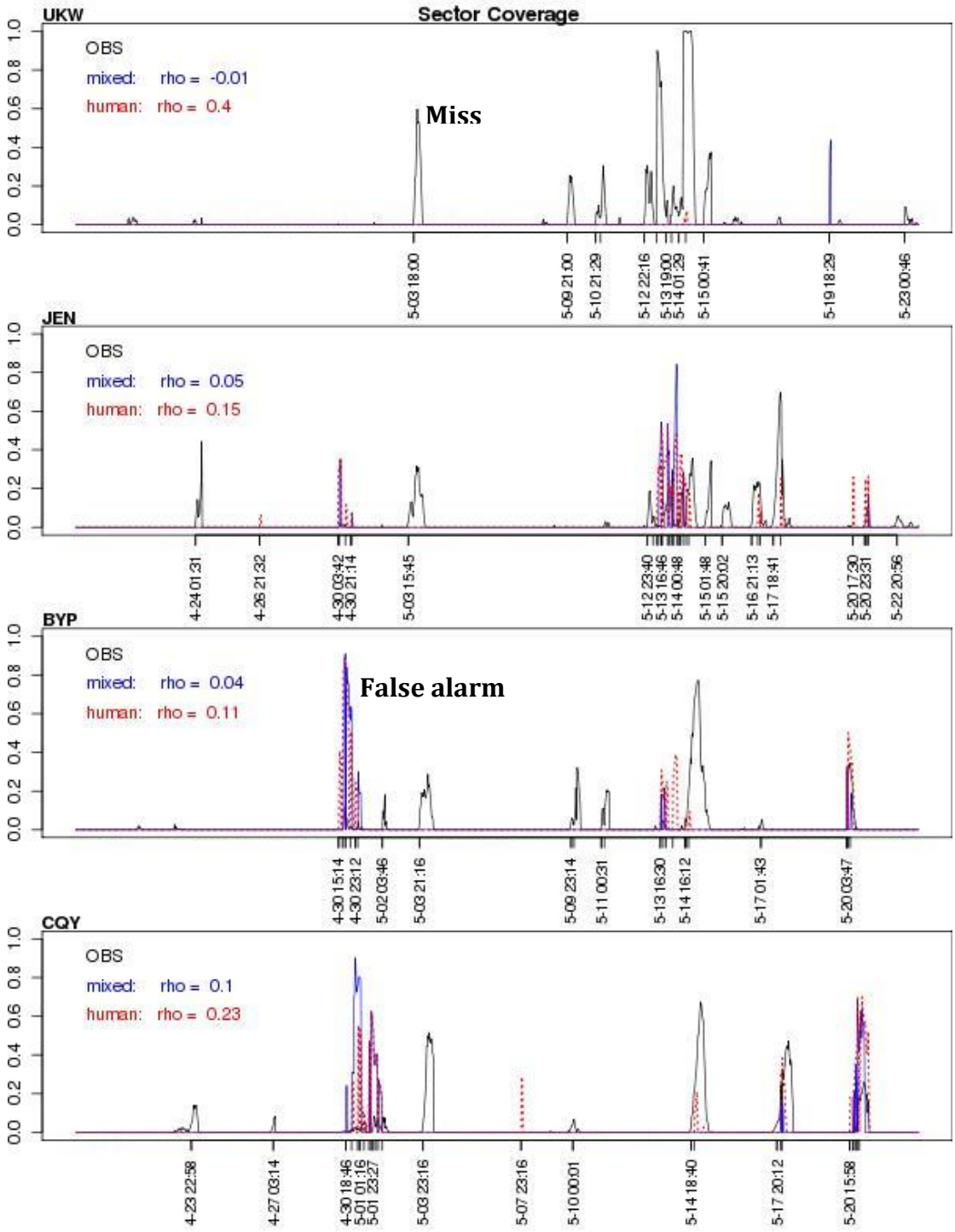

Fig. 14. *Time series of the sector coverage of the observed convection (black), fully automated ANC forecast initiation (blue), and the HITL ANC forecast initiation (red) for the ZFW low-altitude sectors containing the UKW, JEN, BYP, and CQY arrival corner posts. The numbers on the left side of the plots are the correlation coefficients between the observed time series and the fully automated (labeled Mixed) and the HITL (labeled Human) time series, respectively.*

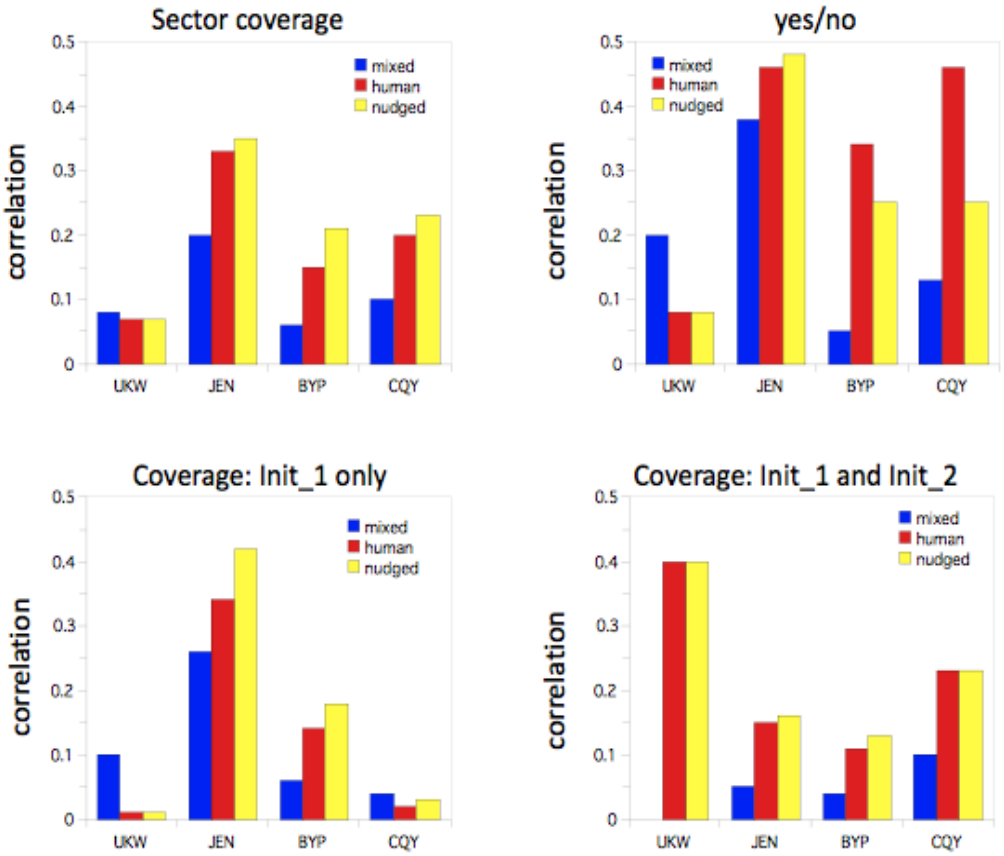

Fig. 15. *Correlation coefficients for time series as in Fig. 14 for the sector coverage of the full initiation forecast (upper left panel), forecasts and observations treated as yes/no (i.e., coverage greater than zero or equal to zero; upper right panel), sector coverage of the initiation 1 only (lower left panel), and sector coverage only when initiation 2 is present (lower right panel). Correlation coefficients are plotted for the fully automated ANC forecasts (labeled mixed; blue), the HITL forecasts (labeled human; red), and the HITL with nudging (labeled nudged; yellow).*

5.4 Notes on Forecast Sensitivity

There are three primary areas in which human involvement can impact the final ANC forecast product: adding boundaries, choosing the ANC regime (e.g., mixed, cold front, dry line, etc.), and the predetermined selection of thresholds for determining initiation levels 1, 2, and 3. Examination of a case study demonstrates that the threshold selection is the most important, followed by boundary additions, and finally, modification to the background regime.

On 4 May 2010 the fully automated mixed regime forecast produces a nearly uniform field of negative initiation potential with a weak boundary in north Texas and a small, circular region of elevated (though still near zero) potential in west Texas (Fig. 16). Changing the regime from mixed to warm front produces only

negligible changes. In addition, the human forecaster adds a frontal boundary extending from the Oklahoma-Arkansas border across north Texas to New Mexico. However, because the background initiation potential field is so low the addition of the boundary is not sufficient to raise the initiation potential above the initiation level 1 threshold. As a result, the final forecast fields in both scenarios are identical. The application of the threshold in this case serves to remove any differences introduced by the human forecaster (for good measure as no convective initiation was observed).

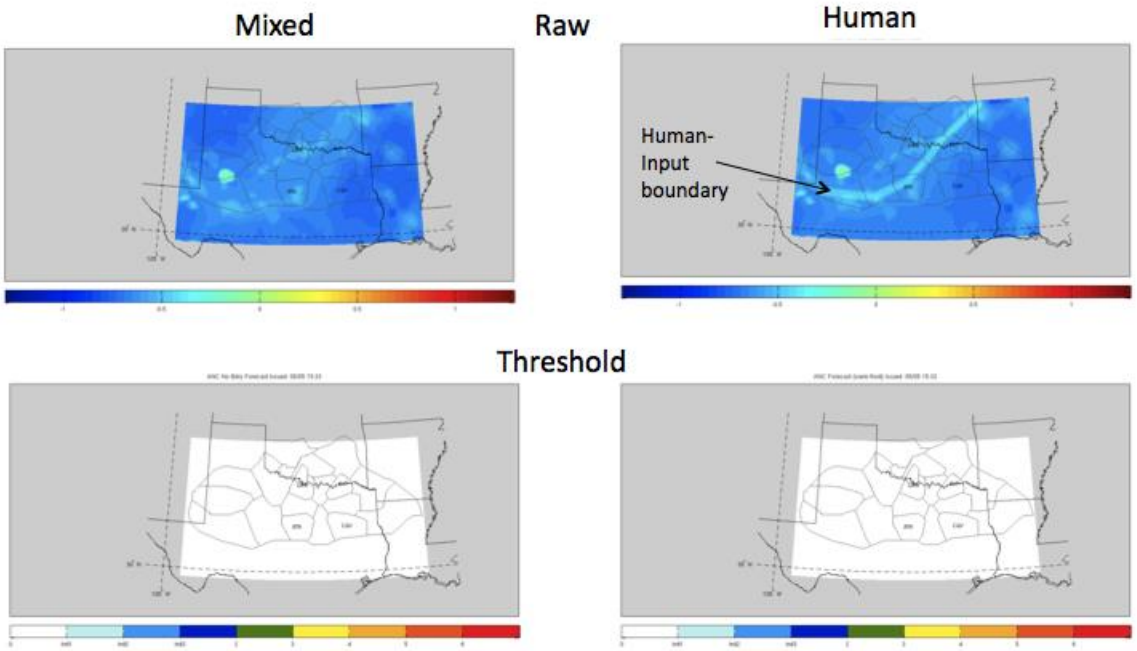

Fig. 16. *Maps of initiation potential on 4 May 2010 before (top) and after thresholding (bottom) for the fully automated, mixed-regime forecast (left) and the HITL forecast (right) in which a large boundary was entered and the regime was changed from mixed to warm front.*

Applying a threshold to the initiation potential field is a nonlinear filter and, as such, can produce inconsistent results. If in the 4 May 2010 case the values in background initiation potential field from the mixed regime were approximately 1.0 larger, the difference between the two continuous fields would remain small. However, applying the categorical filter would introduce a significant change. Both fields would have an initiation 1 region for the circular area in west Texas, while the HITL field would additionally have a large initiation 1 area along the drawn boundary. In other words, the threshold also can act to transform small differences in the initiation potential field into large differences in the final forecast.

It can thus be seen that applying thresholds can both accentuate and minimize differences in the raw initiation potential fields. Examination of time series of the

correlation and mean difference between the fully automated and HITL forecasts before and after applying thresholds (Fig. 17) reveals that differences between the fields are more likely to be strengthened than diminished. The time series also demonstrate that changing the regime in the ANC has only a weak impact on the raw potential field; substantial changes occur only when boundaries are input by the human. Even when boundaries are added the correlation between the fully automated and the HITL forecast fields remain above 0.8, and are typically above 0.95. Low correlations occur only when the continuous fields are discretized by applying the thresholds. Therefore, the selection of threshold values could have a significant impact on the performance of the ANC system.

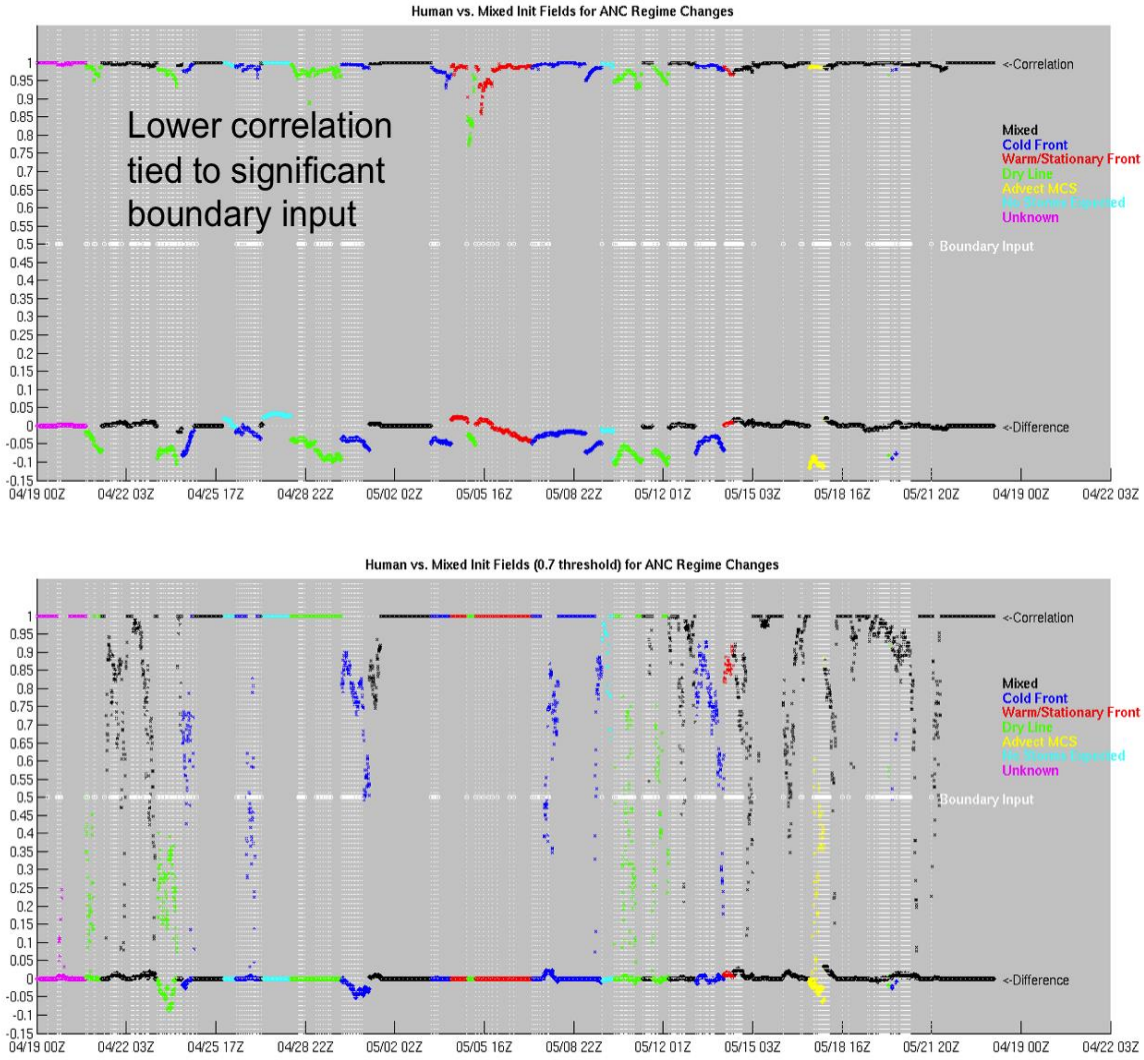

Fig. 17. *Time series plots the correlation (top line) and mean differences (bottom line) between the fully automated and HITL forecasts for the raw initiation potential fields (top panel) and final forecast after applying thresholds (bottom panel). The colors in the time series plots correspond to the human-selected ANC regime and the white marks denote the presence of human-added boundaries.*

6. Conclusions

The AutoNowcaster forecast product was evaluated with a focus on three main areas: the extrapolation component, comparison of the fully automated and HITL initiation component, and for its utility in an aviation-specific capacity. It should be emphasized that, since this evaluation was performed on a demonstration project of relatively short duration, all conclusions are preliminary.

There are two primary conclusions pertaining to the extrapolation component. Firstly, the ANC extrapolation forecasts require calibration. The forecasts by definition represent ongoing convection of 35 dBZ or greater, but in order to eliminate the large underforecast bias the ANC extrapolation needs to be treated as a forecast of 43 dBZ or greater. When verified against 35 dBZ observed reflectivity, CIWS extrapolation significantly outperformed ANC. However, when treated as an extrapolation forecast of 43 dBZ echoes (thereby removing the forecast bias), the gap between CIWS and ANC was reduced substantially.

Compared to the fully automated initiation forecast, human interaction (choosing the ANC regime and adding boundaries) improves the forecasts by reducing the area of initiation 1 forecasts, while increasing both the number and size of embedded initiation 2 and initiation 3. Despite this reduction in the area of forecast initiation, human interaction maintains the same level of skill as the fully automated forecasts, where skill is measured as the distance between forecast and observed initiation cells. About 10% of observed initiation is captured within a forecast region, 25% is within 30 miles, and 75% is within 120 miles. When an initiation forecast contains an initiation 2 or 3 area, the median distance is reduced by half, while the proportion of perfect forecasts is nearly doubled. The continuous initiation potential fields from the fully automated and HITL forecasts are very highly correlated; substantial differences between the two potential fields occur only after they have been discretized into initiation 1, 2, and 3 categories.

Limiting the evaluation to the four low-altitude sectors containing the DFW arrival cornerposts, the increased skill possessed by forecasts with areas of higher embedded potential is once more apparent. The hit rate for occurrence of convection within a sector is nearly double when an initiation 2 area is present compared with an initiation 1 area only. This result holds for both the fully automated and HITL forecasts. Human input does increase forecast association, measured as the correlation between time series of forecast and observed sector coverage, though the correlations for both the fully automated and HITL forecasts are very low. The nudging tool was used sparingly (~10% of all forecasts), but the nudged forecasts did improve slightly upon the HITL forecasts. The correlations are substantially higher, though still not strong, when the sector coverage time series are converted to binary series of the presence or absence of convection.

7. Alternate Convective Initiation Detection Scheme

An alternative approach for detecting initiation can be used in future iterations of the convective initiation assessment. The approach uses the Procrustes technique for matching objects at larger time steps (30 min, 1-h) with a different method of identifying the objects. Identifying objects can be accomplished using band pass filters from a translation into frequency space using a Fourier transform. The band pass filter can be selected to identify objects of a certain spatial scale of interest. Although isolated convection may be missed due to its low signal strength, significant convective initiation, which may be of higher value to a planner focused on air traffic, may be detected in longer time steps. The significant initiation may also be similar in size to the issuance of forecast potential initiation objects from ANC. An example of the identification of convective initiation using a band pass filter approach is shown in Figure 18.

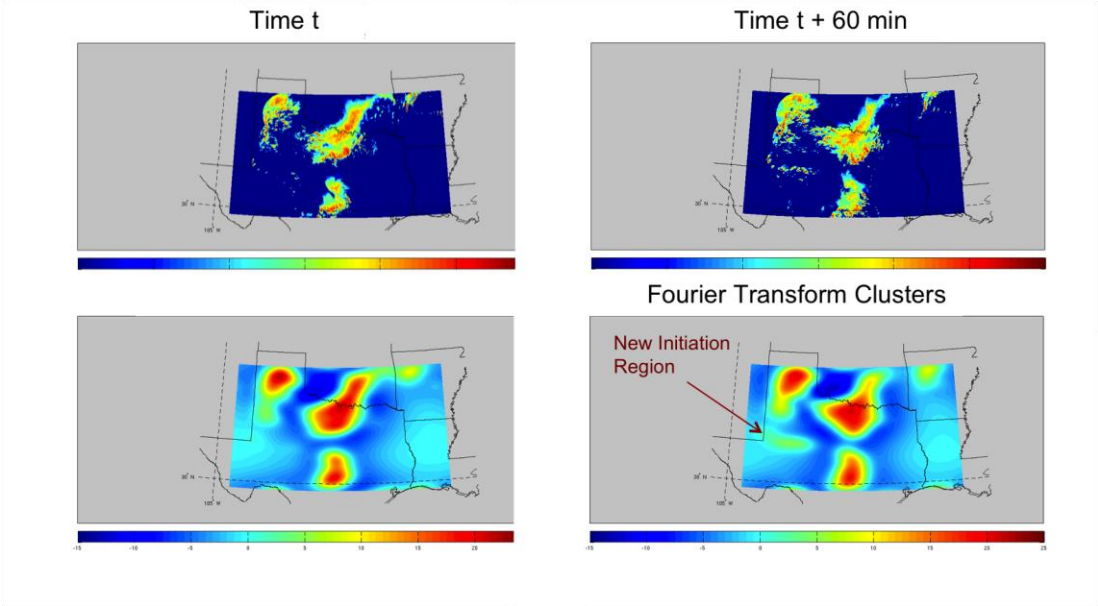

Fig. 18. *An example from 14 May 2010 with observed reflectivity (top) and identified clusters of initiation from the band pass (bottom). The time on the left corresponds to 1745 UTC while the time at the right is 60-min later at 1845Z. A new initiation feature can be seen in W TX.*

8. References

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

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

Mueller, C., T. Saxen, R. Roberts, J. Wilson, T. Betancourt, S. Dettling, N. Oien, J. Yee, 2003: NCAR Auto-Nowcast System. *Wea. Forecasting*, **18**, 545–561.

Troxel, S. W. and C.D. Engholm, 1990: Vertical Reflectivity Profiles: Averaged Storm Structures and Applications to Fan-Beam Radar Weather Detection in the U.S., 16th Conference on Severe Local Storms, Kananaskis Park, AB, Canada, *Amer. Meteor. Soc*.