1. INTRODUCTION

Air Traffic planning is primarily split between two methodologies: strategic (2-6+ hour forecast lead time) and tactical (0-2 hour forecast lead time). Strategic planning occurs daily and is most useful when the weather is highly predictable out to 6-8 hrs in the future. When unanticipated convective weather causes airspace capacity to drop below a sustainable threshold, traffic flow planning becomes less strategic and more tactical. In this case, it is critical that reliable and accurate short-term 0-2 hour forecasts of precipitation and storm tops be available. The Corridor Integrated Weather System (CIWS) was developed to respond to the 0-2 hour forecast needs of the FAA decision makers. The CIWS system produces forecasts of vertically-integrated-liquid water (VIL) and echo tops (ET) over the northeast US at 5-minute increments out to 2 hours (Wolfson & Clark, 2007).

The CIWS system makes extensive use of radar input (NEXRAD, TDWR, and Canadian) for its storm identification and tracking routines. These radars provide nearly complete coverage over the eastern two-thirds of the CONUS (Figure 1a), but areas of degraded coverage exist over the mountain West and off the coasts. In these regions the radar data and CIWS forecasts produced from these data can be adversely affected. On the other hand, the National Lightning Data Network (NLDN) provides nearly continuous coverage of cloud-to-ground (CG) lightning over the CONUS in addition to coverage off the coasts (Figure 1b; Cummins et al. 1998b).

---

†This work was sponsored by the Federal Aviation Administration under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

*Corresponding author address: Haig Iskenderian, MIT Lincoln Laboratory, 244 Wood Street, Lexington, MA 02420-9185; e-mail: haig@ll.mit.edu

In addition to degraded radar coverage caused by terrain blockage or off-shore storms,
radar networks occasionally suffer from outages due to problems in communications or hardware. Figure 2 provides two examples of this.

The first situation (Figure 2a) occurred on 17 August 2007 when the Eastern Region NEXRAD coverage was unavailable to CIWS from 1445 to 2252 UTC, or a span of 8 hours and 7 minutes. The loss of radar data had a high impact on the CIWS forecast and FAA planners who used CIWS since the data loss occurred at a time when there was convective weather near major East Coast airports. Data from the Terminal Doppler Weather Radars (TDWRs) were available, but their range is limited to 90 km; beyond 90 km there was no coverage without the NEXRADs. In this case, the lightning data continued to be received during the entire duration of this radar outage. The NLDN data showed clusters of CG lightning strikes in eastern PA outside the TDWR coverage zone, indicating intense convective activity in the vicinity of several major jet routes.

The second situation (Figure 2b) occurred on 11 May 2007, when communication problems rendered many southern NEXRAD radars unavailable to CIWS from 1840 to 2031 UTC, or a span of 1 hour and 51 minutes. There were numerous thunderstorms in and around the Atlanta-Hartsfield airport during this time period. The loss of NEXRAD radar data caused large areas of degraded coverage in the vicinity of the airport. As will be shown later in this paper, the loss of radar data led to an inferior CIWS forecast over the region. Similar to the 17 August event, the lightning data continued to be received throughout the radar outage. Clusters of cloud-to-ground lightning strikes over central AL indicated strong storms where the radar indicated little if any VIL.

Clearly these clusters of cloud-to-ground lightning strikes depict regions of active convection, and this information could be critically important in the event of loss or degradation of radar data. In this paper, we study whether appropriate translation of the lightning data can supply information regarding the two most important weather radar mosaics for traffic managers: VIL and Echo Tops. This paper uses statistically derived relationships between the 2007 summer lightning strike data and the fields of VIL and ET in the Northeast Corridor domain to illustrate the feasibility of using cloud-to-ground lightning operationally as a proxy for CIWS radar mosaics.

2. BACKGROUND

The determination of a robust relationship between lightning strikes and radar signatures has been sought by several researchers. Tapia et al. (1998) studied 22 summer thunderstorms over Florida and constructed a model for the ratio of rainfall to lightning strikes on the storm scale. They applied their model to estimate the radar rainfall from lightning data and concluded that it was
possible to derive reasonable estimates of precipitation in space and time for heavy rainfall events. Zhou et al. (2002) derived a similar ratio for storms over China and concluded that CG lightning could be used to estimate convective precipitation, especially in regions of insufficient radar coverage. Cheze & Sauvageot (1997) and Soula & Chauzy (2001) found good correlations between lightning and radar data on the storm scale for thunderstorm cases over France. While these case studies indicate the possibility of using lightning data as an estimator of radar signatures in localized regimes, their relationships are not directly applicable to nowcasting systems that must operate over large spatial domains.

There has also been work toward direct implementation of lightning data in nowcasting systems. Weber et al. (1998) performed a regression of VIL in 4 km² pixels and NLDN lightning data for three convective cases and found a modest correlation in the two fields, noting that lightning data could provide benefit in filling in gaps in the NEXRAD coverage. Mueller et al. (1999) visually compared VIL at 4-km resolution and lightning data to derive a relationship to utilize lightning data to improve radar data latency in the National Convective Weather Forecast (NCWF). Megenhardt et al. (2004) mapped the lightning data to a 4-km grid to create relationships between lightning strikes and VIL to serve as input to a hazard detection field called the National Convective Weather Detection (NCWD), upon which NCWF is based. They found that lightning data adds to the NCWD along the leading edges of storms and in regions where radar data are missing.

These studies suggest that lightning data could be useful in a nowcasting system, particularly where radar data are degraded or unavailable. This paper builds upon the prior work by developing a statistical model to relate lightning to VIL and ET at the 1 km² scale over the CIWS domain. There are several basic limitations to any technique involving CG lightning, however. In the early stages of convective development, intracloud (IC) lightning tends to prevail (Solomon & Baker 1998; Williams et al. 1989), so the relationships developed here will not perform well because the NLDN detects only CG and not IC lightning with high detection efficiency. Secondly, CG lightning activity tends to be highest in the vicinity of the cores of the storms (Carte & Kidder 1977), so the relationships developed here will not perform well in stratiform regions surrounding the storms. Lastly, non-convective synoptic-scale storms that do not contain cloud-to-ground lightning will not be depicted via this lightning-derived VIL and ET technique.

3. DATA SETS

The VIL and ET used in this study are from individual NEXRAD radars that are mosaiced to a uniform 1 km grid. (See Dupree et al. 2005 for details on the mosaic procedure). The mosaiced VIL and ET are available every 5 minutes over the CIWS domain (Figure 1). Only VIL and ET within the 230 km range coverage areas were used in this study to ensure good quality data for statistical analysis.

The NLDN provides the locations of CG lightning strikes over the CONUS and adjacent waters. The network provides coverage with a median location accuracy of 500 m. Off the coasts and over southern Canada, the median location accuracy is 2-4 km. The NLDN lightning data set contained all CG strikes collected over the prior 6 minute interval, and was updated every 2 minutes.

Values of convective available potential energy (CAPE) are used in this study to investigate any environmental influence on the relationships between lightning and radar data. CAPE is derived from a combination of upper-air fields from the Rapid Update Cycle (RUC; Benjamin et al. 2004) model and the surface fields from the Space-time Mesoscale Analysis System (STMAS; Xie et al. 2005). The CAPE fields were interpolated to the 1 km radar grid and were available every 15 minutes. Similar interpolation and investigation was performed for the RUC variables of lower-tropospheric relative humidity, lower-tropospheric mixing ratio, and convective cloud top potential, but these fields showed little impact on the lightning-radar relationships.

4. METHODOLOGY

Shown in Figure 3 are the steps to prepare the data for the development of statistical relationships between the lightning and the radar fields. The point locations of the CG lightning strikes (Figure 3a) are aligned in time with the radar data and placed on the 1 km radar grid. To place the CG data on the radar grid, the number of lightning strikes in each square km of the radar grid is calculated to yield the number of lightning strikes per square km. Next, the lightning data are spatially-smoothed by assigning to each pixel the total number of lightning strikes within a radius of 8 km of a pixel (Figure 3b). The result of this
The lightning flash rate field is a field that contains features similar to the radar data, particularly near the convective cores. For example, the two areas of high VIL in the squall line over eastern PA and the gap between them utilized by NY arrivals and departures (Figure 3c) are shown fairly well in the lightning flash rate field (Figure 3b). The northern squall line over eastern NY is not depicted well in the lightning flash rate field since this is a decaying line and there is little lightning activity associated with it.

This procedure to calculate flash rates on the 1 km grid was performed on data from June, July, and August of 2007 at 5-minute intervals over the CIWS domain. A 3-hourly subset of the data containing the flash rate, radar VIL and ET, and CAPE was created for statistical analysis. To determine the points included in this dataset, values of lightning flash rate ($\geq 0$), VIL ($\geq$ level 1), ET ($\geq 0 \text{ kft}$) and CAPE ($\geq 0 \text{ J kg}^{-1}$) within a search radius of 16 km from each lightning strike were included. This search radius, which is twice as large as the radius used to calculate the smoothed flash rates, ensured a representative sample of points in the vicinity of and removed from lightning for model development. The model results were not particularly sensitive to the exact choice of this search radius. The resulting dataset contained more than 3 million points reflecting the variety of synoptic conditions during the summer and capturing the diurnal cycle of convective activity. Two-thirds of the data set was used for model development, the remaining third was set aside for testing the models during development.

A probability matching method (PMM) (Calheiros & Zawadzki 1987) is used to create relationships between $VIL$, $ET$, and lightning flash rate ($L$). A PMM approach has been used by Atlas et al. (1990, 1993); Calheiros & Zawadzki (1987); Crosson et al. (1996); Marks et al. (1993); and Rosenfeld et al. (1993) to derive relationships between radar reflectivity and rainrate, and by Chang et al. (2001) to derive a relationship between lightning and convective rainfall to explore the usefulness of assimilating lightning information in a numerical weather prediction model. The principle of the PMM method is to construct $L - VIL$ (and $L - ET$) relationships based on $L_i, VIL_i$ pairs such that the cumulative distribution functions (CDFs) of $L$ and $VIL$ at the $i$th-probability interval match:

$$\int_{VIL_i}^{VIL} P(VIL) dVIL = \int_{L_i}^{L} P(L) dL$$  \hspace{1cm} (1)$$

$$\int_{ET_i}^{ET} P(ET) dET = \int_{L_i}^{L} P(L) dL$$  \hspace{1cm} (2)$$

where $P(\quad)$ represents a probability density function, and $VIL_i$ and $ET_i$ are low threshold values. To find the low threshold values, we
followed the approach of Rosenfeld et al. (1993). We first use the entire data set as described in Section 5 from which the unconditional CDFs of $VIL_t$, $ET_t$, and $L_t$ are created. We specify $L_t = 1$, which is the lowest detectable lightning flash rate value, and the unconditional CDFs are used to match the unconditional probability of $VIL_t$ and $ET_t$ to the unconditional probability of $L_t$.

Once $VIL_t$ and $ET_t$ have been determined, Eqs. 1 and 2 are applied to the dataset to find $L-VIL$ and $L-ET$ relationships. Shown in Figure 4a are the CDFs for $L$ and $VIL$. From the cumulative distribution functions, the $L-VIL$ relationships are constructed by matching probabilities for the two fields. For example, $L = 6$ and $VIL = 19$ at a probability = 0.7. Therefore, the point (6, 19) is included in the $L-VIL$ relationship (Figure 4b). The PMM relationship in Figure 4b shows that VIL increases as lightning flash rate increases, which is consistent with the case studies of Shafer et al. (2000), and the result of Reap & MacGorman (1989) for a study involving two warm seasons over the Midwest US.

![Figure 4.](image-url)

Figure 4. a) The cumulative distribution function (CDF) of the radar VIL (kg m$^{-2}$) and lightning flash rate (flashes/6-min). b) The lightning flash rate-VIL relationship created from the PMM technique. Points with the same cumulative probability of flash rate and VIL (red and green dashed lines in (a)) are matched (black dashed line in (b)) to create this curve. c) The CDFs of lightning flash rate and echo top height (kft) for all values of CAPE. d) Flash rate-ET relationships created by PMM technique for all CAPE values, and stratified by four CAPE intervals.
In an attempt to include possible effects of surface and atmospheric variables on the lightning-radar relationships, the CDFs and resulting PMM relationships for VIL and ET were constructed from data stratified by time of day, land vs. water, latitude zone, and CAPE. CAPE was the only factor that caused noticeable differences in the PMM relationships and model scores in the case of ET. Figure 4c shows the CDFs for $L$ and $ET$ for all CAPE values and in Figure 4d the PMM relationships have been stratified by CAPE. The PMM $L-ET$ relationships show an increase in ET as flash rate increases, consistent with the results of Watson et al. (1995). A higher ET is diagnosed from a given flash rate as CAPE increases. This result is likely due to stronger updrafts in higher CAPE environments (Weisman & Klemp 1986), which result in higher storm echo tops.

5. RESULTS

a) Application of the proxy model

Figure 5 illustrates an example of the application of the PMM relationships to calculate VIL and ET from the lightning strikes at 2245 UTC 27 July 2007. The locations of the lightning strikes for this case are shown in Figure 3a. The VIL created by the lightning proxy relationships (Figure 5a) shows two broad areas of convective cores over eastern PA depicted by areas of VIP levels 5 and 6, as well as the gap in the squall line. Comparison of Figure 5a with the radar observations in Figure 5b shows the VIL field in the convective cores is captured well in this squall line. The decaying squall line over eastern NY is not captured well in the lightning relationships due to the lack of lightning activity in this decaying line.

Figure 5. Figure shows a comparison of VIL and ET from proxy relationships with VIL and ET from radar. All data are for 2245 UTC on 27 July 2007. a) Proxy VIL calculated from lightning flash rate for level 3+. b) Radar-derived VIL for level 3+. c) Proxy ET height ($\geq 30$ kft) calculated from lightning flash rate. d) Radar-derived ET height. The main features of the squall line are depicted by the proxy VIL and ET.
The ET derived from lightning (Figure 5c) shows the high echo tops in the squall line with localized tops in excess of 50 kft. The gap in the squall line is also evident. The echo top heights derived from lightning compare well with observations (Figure 5d) in the convective cores, although the isolated regions of 50 kft in the lightning cores are higher than observed. Note that the ET in the stratiform regions flanking the convective cores is not captured by the proxy relationships since there is little CG lightning in these regions. From these results, the lightning proxy relationships capture the majority of the operationally-significant features in the radar fields.

To investigate the model performance over the entire 2007 summer season, a validation dataset of over 1 million points was constructed using the methodology in Section 4, except the criteria that data points be within a 16 km search radius of a lightning strike was eliminated. This elimination has the effect of including events that do not contain lightning in the dataset, such as non-convective synoptic-scale storms, and newly-formed or decaying convection that do not contain lightning. To characterize the statistical agreement between the observed radar VIL and ET and the VIL and ET calculated using the proxy relationships, standard contingency table performance scores were calculated using each point and are shown in Table 1. The probability of detection (POD), the false alarm rate (FAR), and the critical success index (CSI) are defined as:

$$POD = \frac{n_{hit}}{n_{hit} + n_{miss}}$$  \hspace{1cm} (3) \\
$$FAR = \frac{n_{falseAlarm}}{n_{hit} + n_{falseAlarm}}$$  \hspace{1cm} (4) \\
$$CSI = \frac{n_{hit}}{n_{hit} + n_{miss} + n_{falseAlarm}}$$  \hspace{1cm} (5)

Table 1: Contingency scores of the PMM model for VIL (VIP level ≥ 3) and ET (≥ 35 kft).

<table>
<thead>
<tr>
<th>Model Score</th>
<th>VIL</th>
<th>ET</th>
</tr>
</thead>
<tbody>
<tr>
<td>POD</td>
<td>0.34</td>
<td>0.32</td>
</tr>
<tr>
<td>FAR</td>
<td>0.36</td>
<td>0.40</td>
</tr>
<tr>
<td>CSI</td>
<td>0.29</td>
<td>0.26</td>
</tr>
</tbody>
</table>

The scores for VIL are lower than reported by Tapia et al. (1998) for their lightning-rainfall model (their scores are POD=0.44, FAR=0.23, CSI=0.39); however their scoring was performed for a 6 hr period over a limited region when convection (with lightning) was known to have occurred, in contrast to our sample which is over the eastern US for an entire summer and many types of events. Since the purpose of this work is to derive relationships for an operational system, it is important to obtain a benchmark for model performance under all conditions. If the scoring is performed on the testing dataset as created in Section 4 (which only included points within a 16 km search radius of lightning strikes), the scores improve to POD=0.69, FAR=0.36, CSI=0.49 for VIL and POD=0.55, FAR=0.41, CSI=0.40 for ET.

b) Blending proxy fields with observed radar

To form a coherent depiction of the weather to display to the user and to serve as input to the CIWS system, the VIL and ET calculated from lightning must be blended with the existing radar VIL and ET where the two sources of data co-exist. In the near range (0-230 km) areas, the radar VIL and ET are of good quality and will be exclusively used. In the no coverage areas (where radar data do not exist), VIL and ET calculated from the lightning proxy relationships will be used. In the far range (230-460 km) areas, a blend of the VIL derived from lightning and radar will be created. The degradation of radar coverage in the presence of the convective weather in the far range is a non-linear function of distance from the radar (Joss & Waldvogel, 1990). We have initially adopted an exponential function for range-dependent weighting of the radar data and the VIL derived from the proxy relationships in the far range. The relationships to determine the VIL at any point are:

$$VIL_{NR} = VIL_r$$  \hspace{1cm} (6) \\
$$VIL_{FR} = VIL_p (1 - w) + VIL_r (w)$$  \hspace{1cm} (7) \\
$$VIL_{NC} = VIL_p$$  \hspace{1cm} (8)

where $VIL_{NR}$, $VIL_{FR}$, and $VIL_{NC}$ are the VIL in the near range, far range, and no coverage areas respectively, $VIL_p$ is the VIL derived from the lightning proxy PMM relationships, and $VIL_r$ is the observed radar VIL. The exponential weighting
function $w$ in the far range is calculated as a function of distance from an operational radar.

As an example of the blending procedure, Figures 6a and 6b show the VIL derived from lightning and the observed VIL for 2030 UTC 11 May 2007. The range-dependent spatial weighting (Figure 6c) is applied to these fields through Eqs. 6 to 8 and the resultant blended VIL product is shown in Figure 6d. The gray scale background in Figure 6d indicates to the user the near range, far range, and no coverage areas. In the far range regions, the influence of the lightning proxy VIL increases as distance increases from the operational radars. In the near range, the observed VIL from the operational radars is used without modification, and in the no coverage areas the VIL field is totally provided by the proxy VIL. Comparison of Figure 6d with Figure 6b shows that the addition of proxy VIL captures the convective storms over central AL that are absent in the radar data alone, providing a more complete depiction of the weather that would be a benefit to traffic managers. The “truth” radar VIL field was available 5 minutes later when the outage ceased, and is shown in Figure 7d.

![Blended radar and proxy VIL](image)

Figure 6. Figure shows the components of the blended radar and proxy VIL field. a) VIL calculated from lightning flash rate for 2030 UTC 11 May 2007. b) Same as (a) except radar VIL. c) Weighting function used to blend observed VIL and lightning-derived VIL. d) Blended radar and lightning-derived VIL. The blended VIL shows convective activity in the far range and no coverage regions where the radar data shows little or no VIL.
A more complete depiction of the long range and otherwise degraded portions of the VIL and ET field would also benefit the CIWS forecast. Figure 7a shows the initial radar VIL field at 2030 UTC upon which a 5-minute CIWS VIL forecast valid at 2035 UTC is based (Figure 6b). The forecasted VIL resembles the radar VIL, with most of the forecasted storms concentrated in the northern portion of the domain. Since the radar data are poor or non-existent in the far range and no coverage regions, there is a lack of forecasted VIL in the central and southern portions of the domain. The storms in this region are shown in the proxy VIL at the initial time (Figure 7c) and are clearly observed 5 minutes later when the radars returned to the CIWS mosaic (Figure 7d). The CIWS 5-minute forecast, which did not capture the storms in the central and southern portions of the domain since they were not in the initial VIL field, would have benefited from addition of the proxy VIL.

The blended VIL field for the 17 August case is shown in Figure 8a. The VIL in the no coverage areas is derived completely from the proxy VIL, and comparison to Figure 2a shows that the proxy VIL provides potentially useful information on the convective cells over eastern PA just beyond the TWDR near range. There is no far range weighting required in this situation as all TDWR data beyond 90 km are considered no coverage areas. Figure 8b shows the VIL field for 17 August 2007 at 2125 UTC, 45 minutes after Figure 8a. The cores that were previously depicted by the proxy VIL over eastern PA have entered the TDWR near coverage and contain VIP level 6, as indicated 45 minutes earlier by the lightning proxy relationships.
6. SUMMARY AND FUTURE WORK

Continuous short-term (0-2 hr) weather forecasts are crucial to FAA traffic managers. In the event of degraded or lost radar data due to radar outages, terrain blockage, and very distant coverage (e.g., off the coasts), the quality of the CIWS mosaics and forecasts can be greatly diminished. Relationships have been sought between cloud-to-ground lightning data and the radar fields of VIL and echo tops for use in the CIWS in the event of degraded or lost radar data. A probability matching methodology has been applied to lightning and radar data to develop proxy relationships for VIL and ET. No influence of the lightning data is incorporated in regions of high quality radar coverage; the lightning influence is incorporated only beyond 230 km range or in no coverage regions. The proxy relationships provide potentially useful weather information to FAA traffic managers for situational awareness, and could serve as input to the CIWS forecast system to produce a forecast under conditions of radar data degradation or loss. In the future, the proxy relationships will be considered for use in the CIWS forecast system.

The regions of VIL and ET derived from the proxy relationships tend to be coherent in time and space, and therefore have the potential to be tracked in the CIWS. We will investigate the possibility of tracking the areas of lightning activity and relating the growth and decay trends in the lightning with growth and decay observed in radar to potentially improve storm decay in CIWS convective weather forecasts. As the domain of the CIWS system expands to the CONUS, statistical relationships using this methodology will be developed for the climatic regions that contain different lightning characteristics, such as over the relatively arid mountainous regions of the West and over the subtropical regions of Florida and the Gulf of Mexico (Williams et al. 2005).

APPENDIX A. VIL thresholds for FAA weather radar color levels (VIP levels).

<table>
<thead>
<tr>
<th>VIL (kg m$^{-2}$)</th>
<th>VIP level</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.14</td>
<td>1</td>
</tr>
<tr>
<td>0.76</td>
<td>2</td>
</tr>
<tr>
<td>3.50</td>
<td>3</td>
</tr>
<tr>
<td>6.90</td>
<td>4</td>
</tr>
<tr>
<td>12.0</td>
<td>5</td>
</tr>
<tr>
<td>32.0</td>
<td>6</td>
</tr>
</tbody>
</table>

7. REFERENCES


