The Impact of Dropwindsonde and Supplemental Rawinsonde Observations on Track Forecasts for Hurricane Irene (2011)

SHARANYA J. MAJUMDAR
Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida

MICHAEL J. BRENNAN
NOAA/NWS/NCEP/National Hurricane Center, Miami, Florida

KATE HOWARD
NOAA/NWS/NCEP/Environmental Modeling Center, College Park, Maryland

(Manuscript received 30 January 2013, in final form 8 June 2013)

ABSTRACT

Because of the threat that Hurricane Irene (2011) posed to the United States, supplemental observations were collected for assimilation into operational numerical models in the hope of improving forecasts of the storm. Synoptic surveillance aircraft equipped with dropwindsondes were deployed twice daily over a 5-day period, and supplemental rawinsondes were launched from all upper-air sites in the continental United States east of the Rocky Mountains at 0600 and 1800 UTC, marking an unprecedented magnitude of coverage of special rawinsondes at the time. The impact of assimilating the supplemental observations on National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) model track forecasts of Irene was evaluated over the period that these observations were collected. The GFS track forecasts possessed small errors even in the absence of the supplemental observations, providing little room for improvement on average. The assimilation of the combined dropwindsonde and supplemental rawinsonde data provided small but statistically significant improvements in the 42–60-h range for GFS forecasts initialized at 0600 and 1800 UTC. The primary improvement from the dropwindsonde data was also within this time range, with an average improvement of 20% for 48-h forecasts. The rawinsonde data mostly improved the forecasts beyond 3 days by modest amounts. Both sets of observations provided small, additive improvements to the average cross-track errors. Investigations of individual forecasts identified corrections to the model analyses of the Atlantic subtropical ridge and an upstream midlatitude short-wave trough over the contiguous United States due to the assimilation of the extra data.

1. Introduction

Because of its size, intensity, and impact on the United States over several days, Hurricane Irene was a notorious tropical cyclone in 2011. Irene made its first landfall in the United States near Cape Lookout, North Carolina, around 1200 UTC 27 August with a maximum 1-min sustained wind speed of 75 kt [1 kt = 0.51 m s⁻¹; Avila and Cangialosi (2011)], a category 1 hurricane on the Saffir–Simpson hurricane wind scale (NWS 2012). Subsequently, Irene moved north-northeastward near the mid-Atlantic coast, and weakened to a tropical storm before making landfall near Atlantic City, New Jersey, and then over New York, New York, on the morning of 28 August. In the United States, Irene produced storm surge, widespread flooding and loss of power, and was directly responsible for over 40 fatalities and caused an estimated $6.5 billion in damage (DOC 2012). The main impacts from Irene in the United States occurred from storm surge and freshwater flooding due to heavy rainfall from North Carolina northward to New England.

During its approach to the United States coast, Irene was one of the most heavily sampled tropical cyclones in history in terms of observations in the inner core, the near-storm environment, and the environment well...
upstream. With the aim of improving numerical model analyses of the steering flow near and upstream of Irene, 10 synoptic surveillance aircraft missions were flown from 23–27 August 2011 to deploy global positioning system (GPS) dropwindsondes. In addition, supplemental 0600 and 1800 UTC rawinsondes were requested beginning at 1800 UTC 22 August in the southeast and mid-Atlantic states. Generally, the coverage of these supplemental rawinsondes is determined subjectively by National Hurricane Center (NHC) forecasters in coordination with the National Weather Service (NWS) Southern and Eastern Regions, and the supplemental rawinsondes are typically requested to coincide with the first synoptic surveillance aircraft mission.

Since 1997, the surveillance aircraft missions have been conducted annually in an effort to reduce track forecast uncertainty for the issuance of hurricane watches and warnings for the United States. Flight tracks are drawn to ensure a symmetric distribution of dropwindsondes (and available rawinsondes) within 3° of the cyclone center, fill gaps in the radiosonde network, and target potential synoptic features of interest (C. Landsea 2012, personal communication). For 175 synoptic surveillance missions flown during 1997–2006, the dropwindsonde data had proven to be useful, yielding an average reduction of National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) track errors by 10%–15% during the first 60 h of the forecast (Aberson 2010). However, it is worth noting from the same study that little improvement had been found in the regional Geophysical Fluid Dynamics Laboratory (GFDL) hurricane model track and intensity forecasts.

There are no comprehensive studies in the literature evaluating the impact of assimilating supplemental rawinsonde data on operational numerical forecasts, although it was recently found that these extra data provided small but positive improvements on NCEP GFS track forecasts during the 2008 season (N. Ramos 2012, personal communication). In an unprecedented move during Irene, the coverage of the supplemental rawinsondes was expanded to include all 59 continental U.S. upper-air stations from the Rocky Mountains eastward starting at 0600 UTC 25 August.

These supplemental observations are routinely integrated into the NCEP data assimilation scheme that produces initial conditions for GFS model forecasts. Given the large costs associated with these supplemental observations,1 a natural question arises as to what impact they had on hurricane forecasts, especially for Irene. Generally, the track forecast errors of the GFS and the NHC official forecast (OFCL) were relatively low for Irene (Table 1). For example, the average OFCL track errors for Irene were 38, 68, 101, 132, and 237 n mi (1 n mi = 1.852 km) at days 1, 2, 3, 4, and 5, respectively (Table 1). These errors were 23%–25% smaller than the 5-yr average OFCL track error forecasts from 2006 to 2010 at days 1–4, but about 10% higher than the 5-yr mean at day 5.

The goal of this study is to quantify the impact of assimilating these supplemental observations on numerical forecasts of the track of Irene with the GFS model. Section 2 will provide details of the supplemental observations, the modeling and assimilation systems, and the methodology for the “data denial” experiments. A synoptic overview of Irene’s evolution and operational model performance will be presented in section 3. Section 4 will contain the overall results of the numerical experiments and illuminate in more detail the impact of the supplemental observations for selected GFS cycles. Section 5 will provide conclusions and suggestions for future work.

2. Data and methodology

a. Datasets

All standard observational data assimilated into the 2011 version of the operational NCEP global modeling system are used in this study. The special dropwindsonde and rawinsonde observations collected during the period of Irene were also assimilated in the operational system.

Ten synoptic surveillance missions were tasked during the period of 23–27 August 2011. These missions were composed of a simultaneous deployment of the National Oceanic and Atmospheric Administration (NOAA) G-IV and U.S. Air Force (USAF) C-130 aircraft centered on 0000 UTC 23 August 2011 (Fig. 1a), the results of which will be discussed in more depth in section 4c, followed by nine twice-per-day missions using only the G-IV. The

---

1 Each synoptic surveillance mission flown by the NOAA G-IV jet costs approximately $42,000 while costs for overtime and materials for the supplemental rawinsonde launches was around $100,000.

<table>
<thead>
<tr>
<th>Forecast hour</th>
<th>OFCL</th>
<th>GFSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>23.8</td>
<td>26.4</td>
</tr>
<tr>
<td>24</td>
<td>38.0</td>
<td>38.7</td>
</tr>
<tr>
<td>36</td>
<td>53.0</td>
<td>51.6</td>
</tr>
<tr>
<td>48</td>
<td>68.4</td>
<td>60.4</td>
</tr>
<tr>
<td>72</td>
<td>101.2</td>
<td>80.6</td>
</tr>
<tr>
<td>96</td>
<td>132.2</td>
<td>99.7</td>
</tr>
<tr>
<td>120</td>
<td>237.1</td>
<td>171.8</td>
</tr>
</tbody>
</table>

2006–10 OFCL avg | 31.0 | 31.0 |

No. of Irene forecasts | 30 | 28 | 26 | 24 | 20 | 16 | 12

TABLE 1. Average track forecast errors (n mi) for OFCL and GFSI for Irene. The 2006–10 mean OFCL forecast errors are given for comparison. The bottom row refers to the number of forecasts for Irene included for each valid time.
rationale for the design of the flight tracks involved a uniform sampling pattern around Irene, with additional dropwindsondes intended primarily to target the evolving upper-level trough and weakness in the ridge to the northwest of Irene during this period. Additional dropwindsonde data were collected by the NOAA P-3 and USAF C-130 aircraft, and these were assimilated into the NCEP system if they were sampled more than
111 km away from the tropical cyclone center. This criterion has been in use in NCEP operations in recent years, to alleviate the concern that the assimilation of these inner-core observations may potentially degrade the forecast due to the insufficient resolution of the global model in the tropical cyclone inner-core region (Aberson 2008). It is worth noting, however, that this concern had been raised based on results using a lower-resolution version of the GFS than the current operational one.

The initial launch of supplemental rawinsondes began at 1800 UTC 22 August 2011 over the southeastern and mid-Atlantic states, and this coverage was maintained with a couple of additions through 1800 UTC 24 August (Fig. 1b). Beginning at 0600 UTC 25 August, the coverage of supplemental rawinsondes was expanded to include all sites along and east of the Rocky Mountains in the expectation that more accurate model initialization of upstream features east of the Rockies would improve the track forecasts of Irene as it approached the U.S. east coast.

b. Methodology

The version of the NCEP Gridpoint Statistical Interpolation (GSI) data assimilation system that was operational in 2011 is a three-dimensional variational method in which model analyses are produced in grid space (Kleist et al. 2009). At the time, the GFS model possessed a horizontal grid resolution of T574 (approximately 27 km) with 64 vertical levels. The following four assimilation–forecast cycles were performed on NCEP’s research and development supercomputer:

CONTROL—the operational version of the GFS and GSI that includes all supplemental data,2
NODROP—a parallel GFS–GSI cycle that withholds all the dropwindsonde data from the assimilation,
NORAOB—a parallel GFS–GSI cycle that withholds all the supplemental 0600 and 1800 UTC rawinsonde data from the assimilation (but includes all 0000 and 1200 UTC rawinsonde data), and
NEITHER—a parallel GFS–GSI cycle that withholds both the supplemental rawinsonde data and the dropwindsonde data.

These cycles were all identical prior to 1800 UTC 22 August, at which time the first supplemental rawinsondes were launched. A few of the earliest dropwindsondes from the first synoptic surveillance missions were also assimilated in this cycle.

---

2 This cycle is referred to as a “control” instead of as “operational,” since very small differences occur between the true operational GFS forecasts and those produced here, due to the computations being performed on a different machine.

3. Synoptic overview of Irene

Irene formed as a tropical storm at 0000 UTC 21 August 2011 about 120 nmi east of Martinique and strengthened to a hurricane as it moved west-northwestward, crossing Puerto Rico early on 22 August and passing near the north coast of Hispaniola on 23 August (Avila and Cangialosi 2011). At this time, the first supplemental dropwindsonde and rawinsonde deployments were made (Fig. 1).

At 0000 UTC 24 August, Hurricane Irene was centered between the north coast of Haiti and the Turks and Caicos Islands, and was moving west-northwestward at 8 kt on the southwestern flank of the 500-hPa Atlantic subtropical ridge A (Fig. 2a). This ridge was in the process of being eroded, as had been noted by NHC in their forecast discussion (available online at http://www.nhc.noaa.gov/archive/2011/al09/al092011.discus.011.shtml) at 0300 UTC 23 August: “A mid-to-upper-level trough located east of North America will lift out within 24 hours...leaving a pronounced weakness in the ridge over the Bahamas.” From the same NHC discussion, it was also noted that 2–3 days into the forecast period, “the track of Irene appears to be sensitive to the timing and amplitude of several shortwave troughs moving eastward across the United States/Canadian border.” In Fig. 2a, the upper-level trough B was present over Saskatchewan, Canada, and the short-wave trough C was situated over northern Nevada and southern Idaho. Both of these features were moving around a large mid- to upper-level anticyclone centered near the Four Corners region.

At 0000 UTC 25 August, Irene was located over the central Bahamas and began to move west-northwestward at 11 kt around the western edge of subtropical ridge A (Fig. 2b). By this time, trough B had moved into the Great Lakes while trough C was moving over the top of the ridge near the Montana–Wyoming border.

During the next 24 h, trough B over the Great Lakes had elongated and moved eastward into the northeastern United States, eroding the western edge of subtropical ridge A (Fig. 2c). As a result, Irene turned northward at 13 kt by 0000 UTC 26 August as it passed through the northwestern Bahamas. Shortwave trough C had moved eastward and was centered over Wisconsin by this time, while a new short-wave trough D near the coast of Oregon was beginning to move around the continental ridge.

By 0000 UTC 27 August, Irene was moving north-northeastward at 11 kt and approaching the coast of North Carolina. Trough C had moved eastward and become elongated as it impinged on the upper-level ridge associated with Irene (Fig. 2d). Upstream, trough
FIG. 2. GFS final analysis of 200–400-hPa potential vorticity (shaded, PVU) and layer-average wind (barbs, kt), 500-hPa geopotential height (black contours every 3 dam), and 900–700-hPa potential vorticity (red contours every 0.25 PVU starting at 0.50 PVU) valid at 0000 UTC (a) 24 Aug, (b) 25 Aug, (c) 26 Aug, (d) 27 Aug, (e) 28 Aug, and (f) 29 Aug 2011. Large black letters indicate features of interest discussed in the text.
D moved to the northern tip of the ridge over northern Montana. At 1200 UTC 27 August, Irene made landfall at 1200 UTC near Cape Lookout in North Carolina as a 75-kt hurricane (Avila and Cangialosi 2011).

Irene had moved offshore of the coast of Virginia by 0000 UTC 28 August and was moving north-northeastward at 13 kt. Short-wave trough D moved quickly south-eastward into the eastern Great Lakes, reinforcing the broad trough in this region (Fig. 2e). The outflow ridge from Irene continued to expand northward, as is evident in the large area where 200–400-hPa potential vorticity (PV) values were less than 1 PVU over much of New York and New England [1 potential vorticity unit (PVU) = \(10^{-6} \text{m}^2 \text{s}^{-1} \text{Kkg}^{-1}\)]. During the day on 28 August, the forward speed of Irene increased to 20–25 kt as the cyclone accelerated between the trough to the west and the subtropical ridge to its east. Irene weakened to a tropical storm by the time it made landfall near Brigantine Island, New Jersey, at 0935 UTC, and made its final landfall near Coney Island in Brooklyn, New York, at 1300 UTC with maximum winds of 55 kt.

By 0000 UTC 29 August, upper-level trough D amplified west of Irene and grew in size as it became negatively tilted, consistent with the amplification of the outflow ridge associated with Irene (Fig. 2f). The low-level vorticity pattern of Irene had clearly become distorted as the cyclone became extratropical while centered near the Vermont–New Hampshire border. Irene was absorbed by a larger extratropical cyclone over Quebec, Canada, shortly after 0000 UTC 30 August (Avila and Cangialosi 2011).

4. Evaluation of track forecasts

a. Operational GFS forecasts

As noted earlier, during Irene’s life cycle, the operational GFS track forecast errors for Irene were small compared with the 5-yr OFCL track forecast errors for the 2006–10 period (Table 1). During the portion of Irene’s life cycle that is the focus of this study (23–28 August 2011), the GFS had smaller errors than OFCL for Irene at 72–120 h (Fig. 3a). When the track forecast errors are decomposed into along- and cross-track components as described by Cangialosi and Franklin (2012), the GFS showed a slow along-track bias that was 40–50 nmi smaller than the bias shown by OFCL by days 4 and 5 (Fig. 3b). However, the GFS developed a pronounced right-of-track bias (i.e., away from the U.S. east coast) beginning at 72 h that grew to 40–60 nmi at days 4 and 5, while OFCL showed little bias through 96 h and a left bias (i.e., toward to the U.S. east coast) of about 40 nmi by 120 h (Fig. 3c). Since along-track errors represent difficulty in forecasting the timing of the arrival of the storm and its impacts, while cross-track errors imply difficulty in identifying where the impacts of the storm will occur, persistent cross-track biases may be considered to be as or more important than similar or even larger along-track errors when making operational forecasts of tropical cyclone track.

b. Error statistics of track forecasts in parallel cycles

The average NCEP GFS track forecast errors, evaluated in nautical miles against the NHC best track, are presented in Fig. 4a and Table 2 for each of the four parallel cycles. Forecasts initialized at 6-hourly intervals between 1800 UTC 22 August and 0000 UTC 28 August 2011 are included. For forecast lead times up to and including 36 h, the differences between the errors of the respective GFS forecasts were negligibly small. However, for forecasts of length 42–60 h, a noticeable difference emerged between the average errors in the CONTROL and NODROP runs. As an example, the assimilation of the additional dropwindsonde data served to reduce the 48-h track forecast errors from 58 to 45 nmi. However, in addition to the operational GFS described in the previous subsection, the GFS forecasts that withheld either or both sets of the supplemental data also possessed substantially lower errors than those of the 5-yr average NHC forecasts. This suggests that the GFS performed considerably better than average for Irene, and the skill of several other operational forecast models suggests that Irene was a case with relatively high predictability in its track forecasts. In contrast to the dropwindsondes, the supplemental rawinsondes only provided minor average improvements to the forecasts, and these were only noticeable beyond 84 h.

The statistical significance of these results is now examined. Using a one-tailed \( t \) test at the 95% level (Aberson and DeMaria 1994), the hypothesis that the track forecasts were improved on average due to the assimilation of the supplemental data is tested, for all forecast lead times up to 120 h. The null hypothesis is that there is no statistically significant average improvement. To determine an effective independent sample size in which the forecasts are not expected to be serially correlated, the method adopted by Aberson and DeMaria

---

3 Here, the operational GFS refers to the “interpolated” (i.e., GFSI) forecast from the previous model cycle, allowing a more fair comparison to the official forecast as described in section 2b of Brennan and Majumdar (2011).

4 For the remainder of this section, the noninterpolated NCEP forecast initialized at the listed time is used.
FIG. 3. (a) Average NHC operational (OFCL) and interpolated previous cycle GFS model (GFSI) Irene track forecast errors (n mi) for all forecast cycles from 0000 UTC 23 Aug through 0600 UTC 28 Aug 2011. (b) As in (a), but for along-track forecast errors and biases (n mi) for OFCL and GFSI where positive (negative) bias indicates that the forecast was too fast (slow) relative to the observed track. (c) As in (b), but for cross-track forecast errors and biases where positive (negative) bias indicates that the forecast was displaced to the right (left) of the observed track. The sample size for each forecast lead time is given below the forecast hour in parentheses.
FIG. 4. (a) Average track forecast errors (n mi) for all GFS Irene data-denial experiments as indicated by line types in the legend. (b) Average GFS Irene track forecast improvement–degradation (n mi) for 0000 and 1200 UTC model cycles compared with CONTROL, where positive (negative) values indicate improvement (degradation) relative to CONTROL. Asterisks refer to forecast times at which the improvements were statistically significant at the 95% level, with the shade of gray corresponding to the respective data-denial experiment denoted in the bar chart. (c) As in (b), but for 0600 and 1800 UTC model cycles. The sample size for each forecast lead time is given below the forecast hour in parentheses.
and explained in their appendix B is used. The effective separation time is deduced, and for the statistics summarized in Fig. 4a, it lies between 6 and 9 h. In addition to considering forecasts that are not serially correlated, it is also instructive to examine the two subsets of model cycles (0000 and 1200 UTC, 0600 and 1800 UTC) due to these considerable differences in the in situ observational network available at these times. Therefore, evaluations of samples initialized 12 h apart are conducted here. First, all forecasts initialized only at 0000 and 1200 UTC are considered in the sample. This sample includes all conventional rawinsonde data and the majority of dropwindsonde data assimilated at those times. Second, all forecasts initialized only at 0600 and 1800 UTC are considered. These forecasts only include the special supplemental rawinsonde data and a small minority of the dropwindsonde data at the initial time. Before evaluating the statistical significance, the impacts of assimilating the dropwindsonde data and the supplemental rawinsonde data are illustrated in Tables 3 and 4 and Figs. 4b and 4c for the two subsamples. Improvements to forecasts of duration 24 h or less are not discussed, given the tiny errors in the track forecasts. Through 72 h, there are only small differences between the average errors of the 0600/1800 UTC and 0000/1200 UTC cycles for all experiments (<12 n mi). Interestingly, the average 96-h track errors for the 0600/1800 UTC CTRL experiment are substantially smaller than the 1200/0000 UTC cycles, while the 0600/1800 UTC cycle NORAOB and NEITHER experiment errors are much larger at 120 h relative to the 0000/1200 UTC cycle errors.

The average percentage improvement over all 48-h forecasts initialized at 0000 and 1200 UTC due to the assimilation of the dropwindsonde data was 20% (Fig. 4b). This value is consistent with the average 48-h forecast improvements realized over the first decade of synoptic surveillance, in which dropwindsondes were centered on the 0000 and 1200 UTC cycles and forecasts initialized at those times were evaluated (Aberson 2010). Beyond 3 days, some average degradation to the track forecasts of Irene was evident due to the assimilation of the dropwindsonde data. The influence of the rawinsonde data appeared to be relatively negligible at all lead times. In contrast to those forecasts initialized at 0000 and 1200 UTC, virtually every forecast initialized at 0600 and 1800 UTC was improved by the assimilation of the supplemental data, for all lead times (Fig. 4c). The improvements are most noticeable at longer lead times, particularly for the supplemental rawinsonde data, although it should again be emphasized that the magnitude of the improvement is not large.

With regard to the impact of the dropwindsonde data (CONTROL versus NODROP) on forecasts initialized at 0000 and 1200 UTC, a statistically significant improvement was found only for forecasts of length 48 h. However, statistically significant improvements were found for forecasts of 36, 42, 72, and 84 h initialized at 0600 and 1800 UTC. For the impact of the supplemental rawinsonde data (CONTROL versus NORAOB), no statistical improvements were found for forecasts of any length initialized at 0000 and 1200 UTC. For forecasts initialized at 0600 and 1800 UTC, when these rawinsondes were launched, statistically significant improvements were found for forecasts of 84 and 114 h. Finally, the impact of assimilating both datasets (CONTROL versus NEITHER) was not found to be statistically significant for any forecasts initialized at 0000 and 1200 UTC. However, statistically significant improvements were found for 42-, 48-, 54-, and 60-h forecasts initialized at 0600 and

<p>| Table 2. Average Irene track forecast errors (n mi) for all GFS experiments from 1800 UTC 22 Aug through 0000 UTC 28 Aug 2011. |</p>
<table>
<thead>
<tr>
<th>Forecast hour</th>
<th>Expt</th>
<th>12</th>
<th>24</th>
<th>36</th>
<th>48</th>
<th>72</th>
<th>96</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTRL</td>
<td>16.7</td>
<td>27.9</td>
<td>37.3</td>
<td>44.7</td>
<td>68.0</td>
<td>103.8</td>
<td>142.7</td>
<td></td>
</tr>
<tr>
<td>NODROP</td>
<td>19.1</td>
<td>29.6</td>
<td>43.4</td>
<td>57.6</td>
<td>71.4</td>
<td>93.8</td>
<td>130.7</td>
<td></td>
</tr>
<tr>
<td>NORAOB</td>
<td>17.6</td>
<td>27.8</td>
<td>39.0</td>
<td>45.6</td>
<td>68.1</td>
<td>106.3</td>
<td>147.1</td>
<td></td>
</tr>
<tr>
<td>NEITHER</td>
<td>20.4</td>
<td>29.3</td>
<td>41.1</td>
<td>54.8</td>
<td>71.3</td>
<td>95.5</td>
<td>135.6</td>
<td></td>
</tr>
<tr>
<td>No. of forecasts</td>
<td>22</td>
<td>22</td>
<td>20</td>
<td>18</td>
<td>14</td>
<td>10</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

<p>| Table 3. Average Irene track forecast errors (n mi) for all 0000 and 1200 UTC cycle GFS experiments from 0000 UTC 23 Aug through 0000 UTC 28 Aug 2011. |</p>
<table>
<thead>
<tr>
<th>Forecast hour</th>
<th>Expt</th>
<th>12</th>
<th>24</th>
<th>36</th>
<th>48</th>
<th>72</th>
<th>96</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTRL</td>
<td>17.4</td>
<td>28.9</td>
<td>37.0</td>
<td>41.1</td>
<td>73.8</td>
<td>126.9</td>
<td>140.3</td>
<td></td>
</tr>
<tr>
<td>NODROP</td>
<td>21.2</td>
<td>32.6</td>
<td>43.0</td>
<td>56.9</td>
<td>71.1</td>
<td>97.0</td>
<td>111.9</td>
<td></td>
</tr>
<tr>
<td>NORAOB</td>
<td>17.4</td>
<td>29.4</td>
<td>39.9</td>
<td>39.9</td>
<td>72.6</td>
<td>115.4</td>
<td>139.0</td>
<td></td>
</tr>
<tr>
<td>NEITHER</td>
<td>20.7</td>
<td>30.5</td>
<td>40.0</td>
<td>51.2</td>
<td>74.9</td>
<td>99.7</td>
<td>114.4</td>
<td></td>
</tr>
<tr>
<td>No. of forecasts</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<p>| Table 4. Average Irene track forecast errors (n mi) for all 0600 and 1800 UTC cycle GFS experiments from 1800 UTC 22 Aug through 1800 UTC 27 Aug 2011. |</p>
<table>
<thead>
<tr>
<th>Forecast hour</th>
<th>Expt</th>
<th>12</th>
<th>24</th>
<th>36</th>
<th>48</th>
<th>72</th>
<th>96</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTRL</td>
<td>16.0</td>
<td>26.8</td>
<td>37.6</td>
<td>48.3</td>
<td>62.2</td>
<td>80.7</td>
<td>145.1</td>
<td></td>
</tr>
<tr>
<td>NODROP</td>
<td>17.0</td>
<td>26.6</td>
<td>43.9</td>
<td>58.2</td>
<td>71.7</td>
<td>90.7</td>
<td>149.4</td>
<td></td>
</tr>
<tr>
<td>NORAOB</td>
<td>17.8</td>
<td>26.2</td>
<td>38.1</td>
<td>51.2</td>
<td>63.6</td>
<td>97.2</td>
<td>155.1</td>
<td></td>
</tr>
<tr>
<td>NEITHER</td>
<td>20.0</td>
<td>28.0</td>
<td>42.2</td>
<td>58.4</td>
<td>67.8</td>
<td>91.2</td>
<td>156.7</td>
<td></td>
</tr>
<tr>
<td>No. of forecasts</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>
1800 UTC. No statistically significant degradations were found for any forecasts, as those degradations noted in Fig. 4b at longer lead times were for a very small sample (six or fewer cases). The main overall conclusion here is that a modest but statistically significant improvement was found due to the assimilation of both dropwindsonde and rawinsonde data, for 42–60-h forecasts initialized at 0600 and 1800 UTC. The inclusion of both datasets yielded marginally superior results to the addition of just one of the special datasets.5

Finally, given that Irene traveled parallel to a heavily populated stretch of the United States coastline, the importance of the cross-track error is magnified in this case. An examination of Fig. 5a illustrates a consistent reduction of cross-track errors for all forecasts verifying at 48 h and beyond due to the supplemental data. For forecasts ranging from 2 through 4 days, the quantitative cross-track improvement due to the dropwindsondes was consistently identical to that due to the rawinsondes. Furthermore, the reduction in error due to the assimilation of both types of supplemental data was greater than if only one type of supplemental data was assimilated. The average reduction in the cross-track errors due to both types of supplemental data ranged from 6 to 27 n mi for 2–5-day forecasts. The similarity of the positive cross-track forecast bias in Fig. 5b to the errors in Fig. 5a indicates a consistent right-of-track bias in the track forecasts, as well as a reduction of this bias due to both types of supplemental data. In contrast, the results are mixed for along-track errors, with degradations due to the dropwindsonde data for forecasts beyond 3 days (Fig. 5c). However, the overall forecast degradations at these lead times due to the dropwindsondes in the 0000 and 1200 UTC model cycles were not statistically significant (Fig. 4b). The forecasts from all four cycles possessed a slow bias (Fig. 5d), and the degradations

---

5 These results are also corroborated by comparing NODROP versus NEITHER and NORAOB versus NEITHER.
were found to be due to forecasts in the CONTROL cycle yielding too strong a ridge to the north of Irene, thereby slowing down the along-track component of Irene’s northward motion (not shown). Overall, the GFS forecasts with and without the supplemental observations assimilated usually possessed a general slow bias to the right of the observed track, consistent with Fig. 3.

c. Case 1: 0000 UTC 23 August 2011—Impact of dropwindsonde data

Of particular interest is the series of forecasts initialized at 0000 UTC August 23, given the relatively high uncertainty in the track forecasts and the initial deployments of supplemental observations at this time. One day prior to this time, the operational GFS forecasts exhibited a strong left bias, bringing the center of Irene close to northern Florida and into southeast Georgia. The subsequent three GFS forecasts shifted Irene eastward and then back westward, exemplifying the uncertainty and lack of run-to-run consistency during the early stages of Irene’s life cycle immediately prior to the first deployments of dropwindsondes and supplemental rawinsondes. In this section, forecasts from the CONTROL cycle, which included the dropwindsonde data from the G-IV and C-130 flights illustrated in Fig. 1, are compared against forecasts from the NODROP cycle, in which these dropwindsonde data were not assimilated.

A small but distinct westward shift in the track forecast due to the assimilation of the dropwindsonde data is evident in Fig. 6a. This westward shift is closer to the best track of Irene, with the most substantial improvements occurring for forecasts between 60 and 90 h (Fig. 6b). As a side note, a negligible influence on this forecast occurred due to the initial handful of rawinsondes that were launched 6 h prior.

In addition to the reduction of track forecast errors in the GFS model, the overall uncertainty as estimated through the spread of track forecasts in different operational models was also reduced. This was noted in real time, as seen in the NHC forecast discussion at 0900 UTC 23 August (http://www.nhc.noaa.gov/archive/2011/al09/al092011.discus.012.shtml): “The 23/00Z G-IV jet aircraft and Air Force C-130 dropsonde data appear to have settled down the models...and there is considerably less difference among the various model solutions now. The overwhelming consensus is that Irene will gradually turn northwestward...then move northward through a developing break in the subtropical ridge over the southeastern United States.”

At the initial time of 0000 UTC 23 August, the assimilation of the data from the supplemental dropwindsondes served to raise the 500-hPa geopotential height in a region about 1000 km to the north and northwest of Irene, corresponding to a slight westward extension of the Atlantic subtropical ridge (labeled A in Fig. 2) and a modest weakening of the southern flank of the trough off the southeast coast of the United States (Fig. 7a).

This particular modification was mostly due to the dropwindsonde data from the C-130 flight, which was designed such that it “will be sampling the weakness in the ridge northwest of Irene” (quoting an e-mail from Chris Landsea, NHC, 2011, prior to the mission). The trend toward higher geopotential heights in this broad area became more pronounced at +24 and +48 h (Figs. 7b and 7c). By +60 h, the westward shift in the track of Irene due to the strengthened ridge in the CONTROL forecast with dropwindsondes is evident. Aside from this ridge, no synoptic features including those upstream of Irene were substantially modified at later forecast times (Fig. 7d).

Although the positive influence of assimilating the dropwindsonde data was clearly evident in the GFS forecast initialized on 0000 UTC 23 August, the corresponding influence was relatively small for forecasts initialized on or after 24 August. Most of the modifications to the forecasts out to 3 days were along the track, with only a modest cross-track modification in each case. This smaller improvement in subsequent cycles could be due in part to the impact of the initial assimilation of the dropwindsondes carrying through into future analysis cycles through the first-guess field.

d. Case 2: 1800 UTC 26 August 2011—Impact of supplemental rawinsonde data

The second case illustrates an example of a moderate improvement to the track forecast of Irene due to the assimilation of data from the supplemental rawinsondes. Although there were no statistically significant improvements for the forecast times examined in this case, the analysis presented here illustrates the broad pattern of small impacts on the GFS that have accrued from the assimilation of the supplemental rawinsonde data, and the subsequent local propagation and growth of the impact.

The 0600 and 1800 UTC rawinsonde coverage was expanded to the eastern two-thirds of the CONUS at 1800 UTC 25 August 2011 (Fig. 1), and the initial time investigated here corresponds to the third such extensive deployment of rawinsondes. The hypothesis here is that the supplemental rawinsondes helped to analyze the trough features upstream of Irene, and that the subsequent merger or interaction of these troughs over the eastern United States may have influenced Irene’s track as it approached the Carolinas and the mid-Atlantic coast.

In the NORAOB cycle in which the supplemental rawinsondes were withheld from the GFS–GSI cycle,
the track forecast was farther to the right than the CONTROL (Fig. 8a). In other words, the assimilation of the supplemental rawinsonde data served to bring the forecast track closer to the observed best track closer to the U.S. east coast. For 36–54-h forecasts, the track forecast errors were reduced by up to 40 n mi (Fig. 8b). At other initial times, the assimilation of the supplemental rawinsonde data was not found to improve the track forecast by more than 20 n mi.

The factors governing this correction to Irene’s track forecast in the model are quite subtle and are more clearly evident in the upper-tropospheric potential vorticity (PV) fields than in the 500-hPa geopotential height patterns. First, the difference between the CONTROL and NORAOB analysis fields of 200–400-hPa PV in Fig. 9a reveals an initially subtle phase shift in the PV maximum over eastern Washington and northern Idaho, associated with short-wave trough D in Fig. 2d. The assimilation of the supplemental rawinsonde data acted to shift the initial position of the PV maximum very slightly to the east (downstream) in the analysis. Although the supplemental rawinsondes were not launched this far upstream, west of the Rockies, it is possible that the influence of assimilating supplemental rawinsonde data...
in earlier cycles propagated upstream and acted to produce the phase shift in the upper-tropospheric PV. This phase shift is also evident in the subsequent differences between the CONTROL and NORAOB forecasts, with the local PV maximum being situated slightly eastward over Montana at +6 h (Fig. 9b) and then farther southward over Minnesota at +18 h as the magnitude of the differences amplifies (Fig. 9c). The spatiotemporal continuity of the difference field is evident at intermediate times not shown here. By +30 h, the upper-tropospheric PV maximum has amplified considerably over the Great Lakes and Midwest, as short-wave trough D sharpens (Fig. 2e). The CONTROL run with the supplemental rawinsonde data possesses a slightly more amplified and negatively tilted trough in the Ohio Valley at this forecast time does than the NORAOB run (Fig. 9d). Although it cannot be stated with absolute certainty, we suggest that the influence of the supplemental rawinsonde data on the western flank of the trough at this time arises from the initial perturbation over eastern Washington and northern Idaho, while the sharpening and negative tilting influence on the eastern flank may arise from a stronger outflow to the northwest of Irene.

At this same time (+30 h), Irene is situated slightly farther northward in CONTROL than in NORAOB. However, the difference between the track locations as exhibited in the PV difference in Fig. 9e shows a more westward shift, consistent with the 36-h positions in Fig. 8a. This is the first time at which a noticeable difference between the track forecasts is evident. We suggest that this westward shift in CONTROL is due to the more negatively tilted trough identified 6 h earlier, which interacts more closely with Irene than that in NORAOB and thereby accelerates Irene’s forward motion and also brings the tropical cyclone farther west.
At +42 h (Fig. 9f) through to +54 h, this overall signal remains consistent, although the improved track forecast in CONTROL is still slower and farther offshore relative to the best track.

A comparison between these forecast fields and the corresponding operational analysis fields (not shown here) indicated that the PV maximum associated with the primary short-wave D was even stronger in the analysis. Hence, the assimilation of the supplemental rawinsonde data acted to modify the PV feature in the correct direction, although a more pronounced modification to this feature may have served to reduce the erroneously slow and offshore track forecast even further than that accomplished in CONTROL.

In summary, Figs. 8 and 9 illustrate that the cumulative impact of assimilating the supplemental rawinsonde data served to moderately curtail the progressively more erroneous slow, offshore drift that would have occurred in this selected forecast had the supplemental rawinsondes not been launched.
Of particular interest is the influence of the dropwindsonde and supplemental rawinsonde data on the GFS track forecasts valid at the times of Irene’s landfall in North Carolina at 1200 UTC 27 August and at 1200 UTC 28 August, the nearest synoptic time to the landfall in New Jersey at 0935 UTC that day. As
discussed previously, it is evident from Fig. 10 that nearly all forecasts beyond 24 h were too slow and often offshore (compared with the two black dots that indicate the landfall locations). However, the supplemental data did produce modest improvements in some cases, and degradations of comparable magnitude were not evident (Tables 5 and 6).

The assimilation of the dropwindsonde data generally served to shift the day 2 track forecasts valid at the time of the North Carolina landfall toward the northwest
(Fig. 10a), with the average track forecast error being reduced from 63 to 41 n mi. It is also evident that the day 2 and day 3 track forecasts valid at the times of the New Jersey landfall were improved due to the assimilation of the dropwindsonde data. The day 2 track forecast errors were reduced from 61 to 41 n mi, while the day 3 errors were reduced from 89 to 61 n mi. The modifications to longer-range forecasts were negligible for both landfalls.

The assimilation of the supplemental rawinsonde data made virtually no difference to the track forecasts valid at the time of the North Carolina landfall. For the New Jersey landfall, modest average improvements of 12 n mi were evident in the day 3–5 forecasts, mostly corresponding again to slightly faster tracks closer to the coast when the supplemental rawinsonde data were assimilated.

It is also worth noting that while there was run-to-run consistency in the actual track forecasts in the model, there was only a moderate run-to-run consistency in the forecast improvements in the GFS model track.

f. Hurricane Weather Research and Forecasting (HWRF) model

In addition to the GFS forecasts, forecasts using the 2011 operational version of NOAA’s regional Hurricane Weather Research and Forecasting (HWRF; Gopalakrishnan et al. 2011) model were also tested. At that time, HWRF comprised a single moving domain with 9-km horizontal grid spacing that followed the tropical cyclone, nested within a large, fixed parent domain with 27-km horizontal grid spacing covering a roughly 80° × 80° area on a rotated latitude–longitude E-staggered grid. Boundary conditions were provided by the GFS, together with a mesoscale vortex initialization procedure, resulting in an interpolated field on the HWRF parent and inner domains. The GSI was then used to assimilate all conventional data, including dropwindsondes and supplemental rawinsondes, into HWRF.

The HWRF track forecasts due to the data-denial experiments were found to be ambiguous and inconclusive, compared with the GFS forecasts. The track forecast errors were generally larger than those in the GFS, and no clear improvement was found in the HWRF forecasts of track due to the extra data. It is particularly important to note that the version of HWRF that is operational in 2013 is vastly different from the 2011 version, with very significant improvements to the nested grids, physics, initialization, and assimilation (V. Tallapragada 2013, personal communication). For this reason, and the inconclusive nature of the results, the 2011 HWRF results are not shown in this paper.

5. Conclusions and recommendations

During most of the life cycle of Hurricane Irene (2011), during which it posed a threat to the east coast of the United States, supplemental aircraft-borne dropwindsondes were deployed and rawinsondes were launched at 0600 and 1800 UTC in order to improve the skill and reduce the uncertainty of operational track forecasts of Irene. In this study, the influence of assimilating these supplemental data was evaluated and analyzed, using the NCEP Global Forecast System (GFS) and the data assimilation scheme that was operational at the time.

Even without assimilation of the supplemental data, errors in the GFS track forecasts were considerably lower than average, thereby leaving little room for improvement. Assimilation of the dropwindsonde data yielded small average improvements to 2–3-day track forecasts of up to 20%, consistent with the findings of Aberson (2010). Some degradation to the forecasts was evident at a few longer lead times, for a small sample of cases. The supplemental rawinsonde data yielded minor improvements overall for all forecasts and particularly for the 4–5-day range. Interestingly, the forecasts initialized at 0600 and 1800 UTC yielded an almost unanimous improvement due to the assimilation of the supplemental data, while degradations were evident at longer lead times for forecasts initialized at 0000 and 1200 UTC. The combined improvement of assimilating dropwindsonde and supplemental rawinsonde data was found to be statistically significant at the 95% level for 42–60-h forecasts initialized at 0600 and 1800 UTC. Importantly for this storm, given its almost parallel track to the U.S. coastline, the supplemental observations acted to consistently reduce the cross-track errors of forecasts beyond 2 days, via a small correction to the right-of-track bias.
Overall, while there was generally run-to-run consistency in the GFS forecasts, a consistent improvement from run to run was not evident, for example when verifications were performed near the time of the landfall locations in North Carolina and New Jersey. Given that the improvements were mostly very small, it was often difficult to separate the true impacts from expected statistical fluctuations. Individual cases such as that initialized at the time of the first significant deployment of dropwindsondes (0000 UTC 23 August) demonstrated a positive benefit, which was directly related to the dropwindsondes acting to strengthen the subtropical ridge to the north of Irene at that time. For the 1800 UTC 26 August initialization, the rawinsonde data cumulatively served to improve the representation of a midlatitude short-wave trough propagating eastward over the continental United States, thereby slightly reducing the slow and offshore bias in the GFS forecast.

It is worth emphasizing that this study was for only one tropical cyclone, in which the track forecast errors were considerably smaller than recent averages. Therefore, although some positive impacts had been found, the conclusions cannot be generalized. Additional cases, such as Hurricane Sandy in October 2012, require a similar detailed evaluation, and the impact of assimilating the supplemental data needs to be evaluated systematically each season with the latest versions of the operational model and data assimilation scheme. Moreover, with the increased focus on intensity forecasting via NOAA’s Hurricane Forecast Improvement Project (HFIP; Gall et al. 2013), data impact studies using the latest operational version of the HWRF model (Gopalakrishnan et al. 2013) with their own advanced data assimilation schemes will also be necessary.

For any study that assesses the impact of assimilating data on NWP, it is necessary to consider whether to use data “denial” versus “adding” a new data type to the assimilation. The concepts are similar but not identical, owing to differences in the combinations of observations that are used in the baseline runs. In this paper, we elected to use the denial approach, for consistency with convention (e.g., Aberson 2010) and since the supplemental data were assimilated into the operational system that was our baseline (CONTROL). An alternative approach is to evaluate the impact of assimilating the supplemental data using the NEITHER cycle as the baseline. Using this approach to compute the errors and test for statistical significance, our conclusions on the effects of assimilating the dropwindsonde data and the rawinsonde data were found to remain largely the same. When a new dataset is to be evaluated for its potential use in operations, the data “addition” approach would be most appropriate.

In parallel, new studies are required to identify optimal strategies for deploying supplemental observations, particularly in an era where reductions to the global observational network are being proposed. For example, the results in this paper suggest that supplemental observations may be more useful at 0600 and 1800 UTC, given the consistent improvements in those model cycles that otherwise typically lack in situ sounding data. Several objective techniques to “target” observations have been developed for short-range forecasts over the past decade, although their reliability has not been evaluated for a large number of tropical cyclone cases. The effectiveness of targeting observations to improve medium-range forecasts also requires investigation; in this regime, it is expected that multiple regions for targeting may be necessary (Brennan and Majumdar 2011; Majumdar et al. 2011a). In particular, several important areas for targeted observations may be outside the range of aircraft and even the rawinsonde network, and therefore the targeting of satellite data (e.g., doubling the density of satellite radiances in particular areas) may be useful (Berger et al. 2011; Bauer et al. 2011; Majumdar et al. 2011b). The relative importance of supplemental observations in high- versus low-predictability regimes also requires investigation. And finally, new observing strategies can be developed to directly improve forecasts of the tropical cyclone wind field (and therefore storm surge), as well as coastal and inland precipitation.

Acknowledgments. The first author gratefully acknowledges financial support from the National Science Foundation (Grant ATM-0848753) and the NOAA Hurricane Forecast Improvement Project. The authors thank Bill Lapenta of the NCEP/Environmental Modeling Center (EMC) for his support of this study and allocation of computational resources. Thanks are also extended to Chris Landsea and Richard Pasch of NHC and Vijay Tallapragada of EMC for reviewing a previous version of this manuscript. We are also grateful to Vijay Tallapragada for providing Hurricane Weather Research and Forecasting (HWRF) model forecasts nested within the parallel cycles for comparison. John Cangialosi of NHC kindly provided the statistics of along- and cross-track errors and biases. Finally, the authors acknowledge the tireless contributions of all personnel involved in the time-consuming process of planning and deploying supplemental observations ahead of Irene and other tropical cyclones.

REFERENCES


