10 Years of Hurricane Synoptic Surveillance (1997–2006)

SIM D. ABERSON
NOAA/AOML/Hurricane Research Division, Miami, Florida

(Manuscript received 10 June 2009, in final form 20 November 2009)

ABSTRACT

In 1997, the National Hurricane Center and the Hurricane Research Division began operational synoptic surveillance missions with the Gulfstream IV-SP jet aircraft to improve the numerical guidance for hurricanes that threaten the continental United States, Puerto Rico, the U.S. Virgin Islands, and Hawaii. During the first 10 yr, 176 such missions were conducted. Global Positioning System dropwindsondes were released from the aircraft at 150–200-km intervals along the flight track in the environment of each tropical cyclone to obtain wind, temperature, and humidity profiles from flight level (about 150 hPa) to the surface. The observations were processed and formatted aboard the aircraft and sent to the National Centers for Environmental Prediction and the Global Telecommunications System to be ingested into the Global Forecast System, which serves as initial and boundary conditions for regional numerical models that also forecast tropical cyclone track and intensity. The results of an observing system experiment using these data are presented.

1. Introduction

The need for additional data acquisition over data-sparse tropical oceans to improve tropical cyclone analysis and forecasting has long been known (e.g., Riehl et al. 1956). Between 1982 and 1996, the National Oceanographic and Atmospheric Administration (NOAA) Hurricane Research Division (HRD) conducted 20 “synoptic flow” experiments to gather observations in the tropical cyclone core and environment in the North Atlantic basin (Burpee et al. 1996). The NOAA WP-3D (P-3) research aircraft released Omega dropwindsondes (ODWs; Govind 1975) to obtain wind, temperature, and humidity profiles below about 400 hPa within 1000 km of the tropical cyclone center. The dropwindsonde observations produced significant improvements in the primary numerical guidance for the National Hurricane Center (NHC) official track forecasts (Burpee et al. 1996). These improvements (16%–30% for 12–60-h forecasts) were as large as those from NHC official forecasts during the previous 20–25 yr and suggested that operational missions would be effective in operational numerical forecast error reduction.

In 1996, NOAA procured a Gulfstream IV-SP jet aircraft (G-IV), and put it to use in operational “synoptic surveillance” missions in the environments of tropical cyclones that threaten the continental United States, Puerto Rico, the U.S. Virgin Islands, and Hawaii. A new dropwindsonde, based on the Global Positioning System, was developed by the National Center for Atmospheric Research to replace the ODW (Hock and Franklin 1999). The first 2 yr of surveillance (1997 and 1998), during which 5 and 19 missions were conducted, respectively, and in which symmetric sampling in all regions of the environment was employed, has been discussed previously (Aberson and Franklin 1999; Aberson 2002). Only modest changes to model forecasts were seen, though individual case studies suggested that the difficulty in operational current storm-motion vector quantification for use in the synthetic vortex procedure was the primary cause of the small average forecast improvements. In those cases in which the operational and postprocessed storm-motion vectors were similar, the surveillance missions showed statistically significant improvements of up to 25% in the Geophysical Fluid Dynamics Laboratory (GFDL) hurricane model. In the complementary subset, the forecasts in all three models tested are degraded. Another important factor in the improvement size was the amount of data coverage: those cases in which data were provided by more than one aircraft had larger improvements, in general, than those in which only the G-IV provided data.
Because of limited aircraft resources, optimal observing strategies for these missions were investigated. In Aberson (2003), the most rapidly growing modes, or targets, were represented by areas of large forecast spread in the NCEP bred-vector ensemble forecasting system; the observing strategy required regularly spaced sampling of the entire target. Assimilation of the data subset from fully sampled targets produced a statistically significant track forecast error reduction of up to 25%. This technique, therefore, produced larger improvements in dynamical model tropical cyclone forecast tracks than were possible by assimilating all available data from the missions. This result highlighted the suboptimality of data assimilation schemes used at the time. The main region of failure of current techniques seems to be at the edges of well-sampled regions. Other targeting techniques have also been investigated (Majumdar et al. 2006; Reynolds et al. 2007).

Aberson (2008) found the causes of other forecast failures due to surveillance. These included the assimilation of erroneous data into the numerical guidance, the assimilation of nonrepresentative dropwindsonde data in the tropical cyclone eyewall into the guidance, and global impacts of data due to the spectral data assimilation and model systems. Modest changes to the operational data processing, quality control, and data assimilation systems have occurred to correct these problems.

Though the surveillance program fully underwent transition to operations by 2007, research into this program has continued. Aberson and Etherton (2006) compared the assimilation of dropwindsonde data into a barotropic model using three-dimensional variational and ensemble transform Kalman filter techniques in two cases in Hurricane Humberto (2001). Qu and Heming (2002) reported that, in a small sample of cases from Hurricane Debby (2000), the dropwindsonde moisture data were as important as the wind vector data in improving forecasts in the Met Office (UKMO) model, with the temperature data being relatively unimportant. Further research on targeting will be reported elsewhere.

The following is an updated impact assessment of numerical track and intensity guidance during the first 10 yr of G-IV operations. General descriptions of the two model systems (GFS and GFDL) and upgrades to them during the 10-yr period are provided in the next section. An in-depth assessment of the impacts from the surveillance missions on model track and intensity forecasts are provided in section 3. Since the initial and boundary conditions for the GFDL are largely based on the GFS, the results of the two models are compared in section 4. Section 5 has a description of the relationship between GFDL track and intensity forecast improvements because of surveillance. Some interesting case studies are presented in section 6, and the final section contains conclusions and a description of future research directions.

2. Overview and procedures

The list of storms around which surveillance missions occurred, and number of such missions around each storm, are provided in Table 1. The results of the 1997 (Aberson and Franklin 1999), 1998 (Aberson 2002), and the Humberto (2001) missions (Aberson and Etherton 2006) have already been reported. During each mission, the G-IV released 25–30 dropwindsondes to sample the atmosphere below flight level (near 150 hPa) at 150–200-km intervals. In those cases in which one or two P-3 or Air Force C-130 aircraft supplemented the G-IV data, 20–25 dropwindsondes were released at the same horizontal resolution from around 400 (P-3) or 300 hPa (C130). The G-IV did not penetrate the inner core of any of the tropical cyclones during surveillance missions, though when the P-3s flew, at least one usually gathered data near the center. Until 2006, HRD meteorologists aboard the aircraft validated the wind and thermodynamic data and generated standard (TEMPDROP) messages for transmission to the National Centers for Environmental Prediction (NCEP) for assimilation into numerical models; starting in 2006, this task was carried out by NOAA/Aircraft Operations Center meteorologists in many cases.

To assess the surveillance impact, the version of the NCEP Global Forecast System (GFS) operational at the time of each mission was used. During the 10 yr, several major upgrades to the quality control algorithm, the tropical cyclone vortex initialization procedure, the data assimilation, and the Global Spectral Model were made. The quality control algorithm involves optimal interpolation and hierarchical decision making to evaluate the observations before input to the analysis (Woollen 1991). Until 2000, the synthetic data procedure (Lord 1991) created observations representative of the tropical cyclone at mandatory levels between 1000 and 300 hPa within 300 km of the storm center, based on operationally estimated position, intensity, and motion inputs as well as nearby observations. This was replaced by a vortex relocation procedure (Liu et al. 2000) in which tropical cyclones in the first-guess field are relocated to the official Tropical Prediction Center position in each 6-h analysis cycle (as in Kurihara et al. 1995) and was refined slightly the next year.

The analysis scheme was the spectral statistical interpolation (Parrish and Derber 1992); the background field (the previous 6-h forecast) is combined with observations with a three-dimensional variational multivariate
formalism. During the 10 yr of this study, radiances and other satellite data such as QuikSCAT winds were added to the system. Before 1998, changes in time interpolation, nonlinear interpolation, and background error covariances were made to the assimilation scheme. In October 2000, new observation error statistics were incorporated and rawinsonde balloon drift in time and space were included in the system. By 2003, background errors were recomputed, and the divergence tendency constraint in tropics was removed. Other minor changes to the assimilation system were implemented in 2004.

In 1997, the GFS horizontal resolution was spectral triangular 126 (T126) with a 284 × 190 Gaussian grid, or an approximately 1° latitude–longitude grid, and the vertical coordinate extended from the surface to about 2.7 hPa with 28 (L28) unequally spaced sigma levels on a Lorenz grid (Caplan et al. 1997; Surgi et al. 1998). The resolution was increased to T170L42 in 2000, to T254L64 (for the first 84 h of model integration) in 2002, and to T382L64 in 2005. Minor changes to model physics were incorporated during this time.

Impact evaluations are also performed on the GFDL model (Kurihara et al. 1998) that uses the GFS initial and forecast fields as initial and boundary conditions, respectively; the operational GFDL model does not have a data assimilation procedure. The GFDL model version operational at the time of each mission is also used. The model modifies the GFS analysis near the storm center with its own vortex initialization scheme; the environmental initial conditions are not modified except near the vortex where a blending between it and the environment is necessary. A summary of updates to the model is provided in Bender et al. (2007).

Two runs of each model are made for each mission: one with no dropwindsonde observations assimilated (hereafter “parallel”), the other with them included (hereafter “operational”). All other observations from the NCEP “final” archive were ingested into the assimilation system for both sets of runs. The results at the synoptic (nominal) times for which each mission was conducted in 1997 and 1998 are identical to those in Aberson (2002). Off-nominal time results are presented only for 1998–2006 cases.

Surveillance missions are usually conducted when a tropical cyclone approaches land, usually starting when a threat during the forecast period is identified and logistics allow the mission, and ending either about 24 h before final landfall, when all preparations should be completed, or when the threat no longer exists. In Aberson (2002, 2003), each surveillance mission was assessed individually: dropwindsonde data obtained in previous missions were not used in the parallel run for the mission being assessed. Since 1999, assessments have mimicked...
the operational data assimilation and forecast cycle: the parallel runs are initiated at the time when the first surveillance data are assimilated into the model for a particular storm or set of storms and continue until 12 h after the last mission nominal time, thus assuming that none of the surveillance missions occurred. Surveillance missions were originally conducted once daily with a 0000 UTC nominal time; the addition of a second aircraft crew in 2002 now allows for twice-daily (0000 and 1200 UTC nominal time) missions, though this frequency is not a requirement. The results reported below are a mixture of the two methods. Using only data from 1999 to 2006 produces similar results.

3. Results

a. Mission nominal time results

Figure 1 shows the average impacts on GFS and GFDL model forecasts for the runs initialized at mission nominal times. Track error is the great-circle distance between the forecast and the concurrent NHC postprocessed best-track positions. Improvements of 10%–15% due to the surveillance missions are seen through 60 h in the GFS (statistical significance >98% through 36 h with serial correlation removed; Aberson and DeMaria 1994); the first 60 h is the critical watch period, and 36 h is the critical warning period, before landfall when preparations must be completed. Differences between the average errors in the two sets of runs are negligible after 72 h. This suggests the possibility of a model predictability limit, or that medium-range track forecasts are impacted more by distant features not sampled by the surveillance missions than by local features. In the GFDL model, the average track forecast improvements increase in time from 24 to 84 h though none of the differences are statistically significant. The GFDL vortex initialization procedure eliminates much of the surveillance data within the tropical cyclone circulation; the data impact remains in the GFS forecast that serves as boundary conditions for the GFDL model, allowing the impact to grow with time in that model. The data impact on GFDL model track forecasts is negligible by 4 days into the forecast as its impact in the GFS wanes.

Figure 2 shows the surveillance data impact on intensity (maximum 1-min sustained wind speed at 10 m above the surface) forecasts for model runs initiated at mission nominal times. The error is the absolute difference between the forecast and the concurrent NHC postprocessed best-track intensities. The GFS intensity forecasts are improved at all forecast times, and the improvements are statistically significant at the 98% level through 84 h. This is likely because the surveillance missions generally surround the vortex and help anchor a region of cyclonic winds in the model initial condition allowing for better representation of the circulation intensity than without them. However, the errors are large at all forecast times due to the difficulty of predicting intensity with a relatively low-resolution global model. Because these forecasts quantitatively lack skill, they are generally not used in operations except to assess possible intensity trends. The GFDL model intensity
forecast errors are smaller than those from the GFS, though the surveillance data degrade the GFDL model intensity forecasts at all forecast times except 24 h; the improvement at 24 h is statistically significant at the 98% level.

b. Results at all initialization times

Figures 3 and 4 show the surveillance data impact on all (mission nominal times and those between) track and intensity forecasts, respectively; the results shown in the previous section are a subset of these. GFS track forecasts are improved by 5%–10% during the critical watch and warning periods before landfall; average GFDL model track forecast impacts are mixed. None of the track improvements are statistically significant. As in the mission nominal time sample, substantial GFS intensity forecast improvements are seen, and these are statistically significant at 98% through 72 h. GFDL model intensity forecast impacts are mixed, though statistically significant (98%) 5%–10% improvements are evident from 96 h onward. These impacts are less positive than those at the mission nominal times.

c. Data assimilation/forecast cycle results

Because the operational data assimilation and forecast cycle interval is 6 h and each surveillance mission lasts about 8 h, the accumulated data from an individual mission are assimilated at two model initialization times, though each datum is assimilated in only one cycle. About three-fourths of the surveillance data are assimilated at the nominal time (hereafter “0 h”), with the remainder assimilated into the preceding cycle (hereafter “−6 h”). The 0-h sample therefore accounts for the cumulative impact of all the dropwindsonde data assimilated into both cycles. Those cases 6 and 12 h after a mission nominal time that do not also qualify as being in other −6- and 0-h samples, are hereafter classified as “+6 h” and “+12 h,” respectively. The 0- and −6-h samples have more than 150 cases at 12 h and more than 70 cases at 120 h; the 6- and 12-h samples have at least 100 cases at 12 h and 40 cases at 120 h.

One may expect the data impact to be largest at mission 0 h and to recede as the observational data for both runs becomes identical (+6 and +12 h). In both models, for track (Figs. 5a,b), the average positive impact of all surveillance data (0 h) is greater than that of the data subset assimilated into the −6-h cycles. This is also the case for the GFS intensity forecasts except at 96 h (Fig. 5c), but the GFDL model intensity forecast results are mixed (Fig. 5d). The +6-h GFS forecast improvements are larger than those in the 0-h sample at all forecast times except 60–84 h (track; Fig. 5a) and 12, 24, and 48 h (intensity; Fig. 5c), possibly due to residual improvements to the first-guess fields in the following model cycle. The +6-h GFDL model forecast track improvements are smaller than the 0-h improvements at all forecast times except 60 h (Fig. 5b), since the possible residual first-guess improvements may be removed by the vortex procedure, though the +6-h GFDL model forecast intensity improvements are larger than those in the 0-h sample at all forecast times except 12, 108, and 120 h (Fig. 5d). The average +12-h GFDL model impact is negative at all forecast times for track (Fig. 5c) and for

---

**FIG. 3.** As in Fig. 1, but for those cases initialized at mission nominal and intermediate synoptic times.

**FIG. 4.** As in Fig. 2, but for those cases initialized at mission nominal and intermediate synoptic times. Sample sizes are the same as in Fig. 3.
intensity except at 96–120 h (Fig. 5d). The +12-h GFS forecast improvements are comparable to or larger than those in the +6-h sample at all forecast times (intensity; Fig. 5c) and from 60 h onward (track; Fig. 5a). The data therefore have a more lasting impact in the GFS, which assimilates the data, than in the GFDL model, which uses the GFS for initial and boundary conditions.

d. Cumulative effect of missions

Figures 6 and 7 show the cumulative impact of missions on track and intensity forecasts for missions conducted in the same storm separated by 24 and 12 h, respectively. Cases classified as “first” are followed, but never preceded, by another case at the specified frequency (12 or 24 h). Those labeled “second” are the second in such a series, and so on. In each set, the first and second samples are homogeneous since the first missions must be followed by a subsequent one. The sample size decreases thereafter as the number of consecutive missions increases. Only those samples with at least eight cases are shown. Because of the small sample sizes, statistical significance tests are not performed.

One might expect that as the cumulative amount of data assimilated into the models increases with subsequent missions, the forecast improvement sizes would also increase; however, given that the impact of each mission decreases in forecasts after the mission nominal time in the GFDL model, this may not be the case. For track forecasts from both models, the impact decreases from the first to the second missions separated by 24 h (Figs. 6a,b), but increases for those separated by 12 h (Figs. 7a,b). For GFS intensity forecasts, the impact increases in time for the missions separated by 24 h (Fig. 6c),
but, for the missions separated by 12 h (Fig. 7c), it increases in time only in the short range. For GFDL model intensity forecasts, the cumulative impact of the missions separated by 24 h is mixed (Fig. 6d), but increases as data accumulates for missions separated by 12 h (Fig. 7d). This suggests that, for optimal impact, future missions should be conducted every 12 h, when possible, but a larger sample size is necessary to make a conclusive recommendation.

4. Relationships between GFS and GFDL model track forecast impacts

Because the GFDL model uses GFS initial and forecast fields as boundary conditions, track forecast impacts of the two models due to surveillance missions are expected to be positively correlated. The correlations between these two datasets is >0.25 for the 36–60-h forecasts, when the surveillance missions improve both GFS and GFDL model track forecast. At all other forecast times except 120 h, the correlations are positive, though small.

Cases are classified as to whether they show consistent improvement or degradation. Those cases in which 9 or 10 of the 10 forecast times (12–120 h, every 12 h) are improved are considered to be improvements; those in which none or only one forecast time shows a forecast improvement are considered degradations. Cases with no 120-h verification (fewer than 10 forecast times are available) are not classified. Because GFS intensities are mainly improved but have no skill, they are not considered here. Table 2 shows the number of cases in which the forecast at 0, 1, 9, or 10 forecast times was improved for each model and parameter (track or intensity) by the surveillance data. Also presented are the numbers of

---

**Fig. 6.** Forecast improvements (%) for the first, second, and third surveillance cases separated by 24 h in sequence with no intermediate missions for (a) GFS track, (b) GFDL model track, (c) GFS intensity, and (d) GFDL model intensity forecasts. Numbers of cases in each sample are shown in parentheses in (a).
times in each set that the other model or parameter is improved 0, 1, 9, or 10 times because of the surveillance missions. GFS track forecasts are found to be improvements (degradations) a total of 19 (16) times by these standards. The GFDL model track forecasts had the same number of improved forecasts, but only 11 degraded forecasts. The GFDL model intensity forecasts are considered improvements in six cases and no degradations are found. The off-diagonal elements show that GFDL model track forecasts are about as likely to be degraded when the GFS track forecasts are improved by the surveillance missions as when they are degraded, and vice versa; this confirms that, though the correlations are small, a strong relationship between GFS and GFDL model track forecast improvements does not exist.

Individual case studies show that this unexpected relationship may be due to differences in the evolution of the large-scale flow in the two models. Figure 8 shows

![Graphs showing comparison between GFS and GFDL forecasts](image)

**Table 2.** The number of GFS track and GFDL track and intensity forecasts in which either all or 9 of 10 (improvements) or in which 1 or none (degradations) show improvement because of a surveillance mission. Blank cells are those that are impossible.

<table>
<thead>
<tr>
<th></th>
<th>GFS track improvement</th>
<th>GFS track degradation</th>
<th>GFDL track improvement</th>
<th>GFDL track degradation</th>
<th>GFDL intensity improvement</th>
<th>GFDL intensity degradation</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFS track improvement</td>
<td>19</td>
<td></td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>GFS track degradation</td>
<td></td>
<td>16</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>GFDL track improvement</td>
<td>3</td>
<td>5</td>
<td>19</td>
<td>1</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>GFDL track degradation</td>
<td>3</td>
<td>3</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GFDL intensity improvement</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>GFDL intensity degradation</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Fig. 7.** As in Fig. 6, but for cases separated by 12 h.
the operational (all dropwindsonde) and parallel (no dropwindsonde) GFS and GFDL model track forecasts for Hurricane Isidore initialized at 0000 UTC 21 September 2002. Though the initial conditions of the operational GFS and GFDL models were similar to each other, as were those of the parallel GFS and GFDL models, the model trajectories were quite different. The surveillance data in the Gulf of Mexico and east of Florida improved (degraded) the GFS (GFDL) model track forecast at all forecast times through 120 h. Isidore moved westward south of the subtropical ridge over the southeastern United States during the first two days of the forecast; it completed a cyclonic loop over the Yucatan Peninsula before moving northward toward Louisiana as the ridge weakened. The surveillance data improved the GFS forecast by slowing the forward motion and moving the storm farther to the south in the operational than in the parallel run. An anticyclonic circulation in the GFS 850–200-hPa deep-layer-mean (DLM) wind (not shown) was initially centered over South Carolina in the operational run and over southeastern Georgia in the parallel run. This circulation weakened into an elongated ridge across the northern Gulf of Mexico in the operational GFS by 12 h into the forecast, allowing for a relatively slow westward motion; the circulation remained strong in the parallel GFS, leading to a more rapid westward motion than in the operational run. By 84 h into the forecast, the ridge dissipated in the operational GFS allowing for northward motion, whereas the ridge remained intact in the parallel GFS, causing the westward motion to continue. Conversely, the surveillance data degraded the operational GFDL model forecast by accelerating the westward motion and moving the storm more to the north late in the forecast than in the parallel run. The anticyclonic circulation over the southeastern United States was located in the same location and had the same intensity in the operational and parallel model pairs, as expected (not shown). Unlike the operational GFS, the operational GFDL model strengthened the ridge, keeping it north of Isidore through 96 h into the forecast, whereas the parallel GFDL model run forecast a weakening ridge, allowing for a relatively slow westward motion and an earlier northward turn than the operational run.

Figure 9 shows the two sets of model track forecasts for Hurricane Floyd initialized at 0000 UTC 12 September 1999, a case in which the surveillance data improved the GFDL model forecast and degraded the GFS forecast at all forecast times. For both models, the operational forecast tracks were to the south of those from the parallel
runs, resulting in forecasts that were farther west and slower than the parallel runs. The parallel GFS forecast was along the best track but slow, so the relatively slow operational run had a degraded forecast. The parallel GFDL model forecast was north and east of the best track and fast, so the relatively slow and westward operational GFDL model forecast showed an improvement. Since the parallel GFS and GFDL model run forecast errors are comparable, especially during the first 48 h of the forecast, the differences in this case may be due to chance, but still contributes to the unexpected correlation.

5. Relationship between GFDL model track and intensity forecast improvements

Since much tropical cyclone intensity change is based on synoptic conditions (DeMaria and Kaplan 1994) and good track forecasts are required for the model tropical cyclone to interact with the correct synoptic environment, a positive correlation between track and intensity impacts may be expected. Because GFS intensity forecasts are generally poor, and forecast improvements are due mostly to improved intensity initialization, this correlation in the GFS is very small. Correlations between the GFDL model track and intensity forecast improvements are negative at most forecast times, and are \(-0.25\) for 98–120-h forecasts. This may be due to small differences in land interactions close to landfall when the surveillance missions are usually conducted, or because the model resolution is not sufficient to accurately forecast intensity. Also, because the track forecasts from the operational and parallel models are not radically different from each other in most cases, differences in the forecast environment may be subtle.

The Hurricane Frances forecast initialized at 1200 UTC 4 September 2004 is the most extreme example of this negative correlation. The surveillance data caused the operational forecast to be south and west of the parallel run, degrading it at all forecast times (Fig. 10a). However, the operational GFDL model storm was weaker than the parallel run at all forecast times. This is due to the operational GFDL model forecasting the first landfall a few hours before the parallel, limiting intensification and leading to the intensity forecast improvement.

6. Successes

Aberson (2008) presented cases in which surveillance missions degraded GFS track forecasts. The following is an examination of cases with the largest GFS track and GFDL model track and intensity forecast improvements in the sample. The only case in which the GFS and GFDL model track forecasts were both improved by the surveillance data at all forecast times through 120 h was Tropical Storm Katrina initialized at 0000 UTC 25 August 2005 (Fig. 11), a little less than 24 h before landfall in South Florida. The surveillance data accelerated Katrina westward allowing for a slight GFS, and a large GFDL model, track forecast improvement. The forecasts all had a slow motion bias, leading to poor track forecasts. Despite track forecast improvements, the additional data degraded the GFDL model intensity forecast at most forecast times. Because the forecasts were too slow, the model storm had about 24 h longer to intensify over the Gulf Stream than the real Katrina allowing for a stronger model storm than reality through 48 h. Though the forecast tracks were forecast to be over south Florida for too short a time period to cause significant weakening, the parallel GFDL model run was closer to Florida from 24 to 48 h into the forecast than the operational model, and the increased land interaction caused it to forecast a relatively weak storm. These two issues caused the additional data to degrade the short-range intensity forecasts. Because both GFDL model runs failed to forecast the rapid intensification in the Gulf of Mexico, and because the operational model run provided a higher intensity forecast than the parallel run throughout the forecast period, the additional data improved the intensity forecast toward the end of the forecast period.

The previous case shows that, though the surveillance data may, on average, improve forecasts, they do not guarantee that the forecasts will be good. The Hurricane Wilma forecast initialized at 1200 UTC 20 October 2005 (Fig. 12) shows a wide range of forecast accuracy in a single case. The operational GFS provided a very good forecast for a late-season recurving hurricane, accurately forecasting slow movement near the Yucatan Peninsula followed by rapid northeastward motion and landfall in Southwest Florida. In contrast, the operational GFDL model produced a poor forecast: the midlatitude trough bypassed the hurricane and left it to meander in the northwestern Caribbean Sea. The surveillance data caused the forecast tracks from both models to be faster than those without, improving the forecast at almost all forecast times. The GFDL model intensity forecasts were similar to each other, except that the operational GFDL model forecast was initially about 10 kt stronger than the parallel forecast. This bias remained throughout the forecast, allowing for the surveillance data to improve the intensity forecast at almost all forecast times. This intensity bias is attributable to the GFDL vortex initialization scheme.

7. Discussion

The entire sample of 176 operational surveillance missions conducted by NOAA in the Atlantic and an eastern and central Pacific basin from 1997 to 2006 is
FIG. 10. Operational and parallel (GFDL and GFNO, respectively) GFDL model (a) track and (b) intensity (kt) forecast and best tracks for Hurricane Frances at 1200 UTC 4 Sep 2004. Values are shown every 12 h from 0 to 120 h.
examined. The missions led to 10%–15% improvements in GFS track forecasts during the critical watch and warning period before possible landfall (within the first 60 h), and small impacts in GFDL track and intensity forecasts, at mission nominal times. These improvements are not as large when all forecasts between mission nominal times are included in the sample, and the dropwindsonde data impact decreases after the mission nominal times. Conducting surveillance missions every 12 h, as is now possible with the second aircraft crew, seems to mitigate this decline in effectiveness, though a larger sample is needed to confirm this result. A complete

![Diagram](image-url)

**Fig. 11.** As in Fig. 10, but for Tropical Storm Katrina at 0000 UTC 25 Aug 2005.
G-IV flight and the required dropwindsondes cost about $40 000 (U.S. dollars), far less than the estimated $1 million (U.S. dollars) average needed to evacuate just 1 mile of coastline for landfall (Aberson et al. 2006). Because of this program’s success, a similar program, Dropwindsonde Observations for Typhoon Surveillance near the Taiwan Region (DOTSTAR), was implemented in 2003 (Wu et al. 2005, 2007). Further research
into improved data collection and assimilation techniques is recommended. For example, ordering additional rawinsonde ascents over land where dropwindsondes are not released when tropical cyclones threaten landfall may further improve forecasts. Other research into optimal sampling techniques is ongoing.

Starting at the beginning of the 2007 hurricane season, the surveillance program was fully transitioned to operations at the National Hurricane Center and Aircraft Operations Center. In previous years, scientists at the Hurricane Research Division drew flight tracks, staffed the NOAA aircraft participating in the surveillance missions, and assessed the data impact while doing research on optimal observing strategies to improve forecasts. Starting with Hurricane Dean in August 2007, these responsibilities were transferred to operational NOAA centers.

Acknowledgments. The author thanks NCEP/EMC for help in running the models and in providing the computer resources to make this study possible. Tim Marchok and Morris Bender of NOAA/GFDL ran many of the GFDL cases and provided many helpful comments, and Kao-San Yeh and Xuejin Zhang for helpful comments on an earlier version of the manuscript. Other fruitful discussions were held with Sharan Majumdar and Brian Etherton. Bob Kohler and Bill Barry provided computer support at HRD. The author also thanks the NOAA/Aircraft Operations Center (AOC) flight crews, AOC G-IV project manager Jack Parrish, and HRD personnel who participated in the flights, in addition to Air Force C-130 crews that also provided surveillance data over the years. Two anonymous reviewers improved the presentation of the material.

REFERENCES