NOTES AND CORRESPONDENCE

Two Years of Operational Hurricane Synoptic Surveillance

SIM D. ABERSON

NOAA/AOML/Hurricane Research Division, Miami, Florida

21 March 2002 and 20 June 2002

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 Virgin Islands, and Hawaii. During the first two years, 24 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 profiles of wind, temperature, and humidity from flight level (nearly 150 hPa) to the surface. The observations were processed and formatted aboard the aircraft and sent to NCEP to be ingested into the Global Data Assimilation System, which subsequently served as initial and boundary conditions for a number of numerical models that forecast the track and intensity of tropical cyclones. The current study is an attempt to mimic this process to assess the impact of these operational missions on the numerical guidance. Although the small number of missions flown in 1997 showed error reductions of as much as 32%, the improvements seen in the 2-yr sample are not promising. The additional dropwindsonde data from the synoptic surveillance missions provided statistically significant improvements in the GFDL forecasts only at 12 h. The ''VBAR'' and Global Forecast System (AVN) forecasts were not significantly improved at any forecast time. Further examination suggests that the AVN synthetic vortex procedure, combined with difficulty in the quantification of the current storm-motion vector operationally, may have caused the mediocre improvements. Forecast improvements of 14%–24% in GFDL forecasts are shown in the subset of cases in which the synthetic vortex data do not seem to be a problem. Improvements in the landfall forecasts are also seen in this subset of cases. A reassessment of tropical cyclone vortex initialization schemes used by forecast centers and numerical modelers may be necessary.

1. Introduction

The need for additional data acquisition over the data-sparse tropical oceans to improve analysis and forecasting of tropical cyclones has long been known (e.g., Riehl et al. 1956). Between 1982 and 1996, the National Oceanic 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) to obtain profiles of wind, temperature, and humidity 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). The improvements (16%–30% for 12–60-h forecasts) were as large as the improvement in NHC official forecasts obtained during the previous 20–25 years and suggested that operational dropwindsonde missions would be effective in the reduction of operational forecast errors.

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 Virgin Islands, and Hawaii. A new dropwindsonde, based on the global positioning system, was developed by the National Center for Atmospheric Research (NCAR) to replace the ODW (Hock and Franklin 1999). The first year of surveillance (1997), during which only five missions were conducted, has been discussed previously (Aberson and Franklin 1999). Dropwindsonde observations improved the set of track forecasts from the Geophysical Fluid Dynamics Laboratory (GFDL) hurricane model by as much as 32% during the hurricane warning period (within 36 h of projected landfall) in this small sample. Forecasts from another dynamical tropical cyclone model (''VBAR'') also showed modest improvements with the

Corresponding author address: Dr. Sim D. Aberson, NOAA/AOML/Hurricane Research Division, 4301 Rickerbacker Cswy., Miami, FL 33149.
E-mail: sim.aberson@noaa.gov

TABLE 1. Synoptic surveillance missions conducted during the 1997 and 1998 hurricane seasons. The nominal time is the synoptic time for the mission (0000 UTC). The initial position and intensity are from the NHC best track.

Storm name	Position	Intensity ($m\ s^{-1}$)	Storm-motion vector: Operational	Storm-motion vector: Best track	Nominal time
Claudette[d]	34.9°N 71.8°W	17	15°/8 kt	35°/11 kt	15 Jul 1997
Erika	15.6°N 55.3°W	28	290°/17 kt	285°/17 kt	5 Sep 1997
Erika[a]	17.5°N 59.2°W	31	295°/10 kt	305°/10 kt	6 Sep 1997
Linda	21.0°N 116.5°W	64	295°/12 kt	290°/11 kt	14 Sep 1997
Linda	22.2°N 124.7°W	39	285°/11 kt	285°/10 kt	15 Sep 1997
Alex	15.9°N 51.7°W	18	275°/12 kt	275°/11 kt	1 Aug 1998
Bonnie[a]	18.7°N 61.3°W	21	285°/23 kt	290°/19 kt	21 Aug 1998
Bonnie[a]	21.1°N 67.3°W	33	295°/16 kt	300°/16 kt	22 Aug 1998
Bonnie[b]	24.8°N 71.8°W	51	335°/4 kt	340°/4 kt	24 Aug 1998
Bonnie[c]	26.9°N 73.2°W	51	330°/5 kt	330°/7 kt	25 Aug 1998
Danielle	23.9°N 66.9°W	39	290°/13 kt	290°/12 kt	29 Aug 1998
Danielle[c]	25.9°N 71.4°W	33	300°/12 kt	300°/10 kt	30 Aug 1998
Danielle[c]	27.9°N 74.1°W	36	310°/8 kt	320°/6 kt	31 Aug 1998
Earl	26.8°N 91.5°W	26	30°/9 kt	40°/13 kt	2 Sep 1998
Hermine[e]	26.9°N 90.3°W	15		315°/10 kt	17 Sep 1998
Georges[a]	13.9°N 49.0°W	46	285°/17 kt	280°/15 kt	19 Sep 1998
Georges[b]	15.7°N 54.9°W	67	285°/16 kt	285°/13 kt	20 Sep 1998
Georges[a]	18.2°N 66.3°W	46	285°/13 kt	290°/12 kt	22 Sep 1998
Georges	18.8°N 70.8°W	36	280°/14 kt	280°/12 kt	23 Sep 1998
Georges	20.5°N 74.9°W	33	295°/10 kt	320°/9 kt	24 Sep 1998
Georges	24.8°N 83.3°W	46	275°/7 kt	285°/8 kt	26 Sep 1998
Georges	27.0°N 86.5°W	49	310°/9 kt	325°/9 kt	27 Sep 1998
Mitch	15.5°N 78.4°W	51	345°/4 kt	330°/5 kt	25 Oct 1998
Mitch	16.4°N 81.0°W	67	270°/9 kt	270°/6 kt	26 Oct 1998

[a] The G-IV and one P-3 conducted the mission.
[b] The G-IV and two P-3s conducted the mission.
[c] The G-IV and one or two P-3s flew missions, though the P-3s were not part of synoptic surveillance mission.
[d] Best-track storm-motion vector used instead of operational vector.
[e] No operational vortex specification available; best-track data used.

assimilation of the dropwindsonde observations. During 1998, 19 additional missions were conducted, making the current sample the largest reported in the literature. The following is an assessment of the impact of the dropwindsonde observations on the 5-day numerical guidance during 1997 and 1998, the first two years of G-IV operations.

2. Overview and procedures

Twenty-four synoptic surveillance missions were conducted during the 1997 and 1998 hurricane seasons in the Atlantic and eastern Pacific basins (Table 1). In this sample, Bonnie, Earl, Hermine, Georges, and Mitch made landfall within 5 days of at least one of the mission nominal times, the synoptic time by which all the data from the mission are available for assimilation (Table 2). In addition, hurricane watches were posted briefly for the northernmost Leeward Islands for Hurricane Erika, and tropical storm warnings were issued for Bermuda during the passage of Hurricane Danielle.

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 along the flightpath. In those cases in which one or two P-3 aircraft supplemented the G-IV data (Table 1), 20–25 dropwindsondes were released at the same horizontal resolution from around 400 hPa. The G-IV did not penetrate the inner core of any of the tropical cyclones, though when the P-3s flew, at least one generally gathered data near the center. 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.

To assess the impact of the dropwindsonde observations collected during the first two years of surveillance missions, the version of the NCEP Global Data Assimilation System (GDAS) operational at the end of the 1998 season was used. The GDAS is composed of a quality-control algorithm, a synthetic data procedure for tropical cyclones, an analysis procedure, and the so-called Aviation Model (AVN). The quality-control algorithm involves optimal interpolation and hierarchical decision making to evaluate the observations before input to the analysis (Woollen 1991). The synthetic data procedure (Lord 1991) creates 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 and nearby observations. The analysis scheme is the spectral statistical interpolation method of Parrish and Derber (1992): the background field (the previous 6-h forecast) is combined with ob-

Table 2. Tropical cyclone landfalls within 120 h of the nominal time of all cases. Landfall times and locations are from NHC preliminary reports. All were during 1998.

Storm name	Nominal time	Landfall location	Landfall time
Bonnie	24 Aug	Near Wilmington, NC	0400 UTC 27 Aug
Bonnie	25 Aug	Near Wilmington, NC	0400 UTC 27 Aug
Earl	2 Sep	Near Panama City, FL	0600 UTC 3 Sep
Hermine	17 Sep	Near Cocodrie, LA	0500 UTC 20 Sep
Georges	19 Sep	3 mi SE of Falmouth, Antigua	0430 UTC 21 Sep
Georges	19 Sep	8 mi SE of Basseterre, St. Kitts	0800 UTC 21 Sep
Georges	19 Sep	20 mi SW of Fajardo, PR	2200 UTC 21 Sep
Georges	19 Sep	84 mi E of Santo Domingo, DR	1230 UTC 22 Sep
Georges	19 Sep	30 mi E of Guantanamo, Cuba	2130 UTC 23 Sep
Georges	20 Sep	3 mi SE of Falmouth, Antigua	0430 UTC 21 Sep
Georges	20 Sep	8 mi SE of Basseterre, St. Kitts	0800 UTC 21 Sep
Georges	20 Sep	20 mi SW of Fajardo, PR	2200 UTC 21 Sep
Georges	20 Sep	84 mi E of Santo Domingo, DR	1230 UTC 22 Sep
Georges	20 Sep	30 mi E of Guantanamo, Cuba	2130 UTC 23 Sep
Georges	22 Sep	84 mi E of Santo Domingo, DR	1230 UTC 22 Sep
Georges	22 Sep	30 mi E of Guantanamo, Cuba	2130 UTC 23 Sep
Georges	22 Sep	Key West, FL	1530 UTC 23 Sep
Georges	23 Sep	30 mi E of Guantanamo, Cuba	2130 UTC 23 Sep
Georges	23 Sep	Key West, FL	1530 UTC 25 Sep
Georges	24 Sep	Key West, FL	1530 UTC 25 Sep
Georges	24 Sep	Biloxi, MS	1130 UTC 28 Sep
Georges	26 Sep	Biloxi, MS	1130 UTC 28 Sep
Georges	27 Sep	Biloxi, MS	1130 UTC 28 Sep
Mitch	25 Oct	72 n mi E of La Ceiba, Honduras	1200 UTC 29 Oct
Mitch	26 Oct	72 n mi E of La Ceiba, Honduras	1200 UTC 29 Oct

servations in a three-dimensional variational multivariate formalism. The forecast model's horizontal resolution is spectral triangular 126 (T126) with a Gaussian grid of 384 $\times$ 190, or an approximately 1° latitude–longitude grid, and the vertical coordinate extends from the surface to about 2.7 hPa with 28 unequally spaced sigma levels on a Lorenz grid (Caplan et al. 1997; Surgi et al. 1998).

Burpee et al. (1996) evaluated the impact of ODWs from the HRD synoptic flow experiments with three dynamical models that use the GDAS output as initial conditions: the GFDL (Kurihara et al. 1998), VBAR (Aberson and DeMaria 1994), and AVN (Caplan et al. 1997) models. The versions of these same three models current at the end of 1998 are used in the present study. The GDAS analysis is the direct input to AVN, whereas VBAR and GFDL modify the analysis near the storm center with their own vortex specification schemes; the environmental initial conditions are not modified by these two models except in the regions close to the vortex where a blending between the vortex and the environment is necessary. Neither the GFDL nor VBAR model directly ingests the dropwindsonde observations. Both models use AVN forecast fields as boundary conditions for the forecast duration. VBAR forecasts are available in the Atlantic basin only. Only the GFDL model provides both track and intensity forecasts. All observations (other than those from dropwindsondes) from the NCEP ''final'' archive were ingested into the GDAS. Two runs of each model were made for each mission: one with none (NO) of the dropwindsonde observations assimilated, the other with the inclusion of all (AL) of these observations.

This impact study represents part of an effort to investigate optimal design of the G-IV missions with targeted observational strategies (Aberson 2000; Aberson 2002, manuscript submitted to *Mon. Wea. Rev.*, hereinafter AMWR). To this end, a number of minor modifications were made to the operational procedures. First, postprocessed dropwindsonde observations (Hock and Franklin 1999) were used, including observations that might not have been received operationally (usually one or two extra soundings in each case). This postprocessing includes automatic and manual flagging of erroneous data, interpolation where data are missing, and filtering. This method is identical to that used operationally on the aircraft, though more time is available to examine and correct errors that may have been undetected or, more often, considered to be insignificant operationally. Also, during postprocessing, other data sources (rawinsondes, ships, dropwindsonde data from other aircraft, etc.) permit a more thorough assessment of data quality than is available operationally. Second, all dropwindsonde observations were assimilated into the GDAS at the nominal mission synoptic time, for consistency with previous studies and to reduce the computational requirements. All dropwindsonde observations were excluded from the data assimilation for at least 24 h before each nominal time to avoid the effects of serial correlation of tropical cyclone track forecasts (e.g., Aberson and DeMaria 1994); assessment of the second and subsequent missions in Erika, Linda, Bon-

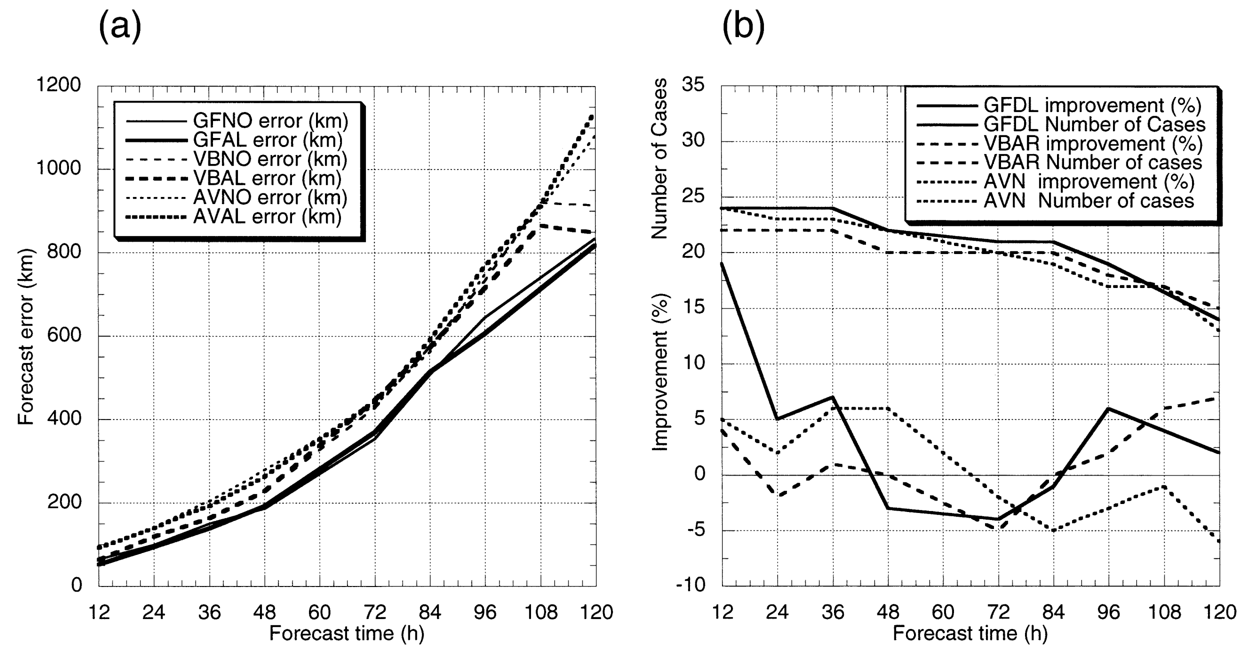

FIG. 1. (a) Absolute track forecast errors of GFDL, VBAR, and AVN forecasts with all (AL) the dropwindsonde data and none (NO) of the dropwindsonde data, and (b) the improvements of the AL forecasts over the NO forecasts. The number of forecasts is also shown in (b).

nie, Danielle, Georges, and Mitch were therefore conducted as though the previous missions had not occurred. Last, neither reconnaissance nor visible satellite imagery were available at the mission nominal time for the Claudette case. Therefore, the operational (that provided operationally for inclusion in the vortex initialization) and best-track (that calculated after an extensive poststorm analysis) storm-motion vectors differed greatly, and the best-track storm-motion vector was used (Aberson and Franklin 1999).

3. Results

a. Forecast track errors

Absolute forecast track errors are defined as the great-circle distance between the forecast and the concurrent postprocessed best-track position provided by NHC. Figure 1a shows the absolute forecast track errors for all available cases with the three models described above. The model errors are representative of average forecast track errors of the models during the 1997 and 1998 hurricane seasons. The comparisons for each model are homogeneous, but the intermodel comparisons are not. The number of forecasts for each model at each forecast time is shown in Fig. 1b. These numbers vary between models for three reasons: VBAR forecasts were not run for the two Linda cases in the eastern Pacific, the Earl and Bonnie days 3 and 4 VBAR forecasts reached the model boundaries before storm dissipation, and the GFDL (all Danielle) and AVN (Claudette, Bonnie day 3, Danielle day 3, and Georges day 6) models predicted dissipation prematurely.

The AL forecast improvements relative to the NO forecasts are shown in Fig. 1b. Differences between the two sets of forecasts are generally less than 5% except in GFDL during the critical first 36 h (when hurricane warnings are posted). The only difference that is statistically significant at the 95% level with a paired *t* test (e.g., Larsen and Marx 1981), with the null hypotheses that the mean of the differences is not significantly different from zero, is the 12-h GFDL forecast. A majority of forecasts were improved with the assimilation of the dropwindsonde data at all times except at 120 h in GFDL, 72 h in VBAR, and 84 and 120 h in AVN. However, only at 12, 72, and 84 h in GFDL and at 120 h in VBAR were at least two-thirds of the forecasts improved by the assimilation of the dropwindsonde data (Table 3).

TABLE 3. Percentage of cases at each forecast time improved by the assimilation of the dropwindsonde data for each model.

Model name	Forecast range 12 h	24 h	36 h	48 h	72 h	84 h	96 h	108 h	120 h
GFDL	71	58	63	50	62	78	73		43
VBAR	59	64	64	60	45	55	56	65	73
AVN	63	61	57	50	50	45	56	53	40

b. Landfall forecasts

The most important statistic is whether the dropwindsonde data are able to improve the landfall position and time forecasts. Landfall errors are investigated in the manner of Powell and Aberson (2001). Forecast

Table 4. Landfall forecast errors (distance and timing) for the three models for all cases shown in Table 2.

Model name	NO distance error (km)	AL distance error (km)	Improvement (%)	NO timing error (h)	AL timing error (h)	Improvement (%)
GFDL	207	201	3	11.9	14.2	−19
VBAR	196	189	4	11.9	15.9	−34
AVN	215	200	7	13.0	14.7	−13

positions are interpolated with splines at half-hour intervals, and the locations and times at which the tracks cross the coastline are noted. All the landfall positions and times are shown in Table 2, and the forecast differences are in Table 4. Because of the small number of landfalls, these are not stratified by forecast time. The average distance errors were improved modestly in all three models by the additional data; however, the timing forecasts were degraded by the dropwindsonde data. The surveillance missions improved 61% of the GFDL landfall location forecasts and 52% of the GFDL landfall timing forecasts. The VBAR landfall location forecasts were improved 65% of the time, and the timing forecasts 52% of the time, by the surveillance missions. The AVN forecasts were improved only 48% of the time (location) and 65% of the time (timing) by the surveillance missions. Though in the majority of cases the landfall forecasts were improved by the additional data, the improvements are not statistically significant, so conclusive statements about landfall forecast improvements cannot be made based on this sample.

c. *Intensity*

Tuleya and Lord (1997) showed modest improvements to GFDL intensity forecasts in the HRD synoptic flow cases. The dropwindsonde observations improve the intensity forecasts by as much as 13% in the current sample (Fig. 2). Only the 96-h forecast improvement is statistically significant at the 95% level. At most forecast times, the forecasts that were improved or degraded by the assimilation of the dropwindsonde data were almost evenly distributed.

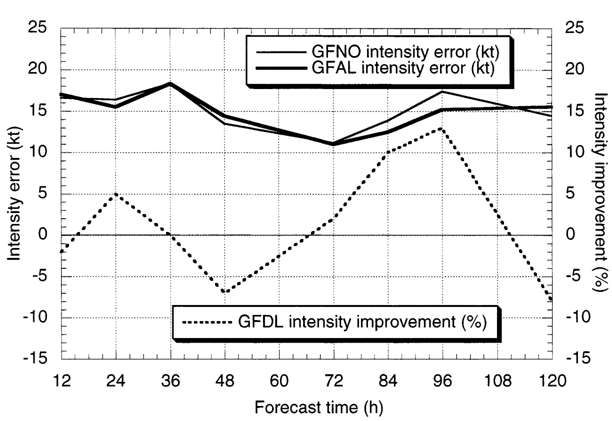

Fig. 2. Intensity error of GFDL forecasts with all (AL) the dropwindsonde data and none (NO) of the dropwindsonde data, and the improvements of the AL forecasts over the NO forecasts.

4. Discussion

The track forecast improvements in the current sample are small in comparison with those from the HRD synoptic flow experiments (Burpee et al. 1996; Tuleya and Lord 1997). Three possible reasons are examined.

a. *Sampling strategy*

Burpee et al. (1996) showed that the three models used in the present study responded favorably to the relatively uniform data distribution in the HRD synoptic flow experiments. Logistical constraints associated with the use of a single aircraft (as opposed to the usual two P-3s in the original experiments) often made obtaining data symmetrically around the storm difficult. The asymmetric sampling in the Erika, Linda, Alex cases and the first, fourth, fifth, and sixth Georges cases (more than one-third of the cases) may not be optimal enough to improve the numerical guidance (Aberson and Franklin 1999).

Figure 3 shows examples of the impact of the dropwindsonde data on the initial 850–200-hPa deep-layer mean steering flow in one case with asymmetric, and another with symmetric, sampling around the storm. The symmetric sampling allows for differences in the steering flow completely around the tropical cyclone. However, because of variations in the accuracy of the first-guess field, interactions between synthetic vortex data and nearby dropwindsonde data, and the spreading of the information from the dropwindsonde data by the data assimilation scheme, differences do not surround the tropical cyclone in every case with symmetric sampling. Because of the spread of the influence of the data into data-sparse regions as shown in Fig. 3a, sampling the environment only partially around the vortex may introduce an asymmetry in the flow, forcing the storm to move with an incorrect velocity (Derber and Bouttier 1999).

Figure 4 shows that in just over one-half of the cases the improvements in both baroclinic models are usually larger for the cases in which the dropwindsonde data are distributed entirely around the vortex than those with an asymmetric distribution. However, at only one forecast time in VBAR did dropwindsonde data completely around the vortex provide a larger forecast improvement

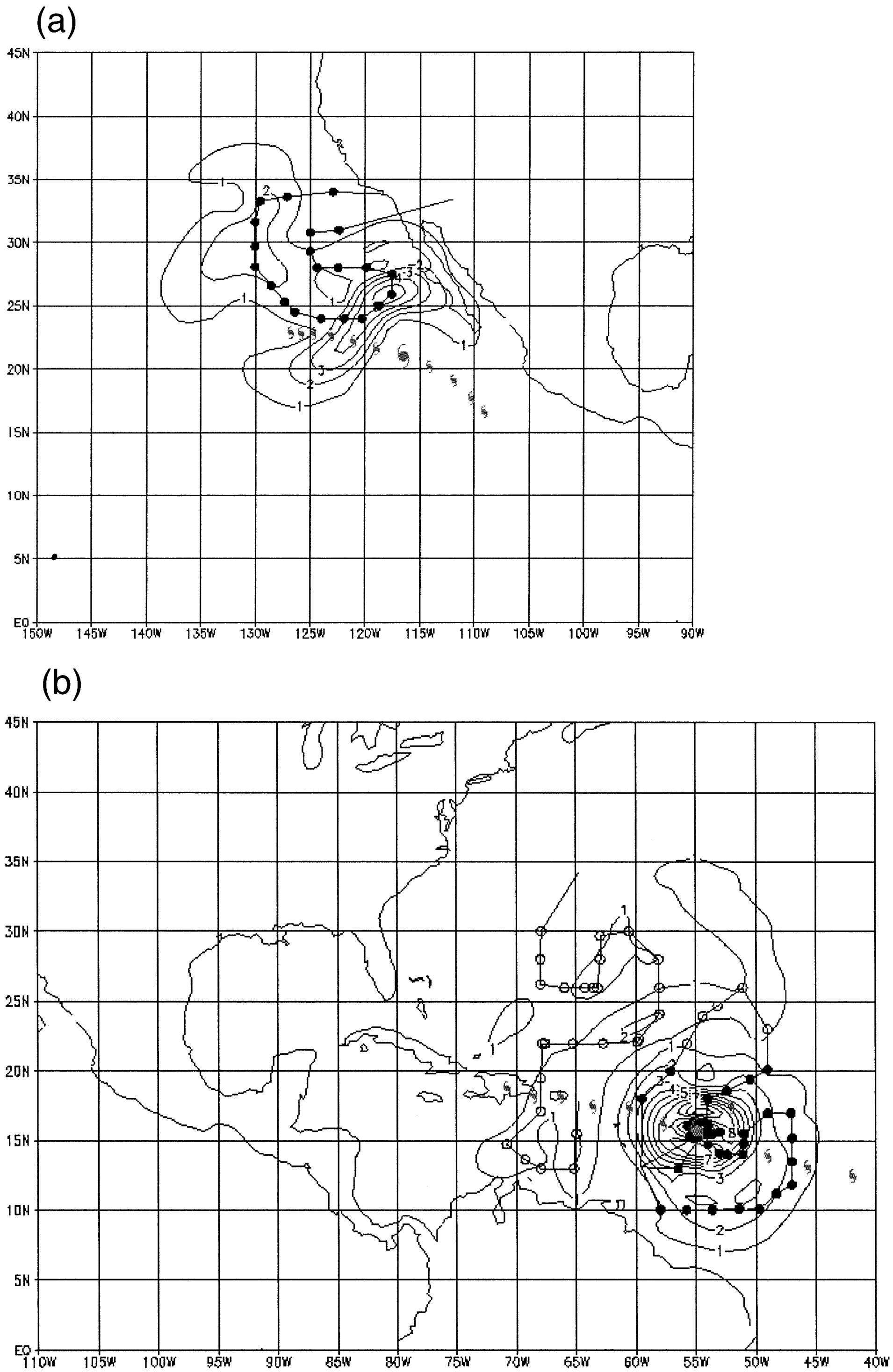

FIG. 3. Initial-condition differences in the AVN model in the vertically averaged 850–200-hPa winds (kt) between the AVN with no (AVNO) and all (AVAL) dropwindsonde data for (a) Hurricane Linda at 0000 UTC 14 Sep 1997 and (b) Hurricane Georges at 0000 UTC 20 Sep 1998. The large hurricane symbol is the location of the hurricane at the nominal time. The small hurricane symbols are the locations of the hurricane every 12 h previous and after the nominal time. The black dots represent the locations of dropwindsonde observations.

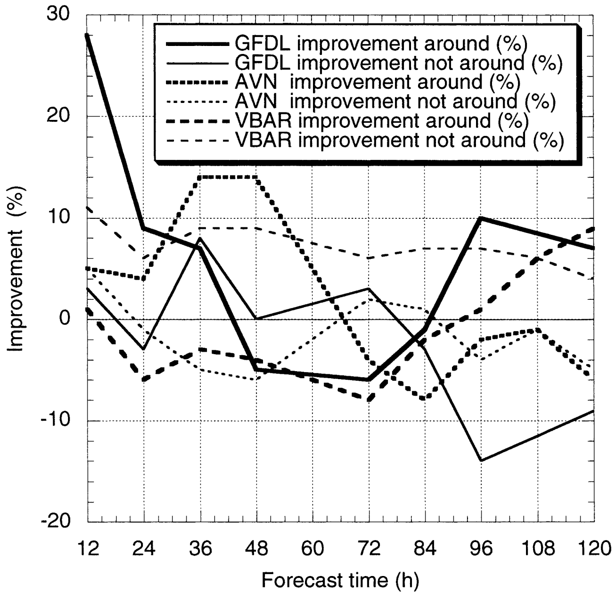

FIG. 4. Improvements of the AL forecasts over the NO forecasts in GFDL, VBAR, and AVN in those cases in which the data were obtained in all directions away from the storm (around) compared with when only some directions away from the storm had data (not around).

than an asymmetric distribution. In GFDL, the 12 h forecasts are statistically significantly improved by the dropwindsonde data in the symmetrically sampled cases, and the 120-h forecasts in the asymmetrically sampled cases are significantly degraded, both at the 99% level. In VBAR, the improvements due to the dropwindsonde data in the cases with asymmetric sampling are significant at the 99% level at 108 h and at the 95% level at all other times except 12, 24, and 72 h; no differences in the sample with symmetric sampling are significant. In AVN, no differences in either sample are statistically significant. Therefore, uniform sampling in all quadrants of the tropical cyclone is not proven to be beneficial to the three numerical models discussed.

b. Data coverage

The G-IV aircraft has a longer range than the slower P-3, and therefore the data distribution from a one-plane G-IV surveillance mission will have greater areal coverage than a one-plane P-3 mission, though less than a two-plane P-3 mission. More than one-half of the missions in Burpee et al. (1996) utilized two P-3 aircraft, whereas only one-quarter of the 1997–98 missions involved more than one aircraft. Since the data coverage in the original sample was usually greater than in the current sample, the amount of areal coverage is investigated as a possible reason for smaller forecast improvement in the current sample than in the synoptic-flow experiment sample.

The sample is separated into subsets based on how many aircraft participated in each mission. Only those cases in which one or more P-3s was directly involved in the surveillance mission were included in the sample with more than one plane, since other research or reconnaissance missions did not augment the synoptic data coverage. Though the first Georges case involved two aircraft, it was removed from the multiplane set because both aircraft completed only abbreviated missions, and the data coverage therefore was approximately that of a one-plane mission.

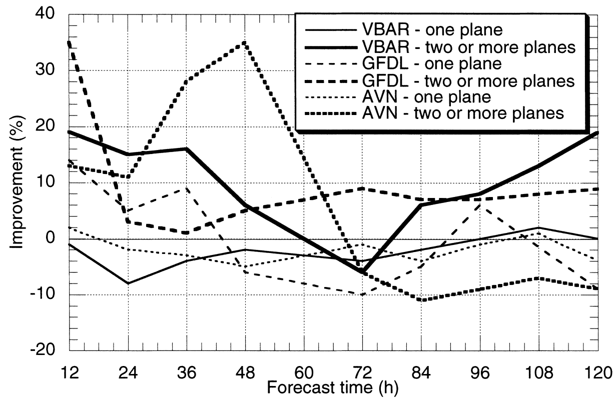

FIG. 5. Improvements of the AL forecasts over the NO forecasts in GFDL, VBAR, and AVN in those cases in which only the G-IV flew compared with those cases in which the G-IV mission was supplemented by at least one P-3 aircraft.

The improvements to the model forecasts due to the assimilation of the dropwindsonde data in the subsets with one and with more than one plane are shown in Fig. 5. The second or third plane improved the GFDL model results between 48 and 84 h but did not substantially change the amount of improvement at other times. In fact, 24- and 36-h forecast improvements with two or more planes were a bit smaller than those with one plane, and significance tests show that the improvements in both samples at 12 h are statistically significant at the 95% level. The second or third plane allowed for substantially improved AVN and VBAR forecasts, and these improvements approached and sometimes surpassed those shown in Burpee et al. (1996). The 48-, 84-, and 96-h improvements in the sample with two or more aircraft were significant at 90%, though no forecast times are significant at 95%. The 12- and 24-h VBAR forecast improvements in the sample with more than one aircraft were statistically significant at the 95% level, whereas no improvements were statistically significant in the sample with one aircraft. Increased data coverage during synoptic surveillance missions seems important for forecast track improvements.

c. Synthetic data

Another possible reason for the small improvements is negative interactions between the mix of data in the hurricane core and near-storm environment. In all numerical models geared toward forecasting tropical cyclone tracks, a vortex scheme is needed to initialize tropical cyclones. In the AVN vortex scheme (Lord 1991), synthetic observations representative of the tropical cyclone based on operationally estimated intensity, position, and motion are created. The operational values differ from the postprocessed best-track values because of the availability of data after operational deadlines and the ability to examine the data without time constraints. Differences in the initial position are almost always less than 30 km, and these do not amplify rapidly in time. Differences in the initial intensity rarely cause track differences larger than a few kilometers even at long forecast lead times. However, the synthetic data include an asymmetry to force the vortex to move in the specified direction, and even small differences in the current storm-motion vector can amplify rapidly with time. For example, even without amplification, a 1 m s^{-1} difference in the initial motion results in a 24-h difference in position of 86.4 km, about two-thirds of the average 24-h forecast error in this sample. Further, if the asymmetry does not agree with the environmental flow near the storm center, a period of adjustment may be required, after which the forecast may be degraded enough that the good dropwindsonde data cannot improve it.

Each forecast is classified as to whether the operational storm-motion vector is either ‘‘good’’ or ‘‘poor,’’ with both subsets of cases having nearly one-half of the total number of cases. The current storm-motion vector is given operationally in direction (°) and speed (kt). The best-track storm-motion vector is calculated from the best-track positions at, and 6 h before, the nominal time; a storm-motion vector based on the most recent 12-h motion does not change the results. If either the directions or the speeds disagree by at least 15° or 3 kt, respectively, then the storm-motion vector is considered to be poor. The operational and best-track storm-motion vectors for all cases are presented in Table 1.

All three models show improvements at all forecast times (except AVN forecasts at 24 h) in cases with the good storm-motion vectors; however, the cases in which the storm-motion vectors are poor are degraded at all times except for 12 and 96 h (GFDL), 24–48 h (AVN), and 120 h (VBAR) (Fig. 6). In GFDL, the improvements due to the dropwindsonde data are statistically significant at the 99% level at 12 and 36 h in the sample with good storm-motion vectors and at 12 h in the sample with poor storm-motion vectors; the improvements are significant at the 95% level at 12 h in the sample with good storm-motion vectors. In VBAR, the improvements in the sample are statistically significant at the 95% level at all times except 24, 36, and 120 h in the sample with the good storm-motion vectors; the degradations in the sample with the poor storm-motion vectors are statistically significant at the 95% level at 72–96 h. In AVN, the only difference that is statistically significant at the 95% level is the 120-h degradation in the sample with the poor storm-motion vectors.

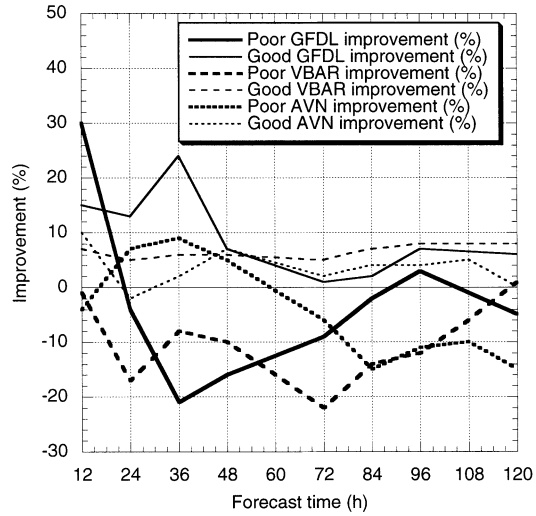

FIG. 6. As in Fig. 5, but for those cases in which the storm-motion vector was good compared with when it was poor.

These statistical tests suggest that the accuracy of the operational storm-motion vector may be the most important factor in predicting the amount of forecast model improvement from the synoptic surveillance missions. Of the six cases in which more than one plane participated, two were in the poor storm-motion vector sample. Improvements in these two cases were not present at all forecast times in any model, unlike the multiplane cases with good storm-motion vectors. This suggests that correction of the storm-motion vector is of paramount importance to model improvement, but also that the amount of improvement may be increased further with additional areal data coverage. Operational constraints prevent very accurate quantification of the storm-motion vector. A new vortex technique that does not utilize the storm-motion vector was implemented by NCEP before the 2000 Atlantic hurricane season, leading to a large improvement in AVN track forecast accuracy during the subsequent seasons (Liu et al. 2000). The introduction of this vortex may allow for improvements in subsequent synoptic surveillance cases on the order of those shown in the sample with the good storm-motion vectors.

5. Conclusions

Only modest changes to model forecasts were achieved in the large forecast sample obtained during the first two seasons of synoptic surveillance operations. The missions resulted in statistically significant improvements in the GFDL track forecasts only at 12 h. No statistically significant improvements are seen in the VBAR and AVN models. Examination of individual cases suggests that the difficulty in the quantification of the current storm-motion vector in real time for use in the synthetic vortex procedure was the primary cause of the small forecast improvements. In those cases in which the storm-motion vector was accurately assessed, the surveillance missions showed statistically significant improvement of up to 25% in the GFDL model. In the complementary subset, the forecasts in all three models are mainly degraded. The ingesting of accurate (dropwindsonde) data into a numerical model alongside marginally accurate (synthetic) data may actually degrade forecasts. Another important factor in the amount of improvement seen 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.

This is the first study showing that adding good (in this case GPS dropwindsonde) data into the mix of observations will not automatically result in improved forecasts, and this may be the case in the extratropics as well. One might suggest that other types of data, such as satellite data, may be more suited to forecast improvements because of their great coverage in space and time. However, tropical cyclones are steered by winds in a layer; since synoptic surveillance missions are generally ordered for mature hurricanes, this layer is likely to extend from 850 to 200 hPa. Dropwindsondes measure momentum, mass, temperature, and moisture over this layer, whereas satellites generally do not provide such soundings. Furthermore, satellite wind data are available only where tracers are available, and satellite data are not as accurate as dropwindsonde data. Though satellite data certainly hold promise for improvements and are an important part of the current mix of observations, an optimal mix between the two data types is preferable to only having one.

In order to improve forecasts with synoptic surveillance missions (and satellite data), not only the data, but also the assimilation of these data, must improve. (Though this study mainly addresses the assimilation of the dropwindsonde data, surely model improvements will contribute to improved forecast accuracy.) The current results should not deter other agencies or nations from planning their own surveillance program. However, they show the need to invest not only in the infrastructure to obtain the data, but also in the techniques to properly assimilate the data into numerical models.

One obstacle to maximal forecast improvement identified in this study has been remedied at NCEP (independent of these results) with the replacement of the synthetic vortex method by the vortex replacement technique before the 2000 Atlantic hurricane season. This change has led to huge improvements in track forecast accuracy during ensuing seasons. Forecast centers and numerical modelers may wish to closely examine the way in which tropical cyclone vortices are initialized in their systems in light of the current results and the recent large increase in forecast skill at NCEP. Preliminary results from subsequent hurricane seasons with the updated vortex procedure and knowledge of where (targeting) and how (sampling strategies) observations should be taken (AMWR) have shown a return to the impressive improvements seen in the Burpee et al. (1996) study.

Acknowledgments. The author thanks Morris Bender and Bob Tuleya of GFDL and Tim Marchok of NCEP for instructions in running the GFDL hurricane model, and Jeff Ator, John Derber, Mark Iredell, Bert Katz, Dennis Keyser, Hua-Lu Pan, Russ Treadon, and Jack Woollen of NCEP for their instruction in running the NCEP GDAS and model. Further thanks go to Tim Marchok of NCEP for implementing the tropical cyclone tracking algorithm for AVN, and Maxine Brown, Janine Clements, Mickey Farley, Harold Lloyd, Mary McCann, Eric Monrad, Carolyn Pasti, George Vandenberghe, and Arthur Wick for helping in data acquisition and computer support on the NCEP computers. The support of Eugenia Kalnay of the University of Maryland and Stephen Lord of NCEP was invaluable. Neal Dorst and Steven Feuer of HRD did much of the dropwindsonde postprocessing, and Joyce Berkeley helped to draft some of the figures. Bob Kohler and Bill Barry provided computer support and guidance at HRD. The author also thanks the NOAA Aircraft Operations Center (AOC) flight crews and dropwindsonde operators, and Sean White, AOC G-IV project manager; HRD personnel who participated in the flights; and Naomi Surgi of NCEP. A special debt of gratitude is owed to Robert Burpee, who originated the concept of the synoptic flow experiments, supported them as director of the Hurricane Research Division, and under whose leadership at the National Hurricane Center the operational dropwindsonde missions became a reality. The operational surveillance missions also came to fruition under the research leadership of James Franklin of NHC who drew the flight tracks and flew in nearly all missions discussed in the manuscript. This work was completed as part of the requirements for a Ph.D. at the University of Maryland under the helpful guidance of Professor Eugenia Kalnay.

REFERENCES

Aberson, S. D., 2000: Three years of tropical cyclone synoptic surveillance in the Atlantic basin. Preprints, *24th Conf. on Hurricanes and Tropical Meteorology,* Fort Lauderdale, FL, Amer. Meteor. Soc., 108–109.

——, and M. DeMaria, 1994: Verification of a nested barotropic hurricane track forecast model (VICBAR). *Mon. Wea. Rev.,* **122,** 2804–2815.

——, and J. L. Franklin, 1999: Impact on hurricane track and intensity forecasts of GPS dropwindsonde observations from the first-season flights of the NOAA Gulfstream-IV jet aircraft. *Bull. Amer. Meteor. Soc.,* **80,** 421–427.

Burpee, R. W., J. L. Franklin, S. J. Lord, R. E. Tuleya, and S. D. Aberson, 1996: The impact of Omega dropwindsondes on operational hurricane track forecast models. *Bull. Amer. Meteor. Soc.,* **77,** 925–933.

Caplan, P., J. Derber, W. Gemmill, S.-Y. Hong, H.-L. Pan, and D. Parrish, 1997: Changes to the 1995 NCEP operational Medium-Range Forecast model analysis–forecast system. *Wea. Forecasting,* **12,** 581–594.

Derber, J., and F. Bouttier, 1999: A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus,* **51A,** 195–221.
Hock, T. F., and J. L. Franklin, 1999: The NCAR GPS dropwindsonde. *Bull. Amer. Meteor. Soc.,* **80,** 407–420.
Kurihara, Y., R. E. Tuleya, and M. A. Bender, 1998: The GFDL hurricane prediction system and its performance in the 1995 hurricane season. *Mon. Wea. Rev.,* **126,** 1306–1322.
Larsen, R. J., and M. L. Marx, 1981: *An Introduction to Mathematical Statistics and Its Applications.* Prentice Hall, 530 pp.
Liu, Q., T. Marchok, H.-L. Pan, M. Bender, and S. J. Lord, 2000: Improvements in hurricane initialization and forecasting at NCEP with global and regional (GFDL) models. NOAA Tech. Procedures Bull. 472, 7 pp. [Available online at http://www.nws.noaa.gov/om/tpb/472.htm.]
Lord, S. J., 1991: A bogusing system for vortex circulations in the National Meteorological Center global forecast model. Preprints, *19th Conf. on Hurricanes and Tropical Meteorology,* Miami, FL, Amer. Meteor. Soc., 328–330.
Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.,* **120,** 1747–1763.
Powell, M. D., and S. D. Aberson, 2001: Accuracy of United States tropical cyclone landfall forecasts in the Atlantic basin 1976–2000. *Bull. Amer. Meteor. Soc.,* **82,** 2749–2767.
Riehl, H., W. H. Haggard, and R. W. Sanborn, 1956: On the prediction of 24-hour hurricane motion. *J. Meteor.,* **13,** 415–420.
Surgi, N., H.-L. Pan, and S. J. Lord, 1998: Improvement of the NCEP global model over the Tropics: An evaluation of model performance during the 1995 hurricane season. *Mon. Wea. Rev.,* **126,** 1287–1305.
Tuleya, R. E., and S. J. Lord, 1997: The impact of dropwindsonde data on GFDL hurricane model forecasts using global analyses. *Wea. Forecasting,* **12,** 307–323.
Woollen, J. R., 1991: New NMC operational OI quality control. Preprints, *Ninth Conf. on Numerical Weather Prediction,* Denver, CO, Amer. Meteor. Soc., 24–27.