The Impact of Dropwindsonde Observations on Typhoon Track Forecasts in DOTSTAR and T-PARC

KUN-HSUAN CHOU
Department of Atmospheric Sciences, Chinese Culture University, Taipei, Taiwan

CHUN-CHIEH WU AND PO-HSIUNG LIN
Department of Atmospheric Sciences, National Taiwan University, Taipei, Taiwan

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

MARTIN WEISSMANN AND FLORIAN HARNISCH
Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

TETSUO NAKAZAWA
Meteorological Research Institute, JMA, Tsukuba, Japan

(Manuscript received 3 August 2010, in final form 1 November 2010)

ABSTRACT

The typhoon surveillance program Dropwindsonde Observations for Typhoon Surveillance near the Taiwan Region (DOTSTAR) has been conducted since 2003 to obtain dropwindsonde observations around tropical cyclones near Taiwan. In addition, an international field project The Observing System Research and Predictability Experiment (THORPEX) Pacific Asian Regional Campaign (T-PARC) in which dropwindsonde observations were obtained by both surveillance and reconnaissance flights was conducted in summer 2008 in the same region. In this study, the impact of the dropwindsonde data on track forecasts is investigated for DOTSTAR (2003–09) and T-PARC (2008) experiments. Two operational global models from NCEP and ECMWF are used to evaluate the impact of dropwindsonde data. In addition, the impact on the two-model mean is assessed.

The impact of dropwindsonde data on track forecasts is different in the NCEP and ECMWF model systems. Using the NCEP system, the assimilation of dropwindsonde data leads to improvements in 1- to 5-day track forecasts in about 60% of the cases. The differences between track forecasts with and without the dropwindsonde data are generally larger for cases in which the data improved the forecasts than in cases in which the forecasts were degraded. Overall, the mean 1- to 5-day track forecast error is reduced by about 10%–20% for both DOTSTAR and T-PARC cases in the NCEP system. In the ECMWF system, the impact is not as beneficial as in the NCEP system, likely because of more extensive use of satellite data and more complex data assimilation used in the former, leading to better performance even without dropwindsonde data. The stronger impacts of the dropwindsonde data are revealed for the 3- to 5-day forecast in the two-model mean of the NCEP and ECMWF systems than for each individual model.

1. Introduction

Starting in 2003, the research program “Dropwindsonde Observations for Typhoon Surveillance near the Taiwan Region” (DOTSTAR) marked the beginning of an era of tropical cyclone (TC) surveillance and targeted observations in the western North Pacific using GPS dropwindsondes (Wu et al. 2005). This program is built upon work pioneered by the National Oceanic and Atmospheric Administration Hurricane Research Division to improve TC track forecasts in the Atlantic (Burpee et al. 1996; Aberson and Franklin 1999; Aberson 2003).
From 2003 to 2009, 45 surveillance flights were conducted around 35 typhoons, with 751 dropwindsondes released. Based on the results of 10 cases conducted in 2004, Wu et al. (2007b) showed that dropwindsonde data from DOTSTAR improved the 72-h track forecast of the ensemble mean\(^1\) of three global models by an average of 22%. Improved methods to combine the dropwindsonde data with bogus vortices showed a clear positive impact on both the TC track and intensity forecasts in a mesoscale model (Chou and Wu 2008). In addition to the impact of DOTSTAR dropwindsonde data on TC track forecasts, detailed aspects such as targeted observations on TCs and validation of remote sensing data have also been studied (Wu et al. 2007a, 2009a,b,c; Yamaguchi et al. 2009; Chou et al. 2010).

In summer 2008, the international THORPEX Pacific Asian Regional Campaign (T-PARC) was conducted in the western North Pacific. The aim of the multinational field campaign was to address short-range TC dynamics and forecast skill in one region and the downstream impacts of TCs on medium-range dynamics and forecast skill in another region (Elsberry and Harr 2008; Parsons et al. 2008). This was the first time that four aircraft [the DOTSTAR Astra jet, the German Aerospace Center (DLR) Falcon 20, a U.S. Navy P-3, and a U.S. Air Force C-130] were used simultaneously to observe typhoons. DOTSTAR Astra and DLR Falcon sampled the TC environment, especially in the high-sensitivity (target) areas, while the P-3 and C-130 conducted reconnaissance flights in the inner core and rainband areas of TCs. Onboard observation equipment and expendables, such as Global Positioning System (GPS) dropwindsondes, wind and water vapor Light Detection and Ranging (LIDARs), Doppler radar, and airborne expendable bathythermographs were deployed. The experiment provided unprecedented, valuable data for studying the physics, dynamics, and thermodynamics of the track and intensity, structure change from genesis through extratropical transition, targeting, and TC predictability. During the

---

\(^1\) The ensemble was comprised of the NCEP Global Forecast System, the Navy Operational Global Atmospheric Prediction System of the Fleet Numerical Meteorology and Oceanography Center, and the JMA Global Spectral Model.
The T-PARC field campaign, the four aircrafts flew in total more than 500 h, including the observations of Typhoons Sinlaku, Hagupit, and Jangmi, and more than 1500 additional soundings were obtained (Weissmann et al. 2011).

Although the overall added value of the dropwindsonde data in improving typhoon track forecasts over the western North Pacific has been demonstrated, the impact of dropwindsondes has not been shown to be statistically significant because of the limited number of DOTSTAR cases studied previously (Wu et al. 2007b). In this paper, a larger sample of cases is examined to obtain more reliable statistics.

The model and analysis method used in this study are presented in section 2. The overall statistics from DOTSTAR cases during 2003–09 are described in sections 3, and the multimodel results 2005–09 are discussed in section 4. The results from two T-PARC typhoon observation cases in 2008 are described in section 5, and the conclusions are given in section 6.
FIG. 2. Track improvement (in km) after the assimilation of dropwindsonde data into the NCEP GFS model for each DOTSTAR mission; (a) 24 h, (b) 48 h, (c) average during the forecast period 6–120 h.
2. The model descriptions and experimental designs

To evaluate the impact of dropwindsonde data on numerical forecasts in the western North Pacific during the DOTSTAR and T-PARC programs, the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) and European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecast System (IFS) modeling systems are used.

a. NCEP Global Forecast System

The operational version of the GFS was used on each mission. This implies that the model resolution and physics processes varied with time from 2003 to 2008 (Aberson 2010). In 2003, the GFS horizontal resolution was T254, and the vertical coordinate extended from the surface to about 2.7 hPa with 64 (L64) unequally spaced sigma levels on a Lorenz grid (Caplan et al. 1997; Surgi et al. 1998). The resolution was increased to T382L64 (38 km horizontally) in 2005.

The NCEP Global Data Assimilation System uses a quality control algorithm, a TC vortex relocation procedure, and the Global Spectral Model. The quality control involves optimal interpolation and hierarchical decision making to evaluate the observations before going into detailed analysis (Woollen 1991). A vortex relocation procedure (Liu et al. 2000) in which TCs in the first guess field are relocated to the analyzed position in each 6-h analysis cycle (as in Kurihara et al. 1995) ensures that the systems are located in the operationally fixed locations. The spectral (prior to 2007) and grid point (since 2007) statistical interpolations are used for the analysis scheme, while the background field (the previous 6-h forecast) is combined with observations using a three-dimensional variational data assimilation (3DVAR) multivariate formalism.

Two runs were conducted to assess the impact of dropwindsonde data on track forecasts. In the control run the dropwindsonde data were assimilated into the model (NCEP-O), whereas in the denial run, the dropwindsonde data were not used (NCEP-N). All other observations from the NCEP final archive were assimilated in both sets of runs. The control runs were made in real time, and the denial runs were completed retrospectively. Within the DOTSTAR program, surveillance missions are performed for one TC at a time, usually at 0000 UTC, whereas during T-PARC observations such missions are conducted at multiple times for each TC during its lifetime. The denial runs are initiated 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 is completed (Aberson and Etherton 2006). Dropwindsonde data were removed globally in the denial runs. Dropwindsonde data in the Atlantic might be expected to influence flow patterns and typhoon track predictions over the Pacific significantly (Aberson 2011). In most years, there were no dropwindsonde data simultaneously in the Atlantic and Pacific basins. Dropwindsonde observations within a radius of 111 km from the TC center are not used in the NCEP analysis (Aberson 2008). For DOTSTAR cases, only 12 (2%) dropwindsondes are rejected in NCEP analysis because they are close to the TC, whereas the number increases to 92 (20%) during T-PARC, mostly due to the dropwindsondes deployed during reconnaissance flights.

b. ECMWF Integrated Forecast System

The 2009 spring version of the ECMWF IFS is used, running with a horizontal resolution of ~25 km (T799), 91 vertical levels, and a 4DVAR data assimilation with 12 hourly windows (0900–2100 and 2100–0900 UTC). Forecasts through 240 h are initialized twice daily at 0000 and 1200 UTC. Further information about the ECMWF analysis and forecasting system is given in Rabier et al. (2000), Mahfouf and Rabier (2000), and Richardson et al. (2009).

The ECMWF assimilation system contains a first-guess check and a variational quality control. The first-guess check is relaxed to a great extent (nearly inactive) for latitudes lower than 30° to avoid high rejections in and near TCs. This modification was extended to cover latitudes up to 40° during T-PARC because Typhoon Sinlaku reintensified near 30°N. The ECMWF IFS also assimilates dropwindsondes data in the TC eye and eyewall (in contrast to NCEP), but a significant percentage of

![Fig. 3. The percentage of cases improved at each forecast time during different time periods in NCEP GFS and ECMWF IFS models. MEAN is the result of two-model mean of NCEP GFS and ECMWF IFS models.](https://example.com/fig3.png)
the observations is usually rejected by the variational quality control (Harnisch and Weissmann 2010).

Consistent with the NCEP GFS modeling system, control (with dropwindsonde data; i.e., EC-O) and denial (without Pacific dropwindsonde data only; i.e., EC-N) runs were conducted to assess the impact of the dropwindsonde data on TC track forecasts. These are all retrospective runs in both DOTSTAR and T-PARC cases.

**FIG. 4.** The JTWC best track (typhoon symbols), the NCEP-N (circles) and NCEP-O (dots) forecast tracks of the three (a)–(c) best and (d)–(f) worst DOTSTAR cases for every 6 h. Track errors are shown at the bottom of the figure. TKE means track error (km) and IMP indicates track error improvement (in km and %).
Most of the DOTSTAR cases are run uncycled, but the six cases in the T-PARC period (Sinlaku, Hagupit, and Jangmi in 2008) are performed in a cycled mode. In addition, only the DOTSTAR cases from 2005 to 2009 are evaluated with the ECMWF modeling system.

3. Results from DOTSTAR during 2003–09 in the NCEP Global Forecast System

From 2003 to 2009, 45 surveillance flight cases were conducted for 35 typhoons. All tracks of observed TCs and locations of deployed dropwindsondes are shown in Fig. 1; information on each case is listed in Table 1. With more cases examined than in Wu et al. (2007b), the statistical confidence level of improved track forecasts from DOTSTAR dropwindsonde data is stressed in this study.

Only 42 cases were examined for the control and denial runs to assess the impact of dropwindsonde data from the NCEP Global Forecast System, because the data were not transferred to the Global Telecommunications System (GTS) in real time in the other three. The forecast tracks from both control and denial runs are compared against the best tracks from Joint Typhoon Warning Center (JTWC). Figure 2 shows track forecast error reduction resulting from the use of dropwindsonde data in each case at 24 and 48 h, and the average during the forecast period (6 to 120 h). At 24 h, 57% of the forecasts were improved by the dropwindsonde data, increasing to 67% at 48 h. Figure 3 shows the percentage of cases with forecast track improvement from the dropwindsonde data at each forecast time (solid line). The percentage of improved cases is between 50% and 70%, and the overall beneficial rate is about 60%. The improvements are generally larger than the degradations.

Figure 4 shows the best three (Meari, Lupit, and Sinlaku) and the worst three (Mindulle, Morakot, and Parma) cases of the 42 GFS runs, based on the averaged track error (Fig. 2c). The dropwindsonde data significantly improved the timing of recurvature of Meari and Lupit (Figs. 4a,c) and the eastward bias of Sinlaku (Fig. 4b), but slightly degraded the track forecasts of those TCs (Mindulle, Morakot, and Parma) that were influenced by the terrain of Taiwan and Luzon (Figs. 4d–f).

The impact of the dropwindsonde data at each forecast time on NCEP GFS track forecasts is shown in Fig. 5. The overall impact of the data is an error reduction of 10% at 24 h, gradually increasing to 22% later in the forecast period. The improvements are statistically significant at the 95% confidence level (paired t-test with one-sided distribution; Larsen and Marx 1981) at 48-, 72-, and 96-h forecast lead times and at the 90% confidence level at 24- and 120-h forecast lead times. The result obtained here is similar to the finding obtained from the ten-year operational synoptic surveillance of 176 missions conducted in the Atlantic (Aberson 2010). These 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) at mission times.

4. Multimodel results from DOTSTAR during 2005–09

The ECMWF IFS system has also been used to evaluate the impact of dropwindsonde data on track forecasts. Because of computational constraints, the ECMWF control and denial runs are only made in 25 cases during 2005 to 2009, as listed in Table 1. Figure 6 shows the best and worst three forecast tracks among the 25 cases. The track forecast improvements and degradations are not significantly different, implying that the impact of dropwindsonde data to the ECMWF IFS is not as significant as to the NCEP GFS. However, the track forecast errors of the ECMWF denial runs are smaller than those from the NCEP GFS, which limits the potential for improvements. A detailed model intercomparison for the T-PARC period is discussed in Weissmann et al. (2011).

Wu et al. (2007b) showed that the three-model mean of the NCEP GFS, U.S. Navy Operational Global Atmospheric Prediction System (NOGAPS), and Japan

---

2 Results of the 72-h forecasts are not shown since in some cases no verification is available.
Meteorological Agency (JMA) Global Spectral Model (GSM) has a larger positive impact than any individual model. Figure 7 shows the best two forecast tracks of the two-model mean. The two-model mean tracks yield the same results with regard to the impact of dropwindsonde data as in the NCEP GFS forecasts; that is, large-track improvement and small-track degradation rates for the most positive and negative cases. Furthermore, the two-model mean could lead to better track forecasts than individual members, such as in the ECMWF IFS model.

Fig. 6. As in Fig. 4, but for the best and worst three track improvement cases among the 2005–09 DOTSTAR cases in the ECMWF IFS model.
case of Sinlaku based on the comparison between Figs. 4b and 7a.

Figure 8 shows the track error reduction by using the dropwindsonde data for each case in the NCEP GFS, the ECMWF IFS and the two-model mean at 48 h and for the average during the forecast period. For the 48-h forecast in ECMWF IFS (Fig. 8b), although the number of cases improved is larger than that of the degraded ones, the magnitude of improvement in positive cases is smaller than that of degradation in negative cases. For the average track error reduction of the ECMWF IFS (Fig. 8e), the number of cases with track improvement is larger than that of the degraded ones, because the impact of the dropwindsonde data in the control run is larger during longer forecast periods in the ECMWF system than in the GFS. For the two-model mean result...
both at 48 h and for the average during the forecast period (Figs. 8c,f), as compared to the result of NCEP GFS (Figs. 8a,d), the percentage of cases with track improvement is similar to that in the NCEP GFS. The percentage of cases improved at each forecast time for the NCEP GFS, the ECMWF IFS, and the two-model mean are shown in Fig. 3. In the ECMWF IFS, except during the 30–72-h forecast period, the percentages of cases improved are higher than 50%, averaging around 62% for the entire forecast period. For the two-model mean, the result is consistent with the NCEP GFS. The percentage of cases improved is higher than 50% during nearly the entire forecast period, with an average of 72%.

The case-averaged impact for cases between 2005 and 2009 at different forecast times in the NCEP GFS, the ECMWF IFS, and the two-model mean is displayed in Fig. 9. For the NCEP GFS (Fig. 9a), the dropwindsonde data lead to 20%–80% mean track error reduction and the statistically significant level is at least 90% at all forecast times. For ECMWF IFS (Fig. 9b), the mean track error reductions are 10%, 20%, and 60% at 24, 96, and 120 h, but are −4% and −30% at 48 and 72 h, respectively. Results at 72, 96, and 120 h are statistically significant at the 90% confidence level.

For the two-model mean result (Fig. 9c), the mean track error reductions are roughly 10%–15% for 24–72-h forecasts, but the statistical significance is below 90%. At 96 and 120 h, a mean track error reduction (significant at the 95% confidence level) of 50% and 90% is achieved, but the sample size is small. The two-model mean has
smaller track errors than individual members during most of the forecast time (especially at 96 and 120 h, explaining the statistical significance). This result demonstrates the advantage of the two-model mean over the individual models.

5. Results from T-PARC program in 2008

The impact of dropwindsonde data on track forecasts during T-PRAC has been studied by Weissmann et al. (2011) using the ECMWF IFS, JMA GSM, and NCEP GFS and the limited area Weather Research and Forecasting (WRF) model. In addition, Harnisch and Weissmann (2010) showed a beneficial influence on track forecast with the ECMWF IFS for Typhoon Sinlaku and Jangmi using mainly DOTSTAR dropwindsonde data in the vicinity of the storm. Aberson (2011) examined the impact of dropwindsonde data from T-PARC and the National Oceanic and Atmospheric Administration (NOAA) Hurricane Field Program on global TC forecasts by the NCEP GFS system. Jung et al. (2010) also conducted experiments examining impacts of dropwindsonde data with regional WRF model, and showed that the assimilation of dropwindsondes data results significantly improves the track forecasts of Typhoon Jangmi. In this section, the impact of dropwindsonde data from the T-PARC field experiment on the NCEP GFS is also examined. In particular, the forecast tracks of Typhoons Sinlaku and Jangmi (with the most abundant data observed during T-PARC) from NCEP GFS are presented.

Figure 10 shows the best and the worst three cases from among the 32 Sinlaku and Jangmi cases. In general, the assimilation of dropwindsonde data usually helps capture the timing of recurvature of Sinlaku and shows more improved tracks than degraded ones. However, the assimilation of dropwindsonde data leads to a westward track bias in Jangmi runs with larger degradations than improvements.

The track error reduction by the dropwindsonde data for each individual case is shown in Fig. 11. For most Sinlaku experiments, the improvement is marginal at the beginning of the forecast, but increases with forecast lead time. In contrast, for most Jangmi cases, the improvement is substantial at the beginning of the forecast period, but becomes negative as forecast time increases. Although negative impacts occur for Jangmi cases, the magnitudes are much smaller than those of the Sinlaku cases. Figure 12 shows the percentage of cases improved at each forecast time for Sinlaku and Jangmi. For Sinlaku,
the percentage of cases improved is above (under) 70% after (before) 36 h, and averages 71% for the entire forecast period. Nevertheless, the result for the Jangmi cases is opposite, and the percentage of improved cases is above (under) 60% before (after) 36 h and averages 47% for the entire forecast period. Because of the larger sample size of Sinlaku cases compared to Jangmi, the percentage of improved cases in T-PARC is closer to that of Sinlaku, with a forecast-period average of 65%.

The case-average track error statistics during T-PARC are shown in Fig. 13. For Sinlaku cases (Fig. 13a), because of the large sample size, the case-averaged track forecast...
error reductions at each forecast time are relatively constant. In addition to minor improvements at 24 h, statistically significant improvements of about 30%–40% are obtained during other forecast times. The Jangmi cases (Fig. 13b), in contrast, have much lower consistency in forecast track error reductions due to the relatively small sample size. The track error reduction of 40% in Jangmi cases at 24-h lead time is significant (although the number of cases in Jangmi is only eight, the t-test calculation still shows it well exceeds the 95% confidence level), whereas the degraded tracks obtained at other forecast times are not statistically significant. For all the T-PARC cases

Fig. 11. As in Fig. 2, but for the T-PARC cases in 2008.
except for the nonsignificant impact at 120 h, the dropwindsonde data significantly improves track forecasts by 20% during the entire forecast period.

6. Concluding remarks

Starting in 2003, a typhoon surveillance program, DOTSTAR, has been obtaining dropwindsonde measurements around TCs near Taiwan. Moreover, in the summer of 2008, the international field project T-PARC was conducted in the same region. Dropwindsondes were the major observation platform in both the surveillance and reconnaissance flights of the T-PARC program.

To further evaluate the impact of dropwindsonde data on typhoon track forecasts over the western North Pacific after Wu et al. (2007b), more cases from both the DOTSTAR and T-PARC projects are investigated. Two major operational global modeling systems, NCEP GFS and ECMWF IFS, are used to assess the impact of the dropwindsonde data. Control and denial runs (with and without dropwindsonde data) were conducted in both systems. The two-model mean of forecast tracks is also evaluated. The forecast tracks are verified against the JTWC best track.

The impact of dropwindsonde data on track forecasts is different between the NCEP and ECMWF systems. For the NCEP system, the assimilation of dropwindsonde data leads to track improvement (degradation) in approximately 60% (40%) of all cases, whereas the improvement in track forecast error is generally larger than the degradation. Overall, the mean 1- to 5-day track forecast error is reduced by about 10%–20% for both DOTSTAR and T-PARC cases (exceeding 90% t-test confidence level). However, for the ECMWF system, the impact is not robust for the entire forecast period. The case-average track error reduction is positive in the beginning and later forecast lead times, but turning negative in between. There are track improvements by using dropwindsonde data at forecast lead times of 96 and 120 h, but the sample size is small. Small average track degradations are significant at 72-h lead time. Larger impacts of the dropwindsonde data are found in 3- to 5-day forecasts when the two-model mean of the NCEP and ECMWF systems is examined, indicating the overall added value of
the dropwindsonde data in improving the track forecasts in the current operational modeling systems.

The influence of additional dropwindsonde observations during the two major typhoon events of T-PARC has also been evaluated by Weissmann et al. (2011), and the current results showing a higher influence in NCEP GFS and less significant impact in ECMWF IFS are consistent. This is likely related to lower-track forecast errors without dropwindsonde data in ECMWF, presumably a result of more extensive use of satellite data and four-dimensional variational data assimilation (4DVAR) in ECMWF in contrast to 3DVAR used in NCEP. In addition, Weissmann et al. (2011) showed that the cycling of analyses is essential to gain forecast improvements by additional observations in the ECMWF system. Thus, the current study may underestimate the full potential of forecast improvements in the ECMWF system as the majority of cases is performed in an uncycled mode.

This study summarizes the most updated results on the impact of dropwindsonde data on track forecasts from both DOTSTAR and T-PARC programs. Based on more cases available from the NCEP GFS system, more reliable statistics of dropwindsonde data on improving the track forecast are obtained. In other words, assimilation of dropwindsonde data could lead to 60% improvements in 1- to 5-day track forecasts and 10%–20% mean track error reduction with at least 90% confidence level.

Acknowledgments. The work is supported by the National Science Council of Taiwan through Grants NSC97-2111-M-002-016-MY3 and NSC-98-2111-M-034-002, the Office of Naval Research Grants N00014-05-1-0672 and N00173-08-1-G007, National Taiwan University Grant 97R0302, the Central Weather Bureau Grant MOTC-CWB-97-6-M-01, and the Academia Sinica. The authors thank all DOTSTAR team members for their efforts and all people who contributed to the observations during T-PARC. Additionally, support from ECMWF and NCEP for the data denial runs is acknowledged. Florian Harnisch and Martin Weissmann are part of the Deutsche Forschungsgemeinschaft (German Research Foundation) research unit PANDOWAE (http://www.pandowae.de).

T-PARC was sponsored by an international consortium from the United States (National Science Foundation, Office of Naval Research, Naval Research Laboratory, Air Force), Germany (DLR, Forschungszentrum Karlsruhe), Japan (Japanese Meteorological Agency), Korea (National Institute of Meteorological Research), and Canada (Environment Canada). The role of the National Center for Atmospheric Research Earth Observing Laboratory (NCAR EOL) in the campaign and data management is acknowledged.

REFERENCES


