A consensus track forecast for southern hemisphere tropical cyclones

Charles R. Sampson and James S. Goerss
Naval Research Laboratory, Monterey, California, USA

and

Ann J. Schrader
Science Applications International Corporation, Monterey, California, USA

(Manuscript received November 2004; revised March 2005)

Tropical cyclone (TC) track forecasting skill of operational numerical weather prediction (NWP) models and their consensus is examined for the southern hemisphere for the period 1992 to 2004. The TC track forecasting skill of the operational NWP models has been steadily improving. For the southern hemisphere, the typical 72 h model forecast error has decreased from roughly 550 km to 400 km over the past ten years and is now comparable to the typical 48 h model forecast error ten years ago. In this study we examine the performance of consensus aids that are formed whenever the TC track forecasts from at least two models from a specified pool of operational NWP models are available. The 72 h consensus forecast error has decreased from about 350 km to roughly 310 km over the past ten years. For a homogeneous dataset from the 2002-2004 seasons, the 72 h forecast errors for a consensus computed from a specified pool of 2, 3 and 5 models were 347, 342 and 312 km, respectively. While the addition of models to the consensus has a modest impact on forecast skill, it has a more marked impact on consensus forecast availability. For the 2002-2004 seasons, the forecast availability for 72 h consensus forecasts computed from a pool of 2, 3 and 5 models was 69%, 74% and 85%, respectively. Preliminary results using the 2003-2004 seasons indicate that the addition of two Australian models and a barotropic model increases the forecast availability by a few per cent at the 24 h, 48 h and 72 h forecast periods.

Introduction

Over the past decade the number of numerical weather prediction (NWP) models capable of producing high-quality tropical cyclone (TC) track forecasts has grown. Today, for the southern hemisphere, forecasters at the Joint Typhoon Warning Center (JTWC) routinely use TC track forecasts from eight operational NWP models. Two of these models are run operationally at the Fleet Numerical Meteorology and Oceanography Center (FNMOC), the Navy Operational Global Atmospheric Prediction System (NOGAPS; Hogan and Rosmond 1991; Goerss and

Corresponding author address: Charles R. Sampson, Naval Research Laboratory, 7 Grace Hopper Ave, Stop 2, Monterey, CA 93943-5502, USA.
Email: sampson@nrlmry.navy.mil

Jeffries 1994), and the Geophysical Fluid Dynamics Laboratory (GFDL) Hurricane Prediction System (GFDN; Kurihara et al. 1993, 1995, 1998; Rennick 1999). The United Kingdom Meteorological Office global model (UKMO; Cullen 1993; Heming et al. 1995), the National Centers for Environmental Prediction (NCEP) global spectral model (GFS; Lord 1993), and the fifth-generation Pennsylvania State University/National Center for Atmospheric Research (NCAR) mesoscale model run operationally by the Air Force Weather Agency (AFWA; Grell et al. 1995) complete the suite of models for which forecasts have been available to JTWC for at least three seasons. The two Australian regional models TC-LAPS (Davidson and Weber 2000) and the Tropical Limited Area Prediction System (TLAPS; Puri et al. 1998) and one barotropic model (WBAR; Weber 2001) have been added recently to bring the total to eight. A timeline indicating when TC track forecasts from these eight NWP models became available to the forecasters at JTWC is shown in Fig. 1.

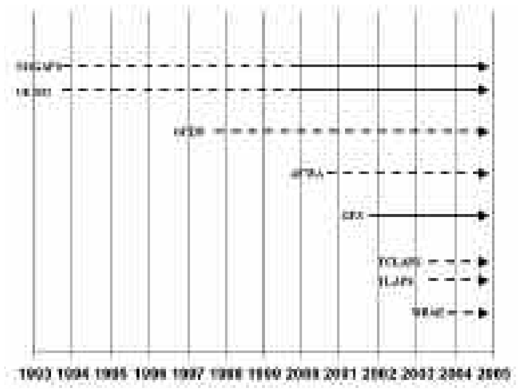
The benefits of consensus forecasting have long been recognised by the meteorological community (Sanders 1973; Thompson 1977). Leslie and Fraedrich (1990) and Mundell and Rupp (1995) applied this approach to TC track prediction and illustrated the forecast improvement that resulted from using linear combinations of forecasts from various TC track prediction models. Goerss (2000) first illustrated the superior TC track forecast performance of consensus forecasts formed using operational NWP models. Currently consensus forecast aids are routinely used by the forecasters at both JTWC and the National Hurricane Center.

In the next section we describe how the TC forecast tracks from the individual NWP models are prepared for use and how the consensus forecast is determined. In subsequent sections we examine the forecast performance for the southern hemisphere of individual NWP models and various consensus forecasts and present a summary of our results followed by a discussion of their implications for the future.

Methods

Forecast tracks discussed in this paper are processed as they are in an operational environment using the Automated Tropical Cyclone Forecasting System (ATCF; Sampson and Schrader 2000). Since forecast track outputs for the NWP models become available to the forecaster six or 12 hours after NWP model runtime, they arrive too late to be used directly. Instead, the NWP model tracks are interpolated to intermediate times, and then interpolated positions are relocated to reflect the forecaster-analysed (best track) position. The version of the interpolator used in this study is similar to that described in Goerss et al. (2004) with one exception – cubic spline interpolation has been replaced by linear interpolation. The names of the interpolated tracks are: NGPI for NOGAPS, EGRI for the UKMO model, GFNI for GFDN, JAVI for the NCEP GFS, AFWI for AFWA, TCLI for the TC-LAPS, TLAI for the TLAPS and WBAI for WBAR.

Fig. 1 **Timeline indicating NWP model forecast track availability at JTWC. Dashed lines indicate forecasts to 72 h while solid lines indicate forecasts to 120 h. Vertical bars indicate start of calendar year.**



A consensus for a given forecast period is a simple average of the interpolated positions which pass the quality control described in Goerss et al. (2004). An attempt is made to compute a consensus forecast at the 12 h, 24 h, 36 h, 48 h, 72 h, 96 h and 120 h forecast periods. This consensus is computed if two or more members exist for a given forecast period. If less than two members exist, the consensus forecast is not computed.

The seasons* are grouped into four distinct periods for statistical analysis: 1993-1997, 1998-2001, 2002-2004 and finally 2003-2004. The purposes of this division are: (a) to obtain sample sizes large enough to produce stable statistical results; and (b) to group years with approximately the same number of available models. The two-model consensus contains the NOGAPS and the UKMO model which were the only NWP models available during the 1993-1997 seasons. The NOGAPS, the UKMO model and GFDN were available during the 1998-2001 seasons and comprise the three-model consensus. The 2002-2004 seasons featured five models (NOGAPS, the UKMO model,

* At JTWC, a southern hemisphere tropical cyclone season is defined as extending from July to June. For example, the 2004 season is from 1 July 2003 to 30 June 2004.

GFDN, the AFWA model and GFS) that comprise the five-model consensus. During the 2003-2004 seasons, TC-LAPS, TLAPS and WBAR became available and brought the suite of available models to eight. Although the sample size for the complete suite of eight models is small, results will be presented as they are of interest to southern hemisphere tropical cyclone forecasters. Statistical analysis is limited to intensities (one-minute mean maximum wind speeds) of 13 m/s (25 kn) or greater in the JTWC best tracks.

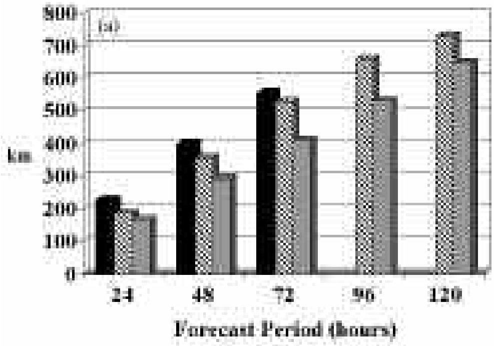
Results

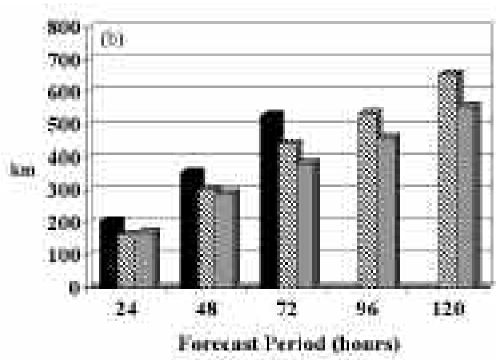
First, we examine the forecast performance of the interpolated versions of the two NWP models that have been available to the JTWC forecasters since the 1993 southern hemisphere season: NOGAPS and UKMO. The forecast errors for the NGPI (the interpolated NOGAPS) and EGRI (the interpolated UKMO model) for the 1993-1997, 1998-2001 and 2002-2004 seasons are displayed in Fig. 2. We see that the 24 h forecast error for NGPI has declined over the decade from about 230 km to 170 km, the 48 h forecast error has declined from about 400 km to about 300 km, and the 72 h forecast error has declined from 550 km to about 410 km. The 24 h, 48 h and 72 h forecast errors for EGRI have declined from about 210 km to 170 km, 350 km to 290 km, and 530 km to 380 km, respectively, over the decade. Generally, these NWP model forecasts have improved such that the 72 h forecasts today are about as accurate as the 48 h forecasts from the early 1990s.

Mean forecast errors for the five-model consensus and each of its members are shown in Fig. 3. For the period of record, the 2002-2004 seasons, the mean forecast errors of the five-model consensus are lower than those of any of its members. Since this comparison is homogeneous, it requires that all of the consensus members are available to form the five-model consensus. Operationally, we only see a consensus computed using all five members about 25 per cent of the time.

A homogeneous comparison of 2002-2004 season forecast performance for consensus forecasts is shown in Fig. 4(a). The two-model consensus is a consensus of the two models available since the 1993 season (NGPI and EGRI), the three-model consensus is a consensus of three models available since the 1998 season (NGPI, EGRI and GFNI), and the five-model consensus is a consensus of five models available since 2002 (NGPI, EGRI, GFNI, AFWI and JAVI). The number of model forecasts available to each consensus varies since we only require forecasts from two or more of the models to compute the consensus. When only NGPI and EGRI

Fig. 2 **Mean position errors (km) for (a) the interpolated NOGAPS forecasts – NGPI and (b) the interpolated UKMO model forecasts – EGRI for the southern hemisphere seasons 1993-1997 (solid black), 1998-2001 (hatched) and 2002-2004 (solid grey).**





are available, the two-model, three-model and five-model consensus forecasts are all the same forecast. It is when GFNI, AFWI and/or JAVI is/are available that the two-model, three-model and five-model consensus forecasts may differ. Inclusion of more models in the consensus has resulted in small but consistent gains in skill. The 24 h two-model, three-model and five-model consensus mean forecast errors were 151 km, 148 km and 140 km, respectively. We see similar results for all forecast lengths out to 72 h where two-model, three-model and five-model consensus mean forecast errors are 347 km, 342 km and 312 km, respectively. Errors for the three-model consensus are not shown at 96 h and 120 h because GFNI was not available beyond 72 h (Fig. 1). On the other hand, the five-model consensus includes JAVI, which extends to 120 h. The five-model consensus mean forecast errors at 120 h are 499 km compared with 529 km for the two-model

Fig. 3 Mean forecast errors (km) for NGPI (black), EGRI (descending hatch), JAVI (ascending hatch), GFNI (vertical hatch), and AFWI (horizontal hatch) and a five-model consensus (grey). The data are a homogeneous set from the 2002-2004 seasons. The number of cases is 402, 300 and 194 for 24 h, 48 h and 72 h, respectively.

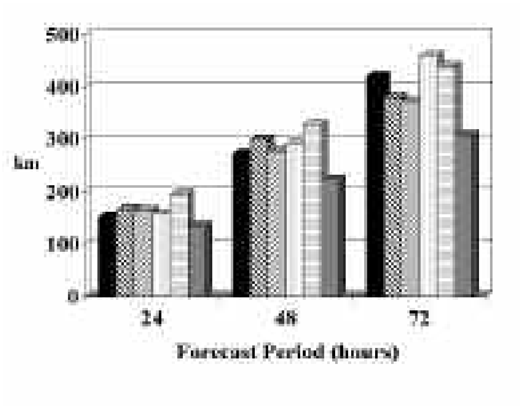
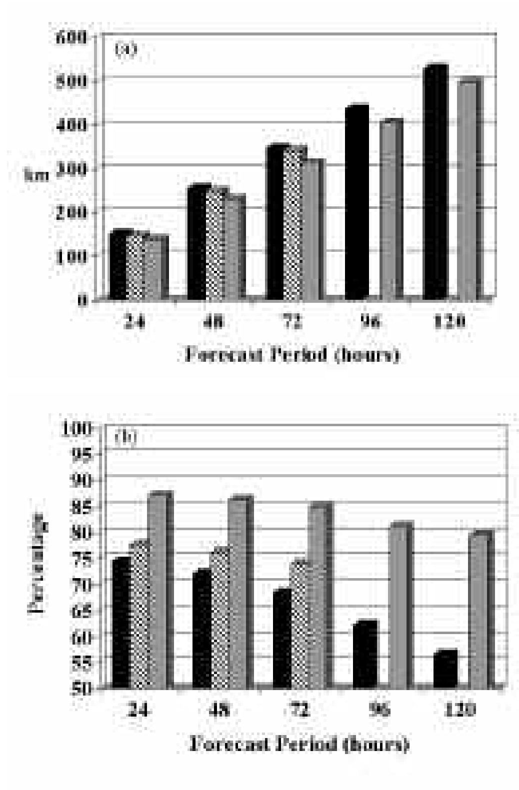


Fig. 4 (a) Mean consensus forecast errors (km) and (b) forecast availability relative to CLIPER (per cent) for the two-model consensus (black), the three-model consensus (hatched), and the five-model consensus (solid grey). The data are a homogeneous set from the 2002-2004 seasons. Until late in 2004, the two-model and three-model consensus forecasts were identical at 96 h and 120 h. Sample sizes are 1197, 979, 763, 555 and 380 at 24 h, 48 h, 72 h, 96 h and 120 h, respectively.



and three-model consensus forecasts. At all forecast lengths, the consensus outperformed the best individual interpolated model in the mean for this period of record. It is possible to find an individual model that outperforms the five-model consensus for a short period, but it is difficult to pre-select that model.

We have seen that addition of models to the consensus results in small but consistent gains in skill. However, this is not the only benefit of making more models available to the consensus. By increasing the number of models in the specified pool available to the consensus, we make it more likely that forecasts from at least two models will be available and that a consensus forecast can be formed. In Fig. 4(b), the availability percentages for the various consensus models for the 2002-2004 seasons are displayed. We define the forecast availability percentage to be the percentage of the time that consensus forecasts were available with respect to a southern hemisphere CLIPER forecast (Neumann and Randrianarison 1976). We choose the southern hemisphere CLIPER for a forecast availability baseline because it is one of very few models that produce forecasts nearly 100 per cent of the time the forecaster runs the objective guidance on the ATCF. Normally we would use the JTWC forecast as the baseline, but JTWC only forecasts for roughly 50 per cent of the six-hourly forecast cycles and usually only forecasts out to 48 h. The availability percentages for the two-model, three-model and five-model consensus forecasts at 72 h were 69, 74 and 85 per cent, respectively. By increasing the number of models in the specified pool from 2 to 5, we have significantly increased the percentage of the time that forecasts from at least two of the models are available so that a consensus forecast can be created. For all forecast lengths, the availability of the two-model consensus ranged from 57 to 75 per cent while that for the five-model consensus ranged from 79 to 87 per cent. To an operational forecaster, this increase in availability may be just as valuable as the increase in forecast skill.

Three additional NWP models have recently become available to JTWC (Fig. 1). Although the period of record is short and the number of forecasts is small, an eight-model consensus increases the forecast availability (using CLIPER as the baseline) from 87 per cent to 90 per cent at 24 h, from 86 per cent to 90 per cent at 48 h, and from 84 per cent to 88 per cent at 72 h for the 2003-2004 seasons. Results of a homogeneous comparison of the eight-model and five-model consensus forecast errors shows less than one per cent difference between the two at all forecast periods.

Conclusions

In conclusion, we have seen that the TC track forecasting skill of NWP models for the southern hemisphere has improved over the past decade. This improvement has contributed to a similar improvement in consensus forecasts created from these NWP models. We have also seen that the addition of models to the consensus results in improvements to both consensus track forecast skill and consensus forecast availability. Finally, we have seen the forecasters successfully integrate consensus forecasting into their operational procedures at JTWC. While a consensus demonstrates superior performance with respect to the individual NWP models in the long run, forecasts from individual models can certainly outperform a consensus in the short run. The forecasters use consensus as a baseline or starting point, then modify the forecast as they see fit. In the future, as the TC track forecasting skill of NWP models continues to improve and as more high-quality NWP models become available, we can look forward to further improvement in consensus track forecast skill and forecast availability.

Acknowledgments

The authors would like to acknowledge Jim Gross, Mark DeMaria, John Knaff and Tim Marchok for their work with the interpolator and statistics code; Mike Fiorino for his invaluable vortex trackers; Ted Tsui, Sam Brand and Andrew Burton for editorial comments; and the entire staff at JTWC for advice, and the care and feeding required to make a consensus work in an operational setting. This work was supported by the Oceanographer of the Navy through the program office at the PEO C4I and Space (PMW-180), Program Element 0603207N.

References

Cullen, M.J.P. 1993. The Unified Forecast/Climate Model. *Met. Mag., 122*, 81-122.

Davidson, N.E. and Weber, H.C. 2000. The BMRC high-resolution tropical cyclone prediction system: TC-LAPS. *Mon. Weath. Rev., 128,* 1245-65.

Goerss, J.S. 2000. Tropical cyclone track forecasts using an ensemble of dynamical models. *Mon. Weath. Rev., 128*, 1187-93.

Goerss, J.S. and Jeffries, R.A. 1994. Assimilation of synthetic tropical cyclone observations into the Navy Operational Global Atmospheric Prediction System. *Weath. forecasting, 9*, 557-76.

Goerss, J.S., Sampson, C.R. and Gross, J.M. 2004. A history of western North Pacific tropical cyclone track forecast skill. *Weath. forecasting, 19*, 633-8.

Grell, G.A., Dudhia, J. and Stauffer, D.R. 1995. A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5). *NCAR Tech. Note NCAR/TN-398 + STR*, 122 pp.

Heming, J.T., Chan, J.C.L. and Radford, A.M. 1995. A new scheme for the initialization of tropical cyclones in the UK Meteorological Office global model. *Met. Appl., 2*, 171-84.

Hogan, T.F. and Rosmond, T.E. 1991. The description of the Navy Operational Global Atmospheric Prediction System's spectral forecast model. *Mon. Weath. Rev., 119*, 1786-815.

Kurihara, Y., Bender, M.A. and Ross, R.J. 1993. An initialization scheme of hurricane models by vortex specification. *Mon. Weath. Rev., 121*, 2030-45.

Kurihara, Y., Bender, M.A., Tuleya, R.E. and Ross, R.J. 1995. Improvements in the GFDL hurricane prediction system. *Mon. Weath. Rev., 123*, 2791-801.

Kurihara, Y., Tuleya, R.E. and Bender, M.A. 1998. The GFDL hurricane prediction system and its performance in the 1995 hurricane season. *Mon. Weath. Rev., 126*, 1306-22.

Leslie, L.M., and Fraedrich, K. 1990. Reduction of tropical cyclone position errors using an optimal combination of independent forecasts. *Weath. forecasting, 5*, 158-61.

Lord, S.J. 1993. Recent developments in tropical cyclone track forecasting with the NMC global analysis and forecast system. Preprints, *20th Conf. on Hurricanes and Tropical Meteorology,* San Antonio, TX, Amer. Met. Soc., 290-1.

Mundell, D.B. and Rupp, J.A. 1995. Hybrid forecast aids at the Joint Typhoon Warning Center: Applications and results. Preprints, *21st Conf. on Hurricanes and Tropical Meteorology*, Miami, FL, Amer. Met. Soc., 216-18.

Neumann, C.J. and Randrianarison, E.A. 1976. Statistical prediction of tropical storm motion over the Southwest Indian Ocean. *Mon. Weath. Rev., 104,* 76-85.

Puri, K., Dietachmayer, G., Mills, G.A., Davidson, N.E., Bowen, R.A. and Logan, L.W. 1998. The new BMRC Limited Area Prediction System, LAPS. *Aust. Met. Mag., 47*, 203-23.

Rennick, M.A. 1999. Performance of the Navy's tropical cyclone prediction model in the western North Pacific basin during 1996. *Weath. forecasting, 14*, 3-14.

Sampson, C.R. and Schrader, A.J. 2000. The Automated Tropical Cyclone Forecasting System (Version 3.2). *Bull. Am. Met. Soc., 81*, 1131-240.

Sanders, F. 1973. Skill in forecasting daily temperature and precipitation: Some experimental results. *Bull. Am. Met. Soc., 54,* 1171-9.

Thompson, P.D. 1977. How to improve accuracy by combining independent forecasts. *Mon. Weath. Rev., 105*, 228-9.

Weber, H.C. 2001. Hurricane track prediction with a new barotropic model. *Mon. Weath. Rev., 129*, 1834-58.