Southern Hemisphere Tropical Cyclone Intensity Forecast Methods Used at the Joint Typhoon Warning Center, Part III: Forecasts Based on a Multi-model Consensus Approach

Charles R. Sampson
Naval Research Laboratory
Monterey, California
USA

John A. Knaff
NOAA/NESDIS
Center for Satellite Research and Applications
Fort Collins, Colorado
USA

Submitted to
Australian Meteorological Magazine
June 10, 2008

Corresponding author address: Charles R. Sampson, NRL,
7 Grace Hopper Ave. Stop 2, Monterey, CA 93943-5502
(email: sampson@nrlmry.navy.mil)

ABSTRACT

Consensus forecast aids are created by combining output from individual forecast aids and are now an integral part of operational tropical cyclone forecasting at the Joint Typhoon Warning Center (JTWC). These consensus aids generally have lower average errors than individual forecast aids and benefit from the skill and independence of their members. This study conducts experiments with intensity forecast aids on three seasons of Southern Hemisphere data (2006-2008). First, the skill of the individual forecast aids is assessed, and then equally weighted consensus aids are developed. A consensus of the top performing intensity forecast aids is found to generally outperform the individual members, but the skill of these aids is still quite low. Adding less skillful members to the consensus generally degrades the skill.

1. Introduction

The meteorological community recognized the benefits of consensus forecasting as far back as the 1970s (Sanders 1973; Thompson 1977). A subjective form of consensus forecasting has been applied to tropical cyclone track forecasting for decades, and more recently objective consensus methods have become popular (Burton et al. 2007). The more successful attempts focused on dynamical track models because they were, on average, the best performers (Goerss 2000; Williford et al. 2003).

Some of these dynamical track models also produce forecasts of tropical cyclone intensity (maximum 1-min mean wind at 10 m elevation). Most are handicapped by resolution, initialization and parameterizations of the smaller scale processes (Knaff et al. 2007), and thus cannot simulate the inner core dynamics of a tropical cyclone. Consequently, the only skillful intensity forecast models are high-resolution models designed specifically for tropical cyclone forecasting (DeMaria et al. 2007). Dynamical models that routinely produce intensity forecasts for the Southern Hemisphere are: the Naval Operational Global Atmospheric Prediction System (NOGAPS; Hogan and Rosmond 1991), the Geophysical Fluid Dynamics Laboratory Hurricane Prediction System run with NOGAPS initial and boundary conditions (GFDN; Rennick 1999), the United Kingdom Meteorological Office global model (UKM; Heming et al. 1995), the National Weather Service (NWS) global spectral model (GFS; Lord 1993), the fifth-generation Pennsylvania State University/National Center for Atmospheric Research Mesoscale Model (MM5; Grell et al. 1995) run operationally by the U. S. Air Force Weather Agency (AFWA), and finally the TC-Limited Area Prediction System (TC-LAPS; Davidson and Weber 2000) and the Tropical eXtended Area Prediction System (TXLAPS; Australian Bureau of Meteorology 2005) run by the Australian Bureau of Meteorology. Table 1 provides a summary of these and other aids used in this study.

There are also two simple models designed specifically to produce intensity forecasts. The Statistical Typhoon Intensity Prediction System (STIPS; Knaff and Sampson 2008b) is a statistical-dynamical model that does not attempt to resolve the tropical cyclone inner core; rather it forecasts changes in intensity through regression of large-scale environmental parameters (vertical wind shear, sea surface temperature, relative humidity, temperature, and low-level vorticity) and an empirical inland decay model (DeMaria et al. 2006). Even though STIPS has lower mean absolute error than the NWP models, it does not forecast rapid intensification (Knaff et al. 2007) since it makes no effort to resolve the inner core. The Coupled Hurricane Intensity Prediction System (CHIPS; Emanuel et al. 2004), on the other hand, permits high radial resolution of the inner core region. Inputs to this model now include a thermodynamic state, wind shear, sea surface temperature and climatological mixed layer dept and sub-mixed layer thermal stratification (Emanuel et al. 2008).

For an intensity skill baseline we will use the simple climatology and persistence statistical model, the 5-day Statistical Hurricane Intensity Forecast (ST5D; Knaff and Sampson 2008a). This model is also a poor predictor of rapid intensification and decay since it is designed to minimize mean forecast errors, yet its' seasonal average performance is still competitive. Other statistical intensity aids (e.g., climatology, climatology and persistence, analogs, extrapolation and hybrids) exist, but they do not perform as well as ST5D (Knaff and Sampson, 2008a) and are not discussed further.

Intensity forecasts for the Northern Hemisphere have been shown to have relatively little skill compared to skill baselines, so it is likely that the benefits of using consensus techniques on intensity forecast aids in the Southern Hemisphere are small, as they are in Northern Hemisphere (Sampson et al. 2008). Still, it is important to perform a study that sets an intensity consensus baseline to assess further improvements. Therein lies the purpose of this study. First, it will first assess the skill of the existing guidance, and then determine whether superior skill can be obtained using a simple equally weighted consensus of the most skillful members. Finally, a consensus aid will be proposed for use in evaluating other more complex consensus techniques.

2. Data

The data used for this study are taken from the operational archive at the JTWC as stored on the Automated Tropical Cyclone Forecasting System (Sampson and Schrader 2000) for which a description is given in JTWC (2008). The authors attempted to get a large homogenous data set for this study so that statistics would be stable. As seen in Table 1, the GFDN intensity forecast aid was available as early as 1998. By 2003 there were five new and interesting intensity forecast aids available for this study. However, the skill baseline (ST5D) became available in 2004 and a few of the better performing aids only became available in 2005 and 2006. Hence, the seasons chosen were 2006-2008. The Southern Hemisphere season at JTWC starts on July 1st of the preceding year and extend through June 30th of the year.

3. Methods

Intensity forecast aids are characterized as either *early* or *late*, depending on whether or not they are available to the JTWC forecaster during the forecast cycle. For example, consider the 1200 UTC (12Z) forecast cycle, which begins with the 12Z synoptic time and ends with the release of an official forecast at 15Z. The 12Z run of the GFDN model is not complete nor is its forecast aid (also named GFDN) available to the forecaster until about 16Z. This is about an hour after the official JTWC forecast is released, and thus the 12Z GFDN would be considered a late forecast aid because it could not be used to prepare the 12Z forecast.

All the dynamical model and specialized model (STIPS and CHIPS) forecast aids available to JTWC are late models. To alleviate the problem, a simple method is used to take the latest available forecast aid from a run of a late model and adjust it to the current synoptic time and initial conditions. For example, the GFDN forecast aid for hours 6-126 from the previous (06Z) run would be adjusted, or shifted, so that the 6-h forecast (valid at 12Z) would exactly match the observed 12Z position and intensity of the tropical cyclone. The adjustment process creates an "early" version of the GFDN forecast aid for the 12Z forecast cycle that is based on the most current available guidance. The adjustment algorithm is called "the interpolator" and the adjusted aids are called "interpolated" aids. The version of the interpolator used in this study is similar to that described in Sampson et al. (2006). The name of the interpolated forecast aid is usually the acronym of the late forecast aid with an "I" substituted for the last letter (Table 1). One exception used in this study is GFDN, for which GFNI is the acronym for the interpolated forecast aid.

All consensus forecasts discussed within are equally weighted averages of forecasts from the available consensus members. In the consensus algorithm, an attempt is made to compute a consensus forecast at each forecast period (12, 24, 36, 48, 72, 96 and 120 h). A consensus is computed if *two or more* consensus members exist for a given forecast period. If fewer than two members exist, the consensus computation is aborted for this and subsequent time periods.

Results presented are from recomputed interpolated aids and consensus forecasts using methods described above. The purpose of this is to ensure that all results are computed using the same version of the interpolator. Average differences in performance between recomputed interpolations and those produced in operations are generally less than 1%.

Forecasts are verified only when the best track intensity is greater than 20 kt (10.3 m/s) and only when the system is a tropical or subtropical cyclone. Interpolated forecast aids are used as described above. If 6-h interpolated forecast aids are not available, 12-h interpolated forecast aids are computed. The 12-h interpolations occur approximately 15% of the time or less for models that are available every six hours. Performance is discussed through use of skill charts. The measure of skill in these charts is defined as:

$$\text{Skill} = 100^* \text{ (baseline error-model error)/baseline error} \quad (1)$$

Thus, skill is positive when the forecast aid error is less than that of the baseline forecast aid. A one-tailed Student's t-test at the 95% level with serial correlation of 30 h removed (Neumann et al. 1977) is also employed as a method to test significance in forecast error differences between individual intensity forecast aids.

3. Results

First an attempt is made to identify forecast aids that may be of use in a consensus. A comparative verification of the various intensity forecast aids is shown in Fig. 1. The intensity forecast skill for these forecast aids is generally less than skill associated with track forecasts. For example, GFNI track forecast skill at 48 h is approximately 25% while its intensity forecast skill at 48 h is negative. However, there are a number of skillful intensity aids at 48 h, and many of them are members of the STIPS ensemble[1] described in detail in Appendix A.

The STIPS ensemble is run in lieu of running STIPS on the JTWC official forecast track. A benefit of running this ensemble is that it is available to the forecasters during their intensity forecast while the traditional STIPS (run on the JTWC official forecast track) is not. Another benefit of the STIPS ensemble is that its forecasts are available every six hours and its forecasts extend beyond 48 h. In contrast, the traditional STIPS is available every 12 hours and usually extends to 48 h. The final benefit of this ensemble is that it presents a range of solutions dependent on tracks. Although the ensemble size is small (at most seven members in the Southern Hemisphere) its members may provide a variety of forecast scenarios; for instance, some tracks may decay over land while others stay over warm ocean water and intensify. The STIPS ensemble performance, measured in mean absolute intensity errors, is within 3% of the traditional STIPS. Of note is that the traditional STIPS performance is significantly superior (2.8%) at 12 h, but it isn't available to the forecasters for use. The STIPS ensemble is generally 1-2% better at 36 and 48 h, but these results are not significant.

Mean forecast error is important in forming a consensus, but so is independence. An equation for the consensus mean error (μ_c) is

$$\mu_c = \mu/(n)^{1/2}, \qquad (2)$$

where μ is the mean of the members (assumed to all be equal to each other) and n is the number of independent members (Sampson et al. 2008). This equation implies that increasing the number of independent members reduces the mean error of the consensus. In operations, the intensity forecast errors are not entirely independent so n is replaced by the effective degrees of freedom n_e. Two members with errors that are completely independent ($n_e = n = 2$) can produce a consensus with a mean error reduction of approximately 30%. On the other hand, two members with errors having little independence ($n_e = 1.1$) would only produce an improvement of approximately 5% over the member mean. Model independence is generally not known a priori, and therefore a trial and error approach to find consensus members is generally required.

The first trial is the STIPS ensemble (the mean of all the xxS1 member forecasts, where xx identifies a particular member listed in Appendix A). The results of this trial are shown in Fig. 2 where the skill baseline is actually the STIPS ensemble, which itself has skill

[1] The term ensemble is reserved for aids constructed from forecasts of a single model while the term consensus

relative to the skill baseline ST5D. An important result from this test is that the STIPS ensemble performs about as well as the members, but certainly no better than the top performers.

It is suspected that the members lack the independence (high n_e in Eq. 2) necessary to reduce the mean ensemble errors. To verify this, each possible two-member ensemble was computed and its mean forecast errors at 48 h was verified against the average of the mean forecast errors for its two input members. For two-member ensembles with 300 or more cases the forecast performance improvements in mean forecast errors ranged from 1% to 3%, indicating that independence is quite low (1.02 to 1.06). On the other hand, forecast improvements for two-member consensus aids computed from other aids used in this study (300 or more cases) are higher (4% to 8%), indicating more independence (1.06 to 1.17). These results are consistent with our expectations and are somewhat lower than the results found for two-model forecast track aids, which averaged 1.54 for aids that included a barotropic model (WBAI) and 1.34 for aids that didn't (Sampson et al. 2006).

Forecast availability is also an important consideration. A forecast aid that performs well may not be as useful to a forecaster if it is only available for 50% of the official forecasts. The availability of the STIPS ensemble at 48 h (651 cases) is approximately 5% higher than the STIPS ensemble member with the highest availability (NGS1), and its availability is more than double the best performing member at 48 h (TCS1).

In the next trial interpolated forecasts are added to the STIPS ensemble members one at a time, resulting in four new consensus aids. The performance of the STIPS ensemble and the four potential consensus aids is shown in Fig. 3. To be consistent with Sampson et al. (2008), GFNI is the first interpolated aid added to form ST11 even though it is not the top performer of the remaining aids. The next three top performers (CHII, TCLI and UKMI) are

---

is used for aids constructed from forecasts of more than one model.

then added one at a time to yield an ST12, ST13 and ST14, respectively. Immediately apparent is that improvements in skill are small. The largest improvement in skill is gained by adding the first interpolated aid (GFNI) in with the STIPS ensemble members. These improvements of approximately 2-3% are significant at the 24- and 48-h forecast periods, but not at the 72-h forecast period[2]. Addition of more aids to this consensus, as done for ST12 through ST14, provides mixed results, none of which are significant. The top performer at 24 h is the STIPS ensemble with the addition of GFNI, CHII and TCLI, while the STIPS ensemble with GFNI and CHII is the top performer at 48 and 72 h. The ST10, ST11, ST12, ST13 and ST14 forecasts were available 74%, 78%, 89%, 90% and 91% of the 465 JTWC forecasts for 48 h, respectively. So from an availability standpoint ST12, ST13 and ST14 provide the best guidance.

4. Summary and Conclusions

Experiments with evenly weighted consensus were conducted on JTWC Southern Hemisphere objective aid data (2006-2008). First, 48-h intensity forecast errors from forecast aids were evaluated to find potential candidates for the formation of a consensus. The STIPS ensemble formed the first trial. Then four of the most skillful aids were added, one at a time, to a consensus and the consensus skill was evaluated. Of the consensus aids attempted, the STIPS ensemble with the addition of GFNI, CHII and TCLI was the top performer at 24 h, and the STIPS ensemble with the addition of GFNI and CHII was the top performer at 48 and 72 h. These consensus forecast aids are likely to have lower mean forecast errors than would the individual models that form the consensus. These consensus aids could serve as deterministic intensity forecast benchmarks for other consensus or ensemble methods such as the methods reviewed in Burton et al. (2007). They may also provide operational forecast

[2] A trial with the STIPS ensemble members and CHII instead of GFNI (not shown) was also performed. The

guidance. It is suspected that improvements in the consensus members and additional members would further benefit this simple, evenly weighted intensity consensus approach.

Acknowledgments: The authors would like to acknowledge Ann Schrader for her work with the ATCF, Jim Gross for development of the interpolator, Mike Fiorino for his vortex trackers, Mark DeMaria for his work with the statistical analysis code, and the JTWC for their efforts to maintain the ATCF best track and objective aid files. This project is supported by grants from the Office of Naval Research and the National Oceanic and Atmospheric Administration. The views, opinions, and findings contained in this report are those of the authors and should not be construed as an official National Oceanic and Atmospheric Administration or U.S. government position, policy, or decision.

---

results were similar to those of ST11.

APPENDIX A

The STIPS Consensus

The STIPS consensus (ST10) is constructed using 2-7NWP model interpolated track forecasts available at approximately synoptic time + 1.5 hours. The interpolated track forecasts chosen were those of the operational track consensus used at JTWC for the 2005-2006 seasons.

Ideally, consensus members should be run through STIPS with thermodynamic and dynamic input from the model corresponding to the interpolated track. This would provide the most independence in the members, which should lead to a larger reduction in the consensus mean. It would also provide model fields with a vortex structure collocated with the interpolated track, and thus should provide for more realistic STIPS computations (e.g., shear computation) for that member. Since the authors could not obtain complete model field input for all of the member models, a compromise solution was constructed. For five of the interpolated model tracks (NGPI, GFNI, UKMI, and AVNI) STIPS is run with dynamic fields (u and v components of the wind) from the model, and NOGAPS data for the other STIPS field data input (temperature, relative humidity and geopotential height And finally, NOGAPS was used for all field data input to run the remaining three interpolated tracks (GFNI, TCLI, AFWI). Table A1 provides an overview of the STIPS consensus members and their input.

The version of STIPS used for ST10 has upgrades regarding decay effects over land (DeMaria et al. 2006). A forecast aid run with this newer version of STIPS on the JTWC track (STFD) was also produced for comparison with ST10. The comparison is not entirely fair since the current operational configuration delays STFD sufficiently so that it is produced about an hour a later than the operational intensity forecast, and is therefore a late forecast aid.

REFERENCES

Australian Bureau of Meteorology 2005. Operational Implementation of TXLAPS_PT375 Bulletin No. 59, May 2005 [Available on-line from http://www.bom.gov.au/nmoc/bulletins/59/].

Burton, A., Caroff, P., Franklin, J., Fukada, E., Lee, T. C. , Sampson, B., and Smith, T. 2007. Sharing experiences in operational consensus forecasting. Sixth International Workshop on Tropical Cyclones, San Jose, Costa Rica [Available on-line from http://severe.worldweather.org/iwtc/].

DeMaria, M, Knaff, J. A., Kaplan, J. 2006. On the Decay of Tropical Cyclone Winds Crossing Narrow Landmasses**.** *J. App. Meteor. and Clim., 45*, 491–499.

____, ____, and Sampson C. 2007. Evaluation of long-term trends in tropical cyclone intensity forecasts. *Meteor. Atmos. Phys*., *97*,19-28.

Davidson, N. E., and Weber, H. C. 2000: The BMRC high-resolution tropical cyclone prediction system: TC-LAPS. *Mon. Wea. Rev*., *128*, 1245-1265.

Emanuel, K., Sundarajan, R. , and Williams, J. 2008. Results from Downscaling IPCC AR4 Simulations. *Bull. Amer. Meteor. Soc.*, *89*, 347-367.

____, Desautels C. , Holloway, C. and Korty, R. 2004. Environmental Control of Tropical Cyclone Intensity. *J. Atmos. Sci*., *61*, 843-858.

Goerss, J. S. 2000. Tropical cyclone track forecasts using an ensemble of dynamical models. *Mon. Wea. Rev.*, *128*, 1187-1193.

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. *Meteor. Appl*., *2*, 171-184.

Hogan, T. F., and Rosmond, T. E. 1991. The description of the Navy Operational Global Atmospheric Prediction System's spectral forecast model. *Mon. Wea. Rev*., *119*, 1786-1815.

Joint Typhoon Warning Center 2008. The Annual Tropical Cyclone Reports [Available on-line from https://metocph.nmci.navy.mil/jtwc.php].

Knaff, J., Guard, C., Kossin, J., Marchok, T., Sampson, B., Smith, T., and Surgi, N., 2007. Operational Guidance and skill in forecasting structure change. Sixth International Workshop on Tropical Cyclones, San Jose, Costa Rica [Available on-line from http://severe.worldweather.org/iwtc/].

_____, and Sampson, C. R. 2008. Southern Hemisphere Tropical Cyclone Intensity Forecast Methods Used at the Joint Typhoon Warning Center, Part I: Control forecasts based on climatology and persistence. *Aust. Met. Mag*., Submitted.

_____, and ____ 2008. Southern Hemisphere Tropical Cyclone Intensity Forecast Methods Used at the Joint Typhoon Warning Center, Part II: Forecasts based on a statistical-dynamical approach. *Aust. Met. Mag*., Submitted.

_____, ____, and DeMaria, M. 2005. An operational statistical typhoon intensity prediction scheme for the Western North Pacific. *Wea. Forecasting*, *20*, 688–699.

____, DeMaria M., Sampson, C. R., and Gross, J. M. 2003. Statistical, Five-Day Tropical Cyclone Intensity Forecasts Derived from Climatology and Persistence. *Wea. Forecasting*, *18*, 80-92.

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. Meteor. Soc., 290-291.

Neumann, C. J.,Lawrence, M. B., and Caso, E. L. 1977. Monte Carlo significance testing as applied to statistical tropical cyclone models. *J. Appl. Meteor.*, *16*, 1165-1174.

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

Sampson, C. R., and Schrader, A. J. 2000. The Automated Tropical Cyclone Forecasting System (Version 3.2). *Bull. Amer. Meteor. Soc.*, *81*, 1131-1240.

_________, Goerss, J. S., and Weber, H. C. 2006. Operational performance of a new barotropic model (WBAR) in the western North Pacific basin. *Wea. Forecasting*, **21**, 656-662.

_________, Franklin, J. L., Knaff, J. K., and DeMaria, M. 2008. Experiments with a simple tropical cyclone intensity consensus. *Wea. Forecasting*, *23*, 304-312.

Sanders, F. 1973. Skill in forecasting daily temperature and precipitation: Some experimental results. *Bull. Amer. Meteor. Soc.*, *54*, 1171-1179.

Thompson, P. D. 1977. How to improve accuracy by combining independent forecasts. *Mon. Wea. Rev.*, *105*, 228-229.

Williford, C. E., Krishnamurti, T. N., Correa-Torres, R., Cocke, S., Christidis, Z., and Kumar, T. S. V. V. 2003. Real-time multimodel superensemble forecasts of Atlantic tropical systems of 1999. *Mon. Wea. Rev.*, *131*, 1878-1894.

Figure Captions

Fig. 1. The 48-h intensity forecast skill (%) for intensity forecast aids available to JTWC relative to a purely statistical model (ST5D). Dataset is from the JTWC 2006-2008 Southern Hemisphere seasons. Acronyms are defined in Table 1 and Table A1. Number of cases is shown in parentheses.

Fig. 2. The 48-h intensity forecast skill (%) of STIPS ensemble members with respect to an average of all members. Dataset is from the JTWC 2006-2008 Southern Hemisphere seasons. Acronyms are defined in Table A1. Number of cases is shown in parentheses.

Fig. 3. Intensity forecast skill relative to a purely statistical model (ST5D) for one ensemble aid and four multi-model ensemble (consensus) aids. Dataset is from the JTWC 2006-2008 Southern Hemisphere seasons. Acronyms are defined in Table A1. Number of cases is shown in parentheses.

Table Captions

Table 1. A list of tropical cyclone intensity forecast aids used in this study. The first column lists the name of the aid, the second column provides the name of the interpolated version of that forecast aid and the final column gives a description of the numerical or statistical model that is the basis for those forecast aids.

Table A1. STIPS ensemble members. The name of the individual ensemble member is given in the first column. The following columns describe the input data used in the STIPS model to create each of the ensemble members. Dynamic forecasts fields refer to the specific forecast model that provides the forecasts of the winds and other forecast fields refer to the model that provides the thermodynamic, moisture and SST fields.

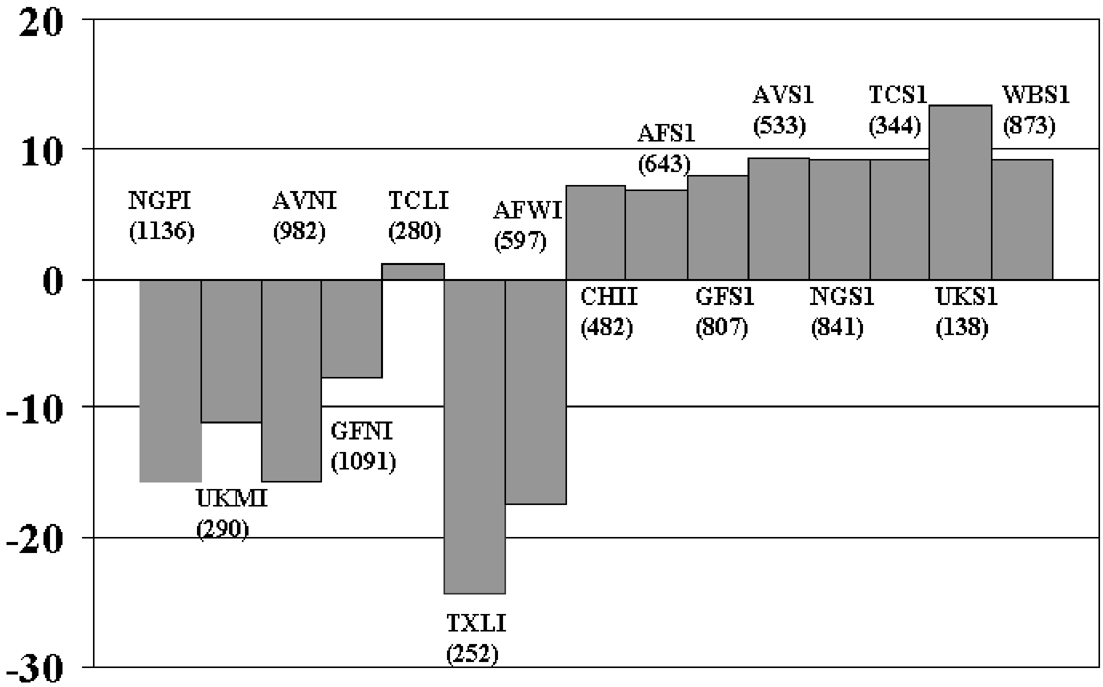

Fig. 1. The 48-h intensity forecast skill (%) for intensity forecast aids available to JTWC relative to a purely statistical model (ST5D). Dataset is from the JTWC 2006-2008 Southern Hemisphere seasons. Acronyms are defined in Table 1 and Table A1. Number of cases is shown in parentheses.

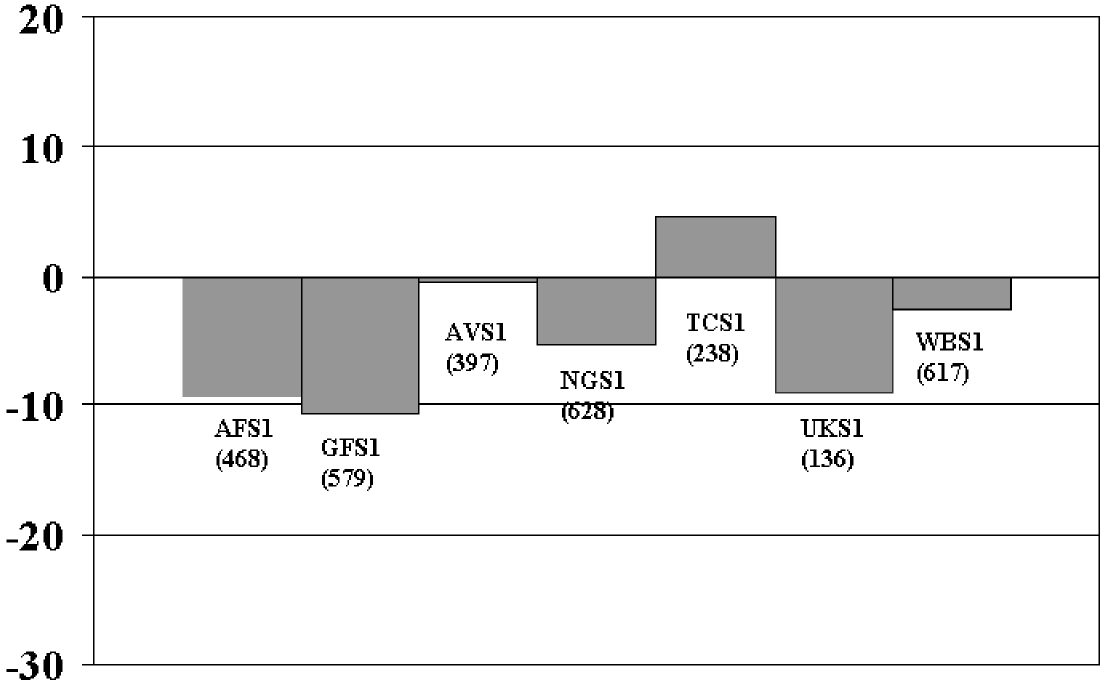

Fig. 2. The 48-h intensity forecast skill (%) of STIPS ensemble members with respect to ST10. Dataset is from the JTWC 2006-2008 Southern Hemisphere seasons. Acronyms are defined in Table A1. Number of cases is shown in parentheses.

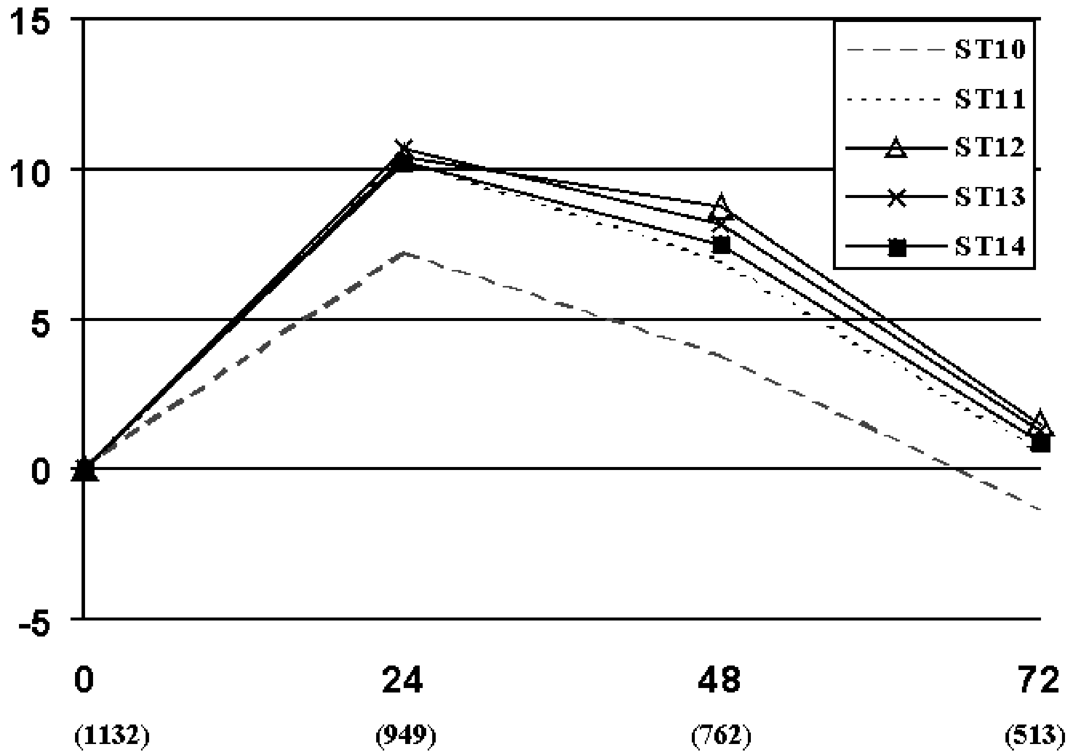

Fig. 3. Intensity forecast skill relative to a purely statistical model (ST5D) for one ensemble aid and four multi-model ensemble (consensus) aids. Dataset is from the JTWC 2006-2008 Southern Hemisphere seasons. Acronyms are defined in Table A1. Number of cases is shown in parentheses.

Table 1. A list of objective tropical cyclone intensity guidance techniques available at the Joint Typhoon Warning Center, its interpolated aid, a brief description, and the year of first availability.

Model	Interpolated	Description	Year first available
NOGAPS	NGPI	U.S. Navy global model (Hogan and Rosmond 1991)	2004
UKM	UKMI	UK global model (Heming et al. 1995)	2003
GFS	AVNI	NWS global model (Lord 1993)	2002
GFDN	GFNI	Geophysical Fluid Dynamic Lab initialized by the Navy Operational Global Analysis and Prediction System model (Rennick 1999)	1998
TC-LAPS	TCLI	Australian TC-Limited Area Prediction System (Davidson and Weber 2000)	2002
TX-LAPS	TXLI	Australian Tropical eXtended Area Prediction System (Australian Bureau of Meteorology 2005)	2005
US. Air Force regional model	AFWI	Air Force mesoscale model (Grell et al. 1995)	2002
ST5D	None	Statistical model (Knaff and Sampson 2008)	2004
STIPS	None	Statistical-dynamical model based on JTWC forecast (Knaff and Sampson 2008)	Not available
S1xx	None	STIPS ensemble members	2006
ST10	None	STIPS ensemble	2006
ST11	None	Multi-model consensus that combines the ensemble members of ST10 and GFNI	2007
ST12	None	ST10 members, GFNI and CHII	Not available
ST13	None	ST10 members, GFNI, CHII and TCLI	Not available
ST14	None	ST10 members, GFNI, CHII, TCLI and UKMI	Not available
CHIPS	CHII	Coupled dynamical hurricane model (Emanuel et al. 2004)	2003

Table A1. STIPS ensemble members. The name of the individual ensemble member is given in the first column. The following columns describe the input data used in the STIPS model to create each of the ensemble members. Dynamic forecasts fields refer to the specific forecast model that provides the forecasts of the winds and other forecast fields refer to the model that provides the thermodynamic, moisture and SST fields.

ST10 member	Track input	Dynamic Forecast Fields	Other Forecast Fields
AFS1	AFWI	NOGAPS	NOGAPS
AVS1	AVNI	GFS	NOGAPS
GFS1	GFNI	NOGAPS	NOGAPS
NGS1	NGPI	NOGAPS	NOGAPS
TCS1	TCLI	NOGAPS	NOGAPS
UKS1	UKMI	UKM	NOGAPS
WBS1	WBAI	NOGAPS	NOGAPS