Multimodel Superensemble Forecasting of Tropical Cyclones in the Pacific

T. S. V. VIJAYA KUMAR AND T. N. KRISHNAMURTI

Department of Meteorology, The Florida State University, Tallahassee, Florida

MICHAEL FIORINO

Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, California

MASASHI NAGATA

National Typhoon Center, Japan Meteorological Agency, Tokyo, Japan

31 January 2002 and 31 July 2002

ABSTRACT

Using currently available operational forecast datasets on the tracks and intensities of tropical cyclones over the Pacific Ocean for the years 1998, 1999, and 2000 a multimodel superensemble has been constructed following the earlier work of the authors on the Atlantic hurricanes. The models included here include forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF), the National Centers for Environmental Prediction/Environmental Modeling Center [NCEP/EMC, the Aviation (AVN) and Medium-Range Forecast (MRF) Models], the U.S. Navy [Naval Operational Global Atmospheric Prediction System, (NOGAPS)], the U.K. Met Office (UKMO), and the Japan Meteorological Agency (JMA). The superensemble methodology includes a collective bias estimation from a training phase in which a multiple-regression-based least squares minimization principle for the model forecasts with respect to the observed measures is employed. This is quite different from a simple bias correction, whereby a mean value is simply shifted. These bias estimates are described by separate weights at every 12 h during the forecasts for each of the member models. Superensemble forecasts for track and intensity are then constructed up to 144 h into the future using these weights. Some 100 past forecasts of tropical cyclone days are used to define the training phase for each basin. The findings presented herein show a marked improvement for the tracks and intensities of forecasts from the proposed multimodel superensemble as compared to the forecasts from member models and the ensemble mean. This note includes detailed statistics on the Pacific Ocean tropical cyclone forecasts for the years 1998, 1999, and 2000.

1. Introduction

This paper is a sequel to an upcoming paper on Atlantic hurricane forecast skills, Williford et al. (2002, manuscript submitted to *Mon. Wea. Rev.*, hereafter WMWR). In that study track and intensity forecasts of Atlantic hurricanes for the years 1998 and 1999 are presented. That study illustrates the use of multimodel forecasts of Atlantic hurricanes, where a weighted ensemble averaging of the results, called the multimodel superensemble, is shown to convey higher 3-day forecast skill for the position and intensity of Atlantic hurricanes. We shall describe this procedure briefly in this paper and demonstrate the application of this methodology for recent tropical cyclones over the Pacific.

Reduction in position and intensity errors of typhoons (or hurricanes) is a vital forecast issue. It turns out that some measurable improvements are possible from this superensemble methodology. This procedure examines the ‘‘collective biases’’ for a number of participating models and arrives at statistical weights from several past forecasts of these storms. These statistics, when applied to forecasts of future storms, appear to reduce the bias, position, and intensity errors. The current state of the multimodel datasets is far from ideal for performing such studies. In the context of a superensemble where past performance of the multimodels is rigorously examined during the training phase, it is important to have a consistent model behavior to systematically remove the bias. During the recent past, the models have been subjected to continuous change by different op-

---

Corresponding author address: T. S. V. Vijaya Kumar, Dept. of Meteorology, The Florida State University, Tallahassee, FL 32310-4520.
E-mail: vijay@met.fsu.edu

erational centers and that makes it a major constraint for studies related to superensemble. An ideal state would be to freeze the multimodels with respect to their physics and dynamics and obtain uniform forecast datasets so that the training phase of the superensemble will be more effective in the collective removal of the bias of these models. Nevertheless, we show in these experimental forecasts that this procedure can be very useful as an operational tool. To illustrate the power of this technique, we show an example from the studies of WMWR where the forecast track of Hurricane Lenny is compared for a few member models and the superensemble (Fig. 1). This figure illustrates the performance of the superensemble methodology applied to Atlantic hurricanes on a real-time basis. The motivation for this study was based on the success of the superensemble technique in the Atlantic region. In these track forecasts, we note improvements for the 1-, 2-, and 3-day positions of the order of 125, 200, and 350 km. This is an example showing rather large improvements in position forecasts. Such contributions were possible from the application of the superensemble approach, which was portrayed in a number of recent studies for weather and seasonal climate forecasts (Krishnamurti et al. 1999, 2000a,b, 2001; T. N. Krishnamurti et al., unpublished manuscript, hereafter K2002; Krishnamurti et al. 2003, hereafter KMWR).

In this paper, the superensemble-based track and intensity forecasts were constructed for the tropical cyclones over the Pacific Ocean for the years 1998–2000. The methodology for the superensemble forecasts follows a similar procedure that has been used for Atlantic hurricanes (WMWR). Forecasts from the currently available operational model datasets on the tracks and intensities of Pacific storms were included in this study. The datasets used here include track and intensity forecasts from the ECMWF, NCEP/EMC (AVN and MRF), the U.S. Navy (NOGAPS), UKMO, and JMA (GSM and TYM). A list of acronyms is given in Table 1.

2. Superensemble methodology for typhoon track and intensity prediction

In the superensemble forecasts the time line is partitioned into two phases. A training phase carries the datasets of past tropical cyclone forecasts from multimodels and the observed best measures of track and intensity for these (past) storms. During this training phase, model-forecasted position (latitude and longitude) and intensity are regressed against the best (observed) position (latitude and longitude) and intensity for each forecast time at intervals of every 12 h. A simple multiple linear regression technique is employed (Krishnamurti et al. 2000a) to generate weights (coefficients) for each model. These bias estimates are described by separate weights for forecasts up to 6 days for each of the member models at every 12-h interval. These coefficients are then passed on to a second phase called the forecast phase. During this phase future forecasts made by the multimodels and the aforementioned statistics are used to construct the superensemble forecasts. Some 100–150 past forecasts of Pacific storm days were used for the training phase for each of the years 1998, 1999, and 2000.

The performance of the superensemble largely depends on the consistency of the member models in their design and is also quite sensitive to changes in the performance of the member models (WMWR). Drastic model changes make a difficult proposition for the construction of superensemble forecasts. A safe way out of that, for this experimental demonstration study on Pacific Ocean typhoon/tropical cyclone forecasts, is to use roughly one-half of a season for the training phase of the superensemble. The premise here is that most model changes occur during the off-season. The statistics thus derived from the training phase were deployed for the forecast phase of the same season. That premise seems to be correct for the years 1998–2000, during which there have been major changes within several of the member models of this study. Thus, this procedure entailed using the early part of the storm forecasts for each of the years 1998, 1999, and 2000 and for each of the basins (EP, WP, SP; see Table 1 for definitions) separately for the definition of the training phases. It has to be noted here that this method's success depends on the consistency in the model forecasts throughout the training phase and also the forecast phase, and hence this method cannot be recommended yet for operational use.

The detailed statistics for the tropical cyclone track and intensity forecasts for 1998–2000 are presented in this paper. Here the errors from individual models, ensemble mean, and the superensemble are shown from the forecast phase of the superensemble. The evaluation of the ensemble mean is done strictly based on an average of its model forecasts and does not invoke any extrapolation. In those cases where the model start times differed by 12 h (i.e., 0000 and 1200 UTC), linear interpolation of forecasts to 1200 UTC was carried out for those models that had a 0000 UTC start time.

3. Data

In this study, the model forecast position (latitude and longitude) and intensity, along with best (observed) track and intensity information, are collected from different operational models listed in Table 2. A relatively large area comprising the east Pacific/northeast Pacific, west Pacific/northwest Pacific, and southwest Pacific/central South Pacific regions was chosen for this study. Table 2 also gives the details of the number of storms, models used, and the number of forecast days (time line) during the training and forecast phases of the superensemble, respectively. The Pacific Ocean storms are divided according to their origin into three parts—EP, WP, and SP—and the exercise is repeated independently over these basins for each of the years of this study.

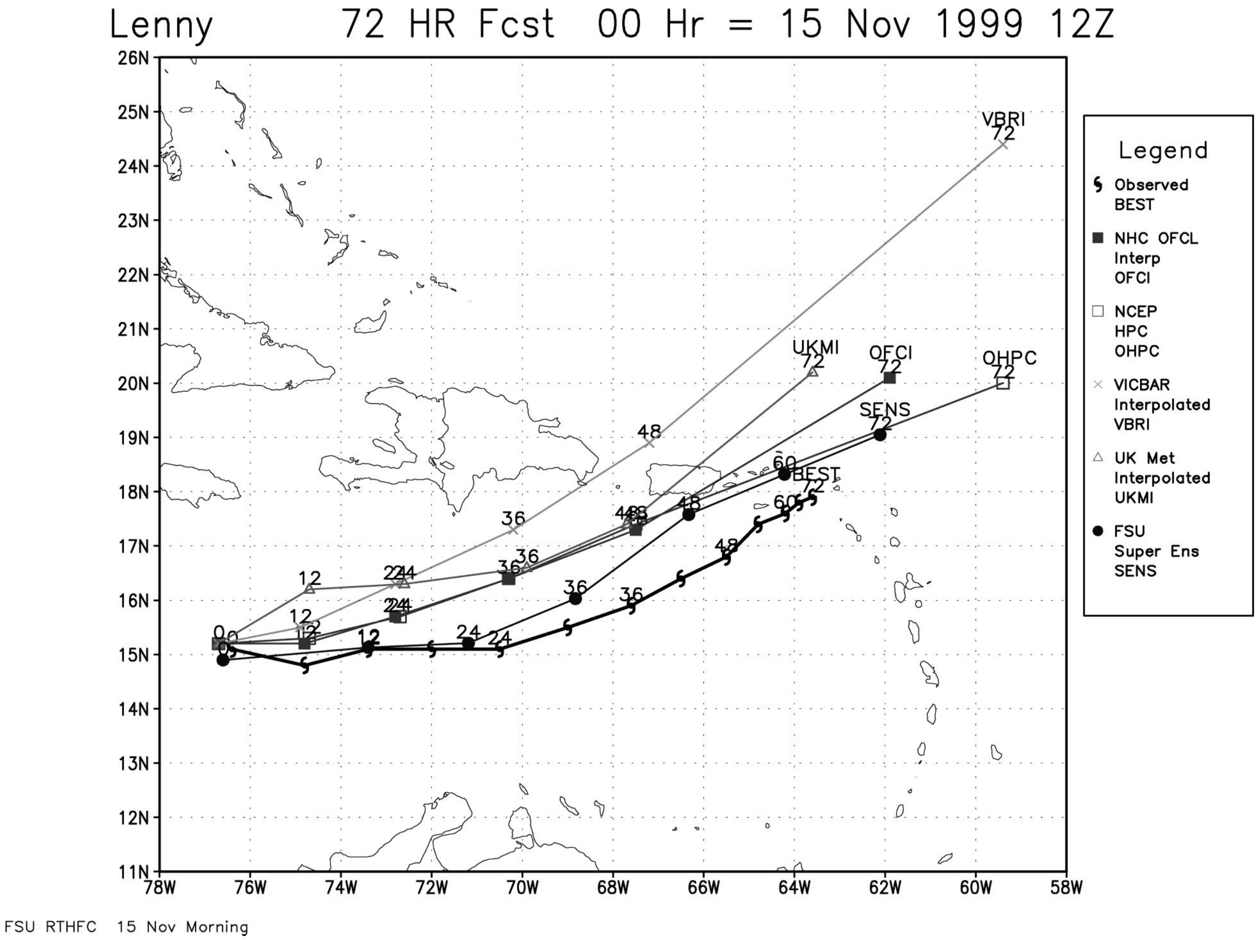

FIG. 1. Superensemble track forecast of Hurricane Lenny. Here the predicted tracks of some member models and superensemble are shown.

Tropical cyclone positions from model datasets

In the present study, the enhanced FNMOC tropical cyclone tracking scheme (M. Fiorino 2002, unpublished manuscript) was used to determine the positions of tropical cyclones from the model datasets. This scheme uses the surface or 10-m winds and the 850-mb relative vorticity maximum to locate centers of cyclonic wind shifts. The input model data are the surface winds (u, v components) on a global 1° latitude grid. The scheme sets a window of five to nine grid points square around the TC position and then two passes are made to locate cyclonic wind shifts. The search window is based on the motion of the storm and increases with increasing speed. The first pass uses a $1° \times 1°$ grid box with four points and the second pass uses a $3° \times 3°$ grid box. The idea is to first look for small-scale cyclonic circulations (one grid box or $\sim$1°), and then look for larger-scale cyclones (three-gridpoint box or $\sim$3°). If the four corners contain a wind vector in range, then a potential position is saved. After these two passes, between one and eight candidates can exist so that for a well-defined cyclone, the number of candidates indicates quality. The final position is the ''best'' center from an ''isogonic'' analysis of the wind fields around the candidate positions. These candidate positions are further picked by comparison of the maximum wind speed and nearest of the candidate and to the initial position. If multiple candidates remain, the final output is a best position based on three quality factors: (a) surface wind circulation confidence is 1–4 (where 1 is highest/best and 4 is lowest/poor), (b) cyclonic wind shift support factor is 3 or 4 points with shifts (4 is highest/best), and (c) intersection support is 2–8 (where 8 is highest/best).

The system starts with the observed TC position and motion at $t = 0$. If a model position is found at $t = 0$, then the observed motion is used to make an estimated position for the next forecast time. For application to global model output, a 12-h forecast time increment is used. The 12-h forecast position search is initialized with a position from a 12-h rhumb-line extrapolation from the initial model position and the observed heading and speed of motion. For the next forecast time (e.g., 24 h), the motion from the current (e.g., 12 h) and 12-h old position (e.g., 0 h) is used for the rhumb-line extrapolation.

TABLE 1. List of acronyms.

Acronym	Definition
AVN	Aviation Global Model of the NCEP
CLIPER	Climatology and persistence
ECMWF	European Centre for Medium-Range Weather Forecasts
EMC	Environmental Modeling Center
EOF	Empirical orthogonal function
EP	East Pacific
FNMOC	Fleet Numerical Meteorology and Oceanography Center
GSM	Global Spectral Model
JMA	Japan Meteorological Agency
JTWC	Joint Typhoon Warning Center
K2002	Krishnamurti et al. (2002, unpublished manuscript)
KMWR	Krishnamurti et al. (2002, manuscript submitted to *Mon. Wea.* Rev.)
MRF	Medium-Range Forecast model
NCEP	National Centers for Environmental Prediction
NCON	Numerical consensus forecast
NHC	National Hurricane Center
NOGAPS	Navy Operational Global Atmospheric Prediction System
NWP	Numerical weather prediction
SP	South Pacific
SVD	Singular value decomposition
TC	Tropical cyclone
TYM	Typhoon Model
UKMO	United Kingdom Met Office
WMWR	Walliford et al. (2002, manuscript submitted to *Mon. Wea. Rev.*)
WP	West Pacific

When the surface wind field scheme fails, an 850-mb relative vorticity maximum exceeding 5×10^{-5} s^{-1} within the search window (based on motion) is sought. The local vorticity maximum with the greatest vorticity nearest the initial (estimated from motion) position is selected.

4. Results and discussion

In this section we present a summary of the track and intensity errors for the years 1998, 1999, and 2000 for the forecast phase defined by Table 2. Here the position and intensity errors at 12-h intervals are shown for the member models, ensemble mean, and the superensemble for the entire Pacific Ocean region and for independent basins. The track errors (km) from the multimodels, ensemble mean and superensemble for the entire Pacific are shown in Figs. 2a–c for the years 1998, 1999, and 2000, respectively. Histograms showing the intensity errors (kt) for the years 1998, 1999, and 2000 from multimodels, the ensemble mean, and the superensemble over the entire region of study are presented in Figs. 3a–c. We note that the skill from the superensemble is consistently high compared to the member models and the ensemble mean.

Model performances from year to year were noted to be quite different in their respective storm forecasts. This is related to some major model changes that were reflected in the different ordering of these models. This aspect does seem to influence the performance of the superensemble over different years. Bogusing (or insertion of synthetic vortex) in many models appears to undergo different spinup features. In some cases even the beta gyres are degraded from the definition of the initial vortex. Models such as ECMWF appear to perform very well after 2–3 days. That is related to an improved definition of the storm environment at scales larger than the tropical cyclone where the better models and the superensemble appear to do better. These are some possible reasons for the subtle differences in the position and intensity errors from year to year and also from time to time.

It is apparent that if no major model changes are present during the training and the forecast phases of this experiment, we can obtain skill improvements on the order of 60, 140, 160, and 200 km for the tropical cyclone position errors over the best models (with smallest mean error) for forecasts at the end of days 1, 2, 3, and 4 respectively. The intensity forecast skills (rms errors) at days 1, 2, 3, and 4 of forecasts by the superensemble exceed those of the best models (with smallest mean error) by about 5, 10, 13, and 20 kt, respectively. The issue of intensity forecasts of the typhoons and hurricanes is a problematic area where further research is clearly warranted. It is however of interest here to note that the superensemble systematically pushes the skill toward a slight improvement over those of the best model and the ensemble mean. These are some of the important results from our study over the whole of the Pacific region.

Independent basin statistics have also been generated and the results for each of the basins confirm the superior performance of the superensemble methodology as revealed by its lesser track and intensity errors com-

TABLE 2. Data used in the study and the time line of the superensemble.

			Time line (no. of forecast days)	
Year	No. of storms	Models used	Training phase	Forecast phase
1998	47	ECMWF, MRF, NOGAPS, UKMO, JMA, GSM, and TYM	100 (23 storms)	113 (24 storms)
1999	56	ECMWF, MRF, NOGAPS, UKMO, JMA, GSM, and TYM	150 (30 storms)	114 (26 storms)
2000	55	ECMWF, MRF, NOGAPS, AVN, and UKMO	148 (30 storms)	116 (25 storms)

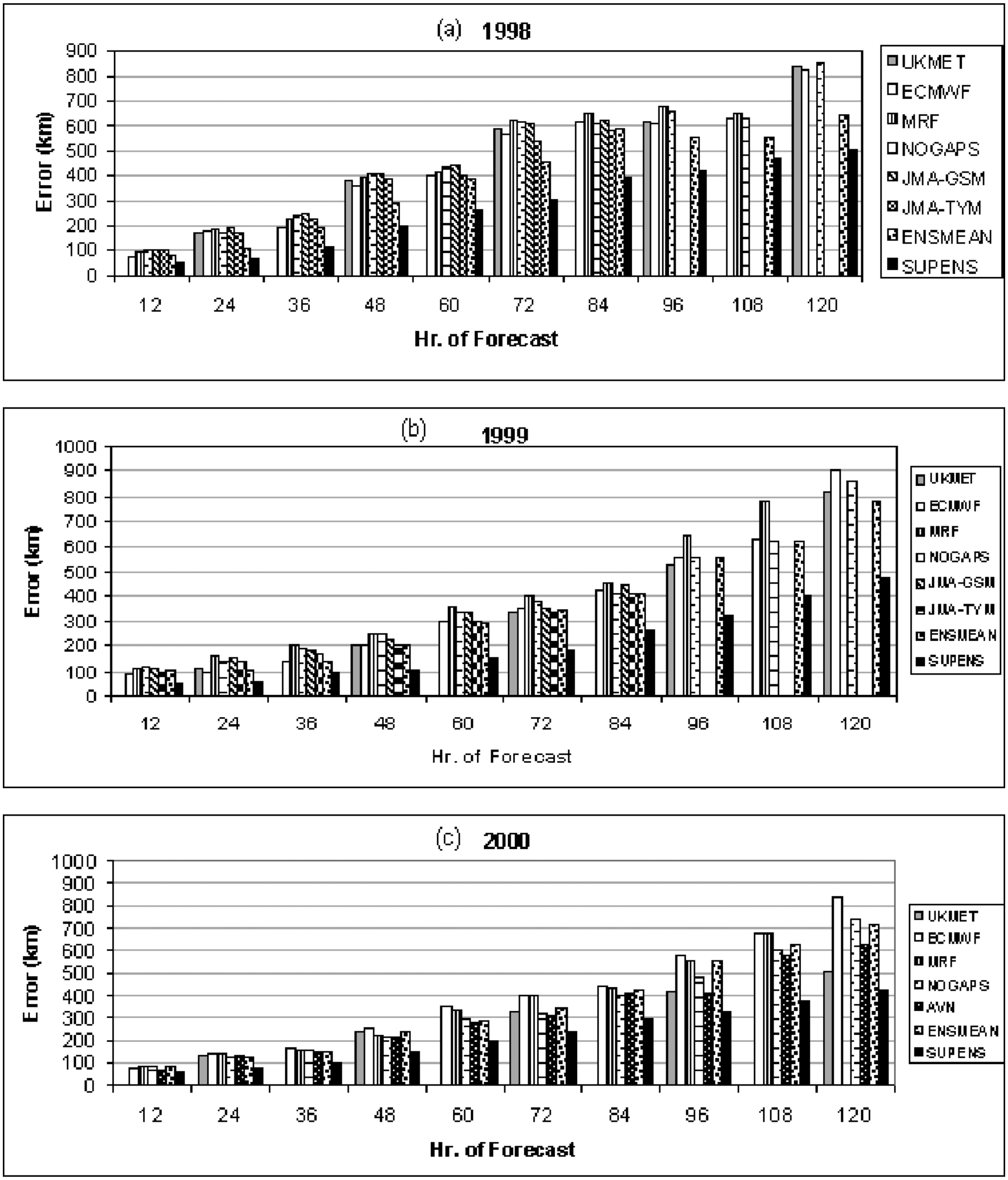

FIG. 2. Mean track errors (km) for tropical cyclones over the Pacific Ocean for (a) 1998, (b) 1999, and (c) 2000.

pared to member models and the ensemble mean. Results from the superensemble track and intensity forecasts for the west Pacific typhoons for the years 1998–2000 are presented in Figs. 4a and 4b. Here the superensemble is constructed using only the multimodel forecasts of typhoons over the west Pacific for each year separately. The results shown in this figure compose independent forecast phases of the superensemble for each year of study. In Fig. 4a the position errors at different times of the forecast for multimodels, the ensemble mean, and the superensemble are displayed. The intensity errors are shown in Fig. 4b. One should note

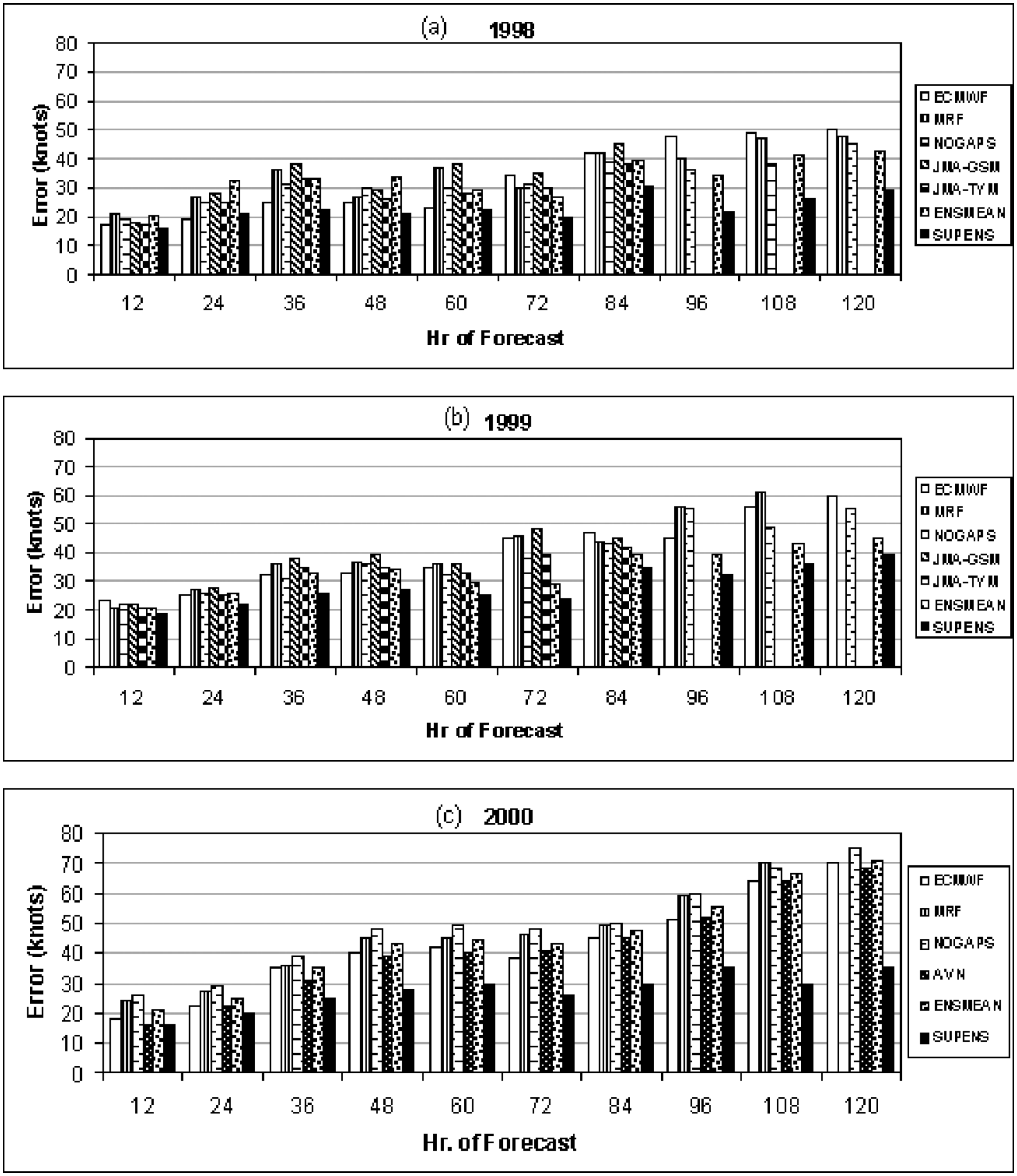

FIG. 3. Mean intensity errors (kt) for tropical cyclones over the Pacific Ocean for (a) 1998, (b) 1999, and (c) 2000.

that the results in Fig. 4 depict the average values for all three years of study, though the superensemble is constructed separately for each year using storms from an earlier part of each year in the training phase and the rest of the season in the forecast phase.

The results from tropical cyclone forecasts over the other basins, east Pacific and South Pacific (not shown here), are similar to the west Pacific and the total Pacific. The error trends were similar for member models and the superensemble outperforms them and the ensemble

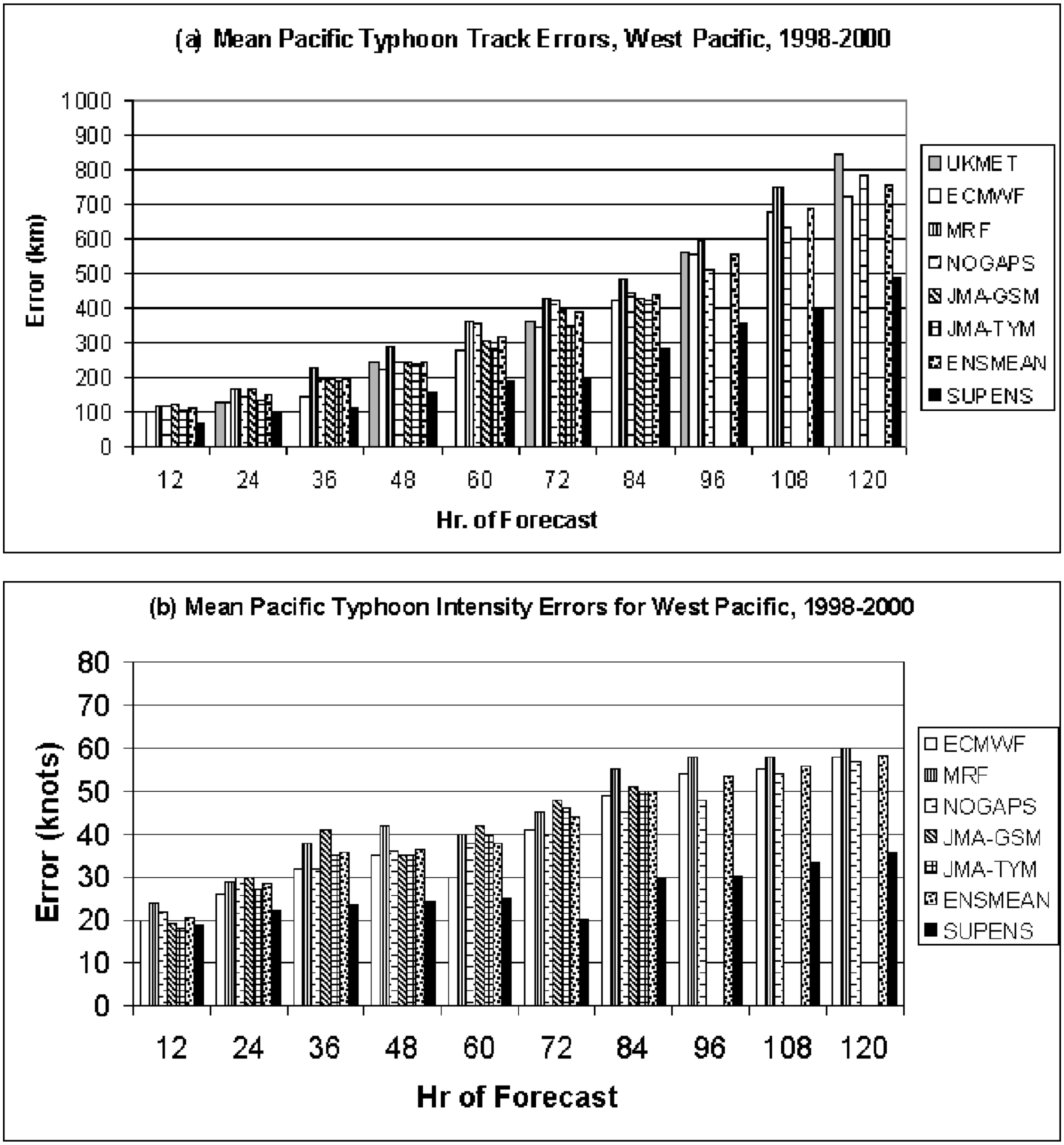

FIG. 4. (a) Mean typhoon track errors (km) for the west Pacific, 1998–2000. (b) Mean typhoon intensity errors (km) for the west Pacific, 1998–2000.

mean for all basins. The magnitudes of error for member models and for the superensemble were slightly higher in the South Pacific.

It is clear from these illustrations that there were no basic differences in the results presented here whether they are applied to the entire Pacific Ocean or to a portion thereof. The proposed superensemble is a well-tested procedure and it is extremely robust. The results shown in Figs. 2–5 are quite similar to results presented by WMWR in their tests with Atlantic hurricanes.

We shall next illustrate some examples of track forecasts from the superensemble. These correspond to 120-h forecasts for (a) Typhoon Faith (Fig. 5), which hit central Vietnam in the west Pacific on 14 December 1998, and (b) Typhoon Olga (Fig. 6), which hit the South Korean coast in the west Pacific on 3 August

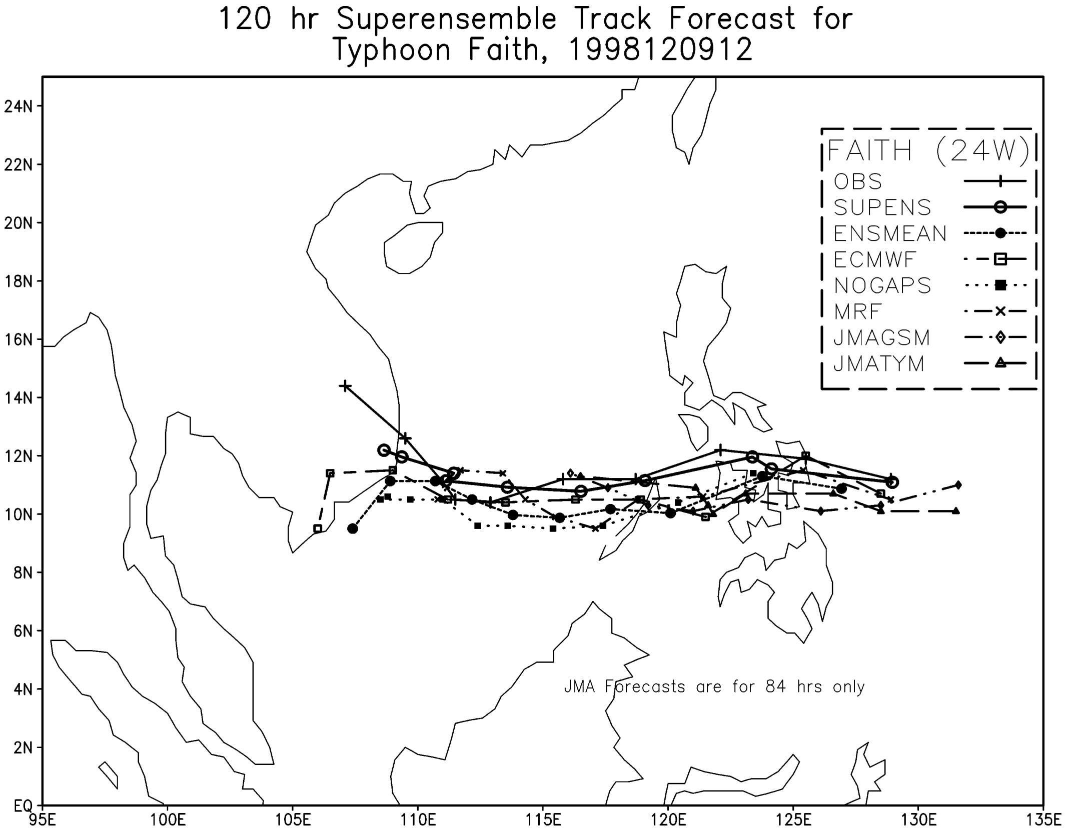

FIG. 5. The 108-h superensemble forecast for Typhoon Faith (24°W), initial conditions: 1200 UTC 12 Dec 1998.

1999. These figures reveal that the superensemble can provide more accurate track forecasts compared to member models and the ensemble mean. Such improvement in track errors, particularly after 96 h of forecast, is remarkable as these are large improvements over the NCON forecasts (McPherson and Zettlemoyer 2001) followed by the JTWC, as well as some other means of operational forecasts like the CLIPER method (used both by the NHC and JTWC), and can be viewed in the light of their contribution to the reduction of track and intensity errors.

5. Summary and future work

Three- to six-day multimodel forecasts for the Pacific Ocean tropical cyclone tracks and intensity during 1998, 1999, and 2000 are examined in this study at every 12-h interval. An annual summary for all tropical cyclones covering the entire Pacific Ocean, along with independent basin statistics for the east and west Pacific regions, is presented for the years 1998–2000. It appears that the superensemble approach for Pacific typhoon track and intensity forecasts shows somewhat improved skill. In particular, the superensemble approach is more useful beyond 72 h when track and intensity errors are much smaller relative to those of all member models and the ensemble mean. Clearly the message of this study is that track improvements of the order of 250 km are possible from the superensemble compared with those of participating member models for 3-day forecasts.

Upon careful examination, we found that the training period hardly affects the superensemble weighting: this applies to both early and late seasons. Experiments where late season forecasts were used in the training phase and earlier season storms were forecast using the weights obtained have shown similar error characteristics. Also in the present study the storm forecasts in the training phase were chosen randomly; they correspond to both early and late seasons in each basin, and also for each year. Though a cross-validation kind of approach (where all the forecasts are used in the training phase except one for which the superensemble forecast is made) gives still better results, for the purpose of testing the method's practical utility, only about half of the number of storm forecasts were used here in the training phase and the coefficients were obtained from all available member models. The optimum sample size of 60 storms for the training phase was arrived at after making several trial and error experiments and was also based on our experience with the Atlantic hurricanes (WMWR).

As an extension to this study, K2002 attempts to use the superensemble methodology for landfall forecasts using an objective method for identifying the position,

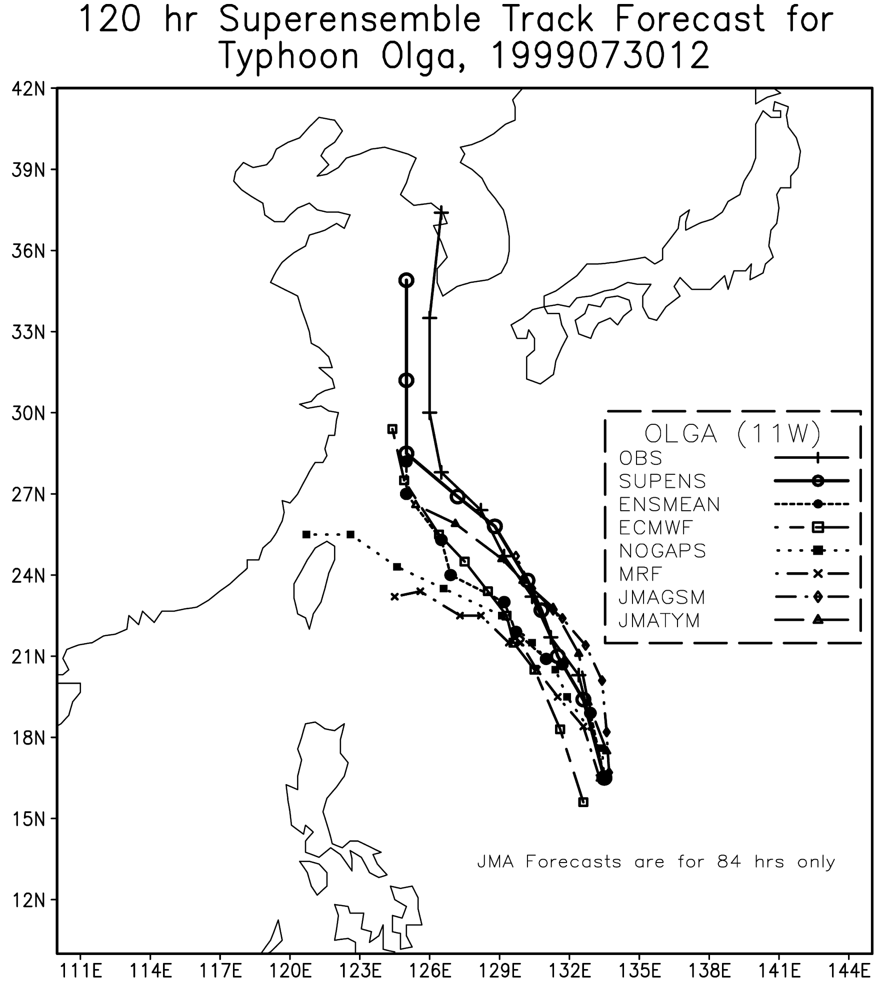

FIG. 6. The 108-h superensemble forecast for Typhoon Olga (11°W), initial conditions: 0000 UTC 30 Jul 1999.

time, and intensity at landfall. The preliminary results from that study indicated some 20%–30% improvement over the official forecasts in the Atlantic and about 20% improvement in the Pacific over the ensemble mean forecasts. The landfall positions are better estimated by the superensemble as seen through reduced errors in the position at the time of landfall (for instance, Typhoons Faith and Olga).

Goerss (2000) gives an excellent demonstration of how the skill of a simple consensus (i.e., uniform weights) depends on both the mean skill of the models used in the consensus and their correlation. If we define the 72-h mean track forecast error as the standard of long-range forecast skill, then our results (Fig. 2) show that the models are highly correlated because the consensus was only slightly better than the best individual model, particularly when the models showed high skill (e.g., 2000). Thus, our results depend on the intermodel correlation, which in turn is likely to be highly dependent on model formulation. This is the reason why the superensemble shows improvement over consensus when trained in the first half of the season and applied to the second half.

Further improvement in the skill may be possible from the superensemble methodology if we were to construct the superensemble on incremental position changes over 12 h rather than the positions themselves (WMWR). It is possible to further improve the skill of the superensemble methodology by eliminating forecast days of lower skill during the training phase. This method has also been pursued for global NWP and shown to have a positive impact on the forecast phase (KMWR). Defining the bias as a forecast mean minus observed mean, Krishnamurti et al. (KMWR) have examined the bias of the superensemble for the real-time NWP superensemble. In their studies they noted a marked reduction of the bias by the superensemble but it still had a nonzero value; thus, there is still some bias in the superensemble forecast. The superensemble model is a linear statistical model. The regression model error and errors due to nonlinearities (random errors) remain in the superensemble forecasts. The error variance of the superensemble forecast was smaller than those of the member model forecasts and this enhances the forecast confidence.

The current research on the construction of the superensemble is exploring several different arenas (Yun et al. 2002, manuscript submitted to *J. Climate*). This research has shown that a singular vector SVD method removes the ill-conditioning of the current Gauss–Jordan method of multiple linear regression and does yield a much superior overall performance for the superensemble. In that study, Yun and Krishnamurti also examined the relative merits of EOF-based methods, a Kalman filter, a Z-transformation method, and back-propagation neural network methods in different areas of research including NWP, seasonal climate, and hurricane/typhoon/tropical cyclone track and intensity forecasts.

The results from this study probably represent an upper bound to real-time tropical cyclone forecasting potential principally for two reasons: 1) training (regression coefficients) is not sufficiently persistent from season to season (model variability) to give consistent improvement over simple consensus and 2) there is variable availability of model tracks, in particular those from ECMWF. While we work to further understand the variability of the coefficients, we remain sanguine that superensembling can give an improvement over a simple consensus partially because of the low ''skill perishability'' of the long-range tracks from the models circa 2000. That is, the current generation of global model forecast tracks are consistent from forecast to forecast and extrapolated/interpolated tracks retain long-range skill (i.e., are not perishable) for 12–18 h (M. Fiorino 2002, personal communication).

This study appears to hold promise for possible use in real-time typhoon forecasts. The real-time implementation of this methodology does require that we reduce or eliminate altogether the need for long training periods from the early storms of the same year. These are issues that need to be addressed by the real-time forecast application groups who wish to pursue this approach.

Acknowledgments. The research reported here was funded by NASA Grants NAG5-9662 and NAG8-1537, NSF Grants ATM-9910526 and ATM-0108741, and FSURF COE. The authors wish to express their special thanks to the Japan Meteorological Agency, Tokyo, Japan, and the Lawrence Livermore National Laboratory, Livermore, California for providing the Pacific typhoon track and intensity datasets. Thanks to Dr. Julian Heming for providing the UKMO data used in this study. The authors are thankful for the useful comments from anonymous reviewers in improving this manuscript.

REFERENCES

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

Krishnamurti, T. N., C. M. Kishtawal, T. LaRow, D. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran, 1999: Improved skills for weather and seasonal climate forecasts from multimodel superensemble. *Science,* **285,** 1548–1550.

——, and Coauthors, 2000a: Multimodel superensemble forecasts for weather and seasonal climate. *J. Climate,* **13,** 4196–4216.

——, C. M. Kishtawal, D. W. Shin, and C. E. Williford, 2000b: Improving tropical precipitation forecasts from a multianalysis superensemble. *J. Climate,* **13,** 4217–4227.

——, and Coauthors, 2001: Real-time multianalysis–multimodel superensemble forecasts of precipitation using TRMM and SSM/I products. *Mon. Wea. Rev.,* **129,** 2861–2883.

——, K. Rajendran, T. S. V. Vijaya Kumar, S. Lord, Z. Toth, X. Zou, J. Ahlquist, and I. M. Navon, 2003: Improved skills for the anomaly correlation of geopotential heights at 500 hPa. *Mon. Wea. Rev.,* in press.

McPherson, T., and M. Zettlemoyer, 2001: Annual tropical cyclone report for the year 2000, 399 pp. [Available from U.S. Naval Pacific Meteorology and Oceanography Center/Joint Typhoon Warning Center, Pearl Harbor, HI 96860.]