Statistical ensemble prediction of the tropical cyclone activity over the western North Pacific

H. Joe Kwon,[1] Woo-Jeong Lee,[1] Seong-Hee Won,[1] and Eun-Jeong Cha[2]

Received 10 October 2007; revised 6 November 2007; accepted 19 November 2007; published 20 December 2007.

[1] This paper presents a statistical model to forecast seasonal tropical cyclone activity. In order to give a comprehensive view of seasonal tropical cyclone activity, we include not only the number of total tropical cyclones but also the number of typhoons and the NTA (Normalized Typhoon Activity) index as the predictands. The model is based on a multiple linear regression model in which the final predictors are selected with respect to minimizing the prediction error rather than simply fitting with past data. The model is expanded into ensemble prediction by considering the uncertainty of the single and deterministic forecast. The probability forecast based on the ensemble model shows reasonably good skill with respect to reliability and relative operating characteristics. **Citation:** Kwon, H. J., W.-J. Lee, S.-H. Won, and E.-J. Cha (2007), Statistical ensemble prediction of the tropical cyclone activity over the western North Pacific, *Geophys. Res. Lett.*, *34*, L24805, doi:10.1029/2007GL032308.

1. Introduction

[2] Even though the western North Pacific is the ocean basin where tropical cyclones (TC) are most active in the whole world, the seasonal prediction problem in this area is relatively unexplored. Two well-known typhoon centers in this basin, namely, the RSMC (Regional Specialized Meteorological Center) Tokyo and the JTWC (Joint Typhoon Warning Center) provide only track and intensity forecasts for an individual TC but do not issue the seasonal prediction.

[3] Since the pioneering work on seasonal prediction of TC activity by *Nicholls* [1979, 1984, 1985] over the two Australian basins, the problem of seasonal prediction has expanded into the North Atlantic [*Gray et al.*, 1992, 1993, 1994] and western North Pacific [*Chan et al.*, 1998, 2001]. These studies are based on the statistical relationships between the interannual variations of TC frequency and the atmospheric/oceanic signals such as El-Nino, Southern Oscillation, Quasi-Biennial Oscillation, *etc*. Although there are many reports on the strong relationship between ENSO (El Nino/Southern Oscillation) and tropical cyclone frequency over the North Atlantic [e.g., *Pielke and Landsea*, 1999, and references therein] leading to great success in the seasonal outlook of NOAA (National Oceanic and Atmospheric Administration), the relationship is not clear over the western North Pacific, at least for the number of TCs. For example, *Lander* [1994] reported that observed annual tropical cyclone totals in the western North Pacific are virtually uncorrelated with any ENSO indices. Nevertheless, the predictions issued by Chan's team since 1998 turn out to be quite successful by including many predefined monthly indices as the predictor, including the ENSO indices issued by Climate Prediction Center and by China Meteorological Administration.

[4] The other track in the seasonal prediction of TC activity is utilizing the global numerical model [*Vitart et al.*, 1997; *Vitart and Stockdale*, 2001; *Vitart*, 2005; *Camargo and Zebiak*, 2002; *Camargo and Sobel*, 2004]. By detecting and counting the model-generated hurricane-like vortices in the numerical model which runs several months ahead of the forecast base time and by scaling the systematic differences between the model climatology and observations, modelers extract the conclusion on the seasonal outlook of the tropical cyclone.

[5] In this study, a prototype statistical model presented by *Lee et al.* [2007] is expanded to an ensemble prediction by taking into account the uncertainty of the deterministic forecast by a single model and generating thirty ensemble members based on the original model. The ensemble prediction makes the categorical and probability forecasts possible. This statistical ensemble model is applied to many predictands such as the number of total TCs, number of typhoons, and a TC activity index that considers the duration and intensity of the TCs in order to give a comprehensive understanding of tropical cyclone activity over the western North Pacific, specifically the responsible area of RSMC Tokyo (100–180°E, 0–60°N).

2. Methodology

2.1. Predictands and the NTA Index

[6] In order to understand TC activity in a comprehensive manner, we need to include not only the number of total TCs but also the number of typhoons (TY) and a certain TC activity index that considers the duration and the intensity of the storms. In this study we present the NTA (Normalized Typhoon Activity) index as similar to the ACE (Accumulated Cyclone Energy),

$$\mathrm{NTA} = \sum \frac{1}{4}(V_{\max}/V_{TY})^2 \tag{1}$$

where $V_{\max}$ is the maximum wind and V_{TY} is TY wind speed, i.e., 64 kt. The normalized TC energy with respect to TY is again divided by four, given the fact that TC information is issued four times a day. This representation is easier to understand than ACE, which is simply the sum of

[1]Department of Atmospheric Sciences, Kongju National University, Kongju, Korea.
[2]Typhoon and Asian Dust Team, Korea Meteorological Administration, Seoul, Korea.

0094-8276/07/2007GL032308

the square of the maximum wind so that the unit becomes kt squared. We may compare two extreme examples (Tracks are not shown for brevity). While the lifespan of the eighth TC in 1988 was very short and the intensity was weak, that of the thirteenth TC in 1987, Freda, was very long, its intensity was very strong, and the maximum wind during its lifespan was recorded as 107 kt from 18 UTC September 9 to 12 UTC the next day. At that time the RSMC did not include the wind data in the TC advisory so we converted the central pressure into wind with the use of a wind-pressure conversion table. In our current representation, the NTA of the former and the latter TCs is 0.3 and 17.5, respectively. This means that the former TC may be considered to have lasted 0.3 days as a 64 kt TC and the latter TC 17.5 days. The correlation between the annual number of total TCs and the NTA index for the period from 1951 to 2006 is only 0.496, which means that there are wide discrepancies between the interannual fluctuations of the two variables (figure not shown for brevity). This also makes very clear the need to define the NTA index in order to have a comprehensive understanding of TC activity. For these reasons, we need to include not only the number of total TCs and the TCs that exceed the TY strength as the predictands, but also the NTA index in a seasonal prediction. In addition, we predict the number of TCs that will affect Korea, requiring the KMA's official seasonal forecast.

[7] For the predictand data such as the number of TCs and the NTA index we have used the best track from the RSMC (Regional Specialized Meteorological Center) Tokyo and the monthly data from NCEP/CPC (National Center for Environmental Protection/Climate Prediction Center) are used in obtaining the predictors. Data prior to 1970 are not used in the analysis since the TC intensity analyses were not trustworthy at that time [*Kwon et al.*, 2006].

2.2. Potential Predictors, Predictor Candidates, and Smart Predictors

[8] The prototype model for this statistical prediction has been presented by *Lee et al.* [2007]. The model consists mainly of three parts, i.e., the potential predictors, the predictor candidates and the smart predictors. Only a brief description of the model is given in this paper. Instead of using the predefined large-scale monthly indices such as the Western Pacific Index, ENSO-related indices, the Arctic index, *etc*., released from the climate centers, we have searched for the predictors independently. First, by examining the lag correlation patterns between the predictands and the synoptic variables such as the sea surface temperature, the mean sea level pressure, 850 hPa temperature, meridional wind and 500 hPa geopotential, we choose the geographical points which show high correlations. The magnitude of the correlations is very different depending upon the combination of the predictands and the synoptic variables as well as the location. It is found that there are several geographic locations that show a significantly high correlation. For example, it is found that the correlation between the March mean sea level pressure at 87.5°W, 5.0°N and the TC frequency from May to December is as high as 0.619. In this way, we collect as many variables as possible that show a correlation coefficient higher than 0.4. We refer to these as the *potential predictors*. The numbers of the total potential predictors vary from approximately thirty to fifty depending upon the combination of the predictands and the synoptic variables. The correlation coefficients of all potential predictors are more than 95% statistically significant by the student-t test. Readers may read *Lee et al.* [2007] for further details (http://www.apjas.org).

[9] The simplest way of constructing the model is to include all potential predictors in the regression equation, which should guarantee the best fit of past data but does not necessarily mean the best prediction. We take a different strategy. We try to find the predictor set among all potential predictors that has produced the best prediction in the past. In other words, we find a set of predictors that minimizes the PRESS (PRediction Error in Sum of Squares),

$$\mathrm{PRESS} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(f_t - y_t)^2} \tag{2}$$

where n is the number of the prediction, f_t and y_t are predicted and observed values, respectively. We call them the *smart predictors*. In order to find the smart predictors, we take the following steps. We construct all possible regression equations that fit the past-year data over a certain period - normally 30 years, apply the equations to the prediction for the next year, and compare the predicted values (f_t) with the observed values (y_t). This procedure is repeated with the one-year shift of the model construction and verification until the most recent year. The model that gives the minimum PRESS for the validation period is finally selected and the predictors of the model are called the smart predictors. Since the model that consists of the smart predictors proves to be smart for the earlier predictions, that smartness will presumably continue in real predictions in the future.

[10] What 'possible' in the previous paragraph means is the inclusion of every combination out of the given number of potential predictors. For example, if the number of the potential predictors is 30, there are 2^{30}-1 combinations out of the thirty potential predictors, so that we need to construct 2^{30}-1 regression equations, which are too large to handle. To reduce the regression equations to a feasible number, we set a limit of three predictors per each of the five synoptic variables, which reduces the total number of predictors to 15. We call them the *predictor candidates*. We have chosen them among all potential predictors in such a way that the lag month is closest to the current month of prediction per each synoptic variable. Now that we have reduced the predictors to a manageable number, we examine 2^{15}-1 regression equations and check which regression equation (model) performed the best prediction based on past data in terms of minimizing the PRESS.

2.3. Ensemble Prediction

[11] When the model described above was applied to the seasonal prediction of TC frequency over several specific periods from May to December, from June to August, from September to November depending upon the user's need, it is found that the model does show some skill in the prediction of TC frequency [*Lee et al.*, 2007]. These

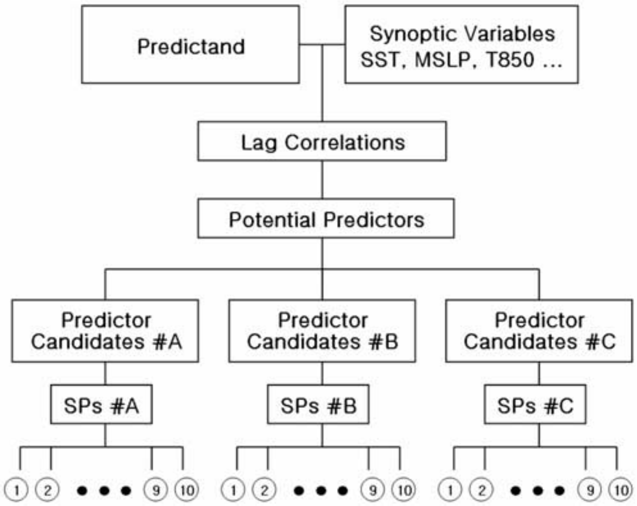

Figure 1. Flow chart of constructing the model and producing 30 ensemble members.

specific time intervals are chosen in accordance with the Korea Meteorological Administration's (KMA) forecast schedule. KMA releases the official three-month seasonal forecasts for tropical cyclones at the end of May and August so that the guidance should be delivered in the early part of the corresponding months.

[12] In order to alleviate a certain danger in a single deterministic forecast, we adopt the probability and category forecast based on the ensemble prediction. Thirty ensemble members are generated as follows (Figure 1). After the 30–50 potential predictors are selected, three different sets of the predictor candidates are carefully chosen. The first choice is the selection of the predictors that show the highest correlation coefficients regardless of the lag months. Second is the selection of the predictors in such a way that the lag month becomes close to the current month of the predictand. Another way may be a subjective choice. The first two choices are for the high correlation and for the close lag month. This procedure is straightforward. But when we examine such chosen predictors, some predictors are too far from the predictand in terms of the geographical location and the lag month. Such predictors are replaced by other predictors which have opposite characteristics, whereby one's intuition may be somewhat involved.

[13] As stated above, the smart predictors are selected so that the regression equation based on those predictors yields the best performance in the past prediction. The performance of this past prediction may vary depending upon the verification periods from the most recent years to past several years. Thus, we have made ten ensemble members per each of the three sets of the predictor candidates by applying to the ten verification periods. Specifically the verification periods are from the most recent two years to the past eleven years, which would produce ten ensemble members per each of the three predictor candidates.

3. An Example of Categorical Probabilistic Prediction

[14] Figure 2 shows an example of the ensemble prediction. Among many predictands, the result for only the number of total TCs during the three fall months (Sep.–Nov.) is shown. The grey bars represent the observations and the white bars represent the predictions whose frequencies are shown on the left and right axes. The TC frequencies observed during 50 years (1951–2000) constitute

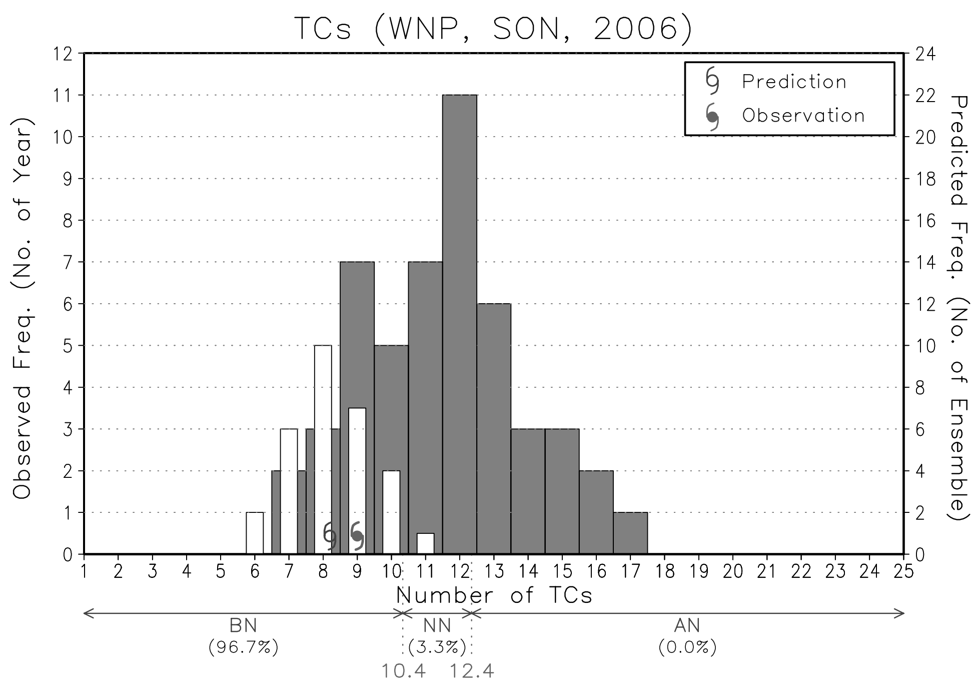

Figure 2. Ensemble prediction for the number of TCs during the three fall months (SON). The grey bars represent the observations and the white bars represent the predictions whose frequencies are shown on the left and the right axes. The range of the three terciles based on the 50-year climatology is indicated at the bottom.

Table 1. Summary of the 2006 TC Activity Predictions

	PRED (ENS MEAN)	OBS	ERROR
MAY–DEC			
TC	26.9	23	**3.9**
TY	14.3	14	0.3
NTA	146.6	148.5	−1.9
JUN–AUG			
TC	13.5	11	**2.5**
TY	5.7	5	0.7
NTA	32.3	55.7	**−23.4**
TC (Korea)	3.0	2	**1.0**
SEP–NOV			
TC	8.2	9	−0.8
TY	6.3	7	−0.7
NTA	62.3	70.8	−8.5
TC (Korea)	1.2	1	0.2

background climatology for defining the upper (AN: Above Normal), middle (NN: Near Normal) and lower (BN: Below Normal) terciles. The normal range of this case is 10.4–12.4, as indicated in the lowest part of Figure 2.

[15] The ensemble prediction states that among all 30 members, two indicate that six TCs will appear during SON, six say seven TCs, ten say eight TCs, seven say nine TCs, four say 10 TCs, and one says 11 TCs. From a probabilistic viewpoint, 29 members out of a total of 30 members (96.7 percent) indicate that the number of TCs that would be generated during the fall of 2006 would be less than or equal to 10, which is in the below normal range. Meanwhile, only 1 member (3.3 percent) predicts a number in the near normal range, and none (zero percent) predicted an above normal number. The average number of TCs predicted by all 30 members, i.e., the ensemble mean value, is 8.2, which is denoted by the open cyclone symbol. The observed TC frequency during this period was nine (closed TC symbol), which is very close to the predicted ensemble mean value and falls within the tercile of the largest categorical probability forecast. All predictions are not as good as the example shown above. The next section describes this statistical ensemble prediction's overall performance for all the forecast cases.

4. Verification

[16] Ensemble predictions for all predictands for the year 2006 have been carried out. Table 1 shows the performance of the 11 ensemble predictions in terms of the difference between the ensemble mean and the observed value. Most forecasts show reasonably good results, except a few cases whose error is close to or exceeds one standard deviation, noted with bold letters.

[17] In verifying the ensemble prediction's performance, we have checked the reliability [*Atger*, 1999] to ensure that the categorical probability forecast is reliable. The reliability is checked for all predictands by comparing the forecast probability category coinciding with the corresponding observed relative frequency for all terciles (Figure 3, left). The grey bars show the frequency with which each probability is actually forecast. In constructing the reliability diagram, the forecast and observed probability are counted for every 10 percent for all events to which this ensemble prediction is applied. The reliability line approximately follows the diagonal – perfect reliability, which means that this probability forecast is indeed reliable. Some over-forecasting tendencies appear over the range of high probability, and vise versa for the lower side. The bias calculated with $^{1}/_{n} \sum N_i(y_i - \overline{o_i})^2$ turns out to be 0.006, which shows slight but negligible over-forecasting, where n is the total number of events, N_i is the number of events of each forecast probability, y_i is the forecast probability category

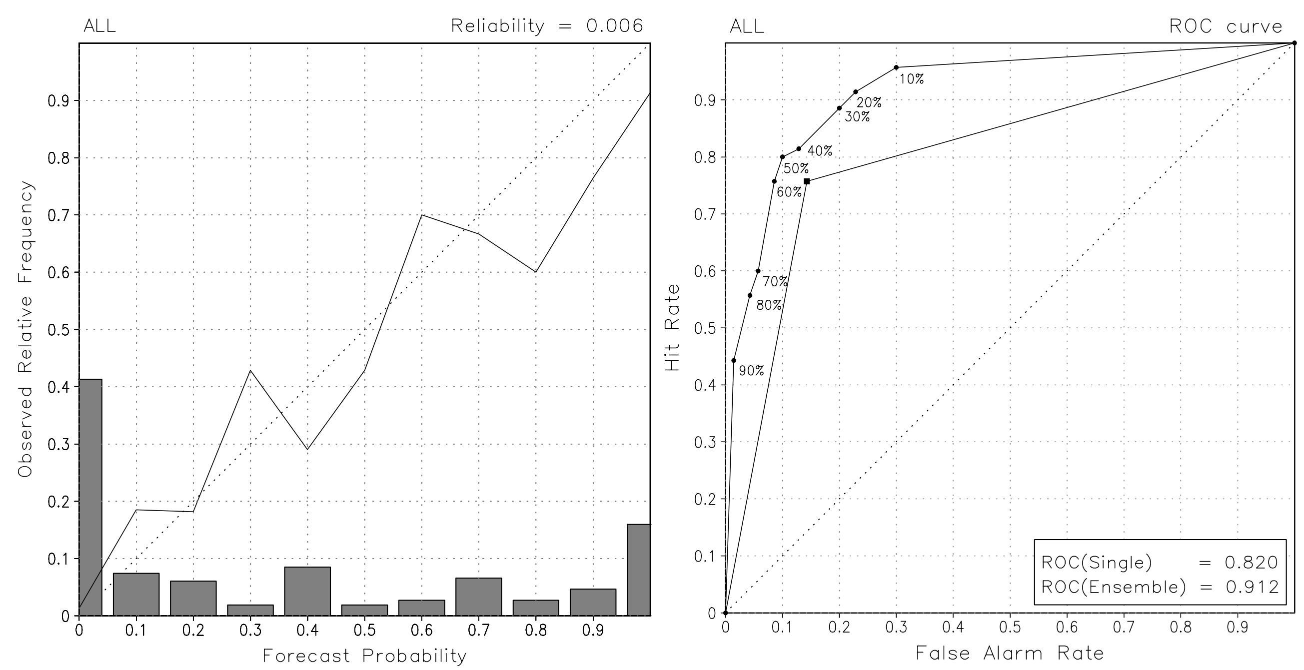

Figure 3. (left) Reliability and (right) ROC of the ensemble prediction. The grey bars in Figure 3 (left) show the relative frequency with which each probability is actually forecast. The HR/FAR line of the single forecast is also shown in Figure 3 (right).

and $\overline{o_i}$ is the mean observed events per each forecast probability category.

[18] Relative operating characteristics (ROC) is one of the standard verification tools used to measure the performance of the ensemble prediction based on the signal detection theory [*Swets*, 1973]. ROC compares the hit rate (HR) and the false alarm rate (FAR) for every forecast probability. It is independent of reliability because the results are insensitive to forecast bias [*Richardson*, 2000]. We obtain the ROC curve (Figure 3, right) by calculating the HR and the FAR for every 10 forecast probability categories with the use of the same verification data as the reliability diagram. The forecast becomes perfect when all the points on the curve concentrate toward the upper left corner. The diagonal means no skill, where the hit rate is equal to the false alarm rate. Figure 3 (right) shows the ROC curve for all forecast probabilities, terciles and predictands, as in Figure 3 (left). The ROC curve shows that HR is larger than the FAR for all cases. The area under the ROC curve is 0.912, which is a high forecast performance, and is an even higher value than *Lee et al.*'s [2007] single forecast (0.820).

5. Summary

[19] This paper presents a probability forecast of seasonal TC activity based on the ensemble statistical model. In order to give a comprehensive view of seasonal TC activity, we include not only the total number of TCs, but also the number of typhoons and the NTA index as predictands.

[20] The model is based on the multiple linear regression model. The model's uniqueness is that: (1) we have chosen the predictors in order to maximize the predictability, rather than using the pre-established monthly climate indices issued by climate centers; (2) the final smart predictors are carefully selected in terms of minimizing the PRESS; and (3) 30 ensemble members are generated to yield the probability forecast. Probability is obtained for each tercile (Above Normal, Near Normal, Below Normal) by counting the number of predictions that fall within the range of each tercile.

[21] We have checked the prediction performance not only by comparing the ensemble mean to the observed value, but also by checking the forecast probabilities on each tercile with the observed frequencies. The most comprehensive view of the model's performance can be seen in the reliability and ROC diagram (Figure 3), which utilize the same forecast results in Table 1. The probability forecast shows reasonably good skill with respect to reliability and relative operating characteristics. In large part, the reliability curve is aligned along the diagonal, which proves it to be reliable. In the ROC curve, the hit rate is much larger than the false alarm rate in most of the forecasts. Because we use data prior to 2006 to build the model to maximize the prediction skill and the prediction is made only for 2006, the model performance shown in Figure 3 is based only on 1 year. Much more years would be needed for the true assessment of the skill of this new model.

[22] **Acknowledgments.** This work was funded by the Korea Meteorological Administration Research and Development Program under Grant CATER 2007–2310. We deeply appreciate the valuable comments by the two anonymous reviewers for improving the paper.

References

Atger, F. (1999), The skill of ensemble prediction systems, *Mon. Weather Rev.*, *127*, 1941–1953.
Camargo, S. J., and A. H. Sobel (2004), Formation of tropical storms in an atmospheric general circulation model, *Tellus, Ser. A*, *56*, 56–67.
Camargo, S. J., and S. E. Zebiak (2002), Improving the detection and tracking of tropical cyclones in atmospheric general circulation models, *Weather Forecasting*, *17*, 1152–1162.
Chan, J. C. L., J. E. Shi, and C. M. Lam (1998), Seasonal forecasting of tropical cyclone activity over the western North Pacific and the South China Sea, *Weather Forecasting*, *13*, 997–1004.
Chan, J. C. L., J. E. Shi, and K. S. Liu (2001), Improvements in the seasonal forecasting of tropical cyclone activity over the western North Pacific, *Weather Forecasting*, *16*, 491–498.
Gray, W. M., C. W. Landsea, P. W. Mielke Jr., and K. J. Berry (1992), Predicting Atlantic seasonal hurricane activity 6–11 months in advance, *Weather Forecasting*, *7*, 440–455.
Gray, W. M., C. W. Landsea, P. W. Mielke Jr., and K. J. Berry (1993), Predicting Atlantic basin seasonal tropical cyclone activity by 1 August, *Weather Forecasting*, *8*, 73–86.
Gray, W. M., C. W. Landsea, P. W. Mielke Jr., and K. J. Berry (1994), Predicting Atlantic basin seasonal tropical cyclone activity by 1 June, *Weather Forecasting*, *9*, 103–115.
Kwon, H. J., S.-H. Won, and S. K. Park (2006), Climatological differences between the two typhoon centers' tropical cyclone information in the western North Pacific, *J. Korean Meteorol. Soc.*, *42*, 183–192.
Lander, M. A. (1994), An exploratory analysis of the relationship between tropical storm formation in the western North Pacific and ENSO, *Mon. Weather Rev.*, *122*, 636–651.
Lee, W.-J., J.-S. Park, and H. J. Kwon (2007), A statistical model for prediction of the tropical cyclone activity over the western North Pacific, *J. Korean Meteorol. Soc.*, *43*, 175–183.
Nicholls, N. (1979), A possible method for predicting seasonal tropical cyclone activity in the Australian region, *Mon. Weather Rev.*, *107*, 1221–1224.
Nicholls, N. (1984), The southern oscillations, sea-surface temperature, and interannual fluctuations in Australian tropical cyclone activity, *J. Climatol.*, *4*, 661–670.
Nicholls, N. (1985), Predictability of interannual variations of Australian seasonal tropical cyclone activity, *Mon. Weather. Rev.*, *113*, 1144–1149.
Pielke, R. A., Jr., and C. N. Landsea (1999), La Niña, El Niño, and Atlantic hurricane damages in the United States, *Bull. Am. Meteorol. Soc.*, *80*, 2027–2033.
Richardson, D. S. (2000), Skill and economic value of the ECMWF ensemble prediction system, *Q. J. R. Meteorol. Soc.*, *126*, 649–668.
Swets, J. A. (1973), The relative operating characteristic in psychology, *Science*, *182*, 990–999.
Vitart, F. (2005), Seasonal forecasting of tropical storm frequency using a multi-model ensemble, *Q. J. R. Meteorol. Soc.*, *132*, 647–666.
Vitart, F., and T. N. Stockdale (2001), Seasonal forecasting of tropical storms using coupled GCM integration, *Mon. Weather Rev.*, *129*, 2521–2537.
Vitart, F., J. L. Anderson, and W. F. Stern (1997), Simulation of interannual variability of tropical storm frequency in an ensemble of GCM integrations, *J. Clim.*, *10*, 745–760.

E.-J. Cha, Typhoon and Asian Dust Team, Korea Meteorological Administration, 45 Gisangcheong-gil, Dongjak-gu, Seoul 156-720, Korea.
H. J. Kwon, W.-J. Lee, and S.-H. Won, Department of Atmospheric Sciences, Kongju National University, Kongju, Chungnam 314-701, Korea. (hjkwon@kongju.ac.kr)