

Climate Prediction of Tropical Cyclone Activity in the Vicinity of Taiwan Using the Multivariate Least Absolute Deviation Regression Method

Pao-Shin Chu^{1, *}, Xin Zhao², Ching-Teng Lee³, and Mong-Ming Lu³

(Manuscript received 21 September 2006, in final form 16 April 2007)

ABSTRACT

In this study, a multivariate linear regression model is applied to predict the seasonal tropical cyclone (TC) count in the vicinity of Taiwan using large-scale climate variables available from the preceding May. Here the season encompasses the five-month period from June through October, when typhoons are most active in the study domain. The model is based on the least absolute deviation so that regression estimates are more resistant (i.e., not unduly influenced by outliers) than those derived from the ordinary least square method. Through lagged correlation analysis, five parameters (sea surface temperature, sea level pressure, precipitable water, low-level relative vorticity, and vertical wind shear) in key locations of the tropical western North Pacific are identified as predictor datasets. Results from cross-validation suggest that the statistical model is skillful in predicting TC activity, with a correlation coefficient of 0.63 for 1970 - 2003. If more recent data are included, the correlation coefficient reaches 0.69 for 1970 - 2006. Relative importance of each predictor variable is evaluated. For predicting higher than normal seasonal TC activity, warmer sea surface temperatures, a moist troposphere, and the presence of a low-level cyclonic circulation

¹ Department of Meteorology, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, Honolulu, Hawaii, USA

² Department of Information and Computer Science, University of Hawaii at Manoa, Honolulu, Hawaii, USA

³ Central Weather Bureau, Taipei, Taiwan, ROC

* *Corresponding author address:* Prof. Pao-Shin Chu, Department of Meteorology, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, Honolulu, Hawaii, USA;
E-mail: chu@hawaii.edu

doi: 10.3319/TAO.2007.18.4.805(A)

coupled with low-latitude westerlies in the Philippine Sea in the antecedent May appear to be important.

(Key words: LAD regression, Seasonal tropical cyclone forecasting, Taiwan, Cross-validation prediction skills, Climate forcings)

1. INTRODUCTION

Taiwan is consistently ravaged by typhoons causing enormous property damage and loss of life. Because of their catastrophic socio-economic consequences, there is an interest in understanding climate controls that are instrumental in year-to-year typhoon variability and developing a basis for seasonal prediction of typhoon activity. Predictive information, if properly made, could be important for decision makers in relevant agencies to mitigate effects of potential flooding, storm surges, and high winds resulting from typhoons.

W. Gray and his associates pioneered the seasonal hurricane prediction enterprise using a regression-based statistical model called the least absolute deviation method (Gray et al. 1992). They showed that nearly half of the interannual variability of hurricane activity in the North Atlantic could be predicted in advance. This is amazing because a hurricane is a small system, and physical mechanisms governing its formation are complicated. Along the same line of statistical modeling, Chan et al. (1998) developed a model to predict typhoon activity over the western North Pacific and the South China Sea. When tested against a 30-yr sample through the jackknife method, skillful forecasts are noted for a suite of predictands (e.g., the annual number of typhoons and the annual number of tropical storms and typhoons). The informative results of Chan et al. (1998) are pertinent to the vast western North Pacific basin and the South China Sea. For a limited geographic domain such as the vicinity of Taiwan, the frequency of typhoon occurrences may be very different from that of basin-wide numbers since factors influencing both TC formation and tracks are vital when the study domain is regional.

In this study, we attempt to predict the seasonal number of tropical cyclones (TCs) in the vicinity of Taiwan on the basis of the Least Absolute Deviation (LAD) regression method. This method has been tested for many years by Gray and his associates and is quite mature. Section 2 discusses the dataset. Section 3 describes TC climatology. Section 4 outlines the LAD model. In section 5, procedures for selecting appropriate predictor variables are described. Section 6 discusses the forecast results, and section 7 evaluates the relative importance of each predictor and diagnoses circulation features for composites of extreme events. A summary is found in section 8.

2. DATA AND DATA PROCESSING

The tropical cyclone series in the vicinity of Taiwan from 1970 to 2003 compiled by the Central Weather Bureau (courtesy of Guay-Hong Chen) is used. This series cover an area between 21 - 26°N and 119 - 125°E, a fairly small geographical domain of five degrees latitude and six degrees longitude. In this study, TC refers to tropical storms and typhoons.

Monthly mean sea level pressure, wind data at 850- and 200-hPa levels, relative vorticity data at the 850 hPa level, and total precipitable water over the western North Pacific (0° - 30° N) are derived from the NCEP/NCAR reanalysis dataset (Kalnay et al. 1996; Kistler et al. 2001). The horizontal resolution of the reanalysis dataset is 2.5° latitude-longitude. Tropospheric vertical wind shear is computed as the square root of the sum of the square of the difference in zonal wind component between 850- and 200-hPa levels and the square of the difference in meridional wind component between 850- and 200-hPa levels (Clark and Chu 2002). The monthly mean sea surface temperatures, at 2° horizontal resolution, are taken from the NOAA Climate Diagnostic Center in Boulder, Colorado. As our purpose is to develop a prototype model, large-scale environmental parameters only for the month of May are derived. Circulation indices such as the North Atlantic Oscillation, Arctic Oscillation, and the Antarctic Oscillation are downloaded from the NCEP's Climate Prediction Center.

3. TROPICAL CYCLONE CLIMATOLOGY

3.1 Annual Cycle

The long-term distribution of monthly TC counts in the vicinity of Taiwan from 1970 to 2003 is shown in Fig. 1. Clearly, an annual cycle is evident, with no activity from December to March, an extended active season from June through October, and a peak in August. From 1970 to 2003, 131 TC occurrences were identified in this target region. Out of these 131 TCs,

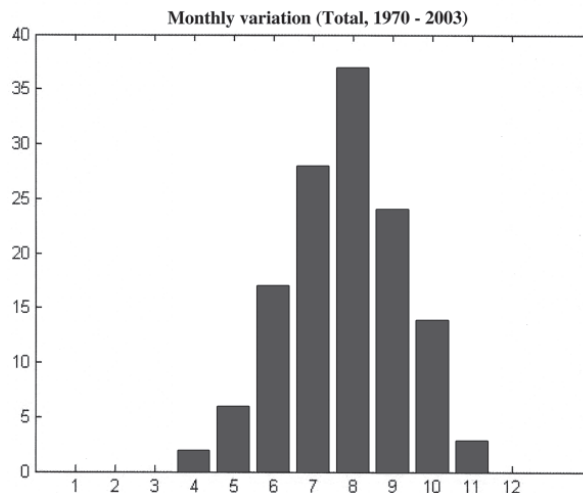


Fig. 1. Monthly tropical cyclone distribution in the vicinity of Taiwan during 1970 - 2003. Total cyclone counts are denoted in the vertical axis, and the month (e.g., 1 for January) is indicated in the horizontal axis.

120 were observed (92%) during the five-month period from June to October, and the TC frequency during these five months will be used as the predictand. One interesting feature in Fig. 1 is that the TC frequency in June slightly outnumbers that in October. This is in contrast to that in the western North Pacific where the climatological TC numbers in October are two to three times greater than those in June (Chia and Ropelewski 2002).

3.2 Interannual Variability

Tropical cyclone counts over the entire western North Pacific during the developing year of the El Niño do not differ much from the long-term climatology, yet the genesis location during the peak and late season of the El Niño developing year is dramatically shifted east- and southward so fewer storms are found to the west of 150°E (Wang and Chan 2002; Chu 2004). Given this pronounced shift in the formation location of TCs, it is instructive to observe how TC frequency in the vicinity of Taiwan evolves during a typical El Niño year. To define an El Niño year, we adopt the definition of the NOAA/NCEP Climate Prediction Center (CPC). That is, the 3-month running mean of sea surface temperature anomalies in the Nino3.4 region (5°N - 5°S, 170 - 120°W) is greater than 0.5°C for at least five consecutive overlapping seasons. Because our interest is the typhoon season, we choose to modify the CPC's definition of an El Niño event by choosing that the time period is extended from June through October. Based on this modification to the CPC's definition, the years 1972, 1982, 1987, 1991, 1994, 1997, and 2002 are identified as El Niño years. The mean annual TC counts in the target area during El Niño years are 4.0, which are very close to the climatological annual mean number of 3.9 during the entire 1970 - 2003 period. However, the annual TC frequency varies from as low as one in 1993 to as high as eight in 2004. This large interannual variability emphasizes that factors other than El Niño must be taken into account if skillful prediction of seasonal TC frequency is sought. Kim et al. (2005) also noted that the record-breaking typhoons (10) observed in Japan in 2004, which is not an El Niño year, can be accounted for by the anomalous large-scale circulation.

4. LEAST ABSOLUTE DEVIATION REGRESSION

A linear regression model can be generally written as

$$y = \sum_{j=1}^K c_j x_j + \varepsilon \quad , \quad (1)$$

where y is the desired predictive variable or predictand; $x_j, j = 1, \dots, K$ represent the predictors; $c_j, j = 1, \dots, K$ are the corresponding regression parameters; and ε is a random variable and denotes the regression deviations (residuals). Least square error (LSE) is probably the best known method for fitting a linear regression model and by far the most widely used. However, LSE is not necessarily the optimal fitting method if the deviation ε is not Gaussian distributed. Moreover, the key statistic in LAD is sample median, whereas in LSE it is sample mean.

Because sample median usually is a more robust estimator for a location parameter than sample mean, LAD regression estimates are less sensitive to large outliers (e.g., extreme values) than the LSE method.

The basic idea of LAD regression problem is generally stated as below. Given N training points, $\{\mathbf{x}_i, y_i\}, \mathbf{x}_i \in R^K, i = 1, \dots, N$, an LAD fitting problem is to find a minimizer, $\hat{\mathbf{c}} \in R^K$, of the distance function (absolute deviation):

$$f(\mathbf{c}) = \sum_{i=1}^N \left| y_i - \sum_{j=1}^K c_j x_{ij} \right| = \sum_{i=1}^N \left| y_i - \langle \mathbf{c}, \mathbf{x}_i \rangle \right|, \quad (2)$$

where $\mathbf{y} = [y_1, y_2, \dots, y_N]'$, $\mathbf{x} = [x_1, x_2, \dots, x_N]'$, $\mathbf{c} = [c_1, c_2, \dots, c_K]'$, such that, $f(\hat{\mathbf{c}}) = \min[f(\mathbf{c})]$.

This problem is solvable because $f(\mathbf{c})$ is continuous and convex. Due to the nonlinearity of absolute operation, solving a LAD regression problem is obviously nonlinear. Many algorithms were developed based on the well-studied linear programming (LP) problem, because LAD and LP are similar in their very basic nature. An LP problem in standard form is to find $\hat{\mathbf{x}}$ which maximizes $f(\mathbf{x}) = \langle \mathbf{c}, \mathbf{x} \rangle$ subject to constraints $\mathbf{Ax} \leq \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$ with given vectors \mathbf{c} and \mathbf{b} and a matrix \mathbf{A} . With some straightforward but long derivations, it can be shown that any LAD curve-fitting can be equivalently expressed as a bounded feasible LP problem. In this study, we choose the Bloomfield-Steiger (BS) algorithm (Bloomfield and Steiger 1980) to find the minimizer. The basic idea of the BS algorithm is to find the normalized steepest direction in each iteration of the algorithm. Suppose the current fit is \mathbf{c} , and $\{\delta_1, \delta_2, \dots, \delta_K\}$ is a set of directions along which the next iteration could move, the optimum descent direction being δ_p along which

$$\min[f(\mathbf{c} + t\delta_p), t \in R] = \min\{\min[f(\mathbf{c} + t\delta_j)], 1 \leq j \leq K\} . \quad (3)$$

In (3), the inner minimization is over t in R . To find this direction, K -weighted median calculations would need to be done [one for each j in the right hand side of (3)]. A pseudo code for the Bloomfield-Steiger algorithm is given in Appendix.

5. PROCEDURES FOR SELECTING PREDICTOR VARIABLES

Tropical cyclone activity, on a seasonal time scale, is modulated by large-scale environmental conditions (e.g., Gray 1977; Watterson et al. 1995). In the following, correlation analysis between seasonal TC occurrences and the environmental parameters (e.g., SST, relative vorticity) in the preceding May over the tropical western North Pacific is used to identify physical relationships. If correlations over a particular area of the western North Pacific are found to be statistically significant at the 95% confidence level, the parameter over this area is retained as a potential predictor variable. For a sample size of 34, this critical value is 0.34 when a two-tailed t-test is applied.

5.1 Sea Surface Temperatures (SSTs)

SSTs are known to be important for TC formation and intensification. Warmer SSTs are expected to fuel the overlying atmosphere with additional warmth and moisture, thereby reducing atmospheric stability and increasing the likelihood of deep tropical convection. Deep convection such as organized thunderstorm clouds acts as the primary mechanism for the vertical coupling of the lower and upper tropospheric flow patterns for TC formation. The contour plot for the correlation between TC counts and SST is shown in Fig. 2a where a large area with positive and significant correlations is found in the Philippine Sea and its adjacent ocean. Small critical regions are also noted in the East China Sea and the extreme northern South China Sea. Thus the average of the SST series over the critical regions is chosen as a predictor.

5.2 Sea Level Pressures (SLPs)

The contour plot for the correlation between the seasonal TC frequency in the vicinity of Taiwan and the May SLP is shown in Fig. 2b. A very limited critical area with negative correlation is found near 132.5°E in the Philippine Sea. This result is physically reasonable as lower SLPs to the southeast of Taiwan in May correspond to higher TC frequency near Taiwan, and vice versa. Dynamically, the juxtaposition of the maximum correlations found in Figs. 2a and b suggests a Rossby-wave type response of atmosphere to equatorial heating as demonstrated in Gill's model.

5.3 Precipitable Water (PW)

The entrainment of drier air in the mid-troposphere results in less buoyancy for the tropical convection systems and diminishes upper-level warming due to decreased release of latent heat (Knaff 1997). Consequently, drier atmosphere tends to suppress deep convection and inhibit TC activity. Positive and strong correlations between PW and TC frequency are found in the core of the tropical western North Pacific where the highest correlations reach more than 0.5 in the Philippine Sea (Fig. 3a). Of particular note is the approximately coincident critical region revealed in Figs. 2a and 3a, suggesting that warmer SSTs are associated with higher atmospheric moisture in the Philippine Sea, and both contribute to greater TC frequency near Taiwan in the following season.

5.4 Low-Level Relative Vorticity (VOR)

The monsoon trough in the western North Pacific is characterized by the strong relative cyclonic vorticity in the lower troposphere and is known to be the birthplace of typhoons. Figure 3b displays the correlation between the seasonal TC counts and the antecedent low-level relative vorticity. Two critical regions, one in the Philippine Sea and the other in the subtropical region to the east of Iwo Jima, are seen. Interestingly, the critical region in the Philippine Sea coincides well with that in Fig. 2b. A synthesis of the correlation maps in Figs. 2 and 3 suggests that lower SLPs in the Philippine Sea induce stronger cyclonic vorticity near the surface

and higher moisture in the atmosphere, leading to greater TC frequency in the vicinity of Taiwan.

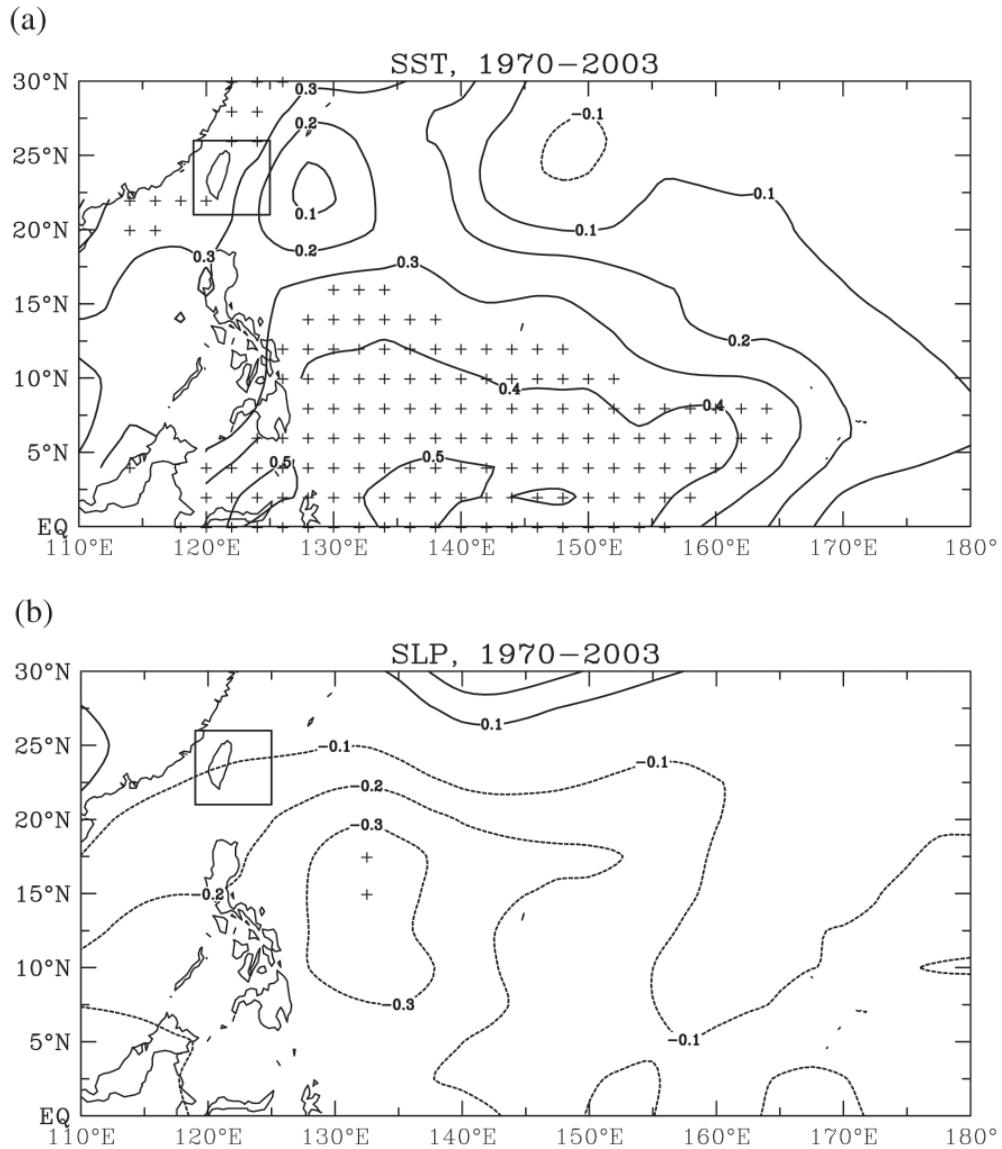


Fig. 2. (a) Correlation map between tropical cyclone count series (JJASO) in the vicinity of Taiwan and the preceding May sea surface temperatures over the tropical western North Pacific. The points with significant correlation are marked by cross. The rectangular box denotes the vicinity of Taiwan. (b) Same as (a) except for the May sea level pressures.

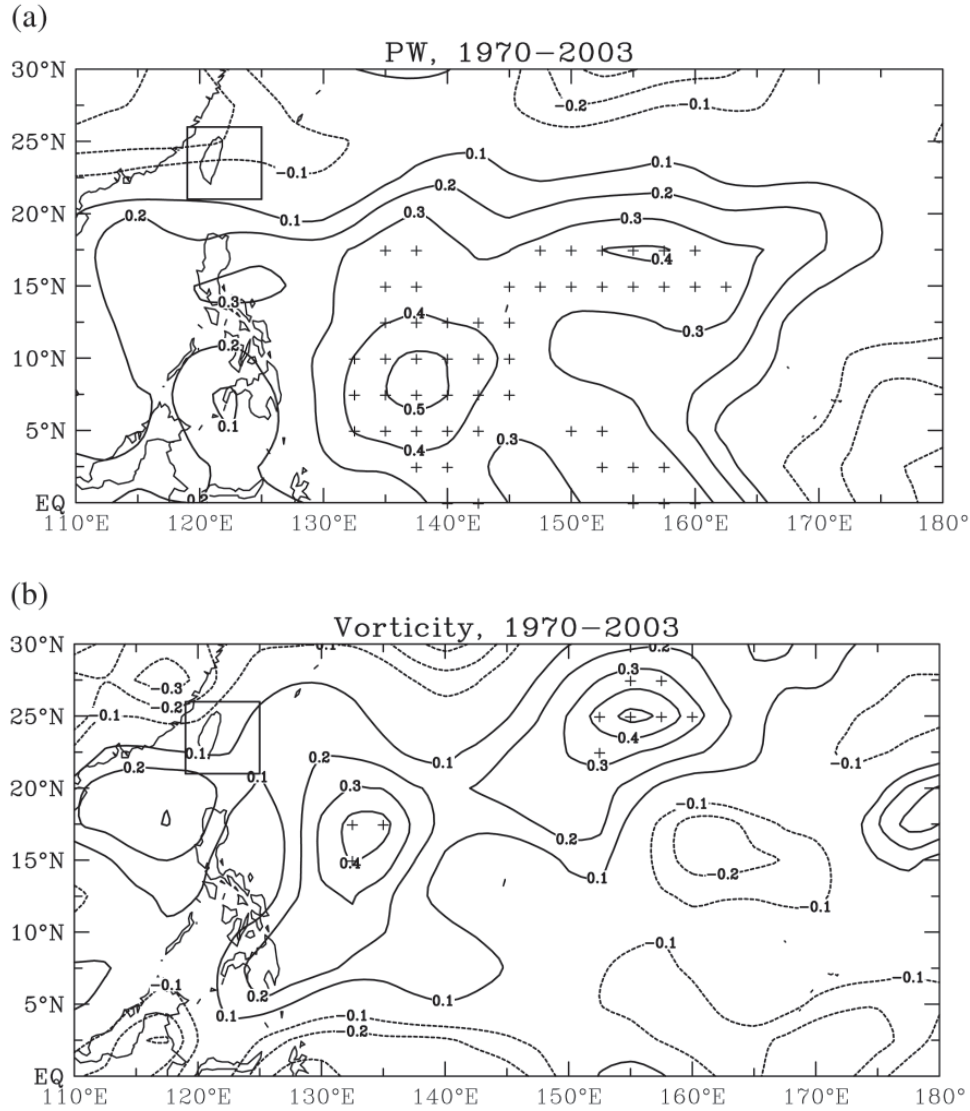


Fig. 3. (a) Same as Fig. 2a, except for the May precipitable water. (b) Same as Fig. 2a, except for the May relative vorticity at 850 hPa.

5.5 Vertical Wind Shear (VWS)

Strong VWS disrupts the organized deep convection (the so-called ventilation effect) which inhibits intensification of the TCs. In Fig. 4, correlation values in the western North Pacific are very weak except for a region near the southern Philippines, around 7.5 - 10°N, 120 - 130°E.

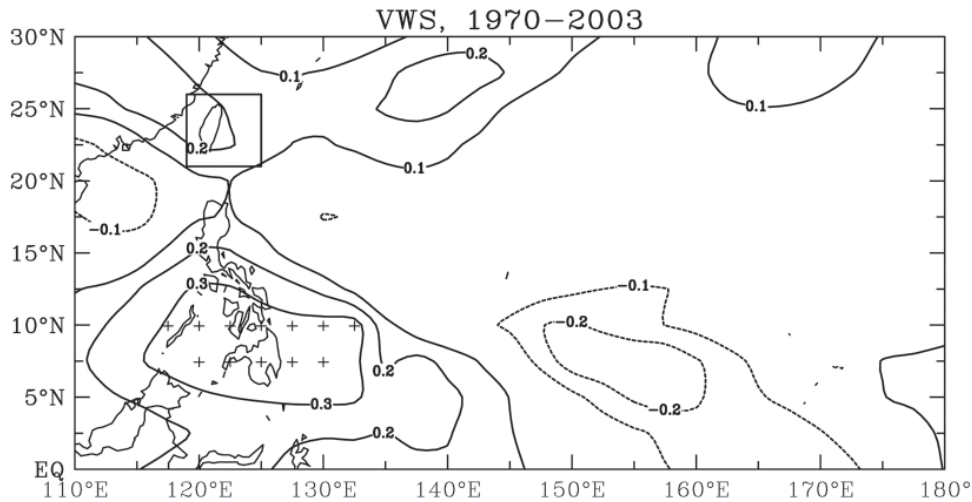


Fig. 4. Same as Fig. 2a, except for the May vertical wind shear.

It is likely that when the VWS is anomalously strong in Mindanao, typhoons tend to veer northwestward instead of tracking westward, leading to more TCs near Taiwan.

5.6 Other Circulation Indices

The predictor variables considered so far could be thought of as “local” parameters since they all reside in the western North Pacific. In the forecasting enterprise, it is also possible that some “remote” parameters, which may be located farther away from the study domain, may affect TC activity. In this study, these “remote” parameters include the North Atlantic Oscillation, Arctic Oscillation, and the Antarctic Oscillation. However, none of these parameters in May exhibit significant correlations with the regional TC counts so they are not employed in this study.

6. PREDICTION RESULTS

With the various predictor variables selected through correlation analysis, we then use a cross-validation method to establish the overall forecasting ability of the LAD model. The approach is as follows (Yu et al. 1997). The predictor and predictand dataset of T time points are divided into L segments. A model is then developed using the data of $L-1$ segments. This model is then used to predict TC frequency in the remaining segment. This process is repeated by changing the segment that has been excluded from the model development. In this study, we remove only one observation at a time for each case. By doing this, we obtain N predictions.

These predicted values are then correlated with N observations and the overall forecast skill can be determined. The cross-validation results are shown in Fig. 5 and a reasonably skillful forecast is seen. In some years, forecast values are smaller than actual observations (e.g., 1982) but in other years they are larger than observations (e.g., 1996). Thus, there is no systematic bias revealed in the prediction scheme. The linear correlation between the cross-validation results and the observed TC data is 0.63.

It is also interesting to compare the correlation skill from the regression-based forecast scheme with that obtained from a random forecast. A random sample is generated from the 34 observed TC values ranging from one to six using a random number generator, and a correlation coefficient between the randomly generated time series and the observed data is computed. This procedure is repeated 10000 times and a histogram of the correlation coefficient can be constructed. Results from the Monte Carlo simulation reveal the correlation coefficient at the 95% (99%) confidence level is 0.34 (0.42). Accordingly, the correlation coefficient (0.63) from the LAD model in this study is deemed to be skillful relative to the benchmark random samples.

To gain further insight about the prediction skill, the observed TC count series was partitioned into three groups. Because of the limited data length and the repeated nature of the observed TC counts over a small region, we set the “below” and “above” thresholds from the

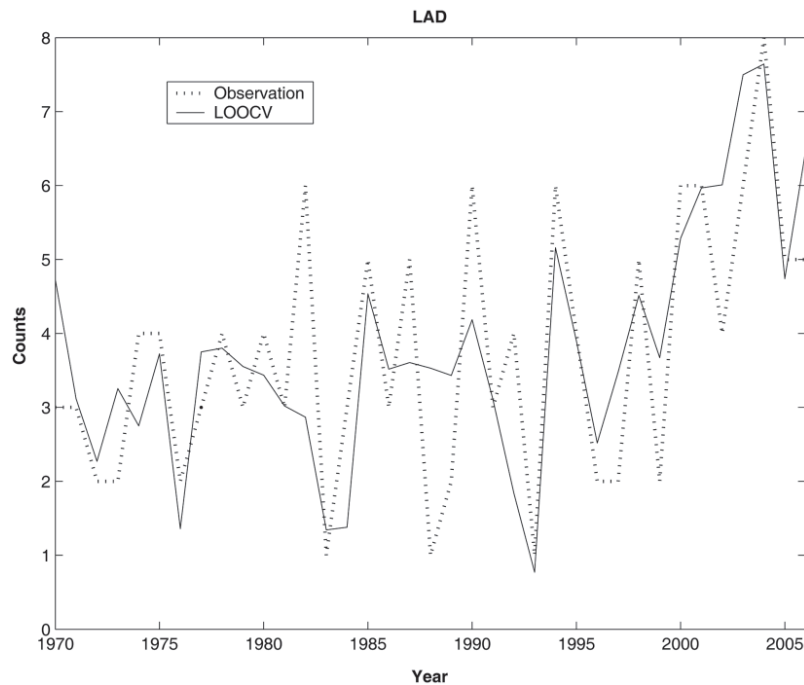


Fig. 5. Time series of the observed and the leave-one-out cross-validated forecasts of tropical cyclone counts.

“normal” category as two and five, respectively. That is, a year is classified as “below” when the seasonal TC count is equal to or less than two. Conversely, a year is chosen as “above” when the seasonal count equals or is greater than five. Accordingly, ten years are chosen as “below” normal years. They are 1972, 1973, 1976, 1983, 1988, 1989, 1993, 1996, 1997, and 1999. The nine “above” normal years are 1982, 1985, 1987, 1990, 1994, 1998, 2000, 2001, and 2003. The remaining 15 years are identified as the normal years. Similarly, the predicted TC count series is classified into three groups according to the approximate 30 : 40 : 30 proportion of the probability distribution. A 3 × 3 verification contingency table for forecasts of seasonal TC counts in terms of a tercile category is shown as Table 1. The highest numbers are along the main diagonal and the Heidke skill score, which is based on the hit rate as the accuracy measure, is 0.34. Note that forecasts equivalent to the reference forecasts receive Heidke score of zero (Wilks 1995). Thus, the LAD forecasts are better than the reference forecasts.

As our forecasts are made up to 2003, there is a concern whether our model perform well for the most recent period. The post-2003 period is particularly interesting because of the record-breaking storms (8) in 2004 and relatively high TC counts (5) in 2005 and 2006. We then generate forecasts sequentially for an independent period, 2004-06, which is not used in the test period. In terms of tercile classification as mentioned previously, our forecasts all fall in the above normal category, consistent with actual observations. Therefore, for this very limited sample size (N = 3), the forecast is perfect. If we use the entire period from 1970 to 2006, the correlation coefficient between forecasts and observations reaches 0.69 and the Heidke score raises to 0.41.

7. RELATIVE CONTRIBUTION OF PREDICTORS AND DIAGNOSTIC OF CIRCULATION FEATURES

This section examines the relative importance of the predictor variables, which is reflected in the contingency relationships between the predictand and the predictors. The predictors and

Table 1. Verification contingency table for seasonal TC forecasts in terms of categorical forecasts. BN denotes for Below Normal, N for Normal, and AN for Above Normal.

		Observed Counts		
		BN	N	AN
Forecast Counts	BN	6	4	0
	N	4	8	3
	AN	0	3	9

predictands are each classified into three categories using data from 1970 - 2003. The predictor categories are based on the 30 : 40 : 30 distribution. Following the definition used in the aforementioned section, ten “below” normal TC years and nine “above” normal years are considered.

Table 2a illustrates that during the “below” than normal TC years, SST, PW, vorticity (VOR) and VWS all tend to be normal or below normal, while SLP variations are almost indifferent to any particular category. During the “above” TC years (Table 2b), SST, PW, and

Table 2. Predictor anomaly categories during the (a) below and (b) above TC years. Symbol “-” represents the below normal category, “+” above normal category, and “O” normal category.

(a) Below-Normal Year

	TC	SST	SLP	PW	VOR	VWS
1972	-	-	0	-	0	0
1973	-	0	+	0	0	-
1976	-	-	-	0	0	0
1983	-	-	+	-	-	-
1988	-	0	0	0	0	0
1989	-	0	0	-	-	0
1993	-	-	+	0	-	-
1996	-	0	-	+	0	+
1997	-	0	+	-	-	-
1999	-	+	0	-	-	-

(b) Above-Normal Year

	TC	SST	SLP	PW	VOR	VWS
1982	+	0	+	-	-	+
1985	+	+	-	+	+	-
1987	+	0	+	-	+	-
1990	+	0	0	+	+	0
1994	+	+	-	+	+	+
1998	+	+	+	0	+	0
2000	+	+	-	+	+	+
2001	+	+	0	+	+	+
2003	+	+	-	+	+	+

VOR show a general tendency to be above normal, and VWS shows a weaker tendency to be above normal. SLP again does not show preference to any category.

The relationship between different predictors and the predictand can be further illustrated by counting the category percentage of each predictor conditioned on the category of the predictand. In Fig. 6a, during the “below” TC years the percentage of below than normal category in PW, VOR, and VWS is very high, reaching 50%. Also during the below TC years, the percentage of above years for the aforementioned three variables is comparatively low, ranging from zero for the VOR to just 10% in PW and VWS. The predictor and predictand relationship is much stronger during the “above” than the “below” TC years. For example, for the “above” TC years (Fig. 6b), the percentage of above normal VOR is extremely high (89%). The percentage of above normal SST and PW is also high (67%), while the percentage of below normal for these two predictors is 0% and 22%, respectively. In summary, VOR, SST, PW, and to a lesser extent, VWS in May appear to be good predictors during active TC years. The predictors deemed important during low TC years are VOR, PW, and VWS.

Among the five predictor variables selected, the VOR appears to exhibit the strongest linear relationship with the TC numbers. That is, when the May vorticity is in the below normal category, the following seasonal TC counts tend to be below normal and vice versa. Note

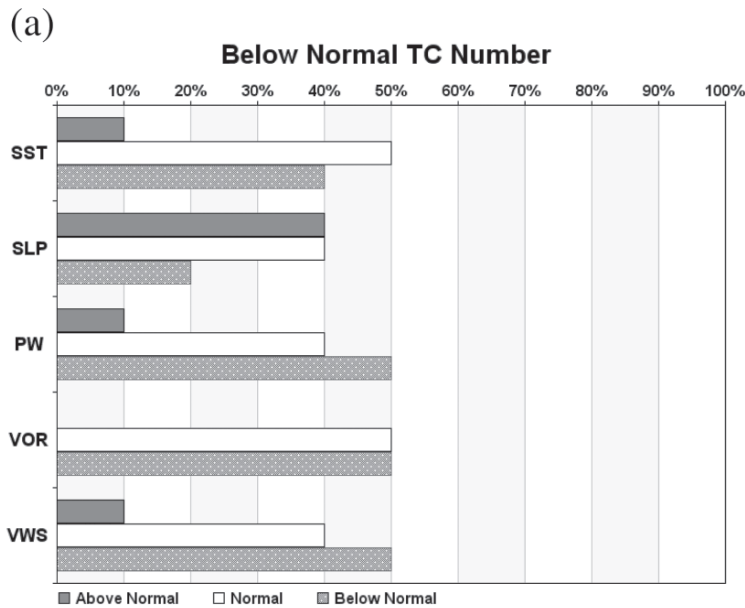


Fig. 6. The category percentage of each predictor during years of (a) below-normal and (b) above-normal tropical cyclone counts. For each panel, filled bar denotes above normal, blank for normal, and bar with small open dots for below normal category for each predictor variable (SST, SLP, PW, VOR, and VWS).

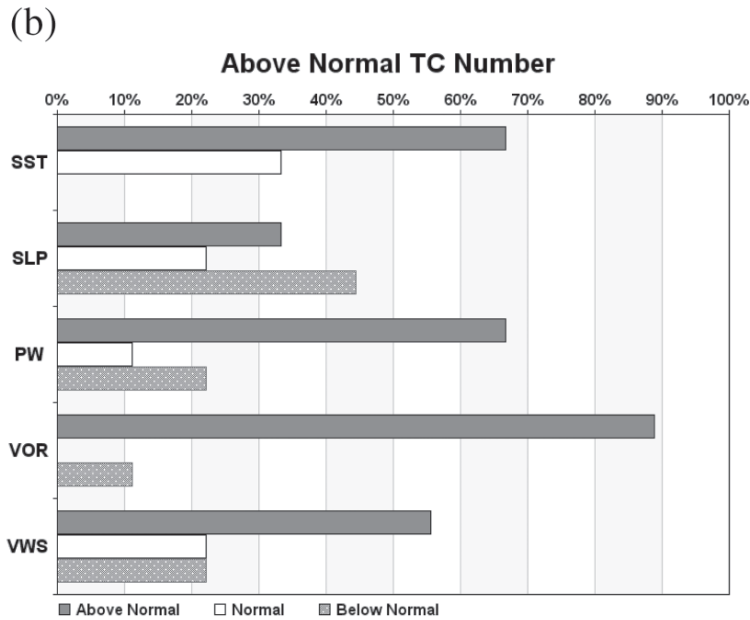


Fig. 6. (Continued)

that in Table 2a, there are five years (1983, 1989, 1993, 1997, and 1999) in which the sign for VOR is negative when the TC count is below normal. When TC counts are above normal (Table 2b), there are eight years (1985, 1987, 1990, 1994, 1998, 2000, 2001, and 2003) in which positive signal appears in VOR. In the following, in order to understand why predictors used in the antecedent May could provide reasonable skill for the following seasonal TC prediction, we will present maps illustrating the evolution of large-scale circulation variables from May through October when the VOR is above or below normal. The composite maps of SST, PW, and low level winds are presented together for these two contrasting groups. To simplify the presentation, the months June through August (JJA) and September to October (SO) are each averaged.

Figure 7a shows a large anomalous anticyclone associated with below normal precipitable water over the tropical northwestern Pacific in May during the below than normal TC years concurrent with the below than normal low-level vorticity. The anticyclone ridge at lower troposphere extends from the Philippine Sea northeastward reaching about 160°E and 27°N. Interestingly, such an orientation of anticyclone is consistent with the correlation map between the seasonal TC frequency and the antecedent vorticity field in Fig. 3b. Probably affected by the overlying anticyclone, the SST anomaly along the southern rim of the anticyclone is negative, while along the northern rim the anomaly is positive when the low-level winds are southerlies. By JJA, the low level anticyclone shrinks but is still evident over the Philippine and Taiwan

(Fig. 7b). The westerly anomalies over Taiwan are unfavorable for steering TCs into Taiwan. The anticyclone appears to expand from JJA to SO, engulfing the western North Pacific in SO (Fig. 7c). Associated with the persistent anti-cyclonic circulation, negative PW anomalies (i.e., the dry condition) dominate over the Philippine Sea from May to SO. Persistent dry atmosphere and anti-cyclonic circulation (i.e., below normal cyclonic vorticity) with westerly or southwesterly wind anomalies in the subtropics provide unfavorable large-scale conditions for tropical cyclone to develop and/or track through the vicinity of Taiwan during the TC season.

For the above normal TC years, Fig. 8 shows that warmer SST over a large area of the northwestern Pacific is persistent from May to October. A cyclonic circulation, which is almost exactly opposite to that shown in the May diagram of Fig. 8a, is observed between the South China Sea and 140°E in Fig. 8a. Low-level westerlies from the South China Sea to the Philippine

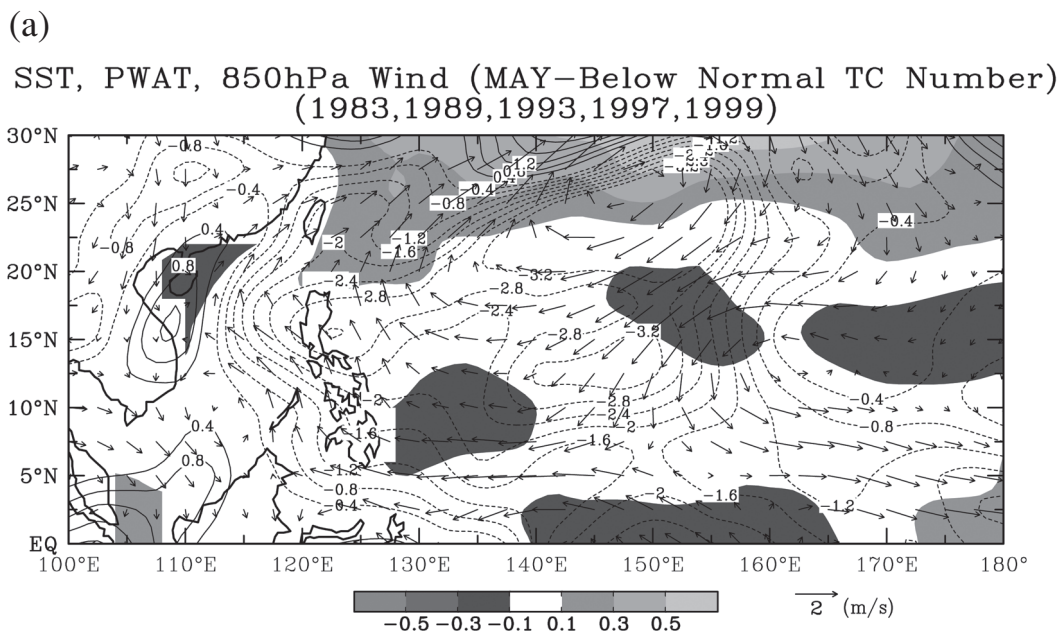
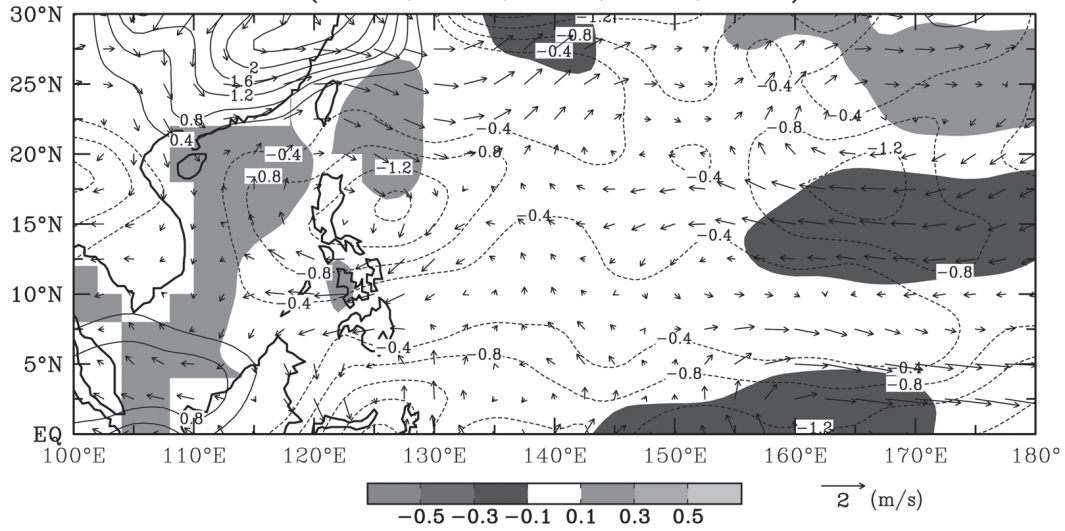


Fig. 7. Composite maps of anomalous sea surface temperatures (shading), precipitable water (contours), and the wind vectors at 850 hPa for the concurrent events of below-normal tropical cyclone years in the vicinity of Taiwan and below-normal vorticity category. The upper panel is for May (a), the middle panel for June-August (b), and the bottom panel for September-October (c). Anomalous values are obtained by taking the difference between the composite values and the respective long-term climatology. For precipitable water, solid (dashed) line denotes positive (negative) differences. For sea surface temperature, the unit is °C; for precipitable water, the unit is kg m^{-2} ; and for vector winds, the unit is m s^{-1} .

(b)
 SST, PWAT, 850hPa Wind (JJA–Below Normal TC Number)
 (1983,1989,1993,1997,1999)



(c)
 SST, PWAT, 850hPa Wind (SO–Below Normal TC Number)
 (1983,1989,1993,1997,1999)

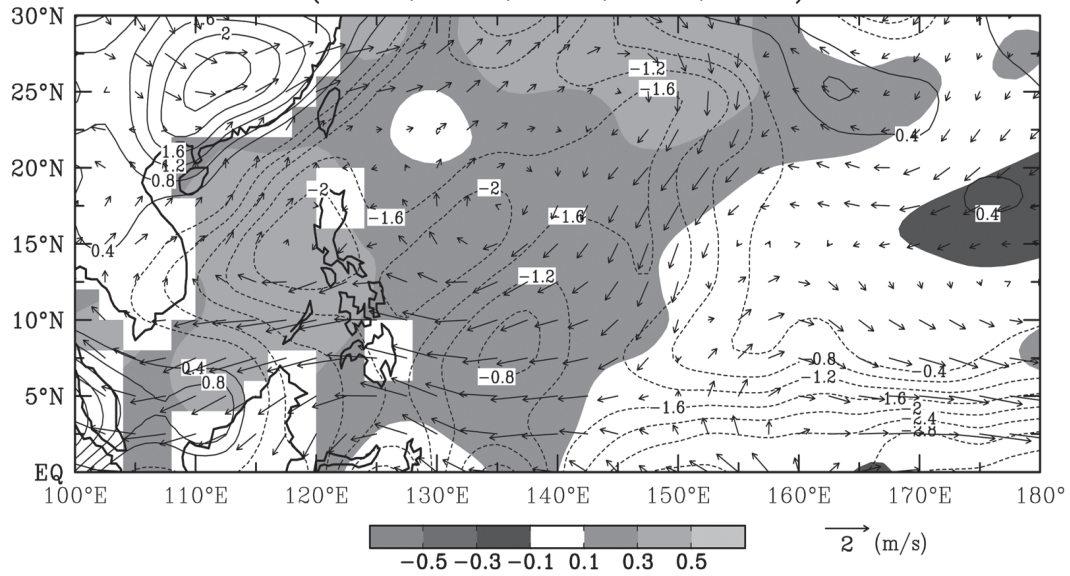
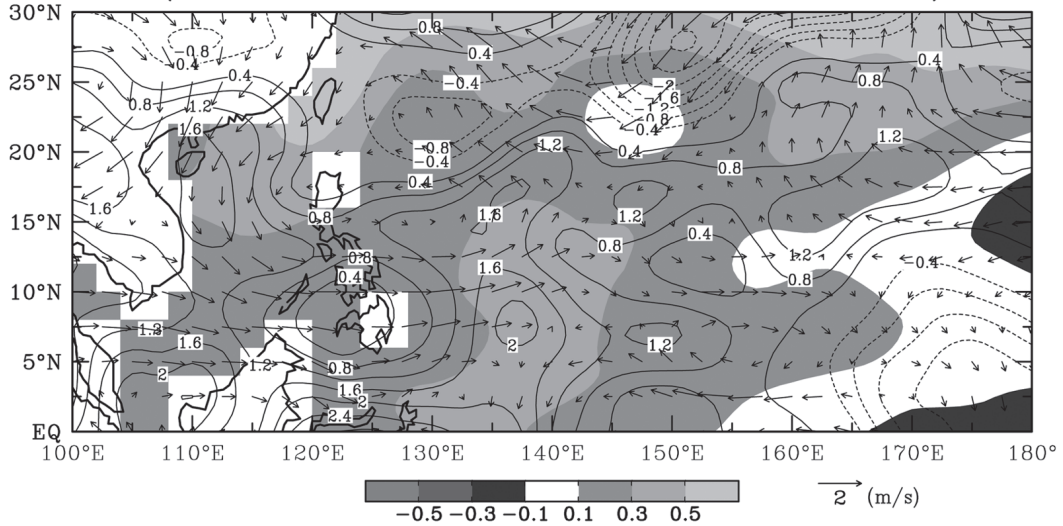


Fig. 7. (Continued)

(a)
 SST, PWAT, 850hPa Wind (MAY–Above Normal TC Number)
 (1985,1987,1990,1994,1998,2000,2001,2003)



(b)
 SST, PWAT, 850hPa Wind (JJA–Above Normal TC Number)
 (1985,1987,1990,1994,1998,2000,2001,2003)

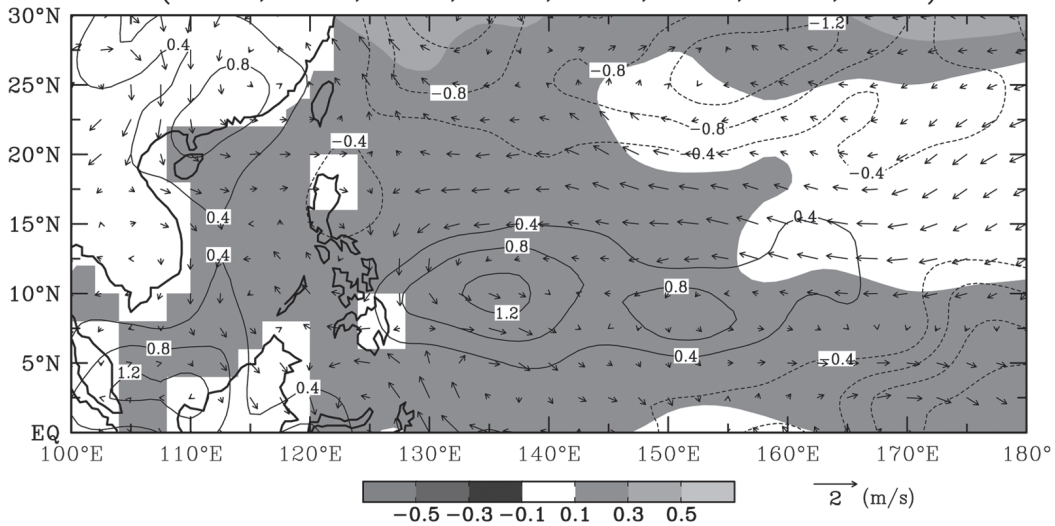


Fig. 8. Same as Fig. 7 but for the concurrent events of above-normal tropical cyclone years in the vicinity of Taiwan and above-normal vorticity category.

(c)

SST, PWAT, 850hPa Wind (SO—Above Normal TC Number)
(1985, 1987, 1990, 1994, 1998, 2000, 2001, 2003)

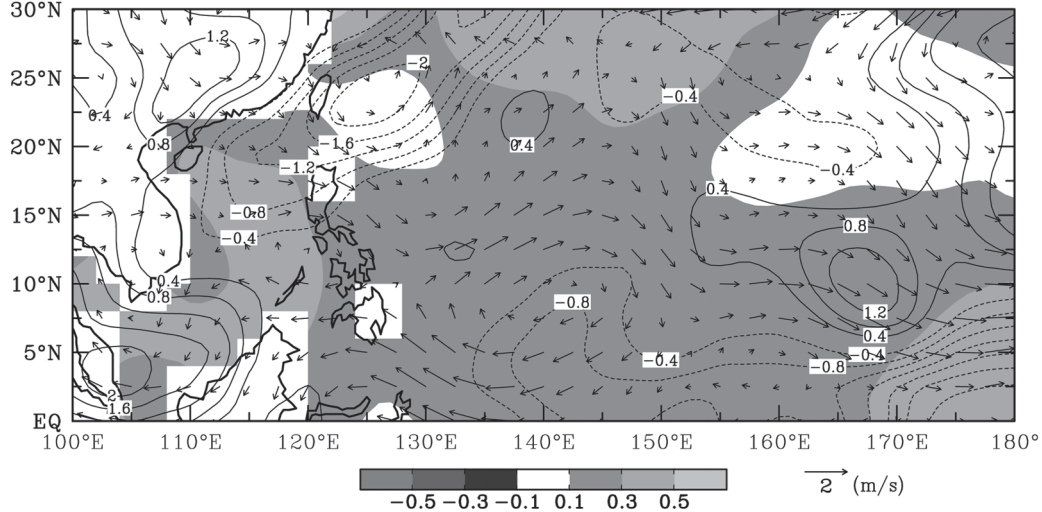


Fig. 8. (Continued)

Sea are evident. Above normal precipitable water appears along the southern rim of the anomalous cyclonic circulation, with three anomaly centers appearing over the South China Sea, the Sulu Sea and the Philippine Sea. The anomalous cyclone is weaker but still can be recognized in JJA. Also note that easterly or southeasterly anomalies in the subtropical western North Pacific tend to steer storms into the vicinity of Taiwan. By SO, the cyclone becomes more confined regionally and centered near Taiwan (Figs. 8b, c). The presence of westerlies in low latitudes and easterlies in the subtropics creates an environment favorable for the generation of low-level cyclonic shear and cyclonic relative vorticity in the western portion of the western North Pacific (Fig. 8c). The above normal PW appears in the critical region from May through SO (Figs. 3a, 8). Since most storms form in the critical region (Chia and Ropelewski 2002), the positive PW anomalies tend to support active tropical convection in the critical region. In short, warm SST, moist atmosphere, and a low-level cyclonic circulation coupled with low-latitude westerly anomalous winds in the Philippine Sea are favorable conditions for tropical cyclones.

8. SUMMARY

Climate prediction of tropical cyclone activity has traditionally been carried out for the North Atlantic and the western North Pacific by various research teams. Because of the vast

expanse of ocean basins and pronounced interannual climate variations in the tropics, there is no guarantee that such basin-wide prediction is also applicable to smaller regions within a basin. In this study, a multivariate least-absolute-deviation regression method is adopted to predict the seasonal (June to October) tropical cyclone frequency in the vicinity of Taiwan using large-scale climate information available in the preceding May. The environmental variables considered include the sea surface temperatures, sea level pressure, precipitable water, low-level relative vorticity, and vertical wind shear. Through lagged correlation analysis between regional TC frequency and individual climate variables over the tropical western North Pacific, we identified critical regions to be used as the predictor datasets. We then used the leave-one-out cross-validation technique to test the predictability of TC frequency. The cross-validation provides a nearly unbiased estimate of true forecast skill.

The linear correlation between the cross-validation predictions and the corresponding actual observations for the test period 1970 - 2003 (1970 - 2006) is 0.63 (0.69), which is significant at the 99% confidence level and much higher than that obtained from a random forecast. Independently, the Heidke score of the prediction when the tercile category of TC counts is used is 0.34 (0.41) for 1970 - 2003 (1970 - 2006). These results imply that it would be possible to predict the seasonal TC counts for a small area with reasonable skill using a physically based regression model. To evaluate the relative contribution of each predictor, the predictors and predictand are each classified into three categories (i.e., below normal, normal, and above normal). Warmer SST, higher precipitable water, and the presence of an anomalous cyclone coupled with low-latitude westerlies in the western portion of the western North Pacific in the subtropics are the precursors to higher-than-normal TC activity near Taiwan. Interestingly, these features are persistent from May throughout the typhoon season and may account for the good predictive skill achieved for the seasonal TC frequency when the pre-season predictors are used. On the other hand, lower precipitable water and low-level anomalous anticyclone with southwesterly or westerly anomalies in the subtropics are characteristics of lower-than-normal TC activity near Taiwan, and these features already begin to show in May. In the future, it would be of interest to determine the predictability of TC frequency when climate variables chosen are for months prior to May. Apart from pure scientific inquiry, if good skills can be obtained, say, when the April predictors are chosen, it would allow decision makers more lead time to respond.

Acknowledgements This project was funded by the Central Weather Bureau, CWB 95-3M-06. Comments and suggestions made by three anonymous reviewers helped improve the presentation of the manuscript substantially. Part of this project is supported by NSC 95-2625-Z-052-006.

REFERENCES

- Bloomfield, P., and W. L. Steiger, 1980: Least absolute deviations curve-fitting. *J. Sci. Stat. Comput.*, **1**, 290-301.

- Chan, J. C. L., J. S. Shi, and C. M. Lam, 1998: Seasonal forecasting of tropical cyclone activity over the western North Pacific and the South China Sea. *Weather Forecast.*, **13**, 997-1004.
- Chia, H. H., and C. F. Ropelewski, 2002: The interannual variability in the genesis location of tropical cyclones in the northwest Pacific. *J. Climate*, **15**, 2934-2944.
- Chu, P. S., 2004: ENSO and tropical cyclone activity. In: Murnane, R. J., and K. B. Liu (Eds.), *Hurricanes and Typhoons: Past, Present, and Potential*, Columbia University Press, 297-332.
- Clark, J. D., and P. S. Chu, 2002: Interannual variation of tropical cyclone activity over the central North Pacific. *J. Meteor. Soc. Japan*, **80**, 403-418.
- Gray, W. M., 1977: Tropical cyclone genesis in the western North Pacific. *J. Meteorol. Soc. Japan*, **55**, 465-482.
- Gray, W. M., C. W. Landsea, P. W. Mielke, and K. J. Berry, 1992: Predicting Atlantic seasonal hurricane activity 6 - 11 months in advance. *Weather Forecast.*, **7**, 440-455.
- Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, B. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, Roy Jenne, and Dennis Joseph, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteorol. Soc.*, **77**, 437-471.
- Kim, J. H., C. H. Ho, and C. H. Sui, 2005: Circulation features associated with the record-breaking typhoon landfall on Japan in 2004. *Geophys. Res. Lett.*, **32**, L14713, doi: 10.1029/2005GL022494.
- Kistler, R., E. Kalnay, W. Collins, S. Saha, G. White, J. Woollen, M. Chelliah, W. Ebisuzaki, M. Kanamitsu, V. Kousky, van den Dool, Huug, R. Jenne, M. Fiorino, 2001: The NCEP-NCAR 50-Year Reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteorol. Soc.*, **82**, 247-267.
- Knaff, J. A., 1997: Implications of summertime sea level pressure anomalies in the tropical Atlantic region. *J. Climate*, **10**, 789-804.
- Wang, B., and J. C. L. Chan, 2002: How strong ENSO affect tropical storm activity over the western North Pacific. *J. Climate*, **15**, 1643-1658.
- Watterson, I. G., J. L. Evans, and B. F. Ryan, 1995: Seasonal and interannual variability of tropical cyclogenesis: diagnostics from large-scale fields. *J. Climate*, **8**, 3052-3066.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.
- Yu, Z. P., P. S. Chu, and T. Schroeder, 1997: Predictive skills of seasonal to annual rainfall variations in the U.S. affiliated Pacific islands: Canonical correlation analysis and multivariate principal component regression approaches. *J. Climate*, **10**, 2586-2599.
-
- Chu, P. S., X. Zhao, C. T. Lee, and M. M. Lu, 2007: Climate prediction of tropical cyclone activity in the vicinity of Taiwan using the multivariate least absolute deviation regression method. *Terr. Atmos. Ocean. Sci.*, **18**, 805-825, doi: 10.3319/TAO.2007.18.4.805(A).

APPENDIX

Pseudo Code for the Bloomfield-Steiger Algorithm:

- [1] Accept $\{\mathbf{x}_i, y_i\}$, $\mathbf{x}_i \in R^k$, $i=1, \dots, n$

- [2] $A \leftarrow \begin{pmatrix} \mathbf{X} & \mathbf{y} \\ \mathbf{I}_k & \mathbf{0} \end{pmatrix}$, where $\mathbf{X}=[\underline{x}_1, \dots, \underline{x}_n]'$, $\mathbf{y}=[y_1, \dots, y_n]'$, \mathbf{I}_k is an $k \times k$ identity matrix
and $\mathbf{0} \in R^k$

- [3] $j \leftarrow 0$

- [4] $\mathbf{Z} \leftarrow \{ i, 1 \leq i \leq n : A_{i, k+1} = 0 \}$

- [5] for $m = 1$ to k
 - $g_m = -\sum_{\mathbf{Z}} |A_{i, m}|$
 - $h_m = \sum_{\mathbf{Z}'} A_{i, m} \text{sign}(A_{i, k+1})$
 - $f_m = \max(g_m - h_m, g_m + h_m) / \sum_{i=1}^n |A_{i, m}|$
 - end

- [6] $p \leftarrow \text{argmin}\{f_i, i=1, \dots, n\}$
if $f_p \leq 0$ go to [10]

- [7] $t \leftarrow A_{q, k+1} / A_{qp}$, the weighted median of $\{A_{i, k+1} / A_{i, p}, i=1, \dots, n\}$ with weights $|A_{i, p}|$
- [8] for matrix \mathbf{A} :
 - column $p \leftarrow$ column p / A_{qp}
 - column $i \leftarrow$ column $i - A_{qi} * \text{column } p$

- [9] $j \leftarrow j+1$ go to [4]

- [10] $\underline{c}_i \leftarrow -A_{n+1, k+1}$, $i=1, \dots, k$