Proceedings of Joint Conference of Korean Data And Information Science Society and Korean Data Analysis Society, June 21-22, 2002, pp. 9-15

Prediction of the number of Tropical Cyclones over Western North Pacific in TC season (여름철 북서태평양 태풍발생 예측을 위한 통계적 모형 개발)

Keon Tae Sohn[1], *Changkon Hong*[2], *H. Joe Kwon*[3] *and Jung Kyu Park*[4]

Abstract

This paper presents the seasonal forecasting of the occurrence of tropical cyclone (TC) over Western North Pacific (WNP) using the generalized linear model (GLM) and dynamic linear model (DLM) based on 51-year-data (1951-2001) in TC season (June to November). The numbers of TC and TY are predictands and 16 indices (the El Nino/Southern Oscillation, the synoptic factors over East asia and WNP) are considered as potential predictors. With 30-year moving windowing, the estimation and prediction of TC and TY are performed using GLM. If GLM forecasts have some systematic error like a bias, DLM is applied to remove the systematic error in order to improve the accuracy of prediction.

1. Introduction

In order to reduce the damage from typhoons the improvement on the accurate prediction of inforamtion about the typhoons is an important target. Major targets of forecasting are (1) typhoon track, (2) typhoon intensity and (3) typhoon activity. Sohn et al. (2001) applied dynamic linear model (DLM) to forecast typhoon tracks. Their DLM uses the numerical model forecasts and observsations as inputs. If numerical model forecasts have some systematic error like a bias, it should be the first step to remove the systematic error in order to improve the accuracy of prediction. In this case DLM may be useful. During the last two decades the forecasting techniques on TC activity have been developed by many authors; Gray et al. (1993, 1994) have repoted the prediction of TC activity in Atlantic using multiple linear regression models. Elsner and Schmertamann (1993) applied generalized linear model (GLM) for Atlantic hurricanes. Chan et al. (1998) and Sohn et al. (2002) considered a projection pursuit regression

[1]Professor, Department of Statistics, Pusan National University
[2]Professor, Department of Statistics, Pusan National University
[3]Professor, Department of Atmospheric Science, Kongju National University
[4]Director, Department of Climate forecating, Korea Meteorological Administration

(PPR) for TC activity over WNP.

This paper concentrates on the prediction of occurrence of the tropical cyclone (TC; tropical storm and typhoon) activity in TC season (Jume to November) over the Western North Pacific (WNP) based on 51 years data (1951-2001). The number of TC and TY are predictands and 16 indices (the El Nino/Southern Oscillation, the synoptic factors over East asia and WNP) are considered as potential predictors. For the prediction, two statistical models are applied; GLM with log link (that is, Poisson regression) and DLM.

2. Data used for our study

2.1 Predictands

This paper investigates the prediction of the number of TC and TY in TC season (June to November). The data set of monthly numbers of TC and TY over the WNP during 51 years (1951-2001) is used for the study. Every predictand are the cumulative number of TC season each year and their values are obtained from RSMC, Tokyo. Chan and Shi (1996) has shown that annual TC activity over the WNP for the 30 years (1965-1994) has a long-term trend and short-term fluctuations with periods of 2 and 7 year. But the number of TC over the WNP for the 51-year (1951-2001) has the autocorrelation with lag 9 according to autocorrelation check. And the number of TC has no significant autocorreation.

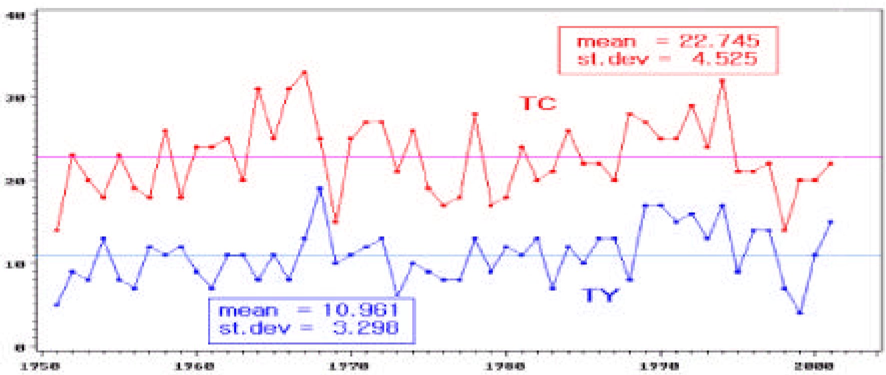

Fig. 1. Time series plots of the number of TC and TY in TC season

2.2 Potential Predictors

In order to choose the potential predictors, Chan et al. (1998) and related papers are referred. 16 indices are considered as potential predictors (Table 1). These values of predictors are obtained from Korea Meteorological Administration (KMA) and NCEP/NCAR reanalysis data. Every index has two values for each year: (1) the average during period June to November of the previous year and (2) the average during period January to March of the current year.

Table 1. Potential predictors

	Predictors
ENSO	SOI, NINO4, NINO3, NINO1+2, NINO West
Synoptic factors	① North Pacific High Index (Average of ZA, 10N-30N, 140E-140W) ② Australian Monsoon Circulation Index(Average of 850V, 5S-20S, 140E-170E) ③ Cross Equatorial Flow Index(Average of 850V, 5S-15N, 140E-170E) ④ QBO Index(Quasi-Biennial Ocillation, 10, 20, 30, 50 hPa) ⑤ India-Burma Trough Index ⑥ Tibetan High Index ⑦ Western Pacific Pattern Index ⑧ SH mean SLP anomaly (5S-35S, 180W-150W(150W-120W))

3. Statistical Models

Two statistical models, GLM and DLM, are considered to predict the number of TC and TY. Because the number of TC and TY is small and nonnegative integer, the GLM with log link, called Poisson regression model, may be useful. With moving 30-year windowing, the estimation and forecasting is performed. For example, in order to forecast the number of TC and TY in 1981, two step procedure is needed; (1) the estimatation of the Poisson regression model is performed based on the previous 30-year data (1951 to 1980) and (2) the forecasts in 1981 are generated using the estimated models. For year 1982 to 2002, the same procedures are repeated. If GLM forecasts have some systematic error like a bias, DLM is applied to remove the systematic error.

3.1 Poisson Regression Model

It is well known that the Poisson distribution is preferred to the normal distribution when responses have nonnegative integer values like TC occurrence. When $Y_1, Y_2, \cdots, Y_n$ are independent Poisson rnadom variables and $(x_{i1}, x_{i2}, \cdots, x_{ik})$, $i = 1, 2, \cdots, n$, are values of regressors, the Poisson regression model is given by

$$\log \mu_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} \text{ where } \mu_i = E(Y_i),\ i = 1, 2, \cdots, n .$$

3.2 Dynamic Linear Model

Since three decades, DLM as a Bayesian forecasting technique has been developed by many authors including West and Harrison (1997). The procedures of DLMs are essentially equivalent to those of the Kalman fitering(Kalman, 1960, 1963) in the normal DLMs with known error variances. Sohn et al. (2001) used the DLM for the mid-term prediction of daily max/min temperatures. Using numerical model forecasts and observations as inputs, the parameters in DLM can move dynamically and then DLM has a charateristics eliminating the systematic error. DLM with a prior distribution of error variance consists of two equations; the state equation and the output equation as follows:

$$\begin{cases} Y_t = F_t'\theta_t + v_t, & v_t \sim N(0, V) \quad \text{(output equation)} \\ \theta_t = \theta_{t-1} + \omega_t, & \omega_t \sim T_{n_{t-1}}(0, W_t) \quad \text{(state equation)} \end{cases}$$

$$\theta_0 | D_0 \sim T_{n_0}(m_0, C_0), \qquad V | D_0 \sim IG(\frac{n_0}{2}, \frac{n_0 s_0}{2})$$

where Y_t is an observarion at time t, F_t is a input vector (1, GLM forecasts and the previous observations), θ_t is the coefficient vector changed dynamically, the output error v_t has the univariate normal distribution with mean 0 and variance V which has an inverse gamma distribution $IG(\alpha, \beta)$, the innovation error w_t has the multivariate T distribution with mean vector 0 and covariance matrix W_t, v_t and w_t are independent to each other. m_t and C_t are the mean vector and covariance matrix of θ_t respectively.

4. Results

The forecasting strategy consists of three step: [Step 1] the estimation and prediction of the nember of TC and TY are performed using GLM and forward variable selection. [Step 2] Check whether GLM forecasts have some systematic error or not. [Step 3] If they have some systematic error, DLM is applied to remove the systematic error.

4.1 Prediction of the number of TC

Fig. 2 is the time series plot of forecasting error of GLM forecasts for TC. Average of Forecasting errors is 0.005 and we decided that there is no significant systematic error in GLM forecasts for TC. Fig. 3 shows the two time series plots of GLM forecasts and observations for TC. The correlation coefficient of TC and GLM forecasts is 0.574 and the root mean square of forecasting errors is 3.78. The forecast for number of TC in TC season in 2002 is 22.3.

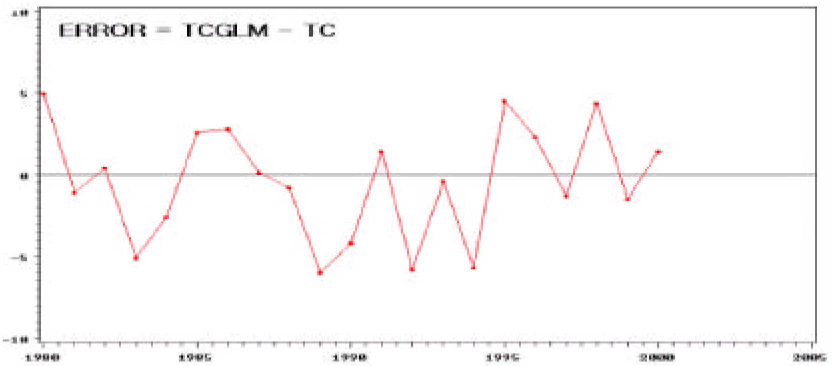

Fig. 2. Time series plot of forecasting errors of TC

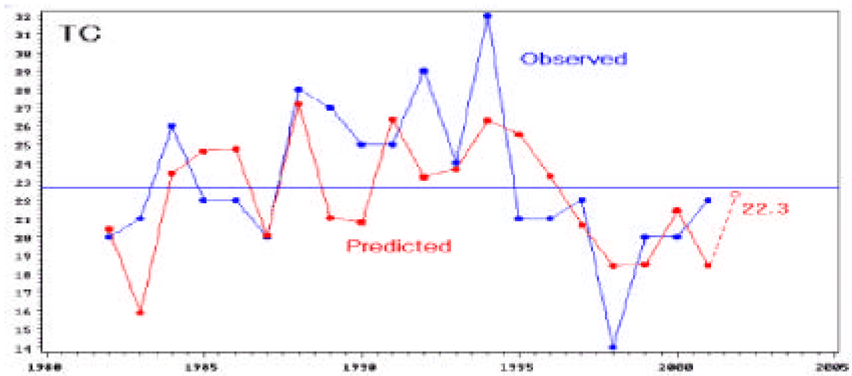

Fig. 3. Time series plots of GLM forecasts and observation for TC

4.2 Prediction of the number of TY

Fig. 4 is the time series plot of forecasting error of GLM forecasts for TY. Because GLM forecasts incline to underestimate the number of TY and the average of GLM forecasting errors is -2.25, we decided to apply DLM with GLM forecasts and observations as inputs. Fig. 5 shows the results of DLM. Average of DLM forecasting error is -0.68 and the correlation coefficient of TY and DLM forecasts is 0.704. The square root of the mean square of GLM forecasting errors is 3.344 and that of DLM forecasting errors is 2.701. The forecast for number of TY in TC season in 2002 is 11.8.

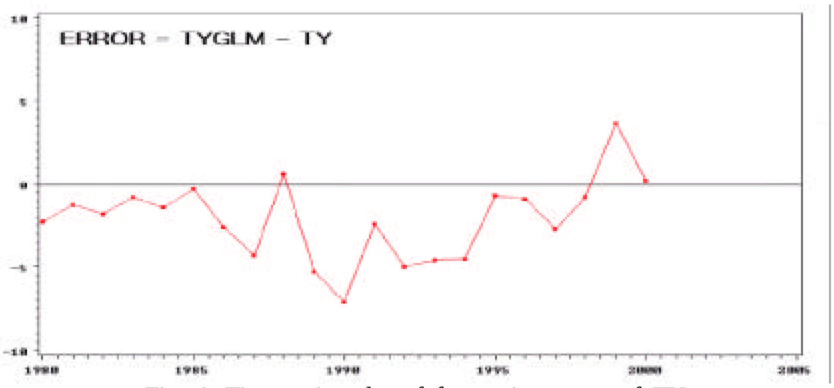

Fig. 4. Time series plot of forecasting errors of TY

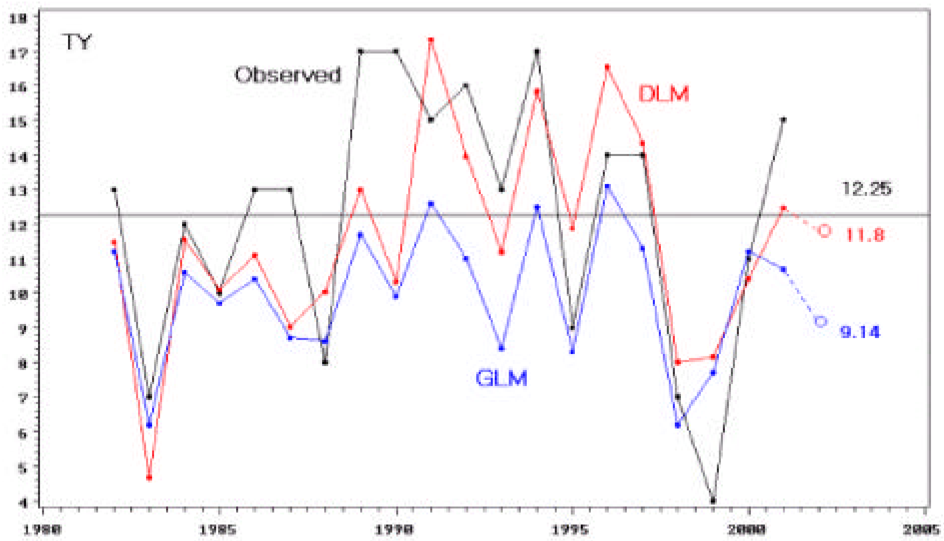

Fig. 5. Time series plots of GLM forecasts, DLM forecasts and observation for TY

5. Concluding Remarks

This paper presents statistical forecasts for the number of TC and TY over the WNP based on 51-year-data (1951-2001) using GLM and DLM. The numbers of TC and TY during TC season in 2002 is expected to be normal. In order to improve the accuracy of prediction we think that the reanalysis of potential predictors is need based on the climatological concepts and the long-term prediction numerical model forecasts should be used as additional inputs. As a further work, the prediction of the numbers of TC and TY in Watch area, Emergency area and Direct effect area will be performed.

Acknowledgements

This study was performed for "El Nino/La Nina monitoring and development of long-range prediction system", funded by Korea Meteorological Administration through Scientific Service Programs in year 2002.

REFERENCES

[1] 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.

[2] Chan, J. C. L. and J.-E. Shi (1997). Application of projection-pursuit principal component analysis method to climate studies. *International Journal of Climatology*, 17, 103-113.

[3] Elsner, J. B., and C. P. Schmertmann (1993). Improving extended- range seasonal prediction of intense Atlantic hurricane activity. *Weather Forecasting*, 8, 345-351.

[4] Gray, W. M., C. W. Landsea and P. W. Mielke (1993). Predicting Atlantic basin seasonal tropical cyclone activity by 1 August. *Weather Forecasting*, 8, 73-86.

[5] Gray, W. M., C. W. Landsea and P. W. Mielke (1994). Predicting Atlantic basin seasonal tropical cyclone activity by 1 June. *Weather Forecasting*, 9, 103-115.

[6] Lander, M. A. (1994). An exploratory analysis of the relationship between tropical storm formation in the Western North Pacific and ENSO. *Monthly Weather Review*, 122, 636-651.

[7] Sohn, K. T., S. D. Kim (2001). Mid-term Prediction of the daily maximum/minimum Temperature in Seoul area using GDAPS outputs and Dynamic linear models, *Journal of Korean Atmospheric Science Society*, 37, 13-20.

[8] Sohn, K. T., H. J. Kwon and C. Hong (2002). Prediction of Tropical Cyclone Activity over Western North Pacific in Summer Season, *Korean Journal of Atmospheric Science*, 5, 55-61.

[9] Sohn, K. T., H. J. Kwon and A. S. Suh (2001). Systematic Approach for Typhoon track forecast, Characteristics of BATS forecasts and Dynamic linear models, *Journal of Korean Atmospheric Science Society*, 37, 295-304.