Prediction of Typhoon Tracks Using Dynamic Linear Models

Keon-Tae SOHN*[1], H. Joe KWON[2], and Ae-Sook SUH[3]

[1] *Pusan National University, Pusan,* 609-735, *Korea*

[2] *Kongju National University, Konju,* 314-701, *Korea*

[3] *Meteorological Research Institute/KMA, Seoul,* 156-010, *Korea*

(Received May 15, 2002; revised January 14, 2003)

P4 A

ABSTRACT

This paper presents a study on the statistical forecasts of typhoon tracks. Numerical models have their own systematic errors, like a bias. In order to improve the accuracy of track forecasting, a statistical model called DLM (dynamic linear model) is applied to remove the systematic error. In the analysis of typhoons occurring over the western North Pacific in 1997 and 2000, DLM is useful as an adaptive model for the prediction of typhoon tracks.

Key words: typhoon track forecast, systematic error, dynamic linear model

1. Introduction

The accurate prediction of typhoon tracks is important in reducing the damage from typhoons. Several numerical models (BATS, GDAPS, RDAPS, etc.) are used to generate the typhoon track forecasts for the Korea Meteorological Administration (KMA). Both GDAPS (Global Data Assimilation and Prediction System) and RDAPS (Regional Data Assimilation and Prediction System) are baroclinic models while the BATS (Barotropic Adaptive-grid Typhoon Simulation) model is a barotropic model which adopts the adaptive grid system. KMA has been using the BATS model in typhoon forecasting for the past 6 years. Even though the relative performance of the BATS model compared to the other baroclinic models of KMA varies by case, BATS has proved to be quite comparable to the other models (Kwon and Shin, 1996; Kwon et al., 2001). The reason why BATS is the chosen model is that it runs every 6 hours while the other models run every 12 hours, meaning that the BATS model is quite suitable to the Administration's new statistical method, DLM (Dynamic Linear Model).

Every numerical model has its own systematic error, like a bias. In order to improve the accuracy of the prediction of typhoon direction, the systematic error of BATS forecasts should be eliminated. Figure 1 shows the track forecast errors of the BATS model for Winnie (9713) for each 6-hour time step. The errors are inclined toward the east side of the best track data. These plots are for track forecasting error $(x_{t,\mathrm{B}} - x_t,\ y_{t,\mathrm{B}} - y_t)$ for each 6-hour-interval forecast where (x_t, y_t) is the location (latitude and longitude) of the best track data, and $(x_{t,\mathrm{B}}, y_{t,\mathrm{B}})$ is for BATS forecasts. (0, 0) indicates an exact forecast.

In order to remove the systematic error of BATS forecasts, the dynamic linear model (DLM) is applied using the BATS forecasts and the best track data as inputs. DLM is useful for forecasting events that move dynamically. For three decades, DLM has been developed as a Bayesian forecasting technique by many authors including West and Harrison (1997). The DLM is a generalization of the Kalman filter model. The variances are assumed to be constant in the Kalman filter model, however these change dynamically in DLM. Sohn and Kim (2001) successfully applied the DLMs to the mid-term prediction of daily max/min temperatures. The parameters in DLM move dynamically using numerical model forecasts and observations, and then the DLMs remove the systematic error by the Bayesian forecasting technique. We expect the DLM to be useful for the improvement of the accuracy of typhoon track predictions.

*E-mail: ktsohn@pusan.ac.kr

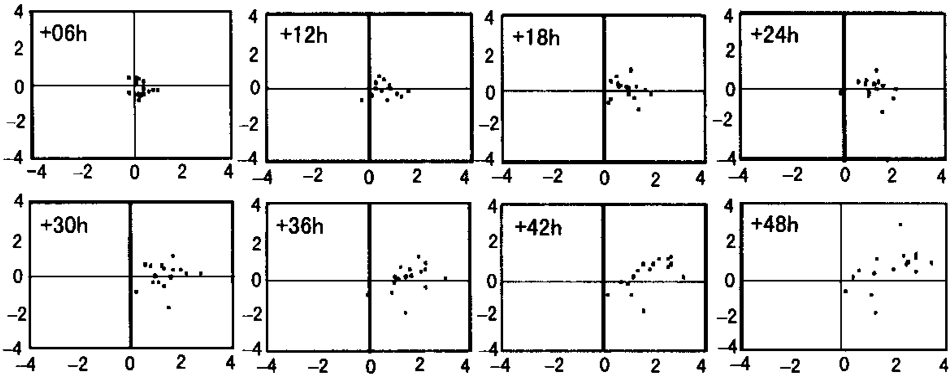

Fig. 1. Track forecast error of BATS for Winnie (relative latitude vs relative longitude in degree).

2. Data

Two datasets for DLM are used for our study: (1) best track data (RSMC, Tokyo) and (2) BATS forecasts (KMA) for typhoons occurring in 1997 and 2000. The first set contains 6-hour-interval data with typhoon-id, time, latitude, longitude, etc. The second is also 6-hour-interval data with typhoon-id, time, current location (latitude and longitude), and forecasts on latitude and longitude up to +60 hours.

3. Dynamic linear model for track forecasting

3.1 *Dynamic linear model*

DLM consists of two equations, the state equation and the output equation given by

$$Y_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t + v_t, v_t \sim N(0, V) \quad \text{(output equation)}$$

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \boldsymbol{w}_t \sim T_{n_t-1}(0, \boldsymbol{W}_t) \quad \text{(state equation)}$$

$$\boldsymbol{\theta}_0|D_0 \sim T_{n_0}(\boldsymbol{m}_0, \boldsymbol{C}_0) \quad V|D_0 \sim IG\left(\frac{n_0}{2}, \frac{n_0 s_0}{2}\right),$$

where Y_t is an observation at time t, $\boldsymbol{F}_t$ is an input vector (including BATS forecasts, the previous observations, and the previous DLM forecasts), $\boldsymbol{\theta}_t$ is the state (coefficient vector that changes dynamically), v_t is the output error term which has a normal distribution with mean 0 and variance V, which has an inverse gamma distribution $IG(\alpha, \beta)$ with two parameters α and β, $\boldsymbol{w}_t$ has a multivariate T distribution with mean vector 0 and variance-covariance matrix $\boldsymbol{W}_t$, v_t is independent of $\boldsymbol{w}_t$, and $\boldsymbol{m}_t$ and $\boldsymbol{C}_t$ are the mean vector and covariance matrix of $\boldsymbol{\theta}_t$ respectively.

The reasons to consider the DLM for forecasting typhoon tracks using BATS forecasts are as follows: (1) DLM, as a physical-statistical model, is useful to eliminate the systematic error of numerical model forecasts, (2) the parameters of DLM are changed dynamically based on Bayesian concepts, (3) the algorithms of the updating procedure and forecasting procedure are easy to program, and (4) BATS has the highest accuracy on short-term prediction among the numerical models used by KMA.

Because the typhoon tracks are 2-dimensional vectors (latitude and longitude), the application of the bivariate DLM seems to be well suited. The bivariate DLM, however, still needs the strong condition without which we cannot make useful results. With comparing univariate DLM outputs to bivariate DLM outputs, we decided to apply the univariate DLM on latitude and longitude separately.

3.2 *Updating procedure*

In order to generate the forecasts using the above DLM, parameters (V, $\boldsymbol{W}_t$, $\boldsymbol{m}_t$, $\boldsymbol{C}_t$, n_t, S_t) should be determined dynamically via the updating procedures given below. See West and Harrison (1997) for detail. Figure 2 shows us how to process the DLM.

Repeat the following steps with varying $t = 1, 2, \cdots$.

[Step 1] Posterior distribution of $\boldsymbol{\theta}_{t-1}$ (the results of the previous procedure):

$$\boldsymbol{\theta}_{t-1}|D_{t-1} \sim T_{n_{t-1}}(\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}),$$

$$V|D_{t-1} \sim IG(n_{t-1}/2, n_{t-1}S_{t-1}/2)$$

where D_t is the total information up to time t.

[Step 2] Compute the prior distribution of $\boldsymbol{\theta}_t$ and the one-step-ahead-forecast:

$$\boldsymbol{\theta}_t|D_{t-1} \sim T_{n_{t-1}}(\boldsymbol{a}_t, \boldsymbol{R}_t),$$

$$Y_t|D_{t-1} \sim T_{n_{t-1}}(\boldsymbol{f}_t, Q_t)$$

where

$$a_t = m_{t-1},\ f_t = F_t' a_t,$$

$$R_t = C_{t-1} + W_t,\ Q_t = F_t' R_t F_t + S_{t-1}.$$

[Step 3] Compute the posterior distribution of θ_t, given

$$Y_t : \theta_t | D_t \sim T_{n_t}(m_t, C_t),$$

where

$$m_t = a_t + A_t e_t,\ C_t = \left(R_t - A_t A_t' Q_t\right) S_t / S_{t-1},$$

$$e_t = Y_t - f_t,\ A_t = F_t R_t / Q_t,$$

$$V | D_t \sim IG(n_t/2, n_t S_t/2),\ n_t = n_{t-1} + 1,$$

$$S_t = S_{t-1} + \left(e_t^2/Q_t - 1\right) / S_{t-1} / n_t$$

and

$$D_t = \{D_{t-1}, Y_t\}.$$

The discount factor δ, $0 < \delta \leq 1$, is considered in order to determine W_t. Let $R_t = C_{t-1}/\delta$. Then $W_t = C_{t-1}(1-\delta)/\delta$. Varying δ from 0.01 to 1.00, we generate the forecasts from DLM with $W_t(\delta)$ and calculate the RMSE (root mean square error) for each δ. Among them, we determine the optimal δ such that the RMSE is minimized.

3.3 *Procedure for developing DLM for predicting typhoon tracks*

Since we focus on the track forecasting of a new typhoon, there are some problems so that the above updating algorithm cannot be used directly. (1) Because an optimal discount factor is determined from the whole track data, we have difficulty in deciding the optimal value of the discount factor for a new typhoon. (2) The effects of the initial values of (m_0, C_0, n_0, S_0) should be considered. In general, DLM is robust to initial values (that is, the effect of bad-guessing the initial values is short-lived). But the lifetime of a typhoon is very short and the effect of initial values may remain in all forecasts. To overcome these problems, we consider the following strategies. (1) For the prediction of new typhoon tracks, we take the average value of the optimal discount factors from past typhoons. (2) In order to reduce the effect of initial values, backcasting is considered. Backcasting means backward forecasting. Before generating DLM forecasts, we should get the data from the first 8 period (up to 48 hours) of observations and BATS forecasts after new typhoon occurs. Then we generate 10 backcasts of observations and BATS forecasts using simple linear regression models separately, and then attach those backcasts to the beginning of the 8 observations and BATS forecasts separately. Using this dataset, the generation of DLM forecasts is performed using the updating and forecasting procedure given previously. Hence, the DLM gives us the track forecasts for the 9th location. We expect that this backcasting strategy can reduce the effect of the initial values. Figure 3 outlines this procedure briefly.

According to the above strategies, a two-stage procedure is considered. The first is for the optimal DLMs with different optimal discount factors from the whole track data of each typhoon, and the second is for the new typhoon using the average discount factor. When the new typhoon occurs, we use only the second procedure for predicting the typhoon track.

The algorithm of the first procedure is as follows.

[Step 1] Generate backcasts using simple linear regression models.

[Step 2] Select initial values of (m_0, C_0, n_0, S_0) randomly.

[Step 3] Varying the components of F_t, perform the following steps.

[3.1] Changing δ from 0.01 to 1.00 by steps of 0.01, perform the updating procedure and generate the one-step-ahead forecast f_t.

[3.2] Determine the optimal discount factor for a given F_t which minimizes RMSE.

[Step 4] Determine the optimal components of F_t which minimize RMSE.

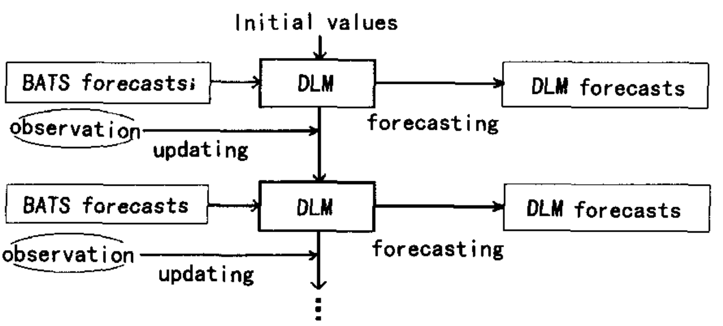

Fig. 2. Procedure for generating forecasts.

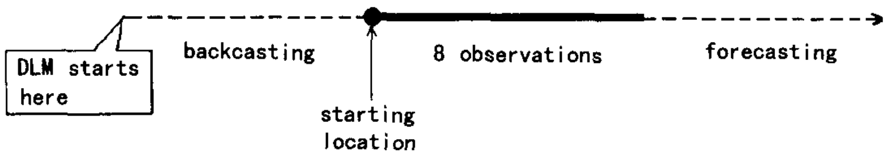

Fig. 3. Backcasting and forecasting procedure.

For each typhoon, DLM is applied and estimated using the above algorithm and then the averages of the optimal discount factors are computed.

The algorithm of the second procedure (a new typhoon) is as follows.

[Step 1] Generate backcasts using simple linear regression models.

[Step 2] Select initial values of $(\boldsymbol{m}_0, \boldsymbol{C}_0, n_0, S_0)$ randomly.

[Step 3] Generate DLM forecasts from DLM with the optimal $\boldsymbol{F}_t$ and the average value of δ.

4. Results

4.1 *Estimation of DLM*

Using the algorithm of the previous section the track forecasting is performed. For the first procedure, (1) with the first 8 periods (up to 48 hours) of observations and BATS outputs, backcasting is performed using simple linear regression for each typhoon occurring in 1997 and 2000, (2) optimal discount factors and components of $\boldsymbol{F}_t$ in the DLMs are determined for each typhoon, and (3) average values of the discount factors for each prediction are computed.

With the optimal components of $\boldsymbol{F}_t$, the average discount factors (obtained from the first procedure) and the same initial values ($\boldsymbol{m}_0 = (0.01, \cdots, 0.01)'$, $C_0 = \mathrm{diag}\,(0.001)$, $n_0 = 1$, $S_0 = 1$), the second procedure is performed for 8 typhoons (Rosie (9709), Tina (9711), Winnie (9713), Oliwa (9719), Bolaven (0006), Jelawat (0008), Saomai (0014), Yagi (0019); 77 cases). Table 1 gives the summary of the estimated optimal DLMs for Winnie, as a new typhoon, for comparison with BATS forecasts. The table shows (1) contains that (1) forecast duration time(hour), (2) components of $\boldsymbol{F}_t$, (3) optimal discount factors, and (4) forecast error (km) between forecasts (from DLM and BATS) and best tracks. For example, y_t(current track) and $\hat{y}^{(6)}_{t,\mathrm{BATS}}$ (+6h forecast generated from BATS) are used as the optimal components of $\boldsymbol{F}_t$ for the +6h forecast. Two discount factors, 0.81 and 0.84, are used for longitude and latitude respectively, and distance from best tracks is reduced (from 62.33 km to 44.69 km).

For the +12h forecast, three values are used as inputs: current location, DLM forecast for the +6h forecast at time t, and the BATS forecast for the +12h forecast at time t. Table 1 shows that the average accuracy of the forecasting is improved by about 54.3% for Winnie.

Figure 4 shows two kinds of plots, BATS forecasts (left) and DLM forecasts (right) for Winnie (9713) and Yagi (0019). The x-axis is for longitude and y-axis for latitude. The long lines are for best track and short lines are for model forecasts. In the case of Winnie, the two plots show that the BATS forecasts have an eastward forecasting error whereas the DLM removes the systematic error successfully. In the case of Yagi, the two plots show that the BATS forecasts have a southwestward forecasting error and again, the DLM reduces the forecasting error of BATS.

5. Concluding remarks

In this paper DLMs are applied to the prediction of typhoon tracks in order to eliminate the systematic error. BATS forecasts and best track data of typhoons occurring in 1997 and 2000 are used. As an initial step, backcasting using a simple linear regression model is considered, and the optimal DLM, which minimizes the value of RMSE, is estimated for each typhoon. For the track forecasting of a new typhoon, the average of the optimal discount factors and the optimal components of $\boldsymbol{F}_t$ are determined from past typhoons. In summary, we can say that the DLM is useful for the reduction of the systematic error of the BATS model.

Giving a better initialization of the typhoon structure should improve the accuracy of the typhoon forecast. The modelers do their best to improve the performance of the numerical models. Even so, the numerical model does have a systematic bias. Our statistical correction technique is still useful even for the best numerical model.

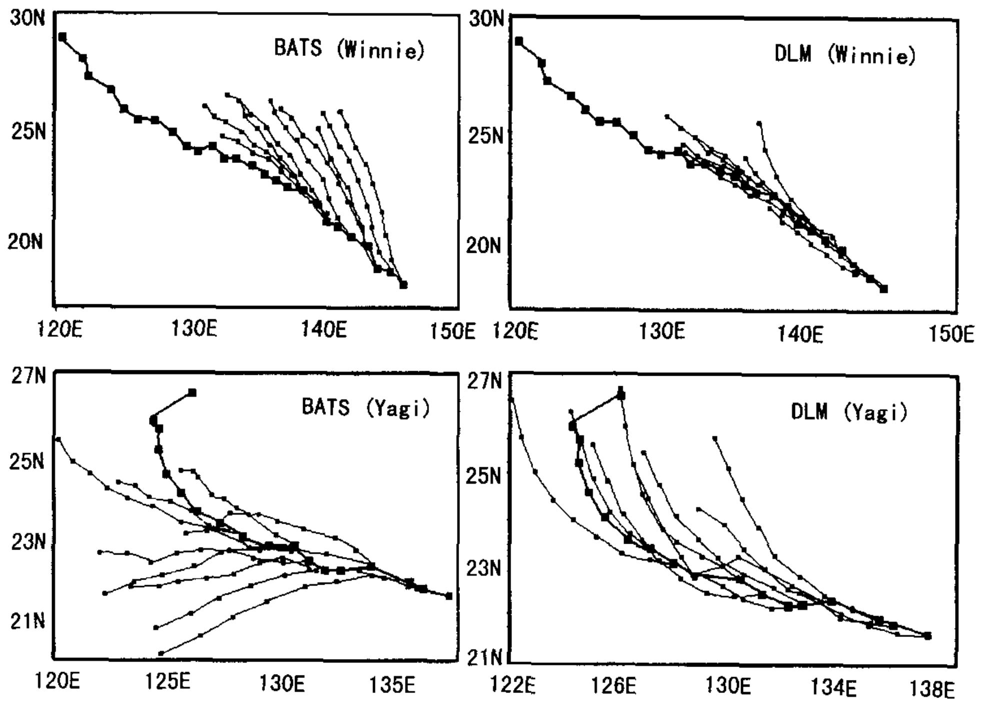

Fig. 4. Track forecasts of BATS and DLM for Winnie and Yagi. Thick and long lines are the best track and the other lines are the model forecasts (latitude N vs. longitude E).

Table 1. Summary of estimated DLMs and track error for Winnie

Forecast duration time (hour)	Components of $F(t)$	Optimal discount factor		Forecast error (km)	
		Longitude	Latitude	BATS	DLM
+6	$y_t, \hat{y}^{(6)}_{t,\mathrm{BATS}}$	0.81	0.84	62.33	44.69
+12	$y_t, \hat{y}^{(6)}_{t,\mathrm{BATS}}, \hat{y}^{(12)}_{t,\mathrm{BATS}}$	0.85	0.87	92.53	58.80
+18	$y_t, \hat{y}^{(12)}_{t,\mathrm{BATS}}, \hat{y}^{(18)}_{t,\mathrm{BATS}}$	0.91	0.82	127.68	74.31
+24	$y_t, \hat{y}^{(18)}_{t,\mathrm{BATS}}, \hat{y}^{(24)}_{t,\mathrm{BATS}}$	0.92	0.92	149.73	85.33
+30	$y_t, \hat{y}^{(24)}_{t,\mathrm{BATS}}, \hat{y}^{(30)}_{t,\mathrm{BATS}}$	0.90	0.86	180.56	98.38
+36	$y_t, \hat{y}^{(30)}_{t,\mathrm{BATS}}, \hat{y}^{(36)}_{t,\mathrm{BATS}}$	0.95	0.90	210.73	110.66
+42	$y_t, \hat{y}^{(36)}_{t,\mathrm{BATS}}, \hat{y}^{(42)}_{t,\mathrm{BATS}}$	0.87	0.89	234.43	126.34
+48	$y_t, \hat{y}^{(42)}_{t,\mathrm{BATS}}, \hat{y}^{(48)}_{t,\mathrm{BATS}}$	0.85	0.91	263.82	142.19

Table 2. Comparison of BATS and DLM forecast errors for 8 typhoons (77 cases) in 1997 and 2000

Forecast Duration time	06h	12h	18h	24h	30h	36h	42h	48h
DLM (km)	47.84	80.65	110.52	131.84	153.26	170.73	191.64	213.81
BATS (km)	52.16	85.02	116.56	146.46	181.41	220.42	265.10	302.44
Reduction Rate of Error (%)	8.28	5.14	5.18	10.0	15.52	22.54	27.71	29.30 (cases:77)

Our statistical model is aimed at the operational forecasts of KMA. In the operation of tropical cyclone track forecasting, one does not know whether a current tropical cyclone will move northward, westward, or go over an anomalous track. This statistical model (dynamic linear model) takes care of all the numerical model traits, i.e., all the biases of the numerical models, whether the current target typhoon moves northward or westward, etc.

In further work, (1) we will generalize the theory of the bivariate DLM with weaker conditions, and (2) for super-ensemble DLM forecasting, the characteristics of several numerical models should be analyzed for using their forecasts together in the DLM.

Acknowledgments. This work is supported by the project "A study on improving forecast skill using a supercomputer" of Meteorological Research Institute, KMA, 2001.

REFERENCES

Kwon, H. J., and K.-S. Shin, 1996: A barotropic typhoon track prediction model by adaptive grid. *11th Conf. on Numerical Weather Prediction*, Aug. 19-23, 1996, Norfolk, VA., 404–406.

Kwon, H. J., S.-H. Han, and K.-S. Shin, 2001: Operational prediction of typhoon track at KMA using barotropic adaptive-grid typhoon simulation (BATS) model. *J. Meteor. Soc.* Japan, **79**, 1157–1168.

Sohn, K. T., and 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 the Korean Meteorological Society*, **37** (1), 3–20. (in Korean)

West, M., and J. Harrison, 1997: *Bayesian Forecasting and Dynamic Models* (2nd ed.), Springer, New York. 680 pp.