The Korean Communications
in Statistics Vol. 12 No. 2, 2005
pp. 295-304

Statistical Correction of Numerical Model Forecasts for Typhoon Tracks[1]

Keon Tae Sohn[2]

Abstract

This paper concentrates on the prediction of typhoon tracks using the dynamic linear model (DLM) for the statistical correction of the numerical model guidance used in the JMA. The DLM with proposed forecast strategy is applied to reduce their systematic errors using the latest observation. All parameters of the DLM are updated dynamically and backward forecasting is performed to remove the effect of initial values.

Keywords : typhoon track forecast, DLM, systematic error

1. Introduction

The objective of this study is to improve the prediction of typhoon tracks by the statistical correction of numerical model forecasts. Nowadays numerical models have been increasingly improved. According to the report of the fifth WMO international workshop on tropical cyclones(IWTC-V, 2002), numerical prediction systems of tropical cyclone tracks have improved by approximately 25% for 24 hours and 50% for 72 hours since 1995. They say that this has been due to (i) the more and better use of observational data to define the large-scale environment and outer structure of storms, (ii) the continuous model development, (iii) higher resolution of forecast systems, and (iv) the use of some form of vortex specification for short lead times. The report shows the current status of numerical track prediction as Table 1. They suggest that a major issue of further improvements in short-term forecasts will be the mesoscale analysis, initialization, a synthetic vortex and the performance of physical parameterization.

Numerical models have still some systematic errors despite their remarkable improvements. For example, Figure 1 shows that the track forecast errors of the TYM, the typhoon model

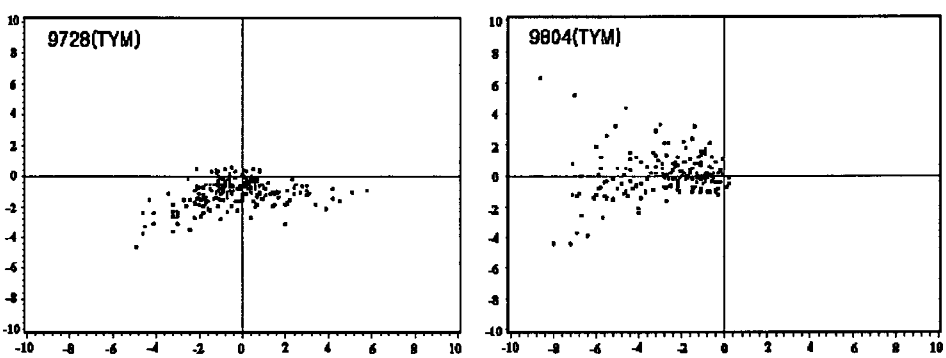
1) This study was supported by the project "A study on improving forecast skill using a supercomputer" funded by the Meteorological Research Institute, KMA, in 2002.
2) Professor, Department of Statistics, Pusan National University, Busan, 609-735, Korea.
E-mail : ktsohn@pusan.ac.kr

used in the Japan Meteorological Administration(JMA), for Paka(9728) and Rex(9804) have some systematic errors. The origin in plots point indicates the exact forecast case. The TYM forecasts for Paka incline toward south and have horizontal dispersion and those for Rex have westward error. There is no doubt that the elimination of their systematic errors can improve the accuracy of prediction.

<Table 1> Mean track forecast errors for the Northwest Pacific up to September 2002 for Joint Typhoon Warning Center(JTWC) consensus and for individual models

Forecast duration (hours)	24	48	72	96	120
Consensus Error (km)	128	216	289	387	503
Range of Individual model errors (km)	140~181	240~303	335~459	435~527	592~666

※ Models: NOGAPS, UKMO, JMA, NCEP, MM5, GFDN, COMPS (from WMO IWTC-V Report)



<Figure 1> Track forecast errors of the TYM for Paka(left) and Rex(right): (relative latitude × relative longitude)

For this purpose, we usually considered two kinds of statistical methods; MOS(model output statistics) and KF(Kalman filter). The MOS techniques can reduce the systematic error using some statistical relationship between observations(predictands) and numerical model forecasts of predictors. Glahn and Lowry(1972) considered the characteristics of numerical model and developed the forecast model. So many authors have considered the MOS for the prediction of temperature and precipitation(e.g., Lemcke and Kruizinga 1988, Ross 1989, Kok and Kruizinga 1992). The MOS however has some disadvantages that the MOS models need long period data and should be re-estimated whenever numerical models are improved or changed.

The KF, proposed by Kalman(1960), needs short period data and maintain the dynamic and nonlinear aspects through a set of statistical equations that provides an efficient solution of the least-square method. The KF is very powerful to estimate past, present and future states, moreover even when the precise nature of modeled system is unknown. Since 1980 the KF have been applied for the analysis and the prediction of the meteorological and climatological data. Ghil and Malanotte-Rizzoli(1991), Simonsen (1991), Persson(1991), Ross and Strudwicke (1994) and Homleid(1995) applied the KF to the prediction of temperature or the probability of precipitation. Verron *et al.* (1999) considered the extended Kalman filter for the purpose of assimilating observations into a high-resolution nonlinear numerical model of the tropical Pacific Ocean.

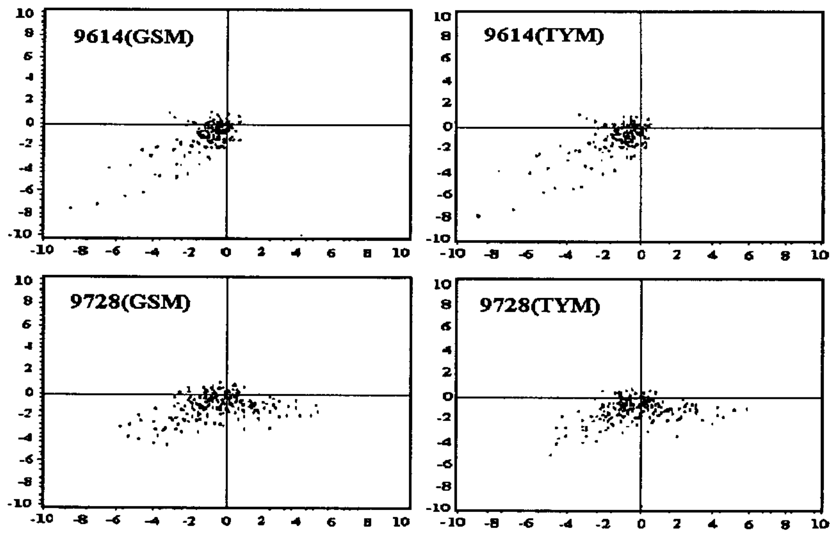
This study concentrates on the typhoon track forecasting using the Bayesian version of Kalman filter model, called the dynamic linear model(DLM). 'Bayesian version' means that all distributions of parameters in the DLM are updated dynamically by the new inputs. There are a few studies of the DLM in meteorological applications. Sohn, Kwon and Suh(2003) applied the DLM to typhoon track forecasting as the statistical correction of BATS. Sohn, Nah and Seo(2003) applied the DLM to the 3-hour-interval prediction of ground-level temperatures in South Korea. Sohn(2004) applied the DLM to the 3-hour weather forecast of six basic climate factors.

In this study the DLM, as a statistical correction of JMA numerical model forecasts, is applied to reduce the systematic errors of the TYM(JMA typhoon model) and the GSM(JMA global model) in order to improve their predictability for typhoon tracks. The data for our study is presented and the systematic errors of the TYM and the GSM are compared in Section 2. The brief introduction to the DLM and the forecast strategy will be presented in Section 3 and the results of the statistical correction using the DLM will be presented in Section 4.

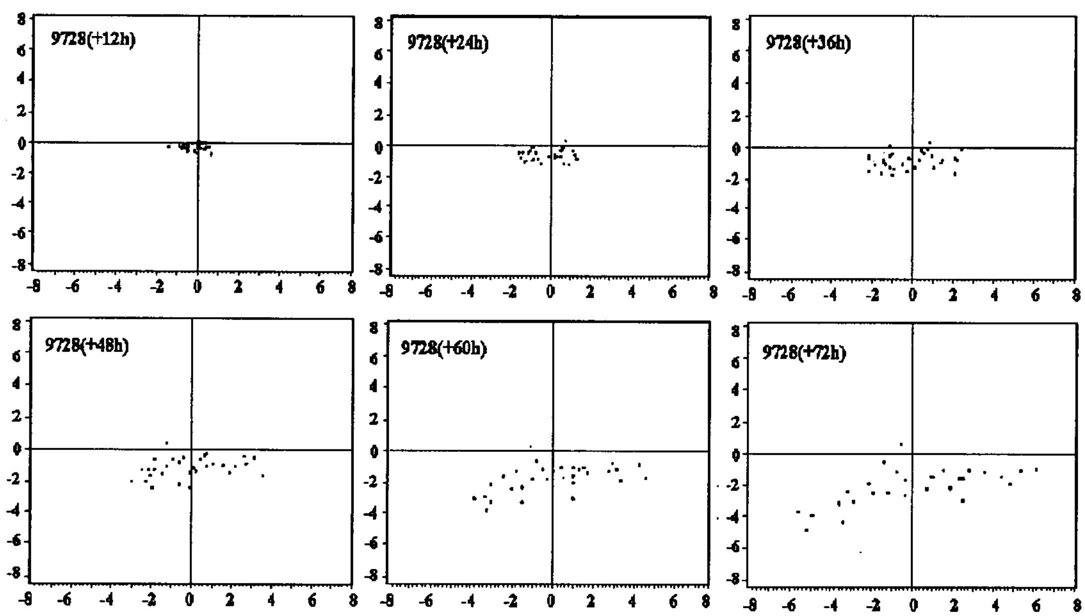
2. Data and Systematic errors of TYM and GSM

Two kinds of data are used for our study. (1) numerical model outputs (TYM forecasts and GSM forecasts) and (2) observed location (best track data from RSMC, Tokyo) for typhoons occurred in year 1996~1999 and 2002. The TYM forecasts and the GSM forecasts have the same formats that consist of 6-hour-interval forecasts up to 84 hours. The GSM is run at 00UTC and 12UTC and the TYM is run at 06UTC and 18UTC. The best track data, as observations, also consist of 6-hour-interval data. Among typhoons in the above period 9 typhoons, which have relatively long lifetime, are selected for our study. They are Orson(9614), Winnie(9713), Keith(9725), Paka(9728), Rex (9804), Rammasun(0205), Chataan (0206), Halong(0207) and Fengshen(0209).

The systematic errors(i.e., track forecast errors) of the TYM (or the GSM) have different shapes for each typhoon. However, according to the comparison of systematic errors of the TYM with those of the GSM, two models have very similar types of systematic errors for each typhoon. For instance, Figure 2 shows that the shapes of the GSM forecast errors and the TYM forecast errors for Orson and Paka are very similar to each other. And their forecast errors have similar shapes for each forecast time. For example, Figure 3 shows that the shapes of track forecast errors for Paka are much alike for each forecast time.



<Figure 2> GSM and TYM forecast errors for Orson(9614) and Paka(9728) : (relative latitude × relative longitude)



<Figure 3> Track forecast errors of GSM and TYM for Paka(9728) for 12h, 24h, 36h, 48h, 60h and 72h forecast duration: (relative latitude × relative longitude)

Therefore we decided to consider the GSM and the TYM together like one model with a view of systematic errors. In this study, we simply call the JM instead of the GSM and the TYM. That is, the JM generates the 6-hour-interval forecasts up to 84 hours for every 6 hours(00UTC, 06UTC, 12UTC and 18UTC). The dynamic linear model is applied to the statistical correction of the JM track forecasts for every 6 hours.

3. Dynamic linear model and Forecast strategy

3.1 Dynamic linear model

The DLM is a Bayesian Kalman filter model. In general, the DLM consists of two equations; the state equation and the output equation given by

(1) (state equation) $\theta_t = \theta_{t-1} + w_t$ where $\theta_t \sim T_{n(t)}(m_t, C_t)$ and $w_t \sim T_{n(t)-1}(0, W_t)$

(2) (output equation) $Y_t = F_t'\theta_t + v_t$ where $v_t \sim N(0, V_t)$ and $V_t \sim IG(n_t/2, n_t s_t/2)$

where θ_t is the dynamic coefficient vector at time t, m_t and C_t are the mean vector and the covariance matrix of θ_t, the evolution error vector w_t has the multivariate T distribution with degree of freedom $n_t - 1$, mean vector 0 and covariance matrix W_t, Y_t is a univariate observation (typhoon track) at time t, F_t is the input vector which consists of observation and model forecasts, the observational error V_t has an inverse gamma distribution with two parameters n_t and s_t, v_t and w_t are independent to each other.

To generate the DLM forecasts, parameters, $\{m_t, C_t, n_t, s_t\}$, should be dynamically updated via the following updating procedures. For the more information on the DLM, see West and Harrison(1997).

[Step 1] Select randomly initial values $\{m_0, C_0, n_0, s_0\}$.

[Step 2] (the updating procedure) Varying $t = 1, 2, \cdots$, perform the following procedure.

[2.1] Compute the mean a_t and variance R_t of θ_t given $\{m_{t-1}, C_{t-1}, n_{t-1}, s_{t-1}\}$.

$a_t = m_{t-1}$, $R_t = C_{t-1} + W_t$ (by the state equation)

[2.2] Compute the mean f_t and variance Q_t of Y_t given $\{m_{t-1}, C_{t-1}, n_{t-1}, s_{t-1}\}$.

$f_t = F_t' m_{t-1}$, $Q_t = F_t' R_t F_t + s_{t-1}$ (by the output equation)

[2.3] Update $\{m_t, C_t, n_t, s_t\}$, given Y_t.

$$A_t = F_t' R_t / Q_t,\ e_t = Y_t - f_t,\ n_t = n_{t-1} + 1,\ s_t = s_{t-1} + \frac{s_{t-1}}{n_t}\left(\frac{e_t^2}{Q_t} - 1\right),$$

$$m_t = a_t + A_t e_t,\ C_t = \frac{s_t}{s_{t-1}}(R_t - A_t A_t' Q_t)$$

In addition, the covariance matrix W_t of evolution error at time t may be estimated by

$W_t = C_{t-1} \cdot (1-\delta) / \delta$ where δ, $0 < \delta \le 1$, is called the discount factor. Varying the discount factor from 0.01 to 1.00, we find the optimal discount factor which minimizes the RMSE (the square root of mean squares of e_t). And then the DLM forecasts are generated by the following forecast strategy and the optimal discount factor.

3.2 Forecast Strategy

For the track forecasting of new typhoon, there are several problems. Firstly, the parameters of DLM should be estimated separately for each typhoon because the systematic errors of one typhoon are different from those of the others. Hence the all parameters of the DLM should be estimated dynamically and automatically for each individual typhoon. Secondly, the effect of initial values should be considered. It is well known that the effect of bad-guessed initial values of the DLM is short-lived from the empirical experiments. However the life of typhoon is short and then the effect of initial values may remain in all the track forecasts. To reduce the effect of initial values, backward forecasting(backcasting) by the simple extrapolation is brought in. Thirdly, the DLM consists of the same structure for all typhoons. That is, the DLM consists of the same components of input vector and the same updating and forecasting algorithm. Fourthly, all processes, collecting data, updating the DLM and forecasting the DLM forecasts, should be automatically executed.

The time flow of the typhoon track forecast system in the KMA(Korea Meteorological Administration) should be considered. In the case of 00UTC forecast, we can get the typhoon location at 00UTC about 01UTC. And GSM forecasts from 00UTC run are collected about 03UTC. And KMA forecasters announce the track forecasts at 02UTC. Therefore we have no choice but to use the latest information(the observed typhoon location at 00UTC) and JM forecasts of the previous 18UTC run.

The procedure of generating DLM forecasts is shown in Figure 4. (1) When a new typhoon occurs, the first observation and the first JM forecasts are collected. (2) Backcasting is performed using the first observation and the first JM forecasts. (3) Varying the value of discount factor, the DLM generates dynamically its forecasts from the beginning of backcasted locations and find the optimal discount factor, which minimizes the RMSE. (4) And then we can generate the new DLM forecasts up to 84 hours for the first DLM forecasts. (5) We collect the new observed track and new JM forecasts after 6 hours. (6) Varying discount factor, the DLM again starts to generate its forecasts from the beginning of backcasted locations and find the optimal discount factor again. (7) And then we generate the new DLM forecasts. Repeat this procedure from (5) to (7).

In short, under the above forecast strategy, the DLM forecasts are generated by the below algorithm.

[Step 1] Select initial values.

[Step 2] Varying the component of input vector among the input vector space, perform the followings

[2.1] Perform backcasting using the first observation and the first JM forecasts.

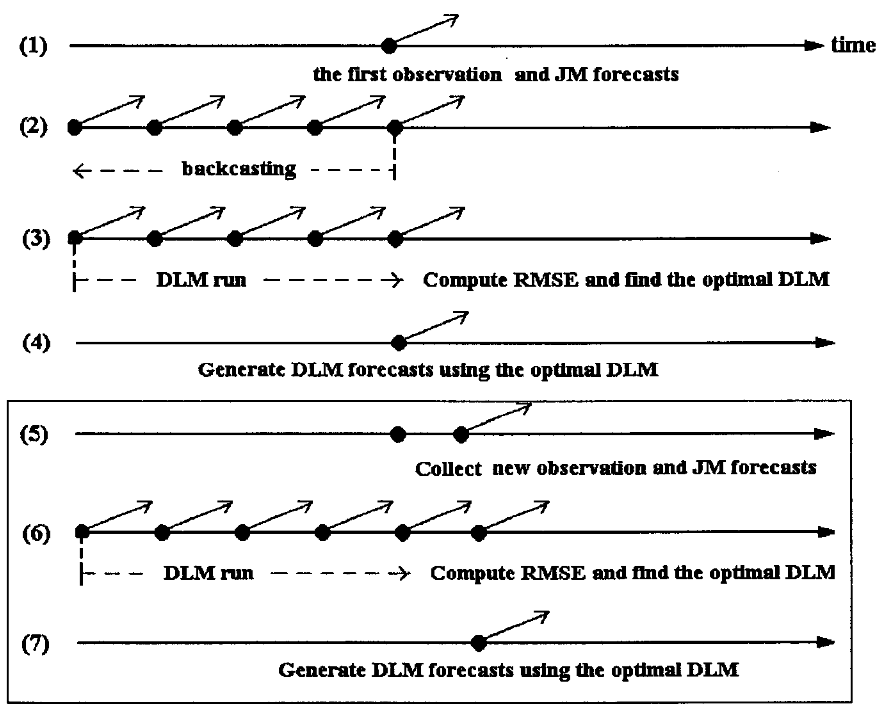
[2.2] Varying discount factor from 0.01 to 1, perform followings.

[2.2.1] Perform the updating algorithm.

[2.2.2] Compute the RMSE.

[Step 3] Determine the optimal discount factor and the optimal input vector that minimize the RMSE.

[Step 4] Generate the DLM forecasts via the optimal DLM.



<Figure 4> Procedure of generating DLM forecasts

4. Results

The DLM with the forecast strategy in the previous section is applied to 9 typhoons. After heuristic experiments, the optimal F_t, which minimizes the RMSE, is determined as follows.

$$F_{t+k} = (1,\ JM^{k+6}_{t-6},\ DLM^{k+6}_{t-6},\ JM^{k}_{t-6}),\quad k=6,12,\cdots,84$$

where 1 is for an intercepts, JM_{t-6}^{k+6} is the JM forecast for time $t+k$ at $t-6$ run, JM_{t-6}^{k} is the JM forecast for time $t+k-6$ at $t-6$ run, DLM_{t-6}^{k+6} is the previous DLM forecast for time $t+k+6$ at time $t-6$. In the case of 12UTC forecast, the DLM uses the JM forecasts of 00UTC run and the previous DLM forecasts at 00UTC.

Table 2 shows the comparison of the RMSE of the JM and the DLM for each forecast duration time and each typhoon. The error reduction rates R, as a measure of improvement, is defined by

$$R = 1 - \frac{RMSE(DLM)}{RMSE(JM)}$$

<Table 2> Comparison of JM and DLM

FSCT time (hour)	Total (358 cases)			9614 (45 cases)			9713 (41 cases)			9725 (38 cases)			9728 (52 cases)		
	RMSE (km)		R (%)	RMSE (km)		R (%)	RMSE (km)		R (%)	RMSE (km)		R (%)	RMSE (km)		R (%)
	JM	DLM		JM	DLM		JM	DLM		JM	DLM		JM	DLM	
6	68.55	47.78	30.3	49.19	32.43	34.0	50.96	49.64	2.5	36.88	32.02	13.1	45.75	39.69	13.2
12	101.06	81.26	19.6	80.41	65.61	18.4	90.39	81.656	9.6	54.35	49.90	8.1	74.01	67.13	9.3
18	129.08	104.89	18.7	106.70	92.30	13.4	119.56	103.75	13.2	73.64	62.27	15.4	101.40	92.02	9.2
24	151.58	127.62	15.8	127.88	120.24	5.9	139.14	124.70	10.3	91.72	76.79	16.2	127.67	117.58	7.9
30	172.37	152.23	11.7	146.54	149.43	-1.9	155.64	148.94	4.3	115.08	90.32	21.5	153.99	142.05	7.7
36	192.81	179.16	7.1	168.32	181.22	-7.6	173.20	168.58	2.6	131.93	110.67	16.1	177.44	165.64	6.6
42	215.27	205.58	4.5	198.90	215.63	-8.4	188.15	187.37	0.4	152.63	130.98	14.1	195.73	187.92	3.9
48	240.32	230.89	3.9	232.56	251.88	-8.3	204.77	202.49	1.1	179.05	153.71	14.1	215.27	213.99	0.5
54	265.12	256.01	3.4	264.03	289.96	-9.8	217.59	216.40	0.5	208.70	174.30	16.4	241.96	244.80	-1.1
60	287.61	281.65	2.1	299.41	328.80	-9.8	225.18	228.24	-1.3	237.12	198.99	16.0	269.32	278.28	-3.3
66	313.68	308.34	1.7	332.61	369.08	-10.9	238.36	240.65	-0.9	278.38	222.03	20.2	302.15	312.48	-3.4
mean	194.31	179.58	10.81	182.41	190.60	1.3	163.90	159.31	3.8	141.77	118.36	15.6	173.15	169.24	4.6

FSCT time (hour)	9804 (46 cases)			0205 (27 cases)			0206 (40 cases)			0207 (32 cases)			0209 (37 cases)		
	RMSE (km)		R (%)	RMSE (km)		R (%)	RMSE (km)		R (%)	RMSE (km)		R (%)	RMSE (km)		R (%)
	JM	DLM		JM	DLM		JM	DLM		JM	DLM		JM	DLM	
6	63.28	42.24	33.2	48.90	55.05	-12.6	193.40	88.56	54.2	71.17	53.39	24.9	58.21	40.20	30.9
12	103.83	85.85	17.3	67.58	63.53	5.9	236.14	140.45	40.5	112.91	101.56	10.0	86.33	74.45	13.7
18	137.17	117.26	14.5	75.65	71.16	5.9	271.98	161.06	40.7	154.26	141.38	08.3	114.44	99.37	13.1
24	165.32	148.23	10.3	76.02	85.11	-11.9	294.48	173.03	41.2	195.05	177.51	08.9	137.75	120.55	12.4
30	193.51	181.25	6.3	79.99	112.77	-40.9	308.86	183.87	40.4	228.55	212.13	07.1	158.53	145.00	8.5
36	227.39	216.22	4.9	88.81	150.69	-69.6	320.34	201.18	37.1	259.99	249.54	04.0	175.11	168.13	3.9
42	263.57	250.41	4.9	106.62	184.77	-73.2	336.01	215.40	35.8	287.87	288.97	-0.3	194.34	191.34	1.5
48	296.37	286.22	3.4	135.86	210.78	-55.1	352.86	220.73	37.4	318.71	327.71	-2.8	215.59	214.93	0.3
54	336.79	324.00	3.7	151.28	230.74	-52.5	367.34	226.94	38.2	351.44	363.93	-3.5	234.40	237.27	-1.2
60	374.05	361.11	3.4	161.16	252.55	-56.7	374.45	231.15	38.2	379.99	403.10	-6.0	253.81	258.02	-1.6
66	413.26	399.69	3.2	170.50	278.39	-63.2	382.00	240.87	36.9	419.08	444.37	-6.0	271.55	274.33	-1.0
mean	234.05	219.32	9.6	105.67	154.14	-38.5	312.53	189.39	40.1	252.64	251.24	4.1	172.73	165.78	7.3

The first block is for the averaged RMSE of JM forecasts and DLM forecasts for 9 typhoons. The averaged RMSE of the JM and that of the DLM are 194.31km and 179.58km respectively. The average of error reduction rates for all duration times is about 10.8%. The error reduction rate is 30.3% for 6 hour-forecast and 15.8% for 24 hour-forecast. That is, the DLM improves the JM forecasts and is more useful for the short-term prediction. The other blocks are for individual typhoons from Orson(9614) to Fengshen(0209). Table 2 says that the DLM improves the JM forecasts more for the short-term prediction except Rammasun(0205)

5. Concluding Remarks

The DLM is applied to the typhoon track forecasting in order to improve the predictability of numerical models, the GSM and the TYM. In this study the DLM works as a statistical correction model of the numerical model forecasts, that is, a post-processing model or a physical-statistical model. Because the structure of the proposed DLM for typhoon track forecasting is very simple, the generating algorithm of the DLM forecasts is easy to be programmed. To reduce the effect of initial values, backcasting via simple extrapolation is considered. All parameters including the optimal discount factor are updated dynamically because the numerical model has different aspects of systematic errors for each typhoon.

Though the DLM is more useful for the short-term prediction, it is needed to consider the another inputs in F_t for the long-term prediction. The proposed forecast strategy and the DLM can be extended to the other applications like the super-ensemble modeling, the typhoon intensity, the central pressure and related synoptic factors.

References

[1] Ghil, M. and P. Malanotte-Rizzoli (1991). Data assimilation in meteorology and oceanography, *Advanced Geophysics*, 33, 141-266.

[2] Glahn, H. R. and D. A. Lowry (1972). The use of model output statistics (MOS) in objective weather forecasting, *Journal of Applied Meteorology*, 11, 1203-1211.

[3] Homleid, M (1995). Diurnal correction of short-term surface temperature forecasts using the Kalman filter, *Weather and Forecast*, 10, 689-707.

[4] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems, *Transaction of the ASME Journal of Basic Engineering*, 35-45.

[5] Kok, K. and S. Kruizinga (1992). Updating probabilistic MOS equations, *Proceedings of the 12th Conferences on Probability and Statistics in Atmospheric Sciences*, American Meteorologist Society, 62-65.

[6] Lemcke, C. and S. Kruizinga (1988). Model output statistics (three years of operational experience in the Netherlands), *Monthly Weather Reviews*, 116, 1077-1090.

[7] Persson, A. O. (1991). Kalman filtering - A new approach to adaptive statistical interpretation of numerical meteorological forecasts, *WMO TD*, 421, xx27-xx37.

[8] Report of the fifth WMO International Workshop on Tropical Cyclone (2002). *Tropical Meteorology Research Programme Report Series*, Report No. 67, Cairns, Australia.

[9] Ross, G. H. (1989). Model output statistics using an updateable scheme, *Proceedings of the 11th Conferences on Probability and Statistics in Atmospheric Sciences*, American Meteorologist Society, 93-97.

[10] Ross, G. H. and C. C. Studwicke (1994). Logistic regression using a Kalman filter within an updateable MOS forecasting system, *Proceedings of the 13th Conferences on Probability and Statistics in Atmospheric Sciences*, American Meteorologist Society, 204-209.

[11] Simonsen, C. (1991). Self adaptive model output statistics based on Kalman filtering, *WMO TD*, 421, xx33-xx37

[12] Sohn, K. T. (2004). Statistical correction of Numerical model forecasts for 3-hour-forecast system, *Journal of the Korean Data Analysis Society*, Vol. 6, No. 2, 453-464

[13] Sohn, K. T., J. Kwon and A. S. Suh (2003). Prediction of Typhoon tracks using dynamic linear models, *Advanced Atmospheric Sciences,* Vol. 20, No. 3, 379-384.

[14] Sohn, K. T., D. K. Rha and Y. K. Seo (2003). The 3-hour-interval prediction of ground-level temperature in South Korea using Dynamic linear models, *Advanced Atmospheric Sciences*, Vol. 20, No. 4, 575-582.

[15] Verron, J., L. Gourdeau, D. T. Pham, R. Murtugudde and A. J. Busalacchi (1999). An extended Kalman filter to assimilate satellite altimeter data into a nonlinear numerical model of the tropical Pacific Ocean: Method and validation, *Journal of Geophysical Research*, 104, C3, 5441-5458.

[16] West, M. and J. Harrison (1997). *Bayesian forecasting and Dynamic models (2nd ed.)*, Springer-Verlag, New York.

[Received December 2004, Accepted April 2005]