A MODEL OF INTEGRATED RANDOM INITIAL VALUES AND RANDOM FORCING AND ITS PRELIMINARY EXPERIMENT

Hu Zengzhen (胡增臻)*, *Shi Jiu'en* (史久恩)

Beijing Institute of Meteorology, Beijing

and *Zhang Jianzhong* (张建中)

Computing Centre, Academia Sinica, Beijing

Received May 29, 1989

ABSTRACT

A simple quasi-geostrophic barotropic vorticity equation model is used as the dynamic frame of the model in this paper. Considering that there are many random errors in model's initial values of meteorological data, and that it is not perfectly complete about model's physical processes (for example, take no account of the interaction between atmosphere and underlying surface, radiation, etc.), we add the random forced term to the model and use the Monte-Carlo method with random initial values. A statistical-dynamic integrated model is thus built up, and a numerical forecasting experiment of 500hPa monthly mean height field of January 1983 has been carried out. The experiment result proves that the forecasting result of the model, considering random forcing and random initial values at the same time, is better than that by the pure dynamic model, the random initial value model and the random forced model.

Key words: Monte-Carlo forecast, random forcing, random initial values, numerical experiment

I. INTRODUCTION

In the 1940s and 1950s, Blinova et al. (Dobylishman, 1959) worked out a series of long-range numerical forecasting and experiment studies using simple linear and nonlinear models. Although the numerical forecasts at that time had a long way to go for the operational demand, they gave many valuable results. For example, it pointed out that the forecasting results calculated by using mean states were better than by using instantaneous states. Afterwards, in about twenty years, there was a little progress in long-range numerical forecasting. From the 1970s, with increasing understanding of the process of ocean-atmosphere interaction and the undeterminism of the atmosphere movement, there raised hope of doing long-range weather forecasting with numerical methods. The study of long-range numerical forecasting entered a flourishing period, and statistical-dynamic integrated methods were even more spectacular.

It was recognized that a more exact portrayal of the rule of atmospheric movement could be made by recognizing the duality of both undeterministic and deterministic features (Lorenz, 1963). In order to describe the twofold nature of the atmospheric movement, meteorologists used statistical-dynamic integrated ways. Epstein (1969) advanced a moment approach which

---

* Present address: Department of Geophysics, Peking University, Beijing

can predict the probability distribution of every atmospheric movement variable. In this method, the dependent variables are various step-moments of the atmospheric movement variables. If the equations are nonlinear, the variation of the low step-moments is related to high step moments. Thus the equations can be closed, and the high step moments must be truncated. As a result, errors are brought into forecasts. Subsequently, there were many studies intending to reduce these errors caused by truncation of high step moments. The real case simulation proved that this method is of positive significance for improving simulation results. However, this method is very complicated, the amount of calculation is very large, and the errors brought by truncation of high step moments of nonlinear equations could not be overcome. In addition, the solution is generally very difficult to find out except that very simple models are adopted. Therefore, it is impossible to be further developed and used in the operation. In order to overcome the difficulty in the moment method, Leith (1974) put forward a statistical-dynamic integrated Monte-Carlo approximate forecasting approach, and also pointed out through simulation that it could significantly reduce the amount of calculation since the ensemble forecast is sufficient when using eight individual forecasts. In the same paper, Leith used a two-dimensional model of homogeneous and isotropic turbulence to prove that the ensemble forecast is the optimum forecast in the sense of the least square. This work laid a foundation for the further development of this approach. Seidman (1981) presented an averaging technique for random perturbation initial values to make long-range numerical weather forecasts. The main points of this method are: first, putting various random perturbations in the observed initial states and giving a series of perturbation initial states; second, using these perturbation initial states to make forecasts respectively through a dynamic model. The forecast needed is the ensemble of all these individual forecasts. The effect of this ensemble forecast technique on the predictability was investigated through the simulation study of a 3-layer general circulation model. It was found that the averaging forecast techniques could increase the predictable time, when comparing with the individual forecasts or climate mean. After that, many people used the Monte-Carlo approach to do experiments of forecast study, and the results were satisfactory (ECMWF, 1980).

Many achievements have been made in the response study of the atmosphere to forced sources (such as topography, heat sources and sinks). Pitcher (1977), by use of a statistical-dynamic way, studied the matter of putting forced term in the spectral forecast equation, and pointed out that the increased errors (external errors), caused by simplifying the model, must be precisely considered. In order to simulate these error sources, a random forced term was put into every spectral equation, and this was used as a parameterization of the increased errors. The experiment results were satisfactory.

In this paper, based on the work of random initial value model and random forced model (Hu et al., 1990; Zhang et al., 1990), the random forced term is put into the model and random initial value is used at the same time. In this way we have done a numerical forecasting experiment of 500hPa monthly mean height field of January 1983, and the results obtained are compared with those by some other models. This work proves that the forecasting result of the model simultaneously considering random forcing and random initial values is an optimum forecast. The dynamic frame of our model is a quasi-geostrophic barotropic vorticity equation (Hu et al., 1990). In this paper, we will first directly describe the statistic part of the model, then give the result of numerical experiment and make the comparison with the results by some other models.

II. A MODEL OF INTEGRATED RANDOM INITIAL VALUES AND RANDOM FORCING

1. *Generation of Random Initial Values*

The generation of random initial values is the key to the random initial value simulation experiment. It is assumed that the observational value is taken to be point A in the phase space. It is thus demanded that all the generated initial values are around point A and constitute a point collection in the phase space which obeys normal distribution (Fig. 1). But

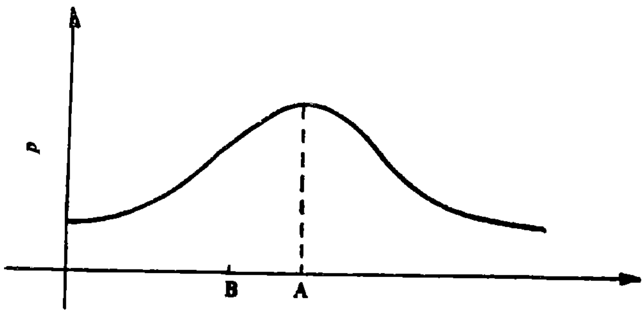

Fig. 1. Sketch figure of possible distribution (p) of all the initial values, observations A and B are points in the phase space.

the real initial state may be point B. The random initial value can be formed by superposing various perturbation values on the observational value. The specific method of generating random initial values is as follows: Assume that $AAK_n^m(t_0)$ represents the observational value at time t_0, and that random initial value is

$$AK_n^m(t_0) = AAK_n^m(t_0) + \sigma \cdot \mu, \tag{1}$$

where σ is the mean square deviation of errors, μ is a pseudorandom number which obeys the distribution of $N(0,1)$, m is zonal wavenumber, and $(n-m)$ is meridional wavenumber. Now the problem of generating the initial values becomes the problem of how to generate σ and μ.

(1) *Generation of the mean square deviation of random errors*

The errors in the meteorological data include systematic and random errors. The systematic errors are caused by the trend errors of instruments used in satellites, and/or large-scale data blank areas in the ocean and the Antarctic. The random errors are produced by the individual measurement of the observation system (for example, satellite observations and routine soundings), interpolation of the observational data into networks and some other factors. In this paper, the system errors are assumed to be zero, namely, only the random errors are considered. Therefore the σ in Eq. (1) becomes the mean square deviation of random errors. Since in practical observations, there is only one observational value at a certain point and time, it is impossible to objectively and accurately decide the mean square deviation of random errors from the observed data. In some countries (for example, in the Meterological Office of United Kingdom), σ is taken as a constant. Because anomaly data are used in this study, we may consider that σ is directly proportional to the observed anomalies. Thus, we can infer the specific proportion relationship through numerical experiments and available data on this subject. Presently, the observation system in continents has a space scale of several hundred kilometers, and a routine observation frequency of twice a day. The observational errors are usually as follows: temperature 1—2°C, wind speed 3—4 m/s and pressure 3—4 hPa (Seidman, 1981). However, the range of errors becomes much bigger in the ocean areas and areas with scarce stations. On the basis of the above analysis, together with the analysis of numerical simulation results, we may conclude that the mean square deviation of random errors is taken as 10% of observed anomalies, namely,

$$\sigma = AAK_n^m(t_0) \times 10\% \ .$$

(2) *Generation of the pseudorandom numbers which obey the distribution of $N(0,1)$*

There are many methods which can generate pseudorandom numbers (The Probability-Statistical Group, 1979). In this paper, we use the multiplicative congruential method to generate the pseudorandom numbers which obey the uniform-distribution. The pseudorandom numbers generated by this method, have many advantages, such as long cycle, excellent statistical characteristics, rapid generating speed, less computer memories occupied, etc. First, the multiplicative congruential method is used to generate the pseudorandom numbers r_n which obey the uniform-distribution in an interval [0, 1], namely, positive integer x_0 is taken as the initial value and another positive integer λ as multiplier, then recurrence formula

$$x_{n+1} = \lambda x_n (\mathrm{mod} M)$$

is used to generate numerical series $\{x_n \mid n = 1,2,\dots\}$. Through the standardization treatment, we get the pseudorandom numbers

$$r_{n+1} = x_{n+1}/M ,$$

where x_0, as the initial value to produce the pseudorandom numbers, is an odd number in an interval $[0,2^{31}]$; $\lambda = 5^{13}$; $M = 3^{31}$; and n is an ordinal number of pseudorandom number series. Then the composite random sampling is used of normal distribution sampling, namely, resolve the normal probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \qquad (-\infty < x < +\infty)$$

into the sum of four probability density functions

$$f(x) = p_1 \times f_1(x) + p_2 \times f_2(x) + p_3 \times f_3(x) + p_4 \times f_4(x) ,$$

where the values of probability $p_i (i = 1, 2, 3, 4)$ are 0.8638, 0.1107, 0.0228002039 and 0.0026997961 respectively, and p_i satisfy the conditions

$$0 < p_i < 1 \text{ and } \sum_{i=1}^{4} p_i = 1 \ .$$

Through the above sampling, we obtain the pseudorandom numbers, i.e. the μ in Eq. (1), which obeys the $N(0,1)$ distribution. Similarly, putting various μ into Eq. (1), we can get various values of random initial values.

2. *Generation of Random Forced Term*

Atmospheric motion is affected by forced sources, such as oceans, mountains, snow cover, ice caps, soil moisture, etc. It is due to the interaction between atmosphere motion and these forced sources that a variety of motions can be produced in the earth's atmosphere. How to correctly consider these forced sources in the numerical forecasting model is still a problem. In the past decades, many meteorologists carried out a series of parameterization experiments and yielded many achievements (Pitcher, 1977), thus laying the foundation for further understanding the role played by the forced sources in atmospheric motion. However, there is still much work to do in order to concretely and accurately put these external forced sources into the numerical forecasting model. In this paper, we put forward a new method which considers the forced sources in combination with the historical data. In the method, first we derive the forced time series through the forecasting equation, then we distinguish, estimate and simulate *ARMA* (*K,L*) (*ARMA:* autoregressive-moving averages; see below Eq. (5)) statistical model of the forced time series, and get the forced term. Subsequently as random forcing, the forced term is put into the equations and the random initial values at the same time are used to do Monte-Carlo simulation. Finally, forecasts are obtained by the integrated model of random initial values and random forcing.

(1) *Generation of the random forced time series*

In order to simulate the forced time series, we must first generate it. Assume the forecasting equation of the model is

$$\frac{dAK_n^m(t_i)}{dt_i}=F_n^m(t_i) \quad . \tag{2}$$

Due to a series of simplifications in the model, which are somewhat different from the real atmospheric motion, when the historical data are substituted into forecasting Eq. (2), the equality generally does not hold. Assuming the remainder is $f_n^m(t_i)$, then we have

$$f_n^m(t_i)=\frac{dAK_n^m(t_i)}{dt_i}-F_n^m(t_i) \quad .$$

Turn into difference form

$$f_n^m(t_i)=\frac{AK_n^m(t_i)-AK_n^m(t_{i-1})}{\Delta t}-F_n^m(t_i) \quad , \tag{3}$$

where $i=1,2,3,\ldots,31$, and $\Delta t=t_i-t_{i-1}$. Based on observational data available and spectral expansion coefficients, Δt is taken as one year. Substituting the historical data into Eq. (3), we can obtain the forced time series

$$f_n^m(t_1),\ f_n^m(t_2),\cdots,f_n^m(t_{31}) \quad . \tag{4}$$

(2) *Distinguishing and estimating the random forced time series*

Having obtained the forced time series, we distinguish and estimate the *ARMA*(*K*, *L*) model:

$$x_t-b_1x_{t-1}-\cdots-b_Kx_{t-K}=a_t-d_1a_{t-1}\cdots-d_La_{t-L} \quad , \tag{5}$$

where a_t is white noise which obeys $N(0,\sigma^2)$. About calculation of σ please refer to Xiang et al. (1986). First, examine whether the random forced time series (4) is a stationary time series or not. If not, we assume for convenience that no forcing exists there. The practical test proves that only 7% are non-stationary among all time series, so that this simplified treatment can not bring about obvious errors. If it is a stationary time series, then we calculate the value of *BIC*:

$$BIC=\ln(ss)+(K_p+L_q)\ln N/N \quad ,$$

where ss is the maximum likelihood estimate of the square deviation of the errors in the *ARMA* (K_p, L_q) model, and N is series length. According to the *BIC* criterion, we decide which is the best in all the below models:

$AR(1)$, $AR(2)$, $\cdots$, $AR(5)$, $MA(1)$, $MA(2)$ and $ARMA$ (1,1).

After the best model is found out, we estimate the parameters in the model, so that we obtain the basic model a certain random forced time series obeys. Repeating all the procedures to various (n,m), we get the best models of all the (n,m).

(3) *Simulation of the random forced time series*

After the best models are found out, through simulation using these best models we can get a random forced time series

$$g_n^m(t_1),\ g_n^m(t_2),\cdots,g_n^m(t_N),$$

where t_N is the length of a random forced series obtained by the simulation. The time interval of the random forced time series here is one year. However, the time interval of the random forced time series needed for Monte-Carlo simulation is assumed as

$$Q_n^m(\tau_1),\ Q_n^m(\tau_2),\cdots,\ Q_n^m(\tau_N),$$

where $\Delta\tau=\tau_i-\tau_{i-1}$ is the time step of integration in our paper, and $\Delta\tau=3$ hours. The problem arises: there are different sampling time intervals between the random forced time series $Q_n^m(\tau_i)$ which the Monte-Carlo simulation needs and the random forced time series $g_n^m(t_i)$ which is obtained from the historical data. Until now, the problem about how to match two time series with different sampling time intervals has not been solved theoretically. Thus we must do a series of parameterized trial experiments. At last we decide that the forced term is put into the equation at each step of integration only for the second ten days of a month, and is suitably weakened. We use $g_n^m(t_i)/4$ as the random forced time series $Q_n^m(\tau_i)$ needed for the integration. The approximately treated random forced term is put into the equation and the random initial values are used, then the Monte-Carlo simulation is made through eight integration calculations. Finally, we get the forecasting ensemble under the influence of a series of random initial values and random forced sources. This forecasting ensemble is the forecasting result of the model combining the random initial values with random forcing.

III. NUMERICAL FORECASTING EXPERIMENT AND COMPARISON WITH FORECASTS OF OTHER MODELS

In order to compare the forecasts of several models, we use an exactly same difference analogue, time smooth, initial data and treated method as in the previous papers (Hu et al., 1990; Zhang et al., 1990), and use the model for doing the same case, 500 hPa monthly mean height anomaly field forecasting experiment of January 1983 (Fig. 2).

Comparing the analysis of Fig. 2 with the forecasts of the random initial value model (Fig. 3), the random forced model (Fig. 4), the pure dynamic model (Fig. 5) and the actual observations (Fig. 6) (Hu et al., 1990; Zhang et al., 1990), we can find that the forecasts of the random initial value and random forced integrated model (Fig. 2) have a marked improvement in relation to those by other models (Figs. 3—5), especially for the forecasted intensity of anomaly centers, and it is the best forecast of the four models. Through careful analysis, we can find that it improves the forecast of the random initial value model. for example, the intensity of positive anomaly centre in the northeast of China has a little reduction, and is closer to the actual observation; and the orientation of axis of this positive anomaly centre changes from northwest-southeast in the forecast of the random initial value model to approximately north-south, and is closer to the actual observation.

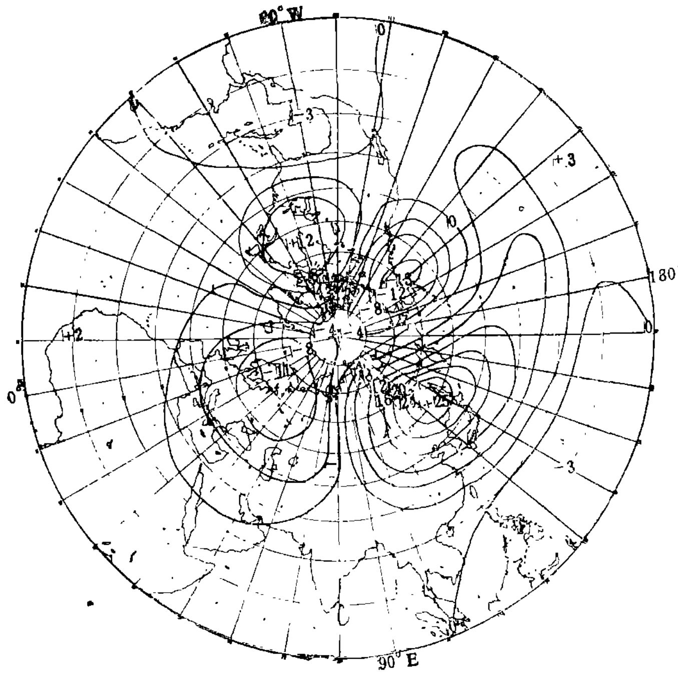

Fig. 2. Forecasted 500 hPa mean height anomaly field of January 1983 by the random initial value and random forced integrated model (unit: 10 gpm).

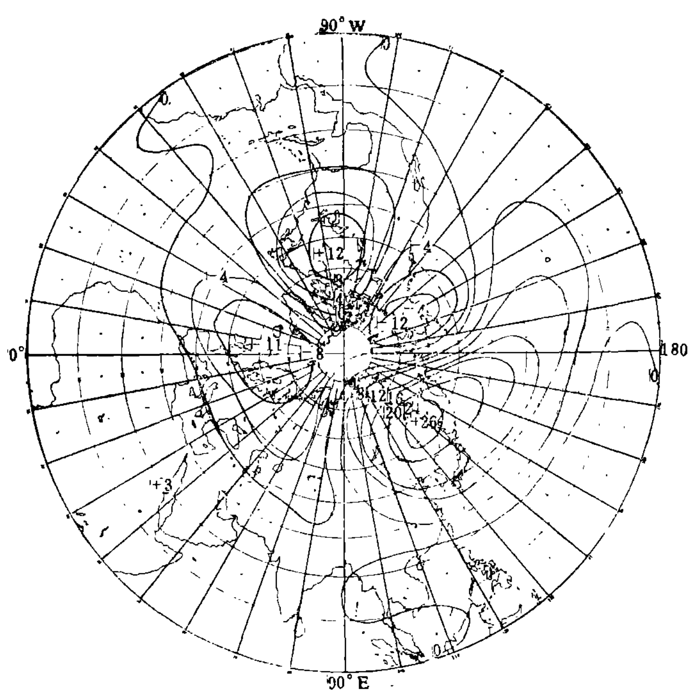

Fig. 3. As in Fig. 2, but by the random initial value model.

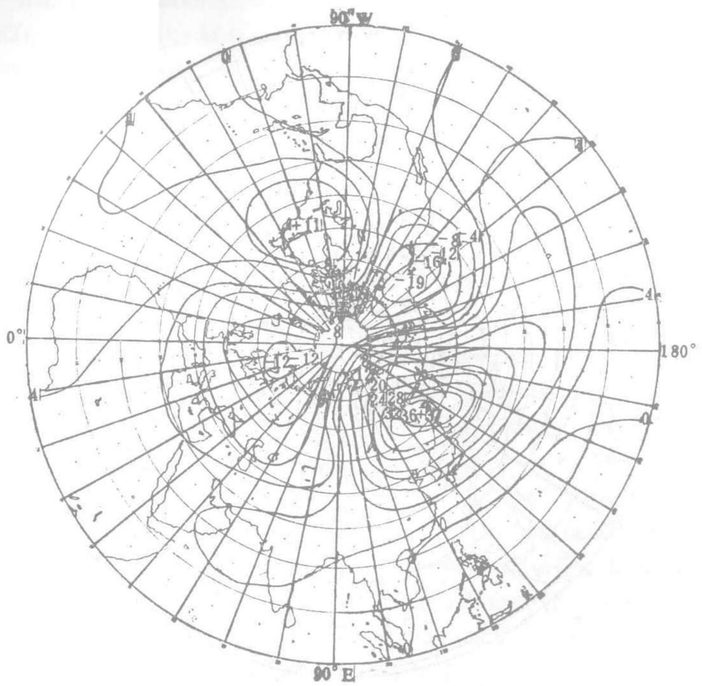

Fig. 4. As in Fig. 2, but by the random forced model.

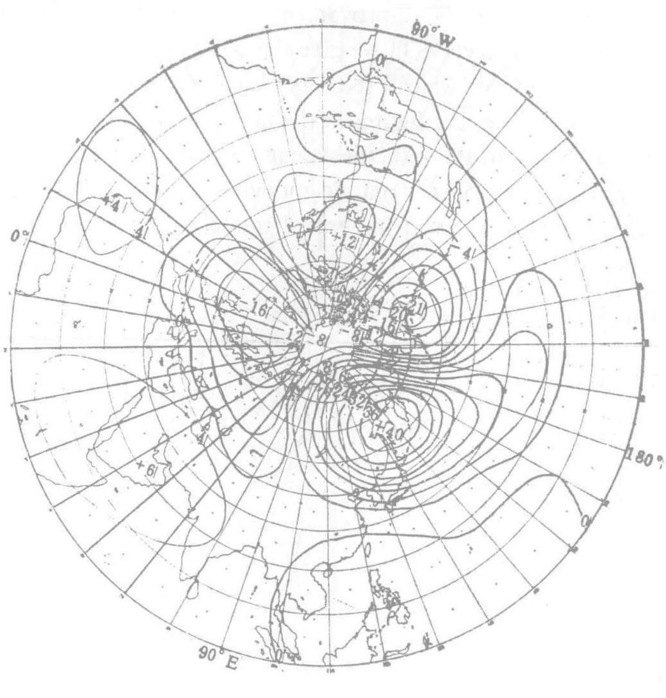

Fig. 5. As in Fig. 2, but by the pure dynamic model.

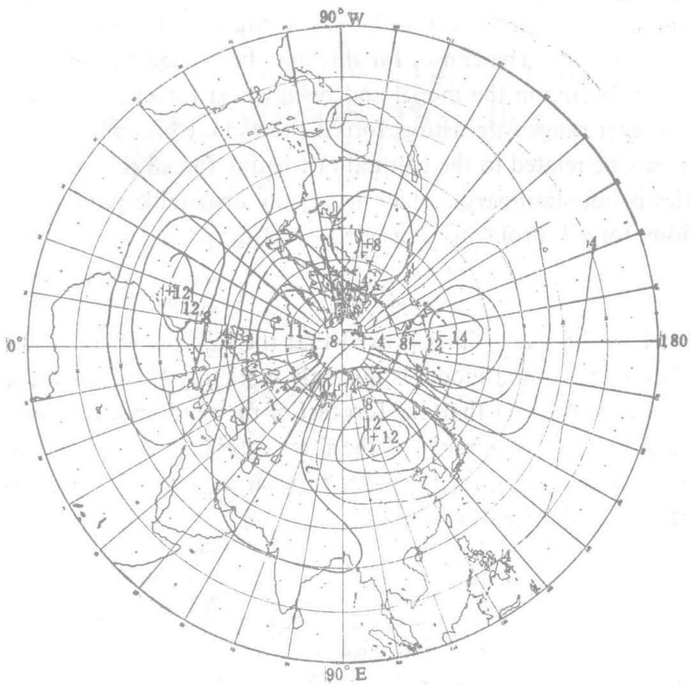

Fig. 6. The actual 500 hPa mean height anomaly field of Januany 1983 (unit: 10 gpm).

In order to objectively analyze the improvement of the several models to the pure dynamic model (only the case of January 1983), we give the comparison result of a few quantitative indexes in Table 1. From the table we can see very clearly the improvement of several models to the pure dynamic model. From the figures in the table, we know that all the models have some forecasting ability (by contrast with the linear correlation coefficient 0.12 of sustained forecast), but the forecast of the pure dynamic model is the worst, and the forecast of the random initial value and random forced integrated model is the best. This improvement is

Table 1. Comparison of Forecasting Effects of Models

Indexes / Models	Sign Correlation Coefficient R_1	Linear Correlation Coefficient R_2	Mean of Error Absolute Value E (10 gpm)
pure dynamic model	0.71	0.26	7.2
random forced model	0.73	0.26	6.7
random initial value model	0.75	0.33	5.1
random initial value and random forced integrated model	0.76	0.40	5.0
sustained forecast		0.12	

mainly due to consideration of random initial values. On the contrary, when the forced term is considered, the improved range is very little. Nevertheless, this can not declare that the forced effect is not important. The causes for this may be related to only considering the interannual scale random forcing in the model and being short forecasting time. For example, if we do seasonal or longer range forecasting, forced effects may become more clear and important, besides this, it may be related to the method of treating forced time series (especially when the forced time series is not stationary), which has many unsuitable points. After all, further study of the random forced problem is needed.

IV. SUMMARY

Through the work in this paper, it is proved that the forecast effect of the random initial value and random forced integrated model is not only markedly better than the forecast of the pure dynamic model, but also better than the forecast of the random forced model and the random initial value model. The comparison of the forecast effects shows that the statistical-dynamic integrated method used in doing long-range weather numerical forecasts has marked superiority over the pure dynamic method.

REFERENCES

Dobylishman (1959), Introduction of climate theory and long-range numerical forecast method, lectures at Central Meteorological Bureau, Beijing (recorded in Chinese).

ECMWF (1980), *Workshop on Stochastic Dynamic Forecasting*, 17—19 October 1979, pp. 1—96.

Epstein, E.S. (1969), Stochastic dynamic prediction, *Tellus*, **21**:739—759.

Hu Zengzhen et al. (1990), Numerical forecast experiment of monthly mean circulation with a random initial state model, *Collected Works of Long-range Forecast*, China Meteorological Press, Beijing, pp. 292—300 (in Chinese).

Leith, C.E. (1974), Theoretical skill of Monte-Carlo forecasts, *Mon. Wea. Rev.*, **102**:409—418.

Lorenz, E.N. (1963), Deterministic nonperiodic flow, *J. Atmos. Sci.*, **20**: 130—141.

Pitcher, E.J. (1977), Application of stochastic dynamic prediction to real data, *J. Atmos. Sci.*, **34**:3—21.

Seidman, A.N. (1981), Averaging techniques in long-range weather forecasting, *Mon. Wea. Rev.*, **109**:1367—1379.

The Probability-Statistic Group of Computer Centre of Academia Sinica (1979), *Probability Statistic Calculation*, Science Press, Beijing, 472 pp. (in Chinese).

Xiang Jingtian et al. (1986), *Dynamic Data Treatment——Time Series Analysis*, China Meteorological Press, Beijing, 433 pp. (in Chinese).

Zhang Jianzhong et al. (1990), The Monte-Carlo method of doing 500 hPa monthly mean flow forecasting by a random forced model, *Collected Works of Beijing Institute of Meteorology*, China Meteorological Press, Beijing (in Chinese).