

带随机初值和随机强迫的简单模式的 集合预报试验*

胡增臻** 史久恩 张建中

(北京气象学院) (中科院计算中心)

提 要

本文用简单的准地转正压涡度方程谱模式作为本文模式的动力框架, 考虑到作为模式初始场的气象资料中存在着大量的随机误差, 以及模式中物理过程的不完善(例如, 没有考虑大气与下垫面的相互作用、辐射等), 采用在模式中加入随机强迫项和使用随机初值的蒙特卡洛方法, 建立了一个统计动力相结合的模式, 并用此模式做了1983年1月500 hPa月平均高度场的数值预报试验。试验结果表明: 同时考虑随机强迫和随机初值的模式预报效果优于纯动力模式、随机初值模式和随机强迫模式的预报效果。

一、引 言

随着自40和50年代数值预报和数值试验研究的增多, 人们认识到大气运动存在着确定性和非确定性^[1], 只有认识了这种双重性, 才能更深刻地刻划大气运动, 为此气象学家提出用统计-动力相结合的方法去研究大气运动。1969年, Epstein^[2]提出了预报大气运动变量概率分布的矩方法。1974年, Leith^[3]为了克服矩方法中的诸多缺陷, 如截断误差大, 计算量大, 仅适用于十分简单的模式等, 提出了蒙特卡洛(Monte-Carlo)近似预报方法; 并且证实集合预报是最小平方意义上的最优预报。Leith还用实例模拟表明: 取集合预报的个数在8个左右即可, 这就极大地减少了计算量。1981年, Seidman^[4]提出了随机扰动初值预报的平均方法, 此方法是在观测的初值状态上叠加不同的随机小扰动产生一系列初值, 对这些初值的预报求集合平均得到所要的预报。并且用一个三层的大气环流模式调查了这种集合平均预报方法对可预报性的影响。在此之后, 一直有人用Monte-Carlo方法做预报试验研究^[5], 结果是令人鼓舞的。

关于大气对强迫源(如地形、热源、热汇等的)响应研究, 已取得了许多成果。1977年, Pitcher^[6]用统计-动力方法研究了在谱预报方程中加入随机强迫项的问题。指出在制做随机预报时必须严格地考虑由于模式简化导致的误差(外部误差)增长。为了模拟这些误差源, 在每一个谱方程中附加一个随机强迫项, 以此作为对误差增长的参数化, 试验结果是令人鼓舞的。

本文在分别考虑随机初值和随机强迫模式^[8,10]的基础上, 在模式中同时加入随机强迫项和使用随机初值。通过对1983年1月500 hPa月平均高度距平场的数值预报试验, 比较了几个模式的预报效果。本文模式的动力部分是准地转正压涡度方程谱模式, 以前

* 本文于1989年6月30日收到, 1989年12月19日收到修改稿。

** 现在通讯地址: 北京中关村中国科学院大气物理研究所。

已做过介绍^[7], 本文首先介绍模式的统计部分, 然后介绍模式的预报试验结果及与其它几个模式的比较分析。

二、模式的统计部分

1. 随机初值的产生

在随机初值模式试验中, 随机初值的产生十分关键。假设观测值为相空间的点 A , 我们要求产生的所有初值是围绕点 A 服从正态分布的相空间的点集(图 1), 而实际的真实的初始状态可能是点 B 。因此随机初值可以通过在观测值上叠加不同的小扰动而形成。具体产生初值的方法是: 设 $AAk_n^m(t_0)$ 为 t_0 时刻的观测值, 则令随机初值

$$AAk_n^m(t_0) = AAk_n^m(t_0) + \sigma \cdot \mu \quad (1)$$

其中:

σ 是误差的均方差。

μ 是服从 $N(0,1)$ 分布的伪随机数。

因此, 产生随机初值的问题变为如何产生 σ 和 μ 的问题。

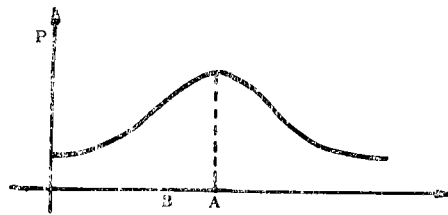


图 1 所有初值的可能分布示意图
(A 为观测点)

1) 随机误差均方差的产生

气象资料中的误差可分为系统误差和随机误差。系统误差可由多种因素产生, 例如, 人造卫星上测量要素的仪器的系统误差以及在海洋和南极地区大面积的资料空白而产生的误差。随机误差是由于观测系统(如卫星和常规探测)的个别测量而产生的。把观测资料内插到各种计算网格上时, 也会产生随机误差; 产生随机误差还有其它一些原因。在本文的研究中, 假定系统误差为零, 即只考虑随机误差, 因此方程(1)中的 σ 是随机误差的均方差。由于在实际气象观测中, 某一点在某一时刻仅有唯一的观测值, 这就不可能客观而精确地确定观测资料中的随机误差的均方差。国外(如英国气象局)的工作是将 σ 取常值。考虑到模式中使用的是距平场资料, 因此可以设想初始场中的随机误差是与距平值本身的大小成正比关系的。通过数值试验和已有的资料去推测具体的比例关系。目前的观测系统在大陆上的空间尺度是几百公里, 常规的观测是一天两次, 误差范围是^[4]: 温度 $1-2^{\circ}\text{C}$, 风速 $3-4 \text{ m/s}$, 气压 $3-4 \text{ hPa}$ 。在海洋上和一些台站稀少的地区误差范围要大得多。根据上面的分析, 结合数值模拟计算的结果分析, 确定把随机误差均方差取为距平观测值的 10% , 即:

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

2) 服从 $N(0,1)$ 分布的伪随机数的产生

产生伪随机数的方法有许多种^[8], 本文采用乘同余法产生均匀分布的伪随机数。利用这种方法产生伪随机数具有周期长、统计性质优、产生速度快和占用计算机内存少等优点。首先利用乘同余法产生 $[0,1]$ 上均匀分布的伪随机数 r_n , 即取正整数 x_0 为初值和另一正整数 λ 为乘子, 利用递推公式

$$x_{n+1} = \lambda x_n \pmod{M}$$

产生数值序列 $\{x_n | n=1, 2, \dots\}$, 经规范化处理, 给出伪随机数

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

然后利用正态分布抽样中的复合随机抽样,即将正态概率密度函数:

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

分解为四个概率密度函数的和

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

其中:

概率 $p_i (i=1, 2, 3, 4)$ 分别取值为 0.8638, 0.1107, 0.0228002039 和 0.0026997961; 满足条件:

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

经过上述抽样后,便得到服从 $N(0, 1)$ 分布的伪随机 $\tilde{\varepsilon}$, 即(1)式中的 μ 。这样将不同的 μ 代入(1)式就得到不同的随机初值。

2. 随机强迫项的产生

本文提出了一个结合历史资料考虑随机强迫的方法。此方法就是首先利用预报方程反推出强迫时间序列,然后对此时间序列进行识别、估计和模拟。将得到的强迫项作为随机强迫加入方程。

1) 随机强迫时间序列的产生

要对强迫时间序列进行模拟,首先需要产生一个强迫时间序列。设模式的预报方程为:

$$\frac{dAk_n^m(t_i)}{dt_i} = F_n^m(t_i) \quad (2)$$

由于模式存在着一系列的简化,与实际大气运动存在着差距,所以将历史资料代入方程(2),等式一般是不成立的。设其余差项为 $f_n^m(t_i)$, 则:

$$f_n^m(t_i) = \frac{dAk_n^m(t_i)}{dt_i} - F_n^m(t_i)$$

化为差分形式:

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

其中:

$i=1, 2, \dots, 31$, 代表由 1951-1982 年资料求得的强迫时间序列长度。

$\Delta t = t_i - t_{i-1}$, 是强迫时间序列的时间间隔。这样将历史资料代入方程(3)可以得到强迫时间序列:

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

2) 随机强迫时间序列的识别和估计

对强迫时间序列(4)进行识别和估计的 $ARMA(k, L)$ 模型:

$$x_t - b_1 x_{t-1} - \dots - b_k x_{t-k} = a_t - d_1 a_{t-1} - \dots - d_L a_{t-L}$$

其中:

a_t 为服从 $N(0, \sigma^2)$ 的白噪声, σ 的求法见文献^[9]。首先检验强迫时间序列(4)是否是

平稳序列,若是非平稳序列,为便于处理暂假定无强迫存在;实际检验表明,非平稳强迫时间序列仅占总序列个数的7%左右。若是平稳序列,计算BIC值:

$$BIC = \ln(ss) + (k_p + L_q) \ln N / N$$

这里 ss 为模型 $ARMA(k_p, L_q)$ 的误差方差的极大似然估计。然后利用 BIC 准则,确定下列模型中哪个是最佳模型:

$$AR(1), \dots, AR(5), MA(1), MA(2), ARMA(1,1)$$

确定出最佳模型后,估计此模型中的参数。这样就得到了某个随机强迫时间序列所符合的基本模型,对不同的 (n, m) 重复以上工作,就可以得到所有 (n, m) 的最佳模型。

3) 随机强迫时间序列的模拟

利用上面得到的最佳模型,模拟得到一个随机强迫时间序列:

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

此随机强迫时间序列的时间间隔是一年,而 Monte-Carlo 模拟需要的随机强迫时间序列,设为:

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

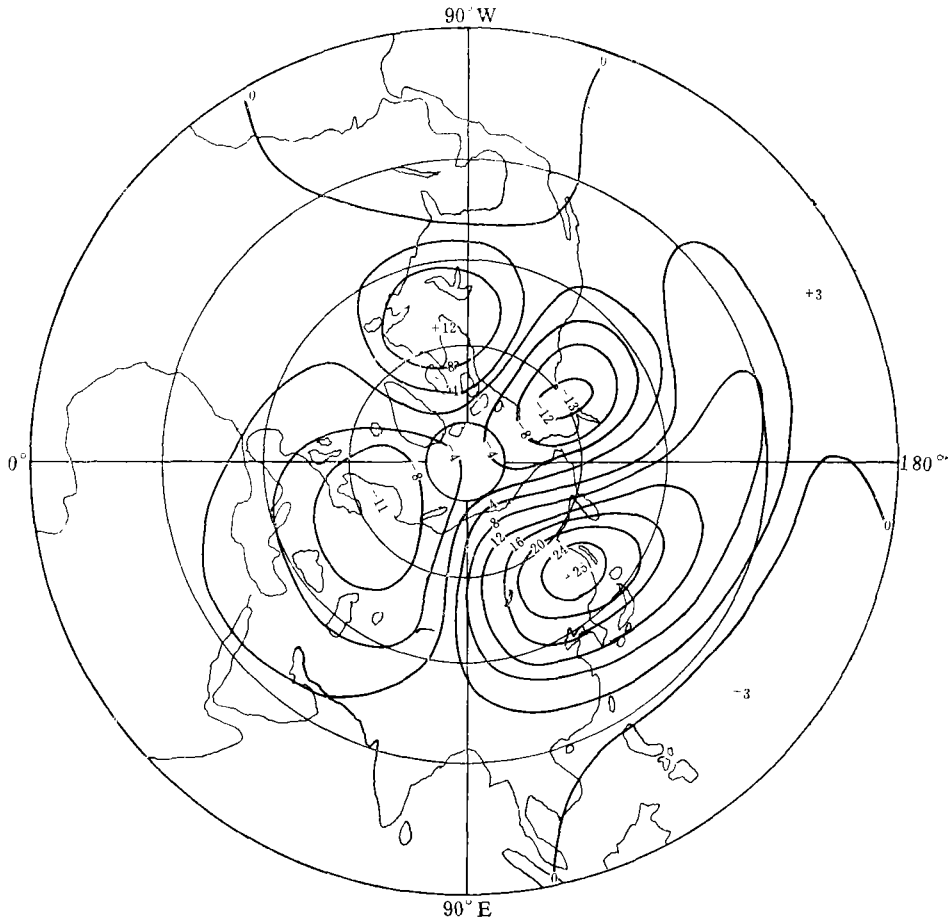


图 2 随机初值与随机强迫相结合模式的预报
(单位: dagpm, 下同)

的时间间隔 $\Delta\tau = \tau_i - \tau_{i-1}$ 取为模式积分的时间步长。在本模式中 $\Delta\tau = 3\text{ h}$ ，这就出现了 Monte-Carlo 模拟所需的随机强迫时间序列 $Q_m^m(\tau_i)$ 与由历史资料求得的随机强迫时间序列 $g_m^m(t_i)$ 具有不同的采样时间间隔。目前，还没有从理论上解决采样时间间隔不同的两时间序列之间的相互匹配问题；为此，我们进行了一系列的参数化式的处理实验。最后确定，只在中旬的积分中每步都加入强迫，并适当地将强迫给予弱化处理：取 $g_m^m(t_i)$ 的四分之一作为积分中所需的随机强迫时间序列 $Q_m^m(t_i)$ 。把近似处理后的随机强迫加入方程中，并采用随机初值，进行8次积分计算的 Monte-Carlo 模拟，最后得到在一系列随机初值和随机强迫作用下的预报集合，作为随机初值和随机强迫相结合的模式预报结果。

三、数值预报试验及与其它几个模式预报效果的对比

为了便于比较，本文采用与文献^[7]完全相同的差分格式、时间平滑、初始资料及处理方法，用本文模式做了同一个例子：1983年1月北半球500 hPa月平均高度距平场的随机初值与随机强迫相结合的预报试验(图2)。



图3 随机初值模式的预报

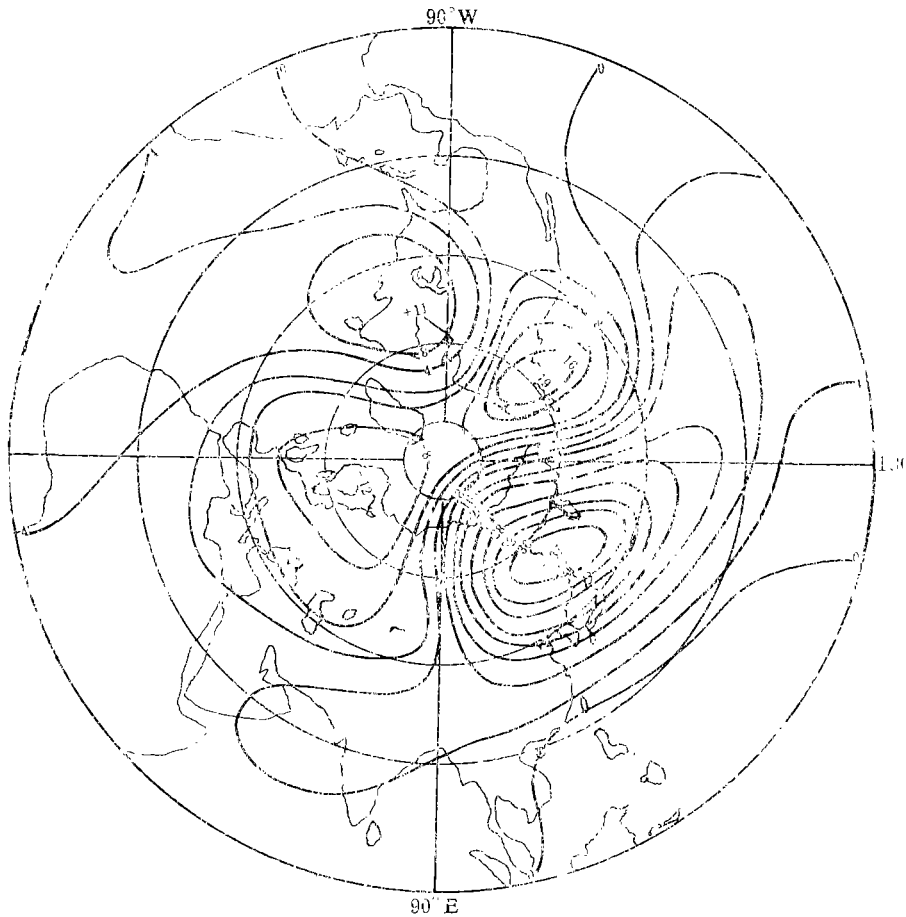


图 4 随机强迫模式的预报

将图 2 与随机初值模式(图 3)^[7]、随机强迫模式(图 4)^[10]和纯动力模式(图 5)^[7]的预报及实况(图 6)对比分析可以发现,随机初值与随机强迫相结合模式的预报不仅比纯动力预报(图 5)有明显改进,而且也比随机强迫模式预报(图 4)和随机初值模式预报(图 3)有改进,是以上几个模式中最好的预报。通过对比对几个距平中心的强度预报,可以发现,随机初值和随机强迫相结合的模式预报对纯动力和随机强迫两模式的预报的改进比较明显,故不做具体分析比较了。较仔细的分析可以发现,它对随机初值模式预报也有改进。例如,在我国东北北部的正距平中心的强度稍减小更接近实况,正距平中心轴向由随机初值模式预报的西北—东南向变为了近似的南北向,更接近于实况。

为了客观地分析以上几个模式对预报效果(仅指 1983 年 1 月份一个预报个例)的改进,下表给出了几个定量指标的对比结果。由此可以明显看出几个模式对预报效果的改进情况。从表中数据可以看出:几个模式均有一定的预报能力(相对于惯性预报的相关系数 0.12 而言),但纯动力模式预报效果最差,随机初值与随机强迫相结合模式的预报效果最好,这种改进主要是考虑了随机初值的结果。考虑随机强迫虽然对预报也有所改进,但改进的幅度不大,这并不能说明强迫作用不重要,很可能与本模式只考虑了年际尺度的随

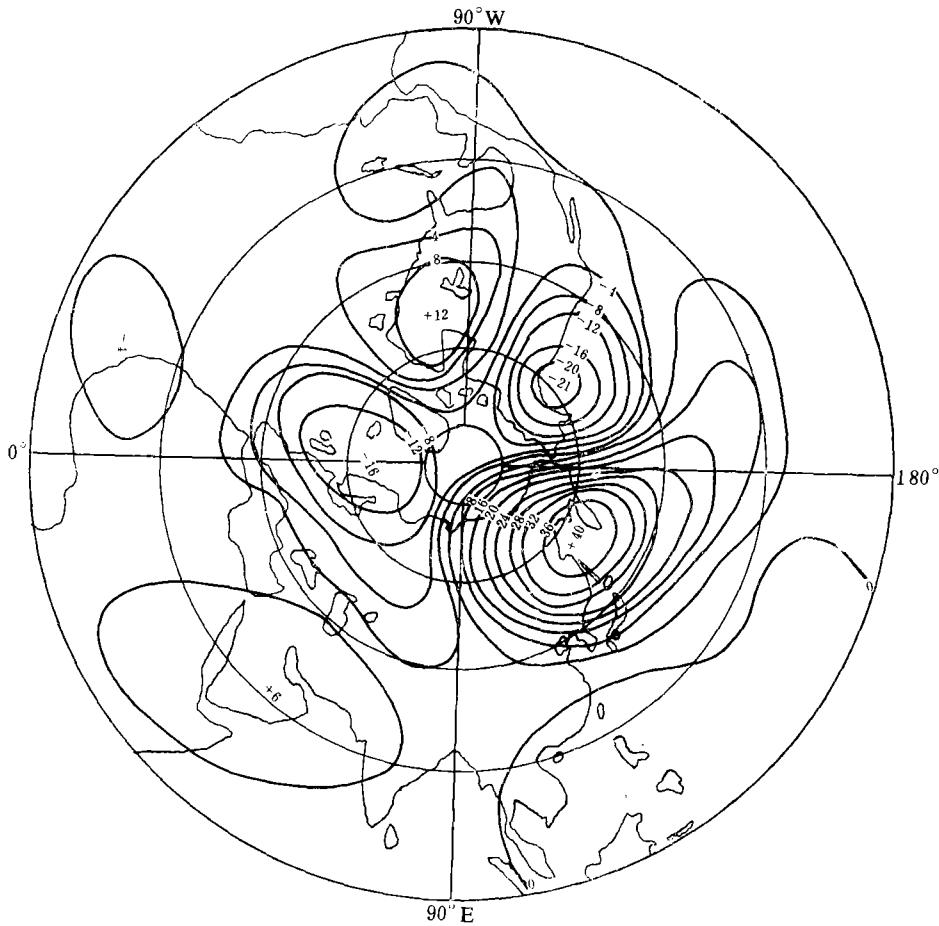


图 5 纯动力模式的预报

表 1 几个模式预报效果的对比

模 式	度量指标	同符号率	线性相关系数	误差绝对值的均值(dagpm)
纯动力模式		0.71	0.26	7.2
随机强迫模式		0.73	0.26	6.7
随机初值模式		0.75	0.33	5.1
随机初值与随机强迫相结合模式		0.76	0.40	5.0
惯性预报			0.12	

机强迫以及预报的时效(一个月)比较短有关。若做季或更长时效的预报,强迫的作用可能会变得更明显。

四、小 结

通过本试验工作说明,随机初值与随机强迫相结合的模式预报效果,不但明显优于纯

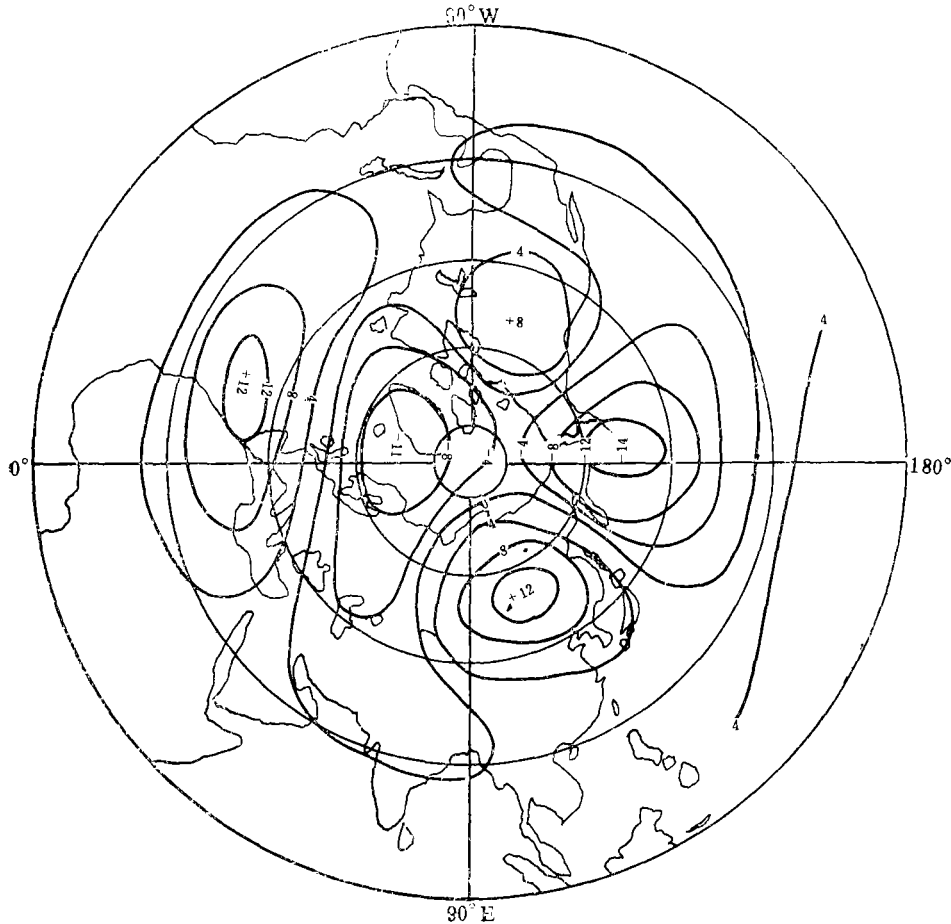


图 6 1983 年 1 月 500 hPa 月平均高度距平场实况

动力模式的预报,而且也优于随机强迫模式和随机初值模式的预报,是最好的预报。

本文所附的例子是计算出的第一个例子,还有待于今后继续计算更多的实例以进行比较。

参 考 文 献

- [1] Lorenz, E. N., Deterministic nonperiodic flow, *J. Atmos. Sci.*, 20, 130—141, 1963.
- [2] Epstein, E. S., Stochastic dynamic prediction, *Tellus*, 21, 739—759, 1969.
- [3] Leith, C. E., Theoretical skill of Monte-carlo forecasts, *Mon. Wea. Rev.*, 102, 409—418, 1974.
- [4] Seidman, A. N., Averaging techniques in long-range weather forecasting, *Mon. Wea. Rev.*, 109, 1367—1379, 1981.
- [5] ECMWF, Workshop on stochastic dynamic forecasting (17—19 October 1979), 1—96, May 1980.
- [6] Pitcher, Eric, J., Application of stochastic dynamic prediction to real data, *J. Atmos. Sci.*, 34, 3—21, 1977.
- [7] 胡增臻、史久恩、张建中, 随机初值模式的月平均环流数值预报试验, 长期天气预报论文集, 292—300, 气象出版社, 1990.
- [8] 中国科学院计算中心概率统计组, 概率统计计算, 1—472, 科学出版社, 1979.
- [9] 项静恬、杜金观、史久恩, 动态数据处理-时间序列分析, 1—133, 气象出版社, 1986.
- [10] 张建中、史久恩、胡增臻, 用随机强迫模式做 500 hPa 月平均环流预报的 Monte-Carlo 方法, 北京气象学

院学报, 1, 17—21, 1990。

THE ENSEMBLE FORECAST EXPERIMENT OF A SIMPLE MODEL WHICH INCLUDES RANDOM INITIAL VALUE AND RANDOM FORCING

Hu Zengzhen* Shi Jiuen

(Beijing Meteorological Institute, Beijing)

Zhang Jianzhong

(Computing Centre of the Chinese Academic Sinica, Beijing)

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.), the random forced term is added to the model, and the Monte-Carlo method with random initial values is used. A statistical-dynamic integrated model is thus built up, and a numerical forecasting experiment of 500 hPa 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.

*Present affiliation: Institute of Atmospheric Physics, Academia Sinica.