

ETKF 方法在区域集合预报中的初步应用

田伟红¹ 庄世宇²

(1. 国家气象中心,北京 100081;
2. 中国气象科学研究院灾害天气国家重点实验室)

提 要: 理论上,集合卡尔曼变换(ETKF—The Ensemble Transform Kalman filter)方法产生的集合扰动在观测空间具有等概率分布的特征,这一特点恰好可以弥补业务集合预报中常用的初值生成方法——Breeding 方法的不足。依据 ETKF 理论和方法,利用 GRAPES-Meso 中尺度模式建立了简单的基于 ETKF 的中尺度集合预报试验平台,并选取 2005 年 11 月 8—12 日出现的一次降雨过程进行集合预报试验,旨在研究 ETKF 方法用于集合预报时,对随时间变化的区域非线性系统,在有限集合数的条件下的特征。试验结果揭示了利用 ETKF 方法进行实际区域模式下集合预报的可行性及其一些基本性质,并指出试验当中存在的不足及今后研究的重点。
关键词: ETKF 方法 集合预报 集合扰动

Application of ETKF Method to Regional Ensemble Forecasts

Tian Weihong¹ Zhuang Shiyu²

(1. National Meteorology Center, Beijing 100081;
2. State Key Laboratory Severe Weather, Chinese Academy of Meteorological Sciences)

Abstract: In theory the disturbance generated by the ETKF (Ensemble Transform Kalman filter) method has the same probability in the observation space, so it can complement the shortage caused by the ensemble generation scheme Breeding. Based on ETKF theory, a simple mesoscale ensemble system is built using the GRAPES-Meso model and a case of rainfall occurred at 8th to 12th Nov 2005 is studied in order to verify whether such method can be applied into an imperfect time-varying regional ensemble model with limited ensemble member. The experiments indicate that this method can be used into the regional ensemble system, and it also can maintenance its priority, and the experimental shortcomings and the study focus in the future are also indicated.

Key Words: ETKF ensemble forecast ensemble disturbs

引 言

集合预报研究中最重要是如何形成初始扰动,包括初值扰动和模式扰动问题。初始扰动的质量好坏直接影响集合预报的质量。初值扰动方法^[1]可划分两类:第一类是估计分析误差的概率分布,如 Monte Carlo 随机扰动法、时间滞后平均法和观测扰动技术。第二类是通过分析数值预报误差在相空间中的增长方向和速度,沿着预报相空间中不稳定的方向来扰动初始条件。第二类扰动方法的原理建立在对大气的可预报性研究成果的基础上,如 NCEP 和 ECMWF 分别在全局模式中应用增长模繁殖法(BGM)和奇异向量法(SV)获得扰动初值,有效地反映了中高纬度斜压不稳定的预报不确定性。

ETKF 方法是由 Bishop^[2]等针对适应性观测问题提出来的,最近又被用于集合预报初值扰动的生成。这种方法基于集合变换方法^[3](ET)和卡尔曼滤波理论(KF),是一种次优的卡尔曼滤波方案,与其它的集合卡尔曼滤波方案不同在于 ETKF 方法可以直接估计出与观测网有关的预报误差协方差矩阵。2003 年 Wang 等^[4]又将 ETKF 方法用于集合预报初值扰动的生成,并与 Breeding 方法做了对比,结果表明利用 ETKF 方法得到的集合发散度比用 Breeding 方法得到的发散度更能反映预报的不确定性。由于 Wang(2003 年)文中介绍的方法给出集合扰动是有偏的,我们一般希望集合平均与最小方差估计得到的“大气真值”相等,故需采用集合中心化方案。2004 年, Wang 等^[5]对比了两种不同的集合中心化方案,试验发现集合扰动中心球形对称的 ETKF 方法要优于常用的集合扰动正负对称法^[1]。2006 年

Wei 等^[6]将 ETKF 方法在实际业务环境中进行试验,揭示了这种方法在业务应用中的优缺点。Harlim 等^[7]将 LEKF(Local Ensemble Kalman Filter)理论中的局地化分析思想和 ETKF 理论结合,提出了 LETKF 理论,来解决由于集合数的限制引起的虚假相关问题。

ETKF 方法作为一种集合预报初值扰动方法在区域中尺度模式(GRAPES-Meso)当中进行了试验,得到了初步的试验结论。主要内容如下:第 1 节简单地介绍 ETKF 方法产生集合扰动的原理;第 2 部分介绍试验设置;第 3 节内容为试验结果分析;第 4 节内容为结论与讨论。

1 ETKF 理论和方法简介

2003 年, Wang^[4]将 ETKF 方法用于集合预报初值扰动的生成。ETKF 方法可以通过将预报扰动量乘以一个变换矩阵 T (T 矩阵的求解见 wang^[4]文中介绍)来得到分析扰动

$$\mathbf{X}^a = \mathbf{X}^f T \quad (1)$$

我们将集合扰动定义为集合预报场与集合预报均值的差,于是有

$$\mathbf{X}^a = (\mathbf{x}_1^a - \bar{\mathbf{x}}^a, \mathbf{x}_2^a - \bar{\mathbf{x}}^a, \dots, \mathbf{x}_k^a - \bar{\mathbf{x}}^a) \text{ 和 } \mathbf{X}^f = (\mathbf{x}_1^f - \bar{\mathbf{x}}^f, \mathbf{x}_2^f - \bar{\mathbf{x}}^f, \dots, \mathbf{x}_k^f - \bar{\mathbf{x}}^f)$$

(k 表示集合数目, \mathbf{X}^a 表示分析量, \mathbf{X}^f 表示预报量)。

当集合预报的代表性较差时,利用 ETKF 方法估计得到的误差协方差矩阵 $\mathbf{P}^a = \mathbf{Z}^f \mathbf{T} \mathbf{T}^T (\mathbf{Z}^f)^T$ (这里 $\mathbf{Z}^f = \mathbf{X}^f / \sqrt{K-1}$), 就会远远小于实际的量。为了解决这个问题,可以将每一个扰动都乘以一个扩大因子 Π_i 来控制预报误差协方差和集合估计的预报误差协方差相等。

$$\mathbf{x}_i^a = \mathbf{x}_i^f T_i \Pi_i \quad (2)$$

$$\text{trace}(\langle \tilde{\mathbf{d}}_{i+1} \tilde{\mathbf{d}}_{i+1}^T \rangle) = \text{trace}(\tilde{\mathbf{H}} \mathbf{P}_{i+1}^e \tilde{\mathbf{H}}^T + \mathbf{I}) \quad (3)$$

其中 $\langle \cdot \rangle$ 表示数学期望算子, \mathbf{I} 为单位矩阵; $\tilde{\mathbf{H}}$ 为标准化的观测算子 $\tilde{\mathbf{H}} = \mathbf{R}^{-\frac{1}{2}} \mathbf{H}$; $\tilde{\mathbf{d}}_{i+1}$ 是一个更新向量 (innovation) $\tilde{\mathbf{d}}_{i+1} = \mathbf{R}^{-1/2} (\mathbf{Y}_{i+1} - \mathbf{H}\mathbf{X}_{i+1}^f)$; \mathbf{Y}_{i+1} 是 t_{i+1} 时刻观测值; \mathbf{X}_{i+1}^f 为控制预报量; \mathbf{P}_{i+1}^e 为集合估计的预报误差协方差矩阵。

在假设观测误差和预报误差不相关的条件下,公式(3)表示更新向量 (innovation)、预报误差协方差和观测误差协方差三者之间的关系。取向量的迹 (trace) 是为了使控制预报误差的方差和集合估计的预报误差方差相等。 $\tilde{\mathbf{d}}_{i+1}^T \tilde{\mathbf{d}}_{i+1}$ 相对于其均值的偏差很小,也就是说 $\tilde{\mathbf{d}}_{i+1}^T \tilde{\mathbf{d}}_{i+1}$ 在其均值 $\langle \tilde{\mathbf{d}}_{i+1}^T \tilde{\mathbf{d}}_{i+1} \rangle$ 附近分布。由于 $\langle \tilde{\mathbf{d}}_{i+1}^T \tilde{\mathbf{d}}_{i+1} \rangle = \text{trace}(\tilde{\mathbf{H}} \mathbf{P}_{i+1}^e \tilde{\mathbf{H}}^T + \mathbf{I})$, $\tilde{\mathbf{d}}_{i+1}^T \tilde{\mathbf{d}}_{i+1}$ 就可以大致估计方程(3)中 $\text{trace}(\tilde{\mathbf{H}} \mathbf{P}_{i+1}^e \tilde{\mathbf{H}}^T + \mathbf{I})$ 的值。于是式(3)可以写为:

$$\tilde{\mathbf{d}}_{i+1}^T \tilde{\mathbf{d}}_{i+1} \approx \text{trace}(\tilde{\mathbf{H}} \mathbf{P}_{i+1}^e \tilde{\mathbf{H}}^T + \mathbf{I}) \quad (4)$$

由于不知道 t_{i+1} 时刻的 $\tilde{\mathbf{d}}_{i+1}$ 及 $\tilde{\mathbf{H}} \mathbf{P}_{i+1}^e \tilde{\mathbf{H}}^T$, 所以只有用其前一时刻 t_i 时刻的值代替, 引入 α_i 来使等号成立。

$$\tilde{\mathbf{d}}_i^T \tilde{\mathbf{d}}_i = \text{trace}(\tilde{\mathbf{H}} \alpha_i \mathbf{P}_i^e \tilde{\mathbf{H}}^T + \mathbf{I}) \quad (5)$$

假设 t_{i-1} 时刻的扩大因子 (inflation factor) 为 $\mathbf{\Pi}_{i-1}$ 且已知, t_i 时刻的为:

$$\mathbf{\Pi}_i = \mathbf{\Pi}_{i-1} \sqrt{\alpha_i} \quad (6)$$

从式(5)中可以得到 α_i 的表达式。

$$\alpha_i = \frac{\tilde{\mathbf{d}}_i^T \tilde{\mathbf{d}}_i - N}{\text{trace}(\tilde{\mathbf{H}} \mathbf{P}_i^e \tilde{\mathbf{H}}^T)} = \frac{\tilde{\mathbf{d}}_i^T \tilde{\mathbf{d}}_i - N}{\sum_{i=1}^{k-1} \lambda_i} \quad (7)$$

N 表示观测点的数目, λ_i 是 $\mathbf{\Gamma}$ 的对角线上的元素 ($i=1, \dots, k-1$)。

于是可以得到

$$\mathbf{\Pi}_i = \sqrt{\alpha_i \alpha_2 \cdots \alpha_i} \quad (8)$$

得到 $\mathbf{\Pi}_i$ 以后就可以利用式(2)来得到分析扰动量, 得到的分析扰动量加到控制预报场上就得到集合预报的初始场。

2 试验设置

试验所选用的预报模式是全球/区域同化和预报系统 (GRAPES, Global and Regional Assimilation and PrEdiction System), 试验选取的区域为 $15 \sim 60^\circ \text{N}$, $70.125 \sim 144.375^\circ \text{E}$, 分辨率为 0.5625° , 水平格点为 133×81 个。控制预报的设置与上述设置一致。本次试验选取了 2005 年 11 月 8—12 日出现的一次降雨过程进行试验, 集合预报的起始时刻为 2005 年 11 月 7 日 12UTC, 到 2005 年 11 月 9 日 18UTC。试验用的分析系统为三维变分同化分析系统, 时间窗为 6 小时, 时间步长为 600s。试验所用的观测资料均为模拟的观测资料, 选取 00 时实际的探空观测站作为试验中所用的观测站, 由分析场插值到站点得到观测值。选取分析的垂直层为 3 层 850hPa、500hPa、200hPa; 选取的分析变量为 \mathbf{U} 、 \mathbf{V} 、 \mathbf{T} 三个变量, 集成员为 15 个。

3 试验结果分析

3.1 特征值分布

由于 ETKF 方法基于卡尔曼滤波理论, 因而具有卡尔曼滤波方法的一些滤波特性。ETKF 方法产生集合扰动的思想和 Breeding 方法是类似的, 这种方法的一个优点是在理想试验设置下其估计得到的分析误差的增长率是一致的。图 1 为在上述试验设置下误差

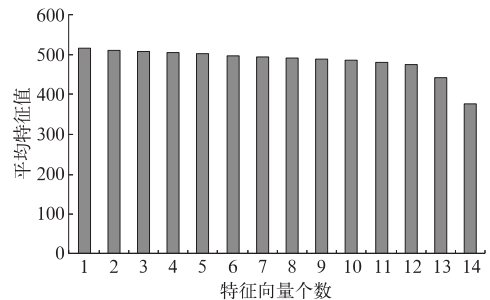


图 1 试验得到的分析误差协方差矩阵在观测空间的特征值谱分布

协方差矩阵的特征值谱分布状况。整体来看分析误差协方差矩阵的特征值谱分布还是比较一致的,这说明通过 ETKF 分析系统可以合理地调整集合初始扰动的增长率及分布,保证集合估计得到的误差概率密度函数的一致性。

3.2 集合数对扩大因子的影响

由于试验集合数的限制,使 ETKF 方法低估了部分预报误差协方差。为了弥补这方面的不足引入了扩大因子(inflation factor)来调整集合扰动的幅度。为了考察集合数的大小对扩大因子的影响,将集合数目扩大一倍进行试验。由表 1、表 2 可以看出,随着集合数由 15 变为 30,试验当中的扩大因子均值由 20.64 变为 9.06。随着集合数增大一倍,扩大因子的数值也减少一半(20.64/9.06)。由此可以看出在集合数增大的过程中扩大因子的作用逐渐减小,这与理论上一致。

表 1 当集合数为 15 时,试验的扩大因子变化情况

时间	扩大因子
2005110718UTC	4.26
2005110800UTC	32.1
2005110806UTC	41.61
2005110812UTC	19.05
2005110818UTC	17.87
2005110900UTC	14.39
2005110906UTC	21.27
2005110912UTC	17.26
2005110918UTC	17.92

表 2 当集合数为 30 时,试验 1 的扩大因子变化情况

时间	扩大因子
2005110718UTC	4.35
2005110800UTC	7.50
2005110806UTC	26.50
2005110812UTC	13.99
2005110818UTC	10.07
2005110900UTC	5.92
2005110906UTC	5.86
2005110912UTC	4.19
2005110918UTC	3.18

3.3 不同扩大因子计算方法得到的集合扰动随时间的变化情况

为比较不同扩大因子计算方法对集合扰动变化的影响,首先利用 ETKF 方法得到初始扰动,然后采用两种不同扩大因子的计算方法调整集合扰动的大小。两组实验分别为:

(1) 第一组试验方案采用 Wang and Bishop^[4]介绍的扩大因子计算方法进行试验。

(2) 第二组实验方案是利用 Breeding 方法的思想计算扩大因子、调整扰动的大小。Toth 和 Kalnay^[8]文中介绍的 Breeding 方法的基本思想是:将每一个预报扰动都乘以一个常数,使得到的集合扰动和根据经验得到的分析误差协方差相等:

$$\mathbf{X}^a = \mathbf{X}^f \times c$$

\mathbf{X}^a 表示分析扰动, \mathbf{X}^f 表示预报扰动, c 是一个常数即我们所需的扩大因子,在试验当中是使集合估计得到的 500hPa 高度上的温度方差与统计值相等得到扩大因子的数值。

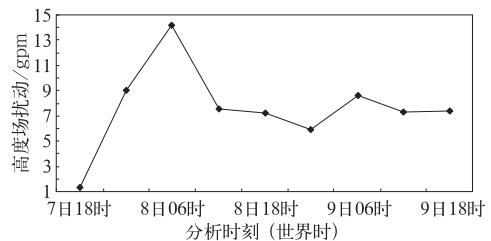


图 2 试验 1 得到的 500hPa 高度场扰动随时间变化情况

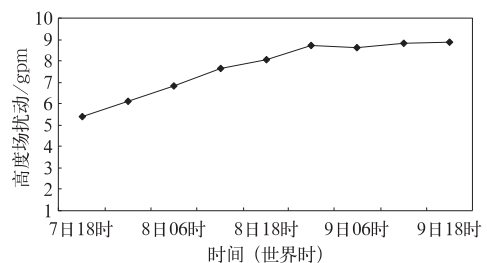


图 3 试验 2 得到的 500hPa 高度场扰动随时间变化情况

由图 2 可以看出试验 1 的 500hPa 高度场的扰动先增大后减少渐渐趋于一定值,图 3 表明试验 2 的 500hPa 高度场扰动是缓慢增长最后趋于稳定。对比图 2 和图 3 可以看出两种试验方案扰动增长方式的不同。

4 结论与讨论

集合变换卡尔曼滤波方法区别于 Breeding 方法一个最大的优点是^[4]在假设系统是线性,且不随时间变化、观测网的地理分布不变、观测误差固定、初始集合扰动代表线性系统误差增长的方向时,其产生的集合扰动的误差分布在彼此正交化的观测空间中的概率分布是一样的。为验证这种方法在一个随时间变化的非线性系统及有限集合数的条件下是否还可以保持其优点,利用 ETKF 方法进行了试验。实验结果表明对于非理想试验设置情况下 ETKF 方法在观测空间当中的误差特征值谱分布还是比较一致的。通过试验 1 和试验 2 的扩大因子变换情况可以看出,集合数增大的过程中扩大因子的作用逐渐减小;试验 1 得到的集合扰动是先增大后减小,最后趋于稳定,试验 2 得到的集合扰动是缓慢地增长,增大到一定程度保持不变。

目前试验当中存在的不足及改进办法:(1)ETKF 系统给出的初始扰动场是由观测加扰动方法给出的,使得集合预报的初始场的发散度偏小,以后试验中这个问题可以通过其它扰动生成方法生成集合预报场来解决。(2)试验当中所用的观测值都是由 GRAPES 模式的分析场插值到观测站点得到,而且假设观测站点的位置不随时间变化,这与实际情况不符。在进一步试验中可以考虑引入实际观测资料进行试验。(3)试验所用的模式是区域模式,所有成员采用相同的边界条件这时边界处的发散度就为零,边界的影响会随着模式积分时间向内传播,最终

影响集合预报场的质量。边界的影响可以通过扩大试验区域或者利用全球模式来克服。(4)在集合数较小的情况下,就会出现虚假相关问题,应在现用计算条件下尽可能多地引入集合预报成员。

在改善集合预报技术方面,要与数值预报其他技术的发展密切配合,协调发展。在集合预报中如何解决初始场不确定性和模式的不确定性仍将是今后研究的重点。

致谢:完成过程中得到奥地利气象局王勇博士和中国气象科学研究院数值中心的薛纪善、陈德辉等老师的热心帮助,在此向他们表示衷心的感谢!

参考文献

- [1] 田华. 国家气象中心中期集合预报概况[J]. 新疆气象, 2004, 27(5): 1-6.
- [2] Bishop C. H., Etherton B. J. and Majumdar S. J. Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects[J]. Mon. Wea. Rev., 2001, 129(3): 420-436.
- [3] Bishop C. H. and Toth Z. Ensemble Transform and Adaptive Observations [J]. J Atmos Sci, 1999, 56(11): 1748-1765.
- [4] Wang X., and Bishop C H. A comparison between Breeding and Ensemble Transform Kalman Filter Ensemble Forecast Schemes[J]. J Atmos sci, 2003, 60(9): 1140-1158.
- [5] Wang, X., CH Bishop and S.J. Julier. Which is better, an ensemble of positive - negative pairs or centered spherical simplex ensemble? [J] Mon. Wea. Rev. 2004, 132(7): 1590-1605.
- [6] Wei Mo zheng Z. Toth, R Wobus et al. Ensemble Transform Kalman Filter - based ensemble perturbation in an operational global prediction system at NCEP[J]. Tellus, 2006, 58A: 28-44.
- [7] Jone Harlim and Brian R Hunt. Local Ensemble Transform Kalman Filter: An Efficient Scheme For Assimilating Atmospheric Data[OL]. 2005, from http://www.atmos.umd.edu/~ekalnay/harlim_hunt05.pdf
- [8] Toth Z., and Kalnay E. Ensemble forecasting at NMC : The generation of perturbations[J]. Bull Amer Meteor soc, 1993, 74: 2317-2330.