

集合卡尔曼滤波数据同化在一维波动方程中的应用

费剑锋 韩月琪

(解放军理工大学气象学院, 南京 211101)

摘要 简要回顾了集合卡尔曼滤波(EnKF: Ensemble Kalman Filter)数据同化方法的发展历史,并介绍了 EnKF 数据同化方法的基本原理,利用一维非线性波动方程进行了数值试验。EnKF 数据同化方法的实现过程简单可行,避免了 EKF 中协方差演变方程预报过程中出现的计算不准确和关于协方差矩阵的大量数据的存储问题,最主要的是 EnKF 可以有效控制模式变量估计误差方差的增长,改善预报效果。

关键词 集合卡尔曼滤波 数据同化 数值试验

引言

扩展卡尔曼滤波(Extended Kalman Filter 或 EKF)在对控制变量较多的非线性系统进行数据同化时存在两个比较重要的困难:①系统控制变量的估计误差协方差随时间演变的计算;②巨大的计算代价和数据储存代价^[1]。为解决以上两个困难, Evensen(1994b)引入了 EnKF 数据同化。

利用 EnKF 进行数据同化,最初的工作是由 Evensen(1994)完成的^[1],并且 Evensen(1994)使用了一个试验个例证明了 EnKF 能够有效地解决 EKF 方法在对非线性系统进行数据同化时,由于使用线性近似和闭合方案而带来的协方差演变的计算不准确问题^[2]。随后,他们又进行了一些应用试验, Evensen 和 Van Leeuwen(1996)讨论了利用 EnKF 同化阿加勒斯流域测高数据的应用^[3],后来 Van Leeuwen 和 Evensen(1996)又使用集合平滑(ensemble smooth 或 ES)重新作了该试验并进行了对比分析,发现利用 ES 方法时,在线性系统中,ES 的估计精度更高,但对于模式的有色噪声滤波效果比 EnKF 差^[4]。

Evensen(1997)使用 Lorenz 系统作了应用举例,试验证明:即使是对于 Lorenz 系统那样一个具有混沌和高度非线性的系统,EnKF 能够有效地追

踪系统的状态转变,计算的误差与实际的误差估计相一致^[5]。

Houtekamer 和 Mitchell(1998)介绍了一种新的 EnKF 应用方法,该方法中将两个模式状态的集合同时向前积分,从一个集合预报中得到的误差统计来更新另一个集合,之所以使用两个集合,是因为这样可以减少分析过程中所遇到的“不繁殖性”问题。针对这一观点, Van Leeuwen(1999)提出了一些有争议的评论,并且 Houtekamer 和 Mitchell(1999)作出了回答^[6]。

Madsen 和 Canizarers(1999)在一个 2D 飓风湍流模式中,对比研究了 EnKF 和 EKF 的降秩方根(RRSQRT)同化方法。这是一个弱非线性问题,在此项试验中发现 EnKF 和 EKF 的结果十分一致^[7]。

Echevin 等(2000)使用普林斯顿海洋模式对 EnKF 进行了研究,且重点研究了多元协方差函数的水平结构和海表面以上的垂直结构,试验结果表明,EnKF 能够跟踪由于海岸线和海岸动力学引起的非各向同性的协方差函数,并且在这个领域中,算法简单是一个特殊的优点^[8]。

Houtekamer 和 Mitchell(2001)继续研究了使用两个集合的同化方法,同时引入了一个方案来进行含有大量观测数据的 EnKF 的全球分析;同时他们也引入了一种方法来同化由于有限的集合尺寸所导

致的最终的虚假长期相关项^[9]。

Reichle 等(2002)成功的讨论了 EnKF 的“描述解”,他们发现,EnKF 成功的向“描述解”逼近,但由于 EnKF 中误差高斯分布的假设,在分析时刻会引起一定的偏差。在所谓的描述方法中克服了这一缺点,描述方法能够获得最大可能性平滑解^[10]。

Natvik 和 Evensen(2003)第一次展示了 EnKF 在海洋生态模式中的理想的 3D 应用。他们的文章证明了如何通过同化 WiFs 海洋有色数据来控制海洋生态模式的演变。除此之外还引入了几种诊断方法,这些方法能够有效的用来检验集合的统计属性^[11,12]。

1 标准的卡尔曼滤波

1.1 模式方程

连续形式的非线性动力学方程可以写成如下形式:

$$\frac{d\psi(t)}{dt} = f(\psi, t) + b'(t) \quad (1)$$

其中 $\psi(t)$ 为 n 维模式状态向量, f 是 n 维非线性向量函数, $b(t)$ 为模式误差向量。离散形式的观测过程可以用下式来表达:

$$\psi_k^o = h(\psi_k) + \varepsilon_k \quad (2)$$

其中 $\psi_k \equiv \psi(t = t_k)$, m 维向量序列 $\psi_k^o = \{\psi_k^o, \psi_{k-1}^o, \dots, \psi_1^o\}$ 为对应的观测向量值, m 维函数 h 为观测算子,可为线性或非线性。假设 n 维模式噪声过程 $\{b(t)\}$ 和 m 维观测噪声过程 $\{\varepsilon_k\}$ 均为零均值高斯白噪声分布,其协方差矩阵为 $Q(t)$ 和 R_k ,而且过程 $\{b(t)\}$ 和过程 $\{\varepsilon_k\}$ 时间上互不相关(上标 t 代表真值,下同)。

1.2 卡尔曼滤波(KF)

KF 是有序的数据同化方法,即模式沿时间向前积分,在能够获得测量的时刻对资料进行同化,在模式连续的积分之前重新初始化。标准的 KF 满足方差最小化估计,即测量 d 和预报的模式状态向量 ψ^f 的线性组合的结果可以使分析估计 ψ^a 的方差最小(上标 f 和 a 分别代表预报和分析,下同)。分析估计的表达式如下(以下所有变量均为时间的函数,暂时省略):

$$\psi^a = \psi^f + K(d - H\psi^f) \quad (3)$$

卡尔曼增益矩阵 K 为

$$K = H^T (HP^f H^T + R)^{-1} \quad (4)$$

其中 P^f 是预报误差协方差矩阵。测量算子矩阵 H 通过式

$$d^f = H\psi^f \quad (5)$$

使模式状态与测量相联系,假设中 H 没有描述误差^[13],则真正的测量可以通过关系

$$d = H\psi^f + \varepsilon \quad (6)$$

定义, ε 是测量误差。观测误差协方差矩阵 R 定义成:

$$R = \overline{\varepsilon\varepsilon^T} = \overline{(d - H\psi^f)(d - H\psi^f)^T} \\ = \overline{(d - d^f)(d - d^f)^T} \quad (7)$$

式中“ $\overline{\quad}$ ”代表期望。矩阵 P^f 通过求解误差协方差矩阵演变方程

$$P^f = F[\psi^f, t] P^f + PF'^T[\psi^f, t] + Q \quad (8)$$

来获取,方程的推导和相关定义可参考文献[14]。滤波后,分析误差协方差矩阵形式如下:

$$P^a = \overline{(\psi^a - \psi)(\psi^a - \psi)^T} \\ = (I - KH) P^f (I - KH)^T + K R K^T \\ = P^f - K H P^f - P^f H^T K^T + K (H P^f H^T + R) K^T \\ = (I - KH) P^f \quad (9)$$

上式的推导利用了式(3)和式(5)。式(3)、(4)、(8)和(9)是 KF 的基本分析方程。

2 EnKF 的基本原理

2.1 误差统计的预报与非线性滤波

实质上, KF 方法是通过解柯尔莫哥洛夫方程(又称 Fokker-Plank 方程):

$$\frac{\partial \phi}{\partial t} + \sum_{i=1}^n g_i \frac{\partial \phi}{\partial \psi_i} = \sum_{i,j=1}^n \frac{Q_{i,j}}{2} \frac{\partial^2 \phi}{\partial \psi_i \partial \psi_j} \quad (10)$$

来进行误差预报的。其中 $\phi(\psi) = \frac{dN}{N}$ 是代表欧氏空间中所有状态向量 ψ 不确定性的概率密度分布函数。通过式(10)求得概率密度分部函数的解来计算不同时间层的各阶误差统计量。在线性动力学系统中,动力学方程式(10)是非耦合的,所以只要初始概率密度是高斯分布的, KF 能够给出最优的线性更新,然而对于非线性动力学模式,不同的统计分量之间是耦合的,所以即使初始误差分布是高斯的,均值和协方差也不可能保持通常的形式 $\phi(\psi, t)$,因而当 KF 方法应用于非线性动力学系统时,会变得十分复杂^[2]。因为大气或海洋动力学模式一般都是高度非线性的,所以需要一种先进的误差分析方案,才能对系统状态给

出准确的估计。

EnKF 应用所谓的马尔可夫蒙特卡罗 (MCMC) 方法求解式(10)。概率密度可用一个大的模式状态的集合来反映,通过向前积分如随机微分方程式(1)所描述的动力学模式,使用 MCMC 方法作集合预报可以近似求解式(10),这一方法构成了 EnKF 的基石^[3]。

2.2 EnKF 中误差统计的描述

在 KF 中依据真实状态定义

$$P^f = \overline{(\psi^f - \bar{\psi})(\psi^f - \bar{\psi})^T} \quad (11)$$

$$P^a = \overline{(\psi^a - \bar{\psi})(\psi^a - \bar{\psi})^T} \quad (12)$$

式中横线代表期望值,由于不知道确切的真值,因此围绕集合均值 $\bar{\psi}$ 来定义集合协方差矩阵

$$P^f \approx P_c^f = \overline{(\psi^f - \bar{\psi})(\psi^f - \bar{\psi})^T} \quad (13)$$

$$P^a \approx P_c^a = \overline{(\psi^a - \bar{\psi})(\psi^a - \bar{\psi})^T} \quad (14)$$

此时,式中横线代表集合均值(下同)。因而 EnKF 中近似的认为:集合均值是最优估计,并且整个集合围绕均值的发散分布状况就定义了集合的误差^[1]。在计算过程中,不用储存一个完整的协方差矩阵,可以使用一个恰当的模式状态的集合来反映同样的误差统计。有限大小的集合将总是提供误差协方差矩阵的一个近似估计,随着集合成员数的增加,蒙特卡罗的抽样误差将以 $1/\sqrt{N}$ 的比例减小,其中 N 为集合成员数^[3]。

2.3 EnKF 分析方案

现在使用如式(13)和(14)所定义的集合协方差来描述分析方案。首先必须将测量看成是随机过程,它的均值是测量的第一初始猜值,协方差为 R_0 ,因此从定义测量的集合开始,

$$d_j = \bar{d} + \varepsilon_j \quad (15)$$

$j = 1, \dots, N$, 保证模拟的随机观测误差的均值为 0, 此时的 j 与(2)式的 k 不同, j 代表的是集合成员(下同)。定义测量的集合协方差矩阵为:

$$R_c = \overline{\varepsilon \varepsilon^T} \quad (16)$$

集合无限大时, R_c 将趋向于标准的 KF 的 R 。

EnKF 的分析步骤包含如下适合于每一个集合成员的数据更新方案:

$$\psi_j^a = \psi_j^f + P_c^f H^T (H P_c^f H^T + R_c)^{-1} (d_j - H \psi_j^f) \quad (17)$$

对于有限大小的集合,式(17)是一个近似方案。进一步讲,当测量数大于集合成员数时,矩阵 $H P_c^f H^T$

和 R_c 将是奇异的,必须使用一个广义逆。注意:式(17)暗示:

$$\bar{\psi}^a = \bar{\psi}^f + P_c^f H^T (H P_c^f H^T + R_c)^{-1} (\bar{d} - H \bar{\psi}^f) \quad (18)$$

其中 $\bar{d} = \bar{d}$ 是测量向量的第一猜值。因此除了 P_c^a 和 R 用 P_c^a 和 R_c 来代替之外,EnKF 中预报和分析集合均值的关系等价于标准 KF 中预报和分析集合之间的关系。无论是使用第一猜值测量来更新集合均值,还是使用扰动测量来更新每一个集合成员,式(18)都是很好的选择。通过对每个集合成员进行更新,将得到一个具有正确误差统计的新的分析集合,再利用新的集合向前积分模式,直到下一个观测时刻,再次进行数据更新。

分析误差协方差 P_c^a 可用标准 KF 的方法推导,由式(17)和(18)得到:

$$\psi_j^a - \bar{\psi}^a = (I - K_c H) (\psi_j^f - \bar{\psi}^f) + K_c (d_j - \bar{d}) \quad (19)$$

$$K_c = P_c^f H^T (H P_c^f H^T + R_c)^{-1} \quad (20)$$

$$\begin{aligned} P_c^a &= \overline{(\psi^a - \bar{\psi}^a)(\psi^a - \bar{\psi}^a)^T} \\ &= (I - K_c H) P_c^f (I - K_c H)^T + K_c R_c K_c^T \\ &= P_c^f - K_c H P_c^f - P_c^f H^T K_c^T + K_c (H P_c^f H^T + R_c) K_c^T \\ &= (I - K_c H) P_c^f \end{aligned} \quad (21)$$

EnKF 分析过程中,必须视观测 d 为随机变量,才能将 R_c 引入,其中已经假定模式状态集合和观测状态集合的分布是独立的^[15]。

3 EnKF 应用试验

本文选取一维非线性波动方程,进行了 EnKF 数据同化试验,试验结果表明:EnKF 数据同化方法对于这样的非线性系统的预报效果有十分明显的改善作用。

本文选取的一维非线性波动方程为:

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = 0 \quad (22)$$

该方程 Cauchy 问题的解为 $u = u_0(x - ut)$, 表示以速度 u 沿 x 方向传播的非线性波动^[16]。取初始条件为 $u(x, 0) = -\sin x$ 时,可以得到式(22)的解析解为:

$$\tilde{u}(x, t) = -2 \sum_{n=1}^{\infty} (nt)^{-1} J_n(nt) \sin(nx) \quad (23)$$

其中 n 阶的 Bessel 积分为

$$J_n(z) = \frac{z}{n\pi} \int_0^\pi \cos(n\theta - z \sin \theta) \cos \theta d\theta \quad (24)$$

利用相应的数值积分方法可以得到任何时刻的解析解^[17],本试验将解析解作为参考解。

取式(22)的数值计算格式为欧拉格式,格距0.5,格点数64,时间步长0.05,积分时间单位4.5;假设函数 $u(x)$ 的特征不相关步长 $l=1.5$;选取100个集合成员数。取初始条件 $u(x,0)=-\sin x$,钢壁边界条件。使用欧拉格式积分式(22)(不经过任何同化,称欧拉解),发现随着积分时间的增加数值计算结果与解析解的偏差增大。

根据 Evensen(1994)介绍的方法,可以产生均值为0,任意方差,协方差由式(25)决定的一维平滑随机场,且该随机场在空间格点上是离散的。

$$P(x_1 - x_2) = \exp\left[-\frac{(x_1 - x_2)^2}{l^2}\right] \quad (25)$$

首先利用上述方法,计算100个均值为0,方差为1.0,协方差由式(25)决定的平滑随机场,叠加到初始猜值 $u(x,0)=-\sin x$,产生成员为100的初始场的集合;相同的方法,计算100个方差为0.25的随机场,作为模式误差初始场的集合;然后采用欧拉方法向前积分式(22),每6步加入一次观测值进行EnKF同化试验。取观测误差方差0.5,计算100个随机场叠加到同化时刻的解析解,产生观测集合,本试验只同化变量的观测。

为考察EnKF的同化效果,首先选取第一次同化后的0.4时间单位(积分第8次)和第二次同化时刻的0.6时间单位(积分第12次)作具体分析,试验结果见图1和图2。

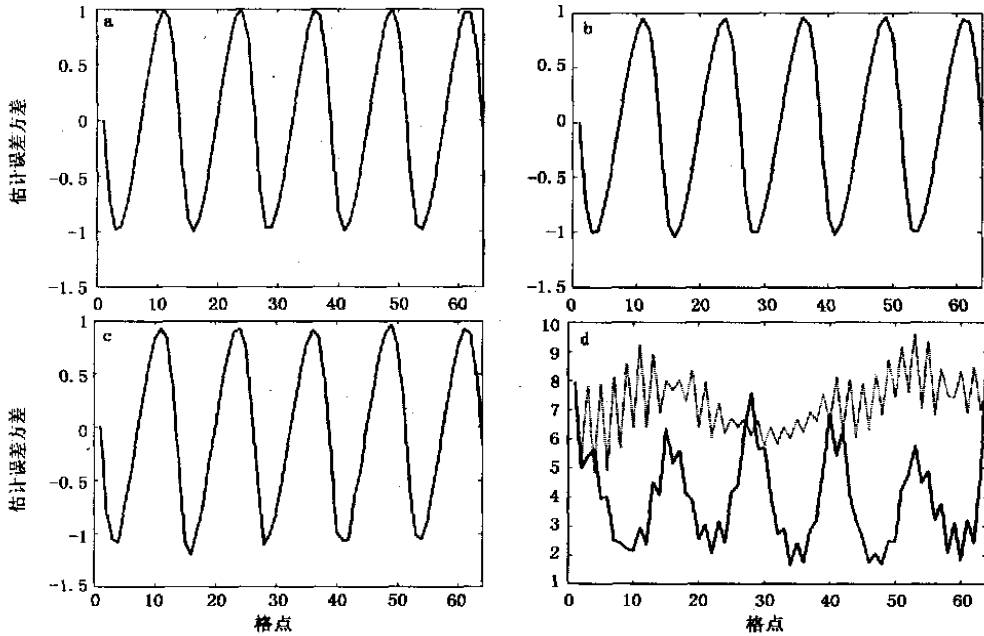


图1 $t=0.4$ 时间单位的数值试验结果

(a.解析解;b.0.3时间单位第一次同化后积分到0.4时间单位的结果;c.欧拉解;d.0.0和0.4时间单位变量 u 的估计误差方差($\times 10^{-3}$)的分布情况,虚线代表初始估计误差方差,实线代表0.4时间单位估计误差方差)

从图1看到,0.4时间单位时,欧拉解(图1c)已经偏离了解析解(图1a),而在0.3时间单位进行一次EnKF同化,继续积分到0.4时间单位的计算结果(图1b)与解析解(图1a)十分接近;另外从估计误差方差(图1d)也可看出,相对初始方差,0.4时间单位的估计误差方差已明显减小。

图2展示了第二次同化时刻的数值试验结果。首先由(图2a)至(2d)发现,随着积分时间的增加,

欧拉解(图2d)更加偏离了解析解(图2a),而在0.3时间单位利用EnKF同化一次后积分到0.6时间单位但还未经第二次同化的数值计算结果(图2b)相对于此时的欧拉解(图2d)已经有了进一步的改进,但与解析解(图2a)相比还有一定的偏差,而经过第二次同化的结果(图2c)已十分接近解析解(图2a),偏差明显减小;其次从(图2e)中发现,经过0.3时间单位第一次同化后,当继续积分式(22)从0.4时

间单位到第二次同化时刻 0.6 时间单位时,变量的预报误差方差增长得很快,超过了 0.4 时间单位的估计误差方差,但经过 EnKF 同化后,变量的估计误差方差得到了显著减小,小于 0.4 时间单位的估计

误差方差。这是因为经过一次 EnKF 同化之后,在模式继续向前积分直到下一次同化时刻的过程中,模式变量的估计误差方差成指数增长^[4],只有经过再次的同化,方差才能得到显著减小。

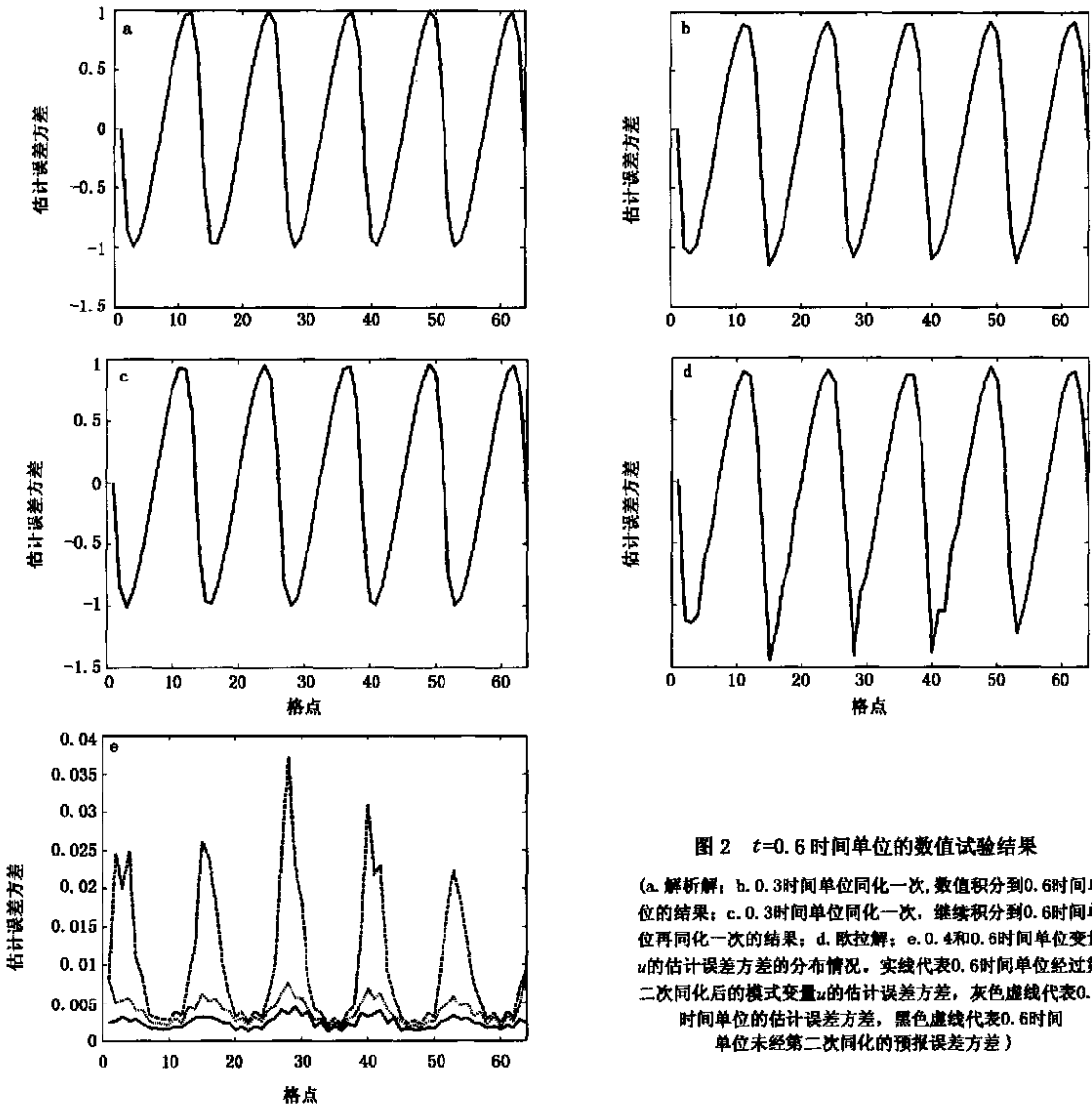


图 2 $t=0.6$ 时间单位的数值试验结果

(a. 解析解; b. 0.3 时间单位同化一次,数值积分到 0.6 时间单位的结果; c. 0.3 时间单位同化一次,继续积分到 0.6 时间单位再同化一次的结果; d. 欧拉解; e. 0.4 和 0.6 时间单位变量 u 的估计误差方差的分布情况,实线代表 0.6 时间单位经过第二次同化后的模式变量 u 的估计误差方差,灰色虚线代表 0.4 时间单位的估计误差方差,黑色虚线代表 0.6 时间单位未经第二次同化的预报误差方差)

最后选取第 4 个格点和第 60 个格点来考察 EnKF 同化对模式变量估计误差方差随时间演变的影响(图 3)。从图 3 看到,在两次同化之间的模式积分预报过程中,某一格点处模式变量估计误差方差迅速的增长,而每经过一次 EnKF 同化,就显著减小一次;且从整个试验过程看,该格点处模式变量估计误差方差的总体形势是减小的。再结合图 1(d) 和图 2(e) 中所有格点的积分预报和分析同化的误差方差的分布,说明以上 2 个格点可以反映全场的

演变情况。以上结果说明,EnKF 数据同化方法能够有效地控制模式变量估计误差方差的增长。

4 结束语

EnKF 数据同化方法的实现过程简单可行,避免了 EKF 中协方差演变方程预报过程中出现的计算不准确和关于协方差矩阵的大量数据的存储问题,最主要的是可以有效的控制估计误差方差的增长,改善预报的效果。

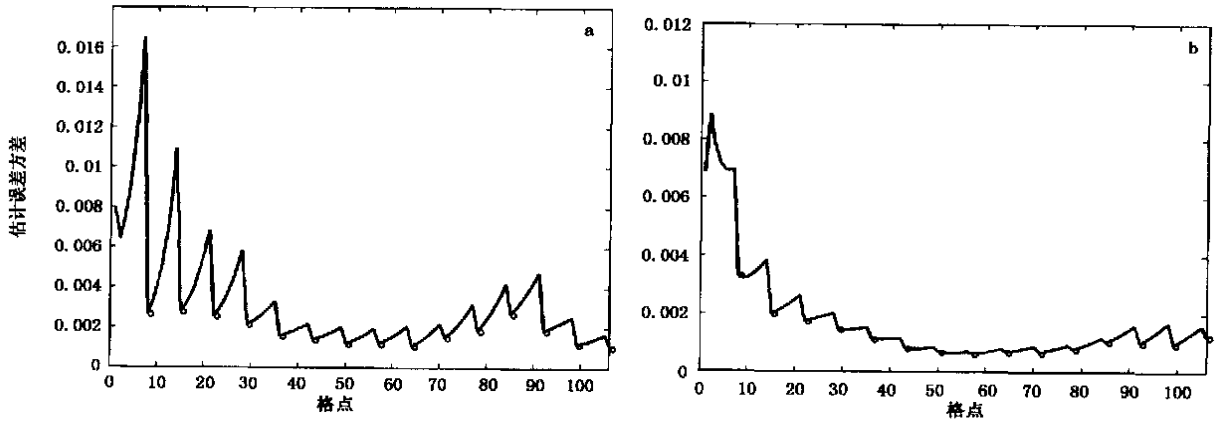


图3 第4个格点(a)和第60个格点(b)变量 u 的估计误差方差随模式积分和同化过程的演变
(因为模式积分预报从0到第90次,再加上其中包含的同化次数15次,所以横轴一共106个点;“。”代表EnKF同化。)

参考文献

- 1 Evensen G. Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, 1994b, 99(C5): 10,143 - 10,162
- 2 Evensen G. Inverse methods and data assimilation in nonlinear ocean models. *Phys. D.*, 1994a, 77: 108 - 129
- 3 Evensen G, van Leeuwen P J. Assimilation of Geosat altimeter data for the Agulhas Current using the Ensemble Kalman Filter with a quasi-geostrophic model. *Mon. Wea. Rev.*, 1996, 124: 85 - 96
- 4 van Leeuwen P J, Evensen G. Data assimilation and inverse methods in terms of probabilistic formulation. *Mon. Wea. Rev.*, 1996, 124: 2898 - 2913
- 5 Evensen G. Advanced data assimilation for strongly nonlinear dynamics. *Mon. Wea. Rev.*, 1997, 125: 1342 - 1354
- 6 Houtekamer P L, Mitchell H L. Reply notes and correspondence. *Mon. Wea. Rev.*, 1999, 127: 1378 - 1379
- 7 Madsen H, Canizares R. Comparison of extended and ensemble Kalman filters for data assimilation in coastal area modelling. *Int. J. Numer. Meth. Fluids*, 1999, 31: 961 - 981
- 8 Echevin V, Mey P D, Evensen G. Horizontal and vertical structure of the representer functions for sea surface measurements in a coastal circulation model. *J. Phys. Oceanogr.*, 2000, 30: 2627 - 2635
- 9 Houtekamer P L, Mitchell H L. A sequential ensemble Kalman filter for Atmospheric data assimilation. *Mon. Wea. Rev.*, 2001, 129: 123 - 137
- 10 Reichle R H, McLaughlin D B, Entekhabi D. Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Wea. Rev.*, 2002, 130: 103 - 114
- 11 Natvik L J, Evensen G. Assimilation of ocean colour data into a biochemical model of the North Atlantic. Part 1. Data assimilation experiments. *J. Marine Sys.*, 2003a, 40 - 41: 127 - 153
- 12 Natvik L J, Evensen G. Assimilation of ocean colour data into a biochemical model of the North Atlantic. Part 2. Statistical analysis. *J. Marine Sys.*, 2003b, 40 - 41: 155 - 169
- 13 Houtekamer P L, Mitchell H L. Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, 1998, 126: 796 - 811
- 14 Burgers G, van Leeuwen P J, Evensen G. Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, 1998, 126: 1719 - 1724
- 15 Evensen G, van Leeuwen P J. An ensemble Kalman smoother for nonlinear dynamics. *Mon. Wea. Rev.*, 2000, 128: 1852 - 1867
- 16 刘式适,刘式达. 非线性大气动力学. 北京:气象出版社,1996
- 17 伍荣生. 大气动力学. 北京:气象出版社,1990

(下转第119页)

Application of Ensemble Kalman Filter Assimilation in One Dimensional Wave Equation

Fei Jianfeng Han Yueqi

(College of Meteorology, University of Science and Technology of PLA, Nanjing 211101)

Abstract: A brief review is given of the development history of the Ensemble Kalman Filter (EnKF) assimilation method and its essential theory. A numerical test was conducted using the one-dimensional nonlinear wave equation. Results indicate that the method has advantages of being simple to realize and being able to avoid the computation inaccuracy occurred in the prediction process using the EKF covariance variation equation and the problem of massive data storage resulted from the covariance matrix; moreover, the EnKF can control effectively the augment of the estimated error variance for model variables and improve the prediction precision.

Key words: ensemble Kalman filter, data assimilation, numerical experiment