

集合滤波和三维变分混合数据同化方法研究*

吴新荣^{1,2,3}, 韩桂军², 李冬², 李威²

1. 中国科学院南海海洋研究所, 广东 广州 510301; 2. 国家海洋信息中心, 天津 300171;

3. 中国科学院研究生院, 北京 100049

摘要: 发展了一种新的混合数据同化方法——基于集合滤波和三维变分的混合数据同化方法。该方法将集合调整卡尔曼滤波(ensemble adjustment Kalman filter, EAKF)得到的集合样本扰动通过一个转换矩阵的形式直接作用到背景场上, 利用顺序滤波的思想得到分析场的一个扰动; 然后在三维变分(three dimensional variational analysis, 3D-Var)的框架下与观测数据进行拟合, 从而给出分析场的最优估计。文中以 Lorenz63 模型为例, 开展了理想数据同化试验, 结果表明, 相比于集合调整卡尔曼滤波, 这种新的混合同化方法可以给出更好的同化结果。

关键词: 混合数据同化方法; 集合调整卡尔曼滤波; 三维变分

中图分类号: P731 文献标识码: A 文章编号: 1009-5470(2011)06-0024-07

A hybrid ensemble filter and 3D variational analysis scheme

WU Xin-rong^{1,2,3}, HAN Gui-jun², LI Dong², LI Wei²

1. South China Sea Institute of Oceanology, CAS, Guangzhou 510301, China;

2. National Marine Data and Information Service, Tianjin 300171, China; 3. Graduate University of CAS, Beijing 100049, China

Abstract: A new hybrid data assimilation scheme based on ensemble adjustment Kalman filter (EAKF) and three-dimensional variational (3D-Var) analysis is developed. In this assimilation scheme, the perturbation of ensemble from EAKF is applied to the background field by using a transformation matrix, thus the perturbation of the analysis field can be obtained by taking advantage of a sequential filter, which will then be optimized by being combined with observations under the framework of 3D-Var. The data assimilation experiment in a perfect case is carried out by using Lorenz-63 model. The results demonstrate that the hybrid data assimilation scheme performs better than EAKF.

Key words: hybrid data assimilation scheme; ensemble adjustment Kalman filter; 3D-Var

基于最优控制论的变分方法^[1]和基于估计理论的集合滤波方法^[2] 是两类主要的现代数据同化方法。三维变分数据同化方法(three-dimensional variational analysis, 3D-Var)是一种经济、灵活以及统计可靠的变分数据同化方法, 其假定预报和观测的误差不随时间变化且服从高斯分布。四维变分(four-dimensional variational analysis, 4D-Var)是在模式的动力约束下, 寻求拟合给定时间窗口内的观测的最佳模式轨迹。4D-Var 的主要问题在于切线性模式和伴随代码的编写, 目前业务化运行时一般都采

用简化的增量形式^[3], 而且其背景场误差协方差矩阵需要其他方法给定。集合滤波(ensemble filter)是广泛使用的基于集合的数据同化方法, 其利用随模式非线性演化的集合成员对流相关(flow-dependent)的误差分布进行估计。如何将变分同化方法和集合滤波方法结合起来, 利用其各自的优点, 得到研究者越来越多的关注。Hamill 等^[4]将 3D-Var 和集合均方根卡尔曼滤波(ensemble square root Kalman filter, EnSRF^[5])结合起来, 指出如果样本足够大, 那么根据集合计算出的背景误差协方差可以完全代替

收稿日期: 2010-01-31; 修订日期: 2010-04-26。孙淑杰编辑

基金项目: 国家重点基础研究发展计划项目(2007CB816001); 国家自然科学基金项目(40776016)

作者简介: 吴新荣(1981—), 男, 江苏省泰州市人, 在读博士生, 主要从事海洋数据同化方法应用研究。E-mail: wuxinrong@mail.nmdis.gov.cn

3D-Var 的静态背景误差协方差。Corazza 等^[6]提出了另一种混合机制, 即利用集合样本对 3D-Var 的背景误差协方差进行膨胀, 从而将流相关的背景误差引入 3D-Var 中。Zupanski^[7]发展了最大似然集合滤波(maximum-likelihood ensemble filter, MLEF), 其使用集合样本计算 Hessian 矩阵和目标泛函的梯度, 改进了集合滤波中非线性观测算符的处理方法。Lorenc^[8]将 3D-Var 和集合卡尔曼滤波(ensemble Kalman filter, EnKF^[2])结合起来, 利用这种混合机制来确定流相关的背景误差协方差。Buehner^[9]采用了类似的混合机制, 指出这种混合数据同化方法的同化效果类似于 EnKF, 优于 3D-Var。Wang 等^[10]将集合转换卡尔曼滤波(ensemble transform Kalman filter, ETKF^[11])和最优插值(optimum interpolation, OI)结合起来, 将由集合样本计算出的背景误差协方差和静态背景误差协方差进行加权, 从而获得流相关的背景误差。Hunt^[12]和 Fertig 等^[13]提出了四维集合卡尔曼滤波(four dimensional ensemble Kalman filter, 4D-EnKF), 试验表明根据集合样本计算出的背景误差协方差可以应用于变分算法中。Liu 等^[14]发展了集合四维变分(ensemble four dimensional variation analysis, En4DVAR), 根据集合样本构建流相关的背景误差协方差, 然后在避免切线性和伴随模式的前提下进行 4D-Var 同化。Zhang 等^[15]将 EnKF 和 4D-Var 相耦合, 即 EnKF 为 4D-Var 提供流相关的背景误差协方差, 4D-Var 用来优化集合平均, 达到防止滤波发散的目的。理想试验表明, 耦合同化方法的同化效果比 EnKF 和 4D-Var 好。Wan 等^[16]发展了覆裹集合卡尔曼滤波(dressed ensemble Kalman filter, DrEnKF), 利用 10 个动力样本以及 90 个由模式积分获得的静态集合成员计算背景误差协方差, 从而更新 10 个动力样本, 结果表明 DrEnKF 与 100 个动力样本的 EnKF 的同化效果相近, 且误差协方差吻合较好。

上面提到的集合滤波和变分同化方法混合机制的切入点都在背景误差协方差上, 即利用先验集合样本为变分方法提供流相关的背景误差协方差, 或利用由动力样本和静态样本计算的误差协方差的加权平均作为背景误差协方差矩阵。本文将集合滤波方法和 3D-Var 结合起来, 发展一种新的混合同化方法, 即得到的预报样本扰动通过一个转换矩阵的形式直接作用到背景场上, 得到分析场的一个扰动; 然后在三维变分的框架下与观测数据进行拟合, 从而给出分析场的最优估计。

本文的结构安排如下: 第一部分介绍本文所发展的混合数据同化方案; 第二部分为理想数据同化试验和结果分析; 第三部分为总结。

1 集合滤波和三维变分混合数据同化方法

一般情况下, 3D-Var 的目标泛函表达式^[17,18]为:

$$J(X) = J_b(X) + J_o(X) \\ = \frac{1}{2}[(X - X_b)^T B^{-1}(X - X_b) + (HX - Y_{obs})^T R^{-1}(HX - Y_{obs})] \quad (1)$$

其中, X 表示由控制变量构成的向量, X_b 表示背景场, B 表示背景场误差协方差矩阵, H 表示观测算符, Y_{obs} 表示观测向量, R 表示观测误差协方差矩阵, J_b 和 J_o 分别表示背景项和观测项。

目标泛函关于控制变量的梯度为:

$$\nabla_X J = B^{-1}(X - X_b) + H^T R^{-1} H (X - H^{-1} Y_{obs}) \quad (2)$$

对应的分析场为:

$$X_a = X_b + B H^T (H B H^T + R)^{-1} (Y_{obs} - H X_b) \quad (3)$$

本文采用的集合滤波方法是 Anderson^[19]提出的集合调整卡尔曼滤波法(ensemble adjustment Kalman filter, EAKF), 其定义联合状态-观测向量

$$Z = [X, HX_t] = [X, Y] \quad (4)$$

其中, X 是模式状态向量, X_t 是状态变量的真值, H 是观测算符, Y 表示观测向量的真值; Z 的维数为 $n+m$, 其中 n 是控制变量的维数, m 是观测向量的维数。观测是模式状态变量的函数, 并且包含相应的观测误差

$$Y_{obs} = H X_t + \varepsilon \quad (5)$$

其中, ε 是 m 维服从均值为 0、协方差为 R 的高斯分布的观测误差。

根据贝叶斯理论, Z 的后验分布可以表示为

$$p(Z^u) = p(Y_{obs} | Z^p) p(Z^p) / D \quad (6)$$

其中, Z^u 和 Z^p 分别表示 Z 的后验和先验取值, D 是标准化因子。在高斯分布的假定下, Z^u 也服从高斯分布, 其均值和协方差分别为

$$\bar{Z}^u = \Sigma^u \left[(\Sigma^p)^{-1} \bar{Z}^p + H^{*T} R^{-1} Y_{obs} \right] \quad (7)$$

和

$$\Sigma^u = \left[(\Sigma^p)^{-1} + H^{*T} R^{-1} H^* \right]^{-1} \quad (8)$$

其中, Σ^u 和 Σ^p 分别表示 X 的后验和先验误差协方差矩阵, \bar{Z}^u 和 \bar{Z}^p 分别表示 Z 的后验和先验均值, H^* 是联合状态空间中的观测算符。(7)和(8)式的推

导过程详见文献[19]。

通过对 Σ^p 以及 $H^T R^{-1} H$ 进行相关变换以及奇异值分解(singular value decomposition, SVD), Z^u 可以写成下式

$$\Sigma^u = A \Sigma^p A^T \quad (9)$$

上式的推导以及 A 的表达式详见文献[19]中的附录 A。

这样, 集合样本即可利用下式进行更新

$$Z_i^u = A^T \left(Z_i^p - \bar{Z}^p \right) + \bar{Z}^u \quad (10)$$

其中, i 表示集合成员。

Houtekammer 等^[20]指出, 如果观测之间是不相关的, 那么可以顺序地对其进行同化而不改变(6)式的结果。EAKF 正是采用了这种顺序同化的思想。对于单个观测 Y^o , 在全局(局部)最小二乘意义下, EAKF 主要包括如下 2 个步骤:

首先, 利用观测算符计算观测的集合成员, 并计算每个成员的增量 ΔY_i ;

其次, 计算每个状态变量的每个集合成员的增量

$$\Delta X_{ij} = \frac{\sigma_{jo}^p}{\sigma_o^p} \Delta Y_i \quad (11)$$

其中, i 表示集合成员, j 表示状态变量, σ_{jo}^p 表示第 j 个状态变量与观测之间的先验协方差, σ_o^p 表示观测的先验方差。方程(10)是两步法的向量形式。

下面推导该两步法框架下状态变量集合成员的更新表达式。

根据文献[21]中的方程(24)

$$X_{ij}^u = X_{ij}^p - \bar{X}_j^p + \sigma_{jo}^p \Gamma \left(Y_i^p - \bar{Y}^p \right) + \bar{X}_j^u \quad (12)$$

其中 \bar{Y}^p 表示 Y^o 的先验均值; Y_i^p 表示 Y^o 的第 i 个先验集合成员; Γ 的表达式参看文献[21]; X_{ij}^p 和 X_{ij}^u 表示 X_{ij} 的先验和后验取值; \bar{X}_j^p 和 \bar{X}_j^u 表示第 j 个状态变量的先验和后验均值, 根据文献[21]的方程(14), 后验均值的表达式为

$$\bar{X}_j^u = \bar{X}_j^p + \frac{\sigma_{jo}^p}{r + \sigma_o^p} \left(Y^o - \bar{Y}^p \right) \quad (13)$$

其中 r 表示 Y^o 的方差。将(13)式代入(12)式, 可得

$$X_{ij}^u = X_{ij}^p + \sigma_{jo}^p \left[\Gamma \left(Y_i^p - \bar{Y}^p \right) + \left(r + \sigma_o^p \right)^{-1} \left(Y^o - \bar{Y}^p \right) \right] \quad (14)$$

此即第 j 个状态变量的第 i 个集合成员的更新表达式。

对 m 个观测, 利用(14)式顺序进行同化, 这样

更新后的状态变量的集合成员严格地服从后验高斯分布。另外, EAKF 上述两步法中仅使用了一阶和二阶矩(线性), 因此先验分布的高阶矩(非线性)结构得以保留。

本文将 EAKF 的机制引入 3D-Var 的背景项中, 用集合本来构建流相关的背景误差协方差; 用顺序滤波来继承状态变量先验分布的非线性信息; 用 3D-Var 来给出状态场的最优估计, 这样就得到了一种新的混合同化方法, 本文称之为集合滤波和三维变分混合数据同化方法(hybrid ensemble filter - 3D-Var)。

由于背景场变量之间是相关的, 为了实施 EAKF 算法, 需要通过一个转换矩阵的形式, 将其旋转到变量独立空间; 然后利用 3D-Var 对观测进行同化, 得到最优分析场; 最后通过反旋转得到一个新的状态变量的分析场。

下面讨论如何将背景场转换到变量独立空间。

$$\mathbf{X} = \left(X_1^T, X_2^T, \dots, X_N^T \right)^T, \mathbf{X}_b = \left(X_{b,1}^T, X_{b,2}^T, \dots, X_{b,N}^T \right)^T,$$

其中 N 表示集合样本大小, X_i ($i=1, \dots, N$) 表示由模式格点组成的控制变量的第 i 个集合成员, 其维数为 $n \times 1$, 其中 n 表示模式格点数; $X_{b,i}$ ($i=1, \dots, N$) 表示由模式格点组成的第 i 个背景集合成员, 其维数为 $n \times 1$ 。另外, 假设观测向量 Y_{obs} 的维数为 $m \times 1$; H 为观测算符; R 为观测误差协方差矩阵。根据 X_b , 可以计算出背景误差协方差矩阵

$$\mathbf{B} = \left(X_{b,1} - \bar{X}_b, \dots, X_{b,N} - \bar{X}_b \right) \left(X_{b,1}^T - \bar{X}_b^T, \dots, X_{b,N}^T - \bar{X}_b^T \right)^T,$$

$$\bar{X}_b = \frac{1}{N} \sum_{i=1}^N X_{b,i} \quad (15)$$

其中, \bar{X}_b 表示集合平均。

对 \mathbf{B} 和 \mathbf{R} 进行 SVD 分解

$$\mathbf{B} = \mathbf{Q} \mathbf{T} \mathbf{Q}^T, \mathbf{R} = \mathbf{G} \mathbf{S} \mathbf{G}^T \quad (16)$$

其中, \mathbf{Q} 和 \mathbf{G} 是正交矩阵, \mathbf{T} 和 \mathbf{S} 为对角矩阵。

对控制变量 X_i 、背景场 $X_{b,i}$ 以及观测场 Y_{obs} 进行如下变量代换:

$$\mathbf{x}_i = \mathbf{Q}^T X_i, \mathbf{x}_{b,i} = \mathbf{Q}^T X_{b,i}, i = 1, \dots, N, \quad (17)$$

$$\mathbf{y}_{\text{obs}} = \mathbf{G}^T Y_{\text{obs}}$$

$$\text{令 } \mathbf{x} = \left(\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_N^T \right)^T, \mathbf{x}_b = \left(\mathbf{x}_{b,1}^T, \mathbf{x}_{b,2}^T, \dots, \mathbf{x}_{b,N}^T \right)^T,$$

$\bar{\mathbf{x}}_b = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_{b,i}$ 。根据 \mathbf{x}_b 可以计算出新的背景误差协方差矩阵

$$\mathbf{B}' = \left(\mathbf{x}_{b,1} - \bar{\mathbf{x}}_b, \dots, \mathbf{x}_{b,N} - \bar{\mathbf{x}}_b \right) \left(\mathbf{x}_{b,1}^T - \bar{\mathbf{x}}_b^T, \dots, \mathbf{x}_{b,N}^T - \bar{\mathbf{x}}_b^T \right)^T \quad (18)$$

将(17)代入(18), 并根据(16)以及 \mathbf{Q} 的正交性, 容易得出:

$$\begin{aligned} \mathbf{B}' &= \mathbf{Q}^T (\mathbf{X}_{b,1} - \bar{\mathbf{X}}_b, \dots, \mathbf{X}_{b,N} - \bar{\mathbf{X}}_b) \\ &\quad (\mathbf{X}_{b,1}^T - \bar{\mathbf{X}}_b^T, \dots, \mathbf{X}_{b,N}^T - \bar{\mathbf{X}}_b^T)^T \mathbf{Q} \\ &= \mathbf{Q}^T \mathbf{B} \mathbf{Q} = \mathbf{Q}^T (\mathbf{Q} \mathbf{T} \mathbf{Q}^T) \mathbf{Q} = \mathbf{T} \end{aligned} \quad (19)$$

因此 \mathbf{B}' 是对角矩阵, 即 $\mathbf{x}_{b,i}$ 中各个模式格点之间是不相关的。

在每一个分析步, 混合数据同化方法主要包含以下步骤。

第一步, 利用先验集合样本计算背景误差协方差矩阵 \mathbf{B} ;

第二步, 对背景误差协方差矩阵 \mathbf{B} 以及观测误差协方差矩阵 \mathbf{R} 进行 SVD 分解;

第三步, 利用(17)式进行变量代换, 将背景场以及观测场转换到独立空间中;

第四步, 对第 $i=1, \dots, N$ 个集合成员, 依次执行下面 3 步:

首先, 利用(14)对背景场 $\mathbf{x}_{b,i}$ 进行顺序滤波

$$\begin{aligned} \mathbf{x}_{ik}^u &= \mathbf{x}_{ik} + \sum_{j=1}^n \left[\Gamma_j + \left(r_j + \sigma_{jj}^p \right)^{-1} \right] (\mathbf{x}_{b,ij} - \bar{\mathbf{x}}_{b,j}) \sigma_{kj}^p, \quad (20) \\ k &= 1, \dots, n \end{aligned}$$

其中 Γ_j 表示第 j 个状态变量的 Γ ; r_j 表示第 j 个状态变量的“观测”误差; \mathbf{x}_{ik}^u 、 \mathbf{x}_{ik} 、 $\mathbf{x}_{b,ij}$ 以及 $\bar{\mathbf{x}}_{b,j}$ 分别表示分析场 \mathbf{x}_i^u 、控制变量 \mathbf{x}_i 、背景场 $\mathbf{x}_{b,i}$ 以及先验均值 $\bar{\mathbf{x}}_b$ 的第 k 、 k 、 j 以及 j 个分量;

其次, 利用 3D-Var 对观测场 \mathbf{y}_{obs} 进行同化。目标泛函采用下式

$$\begin{aligned} J_i(\mathbf{x}_i) &= \frac{1}{2} \left\{ \left(\mathbf{x}_i^u - \mathbf{x}_{b,i} \right)^T \mathbf{T}^{-1} \left(\mathbf{x}_i^u - \mathbf{x}_{b,i} \right) + \right. \\ &\quad \left. \left(\mathbf{H} \mathbf{x}_i^u - \mathbf{y}_{\text{obs}} \right)^T \mathbf{S}^{-1} \left(\mathbf{H} \mathbf{x}_i^u - \mathbf{y}_{\text{obs}} \right) \right\} \end{aligned} \quad (21)$$

其中 \mathbf{x}_i^u 是 \mathbf{x}_{ik}^u 的向量形式, \mathbf{x}_i 是 \mathbf{x}_{ik} 的向量形式, \mathbf{H} 是当前空间下的观测算符。计算出目标函数关于 \mathbf{x}_i 的梯度, 进而利用拟牛顿法^[22]搜索最优解 \mathbf{x}_i^a ;

最后, 对 \mathbf{x}_i^a 进行反正交变换, 从而得到原始空间里第 i 个集合成员的分析场

$$\mathbf{X}_i^a = \mathbf{Q} \mathbf{x}_i^a \quad (22)$$

2 理想数据同化试验

2.1 模型及参数设定

采用 Lorenz-63 模型^[23]作为理想试验模型, 其

控制方程为

$$\begin{cases} \frac{dx_1}{dt} = -\sigma x_1 + \sigma x_2 \\ \frac{dx_2}{dt} = -x_1 x_3 + \kappa x_1 - x_2 \\ \frac{dx_3}{dt} = x_1 x_2 - b x_3 \end{cases} \quad (23)$$

其中, x_1 、 x_2 和 x_3 分别表示对流运动的强度、水平方向以及垂直方向的温度梯度, σ 、 b 以及 κ 分别取为 10、8/3 以及 28, 时间步长为 0.0001s, 差分方案采用双逼近方案^[23]。

由(0.0, 1.0, 0.0)向前积分 2×10^6 步, 取后面 1×10^6 步的积分结果作为真值。同化的初始时刻设在第 1×10^6 步, 总的同化时段为 1×10^6 个积分步, 即 100s。观测是在真场上加上均值为 0、标准差为 2 的高斯噪声所生成; 初始集合成员是在初始真值的基础上给每个模式变量加上均值为 0、标准差为 2 的高斯噪声所形成。需要说明的是, 本文所给出的时间序列图中所示的积分步都是从第 1×10^6 步起算。

为了避免滤波发散^[24], 在 EAKF 和混合数据同化算法中均加入了方差膨胀因子 γ ^[19], 以增加先验分布的协方差, 从而提高观测的权重。目前, γ 的选取一般都是经验性的, 本文中 γ 的选取遵循如下原则: 当集合平均的均方根误差的时间平均值 (time-averaged root mean square (Rms) error of the ensemble mean, RmsEm) 与集合成员的均方根误差的时间平均值 (time-averaged mean Rms error of the individual ensemble members, Mrms) 之比与 $\sqrt{(N+1)/2N}$ (N 表示集合样本大小) 相差不超过 1% 时, 认为此时的 γ 是最佳的^[25]。

2.2 同化结果分析

本文所设计的理想试验均是在完美模式下进行的, 目的是为了考察同化方法在降低由于初始条件的不准确性所引起的误差的能力。

观测区间设为 100 个积分步, 即每 100 个积分步有一个观测。采用 20 个集合成员, 且每个分析步 x_1 、 x_2 和 x_3 均有观测。EAKF 同化试验结果的集合平均的均方根误差的时间平均值 (RmsEm) 为 0.59, 而混合同化试验结果的 RmsEm 为 0.49。从总体误差统计结果来看, 混合同化方法的结果比 EAKF 改进了约 17%。

图 1 和图 2 分别给出 EAKF 和混合同化方法(图中以“Hybrid”表示)在 130000 至 131000 步 x_1 的前 10 个集合成员的时间序列图。相比于混合同化方法,

EAKF 的分析场有较明显的距频,特别是当观测偏离真值较远时,这种距频更明显,它会将更多的噪声引入模式中;另一方面,混合同化方法的分析场相对来说更接近真场且更平滑,模式更稳定。由于加入了方差膨胀因子,观测的权重有所增加,造成了 EAKF 在观测偏离真值较远的地方产生了较大的振荡,从而增加模式的不稳定性。对混合同化方法而言,虽然膨胀因子间接地增加了观测的权重,但是对背景场的顺序滤波使得分析场中保留了背景误差的非线性信息(高阶矩),从而使得分析场更合理、更平滑,模式更稳定。图 3 的集合平均和集合偏差(集合平均 - 真值)的时间序列图更直观地给出了两种同化方法的同化结果。可以看出,几乎在每个分析步,混合同化方法的集合平均比 EAKF 更接近真值,这种差别在观测偏差较大的地方更加明显。从集合平均的偏差来看,混合同化方法的分析场更平滑。

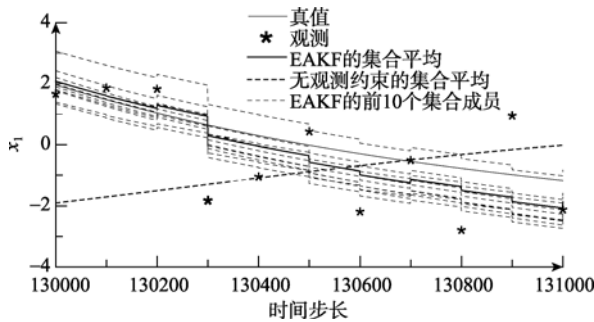


图 1 EAKF 在 130000 至 131000 步, x_1 的前 10 个集合成员的时间序列图

Fig. 1 Time series of first 10 members of x_1 of EAKF between time steps 130000 and 131000

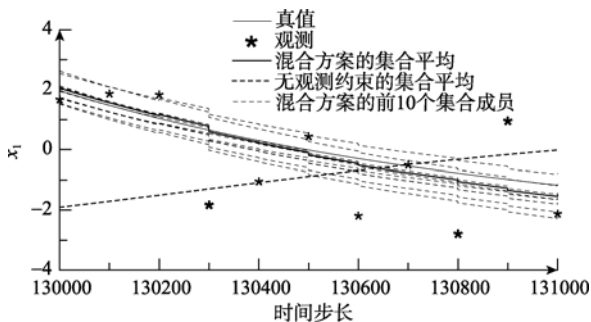


图 2 Hybrid 在 130000 至 131000 步, x_1 的前 10 个集合成员的时间序列图

Fig. 2 Time series of first 10 members of x_1 of hybrid between time steps 130000 and 131000

为了考察 2 种同化方法每个分析步的整体效果,图 4 给出了 130000 至 131000 个积分步,3 个模式变量总体均方根误差的时间序列图。可以很

明显地看出,在每个分析步,混合同化方法的均方根误差相对于 EAKF 而言在每个时间步上均得到了降低。

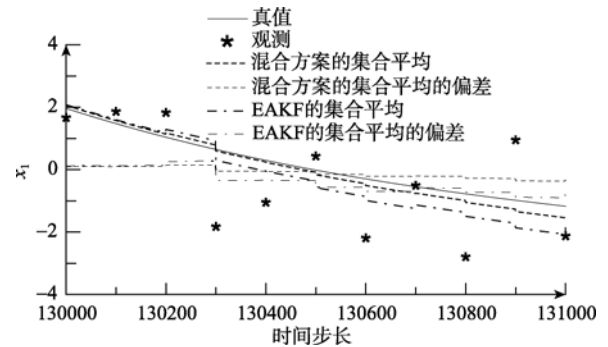


图 3 130000 至 131000 个积分步, x_1 的集合平均和集合偏差的时间序列图

Fig. 3 Time series of ensemble mean and ensemble bias of x_1 between time steps 130000 and 131000

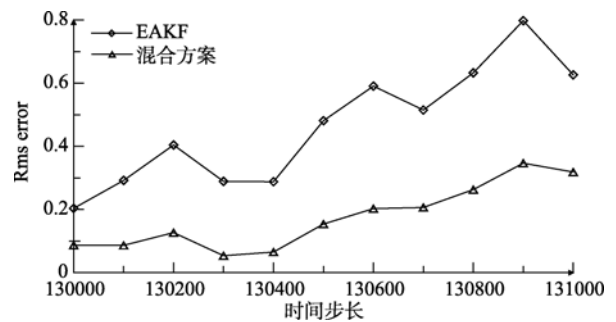


图 4 2 种同化方案在 130000 至 131000 步的均方根误差时间序列图

Fig. 4 Time series of RMS error between time steps 130000 and 131000

图 5—7 分别给出了 (x_1, x_3) 、 (x_1, x_2) 以及 (x_2, x_3) 在 140000 至 155000 个积分步的轨迹图,由于观测分布比较密集,故在图中未画出。在绝大部分时间步,混合数据同化方法与真值吻合得比较好,且分

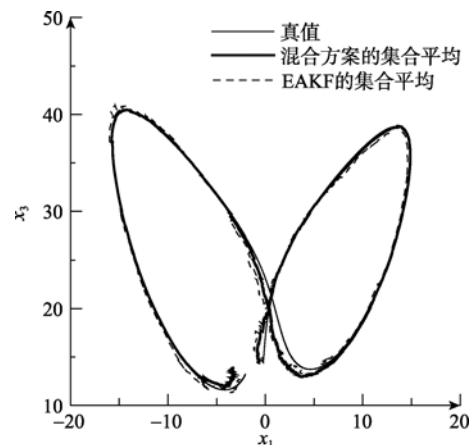


图 5 140000 至 155000 步, (x_1, x_3) 的时间序列图

Fig. 5 Time series of (x_1, x_3) between time steps 140000 and 155000

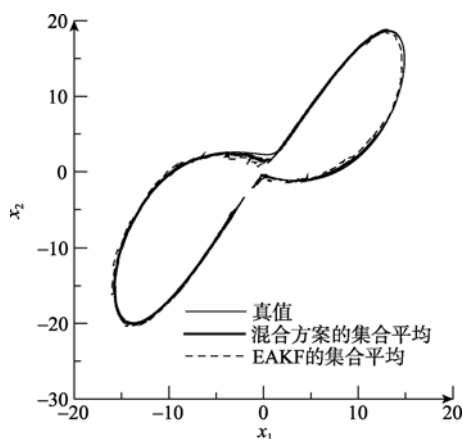


图 6 140000 至 155000 步, (x_1, x_2) 的时间序列图

Fig. 6 Time series of (x_1, x_2) between time steps 140000 and 155000

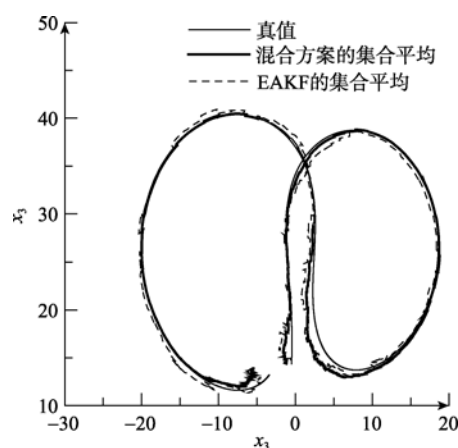


图 7 140000 至 155000 步, (x_2, x_3) 的时间序列图

Fig. 7 Time series of (x_2, x_3) between time steps 140000 and 155000

析结果较平滑。而 EAKF 则存在一些较明显的震荡, 这主要是在观测偏离真值较远的时候发生。当观测偏离真值足够大时, 这种振荡所引入的误差可能会使得模式无法运行下去; 相反, 混合数据同化方法则不会发生类似现象, 这主要得益于对背景场的顺序滤波的作用。

3 总结

本文基于集合滤波和三维变分数据同化方法发展了一种新的数据同化方法。这种同化方法结合了集合调整卡尔曼滤波和三维变分的优点, 即利用集合样本来构建流相关的背景误差协方差, 利用对背景场的顺序滤波来继承先验分布的高阶信息, 利用三维变分来得到分析场的最优估计。理想同化试验结果表明, 这种混合数据同化方法比 EAKF 更具有优越性。

对于实际的海洋或大气数值模式而言, 背景误差协方差矩阵的维数巨大, 此时采用本文的混合数据同化方法可能会耗费大量的计算机内存和计算时间, 因为每个分析步必须先显式地计算出背景误差协方差矩阵, 然后对其进行正交对角化。因此, 下一步的工作将寻找一种在已知集合样本的前提下, 将背景场转换到独立空间中的便捷有效的方法, 例如引入递归滤波的思想, 从而不需要显式地计算背景误差协方差矩阵。

参考文献

- [1] LEDIMET F X, TALAGRAND O. Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects[J]. Tellus, 1986, 38A: 97-110.
- [2] EVENSEN G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics[J]. J Geophys Res, 1994, 99: 10143-10162.
- [3] COURTIER P, THEPAUT J N, HOLLINGSWORTH A. A strategy for operational implementation of 4DVAR, using an incremental approach[J]. Quart J Roy Meteor Soc, 1994, 120: 1367-1387.
- [4] HAMILL T M, SNYDER C. A hybrid ensemble Kalman filter — 3D Variational analysis scheme[J]. Mon Wea Rev, 2000, 128: 2905-2919.
- [5] WHITAKER J S, HAMILL T M. Ensemble data assimilation without perturbed observations[J]. Mon Wea Rev, 2002, 130: 1913-1924.
- [6] CORAZZA M, KALNAY E, J PATIL D, et al. Use of the breeding technique in the estimation of the background error covariance matrix for a quasigeostrophic model[C]//AMS symposium on observations, data assimilation and probabilistic prediction. Orlando, Florida, 2002: 154-157.
- [7] ZUPANSKI M. Maximum likelihood ensemble filter: Theoretical aspects[J]. Mon Wea Rev, 2005, 133: 1710-1726.
- [8] LORENC A C. The potential of the ensemble Kalman filter for NWP: A comparison with 4D-VAR[J]. Quart J Roy Meteor Soc, 2003, 129: 3183-3203.
- [9] BUEHNER M. Ensemble-derived stationary and flow-dependent background error covariances: Evaluation in a quasi-operation NWP setting[J]. Quart J Roy Meteor Soc, 2005, 131: 1013-1043.
- [10] WANG X, HAMILL T M, WHITAKER J S, et al. A com-

- parison of hybrid ensemble transform Kalman filter — optimum interpolation and ensemble square-root filter analysis schemes[J]. *Mon Wea Rev*, 2007b, 135: 1055-1076.
- [11] BISHOP C H, ETHERTON B J, MAJUMDAR S J. Adaptive sampling with the ensemble transform kalman filter. Part I: Theoretical aspects [J]. *Mon Wea Rev*, 2001, 129: 420-436.
- [12] HUNT B R, COAUTHORS. Four-dimensional ensemble Kalman filtering[J]. *Tellus*, 2004, 56A: 273-277.
- [13] FERTIG E J, HARLIM J, HUNT B R. A comparative study of 4D-VAR and a 4D ensemble filter: Perfect model simulations with Lorenz-96[J]. *Tellus*, 2007, 59A: 96-100.
- [14] LIU CHENGSI, XIAO QINGNONG, WANG BIN. An ensemble-based four-dimensional variational data assimilation scheme. Part : Technical formulation and preliminary test[J]. *Mon Wea Rev*, 2008, 136: 3363-3372.
- [15] ZHANG F, MENG Z, JAMES A H. Coupling ensemble kalman filter with four-dimensional variational data assimilation[J]. *Adv Atmos Sci*, 2009, 26(1): 1-8.
- [16] WAN L, ZHU J, WANG H, et al. A “dressed” ensemble Kalman filter using the hybrid coordinate ocean model in the pacific[J]. *Adv Atmos Sci*, 2009, 26(5): 1042-1052.
- [17] COURTIER P, THEPAUT J N, HOLLINGSWORTH A. Variational methods[J]. *J Meteor Soc Japan*, 1997, 75: 211-218.
- [18] LORENC A C. Analysis methods for numerical weather prediction[J]. *Quart J Roy Meteor Soc*, 1986, 112: 1177-1194.
- [19] ANDERSON J L. An ensemble adjustment Kalman filter for data assimilation[J]. *Mon Wea Rev*, 2001, 129: 2884-2903.
- [20] HOUTEKAMER P L, MITCHELL H L. A sequential ensemble Kalman filter for atmospheric data assimilation[J]. *Mon Wea Rev*, 2001, 129: 123-137.
- [21] ANDERSON J L. A local least squares framework for ensemble filtering[J]. *Mon Wea Rev*, 2003, 131: 634-642.
- [22] LIU D C, NOCEDAL J. On the limited memory BFGS method for large scale optimization[J]. *Mathematical Programming*, 1989, 45: 503-528.
- [23] LORENZ E N. Deterministic non-periodic flow[J]. *J Atmos Sci*, 1963, 20: 130-141.
- [24] JAZWINSKI A H. Stochastic processes and filtering theory[M]. New York: Academic Press, 1970: 376.
- [25] ANDERSON J L. A method for producing and evaluating probabilistic forecasts from ensemble model integrations[J]. *J Climate*, 1996, 9: 1 518-1 530.