

基于区域逐步分析的集合变分资料同化方法*

吴祝慧¹⁾ 韩月琪^{2)3)†} 钟中³⁾ 杜华栋³⁾ 王云峰³⁾

1)(金陵科技学院, 南京 211169)

2)(解放军理工大学, 电磁环境效应与光电工程重点实验室, 南京 211101)

3)(解放军理工大学气象海洋学院, 南京 211101)

(2013年11月9日收到; 2013年12月17日收到修改稿)

集合变分数据同化方法的同化效果对集合样本容量具有很强的依赖性, 研究发现此问题的出现是因为其计算过程中分析增量被表示为集合扰动向量或其展开正交基向量的线性组合. 这样的处理方法虽然避免了计算梯度而引入伴随模式, 但是因为物理控制变量个数远大于集合样本容量, 就会导致物理量的同化分析值对集合样本容量很敏感. 根据此原因, 提出了区域逐步分析方法, 减小了同化分析区域内物理变量个数与集合样本容量数之间的比值, 使问题得到解决. 利用浅水方程模式进行资料同化数值试验表明, 基于区域逐步分析的集合变分资料同化方法可以得到较好的结果, 能明显提高同化的精度.

关键词: 区域逐步分析, 集合变分, 资料同化

PACS: 92.60.Wc

DOI: 10.7498/aps.63.079201

1 引言

大气或海洋中的资料同化是指利用已知的信息尽可能精确地描述大气或海洋的真实状态. 信息的来源主要有两个, 一是流体运动所遵循的物理定律; 二是常规和非常规的观测资料. 随着人类对防灾、减灾的迫切需要, 对大气、海洋高精度的预测、预报保障需求, 随着非常规观测资料(卫星遥感资料、雷达资料等)的增多、使用和依赖, 资料同化变得越来越重要. 从目前资料同化的应用来看, 同化方法的功能也不单单是为数值天气预报构造精确的初值, 它在观测系统的评估、资料再分析、目标性观测、集合预报、人工影响天气实施区定位、资料反演研究等方面也都有着广泛的应用.

现行的主要同化方法可分为两大类: 变分同化方法和集合滤波同化方法. 变分同化方法主要包含三维变分(3D-Var)和四维变分(4D-Var)^[1-3].

集合滤波同化方法主要包含 Kalman 滤波法, 集合 Kalman 滤波法(EnKF)及其变种和粒子滤波法^[4-9]. 两类方法有着各自的优势和问题: 变分方法能很好的处理复杂的非线性观测算子, 得到的分析场各变量能较好的满足模式动力和物理约束, 使得分析场更协调; 集合滤波方法能实现背景误差协方差矩阵的随流型演变, 改善特别是时空变化剧烈要素场的分析质量, 不需要花费大量精力完成和维护伴随方程, 使同化程序编写变得简单^[4-13].

近几年来, 汲取变分和集合滤波两种同化方法优点的集合变分数据同化方法成为研究的热点. Qiu等利用奇异值分解(SVD)技术, 提出了一种基于集合的4DVar方法(SVD-En4Dvar)^[14,15]. Liu等将将由预报集合估计的背景误差协方差用于4D-Var, 形成了基于集合的4D-Var方法(En4DVar)^[16]. Zupanski等基于极大似然估计原理提出了能够处理弱不可微问题的三维集合变分同化方法^[17]. 韩月琪等给出了用于大气 Ek-

* 江苏省自然科学基金(批准号: BK20131065)、中国博士后科学基金(批准号: 20110490185)、国家自然科学基金(批准号: 41175090, 41375106, 41105065, 41205073)和气象海洋学院基础理论研究基金资助的课题.

† 通讯作者. E-mail: hanyueqi1975@163.com

man 层湍流系数反演的集合变分梯度计算方案^[18]. Zhang 和张蕾等的试验表明, 集合变分数据同化方法可以得到比 EnKF 和 4D-Var 更好的同化效果. 同时也发现, 这类方法对模式和观测误差不敏感, 但是对预报样本数量的敏感性超过 EnKF, 预报样本过少则同化效果会明显下降^[19-22].

本文从集合变分数据同化方法的原理出发, 分析了此方法对样本敏感的根本性原因, 并且据此提出了解决敏感性的区域逐步分析方法, 然后利用浅水方程模式对提出的解决方案进行了资料同化数值试验, 结果表明: 相比于原来的同化方法, 基于区域逐步分析的集合变分资料同化方法可以得到较好的结果, 能明显提高同化的精度.

2 集合变分数据同化方法

针对三维顺序资料同化方法, 其目标泛函定义为

$$J(\mathbf{x}^k) = \frac{1}{2} (\mathbf{x}^k - \mathbf{x}_b^k)^T \mathbf{B}_k^{-1} (\mathbf{x}^k - \mathbf{x}_b^k) + \frac{1}{2} (\mathbf{y}^k - H_k(\mathbf{x}^k))^T \times \mathbf{O}_k^{-1} (\mathbf{y}^k - H_k(\mathbf{x}^k)), \quad (1)$$

其中 $\mathbf{x}_b^k$ 是在 k 时刻模式预报得到的值, $\mathbf{B}_k$ 为 k 时刻的背景误差协方差矩阵, $\mathbf{y}^k$ 为 k 时刻的观测资料, H_k 为 k 时刻的观测算子, $\mathbf{O}_k$ 为 k 时刻的观测误差协方差矩阵, 上标 T 表示向量转置. 资料同化过程的目的就是寻找和数值模式协调的分析值 $\mathbf{x}^k$, 使得目标泛函 (1) 最小.

Lorenc 在变分算法中用集合预报统计的扰动矩阵来预调节 (1) 式变分中的控制变量^[23,24]:

$$\mathbf{x}^k = \mathbf{x}_b^k + (\mathbf{X}')_b^k \mathbf{w}, \quad (2)$$

其中

$(\mathbf{X}')_b^k = \frac{1}{\sqrt{N}} (\mathbf{x}_1^k - \mathbf{x}_b^k, \mathbf{x}_2^k - \mathbf{x}_b^k, \dots, \mathbf{x}_N^k - \mathbf{x}_b^k)$, $\mathbf{x}_i^k$ ($i = 1, 2, \dots, N$) 为 k 时刻的 N 个集合预报值, N 为集合预报成员个数, $\mathbf{w} = (w_1, w_2, \dots, w_N)^T$ 为权重系数. 将 (2) 式代入 (1) 式, 利用 $\mathbf{B} \approx (\mathbf{X}')_b^k ((\mathbf{X}')_b^k)^T$, 可以得到控制变量为 $\mathbf{w}$ 的目标函数表达式

$$J(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} \left(\mathbf{y}^k - H_k \left(\mathbf{x}_b^k + (\mathbf{X}')_b^k \mathbf{w} \right) \right)^T$$

$$\times \mathbf{O}_k^{-1} \left(\mathbf{y}^k - H_k \left(\mathbf{x}_b^k + (\mathbf{X}')_b^k \mathbf{w} \right) \right). \quad (3)$$

引入信息向量

$$\mathbf{d}^k = \mathbf{y}^k - H_k(\mathbf{x}_b^k), \quad (4)$$

则

$$\begin{aligned} & \mathbf{y}^k - H_k \left(\mathbf{x}_b^k + (\mathbf{X}')_b^k \mathbf{w} \right) \\ &= \mathbf{d}^k - \left(H_k \left(\mathbf{x}_b^k + (\mathbf{X}')_b^k \mathbf{w} \right) - H_k(\mathbf{x}_b^k) \right). \end{aligned} \quad (5)$$

在 H_k 可微的情况下,

$$\begin{aligned} & H_k \left(\mathbf{x}_b^k + (\mathbf{X}')_b^k \mathbf{w} \right) - H_k(\mathbf{x}_b^k) \\ & \approx H_k(\mathbf{X}')_b^k \mathbf{w}, \end{aligned} \quad (6)$$

其中 H_k 为观测算子 H_k 的切线性算子. 将 (5), (6) 式代入 (3) 式可以得到

$$\begin{aligned} J(\mathbf{w}) &= \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} \left(\mathbf{d}^k - H_k(\mathbf{X}')_b^k \mathbf{w} \right)^T \\ & \times \mathbf{O}_k^{-1} \left(\mathbf{d}^k - H_k(\mathbf{X}')_b^k \mathbf{w} \right). \end{aligned} \quad (7)$$

由此可以得目标泛函对于控制变量 $\mathbf{w}$ 的梯度

$$\begin{aligned} \nabla_{\mathbf{w}} J &= \mathbf{w} + \left((\mathbf{X}')_b^k \right)^T (H_k)^T \\ & \times \mathbf{O}_k^{-1} \left(H_k(\mathbf{X}')_b^k \mathbf{w} - \mathbf{d}^k \right). \end{aligned} \quad (8)$$

为了避免使用伴随模式 $(H_k)^T$, 将 (8) 式变形为

$$\begin{aligned} \nabla_{\mathbf{w}} J &= \mathbf{w} + \left(H_k(\mathbf{X}')_b^k \right)^T \\ & \times \mathbf{O}_k^{-1} \left(H_k(\mathbf{X}')_b^k \mathbf{w} - \mathbf{d}^k \right). \end{aligned} \quad (9)$$

将背景场的扰动投影到观测空间上, 根据集合计算结果可以得

$$\begin{aligned} & H_k(\mathbf{X}')_b^k \\ & \approx \frac{1}{\sqrt{N}} (H_k(\mathbf{x}_1^k) - H_k(\mathbf{x}_b^k), H_k(\mathbf{x}_2^k) \\ & - H_k(\mathbf{x}_b^k), \dots, H_k(\mathbf{x}_N^k) - H_k(\mathbf{x}_b^k)). \end{aligned} \quad (10)$$

这样就避免了使用切线性算子 H_k 及其伴随算子 $(H_k)^T$, 从而求得 $\nabla_{\mathbf{w}} J$. 采用拟牛顿或共轭梯度优化迭代算法, 将 $J(\mathbf{w})$ 达到极小时的 $\mathbf{w}$ 代入 (2) 式, 即可得到同化分析值. 由于 $(\mathbf{X}')_b^k$ 为原数值模式集合预报值相对于背景场预报值的扰动值矩阵, 由 (2) 式可知同化分析值为背景场与分析增量之和, 而分析增量是扰动值集合成员的加权线性组合, 因此形成的同化分析值与原数值模式是协调的.

3 集合变分数据同化方法对样本容量敏感的原因分析

由上述的集合变分数据同化方法计算过程可知, 同化结果的分析增量为集合预报扰动向量的加权线性组合, 根据(2)式, 即

$$\mathbf{x}^k - \mathbf{x}_b^k = (\mathbf{X}')_b^k \mathbf{w}. \quad (11)$$

将其展开, 可得

$$\begin{aligned} \mathbf{x}^k - \mathbf{x}_b^k = & w_1 \Delta \mathbf{x}_1^k + w_2 \Delta \mathbf{x}_2^k + \cdots \\ & + w_N \Delta \mathbf{x}_N^k, \end{aligned} \quad (12)$$

其中 $\Delta \mathbf{x}_i^k = \frac{1}{\sqrt{N}} (\mathbf{x}_i^k - \mathbf{x}_p^k)$, ($i = 1, 2, \cdots, N$), 为了更清楚的认识其本质, 将(12)式写成分量的形式:

$$\begin{aligned} x_1^k - x_{1b}^k &= w_1 \Delta x_{11} + w_2 \Delta x_{12} + \cdots + w_N \Delta x_{1N}, \\ x_2^k - x_{2b}^k &= w_1 \Delta x_{21} + w_2 \Delta x_{22} + \cdots + w_N \Delta x_{2N}, \\ x_M^k - x_{Mb}^k &= w_1 \Delta x_{M1} + w_2 \Delta x_{M2} + \cdots \\ &+ w_N \Delta x_{MN}. \end{aligned} \quad (13)$$

其中下标 M 为同化分析值变量的个数. 可以发现(13)式为一个线性代数方程组的形式, 在同化分析结果惟一的情况下, 集合变分数据同化方法就可以看作在数值模式和观测资料约束条件下, 求此线性代数方程组的解向量 $\mathbf{w}$, 从而使得目标泛函达到最小. 对于大气或海洋资料同化来说, 一般分析变量数量都比较大, 因此在分析值变量的个数 M 远大于集合预报成员个数 N 的情况下, 集合变分方法的同化结果必然对集合样本数比较敏感, 集合样本过少甚至得不到理想的同化结果, 导致同化失败.

4 区域逐步分析的集合变分资料同化方法

从前面的分析过程可知, 集合变分方法同化结果对样本容量的敏感性原因是由于分析值变量的个数远大于集合预报的成员个数, 因此降低两者之间的数量比就成为解决问题的关键. 解决的途径主要有两个: 第一就是增加集合预报成员的个数, 这样就要增加数值模式向前集合计算的次数, 计算量增加很大; 第二就是降低分析值变量的个数, 为此

我们提出了基于区域逐步分析的集合变分资料同化方法, 减小了同化分析区域内物理变量个数与集合样本容量数之间的比值, 使问题得到解决.

区域逐步分析方法具体计算步骤如下:

1) 第一次同化分析区域取整个模式区域(如果整个模式区域比较大, 也可以取相比模式区域小的分析区域), 经同化观测资料后得到包含观测场大尺度信息的分析场 $\mathbf{x}_{a1}$.

2) 将 $\mathbf{x}_{a1}$ 作为第二次同化分析的初始背景场, 缩小同化分析区域范围, 在整个模式区域内按照缩小的同化分析区域大小采用依次递进的方式进行逐步分析, 直到整个模式区域范围分析完为止, 分析过程见示意图1所示. 区域逐步分析过程中, 为了保证同化分析变量的连续性及协调性, 将此次前面已分析区域的同化结果作为后面待分析区域同化的背景场. 按照设定的分析区域范围根据已定的方向依次递进对整个模式区域进行同化分析, 因为在此次同化分析过程中, 后面分析区域内的背景场使用了和前面分析区域相重叠部分的同化结果, 因此当模式整个区域范围按照此次设定的分析区域范围被同化分析完后, 得到的分析场即为此次分析整个区域的同化解. 经此次同化观测资料后得到包含观测场中尺度信息的分析场 $\mathbf{x}_{a2}$.

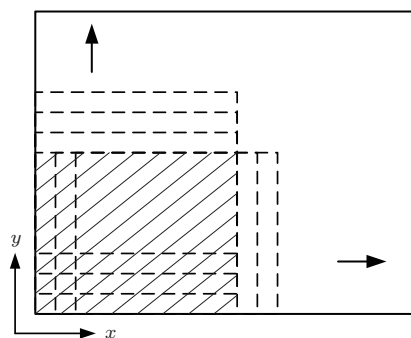


图1 集合变分资料同化方法的区域逐步分析示意图(阴影部分表示区域逐步分析方法中第二次同化区域范围的大小)

3) 将 $\mathbf{x}_{a2}$ 作为第三次同化分析的初始背景场, 再次缩小同化分析区域范围, 进行和2)步骤一样的观测资料同化分析过程, 得到包含观测场小尺度信息的分析场 $\mathbf{x}_{a3}$.

4) 不断缩小同化区域范围进行同化分析, 直到同化结果满足规定条件.

在这要说明的是, 由于整个模式区域范围内观测资料及模式集合预报变量的约束, 区域逐步分析

过程中第一次区域范围选择取的大一些可以从观测资料获得比较光滑的大尺度场信息. 然后逐步缩小区域范围并进行同化分析, 因为同化区域范围的缩小, 减小了区域内物理变量个数与集合样本容量数之间的比值, 而且观测资料的个数相对减小, 所以在得到光滑大尺度场观测信息的基础上可以逐步恢复分析场中的较小尺度分量. 分析过程之所以先由大区域然后逐渐过渡到几个小区域而不是直接分成几个小区域进行同化分析, 它是为了让同化分析得到的较小尺度分量是叠加在一个合理的大尺度观测信息场之上, 这样就保证了同化分析物理变量之间及变量空间分布上的协调性. 如果直接分成几个小区域进行同化分析, 这样虽然降低了小区域内物理变量个数与集合样本容量数之间的比值, 通过同化得到了较小尺度分量观测信息, 但是因为这些小尺度分量和大尺度变量之间的不协调性, 在

数值模式向前积分过程中这些获取的中小尺度信息很快就会被淹没, 起不到资料同化应有的作用.

5 数值试验

用一个二维浅水方程模式和模拟观测来检验基于区域逐步分析的集合变分资料同化方法的同化效果, 评估此方法在降低同化结果对集合样本容量依赖性作用方面的可行性.

考虑地形的浅水方程模式为

$$\begin{aligned} \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} - fv + g \frac{\partial h}{\partial x} &= 0, \\ \frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + fu + g \frac{\partial h}{\partial y} &= 0, \\ \frac{\partial h}{\partial t} + \frac{\partial(uh)}{\partial x} + \frac{\partial(vh)}{\partial y} - \frac{\partial(uh_s)}{\partial x} &- \frac{\partial(vh_s)}{\partial y} = 0, \end{aligned} \quad (14)$$

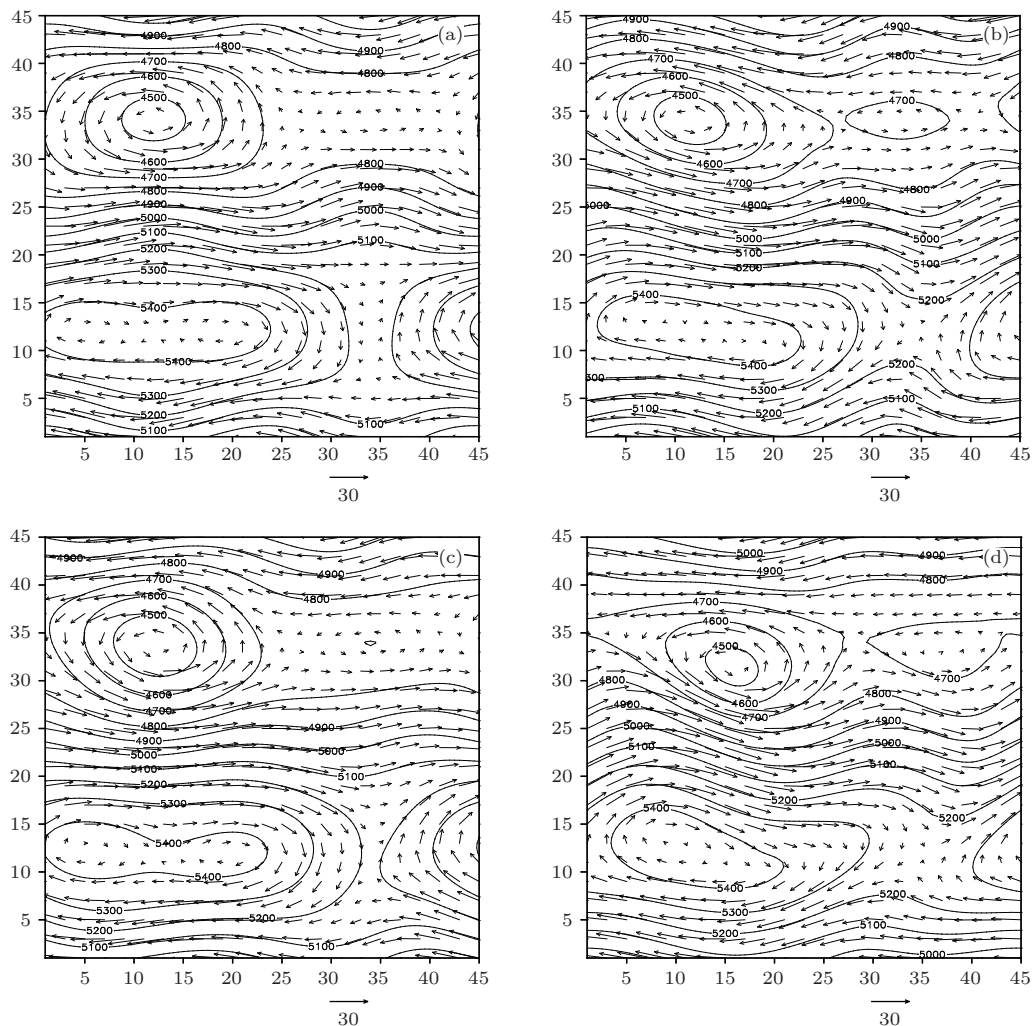


图2 在 $t = 0$ h 时刻初值 (a)、初始集合的平均值 (c), $t = 72$ h 观测场 (b) 及背景场 (d). 水平为东西方向, 垂直为南北方向, 等值线为高度场/m, 矢量为风场/ $\text{m} \cdot \text{s}^{-1}$

其中 $f = 7.272 \times 10^{-5} \text{ s}^{-1}$, h_s 是地形高度, 试验中取为

$$h_s = h_0 \sin(4\pi x/L_x) [\sin(\pi y/L_y)]^2. \quad (15)$$

在用于产生真实物理量场的模式中不考虑地形的影响, 即取 $h_0 = 0 \text{ m}$, 在用于预报和同化的有误差的模式中取 $h_0 = 250 \text{ m}$; $L_x = L_y = 13\,200 \text{ km}$, 是计算区域两个方向的边长, 网格距 $\Delta x = \Delta y = 300 \text{ km}$, 网格点数是 45×45 , 时间步长 $\Delta t = 360 \text{ s}$, 在 x 和 y 方向都取周期性边界条件. 真实物理量场由积分上述模式得到, 初始各物理量场为 [22]

$$\begin{aligned} h = & 5000 + 360 \sin\left(\frac{2\pi y}{L_y}\right) \\ & + 120 \sin\left(\frac{2\pi x}{L_x}\right) \sin\left(\frac{2\pi y}{L_y}\right) \\ & + 60 \cos\left(\frac{4\pi x}{L_x}\right) \\ & \times \sin[4\pi(y - 6\Delta y)/L_y], \end{aligned} \quad (16)$$

$$u = -f^{-1}g \frac{\partial h}{\partial y}, \quad v = f^{-1}g \frac{\partial h}{\partial x}. \quad (17)$$

由上述初值(图 2(a))出发积分不考虑地形影响的模式到同化试验时刻(第 72 小时), 将该时刻的计算结果作为“观测场”(图 2(b)); 由上述初值出发向前积分考虑地形影响的模式 48 h, 将每小时得到的积分结果保存起来作为集合预报的初始集合(因此集合成员个数为 48); 将初始集合的平均值(图 2(c))向前积分考虑地形影响的模式 72 h, 将其积分结果(图 2(d))作为同化试验时刻的背景场.

假定在预报区域内共有 225 个观测点, 观测点位于计算网格上且均匀分布(见图 3 所示). h , u 和 v 的观测误差均方根误差分别为 4.23 m , $0.27 \text{ m} \cdot \text{s}^{-1}$ 和 $0.48 \text{ m} \cdot \text{s}^{-1}$. 资料同化过程中优化算法采用有限内存拟牛顿算法 [25].

首先进行第一组数值试验, 选用传统方法将整个模式区域范围内的变量同时作为同化分析变量, 即同化分析区域范围内的格点数为 45×45 . 目标函数随同化迭代次数的变化见图 4 所示, 可以看到目标函数由初始的 3.892×10^4 降到最后的 1.276×10^4 , 下降了三分之二左右, 迭代进行 10 次以后, 目标函数基本趋于恒定, 不再下降. 第一组试验的同化结果及其相对于背景场的同化分析增量见图 5 所示.

将图 5(a) 与图 2(b) 和 (d) 对比可以发现, 同化分析场结果相对于背景场改善了许多, 这说明这种同化方法是可行有效的. 但同时还可以发现, 同化结果中槽线和脊线的强度、南部高压及北部低压的覆盖范围都还存在一定的偏差. 分析图 5(b), 即同化结果相对于背景场的同化分析增量, 可以看到此组试验同化过程中获得的分析增量相对来说尺度比较大而且比较平滑. 主要原因就是在这组同化分析试验中, 将整个模式区域范围内的变量同时作为同化分析变量, 而且观测资料也是整个模式区域范围内的, 同化分析增量为集合扰动向量的线性组合, 为了平衡模式集合预报各向量及整个区域内的观测资料约束, 在集合成员个数远远小于分析变量个数的情况下, 因此只能得到观测资料比较平滑的大尺度信息.

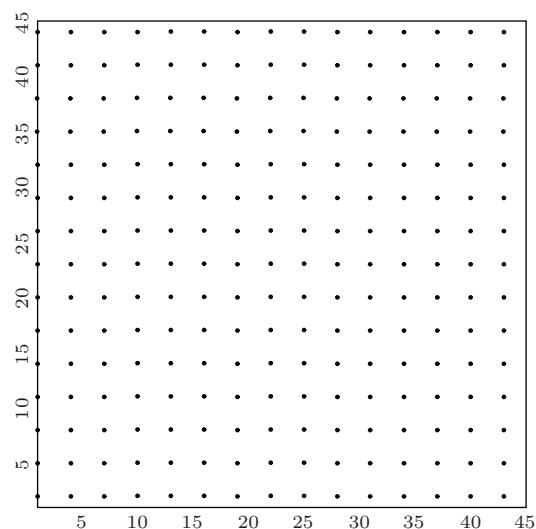


图 3 观测点分布 (水平为东西方向, 垂直为南北方向)

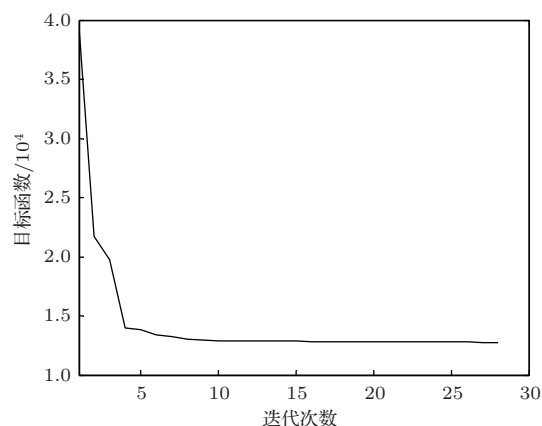


图 4 第一组试验中目标函数随迭代次数的变化

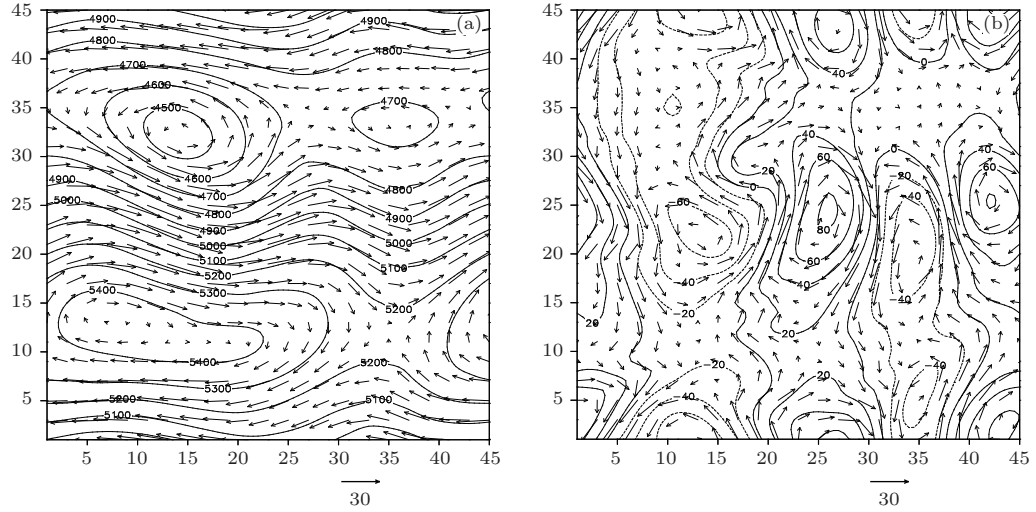


图5 第一组数值试验中同化分析结果(a)及其相对于背景场的同化分析增量(b). 水平为东西方向, 垂直为南北方向, 等值线为高度场/m, 矢量为风场/ $\text{m}\cdot\text{s}^{-1}$

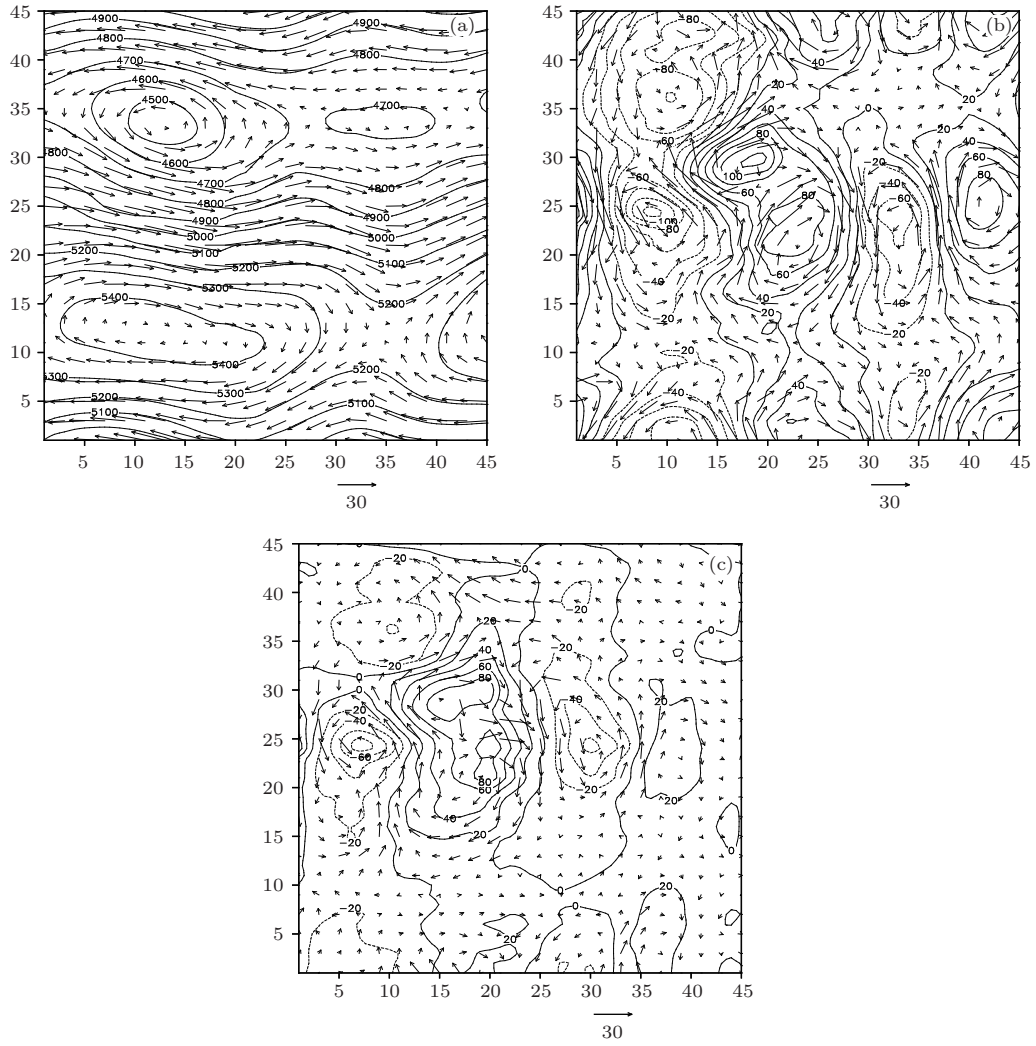


图6 第二组数值试验中同化分析结果(a), 最后同化结果相对于初始背景场的同化分析增量(b)和第二步分析得到的同化分析增量(c). 水平为东西方向, 垂直为南北方向, 等值线为高度场/m, 矢量为风场/ $\text{m}\cdot\text{s}^{-1}$

对于第二组数值试验, 采用区域逐步分析的集合变分资料同化方法. 首先进行和第一组试验相同的同化过程, 完成后将它的同化结果作为第二步同化分析的初始背景场. 第二步同化分析区域范围的格点数选为 20×20 . 第二组试验的同化分析及各分析增量见图 6 所示.

将图 6(a) 和图 2(b) 及图 5(a) 对比可以发现, 第二组试验同化结果中槽线和脊线的强度、南部高压及北部低压的覆盖范围与第一组同化试验结果相比都得到了较大的改善, 更接近于观测场. 由图 6(b) 和 (c) 的同化分析增量, 可以看出通过区域逐步分析方法可以得到比原来第一组试验传统方法更多的中小尺度信息. 在区域逐步分析方法中, 第一次将整个模式区域作为同化分析区域, 因为样本容量的限制, 经同化观测资料后得到包含观测场大尺度信息的分析场; 在此基础上, 缩小同化分析区域范围, 采用依次递进的方式进行逐步同化分析, 得到包含观测场中、小尺度信息的分析场. 使用集合变分数据同化方法, 经过逐步分析过程降低了对样本容量的依赖性, 使得同化结果包含更丰富的观测资料信息.

此外, 为了更进一步比较不同组试验的结果, 还利用下面定义的高度场 h 和场 u 和 v 的均方根误差来评价不同组试验的同化效果:

$$\sigma_{\mu} = \sqrt{\frac{1}{N_g} \sum_{i=1}^{N_g} (\mu_i^a - \mu_i^t)^2}, \quad (18)$$

其中 μ 分别为 h , u 和 v , N_g 为网格点总数 45×45 , 上标 a 表示同化分析场, t 表示模拟真实场. 表 1 为不同组试验中的同化分析结果, 从表中可以看出, 采用区域逐步分析方法的第二组试验结果最好, 无论是高度场还是场, 它们的均方根误差都是最小, 这说明集合变分同化的区域逐步分析方法是可行的.

表 1 浅水方程组数值试验结果

	σ_h/m	$\sigma_u/\text{m}\cdot\text{s}^{-1}$	$\sigma_v/\text{m}\cdot\text{s}^{-1}$
背景场	45.77	3.02	4.92
第一组	26.37	1.70	2.80
第二组	14.68	1.43	1.86

6 结 论

集合变分数据同化方法的同化效果对集合样本容量具有很强的依赖性. 本文通过分析此同化方

法的计算过程, 指出因这种方法分析增量被表示为集合扰动向量或其展开正交基向量的线性组合导致了其结果对样本容量的敏感性. 针对此问题的产生原因, 提出了区域逐步分析方法, 减小了同化分析区域内物理变量个数与集合样本容量数之间的比值, 使问题得到解决. 然后利用浅水方程模式进行资料同化数值试验表明, 基于区域逐步分析的集合变分资料同化方法可以得到较好的结果, 能明显提高同化的精度.

参考文献

- [1] Huang X Y, Xiao Q, Barker D M 2009 *Mon. Wea. Rev.* **137** 299
- [2] Cao X Q, Huang S X, Du H D 2008 *Acta Phys. Sin.* **57** 1984 (in Chinese)[曹小群, 黄思训, 杜华栋 2008 物理学报 **57** 1984]
- [3] Cao X Q, Zhang W M, Song J Q 2012 *Acta Phys. Sin.* **61** 020507 (in Chinese)[曹小群, 张卫民, 宋君强 2012 物理学报 **61** 020507]
- [4] Evensen G 1994 *J. Geophys. Res.* **99** 10143
- [5] Houtekamer P L, Mitchell H L 2001 *Mon. Wea. Rev.* **129** 123
- [6] Lorenc A C 2003 *Q. J. R. Meteorol. Soc.* **129** 3183
- [7] VanLeeuwen P J 2010 *Q. J. R. Meteorol. Soc.* **136** 1991
- [8] Leng H Z, Song J Q, Cao X Q 2012 *Acta Phys. Sin.* **61** 070501 (in Chinese)[冷洪泽, 宋君强, 曹小群 2012 物理学报 **61** 070501]
- [9] Du Z C, Tang B, Li K 2006 *Acta Phys. Sin.* **55** 999 (in Chinese)[杜正聪, 唐斌, 李可 2006 物理学报 **55** 999]
- [10] Snyder C, Zhang F 2003 *Mon. Wea. Rev.* **131** 1663
- [11] Kalnay E, Li H, Miyoshi T 2007 *Tellus Ser. A* **59** 758
- [12] Cao X Q, Song J Q, Zhang W M 2011 *Acta Phys. Sin.* **60** 070511 (in Chinese)[曹小群, 宋君强, 张卫民 2011 物理学报 **60** 070511]
- [13] Wang S C, Li Y, Zhang W M 2011 *Acta Phys. Sin.* **60** 099203 (in Chinese)[王舒畅, 李毅, 张卫民 2011 物理学报 **60** 099203]
- [14] Qiu C J, Zhang L, Shao A M 2007 *Sci. China D* **50** 1232
- [15] Chou J F, Li J P, Gao L 2006 *Chin. Phys.* **15** 882
- [16] Liu C S, Xiao Q, Wang B 2008 *Mon. Wea. Rev.* **136** 3363
- [17] Zupanski M, Navon I M, Zupanski D 2008 *Q. J. R. Meteorol. Soc.* **134** 1039
- [18] Han Y Q, Zhong Z, Wang Y F 2013 *Acta Phys. Sin.* **62** 049201 (in Chinese)[韩月琪, 钟中, 王云峰 2013 物理学报 **62** 049201]
- [19] Wang B, Zhao Y 2005 *Acta Meteor. Sin.* **63** 694 (in Chinese)[王斌, 赵颖 2005 气象学报 **63** 694]
- [20] Zhang F Q, Zhang M, James A 2009 *Adv. Atmos. Sci.* **26** 1
- [21] Zhang L, Qiu C J, Zhang S W 2009 *Acta Meteor. Sin.* **67** 1124 (in Chinese)[张蕾, 邱崇践, 张述文 2009 气象学报 **67** 1124]

- [22] Shao A M, Qiu X B, Qiu C J 2011 *Plateau Meteo.* **30** 583 (in Chinese)[邵爱梅, 邱晓滨, 邱崇践 2011 高原气象 **30** 583]
 [23] Jazwinski A H 1970 *Stochastic Processes and Filtering Theory* (1st Ed.) (New York: Academic Press) p376
 [24] Lorenc A C 1992 *Q. J. R. Meteorol. Soc.* **118** 569
 [25] Liu D C, Nocedal J 1989 *Math. Programming* **45** 503

Ensemble variational data assimilation method based on regional successive analysis scheme*

Wu Zhu-Hui¹⁾ Han Yue-Qi^{2)3)†} Zhong Zhong³⁾ Du Hua-Dong³⁾ Wang Yun-Feng³⁾

1) (Jinling Institute of Technology, Nanjing 211169, China)

2) (National Key Laboratory on Electromagnetic Environmental Effects and Electro-optical Engineering, PLA University of Science and Technology, Nanjing 211101, China)

3) (Institute of Meteorology and Oceanography, PLA University of Science and Technology, Nanjing 211101, China)

(Received 9 November 2013; revised manuscript received 17 December 2013)

Abstract

The ensemble variational data assimilation method may be subject to significant uncertainties due to the size of forecast ensemble. We found that this problem occurs because the analysis increment of this method is expressed as a linear combination of ensemble perturbation vectors or expansion of the orthogonal basis vectors. Though this method avoids introducing adjoint model while calculating the gradient of object function, the number of physical control variables is much larger than the sample size of forecast ensemble, which causes the assimilation results to be sensitive to the number of ensemble members. For this reason, the regional successive analysis scheme of ensemble variational method is proposed. By this scheme, the ratio between the number of physical control variables in analysis region and the sample size is decreased, so that it is expected that the problem can be solved. The results of numerical experiments using shallow water model show that the regional successive analysis scheme can give better assimilation results than traditional method, and the analysis precision is improved appreciably.

Keywords: regional successive analysis, ensemble variational method, data assimilation

PACS: 92.60.Wc

DOI: 10.7498/aps.63.079201

* Project supported by the Natural Science Foundation of Jiangsu Province, China (Grant No. BK20131065), the China Postdoctoral Science Foundation (Grant No. 20110490185), the National Natural Science Foundation of China (Grant Nos. 41175090, 41375106, 41105065, 41205073), the Basic Theory Research Program of Institute of Meteorology and Oceanography, China.

† Corresponding author. E-mail: hanyueqi1975@163.com