

一种基于预报集合的降维资料同化方法的 数值试验研究

邵爱梅 邱崇践*

兰州大学大气科学学院, 兰州 730000

摘 要 对 Qiu 和 Chou (2006) 提出的一种基于预报集合的降维资料同化方法 (4DSVD) 给出了可行的实施方案, 利用中尺度模式 MM5 产生的模拟资料进行数值试验并将其与 MM5/3DVAR 的同化结果进行比较, 分析了不同的观测误差和观测点密度对同化结果的影响。试验表明: (1) 和 3DVAR 相比该方法能更好地从有观测的变量推断无观测的变量 (从温度的观测推测风和比湿); (2) 该方法可以相当有效地滤除观测噪音; (3) 该方法具有更好的将观测信息扩展到资料空缺地区的能力。

关键词 降维 预报集合 资料同化 奇异值分解

文章编号 1006-9895 (2007) 04-0675-10

中图分类号 P468

文献标识码 A

A Numerical Study of an Ensemble-Based Reduced-Dimensional Data Assimilation Method

SHAO Ai-Mei and QIU Chong-Jian

College of Atmosphere Science, Lanzhou University, Lanzhou 730000

Abstract One of the major difficulties in data assimilation is that the degree of freedom in NWP model far exceeds the number of observations collected at any single time, so the problem is usually underdetermined. In order to reduce the degree of underdetermined problem, Qiu and Chou (2006) suggested that the solution should be restricted to the attractor of atmosphere dynamic equations in the phase space. Based on this idea, an ensemble-based reduced-dimensional data assimilation (ERDA) method was developed. In it, one set of short-range forecasts (include the simulated observations) is obtained with the different perturbation initial conditions. The ensemble of atmospheric state is generated by sampling in a time window. Then the singular value decomposition (SVD) technique is used to construct the orthogonal basis vectors from the ensemble. The assimilation is performed in the space spanned by a few leading basis vectors. In this paper the effect of performing data assimilation using ERDA is examined via comparison with three-dimensional variational technique (3DVAR) in an idealized environment. The experiments are performed using the Pennsylvania State University-NCAR Mesoscale Model version 5 (MM5) with $99 \times 99 \times 31$ grid points and the simulated observations. A set of 24-hour forecasts with 20 members is used to generate 180 samples of the ensemble following the precept of selecting a snapshot every 3 hours. The 3DVAR scheme used in the experiments is one designed for the MM5 model. In order to ensure a fair comparison, ERDA and 3DVAR both use the same observational data at a single time level only. Therefore, the four-dimensional data assimilation method proposed by Qiu and Chou (2006) is reduced to a three-dimensional scheme. In all experiments, it is assumed that only

收稿日期 2005-12-29 收到, 2006-04-27 收到修改稿

资助项目 国家自然科学基金资助项目 40505022、40575049

作者简介 邵爱梅, 女, 1976 年出生, 理学博士, 副教授, 主要从事资料同化和多普勒雷达资料的反演研究。E-mail: sam@lzu.edu.cn

* 通讯作者 E-mail: qiucj@lzu.edu.cn

temperature observations are available at sparsely selected grid points. The numerical experiments are performed with the different number of truncated basis vectors, observed density and error level in the observation. The results indicate that (i) the quality of the analysis by ERDA is better evidently than that by 3DVAR. Though 3DVAR can produce a little more precise analysis for the temperature than that by ERDA in most cases, ERDA is more effectual than that of 3DVAR to recover winds, pressure, and humidity these unobserved variables from observed temperature; (ii) in ERDA the observed information can be reasonably extrapolated to the data-void areas by the basis vectors, which represent the basic special configuration of the atmospheric state (increment); (iii) in ERDA, the analysis increment is restricted in a space spanned by a few singular vectors and these vectors are dynamically constrained fields produced by the model while random noises in observations do not fit any of the dynamic constraints. Therefore the errors in the observation have even less damage to the analysis. Nevertheless some useful information in the observation is also lost. So the intensity of analysis increment is light usually than actual one. This is a weakness of ERDA.

Key words reduced-dimension, forecast ensemble, data assimilation, single value decomposition

1 引言

大气资料同化的最大困难是数值模式的自由度远远超过初始时刻的观测资料数,从数学上讲是一个欠定的问题。运用各种遥感观测资料可以在一定程度上克服这一困难,然而这些遥感资料提供的大气信息一般都不够完整,比如气象卫星提供的信息主要限于温度和湿度,而 Doppler 雷达观测的只是速度的一个分量,这仍然不可能从根本上解决欠定的问题。因此,一个好的同化方法应该有能力将观测信息从有限的观测空间延展到整个模式空间。目前,业务上广泛运用的三维变分同化方法(3DVAR)是通过两种途径来实现这种延展,其一是利用背景(预报)场误差协方差矩阵使得一个点的观测能够影响到附近地区;其二是利用适当的平衡方程在不同变量之间建立联系。尽管已经有不少有关依流型而变的误差协方差的研究,目前业务运行的 3DVAR 系统中,背景场误差协方差一般仍然被假定为均匀各向同性并基本不随时间变化的^[1, 2],这显然不符合实际。在 3DVAR 中运用的平衡方程,比如风场和高度场之间的地转或旋转风平衡也有很大的不精确性。另一种有效的资料同化方法是四维变分同化方法(4DVAR)^[3~5],它利用多时刻观测提供的信息来弥补观测在空间上的不足,将资料同化问题表述为微分方程的反问题,通过将预报和观测拟合给出初始场的最优估计,是一种很具吸引力的方法,在业务运用中也有良好表现。然而 4DVAR 需要耗费大量的计算时间,模式的高度非线性和间断性质使线性化遇到困难^[6~8],

书写伴随模式程序的繁杂工作也让人望而生畏。这些原因妨碍了 4DVAR 的广泛运用。由 Evensen^[9]提出的集合卡尔曼滤波(EnKF)方法利用 Monte Carlo 技术可以方便地考虑依流型而变的误差协方差,也不要求切线性模式,已经受到广泛的关注并成功地进行了很多试验^[10~13],不过离业务运用还有一段距离。许多研究指出,大气运动要受到一定的物理规律支配,它最终会收缩到一个维数很低的吸引子上,其实际的维数远远小于大气模式的维数^[14, 15]。据此, Qiu 和 Chou^[16]认为,一个合理的做法应该是在大气模式吸引子上求解资料同化问题。根据这一思想他们提出了一种基于奇异向量分解(SVD)的四维资料同化方法。这一方法建议利用奇异向量分解技术从一个模式输出的样本集合中产生支撑起模式大气吸引子的正交基向量,将观测投影到这些基向量张成的空间得到分析场,以下我们将这种方法称为基于预报集合的降维资料同化方法(ERDA, ensemble-based reduced-dimensional data assimilation)。文献[16]主要从理论上阐述 ERDA 的合理性和可行性,但该方法的有效性还需要通过试验来检验。作为这种检验的第一步,本文将利用模式资料进行数值试验,对 ERDA 和 3DVAR 的同化结果进行比较。为此,本文将 ERDA 简化为一个三维同化方案,并给出一种产生样本集合的方法。下一节首先介绍方法的实施步骤,第 3 节说明试验的设计,第 4 节给出试验的结果,最后是总结和讨论。

2 方法

文献[16]阐明了 ERDA 的理论基础并给出了

一种实施方法,对此首先作一简要说明。设 P 是数值模式的格点数和变量数之积,一个时刻的大气状态是 P 维列向量 $\mathbf{A} \in \mathbf{R}^P$, 即数值模式的状态空间是 P 维向量空间 \mathbf{R}^P 。进一步假设在同化时间窗 $(0, T)$ 内有 Q 个观测值,将其表示为列向量 \mathbf{d} , $\mathbf{d} \in \mathbf{R}^Q$ 。令 $N=P+Q$, 于是,一个时刻的模式大气状态和所有 Q 个观测一起可以表示为一个 N 维列向量 $\mathbf{U} = (\mathbf{d}, \mathbf{A}) \in \mathbf{R}^N$ 。我们将其称为扩展的大气状态。按照非线性大气动力学的理论, \mathbf{A} 应该落入维数为 S 的吸引子 Ω_S 上,而且 $S \ll N$, 就是说大气实际的自由度要远远小于表面上的自由度,在大气吸引子上求解而不是在计算网格点上求解应该是更合理的资料同化方法。文献[16]认为通过对大气模式作长期积分能够得到足够多的大气样本,它们都应该在模式大气的吸引子上,利用 SVD 技术从这组大气样本中可以得到支撑起吸引子 Ω_S 的基向量,资料同化则在这组基向量张成的空间中求解。具体做法是:考虑一个同化时间窗 $(0, T)$ 内有 M 次观测,其中在时刻 t_i 有 r_i 个观测记为 $\mathbf{d}_{i,j} (j=1, 2, \dots, r_i)$, 它可以是模式变量也可以是非模式变量。从 $t=t_0 < 0$ 时刻的大气状态 \mathbf{A}_0 出发,积分数值模式得到 $(0, T)$ 期间的 M 个大气状态 $\mathbf{A}_i = F_i \mathbf{A}_0 (i=1, 2, \dots, M)$, 这里 F_i 是预报算子。根据模式产生的大气状态可以计算观测量的模拟值 $d_{i,j}^{\text{simu}} = H_{i,j}(\mathbf{A}_i) (i=1, 2, \dots, M, j=1, 2, \dots, r_i)$, 这里 H 是观测算子,它实现从模式变量向观测变量的映射。不妨将待定的大气状况设定为 \mathbf{A}_1 , 由于任意 $\mathbf{A}_i, i > 1$, 可以从 \mathbf{A}_1 得到,因此也可以将观测量的模拟值记为 $d_{i,j}^{\text{simu}} = G_{i,j} \mathbf{A}_1$, 这里算子 $G_{i,j} = H_{i,j} F_i$ 。设在期间 $(0, T)$ 内的全体观测值的数量为 Q , 即 $\sum_{i=1}^M r_i = Q$, 将相应的模拟值记为 Q 维列向量 $\mathbf{d}^{\text{simu}} \in \mathbf{R}^Q$ 。 \mathbf{A}_1 和 \mathbf{d}^{simu} 一起构成 $\mathbf{U}_1 = (\mathbf{d}_1, \mathbf{A}_1) \in \mathbf{R}^N$, 这里 $N=P+Q$ 。从 \mathbf{A}_M 继续积分大气模式到时间 $t=kT, k=2, 3, \dots, K$, 按类似的方法得到一组扩展的大气状态样本 $\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_K$, 这样就得到一个 $\mathbf{R}^P \rightarrow \mathbf{R}^N$ 的映射。利用 SVD 技术从这一样本集合中得到一族正交向量,假设它们已按照相应的特征值由大到小的顺序排列,截取前 r 个向量 $\mathbf{b}_i, i=1, 2, \dots, r$, 作为基底,则对 Ω_S 上的任一向量 \mathbf{U} 近似有

$$\mathbf{U} \approx \sum_{i=1}^r \alpha_i \mathbf{b}_i. \quad (1)$$

任意一个基向量 \mathbf{b}_i 可以表示为

$$\begin{aligned} \mathbf{b}_i &= (b_{i,1}, b_{i,2}, \dots, b_{i,Q}, b_{i,Q+1}, b_{i,Q+2}, \dots, b_{i,N})^T \\ &\equiv (\mathbf{D}_i, \mathbf{X}_i), \end{aligned} \quad (2)$$

其中, $\mathbf{D}_i = (b_{i,1}, b_{i,2}, \dots, b_{i,Q})^T$ 对应于观测, $\mathbf{X}_i = (b_{i,Q+1}, b_{i,Q+2}, \dots, b_{i,N})^T$ 对应于模式大气状态。显然有

$$\mathbf{d} = \sum_{i=1}^r \alpha_i \mathbf{D}_i \quad (3)$$

和

$$\mathbf{A} = \sum_{i=1}^r \alpha_i \mathbf{X}_i. \quad (4)$$

将同化时间窗内的实际观测记为向量 \mathbf{d}^{real} , 总可以做到让 \mathbf{d}^{real} 的维数 Q 大于截断阶数 r , 则求下面代数方程组

$$\mathbf{d}^{\text{real}} = \sum_{i=1}^r \alpha_i \mathbf{D}_i \quad (5)$$

的最小二乘解得到系数 α_i , 然后由 (4) 式可以得到大气状态 \mathbf{A} 。这就实现了将观测信息从资料空间向模式空间的延伸。在 ERDA 中同化问题变成一个简单的线性反演问题,不需要用到观测算子的切线性算子,同化各种遥感资料将会很简单。

由上面的叙述可以看到, ERDA 的基本做法是通过积分大气模式得到足够多的大气样本,根据这些大气样本计算出对应于实际观测的“模拟观测”,将它和大气样本拼接在一起得到扩展的大气状态样本,利用 SVD 技术从中提取正交基向量。只要样本足够大,就能保证大气吸引子被由这些基向量张成的空间所覆盖,达到在大气吸引子上寻找同化解的目的。不过文献[16]建议从连续的长时间的预报过程中得到这些样本,相当于想在大气吸引子的全体上来寻找同化解,计算量会很大。实际上,我们只要在同化时刻附近大气吸引子的很小的一个片段上寻找同化解就足够了,不一定要将离同化时刻很远的样本也包含进来。因此,我们认为通过扰动初始场来产生大气样本集合的方法应该更有效。作为第一次数值试验,也是为了与 3DVAR 作比较,我们只利用一个时刻(和同化时刻相同)的观测资料,即采用三维同化的方案。具体做法是:首先从一组不同的初始场出发,给出同化时刻附近的预报场来形成样本集合。由于我们的目的是从这一集合

中提取支撑吸引子的正交基向量,因此没有必要像EnKF那样要求所有样本成员是同一时刻的预报场,这样一次预报可以产生多个样本,从而节省了时间.具体设同化时刻(也是有观测的时刻) $t=\tau$,以 τ 为中心选定一个取样时间窗($\tau-T/2, \tau+T/2$)(需要说明的是,这里的取样时间窗不是上面提到的同化时间窗,在三维同化方案中同化时间窗长度是零).设 $t=t_0$ 时刻($t_0<\tau-T/2$)的初始场(之前的分析场)是 \mathbf{A}_0 ,在 \mathbf{A}_0 上选加适当的扰动,从它出发积分大气模式,在取样时间窗内按照一定的时间间隔 Δt (不妨设为常数)选取模式大气样本,可得到 s 个样本,这里 $s=T/\Delta t+1$.如果有 n 个扰动初始场,共得到 $m=s\times n$ 个样本 $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_m$.与大多数资料同化方法一样,我们也是采用“增量”技术,将未加扰动得到的同化时刻的预报作为背景场 \mathbf{A}_B .将扰动样本扣除背景值得到增量 $\delta\mathbf{A}_i=\mathbf{A}_i-\mathbf{A}_B$.相应的模拟观测的增量是 $\delta d_i^{\text{simu}}=d_i^{\text{simu}}-H(\mathbf{A}_B)$, H 是向前观测算子.由 $\delta\mathbf{A}$ 和 δd^{simu} 构成的扩展大气样本中提取的基向量仍然记为 $\mathbf{b}_i=(\mathbf{D}_i, \mathbf{X}_i)$,这时与(4)和(5)式对应的分别是

$$\delta\mathbf{A}_a(\tau) = \sum_{k=1}^r \alpha_k \mathbf{X}_k \quad (6)$$

和

$$\delta d^{\text{real}} = \sum_{k=1}^r \alpha_k \mathbf{D}_k. \quad (7)$$

根据实际的观测增量,由(7)式解得系数 α_k ,则由(6)式就可以得到大气变量的同化增量 $\delta\mathbf{A}_a$.

3 试验设计

3.1 模式和资料

本文利用模式资料对ERDA进行数值试验,

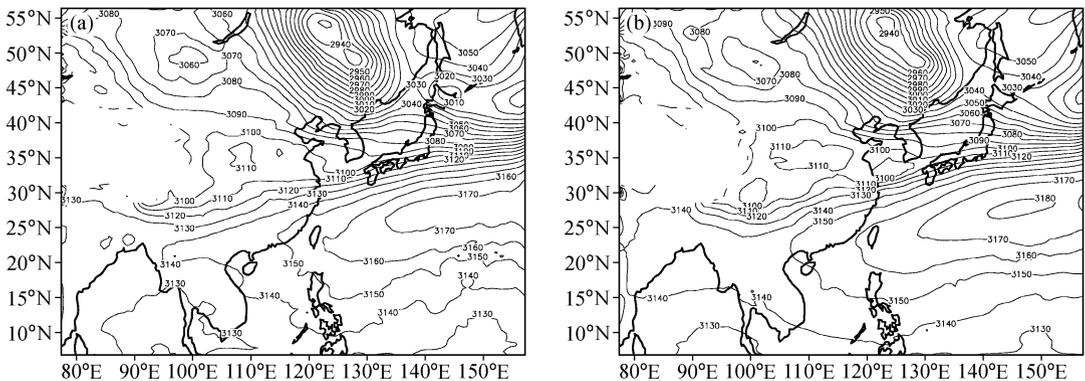


图1 700 hPa 高度场(单位:m);(a)真实场;(b)背景场

Fig. 1 Geopotential height (m) at 700 hPa; (a) True field; (b) background field

用于产生“观测资料”的模式是中尺度模式MM5^[17],格点数是 $99\times 99\times 31$,水平格距54 km,中心点坐标($31.8^\circ\text{N}, 117.3^\circ\text{E}$)。利用美国NCEP(National Center for Environmental Prediction)1°分辨率的再分析资料为模式提供初始场和边界条件,采用KF对流参数化方案.同化时刻是2003年7月5日06:00(国际协调时,下同)。将模式从7月4日06:00开始积分24小时,得到同化时刻的“真实场”。图1a是700 hPa等压面上的高度场,可以看到,在分析区域北部有一个强的低涡,一条西西南-东东北走向的急流带横贯分析区域中部,在这一地区有明显的风切变。

现在的卫星遥感观测可以提供较高水平分辨率的大气温度廓线资料,针对这种情况,在试验中假设仅有气温一个要素的观测,并假定观测在每个模式高度上存在并位于网格点上,但不是每个格点都有观测(具体见下面的试验设计)。“观测值”是在真实场上叠加随机误差得到的,具体算式是:

$$T_{i,k}^o = T_{i,k} + \frac{\vartheta}{\sigma(T_k)} \gamma_{i,k}, \quad (8)$$

这里, $T_{i,k}^o$ 表示在第 k 垂直层的第 i 个温度观测值, $T_{i,k}$ 是相应的真值, $\sigma(T_k)$ 是它的均方差, $\gamma_{i,k}$ 是 $[-1.0, 1.0]$ 之间的随机数, ϑ 是控制误差大小的一个参数。

用于资料同化的背景场也是通过积分MM5产生,与真实场不同的是,所采用的初始场和边界条件是由2.5°分辨率的再分析资料产生,采用的对流参数化方案是Grell方案,意味着预报模式也有误差.也是从7月4日06:00开始作24 h预报得到.将700 hPa等压面背景高度场(图1b)与“真实”

高度场相比可以看到,在背景场中低涡,西风急流和副热带高压都要更强一些。

3.2 3DVAR 系统

用于比较的 3DVAR 系统是 NCAR 发展的 MM5/3DVAR 系统^[18]。该系统的控制变量包括流函数、速度势函数、扰动气压和比湿。我们的试验中质量和风场增量之间采用梯度风平衡关系,背景误差协方差按照 NMC 方法得到^[19],在水平均匀和各向同性的假设下其水平分量近似表达为递归滤波运算^[20],垂直分量用经验正交函数展开处理,有关该系统的详细说明参见文献[18]。在 MM5/3DVAR 系统中采用非跳点的“Arakawa-A”格点系统,然而由 MM5 生成的背景场是采用跳点的 Arakawa-B 格点系统,同化时需要进行插值,但是采用 ERDA 方法,可以让分析场和背景场采用同样的网格系统,不需要进行插值。

3.3 取样方法

如前所述,ERDA 需要在背景场上叠加一组随机扰动再积分大气模式得到样本集合。为此,我们设计了一种产生随机扰动场的方法(见附录),它可以近似地让所产生的随机场在每一空间点上的扰动量服从正态分布,并保证适当的空间相关结构。现在设定 7 月 4 日 18:00 到 7 月 5 日 18:00 之间的 24 小时作为取样时间窗,按照附录所给方法生成 20 个扰动场叠加到初始场后作出预报,在上述时间窗内每隔 3 小时作一次取样共得到 $20 \times 9 = 180$ 个样本。

4 试验结果

针对不同的观测误差和测站空间密度将 ERDA 和 3DVAR 的同化结果进行比较。计算分析增量场的相对均方根误差(ϵ)和分析增量场与真实增量场之间的相关系数(ρ)来评定同化的质量。它们的定义如下:

$$\epsilon(\delta u) = \frac{\sum_{i=1}^N (\delta u_i^T - \delta u_i^A)^2}{\sum_{i=1}^N (\delta u_i^T)^2}, \quad (9)$$

$$\rho(\delta u) = \frac{\sum_{i=1}^N (\delta u_i^T - \langle \delta u^T \rangle)(\delta u_i^A - \langle \delta u^A \rangle)}{\sqrt{\sum_{i=1}^N (\delta u_i^T - \langle \delta u^T \rangle)^2 \sum_{i=1}^N (\delta u_i^A - \langle \delta u^A \rangle)^2}}, \quad (10)$$

其中, N 表示格点数,上标“T”和“A”分别表示真实增量和分析增量, $\langle \rangle$ 是区域平均运算。分析变量包括温度 T , 扰动气压 p' , 水平风 u 、 v (以下合记为 V_h) 和比湿 q 。

从前面的叙述看到,ERDA 的结果会依赖于 SVD 的截断阶数 r 。截断阶数高可以从观测获取较多的信息,但是也会保留更多的观测噪音。截断阶数的选取应该与样本数的大小、资料量的多少以及观测误差大小都有关系,如何确定一个最好的截断阶数仍然是一个有待研究的问题,但是,如果同化结果对截断阶数不敏感,那末截断阶数的选择问题就不会成为 ERDA 运用上的一个困难。因此,在与 3DVAR 比较之前首先做一组试验检验方法对截断阶数的敏感性。这组试验的条件是:每 3 个格距(相距 162 km)有一个温度观测,按照公式(8)给出误差,参数 $\vartheta = 2.0$ (实际得到的整层平均的均方根误差是 0.803 K),经计算背景场的均方根误差是 0.984 K。SVD 的截断阶数依次取 20、30、45。表 1 是不同截断阶数 ERDA 给出的分析场的均方根误差和相关系数的比较。总的来看,同化对截断阶数的取值不算敏感,在 20~45 范围内得到的结果差异不是很大,随着 r 的增加,有观测的变量(温度)分析精度会有所提高,但其他变量分析精度会有所下降。在以下的试验中截断阶数 r 都取为 30。

表 1 不同截断阶数下 ERDA 给出的分析增量的相对均方根误差(ϵ)和相关系数(ρ)

Table 1 Relative rms error (ϵ) and correlation coefficient (ρ) of the analysis increment by ERDA for different truncation order

| r | ϵ | | | | ρ | | | |
|-----|------------|-------|-------|-------|--------|-------|-------|-------|
| | T | p' | V_h | q | T | p' | V_h | q |
| 20 | 0.860 | 0.807 | 0.917 | 1.030 | 0.529 | 0.516 | 0.504 | 0.418 |
| 30 | 0.816 | 0.899 | 0.913 | 1.028 | 0.537 | 0.515 | 0.490 | 0.403 |
| 45 | 0.797 | 0.873 | 0.939 | 1.038 | 0.553 | 0.483 | 0.452 | 0.392 |

表 2 ERDA 和 3DVAR 给出的分析增量的相对均方根误差 (ϵ) 和相关系数 (ρ) 的比较Table 2 Comparison of relative rms error (ϵ) and correlation coefficient (ρ) of the analysis increment obtained by ERDA and 3DVAR

| Expt. | Method | ϵ | | | | ρ | | | |
|-------|--------|------------|-------|-------|-------|--------|-------|-------|-------|
| | | T | p' | V_h | q | T | p' | V_h | q |
| 1 | ERDA | 0.816 | 0.899 | 0.913 | 1.028 | 0.537 | 0.515 | 0.490 | 0.403 |
| | 3DVAR | 0.804 | 1.082 | 1.068 | 1.007 | 0.587 | 0.424 | 0.076 | 0.020 |
| 2 | ERDA | 0.815 | 0.813 | 0.906 | 1.026 | 0.534 | 0.516 | 0.499 | 0.466 |
| | 3DVAR | 0.678 | 1.047 | 1.377 | 1.003 | 0.663 | 0.401 | 0.103 | 0.021 |
| 3 | ERDA | 0.819 | 0.807 | 0.908 | 1.035 | 0.533 | 0.514 | 0.482 | 0.404 |
| | 3DVAR | 0.856 | 1.223 | 1.068 | 1.004 | 0.514 | 0.365 | 0.114 | 0.022 |
| 4 | ERDA | 0.832 | 0.825 | 0.925 | 1.053 | 0.510 | 0.492 | 0.454 | 0.370 |
| | 3DVAR | 0.885 | 1.111 | 1.059 | 1.011 | 0.494 | 0.341 | 0.032 | 0.030 |
| 5 | ERDA | 0.887 | 0.821 | 0.969 | 1.052 | 0.462 | 0.489 | 0.438 | 0.353 |
| | 3DVAR | 0.909 | 1.055 | 1.029 | 1.011 | 0.373 | 0.203 | 0.054 | 0.028 |

共设计了下面 5 组试验来与 3DVAR 进行比较。

试验 1: 每 3 个格距 (相距 162 km) 有一个温度观测, 按照公式 (9) 给出误差, 参数 $\vartheta=2.0$ 。

试验 2: 将观测误差缩小到原来的一半, 其余同试验 1。

试验 3: 将观测误差增加 50%, 其余同试验 1。

试验 4: 将观测的水平间隔增加为 5 个网格距 (270 km), 其余同试验 1。

试验 5: 假设观测只在分析区域中部 1/4 的面积内有效, 水平间隔缩小为 2 个网格距 (108 km), 其余同试验 1。

由表 2 可以看到, 在试验 1 中, 3DVAR 给出的温度分析对背景场有明显改进, 相关系数也较高, 说明对于有观测的变量, 3DVAR 能够给出不错的分析。与之相比, ERDA 给出的温度场分析稍差一些。在 3DVAR 中气压场是根据静力平衡关系由温度场推断出来的, 这是一个精度相当高的平衡关系, 所以得到的扰动气压增量相关系数也比较高。ERDA 没有显式地运用静力平衡关系, 但是它给出的扰动气压场精度比 3DVAR 还要高。对于水平风和比湿, 情况有很大的不同。3DVAR 给出的水平风场分析对背景场几乎没有改进, 增量相关很小, 说明 3DVAR 中运用的流场和质量场的平衡关系并不好, 然而 ERDA 给出的水平风场增量仍然有很高的相关。3DVAR 中温度和比湿被当作两个

独立的量, 不可能由温度观测反演出比湿, 从表 2 中也可看到, 3DVAR 给出的 q 的分析增量几乎为零 (相对误差接近 1, 相关系数接近 0)。然而 ERDA 却能够较好地给出 q 的分析, 相关系数达到 0.403, 不过相对误差略大于 1。图 2~4 是试验 1 给出的 $\sigma=0.75$ 平面上两种同化方法产生的 T 、 V_h 和 q 的增量场和真实增量场的比较。可以看到, 真实的温度增量 (图 2a) 大部分为正值, 结构比较零乱, 比较明显的正增量出现在三个地区, 一个是在东南部海洋上, 另外两个分别与分析区北部的低涡及中部的急流带相对应, 和这两个正增量区相邻一些小的负值区。3DVAR 获得的温度分析增量场 (图 2b) 比较光滑, 但是几个主要的高值和低值中心与真实的增量场很一致。ERDA 给出的温度增量 (图 2c) 也能够很好地表现上述基本特点, 扰动的结构更接近真实场, 不过得到的增量值明显小于实际。真实的水平风增量场 (图 3a) 也表现为一些小尺度的扰动, 主要的扰动分布也是在急流带附近温度增量梯度比较大的地区。3DVAR 给出的 V_h 增量 (图 3b) 比真实场要小得多, 不能反映其基本特征。ERDA 给出的 V_h 增量与真实场却有很多相似之处 (图 3c), 主要的扰动地区和实际相吻合, 风向也大体相符, 不过风速要小一些。前面我们已经指出 3DVAR 不可能由温度观测给出湿度分析, 实际上 3DVAR 给出的湿度分析增量很小 (图略)。从图 4b 看到, ERDA 能够给出与真实场很接近的

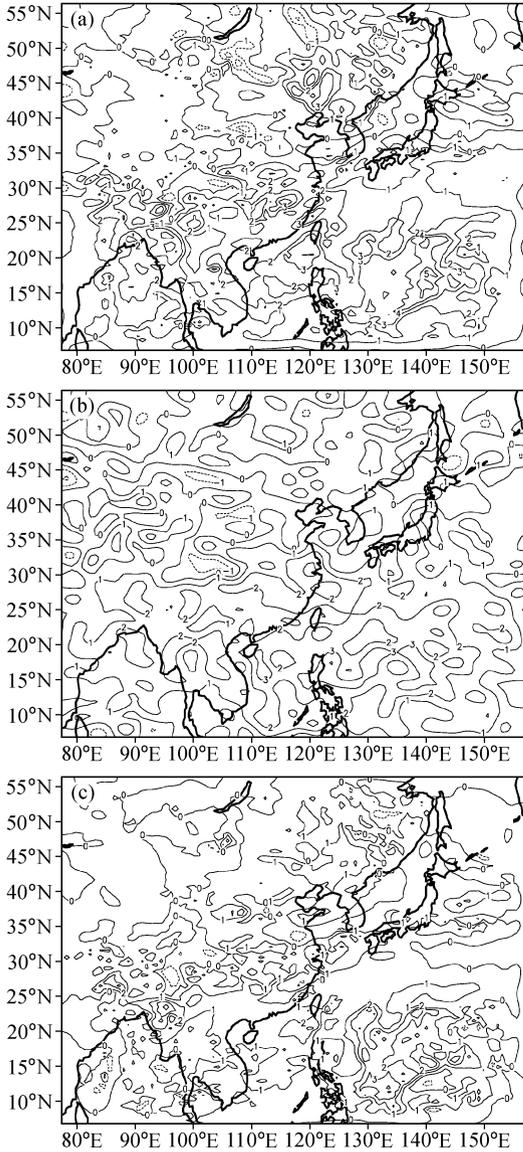


图2 试验1中 $\sigma=0.75$ 平面上的温度增量场(单位:K): (a) 真实场; (b) 3DVAR同化场; (c) ERDA同化场

Fig. 2 Temperature increment (K) at level $\sigma=0.75$ in Expt 1: (a) True field; (b) assimilated field by 3DVAR (three-dimension variational technique); (c) assimilated field by ERDA (ensemble-based reduced-dimensional data assimilation)

湿度分析增量,几个主要的正负增量区都描写得相当不错,数量相差也不大,一个明显的错误是没有反映出(42°N,125°E)附近一个强的负增量团。

比较试验1、2、3可以看出观测误差对同化的影响。误差缩小一半(试验2)或者增加50%(试验3),对ERDA的分析影响轻微,因为随机的观测误差在SVD模上的投影很小,所以观测误差对同化结果影响较小。观测误差对3DVAR的影响则

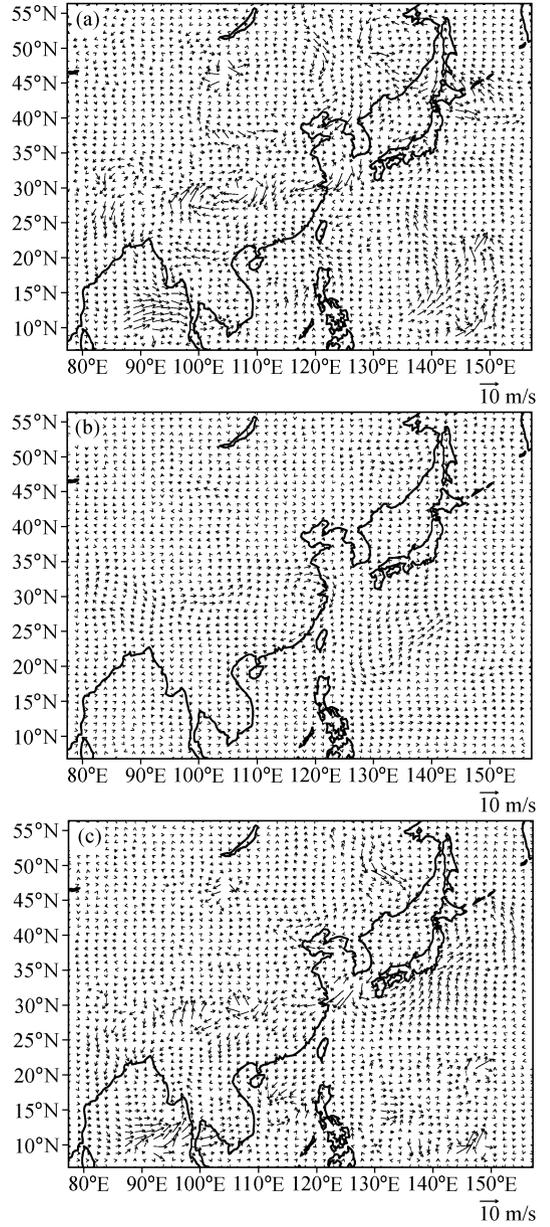


图3 同图2,但为水平风增量

Fig. 3 As in Fig. 2, but for the horizontal wind increment

要大一些,但总的来说它对随机型的观测误差也不算敏感,大部分观测误差在同化过程中被过滤了。当观测误差较小时,3DVAR给出的温度分析会好于ERDA。增加观测点之间的间隔(试验4)同化的精度会有所下降,但不是特别明显。如果只在分析区域中心1/4的面积内存在观测(但是观测密度比较高)(试验5),ERDA仍然可以得到不错的分析结果。图5是试验5给出的 $\sigma=0.75$ 平面上ERDA产生的 T 、 V_h 和 q 的增量场。可以看到,在没有观

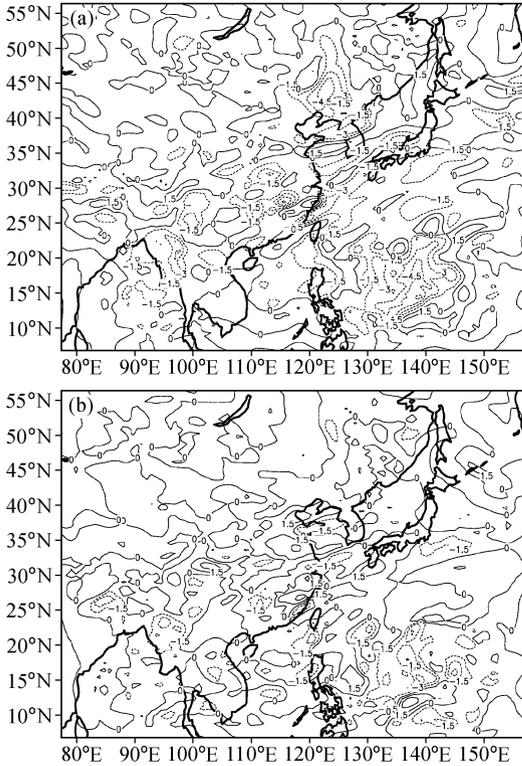


图4 试验1中 $\sigma=0.75$ 平面上的比湿增量场(单位: g/kg):
(a) 真实场; (b) ERDA 同化场
Fig. 4 Specific humidity increment (g/kg) fields at level $\sigma=0.75$ in Expt 1: (a) True field; (b) assimilated field by ERDA

测的地区 ERDA 仍然可以得到比较合理的分析场。3DVAR 的分析增量则局限在有观测的附近地区(图略)。这个结果也是很有意义的, 因为从卫星和雷达遥感观测获得的资料往往只覆盖预报区域的部分地区, 但是在这些地区可以有很高的密度。

5 总结和讨论

本文利用模拟资料检验 ERDA 的能力, 并与 3DVAR 同化结果进行了比较, 这些初步的试验显示 ERDA 确有一些吸引人之处, 与 3DVAR 相比, 主要的好处可以总结如下:

(1) 资料同化的一个目的是希望从较少量的观测来推断所有计算网格点上的模式变量。3DVAR 主要是通过由统计产生背景场误差协方差矩阵来实现。由于目前业务的 3DVAR 系统运用背景场误差协方差来自于气候统计, 并且假设均匀各向同性, 这就大大影响了 3DVAR 的效果。ERDA 是借助预报集合和 SVD 技术达到由观测推断网格点模式变

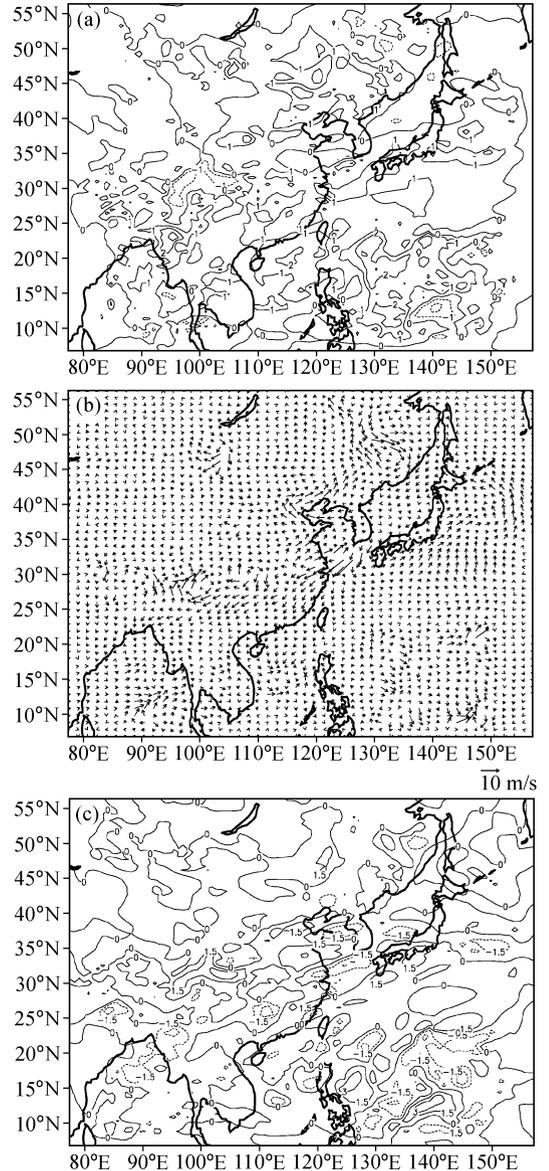


图5 试验5中 $\sigma=0.75$ 平面上 ERDA 同化增量场: (a) 温度(单位: K); (b) 水平风; (c) 比湿(单位: g/kg)
Fig. 5 Increment fields at level $\sigma=0.75$ by ERDA in Expt 5: (a) Temperature (K); (b) horizontal wind; (c) specific humidity (g/kg)

量的目的, 这时候格点变量和观测变量(增量)之间的关系是隐含在 SVD 基函数中的, 这种关系是依流型而变的。

(2) 为了保证同化场在动力上的协调性, 也为了从一种变量的观测推断其他变量, 3DVAR 利用了一些近似的平衡关系。一般来说, 在质量场和水平风之间会利用地转或梯度风平衡关系。我们的试验表明这不是一个好的近似, 从温度观测给出的风

的分析并不好。ERDA 不需要引入这些近似的平衡关系,正如前面讲过的,这种关系是从预报集合中抽取出来的,试验表明,这样得到的关系更可靠一些,ERDA 方法从温度观测推断的风场比 3DVAR 要好得多。还有一些变量之间很难从物理上给出明确的关系,比如温度和比湿之间,3DVAR 就无法从温度观测推测比湿,ERDA 则有这种能力。

(3) 好的同化方法应该有能力滤除观测中的噪音。MM5/3DVAR 利用了背景场约束并且分别运用递归滤波技术(水平方向)和经验正交分解技术(垂直方向)替代背景误差协方差,这可以滤除大部分随机的观测噪音,分析结果对观测误差不敏感。在 ERDA 中分析增量是通过将观测依截断的奇异向量展开得到的,得到的结果是动力协调的。随机的观测噪音大部分不会满足动力协调关系,因此会被滤除,同化结果对观测误差更不敏感。

(4) ERDA 可以很容易做到分析变量和预报变量完全一致,不用特别考虑变量间平衡关系,也不需要切线性算子,因此计算程序非常简单。只是产生一个预报需要花费较多的计算时间,但是按照本文建议的取样技术,所花费时间要比 EnKF 少。

通过我们的试验也发现 ERDA 也存在一些缺陷。因为 ERDA 是在高度截断的谱空间来求解,这固然会减少观测噪音的影响,但是观测中不少有用的信息也被滤除了,这从试验 1~3 的结果也可以看到,当温度的观测误差减小时,3DVAR 给出的温度分析误差会减少得比较快,然而 ERDA 所给出的温度分析误差减少得很少。当观测误差减小时,增加截断阶数可以将分析误差(对有观测的量)减小,然而这可能使其他非观测量的分析误差增加。此外,我们的试验也发现,平均而言 ERDA 有不错的效果,但在一些地区也会出现虚假的扰动,说明 ERDA 在从少量的观测推断所有计算网格点上的模式变量时也有一定的风险,在观测资料过少就有可能出现问题。要让 ERDA 取得好的同化效果,还有一个问题是如何更好地生成样本集合,让表征主要的初始场误差结构的信息能够在该集合中得到反映。上述这些都是需要进一步研究的问题。尽管目前的结果是令人鼓舞的,但是现在的试验还只是初步的,还必须通过更多的利用实际资料的试验对它作进一步的检验,这是我们下一步的

研究目标。

附录:随机扰动场的生成方法

给出一种产生随机扰动场的近似方法,希望在每一个空间点上扰动量服从均值为零,标准差为 σ 的正态分布,并且有适当的空间相关结构。以水平二维场为例说明具体做法。

(1) 设需要产生的样本数为 N ,扰动的标准差为 σ ,将 $[-3\sigma, +3\sigma]$ 作为取值区间,将其划分为 m 个子区间,依照正态分布的概率密度计算每个子区间内的可能取值点数,并按照均匀分布的原则给出每个点的值。这样,只要样本足够大就能够近似地保证在每一个空间点上扰动量服从正态分布。

(2) 将区域中心点作为出发点,可以随机地给出第一个值记为 u_1 ,下一个相邻格点的值在

$$[u_1 - \delta u_{1,2}, u_1 + \delta u_{1,2}]$$

内随机选取,这里 $\delta u_{1,2} = r_{1,2} \sigma / L$, $r_{1,2}$ 是两个点之间的距离, L 是特征长度,本文中 L 取为 2 倍格距。类似的第 k 个格点的取值范围是

$$[u_1 - \delta u_{1,k}, u_1 + \delta u_{1,k}] \cap [u_2 - \delta u_{2,k}, u_2 + \delta u_{2,k}]$$

$$\dots \cap [u_{k-1} - \delta u_{k-1,k}, u_{k-1} + \delta u_{k-1,k}],$$

对所有的点完成取值后得到一个样本。这样使得一个点的扰动值会与相邻点的扰动值有一定的相关性。可以通过改变特征长度来控制相关程度。

(3) 重复步骤 (1) 和 (2),得到所有样本,只是要注意原先已经使用过的值不再重复使用。

致谢 衷心感谢丑纪范教授在本文写作过程中给予作者的鼓励和帮助,也要感谢 Gao Jidong 博士和 Pu Zhaoxia 博士所给予的帮助。

参考文献 (References)

- [1] Courtier P, Andersson E, Heckley W, et al. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Quart. J. Roy. Meteor. Soc.*, 1998, **124** (550): 1783~1807
- [2] Barker D M, Huang W, Guo Y-R, et al. A three-dimensional variational data assimilation system for MM5: Implementation and initial results. *Mon. Wea. Rev.*, 2004, **132**: 897~914
- [3] Lewis J M, Derber J C. The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, 1985, **37A**: 309~322
- [4] Le Dimet F X, Talagrand O. Variational algorithms for analysis and assimilation of meteorological observations: Theoret-

- ical aspects. *Tellus*, 1986, **38A**: 97~110
- [5] Talagrand O. Assimilateion of observations, an introduction. *J. Meteor. Soc. Japan*, 1997, **75** (Special Issue) (No. 1 B): 81~99
- [6] Xu Q. Generalized adjoint for physical processes with parameterized discontinuities. Part I: Basic issues and heuristic examples. *J. Atmos. Sci.*, 1996, **53** (8): 1123~1142
- [7] Xu Q, Qiu C J. Adjoint matching condition for parameterized discontinuities—A derivation using Lagrangian-form cost-function. *Adv. Atmos. Sci.*, 1997, **14** (1): 49~52
- [8] Mu M, Wang J F. A method for adjoint variational data assimilation with physical “on-off” processes. *J. Atmos. Sci.*, 2003, **60** (16): 2010~2018
- [9] Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, 1994, **99** (C5): 10143~10162
- [10] Burgers G, van Leeuwen P J, Evensen G. Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, 1998, **126**: 1719~1724
- [11] Houtekamer P L, Mitchell H L. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, 2001, **129**: 123~137
- [12] Evensen G. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, 2003, **53**: 343~367
- [13] 许小永, 刘黎平, 郑国光. 集合卡尔曼滤波同化多普勒雷达资料的数值试验. *大气科学*, 2006, **30** (4): 712~728
- Xu Xiaoyong, Liu Liping, Zheng Guoguang. Numerical experiment of assimilation of Doppler radar data with an ensemble Kalman filter. *Chinese Journal of Atmospheric Science* (in Chinese), 2006, **30** (4): 712~728
- [14] Li J P, Chou J F. Existence of atmosphere attractor. *Science in China* (Ser. D), 1997, **40** (2): 215~224
- [15] Osborne A R, Pastorello A. Simultaneous occurrence of low-dimensional chaos and colored random noise in nonlinear physical system. *Physics Letters A*, 1999, **181**: 159~171
- [16] Qiu C, Chou J. Four-dimensional data assimilation method based on SVD: Theoretical aspect. *Theor. Appl. Climatol.*, 2006, **83**: 51~57
- [17] Dudhia J. A nonhydrostatic version of the Penn State/NCAR mesoscale model: Validation tests and simulations of an Atlantic cyclone and cold front. *Mon. Wea. Rev.*, 1993, **121**: 1493~1513
- [18] Barker D M, Huang W, Guo Y R, et al. A three-dimensional variational (3DVAR) data assimilation system for use with MM5. NCAR Tech. Note. NCAR/TN-453 + STR, 2003, 68 pp
- [19] Parrish D F, Derber J C. The National Meteorological Center’s spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, 1992, **120**: 1747~1763
- [20] Hayden C M, Purser R J. Recursive filter objective analysis of meteorological fields: Applications to NESDIS operational processing. *J. Appl. Meteor.*, 1995, **34**: 3~15