

王云峰, 费建芳, 袁炳, 等. 2013. 消除系统性观测误差的时空梯度信息同化方法研究 [J]. 大气科学, 37 (1): 54–64, doi:10.3878/j.issn.1006-9895.2012.11234. Wang Yunfeng, Fei Jianfang, Yuan Bing, et al. 2013. Assimilation of temporal and spatial gradient information to eliminate the systematic observation error [J]. Chinese Journal of Atmospheric Sciences (in Chinese), 37 (1): 54–64.

消除系统性观测误差的时空梯度信息 同化方法研究

王云峰 费建芳 袁炳 韩月琪

解放军理工大学气象海洋学院, 南京 211101

摘 要 随着气象观测手段的进步, 各种气象观测资料在数值预报模式中的应用不断发展。然而由于观测资料存在观测误差, 尤其一些非常规资料存在系统性偏差, 且难以对此类误差进行充分订正, 使得观测资料在数值预报模式同化应用过程中的作用没有被充分发挥。文中提出一种消去此类误差的时间及空间梯度信息变分同化方法, 其特点在于不需要知道系统性偏差的具体估计, 而是用一个梯度信息算子对原变量进行变换从而隐性回避此类误差。通过浅水波模式四维变分同化理想试验结果表明, 此同化方法完全消除平整性系统性偏差对同化结果的影响, 本身数值较小的模式变量更能够获得好的同化效果, 大数值变量则可通过估算来确定适用范围。由于最优解不唯一性质的存在, 同化效果更多的吸收观测场的整体时空梯度分布趋势而非观测量值本身, 这对具有较低可信度的观测资料是适用的。

关键词 数值预报 系统性观测偏差 时空梯度信息 同化

文章编号 1006-9895(2013)01-0054-11

中图分类号 P457

文献标识码 A

doi:10.3878/j.issn.1006-9895.2012.11234

Assimilation of Temporal and Spatial Gradient Information to Eliminate the Systematic Observation Error

WANG Yunfeng, FEI Jianfang, YUAN Bing, and HAN Yueqi

Institute of Meteorology and Oceanography PLA University of Science and Technology, Nanjing 211101

Abstract With the advances of meteorological observation instruments, a variety of meteorological observations can be used in numerical prediction models. However, due to the observational error, especially systematic deviations in unconventional observations, the effect of observational data assimilation has not been fully examined. Thus, a variational assimilation method for temporal and spatial gradient information is proposed to eliminate such errors. The principle is that no a priori estimates of the systematic bias are needed, but a gradient information operator is used to transform the original variables so as to implicitly avoid this systematic error. A series of results of four-dimensional variational assimilation ideal experiments based on a shallow water model shows that this assimilation could completely eliminate the impact of smoothness systematic deviation on the assimilation results. The model could provide a good assimilation effect for the variables having a small value, and could estimate the scope of application for the ones having a large value. Due to the uncertainty of the optimal solution, the assimilation effect absorbs more of the overall temporal and spatial

收稿日期 2011-11-27, 2012-08-25 收修定稿

资助项目 国家高技术研究发展计划 (863 计划) 2010AA012304, 国际科技合作项目 2010DFB33880, 国家自然科学基金 41005029、41105065、11271195, 国家公益性行业 (气象) 科研专项 GYHY201106004, 江苏省自然科学基金 BK2010128

作者简介 王云峰, 男, 1972 年出生, 副教授, 主要从事数值模拟及资料同化。E-mail: wangyf@mail.iap.ac.cn

gradient trends of the observation field rather than the observation value itself, which is applicable to observational data with low credibility.

Keywords Numerical weather prediction, Systematic error, Temporal and spatial gradient information, Assimilation

1 引言

随着气象观测手段的进步, 各种气象观测资料在数值预报模式中的应用不断发展。现阶段观测资料主要通过各种同化方法应用到数值预报模式中。前人针对初始场背景误差问题作过很多深入研究, 很大程度解决了观测信息在背景场中的合理性传播及平滑问题, 并通过背景误差协方差矩阵的逆来参与决定观测资料在某个模式变量上的作用程度 (Hollingsworth and Lönnberg, 1986; Parrish and Derber, 1992)。

然而仅仅关注背景误差结构是不够的, 因为雷达、卫星观测资料可信度偏低, 仪器本身及各种定标、定位、数据传输、数据转换等过程各种参数的主观性及不确定性会导致观测误差的出现, 甚至对于常规观测资料, 由于人为及仪器原因同样会出现误差。目前处理观测误差的一般方法是通过质量控制来进行取舍, 比如界定资料的物理意义范围, 剔除离群资料, 考察资料与背景场的一致性、误差结构特性符合程度以及两者的各种交叉检验等等。更一般地, 出于技术实施难度问题及计算量的考虑, 观测误差协方差矩阵大多为对角阵即各种观测资料相关系数为零, 因此某种资料对模式状态的贡献权重实际上就被标为观测误差平方的倒数, 且同一种资料的误差权重被标定为同一值 (Yuan et al., 2010)。这意味着, 当观测误差中包含平整性的系统性误差时, 由于系统性误差是全局性的而非个别资料问题, 故不能被质量控制环节所识别及处理, 因而会对同化效果产生不利影响。在一些非常规观测资料同化工作中, 调整后的初始场与探空资料等与可信度较高的常规观测资料相比出现整体性偏差, 究其原因, 除了观测算子及预报模式自身的计算误差外, 还与资料的系统性观测误差有关。而观测误差、模式误差和观测代表性误差是分析误差的主要来源 (王金成等, 2008; Wang and Li, 2009), 一些研究也表明模式误差演变并不服从马尔科夫链分布且短期内模式预报误差均值随时间呈二次增长 (钟剑等, 2011), 这些误差的相互作用造成的非线性增长会对可预报期限产生不利影响 (丁瑞强

和李建平, 2007)。

正是认识到观测误差的处理在同化工作中的重要性, 一些学者试图寻找不同于常规质量控制的处理方法。Wang et al. (2011) 基于 GRAPES 模式变分系统, 对关于资料映射误差的变分偏差订正 (Variational Bias Correction) 方法进行了深入探讨, 通过迫使模式初始状态所表征的天气状况下的映射偏差与卫星遥感观测事件发生时的天气条件所导致的气团偏差 (Air-mass Bias) 相一致, 从而达到偏差订正效果, 此过程虽具随流型动态性质, 但前提是偏差产生机制已知, 故仍难以消除系统性偏差所带来的影响。Andersson and Järvinen (1999) 曾经对变分质量控制 VarQC (Variational Quality Control) 进行一些研究并在欧洲中期天气预报中心 ECWMF (European Centre for Medium-Range Weather Forecasts) 的集成预报系统 IFS (Integrated Forecasting System) 中得到应用, 其同化系统不会毫无余地的拒绝某些具有较大偏差的观测资料, 而是在往后迭代过程中如果这些资料复而能够被其周围资料所接受, 即能够再与周围资料相匹配, 则将会重新获得其对分析场的影响作用, 所以 VarQC 事实上是 Buddy 检查的一个变相类型, 最终仍难以解决系统性偏差的消除问题。

系统性误差难以准确估计并消除, 所以发展一种回避此类误差的同化方法成为一项亟待课题。文中先用一个梯度信息算子对模式变量进行变量变换, 然后基于浅水方程模式, 利用 4DVAR (Four-Dimensional Variational data assimilation) 方法, 对观测空间中的时间及空间梯度信息进行同化试验, 并进行结果分析。

2 时空梯度信息同化原理

为了不失一般性, 假设存在如下某种观测资料真值矢量:

$$\mathbf{Y} = (y_1, y_2, \dots, y_N)^T, \quad (1)$$

其中, $\mathbf{Y}$ 为观测真值矢量, y_i 为某个资料 ($i=1, 2, \dots, N$), N 为资料序列总数, 上标 “T” 表示转置。假设资料存在整体一致性系统误差 δ , 也存在随机误差向量 $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_N)^T$, 每个资料的误差

$\beta_i (i=1, 2, \dots, N)$ 具有无偏的随机分布特性。记模式状态变量为 $\mathbf{X}$, 记观测算子为 $\mathbf{H}$, 则有残差期望 $\langle \mathbf{e}_0 \rangle$ 如下:

$$\langle \mathbf{e}_0 \rangle = \langle \mathbf{H}\mathbf{X} - \mathbf{Y} \rangle \neq 0, \quad (2)$$

记具有误差的观测资料矢量为 $\hat{\mathbf{Y}}$, 观测矢量中各个观测资料记为 $\hat{y}_i (i=1, 2, \dots, N)$, 则 $\hat{\mathbf{Y}}$ 可表示为:

$$\hat{\mathbf{Y}} = (y_1 + \delta + \beta_1, y_2 + \delta + \beta_2, \dots, y_N + \delta + \beta_N)^T = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_N)^T. \quad (3)$$

为了消除 δ , 重新构建模式变量及观测矢量。把模式状态变量 $\mathbf{X}$ 到观测空间的映射矢量 $\mathbf{H}\mathbf{X}$ 表示为 $\hat{\mathbf{X}} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_N)^T$, $\hat{x}_i (i=1, 2, \dots, N)$ 为模式变量到观测空间的各个映射。然后对其进行梯度信息变换, 即进一步改写为

$$\tilde{\mathbf{X}} = \left(\frac{\hat{x}_2 - \hat{x}_1}{s_1}, \frac{\hat{x}_3 - \hat{x}_2}{s_2}, \dots, \frac{\hat{x}_N - \hat{x}_{N-1}}{s_{N-1}} \right)^T = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_{N-1})^T, \quad (4)$$

其中, s_i 为两个资料位置的时间或空间距离。

相应的把观测矢量 $\hat{\mathbf{Y}}$ 改写为

$$\tilde{\mathbf{Y}} = \left(\frac{\hat{y}_2 - \hat{y}_1}{s_1}, \frac{\hat{y}_3 - \hat{y}_2}{s_2}, \dots, \frac{\hat{y}_N - \hat{y}_{N-1}}{s_{N-1}} \right)^T = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_{N-1})^T + \mathbf{0}(\delta), \quad (5)$$

符号 $\mathbf{0}$ 表示不相关。经过变换, 消除了 δ 的影响。且变换后新变量的残差期望 $\langle \tilde{\mathbf{e}}_0 \rangle = \langle \tilde{\mathbf{X}} - \tilde{\mathbf{Y}} \rangle = \mathbf{0}$ 。

令四维变分同化总体目标泛函 J 如下:

$$J = J_B + J_D, \quad (6)$$

$$J_B = \frac{1}{2} (\mathbf{X}_0 - \mathbf{X}_b)^T \mathbf{B}^{-1} (\mathbf{X}_0 - \mathbf{X}_b), \quad (7)$$

$$J_D = \frac{1}{2} \sum_{i=1}^N \sum_{k=0}^K [\mathbf{DHM}_k(\mathbf{X}_0) - \mathbf{DY}]^T \mathbf{W} [\mathbf{DHM}_k(\mathbf{X}_0) - \mathbf{DY}], \quad (8)$$

其中, J_B 表示初始时刻模式控制变量 $\mathbf{X}_0$ 与背景场 $\mathbf{X}_b$ 的偏差, $\mathbf{B}$ 表示背景误差协方差矩阵, 上标“T”表示转置, “-1”表示求逆。 D 是一个梯度信息算子, 可以是空间上任意方向的梯度转换, 也可以为时间方向上的梯度变换, 对于时间而言是一种时变趋势, 有别于实际意义上的空间梯度。 $\mathbf{W}$ 是权重系数矩阵, 反映了观测资料的质量可信度。 $\mathbf{M}_k$ 表示预报模式积分到第 k 时刻。而 J_D 表示经过梯度信息变换后新控制变量与新观测变量 $\mathbf{DY} = (\tilde{y}_1, \dots, \tilde{y}_{N-1})^T$ 的偏差, 因此, 式 (8) 等价于新变量 $\tilde{\mathbf{X}}$ 的线性无偏估计式:

$$J_D = \frac{1}{2} \sum_{i=1}^{N-1} \sum_{k=0}^K (\tilde{\mathbf{X}} - \tilde{\mathbf{Y}})^T \mathbf{W} (\tilde{\mathbf{X}} - \tilde{\mathbf{Y}}), \quad (9)$$

其中 k 代表某时刻, K 表示总时次。 $\mathbf{W}$ 为权重矩阵, 数据进行时空梯度信息转换后, 原误差方差对权重不再具有直接参考意义, 此时可通过对原始观测误差标准差进行与变量本身相同的变量变换后获得, 这样既可使得方法与传统的变分同化方法相一致, 同时也避免了主观确定权重的不确定性。另外, 式 (8) 或式 (9) 中也不需要式 (4)、(5) 中的 s_i 进行显示计算, 因为计算过程中其对目标泛函的贡献相抵。

3 验证试验

3.1 浅水方程模式

为了测试同化方法的有效性, 进行一系列理想模型同化试验。模式采用 p 坐标系中无地形的浅水方程模式, 模式方程组为:

$$\begin{cases} \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} - fv + \frac{\partial \Phi}{\partial x} = 0 \\ \frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + fu + \frac{\partial \Phi}{\partial y} = 0 \\ \frac{\partial \Phi}{\partial t} + u \frac{\partial \Phi}{\partial x} + v \frac{\partial \Phi}{\partial y} + \Phi \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) = 0. \end{cases} \quad (10)$$

试验中采用直角坐标系, 网格距 $d=300$ km, 南北向格点数为 17, 东西向格点数为 21, u 、 v 和 z 都定义在网格点上。地转参数 f 取为常数 $1.03 \times 10^{-4} \text{ s}^{-1}$, 地图投影放大系数取为 1, 位势高度 $\Phi = gz$ 。对模式进行积分时取固定边界条件, 时间积分方案为迭代格式 (松野格式), 空间差分方案为中央差分, 时间积分步长为 600 s。

初始位势高度场定义为 (Han et al., 2009):

$$\Phi_{i,j}^0 = gz_{i,j}^0 = \Phi_0 - fUqd + rf \sin \left(\frac{4\pi(i-1)}{M-1} \right) \sin \left(\frac{\pi j}{L-1} \right), \quad (11)$$

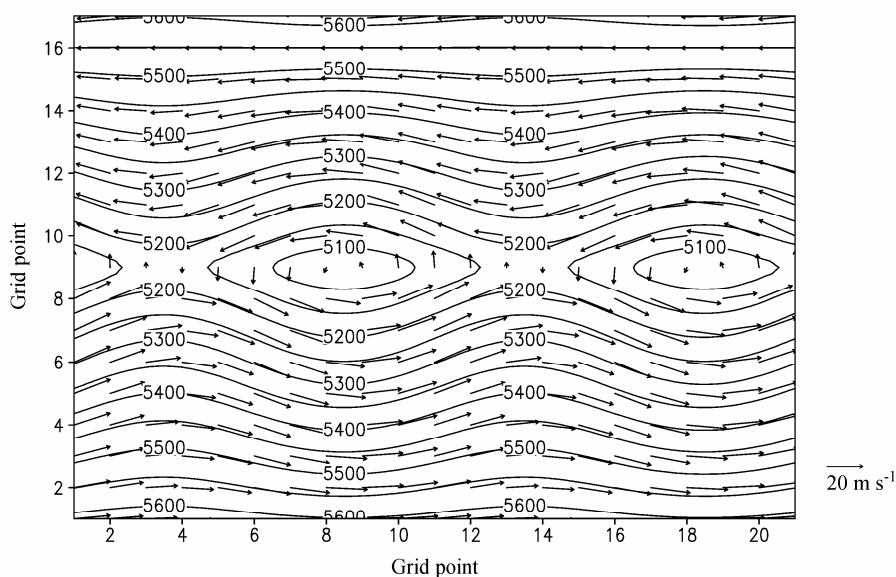
式中 $g=9.8 \text{ ms}^{-2}$, $\Phi_0=5.5 \times 10^4 \text{ m}^2 \text{ s}^{-2}$, $U=20 \text{ m/s}$, $r=(1.5/\pi) \times 10^7 \text{ m}^2/\text{s}^2$ 。 M 为东西向格点数, L 为南北向格点数, i 和 j 分别为东西、南北向的格点序列, 取

$$q = \begin{cases} j-1 & 1 \leq j \leq (L-1)/2+1 \\ L-j & (L-1)/2+1 < j \leq L \end{cases}, \quad (12)$$

初始风场则采用地转初值, 由初始高度场直接导出。初始风场 u_0 、 v_0 和高度场 z_0 的分布如图 1。

3.2 系统性观测误差影响试验

为了突出观测资料系统性误差对同化结果的

图 1 初始风场 (矢量箭头, 标尺单位: m s^{-1}) 和高度场 (等值线, 单位: m) 的分布Fig. 1 Initial wind (vector arrows) and height (contour, units: m) fields

影响, 在背景场和观测资料随机误差不变的情况下, 通过改变观测资料系统性误差来分别进行同化试验。同化方案中, 不考虑背景误差协方差约束作用。假设所有格点皆有观测资料, 且即为模式变量。

将未加扰动的初始场积分 36 步, 得到各时次物理量真值。同化窗口为 7 步时间积分, 用所得最优场进行 36 步预报积分。观测资料是在真值基础上添加均值为零的随机误差 (u, v, z 的扰动均方差分别为 0.4 m/s 、 0.4 m/s 、 4 m) 并全局地分别加上一组递增常数作为系统性误差而得到。 u, v 的系统性误差递增序列分别为 0 、 0.5 、 2 、 5 m/s , z 的系统性误差序列数值上为 u, v 的 10 倍, 之后对这组系统性误差分别逐个进行同化试验。

对于初始背景场的设置, 假设初始场存在振幅扰动, 即 $\delta u_0 = 0.1u_0$, $\delta v_0 = 0.1v_0$, $\delta z_0 = 0.01z_0$ 。此初始场误差大于观测场随机误差, u, v, z 其误差平均值分别约 1.8 m/s 、 0.4 m/s 、 53 m 。

试验设置三个同化方案: (1) 同化观测值本身 (记为 O); (2) 同化东西向、南北向、时间向梯度及观测值本身 (梯度信息作为弱约束, 记为 O+G); (3) 同化东西向、南北向、时间向梯度 (记为 G)。而控制试验记为 C, 其直接以叠加振幅扰动后的背景场作为初始场进行积分。

同化量为观测场时空梯度信息时, 原始数据误差均方差对权重系数的设定不再有意义。本着研究系统性误差的原则, 试验是在假设系统性误差并未

获知的基础上进行的, 于是在直接同化观测量时观测资料被认为是准确的 (均方差已知), 而在同化其时空梯度信息时更认为是准确的。所以不同观测资料之间权重设置遵循对总体目标泛函贡献均等的原则。另一方面, 对于梯度信息的同化, 其在时间和空间中的特性完全不同, 时间上是一种时间倾向, 而空间中则为实际意义上的梯度变化, 时间趋向通常是个小于空间梯度的量, 在贡献对等原则下其权重必然稍大。另一方面, 空间梯度在各个方向并不满足各项同性。因此, 依据对目标泛函贡献均等原则来选择权重不失为一个好的选择。

下降算法采用有限内拟牛顿法 (Nocedal, 1980; Liu and Nocedal, 1989)。迭代停止标准是目标泛函梯度下降两个量级。

图 2 为三个同化方案在不同系统性误差干扰下, 用同化得到的最优场进行预报的结果与真值之间的均方根误差 (RMSE)。基于考察梯度信息同化方法特性的目的, 并没有给出控制试验结果。图中可看出, 由于未考虑背景误差协方差的平衡约束, 对于直接同化观测值本身 (试验 O), 其 RMSE 值与观测资料系统性误差本身相接近, 而由于随机误差、迭代收敛精度的存在, 又不完全等同, 总之其误差取决于观测资料。而当把梯度信息作为弱约束参与同化时 (试验 O+G), 在同化窗内, 误差幅度与试验 O 很接近, 然而在往后的预报中, 误差却比试验 O 小很多, 误差增长变缓。单独同化梯度信息

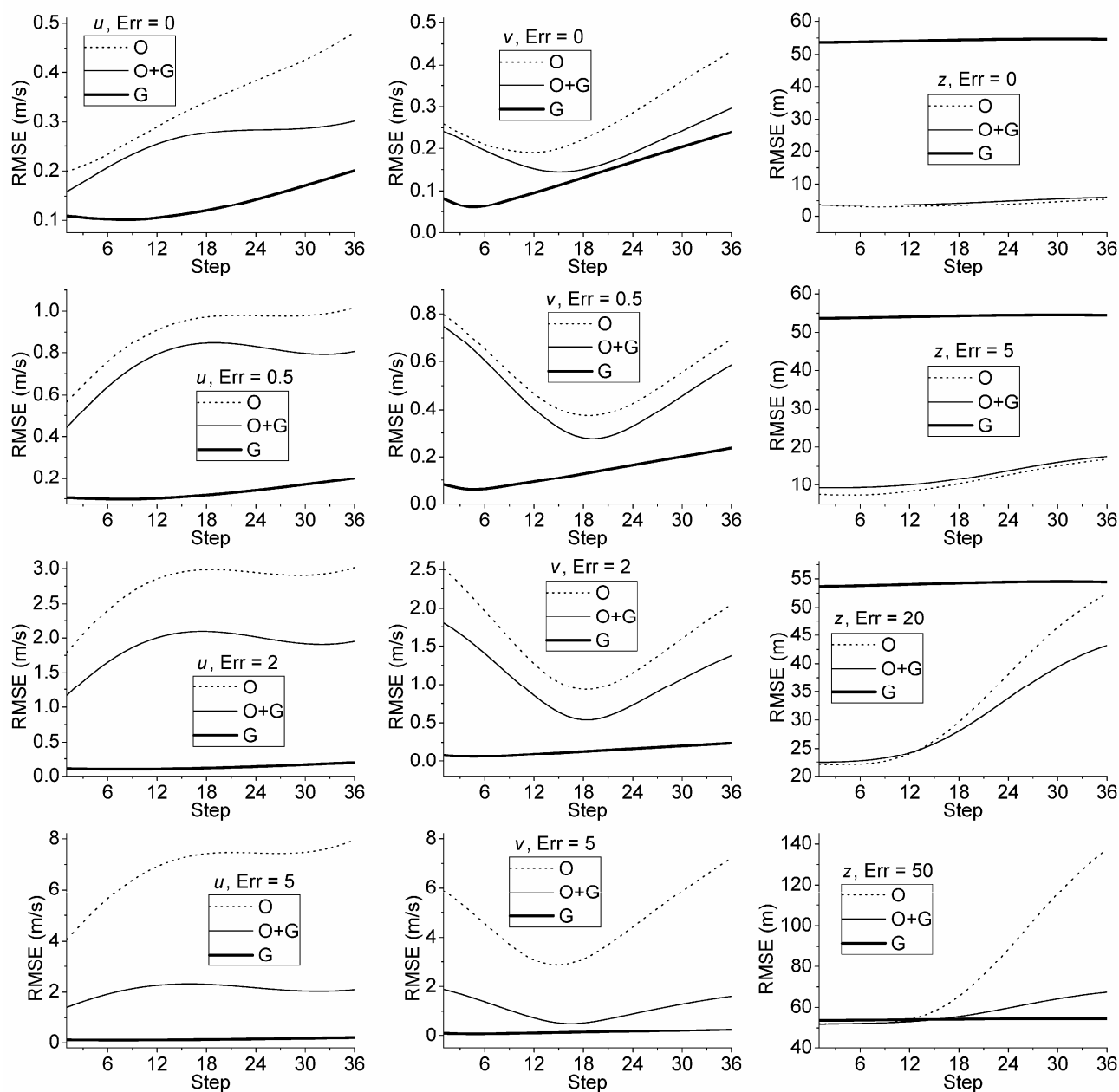


图2 观测值同化试验(O)、梯度信息作为弱约束同化试验(O+G)以及梯度信息同化试验(G)在不同系统性误差(以Err表征)干扰下,用同化得到的最优场进行预报的结果与真值之间的均方根误差(RMSE)。风速 u , v 系统性误差单位为m/s,高度场 z 的系统性误差单位为m

Fig. 2 Under different systematic errors (characterized by Err) setting, for observations assimilation experiment (O), assimilation experiment (O+G) that the gradient information is a weak constraint, and gradient information assimilation experiment (G), the root mean square error (RMSE) between the true values and the results from prediction with assimilation optimal fields. The systematic error units are m/s for u , v and m for z

的试验G,对于不同的系统性误差,其误差结果完全相同(图中由于坐标尺度原因形状有所不同,事实上 u , v , z 中任何一个变量,对于每个系统性误差设置,其RMSE值皆相同),即不受系统性误差的干扰。且对于 u 和 v ,误差一直远低于其他两个试验方案,然而对于 z ,却出现了另外一种较大的额外误差,且数值恒定在53 m左右,即使系统性误差为零时也是如此,且此时误差反而远高于观测资

料均方差为4 m的随机误差扰动,远达不到所需求的优化精度。为了追究此类误差的根源,有必要对同化过程进行更细致的剖析。

试验G既然与系统性误差不相关,则其误差态势势必取决于初始背景场状态或观测资料随机误差分布。图3为控制试验(C)与梯度信息同化试验(G)的误差曲线对比。可看到,控制试验的误差曲线由初始场振幅扰动误差开始随着积分时步

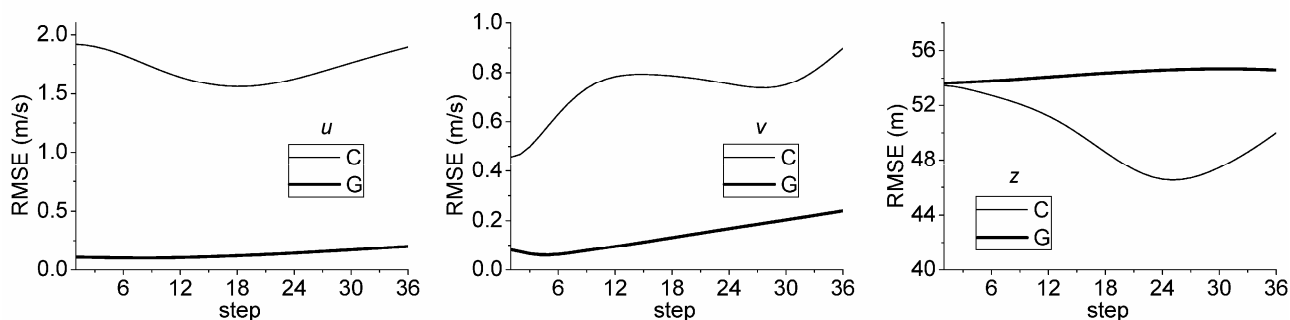


图3 控制试验(C)与梯度信息同化试验(G)的误差曲线对比。试验G中RMSE曲线随时间变化趋势不受系统性误差干扰

Fig. 3 Error curve comparison between control experiment (C) and gradient information assimilation experiment (G). The time trend of RMSE curve in experiment G is not changed by any systematic errors

稍有变化。而试验G中, u 和 v 的误差基本与观测资料随机误差扰动幅度相当, 即尽管不受系统性观测误差干扰, 但仍然受观测资料随机误差影响, 而 z 的初始误差与初始场振幅扰动误差相当, 且一直保持此误差幅度变化甚小。简言之, 梯度同化试验中, u 和 v 的误差较多的受观测资料随机误差决定, 而 z 的误差却更多的偏向于初始场误差状况。

为了剖析此现象的根源, 从式(9)式容易看出, 使 J_D 最小化的理想解集即为满足 $\nabla J_D = 0$ 的模式状态向量。由式(4)、(5)不难得到与 $\nabla J_D = 0$ 等价的一个方程组:

$$\begin{cases} \hat{x}_2 - \hat{x}_1 = \hat{y}_2 - \hat{y}_1 = b_1 \\ \hat{x}_3 - \hat{x}_2 = \hat{y}_3 - \hat{y}_2 = b_2 \\ \dots \\ \hat{x}_N - \hat{x}_{N-1} = \hat{y}_N - \hat{y}_{N-1} = b_{N-1} \end{cases}, \quad (13)$$

式(13)可改写为线性方程组 $A\hat{X} = b$, A 为系数矩阵, $b = (b_1, b_2, \dots, b_{N-1})^T$ 。显然, $\hat{Y} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_N)^T$ 是式(13)的一个特解, 于是不难得出式(13)的一般解为

$$\hat{X}^* = \hat{Y} + cI, \quad (14)$$

c 为任意常数, I 为维数与 $\hat{Y}$ 相同且所有元素皆为1的常向量。于是会有无穷多组解近似满足 J_D 极小条件, 当把观测资料梯度信息作为弱约束与观测值本身一起进行同化时, 观测值本身是式(13)的一组解, 于是其梯度信息失去作用, 最优解仍为观测值, 然而由于模式动力约束及计算精度的存在使得最优解与观测值又存在微小差异, 这与图2中的情形是一致的。当单独同化梯度信息时, 下降算法会从初始模式状态变量 X (背景场) 开始为式(14)优先搜寻离 X 最近的近似解集。因此即使最优场梯度场已经接近观测场梯度分布, 而其本真值仍有可能较为靠近背景场一侧。这有可能成为试验G中 z 存

在较大误差的原因之一。

尽管分析至此, 仍难以确定造成梯度同化额外产生误差的原因是否与初始背景场配置有关, 因此有必要进一步进行试验分析。

3.3 初始场配置影响试验

为了考察不同初始场配置对同化效果的影响, 在观测确定的情况下, 通过改变背景场配置来进行一组试验。试验中观测资料存在一个均值为零的随机扰动高斯误差 (u, v, z 的扰动均方差为 $\sqrt{3}$ m/s、 $\sqrt{2}$ m/s、 $\sqrt{40}$ m), 并叠加一个与随机误差方差相同的系统性误差。因为把梯度信息作为弱约束的同化方法与单独同化观测资料值本身等价, 故本次试验中只设置两个同化试验: 同化观测值 (仍记为试验O) 和同化观测场东西、南北、时间向梯度信息 (仍记为试验G)。而控制试验仍记为试验C。

背景场的5种不同设置方案(SCHEME)如下:

- (1) 与真值相同 (记为 SCH = 1);
- (2) 叠加均值为零的随机误差, u, v, z 的扰动方差为 9 m/s、4 m/s、400 m (记为 SCH = 2);
- (3) 在 (2) 基础上叠加常数误差, u, v, z 分别对应误差 5 m/s、4 m/s、20 m (记为 SCH = 3);
- (4) 在 (2) 基础上叠加振幅误差, u, v, z 分别对应系数 -0.1、-0.2、-0.005 (记为 SCH = 4);
- (5) 构造一个位相误差, 使浅水波模式初始场分布东西方向上偏移半个波周期, 此时原为波峰的位置变为波谷, 反之亦然, 图4为位相调整后的初始场分布 (记为 SCH = 5)。

图5为在各种背景场配置方案下, 直接同化观测值的试验O和同化观测场梯度的试验G中用同化得到的最优场进行预报的结果与真值之间的均方根误差(RMSE)。可以看到, 控制试验C中, 当背景场为真值场时 (SCH = 1), u, v, z 的误差为零,

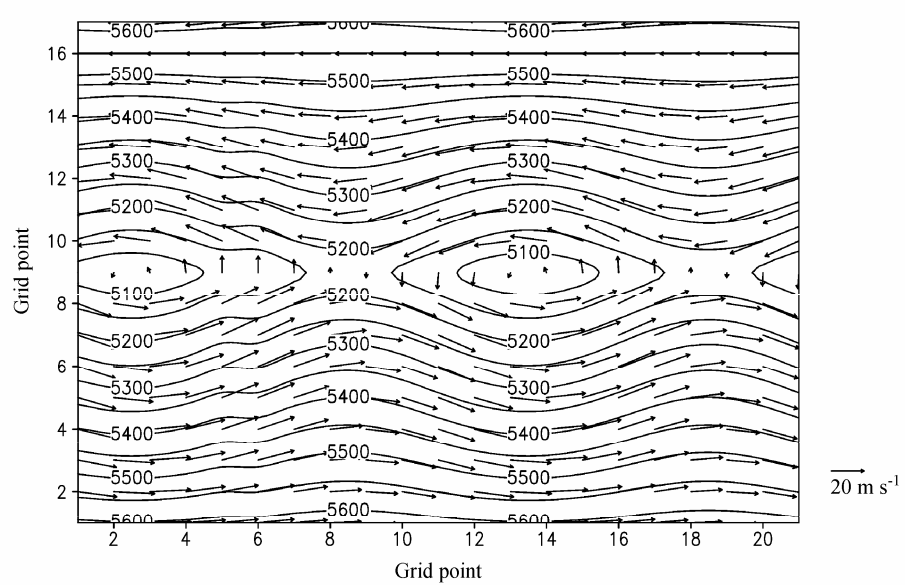


图 4 位相调整后的初始风场 (矢量箭头, 标尺单位: m s^{-1}) 及高度场 (等值线, 单位: m) 分布, 与原初始场相比在东西方向上移动了半个波周期
Fig. 4 Distribution of initial wind (vector arrows) and height (contour, units: m) fields with a phase adjustment. The initial fields move a half-wavelength cycle in the east-west direction compared with the original initial fields

其余背景场设置方案中则在初始误差基础上随时间变化。且以增长态势居多。试验 O 中, 与图 2 中结果一致, RMSE 在同化窗内由观测误差幅度决定, 且由于观测误差被与模式不相协调地累加到背景场上, 出了同化窗后, 积分结果误差大多急剧增长。试验 G 中, 由于去除了观测资料的系统性误差部分, 对于 u, v , 其同化窗内误差与观测资料随机误差大致相当, 换言之, 几种初始场配置中, u, v 的最优解与去除系统性误差之后的观测值相接近, 即式 (13) 中的最优解是确定的, 且出了同化窗之后, 其误差增长都较为缓慢, 效果较佳。而试验 G 中, z 在整个积分过程中误差增长也较为缓慢, 当背景场等于真值时 ($\text{SCH} = 1$), 由于背景场与观测场梯度分布非常接近, 在原背景场上稍作调整即可满足观测场梯度分布, 故其误差与观测场随机误差相当 (去除了系统性误差), 当添加随机扰动误差 ($\text{SCH} = 2$) 和常数误差 ($\text{SCH} = 3$) 时, z 的初始误差仍较大, 与试验 O 相当, 事实上, 根据上文试验结果, 此时已经去除系统性观测误差的影响, 出现的是一个额外误差, 而当初始背景场中变量梯度分布被设置成与原分布状态完全反向时 (方案 $\text{SCH} = 5$), 梯度信息配置相比观测场已完全改变, 然而 z 的初始 RMSE 仍较大, 达不到去除系统性误差的效果。试验 G 同化窗内误差最明显的情形无疑是背景场添加振幅误差的情形 ($\text{SCH} = 4$), 这与图 2 中的现象一致。

然而, 试验 G 中 u, v 同样添加了幅度不小的振幅误差, 却仍能获得较好的同化效果。表 1 为各种背景场误差设置下控制试验 (C)、观测值同化试验 (O) 及梯度信息同化试验 (G) 中, 同化窗口内模式场与观测场间的 RMSE, 此误差很小时, 表明不合理的接近存在较大误差的观测场状态, 而此值大于观测场误差水平时也表明变量与真值之间的误差不合理的超过了观测误差。可以看到, 控制试验中的状况与背景场误差设置相对应, 而试验 O 中, 则过于充分接近观测资料。方案 G 中, u, v 误

表 1 不同初始场误差配置下控制试验 C、观测值同化试验 O 及梯度信息同化试验 G 的模拟值与观测值的均方根误差 RMSE (对同化窗内 7 个时步结果总体计算得到)

Table 1 The RMSE (calculated from results for overall seven time steps within the assimilation window) between the simulation and the observation in control experiments C, observations assimilation experiments (O), and gradient information assimilation experiments (G) under different initial error setups

	u RMSE (m/s)			v RMSE (m/s)			z RMSE (m)		
	C	O	G	C	O	G	C	O	G
SCH=1	2.75	0.86	2.67	2.25	0.70	2.27	10.0	2.64	9.81
SCH=2	1.58	0.82	2.51	1.79	0.69	2.16	17.2	3.29	3.72
SCH=3	4.89	0.88	2.47	2.02	0.66	2.17	27.9	3.24	19.7
SCH=4	2.40	0.82	2.50	1.98	0.67	2.15	31.6	3.92	26.5
SCH=5	3.97	0.79	2.55	9.03	0.75	1.82	59.3	3.63	13.0

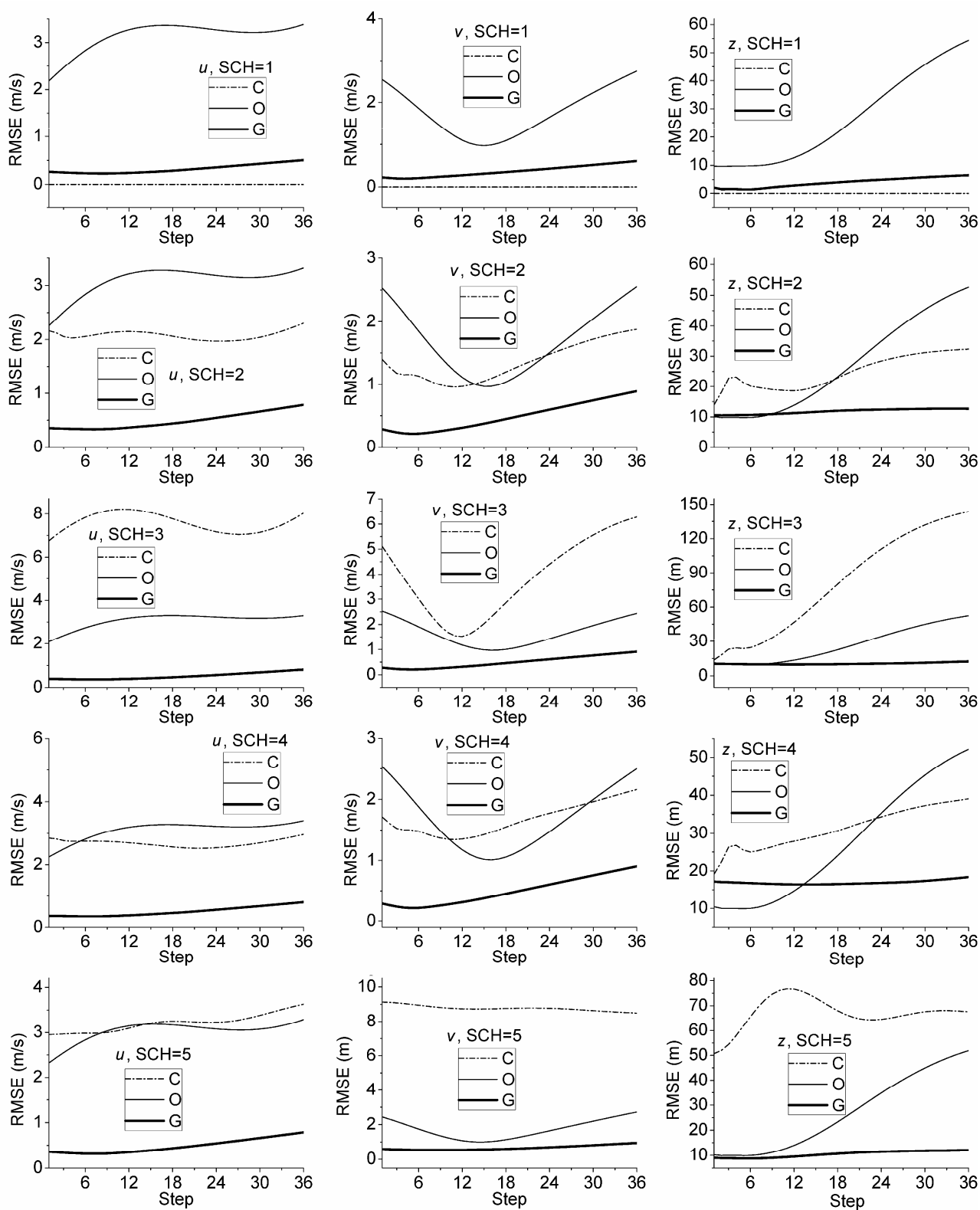


图5 各种背景场配置方案 (SCH) 下, 直接同化观测值试验 O 和同化观测场梯度试验 G 中用同化得到的最优场进行预报的结果与真值的均方根误差 (RMSE)。C 表示控制试验

Fig. 5 Under a variety of background field setup schemes (SCH), the root mean square error (RMSE) between the true values and the results from prediction with optimal fields in observations assimilation experiment O and observation gradient information assimilation experiment G

差稳定在一个与观测误差相吻合的较佳数值上, 与其同化效果极佳相对应, 而 z 与观测的偏离虽无规律可寻, 但却随着背景场设置不同而改变。为此, 需要进一步探究同化试验中变量 z 出现此类误差的原因。

3.4 同化结果的变量基值依赖分析

上文两组试验皆表明, z 的效果也是不受系统性观测误差所影响的, 但梯度信息同化时 u, v 皆能取得非常好的效果时, z 则一直不能得到较好调整, 甚至会比其它试验还差。根据上文对式 (14) 的分析, 虽优化场梯度分布与观测场梯度分布相符合, 但数值却可以存在差异, 比如两个数 a, b , 另外两个数 $c = a + k, d = b + k$, 无论 k 如何变化, $b - a \equiv d - c$, 这也符合两组试验中背景场添加振幅扰动时 z 误差巨大的情形。

然而这并非唯一原因, 因为背景场进行各种设置尤其使其梯度分布信息与观测场反向时, 仍不能取得好的同化效果, 且 u, v 皆同样叠加振幅误差并共同面临最优解不唯一性问题时, 却能取得较好效果。从另一个角度考虑, z 与 u, v 的最大差别就是数

值上的差异, u, v 数值绝对值分别维持在 27、13 以内, 而 z 的数值高于 5000。因此我们以上一小节 SCH=4 的试验 G 中初始场 u, v, z 作为背景场, 并将其初始时刻观测作为资料约束, 进行一个无背景场约束空间梯度信息三维变分同化试验 (假设同化变量无量纲), 不妨假设 u, v, z 的观测均方差分别为 2、2 及 20。初始场及观测资料分别同时乘以一系列的规范化因子进行规范化, 规范化因子皆从一个很大的负数递增到一个很大的正数。试验中, 三种资料梯度信息的权重设置皆遵循对目标泛函的贡献均等原则, 即各类资料可信度相当。

图 6 所示为各变量同化后与真值的均方差 (RMSE) 随着基本量值被规范化到不同量级时的变化曲线。可以看出当规范化到数值很小时, 每种变量同化值与真值之间的 RMSE 皆非常小, 甚至远小于观测误差方差, 随着规范化因子绝对值的增大, 误差开始线性增长, 其增加机制纯粹是由于量值本身被规范化到很大绝对值后所产生, 此误差产生的根本原因可能是最优下降算法所导致, 此内容本文不再深入研究。

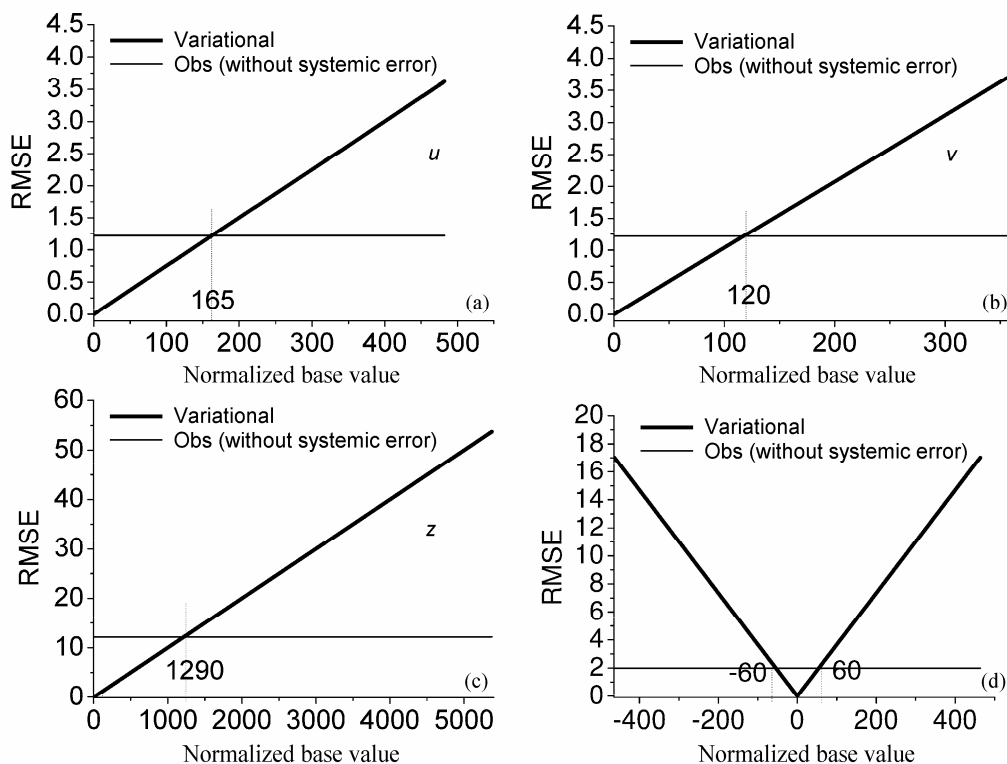


图 6 (a, b, c) 无量纲三维变分梯度同化试验最优场 (u, v, z) 与真值的均方差 (RMSE) 随着基本量值被规范化到不同量级时的变化趋势; (d) 分别产生两组随机数作为背景场和观测, 并规范化后进行梯度信息最小化计算时两者均方根误差随规范化基值的变化曲线。

Fig. 6 (a, b, c) The variations of RMSEs by the scale of basic amount, the RMSEs are calculated between the optimal fields (u, v , and z) and the true values in dimensionless gradient information three-dimensional variational assimilation experiments; (d) the RMSE change curve by the base value scale of the two random numbers which are generated and normalized as a background field and observations for the gradient information minimization calculation.

图 6 中可看出, 在给定的观测随机误差均方差条件下, 对于 u , v , 梯度同化的有效范围为它们的基本量值分别处于 165 及 120 以下, 这是可以满足的。然而 z 却在基值超过约 1290 之后, 其滋生的额外误差就超过了随机观测均方差, 当达到 5000 多量值时, 50 多的误差与图 2 中结果相对应, 但却无法与图 5 中结果相对应。因此有理由相信此额外误差的滋生还与初始背景场与观测场的整体梯度趋势匹配程度相关。图 5 中各试验背景场误差设置皆比图 2 所示试验中背景误差小, 因而相同量值下其同化效果稍好。

图 6d 是先产生一组均值为 0.5 方差为 1 的随机数作为初始场, 另外同样产生一组随机数作为观测场, 进行量值规范化调整后, 同化产生的额外误差与基值的关系。图中, 基值整体从负数切换到正数, 且关于零点对称, 而两侧误差曲线也对称, 额外误差与整体正负性无关 (存在正负交替分布的 u 和 v 额外误差同理)。此时观测场与背景场的梯度分布趋势完全不相吻合, 进行梯度信息同化所滋生出来的额外误差, 需要极小的量值 (−60 到 60) 才能在 2 个单位的观测场随机误差条件下获得满足精度的同化效果。事实上, 大气状态具有连续及平滑性质, 背景场与观测场的整体分布趋势通常也大致吻合, 故主要考虑在给定观测随机误差方差条件下, 能够取得好的梯度同化效果的大致量值, 若不满足这一要求, 可将梯度信息作为弱约束进行同化, 以取得使模式预报误差增长缓慢的效果, 也可以对某个观测资料可信度较高的时次直接同化观测值, 而其他时次则同化梯度信息。另外, 在实际模式动力学及热力学机制仍存在误差时, 一味追求接近观测资料, 造成模式误差缺失, 模拟结果不一定最佳。

4 结论

为了消除观测资料系统性误差, 提出时空梯度信息同化方法, 采用一个梯度变换算子把观测资料转换到时间及空间梯度信息空间中进行同化试验。基于浅水波模式的一系列验证试验表明: (1) 此方法能消除整体一致性的系统性观测误差, 且同化窗口之外的预报误差增长缓慢; (2) 此同化方法存在最优解不唯一性问题, 即使优化场梯度分布状态与观测场相一致, 而数值上有可能并不向观测场靠近, 而是存在一个整体差异, 这种差异并不影响预报误差增长缓慢的特性; (3) 此方法对于原本数

值较小的变量能体现出较好的同化优势, 对于基值较大的变量, 会出现一个不确定性额外误差, 此误差一方面与初始模式状态与观测场之间整体梯度趋势的匹配程度有关, 另一方面与变量基值大小存在密切联系, 这种联系可以近似估测, 比如以观测资料随机误差为标准, 可以估测出特定基值的变量是否满足条件: 梯度信息同化时既去除了资料系统性误差, 又使得所产生的额外误差小于观测资料均方差。

事实上, 在实际模式中, 模式呈复杂的高度非线性, 且非常规资料观测算子也是高度非线性的, 此时唯一性解存在与否, 则需要进一步研究。另一方面, 在实际模式动力学及热力学机制仍存在误差时, 一味追求接近观测资料, 造成模式误差缺失, 模拟结果不一定最佳, 一些可信度较低的遥感观测资料的时空梯度信息比其值本身更具有价值。

参考文献 (References)

- Andersson E, Jrvinen H. 1999. Variational quality control [J]. *Quart. J. Roy. Meteor. Soc.*, 125: 697–722.
- 丁瑞强, 李建平. 2007. 误差非线性的增长理论及可预报性研究 [J]. *大气科学*, 31 (4): 571–576. Ding Ruiqiang, Li Jianping. 2007. Nonlinear error dynamics and predictability study [J]. *Chinese Journal of Atmospheric Sciences (in Chinese)*, 31 (4): 571–576.
- Han Yueqi, Zhang Yaocun, Wang Yunfeng, et al. 2009. A new sequential data assimilation method [J]. *Science in China (Series E: Technological Sciences)*, 52 (4): 1027–1038, DOI: 10.1007/s11431-008-0189-3.
- Liu D C, Nocedal J. 1989. On the limited memory method for large scale optimization [J]. *Mathematical Programming B*, 45 (3): 503–528.
- Lönnberg P, Hollingsworth A. 1986. The statistical structure of short-range forecast error as determined from radiosonde data. Part I: The wind field [J]. *Tellus*, 38: 111–136.
- Nocedal J. 1980. Updating quasi-Newton matrices with limited storage [J]. *Mathematics of Computation*, 35: 773–782.
- Parrish D, Derber J. 1992. The national meteorological center's spectral statistical interpolation analysis system [J]. *Mon. Wea. Rev.*, 120: 1747–1763.
- 王金成, 李建平, 丑纪范. 2008. 两种四维 SVD 同化方法的比较及误差分析 [J]. *大气科学*, 32 (2): 277–288. Wang Jincheng, Li Jianping, Chou Jifan. 2008. Comparison and error analysis of Two 4-Dimensional singular value decomposition data assimilation schemes [J]. *Chinese Journal of Atmospheric Sciences (in Chinese)*, 32 (2): 277–288.
- Wang J C, Li J P. 2009. A four-dimensional scheme based on singular value decomposition (4DSVD) for chaotic-attractor-theory-oriented data assimilation [J]. *J. Geophys. Res.*, 114, D02114, doi:10.1029/2008JD010916.
- Wang Xiang, Li Gang, Zhang Hua, et al. 2011. The GRAPES variational bias correction scheme and associated preliminary experiments [J]. *Acta*

- Meteorological Sinica, 25(1): 51–62, doi:10.1007/s13351-011-0004-0.
- Yuan Bing, Fei Jianfang, Wang Yunfeng, et al. 2010. Experimental study on a dynamic asymmetrical typhoon initialization scheme based on 4D-Var [J]. Journal of Tropical Meteorology, 16 (4): 363–371.
- 钟剑, 黄思训, 费建芳, 等. 2011. 模式误差动力特征: 模式参数误差和物理过程描绘缺失误差 [J]. 大气科学, 35 (6): 1169–1176. Zhong Jian, Huang Sixun, Fei Jianfang et al. 2011. Dynamics of model errors: Accounting for parameters error and physical processes lacking error [J]. Chinese Journal of Atmospheric Sciences (in Chinese), 35 (6): 1169–1176.