

# 集合预报误差在 GRAPES 全球四维变分同化中的应用研究 I：局地化方案设计与动力平衡特征分析\*

龚建东 王瑞春

GONG Jiandong WANG Ruichun

1. 国家气象中心, 北京, 100081

2. 中国气象局数值预报中心, 北京, 100081

1. *National Meteorological Centre, Beijing 100081, China*

2. *Numerical Weather Prediction Center of CMA, Beijing 100081, China*

2020-04-24 收稿, 2020-10-13 改回.

龚建东, 王瑞春. 2021. 集合预报误差在 GRAPES 全球四维变分同化中的应用研究 I: 局地化方案设计与动力平衡特征分析. 气象学报, 79(1): 31-47

Gong Jiandong, Wang Ruichun. 2021. The study of introducing ensemble forecast errors into the GRAPES global 4DVar Part I: Localization scheme and dynamic balance analysis. *Acta Meteorologica Sinica*, 79(1):31-47

**Abstract** Introducing flow-dependent ensemble forecast errors into the assimilation system to make the assimilation closely related to weather situations is an important way to improve the quality of the initial analysis field. In this paper, we investigate how to effectively apply ensemble forecast errors into the global four-dimensional variational data assimilation (4DVar) of GRAPES (Global Regional Assimilation and Prediction System), including how to reduce the computational cost when adding extended alpha control variables and how to maintain the dynamic balance between variables during localization. This paper uses the Gaussian shape spectral filter to realize horizontal localization, uses the EOF decomposition to realize vertical localization, and uses the first 8 leading eigenvectors to limit the increase of extended alpha control variable dimension. With the introduction of 20 to 180 ensemble members, the increase in the total number of control variables can be limited to about 1.1 to 1.8 times in the two-dimensional horizontal localization case, and about 1.7 to 7.1 times in the three-dimensional localization case. For 60 ensemble samples and  $1.0^\circ$  horizontal resolution inner loop, after the introduction of the extended alpha control variables, the wall clock time of 4DVar run time increases by about 30%. Furthermore, by performing horizontal localization on unbalanced analysis variables and then adding the geostrophic balance back to the unbalanced analytical variables, this study allows the analysis to better maintain geostrophic balance and the surface pressure tendency of the initial field is reduced. In addition, although the vertical localization has a large impact on the analysis balance, the analysis can well keep geostrophic balance due to the weak constraint formulation of a digital filter in the cost function. The methods developed in GRAPES Global 4DVar such as adding extended alpha control variables, spectral filtering for horizontal localization and localization on unbalanced analysis variables are suitable for the case where the number of ensemble samples exceeds 100, and the quality of analysis is improved without significantly increasing the computational and storage cost of the 4DVar system.

**Key words** Ensemble forecast errors, Horizontal and vertical localization, Geostrophic balance of analysis, GRAPES global 4DVar, Numerical experiment

**摘要** 通过引入流依赖的集合预报误差, 使得同化分析与天气形势紧密相关, 是改善初值分析质量的重要途径。文中在

\* 资助课题: 国家重点研发计划项目(2018YFC1506700、2018YFC1506705)。

作者简介: 龚建东, 主要从事数值预报资料同化、集合预报研究。E-mail: gongjd@cma.gov.cn

GRAPES(Global Regional Assimilation and Prediction System)全球四维变分资料同化(4DVar)中研究了如何有效应用集合预报误差,包括增加扩展控制变量时如何降低其计算消耗以及如何在局地化过程中保持不同变量之间的动力平衡。利用高斯分布的谱滤波实现水平局地化,利用垂直正交经验函数分解实现垂直局地化,并采用前8个主导特征模态来限制控制变量空间维数增加。引入20至180个集合样本,在水平二维局地化情形下,控制变量总数的增长可以限制在1.1—1.8倍,而在三维局地化情形下,控制变量总数的增长限制在1.7—7.1倍。对60个集合样本和1°水平分辨率内循环,4DVar引入扩展控制变量后墙钟时间增加了约30%。进一步,通过采用在非平衡分析变量上进行水平局地化,然后再将风压地转平衡关系重新叠加到非平衡分析变量上,使得分析更好地保持了风压平衡关系,初始场地面气压倾向变化减小。此外,虽然垂直局地化对分析平衡影响较大,但依靠目标函数中的数字滤波弱约束,分析变量之间仍能较好满足动力平衡关系。结果表明,GRAPES全球4DVar中发展的增加扩展控制变量、谱滤波实现水平局地化、非平衡分析变量进行水平局地化等有效应用集合预报误差的方法,适合集合样本数超过100个的情况,在分析质量改善的同时,4DVar系统的计算和存储消耗没有显著增加。

**关键词** 集合预报误差,水平与垂直局地化方案,分析场地转平衡特征,GRAPES全球4DVar,数值试验

**中图法分类号** P435 P456.7

## 1 引言

背景误差协方差简称背景误差,它的准确估计直接影响资料同化的分析性能。背景误差一般使用气候历史样本进行统计,并考虑到误差协方差结构的复杂性,采用高度模型化的误差自相关、协相关等模型和各向同性均质误差假设来简化协方差结构,使得背景误差可以显式表达并用于资料同化系统中(Lorenc, 1986; Bannister, 2008)。这样处理的不足是在变分资料同化的每个分析循环中,虽然数值预报背景场在不断更新,误差特征也在不断变化,但同化框架中背景误差的特征却固定不变。导致背景误差在不同时刻不断变化的因素较多,但其明显受到所在地天气尺度系统的影响,天气系统的斜压不稳定成为背景误差增长的主导因素,呈现出随不同天气流型变化的“流依赖”特征(Kalnay, 2003, 第5.6节)。如果能在同化框架中引入具备流依赖特征的背景误差,将会帮助同化框架更加有效地确定背景场和观测的相对权重,也能更加合理地将观测信息在空间上传递出去,从而提高分析质量和预报技巧。

随着集合卡尔曼滤波(EnKF, ensemble Kalman filtering)等集合资料同化技术的快速发展与逐步成熟(Evensen, 1994; Anderson, 2001; Houtekamer, et al, 2005),利用集合样本的短期预报来估计集合预报误差协方差(简称集合预报误差)作为背景误差更为合理。从使用集合预报误差角度分析,在现有资料同化方法中,三维变分资料同化(3DVar)使用气候背景误差协方差。4DVar在同化时间窗起始时刻使用气候背景误差协方差,并利用切线性和

伴随模式在同化时间窗内实现背景误差的时间演变。基于4DVar,通过扰动观测资料和海温、叠加随机物理过程等来表示观测资料、海洋下边界条件及模式的不确定性,产生多组分析和预报的集合本来估计集合预报误差,这种方法被称为集合方式的四维变分资料同化方法,简称集合资料同化(EDA, ensemble of data assimilations)方法。在4DVar同化时间窗的起始时刻,同时使用气候背景误差和EnKF或EDA方法生成集合样本估计的集合预报误差,使得在同化时间窗的起始时刻就具有随天气形势演变特征的流依赖动态背景误差,这种方法被称为集合四维变分同化(En4DVar, ensemble four dimensional variational data assimilation),其与4DVar的主要差别在于是否使用集合预报误差。而在同化时间窗内,使用集合预报在时间序列上的多组集合本来直接估计整个同化时间窗内不同时刻的背景误差,这种方法被称为四维集合变分同化(4DEnVar, four dimensional ensemble variational data assimilation),其与4DVar和En4DVar的主要区别是不再引入切线性和伴随模式。熊春晖等(2013)、Fairbairn等(2014)对这些方法的差异做了详细介绍。

目前,国际上领先的业务变分资料同化系统大多已实现从静态的、表示气候特征的背景误差向基于集合样本估计的、能表示随天气形势演变特征的流依赖动态背景误差的升级(熊春晖等, 2013)。英国气象局采用在目标函数中增加扩展控制变量的方式将集合转置卡尔曼滤波(ETKF)估计的集合预报误差与气候背景误差结合,发展了En4DVar系统(Clayton, et al, 2013),加拿大气象局也发展了EnKF

与 4DVar 结合的 En4DVar 同化系统 (Buehner, et al, 2010a, 2010b)。业务数值预报中心的试验结果表明, 考虑集合预报误差改善了 4DVar 的分析质量, 特别是在南半球、热带地区, 并对台风路径预报也有显著改进。欧洲中期天气预报中心与法国气象局使用基于 4DVar 的集合资料同化方法来生成集合样本, 采用谱滤波来滤除噪声以保留有用的天气系统变化的集合预报误差并用于 4DVar, 显著地改善了分析和模式预报质量, 且正贡献在全球范围 10 d 预报中都有效 (Bonavita, et al, 2012)。此外, 英国气象局 (Lorenc, et al, 2015; Bowler, et al, 2017)、美国环境预报中心 (NCEP) 还进一步采用扩展控制变量方式将集合预报误差用于全球 4DEnVar (Wang, et al, 2013; Kleist, et al, 2015a, 2015b), 加拿大气象局也发展并业务应用了 4DEnVar 系统 (Buehner, et al, 2013)。

应用集合预报误差虽然可以通过不同方式实现, 但事实证明这些方式之间是等价的 (Buehner, 2005; Wang, et al, 2007)。可以看到, 在上述业务实践中, Lorenc (2003a, 2003b) 提出的增加扩展控制变量方法是引入集合预报误差的常用方案。增加扩展控制变量可以使得分析增量有更大的空间自由度, 从而帮助更好地引入观测信息。使用集合本来估计集合预报误差有优势, 但其不足也很明显。Lorenc (2003a) 指出, 当集合样本数较少时预报误差容易被低估, 且存在虚假遥相关, 需要进行误差协方差的局地化, 而局地化容易破坏集合样本中所包含的动力平衡关系 (Kepert, 2009)。

GRAPES 于 2018 年 6 月实现了全球 4DVar 业务运行 (Zhang, et al, 2019), 中国成为世界上为数不多的采用四维变分同化的全球业务预报中心之一。4DVar 方法的基本原理表明, 虽然其能在同化时间窗内通过切线性与伴随模式积分来隐式计算流依赖背景误差, 但其初始时刻的背景误差一直是气候态的。要克服这一缺陷, 需要在 4DVar 中引入集合预报误差, 发展 GRAPES 全球 En4DVar 技术, 从而实现在同化时间窗起始时刻就应用流依赖背景误差。

在 GRAPES 全球 4DVar 基础上进行拓展, 有效使用集合样本估计的集合预报误差来改进全球分析质量是本研究的重点。文中主要研究这几个

方面的问题: 首先, GRAPES 全球 4DVar 采用增量分析方案 (Courtier, et al, 1994), 在具体实施上有自己的特殊性 (薛纪善等, 2008), 有必要研究在该系统框架中增加扩展控制变量的科学方案。虽然已有一些研究基于 GRAPES 区域 3DVar 开展了使用集合预报误差的工作 (Chen, et al, 2015), 但全球 4DVar 中尚未开展相关研究。其次, 通过增加集合样本数来降低抽样误差, 并对集合预报误差给予较大权重是各业务中心同化技术发展的趋势。考虑到扩展控制变量的空间维数与集合样本数有关, 集合样本数越大对应的扩展控制变量的空间维数越大。在大于 100 个集合样本数的条件下, 如何在 GRAPES 全球 4DVar 中既能有效使用集合预报误差, 又使得扩展控制变量的维数增加受到限制, 以减少 4DVar 系统的计算与存储消耗。最后在 GRAPES 全球 4DVar 引入扩展控制变量并对其进行局地化时, 需要研究局地化对分析场平衡造成的影响, 并发展减缓局地化影响的办法。

## 2 GRAPES 全球 4DVar 背景误差公式表达

GRAPES 全球模式为水平经纬度格点模式。GRAPES 全球 4DVar 的气候背景误差矩阵 ( $\mathbf{B}$ ) 可以表示为  $\mathbf{B} = \mathbf{B}^{1/2} \mathbf{B}^{T/2} \approx \mathbf{U} \mathbf{U}^T$ 。  $\mathbf{B}^{1/2}$  是  $\mathbf{B}$  的平方根矩阵, 在具体实现时通过一系列误差结构函数 (或变量变换矩阵) 来完成。从极小化求解的控制变量 ( $\mathbf{v}$ ) 到分析变量 ( $\delta \mathbf{X}_1$ ) 的变换可以表示为

$$\delta \mathbf{X}_1 = \mathbf{U} \mathbf{v} = \mathbf{U}_p \mathbf{U}_k \boldsymbol{\varepsilon}_b \mathbf{U}_h \mathbf{U}_v \mathbf{v} \quad (1)$$

式中,  $\delta \mathbf{X}_1 = (\delta u, \delta v, \delta \pi, \delta q)^T$  分别是纬向风与经向风、无量纲气压及比湿的分析增量, 为  $n$  维空间向量。  $\mathbf{v} = (v_\psi, v_{\chi_a}, v_{\pi_a}, v_q)^T$  为控制变量, 包括流函数、非平衡速度势函数、非平衡无量纲气压及比湿。  $\mathbf{U}_h$  是水平变量变换, 在全球模式中采用水平谱滤波形式来模拟背景误差水平相关的各向同性分布特征, 实现分析增量信息由观测点向周围模式格点的传播。水平谱滤波需要采用水平高斯网格, 与水平经纬度网格不同, 因而在 GRAPES 全球 4DVar 中包含水平经纬度网格与水平高斯网格两套网格。  $\mathbf{U}_v$  为垂直变量变换, 采用经验正交函数分解 (EOF) 方式表达背景误差垂直相关结构。  $\boldsymbol{\varepsilon}_b$  为对角矩阵, 表示非平衡部分的背景误差均方根误差。式 (1) 中水平谱滤波和经验正交函数分解的非平衡分析变量的具体

实现过程可由下式表示

$$\delta X_u = \varepsilon_b U_h U_v \nu = \varepsilon_b G_{s \rightarrow d}^1 S^{-1} \sqrt{B_s} E \Lambda^{\frac{1}{2}} S W_g^{-\frac{1}{2}} \nu \quad (2)$$

式中,  $\delta X_u$  是非平衡分析变量,  $E$ 、 $\Lambda$  是背景误差垂直相关矩阵 ( $R_v$ ) 通过经验正交函数分解求得的特征向量与特征值, 其表达与模式垂直层次有关, 而与格点空间或谱空间无关, 可作用在每个谱系数上。  $S$ 、 $S^{-1}$  分别是格点空间到谱空间及从谱空间到格点空间的变换。  $B_s$  是谱空间的水平相关系数矩阵。  $W_g$  是勒让德变换权重系数。式(2)表明,  $\nu$  是垂直特征模式上的投影系数。  $G_{s \rightarrow d}^1$  是从控制变量 ( $\nu$ ) 所在的水平高斯格点到水平经纬度格点的插值算子。

式(1)中  $U_k$  表示背景误差中变量间的协相关部分, 也即平衡算子, 可表示为

$$\delta \pi = \delta \pi_b + \delta \pi_u = N \delta \psi + \delta \pi_u \quad (3)$$

$$\delta \chi = \delta \chi_b + \delta \chi_u = M \delta \psi + \delta \chi_u \quad (4)$$

式中,  $N$  是流函数对无量纲气压解释部分的平衡算子, 通过风压线性平衡方程或平衡方程与统计回归求解两种方法结合的方式给出 (王瑞春等, 2015)。  $M$  是流函数对势函数解释部分的平衡算子, 通过统计方式给出。通过  $U_k$  变量变换将相互独立非平衡分析变量  $(\delta \psi, \delta \chi_u, \delta \pi_u, \delta q)^T$  变换为分析变量全量  $(\delta \psi, \delta \chi, \delta \pi, \delta q)^T$ 。考虑到实际观测及模式预报变量是风场, 因而还需进行物理变换  $U_p$ , 从流函数、速度势函数变换为纬向风与经向风。此外,  $U_p$  还包括由静力平衡关系从无量纲气压 ( $\delta \pi$ ) 诊断得出位温 ( $\delta \theta$ )

$$\delta \theta_v = \frac{c_p \theta_{vb}^2}{g} \left( \frac{\partial (\delta \pi)}{\partial z} \right) \quad (5)$$

式中,  $c_p$  是摩尔定压热容,  $g$  是重力加速度。

式(2)中经验正交函数分解和水平谱滤波的参数由背景误差垂直相关模型和水平相关模型决定。误差垂直相关矩阵 ( $R_v$ ) 的构造由垂直相关模型 ( $\rho_{l,k}$ ) 决定, 与垂直层次所在的高度有关, 表示为

$$\rho_{l,k} = \frac{1}{1.0 + K_z (z_l - z_k)^2} \quad (6)$$

$$K_z = K \left( \frac{RT}{g} \right)^2 \quad (7)$$

式中,  $z_l$ 、 $z_k$  是任意模式层  $l$ 、 $k$  的平均高度。  $K$  是经验系数, 当  $K=0$  时, 所有垂直层次间的相关均为 1, 当其数值增大时, 垂直相关结构逐步变得更加局

地。垂直相关结构也可以通过误差样本统计给出 (王金成等, 2014)。龚建东等 (2020) 在 GRAPES 全球 4DVar 基础上发展了 EDA 方法生成多组集合样本, 也可以用于统计垂直相关结构。

定义新息向量  $d_n = y_n^o - \mathcal{H}_n M(X^b)$ , 其中  $y_n^o$  是  $t_n$  时的观测值, 在高分辨率外循环进行计算。定义  $\mathcal{H}$  为从模式状态变量到观测空间的观测算子, 其线性化算子表示为  $H$ 。定义  $M$  为非线性预报模式, 其切线性模式表示为  $M$ 。于是 GRAPES 全球 4DVar 的目标函数可以表示为

$$J(\nu) = \frac{1}{2} \nu^T \nu + \frac{1}{2} \sum_{n=0}^{L} (d_n - H_n M U \nu)^T R^{-1} (d_n - H_n M U \nu) + J_c \quad (8)$$

式中,  $R$  是观测误差协方差。  $t_0$  与  $t_L$  分别为同化时间窗起始与终止时刻。通过极小化求解式(8)可以求得  $t_0$  时刻分析增量  $\delta X_1$ 。  $J_c$  是目标函数约束项, 采用数字滤波方案约束重力波引发的不平衡 (刘艳等, 2019)。后文为公式表述简洁而略去  $J_c$  项。

由式(1)与(8)可知, 在 4DVar 的  $t_0$  时刻, 分析增量 ( $\delta X_1$ ) 利用气候背景误差 ( $B$ ) 或变量变化矩阵 ( $U$ ) 只能获得反映气候背景误差的分析增量, 而在  $t_n$  时刻的分析增量表示为  $\delta X_n = M \delta X_1$ , 通过模式动力约束可以部分反映流依赖特征。

### 3 GRAPES 全球 4DVar 引入集合预报误差

#### 3.1 扩展控制变量的引入

定义预报变量  $Z^f = \frac{1}{\sqrt{N-1}} \{\delta X_1^f, \dots, \delta X_N^f\}$ ,  $N$  为集合样本总数,  $\delta X_i^f$  为第  $i$  个集合样本相对于集合平均的偏差, 则集合预报误差可以表示为  $P_e^f = Z^f Z^{fT}$ 。定义分析增量  $\delta X_2$ , 它表示由  $\{\delta X_1^f, \dots, \delta X_N^f\}$  为基底所张成  $N$  维空间上的分析增量,  $\delta X_2 = Z^f \alpha$ 。向量  $\alpha = (\alpha_1, \dots, \alpha_N)^T$  表示分析增量  $\delta X_2$  在  $N$  维基底上的投影系数。考虑到集合预报误差  $P_e^f$  是  $n \times n$  维矩阵, 但其矩阵的秩只有  $N$  维, 直接使用  $P_e^f$  会造成虚假的遥相关, 在距观测位置较远的位置上产生虚假的分析增量。引入“Schur”乘积 (元素对元素的乘积, 表示为“ $\circ$ ”), 抑制虚假的遥相关, 缓解  $P_e^f$  的不满秩问题 (Lorenc, 2003a, 2003b)

$$\delta X_2 = \frac{1}{\sqrt{N-1}} \sum_{i=1}^N \delta X_i^f \circ \alpha_i \quad (9)$$

这时  $\alpha_i$  是  $\delta X_i^f$  对应的局地化分析变量。受限于高性能计算机的存储能力, 引入局地化分析变量后需要考虑对其维数进行控制。GRAPES 全球模式水平采用 Arakawa C 网格, 垂直方向采用 Chaney-Philips 模式分层, 不同分析变量 ( $\delta u$ 、 $\delta v$ 、 $\delta \pi$ 、 $\delta q$ ) 所在的模式网格空间位置不同, 如果对同一集合样本  $\delta X_i^f$  的不同分析变量都取同一个局地化分析变量  $\alpha_i$ , 并视质量点  $\delta \pi$  所在位置为共同位置, 可以降低  $\alpha_i$  的长度。

定义扩展控制变量  $v_i^\alpha$ , 与局地化分析变量  $\alpha_i$  的关系是

$$\alpha_i = C^{1/2} v_i^\alpha = U^\alpha v_i^\alpha \quad (10)$$

式中,  $C$  为局地化矩阵,  $C^{1/2}$  为其平方根矩阵, 可通过变量变换矩阵  $U^\alpha$  构造。  $U^\alpha$  可进一步表示为垂直局地化和水平局地化矩阵  $U^\alpha = U_h^\alpha U_v^\alpha$ 。  $U_h^\alpha U_v^\alpha$  的作用与  $U_h U_v$  的作用一致, 同样可以采用水平谱滤波和垂直经验正交函数分解实现。于是式(9)表示为

$$\delta X_2 = \frac{1}{\sqrt{N-1}} \sum_{i=1}^N (\delta u, \delta v, \delta \pi, \delta q)_i^{Tf} \circ U_h^\alpha U_v^\alpha v_i^\alpha \quad (11)$$

结合式(11), 以  $v_i^\alpha$  为扩展控制变量的目标函数定义为

$$J(v_1^\alpha, \dots, v_N^\alpha) = \frac{1}{2} \sum_{i=1}^N (v_i^\alpha)^T v_i^\alpha + \frac{1}{2} \sum_{n=0}^L (d_n - H_n M \delta X_2)^T R^{-1} (d_n - H_n M \delta X_2) \quad (12)$$

式(12)相较于式(8)的定义表明, 对以  $v_i^\alpha$  为控制变量的变分同化系统, 其计算量大小与集合样本数有关。需要构造  $v_i^\alpha$ , 使得其维数远小于  $\alpha_i$  的维数, 以减少存储和计算的负担。从局地化分析变量  $\alpha_i$  到扩展控制变量  $v_i^\alpha$  的变量变换(称为  $T^\alpha$  变换)可以表示为

$$v_i^\alpha = T^\alpha \alpha_i = T_v^\alpha T_h^\alpha \alpha_i \quad (13)$$

由于水平与垂直的局地化的尺度都要明显大于气候背景误差相关的尺度, 因而水平局地化可以只取前若干波数, 垂直局地化只取主导特征向量, 这样  $T^\alpha$  变换是一个降低维数的过程,  $v_i^\alpha$  的空间维数可以远小于  $\alpha_i$  的空间维数。  $T_v^\alpha T_h^\alpha$  的逆变换是  $U_h^\alpha U_v^\alpha$ , 用在式(11)中。

### 3.2 局地化分析变量实现方式

在 GRAPES 全球 4DVar 中实现局地化分析变量有两种方式, 一是只进行水平局地化而不进行垂

直局地化。此时, 扩展控制变量 ( $v_i^\alpha$ ) 蜕变为水平二维变量,  $\alpha_i = U_h^\alpha v_i^\alpha$ ,  $v_i^\alpha$  的维数显著减少, 因而可以不考虑水平局地化的降维处理。也即, 水平局地化所用的谱滤波的波数与 4DVar 内循环水平谱滤波的波数相同,  $v_i^\alpha$  定义在控制变量 ( $v$ ) 所在的水平高斯网格点上,  $\alpha_i$  定义在内循环低分辨率水平经纬度网格点上。这在 GRAPES 全球 4DVar 中表示为

$$\alpha_i = G_{s \rightarrow \alpha}^1 S^{-1} \sqrt{B_{sa}} S W_g^{-\frac{1}{2}} v_i^\alpha \quad (14)$$

式中,  $B_{sa}$  表示水平局地化变量在谱空间的水平相关系数矩阵。

另一种办法是进行三维局地化——同时考虑水平和垂直方向的局地化。此时, 通过在式(13)的  $T_h^\alpha$  水平谱滤波中取更少波数,  $T_v^\alpha$  垂直经验正交函数分解中取前几个主导特征向量来大幅度缩减控制变量 ( $v_i^\alpha$ ) 的维数。在 GRAPES 全球 4DVar 中表示为

$$\alpha_i = G_{s \rightarrow \alpha}^2 S_\alpha^{-1} \sqrt{B_{sa}} E_c A_c^{\frac{1}{2}} S_\alpha W_g^{-\frac{1}{2}} v_i^\alpha \quad (15)$$

式中,  $v_i^\alpha$  定义在比内循环水平高斯网格更低分辨率的水平高斯网格点上, 而  $\alpha_i$  定义在内循环低分辨率水平经纬度网格点上。在水平方向采用低分辨率谱空间, 则变量  $v_i^\alpha$  的水平格点数也相应减少。下标 c 表示对垂直局地化相关矩阵  $C_s = E A^{\frac{1}{2}} E^T$  进行截断, 仅取前若干个主导特征向量与特征值, 对应垂直方向相关“更宽”的模式。算子  $G_{s \rightarrow \alpha}^2$  为从高斯格点到低分辨率经纬度格点的双线性插值算子。

在实际应用时, 当局地化所采用网格的水平分辨率与内循环的分辨率不一致时, 需要重新定义一组网格用于局地化变量, 这会使得 GRAPES 的网格定义更为复杂。一种替代的办法是在水平局地化网格分辨率不变的情况下, 仅在垂直局地化时使用较少的主导特征向量。Clayton 等(2013)的研究表明, 当经验正交函数分解前若干个特征值对总误差的解释比例达到 95% 时可以获得满意的结果。目前的 GRAPES 全球 4DVar 中, 水平局地化与内循环的网格一致, 垂直方向按照解释比达到 90% 来截取主导特征向量。试验表明前 8 个特征值占比已经超过 90%, 后续试验中取前 8 个特征模态进行垂直局地化。

### 3.3 GRAPES 全球 En4DVar 公式表达

采用增加扩展控制变量, 是在具有气候背景误差特征的分析增量上增加扩展控制变量给出的具

备随流型演变分析增量, 最终获得的分析增量是这两部分增量的加权之和(Lorenc, 2003b; Wang, et al, 2007)

$$\delta X_1 = \sqrt{\beta_c} U v \quad (16)$$

$$\delta X_2 = \sqrt{\beta_e} \frac{1}{\sqrt{N-1}} \sum_{i=1}^N \delta X_i^f \circ U^\alpha v_i^\alpha \quad (17)$$

其中, 气候背景误差( $B$ )与集合预报误差( $P_c^f$ )所占的比重分别为 $\beta_c$ 、 $\beta_e$ , 满足关系 $\beta_c + \beta_e = 1$ 。

由新息向量定义, 分析值在观测空间投影( $y_n$ )表示为

$$y_n = \mathcal{H}_n M(X^b + \delta X_1 + \delta X_2) \quad (18)$$

于是 $y_n$ 与观测( $y_n^o$ )的偏差表示为

$$y_n^o - y_n = d_n - H_n M(\delta X_1 + \delta X_2) \quad (19)$$

考虑了扩展控制变量的目标函数表示为

$$J(v, v_1^\alpha, \dots, v_N^\alpha) = \frac{1}{2} v^T v + \frac{1}{2} \sum_{i=1}^N (v_i^\alpha)^T v_i^\alpha + \frac{1}{2} \sum_{n=0}^L (y_n^o - y_n)^T R^{-1} (y_n^o - y_n) \quad (20)$$

由式(20)及式(16)–(17), 对控制变量( $v$ )和扩展控制变量( $v_i^\alpha$ )的梯度表示为

$$\nabla_v J = v + \sqrt{\beta_c} U^T \sum_{n=0}^L M^T H_n^T R^{-1} (y_n^o - y_n) \quad (21)$$

$$\nabla_{v_i^\alpha} J = v_i^\alpha + \sqrt{\beta_e} \frac{1}{\sqrt{N-1}} (\delta X_i^f \circ U^\alpha)^T \sum_{n=0}^L M^T H_n^T R^{-1} (y_n^o - y_n) \quad (22)$$

对式(21)和(22)再次求偏导数, 计算目标泛函相对于控制变量的二阶偏导数(公式略), 构成海森矩阵可以发现, 将权重系数 $\beta_c$ 、 $\beta_e$ 放在分析增量的计算中(式(16)、(17)), 而不是放在目标函数(式(20))背景误差与集合预报误差项上, 主要是改进海森矩阵的条件数, 加快变分同化系统收敛。试验表明, 在 GRAPES 全球 4DVar 中若将权重放在背景误差与集合预报误差项上, 当 $\beta_c$ 、 $\beta_e$ 取值差异大时将显著影响极小化迭代收敛。

最后, 在 GRAPES 全球 4DVar 中具体实现的步骤是: 首先由外循环计算新息向量 $d_n$ , 并按照下列步骤计算

(1) 给定 $v$ 和 $v_1^\alpha, \dots, v_N^\alpha$ 初值为 0;

(2) 由控制变量( $v$ )通过式(16)计算 $\delta X_1$ , 由扩展控制变量( $v_1^\alpha, \dots, v_N^\alpha$ )通过式(17)计算 $\delta X_2$ ;

(3) 由式(19)计算分析值在观测空间投影( $y$ )与观测( $y^o$ )的偏差( $y_n^o - y_n$ );

(4) 由式(20)计算目标函数 $J(v, v_1^\alpha, \dots, v_N^\alpha)$ ;

(5) 由式(21)计算 $\nabla_v J$ 并由式(22)计算 $\nabla_{v_i^\alpha} J$ ; 从而获得目标函数对所有控制变量的梯度( $\nabla_{v, v_1^\alpha, \dots, v_N^\alpha} J$ );

(6) 利用目标函数( $J(v, v_1^\alpha, \dots, v_N^\alpha)$ )及梯度( $\nabla_{v, v_1^\alpha, \dots, v_N^\alpha} J$ ), 由最优下降算法求解 $v$ 和 $v_1^\alpha, \dots, v_N^\alpha$ ;

(7) 回到第(2)步, 进入下一次循环。如精度满足需求, 则退出循环。

由扩展控制变量( $v_1^\alpha, \dots, v_N^\alpha$ )通过式(17)计算 $\delta X_2$ 时, 实际计算是先由式(14)或式(15)计算并存储 $\alpha_i$ 值, 再由式(9)通过“Schur”乘积计算 $\delta X_2$ 。当 $\alpha_i$ 是二维空间变量时可以将所有的 $\alpha_i$ 叠放形成三维数组, 这时数组垂直方向表示不同集合样本对应的局地化分析变量。而当 $\alpha_i$ 是三维空间变量时, 在大规模高性能计算机上存储 $\alpha_i$ 将变得非常困难。在实际计算时对每个集合样本利用式(15)由 $v_i^\alpha$ 计算 $\alpha_i$ 后, 利用式(9)计算该集合样本对应的分析增量( $\delta X_2^i$ ), 每次对集合样本循环计算时, 每增加集合样本后获得的分析增量由下式求得

$$\sum \delta X_2^i = \sum \delta X_2^{i-1} + \delta X_2^i \quad i = 1, N \quad (23)$$

采用以上的计算策略时, 通过循环集合样本, 全部样本只需要存储一个 $\alpha_i$ 即可, 大幅度减少了对变量存储的需求。然而这种处理需要采用循环计算的方式反复计算 $U_h^\alpha U_v^\alpha$ 并依次求得每个 $\alpha_i$ , 在程序设计上更为复杂, 这与二维空间变量的处理方式有明显不同。

## 4 水平局地化对地转平衡约束关系的影响

### 4.1 水平局地化尺度影响

GRAPES 全球 4DVar 的气候背景误差的水平相关模型采用融合水平相关模型(龚建东等, 2020), 水平局地化模型采用高斯型相关函数。借鉴 Xu 等(2008a, 2008b)提出的在分析时刻前后利用集合样本在时间序列上的扩展来增加集合样本数的方法, 试验利用 20 个 NCEP 全球集合预报的 6 与 9 h 预报的集合样本, 并采用扣除对应时间集合平均的扰动样本来减缓因预报时长不一致带来的大尺度场存在的相位误差, 总计 40 个样本。任选 2018 年

7月9日太平洋北部地区一次锋面天气过程进行分析。由集合样本估计出来的集合预报误差如图1所示,在300 hPa 高度水平面上(图1a、b,约为模式35层)集合样本估计的气压与风场误差最大的位置位于冷锋槽及槽后位置,在沿54°N的垂直剖面上(图1c、d),气压误差随高度升高而递减,槽后明显大于槽前,纬向风场的误差主要位于风速急流区附近。

将单个观测资料放置在冷锋锋面后的西北平直气流上,位于(54°N, 173°W)的模式35层(约250 hPa)位置(图1)。取气候背景误差与集合预报误差的权重分别为0.1与0.9,这时分析增量主要体现集合预报误差的作用,并考虑不同水平局地化尺度参数来研究其对分析增量的影响,试验结果如图2所示。水平局地化采用高斯相关分布特征,在水平局地化

尺度的4倍距离的位置,水平相关系数基本降为0。当水平相关尺度参数取值较大时(如28000 km),地球上最远大圆距离(20000 km)处的分析增量受局地化作用将会衰减到原值的70%,水平局地化基本不起作用,可以作为不进行水平局地化的试验结果。由于集合样本数有限,直接由集合样本估计 $P_e$ 会产生虚假的遥相关,在距观测位置较远的距离上会产生虚假的分析增量(图略)。在单点试验中分析增量不仅分布在观测点周围,在远离观测点的位置也有较大的分析增量(图2a)。试验发现当局地化水平相关尺度在1500 km以下时距离观测点较远处的虚假分析增量明显受到抑制。例如:图2b给出的水平相关尺度在850 km时的分析增量,由于所选单点在锋面后且集合预报误差的权重占主

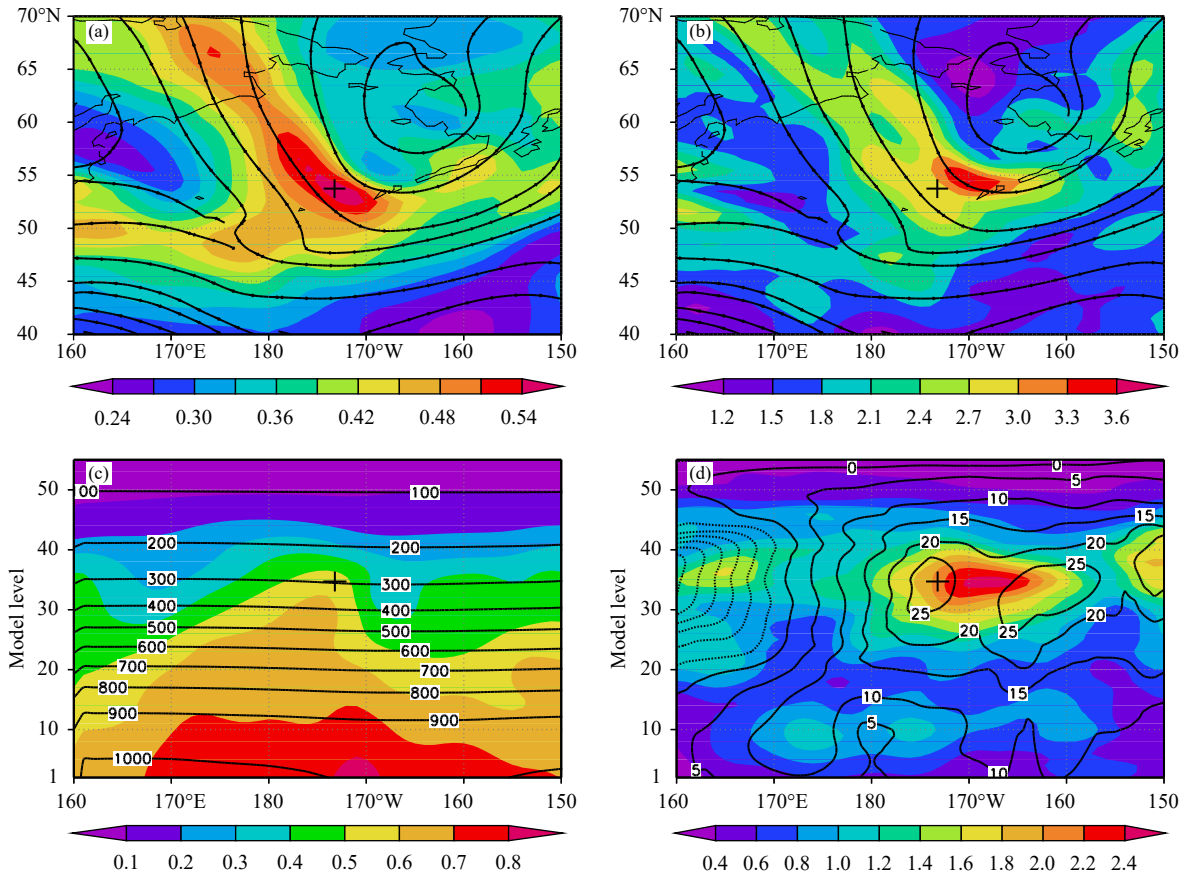


图1 40个集合样本估计的 $t_0$ 时刻集合预报误差(色阶)(a. 气压误差,单位: hPa; b. 风场误差,单位: m/s; c. 气压误差垂直剖面,单位: hPa; d. 风速误差垂直剖面,单位: m/s; a、b中流线为模式背景风场, c、d中等值线分别为模式背景气压与纬向风速;图中“+”为单点观测位置,水平位置为(54°N, 173°W),位于模式面35层)

Fig. 1 Estimated ensemble forecast errors at time  $t_0$  with 40 ensemble samples (shaded)(a. pressure error, unit: hPa; b. wind error, unit: m/s; c. vertical cross section of pressure error, unit: hPa; d. vertical cross section of wind error, unit: m/s; streamlines in (a) and (b) are background wind field, contours in (c) and (d) are background pressure and zonal wind, respectively, the cross "+" denotes observation site, which is located at(54°N, 173°W)and on model level 35)

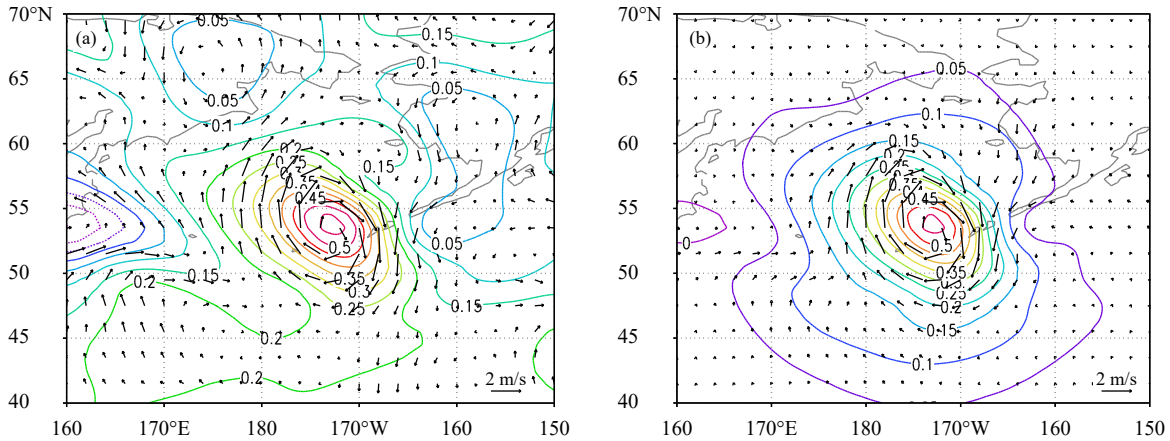


图2 局地化水平相关尺度(a. 28000 km, b. 850 km)对单点气压观测产生的分析增量的影响  
(新息向量大小为 1 hPa, 观测时间位于  $t_0$  时刻, 观测位置如图 1 所示, 图中等值线为气压, 单位: hPa)

Fig. 2 Impacts of correlation scale of horizontal localization on analysis increment when assimilating single-point pressure observations (a. 28000 km, b. 850 km; the innovation is 1 hPa at time  $t_0$ , the observation site is the same as that shown in Fig. 1; the contours are for pressure, unit: hPa)

导, 因而分析增量呈现出随天气流型演变分布的特征——即分析增量沿气流方向延展较长, 而在垂直气流方向分析增量范围受到压缩。当不考虑集合预报误差, 仅有气候背景误差时在同化时间窗起始 ( $t_0$ ) 分析增量的特点如图 9a 所示。

#### 4.2 非平衡变量空间局地化

Mitchell 等 (2002)、Clayton 等 (2013) 研究表明, 直接对水平风场及气压进行局地化时, 由于集合样本中的水平风与气压变量之间已经包含大尺度准地转平衡约束关系, 进行水平局地化容易破坏这种关系。以纬向风 ( $\delta u$ ) 为例, 其地转分量 ( $\delta u_g$ ) 与气压 ( $\delta p$ ) 的关系可以表示为

$$\delta u_g = -\frac{1}{f\rho} \frac{\partial \delta p}{\partial y} \quad (24)$$

式中,  $f$  为科里奥利参数,  $\rho$  为空气密度。当对无量纲气压和风场分量进行局地化时, 等式右侧包含了对局地化函数的偏导数项而左侧没有, 从而破坏对准地转平衡关系的表述。Clayton 等 (2013) 研究表明, 如果对相互独立的分析变量或非平衡变量进行局地化时, 将减缓局地化造成的影响。在 GRAPES 全球 4DVar 中, 先对  $\delta X_i^f$  进行物理逆变换, 再将其中所包含的准地转平衡部分抽取出来, 仅保留变量间不相关的非平衡部分。

$$\delta X_u^i = (\delta\psi, \delta\chi_u, \delta\pi_u, \delta q)_i^T = \mathbf{T}_K \mathbf{T}_P \delta X_i^f \quad (25)$$

$\mathbf{T}_P$  主要实现从分析变量  $(\delta u, \delta v, \delta p, \delta q)^T$  到分析变

量  $(\delta\psi, \delta\chi, \delta\pi, \delta q)^T$  的物理逆变换。 $\mathbf{T}_K$  变换是抽取  $(\delta\psi, \delta\chi, \delta\pi, \delta q)^T$  中流函数与势函数平衡关系, 以及流函数与无量纲气压平衡关系, 给出相互独立的非平衡分析变量  $(\delta\psi, \delta\chi_u, \delta\pi_u, \delta q)^T$ 。此时对非平衡分析变量进行水平局地化, 由于平衡部分已经被抽取出来, 就不会影响平衡特征。在水平局地化后, 重新使用  $\mathbf{U}_K$  变量变换, 计算流函数对应的气压平衡部分。于是式 (11) 可以表示为

$$\delta X_2 = \mathbf{U}_P \mathbf{U}_K \left\{ \frac{1}{\sqrt{N-1}} \sum_{i=1}^N (\mathbf{T}_K \mathbf{T}_P \delta X_i^f) \circ \mathbf{U}_h^\alpha \mathbf{U}_v^\alpha \mathbf{v}_i^\alpha \right\} \quad (26)$$

图 3 给出了分别在分析变量空间与非平衡分析变量空间进行局地化时, 对分析平衡影响的差异。为方便比较起见, 同时给出水平局地化相关尺度 28000 km 的结果, 作为不进行水平局地化的对比试验。由图 3c、f 可见, 当不进行局地化时, 同化时间窗起始时刻 ( $t_0$ ) 的地面气压分析增量与 Z=35 层 (约 250 hPa) 的分析增量对应, 但随着切线性模式积分, 在同化时间窗终止时刻 ( $t_L$ ) 分析增量最大值区向东北移动, 同时在锋面前端出现接近 -0.6 hPa 的气压增量。由图 1c 可知, 集合样本估计的集合预报误差在模式低层的数值远大于 35 层, 单点分析增量在模式低层出现比观测位置更强的分析增量。而当水平局地化尺度减小时, 以水平局地化相关尺度 850 km 的结果为例, 在同化时间窗



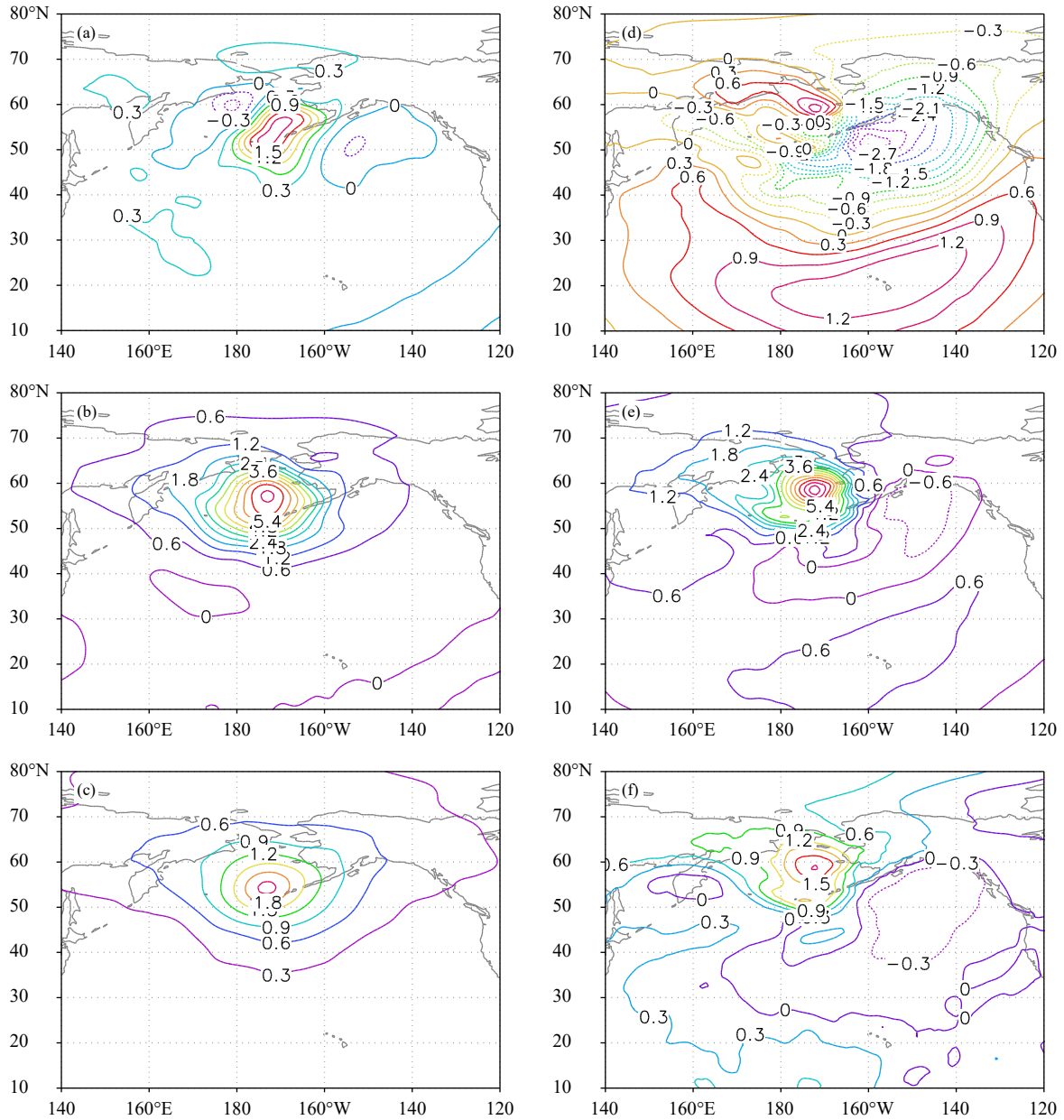


图3 水平局地化对分析增量准地转平衡的影响 (a、d. 分析变量空间分布; b、e. 非平衡分析变量空间分布, 局地化水平尺度为 850 km; c、f. 分析变量空间分布, 水平局地化尺度为 28000 km; a、b、c.  $t_0$ , d、e、f.  $t_L$ )

Fig. 3 Impacts of horizontal localization on geostrophic balance of analysis increments (a, d. localization on analysis variables; b, e. unbalanced analysis variables, and the horizontal correlation scale of localization is 850 km; c, f. analysis variables, and the scale of localization is 28000 km; a, b, c.  $t_0$ , d, e, f.  $t_L$ )

起始时刻 ( $t_0$ ), 模式第一层同样出现与高层分析增量对应的分析增量正值区, 但在同化时间窗终止时刻 ( $t_L$ ), 在锋面前端出现 -3 hPa 的气压增量, 并在远离观测位置区域出现大片区域的气压增量。该气压增量围绕观测位置以环形分布, 其与观测位置的距离随水平局地化尺度变化。进一步分析发现其位置与高斯水平相关模型二阶导数的最大值处对

应。高斯相关模型的二阶导数是表征气压与风场的准地转耦合关系, 分析图 3d 表明, 在分析变量空间进行局地化将产生虚假的分析增量。而当在非平衡分析变量空间进行局地化时 (图 3b、e), 可以发现其结果与不进行水平局地化的结果相似, 分析增量正值区向东北方向移动, 同样在锋面前出现 -0.6 hPa 的分析增量。上述结果表明将集合预报误

差引入 GRAPES 全球 4DVar 时, 采用在非平衡变量空间进行局地化来保证气压与风场的准地转耦合关系非常重要。

## 5 三维局地化对分析平衡的影响

### 5.1 垂直局地化特征尺度

GRAPES 全球 4DVar 采用静力平衡分析方案, 根据静力平衡关系通过式(5)由无量纲气压增量计算虚位温增量。垂直局地化的相关结构可以采用式(6)和(7)所给出的垂直相关模型来实现, 也可以直接利用模式预报的统计样本来估计。王金成等(2014)比较了垂直相关模型与同一检验时刻的不

同预报时效统计样本估计出的垂直相关结构的差异, 结果表明统计样本估计的垂直相关结构更加合理。文中采用 EDA 方法生成的集合样本来估计垂直相关结构( $R_v^s$ )存在因样本数有限造成的虚假遥相关, 文中利用相关范围较宽的垂直相关模型( $R_v$ )对距离远的统计噪音进行衰减限制, 表示为  $R_v \circ R_v^s$ , “ $\circ$ ”表示“Schur”乘积, 结果如图 4b 所示。采用垂直相关模型给出的垂直相关结构( $R_v$ )在不同层次变化较为平滑, 而集合样本统计获得的垂直相关结构( $R_v^s$ )在边界层相关较强, 在边界层之上对流层中层垂直相关迅速减弱, 到对流层中高层垂直

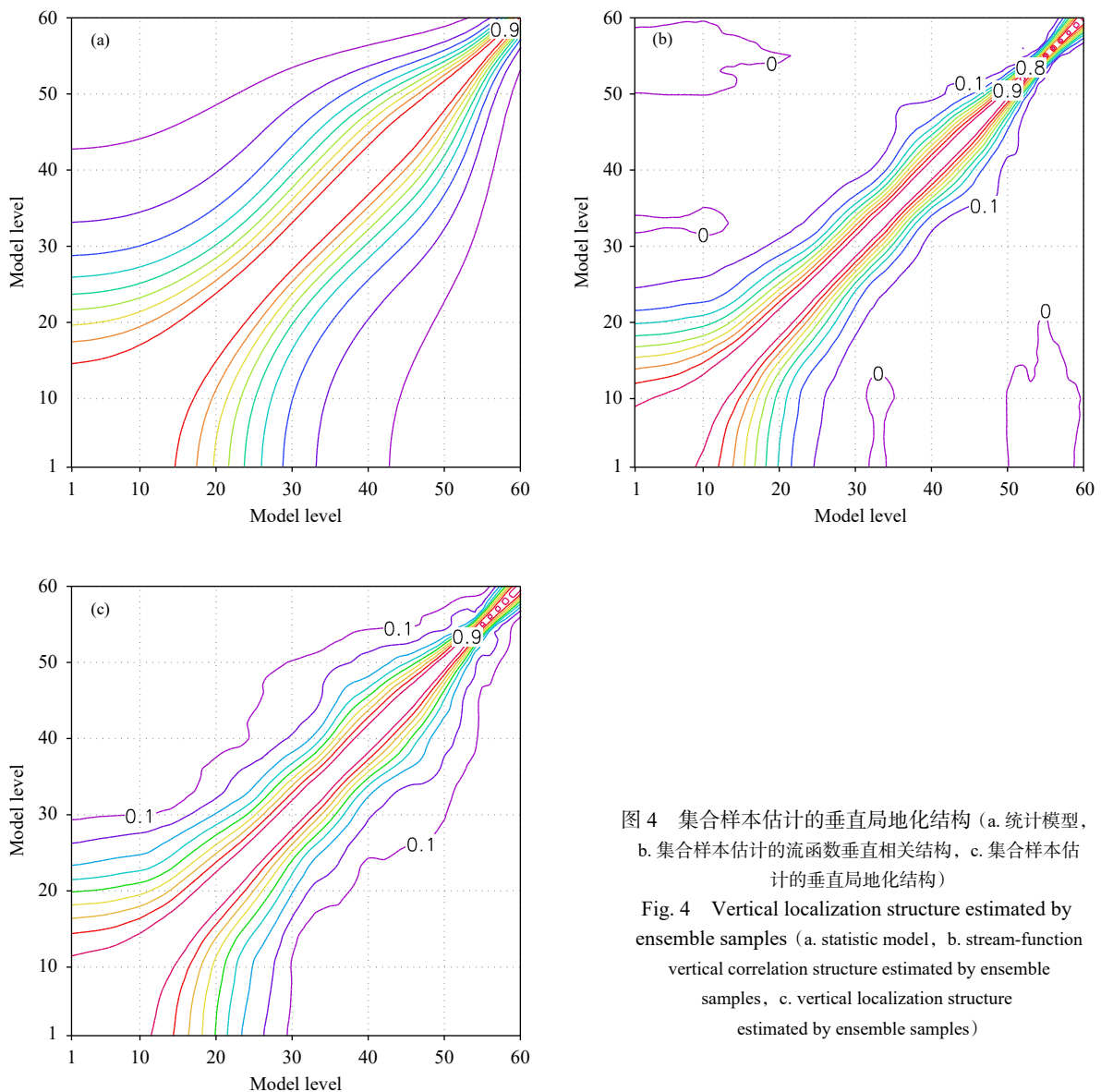


图 4 集合样本估计的垂直局地化结构 (a. 统计模型, b. 集合样本估计的流函数垂直相关结构, c. 集合样本估计的垂直局地化结构)

Fig. 4 Vertical localization structure estimated by ensemble samples (a. statistic model, b. stream-function vertical correlation structure estimated by ensemble samples, c. vertical localization structure estimated by ensemble samples)

相关又进一步变强。总体上,统计的垂直相关结构明显窄于相关模型。借鉴 Clayton 等(2013)的方法,使用流函数的垂直相关结构进行所有变量的垂直局地化,垂直局地化结构表示为  $C_s = (R_v \circ R_v^s)^P$ , 这里使用参数  $P$  控制垂直局地化的幅度,当参数  $P < 1$  时,垂直相关变宽,试验中取为 0.5。这时,局地化垂直相关结构如图 4c 所示。比较垂直相关模型(图 4a)给出的垂直局地化结构,可以发现两者在边界层比较类似,而在对流层中层与高层差异较大。在实际应用垂直局地化矩阵( $C_s$ )时,为了避免模式层顶附近的层次存在较大噪音的干扰,不考虑模式层顶最高的 5 层。

进行垂直局地化时,要对无量纲气压增量( $\delta\pi$ )进行局地化,由于式(5)右端涉及到  $\delta\pi$  在垂直方向的梯度,对  $\delta\pi$  进行局地化会显著改变梯度的计算,进而影响到虚位温的计算。但由于虚位温是通过

静力平衡关系计算得到的,因而垂直局地化后的分析场仍然满足静力平衡关系。使用单点试验来分析垂直局地化及局地化相关尺度对分析的影响。由图 5 可见,如果垂直局地化相关尺度较大(式(7)中经验系数  $K$  取为 1)时,单点观测产生的气压分析增量分布在整个模式层次。对比图 1c 可以发现,由于集合预报气压误差从观测位置向地面呈现出逐步增大的趋势,垂直局地化对气压分析增量的影响非常大。对比图 1d 可以发现,纬向风场误差大值区在垂直方向位于模式 30—40 层(400—200 hPa),相应的纬向风场分析增量也在 30—40 层,随垂直相关尺度的影响比较小。当垂直局地化相关尺度适当时( $K$  值为 7—20),分析增量被约束在观测点附近,气压分析增量与风场分析增量对应较好。采用集合样本估计的垂直局地化结构对气压分析增量在垂直方向的限制作用更强。垂直局地化的影

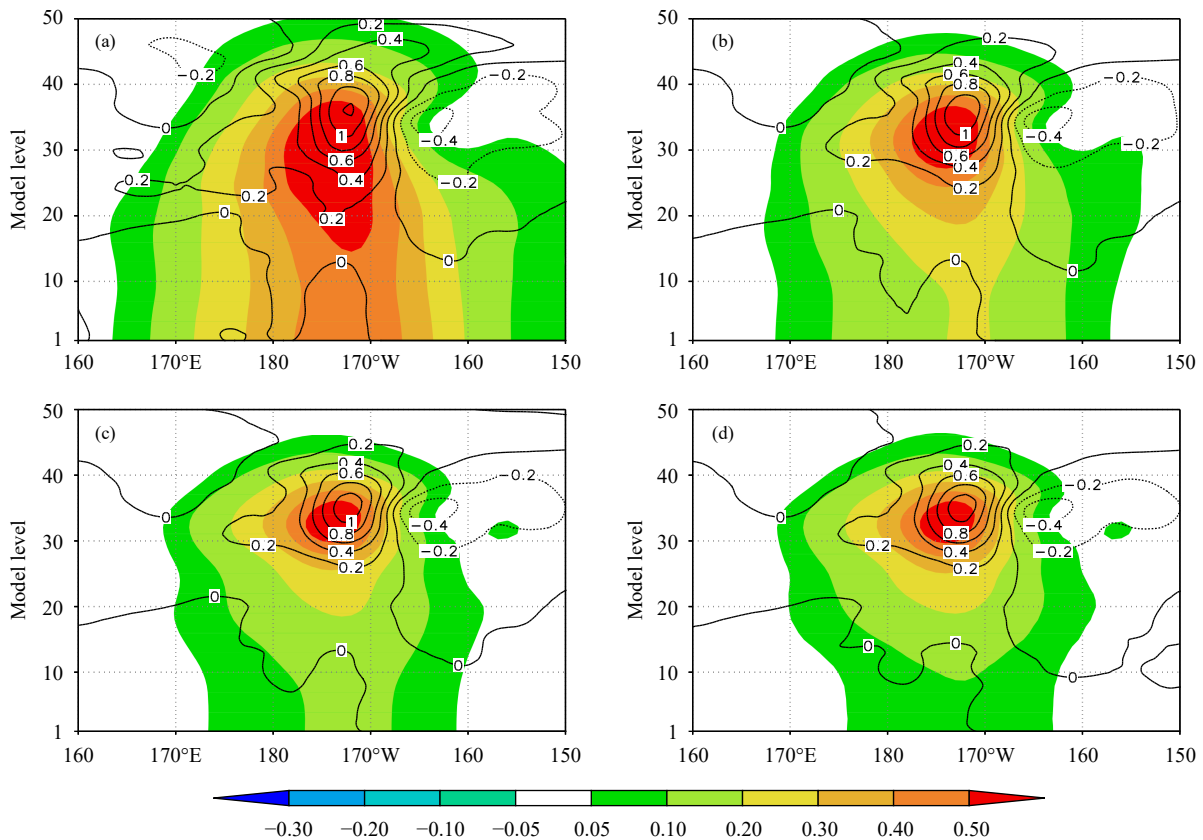


图 5 垂直局地化相关尺度对单点气压观测分析增量的影响 (a.  $K=1$ , b.  $K=7$ , c.  $K=20$ , d. 集合样本估计的局地化尺度; 等值线为纬向风分析增量, 单位: m/s; 色阶为气压分析增量, 单位: hPa)

Fig. 5 Impacts of correlation scale of vertical localization on analysis increment when assimilating single-point pressure observations (a.  $K=1$ , b.  $K=7$ , c.  $K=20$ , d. correlation scale of vertical localization estimated by ensemble samples; contours are for zonal wind analysis increment, unit: m/s; shaded are for pressure analysis increment, unit: hPa)

响表明,影响观测信息在空间传播的重要因素来自集合预报误差本身,垂直局地化主要起到约束虚假遥远相关的作用。这也表明当考虑集合预报误差在GRAPES全球4DVar中应用时,集合样本质量及其估计的预报误差的合理性,对集合预报误差的应用起到关键作用。

进一步考察垂直局地化对位温分析增量的影响,结果如图6所示。由式(5)可知,位温分析增量一般位于无量纲气压增量( $\delta\pi$ )的垂直梯度最大的位置。当相关尺度取不同数值时, $\delta\pi$ 的垂直梯度发生了明显变化,由静力平衡关系导出的位温分析增量同样也发生明显变化。当垂直局地化相关尺度较大时位温分析增量最大在1 K左右,而当垂直局地化相关尺度较小时位温分析增量可以达到1.4 K左右。采用集合样本估计的垂直相关模型,其对无量纲气压与虚位温分析增量的影响,与垂直相关模型的影响较为一致。这表明垂直局地化结构对分

析增量的影响较小,影响较大的是垂直相关尺度,需要合理给定。

## 5.2 分析场平衡特征分析

4DVar中低分辨率切线性模式积分时存在由诸多因素引起的高频重力振荡,当集合预报误差引入4DVar后,与之相随的水平与垂直局地化也是造成分析不平衡、激发高频重力振荡的原因之一。刘艳等(2019)研究表明,通过在4DVar中引入基于数字滤波方案的初始化过程,可以很好地抑制高频振荡过程,使得分析场在动力上更为平衡。由式(8), $J_c$ 是目标函数约束项,其数值反映了对重力波的控制幅度。以单点试验为例比较4DVar与En4DVar,分析不同局地化方案对 $J_c$ 项收敛的影响,结果如图7所示。4DVar在单点观测的目标函数迭代下降27步后, $J_c$ 项数值减小到0.14。而在En4DVar中, $J_c$ 项收敛过程要慢于4DVar,最终收敛数值在0.2左右,也要明显大于4DVar的结果。采用不同

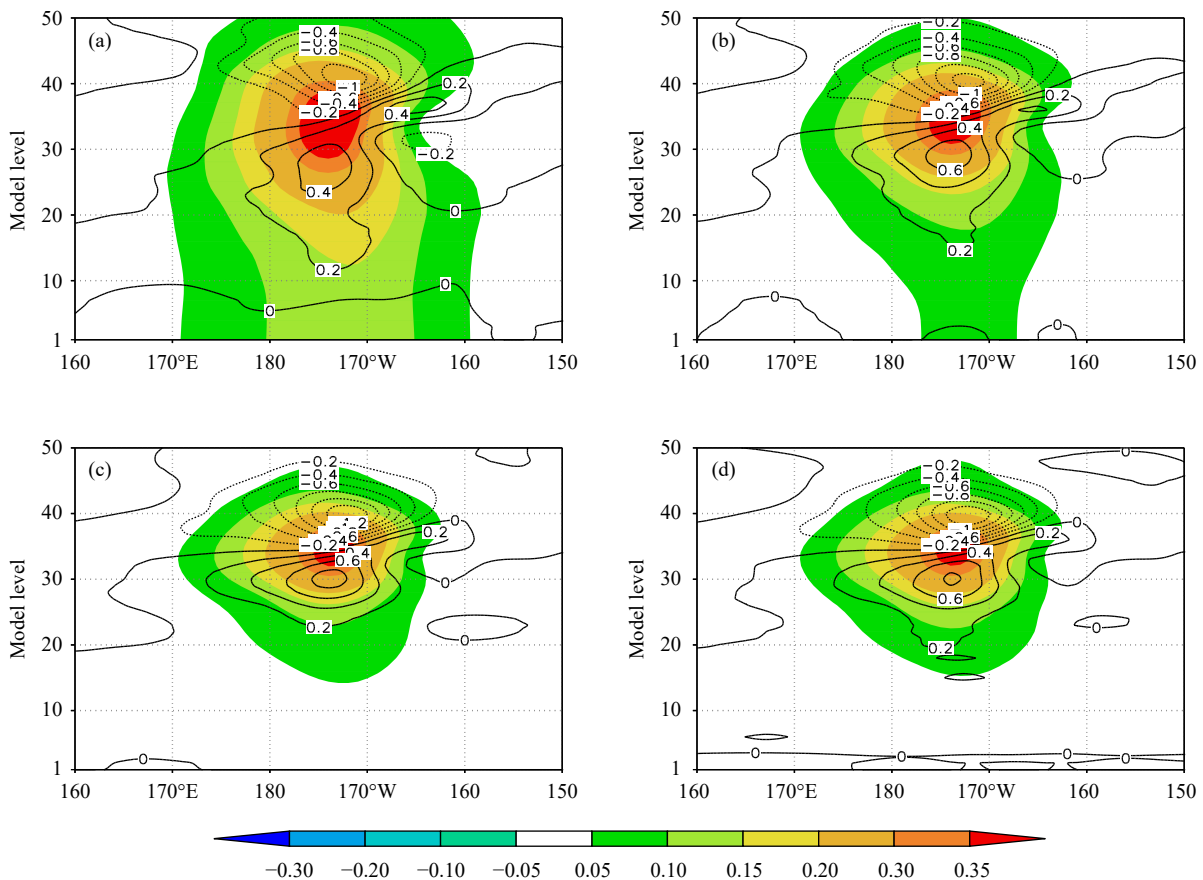


图6 同图5,但为虚位温分析(等值线为虚位温分析增量,单位:K;色阶为无量纲气压分析增量)

Fig. 6 Same as Fig. 5 but for virtual potential temperature (Contours are for virtual potential temperature analysis increment, unit: K; shaded are for non-dimensional pressure analysis increment)

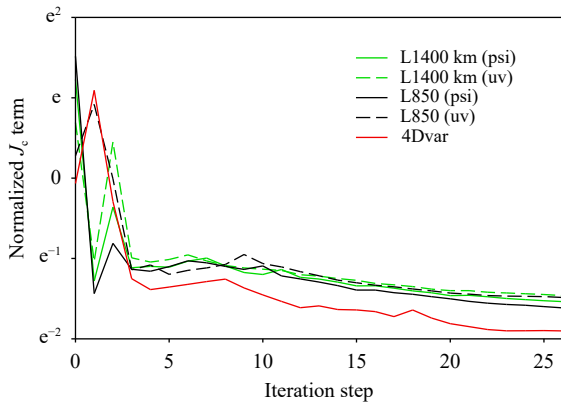


图 7 集合预报误差引入 4DVar 对目标函数约束项  $J_c$  的影响

Fig. 7 Ensemble forecast errors introduced into 4DVar and their impact on  $J_c$  constraint term in cost-function

局地化相关尺度的结果类似, 不过在非平衡分析变量空间进行局地化的结果要略好于在分析变量空间进行局地化。这表明采用集合预报误差后, 切线性模式积分产生的高频重力振荡要强于 4DVar。不过由于目标函数中  $J_c$  项的存在, 分析场仍能保证动力平衡约束关系。

这里将 4DVar 退化为 3DVar, 不考虑切线性模式积分的影响, 以进一步评估水平与垂直局地化对分析场平衡特征的影响。利用引入集合预报误差的 3DVar 同化实际常规资料与卫星资料, 并用分析场驱动 GRAPES 全球模式, 通过模式启动初期地面气压的时间倾向来分析局地化对分析场平衡特征的影响(图 8)。由于背景场中未引入新的同化增量, 由其驱动的模式轨迹平衡特征较好, 这里作为对照试验。从图 8 中可以看到, 当在分析变量空间进行水平局地化时, 模式积分初始时刻的地面气压倾向明显大于背景场启动的结果, 且在 5 h 积分后仍然高于背景场结果。而在非平衡分析变量空间进行水平局地化时, 平衡特征明显改善。这进一步说明, 在非平衡分析变量空间进行水平局地化更有优势。当在水平局地化基础上进一步考虑垂直局地化时, 地面气压倾向增加十分显著。这表明相较于水平局地化, 垂直局地化是造成分析不平衡的主要原因。对不同大小的垂直局地化相关尺度, 随局地化特征尺度变宽, 不平衡程度会略有所减缓, 但改善不大。这表明垂直局地化造成的不平衡与垂直局地化的相关尺度有关, 但其影响较小。此外,

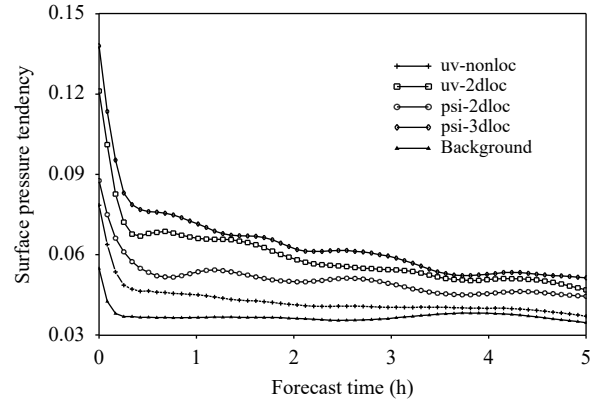


图 8 背景与分析地面气压倾向随预报时间的演变(水平局地化尺度为 1500 km, background 为背景, uv-nonloc 为引入集合预报误差但不进行局地化, uv-2dloc 为对分析变量进行二维局地化, psi-2dloc 为对非平衡分析变量进行二维局地化, psi-3dloc 为对非平衡分析变量进行三维局地化)

Fig. 8 Temporal evolution of background and analysis surface pressure tendency (the correlation scale of horizontal localization is 1500 km; "background" indicates background field, "uv-nonloc" indicates the introduced ensemble forecast errors without non-localization, "uv-2dloc" indicates horizontal localization on analysis variables, "psi-2dloc" is for horizontal localization on unbalanced analysis variables, "psi-3dloc" is for three-dimensional localization on unbalanced analysis variables)

采用集合样本统计给出的垂直局地化结构对分析平衡的改善也有限(图略)。

## 6 集合预报误差对分析的影响

### 6.1 集合预报误差权重影响

气候背景误差与集合预报误差所起的作用由其权重系数  $\beta_c$ 、 $\beta_e$  来决定。当  $\beta_e$  逐步增大时, 集合预报误差逐步占主导。试验利用 40 个集合样本, 当  $\beta_e = 1$  而  $\beta_c = 0$  时, 同化时间窗起始时刻的分析增量呈现各向同性(图 9a)。随着  $\beta_e$  的值逐步增大, 分析增量明显沿着风速较大的方向进行延伸, 而与风速垂直的方向分析增量受到抑制(图 9c)。起始时刻分析增量呈现出随集合背景误差变化的特征, 也即背景误差与背景场的动力演变机制有关, 具备随天气系统流型而变的特征。对于仅采用气候背景误差的分析增量, 由于切线性模式动力演变机制的存在, 在同化时间窗终止时刻分析增量也呈现出随流型变化的特点(图 9b), 而对于集合预报误差占主导作用的分析增量(图 9c、d), 终止时刻的分析增量随流型变化的特点更加明显。对比图 9c 和 d, 终止

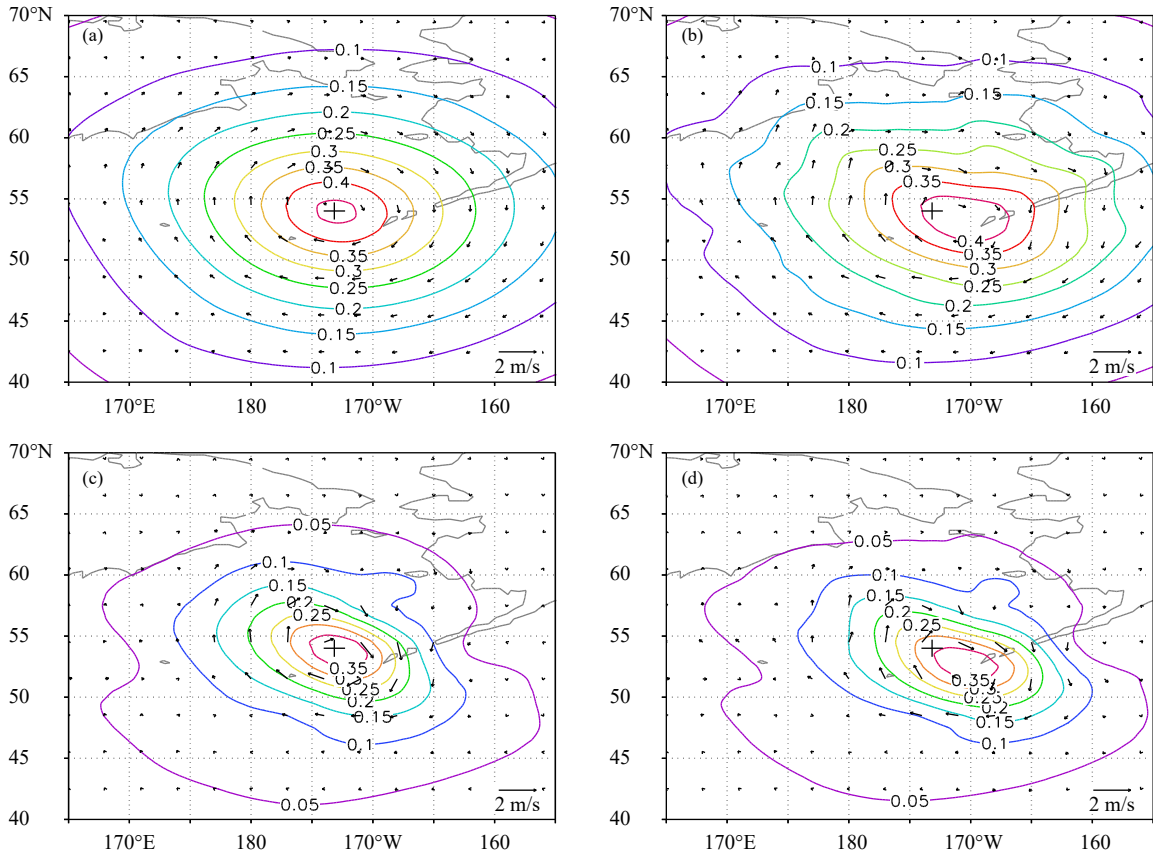


图9 4DVar同化时间窗内分析增量的变化

(a, b. 气候背景误差在 $t_0$ 与 $t_L$ 时的分析增量, c, d. 引入集合预报误差后 $t_0$ 与 $t_L$ 时的分析增量; 等值线为气压, 单位: hPa, 箭头为风)

Fig. 9 Evolution of analysis increment in the 4DVar assimilation time window (a, b. analysis increments at  $t_0$  and  $t_L$  with climatic background error; c, d. increments at  $t_0$  and  $t_L$  after introducing ensemble forecast errors; the contours are for pressure, unit: hPa, arrows are wind speed)

时刻的分析增量与起始时刻的分析增量在特征上更加一致, 流依赖分析增量的中心位置随气流向下游传播, 这是使用集合预报误差对4DVar改进的一个主要特征。

## 6.2 计算效率影响

对  $N$  个集合样本的抽样统计, 则对应的蒙特卡罗抽样误差是  $\frac{1}{\sqrt{N}}$ , 当集合样本数越小时, 抽样误差越大。当集合样本数大于 100 时, 抽样误差低于 10% (图 10)。文中研究的重点是在大于 100 个集合样本情况下引入扩展控制变量后对计算效率的影响。由于在 4DVar 目标函数迭代循环的总计算时长中, 切线性模式和伴随模式的计算占主导。首先关闭切线性模式和伴随模式将 4DVar 退化为 3DVar, 利用实际常规与卫星观测资料评估增加扩展控制变量及不同规模集合样本数对并行计算效

率的影响。此时在内循环水平分辨率设为  $1.0^\circ$ , 气候控制变量 ( $\mathbf{v}$ ) 定义在水平高斯网格上, 垂直方向取 58 个特征模态。迭代收敛判据取为目标函数梯度下降到初始梯度的  $10^{-4}$ , 采用 256 个 CPU 的情形下, 完成一次分析需迭代 65 步, 内循环墙钟时间 194 s。在此基础上引入三维扩展控制变量 ( $\mathbf{v}_i^e$ ), 水平方向采用与气候控制变量 ( $\mathbf{v}$ ) 一致的高斯网格, 垂直方向采用前 8 个主导特征模态, 此时每个集合样本对应的三维扩展控制变量的数目为水平高斯网格数的 8 倍。当集合样本取 20、40、60、120、180 个时, 气候控制变量与扩展控制变量之和相对于气候控制变量分别增加至 1.7、2.4、3.1、5.1、7.1 倍, 墙钟时间分别增加至 1.4、1.8、2.1、2.9、4.1 倍。因为垂直局地化所选择的垂直模态数明显小于气候控制变量 ( $\mathbf{v}$ ) 的垂直相关模态数, 因而扩展控制变量的格点数增加有限。而当采用二维局地化扩展

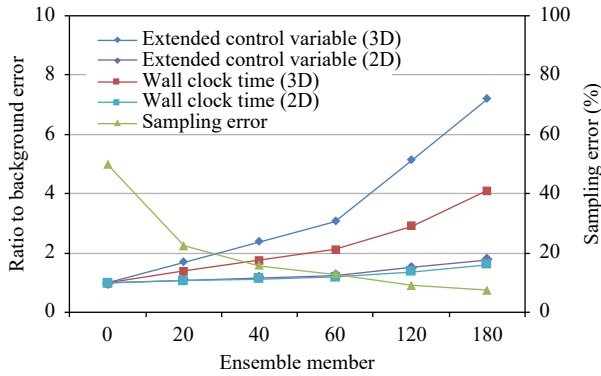


图 10 增加扩展控制变量与集合样本数对计算效率的影响 (采用 GRAPES 3DVar 参数配置的数值试验; 右侧纵轴为集合样本数对应的抽样误差)

Fig. 10 Impacts of adding extended control variables and the number of ensemble samples on computational efficiency (numerical experiment with GRAPES 3DVar configuration; the vertical axis on the right shows the sampling error corresponding to the number of ensemble samples)

控制变量时, 同样 20—180 个集合样本, 气候控制变量与扩展控制变量之和相对于气候控制变量增加至 1.1、1.2、1.3、1.5、1.8 倍, 墙钟时间增加至 1.1、1.1、1.2、1.4、1.6 倍。比较三维与二维扩展控制变量, 在集合样本数较大时二维扩展控制变量的计算效率更高。然而正如前文分析, 采用三维扩展控制变量时分析增量在垂直方向的分布更加合理。这表明文中所采用的扩展控制变量引入方式在有限增加扩展控制变量数量的同时, 较好地解决了计算量和计算效率的问题, 具有实用性。对于采用气候控制变量的 4DVar, 由于切线性模式和伴随模式的计算占主导, 相对于 3DVar 的 194 s, 内循环墙钟时间增加到 1340 s, 计算时间增加至 6.9 倍。比较采用扩展控制变量的 4DVar 和仅采用气候控制变量的 4DVar, 对 60 个集合样本, 引入扩展控制变量后墙钟时间从 1340 s 增加到 1751 s, 墙钟时间增加约 30%。

## 7 总结与讨论

GRAPES 模式是中国自主发展的非静力格点模式, 全球模式采用三/四维变分资料同化作为分析方法。文中研究在现有四维变分资料同化的基础上, 有效使用集合样本估计预报误差扩展变分资料同化的分析能力。围绕 GRAPES 全球 4DVar 在具体实施上的特殊性, 探讨了在引入扩展控制变量过

程中, 如何在大于 100 个集合样本数条件下, 增加扩展控制变量的同时能显著降低扩展控制变量的维数增加, 并减少对计算机存储和计算的压力, 提高算法的实用性。此外, 重点研究了水平局地化与垂直局地化对分析场平衡性的影响, 通过单点试验、常规与卫星观测资料试验, 分析了局地化对地面气压倾向的影响。

研究表明, 可以充分利用 GRAPES 全球 4DVar 气候背景误差结构模型在水平方向采用谱滤波和在垂直方向采用经验正交分解的特点, 利用谱滤波实现水平局地化, 利用经验正交分解实现垂直局地化。考虑到水平局地化所采用的相关模型更宽, 对谱滤波可以采用更少的波数、更粗的高斯网格来降低扩展控制变量的空间维数。考虑到需要独立定义扩展控制变量所依托的网格, 在实施上较为复杂, 目前 GRAPES 变分同化系统只用了较宽的水平特征长度模型, 没有再定义独立的高斯网格。对垂直局地化, 可以通过仅采用前 8 个主导特征模态来降低空间的维数, 此时主导特征模态方差解释占比超过 90%。在引入 20—180 个集合样本的情形下, 对二维局地化, 控制变量增加至 1.1—1.8 倍, 墙钟时间增加至 1.1—1.6 倍。对三维局地化, 仅考虑垂直局地化降低空间维数, 扩展控制变量增加至 1.7—7.1 倍, 墙钟时间增加至 1.4—4.1 倍。对 100—200 个集合样本, 引入扩展控制变量的计算机墙钟时间增加可控, 文中所发展的引入扩展控制变量的方式具备实用性。

为了减少局地化操作对分析平衡的负面影响, 发展水平局地化方案时, 先将风压准地转平衡关系抽取出来, 然后在非平衡分析变量上进行局地化操作, 最后再将风压准地转平衡关系加回到非平衡分析变量上。试验结果表明, 相比直接在分析变量空间进行水平局地化, 本研究发展的在非平衡分析变量空间进行局地化的方案能更好保持分析场的平衡特征。研究还发现, 引入垂直局地化会对分析场平衡特征产生较大影响, 但依靠 GRAPES 全球 4DVar 目标函数中重力波弱约束控制, 分析场仍能保证动力平衡约束关系。

气候背景误差与集合预报误差所起的作用由其权重系数  $\beta_c$ 、 $\beta_e$  来决定。随着  $\beta_c$  的值逐步增大, 单点试验表明分析增量明显沿着风速较大的方向延

伸,而在垂直风速方向分析增量受到抑制。分析增量呈现出随天气形势变化的特征。将权重系数 $\beta_c$ 、 $\beta_e$ 放在分析增量的计算中,而不是放在目标函数的背景误差与集合预报误差项上,主要是改进海森矩阵的条件数,加快变分同化系统收敛。

致 谢:感谢数值预报中心苏勇、孙健、王金成博士对本研究给予的支持。

## 参考文献

- 龚建东, 张林, 王金成. 2020. 背景误差水平相关结构对四维变分资料同化的影响研究. *气象学报*, 78(6): 988-1001. Gong J D, Zhang L, Wang J C. 2020. The impact study of background error horizontal correlation structure on 4DVar. *Acta Meteor Sinica*, 78(6): 988-1001 (in Chinese)
- 刘艳, 薛纪善. 2019. GRAPES 的新初始化方案. *气象学报*, 77(2): 165-179. Liu Y, Xue J S. 2019. The new initialization scheme of the GRAPES. *Acta Meteor Sinica*, 77(2): 165-179 (in Chinese)
- 王金成, 庄照荣, 韩威等. 2014. GRAPES 全球变分同化背景误差协方差的改进及对分析预报的影响: 背景误差协方差三维结构的估计. *气象学报*, 72(1): 62-78. Wang J C, Zhuang Z R, Han W, et al. 2014. An improvement of background error covariance in the global GRAPES variational data assimilation and its impact on the analysis and prediction: Statistics of the three-dimensional structure of background error covariance. *Acta Meteor Sinica*, 72(1): 62-78 (in Chinese)
- 王瑞春, 龚建东, 张林等. 2015. 热带风压场平衡特征及其对 GRAPES 系统中同化预报的影响研究 II: 动力与统计混合平衡约束方案的应用. *大气科学*, 39(6): 1225-1236. Wang R C, Gong J D, Zhang L, et al. 2015. Tropical balance characteristics between mass and wind fields and their impact on analyses and forecasts in GRAPES system. Part II: Application of linear balance equation-regression hybrid constraint scheme. *Chinese J Atmos Sci*, 39(6): 1225-1236 (in Chinese)
- 熊春晖, 张立凤, 关吉平等. 2013. 集合-变分数据同化方法的发展与应用. *地球科学进展*, 28(6): 648-656. Xiong C H, Zhang L F, Guan J P, et al. 2013. Development and application of ensemble-variational data assimilation methods. *Adv Earth Sci*, 28(6): 648-656 (in Chinese)
- 薛纪善, 庄世宇, 朱国富等. 2008. GRAPES 新一代全球/区域变分同化系统研究. *科学通报*, 53(22): 3446-3457. Xue J S, Zhuang S Y, Zhu G F, et al. 2008. Scientific design and preliminary results of three-dimensional variational data assimilation system of GRAPES. *Chinese Sci Bull*, 53(22): 3446-3457
- Anderson J L. 2001. An ensemble adjustment Kalman Filter for data assimilation. *Mon Wea Rev*, 129(12): 2884-2903
- Bannister R N. 2008. A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics. *Quart J Roy Meteor Soc*, 134(637): 1971-1996
- Bonavita M, Isaksen L, Hölm E. 2012. On the use of EDA background error variances in the ECMWF 4D-Var. *Quart J Roy Meteor Soc*, 138(667): 1540-1559
- Bowler N E, Clayton A M, Jardak M, et al. 2017. The effect of improved ensemble covariances on hybrid variational data assimilation. *Quart J Roy Meteor Soc*, 143(703): 785-797
- Buehner M. 2005. Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Quart J Roy Meteor Soc*, 131(607): 1013-1043
- Buehner M, Houtekamer P L, Charette C, et al. 2010a. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Mon Wea Rev*, 138(5): 1550-1566
- Buehner M, Houtekamer P L, Charette C, et al. 2010b. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. *Mon Wea Rev*, 138(5): 1567-1586
- Buehner M, Morneau J, Charette C. 2013. Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction. *Nonlin Processes Geophys*, 20(5): 669-682
- Chen L L, Chen J, Xue J S, et al. 2015. Development and testing of the GRAPES regional ensemble-3DVar hybrid data assimilation system. *J Meteor Res*, 29(6): 981-996
- Clayton A M, Lorenc A C, Barker D M. 2013. Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quart J Roy Meteor Soc*, 139(675): 1445-1461
- Courtier P, Thépaut J N, Hollingsworth A. 1994. A strategy for operational implementation of 4D-Var: Using an incremental approach. *Quart J Roy Meteor Soc*, 120(519): 1367-1387
- Evensen G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J Geophys Res*, 99(C5): 10143-10162
- Fairbairn D, Pring S R, Lorenc A C, et al. 2014. A comparison of 4DVar with ensemble data assimilation methods. *Quart J Roy Meteor Soc*, 140(678): 281-294
- Houtekamer P L, Mitchell H L. 2005. Ensemble Kalman filtering. *Quart J Roy Meteor Soc*, 131(613): 3269-3289
- Kalnay E. 2003. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge: Cambridge University Press, 175-185
- Keper J D. 2009. Covariance localisation and balance in an Ensemble Kalman Filter. *Quart J Roy Meteor Soc*, 135(642): 1157-1176
- Kleist D T, Ide K. 2015a. An OSSE-Based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part I: System description and 3D-hybrid results. *Mon Wea Rev*, 143(2): 433-451
- Kleist D T, Ide K. 2015b. An OSSE-Based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part II: 4DVar and hybrid variants. *Mon Wea Rev*, 143(2): 452-470
- Lorenc A C. 1986. Analysis methods for numerical weather prediction. *Quart J Roy Meteor Soc*, 112(474): 1177-1194



- Lorenc A C. 2003a. Modelling of error covariances by 4D-var data assimilation. *Quart J Roy Meteor Soc*, 129(595): 3167-3182
- Lorenc A C. 2003b. The potential of the ensemble Kalman filter for NWP: A comparison with 4D-Var. *Quart J Roy Meteor Soc*, 129(595): 3183-3203
- Lorenc A C, Bowler N E, Clayton A M, et al. 2015. Comparison of hybrid-4DVar and Hybrid-4DVar data assimilation methods for global NWP. *Mon Wea Rev*, 143(1): 212-229
- Mitchell H L, Houtekamer P L, Pellerin G. 2002. Ensemble size, balance, and model-error representation in an Ensemble Kalman filter. *Mon Wea Rev*, 130(11): 2791-2808
- Wang X G, Snyder C, Hamill T M. 2007. On the theoretical equivalence of differently proposed ensemble: 3DVAR hybrid analysis schemes. *Mon Wea Rev*, 135(1): 222-227
- Wang X G, Parrish D, Kleist D, et al. 2013. GSI 3DVar-based ensemble-variational hybrid data assimilation for NCEP global forecast system: Single-resolution experiments. *Mon Wea Rev*, 141(11): 4098-4117
- Xu Q, Lu H J, Gao S T, et al. 2008a. Time-expanded sampling for ensemble Kalman filter: Assimilation experiments with simulated radar observations. *Mon Wea Rev*, 136(7): 2651-2667
- Xu Q, Wei L, Lu H J, et al. 2008b. Time-expanded sampling for ensemble-based filters: Assimilation experiments with a shallow-water equation model. *J Geophys Res*, 113(D2): D02114
- Zhang L, Liu Y Z, Liu Y, et al. 2019. The operational global four-dimensional variational data assimilation system at the China Meteorological Administration. *Quart J Roy Meteor Soc*, 145(722): 1882-1896