

陈耀登,郭闪,王元兵,等. 基于优化样本组合的集合-变分混合同化方案研究[J]. 热带气象学报, 2020, 36(4): 464-476.

文章编号: 1004-4965(2020)04-0464-13

# 基于优化样本组合的集合-变分混合同化方案研究

陈耀登<sup>1</sup>, 郭闪<sup>1</sup>, 王元兵<sup>1</sup>, 臧增亮<sup>2</sup>, 潘晓滨<sup>2</sup>

(1. 南京信息工程大学气象灾害教育部重点实验室/气候与环境变化国际合作联合实验室/气象灾害预报预警与评估协同创新中心, 江苏 南京 210044; 2. 国防科技大学气象海洋学院, 江苏 南京 211101)

**摘要:** 为有效引入“流依赖”的背景场误差协方差,同时降低集合预报带来的计算量,尝试通过优选与同化时刻天气形势更相似的历史预报样本,并结合预报过程中的时间滞后样本,将两种样本引入集合-变分混合同化系统中,构建基于优选历史预报样本和时间滞后样本的集合-变分混合同化方案。单点观测理想试验表明,优选历史预报样本结合时间滞后样本,既能够缓解样本不足所导致的采样误差,又能够为同化系统提供“流依赖”的背景场误差协方差。连续一周的循环同化及预报试验结果显示,相较于ERA5资料和探空资料,三维变分方案整体表现稍差,样本组合混合同化方案分析场和预报场的均方根误差最小,且比仅用时间滞后样本的混合同化方案有所改进;降水评分整体也表现最优,尤其对中雨和暴雨的模拟改进较明显,较好地模拟出了强降雨中心的强度和位置,且改善了降水过报的问题。

**关键词:** 数值天气预报;资料同化;混合同化;集合样本;流依赖

**中图分类号:** P435      **文献标志码:** A      **Doi:** 10.16032/j.issn.1004-4965.2020.043

## 1 引言

通过资料同化的方法改善数值模式初始场的精度是提高数值预报准确性的主要手段之一。目前,三维变分方法<sup>[1-5]</sup>是应用较广泛的资料同化方法,但其静态的、各向同性的背景场误差协方差,不符合真实大气“流依赖”的特征。张旭斌等<sup>[6-7]</sup>针对三维变分方法的这一缺陷,将时间依赖信息及多尺度信息引入背景场误差协方差,从而有效提取观测资料的多尺度信息,对降水预报改善明显。陈耀登等<sup>[8]</sup>构造了具有纬度依赖特征的背景场误差协方差,使模式的动力平衡关系更为合理。四维变分方法<sup>[9-12]</sup>虽然可以隐性提供“流依赖”的背景场误差协方差,但其求取切线性模式和伴随模式的过程较为复杂。集合卡尔曼滤波同化方法中

背景场误差协方差不用直接表达,且具有依天气形势而变的流依赖特征<sup>[13-16]</sup>。但是集合卡尔曼滤波法存在着有限的集合样本较难合理估计背景场误差、矩阵不满秩,以及分析场中的变量难以平衡等问题。为结合变分方法和集合方法的优点,同时回避各自的不足,研究人员提出和采用了将变分方案(3DVar或4DVar)与集合方案结合的混合(Hybrid)同化方案<sup>[17-23]</sup>。混合同化方案改善了变分方案背景场误差方差模型化、无法依天气形势而变的问题,也一定程度上缓解了集合方案不满秩、变量不协调问题。

然而,集合-变分混合同化方法中如果采用的集合成员数量过少,依然会存在采样误差、流依赖特征不一定合理等问题,为了获得足够数量的集合样本,集合预报过程同样需要消耗较大的计算量<sup>[17,24]</sup>。为缓解集合-变分混合同化方案中集合预

收稿日期: 2019-12-04; 修订日期: 2020-04-16

基金项目: 国家重点研发计划项目(2017YFC1502102); 国家自然科学基金项目(41675102)共同资助

通讯作者: 陈耀登,男,福建省人,教授,主要从事资料同化和数值预报相关研究。E-mail: keyu@nuist.edu.cn

报所需大量计算资源的问题,杨毅等<sup>[25]</sup>尝试将时间扩展取样的预报样本引入集合卡尔曼滤波同化系统,效果好于样本数较少的控制试验。Wang等<sup>[26]</sup>将时间滞后的预报场应用于集合-变分混合同化系统,通过不同起报时间获得同一时刻的预报场,构建了流依赖的时间滞后预报样本,从而避免了实际集合预报带来的计算问题,然而由于较长预报时效的时间滞后预报样本的预报偏离等问题,时间滞后集合有效样本数量依然较为有限。Chen等<sup>[27]</sup>尝试利用NMC(National Meteorological Center)<sup>[28]</sup>方法,将不同预报时效的历史样本引入集合-变分混合同化系统中,节省计算量的同时获得了大量样本,构建了“各向异性”的背景场误差协方差,然而该背景场误差协方差具有较为明显的气候态特征,并不能很好地表示“依流而变”的

真实大气状态。

为充分利用时间滞后预报样本流依赖的特性,和历史预报样本数量大的优点,本文尝试在集合-变分混合同化中,利用优选的历史预报本来扩展时间滞后预报样本,既能够避免时间滞后预报样本数不足的问题,又能够通过优选历史预报样本引入更多的流依赖样本信息,同时继续保持计算资源消耗较少的优势。

## 2 集合-变分混合同化方法及样本组合

### 2.1 集合变分混合同化方法

对于集合-变分混合同化方法,其代价函数定义如下<sup>[18-19]</sup>:

$$J(\delta x_1, \alpha) = \beta_1 \frac{1}{2} \delta x_1^T B^{-1} \delta x_1 + \beta_2 \frac{1}{2} \alpha^T A^{-1} \alpha + \frac{1}{2} (d - H\delta x)^T R^{-1} (d - H\delta x) \quad (1)$$

在代价函数中,右边第一项表示为静态背景场误差协方差相关项,其中  $B$  为静态的背景场误差协方差矩阵。第二项表示集合误差协方差相关项,  $\alpha$  为集合扩展控制变量,  $A$  为空间协方差矩阵。  $\beta_1$  和  $\beta_2$  分别为前两项的权重系数,且满足关系式  $\frac{1}{\beta_1} + \frac{1}{\beta_2} = 1$ 。第三项为观测项,  $H$  表示线性观测算子,  $R$  表示观测误差协方差。式中,分析增量  $\delta x$  被定义为:

$$\delta x = \delta x_1 + \sum_{k=1}^N (\alpha_k \circ x_k^e) \quad (2)$$

式中,  $\delta x_1$  表示静态的分析增量,  $x_k^e$  表示集合预报误差的无偏估计,  $\circ$  表示舒尔积(Schur product)。

### 2.2 样本组合方法

如前所述,为保持计算资源消耗较少的优势,又能充分利用时间滞后预报样本流依赖的特性和历史预报样本数量大的优点,本文尝试在集合-变分混合同化中,利用优选的历史预报本来扩展时间滞后预报样本。两类样本组合后引入集合-变分混合同化系统的流程如图1所示。时间滞后样本,历史预报样本的构建及组合方法具体如下。

#### (1) 时间滞后样本。

时间滞后样本由预报过程中不同起报时间、同一时刻的预报场构成,因此时间滞后样本所生成的集合预报误差可被定义为<sup>[23]</sup>:

$$x_{T,k}^e = \frac{(x_j - x_i)}{\sqrt{N-1}}, 0 < i < j \leq N \quad (3)$$

式中,  $x_i, x_j$  表示不同时刻初始场经过不同预报时长积分后得到的同一时刻的预报场,  $N$  表示时间滞后样本数,第  $k$  个差异场生成的预报误差可表示为  $x_{T,k}^e$ 。以间隔3 h同化并向前预报24 h的时间滞后方案为例,若每3 h保存一次预报场结果,则共有8个时间滞后样本。由8个时间滞后样本两两求差异,可得到  $x_8 - x_7, x_8 - x_6, \dots, x_8 - x_1, \dots, x_2 - x_1$ ,共计28个差异场,再依据公式(3)可得到28个时间滞后样本估计的集合预报误差。

#### (2) 历史预报样本。

历史预报样本由过去一段时间(如一个月)模拟的历史预报数据,每6 h保存的12 h和24 h预报场构建,并将同一时刻24 h预报场与12 h预报场的差异,作为历史预报误差的计算样本。历史预报样本所生成的集合预报误差可由如下公式定义<sup>[27]</sup>:

$$x_{H,l}^e = \frac{x_{H,l}^{diff} - \overline{x_H}}{\sqrt{M-1}}, 0 < l \leq M \quad (4)$$

$$x_{H,l}^{diff} = x_{H,l}^{T1} - x_{H,l}^{T2} \quad (5)$$

其中,  $x_{H,l}^e$  表示由优选历史预报样本估计的集合预报误差,  $M$  表示优选历史预报样本个数,  $x_{H,l}^{diff}$  表示预报时效分别为  $T_1, T_2$  的历史预报差异场,  $\overline{x_H}$  表示

该差异场的平均。

历史预报样本数量较多,且包含气候态信息,若直接全部引入,则与大气的“流依赖性”不符。因此,在同化阶段之前,本研究需优选出与预报时刻天气形势更相符的历史预报样本。具体做法是:首先,对于同化前的背景场  $\mathbf{x}^b$ ,与历史预报样本  $\mathbf{x}_m$  ( $0 < m \leq M$ ),其中  $M$  表示历史预报样本总数,提取两者中低层(500 hPa、700 hPa及850 hPa)的纬向风场  $U$ 、经向风场  $V$  及温度场  $T$ ,一一计算各个变量皮尔森相关系数,并以其加权平均  $Cor_m = \overline{|Cor_{U,m}| + |Cor_{V,m}| + |Cor_{T,m}|}$  ( $0 < m \leq M$ ),视为两者的相关。其次,将历史预报样本按相关系数从高到低进行排序,优选出相关系数更高的历

史预报样本进行后续计算。

(3) 样本组合。

综合两类样本的分析增量可由以下公式表示:

$$\delta \mathbf{x} = \delta \mathbf{x}_1 + \sum_{k=1}^N (\alpha_k \circ \mathbf{x}_{T,k}^e) + \sum_{l=1}^M (\alpha_l \circ \mathbf{x}_{H,l}^e) \quad (6)$$

其中,右侧第一项为静态的分析增量,第二项中的  $\mathbf{x}_{T,k}^e$  表示时间滞后样本估计的集合预报误差,第三项中的  $\mathbf{x}_{H,l}^e$  表示由优选历史预报样本估计的集合预报误差。 $N$  表示时间滞后样本数, $M$  表示历史预报样本数。式(6)为式(2)基于扩展控制变量方法<sup>[29]</sup>的多样本拓展形式,参考 Wang 等<sup>[30]</sup>中的式(6-7),所对应的代价函数为:

$$J(\delta \mathbf{x}_1, \alpha) = \beta_1 \frac{1}{2} \delta \mathbf{x}_1^T \mathbf{B}^{-1} \delta \mathbf{x}_1 + \beta_2 \frac{1}{2} \langle \alpha_l, \alpha_k \rangle^T \mathbf{A}^{-1} \langle \alpha_l, \alpha_k \rangle + \frac{1}{2} (\mathbf{d} - \mathbf{H} \delta \mathbf{x})^T \mathbf{R}^{-1} (\mathbf{d} - \mathbf{H} \delta \mathbf{x}) \quad (7)$$

其中  $\alpha_l$  和  $\alpha_k$  分别表示历史预报样本和时间滞后样本对应的集合扩展控制变量的向量。

水平格点数为  $271 \times 271$ ,垂直分41层,模式顶气压50 hPa。试验的初边界条件来自  $0.5^\circ \times 0.5^\circ$  的 GFS(Global Forecast System)资料,模式参数采用:WSM-5 微物理方案<sup>[31]</sup>、YSU 边界层方案<sup>[32]</sup>,Kaine-Fritsch 积云参数化方案<sup>[33]</sup>、Dudhia 短波辐射方案<sup>[34]</sup> 以及 RRTM (Rapid Radiative Transfer Model) 长波辐射方案<sup>[35]</sup>。观测资料来自 GTS (Global Telecommunications System) 观测数据集(图2)。

### 3 模式参数设置及试验方案设计

研究所用的模式为 WRF (Weather Research and Forecasting Model) V 3.9.1 以及 WRFDA (Weather Research and Forecasting Model Data Assimilation System) V 3.9.1。区域分辨率 9 km,

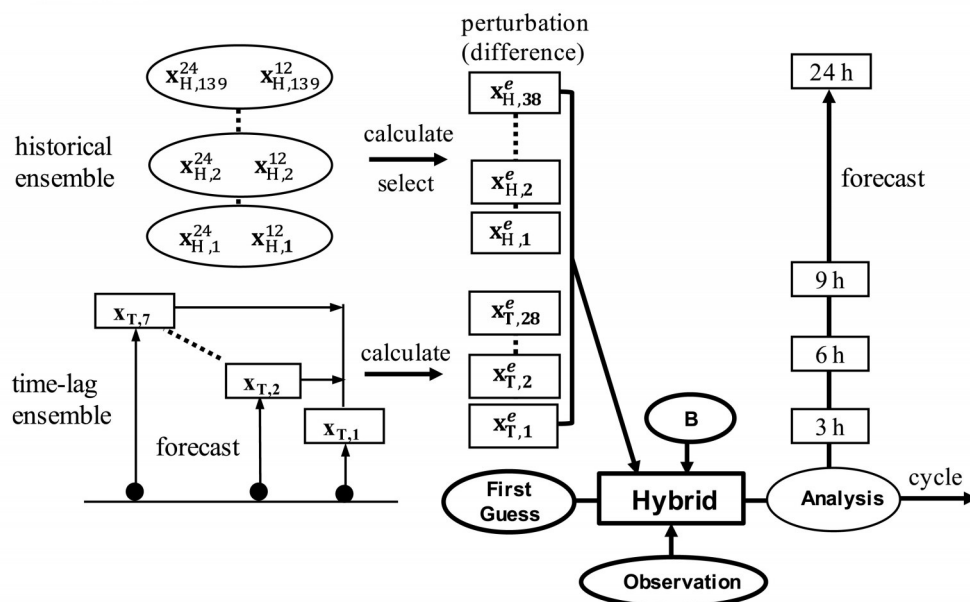


图1 优选历史预报样本与时间滞后样本组合的集合-变分混合同化方案流程图

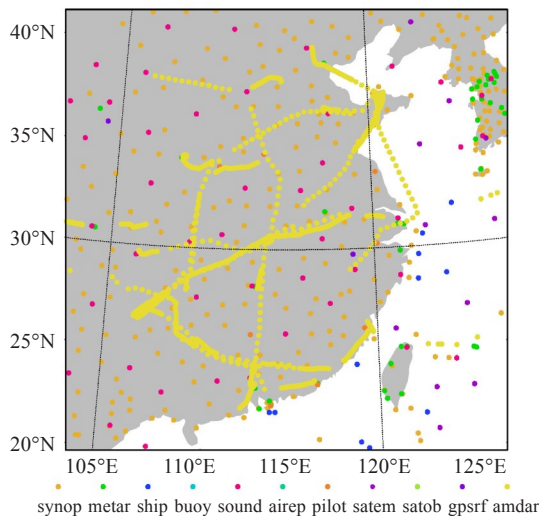


图2 试验区域及GTS观测数据分布  
 观测时间为2016年7月1日00时(世界时),synop为地面站观测,metar为自动站的航空观测,ship为船舶观测,buoy为浮标观测,sound为无线电探空观测,airep为飞机报观测,satem、satob为卫星探测资料,gpsrf为GPS折射率观测,amdar为自动飞机报观测。

研究中静态的背景场误差协方差由NMC(National Meteorological Center)<sup>[28]</sup>方法估计得到。统计样本为2015年7月1日00时—31日12时(世界时,下同)连续一个月每隔12 h WRF模拟的12 h预报场与24 h预报场。同化系统所用的控制变量为CV5方案<sup>[6]</sup>,包括流函数 $\psi$ 、非平衡速度势 $\chi_u$ 、非平衡温度 $T_u$ 、假相对湿度 $RH$ 、及非平衡地表气压

$$P_{s,u \circ}$$

根据同化试验中不同样本来源及样本估计的集合预报误差数,设计了四组对比试验(表1),时间段为2016年7月1日00时—7日00时,每隔3 h同化一次,整个模拟过程共循环同化49次。分别是基于三维变分同化方法的控制试验(3DVar);使用7个时间滞后样本得到28个时间滞后预报误差的混合同化试验(TL28);使用12个时间滞后样本,得到66个时间滞后预报误差的混合同化试验(TL66);以及结合了时间滞后样本与优选历史预报样本的混合同化试验(TH66)。在循环同化过程中,TH66试验在每次同化前均从所有历史预报样本中优选与当前天气形势相近的历史预报样本,引入同化系统。对比样本较为有限的TL66与TL28试验,TH66试验扩展了样本个数,且几乎不需要增加额外的计算资源。试验期间,3DVar、TL28及TL66预报时长为24 h,而在同化间隔均为3 h前提下,为获得更多的时间滞后样本,TL66的预报时长为36 h。

所有历史预报样本来自2015年6月7日00时—7月12日00时的预报数据,每隔6 h得到一组预报时效分别为12 h与24 h的同一时刻预报场,共进行了为期35天的样本构建,生成139组历史预报样本及对应的历史预报误差。所用的初边界条件均为GFS资料,不同化观测资料,且其余参数设置与试验模拟过程相同。

表1 试验名称、样本来源、个数及相应预报误差个数

试验名	样本来源及个数	时间滞后预报误差个数	优选历史预报误差个数	最长预报时效/h
3DVar	/	/	/	24
TL28	时间滞后样本(7)	28	/	24
TL66	时间滞后样本(12)	66	/	36
TH66	时间滞后样本(7)与优选历史预报样本(38)	28	38	24

历史预报样本与同化时刻背景场的相关系数随时间的变化如图3所示,黑色点线表示最低相关系数的变化,黑色细竖线表示各时刻38个优选样本的相关系数数值范围,蓝色点线表示各时刻引入的38个优选样本的平均相关系数。从图3可以看出,优选样本的平均相关系数(蓝色点线)在7月1—4日的平均相关系数较为稳定,始终保持在0.7左右,7月4—7日平均相关系数有明显下降,最小值约为0.5,而最低相关系数数值(黑色实线)则在

0.4左右波动。造成该现象的主要原因是本研究中可供优选的历史预报样本总数较为有限,当天气形势发生明显变化时难以始终选出高相关系数的历史预报样本。另外,由于不同时段构建的历史预报样本所包含的天气形势也不相同,这种差异会对同化结果产生一定影响,实际业务中若构建足够庞大的历史预报样本数据库,理论上可持续引入相关系数更高的优选样本。

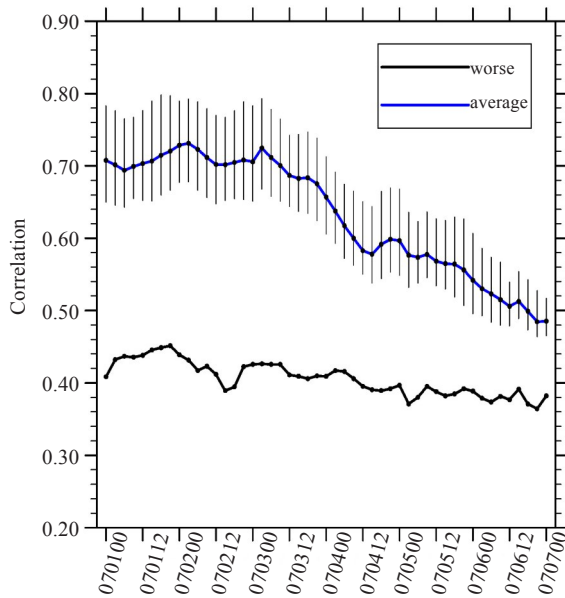


图3 优选历史预报样本相关系数随时间的变化  
每隔3 h计算一次相关系数。

其中蓝线表示优选38个样本平均相关系数,黑色细竖线表示  
优选38个样本相关系数范围,黑色点线表示历史预报  
样本中相关系数最低值。

## 4 单点观测理想试验

为探究不同预报样本对于同化系统的影响,本研究首先进行单点观测理想试验的分析。在模式层第9层(约800 hPa)的112°E,31°N处放置一个新息向量为1.0 m/s的纬向风 $U$ 理想观测,同化时间为2016年7月2日00时,集合误差协方差权重设置为0.75,局地化尺度为200 km。风场的分析增量如图4所示,可以看出,3DVAR单点试验的风场增量为各向同性(图4a),而基于集合-变分混合同化方法的风场分析增量均表现为各向异性的特征:由于TL28试验(图4b)的时间滞后样本过少,带来了较大的采样误差,分析增量分布较为杂乱;在TL66试验中当时间滞后样本数增加后,采样误差的问题被一定程度改善(图4c);TH66试验,通过时间滞后样本结合优选历史预报样本,分析增量表现出了与当前天气形势较好的关联性(图4d)。

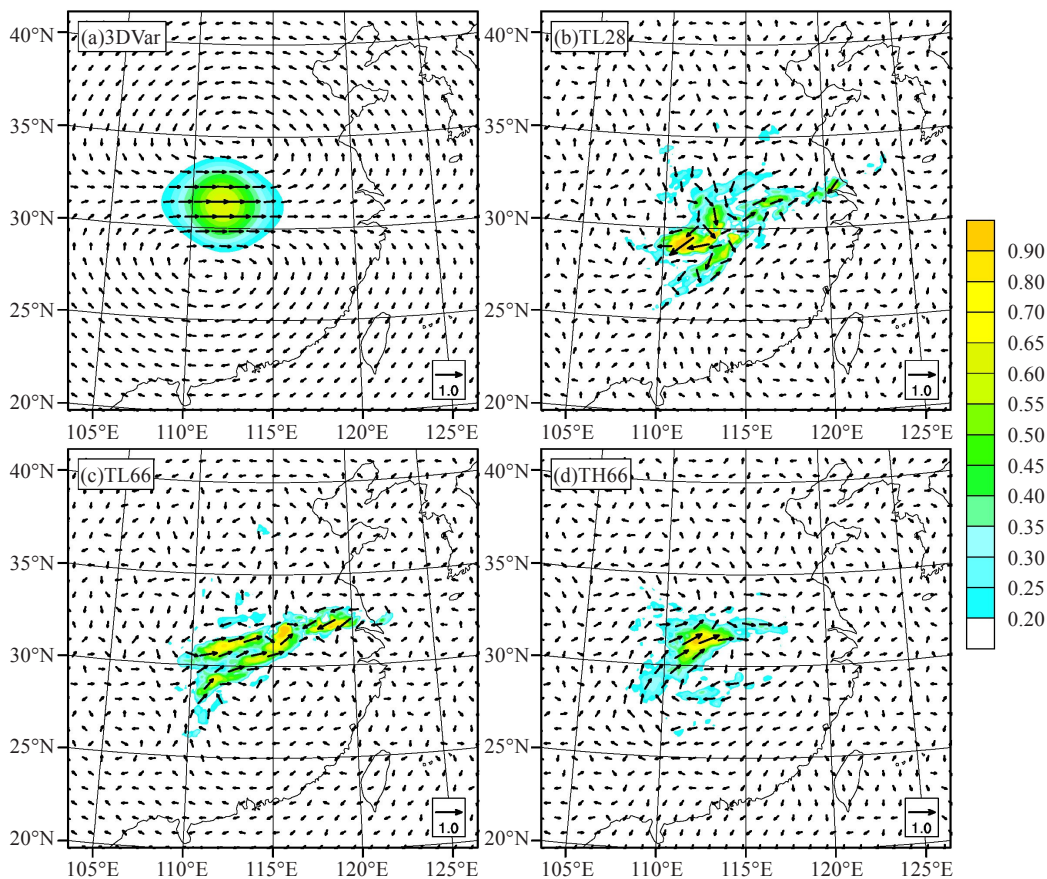


图4 2017年7月2日00时不同试验纬向风 $U$ 单点试验中风场的分析增量(矢量、阴影,单位:m/s)

a. 3DVAR;b. TL28;c. TL66;d. TH66。

在模式第16层(约500 hPa)的118 °E,31 °N位置处放置一个新息向量为1 K的温度  $T$  理想观测,时间为2017年7月2日00时,集合权重及局地化尺度的设置与纬向风  $U$  的单点理想观测试验一致。分析水汽场的增量特征可知,3DVAR中CV5

方案的背景场误差协方差不存在温度  $T$  与水汽  $q$  之间的变量相关(图5a),混合同化试验存在  $T$ 、 $q$  之间的变量相关性,且水汽增量的分布形态与背景场湿度的等值线较为一致(图5b~5d),均能够引入“流依赖”的特征。

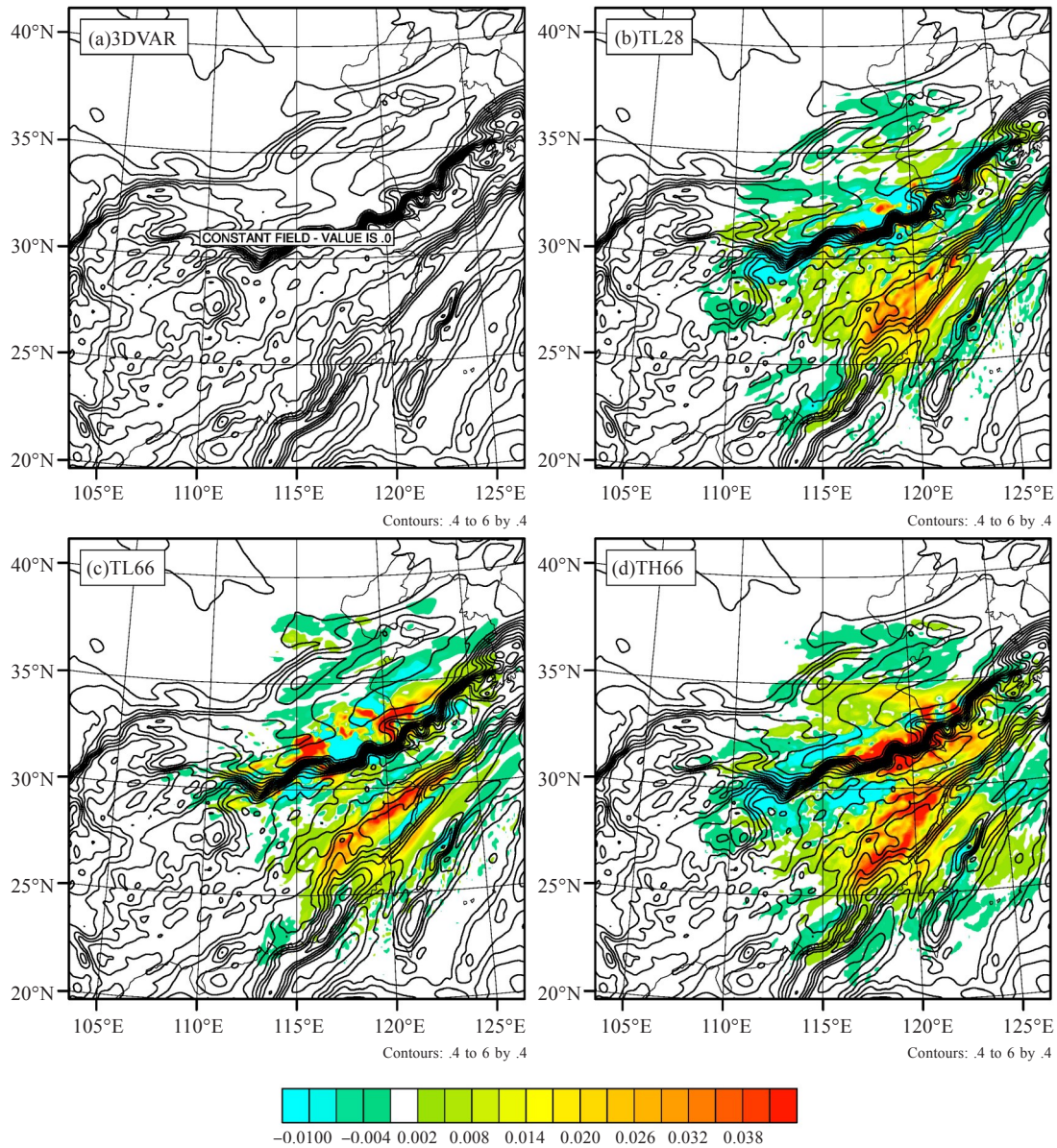


图5 不同试验温度  $T$  单点试验中水汽场的分析增量(阴影,单位:g/kg) 等值线表示2016年7月2日00时模式第16层的背景湿度场。a. 3DVAR;b. TL28;c. TL66;d. TH66。

进一步进行了TH66方案不同时刻的温度  $T$  的单点理想观测试验(7月2—5日每隔24 h进行一次),集合误差协方差权重为0.75,局地化尺度为200 km。从图6可以看出,TH66方案的水汽分析增量随时间不断变化,且分析增量能够随着背

景湿度场发生改变(图6a~6d),其“依流而变”的特征来源于不断更新的时间滞后预报样本及与当前天气形势更相近的历史预报样本,说明本方案能够仅依靠少量计算资源,达到引入合理的“流依赖”的背景场误差协方差的目的。

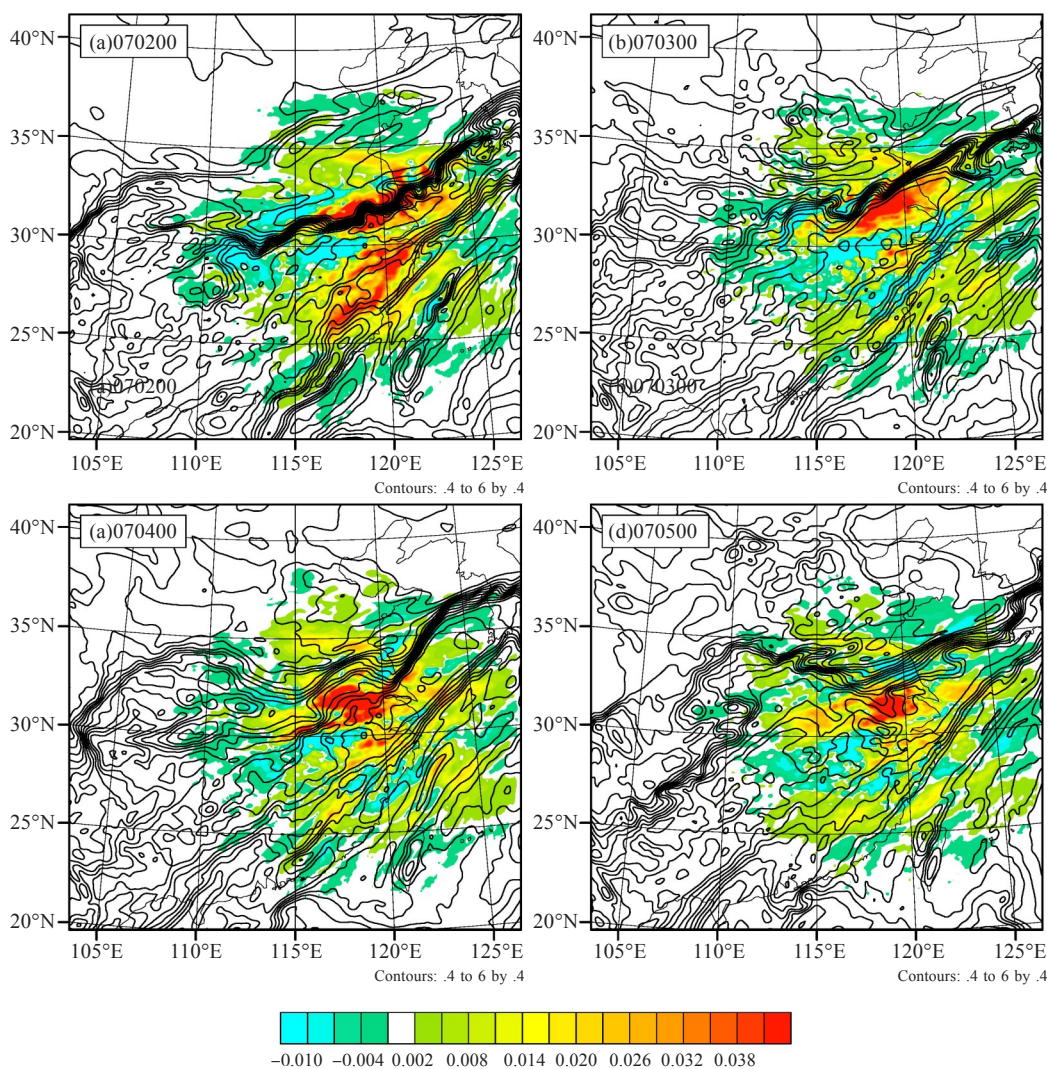


图6 不同时刻温度 $T$ 单点试验中水汽场的分析增量(阴影,单位: $\text{g/kg}$ ) 等值线表示模式第16层的背景湿度场。  
a.7月2日00时;b.7月3日00时;c.7月4日00时;d.7月5日00时。

## 5 梅雨过程循环同化及预报试验

### 5.1 梅雨过程介绍

2016年长江区梅雨期从6月19日入梅,7月21日出梅,雨量为584.3 mm,较常年平均偏多107.9%,位于1951年以来第3位。7月1—6日的总降水量达到270.4 mm,已接近该区整个梅雨期的平均降水量(280.4 mm),其中前三天的降水量分别为60.4 mm、64.3 mm、52.5 mm<sup>[37]</sup>。

本研究的模拟时间段为2016年7月1日00时—7日00时,每隔3 h同化一次,整个模拟过程共

循环同化49次。

### 5.2 均方根误差

图7为四组试验连续一周的分析场及预报场中各变量相对于ERA5(5<sup>th</sup> generation climate reanalysis dataset from ECMWF)的平均均方根误差。从分析场看(图7实线),蓝色实线所表示的TH66的各变量分析结果,明显优于仅依赖于时间滞后样本的混合同化试验TL28、TL66;从12 h预报场看(图7虚线),蓝色虚线所表示的TH66试验各变量仍然能够保持较小的均方根误差,而仅采用时间滞后样本的TL28、TL66相较于3DVAR的改进稍小。

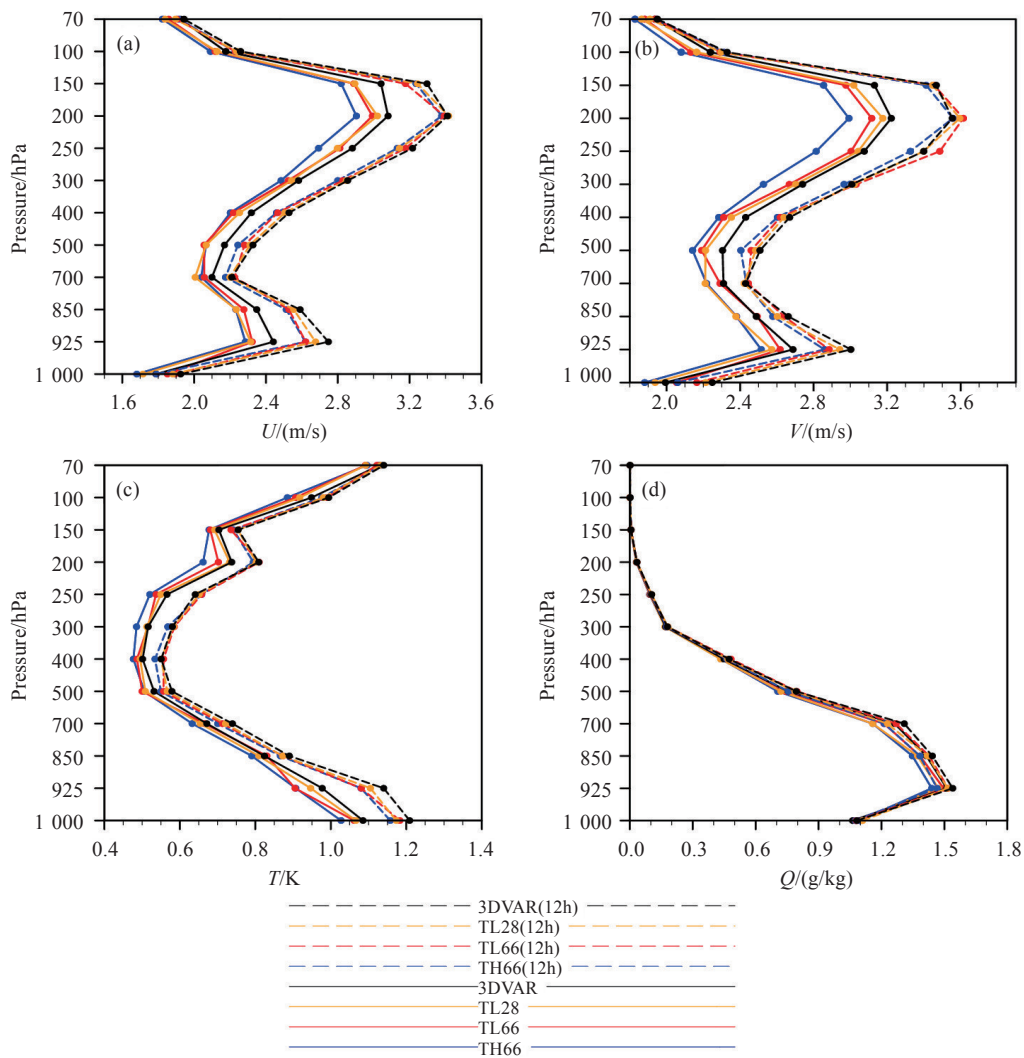


图7 四组试验连续一周的分析场及预报场中各变量相对于ERA5的平均均方根误差  
实线为分析场,虚线为12 h预报场。

四组试验分析场与12 h预报场较探空观测资料平均一周的均方根误差如图8所示。对于U、V变量,蓝线所表示的样本组合的混合同化试验TH66整体表现最优,其分析场和12 h的预报场均与实况探空观测更接近。时间滞后样本数较多的TL66表现略好于TL28,3DVAR与TL28试验整体表现稍差。从水汽q变量的分析场来看(图8d),混合同化试验没有显示出优势,在12 h预报场中,基于样本组合的TH66试验(图8d蓝色虚线)表现略好于其它三组试验。

### 5.3 降水循环同化预报FSS评分

为对比不同试验对降水预报的效果,本研究对四组试验(3DVAR、TL28、TL66、TH66),按不同阈值进行了逐6 h的平均累计降水FSS(Fraction

Skill Score)评分(图9),降水观测为中国地面与CMORPH融合逐小时降水产品(CHMPA)<sup>[38]</sup>。在小雨量级中,3DVAR表现稍差,TL28与TH66表现稍优(图9a)。在中雨量级中,TH66在整个预报时效内均表现最优,尤其在预报时效较长的24 h,TH66能够保持较好的表现,展现出较好的鲁棒性(图9b)。在大雨量级中,预报初期TH66的评分表现稍低于TL28与TL66试验,但在24 h评分与TL66试验接近(图9c)。在暴雨量级中,TH66的评分表现明显优于其它方案(图9d)。综合来看,3DVAR在各降水阈值内的降水评分表现最差,由优选历史预报样本与时间滞后样本结合的混合同化方案TH66表现最佳。



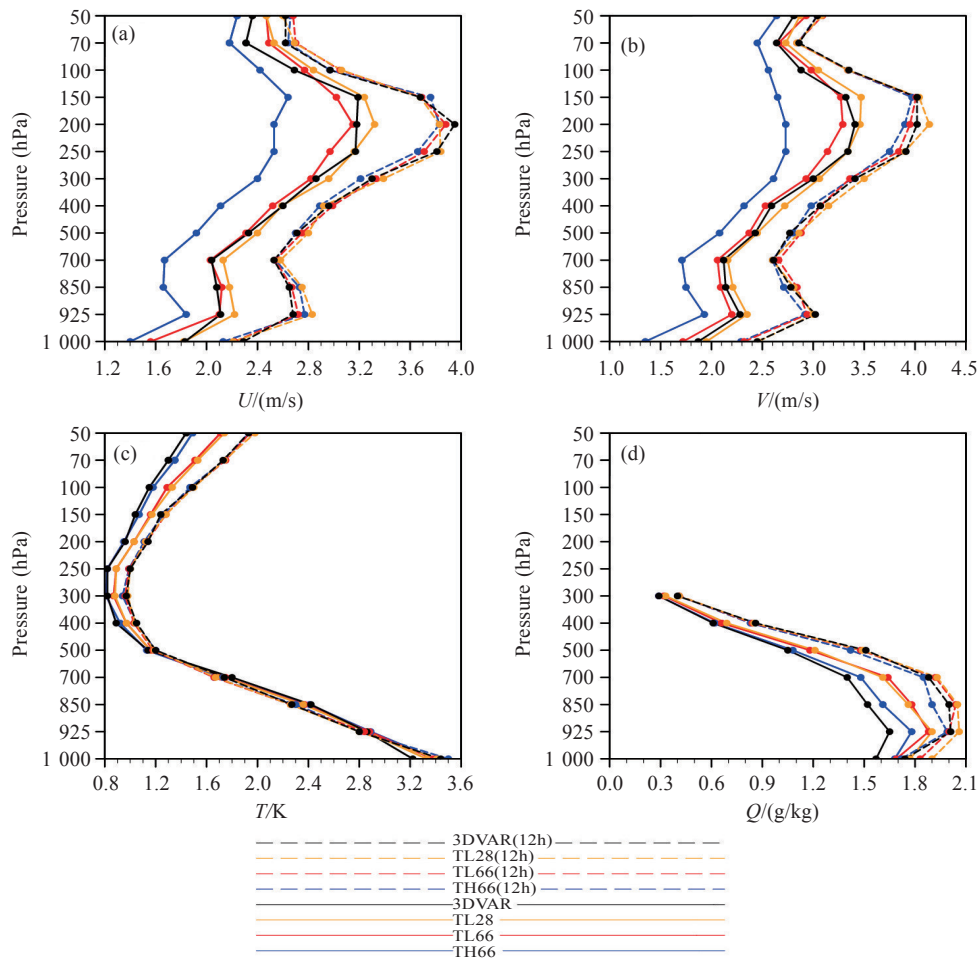


图8 四组试验连续一周的分析场及预报场中各变量相对于探空观测资料的平均均方根误差  
实线为分析场,虚线为12 h预报场。

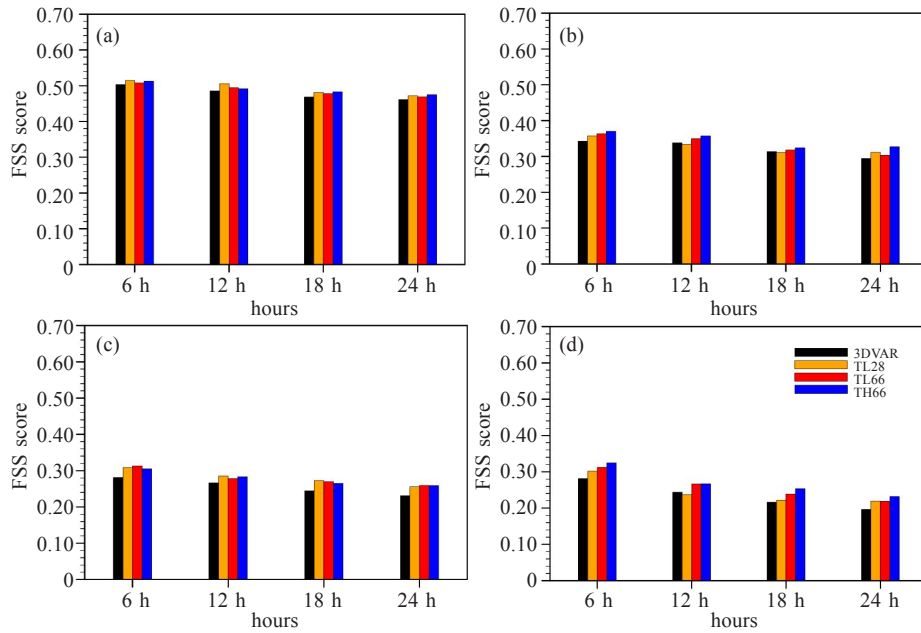


图9 2017年7月1日00时—7日00时四组试验不同降水阈值连续一周的逐6小时平均累计降水评分(FSS)  
a. 小雨(阈值为0.1~3.9 mm);b. 中雨(阈值为4.0~12.9 mm);c. 大雨(阈值为13.0~24.9 mm);d. 暴雨(25.0~59.9 mm)。

#### 5.4 降水分布及诊断

为更好地理解各方案对强降水过程的模拟,本研究挑选了在此次梅雨过程中降水强度较大时间段(2016年7月1日00时—2日00时)的24 h累积降水进行分析(图10)。如图所示观测降水呈西南-东北向,累计24 h的降水量超过150 mm的区域主要集中在安徽中部地区(图10a)。3DVAR试验模拟出了整体的雨带走势,但对于强降水中心的模拟表现较差(图10b)。相较于3DVAR, TL28

对于降水强度较大区域的模拟稍有改善,但在江苏南部地区存在过报的问题(图10c)。增加时间滞后样本数的TL66试验对于安徽中部强降水的模拟有一定改进,但其强降水中心略微偏北,且在江苏南部仍然存在过报的问题(图10d)。由优选历史预报样本组合时间滞后样本的混合同化方案TH66对于强降水中心的模拟在位置和强度上均最接近实况,并且改善了江苏南部降水过报的问题(图10e)。

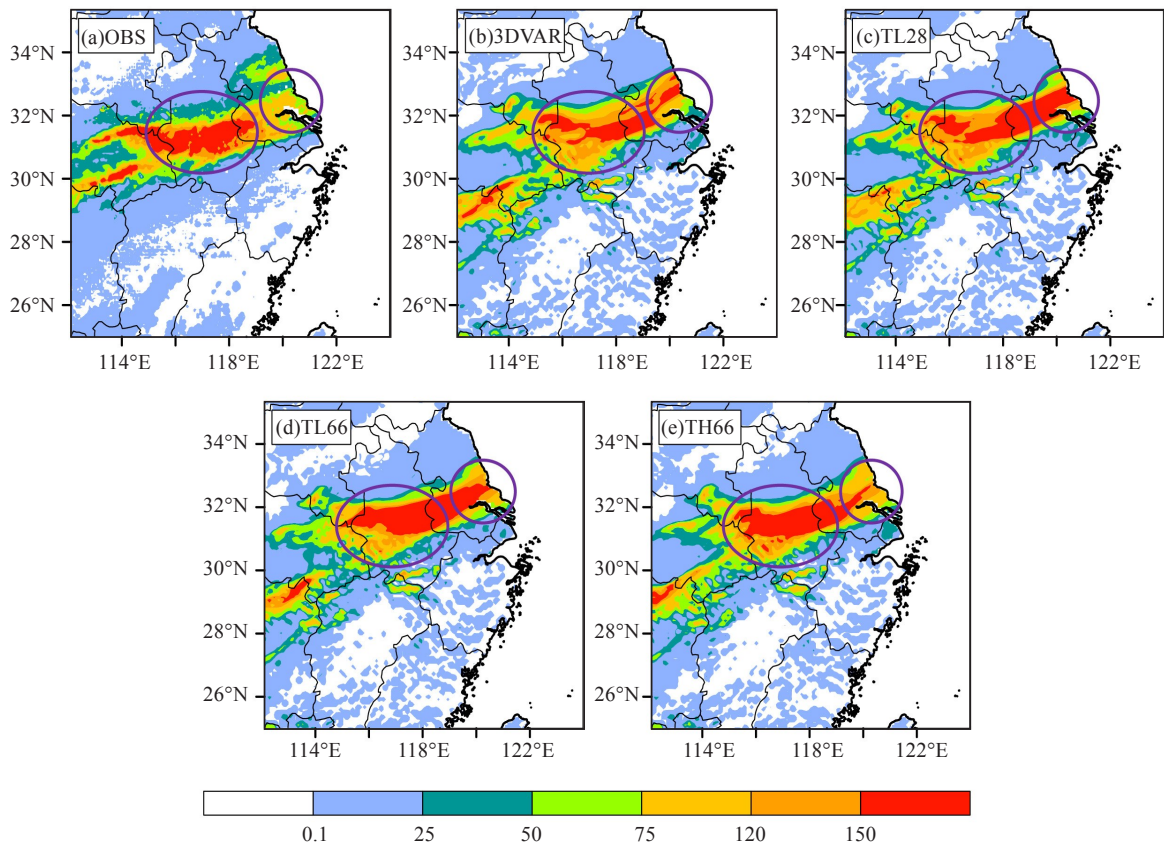


图10 2016年7月1日00时—2日00时累计24小时降水分布(单位:mm) a. 实况降水;b. 3DVAR;  
c. 28个样本误差的时间滞后混合同化方案TL28;d. 66个样本误差的时间滞后混合同化方案TL66;  
e. 时间滞后与优选预报样本组合的混合同化方案TH66。

为进一步探究样本组合的混合同化方案较三维变分方案的降水模拟改进的原因,两组试验在2016年7月1日12时(图11a、11b)和18时(图11c、11d)850 hPa的水汽通量和水汽通量散度如图11所示。参照累计降水分布图可知(图10),两组试验在强降水中心附近均能模拟出较强的水汽通量以及水汽通量散度。在12时的850 hPa层上,3DVAR试验的水汽通量的切变线更弱且稍偏北

(图11a红线),而TH66试验具有更为明显的水汽通量切变(图11b红线)。对比两组试验在18时的水汽通量散度分布可以看出(图11c、11d),3DVAR试验的水汽辐合区明显向东偏移(图11c红框),从而导致3DVAR试验没有较好地模拟出强降水中心,而TH66试验的水汽辐合区仍能够较好地分布在降水中心(图11d红框),这与实况降水分布特征更为接近。

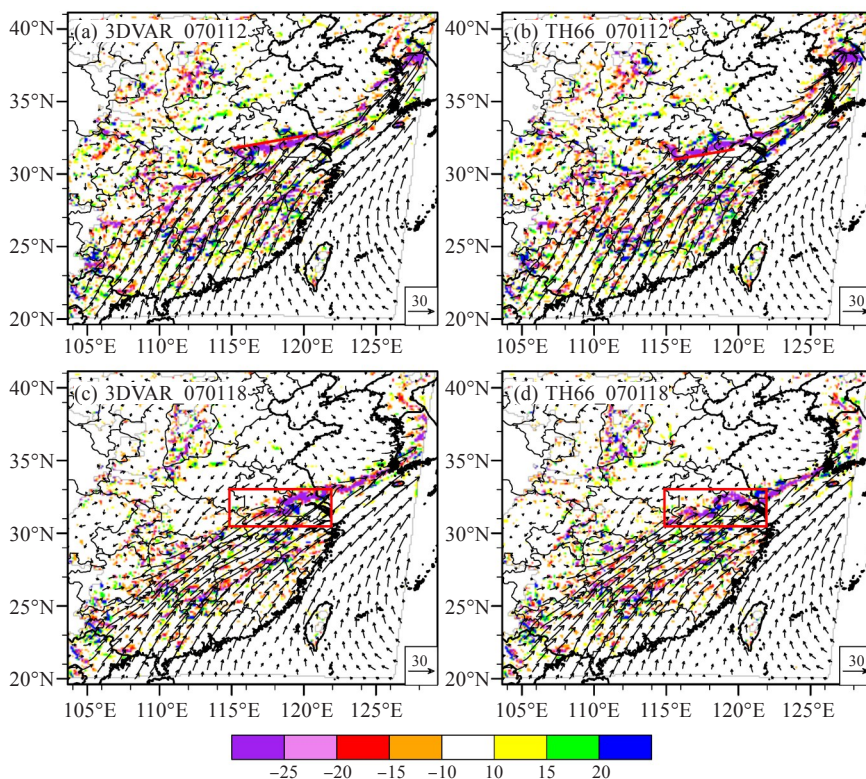


图11 2016年7月1日12时和18时在850 hPa上的水汽通量散度(填色,单位: $\text{g}/(\text{cm}^2 \cdot \text{hPa} \cdot \text{s})$ )和水汽通量(箭矢,单位: $\text{g}/(\text{cm} \cdot \text{hPa} \cdot \text{s})$ ) a. 3DVAR\_070112; b. TH66\_070112; c. 3DVAR\_070118; d. TH66\_070118。

## 6 总 结

集合-变分混合同化方法是资料同化领域的研究热点之一。一方面,该方法能为同化系统提供“流依赖”的背景场误差协方差,另一方面,该同化方法中集合预报需消耗较多的计算资源。本文探究讨论了时间滞后样本的构建方案与历史预报样本的优选方案,通过计算风场与温度场的相关系数,引入与同化时刻更为相近的历史预报样本。研究进行了一系列单点观测理想试验和针对2016年梅雨期强降水过程(2016年7月1日00时—7日00时)的连续循环同化及预报试验。

单点观测理想试验的结果表明,时间滞后样本能够为集合-变分混合同化系统引入“流依赖”的背景场误差协方差;通过优选历史预报样本扩展有限数量的时间滞后样本后,风场增量的表现更为合理,缓解了时间滞后样本量不足所带来的问题。单点试验也表明,不断更新的时间滞后预报样本及优选得到的与当前天气形势更相近的历史预报样本,能够为同化带来“依流而变”的分析

增量,说明本方案能够仅依靠少量计算资源,便达到引入合理的“流依赖”的背景场误差协方差的目的。

对于连续循环同化及预报试验,整体而言,采用时间滞后的集合-变分混合同化方案分析场与预报场均方根误差比三维变分更小;通过优选历史预报样本扩展的时间滞后样本后,其分析场和预报场的均方根误差相对于单纯采用时间滞后样本进一步减小。从降水预报效果上看,优选历史预报样本组合时间滞后样本的混合同化方案效果最好,仅使用时间滞后样本的混合同化方案效果次之,三维变分表现最差。通过分析降水个例过程中的水汽通量条件可知,样本组合方案较三维变分方案能够更好地模拟出水汽通量散度分布,使水汽辐合区与实况强降水区域位置更为接近,从而改善降水预报的表现。

本研究进行了将时间滞后样本与历史预报样本共同引入集合-变分混合同化系统的尝试,并取得了较好的研究结果,但对于该方法的集合协方差权重系数及局地化尺度的参数敏感性试验尚未进行,仍需进一步展开研究。此外,本研究主要构

建了与梅雨过程相关的历史预报样本,并在此基础上进行了针对梅雨期强降水的研究,因此下一步拟构建包含更多天气过程、更长时间序列的历史预报样本,以改进不同类型天气过程的预报效果。

## 参 考 文 献

- [1] WU W S, PURSER R J, PARRISH D F. Three-Dimensional variational analysis with spatially inhomogeneous covariances[J]. *Mon Wea Rev*, 2002, 130(12): 2 905-2 916.
- [2] BARKER D M, HUANG W, GUO Y R, et al. A three-dimensional variational data assimilation system for MM5: Implementation and initial results[J]. *Mon Wea Rev*, 2004, 132(4): 897-914.
- [3] 张华, 薛纪善, 庄世宇, 等. GRAPeS 三维变分同化系统的理想试验[J]. *气象学报*, 2004, 62(1): 31-41.
- [4] 丁伟钰, 万齐林, 闫敬华, 等. 华南地区中尺度模式预报的初值影响分析[J]. *热带气象学报*, 2006, 22(1): 10-17.
- [5] 曹小群, 黄思训, 张卫民, 等. 区域三维变分同化中背景误差协方差的模拟[J]. *气象科学*, 2008, 28(1): 8-14.
- [6] 张旭斌, 薛纪善, 万齐林, 等. 时间依赖的多尺度背景误差协方差研究 I——构造[J]. *热带气象学报*, 2014, 30(4): 675-686.
- [7] 张旭斌, 薛纪善, 万齐林, 等. 时间依赖的多尺度背景误差协方差研究 II——应用[J]. *热带气象学报*, 2015, 31(2): 161-172.
- [8] 陈耀登, 曾腊梅, HUANG Xiang-yu, 等. 纬度依赖背景场误差协方差及其对台风同化和预报的影响研究[J]. *热带气象学报*, 2014, 30(4): 654-662.
- [9] LE DIMET, F X, TALAGRAND O. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects [J]. *Tellus A*, 1986, 38(2): 97-110.
- [10] 龚建东, 邱崇践, 王强, 等. 区域四维变分资料同化的数值试验[J]. *气象学报*, 1999, 57(2): 4-15.
- [11] HUANG X Y, XIAO Q, BARKER D M, et al. Four-Dimensional variational data assimilation for WRF: Formulation and preliminary results [J]. *Mon Wea Rev*, 2009, 137(1): 299-314.
- [12] 龚建东, 刘永柱, 张林. 面向四维变分资料同化的 NSAS 积云深对流参数化方案的简化及线性化研究[J]. *气象学报*, 2019, 77(4): 595-616.
- [13] 刘成思, 薛纪善. 关于集合 Kalman 滤波的理论和方法的发展[J]. *热带气象学报*, 2005, 21(6): 628-633.
- [14] 李霁杭, 高郁东, 万齐林. 基于观测路径的集合预报样本优选对热带气旋的模拟研究[J]. *热带气象学报*, 2019, 35(2): 197-209.
- [15] 李霁杭, 肖辉, 万齐林, 等. 基于 2017 年 5 月 7 日广州特大暴雨分析影响半径对集合卡尔曼滤波方法同化效果的影响[J]. *热带气象学报*, 2019, 35(1): 73-88.
- [16] HOUTEKAMER P L, ZHANG F Q. Review of the ensemble Kalman filter for atmospheric data assimilation[J]. *Mon Wea Rev*, 2016, 144(12): 4 489-4 532.
- [17] HAMILL T M, SNYDER C. A hybrid ensemble kalman filter-3d variational analysis scheme[J]. *Mon Wea Rev*, 2000, 128(8): 2 905-2 919.
- [18] WANG X, BARKER D M, SNYDER C, et al. A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment[J]. *Mon Wea Rev*, 2008, 136(12): 5 116-5 131.
- [19] WANG X, BARKER D M, SNYDER C, HAMILL T M. A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part II: real observation experiments[J]. *Mon Wea Rev*, 2008, 136(12): 5 132-5 147.
- [20] 邱晓滨, 邱崇践. 混合误差协方差用于集合平方根滤波同化的试验[J]. *高原气象*, 2009, 28(6): 1 399-1 407.
- [21] SCHWARTZ C S, LIU Z. Convection-Permitting forecasts initialized with continuously cycling limited-Area 3DVAR, Ensemble Kalman Filter, and “hybrid” variational-ensemble data assimilation systems[J]. *Mon Wea Rev*, 2014, 142(2): 716-738.
- [22] BANNISTER R N. A review of operational methods of variational and ensemble-variational data assimilation: Ensemble-variational data assimilation[J]. *Q J R Meteor Soc*, 2017, 143(703): 607-633.
- [23] WANG Y, MIN J, CHEN Y. Impact of the hybrid gain ensemble data assimilation on meso-scale numerical weather prediction over east China[J]. *Atmos Res*, 2018, 206: 30-45.
- [24] KRETSCHMER M, HUNT B R, OTT E. Data assimilation using a climatologically augmented local ensemble transform Kalman filter[J]. *Tellus A*, 2015, 67(1): 26 617.
- [25] 杨毅, 弓中强, 王金艳, 等. 时间扩展取样集合卡尔曼滤波同化模拟探空试验研究[J]. *气象学报*, 2012, 70(1): 101-108.
- [26] WANG Y, MIN J, CHEN Y, et al. Improving precipitation forecast with hybrid 3DVar and time-lagged ensembles in a heavy rainfall event [J]. *Atmos Res*, 2017, 183(1): 1-16.
- [27] CHEN Y, WANG J, GAO Y, et al. Refinement of the use of inhomogeneous background error covariance estimated from historical forecast error samples and its impact on short-Term regional numerical weather prediction[J]. *J Meteor Soc Jpn Ser II*, 2018, 96(5): 429-446.
- [28] PARRISH, DAVID F, JOHN C DERBER. The National Meteorological Center's spectral statistical-interpolation analysis system[J]. *Mon*

- Wea Rev, 1992, 120(8): 1 747-1 763.
- [29] LORENC A C. The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var[J]. Q J R Meteor Soc, 2003, 129(595): 3 183-3 203.
- [30] WANG Y, CHEN Y, MIN J, et al. Efficient regional hybrid ensemble-variational data assimilation using the Global-Ensemble-Model-Augmented error covariance for numerical weather prediction over Eastern China[J]. Atmosphere, 2020, 11(4): 365.
- [31] HONG S Y, DUDHIA J, CHEN S H. A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation[J]. Mon Wea Rev, 2004, 132(1): 103-120.
- [32] HONG S Y, NOH Y, DUDHIA J. A new vertical diffusion package with an explicit treatment of entrainment processes[J]. Mon Wea Rev, 2006, 134(9): 2 318-2 341.
- [33] KAIN J S, FRITSCH J M. A one-dimensional entraining/detraining plume model and its application in convective parameterization[J]. J Atmos Sci, 1990, 47(23): 2 784-2 802.
- [34] DUDHIA J. Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale model[J]. J Atmos Sci, 1989, 46(20): 3 077-3 107.
- [35] MLAWER E J, TAUBMAN S J, BROWN P D, et al. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave[J]. J Geophys Res Atmos, 1997, 102(D14): 16 663-16 682.
- [36] BARKER D M. Southern high-latitude ensemble data assimilation in the Antarctic Mesoscale Prediction System[J]. Mon Wea Rev, 2005, 133(12): 3 431-3 449.
- [37] 赵俊虎, 陈丽娟, 王东阡. 2016年我国梅雨异常特征及成因分析[J]. 大气科学, 2018, 42(5): 1 055-1 066.
- [38] SHEN Y, ZHAO P, PAN Y, et al. A high spatiotemporal gauge-satellite merged precipitation analysis over China[J]. J Geophys Res Atmos, 2014, 119(6): 3 063-3 075.

## STUDY ON ENSEMBLE-VARIATION HYBRID DATA ASSIMILATION BASED ON OPTIMIZED COMPOSITE SAMPLES

CHEN Yao-deng<sup>1</sup>, GUO Shan<sup>1</sup>, WANG Yuan-bin<sup>1</sup>, ZANG Zeng-liang<sup>2</sup>, PAN Xiao-bing<sup>2</sup>

(1. Key Laboratory of Meteorological Disaster of Ministry of Education/Joint International Research Laboratory of Climate and Environment Change/Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science & Technology, Nanjing 210044, China; 2. College of Meteorology and Oceanography, National University of Defense Technology, Nanjing 211101, China)

**Abstract:** In order to introduce flow-dependent background error covariance effectively and reduce the computing cost of ensemble forecast, the present research attempts to select historical forecasting samples similar to the weather condition at the time of assimilation and combine them with time-lagged samples in hybrid assimilation system. The single observation tests indicate that using the combination of selected historical samples and time-lagged samples, the sample error caused by limited samples is mitigated and flow-dependent background error covariance could be introduced in hybrid assimilation system. Compared with ERA5 and sound dataset, cycling assimilation and forecasts for a week show that 3DVAR has slightly poor performance. Hybrid experiment based on composite samples has the lowest root mean square error (RMSE), and performs better than hybrid experiments which only use the time-lagged samples. Its overall precipitation score is the best, and performs especially well in the assessment of moderate and heavy precipitation. The intensity and location of heavy precipitation center could be predicted well and the problem of over-estimating is alleviated in the composite sample hybrid assimilation scheme.

**Key words:** numerical weather prediction; data assimilation; hybrid assimilation; ensemble sample; flow dependent