

张钰婷, 沈浙奇, 伍艳玲. 基于地球系统模式的局地化粒子滤波器与集合卡尔曼滤波器同化实验[J]. 海洋学报, 2021, 43(10): 137–148, doi:10.12284/hyxb2021139

Zhang Yuting, Shen Zheqi, Wu Yanling. Data assimilation experiments using localized particle filter and ensemble Kalman filter with community earth system model[J]. Haiyang Xuebao, 2021, 43(10): 137–148, doi:10.12284/hyxb2021139

# 基于地球系统模式的局地化粒子滤波器 与集合卡尔曼滤波器同化实验

张钰婷<sup>1</sup>, 沈浙奇<sup>1,2,3\*</sup>, 伍艳玲<sup>1,2,3</sup>

(1. 自然资源部第二海洋研究所 卫星海洋环境动力学国家重点实验室, 浙江 杭州 310012; 2. 河海大学 海洋学院 资料同化与预测研究所, 江苏 南京 210098; 3. 南方海洋科学与工程广东省实验室(珠海), 珠海 广东 519080)

**摘要:** 粒子滤波器 (PF) 是一种非常具有应用前景的非线性资料同化方法。但由于其算法本身存在的粒子退化问题, 目前尚未被广泛地应用于大型地球物理模式。目前主流的集合同化系统仍然倾向于使用集合卡尔曼滤波器 (EnKF) 及其衍生方法。一种新近被提出的局地化粒子滤波器 (LPF) 在经典的粒子滤波器算法中引入局地化技术, 可以使用较小的计算成本有效地避免退化问题, 具有非常大的业务应用潜力。本文在全耦合的通用地球系统模式中开展了 LPF 和 EnKF 的同化实验, 同化资料为模拟的卫星海表温度资料。着重考察了不同局地化参数对两种方法的不同影响, 对比了局地化粒子滤波器与集合卡尔曼滤波器的同化效果差异。比较的结果表明, LPF 的同化效果对于局地化参数的选择非常敏感, 在使用最优局地化参数的条件下, LPF 能达到与 EnKF 相当甚至优于后者的同化效果, 并具有较大的改进空间。

**关键词:** 资料同化; 局地化粒子滤波器; 集合卡尔曼滤波器; 通用地球系统模式; 局地化

中图分类号: P456.7

文献标志码: A

文章编号: 0253-4193(2021)10-0137-12

## 1 引言

作为研究大气、海洋科学热门课题之一, 资料同化技术不仅可以为海洋数值预报模式提供初始场, 还可以构造海洋再分析资料集, 为海洋观测计划和数值预报模式物理量及参数等提供设计依据, 近年来得到了广泛的研究和应用<sup>[1]</sup>。集合卡尔曼滤波器 (EnKF) 是一种有效的资料同化方法, 自 Evensen<sup>[2]</sup> 于 1994 年首次提出以来经过了 20 多年的发展和改进, 已经在海洋资料同化中得到了广泛的研究和应用。近年来, 随着动力模式的不断发展和计算能力的提高, 粒子滤

波器 (PF) 由于不受模型线性和误差高斯分布假设的约束, 也逐渐成为了当前资料同化算法研究的热门<sup>[3]</sup>。

EnKF 是集合预报同卡尔曼滤波器 (KF) 的结合, 它使用集合表示模式变量的概率密度分布, 并采用 KF 的更新公式提供线性模型高斯分布假设下的同化最优解。由于 EnKF 避免了卡尔曼滤波器和扩展卡尔曼滤波器 (EKF) 中的协方差更新模型, 使得它能够被应用于大型地球物理模式。集合调整卡尔曼滤波器 (EAKF) 是在 EnKF 基础上发展起来的衍生方法, 它一般被视为一种确定性的 EnKF 格式<sup>[4]</sup>。EAKF 从滤波理论出发推导了一个用于模式变量更新的算子,

收稿日期: 2020-07-28; 修订日期: 2021-01-14。

基金项目: 国家重点研发计划“海洋环境安全保障”重点专项 (2016YFC1401701); 自然资源部第二海洋研究所基本科研业务费专项 (QNYC1903); 国家自然科学基金 (41606012, 41690124, 41805066, 41806032)。

作者简介: 张钰婷 (1996—), 女, 浙江省余姚市人, 主要从事耦合模式资料同化研究。E-mail: zyt\_ocean@163.com

\* 通信作者: 沈浙奇 (1984—) 男, 浙江省杭州市人, 副教授, 主要从事耦合资料同化系统研发, 资料同化方法和目标观测方法研究。E-mail: zqshen@sio.org.cn

取代了传统 EnKF 中的增益矩阵,且无需增加额外的观测扰动,在计算量上有一定的优势<sup>[4-5]</sup>。作为目前海洋资料同化所采用的主流方法之一——集合卡尔曼滤波器以及衍生方法隐含了预报集合高斯分布的假设,适用于线性系统的同化<sup>[6]</sup>。

相较于集合卡尔曼滤波,粒子滤波不含高斯假设,对非线性非高斯同化系统能产生更好的同化效果。粒子滤波算法基于贝叶斯估计理论,是贝叶斯公式的蒙特卡罗算法近似。“粒子”与集合卡尔曼滤波器中的集合成员相同,是用于表示模式变量的概率分布的集合样本。如果计算资源充足,随着粒子数目的增加,粒子的概率密度分布会逐渐趋向于真实状态场的概率密度分布,粒子滤波器能够实现最优贝叶斯估计的同化效果<sup>[3,7]</sup>。经典的粒子滤波器使用似然函数计算每个粒子的标量权重,因此当状态场空间的维数较大时,状态场数值的微小变化会引起对应权重在量级上的巨大变化,多数集合成员会因权重过小而失效,从而导致粒子滤波器的退化<sup>[8]</sup>。相当多的方法已经被提出来处理粒子滤波器中的退化<sup>[9]</sup>,如最优重要性粒子滤波器<sup>[10]</sup>、等权重粒子滤波器<sup>[11]</sup>、集合卡尔曼粒子滤波器<sup>[12]</sup>、局地化粒子滤波器<sup>[8,13]</sup>等。本文主要考察的局地化粒子滤波器是最近才被提出的一种同化方法,它通过在经典粒子滤波器中引入 EnKF 中常用的局地化方法来解决粒子滤波器的退化问题,得到了广泛关注。

局地化方法自 21 世纪初在集合卡尔曼滤波器的同化中被提出以来<sup>[4]</sup>,已经被广泛地应用于各种业务化集合同化系统中,得到了普遍认可。由于模式集合成员的数量非常有限,在计算背景误差协方差和增益矩阵的过程中往往会出现由于样本不足而造成的虚假远距离相关,造成错误的同化更新。通过引入局地化可以使得同化的更新过程在一个较小子空间中进行,从而抑制这种虚假相关,并大大降低计算量。利用相似的思想,最近的一些工作开始在粒子滤波器中使用局地化,例如, Poterjoy<sup>[13]</sup>、Shen 等<sup>[15]</sup>、Penny 和 Miyoshi<sup>[8]</sup> 改善的局地化粒子滤波器已经被初步验证能够使用与集合卡尔曼滤波器相当的集合成员数来避免粒子退化现象,也逐渐地被用于地球物理模式的资料同化中。

本文在耦合的通用地球系统模式 (Community Earth System Model, CESM) 中开展了集合卡尔曼滤波器和局地化粒子滤波器的观测系统模拟试验。通过同化模拟的卫星海表温度 (SST) 资料,考察不同局地化参数对于两种滤波器方法的不同影响。在此基础

上,进一步比较了两种滤波器方法的同化效果,探讨了同类方法的优缺点以及发展前景。

## 2 数据与方法

### 2.1 模式和同化系统

实验使用的模式为 CESM,它是由美国国家大气研究中心 (NCAR) 于 2010 年发布的新一代地球系统模式,是目前最先进、使用最广泛的地球系统模式之一。该模式采用模块化框架,主体由大气、海洋、陆地、海冰、河流等几大模块组成,并由耦合器 (CPL7) 管理模块间的数据信息交换和模式运行。实验使用了 CESM1.2.1 版本,它的海洋模式为 POP2,大气模式为 CAM4。本实验使用全耦合的模式设置 (B compset) 和当代 (present day) 的外强迫。采用的模式分辨率为  $0.9 \times 1.25_{\text{gx1v6}}$ ,即大气模式水平分辨率为  $0.9^\circ \times 1.25^\circ$ ,垂向 26 层;海洋模式水平分辨率接近  $1^\circ$ ,在赤道区域纬向加密约为  $0.5^\circ$ ,垂向 60 层。

实验所涉及的两种同化方法都是借助 NCAR 开发的资料同化研究平台 (Data Assimilation Research Testbed, DART) 实现。DART 是由 NCAR 的数据同化研究部门开发和维护的一个开源软件。它提供了多种确定性和随机性滤波器算法,包括集合调整卡尔曼滤波器、集合卡尔曼滤波器、核滤波器和粒子滤波器等<sup>[16]</sup>。我们建立了 DART 和 POP2 模式的接口,实现了耦合模式框架下的海洋观测资料的弱耦合同化。鉴于计算量的考虑,我们使用了 20 个集合成员。在所有的实验中,我们也采用了固定系数为 1.02 的协方差膨胀和针对 20 个集合成员的样本误差订正 (Sampling Error Correction) 方案<sup>[17]</sup>。

### 2.2 数据

本文采用的实验方案为观测系统模拟试验 (OSSE),其基本思想是使用一组真值实验作为参考值,然后根据实际存在的观测系统对真值取样,并且叠加具有给定方差的随机误差来模拟观测。在实验中同化上述模拟的观测资料,并利用真值实验得到的参考值来评估不同同化方法得到的分析场。

图 1 显示了 OSSE 的流程图,其中我们利用真值实验生成每周一次的全球海表温度观测,其空间分辨率为  $1^\circ \times 1^\circ$ ,并叠加符合高斯分布的随机误差作为观测误差。为简化讨论,我们针对所有观测点使用相同的观测误差标准差。参考 OISST<sup>[18]</sup> 中的 SST 观测误差的全球平均标准差,假设所有位置的误差标准差均为  $0.3^\circ\text{C}$ 。

此外,本文使用了哈德雷中心 (Met Office Hadley

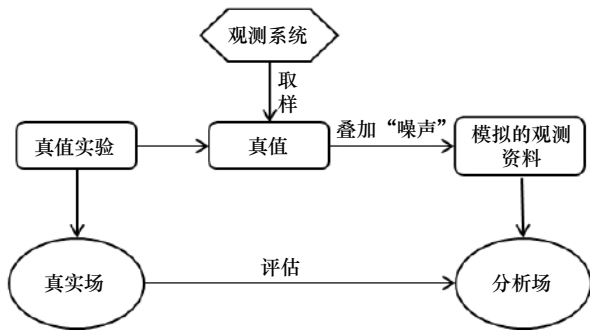


图1 观测系统模拟试验流程图设计

Fig. 1 Flow chart design of observation system simulation experiment

Center) 提供的长度为 100 a (1919–2018) 的月平均 HadISST (Hadley Centre Sea Ice and Sea Surface Temperature data set) 再分析资料<sup>[9]</sup>, 主要用于计算 SST 资料对不同距离的变量的相关系数, 并与模式 20 个集合成员计算的样本相关系数进行比较, 为局地化系数的选取提供理论依据(见 3.1.2 节)。

### 2.3 同化方法和局地化

#### 2.3.1 EAKF 中的局地化

集合调整卡尔曼滤波器(EAKF)<sup>[4]</sup>是在集合卡尔曼滤波器<sup>[2]</sup>的基础上发展起来的。EAKF 避免了集合卡尔曼滤波器中对观测资料的扰动, 且在集合样本数目较少时仍能得到较为满意的结果。EAKF 基于线性回归理论逐个顺次地同化观测资料, 使用观测空间的局地化后可以将每个观测点的更新范围限制在一个较小的子空间中进行, 节省了计算量。

EAKF 的同化步骤可以表示如下, 若用  $x$  表示状态向量, 用  $y^o$  表示误差方差为  $\sigma_o^2$  的观测标量,  $h$  表示观测算子, 则 EAKF 公式的第一步是使用观测算子  $h$  将状态空间的每个集合成员投影到观测点上, 作为每个成员对应观测的先验估计, 如下

$$\rho(d_n; c) = \begin{cases} -\frac{1}{4}\left(\frac{d_n}{c}\right)^5 + \frac{1}{2}\left(\frac{d_n}{c}\right)^4 + \frac{5}{8}\left(\frac{d_n}{c}\right)^3 - \frac{5}{3}\left(\frac{d_n}{c}\right)^2 + 1 & 0 \leq d_n < c \\ \frac{1}{12}\left(\frac{d_n}{c}\right)^5 - \frac{1}{2}\left(\frac{d_n}{c}\right)^4 + \frac{5}{8}\left(\frac{d_n}{c}\right)^3 + \frac{5}{3}\left(\frac{d_n}{c}\right)^2 - 5\left(\frac{d_n}{c}\right) + 4 - \frac{2}{3}\left(\frac{d_n}{c}\right)^{-1} & c \leq d_n < 2c \\ 0 & d_n \geq 2c \end{cases}, \quad (7)$$

式中,  $d_n$  代表模式点  $x_{m,n}$  和观测点  $y_{p,n}$  的距离;  $c$  是一个局地化参数, 可以控制  $\rho$  的去相关距离。由式(7)可知, 当  $d_n$  大于  $2c$  时,  $\rho = 0$ , 从而观测点  $y_{p,n}$  和模式点  $x_{m,n}$  完全不相干。

#### 2.3.2 LPF 中的局地化

相比于 EAKF, 局地化的粒子滤波器(LPFF)最近才被提出来解决传统粒子滤波器中的退化问题<sup>[13]</sup>, 因此关于它的研究结果相对较少。在作者所了解的文

$$y_{p,n} = h(x_{p,n}) \quad n = 1, \dots, N, \quad (1)$$

式中, 下标  $p$  代表先验值;  $n$  代表集成员。

EAKF 使用观测误差方差  $\sigma_o^2$  更新先验误差方差  $\sigma_p^2$ , 得到后验误差方差为

$$\sigma_u^2 = [(\sigma_p^2)^{-1} + (\sigma_o^2)^{-1}]^{-1}, \quad (2)$$

式中, 下标  $u$  代表后验值。

接着使用以下公式计算观测空间的每个成员的后验估计

$$y_{u,n} - \bar{y}_u = \left(\frac{\sigma_u}{\sigma_p}\right) (y_{p,n} - \bar{y}_p), \quad (3)$$

式中,  $\bar{y}_p$  为先验集合的平均;  $\bar{y}_u$  为后验集合的平均且由以下公式计算

$$\bar{y}_u = \sigma_u^2 \left( \frac{\bar{y}_p}{\sigma_p^2} + \frac{y^o}{\sigma_o^2} \right). \quad (4)$$

EAKF 最后利用相关系数将每个观测的增量回归到模式变量中

$$x_{m,n}^u = x_{m,n} + \frac{\sigma_{x_m y}}{\sigma_p^2} (y_{u,n} - y_{p,n}) \quad n = 1, \dots, N, \quad (5)$$

式中,  $x_{m,n}$  表示模式第  $m$  个分量的第  $n$  个集成员;  $\sigma_{x_m y}$  为  $x_m$  与  $y_p$  的协方差。

由于集合模式成员数有限, 因此所模拟的背景误差协方差会存在一定的虚假信息。这些虚假信息随着距离的增加而增加, 且会在同化过程中造成虚假的更新量, 影响结果的准确性。为了避免这些问题, 我们在 EAKF 中引入了局地化方案, 使用一个数值介于 0 和 1 之间且随距离的大小单调递减的因子  $\rho$ , 并将式(5)替换为

$$x_{m,n}^u = x_{m,n} + \rho \frac{\sigma_{x_m y}}{\sigma_p^2} (y_{u,n} - y_{p,n}), \quad n = 1, \dots, N. \quad (6)$$

公式(6)中的  $\rho$  被称为局地化因子, 是一个依赖于距离的函数, 一般可以由以下公式<sup>[20]</sup> 计算得到

献中, 目前尚没有将 LPF 应用于 CESM 模式的研究, 也鲜有对 LPF 的局地化参数进行的讨论。

从原理上说, 在粒子滤波器中采用局地化是为了使用较小的计算本来避免滤波退化问题。粒子滤波器的原理是给每个集成员配给一个权重, 用集成员的加权组合来表示同化变量的完整概率分布密度函数, 并随着同化的进行不断更新权重。经典的粒子滤波器使用一个标量的数来代表权重, 因此当模式

的状态变量维数巨大时,很容易造成其中一个成员占据大部分权重,而其他成员权重都几乎为 0 的现象,这就是滤波退化<sup>[2]</sup>。

在经典粒子滤波器中,状态场 $X^n$ 的标量权重可由以下公式计算

$$w^n \propto \prod_{i=1}^m \exp \left\{ -\frac{1}{2} [y_i - h_i(X^n)]^2 / \sigma_i^2 \right\}, \quad (8)$$

式中,  $w^n$  表示第  $n$  个粒子的标量权重;  $m$  是所有观测点的总数; 观测算子  $h_i$  用于将模式预报场投影到其中的某个观测点  $y_i$  上;  $\sigma_i^2$  是  $y_i$  的方差, 这里的正比例符号“ $\propto$ ”意味着计算等式右边之后还需要进行一次标准化使得所有权重的和为 1。当  $m$  很大时, 多个指数函数相乘导致不同粒子的权重很容易有量级上的差异。

在粒子滤波器中引入局地化的基础是将粒子的权重扩展成为矢量, 即不同的模式网格点使用不同的权重。在此基础上应用粒子滤波器可以将状态分析过程转移到一个较小子空间中进行, 由于子空间的差异性会增加集合成员的多样性, 从而降低集合退化的可能。粒子滤波器中的局地化主要用于计算矢量权重, 使其适用于大型地球物理模式。但是由于粒子滤波器的同化原理与卡尔曼滤波器有本质的不同, 局地化的引入方式也有所不同<sup>[21]</sup>。局地化粒子滤波器的第一步是将式(8)中的标量权重扩展到矢量权重, 也就是说状态场中各个不同的分量  $x_j^n$  采用不同的局部权重  $w_j^n$ , 与此同时, 集合成员的重取样也都在局地进行。Poterjoy<sup>[13]</sup> 借助局地化公式, 利用局地化因子  $\rho$  给出不同分量的权重公式如下

$$w_j^n \propto \prod_{i=1}^m [p(y_i | X^n) \rho + (1 - \rho)], \quad (9)$$

式中,  $p(y_i | X^n)$  为式(8)右侧指数项, 对应每个观测点造成的权重分布,  $\rho$  由式(7)得到, 对应与距离相关的局地化因子。  $m$  在这里代表所有可以影响该模式网格点的观测数目, 根据使用不同的局地化参数而不同。式(9)的本质是: 将各观测点标量权重, 按其于模式点距离的差异, 利用局地化因子  $\rho$  插值到模式网格中, 从而给出非观测点的权重。为了在数值模式中执行 LPF 算法, Poterjoy<sup>[13]</sup> 引入了一种融合采样 (merging and sampling) 的方法, 在序贯的局部重取样过程中, 使用两个与距离相关的权重系数来对重采样粒子和原始粒子进行加权组合, 从而减小因局部重采样而造成的对变量连续性的破坏, 目前该算法也已经在 DART 中被采纳。具体算法流程可参考 Poterjoy<sup>[13]</sup> 和 Shen 等<sup>[15]</sup>。

## 2.4 同化实验设计

本文首先通过敏感性实验考察局地化参数对于 EAKF 和 LPF 的同化效果的影响。局地化的参数数值对应了观测点和模式网格点的去相关距离: 假设局地化参数的值为  $c$ , 那么观测点与距离  $2c$  以上的模式网格点的相关系数为 0。CESM 模式中的距离使用弧度制单位表示, 例如  $d=0.1$  rad, 那么它实际对应的赤道经度为  $\frac{0.1 \times 180^\circ}{\pi} \approx 5.7^\circ$ ; 为了处理海洋在垂直和水平方向的不同尺度, 同化算法使用一个垂向归一化尺度系数 (Vertical Normalization Factor, 以下记做  $\nu$ , 单位为 m/rad) 来进行垂向距离的转化。假设水平方向的距离为  $d_h$ , 单位为 ( $^\circ$ ), 垂直方向的距离为  $d_v$ , 单位为 m, 那么两点之间的模式距离如下计算

$$d = \sqrt{d_h^2 + \left(\frac{d_v}{\nu}\right)^2}. \quad (10)$$

根据  $d$  和  $c$  的比值, 可以使用式(7)计算局地化因子  $\rho$ 。显然  $\nu$  的数值越大, 相同的实际垂向距离  $d_v$  条件下式(10)右端的第二项越小, 相同的  $c$  可以影响到越远的水平距离。而当  $\nu$  为无穷大的时候, 实际上就关闭了垂直方向的局地化, 局地化因子仅由  $d_h$  决定。

实验所选取的同化数据为海表温度, 集合成员数为 20, 同化频率为每 7 d 同化一次, 同化实验的时间为 12 个月。同化的初始集合由如下方法产生: 我们首先对 CESM 模式进行了 100 a 的自由积分, 然后在得到的初始场上层 30 层温度变量上叠加伪随机场<sup>[22]</sup> 的扰动, 构成 20 个集合成员。我们对该集合进行 2 a 的模式积分, 使得每个集合成员各变量之间保持一定的动力平衡, 这些积分的结果则被用来作为同化实验的初始集合。

设计的敏感性实验如表 1 所示: 我们首先考察 EAKF 中的局地化方案, 比较了不同局地化参数  $c$  和垂向归一化尺度系数  $\nu$  对于同化效果的影响。我们先固定  $c$  为 0.1 rad, 然后分别设定  $\nu=1\ 000$  m/rad, 1 500 m/rad, 2 000 m/rad, 以及  $+\infty$  (即垂直方向不采用局地化) 来考察垂向局地化方案对于同化效果的影响。并且分别将实验名称记作 Kc0.1v1000、Kc0.1v1500、Kc0.1v2000 以及 Kc0.1vinf。根据 CESM 气候模式的距离设置, Kc0.1v1000 试验中, 观测点在海表与水平距离超过大约  $11.5^\circ$  的模式格点相关系数为 0, 同时与其正下方垂直距离超过 200 m 的模式格点相关系数也为 0。然后, 我们固定最优垂向归一化尺度系数  $\nu$ , 来考察水平局地化方案对同化效果的影响。由于前一组实验得出的结论为 EAKF 的最优垂向局地化方案为垂向

表 1 实验列表  
Table 1 Experimental list

实验名称	同化方法	$\nu/(m \cdot rad^{-1})$	$c/rad$
EAKF垂向局地化方案实验	EAKF	Kc0.1v1000	1 000
		Kc0.1v1500	1 500
		Kc0.1v2000	2 000
EAKF局地化参数实验	EAKF	Kc0.1vinf	$\infty$
		Kc0.05vinf	$\infty$
		Kc0.1vinf	$\infty$
		Kc0.2vinf	$\infty$
		Kc0.3vinf	$\infty$
		Kc0.1v500	500
LPF垂向局地化方案实验	LPF	Pc0.1v1000	1 000
		Pc0.1v1500	1 500
		Pc0.1vinf	$\infty$
LPF局地化参数实验	LPF	Pc0.05v2000	2 000
		Pc0.1v1000	1 000
		Pc0.2v500	500
控制实验	FREERUN	无同化	

关闭局地化(见 3.1.1 节), 在此基础上, 分别设置局地化参数  $c$  为 0.05 rad、0.1 rad、0.2 rad、0.3 rad, 分别将对应的实验名称记为 Kc0.05vinf、Kc0.1vinf、Kc0.2vinf、Kc0.3vinf。类似地, 我们也针对 LPF 探讨了系数  $\nu$  和参数  $c$  的相关问题, 利用敏感性实验寻找最优的局地化方案, 揭示 LPF 和 EAKF 对于局地化的不同要求。最后, 基于使用最优的局地化方案, 我们比较了 LPF 和 EAKF 的同化效果, 揭示 LPF 的潜在优势与不足。同时, 为了显示同化效果, 本实验设置了不进行任何同化的控制试验(下文简称 FREERUN)进行对照。

### 3 实验结果与讨论

#### 3.1 EAKF 局地化对同化效果影响

##### 3.1.1 EAKF 的垂向局地化方案

我们首先固定局地化参数  $c$  为 0.1 rad, 考察不同垂向归一化尺度系数  $\nu$  对于同化效果的影响。实验采用的系数  $\nu$  为 1 000 m/rad、1 500 m/rad、2 000 m/rad, 与局地化参数  $c$  相乘以后, 对应的距离分别为 100 m/150 m/200 m, 也就是说, 局地化系数  $\rho$  在超过 200 m/300 m/400 m 的数值为 0, 因此, SST 观测最多能影响的垂直深度为 200 m/300 m/400 m。实验选取了均方根误差(RMSE)作为评判同化效果优劣的标准, 其计

算公式为

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i^{assim} - x_i^{true})^2}, \quad (11)$$

式中,  $x_i^{assim}$  代表同化结果的集合平均值;  $x_i^{true}$  表示真实场数据, 下标  $i$  可分别遍历不同空间位置或者不同时间。为了避免同化初始阶段的较大误差对于比较结果的影响, 我们只使用后 6 个月的分析场误差计算 RMSE。同时, 由于模式在高纬度地区的模拟效果较差, 为了凸显局地化的不同影响, 我们只考察 60°S 和 60°N 之间区域的变量。

我们首先根据 RMSE 的垂直分布讨论系数  $\nu$  对同化效果的影响。图 2 为设置不同  $\nu$  时, EAKF 同化 SST 资料后的垂向温盐 RMSE。由于同化资料为 SST, 因此温度的同化效果要明显强于盐度的, 在深度较浅时其优势更加显著。对于温度变量, 使用不同的系数  $\nu$  同化 SST 得到的 RMSE 在 100 m 以浅并没有显著差异, 且都远小于控制实验的 RMSE。而随着深度的增加, 变量与观测之间的距离增大, 一方面所有实验的同化效果逐渐削弱, 另一方面不同实验的

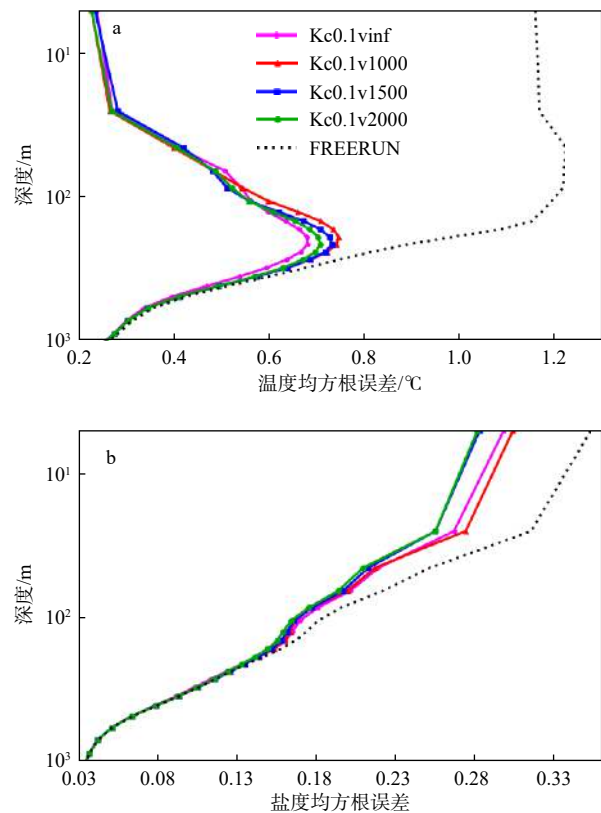


图 2 不同垂向局地化方案 EAKF 实验中区域平均 (60°S~60°N, 环地球) 垂向均方根误差

Fig. 2 Regional mean (60°S–60°N, ring the earth) root mean square error in EAKF experiments with different vertical localization schemes

RMSE 出现差异。特别地,当深度约为 150~300 m 时,在关闭垂向局地化的 Kc0.1vinf 实验产生的 RMSE 是最小的。根据式(6)可知,EAKF 利用 SST 更新深层温度变量的原理是使用 SST 和给定层温度之间的相关性将表层的观测增量回归到深层。因为深层温度的离散度(标准差)很小,仅有  $O(10^{-2})$ ,所以即使只使用 20 个集合成员计算相关系数,表层和深层的相关系数也能够正确表达。所以,虽然 Kc0.1vinf 实验没有引入垂向局地化,较深层的温度也能够得到正确更新。

另一方面,同化 SST 对于盐度的改进基于温盐相关得到。在 150 m 以浅,仍然可以发现盐度变量的 RMSE 小于控制实验的 RMSE,但是不同垂直局地化方案对于盐度同化效果的差异主要在较浅层而非深层出现。同时,图 2b 的结果表明同化 SST 时  $\nu=1\ 500\ \text{m}/\text{rad}/2\ 000\ \text{m}/\text{rad}$  的局地化能更好地改进浅层的盐度,由式(10)可知,在局地化参数  $c$  固定的条件下,对于浅层的变量,尺度系数  $\nu$  越大,  $d_h$  起到的作用也相对越大。体现在盐度上就是相同深度的盐度变量获取海表观测的信息就越多。所以我们可以断言,浅层的盐度同化差异实际上受到水平局地化的影响更大。综合两者的讨论,在垂直方向不采用局地化是相对更优的方案。

### 3.1.2 EAKF 局地化参数

对 EAKF 垂直局地化方案的敏感实验结果显示,EAKF 的同化效果在关闭垂直方向的局地化时相对较好。因此,我们在关闭垂直局地化的基础上,分析不同参数  $c$  对应的水平局地化方案对 EAKF 同化效果的影响,第 2 组实验采用的局地化参数分别为 0.05 rad、0.1 rad、0.2 rad、0.3 rad。图 3 从时间尺度上分析各物理量 RMSE 的变化情况。首先分析表层温度、盐度(SSS)和海表高度(SSH)变量的 RMSE 变化情况,如图 3a 所示,SST 从启用同化的第 2 个月开始就一直保持较小的 RMSE,且使用不同参数实验的结果没有明显差异。这是由于 SST 的观测资料比较密集,即使在某个点的同化中引入虚假更新,也很容易被其他点的正确更新所抵消。而对其他两个变量来说,同化 SST 并不能立刻减小误差,需要进行一段时间的持续同化和模式积分,误差才能显著减小,这点对于 SSS 特别明显。图 3b 显示,SSS 的 RMSE 尚处于下降阶段,没有达到稳定,其同化效果与局地化参数的大小并没有明显的相关关系,在最后 3 个月,几种参数实验的结果没有显著差别。而 SSH 在局地化参数稍大时同化效果较好,这说明 SST 和 SSH 之间的相关性较好,不容易产生虚假相关。

我们进一步分析较深层的同化效果发现,当深度

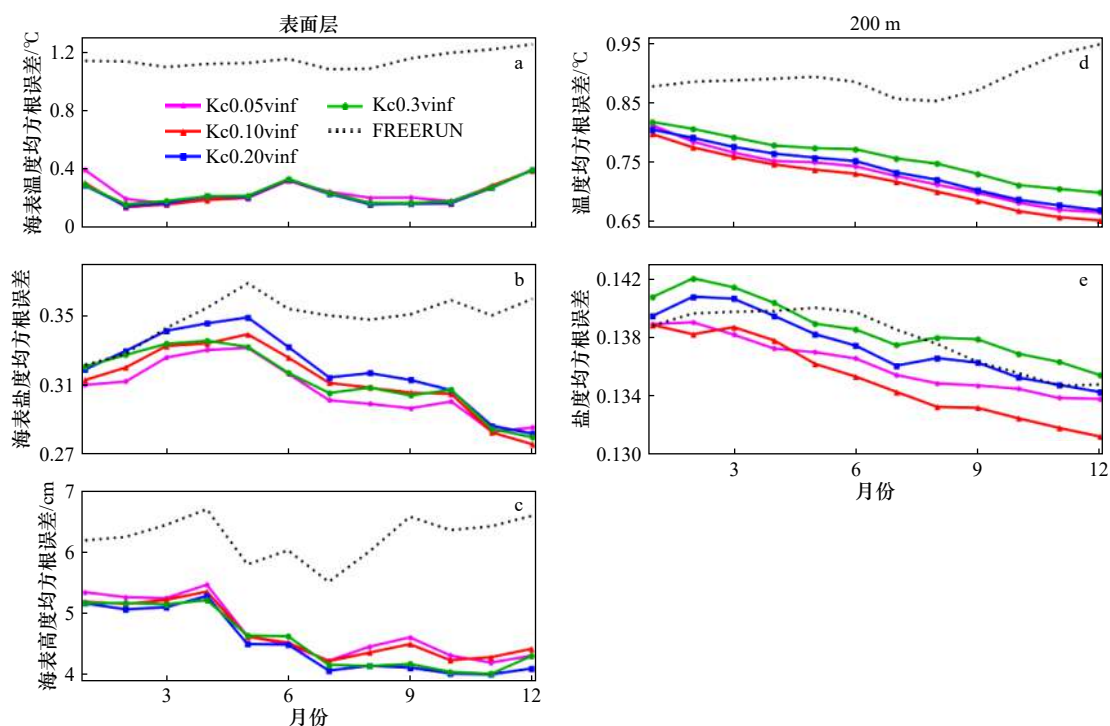


图 3 不同局地化参数 EAKF 实验中区域平均(60°S~60°N, 环地球)的均方根误差时间序列

Fig. 3 RMSE time series of regional mean (60°S~60°N, ring the earth) in EAKF experiments with different local parameters

a. 海表温度; b. 海表盐度; c. 海表高度; d. 200 m 的温度; e. 200 m 的盐度

a. Sea surface temperature; b. sea surface salinity; c. sea surface height; d. temperature at 200 m; e. salinity of 200 m

为 200 m 时, 图 3d 和图 3e 表现出最优同化效果的局地化参数  $c=0.1$  rad。这是因为我们关闭了垂向局地化功能, 因而使得一定水平距离之外的海表观测资料能够无差别地影响整个水柱, 随着深度的增加, 实际的距离增大, 远距离观测带来的虚假相关也会出现。从温度和盐度两个变量, 都可以看出  $c=0.1$  rad 的同化效果好于  $c=0.05$  rad, 这是因为更多的观测资料被用来更新 200 m 深度的变量; 而它的同化效果也好于局地化参数更大的另外两个实验, 这是因为后两者会因为虚假相关而带来虚假的更新。

由于集合滤波器的集合成员数量有限, 在相距较远的两个点之间会产生虚假的相关, 造成错误的更新, 这是在 EAKF 中引入局地化的主要原因。因此, 如果可以量化表示给定集合成员数可能造成的虚假相关, 就可以相应地选择最优的局地化参数。针对某个给定点上的观测, 以 1 月的控制实验预报值为例, 我们使用以下公式定义对应局地化参数  $c$  的虚假相关占比 (Spurious Correlation Ratio, SCR)

$$\text{SCR}(c) = \frac{1}{m} \sum_i^m \left| \frac{\rho(c)r_i^{\text{sample}} - r_i^{\text{true}}}{r_i^{\text{true}}} \right|, \quad (12)$$

式中,  $r_i^{\text{sample}}$  代表使用 20 个集合成员计算的月平均 SST 的样本相关系数, 乘上局地化因子  $\rho(c)$  就是实际同化中用到的相关系数, 我们将两者的乘积称之为有效相关系数;  $r_i^{\text{true}}$  表示“真实”的相关系数, 实践中一般可以使用再分析资料计算的分析相关系数近似表示。这里我们使用月平均的 HadISST 资料来计算该相关系数。具体如下: 使用 1919–2018 年每年 1 月份减去气候态 SST 的资料表示 100 个集合成员, 并且计算这 100 个成员的相关系数作为分析相关系数。 $m$  表示有效相关系数大于 0 的点的数量, 即该观测点在一次同化中对模式产生虚假影响的网格点数目, 为了显示局地化对虚假相关的抑制作用, 本实验中仅把  $\rho(c)r_i^{\text{sample}} - r_i^{\text{true}} > 0$  的情况纳入考量。

图 4 为 SST 相关系数的空间分布, 我们分别选取了位于太平洋的两个点 ( $0^\circ, 180^\circ$ )、( $20^\circ\text{S}, 120^\circ\text{W}$ )、位于印度洋的一个点 ( $20^\circ\text{S}, 60^\circ\text{E}$ )、以及位于大西洋的一个点 ( $20^\circ\text{N}, 40^\circ\text{W}$ ), 计算了使用不同局地化参数下的有效相关系数, 样本相关系数和分析相关系数, 分别对应图 4 中的前 4 列、第 5 列以及第 6 列, 图中的白色区域表示相关系数小于 0 部分。从图中可以

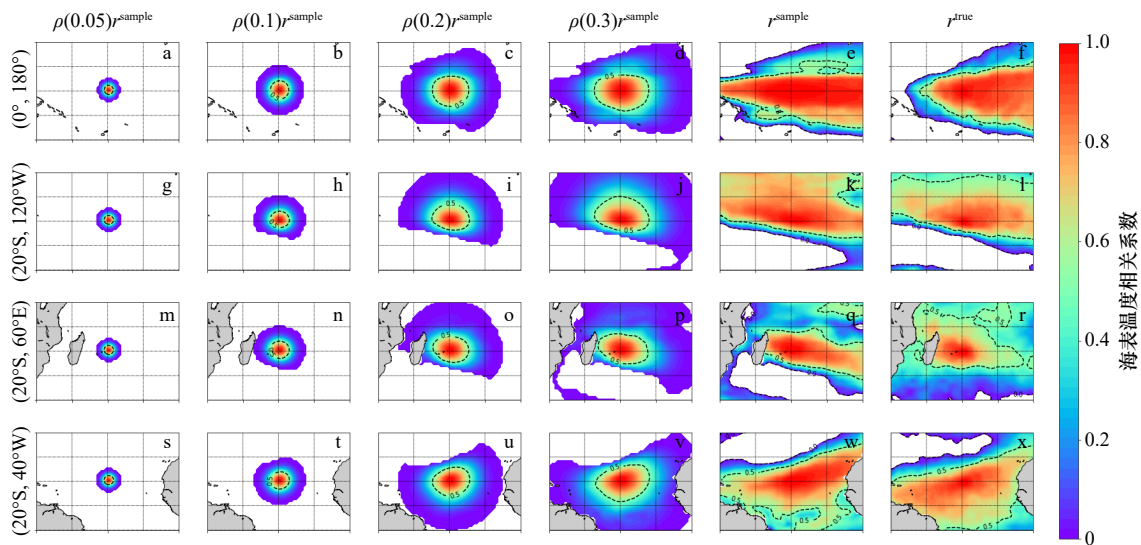


图 4 海表温度相关系数

Fig. 4 Correlation coefficient of sea surface temperature

a–f. 相对于 ( $0^\circ, 180^\circ$ ) 位置处海表温度相关系数; g–l. 相对于 ( $20^\circ\text{S}, 120^\circ\text{W}$ ) 位置处海表温度相关系数; m–r. 相对于 ( $20^\circ\text{S}, 60^\circ\text{E}$ ) 位置处海表温度相关系数; s–x. 相对于 ( $20^\circ\text{N}, 40^\circ\text{W}$ ) 位置处海表温度相关系数; a, g, m, s. 局地化参数为 0.05 时海表温度有效相关系数; b, h, n, t. 局地化参数为 0.1 时海表温度有效相关系数; c, i, o, u. 局地化参数为 0.2 时海表温度有效相关系数; d, j, p, v. 局地化参数为 0.3 时海表温度有效相关系数; e, k, q, w. 无局地化时海表温度样本相关系数; f, l, r, x. HadISST 资料海表温度分析相关系数。空白区域代表相关系数小于 0 的区域

a–f. Relative to ( $0^\circ, 180^\circ$ ); g–l. relative to ( $20^\circ\text{S}, 120^\circ\text{W}$ ); m–r. relative to ( $20^\circ\text{S}, 60^\circ\text{E}$ ); s–x. relative to ( $20^\circ\text{N}, 40^\circ\text{W}$ ). a, g, m, s. Effective correlation coefficient of sea surface temperature when local parameter is 0.05; b, h, n, t. effective correlation coefficient of sea surface temperature when the local parameter is 0.1; c, i, o, u. effective correlation coefficient of sea surface temperature when the local parameter is 0.2; d, j, p, v. effective correlation coefficient of sea surface temperature when the local parameter is 0.3; e, k, q, w. sample correlation coefficient of sea surface temperature without localization; f, l, r, x. analytical correlation coefficient of sea surface temperature. The areas in which correlation coefficients are smaller than 0 are blanked

看出,几乎所有位置的样本相关系数都大于分析相关系数,因此在某些较远区域,样本会夸大 SST 的相关性,进而引入虚假相关。我们也随机选取了其他多个格点进行相同的计算分析,得到的结论类似。通过引入局地化算法,设置恰当的局地化半径可以在保证近距离相关的前提下去除远距离虚假相关,避免远距观测引起的错误更新<sup>[23]</sup>。

我们进一步计算了几种不同局地化方案下的虚假相关占比(表 2),发现局地化方案的引入大大减小了虚假相关的占比。其中当局地化参数  $c=0.1$  rad 时,既不会因为半径过大而引入一些不必要的虚假相关,也不会因为局地化半径过小而滤去大部分正确相关。因此  $c=0.1$  rad 相对而言是更优的局地化参数,与前面敏感性实验的结果一致。

### 3.2 LPF 局地化对同化效果影响

#### 3.2.1 LPF 垂向局地化方案

第 3 组实验考虑了 LPF 垂向局地化方案对同化效果的影响,类似 EAKF 局地化实验,我们首先固定  $c$  为  $0.1$  rad,设置垂向归一化系数  $v=500/1\ 000/1\ 500/\infty$  m/rad。首先,我们从海洋温度和盐度 RMSE 的垂向分布分析 LPF 垂直局地化对同化效果的影响。如图 5 所示,在浅层,当  $v=1\ 000/1\ 500$  m/rad,即对应最大垂向同化距离为  $200$  m/ $300$  m 时,LPF 的同化效果较好。由式(10)可知,当局地化参数  $c$  固定,垂直距离  $d_v$  较小且固定时,观测点到它能同化到的最远模式网格点的水平距离  $d_h$  随着  $v$  的增大而增大。在较浅的深度,  $v$  的值越大就能在水平方向取得越多的观测资料来计算权重,如果不发生退化,就能获得越好的同化效果。而当深度较大时,无论对于温度还是盐度来说,都可以发现关闭垂向局地化的 LPF 的效果较差,甚至会造成负面同化效果——即同化后的误差反而大于控制实验(例如大于  $500$  m 的温度误差)。这是由于

表 2 SST 相关系数虚假相关占比

Table 2 Proportion of false correlation in SST correlation coefficient

位置	虚假相关占比/%				
	局地化参数 $c$ /rad				
	0.05	0.1	0.2	0.3	无局地化
(0°, 180°)	0.003 4	0.000 9	3.699 6	14.947 4	264.448 6
(20°S, 120°W)	0.014 4	0.004 0	0.556 4	1.110 1	381.896 4
(20°S, 60°E)	0.013 9	0.004 3	0.059 3	1.045 0	258.742 8
(20°N, 40°W)	0.011 0	0.006 3	0.191 7	1.565 5	250.448 1
平均	0.010 7	0.003 9	1.126 8	4.667 0	288.884 0

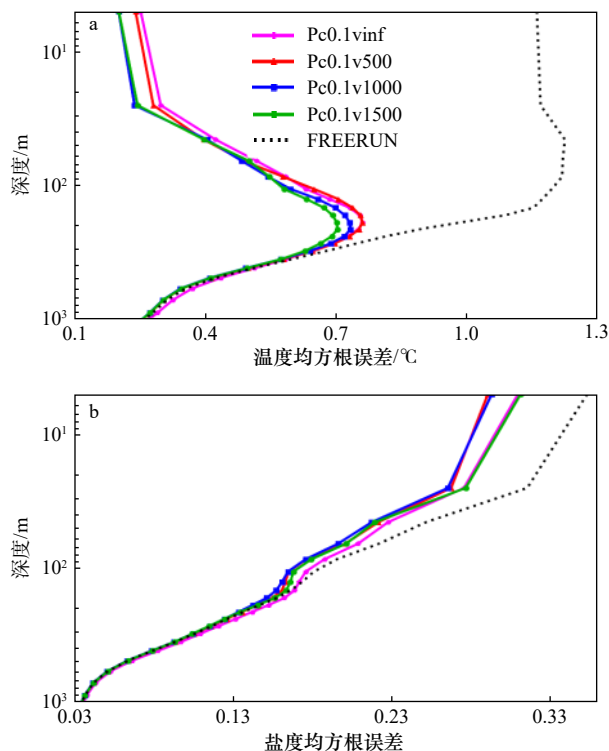


图 5 不同垂向局地化方案 LPF 实验中区域平均 ( $60^{\circ}\text{S}\sim 60^{\circ}\text{N}$ , 环地球)垂向均方根误差

Fig. 5 Regional mean ( $60^{\circ}\text{S}\sim 60^{\circ}\text{N}$ , ring the earth) root mean square error in LPF experiments with different vertical localization schemes

粒子滤波器依赖权重的计算和重采样来更新变量,关闭垂向局地化意味着所有层都会根据表层观测计算的权重来重分配样本,这会过度夸张表层观测的影响范围,从而给深层的变量带来错误的更新<sup>[24]</sup>。综合温度和盐度的实验结果,可以发现当  $v=1\ 000$  m/rad,也就是垂向局地化距离为  $100$  m 时,LPF 的同化效果最好。

#### 3.2.2 LPF 局地化参数

接下来,我们固定 LPF 最优垂向局地化距离为  $100$  m,分析 LPF 的不同局地化参数  $c$  对同化效果的影响。由式(10)可知,考虑垂向局地化的情况下,为了保持垂直方向的局地化距离为  $100$  m,针对不同的  $c$  需要采用不同的  $v$ 。图 6 为采用不同局地化参数的 LPF 同化 SST 资料后的垂向 RMSE,从图中可以很明显的看出,LPF 的效果对局地化参数  $c$  非常敏感。当  $c=0.2$  rad 时,同化效果非常不理想,这主要是由滤波退化造成的。 $c$  的数值越大,越多的观测能同时影响给定点的权重,从而导致该点的权重更容易退化。集合退化会大大降低粒子滤波器的效率,特别地,我们可以发现  $c=0.2$  rad 时的盐度没有同化效果。另一方面,当  $c$  的数值太小时,由于获取资料过少,无法正确表达状态变量的概率分布密度函数,也会导致同化效果不佳。图 7 进一步展示了不同局地化参数得到的

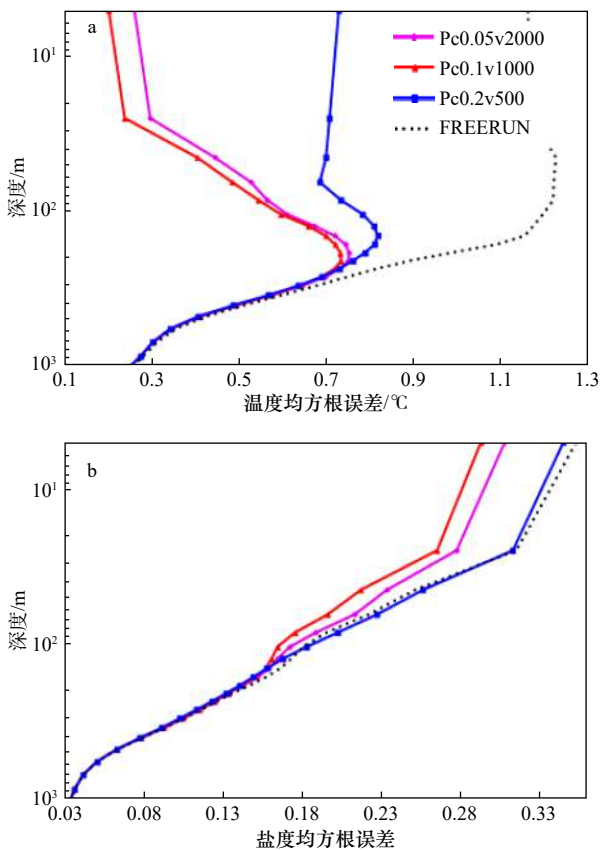


图6 不同局地化参数 LPF 实验中区域平均  
(60°S~60°N, 环地球)垂直向均方根误差

Fig. 6 Regional mean vertical root mean square error  
(60°S~60°N, ring the earth) in LPF experiments with different  
local parameters

分析场的表层变量月平均 RMSE 演变, 不同于图 3, 即使只考虑表层变量, 局地化参数对于同化效果也有很大的影响。我们可以进一步确认  $c=0.2$  rad 会导致 LPF 的退化。综合图 6 和图 7 的结果, 本实验中最优的 LPF 局地化参数  $c=0.1$  rad。

### 3.3 EAKF 与 LPF 的同化效果比较

最后我们采用最优局地化参数的 EAKF 与 LPF 进行对比。图 8 比较了 EAKF 与 LPF 的 RMSE 垂向分布。从温度 RMSE 可以看出, 当深度小于 100 m 时, LPF (蓝线) 同化效果要优于 EAKF (紫线), 但当深度在 100~300 m 时, EAKF 同化效果要优于 LPF 滤波器。从前面的讨论我们很容易理解这个结果, 这是因为 LPF 需要启用垂向局地化来避免深层的错误更新, 而 EAKF 在没有垂向局地化的条件下可以通过相关系数(即使很小)来有效更新深层变量。如果 EAKF 也使用  $\nu=1\ 000$  m/rad 的局地化, 其在 100 m 以下的同化效果和 LPF 是一样的(红线)。对于盐度变量, LPF 的同化效果整体要略优于 EAKF, 并且在浅层的效果更加明显, 这从一方面也说明温度和盐度的关系具有

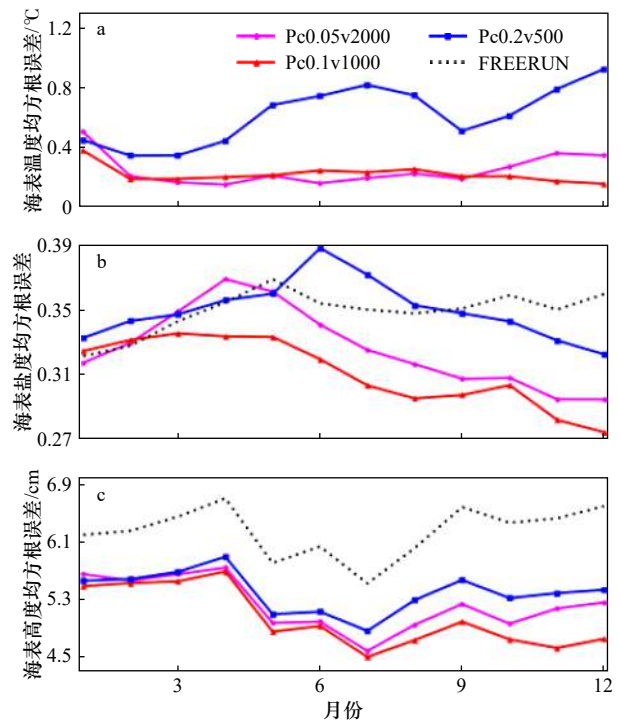


图7 不同局地化参数 LPF 实验中区域平均  
(60°S~60°N, 环地球)均方根误差时间序列

Fig. 7 Regional mean (60°S~60°N, ring the earth) root mean square error time series in the experiments of LPF with different local parameters

一定的非线性, 因而粒子滤波器的表现更好。

图 9 为最优局地化参数的 EAKF 和 LPF 实验的 RMSE 之差在海表的空间分布, 其中差异由以下公式得到

$$\text{Difference} = \text{RMSE}_{\text{EAKF}} - \text{RMSE}_{\text{LPF}}. \quad (13)$$

由定义, 若它们之差大于 0, 则表示 LPF 的同化效果优于 EAKF; 若他们之差小于 0, 则相反。图中可以明显的看出 LPF 对于 SST 的同化效果要优于 EAKF, 除了北太平洋和赤道东太平洋的小部分区域外, LPF 的误差都相对较小。对于 SSS 和 SSH 来说, 两种滤波器同化效果互相有优劣。在一些典型海区(如印度洋、大西洋), LPF 对盐度同化略有优势。而对于赤道太平洋, LPF 对于 SSH 的同化效果更好。我们在表 3 进一步计算了不同区域的平均 RMSE 以及同化影响 (Influence of Assimilation, IOA) 并进行比较。其中 IOA 定义如下:

$$\text{IOA} = \frac{\text{RMSE}_{\text{FREERUN}} - \text{RMSE}_{\text{ASSIM}}}{\text{RMSE}_{\text{FREERUN}}} \times 100\%. \quad (14)$$

显然 LPF 的 SST 同化效果在所有海区都优于 EAKF。LPF 对 SSS 的同化效果在印度洋和大西洋的优势也非常显著。而对于 SSH 来说, EAKF 的同化效

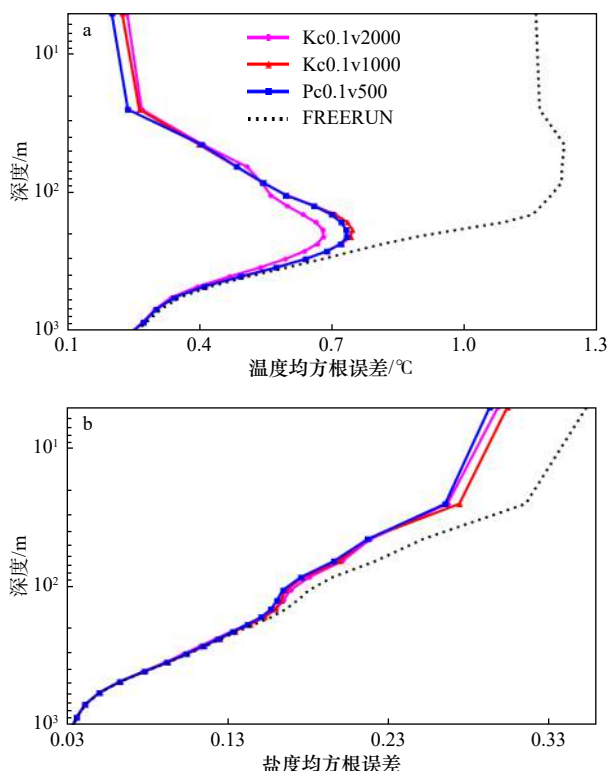


图 8 最优局地化参数 EAKF、LPF 对比实验中区域平均 ( $60^{\circ}\text{S}\sim 60^{\circ}\text{N}$ , 环地球) 垂向均方根误差  
Fig. 8 Regional mean ( $60^{\circ}\text{S}\sim 60^{\circ}\text{N}$ , ring the earth) vertical root mean square error in the experiments of EAKF and LPF with best local parameters

果全面优于 LPF。实际上,我们发现使用集合成员有限的 EAKF 同化 SST 对于 SSH 的改进效果本身就不是非常显著,其结果受到了局地化以外的很多其他因素(如平均动力地形 MDT, 集合离散度, 跨变量相关系数等)的影响,因此还需要更多的研究来进一步解释两种不同同化格式对于 SSH 的同化效果差异。

#### 4 讨论与展望

本文在大型地球系统模式 CESM 中开展了 SST 资料的同化实验。考察了局地化方案和参数对于集合调整卡尔曼滤波器和局地化粒子滤波器两种同化方法的效果的影响,并对两种方法进行了比较。

其中,集合卡尔曼滤波器使用变量之间的相关性来传递观测点的更新量。为此我们提出了一个虚假相关占比公式,对集合样本的相关和真实的相关进行比较,判别局地化对于抑制远距离虚假相关的贡献,并以此为依据选择局地化参数。在垂直方向,我们发现由于深层变量的集合离散度较小,在同化 SST 资料的时候不需要特别采用垂向局地化就能达成效果。而局地化粒子滤波器基于使用局部观测来决定不同

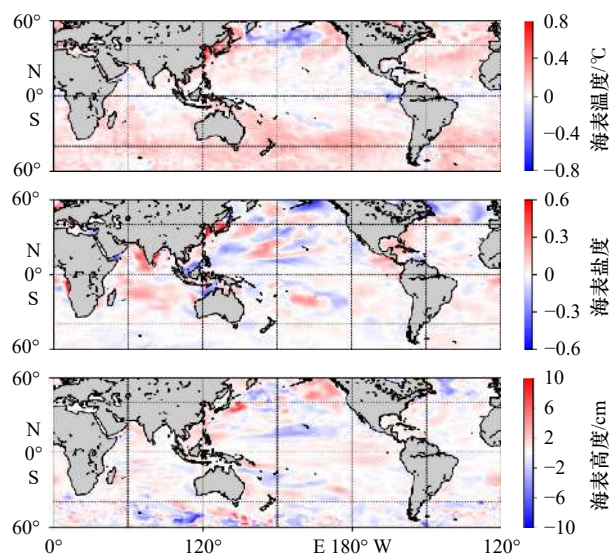


图 9 最优局地化参数 EAKF、LPF 实验均方根误差之差空间分布

Fig. 9 Spatial distribution of the difference between the root mean square error of the EAKF and LPF experiments

位置变量的权重,并利用重采样方法进行更新,因此局地化参数对于粒子滤波器的影响更大。我们必须谨慎选择局地化参数来避免粒子滤波器的退化,否则会产生负面的同化效果。此外我们也验证了垂直方向的局地化对于 LPF 是必要的。

选用最优局地化方案的两种滤波器的对比表明,LPF 在浅层的同化效果比 EAKF 好,在较深层,由于垂向局地化的影响,LPF 的效果略差于 EAKF。由于粒子滤波器在理论上不含高斯分布假设,比较适用于非线性的系统和误差,在实验中 LPF 对于盐度的变量的同化优势也比较明显。

综上,本文验证了粒子滤波器在全球耦合气候模式中的可行性,强调了有效的局地化策略对于避免滤波退化 and 改进同化效果方面的重要性。通过与集合卡尔曼滤波器的同化效果对比,我们进一步揭示了局地化粒子滤波器这一种新兴算法在未来具有非常大的潜力被应用于复杂模式的业务化同化系统中。

当前的实验仅考虑了同化 SST 资料过程中的局地化方案,对于其他的观测资料类型,如温盐廓线资料和海表面高度异常资料,其结论不一定相同。另一方面,当前的工作仅考虑使用 20 个集合成员进行计算,与局地化相关的结论也有一定的限制。我们将在下一阶段的工作中,进一步考虑使用不同的滤波器对于多源海洋观测资料的同化效果差别以及局地化方案的选择策略,也将进一步尝试不同集合成员数对于同化效果的影响。

表3 最优局地化参数 EAKF、LPF 不同区域平均均方根误差 (RMSE)、同化影响 (IOA) 对比

Table 3 Comparison table of mean root mean square error (RMSE) and influence of Assimilation (IOA) of EAKF and LPF in different regions

a.最优EAKF与LPF全球海洋不同区域海表温度RMSE及IOA对比表(后6个月平均)					
区域	FREERUN	EAKF		LPF	
	RMSE/°C	RMSE/°C	IOA/%	RMSE/°C	IOA/%
全球	1.161	0.237	79.59	0.202	82.60
太平洋	1.179	0.218	81.51	0.189	83.97
大西洋	1.183	0.257	78.28	0.217	81.66
印度洋	1.038	0.251	75.82	0.204	80.35
b.最优EAKF与LPF全球海洋不同区域海表盐度RMSE及IOA对比表(后6个月平均)					
区域	FREERUN	EAKF		LPF	
	RMSE	RMSE	IOA/%	RMSE	IOA/%
全球	0.353	0.298	15.58	0.293	17.00
太平洋	0.382	0.312	18.32	0.314	17.80
大西洋	0.343	0.286	16.62	0.267	22.16
印度洋	0.290	0.285	1.72	0.283	2.41
c.最优EAKF与LPF全球海洋不同区域海表高度RMSE及IOA对比表(后6个月平均)					
区域	FREERUN	EAKF		LPF	
	RMSE/cm	RMSE/cm	IOA/%	RMSE/cm	IOA/%
全球	6.25	4.35	30.40	4.73	24.32
太平洋	7.46	5.35	28.28	5.77	22.65
大西洋	4.54	2.97	34.58	3.40	25.11
印度洋	5.25	3.43	34.67	3.74	28.76

## 参考文献:

- [1] 李宏, 许建平. 资料同化技术的发展及其在海洋科学中的应用[J]. *海洋通报*, 2011, 30(4): 463-472.  
Li Hong, Xu Jianping. Development of data assimilation and its application in ocean science[J]. *Marine Science Bulletin*, 2011, 30(4): 463-472.
- [2] Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics[J]. *Journal of Geophysical Research: Oceans*, 1994, 99(C5): 10143-10162.
- [3] van Leeuwen P J. Particle filtering in geophysical systems[J]. *Monthly Weather Review*, 2009, 137(12): 4089-4114.
- [4] Anderson J L. An ensemble adjustment Kalman filter for data assimilation[J]. *Monthly Weather Review*, 2001, 129(12): 2884-2903.
- [5] Zhang S, Harrison M J, Rosati A, et al. System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies[J]. *Monthly Weather Review*, 2007, 135(10): 3541-3564.
- [6] Le Gland F, Monbet V, Tran V D. Large sample asymptotics for the ensemble Kalman filter[M]//Crisan D. *The Oxford Handbook of Nonlinear Filtering*. Oxford University Press, 2011: 598-634.
- [7] Shen Zheqi, Zhang Xiangming, Tang Youmin. Comparison and combination of EAKF and SIR-PF in the Bayesian filter framework[J]. *Acta Oceanologica Sinica*, 2016, 35(3): 69-78.
- [8] Penny S G, Miyoshi T. A local particle filter for high-dimensional geophysical systems[J]. *Nonlinear Processes in Geophysics*, 2016, 23(6): 391-405.
- [9] Farchi A, Bocquet M. Review article: Comparison of local particle filters and new implementations[J]. *Nonlinear Processes in Geophysics*, 2018, 25(4): 765-807.
- [10] Doucet A, Godsill S, Andrieu C. On sequential Monte Carlo sampling methods for Bayesian filtering[J]. *Statistics and Computing*, 2000,

- 10(3): 197–208.
- [11] Zhu Mengbin, van Leeuwen P J, Amezcua J. Implicit equal-weights particle filter[J]. *Quarterly Journal of the Royal Meteorological Society*, 2016, 142(698): 1904–1919.
- [12] Shen Zheqi, Tang Youmin. A modified ensemble Kalman particle filter for non-Gaussian systems with nonlinear measurement functions[J]. *Journal of Advances in Modeling Earth Systems*, 2015, 7(1): 50–66.
- [13] Poterjoy J. A localized particle filter for high-dimensional nonlinear systems[J]. *Monthly Weather Review*, 2016, 144(1): 59–76.
- [14] Whitaker J S, Hamill T M. Evaluating methods to account for system errors in ensemble data assimilation[J]. *Monthly Weather Review*, 2012, 140(9): 3078–3089.
- [15] Shen Zheqi, Tang Youmin, Li Xiaojing. A new formulation of vector weights in localized particle filters[J]. *Quarterly Journal of the Royal Meteorological Society*, 2017, 143(709): 3269–3278.
- [16] Anderson J, Hoar T, Raeder K, et al. The data assimilation research testbed: A community facility[J]. *Bulletin of the American Meteorological Society*, 2009, 90(9): 1283–1296.
- [17] Anderson J L. Localization and sampling error correction in ensemble Kalman filter data assimilation[J]. *Monthly Weather Review*, 2012, 140(7): 2359–2371.
- [18] Reynolds R W. Impact of Mount Pinatubo aerosols on satellite-derived sea surface temperatures[J]. *Journal of Climate*, 1993, 6(4): 768–774.
- [19] Rayner N A, Parker D E, Horton E B, et al. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century[J]. *Journal of Geophysical Research: Atmospheres*, 2003, 108(D14): 4407.
- [20] Gaspril G, Cohn S E. Construction of correlation functions in two and three dimensions[J]. *Quarterly Journal of the Royal Meteorological Society*, 1999, 125(554): 723–757.
- [21] 沈浙奇, 唐佑民, 高艳秋. 集合资料同化方法的理论框架及其在海洋资料同化的研究展望[J]. *海洋学报*, 2016, 38(3): 1–14.  
Shen Zheqi, Tang Youmin, Gao Yanqiu. The theoretical framework of the ensemble-based data assimilation method and its prospect in oceanic data assimilation[J]. *Haiyang Xuebao*, 2016, 38(3): 1–14.
- [22] Evensen G. The ensemble Kalman filter: Theoretical formulation and practical implementation[J]. *Ocean Dynamics*, 2003, 53(4): 343–367.
- [23] Hamill T M, Whitaker J S, Snyder C. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter[J]. *Monthly Weather Review*, 2001, 129(11): 2776–2790.
- [24] Frei M, Künsch H R. Bridging the ensemble Kalman and particle filters[J]. *Biometrika*, 2013, 100(4): 781–800.

## Data assimilation experiments using localized particle filter and ensemble Kalman filter with community earth system model

Zhang Yuting<sup>1</sup>, Shen Zheqi<sup>1,2,3</sup>, Wu Yanling<sup>1,2,3</sup>

(1. State Key Laboratory of Satellite Marine Environmental Dynamics, Second Institute of Oceanology, Ministry of Natural Resources, Hangzhou 310012, China; 2. Institute of Data Assimilation and Prediction, School of Oceanography, Hohai University, Nanjing 210098, China; 3. Guangdong Laboratory of Southern Ocean Science and Engineering (Zhuhai), Zhuhai 519080, China)

**Abstract:** Particle filter (PF) is a very promising nonlinear data assimilation method. However, due to the particle degeneracy problem, it has not been widely used in large geophysical models. In contrast, the ensemble Kalman filter (EnKF) and its derivative methods have been widely used in operational data assimilation systems in recent years. A newly proposed local particle filter (LPF) which employs the localization technique in particle filter, can effectively avoid the degeneracy problem with low computational costs and has great potential for practical applications. In this paper, data assimilation experiments using LPF and EnKF are conducted in a fully coupled Community earth system model. The synthetic satellite sea surface temperature data are assimilated with each method. Different impact of local parameters on each method is investigated, and the data assimilation performances of LPF and EnKF are compared. The comparison results show that the performance of LPF is more sensitive to localization parameter. With the optimal localization strategy, it is shown that LPF can be better than EnKF, and have a potential to be further improved.

**Key words:** data assimilation; localized particle filter; ensemble Kalman filter; community earth system model; localization