

大气边界层模式中随机参数的反演与不确定性分析

颜冰 黄思训 冯径

Retrieval and uncertainty analysis of stochastic parameter in atmospheric boundary layer model

Yan Bing Huang Si-Xun Feng Jing

引用信息 Citation: *Acta Physica Sinica*, 67, 199201 (2018) DOI: 10.7498/aps.67.20181014

在线阅读 View online: <http://dx.doi.org/10.7498/aps.67.20181014>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn/CN/Y2018/V67/I19>

您可能感兴趣的其他文章

Articles you may be interested in

[随机平衡设计傅里叶振幅敏感性分析方法和拓展傅里叶振幅敏感性分析方法在陆面过程模式敏感性分析中的应用探索](#)

Applications of random balance design Fourier amplitude sensitivity test and extended Fourier amplitude sensitivity test in the parameter sensitivity analysis of land surface process model

物理学报.2015, 64(6): 069201 <http://dx.doi.org/10.7498/aps.64.069201>

[黄土高原地表能量平衡分量年际变化及其对气候波动的响应](#)

Plateau surface energy balance components and interannual variability in response to climate fluctuations

物理学报.2014, 63(11): 119201 <http://dx.doi.org/10.7498/aps.63.119201>

[半干旱草原下垫面动量和感热总体输送系数参数化关系研究](#)

Bulk transfer coefficients of momentum and sensible heat over semiarid grassland surface and their parameterization scheme

物理学报.2013, 62(9): 099202 <http://dx.doi.org/10.7498/aps.62.099202>

[梯度计算的集合变分方案及其在大气 Ekman 层湍流系数反演中的应用](#)

Gradient calculation based ensemble variational method and its application to the inversion of the turbulent coefficient in atmospheric Ekman layer

物理学报.2013, 62(4): 049201 <http://dx.doi.org/10.7498/aps.62.049201>

[GPS/LEO 掩星正则变换反演方法](#)

Canonical transform method for GPS/LEO radio occultation

物理学报.2013, 62(7): 079201 <http://dx.doi.org/10.7498/aps.62.079201>

大气边界层模式中随机参数的反演与不确定性分析*

颜冰¹⁾ 黄思训^{1)2)†} 冯径¹⁾

1) (国防科技大学气象海洋学院, 南京 211101)

2) (上海财经大学, 计算科学与金融数据研究中心, 上海 200433)

(2018年5月24日收到; 2018年7月10日收到修改稿)

在大气边界层气象中湍流黏性系数是一个很重要的参数, 通过直接观测往往无法得到其准确值, 仅能通过间接观测获得大致范围. 本文选用随机广义 Ekman 动力近似模式中的湍流黏性系数进行反演研究与不确定性分析. 首先利用风速观测数据, 并采用基于混沌多项式的集合 Kalman 滤波方法对系数进行反演, 降低其不确定性, 缩小可能取值的范围, 该方法的核心思想是将集合 Kalman 滤波方法中求解模式不确定性传播的方法由蒙特卡罗法改为混沌多项式展开, 从而避免大规模采样带来的计算资源耗费. 然后进行数值实验, 结果表明该方法能够有效且快速地求解出湍流黏性的后验概率分布, 从而达到降低系数不确定性的目的. 根据系数的先验分布计算出风速的先验分布, 从而找到风速不确定性大的区域, 且揭示了在不确定性大的区域内的观测数据进行系数反演可得到十分明显的效果, 这对于观测点位置的选择提供了重要的指导.

关键词: 混沌多项式展开, 集合 Kalman 滤波, 随机广义 Ekman 动量近似模式, 不确定性

PACS: 92.60.Fm, 02.30.Zz, 05.10.Gg

DOI: 10.7498/aps.67.20181014

1 引言

行星大气边界层的研究对于气候模式和数值预报模式的发展很重要, 边界层模式的选择非常关键^[1]. 完整的边界层模式过于复杂不利于计算与分析, 简化的 Ekman 模式过于简单无法描述边界层的非线性特征. Tan 和 Wang^[2] 提出的广义 Ekman 动量近似模式既保留了惯性项的作用, 又考虑了湍流黏性系数随高度的变化, 是一个研究大气边界层实用且有效的模式. 在该模式中, 一个非常重要的参数就是湍流黏性系数, 它是反映湍流效应的参数^[3]. 通常, 直接测量出准确的湍流黏性系数较为困难, 所以人们希望得到不确定性较小的湍流黏性系数.

根据湍流黏性系数的分布计算出风速的分布, 叫做“正向不确定量化”或者“预报模式”^[4]. 实际上湍流黏性系数是确定的, 但是由于观测误差等原因无法准确获知其准确值, 仅能得到大致的分布, 所以必须收集更多的信息来缩小分布的范围. 通过风速的观测反演降低湍流黏性系数的不确定性, 叫做“逆向不确定量化”或者“逆模式”^[5]. 不确定性问题的研究比确定性问题困难, 同时, 因为反问题是不适定问题, 这意味着解可能不存在、解存在但是多解或解不稳定. 反问题的研究比正问题的研究困难很多^[6], 所以逆向不确定问题的研究更加困难, 而且由于观测数据存在着不可避免的误差, 观测中包含的信息通常不足以确定所有的不确定参数, 所以参数的不确定性不能被消除, 只能降低.

反问题一般根据贝叶斯理论来求解, 利用给定

* 国家自然科学基金 (批准号: 91730304, 41575026, 61371119) 资助的课题.

† 通信作者. E-mail: huangxsp@163.com

的参数先验分布结合模式输出量计算得到参数的后验分布. 但是它在大部分情况下都不能被解析地表达, 所以必须通过采样的方法来表达后验分布的不确定性, 常用的方法有马尔可夫链蒙特卡罗抽样方法 (Markov chain Monte Carlo, MCMC), 基于 Metropolis-Hastings 算法 [7], Gibbs 采样器 [8], 粒子滤波 [9]. 但是这些采样的方法需要大量的样本, 每一个样本均需要进行一次模式的计算, 所以占用了大量的计算资源.

一种常用的能够解析计算后验分布的方法叫做 Kalman 滤波 [10], 它的要求是模式必须是线性的, 而且先验分布与观测误差分布必须是正态分布. 这种情况下模式参数的后验分布同样是正态分布, 它的特征能够用前两阶矩即均值与方差来表征. 但是实际大部分大气海洋数值模型是非线性的, 而且观测模型也是非线性的. 对于非线性系统, 可以用 Kalman 滤波得到一个较为准确的近似解, 其难点在于在耗费较少计算资源的情况下怎样由参数的分布计算得到模式输出的分布. 扩展 Kalman 滤波方法 (extended Kalman filter, EKF) 是 Kalman 滤波的近似, 将非线性模式在分析值处进行 Taylor 展开到一阶, 略去高阶项, 再进行 Kalman 滤波过程 [11]. 但是在计算过程中由于非线性会导致协方差矩阵的不稳定性, 而且计算量大. 集合 Kalman 滤波方法 (ensemble Kalman filter, EnKF) 首先利用蒙特卡罗法来求解不确定性在非线性系统中的传播 [12,13], 然后计算得到 Kalman 增益矩阵. Evensen [13] 最早把 EnKF 用到数据同化中, 他利用概率论上一些理论与简单数值实验说明该方法的可行性. 由于 EnKF 便于实施, 故在很多领域都得到了广泛的应用. 但是由于 EnKF 基于蒙特卡罗方法, 所以其准确性就与样本的大小密切相关, 而大模式样本需要非常多的计算资源. Kalman 滤波类的方法通常是采用递归的方式来进行, 对于每个数据序列上的数据均进行一次滤波计算, 这样前一次的后验分布作为后一次的先验分布, 这种过程被称为“顺序数据同化”.

混沌多项式展开 (polynomial chaos expansion, PCE) 可以作为蒙特卡罗法的代替, 其本质就是利用一组正交随机多项式展开来表达不确定性 [14,15]. 根据 Askey 方案 [16], 当分布已知的情况下, 最优的正交基函数就是确定的, 所要求解的就是待定的多项式系数. 当系数求解出来后, 就会很快求出随

机变量的统计矩 [17]. 所以混沌多项式展开方法的关键就是求解多项式系数. 有很多方法可以求解此系数, 例如随机 Galerkin 方法 [15], 回归法 [18], 概率配点法 [19]. 本文中利用回归法来求解系数, 它是一种非浸入式方法, 不需要对原模式计算代码进行更改, 只需要进行少量的采样, 然后求解方程组即可求得系数. 该方法易于理解, 便于实施, 而且在消耗较少资源的情况下能达到较高的精度. 模式输入与输出变量采用混沌多项式表达, 利用基函数的正交性求出输出变量的统计矩, 避免了使用蒙特卡罗法, 能够节省计算资源. 这种方法就被称为基于混沌多项式的集合 Kalman 滤波方法 (polynomial chaos-ensemble Kalman filter, PC-EnKF) [20].

本文试图利用 PC-EnKF 求解观测值为风速的情况下湍流黏性系数的后验分布, 以降低系数不确定性和提供更可靠的范围. 而且, 还可以根据先验的风速分布确定出有效观测的区域, 为观测点位置的选择提供强有力的指导. 由于实际观测数据难以获得, 所以本文采用数值实验的方式验证该方法的有效性和可靠性. 本文第 2 部分介绍随机广义 Ekman 动量近似模式以及预报模式的求解方法; 第 3 部分介绍逆模式以及 Kalman 滤波类算法; 第 4 部分介绍数值实验及其分析; 第 5 部分给出总结与讨论.

2 随机广义 Ekman 动量近似模式

2.1 模 式

本文采用 Tan 和 Wang [2] 提出的广义 Ekman 动量近似模式 (general Ekman momentum approximation model, GEM), 该模式简单且同时保留了系统的非线性特征. 将模式中湍流黏性系数由确定性的参数变为符合某种分布的不确定性参数, 此时该模式演变成广义 Ekman 动量近似的随机边界层模式 (stochastic general Ekman momentum approximation model, SGEM). 模型如下:

$$\begin{cases} \frac{\partial}{\partial z} \left(K(z, \xi) \frac{\partial u}{\partial z} \right) + a_1 u + b_1 v = c_1, \\ \frac{\partial}{\partial z} \left(K(z, \xi) \frac{\partial v}{\partial z} \right) + a_2 u + b_2 v = c_2, \end{cases} \quad (1)$$

其中湍流黏性系数 K 是随高度 z 变化的随机变量, ξ 代表的是随机变量, (u, v) 代表 (x, y) 方向上的风速分量;

$$\begin{cases} a_1 = -\frac{\partial u_e}{\partial x}, & b_1 = f - \frac{\partial u_e}{\partial y}, \\ c_1 = f v_g + \frac{\partial u_e}{\partial t}, \\ a_2 = -\left(f + \frac{\partial v_e}{\partial x}\right), & b_2 = -\frac{\partial v_e}{\partial y}, \\ c_2 = -f u_g + \frac{\partial v_e}{\partial t}. \end{cases} \quad (2)$$

(2) 式中, f 代表 Coriolis 参数, 在本文中假定为常数; (u_e, v_e) 为经典 Ekman 模式的解, (u_g, v_g) 为地转风. 方程 (1) 的边界条件为: $u|_{z=0} = 0, v|_{z=0} = 0; u|_{z=H_b} = u_T, v|_{z=H_b} = v_T$. 其中 H_b 代表边界层的厚度, 在本文中假设 $H_b = 1500$ m; (u_T, v_T) 代表边界层顶部的风速, 可由地转动量近似模型求得. 本文在空间、时间方向上将方程 (1) 离散化, 然后直接进行数值计算.

2.2 SGEM 正模型求解

对于此随机模式的解法, 本文拟采用混沌多项式展开的方法. 根据 Askey 方案 [16], 若湍流黏性系数 K 的分布已知, 其对应的最优混沌多项式基函数 Ψ 也就确定下来了. 然后将输入随机变量 K 与输出随机量 (u, v) 在此基函数上展开, 不同的是 K 的系数是已知的, 而 (u, v) 的系数是需要求解的. 所以问题由求解 (u, v) 的分布变为求解其多项式展开的系数. 本文中采用回归法求解展开系数, 此方法简单有效, 易于理解, 在使用较少计算资源的情况下也能有不错的效果. 主要的思想就是对随机变量采样 (样本数应该大于两倍的混沌多项式的项数) [17], 然后计算得到解的样本集和多项式基函数的样本集, 最后求解超定方程组, 结果为多项式展开的系数. 将 SGEM 模式记为 $U = g(K)$, 输入随机变量 K 为湍流黏性系数, 输出随机变量 $U = (u, v)$, 其展开形式为 $U(\xi) \approx \sum_{i=0}^P U_i(z) \Psi_i(\xi)$. $\Psi_i(\xi)$ 为多项式基函数, $\xi = [\xi_1, \xi_2, \dots, \xi_d]^T$; $P+1$ 为多项式的个数, 它与随机变量的维数 d 和截断的阶数 r 有关, 即 $P+1 = (d+r)!/(d!r!)$ [4]. 求解流程如图 1 所示.

3 逆模式及 Kalman 滤波类算法

若模式参数 K 是未知的, 但我们可以根据专家猜值或者一些直接的但精度有限的观测得知其

先验的概率密度函数 $p(K)$, 风速 U 可以通过观测得到. 为了降低参数的不确定性, 得到一个更加接近真值的估计值, 可以利用带误差的观测 U^* . $U^* = U + e_U$, 误差 e_U 符合另一概率密度函数. 求解逆模式的目的就是估计出参数 K 在已知观测 U^* 的情况下的后验概率密度 $p(K|U^*)$.

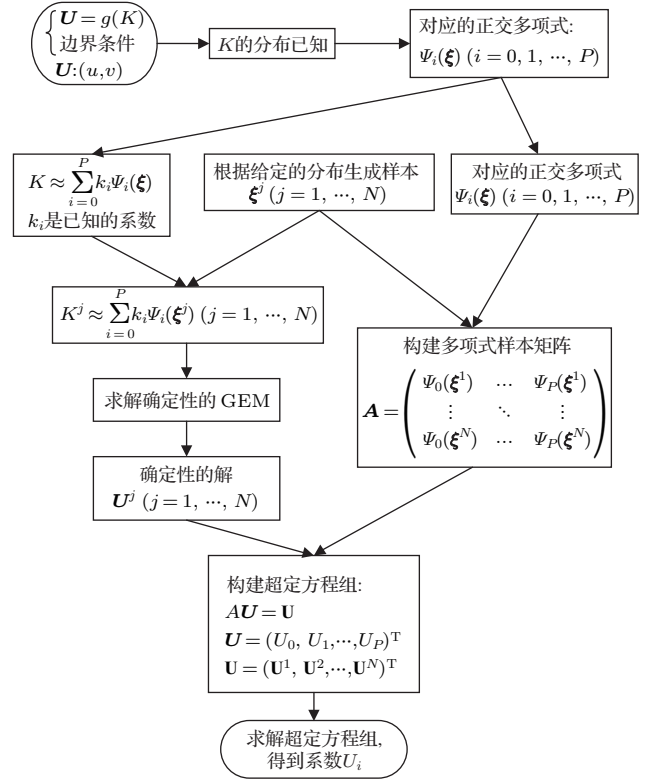


图 1 SGEM 预报模式求解流程图

Fig. 1. The flow chart for solving the SGEM.

3.1 Kalman 滤波

Kalman 滤波能给出准确的后验概率密度函数, 但它有着较高的限制条件: 1) 模式必须是线性的; 2) 参数的先验密度函数 $p(K)$ 与误差 e_U 均服从正态分布. 此时, 参数的后验概率密度函数同样服从正态分布, 其特征可以用均值 $\mu_{K|U^*}$ 与协方差 $C_{KK|U^*}$ 来表达. 其求解过程如下.

第一步 预测, 利用参数 K 先验分布通过线性随机模式求出输出量 U 的分布, 又叫做预报模式求解过程. 由于模式是线性的且均服从正态分布, 所以很容易得到 μ_U, C_{UU}, C_{UK} . μ_U 代表 U 的先验/预测均值, C_{UK} 是 U 和 K 的先验协方差, C_{UU} 是向量 U 的先验/预测协方差. 然后计算得到 Kalman 增益矩阵 G . $G = C_{KU}(C_{UU} + R)^{-1}$, 式

中, $C_{KU} = (C_{UK})^T$, R 是误差 e_U 的协方差矩阵, 通常是预先给定的.

第二步 分析. 通过观测值来调整参数 K 的分布, 求出后验分布, 具体来说就是求出后验均值与后验方差, 可以通过下列公式得到:

$$\begin{aligned} \mu_{K|U^*} &= \mu_K + G(U^* - \mu_U), \\ C_{KK|U^*} &= C_{KK} - GC_{UK}. \end{aligned} \quad (3)$$

其中 μ_K 和 C_{KK} 分别代表参数 K 的先验的均值和协方差.

通常, 观测数据是一个序列, 所以 Kalman 滤波也会通过递归进行, 每一个观测数据均会进行一次 Kalman 滤波过程. 也就是说, 根据前一次观测数据而求得后验分布会作为后一次 Kalman 滤波的先验分布.

3.2 EnKF

Kalman 滤波方法只适用于线性模式, 而大气海洋科学中大部分模式都是非线性的. 对于非线性不太强的模式, 我们可以用 EnKF 方法给出一个近似解, 得出的后验分布不一定符合正态分布, 但也是单峰型的, 可以用前两阶矩 (均值和方差) 来表示其特征. 对于非线性模式, 通常采用蒙特卡罗法来求解输出变量的先验/预测分布, 这就是 EnKF. 首先根据输入参数的先验分布采样, 得到一组参数的样本集, 然后对每一个样本进行确定预报模式的计算, 得到输出变量的样本集.

$$U_i = g(K_i) \quad i = 1, 2, \dots, N. \quad (4)$$

在输出变量样本集的基础上可以计算得到均值、协方差以及与输入参数的协方差.

$$\begin{aligned} \hat{\mu}_U &= \frac{1}{N} \sum_{i=1}^N U_i, \\ \hat{C}_{UU} &= \frac{1}{N-1} \sum_{i=1}^N (U_i - \hat{\mu}_U)(U_i - \hat{\mu}_U)^T, \\ \hat{C}_{UK} &= \frac{1}{N-1} \sum_{i=1}^N (U_i - \hat{\mu}_U)(K_i - \hat{\mu}_K)^T. \end{aligned} \quad (5)$$

然后是分析步, 由于是采样的方法, 需要对每一个样本进行更新:

$$\begin{aligned} K_i^{\text{update}} &= K_i + G(U_i^* - g(K_i)) \\ i &= 1, 2, \dots, N. \end{aligned} \quad (6)$$

之后同样进行递归的过程. 可以看到在 (6) 式中需要对观测 U^* 根据给定的协方差进行扰动后采样得到 U_i^* , 这就会产生不必要的采样误差. Whitaker 和 Hamill^[21] 提出了集合均方根滤波方法 (ensemble square root filter, EnSRF), 此方法可以使用确定性的观测数据, 这样就可以避免在观测时产生采样的误差. 这种方法将参数 K 的均值和扰动分别进行更新. 均值的更新使用标准的 Kalman 增益矩阵, 而扰动的更新使用修正的增益矩阵:

$$\begin{aligned} \tilde{G} &= C_{KU}((\sqrt{C_{UU} + R})^{-1})^T \\ &\times (\sqrt{C_{UU} + R} + \sqrt{R})^{-1}. \end{aligned} \quad (7)$$

在本文中, 我们 EnSRF 方法来进行反演的过程.

3.3 PC-EnKF

EnKF 使用蒙特卡罗法进行输出变量的不确定量化, 而蒙特卡罗法的收敛速度慢且需要足够大的样本来保证结果的准确性, 这就造成了巨大的计算负担, 特别是对于大型非线性模式. 为了在不显著增加计算量的情况下增加逆模式的求解精度, 本文采用 EnKF 的另外一个变种: 基于混沌多项式的集合 Kalman 滤波方法 (PC-EnKF). 此方法类似于 EnKF, 只是在预报模式中采用混沌多项式展开来代替蒙特卡罗法进行不确定性量化. 由于混沌多项式展开的关键在于多项式的系数, 所以将 EnKF 中样本的更新改变为多项式系数的更新. 具体实施步骤如下.

第一步 预测. 首先根据参数 K 的先验分布找出其最优的多项式基函数 $\Psi_i(\xi)$, $\xi = [\xi_1, \xi_2, \dots, \xi_d]^T$ 是由一组给定的概率密度函数的相互独立的随机变量组成的, 例如标准正态分布、均匀分布等. 然后在此基函数上展开得到: $K(\xi) \approx \sum_{i=0}^P k_i \Psi_i(\xi)$. 由于其先验分布是已知的, 所以系数 k_i 也是已知的. 然后利用上文说到的预报模式求解方法, 求出 $U(\xi) \approx \sum_{i=0}^P U_i \Psi_i(\xi)$, 其中 U_i 是多项式展开系数. 在本文中, 假设参数服从正态分布, 其对应的最优多项式为 Hermite 多项式^[4,22]. 混沌多项式都有着标准正交的特性,

$E(\Psi_i(\xi)\Psi_j(\xi)) = \delta_{ij}$, δ_{ij} 是 Kronecker delta 函数.

$$\delta_{ij} = \begin{cases} 0 & i \neq j, \\ 1 & i = j. \end{cases} \quad (8)$$

第一项基函数 $\Psi_0(\xi) = 1$ 代表均值项, 之后的基函数均代表扰动项, 各个扰动项的均值为零. 根据这些性质, 我们可以很快得到其统计特征:

$$\begin{aligned} \mu_U &= \sum_{i=0}^P U_i E(\Psi_i(\xi)) = U_0, \\ C_{UU} &= E((U - \mu_U)(U - \mu_U)^T) \\ &= E\left(\left(\sum_{i=1}^P U_i \Psi_i(\xi)\right)\left(\sum_{j=1}^P U_j \Psi_j(\xi)\right)^T\right) \end{aligned}$$

$$\begin{aligned} &= \sum_{i=1}^P U_i U_i^T, \\ C_{UK} &= E((U - \mu_U)(K - \mu_K)^T) \\ &= E\left(\left(\sum_{i=1}^P U_i \Psi_i(\xi)\right)\left(\sum_{j=1}^P k_j \Psi_j(\xi)\right)^T\right) \\ &= \sum_{i=1}^P U_i k_i^T. \end{aligned} \quad (9)$$

结合误差 e_U 的协方差 R , 而且注意到 $C_{KU} = (C_{UK})^T$, 可以计算出 Kalman 增益和修正的增益.

第二步 分析. 将观测数据同化进来, 由此调整得到参数 K 的后验分布的多项式展开系数. 本文中我们采用 EnSRF 策略进行更新, 所以分为均值更新与扰动更新:

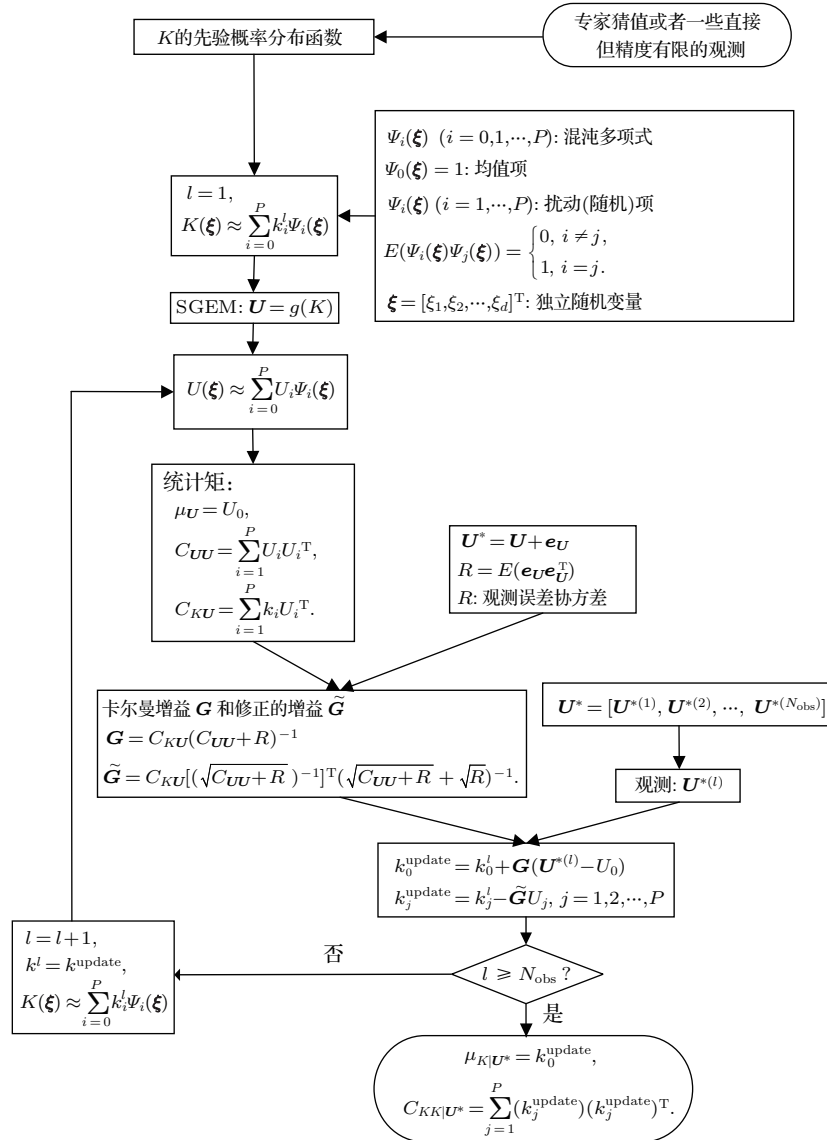


图2 利用 PC-EnKF 减少参数不确定性求解流程图
Fig. 2. The flow chart for reducing the uncertainty by PC-EnKF.

$$k_0^{\text{update}} = k_0 + \mathbf{G}(\mathbf{U}^* - U_0),$$

$$k_j^{\text{update}} = k_j - \tilde{\mathbf{G}}U_j \quad j = 1, 2, \dots, P. \quad (10)$$

如有多次的观测, 可利用递归的方式进行计算, 根据前一次观测数据而求得后验分布会作为后一次 Kalman 滤波的先验分布.

后验分布的均值与方差容易求得:

$$\mu_{K|U^*} = k_0^{\text{update}},$$

$$C_{KK|U^*} = \sum_{j=1}^P (k_j^{\text{update}})(k_j^{\text{update}})^T. \quad (11)$$

其流程如图 2 所示.

4 数值实验

为了验证 PC-EnKF 在反演 SGEM 中湍流黏性系数 K 不确定性时的有效性. 本文选取一类典型的气旋性地转切变流进行实验, 其表达式为 [2]:

$$u_g = u_0 - \alpha y, \quad v_g = 0, \quad (12)$$

其中 $u_0 = 20 \text{ m}\cdot\text{s}^{-1}$; α 为常数, 取 $\alpha = 0.4f$, 而 $f = 10^{-4} \text{ s}^{-1}$. 由于实际观测数据较难获得, 本文决定采用模拟观测数据进行数值实验, 在后续的研究中再使用实际观测数据.

我们假设反演两种类型下的湍流黏性系数; 第一种, 湍流黏性系数是常数, 不随高度变化; 第二种, 湍流黏性系数是随高度变化的量.

4.1 实验一

为了避免湍流黏性系数出现负值的情况, 我们假设 $\pi = \log(K)$, π 服从正态分布. 根据湍流黏性系数取值的大致范围确定先验分布 $\pi \sim N(2, 0.4^2)$, 然后按照先验分布随机采样得到一个样本作为参考的真值参数, 然后根据确定性的 GEM 模式计算得到相应的风速, 在此基础上叠加符合正态分布的观测误差 $R \sim N(0, 0.2^2)$, N 代表正态分布, 各点的误差假定是相互独立的, 由此生成模拟观测数据. 观测数据随高度变化 $\mathbf{U}^{*(l)} = \mathbf{U}^*(z_l), l = 1, 2, \dots, N_H$, N_H 为垂直方向格点数, 在每个点上均进行一次 Kalman 滤波, 这样就形成一个递归的过程. 我们将垂直方向分成 51 个格点, 这样就会有 49 个 Kalman 滤波过程 (因为要去除两个固定边界条件). 按照图 2 所示的流程图进行反演, 为了直观地显示递归过程中的不确定性降低, 每进行十六次循环, 绘制一次模式输入量 π 和对应的模式输出量 \mathbf{U} 的后验分布的均值和方差图形.

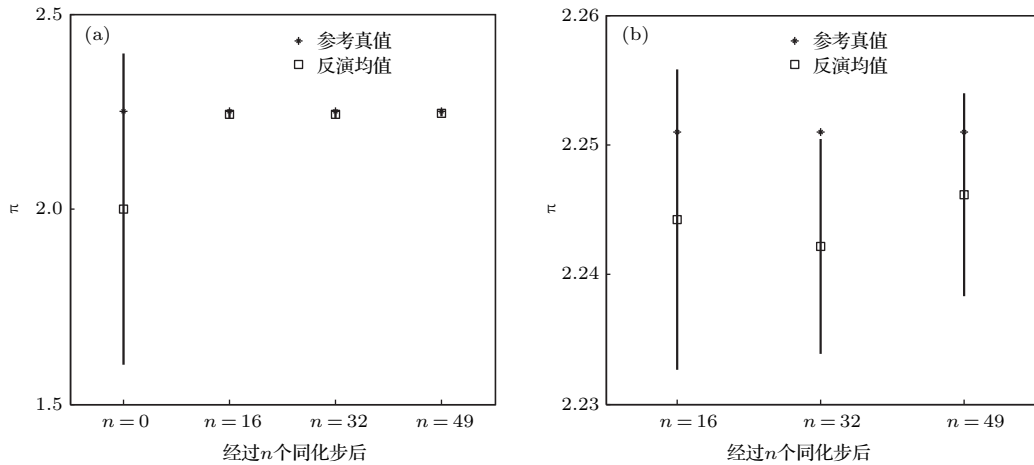


图 3 模式参数 π 的置信区间 (均值 \pm 标准差) 横坐标代表经过的迭代步数, 0 步代表并未经过同化的先验分布, 其余为经过同化的后验分布; 纵坐标代表 π 的取值; 星号代表 π 的参考真值, 方块代表 π 的后验分布的均值; 注意: 在 (a) 中, 由于后验分布的不确定性很小, 而 (a) 的纵坐标范围过大不能显示其中的细节, 故将后验分布的置信区间放大放在 (b) 中

Fig. 3. The estimated confidence interval of model parameter π (mean \pm standard deviation). The horizontal coordinate is the iterative number, where $n = 0$ represents the prior distribution without assimilation, and the others are the posterior distribution after assimilation. The vertical coordinate is the value of π . The asterisk represents the reference value, the square represents the mean of posterior distribution. Note that the uncertainty of posterior distribution is small, while the range of vertical coordinate of Fig. (a) is too large to display details of posterior distribution. So, these details are shown in Fig. (b).

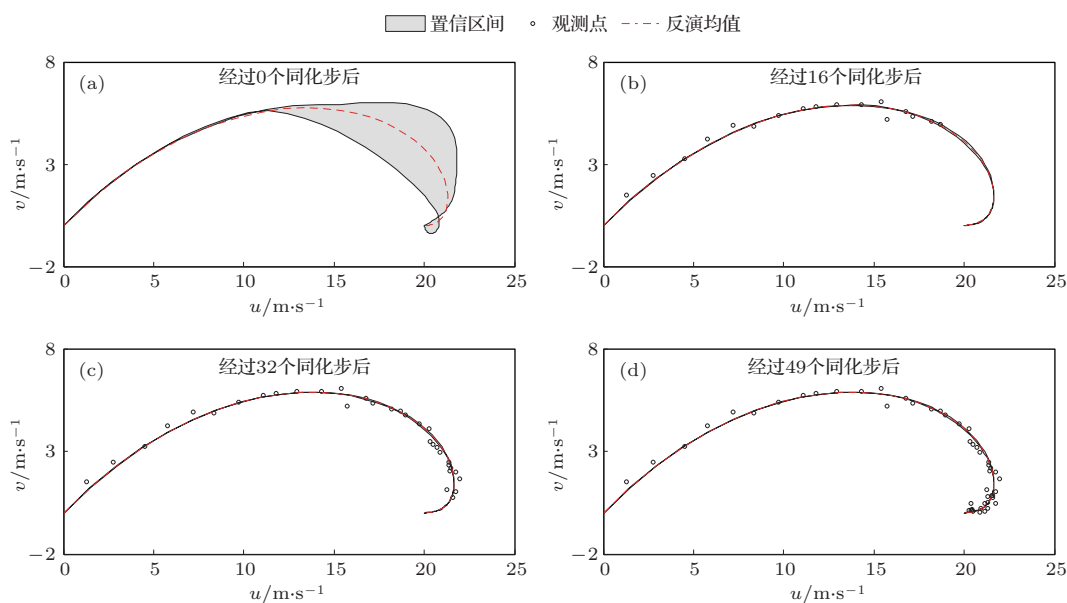


图4 风速 Ekman 螺旋线的观测数据以及风速分布示意图 灰色区域是分布的置信区间 (均值 \pm 标准差), 虚线代表 Ekman 螺旋线分布的均值, 空心圆点代表观测点, 也是数据同化点

Fig. 4. The wind observations of Ekman spirals and the distribution of wind. The grey shaded area is the confidence interval of wind speed (mean \pm standard deviation), the dotted line is the mean of wind distribution, the circular point is the observation point and also the data assimilation point.

从图 3(a) 可以看出, PC-EnKF 能有效利用风速的观测信息进行反演, 显著降低参数的不确定性. 从图 3(b) 可以看出, 随着同化过程的进行, 参数不确定性在逐步减小且其后验分布的均值越来越靠近参考真值.

图 4(a) 中 Ekman 螺旋线先验分布的不确定性较大, 在经过同化步后 (图 4(bd)), 风速的不确定性明显降低, 而且与观测数据的拟合程度较高, 说明了 PC-EnKF 在降低参数不确定性方面是有效的. 经过 16 个同化步之后预测的 Ekman 螺旋线与观测已经比较接近, 也对应着图 3(b) 中经过 16 步的同化过程后模式参数的后验分布小的不确定性.

本文在风速先验分布中, 不确定性较大的区域与不确定性小的区域各选取了 10 个观测点, 在此基础上分别进行反演, 比较其反演结果.

根据图 5 所示的观测点上的观测数据分别进行反演, 可以分别求得两种观测下的参数后验分布.

从图 6 中可以明显地看出, 相比于 b 区域, a 区域观测的反演效果更好, 均值更靠近参考真值且标准差更小. 因为反演属于反问题, 反问题的一个重要特点就是不稳定, 不稳定即模式输出值小的变化就会引起参数的反演结果很大的误差, 对应到不确定性分析中就是模式参数的先验分布不确定性较大, 而对应的模式输出的先验分布不确定性较小.

由一个小范围的分布去反演一个大范围的分布就会引起反演结果的不稳定. 由此我们认为在模式参数先验分布不变的情况下, 选取输出值先验分布不确定性较大的区域的观测进行反演会更加有效率, 图 6 的结果也证明了我们的观点. 这个发现对于风速观测点的位置的选择具有指导意义.

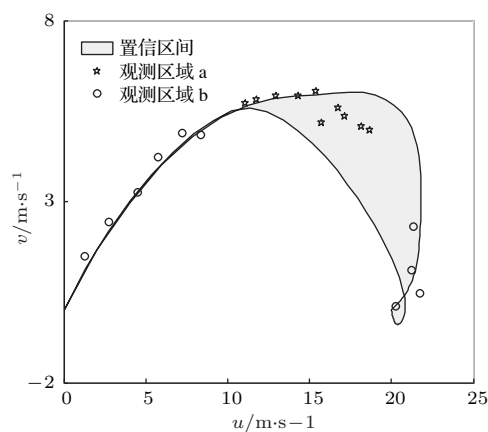


图5 选取的观测点示意图 灰色区域是风速先验分布的置信区间 (均值 \pm 标准差), 星号代表观测区域 a (不确定性大的区域) 的观测, 圆圈代表观测区域 b (不确定性小的区域) 的观测

Fig. 5. The selected observation points. The grey shaded area is the confidence interval of wind speed (mean \pm standard deviation). The star represents observation in area a (heavy uncertainty), the circle represents observation in area b (small uncertainty).

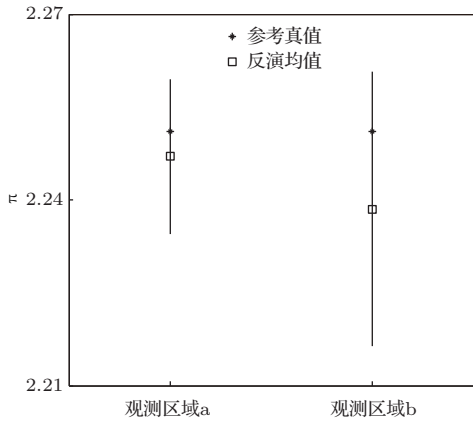


图6 由不同区域的观测反演的结果 横坐标代表不同的区域, 纵坐标的模式参数 π 的取值, 星号代表 π 的参考真值, 方块代表 π 的后验分布的均值

Fig. 6. The result retrieved by different observations. The horizontal coordinate represents the area, the vertical coordinate is the value of π . The asterisk represents the reference, the square represents the mean of posterior distribution.

4.2 实验二

考虑湍流黏性系数一般是随高度变化的量, 我们假设每个点的湍流黏性系数都是一个随机变量. 因此系统成为高维的, 混沌多项式的项数会急剧增加, 计算量也会急剧增加. 在本文

中, 我们假设垂直方向上格点的湍流黏性系数是存在相关性. $K(z, \xi)$ 是模式参数, 假设其先验分布为对数正态分布, 假设 $A = \log(K)$, A 符合正态分布, $E(A) = 2.08$, $Cov(A(z_1), A(z_2)) = \exp(-|z_1 - z_2|/c)$, $a = 25, c = 0.004$. 我们在此基础上随机生成参考真值, 然后经过确定性模式得到模式输出值, 再加上符合观测误差协方差 $R \sim N(0, 0.2^2)$ 的误差变成了带误差的观测值, N 代表正态分布, 且假设各点的观测误差独立.

在利用 PC-EnKF 进行反演之前, 必须将模式参数 A 的先验分布表达成混沌多项式展开的形式, 利用 K-L 分解 (Karhunen-Loève expansion) [4], 由于 A 的先验均值与协方差函数已知, 可以将其近似表达为

$$A(z, \omega) = \bar{A}(z) + \sum_{j=1}^n \sqrt{\lambda_j} \phi_j(z) \xi_j, \quad (13)$$

其中, $\bar{A}(z)$ 为均值, 在本文中其值取 2.08. 特征对 $\{\lambda_j, \phi_j\}$ 可以通过求解如下积分方程求得:

$$\int_{-a}^a e^{-|z_1 - z_2|/c} \phi_j(z_2) dz_2 = \lambda_j \phi_j(z_1). \quad (14)$$

详细的步骤可以参阅文献 [4, 22].

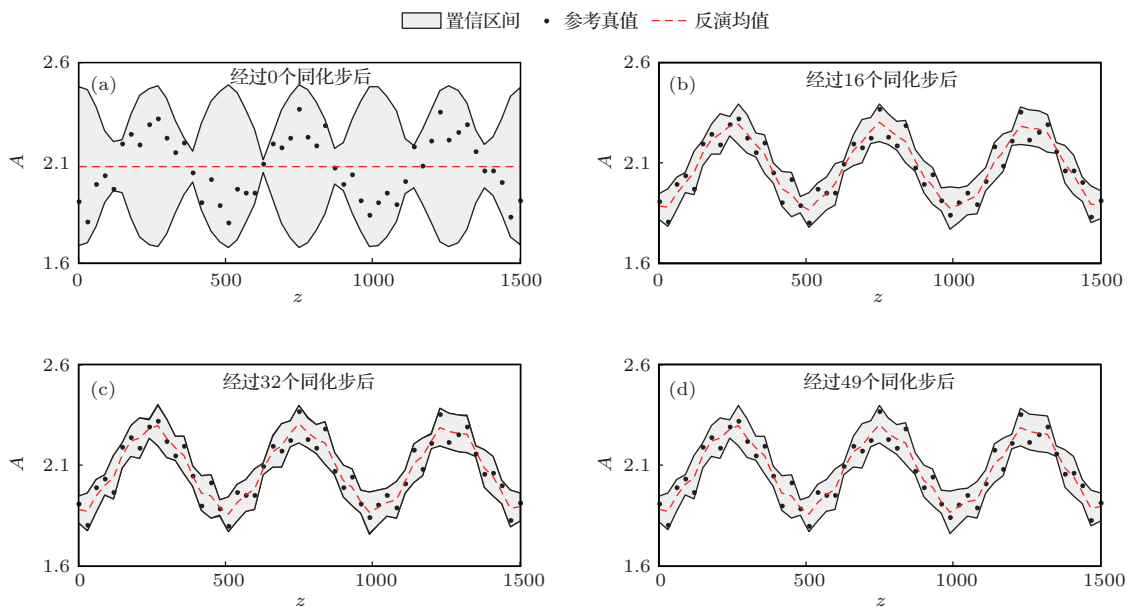


图7 模式参数 A 的分布示意 横坐标代表高度, 纵坐标是参数的取值; 灰色区域是分布的置信区间 (均值 \pm 标准差), 点代表模式参数 A 的参考真值, 虚线代表的是分布的均值

Fig. 7. The distribution of model parameter A . The horizontal coordinate represents the height, the vertical coordinate is the value of model parameter. The grey shaded area is the confidence interval of A (mean \pm standard deviation), the solid point is the reference value of A , the dotted line is the mean of distribution.

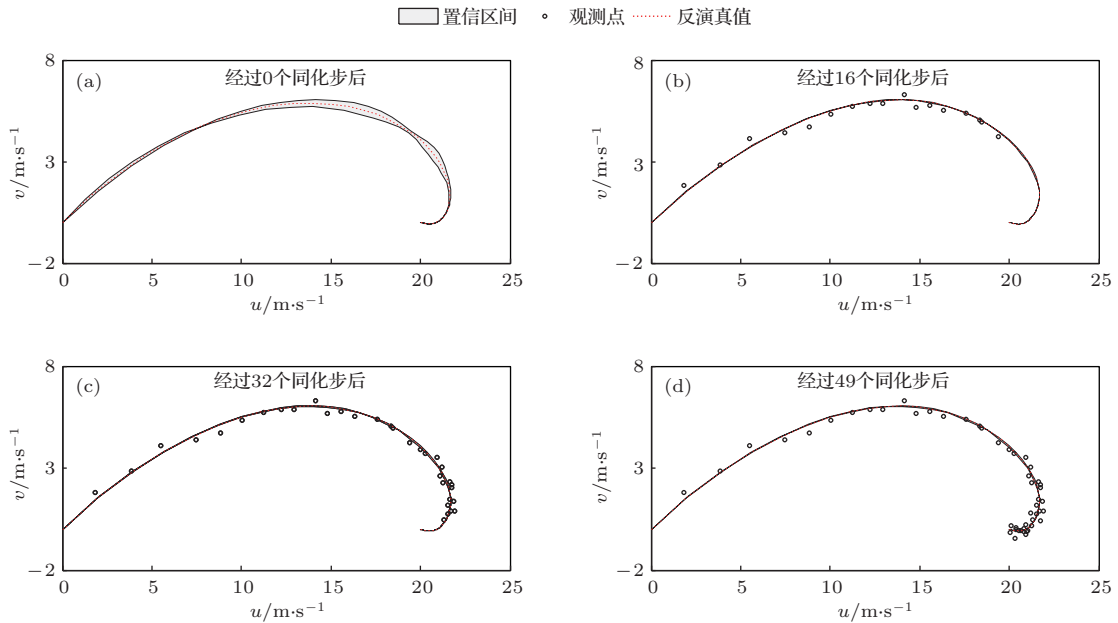


图8 风速 Ekman 螺线的观测数据以及风速分布示意图 灰色区域是分布的置信区间 (均值 \pm 标准差), 虚线代表 Ekman 螺线分布的均值, 空心圆点代表观测点, 也是数据同化点
 Fig. 8. The wind observations of Ekman spirals and the distribution of wind. The grey shaded area is the confidence interval of wind speed (mean \pm standard deviation), the dotted line is the mean of wind distribution, the circular point is the observation point and data assimilation point.

所以, 模式参数 A 的先验分布的混沌多项式展开为: $A = A_0 + \sum_{i=1}^d A_i \psi_i(\xi)$, 其中 $A_0 = \bar{A}(z)$, $A_i = \sqrt{\lambda_i} \phi_i(z)$, $\psi_i(\xi) = \xi_i, i = 1, \dots, d$, 在本文中 $\xi_i \sim N(0, 0.2^2), i = 1, \dots, d$.

在得到先验分布的多项式展开之后, 就可以实施 PC-EnKF 反演过程. 与实验一类似, 实验二同样是每隔 16 次绘制一次模式输入量 π 和对应的模式输出量 U 的后验分布的均值和方差的图形.

从图 7 (a) 可以看出参数的先验分布不确定性是相当大的, 参考真值与分布的均值相去甚远. 经过同化过程后, 不确定性有了明显的减小, 特别是均值与参考真值已经很接近了.

图 8 (a) 中 Ekman 螺线先验分布的不确定性不是很大. 在经过同化步后 (图 8 (b)–(d)), Ekman 螺线的不确定性有所降低, 与观测数据的拟合程度高, 说明了 PC-EnKF 在降低参数不确定性方面是有效的. 在实验二中我们同样选取了不确定性大与不确定性小两个区域的观测, 并分别进行反演. 观测点的位置如图 9 所示.

可以很明显地看出, 图 10 (a) 中的反演效果更好, 具有更小的不确定性, 进一步缩小了参数分布的区间. 这也佐证了我们的观点, 在不确定性大的

区域的观测数据在反演时具有更好的效果. 而且实验二中风速的先验分布 (图 9) 的不确定性明显比实验一 (图 5) 的小, 根据我们上文中的分析, 实验二中模式参数的反演的不稳定性更强, 反演难度大于实验一. 比较图 7 与图 3, 可以看出实验二的反演结果也较实验一差.

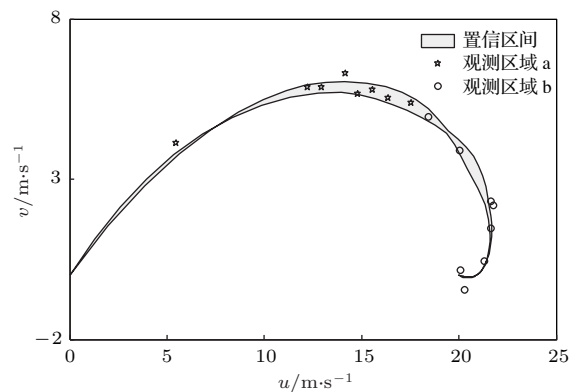


图9 选取的观测点示意图 灰色区域是风速先验分布的置信区间 (均值 \pm 标准差). 星号代表观测区域 a (不确定性大的区域) 的观测, 圆圈代表观测区域 b (不确定性小的区域) 的观测

Fig. 9. The selected observation points. The grey shaded area is the confidence interval of wind speed (mean \pm standard deviation). The star represents observation in area a (heavy uncertainty), the circle represents observation in area b (small uncertainty).

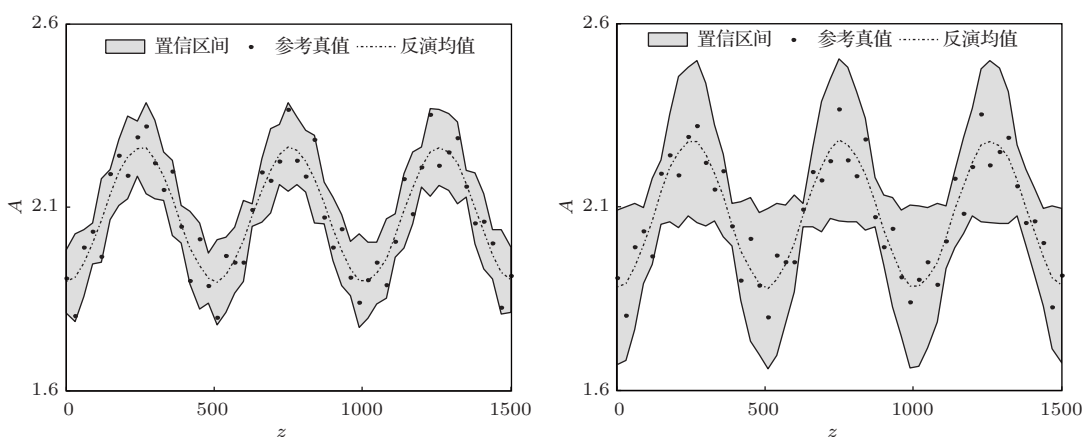


图 10 由不同区域观测反演参数 A 的结果 横坐标代表高度, 纵坐标是参数的取值, 灰色区域是分布的置信区间(均值 \pm 标准差), 点代表模式参数 A 的参考真值, 虚线代表的是分布的均值; (a) 代表由观测区域 a 中的观测反演的结果; (b) 代表由观测区域 b 中的观测反演的结果

Fig. 10. The result retrieved by different observations. The horizontal coordinate represents the height, the vertical coordinate is the value of model parameter. The grey shaded area is the confidence interval of A (mean \pm standard deviation), the solid point is the reference of A , the dotted line is the mean of distribution. (a) display the result retrieved by observations in area a, (b) display the result retrieved by observations in area b.

5 总结与讨论

本文利用 PC-EnKF 方法以及风速的观测, 对 SGEM 模式中湍流黏性系数的不确定性进行了反演, 求出了其后验概率密度函数. 采用两类不同类型的湍流黏性系数进行数值实验, 实验一中假定待反演的系数是常数, 实验二中假定是符合先验协方差函数的随高度变化的系数, 均采用 PC-EnKF 进行反演. 实验结果表明, PC-EnKF 能够很有效地求解后验概率密度函数, 混沌多项式展开能够有效进行不确定性的表达, 将 PC 与 EnKF 结合起来能够用于不确定性的反问题. 而且我们在实验中发现, 根据系数的先验分布计算出风速的先验分布, 在此基础上找到风速不确定性大的区域, 在此区域内的观测数据在反演系数时具有更好的效果, 这对于观测点的位置的选择具有指导意义.

本文利用混沌多项式展开的方法代替 EnKF 中的蒙特卡罗法, 能够很有效地进行不确定性问题的反问题. 但是由于混沌多项式展开对于高维的不确定性问题的计算量急剧增加, 所以 PC-EnKF 对于高维问题的求解还存在着一些问题, 下一步的工作中将重点关注不确定性降维方面.

参考文献

[1] Baklanov A, Grisogono B, Bornstein R, Mahrt L, Zilitinkevich S, Taylor P, Larsen S, Rotach M, Fernando H 2011 *Bull. Am. Meteorol. Soc.* **92** 123

[2] Tan Z M, Wang Y 2002 *Adv. Atmos. Sci.* **19** 266
 [3] Han Y Q, Zhong Z, Wang Y F, Du H D 2013 *Acta Phys. Sin.* **62** 049201 (in Chinese) [韩月琪, 钟中, 王云峰, 杜华栋 2013 物理学报 **62** 049201]
 [4] Xiu D 2010 *Numerical Methods for Stochastic Computations: A Spectral Method Approach* (Princeton: Princeton University Press) p126
 [5] Li W X, Lin G, Zhang D X 2014 *J. Comput. Phys.* **258** 752
 [6] Yan B, Huang S X 2014 *Chin. Phys. B* **23** 109402
 [7] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H, Teller E 1953 *J. Chem. Phys.* **21**
 [8] Geman S, Geman D 1984 *IEEE Trans. Pattern Anal. Mach. Intellig.* **20** 721
 [9] Leng H Z, Song J Q, Cao X Q, Yang J H 2012 *Acta Phys. Sin.* **61** 070501 (in Chinese) [冷洪泽, 宋君强, 曹小群, 杨锦辉 2012 物理学报 **61** 070501]
 [10] Kalman R E 1960 *J. Basic Engin. Trans.* **82** 35
 [11] Gelb A 1974 *Applied Optimal Estimation* (Cambridge: MIT Press)
 [12] Evensen G 2006 *Data Assimilation: The Ensemble Kalman Filter* (New York: Springer) p38
 [13] Evensen G 1994 *J. Geophys. Res.: Oceans* **99** 10143
 [14] Ghanem R G, Spanos P D 1992 *Stochastic Finite Element: A Spectral Approach* (New York: Springer) p214
 [15] Xiu D, Karniadakis G S 2003 *J. Comput. Phys.* **187** 137
 [16] Schoutens W 2000 *The Askey Scheme of Orthogonal Polynomials. In: Stochastic Processes and Orthogonal Polynomials* (New York: Springer) pp1–13
 [17] Sun N Z, Sun A 2015 *Model Uncertainty Quantification. In: Model Calibration and Parameter Estimation: For Environmental and Water Resource Systems* (New York: Springer) pp407–458
 [18] Isukapalli S S, Roy A, Georgopoulos P G 1998 *Risk Anal.* **18** 351
 [19] Tatang M A, Pan W, Prinn R G, McRae G J 1997 *J. Geophys. Res. -Atmos.* **102** 21925 doi: 10.1029/97jd01654

- [20] Wang Y P, Cheng Y, Zhang Z Y, Lin G 2018 *Math. Model. Nat. Phenom.* **13** doi:101051/mmnp/2018023
 [21] Whitaker J S, Hamill M 2003 *Mon. Weather. Rev.* **130** 1913

- [22] Maître O P L, Knio O M 2010 *Spectral Methods for Uncertainty Quantification* (Netherlands: Springer) p536

Retrieval and uncertainty analysis of stochastic parameter in atmospheric boundary layer model*

Yan Bing¹⁾ Huang Si-Xun^{1)2)†} Feng Jing¹⁾

1) (College of Meteorology and Oceanography, National University of Defense Technology, Nanjing 211101, China)

2) (Center for Computational Science and Finance, Shanghai University of Finance and Economics, Shanghai 200433, China)

(Received 24 May 2018; revised manuscript received 10 July 2018)

Abstract

The eddy viscosity is an important parameter in the atmospheric boundary layer meteorology, and we usually cannot determine their exact values by direct measurements, but we can only obtain an approximate range by indirect approximate method. In this paper, the eddy viscosity in the stochastic general Ekman momentum approximation model is used for the retrieval research and uncertainty analysis. The main purpose of retrieval is to reduce the uncertainty and narrow the approximate range of eddy viscosity. First, the polynomial chaos-ensemble Kalman filter and the wind observations are used for eddy viscosity retrieval and uncertainty reduction. The main idea of this method is to replace the Monte-Carlo method with polynomial chaos in the uncertainty quantification of ensemble Kalman filter, and thus avoiding the consumption of computing resources brought by massive samples. The goal of uncertainty quantification is to investigate the effect of uncertainty in the eddy viscosity on the model and to subsequently provide a reliable distribution of simulation results. Then two numerical experiments are implemented, i.e. experiment I in which the eddy viscosity is assumed to be constant, and experiment II in which the eddy viscosity is assumed to be a vertically varying random parameter. The uncertainty of eddy viscosity in experiment I is reduced quickly, at the same time the mean of eddy viscosity can converge to a reference value. The effect in experiment II is also remarkable after 16 data assimilation steps. These results show that the polynomial chaos-ensemble Kalman filter is an effective and fast method of solving the posterior distribution of eddy viscosity and reducing the uncertainty of eddy viscosity. Finally, we calculate the prior distribution of wind speed according to the prior distribution of eddy viscosity and identify the heavy uncertainty area in wind speed. The results indicate that the posterior distribution of eddy viscosity solved with wind observations in the big uncertainty area is more accurate, which provides an important guidance for selecting the location of observation points.

Keywords: polynomial chaos expansion, ensemble Kalman filter, stochastic general Ekman momentum approximation model, uncertainty

PACS: 92.60.Fm, 02.30.Zz, 05.10.Gg

DOI: 10.7498/aps.67.20181014

* Project supported by the National Natural Science Foundation of China (Grant Nos. 91730304, 41575026, 61371119).

† Corresponding author. E-mail: huangexp@163.com