

文章编号:1001-8166(2011)08-0 -07

陆面数据同化系统误差问题研究综述*

摆玉龙^{1,2}, 李新¹, 韩旭军¹

(中国科学院寒区旱区环境与工程研究所, 甘肃 兰州 730000;

2. 西北师范大学物理与电子工程学院, 甘肃 兰州 730070)

摘要:同化系统中的误差问题一直被认为是制约数据同化性能的瓶颈问题。从分析陆面数据同化系统的误差问题研究现状出发, 统一定义了同化系统的误差来源及误差表现, 简要综述了顺序同化方法及连续同化方法中的误差定义和相关理论问题。从误差估计的角度, 重点介绍了目前研究中各种误差估计的方法和面临的困难。针对误差处理方法的研究, 介绍了在集合数据同化中为减小误差常用的乘数放大法、附加放大法和松弛先验法等模型误差参数化方案, 并且介绍了在实际数据同化系统中为减小系统偏差常采用状态增广法。最后总结讨论了各种误差估计与处理方法的特点及其在陆面数据同化中的应用前景和发展方向。

关键词:数据同化; 误差估计; 误差参数化方案

中图分类号: TP79; P237 **文献标志码:** A

1 引言

地球表层系统是一个复杂的巨系统, 主要表现在地球表层的各种过程复杂的相互作用以及各种状态变量在不同的时间和空间尺度上都具有高度的异质性。数据同化方法作为一种优化融合地球观测信息和地学模型信息的重要方法论, 在集成复杂地球系统的模拟经验和融合地球观测时代的多源观测信息方面扮演着重要角色。尽管陆面数据同化系统的建立与研究取得了蓬勃的发展^[1~4], 鉴于地球表层系统的复杂性, 仍有很多内容需要深入的研究, 陆面数据同化系统的误差估计与处理是受到广泛关注的问题之一。McLaughlin^[5]指出(水文)数据同化问题的实质是在概率框架下强调不同来源信息的不确定性; 指出高维问题以及模型误差和观测误差的准确估计, 将是今后(水文)数据同化研究所关注的要点。Reichle^[3]系统综述了地球系统科学领域数据

同化方法的研究现状、应用情况及潜在的问题。文中在讨论数据同化目的及方法的基础上, 指出目前国际数据同化领域的研究前沿集中在输入误差协方差矩阵研究、质量控制和模型偏差等方面。国内研究中, 文献[6]利用经典的 Lorenz 模型系统地比较了集合数据同化中的各种误差处理方案, 并提出以控制工程领域的进化计算方法处理误差演进, 利用遗传算法得到最优误差调节因子。韩旭军^[7]、晋锐等^[8]和韩丽娟^[9]分别在其以陆面数据同化系统研究为主题研究工作中, 针对同化土壤水分、土壤温度和冻土参数等目标在不同的陆面过程模型(Variable Infiltration Capacity, VIC 模型、Simultaneous Heat and Water, SHAW 模型和 Community Land Model, CLM 模型)上, 深入研究了实际同化系统中的误差处理。综上所述, 在同化系统的设计中不确定性与误差估计被认为是亟需解决的重中之重的问題^[3,10~12]。

总体来讲, 陆面数据同化中的误差估计与处理

* 收稿日期:2011-01-25; 修回日期:2011-07-03.

* 基金项目: 国家高技术研究发展计划重点项目“全球陆表特征参量产品生成与应用研究”子课题“陆面模拟与同化系统示范研究”(编号:2009AA122104); 国家自然科学基金项目“基于鲁棒滤波方法的陆面数据同化系统误差估计与处理”(编号:41061038); 国家杰出青年科学基金项目“流域尺度陆面数据同化系统研究”(编号:40925004)资助。

作者简介: 摆玉龙(1973-), 男, 甘肃会宁人, 副教授, 主要从事数据同化、参数估计研究。E-mail: yulongbai@gmail.com

方法研究起步较晚,并且与大气和海洋数据同化相比较,又有自身独特的性质。本文将首先从整体的角度定义了同化系统各种误差表现,指出误差问题始终贯穿于数据同化方法研究中;在此基础上,介绍了顺序数据同化和连续数据同化在误差定义和处理方面的特点及应用;深入分析了集合数据同化过程中的误差来源、误差估计的方法和目前的研究状况;进而针对模型结构误差,介绍了在集合数据同化中常用的乘数放大法、附加放大法和松弛先验法等处理方案。最后总结讨论了陆面数据同化中误差问题的特点、难点以及重要研究方向。

2 现代数据同化方法中的误差问题

2.1 误差定义及来源

为了确定数据同化系统中的误差来源,首先需要定义误差及其相关问题。本文所遵循的符号系统以 Ide 等^[13]文献为标准。在国际数据同化领域的研究中,该文献被一致认为是描述数据同化系统问题的“标准语言”。按照状态空间方法可以给出描述动力学系统的状态预报模型和观测模型,二者在数据同化系统中又分别被称为模型算子和观测算子。

状态空间的非线性预报模型——模型算子表示为:

$$X^i(t_k) = M_k(X^i(t_{k-1}), u_k, \theta) + \eta(t_k) \quad (1)$$

$$X \in \mathbb{R}^{n \times 1}, \eta \in \mathbb{R}^{n \times 1}$$

式中, X 是系统的状态向量(如土壤水分廓

线); n 是系统的维数;上标 i 代表真值; $\eta(t_k)$ 是模型误差;状态向量后的 (t_k) 或 M 的下标 k 表示时间; u 是外部强迫数据(如大气驱动数据), θ 为模型参数(如土壤水力和热力参数); M 是非线性模型, $M(\cdot): \mathbb{R}^{n \times 1} \times \mathbb{R}^{n \times 1} \rightarrow \mathbb{R}^{n \times 1}$, 它可能是依赖于时间的,因此使用了时间下标 k ,但一般在数据同化系统中, M 不随时间变化。

由于数据同化方法是将观测数据同化到动态模型中,因此观测模型表示为:

$$Y^o(t_k) = H_k(X^i(t_k), \varphi) + \varepsilon(t_k)$$

$$Y \in \mathbb{R}^{m \times 1}, \varepsilon \in \mathbb{R}^{m \times 1} \quad (2)$$

式中, $Y^o(t_k)$ 是 t_k 时刻的观测; $\varepsilon(t_k)$ 是观测误差; m 是观测向量的维数; H 是观测算子,满足 $H(\cdot): \mathbb{R}^{n \times 1} \times \mathbb{R}^{m \times 1} \rightarrow \mathbb{R}^{m \times 1}$; φ 为观测算子参数。

由系统动力学方程可见,陆面数据同化系统的误差来源包括模型误差、观测误差和同化算法误差。其中,模型误差包括模型结构误差、参数误差、驱动数据带来的误差和模型计算误差等多种来源。模型计算误差是指空间、时间差分、截断误差以及其他计算过程中表现出来的误差。观测误差由观测算子误差、仪器误差和代表性误差组成。此外,不同的数据同化方法会带来同化算法误差。如 Monte Carlo 类方法中的有限样本数产生的采样误差等,具体的讨论将在第三节展开。

综上,以集合数据同化系统为例,图1从模型演进和误差表现的角度给出了数据同化系统的基本组成和算法流程。图1中的箭头及序号表示模型向前

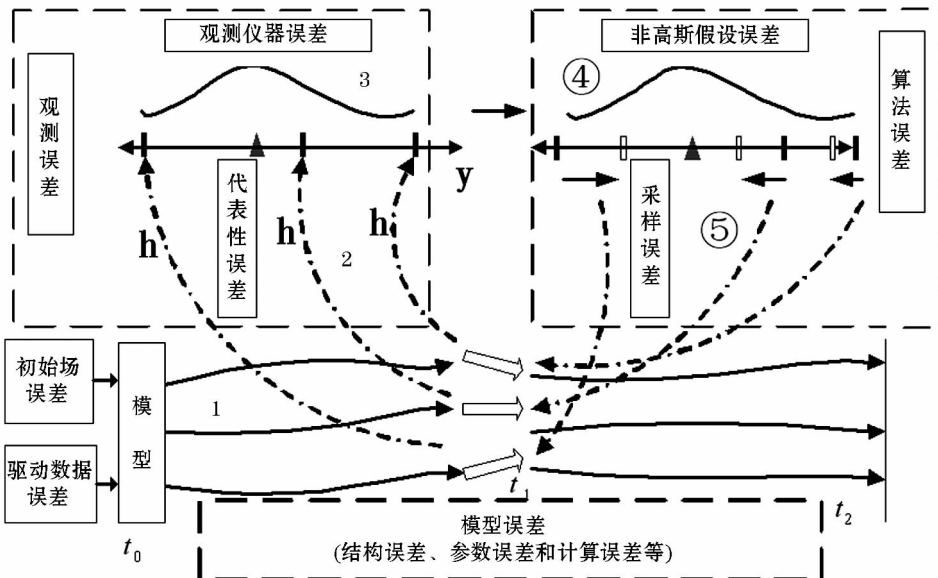


图1 数据同化系统误差来源及误差表现

Fig. 1 Error sources and error characteristic of data assimilation systems

演进的同化流程,框图表示每一步骤中的误差表现。

就陆面数据同化而言,其最终目标是“平衡”模型输出和观测等多源信息,在模型的动力框架内,融合不同来源和不同分辨率的直接与间接观测,将陆面过程模型和各种观测算子(如辐射传输模型)集成为不断地依靠观测而自动调整模型轨迹,并且减小误差的预报系统,从而获得更加准确一致的模型预报量^[10,14,15]。因此,误差问题的研究始终贯穿于数据同化方法研究中。从某种程度上讲,数据同化方法的发展历程中的主题之一是:“如何正确合理地给出同化系统的各种误差信息”,在方法论上主要借助于估计理论、控制论、优化方法来处理不确定性和误差估计问题。

目前,现代数据同化方法可被分为顺序数据同化和连续数据同化 2 类^[15,16]。以下,将从数据同化方法的角度讨论误差定义及误差处理问题。

2.2 顺序数据同化方法中的误差定义

顺序数据同化方法是指在系统运行过程中,利用观测信息在观测和模型误差分别加权的基础上对模型状态进行更新,从而序贯获得模型状态的后验优化估计。此类方法最早从经典 Kalman 滤波(Kalman Filter, KF)^[17]在数据同化系统中的应用开始,已经发展出扩展 Kalman 滤波方法(Extended Kalman Filter, EKF)、集合 Kalman 滤波方法(Ensemble Kalman Filter, EnKF)^[2]等多种衍生类型,应用对象从理想线性系统扩展到非线性系统,从高斯误差扩展到非高斯误差,并且在误差处理和计算效率方面不断得到改善。

2.2.1 经典 Kalman 滤波

在经典的 Kalman 滤波增益的确定中^[17],预报协方差矩阵的正确性是关键的要害,预报和分析误差协方差矩阵是按照如下形式随时间演进的:

$$P_k^f = M_k P_{k-1}^a M_k^T + Q_k \quad (3)$$

$$P_k^a = (I - K_k H_k) P_k^f \quad (4)$$

式中, M_k 是非线性模型的线性正切项, P_{k-1}^a 是在 $k-1$ 时刻的分析误差协方差矩阵,模型结构误差协方差矩阵 Q_k 由以下 2 式给出:

$$\eta_k = X_k^i - M(X_{k-1}^i) \quad (5)$$

$$Q_k = \langle \eta_k \eta_k^T \rangle \quad (6)$$

式中, η_k 是模型误差, X_k^i 是在特定分析时刻 k 未知的状态“真值”。按照文献[18],式(3)中隐含了整体的预报误差是由 2 部分组成的:一部分是在预报过程中由于初始状态和它的动态演进带来的

“内在误差(internal error)”,即 $M_k P_{k-1}^a M_k^T$;另一部分是由于模型的描述缺陷带来的“外在误差(external error)”,即 Q_k 。

在经典的 Kalman 滤波中,模型误差 η_k 假定是零均值的 Gauss 白噪声。在线性系统中,如果模型和观测误差是零均值(无偏)且时间不相关,则 KF 是最优无偏估计。然而,在实际的非线性陆面过程中,许多模型误差信息并不是白噪声,而是具有时间、空间相关性和时变性的,难以建立统一的误差模型,有时甚至根本无法获得其统计特性^[19]。这些情况下 KF 不再是最优无偏估计。

2.2.2 扩展 Kalman 滤波

针对非线性系统的 EKF 滤波需要发展系统动力模型和观测模型的切线性算子^[20]。在误差处理方面,EKF 算法和经典 KF 方法一样需要假定观测与模型状态误差呈正态分布。以 EKF 算法为代表的滤波方法,在高维数据同化系统时,其误差协方差矩阵的计算量和存储量都非常大。因此,该方法在高维陆面数据同化系统中的应用范围并不广泛^[20,21]。

2.2.3 集合 Kalman 滤波

EnKF 的核心思想是利用 Monte Carlo 方法实现集合预报和集合滤波,并近似计算误差矩阵^[11],一方面改善了 EKF 方法中的误差矩阵计算非常消耗时间的缺点;另一方面,避免了模型的线性化,保持了模型的所有动力特征。这种方法已经在海洋、陆面和大气数据同化领域得到了广泛应用^[2,11]。

在 EnKF 中,经典 KF 滤波中所指的“内在误差”的部分可以由集合积分来改进。集合协方差矩阵定义为:

$$P_e^f(t_k) = \frac{1}{N-1} \sum_{i=1}^N [X_i^f(t_k) - \overline{X^f(t_k)}] \times [X_i^f(t_k) - \overline{X^f(t_k)}]^T, P \in \Re^{n \times n} \quad (7)$$

$$\text{其中, } X^f(t_k) = \frac{1}{N} \sum_{i=1}^N X_i^f(t_k) \quad (8)$$

原理上讲,如果设计的集合能够满足所有实际地学状态的各种可能实现, P_e^f 就是“真实”预报误差协方差的最优估计。然而事实上,集合成员经常是使用同一个模型在不同的起始时间向前积分而得的。 P_e^f 中缺失的是模型“外在误差”,即式(3)中的误差协方差项 Q_k 。反之,如果预报集合是利用不同的模型参数或者不同的模型产生的, P_e^f 中包含一定模型结构误差协方差的估计信息。

因此,在 EnKF 中完整的预报协方差矩阵应当

表示成为:

$$P^f = P_e^f + Q \quad (9)$$

式中,模型结构误差协方差 Q 是应该明确的。然而,在复杂地球表层系统中,其值很难确定。因此在数据同化领域发展了附加放大法等模型误差参数化改进方案,相关内容将在第四节详细介绍。

2.2.4 非线性非高斯滤波方法

为了更好地应对 KF 类同化方法在误差的 Gauss 假设和对于非线性问题处理能力有限等缺点,最早发展于信号处理与控制领域的非线性滤波方法逐渐被引入到数据同化领域,体现了“非线性估计算法应更接近系统的非线性系统本质”的先进思想^[22]。近年来,以粒子滤波、鲁棒滤波为代表的非线性、非高斯滤波方法的优异性能引起了高度关注^[23],这些方法主要改进的是 KF 类方法中的 Gauss 假设和线性系统假设等“算法误差”。

粒子滤波(Particle Filter, PF)是一种基于 Monte Carlo 模拟的 Bayes 滤波方法。其主要思想是利用状态空间的一组加权随机样本粒子逼近状态的后验概率分布。其优点是不受模型的线性和高斯假设约束,适用于任意非线性非高斯动态系统;在计算上无需矩阵求逆,误差演进易于实现。同时,它使用了模型误差的全概率分布,因而理论上 PF 方法对于误差分布的尾部比 EnKF 方法更敏感,这对于某些典型陆面过程研究(如洪水预报)来说是很重要的,但它要求有比 EnKF 方法更多的集合成员(粒子),以取得可靠的模型误差估计,因而计算量更大,造成实时应用的困难^[23]。

2.3 连续数据同化方法中的误差定义

连续数据同化方法包括变分方法(三维变分:3DVar 和四维变分:4DVar)和 Monte Carlo 全局优化方法 2 类^[24]。

4DVar 方法可以归结为求如下目标函数的极小化问题^[25]:

$$J(x_0) = \frac{1}{2} (x_0 - x_0^b)^T B^{-1} (x_0 - x_0^b) + \frac{1}{2} \sum_{k=0}^N [H_k(x_k) - y_k]^T R_k^{-1} [H_k(x_k) - y_k] \quad (10)$$

式中, x_0^b 是在 t_0 时刻的背景值或初猜值,4DVar 同化窗口的长度为 $t_N - t_0$; y_k 和 R_k 分别为在时刻 t_k 的观测向量和其对应的观测误差协方差矩阵; B 是背景协方差矩阵; $x_k = M_k(x_0)$ 是在观测时刻 t_k 由非线性模型向前积分获得的模型状态; H_k 是非线性观测算子。由于背景协方差矩阵是由预报误差的概率

密度函数描述的,和顺序同化方法相比,如果公式(3)中的 P_k^f 已知,它可以作为变分同化中背景协方差矩阵 B 的最优选择。然而,由于 Kalman 滤波中的 P_k^f 是不能显式求取的,因此变分同化方法中的背景协方差矩阵 B 只能利用分析新息向量、采用不同的预报长度(NMC methods)和集合(Monte Carlo)等方法获取。Bannister^[26] 系统综述了变分同化方法中的预报误差协方差矩阵的特点和上述估计方法,代表了目前该研究的最前沿方法的评价和综述。

在变分数据同化的应用方面,3DVar 在整个分析过程只对模型误差协方差矩阵做一次先验的估计,并采用空间均匀与各向同性等假定,这种误差处理与真实的大气或者陆面过程的误差情形有很大差别,从而大大影响了同化的结果^[27]。在欧洲中期天气预报中心(European Centre for Medium-Range Weather Forecasts, ECMRWF)等单位,4DVar 的技术已比较成熟。在 4DVar 中,模型误差的结构在分析时间窗中被隐式地调整,从而在一定程度上反映了它随着大气环流变化的特点,但是 3DVar 和 4DVar 需要求解复杂的切线伴随模式。然而,由于大部分陆面过程模型和观测算子都是强非线性的,因此传统的变分法很难得到很好的应用^[28,29]。

3 误差估计综述

数据同化过程中的误差估计研究是其技术向前发展的必经之路,如何减少预报中的不确定性,是目前陆面同化研究中研究的一个前沿问题。按照误差定义及表现,各种误差的估计方法主要有如下方面:

3.1 模型误差

3.1.1 模型输入端误差估计

数据同化系统在输入端的误差包括驱动误差和初始场误差^[1]。模型驱动数据误差对同化结果有着至关重要的影响。如降雨数据时空变异性与观测站网固定点观测之间的矛盾是误差的主要来源^[30]。同时,人们对驱动数据的误差来源,如空间变异程度、流域特征和前期驱动数据状态等复杂关系的认识还不是很清楚。估计输入端误差的合适方法是将输入以随机变量的形式提供给模型。早期的陆面数据同化研究中,把这些误差项都按照扰动集合的方法来处理,在每一个预报或者分析集合上加上一个给定期望和协方差的随机数来增大集合的离散程度,这样既保持了集合的离散分布,又可以引入非加性噪声,比简单的加性噪声更合理。另外,地统计学方面的知识也会对误差的空间变异性描述提供理论

支持,将会成为未来的研究重点^[31]。

3.1.2 模型参数误差估计

由于模型参数的准确性在某种程度上决定了模型预报的准确性,因此模型参数误差估计在陆面数据同化中至关重要。陆面过程模型中的模型参数种类繁多,既包括具有明确的物理意义,可以通过动态方程求解的模型参数,也包括具有一定物理基础但经验性很强的概念性模型参数。因此,模型参数误差的表现是多样化的。如在一些陆面过程模型中往往忽略了气候因子和下垫面均呈现空间分布不均匀的事实,只能给出空间均匀化的模拟结果,必然使其结构和参数的物理意义不明确。模型参数误差的估计方法主要有:①参数敏感性分析方法。敏感性分析方法包括传统参数敏感性分析方法和区域敏感性分析法。前者一般也称为扰动分析方法,在某个参数最佳估计值附近给定一个人工干扰,并计算参数在这一很小范围内产生波动所导致模型输出的变化率。因此,该方法只能反映参数局部的敏感性,不能反映出模型结构特征与参数相关性。区域灵敏度分析承认参数空间分布的复杂性与相关性,是一个在一定准则下对模型结构随机参数大样本发生响应的统计分析过程。②参数不确定性分析方法。由于参数响应面的多峰性和参数相关性,参数一般无法识别或可识别性很弱,即使是采用全局优化的算法,最佳估计参数和目标函数都很难收敛到单一数值。针对传统参数识别的缺陷,模型参数不再当作确定值来处理,而是当作随机变量、区间变量、模糊变量等来处理。随着马尔科夫链蒙特卡罗方法(Markov Chain Monte Carlo, MCMC)的发展,对后验分布进行蒙特卡罗取样在很大程度上克服了这些困难。MCMC方法从参数的联合后验分布中提取样本,提供了比单点估计更多的信息,而且避免了用一个正态近似后验分布进行推断的必要。Beven等^[32]在1992年提出GLUE(General Likelihood Uncertainty Estimation)采用集合预报的方法,代表目前模拟参数误差的最新进展。GLUE方法也存在一些缺陷,如计算要求高,参数集维数的选择具有主观性等^[33]。

3.1.3 模型结构误差估计

在现实中模型结构误差最难给定,其原因是:①模型的物理机制认识缺陷。由于对模型本身的物理过程认识不足,现有陆面过程模型均对实际物理过程进行了一定简化和抽象,再加之模型参数及驱动数据的不确定性,随着模型向前运行,误差不断积

累,不确定性增强。②模型的自由度问题。数据同化系统的自由度是指用于完全描述系统行为的一组无关量的个数。一般而言,模型的自由度往往比观测值多,因此很难把模型的所有状态变量同观测值进行比较。③模型的时空连续性问题。模型的运行是连续的,而观测一般是离散的。因此,定量模型误差需要的观测值不可能与模型的时间步长完全一致。④状态变量的相关性问题。一个状态变量的误差与其他变量有关,如地表温度与叶面积指数、地表热量等有关,这使误差的分析及确定非常困难。在模型结构误差估计方面目前已有的研究包括:①将带有观测信息与不带有观测信息的数据同化系统加以比较^[27];②以同一验证时间结束,而起始时间不同的模型预报之间的比较;③同化增量或更新的分析^[34];④集合积分等内容^[3]。

综上,从根本上讲,模型误差协方差矩阵的确定还是具有相当的主观性。

3.2 观测误差

如前所述,观测误差主要包括观测算子误差、代表性误差和仪器误差等方面。

3.2.1 观测算子误差

陆面数据同化中的观测算子包括:针对直接观测而言,如空间插值;以及对于遥感等非直接观测而言,观测算子一般是联系模型状态变量(如土壤水分、土壤温度)和观测量(如反射率、亮温)之间的正向模型,如遥感辐射传输模型。观测算子误差是指对辐射传输等机理的理解不足等导致的误差。这些观测算子往往包含一定的假设条件,使误差的传递具有很多的不确定性,而这种不确定性极难估计。

3.2.2 代表性误差

代表性误差是源自尺度匹配的陆面数据同化特有问题之一。如微波遥感信号如AMSR-E的空间分辨率为 0.25° ,而陆面模型是采用点的驱动数据和参数运行,虽然 0.25° 的遥感信息包括了点的信息,但当模型同化遥感信息时,其空间尺度是不匹配的。同样,在大气科学中,测量温度就代表点的信息,而陆面模型计算的是一个大网格的平均温度,其间也存在尺度不匹配的问题。

3.2.3 仪器误差

观测误差的其他来源包括观测仪器误差、环境噪声、采样误差和传感器误差等。如不同轨道的卫星观测信息由于定标系统的不一致而是有偏的^[35]。

总体而言,上述各种观测误差在国内外研究中还未有较好的解决办法。①一般都是假设观测误差

符合高斯分布。根据观测误差的统计结果对观测数据进行随机扰动,进行物理范围越界检查后生成观测矢量集合。②引入局地分析减小虚假相关。在观测误差协方差矩阵中引入地统计中误差的空间变异性表达方法。这种方法对于流域尺度数据同化系统来讲可以说是一种非常有效的分析方案^[7]。③仪器误差一般可利用并行观测减小。所谓并行观测即预先给定一批数据,让仪器测定,将测定结果和原值比较,从而确定仪器的性能。同时还可以通过分析数据剔除观测偏差。

3.3 集合同化中的算法误差

算法误差主要是指利用 Monte Carlo 方法时的采样误差和各种算法中的计算误差等。

采样误差是集合数据同化中的主要误差来源,其表现在样本数目的选取上。大量的研究表明,集合数增加会显著减少采样误差,但会带来更多的计算负担^[25]。第四节将要介绍的误差处理方法在某种程度上就是寻找在小样本前提下提高同化质量的方法^[36,37]。

所有上述提到的误差都共同作用在数据同化系统的运算进程中。因此量化这些误差的相关项非常困难。在集合数据同化中,这些误差的“净”作用将会使得误差协方差矩阵产生小偏差,同时在预报误差中产生了一个新的特征空间。因此,各种时新的方法被引入来处理这些误差问题。

4 集合数据同化中的误差处理方法

如上所述,在实际的陆面数据同化中,无论是模型误差还是观测误差,都很难准确地估计。在现代数据同化方法中,发展了一系列的误差处理方法,这些方法并不估计模型先验误差和观测误差,只是在整体上减少各类误差对同化系统的影响。如①在集合数据同化中,可以采用协方差“inflation”的方法处理以上各类误差,包括旨在消除集合二阶矩中偏差的乘数放大法(Multiplicative inflation)、附加放大法(Additive inflation)和松弛—先验法(Relaxation to prior)。这类方法通过扩大或者改变集合方差的方法,可以在集合样本数较少的情形下,获得更好的同化效果。目前已在陆面数据同化系统中有所尝试^[38]。②针对模型偏差处理常采用状态增广法(State Augmentation Methods)。这类方法通过改变新息向量减少同化系统模型偏差,目前只在理想模型中有所试验。③协方差矩阵局地化方法也可用来消除集合数据同化中的虚假噪音等问题,比较典型

的有 Schur 乘积算法等^[36]。

4.1 基于集合二阶矩误差处理方法

第一种方法称为协方差或乘数放大法^[39],这种方法是在集合均值的基础上扩大预报集合的方差。

此方法中,为了模拟模型误差 Q ,在该时刻观测信息被同化之前,基于其均值的集成员方差按照(11)式所示,通过一个略大于0因子 r 放大:

$$\tilde{P}_e^f \leftarrow (1+r) \times P_e^f = P_e^f + r P_e^f = P_e^f + Q_{multi} \quad (11)$$

式中可以看出, Q_{multi} 项模拟了模型误差 Q ,而且这一项和 P_e^f 有相同的误差子空间。乘数放大法被广泛的应用于大气和陆面数据同化中。在理想模型实验^[36]和实际数据仿真试验^[37]中使用该方法提高了同化系统的精确度。国内 Qin 等^[38]在同步估计土壤水分和模型参数的实验中,使用乘数放大法作为误差参数化方案描述模型不确定性。

第二种处理模型误差的方法称为附加放大法。在 EnKF 中,需要对每一个成员引入随机扰动。这种随机扰动是从已知的协方差统计特性的分布中采样得到的。在更新周期开始前,每个背景预报集成员都会加上一个同模型状态维数相同噪音 r^* :

$$\tilde{x}_i^f = x_i^f + r^* \quad (12)$$

进而,背景误差协方差矩阵变为:

$$\tilde{P}_e^f = \frac{1}{N-1} \sum_{i=1}^N (\tilde{x}_{e(i)}^f - \tilde{x}_e^f) (\tilde{x}_{e(i)}^f - \tilde{x}_e^f)^T = P_e^f + Q_{add} \quad (13)$$

和乘数放大法相比较,附加放大法调整以后的集合可能会扩展到不同于未调整集合的误差子空间。进而,这种方法不会像乘数放大法一样任意地扩大集合扩展度。

附加放大法作为一种广泛使用的误差处理方法,在加拿大气象服务中心(Meteorological Service of Canada, MSC)的理想通用循环系统中,这种方式的误差处理显示了很好的效果^[27]。Houtekamer 等^[40]将模型误差用“附加误差”的方法参数化,即给背景误差协方差模型的每个集合增加噪音,噪音的结构同于 3DVar 协方差矩阵。初步结果显示,这种方法是具有竞争力的,同实际运行的 3DVar 方法相比稍差。

第三种处理模型误差的方法是由 Zhang 等^[41]提出的松弛先验法,其本质可以认为是乘数放大法的一种变形算法。其基本含义是在每一个分析点通过式(14)的方式,将分析扰动 x^a 减小并独立地加入先验扰动 x^f 的影响,形成调整后的分析扰动 x^a 。

$$x^a = (1-\alpha)x^a + \alpha x^f \quad (14)$$

式中, α 为 0 ~ 1 之间的调整因子。这种调整是在每次状态更新后开始的, 调整后的分析扰动将向前积分, 从而形成下一个分析时刻的背景集合。

同乘数放大法相比, 此方法的优点是它只是在观测信息对分析有影响的位置调整集合扰动, 因而避免了在没有观测信息影响时, 集合方差无界增长的趋势。这里的调整因子具有一定的通用性, 其值的恰当选择可以提高同化系统的精度。Zhang 等^[41]在其理想模型同化实验中, 选择调节因子 α 为 0.5, 但是 Meng 等^[42]在显著模型误差出现时将调节因子增大到 0.7。Whitaker 等^[37]发现调节因子的最优选择数值为 0.88, 意味着给先验集合扰动的权重为 88%, 而从状态更新中计算出来的扰动的权重仅为 12%。

针对以上 3 种方法的对比研究, Hamill 等^[27]比较了各种参数化模型误差方案; Li 等^[43, 44]将常用的误差处理方法应用于大气 Speedy 模式, 验证了各种方法的同化效果; Bai 等^[6]将上述 3 种常用的方法在理想数值模拟试验中做了深入的比较研究, 结果显示误差调节因子在可行解的范围内呈现出多峰性, 在统一的适应度函数的约束下, 利用遗传算法可以得出最优误差因子。并且在松弛先验法的基础上, 提出了基于进化计算的处理误差演进的新方法。

4.2 对模型偏差的处理方法

状态增广法是最早发展于控制领域的处理系统方程中未知误差的方法, 其核心是将未知的误差信息看作是由系统激励下的一个系统输出, 同时把它扩大到待估计的状态变量中去, 与状态变量一起进行估计^[45]。Dee 等^[46, 47]提出用新息向量的方法来估计模式误差的参数, 提出了一种在 Kalman 滤波更新中, 实时估计和修正模型偏差的方法。这种偏差修正方法已经得到了成功的验证^[48, 49]。Baek 等^[50]发展了另一种偏差消除方法, 这种方法考虑了模型状态中不确定性的相关性。假设对于同化的观测信息是无偏的, 模型偏差和模型参数偏差可以合并成为一个状态向量。这样的数据同化系统可以进行模型偏差估计。该方法已成功应用于 Lorenz-1996 模型中, 并在同化地表温度和同化流域土壤水分观测方面得到了验证^[50]。

5 结 语

随着数据同化理论与实践研究的发展, 误差的估计与处理问题逐渐成为制约数据同化系统发展的瓶颈。陆面数据同化系统中的误差来源主要包括模

型误差、观测误差和算法误差等。由于陆面过程模型只是对真实地表过程的近似描述, 因此模型先验误差的估计是一个非常具有挑战性的课题。观测作为一种重要的手段可以在数据同化方法的框架内提高模型的准确性和可预报性。然而, 地表观测信息往往是所感兴趣的同化变量的间接描述, 误差信息表现出强烈的空间异质性和相关性。

文中, 以误差理论为主线系统地分析了顺序数据同化等现代数据方法中的误差定义与特征, 从中可以看出同化方法的发展在很大程度上依赖于误差问题的正确或者更加接近真实的表达; 同时侧重于目前流行的基于集合的数据同化方法, 按照顺序同化进程分析了各个环节的误差来源和一般处理方法。

结合国内外的研究现状, 目前误差研究主要研究方向包括: ①模型误差的正确给定。综合已有的研究结果, 在以后具体问题研究中需要从陆面过程模型本身物理过程的认识出发, 针对不同的应用背景确定感兴趣的状态变量, 深入研究误差协方差矩阵的构造; ②观测误差中的代表性误差的研究是目前需要攻克的难点问题, 主要表现在观测算子的物理特性及观测空间和状态空间的转换方面, 地统计学的方法或许会提供一个新思路。③由于所有误差在同化进程中都共同作用在研究对象上, 因此可以从整体上加以考虑和处理。在大气数据同化领域, 发展了一些值得借鉴的方法。大量的研究工作表明这些方法是可行的, 而且越来越受到同化研究领域的重视^[5~54]。然而, 这些方法在陆面数据同化系统中的应用与探索还有待进一步展开。

参考文献 (References):

- [1] Li Xin, Huang Chunlin, Che Tao, et al. Development of a Chinese land data assimilation system: Its process and prospects [J]. *Process in Natural Science*, 2007, 17(8): 881-892. [李新, 黄春林, 车涛, 等. 中国陆面数据同化系统研究的进展与前瞻 [J]. 自然科学进展, 2007, 17(2): 163-173.]
- [2] Evensen G. *Data Assimilation: The Ensemble Kalman Filter* [M]. Berlin, Heidelberg: Springer, 2007: 279.
- [3] Reichle R H. Data assimilation methods in the Earth science [J]. *Advances in Water resources*, 2008, 31: 1 411-1 418.
- [4] Mitchell K E, Lohmann D, Houser P R, et al. The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system [J]. *Journal of Geophysical Research*, 2004, 109(D07): 32; doi: 10.1029/2003JD003823.
- [5] McLaughlin D. An integrated approach to hydrologic data assimila-

- tion; interpolation, smoothing and filtering [J]. *Advances in Water Resources*, 2002, 25: 1 275-1 286.
- [6] Bai Y L, Li X. Evolutionary algorithm-based error parameterization methods for data assimilation [J]. *Monthly Weather Review*, 2011, doi: 10.1175/2011MWR3641.1
- [7] Han Xujun. Algorithm Development and Application of the Land Data Assimilation at Catchment Scale [D]. Beijing: Graduate University of the Chinese Academy of Science, 2008. [韩旭军. 流域尺度陆面数据同化方法及其应用研究[D]. 北京:中国科学院研究生院, 2008.]
- [8] Jin Rui, Li Xin. Improve the estimation of hydrothermal state variables in the active layer of frozen ground by assimilating in situ observations and SSM/I data [J]. *Science in China (Series D)*, 2009, 52(11): 1 732-1 745. [晋锐, 李新. 同化站点观测和SSM/I亮温改善冻土活动层状态变量的模拟精度[J]. 中国科学: D辑, 2009, 39(9): 1 220-1 231.]
- [9] Han Lijuan. Estimation of Evapotranspiration by Assimilating MODIS LST Product into the CLM [D]. Beijing: Beijing Normal University, 2006. [韩丽娟. 同化MODIS地表温度产品和陆面过程模型研究地表蒸散[D]. 北京:北京师范大学, 2006.]
- [10] Talagrand O. Assimilation of observations, an introduction [J]. *Journal of the Meteorological Society of Japan*, 1997, 75(1B): 191-209.
- [11] Evensen G. The Ensemble Kalman Filter; Theoretical formulation and practical implementation [J]. *Ocean Dynamics*, 2003, 53: 343-367.
- [12] Huang C L, Li X, Lu L, et al. Experiments of one-dimensional soil moisture assimilation system based on ensemble Kalman filter [J]. *Remote Sensing of Environment*, 2008, 112(3): 888-900.
- [13] Ide K, Courtier P, Ghil M, et al. Unified notation for data assimilation: Operational, sequential and variational [J]. *Journal of the Meteorological Society of Japan*, 1997, 75(1B): 181-189.
- [14] Li X, Koike T, Mahadevan P. A very fast simulated re-annealing (VFSA) approach for land data assimilation [J]. *Computers and Geosciences*, 2004, 30(3): 239-248.
- [15] Daley R. Atmospheric Data Analysis [M]. New York: Cambridge University Press, 1991.
- [16] Li Xin, Bai Yulong. A Bayesian filter framework for sequential data assimilation [J]. *Advance in Earth Science*, 2010, 25(5): 515-523. [李新, 摆玉龙. 顺序数据同化的Bayes框架[J]. 地球科学进展, 2010, 25(5): 515-523.]
- [17] Kalman R E. A new approach to linear filtering and prediction problems [J]. *Transactions of the ASME Journal of Basic Engineering*, 1960, 82 (Series D): 35-46.
- [18] Li Hong. Local Ensemble Transform Kalman Filter with Realistic Observations [D]. Maryland: University of Maryland, 2007.
- [19] Crow W T, Van Loon E. Impact of incorrect model error assumption on the sequential assimilation of remotely sensed surface soil moisture [J]. *Journal of Hydrometeorology*, 2006, 7: 421-432.
- [20] Kumar P, Kaleita A L. Assimilation of near-surface temperature using extended Kalman filter [J]. *Advances in Water Resources*, 2003, 26: 79-93.
- [21] Pham D T, Verron J, Roubaud M C. A singular evolutive extended Kalman filter for data assimilation in oceanography [J]. *Journal of Marine Systems*, 1998, 16(3): 323-340.
- [22] Chai Lin, Yuan Jianping, Luo Jianjun, et al. New development in nonlinear systems estimation [J]. *Journal of Astronautics*, 2005, 26(3): 380-384. [柴霖, 袁建平, 罗建军, 等. 非线性估计理论的最新进展[J]. 宇航学报, 2005, 26(3): 380-384.]
- [23] Han Xujun, Li Xin. Review of nonlinear filters in the land data assimilation [J]. *Advances in Earth Science*, 2008, 23(8): 813-820. [韩旭军, 李新. 非线性滤波方法与陆面数据同化[J]. 地球科学进展, 2008, 23(8): 813-820.]
- [24] Courtier P. Variational methods [J]. *Journal of the Meteorological Society of Japan*, 1997, 75(1B): 211-218.
- [25] Kalnay E, Li H, Miyoshi T, et al. 4D-Var or ensemble Kalman Filter? [J]. *Tellus A*, 2007, 59A(5): 758-773.
- [26] Bannister R N. A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariance [J]. *Quarterly Journal of the Royal Meteorological Society*, 2008, 134(11): 1 951-1 970.
- [27] Hamill T M, Whitaker J S. Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches [J]. *Monthly Weather Review*, 2005, 133(11): 3 132-3 147.
- [28] Qiu Chongjian. Four dimensional variational data assimilation with discontinuous models [J]. *Journal of Lanzhou University (Natural Sciences)*, 1997, 33(1): 115-119. [邱崇践. 变分四维同化方法中的不连续问题[J]. 兰州大学学报: 自然科学版, 1997, (3): 115-119.]
- [29] Reichle R H, Walker J P, Koster R D, et al. Extended versus ensemble filtering for land data assimilation [J]. *Journal of Hydrometeorology*, 2002, 3(12): 728-740.
- [30] Goodrich D C, Faures J, Woolhiser D A, et al. Measurement and analysis of small-scale convective storm rainfall variability [J]. *Journal of Hydrology*, 1995, 173(4): 283-308.
- [31] Han X J, Li X. On the representation of spatial uncertainty with stochastic simulation in land data assimilation [C] // the 8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Vol. 1: Shanghai: World Academic Press, 2008: 221-227.
- [32] Beven K J, Binley A. The future of distributed models: Model calibration and uncertainty prediction [J]. *Hydrological Processes*, 1992, 6: 279-298.
- [33] Yin Xiongrui, Xia Jun, Zhang Xiang, et al. Recent progress and prospect of the study on uncertainties in hydrological modeling and forecasting [J]. *Water Power*, 2006, 32(10): 27-31. [尹雄锐, 夏军, 张翔, 等. 水文模拟与预测的不确定性研究与展望[J]. 水力发电, 2006, 32(10): 27-31.]
- [34] Dee D P, Todling R. Data assimilation in the presence of forecast bias: The GEOS moisture analysis [J]. *Monthly Weather Review*, 2000, 128(9): 3 268-3 282.

- [35] Zhang Peng, Yang Jun, Dong Chaohua, *et al.* General introduction on payloads, ground segment and data application of Fengyun 3A[J]. *Front Earth Science China*, 2009, 132: 1 238-1 253.
- [36] Hamill T M, Whitaker J S. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter [J]. *Monthly Weather Review*, 2001, 129(11): 2 776-2 790.
- [37] Whitaker J S, Hamill T M. Ensemble data assimilation with the NCEP Global forecast system [J]. *Monthly Weather Review*, 2007, 136(2): 463-482.
- [38] Qin J, Liang S, Yang K, *et al.* Simultaneous estimation of both soil moisture and model parameters using particle filtering method through the assimilation of microwave signal[J]. *Journal of Geophysical Research*, 2009, 114, D15103, doi: 10. 1029/2008JD011358.
- [39] Anderson J L, Anderson S L. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilation and forecast [J]. *Monthly Weather Review*, 1999, 127(12): 2 741-2 758.
- [40] Houtekamer P L, Mitchell H L. Ensemble Kalman filter [J]. *Quarterly Journal of the Royal Meteorological Society*, 2005, 131:3 269-3 289.
- [41] Zhang F, Snyder C, Sun J. Impacts if initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter[J]. *Monthly Weather Review*, 2004, 132(5): 1 238-1 253.
- [42] Meng Z, Zhang F. Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation Part II: Imperfect model experiments [J]. *Monthly Weather Review*, 2007, 135(4): 1 403-1 423.
- [43] Li H, Kalnay E, Miyoshi T. Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter [J]. *Quarterly Journal of the Royal Meteorological Society*, 2009, 135: 523-533, doi: 10. 1002/qj. 371.
- [44] Li H, Kalnay E, Miyoshi T, *et al.* Accounting for model errors in ensemble data assimilation [J]. *Monthly Weather Review*, 2009, 137(10): 3 407-3 419.
- [45] Jazwinski A H. *Stochastic Processes and Filtering Theory* [M]. New York: Academic Press, 1970.
- [46] Dee D P, Gaspari G, Redder C, *et al.* Maximum-likelihood estimation of forecast and observation error covariance parameters. Part II: applications [J]. *Monthly Weather Review*, 1999, 127(8): 1 835-1 849.
- [47] Dee D P, da Silva A M. Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology [J]. *Monthly Weather Review*, 1999, 127(8): 1 822-1 834.
- [48] Martin M J, Bell M J, Nichols N K. Estimation of systematic error in an equatorial ocean model using data assimilation [J]. *International Journal for Numerical Methods in Fluids*, 2002, 40(3): 435-444.
- [49] Carton J A, Chepurin G, Cao X, *et al.* A simple ocean data assimilation analysis of the global upper ocean 1950-95, Part I: Methodology [J]. *Journal of Physical Oceanography*, 2000, 30(2): 294-309.
- [50] Baek S J, Hunt B R, Kalnay E, *et al.* Local ensemble Kalman filtering in the presence of model bias [J]. *Tellus*, 2006, 58A: 293-306.
- [51] Zupanski D, Zupanski M. Model error estimation employing an ensemble data assimilation approach [J]. *Monthly Weather Review*, 2006, 134(5): 1 337-1 354.
- [52] Tremolet Yannick. Model-error estimation in 4D-Var [J]. *Quarterly Journal of the Royal Meteorological Society*, 2007, 133: 1 367-1 380.
- [53] Carrassi A, Vannitsem S, Nicolis. Model error and sequential data assimilation: A deterministic formulation [J]. *Quarterly Journal of the Royal Meteorological Society*, 2008, 134: 1 297-1 313.
- [54] Houtekamer P L, Herschel L, Mitchell, *et al.* Model error representation in an operational ensemble kalman filter [J]. *Monthly Weather Review*, 2009, 137(7): 2 126-2 143.

A Review of Error Problems for Land Data Assimilation Systems

Bai Yulong^{1, 2}, Li Xin¹, Han Xujun¹

(1. *Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou 730000, China*; 2. *College of Physics and Electrical Engineering, Northwest Normal University, Lanzhou 730070, China*)

Abstract: As an important methodology for optimally merging Earth observation information and geophysical model output information, data assimilation has played an important role in the area of Earth observation. At present, great progress has been made in the theoretical and methodological exploration and foundation of the operational land data assimilation system. Due to the complexity of research objectives, error problems are thought to be the bottleneck for improving the performance of data assimilation systems. Firstly, the research statuses of error prob-

lems of Land Data Assimilation Systems are reviewed. Based on the mathematical descriptions of land surface process model and measurement process, error sources and error characteristic are unifying defined. In a word, data assimilation systems include model errors, observation errors and the algorithm errors. Secondly, with respect to the sequential and variational data assimilation methods, error definitions and the related theoretical problems of those methods are briefly introduced with the emphasis on the error sources and the fundamental error parameterization methods. Moreover, from the perspective of error estimation, several novel methods for estimating model errors are reviewed from three parts: the model input error estimation, the model parameters error estimation and the model structure error estimations. As for the observation errors, the error sources can be divided with the observation algorithm errors, the representative errors and the instrument errors. Beside some exiting methods, there are no more effectively methods to deal with those kinds of error. Meanwhile, the difficulties for implementing all those methods are clarified. Thirdly, in order to reduce the errors for ensemble data assimilation systems, the common error parameterization methods, such as multiplicative inflation methods, additive inflation methods and the relax-to-prior methods, are employed. All these methods for dealing with model errors are meant to ameliorate the bias error in ensemble second moment. As far as the model bias is concerned, the state augmentation methods are discussed. A new scheme to obtain the optimal estimation of the state and model bias variable simultaneously is reviewed. Finally, the characteristic of all kinds of error estimation and processing methods and the surveys for the future implementation of all above methods in Land data assimilation are given.

Key words: Data assimilation; Error estimation; Error parameterization scheme.