中国科学: 地球科学 2020 年 第 50 卷 第 9 期: 1185 ~ 1194
SCIENTIA SINICA Terrae earthcn.scichina.com
《中国科学》杂志社 SCIENCE CHINA PRESS

评 述

模型与观测的和弦: 地球系统科学中的数据同化

李新[1,2*], 刘丰[3], 方苗[3]

1. 中国科学院青藏高原研究所, 北京 100101;
2. 中国科学院青藏高原地球科学卓越创新中心, 北京 100101;
3. 中国科学院寒区旱区环境与工程研究所, 兰州 730000
* 通讯作者, E-mail: xinli@itpcas.ac.cn

收稿日期: 2019-11-28; 收修改稿日期: 2020-02-29; 接受日期: 2020-04-23; 网络版发表日期: 2020-06-01
中国科学院战略性先导科技专项项目(编号: XDA19070104)、国家自然科学基金项目(批准号: 41801270、41701046)和中国科学院“十三五”信息化专项项目(编号: XXH13505-06)资助

摘要 模型与观测是地球系统科学中两种基本的研究手段, 它们协同演进、相映成辉; 但同时, 两者都未臻完美, 也因为方法论的差异, 表现出相互不协调的一面. 数据同化的出现, 使得模型与观测协力前行, 演奏出一曲和谐的地球系统科学方法论, 也因此在地球系统科学中展现出鲜活的生命力和应用价值. 文章介绍了数据同化在地球系统科学主要分支领域的应用, 追溯了数据同化与理性主义和经验主义方法论的协同演进, 分析了它与估计理论和控制论的渊源, 回顾了国内数据同化近期研究进展, 展望了走向统一的地球系统数据同化所面临的挑战. 数据同化理论与方法将不断进化, 对增强地球系统的理解和预测提供越来越成熟的方法论.

关键词 数据同化, 模型, 观测, 地球系统科学, 方法论

1 引言

地球系统科学的进步有赖于模型和观测的协力前行. 模型是对地球系统科学认知的形式化知识的集大成者, 但模型只是真理的近似, 还远远未臻完美; 观测越来越多源和丰富, 地球大数据洪流滚滚而来, 但所有观测都有特定的时空代表性, 其代表性误差常常难以估计, 这是导致模型和观测不一致的主要原因之一. 然而, 这种不一致性是表象还是真相? 是模型更可信还是观测更可信, 常常难以甄别. 模型和观测可以共奏出美妙的和弦. 从高斯、维纳到洛伦茨, 科学大家一脉相承, 发展了估计理论、控制论、混沌理论, 为模型和观测的融合提供了坚实的方法论基础. 数据同化方法正是植根于这些理论, 它站在巨人的肩膀上, 已经根深叶茂, 成长为地球系统科学的关键方法论, 其核心思想是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, 从而增强系统的可预报性和可观测性.

数据同化具备成为地球系统科学整体及各个分支领域的共同方法论的潜力, 因此, 本文从数据同化在地球系统科学主要领域的应用、研究范式溯源、理论方法演进、未来趋势和挑战等角度对数据同化理论与方法的发展进行了剖析, 并简介了中国地球系统科学领域中的数据同化研究的近期进展.

中文引用格式: 李新, 刘丰, 方苗. 2020. 模型与观测的和弦: 地球系统科学中的数据同化. 中国科学: 地球科学, 50: 1185–1194, doi: 10.1360/SSTe-2019-0280
英文引用格式: Li X, Liu F, Fang M. 2020. Harmonizing models and observations: Data assimilation in Earth system science. Science China Earth Sciences, 63: 1059–1068, https://doi.org/10.1007/s11430-019-9620-x

2 数据同化与地球系统科学

数据同化已经在地球系统科学的分支学科被广泛应用并获得了巨大的成功(图1), 本节简要回顾了数据同化在地球系统科学的主要分支——大气、海洋、陆地、固体地球领域的应用.

(ⅰ) **大气科学**. 数据同化起源于大气科学领域的数值天气预报, 其目的是同化各种来源的观测数据为模型预报提供尽可能准确的初始场. 之后数据同化又逐渐延伸到了大气科学的方方面面, 特别是在大气再分析中起到关键作用. 再分析是指利用数据同化系统把各种来源与不同类型的历史观测数据同数值预报模型进行最优融合, 弥补观测数据时空分布不均匀的缺陷, 以得到时空一致、物理一致的长序列大气再分析数据产品. 从最早的NCEP再分析(Kalnay等, 1996), 到ECMWF再分析, 再到21世纪再分析(Compo等, 2011), 都已在全球及区域气候变化分析、气候诊断、模型评估等领域被广泛使用(Uppala等, 2005; Kobayashi等, 2015; Gelaro等, 2017). 其中, NCEP再分析的介绍论文(Kalnay等, 1996)是地球系统科学领域被引用最多的论文之一(至2019年被引用>25000次), 充分印证了再分析数据在气候变化研究中的广泛应用, 也证明了数据同化的巨大价值. 数据同化在大气科学领域的另一重要应用是为短期数值天气预报提供最优初始场以提高预报的精度, ECMWF和NOAA的短期天气预报系统就是数据同化在短期天气预报中的两个成功实践, 其通过同化各种类型的观测, 对未来一段时间全球的大气与海洋状态进行预报, 同时实时发布预报结果. 该方面一个成功的案例是NOAA的飓风预报系统, 该系统通过每3h同化常规观测、NASA的探空仪、NOAA的多普勒雷达等观测, 显著提高了飓风预报精度和可预报性(Zhang等, 2011). 古气候数据同化是大气数据同化在近些年进军的一个新兴领域, 它的基本思想就是在气候模型或地球系统模型的动力学框架内融合不同类型的气候代用观测(比如, 树轮、冰芯、湖泊沉积物、洞穴沉积物等), 以期对过去的气候变化状态进行最优估计(方苗和李新, 2016), 其典型的应用包括过去千年气候再分析(Hakim等, 2016)、过去千年水文气候重建(Steiger等, 2018). 综上, 大气数据同化在地球系统科学的同化研究中占有基础性地位, 这不仅表现在上述具有影响力的大气数据同化研究, 也表现在应用到其他领域的数据同化理论方法多源于大气科学.

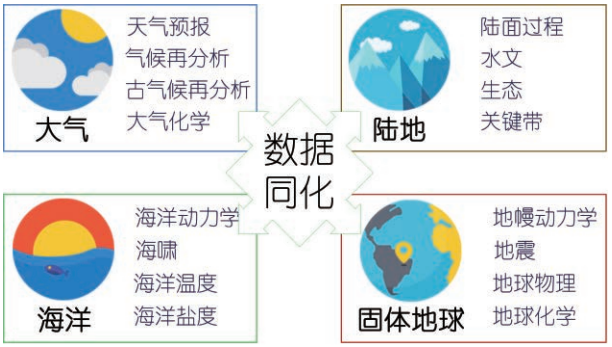

图 1 数据同化在地球系统科学主要分支领域中的应用

(ⅱ) **海洋科学**. 数据同化在海洋科学的众多领域如海洋动力学、海啸、海洋温度、海洋盐度预报和分析等领域被广泛应用(Cummings, 2005; Martin等, 2007; Drenkard和Karnauskas, 2014). 近20年来最被广泛应用的数据同化方法——集合卡尔曼滤波(EnKF)(Evensen, 1994)就是在海洋科学应用中诞生. 由于海洋物理学研究一直受到直接观测不足的制约, 因此海洋数据同化从发展初期就注重使用遥感观测改进海洋动力模型预报精度, 广泛使用了海洋卫星遥感数据和全球海洋观测数据, 如世界海洋数据库(WOD)和全球海洋观测网(Argo). 海洋数据同化除了用于短期海洋预报之外, 在海洋再分析数据的生产上也扮演了极为重要的作用(Hurlburt等, 2009). 从20世纪90年代开始, 美国和欧洲等海洋强国以及中国相继发起了一系列海洋再分析项目(Chassignet等, 2007; Forget等, 2015; Zuo等, 2015), 所产生的海洋再分析数据产品一方面弥补了深层海洋观测的稀缺, 另一方面为海洋动力学研究、海洋气候变化、海洋短期数值预报和气候预测提供了基础数据(Swift等, 2005; Karspeck等, 2017).

(ⅲ) **陆地表层科学**. 陆地数据同化的发展自20世纪90年代中期起步, 目前已经在陆面过程、水文水资源、陆地生态系统等领域取得了巨大成功(McLaughlin, 1995; Xia等, 2012). 其中, 面向大陆尺度和流域尺度的数据同化研究逐步升温, 北美和全球陆面数据同化系统的发展(Mitchell等, 2004; Rodell等, 2004)掀起了包括欧洲、中国在内的多个大尺度的陆面数据同化系统研究的热潮. 陆地数据同化有其自身鲜明的特点: 源于陆地表层系统的高度异质性, 陆地数据同化往往

需要同化多源、多尺度的地面观测数据和卫星遥感数据. 因此, 在发展陆面数据同化系统时, 多尺度的地表异质性所导致的空间代表性误差的估计得到了高度的重视(李新, 2013).

(iv) **固体地球**. 近10年来, 固体地球科学领域中对于数据同化的应用已经起步, 典型的应用包括地震预报、地磁场反演和预报、地幔动力学等领域. 在短时地震预报中, Hoshiba和Aoki(2015)针对当前地震预报工作中对地震震中、震级、地面移动强度的预报低效问题, 提出数值地震预报概念并通过数据同化技术融合实时地震台站观测, 以精确估计当前的地震波场并模拟未来地震波演进. 在对2011年日本东北大地震以及2014年新潟县中部地震进行回报的工作中, 该系统对地震强度提前20s的预报表现出了更准确的、更符合实际观测的精度. 中国地震局所构建的三维空间数值地震预报方法(Wang等, 2017)在Hoshiba和Aoki (2015)的基础上的进一步延伸, 把地震波的预报从二维扩展到了三维空间, 从而进一步提高了对地面移动强度的预报精度. 对地磁场长期变化进行分析与预报是数据同化在固体地球领域的另外一个典型应用(Fournier等, 2010), 其核心思想是在地磁场模型运行过程中不断同化历史观测数据、台站观测、卫星遥感观测, 以构建过去几十年地球磁场并预报未来几年内地球磁场的变化.

总结起来, 数据同化已经成为地球系统科学各个分支领域所通用的一个方法论. 尽管数据同化会随着不同领域的特殊性演化出相应的特点, 但其核心的方法论是一致的, 即综合动力模型和多源观测的信息, 得到更高精度和更一致的分析结果, 并且提高模型的预报精度和可预报性.

3　模型与观测融合: 源流与方法

数据同化所蕴含的方法论体现了科学哲学思想的进化. 本节结合近现代科学哲学思想史, 对数据同化的方法追本溯源.

3.1　源流

模型与观测分别代表了近现代科学哲学中的理性主义(rationalism)和经验主义(empiricism)源流, 它们是针锋相对而最终又相得益彰的两类科学思潮. 理性主义主张唯有理性推理而非经验观察才提供了最确实的理论知识体系. 承继希腊哲学传统, 笛卡儿(René Descartes, 1596~1650)以“我思故我在”树立起理性主义的旗帜. 然而, 笛卡儿认为只有一些永恒真理(包括数学以及形而上学基础)可以单纯靠推理得到, 其余的知识需要借助生活经验以及必要的科学手段. 因此, 更准确地说笛卡儿是一位崇尚形而上学的理性主义者, 也是一位重视科学的经验主义者.

经验主义通常指相信现代科学方法, 认为理论应建立在对于事物的观察, 而不是直觉或迷信. 以培根(Francis Bacon, 1561~1626)为代表的经验主义者, 否定了人拥有与生俱来的知识的观点或不用借由经验就可以获得的知识. 经验主义开创了注重从实验中获得知识的归纳法式的现代科学认识论. 但值得注意的是, 经验主义并不主张人们可以从实务中自动地取得知识; 根据经验主义者的观点, 经由感受到的经验, 必须经过适当归纳或演绎, 才能铸成知识.

这两个思想源流的对立本身就是错误的. 康德(Immanuel Kant, 1724~1804)以后的科学哲学学者, 批判理性主义所坚持的不与经验相结合的旧式形而上学, 同时批判否定必然真理的经验主义, 强调知识是通过调和或折衷理性和经验的矛盾这一综合命题而获取. 正如康德所述, “没有理论的经验是盲目的, 没有经验的理论是空洞的”, 他强调实验加数学, 经验与理性的结合. 康德所提出的先验(a priori)、后验(a posteriori)、综合(synthesis)等概念也都已成为综合方法论的概念基石.

对纯粹理性(pure reason)和纯粹经验(pure experience)的批判, 也体现在对模型和观测的不确定性的认识, 是模型-观测综合的另一思想源流. 在地球系统科学领域, 模型与观测从来就伴生着高度的不确定性, 模型是对地球系统科学认知的形式化知识的集大成者, 但模型只是真理的近似, 还远远未臻完美. 对模型不确定性的最典型的认知莫过于“蝴蝶效应”, 模型内蕴的不确定性可以通过诞生于大气科学领域的Lorenz模型(Lorenz, 1963)来理解, 即一个完全确定性的微分方程模型, 其预报结果却可能完全是随机的, 微小到可以忽略的初值的变化会对结果造成翻天覆地的影响, 从而使得模型的可预报性显著减弱. “蝴蝶效应”深刻地揭示了模型内蕴的不确定性, 引发了混沌研究的热潮, 改变了模型预报中确定论的认识方式. “蝴蝶效应”由

气象学家洛伦茨(Edward N. Lorenz, 1917~2008)提出, 是地球科学对于数学的突出贡献. 观测的不确定性同样存在于地球观测的方方面面, 其不确定性也是内蕴的. 无数的科学实践已经告诉我们, 没有任何观测是绝对精确的, 无论是直接观测还是间接观测, 必然伴随着不确定性. 但传统上, 对器测误差研究较多, 但对由观测的时空代表性以及间接观测所使用的观测算子所引起的代表性误差关注很少, 而所有地球观测都有特定的时空代表性, 其中, 空间代表性误差是指将模型单元的模型状态映射到某一观测在其所代表性空间上的观测值, 或是将地表变量映射到遥感原始观测的误差, 它常常难以估计(李新, 2013), 这是导致模型和观测不一致的主要原因之一. 今天, 为什么存在观测误差的深刻机理还有待揭示, 但高斯在发明最小二乘法时所提出的以下建议早已成为共识: 需要依赖远超过估计数量的观测, 也需要依靠动力系统的信息, 以提高估计精度并控制观测中的误差.

总之, 模型和观测都内蕴不确定性. 针对复杂的地球系统, 是模型更可信还是观测更可信, 常常难以甄别, 这凸显了通过理性和经验的综合, 来控制和减少不确定性是多么重要.

3.2 方法

数据同化的方法论植根于现代估计理论. 贝叶斯(Thomas Bayes, 1701~1761)公式, 作为现代估计理论的一个基石, 可以被看作是康德的综合命题在概率论下的具体形式, 它蕴含了经验(观测)对于理性(模型)的概率分布的影响, 可以被解释为理性和经验的综合, 即后验信息正比于模型(先验信息)和以模型为前提条件的观测(似然函数).

现代估计理论的更可操作的方法论基石, 是由数学王子——高斯(Karl Friedrich Gauss, 1777~1855)奠定的. 在他的著作《天体运动论》中, 高斯发现并详尽阐述了最小二乘法, 其核心思想是最小化所有观测数据间、或观测数据与可能的动力函数间的欧氏距离以获得最优的数值或函数估计, 本质上是一个函数的极值问题. 最小二乘法通过对状态演进轨迹的近似认知以及远超过估计量的观测, 提高估计精度, 这构建了估计理论的基础.

随着人类认知的演进, 充分调和依靠理性推理而来的数理模型和经验观察到的观测成为更为先进的科学哲学思潮. 变分法(Calculus of Variations)的出现为观测信息融合到模型的演进轨迹中铺平了道路. 在变分法的发展历史中, 欧拉(Leonhard Euler, 1707~1783)等数学家作出了里程碑式的贡献. 变分法在最小二乘法原理的基础上, 进一步将最优的数值估计推广到最优的函数估计, 本质是求泛函的极值以获得最优函数解. 变分法是数据同化的主要方法——变分数据同化的理论基础(Talagrand和Courtier, 1987). 变分数据同化通过融合模型和观测信息, 获得无限接近于状态真实演进过程的模拟轨迹. 目前基于变分理论的三维、四维变分数据同化(图2a)是在气象和海洋业务预报中被广泛应用的现代数据同化方法.

数据同化方法的发展, 从20世纪50年代起也受到控制论的不断启发. 与最小二乘法一脉相承, 柯尔莫果洛夫(A. N. Kolmogorov, 1903~1987)和控制论创始人维纳(Wiener, 1894~1964)分别独立提出了线性最小均方估计法; 在此基础上, Kalman滤波(Kalman, 1960)最终集大成, 奠定了线性控制的基石. Kalman滤波可视为是一种更易于计算机实现的递推最小二乘法, 它在离散时间维上采用局部更新的方式动态地接收观测对系统演进的反馈, 通过逐步递推实现系统的最优. 近20多年来出现的集合(Ensemble)Kalman滤波(EnKF)(Evensen, 1994)和粒子滤波(Gordon等, 1993; Han和Li, 2008), 则在贝叶斯估计的框架下, 进一步把估计问题泛化为根据当前时刻之前所获得的所有观测信息和系统动力信息, 求得系统状态后验概率密度函数, 并引入随机动态预报理论, 解决了复杂非线性系统的数据同化问题.

从图2可以看出, 两大类现代数据同化方法——变分法和贝叶斯滤波(集合Kalman滤波、粒子滤波等)在操作和实现流程上不尽一致. 但在概率论的意义下, 两大类方法都可被视为是通过观测的反馈(似然函数)获得模型的后验概率分布, 而且变分法的目标泛函(图2a), 可以由对贝叶斯滤波求取系统状态最大后验概率(图2b)而得到. 因此, 贝叶斯估计可以被认为是各种数据同化方法共同的基础(李新和摆玉龙, 2010).

总体上, 数据同化提供了一套切实可行的方法论来实践理性主义和经验主义的调和. 它将不同来源、不同分辨率的观测数据直接或通过观测算子间接映射到地球系统模型的动力演进空间上, 同时权衡模型和观测的不确定性以减小整个数据同化系统的误差并纠

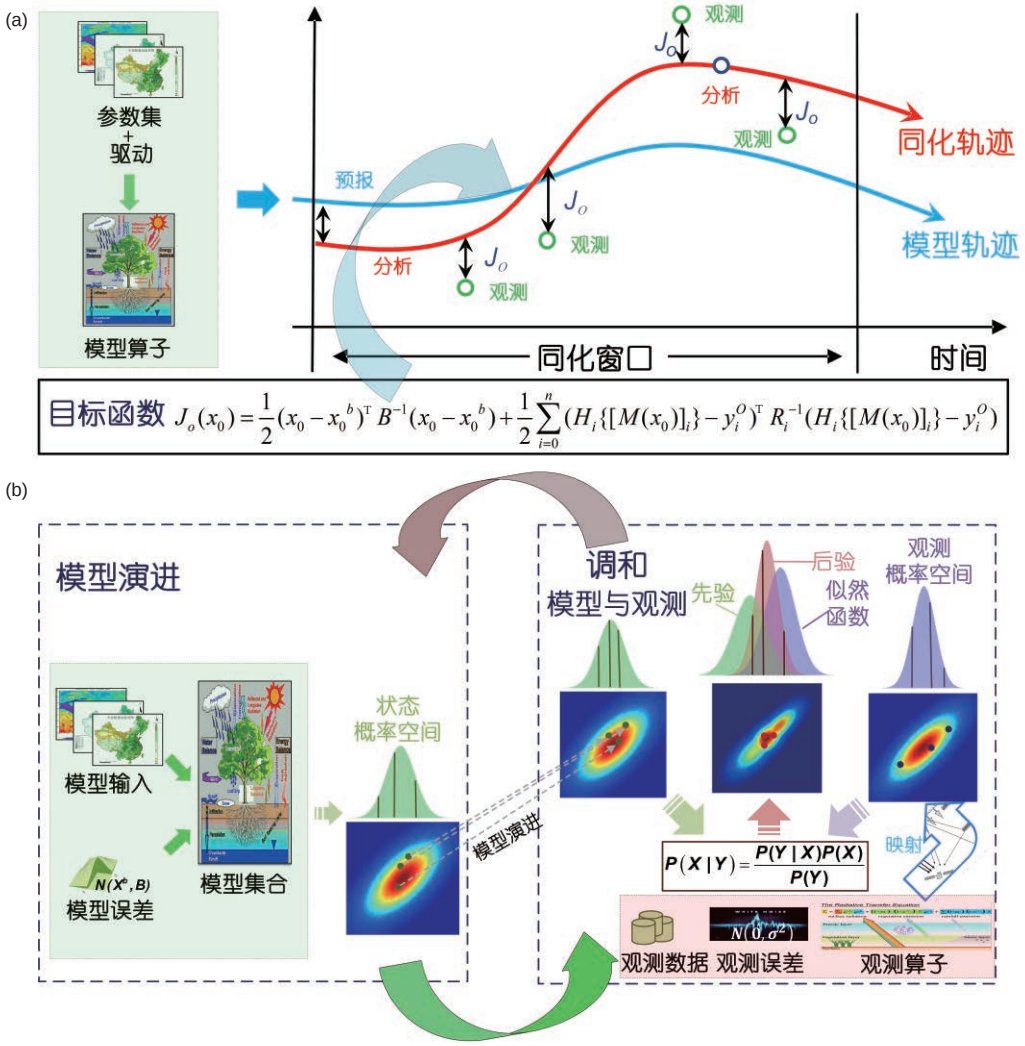

图 2　四维变分法(a)和贝叶斯滤波(b)数据同化示意

正模型的运行轨迹, 从而对地球系统的状态进行更加真实的模拟与预报(Talagrand, 1997; Li等, 2007). 模型和观测可以共奏出美妙的和弦. 数据同化在增强地球系统的可观测性和可预报性中起到了关键作用, 就是因为数据同化与现代科学哲学方法论的进化是一脉相承的, 其具体的方法则以贝叶斯理论、最小二乘法、变分法和控制论为基石.

4　国内数据同化研究进展

中国学者在在非线性非高斯贝叶斯递推滤波、代表性误差估计、变分和集合滤波方法的结合方面取得了创新进展. 针对集合同化理论和方法中的初始场误差演化问题, 穆穆研究小组研究了初始场误差的演化特征并应用到春季ENSO预报问题(Mu等, 2007; Duan等, 2009). 针对变分同化方法计算代价大以及伴随矩阵难以构建问题, Cao等(2007)提出通过特征正交分解法(proper orthogonal decomposition)对动力模型中能最大程度代表动力系统特征的物理量进行识别, 以降低动力系统的维数并从而降低变分同化的计算代价. Wang等(2010)也基于维度减少投影法(dimension reduced projection, DRP)提出了一种计算更经济的四维变分同化方法. 发展结合集合与变分数据同化方法优势的新方法是数据同化方法研究的前沿, Tian等(2011)

发展的基于本征正交分解的集合-四维变分同化方法(PODEn4DVar), 通过本征正交分解避免了切线性算子的计算, 并利用集合模拟的优势直接估计背景场误差以避免线性化假设带来的影响, 显著提高了变分数据同化方法的计算效率和精度; 他们进一步发展了非线性集合四维变分同化方法, 采用了更高效的局地化方案和多重网格方案(Tian等, 2018), 特别是为了克服局地化的巨大计算瓶颈问题, 而针对性地发展了高效的局地相关矩阵分解方法(Zhang和Tian, 2018). 近期, 数据同化理论和方法的又一新进展是基于测度论和随机微积分理论建立了模型与观测的代表性空间尺度转换理论并将其应用于数据同化中, 构建代表性误差的演进模型, 解析得到因尺度转换而产生的代表性误差的数学表达(Liu和Li, 2017).

在数据同化应用方面, 国内近10年来在大气、海洋、陆面数据系统的发展方面取得了长足的进步.

在很长一段时期内, 大气再分析数据的生产方面缺乏中国的贡献. 针对这一问题, 中国开展了中国第一代全球大气再分析资料(CAR-Interim)的生产, 并于2018年推出了10年长度(2007~2016年)的中国全球大气再分析产品, 弥补了中国在再分析数据产品方面的空白(廖捷等, 2018). 当前, 中国已经正式着手开展了40年(1979~2018年)中国全球大气再分产品(CRA-40)生产工作(王旻燕等, 2018), CRA-40同化了更多中国地区的地面站、常规探空观测资料, 将会有效弥补目前由其他机构所发布的再分析数据产品在中国区域尤其是青藏高原地区精度不理想的不足.

国家海洋环境预报中心建立了全球-大洋-近海3级尺度嵌套的全球业务化海洋预报系统(王辉等, 2016). 该系统采用变分方法和集合最优同化方法, 开发了多源数据同化模块并发展了多时空尺度数据融合同化技术, 实现了对海表面温度、卫星高度计数据和Argo全球海洋观测网中的温盐剖面数据等多源观测的协调同化, 显著提高了海洋温盐流业务化预报的水平和质量. 该系统实时发布全球多尺度、多要素的海流、海浪、海温、海冰、海面风场等预报和诊断分析产品, 实现了全球海域范围内从百公里级到公里级空间分辨率的一体化预报业务全覆盖.

在陆面数据同化方面, 国内建立了中国陆域高分辨率多源遥感数据同化系统(Li等, 2007; Huang等, 2008; Che等, 2014)、青藏高原陆面数据同化系统(He等, 2019)、流域水文数据同化系统(Han等, 2012), 在非线性非高斯贝叶斯递推滤波、代表性误差估计、多源遥感数据同化等方面实现了方法和应用突破(Han和Li, 2008; Huang等, 2016), 开发了通用多源遥感数据同化高性能计算软件系统(Liu等, 2020). 中国气象局高分辨率陆面数据同化系统(http://data.cma.cn/, 师春香等, 2011) 实时或近实时生成0.01°东亚区域土壤水分、降水、气温等关键变量数据产品, 促进了陆面数据同化的业务应用.

5 发展趋势与挑战

5.1 发展趋势

数据同化是目前地球系统科学中一个广义的方法论. 图3勾画了数据同化在地球系统科学中的发展历程. 可以看出, 无论数据同化在地球系统科学各分支中如何发展, 其归宿都是要形成统一的地球系统数据同化(Buizza等, 2018; Ruti等, 2020).

地球系统数据同化不仅需要发展多时空尺度数据同化方法, 也需要发展多源、多尺度、异构观测数据处理方法和技术, 以协调地球系统多圈层观测数据纳入统一的数据同化框架中. 此外, 高分辨率地球系统模型的运行和对地球观测系统多尺度多源数据的全面同化, 对计算资源的需求远远高于各分支学科模型的计算需求, 这也是地球系统数据同化提上实际应用所需要解决的一个难题.

5.2 挑战

(ⅰ) **广义和严谨的数据同化数学框架**. 以变分法和Bayes滤波为代表的数据同化方法在数据同化实践中取得了巨大成功, 然而, 目前还缺乏更加统一的、广义和更严谨的数学框架来作为进一步引领数据同化研究的核心理论. 这一理论首先应当通过现代数学理论对模型与观测的调和过程进行严谨的表述, 其中随机微积分、遍历理论、博弈论分别为贝叶斯滤波、蒙特卡罗集合预报与集合同化、稳健(robust)控制奠定理论基础; 其次, 数据同化的广义理论应能够广延到数据同化的各种具体方法和各个研究领域.

(ⅱ) **自然-社会系统的数据同化**. 地球已全面进入人类世, 自然-社会过程的相互作用越来越深入地影响着地球系统的演进(陈发虎等, 2019), 因此, 对“水-土-

图 3　数据同化在地球系统科学中的发展历程与方向

气-生-人”——特别是对社会科学中大量非结构化数据的同化成为挑战. 在自然-社会集成领域中, 比贝叶斯理论更加广义的DS理论(Dempster-Shafer theory)有望在自然-社会系统模型对结构化与非结构化数据的同时同化中发挥更大的作用.

(iii) **数据同化的不确定性研究**. 模型和观测误差的准确估计是调制模型-观测和弦的关键, 也是目前数据同化领域的最大挑战, 要实现模型和观测的物理和谐, 就必须突破这一重大挑战. 在这一挑战中, 尤以观测误差的估计的研究更为缓慢. 对观测误差的估计在很大程度上决定了能否更优地融入多源观测, 而观测代表性误差的“先验”本质, 决定了需要通过系统的观测试验特别是密集的多尺度试验(Li等, 2013)来理解代表性误差的统计特征和空间相关特征. 此外, 在误差的统计特性未知甚至不可知的情形下, 如何基于博弈理论——如H无穷滤波(Luo和Hoteit, 2011), 最大程度地减小系统对不确定性的敏感度, 实现稳健估计, 也是一个挑战.

(iv) **顺应大数据和人工智能时代的发展趋势**. 地球大数据洪流滚滚而来, 大数据分析在一定程度上模糊了模型与观测的界限, 对于极巨量大数据的同化(Miyoshi等, 2016)和利用大数据方法同化多源观测(Chang和Zhang, 2019; Reichstein等, 2019)都是新的挑战. 伴随大数据而来的数据洪流、超高维数据、非结构化数据等问题, 对数据同化而言既是机遇——极大丰富了观测信息, 也是挑战, 特别是“维度灾难”和“信息爆炸”. 而深度学习和人工智能等数据驱动的大数据方法可能会具备比传统物理模型更强的预报能力, 极有可能在大数据时代调和模型和观测中扮演重要角色. 从信息融合的角度, 机器学习和数据同化在数学原理上同构, 都采用Bayes估计策略(Tenenbaum, 1999; Ghahramani, 2015), 因此, 融合机器学习和数据同化的新方法可能极具潜力, 而大数据时代新的认知浪潮也可能会对建立在理性推演+实验观测范式上的传统数据同化方法论造成冲击.

6　结语

数据同化从现代科学哲学方法论思潮汲取营养, 深深植根于估计理论和控制论, 实践于地球系统科学的主要领域并因这些领域的需求而不断外延其应用. 数据同化严谨而优美的数学框架体现了理性和经验的和谐之美, 但是, 我们在本文中希望强调的一点是, 数据同化对模型和观测的调和是以对两者的批判为前提的, 从科学哲学的角度可以认为是理性批判主义(Critical rationalism)的可操作的方法论, 正如波普尔(Karl

Popper)主张的, 对理性应该采取批判的态度, 而科学理论也并不来自经验归纳, 科学理论是通过不断的证伪、反驳、批判而向前发展的.

数据同化成功地融合了先验的模型信息(理性)和大量观测信息(经验), 以概率方式调和了模型和观测, 批判式地渐进真值. 数据同化已成为地球系统科学方法论的重要组成部分, 它以新的范式改进了地球系统的可观测性和可预报性, 在地球系统科学各个领域取得了巨大的成功, 成长为地球系统科学整体及各个分支领域的一个共同方法论. 但数据同化还急需发展更加统一和严谨的数学框架, 也更需要在地球系统模型与地球观测系统的协力前行的道路上, 与它们同步进化, 在更广阔的应用领域中展现鲜活的生命力.

参考文献

陈发虎, 傅伯杰, 夏军, 吴铎, 吴绍洪, 张镱锂, 孙航, 刘禹, 方小敏, 秦伯强, 李新, 张廷军, 刘宝元, 董治宝, 侯书贵, 田立德, 徐柏青, 董广辉, 郑景云, 杨威, 王鑫, 李再军, 王飞, 胡振波, 王杰, 刘建宝, 陈建徽, 黄伟, 侯居峙, 蔡秋芳, 隆浩, 姜明, 胡亚鲜, 冯晓明, 莫兴国, 杨晓燕, 张东菊, 王秀红, 尹云鹤, 刘晓晨. 2019. 近70年来中国自然地理与生存环境基础研究的重要进展与展望. 中国科学: 地球科学, 49: 1659–1696

方苗, 李新. 2016. 古气候数据同化: 缘起、进展与展望. 中国科学: 地球科学, 46: 1076–1086

李新. 2013. 陆地表层系统模拟和观测的不确定性及其控制. 中国科学: 地球科学, 11: 1735–1742

李新, 摆玉龙. 2010. 顺序数据同化的Bayes滤波框架. 地球科学进展, 25: 515–522

廖捷, 胡开喜, 江慧, 曹丽娟, 姜立鹏, 李庆雷, 周自江, 刘志权, 张涛, 王蕙莹. 2018. 全球大气再分析常规气象观测资料的预处理与同化应用. 气象科技进展, 8: 133–142

师春香, 谢正辉, 钱辉, 梁妙玲, 杨晓春. 2011. 基于卫星遥感资料的中国区域土壤湿度EnKF数据同化. 中国科学: 地球科学, 41: 375–385

王辉, 万莉颖, 秦英豪, 王毅, 杨学联, 刘洋, 邢建勇, 陈莉, 王彰贵, 仉天宇, 刘桂梅, 杨清华, 吴湘玉, 刘钦燕, 王东晓. 2016. 中国全球业务化海洋学预报系统的发展和应用. 地球科学进展, 31: 1090–1104

王旻燕, 姚爽, 姜立鹏, 刘志权, 师春香, 胡开喜, 张涛, 张志森, 刘景卫. 2018. 我国全球大气再分析(CRA-40)卫星遥感资料的收集和预处理. 气象科技进展, 8: 158–163

Buizza R, Brönnimann S, Haimberger L, Laloyaux P, Martin M J, Fuentes M, Alonso-Balmaseda M, Becker A, Blaschek M, Dahlgren P, de Boisseson E, Dee D, Doutriaux-Boucher M, Feng X, John V O, Haines K, Jourdain S, Kosaka Y, Lea D, Lemarié F, Mayer M, Messina P, Perruche C, Peylin P, Pullainen J, Rayner N, Rustemeier E, Schepers D, Saunders R, Schulz J, Sterin A, Stichelberger S, Storto A, Testut C E, Valente M A, Vidard A, Vuichard N, Weaver A, While J, Ziese M. 2018. The EU-FP7 ERA-CLIM2 project contribution to advancing science and production of earth system climate reanalyses. Bull Amer Meteorol Soc, 99: 1003–1014

Cao Y, Zhu J, Navon I M, Luo Z. 2007. A reduced-order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. Int J Numer Meth Fluids, 53: 1571–1583

Ghahramani Z. 2015. Probabilistic machine learning and artificial intelligence. Nature, 521: 452–459

Chang H, Zhang D. 2019. Identification of physical processes via combined data-driven and data-assimilation methods. J Comput Phys, 393: 337–350

Chassignet E P, Hurlburt H E, Smedstad O M, Halliwell G R, Hogan P J, Wallcraft A J, Baraille R, Bleck R. 2007. The HYCOM (HYbrid Coordinate Ocean Model) data assimilative system. J Marine Syst, 65: 60–83

Che T, Li X, Jin R, Huang C. 2014. Assimilating passive microwave remote sensing data into a land surface model to improve the estimation of snow depth. Remote Sens Environ, 143: 54–63

Compo G P, Whitaker J S, Sardeshmukh P D, Matsui N, Allan R J, Yin X, Gleason B E, Vose R S, Rutledge G, Bessemoulin P, Brönnimann S, Brunet M, Crouthamel R I, Grant A N, Groisman P Y, Jones P D, Kruk M C, Kruger A C, Marshall G J, Maugeri M, Mok H Y, Nordli Ø, Ross T F, Trigo R M, Wang X L, Woodruff S D, Worley S J. 2011. The Twentieth Century reanalysis project. Q J R Meteorol Soc, 137: 1–28

Cummings J A. 2005. Operational multivariate ocean data assimilation. Q J R Meteorol Soc, 131: 3583–3604

Drenkard E J, Karnauskas K B. 2014. Strengthening of the Pacific equatorial undercurrent in the SODA reanalysis: Mechanisms, ocean dynamics, and implications. J Clim, 27: 2405–2416

Duan W, Liu X, Zhu K, Mu M. 2009. Exploring the initial errors that cause a significant “spring predictability barrier” for El Niño events. J Geophys Res, 114: C04022

Evensen G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J Geophys Res, 99: 10143–10,162

Forget G, Campin J M, Heimbach P, Hill C N, Ponte R M, Wunsch C. 2015. ECCO version 4: An integrated framework for non-linear inverse modeling and global ocean state estimation. Geosci Model Dev, 8: 3071–3104

Fournier A, Hulot G, Jault D, Kuang W, Tangborn A, Gillet N, Canet E, Aubert J, Lhuillier F. 2010. An introduction to data assimilation and predictability in geomagnetism. Space Sci Rev, 155: 247–291

Gelaro R, McCarty W, Suárez M J, Todling R, Molod A, Takacs L, Randles C A, Darmenov A, Bosilovich M G, Reichle R, Wargan K, Coy L, Cullather R, Draper C, Akella S, Buchard V, Conaty A, da Silva A M, Gu W, Kim G K, Koster R, Lucchesi R, Merkova D, Nielsen J E, Partyka G, Pawson S, Putman W, Rienecker M, Schubert S D, Sienkiewicz M, Zhao B. 2017. The Modern-Era retrospective analysis for research and applications, Version 2 (MERRA-2). J Clim, 30: 5419–5454

Gordon N J, Salmond D J, Smith A F M. 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proc F-Radar Signal Process UK, 140: 107

Hakim G J, Emile-Geay J, Steig E J, Noone D, Anderson D M, Tardif R, Steiger N, Perkins W A. 2016. The last millennium climate reanalysis project: Framework and first results. J Geophys Res-Atmos, 121: 6745–6764

Han X, Li X. 2008. An evaluation of the nonlinear/non-Gaussian filters for the sequential data assimilation. Remote Sens Environ, 112: 1434–1449

Han X J, Li X, Hendricks Franssen H J, Vereecken H, Montzka C. 2012. Spatial horizontal correlation characteristics in the land data assimilation of soil moisture. Hydrol Earth Syst Sci, 16: 1349–1363

He J, Zhang F, Chen X, Bao X, Chen D, Kim H M, Lai H W, Leung L R, Ma X, Meng Z, Ou T, Xiao Z, Yang E G, Yang K. 2019. Development and evaluation of an ensemble-based data assimilation system for regional reanalysis over the Tibetan Plateau and surrounding regions. J Adv Model Earth Syst, 11: 2503–2522

Hoshiba M, Aoki S. 2015. Numerical shake prediction for earthquake early warning: Data assimilation, real-time shake mapping, and simulation of wave propagation. Bull Seismol Soc Am, 105: 1324–1338

Huang C, Li X, Lu L. 2008. Retrieving soil temperature profile by assimilating MODIS LST products with ensemble Kalman filter. Remote Sens Environ, 112: 1320–1336

Huang C, Chen W, Li Y, Shen H, Li X. 2016. Assimilating multi-source data into land surface model to simultaneously improve estimations of soil moisture, soil temperature, and surface turbulent fluxes in irrigated fields. Agric For Meteorol, 230-231: 142–156

Hurlburt H, Brassington G B, Drillet Y, Masafumi K, Mounir B, Bourdalle-Badie R, Chassignet E, Jacobbs G A, Le Galloudec O, Lellouche J M, Metzger E, Oke P, Pugh T F, Schiller A, Smedstad O, Tranchant B, Tsujino H, Usui N,Walcraft A J. 2009. High-resolution global and basin-scale ocean analyses and forecasts oceanography. Oceanography, 22: 80–97

Kalman R E. 1960. A new approach to linear filtering and prediction problems. J Basic Eng, 82: 35–45

Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Leetmaa A, Reynolds R, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo K C, Ropelewski C, Wang J, Jenne R, Joseph D. 1996. The NCEP/NCAR 40-year reanalysis project. Bull Amer Meteorol Soc, 77: 437–471

Karspeck A R, Stammer D, Köhl A, Danabasoglu G, Balmaseda M, Smith D M, Fujii Y, Zhang S, Giese B, Tsujino H, Rosati A. 2017. Comparison of the Atlantic meridional overturning circulation between 1960 and 2007 in six ocean reanalysis products. Clim Dyn, 49: 957–982

Kobayashi S, Ota Y, Harada Y, Ebita A, Moriya M, Onoda H, Onogi K, Kamahori H, Kobayashi C, Endo H, Miyaoka K, Takahashi K. 2015. The JRA-55 reanalysis: General specifications and basic characteristics. J Meteorol Soc Jpn, 93: 5–48

Li X, Cheng G, Liu S, Xiao Q, Ma M, Jin R, Che T, Liu Q, Wang W, Qi Y, Wen J, Li H, Zhu G, Guo J, Ran Y, Wang S, Zhu Z, Zhou J, Hu X, Xu Z. 2013. Heihe watershed allied telemetry experimental research (HiWATER): Scientific objectives and experimental design. Bull Amer Meteorol Soc, 94: 1145–1160

Li X, Huang C, Che T, Jin R, Wang S, Wang J, Gao F, Zhang S, Qiu C, Wang C. 2007. Development of a Chinese land data assimilation system: Its progress and prospects. Prog Nat Sci, 17: 881–892

Liu F, Li X. 2017. Formulation of scale transformation in a stochastic data assimilation framework. Nonlin Processes Geophys, 24: 279–291

Liu F, Wang L, Li X, Huang C L. 2020. ComDA: A common software for nonlinear and non-Gaussian land data assimilation. Environ Model Software, 127: 104638

Lorenz E N. 1963. Deterministic nonperiodic flow. J Atmos Sci, 20: 130–141

Luo X, Hoteit I. 2011. Robust ensemble filtering and its relation to covariance inflation in the ensemble Kalman filter. Mon Weather Rev, 139: 3938–3953

Martin M J, Hines A, Bell M J. 2007. Data assimilation in the FOAM operational short-range ocean forecasting system: A description of the scheme and its impact. Q J R Meteorol Soc, 133: 981–995

McLaughlin D. 1995. Recent developments in hydrologic data assimilation. Rev Geophys, 33: 977–984

Mitchell K E, Lohmann D, Houser P R, Wood E F, Schaake J C, Robock A, Cosgrove B A, Sheffield J, Duan Q, Luo L, Higgins R W, Pinker R T, Tarpley J D, Lettenmaier D P, Marshall C H, Entin J K, Pan M, Shi W, Koren V, Meng J, Ramsay B H, Bailey A A.

2004. The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. J Geophys Res, 109: D07S90

Miyoshi T, Kunii M, Ruiz J, Lien G Y, Satoh S, Ushio T, Bessho K, Seko H, Tomita H, Ishikawa Y. 2016. “Big Data Assimilation” revolutionizing severe weather prediction. Bull Amer Meteorol Soc, 97: 1347–1354

Mu M, Xu H, Duan W. 2007. A kind of initial errors related to “spring predictability barrier” for EL Niño events in Zebiak-Cane model. Geophys Res Lett, 34: L03709

Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N, Prabhat N. 2019. Deep learning and process understanding for data-driven Earth system science. Nature, 566: 195–204

Rodell M, Houser P R, Jambor U, Gottschalck J, Mitchell K, Meng C J, Arsenault K, Cosgrove B, Radakovich J, Bosilovich M, Entin J K, Walker J P, Lohmann D, Toll D. 2004. The global land data assimilation system. Bull Amer Meteorol Soc, 85: 381–394

Ruti P M, Tarasova O, Keller J H, Carmichael G, Hov Ø, Jones S C, Terblanche D, Anderson-Lefale C, Barros A P, Bauer P, Bouchet V, Brasseur G, Brunet G, DeCola P, Dike V, Kane M D, Gan C, Gurney K R, Hamburg S, Hazeleger W, Jean M, Johnston D, Lewis A, Li P, Liang X, Lucarini V, Lynch A, Manaenkova E, Jae-Cheol N, Ohtake S, Pinardi N, Polcher J, Ritchie E, Sakya A E, Saulo C, Singhee A, Sopaheluwakan A, Steiner A, Thorpe A, Yamaji M. 2020. Advancing research for seamless earth system prediction. Bull Amer Meteorol Soc, 101: E23–E35

Steiger N J, Smerdon J E, Cook E R, Cook B I. 2018. A reconstruction of global hydroclimate and dynamical variables over the Common Era. Sci Data, 5: 180086

Swift J H, Aagaard K, Timokhov L, Nikiforov E G. 2005. Long-term variability of Arctic Ocean waters: Evidence from a reanalysis of the EWG data set. J Geophys Res, 110: C03012

Talagrand O. 1997. Assimilation of observations, an introduction. J Meteorol Soc Jpn, 75: 191–209

Talagrand O, Courtier P. 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. Q J R Meteorol Soc, 113: 1311–1328

Tenenbaum J B. 1999. Bayesian modeling of human concept learning. Cambridge: Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems II. 59–68

Tian X, Xie Z, Sun Q. 2011. A POD-based ensemble four-dimensional variational assimilation method. Tellus A-Dyn Meteorol Oceanogr, 63: 805–816

Tian X, Zhang H, Feng X, Xie Y. 2018. Nonlinear least squares En4DVar to 4DEnVar methods for data assimilation: Formulation, analysis, and preliminary evaluation. Mon Weather Rev, 146: 77–93

Uppala S M, KÅllberg P W, Simmons A J, Andrae U, Bechtold V D C, Fiorino M, Gibson J K, Haseler J, Hernandez A, Kelly G A, Li X, Onogi K, Saarinen S, Sokka N, Allan R P, Andersson E, Arpe K, Balmaseda M A, Beljaars A C M, Berg L V D, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Hólm E, Hoskins B J, Isaksen L, Janssen P A E M, Jenne R, Mcnally A P, Mahfouf J F, Morcrette J J, Rayner N A, Saunders R W, Simon P, Sterl A, Trenberth K E, Untch A, Vasiljevic D, Viterbo P, Woollen J. 2005. The ERA-40 re-analysis. Q J R Meteorol Soc, 131: 2961–3012

Wang T, Jin X, Huang Y, Wei Y. 2017. Real-time 3-D space numerical shake prediction for earthquake early warning. Earthq Sci, 30: 269–281

Wang B, Liu J, Wang S, Cheng W, Juan L, Liu C, Xiao Q, Kuo Y H. 2010. An economical approach to four-dimensional variational data assimilation. Adv Atmos Sci, 27: 715–727

Xia Y, Mitchell K, Ek M, Sheffield J, Cosgrove B, Wood E, Luo L, Alonge C, Wei H, Meng J, Livneh B, Lettenmaier D, Koren V, Duan Q, Mo K, Fan Y, Mocko D. 2012. Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. J Geophys Res, 117: D03109

Zhang F, Weng Y, Gamache J F, Marks F D. 2011. Performance of convection-permitting hurricane initialization and prediction during 2008–2010 with ensemble data assimilation of inner-core airborne Doppler radar observations. Geophys Res Lett, 38: L15810

Zhang H, Tian X. 2018. An efficient local correlation matrix decomposition approach for the localization implementation of ensemble-based assimilation methods. J Geophys Res, 123: 3556–3573

Zuo H, Balmaseda M A, Mogensen K. 2015. The new eddy-permitting ORAP5 ocean reanalysis: Description, evaluation and uncertainties in climate signals. Clim Dyn, 49: 791–811

(责任编委: 陈发虎)