



地球信息科学基础丛书

Geo-information Science

数据同化算法研发与实验

(附算法程序)

马建文 等 著



同化算法



驱动模型



科学出版社

013028087

TP751.1

14

地球信息科学基础丛书

数据同化算法研发与实验

(附算法程序)

马建文 等 著

本书出版由下列项目经费支持

- 国家重大科技基础设施工程项目“航空遥感系统”
[编号: ZGHG-2009-1-1]
- 国家 973 项目“陆表生态环境要素主被动遥感协同反演理论与方法”
[编号: 973-2007CB714406]



TP751.1

14

科学出版社

北京



北航

C1636971

内 容 简 介

本书基于国内外相关研究和作者在数据同化领域的研究心得,重点介绍数据同化算法的研发与实验。围绕数据同化算法研发和实验这一主题,本书从陆面数据同化理论和陆面过程模型切入,提炼数据同化的“一个框架、四个基本要素”架构,在选择可变渗透能力模型(VIC模型)的基础上开展数据同化算法实验,实现了三维变分算法、四维变分算法和集合卡尔曼滤波算法等三种经典数据同化算法,以及粒子滤波算法和层状贝叶斯方法等两种智能数据同化算法的研发与实验,并且通过具体实例详细介绍各个算法的研发、实验步骤以及算法结果的分析评价。最后,介绍了作者课题组开发的数据同化集成软件系统。

本书旨在对从事陆面数据同化研究的工作者提供入门参考和思路借鉴;同时,也适合定量遥感、全球环境变化及地球系统科学等领域的科研工作者以及高等院校师生参考。

图书在版编目(CIP)数据

数据同化算法研发与实验(附算法程序)/马建文等著. —北京:科学出版社, 2013

(地球信息科学基础丛书)

ISBN 978-7-03-037004-4

I. ①数… II. ①马… III. ①遥感数据—数据处理 IV. ①TP751.1

中国版本图书馆 CIP 数据核字(2013)第 045823 号

责任编辑:彭胜潮 沈睿媛/责任校对:朱光兰
责任印制:钱玉芬/封面设计:王浩

科学出版社出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

源海印刷有限责任公司印刷

科学出版社发行 各地新华书店经销

*

2013年3月第一版 开本:787×1092 1/16

2013年3月第一次印刷 印张:14 1/2 插页:4

字数:320 000

定价:58.00元

(如有印装质量问题,我社负责调换)

本书作者

马建文 秦思娴 王皓玉 张红娟

毕海芸 布日古都 尤红建 王思远

前 言

全球变化主要是指人类活动对地球环境的影响超过自然变化的强度和速率,对未来生存环境产生影响,当前全球变化主要表现为全球变暖。应对全球变暖必须回答以下几个方面的根本问题:如何确定自然与社会变化的临界点?当自然与社会发生急剧变化,人类会面临怎样的风险?未来全球温度提升 $3\sim 4\text{ }^{\circ}\text{C}$,人类该如何生存与适应?如何保护和共享地球自然资源以及生态系统以造福人类?如何在资源能源可持续发展前提下解决粮食安全问题并消除贫困?

为了实现对全球变化进行长期观测和研究,一系列全球范围的观测计划相继被提出,其中以美国国家航空航天局(National Aeronautics and Space Administration, NASA)从1999年开始实施的地球观测卫星计划(Earth Observing Satellites, EOS)和欧洲航天局(European Space Agency, ESA)从2003年开始构建的全球环境及安全监测系统计划(Global Monitoring for Environment and Security, GMES)作为主要代表,这两个对地观测计划的实施标志着全球变化研究进入到以探测全球变化和地球系统科学研究的新阶段。与此同时,多种卫星平台多载荷多种新的遥感器获取数据、地面系统数据同化、模拟计算和多要素全球动画展示等技术都取得了快速进展,开拓了人类认识地球的新技术体系,将传统的地球科学拓展到地球系统科学,是地球科学研究发展历史的里程碑。2005年11月召开的第三次国际地球观测峰会,通过了建立全球综合地球观测系统(Global Earth Observation System of Systems, GEOSS)和发布的对地观测数据共享的宣言与十年计划,开启了全球联合共同观测和模拟研究以应对全球气候变化的新阶段。2012年国际科联(International Council of Scientific Unions, ICSU)发布未来地球十年计划和宣言,呼吁科学界、企业和政府共同关注人口增长与可持续发展,强调全球环境变化研究应当与粮食安全、资源能源安全和环境安全相关联。近年来,国际遥感会议也围绕全球变化专门设立专题,如地球过程与气候变化的动态过程模拟、数据同化、地球综合观测系统和新卫星计划。

全球变化与区域可持续发展研究对我国具有非常重要的现实意义。我国基本国情是人口多、底子薄,用世界7.9%的耕地和6.5%的淡水资源养活着世界近20%的人口,全球变暖对于支撑我国经济社会发展的脆弱生态环境所产生的影响还具有一定的不确定性,应对全球变暖也是我国在2020年全面实现小康社会面临的突出问题和重大挑战。问题就是机遇,迎接挑战才能取得重大成果。为了应对全球变暖对我国经济和社会发展所产生的影响,我国积极参与国际GEOSS计划,并签署了ICSU地球未来宣言,同时在“十一五”和“十二五”中,国家对空间对地观测科技基础设施的投入也以超过20%的速率增加,空间对地观测科技整体能力得到了大幅度提升,空间对地观测技术在气象、国土、测绘、农林、水利、海洋、地震、城市与健康、灾害等方面的应用也在迅速拓展,研究与应用水平不断提升,并且逐步建立了国际化、标准化的运行平台系统与数据产品

生产体系, 不断向形成完整空间信息价值链的方向迈进。

模拟全球变化的动态过程, 开展数据同化研究生成多种参量产品在支持区域和全球可持续发展方面具有普遍的应用价值和需求。首先从应用角度分析, 数据同化现在已经被广泛应用于大气、海洋和陆面系统或耦合系统的参数与状态量分析; 其次从技术层分析, 数据同化在全球变化研究中具有不可或缺性, 主要体现在以下两个方面:

(1) 数据同化可以将新的观测数据引入过程模型中, 不断减少或者滤掉过程模型的噪声, 使得模型模拟预测轨迹更加接近自然界的真实情况。

(2) 由于生物/物理参量之间通常存在复杂的函数关系, 很难直接观测, 通过数据同化可以获得许多没有观测技术条件的状态参量, 如根区的土壤水分。

目前, 国际上主流的数据同化系统共享平台中 NASA 的北美与全球陆面数据同化系统 (GLDAS、NLDAS) 具有典型性。NASA-GSFC (Goddard Space Flight Center) 开发的 Giovanni 系统中嵌入了 GLDAS 系统与 NLDAS 系统以及相应的网络共享模型库和数据库。NLDAS 和 GLDAS 支持 HDF、netCDF、ASCII 等多种数据格式, 其中 kmz 输出格式产品可以直观地在 Google Earth 上查看。NLDAS 和 GLDAS 数据同化系统使用了 MOSAIC 模型、NOAH 模型、SAC 模型和 VIC 模型等四种陆面过程模型, 数据同化方法主要采用了三维变分方法和集合卡尔曼滤波方法等算法, 其目标是利用数据同化算法将地面观测数据和卫星遥感观测数据融入到陆面过程模型, 提高各陆面过程模型模拟与预测的轨迹, 获得北美地区 $1/8^\circ$ 分辨率和全球范围 $1/4^\circ$ 分辨率的多变量、高精度、高时空分辨率的陆面过程参数。

虽然国外的网络共享平台技术系统具有先进性、引领性和参照性, 但是由于数据同化是个很复杂的过程, 包含陆面过程模型、数据同化算法、观测数据以及驱动陆面过程模型运行的基础参量等数据, 必须根据研究区域的实际情况和观测数据设置陆面过程模型和同化算法参数, 因此需要在具体的研究区域针对具体的问题和数据特点开发相应的数据同化系统。我们研究团队在国家发改委“十二五”建设的国家重大科技基础设施工程项目“航空遥感系统”项目支持下, 以 GLDAS、NLDAS 数据同化系统作为参照, 针对国家航空遥感系统项目要求, 将要载荷 11 种传感器, 按项目设计数据同化产品要求, 自主开发了基于 VIC 水文过程模型的数据同化系统, 并生成相应的地表参量产品。

本书的定位在于数据同化算法的研发与实验, 这样定位主要出于以下两方面的考虑: 首先, 数据同化是个非常复杂的过程, 在研究过程中必须明确定位, 重点突出; 其次, 当前主流的三维变分算法、四维变分算法和集合卡尔曼滤波算法等传统的经典数据同化算法在实际应用中还存在某些不足, 需要引入新的数据同化算法。为此, 我们首先自主开发了三维变分算法、四维变分算法和集合卡尔曼滤波算法等传统的经典数据同化算法。同时, 针对这些经典算法存在的问题, 我们还开发了两种现代智能方法, 粒子滤波方法和层状贝叶斯方法, 通过实验证实了这两种算法用于数据同化的可行性和可靠性。最后, 将上述算法集成到软件系统中, 形成了完整的数据同化系统。IGARSS 2012 年德国慕尼黑会议与将要在墨尔本召开的 IGARSS 2013 年会议都设立了数据同化专题会场和分会场交流, 就数据同化的主要进展和问题展开讨论。总体上看, 当前数据同化算法方面没

有突破性的进展,相比之下,我们的研究处于较为全面和领先位置。

数据同化算法研发的实践证明,同化算法程序是从文献中找不到的,通过合作也“引进”不来,必须走自主开发的道路。为此,我们耗费了大量的人力和时间,通过大量的实验才完成了算法的开发。为了减少同水平重复工作,我们决定将主要算法源代码公开,为从事相同工作的科技人员提供参考,共同努力为普及数据同化方法和提高研究水平作点力所能及的贡献。同时,我们的研究结果进一步证实了当前数据同化算法存在的问题,因此还需要将数学领域新的研究成果引入到数据同化领域中,不断推动数据同化算法的发展。

全书分为7章。第1章绪论由马建文完成,首先从以下三个方面说明我们开展数据同化研究和算法开发的重要性和必要性:①全球环境变化与地球系统科学研究对数据同化研究的需求背景;②“大数据”的同化技术和数据同化产品在应对全球变暖、气候变化、全球变化等全球性问题的基础地位,以及新算法研究在数据同化研究领域的前沿探索地位;③国家重大科技基础设施工程项目“航空遥感系统”的数据同化系统研发任务。同时,本章还概括总结了数据同化的基本理论框架。数据同化技术不同于一般的遥感数据反演或特征提取技术,而是帮助研究地球系统复杂过程的工具,经过大量的资料调研和欧美等国家的数据同化系统剖析后,我们将数据同化系统构成概括为“一个框架和四个基本要素”:一个框架是陆面过程驱动模型,四个基本要素是陆面过程驱动模型、数据同化算法、观测数据和驱动陆面过程模型运行的基础参量。根据这个提炼,我们定位在数据同化算法的研究,四个基本要素的划分在我们算法实验边界条件、数据分析和实验结果改进方面提供了极大的便利。

第2章综述了当今国内外同化算法研发现状,从中梳理出我们的研发主线和应当注意的问题,本章由马建文和秦思娴完成。首先从大量的资料调研中梳理出一条同化算法发展进步主线,基本确定数据同化算法的发展趋势是从简单到复杂,从数学方法到智能方法,由考虑驱动模型、观测数据与同化算法结合,到全面考虑模型与算法的组合优化、数据之间的时间/空间和物理一致性问题。根据所确定趋势从而确定我们以三维变分算法、四维变分算法、集合卡尔曼滤波算法、粒子滤波算法和层状贝叶斯方法作为我们重点要进行研究的数据同化算法。

第3章首先介绍了当前主要的陆面过程驱动模型及其发展历史,重点介绍了我们所选择的VIC水文过程模型,主要包含为什么选择VIC水文过程模型、VIC水文过程模型的原理、VIC水文过程模型运行所需要的数据准备、对VIC水文过程模型用于数据同化的应用性改进,并通过实验结果进行了分析和说明。本章中VIC水文过程模型的应用性改进由王皓玉完成,由秦思娴、王皓玉、张红娟和毕海芸完成VIC水文过程模型运行所需要的基础参量数据准备,文字部分由秦思娴整理完成。

第4章~第6章重点介绍数据同化算法的研发与实验过程和结果。

第4章重点介绍了三维变分算法、四维变分算法、集合卡尔曼滤波算法三种数据同化算法开发。这三种算法是已经在运行的数据同化系统中普遍采用的算法,也是经典的数据同化算法。但是算法软件需要立足自主开发。本章介绍了这三种算法的同化算法理论和步骤、技术路线概括、程序实现以及开展的实验结果分析。其中三维变分算法和四

维变分算法的算法研发由王皓玉完成, 集合卡尔曼滤波算法由秦思娴完成, 中国科学院电子所尤红建研究员在算法开发过程中发挥了积极的指导作用。

第 5 章~第 6 章介绍了数据同化中的现代智能方法, 由于算法理论、算法程序和实验体量比较大, 又分为粒子滤波算法 I 和层状贝叶斯网络算法 II。粒子滤波同化算法和层状贝叶斯网络算法是当前数据同化算法前沿研究中的两个主流方向。

第 5 章粒子滤波数据同化算法 I 主要由张红娟完成。我们团队在粒子滤波用于小目标跟踪方面有一篇博士论文和若干文章的积累, 为粒子滤波同化算法研究提供了理论基础。由于有秦思娴和王皓玉的前期工作经验累积, 再加之张红娟良好的数学基础, 粒子滤波同化算法理论研究和程序开发进行比较顺利。在粒子滤波同化算法开发过程中, 张红娟、秦思娴、王皓玉一同开展了粒子滤波同化算法的理论分析、算法程序开发, 咬定问题不放松, 集中时间攻关, 通过多次细致的算法与模型分析实验, 以实验结果出现的问题作为分析线索逐步回溯, 发现问题所在, 逐步实现粒子滤波同化算法与 VIC 水文过程模型的组合优化, 取得了较为满意的结果。后来加入的毕海芸也参与了粒子滤波同化算法的部分研究工作。

第 6 章的层状贝叶斯网络算法 II 主要由秦思娴完成。根据我们团队在贝叶斯网络方法研究方面的长期积累和近期文献调研得知, 层状贝叶斯方法在时空分析与不确定性推理方面表现出了极大的潜力, 这两点对于数据同化来说尤其重要, 同时这也是其他算法不具备的特性。在挖掘层状贝叶斯方法方面, 秦思娴作出了突出的贡献, 首先, 细致完整地推导了层状贝叶斯方法的理论公式, 在此基础上提炼出实现算法的技术流程; 其次, 将技术流程与 VIC 水文过程模型连接, 形成组合优化程序; 最后, 经过一年多埋头工作, 通过反复多次试验和软件调整的艰苦过程, 取得较好的创新成果。秦思娴的论文“Construction and Experiment of Hierarchical Bayesian Network in Data Assimilation”在 2012 年得到 *International Journal of Remote Sensing* 高度评价, 称“A very nicely written paper with significant value to the research community, all-in-all a pleasure to read a high quality research”。秦思娴的研究工作为本书增加了学术亮点。

第 7 章数据同化集成系统主要介绍了数据同化算法集成软件系统, 以及系统各种功能模块的使用方法。由项目主要参加人员共同参与, 王皓玉“操刀”开发完成。数据同化算法集成软件系统开发过程中, 以国外主要网络数据同化系统作为参照, 重点在于将 VIC 水文过程模型、基础参量与参量准备、开发完成的同化算法、观测数据与数据处理、系统界面和可视化功能的集成, 其中在结果可视化展示方面, 我们增加了图像和空间点曲线两种可视化功能, 使处理结果可以更直观的展现。

在完成国家重大科技基础设施工程项目“航空遥感系统”的数据同化系统研发任务过程中, 秦思娴完成博士论文, 王皓玉、张红娟完成硕士论文, 顺利毕业, 毕海芸开展了博士论文研究, 他们的论文内容和研究的中间结果都尽可能地呈现在本书中。在完成论文的过程中, 他们还在国内外刊物上发表了多篇文章, 国内外的评审专家提出了许多修改意见和建议, 这些意见和建议对于提高他们的毕业论文质量起到了不可或缺的作用, 同时也为完善本书的内容和提高本书的质量起到了非常重要的作用。本书在全球变化与区域可持续发展方面的内容重点参考了布日古都博士论文研究成果。

《数据同化算法研发与实验》是继 2000 年《遥感数据智能处理与程序设计》(第二版)与 2010 年《遥感数据自动化处理与程序设计》之后的又一本奉献给遥感界的专著。这三本专著是我们团队在过去 10 余年里,紧紧围绕遥感数据处理智能化、自动化与同化等新的发展方向开展了一系列的研究,并且将我们的理论、算法和程序成果和深入思考等整理出版,与同行进行交流共享和共进,共同推动遥感数据处理的发展。

《数据同化算法研发与实验》是国家重大科技基础设施工程项目“航空遥感系统”的成果,感谢项目组长张兵研究员、课题负责人王思远研究员、课题主要成员尤红建研究员、黄俊副研究员和李利伟副研究员对完成本书的支持。感谢国家 973 项目“陆表生态环境要素主被动遥感协同反演理论与方法”(编号:973-2007CB714406)给予的出版经费支持。

马建文

2012 年 7 月 26 日

目 录

前言

第 1 章 绪论	1
1.1 全球变化研究与数据同化	1
1.2 数据同化基本构成	3
1.3 数据同化算法分类	5
1.4 陆面数据同化研究进展	6
1.5 遥感数据同化研究进展	8
1.6 本书主要内容	8
1.7 本章小结	9
主要参考文献	10
第 2 章 数据同化算法发展与进步	11
2.1 变分方法	11
2.1.1 三维变分算法	11
2.1.2 四维变分算法	12
2.2 卡尔曼滤波算法	13
2.3 集合卡尔曼滤波算法	15
2.4 粒子滤波算法	16
2.5 层状贝叶斯方法	17
2.6 数据同化算法基本公式、机制与特点	20
2.7 本章小结	21
主要参考文献	21
第 3 章 过程模型选择与应用改进	24
3.1 陆面过程模型发展阶段	24
3.2 陆面过程模型比较与选择	25
3.3 VIC 水文过程模型原理与应用改进	28
3.3.1 VIC 水文过程模型原理	28
3.3.2 VIC 水文过程模型代码移植与编译	30
3.3.3 VIC 水文过程模型应用改进	31
3.4 VIC 水文过程模型基础参量准备	32
3.4.1 大气驱动数据	32
3.4.2 土壤参数	33
3.4.3 植被参数	34
3.4.4 全局参数	35
3.4.5 基础参量与数据来源	36
3.5 VIC 水文过程模型数据准备与程序代码	37

3.6	VIC 水文过程模型运行与校验	37
3.6.1	VIC 水文过程模型运行	38
3.6.2	VIC 水文过程模型校验	38
3.7	VIC 水文过程模型实验	38
3.7.1	VIC 水文过程模型实验一	39
3.7.2	VIC 水文过程模型实验二	41
3.7.3	VIC 水文过程模型实验三	45
3.8	本章小结	47
	主要参考文献	47
第 4 章	经典数据同化算法开发与实验	49
4.1	三维变分算法	49
4.1.1	算法原理	49
4.1.2	算法流程	51
4.1.3	算法实现	52
4.2	三维变分算法同化实验	52
4.3	四维变分算法	53
4.3.1	算法原理	53
4.3.2	算法流程	54
4.3.3	算法实现	55
4.4	四维变分算法同化实验	55
4.5	集合卡尔曼滤波算法	55
4.5.1	算法原理	56
4.5.2	算法流程	58
4.5.3	算法实现	59
4.6	集合卡尔曼滤波算法同化实验	59
4.6.1	实验一 站点观测数据与 VIC 水文过程模型数据同化	59
4.6.2	实验二 微波亮温数据与 VIC 水文过程模型数据同化	66
4.7	本章小结	74
	主要参考文献	75
第 5 章	现代智能数据同化算法 I：粒子滤波算法	77
5.1	粒子滤波算法理论基础	77
5.1.1	贝叶斯滤波基本原理	78
5.1.2	粒子滤波算法原理	80
5.2	重要性采样	80
5.2.1	贝叶斯重要性采样	80
5.2.2	序贯重要性采样	82
5.3	粒子退化与重采样	84
5.3.1	粒子退化	84
5.3.2	重采样	85
5.4	粒子滤波算法流程与实现	86
5.4.1	粒子滤波算法流程	87

5.4.2 粒子滤波算法实现	88
5.5 粒子滤波算法同化实验	88
5.5.1 实验一 站点观测数据与 VIC 水文过程模型数据同化	88
5.5.2 实验二 微波亮温数据与 VIC 水文过程模型数据同化	95
5.5.3 实验三 数据同化与 VIC 水文过程模型参数同步估计	98
5.6 本章小结	104
主要参考文献	105
第 6 章 现代智能数据同化算法 II：层状贝叶斯网络算法	107
6.1 层状贝叶斯方法理论基础	108
6.1.1 数据模型	109
6.1.2 过程模型	110
6.1.3 参数模型	111
6.1.4 贝叶斯推理	111
6.2 层状贝叶斯网络算法	112
6.2.1 数据描述	112
6.2.2 层状贝叶斯网络构建	113
6.2.3 层状贝叶斯网络结构	117
6.2.4 层状贝叶斯网络学习、校验与预测	118
6.2.5 层状贝叶斯网络算法流程	119
6.3 基于多尺度回归模型的层状贝叶斯网络算法	120
6.3.1 基于多尺度回归模型的层状贝叶斯网络构建	121
6.3.2 基于多尺度回归模型的层状贝叶斯网络结构	124
6.3.3 基于多尺度回归模型的层状贝叶斯网络推理	125
6.3.4 最大似然参数估计	129
6.3.5 基于多尺度回归模型的层状贝叶斯网络算法技术流程	132
6.4 层状贝叶斯网络同化站点观测与 VIC 水文过程模型数据实验	133
6.4.1 数据预处理	133
6.4.2 层状贝叶斯网络学习与校验	135
6.4.3 层状贝叶斯网络预测	139
6.4.4 程序代码	140
6.5 本章小结	141
主要参考文献	141
第 7 章 数据同化集成系统	143
7.1 系统结构与功能设计	143
7.2 系统详细设计	145
7.2.1 输入输出模块	146
7.2.2 陆面过程模型模块	146
7.2.3 数据同化算法模块	147
7.2.4 数据可视化模块	148
7.2.5 精度评价模块	149
7.3 系统功能实现与界面	149
7.3.1 数据同化功能与界面	149

7.3.2 数据可视化功能与界面	156
7.3.3 精度评价功能与界面	158
7.4 本章小结	159
主要参考文献	159
附录一 VIC 水文过程模型与数据准备的 C/C++代码	160
附录二 三维变分算法(3DVAR)开发 C++代码	174
附录三 四维变分算法(4DVAR)开发 C++代码	183
附录四 集合卡尔曼滤波算法(EnKF)开发 C++代码	190
附录五 粒子滤波算法(PF)开发 C++代码	202
附录六 层状贝叶斯网络算法(HBN)开发 WinBUGS 和 Matlab 代码	210
彩图	

第1章 绪 论

近年来,全球环境变化对人类生存影响日益突出,为了加强对陆地-大气-海洋全球变化进行长期观测和研究,一系列全球对地观测计划相继被提出,其中以1999年12月开始实施的地球观测卫星计划(EOS)和2003年开始构建的全球环境及安全监测系统计划(GMES)作为主要代表,标志着全球变化研究进入一个以探测全球变化为目标的数据获取、数据同化和模拟计算的新阶段。2005年11月召开的第三次国际地球观测峰会,通过了建立全球综合地球观测系统(GEOSS)共同应对全球气候变化,提升综合、定量处理对地观测数据水平和推动对地观测数据共享的宣言。2012年国科联(ICSU)在伦敦举行会议,会议以“压力下的地球”为主题,以“新的科学知识实现全球可持续发展挑战和提出解决方案”为聚焦,提出“未来地球10年计划”。

作为GEOSS计划的重要成员国之一,我国积极响应全球环境变化研究,采取了一系列措施。2008年我国发射风云三号卫星,载荷了11种高性能传感器和探测仪器,可以实现对大气、海洋和陆地协同观测,能够得到40多种要素产品;我国“十二五”建设的“航空遥感系统”将载荷11种先进遥感器,获取区域地表、大气和水体变化参量;同时,中国卫星遥感地面站存有1986年以来的各类对地观测卫星数据资料,总计达250余万景、250TB,是我国最大的对地观测卫星数据历史档案库。如何有效利用这些观测数据成为当前全球环境变化研究的关键问题,通过数据同化充分利用这些数据和信息产品,有助于模拟预测全球环境变化和开展地球系统研究。相比于传统的遥感反演方法,数据同化将观测数据与地球动力模型有效地结合起来,成为连接观测数据与驱动模型预测的关键桥梁,得到更加客观的结果,而传统的反演算法则是对单一来源、单一时刻的遥感数据进行处理,反演结果的质量直接由遥感数据和反演算法决定。全球环境变化研究的迫切性和数据同化本身的优越性极大地带动了数据同化的发展,数据同化算法在不断完善过程中也在引入新的数学研究成果,展示出创新发展的潜力。

1.1 全球变化研究与数据同化

由于人类活动的影响,21世纪温室气体和硫化物气溶胶的浓度急速增加,据科学家预测,未来50~100年人类将完全进入一个变暖的世界。未来100年,全球、东亚地区和我国的温度将迅速上升,全球平均地表温度将上升1.4~5.8℃,到2050年,我国平均气温将上升2.2℃,全球气候升温,地下水水位下降,粮食减产,从而造成海平面升高,沿海地区被淹没,生态环境退化,给人类生存条件带来潜在的影响。我国基本国情是人口多、底子薄,用世界7.9%的耕地和6.5%的淡水资源养活世界近20%的人口。全球变暖对于支撑我国经济社会发展的脆弱生态环境所产生的影响还具有一定的不确定性,应对全球变暖是我国在2020年全面实现小康社会面临的重大挑战之一。在全球变暖背景

条件下,利用尽可能多的空间对地观测资料,增加对全球变化和区域可持续发展的科学认识,为帮助国家解决好我国经济社会发展与生态环境,资源能源出现的“不平衡、不协调和不可持续”区域可持续发展问题提供支撑。

为了监测并预测全球环境变化,以1999年12月开始实施的地球观测卫星计划和2003年开始构建的全球环境及安全监测系统计划作为标志,全球环境变化研究进入一个以探测全球变化为目标的数据获取、数据同化和模拟计算的新发展阶段。2005年11月召开的第三次国际地球观测峰会,通过了建立全球综合地球观测系统共同应对全球气候变化、提升综合定量处理对地观测数据水平和推动对地观测数据共享的宣言。我国是全球综合地球观测系统的主要成员,与其他成员国提供空间对地观测数据共享服务,携手共同应对全球变化和区域可持续、均衡发展问题。我国各种分辨率的空间对地观测数据已经积累了40多年,对地观测平台已经发展到七种系列,包括气象卫星系列、海洋卫星系列、中巴地球资源卫星系列、灾害与环境监测预报卫星星座系统、探月工程——嫦娥系列、北京1号小卫星和天宫一号高光谱数据,并且与多个拥有对地观测卫星的国家签署卫星数据获取协议或使用协议,具备了接收超过15颗对地观测卫星的能力,包括密云、喀什和三亚等地面接收站网。IGARSS 2012年德国慕尼黑学术会议将重点聚焦在应对全球气候变化,专门设立了四个科学主题:地球过程动力和气候变化、数据同化、综合地球观测系统和新卫星计划。为应对全球变化与区域可持续、均衡发展问题提供可靠和持续的技术支持是地球科学和空间对地观测领域需要突破的技术难点和长期任务,需要共同努力完成。

从需求层面看,全球变化动态过程模型和参量估计在支持区域可持续发展方面具有需求的普遍性。我国对全球变暖资料产品和核心关键技术突破的需求主要来自三个方面:为国家应对全球变暖决策提供科技支撑;为不同部门运行系统提供与全球变暖背景相关的信息产品,改善运行系统质量;国家制定与全球环境变化相关的重大科技计划时,需要对全球变暖信息产品和信息平台提出需求。从应用层面看,数据同化适合于大气、海洋和陆面系统或耦合模型驱动系统的参数估算与分析,也是我国制定的国家全球变化相关重大科技计划的关键技术攻关内容。为此,“十二五”期间需要加强相关实验室和大学的联合攻关,突破关键核心技术,加快将我国低端数据产品服务提升到高端信息产品和服务能力,迈向从数据产品服务提升到模型预测决策信息服务的新台阶,主要包括以下两点。

(1) 在网络数字地球科学平台的基础上增加数据同化系统,实现过程模型、同化算法、驱动数据、输入数据和预测结果的网络化共享。数据同化是在陆面、海面和大气过程模型或耦合模型框架基础上完成的,通过数据同化可以修正过程模型的模拟预测轨迹,也可以获得其他不能直接观测或当前不具备观测条件的参量。

(2) 按国际全球变化建议的54种基础参量产品,开发和引进评价信息产品,生成物理模型和统计模型,对信息产品加工流程的每一个环节进行检验。空间对地观测和反演的五十多个基础参量涉及地球系统的大气、海洋和陆地等系统,包括大气温度、风速风向、水汽、压力、降雨、表层辐射通量、温度、云性质、太阳辐射通量、二氧化碳、甲烷、臭氧和气溶胶、其他长寿命温室气体等大气参数,海洋表层温度、盐度、水位、海浪、浮冰、表流、水色、二氧化碳颗粒压强、酸度等海洋参数,径流、反照度、土壤水分、雪水当量等地表参数。

陆地是地球系统的重要组成部分,陆面过程模型模拟地表的生物/物理变化过程,加深人们对地表变化的理解,是全球环境变化的重要模型之一。数据同化作为检验、改善陆面过程模型的重要方法,在全球变化研究中具有极大的重要性,主要表现在以下两方面:

(1) 数据同化利用观测数据不断降低过程模型模拟过程中的不确定性,使得模型模拟预测结果更接近真实情况。

(2) 观测的物理参量与其他地表状态参量之间通常有复杂的函数关系,通过数据同化可以获得许多没有技术条件观测到的状态参量。例如,土壤水分对于全球的能量和水平平衡起到重要作用,但是通常我们只能获得表层的土壤水分,对于根区和底层的土壤水分必须通过不同深度之间的土壤水分函数关系计算得到。又如海水盐度也是海洋过程模型中不可或缺的物理参量,但是大尺度区域的海水盐度需要从海流、海冰和温度等状态中估算。

为了推动全球变化监测数据共享,国际上相继开展了一系列同化运行系统研发,实现模型、算法和数据共享的研究。其中,以美国国家航空与航天局哥达德空间飞行中心、美国国家海洋大气局的国家环境预报中心和水文科学部、普林斯顿大学、华盛顿大学开发的北美陆面数据同化系统(North-America Land Data Assimilation System, NLDAS; Luo et al., 2003)和全球陆面数据同化系统(Global Land Data Assimilation System, GLDAS; Rodell et al., 2004)为典型代表。NLDAS和GLDAS数据同化系统集成了MOSAIC模型、NOAH模型、SAC模型和VIC模型等四种陆面过程模型,主要利用三维变分算法和集合卡尔曼滤波算法进行数据同化,旨在提供北美和全球范围多种参量、高时空分辨率的陆面过程参量产品。

1.2 数据同化基本构成

由于数据同化可以应用于地球系统科学研究的多个领域,因此不同领域专家对数据同化的内涵与外延有各自的表述(Robinson and Lermusiaux, 2000; 黄春林和李新, 2004; 李新和黄春林, 2004; 宫鹏, 2009; 李新和摆玉龙, 2010)。Robinson和Lermusiaux认为数据同化是一种对海洋变量进行估计的通用方法,它联合观测数据和海洋学动态模型提供精确的海洋变量估计值,而这是其他方法所不具备的(Robinson and Lermusiaux, 2000)。李新认为陆面数据同化是集成多源地理空间数据的新思路,其核心思想是把不同来源、不同分辨率、直接和间接的观测数据与模型模拟结果集成,生成具有时间一致性、空间一致性和物理一致性的各种地表状态的数据集,并且认为在陆面数据同化系统上发展起来的陆地信息系统将会发展为未来GIS中的重要工具(李新和黄春林, 2004; 黄春林和李新, 2004; 李新和摆玉龙, 2010)。宫鹏定义数据同化是在考虑数据时空分布和对模型、观测做出误差估计的基础上,在数值模型的动态运行过程中融合新的观测数据的方法,其目的是生成具有时间、空间和物理一致性的数据集,他认为遥感数据的引入可以辅助改善环境模型的模拟精度,尤其是在提高模型对参数的空间不均一分布(异质性)的模拟方面,同时数据同化也可以改善遥感数据产品的精度(宫鹏, 2009)。上述定义方法表述有差异,但所描述的数据同化本质是一样的,总结上述定义数据同化包括4个基本要素:模拟自然界真实过程的动力模型;状态量的直接或间接观测数据;不断将新观

测的数据融入过程模型计算中、校正模型参数、提高模型模拟精度的数据同化算法；驱动模型运行的基础参量数据。

数据同化技术不同于一般的遥感数据反演或特征提取技术，是帮助研究地球系统复杂问题的工具(Daley, 1991)。经过资料调研和对主要国家的数据同化系统剖析后，将数据同化系统概括为一个框架和四个基本要素构成(Reichle, 2008; Gao et al., 2009; 马建文和秦思娴, 2012)。一个框架是陆面过程动力模型，数据同化就是在模型的动力框架内完成的；四个基本要素包括过程驱动模型、观测数据、数据同化算法和驱动过程模型运行所必须的基础参量。本书所用到的数据同化基本构成如图 1.1 所示。

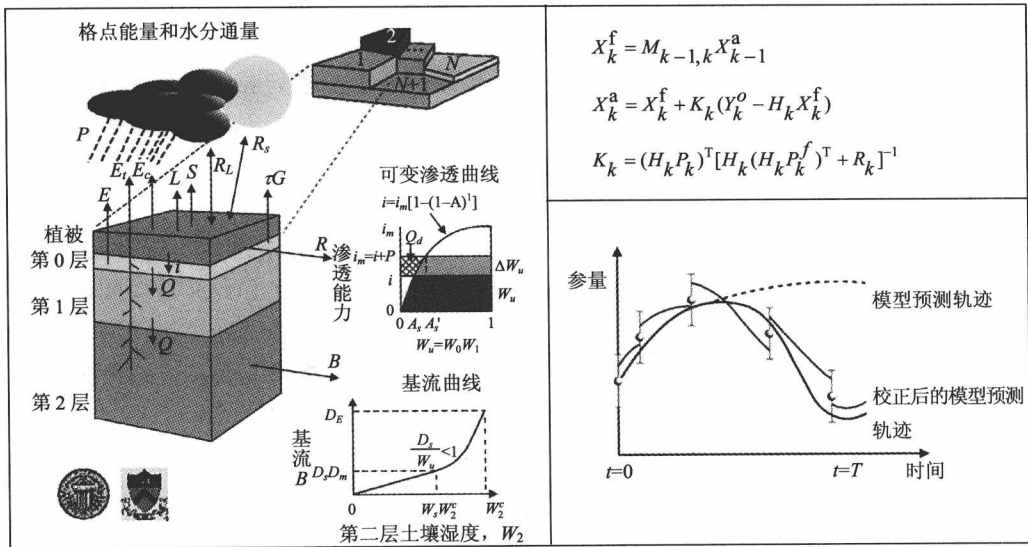


图 1.1 数据同化框架图

所选择的陆面过程模型是 VIC 水文过程模型(Gao et al., 2009)，VIC 水文过程模型原理见左图；

右上图为数据同化算法；右下图为模型与观测数据的作用机制图(Reichle, 2008)

本书所选择的四个基本要素及其作用如下：

(1) 陆面过程驱动模型，陆面过程驱动模型是数据同化核心。书中选择的是 VIC 水文过程模型，VIC 水文过程模型被认为是当前比较成熟的模型之一。

(2) 基础参量，驱动陆面过程模型需要的基础参量，描述研究区域内气候、土壤和植被特性的参数，主要包括反照率、土壤类型、高程等参数。

(3) 数据同化算法，包括以三维变分算法、四维变分算法和集合卡尔曼滤波算法为代表的经典数据同化算法，以及以粒子滤波和层状贝叶斯方法为代表的前沿探索方法。

(4) 观测数据，根据观测数据与待同化状态参量之间的一致性，数据同化通常可以分为直接同化和间接同化两类：一类是利用地面观测数据或者遥感反演产品直接进行同化，即直接同化；另一类是直接利用遥感观测数据进行同化，即间接同化，此时往往需要辐射传输模型或者其他模型表示观测数据与状态数据之间的关系。

将数据同化划分为一个框架和四个独立要素的好处是多方面的。一是将复杂问题简单化条理化，便于通过简明途径认识问题本质；二是便于利用“黑箱”原理开展大量的实验，