

数据同化——一种集成多源地理空间数据的新思路

李 新¹ 黄春林²

(中国科学院寒区旱区环境与工程研究所, 研究员、博士生导师¹; 博士生² 兰州 730000)

〔摘要〕 陆面数据同化是一种集成多源地理空间数据的新思路。介绍了陆面数据同化系统的概念和组成,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为具有时间一致性、空间一致性和物理一致性的各种地表状态的数据集。介绍了国际上目前正在发展的几个陆面数据同化系统,包括 NASA 发展的全球和北美陆面数据同化系统、欧洲陆面数据同化系统和中国西部陆面数据同化系统。认为在陆面数据同化系统基础上发展起来的陆地信息系统将可望成为未来 GIS 中的重要工具。

〔关键词〕 数据同化 地理空间数据 陆面数据同化系统

〔中图分类号〕 K9 P22

〔文献标识码〕 A

〔文章编号〕 1000-7857(2004)12-0013-04

DATA ASSIMILATION: A NEW MEANS FOR MULTI-SOURCE GEOSPATIAL DATA INTEGRATION

LI Xin HUANG Chun-lin

(Cold Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou 730000, China)

Abstract: Land data assimilation is a new means for multi-source geospatial data integration. We first introduced the conception and composition of the land data assimilation system (LDAS). Data assimilation is the method to integrate observations with model simulation results. The observations can be both direct and indirect data, from different sources and with variant resolutions. The output from a LDAS is datasets of land surface variables with spatiotemporal and physical consistence. We then reviewed major LDASs under development, which include the Global and North American Land Data Assimilation Systems developed by NASA (National Aeronautics and Space Administration), European Land Data Assimilation System, and Land Data Assimilation System for West China. We think that the Land Data Information System derived from LDAS will play a very important role in the new generation GIS.

Key Words: data assimilation, geospatial data, land data assimilation system

CLC Numbers: K9 P22

Document Code: A

Article ID: 1000-7857(2004)12-0013-04

1 引言

多源、多分辨率空间数据的融合、集成和尺度推绎是当前 GIS 研究中的热点问题,但缺乏方法论支持。例如,研究者从不同的角度提出了分别针对多种数据模型或针对不同

比例尺/网格分辨率数据的融合方法,但如何在统一的方法论框架内解决以下数据融合和集成的问题尚有待研究。这些问题包括:(1)如何集成来自于不同观测系统(常规观测和遥感观测)的数据;(2)如何集成直接观测(如对地表土壤水分和土壤温度的直接观测)和间接观测(如利用被动微波

收稿日期:2004-09-09。

基金项目:国家重点基础研究发展项目(2001CB309404)、国家自然科学基金项目(90202014)。

作者简介:李新,男,1962~,兰州市中国科学院寒区旱区环境与工程研究所(730000),主要从事四维数据同化、空间分析、环境模拟、遥感和 GIS 在冰冻圈研究中的应用、遥感和 GIS 在水文水资源研究中的应用研究, E-mail:lixin@ns.lzb.ac.cn。

152

[20] John G Lyon, Ding Yuan, Ross S Lunetta, et al. A change detection experiment using vegetation indices [J]. PE&RS, 1998, 64(2): 143~150

[21] Lambin E F. Change-vector analysis in multitemporal space: a tool to detect and categorize land-cover change processes using

high temporal resolution satellite data [J]. Remote sensing of environment, 1994, 48(2): 231~244

[22] Merrill K Riehl, Jiajun Liu. A comparison of four algorithms for change detection in an urban environment [J]. Remote sensing of environment, 1998, 63(2): 95~100

(责任编辑 肖庆山)

观测得到的亮温数据间接反演土壤水分);(3) 如何集成观测数据和模型模拟结果(前者在时间和空间上都不连续,而后者具有时空连续的特征);(4) 如何在融合多源数据的同时,解决数据分辨率不一致的问题,即 Scaling-up 和 scaling-down 的问题。

20 世纪 90 年代末期发展起来的陆面数据同化方法为解决以上问题提供了一条可行的思路。本文将介绍陆面数据同化系统的概念和组成,以及国际上目前正在发展的几个陆面数据同化系统。

2 陆面数据同化系统的概念和组成

2.1 陆面数据同化系统的概念与发展

观测与模拟是地球表层科学研究的两种基本手段,二者缺一不可,但又是一对矛盾体。如果仅就观测而言,无论是常规观测还是遥感观测都无法完整和连续地表达地表时空信息。这主要有以下几方面原因。(1) 观测获得的都是瞬时值,而地表过程无论在时间还是空间上都是连续的,因此迫切需要借助于模型,把观测“转化”为具有时空一致性的数据集。(2) 遥感对地表水文变量的观测都是间接的。一般来说,可以显式地建立地表参数和卫星观测值(如亮温)之间的正向模型。但是,由于参数类型往往多于要反演的变量,而正向模型通常又是十分复杂的非线性模型,这导致反演十分困难或“病态”反演,需要增加先验信息,以提高反演的可能性并增加反演精度。陆面过程模型和分布式水文模型,作为反演中的一种物理约束,可以起到先验信息的作用。(3) 遥感一般无法观测浅层地表以下的信息。例如,对于湿润土壤,微波遥感仅能探测表层几厘米的水分含量,对于根区和深层土壤水分的反演则无能为力。(4) 观测存在误差。如果仅就模拟而言,现有的陆面过程模型通常都包含了复杂的陆面过程参数化方案,但是,它们的模拟精度依然较差,这是因为:很难确定某一特定地区陆面状况的初始值和水文、生物物理参数值;模型不是完美的,存在着模型的物理机制、参数化方案和数值计算方法等方面的缺陷。

要充分发挥观测与模拟各自的优点,就必须对它们进行有机集成。可以设想:如果能够利用陆面过程模型来约束遥感反演模型,利用高分辨率的遥感数据来调整模型的运行轨迹,使积累的误差得到“释放”,最大限度地利用不同来源、不同空间和时间分辨率的观测数据并将它们有机融合,根据数据的不同特性合理地估计其误差分布,那么我们就可以获得更高分辨率,并且具有物理一致性和时空一致性的数据集,就能更好地表达各种尺度上的水分和能量循环。陆面数据同化方法的出现和渐趋实用化,为我们达到这一目标提供了一条可行的途径。

四维数据同化系统(Four-dimensional Data Assimilation System,简称数据同化系统)是指在考虑数据时空分布的基础上,在数值模型的动态运行过程中融合新的观测数据的方法^[1-3]。其理论、概念和方法来源于大气和海洋科学,在数学上主要借助于估计理论、控制论、优化方法和误差估计理论。在过去 20 年间,数据同化系统日趋成熟。目前,美国国家环境预测中心(NCEP)、欧洲中尺度天气预报中心(ECMWF)、中国气象局、日本气象厅(JMA)等部门都采用优化内插或三维变分方法作为其业务运行系统,四维变分方法也逐渐由理论研究走向实际应用。

相对于大气和海洋科学中的数据同化研究而言,陆面和水文数据同化系统的研究则开始较晚。在 IUGG 1991~1994 美国国家报告中,麻省理工学院教授 Dennis McLaughlin 总结到:“数据同化在水文学中迄今还未成为一个独特的领域,因此,在目前鉴别这一领域的研究活动都带有某种主观性和武断性……作者还没有见到一个利用分布式水文模型同化观测数据的研究”^[4]。直到 20 世纪 90 年代后期,陆面数据同化系统的研究才日益活跃起来,一些颇受关注的文章陆续被发表^[4-20];使用“离线”的陆面过程模型同化微波遥感观测数据也逐渐成为陆面数据同化系统的明显特征。最具代表性的事件是:美国国家航空与宇航管理局哥达德空间飞行中心(NASA GSFC)水文科学部和数据同化办公室(DAO)联合支持了陆面数据同化项目,目前已开始准实时地提供北美 1/8° 和全球 1/4° 陆面同化数据集;美国国家大气海洋局(NOAA)也支持了美国大陆尺度的、基于水文学的陆面数据同化系统研究^[21]。此外,ECMWF 也在欧盟的支持下于 2002 年启动了欧洲陆面数据同化系统^[22]、JMA 和日本宇宙开发局(NASDA)也在筹建区域性的陆面数据同化中心。

2.2 陆面数据同化系统的组成

陆面数据同化系统通常由陆面过程模型、水文模型、观测算子(一般为辐射传输模型或客观分析方法)、误差估计模型和优化算法组成。它以极小化以下泛函为目标^[5-7]:

$$J(x_0, \alpha) = \frac{1}{2} \sum_{i=1}^n (H_i \{ [M(x_0, \alpha)]_i \} - y_i^o)^T R_i^{-1} (H_i \{ [M(x_0, \alpha)]_i \} - y_i^o) + \frac{1}{2} (x_0 - x_0^b)^T B^{-1} (x_0 - x_0^b)$$

式中, J 是目标函数;下标 i 是离散后的数值模型积分步长; x_0 是状态矢量的初始条件; α 是数值模型的待定参量,它们是被同化或被反演变量(本项研究中是土壤水分、温度、积雪和冻土状态等)组成的列矩阵;下标 0 表示同化周期开始时的状态; x_0^b 是背景场; y_i^o 是 i 时刻的观测,它可以是与 x 不同物理意义、不同维数的矢量,如多波段、多极化的亮度温度; M 被称为模型算子,在本研究中是一个陆面过程模型; H_i 被称为观测算子,在本研究中是微波辐射传输模型; R_i 是观测误差矩阵; B 是背景场的误差矩阵。

(1) 陆面过程模型和分布式水文模型。它是模拟地表(大气-植被-土壤界面)能量和水分平衡以及生物地球化学循环的模型。目前较为成熟的陆面过程模型包括 SiB2 (Simple Biosphere Model) 模型^[23-25]、BATS (Biosphere-Atmosphere Transfer Scheme) 模型^[26-28] 和当前最为广泛认可的新一代陆面过程模型 CLM (Common Land Model) 模型^[29-31]。

(2) 各种自然介质的微波辐射传输模型。微波辐射传输模型在同化系统中被作为观测算子,陆面过程模型或水文模型输出的部分状态变量(LAI、生物量、土壤湿度、地表温度等)通过辐射传输模型转化为卫星传感器的观测值(亮度温度、后向散射强度等)^[14-16]。

(3) 陆面数据同化方法。一般采用变分法^[1-2]、卡尔曼滤波^[7,32] 和启发式优化算法(如模拟退火法和遗传算法)^[17-20,33]。集合卡尔曼滤波是近年来在传统卡尔曼滤波的基础上发展起来的一种新方法,它采用集合预报的方式,用观测矩阵的概率分布初始化模型,并同步预报,以得到预报结果的概率分布,误差分析依赖于预报结果的集合(符合某种概率分布

的 N 个预测场)而得出^[33]。集合卡尔曼滤波很好地解决了在同化系统中如何处理模型算子和观测算子的非线性和不连续性的问题,又可以依据集合预报结果得出误差估计,从而避免了超大矩阵的数值运算,因此是目前重点发展的同化方法。

(4)误差估计模型。包括观测误差估计模型和背景场误差估计模型。观测误差由仪器误差和代表性误差组成,一般采用高斯噪声模型模拟仪器误差,用空间自相关模型模拟代表性误差。对陆面过程模型和微波辐射传输模型误差传递的有限知识,也是影响进一步提高四维数据同化系统性能的重要因素,但这类误差很难估计。

3 陆面数据同化系统介绍

3.1 北美陆面数据同化系统和全球陆面数据同化系统

北美陆面数据同化系统(NLDAS)和全球陆面数据同化系统(GLDAS)项目于1998年启动,主要以北美洲大陆(包括美国全部、墨西哥和加拿大部分地区)和全球两个不同尺度的陆面状态和通量的建模为目标,强调利用陆面数据同化系统提供全球和区域性的陆面同化数据集。它由NASA哥达德空间飞行中心(GSFC)、美国国家海洋大气局的国家环境预报中心(NOAA/NCEP)和水文科学部、普林斯顿大学、华盛顿大学联合开发。其中,NLDAS空间分辨率为 $1/8^\circ \times 1/8^\circ$ 、时间分辨率为1小时;GLDAS的空间分辨率为 $1/4^\circ \times 1/4^\circ$ 、时间分辨率为3小时^[21]。

北美(全球)陆面数据同化系统包括以下几方面。(1)驱动数据。包括NCEP的全球数据同化系统和NASA的哥达德的数据同化系统生成的气象驱动数据等。(2)陆面模型。包括Mosaic、VIC、NOAH、Sacramento 4种陆面过程模型。(3)陆面参数集。包括北美和全球的植被分类图、叶面积指数、土壤数据集、高程数据集等。(4)观测数据。静止气象卫星红外观测、SSM/I、TRMM等遥感数据以及地面站点观测数据。(5)数据同化算法。四维变分、Kalman滤波、集合Kalman滤波算法。(6)输出数据集。通过同化地表观测和遥感数据,最终将输出土壤水分、蒸散发、能量通量、径流、积雪等数据集^[21]。

3.2 欧洲陆面数据同化系统

欧洲陆面数据同化系统(European Land Data Assimilation System to Predict floods and droughts,ELDAS)是欧盟资助项目,于2001年12月1日正式启动。该项目的目标是设计和实现数值天气预报环境下的土壤水分数据同化系统,评价ELDAS对于水文预报(洪水、季节性干旱)的改进效果。ELDAS最终输出每日、空间分辨率为 $1/5^\circ \times 1/5^\circ$ 的欧洲大陆的土壤水分数据。目前该系统正处于完善中,2005年将全面完成^[22]。

3.3 中国西部陆面数据同化系统

中国西部陆面数据同化系统(West China Land Data Assimilation System,WCLDAS)的研究是国家自然科学基金委员会资助的“中国西部环境和生态科学重大研究计划”中的项目,于2003年1月正式启动,它由中国科学院寒区旱区环境与工程研究所和兰州大学资源环境学院大气科学系合作研究和开发。通过建立中国西部地区陆面数据同化系统,使用分布式水文模型VIC和CLM模型同化被动微波观测资料SSM/I、TMI和AMSR,获得1991年以来高分辨率的

中国西北干旱区和青藏高原土壤水分、土壤温度、积雪和冻土的同化资料;为研究西部地区大尺度和流域尺度的水循环提供具有时空和物理一致性的数据。该项目正在实施中,预计在2005年12月将建成实用的中国西部陆面数据同化系统^[35]。

中国西部陆面数据同化系统包括以下几方面。(1)驱动数据。发展一个三维变分分析系统,以GAME、ECMWF、和NCEP再分析资料为基础,融入台站观测资料和卫星反演资料,最终生成中国西部地区 $0.25^\circ \sim 0.5^\circ$ 分辨率的辐射、气温、降水、水气压、风速等驱动数据。(2)陆面模型。分布式水文模型(VIC模型)、改进的CLM模型。(3)陆面参数集。针对VIC、CLM模型所需参数制备。(4)观测数据。SSM/I、TMI和AMSR被动微波观测资料。(5)数据同化算法。四维变分、集合Kalman滤波算法、模拟退火算法^[17-20]。(6)输出数据集。通过同化地表观测和遥感数据,最终将输出空间分辨为 $1/4^\circ \times 1/4^\circ$ 、时间分辨率为6小时的中国西北干旱区和青藏高原土壤水分、土壤温度、积雪和冻土的同化(再分析)资料(见图1)。

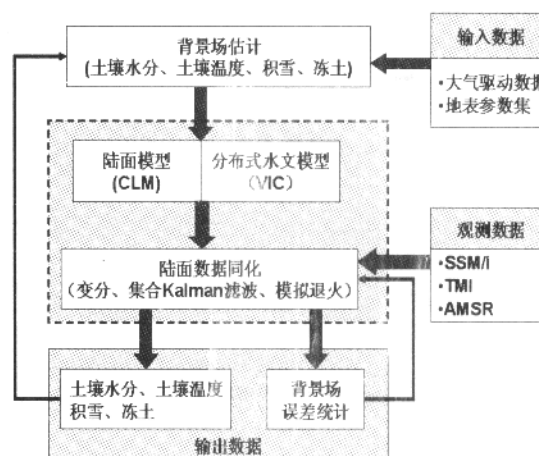


图1 中国西部陆面数据同化系统框架

4 结语

本文介绍了一种集成多源空间数据的新方法——陆面数据同化。陆面数据同化系统可以把不同来源、不同分辨率、直接和间接的观测数据与模型模拟结果集成为具有时间一致性、空间一致性和物理一致性的地表各种状态的数据集。它可望成为未来GIS中模拟地表动力学过程和管理时空动态数据的重要工具,而这些特征都是GIS当前所欠缺的。2002年,NASA DAO(Data Assimilation Office)和水文科研部门提出了发展陆地信息系统的设想,他们期望显著提高陆面数据同化系统的空间分辨率(1km),并且发展集成了数据挖掘、数值模拟和可视化工具的web用户界面以访问陆面数据同化系统^[36]。有理由相信,陆地信息系统将可望成为未来GIS中的重要工具,甚至未来GIS的雏形。

参考文献(References)

- [1] Daley R. Atmospheric Data Analysis [M]. New York: Cambridge University Press, 1991
- [2] Talagrand O. Assimilation of observations, an introduction [J]. Journal of Meteorological Society of Japan, 1997, 75(1):191~209

- [3] Morel P. An overview of meteorological data assimilation. In: Bengtsson L, Ghil M, Kallen E. (Eds), *Dynamic Meteorology: Data Assimilation Methods*[M]. New York: Springer-Verlag, 1981. 5~16
- [4] McLaughlin D. Recent development in hydrologic data assimilation[J]. *Reviews of Geophysics*, 1995:977~984
- [5] Ide K, Courtier P, Ghil M, Lorenc A C. Unified notation for data assimilation: operational, sequential and variational [J]. *Journal of the Meteorological Society of Japan*, 1997: 181~89
- [6] Zupanski M, Kalnay E. Principles of data assimilation. In: Browning K A, Gurney R J. (Ed.), *Global Energy and Water Cycles*[M]. New York: Combridge University Press, 1999
- [7] Kumar P. Assimilation of near-surface temperature using extended Kalman filter [J]. *Advances in Water Resources*, 2003,26:7~93
- [8] Houser P R, Shuttleworth W J, Famiglietti J S, Gupta H V, Syed K H, Goodrich D C. Integration of soil moisture remote sensing and hydrologic modeling using data assimilation [J]. *Water Resources Research*,1998,34(12): 3 405~3 420
- [9] Walker J P, Willgoose G R, Kalma J D. One-dimensional soil moisture profile retrieval by assimilation of near-surface observations: a comparison of retrieval algorithm [J]. *Advances in Water Resources*, 2001, 24: 631~650
- [10] Shuttleworth W J. Combining remotely sensed data using aggregation algorithms [J]. *Hydrology and Earth System Sciences*, 1998(2): 149~158
- [11] Schuurmans J M, Troch P A. Assimilation of remotely sensed latent heat flux in distributed hydrological model[J]. *Advances in Water Resources*, 2003, 26:151~159
- [12] Bach H, Mauser W. Methods and Examples for Remote Sensing Data Assimilation in Land Surface Process Modeling[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2003,41(7): 1 629~1 637
- [13] Lakshmi V. A simple surface temperature assimilation scheme for use in land surface models [J]. *Water Resource Research*, 2000, 36(12): 3 687~3 700
- [14] Entekhabi D, Galantowicz J F, Njoku E G. Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 1994, 32, (2): 438~448
- [15] Galantowicz J F, Entekhabi D, Njoku E G. Test of sequential data assimilation for retrieving profile soil moisture and temperature from observed L-band radiobrightness[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 1999, 37(4): 1 860~1 870
- [16] Hoeben R, Troch P A. Assimilation of active microwave observation data for soil moisture profile estimation [J]. *Water Resources Research*, 2000, 36(10): 2 805~2 819
- [17] 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
- [18] 李新, 小池俊雄, 程国栋. 一个基于模拟退火法的陆面数据同化算法[J].*地球科学进展*, 2004,18(4):632~636
- [19] Pathmathevan M, Koike T, Li X. A new satellite-based data assimilation algorithm to determine spatial and temporal variations of soil moisture and temperature profiles [J]. *Journal of the Meteorological Society of Japan*,2003, 81 (5): 1 111~1 135
- [20] Pathmathevan M, Koike T, Li X, Fujii H. A simplified land data assimilation scheme and its application to soil moisture experiments in 2002 (SMEX02) [J]. *Water Resources Research*, 2003, 39 (12): 1 341
- [21] <http://ldas.gsfc.nasa.gov/>
- [22] Development of a European Land Data Assimilation System to predict floods and droughts. Description of Work [C]. <http://www.knmi.nl/samenw/eldas/>
- [23] Sellers P J, Mintz Y, Sud Y C and Dalcher A. A simple biosphere model (SiB) for use within general circulation models [J]. *Atmos. Sci.*,1986, 43: 305~331
- [24] Sellers P J, Randall D A, Collatz G J, Berry J A, Field C B, Dazlich D A, Zhang C, Collelo G D, Bounoua L. A revised land surface parameterization (SiB2) for atmospheric GCMs, Part I: model formulation[J]. *Journal of Climate*, 1996, 9: 676~705
- [25] Sellers P J, Los S O, Tucker C J, Justice C O, Dazlich D A, Collatz G J, Randall D A. A revised land surface parameterization (SiB2) for atmospheric GCMs, Part II: The generation of global fields of terrestrial biophysical parameters from satellite data[J]. *Journal of Climate* 1996,9: 706~737
- [26] Dickinson R E, Henderson-Sellers A, Kennedy P J, Wilson M F. Biosphere Atmosphere Transfer Scheme (BATS) for the NCAR Community Climate Model. NCAR Technical Note, NCAR, TN275+STR, 1986, 69
- [27] Dickinson R E, Henderson-Sellers A, and Kennedy P J. Biosphere-Atmosphere Transfer Scheme (BATS) Version 1e as coupled to the NCAR Community Climate Model. NCAR Technical Note NCAR/TN-387+STR, 1993, 72
- [28] Wilson M F, Henderson-Sellers A, Dickinson R E, Kennedy P J. Sensitivity of the biosphere-atmosphere transfer scheme (BATS) to the inclusion of variable soil characteristics[J]. *J. Clim. Appl. Meteor.*, 1987, 26: 341~362
- [29] Dai Y, Zeng QC. A land surface model (IAP94) for climate studies. Part I: Formulation and validation in off-line experiments [J]. *Advances in Atmospheric Science*, 1996, 4:433~460
- [30] Dai Y, Zeng X, Dickinson RE, Baker I, Bonan GB, Bosilovich MG, Denning AS, Dirmeyer PA, Houser PR, Niu G, Oleson KW, Schlosser CA, Yang Z-L. The Common Land Model [J]. *Bulletin of the American Meteorological Society*, 84:1 013~1 023
- [31] Zeng X, Shaikh M, Dai Y, Dickinson RE, Myneni R. Coupling of the Common Land Model to the NCAR Community Climate Model[J]. *Journal of Climate*, 2002, 4: 1 832~1 854
- [32] Kalman R E. A new approach to linear filtering and prediction problems [J]. *Trans. ASME, Series D, J. Basic Eng.* 1960,V 82: 35~45
- [33] Kirkpatrick S, Gelatt Jr C D, Vecchi M P. Optimization by Simulated Annealing[J]. *Science*, 1983, 220(4 598): 671~680
- [34] Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics[J]. *Journal of Geophysical Research*, 1994, 99(C5): 10 143~10 162
- [35] <http://ldas.westgis.ac.cn/>
- [36] Houser P R, Peters-Lidard C, Dirmeyer P, et al. Land Information System [M]. NASA's Goddard Space Flight Center (GSFC), 2002

(责任编辑 肖庆山)