

文章编号: 1001-8166 (2010) 03-0306-11

数字黑河的思考与实践 2: 数据集成

李 新, 吴立宗, 马明国, 盖迎春, 冉有华, 王亮绪, 南卓铜
(中国科学院寒区旱区环境与工程研究所, 甘肃 兰州 730000)

摘 要: 数据集成是流域集成研究的核心环节, 是发展、改进和验证模型的基础。介绍了“数字黑河”数据集成研究的进展: 定义了数据集成; 提出数据库集成应具有数据的完整性、可获取性和有效管理 3 个基本特征; 而模型数据集可根据流域集成模型对科学数据的需求, 概化为驱动数据、参数集、验证和诊断数据三大类。进一步丰富和完善了“数字黑河”信息系统; 扩展了黑河数据库, 较为系统地收集了黑河流域的基础地理数据、各类观测数据、科学试验数据和现有的模型数据集, 在线数据量已扩充至 1 000 GB, 并且广泛共享。介绍了模型数据集取得的进展及目前存在的问题; 发展了制备高分辨率驱动数据的统计和动力降尺度方法; 依靠数据融合方法制备了适用于模型的黑河流域土地覆盖图和土壤图。指出了数据集成所面临的挑战和需要优先完成的迫切任务, 核心是以发展流域集成模型为导向, 进一步完善“数字黑河”信息系统, 并且制备高分辨率和可靠的模型数据集。

关 键 词: 数字流域; 数据集成; 模型数据集; 黑河流域

中图分类号: P334 **文献标志码:** A

1 引 言

数据是流域信息基础设施中最主要的组成部分, 是分析的基础、模型的燃料、决策的依据, 流域科学的进展依赖于对于各种观测数据的获取和集成^[1,2]。遥感等新兴观测技术的快速发展正在把水文学由传统上“数据稀缺的科学”推进到数据丰富的科学^[3]。但面对地球观测时代蜂拥而至的各类数据, 如何从信息海洋中提取出有助于更好地理解流域水文与生态等地表过程——并支持综合模型的发展的准确和一致的数据, 是数据集成所要回答的核心问题。因此, 数据集成是流域集成研究的核心环节之一, 也是模型集成的前提。

本文将在“数字黑河”总体设计一文^[4]的基础上, 介绍“数字黑河”中数据集成方面的研究进展,

在对数据集成——包括数据库集成和数据科学内容集成的内涵进行剖析之后, 回顾了“数字黑河”数据库建设及网络信息系统的发展和相关技术集成工作, 介绍了针对发展模型数据集的数据科学内容集成。

2 数据集成的整体设计

2.1 数据集成的定义

“数字黑河”研究中, 我们将数据集成定义为:

第一个层次上, 是数据库集成, 即利用先进的数据库技术, 对各类空间数据统一建库, 实现对异构数据的统一访问; 每种数据都有完备的元数据和数据文档, 有基本的质量控制; 同时, 建立分布式的信息系统, 并通过多种 Web 服务方式实现数据共享。

第二个层次上, 利用多种来源、多分辨率的数

收稿日期: 2009-12-14; 修回日期: 2010-01-23

* 基金项目: 中国科学院西部行动计划 (二期) 项目“黑河流域遥感—地面观测同步试验与综合模拟平台建设”(编号: KZCX2-XB2-09); 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); 中国科学院知识创新工程重要方向项目“地表过程建模环境和模型集成研究”(编号: KZCX2-YW-Q10-1) 资助。

作者简介: 李新 (1969-), 男, 甘肃酒泉人, 研究员, 主要从事陆面数据同化、遥感和 GIS 在冰冻圈和水文水资源研究中的应用、流域集成研究。E-mail: lixin@lzh.ac.cn

据,在质量控制的基础上,整合成为能够广泛地支持各种分析,并服务于模型发展、验证和改进的数据集/数据库。

第三个层次上,融合来自于地面观测、遥感观测、模型输出的多种数据,产生创新性的新的数据产品,如可用于多种模型的标准的参数集,又如可被广泛地应用于各种分析的同化数据产品和长序列遥感数据产品。产生新的数据产品的方法包括信息融合、数据同化、国际数据集的区域化等。

本文主要讨论数据库集成和模型数据集,后者涉及到本文所定义的数据集成的第二个层次的全部和第三个层次的部分。数据同化和国际数据集的区域化不是本文的讨论内容。

2.2 数据库集成

图 1 概述了“数字黑河 数据库集成的框架。

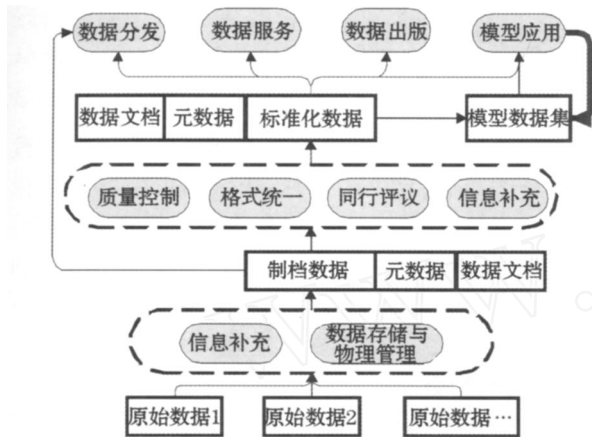


图 1 数字黑河中的数据库集成

Fig 1 The scheme of the database integration in the Digital Heihe River Basin

数据库集成应具有以下 3 个方面的特征,分别是数据的完整性、可获取性和有效管理^[5]。

(1) 数据的完整性,即数据及其描述信息的完整、完全和真实。“数字黑河”信息系统在数据完整性方面应具有以下特征:

- 数据的绝对真实性。
- 基本的质量控制。
- 完备的元数据定义,辅之以必要的文档。
- 逐步实现标准的、自描述的数据格式。
- 重要数据集逐步采用唯一数字编码(DOI Digital Object Identification)。
- 逐步开展对数据的同行评议。
- 逐步实现数据出版。

(2) 数据的可获取性,指实现数据共享的政策和技术手段。主要特征为:

- 完全与开放(Full & Open)的数据共享政策。
- 统一的物理组织/存储,多样的逻辑组织形式。
- 空间数据的管理不局限于特定的GIS软件的框架。
- 异质性数据的互操作。
- 基于服务器/浏览器结构,使用Web用户界面和WebGIS访问数据,实现对数据随时随地的便捷获取。
- 提供数据工具。
- 提供基本的数据可视化。

(3) 数据的有效管理,是指长期保护数据,并使得数据发挥持续价值的管理工作^[5]。包括以下措施:

- 建立数据中心。
- 专职和专业的数据管理人员。
- 保证数据长期存储、数字介质定期更新的管理工作。

2.3 模型数据集

信息系统中的数据,往往并不能直接用于各类模型,而是需要综合多种来源、多种分辨率的数据,才能制备出用于各类流域模型的发展、验证和改进的模型数据集。可根据流域集成模型、水文模型、生态模型、社会经济模型和其他模型对科学数据的需求,将模型数据集概化为驱动数据、参数集、验证和诊断数据三大类。模型数据集是目前流域集成研究中最缺乏的数据集,对流域模型建模等研究工作构成了巨大障碍。针对这一问题,我们提出了如图 2 所示的总体研究思路,目标是为流域综合模型制备适用、完备、高质量和高分辨率的模型数据集。

模型数据集中,第一大类数据是验证和诊断数据(各类观测数据)。它们是发展、验证和改进综合模型的基础,也是模型初始化和数据同化的原料。基础性的观测和综合试验必须进一步加强,以获取更为完备的观测数据。这一方面的讨论可参考作者对于流域观测系统的构想^[7]及本系列论文的第 4 篇。在这里,我们不再就此类数据展开讨论。

第二大类数据是驱动数据。在“水—土—气—生—人”各要素中,水文、土壤和生态模型都需要近地表大气状况,包括气温、降水、气压、水气压、风场、短波辐射和长波辐射作为边界条件,我们通常称这一类数据为驱动数据。格网化的高分辨率(包括时

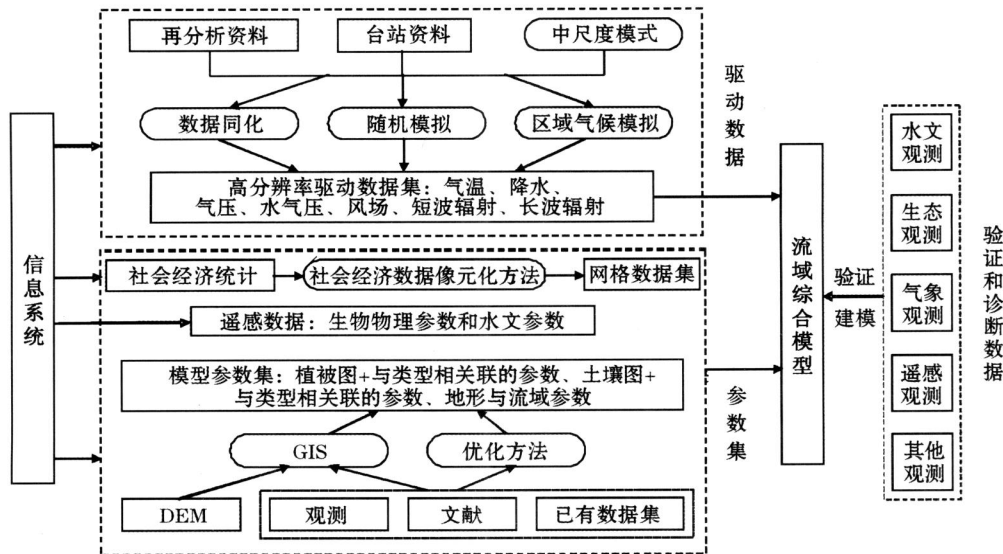


图 2 数字黑河中模型数据集制备的总体框架 (修改自李新^[6])

Fig 2 General scheme of the model datasets preparation in the Digital Hehe River Basin

间和空间分辨率)驱动数据是运行各类模型的基础,但它们通常难以获得。这是因为,台站资料在空间上分布稀疏,中国西北干旱区更是如此;在时间上,常规台站的观测频率一般为一天4次。因此,依据台站资料而内插的格网化数据集通常无法满足模型对驱动数据的质量和时空分辨率要求。另一类可广泛获取的驱动数据是各气候中心,如美国国家大气研究中心(National Center for Atmospheric Research, NCAR)和欧洲中期天气预报中心(European Center for Medium-Range Weather Forecasts, ECMWF)提供的再分析资料。再分析资料的优点是,它们是大气模型同化了观测资料的结果,因此具有严格的物理基础,在大尺度上偏差较小,且时间分辨率较高;但缺点是空间分辨率较粗,全球再分析资料分辨率最好也就是 1° ;一些区域再分析资料空间分辨率可达到 0.25° ;但对于流域尺度的应用,分辨率依然较粗。因此,如何得到流域尺度上的高分辨率驱动数据,是一个巨大的挑战。可行的途径是融合再分析资料、台站资料和中尺度模式,依靠统计降尺度或动力降尺度(数据同化)方法,在再分析资料这一大的背景场基础上,叠加台站观测资料和区域气候模拟的高频信息,获得高分辨率的驱动数据。近期的目标是在流域尺度上得到空间分辨率为 $1\sim 5\text{ km}$,时间分辨率为1小时的驱动数据集。

第三大类数据是参数。几乎所有的水文模型、生态模型和陆面过程模型,都需要大量的土壤、植被

及其他参数^[8]。参数又可分为静态参数、动态参数和社会经济参数(图2)。静态参数指不随时间变化或变化很小的参数,水文、生态和陆面过程模型中,一般假设植被的部分生理参数和生物物理参数是和植被类型一一对应,而土壤的热力和水力参数与土壤质地类型相关。动态参数指随时间而变化的参数,如叶面积指数、光合作用有效能量比例等,多可从遥感数据获得,如从甚高分辨率辐射计(Advanced Very High Resolution Radiometer, AVHRR)、法国地球观测卫星上的一个中等分辨率传感器(SPOT VEGEATION)和中分辨率成像光谱仪(Moderate Resolution Imaging Spectroradiometer, MODIS)提取的NDVI序列和各种生物物理参数产品等。社会经济模型所需参数传统上一般来源于统计资料,但为了与生态和水文模型耦合,需要发展以网格为计算单元的空间显式的社会经济模型,因此需要实现社会经济数据的像元化^[9]。

在参数集制备方面,目前最急需的是高质量的土地覆盖(或植被图、植被功能型图)和土壤质地图及与它们的类型相关联的参数。在现有的陆面过程和大尺度水文模拟中,大多数参数集来自于国际数据集。但这些数据一方面分辨率相对较粗,另一方面质量欠佳,使用的中国区域的信息较少,经验成分很大,与真实的植被、土壤分布及其相应性质的真实情况可能有较大差别^[10]。如目前可用的陆面过程模型中的土壤图,多从全球数据集中切割,其真实信

息源于联合国粮农组织 1 500 万全球土壤图,使用的中国境内的土壤剖面数据甚少,因此,其可靠性十分值得怀疑。目前,研究人员已经认识到,必须充分利用已有的各种数据,制备可靠的高分辨率参数集,以满足陆面过程研究的需求。当前存在的主要问题是,我国所采用的土地覆盖 植被类型和土壤类型分类与国际分类标准不一致,特别是土壤分类,我国采用发生学分类,而陆面和水文模型中通常使用质地分类,如何实现二者之间的转换,是一个复杂的问题^[11]。类似地,土地覆盖 植被数据也需要参比 IG-BP 等标准,实现类型转换。可行的途径是综合已有的数据集、散布在文献中的观测资料、对参数的定点和遥感观测,制备更高分辨率和更可靠的模型参数集。为此,需要开展以下研究工作: 数据类型转换; 数据空间分布的校正,我国现有的大量数据是通过传统手段调查获得的(即通过野外调查、手工填图、地图转绘和手工量算等工序),受技术手段和原始资料的限制,这些数据往往存在较大空间位置误差,同一种地物在不同的资源类型图中的位置不同,从而导致在陆面过程参数化过程中同一种类型出现不同的性质,因此必须对这些数据进行空间校正; 数据集成与信息融合,资源调查时各有侧重,如土壤图中有关湖泊、湿地和冰川的描述就不如相应的专题图全面,有鉴于此,应对各种专题数据进行信息融合,以获得更加准确的地表数据集; 数据内容更新,随着时间的迁移,数年前甚至几十年前获取的数据已经不能代表目前的真实情况,因此必须对这些数据进行更新。数据的区域化研究,充分利用已有的国际数据集,特别是其成熟的分类和参数化方案,在其框架内补充高分辨率的区域信息。

3 研究进展

3.1 数字黑河信息系统

“数字黑河”信息系统始于 1996—1998 年间开展的国家重点科技攻关项目“黑河流域水资源合理利用与社会经济和生态环境协调发展研究”^[12],该项目在研究后期,对收集的各类原始资料进行了整理,同时整合了项目产生的大量科学数据,建成了比较完善的黑河流域水资源信息系统^[13]。作为“数字地球”在流域尺度上的一次实践性尝试,中国科学院寒区旱区环境与工程研究所(以下简称寒旱所)于 2000 年提出了建立“数字黑河”的构想,“数字黑河”网络信息系统始建于 2001 年,至今已发布了 3 个正式版本。

第一版本中,发布了黑河流域基础地理信息数据及部分观测数据,比较并实现了多种 WebGIS 功能^[14],初步开发了元数据系统^[15],采用数据与信息存取连接(Data and Information Access Link, DIAL)工具以地球观测系统—层次数据格式(Earth Observing System-Hierarchical Data Format, EOS-HDF)格式发布了网格数据^[16],建立了黑河万象,开展了在网络上三维虚拟显示黑河流域全景的尝试^[17]。

第二版本中,系统整理并发布了黑河流域气象和水文观测数据以及黑河地—气相互作用野外观测实验(HEIFE)数据,系统收集了遥感数据,开发了较为完善的 WebGIS 数据查询工具,开发了 EOS 数据自动下载和重发布系统^[18],新增了流域研究和知识积累等新板块。

第三版本与前两个版本相比进行了较大的改版,改版后的信息系统采用了先进的网站内容管理系统——DoNeNuke,在数据下载上做了方便使用的简化,并通过链接的方式融合到 DoNeNuke 中,发布了更多黑河流域的科学实验数据,对水文和地下水数据做了详尽的补充,大力扩充了遥感数据。依托“中国西部环境与生态科学数据中心”^[19],开发了更加先进的空间数据共享平台^[20],完善了数据服务,实现了数字黑河信息系统向数据中心的过渡^[19]。

图 3 是“数字黑河”信息系统第三版的首页。

3.2 数字黑河数据库

“数字黑河”已建成一个数据内容翔实的数据库,数据量超过 1 000 GB,主要包括以下内容。

(1) 基础地理数据:包括流域边界、研究区边界、1 5 万(局部)、1 10 万和 1 25 万比例尺的数字等高线、30 ~ 1 000 m 不等分辨率的数字高程模型(DEM)和数字地形模型(DTM)、各级行政边界、河流、湖泊、道路、居民点。

(2) 专题数据:包括 1980 年代末至 2007 年的多期 1 10 万土地利用 覆盖数据、1 10 万冰川图、1 10 万沙漠图、1 100 万的植被、土壤、地质、地貌、林场、草场等专题图。

(3) 水文水资源空间数据:黑河流域东部水系 1 5 万灌溉渠系分布^[21,22]、黑河流域东部水系机井分布、黑河流域西部水系部分灌区图、黑河中游部分河道剖面实测数据、全流域水库分布、流域不同区域的水文地质和地下水埋深图、流域潜水水化学图。

(4) 常规气象水文观测数据:黑河流域及其周边地区 22 个常规气象站从建站至今的逐日数据、流



图 3 数字黑河信息系统首页

Fig 3 Homepage of the Digital Heihe River Basin information system

域内 30 个水文站 1978—2004 年逐日水文常规观测资料(降水量、水面蒸发、水位、流量、含沙量以和洪水过程)、张掖盆地(张掖、临泽、高台)逐日或 5 日地下水水位(或埋深)观测数据,黑河中游水并水位动态监测资料、张掖和酒泉盆地泉水历年观测资料、黑河流域调水记录。

(5) 寒旱所黑河流域观测系统数据:寒旱所设置在黑河流域的野外站和观测系统所获取的数据。这部分数据将在“数字黑河”系列论文的第 4 篇中给予详细介绍。

(6) 科学试验数据:包括 HEIFE 数据^[23,24]、祁连山冰沟小流域 1980 年代开展的积雪观测(包括雪深、雪密度、积雪光谱、融雪径流)、黑河流域内数次小规模遥感试验观测数据、1986—1995 年在黑河干流中游平原区开展的包气带水分运移观测试验资料。

(7) 社会经济数据:“黑河流域水资源合理利用与社会经济和生态环境协调发展研究”项目对黑河流域三级地区各种统计数据进行了收集,共生成 108 张各类统计表格,“数字黑河”数据库对该数据进行了收集和整理;此外,“数字黑河”中还收录了部分社会经济和人口调查数据。

(8) 遥感数据:全流域 1980 年代、1990 年代和 2000 年共 3 期的 MSS/TM/ETM+ 影像资料,2001—2003 年的 ASTER 数据、2002 年逐日 NOAA AVHRR 影像数据、全球 AVHRR, SPOT Vegetation, MODIS, SSM/I 等数据及其归一化植被指数(Normalized Difference Vegetation Index, NDVI)等产品的子集、局部地区的极高分辨率遥感影像。

(9) 模型数据集:包括 2000 年度 1 km 分辨率逐小时的统计降尺度高分辨率大气驱动数据,黑河流域人口数据的空间化结果^[25],以及针对 SWAT、PRMS、VIC 等模型发展的驱动和参数数据集。

“黑河综合遥感联合试验”^[8]期间获取的观测数据,将专门制档,并将通过黑河综合遥感联合试验信息系统发布,并最终集成到“数字黑河”中。

“数字黑河”实现了科学数据的完全共享,其中,大部分数据可以通过网站浏览并且完全下载,不作任何权限限制;高精度数据、行业部门数据和国家法律法规规定不能在线发布的数据,也对国内用户提供离线服务。用户可以在网上查询基本信息,在与黑河数据管理委员会签订数据协议后,完全免费使用这些数据。

3.3 模型数据集

3.3.1 驱动数据

高分辨率驱动数据的制备一般依靠插值方法、统计降尺度和动力降尺度方法。其中,降尺度方法由于兼顾了大尺度大气模型(再分析资料)和区域上的气象观测信息,因此一般能够获得更可靠、分辨率更高的驱动数据。发展制备高分辨率驱动数据的方法是“数字黑河”数据集成的重要组成部分。

其中,高艳红等发展了由美国国家环境预测中心(National Centers for Environmental Prediction, NCEP)再分析资料内插的统计降尺度方法,用于黑河流域的复杂下垫面条件下高分辨率驱动数据的制备。该方法以 NCEP 再分析资料作为背景场,通过客观分析(最优插值方法)对模型(再分析资料)和观测场的偏差进行分析,从而修正背景场,并实现降尺度。利用这一统计降尺度方法计算得到了黑河流域 2002 全年 1 km 分辨率的逐小时近地表气象数据,包括,2 m 高度气温、相对湿度、10 m 高风场 U 分量和 V 分量以及降水。结果表明,与直接插值相比较,应用观测资料的客观分析方法可以更好地反映近地层大气要素场的变化,特别是对于山区复杂下垫面,气温、湿度场都有明显的改进,精细的网格值也能描述山区风场的特征^[26]。

动力降尺度方法中由于引入了局地气候模型的信息,更能捕捉到小尺度——特别是山区复杂地形条件下的近地表气象特征。我们也初步发展了动力降尺度方法,采用中尺度气象预报模式天气研究与预测模型(Weather Research & Forecasting Model, WRF),对一天 4 次、空间分辨率为 1 的美国大气与海洋局的全球预测系统(Global Forecast System, GFS)数据进行动力降尺度,制备了黑河流域空间分辨率为 5 km,时间分辨率为 1 小时的大气驱动数据。分别采用黑河流域的常规和加强观测资料进行逐天和逐时比较,以验证降尺度资料的精度。结果表明,气温、压强、比湿和风场与观测资料吻合得很好;辐射总体上与观测资料相关性较高,但模拟结果在时间上过于平滑;降雨则基本上不能抓住研究区域的降雨特征,其制备寄望于依靠雷达观测和遥感资料,以及发展动力和统计相结合的降尺度方法。

辐射在山区受到地形的显著影响,必须考虑地形(高程、坡度、坡向、遮蔽)对辐射的影响。陈仁升等建立了一个任意地形和实际天气条件下计算大范围、长时间、高时空分辨率的太阳入射短波辐射的模型。该模型采用简化的辐射传输参数化方案和

NCEP/NCAR再分析资料相结合的方法,并被成功地应用于黑河流域,计算得到了 2002 年 1 km 分辨率、逐小时的总辐射、直接辐射和散射辐射资料^[27]。

3.3.2 参数集

我国现有的土地覆盖分类与国际上通行的 IGBP 等标准都不尽一致,需要进行类型的参比和重分类,才能得到适用的土地覆盖或植被功能型图。冉有华等应用基于证据理论的数据融合方法,将 2000 年中国 1:10 万土地利用数据、1:100 万中国植被图、1:10 万中国冰川图、1:100 万中国沼泽湿地图和 MODIS 2001 年土地覆盖产品 (MOD12Q1) 进行了融合,最终基于最大信任度原则进行决策,产生了基于 IGBP 分类系统的土地覆盖数据。新的土地覆盖数据在保持了我国土地利用数据的总体精度的同时,补充了中国植被图中有关植被类型和植被季相的信息,以及来自于中国湿地图和中国冰川图的更为可靠的湿地和冰川信息^[28]。刘伟等^[29]也将中国植被图中的植被分类等信息,融合到高分辨率的黑河流域土地覆盖图中,制备了基于 USGS 分类的黑河流域土地覆盖图,这一新产品比相应的全球数据在分类精度和空间分辨率方面都有很大的提高。这些新的土地覆盖图,显然更加适用于黑河流域的陆面过程及水文和生态模拟。

我国的土壤分类与通行的国际标准也不一致,此外,全球土壤质地图的分辨率过粗,在黑河流域完全不能体现出流域不同海拔高度土壤分布的特征。为了定量表征黑河流域土壤的非均匀分布,高艳红等在流域范围收集了 53 个土壤亚类剖面的质地数据,根据基于土壤发生学分类的 1:100 万比例尺黑河流域土壤图,对照土壤三角,与国际通用的土壤质地类型分类标准参比,建立了黑河流域 30 s 分辨率的土壤质地类型分布图。这套数据可以描述出原全球土壤数据不能表现出的山区和绿洲区的土壤非均匀分布的实际特征^[30]。

改进后的黑河流域高分辨率土地覆盖数据和土壤数据,在流域模拟中发挥了重要作用。例如,高艳红等采用上述土地覆盖图和土壤图,分别替换了 NCAR 的 (NCAR Meso-scale Model, MM5) 中尺度大气模型 MM5 中默认的土地覆盖和土壤数据;同时,用 2002 年一个完整年度的 1 km AVHRR 植被指数计算得到的植被覆盖度数据集替换了 (NCAR Meso-scale Model, NCAR 的 MM5) 中默认的 0.15 分辨率的相应数据。之后,分别比较了原数据集和改进后的高分辨率参数集对模拟结果的影响。结果表明,

中尺度陆面—大气耦合模拟的结果显著依赖于基础参数集,其改进对于提高山区和平原地区的气温、湿度和降水的模拟精度都有明显的贡献^[31]。这一实证研究,进一步说明了发展高质量的参数集是提高流域综合模拟能力的重要途径之一。

4 小结和讨论

数据,正在成为与实验室、观测设备和网络所并行的科学研究基础设施^[5]。数据集成,在流域科学中起着重要的作用,它是流域集成研究的重要环节,是发展、改进和验证模型的基础之一。我们以服务于流域科学——特别是服务于流域集成研究为目标,开展了黑河流域数据集成研究。其中,在数据库集成方面,建立了“数字黑河”信息系统,该系统内容翔实,数据类型全面,共享数据量已超过 1000 GB,在国内外都有大量用户,已经成为黑河流域研究的重要信息基础设施。在数据科学内容集成方面,重点开展了模型数据集研究。我们提出了针对流域综合模型及水文和生态模型的数据分类(验证和诊断数据、驱动数据、参数集)及其集成框架,初步完成了适用于各种水文、生态和陆面过程模型的黑河流域土地覆盖、土壤、遥感等参数集的制备。这些模型数据集,有力地支持了黑河模型集成研究工作^[32],并且可为其他流域的类似研究提供借鉴。

“数字黑河”的数据集成,历经过去 10 年的工作,已初具成效,但依然有许多需要进一步急迫开展的研究任务。

在数据库集成方面:

(1) 完善元数据。科学数据来源多样、管理复杂,数据的标准化不可能一蹴而就,特别是存在大量难以结构化的观测数据。因此,元数据发挥着核心的作用,其作用不可低估。必须着力完善数据的元信息描述,才能保证对各种数据的正确解释、广泛传播和持续利用。对于关键的数据集,还应专门撰写详细的数据文档。

(2) 数据的逻辑组织形式要更加丰富。应完善数据的导航、搜索和浏览,以便于用户通过各种类型的关键词、WebGIS、Google Earth 等方式,检索到所需的数据。检索结果应能够表现出数据的多元特征,按用户所期望的组合方式将相互关联的数据展现给用户。

(3) 完善数据服务。提供各种开放地理信息联盟 (Open Geospatial Consortium, OGC) 标准服务,如网络图层服务、网络地图服务、网络要素服务、数据

目录服务等。

(4) 开发数据质量控制工具。“数字黑河”信息系统中数据的质量控制,目前主要以手工为主,自动化的质量控制应是重点发展的方向。

(5) 完善数据共享政策。数据共享也是数据集成的重要组成部分^[33],共享需要完善的政策和机制保证,但观念的转变也是至关重要的,应率先示范,积极推动形成数据共享的共识。同时,在尽可能保持开放的数据共享政策的同时,逐步形成数据汇交的机制,完善在线数据提交工具,以及数据下载、使用和引用的追踪工具。

在模型数据集方面,驱动数据和参数集的质量和分辨率都有待于进一步的提高。

(1) 驱动数据。制备流域尺度的驱动数据在方法上已经没有大的障碍,关键在于提高数据的质量——特别是辐射和降水数据的质量。可行的途径是充分利用卫星和雷达气象学的成果,在区域气候模型中同化雷达降水、卫星辐射观测等资料。

(2) 土地覆盖数据。定义适合于多种模型模拟目标的土地覆盖 植被功能型数据,至少应与 IGBP 和美国地质调查局 (USGS) 的土地覆盖分类^[34,35],以及公用陆面模型 (Community Land Model, CLM) 模型所使用的植被功能型分类^[36]兼容。分辨率应达到 100 m,以满足水文和陆面过程模型中亚网格模拟的需求。更具有挑战性的是:全球土地覆盖 植被功能型划分,在流域尺度上,可能不足以表达植被在生理特征等方面的差异,应进一步扩展其分类和特征表述^[37]。这就需要在类型细分的基础上制备区域和流域尺度上的植被功能型图^[38],同时,通过查阅大量文献和汇总试验结果,给定各种植被功能型的参数取值范围。

(3) 土壤数据。全球土壤制图在近年来取得了很大的进展^[39],如整合的世界土壤数据库^[40],其分辨率,相比联合国粮农组织原世界土壤图,分辨率空前的提高了(1 500万到 1 100万或更高),精度和属性的丰富程度也有不同程度的提高。同时,我国 1 100万土壤图也基本实现了与国际通用的质地类型分类的参比^[41]。以上这些进展,都为制备更加可靠的流域尺度土壤图及相关属性参数提供了更好的条件,但流域尺度上对土壤制图分辨率的要求更高,这就意味着需要收集和采集更多的土壤剖面,并在 GIS 的支持下,综合植被、地形、气候等环境要素,制备精细的流域土壤图^[42]。

(4) 此外,应开发生成模型数据集的工具。包

括从基础数据集派生出模型所需数据的工具,以及由一个模型数据集的中间产品派生出各类模型所需的数据集的工具,并提供格式转换等功能。

本文没有讨论数据集成的第三个层次中的数据同化,事实上,流域科学从基础研究角度可以被看作是地球系统科学研究方法在流域尺度上的具体体现,这就意味着流域科学所需的数据复杂而多样。因此,应充分利用地球系统科学各分支学科已经积累的、具有不同时空尺度的各种资料,并在统一规范下,对这些资料进行同化、融合等更高层次的集成^[43],才能更好服务于流域科学乃至地球系统科学研究。

致 谢:在数字黑河信息系统的建设过程中,盖春梅、包亮、黄春林等参加了不同阶段的工作。在此致谢!

参考文献 (References):

- [1] Committee on Watershed Management, National Research Council. New Strategies for America's Watersheds [M]. Washington DC: National Academies Press, 1999.
- [2] Maidment D R. CUAHSI (The Consortium of Universities for the Advancement of Hydrologic Science) Hydrologic Information System: Overview of version 1.1 [R]. CUAHSI, 2008.
- [3] MLaughlin D. An integrated approach to hydrologic data assimilation: Interpolation, smoothing, and filtering [J]. *Advances in Water Resources*, 2002, 25: 1 275-1 286.
- [4] Li Xin, Cheng Guodong, Wu Lizong. Digital Heihe River Basin 1: An information infrastructure for the watershed science [J]. *Advances in Earth Science*, 2010, 25 (3): 297-305. [李新, 程国栋, 吴立宗. 数字黑河的思考与实践 1: 为流域科学服务的数字流域 [J]. *地球科学进展*, 2010, 25 (3): 297-305.]
- [5] Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age, National Academy of Sciences. Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age [M]. Washington DC: National Academies Press, 2009.
- [6] Li Xin. Arid region in northwest China [C]. Sun Jiulin, Lin Hai. *Earth System Science and Scientific Data*. Beijing: Science Press, 2009: 399-414. [李新. 西北干旱区 [C]. 孙九林, 林海. *地球系统研究与科学数据*. 北京: 科学出版社, 2009: 399-414.]
- [7] Li Xin, Cheng Guodong. On the watershed observing and modeling systems [J]. *Advances in Earth Science*, 2008, 23 (7): 756-764. [李新, 程国栋. 流域科学研究中的观测和模型系统建设 [J]. *地球科学进展*, 2008, 23 (7): 756-764.]
- [8] Li Xin, Ma Mingguo, Wang Jian, et al. Simultaneous remote sensing and ground-based experiment in the Heihe River Basin: Scientific objectives and experiment design [J]. *Advances in Earth Science*, 2008, 23 (9): 897-914. [李新, 马明国, 王建, 等. 黑

- 河流域遥感—地面观测同步试验:科学目标与试验方案[J]. 地球科学进展, 2008, 23(9): 897-914.]
- [9] Committee on the Human Dimensions of Global Change, NRC. People and Pixels: Linking Remote Sensing and Social Science [M]. Washington, DC: National Academies Press, 1998.
- [10] Sun Shufen. Physical and Geochemical Mechanisms and Parameterization Models for Land Surface Processes [M]. Beijing: China Meteorological Press, 2005. [孙淑芬. 陆面过程的物理、生化机理和参数化模型 [M]. 北京: 气象出版社, 2005.]
- [11] Shi Xuezheng, Yu Dongsheng, Sun Weixia, *et al*. Reference benchmarks relating to great groups of genetic soil classification of China with soil taxonomy [J]. *Chinese Science Bulletin*, 2004, 49(14): 1 507-1 511. [史学正, 于东升, 孙维侠, 等. 中美土壤分类系统的参比基准研究: 土类与美国系统分类土纲间的参比 [J]. 科学通报, 2004, 49(13): 1 299-1 303.]
- [12] Cheng Guodong, Liu Fengjing, Wang Genxu, *et al*. Research Report on the Rational Utilization of Water Resources and the Coordination between Socioeconomic Development and Natural Environment in the Heihe River Basin, Northwest China [R]. Lanzhou: Lanzhou Institute of Glaciology and Geocryology, CAS, 1999. [程国栋, 刘凤景, 王根绪, 等. 黑河流域水资源合理利用与社会经济和生态环境协调发展研究 [R]. 兰州: 中国科学院兰州冰川冻土研究所, 1999.]
- [13] Li Xin, Cheng Guodong, Ding Yongjian, *et al*. Design of the water resource information system of the Heihe River Basin, northwest China [J]. *Journal of Desert Research*, 2000, 20(4): 378-382. [李新, 程国栋, 丁永建, 等. 黑河流域水资源信息系统设计 [J]. 中国沙漠, 2000, 20(4): 378-382.]
- [14] Huang Chunlin, Li Xin. Comparison of three different methods to distribute spatial data on the Internet [J]. *ACTA Geographica Sinica*, 2002, 57(suppl): 44-51. [黄春林, 李新. 比较三种不同的因特网空间数据发布方法 [J]. 地理学报, 2002, 57(增刊): 44-51.]
- [15] Ma Minguo. Database of the Digital Heihe River Basin and Its Application on Ecology and Hydrology [D]. Beijing: Graduate School of the Chinese Academy of Sciences, 2003. [马明国. 数字黑河数据库及其在生态—水文方面的应用研究 [D]. 北京: 中国科学院研究生院, 2003.]
- [16] Huang Chunlin, Li Xin. Application of HDF-EOS data format in manipulating spatial data [J]. *Remote Sensing Technology and Application*, 2001, 16(4): 252-259. [黄春林, 李新. HDF-EOS 数据格式在处理空间数据中的应用 [J]. 遥感技术与应用, 2001, 16(4): 252-259.]
- [17] Lu Ling, Cheng Guodong. Application of VRML technology in the decision support system of the Heihe River Basin [J]. *Remote Sensing Technology and Application*, 1999, 14(2): 15-20. [卢玲, 程国栋. VRML 技术在黑河水资源决策支持系统中的应用 [J]. 遥感技术与应用, 1999, 14(2): 15-20.]
- [18] Nan Zhuotang, Wang Liangxu, Li Xin. EOS data dumper: An automatic downloading and re-distributing system for free EOS data [J]. *Journal of Glaciology and Geocryology*, 2007, 29(3): 463-469. [南卓铜, 王亮绪, 李新. EOS Data Dumper—EOS 免费数据自动下载与重发布系统 [J]. 冰川冻土, 2007, 29(3): 463-469.]
- [19] Li Xin, Nan Zhuotang, Wu Lizong, *et al*. Environmental and ecological science data center for west China: Integration and sharing of environmental and ecological data [J]. *Advances in Earth Science*, 2008, 23(6): 628-637. [李新, 南卓铜, 吴立宗, 等. 中国西部环境与生态科学数据中心: 面向西部环境与生态科学的数据集成与共享 [J]. 地球科学进展, 2008, 23(6): 628-637.]
- [20] Wang Liangxu. Research on Open Source Software Based Geospatial Data Sharing Platform [D]. Beijing: Graduate University of Chinese Academy of Sciences, 2008. [王亮绪. 基于开源技术的空间数据共享平台研究 [D]. 北京: 中国科学院研究生院, 2008.]
- [21] Ma Mingguo, Hu Xiaoli, Song Yi, *et al*. Development and application on the information system of water conservancy project statues for Zhangye city [J]. *Remote Sensing Technology and Application*, 2009, 24(5): 559-566. [马明国, 胡小莉, 宋怡, 等. 张掖市水利工程现状信息系统开发与应用研究 [J]. 遥感技术与应用, 2009, 24(5): 559-566.]
- [22] Hu Xiaoli, Lu Ling, Ma Mingguo, *et al*. The irrigation channel system mapping and its structure analysis for the Zhangye oasis in the middle Heihe River Basin [J]. *Remote Sensing Technology and Application*, 2008, 23(2): 208-214. [胡晓利, 卢玲, 马明国, 等. 黑河中游张掖绿洲灌溉渠系的数字化制图与结构分析 [J]. 遥感技术与应用, 2008, 23(2): 208-214.]
- [23] Hu Yinqiao, Gao Youxi. Some new understandings of processes at the land surface in arid area from the HEIFE [J]. *Acta Meteorologica Sinica*, 1994, 52(3): 285-296. [胡隐樵, 高由禧. 黑河实验 (HEIFE)——对干旱地区陆面过程的一些新认识 [J]. 气象学报, 1994, 52(3): 285-296.]
- [24] Tao Zehong, Zuo Hongchao, Hu Yinqiao. HEIFE data base (HDB) [J]. *Plateau Meteorology*, 1994, 13(3): 369-376. [陶泽宏, 左洪超, 胡隐樵. 黑河实验数据库 (HDB) [J]. 高原气象, 1994, 13(3): 369-376.]
- [25] Wang Xuemei, Li Xin, Ma Mingguo. Pixelizing the population statistics of inland river basin in arid regions—A case study of the Heihe River Basin [J]. *Journal of Arid Land Resources and Environment*, 2007, 21(6): 39-47. [王雪梅, 李新, 马明国. 干旱区内陆河流域人口统计数据的空间化——以黑河流域为例 [J]. 干旱区资源与环境, 2007, 21(6): 39-47.]
- [26] Gao Yanhong, Cheng Guodong, Li Xin, *et al*. Comparison of atmospheric data scaling transfer scheme in the Heihe River Basin [J]. *Plateau Meteorology*, 2005, 24(4): 563-569. [高艳红, 程国栋, 李新, 等. 黑河流域大气资料尺度转换的对比分析 [J]. 高原气象, 2005, 24(4): 563-569.]
- [27] Chen Rensheng, Kang Ersi, Li Xin, *et al*. Hourly incident solar radiation model under actual weather and terrain conditions: A case study in the Heihe River Basin [J]. *Journal of Desert Research*, 2006, 26(5): 773-779. [陈仁升, 康尔泗, 李新, 等. 任意地形实际天气条件下小时入射短波辐射模型——以黑河流域为例 [J]. 中国沙漠, 2006, 26(5): 773-779.]

- [28] Ran Youhua, Li Xin, Lu Ling. China land cover classification at 1 km spatial resolution based on a multi-source data fusion approach[J]. *Advances in Earth Science*, 2009, 24(2): 192-203. [冉有华, 李新, 卢玲. 基于多源数据融合方法的中国 1 km 土地覆盖分类制图[J]. 地球科学进展, 2009, 24(2): 192-203.]
- [29] Liu Wei, Gao Yanhong, Li Haiying, et al. Landuse pattern of the Heihe River Basin and its impact modeling[J]. *Plateau Meteorology*, 2007, 26(2): 278-285. [刘伟, 高艳红, 李海英, 等. 黑河流域土地覆盖分类数据的建立及其影响的模拟[J]. 高原气象, 2007, 26(2): 278-285.]
- [30] Gao Yanhong, Liu Wei, Cheng Guodong, et al. Setup and validation of the soil texture type distribution data in the Heihe River Basin[J]. *Plateau Meteorology*, 2007, 26(5): 967-974. [高艳红, 刘伟, 程国栋, 等. 黑河流域土壤质地分类数据建立及其模拟效果检验[J]. 高原气象, 2007, 26(5): 967-974.]
- [31] Gao Y H, Chen F, Barlage M, et al. Enhancement of land surface information and its impact on atmospheric modeling in the Heihe River Basin, northwest China[J]. *Journal of Geophysical Research*, 2008, 113(D20S90), doi: 10.1029/2008JD010359.
- [32] Cheng Guodong, Li Xin, Kang Ersi, et al. Integrated Model Development and Modeling Environment Building for Interdisciplinary Studies in the Heihe River Basin[R]. Lanzhou: Cold and Arid Regions Environmental and Engineering Research Institute, CAS, 2008. [程国栋, 李新, 康尔泗, 等. 黑河流域交叉集成研究的模型开发和模拟环境建设结题报告[R]. 兰州: 中国科学院寒区旱区环境与工程研究所, 2008.]
- [33] Huang Dingcheng, Guo Zengyan. Study on the Management of Scientific Data Sharing[M]. Beijing: China Science and Technology Press, 2002. [黄鼎成, 郭增艳. 科学数据共享管理研究[M]. 北京: 中国科学技术出版社, 2002.]
- [34] Loveland T R, Reed B C, Brown J F, et al. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data[J]. *International Journal of Remote Sensing*, 2000, 21(6/7): 1303-1330.
- [35] Dai Y, Zeng X, Dickinson R E, et al. The common land model[J]. *Bulletin of American Meteorological Society*, 2003, 84(8): 1013-1023.
- [36] Bonan G, Levis S, Kergoat L, et al. Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models[J]. *Global Biogeochemical Cycles*, 2002, 16(2): 5-15-23.
- [37] Smith B, Prentice I C, Sykes M T. Representation of vegetation dynamics in the modeling of terrestrial ecosystems: Comparing two contrasting approaches within European climate space[J]. *Global Ecology and Biogeography*, 2001, 10: 621-637.
- [38] Ni Jian. Plant functional types and biomes of China at a regional scale[J]. *Acta Botanica Sinica*, 2001, 43(4): 419-425. [倪健. 区域尺度的中国植物功能型与生物群区[J]. 植物学报, 2001, 43(4): 419-425.]
- [39] Sanchez P A, Ahamed S, Carre F, et al. Digital soil map of the world[J]. *Science*, 2009, 325(5941): 680-681.
- [40] FAO/IIASA/ISRIC/ISS-CAS/JRC. Harmonized World Soil Database (version 1.1)[R]. Rome, Italy: FAO and Laxenburg, Austria: IIASA, 2009.
- [41] Shi X Z, Yu D S, Wamer E D, et al. Cross-reference system for translating between genetic soil classification of China and soil taxonomy[J]. *Soil Science Society of America Journal*, 2006, 70(1): 78-83.
- [42] Zhu A xing. Fine-Resolution Digital Soil Surveying: Model and Method[M]. Beijing: Science Press, 2008. [朱阿兴. 精细数字土壤普查模型与方法[M]. 北京: 科学出版社, 2008.]
- [43] Zhou Xiujie. Thoughts on the earth system science[J]. *Advances in Earth Science*, 2004, 19(4): 513-515. [周秀骥. 对地球系统科学的几点认识[J]. 地球科学进展, 2004, 19(4): 513-515.]

Digital Heihe River Basin. 2: Data Integration

L I Xin, WU Lizong, MA Mingguo, GE Yingchun,
RAN Youhua, WANG Liangxu, NAN Zhuotong

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

Abstract: Data integration is fundamentally important for the development, modification and verification of integrated watershed models and therefore is a key research topic in the watershed science. This paper provides an overview of data integration in the Digital Heihe River Basin research. First, the data integration is defined, which should have three basic characteristics as data integrity, accessibility and stewardship. The preparation of model data sets, which can be categorized into forcing data, parameter data and validation data from a viewpoint of supporting integrated watershed modeling, is also an important means of data integration. Secondly, it is introduced that

the basic geographic data, various kinds of observation data, scientific experiment data, and existing model data sets of the Heihe River Basin have been systematically collected, and a comprehensive web-based information system of the Digital Heihe River Basin has been developed to support data distribution, with over 1000 GB of data available on the Internet for free downloading. Most of the data were archived with metadata and documentation when necessary. Thirdly, the preparation of model data sets in the Heihe River Basin is introduced. The statistical and dynamic methods for atmospheric data downscaling were developed. The new land cover map and soil map suitable for hydrological and ecological modeling were developed by data fusion method. At last, the paper clears the challenges and demands of data integration in the watershed science studies in the Heihe River Basin, and proposes that a more reliable and higher resolution model data sets needs to be developed to support the integrated watershed modeling.

Key words: Digital River Basin; Data integration; Model data sets; Heihe River Basin

旨在减缓全球变化的海洋地球工程有可能加速气候变暖

一方面将深海的富含养分的水送到阳光充足的海表面,促进海藻的生长;另一方面将大气中的 CO_2 输送到海底,这是一种减缓全球变暖的途径。然而,莱布尼兹海洋科学研究中心 (IFM-GEOMAR) 的 Andreas Oschlies 教授的最新研究却给这种观点泼了冷水。

该研究的合作者、美国国家海洋学中心 (NOCS) 的 Andrew Yool 博士表示,计算机模拟显示,地球工程计划有可能加速气候变化,从而面临失败。

如果国际社会不能将 CO_2 排放量减少到使其不对全球变化产生影响的限度,那么地球工程计划有可能被其他计划所取代。转而实施一个或多个规模更大的地球工程计划,以减少 CO_2 的排放。

一个可能的方法是:将几百米之下的富含养分的海水输送到海表面,促进浮游植物的生长。大量的微小海藻主宰了海洋表面的大部分生产力。其目的是模仿海洋的上升流,通过光合作用,增加海洋对 CO_2 的吸收量。然后通过浮游植物死亡和下沉到海底,将 CO_2 送到深海,从而达到长期固定 CO_2 的目的。

在先前进行的一项由 Yool 领导的研究中,使用了海洋环流模型计算显示,要想对全球变化产生有效地影响,需要数以千万计的输送管道。即使不考虑制造这些管道的技术难度和经济成本,海洋中到底可以封存多少 CO_2 以及这种方式有何风险仍然是需要回答的问题。

在该项新的研究中,科学家们采用更加综合的全球数值模式模拟这些问题。结果显示,在最乐观的前提下,每年将会有大约 3 亿 t CO_2 被输送到海底。而这个数字还不到人类年均 CO_2 排放量 (36 亿 t) 的 10%。

一个令人吃惊的模拟结果是,这种地球工程对大陆的影响大于对海洋的影响。冷水来到海表面,冷却了大气以及地表,减缓了土壤有机质的分解,最后导致大约 80% 的 CO_2 被封存在陆地上。研究者指出,这种远程的、分布式的固碳方式将会使得监测工作具有很大的难度。

模拟结果的另一个重要发现是,一旦这些模拟的管道被关闭,大气中 CO_2 的含量和海表面温度将会以更快的速度升高。这些发现表明,该地球工程计划在由于某种原因关停的时候,具有额外的环境成本。

所有模式都是基于假设的,现实中的情况还会面临许多不确定性。基于该研究结果,很难发现人造管道可以通过提升海表面生产力而减缓全球变化。

(王金平 编译)

原文题目: Ocean Geoengineering Scheme No Easy Fix for Global Warming

来源: <http://www.sciencedaily.com/releases/2010/02/100216221305.htm>