

黑河综合遥感联合试验地面观测 数据质量控制与评价

马明国

(中国科学院寒区旱区环境与工程研究所,甘肃 兰州 730000)

摘要:本文主要介绍了“黑河综合遥感联合试验”(Watershed Allied Telemetry Experimental Research, WATER)的地面观测数据(气象水文和涡动相关数据除外)的数据质量控制和处理过程。从观测规范的制定和执行、数据处理和规范化、数据文件命名规范化、数据质量评价几个方面进行详细阐述。可以为数据用户提供数据质量参考信息,也可以为以后类似试验的开展提供一些数据质量控制和处理的借鉴。通过最后的数据评价可以看出 WATER 试验的地面观测数据总体上具有较高的数据质量。这些质量控制措施为最后获取高质量的试验数据提供了重要的保证。利用这一套多尺度的、标准化的高质量综合数据集开展大量数据分析和应用研究是下一步工作的重点。

关键词:WATER;黑河流域;地面观测数据;数据质量控制

中图分类号:TP 79 **文献标志码:**A **文章编号:**1004-0323(2010)06-0766-06

1 引言

2008年3月至7月开展的“黑河综合遥感联合试验”(Watershed Allied Telemetry Experimental Research, WATER)分为两个阶段,第一阶段是以冻土和积雪为主要研究对象的上游寒区水文试验;第二阶段是以森林水文过程及干旱区生态水文过程为主要目标的森林水文试验和中游干旱区水文试验。李新^[1]等详细介绍了 WATER 的科学背景、科学问题、研究目标以及观测试验方案和观测系统布置。王建^[2]等介绍了 WATER 上游寒区水文试验的目标与研究内容、试验区选择、样方布设、航空飞行试验和地面同步观测试验以及寒区水文长期观测试验等信息,并针对积雪、冻土和寒区水文过程总结了基于试验的初步研究成果。马明国^[3]等对 WATER 中游干旱区水文试验的研究目标和研究内容进行了简要阐述,重点介绍了航空飞行试验、地面同步试验和加密观测试验的样方布置、数据获取情况。本文则侧重于介绍上游寒区水文试验和中游干旱区水文试验地面观测部分的数据质量控制过程。地面观测试验包含航空和卫星地面同步观测试验和加密

观测试验,观测内容包含各种变量和参数的观测,还包括水气和 CO₂ 通量、气象水文、降雨、径流和其他水文要素的观测。本文涉及的观测内容以地面同步或者加密开展的各种变量和参数的观测为主,张智慧^[4]等对涡动相关通量数据的处理和分析过程及产品生成进行了介绍,黄广辉^[5]等则对水文气象资料处理、质量控制与产品生成进行了详细介绍。数据质量直接影响到数据的科学价值和使用价值,例如国际上公开发布的长时间系列遥感数据集大都进行了严格的质量控制,有些数据还逐条给出了质量控制符,例如 SPOT/VEGETATION 数据集和 MODIS 数据产品。同样,国际上的一些陆面过程试验产生了大量地面观测数据,也进行了数据质量控制,例如寒区陆面过程试验(CLPX)^[6]和欧洲陆地碳平衡评估(CarboEurope)^[7]。本文介绍 WATER 试验中地面观测数据的质量控制和规范化过程,包括观测规范的制定和执行、数据处理和规范化、数据文件命名规范化和数据质量评价4个部分。目的有两个,一是为数据使用人员提供数据质量参考信息,便于更好地将数据应用到研究中去;二是为以后类似试验的

收稿日期:2010-06-16;修订日期:2010-11-23

基金项目:中国科学院西部行动计划(二期)项目(KZCX2-XB2-09-03)、国家重点基础研究发展计划项目(2007CB714401)和中国科学院“西部之光”人才培养计划项目(CACXO728501001)资助。

作者简介:马明国(1976-),男,研究员,主要从事中国西北生态遥感研究。E-mail: mmg@lzb. ac. cn.

开展提供一些数据质量控制和处理的借鉴。

2 观测规范制定与执行

由于本子试验参加单位和人数众多,参加人员学科背景多样化,在开展上游寒区水文试验和中游干旱区水文试验之前,为了保证数据质量,提高数据的统一性和可对比性,针对不同的观测项目分别制定了统一的数据观测规范供大家执行。本次试验所编写的规范包括《作物光谱测量规范》、《Delta 土壤水分速测仪测量规范》、《Steven POGO 便携式土壤水分传感器测量规范》、《地表粗糙度测量规范(土壤部分)》、《地表粗糙度测量规范(积雪部分)》、《地基微波辐射计观测规范》、《冻土温度自动观测系统观测规范》、《冻融深度观测规范》、《玻璃液体地温计观测规范》、《地表冻融卫星同步观测样方规范》、《土壤含水量环刀法测量规范》、《反照率表观测规范》、《光子计观测规范》、《光学和热红外 BRDF 测量的规范》、《航空和卫星同步定标规范》、《红外波谱测量规范》、《红外光谱仪 102F 观测规范》、《中科院地理所自制比辐射率观测规范》、《星载热红外遥感数据测量规范》、《积分球操作规范》、《积雪观测规范》、《农业相关的作物和土壤测量规范》、《地表通量与气象水文要素观测规范》、《雷达及地面降水观测规范》、《森林相关参数测量规范》、《生态调查及实验规范》、《叶面积仪测量规范》、《叶倾角等参数测量规范》、《真实性检验相关的地面测量规范》、《植被占空比测量规范》。观测规范的制定遵循以下几个原则:

(1) 尽量使用和参照已有国家或者行业测量规范,但为了增加可操作性,将规范中与本试验有关的项目和内容挑选出来,形成简化本。

(2) 由对某一观测项目具有丰富观测和研究经验的专家来制定该项目的观测规范。

(3) 观测规范要充分考虑到野外的可操作性,对原理部分不需占用大量篇幅进行详细论述,可以通过建议阅读文献的方式给出,重点介绍实际操作过程的详细步骤和注意事项。

(4) 观测项目如果同时由多种仪器进行观测,可以针对观测仪器分别制定相应观测规范,也可针对观测项目制定总的观测规范,并在其中针对每个仪器给出详细操作流程和注意事项。

(5) 每个观测项目都设计固定的数据记录表格,包含表头信息和记录表格。表头信息必须包含观测的必要描述信息,例如测量日期、地块编号(样方、采样点代码)、测量人员、记录人员、测点位置信

息(经度、纬度、海拔)和气象状况等。其他表头信息则根据实际观测项目进行设计,例如仪器名称和编号、站点场景照片文件名、测量角度、测量高度等。要求记录人员在观测时必须认真填写表头信息和观测记录,保证信息完整,方便查询。实地测量时,以打印的纸质表格为主,用 2B 铅笔填写。

在实际观测过程中,为了保证观测规范的有效执行和观测的井然有序,在以下几个方面开展了细致的统一组织和管理工作的。

(1) 首先是通过预试验熟悉仪器,并检验各种观测规范的合理性和可操作性,在预试验中及时发现问题,进一步完善观测规范。

为了保证观测数据的质量,在野外观测实施之前,组织观测人员集中学习观测规范,并对某些观测项目集中在野外进行操作讲解和示范。在观测过程中,安排固定的人员实地巡检,对不规范的观测进行及时纠正。

(2) 为了保证数据的完整性,尽量避免数据丢失,要求每个加密观测区当天处理观测数据,分析观测样品,当天录入所有记录表格和处理后的数据信息,由加密观测区负责人审核后,原始纸质表格和电子文档交由专人保管、存储和备份。

自动记录仪器的数据当晚需要导出,存储。

(3) 所有观测仪器统一编号,用防水标签粘贴在仪器显眼的空处,填写记录表格表头信息时必须准确记录仪器的名称和编号信息。

(4) 在观测开始前和观测结束后,针对部分观测仪器进行统一标定,给出标定系数,应用到最后的观测数据的处理中,增加数据的可比性。例如红外温度计、光谱仪、便携式土壤水分速测仪。

3 数据处理与规范化

试验结束后,所有的数据最终汇总到“黑河综合遥感联合试验”数据信息系统,但在数据入库到最后提供用户共享前,还需要对数据进行必要的预处理和规范化工作。主要包含以下步骤:

(1) 纸质和电子版数据表格之间的一一对应检查。首先检查纸质数据表格是否全部录入计算机,录入计算机的电子表格是否能够找到原始记录表格。然后对纸质数据和电子版数据记录的录入对照检查,保证录入计算机的数据是纸质数据原始记录。

(2) 数据完整性检查。检查数据的表头信息是否完整,数据记录是否有缺失。发现问题,尽量找到数据观测人员或者观测记录人员进行补充。

(3) 数据预处理。包括利用定标系数将原始观测值转换成具有物理意义的值,例如反照率是将观测到电压值计算成辐射通量;光谱仪观测是将观测到 DN 值计算成反射率,这个过程中需要考虑白板反射率和试验前后光谱仪对比校正的标定系数;固定自记点温度计和手持式红外温度计数据是将观测数据经过辐射定标和仪器比辐射率纠正后计算为温度数据;CE318 太阳分光光度计的数据是利用太阳辐射数据反演获得光学厚度、瑞丽散射、气溶胶光学厚度等一系列大气参数。预处理过程在元数据或相应数据文档中进行了说明。

(4) 数据格式的规范化。虽然试验前期制定了大量规范,实际操作时每个观测小组由于主观性和随意性仍然会有一些的差别,例如同样的数据记录表格,有的小组用 Word 格式存贮,有的利用 Excel。观测数据的存储更是差别甚大,例如叶面积指数,存储格式包含文本、Word 和 Excel 三种格式,即使同种格式版本也不尽相同。针对这些格式差别的问题,尽量将其统一成同一数据格式。而部分不能够

转换成 Excel 表格的数据项目,至少同一观测项目的数据存储为同一格式。该项工作难度较大,工作量也非常大,目前本次试验的数据还没有实现这一步骤。但建议下一次的试验在最开始规范制定和试验设计时就提前考虑这一点,事先做好格式统一的电子表格,尽量做到大家汇交的数据格式保持一致,只做少量完善工作即可达到规范化的目的。

4 数据文件命名规范

由于此次地面试验涉及众多观测项目,由多个观测小组完成,在命名上各有不同,同一要素在不同的加密观测区也不尽相同,从而导致数据只能存贮在每个加密观测区专用目录下面。无论是数据管理还是数据应用都存在较多不便。为了方便管理人员管理数据,也方便用户从文件名来了解数据的内容,我们制定了地面观测数据的命名规范。文件名的形式为 WATER_YK_LAI_20080406_YMS1_Sync-ASAR&PALSAR_Data.xls。各项的含义见表 1,各项关键字之间用“_”连接起来。

表 1 数据文件命名规范中各项含义定义

Table 1 Meaning of each item for the name notation of the data files

编号	示例	含义	是否是必须项
1	WATER	试验名称,针对此次试验全部使用 WATER。	是
2	YK	加密观测区名称,针对此次试验的加密观测区代码分别是:(1)盈科绿洲 YK;(2)花寨子荒漠 HZZ;(3)临泽站 LZS;(4)临泽草地 LZG;(5)阿柔草地 AR;(6)冰沟流域 BG;(7)大野口流域 DYK;(8)排露沟流域 PLG;(9)扁都口 BDK;(10)张掖市 ZY;(11)张掖观象台 ZYNOC。	是
3	LAI	观测项目名称,针对每个观测项目进行统一命名,详细信息参考数据命名规范文档 ^[2] 。	是
4	20080406	观测时间年月日,采用 yyyymmdd 的形式	是
5	YM	观测样地、样带和样方的名称,如果是在固定的样方样带上开展观测,就增加第 5 项,而没有在固定样方样带内开展观测,属于临时布置的观测点,但有 GPS 坐标,则第 5 项用“LS”代表,没有 GPS 坐标的话,则第 5 项用“XX”代表。样地、样带和样方的详细信息参考数据命名规范文档 ^[2] 。	否
6	S1	盈科绿洲加密观测区样地下设计了样方。	否
7	Sync-ASAR+PALSAR	同步的传感器名称,“Sync-”代表同步,后面接传感器名称,多个传感器之间用“+”连接	否
8	DATA	补充说明,如果数据项既有记录表格,也有照片或者文档信息,利用补充说明进行区分。	否

利用规范后的文件命名,可以快速实现数据的检索和查询,例如只要输入“*_LAI_*”就可以提取所有加密观测区的叶面积指数观测数据,而输入“*_20080528_*_Sync-ASTER_*”就可以提取所有 2008 年 5 月 28 日和 ASTER 卫星过境同步的观测

数据。同时,利用这些关键字信息,还可以实现关系数据库来管理这些数据,这个关系数据库的字段对应这些关键字信息,每个数据文件对应一条记录,前 4 个字段对应的必填项,每条记录都有值,而后面的字段,根据文件名中该关键字是否有这项值而对于

表 2 黑河综合遥感联合试验地面数据质量评价结果汇总表

Table 2 Summary table of the quality evaluation for the ground observation data of the Watershed Allied Telemetry Experimental Research

观测变量	数据质量描述
微波辐射计和散射计 CE318	多数可用,仅 2008 年 3 月 12 日的的数据仅供参考。 多数可用,个别数据质量不好,仅供参考。
地物光谱	多数可用,部分存在没有经过预处理和缺少地物描述信息的问题。
地面观测辐射温度	阿柔加密观测区部分可用存在信息描述不全的问题;扁都口加密观测区多数可用,其他为部分可用存在信息描述不全的问题;临泽站加密观测区多数数据可用,预试验期间的数据缺少描述信息不可用;临泽草地加密观测区数据可用,但需要将观测目标信息添加到观测数据表格中;盈科绿洲和花寨子加密观测区多数数据可用,少数数据需要补充必要的说明信息,个别数据由于仪器问题不可用。
热像仪	可用。
反照率	多数可用,个别因缺少探头高度、地物照片等信息仅供参考,个别因云的影响而不稳定不可用。
叶面积	多数可用,临泽草地站部分数据因仪器原因不可用,部分数据未做预处理。
光合作用	多数可用,个别数据未做预处理不可用。
fPAR	盈科加密观测区数据除了 2008 年 5 月 30 日不可用以外其他数据均可用,但需要补充样地记录数据;临泽站和临泽草地加密观测区数据由于仪器精度原因,仅供参考。
生物量	除了个别数据缺少干重数据外,其他均可用
叶绿素、覆盖度及光合 样方调查	多数数据可用,部分数据缺少样地信息。 多数数据可以,个别数据存在无照片或者照片与样方不对应以及缺乏干重数据的问题。
土壤水分	多数可用,同步测量的地表温度数据需要进行标定处理;部分观测需要将观测目标信息添加到观测数据表格中;预试验数据部分因缺少位置信息不可用;部分烘干法测量结果缺失干重不可用。 临泽草地加密观测区手持水分速测仪和 PR2 的测量由于受到土壤盐分影响,测量结构明显偏大,仅供参考。
土壤温度	可用,需要对针式温度计标定。
土壤质地和容重	多数可用,个别因缺少位置信息不可用。
地表粗糙度	多数可用,个别缺少原始照片。
积雪温度	多数可用,但是需要对针式温度计校正,很多温度高于 0℃。
积雪粒径	多数可用,但是不同人观测结果之间可比性较差。
积雪深度	可用。
积雪剖面	多数可用,但是因为是固定分层采样会丢失一些薄层信息。
雪特性分析仪	可用。
雪深花杆观测	可用,但是观测时间密度稍差。
森林测树调查	可用。
森林站超级样地全站仪测量	可用。
土壤冻结深度	可用。
降雨截留	多数可用,个别因超过集雨器容量而不可用。
自制 lysimeter 测量蒸散发	大野口和盈科加密观测区可用,临泽草地加密观测区仅供参考。
凝结水	可用。
样方位置信息	可用。
灌区统计年报	可用。
水库收支	可用。
地下水	可用。
土地利用调查	可用。

该字段是否为空,最后两个字段分别是该数据文件的文件名和存储位置。这样可以方便管理人员管理试验获得的数据资源,同时还为用户提供通过 SQL 对数据记录进行查询的功能,方便用户快速寻找和

提取感兴趣的数据资源。

5 数据质量评价

为了保证数据科学性和可用性,在数据发布和

共享之前,邀请邀请对本领域具有丰富观测和研究经验的专家对地面观测数据进行了细致全面的质量评价工作。在专家遴选上,尽量是其熟悉的领域。正对逐条数据评价的步骤包括数据检查、可用性建议和总体评价等。数据检查包含数据实体和元数据间的一一对应关系检查、数据命名规范的检查、数据说明文档检查、数据预处理检查、数据一致性检查、数据记录完整性检查等方面。数据质量可用性建议分为可用、部分可用、仅供参考和不可用 4 种类型,主要是专家在对数据测量和预处理规范检查的基础上,根据自己掌握的观测要素在不同地表类型上的取值范围来分析数据测量值的精度,判断数据的可用与否。可用指的是该数据各方面指标都达到要求,总体质量较好,可以用于分析、建模和验证等研究中,此类数据可以最终发布。部分可用是指观测数据的部分记录质量较好,部分数据记录则存在问题,用户可以通过数据标识符来判断该数据的可用性,此类数据可以发布,质量较好的数据可以用于最终发表。仅供参考指的观测数据的存在一定的偏差,数据可信度较低,研究人员可以参考该测量值,但不能用于最终发表。不可用是指数据的质量较差,存在测量信息不完整或者缺失、测量不规范、数据明显偏差等问题,数据不能被用于相关研究,这类数据不出现在数据产品中。数据质量总体评价是在以上数据检查和可用性分析的基础上,对数据的质量给出一个总体的评价,以使用户在使用数据时参考。由于观测项目众多、数据记录量庞大和时间有限,专家不可能深入分析和应用数据,因此专家的评价意见也只是一参考信息。研究人员在使用数据过程中,可能还会发现大量问题,这些信息可以考虑通过数据信息系统建立的反馈机制来收集和汇总,然后附注到数据说明文档里,供后续用户参考。

专家对数据评价的结果表明,此次试验的数据多数都是可用的,数据的总体质量较高(见表 2)。但存在较多数据处理不完善和表头信息不完整的问题,需要开展进一步的处理工作。

6 结 语

在 WATER 地面观测试验开展过程中,通过试验前期的观测规范制定、观测期间观测规范执行和后期数据处理及数据形式的规范化等数据质量控制工作,均对最后获取高质量的试验数据提供了重要的保证。但也存在着较多问题和有待改进的地方,主要包含以下几个方面。首先是相关的观测规范、

数据格式、数据命名规则、元数据标准都需要尽早制定,便于观测人员在试验开展之前有足够的时间进行学习 and 训练,在试验执行期间必须严格按照相关规范制定的流程进行观测,数据表格格式和录入都需要严格统一标准,每一个加密观测区需要安排专人负责测量规范执行情况检查,对每天汇交的观测数据进行质量检查和控制,对数据质量有问题的要当场及时纠正,这样既可以保证数据质量,也可以大大减少后期数据处理工作。其次是对可以使用自动仪器进行观测的项目,尽量利用自动仪器进行观测,以减少因为人为因素导致的数据质量问题。第三,必须要对所有观测仪器进行标定,对观测中出现问题的仪器要及时进行修理和重新标定,某一仪器的操作最好固定为专门人员,而不要频繁更换观测人员,这样既减少因为人员更新需要重新熟悉仪器导致的数据不一致,也减少因不同人员观测习惯导致的观测结果不一致。第四,某一项目的观测人员尽量是熟悉该观测项目的研究人员或者是研究生,可以在观测过程中随时对观测结果进行分析和质量评价,边观测边分析,及时发现问题和解决问题,以提高数据质量。得到一套高质量的流域综合数据集是试验的重要目标之一,在今后很长一段时间内,我们将致力于数据的推广和应用,和广大同行一起利用这一套高分辨率的、多尺度的、较高质量的数据集,发展从航空和卫星遥感资料反演和间接估计水循环各分量(例如水热通量、蒸散发、土壤水分等)及其他地表过程分量(例如地表温度、生物物理参数等)的模型和算法,利用地面观测资料进行结果验证和算法的改进。我们热忱欢迎同行使用 WATER 试验数据并对数据提出宝贵建议。

致谢:感谢车涛、陈尔学、楚荣忠、胡泽勇、晋锐、李弘毅、刘良云、刘强、刘绍民、任华忠、田静、王建、肖青、阎广建、张立新、张仁华、赵永超、周春艳等参与地面测量规范的编写;感谢陈玲、杜永明、范闻捷、盖迎春、胡晓利、康国婷、李丽、刘思含、宋怡、谭俊磊、陶欣、王大成、王颢星、王旭峰、吴朝阳、吴月茹、辛晓洲、张智慧、周春艳等参与数据处理和规范化工作;感谢车涛、陈尔学、晋锐、刘良云、刘强、王锦地、王维真、辛晓洲、阎广建、张立新等老师参与数据数据质量评价工作;感谢中科院寒区旱区环境与工程研究所、遥感应用研究所、地理科学与资源研究所、研究生院、新疆生态与地理研究所、对地观测与数字地球中心,北京师范大学、成都电子科技大学、北京大学、

南京大学,中国农业科学院,甘肃省祁连山水源涵养林研究院、新疆气象局、甘肃省气象局和张掖市气象局,德国海德堡大学等兄弟单位的科研人员和学生共同参与地面观测试验;感谢其他参与和关心本次试验的单位和个人。

参考文献:

- [1] Li X, Li X W, Li Z Y, *et al.* Watershed Allied Telemetry Experimental Research [J]. *Journal of Geophysical Research*, 2009, 114(D22103), doi:10.1029/2008JD011590.
- [2] Wang Jian, Che Tao, Zhang Lixin, *et al.* Introduction on the Experiment of Cold Region Hydrological Remote Sensing and Ground-based Synchronous Observation in the Upstream of Heihe Basin[J]. *Journal of Glaciology and Geocryology*, 2009, 31(2):189-197. [王建,车涛,张立新,等.黑河流域上游寒区水文遥感—地面同步观测试验介绍[J].冰川冻土,2009,31(2):189-197.]
- [3] Ma Mingguo, Liu Qiang, Yan Guangjian, *et al.* Simultaneous Remote Sensing and Ground-based Experiment in the Heihe River Basin; Experiment of Forest Hydrology and Arid Region Hydrology in the Middle Reaches [J]. *Advances in Earth Science*, 2009, 24(7):681-695. [马明国,刘强,阎广建,等.黑河流域遥感—地面观测同步试验:森林水文和中游干旱区水文试验[J].地球科学进展,2009,24(7):681-695.]
- [4] Zhang Zhihui, Wang Weizhen, Ma Mingguo, *et al.* The Processing Methods of Eddy Covariance Fluxes Data and Products in WATER [J]. *Remote Sensing Technology and Application*, 2010, 25(6):xxx-xxx. [张智慧,王维真,马明国,等. WATER 试验涡动相关通量数据处理方法及产品生成[J].遥感技术与应用,2010,25(6):xxx-xxx.]
- [5] Huang Guanghui, Ma Mingguo, Tan Junlei, *et al.* Data Quality Control Procedures and Results of Automatic Weather Stations from Watershed Allied Telemetry Experimental Research (WATER) [J]. *Remote Sensing Technology and Application*, 2010, 25(6):xxx-xxx. [黄广辉,马明国,谭俊磊,等. WATER 试验自动气象站数据质量控制与产品生成[J].遥感技术与应用,2010,25(6):xxx-xxx.]
- [6] Parsons M A, Brodzik M J, Rutter N J. Data Management for the Cold Land Processes Experiment: Improving Hydrological Science [J]. *Hydrological Processes*, 2004, 18, 3637-3653.
- [7] Göckede M, Clement R, Moncrieff J B. Quality Control of CarboEurope Flux Data-Part 1: Coupling Footprint Analyses with Flux Data Quality Assessment to Evaluate Sites in Forest Ecosystems [J]. *BioGeoSciences*, 2008, 5, 433-450.

Quality Control and Evaluation for the Ground Observation Data of the Watershed Allied Telemetry Experimental Research

MA Ming-guo

(Cold and Arid Regions Environmental and Engineering Research Institute, CAS, Lanzhou 730000, China)

Abstract: In this paper, the data quality control and management processes were introduced for the ground observation data of the Watershed Allied Telemetry Experimental Research (WATER), except for the eddy covariance fluxes data, meteorological and hydrological data. The main contents include the establishment and implement observation specifications, data standardization processing, data standardization of the data file names, data quality evaluation. There are two objectives of this paper: to provide detailed data quality evaluation information for the data users; and to give some enlightenment on the data quality control and management for the later similar experiments. The data evaluation indicates that the data quality of WATER ground observation data is reliable, which is mainly supported by these data quality control works. The further work is needed to focus on the data analysis and application researches by using these data sets.

Key words: WATER; Watershed Allied Telemetry Experimental Research; Heihe River Basin; Ground observation data; Data quality control