

全国农业气象资料数据模式的研制

郭发辉 宋超辉

(国家气象中心,北京 100081)

摘要 在分析我国农业气象观测资料现状与特点的基础上,提出了制定“全国农业气象资料数据模式”必须遵循的基本原则,即以 1993 年版的《农业气象观测规范》为依据,以现行“农业气象观测记录年报表”的内容为基础,符合气象行业标准编写要求,具有延续性与可扩展性,面向现代化数据管理和应用服务等,对数据模式的结构和主要技术特点也进行了比较详细的论述。数据模式的研制将有助于现存的农业气象报表资料的数字化以及今后农业气象观测数据的采集、传输和处理自动化的实现。

关键词 农业气象资料 数据模式 层结构

引言

农业气象观测资料是我国气象监测网所收集的宝贵气象资料之一。它包括农作物、土壤、畜牧业、果树、林木、养殖、渔业、自然物候、自然灾害等方面的记录。由于诸多原因,农业气象观测资料至今仍只以纸质形式存档,不便于计算机快速处理、数据共享和应用,同时也致使农业气象资料中反映生态环境的记录未能充分开发利用,影响了农业气象和气候变化的研究。要改变目前这一落后状况的唯一途径是尽快地对已有的人工观测资料进行数字化,这就要求研制一个科学、合理的“全国农业气象资料数据模式”(以下简称模式)。

1 农业气象资料的现状和特点

我国国家气象档案馆馆藏着各个时期的农业气象资料,它可以代表我国不同时期农业气象资料的基本情况。因此,本文以国家气象档案馆的档案资料为依据,对我国的农业气象资料主要状况和特点进行分析。

1.1 农业气象资料种类

国家气象档案馆存放的农业气象资料种类按不同时期分为以下 4 种^[1]:

(1) 建国前的自然物候资料,以“农业气象物候调查表”的形式保存,约 880 站年,最早年份是 1934 年。含草本植物如小麦、棉花等 27 种作物的播种、移植、开花、收获的物候期;木本植物如桃、李、杏、梅等 24 种植物的发芽、叶盛、始花、盛花、果实成熟、落叶等物候期;动物如燕、雁、布谷鸟等 11 种动物的始见(或始鸣)、绝见物候期。

(2) 1950~1964 年的农业气象资料,以农业气象报表的形式保存,包括农作物物候和土壤湿度记录,如农作物物候观测记录(农气表-1)、土壤湿度资料(农气表-2-4)。

(3) 1965~1980 年无农业气象资料。

(4) 1981 年以后的农业气象资料,以每年农业气象年报表的形式保存国家级农业气象观测站观测的农作物生育状况、土壤水分、自然物候、畜牧气象的观测记录,如“作物生育状况观测记录年报表(农气表-1)”、“土壤水分(烘干称重法)观测记录年报表(农气表-2-1)”、“土壤水分(中子仪法)观测记录年报表(农气表-2-2)”、“自然物候观测记录年报表(农气表-3)”及“畜牧气象观测记录年报表(农气表-4)”等。

1.2 资料状况

(1) 1965 年及以前(含建国前)的农业气象资料

目前尚未系统清理,资料杂乱,无登记造册,这反映了该时期的资料几乎未得到利用的事实。

(2)1965~1980年无农业气象资料。

(3)1980~1993年农业气象资料质量比以前好,但台站观测资料报送不齐,站址变化较多,同一地点观测资料的连续性较差,而且,在观测记录内容上,这一时期与1993年6月以后不统一。

(4)1993年6月以后执行现行的《农业气象观测规范》,观测资料完整,记录格式统一。

1.3 主要特点

(1)农业气象观测内容多,包含作物、土壤、自然物候、畜牧、果树、林木、蔬菜、养殖渔业、农田小气候9大类观测内容。由于许多观测项目属于人工定性观测,其观测内容描述的复杂程度高于以器测为主的高空、地面气象观测资料,仅现行的《农业气象观测规范》的文字就有420多页,相当于现行《地面气象观测规范》(186页)的两倍多。

(2)农业气象观测标准不统一,记录形式杂乱。《农业气象观测规范》^[2]1993年6月才颁布执行,以前使用的是《农业气象观测方法》。不同时期记录格式不统一增加了对农业气象资料的数据处理、统计等业务的难度。

(3)农业气象资料报表种类、格式多。9大类农业气象资料中已形成固定报表报送的有农作物、土壤、自然物候、畜牧等5种。此外,还有果树、林木、蔬菜、养殖渔业、农田小气候没有固定、统一格式和报送制度。

(4)农业气象资料项目、子项目多。每种农业气象报表均含有各不相同的多种项目、子项目,例如“作物生育状况观测记录年报表(农气表-1)”除含有近40种不同观测作物外,每一种观测作物又含有若干个各不相同的项目(如发育期项目、生长状况项目等)和子项目(如发育期项目下又分始期子项目、普遍期子项目等)。

(5)农业气象资料汉字信息与字符信息并存。每种资料中汉字信息和字符(包括字母、数字、符号等)在报表中交叉出现。

2 模式设计的基本原则

模式制定是一项重要的基础性工作,它要全面、完整、正确地表达农业气象观测在各个时期各类不同观测实际,又要便于计算机信息化处理。它既涉

及到全国农业气象资料的自动化收集、传输、处理、存储、检索和服务应用,又涉及到不同时期资料的连续使用。

(1)以现行规范为依据。模式应以1993年版的《农业气象观测规范》为依据,其它时期的农业气象资料的内涵都按此规范进行整理,以保证模式规定的唯一性。

(2)以现行报表为基础。模式以现行“农业气象观测记录年报表”的内容为基础,各项目数字化资料排列的顺序基本上按现行农业气象报表中的排列顺序进行。这一方面是考虑历史资料向现行的农业气象观测记录年报表格式靠拢;另一方面是考虑现行“农业气象观测记录年报表”不仅全国统一且其内容也是最全面的。1993年以前的农业气象资料可按现行的报表格式相应进行整理。

(3)文件结构标准化,适应计算机处理的技术要求。模式的设计是按照中华人民共和国气象行业标准编写要求进行的,既适应农业气象观测资料的数据特点,又与全国其它气象资料的数据模式基本一致,其文件结构按文件首部、文件体、文件尾部三部分组织。数据文件以站年为单位,这样更便于标准化管理和使用。

(4)文件设置面向数据管理,有利于业务与应用服务。在文件名和文件结构、内容的设置上要面向数据管理,既便于资料收集、传输、存储,又便于检索、数据处理和信息应用、共享。

(5)有延续性与可扩展性,适应今后的业务发展。模式要具有延续性与可扩展性,除了要适应现行的观测内容与业务流程以外,还要考虑业务的发展和观测内容的增加,如今后果树、林木、蔬菜、养殖渔业、农田小气候等农业气象资料可能会统一报表格式与报送。

3 模式的基本结构与主要内容

按模式规定所形成的数据文件被命名为N格式文件,简称N文件。N文件由文件首部、文件体、文件尾部三部分构成。

3.1 文件首部

由反映观测台站观测条件、环境等基本参数组成。首部有11个字段,依次为区站号、经度、纬度、海拔高度、年份、省市自治区名称、台站名称、地址、寄出日期、质量检查状况码、经纬度和海拔高度参数。

3.2 文件体

在文件首部和尾部之间为文件体,由排列顺序固定的年报表所有数据、文字组成。文件体的现行框架由作物生育状况、土壤水分(烘干称重法)、土壤水分(中子仪法)、自然物候、畜牧气象五块数据,即五种农业气象观测记录年报表数据构成,具体的数据组织原则是:

(1) 严格按照五种农业气象观测记录年报表数据顺序分块排列。每一块数据均为独立的数据实体,块与块之间无任何联系,若某块数据全年无观测,则该数据块作缺省处理。一个综合文件的文件体至少有 5 个数据块,一个单种文件的文件体至少有 1 个数据块。块标识符、数据块内容及相应农气报表如表 1。

(2) 每一数据块根据本块特点分为若干个数据段,段以下可分为若干个记录,一个记录可由若干个字段组成,一个字段可由若干个字符或汉字组成。

(3) 每一数据块的第一个数据段为无段落标识符数据段,含其年报表封面的部分项目,由一个记录若干个字段组成,其字段数与字段长度和类型根据年报表封面的内容而定,例如,Cy 块即作物生育状况年报表(农气表-1)的无段落标识符数据段的一个记录由 9 个字段组成,字段排列顺序为“作物名称”、“品种名称”、“品种类型、熟性、栽培方式”、“台站长”、“抄录员”、“观测员”、“校对员”、“预审员”、“审核员”。每个字段长依次为 4C(指 4 个汉字,下同)、10C、15C、12C、12C、12C、12C、12C、12C。而 Ry 块即畜牧气象观测年报表(农气表-4)的无段落标识符数据的一个记录只有“台站长”、“抄录员”、“观测员”、“校对员”、“预审员”、“审核员”6 个字段,字段长均为 12C。

表 1 各数据块标识符、数据内容及相应的农业气象报表

块标识符	内 容	相应报表
Cy	作物生育状况年报数据	农气表-1
Sy	作物土壤年报数据(烘干)	农气表-2-1
Zy	土壤年报数据(中子仪)	农气表-2-2
Ny	自然物候年报数据	农气表-3
Ry	畜牧年报数据	农气表-4

(4) 每一数据块的第二个数据段开始为有段落标识符数据段,为年报表封面以外的观测内容,一个段标识符后为某一本报表中一个相关表格中的所有数据信息,不同数据段标识符可以区别同一块数据,

即同一本报表中不同相关表格的数据信息。不同种类报表所含有段落标识符数据段的个数是不同的。如 Cy 块即作物生育状况年报表(农气表-1)包含 10 个有段落标识符数据段,其标识符及内容如表 2。

而 Sy 块“土壤水分(烘干称重法)观测记录年报表(农气表-2-1)”则包含 13 个有段落标识符数据段,它的标识符及内容如表 3。

表 2 各数据段的段标识符和数据内容

段标识符	内 容
C01x	发育期、生长状况、生长高度、密度、产量因素、产量结构
C02x	主要田间工作
C03x	杂项类
C04x	灾害
C05x	灾害调查
C06x	地段说明
C07x	纪要
C08x	大田生育状况调查
C09x	生长量测定
C10x	农业气象条件鉴定

表 3 Sy 块数据段标识符及内容

段标识符	内 容
S01x	观测日期
S02x	重量含水率
S03x	相对湿度
S04x	总储存量
S05x	有效水分储存量
S06x	地下水位深度与干土层厚度
S07x	作物发育期
S08x	渗透深度
S09x	降水与灌溉
S10x	水文物理及冻结解冻
S11x	水分变化评述
S12x	观测地段说明
S13x	纪要

3.3 文件尾部

反映文件结束的标志为文件尾部,是文件的最后一个记录,由 6 个问号“?????”组成,作为一个站年数据的年结束标志。

4 模式的主要技术特点

4.1 文件名采用综合文件名和单种文件名

每个数据文件的文件名第一个字母均为“N”。综合文件名为“NIIiiiYYYY.dat”,由 14 个字母、数字组成;单种文件名为“NXIiiiYYYY.dat”,由 15 个字母、数字组成。其中:“N”为文件类别(即农业气象数字化资料)的标识符;“X”为种类区别符(X=1,2,3,4 分别为农气表-1、农气表-2、农气表-3、农气

表 4) ;“IIiii”为区站号 ;“YYYY”为资料年份 ;“.”为分隔符 ;“dat”为后缀。这样的文件名虽字母、数字位数较多,但它可以明确表示该文件的内容。

4.2 文件体采用层结构设计

文件体采用了块-数据段-记录-字段-字符(或汉字)的层结构(图 1),其块结束符为“????? <CR)” ,段结束符为“ = <CR)” ,记录结束符为“ <CR)” ,字段结束符(又为字段与字段间的间隔符)为“|”。这样的层结构可清晰反映农业气象观测资料特点。

4.3 采用块标识符区分不同报表的资料

文件体中的第一层是“块”,一个块标识符记录后紧跟着某一种类的资料,即一本年报表的数据信息。一块信息为一个独立的数据实体。不同块标识符不仅可以区别不同类型、种类农气报表的数据信息,也不影响资料统一格式和资料补充和追加。

块标识符由两位组成,其中第一位为类别符,第二位为参数符。例如块标识符为“Cy”,第一位“C”,表示“作物生育状况观测记录年报表(农气表-1)”,

第二位为顺序号参数“y”,表示作物顺序号。当块标识符为 Cy 和 Sy 时,y = A,B,C, … …,Z 表示作物顺序号,当块标识符为 Zy、Ny 和 Ry,且均有记录时,y 为“0”。

4.4 用段标识符区分同一报表中不同的数据内容

文件体中的第二层是数据段,一个段标识符后为某一本报表中一个相关表格中的所有数据信息,不同数据段标识符可以区别同一块数据即同一本报表中不同相关表格的数据信息。

数据段标识符由 4 位字符组成,其中第一位为块标识符的第一个字母,第二至三位为该段在该农气报表中的位置顺序号,第四位为参数符“x”,按相关的说明取值。例如:数据段标识符“C08x”中,“C”表示为农气表-1 ;“08”表示该数据段为农气表-1 中的第 8 个数据段,即“大田生育状况观测调查”栏内的所有数据 ;“x”为参数,当 x 为“1”时,表示只有上半段数据 ;当 x 为“0”时表示上、下半段数据均有。

第一层(块) | 第二层(数据段) | 第三层(记录) | 第四层(字段) | 第五层(汉字或字符)

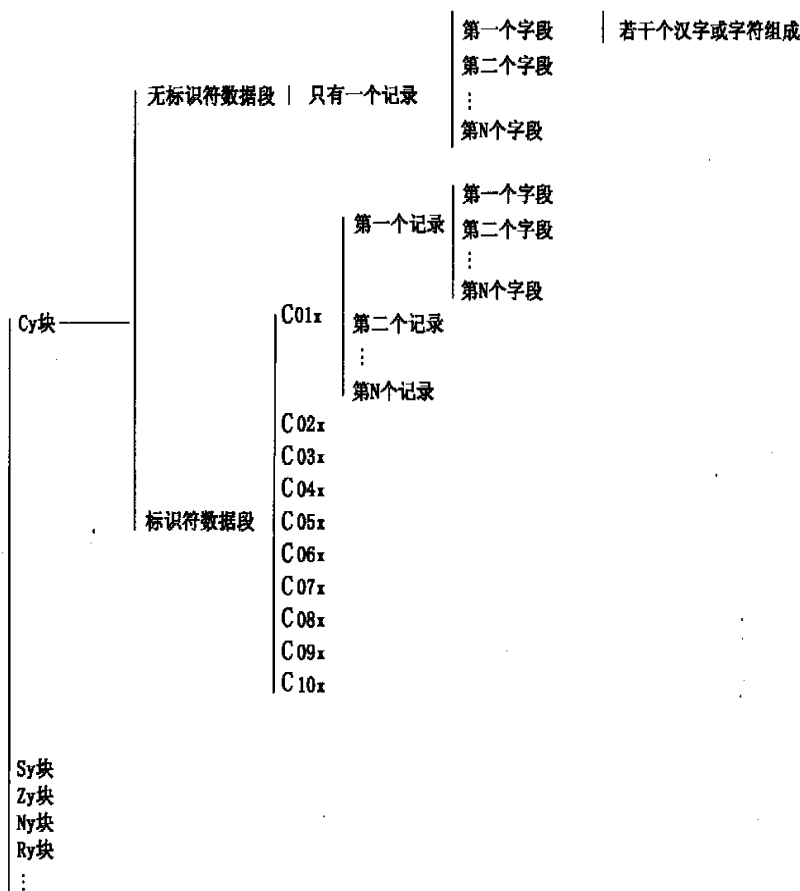


图 1 文件体层结构示意图

4.5 不同数据采用不同的描述方法

为适应农业气象观测资料既有汉字信息又有字符且其长度不同的特征,在数据描述上采用了与其它数据模式不同的描述方法。主要有以下区别:

(1) 采用了字段名的描述方法,即各个数据均有相对应的字段名,字段结束符为“|”;

(2) 数据类型和长度采用文字描述方式,如 3C 表示 3 个汉字,5T 表示 5 个字符,这与其它数据模式直接采用若干个“×”的描述方法是不同的;

(3) 在字段名下采用了 3 个下标,用它表示该字段所在的位置、长度和类型。例如字段名“C01_{01,01,8c}”,其中,“C01”为数据段标号,第 1 下标为该字段所在记录的记录顺序号,第 2 下标为该字段在该记录中的字段顺序号,第 3 下标为本字段的字段长及类型。例如,“C01_{01,01,8c}”为作物生育状况年报数据块的第一个数据段第一个记录的第一个字段,字段长及类型为 8 个汉字。

4.6 采用面向数据管理的文件设置

为便于收集、传输、处理、存储、检索和信息共享,本数据模式最终将形成两类文件,即综合文件和单种文件。在资料收集、传输、存储时可使用综合文件,在检索、处理和信息应用、共享时使用单种文件会更方便些。

4.7 数据模式要有相配套的计算机软件支持

由于农业气象观测内容多、规定复杂,所以模式的格式细节很难为数字化人员掌握。在资料的数字化过程中,录入、校对、修改等操作均面向固定的农

气报表相关格式而不必熟悉该数据格式与排列顺序,这样可极大地减少数字化中过失误差的机率。在资料的审核和形成数字化资料时,须依赖一定的计算机软件支持。与模式相配套的计算机程序有输入程序、恢复、输出打印报表格式程序、错情修改程序和转换模式文件程序。

5 结语

(1) 虽经过深入调研,广泛征求各方面技术人员、专家的意见,多次对模式进行修改,但由于农业气象观测资料规定复杂,涉及面广,可能还存在遗漏之处。

(2) 为适应我国农业气象资料数据采集的需要,并对历史的农业气象资料进行数字化,模式还需要经过业务试验,在实践中不断加以完善。

(3) 全国农业气象资料数字化及应用工作是一个较大的系统工程,涉及建国至 1980 年和建国以前农业气象资料的整理、质量控制、资料分类合并、检索软件的研制、农业气象资料数据库设计、资料整编及服务系统的研制。此外,其它渠道如卫星遥感的有关农业气象数据、图表和农业气象试验站试验资料的收集处理也有大量工作要做。

参考文献

- 1 中国气象局气象档案馆. 中国气象局气象档案馆指南. 北京:气象出版社,2003
- 2 国家气象局. 农业气象观测规范. 北京:气象出版社,1993

Development of Digitizing Mode of National Agrometeorological Data

Guo Fahui Song Chaohui

(National Meteorological Center, Beijing 100081)

Abstract: Based on the analysis of the current situation and features of agrometeorological data, the data mode of agricultural meteorological data for whole China has been developed. The basic principles for developing data mode are: being in accordance with “Guidelines for Agrometeorological Observation” published in 1993, based on the existing annual report forms of agrometeorological observation records, being in accordance with the standards of the meteorological trade, being expandable and persistent, being suitable to the modernization of data management and service. Its configuration, content and technological features are also described. It will help to realize the digitization of the existing report forms and data and the automatic collection, transmission and processing of agrometeorological data.

Key words: agrometeorological, observations data, data mode, layer structure