

彭江华, 黄祖辉. 降水自记纸彩色图形数字化的技术处理[J]. 气象, 2011, 37(2): 249-253.

降水自记纸彩色图形数字化的技术处理

彭江华¹ 黄祖辉²

1 江西省气象信息中心, 南昌 330046

2 江西省南昌市气象局, 南昌 330046

提 要: 利用降水自记纸彩色图形数字化处理技术, 实现降水自记纸图形数字化是对气象历史档案拯救技术的有效探索。降水自记纸彩色图形数字化处理是采用图形扫描和数据处理技术, 对降水自记纸进行图形扫描, 利用降水自记纸数字化处理系统软件, 提取降水曲线, 将降水自记纸转换成标准分钟降水数字化资料, 获取一系列满足降水时空精细化分析要求的降水自记纸图像数据集、降水曲线数据集、以及分钟、小时尺度的降水数据集。最终形成分钟和小时降水强度的标准文件, 实现对降水自记信息的完整、真实、安全的保存。它是改变目前状态、使其信息的内在含义得到充分利用、价值得以充分体现的唯一途径。降水自记纸彩色图形数字化技术处理, 从 5 个环节(自记纸预处理、扫描、扫描检查、降水曲线提取、降水强度数据转换)入手, 结合实际工作经验, 详细分析降水自记纸数字化处理过程中的技术问题及解决方法。

关键词: 降水自记纸, 数字化, 彩色图形

The Technical Processing of Color Graph Digitizing of Precipitation Recording Papers

PENG Jianghua¹ HUANG Zuhui²

1 Meteorological Information Center of Jiangxi, Nanchang 330046

2 Nanchang Meteorological Bureau of Jiangxi, Nanchang 330046

Abstract: The use of the color graph digitizing processing technology of the precipitation recording paper, is an effective exploration for rescuing the archives of climate history. The technology applying graph scanning and data processing, can scan pictures of precipitation recording paper, and can extract dropdown curve and convert precipitation recording papers into digitizing data of precipitation per minute, by the use of digitizing processing system of the precipitation recording paper. Then we can achieve a series of data set, including precipitation graph and precipitation curve, which based on minute-index or hour-index. Finally we can get the standard file of precipitation intensity per minute or hour, so as to achieve the goals of saving data of precipitation autographic records wholly, actually and safely. The technology is the only way to change the current situation, and it makes full use of the implications of the information and reflects the value fully. By the use of the color graph digitizing technologic processing of the precipitation recording paper, and in the light of practical work experiences, we can analyze in detail the technologic problems during the course of digitizing processing and find their solutions.

Key words: precipitation recording paper, digitization, color graph

引 言

降水自记纸是详细记录气象台(站)降水特征的气象资料之一。各气象台(站)积累的降水自记纸,

由于年代长,保管条件差,已经出现不同程度的纸张变质、破损、字迹变淡(模糊)等现象,急需采取措施抢救。降水自记纸彩色图形数字化处理是前期手工进行降水自记纸预处理,对降水自记纸进行彩色图形扫描,建立完整的扫描图像文件,把扫描图像文件,

采用人机交互方式,自动进行曲线跟踪,完成降水曲线的提取,形成降水曲线数据文件,将降水曲线数据转换成降水分钟强度数据,进行数据质量控制,最后形成分钟和小时降水强度标准文件。它采用先进的图形扫描和数据处理技术,实现对降水自记记录信息完整、真实、安全的保存。

1 降水自记纸彩色图形数字化处理系统

降水自记纸彩色图形数字化处理系统是对降水自记纸进行扫描输入、自动跟踪提取降水曲线、形成降水强度标准数据的降水自记纸图形数字化计算机处理系统(包括手工进行的降水自记纸预处理)。处理系统完成将降水曲线数据转换成降水分钟强度数据,并对数据进行质量检查和控制,最终形成标准分钟数据文件;完成将标准分钟强度数据向标准小时强度数据的转换。系统在对数据处理中采用了虹吸订正、累计进位、数据滑动平均等一系列数据处理技术以及对降水曲线数据中的各种异常迹线数据的特殊处理方法,保证了标准数据的真实性和准确性。降水分钟、小时强度数据转换和质量控制系统是降水自记纸数字化处理系统的重要组成部分。系统集成降水自记纸扫描曲线文件向降水分钟强度数据的转换,小时强度数据统计数据分类查询、检索、数据质量控制、数据错情可视化定位显示、数据文件合并等功能^[1]。降水自记纸彩色图形数字化处理系统操作简单、自动化程度高,形成清晰、完整的降水自记纸彩色扫描图像文件,及高精度的降水分钟、小时强度数据标准文件。

2 降水自记纸彩色图形数字化的预处理

降水自记纸预处理是降水自记纸图形数字化处理的一个重要环节,是数字化处理的基础。在降水自记纸进行扫描输入和数字化处理之前,必须对降水自记纸进行一系列技术处理工作,为降水自记纸的扫描和图形数字化做好准备。降水自记纸的预处理包括降水自记纸整理和图面处理两部分。

2.1 降水自记纸整理

为了扫描方便要将归档入库并已装订成册的降

水自记纸拆开成单页,若有 2 种以上降水自记仪器的,先确定一种作为信息化,其他的只作为仪器故障时记录的代用。在拆分时按时间顺序整理,检查是否完整,将降水自记纸中需扫描或不需扫描(如无或很小降水,一般指日降水量 $<0.1\text{ mm}$)的自记纸分开,同时注意日期的连续性。若发现日期不连续或日期跨度与曲线数不符时,应立即查明原因纠正处理。如果丢失,则应在降水自记预处理登记表中登记。需扫描的降水自记纸要标注区站号、日期、时间;不扫描降水自记纸登记时,常出现连续多日或多张自记纸不需扫描,为减少无降水文件数目,只需登记起、止时间。遇停用或缺(漏)测(指无自记纸)情况,则须在降水自记预处理登记表中登记。

自记纸换上和换下时间关系到计算降水量的时间段,影响降水曲线提取的准确度。按照江西降水自记纸基本上是 13:00 换纸的特点,一是将没有标识换上、换下时间的前后两张纸做对比分析,前一张纸是 13:00 换下,后一张纸是在 13:00 或 13:00 以后换上;换下时间一般根据后一张换上时间判断,后一张若是 13:10 换上,前一张应该在 13:00—13:10 换下;二是根据曲线的走势,按自记纸格数计算时间;三是根据无标识前后两张自记纸的换上和换下时间、曲线的起始位置综合判断。

2.2 降水自记纸图面处理

需扫描输入的自记纸,要逐张进行图面清洁、参数标记和曲线异常处理。降水自记纸除了降水曲线以外,还有许多书写笔迹或污迹,应尽量予以清除,使图面清洁、清晰。破损、残缺的自记纸,将其拼接修复。降水自记纸上有两根曲线时,要将降水曲线的起、止时间圈出,以便确认。扫描图像必须包含自记纸上有效信息。台站使用最多的 21 号降水自记纸,长度超过了 EPSON /GT/10000+扫描仪的最大扫描长度,常造成自记纸前部的台站名称、日期或降水曲线的尾部信息无法被完全扫描。为此,统一将台站名称、日期用黑笔重新标记在降水曲线起始点附近,扫描时只需注意降水曲线的尾部信息能完全被扫描即可。

2.3 降水曲线异常处理

曲线有短暂中断或不清晰,可用铅笔稍作修描,使起、止点明确,曲线清晰完整。曲线修描线条粗、细要适中,过细易出现断点,影响曲线提取,过粗影

响精度。降水曲线中断无法修描,可用红笔标明各个中断点,采用时间内插法算出每两个相连点的降水量,标在相应的中断处;曲线未中断,但不能真实反映降水情况的,应根据降水自记纸背面说明,用红笔把相应的降水曲线段圈出来,并标上降水量;遇到降水自记有虹吸订正的,用红笔在自记纸左边填明虹吸订正总量^[2]。遇因换纸时正在下雨推迟换自记纸的情况,要注意降水自记纸尾部与头部的曲线连续性,不连续的标明降水量。如果一张降水自记纸中,两天或两天以上都有降水曲线、或者降水迹线不是最后一条(天)时,为了保证转换程序正确识别有降水曲线的日期,应把日期拆开。有两条降水曲线或一条降水曲线夹在无降水迹线中间的现象。由于自记降水数字化处理程序默认的是一张自记纸有一条降水曲线,而且最后一条迹线是降水曲线,若把有两条降水曲线或一条降水曲线夹在无降水曲线中间的自记纸只扫描一次,就会出现降水天数与实际降水日数不符或不连续。对有两条降水曲线的自记纸做两次扫描,分别命名两条降水曲线的时间;经过扫描和重命名,分别形成两个图像文件和两个数据文件;对一条降水曲线夹在无降水曲线中间的自记纸,更改重命名日期,在无降水不扫描文件中加入降水迹线后几日的时间即可。例 1:有一张自记纸有 15—19 日共 5 条降水曲线,其中,16 日、19 日是有降水的,日期应分别拆为:××××年×月 15—16 日、××××年×月 16—19 日;例 2:有一张自记纸有 20—25 日共 6 条降水曲线,其中 23 日有降水,日期应记为××××年×月 20—23 日,23 日接着为后一张自记纸的起始日期;两张相连有降水曲线,当前一张换下时间和后一张换上时间有矛盾或者不明时,时间以“?”不明处理。

中断无标识的处理:曲线中断都应有标识,当曲线中断无订正值时,首先要判断曲线中断的地方有无降水,若有降水,采用气表-1 中的定时观测降水量代替;如果出现有降水曲线部分中断,要用定时观测的降水量减去无中断部分迹线降水量的差值做异常处理的输入值^[3]。

对无法辨析自记纸的处理:20 世纪 60 年代有些强降水,墨水覆盖了自己记纸曲线,夹杂时间订正的草算式和不规范的勾画,无法判断迹线的走向,也没有降水量的订正值,则以定时观测值代替。

降水自记仪器故障或自记记录出现异常情况的,台站会要在自记纸的背面备注。备注内容要写

在正面的合适位置,站名和日期写全,换纸时间必须在曲线起、止旁标记清楚,以减少信息化时的再处理,保持其原始性。

3 降水自记纸图形数字化扫描处理

降水自记纸图形扫描是对有降水的自记纸逐张扫描。建立台站信息表、台站降水状态表,对降水自记纸扫描质量检查、文件重命名、不扫描文件及缺自记纸文件的生成、文件映像检查。降水自记纸由纸质转换为电脑可识别、存储的图形数字化文件。降水自记纸图形扫描流程:

(1) 扫描效果试验。为达到理想的扫描效果,在进行图像扫描之前,必须进行扫描试验,以便确定最佳的彩色扫描参数(对比度、亮度、灰度等)和图形数据的压缩率,保证扫描图像清晰,曲线与底色差别明显,图像失真度小,压缩率高的要求^[3]。

(2) 扫描图像质量检查及重命名。对每个图像文件进行扫描质量、图面信息的完整性检查,将文件名更改成规定格式,便于检索。主要内容有:建立台站降水状态表,用于对自记纸降水情况和数字化处理结果的对比,保证降水自记纸数字化处理的质量,包括区站号、站名、日期、时间和逐日降水量信息;扫描图像质量检查,提供图像显示、水平、垂直校准线、多比例尺图像缩放功能;降水自记纸区站号、日期的输入,自动完成图像文件命名。

降水自记纸扫描时,遇到降水起止时间不清时,将有降水曲线起止时间突出标注出来,对时间准确定位,时间的分辨率只能用前 1 张或后 1 张降水自记纸的分辨率,起止时间输入“—”或“?”即可。

(3) 不扫描文件的生成及全年文件完整性检查。按登记表中登记的不扫描日期建立 1 KB 文件(系统自动生成),然后通过对文件名的检索进行完整性检查。一般只需检查无自记纸的日期与登记表是否一致,文件名、日期是否重叠。

(4) 建立备注文件。按年代顺序将预处理登记表中停用、缺测、自记纸丢失、用定时降水量减代及用累积量代替等内容分类输入备注文件中。

(5) 扫描参数的设置。主要通过 Gamma 参数的调整来提高扫描效果。纸质好、图像清晰的图像文件大小在 220 KB 左右;纸质差、图像模糊的图像文件大小在 250 KB 左右。预览后,先取定一个扫描框,设置好参数再复制另外一个扫描框,以保证

两个扫描框参数一致。

(6) 扫描区域的固定和换纸。采用挡条固定自记纸的安放位置,用透明塑料板按压自记纸,保证把自记纸有价值部分完全置于扫描框内。在扫描过程中,由于降水自记纸存在着大小不一,自记纸头部和尾部边线不易对齐,达不到扫描要求,要反复几次扫描才能达到要求,影响了扫描处理速度。处理方法是把两张自记纸分别放在第一框、第二框扫描区域内固定好,当第一框、第二框扫描完后,先检查第一框自记纸头部边线与尾部边线是否对齐;若对齐,表明扫描达到要求;若不对齐,采取以下方法:把第一框自记纸尾部固定紧靠档条不变,自记纸头部往档条外面方向稍移动一点,接着重扫第一框自记纸,再检查该自记纸头部、尾部边线是否趋于对齐。如果它们是趋于对齐,以后出现这种情况时就根据头部、尾部不对齐程度来确定头部外移程度;把自记纸头部固定紧靠档条不变,自记纸尾部往档条外面方向稍移动一点即可。

同样道理,检查第二框自记纸时:把自记纸头部固定紧靠档条不变,自记纸尾部往档条外面方向稍移动一点;把自记纸尾部固定紧靠档条不变,自记纸头部往档条外面方向稍移动一点。

注意每次放置纸时要根据以上出现的情况来处理,一张自记纸扫描一般不会超过两次。放(换)纸按降水自记纸时间顺序进行,两张自记纸的扫描分两次进行的,放纸也可分别在一张自记纸扫描完毕后进行。

扫描完成后必须作一次质量检查,目测屏幕显示图像的清晰度,检查扫描图像的偏斜程度,首尾是否完整。

4 降水自记纸图形数字化降水曲线的跟踪提取

4.1 降水曲线跟踪提取

降水曲线跟踪提取是把降水曲线从自记纸彩色扫描图像中分辨出来,以纵坐标(降水量)和横坐标(时间)方式表征。其自动识别和人机交互的能力是决定降水自记纸图形数字化处理精度和影响降水自记纸图形数字化处理工作量的关键因素。降水曲线跟踪提取采用人机交互方式,自动进行曲线跟踪,完成降水曲线的提取,实现由降水自记纸彩色扫描图像

文件到降水曲线数据文件的转换^[4]。具体内容包括:

(1) 设定降水、时间分辨率和跟踪方法,输入降水曲线的起止时间,进行降水曲线的自动跟踪识别,在降水自记纸彩色扫描图像上重叠显示自动跟踪曲线。

(2) 不能一次性完成自动跟踪时,采用多种曲线跟踪的交互修改。降水曲线密集的超强降水的跟踪,采用虹吸线端点连接,虹吸点和归零时间点作为识别后的线点,中间不再进行插点。当有一段跟踪效果不是很好时,可以从某一位置重新开始补跟踪,对跟踪曲线的点、线修改。

(3) 为保证曲线提取质量,当发现图像扫描时放置不正,出现较明显的歪斜,要进行倾斜旋转校正,虹吸下降线不正(水平歪斜),可认为是钟轴倾斜,进行水平(时间)订正^[5-6]。

4.2 降水曲线跟踪提取特殊情况的技术处理

降水曲线在跟踪提取时,经常遇到自记纸在换纸期间没有避开中等强度的降水,使自记纸记录不能真实地反映出实际降水量。例如:在 13:00 换纸时有中等强度的降水,自记纸在 13:05 才开始有记录,即 13:00—13:05 的记录缺测。在处理时,可根据该自记纸备注说明或地面气象记录月报表中定时观测计算出该时段的降水量(假设为 0.3 mm),可在时间坐标为 13:00、降水量坐标为 13:05,降水值减去 0.3 mm 处加 1 个点作为这条迹线的起点即可。

降水曲线在某一段出现异常且中间有虹吸过程的降水曲线跟踪提取处理。例如:降水曲线在 10:30—11:20 异常,且中间有虹吸过程,备注说明 10:00—12:00 时降水量为 5.3 mm,可选择“特殊处理按钮”,用鼠标拉 1 条直线,一端作为异常起点,另一端为异常终点(这条直线与降水实际迹线相连),降水量值输“—1”,先按下 Ctrl 键,用鼠标在曲线的 10:00 位置(如有人工铅笔标记,应对准该标记)和 12:00 位置(同样对准人工标记)拉 1 条直线,然后输入该 10:00—12:00 的降水量,计算出实际异常值。实在不能计算出异常量的以缺测处理即降水量输入“1000”。这样既保留了有效的曲线,又省去了计算真正异常的雨量,提高了处理速度,减轻了工作强度^[7-9]。

5 降水自记纸图形数字化降水分钟强度数据转换和质量控制

降水分钟强度数据转换和质量控制是将降水曲

线数据转换成降水分钟强度数据,进行数据质量控制,形成降水强度标准数据文件。

5.1 降水强度数据转换和质量控制的流程

(1) 降水曲线数据向降水分钟强度数据的转换,以降水曲线为界限的时间-降水量坐标数据转换为以北京时间 20:00 为日界的每分钟降水量的数据,进行虹吸订正处理、累计进位和数据平滑技术处理。

(2) 降水分钟强度数据的质量检查,进行自记降水分钟强度累计日降水量与雨量筒观测日降水量的比较,有疑问输出相应的信息,恭请用户查证核实该数据的正确性,提供疑问数据文件。

(3) 降水分钟强度数据文件的纠错和合并,为修正降水分钟强度数据文件中的错误数据,提供新的降水分钟强度数据文件对原降水分钟强度数据文件中错误数据的纠错覆盖,具有时间连续检查、日期连续检查的两个降水分钟强度数据文件。

(4) 降水分钟强度数据文件检索,从降水分钟强度数据文件中检索所需的降水分钟强度数据文件^[10-11]。

5.2 数据转换和质量控制中特殊情况的技术处理

(1) 在输入台站信息时,台站号不能空缺,在输入台站名称后以“/////”代替^[4]。

(2) 降水曲线相邻日时间重叠,如某降水自记纸换下时间 12:55,下一日自记纸换上时间 12:45。先查找换下和换上时间,查明原因进行时间修正,把时间有误的自记纸重新跟踪一遍。

(3) 降水曲线日期不连续,把不连续日期的降水曲线文件查找出来,重新跟踪即可。

(4) 降水曲线有虹吸,虹吸前后时间坐标偏差大于 2 分钟,把该张降水自记纸进行效果回放,用鼠标光标检查每条跟踪虹吸线是否水平,把不水平的虹吸线用微调按钮进行微调,使之水平保存即可。

(5) 降水曲线有异常时,曲线异常数据的前后时间坐标不一致。把异常降水自记纸进行效果回放,选择曲线数据跟踪效果按钮,让光标停在曲线异常点,微调使曲线异常数据的前后时间坐标一致,保存即可。

(6) 降水曲线在分钟强度转换时,系统软件出现死机或中途不能转换又没有错误信息提示,打开已经转换的分钟强度降水文件,查找到不能转换的自记纸,查明原因重新跟踪,再转换即可。另一种是

平均分割定位法,遇到以上情况时,把转换的降水曲线按年限或日期平均分为两部分,分别进行转换。不能转换的部分,再平均分为两部分,分别进行转换。以此类推,就可以查找到不能转换的自记纸,查明原因重新跟踪,再转换即可^[1,12-14]。

6 结 语

降水自记纸彩色图形数字化处理,将降水自记纸转换成标准分钟降水数字化资料,获取了一系列满足降水时空精细化分析要求的降水自记纸图像数据集、降水曲线数据集、以及分钟、小时尺度的降水数据集。它对于研究其他气象纸质记录的数字化处理及江西省降水强度、暴雨特征研究具有重要意义,尤其对暴雨的观测事实进行 24 小时精细化雨强分析作用更大。

参考文献

- [1] 王伯民,吕勇平,张强.降水自记纸彩色扫描数字化处理系统[J].应用气象学报,2004,15(6):737-744.
- [2] 吴名杰.降水自记纸数据化处理常见问题及解决方法[J].气象研究与应用,2007,28(9):175-176.
- [3] 朱尽文,王志峻,汪青春.“降水自记纸数字化处理系统”简介及数字化处理时应注意的问题[J].青海气象,2006,28(2):87-88.
- [4] 中国气象局预测减灾司.降水自记纸数字化处理系统使用手册[M].北京:2002,46-62.
- [5] 陶士伟,徐枝芳.加密自动站资料质量保障体系分析[J].气象,2007,33(2):34-41.
- [6] 金燕.网络信息资源整合研究[J].现代情报,2007,27(7):40-43.
- [7] 马文彦,冯新,杨芙蓉.地面资料在侦测暴雨天气过程中的应用[J].气象,2010,36(1):41-48.
- [8] 郭锐,李泽椿,张国平.ATOVS 资料在淮河暴雨预报中的同化应用研究[J].气象,2010,36(2):1-12.
- [9] 刘旭林,赵文芳.气象观测数据等值线自动绘制系统[J].气象,2009,35(4):102-107.
- [10] 王海军,杨志.自动气象站实时资料自动质量控制方法及其应用[J].气象,2007,33(10):102-109.
- [11] 任芝花,熊安元.地面自动站观测资料三级质量控制业务系统的研制[J].气象,2007,33(1):19-24.
- [12] 杨贤为.气候应用专用数据库气象资料的质量检验[J].气象,1998,24(12):33-36.
- [13] 王伯民.基本气象资料质量控制综合判别法的研究[J].应用气象学报,2004,12,15(增刊):50-59.
- [14] 王海军,杨志彪,杨代才,等.自动气象站实时资料自动质量控制方法及其应用[J].气象,2007,33(10):20-23.