

彩色扫描图形数字化处理技术的研究 ——气象历史档案拯救技术探索之一*

王伯民

(国家气象中心,北京 100081)

摘 要

利用彩色扫描图形数字化处理技术实现降水自记纸图形的数字化处理是对气象历史档案拯救技术的有效探索。该文从降水自记纸扫描图形的数字化精度分析和彩色扫描的优势特征分析出发,重点研究分析了存储方式、降水曲线彩色表征区分技术和降水自记纸扫描图形数字化数据处理技术,用以为其它图形数字化处理提供一种技术思路和有有效的技术参考。
关键词:彩色扫描 图形数字化 降水自记纸

引 言

气象档案是气象事业发展过程的历史印记,是各地气象要素、天气气候状况及变化的客观记录。国家气象档案馆保存的气象档案最早始于19世纪中叶,至今已有100多年的历史。经过几代气象工作者的耕耘,气象档案馆存档的气象档案种类和载体形式越来越多,这些气象档案不仅是气象部门的财富,也是国家重要的科技档案信息资源。

由于经历历史沧桑和保管条件所限,国家气象档案馆保存的1949年前形成的气象档案已普遍出现纸质老化、变脆、破损以及字迹退化现象。1949年后的气象档案,已经对地面、高空、辐射等观测资料进行了信息化处理,但是,对于降水、风自记纸等非数字的图形资料依然缺乏有效的数字化处理手段,也存在着资料保管中的自然受损和数据的有效使用问题。中国气象局所属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、字迹变淡(模糊)等现象。这些气象档案是宝贵的历史遗产,具有极其重要的科学价值和历史价值,急需尽快地拯救、保护和开发利用。

在我国,降水、风、气压、气温自记纸已经积累有40多年,是数量十分庞大的图形资料档案。以降水自记纸为例,按全国地面基本站和基准站680站40年、平均每年130个雨日计算,约有降水自记纸3536000张。按全国气象台站2700站40年、平均每年130个雨日计算,约有降水自记纸14040000张。降水自记纸是客观记录降水情况的十分宝贵的气象实测资料之一,对于研究我国降水特征,尤其是暴雨特征具有重要意义,并且对农业、城市建设、电讯、交通、水利等国民经济部门,对防灾减灾策略的制定具有重要作用。由于降水自记纸的图形表现方式和纸张载体的限制,其信息无法得到广泛和深入的应用。因此,降水自记纸的数字化处理,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

* 2003-02-18收到,2003-04-17收到修改稿。

以充分体现的唯一途径。

对降水自记纸图形的数字化有三种方式:一是用数字化仪取降水曲线特征点;二是用扫描仪扫描降水自记纸图形,自动跟踪降水曲线图形;三是用照相机拍摄降水自记纸图形,自动跟踪降水曲线图形。用数字化仪处理方式的主要弱点是:人工取点主观因素影响大(误差大),工作量也大。用照相机拍摄降水自记纸图形处理方式的主要弱点是:由于自记纸长,拍摄图形有变形,影响精度。相对而言,用扫描仪扫描降水自记纸图形是一种比较合适的方式。

中国气象局一向十分重视降水自记纸的数字化工作。20世纪80年代,曾用数字化仪进行试验研究,由于主观误差大和工作量太大的缘故,无法推广全面开展。90年代初,曾用黑白滚筒扫描仪进行扫描和程序识别跟踪曲线的方式,进行降水自记纸的数字化工作。可是由于扫描仪的输入性能和操作不便,也由于黑白扫描方式降水自记纸坐标线和数字等其它痕迹对降水自记曲线的干扰太大,曲线跟踪的人工干预操作多,工作量大,而没能全面开展。

随着信息技术的不断发展和完善,存储技术和彩色扫描技术有了长足的发展。如何利用先进的技术,尤其是利用彩色扫描技术进行降水自记纸图形的数字化处理是我们面临的新的研究课题。

1 降水自记纸彩色扫描数字化处理技术整体设计面临的技术问题

从90年代初的降水自记纸黑白扫描数字化处理到降水自记纸彩色扫描数字化处理,从整体设计的角度,我们面临有三个必须首先解决的技术问题。

一是扫描输入要求,必须真实不变形、快速和操作方便。解决这个问题在于要有合适的彩色扫描仪。由于降水自记纸的长度比A3纸的长度还要长,一般的A3纸扫描仪无法适用。经过多次市场调研,采用EPSON/GT/10000+(爱普生)彩色扫描仪,基本满足要求,加上采取降水自记纸预处理的措施,解决了第一个问题。

二是降水自记纸彩色扫描图像的存储空间不能太大,必须在可以容忍的范围内,以解决降水自记纸图像归档的问题。

三是区分降水自记曲线与坐标线的有效技术方法,以及曲线自动跟踪技术。

降水自记纸扫描图形曲线提取后的数据处理也是重要的技术之一。

2 降水自记纸扫描图形的数字化精度分析和存储技术

降水自记纸彩色扫描图像的存储空间,取决于扫描参数和存储方式,前提是保证数字化的高精度。

扫描图形数据属离散型点阵数据,其精度与分辨率有直接关系。以21号自记纸为例,取扫描分辨率150 dpi(点/吋)计算,每一像素(点)的间距意味着0.017 mm的降水量,相当于0.027 mm/min,1.616 mm/h,19.39 mm/12 h,38.79 mm/24 h。在雨量强度分级中,2.6 mm/h为中雨,15.1 mm/12 h、25 mm/24 h为大雨。在降雨比较均匀时,当

雨强小于 0.027 mm/min 时,雨量在各个像素点上的值便会人为地出现时有时无的现象。当雨强大于 0.027 mm/min 时,雨量在各个像素点上的值也会人为地出现时大时小的现象。

扫描分辨率的增大与曲线数字化精度的提高应该是正相关的关系。降水曲线自身的精度,曲线的宽度在 0.3 mm 以上,相当于降水量 0.03 mm 以上,超过了扫描分辨率 150 dpi 时每一像素(点)的间距 0.017 mm 降水量,也就是说,曲线自身的精度不高于扫描分辨率 150 dpi 的精度;而且,曲线宽度的变化常常在 0.2 mm 以上,也就是说,曲线自身的精度变化宽度可能造成的曲线跟踪的误差在 0.02 mm 降水量以上,因曲线附近污迹的干扰可能造成的误差更大。因此,扫描分辨率达到一定程度后,因降水曲线自身的精度限制对于降水曲线数字化精度的提高不再有明显的作用。而扫描分辨率的增大会使自记纸扫描图像存储空间大幅度增加(存储空间与分辨率平方成正比)。因此,分辨率的增大是有限度的。

降水自记纸彩色 24 位扫描图像的全信息存储空间(位图 BMP 方式)和黑白灰度 8 位扫描图像的全信息存储空间都非常大。用 150 dpi 分辨率扫描 1 张 21 号降水自记纸,用位图 BMP 方式存储数据,数据量分别是 4881 KB (彩色 24 位)和 1628 KB (灰度 8 位),1 张光盘可存 130 张(彩色)或 390 张(灰度)自记纸图像。按全国地面基本站和基准站 680 站 40 年、平均每年 130 个雨日计算,约有降水自记纸 353600 张,需光盘 27200 张或 9067 张。数据量非常大,进行存储代价很大。必须采用压缩技术,在基本确保质量的情况下进行数据压缩。表 1 给出了在选取同一压缩质量参数条件下对同一张降水自记纸扫描数据的压缩情况。

表 1 图像压缩试验表

	压缩质量	BMP 数据量	JPEG 数据量	压缩率
彩色 24 位	良好	4881 KB	253 KB	95 %
灰度 8 位	良好	1628 KB	227 KB	87 %

显然,压缩的效果相当不错。压缩后的彩色 24 位扫描图像存储空间和黑白灰度 8 位扫描图像存储空间相差不大。采用彩色 24 位扫描真实的纪录降水自记纸图像,采用 JPEG 压缩存储方式减小存储空间,是改变降水自记纸归档载体的有效途径。

3 降水曲线彩色表征区分技术——曲线自动跟踪提取关键技术之一

对降水曲线自动跟踪提取,是实现降水曲线数字化的基础和关键,是决定降水曲线图形数字化精度的基础因素。

降水曲线自动跟踪提取的关键技术有两项,一是降水曲线彩色表征区分技术——把降水曲线与底图坐标线的彩色数字表征有效的区分开来,二是随降水曲线数字表征跟踪提取的技术。降水曲线彩色表征区分技术是降水曲线自动跟踪提取取得好效果的基础条件。

图像的扫描有彩色和灰度两种方式。灰度图像是红、绿、蓝三色的综合反映。降水自

记图像色彩有其自身的特点。以 21 号自记纸为例,其底色和网格坐标线偏黄色,曲线大多偏蓝黑色。用灰度方式,网格坐标线和降水曲线同时存在,基本上无法找到明显区分网格坐标线和降水曲线灰度控制的阈值。

由于降水自记纸底色、网格坐标线和降水曲线色彩的明显差异,寻找其色彩数字表征的特征,是明显区分网格坐标线和降水曲线的关键。我们从亮度、对比度、色彩平衡等几个方面对降水自记纸彩色图形进行了试验。在试验中发现,消除降水自记纸原图像色彩中绿和蓝色的程度差异,强化图像色彩中红色的差异(只表现为红色和黑色二值),可以有效的去除网格坐标线的痕迹,曲线的色彩呈现为黑色,底色、坐标色则呈现为红色,二者差异显著。也就是说,在图像中,底色和坐标的红色程度强,绿、蓝色的程度弱,而曲线的红色程度弱,绿、蓝色的程度强;其综合效果,曲线与底色、坐标色的灰度差异不如单纯红色的差异明显。用红色单色,可以找到明显区分网格坐标线和降水曲线控制的阈值。

因此,可以以红色阈值的有效选取作为降水曲线自动跟踪提取基本手段。

4 降水自记纸扫描图形数字化数据处理技术

提取的降水曲线数据是用纵坐标(降水量)和横坐标(时间)表现的,必须转换成降水分钟强度存储和应用。降水分钟强度数字精确度的确定,必须考虑降水曲线数据自身的精度和数据的空间存储量。鉴于扫描分辨率为 150 dpi 时,每一像素(点)的间距为 0.017 mm 的降水量,故降水分钟强度数字精确度确定为 0.01 mm/min。

降水曲线数据转换成降水分钟强度是将降水曲线数据(纵坐标降水量和横坐标时间)转换成降水分钟强度。数据的转换必须遵循降水总量不变的基本原则。将降水曲线数据两点之间降水量平均分配到相应的分钟时间段上,跨点的分钟时间段降水量,用相应的时间段降水量合计。为了保证降水总量不变,统计值尾数不能采用通常的四舍五入或尾数进、舍的办法取舍,而采用尾数累计进位的办法。

统计中发现尾数累计进位有时间延后的现象。对数据分析表明,是由于两个数除不尽,在运算过程中,可能出现极小的数值误差,影响进位。计算的误差上限,与有效数字的长度有关。采用微量增值的办法,可以解决计算偏差的问题,但要注意微量累计不能达到分钟强度的 1 个单位。

对降水曲线自动跟踪提取形成分钟降水强度的试验中,我们发现分钟降水强度的振荡式波动变化特征。对照自记纸降水曲线,可以发现这种振荡式波动变化,大多不是降水真实情况的反映,而是因为降水曲线自身表征(线条粗细、颜色深浅)的变化和其它痕迹的干扰,造成降水曲线跟踪提取数据的偏差。减少这种偏差的技术处理方法是进行适当的平滑。

平滑的方法很多,采取何种方法取决于平滑的要求。分析降水分钟强度数据的特点,首先确定了平滑方法的基本选取原则,一是保持降水总量不变;二是保持降水时间段不变;三是在降水曲线宽度变化的一定范围内进行平滑。现有的平滑方法无法完全满足上述基本选取原则,在试验的基础上,研究改进了二项式滑动平均方法以适应当前的需求。现将平滑方法表述如下:

降水曲线分钟强度某时段原数据序列为:

$$x_1, x_2, \dots, x_n$$

降水曲线分钟强度某时段平滑后数据序列为:

$$y_1, y_2, \dots, y_n$$

其中:

$$y_i = x_{i-1,1} + x_{i,0} + x_{i+1,-1} \quad (i = 1, \dots, n)$$

$$x_{i,0} = x_i / 2 \quad \text{尾数进 1}$$

$$x_{i,1} = (x_i - x_{i,0}) / 2 \quad \text{尾数进 1}$$

$$x_{i,-1} = x_i - x_{i,0} - x_{i,1}$$

取 $x_{0,1} = x_{1,-1}$

$$x_{n+1,-1} = x_{n,1}$$

x_i, y_i, i, j 均为整数。

上述方法在进行滑动平均的同时,注意到非整除可能带来的误差,其结果完全符合平滑的三条基本选取原则。

经上述方法平滑后的降水分钟强度数据,可以在一定程度上减少降水曲线跟踪提取数据的偏差,但无法消除这种偏差。这是因为我们无法明确分辨实况和曲线提取偏差的界限,因此,只能在曲线可能表现的精度内作一些弥补工作,特别是在降水强度大的时候。

后 记

在对我国降水、暴雨的研究中,确定时段的降水强度及其分布特征是其最基本的、也是最重要的内容。作为基础资料,过去只有 10 min 最大和 1 h 最大降水强度,无法满足对降水和暴雨深入研究的需要。

降水自记纸经数字化处理以后,形成每分钟降水量的数字化基础资料,并最终形成长年代每分钟降水量的长序列资料。在此基础上,可以统计产生任意时段的降水强度及其分布特征;可以滑动统计任意时段的最大降水强度。这些资料,对农业、城市、电讯、交通、水利等领域的规划和建设,对防灾减灾策略的制定提供了定量的数字依据,具有重要的意义。

由此研究开发的降水自记纸彩色扫描数字化处理系统,已于 2002 年底开始由中国气象局预测减灾司组织对全国各省市的推广应用。该项技术适用于其它部门的同类工作,也为其它类型自记纸的数字化处理奠定了良好的基础。

RESEARCH ON PROCESSING TECHNOLOGY OF DIGITIZING GRAPH OF
COLOR SCANNING——THE TESTING OF RESCUING TECHNOLOGY
OF HISTORIC METEOROLOGICAL ARCHIVES

Wang Bomin

(*National Meteorological Center , Beijing 100081*)

Abstract

The digitizing processing of precipitation recording graph using the technology of color scanning graph is an effective test for rescuing historic meteorological archives . According to the precision analysis of digitizing precipitation recording graph and the features of color scanning ,the research of processing technology of data access mode ,distinguishing precipitation curve color and data processing are stressed to explore . The processing method of precipitation recording graph may furnish a thinking way and reference to other digitizing graph processing .

Key words : Color scanning Digitizing graph Precipitation recording graph