

珍贵气象档案数字化技术与应用

■ 鞠晓慧 王妍 李俊

我国自近代气象开展器测观测以来，积累了各类珍贵的纸质气象观测记录档案。2000年之前，主要采用数据录入和缩微技术进行数据提取和异质备份；2000年之后，通过档案修复、图像扫描和元数据著录等技术手段，建立了近2亿页的数字气象档案资源，提升了气象信息化水平。近年来，通过对历史气象档案中的自记纸进行迹线跟踪和数据提取，获得了分钟级分辨率的历史数据，大大丰富了历史气象数据产品，为气候变化研究奠定了重要基础。

DOI: 10.3969/j.issn.2095-1973.2023.03.012

在气象观测活动中形成的各种原始记录、报表、天气图、整编成果等统称为气象档案。根据中国气象局发布的《珍贵气象档案分级鉴定办法》，观测时间序列长、观测数据完整性好、观测要素种类多的原始观测记录档案称为珍贵气象档案。目前中国气象档案馆保存的纸质气象档案主要为2012年之前各类气象台的观测记录档案，其中历史最久的档案可以追溯到19世纪。

早在20世纪70—80年代，中国气象局就开展了气象档案保护和数字化工作。截至目前，已经持续开展了地面、高空、辐射、农气等主要气象观测报表的图像扫描和数据录入，建立了丰富的数字化成果。中国气象局还开展了部分19世纪中叶至20世纪中叶器测资料的数字化，建立了部分重点城市气温、降水、气压等基本要素长序列资料，为近百年气候变化研究提供基础数据。

1 珍贵气象档案

中国气象档案馆馆藏的纸质气象档案从时间上可以分为19世纪中叶至1950年和1951—2012年两部分。19世纪中叶至1950年的纸质档案多为孤本，档案非常珍贵，涉及600多个台站。1951—2012年纸质档案涉及2400多个台站，台站年代比较完整，数量庞大，是档案馆的主体档案。1951年前地面各类气象档案的详情见表1。

中国近代海关气象档案是1951年以前各类地面气象档案中比较重要的档案之一。清朝海关总税务司从1870年开始在中国沿海各口海关和主要灯塔所在地逐

表1 1951年以前各类地面气象档案

档案名称	观测机构	记录时段	记录内容
海关月总簿和月报表	海关测候所	1876—1949年	地面多要素和海洋部分要素
天主教堂月报和年报	国外教会组织	1881—1948年	地面多要素
地面月总簿和月报表	多个机构	1904—1950年	地面多要素
地面年总簿和年报表	多个机构	1906—1950年	地面多要素
地面观测簿	多个机构	1863—1950年	地面多要素
统计底稿	多个机构	1875—1950年	地面和高空多要素
整编资料	多个机构	1841—1950年	地面和高空多要素
地面自记纸	多个机构	1929—1950年	地面单要素
其他	多个机构	1841—1950年	其他

步设立了气象观测站（海关测候所），开展定时气象观测，先后建立的站点有70多个。中国气象档案馆馆藏的近代海关气象档案约有145卷（册）、17万页，涉及66个站点，观测记录30年以上的有40多个，主要有原始纸质海关月总簿（monthly meteorological return）和海关月报表（monthly report）两类。除中国大陆馆藏了绝大部分近代海关气象档案外，海外有少量馆藏，主要是中国海岸气象登记册（China coast meteorological register）（图1）。

2 数字化技术

2.1 纸质档案修复技术

由于历史久远或保管条件所限，珍贵的纸质档案（如19世纪至1950年的各类气象记录档案）出现了不同程度的发黄、破损、酸化等状况，个别档案破损较为严重，必须对档案原件进行抢救和修复。

对纸质档案进行修复和保护的主要措施有：除尘

收稿日期：2021年6月3日；修回日期：2022年8月25日

第一作者：鞠晓慧（1973—），Email: juxh@cma.gov.cn

资助信息：国家科技基础条件平台专项资金项目；山洪地质灾害防治气象保障工程项目（2021年度）

METEOROLOGICAL OBSERVATIONS TAKEN AT *Anping* FOR THE MONTH OF *January 1894.*

RS	BAROMETER	ATTAC. THERM.	WIND	TEMPERATURE	RAIN	WEATHER & REMARKS	DATE	HOURS	BAROMETER	ATTAC. THERM.	WIND	TEMPERATURE	RAIN	WEATHER & REMARKS	
				Metric. Metric.											
A.M.	30.20	66	N.E. 1			b	9	3 h. A.M.	30.19	52	N.E. 1			b	
	28	64	S.W. 2	51	69			5 h.	30.18	52	N.E. 5	46	50		
	19	69	N 3					9 h. P.M.	28	62	N.W. 2				
P.M.	20	64	N.E. 3			c		9 h.	34	58	N.E. 3				
	17	67	N 2			b		9 h. A.M.	33	61	N 1				
A.M.	19	64	N.E. 3	54	68	c		9 h.	33	57	Calcu 0	43	63		
	12	72	N 7			b	10	9 h. P.M.	25	64	N.W. 3				
P.M.	18	68	N.E. 2			c			9 h.	34	60	Calcu 0			
									9 h.						

CHINA COAST METEOROLOGICAL REGISTER.

9th JANUARY, 1894, at 4 P.

STATION.	BAROMETER.	TEMPERATURE.	HUMIDITY.	WIND.		WAVE.	RAIN.
				Direction.	Force.		
Wladivostok,.....	30.10	5	..	NNE	3	b	..
Tekin,.....	29.80	NW	3
Nagasaki,.....	30.15	41	63	NNE	3	e	..
Shanghai,.....	30.33	52	53	NW	2	e	..
Poochow,.....
Amoy,.....	30.25	57	68	NE	1	b	..
Anping,.....	30.19	62	..	SNW	2	b	..
Swatow,.....	30.25	60	..	E	2	b	..

图1 中国馆藏与美国馆藏的1894年1月9日台湾安平 (Anping) 海关气象观测记录 (左图: 中国气象档案馆馆藏的海关气象月总簿; 右图: NOAA图书馆馆藏的海岸气象登记册)

去污、局部修复、字迹去污、修裱与加固、测酸去酸等。

2.2 缩微技术

缩微技术是在感光材料 (通常指胶片) 上记录缩微影像的技术过程。缩微技术不是数字化处理技术, 但它是纸质档案为加工对象, 是档案异质保存的重要手段。20世纪70—90年代, 气象部门利用缩微摄影技术对历史天气图、气象记录报表和解放前部分历史气象记录档案进行了载体转换, 数量达10万张 (盘), 保存在中国气象档案馆延庆分馆。

2.3 纸带穿孔数字化技术

数据纸带是早期计算机的输入输出手段。通过编制好的程序将气象观测资料使用穿孔机穿成数据纸带的过程称为纸带穿孔。纸带穿孔是早期计算机应用的数字化方式。最早使用纸带穿孔技术进行数字化的气象档案主要是国家级台站地面气象记录月报表, 包括地面观测的19类要素项目。

2.4 图像扫描技术

图像扫描是使用黑白或彩色扫描仪对纸质气象档案进行逐页扫描, 经纠偏、去黑点、压缩等系列操作, 保存为TIFF或JPEG文件的处理过程。扫描技术参数的选择决定着扫描质量。因此, 在扫描处理过程中要根据档案原件的质量和利用需求来设置扫描技术参数和存储格式。扫描后的图像文件也是数字档案的重要档案资源, 用于物联网时代的档案资源共享。

2.5 迹线数据提取技术

与纸质气象报表人工键盘录入不同, 自记纸迹线提取基于较为成熟的计算机技术, 研制人机交互软件, 才可以实现批量规模化处理。自记纸迹线提取软件主要包括图像文件预处理、自记迹线识别、数据提取、标准化数据转换、数据质量控制等技术。在以上的技术处理中, 迹线识别和数据提取是关键, 决定了

提取是否有效以及提取后的数据是否准确。

随着人工智能技术的发展, 基于深度学习的图像跟踪技术已应用到因风自记纸迹线提取软件中, 减少了跟踪误差, 从而减少了人工核查的工作量, 大大提高了数据提取的效率。

自记纸迹线提取时由人工操作提取软件, 逐站逐张提取并进行人工干预和修正, 提取完一个站所有自记纸后进行数据格式转换和数据保存。为了保证提取数据的质量, 需要对提取后的数据进行多次质检和评估。评估不合格的台站需经过整改和重新提取, 直至数据合格。图2为一张降水自记纸迹线提取后的回放效果图, 提取的逐小时降水数据已标识在图像上。

3 数字化技术应用

从1979年1月开始, 中国气象局启动了气象档案信息化处理工作。2000年之前, 以纸带穿孔技术和键盘录入技术为主, 目的是获得时间序列较长的历史观测数据, 以地面观测数据为主。2000年之后, 随着扫描仪的发展普及, 同时认识到纸质档案进行双备份的重要性, 数字化的技术以图像扫描和数据录入为主, 数字化处理从地面资料扩展到高空资料、辐射、农业气象等。近年来, 随着计算机图像处理技术的发展, 对自记纸图像图形进行自动识别和数据提取成为数字化工作的重点。另外, 随着全球气候变化成为热点问题, 中国气象局也加强了对馆藏的19世纪中叶至20世纪中叶长达百年的珍贵气象档案的数字化工作。

中国近代以来至1951年各类气象观测档案主要包括: 天主教会观测月报和年报, 海关总署海关月总簿和月报表, 日本在华观测档案, 国民政府或一般月总簿、一般月报表、自记纸、出版物以及各类统计表 (示例见图3)。这些珍贵档案涉及台站约600个 (分布在大约500个城市), 观测时间为1841—1950年, 但是绝大多数台站观测不连续。由于这些档案来自不同国家或机构、使用不同的观测仪器和单位、采用不同

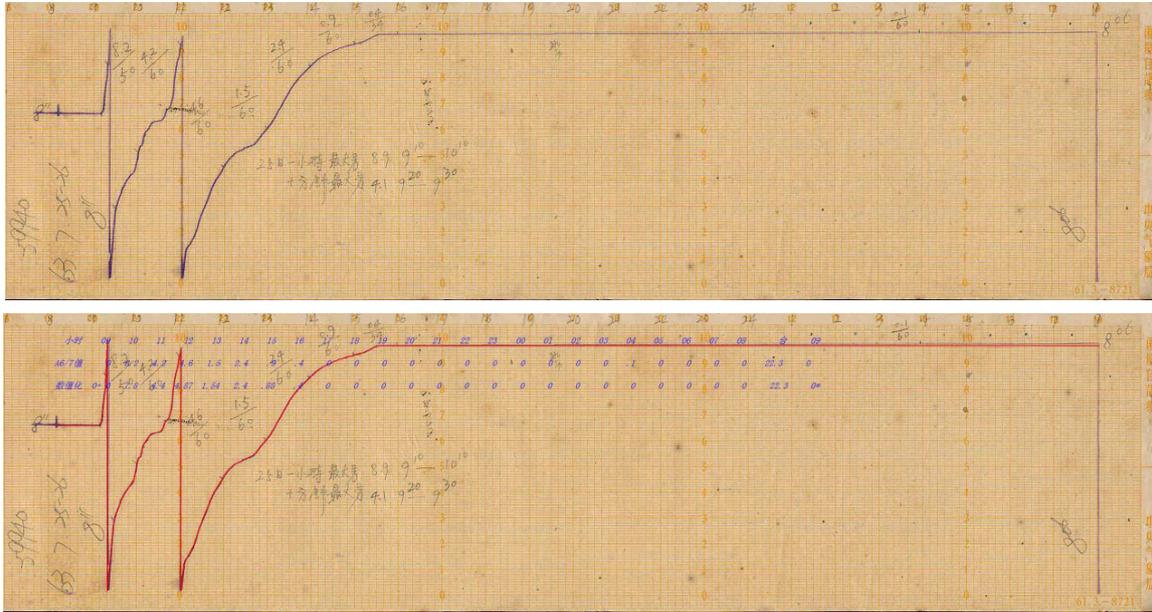


图2 降水自记纸迹线提取效果(上图:原图;下图:迹线提取后效果图)

102

Пекинъ.					Мартъ.					1872.						
Число. Datum.	Температура воздуха. Lufttemperatur.				Абсолютная влажность. Absolute Feuchtigk.			Относительная влажность. Relative Feuchtigk.			Барометр. Barometer.					
	7	1	9	Средне. Mittel.	7	1	9	Средн. Mittel.	7	1	9	Средне. Mittel.	7	1	9	Средне. Mittel.
1	-5.7	+2.6	-0.8	1.30	2.8	3.1	2.9	3.03	96	97	98	78.7	772.0	771.7	772.1	771.98
2	-2.1	+5.6	-0.5	+0.27	2.9	3.5	3.5	3.08	90	92	92	81.2	773.4	773.8	771.4	774.37
3	-2.2	+5.5	-0.1	+1.00	3.5	3.8	3.5	3.08	85	87	78	78.5	770.7	769.5	769.0	769.67
4	+1.6	+0.8	+0.9	-0.59	3.0	3.9	3.7	3.38	74	66	75	71.7	767.0	765.9	765.7	765.87
5	+0.1	+6.5	+1.7	+2.10	2.9	3.4	1.9	2.73	64	54	37	51.7	761.9	760.1	763.1	761.70
6	-0.5	+6.2	+0.5	+2.27	1.8	1.8	2.6	2.07	41	24	54	39.7	764.3	763.5	763.8	763.70
7	-4.9	+11.9	+7.1	+4.67	2.4	2.4	2.8	2.70	73	24	44	47.0	761.3	759.1	759.5	759.37
8	+0.5	+16.5	+7.6	+8.13	2.9	4.1	4.0	3.67	61	30	31	47.5	755.4	754.8	756.7	755.63
9	+2.2	+8.8	+4.5	+3.10	3.7	2.8	3.7	3.40	68	33	69	55.7	760.4	759.6	761.2	760.40
10	-0.8	+9.3	+2.9	+3.97	3.6	4.0	4.1	3.90	81	45	73	66.5	763.2	762.0	760.4	761.68
11	+0.1	+9.9	+4.5	+4.83	4.0	5.1	5.0	4.90	87	56	89	77.3	759.0	757.1	756.0	757.37
12	-1.0	+12.4	+6.1	+5.83	4.0	5.4	4.5	5.37	84	60	90	81.2	758.5	756.9	755.8	756.40
13	+1.7	+16.1	+10.4	+9.40	5.0	1.8	1.8	2.00	96	14	19	43.0	753.5	752.0	754.8	752.68
14	+3.1	+12.0	+5.8	+6.37	3.1	3.8	3.6	3.68	87	28	32	37.0	753.9	755.5	755.9	755.77
15	+0.8	+10.5	+4.5	+6.03	3.6	4.0	4.6	4.07	76	43	73	54.0	750.6	750.5	750.8	750.63
16	+0.9	+8.3	+4.8	+4.67	3.6	3.9	4.6	4.08	72	48	71	63.7	762.0	759.7	758.1	759.28
17	+6.9	+6.9	+3.7	+4.25	4.2	3.9	3.5	4.20	74	67	67	76.0	753.5	753.4	753.7	753.57

图3 1951年前珍贵档案数据示例(左图:俄国气象年报记录的北京1872年3月的气象数据;右图:海关月总簿记录的汉口1880年5月的气象数据)

的时制,无论是纸质档案的物理状况还是记录的数据状况都非常复杂,扫描和数据录入的处理难度都比较大。

中国气象局在全国范围内选取有代表性的116个城市,对19世纪至20世纪中叶的观测记录档案中气压、降水、气温、风向、风速和湿度等要素数据进行了人工录入,形成了重要的数字化成果。

百年长度的历史长序列是研究长期气候变化的基础。相关学者基于已数字化的百年气温、降水等要素的日值、月值资料,研究得出了中国1841年以来的气候变化趋势和特征,相关研究成果已发表在国内外科技期刊上。在中国近代以来珍贵档案数字化成果基础上,研究气象站建站时间、观测情况及观测资料气候代表性等,从而为中国百年气象站认定工作提供重要依据。在国际合作方面,中国气象局参与了中英合作项目“国际大气环流重建计划(ACRE)”中国子计划,在中国区域18世纪以来早期器测气象资料的挖掘、数

字化和应用方面取得了进展,项目成果已应用在20世纪再分析资料(20CR)。

4 结语

中国历史气象档案拯救与数字化工作始于1979年,并在2006年以后快速发展。40多年的工作历程取得了丰硕的成果,基本满足了气象业务、科研和社会各行业对基础气象数据的需求。但是,与中国气象现代化和信息化发展需求相比,与国内档案行业发展水平相比,气象档案数字化还存在一定的差距,主要表现在以下两个方面:一是气象档案信息化程度不够。目前,国省气象档案馆馆藏档案还有接近50%的馆藏档案没有进行图像扫描处理,地面自记纸仅完成了约15%的数据化处理,中国国家级地面站1951年以来的逐小时气温、气压和相对湿度数据序列还未建立。二是多种类、多形式、高质量的历史基础数据集产品研发能力有待提高。需要通过对扫描图像文件元数据整编,实现电子图像文件100%在线访问、检索和应用,

从根本上实现对纸质档案的保护和充分利用。

针对以上不足，今后中国历史气象档案拯救与数字化工作有三个重点：一是开展气温、气压和湿度自记纸迹线提取以及近代气象观测以来的重要站点资料的数字化，建立1951年以来国家级台站分钟和小时分辨率历史数据，以及百年气象台站历史气象数据；二

是基于数字化原始成果，研制形式多样（例如格点资料、图形产品以及定制产品等）的气候基础数据集产品，满足多个行业部门的应用需求；三是依托数字气象档案馆建设，提高历史气象档案扫描图像成果应用能力，改变传统的纸质档案服务方式，提高档案的应用效率。

深入阅读

吴增祥, 2007. 中国近代气象台站. 北京: 气象出版社.
范邵华, 余子, 鞠晓慧, 等, 2018. 气象资料数字化进展及其应用. 气象科技进展, 8(1): 64-70.
兰平, 2019. 气象档案事业70年回顾与展望. 气象科技进展, 9(4): 67-69.
何溪澄, 冯颖竹, 2017. 《海关医报》与1877—1894年广州气象观测记录. 气象科技进展, 7(3): 71-73.
宋建萍, 何晓, 苏秀梅, 等, 2016. 近代湖北海关气象观测档案初探——以江汉关、宜昌关、沙市关为例. 气象科技进展, 6(6): 71-74.
陈永生, 李娜娜, 2017. 中国近代海关档案的分布与现状. 中国档案, (8): 58-59.
鞠晓慧, 马楠, 王妍, 等, 2022. 基于深度学习的气象资料迹线识别.

科学技术与工程, 22(21): 9215-9222.
王伯民, 吕勇平, 张强, 2004. 降水自记纸彩色扫描数字化处理系统. 应用气象学报, 15(6): 737-744.
Ju X H, Huang S P, Li C J, et al, 2019. Development of the Self-recording Per-minute Precipitation Dataset for China. Journal of Meteorological Research, 33(6): 1157-1167.
Cao L J, Yan Z W, Zhao P, et al, 2017. Climatic warming in China during 1901—2015 based on an extended dataset of instrumental temperature records. Environmental Research Letters, 12: 064005.
战云健, 陈东辉, 廖捷, 等, 2022. 中国60城市站1901—2019年日降水数据集的构建. 气候变化研究进展, 18(6): 670-682.

(作者单位: 鞠晓慧、王妍, 国家气象信息中心; 李俊, 中国气象局预报与网络)

(上接68页)

视频方便快捷的优势，可以提前告知民众气象预警信息，从而指导民众科学、理性地对待气象灾害。

4) 建立气象科普品牌特色是科学普及的长效机制

气象科普工作应积极建立气象科普特色品牌。气象科普特色品牌的创立，是促进气象科普工作发展不可或缺的部分，不仅有利于气象权威的树立和传播，还能扩大气象科普工作的影响力。在科普基地建设、校园气象、科普活动、文创产品研发等领域多点开花，形成主题鲜明、特色突出、影响深远的气象科普品牌。

比如中国气象局创设的“象博士”科普特色品牌，通过组建一支专业的气象专家团队，为科普产品提供权威保障，利用“象博士”的形象发布气象预警信息和科普资讯。上海气象博物馆的吉祥物“气象本象”也借鉴其理念，依靠其可爱的卡通形象和科普活动直播互动、短视频、科普漫画等系列作品，积极为民众答疑解惑，与民众之间建立了沟通交流的桥梁。并在3·23世界气象日、5·12防灾减灾日、全国科普日等重大气象节日内开展科普活动的同时宣传品牌理念，给科普品牌的影响力带来了巨大的提升。

此外，文创产品的设计与发售能将品牌的理念和价值观通过实物输送给民众。打造以气象知识、气象科普场馆为主题的文创产品，如手提袋、书签、钥匙扣等轻便的小物件，能增加用户对品牌的熟悉感、亲

切感和信赖感。

4 小结

气象科普工作仍然任重道远。气象科普基地作为气象科普工作的载体，承担着给民众普及、宣传气象科普知识的重要任务。但是，现阶段民众的防灾减灾意识较为淡薄，且气象科普基地对民众的科普效果有待定量。要加强民众对科普工作的反馈，需以气象科普基地为中心，结合多种科普工作方式，对各层面的民众进行科普服务。

深入阅读

田艳芳, 2020. 关于加强气象科普工作的几点思考. 内蒙古科技与经济, (20): 47-48.
朱晔, 朱定真, 2019. 浅谈上海超大城市气象科普实践与启示//中国科普研究所、安徽省科学技术协会. 中国科普理论与实践探索——第二十六届全国科普理论研讨会论文集. 北京: 中国科普研究所.
田依洁, 刘晓晶, 袁旷卓, 2020. 虚拟现实技术在气象科普工作中的应用. 科技传播, 12(20): 52-54.
孙镛涵, 陆明明, 魏磊, 等, 2019. 浅谈提升气象科普传播能力. 黑龙江气象, 36(3): 43-46.
韩志鹏, 袁园, 谢晶心, 等, 2020. 融媒体时代下短视频在气象宣传科普工作中的应用思考. 黑龙江气象, 37(2): 33-34.
张岚, 孙燕, 2021. 融媒体环境下气象科普品牌构建研究. 科技传播, 3(3): 63-65.

(作者单位: 严梓畅、邹嘉南、周笑迁、何南腾, 南京信息工程大学中国气象局气溶胶-云-降水重点开放实验室, 教师教育学院; 武欣蕊, 中国气象局气象干部培训学院)