

# 基于自记观测和自动站观测的四川省降水数据集研究与应用

王凌<sup>1,3</sup> 鞠晓慧<sup>2</sup> 刘霄<sup>1,3</sup> 杜冰<sup>1,3</sup>

(1 四川省气象探测数据中心, 成都 610072; 2 中国气象局国家气象信息中心, 北京 100081;  
3 高原与盆地暴雨旱涝灾害四川省重点实验室, 成都 610072)

**摘要:** 历史气象资料的数字化技术和科学研究不断发展和进步, 通过对四川省国家级气象站近80万页降水自记纸迹线的数字化, 并融合气象站升级自动观测系统后的降水观测数据, 研制四川省序列最长、时间密度最大的降水强度数据集, 将四川省降水分钟数据提前到1951年; 通过对数据集的完整性、规范性、数据约束等质量控制以及各相同时段降水过程总量最大值和对应时间过程等一致性对比分析, 评估该数据集的适用性和准确性; 广泛开展历史气象资料的拯救、数字化等研究活动, 深度挖掘珍贵历史资料最大信息, 为气象服务和科学研究提供更好的数据支撑, 实现历史气象观测记录的科学价值。

**关键词:** 数字化, 降水, 质量评估及应用, 四川省

DOI: 10.3969/j.issn.2095-1973.2023.04.022

## Research and Application of Precipitation Dataset in Sichuan Province Based on Self-recording Observation and Automatic Station Observation

Wang Ling<sup>1,3</sup>, Ju Xiaohui<sup>2</sup>, Liu Xiao<sup>1,3</sup>, Du Bing<sup>1,3</sup>

(1 Sichuan Meteorological Observation Data Center, Chengdu 610072; 2 National Meteorological Information Center, Beijing 100081; 3 Sichuan Provincial Key Laboratory of Heavy Rain and Drought-Flood Disasters in Plateau and Basin, Chengdu 610072)

**Abstract:** The digitization technology and scientific research of historical meteorological data have undergone continuous development. Through the digitization of nearly 800,000 pages of precipitation self-recording paper traces of the national meteorological stations in Sichuan Province, and the integration of precipitation observation data after the upgrading of the automatic observation system of the meteorological stations, the precipitation intensity dataset with the longest sequence and the highest time density in Sichuan Province was developed, and the precipitation minute data of Sichuan Province was advanced to 1951. The applicability and accuracy of the dataset was evaluated by the quality control of the dataset, including the completeness, normativity, and data constraints, and by comparing the consistency of the maximum values of the total precipitation processes in the same time periods and the corresponding time processes. Extensive research activities such as the rescue and digitization of historical meteorological data are carried out to dig deep into the maximum information of precious historical data, provide better data application support for meteorological services and scientific research, and realize the scientific value of historical meteorological observation records.

**Keywords:** digitalization, precipitation, quality assessment and application, Sichuan Province

### 0 引言

早期气象观测资料多以纸质记录为主, 应用处理困难, 而气象纸质历史资料有很高的科学研究价值和很重要的历史保留意义<sup>[1-2]</sup>。从20世纪70年代末至今,

国内气象部门依靠“气候变化应对工程”“山洪地质灾害防治气象保障工程”等多个建设项目, 开展历史纸质气象资料的数字化工作。数字化技术从早期的人工录入到现在的利用图像识别技术进行报表数据识别和自记纸的迹线提取。随着计算机和网络技术的发展, 针对报表资料, 韩瑞等<sup>[2]</sup>开发了基于DBnet和众筹策略的气象纸质表格快速数字化方法及系统; 针对自记迹线的数字化, 中国气象局气象信息中心开发了基于深

收稿日期: 2022年4月24日; 修回日期: 2022年9月9日  
第一作者: 王凌(1968—), Email: wang\_ling126@163.com  
通信作者: 鞠晓慧(1973—), Email: juxh@cma.gov.cn.  
资助项目: 中国气象局山洪地质灾害防治气象保障工程建设项目

深度学习模型的人工智能自动识别方法,实现自记迹线的高精度识别<sup>[3]</sup>。气象数据再分析技术(RA)和气候变化研究,催生抢救早期气象观测记录的国际合作项目,2019年11月,国际地球大气环流重建(ACRE)计划中国、东南亚和日本子计划联合学术研讨会召开,中国气象局积极参与了“ACRE CHINA”子计划项目,让我们有机会进一步认识早期气象观测记录的价值所在,持续对早期观测记录开展拯救、数字化和分析等研究活动<sup>[4]</sup>。

降水现象是常规地面气象观测业务的基本内容之一<sup>[5]</sup>。2000年前后,中国气象局开始在台站布设地面自动气象观测系统,逐步实现了人工观测向自动观测方式的转变<sup>[6]</sup>,使得气象观测数据直接以数字的方式呈现。早期的降水自记纸虽然记录了观测时段降水的全过程信息,但一般只是人工读取了自记纸的部分定时值,还远没有发掘出更多自记纸上气象观测数据的价值。本文将四川省有降水自记观测的151个气象站历史降水自记纸记录,通过技术方法实现数据化,整合自动站正式运行以来的降水数据,最终建立了四川省159个国家级气象站自建站有降水观测以来的完整序列的降水强度数据集,使四川省最早的降水分钟数据从最早建立自动气象站的2004年提前到1951年。

## 1 数据集概述

### 1.1 数据集简介

四川省分钟降水数据集(SURF\_CLI\_SC\_PRE\_MIN)和小时降水数据集(SURF\_CLI\_SC\_PRE\_HOR)包含了四川省区域内(98.10°—108.03°E, 26.5°—33.58°N)159个国家级地面气象站有降水自记观测(最早1951年)至2020年逐分钟和小时降水数据文件以及数据集说明、数据格式说明、台站信息表、异常处理情况登记表和元数据文档,其中包含了成都、内江、泸州3个已经撤销的国家站近50年的自记观测数据。数据集的站点(图1)涵盖了四川省各区(市、县),分布较为均匀,对于解析四川省盆地、山地、高原等不同地形地貌降水特征,具有重要意义。

### 1.2 数据研发

#### 1.2.1 数据来源

数据来源于以下几部分:1)四川省151个国家级气象站(个别高原站无降水自记观测)建站有降水自记观测—自动站正式观测前一年的自记纸迹线提取分钟、小时降水数据(R01和R60文件);2)四川省156个国家级自动气象站运行以来分钟和小时地面要素月数据(J和A文件);3)个别站自动观测时J文件丢失而

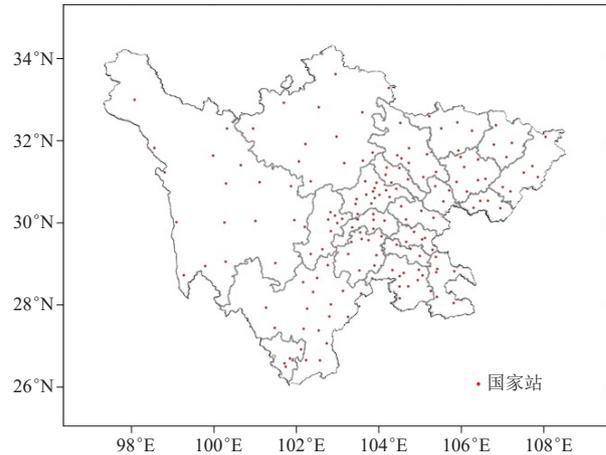


图1 四川省分钟、小时降水强度数据集站点分布  
Fig. 1 Station distribution in the dataset of precipitation intensity per minute or hour in Sichuan Province

有相应降水自记纸,通过补充提取自记纸迹线而形成的分钟数据(R1文件)。

#### 1.2.2 技术路线

通过对四川省历史降水自记纸扫描图像文件中降水迹线的数字化提取,形成降水分钟、小时数据文件,结合自动观测以来的降水数据,形成有自记降水观测以来长序列、全时段省内国家地面气象站分钟、小时降水数据<sup>[7]</sup>。

##### 1.2.2.1 降水自记记录的数字化提取

在历史气象资料拯救与数字化实施过程中,为保证资料安全及数字化质量、节约成本,需遵循安全性、真实性、内容选择、全程规范化管理等原则<sup>[8]</sup>。

利用中国气象局推广的“降水自记纸数字化处理系统软件”,对四川省气象档案馆馆藏的70余万页降水自记纸迹线进行逐张跟踪提取。最终形成降水曲线文件(\*.zjr)以及各站参数文件(\*.lib)<sup>[9]</sup>。图2为提取示意图。

利用“降水自记纸数字化成果质量检查和评估软件”,通过人工和软件相结合的方式,回放检查迹线跟踪准确性,时间和异常处理是否正确,通过软件统计和挑取的小时数据、趋势、过程总量、年度最大降水、最长连续降水是否吻合等三级质量检查,确认提取迹线正确。图3为质量检查示意图<sup>[10-11]</sup>。结合降水曲线文件和参数文件完成数据转换,形成自记降水小时(R60)、分钟(R01)数据文件<sup>[9]</sup>。

##### 1.2.2.2 自动站降水分钟数据的获取

受早期自动站和质量控制技术的局限性影响,各时期对自动站地面观测分钟月数据文件(J文件)的质量检查把握尺度不一,与经过严格质量控制的小时月

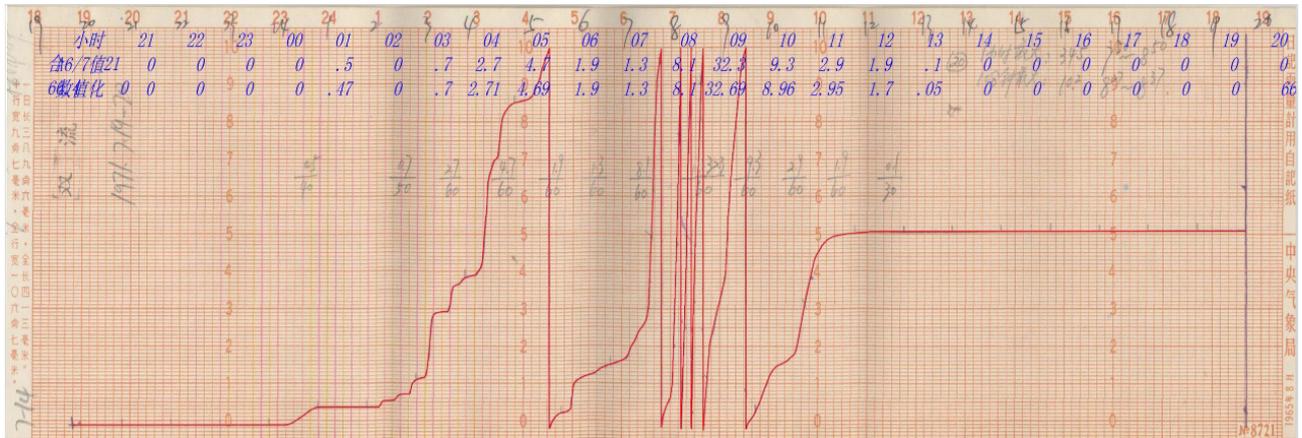


图2 降水自记纸迹线整张提取完成示意图

Fig. 2 Complete extraction of self-recording precipitation traces

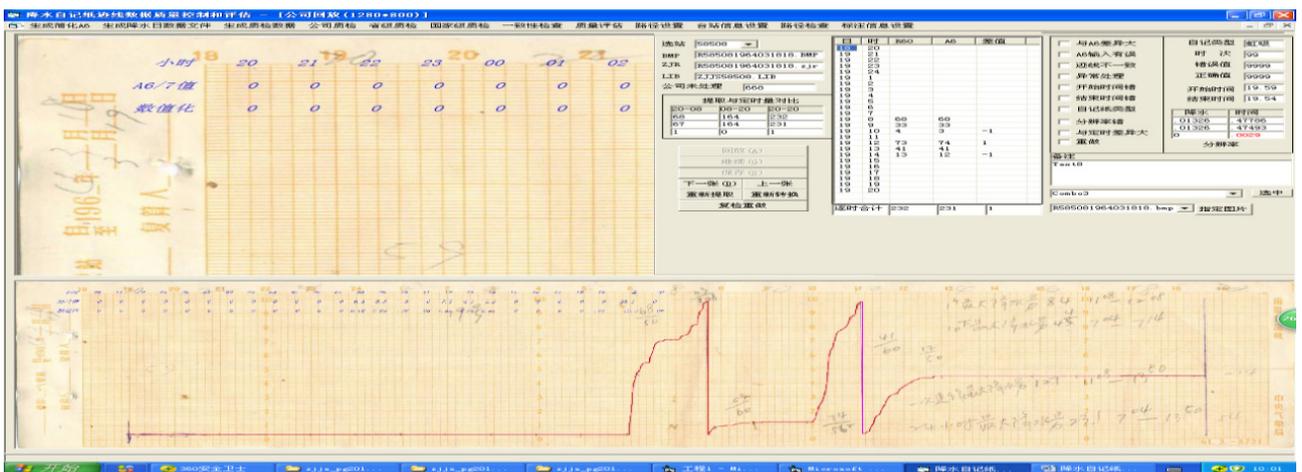


图3 降水自记纸迹线提取质量检查示意图

Fig. 3 Quality inspection for the extraction of self-recording precipitation traces

数据文件(A文件)的降水数据存在差异或不匹配情况。数据研发过程中对四川省156个国家级气象站自动站正式运行后逐月降水数据进行了严格的检查和数据修正(包括格式、气候界限值、极值以及内部一致性检查等)。

### 1.2.2.3 两类数据融合

降水自记纸迹线提取形成的R01、R60文件分别和J文件中的分钟降水数据、A文件中的小时降水数据进行格式转换、融合,分别形成四川省有自记降水观测以来完整序列的分钟降水记录和小时降水记录。

根据降水数据集制作技术规范,编制数据集格式说明、台站信息等说明文档和元数据文档,建立四川省分钟、小时降水强度数据集<sup>[11-12]</sup>。

## 2 数据集的质量评估

对数据集的评估重点考虑数字化过程中人为或技术的漏、误等影响因素以及降水数据本身的属性,根据质检规则和数据约束,对数据进行单点极值阈值的

质量控制并且对站点时间、空间数据进行比对分析,评估数据的准确性和适用性。

### 2.1 数据集质量控制

#### 2.1.1 完整性、规范性检查

该数据集在中国气象局的评估结果中达到完成率100%,合格率100%。通过了中国气象局数据成果完整性、规范性、一致性检查。

#### 2.1.2 数字化部分数据检查

数字化过程中三级质检,通过软件统计核实降水定时、趋势、过程总量、最大降水、最长连续降水与已有人工观测记录的吻合情况,确认了数据正确性。

为避免数字化过程中缺漏提取或自记纸缺失的问题,要检查自记纸数字化提取的降水数据的完整性。方法是利用数据集分钟降水数据计算日降水合计值,与同期人工观测雨量器日降水量 $R$ 进行比较,检查降水日数和降水量一致性。对超出误差范围的样本,计为数据不一致,判为疑误,进行核实修正。根据技术

规定误差范围如下：雨量器日降水量 $R \leq 5.0$  mm，绝对误差 $\leq 0.5$  mm；雨量器日降水量 $R > 5.0$  mm，绝对误差 $\leq R \times 10\%$ 。

通过对该数据集151个站70多万页自记纸的一致性统计对比，共检测出超误差范围的样本数为5755个，约占总样本数的0.8%。经人工核实，其中包含有未提取的周记自记纸和部分历史原因批量虫蛀和损坏的自记纸，其它经核查，或确认与人工雨量器观测雨量存在误差，数字化提取数据无误，或进行了修正。

### 2.1.3 数据极值检查

降水要素的极值问题一直是研究暴雨过程的重要因素，保证数据的准确性尤为关键。选取分钟降水界限值为6 mm，针对数据集中分钟降水量 $\geq 6$  mm的数据提出疑误，进行人工核实。经查，该数据集中分钟降水量 $\geq 6$  mm的记录仅为16条，全部进行了人工校对修正。

## 2.2 数据适用性评估

自动站J文件的分钟降水数据经过了研发过程中的严格审核，数据集的数据适用性主要集中在对降水自记迹线数字化提取的分钟降水自记数据（以下简称自记提取）和同期人工读取自记迹线降水数据（以下简称人工记录）的一致性分析，来验证数字化提取的分钟降水数据的适用性，进而检验数据集的准确性<sup>[13]</sup>。

以数据集中1980—2000年四川省盆地温江、德阳、蒲县、乐山、自贡、安岳6个国家站分钟降水数据为基础，用逐分钟滑动求和分别计算迹线提取的逐年逐5、10、15、20、30、45、60、90、120、180、240、360、540、720、1440 min共15个时段的降水强度极值和时间过程，与对应人工记录地面气象观测年报表相同历时下降水强度的极值和过程对比分析。

### 2.2.1 降水极值的分析

图4a和4b分别是蒲江站各年5 min和30 min历时

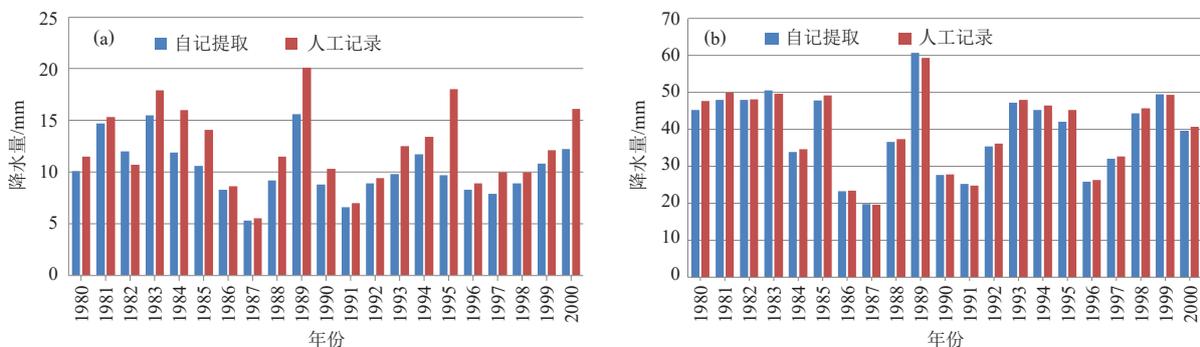


图4 1980—2000年蒲江站5 min历时(a)和30 min历时(b)自记提取和人工记录降水强度极值对比

Fig. 4 Comparison of 5-min duration (a) and 30 min duration (b) of precipitation intensity extremes by self and manual records at Pujiang Station from 1980 to 2000

自记提取和人工记录降水强度极值的对比图，直观可见5 min历时个别年份存在一定差异，30 min历时差异减小。

图5a~5d分别是温江站1980—2000年5 min、10 min、30 min和60 min历时自记提取和人工记录降水强度极值的相关性。可见随着降水历时的延长，两者相关性线性收敛。

统计分析1980—2000年6个国家级气象站自记提取和人工记录降水强度极值的相关系数(表1)可见，通过数字化提取的降水数据和人工记录数据相关性比较好，并且历时越长，相关性越好，相关系数超过0.99。

### 2.2.2 时间过程分析

分析各历时最大降水量时间过程的差异，计算1980—2000年各历时自记提取和人工记录最大降水过程的一致性。分别统计自记提取和人工记录某一年某个时段最大降水出现的时间，如果两者相同，表示两种方式选取的此次降水时间过程一致。考虑到自记纸迹线数字化提取和人工记录两种方式的误差，以及大降水时设备异常偏多，自记和人工对异常处理的偏差，将两者时间差异在60 min以内算作过程一致。多年中某时段最大降水时间过程一致率的算法如下：一致率=(最大降水过程时间差 $< 60$  min出现的次数)/统计年数

以德阳站为例，统计1980—2000年15个历时最大降水的时间过程，结果见图6。各历时的最大降水过程一致率平均为84%。

其中45 min降水过程一致率只有29%，说明1980—2000年21年中45 min最大降水自记提取和人工记录的时间段不一致的较多。核实差异最大的一条记录1993年45 min最大降水，发现1993年两次大的降水过程出现在7月15日和7月29日，在5~30 min历时情况下，自记提取和人工记录的最大降水过程都选取了7

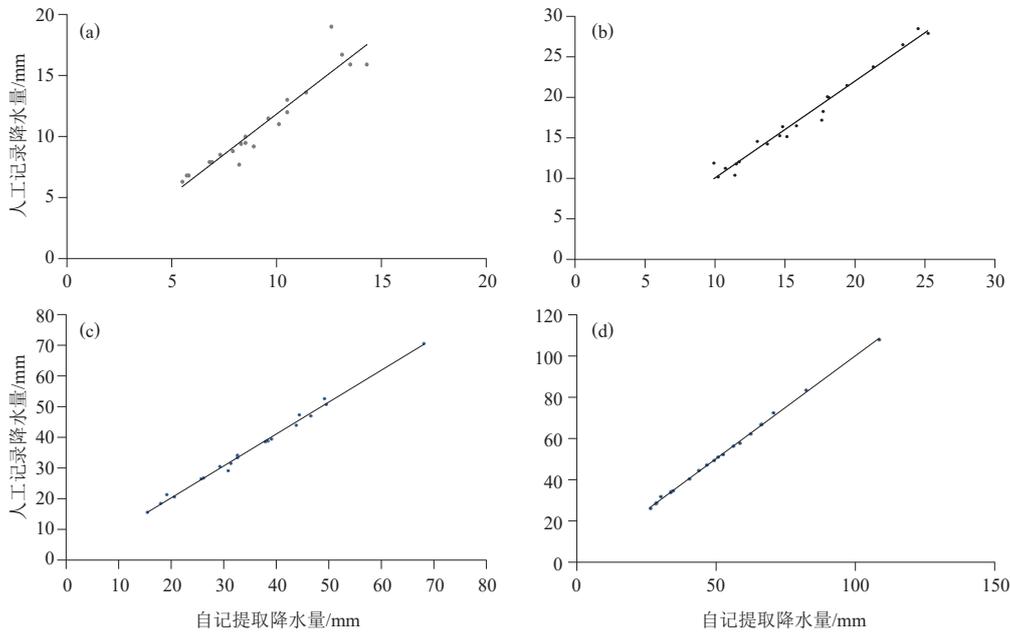


图5 1980—2000年温江站5 min (a)、10 min (b)、30 min (c)和60 min (d) 历时自记提取和人工记录降水强度极值相关性  
Fig. 5 Correlation of 5-min duration (a), 10-min duration (b), 30-min duration (c) and 60-min duration (d) of precipitation intensity extremes by self and manual records at Wenjiang Station from 1980 to 2000

表1 1980—2000年6个国家级气象站不同历时自记提取和人工记录降水强度极值相关系数

Table 1 Correlation coefficient of extreme values of precipitation intensity by self and manual records at 6 national meteorological stations during different periods from 1980 to 2000

降水历时/min	5	10	15	20	30	45	60	90	120	180	240	360	540	720	1440
相关系数	0.84	0.94	0.97	0.94	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99

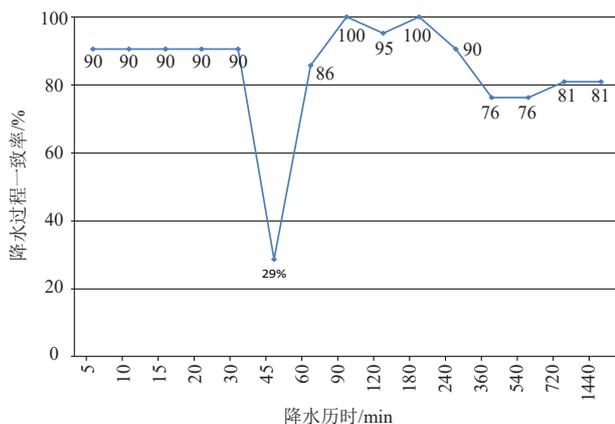


图6 1980—2000年德阳站不同历时自记提取和人工记录最大降水过程一致率

Fig. 6 Consistency rate of self-recorded and manual recorded maximum precipitation process at Deyang Station in different periods from 1980 to 2000

月15日, 60~1440 min的最大降水过程都选取的7月29日, 而45 min最大降水自记提取选择了7月15日降水量为16 mm, 人工记录选择了7月29日的降水, 降水量为15.9 mm。这是因为人工记录和信息化提取之间的差异所致。

根据以上对比分析, 说明通过自记迹线提取数字化的分钟降水数据与人工记录数据的结果在整体及单次降水过程中的表现较为一致, 且历时越长数据稳定性越高。数字化结果比人工记录结果更为准确和客观, 能够较好地替代早期人工记录的数据结果<sup>[14]</sup>。

### 3 数据集产品应用

四川省降水强度数据集丰富了历史气象观测降水分钟数据, 并在实际应用中得到验证。在气象预报预测业务和数据再分析、气候变化、灾害预警等科学研究和气象服务中将起到重要的数据支撑作用。

#### 3.1 在气候业务研究中的应用

降水强度对于研究降水特征, 尤其是对暴雨特征的研究具有重要意义<sup>[15]</sup>。该数据集能够准确为历年、累年降水地区分布、年际变化及暴雨等级特征分析等气候业务、气候变化研究提供数据支持。

2017年中国气象部门联合住建部门进行暴雨强度公式的编制和修订, 为城市规划、径流控制、“海绵城市”建设等提供重要科学依据。四川省气候中心和部分地市气象局在暴雨强度公式编制和修订项目中, 利用该数据集的分钟和小时降水数据, 因为数据时间精

度高，序列长，准确实现了“不漏场次、不漏最大值”的挑选降雨场次的原则，统计样本准确可靠。

以雅安市（俗称“雨城”）为例，根据1980—2014年雅安暴雨强度资料，按《室外排水设计规范》（GB50014—2006，2014版）对各历时降水的概率分布进行拟合，对单一重现期暴雨强度公式、区间参数公式、暴雨强度总公式进行推求，并对计算结果进行比

较分析。最终选择满足规范要求且误差最小的暴雨公式，公式拟合结果见图7和图8<sup>[16]</sup>。

### 3.2 在灾害预警气象数据分析中的应用

四川盆地四面环山，地形特殊，强降水易引发滑坡泥石流灾害<sup>[17]</sup>。降水强度、分布和变化等降水特征分析在防灾减灾策略特别是汛期洪涝灾害预警研究和服务中尤为重要。该数据集能够满足任意时段降水

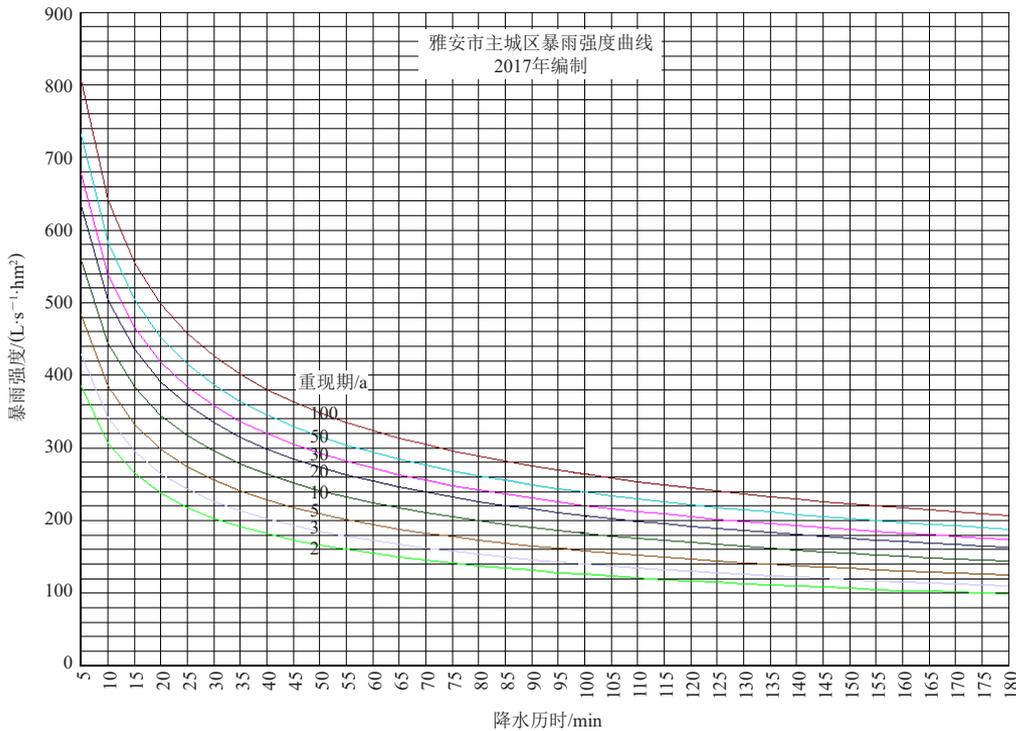


图7 1980—2014年雅安市暴雨强度—重现期—历时关系曲线图

Fig. 7 Relationship curve of rainstorm intensity-return period-duration in Ya'an City from 1980 to 2014

过程个例、累计雨量及极值等数据分析对基础数据的要求。

1981年7月四川发生了一次强烈发展的西南涡暴雨天气过程，“81·7”大暴雨持续时间长，雨区大，山洪灾害造成严重伤亡和重大经济损失。图9是基于该数据集的分钟降水数据制作的1981年7月最大30 min降水量的空间分布。

## 4 结论与讨论

基于降水自记观测和自动站观测的四川省降水数据集，通过对四川省有降水记录以来151个国家气象站降水自记纸迹线的数字化提取，结合自动站观测的降水数据，形成了四川省有降水记录以来时间序列最长、时间密度最大的数据产品。通过数据集的研制，最大限度提取了纸质降水记录的信息，这部分数据可谓首次面世，填补了历史气象记录中逐分钟降水数据的空白。未来将在此基础上深度挖掘，编研有关

四川省以及不同地区、地形地貌的特色降水数据集，并尝试做一些降水和人类活动、城市化进程的关联性研究。

历史气象资料数字化的目标是挖掘历史气象资料的信息，进行丰富、准确、稳定的历史气象资料基础储备。依托现代信息技术，我国历史气象资料的拯救、数字化等研究工作正在全面开展，已经完成了地面、高空、辐射、农气等常用历史气象资料报表的扫描、录入，压、温、湿、降水和风自记纸的扫描，形成了诸多的数字化产品。但由于历史气象资料载体和记录状况的繁杂性、技术发展成熟度制约、数字化人才缺乏等因素<sup>[1]</sup>，类似降水自记迹线的数字化提取也仅仅是其中一种类型的数字化尝试，挖掘隐藏在各种档案材料中的气象观测原始数据，工程量浩大<sup>[4]</sup>，任重道远。数字化成果纳入气象数字档案馆、大数据云平台，应用于数据融合、均一化和数据再分析研

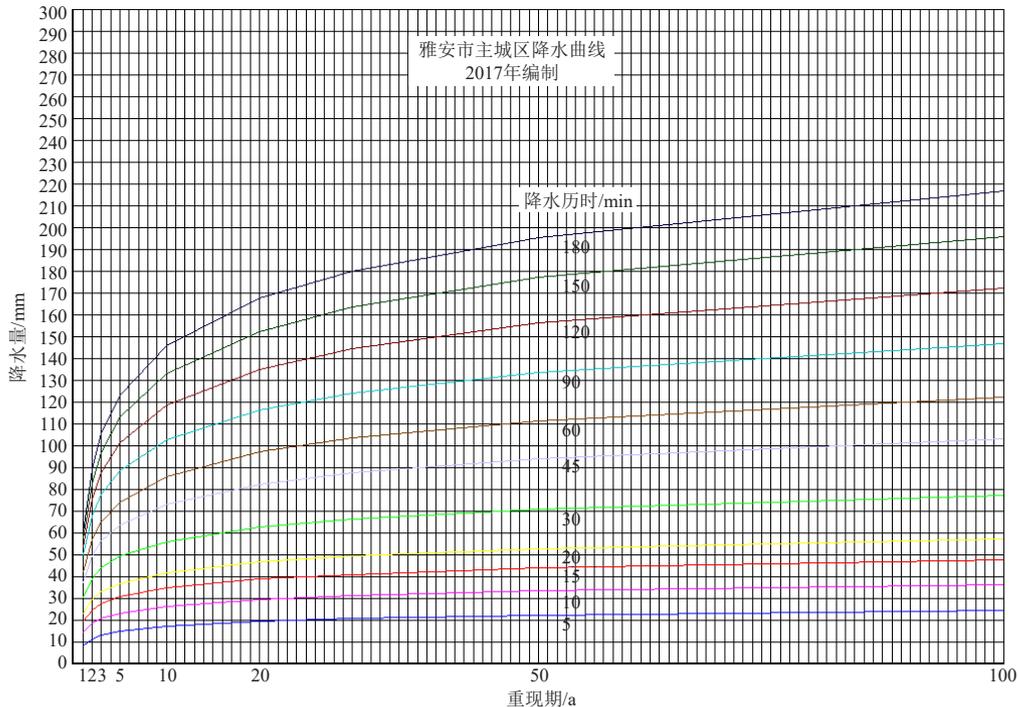


图8 1980—2014年雅安市不同历时的暴雨强度频率曲线

Fig. 8 Rainstorm intensity frequency curve of different durations in Ya'an City from 1980 to 2014

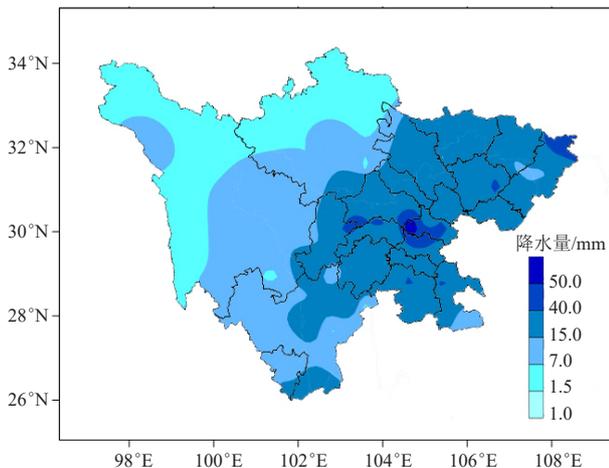


图9 1981年7月四川省最大30 min降水量空间分布图

Fig. 9 Spatial distribution of maximum 30 min precipitation in Sichuan Province in July 1981

究, 服务于气象预报预测、防灾减灾决策服务、气候变化研究, 在广泛的领域得到科学共享, 才能够真正体现这些宝贵的历史气象资料的科学价值。

#### 参考文献

- [1] 范邵华, 余予, 鞠晓慧, 等. 气象资料数字化进展及应用. 气象科技进展, 2018, 8(1): 64-70, 78.
- [2] 韩瑞, 李强, 顾春利, 等. 基于DBnet和众筹策略的气象纸质表格快速数字化方法及系统. 气象科技, 2022, 50(2): 282-289.
- [3] 鞠晓慧, 马楠, 王妍, 等. 基于深度学习的气象资料迹线识别. 科学

- 技术与工程, 2022, 22(21): 9215-9222.
- [4] 贾朋群, 温康民. 早期定量气象观测价值的再开发: 从ACRE说开. 气象科技进展, 2019, 9(6): 122-124
- [5] 中国气象局. 地面气象观测规范. 北京: 气象出版社, 2003.
- [6] 任芝花, 冯明农, 张洪政, 等. 自动与人工观测降雨量的差异及相关性. 应用气象学报, 2007, 18(3): 358-364.
- [7] 王丽吉, 魏爽, 刘熔熔. 浙江省国家地面气象站分钟降水数据集研制. 现代计算机, 2019, (22): 80-83.
- [8] 臧海佳, 吴显中. 气象记录档案数字化工作实践与分析. 中国档案, 2008, (5): 34-36.
- [9] 吕勇平, 王伯民, 吴忠义, 等. 降水自记纸记录数字化 (GB/T 31165-2014). 北京: 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会, 2014.
- [10] 李长军, 黄少平, 陈益玲, 等. 降水自记纸数字化成果质量控制与评估技术. 北京: 气象出版社, 2019.
- [11] 张强, 赵煜飞, 范邵华. 中国国家级气象台站小时降水数据集研制. 暴雨灾害, 2016, 35(2): 182-186.
- [12] Ju X H, Huang S P, Li C J, et al. Development of a self-recording per-minute precipitation dataset for China. Journal of Meteorological Research, 2019, 33(6): 1157-1167.
- [13] 谌穗, 孙立群, 李晴岚, 等. 近38年中国气温和降水的1 km网格数据集. 气象科技, 2021, 49(3): 355-361.
- [14] 刘樱, 马浩, 杨明. 浙江省分钟降水自记纸信息化数据与人工读取数据的对比及适用性分析. 沙漠与绿洲气象, 2019, 13(4): 128-134.
- [15] 高杰, 王慧瑜. 辽宁省降水自记纸数字化成果应用. 安徽农业科学, 2009, 37(33): 16672-16675.
- [16] 雅安市气象局. 雅安市主城区暴雨强度公式编制报告. 雅安: 雅安市住房和城乡建设部, 2021
- [17] 王佳津, 陈朝平, 刘莹, 等. 四川省持续性暴雨定义及时空分布特征. 气象科技, 2017, 45(2): 331-341