

我国地面风自记纸及数字化处理

鞠晓慧

国家气象信息中心

(国家气象科学数据中心)

气象记录档案拯救技术小组

2021.10.21 (贵阳)

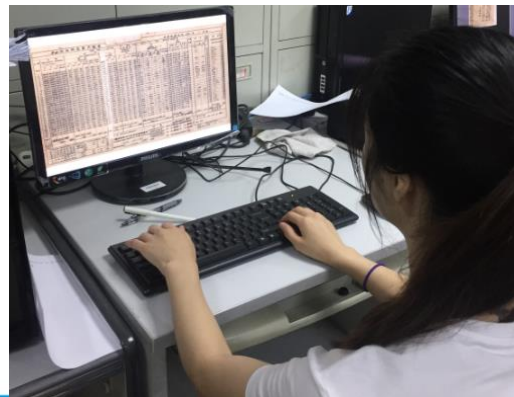
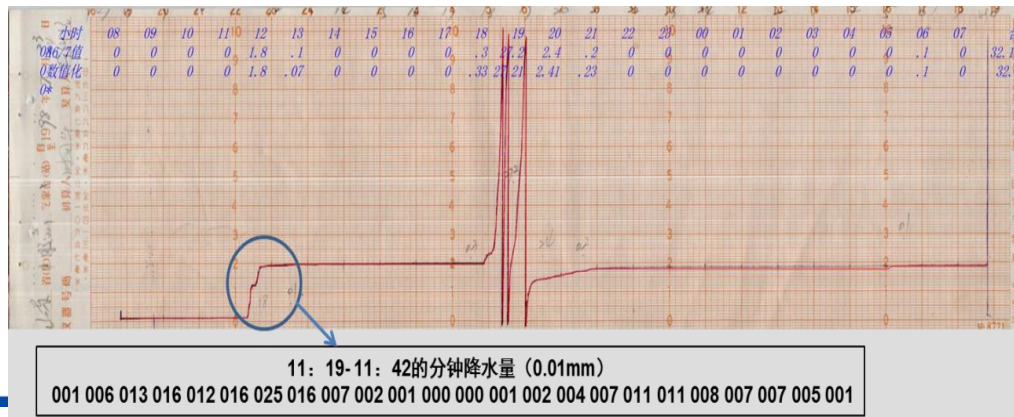
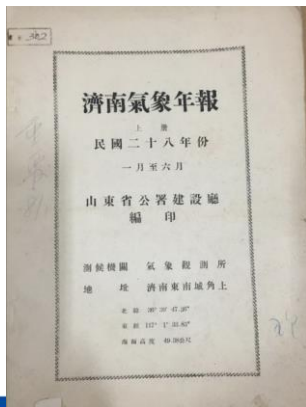


历史气象资料及数字化

人工与自动观测：2000年之前的气象观测以人工观测为主，形成各种纸质记录档案。

历史与实时资料：历史长序列数据集与数据库资料

档案数字化：扫描仪将纸质档案转化成图像的形式；使用计算机键盘录入或计算机软件图像识别将纸质档案上的信息提取出来，转换为计算机可编辑处理、存储在电子介质上的数据文件。通过数字化处理，可以保护原纸质档案原件。 **数字化技术手段：**图像扫描、人工录入、图像字符识别（簿表类档案）和计算机迹线提取（自记纸类档案）。





地面、高空、辐射和农气基础数据产品

基于数字化成果研制的各类数据产品发布在国家气象业务内网和中国气象数据网（外网），提供气象部门内外业务和科研用户使用。

基于地面、高空、辐射基础资料数字化成果研制的数据产品

基础数据产品	中国国家级地面气象站基本气象要素定时/日/旬/月/年值数据产品（V3.0）
	中国高空气象站规定等压面层压温湿定时/旬/月/年值数据集（V2.0）
	中国高空气象站温湿特性层压温湿定时/旬/月/年值数据集（V1.0）
	中国高空气象站规定高度层风速风向定时/旬/月/年值数据集（V1.0）
	中国气象辐射站基本要素逐时/日/旬/月/年值曝辐量数据集（V2.0）
	雾、轻雾和霾现象等雾霾专题数据集（V1.0）
	沙尘暴、扬沙、浮尘等沙尘专题数据集（V1.0）
	雷暴、闪电、大风、冰雹等对流性天气专题数据集（V1.0）
	霜、雾凇、雨凇、结冰等冰冻专题数据集（V1.0）
	暴雨数据集（V1.0）
	高温数据集（V1.0）
	中国国家级地面气象站逐小时降水数据集（V2.0）
	中国国家级地面气象站逐小时 10min 风数据集（V1.0）
气候整编产品	中国 24 城市长年代气温数据集
	中国 24 城市长年代降水数据集
	1981-2010 中国地面气候标准值数据产品
	1981-2010 中国高空气候标准值数据产品
均一化产品	1981-2010 中国辐射气候标准值数据产品
	1981-2010 中国地面气候图集
	中国国家级地面气象站均一化气温数据集（V1.0，2400 余站，日值/月值）
	中国国家级地面气象站均一化降水数据集（V1.0，2400 余站，月值）
	中国国家级地面气象站均一化风速数据集（V1.0，2400 余站，月值）
网格化产品	中国国家级地面气象站均一化日照数据集（V1.0，2400 余站，月值）
	中国国家级地面气象站均一化总辐射数据集（V1.0，98 余站，月值）
	中国地面气温 0.50×0.50 格点数据集（V2.0，日值/月值）
	中国地面降水 0.50×0.50 格点数据集（V2.0，日值/月值）

CONTENTS

01

地面自记风观测及记录

02

地面风自记纸数字化

03

地面分钟风数据质量检测与控制

04

小结与展望

01

地面自记风观测及记录





01 地面自记风观测及记录

测风仪器主要有：

- ✓人工时期：EL型电接风向风速计、达因式风向风速仪、（EN型测风数据处理仪）
- ✓自动时期：单翼风向风速传感器、螺旋桨式风向风速传感器。

要素	2分钟风	10分钟风	最大风	极大风
风速定义	2分钟平均风速	10分钟平均风速	最大10分钟平均风速	瞬时最大风速
风向定义	2分钟最多风向	10分钟最多风向	最大10分钟平均风速的最多风向	极大风速的风向
说明	定时风	自记风		



01 地面自记风观测及记录

EL型电接风向风速计及自记纸

风向： 每2.5分钟由1-2个风向划线组合成瞬时风向。10分钟有5组风向组合，取最多风向。

风速： 测量风程，200米电接一次；10分钟跳动1次，风速是0.3m/s；跳动2次，风速0.7m/s；跳动3次，风速1.0m/s.

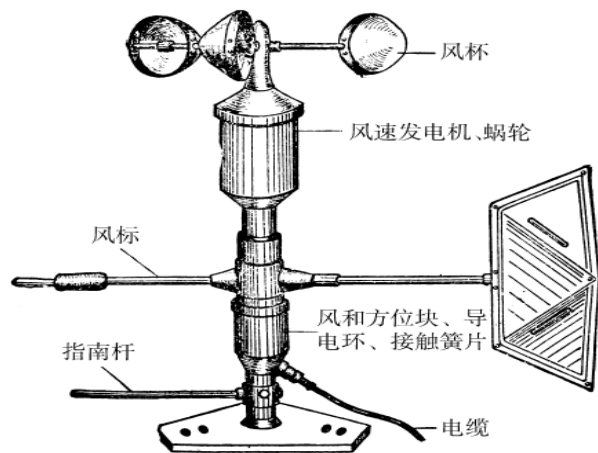
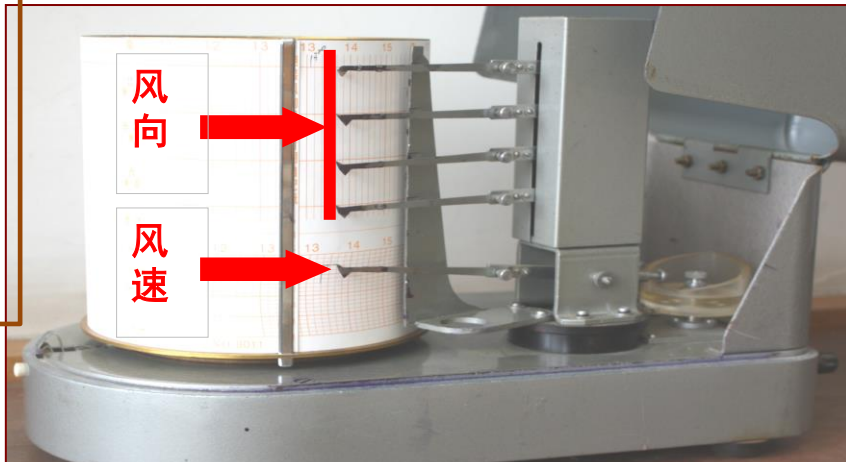
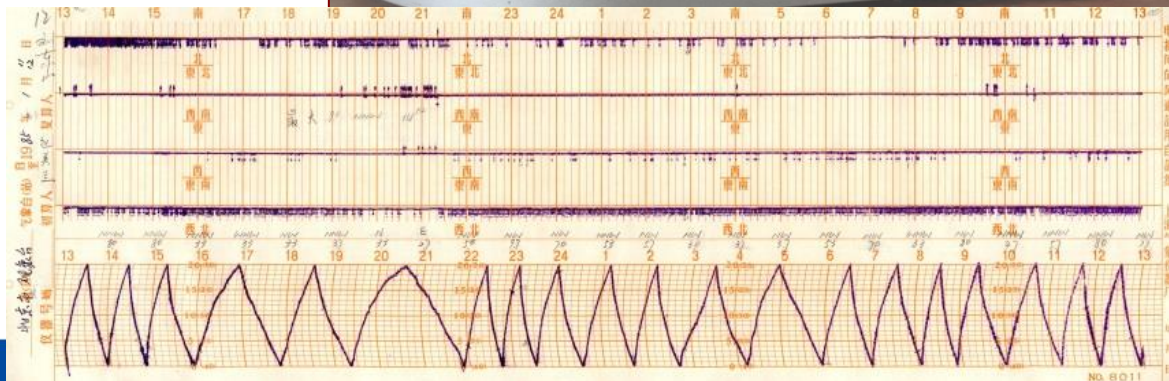


图9-1 EL型风向风速计感应部分





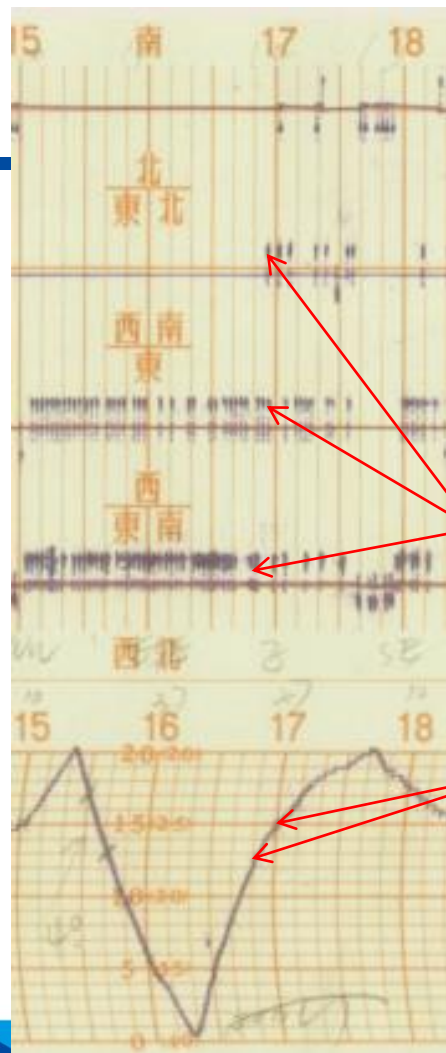
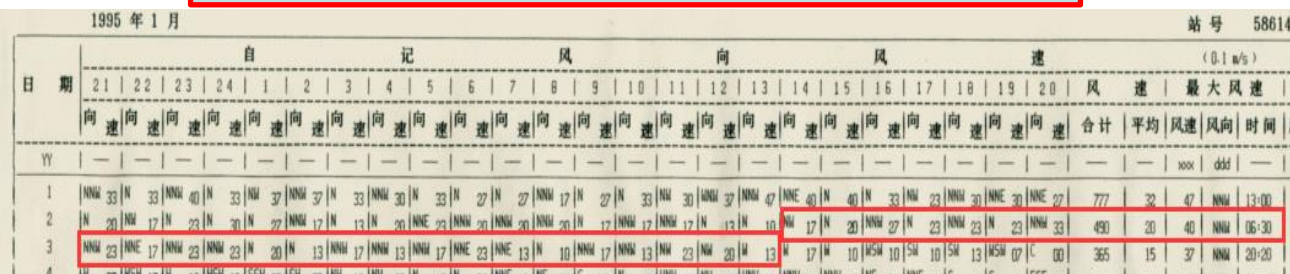
01 地面自记风观测及记录

自记纸记录的曲线信息，观察员如何整理成风向风速数据？

10分钟风向：取10分钟内的5次风向组合次数最多的。如次数最多的2个，去掉最左面的1次风向组合，在剩余的4次风向组合中取最多；。。。。。

10分钟风速：10分钟内迹线通过自记纸上平分格线的格数（1格=1.0m/s,1格=3次跳））计算.

E从自记纸上整理出来的逐小时风向风速



17点风向:
5次风向
ESE,
E,
E,
E,
SE.
取最多

17点风速：
不到3小
格，2.7

02

地面风自记纸数字化





02 地面风自记纸数字化-人工数字化

(1) 人工数字化方式：人工整理、测量和计算逐小时风向风速数据，抄录在气表上；通过人工键盘录入成信息化数据文件。

特点：人工测量个性化差异大；人工整理费时费力，整理到**小时分辨率**。

1995 年 1 月

日期	自 记 风 向													
	21	22	23	24	1	2	3	4	5	6	7	8	9	10
	向	速	向	速	向	速	向	速	向	速	向	速	向	速
YY	—	—	—	—	—	—	—	—	—	—	—	—	—	—
1	NNW 33	N 33	NNW 40	N 33	NNW 37	NNW 37	N 33	NNW 30	N 33	N 27	N 27	NNW 17	N 27	N 33
2	N 20	NNW 17	N 23	N 30	N 27	NNW 17	N 13	N 20	NNE 23	NNW 20	NNW 20	N 17	NNW 17	NNW 17
3	NNW 23	NNE 17	NNW 23	NNW 23	N 20	N 13	NNW 17	NNW 13	NNW 17	NNE 23	NNE 13	N 10	NNW 17	NNW 23
4	NNW 23	NNW 17	NNW 23	NNW 23	N 20	N 13	NNW 17	NNW 13	NNW 17	NNE 23	NNE 13	N 10	NNW 17	NNW 23

A658614.J95 - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

58614 282311616 00332 094 097 1995 01
FZ4

NNW	PPN	NNW	PPN	PNW	NNW	PPN	NNW	PPN	PPN	PPN	NNW
PPN	PPN	PNW	WNW	NNW	NNE	PPN	PPN	PNW	NNW	NNE	NNE
033	033	040	033	037	037	033	030	033	027	027	017
027	033	030	037	047	040	040	033	023	030	030	027
047	NNW	1300.									
PPN	PNW	PPN	PPN	PPN	NNW	PPN	PPN	NNE	NNW	NNW	NNW
PPN	NNW	NNW	PPN	PPN	PNW	PPN	NNW	PPN	NNW	PPN	NNW
020	017	023	030	027	017	013	020	023	020	020	020
017	017	017	013	010	017	020	027	023	023	023	033
040	NNW	0630.									
NNW	NNE	NNW	NNW	PPN	PPN	NNW	NNW	NNW	NNE	NNE	PPN
NNW	NNW	PNW	PNW	PPW	PPW	PPW	WSW	PSW	PSW	WSW	PPC
023	017	023	023	020	013	017	013	017	023	013	010
017	013	023	020	013	017	010	010	010	013	007	000



02 地面风自记纸数字化-计算机迹线提取

(2) 计算机迹线提取：二值化图像识别技术+数据化核心算法

特点：使用专业软件统一识别和提取迹线，提取**分钟分辨率**风向风速数据。

通过颜色空间转换和图像二值化处理，依次得到自记纸网格、风向风速网格、风向基线、风向迹线、风速迹线。

图像处理 核心算法

Canny算法和
Radon变换

颜色空间的转换；
最大类间方差法
(大津法OTSU)
二值化

数据化 核心算法

风向数据化：

- 分别识别4条风向基线和2.5min时间步长的风向线；
- 判断每2.5分钟时刻的风向组合；
- 统计某时间点10min内的5次风向，取最多风向

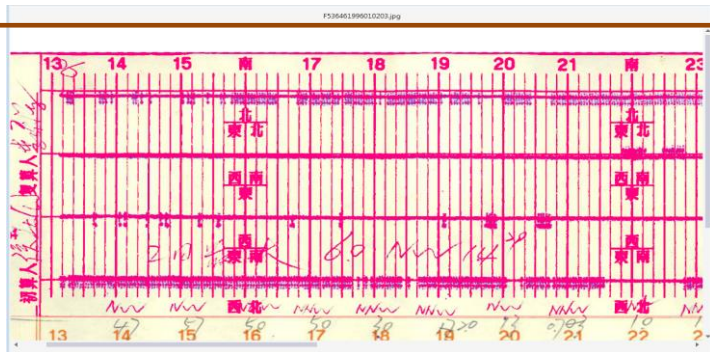
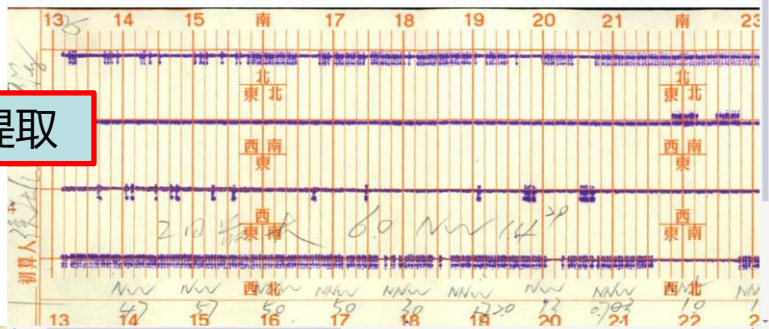
风速数据化：计算某10min的曲线的纵坐标位置差的绝对值，小数位只能是0、3和7



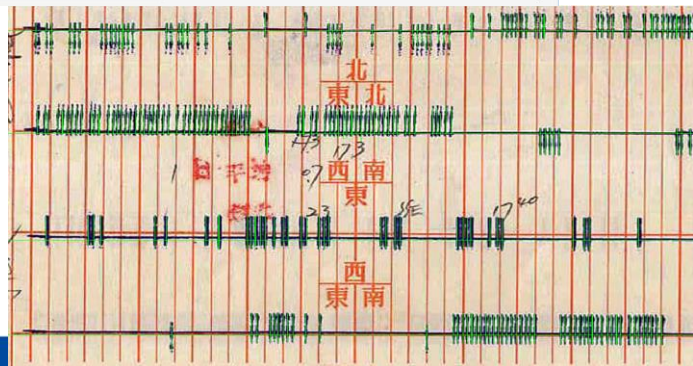
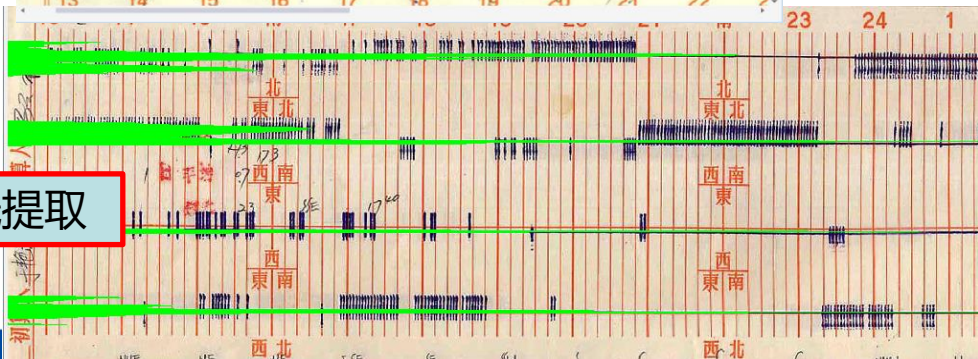
02 地面风自记纸数字化-计算机迹线提取

通过颜色空间转换和图像二值化处理，依次得到自记纸网格、风向风速网格、风向基线、风向迹线、风速迹线。

网格提取



迹线提取

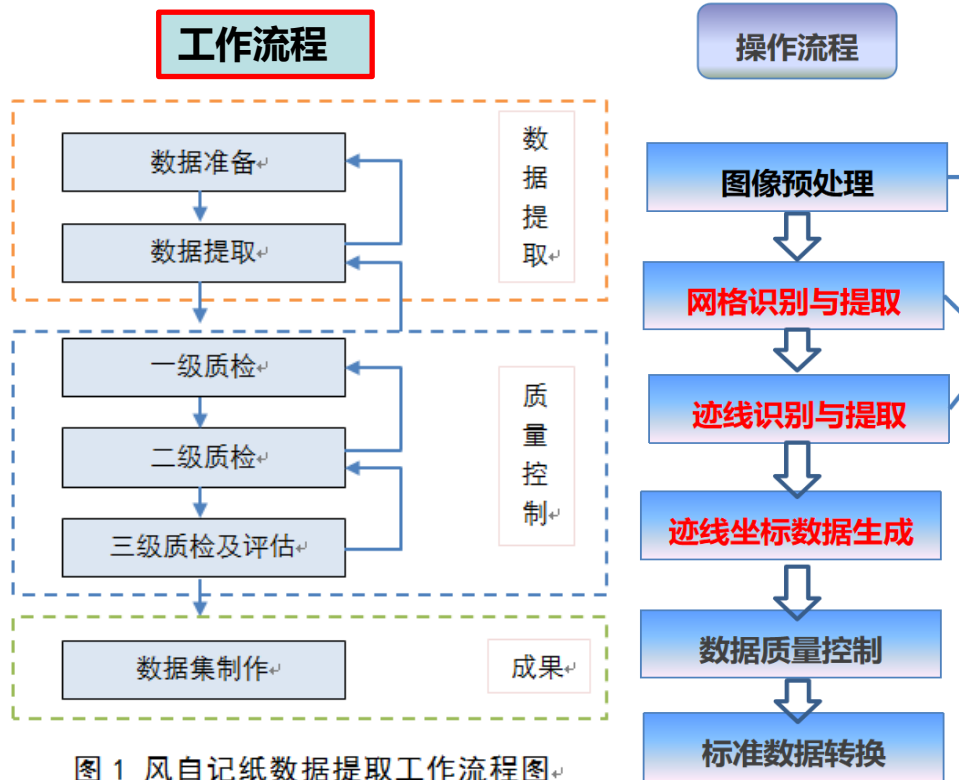




02 地面风自记纸数字化-工作流程

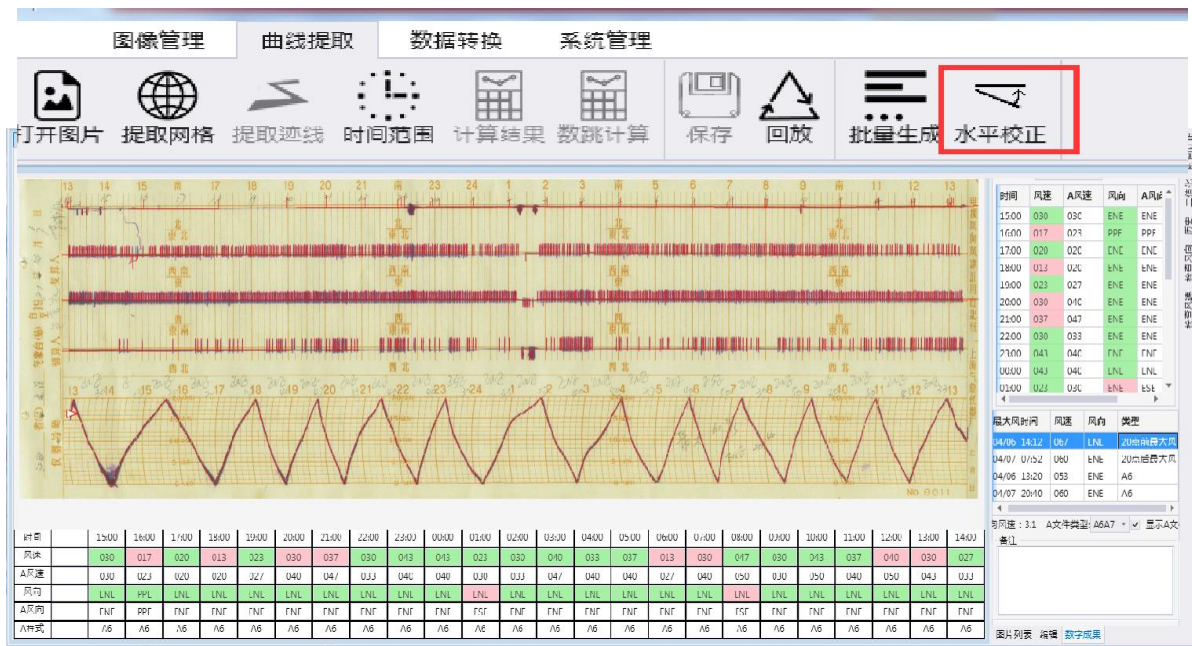
自记纸逐张提取流程

- ✓ **采用国省联动工作模式：**省级负责提取、一级质检、二级质检；**国家**负责三级质检、评估。
- ✓ **全国气象记录档案拯救技术组**研发数字化技术方案，研制数字化提取软件，编制数字化提取技术规定或气象行标；
- ✓ 数字化人员按规范操作软件，通过人机交互，逐张提取，建立逐分钟风向风速。



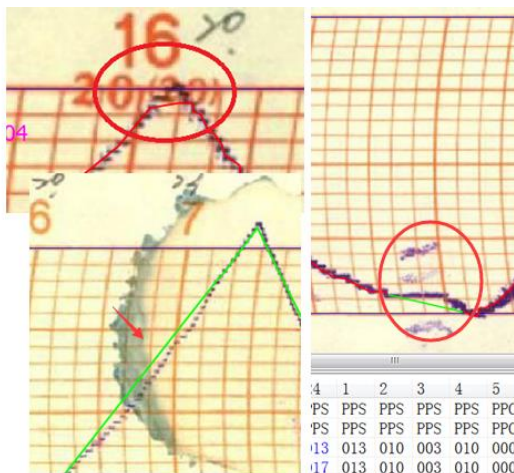
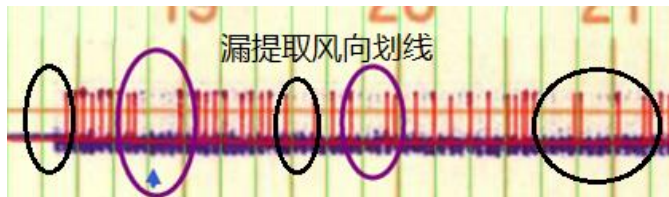
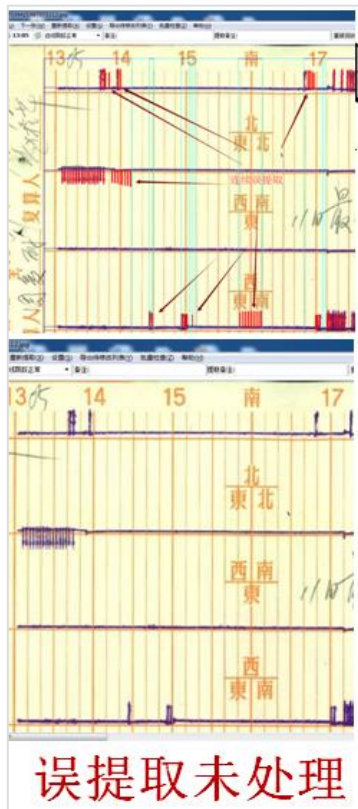


软件提取界面



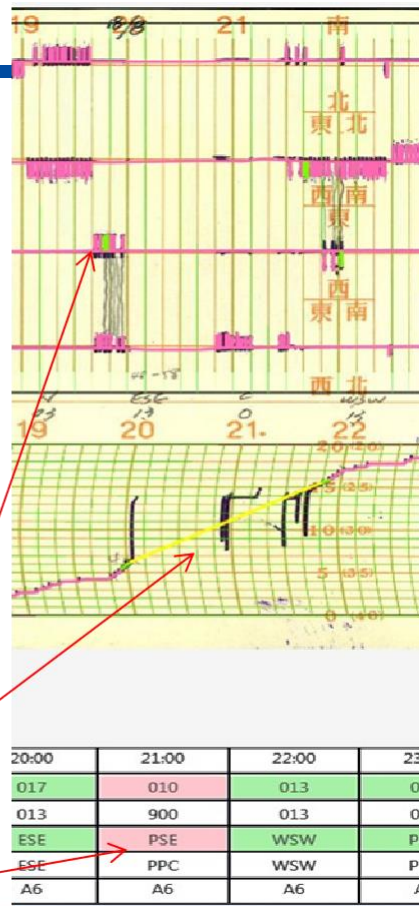


02 地面风自记纸数字化-人机交互



绿线：人工补点
黄线：迹线中断，做异常

与A6小时数据对比
绿色：一致
红色：差异大





02 地面风自记纸数字化-提取数据的质量检查

(1) 提取的风向(速)网格与底图左右边框偏差超过2min、提取的风速上下边框与风速边框偏差超过半跳(0.15 m/s)、图像倾斜或变形的最大偏差大于风速网格1个刻度, 出现一处错误算1.0个错情; ---网格提取要求

(2) 提取风速迹线起(止)位置偏差时间超过2 min或风速超过1/3格、换上(换下)时间错误, 出现一处错误算1.0个错情; ---风速起止跟踪提取要求

(3) 每张纸风向错、漏跟踪风向划线数>30根或连续漏、多跟踪风向划线多于2根, 算1.0个错情; ---风向迹线跟踪提取要求

(4) 风速跟踪线偏离原始迹线外缘1分钟以上或风速跟踪线偏离原始迹线最高、最低点超过半跳(0.15 m/s), 出现一处算0.5个错情; ---风速迹线提取要求

(5) 异常记录处理错误, 错一处算1.0个错情; --异常处理要求

(6) 软件提示的疑误信息未按要求处理的, 每1条算1.0个错情;
---计算机软件检查要求

(7) 比对数据录入错误或影响后续质量检查的修正未备注, 漏备注1处算0.1个错情。---备注信息要求

(8) 其它明显影响提取数据质量或操作失误的, 每出现一处算1.0个错情。---其它



02 地面风自记纸数字化-数字化成果

经过逐站逐张迹线提取，最后形成以台站为单位的数字化成果。

- ✓ EL型电接风自记纸图像文件数据集
- ✓ EL型电接风自记纸曲线文件数据集
- ✓ EL型电接风自记纸分钟风向风速数据集
- ✓ EL型电接风自记纸小时风向风速数据集



02 地面风自记纸数字化-达因风自记纸

工作背景：达因风观测精度是20秒-1分钟，可以记录极大风。研制了基于AI技术的达因风提取软件；创新使用了深度学习技术的迹线跟踪，提高跟踪准确性，减少人工干预，节省约15%-20%的人力成本；

建立部分站点历史上1min风向风速、极大风向风速，强对流天气研究提供数据支撑。

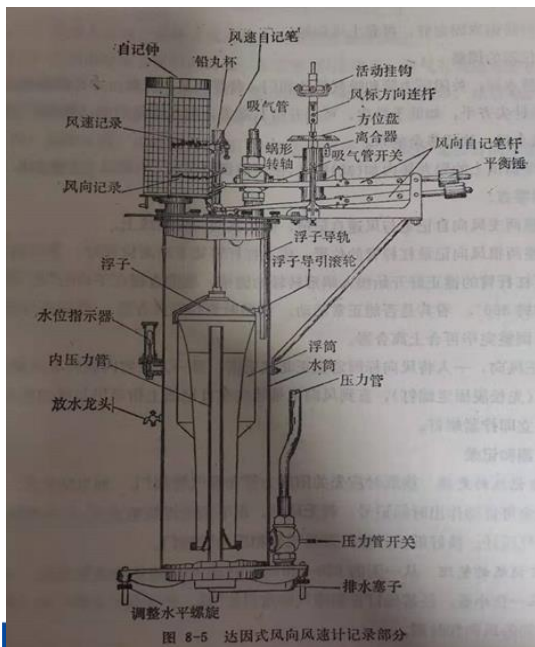
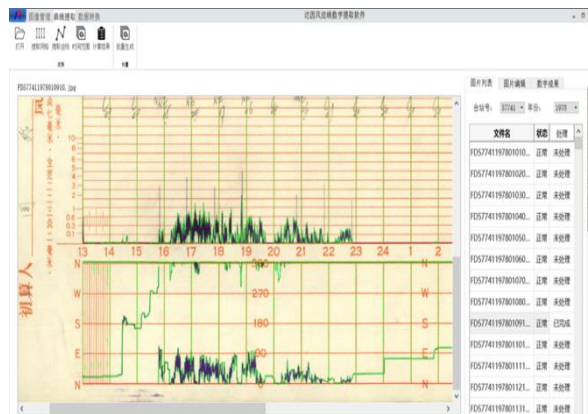
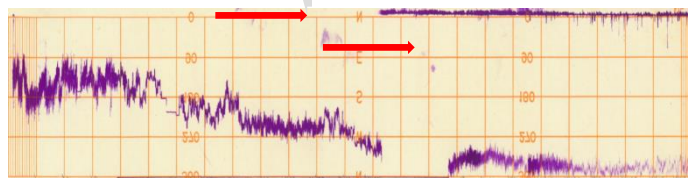


图 8-6 达因式风向风速计记录部分

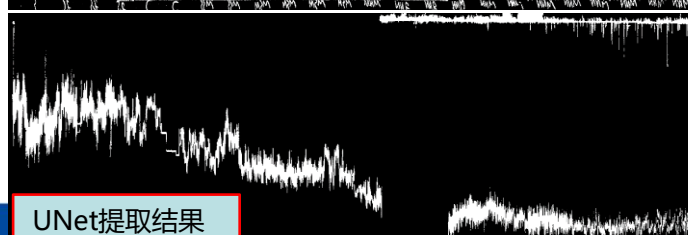


$$Fd = \frac{\sum_{i=x_1}^{x_2} (y_i^u + y_i^d)}{2 \times (x_2 - x_1)}$$

达因风提取软件



自适应阈值提取
结果



UNet提取结果

03

地面分钟风数据质量检测与 控制





03 地面分钟风数据质控-自记纸分钟风数据质量检测

（一）人工观测数据源质控方案

1. 检查对象、内容和对比数据源

本方案检查对象为 EL 风自记纸迹线提取后的分钟风向风速数据文件（Fm）、小时风向风速数据文件（Fh）。

检查内容包括：数据文件格式检查、数据文件完整性检查和数据质量检查。

比对数据有三个，分别为国家气象信息中心发布的“中国国家级地面气象站逐小时 10min 风数据集”（逐时 10min 风向风速数据）、“中国地面基本气象要素定时值数据集（V3.0）”（四次定时 2min 风向风速数据）、“中国地面基本气象要素日值数据集（V3.0）”（日极值）。

3. 检查方法

3.1. 数据文件完整性

根据省级报送的台站起止年月信息，检查是否存在非整年缺少 Fm（Fh）文件、整年缺失情况是否有备注。

3.2. 数据质量检查

3.2.1. 风向格式和界限值检查

- （1）风向应为指定的 17 种字符；
- （2）风速缺测时，风向也应缺测。

3.2.2. 风速格式和界限值检查。

风速在 0~40 m/s（自动站测量范围是 75 m/s）。

3.2.3. 内部一致性检查。

- （1）整点分钟风速、风向=整点小时风速、风向；
- （2） f_m 分钟、 f_h 小时风速 $\leq f_h$ 日最大风速；
- （3） f_h 小时风速、风向与小时风数据集、定时风速、风向比较：
 - 当风速 < 3.0 m/s 时，差异 > 0.4 m/s；当风速 3.0~6.9 m/s 时，差异 > 0.7 m/s；当风速 ≥ 7.0 m/s 时，差异 $> 10\%$ ；
 - 风向差异大于 1 个方位。



03 地面分钟风数据质控-自动站分钟风数据整理

J 文件分钟风质量检查评估方案

1. 检查的内容



本方案主要检查 J 文件分钟风质量，检查使用的比对数据为中国国家级地面气象站逐小时 10min 风数据集；日值 3.0 数据集中定时风（2min 平均）。

2. 检查的指标

2.1. 数据文件完整率

数据完整率=（实有文件数/文件数）×100%

2.2. 数据缺测率

分别计算风向和风速的缺测率。

风向缺测率=（实有风向个数/应有风向个数）×100%

风速缺测率=（实有风速个数/应有风速个数）×100%

2.3. 风速界限值检查

根据《气象资料业务系统（MDOS）省级地面气象资料质量控制技术规程》，风速界限值在 0~100 米/秒之间。后期可根据逐站逐月设定风速界限值。



03 地面分钟风数据质控-质控方法和流程

(1) 气候学界限值检验

$0 \leq \text{风速} < 75 \text{ m/s}$

$1 \leq \text{风向} \leq 17$

(2) 内部一致性检查

a 与其它数据源的比对。当比对数据风速 $< 5.0 \text{ m/s}$ 时，差异 $\leq 1 \text{ m/s}$ ；当比对数据风速 $\geq 5.0 \text{ m/s}$ 时，差异 $\leq 20\%$ ；b 风速风向的一致。

风速缺测，风向也缺测；风速 $\leq 0.2 \text{ m/s}$ ，风向为静风(17)。

(3) 10 分钟风向风速时间一致性检查：

a 连续无变化

针对风速：连续 5 小时以上非静风风速不变时，全部连续数据判为可疑。

b 快速变化

针对风速：10 分钟风速快速变化时，快速变化的数据判为可疑。判定标准：

$|V_i - V_{i-1}| \leq 20 \text{ m/s}$ ，若不符合以上关系，则可以判定该分风速数据可疑。

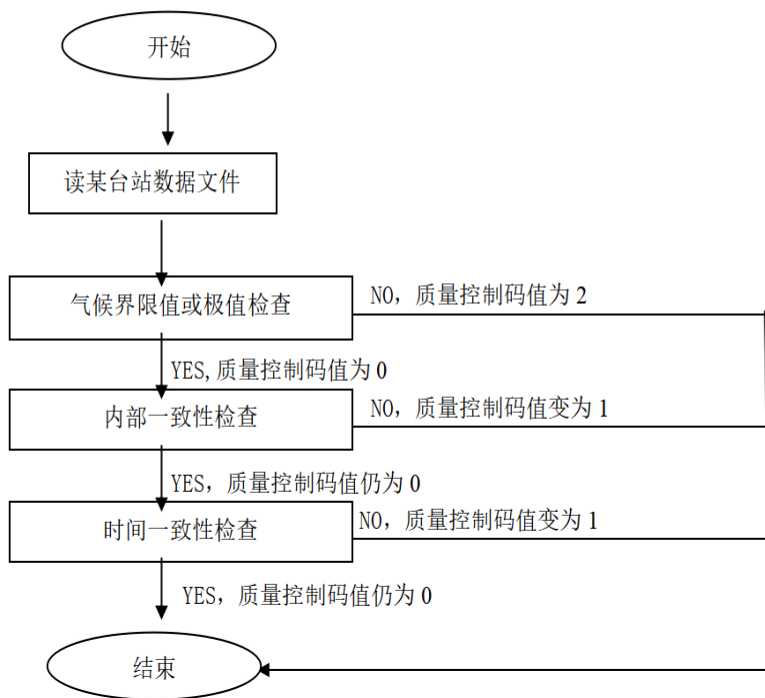
(4) 小时风向风速时间一致性检查：

针对风速：连续 N 个及以上风速不变时，全部连续数据判为可疑。判定连续无变化的标准：

$V_i > 0$, $\text{stdev}(V_i) = 0.0$, $i = [0.8]$ 或者 $V_i > 0$, $i = [0.10]$ 或者 $\text{stdev}(V_i) < 0.2$, $i = [0.10]$,

V_i 为风速， stdev 为标准差

数据集质量控制流程图：





04

小结与展望





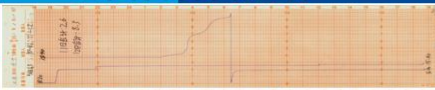
04 小结与展望

- ✓ 全国1951年以来到自动站运行共有2102站2242万页EL风自记纸，300多站80万页的达因风自记纸。
- ✓ 已经完成1212站1123万页的EL风自记纸迹线数据提取。2021年开始达因风自记纸数字化扫描和迹线提取。
- ✓ 十四五期间拟开展气压、气温和湿度自记纸迹线提取。实现地面5要素1951年以来的分钟分辨率历史长序列。（降水自记纸2018年已完成）

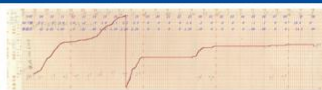


地面各种自记纸

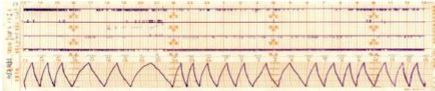
降水自记纸-虹吸式



降水自记纸-翻斗式



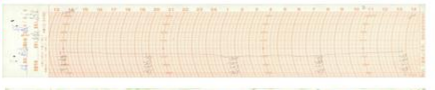
EL型风自记纸



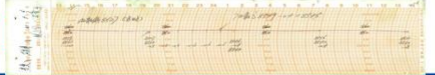
相对湿度自记纸



气温自记纸



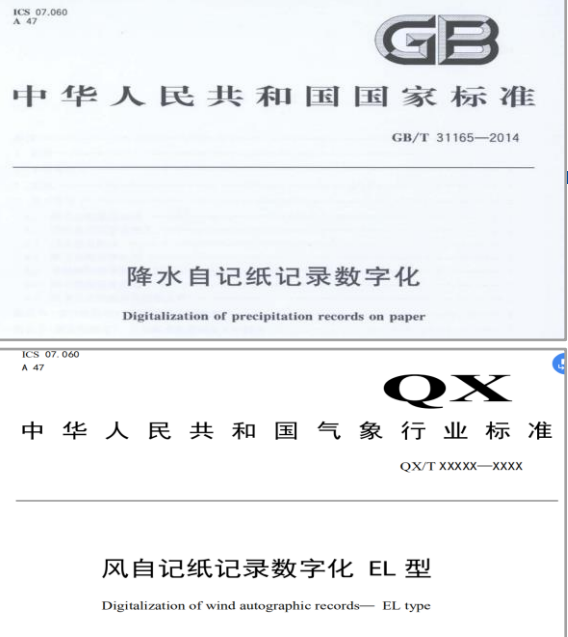
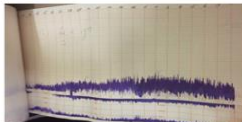
气压自记纸



EN型风自记纸



达因风自记纸



感谢聆听，请批评指正！