

任芝花,张志富,孙超,等. 2015. 全国自动气象站实时观测资料三级质量控制系统研制. 气象, 41(10):1268-1277.

全国自动气象站实时观测资料三级质量 控制系统研制^{*}

任芝花¹ 张志富¹ 孙 超¹ 刘一鸣¹ 李 俊¹ 鞠晓慧¹ 赵煜飞¹
李志鹏² 张 玮² 李洪康² 曾行吉³ 任晓炜³ 刘 莹⁴ 王海军⁴

1 国家气象信息中心, 北京 100081

2 江西省气象信息中心, 南昌 330046

3 广西区气象信息中心, 南宁 530022

4 湖北省气象信息与技术保障中心, 武汉 430074

提 要: 迄今, 中国气象局已在全国建立了 55000 余个地面自动气象站(含国家级自动站和区域自动站)。为了促进众多台站观测资料质量提高, 在实时业务中保障用户用到基本正确有效的数据, 2009 年底中国气象局启动了“全国自动站实时资料质量控制与综合评估系统建设”工程, 旨在研制一套基本适用于全国自动站实时观测数据的质量控制技术方法, 建立台站、省级、国家级资料部门的地面自动站实时资料三级质量控制与反馈业务系统。在探究自动站错误数据表现形式的基础上, 研制了自动站实时观测数据质量控制技术; 从业务分工与应用角度, 设计建立了台站负责质量监控、省级负责质量控制和国家级主要负责质量评估的自动站实时资料三级质量控制系统。该系统已在 2400 多个国家级气象台站、31 个省级和国家级资料部门安装部署与业务应用。通过系统建设与应用, 实时上传的自动站数据质量得到明显改善。主汛期全国自动站逐小时气温、降水数据可用率分别从 2009 年的 88%、83% 提升到 2012、2013 年的 98% 左右。

关键词: 自动站, 实时观测数据, 三级, 质量控制, 质量评估, 质量监控, 质量反馈

中图分类号: P413

文献标志码: A

doi: 10. 7519/j. issn. 1000-0526. 2015. 10. 010

Development of Three-Step Quality Control System of Real-Time Observation Data from AWS in China

REN Zhihua¹ ZHANG Zhifu¹ SUN Chao¹ LIU Yiming¹ LI Jun¹ JU Xiaohui¹
ZHAO Yufei¹ LI Zhipeng² ZHANG Wei² LI Hongkang² ZENG Xingji³
REN Xiaowei³ LIU Ying⁴ WANG Haijun⁴

1 National Meteorological Information Centre, Beijing 100081

2 Jiangxi Meteorological Information Centre, Nanchang 330046

3 Guangxi Meteorological Information Centre, Nanning 530022

4 Hubei Meteorological Information and Technical Support Centre, Wuhan 430074

Abstract: China Meteorological Administration (CMA) commenced building surface automatic weather station (AWS) around 2000. More than 2400 national AWSs have been built to observe precipitation, air temperature, wind direction, wind speed, air pressure, air humidity and soil temperature, etc. Moreover, more than 55000 regional AWSs that are mainly applied to disaster prevention and reduction have been also set up to monitor precipitation, air temperature, wind direction, wind speed, air pressure or humidity.

^{*} 公益性行业(气象)科研专项(GYHY201106038)和中国气象局业务建设项目“全国自动站实时资料质量控制与综合评估系统建设”共同资助

2014 年 8 月 26 日收稿; 2015 年 7 月 8 日收修定稿

第一作者: 任芝花, 主要从事气象数据质量控制和分析评估工作. Email: rzh@cma.gov

The data quality of the above large number of AWSs could not reach very correct degree. So, at the end of 2009 CMA launched a project named Construction of Quality Control and Comprehensive Evaluation System for Real-time Observation Data of AWS. (1) After analyzing various performance ways and causes of the erroneous meteorological data from AWS, the quality control technique was developed for real-time observed and transmitted hourly data of precipitation, air temperature, wind direction, wind speed, air pressure and air humidity. The quality control method includes climatic range check, internal consistency check, temporal consistency check, spatial consistency check, man-computer interaction control, etc. Each kind of inspection process contains a number of control modes. (2) Based on the operational process of AWS observation data acquisition, transmission, collection, decode and real-time service, three-step quality control system of AWS real-time observation data was developed. In the system, the national-level system is mainly responsible for data quality evaluation, the provincial one including 31 provincial data management departments is in charge of data quality control, and more than 2400 national stations are responsible for data quality monitoring and erroneous data correction. The three-step quality control system operates based on national and provincial real-time databases, organized mainly by one real-time quality control software, man-compute interaction quality monitoring platform, data quality evaluation platform, nation-province-station three-step quality automatic feedback process for suspect data information. (3) The three-step quality control system of AWS real-time observation data has been installed and operational applied at more than 2400 national AWSs, 31 provincial meteorological information departments and one national meteorological information center. Quality control, quality monitoring, quality feedback and quality evaluation are operated for meteorological data of real-time observation and transmission from above national and regional AWSs. The quality of all AWS observation data is greatly improved year by year. For example, data availability of hourly air temperature and precipitation real-time observed by above AWSs in main flood season increases from 88%, 83% respectively in 2009 to about 98% in 2012 and 2013.

Key words: automatic weather station (AWS), real-time observation data, three-step, quality control, quality evaluation, quality monitoring, quality feedback

引 言

气象数据质量控制(QC)是质量管理体系中最为重要的组成部分,它包括在气象台站和数据中心的资料审核,以便发现资料中的缺测数据和错误数据而对其补充或修正,从而保证资料最大程度的完整和最大可能的准确。为保证达到这一目的,设计好的质量控制系统包括质量控制技术是必需的。一套质量控制系统应该包含将错误资料返回源地核实,并防止错误再次发生的程序(WMO,2008)。在资料提供给用户前,对其进行质量检查,是提高资料的质量,保证资料发挥高效益的关键环节(任芝花等,2006)。

国内外针对地面气象资料质量控制技术的研究与应用主要包括气候学界限值检查、区域界限值检查、要素间内部一致性检查、时间一致性检查和空间一致性检查等(任芝花等,2005;2006;2007b;刘小宁

等,2005;陶士伟等,2007;2009;王海军等,2007;2012;中国气象局,2010;李亚丽等,2015;Sciuto et al,2009;Zahumensk,2004;Peterson et al,1998;Feng et al,2004;Vejen et al,2002;WMO,1992)。上述技术可用于观测台站、数据中心。质量控制适用于实时态,但也可非实时运作,例如定期的质量控制。随着地面观测自动化技术的发展,西方发达国家纷纷研制了地面自动观测资料质量控制系统。北欧五国(丹麦、芬兰、冰岛、挪威、瑞典)对自动站资料控制流程分4级:台站级资料质量控制、入库前实时资料质量控制、入库后非实时资料质量控制以及人工审核(熊安元,2003)。美国由1100多个观测站组成的地面自动观测系统,在业务运行中,通过对资料进行台站级、州级、国家级三级质量控制,及时发现并反馈观测资料的质量问题,保证资料的准确性(周林等,2002)。截至2005年底,中国气象局已建立了1800多个国家级地面自动气象站,并将自动站每月所有

观测数据及相关信息以单站月报数据文件(中国气象局,2005)的形式滞后 1 至 2 个月进行传输与归档。针对上述自动站整月观测数据,国家气象信息中心联合安徽省气象局、湖北省气象局研制了图 1 所示的由台站,到省级、国家级资料部门的地面自动

站非实时资料三级质量控制业务系统(任芝花等,2007a),并于 2007 年开始业务应用。该系统中各级质量控制融合了自动控制技术和交互式应用技术,提高了自动进程能力,允许在必要时对疑误数据进行详细的人工分析、判断与修正。

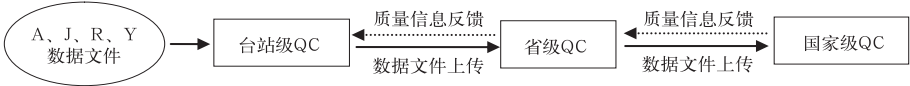


图 1 地面自动站非实时资料三级质量控制业务系统

Fig. 1 Operational system of three-step quality control of AWS non-real-time observation data

至 2009 年,全国已建地面自动站(含国家级自动站和区域自动站,下同)30000 余个,迄今已 55000 多个。自动站观测资料虽然 2006 年就开始实时上传,且台站及数据量在不断增加,但是至 2009 年,尚未建立针对自动站实时上传资料的质量控制业务,严重阻碍了该资料的科研业务应用。2009 年底中国气象局成立了由国家气象信息中心、江西、广西和湖北气象局等单位有关技术人员组成的全国自动站实时资料质量控制与综合评估系统研发团队,旨在研制一套基本适用于全国自动站实时资料的质量控制技术方法,建立台站、省级、国家级资料部门的地面自动站实时资料三级质量控制与反馈业务系统,以提高自动站观测数据的质量,在实时业务中保障用户用到基本正确的有效数据。本文从质量控制技术研制、系统功能设计与应用等角度进行阐述。

1 质量控制对象

质量控制对象为国家级自动站和区域自动站实时观测与上传的逐小时正点气温、气压、湿度、风向风速和降水要素数据。各要素数据内容如下:(1) 气温:正点气温、小时内最高气温和最低气温;(2) 气压:正点本站气压、小时内最高本站气压和最低本站气压;(3) 空气湿度:正点相对湿度、小时内最小相对湿度;(4) 风向风速:正点 2 min 平均风速风向、正点 10 min 平均风速风向、小时内最大风速风向、小时内极大风速风向;(5) 降水:正点前 1 h 雨量。

2 质量控制技术

2.1 技术思路

质量控制的目的是尽可能将隐藏在数据集中的

错误数据得到有效检测,保障用户用到基本正确可靠的资料。假设提前知道错误数据表现形式,针对不同表现形式的错误数据采取相应的检测技术,其检测效果则显而易见。因此质量控制技术方案的研制必须深入数据内部,了解数据本身的规律以及各种原因引起的错误数据的表现形式。自动站小时资料质量控制技术的研制是基于上述认识,基于全国气象部门区域自动站及国家级自动站运行后实时上传的所有小时资料开展的。

通过对 2006 年以来实时上传的地面自动站逐小时观测的温、压、湿、风、雨量数据的应用分析,发现资料中错误数据表现形式如下:(1) 普遍以大值、小值形式存在的奇异数据;(2) 数据长时间连续无变化或微变;(3) 小时数据时间变率过大;(4) 正点数据与相应小时内测量的极值不一致;(5) 数据空间分布明显异常等现象。

在分析错误数据表现形式及质量控制技术调研基础上,确定了图2所示的全国自动站温、压、湿、

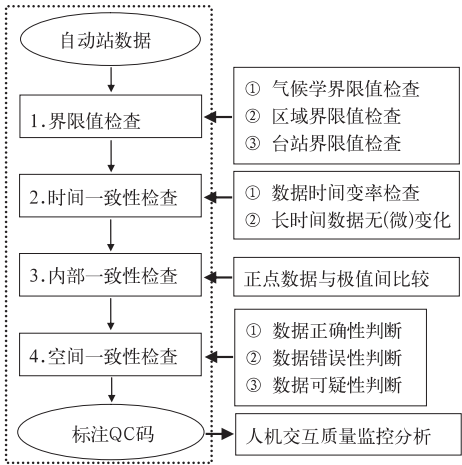


图 2 自动站数据质量控制技术基本思路

Fig. 2 Basic thought on quality control technology of AWS data

风、雨实时资料质量控制技术的基本思路,包括界限值检查、时间一致性检查、内部一致性检查、空间一致性检查等,每种检查过程又包含多项控制方式。图 2 所示的质量控制技术均可通过软件自动完成,但是自动质量控制后,仍有一部分可疑数据无法明确其对与错,该部分数据需要有经验的质量分析方面的专家,借助人机交互质量监控界面提供的多种信息进一步分析判断。

2.2 质量控制流程

经质量控制的每个观测数据应给出表征该数据质量状况的信息,以便数据处理者和用户参考。实际应用中,用质量控制码(QC 码)表示数据的质量状况,QC 码的规定(中国气象局,2005;2010)如下:

- 数据正确: QC 码=0
- 数据可疑: QC 码=1
- 数据错误: QC 码=2
- 无观测数据: QC 码=8
- 数据未作质量控制: QC 码=9

本质量控制为顺序控制。质量控制过程中,按照图 3 流程进行质量控制,首先设置数据的初始 QC 码=0,当无观测数据或数据缺测时 QC 码=8,

然后依次进行质量检查。图 3 中,中间箭头表示被检数据通过相应 QC 模块检查,左右两侧的箭头代表未通过检查。气候界限值检查、区域界限值检查、时间一致性检查和空间一致性检查过程均可自动检测出部分错误数据,当某个数据被判为无观测或错误时,直接自动标注该数据的最终 QC 码,该数据不再需要进行后续的自动检测。当检测过程中,发现某数据可疑时,先暂时标注该数据的 QC 码=1。如图 3 所示,QC 码=0 或 1 的数据,需要继续后续的检测,直到被检测为错误或空间一致性检查完成为止。

经自动质量控制后的数据均有质量控制码,通过后文图 7 所示的人机交互质量监控系统,对疑误数据质量进一步分析辨别,错误或缺测数据可通过发送更正报进行纠正,最后生成数据最终质量信息。

2.3 技术方法

本节中涉及到的质量控制参数是基于中国 1951 年以来的 2400 多个国家站和 2006 年以来的区域自动站长期观测的小时资料、日极值分析而得(鞠晓慧等,2010;任芝花等,2010;张志富等,2013;赵煜飞等,2011),并在实际业务中得到检验与应用。

2.3.1 界限值检查

界限值检查包括气候学界限值或值域范围检查、区域界限值检查、台站界限值检查。

(1) 气候学界限值和值域范围检查:从气候学角度上不可能出现的气象要素临界值称之为要素的气候界限值;另外数据形成时规定采用的表示方式和范围,为要素的值域范围。超出要素的气候界限值或值域范围的数据为错误数据。本系统中各要素气候界限值或值域范围见表 1。

表 1 气候学界限值或值域范围	
Table 1 Climatic range or numerical range	
要素	范围
气温	-55~40℃(1—2 月、10—12 月)
	-50~45℃(3—4 月)
	-25~50℃(5—9 月)
本站气压	520~1085 hPa
相对湿度	0~100%
风速	0~65 m·s ⁻¹
风向	0°~360°
1 h 雨量	0~150 mm

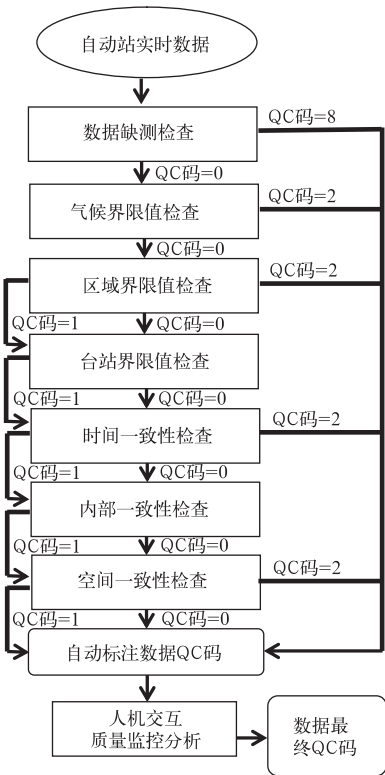


图 3 质量控制流程
Fig. 3 Procedure of quality control

(2) 区域界限值检查:根据要素空间分布特征,全国划分若干个区域,对各区域分别设定界限值范

围。根据界限值设置的宽严程度,规定超越该界限值范围的数据为错误或可疑。进行该项检查的要素有雨量、气温、气压、风速。如针对小时降雨量,全国划分 6 个区域,对各区域分别制定图 4 所示的上限值,大于该上限值的数据为错误(任芝花等,2010);而本站气压的大小与台站海拔高度最为密切,在海拔高度 5000 m 内以 50 m 为间隔,研制了不同高度界限值范围(表 2),本站气压数据超越对应高度界限值范围则为错误数据。

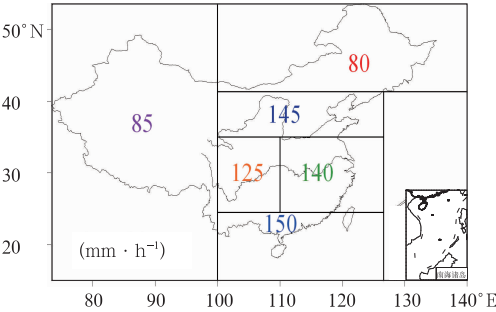


图 4 各区域小时雨量上限值
Fig. 4 The upper limit of hourly rainfall amount at various area

表 2 各高度本站气压界限值范围

Table 2 Limits of air pressure at stations with various altitudes

海拔高度/m	界限值范围/hPa
$H<0$	955~1085
$0\leq H<50$	945~1075
$50\leq H<100$	935~1075
$100\leq H<150$	935~1065
\vdots	\vdots
$4800\leq H<4850$	545~595
$H\geq 4850$	535~585

(3) 台站界限值检查:气象要素值(x)是否在本站要素特定的变化范围 $[x_1, x_2]$ 内的检查为台站界限值检查,即要求 $x\in[x_1, x_2]$ 。超出测站要素变化范围的观测资料为可疑资料,应做进一步检查,以判断资料正确与否。要素变化范围可用测站历史记录中曾出现过的最小值和最大值或其他出现概率很小的最小值和最大值代替。对于观测历史较长的台站,本系统中均给出了每个站各月气温、气压和风速的界限值范围。建站较短的台站,包括区域站可参考邻近的国家站界限值范围进行检查。由于气温具有非常明显的日变化,因此研制给出了气温各月各定时(00:00、06:00、12:00、18:00 UTC)台站界限值

范围。在质量控制过程中,介入上述 4 个时次之间的任何时次的气温界限值取其临近两个定时观测时次涵盖的最大极值范围作为该时次的极值范围,如:01:00、02:00、03:00、04:00、05:00 UTC 的台站界限值范围中,最大值(最小值)为 00:00 和 06:00 UTC 2 个时次的最大值(最小值)。

2.3.2 时间一致性检查

气象要素随时间的变化具有一定的规律。对气象数据变化是否符合这种规律的检查称之为时间一致性的检查。本系统中时间一致性检查包括:数据持续无变或微变检查,当前时次与上个时次数据间变率检查。气温、气压、湿度在时间上均属于连续性要素,因此上述 3 个要素数据均进行连续多个时次数据无变化检查,当前时次与上个时次数据间变率检查。对气温而言,全国划分 4 个区域(图 5),各区域分别设置最长容许持续无变化的时次界限 N (张志富等,2013),若区域中某站连续 N 个以上时次正点气温、最高气温、最低气温均无变化,则所有无变化的气温数据均判为错误;若当前时次气温与上个时次相差 8°C 以上时,则当前时次气温数据作为可疑数据,需作进一步检查。

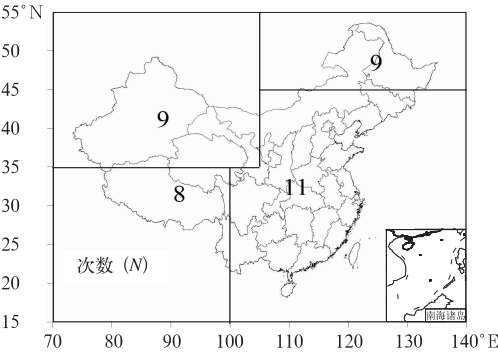


图 5 各区域小时气温数据最长持续无变化时次
Fig. 5 The longest data sequence of invariant hourly air temperature in various areas

降水、风向风速在时间分布上均属于离散型要素,因此仅进行连续多个时次数据无变化或微变检查。考虑到上述离散型要素不同量级,其无变化可持续的时间不同,按照要素数据大小不同,对其进行无变化(微变)检查所允许的持续时间不同。如针对小时降水数据的质量控制,按照降水量属于 $(0, 0.5)$ mm、 $[0.5, 1.0)$ mm、 ≥ 1.0 mm 3 个级别,进行如下时间一致性检查(任芝花等,2010): (1) 连续 $N1$ 个及以上小时降水量(简称 R , 下同)相等且 $R\in(0, 0.5)$ mm 时,则相应数据均判为错误; (2)

连续 N_2 个及以上小时降水量相等且 $R \in [0.5, 1.0)$ mm 时,则相应数据均判为错误;(3) 连续 N_3 个及以上小时降水量相等且 $R \geq 1.0$ mm 时,则相应数据均判为错误。其中 N_1 、 N_2 和 N_3 取值分别为 15、10 和 6。

而针对风向风速的检查则根据 2 min 风速数据大小分级设计了 17 项控制规则,包括因雷击或冰冻

等原因造成的风向长时间变化微弱现象的检查,因篇幅所限,在此不细述。如图 6 所示,C3006 站 2011 年 6 月尽管平均风速普遍在 $4 \text{ m} \cdot \text{s}^{-1}$ 左右,但整月其各项风向数据变化均在 5° 以内,由此判断各项风向数据全月异常,造成该现象的原因为风向标转动不灵或风向传感器遭受雷击。

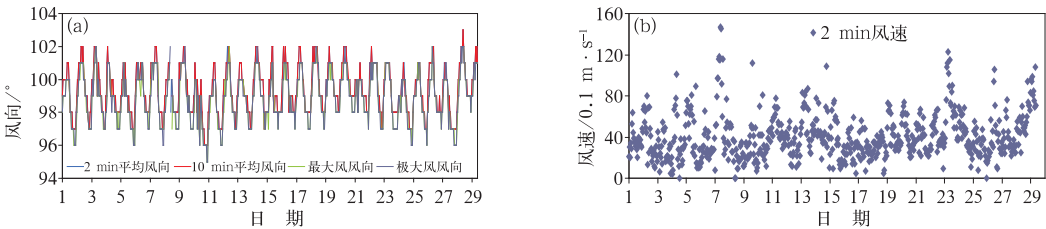


图 6 C3006 站 2011 年 6 月逐小时(a) 风向,(b) 风速的风数据
Fig. 6 Hourly wind data observed from Station C3006 in June 2011
(a) wind direction, (b) wind speed

2.3.3 内部一致性检查

对固定测站不同要素或项目之间是否符合某种物理联系的检查称之为内部一致性检查。目前系统涉及到的内部一致性检查仅为要素正点观测数据与小时内极值间的比较,如:正点气温 ≥ 1 h 内最低气温;1 h 内最高气温 \geq 正点气温;1 h 内最高气温 $>$ 小时最低气温。除了气温外,还有气压、风速、相对湿度进行该项检查。

2.3.4 空间一致性检查

空间一致性检查是根据气象参数具有一定的空间分布特点而进行的检查。其有效性取决于观测站网密度和被检参数与空间的相关程度。通常利用与被检站邻近的台站同一时间观测的气象要素值进行比较;或利用邻近测站观测值计算被检站的估计值,观测值与估计值进行比较检查。考虑到天气系统的移动与渐变,在对被检站当前时次数据进行空间一致性检查时,参与检查的邻近站数据除了当前时次数据外,还应用到上个时次观测数据(任芝花等,2010)。邻近站的选取考虑了与被建站间的相对高度、水平距离及方位等因素(何志军等,2010;任芝花等,2010)。考虑到实时业务的时效性要求,空间一致性检查方法力求简单有效,本系统用的主要方法如下:

(1) 空间平均值比较法

即被检站数据与选取的 N 个邻近站数据的平均值之差值超过规定的阈值,则被检站数据错误。

(2) Madsen-Allerupt 方法(Vejen et al,2002)

Madsen-Allerupt 方法是对周围若干邻近站同一时刻某要素观测值进行由小到大排序,利用排序后的 $\frac{1}{4}$ 、 $\frac{2}{4}$ 、 $\frac{3}{4}$ 分位值,计算统计量 T_u :

$$T_u = (x_u - q_{t,2/4}) / (q_{t,3/4} - q_{t,1/4})$$

式中, x_u 为被检站 t 时刻某要素观测值, $q_{t,1/4}$ 、 $q_{t,2/4}$ 、 $q_{t,3/4}$ 为邻近站 t 时刻该要素观测值进行由小到大排序后的 $\frac{1}{4}$ 、 $\frac{2}{4}$ 、 $\frac{3}{4}$ 分位值。

当 $|T_u|$ 超过规定值时,认为被检站 t 时刻该要素观测值可疑。

(3) 空间极值比较法(任芝花等,2010)

由 N 个邻近站要素值形成数据序列,计算序列的最大值(VALUE_MAX)和最小值(VALUE_MIN)。当被检站数据 R 不符合下列原则且 $N \geq 3$ 时,被检数据可疑:

$$a1 \times \text{VALUE_MAX} \geq R \geq a2 \times \text{VALUE_MIN} \quad (a1 > 1, a2 < 1)$$

本系统中气温、气压、湿度等连续性要素采用了上述空间平均值比较法及 Madsen-Allerupt 方法,当空间平均值比较法检测数据为错误时,该数据不再进行 Madsen-Allerupt 方法检测,否则应继续 Madsen-Allerupt 方法检测;而降水、风速等离散型要素采用空间极值比较法,且针对被检数据不同量级,参数 $a1$ 、 $a2$ 取值不同。如针对降水而言,不符合下列控制规则,被检站数据可疑:当 $R = 0$ mm 时, $\text{VALUE_MIN} \leq 1$ mm;当 $0 \text{ mm} < R \leq 5$ mm 时,

$VALUE_MIN \leq 10 \text{ mm}$; 当 $5 \text{ mm} < R \leq 10 \text{ mm}$ 时, $VALUE_MIN \leq 2R$ 且 $VALUE_MAX > 0 \text{ mm}$; 当 $10 \text{ mm} < R \leq 50 \text{ mm}$ 时, $R \geq 0.5 \times VALUE_MIN$ 且 $R \leq 2 \times VALUE_MAX$; 当 $R > 50 \text{ mm}$ 时, $R \geq 0.7 \times VALUE_MIN$ 且 $R \leq 1.5 \times VALUE_MAX$ 。

2.3.5 人机交互质量监控

上述质量控制方法均可通过软件自动完成,但是完全自动化的质量控制不可能解决所有的数据质量问题,对于一些特殊情况或问题的判断还要辅以人工检查分析,即在上述自动控制后对可疑或错误

数据用人机交互的方式进行质量控制。一般利用人机交互质量监控平台提供的功能,进行相关数据比较、图形分析、台站查询、邻站比较和人工分析判断,最终确定数据质量状况。

本系统提供了如图 7 所示的多种人机交互与监控界面。其中,图 7a 直观地显示了当前时次观测数据中,各要素疑误数据的统计结果及其空间分布情况;图 7b~7c 为人工对疑误数据的分析判断提供了时空及要素间的比较分析功能;图 7d 提供了数据质量控制码的修改功能。

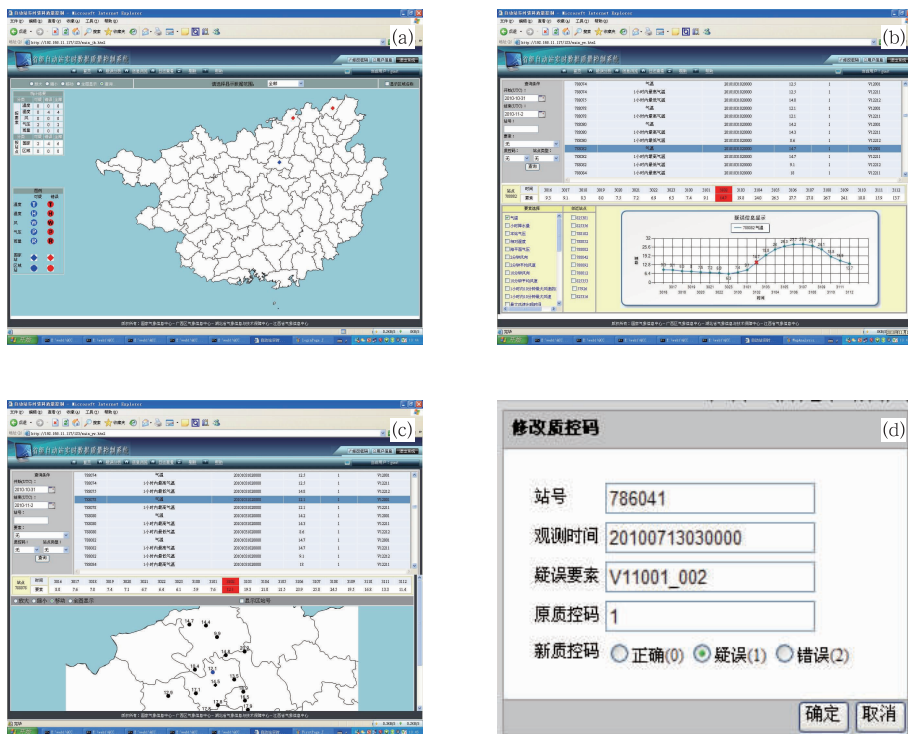


图 7 质量控制系统中部分人机交互分析界面

Fig. 7 Partial man-computer interfaces of the QC system

3 系统功能设计

随着我国国家级自动气象站的建设以及伴随的大量区域自动站的建成,众多台站观测数据开始实时上传。国家级和省级资料服务部门及时将收集到的数据解码并录入到各自的实时气象资料应用数据库(简称实时库),以保障资料服务与应用的及时性。这就要求实时数据质量控制环节不仅要做到有效,还要力求实时、快速。中国气象局发布的《气象信息网络系统发展规划(2011—2015 年)》(中国气象局, 2011)中,规定了国家级主要负责质量评估、省级负

责质量控制、台站负责质量监控与修正的国内资料三级质量控制业务体系。按照该业务分工,在全国自动站观测数据采集、传输、收集、解码入库与实时服务业务流程(图 8 黑线流)基础上,设计并增加全国自动站实时观测资料三级质量控制业务系统(图 8 蓝线流)。该系统主要由自动站资料实时质量控制、人机交互质量监控、质量评估、国家级-省级-台站级三级质量控制疑误信息自动反馈功能块构成,各功能块均基于实时库中的自动站数据而运行。

3.1 质量控制

基于第二节技术方法开发的质量控制软件,分

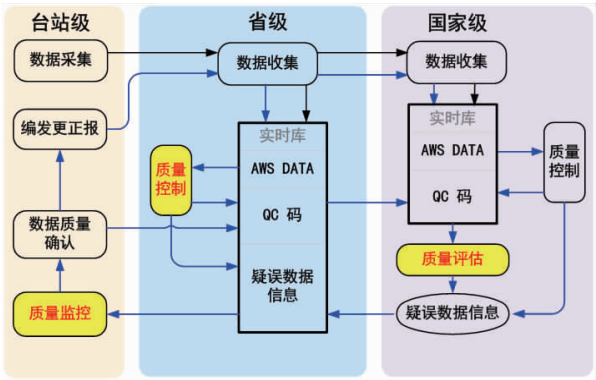


图 8 全国自动站实时资料三级质量控制系统
Fig. 8 The three-step quality control system of AWS real-time observation data in China

别基于国家级和省级实时库中的自动站数据库表统一部署运行。软件质量控制过程中,对实时库中的每个观测数据自动标注 QC 码。国家级和省级质量控制的区别在于国家级进行空间一致性检查时,部分台站可跨省(区、市)应用邻近站数据。

3.2 质量监控与疑误信息反馈

虽然数据质量监控在国家级和省级均可进行,但是由于台站级人员处在观测一线,对数据采集过程、仪器状态和台站周围环境最为熟悉,因此数据质量监控由台站级负责更具优势。在本系统设计建立的国家级-省级-台站级三级质量控制疑误信息自动反馈流程中,确定无论是国家级还是省级业务系统检测出的疑误数据,均由台站进行质量监控、数据质

量确认与错误数据更正。基本思路为:自动站质量控制软件在国家级和省级分别运行,检测出的疑误数据信息都统一到省级实时库进行汇总处理,以确保疑误信息记录的唯一性与完整性。汇总后的疑误信息通过 web 方式到达所属台站监控平台进行处理;台站将处理结果直接反馈给省级实时库,并由省级将反馈结果上传至国家级,国家级和省级实时库根据台站反馈结果,自动更正观测数据或数据质量控制码,从而保证国家级和省级质量控制数据的一致性。

在国家级-省级-台站级三级质量控制疑误信息自动反馈流程中,主要环节在省级实时库与台站监控平台间,流程软件采用了 C/S 与 B/S 混用的架构方式,运行于省级实时库(即服务器端)与台站监控平台(即客户端)之间的气象广域网络环境(李志鹏等,2012)。省级实时库根据国家级和省级汇总后的各条疑误数据信息实现即时告警信息的生成与发送功能(C/S),并负责向台站提供分析检查可疑数据所需的基本资料、接收台站数据检查修订结果(B/S)。台站监控平台部署于各国家级气象台站,实现本站及辖区内区域站疑误数据告警信息的监测与显示功能(C/S),并负责本地疑误数据的分析检查与修订反馈(B/S)。在台站客户端,B/S 应用部分设计为一个功能模块,嵌入在 C/S 模式的主体程序中。

当测站(含国家级站及所辖区域站)被检测出疑误数据时,系统会自动在台站监控桌面右下角弹出图 9a 所示告警对话框并伴有告警声音。在省级实



图 9 台站数据质量监控、反馈界面
Fig. 9 Some interfaces for AWS data quality monitoring and feedback

时库部署 MAS 信息机的条件下,系统也可自动向台站业务人员发送手机短信告警。点击图 9a 警告框则弹出“自动气象站数据质量检测结果反馈”功能框。为便于分析数据的质量状况,该功能框提供了被检站及周边站点同要素、同时次观测数据空间分布图 9b、被检站及周边站点多要素时间变化曲线图 9c,同时可查看近三天疑误数据反馈情况 9d。当确认数据错误时,如图 9e 所示,则需对数据错误原因进行分析选择。最后点击“反馈”按钮,对确认的数据质量信息向省级实时库自动反馈,至此测站用户完成反馈工作。区域自动站数据只进行数据质量的分析判断与反馈。国家级台站如果数据确认错误,还需要通过编发更正报的方式更正错误数据。如图 8 所示,更正报数据通过国内数据传输与收集业务系统,到达省级和国家级实时库对错误数据进行更正。

3.3 质量评估

如图 10 所示,数据质量评估主要根据质量控制后的数据质量控制码,以省(区、市)为单位,从各要素数据可用率以及疑误信息反馈率等角度进行评估(气预函,2012;2014),并对数据质量问题进行分析,以警惕后续相关工作质量的改进。其中,数据可用率指标用来评价自动站数据总体质量,其为 QC 码为 0 的数据总量与应观测数据总量之比。除给出分省的质量统计结果外,增加了单站分要素质量统计结果,有利于各省掌握本省各站的质量状况,及时分析查找问题台站的原因。

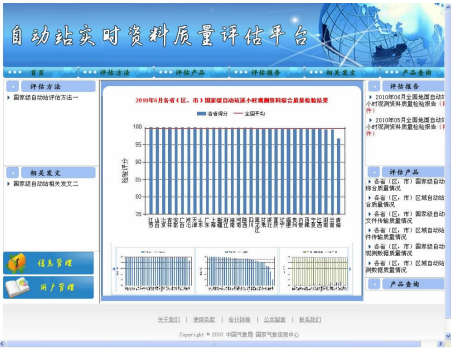


图 10 自动站实时数据质量评估平台
Fig. 10 Quality evaluation interface for
AWS real-time observation data

4 业务应用

全国自动站实时观测数据质量控制系统研制与

建设历时三年。在质量控制技术方法研制的基础上,2010 年 6 月完成了质量控制软件的开发,并在国家级部署试运行,实时库中质量控制后的自动站数据及其 QC 码开始面向国家级用户使用。同年 10 月完成了自动站实时数据质量评估平台的设计开发,开始定期提供自动站实时数据质量评估报告。2011 年 6 月质量控制软件开始在全国 31 个省级业务试运行。在国家级-省级-台站级三级质量控制疑误信息自动反馈流程设计与建设完成的基础上,2012 年 7 月开始,全国自动站实时观测数据三级质量控制系统在全国准业务运行,台站开始利用监控平台,对国家级和省级检测的自动站疑误数据信息进行质量监控与反馈。

通过对自动站实时观测数据三级质量控制,尽可能地避免了错误数据的应用,并进一步通过数据质量评估考核,上传的数据质量得到显著提高。从图 11 可见,在本系统应用过程中,主汛期(7—9 月)全国自动站实时观测的逐小时气温、降水数据的质量逐年得到提高,数据可用率分别从 2009 年的 88%、83% 提升到 2012、2013 年的 98% 左右。

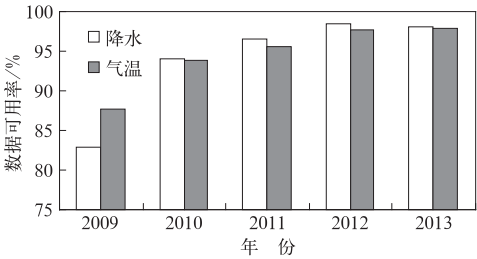


图 11 2009—2013 年 7—9 月全国自动站
实时小时数据质量评估结果
Fig. 11 Quality evaluation result for
AWS real-time hourly data collected
from July to September of 2009—2013

5 结论与讨论

(1) 该系统研制确定的质量控制技术包括界限值检查、时间一致性检查、内部一致性检查、空间一致性检查、人机交互监控等,每种检查过程又包含多项控制方式。

(2) 该系统基于国家级和省级实时库运行,主要由实时质量控制软件、人机交互质量监控平台、质量评估平台、国家级-省级-台站级三级质量控制疑误信息自动反馈流程构成。针对质量控制软件监测

出的疑误数据信息,进一步通过台站质量监控、数据信息反馈、编发更正报的方式明确数据质量或更正数据。

(3) 全国自动气象站实时观测资料三级质量控制系统在业务运行过程中,数据质量得到显著提高。主汛期(7—9月)全国自动站实时观测的逐小时气温、降水数据的质量逐年得到提高,数据可用率分别从2009年的88%、83%提升到2012、2013年的98%左右。用户不仅从实时库中可获取自动站观测数据,还可获得每个数据的质量信息,尽可能地避免了错误数据的应用。

该系统未涉及到不同要素之间的内部一致性检查,比如气温、湿度变化与降水过程的相关性检查等,这是下一步需要研究改进的方面。另外,降水属于局地性非常强的要素,在站网密度稀疏的区域对降水数据进行空间一致性检查,效果很不理想,因此应考虑借助雷达、卫星数据进行辅助判断技术。

参考文献

- 何志军,封秀燕,何利德,等.2010.气象观测资料的四方位空间一致性检验.气象,36(5):117-122.
- 鞠晓慧,任芝花,张强.2010.自动站小时气压的质量控制方法研究.安徽农业科学,38(27):15130-15133.
- 李亚丽,任芝花,陈高峰,等.2015.自动与人工观测气温差异偏大的原因及影响分析——以143个国家基准站为例.气象,41(8):1007-1016.
- 李志鹏,张玮,黄少平,等.2012.自动气象站数据实时质量控制业务软件设计与实现.气象,38(3):371-376.
- 刘小宁,任芝花.2005.地面气象资料质量控制方法研究概述.气象科技,33(3):199-203.
- 气预函[2012]42号.2012.关于下发《全国自动站实时观测资料质量考核办法》的通知.
- 气预函[2014]23号.2014.关于印发《全国地面自动站实时观测资料质量评估办法》的通知.
- 任芝花,刘小宁,杨文霞.2005.极端异常气象资料的综合性质量控制与分析.气象学报,63(4):526-533.
- 任芝花,许松,孙化南,等.2006.全球地面天气报历史资料质量检查与分析.应用气象学报,17(4):412-420.
- 任芝花,熊安元.2007a.地面自动站观测资料三级质量控制业务系统的研制.气象,33(1):19-24.
- 任芝花,熊安元,邹风玲.2007b.中国地面月气候资料质量控制方法的研究.应用气象学报,18(4):516-523.
- 任芝花,赵平,张强,等.2010.适用于全国自动站小时降水资料的质量控制方法.气象,36(7):123-132.
- 陶士伟,徐枝芳.2007.加密自动站资料质量保障体系分析.气象,33(2):34-41.
- 陶士伟,仲跻芹,徐枝芳,等.2009.地面自动站资料质量控制方案及应用.高原气象,28(5):1202-1209.
- 王海军,刘莹.2012.综合一致性质量控制方法及其在气温中的应用.应用气象学报,23(1):69-76.
- 王海军,杨志彪,杨代才,等.2007.自动气象站实时资料自动质量控制方法及其应用.气象,33(10):102-106.
- 熊安元.2003.北欧气象观测资料的质量控制.气象科技,3(5):314-320.
- 张志富,任芝花,张强,等.2013.自动站小时气温数据的质控系统研究.气象与环境学报,29(4):64-70.
- 赵煜飞,任芝花,张强.2011.适用于全国自动站正点相对湿度资料的质量控制方法.气象科学,31(6):687-693.
- 中国气象局.2005.地面气象观测数据文件和记录簿表格式.北京:气象出版社,18-65.
- 中国气象局.2010.中华人民共和国气象行业标准:地面气象观测资料质量控制.
- 中国气象局.2011.《气象信息网络系统发展规划(2011—2015年)》.
- 周林,李湘.2002.关于美国自动地面观测系统(ASOS)的考察报告.气象科技合作动态,(4):18.
- Feng Song, Qi Hu, Weihong Qian. 2004. Quality control of daily meteorological data in China, 1951—2000: A new dataset. Int J Climatol, 24:853-870.
- Peterson T C, Vose R S, Schmoyer R, et al. 1998. Global historical climatology network (GHCN) quality control of monthly temperature data. Int J Climatol, 18:1169-1179.
- Sciuto G, Bonaccorso B, Cancelliere A, et al. 2009. Quality control of daily rainfall data with neural networks. J Hydro, 364:13-22.
- Vejen F, Jacobsson C, Fredriksson U, et al. 2002. Quality control of meteorological observations automatic methods used in the Nordic countries. Climate Report, No. 8/2002, KLIMA.
- WMO. 1992. Manual on GDPS. WMO-No. 485, Volume 1: II. 1-4.
- WMO. 2008. Guide to Meteorological Instrument and Methods of Observation. WMO-No. 8.
- Zahumensk I (Slovakia). 2004. Guidelines on Quality Control Procedures for Data from Automatic Weather Stations. Expert Team on Requirements for Data from Automatic Weather Stations, Third Session, WMO.