

# 基于 CIMISS 的土壤水分质控应用研究

王立俊<sup>1,2\*</sup>, 施晨晓<sup>1,2\*</sup>, 王旭<sup>1\*</sup>, 贺永兴<sup>1</sup>, 吴军<sup>1</sup>

(1. 海南省气象信息中心, 海南海口 570203;

2. 海南省南海气象防灾减灾重点实验室, 海南海口 570203)

**摘要:**气象资料业务系统 MDOS (meteorological data operation system) 的质控方法使用通用阈值去质控土壤水分数据, 会漏检和误检出部分疑误数据, 且疑误结果多以“未通过降水关系检查”为主, 降低值班人员的数据审核效率。为了提高土壤水分质控效率, 结合本地土壤特性, 分析海南土壤水分历史数据, 总结适合本地的阈值范围参数, 提出基于 CIMISS (China integrated meteorological information service system) 的 MQCSM (multiple quality control method of soil moisture) 算法。该算法引入时变检查、持续性检查等检查方法, 多重质控原始土壤水分数据, 能快速、准确地质控出疑误数据, 并分类展示疑误结果于 web 监控页面工值班人员筛查。业务试用结果表明, 对比现有质控方法, 该算法能准确、有效地筛查出疑误土壤水分疑误数据, 且质控监视平台实时展示疑误结果, 提高了值班人员对疑误数据的审核效率。

**关键词:**全国综合气象信息共享平台; 土壤水分; 疑误数据; 阈值选定; 多重质控

中图分类号: TP311.1

文献标识码: A

文章编号: 1673-629X(2021)04-0210-06

doi: 10.3969/j.issn.1673-629X.2021.04.036

## Application Research of Quality Control of Soil Moisture Based on CIMISS

WANG Li-jun<sup>1,2\*</sup>, SHI Chen-xiao<sup>1,2\*</sup>, WANG Xu<sup>1\*</sup>, HE Yong-xing<sup>1</sup>, WU Jun<sup>1</sup>

(1. Meteorological Information Center of Hainan Province, Haikou 570203, China;

2. Key Laboratory of South China Sea Meteorological Disaster Prevention and Mitigation of Hainan Province, Haikou 570203, China)

**Abstract:** The general threshold is used in the quality control methods of MDOS (meteorological data operation system) to control soil moisture data so that some suspected data are missed and detected by mistake and most of suspected result are mainly failing the inspection of precipitation relationship which leads to data audit by duty staff with lower efficiency. The suitable threshold range parameters are summarized by considering local soil characteristics and analyzing the historical data of soil moisture in Hainan. With introducing inspection methods such as time-varying and time-continuity, an algorithm of Multiple Quality Control of Soil Moisture based on China Integrated Meteorological Information Service System is proposed to discover suspected data of soil moisture quickly and accurately by multiple quality control and the result are classified and displayed in the web page for the screening of personnel on duty. Compared with the existing quality control methods, the result of business trial showed that suspected data of soil moisture were screened accurately and effectively through MQCSM algorithm and were displayed on the monitoring platform for improving the efficiency of data audit by personnel on duty.

**Key words:** CIMISS; soil moisture; suspected data; threshold selection; multiple quality control

## 0 引言

土壤水分是土壤的重要物理参数, 准确地测量土壤水分及其变化, 对了解土壤水分状况, 指导灌溉施肥和研究土壤水分运动具有十分重要的意义<sup>[1]</sup>。中国是一个农业大国, 农业生产对气象服务的需求与日俱增。土壤水分观测则是农业气象观测的重要组成部分, 其

资料的应用, 对掌握土壤水分变化规律, 提供实时农业生产服务和气象防灾减灾具有重要意义<sup>[2]</sup>。

海南省自 2010 年开始安装、建设土壤水分自动观测站; 2012 年, 第一批自动土壤水分站通过了中国气象局的考核验收, 开始业务化运行。截至 2013 年底, 海南已建成自动土壤水分站 18 个, 全部通过中国气象

收稿日期: 2020-04-22

修回日期: 2020-08-24

基金项目: 国家自然科学基金(41775011); 海南省气象科技创新项目(HNQXQN201703)

作者简介: 王立俊(1989-), 男, 助理工程师, 研究方向为气象数据理论和应用研究。

局验收,成为业务考核站,其分布情况如图1所示。

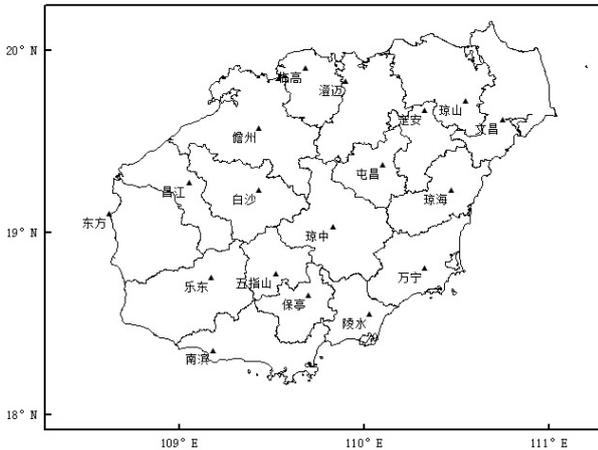


图1 海南省土壤水分自动观测站分布图

中国气象局与各省、市级气象局设计研发了国家级-省级-台站三级质控的 MDOS (meteorological data operation system) 气象资料业务系统<sup>[3]</sup>, 实现对各种类型的气象原始观测数据进行质控和以人机交互的形式审核质控后产生的疑误数据<sup>[4-5]</sup>。

气象信息共享系统 CIMISS (China integrated meteorological information service system) 由中国气象局设计、研发, 整个系统由五个功能模块组成, 分别是: 收集与分发模块 CTS、数据加工处理模块 DPC、数据存储管理模块 SOD、综合业务监控模块 MCP、数据共享服务模块 GDS。这些模块分别部署在国家和各省级节点, 对各种气象资料和产品进行收集、加工、存储及服务, 是气象业务、服务和管理的核心基础数据支撑平台<sup>[6]</sup>, 是 MDOS 系统的数据来源。

目前, MDOS 系统使用通用阈值来质控土壤水分观测数据中 0~100 cm 土层要素, 筛选出的疑误数据以“未通过降水关系检查”为主提示到值班人员, 许多疑误数据被误检、漏检, 使值班人员的审核效率低下。

针对上述问题, 通过分析海南各台站历年小时土壤水分观测数据不同土层的特异性, 总结出不同土层各要素的阈值范围, 引入气候极值检查、时变检查、持续性检查等检查方法, 提出基于 CIMISS 的 MQCSM (multiple quality control method of soil moisture) 算法。该算法采用多时次、多方法的方式质控原始土壤水分观测数据, 能快速、准确地筛查出疑误数据, 值班人员阅览 web 页面查看疑误信息。业务试用结果表明, 系统能有效地质控出更多的误检或漏检土壤水分疑误数据, 提高值班人员审核疑误数据的效率。

## 1 相关技术

### 1.1 质控方法

目前, 国内外气象研究者设计研发了各种质量控

制系统, 用于对不同类型的气象数据进行质量控制。例如: 在地面观测业务中, 北欧国家使用台站级质控、入库前实时质控、入库后非实时质控和人工质控等多种方式质控地面自动站观测数据<sup>[7]</sup>, 美国使用台站-州-国家三级质控方式来质控观测数据<sup>[8]</sup>, 中国使用台站级、省级和国家三级质控方式, 由下至上地对观测数据进行质量控制。

在不同的质量控制系统中, 所使用的数据质量控制方法分为以下五种<sup>[9-14]</sup>, 分别是气候界限值检查、台站极值检查、内部一致性检查、时间一致性检查和空间一致性检查。

(1) 气候界限值检查: 一般是从气候学的角度去判断某个气象要素的数值是否超过了不可能出现的气象要素临界值<sup>[9]</sup>, 通过设置极值上下界来判定要素数值是否正确。

(2) 台站极值检查: 通常是检查原始观测数据中某个要素值是否超出台站已出现过的历史极值, 若数值超过历史极值, 由观测员分析站点历史数据, 按照实际情况和工作经验判断观测数据的正确性<sup>[10]</sup>。

(3) 内部一致性检查: 在同一时次或时段内, 不同气象要素之间是否满足一定的物理关系。若不满足这些关系, 则至少有一个气象要素为错误数据<sup>[11]</sup>。

(4) 时间一致性检查: 在连续时间段内, 某个气象要素必须满足一定的规律性变化。如最高气温存在明显的日变化, 若连续 24 小时数值不发生变化, 则可判定为疑误或错误。

(5) 空间一致性检查: 在同一空间上 (相邻站点), 同一气象要素满足一定的连续性, 常用方法有空间插值法、Madsen 方法<sup>[13]</sup>、空间回归检查法<sup>[14]</sup>等。

### 1.2 质控平台

该文设计的质控平台是基于 CIMISS 接口服务进行研发的。在 CIMISS 系统中, 观测数据的处理流程如下: CTS 模块实时收集市、县台站及无人自动站上传至省级节点的各类气象观测数据, 自动对上行数据进行格式检查, 快速质控等操作, 接着分别分发至国家级和 DPC 等分发地址; DPC 模块实时对分发数据进行解码入库, 由数据简约处理流程负责入库地面资料、高空资料等结构化数据, 雷达、卫星云图等非结构化数据由 SOD 模块中的处理流程负责入库; MCP 模块实时接收 CTS、DPC 等模块发送的处理及告警信息, 并监控、预警各类观测数据的处理状态; 将数据库中的数据封装成 MUSIC 接口, 对外提供数据访问服务。

### 1.3 SSH 框架

该文设计的质控平台的开发架构采用 SSH 框架技术, 以 B/S 模式来跨平台呈现逐小时土壤水分质控结果。SSH 框架是由 Struts 2、Spring 和 Hibernate 组成

的,属于主流的轻量级 J2EE 软件开发架构。采用该框架开发应用,能较好地对应用功能模块分层、解耦,简化系统开发的复杂度,缩短系统开发时间,使系统具备良好的可拓展性和可移植性。

其中,Struts2 是以 Webwork 为核心的逻辑控制器,采用拦截器机制来响应前端用户提交的 web 请求,将 Servlet 与业务逻辑控制器分离<sup>[15]</sup>。Spring 是属于轻量级的 Java Web 框架,通过配置文件及事务管理机制,可灵活管理多种数据库,提供多元化的业务逻辑<sup>[16]</sup>。Hibernate 采用 O/R Mapping 技术,通过配置 XML 文件或 Annotation 注解为 Java 对象和各类数据库中的表结构建立一种或多种映射关系,操控对象即操作数据库<sup>[17]</sup>。

## 2 MQCSM 算法

目前,MDOS 系统使用通用阈值来质控土壤水分观测数据,使得部分疑误数据被漏检和误检,降低值班人员的疑误数据筛查效率。因此,通过分析、研究海南地区各台站土壤水分观测数据不同土层要素的特异性,总结出不同土层的阈值范围,并引入时变检查、持续性检查等检查方法,提出基于 CIMISS 的 MQCSM 算法。该算法对不同土层的要素值进行质控检查,能快速、有效地质控出土壤水分疑误数据。

### 2.1 选取阈值范围

海南地区的原始土壤水分逐小时观测数据仅在 0~10 cm、10~20 cm、20~30 cm、30~40 cm、40~50 cm 五个土层有实际的观测数值,这些观测参数分别是土壤体积含水量  $Q$  (%)、土壤体积重量含水率  $w$  (%)、土壤相对湿度  $R$  (%) 和土壤有效水分贮存量  $u$  (mm),其中通过  $Q$  可计算出  $w$ ,  $R$  和  $u$  的数值,具体公式见《土壤水分观测 频域反测法》<sup>[18]</sup>。

该文使用的具体质量控制方法共四种:(1)界限值检查;(2)极值检查;(3)时变检查;(4)持续性检查。选取 2018 年 1 月至 2019 年 12 月时间段,共计 314 908 条原始土壤水分逐小时观测数据,分析、总结出 0~50 cm 五个土层各土壤要素的阈值范围,用于检查不同土层中的不同要素是否疑误。

#### 2.1.1 界限值检查阈值选取

根据《自动土壤水分观测规范(试行)》的规定,自动土壤观测仪传感器对各层土壤体积含水量的阈值范围是 0%~50%。分析海南地区 2018 至 2019 年的逐小时观测数据分析,发现大多数台站的各层土壤体积含水量数值(尤其是表层)往往大于 50%,结合样本的数据特征,将界限值检查中土壤体积含水量的阈值范围设为 0%~60%,当某一层次的  $Q$  被判定为疑误时,同一层次的  $w$ ,  $R$  和  $u$  均为疑误。

#### 2.1.2 极值检查阈值选取

统计、分析 2018 年至 2019 年期间的各台站 0~50 cm 土层中各土壤水分要素的极值,得出不同土层的体积含水量数值变化范围较大,最小值有 4%,最大值有 99%,且最大值基本分布在 30~40 cm 土层,确定不同土层各要素的阈值范围,如表 1 所示。

表 1 不同土层各要素的阈值范围

土壤水分要素	下限	上限
0~20 cm 相对湿度 $R_{10}$ $R_{20}$	0	150
30~50 cm 相对湿度 $R_{30}$ $R_{40}$ $R_{50}$	0	200
0~10 cm 重量含水率 $w_{10}$	0	60
20~40 cm 重量含水率 $w_{20}$ $w_{30}$ $w_{40}$	0	55
40~50 cm 重量含水率 $w_{50}$	0	50
0~30 cm 有效水分贮存量 $u_{10}$ $u_{20}$ $u_{30}$	-5	45
40~50 cm 有效水分贮存量 $u_{40}$ $u_{50}$	-10	45

#### 2.1.3 时变检查阈值选取

时变检查是检查 0~50 cm 土层中某土壤水分要素前后 2 个时次的变化值是否落在合理的阈值范围内,这是因为无降水、观测仪器传感器无故障或其他外界因素影响时,要素的变化值在一定时间内波动较小。对所选取的业务样本数据进行个例统计和分析,总结出适用 0~50 cm 土层各要素相邻时次变化值的阈值范围,如表 2 所示。

表 2 各要素相邻时次变化值的阈值范围

相邻时次变化值	下限	上限
体积含水量 $\Delta Q_{0-20}$	-10	15
体积含水量 $\Delta Q_{30-50}$	-12	20
重量含水率 $\Delta w_{0-20}$	-8	12
重量含水率 $\Delta w_{30-50}$	-10	9
相对湿度 $\Delta R_{0-20}$	-20	45
相对湿度 $\Delta R_{30-50}$	-10	30
有效水分贮存量 $\Delta u_{0-20}$	-10	15
有效水分贮存量 $\Delta u_{30-50}$	-8	25

其中,  $\Delta Q_{0-20}$  表示 0~20 cm 土层体积含水量当前时次与上一次的变化值,其中  $\Delta Q_{30-50}$  表示 30~50 cm 土层体积含水量当前时次  $i$  与上一时次  $i-1$  的变化值,如式(1)所示,其他要素类似,不逐一详述。

$$\begin{cases} \Delta Q_{0-20} = Q_i - Q_{i-1}, \Delta Q_{30-50} = Q_i - Q_{i-1} \\ \Delta w_{0-20} = w_i - w_{i-1}, \Delta w_{30-50} = w_i - w_{i-1} \\ \Delta R_{0-20} = R_i - R_{i-1}, \Delta R_{30-50} = R_i - R_{i-1} \\ \Delta u_{0-20} = u_i - u_{i-1}, \Delta u_{30-50} = u_i - u_{i-1} \end{cases} \quad (1)$$

#### 2.1.4 持续性检查阈值选取

持续性检查是检查土层中各土壤要素的数值在一

段连续时间段内是否发生变化,若数值未发生变化,则有可能是观测仪器存在故障问题,或是观测仪器周围受到积水影响。此外,由于深层土壤各要素数值变化幅度不大和土壤特性,该将持续性检查中的时间时长设为 48 小时,若某个土壤要素在设定的时长内未发生变化,则判定为“疑误”。

### 2.2 算法思路

MQCSM 算法采用上述检查方法对海南土壤水分站 0~50 cm 土层的各土壤要素进行质量控制,实现对土壤水分数据的多重质控,及时、有效地筛选出疑误数据。根据选取的检查方法,算法思路主要分为三个部分,如图 2 所示。

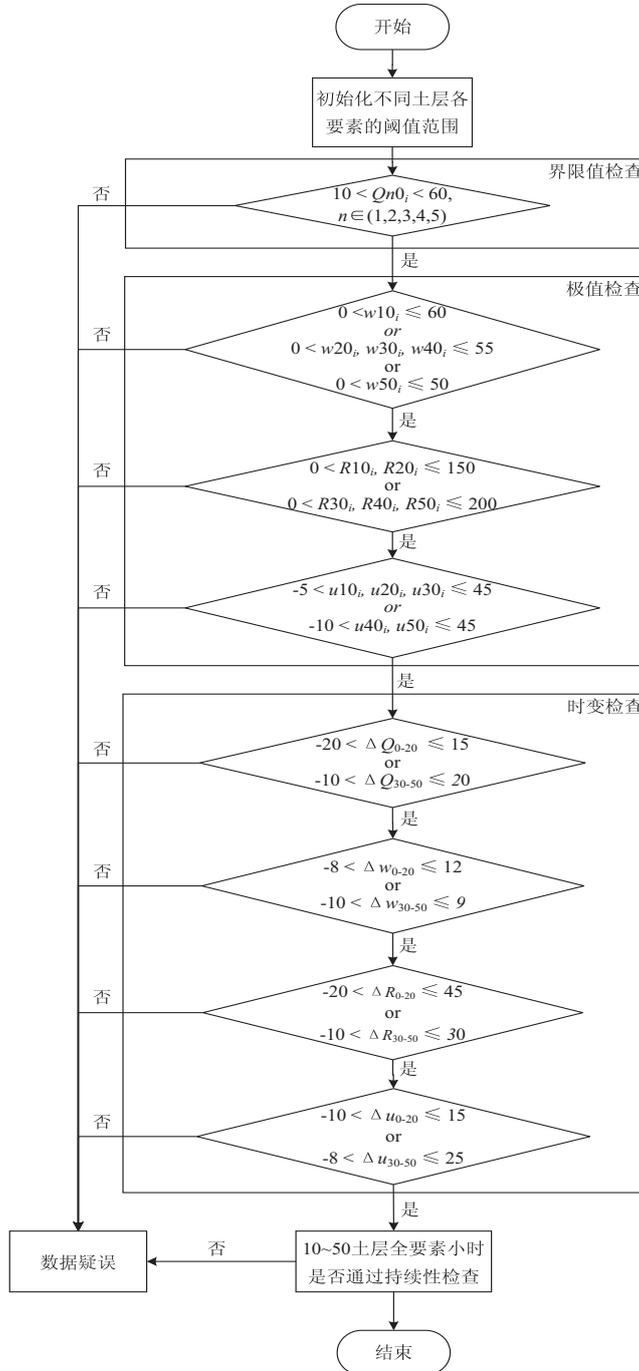


图 2 算法流程

(1) 阈值初始化:使用上小节的阈值选取来初始化 0~50 cm 土层各检查方法的阈值范围,阈值范围可根据实际情况自定义调整。

(2) 多重质控检查:依次对待质控时次的土壤水分数据进行质控检查,整个流程按检查方法分为四个

部分,分别是:(a)对 0~50 cm 土层中的体积含水量  $Q_{n0i}$  ( $n \in (1,2,3,4,5)$ ) 进行界限值检查,其中  $Q_{10i}$  表示某一时次的 10 cm 土层体积含水量,其他土层类似。若  $Q_{n0i}$  的数值不在阈值范围内,则判定  $Q_{n0i}$  为疑误;(b)对 0~50 cm 土层的重量含水率  $w_{n0i}$ 、相对湿度

$R_{n0i}$ 、有效水分贮存量  $u_{n0i}$  进行极值检查,  $w_{10i}$ 、 $R_{n0i}$  和  $u_{n0i}$  的表示与  $Q_{10i}$  类似,  $n \in (1, 2, 3, 4, 5)$ 。若  $w_{10i}$ 、 $R_{n0i}$  或  $u_{n0i}$  的数值不在阈值范围内, 则可判定该要素为疑误; (c) 对 0 ~ 50 cm 土层的体积含水量时变值  $\Delta Q_{n0}$ 、重量含水率时变值  $\Delta w_{n0}$ 、相对湿度时变值  $\Delta R_{n0}$  和有效水分贮存量时变值  $\Delta u_{n0}$  进行时变检查, 这里用 10 cm 土层体积含水量时变值  $\Delta Q_{10}$  来举例,  $\Delta Q_{10}$  的数值为当前时次的要素值  $Q_{10i}$  减去前一时次的要素值  $Q_{10i-1}$ , 若差值不在阈值范围, 则判定  $Q_{10i}$  为疑误, 其他时变值的判定与  $Q_{10i}$  类似,  $n \in (1, 2, 3, 4, 5)$ ; (d) 对 0 ~ 50 cm 土层各土壤要素进行连续 48 小时的持续性检查, 若某个时次某土壤要素连续 48 个时次未发生改变, 则可判定为疑误。上述 4 个检查判定为疑误的土壤要素用于生成下阶段的疑误记录。

(3) 生成疑误记录: 后台自动获取上阶段产生的疑误土壤要素关联的台站号、观测时次、疑误状态、疑误要素及要素值等关键字段, 生成最终的疑误记录。

### 3 平台实现

根据审核人员的值班需求, 分析、总结出逐小时土壤水分实时质控的流程, 使用 SSH 开发框架, 研发了一个集数据质控、疑误数据展示及搜索的土壤水分实时质控平台。

平台的工作流程如图 3 所示: 值班人员打开土壤水分监控页面后, 后台会自动调用数据质控模块: (1) 数据入库: 若数据库为新库, 则获取系统时间, 自动生成相应的 Music 语句来调用接口获取相应时间段内的原始观测数据, 对数据预处理后, 初始化数据库; 若数据库已有数据, 后台获取库中最新的观测时次, 并生成相应的 Music 语句来获取待入库的原始观测数据。

(2) 数据质控: 后台实时对未质控过的观测数据进行多重质控检查, 筛选出疑误数据, 生成相应的质控记录。

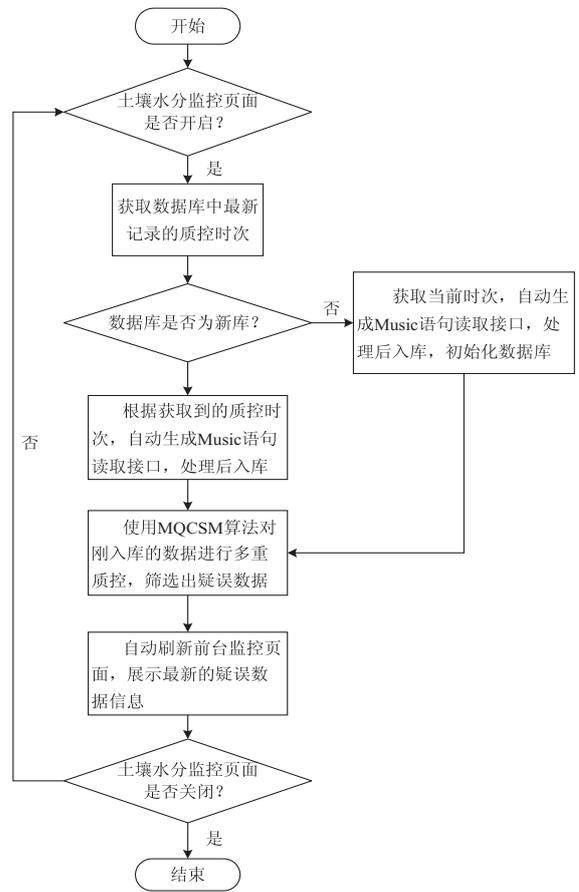


图 3 平台工作流程

土壤水分质控结果展示页面如图 4 所示, 值班人员能实时看到最新时次的土壤水分质控结果, 质控结果内容包含疑误数据的相关信息, 便于值班人员定位疑误数据产生的原因。

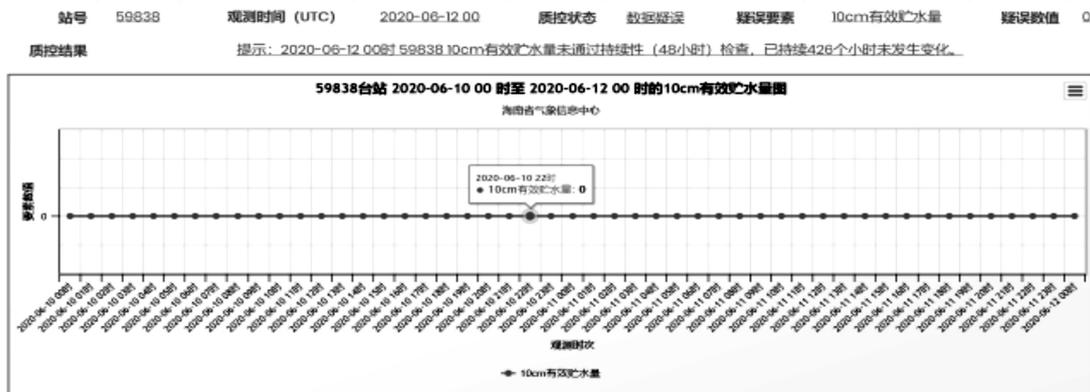


图 4 质控结果展示页面

## 4 实验结果分析

### 4.1 实验环境

使用 Java 编程语言实现 MQCSM 算法, 系统参数为: (1) CPU: Intel (R) Core (TM) i7-6700HQ 2.60

GHz; (2) 内存: 8 G; (3) 硬盘: 500 G; (4) 操作系统: Windows 10 x64 专业版。

### 4.2 业务数据集

算法处理的业务数据集为 2020 年 1 月至 2020 年 2 月期间的所有土壤水分站的逐小时数据, 每个逐小

时数据有 5 个土层数据(0~50 cm),每层包含 4 个土壤要素,共计 518 400 个数据,调用 CIMISS 接口获取该时间段原始数据,预处理后入本地数据库。

### 4.3 结果分析

使用 MQCSM 算法质控后的结果与 MDOS 系统质控后的结果,对比结果如表 3 所示,表中误检率表示数据本身正确,经人工核实确认正确的,但被程序质控出来的疑误记录占总记录数的比值;漏检率表示数据本身可疑或错误,经人工核实确认可疑,但未被程序质控出来的疑误记录占总记录数的比值。

表 3 对比结果

操作	总数据/个	疑误数/个	误检率/%	漏检率/%
MDOS	518 400	18 465	0.3	1.6
MQCSM	518 400	7 721	0.13	0

由对比结果可知,与 MDOS 系统的质控结果对比,MQCSM 算法在数据质控中具有更低的误检率和漏检率,这是因为 MDOS 系统中检查方法采用通用阈值来对土壤水分数据进行质量控制,筛查出较多的疑误数据,但部分疑误数据在本地属于正常数据范围,人工审核后为正确数据,且多以“未通过降水关系检查”提示,而 MQCSM 算法的 0.13% 误检率都是未通过持续性检查的疑误数据,这些数据的数值是正确的,但在实际日常业务中,需要台站人员去检查仪器周边确认是否有干扰因素,导致数值长时间未发生变化。所以 MQCSM 算法采用分析出的要素阈值范围,引入时变检查、持续性检查等检查方法,对观测数据进行多重质控,能更准确地发现疑误数据。

## 5 结束语

MDOS 系统使用通用阈值来质控土壤水分数据,会漏检或误检较多疑误数据,且疑误结果多以“未通过降水关系检查”提示为主。为提高土壤水分疑误数据的质控效率,通过分析海南历史数据,结合海南本地的气候特点,总结出适用本地的阈值范围,引入极值检查、时变检查和持续检查等检查方法,提出了 MQCSM 算法。该算法对土壤水分数据进行多重质控,与现有 MDOS 系统的土壤水分质控效果对比,能更有效、准确地质控出疑误数据。在此基础上,研发了一个土壤水分质控监控平台,使值班人员能快速地筛查土壤水分疑误数据。经业务试用,各模块运行稳定,能有效地提高土壤水分数据审核效率。

### 参考文献:

[1] 王晓东,杨天明,吴必文,等.安徽省自动观测土壤水分质量控制方法[J].气象科技,2015,43(3):399-404.  
[2] 郭旭,龙柯吉,赵旋,等.四川地区自动土壤水分站数

据质量控制方法研究[J].高原山地气象研究,2017,37(2):90-95.  
[3] 王立俊,江益,程洪涛,等.南海区域站实时气象数据质控系统研究[J].计算机技术与发展,2017,27(8):177-181.  
[4] 王佳强,赵煜飞,任芝花,等.中国自动土壤水分观测资料质量控制方法设计与效果检验[J].气象,2018,44(2):244-257.  
[5] 任芝花,张志富,孙超,等.全国自动气象站实时观测资料三级质量控制系统设计[J].气象,2015,41(10):1268-1277.  
[6] 李志鹏,胡佳军,杨立苑,等.基于 CIMISS 的气象数据处理时效监视系统设计与实现[J].气象与减灾研究,2016,39(4):309-313.  
[7] CHAE J H,PARK M S,CHOI Y J. The WISE quality control system for integrated meteorological sensor data[J]. Atmosphere,2014,24(3):445-456.  
[8] QI Y,MARTINAITIS S,ZHANG J,et al. A real-time automated quality control of hourly rain gauge data based on multiple sensors in MRMS system[J]. Journal of Hydrometeorology,2016,17(6):1675-1691.  
[9] 林磊,任义方,张建飞,等.气象观测数据质量控制方法的探讨[J].农业与技术,2018,38(17):126-127.  
[10] 王海军,闫莽莽,向芬,等.逐时气温质量控制中界限值检查算法的设计[J].高原气象,2014,33(6):1722-1729.  
[11] 岳艳霞,智利辉,王磊,等.石家庄市区域自动气象站气温数据适用性分析[J].气象与环境学报,2017,33(2):53-62.  
[12] 闵锦忠,王晨珏,贾瑞怡.苏皖地面自动站资料的质量控制及结果分析[J].大气科学学报,2018,41(5):637-646.  
[13] LANZANTE J R. Resistant,robust and non-parametric techniques for the analysis of climate data:theory and examples, including applications to historical radiosonde station data [J]. International Journal of Climatology,1996,16(11):1197-1226.  
[14] 闫莽莽,王海军,严靖,等.逐时本站气压数据分阶段空间一致性检查方法研究[J].气象与环境学报,2018,34(4):36-44.  
[15] 王立俊,李晋峰,程洪涛,等.气象资料实时传输监控系统应用研究[J].计算机技术与发展,2018,28(3):183-186.  
[16] PANDEY H,RASTOGI H,GUPTA C. Web-based network management system implemented using Hibernate,JBoss and spring framework[C]//2017 2nd international conference on telecommunication and networks (TEL-NET). Noida, India:IEEE,2017:1-5.  
[17] 杨有林,陈海波,王建林,等.宁夏智能化综合气象业务服务共享管理平台设计与实现[J].气象,2018,44(7):961-968.  
[18] GB/T 33705-2017,土壤水分观测频域反射法[S].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,2017.