学号 ：15103402

**常 州 大 学**

**硕 士 专 业 学 位 论 文**

基于数理统计的污水监测

数据真伪鉴别研究

|  |  |
| --- | --- |
| 研究生姓名 | 邹 涛 |
| 指导教师姓名 | 万玉山 |
| 申请学位类别 | 工程硕士 |
| 专业领域名称 | 环境工程 |
| 研究方向 | 水污染控制 |

年 月

**Research on authenticity identification of sewage monitoring data based on mathematical statistics**

A Dissertation Submitted to

**Changzhou University**

**By**

**Zou Tao(作者姓名)**

**(Environment Engineering) (专业领域名称)**

Dissertation Supervisor: Prof. Wan Yu-shan(指导教师)

**常州大学学位论文原创性声明**

本人郑重声明：所呈交的学位论文是本人在导师指导下独立进行的研究工作及取得的研究成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在论文中以明确方式标明。本人已完全意识到本声明的法律结果由本人承担。

作者签名： 签字日期： 年 月 日

**学位论文版权使用授权的说明**

本学位论文作者完全了解 常州大学 有关保留、使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属常州大学。学校有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。学校可以公布学位论文的全部或部分内容，可以采用影印、缩印或扫描等复制手段保存、汇编本学位论文。

保密论文注释：本学位论文属于保密范围，在 年解密后适用本授权书。非保密论文注释：本学位论文不属于保密范围，适用本授权书。

学位论文作者签名： 签字日期： 年 月 日

导师签名： 签字日期： 年 月 日

**中 文 摘 要**

污水在线监测是监测中心对污水处理厂排放废水监督管理的主要手段，其监管力度将直接影响城市水环境的优劣程度，因此对污水处理厂生产废水监控和管理显得十分必要。国家和各省和地区越来越重视生态环境的保护，加强水污染的管理，但由于许多污水处理厂环境保护意识淡薄或者受利益驱动因素，存在监测点稀释污水，私接暗管，篡改监测仪器重要参数等不法行为，导致污水监测数据失真，给监测部门的工作带来很大的麻烦。再加上许多监测人员本身业务素质较低，缺少将数据进行科学有效的分析和处理的能力，导致数据的可信度不高，失去评价意义。

针对城市污水处理厂生产污水监管过程中可能存在的各种生产废水监测数据造假行为，依据数理统计相关理论，实现了基于数理统计的污水监测数据真伪鉴别流程，可以对指定时间段的污水监测指标数据的真实度和可靠性进行有效鉴别，同时也能修正数据，对未知时间段的数据范围做出合理预测，并可从限定的条件下匹配出与待测企业数据变化趋势相近的企业，从而实现与同时间段同行业监测数据的比较分析。基于数理统计的污水监测数据真伪鉴别流程联合运用数理统计的方法，例如格拉布斯法、箱线图法、t检验法等，对数据原有特性进行深度剖析，有效避免了多种外界影响因素的干扰。

确定污水监控指标，熟知各种指标国家排放规定，分析污水处理厂典型污水处理工艺，了解监控部门监控手段，并进行分析需求，数据库设计和系统功能的实现。基于Java语言平台实现系统后台的搭建，基于html+css+javascript实现系统前端功能搭建，基于Oracle数据库实现污水监测指标原始数据和鉴别数据的储存及修改，构建人机交互网页，实现鉴别流程的逻辑功能。

采用B/S网络结构模式构建污水监测数据真伪鉴别系统，并结合某城市污水处理厂2016年年度各监测指标监测数据进行实例验证。结果表明：该系统可以实现对原始数据的增删查改，对监测数据的检测并给出数据真实性和可靠度的鉴定结果。该系统的实现将减轻虚假污水监测数据给环境监测工作带来的巨大压力，为上级监管部门提供有力的科技保证和技术支撑，对于提高监管部门工作人员的业务素质，实现污水处理厂间多元化信息交流有着重要意义。

**关键词:** 真伪鉴别流程；数理统计；B/S模式；鉴别系统

**ABSTRACT**

On line monitoring of sewage is the main means of monitoring and controlling the discharge of wastewater from sewage treatment plants．Supervision will directly affect the quality of urban water environment，so it is necessary to monitor and manage the production wastewater of the sewage treatment plant. Countries and regions pay more and more attention to the protection of ecological environment, strengthen the management of water pollution, but because of weak environmental protection consciousness or driven by the profit factor, many sewage treatment plants have some illegal behaviors, such as diluting sewage at the monitoring points, route underground pipes privately, tampering with the important parameters of monitoring instruments and so on, which causes the distortion of the sewage monitoring data, which brings great trouble to the monitoring department. What 's more, the professional quality of monitoring personnel is low, lack of scientific and effective data analysis and processing capabilities, resulting in the credibility of the data is not high and the loss of evaluation significance.

Directing against various production wastewater monitoring data fraud existing in the process of supervising municipal wastewater treatment plants, this paper realizes the process of identifying the authenticity of sewage monitoring data based on mathematical statistics，which can be used to identify the authenticity and reliability of the monitoring data in the specified time, can also modify the data to make a reasonable prediction of the data range of the unknown time period in the same time, and match the enterprises whose change trend of the wastewater monitoring data is similar under the limited conditions in order to realize the comparison and analysis of monitoring data in the same industry in the same time period. The sewage monitoring data authentication process combination based on mathematical statistics combines the principles of mathematical statistics, such as Grubbs method, the box plot method and the T-Test, to analyze deeply the original characteristics of the data, which effectively avoids the interference of a variety of external factors.

Determine the sewage monitoring indexes, be familiar with a variety of indicators of national emission regulations, analyze typical wastewater treatment technology, understand the monitoring departments’ monitoring means, and analyze the demand, design database and realize system function. Build the system background based on Java language platform, complete the head webpage designing based on html + css + javascript model, realize the storage and modification of the original data and the identification data of the sewage monitoring indexes based on Oracle database. Structure the human-computer interaction web pages to realize the logic function of the identification process.

The B/S network structure model was used to construct the sewage monitoring data authenticity identification system, and the identification results of each monitoring index of a city sewage treatment plant in 2016 are verified by an example. The results show that the system can realize the addition and deletion of the original data, the detection of the monitoring data and the identification of the authenticity and reliability of the data. The realization of this system will reduce great pressure that the false sewage monitoring data brings to environmental monitoring work, and provide strong technical support and technical support for the higher authorities. It is of great significance to improve the professional quality of monitoring personnel and realize communication for pluralistic information between the sewage treatment plants.

**KEY WORDS:** authenticity identification process; mathematical statistics; B/S model; identification system

目 录

[1 诸 论 1](#_Toc479625294)

[1.1 课题背景和研究目的及意义 1](#_Toc479625295)

[1.1.1 课题背景 1](#_Toc479625296)

[1.1.2 研究目的及意义 1](#_Toc479625297)

[1.2 国内外研究现状 2](#_Toc479625298)

[1.2.1 典型污水处理工艺研究现状 2](#_Toc479625299)

[1.2.2 异常数据检测方法研究现状 5](#_Toc479625300)

[1.2.3 基于B/S网络结构模式鉴别系统研究现状 8](#_Toc479625301)

[1.3 主要研究内容及技术路线 9](#_Toc479625302)

[1.3.1 研究目标 9](#_Toc479625303)

[1.3.2 研究内容 9](#_Toc479625304)

[1.3.3 技术路线 10](#_Toc479625305)

[2 研究对象与方法 11](#_Toc479625306)

[2.1 研究对象 11](#_Toc479625307)

[2.1.1 待测污水处理厂污水处理工艺介绍 11](#_Toc479625308)

[2.1.2 污水监测数据指标及来源 15](#_Toc479625309)

[2.2 异常检测研究方法 15](#_Toc479625310)

[2.2.1 格拉布斯法 15](#_Toc479625311)

[2.2.2 箱线图法 18](#_Toc479625312)

[2.2.3 t检验法 21](#_Toc479625313)

[2.2.4 χ2检验法 24](#_Toc479625314)

[2.2.5 F检验法 26](#_Toc479625315)

[2.3 系统开发平台 28](#_Toc479625316)

[2.3.1 系统开发语言 28](#_Toc479625317)

[2.3.2 系统开发工具 28](#_Toc479625318)

[2.3.3 网络结构模式 29](#_Toc479625319)

[3 基于数理统计的污水监测数据真伪鉴别研究 31](#_Toc479625320)

[3.1 引言 31](#_Toc479625321)

[3.2 数据真伪鉴别流程 31](#_Toc479625322)

[3.3 基于数理统计的污水监测数据真伪鉴别 32](#_Toc479625323)

[3.3.1 格拉布斯法修正时观测值 32](#_Toc479625324)

[3.3.2 筛选达标日观测值并修正 37](#_Toc479625325)

[3.3.3 数据纵向对比 40](#_Toc479625326)

[3.3.4 数据横向对比 49](#_Toc479625327)

[3.4 本章小结 58](#_Toc479625328)

[4 基于B/S网络结构模式的污水监测数据真伪鉴别系统 59](#_Toc479625329)

[4.1 引言 59](#_Toc479625330)

[4.2 系统需求分析 59](#_Toc479625331)

[4.3 系统设计思想 60](#_Toc479625332)

[4.3.1 实用性 60](#_Toc479625333)

[4.3.2 安全性与可恢复性 60](#_Toc479625334)

[4.3.3 灵活性与可扩展性 61](#_Toc479625335)

[4.3.4 稳定性与可靠性 61](#_Toc479625336)

[4.4 数据库设计 62](#_Toc479625337)

[4.4.1 概念设计 62](#_Toc479625338)

[4.4.2 逻辑设计 68](#_Toc479625339)

[4.4.3 物理设计 71](#_Toc479625340)

[4.5 系统结构设计 79](#_Toc479625341)

[4.6 系统功能设计 81](#_Toc479625342)

[4.6.1 用户登录界面设计 81](#_Toc479625343)

[4.6.2 原始数据操作模块设计 82](#_Toc479625344)

[4.6.3 鉴别流程模块设计 84](#_Toc479625345)

[4.7 实例操作说明 85](#_Toc479625346)

[4.8 本章小结 89](#_Toc479625347)

[5 结论与建议 91](#_Toc479625348)

[5.1 结论 91](#_Toc479625349)

[5.2 建议 92](#_Toc479625350)

[参 考 文 献 93](#_Toc479625351)

[攻读学位期间研究成果 98](#_Toc479625352)

[致 谢 100](#_Toc479625353)

# 诸 论

## 课题背景和研究目的及意义

### 课题背景

水污染是我国乃至世界面临的一个十分严重的环境问题，它直接影响着人类的生存和发展。保护环境是我国的一项基本国策，积极防治水污染是保护环境的一个重要环节。[1]据统计，我国每年约有1/3的工业废水和90％以上的生活污水未经处理就排入水域。据环保部门监控，目前全国城镇每天至少有l亿t污水未经处理直接排入水体。[2]

随着我国工业园区在各地大规模兴起，依靠管理部门对生产企业污水流量及污染程度的人工监控已经远远不能满足工业发展的需要。[3]我国城市污水处理领域已逐步应用在线监控系统,在线监控系统包括在线监控仪器、数据传输网、数据处理、应用设施和业务信息系统是集环境保护科学、在线监控、现代语音和数据通信、现代网络和信息系统为一体的新技术。[4]

国家和各省和地区越来越重视生态环境的保护，加强水污染的管理，但由于环境保护意识淡薄、利益驱动因素很多，如对污水监测数据造假和非标准的排水，导致严重的环境污染。因此投资于监控和监管，提供一个有效、实用先进的监控系统和解决方案，加强环境监控变得迫在眉睫。[5]

### 研究目的及意义

污水处理是现代社会发展和水资源保护不可缺少的组成部分。进行污水处理的在线监控对改善国家水环境，保障国家社会经济发展起着举足轻重的作用。国家要对全国1.8万个重点污染企业实施主要污染物排放总量控制和消减，以改善环境质量，工厂污水在主要污染物中又占主导地位．因此国家要求1.8万个污染大户要逐步安装污水综合监控分析系统，并实现计算机联网管理。[6]

目前，我国政府部门获取的城市排水与污水处理系统设施的相关数据建立在行政层面上报的基础上，存在上报数据科学严谨性差、缺乏可信度、现势性差以及数据关联分析能力弱等问题。[7]一些排放污水的企业在经济利益驱使下，采用各种手段对在线监测数据进行造假，例如在监测点稀释污水，私接暗管，篡改监测仪器重要参数等。这些不法行为导致污水监测数据失真，给监测部门的工作带来很大的麻烦。许多监测人员本身业务素质较低，缺少将数据进行科学有效的分析和处理的能力，导致数据的可信度不高，失去评价意义。

因此，为减轻虚假污水监测数据给环境监测工作带来的巨大压力，十分有必要开发一套可以鉴别出污水监测数据真伪系统，用于指导城市污水处理厂的日常管理，发现虚假数据，为上级监管部门提供有力的科技保证和技术支撑，对于提高城市污水处理厂的运行管理水平有着重要意义。

## 国内外研究现状

### 典型污水处理工艺研究现状

#### 间歇活性污泥法研究现状

间歇活性污泥法也称序批式活性污泥法(Sequencing Batch Reactor-SBR)，它由单个或多个SBR池组成，运行时，废水分批进入池中，依次经历5个独立阶段，即进水、反应、沉淀、排水和闲置。进水及排水用水位控制，反应及沉淀用时间控制，一个运行周期的时间依负荷及出水要求而异，一般为4～12h，其中反应占40%，有效池容积为周期内进水量与所需污泥体积之和。  
　　比连续流法反应速度快，处理效率高，耐负荷冲击的能力强;由于底物浓度高，浓度梯度也大，交替出现缺氧、好氧状态，能抑制专性好氧菌的过量繁殖，有利于生物脱氮除磷，又由于泥龄较短，丝状菌不可能成为优势，因此，污泥不易膨胀;与连续流方法相比，SBR法流程短、装置结构简单，当水量较小时，只需一个间歇反应器，不需要设专门沉淀池和调节池，不需要污泥回流，运行费用低。

宋小燕[8]采用间歇曝气序批式活性污泥法处理养猪沼液，研究在控温30℃ 、分步进水条件下的短程脱氮性能，发现分步进水间歇曝气序批式活性污泥法系统短程硝化稳定性好，脱氮效率高。[魏武强](http://www.cnki.net/KCMS/detail/%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20/kcms/detail/search.aspx?dbcode=CJFQ&sfield=au&skey=%e9%ad%8f%e6%ad%a6%e5%bc%ba&code=25375796;25468060;)[9]采用序批式活性污泥法（SBR）和人工湿地（CW）来确定最佳的操作参数的组合工艺，发现常温下，第三周期（SBR的运行周期为8 h：进水1 h，限制曝气3 h，沉淀1 h，1 h和流出，怠速运转2小时；连续的水力停留时间（HRT）为24.8 h，水力负荷为24.5m3/m2.d）是最好的循环模式。在这个周期，连续的效率最高，SBR的能耗最低。在低溶解氧条件下，化学需氧量（COD）和氨氮去除率没有明显影响。同时硝化反硝化现象明显，磷吸收明显。[SHIGENORI SHOJI](http://scholar.cnki.net/result.aspx?q=%e4%bd%9c%e8%80%85%3a(SHIGENORI+SHOJI)&uid=WEEvREcwSlJHSldRa1FhcTdWZDlsYkJhSFVidDZ3alBmcmdWcjBHWFA2bz0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4ggI8Fm4gTkoUKaID8j8gFw!!&UID=WEEvREcwSlJHSldRa1FhcTdWZDlsYkJhSFVidDZ3alBmcmdWcjBHWFA2bz0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4ggI8Fm4gTkoUKaID8j8gFw!!)[10]将超滤膜工艺结合间歇曝气序批式活性污泥法实现稳定的去除污水中的BOD、氮、磷，并采用实验室规模的厂SBR系统比较混合动力系统的影响，发现在混合系统的BOD和TN的去除率均高于SBR系统，而总磷去除率在这两个系统是相似的。

#### 吸附再生法研究现状

这种方式充分利用活性污泥的初期去除能力，在较短的时间里(10～40min)，通过吸附去除废水中悬浮的和胶态的有机物，再通过液固分离，废水即获得净化，BOD5可去除85%～90%左右。吸附饱和的活性污泥中，一部分需要回流的，引入再生池进一步氧化分解，恢复其活性;另一部分剩余污泥不经氧化分解即排入污泥处理系统。  
　　分别在两池(吸附池和再生池)或在同一池的两段进行。它适应负荷冲击的能力强，还可省去初次沉淀池。主要优点是可以大大节省基建投资，最适于处理含悬浮和胶体物质较多的废水，如制革废水、焦化废水等，工艺灵活。但由于吸附时间较短，处理效率不及传统法的高。

[陈孟林](http://www.cnki.net/KCMS/detail/%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20/kcms/detail/search.aspx?dbcode=CJFQ&sfield=au&skey=%e9%99%88%e5%ad%9f%e6%9e%97&code=06879544;10836257;06883404;06881511;06881586;)[11]研究了树脂吸附与H2O2-V2O5催化氧化再生法处理印染废水的影响因素以及树脂再生的工艺条件,发现树脂适宜的催化氧化再生工艺条件为:pH=3,H2O2的体积浓度为0.2g/L,V2O5的质量浓度为1.6g/L,室温下再生30 min;同时树脂的再生率随再生次数的增加而略有降低,但整体上再生效率均在90%以上。[梁鑫](http://www.cnki.net/KCMS/detail/%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20/kcms/detail/search.aspx?dbcode=CJFQ&sfield=au&skey=%e6%a2%81%e9%91%ab&code=27489906;11034910;)[12]以颗粒活性炭为原料,采用3种有机酸(草酸、酒石酸和柠檬酸)为改性剂制得改性活性炭，采用比表面积及孔径分析仪、扫描电子显微镜及傅里叶转换红外光谱仪考察改性对活性炭物化性质的影响，发现改性后活性炭BET比表面积及总孔容减小,表面有不均匀的粗糙的刻蚀痕迹,同时伴随有白色晶体颗粒生成, 活性炭对甲醇的吸附能均增大。[Saddig D. Jastaniah](http://scholar.cnki.net/result.aspx?q=%e4%bd%9c%e8%80%85%3a(Saddig+D.+Jastaniah)&uid=WEEvREcwSlJHSldRa1FhcTdWZDlsYkJhSFVidDZ3alBmcmdWcjBHWFA2bz0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4ggI8Fm4gTkoUKaID8j8gFw!!&UID=WEEvREcwSlJHSldRa1FhcTdWZDlsYkJhSFVidDZ3alBmcmdWcjBHWFA2bz0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4ggI8Fm4gTkoUKaID8j8gFw!!)[13]将氡浓度接近其最高水平的放置一段时间的自来水充满一个密闭容器，采用颗粒活性炭（GAC）处理，确定了最佳接触时间和最优氡浓度，得出了制备活性炭的最佳工艺。

#### 氧化沟研究现状

氧化沟是延时曝气法的一种特殊型式，它的平面像跑道，沟槽中设置两个曝气转刷，也有用表面曝气机、射流器或提升管式曝气装置的。曝气设备工作时，推动沟液迅速流动，实现供氧和搅拌作用。  
　　与普通曝气法相比，氧化沟具有基建投资省，维护管理容易，处理效果稳定，出水水质好，污泥产量少，还有较好的脱N、P作用，适应负荷冲击能力强等优点。

[刘珮](http://www.cnki.net/KCMS/detail/%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20/kcms/detail/search.aspx?dbcode=CJFQ&sfield=au&skey=%e5%88%98%e7%8f%ae&code=29001948;09067518;29001949;29001950;)[14]采用以可溶性淀粉为碳源的模拟生活污水，探索了实验室模拟氧化沟处理系统内污泥EPS的变化及其与污泥膨胀的关系，发现在低负荷运行的系统且缺乏细菌可快速容易降解碳源的条件下出现的污泥中EPS减少，并非是氧化沟污泥发生丝状膨胀的致因,丝状菌的优势繁殖才是根本原因，不能通过控制EPS来影响氧化沟低负荷下的污泥丝状膨胀。[Vasileios Diamantis](http://scholar.cnki.net/result.aspx?q=%e4%bd%9c%e8%80%85%3a(Vasileios+Diamantis)&uid=WEEvREcwSlJHSldRa1FhcTdWZDlsYkJhSFVidDZ3alBmcmdWcjBHWFA2bz0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4ggI8Fm4gTkoUKaID8j8gFw!!&UID=WEEvREcwSlJHSldRa1FhcTdWZDlsYkJhSFVidDZ3alBmcmdWcjBHWFA2bz0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4ggI8Fm4gTkoUKaID8j8gFw!!)[15]通过具有均匀曝气场的充气通道反应器观察废水处理系统中的流动分层和短路现象，发现分层流动是曝气量与曝气设备的布置密切相关。[Y.Nakamura](http://scholar.cnki.net/result.aspx?q=%e4%bd%9c%e8%80%85%3a(Y.++Nakamura)&uid=WEEvREcwSlJHSldRa1FhcTdWZDlsYkJhSFVidDZ3alBmcmdWcjBHWFA2bz0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4ggI8Fm4gTkoUKaID8j8gFw!!&UID=WEEvREcwSlJHSldRa1FhcTdWZDlsYkJhSFVidDZ3alBmcmdWcjBHWFA2bz0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4ggI8Fm4gTkoUKaID8j8gFw!!)[16]确定了沟中的好氧区和缺氧区的形成途径，并指出pH下降的问题可以通过在沟中同时实现硝化和反硝化和调节碱度平衡来解决。

#### 连续进水周期循环延时曝气活性污泥法(ICEAS) 研究现状

ICEAS反应器前部设有预反应区(占池容积的10%)。反应池由预反应区和主反应区组成，并实现连续进水，间歇排水。预反应区一般处在厌氧和缺氧状态，有机物在此被活性污泥吸附，该区还具有生物选择作用，抑制丝状菌生长，防止污泥膨胀。被吸附的有机物在主反应区内被活性污泥氧化分解。  
　　反应连续进水，解决了来水与间歇进水不匹配的矛盾。但该工艺沉淀效果较差、净化效果变差，易发生污泥膨胀，污泥负荷较低，反应时间长，设备容积增大，投资较大。

张立秋[17]对ICEAS工艺的最优工况的确定进行了研究，发现ICEAS采用连续进水时，最优工况为曝气2.5h，沉淀1h，排水1h；曝气过程中保持DO在3mg/L左右，沉淀30min后开始排泥，排泥0.5h结束，一周期为4.5h。在此运行工况下，ICEAS反应器既能去除有机物，又能达到脱氮除磷的效果。[Farham Karakani[18]为探究间歇循环延时曝气系统对水污染物特别是磷的去除能力，在德黑兰医科大学进行实验，实验所使用的反应器包括两个区：预反应区和主反应区。实验在三种不同的水力停留时间（HRT）和流量下进行，结果发现该方法不是一个专项除磷的过程，但去磷效果却近似专项除磷工艺效果，并且经济成本更低。](http://scholar.cnki.net/result.aspx?q=%e4%bd%9c%e8%80%85%3a(Farham+Karakani)&uid=WEEvREcwSlJHSldRa1FhcTdWZDlsYkJhSHBjTTRtc01aS1hHNjI5RUV6bz0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4ggI8Fm4gTkoUKaID8j8gFw!!&UID=WEEvREcwSlJHSldRa1FhcTdWZDlsYkJhSHBjTTRtc01aS1hHNjI5RUV6bz0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4ggI8Fm4gTkoUKaID8j8gFw!!)[Mortazavi S B](http://scholar.cnki.net/result.aspx?q=%e4%bd%9c%e8%80%85%3a(Mortazavi+S+B)&uid=WEEvREcwSlJHSldRa1FhcTdWZDlsYkJhSHBjTTRtc01aS1hHNjI5RUV6bz0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4ggI8Fm4gTkoUKaID8j8gFw!!&UID=WEEvREcwSlJHSldRa1FhcTdWZDlsYkJhSHBjTTRtc01aS1hHNjI5RUV6bz0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4ggI8Fm4gTkoUKaID8j8gFw!!)[19]探讨了ICEAS法对废水中SDS表面活性剂的去除效果，发现两小时的曝气时间足够去除范围为20mg/l至400mg/l的SDS表面活性剂，并且去除效果可超过98%，说明采用ICEAS法工艺能够处理含高浓度SDS表面活性剂废水。

#### 生物脱氮除磷工艺(A/A/O) 研究现状

污水首先进入厌氧池与回流污泥混合，在兼性厌氧发酵菌的作用下，废水中易生物降解的大分子有机物转化为聚磷菌可以吸收小分子有机物(如VFA)，并以PHB的形式贮存在体内，其所需的能量来自聚磷链的分解。随后，废水进入缺氧区，反硝化细菌利用废水中的有机基质对随回流混合液带入的NO3- 进行反硝化。废水进入好氧池时，废水中有机物的浓度较低，聚磷菌主要是通过分解体内的PHB而获得能量，供细菌增殖，同时将周围环境中的溶解性磷吸收到体内，并以聚磷链的形式贮存起来，随后以剩余污泥的形式排出系统。系统中好氧区的有机物浓度较低，正有利于该区中自养硝化菌的生长。  
　　厌氧、缺氧、好氧三种不同的环境条件和不同种类的微生物菌群的有机配合，能同时具有去除有机物、脱氮除磷的功能;工艺简单，水力停留时间较短;SVI一般小于100，不会发生污泥膨胀;污泥中磷含量高，一般为2.5%以上;厌氧-缺氧池只需轻缓搅拌，使之混合，而以不增加溶解氧为度;沉淀池要避免发生厌氧-缺氧状态，以避免聚磷菌释放磷而降低出水水质和反硝化产生N2而干扰沉淀;脱氮效果受混合液回流比大小的影响，除磷效果则受回流污泥中挟带DO和硝酸态氧的影响，因而脱氮除磷效果不可能提高。

郭小马[20]针对目前城镇污水低碳源脱氮除磷难，污泥处理处置难，城镇污水厂运行成本高等问题，提出一种新型组合式工艺改良A2O侧流除磷工艺，来解决城镇污水面临的难题。新型组合工艺应用改良A2O为处理反应器，结合侧流除磷技术，以城市污水处理厂曝气沉砂池出水为处理对象，探讨系统在长污泥龄(SRT)、低碳源、限氧曝气等条件下脱氮除磷效果，同时采用现代分子生物学技术解析侧流除磷过程中活性污泥微生物群落结构、组成与相对丰度变化。[Giorgio Mannina](http://scholar.cnki.net/result.aspx?q=%e4%bd%9c%e8%80%85%3a(Giorgio+Mannina)&uid=WEEvREcwSlJHSldRa1Fhb09jMjQwK0pZbEFqREU1YjVPOXRXUzdpMkVNdz0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4ggI8Fm4gTkoUKaID8j8gFw!!&UID=WEEvREcwSlJHSldRa1Fhb09jMjQwK0pZbEFqREU1YjVPOXRXUzdpMkVNdz0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4ggI8Fm4gTkoUKaID8j8gFw!!)[21]在开普顿大学通过比值不同的碳氮的膜生物反应器对碳和营养物质的去除，生物的呼吸活性，活性污泥特性及膜污染进行探究，发现碳和营养物质的去除，生物的呼吸活性随碳氮比值的减少而减少，活性污泥的特性随碳氮比值的减少而恶化，膜污染变得严重，过滤性能变差。[J. Carrera](http://scholar.cnki.net/result.aspx?q=%e4%bd%9c%e8%80%85%3a(J.++Carrera)&uid=WEEvREcwSlJHSldRa1Fhb09jMjQwK0pZbEFqREU1YjVPOXRXUzdpMkVNdz0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4ggI8Fm4gTkoUKaID8j8gFw!!&UID=WEEvREcwSlJHSldRa1Fhb09jMjQwK0pZbEFqREU1YjVPOXRXUzdpMkVNdz0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4ggI8Fm4gTkoUKaID8j8gFw!!)[22]通过对影响生物除磷的三个影响因素——进水COD / P比值、厌氧部分和污泥停留时间(SRT)进行探究，发现最影响生物除磷过程中的因素是进水COD／P比值，而厌氧部分影响力最小，并发现最佳的进水COD / P比值在41和48之间。

### 异常数据检测方法研究现状

#### 箱线图法研究现状

箱线图是一种用作显示一组数据分散情况资料的统计图。箱线图法利用数据中的五个统计量：最小值、第一四分位数（下四分位数）Q1、第二四分位数（中位数）Q2、第三四分位数（上四分位数）Q3与最大值来描述数据[23]。

箱线图作为描述统计的工具之一，其功能有独特之处，主要有以下几点：1. 直观明了地识别数据批中的异常值；2. 利用箱线图判断数据批的[偏态](http://wiki.mbalib.com/wiki/åæ)和尾重；3. 利用箱线图比较几批数据的形状。箱线图根据实际数据绘制，既不需要事先假定数据服从特定的概率分布，也没有对数据作任何限制性要求，能够真实、直观地表现数据形状的本来面貌。箱线图判断异常值的标准以四分位数和四分位距为基础，较多数据的变化对四分位数影响不大，检测结果比较客观。[24]

[王新峰](http://social.wanfangdata.com.cn/Locate.ashx?ArticleId=dzktaq201002027&Name=%e7%8e%8b%e6%96%b0%e5%b3%b0)[25]利用箱线图法对锦屏一级水电工程厂房区某排水廊道野外测量数据进行整理分析，发现箱线图的离群值可以有效地避免掩盖效应，清晰直观地指示裂隙的疏密展布，间距均值参数误差平均缩小到3.39%以下，相比概率累计分布曲线拟合得到的均值，其精度平均提高了46.97%，从而得出箱线图法是裂隙间距分析的有效手段。[张海燕](http://social.wanfangdata.com.cn/Locate.ashx?ArticleId=sjdz201504028&Name=%e5%bc%a0%e6%b5%b7%e7%87%95)[26]运用箱线图直观了解不同降水量下含水量发生显著性变化的土层深度，进一步使用秩和检验方法，利用本底值与降水影响情况下样本土壤含水量数值进行比较分析，发现在0.1显著性水平下，污染液不能达到含水层。[刘瑶](http://social.wanfangdata.com.cn/Locate.ashx?ArticleId=6725536&Name=%e5%88%98%e7%91%b6)[27]利用统计分析方法中探索性数据分析方法的箱线图法对长江干流南京段至上海段1995-2005年环境监测指标高锰酸盐指数、五日生化需氧量和氨氮的数据进行了分析，总结出了该河段的水质空间变化趋势，表明箱线图可以直观描述监测点位多年监测数据分布状况，并可以对比分析上下游水质的空间变化特征和异常值。

#### 格拉布斯法研究现状

格拉布斯法是以正态分布为前提，从多次等精度独立测量过程中剔除粗大误差的方法[28]。要有效地剔除[偶然误差](http://www.so.com/s?q=%E5%81%B6%E7%84%B6%E8%AF%AF%E5%B7%AE&ie=utf-8&src=wenda_link)，一般要测量10次以上，兼顾到精度和响应速度，取15次为一个单位。在取得的15个数据中，有些可能含有较大的误差，需要对它们分检，剔除可疑值，提高自适应度。

[池庆玺](http://social.wanfangdata.com.cn/Locate.ashx?ArticleId=cgjsxb200606070&Name=%e6%b1%a0%e5%ba%86%e7%8e%ba)[29]通过引入统计学中剔除异常数据的格拉布斯准则,将其与基于算数平均和递推估计的"一致性校验"算法相结合,提出了一种基于此准则的雷达信号分选方法,并围绕该方法中关键参数的不同选取给分选性能带来的影响,通过MATLAB的仿真分析验证了该方法应用于反辐射导弹(ARM)雷达信号分选的有效性和可行性。[邬晓光](https://vpn.njtech.edu.cn/,DanaInfo=njtech.zhizhen.com+s?sw=author%28%E9%82%AC%E6%99%93%E5%85%89%29&size=15&isort=0&x=0_729)[30]通过运用格拉布斯准则原理确定定额测定不正常数据来代替传统的由专家逐条记录分析判断的方法。利用误差原理中的格拉布斯判别准则，较好地判别所测数据中的坏值，方法简单、可靠，且易编写程序，从而为判别大量观测数据中的坏值提供了一通用性较强的平台，为编制可靠的定额创造了必要条件。[邬蓉蓉](http://social.wanfangdata.com.cn/Locate.ashx?ArticleId=dianqjs201611024&Name=%e9%82%ac%e8%93%89%e8%93%89)[31]提出并应用了基于格拉布斯准则检验变压器油气监测数据随机误差的自适应处理方法，扭转了缺少自动甄别正态分布下监测数据随机误差手段的被动局面，降低了误告警的概率, 改进了监测系统的纠错能力，提升了监测预警的效果。[郭立水](http://social.wanfangdata.com.cn/Locate.ashx?ArticleId=mzjs201603008&Name=%e9%83%ad%e7%ab%8b%e6%b0%b4)[32]从求出拟似异常值的、格拉布斯临界值的查询、异常数据的识别与排除、重复识别与排除异常数据等方面简单介绍了格拉布斯准则法及其现场使用，并采用格拉布斯法对在线灰分仪的异常数据进行处理，用格拉布斯法排除灰分仪异常样本比均方差或均方根差算法更具优越性。

#### t检验法研究现状

t检验法主要用于样本含量较小（例如n<30），总体标准差σ未知的[正态分布](http://wiki.mbalib.com/wiki/æ­£æåå¸)资料。[33]它是用T分布理论来推断差异发生的概率，从而判定两个平均数的差异是否显著。[34]

t检验分为单总体检验和双总体检验。单总体t检验时检验一个样本平均数与一个已知的[总体平均数](http://baike.so.com/doc/3880200.html)的差异是否显著。当[总体分布](http://baike.so.com/doc/2213054.html)是[正态分布](http://baike.so.com/doc/5016783.html)，如总体[标准差](http://baike.so.com/doc/3092013.html)未知且[样本容量](http://baike.so.com/doc/2291610.html)小于30，那么样本平均数与总体平均数的[离差](http://baike.so.com/doc/835770.html)[统计量](http://baike.so.com/doc/3632362.html)呈t分布。双总体t检验是检验两个样本平均数与其各自所代表的总体的差异是否显著。双总体t检验分为两种情况，一是独立样本t检验，一是配对样本t检验。

[彭定赟](http://social.wanfangdata.com.cn/Locate.ashx?ArticleId=6567897&Name=%e5%bd%ad%e5%ae%9a%e8%b5%9f)[35]在多变量回归分析中,在单个系数的显著性检验时采用t检验,而在多个系数和全部斜率系数的显著性检验时采用F检验这一普遍现象,提出了双变量回归的系数显著性检验用t检验,多变量回归的系数显著性检验不分单个系数和多个系数,全部用F检验的建议，并通过理论分析和实证研究两方面,论证了本文关于理顺t检验和F检验建议的合理性。[李简微](http://social.wanfangdata.com.cn/Locate.ashx?ArticleId=yxmxmr-zx2015051277&Name=%e6%9d%8e%e7%ae%80%e5%be%ae)[36]采用多组实例与极端数据对比研究配对 t 检验与成组 t 检验的相关数据，对数据相关性与结果的关系进行分析，发现t 检验方法的优选只取决于 r 值，与相关性无关。两组数据若具有良好的相关性，需要对两种 t 检验方法提高重视，以免在选择方法过程中难以得出正确结论。[董秀玥](http://social.wanfangdata.com.cn/Locate.ashx?ArticleId=slyyxzz201001004&Name=%e8%91%a3%e7%a7%80%e7%8e%a5)[37]通过多组实例数据与极端数据配对t检验与成组t检验结果的比较研究发现，决定优选t检验方法的决定因素不在于相关性好差,而主要取决于r值正负。当两组数据相关性越好时,越要重视两种t检验方法的选择,否则会得出错误的结论。

#### χ2检验法和F检验法研究现状

χ2检验是检验服从正态分布总体N的某个样本的波动性是否显著的一种假设检验方法。F检验是检验两个正态随机变量的总体方差是否相等的一种假设检验方法。χ2检验法和F检验法很类似都是通过检验分布的波动性来推断样本是否来自某个拟定的总体，以检验某种事件出现的理论次数与经验次数之间的差异的显著程度的。

t检验主要用于观察样本观测均值的变化情况；而χ2检验和F检验主要用于观察样本观测值的分散情况。将二者联合运用，可以样本观测结果的变化是否超出了控制限[38]。

[陈创买](http://social.wanfangdata.com.cn/Locate.ashx?ArticleId=rdqxxb200706024&Name=%e9%99%88%e5%88%9b%e4%b9%b0)[39]在时序方差分析周期外推预测、双重检验逐步回归预测及其他相关问题的分析中,发现F检验法可明显提高对因子显著性问题的判别能力和气候预测水平。应用χ2检验和F检验的统计方法，通过定期对计量标准(尤其是准确度较高且较重要的计量标准)进行核查，判断计量标准是否处于统计控制状态。只有测量标准始终处于受控状态，才能确保该计量标准量值的稳定，保证测量结果的可靠性。

### 基于B/S网络结构模式鉴别系统研究现状

B/S（Browser/Server，浏览器/服务器）模式又称B/S网络结构模式。它是随着Internet技术的兴起，对C/S模式应用的扩展。在这种结构下，用户工作界面是通过IE浏览器来实现的。B/S模式最大的好处是运行维护比较简便，能实现不同的人员，从不同的地点，以不同的接入方式（比如LAN, WAN, Internet/Intranet等）访问和操作共同的数据。

共同的数据采用Oracle12数据库存储污水检测指标数据。ORACLE[数据库系统](http://baike.so.com/doc/3054063-3219400.html)是美国ORACLE公司(甲骨文)提供的以[分布式数据库](http://baike.so.com/doc/6591740-6805519.html)为核心的一组软件产品，是目前最流行的客户/服务器(CLIENT/SERVER)或B/S[体系结构](http://baike.so.com/doc/6477885-6691588.html)的数据库之一。ORACLE数据库是目前世界上使用最为广泛的[数据库管理系统](http://baike.so.com/doc/2035924-2154168.html)，作为一个通用的[数据库系统](http://baike.so.com/doc/3054063-3219400.html)，它具有完整的[数据管理](http://baike.so.com/doc/5392609-5629451.html)功能；作为一个关系数据库，它是一个完备关系的产品；作为[分布式数据库](http://baike.so.com/doc/6591740-6805519.html)它实现了[分布式处理](http://baike.so.com/doc/6591676-6805455.html)功能。

胡锡春[40]将Web服务器应用于污水处理中，实现运行数据实时发布，并接受用户的控制命令的输入，实现了操作的便捷，又保证了系统的可靠性。郝雅青[41]开发了基于Web的污水流量排放监控及管理系统，构建了分布式浏览器、Web服务器、数据库服务器三层信息发布系统的B/S架构污水排放监控及管理系统。刘爽[42]对基于Web的远程诊断管理系统关键技术的分析，在总结东北地区玉米低温冷害领域专家经验和知识的基础上，初步实现了一个基于Web技术的玉米低温冷害远程诊断与管理系统，并且与农业环境无线远程监控系统相结合，使之具有实时预测功能。陈胜春[43]分析了城市重点污染源管理工作中存在的问题，确定了系统开发环境、系统的体系结构和系统的功能结构，采用Aptana Studio对系统进行设计，并使用Web窗体技术、数据访问技术、SQL Server 2000数据库和javascript编程语言等来实现。宋庠明[44]提出了基于CAN总线及Web的污水处理在线监控系统，分析了系统的体系蛄构、模块功能层次以及在线监控中的一些关键技术，实现了污水处理在线监控的异地化、网络化，并有传输数据可靠、实时性强和资源共享等优点。张学雷[45]等基于Oracle和SQL Server 2005数据库平台，采用集群、数据仓库、数据挖掘等技术开发出安徽省环境监控数据库，为污染源自动监控系统、地表水自动监控系统、空气质量自动监控系统、噪声自动监控系统、环保业务系统、环保门户办公等系统提供强有力的基础支持,有效提高海量数据的处理能力。

## 主要研究内容及技术路线

### 研究目标

本课题主要着眼于控制企业污水超标排放、伪造篡改环境监控数据等问题，选取污水处理厂污水监测指标数据为研究对象，综合运用数理统计的原理，结合B/S网络结构模式，构建污水监测即污水监测数据真伪鉴别系统，用于指导城市污水处理厂的日常管理，发现虚假数据，为上级监管部门提供有力的科技保证和技术支撑，对于提高城市污水处理厂的运行管理水平有着重要意义。

### 研究内容

#### 基于数理统计方法的污水监测数据真伪鉴别流程技术研究

基于格拉布斯法对污水监测数据时观测值进行异常值的判断和修正，初步鉴别并剔除异常数据，后续鉴别流程均在此修正数据的基础上进行，因此需保证较高可信度；基于《城市污水处理厂污染物排放标准》(GB18918-2002)对修正后的时数据进行达标筛选，不满足排放标准的数据剔除。基于上述两步筛选和鉴定，企业指定月份内的监测数据已经具有很高的可信度，至此可进行数据的纵横对比。

实现数据纵向对比，即基于t检验法对污水监测指标的月均值进行检验，验证抽检数据样本均值是否在整体样本均值的接收域；基于χ2检验法判断样本相对于整体的波动性是否有显著变化，从而判断每个月数据的离散程度是否符合整体样本的离散程度。实现数据横向对比，即基于F检验法在给定的显著水平下，确定F的接受域，从而确定离散程度相近的样本；基于箱线图法逐渐缩小匹配样本范围，直到筛选出一个或几个对比对象为止；基于采用χ2检验法对确定的匹配对象与待测对象的差值对象进行波动性进行分析，观察离散度是否发生大的变动，

最终判断数据的真实性。

#### 基于B/S网络结构模式的污水监测数据真伪鉴别系统构建

确定污水监控指标，熟知各种指标国家排放规定，理解污水处理厂典型污水处理工艺，了解监控部门监控手段和检测指标。明确任务目标，分析需求，实现功能和数据库设计。基于Java语言平台实现系统后台的搭建，基于html+css+javascript实现系统前端功能搭建，基于Oracle数据库实现污水监测指标原始数据和鉴别数据的储存及修改，构建人机交互网页，实现鉴别流程的逻辑功能。

系统主要分为两大模块：原始数据增删查改模块和鉴别流程模块。原始数据增删查改模块实现数据的实时存放，并保存提交数据的原始记录，为后续的鉴别流程提供参照对象。尤其重要的是，对于不同的系统使用用户数据库中设定其不同的权限，只有作为数据库管理者的用户才能在此模块修改原始数据，这可极大的避免工作人员的人为因素影响。鉴别流程模块实现污水监测数据真伪鉴别功能，每个鉴别流程占用一个操作模块，层层递进。整个鉴别流程共四个操作模块：格拉布斯法修正时观测值模块、筛选达标日观测值并修正模块、数据纵向对比模块和数据横向对比模块。每个操作模块的鉴别结果均可以实时查看，数据纵向对比模块和数据横向对比模块在格拉布斯法修正时观测值模块和筛选达标日观测值并修正模块的操作基础上进行。整套流程最终完成对污水监测数据真伪的鉴别。

### 技术路线



图 1‑1 技术路线图

Figure 1‑1 Technology roadmap

# 研究对象与方法

## 研究对象

### 待测污水处理厂污水处理工艺介绍

本文以南方某城市的污水处理厂作为研究对象，该污水处理厂产生的污水为混合废水，主要以化学工业为主，废水主要是染料、医药、农药中间体及其它精细化工产品生产过程中排出的工业废水。水质极其复杂，色度深、酸度大，有机有毒物质含量高。采用传统的物理化学和生化处理工艺对开发区混合废水进行处理，很难达到高效，节能、低成本以及环保等需求，因此以传统物理化学和生化处理工艺为基础，该污水处理厂采用改进型工艺，出水水质符合《城市污水处理厂污染物排放标准》(GB18918-2002)规定的一级B标准。该工艺技术成熟可靠，比改进前的污水处理工艺的脱氮除磷的效率高，且操作简便，可实现自动控制，经济实用，在南方城市已经得到广泛应用。

该污水处理厂的改进型工艺流程，包括废水调节池、厌氧反应池、多级生化池、好氧接触氧化池和砂滤池；所述废水调节池、厌氧反应池、多级生化池、好氧接触氧化池和砂滤池依次连通，如所示。

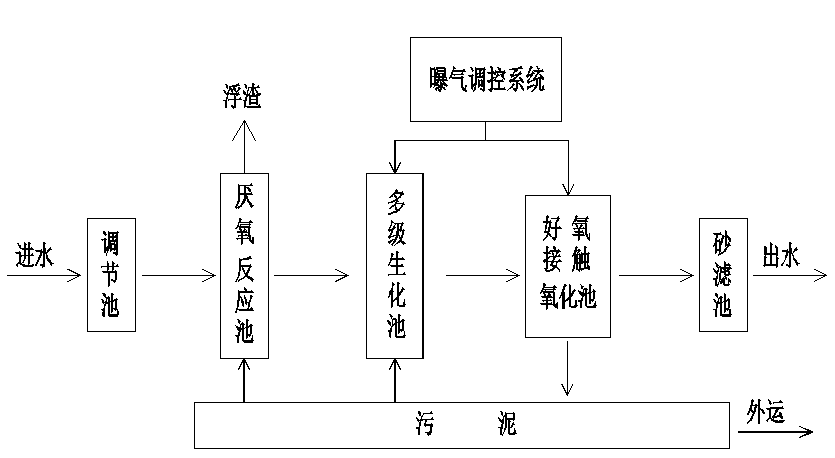


图 2‑1 改进型工艺流程图

Figure 2‑1 Improved process flow chart

废水调节池包括进水管和出水管，用于调节混合废水的水质和水量。

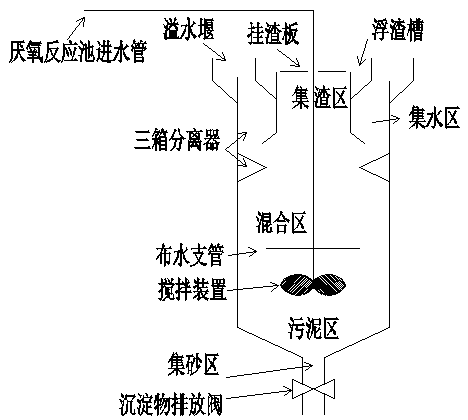


图 2‑2 厌氧反应池结构示意图

Figure 2‑2 Schematic diagram of anaerobic reactor

厌氧反应池（如所示）包括进水管和用于排出处理后水的出水管，厌氧反应池从下至上依次为集砂区、污泥区、混合区和分离区。厌氧反应池包括进水管和用于排出处理后水的出水管，厌氧反应池从下至上依次为集砂区、污泥区、混合区和分离区。分离区包括集水区和位于集水区内的集渣区。厌氧反应池的污泥区和混合区之间设置有搅拌装置，搅拌装置的上方设有布水支管，布水支管连接进水管。分离区内设有三相分离器，三相分离器包括导流板和位于导流板下方与导流板配合使用的三角导流环，三角导流环安装在厌氧反应池的内壁上，导流板的上部与分离区的形状相同，导流板的下部呈喇叭状，导流板的下部的内径大于三角导流环的内径。分离区外壁的上部设有溢水堰，溢水堰与出水管相连。集渣区布设有刮渣板和浮渣槽。废水从下往上溢时，水与浮渣一起通过三角导流环进入导流板的下部，浮渣继续往上进入集渣区，水通过导流板与三角导流环之间的间隙进入集水区。为了废水处理的效果更好，布水支管设置成同心圆形状或十字形状，布水支管上具有水平辐射出水口。由于水、沉淀物的密度不同，在三相分离器作用下实现分离。为了排出处理后的沉淀物，所述的厌氧反应池底部设有沉淀物排放阀。

厌氧反应池的出水管与多级生化池的进水管连通。

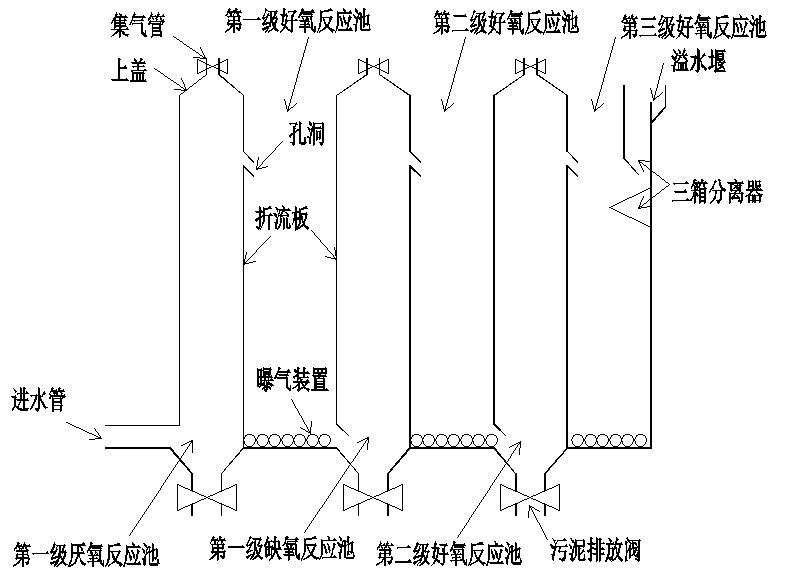


图 2‑3 多级生化池结构示意图

Figure 2‑3 Schematic diagram of multistage biochemical pool

多级生化池（如所示）包括通过折流板分隔成的厌氧反应池，好氧反应池和缺氧反应池，第一级厌氧反应池首端设有用于供入废水的进水管，第一级好氧反应池末端与第一级缺氧反应池首端连通，第二级好氧反应池末端与第二级缺氧反应池首端连通；各级反应池的进水一侧折流板的下部设置有45度的转角，以避免水流进入时产生的冲击作用，从而起到缓冲水流和均匀布水的作用；第三级好氧反应池末端设有三相分离器和溢水堰，溢水堰连接出水管；第一级厌氧反应池和第一级好氧反应池，第一级缺氧反应池和第二级好氧反应池，第二级缺氧反应池和第三级好氧反应池之间均用带孔洞的折流板隔离，好氧反应池的底部均布置有曝气装置，底部设计成锥形结构，锥形结构连接污泥排放阀；多级生化池的第一级厌氧反应池，第一级缺氧反应池，第二级缺氧反应池的上盖设计成圆锥形结构，圆锥形结构顶端都设有甲烷废气集气管。

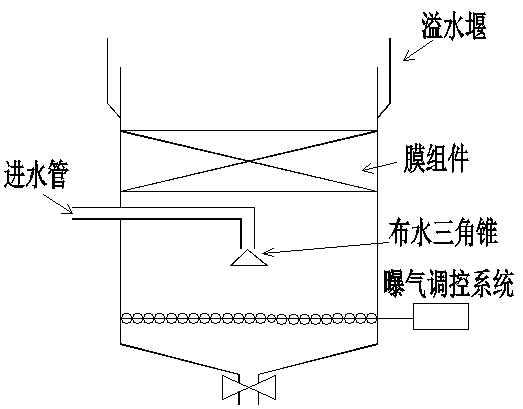


图 2‑4 好氧接触氧化池结构示意图

Figure 2‑4 Schematic diagram of aerobic oxidation tank

好氧接触氧化池(如2-4所示)内中下部设置进水管，进水管下部设有布水三角锥；布水三角锥下部设有曝气调控系统，曝气调控系统包括曝气盘、鼓风机和溶解氧测量调控装置；曝气盘是均匀设置有微孔的微孔式曝气盘；曝气盘通过曝气管连接鼓风机，鼓风机设置在好氧接触氧化池外，好氧接触氧化池的上部、废水水面下设置溶解氧测量调控装置，溶解氧测量调控装置根据氧容量调控鼓风机工作；进水管上部内置有膜组件；好氧接触氧化池的出口处布设有溢水堰，此过程处理后的水进入砂滤池。

好氧接触氧化池中的污泥一部分通过污泥回流泵回流到厌氧反应池中进行厌氧释磷，一部分作为剩余污泥。

该工艺流程的废水处理过程如下：

①混合废水通过进水管进入废水调节池调节水质和水量；

②然后污水通过进水管进入厌氧反应池分离区的中下部；位于厌氧反应池进水管下方的搅拌装置使污水中的固体物产生摩擦，去除固体物上的其他污染物；污水中的上浮物在浮力作用下上升，在厌氧反应池三相分离器的作用下上升至集渣区，在刮渣板的作用下，浮渣进入浮渣槽并被清理外运；沉淀物在重力的作用下下沉到厌氧反应池下部的集砂区，通过厌氧反应池底部的沉淀物排放阀排出；分离处理后的水在厌氧反应池三相分离区导流板作用下进入厌氧反应池集水区，通过溢水堰、出水管和连接管连通多级生化池的进水管；

③废水通过的第一级厌氧反应池的进水管进入多级生化池的下部；废水进入多级生化池后沿折流板上下前进。位于各级好氧区下方的曝气盘产生大量细小气泡使污水中的固体物进一步摩擦，同时把废水中的大分子物质氧化成易于吸收和吸附的小分子物质。废水依次通过多级生化池的每个反应池的污泥床，反应池中的污泥随着废水的上下流动和沼气上升的作用而运动，折流板的阻挡作用和污泥自身的沉降作用又使污泥的流速降低，因此大量的污泥都被截留在反应池中，反应池中的微生物与废水中的有机物充分接触；多级生化池中的异养菌将[废水](http://baike.baidu.com/view/326888.htm)中的淀粉、纤维、碳水化合物等悬浮[污染物](http://baike.baidu.com/view/1036818.htm)和可溶性有机物水解为有机酸，使大分子有机物分解为小分子有机物，不溶性的有机物转化成可溶性有机物；

④通过多级生化池反应后的废水在第三级好氧反应池末端设有的三相分离器实现泥、水、甲烷气的分离，污泥在重力的作用下下沉到多级生化池的下部，多余的污泥通过底部的污泥排放阀排出；多级生化池产生的甲烷废气通过反应池顶部集气管收集排放；废水通过溢水堰、出水管和连接管连通好氧接触氧化池的进水管；

⑤废水通过进水管进入好氧接触氧化池的中下部，在布水三角锥的作用下均匀布水，所述的曝气盘是均匀设置有微孔的微孔式曝气盘，产生大量的微气泡，所述溶解氧测量调控装置根据氧容量调控鼓风机工作，确保好氧接触氧化池水中的溶解氧大于2mg/L，处理后的废水通过溢水堰和出水管流出；

⑥好氧接触氧化池的沉淀污泥一部分污泥回流到厌氧反应池中，一部分作为剩余污泥，池中反应后的水通过出水管进入砂滤池，过滤后达标排放；

⑦厌氧反应池、多级生化池和好氧接触氧化池产生的剩余污泥脱水后外运。

### 污水监测数据指标及来源

通过对该污水处理厂的实际考察，以及查阅相关文献，咨询行业专家，了解到厂区日常运行状况和进出水质情况，并收集了该污水处理厂部分历史污水处理出水水质指标数据，主要包括COD浓度、pH值、SS浓度、氨氮浓度、总氮浓度、总磷浓度等。查得《城市污水处理厂污染物排放标准》(GB18918-2002)规定的一级B标准，如所示。

表 2‑1 污水处理厂出水水质设计要求(单位：mg/L)

Table 2‑1 Design requirements for effluent quality of sewage treatment plant (unit: mg/L)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 指标名称 | COD | pH | SS | 氨氮 | TN | TP |
| 超标浓度 | 60 | 6.0~9.0 | 20 | 8(15) | 20 | 1.0 |

据了解，目前对于国控企业要求每天对某一监测点每两小时监测一次，一天可反馈12个时观测值。对于一般性企业没有硬性规定，通常四小时监测一次。

## 异常检测研究方法

### 格拉布斯法

格拉布斯法是以正态分布为前提，从多次等精度独立测量过程中剔除粗大误差的方法。[46]假设从总体中抽取一个样本X1,X2…Xi…Xn服从正态分布，对其进行从小到大排列，使X1≤X2…≤Xi…≤Xn，假设Xn是需要检验判别的异常数据，S是样本标准差，对应的概率密度为：

所以，

(2-1)

其中，α(0<α<1)为显著性水平，可根据实际情况取值。

根据顺序统计原则，给出标准化顺序统计量g.

当Xn为最小值x1时，则.

当Xn为最小值xn时，则.

式S

根据格拉布斯统计量的分布，在指定的显著水平 (一般=0.05)下，求得判别可疑值的临界值，格拉布斯的判别标准为： ，则可疑值Xn是异常的，应予舍去。

称为格拉布斯系数，其值列于中。

表 2‑2 格拉布斯系数

Table 2‑2 Grubbs ' coefficient

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| n | α | | n | α | | n | α | | n | α | |
| 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 |
| 3 | 1.15 | 1.15 | 10 | 2.41 | 2.18 | 17 | 2.78 | 2.47 | 24 | 2.99 | 2.64 |
| 4 | 1.49 | 1.46 | 11 | 2.48 | 2.24 | 18 | 2.82 | 2.50 | 25 | 3.01 | 2.66 |
| 5 | 1.75 | 1.67 | 12 | 2.55 | 2.29 | 19 | 2.85 | 2.53 | 30 | 3.10 | 2.74 |
| 6 | 1.94 | 1.82 | 13 | 2.61 | 2.33 | 20 | 2.88 | 2.56 | 35 | 3.18 | 2.81 |
| 7 | 2.10 | 1.94 | 14 | 2.66 | 2.37 | 21 | 2.91 | 2.58 | 40 | 3.24 | 2.87 |
| 8 | 2.22 | 2.03 | 15 | 2.70 | 2.41 | 22 | 2.94 | 2.60 | 50 | 3.34 | 2.96 |
| 9 | 2.32 | 2.11 | 16 | 2.74 | 2.44 | 23 | 2.96 | 2.62 | 100 | 3.59 | 3.17 |

通过格拉布斯表可得值，根据异常数据Xn与平均值的残差是否大于倍样本标准差来判断异常数据是否应当剔除。[47]利用格拉布斯法每次只能舍去一个可能值，若有两个以上的可疑数据，应该逐个剔除，舍弃第一个数据后，检测次数变为n-1，以此为基础再判别第二个可疑数据是否应该舍弃。每次均值和标准偏差要重新计算，再决定取舍。[48]

如下做简单算例进行解释说明：

运用格拉布斯法对数列N:8.2、5.4、14.0、7.3、4.7、9.0、6.5、10.1、7.7、6.0进行异常值检测。

排列数据：将上述测量数据按从小到大的顺序排列，得到4.7、5.4、6.0、6.5、7.3、7.7、8.2、9.0、10.1、14.0。可以肯定，可疑值不是最小值就是最大值。

①计算平均值 和标准差S： ＝7.89；标准差S＝2.704。计算时，必须将所有10个数据全部包含在内。

②计算偏离值：平均值与最小值之差为7.89－4.7＝3.19；最大值与平均值之差为14.0－7.89＝6.11。

③确定一个可疑值：比较起来，最大值与平均值之差6.11大于平均值与最小值之差3.19，因此认为最大值14.0是可疑值。

④计算值：＝(Xi－ )/S；其中i是可疑值的排列序号——10号。

因此＝(X10－ )/S＝(14.0－7.89)/2.704＝2.260。由于 X10－ 是残差，而S是标准差，因而可认为是残差与标准差的比值。下面要把计算值与格拉布斯表给出的临界值比较，如果计算的值大于表中的临界值，则能判断该测量数据是异常值，可以剔除。其中，临界值与两个参数有关：检出水平α 和测量次数n。

⑤定检出水平α：如果要求严格，检出水平α可以定得小一些，例如定α＝0.01，那么置信概率P＝1－α＝0.99；如果要求不严格，α可以定得大一些，例如定α＝0.10，即P＝0.90；通常定α＝0.05，P＝0.95。

⑥查格拉布斯表获得临界值：根据选定的P值(此处为0.95)和测量次数n（此处为10），查格拉布斯表，横竖相交得临界值＝2.18。

⑦比较计算值和临界值：＝2.260，＝2.18，＞。

⑧判断是否为异常值：因为＞，可以判断测量值14.0为异常值，将它从10个测量数据中剔除。

⑨余下数据考虑：剩余的9个数据再按以上步骤计算，如果计算的＞，仍然是异常值，剔除；如果＜，不是异常值，则不剔除。本例余下的9个数据中没有异常值。

对固定条件下的一种测量，理论上可以无限次测量下去，可以得到无穷多的测量数据，这些测量数据构成一个容量为无限大的总体；或者换一个角度看，本来就存在一个包含无穷多测量数据的总体。[49]实际的测量只不过是从该无限大总体中随机抽取一个容量为n(例如n＝10)的样本。这种样本也可以有无数个，每个样本相当于总体所含测量数据的不同随机组合。样本中的正常值应当来自该总体。通常的目的是用样本的统计量来估计总体参量。总体一般假设为正态分布，样本中的正常值应当属于同一总体；而异常值有两种情况：第一种情况异常值不属于该总体，抽样抽错了，从另外一个总体抽出一个(一些)数据，其值与总体平均值相差较大；第二种情况异常值虽属于该总体，但可能是该总体固有随机变异性的极端表现，比如说超过3σ的数据，出现的概率很小。用统计判断方法就是将异常值找出来，舍去。

然而每种统计检验法都会犯两种错误。[50]第一种错误：将本来不属于该总体的、第一种情况的异常值判断出来舍去，不会犯错误；将本来属于该总体的、出现的概率小的、第二种情况的异常值判断出来舍去，就会犯错误。第二种错误：还有一种情况，不属于该总体但数值又和该总体平均值接近的数据被抽样抽出来，统计检验方法判断不出它是异常值，就会犯另外一种错误。但在所有方法中，格拉布斯法犯这两种错误的概率最小。

本研究基于Java语言在服务器端编程实现格拉布斯法的计算流程，由计算机完成不同显著水平下的检测计算，高效修正原始监测数据，保障后续检测流程中监测数据的可靠度。

### 箱线图法

箱线图是一种用作显示一组数据分散情况资料的统计图。[51]箱线图法利用数据中的五个统计量：最小值、第一四分位数（下四分位数）Q1、第二四分位数（中位数）Q2、第三四分位数（上四分位数）Q3与最大值来描述数据。[52]

中位数：将一组数据按大小顺序排列。处在最中间位置的一个数叫做这组数据的中位数，如果n为奇数，则中位数为中间值；如果n为偶数，则中位数为中间两个值的平均数。

四分位数：将一组数据按大小顺序排列后并分成四等份。处于三个分割点位置的数就是四分位数。最小的四分位数称为下四分位数(Q1)，表示在所有数值中，有四分之三大于下四分位数， 四分之一小于下四分位数(即第25百分位数)。最大的四分位数称为上四分位数(Q3)，表示所有数值中，有四分之一大于上四分位数， 四分之三小于上四分位数(即第75百分位数)。中点位置的四分位数就是中位数(Q2，即第50百分位数)。

箱线图可用于鉴别数据中的异常值，包括离群值和极端值，鉴别原理如下：假设从总体中抽取一个样本X1,X2…Xi…Xn,对其进行从小到大排列，使X1≤X2…≤Xi…≤Xn，设Qk为第k四分位数，则有：

Qk = (2-2)

计算四分位距IQR，内限值区间（Q1+1.5IQR和Q3-1.5IQR），在区间外的值即视为异常值。[53]

四分位距(IQR)：四分位距为第一个四分位数和第三个四分位数的差。IQR=Q3 - Q1，也就是观测值(箱子)的范围。

触须：从箱体边缘延伸出去的直线称为触须，分别位于Q1+1.5×IQR和Q3-1.5×IQR的位置，在没有异常值的情况下，上下触须即为该批数据的最大值和最小值。

异常值：如果一个值位于箱体外面(大于Q3或小于Q1)，并且距离相应边界大于1.5倍的IQR，那么这个点就被认为是一个温和异常值，用“o”来标识；如果距离相应边界大于3倍的IQR，则称为极端异常值。用“\*”来标识．异常值应位于上下触须以外的位置。平均数是反映数据集中趋势的一项指标，用“+”表示。

箱线图的绘制步骤：

①画数轴。度量单位大小和数据批的单位一致，起点比最小值稍小，长度比该数据批的全距稍长。

②画一个矩形盒。两端边的位置分别对应数据批的上下四分位数(Q3和Q1)。在矩形盒内部中位数位置(Q2)画一条线段为中位线，并标出平均数。

③在Q1-1.5×IQR和Q3+1.5×IQR处画两条与中位线一样的线段，即上下触须，这两条线段亦为异常值截断点，称其为内限。

④从矩形盒两端边向外各画一条线段直到不是异常值的最远点(最大值和最小值)，表示该批数据正常值的分布区间。

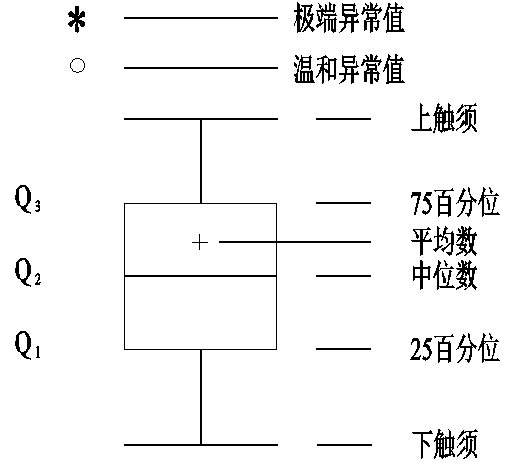
⑤在Q1-3×IQR和Q3+3×IQR处画两条线段，称其为外限。在内限与外限之间用“o”标出温和异常值，在外限以外用“\*”标出极端异常值。相同值的数据点并列标出在同一数据线位置上，不同值的数据点标在不同数据线位置上。

图 2‑5 箱线图

Figure 2‑5 Box diagram

至此一组数据的箱线图便绘出来了()。箱线图的特点是以一图涵盖5个位置统计量，5条横线由上到下分别是．最大值或上触须，上四分位数，中位数，下四分位数与最小值或下触须。

箱线图是数据分布特征直观简洁的表示方法，用于反映一组或多组连续型定量数据分布的中心位置和散布范围。[54]中位数反映了该组数据向中心分布的集中性，越向中位数集中，表示这组数据集中的程度或水平越高。全距（最大值与最小值之差）和四分位距则反映了数据的分散趋势。异常值反映了极端数据的分布情况。如果中位数值线近似位于四分位距箱体的中间，上下触须在长度上相近，可以看出数据具有明显的对称性；中位数不位于四分位距箱体的中间，或某一触须显著地比另一个长，那么数据是偏态的。[55]

在顺序排列样本后，使用箱型图法判断异常值Xe，Xe必然首先出现在最大值或最小值的位置，即Xe∈{ X1 , Xn }。确定了异常值的位置根据样本观测值即可定位出虚假样本。剔除异常值后，进行迭代运算，即可修正样本数据。

箱线图作为描述统计的工具之一，其功能有独特之处，主要有以下几点：

1.直观明了地识别数据批中的异常值

一批数据中的异常值值得关注，忽视异常值的存在是十分危险的，不加剔除地把异常值包括进数据的计算分析过程中，对结果会带来不良影响。[56]箱线图提供了识别异常值的一个标准：异常值被定义为小于Q1-1.5×IQR或大于Q3+1.5×IQR的值。这与识别异常值的经典方法有些不同。众所周知，基于[正态分布](http://wiki.mbalib.com/wiki/%E6%AD%A3%E6%80%81%E5%88%86%E5%B8%83)的3σ法则或z分数方法是以假定数据服从正态分布为前提的，但实际数据往往并不严格服从正态分布。它们判断异常值的标准是以计算数据批的均值和[标准差](http://wiki.mbalib.com/wiki/%E6%A0%87%E5%87%86%E5%B7%AE)为基础的，而均值和标准差的耐抗性极小，异常值本身会对它们产生较大影响，这样产生的异常值个数不会多于总数0.7%。[27]显然，应用这种方法于非正态分布数据中判断异常值，其有效性是有限的。箱线图的绘制依靠实际数据，不需要事先假定数据服从特定的分布形式，没有对数据作任何限制性要求，它只是真实直观地表现数据形状的本来面貌；另一方面，箱线图判断异常值的标准以[四分位数](http://wiki.mbalib.com/wiki/%E5%9B%9B%E5%88%86%E4%BD%8D%E6%95%B0)和四分位距为基础，四分位数具有一定的耐抗性，多达25%的数据可以变得任意远而不会很大地扰动四分位数，所以异常值不能对这个标准施加影响，箱线图识别异常值的结果比较客观。[58]

2.利用箱线图判断数据批的[偏态](http://wiki.mbalib.com/wiki/%E5%81%8F%E6%80%81)和尾重

比较标准正态分布、不同自由度的t分布和非对称分布数据的箱线图的特征，可以发现：对于标准[正态分布](http://wiki.mbalib.com/wiki/%E6%AD%A3%E6%80%81%E5%88%86%E5%B8%83)的大样本，只有 0.7%的值是异常值，中位数位于上下四分位数的中央，箱线图的方盒关于中位线对称。选取不同自由度的t分布的大样本，代表对称重尾分布，当t分布的自由度越小，尾部越重，就有越大的概率观察到异常值。[59]以[卡方分布](http://wiki.mbalib.com/wiki/%E5%8D%A1%E6%96%B9%E5%88%86%E5%B8%83)作为非对称分布的例子进行分析，发现当卡方分布的自由度越小，异常值出现于一侧的概率越大，[中位数](http://wiki.mbalib.com/wiki/%E4%B8%AD%E4%BD%8D%E6%95%B0)也越偏离上下四分位数的中心位置，分布偏态性越强。异常值集中在较小值一侧，则分布呈现左偏态；异常值集中在较大值一侧，则分布呈现右偏态。[60]

3.利用箱线图比较几批数据的形状

同一数轴上，几批数据的箱线图并行排列，几批数据的中位数、尾长、异常值、分布区间等形状信息便一目了然。在一批数据中，哪几个数据点出类拔萃，哪些数据点表现不及一般，这些数据点放在同类其它[群体](http://wiki.mbalib.com/wiki/%E7%BE%A4%E4%BD%93)中处于什么位置，可以通过比较各箱线图的异常值看出。各批数据的四分位距大小，正常值的分布是集中还是分散，观察各方盒和线段的长短便可明了。每批数据分布的偏态如何，分析中位线和异常值的位置也可估计出来。还有一些箱线图的变种，使数据批间的比较更加直观明白。例如有一种可变宽度的箱线图，使箱的宽度正比于批量的平方根，从而使批量大的数据批有面积大的箱，面积大的箱有适当的视觉效果。如果对同类群体的几批数据的箱线图进行比较，分析评价，便是[常模](http://wiki.mbalib.com/wiki/%E5%B8%B8%E6%A8%A1)参照解释方法的可视图示；如果把受测者数据批的箱线图与外在效标数据批的箱线图[比较分析](http://wiki.mbalib.com/wiki/%E6%AF%94%E8%BE%83%E5%88%86%E6%9E%90)，便是效标参照解释的可视图示。[61]

### t检验法

设相互独立的随机变量X,Y满足(n),那么所服从的分布为自由度为n 的t分布，记，概率密度表达式如下：[62]

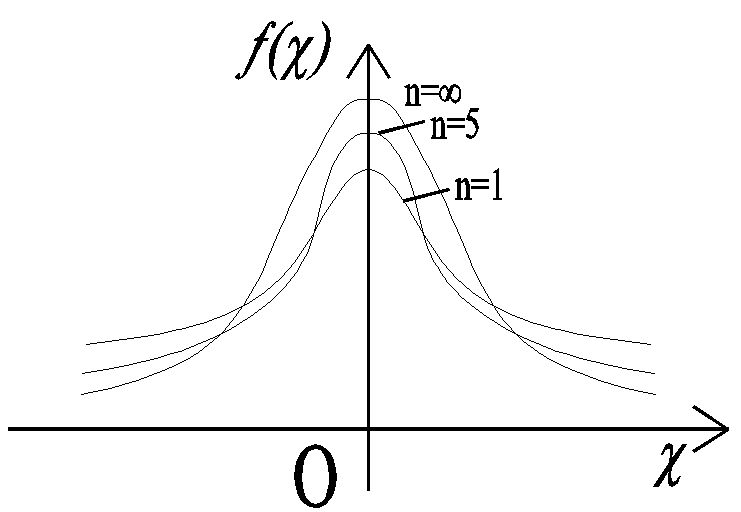


图 2‑6 不同n值下t分布概率密度图像

Figure 2‑6 Probability density diagram of t distribution under different n values

由可以看出，随n的取值而发生变化。。

给定显著水平，如果满足

P=

则称为t分布的上侧分位点，如。



图 2‑7 t分布的上侧α分位点图

Figure 2‑7 Upper quantile of t distribution diagram

如果X1, X2, X3,… Xn是由服从正态分布N的总体中随机抽取的样本，为样本的均值，为样本方差，

现提出原假设和备择假设H1:。

在原假设下，根据实际情况对显著水平取值，对于统计量，有分位数,使：

P= α

当T的观测值t=满足不等式：

= (2-3)

时，就拒绝原假设H0，接收备择假设H1。

在显著水平α下，统计量T的接收域为：

这种利用服从t分布的统计量的检验法叫t检验法。t检验法主要用于样本含量较小（例如n<30），总体标准差σ未知的[正态分布](http://wiki.mbalib.com/wiki/%E6%AD%A3%E6%80%81%E5%88%86%E5%B8%83)资料。[63]它是用T分布理论来推断差异发生的概率，从而判定两个平均数的差异是否显著。[64]

上述检验过程适合单个正态总体参数显著性检验，对于两个正态总体配对t检验过程如下。

假设两个正态总体X1, X2, X3,… Xn和Y1, Y2, Y3,…Yn分别服从N()和N()，且相互独立。X、Y的样本均值及样本方差分别记为

现提出原假设和备择假设。

取统计量，其中.

在

给定显著水平，查t分布表确定

P= α

由样本观测值求得

当

当

在显著水平α下，统计量T的接收域为：

上述检验过程的都是t检验的双侧检验，而一些实际问题往往只需要单侧检验即可。对于单一样本检验和两个正态总体配对t检验，单侧双侧检验总结如和表2-4。

表 2‑3 单个正态总体参数显著性单双侧t检验

Table 2‑3 single and bilateral t test of single normal population parameters’ conspicuousness

|  |  |  |  |
| --- | --- | --- | --- |
| 假设 | 双侧检验 | 左侧检验 | 右侧检验 |
| 假设形式 | H1: | H1: | H1: |
| 统计量 | σ未知： | | |
| 拒绝域 |  |  |  |
| P值决策 |  | | |

表 2‑4 两个正态总体参数显著性配对单双侧t检验

Table 2‑4 single and bilateral t test of double normal population parameters’ conspicuousness

|  |  |  |  |
| --- | --- | --- | --- |
| 假设 | 双侧检验 | 左侧检验 | 右侧检验 |
| 假设形式 | H1: | H1: | H1: |
| 统计量 |  | | |
| 拒绝域 |  |  |  |
| P值决策 |  | | |

### χ2检验法

如果X1, X2, X3,… Xn是由服从标准正态分布N的总体中随机抽取的样本，且相互独立，记[65]

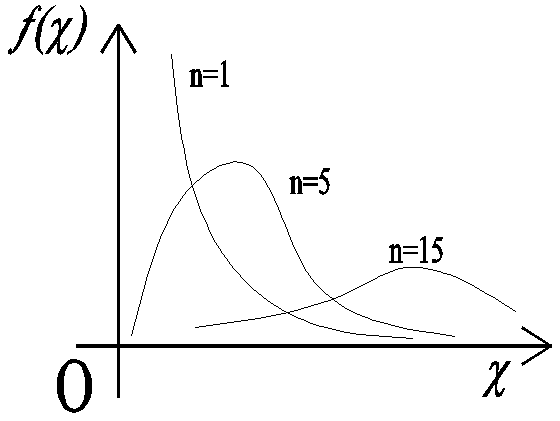


图 2‑8 不同n值下χ2分布的概率密度函数图像

Figure 2‑8 Probability density diagram of χ2 distribution under different n values

由可以看出随n的取值的不同而发生变化。

对于[66]

⑴若

⑵若

⑶若

称为所示。

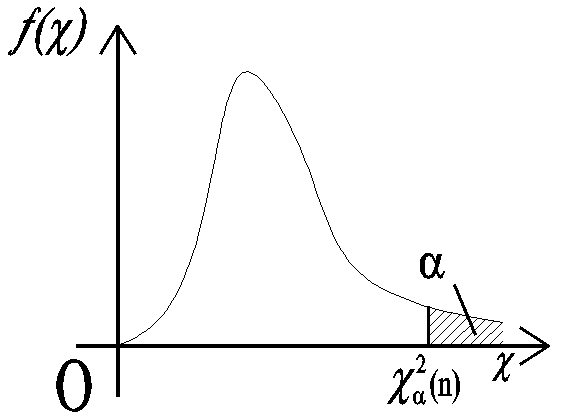


图 2‑9 χ2分布的上侧α分位点图

Figure 2‑9 Upper quantile of χ2 distribution diagram

χ2检验是检验服从正态分布总体N的某个样本的波动性是否显著的一种假设检验方法。[67]设X1, X2, X3,… Xn,是由N()的总体中随机抽取的样本，‾X 和S2的观测值分别为‾x，s2。

现提出原假设H0:和备择假设H1:。

当

对于给定的显著水平,可查χ2分布表确定分位数和，则有

故有

即：

那么置信区间为：

进

对应具体的置信区间为：

若满足

(2-4)

则接受H0，否则拒绝H0。

### F检验法

设， 则称随机变量所服从的分布为自由度为 [68]

F

由可以看出随的取值的不同而发生变化。

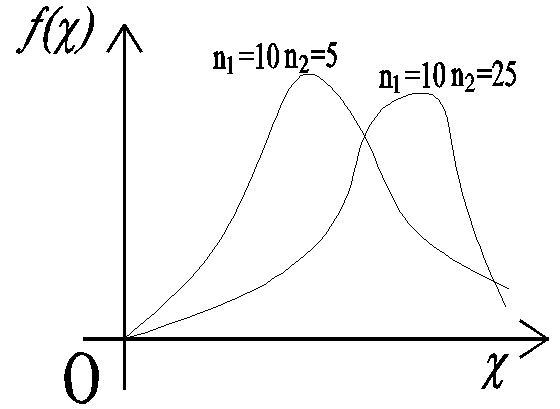


图 2‑10 不同n1、n2值下F分布的概率密度函数图像

Figure 2‑10 Probability density diagram of F distribution under different n1,n2 values

对于[69]

⑴

⑵

在给定显著水平，满足条件

的数

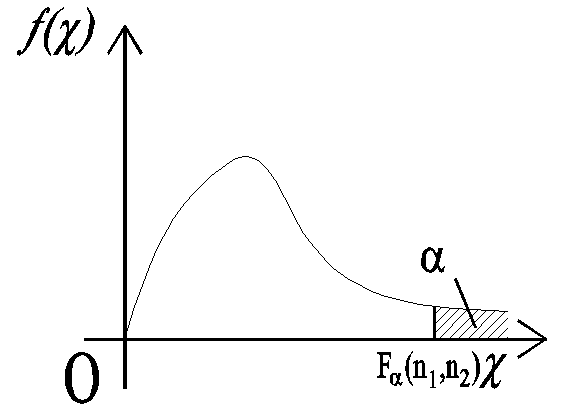


图 2‑11 F分布的上侧α分位点图

Figure 2‑11 Upper quantile of F distribution diagram

F检验是检验两个正态随机变量的总体方差是否相等的一种假设检验方法。[70]设X1, X2, X3,… Xn是由N(的总体中随机抽取的样本，Y1, Y2, Y3,… Yn是由N()的总体中随机抽取的样本,且样本相互独立，样本方差。

现提出原假设H0:和备择假设H0:。

由

当，

对于给定的显著水平,可查F分布表确定分位数和，则有

即：

因此

根。如满足不等式

(2-5)

则接受H0，否则拒绝H0。

## 系统开发平台

### 系统开发语言

web前端开发语言：html5+css+ javascript

服务器开发语言：Java

### 系统开发工具

本系统主要使用Eclipse集成开发环境。Eclipse是一个非常优秀的集成的开放源代码的开发平台，可以用于管理多种开发任务，包括测试、性能调整及程序调试等，而且还可以集成来自多个供货商的第三方应用程序开发工具。Eclipse的出现，为Java开发者提供了免费使用强大的JavaIDE的机会，通过集成大量的插件，Eclipse的功能可以不断扩展，以支持各种不同的应用。[71]

Eclipse提供了公共用户界面模型使用工具。[72]Eclipse软件旨在简化用于多操作系统软件工具的开发过程，它被设计成可以在多个操作系统上运行。基于Eclipse的软件在Linux和微软Windows系统上都可以运行，因此可以省去开发者有时要把Windows应用程序切换到Linux的操作，进而简化了整个开发过程。

除此之外，Eclipse平台是一个成熟的、精心设计的及可扩展的体系结构。Eclipse的价值还在于它为创建可扩展的集成开发环境提供了一个开放源码平台。这个平台允许任何人构建与环境和其他工具无缝集成的工具，而工具与Eclipse无缝集成的关键便是插件。[73]

虽然Eclipse软件基于Java，但根据Eclipse的体系结构，通过开发插件，它能扩展到任何语言的开发，甚至能成为图片绘制工具。因此，Eclipse也支持C、C++等编程语言。Eclipse提供的核心框架和可扩展的插件机制，给广大的程序设计者提供了无限的想象力和创造空间。可见，Eclipse是一个非常不错的开发平台。

本系统数据库主要使用Oracle12。Oracle12是一个功能极其强大的数据库系统。这种类型数据库的关键是理解数据间的关系，然后构造反映这些关系的信息库。Oracle12成功的将关系型数据库转移到桌面计算机上，提供了一个完整的客户/服务器体系结构的商用DBMs。同时它利用SQL\*NET软件层，与多种操作系统支持通信协议相配合，为oracle关系型数据库提供分布式环境，可以实现单点更新，多点查询。Oracle数据库已经被用于各种大型信息系统中，特别是诸如银 行，保险，烟草，石油等大数据量，对安全性要求较高的企业。[74]

Oracle12特点主要体现在：①支持大数据库、多用户的高性能事务处理Oracle支持最大数据库（几百TB），可充分利用硬件设备。支持大量用户同时在同一数据上执行各种应用，并使数据争用最小，保证数据的一致性。[75]②硬件环境独立。Oracle具有良好的硬件环境独立性，支持各种类型的大型，中型，小型和微机系统。③遵守数据存取语言、操作系统、用户接口和网络通信协议的工业标准。④较好的安全性和完整控制。Oracle有用户鉴别、特权、角色、触发器、日 志、后备等功能，有效地保证了数据存取的安全性和完整性以及并发控制和数据的回复。⑤具有可移植性、可兼容性与可连接性。Oracle不仅可以在不同型号的机器上运行，而且可以在同一厂家的不同操作系统支持下运行，具有操作系统的独立性。

### 网络结构模式

在当前互联网领域，浏览器/服务器结构是当前非常流行的客户机/服务器结构，主要是利用不断成熟的WWW浏览器技术，结合浏览器的多种脚本语言，用通用浏览器就实现了原来需要复杂的专用软件才能实现的强大功能，并节约了开发成本，是一种全新的软件体系结构。[76]

B/S结构是一种典型的3层结构模式：表示层、处理层和数据层。其中表示层为浏览器。浏览器仅承担网页信息的浏览功能，以超文本的格式实现信息的浏览和输入，没有任何业务处理能力；功能层由服务器承担业务处理逻辑和页面的储存管理，接收客户浏览的任务请求，并根据任务请求类型执行相应的事务处理程序；数据层有数据库服务器承担数据逻辑，其任务是接收服务器对数据库提出的操作请求，由数据库完成数据的查询、修改、统计、更新等工作，并对数据处理结果交给服务器。

B/S网络结构模式有以下优势：①操作使用简单。②维护和升级方式简单。③成本低，选择多。[77]当然，B/S网络结构模式也有不足之处：①部分需求功能不能实现或者实现较困难。②复杂的应用构造困难。③安全性不是很好。④当网页访问量过大时，可能导致服务器荷载过重。本研究主要设计后台管理系统的设计，不会有过多的用户访问，也没有特别复杂的功能需要实现，再考虑到方便维护，降低经济成本等因素，最终决定采用B/S网络结构模式实现系统的搭建。

# 基于数理统计的污水监测数据真伪鉴别研究

## 引言

人工上报数据和在线监测数据是主管部门实现对污水处理厂监督管理和业绩评价的重要考核指标。在城市污水处理厂正常运行过程中，一些排放污水的企业在经济利益驱使下，采用各种手段对在线监测数据进行造假，例如在监测点稀释污水，私接暗管，篡改监测仪器重要参数等。这些不法行为导致污水监测数据失真，给监测部门的工作带来很大的麻烦。许多监测人员本身业务素质较低，缺少将数据进行科学有效的分析和处理的能力，导致数据的可信度不高，失去评价意义，因此数据的有效性和精确性值得关注。如果能采取一定的方法，将异常数据检测出，将十分有利于管理部门及时发现污水处理厂的异常情况，并第一时间与污水厂取得联系，找出问题原因。城市污水处理厂异常数据检测的实现，将为管理部门对污水处理厂的监管提供有力技术措施。

本文依据统计学相关理论，针对城市污水厂出水水质数据，实现了对城市污水处理厂的监测数据真伪鉴别，可以对来自在线监测设备和人工上报的数据进行有效性甄别，奇异值剔除，判断出水水质数据是否达标等，并可判别两者是否存在显著性差异，从而验证人工上报数据的可靠度。

## 数据真伪鉴别流程

本鉴别流程综合运用数理统计的方法，首先采用格拉布斯法进行异常值判断并修正时监测数据，再根据《城市污水处理厂污染物排放标准》（GB18918-2002）要求的级别标准对修正后的时数据进行达标筛选，不满足排放标准的数据剔除，并采用t检验法来验证抽检数据样本均值是否在整体样本均值的接收域，采用χ2检验法来判断样本相对于整体的波动性是否有显著变化，以此实现污水监测数据的纵向对比；采用F检验法和箱线图法来匹配年数据离散程度相近的样本，以此实现污水监测数据的横向对比。

综合运用数理统计方法的鉴别流程如：

(1)获取污水处理厂污水排放指标时在线监测数据，建立数据表；

(2)对时在线监测数据进行升序排序；

(3)运用格拉布斯法对排序后时在线监测数据进行异常值的判断和观测值的修正，判断出的异常值即为异常数据，找出异常值对应的时刻，确定异常值出现的位置，并进行剔除，保留修正数据；

(4)在修正的时监测数据基础上计算日均值；

(5)根据《城市污水处理厂污染物排放标准》（GB18918-2002）要求的级别标准对日均值进行达标筛选，不满足排放标准的数据剔除；

(6)取监测月份和前11个月的年度观测值作为样本基数，通过月均值检验和月均值波动性检验，实现数据的纵向比较；

(7)匹配年数据离散程度相近的样本，来匹配出具有对比价值的企业，并通过差值波动性分析，实现数据的横向比较；

(8)由最终检测结果鉴别数据真伪。

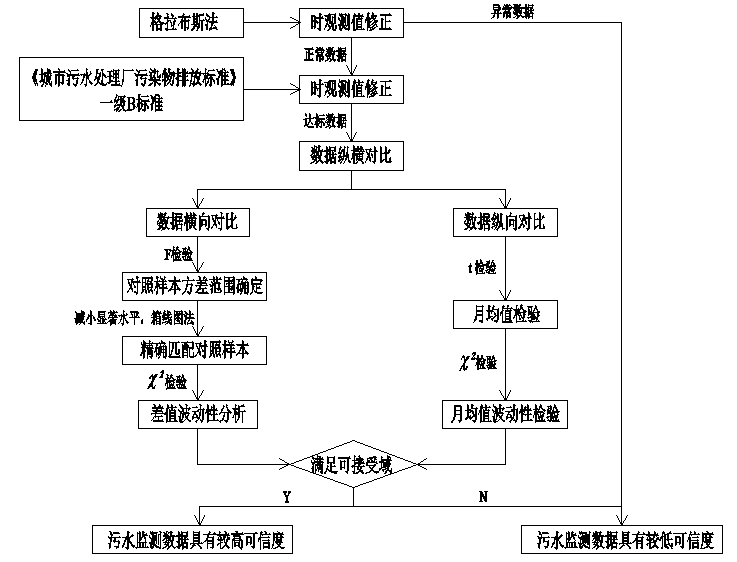


图 3‑1 污水监测数据真伪鉴别流程图

Figure 3‑1 Authenticity identification flow chart of sewage monitoring data

## 基于数理统计的污水监测数据真伪鉴别

目前监测部门对不同类型企业的污水监测指标略有不同。本文为清晰说明本鉴别流程，取其中一项指标COD的观测值作为研究对象，其它指标方法类同不再一一阐述。

### 格拉布斯法修正时观测值

本文利用统计学中的格布拉斯法实现了污水监测数据的异常值检测，其提供了一种精确度较高地识别数据中异常值的方法，识别结果比较客观。据了解，目前对于国控企业要求每天对某一监测点每两小时监测一次，一天可反馈12个时观测值。对于一般性企业没有硬性规定，通常四小时监测一次。为更好的证明本鉴别流程的可行性，实地走访某污水处理厂（国控企业）后取其2016年12月1日的污水部分监测指标数据（）进行实例说明。

表 3‑1 污水处理厂2016年12月1日的污水部分监测指标数据

Table 3‑1 A part of the monitoring indicators data of sewage treatment plant in December 1, 2016

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 时刻 | COD | pH | SS | 氨氮 | TN | TP |
| 00：00 | 34.52 | 6.8 | 10.00 | 2.37 | 15.54 | 0.70 |
| 02：00 | 33.63 | 7.2 | 11.00 | 2.32 | 16.74 | 0.80 |
| 04：00 | 38.41 | 6.8 | 9.00 | 2.56 | 16.82 | 0.60 |
| 06：00 | 36.80 | 7.1 | 12.00 | 2.29 | 16.79 | 0.70 |
| 08：00 | 39.22 | 7.3 | 13.00 | 2.15 | 17.36 | 0.80 |
| 10：00 | 39.87 | 6.9 | 14.00 | 2.21 | 17.84 | 0.80 |
| 12：00 | 36.61 | 7.2 | 14.00 | 1.98 | 18.58 | 0.70 |
| 14：00 | 35.72 | 7.1 | 15.00 | 2.05 | 19.34 | 0.90 |
| 16：00 | 37.94 | 7.2 | 13.00 | 2.25 | 18.56 | 0.80 |
| 18：00 | 34.75 | 7.3 | 12.00 | 2.38 | 17.94 | 0.70 |
| 20：00 | 38.52 | 7.0 | 10.00 | 1.97 | 16.44 | 0.80 |
| 22：00 | 37.68 | 6.8 | 9.00 | 2.29 | 15.89 | 0.60 |

因为时监测数据是本鉴别流程的基础数据，为保证其可信度，所以进行观测值修正是首要任务。用格布拉斯法对数据的异常值剔除是建立在正态分布的数据基础之上，先对监测数据进行正态性检验。

正态性检验的方法主要有Shapiro-Wilk 检验法、Anderson-Darling 检验法、Kolmogorov-Smirnov 检验法等，在此简单介绍Shapiro-Wilk 检验法，并对数据进行正态性检验。

Shapiro-Wilk 检验法也称为W检验，适用于样本的数量n满足(8<n<50)的统计量。W检验法是建立在次序统计量的基础上进行的，将n个独立观测值按照非将次序排列，记为X1, X2, X3,… Xn，检测统计量为：

其中

另外…,据此可将W检验转化为：

当总体趋近于正态分布时，W值接近于1，因此在在显著水平α下，如果W的值小于其分位数，即时，则拒绝原假设。如果计算出，则在显著水平α上未落入拒绝域，可认为这次数据近似服从正态分布，如果计算出，则在显著水平α上落入拒绝域，则认为这次数据不服从正态分布。

如上所述，对数据进行升序排列后如所示。

表 3‑2 升序排列后污水处理厂2016年12月1日的污水部分监测指标数据

Table 3‑2 A part of the monitoring indicators data of sewage treatment plant in December 1, 2016 after ascending order

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| COD | pH | SS | 氨氮 | TN | TP |
| 33.63 | 6.8 | 9.00 | 1.97 | 15.54 | 0.60 |
| 34.52 | 6.8 | 9.00 | 1.98 | 15.89 | 0.60 |
| 34.75 | 6.8 | 10.00 | 2.05 | 16.44 | 0.70 |
| 35.72 | 6.9 | 10.00 | 2.15 | 16.74 | 0.70 |
| 36.61 | 7.0 | 11.00 | 2.21 | 16.79 | 0.70 |
| 36.80 | 7.1 | 12.00 | 2.25 | 16.82 | 0.70 |
| 37.68 | 7.1 | 12.00 | 2.29 | 17.36 | 0.80 |
| 37.94 | 7.2 | 13.00 | 2.29 | 17.84 | 0.80 |
| 38.41 | 7.2 | 13.00 | 2.32 | 17.94 | 0.80 |
| 38.52 | 7.2 | 14.00 | 2.37 | 18.56 | 0.80 |
| 39.22 | 7.3 | 14.00 | 2.38 | 18.58 | 0.80 |
| 39.87 | 7.3 | 15.00 | 2.56 | 19.34 | 0.90 |

考虑到计算的复杂性，这里仅对COD、氨氮和TN指标进行正态性检验，其它指标计算方法类同。

首先计算出：

COD:

氨氮：

TN：

将W的计算过程列入，将安排在同一行，的值查W检验表得来。

表 3‑3 污水部分监测指标数据正态性检验流程

Table 3‑3 The process of normality inspection of sewage monitoring data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| COD | | | | |
| k | xk | x(n+1-k) | dk | ak |
| 1 | 33.63 | 39.87 | 6.24 | 0.5475 |
| 2 | 34.52 | 39.22 | 4.70 | 0.3325 |
| 3 | 34.75 | 38.52 | 3.77 | 0.2347 |
| 4 | 35.72 | 38.41 | 2.69 | 0.1586 |
| 5 | 36.61 | 37.94 | 1.33 | 0.0922 |
| 6 | 36.80 | 37.68 | 0.88 | 0.0303 |
| 氨氮 | | | | |
| k | xk | x(n+1-k) | dk | ak |
| 1 | 1.97 | 2.56 | 0.59 | 0.5475 |
| 2 | 1.98 | 2.38 | 0.40 | 0.3325 |
| 3 | 2.05 | 2.37 | 0.32 | 0.2347 |
| 4 | 2.15 | 2.32 | 0.17 | 0.1586 |
| 5 | 2.21 | 2.29 | 0.08 | 0.0922 |
| 6 | 2.25 | 2.29 | 0.04 | 0.0303 |
| TN | | | | |
| k | xk | x(n+1-k) | dk | ak |
| 1 | 15.54 | 19.34 | 3.80 | 0.5475 |
| 2 | 15.89 | 18.58 | 2.69 | 0.3325 |
| 3 | 16.44 | 18.56 | 2.12 | 0.2347 |
| 4 | 16.74 | 17.94 | 1.20 | 0.1586 |
| 5 | 16.79 | 17.84 | 1.05 | 0.0922 |
| 6 | 16.82 | 17.36 | 0.54 | 0.0303 |

在实际工程中，因为计算量巨大，可以运用MATLAB编程软件进行计算，根据公式3-1求得的W和Wα如，Wα由显著水平α和样本容量n通过查表获取，这里α取0.05。

表 3‑4 污水部分监测指标数据W及Wα

Table 3‑4 Wand Wα ofa part of the monitoring indicators data

|  |  |  |  |
| --- | --- | --- | --- |
| 监测指标 | COD | 氨氮 | TN |
| W | 0.960 | 0.956 | 0.967 |
| Wα | 0.859 | 0.859 | 0.859 |

比较发现，COD、氨氮和TN的时在线监测数据均满足W > Wα，大体呈正态分布。在此基础上，可通过格拉布斯法进行异常值的判断和观测值的修正。假定企业发现COD 浓度有超出行业规定标准趋势，在14:00通过稀释监测点污水的方法降低污染物浓度，使COD浓度变为25.72 mg/L，氨氮浓度为1.45mg/L，TN浓度为14.34 mg/L，现采用格拉布斯法进行异常值判断并修正数据。

可以肯定异常值xe∈{ xmax, xmin}。由确定各监测指标的最大值xmax，最小值xmin，并分别计算出平均值、标准差、最大位置偏离值和最小位置偏离值，结果如。

表 3‑5 格布拉斯法部分计算值

Table 3‑5 Partial calculated value used Grubbs 'Criterion

|  |  |  |  |
| --- | --- | --- | --- |
| 监测指标 | COD | 氨氮 | TN |
| xmax | 39.87 | 2.56 | 18.58 |
| xmin | 25.72 | 1.45 | 14.34 |
| 平均值 | 36.14 | 2.18 | 16.83 |
| 标准差 | 3.82 | 0.28 | 1.30 |
| 最大位置偏离值 | 3.73 | 0.38 | 1.73 |
| 最小位置偏离值 | 10.42 | 0.73 | 2.49 |
| 假定可疑值 | 25.72 | 1.45 | 14.34 |

对于作假后COD、氨氮和TN的监测数据，|xmax-‾x| <‾|x - xmin|，因此假定可疑值均为xmin，即最小值的异常概率最大。

由公式2-1可知，如果异常数据和平均值的残差与标准差的比值（其

中i为可疑值排列序号）大于格拉布斯表的临界值，可判断异常值及其位置。临界值可根据样本观测值个数n和显著水平α查得到。这里取α=0.05，查得=2.285。各监测数据的Gi值见。

表 3‑6 异常值鉴别

Table 3‑6 Outlier identification

|  |  |  |  |
| --- | --- | --- | --- |
| 监测指标 | COD | 氨氮 | TN |
| Gi | 2.728 | 2.607 | 1.915 |
|  | 2.285 | 2.285 | 2.283 |
|  | 是 | 是 | 否 |
| 是否为异常值 | 是 | 是 | 否 |

由可以看出COD和氨氮指标数据均满足确定COD和氨氮在14:00的观测值为异常值。但是对于TN，

数据的异常值未能准确判断出来，这说明判断的精度不够，需要增大显著水平α的值来提高检测精度。因为是同一时刻测出的各个指标的值，有理由认为各个指标的浓度均被稀释，剔除异常值，重复上述流程修正观测值，直到满足检出水平为止。

### 筛选达标日观测值并修正

根据该污水处理厂排入地表水域环境功能和保护目标，以及污水处理厂的处理工艺，出水水质需符合《城市污水处理厂污染物排放标准》(GB18918-2002)要求的一级B 标准，即要求。对修正后的时数据进行达标筛选，不满足排放标准的数据剔除，那么每天的时监测值均为达标数据。这里取其样本平均值作为日观测值。同样采用格拉布斯法对该企业某月日观测值‾x1 ,‾x2 ,‾x3…‾xn进行修正，即可获得具有较高真实度的日监测数据。通过走访实地企业获取到改污水处理厂2016年12月份的监测数据，并通过格拉布斯法修正后的日观测值如。

表 3‑7 污水处理厂2016年12月份污水部分监测指标数据

Table 3‑7 A part of the monitoring indicators data of sewage treatment plant in December, 2016

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 日期 | COD | pH | SS | 氨氮 | TN | TP |
| 2016.12.1 | 36.97 | 7.1 | 11.83 | 2.24 | 17.32 | 0.74 |
| 2016.12.2 | 34.35 | 8.1 | 9.43 | 2.97 | 15.94 | 0.56 |
| 2016.12.3 | 38.64 | 7.6 | 7.86 | 3.56 | 14.66 | 0.41 |
| 2016.12.4 | 40.89 | 6.4 | 8.67 | 3.98 | 13.79 | 0.63 |
| 2016.12.5 | 42.45 | 6.0 | 9.13 | 3.21 | 15.39 | 0.87 |
| 2016.12.6 | 40.53 | 7.2 | 10.78 | 4.12 | 17.55 | 0.75 |
| 2016.12.7 | 38.72 | 8.6 | 12.54 | 4.35 | 18.93 | 0.44 |
| 2016.12.8 | 34.51 | 8.8 | 14.36 | 3.69 | 19.53 | 0.38 |
| 2016.12.9 | 32.18 | 7.5 | 15.69 | 3.45 | 17.46 | 0.57 |
| 2016.12.10 | 30.56 | 7.1 | 16.73 | 3.00 | 19.01 | 0.72 |
| 2016.12.11 | 29.88 | 8.1 | 13.45 | 2.96 | 17.66 | 0.85 |
| 2016.12.12 | 25.47 | 7.2 | 11.44 | 2.53 | 15.93 | 0.89 |
| 2016.12.13 | 28.21 | 6.5 | 10.56 | 2.04 | 17.52 | 0.92 |
| 2016.12.14 | 33.59 | 6.3 | 8.42 | 1.97 | 13.96 | 0.68 |
| 2016.12.15 | 37.58 | 7.1 | 9.74 | 2.44 | 14.20 | 0.96 |
| 2016.12.16 | 39.14 | 6.9 | 7.15 | 1.69 | 18.36 | 0.74 |
| 2016.12.17 | 43.58 | 6.3 | 6.88 | 2.87 | 17.36 | 0.68 |
| 2016.12.18 | 44.80 | 7.2 | 8.69 | 3.22 | 19.40 | 0.93 |
| 2016.12.19 | 41.52 | 6.8 | 12.66 | 3.45 | 13.88 | 0.82 |
| 2016.12.20 | 39.40 | 6.3 | 14.36 | 3.68 | 15.57 | 0.62 |
| 2016.12.21 | 35.12 | 6.2 | 15.89 | 3.79 | 16.08 | 0.57 |
| 2016.12.22 | 33.69 | 7.5 | 16.11 | 3.02 | 18.96 | 0.55 |
| 2016.12.23 | 30.47 | 8.6 | 14.36 | 2.64 | 19.52 | 0.69 |
| 2016.12.24 | 29.16 | 9.0 | 12.69 | 2.44 | 18.47 | 0.84 |
| 2016.12.25 | 28.45 | 7.4 | 10.89 | 2.18 | 19.22 | 0.53 |
| 2016.12.26 | 26.78 | 8.1 | 9.56 | 1.73 | 15.66 | 0.67 |
| 2016.12.27 | 30.45 | 7.9 | 8.63 | 1.68 | 17.20 | 0.76 |
| 2016.12.28 | 34.22 | 6.4 | 7.46 | 2.34 | 19.38 | 0.83 |
| 2016.12.29 | 38.56 | 6.3 | 8.66 | 2.16 | 18.64 | 0.92 |
| 2016.12.30 | 40.52 | 6.2 | 9.32 | 2.77 | 16.29 | 0.93 |
| 2016.12.31 | 40.69 | 6.7 | 10.99 | 3.52 | 18.97 | 0.78 |

计算出各指标的均值和2阶中心矩见。

表 3‑8 2016年12月份各检测指标均值及2阶中心矩

Table 3‑8 The second order centre-to-centre spacing and mean values of each detection indexs in December, 2016

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 监测指标 | COD | pH | SS | 氨氮 | TN | TP |
| ‾x | 35.51 | 7.2 | 11.13 | 2.89 | 17.16 | 0.72 |
| 2阶中心矩 | 27.61 | 0.71 | 8.01 | 0.54 | 3.38 | 0.03 |

仍然用W检验法对该污水处理厂12月份数据进行正态性检验，检验过程如。

表 3‑9 2016年12月份日观测值正态性检验流程

Table 3‑9 The process of normality inspection of daily observations in December, 2016

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| COD | | | | | | | | | |
| k | xk | x(n+1-k) | dk | ak | k | xk | x(n+1-k) | dk | ak |
| 1 | 25.47 | 44.80 | 19.33 | 0.4251 | 9 | 30.56 | 39.40 | 8.84 | 0.1036 |
| 2 | 26.78 | 43.58 | 16.80 | 0.2944 | 10 | 32.18 | 39.14 | 6.96 | 0.0862 |
| 3 | 28.21 | 42.45 | 14.24 | 0.2487 | 11 | 33.59 | 38.72 | 5.13 | 0.0667 |
| 4 | 28.45 | 41.52 | 13.07 | 0.2148 | 12 | 33.69 | 38.64 | 4.95 | 0.0537 |
| 5 | 29.16 | 40.89 | 11.73 | 0.1870 | 13 | 34.22 | 38.56 | 4.34 | 0.0381 |
| 6 | 29.88 | 40.69 | 10.81 | 0.1630 | 14 | 34.35 | 37.58 | 3.23 | 0.0227 |
| 7 | 30.45 | 40.53 | 10.08 | 0.1415 | 15 | 34.51 | 36.97 | 2.46 | 0.0076 |
| 8 | 30.47 | 40.52 | 10.05 | 0.1219 | W | 0.949 | | | |
| pH | | | | | | | | | |
| k | xk | x(n+1-k) | dk | ak | k | xk | x(n+1-k) | dk | ak |
| 1 | 6.0 | 9.0 | 3.0 | 0.4251 | 9 | 6.4 | 7.6 | 1.2 | 0.1036 |
| 2 | 6.2 | 8.8 | 2.6 | 0.2944 | 10 | 6.5 | 7.5 | 1.0 | 0.0862 |
| 3 | 6.2 | 8.6 | 2.4 | 0.2487 | 11 | 6.7 | 7.5 | 0.8 | 0.0667 |
| 4 | 6.3 | 8.6 | 2.3 | 0.2148 | 12 | 6.8 | 7.4 | 0.6 | 0.0537 |
| 5 | 6.3 | 8.1 | 1.8 | 0.1870 | 13 | 6.9 | 7.2 | 0.3 | 0.0381 |
| 6 | 6.3 | 8.1 | 1.8 | 0.1630 | 14 | 7.1 | 7.2 | 0.1 | 0.0227 |
| 7 | 6.3 | 8.1 | 1.8 | 0.1415 | 15 | 7.1 | 7.2 | 0.1 | 0.0076 |
| 8 | 6.4 | 7.9 | 1.5 | 0.1219 | W | 0.924 | | | |
| SS | | | | | | | | | |
| k | xk | x(n+1-k) | dk | ak | k | xk | x(n+1-k) | dk | ak |
| 1 | 6.88 | 16.73 | 9.85 | 0.4251 | 9 | 8.69 | 12.69 | 4.00 | 0.1036 |
| 2 | 7.15 | 16.11 | 8.96 | 0.2944 | 10 | 9.13 | 12.66 | 3.53 | 0.0862 |
| 3 | 7.46 | 15.89 | 8.43 | 0.2487 | 11 | 9.32 | 12.54 | 3.22 | 0.0667 |
| 4 | 7.86 | 15.69 | 7.83 | 0.2148 | 12 | 9.43 | 11.83 | 2.40 | 0.0537 |
| 5 | 8.42 | 14.36 | 5.94 | 0.1870 | 13 | 9.56 | 11.44 | 1.88 | 0.0381 |
| 6 | 8.63 | 14.36 | 5.73 | 0.1630 | 14 | 9.74 | 10.99 | 1.25 | 0.0227 |
| 7 | 8.66 | 14.36 | 5.70 | 0.1415 | 15 | 10.56 | 10.89 | 0.33 | 0.0076 |
| 8 | 8.67 | 13.45 | 4.78 | 0.1219 | W | 0.930 | | | |
| 氨氮 | | | | | | | | | |
| k | xk | x(n+1-k) | dk | ak | k | xk | x(n+1-k) | dk | ak |
| 1 | 1.68 | 4.35 | 2.67 | 0.4251 | 9 | 2.34 | 3.45 | 1.11 | 0.1036 |
| 2 | 1.69 | 4.12 | 2.43 | 0.2944 | 10 | 2.44 | 3.45 | 1.01 | 0.0862 |
| 3 | 1.73 | 3.98 | 2.25 | 0.2487 | 11 | 2.44 | 3.22 | 0.78 | 0.0667 |
| 4 | 1.97 | 3.79 | 1.82 | 0.2148 | 12 | 2.53 | 3.21 | 0.68 | 0.0537 |
| 5 | 2.04 | 3.69 | 1.65 | 0.1870 | 13 | 2.64 | 3.02 | 0.38 | 0.0381 |
| 6 | 2.16 | 3.68 | 1.52 | 0.1630 | 14 | 2.77 | 3.00 | 0.23 | 0.0227 |
| 7 | 2.18 | 3.56 | 1.38 | 0.1415 | 15 | 2.87 | 2.97 | 0.10 | 0.0076 |
| 8 | 2.24 | 3.52 | 1.28 | 0.1219 | W | 0.960 | | | |
| TN | | | | | | | | | |
| k | xk | x(n+1-k) | dk | ak | k | xk | x(n+1-k) | dk | ak |
| 1 | 13.79 | 19.53 | 5.74 | 0.4251 | 9 | 15.93 | 18.93 | 3.00 | 0.1036 |
| 2 | 13.88 | 19.52 | 5.64 | 0.2944 | 10 | 15.94 | 18.64 | 2.70 | 0.0862 |
| 3 | 13.96 | 19.40 | 5.44 | 0.2487 | 11 | 16.08 | 18.47 | 2.39 | 0.0667 |
| 4 | 14.20 | 19.38 | 5.18 | 0.2148 | 12 | 16.29 | 18.36 | 2.07 | 0.0537 |
| 5 | 14.66 | 19.22 | 4.56 | 0.1870 | 13 | 17.20 | 17.66 | 0.46 | 0.0381 |
| 6 | 15.39 | 19.01 | 3.62 | 0.1630 | 14 | 17.32 | 17.55 | 0.23 | 0.0227 |
| 7 | 15.57 | 18.97 | 3.40 | 0.1415 | 15 | 17.36 | 17.52 | 0.16 | 0.0076 |
| 8 | 15.66 | 18.96 | 3.30 | 0.1219 | W | 0.904 | | | |
| TP | | | | | | | | | |
| k | xk | x(n+1-k) | dk | ak | k | xk | x(n+1-k) | dk | ak |
| 1 | 0.38 | 0.96 | 0.58 | 0.4251 | 9 | 0.62 | 0.84 | 0.22 | 0.1036 |
| 2 | 0.41 | 0.93 | 0.52 | 0.2944 | 10 | 0.63 | 0.83 | 0.20 | 0.0862 |
| 3 | 0.44 | 0.93 | 0.49 | 0.2487 | 11 | 0.67 | 0.82 | 0.15 | 0.0667 |
| 4 | 0.53 | 0.92 | 0.39 | 0.2148 | 12 | 0.68 | 0.78 | 0.10 | 0.0537 |
| 5 | 0.55 | 0.92 | 0.37 | 0.1870 | 13 | 0.68 | 0.76 | 0.08 | 0.0381 |
| 6 | 0.56 | 0.89 | 0.33 | 0.1630 | 14 | 0.69 | 0.75 | 0.06 | 0.0227 |
| 7 | 0.57 | 0.87 | 0.30 | 0.1415 | 15 | 0.72 | 0.74 | 0.02 | 0.0076 |
| 8 | 0.57 | 0.85 | 0.28 | 0.1219 | W | 0.949 | | | |

考虑到样本容量较大，这里提高显著水平，取α=0.1，n=31，查可以得到Wα=0.902，该污水处理厂12月份各指标数据均满足W＞Wα，可以近似把该污水处理厂2016年各月份的监测数据看作呈正态分布，在此基础上后续鉴别流程不再对数据做正态性检验。

基于上述两步筛选和鉴定，企业指定月份内的监测数据已经具有很高的可信度。但由于月数据样本基数小，偶然性大，如果企业采用的非法手段（如私接暗道排污）运作时间较长，仅对月数据修正鉴别是不能断定数据真伪的。如果有充足的样本基数（如一年的样本观测值），有参照价值的对比样本，运用合理的检测原理和方法即可达到满意效果。

考虑至此，如下将基于数理统计方法对监测数据进行纵横对比。

### 数据纵向对比

基于纵向对比就是与企业历史数据作比较，可取监测月份和前11个月的年度观测值作为样本基数。由于样本数据量巨大，为了提高鉴别的可信度，既要评价抽检数据样本均值是否在整体样本均值的接收域，又要判断样本相对于整体的波动性是否有显著变化。

#### 月均值检测

基于该污水处理厂一年内污水各指标监测数据近似服从正态分布，‾Xn, ‾Xn-1, ‾Xn-2,… ‾Xn-11,是随机抽取的12个月均值样本，其中Xn代表n月份的某一监测指标月观测值，本文取走访企业2016年度日修正数据的算术平均值作为月观测值，取12月份监测指标月均值为待测样本，并统计计算前十一个月份的月均值，得到。

表 3‑10 2016年各污水监测指标月观测值

Table 3‑10 Monthly observations of each detection indexs in 2016

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 月份 | COD | pH | SS | 氨氮 | TN | TP |
| 1月 | 33.47 | 7.5 | 9.26 | 1.94 | 15.34 | 0.56 |
| 2月 | 35.89 | 8.1 | 11.68 | 3.12 | 17.83 | 0.75 |
| 3月 | 37.64 | 7.4 | 13.25 | 3.56 | 19.35 | 0.84 |
| 4月 | 32.80 | 6.9 | 8.64 | 1.86 | 14.96 | 0.59 |
| 5月 | 39.42 | 7.5 | 15.96 | 3.97 | 20.14 | 0.93 |
| 6月 | 32.84 | 7.2 | 9.02 | 1.62 | 14.28 | 0.48 |
| 7月 | 36.61 | 7.6 | 12.34 | 3.16 | 18.23 | 0.77 |
| 8月 | 33.57 | 7.9 | 9.75 | 1.47 | 15.34 | 0.62 |
| 9月 | 36.97 | 8.4 | 12.68 | 3.25 | 18.35 | 0.80 |
| 10月 | 39.88 | 8.2 | 14.78 | 3.81 | 20.31 | 0.96 |
| 11月 | 34.94 | 7.6 | 13.20 | 2.01 | 16.20 | 0.61 |
| 12月 | 35.51 | 7.2 | 11.13 | 2.89 | 17.16 | 0.72 |

由于企业年度均值离散程度σ2是未知量，本文采用t检验法来验证抽检数据样本均值是否在整体样本均值的接收域，这里将各个监测指标12月份的月均值作为原假设H0，对其观测值的否定值作为备择假设H1，如。

表 3‑11 t检验原假设及备择假设

Table 3‑11 Original hypothesis and Alternative hypothesis of t test

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 监测指标 | COD | pH | SS | 氨氮 | TN | TP |
| 原假设 | H0=35.51 | H0=7.2 | H0=11.13 | H0=2.89 | H0=17.16 | H0=0.72 |
| 备择假设 | H1≠35.51 | H1≠7.2 | H1≠11.13 | H1≠2.89 | H1≠17.16 | H1≠0.72 |

计算出前11个月各指标算术平均值、标准差和统计量，结果见。

表 3‑12 前11个月各指标的算术平均值、标准差和统计量

Table 3‑12 Arithmetic mean values, standard deviation and statistics of each detection indexs in the first 11 months

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 监测指标 | COD | pH | SS | 氨氮 | TN | TP |
| 算术平均值 | 35.82 | 7.7 | 11.87 | 2.71 | 17.30 | 0.72 |
| 标准差 | 2.531 | 0.448 | 2.442 | 0.933 | 2.169 | 0.158 |
| 统计量 | 0.406 | 3.701 | 1.005 | -0.640 | 0.214 | 0 |

对待测月份各指标观测值进行双侧t检验，由公式2-3可知，如果统计量T的观测值t满足| t | < tα/2 (10),可接收原假设H0，即认为12月的月均值在整体样本的接受域内。分位数tα/2 (10)可根据显著水平α查t分布表得到。这里取α=0.05，查得=2.228，各指标的可接受域如下：

COD：34.14~37.50 mg/L； pH：7.4~8.0

SS：10.25~13.49 mg/L； 氨氮：2.09~3.33 mg/L；

TN：12.53~22.07 mg/L； TP： 0.38~1.06 mg/L。

为了更直观地表现该污水处理厂2016年各指标的变化趋势以及预测月份可接受范围，做如下折线图表示。

C:\Users\Administrator\AppData\Roaming\Tencent\Users\625328200\QQ\WinTemp\RichOle\IWAQ@FKD5N2GK]424U`TF(T.png

图 3‑2 2016年COD变化趋势图

Figure 3‑2 The COD trend chart in 2016

C:\Users\Administrator\AppData\Roaming\Tencent\Users\625328200\QQ\WinTemp\RichOle\IWAQ@FKD5N2GK]424U`TF(T.png

图 3‑3 2016年pH变化趋势图

Figure 3‑3 The pH trend chart in 2016

C:\Users\Administrator\AppData\Roaming\Tencent\Users\625328200\QQ\WinTemp\RichOle\IWAQ@FKD5N2GK]424U`TF(T.png

图 3‑4 2016年SS变化趋势图

Figure 3‑4 The SS trend chart in 2016

C:\Users\Administrator\AppData\Roaming\Tencent\Users\625328200\QQ\WinTemp\RichOle\IWAQ@FKD5N2GK]424U`TF(T.png

图 3‑5 2016年氨氮变化趋势图

Figure 3‑5 The ammonia nitrogen trend chart in 2016

C:\Users\Administrator\AppData\Roaming\Tencent\Users\625328200\QQ\WinTemp\RichOle\IWAQ@FKD5N2GK]424U`TF(T.png

图 3‑6 2016年TN变化趋势图

Figure 3‑6 The TN trend chart in 2016

C:\Users\Administrator\AppData\Roaming\Tencent\Users\625328200\QQ\WinTemp\RichOle\IWAQ@FKD5N2GK]424U`TF(T.png

图 3‑7 2016年TP变化趋势图

Figure 3‑7 The TP trend chart in 2016

从2016年污水各监测指标变化趋势图可以发现：各月份出水水质要求均满足《城市污水处理厂污染物排放标准》(GB18918-2002)要求的一级B 标准；12月份COD、SS、氨氮、TN和TP的出水浓度都落在各自的接受域内，说明显著水平为0.05时，污染物的浓度值具有较高的可信度；虽然pH值不在可接受域内，但通过图表可以发现同一月份的各监测指标的变化趋势大体相同，因此对数据的真实性不具有否定意义。

从2016年污水各监测指标变化趋势图还可以发现：5月份的出水水质的各监测指标的值均偏高，而在6月份又明显下降，推测可能是污水处理厂改进了污水处理工艺，提高了污染物去除能力，随后走访该污水处理厂获知该污水处理厂在5月中旬确实对污水处理工艺做了改进(具体实施工艺见2.1.1章内容)。这便证实了本研究采取的鉴别方法不仅能鉴别污水监测数据的真伪，而且对工作人员预测临近时间段的监测数据和企业是否改进工程技术具有一定的指导意义。

#### 月均值波动性检测

要判断样本相对于整体的波动性是否有显著变化，实质是观察每个月数据的离散程度是否符合整体样本的离散程度。样本方差S2是数据离散程度最直观的描述，如果视年度日数据为服从正态分布的总体N，只要检验鉴别样本方差S2的观测值s2是否在总体样本的σ2的接受域内即可。

因为求得总体样本的σ2涉及极其繁重的计算量，本文在获取到走访企业的年度数据后运用数学工具软件Matlab算的各污水监测指标总体长期以来服从的方差σ2和12月份的各指标的方差观测值s2见。

表 3‑13 各指标长期服从方差和待测月份方差观测值

Table 3‑13 The long-term variance and the variance of the month to be measured

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 监测指标 | COD | pH | SS | 氨氮 | TN | TP |
| σ2 | 47.46 | 1.13 | 6.82 | 0.92 | 2.84 | 0.02 |
| s2 | 28.53 | 0.73 | 8.28 | 0.56 | 3.49 | 0.03 |
|  | 18.03 | 19.38 | 36.42 | 18.26 | 36.87 | 45.00 |

采用χ2检验法来判断样本相对于整体的波动性是否有显著变化，因为12月份的日观测值有31个，故样本观测个数n=31。显著水平α仍然取0.05。查χ2分布表得到，χ20.025 (30)=46.979，χ20.975 (30)=16.791。

由公式2-4可以算出各指标的χ2值，见 。当χ2 =∈(16.791，46.979)时，认为样本相对于整体的波动性没有显著变化。经比较，各指标的χ­2值均在可接受域内，可初步认为检测月份的数据比较可信。但由于监测数据的样本容量很大，监测部门想要足够精确的检测出可疑数据是很困难的。此过程的检验容易犯第二种错误，假定犯第二种错误的概率为β，现对此检验过程(即μ, σ2均未知的方差检验)影响β值因素进行分析探究。

当样本服从其中μ, σ2均未知， -∞<μ<+∞，0< σ2<+∞。

①检验问题：H0：σ2≤ ，H1：σ2 >

由上述检验过程可知这种检测问题的拒绝域为：

类似地，

其中，

②检验问题：H0：σ2 ≥ ，H1：σ2 <

同理有：

类似地，

③检验问题：H0：σ2 = ，H1：σ2 ≠

同理有：

类似地，

显然，本鉴别流程符合式3-4的情况，所以。对于一定的待测样本来说，其容量值n为定值，影响β大小的因素只有显著水平σ。当显著水平σ增大时，值增大，故β值减小。从而得出，提高显著水平σ可以有效减小犯第二类错误的概率β。监测部门想要最大程度减小犯第二类错误的机率，可以提高检测过程的显著水平。

尽管提高显著水平可以减小犯第二类错误的机率，但在给定的显著水平下，由于样本的容量过大，想要提高检测的精度，还需要判别置信区间的长度是否达到最短。下面给出单个正态总体方差最短置信区间判别定理。[78]

给定显著水平α，样本服从其中μ, σ2均未知，若方差σ2的一个置信区间为，当a，b满足如下条件：

那么为方差σ2的最短置信区间。

对a、b进行求解：令a=，利用牛顿法[79]可以求得方程

的近似解，设求出的近似解x≈b。

利用复化辛普森公式[79]可以求出积分：

在给定精度ε后，当时，则方差σ2的最短置信区间。

这里取COD指标监测数据进行计算说明。上述检测过程中已经计算出a=χ20.975 (30)=16.791，b=χ20.025 (30)=46.979。给定精度取值ε=0.001,考虑到计算量过大，用数学计算软件Matlab 完成迭代运算，迭代运算图如。

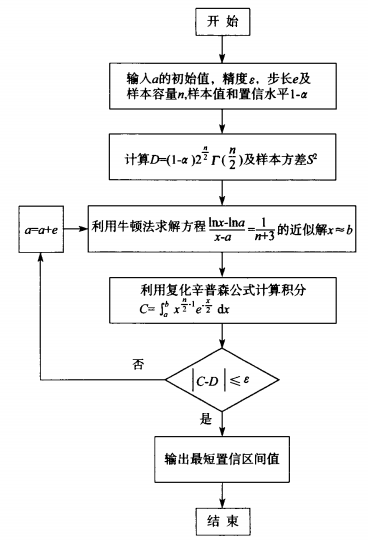
****

图 3‑8 方差σ2的最短置信区间迭代运算图

Figure 3‑8 Iterative operation diagram of the shortest confidence interval of variance σ2

当满足时，求得χ2的最短

短置信区间为(19.329，41.875)，区间长度由30.188缩短到22.546，缩短了25%的接收域长度，大大提高了鉴别的准确性。实际上，取一年的日监测值作为整体样本作此判断计算量过大，可根据实际情况简化至连续几个月的数据，以上仅提供一个鉴定计算方法。

如果企业数据部分丢失导致无法获取足够多的监测数据，或者企业采用非法手段对在线监测数据进行造假已有很长时间，上述的鉴别过程就会因为失去可靠的总体样本观测值而不能准确鉴别出虚假数据，因此下文将通过数据的横向对比来继续完成鉴别流程。

### 数据横向对比

数据横向对比即同行业同时间段的监测数据对比。对于污水处理厂来说，污水处理工艺不同，其监测指标的测定结果也不尽相同。由于厂房规模、处理工艺、操作人员等复杂因素的影响，使得不同的污水处理厂之间很难找到具有对比性的因子。为了想要突破影响因素的束缚，本文从监测数据入手，通过匹配年数据离散程度相近的样本，来匹配出具有对比价值的企业。

#### 对照样本方差范围确定

不同的污水处理厂之间的出水水质监测数据的实际波动情况都会有所差异。可能出现这样的问题：不同污水处理厂的出水水质监测数据平均值相近，但是实际波动很大，对数据的预测干扰很大，这种情况用对照样本检验就不能判断待测月份数据的可信度。因此，有必要在原料检验时对方差进行检验，以发现不同污水处理厂间数据的变化是否相近。

检验每月观测值的接近程度是检验整体数据均值的相似度，不能反映企业整体数据的波动性，对于匹配监测指标数据变化相近的企业来说，数据的整体特性尤其重要。不同污水处理厂之间除了要保证平均值相近以外，方差也要满足可接收域。方差是两个样本整体数据均匀性差异的一种体现形式。从方差可以看出匹配样本和待测样本之间变化趋势的差异，也就能通过匹配样本预测待测样本待测月份数据的可接受取值范围。

然而实际上即使计算出了待测样本整体数据的方差，往往很难判断是否有一个可匹配对象存在，因为在监测中心的数据库中储存有大量的数据，很难制定一个方差的标准来进行匹配。样本方差不超过多少就可以判定样本间变化趋势相近，工程技术上很难论证。所以，判断不同污水处理厂之间是否具有显著差异或者说是否具有相似的波动性更具有预测和对照的意义，因为整体样本的变化趋势可以更直观地描述数据特性，尤其对考核待测数据的真实度更为重要。F检验是比较两个样本观测值之间方差是否显著，即两个样本观测值均匀性的差异是否显著，换句话说，两个样本观测值是否具有相同的变化趋势。同时，F检验也会给出平均值和方差的检验结果，这样就能确定具有较高可信度的方差区间。

假定企业12月份监测数据为待测样本，各监测指标年度服从方差见，近似服从正态分布样本Y为匹配样本，年度服从方差σ2。匹配具有对比价值的企业，可看作对两个独立正态总体方差近似度的检验。考虑到本鉴别系统属于后台管理系统，在系统数据库中储存有许多的污水处理厂企业，在匹配离散程度相近的样本前，首先应当限定一些影响因素，比如污水处理工艺相同、出水水质排放量相近、污水处理厂地址临近、处理污水种类相近等。

在上述影响条件相近的基础上，采用F检验法，在给定的显著水平下，确定F的接受域，从而确定离散程度相近的样本。本文仍在显著水平α=0.05的条件下进行说明。因为样本基数趋于无穷大，F(n1,n2)近似等于1，由公式2-5可知检测失去意义，故根据实际情况考虑，这里取前两个月的样本离散程度进行研究。

查找F分布表，n1=n2=61，α=0.05，可得到 由公式2-7：0.60<<1.67，各监测指标的对照样本方差范围见。

表 3‑14 各监测指标对照样本方差范围

Table 3‑14 The variance range of check sample of each detection indexs

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 监测指标 | COD | pH | SS | 氨氮 | TN | TP |
| 下限值 | 28.476 | 0.678 | 4.092 | 0.552 | 1.704 | 0.012 |
| 上限值 | 79.258 | 1.887 | 11.389 | 1.536 | 4.743 | 0.033 |

为了提高检测的精度，还仍需要判别置信区间的长度是否达到最短。下面直接给出两个正态总体方差比最短置信区间判别定理。

给定显著水平α，样本服从其中、均未知，n、m为X、Y的样本容量，为X、Y的样本方差。若方差的置信区间为，当a，b满足如下条件：

那么为方差σ2的最短置信区间。

对a、b进行求解：令a=，利用牛顿法可以求得方程

的近似解，设求出的近似解x≈b。

利用复化辛普森公式[79]可以求出积分：

在给定精度ε后，当时，则方差比的最短置信区间。

此迭代运算过程与单个正态总体方差最短置信区间判别过程大体相同，这里精度仍然取值0.001，用Matlab计算后直接给出最终结果为(0.60,1.67)，从而计算各监测指标的方差最短置信区间上下限值见。

表 3‑15 各监测指标正态总体方差最短置信区间

Table 3‑15 the shortest confidence interval of normal population variance of each detection indexs

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 监测指标 | COD | pH | SS | 氨氮 | TN | TP |
| 方差比最短置信区间下限值 | 38.917 | 0.927 | 5.592 | 0.754 | 2.329 | 0.016 |
| 方差比最短置信区间上限值 | 68.817 | 1.639 | 9.889 | 1.334 | 4.118 | 0.029 |

由表3-15可以发现，在计算两个正态总体方差比最短置信区间后，各监测指标的区间长度都达到最短：COD的方差区间长度由50.782减短到29.900； pH的方差区间长度由1.209减短到0.712；SS的方差区间长度由7.297减短到4.297；氨氮的方差区间长度由0.984减短到0.580；TN的方差区间长度由3.093减短到1.789；TP的方差区间长度由0.021减短到0.013。各监测指标方差区间长度均比计算前缩短了41.1%。在限制条件下，由各监测指标两个正态总体方差比最短置信区间，从监测中心数据库中的匹配出方差长期服从满足该范围的企业，即可视为具有对比意义的参考对象。

#### 精确匹配对照样本

监测部门的数据库中储存了巨大数据信息，且样本方差变动较大，不能近似看作正态分布，为了精确匹配对象Y，就继续需要缩减F的可接收域。此时可通过减小显著水平α，配合使用箱线图法来缩小范围，直到匹配出一个或几个对比对象为止。

设X有匹配对象Y1，Y2…Y10，每个匹配对象各监测指标对应的方差观测值如下。

表 3‑16 匹配对象各监测指标的方差

Table 3‑16 Variance of each monitoring indexs of match object

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 匹配对象 | COD | pH | SS | 氨氮 | TN | TP |
| Y1 | 47.381 | 1.226 | 5.682 | 0.934 | 3.597 | 0.017 |
| Y2 | 37.698 | 0.928 | 7.375 | 0.985 | 3.008 | 0.020 |
| Y3 | 51.742 | 1.508 | 7.699 | 0.977 | 4.023 | 0.028 |
| Y4 | 68.900 | 1.357 | 8.520 | 1.054 | 4.015 | 0.024 |
| Y5 | 54.237 | 1.478 | 9.888 | 1.114 | 3.525 | 0.024 |
| Y6 | 43.120 | 1.638 | 7.435 | 1.330 | 3.428 | 0.025 |
| Y7 | 51.057 | 1.365 | 8.256 | 1.052 | 2.334 | 0.018 |
| Y8 | 49.052 | 1.447 | 8.541 | 1.079 | 3.872 | 0.019 |
| Y9 | 46.201 | 1.389 | 8.571 | 1.142 | 3.904 | 0.027 |
| Y10 | 45.882 | 1.279 | 8.584 | 0.992 | 3.304 | 0.023 |

由公式2-2可求得各监测指标的最小值、第一四分位值、中位数、第三四分位值、最大值、四分位距、外限、内限，从而有效判断出异常值，及其发生位置，计算结果见。

表 3‑17 箱线图法计算参数值

Table 3‑17 Calculate parameter values by box diagram method

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 监测指标 | COD | pH | SS | 氨氮 | TN | TP |
| Qmin | 37.698 | 0.928 | 5.682 | 0.934 | 2.334 | 0.017 |
| Q1 | 45.192 | 1.266 | 7.420 | 0.983 | 3.305 | 0.018 |
| Q2 | 48.217 | 1.377 | 8.393 | 1.053 | 3.477 | 0.024 |
| Q3 | 52.366 | 1.486 | 8.574 | 1.119 | 3.932 | 0.026 |
| Qmax | 68.900 | 1.638 | 9.888 | 1.330 | 4.023 | 0.028 |
| IQR | 7.174 | 0.220 | 1.154 | 0.136 | 0.627 | 0.008 |
| Q1- 1.5IQR | 34.431 | 0.936 | 5.689 | 0.779 | 2.365 | 0.006 |
| Q3+1.5IQR | 63.127 | 1.816 | 10.305 | 1.323 | 4.873 | 0.038 |
| Q1- 3IQR | 23.670 | 0.606 | 3.958 | 0.575 | 1.425 | -0.006 |
| Q3+3IQR | 73.888 | 2.146 | 12.036 | 1.527 | 5.813 | 0.050 |
| 异常值 | 68.900 | 0.928 | 5.682 | 1.330 | 2.334 | 无 |
| 发生位置 | Y4 | Y2 | Y1 | Y6 | Y7 | 无 |

以不同计数单位作为纵坐标，在二维直角坐标系上做出各监测指标的箱线图，在图中用“ο”标记出异常值的大小，并标记出发生位置。在各箱线图纵向的虚线段表示温和异常值区域，外限以外的极端异常值区域视为极端异常值区域，结果如下：

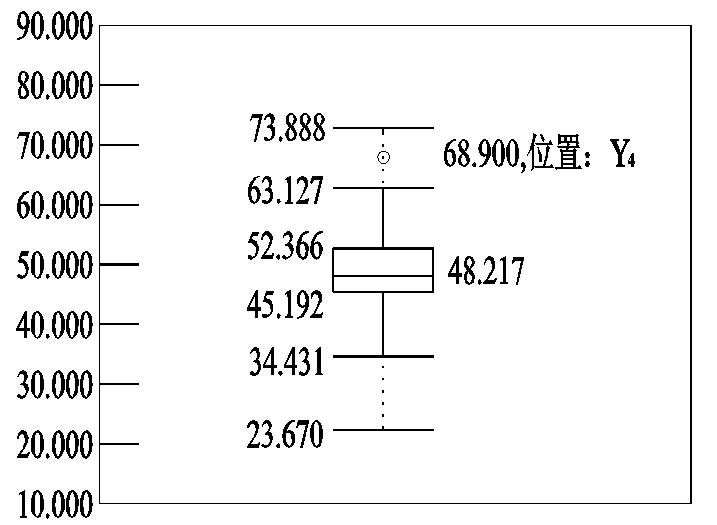
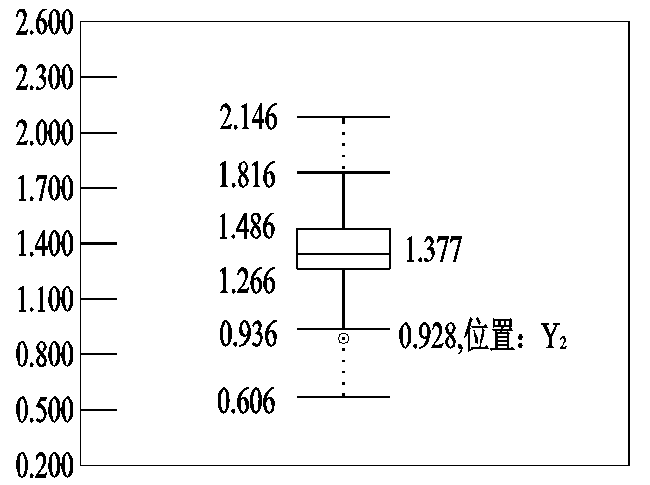
 

图 3‑9 COD方差观测值箱线图 图 3‑10 pH方差观测值箱线图

Figure 3‑9 The box diagram of Figure 3‑10 The box diagram of

COD variance observation pH variance observation

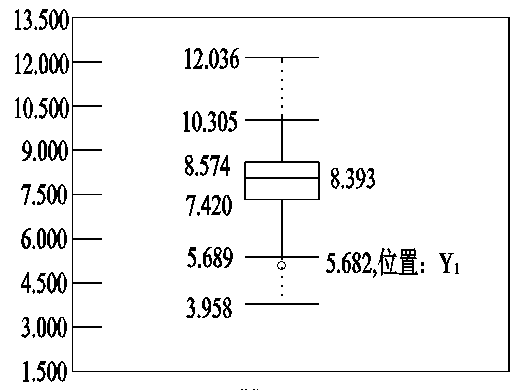
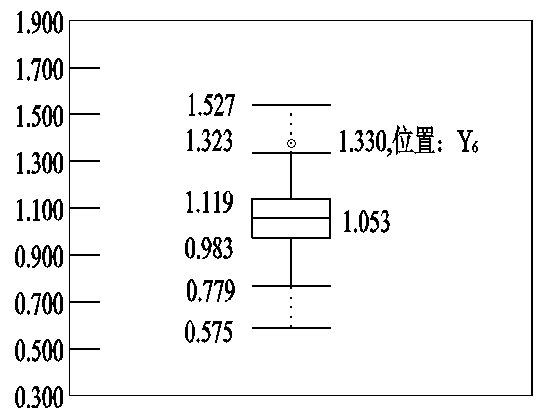
 

图 3‑11 SS方差观测值箱线图  图 3‑12 氨氮方差观测值箱线图

Figure 3‑11 The box diagram of Figure 3‑12 The box diagram of

SS variance observation ammonia nitrogen variance observation

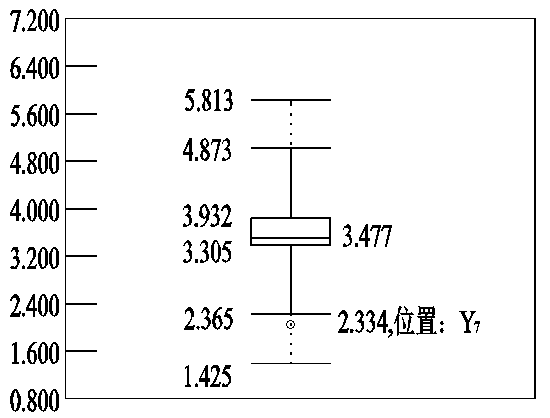
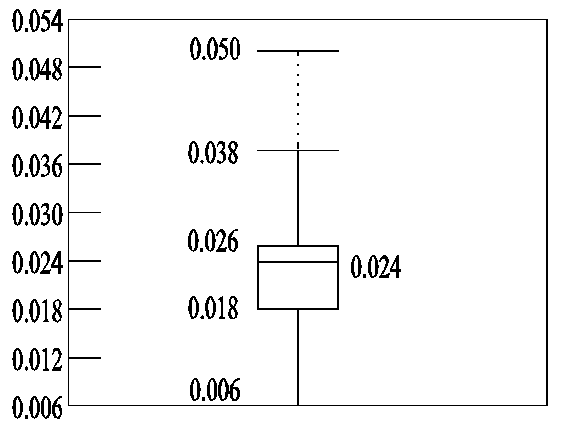
 

图 3‑13 TN方差观测值箱线图  图 3‑14 TP方差观测值箱线图

Figure 3‑13 The box diagram of Figure 3‑14 The box diagram of

SS variance observation TP variance observation

从各监测指标方差观测值的箱线图可知匹配对象Y1 ,Y2 ,Y4 ,Y6 ,Y7均有出现异常值，可以剔除。从图3-14又能发现各匹配对象的TP方差观测值均落在可接受区域内，未出现异常值。然而还剩下5个匹配对象，未能实现精确匹配的目的。对筛选出的5个匹配对象与待测样本的影响因子进行评价分析，计算各影响因子的权重，构建目标函数，最终确定最匹配对象。

在匹配过程中所确定的影响因子共有6个，依次为COD、pH、SS、氨氮、TN、TP，并用Z依次表示影响因子集，则Z={Z1, Z2, Z3, Z4, Z5, Z6}，其中Xi表示第i个影响因子。匹配的最终目的是找出最优匹配对象，各监测指标方差与对照样本方差的差值可以较为直观地反映两者之间的差异程度。设各影响因子的观测值集为z={ z1, z2, z3, z4, z5, z6},其中Xi表示第i个影响因子观测值，即匹配对象的监测指标方差与对照样本长期服从方差的差值绝对值。

匹配对象的监测指标方差与对照样本长期服从方差的差值绝对值互不相等，其影响程度也不同，为确定每种影响因子对匹配结果的影响程度，采用相对比较法确定各影响因子的权重。具体方法如下：将所有影响因子分别按行和列构成1个矩阵，根据三级比例标度，属性两两比较进行评分，并记在矩阵相应位置；再将各属性评分值按行求和，得到各属性评分总和；最后，进行归一化处理，求得各属性的权重值，即对于6个监测指标a1,a2,a3,a4,a5,a6，对前提条件进行描述，按三级比例标度两两比较评分，其中第i行第j列的评分表示为aij。

计算各影响因子的观测值如下。

表 3‑18 匹配对象的监测指标方差与对照样本长期服从方差的差值

Table 3‑18 The difference between the monitoring index variance and the long-term variance of the matched sample

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 匹配对象 | COD | pH | SS | 氨氮 | TN | TP |
| Y3 | 51.742 | 1.508 | 7.699 | 0.977 | 4.023 | 0.028 |
| Y5 | 54.237 | 1.478 | 9.888 | 1.114 | 3.525 | 0.024 |
| Y8 | 49.052 | 1.447 | 8.541 | 1.079 | 3.872 | 0.019 |
| Y9 | 46.201 | 1.389 | 8.571 | 1.142 | 3.904 | 0.027 |
| Y10 | 45.882 | 1.279 | 8.584 | 0.992 | 3.304 | 0.023 |
| X | 47.460 | 1.130 | 6.820 | 0.920 | 2.840 | 0.020 |
| |Y3- X| | 4.282 | 0.378 | 0.879 | 0.057 | 1.183 | 0.008 |
| |Y5- X| | 6.777 | 0.348 | 3.068 | 0.194 | 0.685 | 0.004 |
| |Y8- X| | 1.592 | 0.317 | 1.721 | 0.159 | 1.032 | 0.001 |
| |Y9- X| | 1.259 | 0.259 | 1.751 | 0.222 | 1.064 | 0.007 |
| |Y10- X| | 1.578 | 0.149 | 1.764 | 0.072 | 0.464 | 0.003 |

三级比例标度的含义为：

影响因子数为6，则构建评分值矩阵为：

显然，aii=0.5，。分析比较表3-18中数据，按指标变化波动大小，将重要性指标按从大到小排序为：COD > SS = TN > pH = 氨氮 > TP，即Z1 > Z3 = Z5 > Z2 = Z4 > Z6，则有：

各影响因子权重计算方法如下：

构建目标函数：

由式3-5计算权重集为W={0.343，0.105，0.210，0.105，0.210，0.027}，则目标函数为：

计算各匹配对象的目标函数值如下：

Y3：

Y5：

Y8：

Y9：

Y10：

目标函数值反映的是各匹配对象的监测指标在不同评判权重下与检测对象的差异程度，故以最小值最优的原则对结果进行评定。对上述目标函数值进行升序排序，有：Y10 < Y9 < Y8 < Y3 < Y5，鉴别认为Y10为最佳匹配对象，其它匹配对象按最小值最优的原则设定优先级，在服务器中储存备用。

#### 差值波动性分析

确定最佳匹配对象Y10后，Y企业监测数据的变化与待测企业的变化趋势相近，对X、Y样本的观察值做差Q=X-Y，对差值对象Q的波动性进行分析，如果待测月份的差值波动性较之前显著，则认为检测月份数据有造假的嫌疑。

在鉴别系统中完成上述鉴别流程后，在数据库中匹配到Y10，下表是数据库中储存的匹配对象2016年12月份各监测指标数据及其平均数和方差。

表 3‑19 匹配对象2016年12月份各监测指标数据及其平均数和方差

Table 3‑19 Monitoring data ,average values and variance of matched objects in December,2016

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 监测指标 | COD | pH | SS | 氨氮 | TN | TP |
| 2016.12.1 | 25.85 | 7.5 | 17.67 | 5.61 | 11.36 | 0.57 |
| 2016.12.2 | 25.47 | 7.7 | 14.59 | 5.60 | 11.90 | 0.73 |
| 2016.12.3 | 27.52 | 8.0 | 13.70 | 6.93 | 8.70 | 0.24 |
| 2016.12.4 | 32.01 | 6.0 | 12.83 | 6.61 | 9.75 | 0.80 |
| 2016.12.5 | 31.33 | 6.4 | 14.97 | 6.58 | 9.43 | 0.70 |
| 2016.12.6 | 31.65 | 6.8 | 14.94 | 6.84 | 13.51 | 0.92 |
| 2016.12.7 | 27.60 | 9.0 | 18.38 | 7.72 | 12.97 | 0.27 |
| 2016.12.8 | 25.63 | 8.4 | 18.52 | 6.32 | 15.49 | 0.55 |
| 2016.12.9 | 21.06 | 7.9 | 19.85 | 6.82 | 11.50 | 0.40 |
| 2016.12.10 | 21.68 | 6.7 | 19.89 | 5.63 | 14.97 | 0.89 |
| 2016.12.11 | 18.76 | 8.5 | 19.29 | 6.33 | 11.70 | 0.68 |
| 2016.12.12 | 15.47 | 6.8 | 18.28 | 5.16 | 11.89 | 0.96 |
| 2016.12.13 | 17.09 | 6.9 | 14.72 | 5.41 | 11.57 | 0.85 |
| 2016.12.14 | 24.71 | 6.0 | 14.26 | 4.60 | 9.92 | 0.85 |
| 2016.12.15 | 26.46 | 7.5 | 13.90 | 5.81 | 8.24 | 0.79 |
| 2016.12.16 | 30.26 | 6.5 | 12.99 | 4.32 | 14.32 | 0.91 |
| 2016.12.17 | 32.46 | 6.7 | 11.04 | 6.24 | 11.40 | 0.51 |
| 2016.12.18 | 35.92 | 6.8 | 14.53 | 5.85 | 13.44 | 0.93 |
| 2016.12.19 | 30.40 | 7.2 | 16.82 | 6.82 | 9.84 | 0.93 |
| 2016.12.20 | 30.52 | 6.3 | 19.36 | 6.31 | 9.61 | 0.45 |
| 2016.12.21 | 24.00 | 6.6 | 19.99 | 7.16 | 12.04 | 0.74 |
| 2016.12.22 | 24.81 | 7.1 | 18.26 | 5.65 | 13.00 | 0.38 |
| 2016.12.23 | 19.35 | 9.0 | 18.52 | 6.01 | 15.48 | 0.86 |
| 2016.12.24 | 20.28 | 8.6 | 18.53 | 5.07 | 12.51 | 0.67 |
| 2016.12.25 | 17.33 | 7.8 | 15.05 | 5.55 | 15.18 | 0.70 |
| 2016.12.26 | 17.90 | 7.7 | 13.72 | 4.36 | 9.70 | 0.50 |
| 2016.12.27 | 19.33 | 8.3 | 16.48 | 5.05 | 13.16 | 0.93 |
| 2016.12.28 | 25.34 | 6.0 | 13.30 | 4.97 | 13.42 | 0.66 |
| 2016.12.29 | 27.44 | 6.7 | 12.82 | 5.53 | 14.60 | 0.92 |
| 2016.12.30 | 31.64 | 6.6 | 15.16 | 5.40 | 10.33 | 0.85 |
| 2016.12.31 | 29.57 | 6.3 | 15.15 | 6.89 | 14.93 | 0.95 |
| 平均值 | 25.45 | 7.2 | 16.05 | 5.91 | 12.13 | 0.72 |
| 方差 | 29.76 | 0.80 | 6.53 | 0.72 | 4.47 | 0.04 |

采用χ2检验法来判断样本相对于整体的波动性是否有显著变化，因为12月份的日观测值有31个，故样本观测个数n=31。显著水平α仍然取0.05。查χ2分布表得到，χ20.025 (30)=46.979，χ20.975 (30)=16.791。

由于X、Y样本独立，且服从正态分布，设Q服从于N(μ,σ2)分布，那么，‾q , s2分别为μ和σ2的观测值，‾x,‾y,分别为的观测值，有 , 差值对象Q长期服从方差和2016年12月份的样本方差及χ2见。

表 3‑20 差值对象Q方差

Table 3‑20 Difference object’s Q variance

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 监测指标 | COD | pH | SS | 氨氮 | TN | TP |
|  | 47.46 | 1.13 | 6.82 | 0.92 | 2.84 | 0.02 |
|  | 28.53 | 0.73 | 8.28 | 0.56 | 3.49 | 0.03 |
|  | 45.88 | 1.28 | 8.58 | 0.99 | 3.30 | 0.03 |
|  | 29.76 | 0.80 | 6.53 | 0.72 | 4.47 | 0.04 |
| σ | 93.34 | 2.41 | 15.40 | 1.91 | 6.14 | 0.05 |
|  | 58.29 | 1.53 | 14.81 | 1.28 | 7.96 | 0.07 |
|  | 18.734 | 19.054 | 28.843 | 20.084 | 38.867 | 42.000 |

由上表可以发现差值对象Q各监测指标的χ2值全部落在(16.791,46.979)区间内，说明匹配样本Y10与待测样本X各监测指标在2016年12月份变化趋势相近。至此，待测样本（即12月份的各监测指标数据）通过格拉布斯法的异常值诊断、《城市污水处理厂污染物排放标准》(GB18918-2002)要求的一级B 标准的达标筛选、自身历史数据的纵向比较、同行业同时间段的横向比较，不仅验证从走访污水处理厂获取的污水监测数据具有较高可信度，也证实企业诚信度较高。

## 本章小结

（1）本章对基于数理统计的污水监测数据真伪鉴别流程作出了详细解释说明，主要包括时数据修正，规范达标修正，数据纵向比较和数据横向比较四个部分。

（2）本文基于数理统计原理，针对不同的检测流程运用不同的检测原理。基于格拉布斯法原理，实现了时监测数据中的异常值检测；基于《城市污水处理厂污染物排放标准》(GB18918-2002)划定了日监测数据的各出水指标的达标要求；基于t检验法，验证了抽检数据样本均值是否在整体样本均值的接收域；基于χ2检验法，判断出样本相对于整体的波动性是否有显著变化，完成数据的纵向对比；基于F检验法,锁定了匹配样本各监测指标方差相近的范围；基于箱线图法从许多满足方差范围的企业中筛选出部分变化趋势更相近的匹配对象；再次运用χ2检验法，通过待测对象和匹配对象的监测数据的差值波动性分析得出最终鉴别结果。

（3）横向数据对比过程中，对于匹配出的对象按照三级比例标度评判标准，设定指标的重要性级别，构造评分函数，按最小值最优原则选择最优匹配对象，完成待测对象与匹配对象的差值波动性分析。

（4）结合真实的污水处理厂监测数据，证实了鉴别流程的可行性。

# 基于B/S网络结构模式的污水监测数据真伪鉴别系统

## 引言

针对污水监测数据真伪鉴别流程需求，阐述了基于B/S网络结构模式的污水监测数据真伪鉴别系统的设计思想，并对系统的程序化设计与实现做了详细的介绍，开展了数据库概念设计、逻辑设计和物理设计，建立了原始数据增删查改模块和鉴别流程模块，通过系统功能图描述了各功能模块的具体作用。最后，以走访污水处理厂的真实数据验证了系统的实用性和准确性。

## 系统需求分析

从获得可靠的监测数据和加强污水处理厂监管力度的角度出发，致力于开发一套可以鉴别出污水监测数据真伪，用于指导城市污水处理厂的日常管理，发现虚假数据，为上级监管部门提供有力的科技保证和技术支撑的污水监测数据真伪鉴别系统。

系统数据库数据从当地监测中心以及相关网站获取。系统主要面向污水处理厂企业和监测中心管理者，既方便监测中心对各污水处理厂的管理，对企业监测数据诚信度的考核，对自身监测人员职业素养的提升，又方便各污水处理场之间的资源共享和交流合作。



图 4‑1 污水监测数据真伪鉴别系统框架

Figure 4‑1 Sewage monitoring data authenticity identification system framework

## 系统设计思想

主要从系统实用性、安全性与可恢复性、灵活性与可扩展性和稳定性与可靠性四个方面阐述系统的设计思想。

### 实用性

系统开发要“以人为本”，充分考虑城市环保系统业务管理的实际需要，贴近用户的需求与习惯做法，做到功能强大、界面友好和美观、操作简单、使用灵活。充分实现信息资源的共享，减少工作人员的工作量，实现各项业务办理的计算机协同工作环境，使工作人员办理业务过程中能方便地获得所需的信息。界面是用户接触本系统的“接口”，因此，系统提供的界面必须简单、明了，让用户方便、快捷地使用，同时也要设计得美观、大方，让用户看起来赏心悦目，用起来心情舒畅。秉承“从用户的角度出发”的宗旨，制作方便、美观的页面。系统的实用性特点如下：

（1）系统主要基于B/S架构，用户只需使用浏览器就可以使用系统全部的功能，不需要在客户机上安装任何的软件，而且系统的升级和维护只需要在服务器端完成即可，不需要做客户端的维护；

（2）系统从城市环保部门的角度出发，结合环境保护业务的需要以及我们多年设计环境信息软件系统的经验，设计出符合用户使用习惯和操作流程的界面。

（3）把复杂的操作简单化，通过应用“分步查询”、“向导”等界面，引导用户逐步的输入信息，得到结果。

### 安全性与可恢复性

信息安全在政府解决方案中是非常重要的一个环节，安全保密实现的目标是保证系统数据安全、运行稳定、不易被破坏、系统可监控、并且可恢复。

系统安全有两个层面的内容：首先是，管理制度。管理体制是系统安全的重要组成部分，也是系统安全运行的主要保障。因此，需要建立严格的管理制度，制定完善、有效的管理措施。其次是技术解决方案，在系统运行平台、网站后台操作、系统硬件三个环节上采取安全措施。

系统运行平台安全主要是指网站防黑客攻击和防病毒攻击的防范能力。主要安全措施有：

（1）通过持续不断的系统修补，防止黑客通过系统安全漏洞攻击。

（2）通过病毒防火墙建立防病毒攻击的屏障。

网站后台操作安全主要是要防止取得网站后台控制权的用户，非法操作网站后台系统。主要安全措施有：

（1）建立完善的网站管理安全规章制度

（2）后台系统记录完整的登录和操作日志，任何人对后台系统的任何操作都可以事后追查。

完善的权限管理，系统除了按照传统的用户/用户组的管理方式进行权限管理外，还根据城市环保系统的实际情况，加入按照行政区分级分地区管理的概念，即用户除了根据分配的权限获得相应的功能和访问控制能力之外，还会根据其所属的地区（城市及其下辖各县、市），分配相应地区的数据访问权限，因此，可以把管理员的权限分发给各个地区，由他们各自管理该地区的权限，以减轻中心管理人员的负担，也不会因为中心管理人员的变动而使系统的运行产生问题。

系统硬件安全主要是指由于硬件故障导致的网站服务停止或数据丢失，主要防止措施有：

（1）高效双机热备份：系统支持双机热备份的工作方式，大大提高了系统运行的容错性和不间断性。

（2）备份/恢复：系统提供自动备份和手动备份的功能，可适应用户的安全备份策略，保证数据的安全。假如系统因为硬件、操作系统等不可预知的原因导致崩溃时，系统的管理人员可以重装系统，由备份的数据文件中恢复数据，系统可以挽回从系统初运行到最后一次备份的所有数据。

### 灵活性与可扩展性

系统采用框架技术进行构建，框架设计中很关键的是在应用可能发生变化的地方增加系统的扩展性和灵活性，框架技术的主要优点是模块化、可重用、简洁可维护及可扩展。应用程序以搭积木的方式进行构建，可以非常方便增加自定义的功能，以适应特定的业务需求，在调整与扩展的同时不会对其他业务造成影响。

### 稳定性与可靠性

在系统设计、开发和应用时，从系统结构、技术措施、软硬件平台、技术服务和维护响应能力等方面综合考虑，确保系统较高的稳定性与可靠性，例如在网络环境下对系统多用户并发操作要具有较高的稳定性和响应速度，综合考虑确保系统应用中最低的故障率，确保系统的稳定性。

采用框架技术开发也是提高系统稳定性与可靠性的重要手段，稳定性是使系统持续正确的运行，不间断的提供服务的一种能力。框架技术的模块化设计把应用分割成多个组件或者模块，分而治之，从而隔离了变化的影响范围，降低系统的复杂性，使系统的稳定性得到保证。另外框架是开发工作长期的积累，是经过实践稳定下来的一组可重用的设计和代码，它是经过实践检验的，可以提供极高的可靠性。

## 数据库设计

数据库中的所有数据均以表的形式储存。整个数据库中的表分为原有表和待写表。原有表储存的是最原始的数据信息，整个鉴别流程在原有表基础上进行；待写表是空表，表中的数据随鉴别流程的进行而逐步产生。部分表之间有一定的关联，称之为表链。数据库设计目的是构建原有表、待写表以及表链，主要包括概念设计、逻辑设计和物理设计三个方面。

### 概念设计

概念设计即是利用设计概念并以其为主线贯穿全部设计过程的设计方法。在需求分析的基础上，用实体-联系模型（E-R 模型）表达现实世界的概念结构。通过对整个系统的分析，得到所有实体的E-R 图。

原有表E-R图：



图 4‑2 用户E-R图

Figure 4‑2 User E-R diagram



图 4‑3 时在线监测数据E-R图

Figure 4‑3 Timely online monitoring data E-R diagram



图 4‑4 污水处理厂基本信息E-R图

Figure 4‑4 Basic information of sewage treatment plant E-R diagram



图 4‑5 长期服从方差E-R图

Figure 4‑5 Long-term variance E-R diagram

待写表E-R图：



图 4‑6 修正后日在线监测数据E-R图

Figure 4‑6 The on-line monitoring data after correction E-R diagram



图 4‑7 纵向检测结果E-R图

Figure 4‑7 Longitudinal test results E-R diagram

****

图 4‑8 均值检测结果E-R图

Figure 4‑8 Average detection result E-R diagram



图 4‑9 鉴别流程E-R图

Figure 4‑9 Identification process E-R diagram



图 4‑10 横向检测结果E-R图

Figure 4‑10 Transverse test results E-R diagram



图 4‑11 待测对象E-R图

Figure 4‑11 Object to be tested E-R diagram



图 4‑12 匹配对象E-R图

Figure 4‑12 Match object E-R diagram



图 4‑13 匹配对象待测时间段方差E-R图

Figure 4‑13 variance be measured of the matching object E-R diagram



图 4‑14 差值对象待测时间段方差E-R图

Figure 4‑14 variance be measured of the difference object E-R diagram



图 4‑15 差值对象长期服从方差E-R图

Figure 4‑15 Long-term variance of the difference object E-R diagram



图 4‑16 差值对象E-R图

Figure 4‑16 difference object E-R diagram



图 4‑17 最终鉴别结果E-R图

Figure 4‑17 Final identification result E-R diagram

### 逻辑设计

在概念设计的基础上，将系统E-R 图转化为关系模型，定性的表达实体与实体之间的联系，整个数据库表的设计遵循第三范式进行。每张数据库中的表均设置ID属性作为主键，表中的每条数据都对应不同的ID号，有联系的表之间用箭头链接，如下所示。

原有表关系模型：

|  |  |  |
| --- | --- | --- |
| 用户表 | | |
| Link | userID | INTEGER |
|  | uname | STRING |
|  | upassword | STRING |
|  | urole | INTEGER |

|  |  |  |
| --- | --- | --- |
| 时在线监测数据 | | |
| Link | timedataID | INTEGER |
|  | T\_time | DATETIME |
|  | T\_date | DATETIME |
|  | T\_COD | DOUBLE |
|  | T\_pH | DOUBLE |
|  | T\_SS | DOUBLE |
|  | T\_N | DOUBLE |
|  | T\_TN | DOUBLE |
|  | T\_TP | DOUBLE |

|  |  |  |
| --- | --- | --- |
| 污水处理厂基本信息 | | |
| Link | factoryID | INTEGER |
|  | fname | STRING |
|  | faddress | STRING |
|  | fkind | INTEGER |
|  | ftec | STRING |
| L1 | timedataID | INTEGER |

****

|  |  |  |
| --- | --- | --- |
| 长期服从方差值 | | |
| Link | Long\_[var](http://dict.youdao.com/w/variance/#keyfrom=E2Ctranslation)ID | INTEGER |
|  | fname | STRING |
|  | LV\_COD | DOUBLE |
|  | LV\_pH | DOUBLE |
|  | LV\_SS | DOUBLE |
|  | LV\_N | DOUBLE |
|  | LV\_TN | DOUBLE |
|  | LV\_TP | DOUBLE |

图表 1 原有表关系模型

Chart 4‑1 Original table relation model

|  |  |  |
| --- | --- | --- |
| 修正后日在线监测数据 | | |
| Link | daydataID | INTEGER |
|  | tfname | STRING |
|  | D\_date | DATETIME |
|  | D\_COD | DOUBLE |
|  | D\_pH | DOUBLE |
|  | D\_SS | DOUBLE |
|  | D\_N | DOUBLE |
|  | D\_TN | DOUBLE |
|  | D\_TP | DOUBLE |

****

****

****

****

图表 2 待写表关系模型

Chart 4‑2 Table to be written relation model

### 物理设计

利用Oracle数据库建立数据表，每张表的字段及其存储如下。

表4-1至表4-4为原有表。

表 4‑1 用户表

Table 4‑1 User table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 属性名 | 数据类型 | 是否可以为空 | 字段描述 | 备注 |
| userID | NUMBER(6,0) | NO | 用户编号 | 逻辑主键自增 |
| uname | VARCHAR2(20 BYTE) | NO | 用户姓名 |  |
| upassword | VARCHAR2(20 BYTE) | NO | 用户密码 |  |
| urole | NUMBER(1,0) | NO | 角色 | 1:监测中心  0:污水处理厂 |

表 4‑2 污水处理厂基本信息表

Table ‑ Basic information of sewage treatment plant table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 属性名 | 数据类型 | 是否可以为空 | 字段描述 | 备注 |
| factoryID | NUMBER(6,0) | NO | 企业编号 | 逻辑主键自增 |
| fname | VARCHAR2(40 BYTE) | NO | 企业名称 |  |
| faddress | VARCHAR2(40 BYTE) | NO | 企业地址 |  |
| fkind | NUMBER(1,0) | NO | 企业类型 | 1:国有企业  0:非国有企业 |
| ftec | VARCHAR2(40 BYTE) | NO | 污水处理工艺 |  |
| timedataID | NUMBER(6,0) | NO | 日监测数据编号 | 外键 |

表 4‑3 时在线监测数据表

Table 4‑3 Online monitoring data table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 属性名 | 数据类型 | 是否可以为空 | 字段描述 | 备注 |
| timedataID | NUMBER(6,0) | NO | 日监测数据编号 | 逻辑主键自增 |
| T\_time | DATE | NO | 监测时间 |  |
| T\_date | DATE | NO | 监测日期 |  |
| T\_COD | NUMBER(4,2) | YES | COD时在线监测值 |  |
| T\_pH | NUMBER(4,2) | YES | pH时在线监测值 |  |
| T\_SS | NUMBER(4,2) | YES | SS时在线监测值 |  |
| T\_N | NUMBER(4,2) | YES | 氨氮时在线监测值 |  |
| T\_TN | NUMBER(4,2) | YES | TN时在线监测值 |  |
| T\_TP | NUMBER(4,2) | YES | TP时在线监测值 |  |

表 4‑4 长期服从方差值表

Table 4‑4 Long-term variance table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 属性名 | 数据类型 | 是否可以为空 | 字段描述 | 备注 |
| Long\_[var](http://dict.youdao.com/w/variance/#keyfrom=E2Ctranslation)ID | NUMBER(6,0) | NO | 长期服从方差编号 | 逻辑主键自增 |
| fname | VARCHAR2(40 BYTE) | NO | 企业名称 |  |
| LV\_COD | NUMBER(4,2) | YES | COD长期服从方差 |  |
| LV\_pH | NUMBER(4,2) | YES | pH长期服从方差 |  |
| LV\_SS | NUMBER(4,2) | YES | SS长期服从方差 |  |
| LV\_N | NUMBER(4,2) | YES | 氨氮长期服从方差 |  |
| LV\_TN | NUMBER(4,2) | YES | TN长期服从方差 |  |
| LV\_TP | NUMBER(4,2) | YES | TP长期服从方差 |  |

表4-5至表4-16为待写表。

表 4‑5 修正后日在线监测数据表

Table 4‑5 Daily online monitoring data after correction table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 属性名 | 数据类型 | 是否可以为空 | 字段描述 | 备注 |
| daydataID | NUMBER(6,0) | NO | 日监测数据编号 | 逻辑主键自增 |
| tfname | VARCHAR2(40 BYTE) | NO | 企业名称 |  |
| D\_date | DATE | NO | 监测时间 |  |
| D\_COD | NUMBER(4,2) | YES | COD时在线检测值 |  |
| D\_pH | NUMBER(4,2) | YES | pH时在线检测值 |  |
| D\_SS | NUMBER(4,2) | YES | SS时在线检测值 |  |
| D\_N | NUMBER(4,2) | YES | 氨氮时在线检测值 |  |
| D\_TN | NUMBER(4,2) | YES | TN时在线检测值 |  |
| D\_TP | NUMBER(4,2) | YES | TP时在线检测值 |  |

表 4‑6 纵向检测结果表

Table 4‑6 Longitudinal test result table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 属性名 | 数据类型 | 是否可以为空 | 字段描述 | 备注 |
| verResID | NUMBER(6,0) | NO | 纵向检测结果编号 | 逻辑主键自增 |
| tfname | VARCHAR2(40 BYTE) | NO | 企业名称 |  |
| optname | VARCHAR2(20 BYTE) | YES | 检测员姓名 |  |
| startdate | DATE | NO | 监测数据起始日期 |  |
| enddate | DATE | NO | 监测数据终止日期 |  |
| [ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)ResID | NUMBER(6,0) | NO | 均值检测结果编号 | 外键 |
| [var](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)ResID | NUMBER(6,0) | NO | 方差检测结果编号 | 外键 |
| vRes | NUMBER(1,0) | NO | 纵向检测结果 | 1:数据可靠  0:数据不可靠 |

表 4‑7 均值检测结果表

Table 4‑7 Average detection result table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 属性名 | 数据类型 | 是否可以为空 | 字段描述 | 备注 |
| [ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)ResID | NUMBER(6,0) | NO | 均值检测结果编号 | 逻辑主键自增 |
| startdate | DATE | NO | 监测数据起始日期 |  |
| enddate | DATE | NO | 监测数据终止日期 |  |
| [ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_COD | NUMBER(4,2) | YES | COD均值 |  |
| [ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_pH | NUMBER(4,2) | YES | pH均值 |  |
| [ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_SS | NUMBER(4,2) | YES | SS均值 |  |
| [ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_N | NUMBER(4,2) | YES | 氨氮均值 |  |
| [ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_TN | NUMBER(4,2) | YES | TN均值 |  |
| [ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_TP | NUMBER(4,2) | YES | TP均值 |  |
| [ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)Res | NUMBER(1,0) | NO | 均值检测结果 | 1:数据可靠  0:数据不可靠 |

表 4‑8 方差检测结果表

Table 4‑8 Variance detection result table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 属性名 | 数据类型 | 是否可以为空 | 字段描述 | 备注 |
| [var](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)ResID | NUMBER(6,0) | NO | 方差检测结果编号 | 逻辑主键自增 |
| startdate | DATE | NO | 监测数据起始日期 |  |
| enddate | DATE | NO | 监测数据终止日期 |  |
| [ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_COD | NUMBER(4,2) | YES | COD方差 |  |
| [ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_pH | NUMBER(4,2) | YES | pH方差 |  |
| [ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_SS | NUMBER(4,2) | YES | SS方差 |  |
| [ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_N | NUMBER(4,2) | YES | 氨氮方差 |  |
| [ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_TN | NUMBER(4,2) | YES | TN方差 |  |
| [ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_TP | NUMBER(4,2) | YES | TP方差 |  |
| [var](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)Res | NUMBER(1,0) | NO | 方差检测结果 | 1:数据可靠  0:数据不可靠 |

表 4‑9 横向检测结果表

Table 4‑9 Horizontal test results table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 属性名 | 数据类型 | 是否可以为空 | 字段描述 | 备注 |
| horResID | NUMBER(6,0) | NO | 横向检测结果编号 | 逻辑主键自增 |
| tfname | VARCHAR2(40 BYTE) | NO | 企业名称 |  |
| optname | VARCHAR2(20 BYTE) | YES | 检测员姓名 |  |
| testID | NUMBER(6,0) | NO | 待测对象编号 | 外键 |
| mapID | NUMBER(6,0) | NO | 匹配对象编号 | 外键 |
| gapID | NUMBER(6,0) | NO | 差值对象编号 | 外键 |
| hRes | NUMBER(1,0) | NO | 横向检测结果 | 1:数据可靠  0:数据不可靠 |

表 4‑10 待测对象表

Table 4‑10 Object to be tested table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 属性名 | 数据类型 | 是否可以为空 | 字段描述 | 备注 |
| testID | NUMBER(6,0) | NO | 待测对象编号 | 逻辑主键自增 |
| startdate | DATE | NO | 监测数据起始日期 |  |
| enddate | DATE | NO | 监测数据终止日期 |  |
| Long\_[var](http://dict.youdao.com/w/variance/#keyfrom=E2Ctranslation)ID | NUMBER(6,0) | NO | 长期服从方差编号 | 外键 |
| [var](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)ResID | NUMBER(6,0) | NO | 待测时间段方差编号 | 外键 |

表 4‑11 匹配对象表

Table 4‑11 Match object table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 属性名 | 数据类型 | 是否可以为空 | 字段描述 | 备注 |
| mapID | NUMBER(6,0) | NO | 匹配对象编号 | 逻辑主键自增 |
| mfname | VARCHAR2(40 BYTE) | NO | 匹配企业名称 |  |
| startdate | DATE | NO | 监测数据起始日期 |  |
| enddate | DATE | NO | 监测数据终止日期 |  |
| Long\_[var](http://dict.youdao.com/w/variance/#keyfrom=E2Ctranslation)ID | NUMBER(6,0) | NO | 长期服从方差编号 | 外键 |
| map[Var](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)ID | NUMBER(6,0) | NO | 匹配对象待测时间段方差编号 | 外键 |

表 4‑12 差值对象表

Table 4‑12 Difference object table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 属性名 | 数据类型 | 是否可以为空 | 字段描述 | 备注 |
| gapID | NUMBER(6,0) | NO | 差值对象编号 | 逻辑主键自增 |
| startdate | DATE | NO | 监测数据起始日期 |  |
| enddate | DATE | NO | 监测数据终止日期 |  |
| gapLong\_[var](http://dict.youdao.com/w/variance/#keyfrom=E2Ctranslation)ID | NUMBER(6,0) | NO | 差值对象长期服从方差编号 | 外键 |
| gapVarID | NUMBER(6,0) | NO | 差值对象待测时间段方差编号 | 外键 |

表 4‑13 匹配对象待测时间段方差表

Table 4‑13 Variance to be measured of the matching object table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 属性名 | 数据类型 | 是否可以为空 | 字段描述 | 备注 |
| map[Var](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)ID | NUMBER(6,0) | NO | 匹配对象待测时间段方差编号 | 逻辑主键自增 |
| mfname | DATE | NO | 匹配企业名称 |  |
| startdate | DATE | NO | 监测数据起始日期 |  |
| enddate | NUMBER(4,2) | NO | 监测数据终止日期 |  |
| [ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_COD | NUMBER(4,2) | YES | COD方差 |  |
| [ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_pH | NUMBER(4,2) | YES | pH方差 |  |
| [ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_SS | NUMBER(4,2) | YES | SS方差 |  |
| [ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_N | NUMBER(4,2) | YES | 氨氮方差 |  |
| [ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_TN | NUMBER(4,2) | YES | TN方差 |  |
| [ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_TP | NUMBER(4,2) | YES | TP方差 |  |

表 4‑14 差值对象长期服从方差表

Table ‑ Long-term variance of difference object table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 属性名 | 数据类型 | 是否可以为空 | 字段描述 | 备注 |
| gapLong\_[var](http://dict.youdao.com/w/variance/#keyfrom=E2Ctranslation)ID | NUMBER(6,0) | NO | 差值对象长期服从方差编号 | 逻辑主键自增 |
| tfname | VARCHAR2(40 BYTE) | NO | 待测企业名称 |  |
| mfname | VARCHAR2(40 BYTE) | NO | 匹配企业名称 |  |
| startdate | DATE | NO | 监测数据起始日期 |  |
| enddate | DATE | NO | 监测数据终止日期 |  |
| L[ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_COD | NUMBER(4,2) | YES | COD方差 |  |
| L[ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_pH | NUMBER(4,2) | YES | pH方差 |  |
| L[ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_SS | NUMBER(4,2) | YES | SS方差 |  |
| L[ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_N | NUMBER(4,2) | YES | 氨氮方差 |  |
| L[ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_TN | NUMBER(4,2) | YES | TN方差 |  |
| L[ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_TP | NUMBER(4,2) | YES | TP方差 |  |

表 4‑15 差值对象待测时间段方差表

Table 4‑15 Variance to be measured of difference object table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 属性名 | 数据类型 | 是否可以为空 | 字段描述 | 备注 |
| gapVarID | NUMBER(6,0) | NO | 差值对象待测时间段方差编号 | 逻辑主键自增 |
| tfname | VARCHAR2(40 BYTE) | NO | 待测企业名称 |  |
| mfname | VARCHAR2(40 BYTE) | NO | 匹配企业名称 |  |
| startdate | DATE | NO | 监测数据起始日期 |  |
| enddate | DATE | NO | 监测数据终止日期 |  |
| T[ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_COD | NUMBER(4,2) | YES | COD方差 |  |
| T[ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_pH | NUMBER(4,2) | YES | pH方差 |  |
| T[ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_SS | NUMBER(4,2) | YES | SS方差 |  |
| T[ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_N | NUMBER(4,2) | YES | 氨氮方差 |  |
| T[ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_TN | NUMBER(4,2) | YES | TN方差 |  |
| T[ave](http://dict.youdao.com/w/average/#keyfrom=E2Ctranslation)\_TP | NUMBER(4,2) | YES | TP方差 |  |

表 4‑16 最终鉴别结果表

Table 4‑16 Final result table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 属性名 | 数据类型 | 是否可以为空 | 字段描述 | 备注 |
| finResID | NUMBER(6,0) | NO | 鉴别次数 | 逻辑主键自增 |
| tfname | VARCHAR2(40 BYTE) | NO | 待测企业名称 |  |
| mfname | VARCHAR2(40 BYTE) | NO | 匹配企业名称 |  |
| startdate | DATE | NO | 监测数据起始日期 |  |
| enddate | DATE | NO | 监测数据终止日期 |  |
| optname | VARCHAR2(20 BYTE) | YES | 检测员姓名 |  |
| optdate | DATE | YES | 检测日期 |  |
| verResID | NUMBER(6,0) | NO | 纵向检测结果 | 外键 |
| horResID | NUMBER(6,0) | NO | 横向检测结果 | 外键 |
| Res | NUMBER(1,0) | NO | 最终检测结果 | 1:数据可靠  0:数据不可靠 |

## 系统结构设计

本系统的设计遵守和符合以下最基本的原则：遵循可扩展性和易维护性相结合，可靠性和稳定性相结合，先进性和成熟性相结合，以实用性和经济性相结合的原则以及用户可以积极参与原则。污水监测数据真伪鉴别系统的从系统架构上可分为三部分，即应用程序客户端、网络应用程序和数据后台。软件设计采用三层结构的设计方法，即通过服务器层实现应用前端与后台数据库的数据共享，通过html5、java语言和相关的服务程序实现网络形式的新形态管理系统。



图 4‑18 系统总体结构图

Figure 4‑18 System overall structure chart

污水监测数据真伪鉴别系统的浏览器端采用的是html5+css+javaspcript框架。HTML5（超文本标记语言）结构上清晰明确，取消了一些过时的标记，将内容和展示进行分离，改进相关结构后，生成的网页更干净和易于管理，提升了用户体验，有利于搜索引擎和屏幕阅读器进行抓取。CSS（层叠样式表）是用于控制网页样式并允许将样式信息与网页内容分离的一种标记性语言。它有以下几个显著优势：表现和内容相分离；提高页面浏览速度；易于维护和改版。JavaScript是一种介于Java和HTML之间、能嵌入Web 页面中基于对象和事件驱动的解释性的编程语言。JavaScript使信息和用户之间不再只是简单的显示与浏览的关系，更可以进行实时、动态、可交式地表达。JavaScript可以被嵌入到HTML文件中，不需要经过Web服务器就可以对用户操作作出响应，使网页更好地与用户交互；在利用客户端个人电脑性能资源的同时，适当减小服务器端的压力，并减少用户等待时间。[80]

服务器端分两大模块，而且均为多线程程序：①服务器与客户机通讯模块。服务器选定一空闲端口，启动ServerSocke，等待客户与之通讯，客户与服务器成功建立通信后，启动一线程进入基于JDBC-ODBC桥驱动的JAVA数据库应用程序模块。SQL查询操作完成后，通过客户与服务器通信将结果传送给客户。②用基于JDBC-ODBC桥驱动的JAVA数据库应用程序模块，完成对本地数据库的查询操作，注册JDBC数据库驱动程序，建立应用程序与数据库的连接，等待客户发送SQL语句，以便对数据库进行操作，执行客户端发送的SQL语句。

基于 B/S 系统架构的污水监测数据真伪鉴别系统的逻辑结构规划包含表现层、业务功能处理层和数据层三部分。

表现层：表示层位于第一层，其与用户能够进行直接接触，可以把用户的逻 辑业务请求输入到系统中，表示层将用户的业务请求发送到业 务功能处理层，之后再把业务功能处理层和数据功能处理层处 理的结果反馈给用户，将信息显示在用户终端上，呈现与用户，供其进行浏览。

污水监测数据真伪鉴别系统中表现层主要指对外展示的用户界面。系统管理人员通过界面与系统后台逻辑相交互，登陆用户进入系统要对用户权限进行验证，系统根据管理员给用户分配的角色赋予不同的系统功能，如不同权限的信息展示等。系统中用户可通过图形界面更直观的进行操作。

业务功能处理层：业务功能处理层位于 Web 服务器上，其主要功能是接收表示层所传送来的应用请求进行处理，并在业务逻辑的处理过程中，可以实时的检测到用户的逻辑业务请求，发现污水监测数据真伪鉴别系统中存在的逻辑业务处理功能，实现污水监测数据真伪鉴别系统的数据处理，比如可以解析出来系统相关的SQL处理语言，并且对系统的程序进行操作，反馈给表示层，并把请求处理的结果返回到表示层。

污水监测数据真伪鉴别系统业务功能处理层是系统中核心角色层，包括原始数据操作模块功能和鉴别流程模块。原始数据操作模块包括污水处理厂的各监测指标数据，用户可以在此模块实现数据的增删查改。鉴别流程模块为系统的核心功能模块，包括格拉布斯法修正时观测值模块、筛选达标日观测值并修正模块、数据纵向对比模块和数据横向对比模块。

数据层：数据库功能处理层位于数据库管理系统中，在 B/S 架构 里，数据功能处理层主要是对逻辑层传送来的应用数据请求 进行处理，数据库的操作引擎实现了此层数据处理的过程， 具备庞大的数据操作的性能，可以对数据库进行查询、更新 等操作，并且把数据操作的结果返回于系统逻辑层，进而返 回给客户端的表示层，把操作的结果提供给用户浏览。

污水监测数据真伪鉴别系统采用oracle数据库对结构化的数据进行存储；对非结构化数据运用文件等存储方式，但对整个存储数据无论是存储在关系数据库还是存储在非关系数据库中的数据，它们之间都建立一定映射关系。

## 系统功能设计

系统功能设计包括用户登录界面设计、原始数据操作模块设计和鉴别流程模块设计三个方面。每项功能均设计API接口，用于实现前端和服务器，服务器和数据库之间的交互。每项设计在数据库中均对应有指定的数据表，用于储存用户登录信息和操作信息。

### 用户登录界面设计

系统把用户登录时的身份设定不同权限，数据库中的数据信息按权限分级展示。用户信息全部储存数据库的用户表中，并按照身份设置不同的保密级别。以监测中心身份注册的用户可以查看数据库中所有企业的数据信息，但是不能对企业原始数据进行修改；以污水处理厂身份注册的用户只能查看和改动企业自身的监测数据信息，对其他企业的数据信息只有在得到权限许可后方可查看。用户的登录界面如下图。



图 4‑19 污水监测数据真伪鉴别系统登录界面

Figure 4‑19 The login interface of sewage monitoring data authenticity identification system

### 原始数据操作模块设计

原始数据操作模块包括原始数据的增删查改。在这个模块中，用户可以通过已知条件从数据库中获取原始数据信息。鉴别流程模块运用的数据信息基于原始数据操作模块，因此在企业历史原始数据要发生变动时，企业须通告监测中心并获取数据库的相应权限，数据库中会储存相应修改记录。原始数据操作模块的操作数据来源仅限于数据库中的原有表。原始数据操作模块的各项操作模块如下。







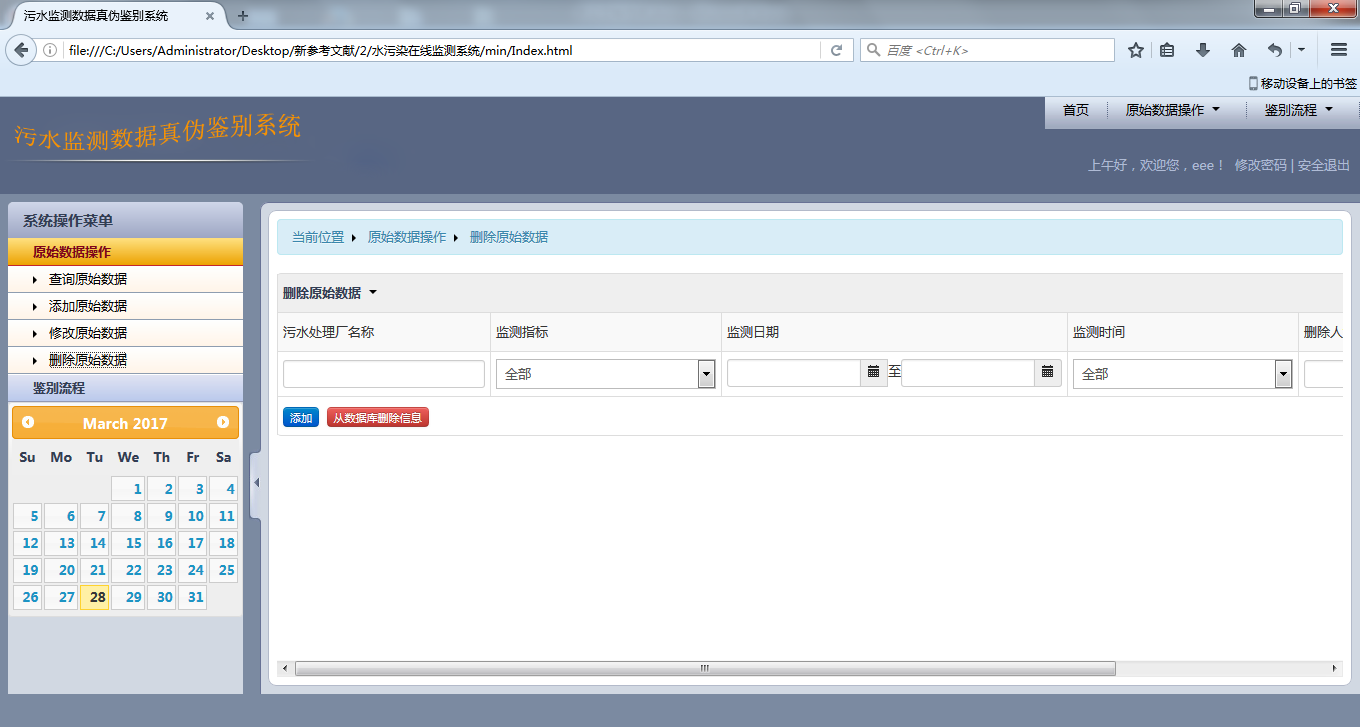
****

图 4‑20 原始数据操作模块界面

Figure 4‑20 Original data operation module interface

### 鉴别流程模块设计

鉴别流程模块设计以鉴别流程为依据，设定包括格拉布斯法修正时观测值模块、筛选达标日观测值并修正模块、数据纵向对比模块和数据横向对比模块四个模块。每个操作模块的鉴别结果均可以实时查看，数据纵向对比模块和数据横向对比模块在格拉布斯法修正时观测值模块和筛选达标日观测值并修正模块的操作基础上进行。每个操作模块产生的数据信息都实时存放于数据库的待写表中，便于操作人员查看历史鉴别信息，具体操作流程以及鉴别系统的运行情况在4.7章中以实例操作说明。

## 实例操作说明

以甲污水处理厂（国控企业）为例，获取其2016年年度监测数据和同地区部分污水处理厂的基本信息及其2016年年度监测数据建立数据库。数据库中已设定相应原有表和待写表。对从甲污水处理厂获取到的2016年12月份污水监测数据按鉴别流程进行检测，最终系统给出数据真伪判断结果。

首先在登录界面以用户名eee，监测中心的身份登录系统。进入鉴别流程模块的格拉布斯法修正时观测值模块，输入污水处理厂名称、显著水平、检测起止日期和检测指标，单击“修正并显示结果”按钮，系统后台自动搜索数据库中指定时间段的监测数据，并在设定的显著水平下完成格拉布斯法计算过程。如果有异常值出现，会在“是否存在异常值”后面显示“是”，并在下面显示检测时间段数据表格，在“出现坐标”后显示异常数据出现的行列数。点击“剔除异常值继续修正”按钮后，即可进行格拉布斯法的迭代计算过程，直到不再出现异常值为止，见。



图 4‑21 格拉布斯法修正时观测值界面

Figure 4‑21 Corrected observation by Grubbs 'Criterion interface

本例假定在14：00在监测点稀释污水，导致污染物浓度有所降低，根据异常值出现的坐标位置可以找到数据记录，如下图。



图 4‑22 异常数据记录

Figure 4‑22 Abnormal data record

从检测结果可以看出，系统很准确地找出了异常值发生位置，确定异常值为14：00时刻的COD浓度和氨氮浓度，与实际情形相符合。当所有异常值都提出以后，进行筛选达标日观测值并修正模块操作。系统中设定以《城市污水处理厂污染物排放标准》(GB18918-2002)为参照依据，对待测时间段的时监测数据达标筛选，取12个时刻监测数据的平均值作为日均值，结果如下图所示。



图 4‑23 筛选达标日观测值并修正模块操作界面

Figure 4‑23 Screening and correcting standard daily observations module interface

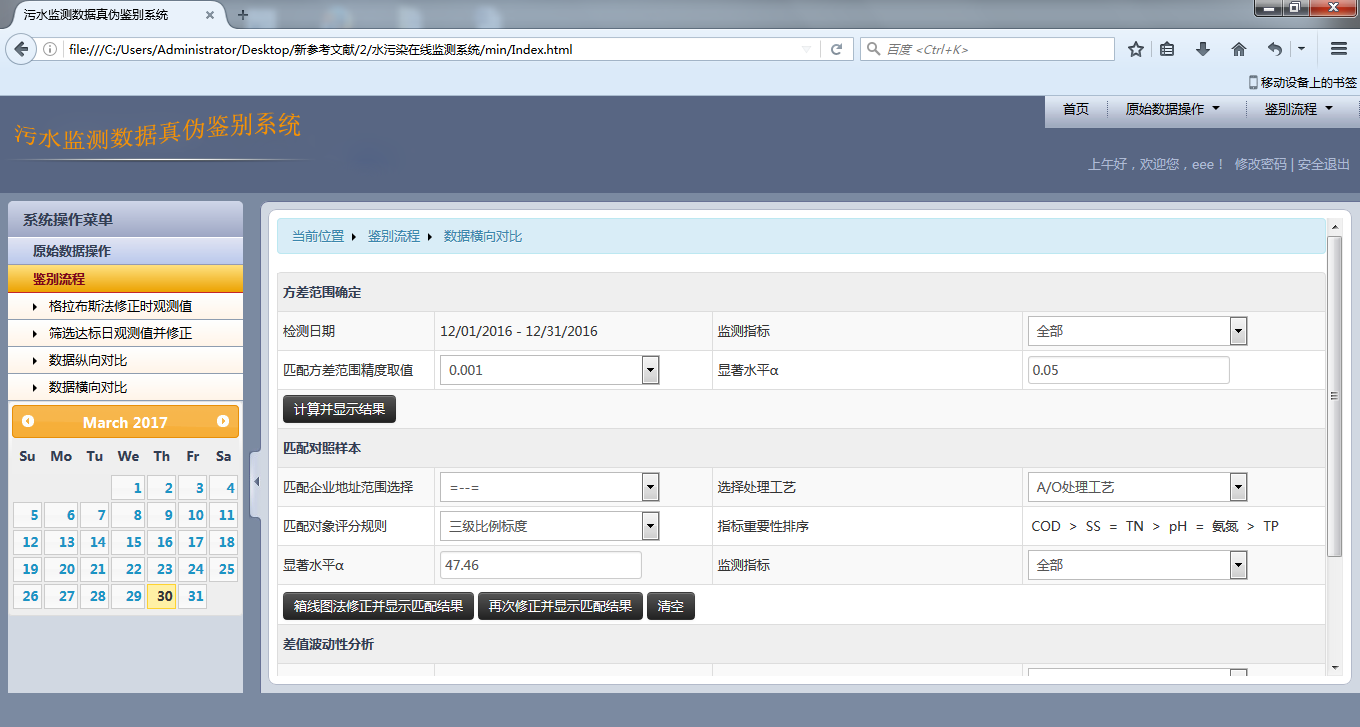
经过上述两个模块检测完毕后，开始数据的纵横对比。数据纵向对比包括月均值检验和月均值波动性检验，都需要选择对比月份。本例取年度数据验证系统实用性，故均选择前11月的均值和方差做实例说明。显著水平仍然选取0.05，当点击“开始检测”按钮后，系统自动完成t检验和χ2检验的计算过程。此过程会访问数据库中的长期服从方差表，并把检测过程的数据结果自动写入数据库的部分待写表中，在界面仅向用户展示检测结果，如下图所示。



图 4‑24 数据纵向对比操作界面

Figure 4‑24 Longitudinal data comparison module interface

数据横向对比包括包括方差范围确定、匹配对照样本和差值波动性分析。在进行方差范围确定时，通过数据库中储存的待测企业长期服从方差用F检验法确定方差可接受范围。本例选定精度为0.001，系统后台自动完成方差最短置信区间的迭代运算，直到达到要求精度为止。点击“计算并显示结果”按钮后，系统根据方差最短置信区间从数据库中匹配满足要求的企业，并弹出对话框显示方差范围和匹配结果，如下所示。



****

图 4‑25 方差范围确定操作界面及结果显示

Figure 4‑25 Variance range determination module interface and results show

由上图可以看出满足方差最短置信区间的污水处理厂在数据库中有许多，继续设定筛选条件，比如地区范围、污水处理工艺等，从数据库的污水处理厂基本信息数据表中，匹配更符合条件的对比对象。再运用箱线图法对匹配对象的各监测指标在异常判别，筛选出变化趋势更相近的对象。运用三级比例标度评分规则，在设定重要性排序下计算各个指标的评分值，为选择最佳匹配对象做参考，如下所示。





图 4‑26 匹配对照样本操作界面及结果显示图

Figure 4‑26 Matched control sample module interface and results show

匹配结果会出现在差值波动性分析的“选择最优匹配对象”的选项中，根据最小值最优的原则，选择乙污水处理厂进行差值波动性分析，最终得到检测结果，认为原始数据的可靠性较高，证实与实际相符，如下所示。



图 4‑27 差值波动性分析操作界面

Figure 4‑27 Difference volatility analysis module interface

## 本章小结

通过本章的研究分析，得出如下结论：

（1）本章详细阐述了基于数理统计的污水监测数据真伪鉴别系统的设计思想，并在开展需求分析、系统详细设计、数据库设计的基础上，程序化实现了基于数理统计的污水监测数据真伪鉴别流程，主要包括原始数据操作和鉴别流程两大模块。

（2）结合某城市污水处理厂2016年12月份各监测指标的数据进行了真伪鉴别，鉴别出在假定某时刻稀释监测点污水后，COD和氨氮浓度发生显著变化，从而导致监测数据失真。通过系统对数据修正后，完成了后续检测步骤，并成功匹配出变化趋势相近的企业，结合实际情况验证了系统的合理性和实用性。

# 结论与建议

## 结论

本文为减轻虚假污水监测数据给环境监测工作带来的巨大压力，以某城市某污水处理厂2016年年度污水监测数据为研究对象，开展了基于数理统计的污水监测数据真伪鉴别研究和基于B/S网络结构模式的污水监测数据真伪鉴别系统研究，最终实现了城市污水污水监测数据真伪鉴别系统的程序化，得出的结论如下：

（1）基于数理统计的污水监测数据真伪鉴别流程主要包括时数据修正，规范达标修正，数据纵向比较和数据横向比较四个部分。每个部分运用不同的数理统计原理：基于格拉布斯法原理，实现了时监测数据中的异常值检测；基于《城市污水处理厂污染物排放标准》(GB18918-2002)划定了日监测数据的各出水指标的达标要求；基于t检验法，验证了抽检数据样本均值是否在整体样本均值的接收域；基于χ2检验法，判断出样本相对于整体的波动性是否有显著变化，完成数据的纵向对比；基于F检验法,锁定了匹配样本各监测指标方差相近的范围；基于箱线图法从许多满足方差范围的企业中筛选出部分变化趋势更相近的匹配对象；再次运用χ2检验法，通过待测对象和匹配对象的监测数据的差值波动性分析得出最终鉴别结果。

（2）在实地走访、网上查阅、专家咨询及数据分析的基础上，确定了六项污水监测指标（COD、pH、SS、氨氮、TN、TP），对走访污水处理厂在2016年内变化趋势进行了模拟，并预测未知时间段的数据波动范围。根据模拟情况，发现待测月份COD、SS、氨氮、TN和TP的出水浓度都落在各自的接受域内，说明显著水平为0.05时，污染物的浓度值具有较高的可信度；虽然pH值不在可接受域内，但通过图表可以发现同一月份的各监测指标的变化趋势大体相同，因此对数据的真实性不具有否定意义。

（3）横向数据对比过程中，对于匹配出的对象按照三级比例标度评判标准，设定重要性指标按从大到小排序为：COD > SS = TN > pH = 氨氮 > TP，构造评分函数，按最小值最优原则选择出最优匹配对象，完成待测对象与匹配对象的差值波动性分析，不仅验证从走访污水处理厂获取的污水监测数据具有较高可信度，也证实企业诚信度较高。

（4）提出了基于数理统计的污水监测数据真伪鉴别系统的设计思想，并在开展需求分析、系统详细设计、数据库设计的基础上，程序化实现了基于数理统计的污水监测数据真伪鉴别流程，主要包括原始数据操作和鉴别流程两大模块。

（5）采用B/S网络结构模式，结合某城市污水处理厂2016年12月份各监测指标的数据进行了真伪鉴别，鉴别出在假定某时刻稀释监测点污水后，COD和氨氮浓度发生显著变化，从而导致监测数据失真。通过系统对数据修正后，完成了后续检测步骤，并成功匹配出变化趋势相近的企业，结合实际情况验证了系统的合理性和实用性。

## 建议

(1) 基于数理统计的污水监测数据真伪鉴别流程要求数据近似呈正态分布，如果是离散型数据，则可能产生与实际情况相违背的鉴别结果，对数据的限制性比较大。对于离散数据的真伪鉴别有待更深入的研究。

(2) 基于数理统计的污水监测数据真伪鉴别流程的检测精度有限，仅仅是改变显著水平和缩减可接受域的上下限来提高精度，对变动差异较小的数据不能准确的鉴别出来，后续可在如何提高检测精度方面进行更深入研究。

(3) 基于数理统计的污水监测数据真伪鉴别流程对于最终的鉴别结果仅仅给出是与否的结论，仅仅介于0和1之间的选择，不够细化，建议在后续研究中拓展为分值鉴定，给各项流程的检测结果打分，最后按权重值给出总的评分，通过评分高低来评定数据的真实性和可靠度。

(4) 基于数理统计的污水监测数据真伪鉴别系统数据库设计不够规范，逻辑联系较复杂，有冗余现象，对服务器访问数据库时长有一定影响，数据库资源开销比较大，对如何消除数据库中各种冗余现象有待继续研究。

(5) 基于数理统计的污水监测数据真伪鉴别流程系统中数据信息不够充足，在匹配变化趋势相近的企业时，对外界的影响因素仅仅考虑了地区性和污水处理工艺流程，未考虑天气因素、人为因素等，后续研究可加以补充。

参 考 文 献

[1]吴文佳,吴丽茹.污水排放的远程监视监控系统[J].环境污染治理技术与设备.2003,(1):85-89.

[2]张丽琼.污水监控系统在城市环境保护控制中的应用[J].低碳世界.2013,(6).

[3]傅其凤,安旭朝,陈国庆.工业园区污水水质和流量在线监控系统的应用[J].给水排水.2015,(11):100-102.

[4]葛玉连,赵保康,喻一萍.浅谈城市污水处理及工业污水在线监控[J].排灌机械.2005,(6):34-37.

[5]李会来.污水监控系统现状研究[J].城市建设理论研究(电子版).2015,(9).

[6]赵小强.基于GSM的远程污水综合监控分析系统的设计[J].微计算机信息杂志.2010,(5):91-92,85.

[7]刘宇.城市污水处理厂A2/O工艺故障诊断专家系统研究[D].哈尔滨工业大学,2013.

[8]宋小燕,刘锐,税勇等.间歇曝气 SBR 处理养猪沼液的短程脱氮性能[J].环境科学,2016,37(5):1873-1879.

[9]魏武强,Wisaam S. Al-Rekabi. Combined process of sequencing batch reactor activated sludge process and constructed wetland for domestic wastewater treatment[J]. Journal of Chongqing University(English Edition),2010,04:201-207.

[10]SHIGENORI SHOJI,MITURU INOUE,OSAMU NISHIMURA,YUHEI INAMORI. Advanced Treatment Using Sequencing Batch Reactor Activated Sludge Process Combined with UF Membrane Process[J]. Japanese Journal of Water Treatment Biology,2000.

[11]陈孟林,宿程远,王全喜,何星存,黄智. 吸附-催化氧化再生法处理印染废水的试验研究[J]. 工业水处理,2010,09:46-49.

[13]Saddig D. Jastaniah,Bassam Z. Shakhreet,Hanan Y. Abbas,Awad M. Elkhadir,Saeed M. Bafaraj. Treatment of Radon Rich Bottled Water by Granular Activated Carbon Adsorption Method[J]. Open Journal of Biophysics,2014,04.

[14]刘珮,袁林江,陈希,鲁文娟. 低负荷氧化沟系统中EPS与活性污泥沉降性能的关系[J]. 环境科学学报,2013,06:1611-1615.

[15]Vasileios Diamantis,Ioannis Papaspyrou,Parasxos Melidis,Alexander Aivasidis. High aeration rate enhances flow stratification in full-scale oxidation ditch[J]. Bioprocess and Biosystems Engineering,2010,332:.

[16]Y. Nakamura,M. A. Ishiksws,H. Nakanishi. An Improvement of an Oxidation Ditch Process for the Low Alkalinity Wastewater Treatment and Various Influent Conditions[J]. Environmental Technology,1997,186:.

[17]张立秋,张可方,李淑更,吴学伟. ICEAS反应器处理城市污水最优工况[J]. 水处理技术,2006,05:58-60.

[18]Farham Karakani ,Amir Hossein Mahvi. Wastewater Phosphorus Removal by Intermittent Cycle Extended Aeration System[J]. Pakistan Journal of Biological Sciences.2005,82:.

[19]Mortazavi S B,Khavanin A,Moussavi G,Azhdarpoor A. Removal of sodium dodecyl sulfate in an intermittent cycle extended aeration system.[J]. Pakistan Journal of Biological Sciences,2008,112:.

[20]郭小马.改良A2O-侧流除磷工艺脱氮除磷及微生物菌落特征研究[D].中国海洋大学,2015.

[21]Giorgio Mannina,Marco Capodici,Alida Cosenza,Daniele Di Trapani. Carbon and nutrient biological removal in a University of Cape Town membrane bioreactor: Analysis of a pilot plant operated under two different C/N ratios[J]. Chemical Engineering Journal,2016,:.

[22] J. Carrera,M. Sarra,F. J. Lafuente,T. Vicent. Effect of Different Operational Parameters in the Enhanced Biological Phosphorus Removal Process. Experimental Design and Results[J]. Environmental Technology,2001,2212:.

[23] Chang[CJ](http://xueshu.baidu.com/s?wd=author%3A%28Chang%2C%20Che-Jung%29%20&tn=SE_baiduxueshu_c1gjeupa&ie=utf-8&sc_f_para=sc_hilight%3Dperson)，Li [DC](http://xueshu.baidu.com/s?wd=author%3A%28Li%2C%20Der-Chiang%29%20&tn=SE_baiduxueshu_c1gjeupa&ie=utf-8&sc_f_para=sc_hilight%3Dperson) ，Huang[YH](http://xueshu.baidu.com/s?wd=author%3A%28Huang%2C%20Yi-Hsiang%29%20&tn=SE_baiduxueshu_c1gjeupa&ie=utf-8&sc_f_para=sc_hilight%3Dperson) et al.A novel gray forecasting model based on the box plot for small manufacturing data sets[J].Applied mathematics and computation,2015,265:400-408.

[24]孙向东,刘拥军,陈雯雯等.箱线图法在动物卫生数据异常值检验中的运用[J].中国动物检疫,2010,27(7):66-68.

[25]王新峰,梁杏,孙蓉琳等.基于野外测量数据的裂隙间距箱线图法初探——以锦屏一级水电工程厂房为例[J].安全与环境工程,2010,17(2):103-107,122.

[26]张海燕,王新民,陈初雨等.干贮粉煤灰降水渗透深度的统计分析[J].世界地质,2015,34(4):1131-1139.

[27]刘瑶,于雷,余向勇等.箱线图在河流水质评价中的应用研究[C].中国环境科学学会2008年学术年会论文集.2008:600-604.

[28]陈任秋,刘军.基于格拉布斯准则的DS证据理论在WSNS中的应用[J].传感器与微系统,2015,(6):153-155,160.

[29]池庆玺,司锡才.基于格拉布斯准则的雷达信号分选方法的探讨[J].传感技术学报,2006,19(6):2625-2629.

[30]邬晓光，姜海涛，张柳煜.运用格拉布斯准则原理确定定额测定中不合理数据[A].第四届全国公路科技创新高层论坛[C],2008

[31]邬蓉蓉,张炜,樊宇璐等.变压器油中气体监测数据随机误差的自适应检验方法[J].电气技术,2016,(11):91-95.

[32]郭立水,王海霞.用格拉布斯法排除在线灰分仪样本数据中的异常数据[J].煤质技术,2016,(3):21-23.

[33]王首绪,龙琰,郭四泽等.基于t分布理论的公路定额数据小样本容量的收敛条件[J].长沙理工大学学报(自然科学版),2007,4(2):13-15,20.

[34]Zhang[R S](http://www.baidu.com/link?url=RUeCwGGFXzEYuP81tm4WctSXQP27HFonRKE9M8fWG2rKpWaW53jlY5nImhxlm3A8jwof_yoNfQ5gKAFNcoVqCvbGBwbz6E2MxpYPGQgKU-F9bwwnUEyCiCfykXXoqJzZ5gGBw_YSMB1ScV89CCNKkHwcMP8-ZXQzs5zjCE7NfjLuxE0eSOWhBNYdHuGy5FnE) , [Li](http://www.baidu.com/link?url=I8fSETN1VJOKE53STQnfF3d3VhCPLhtABT2N1r_rdJJQUg5_UJ0wY4x1QxQmNEjPnJ8MiTpEP6coIZSjaAOftXX0lq_GOu3wuZYW8HYIteMxopcIRBKLESSSKS71Tr3kMYqisfjp2pgni4aa7yu8xnvw1pkf2wW7jm1BuPm0PWq) Y B, Li [X T](http://www.baidu.com/link?url=S8VVIpPTgKqV3V3gkx4HnoNhdGSXxY5meYhgncRJqTkn4SlNnGYP8yiH00EuziQ9fk8zc4O9gOx4DMtmYCc3kZ8N5Bn6p2Z4HvqmuD532v7ULrzw9iqv_SOvdxKlaKBmdEqr8WjfAzrOFarp0DeKNBTTThbHASWmw41kTcw7yYO7x2UoLyb-K8FSaU0QaQy_).Topology Inference With Network Tomography Based on t-Test[J].IEEE communications letters,2014,18(6):921-924.

[35]彭定赟.基于单个系数的t检验和F检验的一致性探讨[C].//中国数量经济学会2007年年会论文集.2007:10-15.

[36]李简微.配对t检验与成组t检验的优选方法[J].医学美学美容（中旬刊）,2015,(5):736-737.

[37]董秀玥.配对t检验与成组t检验优选方法研究[J].数理医药学杂志,2010,23(1):11-14.

[38] 董亮星.t检验和F检验在期间核查中的应用[J].计量技术,2011,(6):59-62.

[39] 陈创买,周文.关于因子显著性问题的F-信度检验法[J].热带气象学报,2007,23(6):679-682.

[40]胡锡春,刘再旺,韩晓磊.无锡市芦村污水处理厂先进工艺及在线监控技术的运用[J].科技信息. 2010,(23).

[41]郝雅青.明渠污水智能流量计量WEB监控平台软件实现[D].太原理工大学, 2010.

[42]刘爽.玉米低温冷害远程诊断管理系统设计与实现[D].中国农业科学院, 2008.

[43]陈胜春.城市重点污染源在线监控数据管理系统设计方案[D].电子科技大学,2010.

[44]宋庠明.基于CAN及Web污水处理在线监控系统[J].中国高新技术企业,2008,(17):153,156.

[45]张学雷.省级环境在线监控数据库平台建设与应用[J]. 安徽职业技术学院学报,2010,03:16-19.

[46]肖树臣,秦玉勋,韩吉庆. 基于格拉布斯法的试验数据分析方法[J]. 弹箭与制导学报,2007,(01):275-277

[47]雷晓平. 浅谈可疑数据的取舍方法—格拉布斯法[J]. 河南建材,2011,(02):163+165.

[48]刘新惠. 用格拉布斯法解决样本数据中的异常数据问题[J]. 财贸研究,2000,(02):78.

[49]魏治文,程琳,来记桃,吴火兵. 几种异常值判别准则在安全监测数据处理中的应用[J]. 大坝与安全,2009,(01):67-69+84.

[50]林乐明. 关于假设检验中两类错误的探讨[J]. 财经问题研究,1985,(04):79-82+32.

[51]杨帆,陈良强,罗汝叶,王和玉,汪地强,王莉. 基于箱线图对大曲质量的判别分析[J]. 酿酒科技,2015,(05):1-3.

[52]查如琴. 简谈几种“箱线图绘制”的描述[J]. 读与写(教育教学刊),2012,(07):54+63.

[53]雍红月. 运用EXCEL进行经济指标的箱线图分析[J]. 统计教育,2007,(02):35-36.

[54]梁飞豹,林同华. 线性模型有偏估计的一种新算法[J]. 合肥工业大学学报(自然科学版),2016,(06):854-858.

[55]赵静波. 社会偏好分布与共识达成的描述性模型——箱线模型的建构及应用[J]. 甘肃行政学院学报,2015,(06):40-48+106+126.

[56]林小妹，舒芳誉.一元在线质量监控图的设计与应用[J].科技创新与应用.2016,(17)：21-22.

[57]陈建兰，胡晓敏.基于教学质量评价的教学方法的探讨[J].教育教学论坛.2015,(25)：177-178.

[58]吕鑫，郝连秀，郭庆彪.安徽省空气质量时空分布特征分析[J].长江科学院院报.2016,(12)：144-147，152.

[59]朱红霞，赵淑莉，阚海东.2013年我国典型城市大气污染物浓度分布特征[J].环境科学与技术.2015,(6)：227-233.

[60]杨斌.正态性检验的几种方法比较[J].统计与决策.2015,(14)：72-74.

[61]米子川.基于图形分析方法的函数型数据异常值检验实证研究[J].统计与信息论坛.2014,(6)：18-24.

[62]冯变英,张旭,张春枝. 关于t检验方差分析及多重比较的研究[J]. 太原师范学院学报(自然科学版),2012,(04):59-61.

[63]査如琴. 基于SPSS的双总体(σ12, σ22未知,n≤30)配对样本t检验与独立样本t检验[J]. 读与写(教育教学刊),2016,(07):44-45.

[64]张勇,朱郭勤. 水准点稳定性F检验法与t检验法异同的研究[J]. 铁道勘察,2010,(04):21-24.

[65]印鉴,谭焕云. 基于χ2统计量的kNN文本分类算法[J]. 小型微型计算机系统,2007,(06):1094-1097.

[66]刘泽显,莫达隆,冯祖针.χ2拟合优度检验在教学中的应用及Matlab实现[J]. 长春大学学报,2012,(04):419-422.

[67]曾珍,张宇. χ2分布、t分布、F分布与正态分布间的关系[J]. 湖北师范学院学报(自然科学版),2015,(03):62-66.

[68]靳庭良,张宝青. 回归分析中t检验与F检验关系的进一步探讨[J]. 统计与决策,2009,(21):7-9.

[69]李志财. 方差齐性检验在检测质量控制中的应用[J]. 现代测量与实验室管理,2014,(06):33-34.

[70]邱红兵,罗季. 线性模型中F-检验的稳健性[J]. 数学学报,2010,(02):385-392.

[71]Chien-Hsiu Lee，Chien-Cheng Lin.Double-lined M dwarf eclipsing binaries from Catalina Sky Survey and LAMOST[J].Research in Astronomy and Astrophysics.2017,(2)：29-34.

[72]David Hull.Value in What is Saved and What is Lost:Textology in Mao Dun＇s Eclipse[J].中国文学研究前沿(英文版).2016,(2)：204-233.

[73] 刘洪星,谢玉山.Eclipse开发平台及其应用[J].武汉理工大学学报（信息与管理工程版）,2005,27(2):89-92.DOI:10.3963/j.issn.1007-144X.2005.02.024.

[74]刘宁. 综合业务系统如何同时发挥C/S和B/S结构优势[J]. 金融电子化,2008,(08):57-59.

[75]魏亚楠,宋义秋.oracle数据库应用系统的性能优化[J].制造业自动化,2010,32(8):204-206,222.

[76] THOMAS KYTE.苏金国,王小振,等译.ORACLE9I&10G编程艺术:深入数据库体系结构[M].人民邮电出版社，2006:265-269.

[77]李云云.浅析B/S和C/S体系结构[J].科学之友,2011,(2):6-8.

[78]蒋书法.最短置信区间的求解[J].上海电力学院学报,2014,30(2):188-192.

[79]李信真，车刚明，欧阳洁.计算方法[M].西安工业大学出版社，2000：31-156.

[80]王政.Web前端开发技术以及优化研究[J].电脑知识与技术,2013,(22):5037-5038.

攻读学位期间研究成果

1.发表论文

1．Yushan WAN, Tao ZOU, Min SHAO, Dinglong LI. Study on Eco-restoration of Polluted Rivers[J].Agricultural science & technology. 2016,17 (10):2386-2409

2．Yushan WAN, Dawei LI, Min SHAO, Tao ZOU, Yanqiu CHEN. Status and Health Risk Assessment of Heavy Metal Contamination in Vegetables[J].Agricultural science & technology. 2016,17 (9):2161-2166

2.发明专利

1.邹涛；万玉山；李大伟；沈梦；温馨；陈艳秋. 一种鉴别污水监测数据真伪的方法(201611114422.7)

2.邹涛；万玉山；李大伟；沈梦；温馨；陈艳秋. 一种开发区混合废水处理系统(201710028502.9)

3.万玉山；邹涛；李大伟；陈艳秋.一种含铊废水处理装置(201610806793.5)

4.万玉山；邹涛；李大伟；陈艳秋；黄利. 一种垃圾渗滤液RO浓水处理装置(201610410133.5)

5.万玉山；邹涛；陈艳秋；李大伟；黄利. 一种基于Android平台的单元刚度矩阵计算系统 (201610194565.7)

6.万玉山；李大伟；邹涛；陈艳秋；黄利. 一种利用沼气供暖草莓大棚的装置（李大伟写）

7.万玉山；邹涛；李大伟；陈艳秋；黄利.一种畜禽养殖废水处理渠(201610099250.4)

8.万玉山；邹涛；李大伟；陈艳秋；黄利. 含重金属离子电镀废水处理装置(201610029893.1)

9.万玉山；邹涛；李大伟；陈艳秋；黄利.一种铁盐脱氮除磷装置(201610016350.6)

10.万玉山；邹涛；李大伟.农村污水处理装置(201510916780.9)

致 谢

时光荏苒，一晃研究生三年生涯已快过去。这三年来，我不仅在科研的过程中学到了很多知识，也收获了与导师、师姐、师妹之间的友谊。

能够顺利完成本论文，首先要感谢校内导师万玉山。在校期间。不仅在科研上提供指导，生活上也给予了诸多关心。在校外期间，也时常短信问候，以便及时给我帮助。万玉山导师对工作的严谨态度以及对学生的和蔼可亲，将成为我今后学习的方向。同时感谢李大伟、陈艳秋、温馨、沈梦等办公室的朋友们在我离校期间对我的支持与帮助。

感谢家人、舍友以及其他朋友的一路互相陪伴，给予我源源不断的动力！尤其是父母对我生活经济上的辛苦付出。

最后，向所有帮助过我的老师、同学、亲人、朋友表示最真诚的感谢，祝大家健康、平安、幸福！