

# Research on Data Error Analysis Algorithm of Agricultural Sensor Network

Jiale Li<sup>1\*</sup>, Bin Wen<sup>1</sup>, Nan He<sup>2</sup>

<sup>1</sup>Department of Communication Engineering, Chengdu University of Information Technology, Chengdu Sichuan

<sup>2</sup>Wenjiang Meteorological Bureau, Chengdu Sichuan

Email: 781767038@qq.com

Received: Dec. 8<sup>th</sup>, 2018; accepted: Dec. 21<sup>st</sup>, 2018; published: Dec. 28<sup>th</sup>, 2018

---

## Abstract

The accuracy of the agrometeorological observation elements has an important impact on the observation effect, especially in the complex environment of the field, especially the factors affecting the temperature change. At present, a large number of devices for field temperature observation are mostly wireless sensor networks with scattered nodes. The sensors used are low-cost, low-precision industrial sensors. Once they are in error, they are difficult to check and have a great impact on subsequent analysis. In order to mitigate the difficulty in searching the dysfunctional wireless sensors and targeting the problematic data, we analyze error from the perspective of agricultural sensor network data, and subsequently check out the errors in the wild wireless sensor network. By fully extracting and analyzing the standard data of the automatic station, it is found that the maximum temperature difference between different regions on the same day is approximately normal distribution on a one-year cycle, and the temperature data characteristics collected by the analog automatic station are collected by industrial sensors. The data were extracted and smoothed, and it is found that this property is also applicable to industrial sensors. Therefore we design a data error analysis algorithm, applied to wild industrial sensor calibration. By validating temperature, which is one of the essential factors in agricultural atmosphere observation, the result demonstrates the efficiency of the algorithm. Instead of directly manipulating on the current wireless sensor network, the algorithm conducts dimensionless processing, followed by extraction, smoothing, and comparison to inspect error, to save the cost of error inspection.

## Keywords

Agricultural, Industrial Sensors, Regional Characteristics, Error Analysis

---

# 农业传感网络数据差错分析算法研究

李嘉乐<sup>1\*</sup>, 文斌<sup>1</sup>, 贺南<sup>2</sup>

<sup>1</sup>成都信息工程大学, 通信工程学院, 四川 成都

<sup>2</sup>温江气象局, 四川 成都

\*第一作者。

Email: 781767038@qq.com

收稿日期: 2018年12月8日; 录用日期: 2018年12月21日; 发布日期: 2018年12月28日

## 摘要

农业气象观测要素精度的高低对观测效果有着重要影响,尤其是在野外复杂的环境下,影响温度变化的要素尤其多。目前大量用于野外温度观测的设备多为节点分散的无线传感网络,所用传感器为成本低、精度差的工业传感器,一旦出错,难以排查,且对后续分析工作影响巨大。为了降低出错无线传感器的排查难度,及时锁定出错数据,从农业传感网络数据的角度入手,对其进行差错分析,从而对野外的无线传感网络进行出错排查。通过对自动站的标准数据进行充分提取分析,发现以一年为周期,同一天不同地区之间的最高温度差成近似正态分布,类比自动站所采集的温度数据特性,通过对工业传感器采集的数据进行提取、平滑等分析,发现了该性质也适用于工业传感器,由此设计了一种数据差错分析算法,应用于野外的工业传感器的检错,通过温度这一农业气象观测的关键要素之一进行验证,验证结果证明了该算法的有效性。该检错方法不直接对现有无线传感网络进行操作,而是对其进行数据无量纲化处理,再对其进行提取、平滑和对比等处理来检错,节省了传感器检错成本,节省人力物力。

## 关键词

农业, 工业传感器, 地域特性, 差错分析

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 前言

我国是农业大国,农业气象观测是我国农业气象服务的基础,也是地面观测的重要环节[1]。农业气象观测数据的质量直接影响到作物的生命活动以及野外农业气候预测的准确性,同时也影响着作物的品质和产量以及病虫害的预防工作。随着我国社会经济和科学技术的不断发展,“高产、优质、高效”已经成为我国农业新的发展方向,这也就对气象为农服务提出了更高的要求,高精度、高密度、多要素、连续稳定的农业气象观测信息是现代农业快速发展基本保证[2]。

随着传感器和无线通信技术的不断进步,现在的无线传感网络也大多由一些低功耗、体积小、功能多的传感器组成。但这种廉价微型的传感器,在野外复杂的环境下,容易造成磨损和测量不准,故传感器的检错至关重要[3]。本文选取温度要素进行观测,提出一种通过对农业气象观测数据进行分析检测对应传感器的方法。

## 2. 我国农业气象观测现状

自20世纪50年代开始,我国开始建设农业气象观测网,至2009年我国共有农业气象观测站631个,其中一级农业气象观测站398个,二级农业气象观测站233个,另外还有68个农业气象试验站[4]。1979年编制和出版了《农业气象观测规范》并在1993年对其进行了修订[5]。但我国的农业气象观测体系是上世纪建立的,观测仪器设备老旧落后,多数农业气象观测任务需要人工手动操作,不仅费时费力,还

消耗了大量的人力物力,造成观测任务的成本较高。为与现代农业发展相适应,自2009年开始,我国开始大量组建自动化农业气象观测网,大力推动自动化农业气象观测的发展[6]。随着农业技术的快速发展,气候的不断变化以及农业种植生产模式的不断改进,以往的农业气象观测模式无论是在站点布局、观测重点、观测频率以及精度上都已经无法满足现代农业发展日益提高的需求。因此,农业气象观测已经成为我国重点关注的问题之一。

### 3. 无线传感网络与观测要素的地域特征

随着自动气象站的推广和普及,农业气象观测相关技术已近成熟,但对于数据获取的规模和数据分辨率方面能力还是比较有限的。这是由于自动气象站用于获取数据的设备都是一些价格昂贵高度精密仪器,具有远距离传输功能的无线通讯设备,而这些设备不可能高密度的部署,所以大多数农业自动气象站点的布局是非常广泛的,那么在监测作物生育期时想要做到精准化控制是远远不够的[7]。这时无线传感器网络的优势就得以体现了,大量传感器节点共同协助,收集周围的环境信息,可以包括温度,光照强度,降水量,定位等类似的信息[8]。本文通过对自动站采集的数据进行分析,并将找到的规律迁移到工业传感器的检错上,选取温度数据进行验证,达到不直接对工业传感器进行操作而是通过数据分析进行检错。

#### 3.1. 观测要素验证的确定

通过对野外环境农业气象要素的观测,不仅可以有效的避免突发自然环境变化对作物的影响,精准全面的数据采集还可以为农业精准化控制提供基础,使作物常年处于最佳的生长条件。作物品种不同对环境要求也不同,同一作物处于不同生长阶段要求的适宜环境也存在差异。本文选取对作物生长最为关键要素之一进行验证——温度。

温度:温度是对作物生长影响最多的要素之一。它主要通过光合作用、呼吸作用、蒸腾作用来影响作物生长。野外内气温分布并不是均匀的。作物的光合作用需要有一定温度范围及三基点,即作物生长过程中的最适温度、最高温度和最低温度[9]。当野外温度超出范围时就会影响作物的生长。三基点会随着作物生长的不同阶段而变化,如:生长前期,作物所需温度较高,否则会出现茎细叶薄的现象;生长后期,为保证果实率则需要较低的温度[9]。

#### 3.2. 温度的地域特征挖掘

由于观测手段的逐步改善,在同一要素的数据记录上存在较大差异[10],故选取观测手段一致的1971年到1980年间的A、B两地每天的最高温度数据作为分析的基本要素,以年为周期对不同地区同一天的最高温度差,以0.1为步进做统计计数,再对结果进行小波变换,做无量纲化处理,再进行提取、平滑、分析等处理来挖掘温度所具有的地域特征。

日高温差:

$$D_{A_n B_n} = T_{A_n} - T_{B_n} \quad (1)$$

其中 $T_{A_n}$ 和 $T_{B_n}$ 分别代表A、B两地的日最高温度, $D_{A_n B_n}$ 代表A与B地之间的日最高温度差,其中平年的 $1 < n < 365$ ,闰年的 $1 < n < 366$ 。

统计计数:再-2.5到2之间,以0.1为步进,统计频数。

小波变换:是一个时间和频域的局域变换的时频分析方法,能有效地从信号中提取信息,通过伸缩和平移等运算功能对函数或信号进行多尺度细化分析,主要特点是通过变换能够充分突出问题某些方面的特征[11],这里主要采用基于小波变换的多尺度空间中模极大值特征的提取方法,利用小波变换的信号

局域化分析能力, 求解小波变换的模极大值特征来检测信号的局部奇异值, 即突显出日最高温度差频数的众数区间。小波是一个均值为零, 且既在时域又在频域存在的函数。小波变换有很多种, 这里用到的是离散小波变换(DWT), 其小波基函数为:

$$\Psi_{a,\tau}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-\tau}{a}\right) \quad (2)$$

其中  $a$ 、 $\tau$  限定在一些离散的点上取值。

对尺度进行幂级数离散化, 令:  $a = a_0^j, a_0 > 0, j \in Z$

则, 此时小波基函数为:

$$\Psi_{a,\tau}(t) = a_0^{-j} \Psi[a_0^{-j}(t-\tau)], \quad j = 0, 1, 2, \dots \quad (3)$$

位移离散化:

$$\tau = ka_0^j \tau_0 \quad (4)$$

故离散小波变换可以定义为:

$$\text{WT}_f(a_0^j, k\tau_0) = \int f(t) \Psi_{a_0^j, k\tau_0}^*(t) dt, \quad j = 0, 1, 2, \dots, k \in Z \quad (5)$$

无量纲化: 取  $a_0 = 2$ ,  $\tau_0 = 1$ , 则小波基为:

$$\Psi_{j,k}(t) = 2^{-\frac{j}{2}} \Psi(2^{-j}t - k) \quad (6)$$

对应的小波离散变换为:

$$\text{WT}_f(j, k) = \int f(t) \Psi_{j,k}^*(t) dt \quad (7)$$

提取: 对与 10 年的日最高温度差频数进行分析, 提取固定存在的地域特性, 并找出频数最大的最高气温差区间作为两地最高气温差的回归区间。

平滑: 为了更好的查看数据走向、查看数据趋势[12], 我们采用曲线五点滑动平均方法对日最高温度频数进行平滑处理, 来提高数据的光滑度, 其公式如下:

非端点:

$$x(n) = \frac{x(n-2) + x(n-1) + x(n) + x(n+1) + x(n+2)}{5} \quad (8)$$

端点:

$$x(1) = \frac{3x(1) + 2x(2) + x(3) - x(4)}{5} \quad (9)$$

$$x(2) = \frac{4x(1) + 3x(2) + 2x(3) + x(4)}{10} \quad (10)$$

$$x(m-1) = \frac{x(m-3) + 2x(m-2) + 3x(m-1) + 4x(m)}{10} \quad (11)$$

$$x(m) = \frac{-x(m-3) + x(m-2) + 2x(m-1) + 3x(m)}{5} \quad (12)$$

其中,  $x(n)$  表示最高温度频数序列,  $3 \leq n \leq m-2$ ,  $m$  为数据总个数。

## 4. 仿真结果及特性分析

### 4.1. 模型建立与仿真

使用 Matlab 结合 Microsoft Excel 建立模型并挖掘。自动站分析模型建立框架如图 1 所示，对自动站资料进行最高温度差和频数分布的计算，通过离散小波变换进行无量纲化处理，再进行平滑处理以便分析趋势与提取特征。同理工业传感器验证模型框架如图 2 所示。采用相同分析方法对工业传感器资料进行处理，类比由自动站资料得出的结论，进行特征分析。如图 3 所示为检错流程图，首先将两地的数据载入模型，对数据进行日最高温度差及其频数的计算，在通过离散小波变换对其进行无量纲化处理，选择合适的平滑系数进行特性分析，若满足地域特性的要求则两组数据是正确的，若无明显地域特性则说明采集的数据有误。原始数据为 Excel 格式，A、B 两地情况分别如表 1、表 2 所示，使用 excel 计算  $D_{A_n, B_n}$ ，得到下表 3。

**Table 1.** A point daily maximum temperature table

**表 1.** A 地日最高温度表

地区	年份	日期	最高温度
A	1971	1971-01-01	6.8
A	1971	1971-01-02	9.3
A	1971	1971-01-03	6.3
.....	.....	.....	.....
A	1980	1980-12-29	8.9
A	1980	1980-12-30	9.8
A	1980	1980-12-31	10.4

**Table 2.** B point daily maximum temperature table

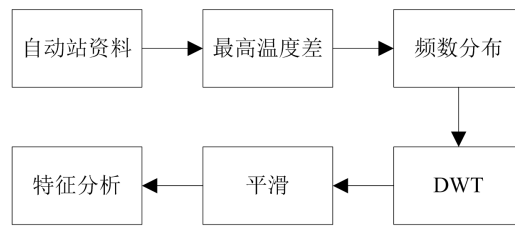
**表 2.** B 地日最该温度表

地区	年份	日期	最高温度
B	1971	1971-01-01	7.9
B	1971	1971-01-02	9
B	1971	1971-01-03	6.4
.....	.....	.....	.....
B	1980	1980-12-29	7.4
B	1980	1980-12-30	10.2
B	1980	1980-12-31	11.7

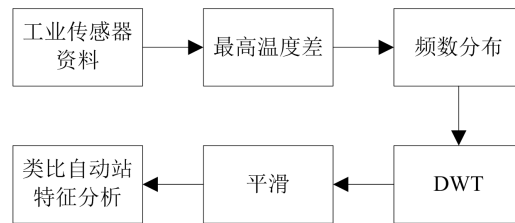
**Table 3.** Maximum temperature difference between A & B

**表 3.** A-B 地最高温度差

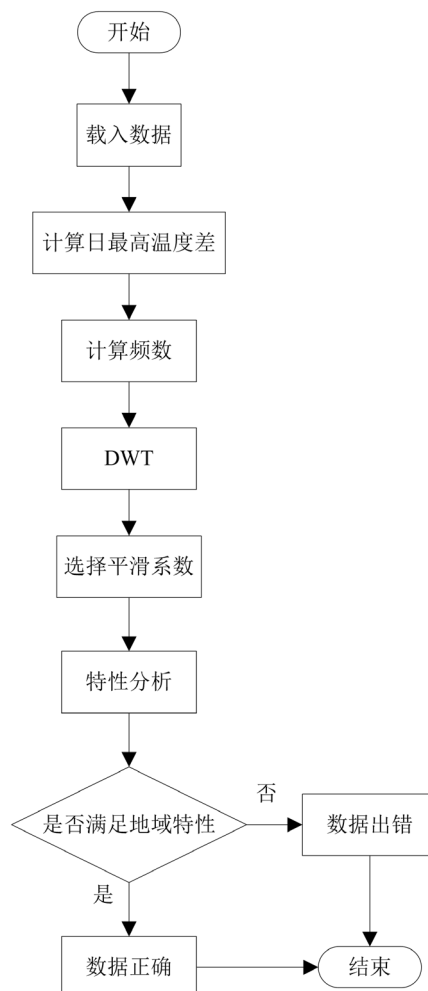
年份	日期	最高温度差
1971	1971-01-01	-1.1
1971	1971-01-02	0.3
1971	1971-01-03	-0.1
.....	.....	.....
1980	1980-12-29	1.5
1980	1980-12-30	-0.4
1980	1980-12-31	-1.3



**Figure 1.** Automatic station analysis model framework  
**图 1.** 自动站分析模型框架



**Figure 2.** Industrial sensor validation model framework  
**图 2.** 工业传感器验证模型框架



**Figure 3.** Error detection flow chart  
**图 3.** 检错流程图

### 4.2. 特性分析

如图 4 所示为，自动气象站测得的最高温度差的频数分布图，从图中可以看出每一年的最高温度差频数分布都呈近似正态分布，并大致保持一致的趋势。对以下几个参数进行分析：

趋势：每条曲线都呈现近似正态分布的趋势，且均存在一个最高值，代表一年中此点对应最高温度差的频数最大，也就是说一年之中有一部分时间的日最高温度差均为一固定值，即为最高温度差的可回归的地域特性提供了依据。

半高宽：又称半峰宽(Full Width at Half Maximum)，这里指的是以频数曲线高度最大处为中心，当高度为一半时曲线的全宽，即峰值高度一半时的最高温度差跨度。对每一年的频数曲线求半高宽，再取交集，可以得出 $[-1^{\circ}\text{C}\sim 0^{\circ}\text{C}]$ 囊括了一年大部分时间的最高温度差，可作为最高温度差地域特性的回归区间。

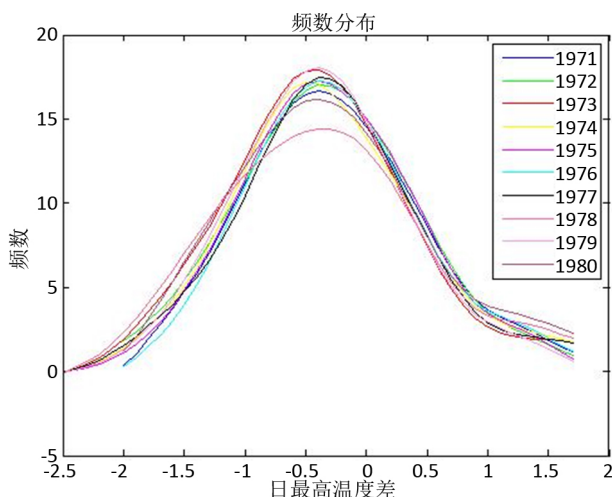


Figure 4. Frequency distribution curve of the highest temperature difference between A and B

图 4. A、B 两地最高温度差的频数分布曲线

### 4.3. 检错可行性验证

选取两个具有差异的环境点 C、D，每个点各放置两个工业传感器 T1、T1’、T2、T2’，假设每个点放置的工业传感器完全相同，且其中 T1、T1’保证读数准确，T2、T2’为不准传感器，其测试环境如图 5 所示。读取两个环境点放置的工业传感器采集的日最高温度数值，计算两个环境点之间的最高温度差，并进行分析，确定其地域特性是否存在回归区间，如图 6 所示为检错判定流程图。

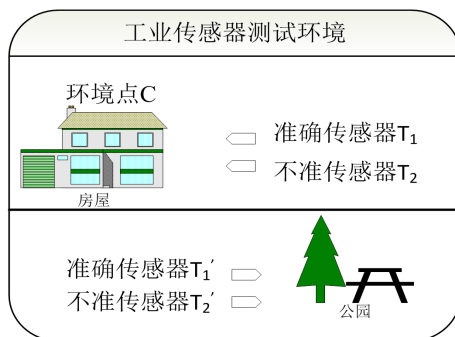


Figure 5. Industrial sensor test environment diagram

图 5. 工业传感器测试环境示意图

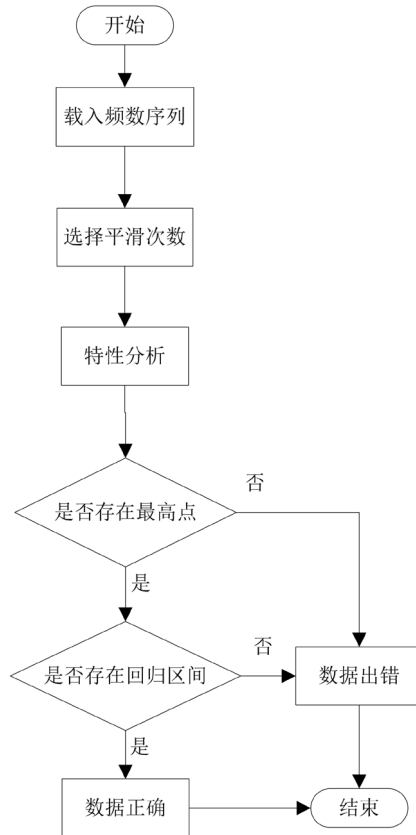


Figure 6. Error detection decision flow chart  
图 6. 检错判定流程图

经过 Matlab 仿真测试，得到如下结果图 7。从图中可以看出蓝色线为 T1 组传感器，即准确传感器组测得的两个环境点最高温度参数差的频数分布曲线，可以看出两环境点间存在一个频数最大的温度差，可以取 $[0^{\circ}\text{C}\sim 1^{\circ}\text{C}]$ 作为回归区间。图中红色线为 T2 组传感器，即不准传感器组测得的两个环境点最高温度参数差的频数分布曲线，由图可看出曲线整体起伏不大并未有明显最高点，无法找到回归区间，故怀疑这组数据有误。

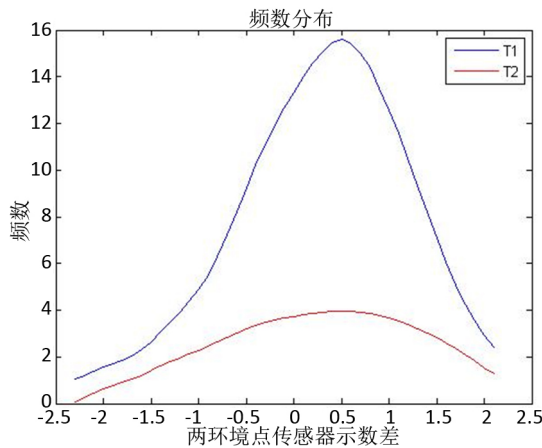


Figure 7. Error detection test result chart  
图 7. 检错测试结果图



## 5. 总结与展望

本文选取观测手段一致的计量级传感器测得的日最高气温数据入手, 挖掘温度存在的地域性特征, 并应用该特征设计了一种数据差错分析算法用于提高野外工业传感器检错效率, 节约检测成本。该算法提供了一种全新的农业传感网络检错视角, 在一定程度上推进了现代农业的数据化发展[13], 但依然存在很多可以深入研究的地方, 未来的研究方向可以从以下几个方面入手:

1) 数据的利用: 由于数据的差异型, 本文只用了人工观测的数据, 未来可以从自动站与人工检测的数据同步方向入手, 将自动站的数据进行相应的平滑、筛选、补偿处理, 将其与人工观测的数据结合起来分析。

2) 工业传感器的矫正: 本文仅是提供了通过数据检测传感器好坏的方法, 但对于出错数据的矫正方法仍有待研究, 如可以采用游走拓扑等方式。

3) 工业传感器的精度提升: 在本文提出的检测方法基础上, 可以利用最高温度差的回归特性, 将工业传感器向计量传感器的精度进行修正, 使成本低的工业传感器经过数据修正达到计量级精度。

## 参考文献

- [1] 谢秀英, 谢秀琴, 宁金鹤, 候向阳, 阿依奴. 农业气象观测技术要点分析[J]. 现代农业科技, 2017(12): 232+235.
- [2] 陈诚, 李必军, 张永博. 基于无线传感器网络的农业信息采集系统设计[J]. 安徽农业科学, 2016, 44(10): 242-245.
- [3] Kanawaday, A., et al. (2017) Machine Learning for Predictive Maintenance of Industrial Machines using IoT Sensor Data. 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 24-26 November 2017, 4. <https://doi.org/10.1109/ICSESS.2017.8342870>
- [4] 余卫东, 杨光仙, 张志红. 我国农业气象自动化观测现状与展望[J]. 气象与环境科学, 2013, 36(2): 66-71.
- [5] 蒋运志, 陈宗行, 马新建, 蒋祖林. 农业气象观测中需注意的问题[J]. 现代农业科技, 2010(15): 338.
- [6] Zhang. C.J. (2014) A Sensor Grouping Method for Industrial Sensor Health Management. *Proceedings of 2014 International Conference on Future Mechanical Engineering and Materials Engineering (MEME 2014)*, Information Engineering Research Institute, USA, 6.
- [7] 赵劲松, 李元, 邱彤. 一种基于小波变换与神经网络的传感器故障诊断方法[J]. 清华大学学报(自然科学版), 2013, 53(2): 205-209+221.
- [8] 丁硕, 常晓恒, 巫庆辉, 杨友林, 胡庆功. 基于 Elman 神经网络的传感器故障诊断研究[J]. 国外电子测量技术, 2014, 33(4): 72-75.
- [9] 姚建成. 气象对农作物生长的影响[J]. 农业与技术, 2017, 37(20): 244.
- [10] 王健. 中美股市联动性——基于极大重叠离散小波变换的研究[J]. 世界经济文汇, 2014(2): 72-89.
- [11] 何乃强, 惠晓滨, 周漩. 基于正态区间估计的改进型无量纲化方法[J]. 计算机工程与应用, 2012, 48(5): 236-238.
- [12] 杨春庄. 改进的滑动窗平滑声能衰减曲线法估值混响时间[C]//中国声学学会. 泛在信息社会中的声学——中国声学学会 2010 年全国会员代表大会暨学术会议论文集. 北京: 中国声学学会, 2010: 2.
- [13] 陈成国, 石慧兰, 王桂兰, 邵志勇, 周圣军. 自动与人工气象观测要素的差异及原因分析[J]. 安徽农业科学, 2010, 38(35): 20222-20224+20235.

**知网检索的两种方式：**

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>  
下拉列表框选择：[ISSN]，输入期刊 ISSN：2161-8801，即可查询
2. 打开知网首页 <http://cnki.net/>  
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：[csa@hanspub.org](mailto:csa@hanspub.org)