西安科技大学

**毕业设计(论文)文献翻译**

题 目 超声波传感器在水流中测量水位数据的异 常值检测和平滑处理

院、系(部) 计算机科学与技术学院

专业及班级 计科1702班

姓 名 赵 辉

指导老 师 秋兴国

日 期

**摘 要**

由超声波传感器在河流尺度通道中获取的水位数据集显示出相对大量的异常值，这些异常值超出超声波传感器和水面之间的测量范围，并且由于随机误差（如水波）导致数据分散约2 cm。因此，本研究针对超声波测得之水位资料，提出一套资料处理演算法，以去除异常值并平滑处理。离群点剔除过程包括一个初始截止过程，将离群点剔除出测量范围，以及一个基于稳健估值器的中位数绝对偏差（MAD）的修正Z分数的离群点检测过程。此外，采用指数加权滑动平均（EWMA）方法对处理后的数据进行平滑处理。对用户主观设置的因素进行敏感性分析，包括MAD异常值检测阶段的窗口大小、修正的Z分数异常值去除阶段的拒绝准则以及EWMA平滑阶段的平滑常数，基于四种不同的水位数据集采集超声波传感器在河川规模实验。

**关键词：**数据平滑；指数加权移动平均；中值绝对偏差；修正Z-分数；离群点检测；超声波传感器；水位监测

# 1 介绍

可靠的流量测量和数据对于水管理、防洪、河流养护、河流修复、河流恢复等是必不可少的世界气象组织[5]和Sauer和Turnipseed[4]建议使用3 mm或有效水位的0.2%（孔口、进水口或传感器暴露于水体的其他点上方的水面高度）中的较大值作为测流站所需的水位测量精度。然而，利用坡度等级计算水库蓄水量变化或流量需要更高的精度，而超声波传感器可用于要求如此高精度的水位测量。与传统的水位测量方法（如消力井和浮子系统）相比，使用超声波传感器的测量方法降低了成本，并且对安装位置几乎没有限制[9]。最近，Mousa等人[10]、Rahmtalla等人[11]、Satria等人[12]、Sunkpho和Ootamakorn等人[13]将超声波传感器用作洪水预报和警报系统的实时水位监测方法。此外，Bae等人[14]在水力实验中应用了超声波传感器来监测水位。

虚假水位数据引入的误差会使流量测量失效，此类误差通常是由人为错误、自动水头记录仪故障或阻塞控制段的树枝或其他碎屑对正常流量造成的障碍造成的[15]。特别是，当用超声波传感器测量水位时，误差可能是由于与传感器安装位置有关的各种环境因素，以及由不规则的水面流引起的不规则的超声波信号造成的。这些误差产生的数据称为异常值，与随机误差不同，这些异常值没有统计意义。因此，必须采用适当的统计方法来检测异常值并将其排除在分析之外[1,15]。异常值分布的特征可能因测量方法、设备和其他条件而异，因此必须根据数据特征使用不同的处理技术[16]。采用超声波传感器测量水位时，必须了解水位数据的异常值分布特征，以便采用适当的统计处理技术去除数据中的异常值。

超声波传感器收集的水位数据显示出大约20–50 mm的离散度[17]，而导致这种数据离散度的主要误差源是水波[3]。水波引起的误差可分为随机误差[18]。通过在相同条件下重复测量，可以将随机误差的影响降至最低[15]。然而，在河流中，

水位随时间变化，重复测量在相对较短的时间内进行。因此，必须用统计方法在相对较短的时间范围内估计观测数据的代表值，在这个时间范围内，水位可以被认为是恒定的。目前还没有一种算法能够自动处理超声波传感器采集的大量水位数据，并针对不同的数据应用目的进行离群点检测和数据平滑。

因此，本研究的目的是发展统计处理的一般化演算法，以去除实际水流条件下超声波传感器所取得的水位资料中的异常值和随机误差，并平滑水位资料，以减少随机误差所造成的水波频散。根据超声波传感器在水流中测量水位的数据特征，本文提出了一种适用于该算法的统计方法：一种初始截断过程，将异常值剔除在测量范围之外；一种基于稳健模型中位数绝对偏差（MAD）的修正Z分数的异常值检测过程具有不受大偏差离群值影响的精确估计功能的估计器，以及通过应用不同的平滑常数来平滑处理数据的指数加权移动平均（EWMA）方法。特别地，本研究提出一个在使用者设定的最小和最大范围内确定最佳视窗大小的具体过程。此外，该算法还专门处理了一个问题，即当水位在优化窗口大小指定的数据集中保持不变时，MAD值变为零。在本研究中，我们利用从设计流量可刻意控制的流标实验水道收集的长期水位资料，进行敏感性分析，以确定算法中与离群点去除和平滑处理相关的因素。这些因素包括离群点检测的移动中值和中值绝对偏差（MAD）的窗口大小，离群点去除的修正Z分数的拒绝标准，以及平滑过程的指数加权移动平均（EWMA）的平滑常数因子，在超声波传感器监测水位的数据处理算法中，需要用户定义。

# 2 材料和方法

## 2.1 数据收集

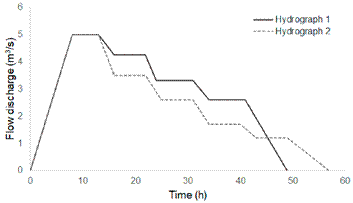
利用超声波传感器在韩国土木工程与建筑技术研究所河流实验中心（KICT-REC）的大型室外水槽中分两轮收集水位数据（图1）。河道长594m，宽11m，下游断面河床坡度约为3/1000，为下游水位测点。通过调节流量供给泵设备，可以控制流道内产生的流量。如图2所示，设计了两种不同形式流量过程线的非定常流实验。虽然两条水文曲线的最大流量设计为5m3/s，但水流事件的总持续时间不同。两个超声波传感器分别安装在渠道测量点（A）的上游堰和测量点（B）的下游桥上，以监测两个不同水位线相继生成时的水位（图1）。因此，本研究共收集了四组超声波传感器的数据集：过程线1的两个测量点（a）和（B）的数据集，以及过程线2的两个测量点（a）和（B）的数据集。通过使用先前建立的水位和流量之间的关系，将测量点（A）上游堰处测量的水位转换为流量，可以跟踪从上游到渠道流量随时间的变化。堰上的流速是堰上水头的函数。因此，矩形堰的流量测量是基于伯努利方程原理的。在实验前对矩形宽堰的流量系数进行了测定和标定，建立了矩形宽堰的流量方程。在测量点（B）下游桥梁处进行水位测量，观察试验航道下游段的水位变化。



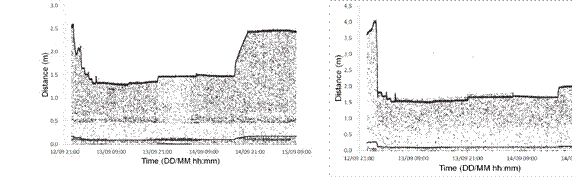
**图1 韩国土木工程与建筑技术研究所河工实验中心（KICT-REC）河工实验航道。**

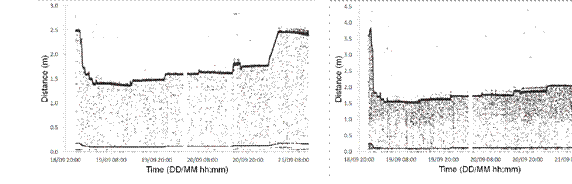
超声波传感器从发射器产生连续的信号，当信号从目标物体表面反射回来并返回到接收器时，它们的往返时间与声速一起转换成距离。本实验中用于水位测量的超声波传感器为HC-SR04型，与传感器单元的距离可测量

0.2–4.0 m，测量分辨率为3 mm，最小采样间隔为60 ms。在本实验中，以2.5 s的间隔对原始数据进行采样。根据用户的需要，这些传感器可以与数据存储设备、通信模块、电源、电平校准器等相结合[19]，在本研究中，超声波传感器与Arduino板和WiFi模块相结合，在室外水槽实时采集数据。该系统获得的原始数据显示了传感器到水面的距离值（图3）。为目标水文线的每个测量位置收集的数据集数量在77040和96954之间（表1）。图3显示了从超声波传感器获得的目标水文曲线每个测量点的所有数据，包括异常值。



**图2 为本研究设计的目标过程线在河川尺度实验水道中。**

  **（a） （二）**



**（c） （四）**

**图3 从超声波传感器到水面大约60小时距离的测量原始数据。（a）水文1情况-下游桥梁；（b）水文1情况-上游堰；（c）水文2情况-下游桥梁；（d）水文2情况-上游堰。**

**表1 每个数据集的离群点检测结果和中位绝对偏差（MAD）过程的移动窗口大小。**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **流动条件** | **传感器位置** | **采集时间（h）** | **最优的**  **移动**  **窗口**  **尺寸wo**() | **原材料数量**  **数据** | **数量**  **离群值**  **用Cuto删除**ff公司 | **数量**  **离群值**  **用MAD删除** | **总离群值** | **%的**  **离群值** | **数量**  **之后的数据**  **离群值**  **拆卸** |
| 过程线1 | 下游桥梁 | 58 | 24 | 86,942 | 22,971 | 3653 | 26,624 | 30.6 | 60,318 |
| 过程线1 | 上游堰 | 52 | 13 | 77,040 | 10,674 | 3169 | 13,843 | 18.0 | 63,197 |
| 过程线2 | 下游桥梁 | 61 | 11 | 87,916 | 9087 | 3601 | 12,688 | 14.4 | 75,228 |
| 过程线2 | 上游堰 | 67 | 19 | 96,954 | 12,650 | 4748 | 17,398 | 17.9 | 79,556 |

## 2.2 数据处理方法

2.2.1. 离群点检测的稳健统计方法

在超声波传感器测量的水位数据中，当超声波在到达水面之前被反射时，可能会出现异常值，并且这种测量值的偏差可能相当大（图3）。因此，为了检测超声波传感器测得的水位数据中的异常值，必须使用对偏差较大的异常值不敏感的鲁棒估计器。根据Huber[20]、Leys et al[21]和Rousseeuw and Croux[22]，样本中位数可以用作位置的稳健估计器，因为中位数对如此大偏差的影响非常不敏感，因此可以最小化由于超声波传感器测量的水位数据分布中的异常值造成的不对称影响。另外，MAD可以作为一个稳健的尺度估计，用于估计以样本中位数为中心的数据离散度。与中位数一样，MAD对异常值偏差非常不敏感[20–23]。

当收集的数据集被给定为X={x1，x2，xn}时，MAD的定义如下[24–26]：

                                             MAD(X)=Med{|X-Med(X)|} (1)

其中Med（）是一个用于查找数据集X的中值的函数。因此，MAD（）显示数据集X和Med（）的单个观测值之间绝对偏差的中值。为了比较方程式（1）中计算的MAD值与标准偏差，方程式（2）用于将MAD（）转换为标准化MAD（MADN）[24]：

MAD(X)= （2）

其中，常数值0.6745是正态分布的75%百分位数，对应于标准正态分布的MAD，标准偏差为1[24,26–28]。MADN（X）对应于由样本中值估计的数据中心位置的正态分布的标准偏差。

为了检测离群值，必须使用无量纲变量，如等式（3），将MADN（x）与以样本中位数估计的位置为中心的个体观察席的偏差进行比较。式（3）中的Mi是修正的Z分数。将其与离群值剔除准则β进行比较，以确定数据点是否为离群值[26]。

|xi-MED(X)

                      Mi =  <β (3)

MADN(X)

Iglewicz和Hoaglin[26]提出了拒绝标准，β=3.5，基于对10、20和40个样本大小使用伪正态观测的模拟结果。此外，Leys等人[21]报告说，根据研究者的观点，β的测定方式可能有所不同。

2.2.2. 数据平滑的指数加权移动平均法

即使测量水位数据中出现的异常值被完全去除，由于流动水中的水波，也会出现大约2 cm的弥散。如果使用超声波传感器在相同条件下短时间间隔重复测量恒定水位，则由于随机误差（不包括异常值）导致的观测值形成正态分布。因此，移动平均可以用作平滑方法。这是信号处理中常用的低通滤波器的最简单形式，用于平滑时间序列数据中的短期变化。但是，在监测期间，实际水位可能会随时间变化。因此，在平滑过程中，最近观察到的数据必须比旧数据反映得更多，EWMA可用于此类情况，以便对旧数据应用较低的权重[29]。当数据集被列为时间t的序列时，EWMA指数地减少旧数据的权重因子，如等式（4）所示[29]：

                                    (4)

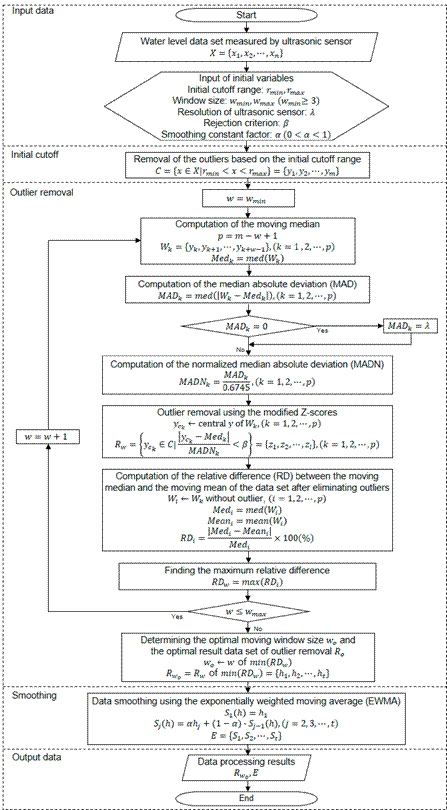
其中α表示作为平滑常数因子的加权递减程度，范围为0<α≤1。如等式（4）所示，通过以较小的α增加过去数据的权重，可以获得高水平的平滑，这使得数据离散度更小。然而，如果α变得太小，数据在短时间内的变化几乎被忽略。因此，有必要根据研究者的主观标准和数据使用目的来改变α。

因为EWMA需要过去的观测值来进行计算，所以初始观测值x1很难确定。因此，一些方法使用初始观测值，如S1（）=，其他方法使用多个初始值的平均值[29]。在这项研究中，初步观察已应用于EWMA。

2.2.3. 水流中超声波传感器测量水位的广义数据处理算法

提出了一种综合数据处理算法，用于消除超声波传感器采集的水位数据中的异常值和平滑数据变化。该算法的主要处理步骤大致分为离群点去除步骤和数据平滑步骤（图4）。离群点处理步骤进一步分为离群点的初始截断和用修正的Z-分数去除离群点。用户必须在监测水位的数据处理算法中定义输入参数和条件因素，如初始截止范围（rmin和rmax）、窗口大小的最大值和最小值（wmin和wmax）、平滑常数因子（α）、超声波传感器的分辨率（λ）和拒绝准则（β）通过超声波传感器。最终输出的数据集为完成离群点剔除（Rwo）后的数据集和同时完成离群点剔除和平滑（E）后的数据集。开发的算法用python3编写，对实验采集的数据集进行数据处理测试。

建议的异常值去除阶段的第一步是初始截止。关于从水面到超声波传感器的距离，假设传感器的安装位置是恒定的，可以通过考虑从传感器到最小水位的距离来确定初始截止范围。如果在分析之前无法确定最低水位，则可以假设最低水位为零。零水位意味着没有水流，与河床水位相同。在本研究中，此区间之外的数据被确定为清晰的离群值，并且该算法的初始截止阶段将这些离群值整体去除。因此，当用户在此阶段输入初始截止范围的rmax和rmin来定义后处理输入变量时，可以移除定义为初始截止范围的该区间之外的多个异常值。因此，在下一阶段可以通过样本中值和MAD估计来减少原始数据中包含的离群点数量，提高离群点检测性能。如果用户不能定义初始截止范围的rmax和rmin，则在算法中，rmax的传感器到床的距离和rmin的零值用作默认值。



**图4 超声波传感器水位测量数据异常点剔除及平滑处理流程图。**

在初始截止阶段之后，使用修改的Z分数完成异常值去除过程。该阶段计算通过初始截止点的数据集中每个数据点的样本中位数和MAD，然后计算相应的修正Z分数，即无量纲值。当该值超过用户输入的β时，相应的数据点被确定为异常值并被删除。在该步骤中，当估计中值和MAD时，应用移动窗口来考虑水位数据随时间的变化，因此，必须确定最佳窗口大小（wo）。本文提出的算法计算了移动中值和移动均值的相对差，作为数据集中心趋势的度量，去除了每个窗口大小w的异常值。具有最小最大相对差的w被确定为wo。为了确定w的计算范围，需要输入wmin和wmax。如果样本数为2，则中位数和平均值计算结果相同。因此，最小窗口大小wmin必须大于3。该算法应用于实时监控系统时，无法实时确定最佳窗口大小。在这种情况下，应采用固定的窗口大小，或根据一定时间段内的数据处理结果更新最佳窗口大小。

如果在剔除离群值期间计算MAD时，窗口大小内数据集中超过50%的值是相同的值，则计算出的MAD值可能变为零，方程（2）和（3）中的MADN和Mi计算将不可能。Machado[30]、Chu等人[31]和Vivcharuk等人[32]提出了使用平均绝对偏差的方法

（MeanAD）作为零疯狂的替代品。但是，如果窗口大小中的所有数据都具有相同的值，则MeanAD也将为零。即使在MeanAD不为零的情况下，如果存在偏差较大的异常值，估计的规模也会很大，并且很难检测出偏差相对较小的异常值。当测量流动水的水位时，在水位保持不变的情况下，通常会出现MAD和MeanAD值为零的情况。因此，在本研究中，为了解决这个问题，使用者输入超声波传感器的最小测量单位作为输入资料。因此，当计算的MAD为零时，使用一种方法将分辨率λ定义为MAD值。由于λ与测量过程中可能出现的最小单位偏差相同，因此它对应于当MAD不为零时可能出现的MAD的最小尺寸。这种处理方法在窗口大小大于50%的数据集值具有相同的值时，将偏差较大的离群点的影响降到最低，从而能够稳健地检测出离群点。

最后，EMWA数据平滑阶段最小化了由于水位数据中的随机误差造成的数据分散，这些数据中的异常值已被完全去除。这一步在Rwo上使用等式（4），Rwo是剔除异常值后的输出数据。该步骤算法的初始阶段需要平滑常数α输入，如第2.2.2节所述，可根据情况使用不同的值。在这项研究中，利用超声波传感器测量的水位数据，根据平滑常数（α）来评估数据的平滑程度，结果见第3节。

# 3 结果

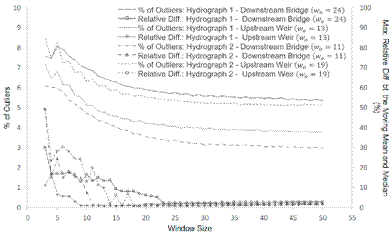
## 3.1 离群点剔除处理

在第2节所述的流标实验中，采用了去除异常值和平滑超声波传感器采集的水位数据的算法。对于下游桥梁获取的数据集，初始截止阶段所需的输入变量rmax和rmin分别设置为2.6 m和1.2 m；对于上游堰获取的数据集，输入变量rmax和rmin分别设置为4.1 m和1.3 m。表1显示了通过初始截止过程删除的异常值的数量。由于每个数据集中的离群点分布特征不同，初始截断剔除的离群点比率也不同。

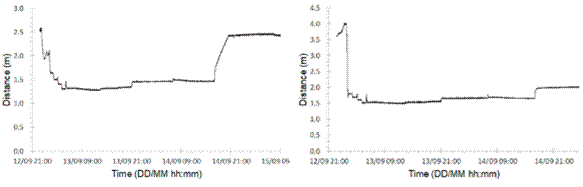
通过计算经过初始截止阶段的数据集的修正Z分数，进一步去除异常值。这里，制造商提供的实验超声波传感器的0.003 m分辨率被用作λ，以处理MAD值为零的情况。根据Iglewicz和Hoaglin（1993）提出的标准，对于修正的Z评分拒绝标准β，使用了3.5的值。为了计算去除异常值的最佳窗口大小wo，wmin和wmax分别使用3和50。窗口大小增加1，并计算移动中值和移动平均值之间的相对差值的最大值，对于总共48个窗口大小中的每一个，去除异常值。图5显示了移动中值和平均值之间的最大相对差异，以及48个窗口大小的异常检测比率。因为最佳窗口大小wo被设置为包含移动中值和平均值之间最大相对差异的最小值的窗口大小，所以四个数据集的最佳窗口大小在11到24之间。异常值检出率随窗口大小的增加呈指数下降趋势。最大相对差值随着窗口大小的增大而减小，在最小值之后略有增大。当w设置得太小时，数据样本中正常值的比率很低，移动中值和MAD的估计精度降低，而不准确的异常值检测的比率增加。最后，图6显示了通过异常值删除过程的数据集Rwo。原始数据集中被检测为异常值的数据比率至少为14.4%，最多为30.6%（表1）。此外，超过70%的异常值在初始截止阶段被去除。

除了视窗大小的设定外，本研究亦根据用户在离群值阶段必须输入的拒绝准则β的变化来分析离群值的侦测程度。虽然通常使用Iglewicz和Hoaglin[26]提出的3.5值，但需要对拒收标准β进行灵敏度分析，以确保3.5值适用于超声波传感器测量的水位数据的异常检测。在这项研究中，β以0.1为单位从1增加到5，并计算了去除的异常值的相应比率；结果如图7所示。对于用于计算移动中值和MAD的窗口大小，使用β设置为3.5时找到的最佳窗口大小。如图7所示，

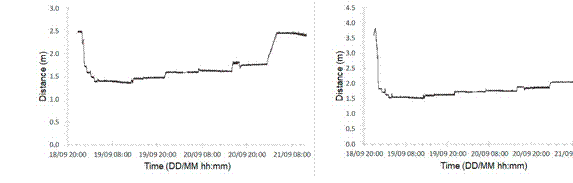
当β增加时，剔除数据的比率呈指数下降，当β超过3.0到3.5的范围时，剔除异常值的百分比下降幅度大大降低。表2显示了对四个数据集应用不同的剔除标准β时剔除的异常值的比率。



**图5 异常检测率和根据窗口大小移动平均值和中值之间的最大相对差**

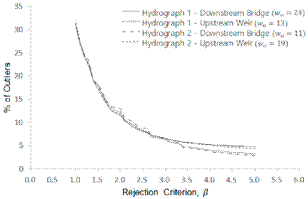


                                          （a） （b）



                                          （c） （d）

**图6 离群点去除过程后的最终数据集。（a） 水文1情况—下游桥梁；（b）水文1情况—上游堰；（c）水文2情况—下游桥梁；（d）水文2情况—上游堰。**



**图7 异常值百分比和拒绝准则之间的关系。**

**表2 根据拒绝标准检测到的异常值的百分比。**

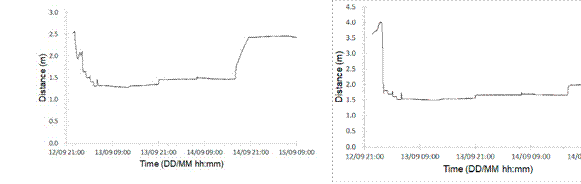


**%异常值数量**

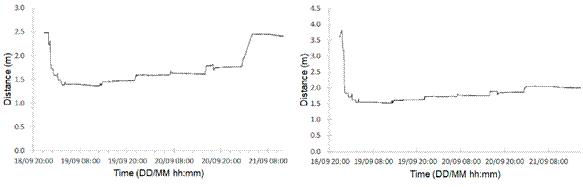
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **拒收标准**β | **过程线1的情况** | | **过程线2的情况** | |
| **下游桥梁** | **上游堰** | **下游桥梁** | **上游堰** |
| 1.0 | 30.92 | 31.32 | 31.22 | 31.83 |
| 2.0 | 11.56 | 13.04 | 12.29 | 11.99 |
| 2.5 | 8.05 | 8.77 | 8.13 | 8.31 |
| 3.0 | 6.52 | 6.37 | 6.23 | 6.60 |
| 3.5 | 5.71 | 4.57 | 4.78 | 5.63 |
| 4.0 | 5.25 | 3.95 | 4.15 | 5.12 |
| 5.0 | 4.76 | 2.87 | 3.28 | 4.42 |

## 3.2 数据平滑

超声波传感器采集的水位数据Rwo在剔除异常值后，由于随机误差，显示出约2cm的数据离散。为了缓解这种数据分散，在算法的最后阶段使用EWMA进行数据平滑。图8显示了平滑常数因子α设置为0.1时的数据平滑结果。减小了随机误差引起的水位数据离散度。



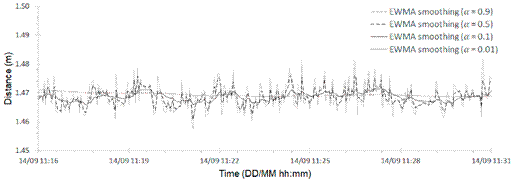
**（a） （二）**



**（c） （四）**

图8。使用EWMA平滑处理后的数据集。（a） 水文1情况—下游桥梁；（b）水文1情况—上游堰；（c）水文2情况—下游桥梁；（d）水文2情况—上游堰。

这些平滑处理结果可以根据平滑常数因子α而变化。图9比较了α为0.01、0.1、0.5和0.9时，平滑过程对整体数据的部分部分部分的影响。当α为0.01时，虽然忽略了短期变化，但前期资料的权重增加，水位资料的离散度大大减小。当α值大于等于0.5时，虽然平滑效果大大降低，但数据的离散度反映得非常紧密，这与本研究平滑过程的目标相反。因此，基于EWMA的平滑效果对α非常敏感。为了根据输入变量α定量检查平滑效果，针对图9（表3）中的数据集计算了根据α的标准偏差变化。当α为1时，等式（4）中先前数据的权重变为零，反映了没有平滑过程效应的情况。因此，当α为1时，基于0.047的标准偏差，通过标准偏差减少率计算平滑度，这表示。对水位数据的平滑效果随着α的减小而增大。



**图9 不同平滑常数因子下平滑水平的比较（过程线1-下游桥梁的情况）**

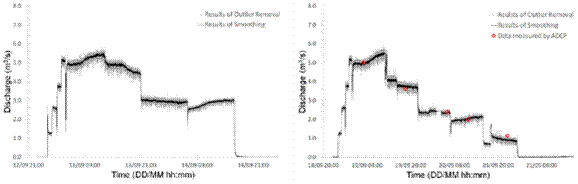
**表3 不同平滑常数因子的平滑水平和标准差（使用14/09 11:16至14/09 11:31的数据集，水文1-下游桥梁）。**

|  |  |  |
| --- | --- | --- |
| **平滑常数因子**α | **标准差（m）** | **平滑度（%）** |
| 1.00 | 0.0047 | 0.0 |
| 0.90 | 0.0043 | 9.3 |
| 0.80 | 0.0039 | 17.6 |
| 0.70 | 0.0035 | 25.3 |
| 0.60 | 0.0032 | 32.5 |
| 0.50 | 0.0028 | 39.4 |
| 0.40 | 0.0025 | 46.1 |
| 0.30 | 0.0022 | 53.0 |
| 0.20 | 0.0019 | 60.5 |
| 0.10 | 0.0014 | 70.6 |
| 0.01 | 0.0007 | 84.9 |

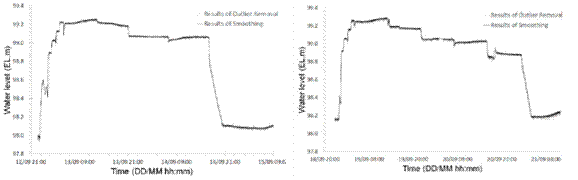
最后，图10显示了当数据集Rwo和EWMA平滑数据集E都转换为流量和水位时的结果。图10a，b显示了将超声波传感器到上游堰水库水面的距离转换为水位值，然后通过描述水位和流量之间关系的堰方程转换为流量的结果。即图10a、b和图2s目标流量过程线1和2。在实验中，通过ADCP（声学多普勒流速剖面仪）测得的流量数据的交叉验证，验证了过程线2图10b中的流量。ADCP是一种水声流速计，用于测量深度范围内的流速并产生流量数据。在过程线2的试验过程中，应用ADCP获取河道断面的流量信息，并将图10b中的结果与上游堰式水库中超声波传感器测得的水位换算流量数据进行比较。ADCP测量在过程线2的五个不同步骤中进行八次，结果通过八次测量的平均值产生。为了进行比较，将超声波传感器产生的流量流量数据平均，对应于过程线2每一步的八次ADCP测量。ADCP测得的流量偏差为−0.05 m3/s、+0.14 m3/s、0.06 m3/s、+0.07 m3/s和0.23 m3/s，然后依次为时间顺序。除最低流量条件外，由超声波传感器测量的水位换算的流量流量偏差范围为ADCP测量流量的−2.17%至3.81%。

图10a，b显示了经过异常值去除的数据和经过平滑处理的数据在离散度上的巨大差异。此外，

当流量较大时，这种差异更大。为了按阶段比较数据处理结果，图11显示了过程线1-下游桥梁数据集中从14/09 11:16到14/09 11:31的间隔数据，其中包括超声波传感器到水面的距离数据，该距离数据尚未转换为水位数据。该区间的异常点剔除和平滑结果的平均值分别为1.46855m和1.46860m，几乎相同。然而，未经数据处理的原始数据的平均值为1.232m，表明了由于未去除的异常值而产生的差异。

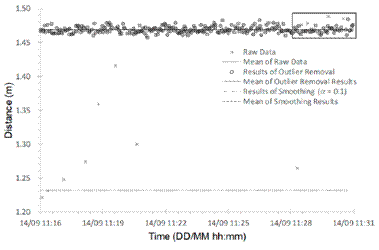


                                              （a） （b）



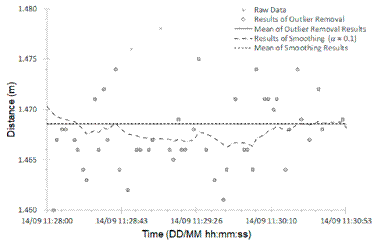
                                              （c） （d）

图10。通过超声波传感器监测上游堰的流量和下游桥梁的水位。（a） 当水位线1-上游堰时，由水位数据转换的流量；（b）当水位线2-上游堰时，由水位数据转换的流量；（c）当水位线1-下游桥时，由实测数据转换的水位；（d） 在水文图2-下游桥梁的情况下，通过测量数据转换的水位。



（一）

**图11 续.**



（b）

**图11 1号过程线下游桥梁超声波传感器测得的水位数据的原始数据、异常点去除结果和平滑结果的比较。（a） 比较部分从14/09 11:16到14/09 11:31；（b）比较部分从14/09 11:28到14/09 11:31。**

# 4 讨论和结论

本研究发展了一套针对使用超音波感测器在河川尺度实验中所收集水位资料的离群值剔除及平滑处理演算法。该研究包括对用户在每个数据处理步骤中主观确定的因素进行敏感性分析，包括异常点检测MAD的窗口大小、异常点去除修正Z分数的拒绝准则以及EWMA的平滑常数。

在本研究中，由超声波传感器收集的水位数据集中，存在许多异常值，这些异常值分布在数据集中，偏差较大，并且由于水波的出现，随机误差较大。考虑到这些数据的特点，本文提出的数据处理算法大致分为离群点剔除过程和平滑过程。该算法包含一个初始截止阶段，该阶段设置初始截止范围，并确定该区间以外的数据为异常值，将其整体删除。在超声波传感器采集的水位数据集中，这个过程检测到超过70%的异常值。初始截止后的异常值去除过程使用基于中位数和MAD的修正Z分数。离群点去除阶段包括确定用于设置移动中值和MAD的最佳窗口大小wo的过程。为了解决MAD为零的情况，例如当相同的值分布在给定窗口大小的数据集中的特定比率以上时，该算法将超声传感器的分辨率λ定义为MAD值。然后在数据平滑过程中使用EWMA，通过输入不同的平滑常数因子α值，可以根据数据处理目标计算不同级别的平滑结果。

在对算法进行敏感性分析的基础上，以剔除异常后移动中值与移动均值的最大相对差作为中心趋势测度的评价标准，确定了剔除异常过程的最佳窗口大小wo。当窗口大小较小时，异常值的检出率较高，样本中位数和MAD的估计精度降低，导致修正Z-score阶段异常值检出率较高。随着窗口大小w的增大，剔除异常值的比率呈指数下降。剔除异常值后，移动中值与移动均值的最大相对差随着w的减小而减小，直到一个固定水平，然后在最低水平后略有增大。β是修正Z-分数的拒绝准则，该算法用于剔除异常值，随着β值的增加，检出异常值的比率呈指数下降趋势。如果β超过3.0到3.5之间的范围，则去除的异常值的百分比减少会大大减少。最后，根据平滑常数因子α为1.0时的标准差值，将平滑过程中EWMA的平滑程度表示为α减小引起的标准差减小率。这项研究证实，随着平滑常数因子α的减小，超声波传感器测量的水位数据的平滑程度增加到标准偏差减少的最大84.9%。

# 参考文献

[1] Herschy，R.W.《河流测量》；CRC出版社：Boca Raton，佛罗里达州，美国，2014年。

[2] Fenton，J.D.；Keller，R.J.《根据水位测量计算流量》，技术报告01/6；集水区水文CRC：澳大利亚堪培拉，2001年。

[3] Boiten，W.《水文测量学：代尔夫特系列讲稿》，联合国教科文组织-国际水文地质学系列讲稿；第3版；泰勒和弗朗西斯：荷兰莱顿，2008年；第247页。

[4] Sauer，V.B.；Turnipseed，D.P.测量站的水位测量。《美国地质调查技术与方法》3-A7；美国地质调查局：弗吉尼亚州雷斯顿，美国，2010年；第45页。[交叉引用]

[5] 世界气象组织。第一卷——实地调查。世界气象组织第1044号。《河流测量手册》；世界气象组织：瑞士日内瓦，2010年；第250页。[交叉引用]

[6] Mcmillan，H.；Krueger，T.；Freer，J.《水文观测不确定性基准：降雨、河流流量和水质》。水解液。过程。2012, 26, 4078–4111. [交叉引用]

[7] Horner，I.；Renard，B.；Le Coz，J.；Branger，F.；McMillan，H.K.；Pierrefeu，G.《水位测量误差对水流不确定性的影响》。水资源。2018年第54号决议，1952-1976年。[交叉引用]

[8] Hamilton，A.S.；Moore，R.D.。量化流量记录中的不确定性。可以。水资源。J.2012，37，3–21。[交叉引用]

[9] Kruger，A.；Krajewski，W.F.；Niemeier，J.J.；Ceynar，D.L.；Goska，R.。桥装式水位传感器（BMRSS）。IEEE访问2016，48948–8966。[交叉引用]

[10] Mousa，M.；Zhang，X.；Claudel，C.。使用超声波和红外传感器在城市进行山洪探测。IEEE Sens.J.2016，167204–7216。[交叉引用]

[11] Rahmtalla，A.；Mohamed，A.；Wei，W.G.。使用超声波的实时无线洪水监测系统。国际科学杂志。2014年第3号决议，2012–2015年。

[12] Satria，D.；Yana，S.；Munadi，R.；Syahreza，S.。使用Arduino和GSM模块的基于谷歌地图的洪水监测系统原型。国际工程技术研究所。2017, 4, 1044–1047.

[13] Sunkpho，J.；Ootamakorn，C.《实时洪水监测和预警系统》。Songklanakarin科学杂志。技术。2011, 33, 227–235.

[14] Bae，I.；Yu，K.；Yoon，B.；Kim，S.采用远红外相机的表面图像测速仪的不可见环境适用性研究。韩国水资源。协会，2017，50597–607。（朝鲜语）[对照]

[15] Clemmens，A.J.；Wahl，T.L.；Bos，M.G.；Replogle，J.A.《用水槽和堰测量水》；国际土地复垦和蓄水研究所：荷兰瓦赫宁根，2001年；第382页。

[16] Cho，H.；Jeong，S.T.；Ko，D.H.；Son，K.-P.水温监测数据的有效异常值检测。韩国社会学杂志。海岸。海洋工程，2014，26，285–291。（朝鲜语）[对照]

[17] 国家灾害管理研究所（NDMRI）。小型河流设施标准改进实验与流量测量技术的发展；国家灾害管理研究所：Jungang dong，韩国，2017年。（韩语）

[18] 国际标准化组织。ISO 772:2011水文词汇和符号，第5版；国际标准化组织：瑞士日内瓦，2011年。

[19] Madli，R.；Hebbar，S.；Pattar，P.；Golla，V.。自动检测和通知道路上的坑洼和驼峰，以帮助驾驶员。IEEE Sens.J.2015，154313–4318。[交叉引用]

[20] Huber，P.J.《稳健统计》；Wiley:Hoboken，NJ，USA，1981年。[交叉引用]

[21] Leys，C.；Ley，C.；Klein，O.；Bernard，P.；Licata，L.检测异常值：不要使用平均值周围的标准偏差，使用中位数周围的绝对偏差。经验Soc。心理学。2013, 49, 764–766. [交叉引用]

[22] Rousseeuw，P.J.；Croux，C.《中位数绝对偏差的替代方法》。杰姆。Stat.Assoc.1993，88，1273–1283。[交叉引用]

[23] Croux，C.；Filzmoser，P.；Oliveira，M.R.《投影寻踪稳健主成分分析算法》。化疗。内尔。实验室系统。2007, 87, 218–225. [交叉引用]

[24] Maronna，R.A.；Martin，D.R.；Yohai，V.J.《稳健统计：理论与方法》；Wiley:Hoboken，NJ，USA，2006年。

[25] Huber，P.J.稳健统计。国际统计科学百科全书；施普林格：柏林，德国，2011；第1248-1251页。

[26] Iglewicz，B.；Hoaglin，D.C.《如何检测和处理异常值》；Asq出版社：美国威斯康星州密尔沃基，1993年；第87页。

[27] Satake，E.《临床科学的统计方法和推理：循证实践》；复数出版：美国德克萨斯州奥斯汀，2014年。

[28] 赫胥黎，T.H.《出其不意——意外的尾巴》。ICEAA 2016国际培训研讨会论文集，英国布里斯托尔，2016年10月17-20日。

[29] 布朗，R.G.离散时间序列的平滑、预测和预测；普伦蒂斯霍尔国际管理系列；普伦蒂斯霍尔：美国新泽西州上马鞍河，1963年。

[30] Machado，J.M.O.《会计数据中的异常值检测》。葡萄牙波尔图波尔图大学硕士论文，2018年。

[31] Chu，J.Y.；Shyr，J.-Y.；Zhong，W.决策树洞察数据挖掘。美国专利2014/0279775 A1，2014年9月18日。

[32] Vivcharuk，V.；Baardsnes，J.；Deprez，C.；Sulea，T.；Jaramillo，M.；Corbeil，C.R.；Mullick，A.；Magoon，J.；Marcil，A.；Durocher，Y.；等人。抗体和蛋白质治疗学（ADAPT）的辅助设计。《公共科学图书馆·综合》2017，12，e0181490。[交叉引用][PubMed]