

# 基于 EEMD 的水资源监测数据异常值检测与校正

方海泉<sup>1</sup> 薛惠锋<sup>1</sup> 蒋云钟<sup>2</sup> 周铁军<sup>3</sup> 万毅<sup>4</sup> 王海宁<sup>1</sup>

(1. 中国航天系统科学与工程研究院研究生部, 北京 100048; 2. 中国水利水电科学研究院水资源研究所, 北京 100038;  
3. 湖南农业大学理学院, 长沙 410128; 4. 水利部水资源管理中心, 北京 100053)

**摘要:** 提出利用中位数法与集成经验模态分解(EEMD)相结合的方法对时间序列数据的异常值进行检测, 首先通过中位数法对明显异常的数据进行初步筛选, 再用 EEMD 对剩余数据进行分解, 通过叠加低频分量可以拟合出大多数数据的整体变化趋势, 而不受异常值的影响, 从而根据偏差比率可有效检测出异常值。然后根据异常值检测后的时间序列数据的凹凸性变化趋势, 用分段曲线拟合对异常值校正。最后, 以 H1 自来水管的日取水量数据为例进行实证分析。结果表明: 提出的中位数法与 EEMD 相结合的方法能够有效地检测异常值, 校正后得到的数据能够真实反映该水厂取用水情况, 可为后续分析提供更加真实可靠的数据。

**关键词:** 水资源; 监测数据; 异常值; 集成经验模态分解; 中位数; 分段曲线拟合

**中图分类号:** N945 **文献标识码:** A **文章编号:** 1000-1298(2017)09-0257-07

## Outlier Detection and Correction for Water Resources Monitoring Data Based on EEMD

FANG Haiquan<sup>1</sup> XUE Huifeng<sup>1</sup> JIANG Yunzhong<sup>2</sup> ZHOU Tiejun<sup>3</sup> WAN Yi<sup>4</sup> WANG Haining<sup>1</sup>

(1. Graduate School, China Aerospace Academy of Systems Science and Engineering, Beijing 100048, China

2. Institute of Water Resources, China Institute of Water Resources and Hydropower Research, Beijing 100038, China

3. College of Science, Hunan Agricultural University, Changsha 410128, China

4. Water Resources Management Center, Ministry of Water Resources, Beijing 100053, China)

**Abstract:** In order to improve the availability and accuracy of online monitoring data of water resources, it is very important to detect and correct the outliers of monitoring data. The water resources monitoring data are non-linear and non-stationary time series data, the outlier detection method of the conventional time series did not take into account the convexity and concavity of time series. A combining median and ensemble empirical mode decomposition (EEMD) method was presented for outlier detection. Firstly, the outliers were preliminarily detected by the median method. And then the remaining data were decomposed by EEMD. The overall trend of most of the data can be fitted by superimposing the low-frequency components, but not affected by outlier, and the outlier can be detected effectively according to the deviation rate. Then, according to change of convexity and concavity of time series data after outlier detection, the method of piecewise curve fitting was used to correct the outliers. Finally, taking the daily water intake data of H1 waterworks as an example, the results showed that the method of combining median and EEMD can detect outliers effectively. The data obtained after correction can truly reflect the actual situation of water intake of waterworks. It can also provide more reliable data for subsequent analysis.

**Key words:** water resources; monitoring data; outliers; ensemble empirical mode decomposition; median; piecewise curve fitting

## 引言

水资源是人类社会赖以生存和发展的基本条

件,是一个国家经济社会发展的重要战略资源。2011年中央1号文件和中央水利工作会议明确要求实行最严格水资源管理制度。为支撑最严格水资

收稿日期: 2017-02-02 修回日期: 2017-02-28

基金项目: 国家自然科学基金委员会-广东联合基金项目(U1501253)和广东省省级科技计划项目(2016B010127005)

作者简介: 方海泉(1985—),男,博士生,主要从事数据分析与系统工程研究, E-mail: fanghaiquan22@126.com

源管理制度更好地实施,2012年国家启动了水资源监控能力建设项目<sup>[1-2]</sup>。在此之前,国家也已开展了很多相关工作,如2005年以来,各地相继开展了城市和流域水资源实时监控与管理系统建设,为及时全面准确掌握水资源信息奠定了良好基础<sup>[3-4]</sup>。通过对水资源在线监测,可以获取大量宝贵的真实客观数据。但水资源在线监测数据需要经历采集、传输、存储以及数据平台交换等环节,其中任何一个环节出现故障都会导致数据获取异常。为了更有效利用监测数据,需要采用合理的数据挖掘方法对异常值进行检测与校正。

水资源监测数据主要包括水量监测数据和水质监测数据。本文主要讨论水量监测数据中的取用水户日取水量监测数据,并以自来水管网的取水量为例进行分析。国内对于水量在线监测数据的分析才开始起步,文献[5]应用A-D检验方法判断取水户水量监测数据是否符合正态分布,从而判断获取的监测数据是否可靠。国外对流域水量水质监测、农业灌溉用水监测数据的研究较多,对取用水户取水量监测数据的分析也不多。文献[6]研究了美国阿肯色州农业灌溉水量问题,文献[7]应用主成分分析法对希腊东北部Kosynthos河的水量和水质数据进行了分析,文献[8]对土耳其流域水量和水质监测数据进行分析得出流量在逐年减少、水温在逐年增加的结论。

异常值目前尚无公认的准确定义,普遍采用的是HAWKINS<sup>[9]</sup>给出的定义。HAWKINS认为异常值是在数据集中与众不同的数据,使人怀疑这些数据并非随机偏差,而是产生于完全不同的机制。本文把取用水监测数据的异常值定义为对于任一时间点的数据相对其邻域时间范围内其他数据偏差较大的数据。该定义不仅考虑了时间序列数据的集合特性,还考虑了任一时刻点数据的时序变化特性。取用水监测数据本质上属于时间序列数据。对于时间序列异常值检测方法的研究,前人已经提出了很多算法,如基于统计、聚类、距离、密度等异常值检测方法<sup>[10-13]</sup>。文献[14]对国内外各种异常检测方法进行了对比,分析了每种异常检测方法的优缺点。近年来,异常值检测在很多领域都得到非常重要的研究和发展;文献[15]提出多步异常值检测方法应用于交通数据分析;文献[16]应用基于统计学的异常值分析方法对结构损伤进行识别;文献[17]应用函数型数据分析方法对水质监测数据进行异常值分析。通过异常值检测之后,把检测出的异常值替换为0值,然后再对0值进行填补,也就是对异常值进行校正。常用的0值填补方法有均值填补法、回归

填补法、热平台填补、多重填补法、人工神经网络等<sup>[18-21]</sup>。

针对取用水户日取水量监测数据呈现出季节波动和年度周期变化的特点,本文首先提出中位数法和集成经验模态分解(Ensemble empirical mode decomposition, EEMD)法相结合的方法对异常值进行检测,再用分段曲线拟合法对异常值校正。

## 1 理论方法

### 1.1 EEMD

经验模态分解(EMD)方法是由HUANG等<sup>[22]</sup>于1998年提出。EMD在机械故障诊断、地球物理探测、生物医学分析、回波检测等方面都得到广泛的应用<sup>[23-26]</sup>。尚未见有文献用于时间序列异常值检测。EMD可将不同尺度的波动或趋势从原信号中逐级分离出来<sup>[22]</sup>。EMD适用于分析非线性非平稳信号,且具有完全自适应性。原始的时间序列 $x(t)$ 经过EMD分解得到一系列固有模态函数(Intrinsic mode function, IMF)和一个余项(residual, 简记res),即

$$x(t) = \sum_{i=1}^n c_i + r_n \quad (1)$$

式中  $c_i$ ——第*i*个IMF  $r_n$ ——余项  
 $n$ ——分解得到的固有模态函数的个数

EEMD是EMD的改进算法,有效解决了EMD的混频现象<sup>[27]</sup>。

### 1.2 分段曲线拟合

曲线拟合的原理是已知样本点 $(x_i, y_i)$  ( $i=1, 2, \dots, n$ ), 求出解析函数 $y \approx f(x)$ , 使 $f(x)$ 在原样本点 $x_i$ 上尽可能接近 $y_i$ 的值<sup>[28]</sup>。曲线拟合方法包括最小二乘曲线拟合(多项式拟合)和线性插值拟合, 本文采用最小二乘曲线拟合方法。最小二乘曲线拟合是使误差平方和最小的多项式拟合<sup>[28]</sup>。即寻找一条曲线使在误差平方和最小的准则下与所有数据点最为接近, 也即找出使 $\sum_{i=1}^n |f(x_i) - y_i|^2$ 最小的 $f(x)$ 。通常, 多项式的次数选用3次, 低于3次, 可能会造成波峰丢失, 高于3次, 拟合时间长, 容易产生虚假波峰<sup>[29]</sup>。

在曲线拟合之前需要判断时间序列数据的凹凸性。若凹凸性没有发生变化, 可用直接曲线拟合; 若凹凸性发生变化, 需要对时间序列数据进行分段, 再对每一段分别进行曲线拟合, 这就是分段曲线拟合。直接曲线拟合可以视为分段曲线拟合的特殊情况, 此时分为一段。凹凸性可根据变化趋势判断, EMD是目前提取时间序列趋势的最好方法<sup>[30]</sup>。

## 2 异常值检测与校正方法

### 2.1 异常值检测与校正模型

面对大量原始的取用水监测数据,对数据进行异常值检测和校正是后序分析必不可少的重要环节,本文建立了取用水监测数据的异常值检测与校正模型,如图1所示。该模型不仅适用于取用水监测数据,也适用于一般的时间序列数据异常值的检测与校正。

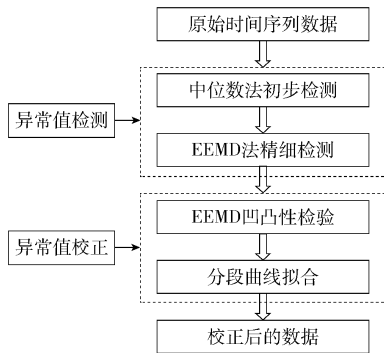


图1 异常值检测与校正模型

Fig. 1 Model of outlier detection and correction

异常值检测与校正模型的总体思路: 首先,应用中位数法对原始时间序列数据进行异常值初步检测,其次,用EEMD法进一步精细检测,把检测出的异常值替换为0,再通过EEMD判断凹凸性检验,根据凹凸性进行分段,最后应用分段曲线拟合方法对0值进行填补。经过异常值检测与校正得到更加接近真实、有效的时间序列数据。

### 2.2 异常值检测

一般情况下,取用水监测数据的异常值包括异常大值、异常小值(大于0)、0值和负值4种类型。根据自来水厂取用水实际情况,一般水厂每天24h都在不间断取水,负值和0值属于异常值。因此,本研究主要讨论异常大值和异常小值(大于0)2种类型的检测,并提出中位数与EEMD结合的异常值检测方法。

#### 2.2.1 中位数法初步检测

对于任一时间序列数据 $\{a_i\}$ ,  $i=1, 2, \dots, t$ 表示时间序列的天数。把 $\{a_i\}$ 中的正数选出来构成一个新的数列,记为 $\{a_{i_k}^+\}$ ,把 $\{a_{i_k}^+\}$ 的中位数记为 $\text{med}\{a_{i_k}^+\}$ , $\{a_{i_k}^+\}$ 中每个数与中位数的比值 $r_{i_k} = \frac{a_{i_k}^+}{\text{med}\{a_{i_k}^+\}}$ ,若 $r_{i_k} > 1$ , $r_{i_k}$ 越大,则该日数据为异常大值的可能性越大;若 $r_{i_k} < 1$ , $r_{i_k}$ 越小,则该日数据为异常小值的可能性越大。

设定 $r_{i_k}$ 的阈值为 $r_1, r_2$ ,若 $r_{i_k} > r_1 > 1$ 或者 $r_{i_k} < r_2 < 1$ ,则 $a_{i_k}^+$ 为异常值,把异常值替换为0,原始的监

测数据 $\{a_i\}$ 经过替换后得到的数列记为 $\{b_i\}$ 。阈值 $r_1, r_2$ 可设置相对宽松一点,主要是为了把明显的异常大值和异常小值检测出来并替换为0,可减少这类明显的异常大值和异常小值对EEMD分解的影响。

#### 2.2.2 EEMD法精细检测

把经过第一次异常值替换后得到的 $\{b_i\}$ 中的正数选出来构成一个新的数列 $\{b_{i_k}^+\}$ ,且每个正数之间的排列顺序不变,把这样的数列称为正数组。再用EEMD对 $\{b_{i_k}^+\}$ 进行分解,得到 $n$ 个分量(包括 $n-1$ 个IMF和1个残余项 $r_n$ ),按频率依次从高到低排列,舍去高频分量,把后面 $m(m=1, 2, \dots, n)$ 个低频分量对应求和可得到一个新的数列,记为 $\{c_{i_k}\}$ 。 $m$ 值的选取需要经过多次尝试,若 $m$ 选取过小, $\{c_{i_k}\}$ 与 $\{b_{i_k}^+\}$ 的偏差过大,会把正常值当做异常值检测,称为错检;若 $m$ 选取过大, $\{c_{i_k}\}$ 与 $\{b_{i_k}^+\}$ 过于接近,异常值就检测不出来,称为漏检。根据经验,一般情况下 $m:n=2:3$ 时比较合适。从几何图像上看,当 $m$ 选取的恰当时, $\{c_{i_k}\}$ 既能很好接近 $\{b_{i_k}^+\}$ 中的正常值,又不受异常值影响,从而既不会错检,也不会漏检。

为了量化 $\{b_{i_k}^+\}$ 与 $\{c_{i_k}\}$ 的偏离程度,定义偏差比率为

$$d_{i_k} = \frac{|c_{i_k} - b_{i_k}^+|}{c_{i_k}} \quad (2)$$

$d_{i_k}$ 越大,则该时间点的数据为异常值的可能性也就越大。

设 $d_{i_k}$ 的阈值为 $d_0$ ,即偏差比率大于 $d_0$ 的数据为异常值。也即当 $b_{i_k}^+ < (1 - d_0) c_{i_k}$ 或者 $b_{i_k}^+ > (1 + d_0) c_{i_k}$ 时, $b_{i_k}^+$ 为异常值。经过异常值检测之后再把异常值替换为0,把 $\{b_i\}$ 经过异常值替换得到的数列记为 $\{g_i\}$ 。

### 2.3 异常值校正

原始的监测数据 $\{a_i\}$ 经过2次异常值检测和替换变为 $\{g_i\}$ ,异常值都被替换为0,继而需要对0值进行填补,也就是对异常值进行校正。填补之前,需要判断 $\{g_i\}$ 的凹凸性,因此需要提取 $\{g_i\}$ 的变化趋势。先把 $\{g_i\}$ 的正数选出来构成一个新的数列 $\{g_{i_k}^+\}$ ,且每个正数之间的排列顺序不变,再用EEMD对 $\{g_{i_k}^+\}$ 进行分解得到 $n_0$ 个分量,选取后 $h$ 个低频分量对应求和构成新的数列作为趋势项 $\{u_{i_k}^+\}$ , $h$ 值的选取也需要经过多次实验确定,根据经验,一般情况下 $h:n_0=1:3$ 时趋势项 $\{u_{i_k}^+\}$ 能较好地体现 $\{g_{i_k}^+\}$ 的变化趋势。若 $\{g_i\}$ 的凹凸性没有改变,可直接曲线拟合;若 $\{g_i\}$ 的凹凸性改变了,需要

用分段曲线拟合。分段点的选取原则是确保分段之后每一段的凹凸性不变,这样,分段点的选取范围就可以比较宽,可避开0值为分段点。

直接曲线拟合方法是把 $\{g_i\}$ 中的正数选出来,得到一组样本点 $(i_+, g_{i_+})$ , $i_+$ 表示 $\{g_i\}$ 中出现正数的位置,采用多项式对样本点 $(i_+, g_{i_+})$ 进行曲线拟合,得到的多项式记为 $f(x)$ 。用曲线拟合值 $f(x_{i_0})$ 替换 $\{g_i\}$ 中的0值得到校正后的时间序列数据,记为 $\{k_i\}$ ,这里 $i_0$ 表示 $\{g_i\}$ 中0值出现的位置。分段曲线拟合就是对每个时间段分别进行曲线拟合,对于每一段曲线拟合方法与直接曲线拟合相同,拟合后再把不同时间段校正后的结果按时间顺序拼接起来得到最后的结果。

### 3 实例研究

#### 3.1 研究对象

选取 H1 自来水管厂的日取水量数据为研究对象。监测数据来源于水资源管理系统数据库。选取时间范围为2015年1月1日—2016年12月31日,共计731 d。

#### 3.2 异常值的检测结果

##### 3.2.1 原始监测数据时间序列图

H1 自来水管厂的原始日取水量时间序列监测数据为 $\{a_i\}$ ,如图2所示。从图2中可以看出,存在异常大值、异常小值和0值,并且存在连续多个0值。

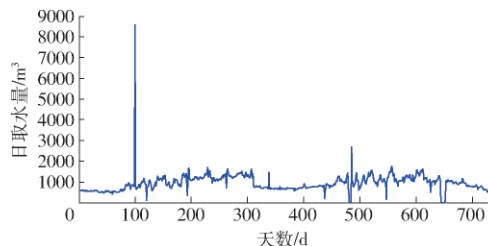


图2 原始监测数据(2015-01-01—2016-12-31)

Fig. 2 Original monitoring data

##### 3.2.2 中位数法异常值初步检测

设定异常值初步检测的条件为监测数据大于中位数的5倍或者小于中位数的0.1倍,经过初步检测后把异常值替换为0。原始的监测数据经过替换后得到序列 $\{b_i\}$ ,如图3所示。

##### 3.2.3 EEMD 法异常值进一步检测

把 $\{b_i\}$ 中的正数选出来构成正数组 $\{b_{i_k}^+\}$ ,并保持正数组中每个正数的次序不变,用EEMD对正数组进行分解,从高频到低频依次得到9个分量,包括8个IMF和1个余项 $r_n$ ,如图4所示。把后面6个低频分量对应相加得到一个新数组 $\{c_{i_k}\}$ ,如图5所示。从图5可以看出,后面6个低频分量叠加构

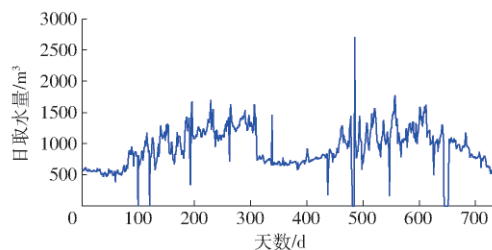


图3 异常值初步检测后的数据  
(2015-01-01—2016-12-31)

Fig. 3 Data after initial outlier detection

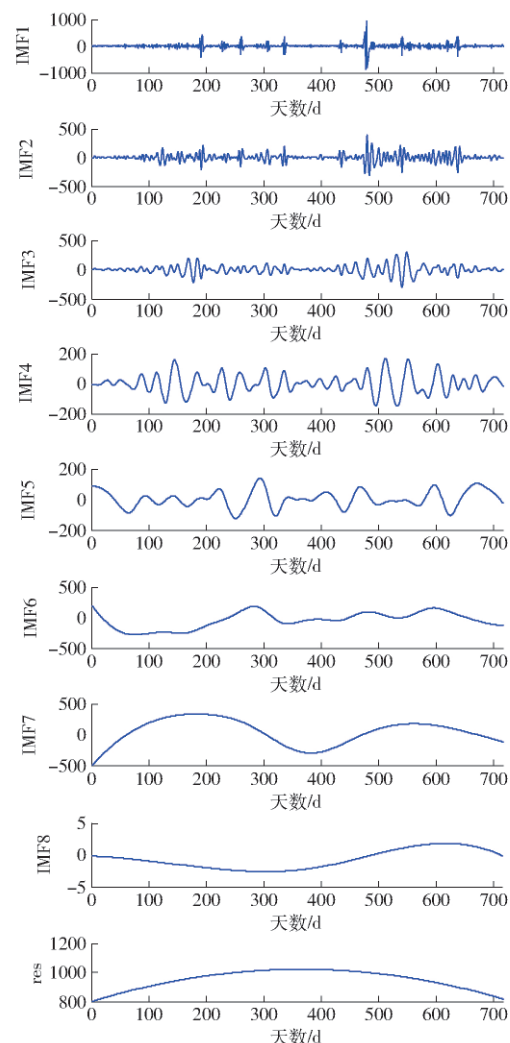


图4 异常值初步检测后集成经验模态分解  
(2015-01-01—2016-12-31)

Fig. 4 EEMD after initial outlier detection

成的数组 $\{c_{i_k}\}$ 能很好地拟合绝大部分正常数据,且不受异常值的影响,其原理可通过图4看出。图4的前3个高频分量的幅度变化剧烈,去掉前3个高频分量相当于滤波,滤除了波动幅度大的值,剩下的后6个低频分量叠加之后得到的曲线相对平缓。

根据式(2)计算偏差比率,结果如图6所示。设 $d_{i_k}$ 的阈值为0.5,即偏差比率大于0.5的数据为异常值,也即当 $b_{i_k}^+ < 0.5c_{i_k}$ 或者 $b_{i_k}^+ > 1.5c_{i_k}$ 时, $b_{i_k}^+$ 为异常值。经过第2次异常值检测之后把异常值替换

为0。 $\{b_i\}$ 经过异常值替换后变为数列 $\{g_i\}$ ,如图7所示。

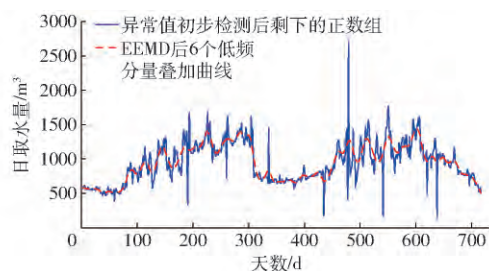


图5 后6个低频分量叠加曲线  
(2015-01-01—2016-12-31)

Fig.5 Curve superposed by six low frequency components

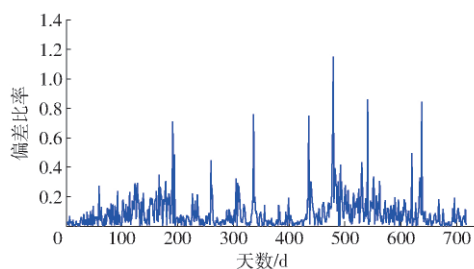


图6 偏差比率(2015-01-01—2016-12-31)

Fig.6 Deviation rate

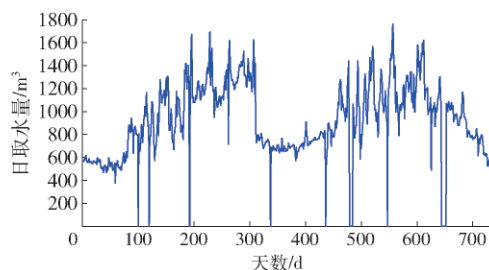


图7 基于中位数与EEMD结合的异常值检测  
(2015-01-01—2016-12-31)

Fig.7 Outlier detection based on method of combining median and EEMD

### 3.3 中位数-EEMD 异常值检测优势

为了体现本文提出方法的优势,与传统的异常值检测方法箱线图对比。箱线图异常值检测方法基于统计学原理<sup>[31]</sup>,能够检测出数据集中异常大和异常小值,但是没有考虑时序变化特性,与图7进行对比,图8中的箭头标出了用箱线图未能检测出的异常值,尤其是第3个箭头(从左到右数)标出的异常值,该异常值相比临近时间点的数据异常大,但是从整个数据集中看不属于异常。证明本文提出检测异常值检测方法非常有效。

### 3.4 异常值校正结果

#### 3.4.1 用EEMD提取异常值检测后的变化趋势

用EEMD对 $\{g_i\}$ 的正数组 $\{g_{i_k}^+\}$ 进行分解得到9个分量,选取后3个低频分量对应求和得到的数列 $\{u_{i_k}^+\}$ 作为变化趋势,如图9所示。可以看出凹凸

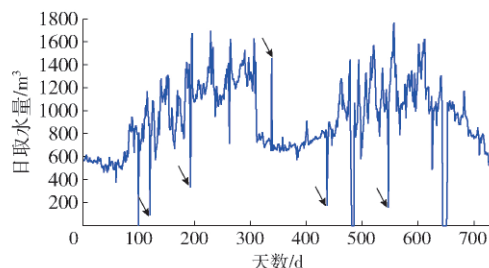


图8 基于箱线图的异常值检测  
(2015-01-01—2016-12-31)

Fig.8 Outlier detection based on boxplot

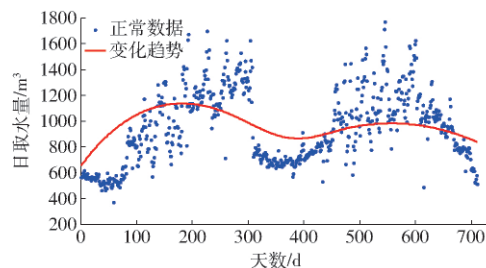


图9 基于EEMD凹凸性检验  
(2015-01-01—2016-12-31)

Fig.9 Convexity and concavity test based on EEMD

性发生了变化,需要用分段曲线拟合填补0值。选取第365天的位置为分段点把 $\{g_i\}$ 分为2段,即每一段为1年数据,分别记为 $\{g_{1i}\}$ 和 $\{g_{2i}\}$ 。

#### 3.4.2 用分段曲线拟合法对异常值校正

把 $\{g_{1i}\}$ 中的正数选出来进行曲线拟合,用曲线拟合的结果替换相应位置的0值,经过校正后获得的时间序列数据记为 $\{k_{1i}\}$ ,结果如图10左半图所示。同样地,对 $\{g_{2i}\}$ 的0值进行填补,结果如图10右半图所示。把2段时间序列数据校正后得到的结果按时间顺序拼接起来得到时间序列数据如图11所示。

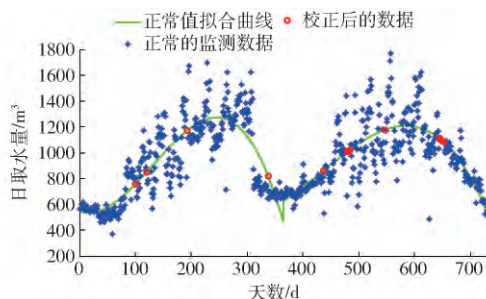


图10 分段曲线拟合校正异常值  
(2015-01-01—2016-12-31)

Fig.10 Outlier correction by piecewise curve fitting

### 3.5 异常值校正结果对比

为了与分段曲线拟合对比,采用直接曲线拟合对异常值校正,结果如图12所示。与图10对比可以看出,对于凹凸性发生改变的时间序列数据,分段拟合曲线能更好地体现时序变化趋势。因此,对于



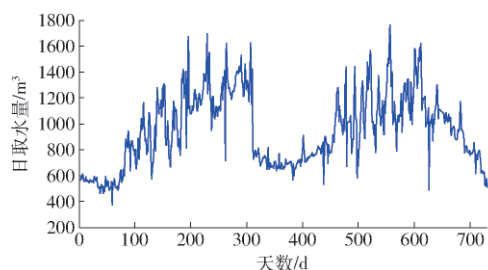


图 11 经过异常值检测和校正后得到的数据  
(2015-01-01—2016-12-31)

Fig. 11 Data after outlier detection and correction

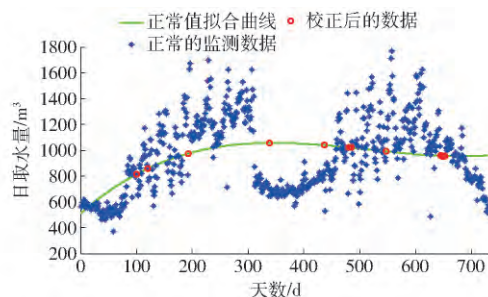


图 12 直接曲线拟合校正异常值  
(2015-01-01—2016-12-31)

Fig. 12 Outlier correction by direct curve fitting

凹凸性发生改变的时间序列数据,对异常值进行校正时,分段曲线拟合结果优于直接曲线拟合结果。

### 3.6 讨论

常规的异常值检测与校正方法对时间序列数据的缺失个数有一定的要求,若数据缺失严重则难以

恢复,连续缺失数据越多恢复越困难。本文提出的中位数-EEMD异常值检测和曲线拟合异常值校正方法在使用之前同样需要考虑数据缺失问题。

## 4 结束语

随着水资源监控能力建设项目的不断完善和发展,水资源管理系统存储的水资源监测数据逐日递增,为充分发挥大量的监测数据在水资源管理业务上的作用,有效的数据预处理方法至关重要。对于时间跨度大且变化趋势复杂的非线性非平稳水资源在线监测数据,本文提出了中位数法与EMMD相结合的异常值检测方法和分段曲线拟合的异常值校正方法。通过以H1自来水厂的日取水量数据为例,并与传统的异常值检测和校正方法进行对比,结果说明本文提出的方法更加优越,经过异常值检测和校正得到更加接近真实的日取水量数据。本文提出的异常值检测和校正方法可为监测数据服务于水资源管理业务提供参考,并且可推广应用于其他领域的时间序列数据的异常值检测和校正。该方法适用于历史监测数据,对于实时在线监测数据异常值的检测和校正需要以此为基础,因为实时在线监测数据异常值检测和校正必须建立在历史监测数据可靠的基础上,所以本文提出的方法对于下一步实时在线监测数据异常值检测和校正也起到很好的理论支撑作用。

## 参 考 文 献

- 1 蔡阳. 国家水资源监控能力建设项目及其进展[J]. 水利信息化, 2013(6): 5-10.  
CAI Y. National water resources monitoring capacity building project and its progress[J]. Water Resources Informatization, 2013(6): 5-10. (in Chinese)
- 2 中华人民共和国水利部, 财政部. 国家水资源监控能力建设项目实施方案(2012—2014)(水资源[2012]411号)[Z]. 北京: 中华人民共和国水利部, 2012.
- 3 蒋云钟, 张小娟, 石玉波, 等. 水资源实时监控与管理系统标准体系建设[J]. 中国水利, 2007(1): 55-58.  
JIANG Y Z, ZHANG X J, SHI Y B, et al. On the construction of standard system of water resources real-time monitoring and controlling system[J]. China Water Resources, 2007(1): 55-58. (in Chinese)
- 4 蒋云钟, 万毅. 水资源监控能力建设功能需求及实施策略[J]. 中国水利, 2012(7): 26-30.  
JIANG Y Z, WAN Y. Demand analysis of water resources monitoring capacity and implementation strategy study[J]. China Water Resources, 2012(7): 26-30. (in Chinese)
- 5 蒋吉发, 刘飞. A-D检验在取水户水量在线监测数据稳定性分析中的应用[J]. 四川水利, 2016(2): 53-55.  
JIANG J F, LIU F. A-D detection in the application of stability analysis on on-line monitoring data about intake water amount[J]. Sichuan Water Conservancy, 2016(2): 53-55. (in Chinese)
- 6 REBA M L, DANIELS M, CHEN Y S, et al. A statewide network for monitoring agricultural water quality and water quantity in Arkansas[J]. Journal of Soil and Water Conservation, 2013, 68(2): 45-49.
- 7 PISINARAS V, PETALAS C, GEMITZI A, et al. Water quantity and quality monitoring of Kosynthos river, North-eastern Greece[J]. Global NEST Journal, 2007, 9(3): 259-268.
- 8 ODEMIS B, EVRENDILEK F. Monitoring water quality and quantity of national watersheds in Turkey[J]. Environmental Monitoring & Assessment, 2007, 133(1-3): 215-229.
- 9 HAWKINS D. Identification of outliers[M]. London: Chapman and Hall, 1980.
- 10 BARNET V, LEWIS T. Outliers in statistical data[M]. Chichester: John Wiley and Sons, 1994.
- 11 WANG J S, CHIANG J C. A cluster validity measure with outliers detection for support vector clustering[J]. IEEE Transactions

- on Systems Man and Cybernetics , Part B-Cybernetics , 2008 , 38( 1) : 78 – 89.
- 12 KNORR E M , NG R T. A unified notion of outliers: properties and computation [C] // Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining , 1997: 219 – 222.
- 13 BREUNIG M , KRIEGER H P , NG R , et al. LOF: identifying density-based local outliers [C] // Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data , 2000: 93 – 104.
- 14 陈运文 , 吴飞 , 吴庐山 等. 基于异常检测的时间序列研究 [J]. 计算机技术与发展 , 2015 , 25( 4) : 166 – 170.  
CHEN Y W , WU F , WU L S , et al. Research on time series based on anomaly detection [J]. Computer Technology and Development , 2015 , 25( 4) : 166 – 170. ( in Chinese)
- 15 MONOWAR H B , BHATTACHARYA D K , KALITA J K. A multi-step outlier-based anomaly detection approach to network-wide traffic [J]. Information Sciences , 2016 , 348: 243 – 271.
- 16 缙百勇 , 陆秋海 , 向志海 等. 损伤识别的模态数据异常值分析方法 [J]. 清华大学学报: 自然科学版 , 2015 , 55( 3) : 356 – 360.  
GOU B Y , LU Q H , XIANG Z H , et al. Damage detection using modal data outlier analysis [J]. Journal of Tsinghua University: Science and Technology , 2015 , 55( 3) : 356 – 360. ( in Chinese)
- 17 DÍAZ M C , GARCÍA N P J , ALONSO F J R. Detection of outliers in water quality monitoring samples using functional data analysis in San Esteban estuary ( Northern Spain) [J]. Science of the Total Environment , 2012 , 439: 54 – 61.
- 18 BELLO A L. Imputation techniques in regression analysis: looking closely at their implementation [J]. Computational Statistics and Data Analysis , 1995 , 20( 1) : 45 – 57.
- 19 RAO J N K , SHAO J. Jackknife variance estimation with survey data under hot deck imputation [J]. Biometrika , 1992 , 79( 4) : 811 – 822.
- 20 RUBIN D B. Multiple imputations in sample surveys [R]. American Statistical Association , 1978: 20 – 34.
- 21 AMER S R. Neural network imputation: a new fashion or a good tool [D]. Corvallis , OR: Oregon State University , 2004.
- 22 HUANG N E , SHEN Z. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis [J]. Proceedings of the Royal Society of London , 1998 , A454: 903 – 995.
- 23 LAU B , RIEMENSCHNEIDER S , XU Y. Gearbox fault diagnosis using empirical mode decomposition and Hilbert spechum [J]. Mechanical Systems and Signal Processing , 2006 , 20( 3) : 718 – 734.
- 24 ZHANG R R , ASCE M , MA S , et al. Hilbert-Huang transform analysis of dynamic and earthquake motion recordings [J]. Journal of Engineering Mechanics , 2003 , 129( 8) : 861 – 875.
- 25 CHAPPELL M A , PAYNE S J. A method for the automated detection of venous gas bubbles in humans using empirical mode decomposition [J]. Annals of Biomedical Engineering , 2005 , 33( 10) : 1411 – 1421.
- 26 曾祥 , 周晓军 , 杨辰龙 等. 基于经验模态分解和 S 变换的缺陷超声回波检测方法 [J/OL]. 农业机械学报 , 2016 , 47( 11) : 414 – 420. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?file\\_no=20161156&flag=1](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20161156&flag=1). DOI: 10.6041/j.issn.1000-4298.2016.11.056.  
ZENG X , ZHOU X J , YANG C L , et al. Ultrasonic defect echoes identification based on empirical mode decomposition and s-transform [J/OL]. Transactions of the Chinese Society for Agricultural Machinery , 2016 , 47( 11) : 414 – 420. ( in Chinese)
- 27 WU Z H , HUANG N E , LONG S R , et al. On the trend , detrending , and the variability of nonlinear and non-stationary time series [J]. Proceedings of the National Academy of Sciences , 2007 , 104: 14889 – 14894.
- 28 胡守信 , 李柏年. 基于 MATLAB 的数学实验 [M]. 北京: 科学出版社 2004: 63 – 65.
- 29 张霞 , 李儒 , 岳跃民 等. 谐波改进的植被指数时间序列重建算法 [J]. 遥感学报 , 2010 , 14( 3) : 442 – 447.  
ZHANG X , LI R , YUE Y M , et al. Improved algorithm for reconstructing vegetation index image time series based on Fourier harmonic analysis [J]. Journal of Remote Sensing , 2010 , 14( 3) : 442 – 447. ( in Chinese)
- 30 邓拥军 , 王伟 , 钱成春 等. EMD 方法及 Hilbert 变换中边界问题的处理 [J]. 科学通报 , 2001 , 46( 3) : 257 – 263.  
DENG Y J , WANG W , QIAN C C , et al. EMD method and transact boundary problem of Hilbert transform [J]. Chinese Science Bulletin , 2001 , 46( 3) : 257 – 263. ( in Chinese)
- 31 盛骤 , 谢式千 , 潘承毅. 概率论与数理统计 [M]. 北京: 高等教育出版社 2008: 132 – 135.