



基于改进累积和控制图的GNSS变形信息的识别与预警

吴昊¹ 刘超¹ 赵兴旺¹

1 安徽理工大学空间信息与测绘工程学院, 安徽 淮南, 232001

摘要: 累积和(cumulative sum, CUSUM)控制图是一种较为成熟的突变点识别方法, 可以较为有效地对全球导航卫星系统(global navigation satellite system, GNSS)坐标序列中的变形信息进行识别与预警。然而, 对于长时间的连续监测, GNSS坐标序列中不可避免地存在粗差或离群值, 其对CUSUM控制图的检测效果有一定的影响。鉴于此, 提出了采用基于中位数构建的改进CUSUM控制图进行GNSS坐标序列中变形信息的识别与预警, 并给出了该方法的基本原理和算法流程。结果表明: 采用改进CUSUM控制图进行变形信息识别和预警的准确度优于传统方法, 其灾害预警的误报率显著减小, 说明了采用改进CUSUM控制图的必要性和优越性。

关键词: 控制图; 粗差; 累积和; GNSS; 变形监测; 灾害预警

中图分类号: P228

文献标志码: A

高层建筑与大型桥梁变形、城市地表沉陷和高速公路边坡滑坡等风险, 对人民生命财产安全构成了巨大的威胁。因此, 有必要对变形体进行有效监测, 获取变形信息, 采用一定的数学方法进行数据处理与分析, 进而进行有效的预警预报, 减小灾害发生概率和影响范围^[1-3]。

全球导航卫星系统(global navigation satellite system, GNSS)作为一种高精度定位技术, 可以为变形体提供连续实时动态监测, 已成为一种变形监测的重要手段, 在相关变形监测领域得到了广泛应用^[4-9]。采用GNSS进行动态变形监测, 其主要目的之一是对变形灾害进行实时识别和预警。对于如何准确识别GNSS坐标序列中的变形信息, 相关学者展开了广泛的研究。胡东^[10]提出了利用添加季节和趋势分解与断点识别(breaks for additive season and trend, BFAST)算法分析GPS坐标序列, 并利用蒙特卡洛方法对数据进行多尺度模拟, 通过实验验证了此方法的有效性。伊廷华等^[11]为了有效地判别GPS坐标序列的突变点, 建立了GPS监测序列的关联模型, 提出了基于关联负选择的坐标序列突变点的分步识别算法。在诸多研究突变点识别理论方

法中, 控制图理论以其较为成熟的理论和良好的检测效果, 在坐标序列的突变点识别中得到了关注。常见的控制图理论包括休哈特控制图、累积和(cumulative sum, CUSUM)控制图和移动加权控制图等方法^[12]。Miao等^[13]应用均值控制图对桥梁的GNSS动态坐标异常变化进行统计模式识别, 得到了很好的识别效果; Ogaja等^[14]采用CUSUM控制图进行结构振动频率突变的检测, 验证了该方法进行较小突变检测的可行性; Iz^[15]采用CUSUM控制图对GPS和甚长基线干涉测量(very long baseline interferometry, VLBI)的差异进行突变检测, 获取地震活动引起的基线的瞬时变化特征; 郭庆^[16]基于CUSUM控制图构建了GPS异常监测数据的预警模型, 实现了不同标准差条件下突变点的有效探测; 伊廷华等^[17]对比分析了休哈特控制图和CUSUM控制图在标准差偏移方面的识别能力, 发现两者各有优势, 相互补充。

虽然GNSS动态变形监测通常采用载波相对定位的方式可以消除其卫星钟差和接收机钟差, 也可以将对流层误差和电离层误差等空间相关误差大大削弱, 但多路径误差和衍射误差等非

收稿日期: 2019-12-27

项目资助: 国家自然科学基金(41474026)。

第一作者: 吴昊, 硕士生, 主要从事变形灾害的预警与预报。haowu0505@gmail.com

通讯作者: 刘超, 博士, 副教授。chaoliu0202@gmail.com

建模系统误差仍然影响监测结果,因此,GNSS坐标序列通常不服从正态分布,难以满足控制图相关研究规定的样本数据服从正态分布的要求。虽然相关学者提出了解决这一问题的途径,如Quesenberry等^[18]提出的 Q 统计量的方法,对原始GNSS坐标序列进行转换,使转换后的坐标序列近似地服从正态分布,但考虑到GNSS长时间监测过程中可能出现的个别粗差或离群值,相关的正态化方法难以对其进行有效处理,其仍会影响采用相关控制图方法进行GNSS变形监测突变点的识别精度。鉴于此,本文以CUSUM控制图方法为例,提出采用基于中位数构建的改进CUSUM控制图进行GNSS坐标序列中变形灾害的识别,并与经典的CUSUM方法进行对比分析,得出了一些有益的结论。

1 基于中位数的累积和控制图

控制图是统计质量管理的一种重要手段和工具,其通常基于采集数据的样本均值和假设检验的原理所构造,用于监测生产过程是否处于控制状态。在变形监测领域,若被监测体处于安全状况或受到较小变形的影响,则GNSS监测数据具有相对稳定的均值和方差;若被监测体出现一定程度的变形,GNSS监测数据相应地出现一定程度的偏移,但其方差未发生改变。因此,可基于控制图理论建立变形灾害的识别模型。

具体描述如下:对于GNSS监测的坐标序列 $X(t)$, $t=1, 2, \dots, n$, n 为样本大小,认为其近似服从正态分布,即 $X(t) \sim N(\mu_0, \sigma)$,其中, μ_0 为坐标序列的均值, σ 为坐标序列的标准差。若监测体发生变形,其监测的坐标序列近似服从 $X(t) \sim N(\mu_0 + \delta\sigma, \sigma)$,其中, $\delta\sigma$ 为坐标序列的均值偏移量^[19]。

令 $\mu_1 = \mu_0 + \delta\sigma$,进行假设检验。原假设 H_0 为坐标序列未发生异常波动,其均值不变;备选假设 H_1 为坐标序列在 $s(s < n)$ 时刻发生异常波动,即 s 时刻发生变形,认为在 s 时刻以后的监测数据的均值发生改变。

在坐标序列服从正态分布的前提下, s 时刻前后的 $X(t)$ 的概率密度函数可以表示为:

$$\begin{cases} f_0(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu_0)^2}{2\sigma^2}\right], H_0 \text{ 成立} \\ f_1(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu_1)^2}{2\sigma^2}\right], H_1 \text{ 成立} \end{cases} \quad (1)$$

构建对数似然比统计量:

$$\lambda_n = \sum_{t=1}^n \ln \frac{f_1(x_t)}{f_0(x_t)} = \sum_{t=1}^n \frac{1}{2\sigma^2} \left[(x_t - \mu_0)^2 - (x_t - \mu_1)^2 \right] \quad (2)$$

令均值偏移量 $\mu_1 - \mu_0 = \delta\sigma = \Delta$, $k = \Delta/2$ 。则

$$\lambda_n = \sum_{t=1}^n \frac{\Delta}{\sigma^2} (x_t - \mu_0 - k) \quad (3)$$

若 $\Delta > 0$,则式(3)等价于 $C_t^+ = \max \sum_{j=1}^t (x_j - \mu_0 - k)$ ^[17],其中, C_t^+ 为 $\Delta > 0$ 条件下的似然比统计量,其可以进一步表示为:

$$\begin{aligned} C_t^+ &= C_{t-1}^+ + (x_t - \mu_0 - k) = \\ &\max(0, C_{t-1}^+ + x_t - \mu_0 - k) \end{aligned} \quad (4)$$

式(4)即为CUSUM的上偏移统计量,同理,可得CUSUM的下偏移统计量为:

$$C_t^- = \min(0, C_{t-1}^- + x_t - \mu_0 + k) \quad (5)$$

为了削弱样本中可能出现的粗差或离群值对变形信息识别精度的影响,采用中位数代替原始样本进行似然比统计量的构建。中位数计算方法如下:

$$\tilde{x}_i = \text{median}(x_{m \cdot (i-1) + 1}, x_{m \cdot (i-1) + 2}, \dots, x_{m \cdot i}) \quad (6)$$

式中, \tilde{x}_i 为原始样本的中位数; i 为形成的中位数样本数目, $i=1, 2, \dots, \text{INT}(n/m)$, $\text{INT}(\cdot)$ 为取整函数, m 为设置的计算中位数的子样本容量; $\text{median}(\cdot)$ 为中位数函数。

采用式(6)得到的中位数样本构建改进的CUSUM(称为CUSUM中位数)的上、下偏移统计量:

$$\begin{cases} \tilde{C}_i^+ = \max(0, \tilde{C}_{i-1}^+ + \tilde{x}_i - \tilde{\mu}_0 - \tilde{k}) \\ \tilde{C}_i^- = \min(0, \tilde{C}_{i-1}^- + \tilde{x}_i - \tilde{\mu}_0 + \tilde{k}) \end{cases} \quad (7)$$

式中, $\tilde{\mu}_0$ 为中位数样本的均值; $\tilde{k} = \frac{\tilde{\mu}_1 - \tilde{\mu}_0}{2}$, $\tilde{\mu}_1$ 为发生变形后的中位数样本的均值。

基于式(7)上、下统计量构建相应的变形灾害预警模型如下:

$$\begin{cases} \tilde{C}_i^+ = \max(0, \tilde{C}_{i-1}^+ + \tilde{x}_i - \tilde{\mu}_0 - \tilde{k}) \\ \tilde{C}_i^+ > h, \text{ 预警} \end{cases} \quad (8)$$

$$\begin{cases} \tilde{C}_i^- = \min(0, \tilde{C}_{i-1}^- + \tilde{x}_i - \tilde{\mu}_0 + \tilde{k}) \\ \tilde{C}_i^- < -h, \text{ 预警} \end{cases} \quad (9)$$

式中, \tilde{C}_0^+ 和 \tilde{C}_0^- 分别为统计量的上、下偏移的初始值,通常取值为0; h 为判断变形灾害发生的阈值,即当上统计量大于 h 或下统计量小于 $-h$ 时,

视为发生变形并进行预警。

为了确定阈值 h , 引入平均运行链长 (average run length, ARL), ARL 为控制图发出失控状态之前的平均样本数。ARL 的计算方法有很多, 如常见的马尔可夫链法和积分方程法。本文拟采用 Tran 等提出的方法进行相关计算^[20]。

当检测过程处于控制状态时, ARL 被表示为 ARL_0 ; 相反, 当检测过程失控时, ARL 被表示为 ARL_1 。 ARL_0 值越大, 表示对异常数据判别中发生的误报率 (第一类错误) 越低; ARL_1 值越小, 表示对异常数据判别中发生的漏报率 (第二类错误) 越低。故而, 关于控制图最优参数的选取, 可以认为是在 ARL_0 固定不变的情况下, ARL_1 的值越小越好, 即:

$$\begin{cases} ARL_1(h, k, \Delta) = \min \\ ARL_0 \geq \tau \end{cases} \quad (10)$$

式中, τ 为允许的最大误报率, 依据传统的 6σ 原理, 通常设 $\tau = 370.4$ ^[21]。在此基础上, 对于不同的偏移值 Δ , 通过模式搜索的方法获取控制限 h 的最优解。为了便于描述, 在构建 \bar{x}_i 中位数统计量时将样本容量设为奇数, 以样本容量为 3、5 和 7 为例, 在 $\Delta \in [0.2, 2.0]$ 时计算最优决策量 (h, k, ARL_1), 结果见图 1。图 1 中, 在偏移取值不同的情况下, k 在样本容量为 3 和 5 时取值近似, 与偏移值 Δ 的关系可表示为 $k = \Delta/2$ 。样本容量为 7 时, 在 $\Delta \leq 1.6$ 之前与样本容量是 3 和 5 时的取值近似, 在 $\Delta > 1.6$ 之后逐渐稳定到 0.8。控制限 h 与 ARL 的取值随着 Δ 的增加而逐渐下降, 样本容量越大, 则数值越低。

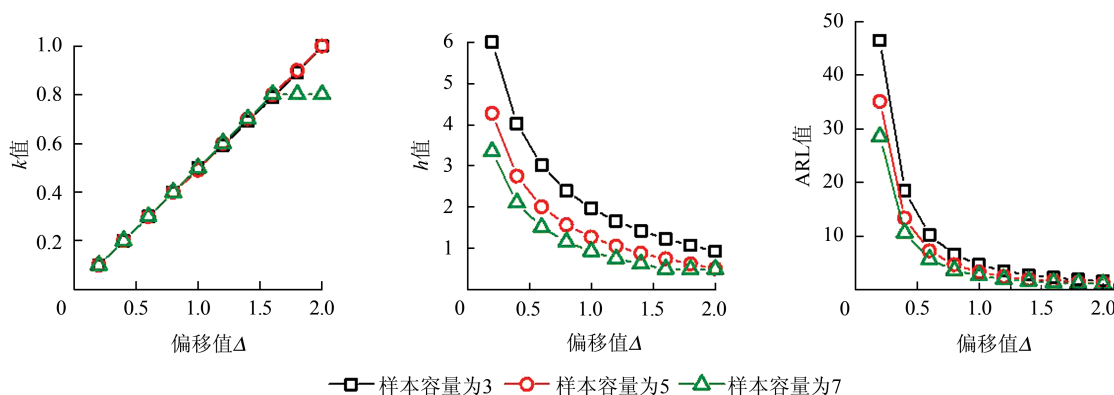


图 1 偏移值 Δ 与最优决策量 (h, k, ARL) 的关系图

Fig.1 Relationship Between Offset Value(Δ) and Optimal Decision Quantity(h, k, ARL)

2 GNSS 变形信息的识别与预警

2.1 正态性检验

由于 CUSUM 控制图理论是基于对服从正态分布的数据进行检验, 因此, 在进行变形数据的识别与预警之前需要对其进行正态性判别。判定方法大致可分为图形判别和构建检验统计量两类。本文选择的方法是构建统计量中的 Shapiro-Wilks 检验^[22]。在本文中, 通过 R 语言采用 Shapiro-Wilks 检验中 p -value 的验证方法。

2.2 数据转换

由于 GNSS 监测数据受到多路径误差及其他非建模误差的影响, 其数据通常不服从正态性, 因此, 在运用控制图进行变形信息的检验和预警之前, 需要对其进行转换。本文采用 Quenberry 等^[18]提出的方法, 用来解决 GNSS 监测数据 $X(t)$ ($t=1, 2, \dots, n$) 不服从正态分布的问

题^[23-24]。设 n 个样本点的概率密度函数为:

$$F(i) = \frac{1}{n} \sum_{l=1}^n K_l(x_i - x_l) = \frac{1}{nl} K\left(\frac{x_i - x_l}{l}\right) \quad (11)$$

式中, $K(\cdot)$ 为核函数; $K_l(\cdot)$ 为缩放核函数, 且 $K_l(x) = 1/(l \cdot K(x))$, 其中 l 为一个平滑参数; $i=1, 2, \dots, n$; $t=1, 2, \dots, n$ 。在此基础上, 对原始 GNSS 监测数据进行 Q 统计量转换:

$$\tilde{Q}(i) = \Phi^{-1}(F(i)) \quad (12)$$

式中, $\Phi^{-1}(\cdot)$ 是标准正态分布的累积分布函数的逆函数。通过转换后的 \tilde{Q} 统计量近似服从正态分布, 已基本满足对 GNSS 变形信息的识别与预警算法的条件。

2.3 算法识别步骤

综上所述, 本文提出的基于 CUSUM 中位数控制图的 GNSS 变形信息识别与预警算法描述如下:

1) 获取监测体的 GNSS 变形监测数据, 并计算其均值和标准差;

2)对所采集的GNSS监测数据进行Shapiro-Wilks正态性检验,将不服从正态分布的数据使用 Q 统计量进行转换;

3)将数据(或转换后的数据)按照式(6)计算中位数;

4)统计步骤3)得到的中位数,根据式(8)、式(9)构建CUSUM中位数控制图的上、下偏移统计量;

5)依据图1,确定异常值识别模型的参数 k 和控制限 h ;

6)对比判断CUSUM中位数上、下偏移统计量是否超过控制限 h ,对于超过控制限的数据定义为异常数据并进行预警。

3 数据处理与分析

数据来源为安徽理工大学某楼顶的GNSS连续跟踪站,采用其中相距约4 m的两个站点的观测数据进行相关算法的验证。GNSS接收机均采用天宝BD980板卡,天线型号为AT300,数据源为GPS/GLONASS/BDS三系统,采样频率设置为3 Hz。截取500 s的 x 方向坐标序列(1 500个观测历元),并减去其真实值(由长时间观测平均得到),如图2所示。由于两个站点的距离较近,通过载波相位双差技术可以较为有效地削弱空间相关系统误差,可以认为其坐标序列中的系统误差主要为多路径误差等非建模系统误差;同时,站点架设在建筑物的高度仅约12 m,且相对稳固,通过长时间观测,未发现明显的变形信息。因此,可以将该段坐标序列作为不含有变形信息的GNSS相对定位的观测坐标序列,并考虑加入不同特征的变形信息,进行后续算法的测试。

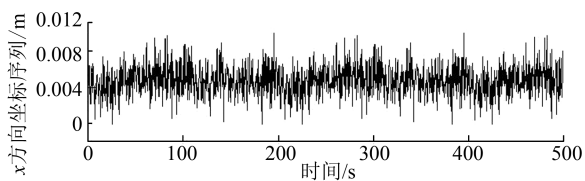


图2 实验采用的GNSS坐标序列

Fig.2 GNSS Coordinate Series Used in the Experiment

进行变形信息识别和预警,需先对原始GNSS坐标序列进行正态性检验,检验结果见表1,其中, P 值小于0.05,拒绝原假设,即认为原始GNSS坐标序列不服从正态分布。采用式(11)、式(12)对原始GNSS坐标序列进行数据转换,得到近似服从正态分布的坐标序列。在原始GNSS坐标序列中分别加入无粗差影响的不同标准差偏移、离散粗差影响的偏移、连续粗差影响的偏

移和连续突变型变形等4种不同形式的变形信息,验证算法对不同条件下变形信息的识别和预警的性能。

表1 GNSS坐标序列的正态性检验结果

Tab.1 Normality Test Results of GNSS Coordinate Series

参数	结果	参数	结果
平均值	0.004 74	峰度	0.141 71
标准误差	0.000 04	偏度	0.094 71
中位数	0.004 61	W值	0.997 21
标准差	0.001 66	P值	0.004 91

3.1 无粗差影响的变形信息的检验

在原始GNSS坐标序列的241~300历元(80~100 s)处分别加入1~3倍的标准差的上偏移,形成3组突变型变形的测试数据,如图3所示。

分别采用经典CUSUM控制图和基于中位数的改进CUSUM控制图两种方法对上述变形信息进行识别和预警。其中,经典CUSUM控制图的控制限的确定参考Siegmund^[25]所提出的ARL的计算方法;基于中位数的改进CUSUM控制图的控制限的确定参见图1,设样本容量为3,并取 $\Delta=1$ 。

两种方法的变形信息识别结果如图4和图5所示,可知:

1)无论采用哪种方法构建统计量,得到的统计量序列均有效地抑制了噪声的影响,变形信息得到了有效增强。以加入1倍标准差偏移为例,由于变形量较小,在原始坐标序列中,其变形信息被噪声淹没(见图3),而在构建的统计量序列(见图4(a)和图5(a))中,可以较为清晰地识别加入变形信息的区间。

2)基于中位数的改进CUSUM控制图的识别精度高于经典CUSUM控制图。表2给出了两种方法对不同变形信息区间的识别结果,其中,对于1~3倍标准差的变形,经典CUSUM控制图的误报历元数和漏报历元数分别为73、143、164和4、3、3,而对应基于中位数的改进CUSUM控制图的误报历元数和漏报历元数分别为3、45、60和17、11、5。新方法的漏报率有所增大,但其误报率得到了较大程度的减小,识别的变形区间明显收窄,与真实区间更为接近,表明在发生变形结束后,新方法构建的统计量具有更为迅速的回归速度。

3)虽然图3中未加入任何下偏移变形信息,但经典CUSUM控制图仍检测出了少量下偏移

变形信息。对于 1~3 倍标准差的变形,经典 CUSUM 控制图分别检测出 21、39、45 个下偏移变形历元,而基于中位数的改进 CUSUM 控制图未检测出任何下偏移变形信息,说明新方法可以有效地减小预报的误警率。

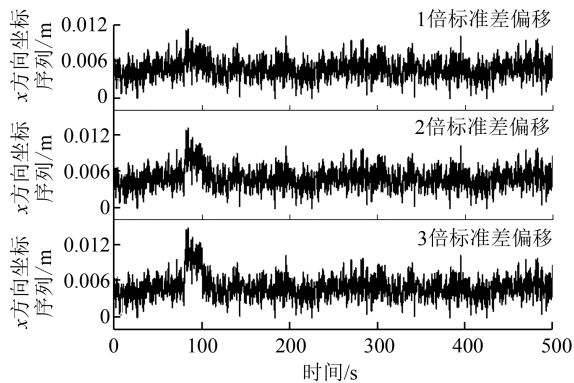


图3 含有不同偏移量的 GNSS 坐标序列

Fig.3 GNSS Coordinate Series with Different Offsets

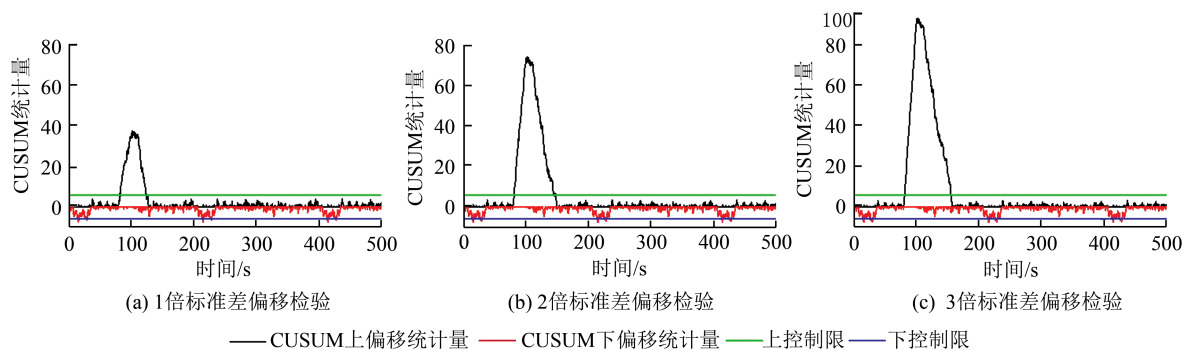


图4 经典 CUSUM 控制图对不同标准差变形数据的识别结果

Fig.4 Recognition Results of CUSUM Control Chart for Different Standard Deviation Deformation Data

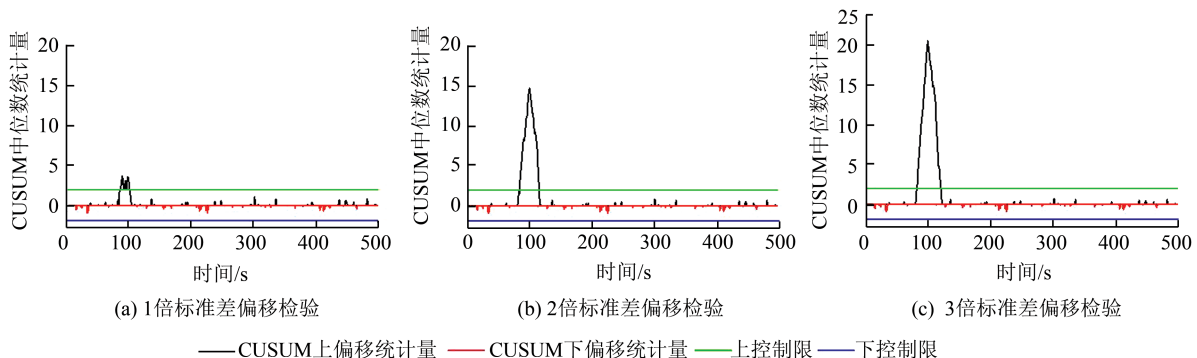


图5 改进 CUSUM 控制图对不同标准差变形数据的识别结果

Fig.5 Recognition Results of Modified CUSUM Control Chart for Different Standard Deviation Deformation Data

2)在 2 倍标准差偏移序列的 67、361、418 历元处分别加入 -5 倍、-8 倍、12 倍标准差大小的粗差;

3)在 3 倍标准差偏移序列的 42、351、1 212 历元处分别加入 20 倍、30 倍、10 倍标准差大小的粗差。

经典 CUSUM 控制图和基于中位数的改进 CUSUM 控制图对含有离散粗差的变形信息的识

表2 CUSUM 及其改进方法对变形信息区间的识别结果/历元

Tab.2 Recognition Results of CUSUM and Its Modified Method for Deformation Information /Epoch

变形信息	识别结果	
	经典 CUSUM	改进 CUSUM
1 倍标准差偏移	245~373	258~303
2 倍标准差偏移	244~443	252~345
3 倍标准差偏移	244~464	246~360
区间真值	241~300	

3.2 离散粗差影响的变形信息的检验

在加入不同标准差偏移(见图 3)的基础上,随机选择不同的位置加入一定大小的粗差,用于进一步测试算法的性能。具体方案为:

1)在 1 倍标准差偏移序列的 16、151、610 历元处分别加入 10 倍、5 倍、-16 倍标准差大小的粗差;

别结果见表 3。由表 3 可知:

1)加入粗差后,两种方法均可以识别出加入变形的区间(241~300 历元),但识别的精度有所差异。主要因为粗差的位置与变形区间临近,影响了变形信息的识别精度;而基于中位数的改进 CUSUM 控制图并未受到粗差的影响,对变形区间的识别结果与未加入粗差的结果基本一致。

表3 CUSUM及其改进方法对含有离散粗差的变形序列的识别结果/历元

Tab.3 Recognition Results of CUSUM and Its Modified Method for Deformation Series With Discrete Gross Errors /Epoch

变形信息	检验方向	识别结果	
		经典CUSUM	改进CUSUM
1倍标准差 偏移	上偏移	16~20, 151~152, 245~373	258~303
	下偏移	45~47, 84, 87, 610~712, 1 243~1 245, 1 282, 1 285	
2倍标准差 偏移	上偏移	244~446	252~345
	下偏移	45~47, 67~108, 361~370, 644~687, 1 243~1 245, 1 277~1 286	
3倍标准差 偏移	上偏移	42~54, 244~491, 1 212~1 218	246~357
	下偏移	40, 41, 87, 639~650, 654~657, 660, 663, 664, 670~704, 1 238~1 240, 1 243~1 249, 1 253~1 256, 1 259, 1 262~1 263, 1 270~1 303	

2)对于加入粗差的位置,经典CUSUM控制图均错误地识别为变形信息,而基于中位数的改进CUSUM控制图未识别出任何变形信息。如对于加入1倍标准差偏差的坐标序列,在其16、151和610历元分别加入了10倍、5倍、-16倍标准差大小的粗差,经典CUSUM控制图错误识别的区间分别为16~20、151~152和610~712历元,而基于中位数的改进CUSUM控制图未作出任何变形预警;其他两个变形序列的识别结果类同。

3)统计两种方法进行变形预警的误报历元数和漏报历元数。对于1~3倍标准差的变形,经典CUSUM控制图的误报历元数和漏报历元数分别为193、258、319和4、3、3;而对应基于中位数的改进CUSUM控制图的误报历元数和漏报历元数分别为3、45、57和17、11、5。

3.3 连续粗差影响的变形信息的检验

为了进一步分析在连续粗差影响下该方法的检验性能,在加入2倍标准差大小偏移(见图3)的基础上,随机选择不同的位置加入一定大小的

连续粗差,具体方案为:

1)在50~51历元处分别加入3倍、5倍的标准差大小的粗差;

2)在209~210历元处分别加入3倍、-5倍的标准差大小的粗差;

3)在726~728历元处分别加入7倍、8倍、-9倍的标准差大小的粗差;

4)在1 269~1 271历元处分别加入7倍、8倍、9倍的标准差大小的粗差。

由于样本容量的选择直接影响中位数的生成,进而影响基于中位数的改进CUSUM控制图方法的识别精度,选择不同的样本容量(3、5和7)进行变形信息的识别,结果如表4所示。由表4可知:

1)与表3相比较,对于离散的粗差点,基于中位数的改进CUSUM控制图(样本容量为3)可以很好地避免粗差点的误识别,而对于连续的粗差(见表4),该方法(样本容量为3)将连续粗差误识别为变形信息,如方案3和4的粗差被识别的区间分别为729~735历元和1 272~1 281历元。

表4 基于改进CUSUM控制图对含有连续粗差的变形序列的识别结果/历元

Tab.4 Recognition Results of Deformation Series with Continuous Gross Error Based on the Modified CUSUM /Epoch

样本容量	方案	改进CUSUM	
		上偏移识别结果	下偏移识别结果
3	1	252~342	—
	2	252~342	—
	3	252~345, 729~735	—
	4	252~342, 1 272~1 281	—
5	1	250~385, 590	80, 85, 90, 390, 680~685, 1 240~1 245, 1 255
	2	250~385, 590	80, 85, 90, 390, 625, 680~685, 1 240~1 245, 1 255
	3	250~385, 590	80, 85, 90, 390, 625, 680~685, 1 240~1 245, 1 255
	4	250~385, 590, 1 274~1 295	80, 85, 90, 390, 625, 680~685, 1 240~1 245, 1 255
7	1	252~378, 1 190	1 106, 1 225, 1 281~1 288
	2	252~378, 1 190	1 106, 1 225, 1 281~1 288
	3	252~378, 1 190	1 106, 1 225, 1 281~1 288
	4	252~378, 1 190	110, 612, 251, 281

2) 样本容量的增大虽然可以削弱连续粗差的误报率, 但对应阈值不断减小, 会造成其他位置误报率的增加。当样本容量设置为 7 时, 粗差位置并未被识别为变形信息, 但是 1 190 历元被误识别为上偏移变形, 1 106、1 225 等位置被误识别为下偏移变形。因此, 对于样本容量的选择, 还需要根据序列的具体特征综合考虑。

3.4 连续突变型变形信息的检验

为了进一步测试该算法对不同类型变形信息的检验性能, 设置一组连续突变型的变形信息进行测试, 具体方案为: 在原始监测数据 (见图 2) 的基础上, 选择在 1 201~1 500 历元处加入 0.01 m 的偏移量作为变形数据, 如图 6 所示。

两种方法的结果分别如图 7 和图 8 所示, 其统计结果见表 5。由图 7、图 8 和表 5 可知, 与经典 CUSUM 控制图方法相比, 基于中位数的 CUSUM 控制图方法的误警率明显减小。如图 7 所示, 虽然在实验中未添加任何下偏移变形信息, 但经典 CUSUM 控制图检验出的下偏移统计量超过控制限的数目高达 230 个, 如 34~149、387、638~748 等历元均被误识别为下偏移变形信息; 而基于中位数的改进 CUSUM 控制图方法未检测出任何的下偏移变形信息 (见图 8), 误警率得到了有效控制。另外, 两种方法均有效地识别出了添加的偏移量区间 (1 201~1 500 历元), 其中基于中位数的改进 CUSUM 控制图的漏报率略高于经典 CUSUM 控制图, 但综合考虑误警率和漏警率, 改进 CUSUM 控制图的识别精度高于经典 CUSUM 控制图。

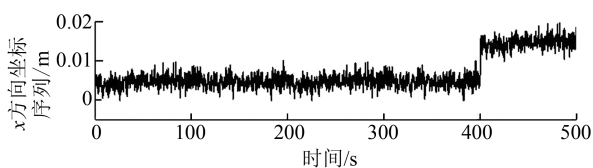


图 6 含有突变型变形信息的 GNSS 坐标序列

Fig.6 GNSS Coordinate Series with an Abrupt Change

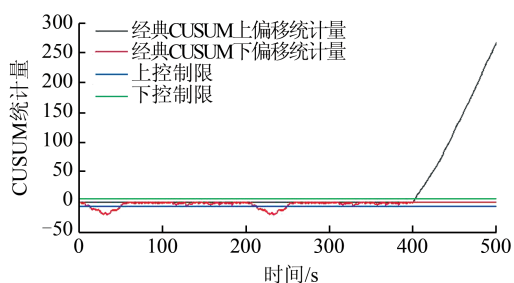


图 7 经典 CUSUM 控制图对连续突变型变形数据的识别结果

Fig.7 Recognition Results of the Abrupt Changes in GNSS Coordinate Series by CUSUM

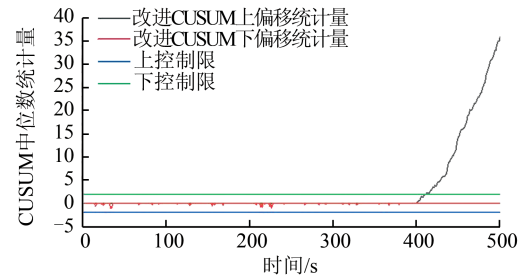


图 8 改进 CUSUM 控制图对连续突变型变形数据的识别结果

Fig.8 Recognition Results of the Abrupt Changes in GNSS Coordinate Series by the Modified CUSUM

表 5 CUSUM 及其改进方法对连续突变型变形序列的识别结果 / 历元

Tab.5 Recognition Results of the Abrupt Changes in GNSS Coordinate Series by CUSUM and Its Improved Method / Epoch

变形信息	上偏移	下偏移
经典 CUSUM	1 209~1 500	34~149、387、633~635、 638~748、986
改进 CUSUM	1 236~1 500	无
区间真值	1 201~1 500	无

4 结 语

考虑到 GNSS 变形监测中可能出现的粗差会影响变形信息的识别与预警, 提出采用基于中位数的 CUSUM 控制图进行变形信息的识别与预警。通过理论推导和数据处理与分析, 得到如下结论:

1) CUSUM 控制图以及基于中位数的改进 CUSUM 控制图均可以有效地抑制噪声对变形信息的影响, 增强监测序列的变形特征;

2) 对于未加入粗差的不同变形序列, 基于中位数的改进 CUSUM 控制图对变形信息的识别精度高于经典 CUSUM 方法;

3) 对于加入离散粗差的变形序列, 基于中位数的改进 CUSUM 控制图优势更为明显, 其表现出了良好的稳健性, 与采用该方法对未加入粗差的变形序列的识别结果基本相当; 对于存在连续粗差的变形序列, 采用基于中位数的改进 CUSUM 控制图进行变形信息的识别与预警, 需要设置合适的样本容量, 以达到最佳的检验效果。

参 考 文 献

- [1] Chen Yongqi, James L. Development of the Methodology for Single Epoch GPS Deformation Monitoring [J]. *Journal of Wuhan University of Surveying and*

- Mapping*, 1998, 23(4): 324-328, 363(陈永奇, James L. 单历元GPS变形监测数据处理方法的研究[J]. 武汉测绘科技大学学报, 1998, 23(4): 324-328, 363)
- [2] Berardino P, Fornaro G, Lanari R, et al. A New Algorithm for Surface Deformation Monitoring Based on Small Baseline Differential SAR Interferograms[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2002, 40(11): 2 375-2 383
- [3] Monserrat O, Crosetto M. Deformation Measurement Using Terrestrial Laser Scanning Data and Least Squares 3D Surface Matching[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2008, 63(1): 142-154
- [4] Im S B, Hurlbaeus S, Kang Y J. Summary Review of GPS Technology for Structural Health Monitoring[J]. *Journal of Structural Engineering*, 2011, 139(10): 1 653-1 664
- [5] Leach M P, Hyzak M. GPS Structural Monitoring as Applied to Cable-Stayed Suspension Bridge[C]. Proceedings of the 20th FIG Congress, Melbourne, Australia, 1994
- [6] Lovse J W, Teskey W F, Lachapelle G, et al. Dynamic Deformation Monitoring of Tall Structure Using GPS Technology[J]. *Journal of Surveying Engineering*, 1995, 121(1): 35-40
- [7] Guo J J. Research of Displacement and Frequency of Tall Building Under Wind Load Using GPS[C]. Proceedings of the 10th International Technical Meeting of Satellite Division of the Institute of Navigation, Missouri, USA, 1997
- [8] Jiang Weiping, Liu Hongfei, Liu Wanke, et al. CORS Development for Xilongchi Dam Deformation Monitoring[J]. *Geomatics and Information Science of Wuhan University*, 2012, 37(8): 949-952(姜卫平, 刘鸿飞, 刘万科, 等. 西龙池上水库GPS变形监测系统研究及实现[J]. 武汉大学学报·信息科学版, 2012, 37(8): 949-952)
- [9] Ye Shirong, Zhao Lewen, Chen Dezhong, et al. Real-Time Deformation Monitoring Data Processing Based on BDS Triple-Frequency Observations[J]. *Geomatics and Information Science of Wuhan University*, 2016, 41(6): 722-728(叶世榕, 赵乐文, 陈德忠, 等. 基于北斗三频的实时变形监测数据处理[J]. 武汉大学学报·信息科学版, 2016, 41(6): 722-728)
- [10] Hu Dong. GPS Time Series Analysis Using BFAST Algorithm[D]. Chengdu: Southwest Jiaotong University, 2016(胡东. 利用BFAST算法分析GPS时间序列[D]. 成都:西南交通大学, 2016)
- [11] Yi Tinghua, Guo Qing, Li Hongnan, et al. Multi-step Identification Algorithm Based on Relational Negative Selection for GPS Abnormal Monitoring Data[J]. *Journal of Vibration Engineering*, 2015, 28(1): 1-8(伊廷华, 郭庆, 李宏男, 等. GPS异常监测数据的关联负选择分步识别算法[J]. 振动工程学报, 2015, 28(1): 1-8)
- [12] Zhou Jixiang, Mao Shisong. Statistical Methods for Quality Management[M]. Beijing: China Statistics Press, 2008(周纪芄, 茆诗松. 质量管理统计方法[M]. 北京: 中国统计出版社, 2008)
- [13] Miao Changqing, Wang Man, Tian Hongjin, et al. Damage Alarming of Long-span Suspension Bridge Based on GPS-RTK Monitoring[J]. *Journal of Central South University*, 2015, 22(7): 2 800-2 808
- [14] Ogaja C, Wang J, Rizos C. Detection of Wind-Induced Response by Wavelet Transformed GPS Solutions[J]. *Journal of Surveying Engineering*, 2003, 129(3): 99-104
- [15] Iz H B. Differencing Reveals Hidden Changes in Baseline Length Time-Series[J]. *Journal of Geodesy*, 2006, 80(5): 259-269
- [16] Guo Qing. Research on Distortion Analysis and Abnormally Recognition of GPS Structural Monitoring Data[D]. Dalian: Dalian University of Technology, 2013(郭庆. GPS结构监测数据失真分析与异常识别方法研究[D]. 大连: 大连理工大学, 2013)
- [17] Yi Tinghua, Guo Qing, Li Hongnan. The Research on Detection Methods of GPS Abnormal Monitoring Data Based on Control Chart[J]. *Engineering Mechanics*, 2013, 30(8): 133-141(伊廷华, 郭庆, 李宏男. 基于控制图GPS异常监测数据检验方法研究[J]. 工程力学, 2013, 30(8): 133-141)
- [18] Quesenberry C P. On Properties of Binomial Q Charts for Variables[J]. *Journal of Quality Technology*, 1995, 27(3): 184-203
- [19] Mertikas S P, Damianidis K I. Monitoring the Quality of GPS Station Coordinates in Real Time[J]. *GPS Solutions*, 2007, 11(2): 119-128
- [20] Tran K P, Castagliola P, Celano G. Monitoring the Ratio of Population Means of a Bivariate Normal Distribution Using CUSUM Type Control Charts[J]. *Statistical Papers*, 2018, 59(1): 387-413
- [21] Hu Xuelong, Zhou Xiaojian, Huang Weidong, et al. Design of the CUSUM Median Chart[J]. *Journal of Applied Statistics and Management*, 2018, 37(3): 469-477(胡雪龙, 周晓剑, 黄卫东, 等. CUSUM中位数控制图设计[J]. 数理统计与管理, 2018, 37(3): 469-477)
- [22] Shapiro S S, Wilk M B. An Analysis of Variance Test for Normality[J]. *Biometrika*, 1965, 52(3):

- 591-611
- [23] Silverman B W. Using Kernel Density Estimates to Investigate Multimodality[J]. *Journal of the Royal Statistical Society Series B (Methodological)*, 1981, 43(1): 97-99
- [24] Parzen E. On Estimation of a Probability Density Function and Mode[J]. *The Annals of Mathematical Statistics*, 1962, 33(3): 1 065-1 076
- [25] Siegmund D. Error Probabilities and Average Sample Number of the Sequential Probability Ratio Test [J]. *Journal of the Royal Statistical Society Series B (Methodological)*, 1975, 37(3): 394-401

Identification and Forewarning of GNSS Deformation Information Based on a Modified Cumulative Sum Control Chart

WU Hao¹ LIU Chao¹ ZHAO Xingwang¹

¹ School of Geomatics, Anhui University of Science and Technology, Huainan 232001, China

Abstract: Cumulative sum (CUSUM) control chart is a more mature method for outlier identification, which can effectively identify and forewarn the deformation information in the global navigation satellite system (GNSS) coordinate series, but for long-term continuous monitoring, gross errors inevitably exist in the GNSS coordinate series, which has a certain impact on the accuracy of the CUSUM control chart. Therefore, a modified CUSUM control chart based on median was proposed to identify and forewarn the deformation information in the GNSS coordinate series, and its principle and algorithm flow were given. The results show that the accuracy of deformation information identification and forewarning using modified CUSUM control chart is better than traditional method, and the false alarm rate of the disaster warning are reduced, which shows the necessity and superiority of using the modified CUSUM control chart.

Key words: control charts; gross error; cumulative sum; global navigation satellite system; deformation monitoring; disaster forewarning

First author: WU Hao, postgraduate, specializes in early warning and forecast of deformation disaster. E-mail: haowu0505@gmail.com

Corresponding author: LIU Chao, PhD, associate professor. E-mail: chaoliu0202@gmail.com

Foundation support: The National Natural Science Foundation of China(41474026).

引文格式: WU Hao, LIU Chao, ZHAO Xingwang. Identification and Forewarning of GNSS Deformation Information Based on a Modified Cumulative Sum Control Chart[J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(10): 1517-1525. DOI: 10.13203/j.whugis20190003 (吴昊, 刘超, 赵兴旺. 基于改进累积和控制图的 GNSS 变形信息的识别与预警[J]. *武汉大学学报·信息科学版*, 2020, 45(10): 1517-1525. DOI: 10.13203/j.whugis20190003)