

基于期望最大算法的空间事件及异常值探测*

刘劲宏^{1,2} 吴晨韵¹ 徐 劲² 杜建丽² 雷祥旭³

1(重庆交通大学智慧城市学院 重庆 400074)

2(中国科学院紫金山天文台 南京 210023)

3(山东理工大学建筑工程学院 淄博 255000)

摘 要 美国 ELSET 数据库提供的 TLE 数据是目前使用最广泛的数据,在热层大气密度反演、弹道系数估计、碰撞预警等领域具有重要作用。受空间环境扰动、空间事件以及 TLE 产生过程等共同影响,ELSET 数据库包含大量亟待清理的异常值和识别的空间事件,例如发布错误的 TLE、轨道根数异常和 Bstar 异常。现有方法在清理异常轨道根数时缺乏统一性,需要使用不同的技术,清理流程较为繁杂,并且仅适用于特定轨道区域的少数目标。为克服现有方法的弊端,提出了一种基于期望最大算法的滑动窗口-多项式拟合预报方法,对含有轨道机动的碎片以及受空间环境影响的碎片进行异常值与空间事件探测。研究表明,该方法能够灵活处理不同空间环境下的异常值与空间事件探测,具有普适性,适用于所有轨道碎片。

关键词 空间碎片,两行根数,期望最大,异常值探测

中图分类号 P171

Space Event and Outlier Detection Based on Expectation Maximization Algorithm

LIU Jinghong^{1,2} WU Chenyun¹ XU Jin²

DU Jianli² LEI Xiangxu³

1(Smart City Institute, Chongqing Jiaotong University, Chongqing 400074)

2(Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210023)

3(School of Civil and Architectural Engineering, Shandong University of Technology, Zibo 255000)

Abstract The United States provide Element Sets (ELSET) database in Two-Line Element (TLE) format for public use, which plays an important role in the inversion of atmospheric density in the thermosphere, ballistic coefficient estimation, early-warning and so on. Due to large uncertainties existing in the TLE generation process, space environment changes and space events, ELSET database contains a

* 重庆市自然科学基金面上项目(CSTB2022 NSCQ-MSX1093),重庆市教委科学技术研究项目(KJQN202200701)和中国博士后科学基金项目(2021 M703487)共同资助

2021-11-19 收到原稿,2022-05-12 收到修定稿

E-mail: liu-jh@whu.edu.cn

large number of abnormal TLE data to be filtered, such as corrected TLE, orbital element outlier and Bstar outlier. The existing methods to filter out the outliers lack general applicability and are very complicated, which are only applicable to a few space targets in certain orbit regions. To overcome the shortcomings of the existing methods, a filtering method is proposed based on Expectation Maximization (EM) algorithm employing a sliding window and polynomial fitting method, which can detect outliers for different orbital elements and space events. The research shows that the algorithm can effectively single out the outliers in TLE sequences and is suitable for all orbital debris.

Key words Space debris, Two-line element, Expectation maximization, Outlier detection

0 引言

美国空间监测网(Space Surveillance Network, SSN)在全球布设了 29 个测站,能够连续跟踪观测空间碎片,监测数据通过轨道处理后以两行根数(TLE)的格式进行轨道编目进入 ELSET 数据库,并同时在网站^{*}发布,是可公开使用的种类最全、数量最多的空间碎片轨道编目数据库^[1-3]。TLE 数据在空间活动研究中具有重要作用,例如利用 TLE 数据反演热层大气密度、估计弹道系数、再入预报和碰撞预警等^[4-8]。然而许多研究表明,TLE精度存在较大波动的现象,轨道根数出现离散程度不一的异常值以及与趋势相违背的连续 TLE 数据等^[9,10]。由于 TLE 精度未提供,直接使用 TLE 数据将引起潜在风险或降低后续研究成果的稳定性和可靠性。因此,必须识别 ELSET 数据库中的异常 TLE 数据以及空间事件。

目前,ELSET 数据库中异常 TLE 的产生机理尚未被完全理解,并且不同研究者仅关注特定轨道区域的少部分空间碎片,造成异常数据的清理方法和流程不同。针对低轨碎片和大椭圆轨道碎片,Águeda 等^[11]和 Dolado-Perez 等^[12]将 TLE 轨道根数转化为分点根数,使其变化平滑,采用迭代最小二乘多项式拟合轨道参数的变化趋势,将离散明显的异常值剔除。针对存在轨道机动的卫星,Patera^[13]提出了滑动窗口-多项式拟合方法,探测卫星轨道机动,通过计算轨道参数与其回归值的离散度确定是否存在变轨,然而离散度越小,越难区分轨道机动和空间环境扰动产生的异常值。针对地球同步轨道的火箭体,Lidtke 等^[3]利用滑动窗口和多项式提出了多步清理策略,给出多种异常值阈值的定义,以便清理不同的轨道参数。然而

以上方法无法统一轨道参数异常值的清理流程和算法,适用的空间碎片对象有限,并且在实际处理过程中存在异常值判定不合理等问题。

为克服以上问题,Liu 等^[14]提出了基于期望最大(Expectation Maximization, EM)算法的滑动窗口-多项式拟合预报技术,解决了现有算法存在的关键问题,重新定义了基于高斯统计分布特征的阈值,统一了轨道参数的异常值探测流程,适用于所有轨道区域空间碎片异常值清理。探测结果表明:对于相对误差幅度 $\geq 5 \times 10^{-3}$ 的平运动和偏心率异常值,可以实现 100% 的探测;对于相对误差幅度 ≥ 0.01 的轨道倾角异常值,可实现 74%~98% 的探测;对于相对误差幅度 $\geq 5 \times 10^{-3}$ 的轨道倾角异常值,可实现 37%~68% 的探测。

在实际清理过程中,由于异常值没有真实的参考值,对 TLE 中探测的异常值进行严格评估通常是不可能的,这是所有清理算法常见的问题。本文在上述工作基础上,基于 EM 算法的滑动窗口-多项式拟合预报技术,识别受复杂空间环境及轨道机动影响的空间碎片,给出该方法的清理步骤和最终清理效果。

1 TLE 异常值清理算法原理

TLE 轨道根数异常值清理算法基于极大验后估计得到多项式的拟合-预报表达式,通过引入 EM 算法精确地估计误差的期望和方差,该算法涉及的变量包括阈值、预报值个数和滑动窗口大小(滑动窗口包含的 TLE 数据数量)^[14]。

1.1 基于极大验后估计的线性回归和预报表达式

假设观测值 y 可表示为如下多项式:

^{*}www.space-track.org

$$y = \Phi\theta + \eta. \quad (1)$$

式中,

$$\Phi = \begin{bmatrix} 1 & \cdots & x_1^n \\ \vdots & \ddots & \vdots \\ 1 & \cdots & x_{N_{\text{TLE}}}^n \end{bmatrix},$$

$$\theta = [a_0 \quad \cdots \quad a_n]^T.$$

η 的方差阵为 Σ_η^{-1} , 待估参数 θ 的方差阵为 Σ_θ^{-1} . 假定窗口内 TLE 轨道根数误差相同且相互独立, 那么观测值和待估参数的方差阵可简化为 $\sigma_\eta^2 \mathbf{I}_{N_{\text{TLE}}}$ 和 $\sigma_\theta^2 \mathbf{I}_{n+1}$, 这里 $\mathbf{I}$ 表示单位矩阵。

根据极大验后估计原理^[15], 如果给定一系列观测值, 对未知参数进行估计, 其极大验后估计的期望和方差分别为

$$\mu_{\theta|y} = E[\theta|y] = \theta_0 + \frac{\left(\frac{\mathbf{I}}{\sigma_\theta^2} + \frac{\Phi^T \Phi}{\sigma_\eta^2}\right)^{-1} \Phi^T (y - \Phi\theta_0)}{\sigma_\eta^2}, \quad (2)$$

$$\Sigma_{\theta|y} = \left(\frac{\mathbf{I}}{\sigma_\theta^2} + \frac{\Phi^T \Phi}{\sigma_\eta^2}\right)^{-1}. \quad (3)$$

相应预报值的期望和方差表达式如下:

$$\mu_Y = \varphi^T(x) \mu_{\theta|y}, \quad (4)$$

$$\Sigma_Y = \sigma_\eta^2 + \varphi^T(x) \Sigma_{\theta|y} \varphi(x) = \sigma_\eta^2 + \sigma_\eta^2 \sigma_\theta^2 \varphi^T(x) (\sigma_\eta^2 \mathbf{I} + \sigma_\theta^2 \Phi^T \Phi)^{-1} \varphi(x). \quad (5)$$

令 $\mu = \mu_Y$, $\sigma = \Sigma_Y$, 预报值与实测值之差将有 95% 的概率落入 $[\mu - 2\sigma, \mu + 2\sigma]$ 区间, 99.7% 的概率落入 $[\mu - 3\sigma, \mu + 3\sigma]$ 区间. 因此, 异常值阈值可设定为 2σ 或 3σ . σ_η^2 和 σ_θ^2 的精确值可通过 EM 算法迭代获得^[16].

1.2 异常值及其空间事件探测原理

图 1 给出了基于 EM 算法的异常值及其空间事件探测流程^[14]. 假定滑动窗口内观测值总数始终为 $N_{\text{TLE}} = 9$, 对滑动窗口内的变化趋势进行拟合, 并剔除离散度较大的值(窗口内阈值设定); i 为窗口内最后一个 TLE 观测值, 假定同时预报窗口外 $k = 6$ 个观测值, $i+1, \dots, i+6$ 为窗口外观测值时刻的预报值, 则根据窗口外阈值设定, 预报结果有以下三种情况。

情况 1 预报值与观测值的差全部落入置信区间内, 滑动窗口向前移动 $k = 6$, 继续向前预报。

情况 2 预报值与观测值的差全部落在置信区间之外, 形成新的子序列且滑动窗口以第 $i+1$ 个实测

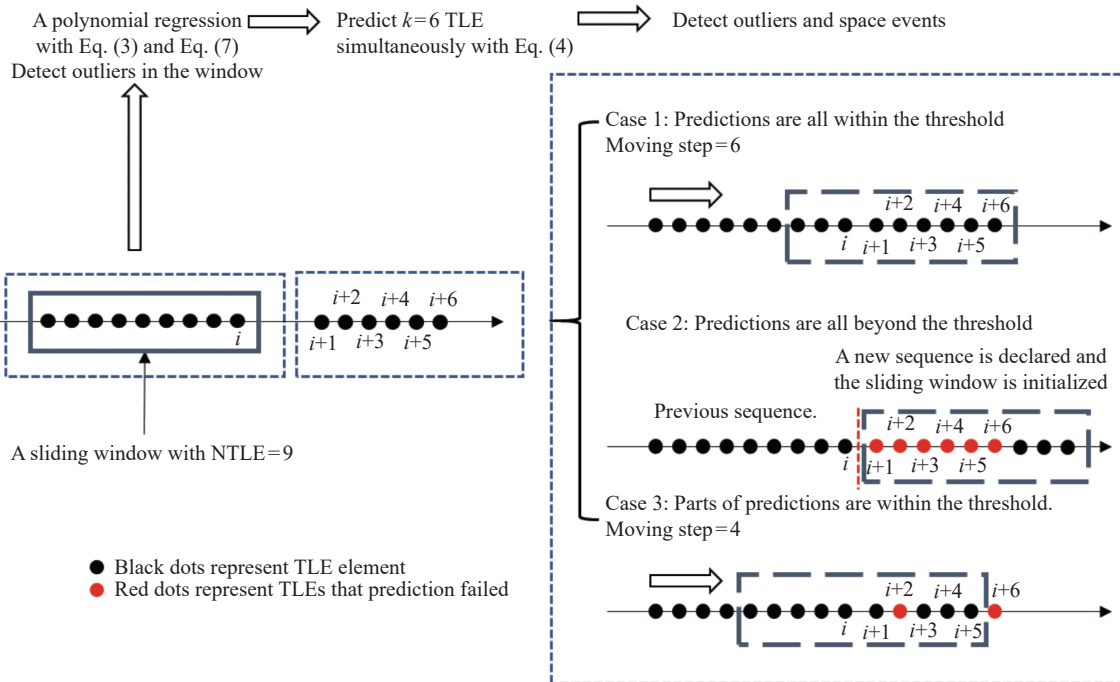


图 1 基于 EM 算法的异常值及空间事件探测流程图 (黑色圆点代表 TLE 观测值, 红色圆点代表预报失败值)

Fig. 1 Flowchart of outlier detection based on EM algorithm (Black dot represents the TLE observation value, and the red dot represents the prediction failure value)

值为起始点,重新初始化滑动窗口。

情况 3 预报值与观测值的差部分落在置信区间之外($i+1$ 和 $i+6$),第 $i+1$ 为异常值,滑动窗口向前移动至第 $i+5$ 个观测值,重新初始化滑动窗口。

通过以上三种情况的区分,可以探测异常值和空间事件,且使用滑动窗口外多个预报值同时进行判断,使得空间事件的探测更加合理。经过以上方式的滑动窗口前移,TLE 序列中明显的异常值被清理,而次级别的异常值可能依然存在于 TLE 序列中,因而整个清理过程需要重复多次。

2 TLE 异常值探测算法应用

2.1 阈值设定策略对异常值探测的影响

Liu 等^[14]讨论了异常值及空间事件探测的影响因素,包括滑动窗口大小、滑动窗口和预报值的阈值设定策略、预报个数以及多项式拟合阶数。研究表明,滑动窗口大小在 60~100 内,预报个数为 8,多项式阶数为 3 时较为合适。阈值策略包括以下三种。

策略 1 滑动窗口内阈值和预报阈值均采用 3σ 。

策略 2 滑动窗口内阈值采用 2σ ,预报阈值采用 3σ 。

策略 3 滑动窗口内阈值和预报阈值均采用 2σ 。

碎片 NORAD ID 为 13025,处理时间范围为 1982 年 1 月 9 日至 1988 年 11 月 20 日,共计 6.9 年,对轨道倾角采用三种阈值策略清理,结果如图 2 所示^[14]。策略 1 适合探测异常值少、变化平滑的 TLE 子序列,产生的异常值数量最少,而部分变化复杂的子序列仍然存在异常值;策略 2 产生的异常值数量适中;策略 3 产生的异常值数量最多,对于变化平滑的子序列,策略 3 过于严格,其对于变化复杂的子序列却很适合。因此,在实际清理过程中,三种策略需交替使用,首先采用策略 1 清理,将变化平滑的子序列分离,然后采用策略 2 或者策略 3 对复杂的子序列做进一步清理。

2.2 TLE 异常值清理流程

针对轨道参数(平运动、偏心率、轨道倾角和 Bstar)中存在的异常值,介绍异常值清理算法的实际步骤,并展示部分碎片清理结果。在无特殊说明时,清理算法涉及的参数预报个数默认为 8,多项式阶数为 3,其他参数如滑动窗口大小、阈值策略,根据实际情况进行选择。选用卫星 HAIYANG 2 A(NORAD ID: 37781),该卫星存在轨道机动,且经历的空间环境

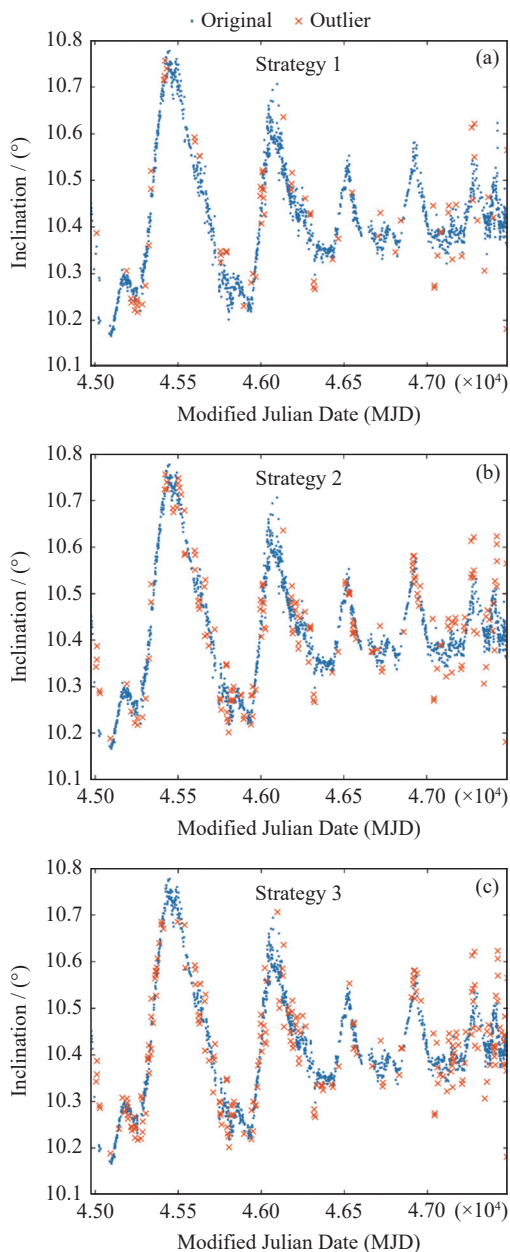


图 2 采用三种阈值策略清理轨道倾角,碎片 NORAD ID13025 (Ariane 1 R/B)

Fig. 2 Inclination outlier detection by three strategies for NORAD ID13025 (Ariane 1 R/B)

包括复杂和平滑变化时期,适用于对异常值清理算法的实际清理流程进行说明,处理的 TLE 数据时间范围为 2018 年 1 月 1 日至 2021 年 9 月 8 日。清理步骤如下。

第 1 步 清理平运动序列中的异常值,滑动窗口大小为 200,阈值采用策略 1,清理 1 次,序列中存在的显著异常值完全被清理,清理结果如图 3 所示。判

定的 TLE 子序列起始点和终止点, 用于辅助清理 TLE 中的异常值。

第 2 步 清理偏心率序列中的异常值, 滑动窗口大小为 200, 阈值采用策略 1, 连续清理 2 次。清理结果见图 4, 共计产生 16 个子序列(编号 1~16), 其中

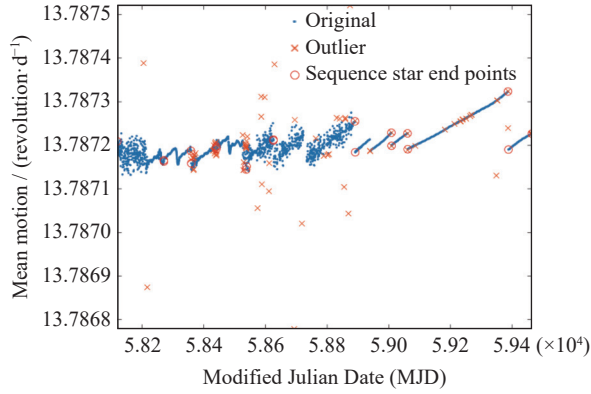


图 3 平运动序列中的异常值清理结果

Fig. 3 Outlier detection in mean motion

第 1, 3, 4, 6 和 14 子序列存在明显异常值, 应进一步对其进行异常值清理。子序列 1 包括 406 个观测值, 采用策略 1, 滑动窗口大小分别用 40 和 80 各清理

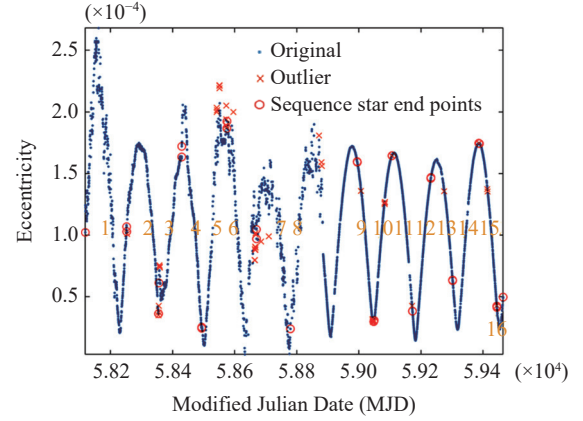


图 4 偏心率序列中的异常值初步清理结果

Fig. 4 Preliminary results of outlier detection in eccentricity series

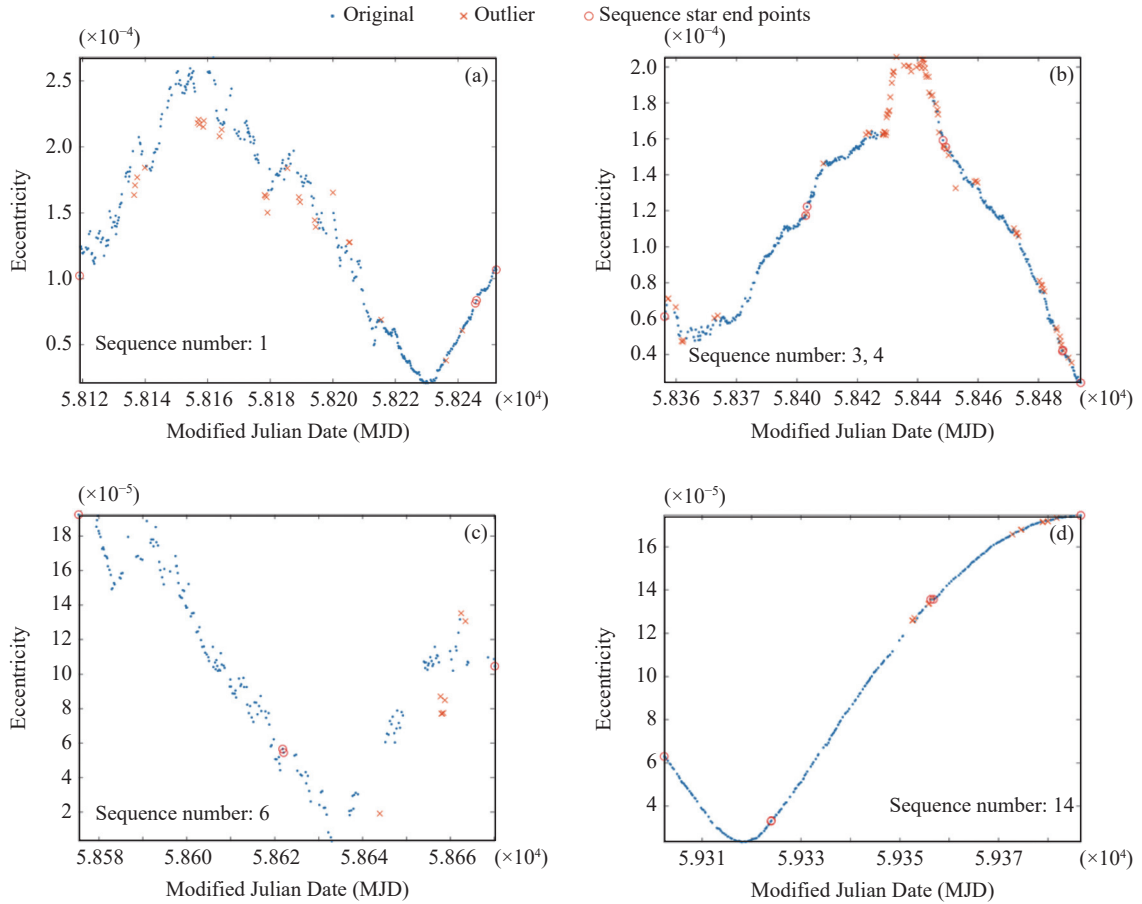


图 5 各个子序列的异常值清理结果

Fig. 5 Outlier detection results of each subsequence

1次。子序列3和4合并为一个子序列,共包括467个观测值,采用策略2,滑动窗口大小为100,连续清理2次。子序列6包括200个观测值,采用策略1,滑动窗口大小为40,清理1次。子序列14包括289个观测值,采用策略1,滑动窗口大小为40,清理1次。各个子序列清理的异常值结果见图5,偏心率异常值最终的清理结果见图6。

第3步 清理轨道倾角序列中的异常值。轨道倾角序列中存在大量多个(8~30)连续观测值近乎水平直线的分布。采用2阶多项式拟合,窗口大小20进行清理,异常值清理结果如图7所示。

第4步 清理Bstar序列中的异常值,滑动窗口大小为80,阈值采用策略1。清理结果如图8所示,由于Bstar变化没有规律,仅清理了3个异常值明显

的观测值。

经过以上步骤,TLE序列中的异常值得以清理。在清理过程中需注意以下几方面问题。

(1)异常值存在复杂性,清理一次未必能成功,需重复多次。

(2)对轨道根数初次清理时,滑动窗口大小宜选用较大值(100~200),便于拟合长期趋势变化,随后根据判定的子序列进一步清理。

(3)子序列的观测值数量一般较少,滑动窗口大小宜选用较小值(20~40),若子序列中出现异常值明显(子序列6)或部分趋势相违背(子序列3,4)的情况,宜选用较大值(100~200)。

(4)针对变化平滑的子序列,阈值采用策略1,而变化稍微复杂的子序列需要采用策略2。

(5)针对变化特殊的子序列,例如轨道倾角,其多项式阶数和窗口阈值需作调整。

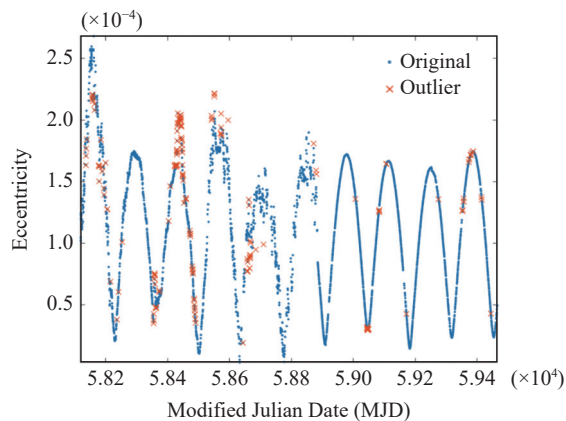


图6 偏心率序列中的异常值最终清理结果
Fig. 6 Final results of outlier detection in eccentricity series

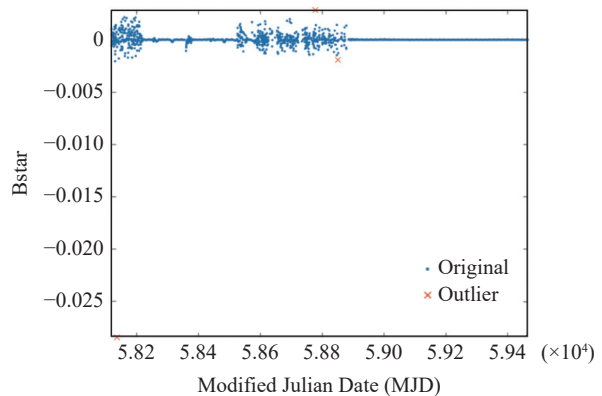


图8 Bstar序列中的异常值清理结果
Fig. 8 Results of outlier detection in Bstar

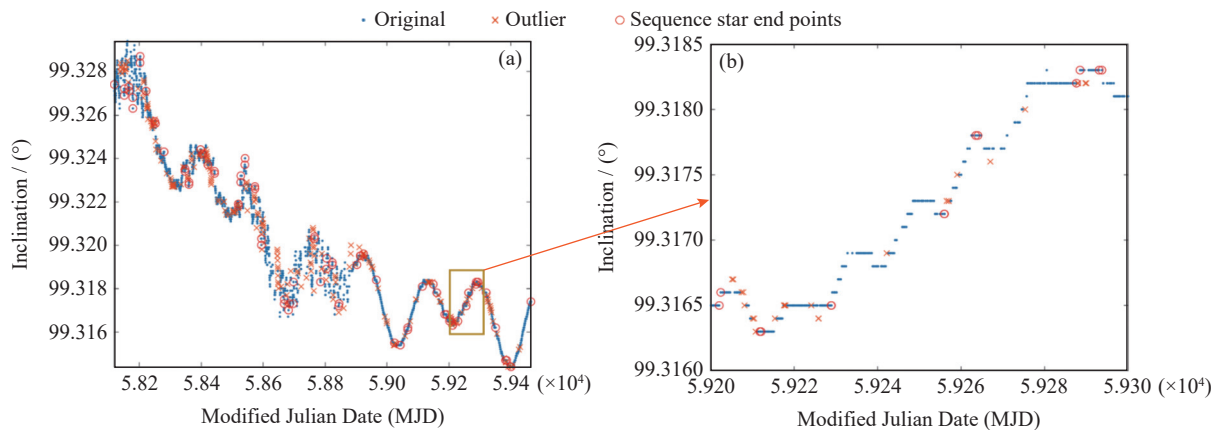


图7 轨道倾角序列中的异常值清理结果
Fig. 7 Results of outlier detection in inclination

2.3 TLE 异常值清理结果

从 ELSET 数据库中选取 4 个碎片, 展示异常值清理结果(见图 9)。从图 9 可知, TLE 序列中并非所有轨道根数都包含异常值, 碎片 27386 是卫星 Envisat, 存在轨道机动, 所探测的异常值数量最多, 其次是卫星 O3BFM9(NORAD ID: 40351)。碎片 12446 是 TITAN 34 B 的火箭体残骸, 异常值主要来源于轨道倾角, 碎片 23358 是 USA 40 R/B 的碎片, 雷达反射面积等级为 SMALL, 所探测的异常值数量最少。

3 结论

针对 TLE 序列中存在的大量异常值, 提出一种基于期望最大算法的滑动窗口-多项式拟合预报方法, 给出清理步骤及部分碎片清理结果, 得到结论如下。

(1) 对于长周期 TLE 序列的清理, 滑动窗口大小应尽量大(100~200), 阈值采用策略 1, 实现异常值的清理和子序列的判定, 并进一步对子序列进行逐个

清理。

(2) 对于短期 TLE 序列的清理, 由于观测值总量较少, 滑动窗口大小宜小(15~40)。

(3) 若 TLE 序列包含与趋势变化相违背的明显的连续多个异常值, 采用滑动窗口 100~200 和阈值策略 1 将其清理。

(4) 对于变化平缓的 TLE 序列, 阈值采用策略 1; 对于变化稍微复杂的 TLE 序列, 阈值采用策略 2。

(5) 由于异常值的存在影响多项式的拟合与预报, 整个过程需要反复清理, 降低异常值的影响, 一般 1~3 次。

由于 TLE 异常值产生的原因复杂, 在实际清理中需结合 TLE 序列的变化特点, 合理选择参数组合, 实现清理结果最优。为便于清理, 在基于 EM 算法的滑动窗口-多项式拟合预报技术基础上开发了可视化异常清理软件, 可通过重复性清理实验找到最佳清理参数设置和最佳清理效果。作者利用卫星(ID:36985) 2011 年的 TLE 数据进行轨道机动探测, 初步发现异常清理后, 能显著降低其他因素的干扰, 提升轨道机

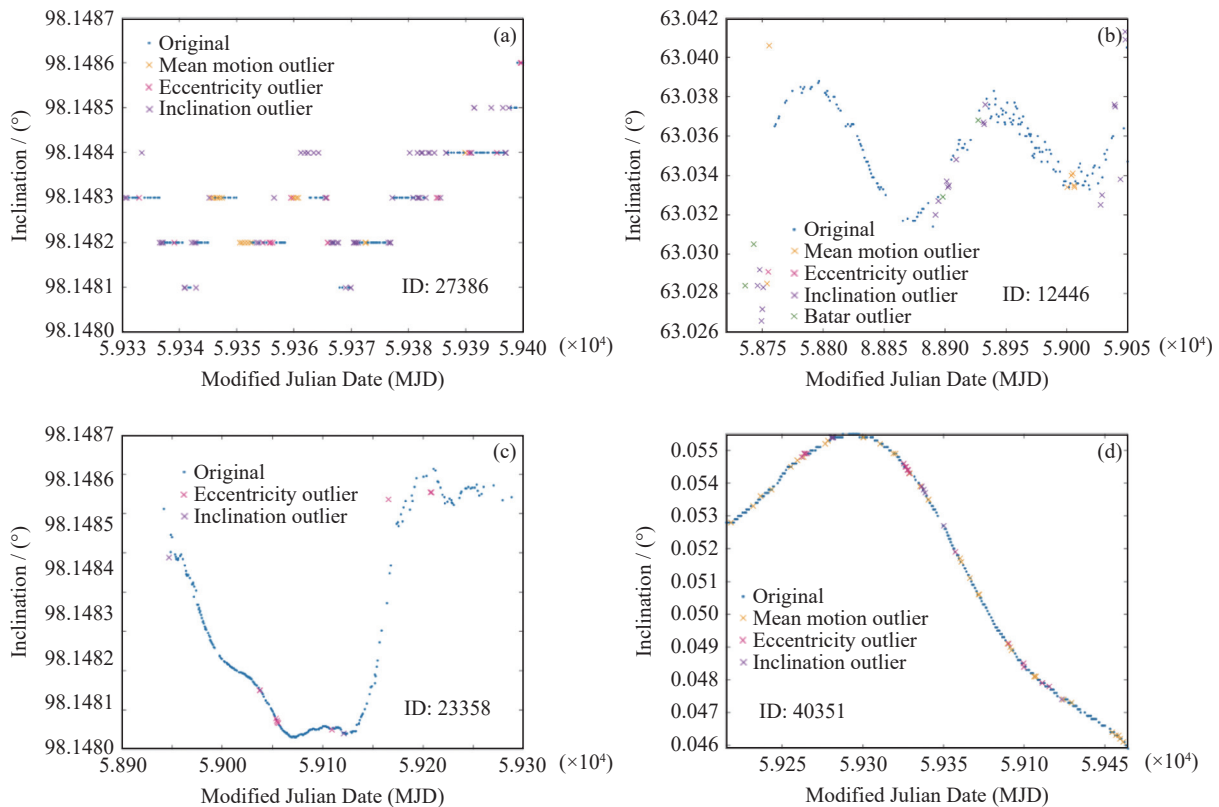


图 9 空间碎片 12446, 23358, 27386 和 40351 轨道倾角异常值清理结果

Fig. 9 Outlier detection of inclination for debris 12446, 23358, 27386 and 40351

动探测成功率,更易识别小推力轨道机动。因此,后续将在已有研究基础上对观测值实时异常判定及轨道机动探测开展更深入的研究。

参考文献

- [1] MULDOON A, ELKAIM G, WEEDEN B, *et al.* Improved orbital debris trajectory estimation based on sequential TLE Processing[C]//60th International Astronautical Congress. Daejeon, Korea, 2009
- [2] VALLADO D A. Fundamentals of Astrodynamics and Applications[M]. 4th ed. California: Microcosm Press, 2013
- [3] LIDTKE A A, GONDELACH D J, ARMELLIN R. Optimising filtering of two-line element sets to increase re-entry prediction accuracy for GTO objects[J]. *Advances in Space Research*, 2019, **63**(3): 1289-1317
- [4] OSWEILER V P. Covariance Estimation and Autocorrelation of NORAD Two-Line Element Sets[D]. Wright-Patterson: Air Force Institute of Technology, 2006
- [5] LEGENDRE P, DEGUINE B, GARMIER R, *et al.* Two line element accuracy assessment based on a mixture of Gaussian laws[C]//AIAA/AAS Astrodynamics Specialist Conference and Exhibit. Keystone, Colorado: AIAA, 2006: 6518
- [6] VALLADO D A, CEFOLA P J. Two-line element sets-Practice and use[C]//63rd International Astronautical Congress. Naples, Italy: International Astronautical Federation, 2012: 29-44
- [7] SAUNDERS A, SWINERD G G, LEWIS H G. Deriving accurate satellite ballistic coefficients from two-line element data[J]. *Journal of Spacecraft and Rockets*, 2012, **49**(1): 175-184
- [8] GEUL J, MOOIJ E, NOOMEN R. GOCE statistical re-entry predictions[C]//Proceedings of 7th European Conference on Space Debris. Darmstadt, Germany: The ESA Space Debris Office, 2017
- [9] SONG W D, WANG R L, WANG J. A simple and valid analysis method for orbit anomaly detection[J]. *Advances in Space Research*, 2012, **49**(2): 386-391
- [10] GONDELACH D, LIDTKE A A, ARMELLIN R, *et al.* Re-entry prediction of spent rocket bodies in GTO[C]//26th AAS/AIAA Space Flight Mechanics Meeting. 2016
- [11] ÁGÜEDA A, AIVAR L, TIRADO J, *et al.* In-orbit lifetime prediction for LEO and HEO based on orbit determination from TLE data[C]//6th European Conference on Space Debris. Darmstadt, Germany, 2013: 22-25
- [12] DOLADO J C, AIVAR-GARCIA L, MATÉ A A, *et al.* OPERA: a tool for lifetime prediction based on orbit determination from TLE data[C]//24th International Symposium on Space Flight Dynamics. Laurel, MD, USA, 2004
- [13] PATERA R P. Space event detection method[J]. *Journal of Spacecraft and Rockets*, 2008, **45**(3): 554-559
- [14] LIU J H, LIU L, DU J L, *et al.* TLE outlier detection based on expectation maximization algorithm[J]. *Advances in Space Research*, 2021, **68**(7): 2695-2712
- [15] CUI Xizhang, YU Zongchou, TAO Benzao, *et al.* Generalized Survey Adjustment[M]. 2nd ed. Wuhan: Wuhan University Press, 2009 (崔希璋, 於宗俦, 陶本藻, 等. 广义测量平差[M]. 2版. 武汉: 武汉大学出版社, 2009)
- [16] LIU Jinghong. Prediction of Atmospheric Reentry of Space Objects Based on TLE Data[D]. Wuhan: Wuhan University, 2020 (刘劲宏. 基于TLE数据的空间目标大气再入预报[D]. 武汉: 武汉大学, 2020)