



基于ARIMA模型的卫星钟差异常值探测的模型选择方法

马朝忠^{1,2} 朱建青³ 韩松辉¹

1 信息工程大学基础部,河南 郑州,450001

2 信息工程大学地理空间信息学院,河南 郑州,450001

3 苏州科技大学数理学院,江苏 苏州,215009

摘要:由于各种不确定因素的干扰,人们获取的卫星钟差数据中经常会出现异常扰动,降低了卫星钟性能分析的可靠性,破坏了钟差建模和预报的有效性,影响了导航定位结果的精准度。对此,以求和自回归移动平均模型为基础,建立了钟差时间序列异常值探测模型;基于Bayes统计原理,将异常值的定位和定值问题转化为模型选择问题;通过模型后验概率的近似计算,构建了模型选择的度量标准,避免了复杂的迭代计算问题。通过全球定位系统和北斗导航卫星系统不同卫星钟差数据的仿真试验,验证了所提出的方法对于卫星钟差序列中异常影响的定位和定值的正确性和有效性。

关键词:卫星钟差;异常值;Bayes原理;求和自回归移动平均模型

中图分类号:P228

文献标志码:A

有效可靠的卫星钟差数据是进行星载原子钟性能分析和卫星钟差建模与预报的前提和基础^[1],也是卫星导航定位系统实现精准定位的关键。然而,由于原子钟本身性能、空间环境的影响和各种不确定因素的干扰以及有计划的卫星钟调相调频操作,所获取的卫星钟差数据中经常会出现异常扰动,从而降低了卫星钟性能分析的可靠性,破坏了钟差建模和预报的有效性,影响了导航定位结果的精准度。因此,需要对这些异常值进行处理。

近年来,已有不少学者对钟差异常值探测的方法开展了研究^[1-19]。目前常见的钟差异常值探测方法大致可以分为两类:一是通过比对直接探测观测数据中的异常值,如中位数法^[1-6]、Allan方差法及其扩展方法^[3-11];二是针对观测数据首先建立相应模型,然后依托模型进行异常值探测,如基于二次多项式模型的递推遗忘因子最小二乘法^[12-13]、基于时间序列自回归(autoregressive, AR)模型的Bayes方法等^[14-19]。这些方法能够在一定程度上实现对异常值的探测,但也存在不足。直接法虽然简单方便,但对异常值的大小不敏感,易出现漏判现象;基于各种模型进行异常

值探测的方法已逐渐为广大学者所认可,但基本上都是基于平稳序列进行研究的,且大部分方法都需要通过数值迭代来实现,不可避免地会受到迭代发散问题的困扰;一些方法还需要进行多重假设检验以及阈值确定等。这些问题的存在都增加了异常值探测的不确定性。

卫星钟差数据序列是一个非平稳时间序列,在长时间段的卫星钟差数据中存在着较为显著的数据异常^[1],这些异常数据主要包括相位跳变、频率跳变、数据间断和数据粗差等^[1-6]。依据时间序列分析的观点,除频率跳变外,数据间断、数据粗差、相位跳变等形成的异常值都可归结为加性异常值(additive outlier, AO)^[5,15,20-21]。基于以上分析,本文基于求和自回归移动平均(autoregressive integrated move average, ARIMA)模型^[20-21]建立异常值探测模型,通过模型选择的方法对卫星钟差数据中的AO类异常值进行探测与处理。假定卫星钟差观测序列中含有多个异常值,针对异常值的不同情况建立相应的异常值探测模型;构造用于异常值探测的候选模型集合,从而将异常值的定位、定值问题转化为模型选择问题;基于Bayes统计原理确定各个候选模型的后验概

率,导出模型选择的度量指标;选择指标值最大的候选模型为期望模型,以实现异常值的定位和定值。通过GPS和北斗导航卫星系统(BeiDou navigation satellite system, BDS)不同卫星钟差异异常值的探测及钟差预报等试验,表明基于ARIMA模型的卫星钟差异异常值探测的模型选择方法能同时解决异常值的定位和异常扰动的定值问题,验证了方法的正确性和有效性。

1 异常值探测模型

若有卫星钟差序列 $\{x_t\}$ 符合如下形式,则称其满足ARIMA(p, d, q)模型^[20-21]:

$$\begin{cases} \phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ \varepsilon_t \text{i.i.d.} N(0, \sigma^2) \end{cases} \quad (1)$$

式中, B 为后移算子,如 $B^k x_t = x_{t-k}$; $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ 和 $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ 分别为 B 的 p 阶和 q 阶多项式,它们的零点都在单位圆外; $\nabla = 1 - B$ 为差分算子; d 为差分次数;经过 d 次差分运算后, $\nabla^d x_t$ 为平稳时间序列; $\varepsilon_t \text{i.i.d.} N(0, \sigma^2)$ 表示 $\varepsilon_t (t = 1, 2, \dots)$ 为相互独立且同分布于均值为零、方差为 σ^2 的正态分布。

对于一组卫星钟差观测数据 y_1, y_2, \dots, y_T ,若其中含有 m 个AO类异常值,建立异常值探测模型 M_{τ_m} :

$$\begin{cases} y_t = x_t + w_{i_1} I_t^{(t_1)} + \dots + w_{i_m} I_t^{(t_m)} \\ \varphi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ \varepsilon_t \text{i.i.d.} N(0, \sigma^2) \end{cases} \quad (2)$$

式中, x_t 为不含任何异常值的纯净时间序列,且满足式(1); $\tau_m = [t_1 t_2 \dots t_m]^T$ 为异常值位置指示向量,即对于 τ_m 的任意一个分量 h ,如果 $t = h$,则 $I_t^{(h)} = 1$,否则 $I_t^{(h)} = 0$; $w_{i_1}, w_{i_2}, \dots, w_{i_m}$ 表示相应位置处异常扰动的大小; $\rho_{\tau_m} = (\varphi_1 \varphi_2 \dots \varphi_p, \theta_1 \theta_2 \dots \theta_q, w_{i_1} w_{i_2} \dots w_{i_m}, \sigma)$ 为模型参数; $p_m = p + q + m + 1$ 为参数个数。异常值探测就是找出观测序列中异常值的位置并计算其扰动大小,即确定 $\tau_m = [t_1 t_2 \dots t_m]^T$ 和 $w_{i_1}, w_{i_2}, \dots, w_{i_m}$ 。

2 异常值探测原理

常见的基于模型的异常值探测方法均需要通过多次迭代来实现。如基于AR模型的Bayes方法在计算时不得不考虑迭代的收敛问题,有时

还需进行多重假设检验;基于二次多项式模型的方法存在临界值选取困难等问题。在此借鉴模型选择的思想^[22-23],从整体上考虑异常值数量 m 、位置 $\tau_m = [t_1 t_2 \dots t_m]^T$ 和扰动大小 $w_{i_1}, w_{i_2}, \dots, w_{i_m}$ 同时确定的问题。

首先,设用于异常值探测的候选模型集合为 $M = \{M_{\tau_0}, M_{\tau_1}, \dots, M_{\tau_m}\}$,其中 $M_{\tau_i} (i = 0, 1, 2, \dots, m)$ 表示观测序列中含有 i 个异常值的探测模型, m 表示观测序列中可能存在异常值的最大个数。因此,候选模型的总数为:

$$S = \binom{T}{0} + \binom{T}{1} + \dots + \binom{T}{m} \quad (3)$$

式中, $\binom{T}{i}$ 表示从 T 个观测值中选择 i 个异常值的组合数, $i = 0, 1, 2, \dots, m$ 。

其次,根据Bayes统计推断原理,计算含有 i 个异常值的候选模型的后验概率:

$$P(M_{\tau_i} | \mathbf{y}) = \frac{P(M_{\tau_i})L(\mathbf{y} | M_{\tau_i})}{f(\mathbf{y})} \quad (4)$$

式中, $i = 0, 1, 2, \dots, m$; $P(M_{\tau_i})$ 是候选模型 M_{τ_i} 的先验分布; $L(\mathbf{y} | M_{\tau_i})$ 为在给定候选模型 M_{τ_i} 条件下的边缘似然函数; $\mathbf{y} = [y_1 y_2 \dots y_T]^T$; $f(\mathbf{y}) = \sum_{j=0}^m \sum_{\tau_j} P(M_{\tau_j})L(\mathbf{y} | M_{\tau_j})$ 是 \mathbf{y} 的无条件似然函数。

最后,根据式(4)的结果选择后验概率最大的模型作为异常值探测模型;当模型确定后,异常值的个数、位置和异常扰动的大小同时确定。由此,将异常值的探测问题转化为模型的选择问题。

3 后验概率的计算方法

由式(4)可以看出,后验概率的计算主要是确定模型的先验分布 $P(M_{\tau_i})$ 和给定模型条件下的边缘似然函数 $L(\mathbf{y} | M_{\tau_i})$ 。在模型未确定之前,每个候选模型的地位相当,因此采用均匀分布作为模型的先验分布,即 $P(M_{\tau_i}) = \frac{1}{m+1}$,其中 $c = \binom{T}{i}$ 。边缘似然函数 $L(\mathbf{y} | M_{\tau_i})$ 可以表示为:

$$L(\mathbf{y} | M_{\tau_i}) = \int L(\rho_{\tau_i} | \mathbf{y}, M_{\tau_i}) p(\rho_{\tau_i} | M_{\tau_i}) d\rho_{\tau_i} \quad (5)$$

式中, $L(\rho_{\tau_i} | \mathbf{y}, M_{\tau_i})$ 为观测序列的似然函数; $p(\rho_{\tau_i} | M_{\tau_i})$ 为给定模型 M_{τ_i} 条件下参数 ρ_{τ_i} 的条件先验概率。鉴于边缘似然函数 $L(\mathbf{y} | M_{\tau_i})$ 的解析表达式难以通过直接计算得到,故采取近似计算。对

似然函数 $L(\boldsymbol{\rho}_{\tau_i}|\mathbf{y}, M_{\tau_i})$ 取对数, 令:

$$l_{\tau_i}(\boldsymbol{\rho}_{\tau_i}) = \ln L(\boldsymbol{\rho}_{\tau_i}|\mathbf{y}, M_{\tau_i})$$

在参数 $\boldsymbol{\rho}_{\tau_i}$ 的精确最大似然估计值 $\hat{\boldsymbol{\rho}}_{\tau_i}$ 处, 将对

$$L(\mathbf{y}|M_{\tau_i}) = \left(\frac{2\pi}{T}\right)^{\frac{p_i}{2}} \exp(l_{\tau_i}(\hat{\boldsymbol{\rho}}_{\tau_i})) p(\hat{\boldsymbol{\rho}}_{\tau_i}|M_{\tau_i}) |H_{\tau_i}(\hat{\boldsymbol{\rho}}_{\tau_i})|^{-\frac{1}{2}} + O\left(T^{-\frac{p_i}{2}-1}\right) \quad (6)$$

式中, p_i 表示参数个数; $p(\hat{\boldsymbol{\rho}}_{\tau_i}|M_{\tau_i})$ 表示参数 $\hat{\boldsymbol{\rho}}_{\tau_i}$ 的条件先验概率; $H_{\tau_i}(\hat{\boldsymbol{\rho}}_{\tau_i})$ 表示 $T^{-1}l_{\tau_i}(\hat{\boldsymbol{\rho}}_{\tau_i})$ 的 Hessian

$$P(M_{\tau_i}|\mathbf{y}) = \frac{P(M_{\tau_i})}{f(\mathbf{y})} \left(\frac{2\pi}{T}\right)^{\frac{p_i}{2}} \exp(l_{\tau_i}(\hat{\boldsymbol{\rho}}_{\tau_i})) p(\hat{\boldsymbol{\rho}}_{\tau_i}|M_{\tau_i}) |H_{\tau_i}(\hat{\boldsymbol{\rho}}_{\tau_i})|^{-\frac{1}{2}} + O\left(T^{-\frac{p_i}{2}-1}\right) \quad (7)$$

从理论研究的角度来看, 采用后验概率 $P(M_{\tau_i}|\mathbf{y})$ 的数值最大的模型作为最终选择的模型是合适的; 在实际应用中, 顾及计算时间和计算

$$\ln(P(M_{\tau_i}|\mathbf{y})) = l_{\tau_i}(\hat{\boldsymbol{\rho}}_{\tau_i}) + \frac{p_i}{2} \ln\left(\frac{2\pi}{T}\right) + \ln(p(\hat{\boldsymbol{\rho}}_{\tau_i}|M_{\tau_i})) - \frac{1}{2} \ln(|H_{\tau_i}(\hat{\boldsymbol{\rho}}_{\tau_i})|) - \ln(f(\mathbf{y})) + O(T^{-1}) \quad (8)$$

由贝叶斯信息准则的推导过程可知, 式(8)的主要部分为其前两项, 其中第 1 项反映的是模型拟合的好坏, 第 2 项反映的是模型参数的多少^[22-23]。考虑到异常扰动项的加入会使参数增加, 为限制参数过度化产生的影响, 引入惩罚因子 $\ln\binom{m}{i}$, 同时为了使表达式的系数为整数, 式(8)的两端同时乘以 2, 得到可用于模型选择的度量指标(measure of detecting outliers, MDO):

$$\text{MDO} = 2l_{\tau_i}(\hat{\boldsymbol{\rho}}_{\tau_i}) - p_i \ln(T) - 2\ln\binom{m}{i} \quad (9)$$

使 MDO 值最大的候选模型即为期望模型。

由式(3)易见, 随着数据量 T 的增加, 候选模型总数 S 将会迅速增长, 使模型选择变得比较复杂。为降低计算复杂度, 可采用两阶段法实施计算。第 1 阶段即假设数据中只有 1 个异常值, 用逐步型法筛选出一部分 MDO 值较大的观测值标记为疑似异常值, 通常疑似异常值数量小于数据总数的 5%; 第 2 阶段是在已选出的疑似异常值的基础上, 再进一步计算其组合的 MDO 值, 从而实现快速异常值探测。

4 算例与分析

为了验证本文提出的基于模型选择的卫星钟异常值探测法的正确性和有效性, 采用国际 GNSS 服务官网发布的 GPS 卫星和 BDS 卫星精密钟差数据进行试验和分析。

4.1 方案 1

提取 G31 卫星钟为期 30 d 的钟差数据, 时间

数似然函数 $l_{\tau_i}(\boldsymbol{\rho}_{\tau_i})$ 进行二阶展开, 从而导出式(5)的 Laplace 近似公式^[22], 故式(5)又可以表示如下:

矩阵; $O(\cdot)$ 表示高阶无穷小。式(4)又可表示如下:

量等问题, 对后验概率 $P(M_{\tau_i}|\mathbf{y})$ 的形式进行进一步简化。对式(7)两端分别取对数, 得到如下公式:

为 2017-09-16 00:00 至 2017-10-15 23:55, 采样间隔为 15 min, 共计 2 880 个历元。从中截取一段为期 3 d 且质量好、无显著粗差的数据, 其中前 88 个数据用于拟合钟差模型, 后 200 个数据用于进行异常值探测试验。在后 200 个历元的第 31、121 和 187 个观测值上分别加上大小为 $3 \mu\text{s}$ 、 $-3 \mu\text{s}$ 和 $6 \mu\text{s}$ 的单个异常扰动。采用本文提出的模型选择方法进行异常值探测, 各种情形下相应的 MDO 值列于表 1。其中, 第 2 列($i=0$)最后一行代表数据中不含异常值时的 MDO 值; 第 3~5 列($i=1$)最后一行代表数据中含有 1 个异常值时最大的 3 个 MDO 值(异常值位于其他位置时, MDO 值均小于 -3.2), 相应的 τ_i 为 3 个异常值出现的位置, w_i 为对应异常扰动值大小; 第 6~8 列($i=2$)最后一行表示数据中含有 2 个异常值时最大的 3 个 MDO 值(异常值位于其他位置时, MDO 值均小于 -2.8), 相应的 τ_i 为 3 对异常值出现的位置, w_i 为对应位置处异常扰动值大小; 第 9 列($i=3$)最后一行表示数据中含有 3 个异常值时的最大 MDO 值(异常值位于其他位置时, MDO 值均小于 20)。

从表 1 可以看出: (1) 该钟差序列的第 31、121 和 187 个观测值为异常值, 异常扰动大小分别为 $2.974 5 \mu\text{s}$ 、 $-2.826 3 \mu\text{s}$ 和 $6.040 1 \mu\text{s}$, 与所加的异常扰动比较接近。(2) 对于固定的异常值数量 i , 候选模型的 MDO 值受到异常值大小和位置的影响。(3) 对于不同的 i , MDO 值既与异常值的数量、大小和位置有关, 也和惩罚因子有关。(4) 最终探测模型的 MDO 值明显大于其他模型, 说明所提方法能较为准确地探测到异常值的位置, 并确定异常扰动的大小。

表1 不同异常值数量的异常扰动定位与定值结果

Tab.1 Locations and Magnitudes of Abnormal Disturbances with Different Numbers of Outliers

异常扰动项	$i=0$		$i=1$		$i=2$		$i=3$	
τ_i	31	121	187	(31, 121)	(31, 187)	(121, 187)	(31, 121, 187)	
$w_i/\mu s$	2.7452	-1.5483	10.4219	2.9609, -2.9734	2.8406, 6.0542	-3.4301, 5.8417	2.9745, -2.8263, 6.0401	
MDO	-1.3377	-2.7610	-2.7565	-1.8445	-2.5353	-1.1496	-1.1433	22.0521

4.2 方案2

基于上述30 d的钟差数据进行异常值探测和钟差预报试验分析。首先对前100个历元的试验数据进行异常值探测并修正钟差序列预报模型,然后对后续2780个历元的钟差进行预报。试验结果分析如下:(1)使用本文提出的基于模型选择的异常值探测方法对前100个数据进行探测时,发现其中存在4处异常,分别位于第5、6、7和38个历元,异常扰动分别为130 ns、120 ns、300 ns和-100 ns。采用基于AR模型的Bayes方法^[14-16]只能探测到第7个历元和第38个历元处的数据受到异常影响,异常扰动分别为180 ns和99 ns。这表明基于ARIMA模型的卫星钟差异常值探测的模型选择方法对于成片异常值的探测也有一定作用。(2)将数据修正后应用于后续钟差预报,并将预报结果与目前广泛使用的抗差二次多项式的预报结果和新近发展的基于AR模型的Bayes方法的预报结果分别进行比较,比较的指标为预测均方根误差(root mean square error of prediction, RMSEP)、平均误差(Mean)和最大偏差绝对值(maximum absolute biases, MAB),结果列于表2。比较发现,在这3个指标下,基于AR模型的Bayes方法的预报精度优于抗差二次多项式模型的预报精度,而基于ARIMA模型的模型选择方法的预报精度又优于基于AR模型的Bayes方法的预报精度。

表2 3种方法的预报精度/ns

Tab.2 Prediction Precisions of Three Methods/ns

预报方法	RMSEP	Mean	MAB
基于AR模型的Bayes方法	0.94	90.2	294
抗差二次多项式法	1.34	103.8	477
模型选择法	0.50	75.8	185

4.3 方案3

为进一步验证所提方法的有效性,采用北斗不同类型卫星的精密钟差数据进行试验分析。选用北斗地球同步轨道(geostationary earth orbit, GEO)卫星C02、倾斜地球同步轨道(inclined geosynchronous satellite orbit, IGSO)卫星C10和

地球轨道(medium earth orbit, MEO)卫星C14的钟差数据,时间从2018-09-09 00:00至2018-09-10 23:55,采样间隔为5 min。在3颗卫星的第一天精密钟差数据中分别随机抽取3个历元的数据加入粗差,以此为背景建立模型,对3颗卫星后一天的钟差分别进行预报。预报结果与实测值的误差分布见图1。可以看出,本文提出的基于ARIMA模型的模型选择方法的预报精度比基于AR模型的Bayes方法的预报精度和采用抗差二次多项式模型的预报精度都高。

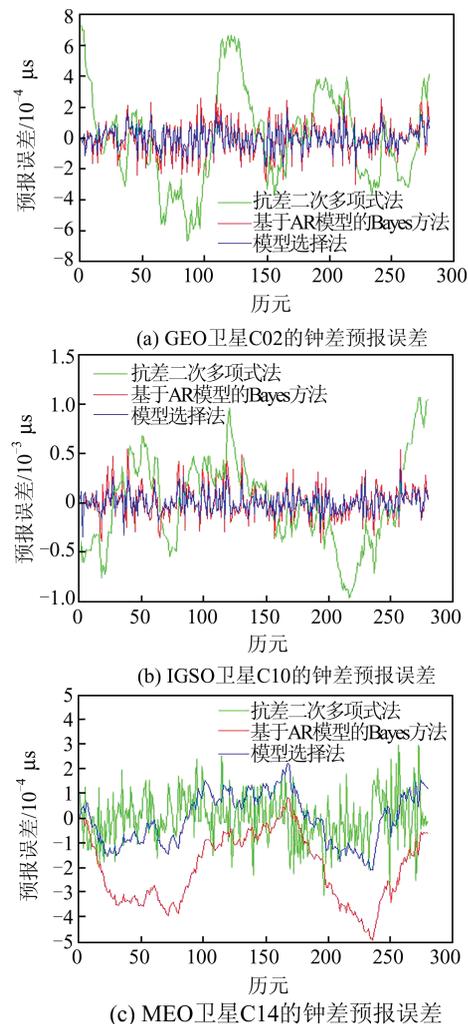


图1 不同卫星3种方法的预报误差

Fig1 Prediction Errors of Three Methods with Different Satellites

5 结 语

卫星钟异常值探测不仅有助于提高定位和时频传递精度,也对卫星完好性监测提供了有力支持。因此,寻求一种高效的钟差序列异常值探测方法显得尤为重要。本文构建了一种异常值探测新模型,提出了基于 ARIMA 模型的钟差异常值探测的模型选择法;以 Bayes 统计原理为理论基础,导出模型的后验分布,将异常值的探测问题转化为候选模型的选择问题;推导出可用于模型选择的度量指标 MDO,解决了后验概率计算复杂的实际问题;针对数据量 T 的增加会使候选模型总数 S 迅速增长,从而导致模型选择变得比较复杂的新情况,提出了两阶段法实施计算,以降低计算复杂度;通过 GPS 和 BDS 卫星钟差数据异常值探测和钟差预报等仿真试验分析,验证了所提方法的正确性和有效性。

参 考 文 献

- [1] Wang Yupu. Research on Modeling and Prediction of the Satellite Clock Bias and Performance Evaluation of GNSS Satellite Clocks[D]. Zhengzhou: Information Engineering University, 2017(王宇谱. GNSS 星载原子性能分析与卫星钟差建模预报研究[D]. 郑州:信息工程大学,2017)
- [2] Winkler G M R. Introduction to Robust Statistics and Data Filtering [OL]. <http://www.stable32.com/robstat.htm#INTRODUCTION>, 1993
- [3] Riley W J. The Calculation of Time Domain Frequency Stability [M/OL]. <http://www.stable32.com/paper1ht.htm>, 2002
- [4] Riley W J. Handbook of Frequency Stability Analysis [M/OL]. <http://www.stable32.com/Handbook.pdf>, 2007
- [5] Guo Hairong. Study on the Analysis Theories and Algorithms of the Time and Frequency Characterization for Atomic Clock of Navigation Satellites[D]. Zhengzhou: Information Engineering University, 2006(郭海荣. 导航卫星原子钟时频特性分析理论与方法研究[D]. 郑州:信息工程大学,2006)
- [6] Huang Guanwen. Research on Algorithms of Precise Clock Offset and Quality Evaluation of GNSS Satellite Clock [M]. Beijing: Surveying and Mapping Press, 2017(黄观文. GNSS 星载原子质量评价及精密钟差算法研究[M]. 北京:测绘出版社,2017)
- [7] Allan D W. The Statistics of Atomic Frequency Standards[J]. *Proceedings of the IEEE*, 1966, 54(2):221-230
- [8] Steigenberger P, Hugentobler U, Hauschild A, et al. Orbit and Clock Analysis of Compass GEO and IGSO Satellites[J]. *Journal of Geodesy*, 2013, 87(6):515-525
- [9] Steigenberger P, Hugentobler U, Loyer S, et al. Galileo Orbit and Clock Quality of the IGS Multi-GNSS Experiment [J]. *Advances in Space Research*, 2015, 55(1):269-281
- [10] Sesia I, Galleani L, Taavella P. Implementation of the Dynamic Allan Variance for the Galileo System Test Bed V2[C]. Frequency Control Symposium, 2007 Joint with 21st European Frequency and Time Forum, Geneva, Switzerland, 2007
- [11] Sesia I, Galleani L, Taavella P. Application of the Dynamic Allan Variance for the Characterization of Space Clock Behavior [J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2011, 47(2):884-895
- [12] Yu Heli, Hao Jiming, Liu Weiping, et al. A Real-Time Anomaly Monitoring Algorithm for Satellite Clock [J]. *Geomatics and Information Science of Wuhan University*, 2016, 41(1):106-110(于合理,郝金明,刘伟平,等. 一种卫星钟异常实时监测算法[J]. 武汉大学学报·信息科学版, 2016, 41(1):106-110)
- [13] Niu Fei, Han Chunhao, Zhang Yisheng, et al. Analysis and Detection on Atomic Clock Anomaly of Navigation Satellites [J]. *Geomatics and Information Science of Wuhan University*, 2009, 34(5):585-588(牛飞,韩春好,张义生,等. 导航卫星星载原子钟异常监测分析[J]. 武汉大学学报·信息科学版, 2009, 34(5):585-588)
- [14] Zhang Qianqian, Gui Qingming. Bayesian Methods for Outliers Detection in GNSS Time Series [J]. *Journal of Geodesy*, 2013, 87(7):609-627
- [15] Zhang Qianqian, Han Songhui, Du Lan, et al. Bayesian Methods for Outliers Detection and Estimation in Clock Offset Measurements of Satellite-Ground Time Transfer [J]. *Geomatics and Information Science of Wuhan University*, 2016, 41(6):772-777(张倩倩,韩松辉,杜兰,等. 星地时间同步钟差异常处理的 Bayesian 方法[J]. 武汉大学学报·信息科学版, 2016, 41(6):772-777)
- [16] Li Tao, Heng Guanghui, Gui Qingming. The Application of Bayesian Method in Autoregressive Model Outliers Detecting of Satellite Clock Error Prediction [J]. *GNSS World of China*, 2010(4):15-20(李涛,衡广辉,归庆明. AR 序列异常值探测的 Bayes 方法在卫星钟差预报中的应用[J]. 全球定位系统, 2010(4):15-20)

- [17] Wang Yupu, Lv Zhiping, Chen Zhengsheng, et al. A New Data Preprocessing Method for Satellite Clock Biased Its Application in WNN to Predict Medium-Term and Long-Term Clock Bias [J]. *Geomatics and Information Science of Wuhan University*, 2016, 41(3):373-379(王宇谱, 吕志平, 陈正生, 等. 一种新的钟差预处理方法及在WNN钟差中长期预报中的应用[J]. 武汉大学学报·信息科学版, 2016, 41(3):373-379)
- [18] Teng Yunlong, Shi Yibing, Zheng Zhi. Time Series Analysis and Its Application in Detection and Correction of Cycle Slip [J]. *Journal of Astronautics*, 2011, 32(3):543-548(滕云龙, 师亦兵, 郑植. 时间序列分析在周跳探测与修复中的应用[J]. 宇航学报, 2011, 32(3):543-548)
- [19] Guo Hairong, Yang Sheng, Yang Yuanxi, et al. Numerical Prediction Methods for Clock Difference Based on Two-Way Satellite Time and Frequency Transfer Data [J]. *Geomatics and Information Science of Wuhan University*, 2007, 32(1):43-46(郭海荣, 杨生, 杨元喜, 等. 基于卫星双向时间频率传递进行钟差预报的方法研究[J]. 武汉大学学报·信息科学版, 2007, 32(1):43-46)
- [20] Koch K R. Bayesian Inference with Geodetic Application [M]. Berlin: Springer, 1990
- [21] Box G E P, Jenkins G M. Time Series Analysis: Forecasting and Control [M]. 4th ed. New Jersey: Hoboken, 2011
- [22] Claeskens G, Hjort N L. Model Selection and Model Averaging [M]. New York: North-Holland, 2008
- [23] Schwarz G. Estimating the Dimension of a Model [J]. *The Annals of Statistics*, 1978, 6(2): 461-464

Model Selection Method Based on ARIMA Model in Outliers Detection of Satellite Clock Offset

MA Chaozhong^{1,2} ZHU Jianqing³ HAN Songhui¹

¹ Basic Department of Information Engineering University, Zhengzhou 450001, China

² Institute of Geospatial Information, Information Engineering University, Zhengzhou 450001, China

³ College of Mathematics and Physics, Suzhou University of Science and Technology, Suzhou 215009, China

Abstract: Due to the interference of various uncertain factors, the abnormal disturbances often occur in the satellite clock offset data, which reduces the reliability of the performance analysis of the satellite clock, destroys the validity of the modeling and prediction of clock offset, and affects the accuracy of the navigation positioning results. As to this problem, on the basis of the autoregressive integrated move average (ARIMA) model, this paper establishes an outlier detection model of clock offset time series. Based on the principle of Bayes statistics, the problems of outliers detection and the outliers magnitudes estimation are transformed into a model selection problem. Through the approximate calculation of the posterior probability of the model, the measurement standard of the model selection is derived so the complex iterative computation is avoided. Simulation Test examples of GPS and BeiDou illustrate that the proposed method can detect the outliers effectively and estimate the magnitudes of outliers accurately in the clock offset sequence; furthermore, it can obtain higher prediction precision when the method is applied in the medium and long term prediction of the satellite clock offset.

Key words: satellite clock offset; outliers; Bayes theory; autoregressive integrated move average model

First author: MA Chaozhong, PhD candidate, associate professor, specializes in surveying error theory and GNSS data processing methods. E-mail: marcz@163.com

Foundation support: The National Natural Science Foundation of China (41474009, 41174005).

引文格式: MA Chaozhong, ZHU Jianqing, HAN Songhui. Model Selection Method Based on ARIMA Model in Outliers Detection of Satellite Clock Offset[J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(2):167-172. DOI:10.13203/j.whugis20180230(马朝忠, 朱建青, 韩松辉. 基于ARIMA模型的卫星钟差异常值探测的模型选择方法[J]. 武汉大学学报·信息科学版, 2020, 45(2):167-172. DOI:10.13203/j.whugis20180230)