

# 基于Python的测量数据精度鉴定的预处理方法

涂鑫, 张小飞, 邱维江, 陈欣

中国人民解放军63819部队, 四川 宜宾

收稿日期: 2022年3月3日; 录用日期: 2022年4月1日; 发布日期: 2022年4月8日

## 摘要

针对测控设备在精度鉴定中存在数据时间未对齐、时间格式错误、数据不连续等问题, 探讨测控设备在精度鉴定中数据预处理的一些具体方法。本文首先对数据预处理的流程及工具进行简要介绍, 然后指出了测量数据与精密轨道数据之间存在的时间未对齐、时间格式错误等问题, 并提出了借助Python的第三方库函数Numpy和Pandas实现对应问题的解决方法, 最终给出处理前后的结果对比。结果证明, 本文给出的数据预处理的方法行之有效, 能够有效解决测控设备精度鉴定中数据间存在的问题, 对精度鉴定的数据预处理方面具有一定的借鉴意义。

## 关键词

测量数据, 精度鉴定, Python, 数据分析

# A Preprocessing Method of Precision Appraisal for Measurement Data Based on Python

Xin Tu, Xiaofei Zhang, Weijiang Qiu, Xin Chen

63819 Troops of the Chinese People's Liberation Army, Yibin Sichuan

Received: Mar. 3<sup>rd</sup>, 2022; accepted: Apr. 1<sup>st</sup>, 2022; published: Apr. 8<sup>th</sup>, 2022

## Abstract

Aiming at the problems of data time unalignment, time format error and data discontinuity in precision appraisal of measurement and control equipment, this paper discusses some specific methods of data preprocessing in precision appraisal for measurement and control equipment. This paper first briefly introduces the flow of data preprocessing and tools, and then points out that

there are some problems between the measured data and the precision orbit data, such as time misalignment, time format error and so on, and puts forward the solution to the corresponding problem with the help of Python's third-party library functions like numpy and pandas, and finally gives the results of the comparison before and after processing.

## Keywords

Measured Data, Precision Appraisal, Python, Data Analysis

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

测控设备的精度鉴定, 主要通过对测量数据与精密轨道数据的综合分析, 从而给出测控设备测量精度效果的过程。在精度鉴定中, 数据的预处理对最终数据分析结果的有效性起着至关重要的作用, 主要包括系统误差修正、电波折射误差修正等修正处理, 同时也包括了数据有效性检验、坐标转换、时标对齐、数据对齐等最基础的处理。在文献[1]中, 对数据的采样时间不一致问题进行了插值算法的数据对齐研究, 在文献[2]和文献[3]在数据预处理方面只讲了时间误差修正的内容, 文献[4]对精度鉴定的4种方法进行了对比研究, 但是没有提及预处理过程, 文献[5]对数据预处理提出了比较全面的标准及部分处理算法, 但是这些精度鉴定研究方面的文献, 对数据的预处理只是多半做了个简要描述, 未进行实际数据研究。

目前, 在利用“天平一号”标校星开展测控设备的精度鉴定时, 主要方法是分析测控设备的测量数据与“天平一号”精轨数据差值的误差量, 但是由于测量数据和精密轨道数据本身的不确定性, 数据可能会出现时间格式有误、丢帧、时间不一等一系列问题, 在上述文献中也未提及这些问题, 直接开展数据分析的话有可能导致精度鉴定结果无效。因此做好数据的预处理也是精度鉴定的关键工作。本文将运用某型号单脉冲雷达跟踪“天平一号”标校星的测量数据, 针对数据中遇到的时间格式错误、数据不连续、时间对不齐等问题, 开展精度鉴定中相应问题的数据预处理方法研究。

## 2. 数据预处理流程及工具

### 2.1. 数据预处理流程

测控设备精度鉴定的基本流程如图1所示, 其中, 数据预处理包括对测控设备测量数据的动态滞后修正、系统误差修正及电波折射修正等修正类预处理, 同时包括卫星精轨数据在内开展的还有合理性检验、平滑滤波等数据处理内容[5]。具体如下:

1) 动态滞后修正及信标时延修正: 动态滞后修正是在跟踪目标的过程中, 由于角度变化快, 容易产生动态滞后, 需要进行波束中心偏角修正, 信标时延修正是对“天平一号”标校星上的脉冲应答机固定时延导致的距离修正;

2) 系统误差修正: 根据标校得到的各项误差数据, 对设备的外测数据实施方位、俯仰、距离进行系统误差的修正;

3) 电波折射修正: 由于无线电波在大气中传播时会受到不均匀介质的影响使电波射线发生弯曲, 传播速度异于光速而产生折射误差的修正;

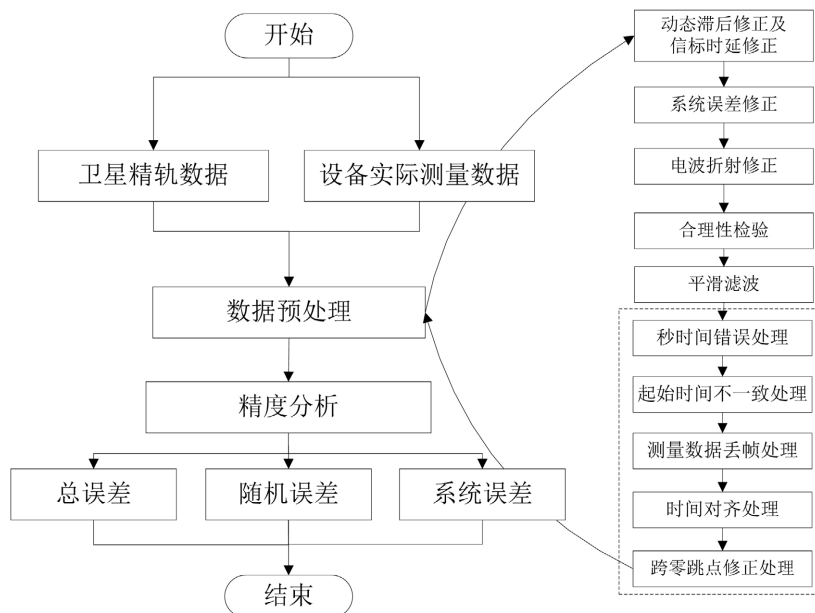


Figure 1. Basic process of data precision appraisal

图 1. 数据精度鉴定基本流程

- 4) 合理性校验：对在跟踪过程中可能会产生异常值进行合理性检验；
  - 5) 平滑滤波：由于测量系统不可避免的会产生测量噪声，从而导致测量数据出现随机误差，为减小误差带来的影响，需要对测量数据进行平滑滤波；
  - 6) 秒时间错误处理：对精轨数据中遇到的秒时间错误，即满 60 秒却未进分钟位的处理；
  - 7) 起始时间不一致处理：对获取到的精轨数据开始的第一帧时间与待分析的测量数据的第一帧时间不重合所进行的处理；
  - 8) 测量数据丢帧处理：对测量数据中存在丢帧，删除对应丢失时段精轨数据的处理；
  - 9) 时间对齐处理：对测量数据和精轨数据精确时间不对齐的数据处理；
  - 10) 跨零跳点修正处理：对测量数据中整周跳点而不同时过  $0^\circ$  (跨零)产生的误差而开展的修正处理；
- 由于系统动态滞后修正系统误差修正、大气折射修正、平滑滤波等处理的研究文献较多，标准比较统一，本文不再赘述，而是主要有针对性地开展秒时间错误、数据对齐等处理方法研究，以解决测量数据与卫星精轨数据的采样时刻不一致、时间格式错误、跨零跳点、丢帧等问题。

## 2.2. 数据预处理工具

在本文的数据预处理的过程中，全部运用 Python 的第三方库来实现。

利用 Python 进行数据分析，Pandas 库是首选库，它提供了大量快速便捷的处理大型数据集(比如 Excel 文件读写)所需的函数和算法[6]，并含有例如 DataFrame、Series 等各类数据结构，能够使数据分析工作简单而高效。具体有以下特点[7]：

- 1) 提供的数据结构既能处理时间序列，也能处理非时间序列；
- 2) 数学运算和约简可以轴向执行；
- 3) 具有强大的缺失数据处理功能；
- 4) 可以合并常见数据库中的关系型运算。

此外 Pandas 还有数据标准化和数据转换的功能等。

NumPy (Numerical Python)是 python 科学计算的基础工具包[8], 包括了 N 维数组对象 Array。拥有成熟的函数库, 例如矩阵数据类型和矢量处理等, 很适合用于多维数组的数据处理。

Numpy 和 Pandas 二者结合起来可以非常方便地实现报表源数据到目标数据文件的快速转换。

### 3. 数据预处理方法

数据预处理首先需要根据设备的测量数据获取对应的卫星精轨数据, 并对精轨数据进行坐标转换, 以保证两类数据在同一坐标系下作比较; 再分别对两类数据本身及数据之间存在的问题做一系列的预处理, 其中问题包括数据第一帧对齐、删除丢失的测量数据、时间不对齐等, 最后再进行坐标转换, 为精度鉴定提供准确的待分析数据。数据中存在的问题及其处理方法如下。

#### 3.1. 问题描述

对同一时段的测控设备测量数据与卫星精轨数据进行初步分析, 表中, Time 为数据的绝对时时标, X、Y、Z、V<sub>x</sub>、V<sub>y</sub>、V<sub>z</sub> 为卫星在地固系下 X、Y、Z 分量及分量速度, V、R、A、E 为测量系下目标的速度、距离、方位角度及俯仰角度, DA、DE 为方位、俯仰角度的变化量。分析后发现二者在做精度鉴定分析前, 存在以下问题:

##### 1) 秒时间有误

在精轨数据中, 时间格式出现 xx:xx:60.xxx 形式, 即满 60 秒却未进分钟位, 导致时间不满足正常格式要求。如表 1 所示, 精轨数据时间出现 18:36:60.000, 其正确时间形式应为 18:37:00.000。

Table 1. Precision orbit data error data in seconds time

表 1. 精轨数据秒时间错误数据

Time	X (m)	Y (m)	Z (m)	V <sub>x</sub> (m/s)	V <sub>y</sub> (m/s)	V <sub>z</sub> (m/s)
2021-09-11T18:36:60.000	-1,979,075.511	5,961,320.698	2,781,960	2422.595	-2435.05	6889.751
2021-09-11T18:37:00.050	-1,978,954.379	5,961,198.936	2,782,304	2422.698	-2435.43	6889.579
2021-09-11T18:36:60.000	-1,978,833.241	5,961,077.154	2,782,649	2422.802	-2435.82	6889.407

##### 2) 起始时间不一

根据测量数据获取到的卫星精轨数据, 其开始的第一帧时间不一定会与测量数据的第一帧时间完全重合, 直接进行差值计算会导致偏差过大, 如表 2 和表 3 所示。测量数据第一帧起始时间为 18:35:42.250, 对应下载的原始精轨数据第一帧起始时间 18:35:42.000。

Table 2. Start time of measurement data

表 2. 测量数据起始时间

Time	V (m/s)	R (m)	A (°)	E (°)	DA (°)	DE (°)
18:35:42.250	-6683.20	1.31E+06	154.366	16.8092	0	-0.00781
18:35:42.300	-6682.84	1.31E+06	154.362	16.8226	0	-0.00781
18:35:42.350	-6682.47	1.31E+06	154.357	16.8284	0	-0.00781

Table 3. Start time of Precision orbit data

表 3. 精轨数据起始时间

Time	X (m)	Y (m)	Z (m)	V <sub>x</sub> (m)	V <sub>y</sub> (m)	V <sub>z</sub> (m)
18:35:42.000	-2,161,428	6,127,781	2,234,803	2249.087	-1830.87	7131.155

Continued

18:35:42.050	-2,161,316	6,127,689	2,235,160	2249.206	-1831.26	7131.018
18:35:42.100	-2,161,204	6,127,598	2,235,516	2249.325	-1831.65	7130.880

## 3) 测量数据丢帧

设备进行跟踪时有可能出现不连续或无效跟踪数据剔除,在此统称为数据丢帧,若直接处理的话会出现数据中途错位继而差值突然增大的情况。从表 4 中可以发现测量数据记录的时间从 18:49:32.700 跳到 18:49:34.550,中间丢失了 37 帧数据。

Table 4. Loss record of measurement data

表 4. 测量数据丢失记录

Time	V (m/s)	R (m)	A (°)	E (°)	DA (°)	DE (°)
18:49:32.650	454.906	489,573	220.917	92.6592	0.085938	0.039063
18:49:32.700	460.474	489,594	220.928	92.6695	0.109375	0.023438
18:49:34.550	651.769	490,618	282.238	85.7685	-0.09375	0.320312
18:49:34.600	657.712	490,649	284.226	85.7390	-0.09375	0.335938

## 4) 时间未对齐

精度鉴定的两类数据都选用同频率采样数据,但由于数据传输会产生时延导致测量数据的时间出现偏差,如表 5 中标准测量数据时间应为 18:28:33.700,但时延存在导致了 0.002 毫秒的偏差。

Table 5. Time delay of measure data

表 5. 测量数据时延

Time	V (m/s)	R (m)	A (°)	E (°)	DA (°)	DE (°)
18:28:33.702	-7020.16	2,237,433	155.9238	3.424301	0.002292	0.014897
18:28:33.752	-7020.13	2,237,086	155.9228	3.435631	-0.00344	0.00573
18:28:33.803	-7020.05	2,236,736	155.9207	3.445244	0.004011	0.011459
18:28:33.853	-7020.00	226,389	155.9190	3.454857	0.005157	0.006303

## 5) 跨零不同步

经过一系列预处理后需要对测量数据和精轨数据作差。其中方位角的范围为  $0^{\circ}\sim 360^{\circ}$ ,由于两份数据不一定同时过  $0^{\circ}$  (跨零),做差之后可能会出现角度差值过大的情况,如表 6 所示,其中,  $A_{jg}$  为卫星精轨数据方位角度,  $A_{cl}$  为测量数据的方位角度, DA 为同一时刻的角度差值。在 18:41:00.950 时精轨数据已跨零,而此时测量数据数值仍在跨零之前,直接计算时其差值为接近于  $360^{\circ}$  的错误差值。

Table 6. Record data across zero at different times

表 6. 不同时跨零记录数据

Time	$A_{jg}$ (°)	$A_{cl}$ (°)	DA (°)
18:41:00.900	0.000211	0.032952	0.032741
18:41:00.950	359.9964	0.028452	-359.968
18:41:01.000	359.9926	0.024294	-359.968

### 3.2. 处理方法

#### 1) 时间错误、起始时间不一致及丢帧问题处理

对于前三个问题，利用 Pandas 库进行统一处理。首先将测量数据读入 dataframe 中，利用库函数将时间转为 datetime [64]格式再进行处理，具体流程如图 2 所示，具体如下：

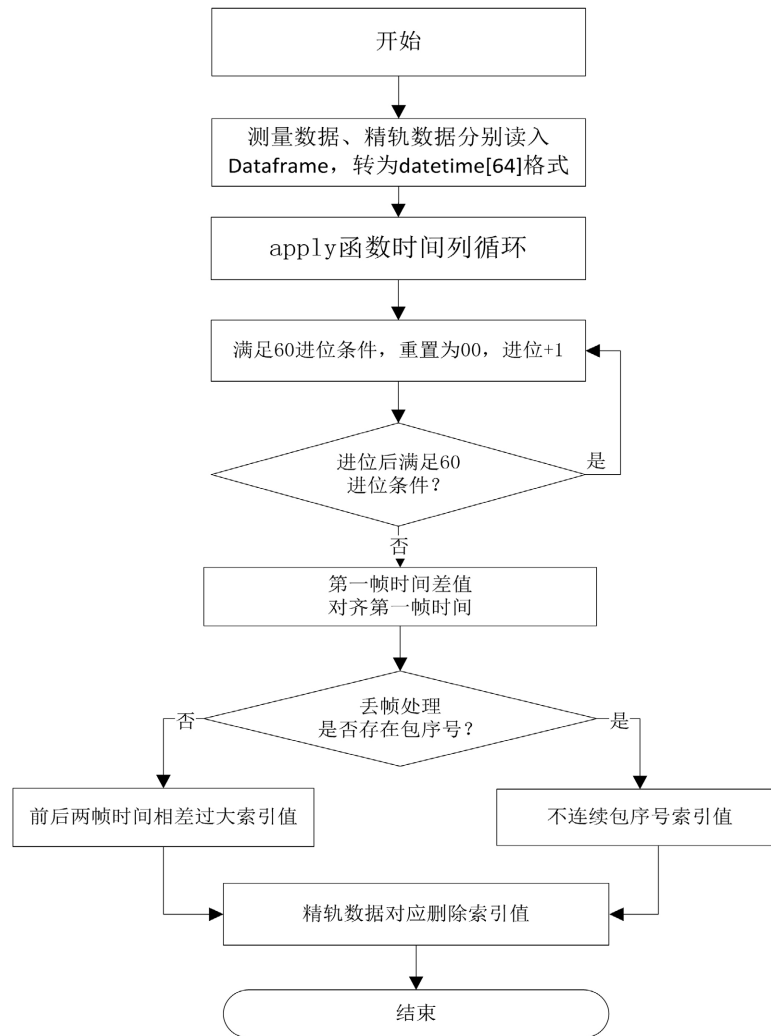


Figure 2. The flow of data alignment processing

图 2. 数据对齐处理流程

a) 利用 apply 函数对时间列循环判断，将 60 秒替换为 00，并将分钟加 1。此时还要注意分钟加 1 后是否满足 60 需要小时进位的情况。

b) 根据两份文件的第一帧时间差值，判断哪份数据时间靠前并计算多出的帧数，然后将靠前的数据上移对应帧数，使两份数据第一帧时间对齐。

c) 找到 PDXP 包序号不连续的索引值，若没有包序号则找出前后两帧时间相差过大的索引值，根据索引值删除对应的精轨数据。

#### 2) 时间对齐处理

对于时间未对齐的问题，为减少时延带来的不必要的误差，需要事先对数据进行时间修正。首先判

断对齐后的数据是否存在微小时延, 若存在需要将时标重新更正与精轨数据的时标相同, 同时将测量数据修正为更改时标后的数据。计算方法为:

$$x' = x - dx \times \Delta T \quad (1)$$

其中

$$dx = \frac{x_2 - x_1}{T_2 - T_1} \quad (2)$$

$$\Delta T = T_1 - T_{\text{轨道}} \quad (3)$$

式中  $x$  为时标修正前的测量数据;  $x'$  为时标修正后的数据;  $dx$  为测量数据两帧之间的变化率;  $\Delta T$  为时间对齐时测量数据相比于轨道数据的时延。

### 3) 跨零跳点修正处理

对于跨零不同步导致差值过大问题, 需要先找到方位角差值过大的值, 再计算它的相对差值。跨零跳点[5]的具体处理过程:

步骤一: 令  $j=1, A'_j = A_j$ 。

步骤二: 判断式(4)和式(5):

$$|A_{j+1} - A'_j| > \delta_0 \quad (4)$$

$$|A_{j+2} - A_{j+1}| \leq \delta_0 \quad (5)$$

式中:  $\delta_0$  为判断门限, 一般取值范围为  $350^\circ$  到  $359^\circ$ 。若式(4)和式(5)同时成立, 则认为  $A_{j+1}$  为跳点, 修正如下:

当  $A_{j+1} - A'_j < -\delta_0$  时, 令  $A'_{j+1} = A_{j+1} + 360^\circ$ 。

当  $A_{j+1} - A'_j > \delta_0$  时, 令  $A'_{j+1} = A_{j+1} - 360^\circ$ 。

若式(4)和式(5)不同时成立, 则认为  $A_{j+1}$  不是跳点, 令  $A'_{j+1} = A_{j+1}$ 。

步骤三: 依次取  $j = 2, 3, \dots, N-2$ , 重复步骤二中各步。

## 4. 预处理结果分析

以某型号单脉冲雷达跟踪“天平一号”标校星为例, 对雷达测量数据进行精度鉴定预处理的验证分析。

### 4.1. 数据预处理计算结果

测量数据与精轨数据经过秒时间处理、第一帧时间对齐、丢帧处理、时间对齐、跨零跳点修正处理等一系列预处理修正后, 便可对两份数据进行误差统计。各处理结果分析如下:

#### 1) 秒时间修正结果

对于表 1 中所示的精轨数据的时间, 由原先的 18:36:60.000 修正为 18:37:00.000。如表 7 所示。

**Table 7.** Comparison of precision orbit data before and after processing in seconds time

**表 7.** 精轨数据秒时间处理前后对比

修正前		修正后	
Time	Date	Time	Date
2020-09-11	18:36:60.000	2020-09-11	18:37:00.000
2020-09-11	18:37:00.050	2020-09-11	18:37:00.050

## 2) 第一帧数据时间对齐结果

如表 2 和表 3 所示, 测量数据起第一帧起始时间 18:35:42.250, 下载的原始精轨数据第一帧起始时间为 18:35:42.000。经数据对齐处理后第一帧数据时间为最晚时间 18:35:42.250, 其计算结果如表 8 所示。

**Table 8.** Time alignment calculation results**表 8.** 数据时间对齐计算结果

$R_{jg}$ (m)	$A_{jg}$ (°)	$E_{jg}$ (°)	$DR_{jg}$ (m/s)	$Time_{jg}$	$Time_{cl}$	dR (m)	dA (°)	dE (°)	dV (m/s)
1,314,425	154.328	16.734	-6682.77	18:35:42.250	18:35:42.250	37.3993	0.0378	0.0169	-0.436
1,314,091	154.324	16.742	-6682.41	18:35:42.300	18:35:42.300	41.5365	0.0376	0.0228	-0.437
1,313,757	154.321	16.750	-6682.05	18:35:42.350	18:35:42.350	35.6521	0.0364	0.0211	-0.429

## 3) 丢帧处理结果

如表 4 所示的测量数据中间丢失了 37 帧数据的问题, 将原本连续的精轨数据对应删除 18:49:32.700 至 18:49:34.550 时间段数据, 以对标测量数据, 其结果如表 9 所示。

**Table 9.** Processing result of frame loss of precision orbit data**表 9.** 精轨数据丢帧处理结果

$R_{jg}$ (m)	$A_{jg}$ (°)	$E_{jg}$ (°)	$DR_{jg}$ (m/s)	$Time_{jg}$	$Time_{cl}$
489,540.6	330.0844	86.36835	436.6873	18:49:32.600	18:49:32.600
489,562.5	330.3054	86.32559	442.2990	18:49:32.650	18:49:32.650
489,584.8	330.5212	86.28279	447.9101	18:49:32.700	18:49:32.700
490,605.1	336.0357	84.67652	654.8979	18:49:34.550	18:49:34.550
490,704.6	336.3396	84.54512	671.6198	18:49:34.700	18:49:34.700

## 4) 时间对齐处理结果

对存在时间误差的测量数据, 将其时间为 18:35:42.304, 对标精轨数据时间将各数据修正到 18:35:42.300 时的数值, 结果如表 10 所示。

**Table 10.** Result of time alignment**表 10.** 时间对齐处理结果

$R_{修正前}$ (m)	$A_{修正前}$ (°)	$E_{修正前}$ (°)	$DR_{修正前}$ (m/s)	$Time_{修正前}$	$Time_{修正后}$	$R_{修正后}$ (m)	$A_{修正后}$ (°)	$E_{修正后}$ (°)	$DR_{修正后}$ (m/s)
1,316,068	154.4187	17.0473	-6682.83	18:35:42.304	18:35:42.300	1,316,094.48	154.4190	17.0467	-6682.85
1,315,726	154.4139	17.0566	-6682.47	18:35:42.353	18:35:42.350	1,315,576.94	154.4141	17.0560	-6682.49
1,315,392	154.4084	17.0649	-6682.11	18:35:42.403	18:35:42.400	1,315,412.04	154.4087	17.0643	-6682.13

## 5) 跨零跳点处理结果

在 18:41:00.900 时, 精轨数据方位角由  $0.000211^\circ$  变为  $359.9964^\circ$ , 此时测量数据已经在此时刻前完成了跨零, 方位角变为  $0.0284520^\circ$ , 对其跨零跳点差值修正, 得到其相对差值为  $0.03205^\circ$ 。如表 11 所示。



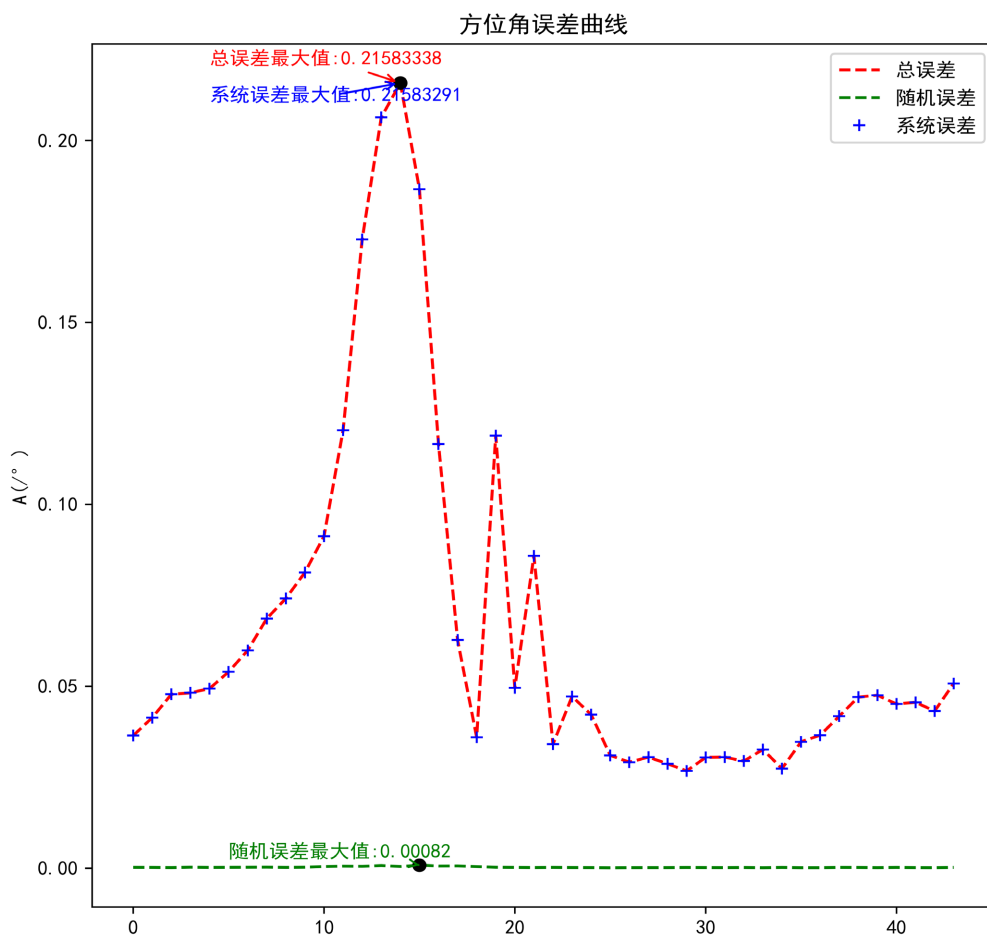
**Table 11.** Difference processing results**表 11.** 差值处理结果

Time	$A_{jg}$ (°)	$A_{cl}$ (°)	dA (°)
18:41:00.900	0.000211	0.032952	0.03274
18:41:00.950	359.9964	0.028452	0.03205
18:41:01.000	359.9926	0.024294	0.031697

## 4.2. 精度鉴定结果

为了进一步检验数据预处理的实际效果,选取某圈次数据开展精度鉴定分析,本文进行精度鉴定时主要采用经典鉴定方法[4],对标校卫星的同一采样时刻将测量数据  $S_1$  和精轨数据  $S_0$  作差得到差分序列  $\{\Delta S_j\}$ ,经统计分析处理得到被鉴定系统测量数据的各类误差统计量(一般为随机误差、系统误差、总误差)。

数据进行一系列预处理后,数据精度鉴定最终处理结果如表 12 所示,其中,数据每隔 10 秒统计一次误差,各误差统计量结果如图 3~5 所示。通过表 12 和图 3~5 的统计结果与该型号单脉冲雷达的性能指标对比,该精度鉴定结果满足设备性能指标要求,说明数据预处理后的数据可用于精度鉴定分析使用,数据预处理方法可靠有效。

**Figure 3.** Azimuth error curve (step length: 10 seconds)**图 3.** 方位角误差曲线(步长 10 秒)

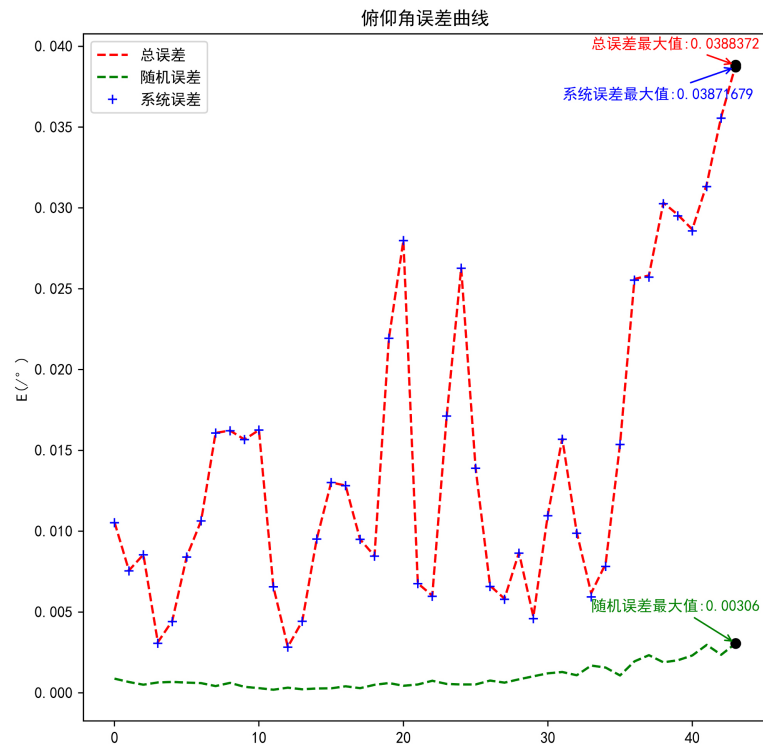


Figure 4. Pitch angle error curve (step length: 10 seconds)

图 4. 俯仰角误差曲线(步长 10 秒)

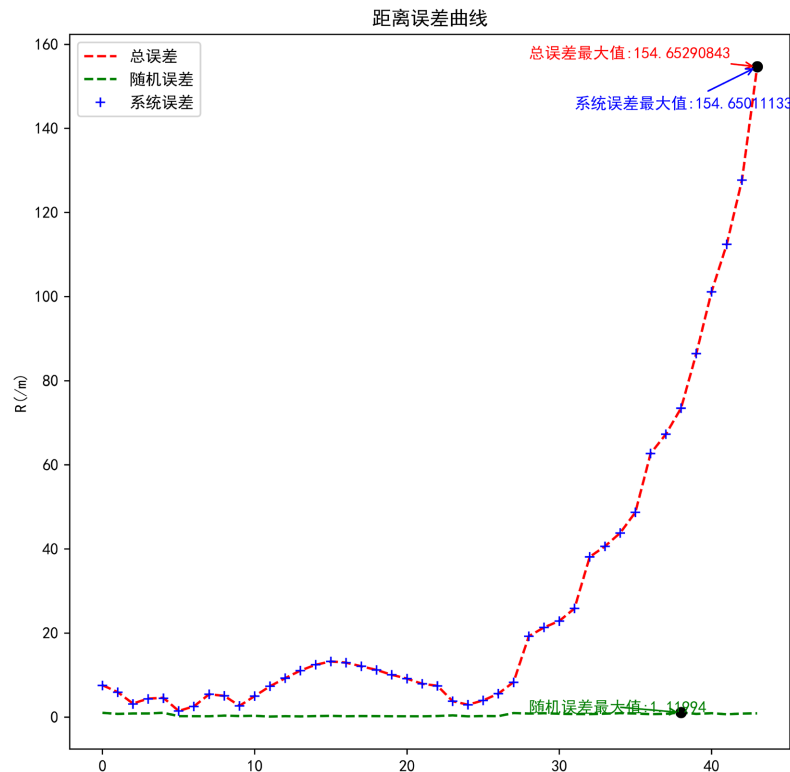


Figure 5. Distance error curve (step length: 10 seconds)

图 5. 距离误差曲线(步长 10 秒)

**Table 12.** Data accuracy appraisal processing results  
**表 12.** 数据精度鉴定处理结果

参数	总误差	随机误差	系统误差
距离(单位: 米)	20.4815	0.6338	20.4717
方位角(单位: °)	0.0802631	0.00031184	0.0802625
俯仰角(单位: °)	0.0622897	0.00098764	0.0622819

## 5. 结束语

本文针对精度鉴定中测量数据与精轨数据所存在的时间格式错误等问题,介绍了针对问题所开展的预处理方法。重点提出了精轨数据时间格式错误问题,测量数据与精轨数据第一帧未对齐问题,测量数据有丢帧、不连续问题,测量数据与精轨数据时间未对齐问题以及跨零跳点问题,给出了对应问题的解决处理方法,并通过跟踪“天平一号”的测量数据进行逐个处理验证,最后开展精度鉴定分析。通过数据验证结果分析,说明各数据处理方法流程正确,切实可行,对精度鉴定中数据预处理方法具有一定的参考借鉴意义。

## 参考文献

- [1] 王鹏宇, 李乃忠, 宋玉红, 刘凯. 无线电测控设备卫星标校关键问题分析[J]. 航天器发射场, 2020, 127(2): 42-46.
- [2] 胡绍林, 许爱华. 脉冲雷达跟踪测量数据处理技术[M]. 北京: 国防工业出版社, 2007.
- [3] 刘利生, 张玉祥. 外弹道测量数据处理[M]. 北京: 国防工业出版社, 2002.
- [4] 刘利生, 赵华, 刘元. 应用标校卫星鉴定测控系统精度方法探讨[J]. 飞行器测控学报, 2014, 33(4): 275-282.
- [5] GJB 1381A-2011. 导弹、航天器实验光电经纬仪和脉冲雷达测量精度评定[S]. 北京: 中国人民解放军总装备部, 2011.
- [6] 郭学兵. 基于 Python 的并行编程技术在批量象规范报表入库处理中的应用[J]. 软件, 2018, 39(7): 33-37.
- [7] 梁丽. Python 编程语言在数据分析中的运用[J]. 信息与电脑, 2020, 12(2): 25-31.
- [8] 李俊华. 基于 Python 的数据分析[J]. 电子技术与软件工程, 2018, 17(1): 56-62.