Research and Application of a Data Processing Method on Outliers in Unmanned Aerial Vehicle (UAV) Tracking Measurement

Kunkun Li[*], *Rui* Cao, *Xiaoyu* Qi, and *Yaxiong* Hu

Northwest Institute of Nuclear Technology, China

Abstract. In unmanned aerial vehicle (UAV) flight test, some complex outliers often appear in tracking measurements due to various factors stemming from environment, instrumentations, and even operators. This seriously affects the reliability of applications and analysis. This paper aims at the problem of outliers elimination to measurement data and proposes a joint data processing method comprising grouped data extraction, data trend modeling, and outliers detection. In the context of accuracy assessment test for measurement and control equipment (MCE), an approach using the measurement parameter for UAV flight is designed to estimate the time system deviation between MCEs, as well as to validate the effect of the proposed joint method. Finally, a simulation is implemented and results show the effectiveness and feasibility of the joint method to outliers elimination and time deviation estimation.

1 Introduction

With the development of advanced technology throughout the world, *Aircraft flight test* (AFT), an indispensable technique in tasks such as aircraft design assistance, quality inspection, and performance appraisal, plays an increasingly important role in the coming into service and capabilities upgrade of aircrafts [1]. As an unusual branch *unmanned aerial vehicle* (UAV) flight continually attracts the interests of scholars from various research domains owing to its widespread availability and immense demand. In such a flight test, often times the tracking measurements serve as a basis of assessment for UAV flight quality, and tracking performance of ground-based *measurement and control equipment* (MCE). Hence, the measurement data quality directly affects the reliability of following applications and analysis.

Regrettably, during data acquisition some data points whose patterns do not conform to expected normal behavior of measurement data often occur in sequential tracking measurement data. These anomalous patterns or abnormal data points are referred to as outliers by analyst [2]. Concerning the causes to outliers, they typically come from internal impact, vibration, system failure from within instrumentations, environmental disturbance, and even some inappropriate operations by people. To separate the useful information from measurement data, it is, therefore, necessary to do some work including outliers detection, elimination, and correction in the preprocessing phase.

Generally, for a simple one-dimensional data set in time series, the outliers in it can be classified into two basic types according to their continuity, namely isolated outliers and spot outliers, respectively. As they are called, the isolated outliers refer to some discontinuous abnormal data points since accidental disturbances arise at times. And the spot outliers refer to some continuous abnormal data points generally with little difference in amplitude due to some persistent-existing causes [3]. In previous works [4-7], the outlier detection approaches towards isolated outliers revolve around some conventional statistical detection methods. They have well elimination performance but are not always tolerant to spot outliers owing to limited recognition capacity, and constraint on the statistical property of data. Thus they suffer from problems of halfway outliers elimination, and even removal of some useful information. As for spot outliers, the techniques mainly include Kalman filtering method, robust filtering with M-estimation method, B-spline approximation by least square method, etc. These methods offer better recognition to spot outliers, whereas they still pose challenges such as erratic processing performance, and data loss after outliers elimination.

The remainder of this paper is organized as follows: First, a joint data processing method on outliers is formulated, where the Pauta criterion is introduced and the involved model or thought are presented. Then, a problem concerning the estimation of time system deviation between MCEs in accuracy assessment test is described. Further, a solution using the tracking measurements of UAV flight is designed, and next the estimation steps are elaborated in details. After that a simulation on an UAV flight test scenario is executed, and the performance of some estimation methods are compared and analysed. Finally, this paper summarizes the main contributions and draws a conclusion.

---

[*] Corresponding author: lkk9419nudt@163.com

2 Joint data processing method

2.1 Pauta criterion

The Pauta criterion [8], a typical and practical method in data preprocessing, is widely used for scientific computation on some engineering problems. According to the criterion, for a long series of sample data points which follow a normal distribution, let u denote the mean value and σ be the standard deviation. Theoretically, u is 0 in terms of the residuals between u and the sample data values, and the probability of residuals within the confidence interval $(u-3\sigma, u+3\sigma)$ is specified over 99.73%. Hence, some of the sample data are regarded as outliers while the corresponding residuals exceeds such interval with a less 0.3% probability.

Consider a sequence of measurement data points denoted by $x(k)\ \{k=1,2,...,N\}$, where N represents the data length. Suppose some outliers are present at uncertain positions in data series. To detect outliers via Pauta criterion, the arithmetic mean value of measurement data, denoted by $\bar{x}$, needs to be calculated as follows

$$\bar{x} = \frac{1}{N}\sum_{k=1}^{N} x(k) \qquad (1)$$

and the standard deviation σ is given by

$$\sigma = \sqrt{\sum_{k=1}^{N} \frac{\left(x(k)-\bar{x}\right)^2}{N-1}} \qquad (2)$$

With $\bar{x}$, the residuals of measurement data can be expressed by

$$x_r(k) = x(k) - \bar{x},\ k=1,2,...,N \qquad (3)$$

Compare the absolute value of residuals above with a threshold 3σ in sequence, and conduct the critical condition. Once $|x_r(k)| \geq 3\sigma$ holds, the data at point k, viz., $x(k)$, is then perceived as an untrusted data or outlier, and thus should be eliminated, otherwise be preserved.

2.2 Data trend modeling by least square

One of the key to data analysis lies in a good mathematical model chosen to enhance the extraction of useful information. However, a congruent model is hardly established in most cases owing to lack of priori information, and limited data collection. With such challenges, numerous methods have been developed and applied for data feature extraction over the past several decades [9-11], and amongst them the *least square* (LS) method is tractable to analyst from various knowledge disciplines due to its simplicity, robustness, and convenience of implementation.

The steps of data trend modeling with least square method are

Step 1. Suppose an approximate polynomial with a certain degree is exploited to data trend fitting. List a matrix equation according to the LS theorem;

Step 2. Solve the equation to yield a coefficient matrix of the above polynomial, and then we derive a curve fitting data trend.

To be more specific, let $x(t_n)\ \{n=1,2,3,...,N\}$ denote the time-series data and t_n be the discrete sampling time. Then we employ a Kth-degree polynomial to fit the data trend, as expressed by

$$y(n) = \sum_{k=1}^{K} b_k \left(t_n\right)^k,\ n=1,2,...,N \qquad (4)$$

where b_k represents the polynomial coefficient matrix. In other words, $y(n)$ is taken to be the approximation of polynomial elements w.r.t. $x(n)$. Under such condition an error function is defined as

$$\begin{aligned} E(h) &= \sum_{n=1}^{N}\left(x(n)-y(n)\right)^2 \\ &= \sum_{n=1}^{N}\left(x(n)-\sum_{k=1}^{K} b_k \left(t_n\right)^k\right)^2,\ n=1,2,...,N \end{aligned} \qquad (5)$$

To find the minimum of $E(h)$, let us take the partial derivative of equation (5) w.r.t. b_j and make it to be zero, then we have

$$\frac{\partial E}{\partial b_j} = \sum_{k=1}^{N} 2\left[x(n)-\sum_{k=1}^{K} b_k \left(t_n\right)^k\right]\left[-\left(t_n\right)^j\right]=0,\ n=1,2,...,N \qquad (6)$$

After algebraic transposition, $K+1$ equations are generated as follows

$$\sum_{k=0}^{K} b_k \sum_{n=1}^{N}\left(t_n\right)^{k+j} = \sum_{n=1}^{N} x(n)\left(t_n\right)^j,\ n=1,2,...,N;\ j=0,1,...,K \qquad (7)$$

where b_k could be got by means of matrix inversion. Ultimately, a resulting polynomial for the data trend is obtained.

2.3 Outliers elimination by grouped data extraction

In the research field of engineering, some knotty problems with large scale computing are likely encountered. Most often, some straight-forward solutions would probably offer the corresponding answers in a general sense. Nevertheless, this do also face a significant drawback that such doing not merely wastes researchers' energy but imposes great burden on the computer's processor in attempt to support fast and accurate calculation. With this challenge, some of the problems are sometimes split into the same but smaller
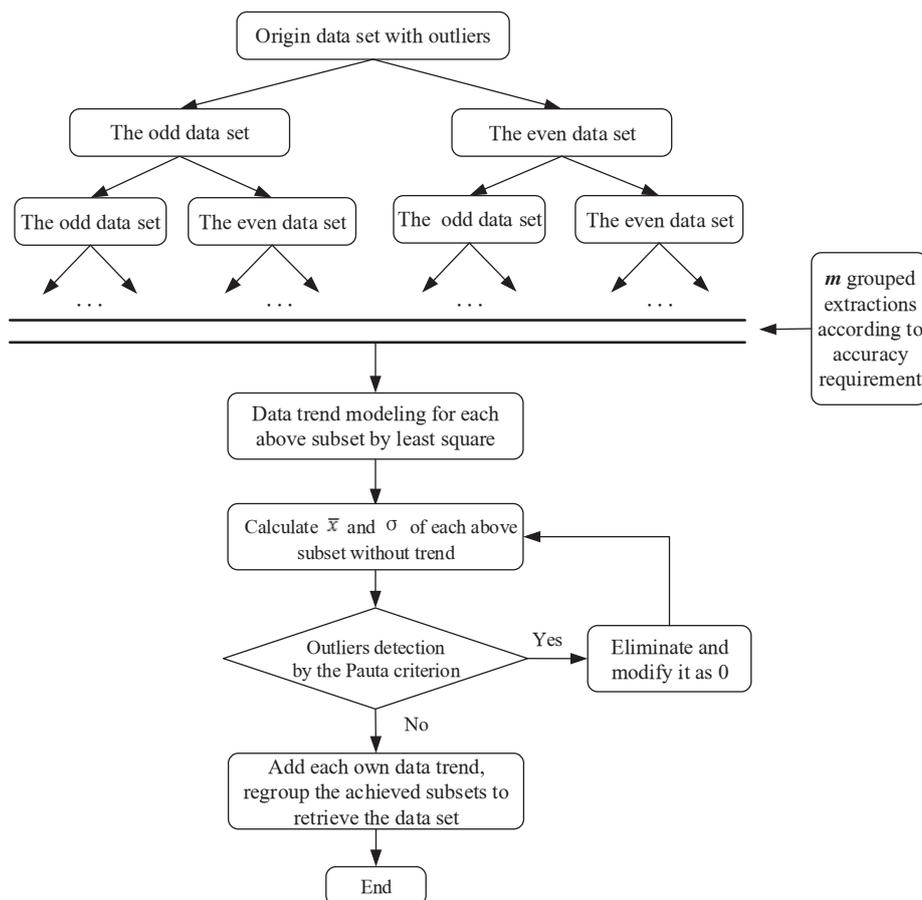
Fig. 1. Flow diagram of outliers elimination by the joint data processing method.

sub-problems by researchers according to the divide and conquer method [12]. As a consequence, the answers of these sub-problems are relatively easier to get, and after a comprehensive analysis, the resolution of original problem is achieved in the end.

With this flexibility in mind, a time-series data set with complicated types of outliers can be decomposed into multiple subsets with smaller sizes. For example, one sequential data set becomes 2^m subsets after m recursive grouped extractions according to the differences of odd and even positions of data points. This way there are two advantages to data processing, as follows

(1) Spot outliers lead to a problem that the calculated σ of original data deviates from the actual σ to some extent. After multiple grouped extractions the number of outliers in each resulting subset decreases, and thus the σ of such a subset tends to the actual σ;

(2) The types of outliers in original data are simplified as all spot outliers resolve into some sparse isolated outliers after several grouped extractions. Thus it is convenient for processing of each subset.

Given all the above argument, for elimination of outliers in a sequential data set we can handle it with a considered joint method on the basis of grouped data extraction, data trend modeling, and outliers detection via Pauta criterion, as shown in Fig. 1.

3 Method application

3.1 Problem description

In general, the *measurement and control equipment* (MCE) consists mainly of some radio or optical measurement equipment (radar, photoelectric theodolite), and some wireless information transmission equipment (telemetry equipment, remote control equipment). Before delivery of a MCE, it is of significance to assess its indicators in a testing outfield with some dynamic conditions in consideration, that is, MCE accuracy assessment test [13]. Such a test relates to ceaseless calibrations for the measurement error of MCE, and most importantly, obtains a comprehensive evaluation of MCE's performance. As such, the capability of MCE is ensured to meet the functional index listed in technical specifications, and customer's requirements.

Nowadays, with the advances of high-precision carrier phase differential GNSS (Global Navigation Satellite System), a GNSS equipment with centimeter-level positional accuracy is generally mounted on the UAV platform as a cooperative target in MCE accuracy assessment test by testers, as illustrated in Fig. 2. The raw information on time-variant position of UAV flight is then transmitted to a receiver aboveground in real time
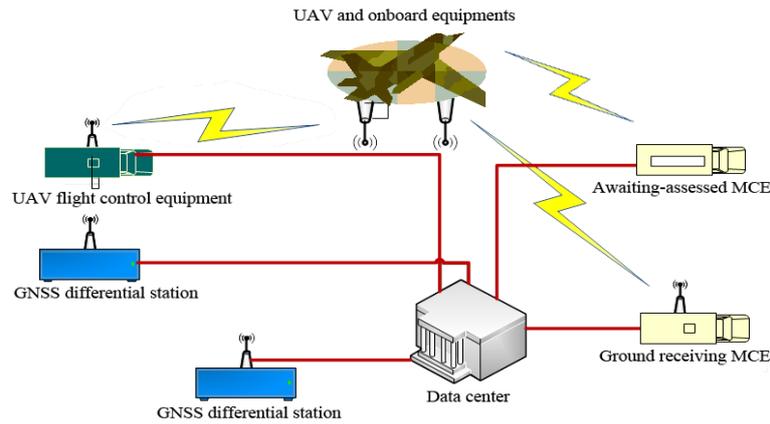
Fig. 2. Schematic diagram of some participant equipments in accuracy assessment test.

via a S-band telemetry transmitter. Finally, high-precision differential positioning information is achieved in ground data center, and acts as a comparison standard with regard to the MCE tracking measurements. Thereby, a statistical analysis of the comparison residuals can be conducted so as to accomplish MCE accuracy assessment.

Suppose a tracking radar is the awaiting-assessed MCE without loss of generality. And some other aided testing MCEs are assumed to have been all tested aforehand to reach their ideal state in test. Before starting a comparison analysis, the radar tracking measurement data, along with the standard data for comparison, must be in a unified time system as far as possible. However, there are several potential causes for disunity of time systems of different MCEs, as follows

(1) The precision of relevant time systems unification is finite in the strict sense. Even if the timing accuracy of the satellite positioning system is nominally up to a nanosecond level, it may appear that the actual time system deviation reaches tens of milliseconds since some differences always exist in signal processing, and integration of terminal timing systems for different manufacturers;

(2) Not all of MCEs, in practice, adopt the same strategy to record time information on measurement data. Some mark time on the original recorded data, while others add time stamps to the post-processing data;

(3) Uncertain time delays in data processing, transmission, and other links may cause the ideal time synchronization not to be achieved even in the case of ideal time system.

Hence, it is of course significant to obtain a well estimate of time deviation between the radar system and other ideal MCE system, and unsatisfactory estimation will directly affect the final success of accuracy assessment test.

In short, what problems to address in this paper are

Problem 1. How to restrain the interference of complex outliers in radar tracking measurements so as to ensure the reliability of applications?

Problem 2. How to obtain a well estimate of the time deviation so as to conduct the comparison analysis under an aligned time system?

3.2 Problem solution

In this part, the two preceding problems which seem uncorrelated are considered together in the context of accuracy assessment test. And an approach using the pitch angles of UAV flight relative to the tracking radar is designed aimed at the estimation of time system deviation while regarding the elimination of outliers in tracking measurements.

In order that the standard data mentioned earlier has a stable and reliable quality for comparison, the UAV route, in accuracy assessment test, is generally designed in form of a horizontal straight line, and UAV flies at a constant speed to stay in a cruise state. Additionally, it remains to require that the tracking capability of the awaiting-assessed radar must be taken into account when we perform planning operation on the route. Subsequently, according to a feature that a turning point of the tracking parameter emerges as the flying-along-route UAV passes a specific waypoint, whose projection on ground is closest to the radar, we get an idea of obtaining the time deviation estimation.

As shown in Fig. 3, the position of radar is denoted by P_o, P_c represents the aforementioned waypoint in route. D is the distance between P_o and the projective point of P_c. θ is the theoretical angle of UAV relative to radar in pitch at time t. h is the flight height of UAV relative to the ground. Let t_0 denote the starting time of UAV entering the straight-line route and t_c be the time of UAV reaching the waypoint P_c. At this point θ can
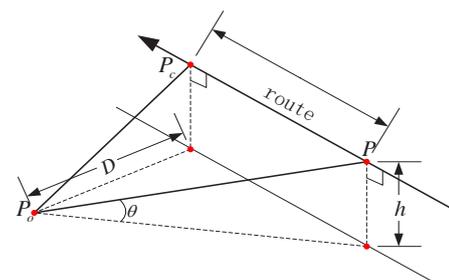

Fig. 3. Diagram of measurement parameter for the awaiting-assessed radar.

be expressed in the following mathematical formula

$$\sin \theta(t) = \frac{h}{\sqrt{v^2 \cdot (t + t_0 - t_c)^2 + D^2 + h^2}} \qquad (8)$$

then we have

$$\sin^{-2} \theta(t) = \frac{v^2 \cdot (t + t_0 - t_c)^2}{h^2} + (1 + \frac{D^2}{h^2}) \qquad (9)$$

As can be seen from formula (9), $\sin^{-2} \theta(t)$ is explicitly characterized by a quadratic polynomial with respect to the parameter t. This heuristically allows us to make data trend modeling by least square method with $K = 2$. Let us refer to $\sin^{-2} \theta_1(t_1)$ and $\sin^{-2} \theta_2(t_2)$ as another two transformed forms of the standard data for comparison, and radar tracking measurements for assessment, respectively. In the presence of time system deviation the two corresponding theoretical curves are indicated in Fig. 4, where the extreme points on two curves uniformly stand for a specific time at which UAV reaches the waypoint P_c.
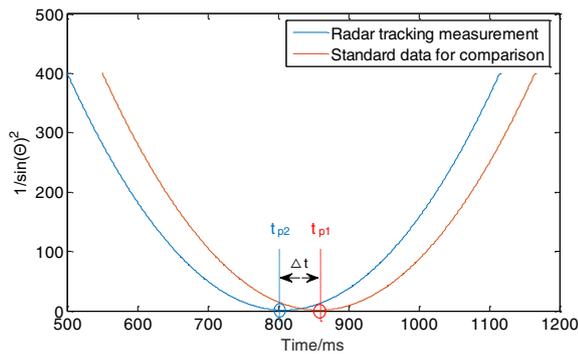

Fig. 4. Theoretical curves of the two $sin^{-2}\theta(t)$ in presence of time deviation.

Following this invariant uniformity, it suffices to consider the deviation between the two associated times of such extreme points, which can be readily calculated by the following steps

Step 1. Use two quadratic polynomial curves to fit $\sin^{-2} \theta_1(t_1)$ and $\sin^{-2} \theta_2(t_2)$, respectively;

Step 2. Find approximate times for extreme points of the two curves above, denoted by t_{p1} and t_{p2}, respectively;

Step 3. Calculate the time system deviation by $\Delta t = t_{p1} - t_{p2}$.

3.3 Solution flow

To get the standard data for comparison, the high-precision GPS information on position of UAV flight is obtained by the real-time kinematic (RTK) positioning technique in data center, and converted to the pseudo pitch angles of the tracking radar after a coordinate transformation. This mathematically results in a standard data sequence, denoted by $E_{S\tan}(t_k)$ $(k = 1, 2, 3, ..., N_{S\tan})$

in discrete form for brevity. In parallel, the radar tracking measurements in pitch are denoted by $E_{Rad}(t_n)$ $(n = 1, 2, 3, ..., N_{Rad})$. With some underlying disturbances inside and outside, $E_{Rad}(t_n)$ can be expressed in the following form

$$E_{Rad}(t_n) = E_{Rad}(t_n)^* + s_n + \omega_n + \delta_n \qquad (10)$$

where $E_{Rad}(t_n)^*$ represents the actual pitch angle sequence; s_n is the systematic error which can be removed by operations such as modeling, compensation, and calibration; ω_n is the random error which can be restrained by filtering, and smoothing methods; δ_n signifies some complicated types of outliers, and the outliers elimination, as one of the two problems stated previously, is concerned in this paper.

Recalling the content in Problem solution, the estimate of time system deviation can be obtained analytically with the proposed joint method, as steps presented below

Step 1. Conduct data trend modeling for $\sin^{-2} E_{S\tan}(t_k)$ by least square method. We derive a fitting quadratic polynomial $f_{S\tan}(t_k) = a_1 t_k^2 + a_2 t_k + a_3$. Then the time corresponding to the extreme point is $t_{S\tan}^P = -a_2 / 2a_1$;

Step 2. Group $\sin^{-2} E_{Rad}(t_n)$ into 2^m data sequences via m recursive extractions, as depicted in Fig. 1;

Step 3. Conduct data trend modeling for each above data sequence by least square method, and subtract each own trend polynomial. We get 2^m data sequences without data trends.

Step 4. Use Pauta criterion to detect the resulting 2^m data sequences above. Eliminate the discovered outliers and correct them as 0;

Step 5. Add each own data trend for the newly resulting 2^m data sequences above and regroup them. We get a new data sequence $\sin^{-2} \hat{E}_{Rad}(t_n)$. Conduct data trend modeling for it by least square method. We derive another fitting quadratic polynomial $f_{S\tan}(t_n) = b_1 t_n^2 + b_2 t_n + b_3$. Then the time corresponding to the extreme point is $t_{Rad}^P = -b_2 / 2b_1$;

Step 6. Calculate the time system deviation by $\Delta t = t_{S\tan}^P - t_{Rad}^P$.

3.4 Simulation and analysis

To verify the effectiveness of the proposed method applied for estimation of time system deviation, a simulation scenario for accuracy assessment test is created and carried out. Suppose the test for tracking radar is executed on a virtual testing outfield. Some other MCEs participate in the test as aided testing equipments in different locations, which are not displayed for simplicity.
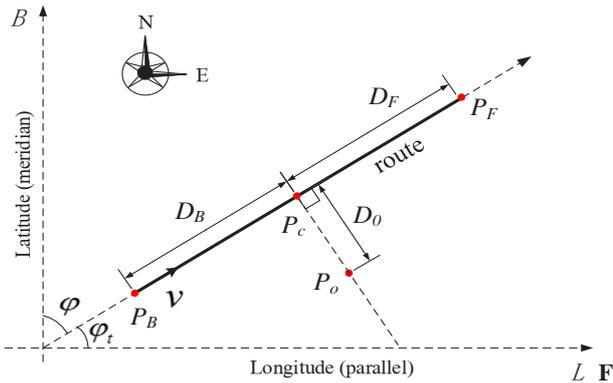
Fig. 5. Accuracy assessment test in the horizontal plane (P_B—initial route point, P_F—final route point, P_O—radar location).

For the planning of UAV route, Fig. 5 shows its schematic diagram in the horizontal plane. At 12:01 PM, UAV flies along the planned route $P_B P_F$ at a constant cruising speed $V = 50m/s$ from southwest to northeast on the side of radar. The flight height is $h = 1500m$. The flight angle is $\varphi = 45°$. The length of route is $P_B P_F = 10km$, and herein we postulate that the projective point P_c is selected as half of the route (i.e., $D_B = D_F = 5km$). The radar, denoted as P_o in Fig. 5, is located a distance of $D_o = 5300m$ away from the route.

For the awaiting-assessed radar, suppose the measurement noise listed in technical specification for angular measurements is zero-mean white Gaussian noise with random walks of $\sigma_E^2 = (300\mu)^2 \text{rad}^2$ in pitch, and $\sigma_A^2 = (300\mu)^2 \text{rad}^2$ in azimuth, respectively. The sampling frequency is $f_s = 20\text{Hz}$. As for the GPS information on UAV position generated by the real-time kinematic (RTK) positioning technique in ground data center, the positional accuracy along three axes in the earth-centered earth-fixed coordinate system are assumed to be all equal to $0.01m$.

Finally, suppose the radar and ground data center work under a nominally unified GPS time reference, but, owing to some of the causes explained previously, there exists a time deviation between them, which here is assumed as $\Delta t = 35ms$. To ensure that the tracking measurement data for comparison analysis really comes from a data collection regarding the straight UAV flight phase, we selectively consider a 7km range in route (i.e., a length of 3.5km for the front, and back sub-routes, respectively) since UAV usually calls for adjustment on its own attitude to fly into and off a route at the start and end phase, respectively. With some various disturbances, there are some complex outliers in measurements, and they appear more intensively than others as UAV approaches P_c. This is mainly due to instable tracking

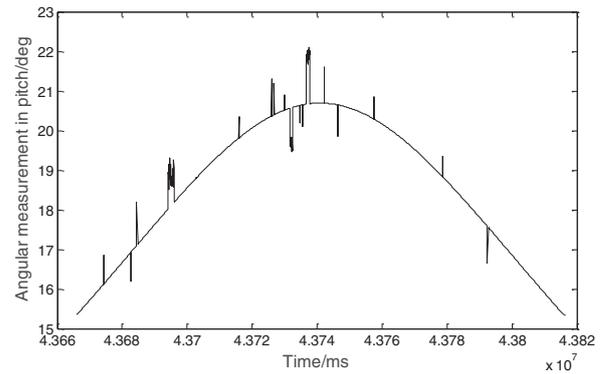quality induced by the increasing azimuth velocity to reach its limit over time.

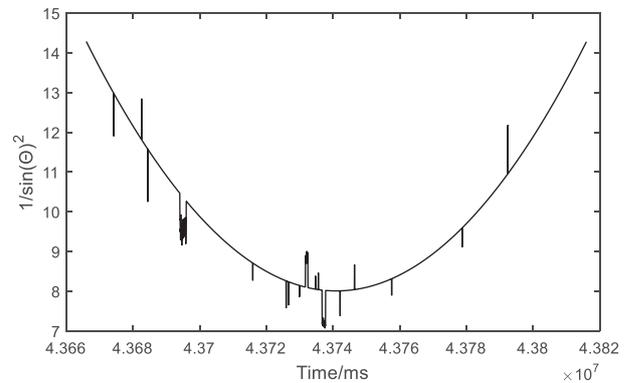**Fig. 6.** Curve of radar tracking measurement in pitch.

Fig. 7. Curve of $sin^{-2}\theta(t)$ by converting the form of radar tracking measurement in pitch.

Figs. 6 shows the changing curve of radar tracking angular measurement in pitch. It can be seen that the curve has a distinct turning, but, in the strict sense, it has no obvious feature of polynomial with an explicit degree. In this case, we convert the form of pitch angle measurements to what as formula (9), and then a curve with the feature of quadratic polynomial for $\sin^{-2} E_{Rad}(t_n)$ is derived, as depicted in Fig. 7.

Figs. 8, 9, and 10 show the curves of $\sin^{-2} \hat{E}_{Rad}(t_n)$ after conducting $m = 1$, 2 and 3 recursive grouped extractions, and outliers elimination, respectively. A quick look at the three Figs reveals that the number of outliers decreases as m increases, thereby showing the effectiveness of the proposed method to outliers elimination. For a clear and fair comparison on the estimation effect, we make 200 independent Monte Carlo runs. Table 1 provides the average estimates of time system deviation by different methods. From Table 1 it is apparent that the estimate, after more recursive grouped extractions, comes closer to the actual value we

Table 1. A comparison of average estimates of time system deviation by different methods.

Methods	Straight-forward polynomial curve fitting method	Joint data processing method by $m = 1$	Joint data processing method by $m = 2$	Joint data processing method by $m = 3$
Average estimates	421.28ms	46.34ms	40.60ms	37.16ms

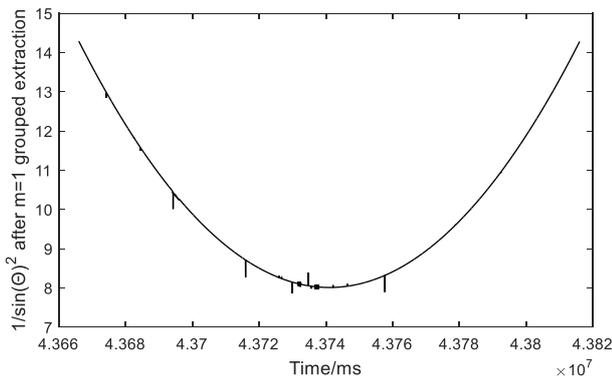| Deviation from $\triangle t$=35ms we set	386.28ms	11.34ms	5.60ms	2.16ms

Fig. 8. Curve of $sin^{-2}\theta(t)$ after m=1 grouped extraction and outliers elimination.
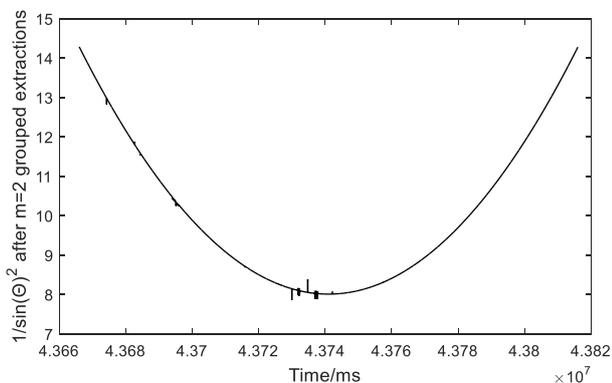

Fig. 9. Curve of $sin^{-2}\theta(t)$ after m=2 grouped extractions and outliers elimination.
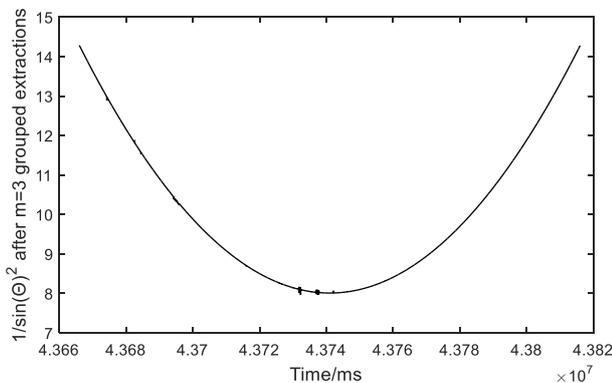

Fig. 10. Curve of $sin^{-2}\theta(t)$ after m=3 grouped extractions and outliers elimination.

set previously. Moreover, the joint method performs noticeably better compared to a straight-forward polynomial curve fitting method, whose estimate being considerably far away from the actual value indicates the adverse impact of complex outliers in case of ignoring outliers elimination. Consequently, the proposed method is feasible to the estimation of time system deviation, and thus helpful to the accuracy assessment test.

4 Conclusion

A joint data processing method, constituted by grouped data extraction, data trend modeling, and outliers detection, was presented to eliminate complex outliers in UAV tracking measurements, and sought to estimate the time system deviation between different measurement and control equipments. A simulation on accuracy assessment test for radar was implemented, and results demonstrated that the proposed joint method was effective to outliers elimination. What's more, the method helped obtain a well estimate of time system deviation, thereby facilitating the accuracy assessment test for the awaiting-assessed equipment.

References

1. M. Zieja, W. Kosiński, J. Wójcik, J. Kon, **20**, 501-510 (2013)

2. V. Chandola, V. Kumar, ACM. Comp. Surv, **41**, 3 (2009)

3. S.H. Cui, M. Wang, J. Wang, *Velocity Tracking Measurement Data Processing* (NDI Press, Beijing, 2017)

4. B.W. Hou, J.Q. Wang, X,Y. Zhou, D. Li, Z.M. He, Aero. SH, **35**, 91-100 (2018)

5. H.B. Zhou, D.D. Chen, Y. Peng, B, Shipb. Elec. Coun, **36**, 84-87 (2013)

6. Z. Liu, J.F. Wang, Y. Wu, *Proceeding of the 32th Chinese Control Conference*, 4652-4655 (2013)

7. Y.P. Duan, EMT, **35**, 40-42 (2012)

8. D. Xiao, B. Shang, B. Wu, *International Conference on Fuzzy Systems & Knowledge Discovery* (2015)

9. P. Theo, *Algorithms for Graphics and Image Processing* (CS Press, 1982)

10. M. Fischler, *Commun Acm*, **24** (1981)

11. W. Zeng, S. Feng, Inf. Sci, **272**, 73-83 (2014)

12. M. Brucato, J.F. Beltran, A. Abouzied, *Proceedings of the Vldb Endowment*, **9**, 576-587 (2016)

13. Y.J. He, L.H. Wu, J.C. Ren, *Introduction to Range Measurement and Control* (SDU Press, Jinan, 2009)