Analysis of statistical methods for outlier detection in telemetry data arrays, obtained from “AIST” small satellites

M A Ivanushkin, S S Volgin, I V Kaurov and I S Tkachenko

Department of Space Engineering, Samara National Research University, 34, Moskovskoye shosse, Samara, 443086, Russia

E-mail: nku-mka@yandex.ru

Abstract. Due to various disturbing factors affecting data reception, telemetry information often contains outliers and lost entries. Detection of outliers and anomalies is one of the key problems of satellite data analysis. A comparison of several statistical methods of outlier detection in telemetric information on the power supply system, obtained from the “AIST” small satellites, was carried out for the purpose of increasing the accuracy of satellite’s systems monitoring based on the telemetry data. The structure of telemetry from the power supply and thermal control systems of the “AIST” small satellites was analyzed. Several statistical methods for outlier detection were studied. Methods, described in the paper, allow to detect fast acting changes in telemetry data from “AIST” small satellites.

1. Introduction

Telemetry data is the main source of information about the events, occurring onboard of an operating spacecraft. Through the process of telemetry analysis, the emergency markers are identified, the satellite’s operation is being monitored, the platform’s performance is evaluated and failures and malfunctions of the onboard equipment are detected.

Telemetry data is transmitted to Earth during the communication sessions between the satellite and the ground control center [1]. However, the received data is prone to having outliers and missing data entries due to the influence of numerous disturbing factors, such as weather conditions and equipment malfunctions. The paper [2] presents a study of methods for imputing missing data entries into telemetry arrays and proposes an algorithm for imputing data into the telemetry from “AIST” small satellite. The paper, however, does not cover the problem of outlier detection.

Detection of outliers and data anomalies is one of the main problems in satellite data analysis. There are several definitions of the term “outlier”. In [3], an outlier is “an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data”, in [4], an outlier is “an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism”.

In [5] a well-structured and elaborate review for outlier and anomaly detection in data arrays is given. Due to the diversity of data types, researchers propose their own methods and algorithms for detecting outliers [6], [7]. According to [5], most of the methods can be divided into classification (neural networks), nearest neighbor, cluster and statistical, as well as methods related to information theory and spectral theory.

This paper discusses application of a number of parametric statistical methods for analyzing the accumulated telemetry information from the “AIST” small satellites. Detection of outliers in the
telemetry information will allow to determine the anomalous readings from onboard equipment sensors, which may indicate a malfunction in any of the spacecraft systems, as well as determine outliers caused by errors when receiving telemetry information, which will reduce the error in assessing the state of the spacecraft onboard systems.

2. Structure of the telemetry data
Telemetry data, received from the “AIST” small satellite [8], is an array of measurements of main parameters, defining the state of the spacecraft. Telemetry data is transmitted to the ground station as a series of frames. Structurally, these frames are series of bites of predefined length with a header in the beginning and a control sum at the end. Thus, telemetry information is a consistently measured data set at equal intervals of time, which makes it possible to characterize telemetry information as a time series.

Let us group the telemetric information based on the purpose of the “AIST” small satellite systems. Table 1 lists parameters of the power supply system and table 2 lists the parameters related to the thermal control system.

### Table 1. List of parameters of the power supply system.

| Parameter                        | Unit      | Example value |
|----------------------------------|-----------|---------------|
| Date and time                    | DD.MM.YYYY hh:mm:ss | 01.04.2019 0:00 |
| On-board voltage                 | V         | 14.04         |
| On-board current                 | A         | 0.21          |
| Solar array current              | A         | 0.93          |
| Electric current of subsystems 1-17 | A | 0            |

Power control system contains date and time of readings, values of the main parameters, onboard voltage and current, solar array current and electric currents of onboard subsystems of the satellite (17 in total).

### Table 2. List of parameters of the thermal control system.

| Parameter                        | Unit      | Example value |
|----------------------------------|-----------|---------------|
| Date and time                    | DD.MM.YYYY hh:mm:ss | 01.04.2019 0:00 |
| Readings from temperature sensors TR1-TR16 | C° | 66 |
| Battery temperature              | C°        | 55            |
| Telemetry controller temperature | C°        | 63            |
| Navigation software temperature  | C°        | 41            |
| Temperature of the second transmitter | C° | 41 |
| Temperature of the first transmitter | C° | 42 |
| Temperature of the onboard computer | C° | 60 |

Telemetry of the thermal control system contains time and date of measurements, readings from temperature sensors, located on external and internal surfaces of the satellite, and also temperatures of certain subsystems, such as battery, telemetry controller, navigation equipment, transmitters and onboard computer. The results of the analysis of the operation of the thermal control system of the “AIST” small satellite are presented in [9, 10, 11].

3. Analysis of statistical methods of outlier detection
In this section we will review several statistical methods on their ability to detect outliers in a test telemetry set, based on the power supply system telemetry from the “AIST” small satellite. The test set was constructed from a regular telemetry file by adding several outlier values to it.
The data array contains recordings of three parameters of the power supply system: on-board voltage (Ubs, V), on-board current (Ibs, A), solar array current (Isun, A). On the voltage chart, sporadic data changes are noticeable, which are beyond the allowable values. Sharp current surges of the on-board network up to 1.5 A, with a duration of about 5-10 minutes, are visible on the graph of the on-board network current; such jumps make it possible to judge that some onboard equipment was operating at that time.

The graph of the current of the solar cells is observed to change from 0 to 2 A, which indicates the frequency of the apparatus entering the Earth’s shadow as it moves in orbit, there is also one sharp jump to 4 A.

3.1. Standard score
Standard score (or z-score) is a measure of relative spread of the observed or measured value, which shows how many standard deviations make up its spread of the relative average value [12]. Z-score is defined as follows:

\[ Z_i = \frac{x_i - \bar{x}}{s}, \]

where \( \bar{x} \) and \( s \) indicate the mean and standard deviation of the sample, respectively.

In order to find outliers in telemetry arrays, the z-score is calculated, which requires centering and normalization of a random value. Then the value points that are too far from the zero are identified as the outliers. The threshold value of 2.5 or -2.5 was used, therefore, if the z-score was below -2.5 or above 2.5, it was identified as an outlier.

This method has a disadvantage – the outliers influence the mean values and the standard deviation. Considering that the z-score method relies on these values for measuring the central tendency and dispersion, the results could be inadequate.

3.2. Modified Z–score method
Modified z-score, described in [13] uses median and the median of the absolute deviation of the median (MAD) instead of the mean and standard deviation of the sample thus reducing the influence of the outliers on the score.

Modified z-score is computed as follows:

\[ M_i = \frac{0.6745(x_i - \bar{x})}{MAD}, \]

where 0.6745 is the 0.75-th quartile of standard normal distribution, that the MAD converges to for large normal data.

3.3. The interquartile range
InterQuartile range (IQR) is a measure of statistical dispersion, much more resilient towards outliers, being equal to the difference between 75th and 25th percentiles, or between upper and lower quartiles.

According to [14] any data point, laying in the range 1.5*IQR lower than first quartile \( Q_1 \), or 1.5*IQR higher that third quartile \( Q_3 \), is an outlier:

\[ x_{\text{outlier}} = (x_i > Q_3 + IQR \cdot 1.5) \lor (x_i < Q_1 - IQR \cdot 1.5), \]

where \( x_{\text{outlier}} \) - outlier data point, \( IQR = Q_3 - Q_1 \).

4. Results
The test telemetry set, based on the power supply system telemetry from the “AIST” small satellite was analyzed using the abovementioned methods.

Table 3 shows outliers, detected using z-score method for the on-board voltage Ubs, V, demonstrating that according to the method values Ubs, V 7,15, 12,0 and 3,61 are outliers, as their z-score is less than 2,5.
Table 3. Outliers, detected with the z-score method.

| Time, min | Datatime          | Ubs,V | z         | outlier |
|-----------|-------------------|-------|-----------|---------|
| 106       | 18.06.2019 2:58   | 7.15  | -15.3203  | True    |
| 107       | 18.06.2019 2:59   | 7.15  | -15.3203  | True    |
| 507       | 18.06.2019 9:39   | 12    | -5.02767  | True    |
| 723       | 18.06.2019 13:15  | 3.61  | -22.8328  | True    |

Therefore, z-score method considers Ubs values lower than 13.38 and higher than 15.68 to be outliers. For on-board current Ibs, the values lower than 0.11 and higher than 0.29 A are considered as outliers. For the solar array current, according to the z-score method every value above 1.7 A is an outlier. Table 4 shows outliers, detected using modified z-score method in the on-board voltage Ubs, V data array.

Table 4. Outliers, detected using modified z-score method.

| Time, min | Datatime          | Ubs,V | z         | outlier |
|-----------|-------------------|-------|-----------|---------|
| 106       | 18.06.2019 2:58   | 7.15  | -37.772   | True    |
| 107       | 18.06.2019 2:59   | 7.15  | -37.772   | True    |
| 353       | 18.06.2019 7:05   | 15.61 | 6.122384615 | True    |
| 354       | 18.06.2019 7:06   | 15.61 | 6.122384615 | True    |
| 507       | 18.06.2019 9:39   | 12    | -12.60796154 | True    |
| 723       | 18.06.2019 13:15  | 3.61  | -56.13915385 | True    |
| 1293      | 18.06.2019 22:45  | 15.68 | 6.485576923 | True    |
| 1362      | 18.06.2019 23:55  | 13.64 | -4.098884615 | True    |
| 1363      | 18.06.2019 23:56  | 13.58 | -4.410192308 | True    |
| 1364      | 18.06.2019 23:57  | 13.45 | -5.084692308 | True    |
| 1365      | 18.06.2019 23:58  | 13.38 | -5.447884615 | True    |
| 1366      | 18.06.2019 23:59  | 13.45 | -5.084692308 | True    |

According to the modified z-score method Ubs values lower than 13.78 and higher than 14.76 to be outliers. For on-board current Ibs, the values lower than 0.11 and higher than 0.29 A are considered as outliers. For the solar array current, according to the modified z-score method every value above 1.7 A is an outlier.

Analysis of the Ubs telemetry data with the InterQuartile range method shows that values lower than 13.64 and higher than 14.76 to be outliers. For on-board current Ibs, the values lower than 0.19 and higher than 0.27 A are considered as outliers. For the solar array current, according to the IQR method every value above 1.79 A is an outlier.

Between the compared methods, modified z-score and the IQR have shown the best and the most similar results. Figure 1 shows the test telemetry array with the outliers, detected with modified z-score and IQR methods marked with dots.

It should be noticed, that by analyzing the data, presented in Figure 1, the modified z-score method and IQR identified the data instances that appear to be caused by errors in the reception of telemetry information as outliers. At the same time, these methods have also marked data instances that are obviously not outliers, namely, increased on-board current values (Ibs, A) corresponding with the spacecraft communication session.

Figure 2 shows the test telemetry from the thermal control system of the “AIST” small satellite with the outliers, detected with modified z-score and IQR methods marked with dots.

The periodic change in the readings from thermal sensors that can be noted on figure 2 is caused by rotation of the satellite around its center of mass. All deviations were successfully detected as outliers by the modified z-score and IQR methods.
5. Conclusion
The article has presented a structural analysis of telemetry array of the power control system and thermal control system of the “AIST” small satellite. Statistical methods, capable to detect outliers in arrays of data, were analyzed.
Methods, considered in the article, allow finding uncharacteristic data values in telemetry arrays. However, they often mark valid data instances as outliers, as it has been shown on the example of the satellite’s on-board current values spiking during the communication sessions. Thus, it can be concluded that such methods are only suitable for studying telemetric information data, identifying data instances that may require specific attention, be it an outlier or a sudden change in the operating parameters of the onboard equipment. For further, more detailed analysis of telemetric information data, it will obviously be necessary to use neural network and cluster methods in order to effectively detect outliers and anomalies.

References
[1] Tkachenko S et al 2017 Experience in Operating a Constellation of AIST Type Small Satellites with Samara University Ground Control Center and Perspective of Managing a Constellation of Satellites using a Distributed Network of Control Centers Based on Multiagent Approach Procedia Engineering 185 359–365
[2] Salmin V V, Ivanushkin M A, Volgin S S and Tkachenko I S 2018 Processing of telemetry data arrays for “AIST” small satellites using the methods of imputing missing data AIP Conference Proceedings 2046
[3] Barnett V and Lewis T 1994 Outliers in Statistical Data 3rd edition (Chichester: John Wiley and Sons) 584
[4] Hawkins D M 1980 Identification of Outliers (London, New York: Chapman and Hall) 188
[5] Chandola V, Banerjee A and Kumar V 2009 Anomaly Detection: A Survey ACM Computing Surveys 41
[6] Zanina M A, Belov A A and Volvenko S V 2018 Estimation of Accuracy of Algorithm for Measuring Radiofrequency Pulse Parameters 2018 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech-2018)
[7] Ovsyannikova A S, Lavrenyuk I I, Makarov S B and Wei X 2019 Error estimation of discrete formation method for optimal FTN signals J. Phys.: Conf. Ser. 1236 012083
[8] Tkachenko S I, Salmin V V, Semkin N D, Kurenkov V I, Abrashkin V I, Prokhorov A G and Tkachenko I S 2010 The design shape and basic performances of small satellite of SSAU and SRP SRC “TSSKB-PROGRESS” Bulletin of the SSAU 2 (22) 154–156
[9] Tkachenko S I, Salmin V V, Tkachenko I S, Kaurov I V and Korovin M D 2017 Verifying Parameters of Ground Data Processing for the Thermal Control System of Small Spacecraft AIST Based on Telemetry Data Obtained by Samara University’s GCS Procedia Engineering 185 205–211
[10] Tkachenko S I, Salmin V V, Tkachenko I S, Safronov S L, Kaurov I V, Korovin M D, Ivanushkin M A and Volgin S S 2016 Improving ground thermal vacuum testing for small satellites of the "AIST" family CEUR Workshop Proceedings 1638 691–699
[11] Tkachenko S I, Salmin V V, Tkachenko I S, Safronov S L, Kaurov I V, Ivanushkin M A, Volgin S S and Korovin M D 2016 AIST small satellites’ family, a joint project between RSC progress and SSAU, thermocontrol system performance analysis Proceedings of 2015 International Conference on Mechanical Engineering, Automation and Control Systems
[12] Kreyszig E 1979 Advanced Engineering Mathematics 4th edition (New York, London: John Wiley and Sons) 1032
[13] Iglewicz B and Hoaglin D 1993 How to Detect and Handle Outliers (Milwaukee, Wis.: ASQC Quality Press)
[14] Tukey J 1977 Exploratory Data Analysis (Reading, Mass.: Addison-Wesley Publishing Company) 688