IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON INTELLIGENT AND COGNITIVE TECHNIQUES FOR INTERNET OF THINGS

Received November 28, 2019, accepted December 22, 2019, date of publication December 25, 2019,
date of current version January 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2962235

Study on the Satellite Telemetry Data Classification Based on Self-Learning

PENG WAN, (Student Member, IEEE), YAFENG ZHAN, AND WEIWEI JIANG, (Student Member, IEEE)

Space Center, Tsinghua University, Beijing 100084, China
Beijing National Research Center for Information Science and Technology (BNRist), Beijing 100084, China

Corresponding author: Yafeng Zhan (zhanyf@tsinghua.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61971261 and Grant 61671263, and in part by the Tsinghua University Independent Scientific Research Project under Grant 20194180037.

ABSTRACT Since great redundancy of telemetry data of spacecraft, telemetry data compression is a good solution for the limited bandwidth and contact wireless links. It is important to obtain accurate data characteristic firstly. State-of-the-art machine learning methods work well on data mining and pattern recognition under conditions of the given test data set, which could be used as the available tools for post-event data processing and analysis, such as trend forecasting and outlier detection, but they have not provided the proper solution from the source on-board. In this paper, four base classes of the telemetry data are suggested and studied through the time series feature and information entropy analysis, then a new on-board lightweight self-learning algorithm named Classification Probability calculation - Window Step optimization (CP-WS) is proposed to obtain the class features and make the decision of each single parameter from the continuous discrete telemetry time series. Simulation results show that, our algorithm correctly classifies the simulation and real mission data into the appropriate base class with advantages of high classification accuracy as 100% and adaptive computational complexity from $O(L^2)$ to $O(L)$, which could be used in satellite on-board data compression for space-to-ground transmission, especially for the deep space explorers to save important status with less on-board storage space.

INDEX TERMS Telemetry data, self-learning, time series, information entropy, classification, sliding window.

I. INTRODUCTION

In current space missions, such as manned-space stations, earth observation satellites, deep space explorers, etc., the on-board telemetry plays an important role in helping Mission Control Center (MCC) to monitor the platform status, discover the abnormal phenomena, and acknowledge the remote control feedbacks. However, more complicated spacecraft with advanced applications challenges current space telemetry system with the conditions of narrow wireless bandwidth and fixed-length frame telemetry, which is difficult to transmit the increasing telemetry volumes. Meanwhile, the discontinuous short-term contacts between spacecraft and ground stations restrict the data transmission capability, which restricts the evaluation of parameter's long-term behaviors. In addition, the monitoring and interpretation of vast telemetry parameters also consume lots of manpower, which requires automatic assistance by machines.

As a fast-developing technology in recent years, machine learning (ML) related technologies have been studied widely in space missions with telemetry in recent years. Yairi *et al.* [1]–[3] studied on the satellite health monitoring based on the probabilistic clustering, dimensionality reduction, hidden markov, regression tree. Tariq *et al.* [4], Hundman *et al.* [5], and Fuertes *et al.* [6] studied on the spacecraft anomaly detection based on the K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Long Short-Term Memory (LSTM) with the exhaustive testing on the telemetry of Centre National d'Etudes Spatiales (CNES) spacecraft. Iverson *et al.* [7] and Robinson *et al.* [8] studied on the space operation assistant based on the data-driven and model-based monitoring techniques applied in several space missions. To sum up, the applications of machine learning in space missions mainly focus on the forecasting and outlier detection in order to provide the flight control procedure

The associate editor coordinating the review of this manuscript and approving it for publication was Min Jia.

with much more information about the spacecraft operation status, most of which are operated in MCC on ground without consideration of space-to-ground wireless communication limits. In addition, current self-learning research focuses on the telemetry behavior prediction, regardless of the specific feature differences of various telemetry types in time series analysis.

In this paper, the self-learning classification algorithm is studied comprehensively in order to achieve the on-board high accuracy telemetry classification with low computational complexity and low time latency, which could be used to obtain the telemetry features for data compression on time-dependent space environment and bandwidth-constrained wireless links. Section II introduces the main characteristics of satellite telemetry concerned with time series and information entropy. The self-learning classification model CP-WS for satellite telemetry data is described in Section III, including the Classification Probability (CP) calculation and Window Step (WS) optimization. Section IV gives the classification performance evaluations based on simulation data and real mission data. Section V concludes this paper.

II. TELEMETRY CHARACTERISTICS

A. TIME SERIES ANALYSIS

1) TIME SERIES FEATURES

Time series is the data set arranged at sequential time intervals [9], either one dimension or multiple dimensions. The behavior of time series could be influenced by many factors: some factors provide the long-term and decisive effects, which show trend and regularity; the other factors provide the short-term and non-decisive effects, which show some irregularity. Real-world time series data is usually composed of several factors as listed below:

- *Trend:* the behavior of time series shows a continuous upward, downward, or steady movement in a certain direction by time, which might be driven by the long-term factors.
- *Cyclic:* the behavior of time series shows a sequence of points circulating above and below the trend line, with regular changes lasting for a relatively long time.
- *Seasonal variation:* the behavior of time series shows a cyclic fluctuation in a fixed period affected by season. For example, the temperature of the satellite solar panel shows a cycle of 'seasonal' changes, that is, the temperature in the sunshine area rises while that in the shadow area decreases.
- *Irregular movement:* the behavior of time series shows irregular fluctuations due to accidental factors.

Time series could be classified into different types according to various criteria, such as: Dimensional criterion, i.e., Univariate time series vs. Multivariate time series; Continuous criterion, i.e., Discrete time series vs. Continuous time series; Statistical criterion, i.e., Stationary time series vs. Non-stationary time series. The low-order moments of time series are usually used to describe their Eigen statistical characteristics, such as mean, variance, and auto-covariance function. For discrete time series $\{X_t, t \in T\}$, X_t is a random variable with probability $p(X_t)$, in which:

a) Mean: Defined as μ_t with $\mu_t = EX_t = \sum_{t=-\infty}^{+\infty} p(X_t) X_t$, for time series $\{X_t, t \in T\}$.

b) Variance: Defined as σ_t with $\sigma_t^2 = DX_t = E(X_t - \mu_t)^2 = \sum_{t=-\infty}^{+\infty} p(X_t)(X_t - \mu_t)^2$.

c) Auto-covariance function: Defined as $\gamma(t, s) = Cov(X_t, X_s) = E(X_t - \mu_t)(X_s - \mu_s), \forall t, s \in T$.

For engineering data analysis, the weak stationary condition is usually used to test the stationarity of data. If the time series could satisfy the following three conditions, it can be judged to be weak stationary:

i $\forall t, j \in T, E(X_t) = E\left(X_{t-j}\right) = \mu = const$
ii $\forall t, j \in T, Var(X_t) = Var\left(X_{t-j}\right) = \sigma^2 = const$
iii $\forall t, j, s \in T, \gamma(t, t-s) = \gamma(t-j, t-j-s) = \gamma(s)$

2) TELEMETRY DATA FEATURES AND BASE CLASSES

The telemetry data with multiple parameters could be regarded as the typical high-dimensional discrete time series data set under fixed sampling time intervals in space missions. The stationary feature and trend behavior are widely studied to analyze the time series, which could be used to classify the telemetry data.

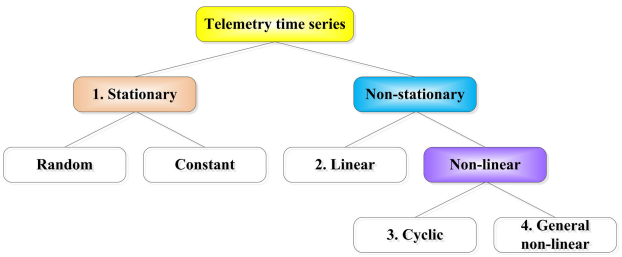


FIGURE 1. The classification tree of the satellite telemetry data.

The binary classification tree of the satellite telemetry data is shown in Figure 1, which could be divided into four base classes as Stationary (including constant), Linear, Cyclic, and General non-linear.

(1) Base class 1: Stationary time series

- Random data

The typical random telemetry data follows either Gaussian distribution or Poisson distribution, whose mathematical expression is $Y_t = \epsilon_t$.

For Gaussian distribution: Gaussian time series could be classified to stationary process obviously with the Eigen statistics as $E(Y_t) = \mu$, $Var(Y_t) = \sigma^2$, $Cov(Y_t, Y_s) = \gamma_{t-s} = 0$. Some zero-valued telemetry parameters in idle state are usually influenced by additive Gaussian white noises, which could be classified into random data with stationary feature.

For Poisson distribution: Some telemetry parameters have characteristics of randomness but not Gaussian. Take the SEU (Single Event Upset) as an example, the parameter usually records the number of the SEU appearance from the launch date till current time, which could be modeled by the homogeneous Poisson process with the Eigen statistics as $E(Y_t) = \mu = \lambda \cdot t$, $Var(Y_t) = \sigma^2 = \lambda \cdot t$, $Cov(Y_t, Y_s) = \gamma_{t-s} = \lambda \cdot \min\{s, t\}$. As described above, the Eigen statistics of Poisson time series are related to time, so that it should be classified into random data with non-stationary feature. However, the first-order difference of homogeneous Poisson process could be classified into random data with stationary feature, which is proved in Appendix A.

- Constant data

The mathematical expression of non-zero constant data is $Y_t = \mu$, the Eigen statistics are $E(Y_t) = \mu$, $Var(Y_t) = 0$, $Cov(Y_t, Y_s) = \gamma_{t-s} = 0$, all of which are time-independent. The mathematical expression of non-zero constant data with additive Gaussian white noises is $Y_t = \mu + \epsilon_t, \mu \neq 0$ and $\epsilon_t \sim N\left(0, \sigma^2\right)$, *i.i.d.*, the Eigen statistics are $E(Y_t) = \mu \neq 0$, $Var(Y_t) = \sigma^2$, $Cov(Y_t, Y_s) = \gamma_{t-s} = 0$, all of which are also time-independent. Therefore, constant data could be classified into stationary time series.

(2) Base class 2: Linear trend time series

In relatively short-term period such as one contact about 10 minutes, the behavior of some telemetry parameters shows an obvious linear trend. The mathematical expression of linear trend data is $Y_t = \mu + Y_{t-1} + \epsilon_t$, which could also be seemed as the difference stationary process. The iterative deduction result gives $Y_t = Y_0 + \mu \cdot t + \sum_{i=1}^{t} \epsilon_i$, and let initial value Y_0 equals to 0, we have combined with one deterministic time trend and a random walk process. The mean value of such time series is $E(Y_t) = \mu \cdot t$, which is time-dependent. Therefore, the linear trend time series should be classified to non-stationary type, which is dominated by the linear trend factor slope μ. The linear trend could be calculated by the linear regression method.

(3) Base class 3: Cyclic trend time series

In relatively long-term time period such as one day, the behavior of some telemetry parameters shows obvious cyclic trend. The mathematical expression of cyclic trend data is $Y_t = \mu_t + \epsilon_t = f(t) + \epsilon_t$, where $f(t)$ is the time dependent deterministic function, especially appearing cyclic feature. The mean value of such time series is $EY_t = Ef(t)$, which is time-dependent. Therefore, the linear trend time series should be classified to non-stationary type, which is dominated by the cyclic trend function $f(t)$. The cyclic trend could be calculated by the fast Fourier transform (FFT) or empirical mode decomposition (EMD) [10], [11].

(4) Base class 4: General non-linear trend time series

Actual time series data in engineering usually contain various behavior trends, therefore the data that could not be clearly classified to certain definite classification attributes to the general non-linear trend.

B. ENTROPY ANALYSIS

1) TIME SERIES ENTROPY

The information entropy analysis could be used to evaluate the changes of the effective information with time, which is important for the telemetry classification algorithm parameters' initialization, such as the sliding window size and step forwarding interval. In most time of its life the satellite is operating in the normal stable status, therefore the telemetry system on board could be seemed as the discrete stable data source. According to the satellite design, the numerical characteristics of the telemetry source could be expressed as $\{a_1, a_2, \ldots, a_K\}$ where K is the number of numerical space, and the time series of such source could be expressed as $\{\ldots, u_{-1}, u_0, u_1, \ldots, u_i, \ldots\}$.

According to stable stochastic process, all the finite-dimensional distribution probabilities of the source output data sequence are independent of the starting point of the time axis, that is, $P\left(u_i, u_{i+1}, \ldots, u_{i+N} = \vec{A}\right) = P\left(u_j, u_{j+1}, \ldots, u_{j+N} = \vec{A}\right)$, where $\vec{A}$ is a particular data sequence.

Assuming that the length of time series is limited to a finite number N, e.g. the time slots in one space-ground contact period or the sliding window size for data processing, then the finite time series is a random vector $(u_1, u_2, \ldots, u_N)$, whose information entropy could be expressed by the joint entropy $H(U_1, U_2, \ldots, U_N)$. Therefore, the statistical information entropy on each element in the numerical value space could be expressed by $H_N(U) = \frac{1}{N} \cdot H(U_1, U_2, \ldots, U_N)$. If the value of $H_N(U)$ converges to the finite ultimate value as $N \to \infty$, then it is defined as the entropy rate of the source, marked as $H_\infty(U)$ with the equation $H_\infty(U) = \lim_{N\to\infty} H_N(U)$. In current space missions, the telemetry data is usually measured by equally spaced discrete value in the range between upper and lower bounds, and we make the quantization number K as the maximum value for N.

2) TELEMETRY DATA ENTROPY

In order to obtain the high accuracy of data classification, the information entropy rate should be reduced as much as possible with less uncertainty, which requires appropriate number of the time slots in one contact or appropriate sliding window size marked as parameter W. The value of W should satisfy the classification requirement of all telemetry types, that is $W = \max\{W_i\}, i = 1, 2, 3, 4$. The mathematical analysis of the information entropy for each telemetry type is described as follows.

(1) *Base class 1:* Stationary time series

- Random data

For random data, the time series is modeled as the independent memoryless stable source, whose joint entropy could be seen as $H(U_1, U_2, \ldots, U_N) = \sum_{i=1}^{N} H(U_i) = N \cdot H(U_i), i = 1, 2, \ldots, N$. Therefore the information entropy rate for random data is $H_\infty(U) = H_N(U) = H(U_i) = H_1(U)$, whose entropy rate does not change with the increase of parameter N.

- Constant data

For consant data, the time series is sampled by only one constant numerical value without changes, therefore the entropy rate is zero.

(2) *Base class 2:* Non-stationary time series

For non-stationary time series, the statistical correlation exists between the historical and future data, which could be described by mathematical function $f(t)$ as the discrete stable source with memory. It could be proved that the information entropy rate $\lim_{N\to\infty} H_N(U)$ certainly exists. Under the circumstance of equally spaced sampling with time slot Δt, the appearance probability p for each quantization rank is proportional to the difference between neighbor ranks as $p \propto \frac{f(t)-f(\Delta t)}{\Delta t} = f'(t)$.

- Linear trend

For linear trend data, $f'(t) = const \neq 0$, therefore the appearance probability for any quantization rank in the numerical space is equal to each other as $p_i = p_j = const, i, j \in K, i \neq j$. Then we have,

$$H_N(U) = -\sum_{i=1}^{K-N+1} p_i \log p_i = \log_2 (K - N + 1) \quad (1)$$

where, $p_i = \frac{1}{K-N+1}$, $U = \{u_1, u_2, \ldots, u_N\}$ is the telemetry time series with N sample points, and the entropy rate $H_N(U)$ decreases with the larger N.

- Cyclic trend

For cyclic trend data, $f'(t) \neq const$, therefore the appearance probability for any quantization rank in the numerical space is not equal to each other, which could be calculated by the numerical difference for each time slot in any complete period.

Some special cyclic trend data could be calculated by mathematical formula, take the sine function as an example, $f(t = i) = \sin\left(\frac{\pi}{2} \cdot \frac{i}{K-1}\right), i = 0, 1, 2, \ldots, K-1$, where K is the number of ranks in numerical quantization space. Then we have the derivative of the sine function as $f'(t) \propto \cos\left(\frac{\pi}{2} \cdot \frac{i}{K-1}\right)$. Let $p_i = \beta \cdot \cos\left(\frac{\pi}{2} \cdot \frac{i}{K-1}\right)$ be the appearance probability for rank i, where β is the adjustable scaling factor, then we could get the information entropy rate in different time series with increasing length as follows, and the entropy rate $H_N(U)$ decreases with the larger N.

$$H_N(U) = -\sum_{i=0}^{K-N} p_i \log p_i \quad (2)$$

where, $p_i = \cos\left(\frac{\pi}{2} \cdot \frac{i}{K-1}\right) / \sum_{i=0}^{K-N} \cos\left(\frac{\pi}{2} \cdot \frac{i}{K-1}\right)$.

- General non-linear trend

For general non-linear trend data, $f'(t) \neq const$, therefore the appearance probability for any quantization rank in the numerical space is not equal to each other, which could be calculated by the numerical difference for each time slot in the entire contact or sliding window size. The entropy rate $H_N(U)$ for general non-linear trend data is $0 \leq H_N(U) \leq \log_2 K$ with the lower bound 0 and the upper bound $\log_2 K$, which is proved in Appendix B.

3) ENTROPY CALCULATION

There are several widely used entropy calculation algorithms for general non-linear trend data and most of the cyclic trend data, such as Approximate Entropy (ApEn) [12], Sample Entropy (SampEn) [13] and Permutation Entropy (PermEn) [14]. Note that ApEn is a model-independent statistic algorithm for distinguishing various classes of time series data, both deterministic and stochastic, in which lower ApEn values mean more persistence and correlation, larger ApEn values mean more independence. In this paper, we choose ApEn algorithm [12] to measure the time-dependent telemetry entropy.

III. CLASSIFICATION MODEL

A. RELATED WORKS

As we all known, it is very important to acknowledge the classification of the non-stationary trend in time series analysis. In order to transfer the time series data from non-stationary to stationary, traditional time series analysis only classify the data into deterministic or stochastic trend at first. Then based on the theory of Box-Jenkins [9], any non-stationary time series with deterministic trend could be expressed by the ARIMA(p, d, q) model. However, it is really time-consuming to calculate the difference order d in current time series analysis architecture, as shown in Figure 2. Meanwhile, the rough classification of deterministic vs. stochastic increases the number of model parameters and reduces the data transmission compression efficiency.

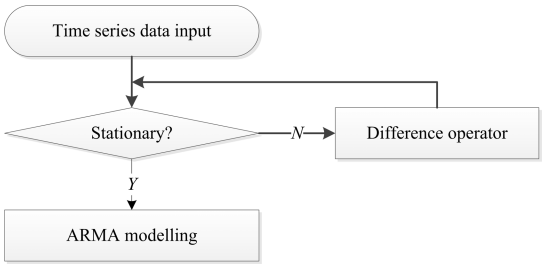


FIGURE 2. The traditional time series modeling procedure.

With regard to data-driven machine learning methods, due to the labor-consuming data labeling work, especially for the time-dependent telemetry time series, unsupervised clustering algorithms are usually used to analyze the features of diverse time series data [15] in most academic and industrial applications, such as making a comprehensive decision by multi-features sensing time series data and sparse feature learning in 5G communications [16]–[19]. However, such methods could only distinguish the time series with different behaviors, which could not provide us with enough available system parameters for further data transmission compression.

The satellite telemetry data usually have no good feature attributes along with a few outliers and random noises, it is difficult for current ARIMA and unsupervised clustering

methods to achieve high accuracy classification with less model parameters.

B. OUR PROPOSED MODEL

In this paper, we propose the CP-WS telemetry classification model with the WS (Window-Step) algorithm embedded in the CP (Classification Probability) algorithm, that is, CP classification utilizes the classification test probability (*cp*) to describe the membership level of the current time series belonging to a certain classification; and WS optimization self-learns proper parameters from the continuous telemetry data stream in order to improve the CP classification performance, which could be further used as the data compression parameters in telemetry transmission missions.

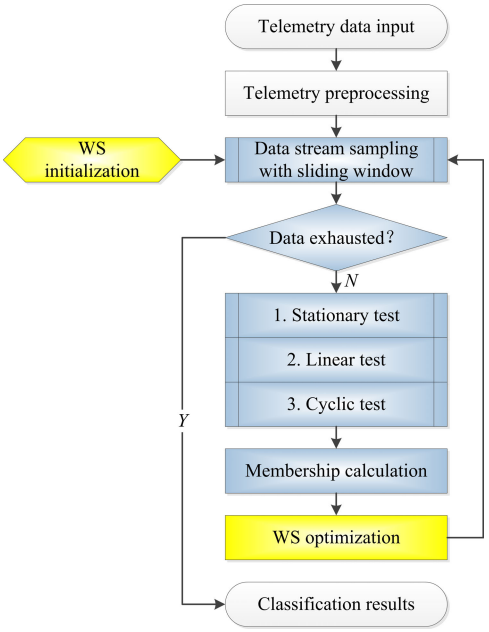


FIGURE 3. The workflow of our proposed CP-WS model.

Our proposed model is shown in Figure 3, the main steps are listed as follows:

1) Telemetry time series are generated as the data stream by the on-board telemetry subsystem.
2) A telemetry preprocessing block is used to truncate the data stream, eliminate local anomalies, and normalize the original telemetry data.
3) Initialize the WS optimization parameters such as window size and step interval, by the calculation results of ApEn algorithm.
4) Classification probability calculation:
   - Utilize the initial or optimized window size and step interval to frame the start and end point of the subset data of input telemetry time series;
   - If the size of the subset data is not empty, then make the stationary test, linear test and cyclic test in turn with the classification probabilities;
   - Calculate and determine the membership of this subset data, record it with the classification probabilities in local storage space labeled by current window size and step interval.
5) Optimize window size and step interval by minimizing the cost function with regularization term.
6) If the input data is not exhausted, repeat step 4) $\sim$ 5); else, calculate the classification result according to the storage records.

As shown in figure 3, a telemetry preprocessing block is introduced in our proposed CP-WS model in order to improve the classification efficiency, including:

- Data truncation: extract enough available data volumes from the infinite on-board telemetry time series data stream for the classification purpose;
- Outlier elimination: detect and repair the local anomalies with high efficiency and low time latency by the SCREEN algorithm [20], which would be further studied in our next paper;
- Data normalization: obtain the normalization time series $\hat{X}$ from the original time series X by formula $\hat{X} = \frac{X - X_{mean}}{X_{max} - X_{min}}$.

C. CP CLASSIFICATION

1) STATIONARY TEST

In stationary test, we utilize the Augmented Dickey-Fuller [21] to get the bool hypothesis test result h_a,

- $h_a = 1$ indicates rejection of the unit-root null in favor of the alternative model, which means the time series is stationary;
- $h_a = 0$ indicates failure to reject the unit-root null, which means the time series is non-stationary.

Then we choose the value of h_a as the stationary test result, that is, $p_1 = h_a$.

2) LINEAR TEST

In linear test, we utilize the residuals of linear fit method [22] to get the p-Value of the first order variable h_p, which indicates rejection probability of the linear trend null in favor of the alternative model, then we choose the value of $(1 - h_p)$ as the linear test result, that is, $p_2 = 1 - h_p$.

3) CYCLIC TEST

In cyclic test, we utilize the ensemble empirical mode decomposition (EEMD) [11] with adaptive controlled stopping condition to obtain the residual value. Then we choose the ratio rho of the residual value to extreme difference as the non-cyclic probability, and the cyclic test result is $p_3 = 1 - rho$.

4) CLASSIFICATION PROBABILITIES

Let cp_i be the *i-th* classification test result, with stationary test as 1-*st* test, linear test as 2-*nd* test, and cyclic test as 3-*rd* test.

Then we have the classification probability for each type as follows:

- base class 1: stationary probability: $cp_1 = p1$
- base class 2: linear trend probability: $cp_2 = p_2 \cdot (1 - p_1)$
- base class 3: cyclic trend probability: $cp_3 = p_3 \cdot (1 - p_2) \cdot (1 - p_1)$
- base class 4: general non-linear trend probability: the residual part, that is, $cp_4 = (1 - p_3) \cdot (1 - p_2) \cdot (1 - p_1)$

5) MEMBERSHIP DETERMINATION

The batch classification probabilities for each base class are obtained by iteration of the CP-WS algorithm, and in this paper the maximum statistical mean value is chosen as the criterion to determine the membership, that is, current time series data should be labeled as the base class with the largest statistical mean value among all the classification probabilities.

D. WS OPTIMIZATION

1) INITIALIZATION

Run the ApEn algorithm iteratively with the adjustable parameter **m** under certain stopping criterion, then the initial window size of our CP-WS algorithm could be set to multiple times of **m**, which might vary among different base classes.

Take the sine function discrete time series simulation data as an example, which would be described in section 4.1 in details, the ApEn values decrease sharply with the increase of parameter **m** ranging from 1 to N. In our simulation, the stopping criterion is set to 1% of the maximum ApEn values, which means the ApEn algorithm runs till the ApEn value calculated in current round is less than the stopping criterion, and the initial window size is set to **m**. Note that, the parameter **r** of ApEn is set to 10% of the extreme value difference, which might also vary among different base classes.

2) OPTIMIZATION OBJECTIVE

In order to improve the performance of our CP classification, the WS optimization is proposed to provide the proper window size and step interval for high accuracy classification. The computational complexity and time latency should also be considered for our algorithm to be applied in space, therefore two regularized factors are introduced into our model in order to constrain the cost function.

The cost function J_i of WS optimization for *i-th* base class ($i = 1, 2, 3, 4$ indicates stationary class, linear class, cyclic class, and general non-linear class respectively) is shown in equation (3).

$$J_i(X|W, S) = \frac{1}{N} \cdot \sum_{j=1}^{N} \left(1 - p_{ij}\right) + c \cdot W \cdot N + W/L \quad (3)$$

where,

- X: telemetry data time series, $X = x_n, n = 1, 2, \ldots, |x_n| < \infty$.
- L: the maximum window size of current data subset.
- W: window size, adjustable parameter, ranging from $W_0 \sim L$.
- S: step interval, adjustable parameter, ranging from $1 \sim W$.
- N: defined as $N = \lceil (\text{L-W})/S \rceil$, which means the numbers of the window sliding from the beginning to the end of current data subset. Due to the limitation of data subset length, the last step might not be the best trained parameter in order to provide full window size data. If $L \gg S$, the mismatch of tail step could be neglected.
- j: sliding sequence order, ranging from $1 \sim N$.
- p_{ij}: the probability of the *i- th* base class based on the *j-th* window data.
- c: coefficient of the computational complexity regularized factor, which is important for WS optimization, that is, if c is too small, the optimization result tends to be larger window, conversely the result tends to smaller.

The equation (3) is composed of three part:

- The first part $\frac{1}{N} \cdot \sum_{j=1}^{N} \left(1 - p_{ij}\right)$ gives the quantitative evaluation of classification uncertainty ranging in [0, 1], which usually decreases with the larger W and N, and the optimization objective is to minimize such part.
- The second part $c \cdot W \cdot N$ gives the quantitative evaluation of computational complexity, the optimization objective is also to minimize such part. In order to match the first part in magnitude ranging in [0, 1], the coefficient c is set to the reciprocal of the maximum $W \cdot N$, exists on the condition $W = L/2, S = 1$. In this paper, the coefficient c is set to $4/L^2$ for the normalization purpose. The computational complexity of the WS optimization is analyzed in Appendix C.
- The third part W/L gives the quantitative evaluation of time latency ranging in [0, 1], which increases with the larger W, the optimization objective is also to minimize such part.

The optimization objective is shown in equation (4), that is, to find the base class with minimum comprehensive loss of classification uncertainty and computational complexity.

$$\arg\min_{i} \left\{ \arg\min_{W^*, S^*} J_i(X|W, S) \right\} \quad (4)$$

3) OPTIMIZATION METHOD

Exhaustive and heuristic methods are widely used in optimization: the former could obtain global optimal solution with high computational complexity, which is usually used in off-line data processing on ground; the latter could obtain local optimal solution with low computational complexity, such as Simulated Annealing (SA), Genetic Algorithm (GA), Particle Swarm Optimization (PSO), etc., which could be used in off-line data processing and on-line static data processing. However, telemetry data is a time-dependent dynamic discrete time series with low time latency requirement, both exhaustive and heuristic methods are not suitable to be applied on board with constrained resources.

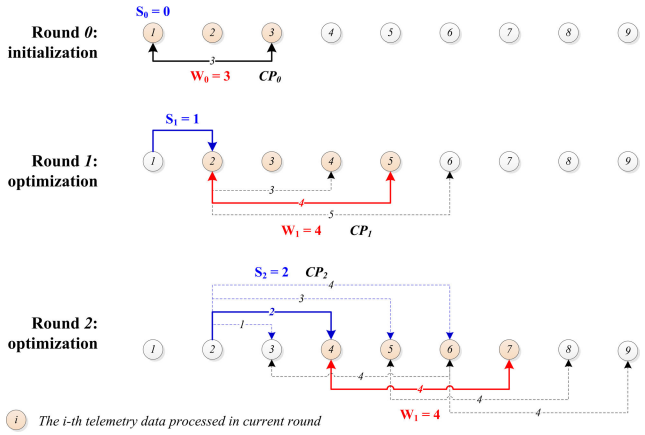


FIGURE 4. An example of the CP-WS iterative optimization procedure.

In this paper, we propose a progressive window constrained step forwarding optimization method, WS optimization, which is embedded in the classification probability calculation procedure, and self-learns the optimal parameters along with the window sliding over the continuous telemetry data stream. As shown in Figure 4, the optimization procedure could be listed as follows:

- Round 0: initialization: utilize the initial window size W0 to calculate the initial classification probabilities $CP_0 = cp_{1,0}, cp_{2,0}, cp_{3,0}, cp_{4,0}$;
- Round 1: fix Step interval as S_1, train the window size from $W_0 \sim L$ until finding the optimal solution W_1 based on the optimization objective;
- Round 2: fix Window size as W_1, train the step interval from $1 \sim W_1$ until finding the optimal solution S_2 based on the optimization objective;
- Next Odd rounds: round sequence $(2n+1), n = 1, 2, \ldots$, similar to round 1, fix Step interval as S_{n+1}, train the window size from $W_n \sim L$ until finding the optimal solution W_{n+1} based on the optimization objective;
- Next Even rounds: round sequence $(2n+2), n = 1, 2, \ldots$, similar to round 2, fix Window size as W_{n+1}, train the step interval from $1 \sim W_{n+1}$ until finding the optimal solution S_{n+2} based on the optimization objective.
- Runs until the data exhaust, during which our algorithm would self-learn from the continuous time-dependent telemetry time series.

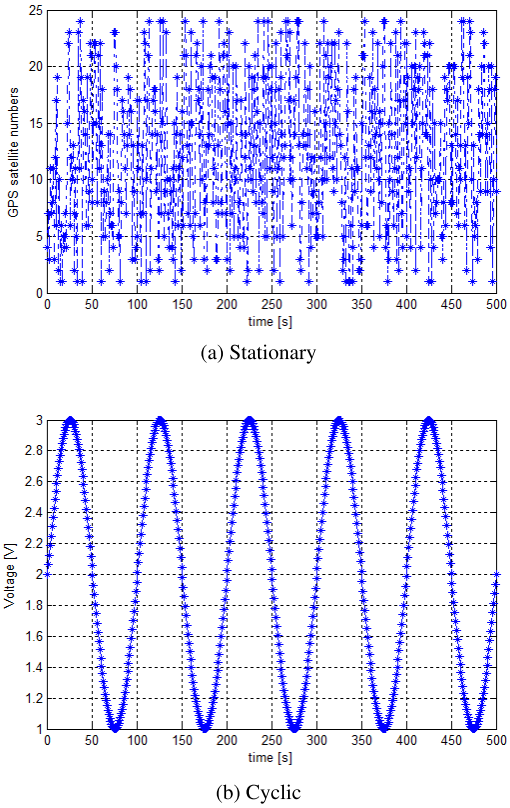
where, S_n indicates the *n-th* optimal step interval result, W_n indicates the *n-th* optimal window size. Apparently, the computational complexity and time latency are both equal to $O(L)$ due to the continuous data input update, which is only concerned with the maximum of window size.

IV. EXPERIMENTS AND RESULTS

A. SIMULATION DATA

1) DATA GENERATION

In order to evaluate the performance of the classification algorithm, we emulate the time data series for some base classes as stationary and cyclic, whose behaviors are similar as real mission data.



(a) Stationary

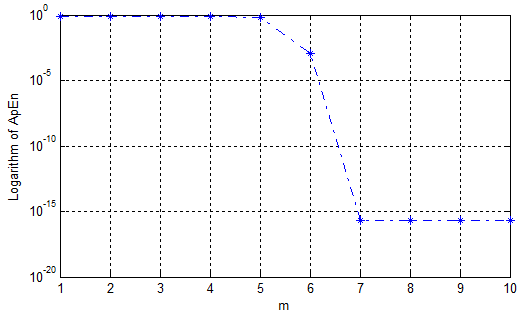
(b) Cyclic

FIGURE 5. Simulation data.

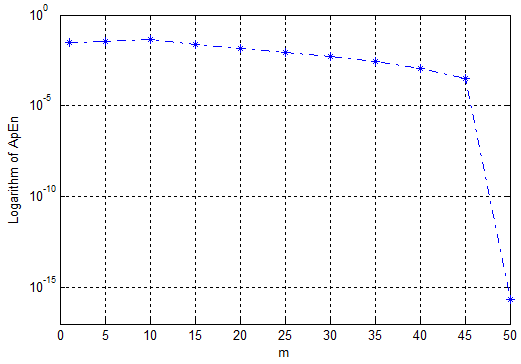
- Stationary data: Gaussian integer distribution ranging in [1, 24] with 500 points, as shown in Figure 5(a), which could be deemed as the available GPS satellite numbers for automatic guidance and navigation on-board.
- Cyclic data: Sine function data ranging in [1, 3] with 500 points, as shown in Figure 5(b), which could be deemed as the available voltages of some equipment on-board.

2) ApEn CALCULATION

- Stationary data: Fix parameter r equals to the root variance of input time series, calculate the ApEn values with the increasing parameter m ranging from $1 \sim 10$. As seen in Figure 6(a), the ApEn value of stationary data falls down dramatically after $m = 5$ then record $m_{0,1} = 5$.
- Cyclic data: Fix parameter r equals to the root variance of input time series, calculate the ApEn values with the increasing parameter m ranging from $1 \sim 50$. As seen in Figure 6(b), the ApEn value of cyclic data falls down dramatically after $m = 45$ then record $m_{0,2} = 45$.
- Initial window size: In order to provide enough data for classification, the initial window size of WS optimization is chosen as the maximum of all the parameter $m_{0,i} (i = 1, 2)$ close to multiple of 50, therefore we get $W_0 = 50$.



(a) Stationary



(b) Cyclic

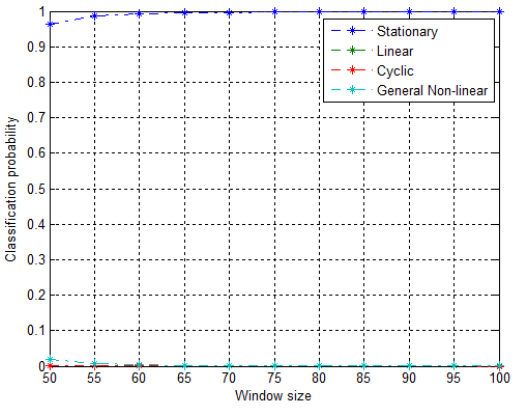
FIGURE 6. ApEn of simulation data.

3) CLASSIFICATION

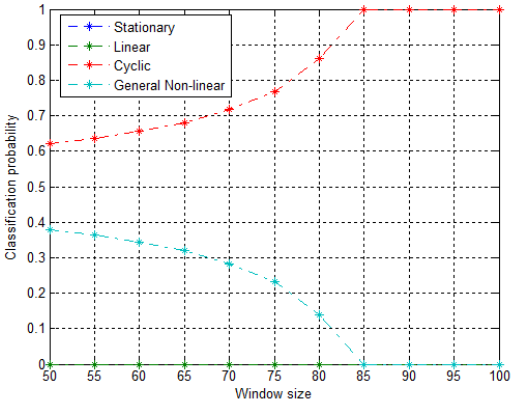
- Stationary data: The classification probabilities of stationary data are shown as Figure 7(a) with $S = 1$. When the window size is larger than 50, the classification probability of base class 1 cp_1 monotonically increases to nearly 100% (as shown in Table 1), which means such time series data could be definitely determined to base class 1 with classification accuracy as 100% by CP classification in high reliability under the membership determination criterion in section III.C. In order to improve our algorithm with proper computational complexity and time latency, the WS optimization should be considered as described in next section.
- Cyclic data: The classification probabilities of cyclic data are shown as Figure 7(b) with $S = 1$. When the window size is larger than 50, the classification probability of base class 3 monotonically increases to 100% (as shown in Table 2), which means the time series could be definitely determined to base class 3 with classification accuracy as 100% by CP classification in high reliability under the membership determination criterion in section III.C. Similar to stationary data, the WS optimization is also considered in next section.

4) OPTIMIZATION

- Stationary data: The optimization procedure of each base class is shown as Figure 8(a). Through the WS optimization as shown in Table 3, we find the best parameters as $W = 50$ with the lowest cost function value as $J_{\min} = 0.4984$, and the classification



(a) Stationary



(b) Cyclic

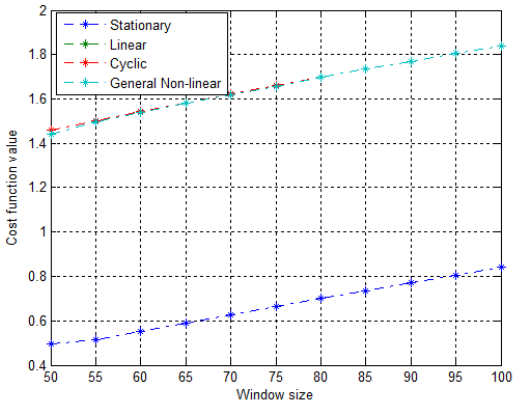
FIGURE 7. Classification probability of simulation data.

TABLE 1. The classification probability of base class 1 (Stationary data).

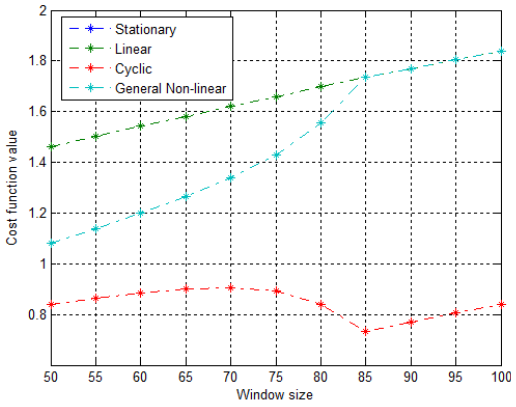
W	cp_1	cp_2	cp_3	cp_4	classification result
50	0.9616	0.0170	0.0001	0.0213	base class 1
55	0.9826	0.0076	0.0000	0.0097	base class 1
60	0.9919	0.0037	0.0000	0.0044	base class 1
65	0.9942	0.0025	0.0000	0.0034	base class 1
70	0.9957	0.0016	0.0000	0.0026	base class 1
75	0.9969	0.0011	0.0000	0.0020	base class 1
80	0.9978	0.0006	0.0000	0.0015	base class 1
85	0.9985	0.0003	0.0000	0.0011	base class 1
90	0.9991	0.0001	0.0000	0.0008	base class 1
95	0.9995	0.0000	0.0000	0.0004	base class 1
100	0.9998	0.0000	0.0000	0.0002	base class 1

TABLE 2. The classification probability of base class 3 (Cyclic data).

W	cp_1	cp_2	cp_3	cp_4	classification result
50	0	0	0.6208	0.3792	base class 3
55	0	0	0.6368	0.3632	base class 3
60	0	0	0.6576	0.3424	base class 3
65	0	0	0.6812	0.3188	base class 3
70	0	0	0.7169	0.2831	base class 3
75	0	0	0.7676	0.2324	base class 3
80	0	0	0.8599	0.1401	base class 3
85	0	0	1.0000	0	base class 3
90	0	0	1.0000	0	base class 3
95	0	0	1.0000	0	base class 3
100	0	0	1.0000	0	base class 3



(a) Stationary



(b) Cyclic

FIGURE 8. Optimization procedure of simulation data.

TABLE 3. The optimization procedure of Stationary data.

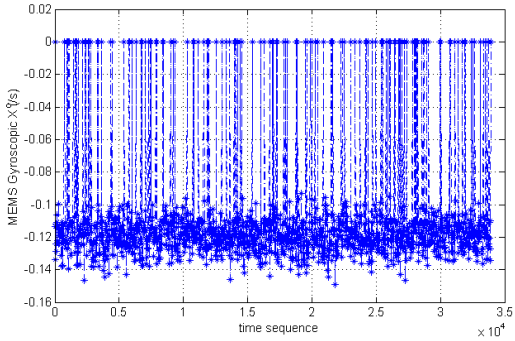
$Window$	J_1	J_2	J_3	J_4
50	0.4984	1.4430	1.4599	1.4387
55	0.5190	1.4940	1.5016	1.4919
60	0.5505	1.5387	1.5424	1.5380
65	0.5882	1.5799	1.5824	1.5790
70	0.6259	1.6200	1.6216	1.6190
75	0.6631	1.6589	1.6600	1.6580
80	0.6998	1.6970	1.6976	1.6961
85	0.7359	1.7341	1.7344	1.7333
90	0.7713	1.7703	1.7704	1.7696
95	0.8061	1.8056	1.8056	1.8052
100	0.8402	1.8400	1.8400	1.8398

probability of base class 1 equals to 96.16% as shown in Table 1, which could provide strongly support to acknowledge the time series as stationary.

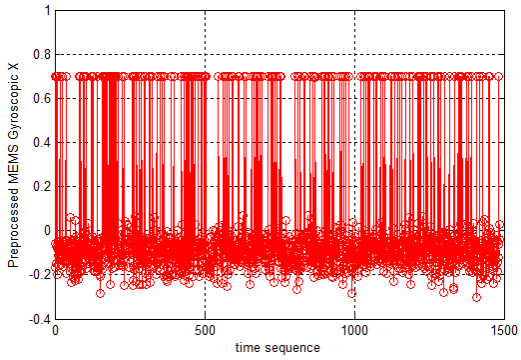
- Cyclic data: The optimization procedure of each base class is shown as Figure 8(b). Through the WS optimization as shown in Table 4, we find the best parameters as $W = 85$ with the lowest cost function value as $J_{\min} = 0.7344$, and the classification probability of base class 3 equals to 100% as shown in Table 2, which could provide strongly support to acknowledge the time series as cyclic.

TABLE 4. The optimization procedure of Cyclic data.

$Window$	J_1	J_2	J_3	J_4
50	1.4600	1.4600	0.8392	1.0808
55	1.5016	1.5016	0.8648	1.1384
60	1.5424	1.5424	0.8848	1.2000
65	1.5824	1.5824	0.9012	1.2636
70	1.6216	1.6216	0.9047	1.3385
75	1.6600	1.6600	0.8924	1.4276
80	1.6976	1.6976	0.8377	1.5575
85	1.7344	1.7344	0.7344	1.7344
90	1.7704	1.7704	0.7704	1.7704
95	1.8056	1.8056	0.8056	1.8056
100	1.8400	1.8400	0.8400	1.8400



(a) Original



(b) Preprocessed

FIGURE 9. MEMS Gyroscopic $X(°/s)$ data.

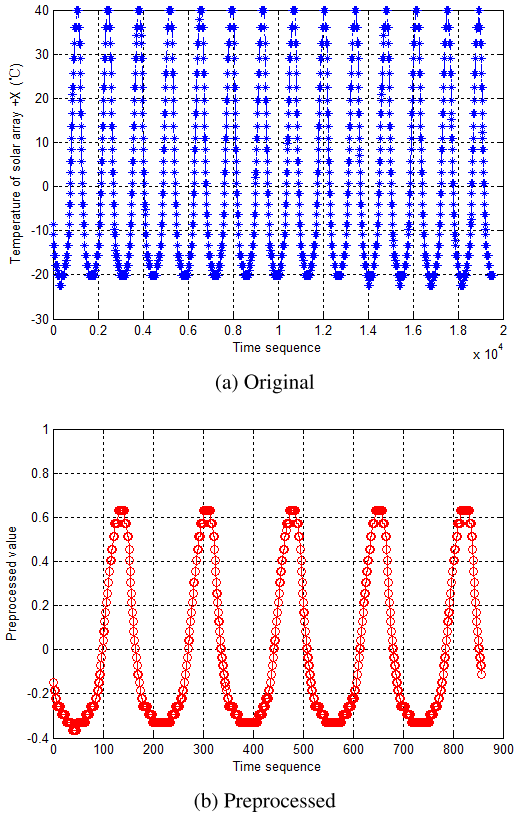
B. MISSION DATA

1) MISSION BACKGROUND

The mission data is the one-day on-board telemetry data of Tsinghua University smart communication satellite recorded on July 20th 2018, the parameters used for flight control is no more than 31. In this paper, we study on 2 typical representative parameters as shown in Figure 9 and Figure 10 in order to evaluate our algorithm, in which the 'MEMS Gyroscopic $X(°/s)$' should be stationary, and the 'Temperature of solar array $+X(°C)$' should be cyclic.

2) ApEn CALCULATION

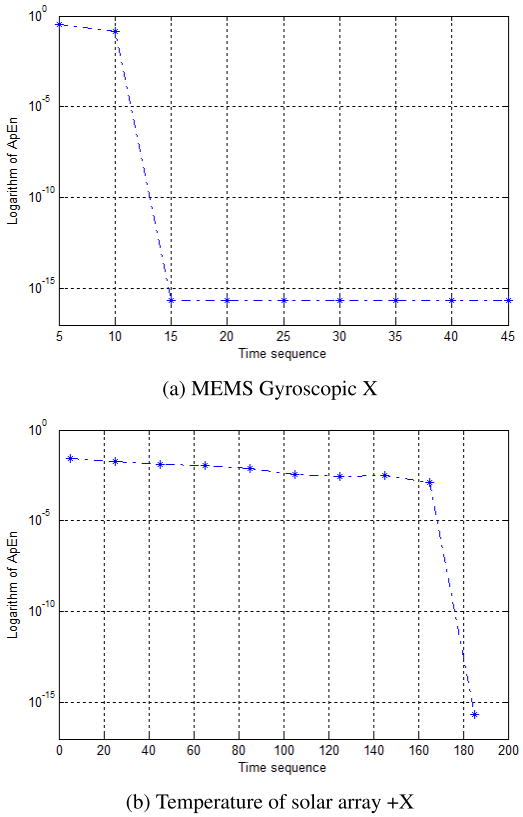
- MEMS Gyroscopic X data: Fix parameter r equals to the root variance of input time series, calculate the ApEn values with the increasing parameter m ranging



(a) Original

(b) Preprocessed

FIGURE 10. Temperature of solar array $+X(°C)$ data.



(a) MEMS Gyroscopic X

(b) Temperature of solar array +X

FIGURE 11. ApEn of mission data.

from $5 \sim 45$. As seen in figure 11(a), the ApEn value of stationary data falls down dramatically after $m = 10$ then record $m_{0,1} = 10$.

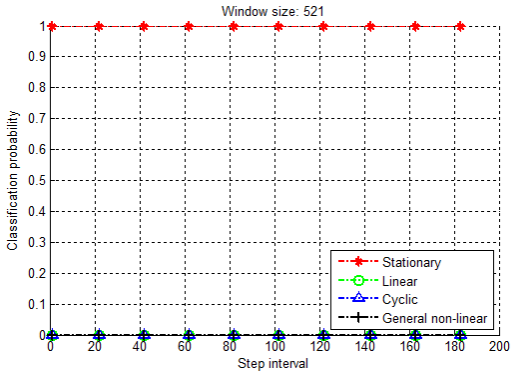
- Temperature of solar array +X data: Fix parameter r equals to the root variance of input time series, calculate the ApEn values with the increasing parameter m ranging from $5 \sim 200$. As seen in Figure 11(b), the ApEn value of cyclic data falls down dramatically after $m = 165$ then record $m_{0,2} = 170$.
- Initial window size: In order to provide enough data for classification, the initial window size of WS optimization is chosen as the maximum of all the parameter $m_{0,i}(i = 1, 2)$ close to multiple of 50, therefore we get $W_0 = 150$.

3) CLASSIFICATION

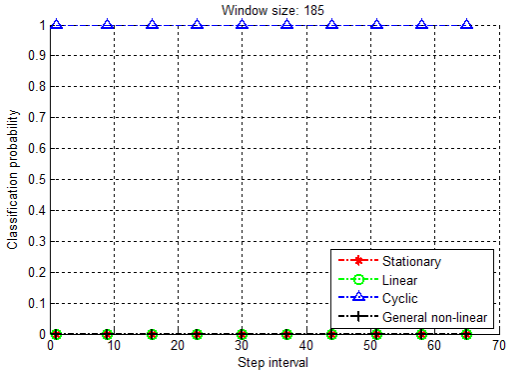
- MEMS Gyroscopic X data: The classification probabilities of telemetry parameter 'MEMS Gyroscopic X' are shown in Figure 12(a), in which the classification probability of base class 1 converses to 100%, that is, the time series data could be definitely determined to base class 1 as Stationary trend data with classification accuracy as 100% by CP classification in high reliability under the membership determination criterion in section III.C.
- Temperature of solar array +X data: The classification probabilities of telemetry parameter 'Temperature of solar array +X' are shown in Figure 12(b), in which the classification probability of base class 3 converses to 100%, that is, the time series data could be definitely determined to base class 3 as Cyclic trend data with classification accuracy as 100% by CP classification in high reliability under the membership determination criterion in section III.C.

4) OPTIMIZATION

- MEMS Gyroscopic X data: The optimization procedure of telemetry parameter 'MEMS Gyroscopic X' is shown as Figure 13(a). Through the WS optimization as shown in Table 5, we find the best parameters as $W = 521$ and $S = 183$ with the lowest cost function value as $J_{\min} = 0.0057$, and the classification probability of base class 1 equals to 100% as shown in Figure 12(a), which could provide strongly support to acknowledge the time series as stationary.
- Temperature of solar array +X data: The optimization procedure of telemetry parameter 'Temperature of solar array +X' is shown as Figure 13(b). Through the WS optimization as shown in Table 6, we find the best parameters as $W = 185$ and $S = 65$ with the lowest cost function value as $J_{\min} = 0.0138$, and the classification probability of base class 3 equals to 100% as shown in Figure 12(b), which could provide strongly support to acknowledge the time series as cyclic.

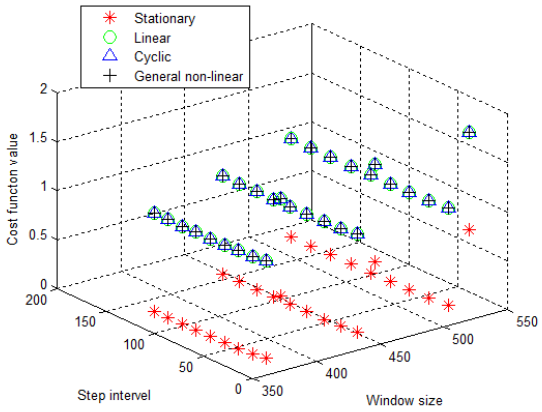


(a) MEMS Gyroscopic X

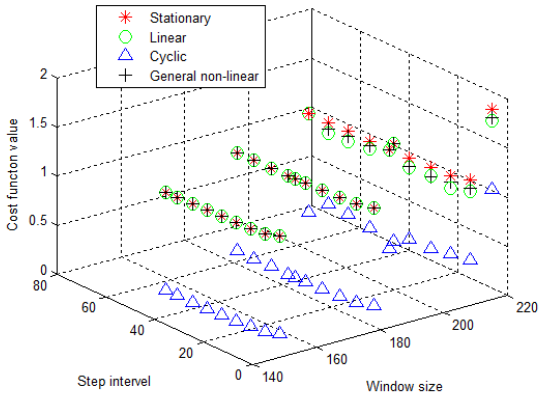


(b) Temperature of solar array +X

FIGURE 12. Classification probability of mission data.



(a) MEMS Gyroscopic X



(b) Temperature of solar array +X

FIGURE 13. Optimization procedure of mission data.

C. DATA TRANSMISSION APPLICATION

In this paper, we utilize the continuous generated simulation data and long-term mission data to make a complete test for our classification algorithm. The final performance should be evaluated by the telemetry compression results in data transmission procedure between the spacecraft and ground stations.

Take the telemetry parameter 'Temperature of solar array +X' as an example, which is classified into base class 3 as cyclic trend, then our proposed data recovery method could be used to obtain the dynamic features and recover the original data. The main steps include:

- Period determination: utilize the fast Fourier transform (FFT) to obtain the approximate period of telemetry data series X;
- Subset data construction: obtain the complete time series data subset in one period;
- Polynomial fitting: utilize n-order ($n = 4$) polynomial fitting method to fit the subset data, obtain the coefficients $p_i, i = 1, 2, \cdots, 5$
- Satellite generates the curve fit parameters on-board and transmits them to ground station, in which the total number is 9 and the total volume is 36 bytes:
  1) Period: 4 bytes, unsigned positive integer
  2) Coefficient p_1: 4 bytes, single-precision float
  3) Coefficient p_2: 4 bytes, single-precision float
  4) Coefficient p_3: 4 bytes, single-precision float
  5) Coefficient p_4: 4 bytes, single-precision float
  6) Coefficient p_5: 4 bytes, single-precision float
  7) Minimum $X_{\min}$: 4 bytes, single-precision float
  8) Mean X_{mean}: 4 bytes, single-precision float
  9) Maximum $X_{\max}$: 4 bytes, single-precision float
- Ground station utilizes the curve fit parameters Period and coefficients $p_i, i = 1, 2, \cdots, 5$ to build the one period fitting time series subset, then repeat the data subset in subsequent cycles to achieve the fitting of the satellite telemetry data X_{fit}.
- The minimum value $X_{\min}$ and maximum value $X_{\max}$ are optional, if you want to get the original physical time series $X_{original}$, it could be calculated by $X_{original} = X_{fit} \times (X_{\max} - X_{\min}) + X_{mean}$.

Figure 14 gives the data recovery results of 'Temperature of solar array +X' with the root mean square error (RMSE) as 0.0519 and crosscorrelation coefficient as 0.9883, that is, the statistical mean recovery biases is limited within 5.19% and the curve fitting data perfroms well on the cross-correlation property with the original telemetry. For the

TABLE 5. The optimization procedure of MEMS Gyroscopic X data.

$Window$	$Step$	J_1	J_2	J_3	J_4
373	1	0.7524	1.7524	1.7524	1.7524
373	16	0.0474	1.0474	1.0474	1.0474
373	30	0.0257	1.0257	1.0257	1.0257
373	45	0.0169	1.0169	1.0169	1.0169
373	59	0.0129	1.0129	1.0129	1.0129
373	73	0.0108	1.0108	1.0108	1.0108
373	88	0.0088	1.0088	1.0088	1.0088
373	102	0.0074	1.0074	1.0074	1.0074
373	117	0.0068	1.0068	1.0068	1.0068
373	131	0.0061	1.0061	1.0061	1.0061
447	1	0.8416	1.8416	1.8416	1.8416
447	19	0.0446	1.0446	1.0446	1.0446
447	36	0.0235	1.0235	1.0235	1.0235
447	53	0.0162	1.0162	1.0162	1.0162
447	71	0.0122	1.0122	1.0122	1.0122
447	88	0.0097	1.0097	1.0097	1.0097
447	105	0.0081	1.0081	1.0081	1.0081
447	122	0.0073	1.0073	1.0073	1.0073
447	140	0.0065	1.0065	1.0065	1.0065
447	157	0.0057	1.0057	1.0057	1.0057
521	1	0.9110	1.9110	1.9110	1.9110
521	22	0.0416	1.0416	1.0416	1.0416
521	42	0.0217	1.0217	1.0217	1.0217
521	62	0.0151	1.0151	1.0151	1.0151
521	82	0.0113	1.0113	1.0113	1.0113
521	102	0.0095	1.0095	1.0095	1.0095
521	122	0.0076	1.0076	1.0076	1.0076
521	143	0.0066	1.0066	1.0066	1.0066
521	163	0.0057	1.0057	1.0057	1.0057
521	183	0.0057	1.0057	1.0057	1.0057

TABLE 6. The optimization procedure of Temperature of solar array +X data.

$Window$	$Step$	J_1	J_2	J_3	J_4
154	1	1.7533	1.7533	0.7533	1.7533
154	7	1.1085	1.1085	0.1085	1.1085
154	13	1.0592	1.0592	0.0592	1.0592
154	19	1.0411	1.0411	0.0411	1.0411
154	25	1.0312	1.0312	0.0312	1.0312
154	31	1.0247	1.0247	0.0247	1.0247
154	37	1.0214	1.0214	0.0214	1.0214
154	43	1.0181	1.0181	0.0181	1.0181
154	49	1.0164	1.0164	0.0164	1.0164
154	54	1.0148	1.0148	0.0148	1.0148
185	1	1.8436	1.8436	0.8436	1.8436
185	9	1.0948	1.0948	0.0948	1.0948
185	16	1.0533	1.0533	0.0533	1.0533
185	23	1.0375	1.0375	0.0375	1.0375
185	30	1.0296	1.0296	0.0296	1.0296
185	37	1.0237	1.0237	0.0237	1.0237
185	44	1.0198	1.0198	0.0198	1.0198
185	51	1.0178	1.0178	0.0178	1.0178
185	58	1.0158	1.0158	0.0158	1.0158
185	65	1.0138	1.0138	0.0138	1.0138
216	1	1.9135	1.8062	1.0994	1.8349
216	10	1.0923	0.9837	0.2767	1.0165
216	18	1.0507	0.9220	0.2499	0.9803
216	26	1.0369	0.9512	0.2123	0.9472
216	35	1.0277	0.9342	0.2051	0.9437
216	43	1.0231	1.0231	0.0231	1.0231
216	51	1.0185	0.9488	0.1354	0.9711
216	60	1.0161	0.9122	0.1722	0.9640
216	68	1.0138	0.9041	0.1840	0.9534
216	76	1.0138	1.0138	0.0138	1.0138

data compression, the total volume generated on board in one day is equal to 19574 parameters $\times$2bytes/parameters = 39148 bytes, so that our algorithm could compress the

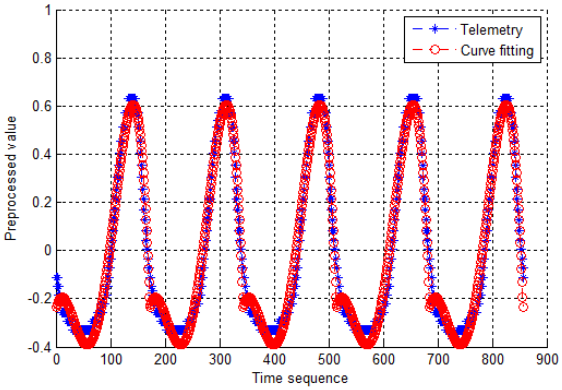


FIGURE 14. Data recovery results of Temperature of solar array +X.

'Temperature of solar array +X' telemetry volume into less than 0.1%.

Through the cyclic trend data recovery method described above, we could utilize only 9 fit parameters to accomplish high accuracy data recovery in which the data compression rate performs better over time. Compared with the fixed length telemetry compression algorithm 'POCKET+' [23] whose best compression rate is no less than 5%, our algorithm could compress the all-day telemetry data into less than 0.1%, which performs 50 times better than 'POCKET+'. In addition, our algorithm could also be used in flexible length telemetry, which is more suitable for the future space mission requirements.

V. CONCLUSION

In this paper, the proposed CP-WS algorithm achieves a high accuracy for telemetry classification based on self-learning, which is suitable for time-dependent space environment and massive data transmission. Based on the time series feature analysis, four base classes including stationary, linear, cyclic and general non-linear are introduced firstly in order to describe and distinguish the telemetry behaviors over time. Next, the information entropy analysis provides the entropy rate characteristics of different base classes, as well as the available entropy calculation method for telemetry time series. Finally, a new on-board lightweight self-learning algorithm CP-WS (Classification Probability calculation ÍC Window Step optimization) is proposed to obtain the class features and make the decision of each single parameter from the discrete telemetry time series, with high classification accuracy as 100% and adaptive computational complexity ranging from $O(L^2)$ to $O(L)$. In our opinion, the new algorithm enriches the research point of telemetry data transmission, which extracts feature of different telemetry types from classification in order to compress the telemetry parameters according to their specific behavior features respectively, and improves the transmission efficiency with high accuracy.

It is worthwhile pointing out that, our algorithm is oriented to single telemetry parameter classification at present, the computational complexity of all the satellite telemetry parameters is simply the mathematical addition of each

telemetry parameter ranging from $O(n \cdot L^2)$ to $O(n \cdot L)$, in which n indicates the number of telemetry parameters. In following studies, we would pay attention to the training parameters shared among different telemetry parameters, in order to further decrease the computational complexity. Meanwhile, the data recovery methods of different base classes would be further studied to improve the compression performance of the whole satellite.

APPENDIXES

APPENDIX A
PROVE ON STATIONARY FEATURE FOR THE FIRST-ORDER DIFFERENCE OF HOMOGENEOUS POISSON PROCESS

Based on the features of homogeneous Poisson process $Y(t)$ with $0 \leq t_0 < t_1 < \ldots < t_n$, it could be equivalent to n independent incremental processes $Y(t_1) - Y(t_0)$, $Y(t_2) - Y(t_1), \ldots, Y(t_n) - Y(t_{n-1})$, with identical distribution. Define the first-order difference as $\Delta Y_n = Y(t_n) - Y(t_{n-1})$, then we have:

1) Mean: $E(\Delta Y_n) = \lambda \cdot t_n - \lambda \cdot t_{n-1} = \lambda$.

2) Variance: $Var(\Delta Y_n) = \lambda$

Detailed analysis is described as follows:

$E(Y_t) = \lambda \cdot t$,

$Var(Y_t) = \lambda \cdot t = E(Y_t^2) - (EY_t)^2 = E(Y_t^2) - \lambda^2 \cdot t^2$

$\Rightarrow E(Y_t^2) = \lambda \cdot t + \lambda^2 \cdot t^2$

$E(\Delta Y_t) = \lambda \cdot t - \lambda \cdot (t-1) = \lambda$

then, we have:

$$\begin{aligned} Var(\Delta Y_t) &= E(\Delta Y_t^2) - (E\Delta Y_t)^2 \\ &= E[Y_t^2 + Y_{t-1}^2 - 2 \cdot Y_t \cdot Y_{t-1}] - \lambda^2 \\ &= E(Y_t^2) + E(Y_{t-1}^2) - 2 \cdot E[Y_t \cdot Y_{t-1}] - \lambda^2 \end{aligned}$$

According to the autocorrelation characteristics as $E[Y_t \cdot Y_{t-1}] = \lambda^2 \cdot t \cdot (t-1) + \lambda \cdot \min(t, t-1)$, we have $Var(\Delta Y_t) = \lambda t + \lambda^2 t^2 + \lambda(t-1) + \lambda^2(t-1)^2 - 2[\lambda^2 t(t-1) + \lambda(t-1)] - \lambda^2 = \lambda$.

3) Auto-covariance function: $Cov(\Delta Y_n, \Delta Y_m) = E(\Delta Y_n \cdot \Delta Y_m) - E(\Delta Y_n) \cdot E(\Delta Y_m) = 0$.

To summarize, all three Eigen statistics of the first-order difference of homogeneous Poisson process are time-independent, so that such time series could be classified to random data with stationary feature. (QED)

APPENDIX B
PROVE ON ENTROPY RATE FOR THE GENERAL NON-LINEAR TREND

In this paper, we use the general stable source model to derive the information entropy rate changes of the general non-linear function.

Theorem: For discrete stable source, if $H_1(U) < +\infty$, then the entropy rate $\lim_{N \to \infty} H_N(U)$ exists.

Proof: Based on the stable feature of the source and the fact that unconditional entropy is not less than conditional entropy, $H(U_{N-1}|U_1U_2 \ldots U_{N-2}) = H(U_N|U_2U_3 \ldots U_{N-1}) \geq H(U_N|U_1U_2 \ldots U_{N-1})$, which means that conditional entropy $H(U_N|U_1U_2 \ldots U_{N-1})$ decreases with the larger N, then we have:

$$\begin{aligned} &H(U_1U_2 \ldots U_N) \\ &= N \cdot H_N(U) \\ &= H(U_1) + H(U_2|U_1) + \ldots + H(U_N|U_1 \ldots U_{N-1}) \\ &= H(U_N) + H(U_N|U_{N-1}) + \ldots + H(U_N|U_1 \ldots U_{N-1}) \\ &\geq N \cdot H(U_N|U_1 \ldots U_{N-1}) \end{aligned}$$

In additional, we have:

$$\begin{aligned} &H(U_1U_2 \ldots U_N) \\ &= N \cdot H_N(U) \\ &= H(U_N|U_1 \ldots U_{N-1}) + H(U_1U_2 \ldots U_{N-1}) \\ &= H(U_N|U_1 \ldots U_{N-1}) + (N-1) \cdot H_{N-1}(U) \end{aligned}$$

From above formulas, we can deduce that $H_{N-1}(U) \geq H_N(U)$. Therefore, for general non-linear function, we have: $+\infty > H_1(U) \geq \ldots \geq \ldots H_{N-1}(U) \geq H_N(U) \geq 0$. With the extreme value of entropy, $H_1(U) \leq \log_2 K$, then we could get the information entropy rate $H_N(U)$ for the general non-linear trend data as $0 \leq H_\infty(U) \leq \ldots \leq H_N(U) \leq \ldots \leq H_1(U) \leq \log_2 K$, with the lower bound 0 and the upper bound $\log_2 K$. (QED)

APPENDIX C
THE COMPUTATIONAL COMPLEXITY OF THE WS OPTIMIZATION

The computational complexity of the WS optimization diverses in different optimization iterations according to time dependant telemetry data stream. Given the maximum window size of current data subset as L, we could analyze the computational complexity as follows.

1) The complexity in each CP classification iteration

In each iteration, the window size W is given as the priori parameter, which is deemed as the length of the data subset with dimension as 1. As described above, the Classification Probability (CP) classification includes three serial test with respective computational complexity as follows:

- Stationary test: the computational complexity of Augmented Dickey-Fuller test is $O(W)$.
- Linear test: the computational complexity of linear fit method is equivalent to classical least square method, which is also $O(W)$.
- Cyclic test: the computational complexity of ensemble empirical mode decomposition (EEMD) method is $O(2 \cdot k \cdot W)$, where k means the number of intrinsic mode fuctions (IMF). In our simulation, k is determined by the termination condition with the maximum value as constant 7. Therefore, the computational complexity of EEMD is equivalent to $O(W)$.

To sum up, the complexity in each CP classification iteration is the combination of all these three serial test, which is equivalent to $O(W)$.

2) The complexity of WS optimization embedded with the CP classification

The computational complexity of WS optimization embedded with the CP classification could be described by the equation $O(W) \cdot N$, where N means the numbers of the window sliding from the beginning to the end of current data subset, defined as $N = \lceil (\text{L-W}) / S \rceil$. Then, the lower bound and upper bound of such computational complexity could be calculated as follows:

- Upper bound: the upper bound of computational complexity exists on the condition of maximum $O(W) \cdot N$, that is, $\max(O(W) \cdot N) = \max(O(W \cdot N)) = O(\max(W \cdot N))$. The maximum value of $W \cdot N = W \cdot (L-W)/S$ is $L^2/4$ on the condition $W = L/2, S = 1$. Therefore, the upper bound is equivalent to $O(L^2)$.
- Lower bound: the lower bound of computational complexity exists on the condition of minimum $O(W) \cdot N$, that is, $\min(O(W) \cdot N) = \min(O(W \cdot N)) = O(\min(W \cdot N))$. The minimum value of $W \cdot N = W \cdot (L - W)/S$ is L on the condition $W = L$, $S = 0$. Therefore, the lower bound is equivalent to $O(L)$.

To sum up, the computational complexity of the WS optimization is adpative according to different data subset, ranging from $O(L^2)$ to $O(L)$. (QED)

REFERENCES

[1] T. Yairi, N. Takeishi, T. Oda, Y. Nakajima, N. Nishimura, and N. Takata, "A data-driven health monitoring method for satellite housekeeping data based on probabilistic clustering and dimensionality reduction," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 3, pp. 1384–1401, Jun. 2017.

[2] T. Yairi, T. Tagawa, and N. Takata, "Telemetry monitoring by dimensionality reduction and learning hidden Markov model," in *Proc. Int. Symp. Artif. Intell., Robot. Automat. Space*, 2012.

[3] T. Yairi, M. Nakatsugawa, K. Hori, S. Nakasuka, K. Machida, and N. Ishihama, "Adaptive limit checking for spacecraft telemetry data using regression tree learning," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Oct. 2004, pp. 5130–5135.

[4] S. Tariq, S. Lee, Y. Shin, M. S. Lee, O. Jung, D. Chung, and S. Woo, "Detecting anomalies in space using multivariate convolutional LSTM with mixtures of probabilistic PCA," in *Proc. 25th ACM SIGKDD Int. Conf.*, Anchorage, AK, USA, 2019.

[5] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, "Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding," in *Proc. 24th ACM SIGKDD Int. Conf.*, London, U.K., 2018, pp. 387–395.

[6] S. Fuertes, G. Picart, J.-Y. Tourneret, L. Chaari, A. Ferrari, and C. Richard, "Improving spacecraft health monitoring with automatic anomaly detection techniques," in *Proc. 14th Int. Conf. Space Oper.*, Daejeon, South Korea, 2016.

[7] D. L. Iverson, R. Martin, and M. Schwabacher, "General purpose data-driven monitoring for space operations," *J. Aerosp. Comput. Inf. Commun.*, vol. 9, no. 2, pp. 26–44, 2012.

[8] P. Robinson, M. Shirley, D. Fletcher, R. Alena, D. Duncavage, and C. Lee, "Applying model-based reasoning to the FDIR of the command and data handling subsystem of the international space station," in *Proc. Int. Symp. Artif. Intell., Robot. Automat. Space (iSAIRAS)*, 2003.

[9] J. D. Hamilton, *Time Series Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 1994.

[10] N. E. Huang, Z. Shen, and S. R. Long, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Math. Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998.

[11] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise assistant data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, 2009.

[12] S. M. Pincus, "Approximate entropy as a measure of system complexity," *Proc. Nat. Acad. Sci. USA*, vol. 88, no. 6, p. 2297, Mar. 1991.

[13] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Amer. J. Physiol.-Heart Circulatory Physiol.*, vol. 278, no. 6, 2000, Art. no. H2039.

[14] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Phys. Rev. Lett.*, vol. 88, no. 17, Apr. 2002, Art. no. 174102.

[15] B. D. Fulcher, "Feature-based time-series analysis," Sep. 2017, *arXiv:1709.08055*. [Online]. Available: https://arxiv.org/abs/1709.08055

[16] X. Liu, M. Jia, Z. Na, W. Lu, and F. Li, "Multi-modal cooperative spectrum sensing based on Dempster-Shafer fusion in 5G-based cognitive radio," *IEEE Access*, vol. 6, pp. 199–208, 2018.

[17] X. Liu, X. Zhang, M. Jia, L. Fan, W. Lu, and X. Zhai, "5G-based green broadband communication system design with simultaneous wireless information and power transfer," *Phys. Commun.*, vol. 25, pp. 539–545, Jun. 2018.

[18] X. Liu, M. Jia, X. Zhang, and W. Lu, "A novel multichannel Internet of things based on dynamic spectrum sharing in 5G communication," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 5962–5970, Aug. 2019.

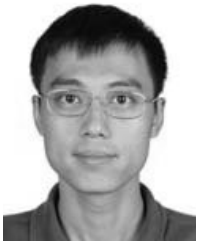
[19] M. Jia, Z. Gao, Q. Guo, Y. Lin, and X. Gu, "Sparse feature learning for correlation filter tracking toward 5G-enabled tactile Internet," *IEEE Trans. Ind. Informat.*, to be published, doi: 10.1109/TII.2019.2906087.

[20] S. Song, A. Zhang, J. Wang, and P. S. Yu, "SCREEN: Stream data cleaning under speed constraints," in *Proc. ACM SIGMOD Int. Conf. Manage. Data (SIGMOD)*, Melbourne, VIC, Australia, 2015, pp. 827–841.

[21] R. Harris, "Testing for unit roots using the augmented Dickey-fuller test: Some issues relating to the size, power and the lag structure of the test," *Econ. Lett.*, vol. 38, no. 4, pp. 381–386, 1992.

[22] R. S. Tsay, "Nonlinearity tests for time series," *Biometrika*, vol. 73, no. 2, pp. 461–466, 1986.

[23] D. Evans and A. Donati, "The ESA POCKET+ housekeeping telemetry compression algorithm: Why make spacecraft operations harder than it already is?" in *Proc. SpaceOps Conf.*, Marseille, France, 2018.



PENG WAN (Student Member, IEEE) received the B.S. degree in physics from Peking University, in 2004, and the M.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 2007, where he is currently pursuing the Ph.D. degree in aeronautical and astronautical science and technology with the Space Center. His research interests include the system design of space network and deep space communications.

YAFENG ZHAN received the B.S.E.E. and Ph.D.E.E. degrees from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 1999 and 2004, respectively. He is currently an Associate Professor with the Space Center, Tsinghua University. His current research interests include communication signal processing and deep space communications.

WEIWEI JIANG (Student Member, IEEE) was born in Nantong, Jiangsu, China, in 1991. He received the B.Sc. and Ph.D. degrees from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2013 and 2018, respectively. He is currently a Postdoctoral Researcher with the Department of Electronic Engineering, Tsinghua University. His current research interests include the intersection between big data and machine learning techniques and space communication signal processing applications.