A Data-Driven Health Monitoring Method for Satellite Housekeeping Data Based on Probabilistic Clustering and Dimensionality Reduction

TAKEHISA YAIRI, Member, IEEE
NAOYA TAKEISHI, Student Member, IEEE
TETSUO ODA
University of Tokyo, Tokyo, Japan

YUTA NAKAJIMA
NAOKI NISHIMURA
NOBORU TAKATA
Japan Aerospace Exploration Agency, Tsukuba, Japan.

In the operation of artificial satellites, it is very important to monitor the health status of the systems and detect any symptoms of anomalies in the housekeeping data as soon as possible. Recently, the data-driven approach to the system monitoring problem, in which statistical machine learning techniques are applied to the large amount of measurement data collected in the past, has attracted considerable attention. In this paper, we propose a new data-driven health monitoring and anomaly detection method for artificial satellites based on probabilistic dimensionality reduction and clustering, taking into consideration the miscellaneous characteristics of the spacecraft housekeeping data. We applied our method to the telemetry data of the small demonstration satellite 4 (SDS-4) of the Japan Aerospace Exploration Agency (JAXA) and evaluated its effectiveness. The results show that the proposed system provides satellite operators with valuable information for understanding the health status of the system and inferring the causes of anomalies.

Manuscript received December 4, 2015; revised September 13, 2016; released for publication December 25, 2016. Date of publication February 20, 2017; date of current version June 7, 2017.

DOI. No. 10.1109/TAES.2017.2671247

Refereeing of this contribution was handled by J. H. Saleh.

This work was partially supported by JSPS KAKENHI under Grant JP22560779 and Grant JP26289320.

Authors' addresses: T. Yairi, N. Takeishi, and T. Oda are with the Department of Aeronautics and Astronautics, The University of Tokyo, Tokyo 1138656, Japan, E-mail: (yairi@ailab.t.u-tokyo.ac.jp; takeishi@ailab.t.u-tokyo.ac.jp; tetuoda0414@gmail.com); Y. Nakajima, N. Nishimura, and N. Takata are with Japan Aerospace Exploration Agency, Tsukuba, Japan, E-mail: (nakajima.yuta@jaxa.jp; swordembryo@kra.biglobe.ne.jp; ayayataka116@yahoo.co.jp)

I. INTRODUCTION

It is a very important task in the operation of artificial satellites to monitor the health state of the system and detect any symptoms of the components in the housekeeping telemetry, which contains miscellaneous sensor and status data. The most fundamental and widely used method for this purpose is limit checking, in which the sensor measurements are checked as to whether they are within predefined ranges. Unfortunately, it is becoming significantly more difficult to set the limit values appropriately for all the sensors in advance, as the number of sensors included in the recent satellite systems is increasing. In addition, there exist anomalies that cannot be found by checking the individual sensor measurements separately. Although rule-based and model-based anomaly detection and diagnosis systems were actively studied to address these issues in some space missions, it is very costly to construct and maintain the sets of rules and system models using expert knowledge.

Meanwhile, in many fields, the mining of useful information in large datasets that used to be difficult to handle has attracted much attention recently. In the context of spacecraft operation, data-driven or learning-based health monitoring/anomaly detection methods have been proposed. In this approach, empirical behavior models of the systems are obtained by applying machine learning or data mining algorithms to the past sensor data, and then, the models are exploited to check whether the new data are normal or not. A major advantage of data-driven health monitoring is that it can be applied to a wide variety of systems even by a person without expert knowledge of them. In other words, this approach is suitable for monitoring the systems *broadly and shallowly* and is expected to reduce the operators' workload. When we apply the data-driven monitoring approach to an actual artificial satellite system, however, it is mandatory to choose an appropriate statistical model for representing the normal system behavior and a learning algorithm for estimating the parameters in the model from the training data. This is not a trivial issue, because a satellite system is very complex and its housekeeping data include many aspects, such as high dimensionality, multimodality, and heterogeneity. Another issue of the data-driven approach is that it has not yet been fully verified in actual satellite operations.

The contribution of this paper is twofold. First, we focus on the high dimensionality and multimodality that are two significant characteristics of the satellite housekeeping data and propose a health monitoring/anomaly detection method based on probabilistic dimensionality reduction and clustering to handle them. Second, we experimentally applied the proposed method to Japan Aerospace Exploration Agency (JAXA)'s small demonstration satellite 4 (SDS-4) in operation, and validated it over more than two years. The result of this experiment shows that the proposed data-driven monitoring method is very useful, not only because it automatically detects "anomalous" patterns that were unseen in the past, but also because it provides the operators with valuable information for understanding the health

status of the system and analyzing the causes of the detected anomalies.

The remaining sections of this paper are organized as follows. Section II briefly summarizes the conventional methods of anomaly detection and diagnosis for spacecraft systems. In Section III, we clarify the important characteristics of the satellite housekeeping data and describe our data preprocessing methods for handling them. In Section IV, we describe a new data-driven health monitoring and anomaly detection method based on probabilistic dimensionality reduction and clustering. Section V presents the experimental results of testing the proposed method in the operation of JAXA's SDS-4. Section VI discusses the validity, limitation, and future direction of the data-driven approach based on the experimental results, and Section VII presents some concluding remarks.

II. RELATED WORKS

Existing health monitoring and anomaly detection methods for spacecraft systems can be divided into two categories: *knowledge-driven* and *data-driven* approaches.

The former approach detects anomalies and infers the causes using rule bases [1]–[5], qualitative models [6]–[9], and probabilistic models [10], [11] that are built from experts' knowledge in advance. Methods of this type have the advantage that they are able to diagnose the anomalies in detail, if the knowledge is accurate and complete. However, it is very costly to construct and maintain such complete rule bases and models. In particular, it is difficult to comprehensively describe the complicated behaviors of the modern space systems, as they are composed of a huge number of components, such as sensors and actuators.

The latter approach first learns the empirical models of the system by applying statistical machine learning algorithms to the past operation data, and then, checks whether the system is normal or not by evaluating the most recent operation data using the learned models. Thus far, a variety of machine learning methods, including regression [12]–[14], classification [15], clustering [16], [17], kernel principal components analysis (PCA) [18], [19], dimensionality reduction, and the hidden Markov model (HMM) [20], have been employed for modeling space systems. The most important advantage of data-driven methods is that they can be easily applied to a variety of systems, because they do not require expensive expert knowledge and automatically learn the statistical models from the data. Conversely, if a sufficient amount of training data is not provided, a data-driven method is unable to learn a proper model and is likely to give many false alarms or miss true anomalies. As compared with knowledge-driven methods, the data-driven approach is suitable for monitoring the health status of the systems and detecting any "suspicious" patterns that are not contained in the past data. In this paper, we focus on data-driven anomaly detection and propose a method based on probabilistic clustering and dimensionality reduction in Section IV.

III. PROPERTIES AND PREPROCESSING OF SATELLITE TELEMETRY DATA

A. Properties of Telemetry Data

Regardless of the problem to which we apply machine learning or data mining methods, it is necessary that we understand the data characteristics and perform appropriate preprocessing of the data. For this purpose, in this section, we first clarify the common characteristics of the housekeeping data of artificial satellites.

1) High Dimensionality: Satellite housekeeping data are apparently very *high dimensional*, usually being composed of hundreds to thousands of variables. In such a high-dimensional data space, it is difficult to compute distances among data samples correctly. This issue is widely known as the *curse of dimension*. In the context of anomaly detection, it is known that the difference in the distance between anomalous and normal samples and that between normal samples becomes indistinct as the dimensionality becomes higher [21]. Therefore, a simple distance-based anomaly detection algorithm is not suitable for satellite housekeeping data. In addition, it should be also noted that the variables in housekeeping data are strongly correlated, which means that the intrinsic dimension of the data is much lower.

2) Multimodality: A satellite system (or each of its subsystems) has a number of different operational modes and changes from one mode to another over time. For example, the most fundamental modes for earth orbiting satellites are daytime and nighttime modes. Between these two modes, there are distinct differences in power generation, surface temperature, and so on. As a result, the distribution of a satellite's housekeeping data becomes multimodal and their samples form multiple clusters in the data space. Fig. 1 shows an example of how the satellite data distributes like, which was obtained by applying PCA to SDS-4 real-valued 89-dimensional data of one month. We performed the *normalization* on the data beforehand as described in Section III-B.3, because the original data have diverse physical units and numerical ranges. The left figure shows the first and second principal components, whereas the right one shows third and fourth components. It can be seen that the data samples are naturally divided into several groups. Each point in the figure indicates the data sample at a time step. This implies that it is unreasonable to model the satellite's housekeeping data by using unimodal distributions such as multivariate Gaussian.

3) Heterogenuity: The variables in a satellite's housekeeping data are divided into two types: Continuous variables that take real values and discrete variables that take categorical values. In addition, continuous variables have a diversity in the physical quantities, units, and ranges. The housekeeping data are very different in terms of this property from other high-dimensional data, such as image data that are represented in homogeneous variables or pixels. To handle such heterogeneous data, we have to apply appropriate preprocessing techniques, such as nondimensionalization, to the original data and employ modeling methods that can accommodate both continuous and discrete variables.

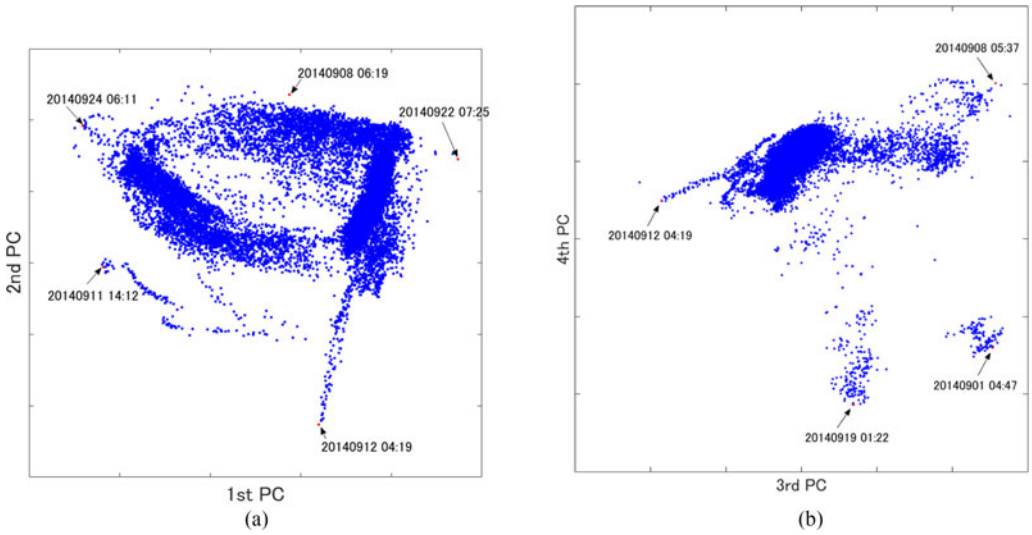


Fig. 1. Distribution of continuous data of SDS-4 for one month in the low-dimensional space obtained by PCA. It can be seen that the distribution has the multimodality. (a) First versus second components. (b) Third versus fourth components.

4) Temporal Dependence: As artificial satellites are dynamical systems, the housekeeping data generated by them are (multidimensional) time series. There is no doubt that this temporal dependence is an essential property of the housekeeping data and is very useful for monitoring the system. For example, if the values of several variables change together at a time point, it is natural to consider that some event has occurred in the system.

In this study, however, we utilized the temporal dependence only for detecting "trivial" outliers for each variable, as described below, and focus rather on the instantaneous relationships among multiple variables at each time step. The main reason why we did not take advantage of the temporal dependence in this study is that it was difficult to apply the typical discrete-time time-series models, such as autoregressive models and HMMs, to the satellite housekeeping data, as the sampling rates of telemetry of small satellites, including SDS-4, are often changed because of the limitation of data transmission capacity.

5) Missing Data: In most spacecraft systems, as the measurement variables differ in sampling periods and timings, not all the values of all the variables at every time step are obtained. For example, measurements such as attitude angles and rates are usually sampled at a higher sampling rate, whereas those such as component temperatures are sampled at a lower rate. As a result, a large part of the housekeeping data is usually missing, as illustrated in Fig. 2. It should be noted that the pattern of missing data is structural rather than random. However, most machine learning algorithms, with a few exceptions, basically assume that the data are complete. Therefore, when we apply them to the housekeeping data, we have to modify them so that they can handle the missing data or perform a preprocessing step to recover the missing elements before applying the learning algorithms.

6) Trivial Outliers: Satellite housekeeping data occasionally contain exceptionally large abnormal values,

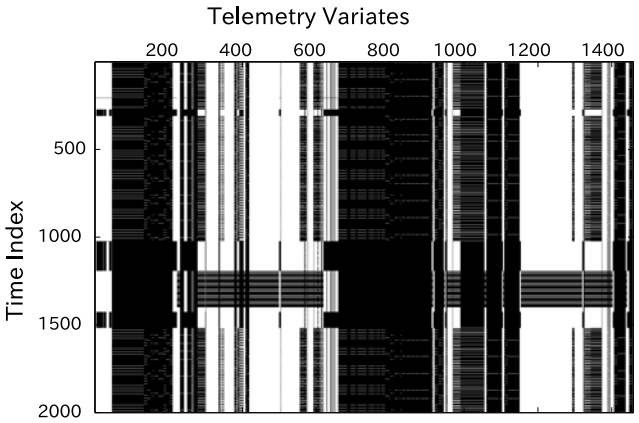


Fig. 2. Telemetry data with structurally missing elements. Elements filled with black indicate that the values are missing.

caused by errors in data conversion or transmission. The left figure of Fig. 3 shows one variable of the SDS-4 housekeeping data containing such abnormal values. As can be seen from the axis scale, the absolute values of these outliers are incomparably large. In contrast, the right figure of Fig. 3 shows the same data after removing the outliers. It can be observed that most of the normal values remain in the range 0.0–1.5.

Such erroneous values or outliers are to a certain extent common to almost all satellites' housekeeping data. The important point here is that these outliers are caused only temporarily by errors in the data conversion or transmission process and do not represent any serious anomalies in the satellite systems. In this sense, they can be called "trivial outliers" and should be discriminated from the truly serious anomalies, the detection of which is our objective.

These trivial outliers have potentially two nontrivial issues for data-driven health monitoring. First, trivial outliers may hide the truly serious anomalies from the operators, because the absolute values of the former are generally

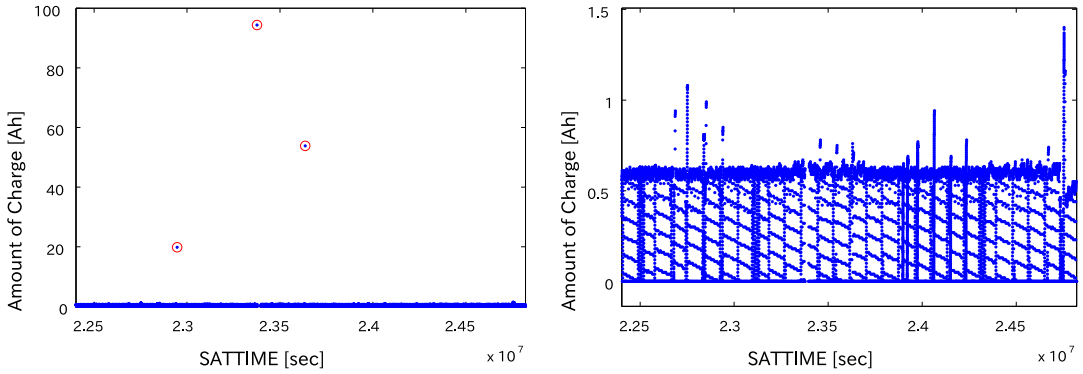


Fig. 3. Telemetry variate with outliers and after they are removed. (a) With outliers. (b) After removed.

much larger than those of the latter. Second, if outliers with extremely large values are contained in the training data, the learned models are severely damaged.

Therefore, it is necessary to remove trivial outliers beforehand in the data preprocessing stage.

B. Data Preprocessing

Some of the characteristics of the satellite housekeeping data can be treated in the preprocessing stage, which facilitates the application of the standard machine learning algorithms to the resultant data. In this section, we explain three types of preprocessing that we applied to SDS-4 housekeeping data. We consider they are also applicable to other satellites.

1) Detection and Removal of Trivial Outliers: As mentioned in Section III-A.6, the trivial outliers contained in the satellite housekeeping data interfere significantly with data-driven anomaly detection, in both the training and testing phases.

These trivial outliers have two distinct characteristics. First, they occur very rarely and abruptly. Second, their values are very different from those of the neighboring time steps. Taking these characteristics into account, we propose a simple method to detect and remove trivial outliers, as explained below.

First, we compute the αth and $(100-\alpha)$th percentiles (denoted as p_α and $p_{100-\alpha}$, respectively) for each continuous variable from the data. Then, we define the upper and lower thresholds for the variable as

$$\begin{aligned}\theta_{\text{upper}} &= p_{100-\alpha} + \beta \cdot (p_{100-\alpha} - p_\alpha) \\ \theta_{\text{lower}} &= p_\alpha - \beta \cdot (p_{100-\alpha} - p_\alpha).\end{aligned} \tag{1}$$

In the case of SDS-4 housekeeping data, we found by trial and error that the combination of $\alpha = 0.1$ and $\beta = 5.0$ works better than more conservative settings (e.g., $\alpha = 1.0$, $\beta = 0.5$) for the large part of continuous variables. This is considered to be due to the characteristics of the data that the majority of normal values are concentrated on narrow ranges, whereas the absolute values of outliers are enormously larger, as can be seen in Fig. 3. We will discuss how to choose suitable values of α and β for each variable individually in future.

If a value of y_t at time step t exceeds the threshold, but its previous and next values, i.e., y_{t-1} and y_{t+1}, do not, we judge that y_t is a trivial outlier and remove it. Removed values are recovered by the method described next.

2) Missing Value Completion: As described in Section III-A.5, satellite housekeeping data contain structural missing parts because of the difference in sampling periods and timings among their variables. In addition, they contain very few randomly missing elements caused by removing trivial outliers, as explained above.

In this study, we employed a simple method, the so-called piecewise constant interpolation, in which each missing value is substituted by the immediately preceding value. The reason why we did not choose more advanced methods, such as linear and polynomial interpolation, is that they are subject to noise and may generate other anomalous values.

3) Normalization: As explained in Section III-A.3, it is inconvenient to regard the housekeeping data as ordinary multivariate data, because the diversity in physical units and ranges of the variables or attributes is wide. For example, it would obviously be inappropriate to treat equally one variable representing electric current that ranges from 0.0[A] to 1.0[A] and another variable representing an attitude angle that ranges from -180° to 180°. Therefore, we have to normalize all the continuous variables in the housekeeping data by converting their ranges appropriately or nondimensionalizing them, which enables us to handle all the variables uniformly. Specifically, we reuse the αth and $100-\alpha$th percentiles for each variable, which were computed for detecting trivial outliers in Section III-B.1

$$y_{\text{norm}} = \frac{2}{p_{100-\alpha} - p_\alpha}\left(y_{\text{org}} - \frac{p_{100-\alpha} + p_\alpha}{2}\right). \tag{2}$$

As for the categorical variables, we convert each categorical value to an integer number starting from 1 in descending order of frequency for each variable. For example, if a categorical variable takes two values, "DIS" and "ENA" with probabilities 0.8 and 0.2, respectively, we encode the former to 1 and the latter to 2.

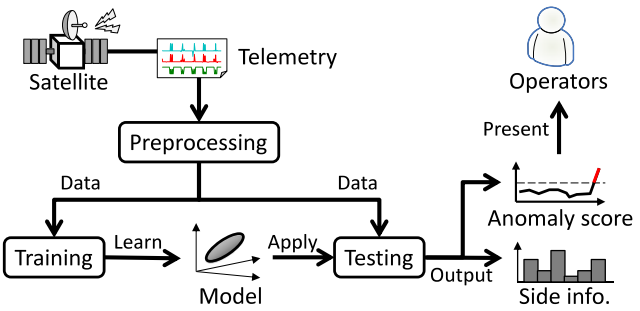


Fig. 4. Flow of data-driven anomaly detection.

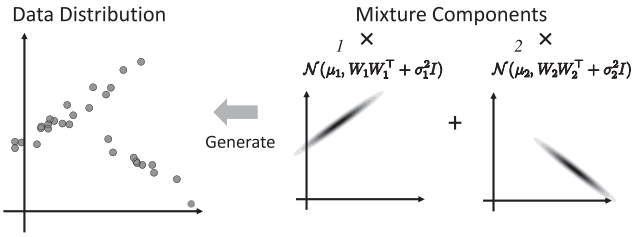


Fig. 5. Modeling of continuous data by MPPCA (in case $K = 2$).

IV. DATA-DRIVEN MONITORING BY CLUSTERING AND DIMENSIONALITY REDUCTION

Fig. 4 illustrates the entire process of the data-driven monitoring and anomaly detection method for satellite housekeeping data proposed in this paper.

The procedure after the data preprocessing described in the previous section is divided into two phases: *training*, in which normal behavior models of the system are learned from the past housekeeping data, and *testing*, in which newly incoming data are monitored using the learned models. While this framework itself is very simple, there is a wide range of possible techniques for modeling the normal behavior of the satellite housekeeping data and learning the models. In particular, it is important to choose the appropriate modeling methods and learning algorithms for the properties of the satellite data, such as high dimensionality, heterogeneity, and multimodality, as discussed in Section III. In this section, we explain the proposed method of modeling, training, testing, and analyzing in detail.

A. Modeling of Satellite Normal Behavior

As described in the previous section, satellite housekeeping data have two major characteristics: high dimensionality and multimodality. In machine learning, these two data properties are frequently treated using techniques called dimensionality reduction and clustering (or mixture models). According to this basic idea, we utilize an integrated model of the mixture of probabilistic principal components analyzers (MPPCA) [22] for modeling the real-valued continuous variables and the categorical mixture distribution for modeling the categorical discrete variables in the housekeeping data. We name this integrated model MPPCACD (mixture of probabilistic principal components analyzers and categorical distributions).

1) Modeling of Continuous Data by Mixture of Probabilistic PCA: Let $\boldsymbol{y}_t \in \mathcal{R}^{D_c}$ denote the data sample vector of the continuous housekeeping data consisting of D_c variables at time t. In the MPPCA model, $\boldsymbol{y}_t$ is assumed to be sampled from the conditional probability distribution $p(\boldsymbol{y}_t|\boldsymbol{x}_t, s_t = k, \boldsymbol{\Theta})$ given a discrete latent variable $s_t \in \{1, 2, \ldots, K\}$ and a low-dimensional continuous latent vector $\boldsymbol{x}_t \in \mathcal{R}^L$ as

$$p\left(\boldsymbol{y}_t|\boldsymbol{x}_t, s_t = k, \boldsymbol{\Theta}\right) = \mathcal{N}\left(\boldsymbol{y}_t|\boldsymbol{W}_k\boldsymbol{x}_t + \boldsymbol{\mu}_k, \sigma_k^2\boldsymbol{I}\right) \tag{3}$$

where $\boldsymbol{\Theta}$ denotes the set of model parameters. The discrete latent variable s_t indicates from which of K clusters (or local models) $\boldsymbol{y}_t$ comes, whereas the continuous latent vector $\boldsymbol{x}_t$ represents the low-dimensional coordinates of $\boldsymbol{y}_t$ in the local model. $\boldsymbol{W}_k$, $\boldsymbol{\mu}_k$, and σ_k^2 denote the factor loading matrix, mean vector, and noise variance for the kth cluster, respectively. We assume the prior distributions of $\boldsymbol{x}_t$ and s_t as follows:

$$p(\boldsymbol{x}_t|\boldsymbol{\Theta}) = \mathcal{N}(\boldsymbol{x}_t|\boldsymbol{0}, \boldsymbol{I}) \tag{4}$$

$$p(s_t|\boldsymbol{\Theta}) = \mathrm{Cat}(s_t|\boldsymbol{\pi}) \tag{5}$$

where $\mathrm{Cat}(\cdot)$ represents the categorical distribution, also called discrete distribution or generalized Bernoulli distribution, and $\boldsymbol{\pi} = [\pi_1, \ldots, \pi_K]^\top$ is the parameter vector whose kth element π_k is the prior probability for the kth cluster. Therefore, the probability that s_t has the value $k \in \{1, 2, \ldots, K\}$ is $p(s_t = k|\boldsymbol{\pi}) = \pi_k$. From the conditional and prior distributions above, the probability density distribution of $\boldsymbol{y}_t$ is derived:

$$p(\boldsymbol{y}_t|\boldsymbol{\Theta}) = \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(\boldsymbol{y}_t|\boldsymbol{\mu}_k, \boldsymbol{C}_k) \tag{6}$$

where $\boldsymbol{C}_k$ represents the covariance matrix for the kth cluster and is computed as

$$\boldsymbol{C}_k \equiv \boldsymbol{W}_k\boldsymbol{W}_k^\top + \sigma_k^2\boldsymbol{I}. \tag{7}$$

As can be seen from (6), the MPPCA model is a special case of the Gaussian mixture model (GMM). The main difference between MPPCA and the more general GMM is that the covariance matrix of each cluster is restricted as (7) in MPPCA. This means that the distribution of data belonging to each cluster is constrained in a low-dimensional (L-dimensional) subspace. It is reasonable to model the continuous housekeeping data by using MPPCA, because the data distribution generally forms a finite number of clusters reflecting the different operational modes of the system and because the variables have strong correlations, which means that each cluster can be approximated by a low-dimensional subspace. Fig. 5 illustrates an example of modeling two-dimensional data distribution by using MPPCA.

2) Modeling of Status Data by Mixture of Categorical Distribution: Let $\boldsymbol{z}_t = [z_t^{(1)}, \ldots, z_t^{(D_s)}]^\top$ denote a categorical status data sample at time t, which is a D_s-dimensional vector consisting of D_s categorical attributes. We assume that each status variable $z_t^{(j)}$ $(j = 1, 2, \ldots, D_s)$

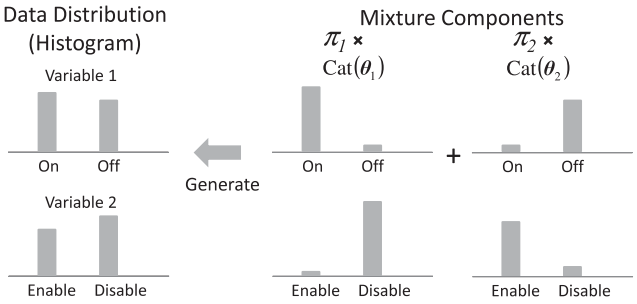


Fig. 6. Modeling of categorical data by categorical mixture (in case $K = 2$).

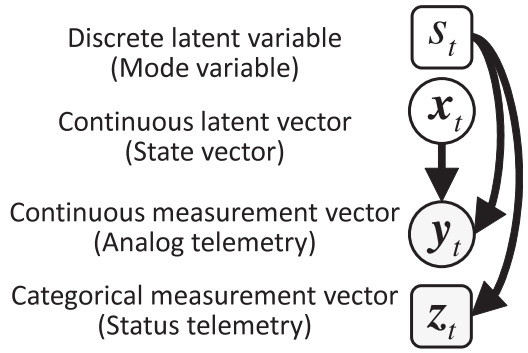


Fig. 7. Graphical representation of the proposed model.

has an integer value ranging from 1 to M_j according to the categorical distribution, the parameters of which depend on the discrete latent variable s_t

$$p\left(z_t^{(j)}|s_t = k\right) = \mathrm{Cat}\left(z_t^{(j)}|\boldsymbol{\theta}_k^{(j)}\right) \tag{8}$$

where $\boldsymbol{\theta}_k^{(j)} = [\theta_{k,1}^{(j)}, \ldots, \theta_{k,M_j}^{(j)}]^\top$ denotes the parameter vector for this categorical distribution. That is to say, $p(z_t^{(j)} = l|s_t = k) = \theta_{k,l}^{(j)}$. In the case of satellite systems, for example, this assumption means that each status variable is likely to take a specific value depending on the operational mode. In addition, we use the so-called "naive" assumption that the categorical variables are conditionally independent given the value of s_t

$$p\left(\boldsymbol{z}_t|s_t = k\right) = \prod_{j=1}^{D_s} \mathrm{Cat}\left(z_t^{(j)}|\boldsymbol{\theta}_k^{(j)}\right). \tag{9}$$

As a result, the distribution of status data is modeled by the multidimensional categorical mixture as

$$p\left(\boldsymbol{z}_t|\boldsymbol{\Theta}\right) = \sum_{k=1}^{K} \pi_k \cdot \prod_{j=1}^{D_s} \mathrm{Cat}\left(z_t^{(j)}|\boldsymbol{\theta}_k^{(j)}\right). \tag{10}$$

Fig. 6 illustrates the manner in which two-dimensional categorical data are modeled by the categorical mixture distribution model.

As seen above, in the proposed method, the real-valued continuous data are modeled by MPPCA, whereas the discrete-valued categorical data are modeled by categorical mixture. It should be noted here that the discrete latent variable s_t is shared by the two models. Fig. 7 shows a graphical representation of the integrated model.

Finally, the MPPCACD model is derived as the joint probability density distribution of the continuous and categorical measurement variables and is expressed as follows:

$$p\left(\boldsymbol{y}_t, \boldsymbol{z}_t|\boldsymbol{\Theta}\right) = \sum_{k=1}^{K} \left(\pi_k \mathcal{N}\left(\boldsymbol{y}_t|\boldsymbol{\mu}_k, \boldsymbol{C}_k\right) \prod_{j=1}^{D_s} \mathrm{Cat}\left(z_t^{(j)}|\boldsymbol{\theta}_k^{(j)}\right) \right) \tag{11}$$

Algorithm 1: EM Algorithm for MPPCACD Model.

1: Initialize parameter estimate $\hat{\boldsymbol{\Theta}} = \{\hat{\pi}_k, \hat{\boldsymbol{\mu}}_k, \hat{\boldsymbol{W}}_k, \hat{\sigma}_k^2, \hat{\boldsymbol{\theta}}_k\}$ $(k = 1, \ldots, K)$ [See the Appendix (A).]
2: **repeat**
3: Compute $r_{t,k}$ or posterior of s_t given $\hat{\boldsymbol{\Theta}}$ $(t = 1, \ldots, T, k = 1, \ldots, K)$ (by (20)).
4: Update $\hat{\pi}_k$ and $\hat{\boldsymbol{\mu}}_k$ $(k = 1, \ldots, K)$ using $\{r_{t,k}\}$ and $\boldsymbol{y}_{1:T}$ (by (22) and 23).
5: Update $\hat{\sigma}_k^2$ and $\hat{\boldsymbol{W}}_k$ $(k = 1, \ldots, K)$ using $\{r_{t,k}\}$, $\boldsymbol{y}_{1:T}$ and updated $\{\hat{\boldsymbol{\mu}}_k\}$ (by (25) and 26).
6: Update $\hat{\boldsymbol{\theta}}_k^{(j)}$ $(j = 1, \ldots, D_s, k = 1, \ldots, K)$ using $\{r_{t,k}\}$ and $\boldsymbol{z}_{1:T}$ (by (29)).
7: **until** Converged

where $\boldsymbol{\Theta}$ is the set of parameters in this model, and is composed as

$$\boldsymbol{\Theta} = \left\{\pi_k, \boldsymbol{\mu}_k, \boldsymbol{W}_k, \sigma_k^2, \boldsymbol{\theta}_k^{(j)}\right\} \quad (j = 1, \ldots, D_s, k = 1, \ldots, K). \tag{12}$$

B. Learning of the MPPCACD Model

In the training phase, the values of model parameters $\boldsymbol{\Theta}$ are learned from the training data, which are assumed to contain only normal past housekeeping data. This is achieved by finding the value of $\boldsymbol{\Theta}$ that maximizes the log-likelihood function

$$L(\boldsymbol{\Theta}) = \sum_{t=1}^{T} \ln p(\boldsymbol{y}_t, \boldsymbol{z}_t|\boldsymbol{\Theta}). \tag{13}$$

As the optimal solution cannot be obtained analytically, we utilize the expectation–maximization (EM) algorithm to obtain the solution numerically.

Briefly stated, the EM algorithm repeats the E-step in which a lower bound for the log-likelihood function of (13) is constructed using the current estimates of the parameters $\boldsymbol{\Theta}$, and the M-step in which the lower bound is maximized with respect to the parameters $\boldsymbol{\Theta}$, until convergence. We summarize the EM algorithm for the MPPCACD model in Algorithm 1. More details, including update equations, are presented in the Appendix (VII). The derivation of

the equations is omitted in this paper, because it is a straightforward extension of that for MPPCA [22].

A nontrivial problem in the learning of discrete and continuous latent variable models is the determination of K—the number of mixture components (or clusters) and L—the dimensionality of continuous latent space. This issue is known as the model selection problem in machine learning, and a variety of approaches, including cross validation, information criteria, and Bayesian model selection, have been studied. In fact, we tested some of the approaches, but the results were not reliable in that they tended to recommend excessive values of K and L. Therefore, we now employ a heuristic rule to determine these parameters, as follows. First, we choose the value of L by using the so-called "elbow law" after applying the ordinary PCA to the continuous housekeeping data. Then, we manually decide K based on the scatter plots of the principal component scores. The development of an automatic and better model selection method is a significant future work that we intend to perform.

C. Monitoring and Analysis

When the model parameters have been estimated from the training data, the learned model can be utilized to evaluate whether new data samples $\{\boldsymbol{y}_t, \boldsymbol{z}_t\}$ $(t = T+1, \ldots, T+T')$ are normal or anomalous. Specifically, we define the anomaly score of a data sample $a(\boldsymbol{y}_t, \boldsymbol{z}_t, |\hat{\boldsymbol{\Theta}})$ as the minus log likelihood when the model is given

$$a\left(\boldsymbol{y}_t, \boldsymbol{z}_t, |\hat{\boldsymbol{\Theta}}\right) \equiv -\ln p\left(\boldsymbol{y}_t, \boldsymbol{z}_t, |\hat{\boldsymbol{\Theta}}\right). \tag{14}$$

The right-hand side of (14) is computed by substituting the parameter estimate $\hat{\boldsymbol{\Theta}}$ into (11). It should be noticed that the anomaly monitor outputs the degree of anomaly of each data sample as a real value, rather than whether the sample is anomalous or normal as a binary value. This is because it is not trivial to set a threshold value that discriminates anomalous samples from normal ones. If we had enough number of real anomaly data samples, we could determine an optimal threshold by considering the trade-off between false positives and false negatives. However, it was impossible in this study, because the housekeeping data of SDS-4 contains only a limited number of anomalous samples, which is the case with many other satellites. Even if we could manage to choose a reasonable threshold value on the score, false alarms may occur because of the noise and trivial outliers that could not be removed by the data preprocessing. In addition, as a fate of the data-driven system monitoring, some normal data samples are considered abnormal unless their similar patterns are contained in the training data. Therefore, even when there are nontrivial increases in the anomaly score, we should not too hastily reach the conclusion that some serious faults have occurred in the system, but rather investigate the reasons for them carefully. In fact, we can obtain useful information about which part is more suspicious than the other by analyzing the contribution of each variable to the anomaly score, as follows. First, we compute the maximum *a priori* (MAP) estimate of the latent variables s_t and $\boldsymbol{x}_t$ when the data sample $\{\boldsymbol{y}_t, \boldsymbol{z}_t\}$ at each time step is given

$$\hat{s}_t = \arg\max_{s_t} p\left(s_t | \boldsymbol{y}_t, \boldsymbol{z}_t, \hat{\boldsymbol{\Theta}}\right) \tag{15}$$

$$\hat{\boldsymbol{x}}_t = \arg\max_{\boldsymbol{x}_t} p\left(\boldsymbol{x}_t | \boldsymbol{y}_t, \hat{s}_t, \hat{\boldsymbol{\Theta}}\right). \tag{16}$$

Then, the reconstruction of continuous data sample $\hat{\boldsymbol{y}}_t$ and the reconstruction error $\boldsymbol{e}_t$ are computed by

$$\hat{\boldsymbol{y}}_t = \hat{\boldsymbol{W}}_{\hat{s}_t}\hat{\boldsymbol{x}}_t + \hat{\boldsymbol{\mu}}_{\hat{s}_t} \tag{17}$$

$$\boldsymbol{e}_t = \boldsymbol{y}_t - \hat{\boldsymbol{y}}_t. \tag{18}$$

As the absolute value of each element of the reconstruction error $cy_t^{(j)} \equiv |e_t^{(j)}|$ is the difference between the actual measurement and the value predicted by the learned model, it can be considered to indicate the extent to which the corresponding variable contributes to the anomaly score. As for the status telemetry $\boldsymbol{z}_t$, it is impossible to compute the reconstruction error, because the variables are categorical or qualitative. Instead, we employ the minus marginal log likelihood defined by the following equation as the contribution to the anomaly:

$$cz_t^{(j)} \equiv -\ln p\left(z_t^{(j)} | \hat{s}_t, \hat{\boldsymbol{\Theta}}\right). \tag{19}$$

V. EXPERIMENT

A. Small Demonstration Satellite 4 (SDS-4)

JAXA has been running the SDS program since 2006 for the purpose of demonstrating new components and technologies in orbit and training young engineers. SDS-4 is a microsatellite with a mass of about 50 kg, launched in May 2012 as a part of the SDS program[23]. The operation of SDS-4 is rather complicated because its main mission is to demonstrate new devices, which makes it difficult to monitor the health status of the system. In November 2012, SDS-4 entered the postmission phase. Its basic operation schedule consists of a four-week cycle with a different demonstration experiment conducted each week, as shown in Table I.

TABLE I
Basic operation Cycle of SDS-4

Week	Main mission	Typical operations
1	Bus system exp.	MCMR image capt., attitude control.
2	SPAISE exp.	AIS signal receipt, attitude control.
3	IST/FOX exp.	FOX heater ON/OFF, attitude control.
4	SPAISE exp. (backup)	AIS signal receipt, attitude control.

While the SDS-4 telemetry was originally composed of a total of 1458 variables, many of these are considered to be irrelevant to the purpose of health monitoring. Therefore, we exclude the following variables.

1) Variables with extremely low sampling rates.

TABLE II
Numbers of Continuous and Status Telemetry Variables Used in Experiment

Subsys.	Continuous	Status
Attitude Control Subsys. (ACS)	17	20
Electrical Power Subsys. (EPS)	17	25
Thermal Control Subsys. (TCS)	40	–
Command and Data Handling Subsys. (CDH)	1	16
Transmitter and Receiver Subsys. (TRX)	4	11
Others	10	284
Total	**89**	**356**

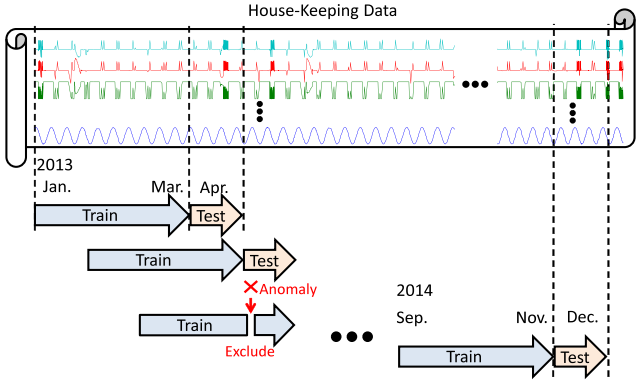


Fig. 8. Training and testing cycle in validation experiment. Data samples detected as anomalous are excluded from the later training data sets.

2) Variables with low variation, i.e., that are almost constant over time.
3) Variables that are available only when the satellite is communicating with the ground station.
4) Variables that are unrelated to the system health.

As a result, 89 continuous and 356 status variables were chosen for monitoring. The numbers of the chosen variables grouped by subsystems are listed in Table II.

B. Overview

In this verification experiment, we tested the proposed data-driven monitoring method on the SDS-4 housekeeping data for about two years from January 2013 to December 2014.

Taking the basic cycle of the operation schedule mentioned above into the consideration, we employed a monitoring cycle comprising three months of training, followed by one month of testing, as illustrated in Fig. 8. If some portion of testing data is detected as anomalous, that part is excluded from the later training datasets.

Based on the heuristic rule described in Section IV-B, we chose the values of K (number of clusters) and L (dimension of continuous latent vector). Fig. 9 shows the proportion of eigenvalues in descending order (blue line) and cumulative proportion of them (red line), when PCA is applied to the normalized continuous data. We can see the "elbow point" lies between 4 and 8, whereas the cumulative proportion reaches 0.9 (90%)

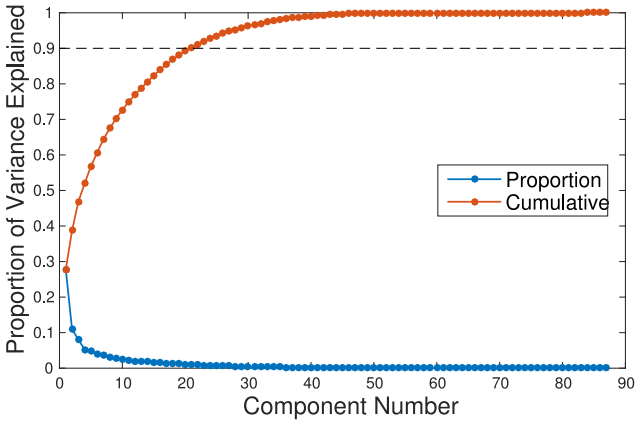


Fig. 9. Proportion and cumulative proportion of variances explained by eigenvalues of covariance matrix in SDS-4 continuous data. Based on this result, we set $L = 15$ in MPPCACD model.

when we take top-20 eigenvalues. Therefore, we set L to 15. Regarding K, we referred to the distribution of samples in a low-dimensional space. As we can observe 5–10 clusters or groups in Fig. 1, we set K to 8.

As a result of applying the proposed method to the data of two years, we obtained 86 cases in which the anomaly score was significantly large. We presented these cases to the SDS-4 operators and asked them to evaluate the detected anomalies. Unfortunately, some of these were determined to be trivial anomalies caused by erroneous values leading to false alarms in which normal but nonfrequent operations were mistakenly detected. Nonetheless, we also found that the proposed method successfully detected many nontrivial anomalies. In the remaining part of this section, we show several typical cases in this experiment.

In cases 1 and 2, we also conducted comparative experiments with one-class support vector machine (OCSVM) [24] and support vector data description (SVDD) [25]. Here, we used LIBSVM (ver. 3.1.8) [26] that includes these algorithms. In these algorithms, RBF kernel were employed, as it was much better than dot (linear) kernel. Their parameters (including kernel parameters) were manually tuned, because automatic selection of parameter values of anomaly detection algorithms is difficult, as we explained before. While we used the same training and testing datasets, we converted the categorical variables to numerical ones by *one-of-K coding* for OCSVM and SVDD so that the RBF kernel could deal with all variables together.

C. Case Study 1: Normal Operation

In the first case, we trained an MPPCACD model with the data for the period from August 2014 to October 2014, and then, evaluated the testing data for November 2014. Fig. 10 shows the transitions of 89 continuous variables for the testing period. Note that they are vertically stacked for visibility. Fig. 11 shows the transitions of 356 status variables for the same period. In this figure, different categorical values except the most frequent one are indicated by different colors, for each status variable.

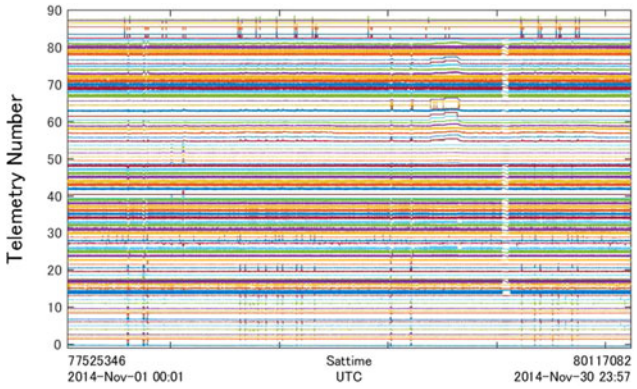


Fig. 10. Continuous telemetry data on November 2014.

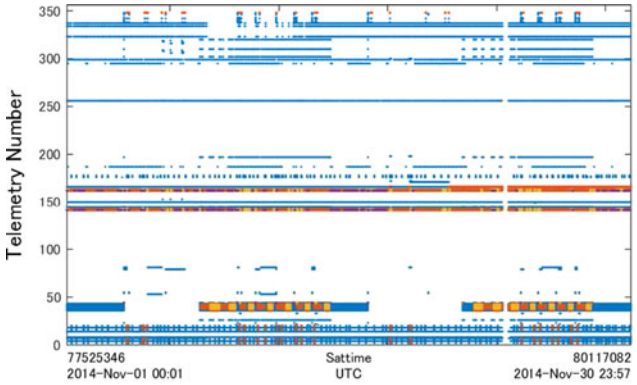


Fig. 11. Status telemetry data on November 2014.

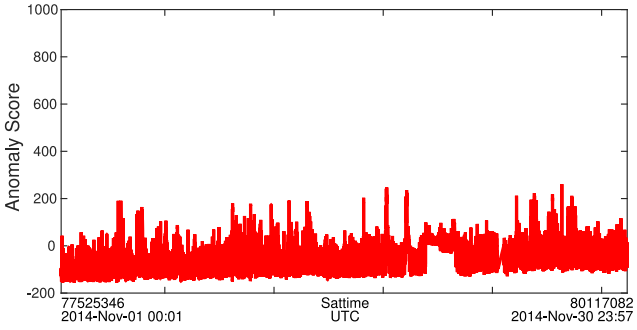


Fig. 12. Anomaly score (MPPCACD) on November 2014.

Fig. 12 shows the resultant anomaly score $a(\boldsymbol{y}_t, z_t|\hat{\boldsymbol{\Theta}})$ for the testing period. In this figure, we can see that the anomaly score most frequently remains within the range from -200 to 200 in general, and that it takes higher values occasionally and remains at a higher level (more than 0) for a short period.

Next, let us examine the results for each day in detail. Fig. 13 shows the transition of the anomaly score for November 3. Note that there were no special operations, such as device experiments, on that day. We confirmed that the proposed method presents a low anomaly score when the satellite is operating normally. Fig. 14 shows the MAP estimate of $\hat{s}_t$ over time, which indicates to which cluster the data point is assumed to belong. We can see that the

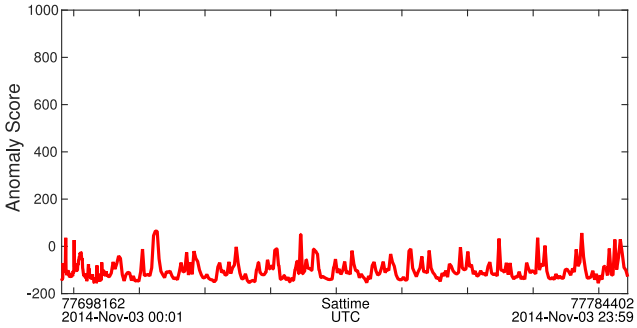


Fig. 13. Anomaly score (MPPCACD) on November 3, 2014 (normal operation).

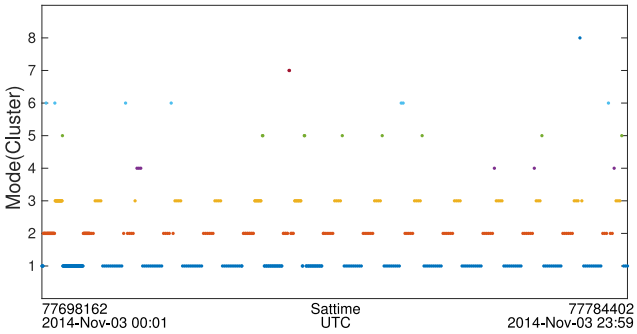


Fig. 14. Estimated discrete latent variable on November 3, 2014.

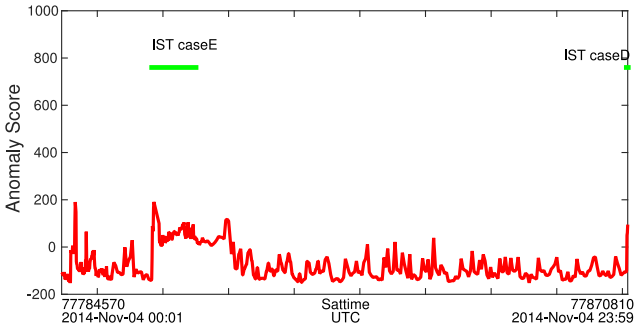


Fig. 15. Anomaly score (MPPCACD) on November 4, 2014 (IST experiment was conducted).

state of the satellite system regularly or periodically moves from one cluster to another.

In contrast, Fig. 15 shows the anomaly score for November 4. On this day, an In-flight experiment of Space materials using THERME (IST) experiment, which is one of the main missions of SDS-4, was conducted, which caused a temporary increase in the score up to about 200. Similar phenomena were observed for other experimental operations, such as SPAISE, FOX: FHP On-orbit Experiment (FOX), and MCMR. While these operations are not anomalies in fact, the anomaly score tends to be higher than for ordinary operations, because the proportion of these experimental operations of the whole data is relatively small.

Fig. 16 shows the anomaly score for the period from November 19 to 22. In this figure, we can see the score remains at a relatively high level during the period from November 20 07:40 to November 21 18:00. It was determined that this was due to the change in threshold for battery

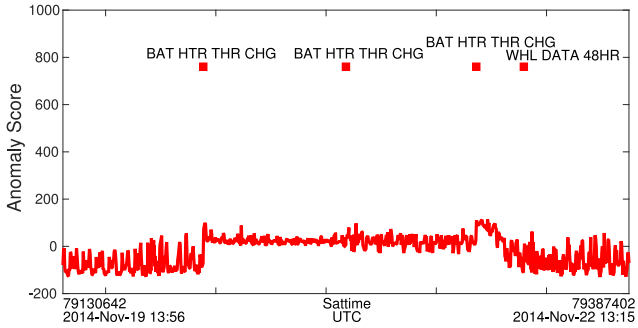


Fig. 16. Anomaly score (MPPCACD) from November 19 to 22, 2014 (battery heater threshold changed).

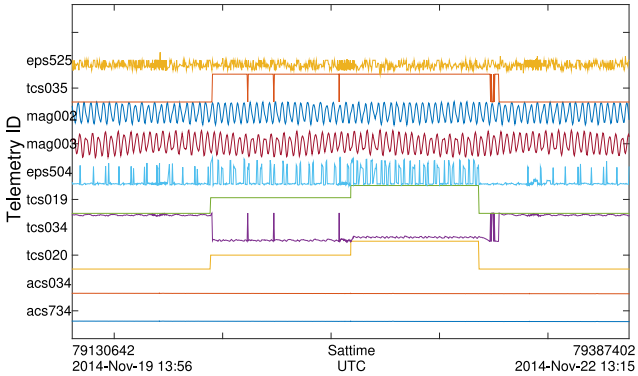


Fig. 17. Top ten contributing continuous variables from November 19 to 22, 2014.

heaters. As this operation is relatively rare (performed approximately once per month on average), this result is reasonable. Fig. 17 shows the transition of the top ten continuous variables contributing to the anomaly score in this period. We can confirm that several variables in the TCS subsystem, which is closely related to the battery heater, indicate characteristic patterns in the corresponding period.

Fig. 18 shows the anomaly scores for November 2014 when OCSVM and SVDD are applied. Compared with the anomaly score by MPPCACD (see Fig. 12), the scores by OCSVM and SVDD varied widely, which might cause false alarms, though there were no significantly large score values. A potential problem here is that it is difficult to understand why the anomaly scores go up or down and how they are related to the health status of the satellite system.

D. Case Study 2: Anomalous and Rare Events

The next example is the result of evaluating the testing data in June 2014 by using the model trained on the data from March 2014 to May 2014. Fig. 19 shows the anomaly score in this month. We can see that the score increased nearly to 500 three times, on June 5, 15, and 20, respectively. Interestingly, the reasons for the high anomaly score vary among these three cases, which we explain in detail below.

1) Detection of a Rare Experiment: Fig. 20 shows the anomaly score on June 5, 2014. We referred to the operation record of SDS-4, and found that the time period in which the score was high coincided with that in which an experiment related to an EPS subsystem named "SA-IV evaluation" was conducted for the first time. In other words, there were no similar operation patterns in the past data of SDS-4. Therefore, it is quite reasonable that any data-driven monitoring method, including ours, would consider this event anomalous. Fig. 21 shows the transition of the top ten continuous variables contributing to the anomaly score on that day. We can confirm that the several variables in the TCS and EPS subsystems that are related to the experiment behaved differently from others.

2) Detection of an Attitude Anomaly: Fig. 22 shows the anomaly score on June 15, 2014. There was no record of any special operation, such as component experiments, having been conducted in the time period in which the score increased. Fig. 23 shows the transition of the top ten contributing continuous variables. There are big changes in some variables, such as attitude angles and angular velocities, that are related to the attitude control subsystem.

The SDS-4 operators later investigated this case in detail and found that an anomaly indeed occurred in which the attitude angle error increased for some reason and caused a change in the attitude control mode from NOMINAL to SHIFT. This means that the proposed method successfully detected the anomaly of the spacecraft system. From this and other similar cases, we learned that we should pay more attention to the cases in which the anomaly score is significantly high but there are no records of particular operations.

3) False Alarm: Fig. 24 shows the anomaly score on June 20. The time period in which the score increased coincides with the recorded time of an FOX experiment. While this pattern seems to be similar to the case in Section V-D.1, it differs in that FOX experiments are common in the operation of SDS-4 and are contained in the training data. Although SDS-4 operators examined this case in detail, they could not find any symptoms of abnormal system behavior and concluded that the increase in the anomaly score was a false alarm.

A reason for this false alarm is that variety in the experiment configuration and parameter settings of the FOX experiment is wide, which makes it difficult for the trained model to cover all the possible normal patterns.

4) Comparison With OCSVM and SVDD: Fig. 25 shows the anomaly scores when OCSVM and SVDD were applied to the same datasets. While these two methods presented high anomaly scores on June 5 (rare event) and June 15 (attitude anomaly), they were not so on June 20 (regular experiment). It indicates that they have lower risk of issuing a false alarm to this event. On this point, we admit that OCSVM and SVDD performed better than MPPCACD. However, the anomaly scores by them varied widely in this case too, which may cause other false alarms.

E. Case Study 3: Training Models Specialized for Experiment Operations

As seen in the above cases, the anomaly score in the time periods in which the four main experiments (IST, SPAISE,

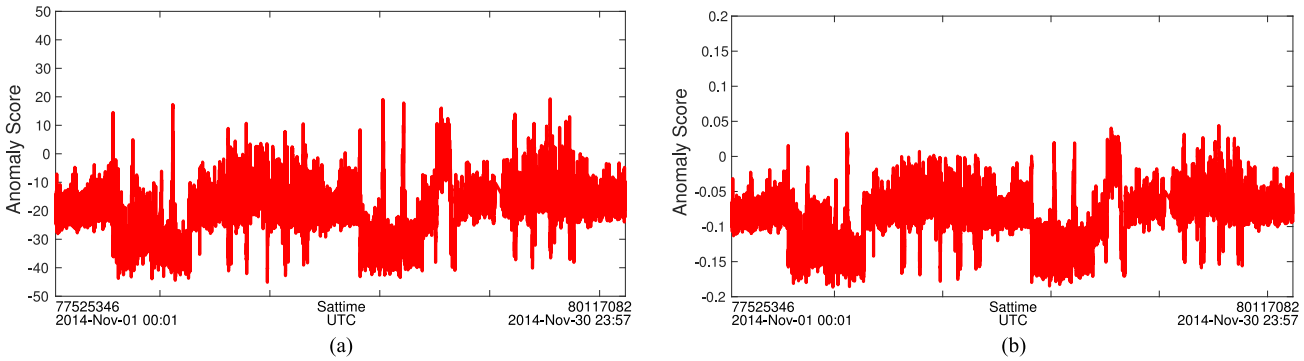


Fig. 18. Anomaly scores (OCSVM and SVDD) on November 2014. (a) OCSVM. (b) SVDD.

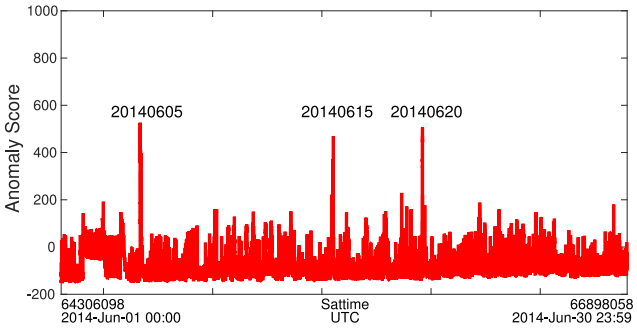


Fig. 19. Anomaly score (MPPCACD) on June 2014.

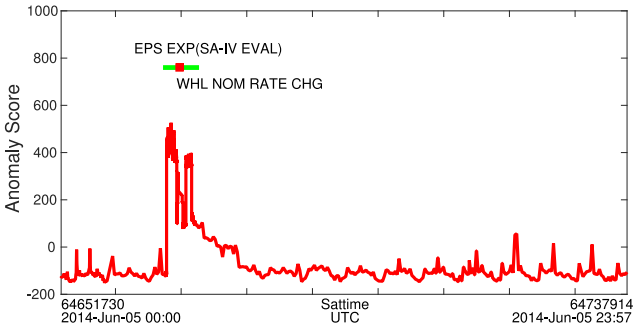


Fig. 20. Anomaly score (MPPCACD) on June 5, 2014.

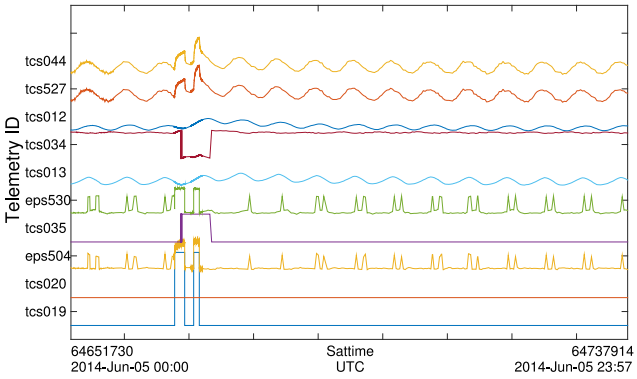


Fig. 21. Top ten contributing continuous variables from June 5, 2014.

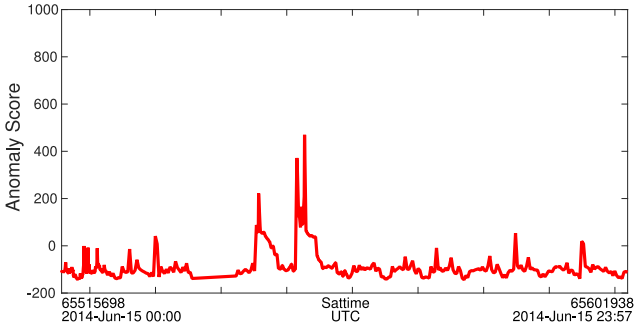


Fig. 22. Anomaly score (MPPCACD) on June 15, 2014.

FOX, and MCMR) were performed tends to be significantly higher than that in the nominal operation, when we simply use the housekeeping data of the past three months for model training. In particular, in some FOX and MCMR experiments, the anomaly score became sufficiently high to be considered a false alarm. We consider the main reason for this phenomenon is that the proportion of the data during the experiments is much smaller than that of the data during the nominal operation. In fact, the proportion of time spent on these experiments is only about 5% of the operation of SDS-4. Therefore, the learned model is expected to emphasize the dominant nominal operation, rather than normal but minor experimental operations.

A simple and reasonable method to obtain a model that accounts better for the spacecraft's behavior during an experimental operation is to increase the proportion of data during the experiments. In the last case study, we report the result of applying this idea. Specifically, instead of using all the data for the three months, as in the previous case, we beforehand extracted 85 short periods in which any of the four experiments were conducted from the time period between March 1, 2014 and May 31, 2014 and concatenated them into a training dataset. We evaluated again the SDS-4 data of June 2014 using the MPPCACD model learned from the concentrated training data. Fig. 26 shows the anomaly score on June 20 in this case. We can see that in comparison with Fig. 24 the anomaly score during the FOX experiment is suppressed as expected, which decreases the risk of false alarms. However, when we examine the anomaly score of the whole of June 2014, as shown in Fig. 27, the score for June 5 and 15 on which a true anomaly or rare events occurred remained high. This result indicates the availability

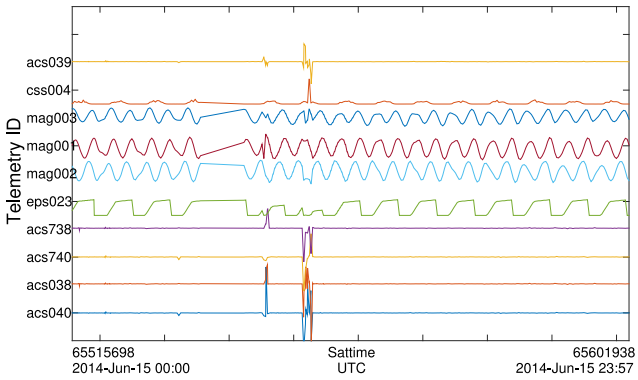


Fig. 23. Top ten contributing continuous variables from June 15, 2014.

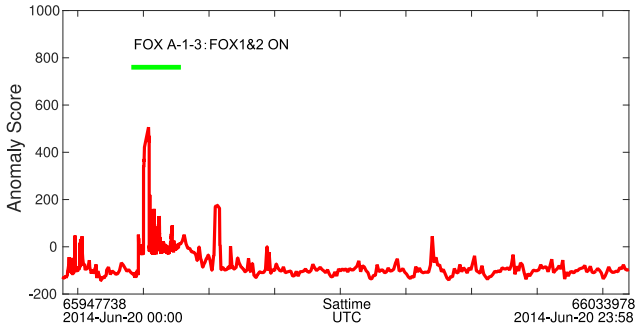


Fig. 24. Anomaly score (MPPCACD) on June 20, 2014.

of more appropriate models for monitoring housekeeping data, by carefully composing the training data with side information, such as operation records and operators' expert knowledge.

F. Case study 4: Monitoring Subsets of Variables

While we have assumed that all the variables are monitored together so far, it is also possible to apply the proposed method to subsets of the variables. Especially when the system has a large number of measurement variables, it is reasonable to divide the whole set into meaningful subsets and monitor them in parallel as it will provide useful information to identify the cause of anomaly.

Therefore, we conducted another experiment to examine the validity of this idea, in which MPPCACD (or MPPCA) was applied to the major three subsystems—ACS, EPS, and TCS in parallel. The numbers of continuous and categorical variables in each subsystem are listed in Table II. The parameter K was set to 8 in all subsystems, while L was set to 5 in ACS and EPS and 10 in TCS, taking the number of continuous variables into consideration. Fig. 28 shows the anomaly scores for these three subsystems in June 2014, where MPPCACD (MPPCA for TCS) models were trained on the data for the period from March 2014 to May 2014.

By comparing these results with Fig. 19 (the case in which MPPCACD was applied to the whole system), we found several interesting things. On June 5, the anomaly scores of EPS and TCS variable groups were significantly high, while that of ACS variables were not. This is considered to be reasonable, because an EPS-related experiment that had never been done before was performed on that day. On June 15, the anomaly score of ACS variables was significantly high. This is reasonable because the satellite encountered an unexpected change of attitude control mode on that day, as explained in Section V-D.2. On June 20, we had a false alarm in which a regular FOX experiment was mistaken for anomaly, as described in Section V-D.3. From Fig. 28(a)–(c), we can tell TCS played a role in this false alarm, because only the anomaly score of TCS was high on this day.

G. Case Study 5: Changing the Length of the Training Period

While we had fixed the training period to 3 months in this experiment as the operation cycle of SDS-4 is every four weeks, we made another case study in which the training period was extended to 12 months. Fig. 29 shows the anomaly scores on June 2014 using 12 months of training data from June 2013 to May 2014. Comparing this result with Fig. 19, we found that using the data of 12 months for training is better in the following two aspects. First, the anomaly score on June 20 is decreased, which means the risk of false alarm on this event is reduced. Second, the baseline of the score remains almost unchanged, while that of the case of using 3 months data for training is gradually increasing. This result is reasonable, because 12 months of training data comprises the seasonal changes caused by the earth's revolution around the sun, as well as because the data of longer period covers the distribution of normal data more completely.

VI. DISCUSSION

From the experimental study described in the previous section, we learned a variety of lessons, in particular, concerning the difficulties involved in applying data-driven health monitoring methods to the housekeeping data of artificial satellites that are actually in operation and clues to how we can overcome them in future. In the remaining part of this section, we focus on three issues that are considered to be very important.

A. False Alarm and Performance Evaluation

An ideal anomaly detector free of false alarms would never consider normal behavior patterns to be anomalies, even if it had not seen similar patterns previously. Data-driven or learning-based anomaly detection methods, including the proposed one, however, are intended to detect any unseen events or patterns that are significantly different from the past ones as anomalies, regardless of the actual severity of their occurrence for the systems. In other words, the data-driven monitoring methods detect *statistical* anomalies, which do not necessarily coincide with the system failures. A difficult problem arises in nonstationary systems such as SDS-4 in that normal patterns that have not been experienced in the past can always exist. They are statistical anomalies, but not system failures in fact. This

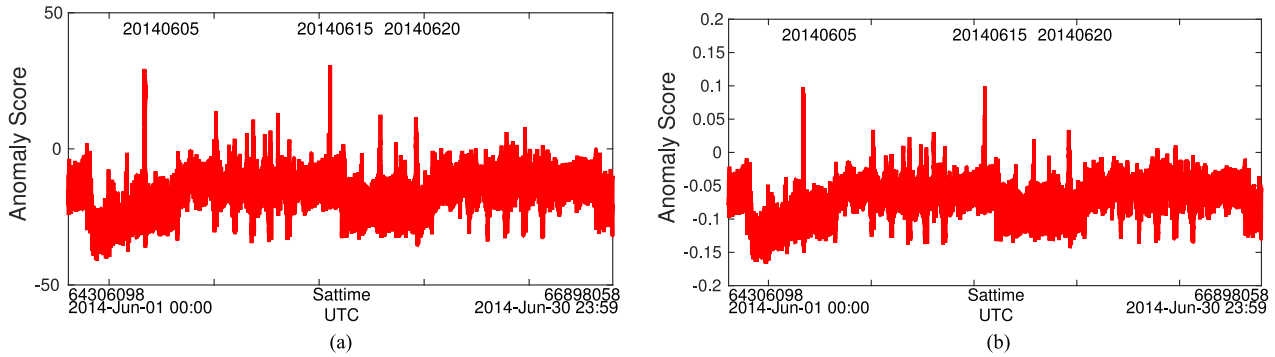


Fig. 25. Anomaly scores (OCSVM and SVDD) on June 2014. (a) OCSVM. (b) SVDD.

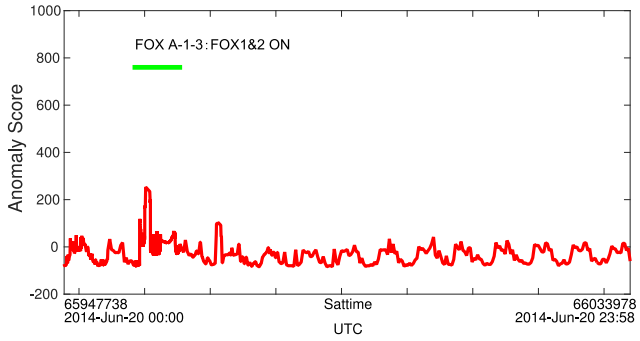


Fig. 26. Anomaly score (MPPCACD) on June 20, 2014 with a specialized model.

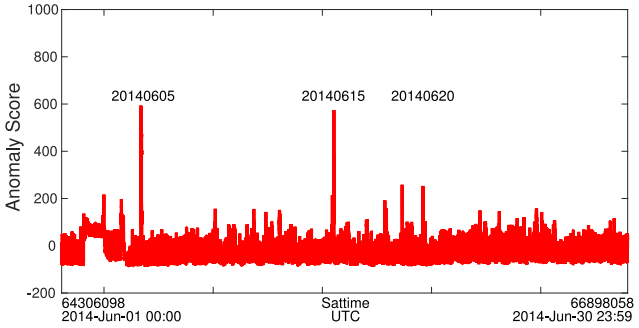


Fig. 27. Anomaly score (MPPCACD) on June 2014 with a specialized model.

is the fundamental reason why the data-driven approach cannot completely avoid false alarms.

Another related issuew is how we can evaluate an anomaly detection method in practice when we do not have enough anomalous samples. Although the goal of an anomaly detection is to detect any anomalous pattern even if it has never been experienced, we need a large number of such unexperienced anomalous data samples in advance if we want to evaluate its performance quantitatively. This is a common dilemma for researchers of anomaly detection and health monitoring, especially for those who are focused on critical systems in the real world. We believe the best things we can do for now are to output the anomaly score indicating the degree of likelihood that each data sample is anomalous and to present and analyze case studies on the limited number of real anomaly examples, as we did above.

The lesson we learned in this discussion is that we should not simply judge whether the system is in serious trouble or not only by the magnitude of the anomaly score, but rather should try to understand what is occurring in the system using the posteriori analysis mentioned in Section IV-C. That is to say, we should utilize the data-driven anomaly detection methods for the purpose of wide and shallow monitoring, allowing a small number of false alarms.

B. From Ofline Learning to Online Learning

The data-driven anomaly detection method we proposed is based on the framework of offline (or batch) learning, in which training and testing datasets are prepared separately and the former is used for learning a model and the latter evaluated by the learned model. Although the offline learning algorithm is easy to use, it has some potential drawbacks as follows. First, it is inefficient, because we have to relearn the model from scratch whenever the training dataset is renewed periodically. In addition, there may be discontinuity between the old and new models. Second, if the environmental conditions change slowly or the data distribution is nonstationary, the anomaly score gradually increases even if the system itself remains normal.

Therefore, an online learning method that updates the model each time a new data sample arrives is preferable. In addition to its efficiency, the online learning algorithm has the merit that it reduces the number of false alarms. When it detects that a data pattern is an anomaly but the operators' investigation determines it to be a false alarm, it modifies the model so that it does not commit similar false alarms.

From another point of view, the online learning algorithm enables the data-driven anomaly detector and human operators to be more interactive and cooperative. The anomaly detector utilizes the operators' feedback to improve the detection performance immediately, whereas the operators are motivated to understand the health status of the satellite systems by examining the detector's outputs.

C. Extensions to Temporal Models

In this study, we focused mainly on high dimensionality, multimodality, and heterogeneity, which are three important

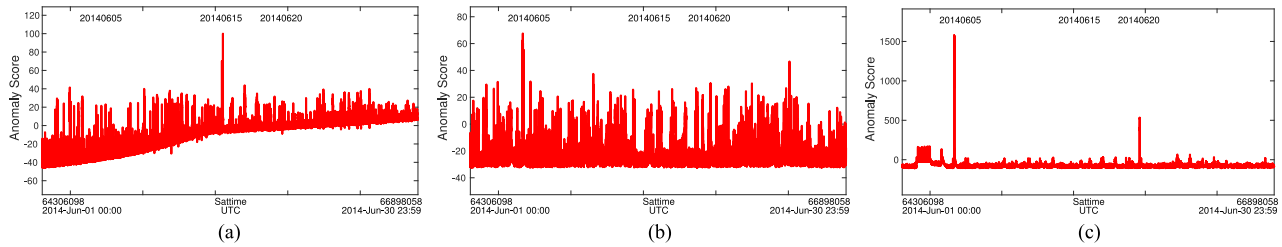


Fig. 28. Anomaly scores in June 2014, when MPPCACD is applied to ACS, EPS, and TCS subsystems. (a) ACS. (b) EPS. (c) TCS.

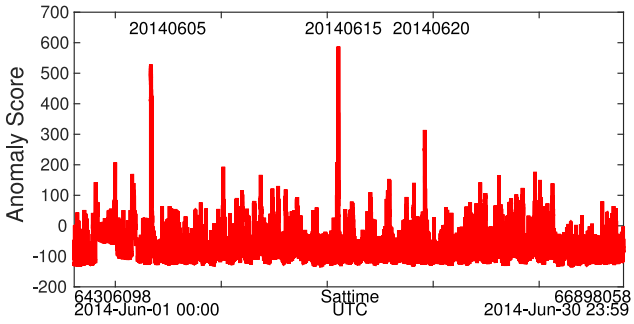


Fig. 29. Anomaly score (MPPCACD) on June 2014 with 12 months of training data from June 2013 to May 2014.

properties of the satellite housekeeping data, and employed the MPPCACD model that combines MPPCA and categorical mixture models as a result. However, we had to ignore the temporal dependency that is another property of housekeeping data, because of the irregularity of sampling periods in the telemetry of the actual satellite (SDS-4), as explained in Section III-A.4. It is, of course, desirable to employ an advanced model that takes the temporal dependency into account, after solving the technical issues in future.

The most straightforward means of handling temporal dependency is to introduce a transition model for the continuous and discrete latent variables $\boldsymbol{x}_t$ and s_t. Specifically, the proposed model can be enhanced by HMM, switching linear dynamical systems, and more general dynamic Bayesian networks [27]. Although this extension seems to be sufficiently reasonable, it requires some prior knowledge about the system, because the model becomes more complex, whereas the training data are finite in practice. In other words, we consider the data-driven health monitoring approach will be hybridized with the knowledge-driven (in particular, model-based) approach in future.

D. Comparison With State-of-the-Art Anomaly Detection Algorithms

In the previous section, we conducted a comparative experiment with OCSVM and SVDD that are state-of-the-arts anomaly detection algorithms. In terms of detection performance, we found there were merits and demerits to both our proposed method and OCSVM and SVDD, but the difference was not crucial. We insist, however, that MPPCACD has some advantages for the purpose of health monitoring of satellite housekeeping data. From the case studies in Sections V-C and V-D, we observed the anomaly score by MPPCACD is constantly low when the system is normal, and is distinguishable from the high score when an anomaly occurs. On the other hand, the anomaly scores by OCSVM and SVM varies widely even when the system is normal. This is considered to be because OCSVM and SVDD (with RBF kernels) are general-purpose methods that can deal with any non-linear distribution, while MPPCACD is a special-purpose method for the satellite housekeeping data that has properties described in Section III-A. In addition, MPPCACD provides useful information that helps operators understand the health status of the system. By comparing the original data and reconstructed one by (17), we can investigate which variables are contributing to the anomaly score, which helps us identify the cause of anomaly. We can also monitor the transition of operational modes by estimating the value of discrete latent variable s_t with (15). As shown in Fig. 14, the transition pattern tends to be regular when the system is normal.

VII. CONCLUSION

In this paper, we proposed a data-driven health monitoring method based on probabilistic clustering and dimensionality reduction for artificial satellites. The proposed method is suitable for modeling spacecraft housekeeping data, which is high-dimensional, multimodal, and heterogeneous. It learns the model parameter values from the training data, and then, evaluates the anomaly score for each data sample in the testing data. It also provides side information, such as reconstruction errors and the responsibilities of each cluster, which helps operators understand the health status of satellite systems. Another major contribution of this paper is that it demonstrated the validity of the proposed method by applying it to the housekeeping data of an experimental artificial satellite for more than two years. We are planning to develop a general-purpose monitoring system for spacecraft based on the lessons learned in this study.

APPENDIX EM ALGORITHM FOR MPPCACD MODEL

As the EM algorithm for the MPPCACD model is a simple extension of that for MPPCA [22] and its derivation is straightforward, we present only the main results necessary for implementation here. Note that we use the eigen-decomposition of a weighted covariance matrix for

estimating each local PPCA model at each M-step, while there is an alternative that estimates each PPCA model also by using EM iterations. Please refer to the original paper on MPPCA by Tipping and Bishop [22] for details.

A. Initialization

The parameter set for an MPPCACD model can be represented as $\boldsymbol{\Theta} = \{\pi_k, \boldsymbol{\mu}_k, \boldsymbol{W}_k, \sigma_k^2 \boldsymbol{\theta}_k\}\ (k = 1, 2, \ldots, K)$. Here, $\boldsymbol{\theta}_k$ represents the parameters of the D_s-dimensional categorical distribution for the kth cluster (local model) and is further described as $\boldsymbol{\theta}_k = \{\boldsymbol{\theta}_k^{(1)}, \ldots, \boldsymbol{\theta}_k^{(D)}\}$, and $\boldsymbol{\theta}_k^{(j)} = \{\theta_{k,1}^{(j)}, \ldots, \theta_{k,M_j}^{(j)}\}$.

As the log likelihood function for a mixture distribution model is nonconvex in general, we have to give a good initial estimate for the parameter to obtain a near-optimal solution. This is the case with other mixture models, such as the well-known Gaussian mixture. In fact, if we set the initial parameter values randomly, it very frequently becomes stuck in local optima that are inferior to the global ones. We, therefore, used a heuristic procedure as below to obtain reasonable initial parameter estimates. First, we apply k-means clustering only to the continuous data $\boldsymbol{y}_{1:T}$, based on Ding's method [28], and assign each sample to one of the clusters temporarily. Then, we perform the M-step below only once to obtain the initial parameter estimates.

B. E-step: Computing Responsibilities

In the E-step, we compute $P(s_t = k|\boldsymbol{y}_t, \boldsymbol{z}_t, \boldsymbol{\Theta})$, which is the probability of each data sample $\{\boldsymbol{y}_t, \boldsymbol{z}_t\}\ (t = 1, 2, \ldots, T)$ belonging to each cluster $k = 1, 2, \ldots, K$ or the responsibility $r_{t,k}$ when the current estimate of parameters $\hat{\boldsymbol{\Theta}}$ is given

$$\begin{aligned} r_{t,k} &\equiv P\left(s_t = k|\boldsymbol{y}_t, \boldsymbol{z}_t, \hat{\boldsymbol{\Theta}}\right) \propto p\left(\boldsymbol{y}_t, \boldsymbol{z}_t, s_t = k|\hat{\boldsymbol{\Theta}}\right) \\ &\propto P\left(s_t = k|\hat{\boldsymbol{\Theta}}\right) p\left(\boldsymbol{y}_t|s_t = k, \hat{\boldsymbol{\Theta}}\right) P\left(\boldsymbol{z}_t|s_t = k, \hat{\boldsymbol{\Theta}}\right) \\ &\propto \pi_k \cdot \mathcal{N}\left(\boldsymbol{y}_t|\hat{\boldsymbol{\mu}}_k, \hat{\boldsymbol{C}}_k\right) \cdot \prod_{j=1}^{D_s} \mathrm{Cat}\left(z_t^{(j)}|\hat{\boldsymbol{\theta}}_k^{(j)}\right) \end{aligned} \tag{20}$$

where

$$\hat{\boldsymbol{C}}_k = \hat{\boldsymbol{W}}_k \hat{\boldsymbol{W}}_k^{\top} + \hat{\sigma}_k^2 \boldsymbol{I}. \tag{21}$$

In practice, we compute the right-hand side of (20) for $k = 1, \ldots, K$, and then, normalize them so that they sum to 1, that is, $\sum_{k=1}^{K} r_{t,k} = 1$.

C. M-Step: Updating Parameter Estimates

First, the estimates of the mixing parameters $\hat{\pi}_k$ and the mean vectors of continuous variables $\hat{\boldsymbol{\mu}}_k$ for clusters $(k = 1, 2, \ldots, K)$ are updated by the following equations:

$$\hat{\pi}_k = \frac{\sum_{t=1}^{T} r_{t,k}}{N} \tag{22}$$

$$\hat{\boldsymbol{\mu}}_k = \frac{\sum_{t=1}^{T} r_{t,k} \cdot \boldsymbol{y}_t}{\sum_{t=1}^{T} r_{t,k}}. \tag{23}$$

Next, we compute the weighted sample covariance matrices $\boldsymbol{S}_k\ (k = 1, 2, \ldots, K)$ using the responsibilities $\{r_{t,k}\}$

$$\boldsymbol{S}_k = \frac{1}{\sum_{t=1}^{T} r_{t,k}} \sum_{t=1}^{T} r_{t,k} \cdot (\boldsymbol{y}_t - \hat{\boldsymbol{\mu}}_k)(\boldsymbol{y}_t - \hat{\boldsymbol{\mu}}_k)^{\top}. \tag{24}$$

Let the eigenvalues of $\boldsymbol{S}_k$ be $\lambda_{k,1} \geq \lambda_{k,2} \geq \cdots \geq \lambda_{k,D_c}$ and the corresponding eigenvectors be $\boldsymbol{u}_{k,1}, \boldsymbol{u}_{k,2}, \ldots, \boldsymbol{u}_{k,D_c}$. Then, the estimate of σ_k^2 is updated as

$$\hat{\sigma}_k^2 = \frac{1}{D_c - L} \sum_{j=L+1}^{D_c} \lambda_{k,j}. \tag{25}$$

The estimate of $\boldsymbol{W}_k$ is also updated by

$$\hat{\boldsymbol{W}}_k = \boldsymbol{U}_{k,L}(\boldsymbol{\Lambda}_{k,L} - \hat{\sigma}_k^2 \boldsymbol{I})^{1/2} \tag{26}$$

where $\boldsymbol{U}_{k,L}$ and $\boldsymbol{\Lambda}_{k,L}$ are defined as

$$\boldsymbol{U}_{k,L} \equiv [\boldsymbol{u}_{k,1}, \boldsymbol{u}_{k,2}, \ldots, \boldsymbol{u}_{k,L}] \tag{27}$$

$$\boldsymbol{\Lambda}_{k,L} \equiv \mathrm{diag}(\lambda_{k,1}, \lambda_{k,2}, \ldots, \lambda_{k,L}). \tag{28}$$

Finally, the estimates of parameters for categorical distributions $\theta_{k,m}^{(j)}\ (k = 1, 2, \ldots, K,\ j = 1, 2, \ldots, D_s,\ m = 1, 2, \ldots, M_j)$ are updated as

$$\hat{\theta}_{k,m}^{(j)} = \frac{\sum_{t=1}^{N} r_{t,k} \cdot \mathbb{I}(z_i^{(j)} = m)}{\sum_{t=1}^{T} r_{t,k}} \tag{29}$$

where $\mathbb{I}(\cdot)$ is the indicator function and defined as

$$\mathbb{I}(x) = \begin{cases} 1 & \text{if } x \text{ is True,} \\ 0 & \text{if } x \text{ is False.} \end{cases}$$

REFERENCES

[1] C. Chang, W. Nallo, R. Rastogi, D. Beugless, F. Mickey, and A. Shoop
Satellite diagnostic system: An expert system for intelsat satellite operations
In *Proc. 4th Eur. Aerosp. Conf.*, 1992, pp. 321–327.

[2] D. P. Tallo, J. Durkin, and E. J. Petrik
Intelligent fault isolation and diagnosis for communication satellite systems
Telematics Informat., vol. 9, nos. 3/4, pp. 173–190, 1992.

[3] M. Rolincik, M. Lauriente, H. C. Koons, and D. Gorney
An expert system for diagnosing environmentally induced spacecraft anomalies
In *Proc. 5th Annu. Space Oper. Appl. Res. Symp.*, 1992, pp. 36–44.

[4] F. Ciceri and L. Marradi
Event diagnosis and recovery in real-time on-board autonomous mission control
In *Proc. 1st Ada-Eur. Symp.*, 1994, pp. 288–301.

[5] N. Nishigori, M. Hashimoto, A. Choki, and M. Mizutani
Fully automatic and operator-less anomaly detecting ground support system for mars probe NOZOMI
In *Proc. 6th Int. Symp. Artif. Intell. Robot. Autom. Space*, 2001.

[6] B. C. Williams and P. P. Nayak
A model-based approach to reactive self-configuring systems
In *Proc. 13th Nat. Conf. Artif. Intell.*, 1996, pp. 971–978.

[7] P. I. 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. Autom. Space*, 2003.

[8] A. Fijany, F. Vatan, A. Barrett, M. James, and R. Mackey
An advanced model-based diagnosis engine
In *Proc. Int. Symp. Artif. Intell., Robot. Autom. Space*, 2003.

[9] A. Finzi, M. Lavagna, and G. Sangiovanni
Fuzzy inductive reasoning and possibilistic logic for space systems failure smart detection and identification
In *Proc. Int. Symp. Artif. Intell., Robot. Autom. Space*, 2003.

[10] F. Hutter and R. Dearden
Efficient on-line fault diagnosis for non-linear systems
In *Proc. 7th Int. Symp. Artif. Intell. Robot. Space.*, 2003.

[11] Y. Kawahara, T. Yairi, and K. Machida
Diagnosis method for spacecraft using dynamic bayesian networks
In *Proc. 8th Int. Symp. Artif. Intell., Robot. Autom. Space*, 2005.

[12] D. DeCoste
Automated learning and monitoring of limit functions
In *Proc. Int. Symp. Artif. Intell., Robot. Autom. Space.*, 1997, pp. 287–292.

[13] T. Yairi, M. Nakatsugawa, K. Hori, S. Nakasuka, and K. Machida
Adaptive limit checking for spacecraft telemetry datausing regression tree learning
In *Proc. Int. Conf. Syst., Man Cybern.*, 2004, pp. 5130–5135.

[14] R. Fujimaki, T. Yairi, and K. Machida
An anomaly detection method for spacecraft using relevance vector learning
In *Proc. 9th Pacific-Asia Conf. Knowl. Discovery Data Mining.*, 2005, pp. 785–790.

[15] D. DeCoste and M. Levine
Automated event detection in space instruments: A case studyusing ipex-2 data and support vector machines
Proc. SPIE., 2000.

[16] M. Schwabacher, N. Oza, and B. Matthews
Unsupervised anomaly detection for liquid-fueled rocket propulsion health monitoring
J. Aerosp. Comput., Inf. Commun., vol. 6, no. 7, pp. 464–482, 2009.

[17] D. L. Iverson *et al.*
General purpose data-driven monitoring for space operations
J. Aerosp. Comput., Inf. Commun., vol. 9, no. 2, pp. 26–44, 2012.

[18] R. Fujimaki, T. Yairi, and K. Machida
An approach to spacecraft anomaly detection problem using kernel feature space
In *Proc. 11th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2005, pp. 401–410.

[19] M. Inui, Y. Kawahara, K. Goto, T. Yairi, and K. Machida
Adaptive limit checking for spacecraft telemetry data using kernel principal component analysis
Trans. Japan Soc. Aeronaut. Space Sci., Space Technol. Japan, vol. 7, no. ists26, pp. Pf-11–Pf-16, 2009.

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

[21] A. Zimek, E. Schubert, and H.-P. Kriegel
A survey on unsupervised outlier detection in high-dimensional numerical data
Statist. Anal. Data Mining, vol. 5, no. 5, pp. 363–387, 2012.

[22] M. Tipping and C. Bishop
Mixtures of probabilistic principal component analyzers
Neural Comput., vol. 11, pp. 443–482, 1999.

[23] Y. Nakamura *et al.*
Small demonstration satellite-4 (sds-4): Development, flight results, and lessons learned in jaxas microsatellite project
In *Proc. 27th AIAA/USU Conf., Small Satell. Constellations*, 2013, Paper SSC13–X–1.

[24] B. Schölkopf, J. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson
Estimating the support of a high-dimensional distribution
Neural Comput., vol. 13, pp. 1443–1471, 2001.

[25] D. M. Tax and R. P. Duin
Support vector data description
Mach. Learn., vol. 54, no. 1, pp. 45–66, 2004. [Online]. Available: http://dx.doi.org/10.1023/B:MACH.0000008084.60811.49

[26] C.-C. Chang and C.-J. Lin
LIBSVM: A library for support vector machines
ACM Trans. Intell. Syst. Technol., vol. 2, pp. 27–1–27-27, 2011. [Online]. Available: http://www.csie.ntu.edu.tw/cjlin/libsvm

[27] K. P. Murphy
Dynamic bayesian networks: Representation, inference and learning
Ph.D. dissertation, Univ. California, Berkeley, CA, USA, 2002.

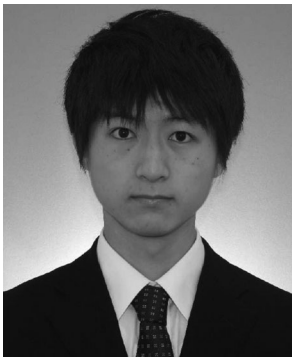
[28] C. H. Ding and X. He
Principal component analysis and effective k-means clustering
In *Proc. 4th SIAM Int. Conf. Data Mining*, 2004, pp. 497–501.

Takehisa Yairi (M'15) received the B.Eng., M.Sc., and Ph.D. degrees in aerospace engineering from the University of Tokyo, Tokyo, Japan, in 1994, 1996, and 1999, respectively.

He is currently a Full-Time Associate Professor in the Department of Aeronautics and Astronautics, School of Engineering, the University of Tokyo. His research interests include anomaly detection, health monitoring, fault diagnosis, learning dynamical systems, nonlinear dimensionality reduction, as well as application of machine learning and probabilistic inference to aerospace systems.

Naoya Takeishi (S'14) received the B.Eng. and M.Sc. degrees in aerospace engineering from the University of Tokyo, Tokyo, Japan, in 2013 and 2015, respectively. He is currently working toward the Ph.D. degree in the Department of Aeronautics and Astronautics, School of Engineering, the University of Tokyo.

His research interests include analyses of nonlinear dynamical systems, machine learning and probabilistic inference on time-series data, anomaly detection for engineering systems and information extraction from unstructured documents, as well as simultaneous localization and mapping for asteroid exploration.

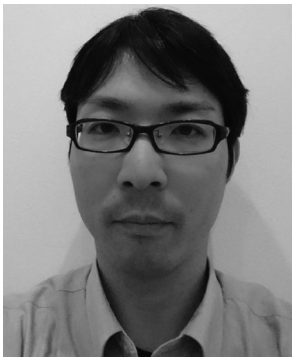


Tetsuo Oda received the B.Eng. and M.Eng. degrees in aerospace engineering from the University of Tokyo, Tokyo, Japan, in 2012 and 2014, respectively.

His research interests include anomaly detection for aerospace systems using machine learning.

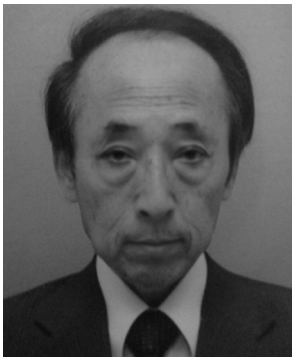
Yuta Nakajima received the Master of Science degree in engineering from Keio University, Tokyo, Japan, in 2009.

He is a Systems Engineer from Tokyo, currently at Systems Technology Unit, Research and Development Directorate in Japan Aerospace Exploration Agency (JAXA), Tsukuba, Japan. He joined the JAXA in 2009 and worked in the small demonstration satellite program as an Attitude and Orbit Control System Engineer. He has studied in the Graduate School of System Design and Management, Keio University, in 2012, and the Inernational Space University in 2013. His research interests include systems architecture, systems engineering, model-based systems engineering, concurrent engineering, design-thinking, decision making, and guidance, navigation, and control of spacecraft.



Naoki Nishimura received the B.Eng. and M.Eng. degrees in electronic engineering from the University of Fukui, Fukui, Japan in 2007 and 2009, respectively.

From 2013 to 2016, he was a Engineer and Researcher with the Japan Aerospace Exploration Agency. He designed and developed the onboard software for the small satellite, SDS-4, launched in 2012, and, was the Operator of that satellite. His research interests include load reduction of small satellite operation using ergonomics and machine learning and concept study of small satellites.



Noboru Takata received the B.S., M.Eng., and Ph.D. degrees in telecommunications engineering from the Tokyo Institute of Technology, Tokyo, Japan, in 1973, 1975, and 1978, respectively.

He had worked at Japan Aerospace Exploration Agency as an Engineer for 37 years, studying and developing telecommunications equipment such as transponders, antennas, and decoders. His research interests include reliability engineering, fault tolerance, anomaly detection, and fault diagnosis for spacecraft systems.