



Analysis of operational and mechanical anomalies in scheduled commercial flights using a logarithmic multivariate Gaussian model

Guoyi Li*, Hyunseong Lee, Ashwin Rai, Aditi Chattopadhyay

School for Engineering of Matter, Transport and Energy, Arizona State University, Tempe, AZ 85287, USA



ARTICLE INFO

Keywords:

Flight safety
Anomaly detection
Scheduled commercial flight
Multivariate Gaussian
Unsupervised learning
Performance evaluation

ABSTRACT

This paper presents a machine learning approach to evaluate the performance of aircrafts using on-board sensor information on commercially scheduled flights with the aim to further improve system health monitoring strategies in air transportation. Logarithmic multivariate Gaussian models are trained to evaluate the performance of aircrafts at different flight phases (takeoff, ascent, cruise, etc.) separately. By including a forward synchronization, feature selection, and mini-batch training process, this model overcomes challenges introduced by the large size and high dimensionality of flight datasets. This framework also addresses the re-sampling issue in existing literature causing difficulties in handling time-series signals with different lengths. For demonstration and validation, the developed model is applied to analyze performance anomalies associated with the mechanical system and pilot operation in a historical flight dataset. Compared with existing literature focusing on similar datasets, this evaluation methodology shows promise in detecting performance anomalies especially at approach and takeoff phases. Therefore, the developed model is expected to be an effective addition to the current anomaly analysis and monitoring technologies for scheduled commercial flights. Applications include assisting transportation management systems by handling large amounts of historical flight datasets to analyze mechanical and operational anomalies, which may potentially improve future aeronautical system design and pilot training.

1. Introduction

Scheduled commercial flights carried 4.1 billion travelers in 2017 (ICAO, 2018) indicating that aviation safety is, and will always be, a significant concern for society (Causse et al., 2013; Fleischer et al., 2015). Unfortunately, comparing with 2016, the ICAO also reported that the accident rate of scheduled commercial flights increased from 2.1 to 2.4 in 2017, while the total number of accidents increased from 75 to 88. Hence, there is an emerging need to develop automated monitoring systems to provide reliable assessment of transportation safety, for both current and next generation scheduled commercial flights (Kuhn, 2018; Sun et al., 2019; Liu and Goebel, 2018). Many studies have been conducted on multiple aspects of air traffic management to improve the air traffic system, including air traffic simulation (Menon et al., 2018), aircraft monitoring (Lee et al., 2019), aircraft modeling (Yu et al., 2018; Zhang et al., 2018; Yu et al., 2019), airport runway simulation (Sun et al., 2019b), information fusion (Wang et al., 2018), signal communication (Wang and Ying, 2018), and risk and uncertainty management (DeCarlo and Bichon, 2018; Zhang and Mahadevan,

* Corresponding author at: Arizona State University, Tempe, AZ 85287-6106, USA.

E-mail addresses: guoyili@asu.edu (G. Li), hlee269@asu.edu (H. Lee), aditi@asu.edu (A. Chattopadhyay).

2019).

Among various emerging research topics that support the vision of increased aeronautical safety, this paper focuses on the development of an automated anomaly analysis framework for single aircraft monitoring. The major motivation is that current aircraft monitoring systems rely on pre-defined safety thresholds and binary exceedance criterions, which may not be sufficient and accurate in reflecting the current health status of the aircraft or to identify operational and mechanical issues (Zhang et al., 2015). This drawback is mainly because inflight malfunction of aircraft hardware systems or inappropriate pilot operations, especially at the very early stage, may not lead to significantly lowered flight performances and unexpected control issues (McNeely et al., 2016). Individually, small deviations may not set off threshold alarms, however non-linear coupling between such deviations may lead to significant performance issues leading to a false sense of security. Hence, to maintain high standards of transportation safety, a robust monitoring system with sufficient sensitivity is highly desired. Such a system should be capable of not only detecting state malfunctions but also providing suggestions for maintenance to reduce unnecessary aircraft downtime (Dalton et al., 2001), which is essential for maintaining the routine of air transportation (Rebollo and Balakrishnam, 2014). However, the development of such a safety monitoring framework is challenging due to the complexity of aircrafts including both hardware and software systems, as well as the large wsize of the dataset recorded from on-board sensors.

1.1. State-of-the-Art in anomaly detection

In order to address the challenges in system health monitoring, many studies have been conducted to develop fault detection methodologies that ensure system reliability in the presence of system anomalies (Oonk et al., 2012), which are generally categorized into two classes – model-based and data-driven approaches (Zheng and Van Zuylen, 2013; Papathanasopoulou and Antoniou, 2015). Model-based fault detection techniques are the most widely used approaches, which describe each component in the monitored system using the terms ‘normal’ or ‘faulty’ based on the observed consequences (Mikaelian et al., 2005). These comprise three classes – quantitative, qualitative, and process history based methods (Venkatasubramanian et al., 2003; Venkatasubramanian et al., 2003; Venkatasubramanian et al., 2003). The quantitative approach explicitly defines the input-output relation through state-space models using Kalman filters and other estimation algorithms based on the observations of consequence. For example, Foo et al. (2013) developed extended Kalman filters for sensor fault detection in synchronous motor driver applications by estimating the phase currents and rotor speed of the motor simultaneously. Zhao et al. (2017) introduced observer-based dynamic algorithms for fault tolerant control of a nonlinear system with potential actuator failures based on adaptive dynamic programming. On the other hand, the qualitative approach utilizes a set of if-then-else rules with their corresponding inferences that find the consequence based on given knowledge, such as digraphs, fault trees and qualitative physics (Bartlett et al., 2009). For instance, Sihombing and Torbol (2018) proposed a parallel fault tree algorithm with a graphical processor unit computing scheme, providing increased reliability in identifying system failures. In addition to model-based algorithms, data-driven approaches have recently been attracting increased attention by the fault detection society due to the rapid development of computation power and artificial intelligence (Ge, 2017); such approaches typically leverage a sufficient amount of historical information with multiple explanatory features to diagnose and predict system performance (Cerrada et al., 2018). As critical aspects of data-driven methods, feature extraction and information fusion techniques have been extensively developed to address the computation efficiency issue associated with the dataset with high dimensionality (Basir and Yuan, 2007); some examples are support vector machine (SVM) and deep neural networks (DNN) (Banerjee and Das, 2012; Vanini et al., 2014). For example, Samanta and Al-Balushi (2003) developed a DNN to address the problem of fault diagnosis of rotating bearing systems using real-time time-domain signals of vibration. Zhang et al. (2006) developed a system health management framework for a rapid state estimation of large-scale structure by employing incremental SVM regression. In addition, the outlier detection algorithm represents another class of fault detection method that shows promising effectiveness and efficiency. Schwabacher et al. (2009) summarized widely used unsupervised anomaly detection algorithms and demonstrated their performance under different anomaly conditions. For example, the distance-based outlier detection algorithm was employed to detect anomalous conditions based on the investigation of the extreme values with respect to the normal condition using clustering methods such as k-nearest neighbor (Bay and Schwabacher, 2003; Knorr et al., 2000). The major advantages of such algorithms are that no explicit distribution is required to establish abnormal conditions, which significantly reduces computational cost, and no strict limit is imposed on feature dimensionality, which is well-suited for transportation monitoring in the aeronautical domain.

1.2. Anomaly detection for aircraft

The success of data-driven approaches for fault detection has motivated the development of a system health management framework for aircraft safety. For safety assurance, the current scheduled commercial flights generally implement the Flight Operations Quality Assurance (FOQA) program designed by the Federal Aviation Administration (FAA) (FAA, 2004). In this program, exceedance analysis was used to identify the fault conditions based on the state of each flight parameter, such as engine fan speed, control surface position, engine power plant performance, etc.; the monitored parameter exceeding the pre-defined threshold will result in a safety alarm under a specified condition (Tsuruta, 2008). For the investigation of discrete flight parameters, longest common subsequence and sequence clustering techniques were developed (Budalakoti et al., 2006; Budalakoti et al., 2009); such discrete flight parameters, e.g. flap position, shows fast estimation due to its beneficial correlation with aircraft dynamics and pilot control. For the further improvement of the detection accuracy, kernel functions are typically used to transform the feature space for obtaining a better representation of the monitored system. Multiple kernel anomaly detection (MKAD) algorithms (Das et al., 2010), a combination of multiple kernel functions used to construct the feature space, are shown to be suitable for analyzing aircraft performance at

the approach phase. Li et al. (2015) suggested clustering-based anomaly detection method, known as ClusterAD-Flight, to automatically detect anomalous conditions based on routine airline flights at approach and takeoff phases. However, both MKAD and ClusterAD-Flight require the full flights information in order to identify the anomalies, which limits their applications in monitoring real-time data stream. In order to address this issue, Li et al. (2016) has reported the development of a Gaussian mixture model, named as ClusterAD-DataSample for flight operation and safety monitoring, which shows the capabilities in detecting instantaneous anomalies at approach and takeoff phases. Motivated by the effectiveness and efficiency of ClusterAD-DataSample, the authors' preliminary work on multivariate Gaussian model has shown to be capable of detecting performance anomalies in aircrafts during the cruise phase with a single and unimodal distribution (Li et al., 2018). However, the computational efficiency of the multivariate Gaussian model, as predicted by Ng (2015) and Li et al. (2016), is challenged by the large size and dimension of the flight dataset (detailed in Section 3.1), which restricts this investigation to a small dataset (458 flights) with an incomplete feature space (20 features).

1.3. Proposed approach

In order to address the aforementioned challenges, this paper presents the development of a logarithmic multivariate Gaussian (LMG) model for real-time anomaly analysis of in-flight scheduled commercial aircraft, with the potential of indicating deviatoric events, which may include unsafe incidents. Under the hypothesis that the on-board signals follow a unimodal multivariate Gaussian distribution at each flight phase, the LMG model evaluates the aircraft performance based on the distributions that are individually trained at each flight phase. The *flight phase* is defined as the operation conditions of an aircraft by the National Transportation Safety Board (NTSB) (NTSB, 2013), which generally comprises of standing, pushback/towing, taxi, takeoff, initial climb, *en route*, etc. Unlike the existing Gaussian models, the LMG is formulated in logarithmic space to handle a large Mahalanobis distance induced by the high dimensionality of the aircraft dataset (Xiang et al., 2008), while the mini-batch learning process eliminates the need for excessive computational memory (Mairal et al., 2009); the detected results will be discussed and interpreted with help of subject matter experts familiar with aviation practices. Such an approach addresses the following two issues in state-of-the-art literature: (i) the re-sampling of information from different flights into the same length at a fixed interval, which challenges the capabilities of the aviation monitoring system in detecting anomalies in flight phases that vary considerably in duration such as the cruise phase (Li et al., 2015; Li et al., 2016); (ii) the limitation of computational memory in handling large flight dataset, especially relevant with the yearly increased number of commercial scheduled flights (Li et al., 2018). Hence, the LMG model presented in this paper is expected to be an effective addition to existing technologies in the assessment of aircraft performance using signals from on-board sensors, while assisting air transportation experts and monitoring systems to identify anomalous aircraft behaviors that may be safety issues.

It should be noted that the *anomaly* in this manuscript is defined as the aircraft behavior that shows a significant deviation from behavior observed in most historical datasets of scheduled commercial flights. Consequently, an anomaly does not necessarily indicate an unsafe incident; this is because of the limitations of the training dataset, which will be introduced later in this manuscript (Section 3.1). The remainder of this paper is organized as follows: the description of LMG model and corresponding pseudo code with a mini-batch learning process are presented in Section 2. This is followed by a description of the investigated dataset as well as the justification and validation of hypotheses made in the LMG model in Section 3. The learning process with respect to the number of mini-batches, and the indication of aircraft performance in different flight phases are demonstrated in Section 4. The detected anomalous events in each flight phase are presented separately in Section 5. Key achievement, limitations and future directions are summarized in the concluding remarks section.

2. LMG model for anomaly detection

The current model utilized in this work is formulated to detect anomalous behaviors in a monitored aircraft (i.e., outliers) by analyzing on-board flight sensor data. The model is formulated to accommodate the following characteristics of the dataset: (i) features recorded by sensors with multiple sampling frequencies, (ii) large dimensionality of feature space and (iii) large size of flight data exceeding the memory limitation of common computers. Forward synchronization is implemented to ensure features to be at the pre-defined sampling frequency, as described in Section 2.1. It is followed by a mini-batch learning algorithm that is used to overcome the computational issues associated with large datasets. The Logarithmic Gaussian formulation, introduced in Section 2.3, is developed to effectively evaluate the high dimensional information; the justification and validation of key hypotheses made in this computation scheme will be discussed in Section 3.

2.1. Forward synchronization

For each mini-batch, the raw signals are defined into $x_{raw}^{(l^j)}$, where $j = 1, 2, \dots, n$, and n is the number of features, l^j is the l^{th} time-step of j^{th} feature under sampling frequency of f , since l will be different across features. An averaging forward synchronization, which shows promising capability, is implemented on the raw signals into the highest sampling frequency (defined as f_{max}) in the dataset; the synchronized feature space can be expressed as follows,

$$x_j^{(s^j)} = x_{raw}^{(l^j)}, \quad \text{if } s \leq R^j \quad (1)$$

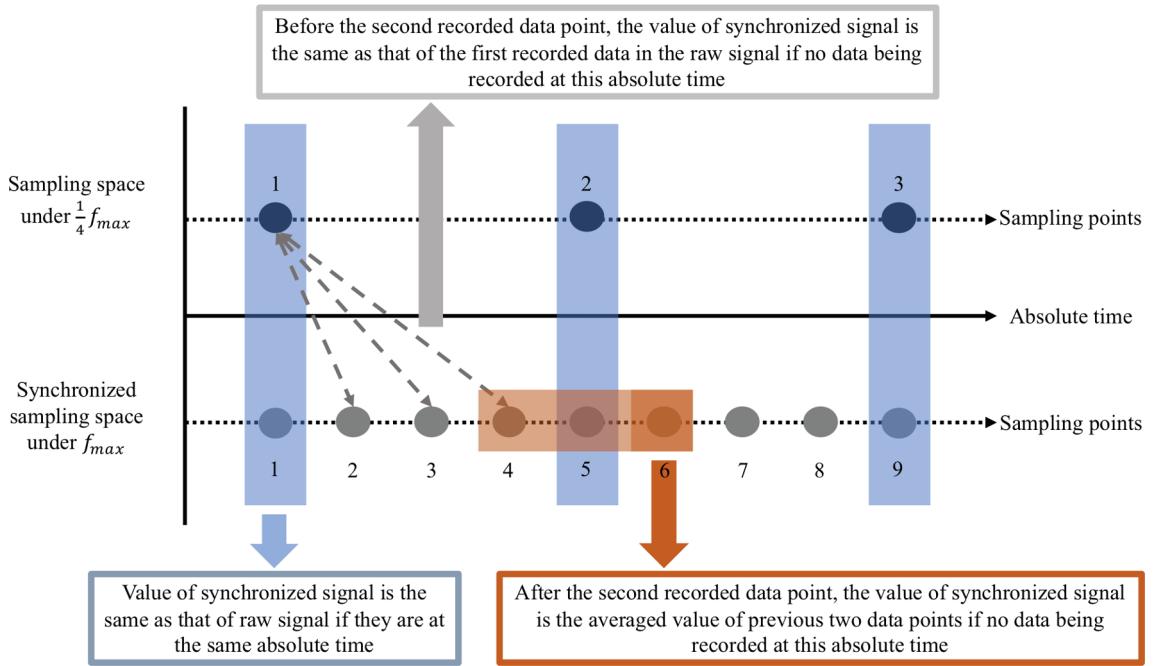


Fig. 1. A demonstration of the proposed forward synchronization scheme.

$$x_j^{(s^j)} = x_{\text{raw},j} \frac{((\frac{s-1}{R^j}+1)^j)}{R^j}, \quad \text{if } s > R^j \& \left(\frac{s-1}{R^j}\right) \text{ is an integer} \quad (2)$$

$$x_j^{(s^j)} = \frac{1}{2} (x_j^{((s-1)^j)} + x_j^{((s-2)^j)}), \quad \text{if } s > R^j \& \left(\frac{s-1}{R^j}\right) \text{ is not an integer} \quad (3)$$

where s is an integer indicating the sampling point of synchronized signal and $R = \frac{f_{\max}}{f}$. It is worth mentioning that the constant R^j denotes the ratio of length between signals under f_{\max} and f of the j^{th} feature, indicating $(R^j - 1)$ data points that need to be generated for each sampling point of x_{raw} under f . In this research, the R^j is assumed to be an integer. A demonstration of the proposed forward synchronization of sensor signals under sampling frequency of $f = \frac{1}{4}f_{\max}$ (i.e., $R = 4$) is shown in Fig. 1.

2.2. Mini-batch learning for multivariate Gaussian model

For a Gaussian model, it is generally assumed that each feature follows an identical Gaussian distribution so that the probability density function is expressed as $p(x_j; \mu_j, \sigma_j^2)$, where μ_j and σ_j are mean and standard deviation of feature j to be learned from the training examples. In order to evaluate correlations in the investigated feature space, which is difficult to be obtained analytically, a multivariate Gaussian model is implemented. The probability density function of a multivariate Gaussian model can be expressed as $p(x; \mu, \Sigma)$, where Σ denotes the covariance matrix. To simplify the notation after forward synchronization, the vector containing all sensor features of the m^{th} mini-batch under sampling frequency f_{\max} is defined as x_j^m , $j = 1, 2, \dots, n$, and n is the number of features. In addition, to maintain computational efficiency, this model is formulated in a vectorized computation scheme. In this case, the μ^m and Σ^m , which are the trained mean vector and covariance matrix of the m^{th} mini-batch (known as *local mean vector* and *local covariance matrix* later) can be expressed as follows.

$$\mu^m = \frac{1}{N_m} \sum_{i=1}^{N_m} (x^m)^i \quad (4)$$

and

$$\Sigma^m = \frac{1}{N_m} \sum_{i=1}^{N_m} ((x^m)^i - \mu^m)((x^m)^i - \mu^m)^T \quad (5)$$

where N_m is the number of samples in the m^{th} mini-batch. For clarification, x^m , μ^m and Σ^m have the dimension of $N_m \times n$, $1 \times n$ and $n \times n$, respectively. After training the m^{th} mini-batch, the local mean vectors and covariance matrix from 1^{st} to m^{th} mini-batch will be used to obtain the final trained mean vectors μ_g and covariance matrix Σ_g (known as *global mean vector* and *global covariance matrix* later and shown in Eqs. (6) and (7)).

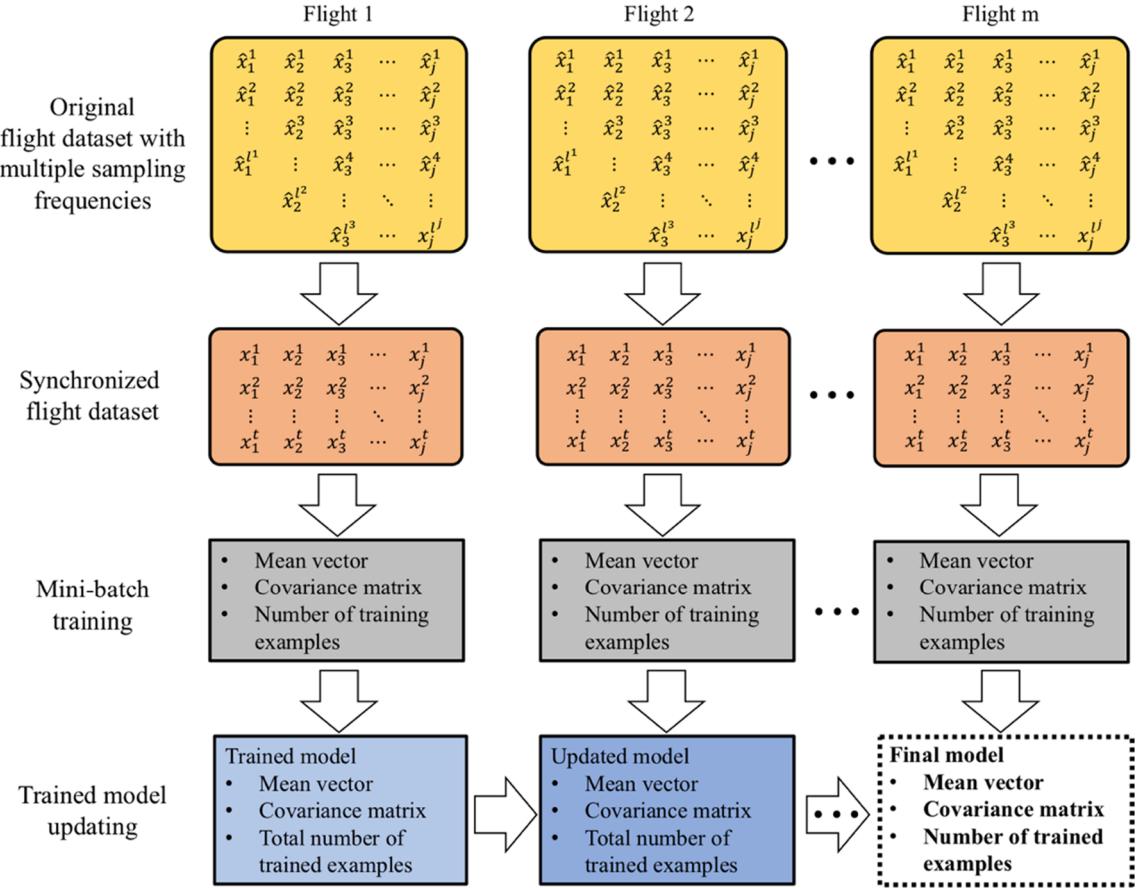


Fig. 2. Demonstration of mini-batch learning for global mean vector and global covariance matrix.

$$(\mu_g)^m = \frac{(\mu_g)^{m-1} \sum_{i=1}^{m-1} N_i + \mu^m N_m}{\sum_{i=1}^m N_i} \quad (6)$$

$$(\Sigma_g)^m = \frac{(\Sigma_g)^{m-1} \sum_{i=1}^{m-1} N_i + \Sigma^m N_m}{\sum_{i=1}^m N_i} + \frac{(\sum_{i=1}^{m-1} N_i)N_m}{(\sum_{i=1}^m N_i)^2} ((\mu_g)^{m-1} - \mu^m)^2 \quad (7)$$

Implementing $\sum_{i=1}^{m-1} N_i$ and $\sum_{i=1}^m N_i$, instead of simply multiplying the batch size and number of batches, the global mean vector and covariance matrix can be trained using multiple mini-batches with varying batch size, proven to be suitable in the subsequent sections. It should be noted that the general Gaussian model is a specific case of the multivariate Gaussian model, whose covariance matrix just contains non-zero values at the diagonal, i.e., σ_j^2 . A schematic is shown in Fig. 2 to demonstrate the mini-batch learning process.

2.3. Anomaly detection using LMG

After obtaining the global mean vectors μ_g and covariance matrix Σ_g with all the training mini-batches, the performance of each time instance will be evaluated with its associated signals using a probability density function. It fuses a multi-dimensional feature space into a single value of probability, indicating the similarity between the investigated point and the trained model (known as *performance index* later in the manuscript). For a multivariate Gaussian model, the original probability density function can be expressed as,

$$p(x; \mu_g, \Sigma_g) = \frac{1}{\sqrt{(2\pi)^M |\Sigma_g|}} \exp\left(-\frac{1}{2}(x - \mu_g)^T \Sigma_g^{-1} (x - \mu_g)\right) \quad (8)$$

where $M = \sum_{i=1}^m N_i$. The $(x - \mu)^T \Sigma^{-1} (x - \mu)$ refers to the Mahalanobis distance which is used as the distance metric between the investigated point and the trained distribution. However, there are two issues in directly applying Eq. (8) to evaluate the performance index, including (i) the covariance matrix is not full rank and (ii) high feature dimension results in a large Mahalanobis distance so that the performance index will be absolute zero due to numerical truncation at some time instances. To address the issue that the

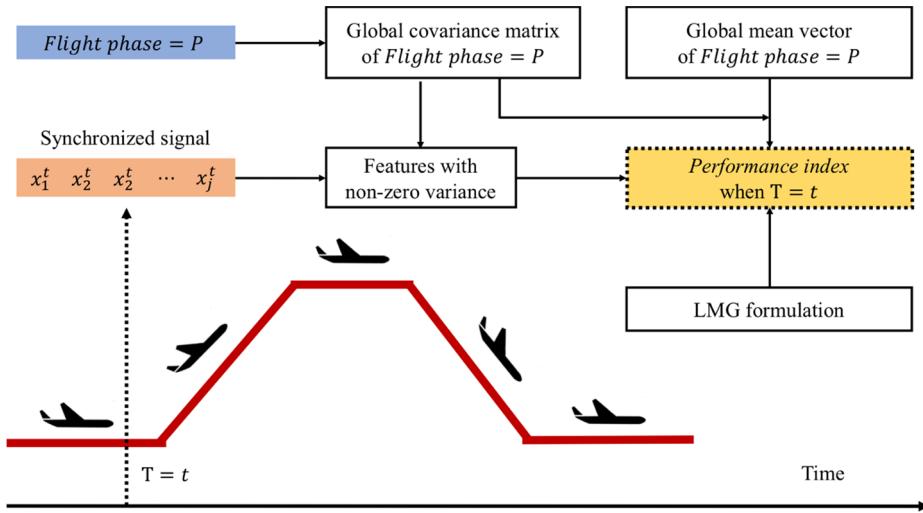


Fig. 3. A schematic showing anomaly detection in flight data using LMG model.

covariance matrix is not a full rank matrix, the diagonal of Σ_g is investigated, and the features with zero-variance will be excluded in the performance evaluation using the *zero-variance threshold* introduced later in [Section 3](#). Then, instead of using Eq. (8), an LMG model is formulated as follows:

$$\text{Logp}(x; \mu_g, \Sigma_g) = -\frac{M}{2} \log(2\pi) - \frac{1}{2} \log(|\Sigma_g|) - \frac{1}{2} (x - \mu_g)^T \Sigma_g^{-1} (x - \mu_g) \quad (9)$$

It is worth mentioning the difference between this LMG model and other approaches. To handle high dimensional information, an efficient way is to compute p and transform it (left-hand side of Eq. (8)) into logscale such as the aforementioned Gaussian mixture model ([Li et al., 2016](#)). However, if the Mahalanobis distance is too large, this results in a negligible exponential term in Eq. (8) (i.e., $\log(y) \approx 0$, if $y = \exp\left(\frac{1}{2}(x - \mu_g)^T \Sigma_g^{-1} (x - \mu_g)\right) \approx 0$). Considering this, the right-hand side of Eq. (8) is initially transformed to logscale, and the performance index is evaluated through Eq. (9), which ensures a non-zero performance index at each time instance. This formulation is originally motivated by the logscale likelihood function in the *maximum a posteriori* approach developed by [Williams and Rasmussen \(2006\)](#) for Gaussian process machine learning. Using Eq. (9), the performance of each flight at each time instance is evaluated based on μ_g and Σ_g learned from different flight phases. Therefore, the anomalies, which possess large distances (less similarity) from the trained model in the feature space, can be identified with a pre-defined anomaly ratio. As shown in [Fig. 3](#), a schematic demonstrating the process of anomaly detection. For clarity, a pseudo code of the complete framework introduced in [Section 2](#) is provided in [Table 1](#). The computation scheme used in this research is programmed in the MATLAB.

3. Dataset, hypotheses and assumption

3.1. Historical airline dataset

The anomaly analysis is performed using flight data recorder (FDR) information containing on-board aircraft sensor data from scheduled commercial flights, which is publicly available at the National Aeronautics and Space Administration (NASA) DASHlink network. The chosen data corresponds to 186 flight parameters, which can be categorized into two classes: sensor signals, and state indicators. Sensor signals refer to the flight parameters that are obtained by the on-board sensors and can represent the performance of the aircraft, such as body longitudinal acceleration, position of rudder, aileron position, elevator position, and engine exhaust gas temperature, etc. The state indicators generally present the flight parameters that are integers denoting an on-off state such as electric control system on/off; this class also include very limited information of the aircraft and flight schedule such as engine series indicator, flight date, flight time, etc.

In this research, 5376 flights have been investigated, and all sensor signals have been evaluated by the introduced LMG model, 96 features in total. The sampling frequencies of the sensor signals vary between 0.25 Hz and 16 Hz, while the flight duration among these flights vary between 1 h to 3.5 h, approximately. The flight phase, characterized by state indicators, is used to indicate different operational states of the aircraft when training the global mean vector and global covariance matrix. Specifically, a complete flight is categorized into flight phases from 1 to 7, indicating parking, taxiing, takeoff, ascent, cruise, descent, and landing, respectively. In addition, to be able to update the trained model easily with new flight data if any, the mini-batch is defined as one flight, which is the reason for formulating the mini-batch learning using a dynamic batch size approach. The dataset presents two limitations: first, no information about the aircraft type is included in the dataset (sanitized dataset), which prevents the cross-checking of the value of signals using existing aircraft documents such as engine manuals. Second, there is no *ground truth* available in this dataset which prevents the tuning of hyperparameters through cross-validation, and directly testing performance of the developed model using

Table 1

Pseudo code of LMG framework introduced in this section.

```

Input
Time-series flights dataset
Flight phase indicator
Sampling frequency
Anomaly ratio of each flight case
Anomaly duration threshold of each flight case
Algorithm
Step 1. Initialize parameters
    Global mean vector
    Global covariance matrix
    Global performance vector
    Global zero variance count vector
    Final performance vector
Step 2. Mini-batch learning & global model updating
    for each flight  $J$  in the dataset
        Synchronize each feature into the same sampling frequency
        for each flight phase  $K$ 
            Count samples
            Compute local mean vector
            Compute local covariance matrix
            Update global mean vector using [ local mean vector; # of local samples; accumulated # of samples across all investigated flight]
            Update global covariance matrix using [ local covariance matrix; # of local samples; accumulated # of samples across all investigated flight]
            Find zero variance terms in diagonal of local covariance matrix
            Update global zero count variance vector based on # of zero variance w.r.t. feature
        end for
    end for
Step 3. Feature selection
    for each flight phase  $K$ 
        Check if: global zero variance count vector [feature] < zero variance threshold
        Save global mean vector with features that pass the check
        Save global covariance matrix with features that pass the check
    end for
Step 4. Anomaly detection and evaluation
    for each flight  $J$  in the dataset
        Synchronize each feature into the same sampling frequency
        for each flight phase  $K$ 
            Local performance vector  $\leftarrow$  LMG formulation [global mean vector, global covariance matrix]
            Append local performance vector to global performance vector
            Sort updated global performance vector in ascending order
            Truncate global performance vector using anomaly ratio and total # of investigated data points
        end for
    end for
    for each flight phase  $K$ 
        for each flight in global performance vector
            if anomaly duration > anomaly duration threshold
                Append global performance vector to final performance vector
            end if
        end for
    end for
end for
Output
For each flight phase in final performance vector:
    Indices of flights with anomalies
    Time indices of anomalies of each flight

```

metrics such as F1 Score. Moreover, as mentioned by [Li et al. \(2016\)](#), there are no catastrophic failure events or even a fatal event included in the dataset, since such information is extremely rare in commercial airlines and are only made public in a limited form (such as the NTSB Aviation Accident Reports). Consequently, the detected anomalies in this research only refer to the mechanical responses of the aircraft or pilot operations that deviate from the common observations. Hence, this work aims at proving that the developed model is able to detect anomalous behaviors from this dataset, based on the consideration that the feature space of an ‘actual’ fault is anticipated to be farther from the trained distribution than the detected ‘anomalies’ in this dataset.

Before training the model, forward synchronization introduced in [Section 2.1](#) is applied to synchronize all features to the sampling frequencies of 16 Hz (i.e., $f_{max} = 16$) for each flight; the corresponding number of data points after synchronization is shown in [Table 2](#) for each flight phase. It should be noted that this synchronization strategy significantly increases the size of data (the sizes of raw data and synchronized data are 6.78 and 15.82 GB respectively). Nonetheless, the maximum size of matrix after synchronization of one batch (one flight) is approximate $192,000 \times 96$, which is affordable for a common desktop or laptop computer. [Fig. 4](#) demonstrates the comparison of the raw signals and the corresponding synchronized signals from one of the flights in the investigated dataset, including the left elevator position whose original sampling frequency is 1 Hz, the exhaust gas temperature whose original

Table 2

Number of data points after synchronization in each flight phase.

Flight Phase	Number of Data Points
Parking	4,515,896
Taxi	99,181,024
Takeoff	2,372,560
Ascent	82,688,608
Cruise	135,314,736
Descent	89,544,912
Landing	1,689,504
Total	415,306,240

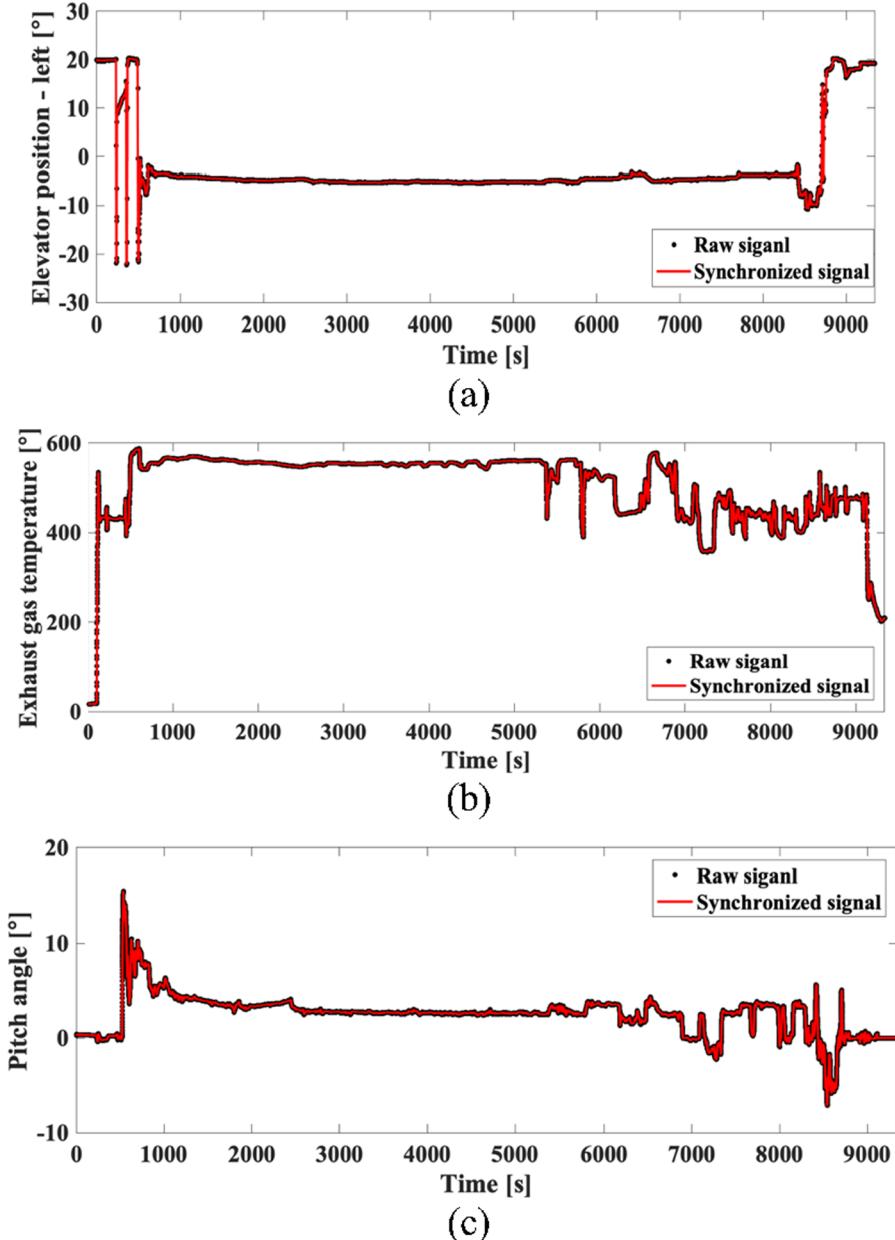


Fig. 4. Comparison of raw signals and synchronized signals: (a) left elevator position whose original sampling frequency is 1 Hz; (a) exhaust gas temperature whose original sampling frequency is 4 Hz; (a) pitch angle whose original sampling frequency is 8 Hz.

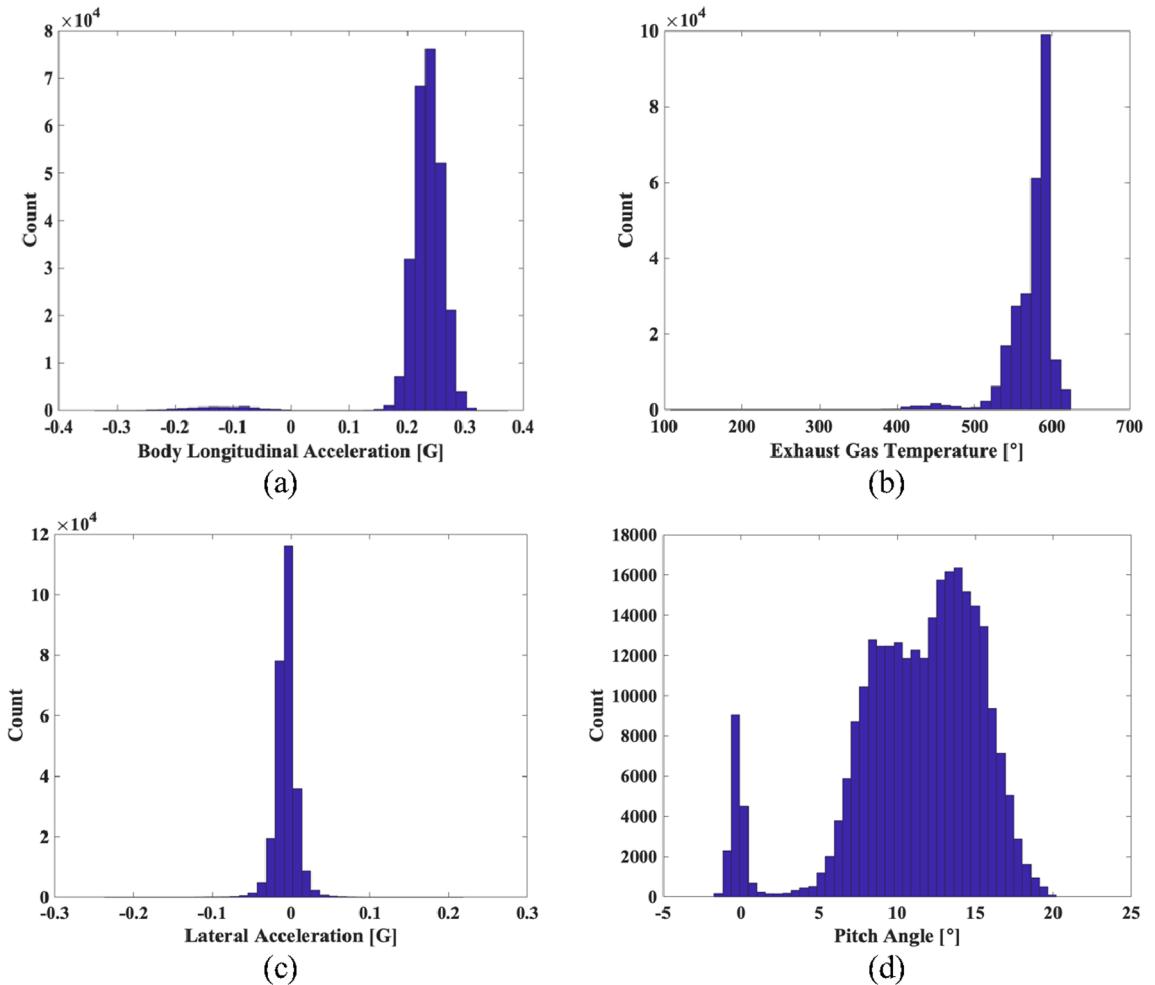


Fig. 5. Distributions of (a) body longitudinal acceleration, (b) exhaust gas temperature from sensor 1, (c) lateral acceleration and (d) pitch angle at takeoff phase.

sampling frequency is 4Hz, and the pitch angle whose original sampling frequency is 8 Hz. The synchronized signals show good agreement with the raw signals without introducing much variance. It is also observed that for a flight with the duration of approximate 2.6 h containing 149,504 sampling points in each feature, the averaged processing time of each feature is 1 ms using a 1.3 GHz Intel Core i5 processor. Therefore, this approach is computationally efficient and can potentially handle streaming data for real-time detection.

3.2. Hypotheses and assumptions

As mentioned in Section 2, the formulations of the LMG model are developed based on a few hypotheses. In this subsection, three key hypotheses will be discussed and verified including the feasibility of using the multivariate Gaussian model and zero-variance criteria for anomaly detection and feature selection, respectively. In addition, considering the repeatability of the presented research and the complexities associated with airline flight information, the assumptions made for an effective anomaly analysis will be described. The assumptions mainly contain three important thresholds used in the formulation, and the metrics of determining these thresholds will be discussed.

The first and the most important hypothesis in the research is that each feature recorded by the on-board sensor, at each flight phase, follows a unimodal Gaussian distribution so that it is feasible to use seven trained models (global mean vectors and covariance matrices) to evaluate performance of aircraft at the corresponding seven flight phases. Under this hypothesis, instead of exhibiting a sequential behavior, the performance of aircraft is governed by the flight phase, which eliminates the dependence of the developed model on time. However, although this hypothesis enables the model to handle flight data from different flight duration without the need for re-sampling, its feasibility must be verified. Therefore, for the purpose of demonstration, four features are chosen, and their distributions across 500 flights are investigated during takeoff (269,984 sampling points), ascent (8,777,465 sampling points) and cruise (13,701,194 sampling points) as shown in Figs. 5, 6 and 7. Although a small number of outliers can be found in some cases

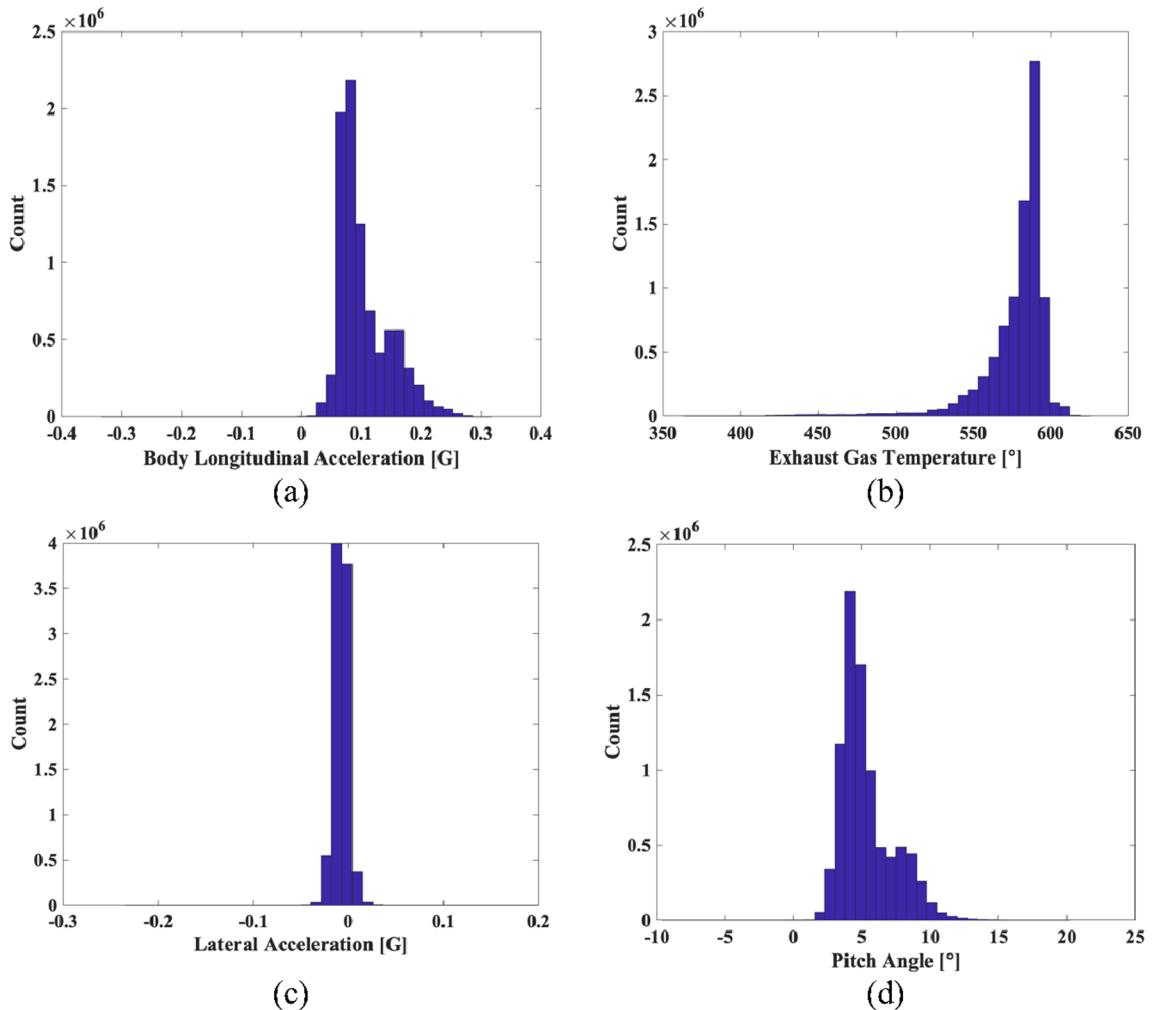


Fig. 6. Distributions of (a) body longitudinal acceleration, (b) exhaust gas temperature from sensor 1, (c) lateral acceleration and (d) pitch angle at ascent phase.

(e.g., pitch angle at takeoff phase), it can be seen that all the distributions are very close to unimodal Gaussian distributions, which proves the feasibility of this hypothesis; the outliers in pitch angle at takeoff phase is due to the operation mode is turned from taxiing to takeoff before the pilot increases the pitch angle, which is a normal aviation operational sequence.

The second hypothesis made in this model is that the features are correlated due to the coupling of complex aircraft systems, which indicates the need of the multivariate Gaussian model; otherwise, a univariate Gaussian model will be a better choice due to its computational efficiency. Therefore, it is necessary to verify the behaviors of feature space through the evaluation of correlations. However, it is observed that the variances of some features are zero at particular phases; based on the correlation definition of variables x and y shown in Eq. (10) the correlation cannot be computed with such features.

$$\text{Corr}(x, y) = \frac{1}{N-1} \sum_{i=1}^N \left(\frac{x_i - \mu_x}{\sigma_x} \right) \left(\frac{y_i - \mu_y}{\sigma_y} \right) \quad (10)$$

where μ_x , σ_x , μ_y and σ_y are the mean and standard deviation of feature x and y , respectively. To maintain consistence in the comparison, the correlations are set to be zero if the variance of one or both corresponding features are zeros, and the features (96 in total) are in the same order among these three flight phases. The correlation matrices are constructed with 500 demonstrative flights and represents the feature space at takeoff, ascent and cruise phases, exhibited in Fig. 8. It can be seen that most features are correlated, but these correlations vary at different flight phases; this observation can be a justification for using global mean vectors and covariance matrices, which are trained by information from particular flight phases, to evaluate the aircraft performance using the LMG model.

Meanwhile, the features with zero-variance prevent the model from obtaining a full rank covariance matrix as mentioned in Section 2. This motivates the third hypothesis that if the variances of a feature are zero in multiple flights, this feature will be

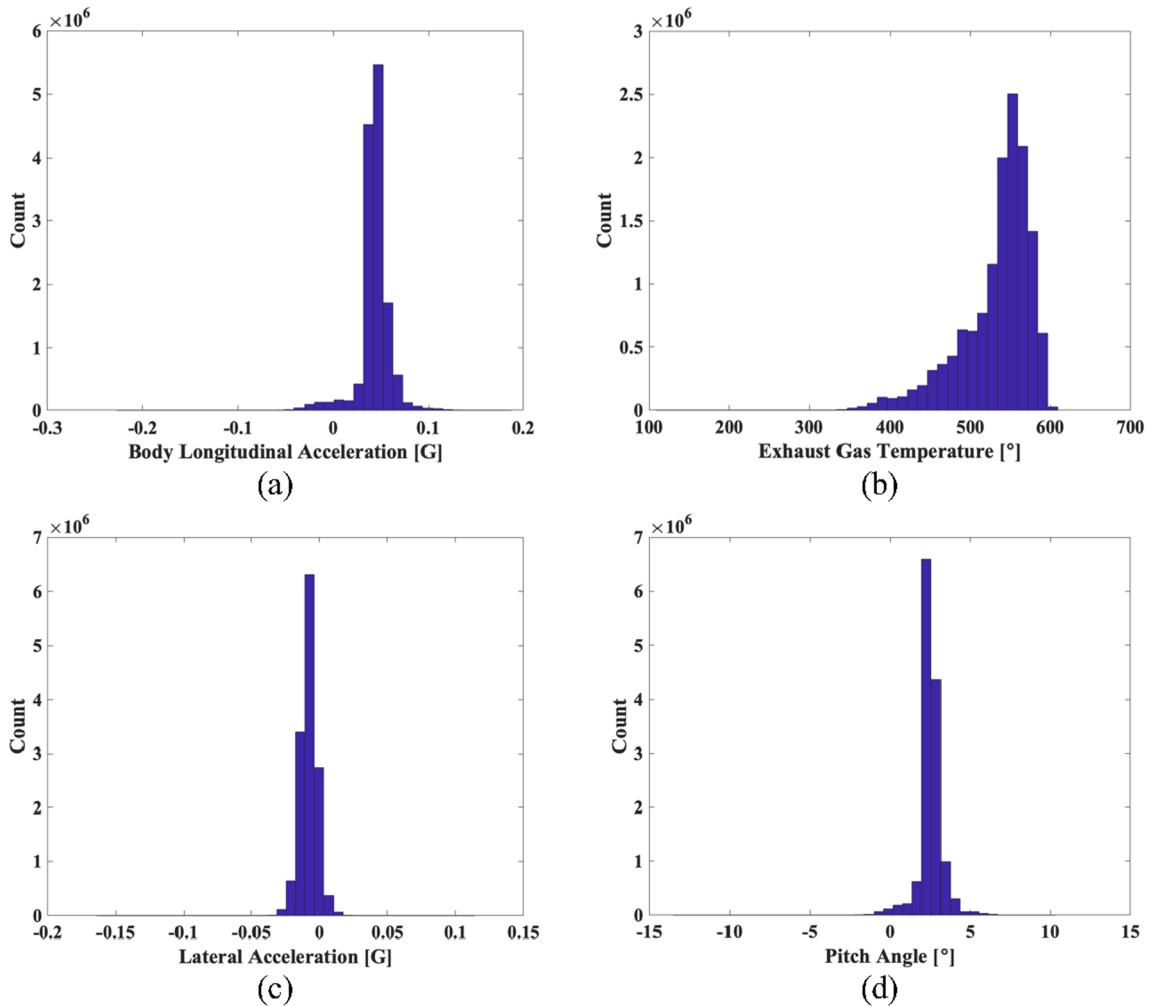


Fig. 7. Distributions of (a) body longitudinal acceleration, (b) exhaust gas temperature from sensor 1, (c) lateral acceleration and (d) pitch angle at cruise phase.

disregarded as containing null information in the context of evaluating aircraft performance using the LMG model for the current flight phase. As an example, variances of air brakes are zero in most of the flights when the aircraft is taking off, which is consistent with normal aircraft operation. However, it should be noted that the term *null* does not indicate that the corresponding signals are not needed at all for the aircraft; instead, such feature(s) can be used to train a binary classifier or a univariate distribution model and integrated with the proposed anomaly detection model as additional modules.

Due to the unlabeled FDR dataset, there are four assumptions made in this model in addition to the mentioned hypotheses for the current study. First, the zero-variance threshold is assumed to be 50, indicating the feature will be eliminated in the LMG model if its variances are zeros in more than 50 flights at a particular phase. The sensitivity of this value has been tested, and the result shows the number of eliminated features is insensitive to this threshold. Although, other dimension reduction methods, such as principal component analysis (PCA) and autoencoder (Burges, 2010), exist and their recovery accuracy could be evaluated, it would be very difficult to evaluate the dimension of orthogonal space in PCA or the number of hidden nodes in autoencoders on the anomaly detection accuracy without sufficiently labeled negative examples. It should be noted that the variances of features including fuel flows and fuel quantities from the four fuel tanks are zero in small percentage of all analyzed flights, which indicates the aircraft does not use all the fuel tanks all the time. As a result, these eight variables are replaced by two variables including averaged fuel flow and averaged fuel quantity. Third, Due to the lack of validation set, the decision boundary of the performance distribution could not be statistically determined. Therefore, considering the size of investigated flight dataset, the anomalous flights are flagged using an assumed anomaly ratio of 0.1%. Additionally, due to the dataset size and the high sampling frequency, a time duration threshold is also expected to reduce the impact of false positive. The presented research will only analyze the detected anomalies with durations of more than 30 s at flight phases excluding takeoff and landing. For takeoff and landing phases, the time duration thresholds are set to be 5 s, which supports the detection of rapidly varying anomalous behavior more characteristic of these shorter flight phases. It should be noted that the *duration* is not necessarily a continuous event but measured by the total number of points flagged as anomaly

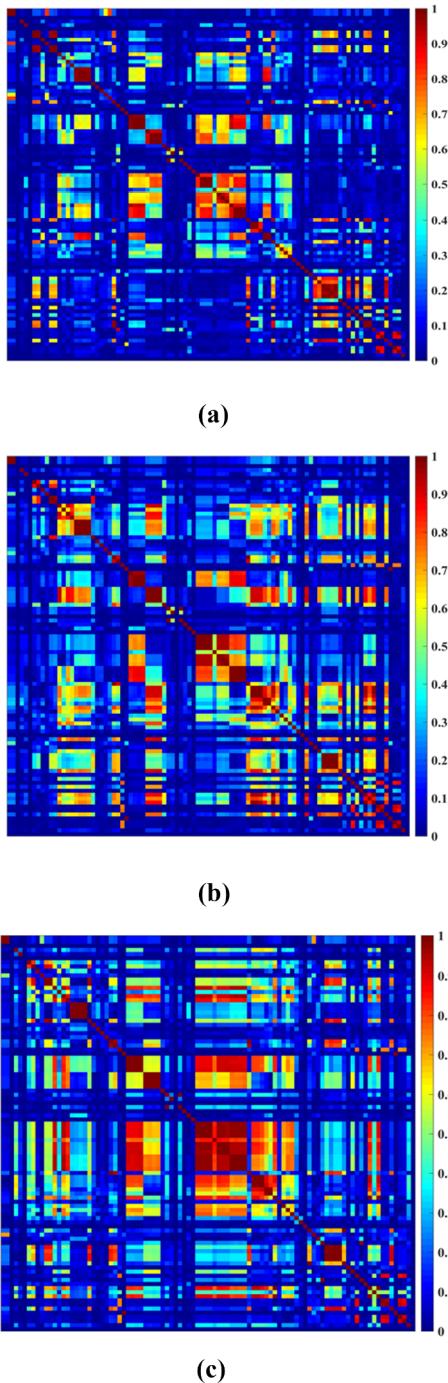


Fig. 8. Correlation matrix of the feature space at (a) takeoff, (b) ascent and (c) cruise phases.

at a particular phase. Using this method, random sensor failures can be captured which may not necessarily be continuous sensor functionality loss, as described in [Section 5](#). The number of features investigated in each flight phase as well as the key assumptions are summarized in [Table 3](#).

4. Model training and anomaly detection

Based on the aforementioned hypotheses and assumptions, the global mean vectors and covariance matrices at different flight phases are trained by the FDR data, and used to evaluate the aircraft performance by the LMG model. In this section, the mini-batch

Table 3

Number of features investigated in each flight phase and key assumptions.

Flight Phase	Number of Features	Zero-Variance Threshold	Anomaly Ratio	Time Duration Threshold [s]
Parking	21	50	0.1%	30
Taxi	38	50	0.1%	30
Takeoff	70	50	0.1%	5
Ascent	81	50	0.1%	30
Cruise	78	50	0.1%	30
Descent	80	50	0.1%	30
Landing	72	50	0.1%	5

mean vector training processes of four representative features will be demonstrated under multiple flight phases, since it is difficult to demonstrate the covariance learning due to its higher dimensionality. It will be followed by the illustration of anomaly detection using the performance index (i.e., logarithmic probability density function) obtained by LMG model. The flight phases that will be shown in this section include takeoff, ascent and cruise, while the representative features are body longitudinal acceleration, exhaust gas temperature for sensor 1, lateral acceleration and pitch angle. Fig. 9 shows the mini-batch learning processes for the mean of the four mentioned features at takeoff, ascent and cruise phases. It can be seen that (i) these features possess different mean values at the three investigated phases; (ii) the mean values vary significantly when they are trained by a smaller number of flights (< 500); (iii) the mean values achieve good convergences once a relatively large number of flights (> 3000) have been trained. It should also be noted that the values of these feature do not necessarily indicate general performance of commercial scheduled flights, since the investigated dataset may be obtained from flights with specific aircraft type(s) and flight plan(s).

The distributions of performance index at takeoff, ascent and cruise phases are shown in Fig. 10. The histograms of takeoff and ascent phases are truncated at $-10,000$ and the histogram of cruise phase is truncated at $-20,000$; the actual minimum value of the

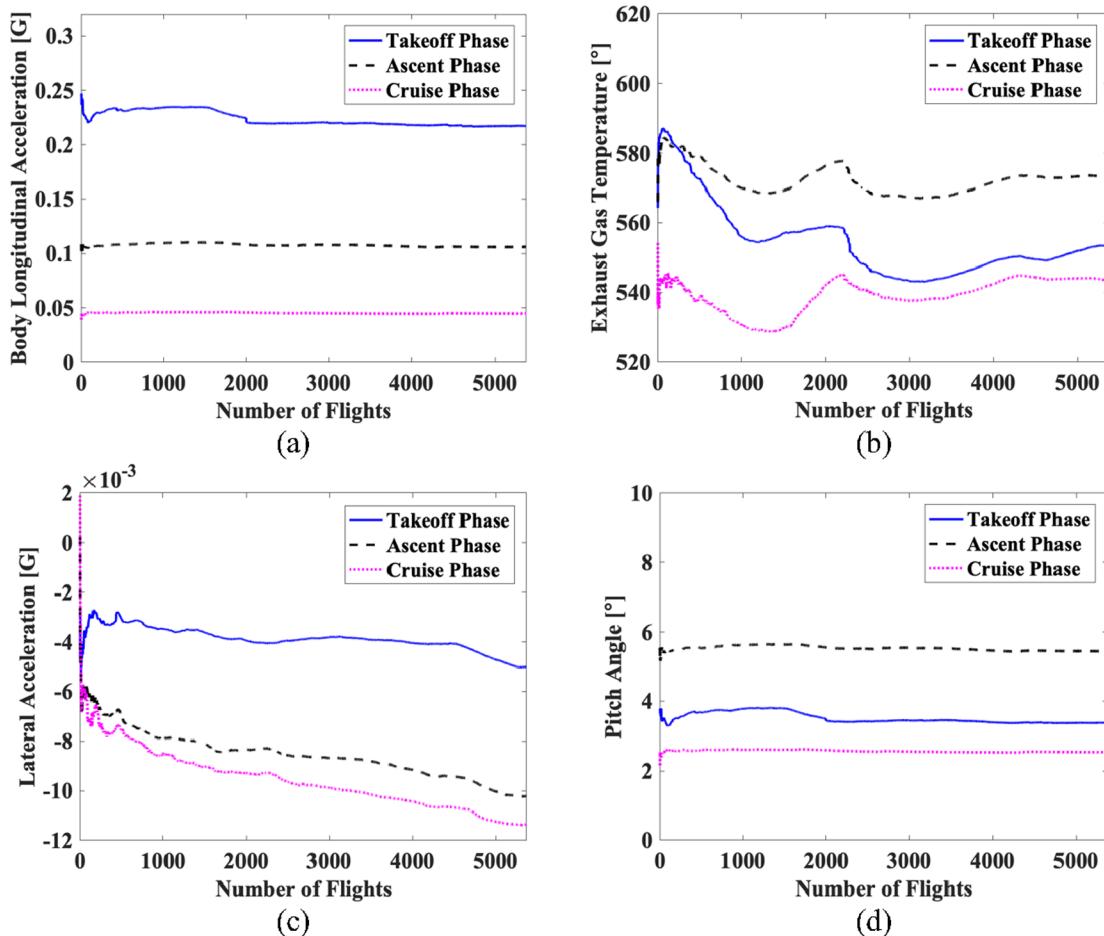


Fig. 9. Demonstration of mini-batch learning processes of (a) body longitudinal acceleration, (b) exhaust gas temperature from sensor 1, (c) lateral acceleration and (d) pitch angle at takeoff, ascent and cruise phases.

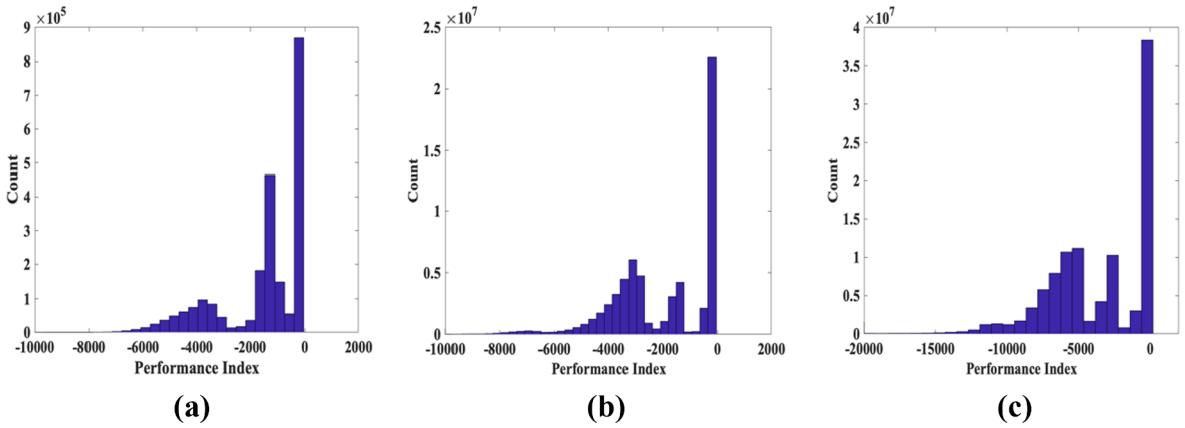


Fig. 10. Distribution of performance index (i.e., Logp) at (a) takeoff, (b) ascent and (c) cruise phases.

performance index at takeoff, ascent and cruise phases are -2.85×10^6 , -2.35×10^6 and -4.04×10^7 , respectively. It should be noted that the truncation made here is only for a better visualization of histogram in Fig. 10 due to the widespread performance index, and these cut-off values do not imply the correlation between flight phases and fraction of anomalies. It is found most of the data points are close to the trained model, which are in the range from the truncated value to zero; 97.46% at takeoff phase, 77.39% at ascent phase and 77.10% at cruise phase. Compared with takeoff and ascent phases in the entire histograms, cruise phase shows a more wide-spread distribution with a less actual minimum value of performance index, indicating the performance of investigated aircraft contains more stochasticity at cruise phase. Based on the pre-defined anomaly ratios and time duration thresholds (see Table 3), the investigation of anomalies at each flight phase will be discussed in the next section.

5. Anomaly analysis

According to the approaches, hypotheses and assumptions mentioned before, the anomalous events in the investigated dataset are evaluated, and the number of flights with detected anomalies are provided in Table 4 for each flight phase. It should be noted that (i) the order of anomalies does not reflect their significance, and (ii) the total number of anomalous cases is larger than the number of flights with anomalies for some flight phases such as cruise, since anomalous events may happen more than once in some flights. It is also worth mentioning that the detected anomalies are not expected to cover all possible anomalies, and the number of flights with detected anomalies does not statistically indicate the frequency of occurrence of an anomaly. This is because the type and number of detected anomalies depend on the assumed anomaly ratio and time duration threshold, which is very difficult to be cross-validated using the current dataset. Instead, the list of these anomalies can be regarded as potential anomalous events in commercial scheduled flights which may then be reviewed by experts in the field of air transportation, to identify and prevent unsafe issues, processes, or maneuvers. Through the investigation of the flight signals, the authors have detailed the anomalies at each phase in Appendix A. The detected anomalies can be categorized into two classes: operational anomalies and mechanical anomalies. The operational anomalies indicate operations conducted by the pilot that deviate from standard operation (observed over most flights), such as low power lever angle during takeoff, sudden changes in spoiler position during ascent, abrupt selected course changes without heading change during cruise, etc. The mechanical anomalies majorly represent the event that the received signal(s) exhibits an abnormal value, which cannot be observed in other flights during a flight phase. It is found that some anomalies in sensor signals at particular time instances including uncommon oil temperature at taxiing, inconsistent angle of attack from two different sensors during takeoff, engine signals (exhaust gas temperature, fan speed, etc.) drop to zero during cruise, etc. Such mechanical anomalies might be caused by temporary/permanent function loss of sensors, unusual weather conditions, or uncommonly observed pilot operations.

Three representative examples of the detected anomalies are exhibited in this section. Fig. 11 presents the anomalous cases associated with abnormal sensor signals. The power lever angle recorded by sensor 1 shows low and unrealistic values throughout the

Table 4
Number of flights with detected anomalies at each flight phase.

Flight Phase	Number of Flights with Detected Anomalies
Parking	3
Taxi	29
Takeoff	5
Ascent	16
Cruise	48
Descent	14
Landing	5

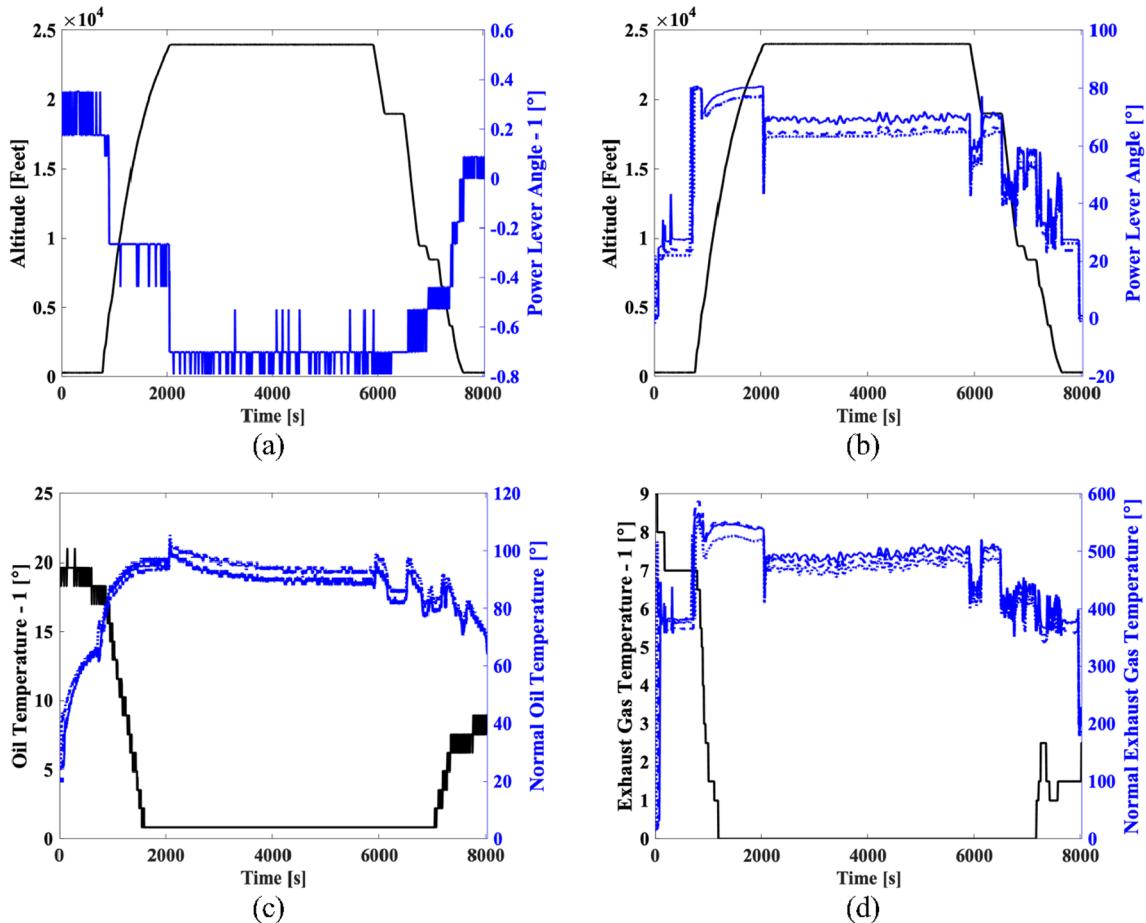


Fig. 11. Demonstration of detected anomaly associated with abnormal sensor signals, illustrated signals including (a) power lever angle from sensor 1, (b) power lever angle from sensor 2, 3 and 4, (c) oil temperature and (d) exhaust gas temperature.

entire flight (see Fig. 11(a)), compared with those from the other three sensors shown in Fig. 11(b). In addition, other engine related parameters such as the oil temperature (Fig. 11(c)) and exhaust gas temperature (Fig. 11(d)) from sensor 1 also exhibit unrealistically low values. It indicates that engine 1 may undergo mechanical issues, or the sensors of engine 1 experience function loss that prevents them from correctly reflecting the status of the engine. Fig. 12 displays an instantaneous anomaly where a sudden drop of flight path

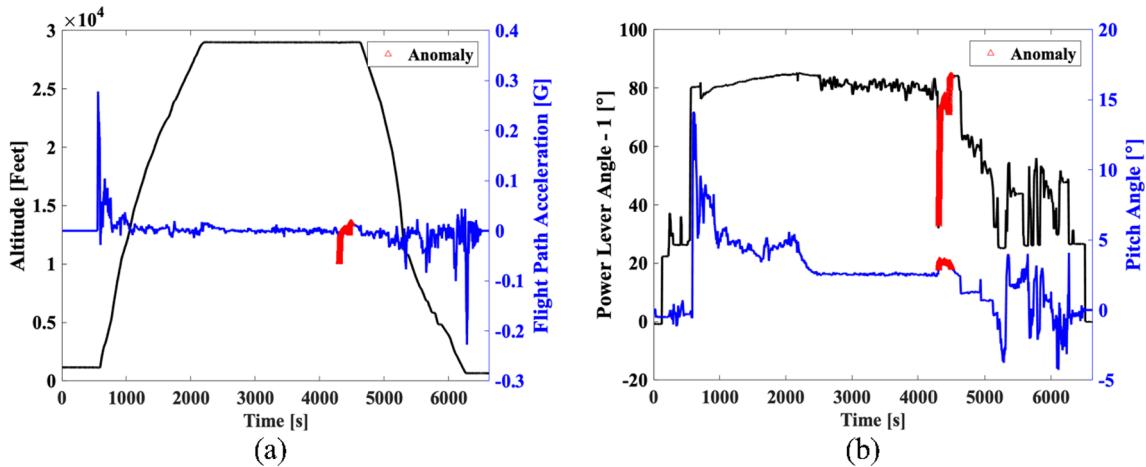


Fig. 12. Demonstration of detected anomaly associated with early power reduction before descent; illustrated signals including (a) flight path acceleration and (b) power lever angle from sensor 1 and pitch angle.

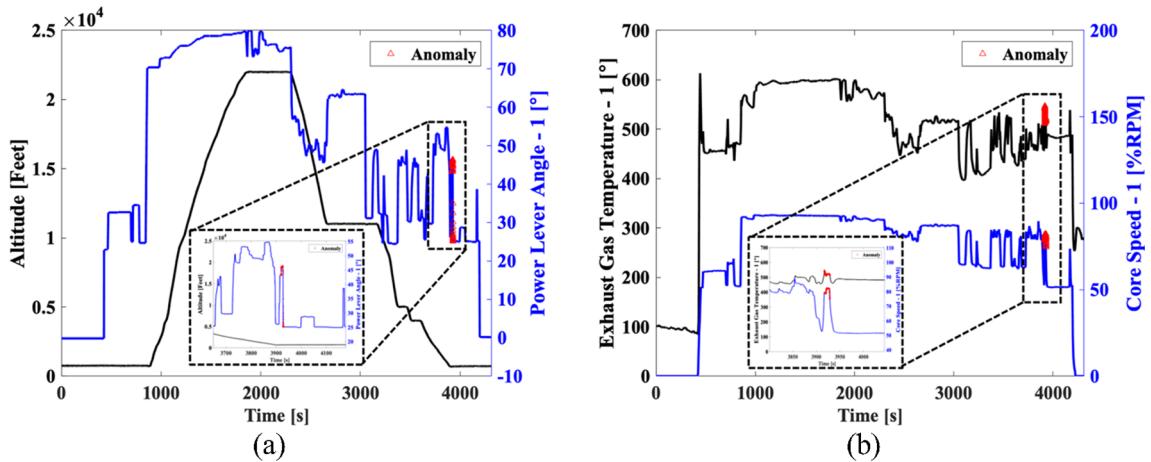


Fig. 13. Demonstration of detected anomaly associated with high power landing; illustrated signals include (a) power lever angle and (b) exhaust gas temperature and core speed from sensor 1.

acceleration is detected at cruise phase (refer to Anomaly 8 in Table A.6). Based on the interpretation of the subject experts and a detailed signal analysis, the anomalous issue is thought to have been triggered when the pilot prematurely (presumably by mistake) reduced the power lever angle before the top-of-descent as shown in Fig. 12(b). In order to compensate for the sudden reduction in the power level angle and maintain a stable cruise altitude, the pilot then chose to increase the pitch angle until the power was recovered. Fig. 13 exhibits an anomalous increase in the power lever angle during landing (refer to Anomaly 2 in Table A.8). Due to the increase in power lever angle, the engine related signals also show relatively high values during landing (Fig. 13 (b)); anomalies such as this may cause unstable touch downs. It should be noted that the signals presented in Figs. 12 and 13 are from sensor 1 only, since the signals from the four engines are almost identical throughout the entire flight.

It is worth mentioning that the presented model was trained using an Intel i5-8600 k processor (six cores with a 3.6 GHz based core clock) in a Windows 10 64-bit operation system, with 32 GB DDR4 memory. The total training and testing time of the mentioned dataset (417,448,448 data points in total) were 16,023 s and 11,559 s, respectively. It indicates that the averaged training and testing time for each data point is 38.4 μ s and 27.7 μ s per data point, respectively. It can be found that the total number of data points mentioned here is larger than that presented in Table 2. It is because that the data for training and testing contains all the recorded data points, but the anomaly analysis (shown in Table 2) ignored the data points without information of flight phase.

Compared with the anomalies reported in existing literature using FOQA datasets (Li et al., 2015; Li et al., 2016), the proposed method shows good similarity in the context of the types of detected anomalies at approach (descent and landing phases in this research) and takeoff phases, including low power takeoff, high energy approach, abnormal values in the engine, etc. Furthermore, the developed model is capable of detecting anomalies for the entire flight, which has not been reported in existing literature, with sufficient sensitivity to both mechanical and operational issues. One noticeable discrepancy is that Li et al. (2016) reported the effect of a strong and quartering tail wind that results in abnormal shift to the center of gravity, which is not detected by the current research, since no sensor information indicating tail wind of the aircraft is available in the FDR dataset. Additionally, as mentioned in Section 3.1, the model has not been validated with actual faulty or fatal issues due to the rarity of such cases and difficulty in acquiring such data. Hence, although the developed model and anomalies summarized in this section can potentially assist air transportation management to achieve better safety, it should be noted that the mentioned anomalies may or may not lead to consequences that can be concretely deemed unsafe.

6. Concluding remarks

A logarithmic multivariate Gaussian model has been developed to evaluate the performance of aircrafts through sensor information from completed commercial scheduled flights. Such a model overcomes the difficulties associated with large size and high dimensionality in the flight dataset using a mini-batch training process and performance evaluation in the logarithmic domain. This model, most importantly, addresses the aforementioned re-sampling issue in the existing literature, which prevents the monitoring framework from handling time-series signals with different lengths. Specifically, with the introduced hypothesis that the performance of aircraft follows a unimodal Gaussian distribution, the proposed multivariate Gaussian model eliminates the dependence of performance evaluation on time and further enables the model to handle flights of any time duration. In addition, the unimodal hypothesis, which has been proven in earlier manuscripts, generalizes the proposed approach: the trained model (i.e., global mean and covariance vectors) only represents commonly observed behaviors. It indicates the proposed approach will flag any behavior that deviates from normal behavior without the need for knowing the actual issue *a priori*. Therefore, it is well-suited for commercial scheduled flights since the potential mechanical or operational issues may not be recorded by historical dataset due to the complexity of aircraft systems and stochasticity in human (pilot and control tower) behavior.

The developed model is expected to be an effective addition to current anomaly analysis and monitoring technologies for

scheduled commercial flight. Moreover, in addition to the direct application, the model is also expected to possess the potential in handling historical sensor data for anomaly detection and monitoring of other similar transportation components such as unmanned aerial vehicle, self-driving car, and high-speed rails. Specifically, it can be used to help transportation management to handle large amounts of historical dataset, and assist them in analyzing mechanical and operational anomalies, which may further improve future sensor systems and pilot training. In this case, the data size will be dramatically increased when applying the model to such applications due to the exponentially increased size of the transportation dataset. Therefore, the authors suggest a parallel computation architecture in the training phase to further improve training efficiency based on the compatible scheme shown in pseudo code. Moreover, the developed model, after training, can be potentially equipped as a real-time on-board monitoring system due to its computational efficiency. To overcome the issue associated with the absence of labeled information for anomaly analysis on commercial scheduled flights, the authors suggest building a hybrid dataset combining realistic FDR dataset and simulated signals for aircraft models.

Acknowledgements

The research reported in this paper was supported by funds from NASA University Leadership Initiative program (Contract No. NNX17AJ86A, Project Officer: Dr. Anupa Bajwa and Dr. Kai Goebel). In addition, the help in anomaly interpretation from Dr. P.K. Menon is sincerely appreciated. All supports are gratefully acknowledged.

Appendix A

(See. [Tables A1–A8](#)).

Table A1
Definitions of abbreviations used in [Tables A.2–A.8](#).

Abbreviation	Definition
EGT	Exhaust gas temperature
PLA	Power lever angle
N1	Fan speed
N2	Core speed
AOA	Angle of attack
PTCH	Pitch angle

Table A2
Anomalies detected at parking phase.

Index	Description of anomaly	Number of flights out of 3 flights with anomalies
1	Uncommon drift and PTCH	1
2	No anomalous event found	2

Table A3
Anomalies detected at taxiing phase.

Index	Description of anomaly	Number of flights out of 29 flights with anomalies
1	Aircraft turns around and induces change in brake pressure, aileron, elevator, etc.	16
2	True airspeed is not consistent with calibrated airspeed	1
3	Increase in static air temperature and total air temperature; drop in static pressure with a sudden change in AOA	1
4	Changes in selected vertical speed and heading; increase in AOA with drop in airspeed	1
5	Uncommon wheel position and oil temperature	1
6	Unrealistic AOA*	3
7	Low EGT*	5
8	Negative oil temperature	1

* Flight does not take off.

Table A4

Anomalies detected at takeoff phase.

Index	Type of anomaly	Number of flights out of 5 flights with anomalies
1	Inconsistency in AOAs from two sensors; spikes in EGT from sensor 3 and 4*	1
2	Zero oil pressure; low EGT; zero airspeed and vertical speed; power fluctuations*	2
3	Relatively low PLA and N1; brake activated; low true airspeed	1
4	Low rudder position; Zero N1, N2, EGT, and PLA from one sensor	1

* Flight does not take off.

Table A5

Anomalies detected at ascent phase.

Index	Type of anomaly	Number of flights out of 16 flights with anomalies
1	Anomalous ascend phase before final descent	1
2	Sudden descent during ascent with air brake activated	3
3	Sudden changes in one of the roll spoilers	5
4	Increased vertical speed and PTCH before cruise phase	1
5	Stop ascending during ascent phase	2
6	No takeoff but still showing ascent phase	3
7	One engine not working	1
8	Negative selected course	1

Table A6

Anomalies detected at cruise phase.

Index	Type of anomaly	Number of flights out of 48 flights with anomalies
1	Flap up and pitch down before descent	11
2	Unstable cruise: fluctuations in AOA; unstable cross track acceleration; sudden change in heading	6
3	Spikes in brake pressure	4
4	Engine loses function (power, N1, N2 and EGT down to zero or near zero) causing drop in airspeed	2
5	Engine signals drop to zero or near zero	8
6	Reduced PLA results in reduced longitudinal speed; increased PTCH and airspeed drop	8
7	Airspeed reduction before descending, especially during short cruise between two descend phases	2
8	Early reduction in PLA angle before top-of-descent; increased PTCH before PLA recovered	4
9	Increased PLA while observed drop in airspeed with increased PTCH	2
10	Abrupt selected course changes without heading change	1
11	Fluctuations in vertical acceleration and altitude rate; probably due to turbulence	3
12	Spikes in impact pressure	1
13	Sudden increase in pitch trim position	1

Table A7

Anomalies detected at descent phase.

Index	Type of anomaly	Number of flights out of 14 flights with anomalies
1	Inconsistent signals from engine sensors and engine function loss	7
2	Multiple takeoff and landing	1
3	No takeoff observed	1
4	Changes in wind speed and drift angle	1
5	Wrong flight phase indicator	4

Table A8

Anomalies detected at landing phase.

Index	Type of anomaly	Number of flights out of 5 flights with anomalies
1	Extremely low PLA	1
2	Increased power: spikes in PLA and EGT from 3 out of four sensors	3
3	Reduced EGT, N1, PLA and N2	1

Appendix B. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.trc.2019.11.011>.

References

- Banerjee, T.P., Das, S., 2012. Multi-sensor data fusion using support vector machine for motor fault detection. *Inf. Sci.* 217, 96–107.
- Bartlett, L.M., Hurdle, E.E., Kelly, E.M., 2009. Integrated system fault diagnostics utilising digraph and fault tree-based approaches. *Reliab. Eng. Syst. Saf.* 94 (6), 1107–1115.
- Basir, O., Yuan, X., 2007. Engine fault diagnosis based on multi-sensor information fusion using Dempster-Shafer evidence theory. *Inform. Fusion* 8 (4), 379–386.
- Bay, S.D., Schwabacher, M., 2003. Mining distance-based outliers in near linear time with randomization and a simple pruning rule. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 29–38.
- Budalakoti, S., Srivastava, A.N., Akella, R., 2006. Discovering atypical flights in sequences of discrete flight parameters. In: 2006 IEEE Aerospace Conference. IEEE, pp. 1–8.
- Budalakoti, S., Srivastava, A.N., Otey, M.E., 2009. Anomaly detection and diagnosis algorithms for discrete symbol sequences with applications to airline safety. *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* 39 (1), 101–113.
- Burges, C.J., 2010. Dimension reduction: a guided tour. *Found. Trends® Machine Learn.* 2 (4), 275–365.
- Causse, M., Dehais, F., Péran, P., Sabatini, U., Pastor, J., 2013. The effects of emotion on pilot decision-making: a neuroergonomic approach to aviation safety. *Transport. Res. Part C: Emerg. Technol.* 33, 272–281.
- Cerrada, M., Sánchez, R.V., Li, C., Pacheco, F., Cabrera, D., de Oliveira, J.V., Vásquez, R.E., 2018. A review on data-driven fault severity assessment in rolling bearings. *Mech. Syst. Sig. Process.* 99, 169–196.
- Dalton, R.P., Cawley, P., Lowe, M.J.S., 2001. The potential of guided waves for monitoring large areas of metallic aircraft fuselage structure. *J. Nondestr. Eval.* 20 (1), 29–46.
- Das, S., Matthews, B.L., Srivastava, A.N., Oza, N.C., 2010. Multiple kernel learning for heterogeneous anomaly detection: algorithm and aviation safety case study. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 47–56.
- DeCarlo, E.C., Bichon, B.J., 2018. Modeling the effects of uncertainty on the national airspace system. In: PHM Society Conference (Vol. 10, No. 1).
- Federal Aviation Administration, 2004. Advisory Circular. 120–82 Flight Operational Quality Assurance. Federal Aviation Administration.
- Fleischer, A., Tchetchik, A., Toledo, T., 2015. Does it pay to reveal safety information? The effect of safety information on flight choice. *Transport. Res. Part C: Emerg. Technol.* 56, 210–220.
- Foo, G.H.B., Zhang, X., Vilathgamuwa, D.M., 2013. A sensor fault detection and isolation method in interior permanent-magnet synchronous motor drives based on an extended Kalman filter. *IEEE Trans. Ind. Electron.* 60 (8), 3485–3495.
- Ge, Z., 2017. Review on data-driven modeling and monitoring for plant-wide industrial processes. *Chemometrics Intell. Laborat. Syst.* 171, 16–25.
- ICAO Safety Report 2018 Edition. International Civil Aviation Organization, Montreal, Canada.
- Knorr, E.M., Ng, R.T., Tucakov, V., 2000. Distance-based outliers: algorithms and applications. *VLDB J.—Int. J. Very Large Data Bases* 8 (3–4), 237–253.
- Kuhn, K.D., 2018. Using structural topic modeling to identify latent topics and trends in aviation incident reports. *Transport. Res. Part C: Emerg. Technol.* 87, 105–122.
- Lee, H., Li, G., Rai, A., Chattopadhyay, A., 2019. Anomaly detection of aircraft system using Kernel-based learning algorithm. In: AIAA Scitech 2019 Forum, pp. 1224.
- Li, L., Das, S., John Hansman, R., Palacios, R., Srivastava, A.N., 2015. Analysis of flight data using clustering techniques for detecting abnormal operations. *J. Aerosp. Informat. Syst.* 12 (9), 587–598.
- Li, L., Hansman, R.J., Palacios, R., Welsch, R., 2016. Anomaly detection via a Gaussian Mixture Model for flight operation and safety monitoring. *Transport. Res. Part C: Emerg. Technol.* 64, 45–57.
- Li, G., Rai, A., Lee, H., Chattopadhyay, A., 2018. Operational anomaly detection in flight data using a multivariate gaussian mixture model. PHM Society Conference. (vol. 10, No. 1).
- Liu, Y., Goebel, K., 2018. Information fusion for national airspace system prognostics. In: PHM Society Conference (Vol. 10, No. 1).
- Mairal, J., Bach, F., Ponce, J., Sapiro, G., 2009. Online dictionary learning for sparse coding. In: Proceedings of the 26th Annual International Conference on Machine Learning. ACM, pp. 689–696.
- McNeely, J., Hatfield, M., Hasan, A., Jahan, N., 2016. Detection of uav hijacking and malfunctions via variations in flight data statistics. In: 2016 IEEE International Carnahan Conference on Security Technology (ICCST). IEEE, pp. 1–8.
- Menon, P.K., Yang, B.J., Dutta, P., Park, S.G., Chen, O., Cheng, V.H., 2018. A computational platform for analyzing the safety of the national airspace system. In: PHM Society Conference (Vol. 10, No. 1).
- Mikaelian, T., Williams, B.C., Sachenbacher, M., 2005. Model-based monitoring and diagnosis of systems with software-extended behavior. In: AAAI, vol. 5. pp. 327–333.
- Ng, A., 2015. Lectures: Machine Learning.
- National Transportation Safety Board, 2013. Phase of Flight – Definitions and Usage Notes, Version 1.3, Washington, DC, U.S.
- Oonk, S., Maldonado, F.J., Figueiroa, F., Lin, C.F., 2012. Predictive fault diagnosis system for intelligent and robust health monitoring. *J. Aerospace Comput. Inform. Commun.* 9 (4), 125–143.
- Papathanasiou, V., Antoniou, C., 2015. Towards data-driven car-following models. *Transport. Res. Part C: Emerg. Technol.* 55, 496–509.
- Rebollo, J.J., Balakrishnan, H., 2014. Characterization and prediction of air traffic delays. *Transport. Res. Part C: Emerg. Technol.* 44, 231–241.
- Samanta, B., Al-Balushi, K.R., 2003. Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. *Mech. Syst. Sig. Process.* 17 (2), 317–328.
- Schwabacher, M., Oza, N., Matthews, B., 2009. Unsupervised anomaly detection for liquid-fueled rocket propulsion health monitoring. *J. Aerospace Comput. Inform. Commun.* 6 (7), 464–482.
- Sihombing, F., Torbol, M., 2018. Parallel fault tree analysis for accurate reliability of complex systems. *Struct. Saf.* 72, 41–53.
- Sun, J., Ellerbroek, J., Hoekstra, J.M., 2019a. WRAP: An open-source kinematic aircraft performance model. *Transport. Res. Part C: Emerg. Technol.* 98, 118–138.
- Sun, Z., Zhang, C., Tang, P., Wang, Y., Liu, Y., 2019. Bayesian Network Modeling of airport runway incursion occurring processes for predictive accident control. In: Advances in Informatics and Computing in Civil and Construction Engineering. Springer, Cham, pp. 669–676.
- Tsuruta, G.M.L., 2008. The Analysis of Flight Operational Quality Assurance (FOQA) Data: Exploration of a Proposed List of Improved Safety Parameters.
- Vanini, Z.S., Khorasani, K., Meskin, N., 2014. Fault detection and isolation of a dual spool gas turbine engine using dynamic neural networks and multiple model approach. *Inf. Sci.* 259, 234–251.
- Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N., 2003b. A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies. *Comput. Chem. Eng.* 27 (3), 313–326.
- Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N., Yin, K., 2003c. A review of process fault detection and diagnosis: Part III: Process history based methods. *Comput. Chem. Eng.* 27 (3), 327–346.
- Venkatasubramanian, V., Rengaswamy, R., Yin, K., Kavuri, S.N., 2003a. A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. *Comput. Chem. Eng.* 27 (3), 293–311.
- Wang, W., Ying, L., 2018. Dynamic data communications for real-time information fusion. In: PHM Society Conference (Vol. 10, No. 1).
- Wang, Y., Liu, Y., Sun, Z., Tang, P., 2018. A Bayesian-entropy network for information fusion and reliability assessment of national airspace systems. In: PHM Society Conference (Vol. 10, No. 1).
- Williams, Christopher K.I., Rasmussen, Carl Edward, 2006. Gaussian processes for machine learning. vol. 2, no. 3. Cambridge, MA: MIT Press.
- Xiang, S., Nie, F., Zhang, C., 2008. Learning a Mahalanobis distance metric for data clustering and classification. *Pattern Recogn.* 41 (12), 3600–3612.
- Yu, Y., Yao, H., Liu, Y., 2018. Physics-based learning for aircraft dynamics simulation. In: PHM Society Conference (Vol. 10, No. 1).
- Yu, Y., Yao, H., Liu, Y., 2019. A hybrid learning approach for the simulation of aircraft dynamical systems. In: AIAA Scitech 2019 Forum, pp. 0436.
- Zhang, X., Kong, Y., Subramanian, A., Mahadevan, S., 2018. Data-driven modeling for aviation safety diagnosis and prognosis. In: PHM Society Conference (Vol. 10, No. 1).
- Zhang, X., Mahadevan, S., 2019. Ensemble machine learning models for aviation incident risk prediction. *Decis. Support Syst.* 116, 48–63.

- Zhang, J., Sato, T., Iai, S., 2006. Support vector regression for on-line health monitoring of large-scale structures. *Struct. Saf.* 28 (4), 392–406.
- Zhang, G., Wang, J., Lv, Z., Yang, Y., Su, H., Yao, Q., Huang, Q., Ye, S., Huang, J., 2015. A integrated vehicle health management framework for aircraft—A preliminary report. In: 2015 IEEE Conference on Prognostics and Health Management (PHM). IEEE, pp. 1–8.
- Zhao, B., Liu, D., Li, Y., 2017. Observer based adaptive dynamic programming for fault tolerant control of a class of nonlinear systems. *Inf. Sci.* 384, 21–33.
- Zheng, F., Van Zuylen, H., 2013. Urban link travel time estimation based on sparse probe vehicle data. *Transport. Res. Part C: Emerg. Technol.* 31, 145–157.