Identifying Instantaneous Anomalies in General Aviation Operations

Tejas G. Puranik[*] and Dimitri N. Mavris[†]

Georgia Institute of Technology, Atlanta, Georgia, 30332-0150

Quantification and improvement of safety is one of the most important objectives among the General Aviation community. In recent years, data mining techniques are emerging as an important enabler in the aviation safety domain with a number of techniques being applied to flight data to identify and isolate anomalous (and potentially unsafe) operations. There are two types of anomalies typically identified – flight-level (where the entire flight exhibits patterns deviating from nominal operations) and instantaneous (where a subset or few instants of the flight deviate significantly from nominal operations). Energy-based metrics provide measurable indications of the energy state of the aircraft and can be viewed as an objective currency to evaluate various safety-critical conditions across a heterogeneous fleet of aircraft. In this paper, a novel method of identifying instantaneous anomalies for retrospective safety analysis using energy-based metrics is proposed. Each data record is split by sliding a moving window across the multi-variate series of evaluated energy metrics. A mixture of gaussian models is then used to perform clustering using the values of energy metrics and their variability within each window. The trained models are then used to identify anomalies that may indicate increased levels of risk. The identified anomalies are compared with traditional methods of safety assessment (exceedance detection).

I. Introduction

Aviation is statistically one of the safest modes of transportation within the US and worldwide.[1] It accounts for less than 2% of the fatalities from transportation related accidents in 2013. However, a significant majority of aviation accidents are within the General Aviation (GA) domain.[1] According to the National Transportation Safety Board (NTSB),[1] the total number of accidents per million flight hours in GA is an order of magnitude higher than commercial operations. Therefore, improving the safety in GA operations is of particular interest to the industry. According to the Federal Aviation Administration (FAA), the demand for air travel and traffic is predicted to grow steadily through 2036 at a rate of approximately 1.8% annually.[2] Per the FAA, commercial operations are expected to double and GA operations are also expected to gain a much-needed revitalization in the coming years. With such a large increase in expected operations, there is an ever-increasing demand for improving safety of all aviation operations (especially GA).

In the past, accidents have been the primary triggers for identifying problems and developing mitigation strategies.[3] However, the industry is now moving towards a more proactive approach to safety enhancement in which potential unsafe events are identified beforehand and mitigation strategies are implemented in order to prevent accidents and loss of life. The FAA have outlined various voluntary safety improvement programs such as the Flight Operational Quality Assurance (FOQA), Aviation Safety Action Program (ASAP) and others. Aviation Safety Information Analysis and Sharing (ASIAS)[4] is a system which aims to connect a number of such safety databases in order to facilitate integrated queries across multiple safety databases. All of these efforts are geared towards enhancing the objective of proactive safety enhancement. While many of these and other programs started for commercial operations, they have trickled into GA as well. Specifically within GA, the GAJSC is a government and industry group that uses the same approach as the CAST on the commercial operations side by analyzing GA safety data to develop intervention strategies to prevent or mitigate problems associated with accident causes, called Safety Enhancements (SE).[5]

[*]Graduate Researcher, Daniel Guggenheim School of Aerospace Engineering, AIAA Student Member

[†]S.P. Langley NIA Distinguished Regents Professor, Daniel Guggenheim School of Aerospace Engineering, AIAA Fellow

The most common and widespread programs in existence for quantitative safety assessment are Flight Operational Quality Assurance (FOQA)[6] or Flight Data Monitoring (FDM).[7] It is defined as: "*The systematic proactive use of digital flight data from routine operations to improve aviation safety within a non-punitive and just safety culture*". Retrospective analysis of flight data in FOQA programs is one of the most important enablers for quantitative safety assessment. FOQA programs aim to improve operational safety with a continuous cycle involving data collection from on-board recorders, retrospective analysis of flight data records, identification of operational safety exceedances, design and implementation of corrective measures, and monitoring to assess their effectiveness. Implementation of FOQA is widespread in commercial aviation, and although it is sparse among GA operators, recent and current efforts seek its introduction and broad adoption in that sector. Exceedance detection is the most common method of analysis using FOQA data currently in existence. An exceedance is the deviation of a single parameter beyond an established threshold. An event is defined by one or more parameter exceedances that take place concurrently over a specified period of time.

With the availability of flight data and advanced computing power, data mining techniques for safety analysis, incident examination, and fault detection are garnering increased interest in the aviation community in recent years. The current practice of exceedance detection performs well on known safety issues but is blind to safety-critical conditions that may be captured by flight data records but not included in the set of defined safety events. Additionally, the events defined in exceedance detection are 'static' in that they do not take into account the information that might be available in a wealth of recorded flight data. Data mining approaches have the potential of revealing safety events of interest as emergent artifacts from within a wealth of flight data records. While formal techniques for flight data analysis are not new, applications of data mining for retrospective operational safety analysis are fairly sparse. Additionally, a large portion of the existing literature is dedicated to commercial aviation even though historically, GA operations have had considerably greater accident and incident rates.

Previous applications of data mining in the aviation safety domain have primarily treated it as an anomaly detection problem with data objects as multivariate time series.[8–12] In the data mining community, anomaly detection is loosely defined as the *"task of obtaining patterns in data that do not conform to a well defined notion of normal behavior"*.[13] The exact notion of anomaly is different for different application domains. In the aviation safety domain, two main types of anomalies are of interest to the analyst – *instantaneous* and *flight level*. Instantaneous anomalies refer to those anomalies where only an instant or small part (a few seconds) of the flight record is abnormal. Flight level anomalies refer to those flights with abnormal data patterns over a an entire phase of flight. The main focus of this paper is identification of instantaneous anomalies in GA flight data.

While the accident rates for GA flight operations are higher than those in commercial aviation, it must be noted that GA contains a very heterogeneous fleet and operations. This includes single and multi-engine piston, turboprop, turbojet powered aircraft as well as helicopters and experimental aircraft. However, accident and incident rates and number of active aircraft are not uniform across all the aircraft classes within GA. Historically single-engine piston aircraft make up a significant proportion of the entire fleet and the number of accidents and fatalities.[14] Therefore, it is of particular importance to examine operations of this class of aircraft within GA and to improve its safety record. Most of these aircraft typically belong to the *normal* category under the FAA airplane categories (14 CFR Part 23.3[15]). This category is limited to airplanes that have a seating configuration, excluding pilot seats, of nine or less, a maximum certificated takeoff weight of 12,500 pounds or less, and intended for non-acrobatic operation. Following are important points that distinguish the category of aircraft and operations which form the focus of this paper:

1. Smaller sized aircraft
2. Less weather-tolerant aircraft
3. Limited or no flight data recording capability
4. Highly variable and heterogeneous mission profiles
5. Variability in pilot certificate and experience level (number of hours flown)
6. Greater variety of airports of operations
7. Operations mostly under Visual Flight Rules (VFR)

All of these factors present important challenges for improving safety of GA operations. Therefore, existing mature methods from commercial operations need to be tailored to GA operations or entirely new methods need to be developed.

The rest of the paper is organized as follows – Section II contains a review of relevant past work related to the identification of anomalies from flight data. Section III presents the details of the methodology developed for identifying instantaneous anomalies and its implementation. Section IV contains the results obtained from the application of this methodology on a set of real flight data from GA operations and the comparison of these results to traditional exceedance detection techniques. Section V provides the conclusions and recommendations from this body of work and outlines avenues for future work.

II. Review of Existing Work

Majority of the prior work in the application of data mining techniques to aviation data focuses on flight level anomalies. SequenceMiner[16] has been shown to detect abnormal switching from a learned model of normal switching. This technique detects flight level anomalies but is limited to discrete data. Gorinevsky et al.[17] have described an application of data mining technology called Distributed Fleet Monitoring (DFM) to Flight Operational Quality Assurance data. This application consists of fitting a large scale multi-level regression model to the data set and finding anomalies using these built models. The algorithm is able to identify anomalies within a flight record (instantaneous), abnormal flight-to-flight trends (flight level anomalies) and abnormally performing aircraft. Hotelling T^2 statistics for residuals from the built models are calculated and used for monitoring and identifying anomalies. While this framework is capable of identifying instantaneous anomalies, it is limited to models fitted in the clean configuration. Also, most of the anomalies detected are in the determination of aerodynamic or propulsion parameter estimates or gross weight. Das et. al.[9] have developed Multiple Kernel Anomaly Detection (MKAD) which applies a one-class support vector machine for anomaly detection. MKAD identifies flight level anomalies well in heterogeneous data but is not built to identify instantaneous anomalies. Moreover, using the normalized Longest Common Sub-sequence (nLCS) kernel is useful for discrete data, but it results in loss of some finer features for continuous data. For instantaneous anomalies, the finer features are of more interest. Matthews et al.[10] have discussed and summarized the aviation knowledge discovery pipeline using various algorithms and attempted to combine the strengths of different approaches. Smart and Brown[18] used a data-mining approach using a number of one-class classification algorithms to identify anomalies in the descent phase for airliners. They tested their method on a labeled data set containing abnormal flights. Li et al.[11] have implemented ClusterAD-Flight, which uses cluste6r based anomaly detection on pre-processed flight data parameters to identify flight level anomalies. One of the issues in this method can be that it isolates each sample in each signal as a unique feature, when in fact, the change over that time may be an important factor. In previous work by the authors (Puranik et al.[19]), the application of energy metrics as features along with clustering techniques to identify flight level anomalies has been explored.

Typically techniques that are used to identify instantaneous anomalies are different than those used to identify flight-level anomalies. Supervised learning methods such as Inductive Monitoring System[8] (IMS) rely on a training set consisting of typical system behaviors which is compared with real-time data to detect anomalies. Each point is monitored standalone and therefore, the temporal aspect of anomalous sub-sequences is lost when identifying anomalies. Orca[20] is a technique that uses scalable k-nearest neighbor approach to detect anomalies in data with continuous and discrete features. Since each data point is a sample in time and treated as independent by the algorithm, Orca struggles to detect anomalies with temporal signatures. Amidan and Ferryman[21] have utilized Singular Value Decomposition (SVD) to identify instantaneous anomalies. They mapped the five seconds before and after each recorded data point and fit a linear regression model to it. The coefficients of the regression model were then used to create a mathematical signature for each recorded data point which was used to identify outliers. Mugtussidis[22] has used Bayesian classification to distinguish between typical data points, that are present in the majority of flights, and unusual data points that can be only found in a few flights. Some of the methods rely on developing approximate models using flight data and detecting those flight records which deviate greatly from this model as outliers. Melnyk et al.[12] developed a vector auto regressive exogenous model to detect anomalies from flight data. Instantaneous anomalies are detected as residuals from the model. However, this technique involves identifying a model for each flight in the database and then calculating the residuals with respect to every other flight. Li et al.[23] developed ClusterAD-Data Sample, which is a technique leveraging a mixture of Gaussian models to identify probability of a sample being anomalous during take off, approach, and landing. However, this method also treats each data sample independently and uses additional context to identify whether a particular gaussian component is appropriate at a given time.

In domains other than aviation safety, techniques for identifying instantaneous anomalies (or anomalous sub-sequences) in time-series have been explored. Majority of these techniques focus on univariate time-series data by moving a sliding window through the time-series. For example, Keogh et al.[24] have used a sliding window based technique along with Symbolic Aggregate ApproXimation (SAX) representation of time series to identify unusual sub-sequences. SAX discretizes the continuous data stream into a word by using average values of the parameter in a given window. One of the important assumptions in some of these applications is that the time series are stationary i.e. their mean value does not change over time. This is not necessarily true for most parameters in flight data. In some applications where multivariate time series are considered, they are first converted into a univariate time series. For example Chandola[13] uses the technique of subspace monitoring to find the distance between successive windows (sub-spaces) using the principal angles of each subspace. Once this has been done, the multivariate time series is converted into a univariate series and a number of techniques are used. While the temporal aspect of anomalies is preserved in these techniques, data in each time series is compared only to the data from the same time series. This loses out on potential insights that can be gained from other similar data available.

There are various challenges in GA operations that preclude the use of some of the existing techniques in literature. First, the fleet of aircraft operated within GA is heterogeneous, even within the specific category of aircraft considered, the gross weight, rated engine power, and other aspects of the aircraft can vary significantly. In addition, some aircraft are possibly limited in terms of parameters recorded (if any) by their digital flight data recorders. Therefore, metrics used in anomaly detection algorithms need to be those that correspond to safety margins and safe operations in GA and can be readily estimated from recorded data. Finally, the types of missions flown in GA can be quite different and non-uniform. Therefore, techniques that do not make assumptions about phases of flight or are more broadly applicable across different phases of flight are desired.

In addition to the challenges specific to GA identified in the previous paragraph, there are certain limitations of existing techniques surveyed previously. In many of the approaches for identifying instantaneoous anomalies, each point is monitored as a standalone independent sample and therefore, the temporal aspect of anomalous sub-sequences may be lost.[8,23] This is also the case in FDM event analysis, where parameter values exceeding certain static thresholds are flagged as exceedances without necessarily considering their context.[6] Moreover, in many anomaly detection applications the multivariate time series is converted into a univariate series or a high dimensional vector thereby causing some information to be lost. In some cases of monitoring, the time-series is only compared to reference thresholds (such as in exceedance detection) or data within the same time-series losing out on potential insights that can be gained from similar data available. Finally, the existing anomaly detection techniques deal with data in which no relationship is assumed among the data points which is not necessarily true for flight parameters at each point. The methodology developed in this paper aims to address some of these limitations and is described in the subsequent section.

III. Methodology and Implementation

Unlike other applications of data mining or anomaly detection, aviation data is typically not labeled. This means that there is no knowledge a-priori as to which flight records (if any) are actually anomalous. Therefore, the anomaly detection algorithm needs to be semi-supervised or unsupervised. Also, there is no universal definition for an anomaly in this context. In the methodology presented in this paper, instantaneous anomalies are those instants that deviate from the nominal operations as identified by an anomaly detection algorithm. The chief assumption in this type of application is that majority of the data contains nominal operations which is a reasonable assumption in this case because of the extremely low accident and incident rates per million flight hours.[1] With these requirements in mind, the data obtained from GA flights is analyzed in a general anomaly detection framework shown in Figure 1. Various components of this framework as utilized in this paper are described further.

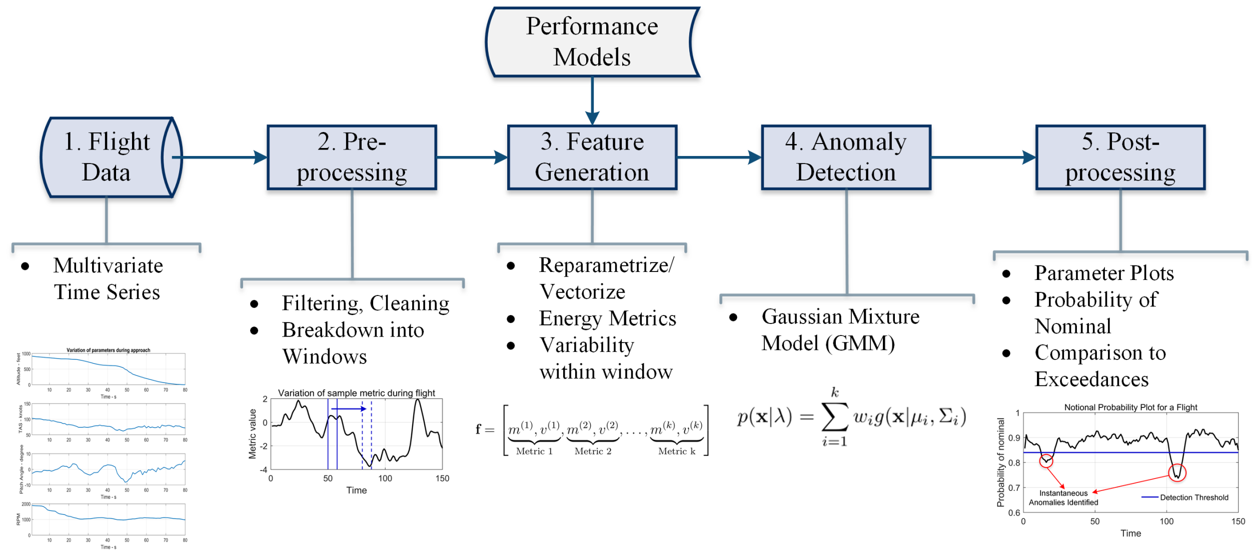


Figure 1. General framework for anomaly detection

A. Flight Data

In any anomaly detection problem it is important to understand the nature of the data captured (from flight data recorder in this case) as it has an impact on the techniques that can be used to analyze the data. Digital Flight Data Recorder (DFDR) data typically contains different channels that record discrete, continuous, and categorical data at a specified frequency (e.g., once per one second interval). Therefore, in analysis, the data obtained is used as a multi-variate time series for flight records, which are typically of different duration. The number of parameters in this multi-variate time series can be as low as 20 - 30 in GA operations to a number as high as thousands in commercial operations.[25–27] As this methodology is intended for application in a retrospective setting, data collected from routine GA operations (such as FOQA data) can be used.

B. Pre-processing

The raw data obtained from flights is initially pre-processed to obtain data suitable for anomaly detection. Sensor noise that may be present in the original data is smoothed using a simple moving average filter. Care is taken to not over-smooth the data as it would result in the loss of finer features in the data. In previous work on identification of instantaneous anomalies, each data sample is treated independently. This results in the loss of temporal nature of data obtained from flight records. It is of interest to capture the variation of key metrics before and after the sample under consideration. Therefore, in this methodology sliding window-based techniques are used for identifying instantaneous anomalies. These type of techniques have been used previously on univariate time series.[21,24,28] It involves sliding a window of a particular duration across the length of the time series (see Figure 2 for notional example).

For each instant, the data contained within the entire window is utilized rather than just the instant under consideration. As noted in literature, the main advantage of window-based techniques is that they need only 1 intuitive parameter (the window length) as opposed to others that can require 3-7 parameters.[24] There are various factors that can affect the choice of length of the window such as total duration of the time series, typical response time of the system, computational concerns etc. Due to all these factors, there is no consensus in the literature as to the appropriate length of the window.[28] In this work, five seconds (two seconds before and after the current point except at the ends) will be used as the window length. An important distinction between existing window-based techniques in other domains and the application in this paper is that window-based anomaly detection typically tries to find anomalous windows in a time series based on the data in that time series itself. On the other hand, the current method aims to also leverage the additional information available from the rest of the data set.

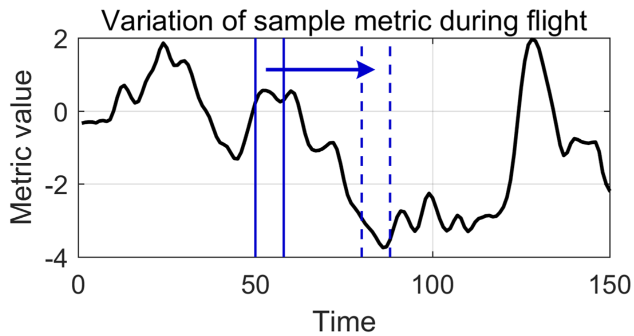


Figure 2. Notional depiction of sliding window across a metric

C. Feature Generation

One of the most important steps in the anomaly detection framework is the generation of *features* for the algorithm. Usually, these features are derived from the manipulation of the raw data collected. This can be as simple as using the raw data directly, to calculating new metrics using a combination of the recorded data and external information. The performance of the algorithm in identifying appropriate anomalies can depend heavily on the types of features used.[13]

In previous work (Puranik et al.[29,30]), it was demonstrated that energy-based metrics such as those quantifying the energy state, rates of change of energy, and their margins and deviations are appropriate metrics that correspond to safety margins and safe operations in GA and can be readily estimated from recorded data. The use of energy-based metrics to identify flight-level anomalies using a clustering algorithm was also shown in previous work.[19] Therefore, in the present work, energy-based metrics are utilized as features for identification of instantaneous anomalies. A summary of implemented metrics, their formulas, and data required for their computation is presented in Table 2 in the appendix. It it noted that since this analysis is not restricted to a particular phase of flight, those metrics that correspond to deviations from nominal profiles in certain phases of flight cannot be considered. The values of these metrics within a window is used to generate a corresponding vector (called feature vector) for each data point in the time series.

In a window-based techniques, each window is a unit of analysis. In order to obtain features of each window, the values of energy metrics at the current point along with their variability within the window is calculated. This allows the temporal aspect of instantaneous anomalies to be preserved while detecting anomalies. A similar window-based approach was used by Amidan and Ferryman,[21] however, their approach consists of fitting a linear regression model to each window and identifying anomalies based on the mathematical signature generated by the coefficients of the regression. In the current approach, the original value of the energy metric is retained as one of the features and additional dimensions corresponding to the variability of the metric within the window are generated. Specifically, the range of the metric values (maximum value minus minimum value) within the window is used as a measure of its variability. Thus, for example, if the window size is five seconds, and there are k metrics being used, then the feature vector for that window will contain $2 \times k$ dimensions (the value of each metric (m) at that point and its variability(v) within the window). Equation 1 shows the feature vector for each window which is then subsequently used in anomaly detection.

$$\mathbf{f} = \left[\underbrace{m^{(1)}, v^{(1)}}_{\text{Metric 1}}, \underbrace{m^{(2)}, v^{(2)}}_{\text{Metric 2}}, \ldots, \underbrace{m^{(k)}, v^{(k)}}_{\text{Metric k}} \right] \tag{1}$$

Each feature vector is thus generated by concatenating the value of all the energy metrics at that point along with its variability within the sliding window. Thus, each point in the time series is transformed into a corresponding feature vector which can then be used for anomaly detection.

D. Anomaly Detection

The next step is to use the features generated for each point within an anomaly detection algorithm. It is noted previously that most anomaly detection techniques for instantaneous anomalies did not deal with multivariate data explicitly and converted it into univariate data prior to analysis. However, it is desired that algorithms deal with multivariate data directly be used. Therefore, a Gaussian Mixture Model (GMM) is used for clustering and anomaly detection. The GMM can cluster normal operations together and help identify anomalies along with their probability based on the data set. One of the main advantages of using a GMM is that it does operate directly on the multivariate series and does not transform it into a univariate series. The other advantage is that GMM allows multiple standard operations to simultaneously exist. This is very important for algorithms that are not specific to a particular phase of flight as the variation of parameters and energy metrics during two different phases of flight is expected to be quite different.

A GMM is a parametric probability density function represented as a weighted sum of Gaussian component densities. Each component in the mixture is a type of standard observation or behavior of the system (example, one component for each phase of flight). The number of components, k, in the GMM determines number of sub-populations or clusters. The relation between predictor variables (or features) is captured in the form of a covariance matrix Σ. If each member of the population (in this case each feature vector) is an m-dimensional vector, then the GMM with k components and a covariance matrix Σ is given by:

$$p(\mathbf{x}|\lambda) = \sum_{i=1}^{k} w_i g(\mathbf{x}|\mu_i, \Sigma_i) \tag{2}$$

where $g(\mathbf{x}|\mu_i, \Sigma_i)$ indicates each of the components of the mixture model which is a multivariate gaussian model and w_i indicates the weighing of the component. The trained GMM is completely defined by the three parameters (w_i, μ_i, Σ_i) and k – the number of components. The parameters of the GMM are typically obtained via an Expectation-Maximization (EM) algorithm using the data set available.[31] This technique is utilized in the work presented in this paper. However, there are a few other important decisions that need to be made regarding the nature of the model before the EM algorithm can be used. These are – the nature of covariance matrix (full or diagonal), whether parameters among gaussian components are shared or not, and the number of components k. Due to the large computational cost of full covariance matrices, diagonal covariance matrix is used and the parameters are not shared among different gaussian components. Finally, the number of components can be set based on prior knowledge or it can be obtained using statistical metrics. Information theoretic metrics such as Akaike Information Criterion (AIC)[32] or Bayesian Information Criterion (BIC) have been used to identify the optimal number of components k.[23] These information theoretic metrics try to provide a balance between model complexity and overfitting. On the other hand, many internal clustering validation measures (such as Calinski-Harabasz index,[33] silhouette criterion, Davies Bouldin index,[34] etc.) rely on information within the data to provide a measure of the goodness of a clustering structure. They are typically based on two important criteria – compactness and separation. Objects within same clusters (or components) are closely related or similar to each other (compactness) and how distinct clusters (or components) are from each other.[35] In the present methodology, the Calinski-Harabasz (CH) index has been used to determine the optimal number of components. The number of components is progressively increased and the C-H index is measured. The number of components with the highest value of the index is chosen.

The advantage of using GMM for clustering is that it can provide statistical inferences about the underlying distributions. Therefore, once the required GMM has been trained using the existing data, it can be used to detect outliers or anomalies among the dataset. Using the values of the parameters obtained for the GMM (w_i, μ_i, Σ_i) the posterior probabilities of any component p for an observation $\mathbf{x}$ can then be calculated as:

$$p(\mathbf{x} \in p) = g(\mathbf{x}|\mu_p, \Sigma_p) \tag{3}$$

The estimated probability density function for each observation is then obtained as a sum over all components of the component density at that observation times the component probability.

E. Post-processing

Using the estimated probability density function of each observation, a profile of the probability density over the entire duration of the flight as shown in Figure 3 can then be constructed. Using appropriate thresholds for the probability enables identification of anomalous sub-sequences or instantaneous anomalies. This threshold can be varied to obtain different number of instantaneous anomalies. The safety analyst can then decide this threshold based on the trade-off between workload and missed detection. Once instantaneous anomalies are identified, plots of variation of flight parameters and energy metrics can be used to visualize and understand the reason for identification of this anomaly. This flexibility enables the analyst to focus attention on a limited number of important anomalies as opposed to a large untenable data set. The identified anomalies can also be compared against traditional exceedance events such as those defined in Table 1 in the appendix.

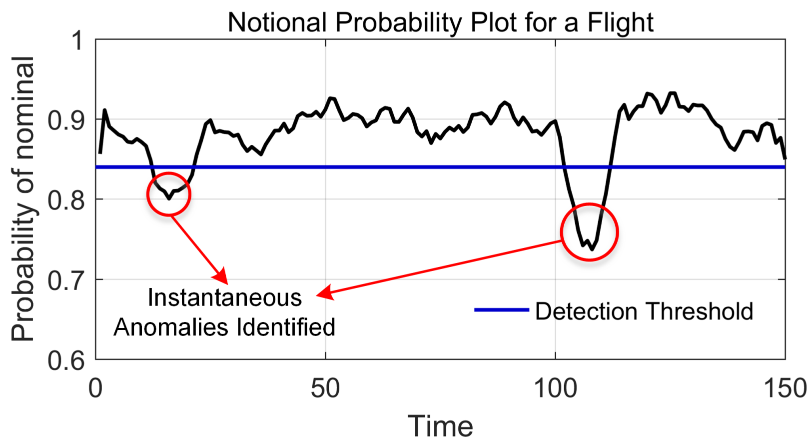


Figure 3. Notional depiction of probability density at each point during a flight record and the detection threshold

IV. Results

The methodology outlined previously in section III is implemented on a set of actual flight data and the results are discussed in this section. The data set used in this paper consists of one thousand flight data records collected from routine training flights on a Cessna 172S aircraft. The flight parameters are recorded using a Garmin G1000 at a frequency of one hertz. The parameters included in this study are continuous parameters related to atmospheric data (airspeed, wind speed, pressure altitude etc.), attitude (roll, pitch, yaw etc.), engine data (RPM, exhaust gas temperatures, fuel flow etc.), Global Positioning System (GPS) information, and others. The data set is pre-processed to generate the window features and evaluate energy metrics for each point in the flight data record. While this methodology does not make any assumptions about the phase of flight, in the current implementation, results from the approach and landing phase are focused upon due to the computational cost. This phase of flight has been recognized as the most important from the point of view of improving safety as maximum accidents have occurred in this phase.[36,37]

Prior to using the GMM for anomaly detection, the number of components for the GMM are identified using the C-H index as noted previously. The GMM is trained with steadily increasing number of components and the C-H index is measured for each trained model. The model with components that gives the highest C-H index is the one displaying the best internal clustering structure and is chosen for this application. Figure 4 shows the variation of C-H index with number of mixture components. Based on this sensitivity analysis, the number of components chosen for the GMM is five. The mean and variance of each gaussian component and the mixing probabilities for the trained GMM with five components is then used in further analysis. These parameters are then utilized to calculate the posterior probability density of being nominal at each point in each flight using Equation 3. The detection threshold for anomalies can be varied depending on the trade off between missed detection and excessive analysis workload. Since the data is obtained from routine operations, there is no ‘ground truth’ available to compare anomalies, however, it is expected that the number of anomalies (if any) will be a very small fraction of the total data. This trade off is examined

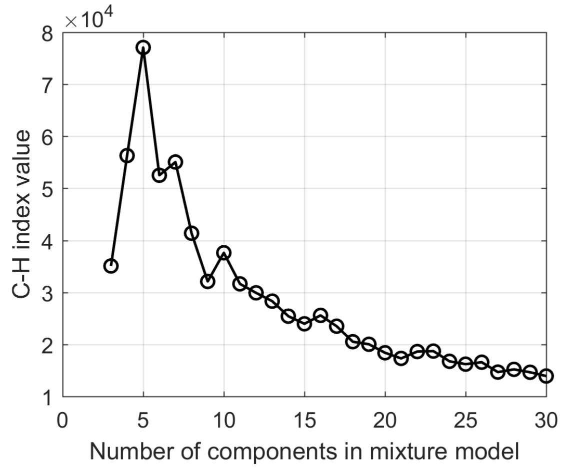


Figure 4. Variation of CH index with number of components

further in the comparison section (B).

A. Instantaneous Anomalies Identified

In this section, an example of a top instantaneous anomaly during approach and landing identified from the data set is elaborated. The detection threshold for the algorithm is set at 0.5%. Figure 5 shows the variation of the probability density as a function of the distance remaining to the runway for a flight record with an instantaneous anomaly. The probability density is depicted as a natural logarithm of the actual density due to the low values typically observed.[23] The dark and light grey bands represent the probability density at those locations for 50% and 95% of the flight data available and the solid black line represents the probability density of the flight record under consideration. Various detection thresholds are shown in different colors. The identified instantaneous anomaly is detected under all thresholds as it has a precipitous drop in probability density in the region of anomaly. The anomaly can be further analyzed by visualizing the energy metrics and flight parameters in this window and comparing them to all other flight records as well as exceedance events.

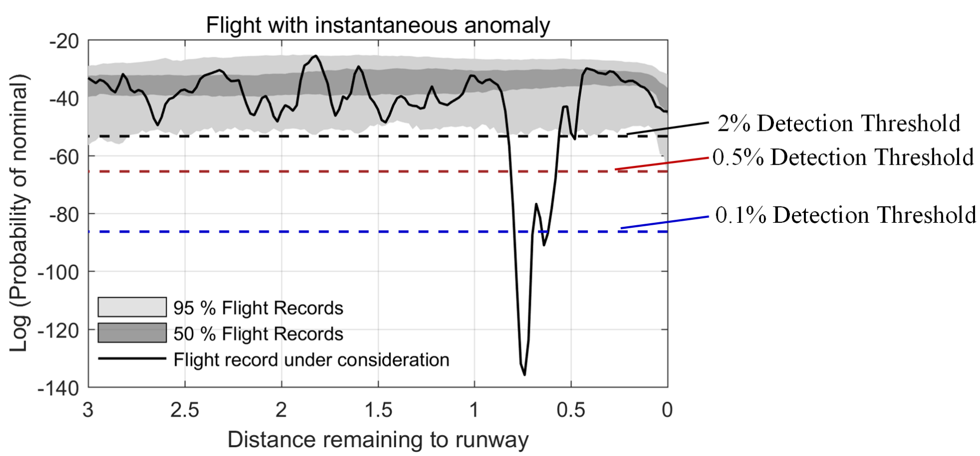


Figure 5. Probability of nominal operations as a function of distance remaining to runway

Figure 6 shows the variation of key energy metrics in the approach and landing phase for 50% of the flight records (light grey band), 95% of the flight records (dark grey band), and the flight record with the identified anomaly (solid black line). As clearly seen from the figure, the variation of the metrics is within nominal values for the most part, except at the location of the identified instantaneous anomaly window (outlined with dotted blue lines). At the beginning of the anomalous window, the total energy and kinetic energy of the flight is too low and therefore, an attempt is made to increase the energy as evidenced by the spike in the specific total energy rate metric. This also results in a very low thrust margin in the anomalous window. It is worth investigating the variation of raw flight data parameters in this window to understand it further

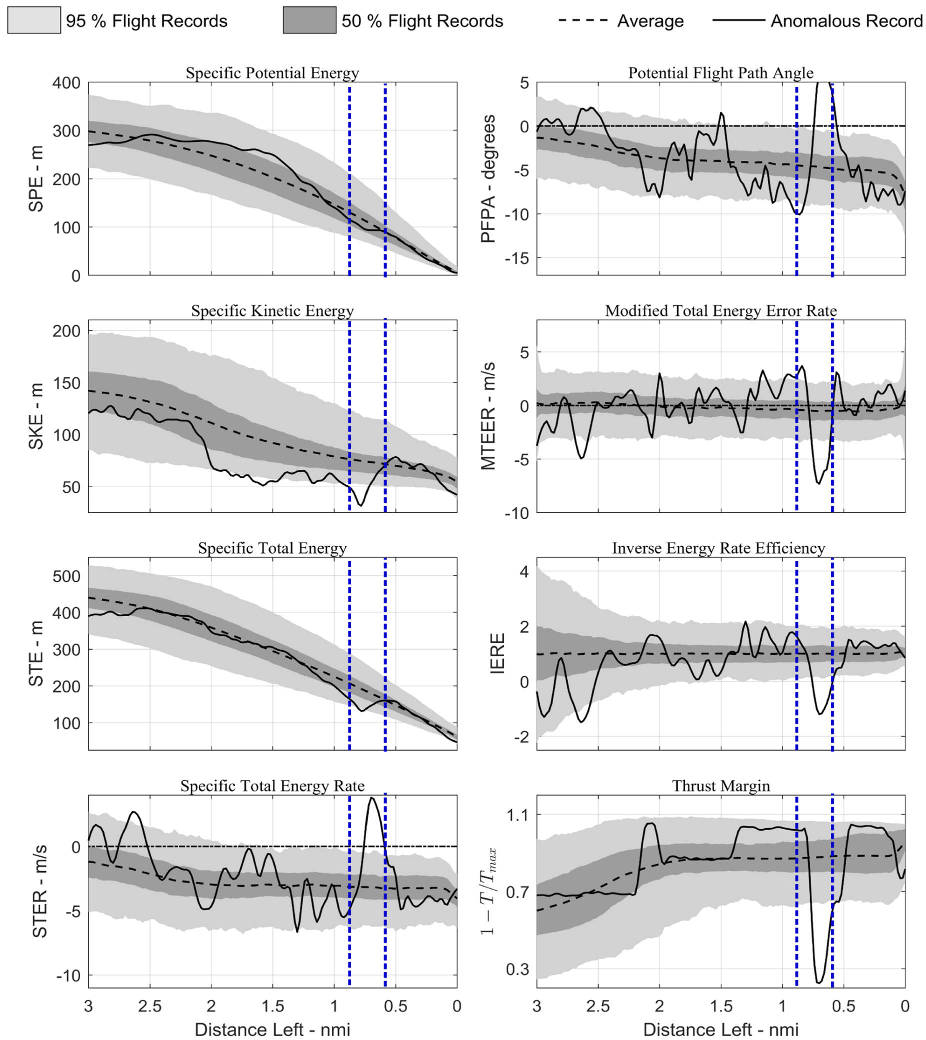


Figure 6. Variation of energy metrics as a function of distance remaining to runway

Figure 7 shows the variation of certain flight parameters during approach and landing phase for all flight records and the one with the anomalous window. From the variation of the true airspeed, it is evident that the aircraft approached at a very low airspeed and therefore, RPM was increased during the anomalous window to increase it. This caused the variation of energy metrics to be significantly different from nominal operations and identified the window as anomalous. This parameter variation has a semblance of an unstabilized approach and therefore, the identified anomaly could warrant further inspection by an expert. It is worth noting that this anomaly was identified without any prior input and matched exceedance events. Further detailed comparison with existing exceedance events is presented in the following section.

B. Comparison with Traditional Methods

Traditionally, unsafe events are identified using exceedance detection. This consists of examining the flight record for specific occurrences of parameters exceeding pre-defined thresholds.[6] An *exceedance* is the deviation of a single parameter beyond an established threshold. An *event* is defined by one or more parameter exceedances that take place concurrently over a specified period of time. Some examples of events during

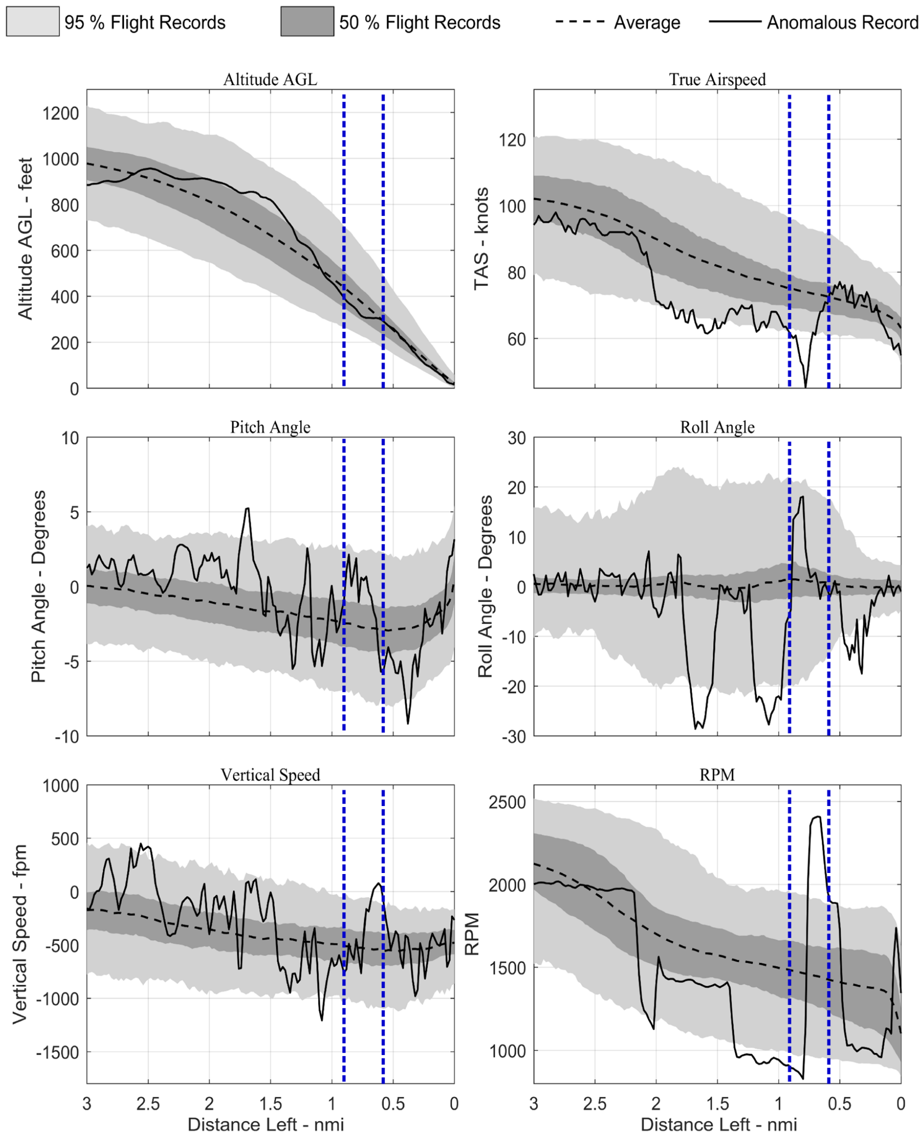
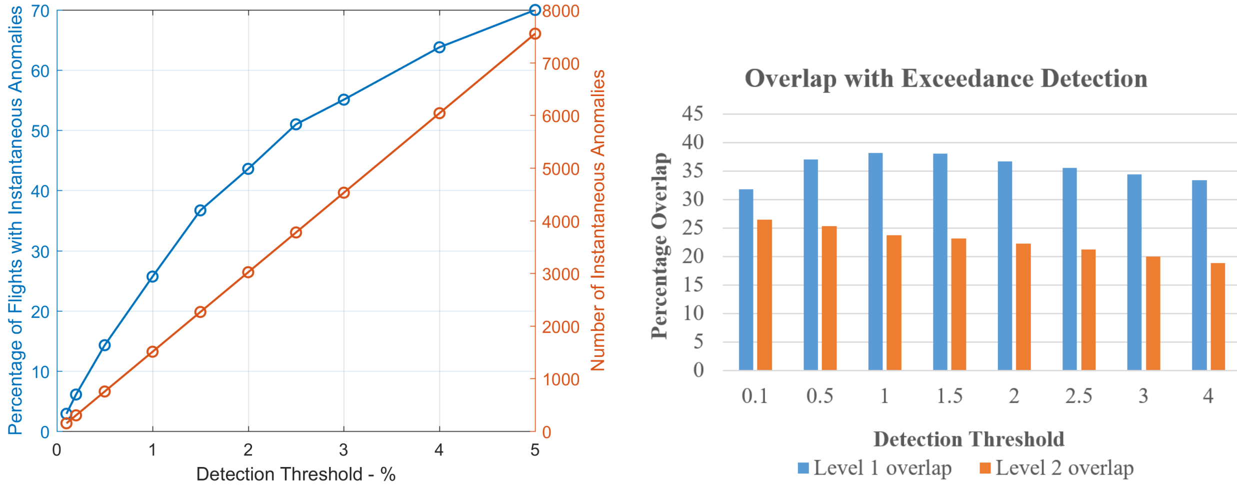


Figure 7. Variation of flight parameters as a function of distance remaining to runway

approach and landing are shown in Table 1 in the appendix. Exceedance detection classifies safety events based on severity of exceedance as level 1, level 2 etc. Flight records with certain number of exceedance events can then be further examined. The thresholds used in exceedance detection can be subjective and vary from aircraft-to-aircraft.[38]

In the anomaly detection problem using retrospective analysis of flight data, there is usually no 'ground truth' available as to what constitutes an actual anomaly. Therefore, it becomes hard to evaluate the performance of these algorithms without extensive subject matter expert review. Therefore, it is of value to compare how the identified instantaneous anomalies compare with traditional exceedance detection techniques as this would indicate the ability of the algorithms to identify known unsafe events. Figure 8 shows the comparison of identified instantaneous anomalies with exceedance events presented in Table 1 at different thresholds for anomaly detection. Figure 8(a) plots the variation of number of anomalies identified with the detection threshold as well as the proportion of flights containing instantaneous anomalous windows. The number of anomalies provides an indication of the total number of instants identified at the detection threshold. On the other hand, the proportion of flights containing anomalous windows is also important because multiple instantaneous anomalies may be present within the same flight record. Figure 8(b) plots the level of agreement between the identified instantaneous anomalies and defined exceedance events (both level 1 and level 2). An agreement is assumed if the identified anomaly contains any of the exceedance events listed in Table 1.



(a)

(b)

Figure 8. Number of instantaneous anomalies obtained and comparison of instantaneous anomalies with exceedance detection

As expected, the number of anomalies and proportion of flights with anomalies increases as the detection threshold is increased. However, as seen in the figure, detection thresholds as low as 5% result in anomalies identified in a majority of flight records. This seems to indicate that the detection threshold should be much lower than that (closer to 1%). The overlap between the anomalous instants identified and the exceedance events outlined in Table 1 is shown in Figure 8(b). The agreement between anomalies obtained from the flight data and the exceedance events outlined in Table 1 is higher for lower detection thresholds and goes on decreasing as the threshold is increased. The reason for this is that as the threshold is increased, increasingly benign events may also get captured as anomalous by the algorithm. One of the reasons high overlap is not observed is because all the exceedance events in the table cannot be easily verified (for example – the flap position is typically not annotated in GA FDR. Therefore all the exceedance events dependent on knowing it cannot be evaluated).

It is noted that improving and calibrating the algorithm to overlap more with the exceedance detection results will affect the algorithm's ability to identify currently known unsafe or anomalous events. However, it will not necessarily improve the algorithm's ability to detect potentially significant events that are not included in the available set of pre-defined events (which is one of the aims of using data mining and anomaly detection techniques).

V. Conclusions and Future Work

In this paper, a novel method for identifying instantaneous anomalies in GA flight data was demonstrated. This method utilized energy-based metrics as features in an anomaly detection framework. A sliding-window based pre-processing technique is formulated to ensure temporal aspect of features is captured. A mixture of gaussian models is trained using the available flight data for cluster analysis and outlier detection. Multivariate series are explicitly treated in GMM and it also allows for multiple standard operations which explicitly addresses some of the limitations identified previously. An example of an instantaneous anomaly identified by the methodology is presented and a comparison of the anomalies identified by the methodology with traditional exceedance events is presented.

The main advantage of using these methods is that the expert review process is cut down due to the specific windows identified. They use the data from multiple standard operations and not just static defined events. While the identified anomalies are merely abnormalities in the metrics of interest, they need to be reviewed further in order to fully understand the reason behind their occurrence and whether they are actually events that warrant further inspection and analysis. The methodology will be fine-tuned further to include results from other phases of flight. Due to the nature of instantaneous anomalies, the computational time required for identifying them increases greatly with increased number of flight records. Therefore, efficient methods of ensuring the scalability of the techniques will be explored in future work.

Acknowledgments

This work was sponsored by the Federal Aviation Administration under Project No. 5 of the FAA Center of Excellence for General Aviation, PEGASAS (Partnership to Enhance General Aviation Safety Accessibility and Sustainability), via grant No. 12-C-GA-GIT-004 to the Georgia Institute of Technology. The authors are grateful to FAA project technical monitor Mr. Michael Vu and project partners. The information in this research does not constitute FAA Flight Standards or FAA Aircraft Certification policy. The authors would like to thank Evan Harrison, Sanggyu Min, and Dr. Imon Chakraborty of the Aerospace Systems Design Lab for their valuable feedback and comments.

A. List of Exceedance Events

Table 1. Sample exceedances set during approach and landing for a Cessna 172 aircraft (Higgins et al.[39])

No	Event	Level 1	Level 2
1	V_{NE} (Never Exceed Velocity)	158 knots	163 knots
2	Vertical g Load	3.0	3.8
3	Vertical g Load - Min	-1.0	-1.52
4	Oil Temperature - Max	–	245 F
5	Oil Pressure - Min	–	20 psi
6	Oil Pressure - Max	–	115 psi
7	RPM - Max	2700 RPM or more for $\geq$ 1s	2700 RPM or more for > 5s
8	Cylinder Head Temperature - Max		500 F
9	Fuel Quantity - Min	8 gal.	5 gal.
10	Bank Angle	60°	$\geq$ 65°
11	Pitch Attitude (positive)	30°	$\geq$ 35°
12	Pitch Attitude (negative)	-30°	$\leq$ -35°
13	Vertical Speed Magnitude Below 1000 AGL	$\geq$ 800 fpm	$\geq$ 1000 fpm
14	Airspeed at or below 200 feet AGL High Speed Full Flaps	66 knots for 2s	71 knots for 2s
15	Airspeed at or below 200 feet AGL High Speed Zero Flaps	75 knots for 2s	80 knots for 2s
16	Airspeed at or below 200 feet AGL Low Speed Full Flaps	60 knots for 2s.	$\leq$56 knots for 1s
17	Airspeed at or below 200 feet AGL Low Speed Zero Flaps	69 knots for 2s	$\leq$65 knots for 2s
18	Extended Centerline deviation at 200 feet AGL	2°	3°
19	Glide angle High (Too steep) at 200 feet AGL	4°	5°
20	Glide angle Low (Too shallow) at 200 feet AGL	2°	1°
21	Bank Angle at or below 200 feet AGL	20°	25°
22	Pitch Attitude at Touchdown (High)	10.5°	12°
23	Pitch Attitude at Touchdown (Low)	3°	1°
24	Airspeed at Touchdown (High - Full Flap)	55 knots	60 knots
25	Airspeed at Touchdown (High - No Flap)	63 knots	68 knots

B. Summary of Utilized Energy Metrics

Table 2. Summary of implemented energy metrics and data required for computation

Metric	Formula	*Can be estimated using:* Flight Data	Flight Data + Ref. Profile	Flight Data + Perf. Model
Specific Total Energy	$h + V^2/2g$	✓	✓	✓
Specific Potential Energy	h	✓	✓	✓
Specific Kinetic Energy	$V^2/2g$	✓	✓	✓
Specific Total Energy Rate	$\dot{h} + V\dot{V}/g = (T - D)V/W$	✓	✓	✓
Specific Potential Energy Rate	$\dot{h} = V \sin \gamma$	✓	✓	✓
Specific Kinetic Energy Rate	$V\dot{V}/g$	✓	✓	✓
Potential Flight Path Angle	$\gamma + \dot{V}/g$	✓	✓	✓
Energy Rate Distribution	$sign(\frac{SKER}{SPER}) \times \exp(-\lvert\frac{SKER}{SPER}\rvert)$	✓	✓	✓
Specific Total Energy Error	$h_{act} - h_{ref} + (V_{act}^2 - V_{ref}^2)/2g$	×	✓	×
Specific Potential Energy Error	$h_{act} - h_{ref}$	×	✓	×
Specific Kinetic Energy Error	$(V_{act}^2 - V_{ref}^2)/2g$	×	✓	×
Normalized Specific Energy Error	$((STE)_{act} - (STE)_{ref})/(STE)_{tol}$	×	✓	×
Specific Total Energy Error Rate	$sign(STEE) \times \delta(STEE)/\delta t$	×	✓	×
Inverse Energy Rate Efficiency	$V_{act}(T - D)/V_{red}W(\gamma_{ref} + \dot{V_{ref}}/g)$	×	✓	×
Max. Potential Flight Path Angle	$T_{max} - D/W$	×	×	✓
Min. Potential Flight Path Angle	$T_{idle} - D/W$	×	×	✓
Thrust Margin	$1 - T/T_{max}$	×	×	✓
Energy Rate Margin	$(W(\gamma_a + \dot{V}_a/g))/(T_{max} - D)$	×	×	✓
Energy Rate Demand	$W(\gamma_c + \dot{V}_c/g)/(T_{idle} - D)$	×	×	✓

References

[1] "National Transportation Safety Board Website (Retrieved 10/2016)," url: `http://www.ntsb.gov/investigations/data/Pages/AviationDataStats.aspx`.

[2] "Federal Aviaition Administration Aerospace Forecasts Fiscal Years 2016-2036. (Retrieved 3/2017)," url: `https://www.faa.gov/data_research/aviation/aerospace_forecasts/media/FY2016-36_FAA_Aerospace_Forecast.pdf`.

[3] Logan, T. J., "Error Prevention as Developed in Airlines," *International Journal of Radiation Oncology*Biology*Physics*, Vol. 71, No. 1, 2008, pp. S178–S181, doi:10.1016/j.ijrobp.2007.09.040.

[4] "Federal Aviation Administration - Aviation Safety Information Analysis and Sharing (ASIAS)," url: `http://www.asias.faa.gov`.

[5] "General Aviaition Joint Steering Committee (GAJSC)," url: `http://www.gajsc.org/`.

[6] "Advisory Circular, 120-82 - Flight Operational Quality Assurance," April 2004, url: `https://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentID/23227`.

[7] "Civil Aviation Authority CAA, Flight Data Monitoring CAP 739, Second Edition," June 2013, ISBN 978-0-11792-840-4. url: `https://publicapps.caa.co.uk/docs/33/CAP739.pdf`.

[8] Iverson, D. L., "Inductive System Health Monitoring," Tech. rep., National Aeronautics and Space Administration, 2004, url: `https://ntrs.nasa.gov/search.jsp?R=20040068062`.

[9] Das, S., Matthews, B. L., Srivastava, A. N., and Oza, N. C., "Multiple Kernel Learning for Heterogeneous Anomaly Detection: Algorithm and Aviation Safety Case Study," *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2010, pp. 47–56, doi:10.1145/1835804.1835813.

[10] Matthews, B., Das, S., Bhaduri, K., Das, K., Martin, R., and Oza, N., "Discovering Anomalous Aviation Safety Events Using Scalable Data Mining Algorithms," *Journal of Aerospace Information Systems*, Vol. 10, No. 10, Oct. 2013, pp. 467–475.

[11] Li, L., Das, S., John Hansman, R., Palacios, R., and Srivastava, A. N., "Analysis of Flight Data Using Clustering Techniques for Detecting Abnormal Operations," *Journal of Aerospace Information Systems*, Vol. 12, No. 9, Sept. 2015, pp. 587–598, doi:10.2514/1.I010329.

[12] Melnyk, I., Matthews, B., Valizadegan, H., Banerjee, A., and Oza, N., "Vector Autoregressive Model-Based Anomaly Detection in Aviation Systems," *Journal of Aerospace Information Systems*, 2016, pp. 161–173.

[13] Chandola, V., Banerjee, A., and Kumar, V., "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, Vol. 41, No. 3, 2009, doi:10.1145/1541880.1541882.

[14] "24th Joseph T. Nall Report - Aircraft Owners and Pilots Association," 2012, url: https://www.aopa.org/-/media/files/aopa/home/pilot-resources/safety-and-proficiency/accident-analysis/nall-report/15-fn-0022-1-24th-nall-v6.pdf.

[15] "Federal Aviation Administration, 14 CFR §23.3 Airplane Categories," url: https://www.ecfr.gov/cgi-bin/text-idx?node=pt14.1.23.

[16] Budalakoti, S., Srivastava, A. N., and Otey, M. E., "Anomaly Detection and Diagnosis Algorithms for Discrete Symbol Sequences with Aapplications to Airline Safety," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, Vol. 39, No. 1, 2009, pp. 101–113, doi:10.1109/TSMCC.2008.2007248.

[17] Gorinevsky, D., Matthews, B., and Martin, R., "Aircraft anomaly detection using performance models trained on fleet data," *Intelligent Data Understanding (CIDU), 2012 Conference on*, IEEE, 2012, pp. 17–23, doi:10.1109/CIDU.2012.6382196.

[18] Smart, E. and Brown, D., "A Two-Phase Method of Detecting Abnormalities in Aircraft Flight Data and Ranking their Impact on Individual Flights," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13, No. 3, 2012, pp. 1253–1265, doi:10.1109/TITS.2012.2188391.

[19] Puranik, T. G., Jimenez, H., and Mavris, D. N., "Utilizing Energy Metrics and Clustering Techniques to Identify Anomalous General Aviation Operations," *AIAA SciTech Forum*, Jan. 2017, Paper No. AIAA 2017-0789, doi:10.2514/6.2017-0789.

[20] Bay, S. D. and Schwabacher, M., "Mining Distance-based Outliers in Near Linear Time with Randomization and a Simple Pruning Rule," *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2003, pp. 29–38, doi:10.1145/956750.956758.

[21] Amidan, B. G. and Ferryman, T. A., "APMS SVD Methodology and Implementation," Tech. rep., U.S. Department of Energy PNWD-3026, 2000, doi:10.2172/753847.

[22] Mugtussids, I. B., *Flight Data Processing Techniques to Identify Unusual Events*, Ph.D. thesis, Virginia Polytechnic Institute and State University, 2000, url: http://hdl.handle.net/10919/28095.

[23] Li, L., Hansman, R. J., Palacios, R., and Welsch, R., "Anomaly detection via a Gaussian Mixture Model for flight operation and safety monitoring," *Transportation Research Part C: Emerging Technologies*, Vol. 64, 2016, pp. 45–57.

[24] Keogh, E., Lin, J., and Fu, A., "HOT SAX: Efficiently finding the most unusual time series subsequence," *Fifth IEEE International Conference on Data Mining (ICDM'05)*, IEEE, 2005, p. 8 pp., doi:10.1109/ICDM.2005.79.

[25] "Federal Aviation Administration, 14 CFR §121.344 Digital Flight Data Recorders for Transport Category Airplanes, 2011," url: https://www.ecfr.gov/cgi-bin/text-idx?tpl=/ecfrbrowse/Title14/14cfr121_main_02.tpl.

[26] "Federal Aviation Administration, 14 CFR §91.609 Appendix E, Flight Data Recorders and Cockpit Voice Recorders," url: https://www.ecfr.gov/cgi-bin/text-idx?node=14:2.0.1.3.10.

[27] Campbell, N., "Flight Data Analysis - An Airline Perspective," url: http://www.asasi.org/papers/2003/Flight%20Data%20Analysis_Campbell_V2.pdf.

[28] Chandola, V., *Anomaly Detection for Symbolic Sequences and Tme Series Data*, Ph.D. thesis, University of Minnesota, 2009, url: http://hdl.handle.net/11299/56597.

[29] Puranik, T., Harrison, E., Min, S., Jimenez, H., and Mavris, D., "Energy-Based Metrics for General Aviation Flight Data Record Analysis," *16th AIAA Aviation Technology, Integration, and Operations Conference*, 2016, Paper No. AIAA 2016-3915, doi:10.2514/6.2016-3915.

[30] Puranik, T. G., Jimenez, H., and Mavris, D. N., "Energy Based Metrics for Safety Analysis of General Aviation Operations," *Journal of Aircraft*, Accepted – March 2017.

[31] Dempster, A. P., Laird, N. M., and Rubin, D. B., "Maximum Likelihood from Incomplete Data Via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39, No. 1, 1977, pp. 1–38, doi:10.2307/2984875.

[32] Schwarz, G., "Estimating the dimension of a model," *The annals of statistics*, Vol. 6, No. 2, 1978, pp. 461–464.

[33] Caliński, T. and Harabasz, J., "A dendrite method for cluster analysis," *Communications in Statistics*, Vol. 3, No. 1, Jan. 1974, pp. 1–27, doi:10.1080/03610927408827101.

[34] Davies, D. L. and Bouldin, D. W., "A Cluster Separation Measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-1, No. 2, 1979, pp. 224–227, doi:10.1109/TPAMI.1979.4766909.

[35] Liu, Y., Li, Z., Xiong, H., Gao, X., and Wu, J., "Understanding of internal clustering validation measures," *2010 IEEE 10th International Conference on Data Mining (ICDM)*, IEEE, 2010, pp. 911–916, doi:10.1109/ICDM.2010.35.

[36] "National Transportation Safety Board Accident Statistics," Retrieved 04/2017. url: https://www.ntsb.gov/investigations/data/Pages/AviationDataStats2014.aspx.

[37] "Federal Aviaition Administration Fact Sheet - General Aviation Safety," Retrieved - 04/2017. url: https://www.faa.gov/news/fact_sheets/news_story.cfm?newsId=16774.

[38] Fala, N. and Marais, K., "Detecting Safety Events During Approach in General Aviation Operations," *AIAA Aviation*, American Institute of Aeronautics and Astronautics, June 2016, Paper No. AIAA-2016-3914, doi:10.2514/6.2016-3914.

[39] Higgins, J., Clachar, S., and Hennselek, K., "Flight Data Monitoring (FDM)/Flight Operational Quality Assurance (FOQA) for General Aviation," U.S. Department of Transportation - Federal Aviation Administration Technical Report DOT/FAA/AR-13/21, December 2013.