Utilizing Energy Metrics and Clustering Techniques to Identify Anomalous General Aviation Operations

Tejas G. Puranik[*], Hernando Jimenez[†], and Dimitri N. Mavris[‡]

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

Among operations in the General Aviation community, one of the most important objectives is to improve safety across all flight regimes. Flight data monitoring or Flight Operations Quality Assurance programs have percolated in the General Aviation sector with the aim of improving safety by analyzing and evaluating flight data. 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. The use of data mining techniques for safety analysis, incident examination, and fault detection is gaining traction in the aviation community. In this paper, we have presented a generic methodology for identifying anomalous flight data records from General Aviation operations using energy-based metrics and clustering techniques. The sensitivity of this methodology to various key parameters is quantified using different experiments. A demonstration of this methodology on a set of actual flight data records as well as simulated flight data is presented highlighting its future potential.

I. Introduction

One of the most important objectives among operations in the General Aviation (GA) community is to improve safety across all flight regimes. In the past, accidents have been the primary triggers for identifying problems and developing mitigation strategies.[1] However, the industry is now moving towards a more pro-active 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. According to the National Transportation Safety Board (NTSB),[2] the number of total accidents per million flight hours in GA is an order of magnitude higher than that of commercial operations. With air traffic expected to grow tremendously over the next decade and GA set to receive a significant impetus, improving safety is of paramount interest to the aviation industry. Flight Data Monitoring (FDM) or Flight Operations Quality Assurance (FOQA)[3] programs, which are well-established in commercial operations have percolated into GA with the aim of improving operational safety. Typical FOQA programs involve a continuous cycle of 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.

Data mining techniques for safety analysis, incident examination, and fault detection have garnered increased interest in the aviation community in recent years. Current practice in FOQA is chiefly underpinned by a-priori definition of safety events known as 'exceedances'.[3] This method 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 pre-defined events. Data mining approaches have the potential of revealing safety events of interest as emergent artefacts 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. A large portion of the existing literature is dedicated to commercial aviation despite the fact that GA operations have historically had considerably greater accident and incident rates.[2]

Additionally, transport category airplanes have certain minimum requirements related to Digital Flight Data Recorder Systems (DFDRS) set by the Federal Aviation Administration (FAA) for U.S carriers.[4]

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

[†]Research Faculty, Daniel Guggenheim School of Aerospace Engineering, AIAA Senior Member

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

Therefore, data mining techniques have lent themselves well to data obtained from FOQA programs in commercial operations due to the wealth of information available. In GA operations that have been considered in this work (small airplanes less than 12,500 lbs and 10 seats),[5] a large number of parameters may not be recorded due to limited capability of their data collection hardware. Therefore, using recorded flight data with quantitative aircraft performance models (such as those developed in Harrison et al.[6] and Min et al.[7]) are a key enabler for evaluating a number of energy based metrics for safety analysis. These energy metrics provide a measurable quantification of the energy state of the aircraft and can be viewed as an objective currency to evaluate various safety-critical conditions across a large fleet of aircraft with minimal amount of recorded parameters. In previous work (Puranik et al.[8,9]), the authors have demonstrated the implementation of energy metrics and associated challenges for GA application.

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. Also, there is no set definition for what an anomaly in this context is. Therefore, unsupervised or semi-supervised algorithms need to be used to identify anomalies. Identifying anomalies as outliers of a clustering algorithm is useful way of identifying anomalies as it allows the possibility of multiple standard patterns. There are a number of ways in literature by which the "anomalousness" (or anomaly score) of an outlier can be quantified. Campos et al.[10] have provided a review of some of these scores using different data sets to quantify their relative performance. In most cases, the performance of an anomaly score is dependent on the type of data. In the absence of labeled data, the best way to proceed is to choose methods that perform well across a broad range of anomaly scores.

Considering the above observations we articulate the needs motivating the work in this paper as follows. First, we present a methodology that utilizes energy metrics in conjunction with clustering techniques to identify anomalous flight data records in the approach and landing phase. Second, we propose to test this methodology using real world data with the intention of achieving the objectives listed below:

1. Demonstrate the use of energy metrics (or a subset of energy metrics) to identify anomalous GA flight data records using clustering techniques

2. Quantify the sensitivity of the results obtained to various key parameters and steps in the methodology

3. Identify anomalous GA flight data records using different subsets of energy metrics based on data/models required for computation of these metrics (refer Appendix A)

4. Demonstrate the performance of the method across a broad range of anomaly scores in order to ensure that the most consistent results are obtained

The following sections will build the methodology and discuss results with the aim of addressing the objectives mentioned above. The rest of the paper is organized as follows: Section II provides a review of existing data mining techniques applied in the aviation domain. Section III contains the outline and description of the key elements of the methodology used in this work. Section IV discusses various experiments carried out to fine tune the methodology and insights obtained. Section V presents the results of the application of this methodology and Section VI draws conclusions and outlines future avenues of work being pursued.

II. Review of Existing Work

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.[11–15] In the broader 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*".[16] In the aviation safety domain, two main types of anomalies are interesting to the analyst - instantaneous and flight-level. Instantaneous anomalies refer to specific instances within a particular flight that seem to be abnormal compared to the rest of the flight, whereas flight-level anomalies refer to those flights with abnormal data patterns that persist over a period of time compared to other flights. In this work, we have focused on identification of flight-level anomalies. Instantaneous anomalies will be treated separately in future work.

Chandola et al.[16] have provided a comprehensive survey of anomaly detection that covers techniques applied across all domains. Liao et al.[17] have specifically surveyed clustering techniques pertaining to time series data and provided a taxonomy of techniques. Within the aviation domain, Gavrilovski et al.[18]

have surveyed data mining techniques and provided a review of published work on its application to flight data. They have also identified challenges and opportunities for its application to fixed and rotary wing GA applications and have also noted salient distinctions between application of these methods to commercial aviation against rotorcraft and fixed wing GA operations. Anomaly detection techniques used in literature can be broadly classified into two categories - supervised learning and unsupervised learning.

Supervised learning methods such as Inductive Monitoring System (IMS)[19] 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. SequenceMiner[20] has been shown to detect anomalies in discrete parameter sequences by learning from a model of normal switching. This technique detects flight-level anomalies but is limited to discrete data.

Some of the methods rely on developing an approximate model using flight data and detecting those flight records which deviate greatly from this model as outliers. For example, Chu et al.[21] have proposed an approach for detecting anomalies from aircraft cruise flight data using a model trained using historical data of a fleet of aircraft. Anomalies are detected as outliers that exceed the scatter caused by turbulence and the modeling error. Gorinevsky et al.[15] 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 (aerodynamically) clean configuration. Also, most of the anomalies detected are in the determination of aerodynamic or propulsion parameter estimates or gross weight. Melnyk et al.[14] have treated each multivariate time series using a vector auto-regressive exogenous model. Dissimilarity between two flights is measured as the residuals obtained by using the model of one flight on the data of another. Outliers are identified using Local Outlier Factor (LOF) which is a nearest neighbor based anomaly detection method. This method requires that a different model be built for each flight record being analyzed. This method also requires that pilot inputs be recorded which may not necessarily be the case for GA data since the required instrumentation is usually not present.

On the other hand, unsupervised learning methods do not rely on a training set and try to obtain anomalous records from a large data set using techniques such as clustering. Bay and Schwabacher[22] have described a method called Orca which uses Euclidean distance of a point to its k-nearest neighbors to gauge the anomalousness of each point. However, as with IMS, this method treats each point independently, therefore it loses the temporal aspect of abnormal flights. Das et. al.[11] have developed Multiple Kernel Anomaly Detection (MKAD) which applies a one-class support vector machine for anomaly detection. MKAD uses the normalized Longest Common Sub-sequence (nLCS) kernel which is useful for discrete data, but it results in loss of some finer features for continuous data (when it is discretized). Matthews et al.[12] have discussed and summarized the aviation knowledge discovery pipeline using various algorithms. Li et al.[13] have implemented ClusterAD, which uses cluster-based anomaly detection on pre-processed flight data parameters to identify abnormal operations.

III. Methodology

The methodology followed in this work is outlined in Figure 1. The various steps in this methodology correspond to steps in a general anomaly detection framework. The details of analyses performed in each step are provided in the following subsections. There are three main inputs to this methodology. The first input is the set of flight data records. This can be obtained using a Digital Flight Data Recorder such as Garmin G1000. Alternatively, a flight data record can also be simulated using a flight simulator. A typical flight data record contains multiple parameters related to different systems of the aircraft and the environmental conditions in which it is operating. These parameters are recorded at a specific frequency (e.g., once per one second interval). Figure 2 shows an example of part of a flight data record during approach and a subset of the parameters recorded. Each flight record is thus a multivariate time-series consisting of continuous or discrete parameters. For the purpose of this work, only continuous parameters have been considered.

In addition to flight data records, reference energy profiles for the approach and landing phase are also required for the methodology. In previous work (Puranik et al.[8]), we have demonstrated a data-driven

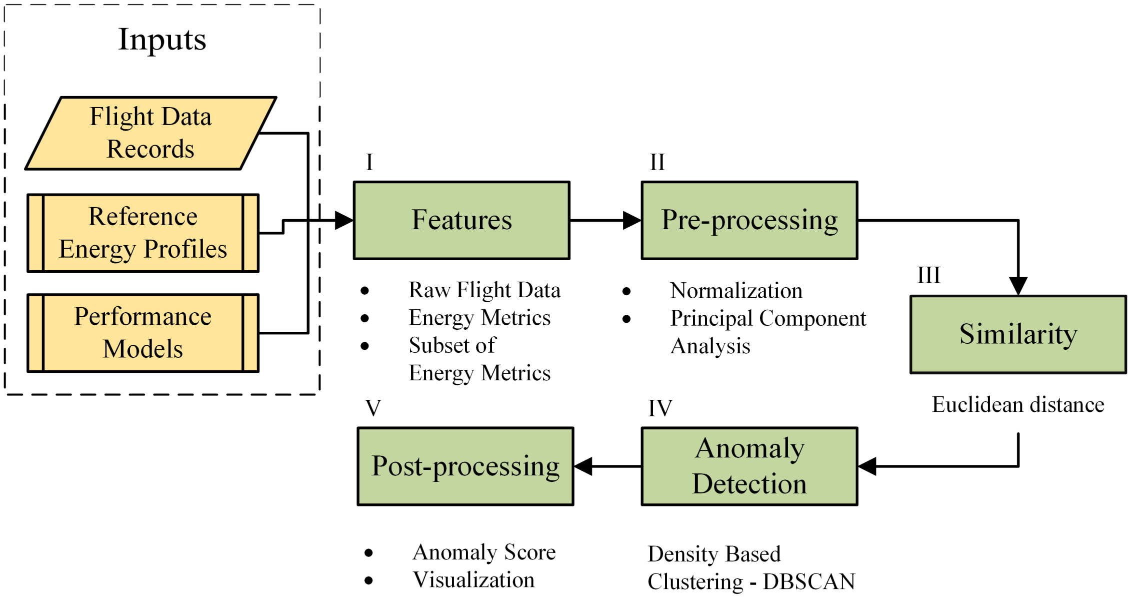


Figure 1. Methodology followed to identify anomalous records using energy metrics

approach to obtain reference energy profiles using actual flight data records. A reference profile, in this context, is the set of nominal values of a particular metric (for example specific potential energy) as a function of distance remaining to the runway threshold during approach and landing. In most cases, an actual value close to this reference profile is desirable. Aircraft performance models for the aerodynamic and propulsion characteristics make up the third and final set of inputs required. The models developed in Harrison et al.[6] and Min et al.[7] have been utilized in this work.

I. Features

The first step of the methodology is generation of a feature vector to be used for anomaly detection. These feature vectors are directly generated using the information contained in the flight data record and the reference energy profiles. The data obtained from the FDR is initially cleaned up by smoothing noisy data, filtering out records with incomplete data etc. In previous work, (Puranik et al.[9]) the authors demonstrated the implementation of a number of energy based metrics for safety analysis. Appendix A contains a table summarizing the implemented metrics, and data required for computing them. For a FDR, the value of each energy metric is calculated at all time steps where data has been recorded. This creates a time series for each energy metric similar to those shown in Figure 2. Approach and landing operations are often expressed in terms of distance to the runway threshold, rather than in terms of time,[13] so that flight data record parameters (and energy metrics) are sampled here according to ground-track distance to the runway threshold. Data of different flights thus become comparable since each flight parameter is sampled at fixed distance-based intervals.

In order to generate features for clustering, each flight data record needs to be represented as a vector. Therefore the energy metrics are transformed into high dimensional vectors for each flight record. Consider a data set of n flight records, focusing on k energy metrics and discretizing the approach phase in p small segments. Then, the feature vector for each flight record can be represented by concatenating the contribution from each metric as:

$$\mathbf{f} = [\underbrace{m_1^{(1)}, m_2^{(1)}, .., m_p^{(1)}}_{\text{Metric 1}}, \underbrace{m_1^{(2)}, m_2^{(2)}, .., m_p^{(2)}}_{\text{Metric 2}}, ..., \underbrace{m_1^{(k)}, m_2^{(k)}, .., m_p^{(k)}}_{\text{Metric k}}] \tag{1}$$

where m_j^i is the value of the i^{th} metric at the j^{th} location in the approach and landing.

Therefore all flights are represented by a feature vector of the same length and each element of the vector can be compared to the corresponding element of another vector. Another option for the feature vector is

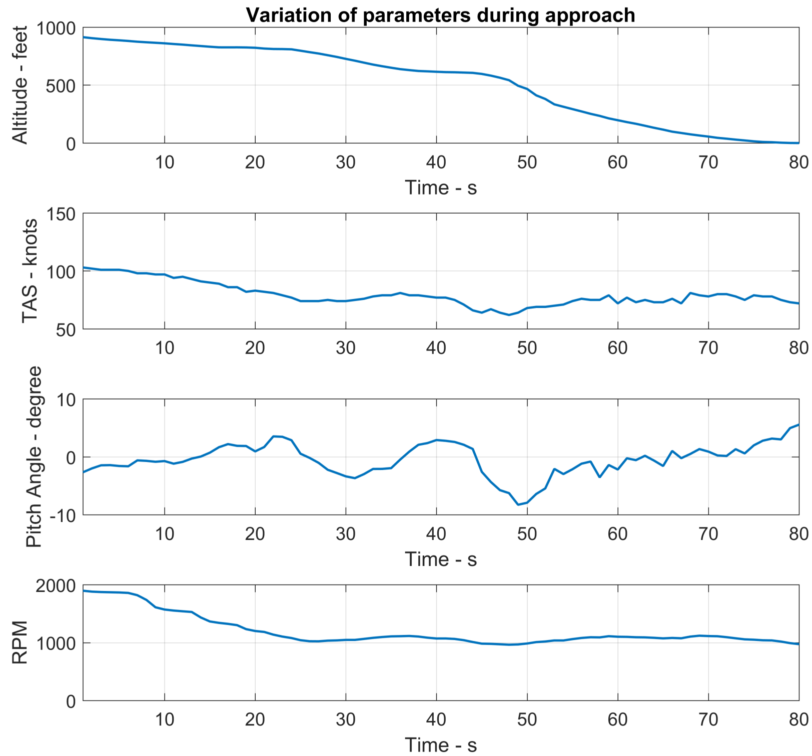


Figure 2. Example of flight data record parameters

to use only a subset of the energy metrics defined in Appendix A which are the most relevant for identifying anomalous flights. In a similar manner, a feature vector (FV) for each flight record can also be generated using the raw flight data parameters. These choices for the feature vector and their implications on the quality of results are explored later in the paper.

II. Pre-processing

The second step in the methodology consists of pre-processing the feature vectors obtained in Step I. Even though all the feature vectors have the same length, different metrics have different magnitudes. Therefore, each metric is normalized such that it has a zero mean and unit variance. This is achieved using z-score normalization. Let $m_{j,1}^{(i)}, m_{j,2}^{(i)}, ...m_{j,n}^{(i)}$ be the values of a particular metric i for all n flights at a particular distance j from the runway. The z-score normalized values are given by:

$$m_{j,1}^{(i)} = \frac{m_{j,1}^{(i)} - \sum_{p=1}^{n} m_{j,p}^{(i)}}{std(m_{j,1...n}^{(i)})} \tag{2}$$

The feature vectors obtained can contain hundreds of dimensions. However, while identifying outliers and clusters, the variability is typically embedded in a smaller number of dimensions. Principal Component Analysis (PCA) is a linear transformation that is used to transform data into a new orthogonal coordinate system.[23] The coordinates in the new system are ranked in descending order of the amount of embedded information (variance) they contain. Dimensionality reduction is achieved by only retaining the first few components that explain majority of the variance (in this method the number is chosen such that 99% of the variance is captured). If the reduced dimensional vector contains p elements, each flight record can now be represented as:

$$\mathbf{f'} = [v_1, v_2, ...v_p] \tag{3}$$

While the set of energy metrics is not as prohibitively large as the set of flight data parameters, it is still useful to reduce dimensionality of the problem. Also, since many of the metrics are manifestations of the potential, kinetic, and total mechanical energy in different forms, there is a possibility of strong correlation among these. PCA also yields basis vectors which result in reduction of this correlation between dimensions. However, this representation in reduced dimensions has the drawback that each element of the vector no longer has any physical interpretation as was the case earlier.

Since different energy metrics are concatenated to form the feature vector, PCA can also be performed metric-wise and the reduced dimensional representation of each metric can then be concatenated to form a reduced-dimensional feature vector. However, performing PCA metric-wise has the disadvantage that correlations between different metrics may not be eliminated. Both these approaches are explored later in this paper.

III. Similarity

Once the feature vectors are finalized, each flight data record needs to be compared against every other flight data record. This comparison is achieved using a similarity function (S) or dissimilarity function (D). In this approach, clustering requires a distance between any two points. Therefore, a Euclidean distance between the reduced dimensional feature vectors is used as the dissimilarity function. Let the feature vectors corresponding to two flight records be $\mathbf{v} = [v_1, v_2, ..., v_p]$ and $\mathbf{w} = [w_1, w_2, ..., w_p]$. Then the euclidean distance between $\mathbf{v}$ and $\mathbf{w}$ is given as:

$$D(v, w) = \sqrt{\sum_{i=1}^{p} (v_i - w_i)^2} \tag{4}$$

A distance function of this nature is, by its form, symmetric, i.e $D(v, w) = D(w, v)$. Many other similarity functions can be utilized here and a detailed discussion will be included in future work.

IV. Anomaly Detection

The fourth step of the methodology is the actual anomaly detection. There are various algorithms that can be used for anomaly detection. In this paper we have utilized density-based clustering (DBSCAN).[24] DBSCAN has the ability to automatically determine the number of clusters and also detect outliers (anomalies) based on a user specified threshold. One of the drawbacks of DBSCAN is that its performance can suffer if there are multiple clusters with varying densities. This will be addressed in experiment 1.

Given a set of points (flight data records), DBSCAN groups together instances that are closely packed together while marking points in low-density regions as outliers. A cluster forms when there are at least a minimum number of points (hereafter called $MinPts$) within a user specified threshold (hereafter called ϵ) of a given point. Clusters grow when additional points satisfy the density criterion specified by the algorithm until all the points have been allotted to a cluster or labeled as outliers.

There are two parameters that need to be supplied to DBSCAN - ϵ and $MinPts$. ϵ depends heavily on the similarity function used, normalization of data, and other factors. In many cases, rather than providing ϵ, its value is varied from the minimum distance observed among all flights in the data set to the maximum distance observed. Instead of ϵ, the user provides the percentage of flights that will be tagged as outliers. This number has a direct correlation with the value of ϵ but is more intuitive to the user of the methodology. $MinPts$ on the other hand, depends on the homogeneity of operations and how similar flights are to each other in terms of the features chosen in Step I. Typically, $MinPts$ has a less significant effect on the algorithm than ϵ if it is within a nominal range of values.[24] Experiment 1 will provide more details on the choice of this parameter.

V. Post-processing

The final step of the methodology is post-processing of the results obtained. The main aim of this step is to provide the user of this methodology with a list of anomalous or abnormal flights observed in the data set along with a quantification of how anomalous each flight record is. The anomalousness of a flight can be inspected in various ways. Qualitative assessment is done by visualizing the data records whereas quantitative assessment is done in terms of different types of anomaly scores.

Visualization

Visualization of the anomalous flight data record parameters compared to other flights is very useful to understand why the flight is being tagged as anomalous. Since the data has been standardized according to distance remaining, a distribution of the values of the parameters can be obtained. Similar to the visualization by Li et al.,[13] a visualization of the anomalous flight's parameters during approach along with the values of 50^{th} and 90^{th} percentile of all flight records are shown for comparison.

Anomaly Scores

While the algorithm used in Step IV provides a list of anomalous flights, it does not necessarily quantify how abnormal each outlier is. Various outlier scores can be used to quantify the anomalousness of each flight data record. These scores can be evaluated using the feature vector defined earlier in Step I and the similarity/dissimilarity functions discussed in Step III. Campos et al.[10] have provided a review of commonly used anomaly scores in outlier detection applications on unsupervised data. In the absence of ground truth about which flight records are anomalous, it is important to obtain different anomaly scores for the data set and compare them. From the perspective of the safety analyst or user of this methodology, flights that have consistently high anomaly scores would warrant further inspection.

In this methodology, we have examined a set of anomaly scores frequently used in the data mining community in order to quantify the anomalousness. Anomaly scores are typically classified into two categories - global and local, based on the number of nearest (from the point of view of the similarity function) neighbors utilized in calculating them. All scores typically work with a set of k neighbors of a particular data point and their properties. Local methods have a lower value of k than global methods. A mix of most common global and local scores are used in this method. They are enumerated below:

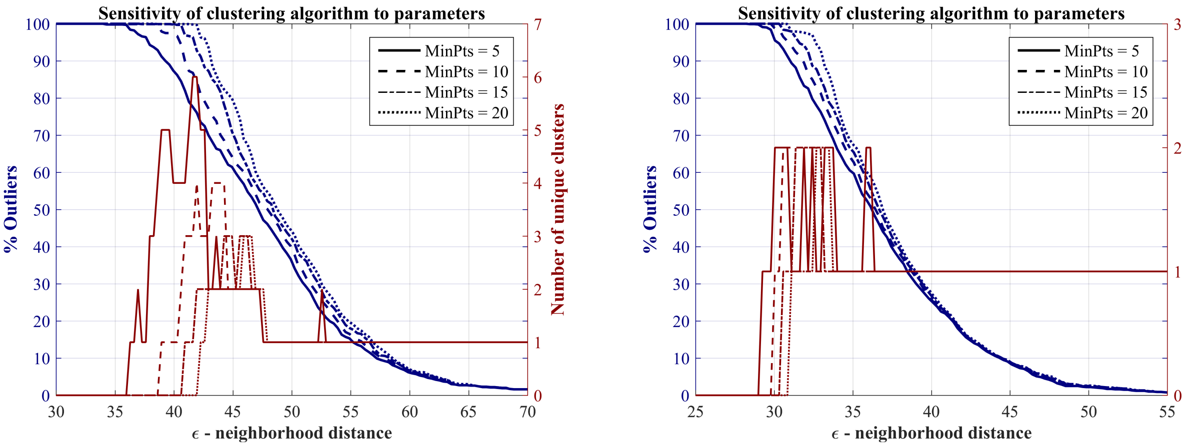
1. Average distance to k-nearest neighbors (kNN-avg)
2. Local Outlier Factor (LOF)[25]
3. Simplified Local Outlier Factor (LOF)[10]
4. Local Distance-based Outlier Factor (LDOF)[26]

IV. Experiments

For the purpose of demonstration of our methodology and conducting experiments we have utilized flight data from training flights on a Cessna C172S equipped with Garmin G1000 glass cockpit displays. Each data record contains information about aircraft state characteristics such as altitude, true airspeed, indicated airspeed, latitude, longitude, and engine RPM among others, collected at one second intervals. Over one thousand flight records were collected for this study, and properly de-identified with respect to operator, pilots onboard, date, and time of day, prior to any research efforts. All the records correspond to operations from the same airport, but for flights landing on different runways. For this study the authors have access to the identity of the airport, and can access runway information such as the latitude-longitude location of runway thresholds from publicly available sources, which coincide with latitude-longitude data in the flight data records. Using this data and aircraft performance models developed previously (Harrison et al.[6] and Min et al.[7]), various experiments are conducted to provide further insights into the working of the methodology.

Experiment 1: Tuning of Clustering Parameters

The first experiment carried out is tuning the parameters of the clustering algorithm. The main objective of this experiment is to determine appropriate values for the clustering parameters their overall effects on the result. As noted earlier, there are two main parameters in the DBSCAN algorithm - ϵ (the minimum neighborhood distance to form a cluster) and $MinPts$ (the minimum number of points needed in the vicinity of a point to start a cluster). However, the value of ϵ can be automatically chosen if the outlier significance level is specified (e.g 1% of flights are outliers). Therefore, it is important to identify the sensitivity of the algorithm to the choice of $MinPts$ for various feature vectors. In Expt. 3, different options of feature vectors have been discussed in Table 1. For the purpose of this experiment, two options for feature vectors have been considered for the purpose of demonstration - FV-1 (all raw parameters), FV-4 (all energy metrics). Figure 3 shows the percentage of flights identified as outliers along with the number of unique clusters at each setting for various combinations of ϵ and $MinPts$. The value of $MinPts$ is set at four discrete settings: [5, 10, 15, 20]. Typically in literature a default value of 5 is used.[13,24] On the other hand ϵ is varied between the minimum and maximum pairwise distance value observed in the current data set. Figure 3a shows the results of this experiment using raw flight data parameters (FV-1) whereas Figure 3b shows the results using all energy metrics (FV-4).



(a) Using Feature Vector 1 - Raw Parameters

(b) Using Feature Vector 2 - Energy Metrics

Figure 3. Sensitivity of algorithm to clustering parameters and number of unique clusters for two feature vector options

In both figures, we observe that, for all values of $MinPts$, as the value of ϵ increases from the minimum to the maximum value, the proportion of flights that are identified as outliers (blue curves) steadily decreases. This is expected as more points will be included in the clusters when the cluster radius is increased. The curves for different values of $MinPts$ eventually collapse almost into a single curve as ϵ is increased further. For the type of data dealt with in aviation safety, the proportion of anomalous or abnormal flights is expected to be very low (as seen from the low accident and incident rates per million flight hours).[2] Therefore, at the values of ϵ that correspond to low outlier percentages, any value of $MinPts$ from the set chosen gives almost equivalent results. This important observation leads to the conclusion that the value of $MinPts$ can be set to a default of 5 for the purpose of this work.

The second set of curves (orange) from Figure 3 corresponds to the number of unique clusters at each setting of ϵ and $MinPts$. The common trend in both figures is that as the value of ϵ increases, the number of clusters settles at one. However at lower values of ϵ, the use of raw flight data parameters (Figure 3a) shows a lot of variability as well as more number of unique clusters. As opposed to this, the use of the energy metrics (Figure 3b) results in either one or two clusters at low ϵ values and settles at a single cluster much faster. However, as noted earlier, at the outlier significance levels of interest there is only a single cluster present among flight data records for both types of feature vectors. It was noted earlier that DBSCAN performance can suffer if there are multiple clusters with varying densities. However, the results of Expt. 1 indicate that there is a single cluster and those drawbacks of DBSCAN will not be an issue for this application.

Experiment 2: Effect of Principal Component Analysis

In this experiment, the effect of using PCA on the feature vectors is quantified and the results obtained with and without PCA are compared. The same two set of feature vectors as the previous experiment are used in this experiment. During step II of the methodology, data is pre-processed to normalize and reduce dimensionality using PCA. PCA will result in some time savings as a smaller dimensional feature vector will be used each time the algorithm is executed. As mentioned earlier, data from the approach and landing phase is re-sampled based on distance remaining to the runway threshold. In this work, the flight data from up to three nautical miles from the runway threshold are divided into 151 distance-based snapshots. Prior to PCA, FV-1 contains 24 flight data parameters of interest recorded at each distance-based snapshot, whereas FV-4 contains 12 energy metrics evaluated at each distance-based snapshot. Therefore FV-1 has a total of 3624 dimensions and FV-4 has a total of 1812 dimensions. For both feature vector options, PCA is used to reduce dimensionality such that 99% of the variance is captured in the reduced dimensional space. This results in FV-1 being reduced to 387 dimensions and FV-4 being reduced to 216 dimensions.

However, one of the end-results of this methodology is a list of anomalous flights. Using PCA to reduce dimensionality should not result in a change in the end-result obtained. Therefore, the final list of anomalous flights obtained with and without PCA (for the same feature vector option) is compared at the 3% outlier level. The method yields the exact same set of flights with and without PCA for both feature vector options considered, thus indicating the effectiveness of PCA for reducing the dimensionality while maintaining the quality of results.

While PCA is a useful dimensionality reduction technique, it also provides additional information about the data being compressed. PCA transforms the data into the new coordinates by taking a linear combination of the distance-based snapshots. Each principal component is arranged in decreasing order of the variance of the data set captured. The coefficients of each distance-based sample within PCA are an indicator of its contribution to the overall variability. Averaging the absolute value of these coefficients at a particular distance-based sample for the entire feature vector provides an indication of the contribution of that location to the overall variability. This averaging can also be done on the coefficients of each individual energy metric at the same distance-based locations to obtain the contributions of these metrics to the overall variability. In this experiment, feature vector 4 containing all energy metrics is utilized.

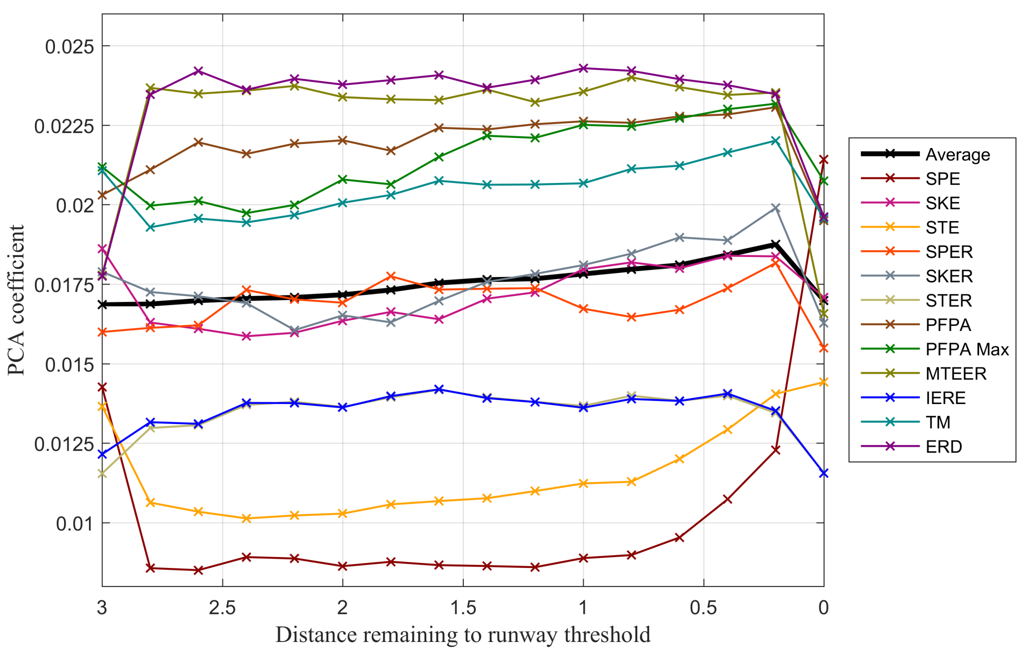


Figure 4. Average absolute value of PCA coefficients during approach and landing for FV-4 (all energy metrics)

Figure 4 shows the trend of the average absolute value of the coefficients when PCA is applied to the entire feature vector during approach and landing. This type of sensitivity analysis permits interpretations regarding the location-wise contribution of all metrics to the reduced dimensional vector and the relative contribution of each individual metric to the reduced dimensional vector. A relatively flat line for any metric (or the average) indicates that all locations along the approach contribute equally to the reduced dimensional vector. This trend is observed for most metrics as well as the average. On the other hand, if the slope of the line increases towards the end, it indicates that locations closer to the runway threshold contribute more to the reduced dimensional vector. This trend is observed for some of the metrics such as Specific Potential Energy (SPE), Specific Total Energy (STE), Thrust Margin (TM) etc.

The other interpretation obtained from this analysis is with regard to the relative magnitude of the coefficients of each metric. As is evident from the figure, some metrics such as Energy Rate Distribution (ERD), Thrust Margin (TM), and Modified Total Energy Error Rate (MTEER) have higher coefficients than others like Specific Potential Energy (SPE) or Specific Total Energy (STE). Therefore, the metrics with higher coefficients contribute more to the reduced dimensional vector than those with lower coefficients. This can have an important effect on which flight records are identified as anomalous when only a subset of the metrics are used in the feature vector.

Experiment 3: Feature Vector Options

As noted earlier in Section III, flight data records can be represented using different feature vectors. In this experiment, we have implemented the methodology using different sets of feature vectors in order to understand which flights are being identified as anomalous by the methodology in each case. Different anomaly scores using these feature vectors have also been calculated in each case to observe how separable the anomalous and normal flights are in the feature vector space. Table 1 outlines the different feature vector options and a description of how the features are computed.

The first feature vector option (FV-1) includes all parameters (24 in this case study) obtained from flight data records. This is similar to those used by others such as Li et al.[13] for commercial FOQA data. The first subset of energy metrics (FV-2) is all those metrics that can be obtained using combinations of parameters from the flight data record. These include kinematic metrics such as the kinetic, potential, and total energy and their rates. All metrics marked with a 'Yes' in the fourth column and 'No' in the last column of Table 2 (Appendix) are included in this subset. The second subset of energy metrics (FV-3) include all those metrics from FV-2 and those that require reference energy profiles as defined in earlier work.[8] With respect to Table 2, these are all rows that have 'Yes' in the fourth column. The final feature vector of energy metrics includes all metrics listed in Table 2. In addition to FV-3, this set includes those metrics that require an aircraft performance model in order to be evaluated. Therefore, the feature vectors are arranged in such a way that from FV-1 to FV-4, each feature vector subset requires additional data and/or models than the previous one in order to be evaluated.

Table 1. Feature vector options and description of contents

Name	Features	Description
FV-1	Raw Flight Parameters	Set of all flight parameters collected from flight data
FV-2	Energy Metrics Subset 1	Set of all energy metrics from Table 2 that can be obtained from basic flight data alone
FV-3	Energy Metrics Subset 2	Set of all energy metrics from Table 2 that can be obtained from basic flight data and reference energy profiles
FV-4	All Energy Metrics	Set of all energy metrics listed in Table 2

The results shown here contain a summary of anomalous flight data records identified by the described methodology at the 5% significance level using each of the feature vector options listed above in Table 1. This corresponds to roughly 50 anomalous flight data records. Figure 5 shows the number of anomalous flight data records that are common across different feature vector options and those that are unique to some feature vector combinations. Apart from the small number of unique flights identified as anomalous using only one of the various options, the main observation is that there is a list of 16 flight records that are identified as anomalous using all feature vector options. Further, 17 other flights are identified as anomalous using feature vectors 2,3,4 (different subsets of energy metrics) which are not captured by feature vector 1. Therefore, the different subsets of energy metrics have more consistency among themselves. On the other hand, feature vector 1 identifies 25 other flight records as anomalous that are not captured by any other feature vector option. A reason for this might be that FV-1 includes a number of raw parameters which do not affect the energy state of the aircraft.

The results of this experiment indicate that there is a large overlap in the anomalous flight records identified by different energy metrics. This is encouraging as some of the metrics can be computed using the most basic flight data captured. For airplanes that do not have sophisticated flight data capturing capability, this offers an attractive opportunity to use the described methodology to identify anomalies. On the other hand, a smaller agreement in the anomalous flights identified using raw parameters versus the other feature vectors is observed. Therefore, quantifying the anomaly score and visualizing it for these flight records will provide further insight into why certain flights are tagged as anomalous using only certain features.

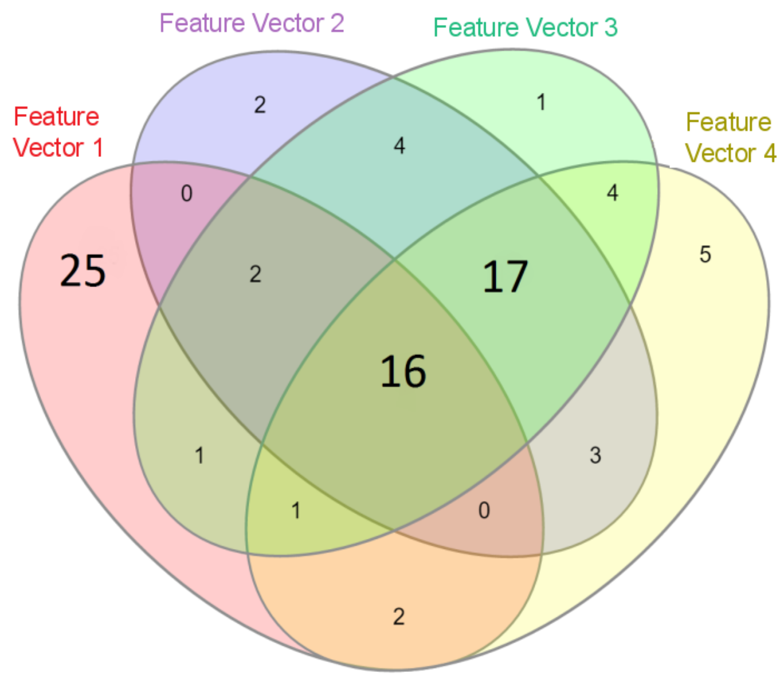


Figure 5. Venn diagram showing overlap of anomalous flights as classified by the different feature vector options

V. Results

In this section, we have described the results obtained from implementing the methodology on the data set introduced earlier. It is important for the methodology to be able to identify known abnormalities as well as emergent artefacts. Therefore, in this section, two subsections are presented which tackle each of these possible types of anomalous flight data records. Feature Vector 4 from Table 1 is chosen to identify anomalous records. The metrics are normalized using z-score and PCA is performed such that 99% of the variability is preserved in the reduced dimensional vectors. Euclidean distance is used as the similarity metric for comparison. In the clustering algorithm, $MinPts = 5$ and ϵ is chosen such that 5% of the flight records are tagged as anomalous.

For both the subsections contained in this section, a scatterplot matrix consisting of three different anomaly scores is plotted in Figure 6. In this figure, the blue dots represent flight records that were tagged as normal by the methodology. The red dots represent flight records which were tagged as anomalous at the 5% significance level. The upright green triangle represents the artificial flight data record introduced in the following subsections. The inverted green triangle represents a representative anomalous flight record from those identified at the 5% significance level. The distribution of scores indicates that most the flights having higher values across all anomaly scores are identified as anomalous by the methodology. While there is a high positive correlation between different anomaly scores, some spread is also observed and therefore, it is important to visualize all of them simultaneously while analyzing anomalies.

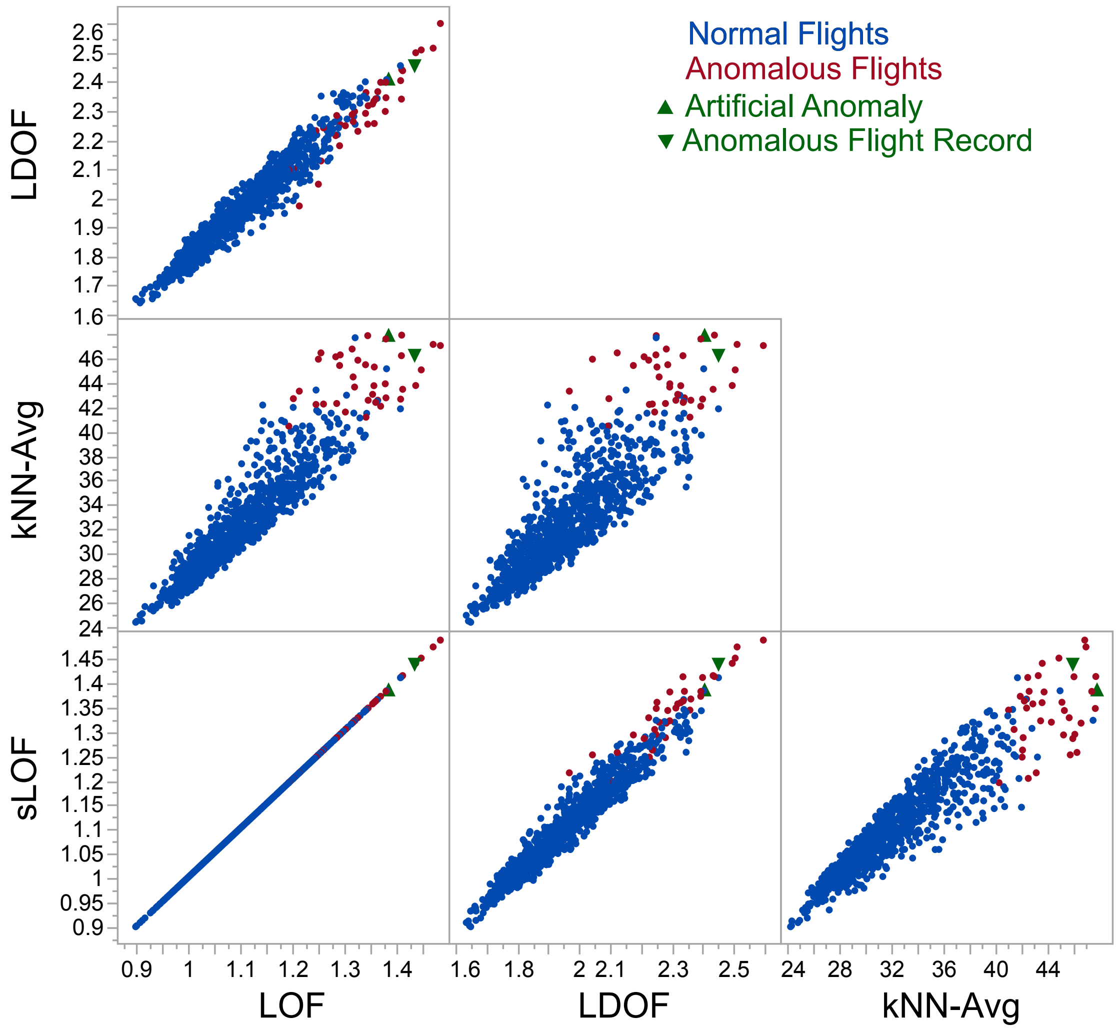


Figure 6. Visualization of various anomaly scores for set of flight records

A. Artificial Anomaly

Within a set of flight data records, flights that feature anomalous occurrences are not flagged or labeled as such. This makes it difficult to validate different methodologies for anomaly detection using such flight data records alone. Therefore, in order to facilitate such validation efforts, an artificial flight data record, generated using a flight simulation model, is included in the data set. The dynamics of the aircraft are simulated using a MATLAB/Simulink model which is connected to FlightGear Flight Simulator to allow visual rendering of the motion of the aircraft. Further details of the model can be found in Chakraborty et al.[27] The simulated approach and landing are flown by a private pilot instructed to deliberately fly an unstabilized 3 NM final approach with poor energy management. The anomaly scores of this record are visualized in Figure 6. A visualization of different energy metrics during this approach and landing, along with variation of flight parameters is shown in Figures 7,8. As is clearly evident, this approach and landing is anomalous and is correctly captured as having a high score (shown by upright green triangle) using this methodology. Therefore, this methodology is able to identify known abnormalities and place them in the higher regions of anomaly scores.

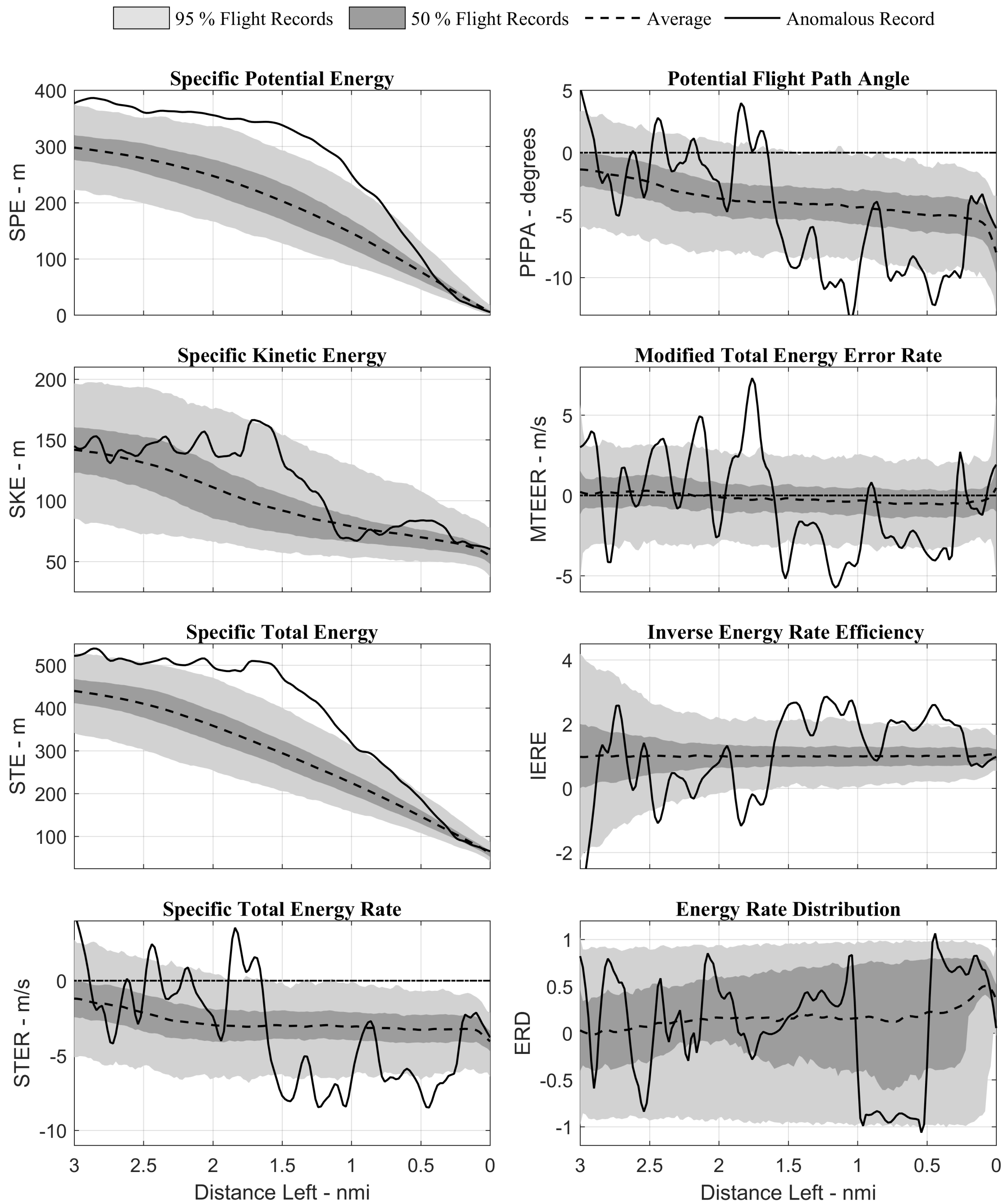


Figure 7. Visualization of energy metrics for artificial anomaly

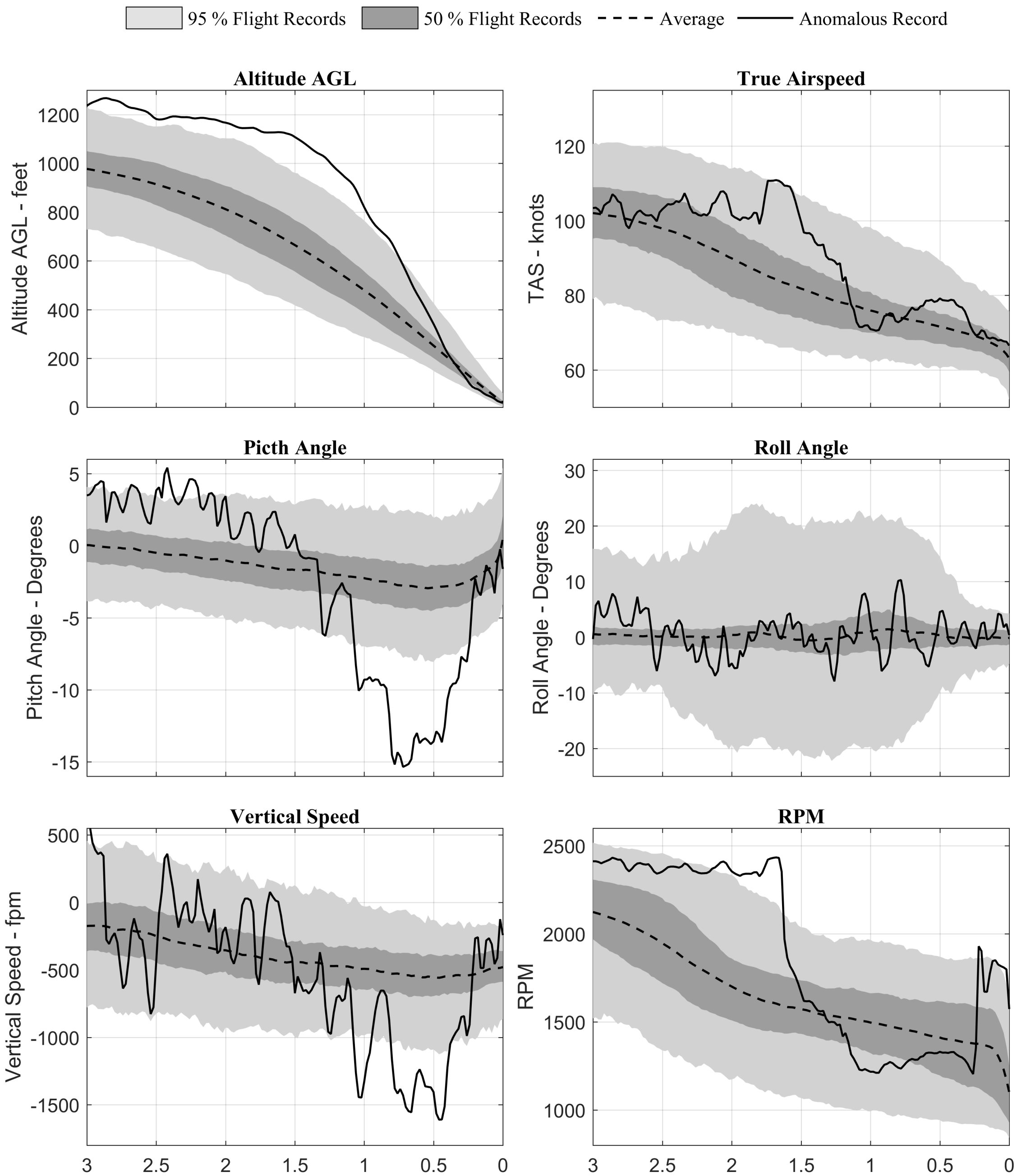


Figure 8. Visualization of a subset of flight data parameters metrics for artificial anomaly

B. Anomalies in Actual Flight Data

This methodology is now applied to the entire data set to identify anomalous flight data records and their scores. The list of flight records tagged as anomalous by the methodology are highlighted in red in Figure 6. The only information provided to the algorithm is the percentage of flights to be marked anomalous. Once the list of anomalous flight records is obtained, one of the top anomalous flights from the data set having a consistently high score for LOF, LDOF, and kNN is visualized in Figures 9, 10.

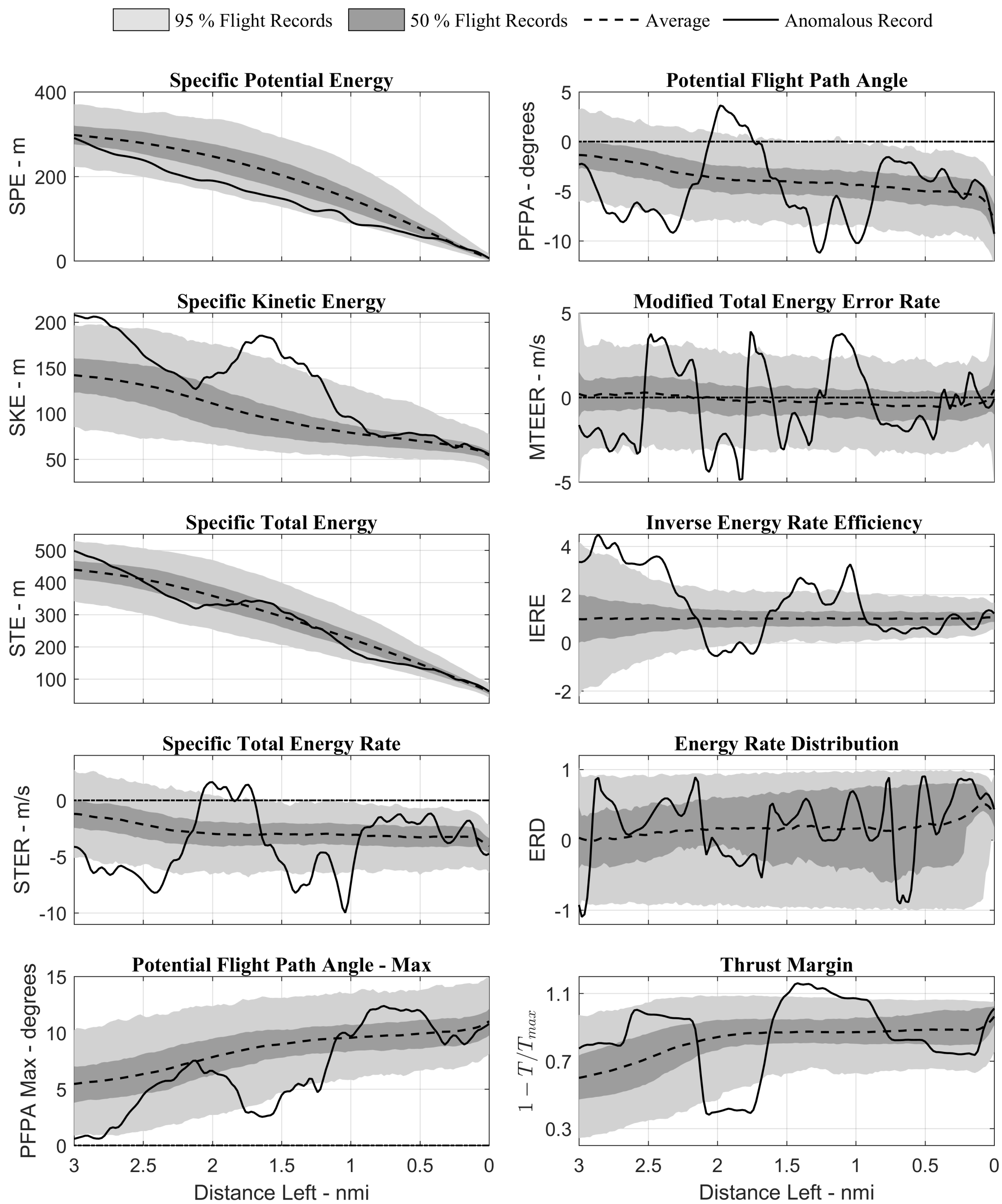


Figure 9. Visualization of energy metrics for a flight data record with low and fast approach

Based on the visualization of parameters and energy metrics, it can be classified as a low and fast approach and could be considered unsafe operations. The visualization of parameters shows that the RPM, fuel flow rate, exhaust gas temperature, and airspeed variation is outside of the bounds for majority of flight records.

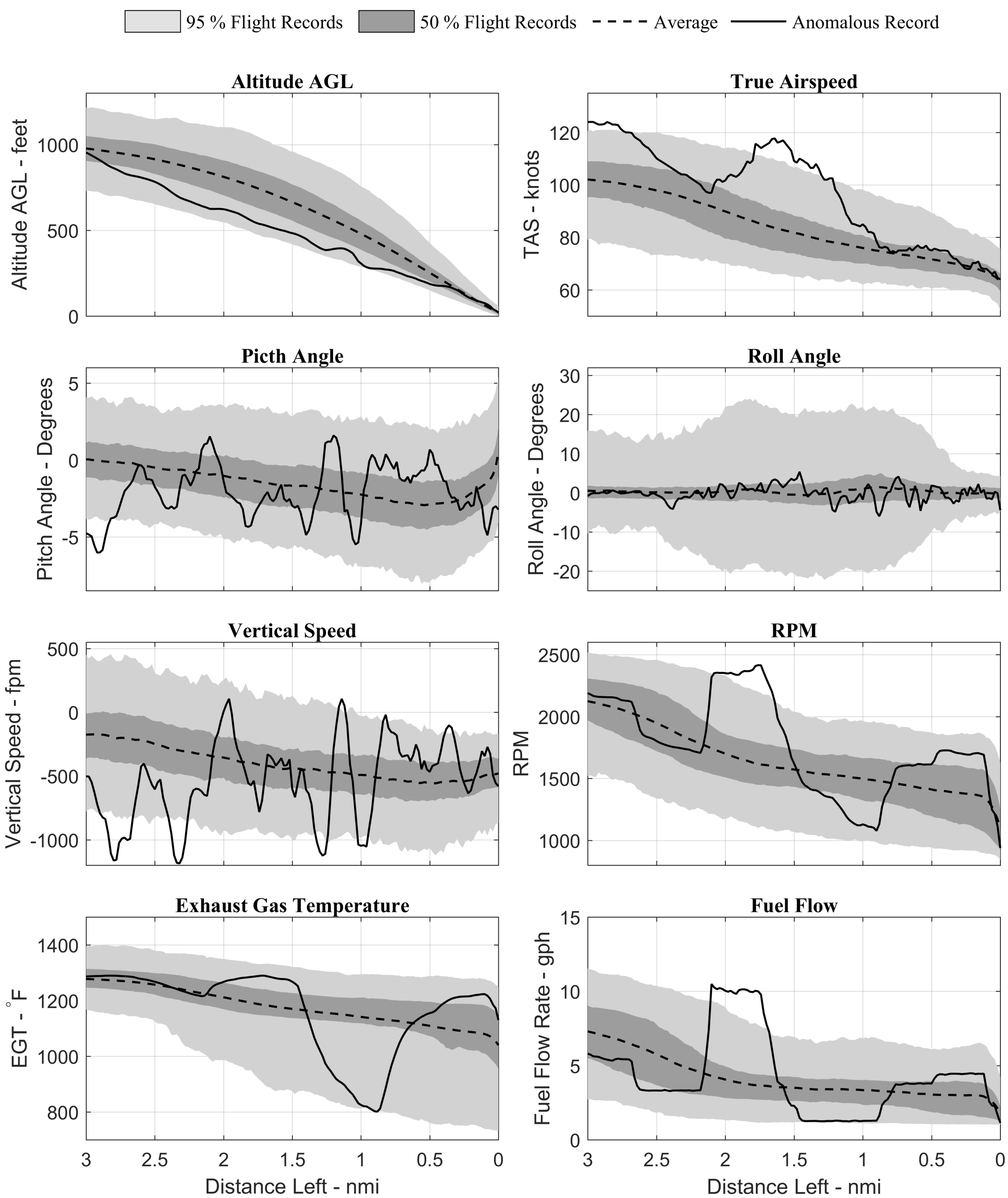


Figure 10. Visualization of a subset of flight data parameters metrics for a flight data record with low and fast approach

Therefore, we see that this methodology is able to identify flight records that have high anomaly scores and upon visualization corroborate that the approach and landing is potentially unsafe and worthy of further inspection.

VI. Conclusions and Future Work

In this paper, we have demonstrated the implementation of a generic anomaly detection methodology which utilizes energy-based metrics as features. Through specific examples, the implementation of this methodology and its potential in identifying known anomalies as well as emergent artefacts is demonstrated. The sensitivity of the methodology to the clustering parameters is quantified in experiment 1. The effect of PCA and the relative importance/contribution of individual metrics is examined in experiment 2. Experiment 3 demonstrated the effect of using different feature vectors including raw flight data and subsets of energy metrics. Thus, energy-based metrics show plenty of promise for improving operational safety when used in conjunction with data mining techniques.

A number of avenues for enhancing the applicability of this method are identified which will be pursued in future work. The steps of the generic anomaly detection framework remain the same, but different alternatives can be explored within each step. Various similarity measures such as radial basis kernel, Minkowski distance, normalized longest common sub-sequence kernel etc., can be utilized and their effect on the quality of the results can be quantified. Other machine learning algorithms such as support vector machines can be used to identify anomalies in Step IV of the methodology. One of the main advantages of energy metrics is that they can be considered as an objective metric across different aircraft types. Therefore, demonstrating the use of flight data from different aircraft will enable the scope of this methodology to be widened. The dependence of the methodology on distance-aligned feature vectors should be removed in order to make its application to all phases of flight easier. Lastly, comparison of the obtained anomalous flight records to traditional exceedance detection will enable further validation of the methodology in identifying known unsafe situations.

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 also like to thank Evan Harrison, Sanggyu Min, and Dr. Imon Chakraborty for their feedback and assistance.

A. Appendix - Implemented energy metrics

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

			Can be estimated using		
	Metric	**Formula**	**Flight Data**	**Flight Data + Performance Models**	**Requires Reference Profile**
1	Specific Potential Energy (SPE)	h	Yes	Yes	No
2	Specific Kinetic Energy (SKE)	$\frac{V^2}{2g}$	Yes	Yes	No
3	Specific Total Energy (STE)	$h + \frac{V^2}{2g}$	Yes	Yes	No
4	Specific Potential Energy Error (SPEE)	$h_{act} - h_{ref}$	Yes	Yes	Yes
5	Specific Kinetic Energy Error (SKEE)	$\frac{V_{act}^2 - V_{ref}^2}{2g}$	Yes	Yes	Yes
6	Specific Total Energy Error (STEE)	$h_{act} - h_{ref} + \frac{V_{act}^2 - V_{ref}^2}{2g}$	Yes	Yes	Yes
7	Specific Potential Energy Rate (SPER)	$\dot{h} = V \sin\gamma$	Yes	Yes	No
8	Specific Kinetic Energy Rate (SKER)	$\frac{V.\dot{V}}{g}$	Yes	Yes	No
9	Specific Total Energy Rate (STER)	$\dot{h} + \frac{V.\dot{V}}{g} = \frac{(T-D)V}{W}$	Yes	Yes	No
10	Potential Flight Path Angle (PFPA)	$\gamma + \frac{\dot{V}}{g}$	Yes	Yes	No
11	Maximum Potential Flight Path Angle (PFPA-Max)	$\frac{T_{max} - D}{W}$	No	Yes	No
12	Minimum Potential Flight Path Angle (PFPA-Min)	$\frac{T_{idle} - D}{W}$	No	Yes	No
13	Modified Total Energy Error Rate (MTEER)	$sign(\delta E) \times \frac{\Delta(\delta E)}{\Delta t}$	Yes	Yes	Yes
14	Inverse Energy Rate Efficiency (IERE)	$\frac{V_a(T-D)}{V_c W(\gamma_c + \frac{\dot{V_c}}{g})}$	Yes	Yes	Yes
15	Thrust Margin (TM)	$1 - \frac{T}{T_{max}}$	No	Yes	No
16	Energy Rate Distribution (ERD)	$sign(\frac{SKER}{SPER}) \times \exp(-\lvert\frac{SKER}{SPER}\rvert)$	Yes	Yes	No

References

[1]Logan, T. J., "Error prevention as developed in airlines," *International Journal of Radiation Oncology* Biology* Physics*, Vol. 71, No. 1, 2008, pp. S178–S181.

[2]"National Transportation Safety Board Website (Retrieved 10/2016)," .

[3]"Federal Aviation Administration, Advisory Circular, 120-82 - Flight Operational Quality Assurance," April 2004.

[4]"Federal Aviation Administration, 14 CFR 121.344 Digital Flight Data Recorders for Transport Category Airplanes, 2011," .

[5]"Federal Aviation Administration, 14 CFR 23.3 Airplane categories," .

[6]Harrison, E., Min, S., Jimenez, H., and Mavris, D. N., "Implementataion and Validation of an Internal Combustion Engine and Propeller Model for General Aviation Aircraft Performance Studies," *AIAA Aviation*, American Institute of Aeronautics and Astronautics, June 2015.

[7]Min, S., Harrison, E., Jimenez, H., and Mavris, D. N., "Development of Aerodynamic Modeling and Calibration Methods for General Aviation Aircraft Performance Analysis - a Survey and Comparison of Models," *AIAA Aviation*, American Institute of Aeronautics and Astronautics, June 2015.

[8]Puranik, T., Harrison, E., Min, S., Jimenez, H., and Mavris, D., "General Aviation Approach and Landing Analysis using Flight Data Records," *16th AIAA Aviation Technology, Integration, and Operations Conference*, 2016.

[9]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.

[10]Campos, G. O., Zimek, A., Sander, J., Campello, R. J., Micenkov, B., Schubert, E., Assent, I., and Houle, M. E., "On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study," *Data Mining and Knowledge Discovery*, 2015, pp. 1–37.

[11]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.

[12]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.

[13]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.

[14]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.

[15]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.

[16]Chandola, V., Banerjee, A., and Kumar, V., "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, Vol. 41, No. 3, 2009, pp. 15–.

[17]Liao, T. W., "Clustering of time series data - a survey," *Pattern Recognition*, Vol. 38, No. 11, 2005, pp. 1857 – 1874.

[18]Gavrilovski, A., Jimenez, H., Mavris, D. N., Rao, A. H., Shin, S., Hwang, I., and Marais, K., "Challenges and Opportunities in Flight Data Mining: A Review of the State of the Art," *AIAA SciTech*, American Institute of Aeronautics and Astronautics, Jan. 2016.

[19]Iverson, D. L., "Inductive System Health Monitoring," 2004.

[20]Budalakoti, S., Srivastava, A. N., and Otey, M. E., "Anomaly detection and diagnosis algorithms for discrete symbol sequences with applications to airline safety," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, Vol. 39, No. 1, 2009, pp. 101–113.

[21]Chu, E., Gorinevsky, D., and Boyd, S., "Detecting aircraft performance anomalies from cruise flight data," 2010, pp. –.

[22]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.

[23]Hotelling, H., "Analysis of a complex of statistical variables into principal components." *Journal of educational psychology*, Vol. 24, No. 6, 1933, pp. 417–.

[24]Ester, M., Kriegel, H.-P., Sander, J., and Xu, X., "A density-based algorithm for discovering clusters in large spatial databases with noise." *Kdd*, Vol. 96, 1996, pp. 226–231.

[25]Breunig, M. M., Kriegel, H.-P., Ng, R. T., and Sander, J., "LOF: identifying density-based local outliers," *ACM sigmod record*, Vol. 29, ACM, 2000, pp. 93–104.

[26]Zhang, K., Hutter, M., and Jin, H., "A new local distance-based outlier detection approach for scattered real-world data," *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Springer, 2009, pp. 813–822.

[27]Chakraborty, I., Lozano, B., and Mavris, D. N., "Pilot-Friendliness Considerations for Personal Air Vehicle Flight Control Systems," *AIAA Aviation*, American Institute of Aeronautics and Astronautics, June 2015.