Application of Machine Learning Techniques to Parameter Selection for Flight Risk Identification

Eugene Mangortey*, Dylan Monteiro*, Jamey Ackley*, Zhenyu Gao*, Tejas G. Puranik[†], Michelle Kirby[‡], Olivia J. Pinon[‡], and Dimitri N. Mavris[§]
Georgia Institute of Technology, Atlanta, GA 30332

In recent years, the use of data mining and machine learning techniques for safety analysis, incident and accident investigation, and fault detection has gained traction among the aviation community. Flight data collected from recording devices contains a large number of heterogeneous parameters, sometimes reaching up to thousands on modern commercial aircraft. More data is being collected continuously which adds to the ever-increasing pool of data available for safety analysis. However, among the data collected, not all parameters are important from a risk and safety analysis perspective. Similarly, in order to be useful for modern analysis techniques such as machine learning, using thousands of parameters collected at a high frequency might not be computationally tractable. As such, an intelligent and repeatable methodology to select a reduced set of significant parameters is required to allow safety analysts to focus on the right parameters for risk identification. In this paper, a step-by-step methodology is proposed to down-select a reduced set of parameters that can be used for safety analysis. First, correlation analysis is conducted to remove highly correlated, duplicate, or redundant parameters from the data set. Second, a pre-processing step removes metadata and empty parameters. This step also considers requirements imposed by regulatory bodies such as the Federal Aviation Administration and subject matter experts to further trim the list of parameters. Third, a clustering algorithm is used to group similar flights and identify abnormal operations and anomalies. A retrospective analysis is conducted on the clusters to identify their characteristics and impact on flight safety. Finally, analysis of variance techniques are used to identify which parameters were significant in the formation of the clusters. Visualization dashboards were created to analyze the cluster characteristics and parameter significance. This methodology is employed on data from the approach phase of a representative single-aisle aircraft to demonstrate its application and robustness across heterogeneous data sets. It is envisioned that this methodology can be further extended to other phases of flight and aircraft.

I. Nomenclature

ANOVA	=	Analysis Of Variance
ASAP	=	Aviation Safety Action Program
ASIAS	=	Aviation Safety Information Analysis and Sharing
ASRS	=	Aviation Safety Reporting System
CLARA	=	Clustering for Large Applications
CSV	=	Comma Separated Values
CVR	=	Cockpit Voice Recorder
DFDR	=	Digital Flight Data Recorders
DIANA	=	Divisive Analysis
DM	=	Dissimilarity Matrix
DNE	=	Does Not Exist
FAA	=	Federal Aviation Administration

---

*Graduate Research Assistant, Aerospace Systems Design Laboratory, School of Aerospace Engineering, AIAA Student Member

[†]Research Engineer II, Aerospace Systems Design Laboratory, School of Aerospace Engineering, AIAA Member

[‡]Senior Research Engineer, Aerospace Systems Design Laboratory, School of Aerospace Engineering, AIAA Senior Member

[§]S.P. Langley NIA Distinguished Regents Professor and Director of Aerospace Systems Design Laboratory, School of Aerospace Engineering, AIAA Fellow

FDR	=	Flight Data Recorders
$FOQA$	=	Flight Operational Quality Assurance
GA	=	General Aviation
GPS	=	Global Positioning System
HAT	=	Height Above Touchdown
$NASA$	=	National Aeronautics and Space Administration
NaN	=	Not a Number
$NTSB$	=	National Transportation Safety Board
ODM	=	Ordered Dissimilarity Matrix
PAM	=	Partitioning Around Medoids
PCA	=	Principal Component Analysis
SME	=	Subject Matter Experts
$SOTA$	=	Self Organizing Tree Algorithm
SSE	=	Sum Squared Error
$t-SNE$	=	t-distributed Stochastic Neighbor Embedding
VAT	=	Visual Assessment of cluster Tendancy

II. Introduction

In recent years, the use of data mining and machine learning techniques for safety analysis, incident and accident examination, and fault detection has been gaining traction among the aviation community. Modern commercial airplanes are capable of recording thousands of heterogeneous parameters at various sampling rates [1]. Regulatory agencies such as the Federal Aviation Administration (FAA) impose certain requirements defining parameters that must be recorded by commercial airliners during operations, including parameters related to the state, attitude, control surface deflections, engine information, environmental conditions, Global Positioning System (GPS) information, and others [2]. Data-driven safety programs such as Flight Operational Quality Assurance (FOQA) [3] aim to improve safety through the collection and analysis of routine flight data and has been widely adopted in commercial operations. Modern commercial aircraft are thus equipped to record thousands of parameters at a high frequency throughout the duration of the flight [1]. This recording capability coupled with the high volume of operations results in an explosion of flight data available for safety analysis to airliners and operators. Indeed, the number of operators using flight data collected on-board through Digital Flight Data Recorders (DFDR) has increased exponentially in the past few decades [1]. Airliners have established flight safety divisions to analyze flight data, investigate safety issues, and proactively identify risks. Similarly, collaborations between industry and government such as the Aviation Safety Information Analysis and Sharing (ASIAS) [4] have spurred research in data-driven techniques for aviation safety. This data is typically used during retrospective analysis to identify anomalies during routine operations using various machine learning techniques [5–13]. These safety analyses are conducted at a system-level whereas sub-system level safety analysis for existing and upcoming technologies has been subject of previous work and is beyond the scope of the current work [14].

However, among the data collected, not all information is equally important from a safety and risk perspective. Additionally, there are often challenges associated with leveraging machine learning techniques and available datasets. Many of the parameters collected can be highly correlated due to the physics of the airplane, redundancy of sensors, redundancy of parameters created by varying units, derived parameters, etc. Some data varies throughout the flight while some stays constant. There might also be parameters that have no correlation with safety. As millions of flights are flown and the size of the data grows, knowing which parameters analysts need to pay attention to becomes very critical. As such, there is a need to identify the most influential parameters from among the thousands available in order to build better risk prediction models. Finally, while the capabilities of modern commercial aircraft in terms of their data recording are exceptional, some variability may exist among different aircraft types. Therefore, developing and implementing a repeatable and robust method to down-select parameters most critical to safety can help harmonize analyses across a heterogeneous fleet of aircraft.

III. Background

A. Review of prior efforts related to aviation safety

In the past, accidents have been the primary triggers for identifying problems and developing mitigation strategies [15]. However, with the industry moving towards a more proactive approach, data collected from routine operations is now analyzed for unusual behaviors or pattern identification. The data and parameters available for such analyses can vary according to their application or purpose. For example, in the General Aviation (GA) domain, there are few parameters recorded on board the airplane or helicopter. In such instances, important parameters such as energy-based metrics are identified from among those recorded using domain expertise [16–19]. In other instances, safety events are defined using thresholds on specific parameters that have been identified by subject matter experts. An example of this type of definition is the well-known stabilized approach criteria [20]. Similar definitions exist for helicopter approach stability [21]. In commercial airline operations, data collected from on-board recorders contains thousands of channels of discrete, binary, continuous, and categorical values. These parameters may be organized in the form of a multi-level hierarchy, which consists of multiple levels of abstraction for the data collected.

Efforts have also been made over the years to leverage machine learning techniques to identify risks in aviation operations. First, Li et al. [9] leveraged clustering algorithms to detect abnormal flights without requiring the use of predefined criteria. Results obtained from their work showed that cluster-based anomaly detection identified significantly more anomalies when compared to results from exceedance analysis, an approach that is widely used by airlines. Even though their research did not provide insights into parameters critical to risk identification, it highlighted the use of clustering algorithms to identify risks in flight safety.

Second, Puranik et al. [12] used energy-based metrics to evaluate various safety-critical flight conditions. Many loss-of-control accidents are correlated to poor energy management. This work highlighted the potential value of using energy metrics and clustering techniques to identify anomalous General Aviation flight operations. The scope of this work can be further expanded to identify parameters critical to flight safety in commercial aviation.

Finally, Olive et al. [22] developed a framework to identify and characterize anomalies in en-route Mode S trajectories. This involved using a trajectory-clustering method to obtain the main flows in an airspace and using auto-encoding artificial neural networks to perform anomaly detection in flown trajectories. Even though their work provided valuable insights into estimating risks of midair collision, it did not focus on identifying parameters critical to risk identification.

B. Research Gaps & Objective

The review of efforts related to improving flight safety highlights a few limitations and/or gaps. Commercial aviation aircraft are fitted with state of the art sensors and record a large number of parameters which may provide insight into safety risks. However, there are limited studies on automated identification of important or critical parameters from a flight safety perspective. Second, previous efforts have focused solely on identifying anomalous flights and limited work has been carried out to identify parameters that are indicative of anomalies. This work addresses this gap by developing and implementing a methodology for identifying these parameters that may help identifying root causes of these anomalies. Third, identification of critical or important parameters manually by subject matter experts (SME) can be a tedious task considering the high dimensionality of the data sets. This research addresses this gap by developing a systematic methodology that minimizes human effort. Finally, the review of prior research highlighted the use of clustering algorithms in the context of flight safety. However, a rigorous benchmarking of different clustering algorithms is lacking. There is a need to perform a benchmarking exercise to determine the most appropriate clustering algorithm for the task at hand. Based on the research gaps identified, the focus of this research is to:

Develop a repeatable and robust methodology for 1) categorizing flight data and 2) identifying parameters that are indicative of anomalies

IV. Methodology

The methodology developed in this work is outlined in Figure 1. The following subsections provide further details of each step of the methodology and the specific choices made at each step.
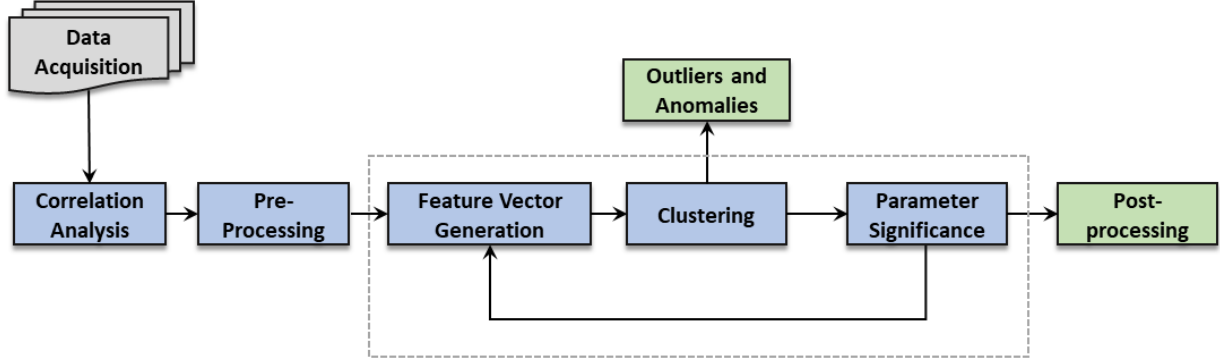

Fig. 1 Steps in the parameter significance identification and clustering methodology. Those highlighted in the dotted box are repeated iteratively for various feature vector options

A. Data Identification and Acquisition

Aviation safety data can be obtained from five primary sources: Flight Data Recorders (FDR), Cockpit Voice Recorder (CVR), Aviation Safety Action Program (ASAP), Aviation Safety Reporting System (ASRS), and Flight Operations Quality Assurance (FOQA).

- **Flight Data Recorders (FDR)** monitor and record aircraft parameters such as time, altitude, heading, airspeed, etc. Some FDRs record flight characteristics such as power settings, instrument readings, etc. that can be used by the National Transportation Safety Board (NTSB) to develop animations for the visualization of the last moments of a flight [23]
- **Cockpit Voice Recorders (CVR)** record flight crew voices and other sounds such as stall warnings, landing gear extensions and retractions, etc. in the cockpit of an aircraft. These recordings are typically used by investigators to investigate accidents, in addition to other data sources such as FDR [23]
- **Aviation Safety Action Program (ASAP)** is a voluntary reporting system used by airlines and the Federal Aviation Administration (FAA) to detect safety-related issues before they cause an accident. Flight crews anonymously report safety-related events, which can be reviewed by analyzing flight data [24]
- **Aviation Safety Reporting System (ASRS)** is a confidential system developed by the Federal Aviation Administration (FAA) that allows pilots and other flight crews to report incidents such as near misses, which may not be recorded. These reports, which are managed by the National Aeronautics and Space Administration (NASA), are analyzed and used to make improvements to aviation safety [25, 26]
- **Flight Operations Quality Assurance (FOQA)** consists of regularly recorded aircraft sensor measurements and switch settings. FOQA data is similar to FDR data. However, airlines decide which parameters to record for the FOQA dataset

Among these, *Flight Operations Quality Assurance* data consists of regularly recorded aircraft sensor measurements and switch settings. The flight data obtained from the flight data recorder are a multivariate time series, whose lengths typically vary between records due to varying duration of the flight. The data collected consists of thousands of parameters (numerical, discrete, categorical, text, etc.) recorded at a frequency of up to 16 Hz. 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.

In the present work, de-identified FOQA data obtained from commercial airline routine operations is utilized. The data consists of comma separated variable (CSV) files of approximately 5,000 flights with 623 parameters collected at a

frequency of 1 Hz. The total size of the data set is over 150 GB with each file averaging around 35 MB. The data from each flight can be further divided into various phases of flight such as Taxi Out, Take off, Climb, Cruise, Approach, Landing, Roll Out, etc. The data is extracted in the form of text files and converted to CSV along with some data cleaning (such as removal of incomplete or corrupted files etc.). Consequently, flight data from the approach phase of flight, comprised of 623 parameters was used for this research.

B. Correlation Analysis

Correlation analysis is commonly used in feature selection processes to identify feature relevance as well as feature redundancy. Redundancy in high-dimensional data sets is not always clear, especially when a feature is correlated to a set of features [27]. If left unaddressed, high correlations may lead to skewed or misleading results [28]. A correlation analysis conducted prior to feature selection can provide valuable insight into the pairwise relationships between features. Results of the analysis can help maintain features that are highly correlated with the machine learning application while selecting feature subsets that have minimal inner-correlation [29].
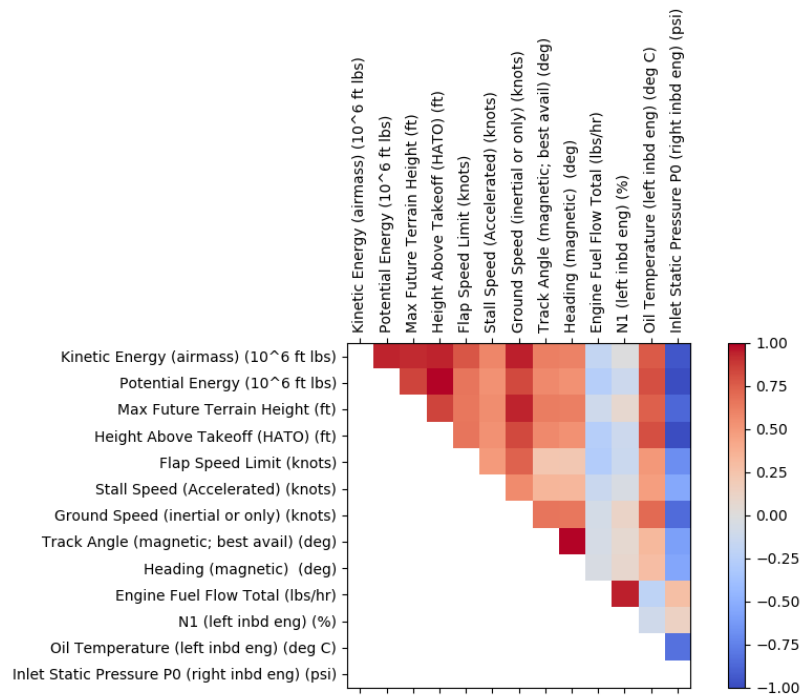

Fig. 2   Sample Correlation Matrix

An analysis is conducted with Python scripts that identify linear pairwise correlations using the Pearson correlation coefficient. The domain of the coefficient (r), $\{r \mid -1 \leq r \leq 1\}$ indicates a perfectly negative correlation at r = -1 and a perfectly positive correlation at r = 1. A value of zero indicates no measurable linear correlation. Figure 2 shows a sample correlation matrix illustrating the magnitude of the relationship between features in a FOQA dataset.

Correlation coefficients with magnitude near |1| often indicate duplicate and/or redundant features. Temperature measurements captured in multiple units is one example that displays this magnitude. Some derived features also exhibit perfect or near perfect linear correlations. When conducting a pairwise trade-off of derived and original features, the default is to select the original in lieu of the derived as this is thought to maintain the greatest amount of orthogonal information. By reducing the absolute value of the correlation coefficient in small increments and reanalyzing the data, additional redundant or near-redundant feature pairs emerge. The iterative process of selecting lower values for the correlation coefficient and evaluating highly correlated feature pairs continues until the analyst is confident that the remaining features are independent enough to obtain accurate and clear results from the machine learning application.

C. Pre-processing

The pre-processing stage in the methodology involves down-selecting parameters based on correlation analysis, SME review, and other miscellaneous methods. The output of this stage is a series of parameter subsets which are unique combinations of parameters used to generate feature vectors in the next step. Parameter sets were generated for the purpose of this research to identify the sensitivity of clustering performance and results to the combinations of parameters in the sets. Furthermore, comparing the clusters and parameter significance across parameter sets could validate the assumptions and SME review part of this pre-processing section. A summary of the pre-processing steps is shown in Table 1. Figure 3 graphically depicts the number of parameters remaining during each pre-processing step. It also highlights the significant reduction of parameters in certain steps and emphasizes the importance of pre-processing in parameter selection.

Table 1   Summary of Parameter Pre-Processing Steps

| Step	Description	Parameters Start	Parameters Removed	Parameters End	Parameter Set ID
1	Remove metadata	623	228	395	
2	Remove discrete	395	185	210	
3	Remove correlation = 1	210	6	204	
4	Remove correlation >= 0.99	204	38	166	
5	Remove correlation >= 0.97	166	19	147	
6	Remove correlation >= 0.95	147	6	141	
7	Remove misc. parameters	141	79	62	0
8	SME parameter removal	62	12	50	
9	Remove correlation >= 0.90	50	2	48	1
10	SME parameter removal	48	14	34	2
11	SME parameter removal	34	10	24	3
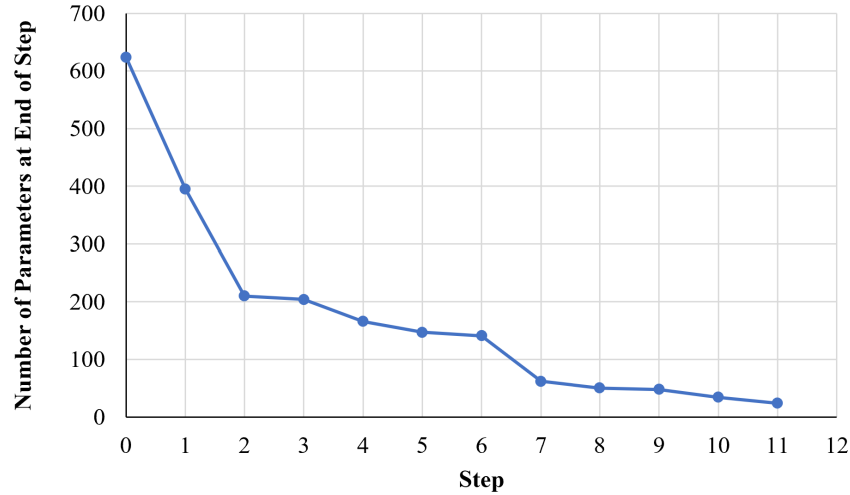
Fig. 3   Number of Parameters at Each Pre-Processing Step

Step 1 of the pre-processing consists of removing metadata parameters from the FOQA data. Metadata parameters are characteristics of the flight itself and do not vary during the flight. Examples include *Departure Airport Name*, *Departure Runway*, etc. Step 2 involves removing the discrete parameters from the set. It is noted that discrete parameters are certainly important from a flight-safety perspective, since they could include various switches in the cockpit such as *Autopilot*, *Autothrottle*, *Mode Selectors*, etc. However, for the purpose of this research, a decision

was made to include only continuous parameters as the current feature vector generation (presented in Section IV.D), normalization methods, and clustering algorithms (both presented in Section IV.E) work for continuous parameters only. The inclusion of discrete parameters is scope for future work, which would include additional research into various encoding methods, normalization methods, and algorithms to handle mixed data types.

Steps 3-6 and 9 use the results of the correlation analysis to remove highly correlated parameters at various thresholds. The focus of step 7 is the removal of miscellaneous parameters that were clearly not relevant for this analysis. These parameters included any calculated parameters (e.g. *Lift*), empty parameters, directional parameters (e.g. *Compass Heading*), and any duplicate parameters.

Steps 8, 10, and 11 involve SME review of the parameter sets. For example, parameters that are not related to the aircraft (mainly atmospheric parameters such as *Headwind*, etc.) are removed during step 10. These reductions driven by SME reviews are inherently more subjective and therefore must be made cautiously. Collaborative consideration from multiple SME sources should be considered for these steps in the approach. A regulatory review is also beneficial as regulatory parameter requirements have been established by experienced SMEs. The results and comparison across the different parameter sets could yield additional insight and validation of the SME review. A summary of which parameters are included in the different parameters sets is provided in Table 5 in the Appendix.

D. Feature Vector Matrix Generation

This stage of the methodology (from Figure 1) involves generating feature vectors using the FOQA data and parameter combinations from the previous stage. A feature vector is a vector that contains information describing an object's important characteristics. In this case, a feature vector for a given flight involves mapping FOQA entries from the flight into a single row vector. A feature vector matrix has all of the feature vectors for a set of flights, with each row representing a given flight. Each column in the feature vector is a parameter-altitude combination and the naming convention for each column is *Parameter_i*, where *Parameter* is the name of the parameter and *i* is an integer counter that represents a particular altitude above touchdown. For this analysis, the feature vectors were sampled using HAT (Height Above Touchdown). For example, *Airspeed (true)_10* signifies the 10th HAT interval for the parameter *Airspeed (true)*. Therefore, each column represents the value of a parameter across all flights at the sample height above touchdown. The format of a feature vector and feature vector matrix is shown in Figure 4. Note that if a column (parameter-altitude combination) in the feature vector has the same value for all flights, then that column is removed from the analysis because that column would have no effect on the clustering results. The feature vector matrices are the inputs to the clustering algorithms.
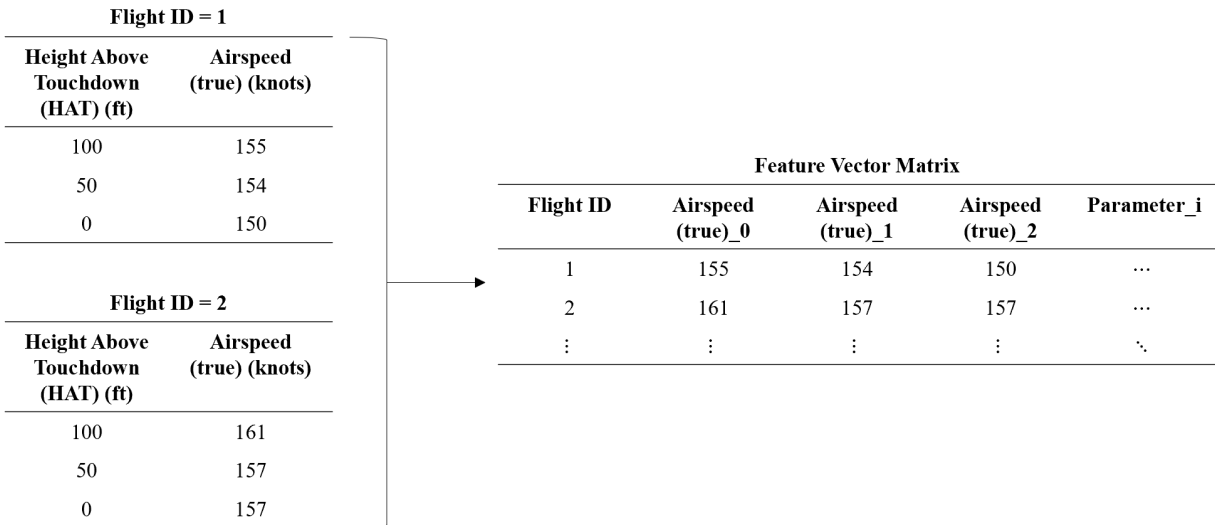
Flight ID = 1

Height Above Touchdown (HAT) (ft)	Airspeed (true) (knots)
100	155
50	154
0	150

Flight ID = 2

Height Above Touchdown (HAT) (ft)	Airspeed (true) (knots)
100	161
50	157
0	157

Feature Vector Matrix

Flight ID	Airspeed (true)_0	Airspeed (true)_1	Airspeed (true)_2	Parameter_i
1	155	154	150	⋯
2	161	157	157	⋯
⋮	⋮	⋮	⋮	⋱

Fig. 4 Notional Example of a Feature Vector Matrix

In order to test the sensitivity of the clustering results to the definition of the approach phase, three different altitude cutoffs were used to extract the data for the feature vectors. In parallel work [5], a more thorough treatment of the effect

of approach phase length on safety risk and anomalies is presented. In the present work, all feature vectors ended at the touchdown point (i.e. 0 height above touchdown) and all feature vectors used an altitude interval of 50ft. The three altitude cutoffs were:

- 1,000 ft above touchdown to the touchdown point, in 50ft intervals
- 3,000 ft above touchdown to the touchdown point, in 50ft intervals
- 5,000 ft above touchdown to the touchdown point, in 50ft intervals

As such, a total of 12 feature vectors are created using the combinations of the three altitude segments and four parameter sets described in Table 5 in the Appendix.

E. Clustering

The next step in the methodology focuses on leveraging clustering algorithms to achieve the objectives of this research. Clustering is an unsupervised Machine Learning technique that has been widely used to identify trends and patterns in datasets. Clustering is achieved by partitioning datasets into clusters, where objects in one cluster are similar to each other compared to objects in other clusters [30–34]. Figure 5 provides an overview of the steps taken to cluster the feature vectors highlighted previously.

Fig. 5   Overview of clustering process

1. Normalize data

The clustering process can be skewed by parameters with larger ranges of values, as they will completely dominate the other parameters. It is thus important to normalize the feature vectors using methods such as z-score standardization. Z-score standardization scales parameters to ensure that they have a mean of zero and a standard deviation of one [35–37] and is calculated using:

$$Z = \frac{Value - Mean}{Standard\ Deviation}$$

2. Reduce dimensionality of data

Using datasets with many features makes it difficult to explore and understand the relationships between the features. To address these challenges, the dimensionality of datasets should be reduced to minimize the number of random variables under consideration. Dimension reduction techniques such as the t-Distributed Stochastic Neighbor Embedding (t-SNE) produce a set of principal components, which reduces the likelihood of overfitting models, and makes it easier to explore and visualize relationships between features [38–40]. In particular, the t-SNE technique calculates the probability of the similarity of points in high and low dimensional spaces. The similarity of points is calculated as the conditional probability that a point X would choose point Y as its neighbor if neighbors were picked in proportion to their probability density under a normal distribution centered at X. The difference between the conditional probabilities in the higher dimensional and lower dimensional spaces is then minimized to produce a representation of data points in the lower dimensional space [41–43]. In the context of this research, the t-SNE technique was implemented after the highly dimensional dataset produced poorly performing clusters.

3. Assess clustering tendency of data

The majority of clustering algorithms split up datasets into predefined numbers of clusters, even if no meaningful clusters exist. It is thus important to assess the clustering tendency of a dataset to determine if meaningful clusters

can be created [44–47]. Hopkins statistic [48] and Visual Assessment of cluster Tendency (VAT) [44–47, 49] are two methods that are commonly used to determine if a dataset has a non-random structure and will produce useful clusters [49]. The null hypothesis for the Hopkins statistic is defined as the dataset being uniformly distributed. The alternative hypothesis is defined as the dataset not being uniformly distributed. Thus, a Hopkins statistic close to zero means that the null hypothesis is rejected and that the dataset has a high clustering tendency [50]. The Visual Assessment of cluster Tendency (VAT) on the other hand is an image that indicates the presence of meaningful and well separated clusters, represented by dark boxes along the main diagonal of the image, as seen in Figure 6. The VAT is implemented by computing the Dissimilarity Matrix (DM) between objects in the dataset using the Euclidean distance measure [51, 52]. An Ordered Dissimilarity Matrix (ODM) is then created by reordering the original Dissimilarity Matrix so that similar objects are close to one other. The Ordered Dissimilarity Matrix is then displayed as the VAT for the dataset [44–47].

Fig. 6   Visual Assessment of cluster Tendency (VAT)

4. Benchmark algorithms

   Clustering algorithms have different methodologies and assumptions. It is thus important to benchmark different clustering algorithms to identify the best suited one for the dataset under consideration. The clustering algorithms benchmarked for this research are:

   1) **Agglomerative Hierarchical Clustering Algorithms**: Agglomerative hierarchical clustering algorithms group similar objects in multidimensional spaces into categories by assigning each object to a cluster and then merging similar clusters by their proximity to each other [53]. The agglomerative hierarchical clustering algorithms leveraged for this research are:
      • Complete Linkage: The distance between two clusters is defined as the longest distance between two objects in each cluster [54]
      • Average Linkage: The distance between two clusters is defined as the average distance between each object in one cluster to every object in the other cluster [54]
      • Centroid Linkage: The distance between two clusters is defined as the distance between the centroids of the clusters
      • Single Linkage: The distance between two clusters is defined as the shortest distance between two objects in each cluster [54]
      • Ward: The distance between two clusters is defined as how much the sum of squares will increase when the clusters are merged [55]
   2) **Divisive Analysis (DIANA)**: This algorithm initially places all objects into the same cluster. At each point in time, the algorithm then splits the largest available cluster into two smaller clusters until each cluster contains at least one object [56]
   3) **Self Organizing Tree Algorithm (SOTA) Clustering Algorithm**: This algorithms splits objects into clusters without specifying a predetermined number of clusters. It does so by detecting patterns in the dataset without any human interaction [57]

4) **Kmeans Clustering Algorithm**: The Kmeans algorithm assigns objects to a predetermined number of clusters, where the differences between objects in each cluster are minimized, and the differences between objects in different clusters are maximized [58]

5) **Partitioning Around Medoids (PAM) Clustering Algorithm**: The Partitioning Around Medoids (PAM) or k-medoids clustering algorithm is similar to the Kmeans clustering algorithm. However, the PAM algorithm clusters objects into a predetermined number of clusters around medoids or centers [56]

6) **Clustering for Large Applications (CLARA) Clustering Algorithm**: This algorithm works similarly to the Partitioning Around Medoids (PAM) or k-medoids clustering algorithm, where objects are clustered around centers or medoids. However, the CLARA algorithm only clusters a sample of the large dataset and then assigns the remaining objects in the dataset to the clusters obtained from the sample [56]

7) **Model-based Clustering Algorithm**: This algorithm is a statistical model made up of a combination of Gaussian distributions that are used to fit the data, where each combination of Gaussian distributions represents a cluster [59]

5. Evaluate clusters

Most clustering algorithms require users to arbitrarily select the number of clusters to be used. In order to ensure that the combination(s) of optimal number of clusters and best suited algorithms are identified, there is a need to vary the number of clusters used while benchmarking the different algorithms. This combination is identified using these evaluation metrics:

- **Connectivity:** "This measures the extent to which items are placed in the same cluster as their nearest neighbors in the data space" [57, 60, 61]. Connectivity ranges from zero to infinity and should be minimized
- **Dunn Index:** This measures the ratio between the smallest distance between items in different clusters and the largest distance between items in the same cluster [60–62]. The Dunn Index ranges from zero to infinity and should be maximized
- **Silhouette:** This measures the average distance between different clusters [60, 61, 63] and ranges from 1 to -1. A Silhouette score of -1 refers to poorly clustered items while a score of 1 refers to well clustered items

F. Parameter Significance

After different clustering algorithms are applied and evaluated, the one-way analysis of variance (ANOVA) [64] is performed on selected clustering results to identify parameter significance. The goal of the one-way ANOVA is to determine whether a parameter has a common mean across different clusters. In other words, parameter's significance is identified if the mean values of its distributions across multiple clusters are different enough. The ANOVA for the one-way layout can be derived as follows. Using the linear model [64]:

$$y_{ij} = \hat{\eta} + \hat{\tau}_i + r_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i \tag{1}$$

where y_{ij} is the jth observation in cluster i, $\hat{\eta}$ is the estimated overall mean, $\hat{\tau}_i$ is the estimated ith cluster effect, the errors r_{ij} are independent $N(0, \sigma^2)$ with mean 0 and variance σ^2, k is the number of clusters, and n_i is the number of observations in cluster i. Then using the dot subscript to indicate the summation over the particular index, the following decomposition is obtained:

$$\hat{\eta} = \bar{y}_{..}, \quad \hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}, \quad r_{ij} = y_{ij} - \bar{y}_{i.} \tag{2}$$

Equation 1 then becomes:

$$y_{ij} - \bar{y}_{..} = (\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.}) \tag{3}$$

Squaring both sides of Equation 3 and summing over i and j yield

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^{k} n_i (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 \tag{4}$$

Note that for each i,

$$(\bar{y}_{i.} - \bar{y}_{..}) \sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.}) = 0 \tag{5}$$

so that these cross-product terms do not appear in Equation 4. Equation 4 indicates that the corrected total sum of squares equals the treatment sum of squares (SST) plus the error sum of squares (SSE). In this specific case, the treatment sum of squares is also the *between-cluster sum of squares* and the error sum of squares is the *within-cluster sum of squares*.

The F statistic for the null hypothesis is that there is no difference between the clusters, $H_0 : \tau_1 = \ldots = \tau_k$, is,

$$F = \frac{\sum_{i=1}^{k} n_i(\bar{y}_{i.} - \bar{y}_{..})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2/(N-k)} \tag{6}$$

where N is the total number of observations. Under the null hypothesis H_0, this F statistic has an F distribution with parameters $k-1$ and $N-k$, and the area under its curve to the right of the observed F value is defined as the p value

$$p = \text{Prob}(F_{k-1,N-k} > F) \tag{7}$$

A small p value provides evidence that there is an cluster-to-cluster difference for the current parameter. In the rest of this work, the F test rejects the null hypothesis H_0 at the 0.01 level.
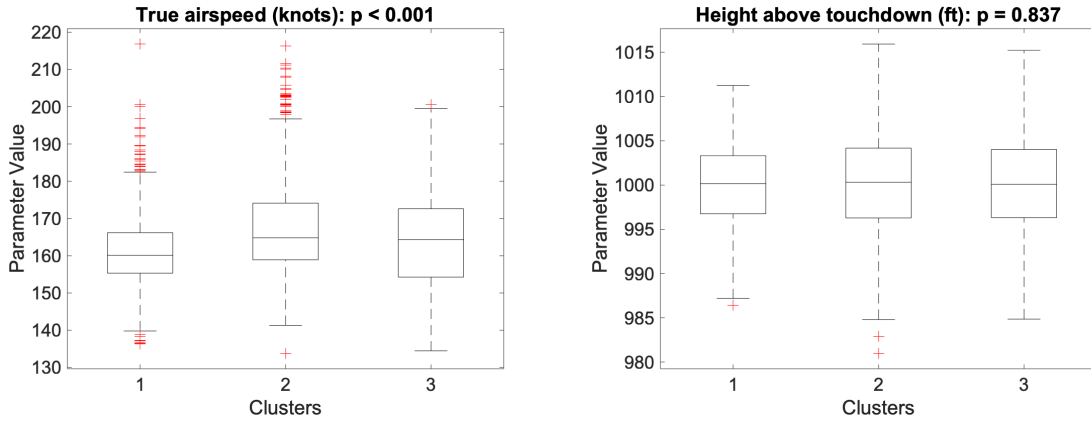

Fig. 7 Examples of significant (left) and nonsignificant (right) parameters under ANOVA

The two examples in Figure 7 illustrate how the one-way ANOVA is used to identify parameters' significance. The left panel of Figure 7 shows an example in which the parameter "True airspeed (knots)" distributes differently between the three clusters. The p value in this case is smaller than 0.001, which rejects the null hypothesis $H_0 : \tau_1 = \tau_2 = \tau_3$ at the 0.01 level. The right panel of Figure 7 is another example in which the parameter "Height above touchdown (ft)" is insignificant under the ANOVA test. In this example, a p value of 0.837 is a strong indication that the distributions of this parameter have the same mean and are similar across all three clusters.

Once H_0 is rejected in a case, two follow-up questions emerge: (1) what pairs of clusters are different, and (2) how does the parameter behave in each cluster? These questions will be addressed by further in-depth comparisons in the post-processing section.

G. Post-processing and Decision-Making Framework

The final step in the methodology presented in Figure 1 is to evaluate the results of the clustering and post-process the data. There are two main parts to this step. The first is a high-level analysis of the clustering metrics results from the clustering algorithms presented in Section IV.E for all of the feature vectors generated in Section IV.D. The second part is a detailed cluster analysis for a particular feature vector case to identify cluster characteristics and significant parameters. To aid in the post-processing, interactive dashboards were created for both parts using Tableau version 2019.3.

11

1. Analyzing Clustering Metrics

The objective of this part is to evaluate the clustering performance of the various feature vectors generated using the Connectivity, Dunn Index, and Silhouette Score outlined in Section IV.E. For each feature vector, the combination of clustering algorithm and number of clusters that had the best score for these three metrics was selected as the output from the clustering step. This output represents a case that is the best performing combination of algorithm and number of clusters for a given feature vector. The metrics for each case were plotted on an interactive dashboard so that each case can be compared and contrasted. Furthermore, the t-SNE plot was also saved for each case and incorporated in the dashboard. An example of the clustering metrics dashboard indicating the results for different feature vectors is shown in Figure 8.
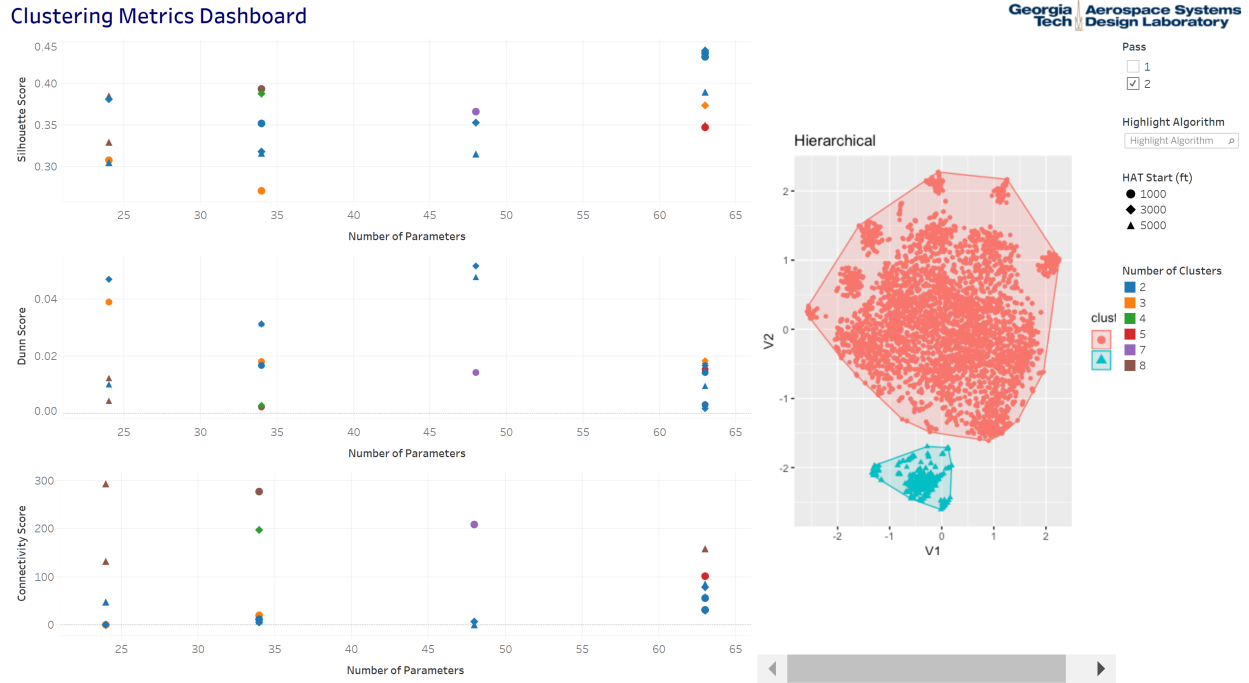
Fig. 8   Clustering Metrics Decision-Making Dashboard

The core capabilities of the interactive dashboard are summarized below:

- Displays the variation of Connectivity, Dunn Index, and Silhouette Score as a function of number of parameters in each feature vector
- A user can click on a particular case data-point and the corresponding t-SNE chart will be displayed
- Highlight the results of a selected clustering algorithm, feature vector altitude cutoff, or number of optimal clusters to observe how clustering performance varies across these dimensions
- When the user clicks on a case data-point, it will highlight on all three of the metric charts

As such, this dashboard allows the user to compare the clustering performance of the different cases and then select which cases should be explored further to analyze the characteristics of the resulting clusters.

2. Analyzing Cluster Characteristics

Once the user has selected a case of interest from the dashboard presented in Figure 8, the characteristics of the clusters themselves can be analyzed using the detailed parameter analysis dashboard created in this part of the methodology. An example of this dashboard is shown in Figure 9. The dashboard incorporates the cluster labels for the selected case, parameter distributions, parameter time-series, and the ANOVA results from Section IV.F.

The main capabilities of this interactive dashboard are summarized below:

- Rank order the parameters by their ANOVA significance score
- Display the number of clusters and how many flights are in each cluster
- The user can click on a parameter in the *Parameter ANOVA Significance* pane and the corresponding plots for that parameter are displayed
- Display of a box plot that shows the distribution of the selected parameter across the different clusters
- The user can apply an altitude filter to the box plot to only show data for a particular altitude range
- Display of a line graph that shows the variation of each parameter as a function of height above touchdown (each line represents a flight and its color-coded by cluster)
- Filter to hide/show certain clusters from the box plot and line graph
- The user can click on a cluster label in the *Number of Flights per Cluster* section and the corresponding cluster of flights is highlighted in the line graph
- The user can select an individual flight and highlight by Flight ID - this is useful to compare the behavior of individual flights across multiple parameters

This dashboard allows the user to analyze the characteristics of each cluster and identify the parameters that drive the formation of the clusters.
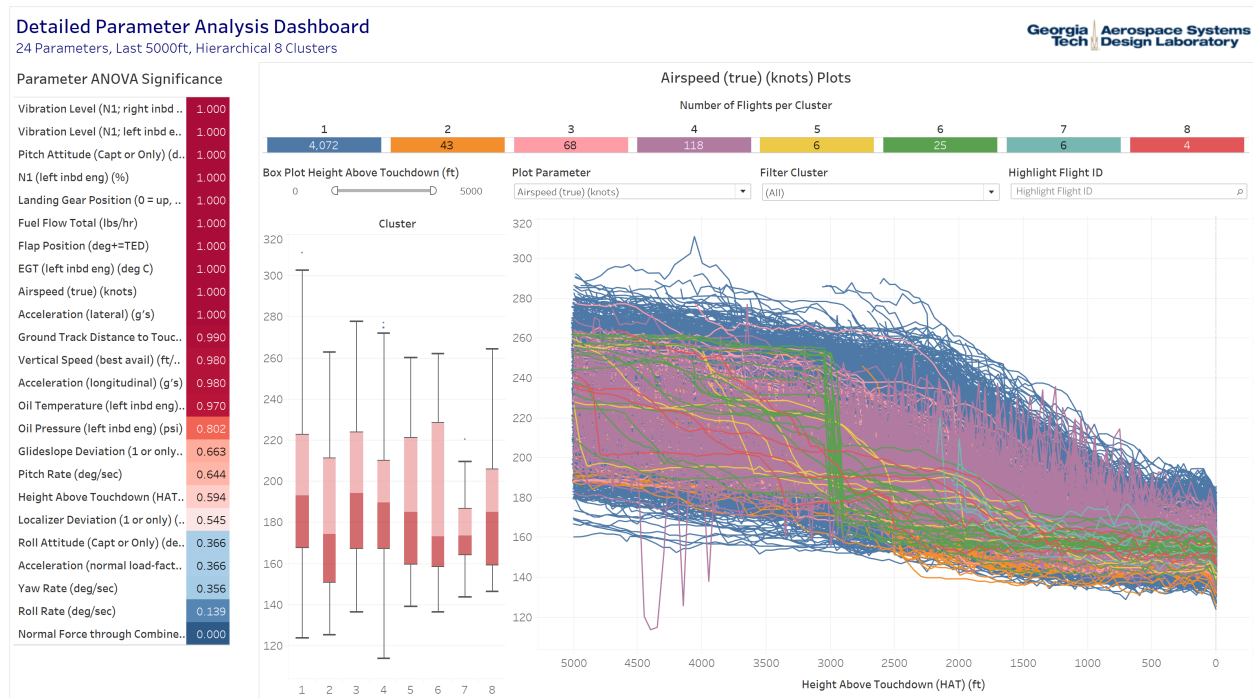

Fig. 9   Example of Detailed Parameter Analysis Dashboard for a Particular Case

V. Implementation and Results

This section discusses results obtained for a use case focused on identifying an appropriate set of parameters for detecting anomalies in the approach phase of flight using the methodology presented in Section IV.

A. Use Case Description: Approach and Landing Phases

Improving safety during the approach and landing phases has been a subject of interest due to the critical nature of these phases of flight. Errors or deviations from the flight path during approach and landing are more likely to end in accidents or incidents, as they have a smaller safety margin due to the limited time for a pilot to correct an error or react to a deviation [65]. The frequency of safety event occurrences also varies with respect to the phase of flight. Table 2 provides a breakdown of events by phase of flight for the de-identified flight data available. Specifically, it shows that

13

approximately 80% of events occurred during the approach phase of flight. Thus, while it is important to analyze all phases of flight, this research focuses on identifying outliers and significant parameters during this phase. Consequently, each of the feature vectors were generated using data from the approach phase. The remainder of this section discusses results obtained from the feature matrices generated in the approach and landing phases using different numbers of parameters and approach phase lengths.

<p align="center">Table 2   Events by phase of flight</p>

Phase of flight	Frequency (%)
Start & Push	2.41
Taxi Out	0.09
Takeoff	4.77
Rejected takeoff	0.06
Initial Climb	5.65
Climb	2.01
En route	2.71
Descent	2.01
Approach	79.41
Roll Out	0.01
Unknown State	0.87

B. Clustering Results

1. Assess clustering tendency of data

The clustering tendency of each feature matrix was assessed using Hopkins Statistic and the Visual Assessment of cluster Tendency (VAT). This was done after normalizing and reducing their dimensionality using t-SNE. The Hopkins statistic obtained for the feature matrix generated with 24 parameters and an altitude cutoff of 5,000 ft above touchdown was **0.051**, which indicates very high clustering tendency of the data. The presence of multiple dark boxes along the diagonal of Figure 10 also indicates that the clustering tendency of the dataset is very high.
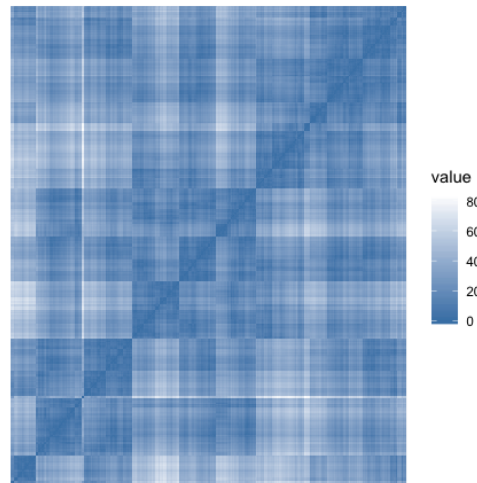

Fig. 10   Visual Assessment of cluster Tendency (VAT) for feature matrix generated with 24 parameters and an altitude cutoff of 5,000 ft above touchdown

2. Benchmarking and evaluation of clustering algorithms

Figure 16 in the Appendix shows the results obtained from the evaluation metrics used to determine the appropriate clustering algorithm and optimal number of clusters for the feature matrix generated with 24 parameters, 5,000 ft. In particular, it shows that the appropriate algorithm for this feature matrix is one of the Hierarchical algorithms, and the optimal number of clusters is two. It is worth noting that the combination of Hierarchical algorithms with two clusters was identified as the best scoring combination for each of the feature matrices.

Analysis of the clusters revealed that 4,371 flights were placed in one large cluster, while 24 flights were outliers for each of the feature matrices, as seen in Figure 11. Further analysis also revealed that the 24 outlier flights were characterized by abnormal values of the parameter *Acceleration (normal load-factor) (g's)*. This parameter was constant at a value of -3.375 g's for the entire length of the approach phase. It was also observed that these flights were operated with the same aircraft, and that the observed patterns may have been a result of faulty sensors, such as a faulty accelerometer. This finding was noted, and consequently the 24 flights were isolated and the clustering process was repeated with the remaining 4,371 flights in cluster 1, as seen in Figure 11. This re-clustering of cluster 1 leads to additional insight of all the flights without an erroneous sensor.
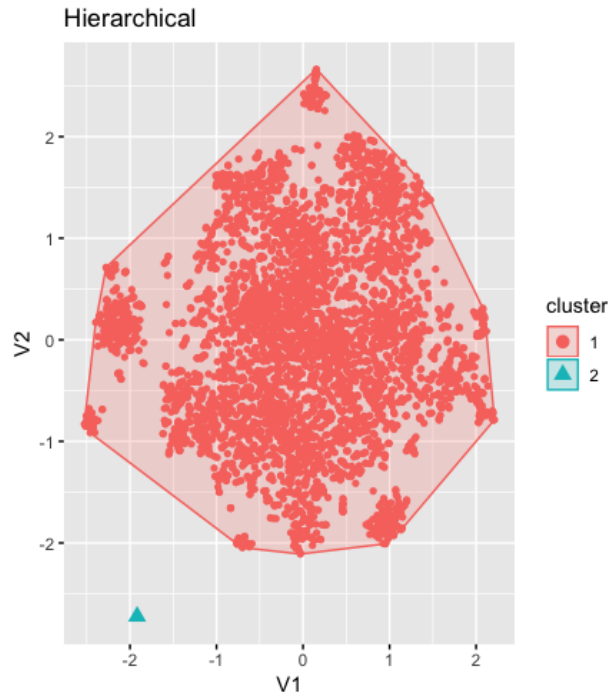

Fig. 11   Visualization of clusters obtained from the Hierarchical algorithms for the first pass of all feature vectors

All of the results and discussion henceforth are for the second pass of clustering on the main group of 4,371 flights. The metric results for this second pass are shown in Figure 17 in the Appendix. In this case, the optimal algorithm-cluster combination was not clear through the metrics, as each metric had a different ranking. For the best connectivity score, the hierarchical clustering algorithm with 2 clusters was optimal. For the best Dunn index, the hierarchical clustering algorithm with 8 clusters was optimal. For the best Silhouette score, the k-means clustering algorithm with 8 clusters was optimal. The resulting t-SNE plot for these three results are shown in Figures 12. The decision-making dashboards outlined in Section IV.G were used to decide which results to analyze in detail. The results are discussed below in Section V.D.
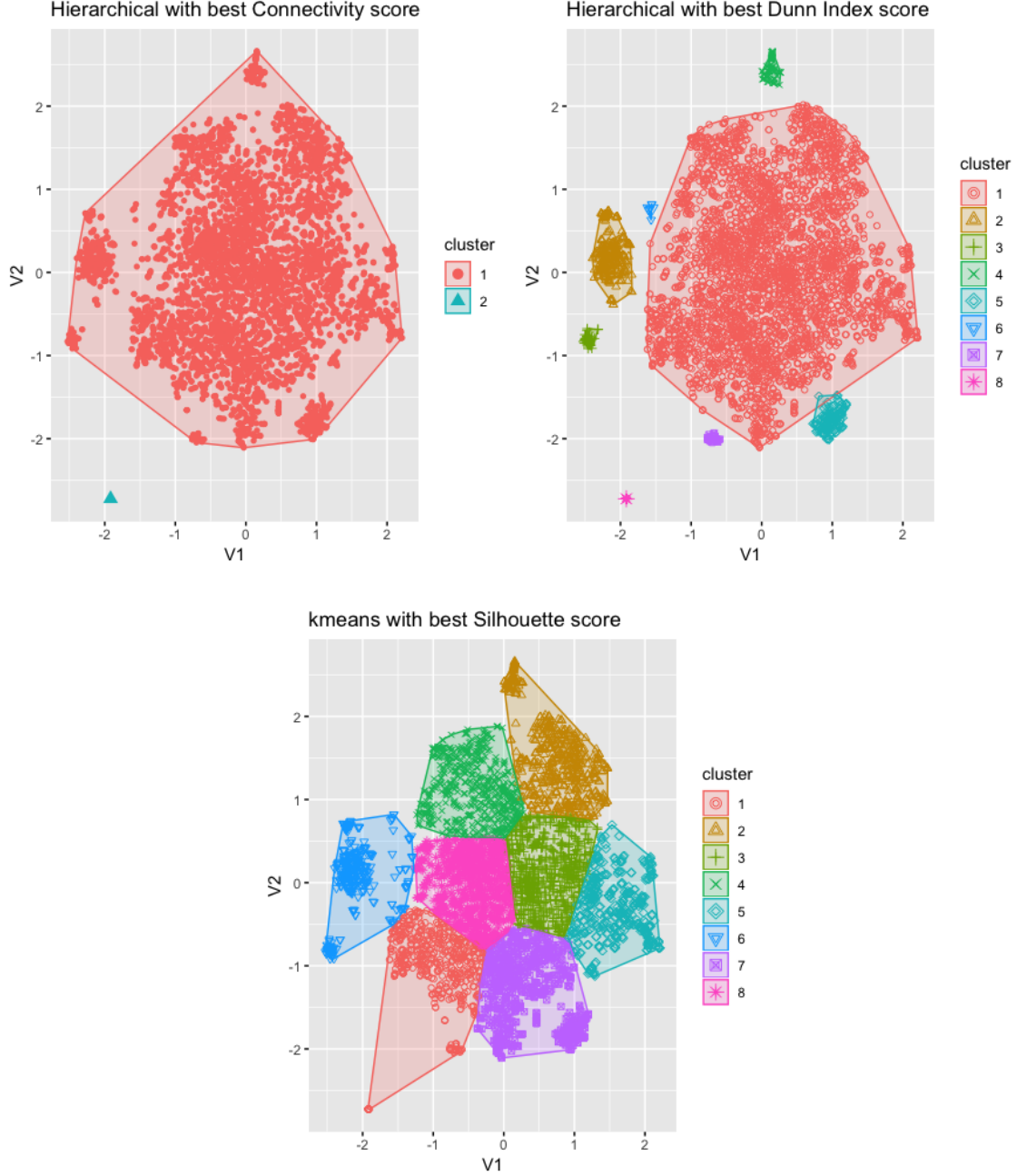
Fig. 12   t-SNE Plots of Optimal Algorithm-Cluster Combinations for Second Pass of 24 Parameter, 5,000 ft Feature Matrix

C. Parameter Significance

In this section, the ANOVA-based parameter significance approach introduced in Section IV.F is extended to identify influential parameters throughout the entire approach phase. Here, we introduce the procedure of arriving at an ANOVA summary table for each case, and discuss the results from two example cases using the same feature matrix of 24 parameters, 5,000 ft (for the second clustering pass).

Figure 13 shows an ANOVA summary table for 24 parameters, 5,000 ft, and 2 hierarchical clusters. In the majority of this table, each cell contains the parameter significance information for a specific parameter and height above touchdown

combination. Note that in a 5,000 ft approach case, there are 101 altitude levels (every 50 ft, indexed 0-100). Due to limited space, each table only displays 21 altitude levels using an increment of 5. Each cell corresponding to a parameter and altitude level combination indicates the one-way ANOVA comparison result similar to Figure 7. A result of '1' indicates that the ANOVA test has a p value smaller than 0.01 and the parameter is identified as significant at this altitude level. Similarly, a result of '0' indicates an insignificant result, and an 'N/A' indicates that the parameter is not applicable in this altitude level. The 'Significance Score' at the bottom of the table averages the applicable scores for each parameter. For example, in Figure 13 the parameter 'True Airspeed (knots)' has '1's in 17 out of 21 altitude levels, and a significance score of 0.81 is calculated. Under this evaluation method, a parameter with higher significance score is deemed more influential in the approach operation. In the case of Figure 13, six parameters have a significance score higher than 0.5, indicating that they are overall more influential than the other parameters under the current altitude levels and clustering methods. From the table in Figure 13, one can also observe the distribution of parameter significance. Some parameters are significant only in certain segments during the whole approach phase. This also provides insights on segment-level parameter significance.

Similar results from 24 parameters, 5,000 ft, and 8 hierarchical clusters are shown in Figure 14. As we increase the number of clusters on the exact same data set, one prominent observation is that the majority of ANOVA test results becomes '1', which means that more parameters are now identified as significant. In Figure 14, among the 24 parameters, 19 parameters have a significance score of higher than 0.5. Some of them are significant in all the altitude levels. This result is expected because increasing the number of clusters leads to a better differentiation of the flights.

D. Detailed Cluster Analysis Using Decision-Making Framework

The decision-making dashboards from Section IV.G were applied to the second-pass results of the 24 parameter, 5,000 ft approach feature matrix.

1. Analyzing Clustering Metrics

First, the high-level clustering metrics dashboard (outlined in Section IV.G.1) was used to determine which clustering result of the 24 parameter, 5,000 ft approach feature matrix should be examined in further detail. Recall that the second pass of this feature matrix resulted in three different optimal algorithm-cluster combinations, as outlined in Section V.B.2 and illustrated in Figure 12.

Figure 15 shows the metric dashboard with the hierarchical algorithm with 8 clusters case highlighted. This result was highlighted because the hierarchical clustering algorithm seemed to determine seven different "outlier" groups of flights that are relatively distinct from the main group of flights (cluster 1). After comparing this 8-cluster-hierarchical result to the 2-cluster-hierarchical result in Figure 12, the 8-cluster version was chosen for multiple reasons. It captures cluster 2 in the 2-cluster result and also adds additional insight into six other clusters not captured in the 2-cluster result. The three metrics shown in Figure 15 have acceptable values with no extreme outliers. And finally, the 8-cluster-hierarchical result is the most optimal for this feature matrix in terms of Dunn Index (as shown in Figure 17 in the Appendix). Additionally, the 8-cluster-hierarchical result might be more meaningful than the 8-cluster-kmeans result shown in Figure 12 since the clusters in the hierarchical result seem to have more distinct, dense clusters. Taking into account these considerations, the result of the hierarchical algorithm with 8 clusters for the second pass of the 24 parameter, 5,000 ft approach feature matrix was used for the rest of this analysis.

2. Analyzing Cluster Characteristics

The detailed cluster analysis dashboard (outlined in Section IV.G.2) was built for the 8-cluster-hierarchical result for the 24 parameter, 5,000 ft feature matrix. A snapshot of the dashboard is presented in Figure 9. This dashboard was used to determine the characteristics of each cluster as well as what parameters drove the formation of the various clusters. A summary of the cluster characteristics is shown in Table 3.

The ANOVA result from Section V.C is summarized in Table 4. The 19 highlighted parameters have an ANOVA significance score greater than 0.5, and are thereby labelled as "critical parameters" for this particular feature matrix.

\ Parameters Altitude Level Indices	Lateral Acceleration (g's)	Longitudinal Acceleration (g's)	Normal Load Factor (g's)	True Airspeed (knots)	EGT - left inbd eng (deg C)	Flap Position (deg+=TED)	Total Fuel Flow (lbs/hr)	Glideslope Deviation (1 or only dots)	Ground Track Distance to Touchdown (nm)	Height Above Touchdown (ft)	Landing Gear Position (0=up, 1=in transit, 2=down)	Localizer Deviation (1 or only, dots)	N1 - left inbd eng	Normal Force through Combined Gear (lbs)	Oil Pressure - left inbd eng (psi)	Oil Temperature - left inbd eng (deg C)	Pitch Attitude - Capt or Only (deg)	Pitch Rate (deg/sec)	Roll Attitude - Capt or Only (deg)	Roll Rate (deg/sec)	Vertical Speed - best avail (ft/min)	Vibration Level - N1 - left inbd eng	Vibration Level - N1 - right inbd eng	Yaw Rate (deg/sec)
0	0	0	0	1	1	1	0	1	0	0	0	0	1	N/A	1	1	1	0	0	0	1	0	0	0
5	0	0	0	1	1	1	0	1	0	0	0	0	1	N/A	0	1	0	0	0	0	0	0	0	0
10	0	0	0	1	1	1	0	1	0	0	0	0	1	0	1	1	1	0	0	0	1	0	0	0
15	0	0	0	1	1	1	0	0	0	0	0	0	1	N/A	1	1	1	1	0	0	1	0	0	0
20	0	0	1	0	1	1	0	1	0	0	0	0	0	N/A	0	1	1	1	0	0	1	0	0	0
25	0	0	0	0	1	1	0	1	0	0	0	0	0	N/A	0	1	0	1	0	0	0	0	0	0
30	0	0	0	0	1	0	0	1	0	0	0	0	0	N/A	0	1	0	1	0	0	0	0	0	0
35	0	0	0	0	1	0	0	0	0	0	0	0	0	N/A	0	0	0	0	0	0	0	0	0	0
40	0	0	0	1	1	0	0	0	0	0	0	0	0	N/A	0	1	0	0	0	0	1	0	0	0
45	0	0	0	1	1	0	0	1	0	0	0	1	0	N/A	1	1	0	0	0	0	0	0	0	0
50	0	0	0	1	0	1	1	0	0	0	1	0	1	N/A	1	1	1	0	1	1	1	0	1	1
55	0	1	0	1	0	1	0	0	0	0	1	0	0	N/A	0	1	1	0	0	0	0	0	0	0
60	0	1	0	1	0	1	0	0	1	0	1	1	0	N/A	0	1	1	0	0	0	0	0	0	0
65	0	0	0	1	0	1	1	1	0	0	1	0	0	N/A	0	1	0	0	0	0	0	0	0	0
70	0	0	0	1	0	1	1	1	0	0	0	1	1	N/A	1	1	0	0	0	0	0	0	0	0
75	0	0	0	1	1	1	0	1	0	0	0	1	0	N/A	0	1	0	0	0	0	0	0	0	0
80	1	1	1	1	1	1	0	1	0	0	0	1	0	N/A	0	1	1	1	1	1	0	0	1	1
85	1	1	1	1	0	1	0	1	0	0	0	1	0	N/A	0	1	0	1	1	0	0	0	0	1
90	0	0	1	1	1	1	0	0	0	0	0	1	0	N/A	0	1	1	1	1	1	1	0	0	1
95	0	1	0	1	1	1	0	1	0	0	0	0	1	N/A	1	1	1	0	1	0	1	0	0	0
100	0	0	0	1	1	1	0	1	0	0	N/A	0	0	0	0	1	1	0	0	0	0	0	0	0
Significance Score	0.10	0.24	0.19	0.81	0.71	0.81	0.14	0.67	0.05	0.00	0.20	0.38	0.33	0.00	0.33	0.95	0.52	0.33	0.24	0.14	0.38	0.00	0.10	0.19

Fig. 13   ANOVA summary table for 24 parameters, 5,000 ft, and 2 hierarchical clusters

\ Parameters Altitude Level Indices	Lateral Acceleration (g's)	Longitudinal Acceleration (g's)	Normal Load Factor (g's)	True Airspeed (knots)	EGT - left inbd eng (deg C)	Flap Position (deg+=TED)	Total Fuel Flow (lbs/hr)	Glideslope Deviation (1 or only dots)	Ground Track Distance to Touchdown (nm)	Height Above Touchdown (ft)	Landing Gear Position (0=up, 1=in transit, 2=down)	Localizer Deviation (1 or only, dots)	N1 - left inbd eng	Normal Force through Combined Gear (lbs)	Oil Pressure - left inbd eng (psi)	Oil Temperature - left inbd eng (deg C)	Pitch Attitude - Capt or Only (deg)	Pitch Rate (deg/sec)	Roll Attitude - Capt or Only (deg)	Roll Rate (deg/sec)	Vertical Speed - best avail (ft/min)	Vibration Level - N1 - left inbd eng	Vibration Level - N1 - right inbd eng	Yaw Rate (deg/sec)
0	1	1	1	1	1	1	1	1	1	1	1	0	1	N/A	1	1	1	1	0	0	1	1	1	0
5	1	1	0	1	1	1	1	0	1	1	1	0	1	N/A	0	1	1	1	0	0	1	1	1	0
10	1	1	0	1	1	1	1	0	1	1	1	0	1	0	1	1	1	1	0	0	1	1	1	0
15	1	1	1	1	1	1	1	0	1	1	1	0	1	N/A	1	1	1	1	0	0	1	1	1	0
20	1	1	1	1	1	1	1	0	1	1	1	0	1	N/A	0	1	1	1	0	0	1	1	1	0
25	1	1	0	1	1	1	1	1	1	1	1	0	1	N/A	1	1	1	1	0	0	1	1	1	0
30	1	1	0	1	1	1	1	0	1	1	1	0	1	N/A	1	1	1	1	0	0	1	1	1	0
35	1	1	1	1	1	1	1	0	1	1	1	0	1	N/A	1	1	1	1	0	0	1	1	1	0
40	1	1	0	1	1	1	1	0	1	1	1	0	1	N/A	1	1	1	0	0	0	1	1	1	0
45	1	1	0	1	1	1	1	1	1	1	1	1	1	N/A	1	1	1	1	0	0	1	1	1	0
50	1	1	0	1	1	1	1	1	1	1	1	1	1	N/A	1	1	1	1	1	0	1	1	1	1
55	1	1	0	1	1	1	1	1	1	1	1	1	1	N/A	1	1	1	0	0	0	1	1	1	0
60	1	1	0	1	1	1	1	1	1	0	1	1	1	N/A	1	1	1	0	0	0	1	1	1	0
65	1	1	0	1	1	1	1	1	0	0	1	1	1	N/A	1	1	1	0	1	0	1	1	1	1
70	1	1	0	1	1	1	1	1	1	0	1	1	1	N/A	1	1	1	0	1	0	1	1	1	1
75	1	1	0	1	1	1	1	1	1	0	1	1	1	N/A	1	1	1	0	1	0	1	1	1	1
80	1	1	1	1	1	1	1	1	1	0	1	1	1	N/A	1	1	1	1	1	1	1	1	1	1
85	1	1	1	1	1	1	1	1	1	0	1	1	1	N/A	1	1	1	1	1	0	1	1	1	1
90	1	1	1	1	1	1	1	1	1	0	1	1	1	N/A	1	1	1	1	1	1	1	1	1	1
95	1	1	0	1	1	1	1	1	1	0	1	0	1	N/A	1	0	1	0	1	0	1	1	1	1
100	1	0	0	1	1	1	1	1	0	1	N/A	1	1	0	0	1	1	0	0	0	0	1	1	0
Significance Score	1.00	0.95	0.33	1.00	1.00	1.00	1.00	0.67	0.95	0.62	1.00	0.52	1.00	0.00	0.86	0.95	1.00	0.67	0.38	0.10	0.95	1.00	1.00	0.38

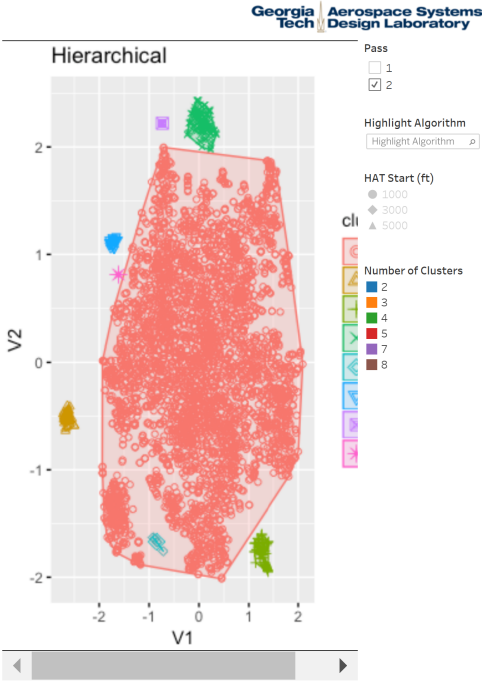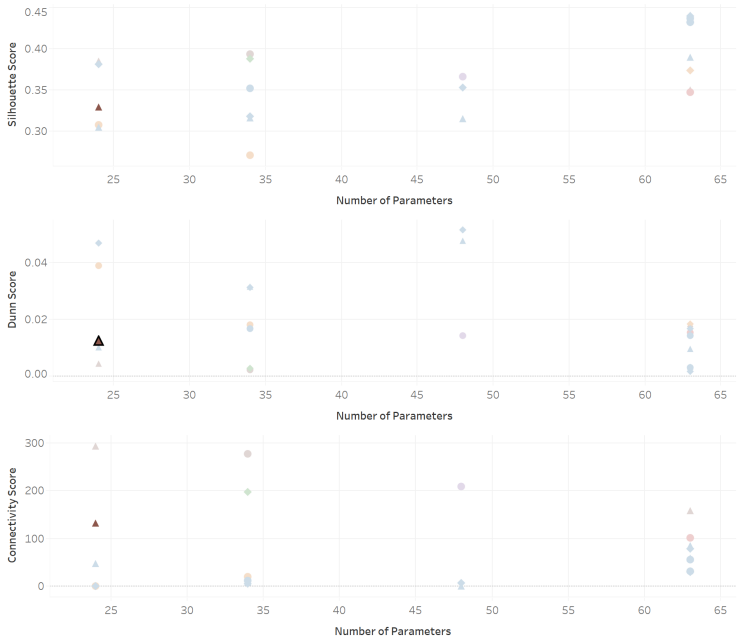Fig. 14   ANOVA summary table for 24 parameters, 5,000 ft, and 8 hierarchical clusters

Fig. 15 Clustering Metric Dashboard Showing the Performance of the Hierarchical Algorithm with 8 Clusters for the Second Pass of the 24 Parameter, 5,000 ft Approach Feature Matrix

Table 3 Cluster Characteristics for 24 Parameter, Last 5,000 ft, Hierarchical 8 Clusters Pass 2 Results

Cluster	Number of Flights	Cluster Characteristics
1	4,072	- Main cluster
2	43	- Steady N1/vibration until 2,500ft then increase in magnitude and variance in vibration - Predominantly stepped approaches with a step around 2,500ft - Landing gear deployed around 2,500ft - 3,000ft - Deceleration (airspeed decrease) around 2,500ft - Relatively lower median airspeed below 2,000ft - Vertical speed median around -1,000ft/min with some large positive outliers - Large positive glideslope deviation around 1,500ft then changes to large negative deviation at 50ft - potential overcorrection - Large positive localizer deviation around 700ft, corrected by around 300ft - Roll angle of -30deg around 1,000ft to 300ft
3	68	- Relatively more negative lateral acceleration throughout the approach
4	118	- Large oscillatory behavior in engine-related quantities - Likely go-around flights, as confirmed by trajectory plot
5	6	- Relatively low N1 and fuel flow - Relatively high oil temperature - Relatively low oil pressure
6	25	- Predominantly stepped approaches with a step around 3,000ft - Spike in pitch attitude at 3,000ft - Spike in N1 an other engine related quantities at 3,000ft - Flaps and landing gear deployed at 3,000ft - Deceleration (airspeed decrease) around 3,000ft - Relatively lower median airspeed below 3,000ft - Spike in vertical speed to 0ft/min at 3,000ft - Positive glideslope deviation at 3,000ft and until around 500ft - Localizer deviation at -3 dots from 3,000ft until around 600ft then returns to zero around 300ft
7	6	- Approach phase starts at around 2,000ft - Short final approaches; configuration set by 1,500ft
8	4	- High oscillatory longitudinal acceleration in the last 1,500ft - Localizer and glideslope deviations medians around 0, with very little variance

Table 4   Parameter ANOVA Significance for 24 Parameter, Last 5,000 ft, Hierarchical 8 Clusters Pass 2 Results

Parameter Name	ANOVA Significance Score
Acceleration (lateral) (g's)	1.000
Airspeed (true) (knots)	1.000
EGT (left inbd eng) (deg C)	1.000
Flap Position (deg+=TED)	1.000
Fuel Flow Total (lbs/hr)	1.000
Landing Gear Position (0 = up, 1 = in transit, 2 = down)	1.000
N1 (left inbd eng) (%)	1.000
Pitch Attitude (Capt or Only) (deg)	1.000
Vibration Level (N1; left inbd eng)	1.000
Vibration Level (N1; right inbd eng)	1.000
Ground Track Distance to Touchdown (nm)	0.990
Acceleration (longitudinal) (g's)	0.980
Vertical Speed (best avail) (ft/min)	0.980
Oil Temperature (left inbd eng) (deg C)	0.970
Oil Pressure (left inbd eng) (psi)	0.802
Glideslope Deviation (1 or only) (dots)	0.663
Pitch Rate (deg/sec)	0.644
Height Above Touchdown (HAT) (ft)	0.594
Localizer Deviation (1 or only) (dots)	0.545
Acceleration (normal load-factor) (g's)	0.366
Roll Attitude (Capt or Only) (deg)	0.366
Yaw Rate (deg/sec)	0.356
Roll Rate (deg/sec)	0.139
Normal Force through Combined Gear (lbs)	0.000

3. Details of Cluster Characteristics

This section highlights the key findings from the detailed cluster analysis dashboard (outlined in Section IV.G.2) that informed the cluster characteristics in Table 3. The figures that are referenced in this section are located in the Appendix (Section VI.C).

Cluster 1 is labelled as the main baseline cluster with 4,072 flights. However, it is interesting to note that there seems to be additional sub-clusters and outliers within cluster 1, as can be seen from the t-SNE plot in Figure 15. For example, the Exhaust Gas Temperature (EGT) data for cluster 1 in Figure 18 highlights one particular flight with significant anomalous readings. Further analysis of this cluster, including a third pass of re-clustering could reveal additional insights and groups of these remaining 4,072 flights in cluster 1.

Cluster 2 consists of 43 flights with some relatively distinct characteristics highlighted in Table 3. These flights are predominantly stepped approaches with a level-off segment at approximately 2,500 ft above touchdown. During this level-off at 2,500 ft, there is a large decrease in airspeed (deceleration), as highlighted in Figure 19. Furthermore, the figure also shows that the mean airspeed for this cluster is low when compared to the rest of the flights between 2,000 ft and touchdown. The glideslope deviation plots for this cluster (shown in Figure 20) also provide some insight. The majority of the flights in this cluster have a large positive glideslope deviation from 2,000 ft to approximately 1,500 ft. This large positive deviation then decreases to a large negative deviation at 50 ft before going back towards zero deviation at touchdown. This could suggest that these flights were too high after the level-off segment and then over-corrected to try and get back on the glideslope, potentially signifying an unstable approach.

Another interesting insight from cluster 2 comes from the localizer deviation (Figure 21) and the roll attitude (Figure 22) plots. The localizer deviation plot shows that the majority of the flights have a large positive deviation from 2,000 ft until approximately 700 ft. The aircraft then turns such that a roll attitude of -30 degrees is maintained from approximately 1,000 ft to 300 ft. The localizer deviation approaches zero at around 300 ft above touchdown. This suggests that these flights all made moderately steep turns very close to the ground (within 1,000 ft above touchdown). Further investigation of these flights may help determine the cause of this maneuver, i.e. if it was an air traffic constraint, airport-related constraint, operational constraint, or some other reason that could influence flight safety.

Cluster 3 is comprised of 68 flights that all have relatively more negative lateral acceleration throughout the approach phase as shown in Figure 23. Some of the flights even reach peaks of -0.5 g's of lateral acceleration during the approach. This analysis identified the critical parameter that led to the formation of this cluster, and these flights could be investigated further to determine the cause of the lateral acceleration anomalies. For example, the anomalies could be caused by high crosswinds or some other precursor.

Cluster 4 has 118 flights that display significant oscillatory behavior in the majority of parameters (oscillatory with respect to height above touchdown). The oscillatory behavior seems most common in the engine related parameters such as the N1 (fan speed percentage), as shown in Figure 24. Further investigation of this cluster's trajectories (Figure 25) show that the majority of the flights are likely go-around or missed approach flights.

Cluster 5 contains 6 flights. These flights have relatively low N1 and fuel flow, relatively high oil temperature, and relatively low oil pressure (all in the left engine). For example, the oil temperature values of this cluster relative to the others is displayed in the box plot and line graph in Figure 26. Further analysis could be conducted to identify the precursors and other contributing factors, such as oil quantity.

Cluster 6 is quite similar to cluster 2 in the fact that the majority of flights are stepped approaches. Cluster 6 has 25 flights that all have a level-off segment around 3,000 ft above touchdown, as illustrated in Figure 27. The majority of these flights undergo significant deceleration in this level-off phase, as shown in the airspeed plots in Figure 28. This level-off phase can also be seen by a spike in the vertical speed to 0ft/min at 3,000 ft, as shown by Figure 31. The engine fan speed plot in Figure 30 illustrates that the N1 was constant until the level-off segment at 3,000 ft, then increased to maintain altitude before decreasing and varying until touchdown. Moreover, the localizer deviation plots in Figure 29 show that the flights predominantly had a localizer deviation of -3 dots from 3,000 ft until around 600 ft and then a return to zero around 300 ft above touchdown. These were relatively large localizer deviations close to the touchdown point, potentially indicating an unstable approach or some other event. These flights could be investigated further to determine the root causes of this deviation.

Cluster 8 contains 4 flights, making it is a relatively small cluster. However, the flights have significant differences in the distribution of the longitudinal acceleration, demonstrated by the box plot in Figure 32. The line graph in the figure also suggests that the longitudinal acceleration was highly oscillatory during the last 1500 ft of the approach phase.

In summary, the detailed cluster analysis dashboard facilitated an interactive and thorough analysis of the different clusters as well as the parameters that were significant in the formation of these clusters. These results could be used to investigate some flights in more detail and to potentially identify new events and insight relevant to flight safety.

VI. Conclusion and Future Work

The use of machine learning techniques for safety analysis, incident and accident investigation, and fault detection has gained increased traction among the aviation community. In this paper, a robust and repeatable method for categorizing heterogeneous flight data obtained from operations and identifying critical parameters that are indicative of anomalies is identified. The methodology consists of a series of steps that can be repeated for similar data sets to yield valuable insights into operations and safety.

A systematic process of down-selecting parameters based on removal of highly correlated parameters, subject matter expert opinions, removal of discrete parameters and metadata categorization is demonstrated to yield smaller parameter subsets. Feature vectors are generated using the parameters subsets and variations in approach phase length. Following feature vector generation and feature matrix construction, a rigorous benchmarking of different clustering algorithms is demonstrated in order to identify the best algorithm and the optimal number of clusters in which to divide the available dataset. Finally, a one-way analysis of variance is conducted to identify the significance of various parameters in the clustering results. In order to use this work in a decision-making context, a dashboard that enables interactive analysis of the data is constructed and leveraged.

The implementation of the methodology on the dataset reveals various interesting clusters within the data analyzed. Along with a large cluster of nominal operations, a variety of smaller clusters are formed based on abnormal or unusual variations in the parameters. A rank-ordered list of parameters leads to the formation of these clusters providing further insight into why such clusters are formed and how they can aid in the analysis of results by safety experts.

In the future, this work can be readily extended to other phases of flight as well as other airframes. This can lead to a holistic understanding of the fleet-level operations and identify trends that could lead to problems or difficulties in the future. Another avenue of future work is to modify this approach to a time increment-based feature vector definition rather than a height above touchdown-based feature vector definition. Finally, future work can explore the possibility of incorporating subject matter expert feedback into the clustering process, thereby increasing the likelihood of uncovering potential safety issues and reducing the frequency of false positives.

Acknowledgments

References

[1] Campbell, N., "Flight Data Analysis - An Airline Perspective," *Australian and New Zealand Societies of Air Safety Investigators Conference*, 2003.

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

[3] "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.

[4] FAA, "Federal Aviation Administration - Aviation Safety Information Analysis and Sharing (ASIAS)," , 2017. URL http://www.asias.faa.gov, url: http://www.asias.faa.gov.

[5] Sheridan, K., Puranik, T. G., Mangortey, E., Pinon, O. J., Kirby, M., and Mavris, D. N., "An Application of DBSCAN Clustering For Flight Anomaly Detection During The Approach Phase," *AIAA SciTech Forum*, 2020.

[6] Puranik, T. G., and Mavris, D. N., "Anomaly Detection in General-Aviation Operations Using Energy Metrics and Flight-Data Records," *Journal of Aerospace Information Systems*, Vol. 15, No. 1, 2018, pp. 22–35. doi:10.2514/1.I010582.

[7] Puranik, T. G., and Mavris, D. N., "Identification of Instantaneous Anomalies in General Aviation Operations using Energy Metrics," *Journal of Aerospace Information Systems*, Vol. Article in Advance, 2019. doi:10.2514/1.I010772.

[8] 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.

[9] 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, 2015, pp. 587–598. doi:10.2514/1.I010329.

[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, 2013, pp. 467–475. doi:10.2514/1.I010080.

[11] Puranik, T. G., and Mavris, D. N., "Identifying Instantaneous Anomalies in General Aviation Operations," *17th AIAA Aviation Technology, Integration, and Operations Conference*, 2017. doi:10.2514/6.2017-3779, Paper No: AIAA-2017-3779.

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

[13] 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. doi:10.2514/1.I010394.

[14] Bendarkar, M. V., Behere, A., Briceno, S. I., and Mavris, D. N., "A Bayesian Safety Assessment Methodology for Novel Aircraft Architectures and Technologies using Continuous FHA," *AIAA Aviation 2019 Forum*, 2019. doi:10.2514/6.2019-3123.

[15] 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.

[16] Puranik, T., Jimenez, H., and Mavris, D., "Energy-Based Metrics for Safety Analysis of General Aviation Operations," *Journal of Aircraft*, Vol. 54, No. 6, 2017, pp. 2285–2297. doi:10.2514/1.C034196.

[17] Puranik, T. G., 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. doi:10.2514/6.2016-3915, Paper No. AIAA 2016-3915.

[18] 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. Paper No. AIAA 2016-3913, doi:10.2514/6.2016-3913.

[19] Puranik, T. G., "A Methodology for Quantitative Data-driven Safety Assessment for General Aviation," Ph.D. thesis, Georgia Institute of Technology, 2018. URL http://hdl.handle.net/1853/59905.

[20] "Federal Aviation Administration Advisory Circular 120–71A," https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC120-71A.pdf, 2003. Retrieved: 11/2019.

[21] Payan, A. P., Lin, P.-N., Johnson, C., and Mavris, D. N., "Helicopter Approach Stability Analysis Using Flight Data Records," *17th AIAA Aviation Technology, Integration, and Operations Conference*, 2017. doi:10.2514/6.2017-3437, Paper No: AIAA 2017-3437.

[22] Olive, X., and Basora, L., "Identifying Anomalies in past en-route Trajectories with Clustering and Anomaly Detection Methods," *Thirteenth USA/Europe Air Traffic Management Research and Development Seminar*, 2019.

[23] "Cockpit Voice Recorders (CVR) and Flight Data Recorders (FDR)," , 2019. URL https://www.ntsb.gov/news/Pages/cvr_fdr.aspx.

[24] "Aviation Safety Action Program," , 2019. URL https://www.faa.gov/about/initiatives/asap/.

[25] "Aviation Safety Reporting System," , 2019. URL https://asrs.arc.nasa.gov.

[26] Reynard, W., *The Development of the NASA Aviation Safety Reporting System*, Vol. 1114, National Aeronautics and Space Administration, 1986.

[27] Yu, L., and Liu, H., "Efficient Feature Selection via Analysis of Relevance and Redundancy," *J. Mach. Learn. Res.*, Vol. 5, 2004, pp. 1205–1224. URL http://dl.acm.org/citation.cfm?id=1005332.1044700.

[28] Raschka, S., and Mirjalili, V., *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow*, 2nd ed., Packt Publishing, 2017.

[29] Khalid, S., Khalil, T., and Nasreen, S., "A survey of feature selection and feature extraction techniques in machine learning," *2014 Science and Information Conference*, 2014, pp. 372–378. doi:10.1109/SAI.2014.6918213.

[30] Mangortey, E., Puranik, T., Pinon, O., and Mavris, D., "Classification, Analysis, and Prediction of the Daily Operations of Airports Using Machine Learning," *AIAA Science and Technology Forum (AIAA Scitech)*, 2020.

[31] Bezdek, J. C., Ehrlich, R., and Full, W., "FCM: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, Vol. 10, No. 2-3, 1984, pp. 191–203.

[32] Likas, A., Vlassis, N., and Verbeek, J. J., "The global k-means clustering algorithm," *Pattern recognition*, Vol. 36, No. 2, 2003, pp. 451–461.

[33] Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., and Wu, A. Y., "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, , No. 7, 2002, pp. 881–892.

[34] Guha, S., Rastogi, R., and Shim, K., "ROCK: A robust clustering algorithm for categorical attributes," *Information systems*, Vol. 25, No. 5, 2000, pp. 345–366.

[35] Struyf, A., Hubert, M., and Rousseeuw, P., "Score normalization in multimodal biometric systems," *Pattern Recognition, Volume 38, Issue 12, Pages 2270-2285*, 2005.

[36] Jain, A., Nandakumar, K., and Ross, A., "Score normalization in multimodal biometric systems," *Pattern recognition*, Vol. 38, No. 12, 2005, pp. 2270–2285.

[37] Patro, S., and Sahu, K. K., "Normalization: A preprocessing stage," *arXiv preprint arXiv:1503.06462*, 2015.

[38] Roweis, S. T., and Saul, L. K., "Nonlinear dimensionality reduction by locally linear embedding," *science*, Vol. 290, No. 5500, 2000, pp. 2323–2326.

[39] Tenenbaum, J. B., De Silva, V., and Langford, J. C., "A global geometric framework for nonlinear dimensionality reduction," *science*, Vol. 290, No. 5500, 2000, pp. 2319–2323.

[40] Van Der Maaten, L., Postma, E., and Van den Herik, J., "Dimensionality reduction: a comparative," *J Mach Learn Res*, Vol. 10, No. 66-71, 2009, p. 13.

[41] Maaten, L. v. d., and Hinton, G., "Visualizing data using t-SNE," *Journal of machine learning research*, Vol. 9, No. Nov, 2008, pp. 2579–2605.

[42] Wattenberg, M., Viégas, F., and Johnson, I., "How to use t-SNE effectively," *Distill*, Vol. 1, No. 10, 2016, p. e2.

[43] Van der Maaten, L., and Hinton, G., "Visualizing data using t-sne," *Journal of machine learning research*, Vol. 9, No. Nov, 2008.

[44] Bezdek, J. C., Hathaway, R. J., and Huband, J. M., "Visual assessment of clustering tendency for rectangular dissimilarity matrices," *IEEE Transactions on fuzzy systems*, Vol. 15, No. 5, 2007, pp. 890–903.

[45] Huband, J. M., Bezdek, J. C., and Hathaway, R. J., "bigVAT: Visual assessment of cluster tendency for large data sets," *Pattern Recognition*, Vol. 38, No. 11, 2005, pp. 1875–1886.

[46] Hathaway, R. J., Bezdek, J. C., and Huband, J. M., "Scalable visual assessment of cluster tendency for large data sets," *Pattern Recognition*, Vol. 39, No. 7, 2006, pp. 1315–1324.

[47] Bezdek, J. C., and Hathaway, R. J., "VAT: A tool for visual assessment of (cluster) tendency," *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)*, Vol. 3, IEEE, 2002, pp. 2225–2230.

[48] Banerjee, A., and Dave, R., "Validating clusters using the Hopkins statistic," *2004 IEEE International Conference on Fuzzy Systems (IEEE Cat. No.04CH37542)*, 2004.

[49] Bezdek, J., Hathaway, R., and Huband, J., "Visual Assessment of Clustering Tendency for Rectangular Dissimilarity Matrices," *IEEE Transactions on Fuzzy Systems, vol. 15, no. 5, pp. 890-903*, 2007.

[50] Statistical Tools For High-Throughput Data Analysis, "Assessing clustering tendency: A vital issue - Unsupervised Machine Learning," , 2008. URL http://www.sthda.com/english/wiki/print.php?id=238#a-single-function-for-hopkins-statistic-and-vat.

[51] Danielsson, P.-E., "Euclidean distance mapping," *Computer Graphics and image processing*, Vol. 14, No. 3, 1980, pp. 227–248.

[52] Wang, L., Zhang, Y., and Feng, J., "On the Euclidean distance of images," *IEEE transactions on pattern analysis and machine intelligence*, Vol. 27, No. 8, 2005, pp. 1334–1339.

[53] Salvador, S., and Chan, P., "Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms," *6th IEEE International Conference on Tools with Artificial Intelligence,pp. 576-584,doi: 10.1109/ICTAI.2004.50*, 2004.

[54] Olson, C., "Parallel algorithms for hierarchical clustering," *Parallel Computing 21, pp. 1313-1325*, 1995.

[55] Murtagh, F., and Legendre, P., "Ward's Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm," *arXiv e-prints*, 2011.

[56] Struyf, A., Hubert, M., and Rousseeuw, P., "Clustering in an Object-Oriented Environment," *Journal of Statistical Software, Volume 1, Issue 4*, 1997.

[57] Brock, G., Pihur, V., and Datta, S., "clValid: An R Package for Cluster Validation," *Journal of Statistical Software*, 2008.

[58] Lantz, B., *Machine Learning with R: Discover How to Build Machine Learning Algorithms, Prepare Data, and Dig Deep into Data Prediction Techniques with R*, Packt Publishing, 2015.

[59] Yeung, K. Y., Fraley, C., Murua, A., Raftery, A. E., and Ruzzo, W. L., "Model-based clustering and data transformations for gene expression data," *Bioinformatics*, Vol. 17, No. 10, 2001, pp. 977–987.

[60] Statistical tools for high-throughput data analysis (STHDA), "How to choose the appropriate clustering algorithms for your data? - Unsupervised Machine Learning," , 2019. URL http://www.sthda.com/english/wiki/print.php?id=243.

[61] Jun, S., "An Ensemble Method for Validation of Cluster Analysis," *IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 6, No 1*, 2011.

[62] Dunn, J., "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *J. Cybern.,vol. 3,pp. 32-57,1973*, 2008.

[63] J.Rousseeuw, P., "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics Volume 20, Pages 53-65*, 1987.

[64] Wu, C. J., and Hamada, M. S., *Experiments: planning, analysis, and optimization*, Vol. 552, John Wiley & Sons, 2011.

[65] Rao, A. H., and Puranik, T. G., "Retrospective Analysis of Approach Stability in General Aviation Operations," *18th AIAA Aviation, Technology, Integration, and Operations Conference, Atlanta, GA. June*, 2018. doi:10.2514/6.2018-3049.

Appendix

A. Summary of Parameter Sets Contained in Feature Vectors

Table 5   Parameter Sets for the Feature Vector Generation

Parameter	Set 0	Set 1	Set 2	Set 3
Acceleration (lateral) (g's)	1	1	1	1
Acceleration (longitudinal) (g's)	1	1	1	1
Acceleration (normal load-factor) (g's)	1	1	1	1
Aileron Position (asymmetric) (deg+=RWD)	1	1	1	
Air Density (ambient) (slugs/ft³)	1			
Air Temperature (outside) (deg C)	1			
Airspeed (true) (knots)	1	1	1	1
Angle of Attack (Best Available) (deg)	1	1	1	
Best Estimate of Terrain Elevation (ft)	1			
Coverage of Closest Cloud Layer	1	1		
Crosswind (knots)	1	1		
EGT (left inbd eng) (deg C)	1	1	1	1

Elevator Position (deg+=TEU)	1	1	1	
Flap Position (deg+=TED)	1	1	1	1
Flight Director Pitch (degrees)	1			
Flight Path Acceleration (g's)	1	1		
Flight Path Load-Factor (g's)	1	1		
FMS Lateral Deviation (1, Left, or only) (ft)	1	1		
FMS Vertical Deviation (1, Left, or only) (ft)	1	1		
Fuel Flow Total (lbs/hr)	1	1	1	1
Glideslope Deviation (1 or only) (dots)	1	1	1	1
Glideslope Deviation (2) (dots)	1	1		
Ground Speed (best avail) (knots)	1			
Ground Track Distance to Touchdown (nm)	1	1	1	1
Headwind (knots)	1	1		
Height Above Touchdown (HAT) (ft)	1	1	1	1
Horizontal Stabilizer (Pitch Trim Surface) Position (deg+=TEU)	1	1	1	
Hydraulic Pressure (A, 1, or Left) (psi)	1	1	1	
Hydraulic Pressure (B, 2, or Center) (psi)	1	1	1	
Landing Gear Position (0 = up, 1 = in transit, 2 = down)	1	1	1	1
Localizer Deviation (1 or only) (dots)	1	1	1	1
MSL Altitude of Closest Cloud Layer (ft)	1			
N1 (left inbd eng) (%)	1	1	1	1
N1 Target (left inbd eng) (%)	1			
Normal Force through Combined Gear (lbs)	1	1	1	1
Normal Force through Nose Gear (lbs)	1			
Oil Pressure (left inbd eng) (psi)	1	1	1	1
Oil Quantity (left inbd eng) (quarts)	1	1	1	
Oil Quantity (right inbd eng) (quarts)	1			
Oil Temperature (left inbd eng) (deg C)	1	1	1	1
Pitch Angle to Sun (degrees up from nose of aircraft)	1	1		
Pitch Attitude (Capt or Only) (deg)	1	1	1	1
Pitch Command (deg+=Nose Up)	1			
Pitch Rate (deg/sec)	1	1	1	1
Roll Attitude (Capt or Only) (deg)	1	1	1	1
Roll Rate (deg/sec)	1	1	1	1
Rudder Pedal Pos (+=Nose Right)	1			
Rudder Position (lower) (deg+=TER)	1	1	1	
Speed Brake Handle (deg)	1			
Speed Brake Position (deg += TEU)	1	1	1	
Spoiler Position Asymmetry (roll assist) (deg+=RWD)	1	1	1	
Sun Angle from Nose (degrees from nose of aircraft)	1	1		
Total Fuel Quantity (lbs)	1			
Vertical Separation from Closest Cloud Layer (ft)	1	1		
Vertical Speed (best avail) (ft/min)	1	1	1	1
Vertical Wind (ft/min)	1			

Vibration Level (N1; left inbd eng)	1	1	1	1
Vibration Level (N1; right inbd eng)	1	1	1	1
Vibration Level (N2; left inbd eng)	1	1		
Vibration Level (N2; right inbd eng)	1	1		
Yaw Angle to Sun (degrees to right of nose of aircraft)	1	1		
Yaw Rate (deg/sec)	1	1	1	1
Total number of parameters	**62**	**48**	**34**	**24**

B. Summary of clustering results

```
Algorithm      Metrics         2        3        4        5        6        7        8

hierarchical Connectivity    0.0000   0.1111  36.2107  70.7377  90.2607 100.4706 102.6246

             Dunn            0.2800   0.0562   0.0111   0.0132   0.0165   0.0168   0.0168

             Silhouette      0.4504   0.3364   0.3133   0.3381   0.3390   0.3192   0.3080

kmeans       Connectivity  108.2361 170.1397 222.3020 193.9306 240.2591 271.9734 278.6075

             Dunn            0.0022   0.0031   0.0014   0.0041   0.0049   0.0028   0.0039

             Silhouette      0.3417   0.3664   0.3536   0.3589   0.3778   0.3833   0.3890

pam          Connectivity  105.6583 157.0500 209.8623 234.2238 238.5563 254.0575 300.7655

             Dunn            0.0015   0.0022   0.0024   0.0024   0.0018   0.0017   0.0039

             Silhouette      0.3413   0.3619   0.3313   0.3674   0.3697   0.3837   0.3658

diana        Connectivity  106.1075 168.9401 216.5250 216.5250 240.2706 291.8333 312.7409

             Dunn            0.0029   0.0006   0.0007   0.0010   0.0010   0.0010   0.0011

             Silhouette      0.3424   0.3219   0.3510   0.3544   0.3403   0.3259   0.3321

sota         Connectivity  106.6310 162.0627 227.9448 274.4302 305.2183 311.1579 350.4698

             Dunn            0.0012   0.0012   0.0015   0.0013   0.0013   0.0013   0.0013

             Silhouette      0.3390   0.3102   0.3467   0.3178   0.3252   0.3188   0.3138

clara        Connectivity  112.8218 214.7056 255.7615 195.7667 268.3810 276.3766 309.7071

             Dunn            0.0012   0.0033   0.0027   0.0025   0.0013   0.0020   0.0017

             Silhouette      0.3404   0.3126   0.3097   0.3480   0.3635   0.3609   0.3433

model        Connectivity   62.7901 220.5393 164.7067 207.4143 310.9389 263.1421 253.5532

             Dunn            0.0019   0.0006   0.0016   0.0014   0.0020   0.0016   0.0026

             Silhouette      0.2898   0.3162   0.2016   0.2681   0.2832   0.2862   0.2658

Optimal Scores:

 Metric        Score     Method       Clusters

Connectivity 0.0000   hierarchical      2

Dunn         0.2800   hierarchical      2

Silhouette   0.4504   hierarchical      2
```

Fig. 16   Identification of appropriate clustering algorithm and optimal number of clusters for feature vector generated with 24 parameters, 5,000 ft, first pass

```
Algorithm    Metrics          2        3        4        5        6        7        8

hierarchical Connectivity   46.7905  97.7187 104.2230 104.2230 104.2230 122.5325 132.4607

             Dunn            0.0102   0.0092   0.0098   0.0098   0.0098   0.0100   0.0123

             Silhouette      0.3041   0.3401   0.3286   0.3185   0.3018   0.3189   0.3291

kmeans       Connectivity  110.1254 142.9385 231.9028 222.6274 259.6540 282.3921 293.5496

             Dunn            0.0014   0.0025   0.0026   0.0030   0.0040   0.0038   0.0045

             Silhouette      0.3247   0.3696   0.3537   0.3613   0.3611   0.3828   0.3845

pam          Connectivity  108.6071 145.3302 218.1437 221.1901 243.4468 252.0024 312.9766

             Dunn            0.0025   0.0033   0.0022   0.0023   0.0027   0.0037   0.0026

             Silhouette      0.3220   0.3695   0.3444   0.3617   0.3721   0.3843   0.3705

diana        Connectivity  113.0361 157.1028 203.4075 220.1968 253.1571 300.9488 309.3873

             Dunn            0.0034   0.0038   0.0049   0.0054   0.0057   0.0058   0.0058

             Silhouette      0.3249   0.3051   0.3285   0.3258   0.2882   0.2799   0.2792

sota         Connectivity  137.9048 200.4393 249.4552 277.8790 333.2802 382.9270 397.7563

             Dunn            0.0018   0.0018   0.0023   0.0027   0.0015   0.0015   0.0015

             Silhouette      0.3232   0.2804   0.3178   0.3298   0.2981   0.2926   0.2740

clara        Connectivity   94.0016 169.4349 267.9183 272.9837 291.4726 298.9865 360.3492

             Dunn            0.0023   0.0005   0.0028   0.0027   0.0030   0.0050   0.0018

             Silhouette      0.3246   0.3613   0.3081   0.2975   0.3356   0.3613   0.3424

model        Connectivity  124.1881 160.2353 204.0825 212.9687 251.6087 256.8107 301.3294

             Dunn            0.0021   0.0040   0.0014   0.0011   0.0002   0.0011   0.0016

             Silhouette      0.3202   0.2786   0.2687   0.3295   0.3092   0.2643   0.2221

Optimal Scores:

             Score    Method       Clusters

Connectivity 46.7905 hierarchical 2

Dunn          0.0123 hierarchical 8

Silhouette    0.3845 kmeans       8
```

Fig. 17   Identification of appropriate clustering algorithm and optimal number of clusters for feature vector generated with 24 parameters, 5,000 ft, second pass

C. Figures for Details of Cluster Characteristics
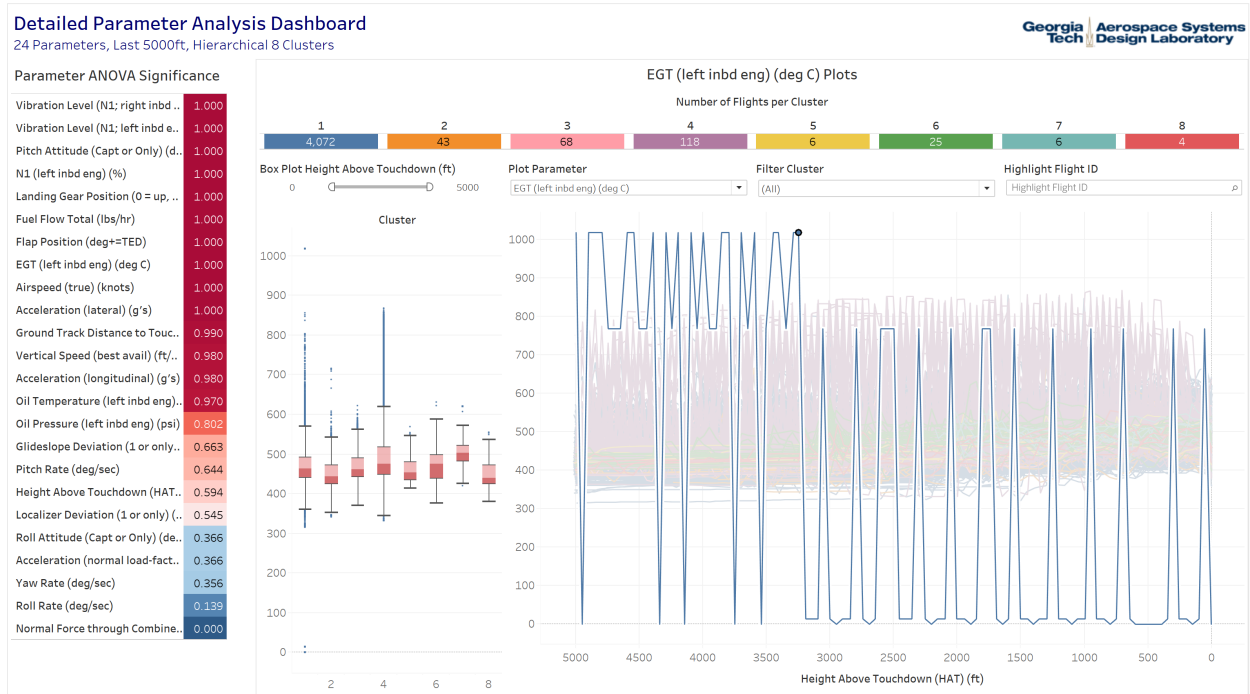
Fig. 18   Cluster 1 EGT Anomaly
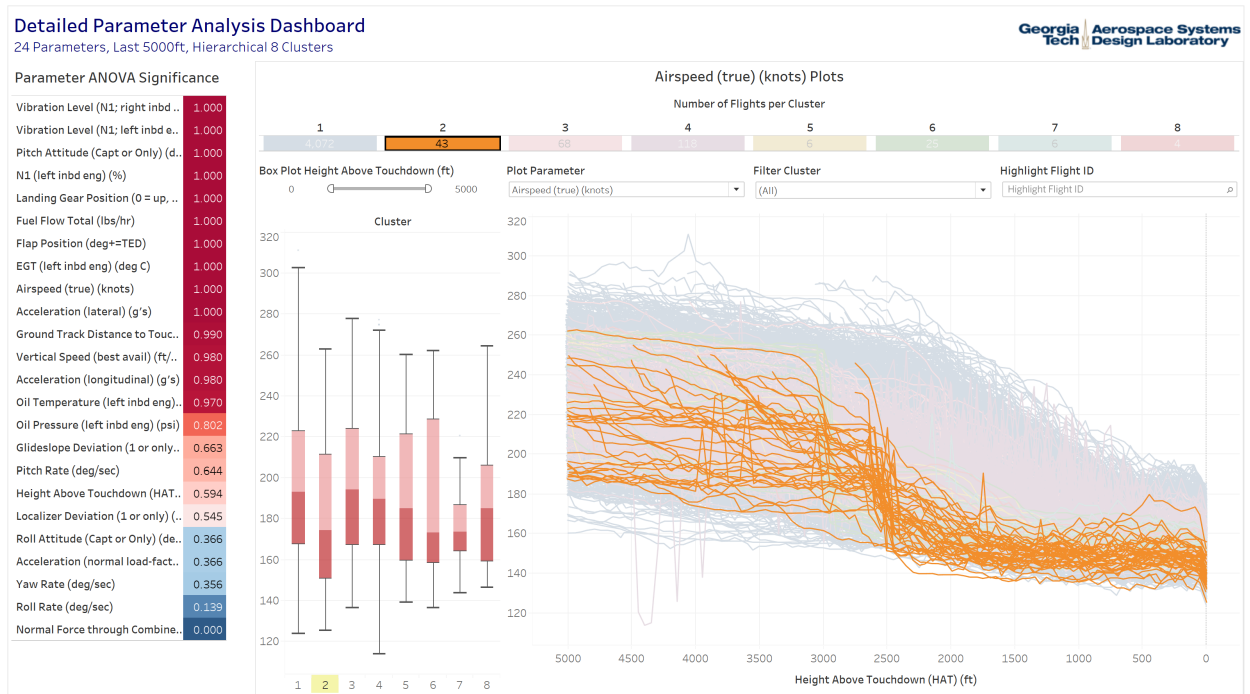
Fig. 19   Cluster 2 Airspeed
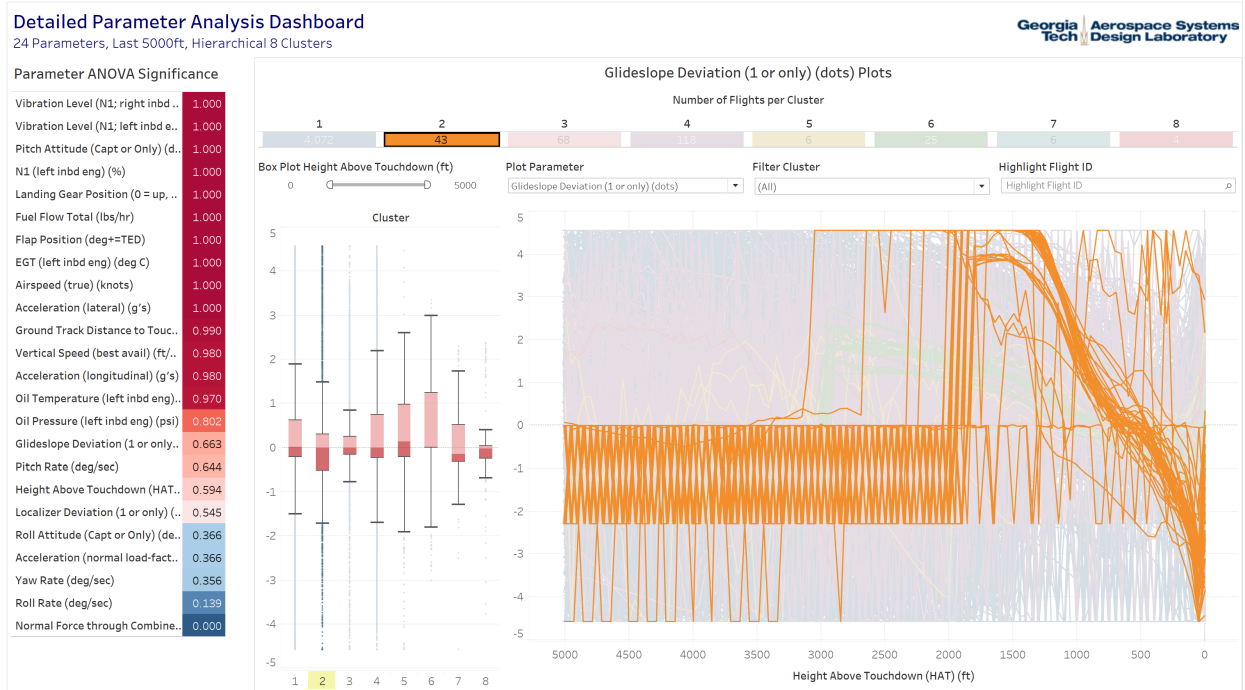
Fig. 20   Cluster 2 Glideslope Deviation
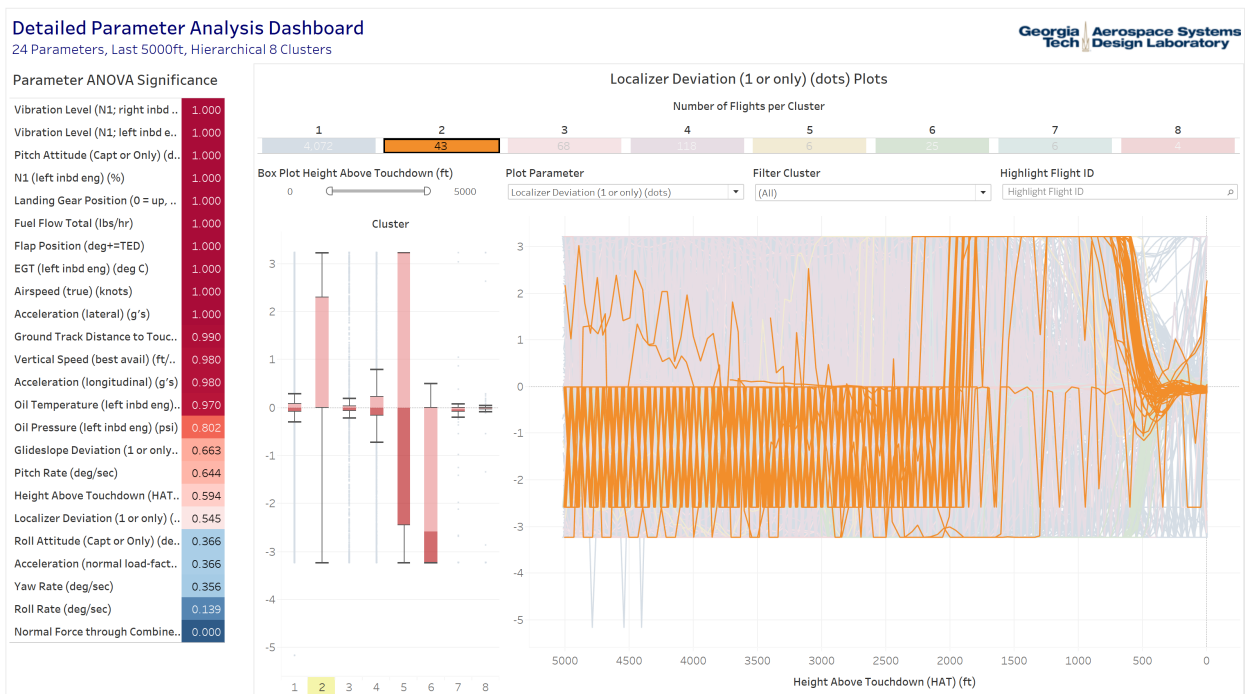
Fig. 21   Cluster 2 Localizer Deviation
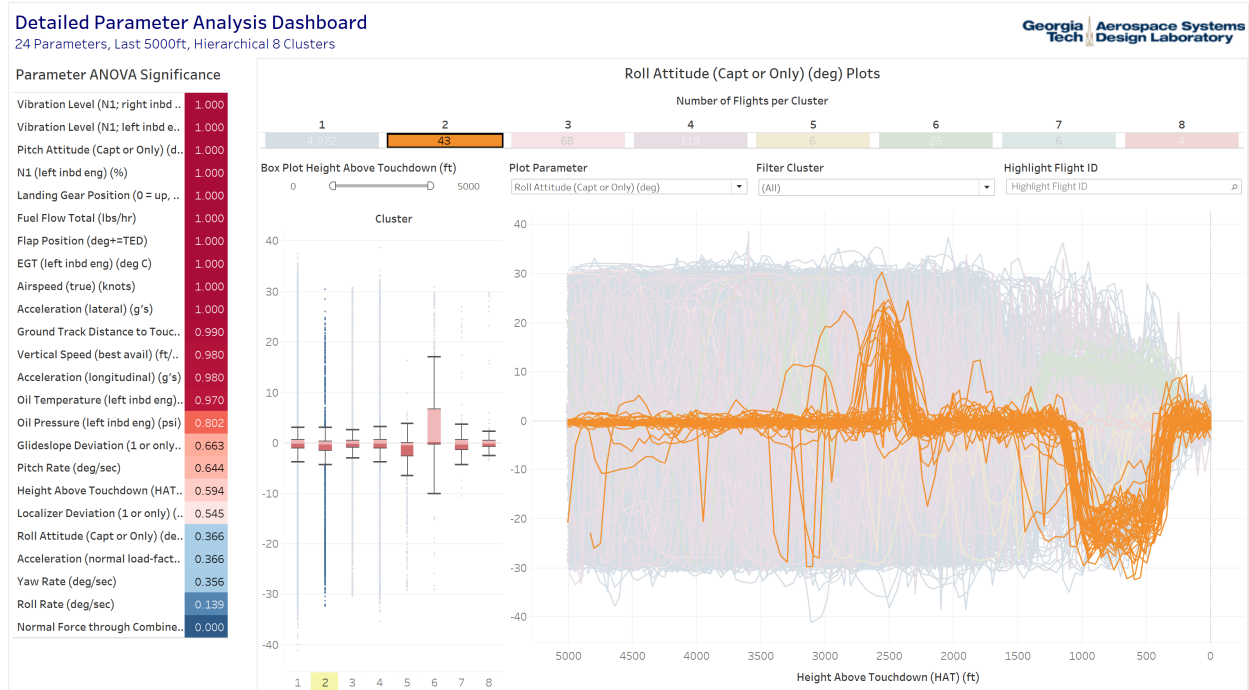
Fig. 22   Cluster 2 Roll Attitude
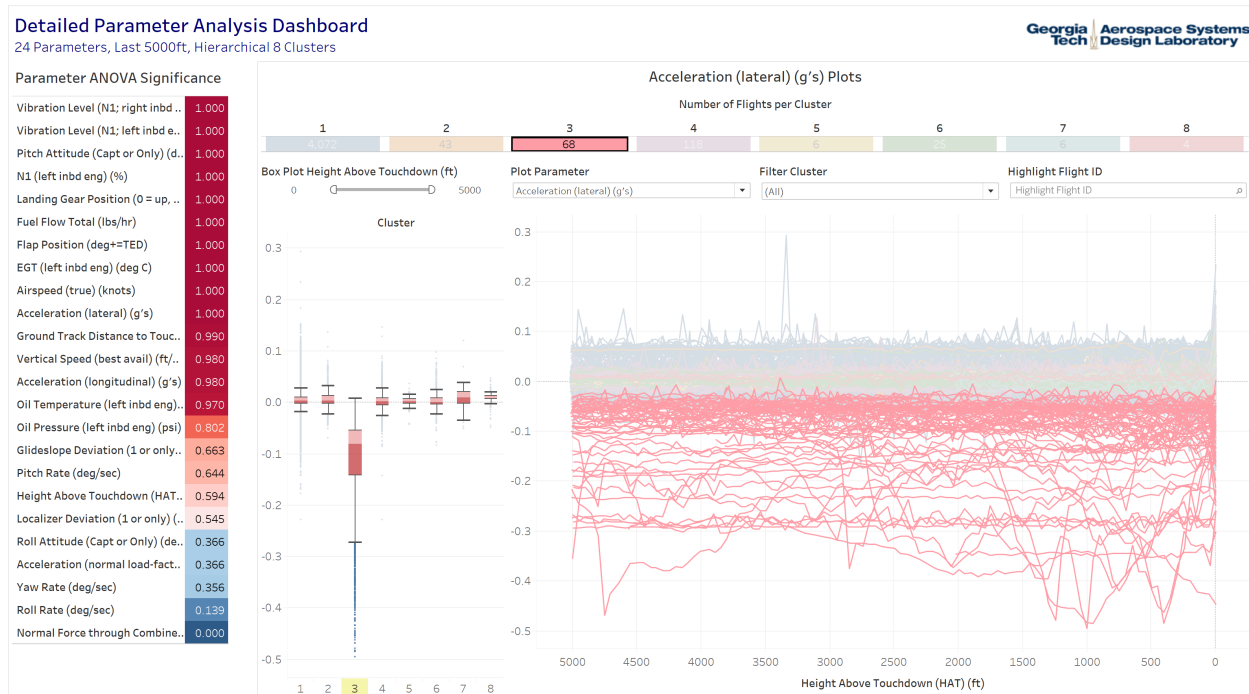
Fig. 23   Cluster 3 Lateral Acceleration
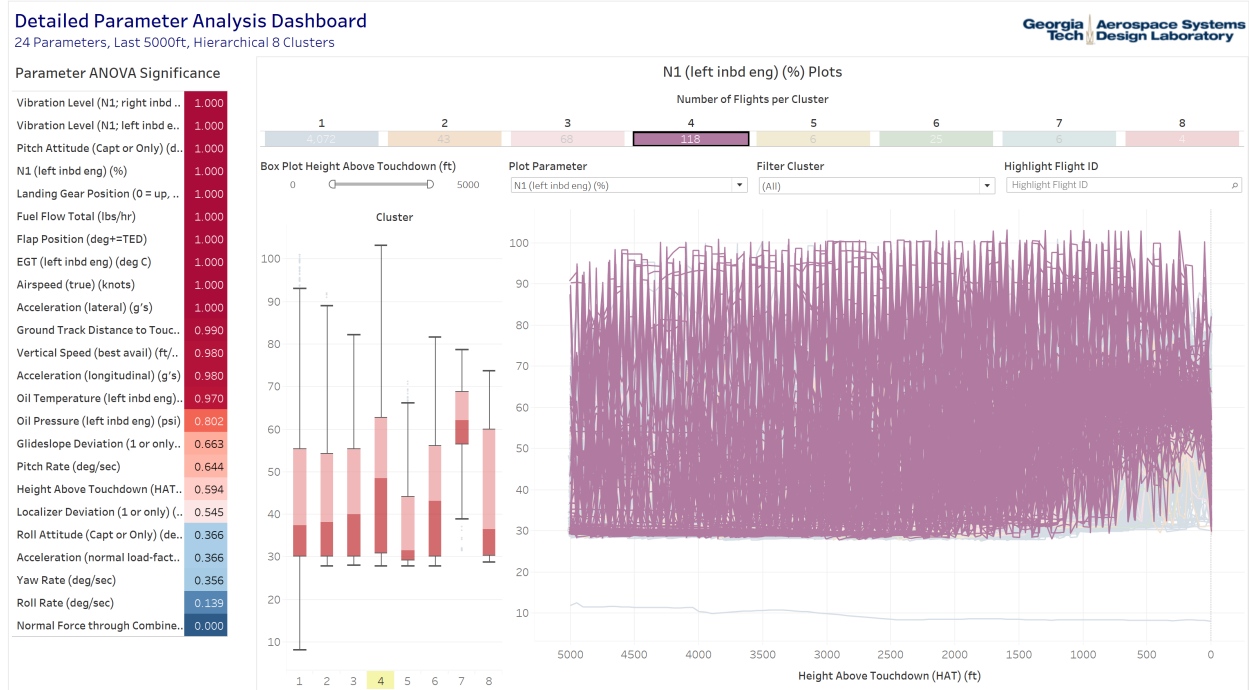
Fig. 24   Cluster 4 N1
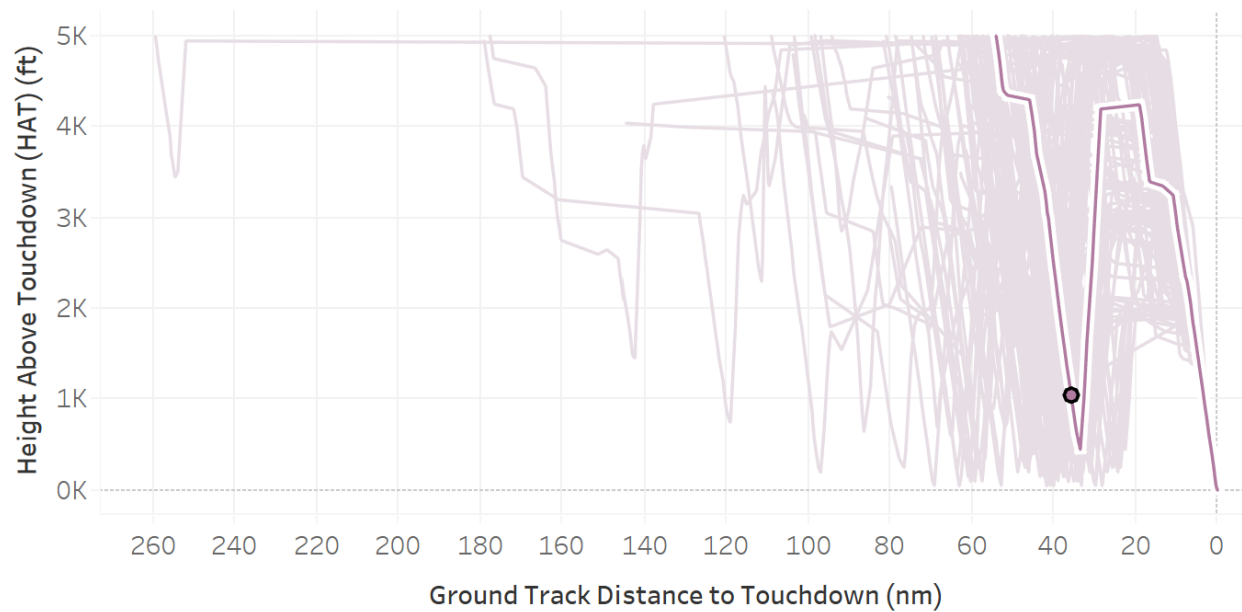
Fig. 25   Cluster 4 Trajectories
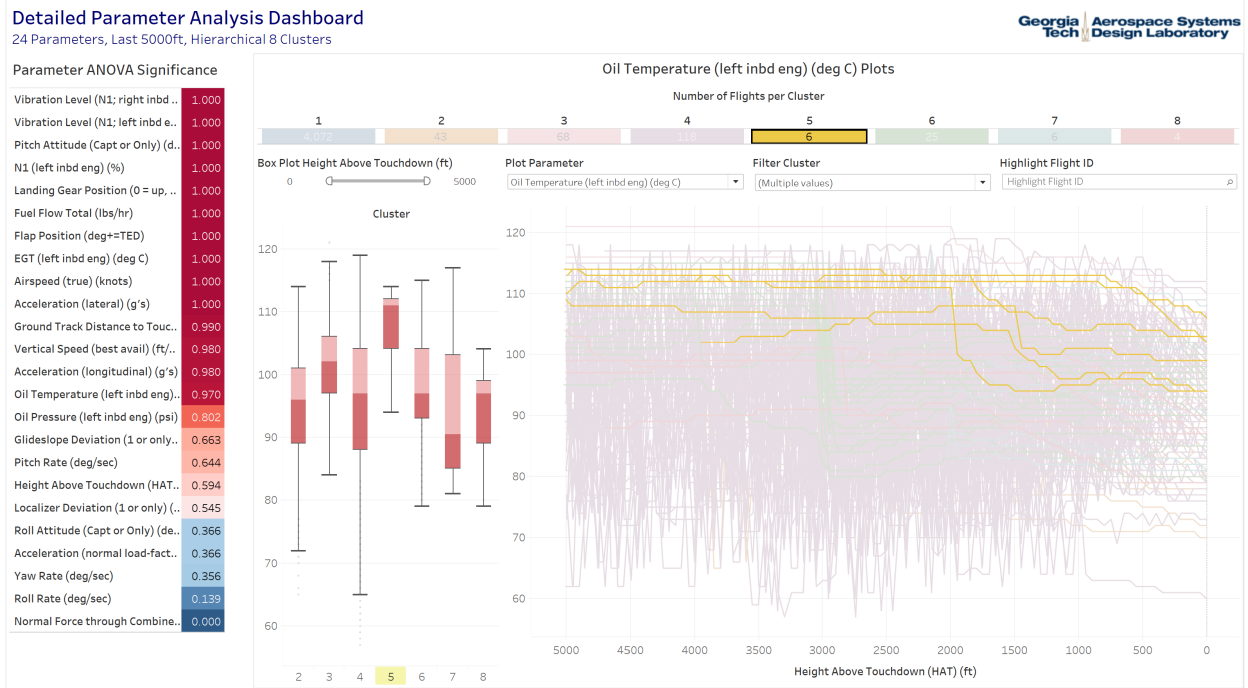
Fig. 26   Cluster 5 Oil Temperature
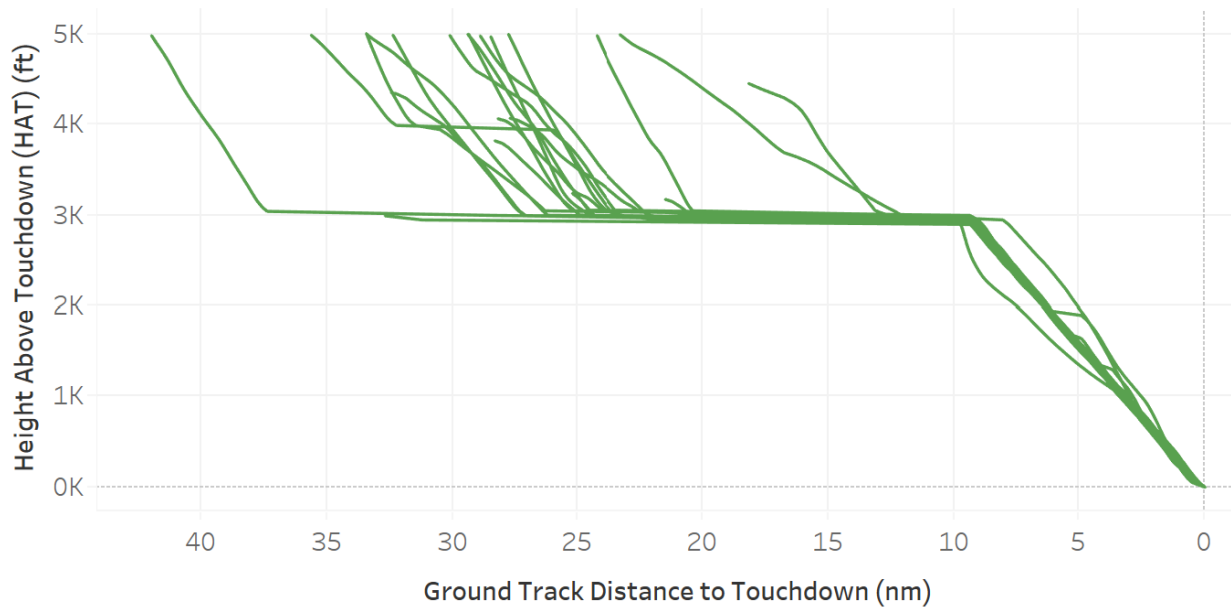
Fig. 27   Cluster 6 Trajectories
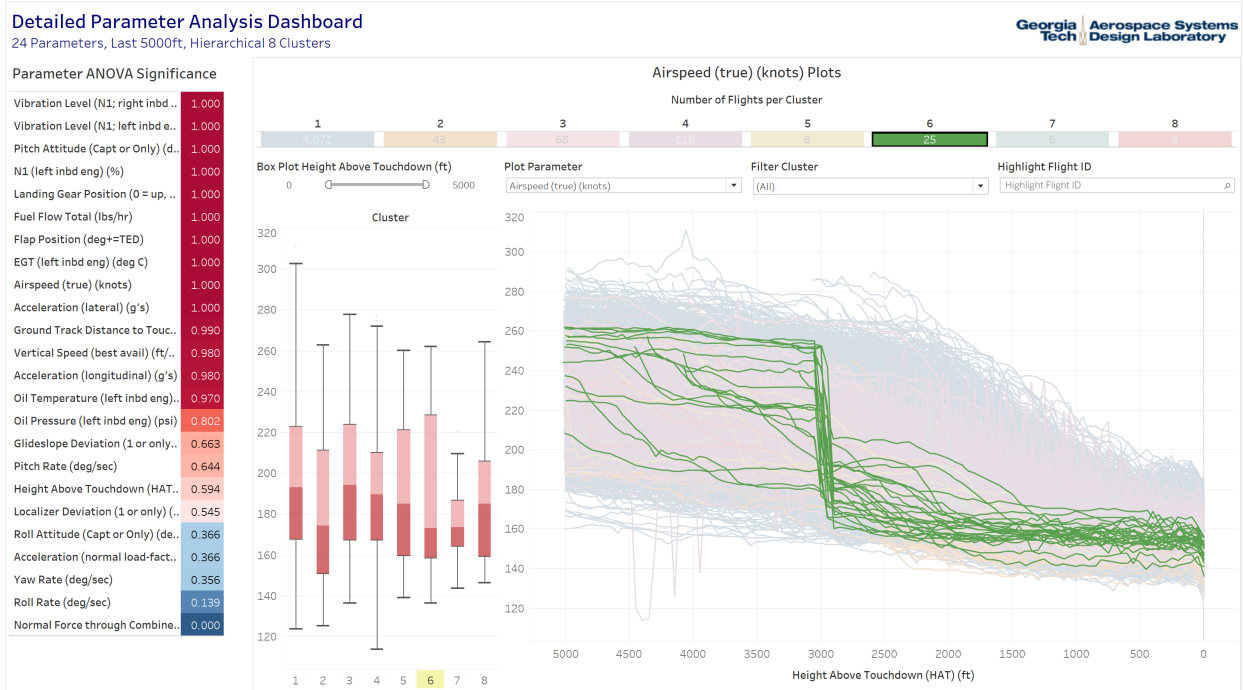
Fig. 28　Cluster 6 Airspeed
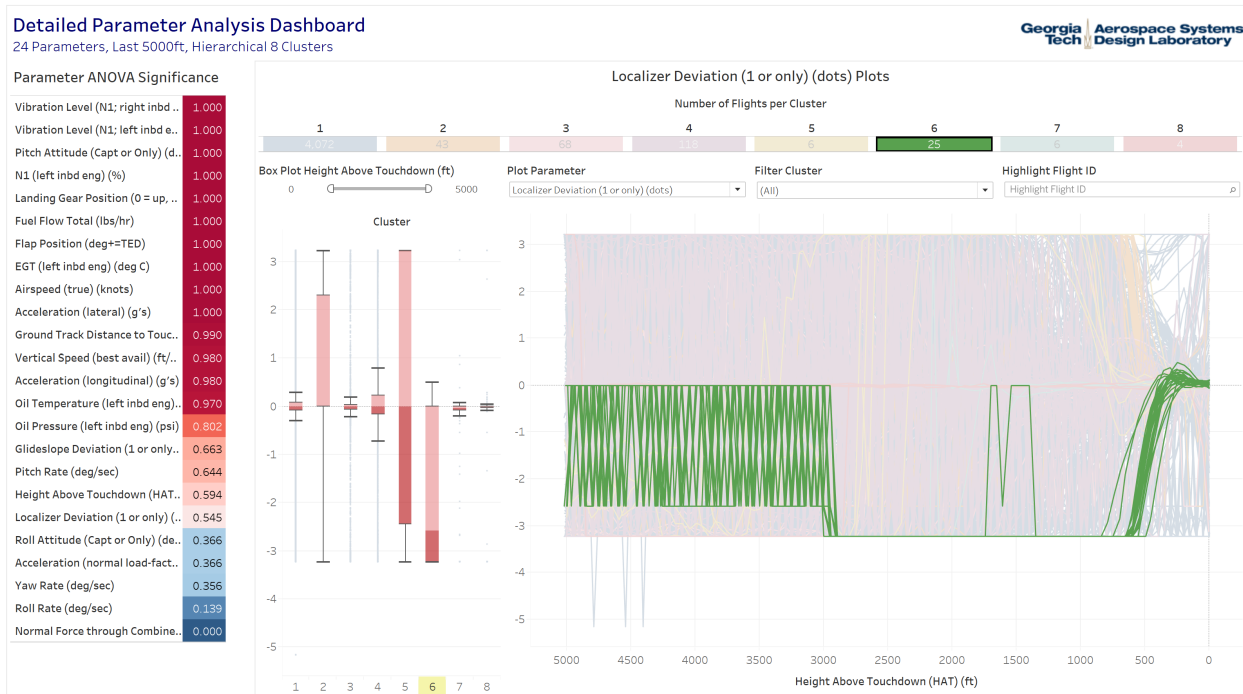
Fig. 29　Cluster 6 Localizer Deviation
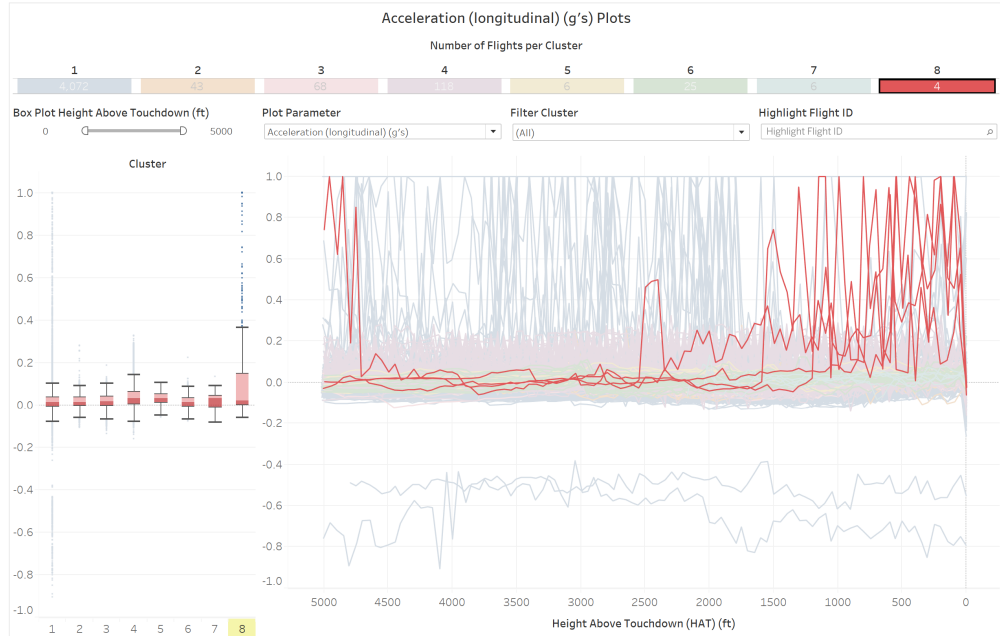
Fig. 30   Cluster 6 N1

Fig. 31   Cluster 6 Vertical Speed

Fig. 32   Cluster 8 Longitudinal Acceleration