Analysis of Flight Data Using Clustering Techniques for Detecting Abnormal Operations

Article *in* Journal of Aerospace Computing, Information and Communication · September 2015

DOI: 10.2514/1.I010329

Lishuai Li
City University of Hong Kong

R. John Hansman
Massachusetts Institute of Technology

Rafael Palacios
Universidad Pontificia Comillas

Ashok N. Srivastava
Verizon Communications

Journal of Aerospace Information Systems
Vol. 12, No. 9, September 2015

Analysis of Flight Data Using Clustering Techniques for Detecting Abnormal Operations

Lishuai Li*
City University of Hong Kong, Kowloon, Hong Kong, People's Republic of China
Santanu Das†
Verizon Labs, Palo Alto, California 94301
R. John Hansman‡
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
Rafael Palacios§
Comillas Pontifical University, 28015 Madrid, Spain
and
Ashok N. Srivastava¶
Verizon Labs, Palo Alto, California 94301

DOI: 10.2514/1.I010329

The airline industry is moving toward proactive risk management, which aims to identify and mitigate risks before accidents occur. However, existing methods for such efforts are limited. They rely on predefined criteria to identify risks, leaving emergent issues undetected. This paper presents a new method, cluster-based anomaly detection to detect abnormal flights, which can support domain experts in detecting anomalies and associated risks from routine airline operations. The new method, enabled by data from the flight data recorder, applies clustering techniques to detect abnormal flights of unique data patterns. Compared with existing methods, the new method no longer requires predefined criteria or domain knowledge. Tests were conducted using two sets of operational data consisting of 365 B777 flights and 25,519 A320 flights. The performance of cluster-based anomaly detection to detect abnormal flights was compared with those of multiple kernel anomaly detection, which is another data-driven anomaly detection algorithm in recent years, as well as with exceedance detection, which is the current method employed by the airline industry. Results showed that both cluster-based anomaly detection to detect abnormal flights and multiple kernel anomaly detection were able to identify operationally significant anomalies, surpassing the capability of exceedance detection. Cluster-based anomaly detection to detect abnormal flights performed better with continuous parameters, whereas multiple kernel anomaly detection was more sensitive toward discrete parameters.

Nomenclature

K = number of principal components kept after dimension reduction
m = number of flight parameters
N = total number of principal components
n = number of samples for every flight parameter
$\boldsymbol{v}$ = high-dimensional vector to represent a flight
x_j^i = value of the ith flight parameter at sample time j
ε = maximum distance between two points for them to be considered in the same cluster
λ_i = variance explained by principal component i

I. Introduction

IN HISTORY, improvements in airline safety were made from hard lessons. Accidents triggered the development and implementation of mitigation strategies [1]. After in-depth analysis of accidents, root causes were identified; corrective actions were proposed to prevent similar events from occurring. Recently, the airline industry has moved toward a proactive approach to safety: this approach identifies risks and informs actions before accidents occur. The digital flight data recorder (FDR) data are routinely analyzed by many airlines with the purpose of identifying risks [2]. However, current data analysis methods can only reveal a fraction of the information embedded in the FDR dataset.

FDR data consist of tens to thousands of flight parameters recorded throughout a flight. These parameters include altitude, airspeed, accelerations, thrust, engine pressures, engine temperatures, control surfaces, autopilot modes, etc. The flight operations quality assurance (FOQA) program, also called flight data monitoring in Europe, aims to use FDR data to improve safety in airline operations [2]. In current FOQA programs, the analysis of flight data is conducted using special-purpose software based on the approach called "exceedance detection" [2], in

Presented as Paper 2012-5593 at the 2011 IEEE/AIAA 30th Digital Avionics Systems Conference (DASC), Seattle, WA, 16–20 October 2011; received 4 September 2014; revision received 28 May 2015; accepted for publication 28 July 2015; published online 25 September 2015.

*Assistant Professor, Department of Systems Engineering and Engineering Management, Tat Chee Avenue. Member AIAA.
†Member AIAA.
‡Professor, Department of Aeronautics and Astronautics, 77 Massachusetts Avenue. Fellow AIAA.
§Tenured Assistant Professor, Institute for Research in Technology, CL Alberto Aguilera 23.
¶Associate Fellow AIAA.

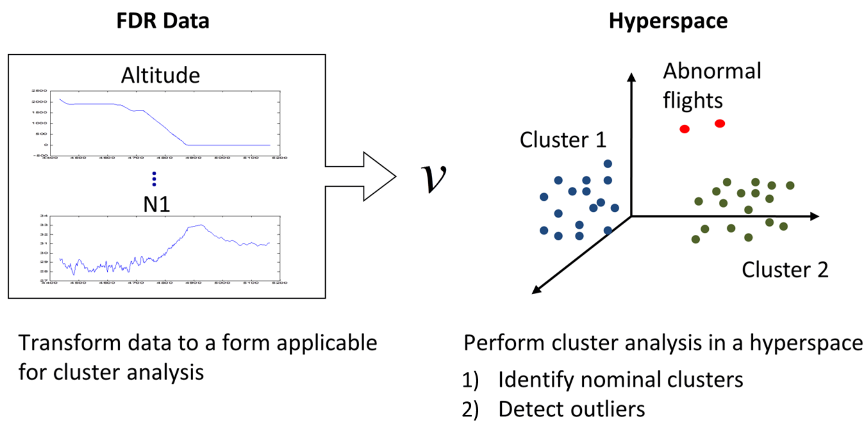


Fig. 1 Concept of cluster-based anomaly detection.

which selected flight parameters are compared with predefined thresholds. Exceedance events are detected when a parameter exceeds such thresholds. The exceedance detection method performs well on known safety issues, but it is incapable of identifying unknown issues.

Relevant literature exist in two groups: each has its limitations and cannot be readily employed in airline operations. The first group focuses on anomaly detection methods for aviation systems. The Morning Report software package was one of the earliest efforts made to detect anomalies from routine FDR data [3]. The software models the time-series data of selected flight parameters using a quadratic equation. Each flight is mapped into a point that is described by the coefficients of the quadratic equations in the feature space. An "atypical score" for each flight is calculated using the distance between the point and the mean of the distribution in the feature space. This method is innovative yet relatively simple (quadratic equations to model time series). Later studies apply data-mining techniques to detect data anomalies in aerospace systems [4–9]. Some adopt the supervised learning methods [7], such as the Inductive Monitoring System (IMS) software, which summarizes the data distributions of typical system behaviors from a presanitized training dataset, which is then compared with real-time operational data to detect abnormal behaviors. However, the IMS is limited in accounting for the temporal patterns and it cannot function without a training dataset labeling the norms. Others adopt the unsupervised approach. The sequence miner algorithm focuses on discrete flight parameters to monitor pilot operations, such as cockpit switch flips [4,5]. The algorithm can discover abnormal sequences in the switch operations based on the longest common subsequence measures. Srivastava developed a statistical framework to incorporate both discrete and continuous flight parameters in FDR data [6]. Built on this framework, Das et al. developed multiple kernel anomaly detection (MKAD), which applies a one-class support vector machine for anomaly detection [9]. MKAD assumes one type of data pattern for normal operations, which is not always valid in real operations, since standards vary according to flight conditions. Moreover, how to characterize the temporal structure during various flight phases remains unresolved. Most recently, Matthews et al. summarized the knowledge discovery pipeline for aviation data using the previously discussed algorithms [10].

The second group of literature consists of anomaly detection techniques emerged in multiple domains other than aviation. Most anomaly detection techniques can only solve problems of a domain-specific formulation [11,12]. For example, specific techniques are developed for intrusion detection in computer systems [13–15]; fault detection in mechanical units and structures [16,17]; and fraud detection related to credit cards, mobile phones, insurance claims [18,19], etc. Additionally, two groups of techniques are developed for time-series data depending on how dissimilarities are measured: data based and model based [20,21]. The former measures the dissimilarity based on data observations. The dissimilarity is measured by a variety of distance functions such as Euclidean distance [22,23], dynamic time warping distance [24], probability-based distance [25], correlation-based distance [26], attribute-based distance [27], etc. The latter uses the temporal structure to construct models. Time series are then compared with each other based on model parameters or residuals. The regression-based model is the earliest and is widely used [28–30], whereas the Markovian models are popular in analyzing sequential behaviors in time series that are not perfectly aligned [31–33].

This study aims to develop a new data-driven method to identify potentially emergent safety issues from routine airline operations. The objective is to detect abnormal flights using cluster analysis, without knowing the standard of the norm. Compared with existing methods, the advantages of the new method lie in that it can 1) detect unknown issues; 2) operate across a range of applications (e.g., another airline, aircraft type) free of initial configuration or toning; and 3) better recognize abnormal behaviors by allowing multiple types of standard operations. Once detected, abnormal flights will be handed over to domain experts to identify latent risks and inform operations before accidents occur.

II. Cluster-Based Anomaly Detection Method

This paper presents a new method known as ClusterAD-Flight, which stands for cluster-based anomaly detection to detect abnormal flights.

A. Concept of Cluster-Based Anomaly Detection

ClusterAD-Flight is based on cluster analysis, which is a commonly used data-mining technique to identify common patterns in a dataset. We assume that a majority of flights exhibit common patterns under routine operations; a few outliers that deviate from those common patterns are of interest to airline safety management. The first step is to transform FDR data into high-dimensional vectors, which capture the multivariate and temporal characteristics of each flight. In the second step, the dimensions of the aforementioned vectors are then reduced to address issues related to data sparseness and multicollinearity. The third step is to apply cluster analysis on the aforementioned vectors of reduced dimensions. Groups of proximate vectors are clusters, or the common patterns; standalone vectors are anomalies, or abnormal flights as illustrated in Fig. 1. This paper applies ClusterAD-Flight in analyzing the flight phases of takeoff and final approach, which are critical to safety, since 53% of fatal accidents and 47% of onboard fatalities happened during those two phases for the worldwide commercial jet fleet from 2002 to 2011 [34].

B. Process and Techniques of ClusterAD-Flight

The ClusterAD-Flight method consists of three key steps, as illustrated in Fig. 2:

1) Step 1 is data transformation, which is transforming the time series into high-dimensional vectors.
2) Step 2 is dimension reduction, which is addressing problems of data sparseness and multicollinearity.
3) Step 3 is cluster analysis, which is identifying clusters and outliers in high-dimensional space.

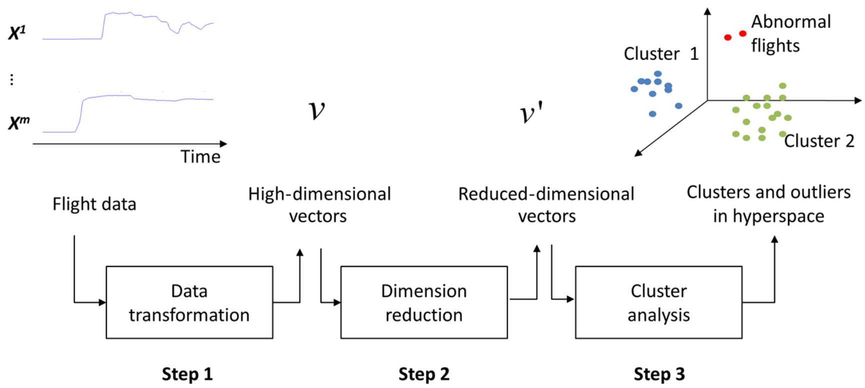


Fig. 2 Three key steps in ClusterAD-Flight.

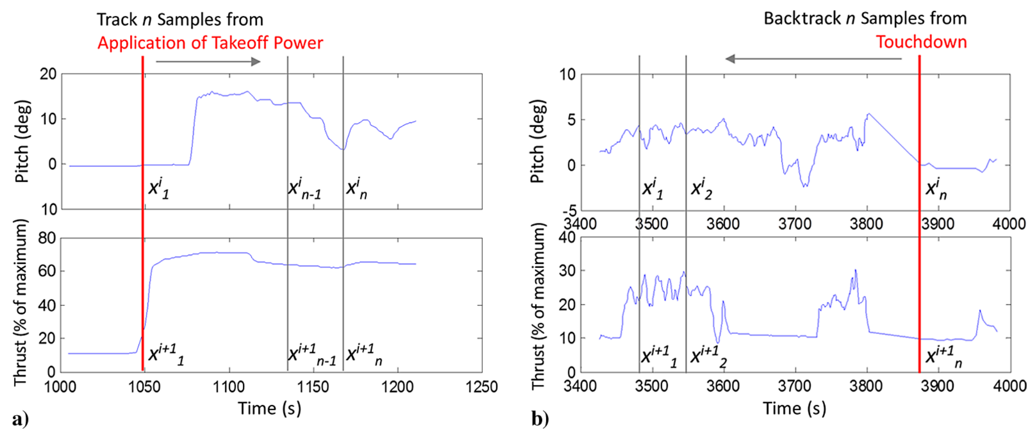


Fig. 3 Sampling time series: a) takeoff phase, and b) approach phase.

1. Step 1: Data Transformation

Data preprocessing is performed before the first step to clean the raw FDR data, normalize the continuous flight parameters, and binarize the discrete ones. Continuous parameters are normalized to have "zero mean and unit variance." The discrete parameters are transformed into Boolean features so that they can be treated as a vector in the Euclidean space together with continuous parameters.

FDR data are then mapped into comparable vectors in a high-dimensional space, anchored by a specific event in time. Data of different flights become comparable, since each flight parameter is sampled at fixed temporal or distance-based intervals starting from the anchoring event. All sampled values are arranged to form a vector for each flight:

$$\boldsymbol{v} = [x_1^1, x_2^1, \ldots, x_n^1, \ldots, x_j^i, \ldots, x_n^m] \tag{1}$$

where x_j^i is the value of the ith flight parameter at sample time j. The total dimensionality of every vector is m(the number of flight parameters) $\times$ n(the number of samples for every flight parameter). The Euclidean distance between the vectors measures the dissimilarity between flights: greater distance means less similarity.

Figure 3 shows how the takeoff and approach phases are anchored. During the takeoff phase, the anchor is the point in time when takeoff power is applied; each flight is sampled at a fixed temporal interval (e.g., 1 s) for all parameters. For the approach phase, the anchor is the point in distance of touchdown; each flight is sampled at a fixed interval in distance instead of in time. The reason is because procedures during the approach phase are often distance specific or height specific rather than time specific.

2. Step 2: Dimension Reduction

The vectors formed in the first step usually feature thousands of dimensions, which are products of the number of flight parameters timed by the number of sampling points. For example, we will arrive at an analysis space with 10,000 dimensions if we measure 100 parameters over 100 time steps. However, flights are sparsely distributed across dimensions, making it difficult to identify data clusters. To consolidate the data, we use the principal component analysis (PCA), which is a commonly used procedure to transform data into a new orthogonal coordinate system [35], to reduce the number of dimensions. The coordinates in the new system, referred to as components, are ranked by the amount of embedded information, e.g., the first component contains the largest variance, the second contains the next largest variance, and so on. We keep the first few components with the majority of the information, therefore reducing the number of dimensions. In this study, the first K components that capture 90% of the variance in the data are kept:

$$\frac{\sum_{i=1}^{K} \lambda_i}{\sum_{i=1}^{N} \lambda_i} > 90\% \tag{2}$$

where λ_i is the variance explained by principal component i. N is the total number of principal components, which equals the original number of dimensions. K is the number of principal components kept. The magnitude of dimensional reduction varies by dataset but can be significant. In this research, we reduced the vector dimensions from 6188 to 77 for the takeoff phase and from 6279 to 95 for the approach phase in the initial testing of ClusterAD-Flight.

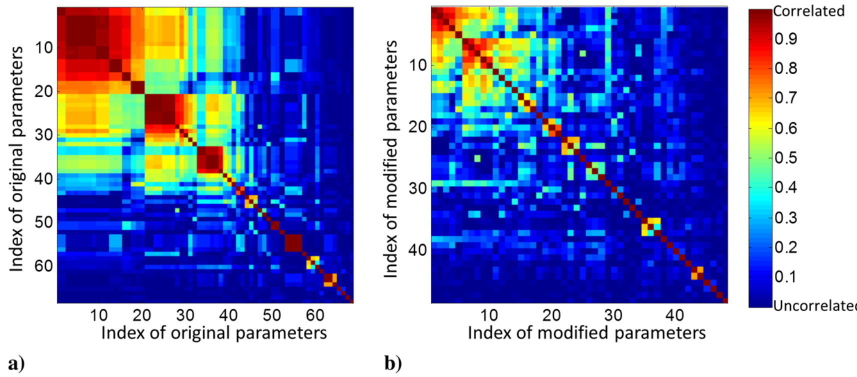


Fig. 4 Correlation matrix: a) before decorrelation, and b) after decorrelation.

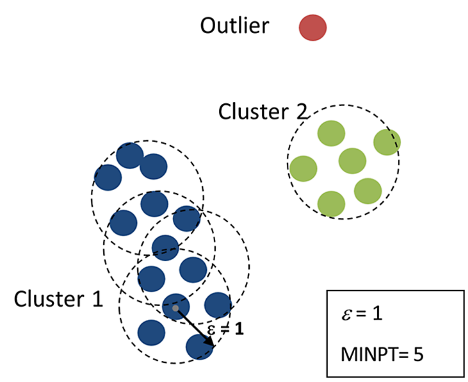


Fig. 5 Illustration of DBSCAN clustering process [42].

Another problem is the correlations between parameters in FDR datasets. As an example, a large number of flight parameters are linearly correlated with each other in the dataset used in the initial testing of ClusterAD-Flight, as shown in Fig. 4a. Performing PCA solves this problem, since it converts possibly correlated variables into linearly uncorrelated variables. However, for oversized datasets, the use of PCA becomes impractical, leading to potential biases in anomaly detection. This is caused by the PCA embedded singular value decomposition of the matrix, which is computationally expensive for a large dataset [36–38].

In the absence of PCA, we apply the following two-step procedure to weaken the effect of multicollinearity. The first is to identify sets of parameters that are closely correlated; the second is to compress correlated parameters of each set into two values at each sampling point: the average and the maximum difference among themselves. During data preprocessing, all continuous flight parameters are normalized to have zero mean and unit variance. The average value captures the general trend, whereas the maximum difference examines abnormal patterns. We use Pearson correlation coefficients, which are a measure of linear dependence between two variables developed by Karl Pearson from a related idea introduced by Francis Galton in the 1880s [39–41], to identify sets of correlated parameters. Correlated parameters are modified into new variables, which have much weaker linear dependence among each other. Figure 4b shows the linear correlations among modified parameters after decorrelation.

3. *Step 3: Cluster Analysis*

We apply the density-based spatial clustering of applications with noise (DBSCAN) algorithm [42] to perform the cluster analysis because it can 1) automatically determine the number of clusters; 2) handle data with outliers; and 3) identify multiple clusters and detect outliers. The DBSCAN algorithm identifies clusters based on a density criterion, where a cluster forms if at least the minimum number of points required to form a cluster (hereafter referred to as MINPT) are within ε radius of a circle as illustrated in Fig. 5. One cluster grows by adding points in the neighborhood that also meet the same density criterion until no other point can be added further. New clusters form as long as it satisfies the density criterion. The algorithm labels points that do not belong to any cluster as outliers.

This algorithm needs only two parameters as input: MINPT and ε. MINPT is the minimum number of flights that constitutes a nominal group, and ε affects the percentage of abnormal flights that can be detected. The values of MINPT and ε are selected based on a sensitivity analysis. We run DBSCAN multiple times with fixed MINPT while allowing ε to vary from the minimum pairwise distance to the maximum in the data. We start with a number for MINPT between 3 and 15, and results show that the abnormal flights detected are insensitive to changes of MINPT in the aforementioned range. Yet, if MINPT are set to be a relatively large number, some flights will be considered as abnormal while representing a type of normal operations. After MINPT are determined, the value of ε is set according to the acceptable percentage of abnormal flights to be detected.

III. Experiment 1: Initial Test of ClusterAD-Flight on Airline Data

ClusterAD-Flight was tested on a FDR dataset provided by an international airline company, which contained 365 B777 flights. Sixty-nine flight parameters were available in the dataset, including but not limited to engine parameters, aircraft position, speeds, accelerations, attitudes, control surface positions, winds, and environmental pressures and temperatures. Anomaly detection was conducted separately for the approach phase and the takeoff phase. Abnormal flights were detected using three sets of detection threshold: the top 1, 3, and 5% outliers. Outlier flights were further analyzed to determine operational anomalies and characterize the abnormal behaviors. We also investigated the causes if more than one cluster were present.

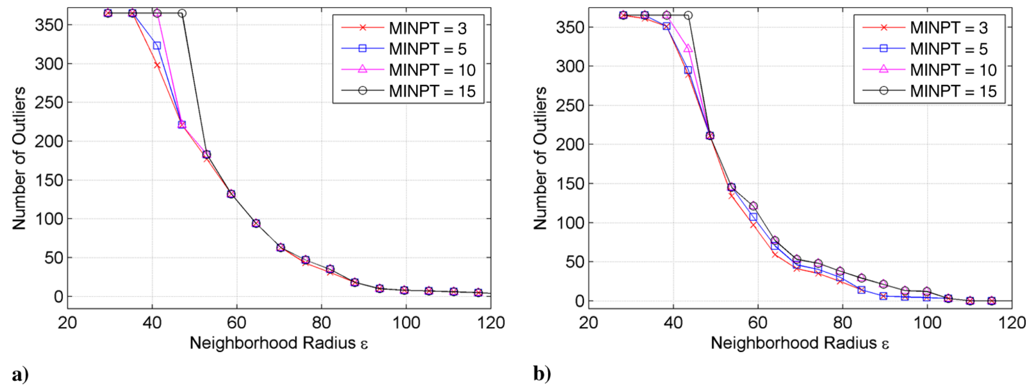


Fig. 6 Sensitivity to ε and MINPT: a) approach phase, and b) takeoff phase.

A. Data Preprocessing and Algorithm Settings

The first step was to set the sampling rate at takeoff and approach phase, with each at a fixed 91 samples. For the takeoff phase, observations were obtained at 1 s intervals, from the pilot applying takeoff power up to 90 s afterward; for the approach phase, the same number of observations was obtained from 6 n miles before touchdown. The second step was principal component analysis. By keeping the first K components that captured 90% of the variance in the data, the number of dimensions was reduced from 6188 (68 flight parameters $\times$ 91 samples) to 77 for the takeoff phase and from 6279 (69 flight parameters $\times$ 91 samples) to 95 for the approach phase.

Input parameters of clustering algorithm, MINPT, and ε were set based on the sensitivity analysis described previously. Results indicated the selection was insensitive to MINPT (between 3 and 15) but that fewer flights were identified as outliers when ε increased, as shown in Fig. 6. Therefore, MINPT was set at a value of five, and the value of ε was set to find the top 1, 3, and 5% outliers.

B. Results

1. Abnormal Flights Detected During Approach Phase

For the approach phase, three abnormal flights were detected under a 1% detection threshold: 10 under 3% and 16 under 5%. Two abnormal flights are plotted against normal ones as an example in the following graphs (Fig. 7), in which lines show parameter values of abnormal flights and bands depict the value range of the “common patterns,” where darker-colored bands indicate the 25th to the 75th percentile of all flights and lighter-colored bands encompass the 5th to the 95th percentile. The darker-colored region covers 50% of the data, whereas the lighter-colored region captures 90% of the data.

Figure 7a shows an abnormal flight of low and slow approach. Its vertical profile remained below the common glide slope until 2 n miles before touchdown; the calibrated airspeed was lower than most other flights until 3 n miles before touchdown; the flap was set to the landing configuration from at least 6 n miles before touchdown, which was earlier than normal; and the pilot used a much higher thrust than others until touchdown and a higher than normal pitch attitude to catch up with the glideslope between 3 and 2 n miles before touchdown.

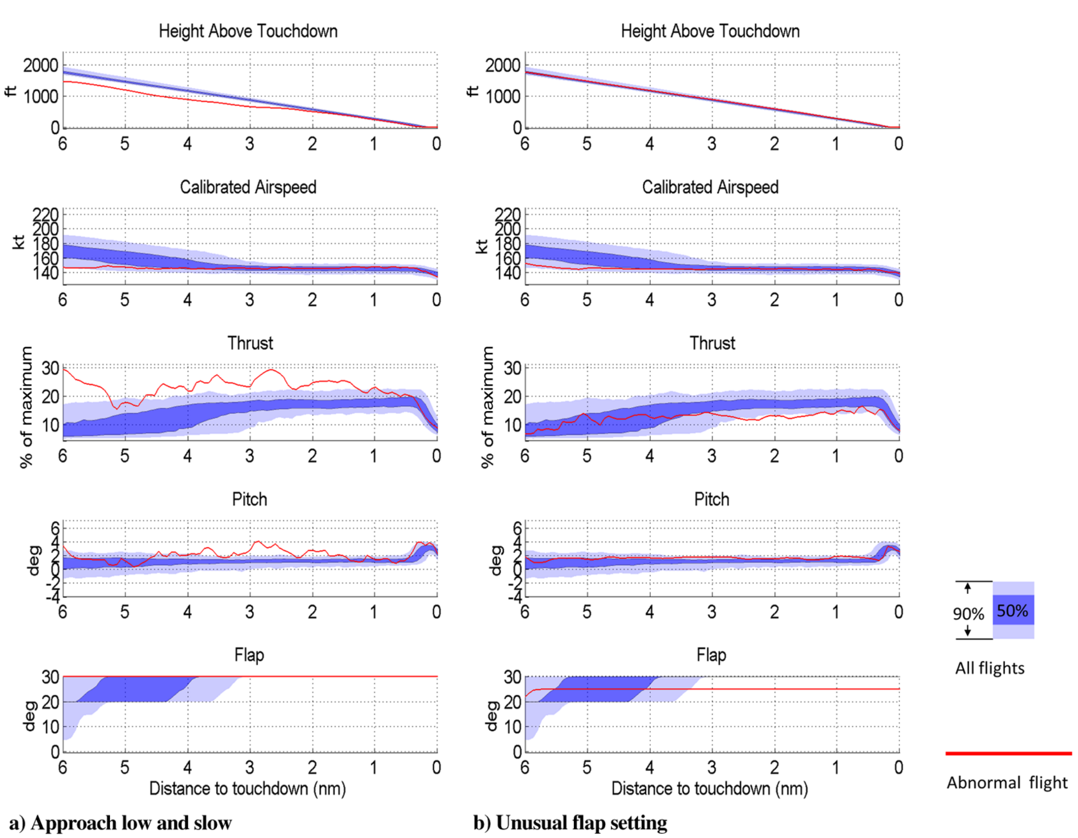
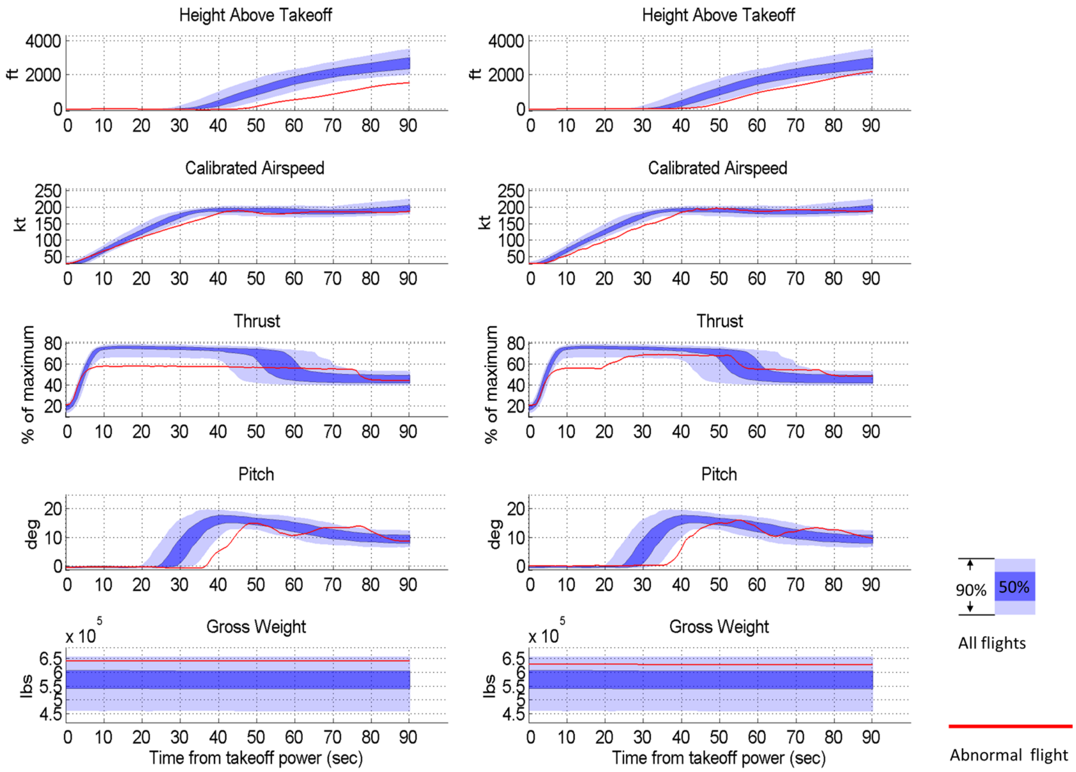


Fig. 7 Two abnormal flights detected during approach phase.



a) Reduced power takeoff **b) Takeoff power change**

Fig. 8 Two abnormal flights detected during takeoff phase.

Figure 7b shows a second example of unusual flap setting. The pilot kept the flap setting at 25 from 6 n miles before touchdown until landing, whereas most others used flap 30 for the landing configuration. The flight used less thrust for the final part of the approach, whereas major indicators of the approach performance, the altitude, the airspeed, and the pitch were within the 90% normal range.

2. *Abnormal Flights Detected from Takeoff Phase*

Four abnormal flights were detected using the 1% detection threshold: 9 under 3%, and 22 under 5% in the takeoff phase. Two abnormal flights are illustrated as examples. Figure 8a shows an abnormal flight with a lower takeoff power, despite that it was heavily loaded. The aircraft accelerated slowly and the rotation happened late, as it took much longer to reach the required airspeed. After the initial rotation, the pitch reached 15 deg again at 80 s after applying takeoff, which was similar to the angle during the initial rotation. Figure 8b shows another example that behaved similarly to the flight in Fig. 8a for the first 20 s. Its power setting was changed back to the normal level before the initial rotation, yet its climb rate and acceleration were still lower than most other flights.

3. *Summary of Abnormal Flights Detected*

The same type of analysis was performed for all top 5% abnormal flights detected during takeoff and approach. All identified flights exhibited some identifiable degree of anomaly. Not all abnormal flights implied safety concerns; some benign cases including takeoffs in strong wind or takeoffs with an early turn. Table 1 summarizes abnormal behaviors in the approach phase: mostly high-energy approaches and low-energy approaches. Some flights featured abnormally high pitch, unusual flap settings, and lining up with localizer relatively late. Weather-related anomalies were also found, including strong crosswinds and high atmospheric temperature. Table 2 summarizes abnormal behaviors in takeoff

Table 1 Abnormal behaviors in approach phase

Observed abnormal behaviors	No. of flights
High-energy approaches	
Fast	3
Fast and unstable airspeed	1
High and lined up late	1
Initially fast, then normal	1
Initially fast, then slow	1
Low-energy approaches	
Low and slow	1
Low and high power	1
Low and unusual yaw trim	1
Weather related	
High temperature	1
Strong crosswind	1
Other unusual	
Unusual flap setting	2
Abnormal high pitch	1
Line up late	1

Table 2 Abnormal behaviors in takeoff phase

Observed abnormal behaviors	No. of flights
High-power takeoffs	
Early rotation, fast climbout	3
Early rotation, early turn	2
Early rotation, crosswind	1
High pitch rotation, climb out high	2
Low-power takeoffs	
Reduced power, slow climbout	6
Reduced power, low and slow climbout, extended period of high pitch	1
Reduced power	1
Power setting related	
Excessive power reduction after takeoff	1
Start with reduced takeoff power, then switch to normal takeoff power, with low and slow climb out	1
Extended period of takeoff power	1
Weather related	
Rise of spoiler, strong wind	1
Other unusual	
Double rotation	1
Early turn after takeoff	1

phase. Typical abnormal behaviors were high- or low-power takeoffs associated with other notable factors such as excessive power reduction after takeoff, double rotation, and high pitch attitude during takeoff.

4. *Clusters with Nominal Data Patterns*

This initial testing also demonstrated that ClusterAD-Flight was able to recognize different nominal patterns in the data. Figure 9 shows three nominal patterns, or clusters, identified during the takeoff phase. Most takeoffs shared a common data pattern (cluster 1), and two small groups of takeoffs involved other patterns (cluster 2 and cluster 3). We found that all cluster 2 flights were departing from O.R. Tambo International Airport (International Civil Aviation Organization code FAJS), which is a high-altitude (5558 ft mean sea level) airport near the city of Johannesburg, South Africa. Cluster 3 flights were reduced-power or derated takeoffs: standard and safe procedure taken under certain circumstances to reduce operating and maintenance costs.

IV. Experiment 2: Comparison of ClusterAD-Flight, MKAD, and Exceedance Detection

The evaluation was conducted through a comparative study among three methods: ClusterAD-Flight, multiple kernel anomaly detection, and exceedance detection (ED). In the absence of standard FDR datasets with predefined normal flights and abnormal flights, the comparative study provided relative degrees of confidence on the performance of ClusterAD-Flight. MKAD is another anomaly detection algorithm based on data-mining techniques that was recently developed [9]. Exceedance detection is used as a baseline for the evaluation. It is widely used in airline FOQA programs, and it detects exceedance events at three levels of severity: level 3 (severe exceedance), level 2 (moderate exceedance), and level 1 (mild exceedance).

A. Dataset

We used a FDR dataset from a commercial passenger airline's normal operations consisting of 25,519 A320 flights landing at a standard European airport. The analysis focused on the approach phase (from 6 n miles before touchdown to touchdown). Each flight consisted of 367

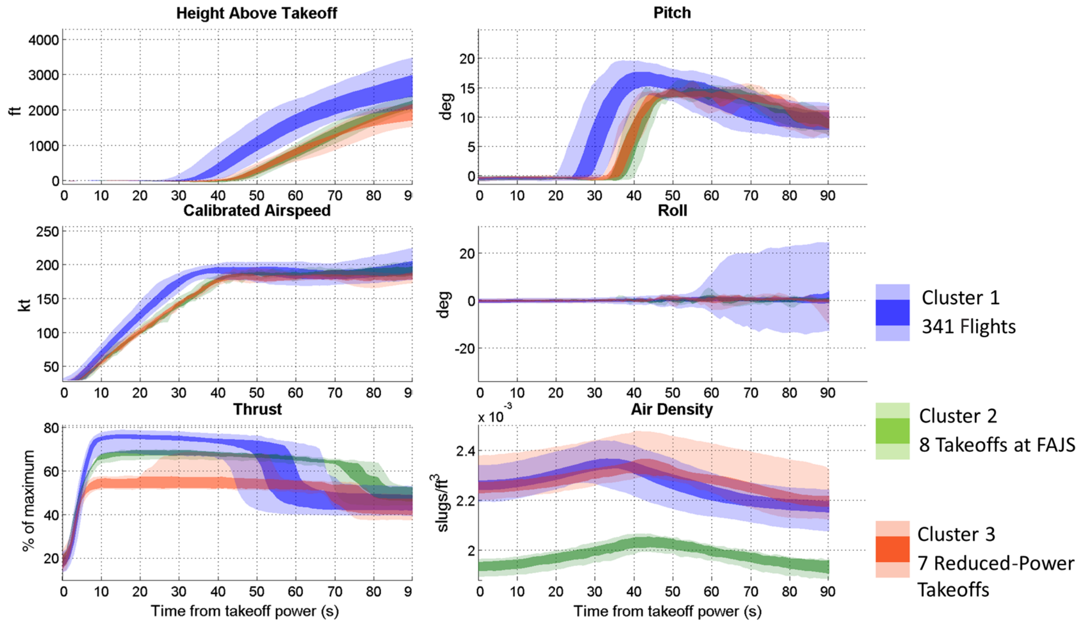


Fig. 9 Multiple nominal patterns in takeoff phase.

Table 3 Number of abnormal flights detected by ClusterAD-Flight and MKAD

	Detection threshold			
	1%	3%	5%	10%
ClusterAD-Flight	277	753	1,274	2,539
MKAD	203	704	1,206	2,483

Table 4 Number of abnormal flights detected by Exceedance detection

	Severity level		
	Level 3	Level 2	Level 1
Exceedance detection	729	3,581	18,888

discrete and continuous parameters sampled at 1 Hz, with the average flight length between 2 and 3 h. A subset of the flight parameters was selected based on domain expert's feedback in order to focus on detecting abnormalities in crew operations.

B. Results

1. Total Number of Abnormal Flights Detected by Each Method

Tables 3 and 4 summarizes the number of abnormal flights detected by ClusterAD-Flight, MKAD, and exceedance detection at various "detection thresholds." The detection threshold determines the percentage of abnormal flights in ClusterAD-Flight and MKAD. Similarly, the "severity level" regulates the number of flights with exceedance events in ED. Less than 3% of the flights were found with level 3 exceedance events, whereas most flights (18,888 out of 25,519) were detected with level 1 exceedance events.

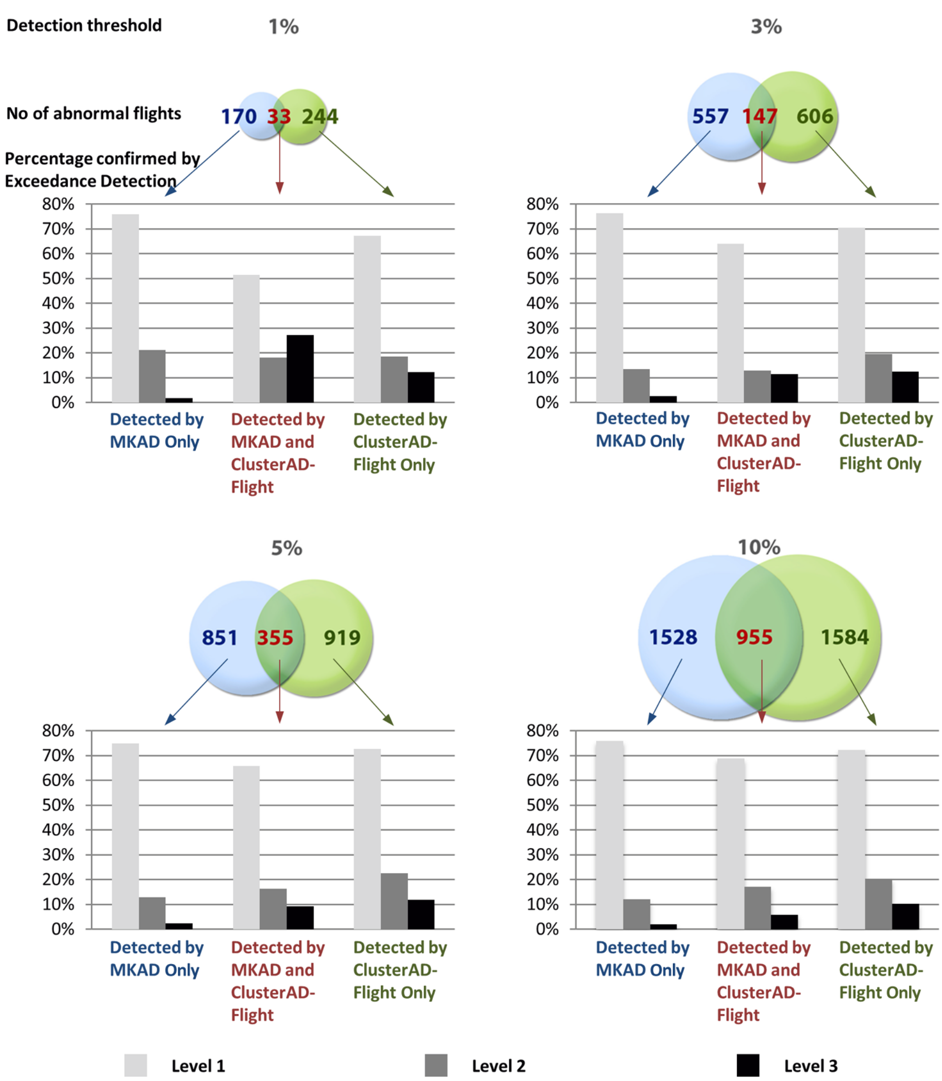


Fig. 10 Comparing the performance of ClusterAD-Flight and MKAD in detecting exceedance events.

2. Comparison of ClusterAD-Flight and MKAD Using Exceedance Detection as the Baseline

To compare the performance in detecting known anomalies, ClusterAD-Flight and MKAD were compared against ED, which can only detect predefined safety issues. Abnormal flights detected by ClusterAD-Flight or MKAD were categorized into three groups: 1) flights detected by MKAD only; 2) flights detected by both MKAD and ClusterAD-Flight; and 3) flights detected by ClusterAD-Flight only. These three groups were compared based on the percentages of flights with exceedance events, as shown in Fig. 10. A higher percentage of flights with exceedance events would indicate a higher degree of alignment with ED.

The results showed that ClusterAD-Flight was more sensitive to detect flights with moderate to severe exceedance events than MKAD. It was more likely to find the same results as ED. The percentages of flights with level 2 and level 3 exceedance events were consistently higher in group 3 (flights detected by ClusterAD-Flight only) than in group 1 (flights detected by MKAD only) at all detection thresholds.

The results also showed that using a combination of ClusterAD-Flight and MKAD at a tight detection threshold was most likely to detect severe exceedance events than using either method alone. At the detection threshold of 1%, the percentage of flights with level 3 exceedance events in group 2 (flights commonly detected by both ClusterAD-Flight and MKAD) was 27%, which was significantly higher than the percentage of any other group at other detection thresholds.

It is worth noting that these comparisons were made using ED as the baseline, which only shows how sensitive ClusterAD-Flight and MKAD were toward known safety issues that were captured in ED. The capability of ClusterAD-Flight and MKAD to detect unknown issues could not be compared directly in this comparative study.

3. Domain Expert Review to Determine the Operational Significance of Abnormal Flights

We conducted a detailed review of flights detected by ClusterAD-Flight or MKAD with two domain experts. The first expert was a retired airline pilot of a major U.S. carrier with over 35 years of flight experience; the second domain expert had over 30 years of research and operational experience in human factors, aviation, and pilot performance. The objective was to better understand the capabilities of these two methods in detecting operational anomalies, especially the ones that could not be captured by ED. Domain experts were asked to identify whether abnormal behaviors existed based on their operational experience; they were not informed which flight was detected by which anomaly detection method. During the review, information of the flights detected was shown in graphs of the same format described earlier. Limited by available resources, only a small portion of flights detected were selected to be reviewed. The selections were made based on the following criteria: group 1 had 57 flights detected by ClusterAD-Flight at all detection thresholds but not detected by MKAD at any detection threshold; group 2 had 88 flights detected by MKAD at all detection thresholds but not detected by ClusterAD-Flight at any detection threshold; and group 3 had 33 flights detected by both ClusterAD-Flight and MKAD at all detection thresholds. An abnormal flight was presented from each group as an example to illustrate the review process.

4. Example 1: High Airspeed Approach—Detected by ClusterAD-Flight Only

Figure 11 shows a flight of excessive high-airspeed instrument landing system (ILS) approach, which might result in a rushed and unstable approach. The rushed and unstable approach has been identified as one of the contributory factors in controlled flight into terrain and other approach-and-landing accidents, since it leaves insufficient time for the flight crew to plan, prepare, and execute a safe approach.

This flight was detected by ClusterAD-Flight under all detection thresholds, yet it was not detected by MKAD under any detection threshold. The airspeed profile was much higher than the normal and the targeted airspeed until 1.6 n miles before touchdown; the engine was set to idle until 3 n miles before touchdown. A number of flight parameters was changed abruptly around 3.5 n miles before touchdown, e.g., the pitch, the target airspeed, the stabilizer position, the vertical speed, etc. It was unclear what caused these changes, but we were confident that these were “unusual” operations.

ED also detected the same flight as “speed high in approach (at 1000 ft)” (level 3), “speed high in approach (at 500 ft)” (level 3), “flaps late setting at landing” (level 2), “deviation below glideslope (1000–300 ft)” (level 2), and five level 1 events.

5. Example 2: Unusual Auto Landing Configuration, Detected by MKAD Only

This abnormal flight used the autolanding system yet was apparently not following the standard procedure. The aircraft started the approach using a single autopilot and kept it engaged for the entire approach and landing. The standard procedure specifies two if autolanding is required so that, if one autopilot fails, the other autopilot can still complete the landing, or start with only one autopilot and disengage it when the runway is in

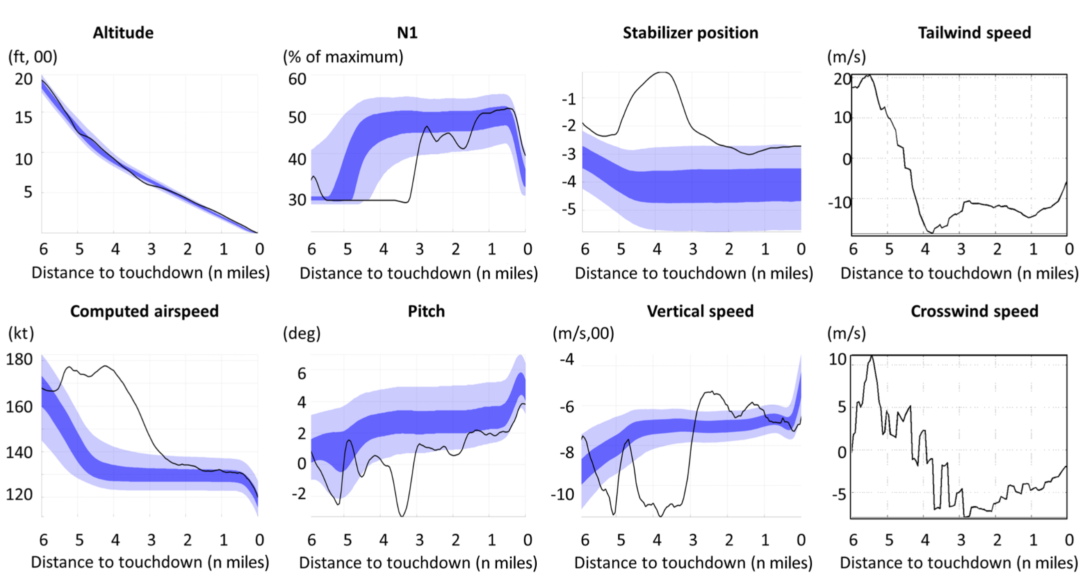


Fig. 11 Abnormal flight detected by ClusterAD-Flight: high airspeed approach.

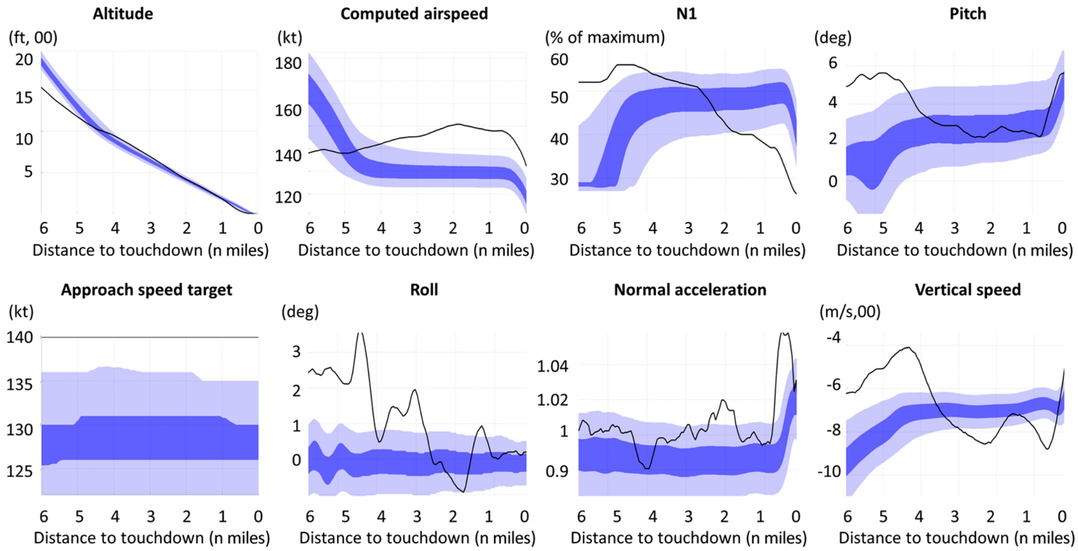


Fig. 12 Abnormal flight detected by ClusterAD-Flight and MKAD: high energy approach.

sight, or at least by the minimum charted descent altitude if autolanding is not required. From an operational perspective, this abnormal flight indicated a high level of risks if the pilot was relying on the single autopilot; the autolanding capability could be lost at an extremely inopportune time.

This flight was detected by MKAD directly at the cause of the problem: the unusual settings of autopilots. ED identified this flight with exceedance events "speed low at touchdown" (level 2) and "flaps questionable setting at landing" (level 3). Interestingly, ED only detected the low speed and the ill-configured flaps, but not the direct cause.

6. *Example 3: High-Energy Approach, Detected by Both ClusterAD-Flight and MKAD*

This abnormal flight had an unusually high airspeed during approach. A domain expert suggested that this flight might have had an energy state awareness problem. The speed profile and power profile of this flight could be precursors to a runway excursion for shorter runways. A domain expert stated the following: "Ideally, in these cases, it should be a go-around (aborted landing)." The landing operation was performed in a cloudy weather condition with an average visibility of 8.2 miles and with almost no wind. The aircraft intercepted the glideslope (see altitude plot in Fig. 12) from below. Before the interception, it was slower than other flights and the pitch was high. Then, the pilot decreased the pitch at 4 n miles before touchdown, which resulted in a high and unstable airspeed. At 500 ft, the pilot pulled the nose up slightly early and reduced power to ensure rapid deceleration. The effect of that could be clearly seen in the normal acceleration and vertical speed profiles.

This flight was detected by ClusterAD-Flight and MKAD under all three detection thresholds. Yet, inspectors would easily miss this anomaly using ED., which reported only level 1 exceedance events: "speed high in approach (at 50 ft)," "pitch high at touchdown," "path low in approach (at 1200 ft)," "long flare time," and "approach fast 500 radio altitude (RAD)." Level 1 events are commonly ignored from further inspection because they were reported in 18,888 out of 25,519 flights (74%) in this dataset.

C. Summary of Flight Review

Review results showed that both ClusterAD-Flight and MKAD were capable of detecting abnormal flights with operational significance. Each had its strengths: ClusterAD-Flight worked better with continuous flight parameters, as shown in example 1; whereas MKAD was more sensitive to the atypical sequence of discrete flight parameters, as shown in example 2. Both could detect flights that could not be flagged by the traditional method (exceedance detection), such as the flight detected in example 3. All flights selected for review were identified with abnormal behaviors confirmed by domain experts. Some indicated a higher level of risks suggested by domain experts; some were benign low-occurrence events.

V. Conclusions

Airlines currently collect flight data recorder data from aircraft on a regular basis, but it is challenging to analyze and extract useful conclusions from such a large amount of data. This paper presents a new method to analyze such data and extract information to support proactive safety management. The anomaly detection method, referred to as ClusterAD-Flight, can automatically detect abnormal flights from routine airline flights using cluster analysis. These flights can then be referred to domain experts for operational significance and latent risks.

ClusterAD-Flight was tested on flight data provided by airlines. The first test was performed on a dataset of 365 B777 flights. Results showed that the proposed method was able to detect abnormal flights without previous knowledge of anomalies. The second test was performed on a dataset of 25,519 A320 flights. In this test, ClusterAD-Flight was compared with multiple kernel anomaly detection, which is another emerging anomaly detection method; and with exceedance detection, which is the method in use by the airline industry. Results showed that both ClusterAD-Flight and multiple kernel anomaly detection (MKAD) were able to help domain experts identify abnormal behaviors from a large amount of routine airline flights, some of which might indicate high levels of risks. The results also suggested different strengths of ClusterAD-Flight and MKAD. ClusterAD-Flight was more sensitive in detecting abnormal patterns of continuous parameters and known safety issues that are currently included in exceedance detection (ED); MKAD was more sensitive toward flights with atypical sequences of discrete parameters. Looking forward, a hybrid method that combines the strengths of ClusterAD-Flight and MKAD could be a better way to capture a vast array of operationally abnormal flights while reducing the number of benign cases.

A future enhancement to this work is to specify the algorithm settings to generate the most desirable outcome in practice. Like many other anomaly detection methods, including MKAD and ED, ClusterAD-Flight requires a detection threshold (top x% of flights to be detected) to be set ahead of time. An overly tight detection threshold will miss true anomalies and emerging risks. However, an overly relaxed detection threshold will trigger false alarms. Rather than setting a perfect detection threshold, a potential direction for future work is to develop reinforcement learning capabilities, where initial domain expert reviews can inform recurrent reviews with established baselines. In this way, abnormal flights of identical symptoms can be categorized automatically without repeated expert review. Thus, one can afford to use a relative relaxed detection threshold to minimize true anomalies being missed.

Another direction is to extend the proposed method to other flight phases. Currently, ClusterAD-Flight is limited to takeoff and approach phases. Given the diversity of temporal patterns in other flight phases, new techniques are needed to make the time series of different flights comparable.

Acknowledgments

The work was supported by the Federal Aviation Administration under the Joint University Project (FAA 11-G-016), NASA under grant NNA06CN23A, and the City University of Hong Kong under the Start-Up Grant for New Faculty number 7200418. The authors would like to thank Robert Lawrence and Michael Feary at the NASA Ames Research center for sharing their knowledge in flight operations and for insightful discussions along this study.

References

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

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

[3] Amidan, B. G., and Ferryman, T. A., "Atypical Event and Typical Pattern Detection within Complex Systems," *2005 IEEE Aerospace Conference*, IEEE Publ., Piscataway, NJ, March 2005, pp. 3620–3631.
doi:10.1109/AERO.2005.1559667

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

[5] Budalakoti, S., Srivastava, A. N., and Akella, R., "Discovering Atypical Flights in Sequences of Discrete Flight Parameters," *2006 IEEE Aerospace Conference*, IEEE Publ., Piscataway, NJ, March 2006, pp. 1–8.
doi:10.1109/AERO.2006.1656109

[6] Srivastava, A. N., "Discovering System Health Anomalies Using Data Mining Techniques," *Proceedings of the 2005 Joint Army Navy NASA Air Force Conference on Propulsion*, Charleston, SC, June 2005, pp. 1–11, https://c3.nasa.gov/dashlink/static/media/publication/Data_Mining_for_ISHM.pdf.

[7] Iverson, D. L., "Inductive System Health Monitoring," *Proceedings of the 2004 International Conference on Artificial Intelligence (IC-AI04)*, CSREA Press, June 2004.

[8] Schwabacher, M., Oza, N., and Mattews, B., "Unsupervised Anomaly Detection for Liquid-Fueled Rocket Propulsion Health Monitoring," *Journal of Aerospace Computing, Information, and Communication*, Vol. 6, No. 7, 2009, pp. 464–482.
doi:10.2514/1.42783

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

[10] Matthews, B. L., 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] Hodge, V., and Austin, J., "A Survey of Outlier Detection Methodologies," *Artificial Intelligence Review*, Vol. 22, No. 2, Oct. 2004, pp. 85–126.
doi:10.1023/B:AIRE.0000045502.10941.a9:AIRE.0000045502.10941.a9

[12] Chandola, V., Banerjee, A., and Kumar, V., "Anomaly Detection: A Survey," *ACM Computing Surveys*, Vol. 41, No. 3, July 2009, pp. 15:1–15:58.
doi:10.1145/1541880

[13] Hofmeyr, S. A., Forrest, S., and Somayaji, A., "Intrusion Detection Using Sequences of System Calls," *Journal of Computer Security*, Vol. 6, No. 3, 1998, pp. 151–180.
doi:10.1.1.43.6197

[14] Lee, W., and Stolfo, S., "Data Mining Approaches for Intrusion Detection," *Proceedings of the 7th USENIX Security Symposium*, USENIX Assoc., Berkeley, CA, Jan. 1998.

[15] Mukherjee, B., Heberlein, L. T., and Levitt, K. N., "Network Intrusion Detection," *IEEE Network*, Vol. 8, No. 3, 1994, pp. 26–41.
doi:10.1109/65.283931

[16] Bishop, C. M., "Novelty Detection and Neural Network Validation," *IEE Proceedings of Vision, Image and Signal Processing*, Vol. 141, No. 4, 1994, pp. 217–222.
doi:10.1049/ip-vis:19941330-vis:19941330

[17] Worden, K., Manson, G., and Fieller, N. R. J., "Damage Detection Using Outlier Analysis," *Journal of Sound and Vibration*, Vol. 229, No. 3, 2000, pp. 647–667.
doi:10.1006/jsvi.1999.2514

[18] Brause, R., Langsdorf, T., and Hepp, M., "Neural Data Mining for Credit Card Fraud Detection," *Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence*, IEEE Publ., Piscataway, NJ, 1999, pp. 103–106.
doi:10.1109/TAI.1999.809773.

[19] Bolton, R. J., and Hand, D. J., "Statistical Fraud Detection: A Review," *Statistical Science*, Vol. 17, No. 3, 2002, pp. 235–255.
doi:10.1214/ss/1042727940

[20] Liao, T. W., "Clustering of Time Series Data—A Survey," *Pattern Recognition*, Vol. 38, No. 11, Nov. 2005, pp. 1857–1874.
doi:10.1016/j.patcog.2005.01.025

[21] Gaber, M. M., Zaslavsky, A., and Krishnaswamy, S., "Mining Data Streams: A Review," *ACM Sigmod Record*, Vol. 34, No. 2, 2005, pp. 18–26.
doi:10.1145/1083784

[22] Gariel, M., Srivastava, A. N., and Feron, E., "Trajectory Clustering and an Application to Airspace Monitoring," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 12, No. 4, 2011, pp. 1511–1524.
doi:10.1109/TITS.2011.2160628

[23] Chan, P. K., and Mahoney, M. V., "Modeling Multiple Time Series for Anomaly Detection," *Fifth IEEE International Conference on Data Mining (ICDM'05)*, IEEE Computer Soc., Houston, TX, 2005, pp. 90–97.
doi:10.1109/ICDM.2005.101

[24] Keogh, E., Lonardi, S., and Ratanamahatana, C. A., "Towards Parameter-Free Data Mining," *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, New York, 2004, pp. 206–215.
doi:10.1145/1014052.1014077

[25] Kumar, M., and Woo, J., "Clustering Seasonality Patterns in the Presence of Errors," *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, New York, 2002, pp. 557–563.
doi:10.1145/775047.775129
[26] Möller-Levet, C., Klawonn, F., Cho, K.-H., and Wolkenhauer, O., "Fuzzy Clustering of Short Time-Series and Unevenly Distributed Sampling Points," *Advances in Intelligent Data Analysis V*, Vol. 2810, Springer, Berlin, 2003, pp. 330–340.
doi:10.1007/978-3-540-45231-7-3-540-45231-7_31
[27] Au, W. H., Chan, K. C. C., Wong, A. K. C., and Wang, Y., "Attribute Clustering for Grouping, Selection, and Classification of Gene Expression Data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Vol. 2, No. 2, 2005, pp. 83–101.
doi:10.1109/TCBB.2005.17
[28] Rousseeuw, P. J., and Leroy, A. M., *Robust Regression and Outlier Detection*, Wiley, New York, 1987, pp. 216–247.
[29] Fox, A. J., "Outliers in Time Series," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 34, No. 3, 1972, pp. 350–363.
[30] Bianco, A. M., Garcia Ben, M., Martinez, E. J., and Yohai, V. J., "Outlier Detection in Regression Models with ARIMA Errors Using Robust Estimates," *Journal of Forecasting*, Vol. 20, No. 8, 2001, pp. 565–579.
doi:10.1002/(ISSN)1099-131X
[31] Smyth, P., "Markov Monitoring with Unknown States," *IEEE Journal on Selected Areas in Communications*, Vol. 12, No. 9, 1994, pp. 1600–1612.
doi:10.1109/49.339929
[32] Smyth, P., "Clustering Sequences with Hidden Markov Models," *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, 1997, pp. 72–93.
[33] Warrender, C., Forrest, S., and Pearlmutter, B., "Detecting Intrusions Using System Calls: Alternative Data Models," *Proceedings of the 1999 IEEE Symposium on Security and Privacy*, IEEE Publ., Piscataway, NJ, May 1999, pp. 133–145.
doi:10.1109/SECPRI.1999.766910.
[34] *Statistical Summary of Commercial Jet Airplane Accidents: Worldwide Operations 1959–2011*, Boeing Commercial Airplanes, Seattle, WA, 2012, p. 20.
[35] Hotelling, H., "Analysis of a Complex of Statistical Variables into Principal Components," *Journal of Educational Psychology*, Vol. 24, No. 6, 1933, pp. 417–441.
doi:10.1037/h0071325
[36] Ordonez, C., Mohanam, N., and Garcia-Alvarado, C., "PCA for Large Data Sets with Parallel Data Summarization," *Distributed and Parallel Databases*, Vol. 32, No. 3, Sept. 2013, pp. 377–403.
doi:10.1007/s10619-013-7134-6-013-7134-6.
[37] Rokhlin, V., Szlam, A., and Tygert, M., "A Randomized Algorithm for Principal Component Analysis," *SIAM Journal on Matrix Analysis and Applications*, Vol. 31, No. 3, 2009, pp. 1100–1124.
doi:10.1137/080736417
[38] Halko, N., Martinsson, P., Shkolnisky, Y., and Tygert, M., "An Algorithm for the Principal Component Analysis of Large Data Sets," *SIAM Journal on Scientific Computing*, Vol. 33, No. 5, 2011, pp. 2580–2594.
doi:10.1137/100804139
[39] Pearson, K., "Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity and Panmixia," *Philosophical Transactions of the Royal Society of London, Series A*, Vol. 187, Jan. 1896, pp. 253–318.
doi:10.1098/rsta.1896.0007
[40] Rodgers, J. L., and Nicewander, W. A., "Thirteen Ways to Look at the Correlation Coefficient," *American Statistician*, Vol. 42, No. 1, 1988, pp. 59–66.
doi:10.2307/2685263
[41] Stigler, S. M., "Francis Galton's Account of the Invention of Correlation," *Statistical Science*, Vol. 4, No. 2, 1989, pp. 73–79.
doi:10.1214/ss/1177012580
[42] Ester, M., Kriegel, H. P., Sander, J., and Xu, X., "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, Vol. 96, No. 34, AAAI Press, Portland, OR, Aug. 1996, pp. 226–231.

A. Banerjee
Associate Editor