12th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference and 14th AIAA/ISSM
17 - 19 September 2012, Indianapolis, Indiana

AIAA 2012-5593

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

Comparison of Algorithms for Anomaly Detection in Flight Recorder Data of Airline Operations

Santanu Das.*
UARC, NASA Ames Research Center, Moffett Field, CA 94035

Lishuai Li†
Massachusetts Institute of Technology, Cambridge, MA, 02139

Ashok N Srivastava‡
NASA Ames Research Center, Moffett Field, CA 94035

R. John Hansman§
Massachusetts Institute of Technology, Cambridge, MA, 02139

I. Introduction

In order to improve the current high level of safety in air carrier operations, there is increasing emphasis on developing proactive safety management systems to identify and mitigate risk areas before they manifest in aircraft accidents or incidents. One way to conduct proactive safety management for airline operations is to utilize the operational data archived in modern Flight Data Recorders (FDRs) equipped on aircraft. Recently, efforts have been made to develop algorithms to detect anomalies in sensor data from a complex engineered system in a dynamic operating environment.[2,5,1] These algorithms take a data-driven approach to build a model for detecting anomalies directly from data collected during system operation, rather than building it based on domain knowledge, standard operating procedure or human expertise. The knowledge discovery processes, in general, face the challenge of validating new discoveries from real-world data as in many cases there exists no "ground truth" that can confirm the presence of these anomalies.

In this study, we compared two data-driven anomaly detection algorithms and contrasted the two data-driven methods with the traditional flight data analysis method - Exceedance Detection. The two data-driven anomaly detection algorithms are Cluster-based Anomaly Detection (ClusterAD)[5] and Multiple Kernel Anomaly Detection (MKAD).[2] Both algorithms were developed to detect anomalous flights in recorded flight data.

The Exceedance Detection method is a Flight Operational Quality Assurance (FOQA) analysis tool widely used in the airline industry. It detects exceedance events when certain flight parameters exceed pre-specified thresholds. Only known safety concerns are examined by this method.

All three methods were independently tested on the same set of flight data from an airline's normal operations. The entire dataset contains recorded flight data of a narrow-body aircraft. The data are from short to medium range flights of a commercial passenger jet airline. The comparison results of the three

*UARC, NASA Ames Research Center, Moffett Field, CA 94035, AIAA Member, Email: santanu.das-1@nasa.gov

†PhD candidate, Department of Aeronautics and Astronautics, MIT, 77 Massachusetts Avenue, Cambridge, MA, 02139, Email: lishuai@mit.edu

‡NASA Ames Research Center, Moffett Field, CA 94035, AIAA Member, Email: ashok.srivastava@nasa.gov

§Professor, Department of Aeronautics and Astronautics, MIT, 77 Massachusetts Avenue, Cambridge, MA, 02139, AIAA Fellow, Email: rjhans@mit.edu

Copyright © 2012 by the American Institute of Aeronautics and Astronautics, Inc. The U.S. Government has a royalty-free license to exercise all rights under the copyright claimed herein f

methods is an empirical evaluation of ClusterAD and MKAD. In addition, the detected flights were organized and referred to domain experts in order to check the operational significance.

II. Background

This study focuses on two recently developed algorithms for anomaly detection in flight data. Both algorithms detect anomalous flights based on a model built from the flight data, rather than domain knowledge, standard operating procedure (SOP), or other prior knowledge. The traditional method, Exceedance Detection, by contrast, is based on flight manuals, SOPs, and other domain knowledge. Thus, it is used as a baseline to contrast the two data-driven anomaly detection algorithms. A description of the flight data and introductions of the three methods are given in the following paragraphs.

A. Flight Data

Flight data refer to the digital flight data recordings generated during aircraft operations. Various flight parameters (on average more than 300 flight parameters) are measured by sensors in an airplane and recorded on-board by the Digital Flight Data Recorder (DFDR) or Quick Access Recorder (QAR) during every commercial airline flight. Airlines collect the data on a regular basis and conduct analysis to evaluate daily operations and identify risks.

The data recorded on DFDRs or QARs include multiple flight parameters sampled at different rates. The specification of the parameters are to be recorded and the sampling rates vary by the type of recorder and the configuration requirements of the airline. The typical flight parameters include altitude, airspeed, thrust, accelerations, etc. The number of flight parameters can go up to 2000 depending on recording capabilities of modern airplanes. The size and complexity of the flight data create challenges in the data analysis.

B. Data-Driven Anomaly Detection Methods

The two algorithms studied in this paper are recently developed data-driven methods for anomaly detection in flight data. They are Cluster-based Anomaly Detection (ClusterAD)[5] and Multiple Kernel Anomaly Detection (MKAD).[2] Both algorithms build a model to characterize normal operations based on recorded flight data, and then detect anomalous flights based on the model, rather than domain knowledge, SOPs, or other prior knowledge.

1. ClusterAD

The Cluster-based Anomaly Detection (ClusterAD) algorithm is developed to group nominal flights by clusters and to detect anomalous flights that do not belong to any cluster.[5] As flight operations are highly standardized, most flights share similar patterns in the flight data and only a few of them have data patterns different from the majority. This method considers the recordings of multiple flight parameters simultaneously to look for patterns in the data.

ClusterAD performs well with flight phases that have standard procedures and clear time anchors, such as take-off and final approach. In this method, the time series of flight parameters are anchored by a specific time (e.g. the application of power during take-off or the touchdown in final approach). Then the time series are transformed to a vector in a high-dimensional space. Since the vector captures the information of all available flight parameters during the phase of flight, vectors representing similar flights are “close” to each other in the high-dimensional space. Then, cluster analysis based on DBSCAN[3] is performed to identify the clusters of proximate vectors, which are the nominal flights, and to detect outliers far away from any cluster, which are considered the anomalous flights. Both standard operations and anomalous operations can be identified in this cluster analysis. In addition, ClusterAD can handle situations when multiple standard operations exist in the data and the number of standard operations is unknown.

ClusterAD tends to work well with continuous flight parameters. However, it is not sensitive to the sequence of the discrete parameters (e.g. sequences of switches), as the discrete flight parameters are processed in the same way as the continuous ones but only state differences are observable in the algorithm.

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

2. MKAD

Multiple Kernel Anomaly Detection (MKAD) is a unique software system that can process and combine information from vast resources in varieties of semantic structures (e.g. discrete, continuous, text, and network data) simultaneously to identify aviation safety anomalies.[2] In aviation data heterogeneity may result from the presence of multiple attributes, where the attributes do not belong to the same data type. For example the attributes can be either continuous or discrete, or a mixture of both, or in sequential form. Since standard operating procedures exist for flying aircraft, the sequence of the discrete pilot inputs along with the measured quantities or parameters are extremely meaningful. Both deviations in values of continuous parameters and abnormalities in sequence of discrete parameters are considered in the MKAD algorithm.

MKAD is essentially a one-class Support Vector Machines (SVMs) based anomaly detection algorithm. MKAD can integrate knowledge from various heterogeneous data sources by virtue of the "multiple kernel" approach where new kernels can be derived from existing separate kernels built on different data types and thus incorporating the combined knowledge in the learning process. However each kernel has to be continuous, symmetric, and positive definite. For example, the resultant kernel K can be a convex combination of all kernels computed over multiple features i.e. $K(\vec{x}_i, \vec{x}_j) = \sum_{p=1}^{n} \eta_i \hat{K}_p(\vec{x}_i, \vec{x}_j)$, with $\eta_i \geq 0$ and $\sum_{i=1}^{n} \eta_i = 1$. Here $\hat{K}_p(\vec{x}_i, \vec{x}_j)$ represents the p^{th} kernel computed over either discrete or continuous parts of data points x_i and x_j, and η_i is to weight individual kernels. Once the kernel is formed, MKAD solves the optimization problem to construct an optimal hyperplane[6] in high dimensional feature space to separate the abnormal patterns from the normal ones. Once the optimization is solved, the model can be used to compute a decision function $f(\vec{z}) = sign(\sum_i \alpha_i \sum_p \eta_p \hat{K}_p(\vec{z}, \vec{x}_i) - \rho)$ to predict positive or negative labels for a given test vector $\vec{z}$. Examples with negative labels are classified as outliers whereas examples with positive labels are classified as normal. The absolute value of $f(\vec{z})$ gives an indication of how normal/abnormal the data point is, and can be used to rank the data points.

C. Traditional Anomaly Detection Method: Exceedance Detection

Exceedance detection is the traditional flight data analysis method widely used in the airline industry. It consists of checking whether particular flight parameters exceed predefined limits under certain conditions. The list of flight parameters to watch and the limits of those parameters are specified by safety specialists in advance. The watch list is always chosen to coincide with the airline's standard operating procedures, such as the pitch at takeoff, the speed at takeoff climb, the time of flap retraction, etc.[4] Therefore, this approach requires a pre-defined watch list of key parameters under certain operational conditions and, in addition, precisely defined thresholds of the key parameters. Known safety issues can be accurately examined by Exceedance Detection; however, the unknown emerging risks remain latent.

In this paper, we leveraged the results from a standard Exceedance Detection currently used by an airline. The standard Exceedance Detection detects three levels of exceedance events. Level 1 indicates minor deviations from the performance target, Level 2 indicates moderate deviations, while Level 3 indicates the severest deviation from the target value.

III. Approach and Experiment Setup

A. Experiment Design

The objective of this study is to compare two data-driven anomaly detection algorithms, ClusterAD and MKAD, and to evaluate both algorithms with a baseline method, Exceedance Detection. The three anomaly detection methods were applied on the same set of flight data. Each method detected a list of anomalous flights in the dataset.

The anomalous flights detected by different algorithms were compared to assess the commonalities and differences (1) between MKAD and ClusterAD and (2) between the data-driven methods and the Exceedance Detection method. The comparison between ClusterAD and MKAD in this study was focused on evaluating the flights commonly detected by both methods and the flights only detected by one of the methods, as shown in Table 1. In addition, the detected flights of interest were discussed with domain experts to understand the operational characteristics and implications.

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

Table 1. **Comparison Between ClusterAD and MKAD**

	MKAD - *Anomalous*	**MKAD** - *Normal*
ClusterAD - *Anomalous*	Common Detection	Unique ClusterAD Detection
ClusterAD - *Normal*	Unique MKAD Detection	Common Rejection

B. Algorithm Settings

Algorithms can be set at different sensitivity levels. ClusterAD and MKAD use the "detection threshold" setting to determine how many flights will be identified as anomalous. For example, by specifying a detection threshold ($x\%$), ClusterAD and MKAD detect the top $x\%$ anomalous flights from a set of flights. We compared ClusterAD and MKAD using a series of detection thresholds (detection threshold $= 1\%, 3\%, 5\%, 10\%$) to test algorithm performance on different conditions. In addition, the inter-comparison between ClusterAD and MKAD was always made on the same detection threshold to produce comparable results .

Exceedance Detection detects "exceedance events" when aircraft performance fails to meet target value ranges during specific maneuvers. Normally, airlines use three levels to detect the "exceedances". Level 1 indicates minor variation from the performance target, while Level 3 indicate the severest deviation from the target value. Because Level 3 events are the issues and problems that are of greatest concern to airlines, the comparisons with Exceedance Detection were focused on Level 3 in this study.

The overall design of the experiment and the algorithm settings are shown in Figure 1. Each of the boxes is an experiment scenario we have tested with the corrosponding algorithm settings.

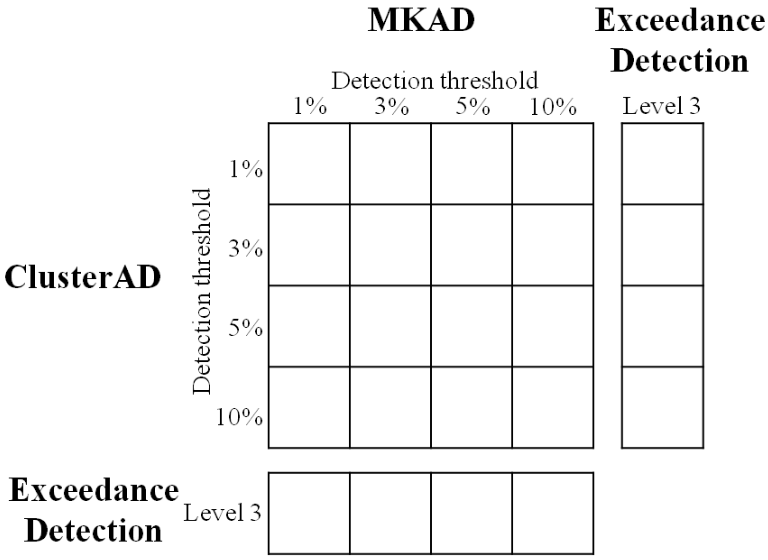

Figure 1. **Experiment Design and Algorithm Settings**

C. Data Pre-Processing and Parameter Selection

The flight data used in this study to compare all three methods are from a commercial passenger jet airline. All aircraft analyzed are of the same fleet and type. Each flight consists of 367 discrete and continuous parameters sampled at 1 Hz with the average flight length between 2 and 3 hours. However, we used a subset of the flight parameters (see Table 2) based on domain expert's feedback in order to focus on detecting operational problems. Using information from the domain expert in conjunction with the statistics from the data, the flap parameter, which is categorical in nature, was decomposed into 4 binary state variables. The mapping of the original flap parameter to the binary state variables is shown below (Table 3).

As a pre-processing step, we filtered the flight data by destination and phase of flight, in order to obtain

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

Table 2. List of Discrete and Continuous Parameters Used in This Study (please update)

Attribute Type	Variable Names
Discrete	Autopilot and all Autopilot related modes, Auto-throttle, Flight Director, Glide Slope, Stall Indicator, Flap Positions (derived parameter), Ground Proximity Warning System, Altitude Mode, Flare Mode, Flight Path Angle Mode etc.
Continuous	Altitude, Target Air Speed, Computed Air Speed, Engine-related Measures, Pitch Angle, Roll Angle, Rudder Position, Angle of Attack, Aileron Position, Stabilizer Position, Aircraft Gross Weight, Latitude, Longitude and Normal Accelerations, Derived parameters like Above Stall Speed, Vertical Speed etc.

Table 3. Relation of Derived Flap Parameters with Flap Positions.

	Flap0	Flap1	Flap2	FullFlaps
Flap Positions (in degree)	10	15	20	40

segments with comparable characteristics. The data used in this study are 25519 flights landing at a standard European airport. Only the part from 6 nmi before touchdown to touchdown of each flight was used for the analysis.

IV. Comparison Results

A. Overview

The number of flights detected by each method is summarized in Table 4 and Table 5. Since the number of anomalous flights detected in the data-driven methods, ClusterAD and MKAD, is controlled by the detection threshold, more flights are considered anomalous when a higher detection threshold is used, as shown in Table 4. While in the Exceedance Detection method, the severity level is the main factor impacting the number of flights being detected with exceedance events. As shown in Table 5, almost all flights are found to have at least one Level 1 exceedance event, while only less than 3% flights have at least one Level 3 exceedance event.

Table 4. Number of Flights Detected by Data-Driven Methods

	Detection Threshold			
	1%	3%	5%	10%
ClusterAD	277	753	1271	2539
MKAD	203	704	1206	2483

Table 5. Number of Flights Detected by Exceedance Detection

	Level 3	Level 2	Level 1
Exceedance Detection	729	3581	18888

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

B. Comparison between ClusterAD and MKAD

Because ClusterAD tends to work well with continuous flight parameters, while MKAD is better to incorporate the sequence of discrete flight parameters in anomaly detection, we expected the results of the two algorithms would be different. This is confirmed by the results. Comparing the flights detected by ClusterAD and the ones detected by MKAD, the number of common flights which are detected by both MKAD and ClusterAD vary from 33 (Detection Threshold = 1%), 147 (Detection Threshold = 3%), 355 (Detection Threshold = 5%), to 955 (Detection Threshold = 10%). The agreement between the two methods increases as the detection criteria become more relaxed, as shown in Fig. 2

The difference between the types of flights detected by ClusterAD and the ones detected by MKAD also confirms that ClusterAD is more influenced by deviations in continuous parameters, while MKAD is impacted by sequences anomalies in discrete parameters. Flights with corrupted data, especially corrupted data in continuous parameters, are expected to be picked out by ClusterAD. Some flights with corrupted data have parameters with a constant offset or with missing values. They were in the dataset because they cannot be filtered out in advance without a pre-defined normal range of every flight parameter. As shown in Fig. 3, ClusterAD detects more flights with corrupted data than MKAD, as ClusterAD is more influenced by the deviations in continuous parameters. Meanwhile, Table 6 shows that MKAD detects more autolanding flights because this approach features more activity in discrete flight parameters than other types of approach.

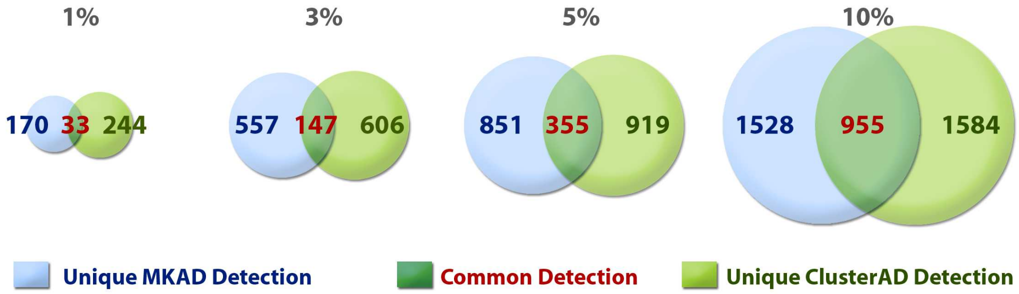

Figure 2. Comparison between ClusterAD Detection and MKAD Detection

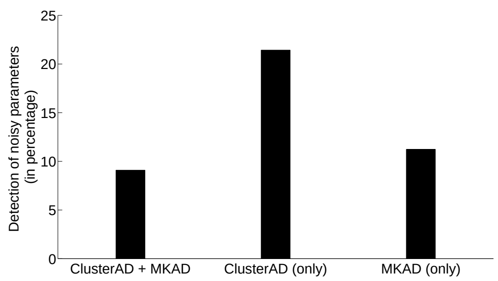

Figure 3. Percentage of Corrupted Data Detected by ClusterAD and MKAD

C. Comparison between Exceedance Detection and Data-Driven Methods

Among different levels of exceedance, Level 3 exceedances are the severest and raise the most concerns for airlines. Thus, Exceedance Detection at Level 3 was used as the baseline to compare the performance of ClusterAD and MKAD. Table 7 shows the number of common flights detected by Exceedance Level 3 and ClusterAD and the number of common flights detected by Exceedance Level 3 and MKAD. It is noted that ClusterAD has detected more common flights than MKAD for all of the tested detection thresholds.

This result is expected since ClusterAD can be considered as a variation of Exceedance Detection. ClusterAD looks for deviations between anomalous values and nominal values. When most operations are fol-

Table 6. Automation Summary of all Landing Flights and top 10% Anomalous Flights Detected by Data-Driven Methods

	Automation type				
	ILS Approach	Autolanding	Non-precision Landing	Visual Landing	Others
All Flights	21960	568	2	2987	2
ClusterAD	1141	20	2	1160	0
MKAD	961	103	2	1416	1

lowing the standards, the nominal values are close to the target values used in Exceedance Detection. A significant deviation from the nominal value detected in ClusterAD should also be detected in Exceedance Detection. The difference between ClusterAD and Exceedance detection is that ClusterAD considers all the available flight parameters simultaneously and does not need a pre-specified list of queries.

Table 7. Compare Exceedance Detection with ClusterAD and MKAD

		Number of flights detected by both Exceedance Detection (Level 3) and	
		ClusterAD	**MKAD**
Detection threshold	1%	39	12
	3%	93	31
	5%	143	53
	10%	220	86

V. Operational Characteristics of Flights Detected by Data-Driven Methods

In order to further understand the strength and the limitations of ClusterAD and MKAD, we reviewed the detected flights with domain experts in detail and cross-checked with the Exceedance detection results, to look for operational significance. We selected three groups from all the flights detected by any method:

- Group (1) - Common flights always detected by both ClusterAD and MKAD on all detection thresholds
- Group (2) - Unique flights always detected by ClusterAD on all detection thresholds, but not detected by MKAD on any detection threshold
- Group (3) - Unique flights always detected by MKAD on all detection thresholds, but not detected by ClusterAD on any detection threshold

In this section, we present several representative examples in each group. Three types of graphs are used to show the information of a flight: (1) Speed and flap setting during final approach; (2) Autopilot mode transitions during final approach; (3) Time-series plots of most distinctive flight parameters. Regarding to the third type, the same format is used to show the most distinctive flight parameters. The detected flight is shown by red lines. The patterns of most flights are depicted by blue bands. The dark blue bands indicate the 25th to the 75th percentile of all flights data; the light blue bands encompass the 5th to the 95th percentile. The dark blue region contains 50% of the data, while the light blue region covers 90%. The wind plots and latitude & longitude plots in the Type 3 graph do not have blue bands because those parameters (wind, latitude, longitude) were not included in the parameter list for data-driven anomaly detection.

A. Common Flights Detected by Both ClusterAD and MKAD

Common flights detected by both ClusterAD and MKAD are the ones that have significantly different data patterns from other flights. A number of flight parameters of these flights are distinctively different compared to the nominal values of most flights. Some of the common flights indicate interesting operational

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

implications, while some of them are benign as they are abnormal due to the low occurrence of the operation. Four examples of the common flights are presented in detail: High-Energy Approach, High-Airspeed with Low-Power Approach, Recycling Flight Director, and Influence of Wind.

1. *High Energy Approach*

The high-energy approach is a flight that is detected by both ClusterAD and MKAD on all detection thresholds. There are three basic conditions that may lead to a high-energy approach: the aircraft may be too high, too fast, or both. This flight has been categorized as a high energy approach with unusually high air speed when compared to a set of reference flights landing at that airport.

The flight review with domain expert suggested that this flight might indicate an energy state awareness problem. The speed profile and power profile of this flight could be precursors to runway excursion for shorter runways. The high speed was not due to the wind. Ideally, in these cases it should be a go around. The landing operation was performed in a cloudy weather condition with average visibility of 8.2 miles and with almost no wind. “Flap 0”, “Flap 1” and “Flap 2” (see Table 3) along with the landing gear were deployed before 1800 ft (or 6 nmi from touch down). During this part of the flight, a gradual turn was initiated to align with the runway in preparation for landing. This flight intercepted the glide slope (see altitude plot in Fig. 4) from below and was slower than most other flights at the beginning of their approach. During this period the pitch was high. Immediately after this, the pilot spooled up the engines for some time to increase the speed. This was followed by lowering the pitch, which further accelerated the aircraft. Although the target airspeed (140 kts) was higher than most others (126-130 kts), the high power used until 3 nmi before touchdown resulted in a high and unstable airspeed and a significant decrease in N1 for the rest of the approach. In addition, the pitch angle profile and altitude profile showed signs of unstable approach. At 500 ft the pilot pulled the nose up slightly early to ensure rapid deceleration. The effect of that maneuver can be clearly seen in the normal acceleration and vertical speed profiles.

The anomaly detection algorithms found this flight atypical due to the combined effect of the deviations of several continuous parameters. This is an interesting example of what can be detected by data-driven methods, but may be overlooked by the Exceedance Detection method. The exceedance based approach reported Level 1 exceedances, namely “Speed High in Approach (at 50ft)”, “Pitch High at Touchdown”, “Path Low in Approach (at 1200ft)”, “Long Flare Time”, and “APPROACH FAST 500 RAD”. The findings of the anomaly detection methods provide a clear picture on the unusual energy management scenario, however the Exceedance Detection method categorized it as a normal flight.

2. *High-Airspeed with Low-Power Approach*

One type of anomalous approaches that can be detected by both ClusterAD and MKAD is the high-airspeed and low-power approaches. Figure 5 is an example of this style of approach. It was a visual landing. The airspeed was always high and the engine was set to idle until 1 nmi before touchdown. Procedure calls for the engines to be spooled up for the entire final approach so that instantaneous power adjustments can be made. Other flight parameters also show abnormal patterns compared to the patterns in the majority of flights. For example, the altitude profile was above the normal altitude profile from 5 nmi to 1 nmi before touchdown, the pitch was relatively low until 2 nmi before touchdown, the roll angle had a significant amount of activity at the beginning of the final approach, etc.

Although this flight landed safely, this type of approach is not recommended. The test on this dataset shows that both ClusterAD and MKAD can catch this type of anomaly. Moreover, Exceedance detection confirms that this type of anomaly is operationally significant. The exceedance detection identifies four Level 3 events, one Level 2 event, and four Level 1 events in the approach part for this flight. The Level 3 events are “Speed High in Approach (at 1000ft)”, “Speed High in Approach (at 500ft)”, “ Low Power on Approach” and “APPROACH FAST 500 RAD”. The Level 2 exceedance is “Pitch High at Touchdown”.

3. *Recycling Flight Director*

For this particular flight, the main contribution toward the anomaly score came from an atypical event in discrete parameters as a result of a change of runway as well as another error of commission. The first event is related to automation disconnection. This flight was completely hand flown and was initially configured for the right runway. The latitude-longitude profile (Fig.8) and the presence of the Autopilot modes (G/S

and *LOC*) in Fig.6 confirm this fact. Once the new runway was assigned and the required maneuvering was initiated to align with the left runway, the crew had to recycle the flight directors in order to get to the default modes of Heading and Vertical Speed. The second event was related to mode switching and we are unable to reach any hypothesis as to why the pilot would take such an action. The transition from "vertical speed mode" to "open climb mode" around 1500 ft was an inappropriate move by the pilot as the missed approach altitude has already been set and the "open climb mode" would spool up the engines to climb to that altitude right away. However it is clear from Fig. 6 that the pilot immediately corrected this mistake and continued to hand fly the aircraft appropriately. A level 2 exceedance in "Speed High in Approach (at 1000ft)" was reported, but may have been unrelated to all the forgoing actions.

4. Influence of Wind

Both ClusterAD and MKAD found this flight (Fig. 10) as anomalous for several reasons. Deviations of multiple continuous parameters combined with various mode transitions contributed to the atypicality. Some of these deviations are not immediately obvious when examining individual parameter plots; however together they can combine to create an atypical flight. The discrete parameter plot (Fig. 9), shows that one autopilot (AP 1) was used and later the pilot had disconnected the autopilots and proceed manually with the rest of the approach and landing. Moreover, both the flight directors were recycled immediately after that. The first hypothesis is that there have been a change in parallel runway, and the pilot has to disconnect the automation while aligning the aircraft to the new runway assigned to him. But the latitude-longitude plot confirms that there was no "runway change". The second and more likely hypothesis, is that the switching of the autopilot and flight directors could have been part of an auto approach process. The pilot first disconnected the autopilots once the necessary visibility of the runway was achieved, then recycled the flight directors to engage the default modes and later decided to hand fly the aircraft. Another interesting observation is the missing autoflight lateral mode. Further investigations revealed that the "NAV" mode was active throughout the earlier part of the flight and was deactivated right before the final approach. This would not happen and is probably an artifact of the recording process. Any time either an autopilot or a flight director is engaged both a lateral mode and a vertical mode must be in use.

The entire operation was performed in an extremely windy condition. Though wind was not part of the analysis but the wind plots in Fig. 10 help explain atypical fluctuations in some of the parameters like target airspeed, rudder and lateral/normal acceleration. Exceedance based model didn't detect any event for this flight.

B. Unique Flights Detected by ClusterAD Only

The Cluster-based Anomaly Detection (ClusterAD) algorithm groups nominal flights by clusters, to detect anomalous flights based on the sample-by-sample difference of each flight parameter. It requires the part of the data being analyzed to have specific time anchors to make the comparison. The time series of flight parameters are anchored by a specific time (e.g. the application of takeoff power during take-off, or the touchdown in final approach). Detailed description of the method can be found in one of our earlier paper.[5]

ClusterAD tends to perform well with continuous flight parameters and is influenced by the magnitude of deviations. In this section, we present two examples of flights detected by ClusterAD, yet not by MKAD. Both flights had significant deviations in continuous flight parameters. Therefore, ClusterAD can be considered as a variation of Exceedance Detection, as it works in a similar way when considering flight parameter deviations; however, ClusterAD automatically inspects all available flight parameters and make the comparisons based on nominal values summarized from the data itself, rather than pre-specified parameters. In addition, ClusterAD can handle situations when multiple standard operations exist in the data as well as when the number of standard operations is unknown.

1. Very High Airspeed

This flight type was detected by CLusterAD at all detection thresholds, yet not detected by MKAD on any detection threshold. It was a very high airspeed ILS approach (see Fig. 11 and Fig. 12). The airspeed profile was always much higher than the normal airspeed and also than the target airspeed until less than 2 nmi before touchdown. Because the airspeed was too high, the engine was set to idle until 3 nmi before touchdown. Moreover, many flight parameters, e.g. the pitch, the target airspeed, the stabilizer position,

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

the vertical speed, etc. had an abrupt change around 3.5 nmi before touchdown. It is not clear what caused this change and why there was a significant drop in pitch even though the airspeed was too high.

This flight was also detected in Exceedance Detection. The detected events are "Speed High in Approach (at 1000ft)" (Level 3), "Speed High in Approach (at 500ft)" (Level 3), "Flaps Late Setting at Landing" (Level 2), and "Deviation below Glideslope (1000ft - 300ft)" (Level 2). In addition, five Level 1 events are also found in this flight.

This example shows that ClusterAD can detect approaches with excessive airspeed, which is one type of rushed and unstabilized approach. The rushed and unstabilized approaches are one of the contributory factors in Controlled Flight Into Terrain and other approach-and-landing accidents, because they can result in insufficient time for the flight crew to correctly plan, prepare, and execute a safe approach.

2. Landing Runway Change

Another type of flight detected by ClusterAD but not by MKAD is the flight with landing runway changes during final approach. The flight shown in Fig. 13 is a representative example of this type. The flight was originally lined up for the right runway. Then, it turned and landed on the left runway. Although the ground reference position information (e.g. latitude and longitude) was not included in the anomaly detection analysis by ClusterAD and MKAD, ClusterAD was able to capture the abnormal behaviors in other flight parameters caused by the change of runway turn and identified the flight as abnormal on all detection thresholds.

None of these abnormal behaviors is severe by the standard of Exceedance Detection. No Level 3 or Level 2 events are detected in the flight for this flight phase. Only four Level 1 events are detected. This type of anomalous flight could be operationally benign, because the change of landing runway happens due to many reasons (e.g. ATC assignment to accommodate traffic flows, ILS instrument limitations, etc). However, to identify this type of abnormal operations and then to track the trend can help to understand the operations better, such as whether it happens at a particular airport, during a specific time of the day, or under certain weather conditions. Moreover, further analysis may bring insights on whether there is a correlation between the approaches with runway change and the unstable approaches.

C. Flights by MKAD

In the Multiple Kernel Anomaly Detection (MKAD) algorithm, we model discrete (switching) sequences and continuous sequences for a given process using a normalized Longest Common Subsequence (nLCS) based kernel. Further details on the preprocessing steps of discrete and continuous parameters can be found in the following paper.[2] For FOQA data type analysis, sequential features in many cases provide valuable information since the order of the switching may provide justifications on performing pilot's activities to achieve an objective. It is important to note that like ClusterAD, MKAD identifies anomalies at the fleet level, meaning individual flights were labeled anomalous by the algorithm using the combined information of discrete and continuous parameters.

1. Unusual Autolanding Configuration

In this section we will describe two flights identified by MKAD mostly due to some atypical patterns in the switching sequences generated from the discrete parameters. In both cases the flights used autoland systems which have been designed to control the aircraft automatically during approach and landing. In this data set we have a small fraction (around 2% of the 25,519 flights) of autolanding examples, and the deviations from normal switching behaviors for autolanding make them statistically significant compared to the rest of the flights. Autolanding is mostly used in poor visibility conditions and/or bad weather where the crew can only see the runway lights just few seconds before landing. In many instances with poor visibility conditions, visual landing may not be possible or may be considered unsafe, and therefore autolanding is preferred. The presence of automatic guidance systems with human in-the-loop makes the autoland an extremely accurate and safe maneuver. However there are strict requirements which are imposed by the authorities on airborne elements and ground environment, as well as special crew qualification for autolanding.

The first flight (Fig.14), reported by MKAD, engaged autolanding without the full flap setting. Under normal circumstances the autoland is executed with both autopilots engaged and with flaps configured as full. The use here of the flap setting prior to full introduced some differences from the usual autoland

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

patterns. Out of all autolanding examples, more than 90% of flights performed this operation with full flap settings. While legal from an operational standpoint this was still reported by the algorithm due to some statistically significant activities (or signatures) in parameters like autopilot and autopilot modes and flight directors. The Exceedance-based method indicated “Pitch High at Touchdown”, “Short Flare Time” and “Short Flare Distance”with considerable severities (Level 2-s & 3-s).

In the second example (Fig. 15) we demonstrate another atypical autoland configuration. In this flight the flaps were configured full. The weather was reported as foggy with 0.1 mi visibility. It is common to use only one autopilot for an approach that does not require an autoland. The autopilot is then disengaged when the runway is in sight or at least by the minimum charted descent altitude. Procedures specify that an approach which requires an autoland, however, must be started with two autopilots. If one autopilot then fails, a “fail-operational state” exists and the automatic landing can be completed. This flight departed from normal operational requirements by utilizing only one autopilot for the entire approach and autoland. This scenario could be of interest because under bad weather conditions (like extremely low visibility conditions) with further degradation of the system the autolanding capability may be lost at an extremely inopportune time. The algorithm was able to find it due to the uncharacteristic settings of autopilots for landing aircraft. For example, out of all autolanding examples, only 2 flights with different tail numbers performed this kind of operation. The exceedance based method indicated “Speed Low at Touch down”and “Flaps Questionable Setting at Landing”with considerable severities (Level 2-s & 3-s).

D. Unique Flights Detected by Exceedance Detection

In Exceedance-based detection severity Level 2 and 3 are of concern to airlines. To make a fair comparison, in this section we present a couple of flights that have been detected by Exceedance Detection but not by any of the data-driven methods described above. For this study we chose the top two flights with the maximum number of Level 3 exceedances for the phase of the flight described above. Table 8 shows the type of exceedances identified by the Exceedance Detection method for all three severities.

Both ClusterAd and MKAD are multivariate methods, which identify anomalous flights by considering all available flight parameters. The abnormality level of a flight is the combined effect of how abnormal a flight parameter is at an instance, how long the abnormality lasts for a flight parameter, and how many flight parameters are abnormal. The flight parameter plots of Flight 1 and Flight 2 show that most of the parameters are within the blue bands for most of the approach. The events detected by Exceedance detection for Flight 1 and Flight 2 are specific deviations at a particular time, such as landing, touchdown, 1000 ft, 500 ft, etc. A short time deviation of a few flight parameters may not be able to bring the flight to top of the abnormal list generated in ClusterAD and MKAD. Therefore, neither ClusterAD or MKAD could detect those flights.

FOQA programs are typically designed to search for foreseeable problems. These data-driven algorithms on the other hand, were created to search for atypicalities which were unforeseen. For example, in the case of Flight 1 (Fig. 16), a finely tuned FOQA tool picked out the high pitch rate, but that is because it was specifically looking for this problem. Pitch rate was not part of the data driven analysis and so Flight 1 was not picked up by any of the data driven methods. Though the overall profile of computed airspeed looks normal from the Fig. 17, there are two small deviations in computed airspeed. The difference between computed airspeed and target airspeed around those deviations resulted in the speed-related exceedances. Data driven techniques are not sensitive to such small deviations and thus didn’t identify Flight 2 as anomaly.

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

Table 8. Top 2 Unique Flights reported by Exceedance Detection

	Level 3	Level 2	Level 1
Flight 1	Pitch Rate High at Landing, Short Flare Time, Tail Strike Risk at Landing	Pitch High at Touchdown	Height High at Threshold, Short Flare Distance
Flight 2	Speed High in Approach (at 1000ft), Speed High in Approach (at 500ft), Low Power on Approach	Deviation above Glideslope, (1000ft - 300ft), Go Around	Pitch High at Touchdown, Rate Of Descent High in Approach (2000ft - 1000ft), Approach Fast 500 RAD

VI. Conclusions

In this paper, we have compared results from two data driven algorithms with results from traditional exceedance based methods on a common aviation data set. The overall results indicate that ClusterAD tends to work well with continuous flight parameters while MKAD is more sensitive to the sequence of the discrete parameters. Exceedance Detection can be very efficient in detecting foreseen anomalies using continuous and discrete data. Each method overlaps the others to some extent, but there remain unique strengths which set them apart as well. The aim is not to continue to find known anomalies but to combine strengths from different methods into a robust approach enable detection of unknown operationally significant anomalies in aviation data sets.

From this study, two generalizations come to mind. First, there will always be a range of appropriate and acceptable operating conditions. The fact that a flight is operating near the limit of this range may make it atypical but does not make it operationally incorrect. An example would be a maximum gross weight landing, with higher target and computed airspeeds. Data driven algorithms do find atypicalities based on baseline statistics, but Exceedance Detection systems may evaluate these situations more appropriately since they can take into account some of the applicable physical laws.

Secondly, if enough pilots make the same mistake it will no longer be atypical. One example of this might be extended time in the flare after main gear touchdown. This runs the risk of hard nosewheel touchdown which is not an acceptable practice, but is nonetheless fairly common. Data driven algorithms might have difficulty finding this phenomenon.

Additionally, FOQA programs generally examine data in temporal slices, finding exceedences which might be very temporary and which must be analyzed to determine their relevance. On the contrary a data-driven algorithm provides some perspective, comparing the flight to a baseline comprised of many other flights. Domain analysis is still required, but from a different point of view.

Acknowledgments

This project was supported by the NASA Aviation Safety Program, System-Wide Safety and Assurance Technologies (SSAT) Project and FAA Joint University Program. The authors would like to thank Robert Lawrence and Dr. Michael Feary for their invaluable domain expertise and insightful discussions.

Appendices

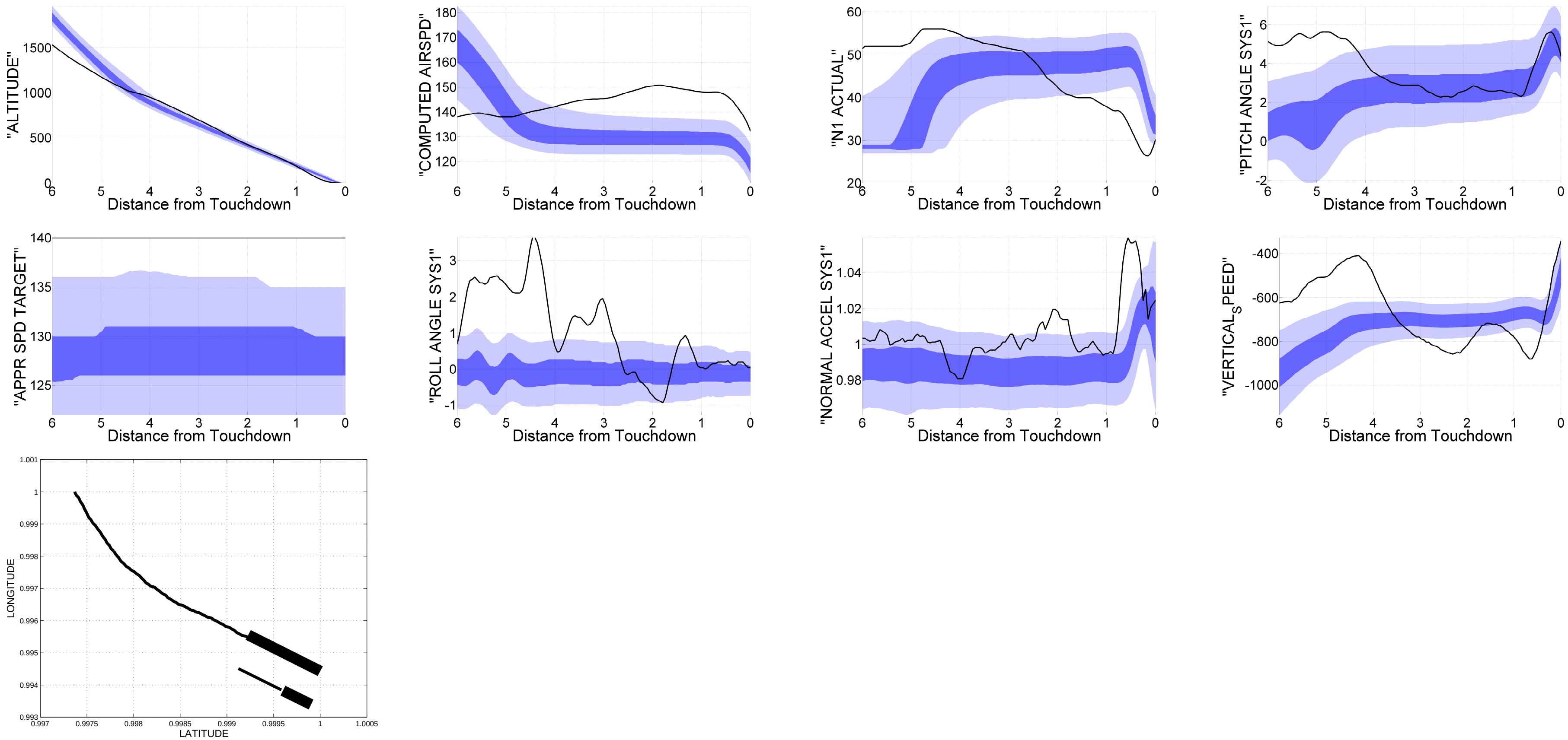

Figure 4. Parameter anomalies discovered by both Data-Driven Method ($ClusterAD/MKAD$): High energy approach - Continuous Parameters.

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

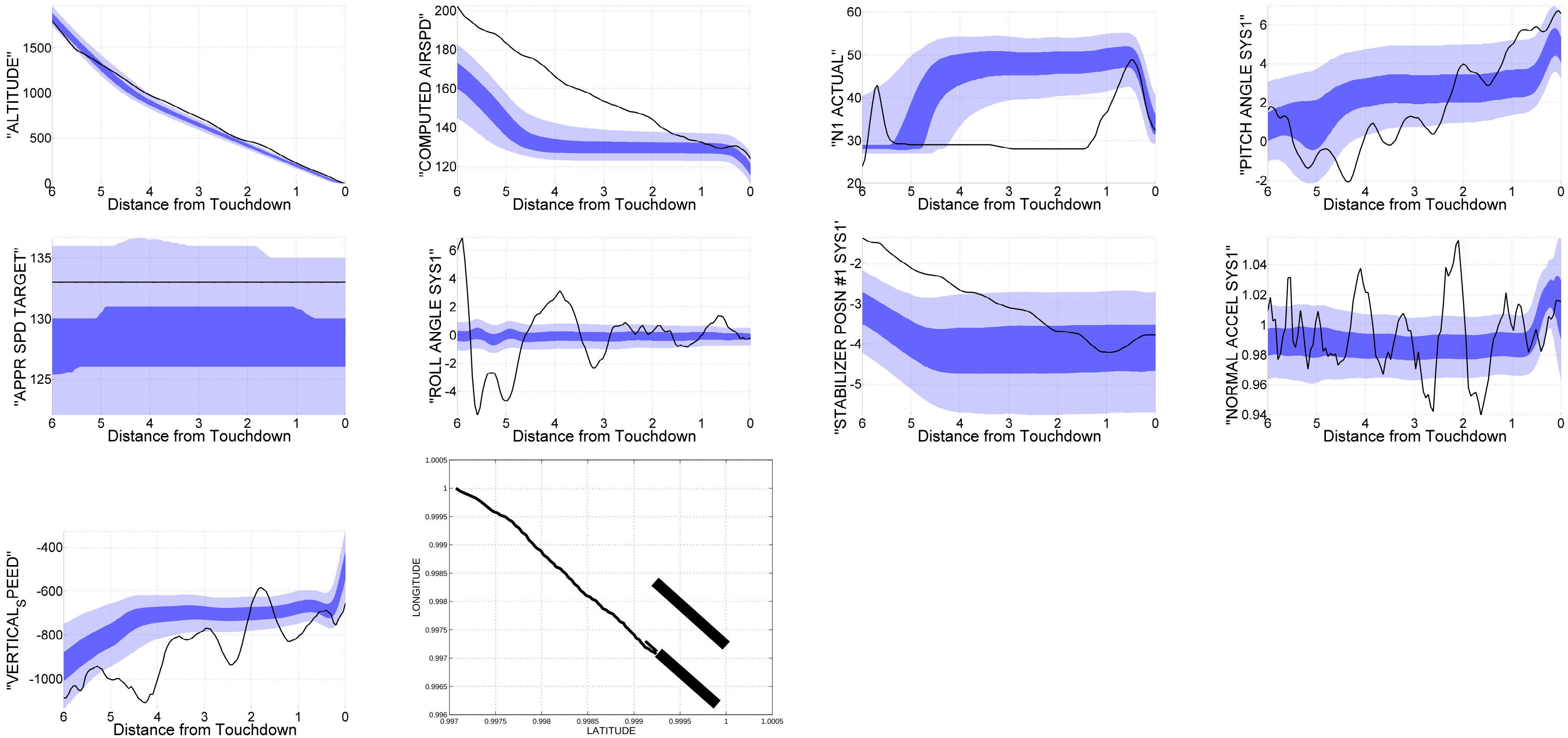

Figure 5. Parameter anomalies discovered by both Data-Driven Method (*ClusterAD/MKAD*): High-Airspeed with Low-Power Approach - Continuous Parameters.

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

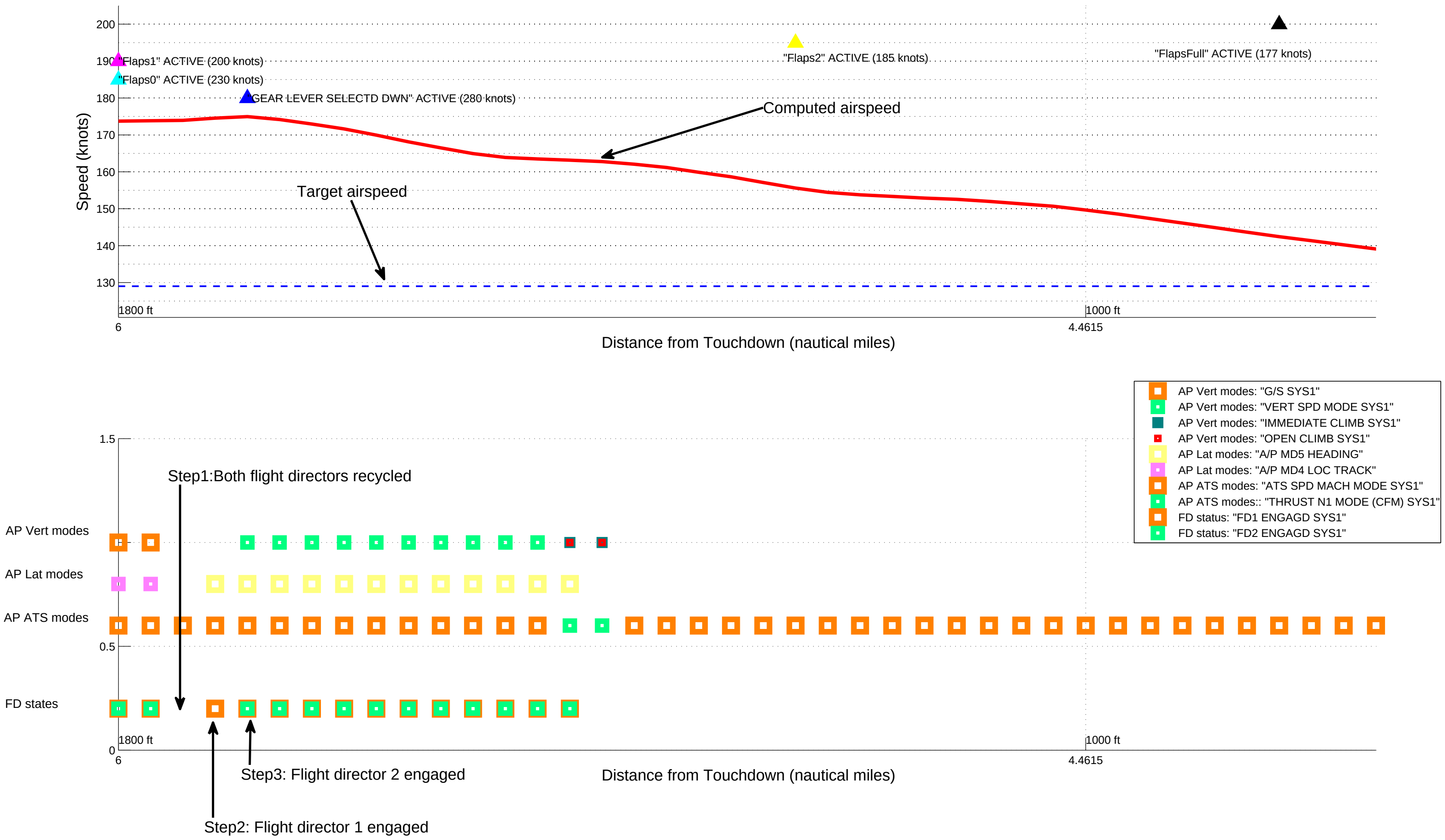

Figure 6. Flight detected by both Data-Driven Method ($ClusterAD/MKAD$): Recycling Flight Director - Discrete parameters

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

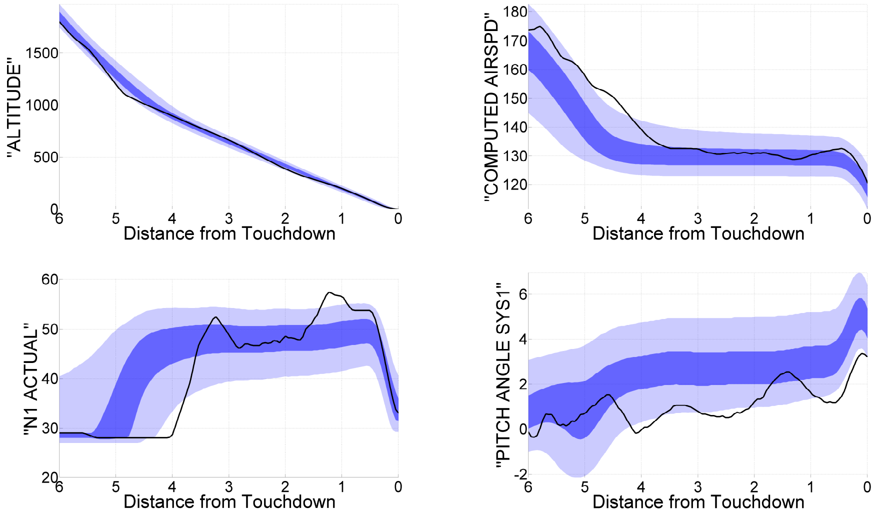

Figure 7. Parameter anomalies discovered by both Data-Driven Method ($ClusterAD/MKAD$): Recycling Flight Director - Continuous Parameters.

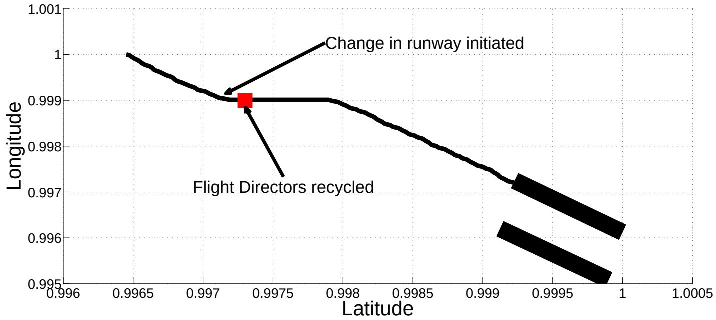

Figure 8. Change in Runway for the Recycling Flight Director Example.

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

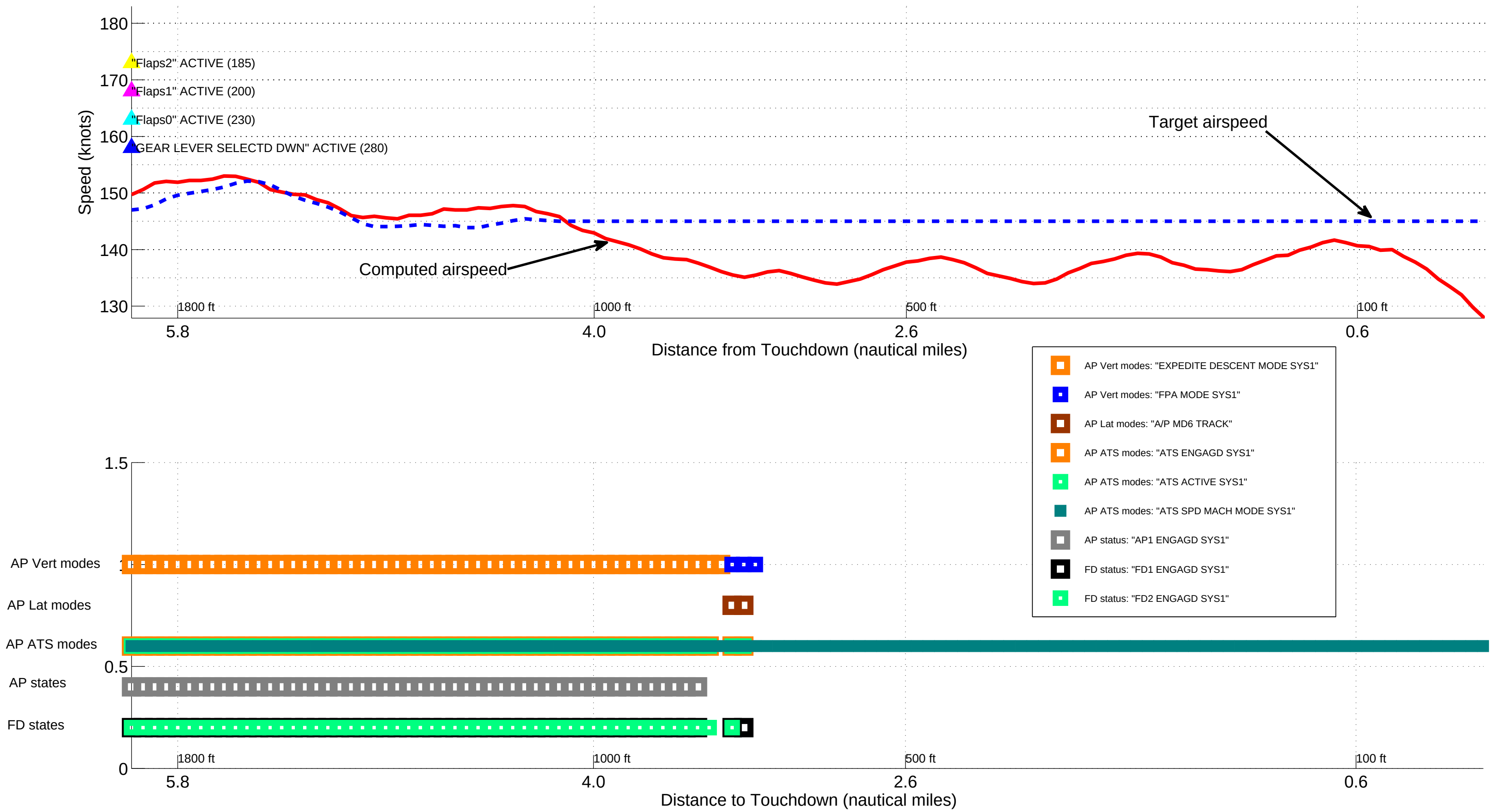

Figure 9. Flight detected by both Data-Driven Method ($ClusterAD/MKAD$): Wind Effect: Discrete Parameters showing Mode Transitions.

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

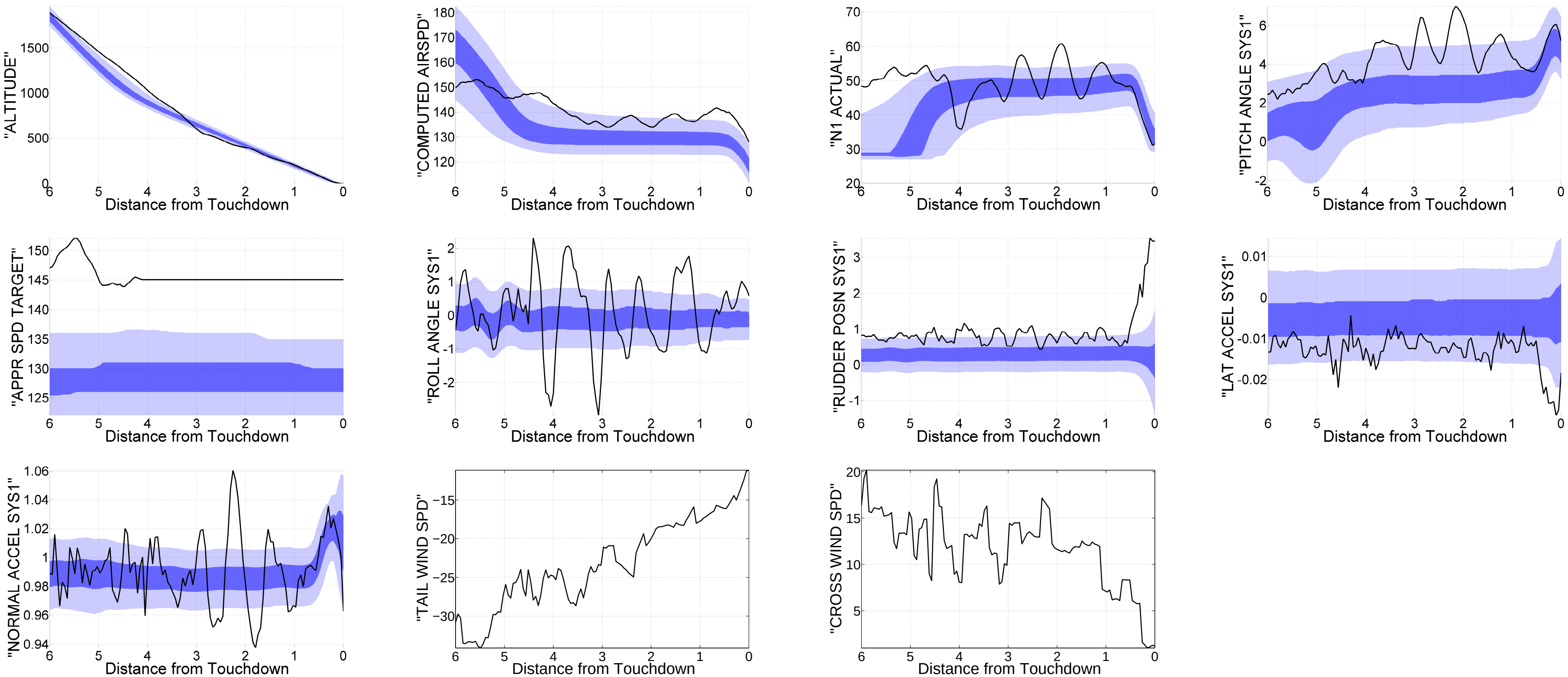

Figure 10. Parameter anomalies discovered by both Data-Driven Method ($ClusterAD/MKAD$):Wind effect - Continuous Parameters.

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

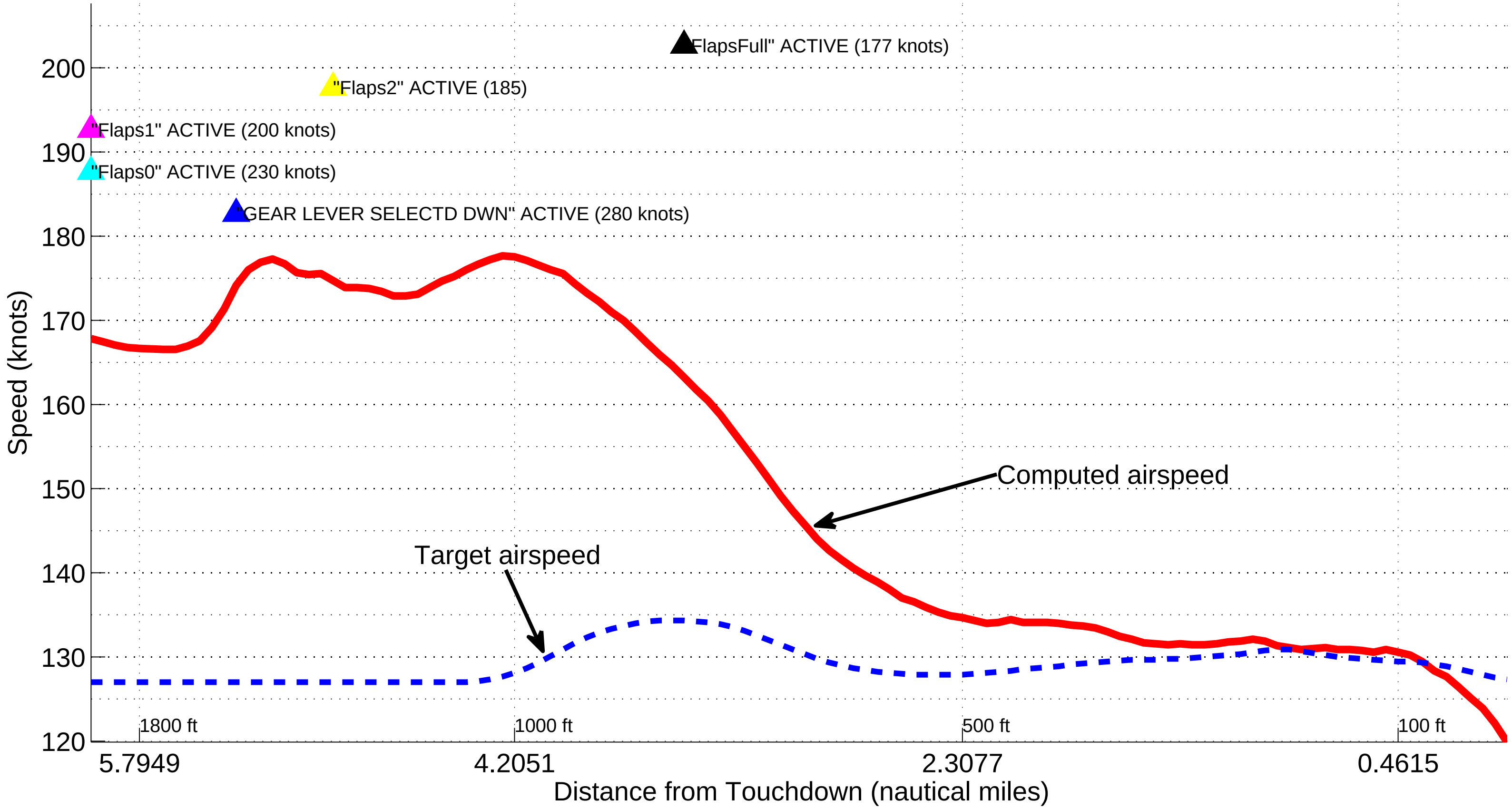

Figure 11. Flight detected by ClusterAD: Very High Airspeed - Discrete Parameters

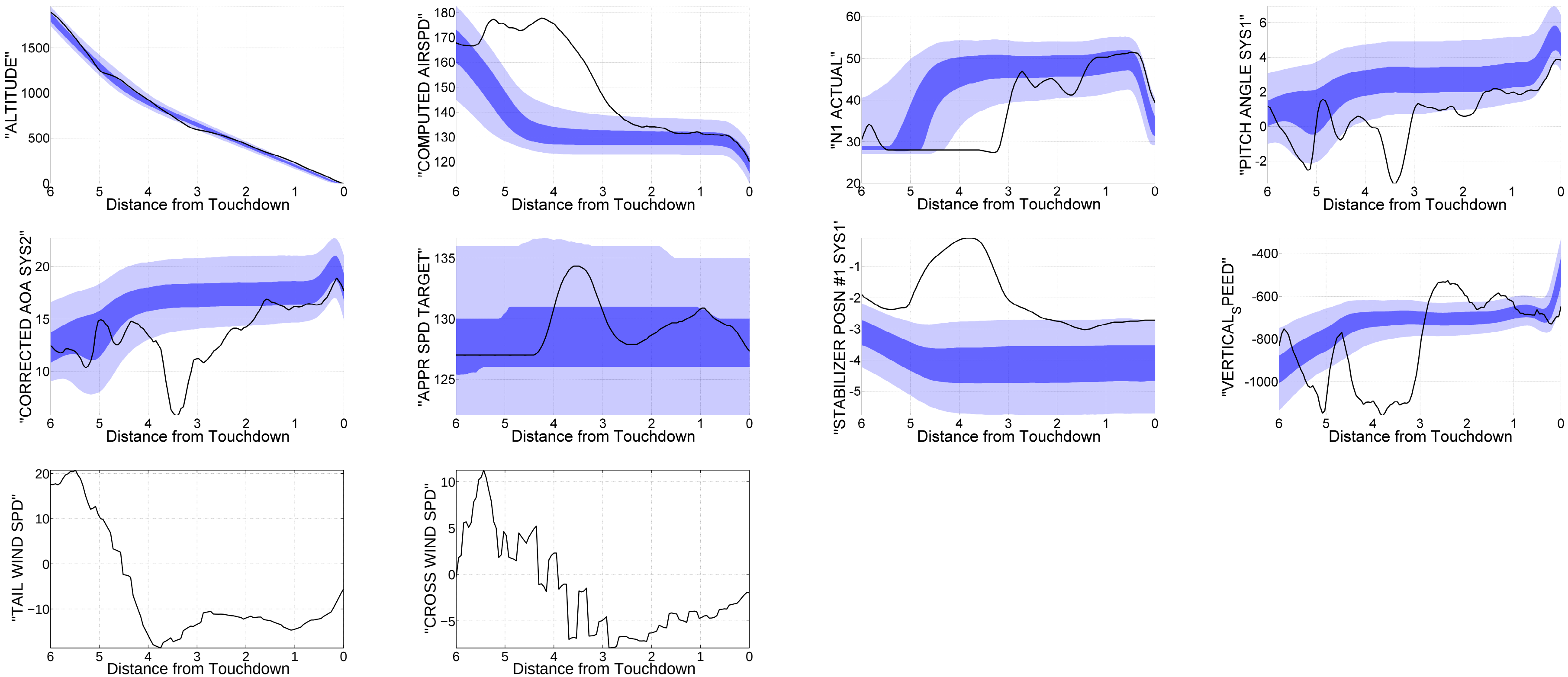

Figure 12. Parameter anomalies discovered by ClusterAD: Very High Airspeed - Continuous Parameters.

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

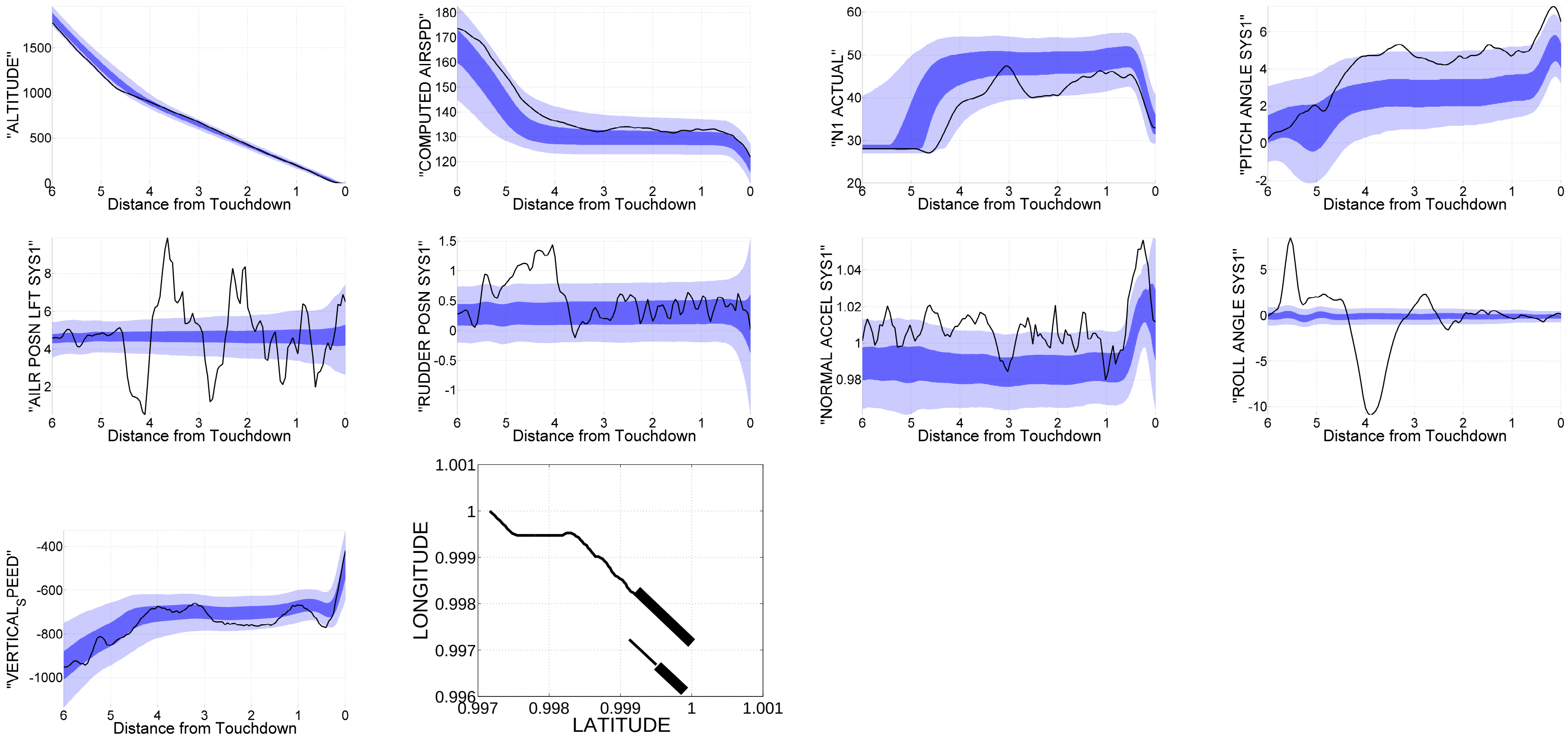

Figure 13. Parameter anomalies discovered by ClusterAD: Change in runway - Continuous Parameters.

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

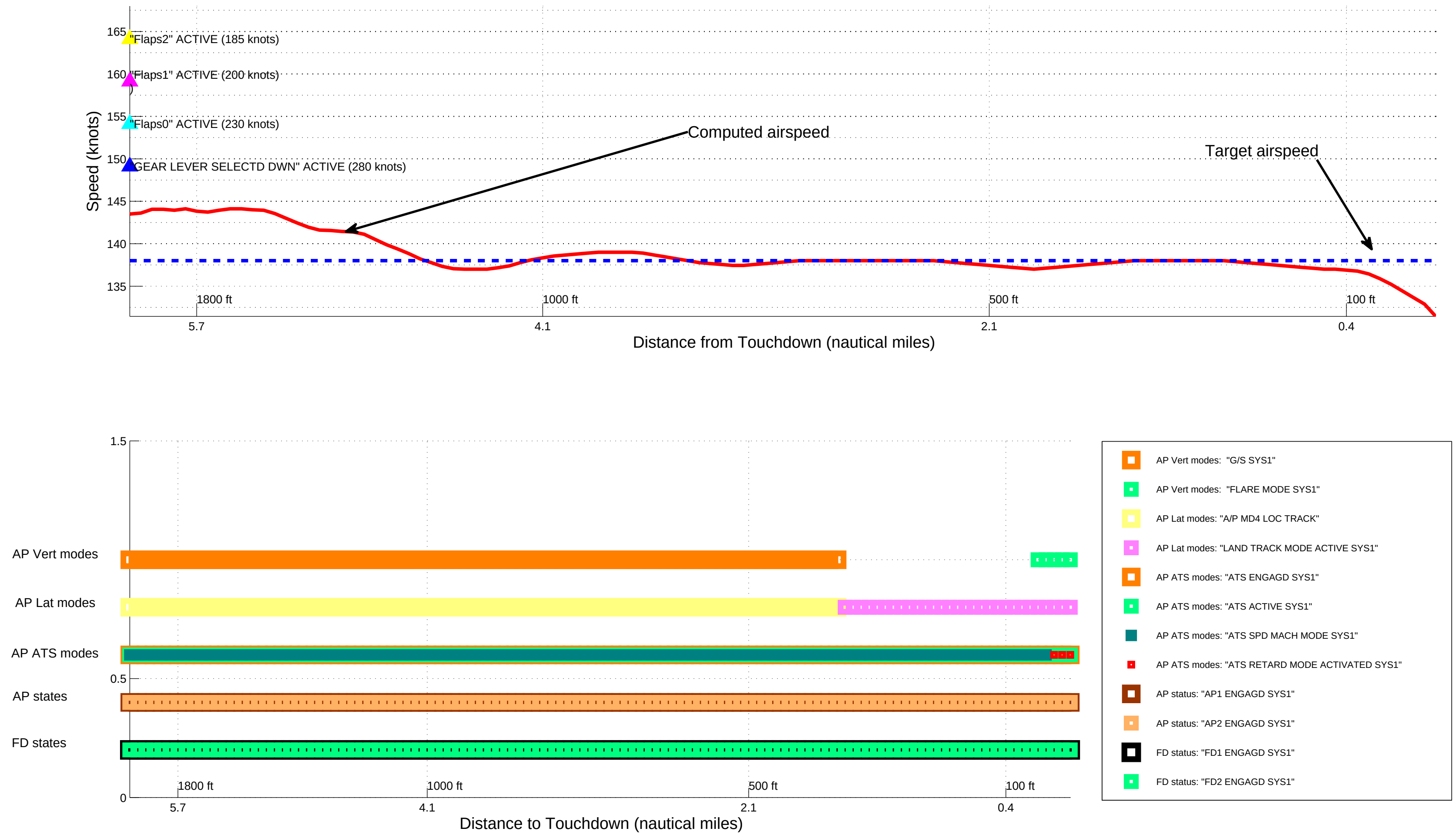

Figure 14. Summary of flight detected by MKAD: Unusual Auto Landing Configuration (Without Full Flaps) - Discrete Parameters

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

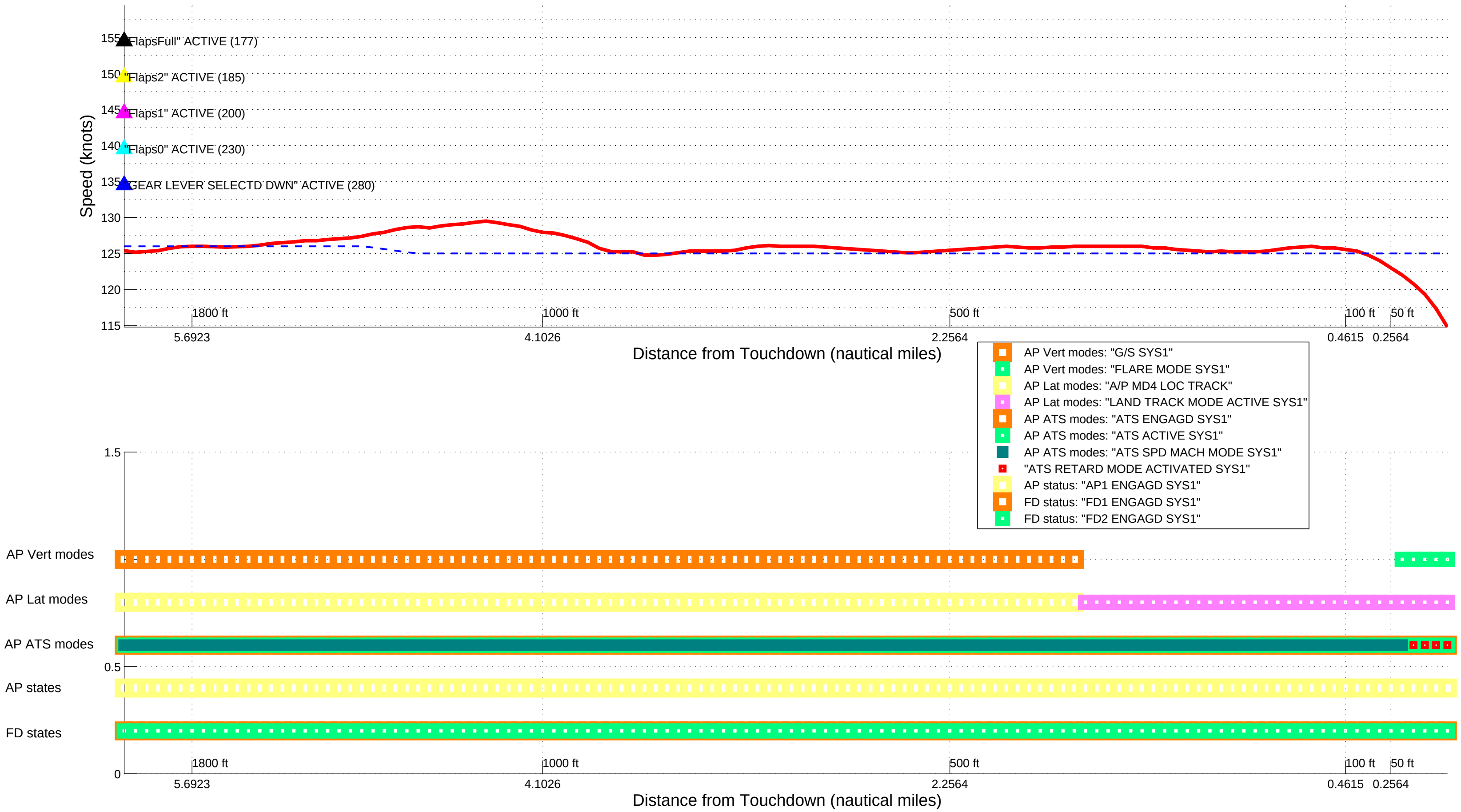

Figure 15. Summary of flight detected by MKAD: Unusual Auto Landing Configuration (Without both Autopilots Engaged) - Discrete Parameters

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

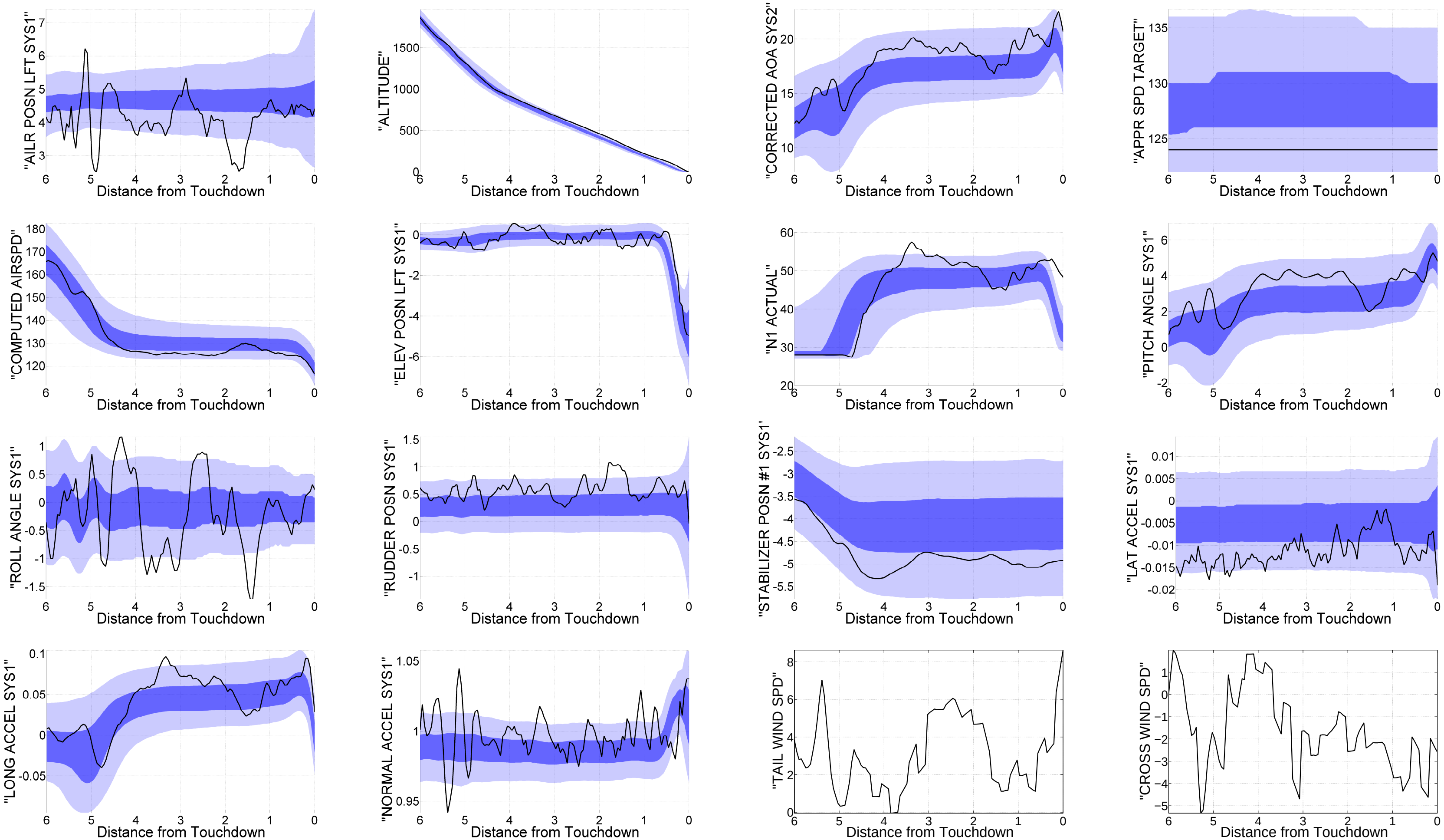

Figure 16. Anomalous Flight reported by Exceedance Detection : Flight 1 - Continuous Parameters

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593

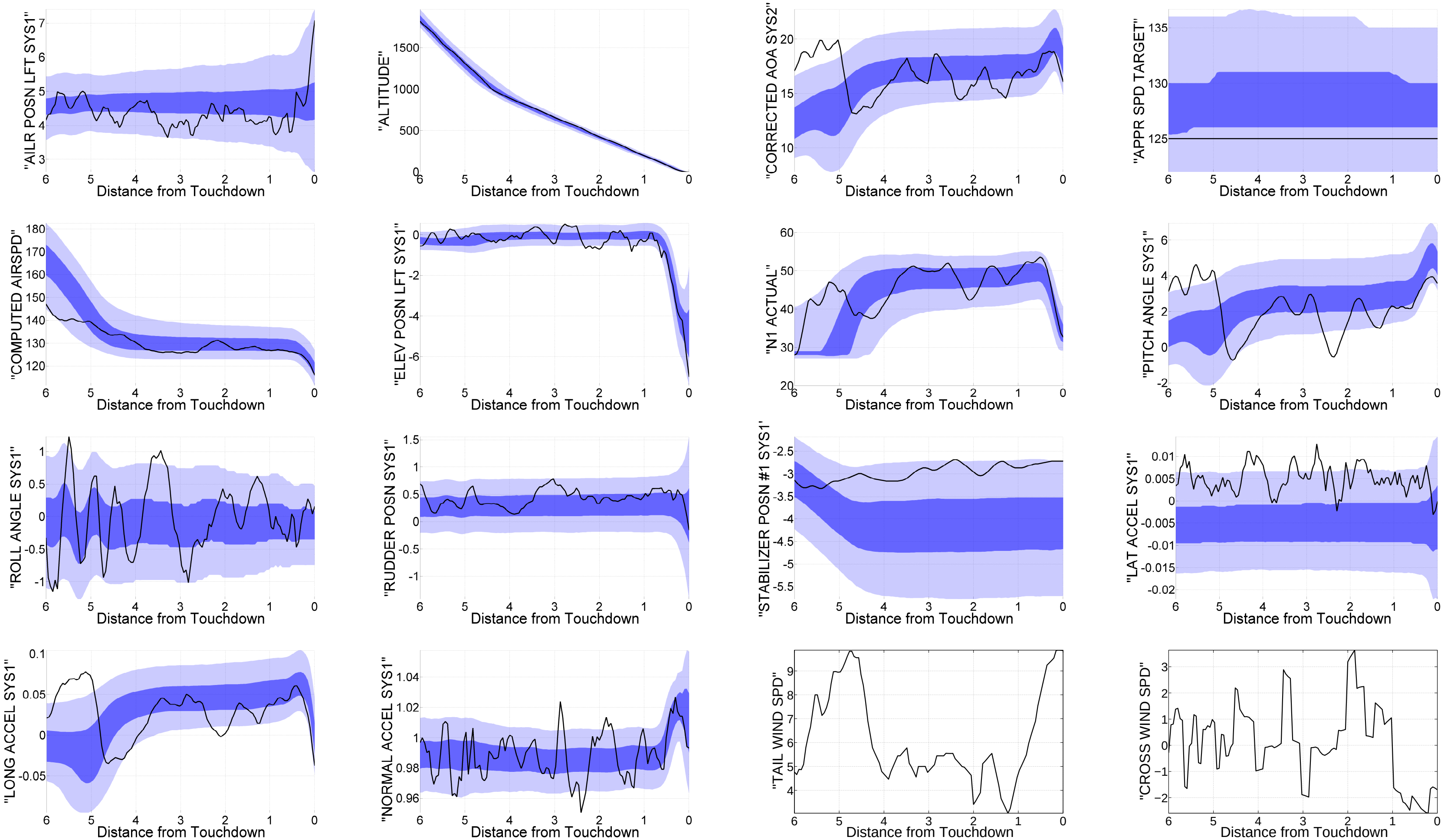

Figure 17. Anomalous Flight reported by Exceedance Detection : Flight 2 - Continuous Parameters

References

[1]S. Das, B.L. Matthews, and R. Lawrence. Fleet level anomaly detection of aviation safety data. In *Prognostics and Health Management (PHM), 2011 IEEE Conference on*, pages 1 –10, june 2011.

[2]Santanu Das, Bryan L. Matthews, Ashok N. Srivastava, and Nikunj C. Oza. Multiple kernel learning for heterogeneous anomaly detection: algorithm and aviation safety case study. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '10, pages 47–56, New York, NY, USA, 2010. ACM.

[3]Martin Ester, H.P. Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd international conference on knowledge discovery and data mining*, volume 1996, pages 226–231. Portland: AAAI Press, 1996.

[4]U.S. Department of Transportation Federal Aviation Administration. Flight operational quality assurance (ac 120-82). 2004.

[5]Lishuai Li, Maxime Gariel, R. John Hansman, and Rafael Palacios. Anomaly detection in onboard-recorded flight data using cluster analysis. In *Digital Avionics Systems Conference (DASC), 2011 IEEE/AIAA 30th*, pages 4A4–1 –4A4–11, oct. 2011.

[6]Bernhard Schölkopf, John C. Platt, John C. Shawe-Taylor, Alex J. Smola, and Robert C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, 2001.

Downloaded by NASA Ames Research Center on December 20, 2012 | http://arc.aiaa.org | DOI: 10.2514/6.2012-5593