Aircraft Anomaly Detection using Performance Models Trained on Fleet Data

Dimitry Gorinevsky*
Mitek Analytics LLC
Palo Alto, CA 94306
dimitry@mitekan.com

Bryan Matthews
SGT Inc., NASA Ames Research Center
Moffett Field, CA 94305
bryan.l.matthews@nasa.gov

Rodney Martin
NASA Ames Research Center
Moffett Field, CA 94305
rodney.martin@nasa.gov

***Abstract*—This paper describes an application of data mining technology called Distributed Fleet Monitoring (DFM) to Flight Operational Quality Assurance (FOQA) data collected from a fleet of commercial aircraft. DFM transforms the data into a list of abnormaly performing aircraft, abnormal flight-to-flight trends, and individual flight anomalies by fitting a large scale multi-level regression model to the entire data set. The model takes into account fixed effects: flight-to-flight and vehicle-to-vehicle variability. The regression parameters include aerodynamic coefficients and other aircraft performance parameters that are usually identified by aircraft manufacturers in flight tests. Using DFM, a multi-terabyte airline data set with a half million flights was processed in a few hours. The anomalies found include wrong values of computed variables such as aircraft weight and angle of attack as well as failures, biases, and trends in flight sensors and actuators. These anomalies were missed by the FOQA data exceedance monitoring currently used by the airline.**

I. INTRODUCTION

Flight Operations Quality Assurance (FOQA) programs collect high-rate aircraft data from each flight of hundreds of aircraft. This paper demonstrates a data processing approach for finding subtle anomalies in aircraft performance from very large FOQA data sets, automatically, accurately, and quickly. The anomalies are not characterized in advance of the processing, instead, they are detected as deviations from the performance observed for most aircraft in the fleet. As an interim step, a physically meaningful model of aircraft performance is built from the data. The anomalies are then detected as excessively large deviations from this model.

A. FOQA monitoring

A FOQA dataset for a single flight includes the same parameters that are usually collected by the crash-protected aircraft recorder and some additional data channels. Some 1000 channels are sampled at 1 sec interval through the flight duration and logged to yield tens of megabytes of data per flight. After the flight, the airborne data collected by a Digital Flight Data Acquisition Unit is transferred from the aircraft to a ground computing system. Most airlines process the collected FOQA data in a centralized way. For a medium size airline, data for all flights of all aircraft add up to a few terabytes per year.

*Corresponding author
This work was supported by NASA Contract NNX12CA02C

Airlines typically employ just a few FOQA analysts who do not have time to look at the data for hundreds of thousands of flights. This calls for automated FOQA data processing. A FOQA system must find a small number of anomalous flights and allow the analyst to focus on more detailed review of the anomalous data sets.

In the currently deployed FOQA systems, the automated processing first detects when the selected parameters exceed predefined thresholds. The thresholds have to be sufficiently large so that a small number of the exceedance events is generated. FAA Advisory Circular No 120-82 [1], which introduced FOQA, recommends establishing a Routine Operational Measurement - a sample of a chosen parameter at predefined points in time or space during every flight being analyzed. The baselines for normal operation are determined as mean, minimum, and maximum statistics of such data.

B. Anomaly detection

The simplest form of monitoring, known as Statistical Process Control (SPC), has been used in practice for several decades. SPC has been introduced for quality assurance of manufacturing processes. The classical SPC methods are univariate: a time series for a selected measured or computed process parameter is compared against control limits. The exceedances of the control limits are reported as anomalies. The FOQA exceedance monitoring approach closely resembles the classical univariate SPC.

Multivariate Statistical Process Control (MSPC) methods monitor many data channels simultaneously. MSPC can provide significant improvement over univariate SPC monitoring when the monitored channels are strongly correlated, as is often the case in practice. In the MSPC framework, the anomalies are commonly detected by computing Hotelling T2 statistics for the multivariable data [2].

MSPC is well established in industrial process areas such as refineries and semiconductor manufacturing. These are stationary processes operating around fixed setpoints. In contrast, aircraft data is nonstationary and has to be sampled at predefined conditions or preprocessed before MSPC can be used. One type of such preprocessing is computing the deviations from an aircraft performance model.

Some large airlines use proprietary performance models provided by aircraft manufacturers and monitor mismatches between these models and the aircraft data sampled in cruise regime. The model mismatches found are used to drive fuel performance improvements and to support aircraft maintenance. To use such models consistently, the airlines introduce focused programs in flight performance model maintenance to make sure they reflect actual aircraft performance. Such performance model maintenance programs can be afforded by large airlines only. There is no indication in the literature that they use multivariate approaches, such as MSPC, for monitoring the performance model deviations.

An alternative to using the proprietary models is presented by data driven models trained using massive amounts of historical data. Regression models of FOQA data [3] and clustering models [4], [5] have been considered in earlier NASA work. Use of clustering models for aircraft operation anomaly detection was considered in MIT work [6]. Regression models for aircraft performance were considered in a Stanford paper [7].

C. Distributed Fleet Monitoring

This paper demonstrates an application of a data driven monitoring approach called Distributed Fleet Monitoring (DFM) to FOQA data. DFM analytical software was developed by Mitek Analytics LLC under a series of NASA-funded projects. DFM has been verified using realistic simulated FOQA data in an earlier unpublished study [8].

DFM builds performance models of aircraft as a regression fit of the FOQA data. The nonlinear regressors used for aircraft performance modeling have well understood structure - the same as for proprietary aircraft performance models developed by aircraft manufacturers. The standard approach is to fit the aircraft performance models to flight test data. DFM differs by fitting the models to the historical operational data.

The data-driven performance models in DFM are used to remove the bulk of the data variability by computing the model prediction residuals. MSPC methods are then applied to the residuals. Determining a linear regression model and then applying MSPC to the model prediction residuals is related to the Partial Least Squares (PLS) method of MSPC.

DFM is a fleet-wide multi-level MSPC method. It extends the known approaches of data-driven regression modeling of performance, model-based calculation of the residuals, and MSPC monitoring of the residuals to include fixed effects in the model. DFM uses a three-level regression model for aircraft performance. The first level describes time inside the flight, the second level described flight-to-flight variability and trends, and the third level describes the vehicle-to-vehicle variability.

Multi-level regression models have been earlier used in social science, drug testing and other applications, see [9], [10]. Much smaller data sets were used and no scalable exact solution methods proposed. The processing of aircraft fleet data is considered in [11] using a two-level regression model, which is related to the three-level model used by DFM.

D. Contributions

This paper demonstrates an application of DFM to an airline FOQA data set of half million flights. DFM was able to detect a number of anomalies of interest that were missed by existing airline FOQA data analysis methods.

The demonstrated approach complies with the FAA recommendations [1] in how it handles the secured raw data. As the initial step, the raw FOQA data is preprocessed into de-identified compressed data that does not contain the details of the flight. The subsequent post-processing of the compressed data yields a report of anomalies in aircraft performance. As required by [1], the data is analyzed in relation to existing aggregate information (the fleet performance model). The reported approach emphasizes support of airline Engineering Maintenance and Flight Safety functions.

The demonstrated DFM approach extends the existing FOQA practices. A data-driven performance model is established as the normal operation baseline. The anomalies are detected as deviations from the baseline. DFM novelty compared to the existing FOQA practices is in using a nonlinear model with physics-based structure that is multivariate and multi-level. The existing FOQA systems watch for univariate exceedances. DFM looks at cross-fleet multivariate statistics for abnormal models, model prediction residuals, and trends. Similar to the existing FOQA systems, DFM reporting is automated so that the analyst can focus on a small number of important anomalies.

DFM algorithms can work in a grid computing environment where the data are distributed over multiple nodes and the bulk of the computing is collocated with the data. The grid computing aspect of DFM was outside of the scope of this study. In this study, the DFM algorithms were implemented on a NASA ARC Unix cluster with substantial shared disk storage capacity. The entire 5 Tb data set was stored on the disk and accessible through the cluster file system.

The contributions of this paper are as follows. (i) The paper demonstrates practical efficacy of regression modeling of airframe performance using a large set of FOQA data. (ii) It shows that computational implementation of multilevel regression modeling can be scalable. FOQA data for half million flights is processed in a few hours. (iii) The approach is demonstrated to detect several classes of real life FOQA data anomalies missed by the standard FOQA analysis.

II. Regression Modeling

This section describes the regression modeling in the center of the DFM data processing approach.

A. Simple regression

In this work a regression model of the nominal aircraft flight performance is trained using the FOQA data. As a starting point consider a simple linear regression model of the aircraft dynamics relating the FOQA record data and the performance data derived from the FOQA data at a given instant (time sample) t. The model has the form

$$y(t) = Bx(t) + v(t), \tag{1}$$

where y is the vector of performance variables (nonlinear combinations of the data channels), x is the vector of regressors variables, B is the matrix of regression parameters encoding the performance model, and v is the vector of residuals (noise). The regression parameters include aerodynamic coefficients, engine thrust coefficient, etc. A cruise flight model with similar structure is considered in [7].

In this paper we consider aircraft flight performance in enroute flight including climb, cruise, level turns, and descent. We extract the segment of the FOQA data limited by the altitude and the range of the Mach number. The enroute flight can be described using models of the aircraft steady flight performance discussed in [12]. It is well known that the aerodynamic forces acting on the aircraft can be represented in the form

$$F_{aero} = \bar{q}C_{a,0} + \bar{q}C_{a,1}a + \bar{q}C_{a,2}u_1 + \ldots + \bar{q}C_{a,n+1}u_n, \quad (2)$$

where a is angle of attack (AOA), $u_1, \ldots, u_n$ are control surface deflections, and $C_{a,0}, C_{a,1}, \ldots, C_{a,n+1}$ are the model coefficients (aerodynamic coefficients times the cross section area). The dynamic pressure $\bar{q} = \frac{1}{2}\rho_{air}V^2$, where ρ_{air} is the air density and V is the airspeed. The engine thrust was modeled to be proportional to air density and fan (propeller) speed, see [12].

Regression parameters in B include the aerodynamic coefficients C in (2), aircraft mass model parameters, and thrust model parameters. The results of this work demonstrate that a fixed regression model B can describe the entire enroute flight segment with aircraft in the clean aerodynamic configuration.

The columns of matrix B can be computed by least squares regression fit of the actual FOQA data. One possible approach is to fit the model for a single flight data set. This will not be very accurate because such data insufficiently covers the operation of the aircraft fleet. As solutions to ill-conditioned problems with noisy data, the models fitted to different flight data sets might differ substantially. Another possible approach is to fit the model to the pooled data for the entire fleet. Such model will be much more stable but would completely miss the fixed effects, the fact that all individual aircraft have slightly different performance. The pooled model would also miss the flight-to-flight trends in the aircraft performance. The problem formulation in the next section addresses these issues.

B. Three-level regression

We consider a three-level regression model for a fleet of aircraft. The FOQA data are used to compute response variables $y \in \Re^m$ and explanatory variables $x \in \Re^n$ (regressors) as discussed in the previous subsection. Computing these variables (e.g., the dynamic pressure) is a nonlinear transformation of the raw FOQA data.

The three-level model considers the response (output) variables $y_{i,kf}(t)$ that depend on

t - a sample number inside a given flight record
i - a vector component number,
k - a tail number
f - a flight number

The three-level model has similar indexing for the explanatory (regression) variables $x_{j,kf}(t)$, and model residuals $v_{i,kf}(t)$. The model can be written in the form

$$y_{kf}(t) = B_k x_{kf}(t) + a_{kf} z(t) + v_{kf}(t), \quad (3)$$

where $x_{kf}(t) \in \Re^m$ is the regressor vector, $y_{kf}(t) \in \Re^m$ is the response variable vector, $B_k \in \Re^{m,n}$ is the model for tail k, each flight has a bias $a_{kf} \in \Re^m$, and $z(t) = 1$ describes the bias that is fixed inside the flight.

The regression fit problem could be posed as minimization of the least-squares loss index L

$$L = \sum_{k=1}^{K}\sum_{f=1}^{F_k}\sum_{t=1}^{T_{kf}} \|y_{kf}(t) - B_k x_{kf}(t) - a_{kf}z(t)\|^2 + \rho\sum_{k=1}^{K}\sum_{f=2}^{F_k}\|a_{kf} - a_{k,f-1}\|^2 + \mu\sum_{k=1}^{K}\|B_k - B_*\|_F^2, \quad (4)$$

where F_k is the number of flights for tail k in the data set, T_{kf} is the number of the samples in the flight data collected in flight f of tail k, and B_* is the unknown central model for the fleet.

The problem of minimizing (4) can be interpreted as optimal Bayesian estimation of the three-level regression parameters from the pooled fleet data. Loss index (4) describes three levels of the regression fit and includes three main components corresponding to posterior and prior probabilities in the Bayesian model

1) The data for each flight f of each tail k is described by the model fit residuals $y_{kf}(t) - B_k x_{kf}(t) - a_{kf}z(t)$. The first sum in (4) describes the accuracy of the three level regression fit pooled across all flights of all tails. This term corresponds to the negative log posterior probability of the observation noise.
2) The prior for flight-to-flight trend (bias change) is defined by the quadratic penalties in the flight-to-flight trend increments $\|a_{kf} - a_{k,f-1}\|^2$. This second sum in (4) corresponds to the negative log priors in independent random walk models for trends a_{kf}.
3) Tail-to-tail model prior is defined by the quadratic penalty $\|B_k - B_*\|_F^2$, where B_* is an unknown fleet average model. The third sum in (4) corresponds to the negative log prior for normal distribution of the models B_k with the mean B_*.

Computing parameters a_{kf}, B_k, and B_* of the three-level regression (3) by minimizing (4) is a batch problem. The solution is discussed in the next section. The solution could be also implemented incrementally as an extension of the well known recursive least squares method.

III. DFM Algorithm

Figure 1 illustrates a functional decomposition of the Distributed Fleet Monitoring (DFM) logic described in [8]. The overall monitoring data function takes the aircraft data (raw FOQA data) and reports monitoring results, such as anomalies. The collection of these functions fits a three-level regression to the fleet data and reports anomalous deviations from this model as described in more detail below. The word Distributed is in the DFM name because the preprocessing is done for one flight at a time and can be implemented as distributed and parallel computations.

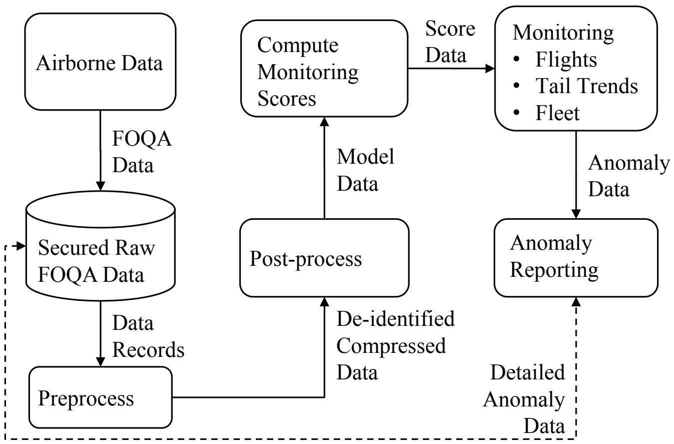

Fig. 1. Computational logic flow of the DFM algorithm.

A. Regression model training

The regression fit problem is to minimize loss index (4) with respect to a_{kf}, B_k, and B_*. To do that, consider quadratic form reduction for the first term in (4). By expanding the norms in (4) one can see that the contribution of all data for the entire flight into the loss index L can be described through the scatter matrices

$$\sum_{t=1}^{T_{kf}} y_{kf}(t)y_{kf}^T(t), \quad \sum_{t=1}^{T_{kf}} x_{kf}(t)x_{kf}^T(t), \tag{5}$$
$$\sum_{t=1}^{T_{kf}} x_{kf}(t)y_{kf}^T(t), \quad \sum_{t=1}^{T_{kf}} x_{kf}(t)z(t), \quad \sum_{t=1}^{T_{kf}} y_{kf}(t)z(t).$$

The dimensions of these matrices are n, m, or 1.

The solution of the regression problem now looks as follows: first, compute the scatter matrices for all flight records, then, minimize the loss index L with respect to a_{kf}, B_k, and B_*. The FOQA data collected in one flight might contain some 1000 channels sampled at 12,000 instances to yield some 100 MB of the data. For a couple dozen regressors and outputs, the scatter matrices would take about a KB of memory. Thus, computing these matrices provides data reduction on the order of 100,000:1. The original 5 TB of raw FOQA data are reduced to about 50 MB of the scatter matrices. The scatter matrix data fit into computer memory, which allows solving the minimization problem efficiently. A version of the algorithm described in [11] could be used for solving the problem of minimizing (4).

Computing the scatter matrix data corresponds to the Preprocess step in Figure 1. Solving the minimization problem to obtain the trends a_{kf}, and models B_k, B_*, from the scatter matrix data is shown as Post-process in Figure 1.

B. Monitoring

Monitoring of the anomalies relies on the knowledge that the majority of the aircraft in the fleet and data sets for each aircraft are nominal. A small percentage of the aircraft and of the flights might be abnormal and need to be reported as such. The DFM automated monitoring system processes the data without human intervention and provides anomaly reports in the end. These reports provide decision support and can be reviewed or acted upon by a human operator.

Minimizing index (4) yields estimates of the regression models B_k, and trends a_{kf} for all vehicles in the fleet. These estimates and the data $y_{kf}(t)$ allow computing model prediction residuals $v_{kf}(t) = y_{kf}(t) - B_k x_{kf}(t) - a_{kf}z(t)$.

The processing results are used to compute three types of the monitoring scores as Hotelling T2 statistics for the respective multivariable data.

$$T^2_{v,kf} = \frac{1}{T_{kf}} \sum_{t=1}^{T_{kf}} T^2(v_{kf}(t)), \tag{6}$$
$$T^2_{a,kf} = T^2(a_{kf}), \tag{7}$$
$$T^2_{B,k} = T^2(B_k). \tag{8}$$

These T2 statistics are based on the empirical means and empirical covariances of the data. The score $T^2_{v,kf}$ (6) for the model prediction residual is the Hotelling T2 statistics for the residuals $v_{kf}(t)$ averaged across all data points in a flight. Using the score $T^2_{a,kf}$ (7) for the trend corresponds to the Multivariate Exponentially Weighted Moving Average (MEWMA) method of MSPC [2]. The use of residual statistics $T^2_{v,kf}$ corresponds to MEWMA Wandering Mean prediction error used for detecting an abrupt change in conjunction with the MEWMA method, see [2]. The score $T^2_{B,k}$ (8) is the Hotelling T2 statistic describing the deviation of the tail model from the population average. Figure 1 shows the calculation of the T2 statistics (6), (7), (8) for the residuals, trends, and models (Score Data) as Compute Monitoring Scores block.

C. Anomaly reporting

Computing Hotelling T2 statistics (7) for the estimated trends a_{kf} for all aircraft tail and comparing it with a threshold allows detecting trend anomalies. The abrupt change anomalies can be similarly detected through Hotelling T2 statistics (6) for the regression model residuals $v_{kf}(t)$. The model anomalies can be computed by thresholding T2 statistics (8) for the models B computed for all aircraft tails.

The anomalies are detected when the T2 statistics (Score Data) exceed the respective thresholds. The thresholds are established from the false positive/false negative alarm tradeoff. The Monitoring block in Figure 1 detects three types of anomalies from the Score Data: (i) anomalous single flight residuals, (ii) anomalous tail trends, and (iii) aircraft performance models that are anomalous compared to other tails in the fleet.

The Anomaly Reporting block in Figure 1 takes the Anomaly data produced by the Monitoring block and generates more detailed anomaly reports in a form accessible to human operators. The report includes summary conclusions for operators and maintenance personnel. The report can also include detailed engineering information in the form of detailed graphs, charts, and tables for the engineering personnel.

The Anomaly Reporting block could also produce detailed anomaly reports. For a particular anomalous FOQA flight record, detailed charts of the monitored variables $y_{kf}(t)$ (flight f for tail k), of the regression model fit, and of raw FOQA data variables could be displayed to help with establishing the root cause of the anomaly. This requires fetching the selected Detailed Anomaly Data - a raw FOQA data record - from the Secured Raw FOQA Data set as illustrated in Figure 1. Only a small number of the raw data records generally need to be fetched to investigate the anomalies.

IV. FOQA Dataset

This section provides an overview of DFM application to commercial airline data for a fleet of aircraft.

A. Data set description

We used DFM to process data for a medium-size airline fleet that included Airbus A319 and A320 aircraft. The fleet included 188 aircraft (tail numbers). The data was collected over two years: 2010 and 2011. In this period each aircraft has made between 200 and 3000 flights. The overall data set included about half million flights.

The A319 and A320 aircraft have the same body and wing shape, but differ in length (123ft for A320 vs 111ft for A319). The weight ranges and engine thrust ranges of the two aircraft types mostly overlap, but are some 5% higher for A320.

The data channels used for processing and monitoring included: time stamps, aileron left and right, elevator left and right, rudder, stabilizer, angle of attack, lateral, longitudinal and normal accelerations, aircraft weight, pitch, roll, fan speeds for two engines, mass fuel flow for two engines, total temperature, altitude, and airspeed.

In addition to the described raw data, derived parameters were used. These included dynamic pressure, Mach, and air density computed based on atmosphere model.

B. Implementation overview

DFM algorithms are implemented using a pipeline computational architecture. The architecture consists of data pipes each reading inputs from a buffer (disk files) and writing outputs into another buffer. Each pipe implements a stage of the data processing. By going through the input data files sequentially, the pipe can process large amounts of data that do not fit into memory. The DFM data pipes roughly correspond to what is shown in Figure 1. The first datapipe ingests the raw FOQA data, the last datapipe outputs the Anomaly Reports.

In this work, DFM software was implemented on a Unix cluster. Embarrassingly parallel implementation of the algorithm was used. The parallel and sequential implementation allowed multi-Terabyte data processing with the standard limited memory of a few Gb at each processor. The computational logic implementing the datapipes was programmed in Matlab and deployed as Matlab-generated Unix binaries. The parallel execution of 16 threads on as many cores allowed to complete the DFM processing of the 5 TB data set in about 10 hours.

C. Processing results overview

The anomaly detection thresholds were initially calculated based on 5% p-values for the T2 statistics. Later, it was found that the outlier distribution in the data is, in fact, heavy tailed and the thresholds were empirically increased by an order of magnitude to limit the number of the flagged anomalies such that they can be surveyed in detail.

No model anomalies were flagged. A possible reason is that the fleet was a mixture of A319 and A320 aircraft. The differences between these two types of aircraft established the large range of the normal model variability. No vehicle in the fleet differed from the rest by much more than this variability.

For reporting, the monitoring scores were scaled by 5% p-values of the T2 statistics. For the estimated trends, the bulk of the scaled monitoring scores was less than 0.1. The 15 tails flagged for the trend anomalies had monitoring scores exceeding 10.

The bulk of the scaled monitoring scores for the residual anomalies was less than 0.2. The 20 tails flagged for the residual anomalies had scores exceeding 6. Most of the flights with residual anomalies were also flagged as trend anomalies.

Many intermediate-sized anomalies have smaller magnitude. They clearly stand out compared to the bulk of the data, but were left out in this study. These anomalies could be of interest to aircraft operator once all larger anomalies are addressed. The flagged anomalies with large trend and residual monitoring scores involve 24 tails. These anomalies were analyzed in more depth and are described in more detail below.

V. Anomalies Found

The following large anomalies were uncovered as a result of the DFM processing of the described FOQA dataset.

A. Angle of Attack

Wrong value of Corrected Angle of Attack (AOA) was recorded for three tails. In these cases the Indicated AOA value looks normal. The corrected AOA value is stuck at zero for one aircraft. The corrected AOA values are stuck at 43 and at 42 for two other aircraft. For each of these three aircraft, the anomalies persisted for several sequential flights then disappeared. The problem is illustrated in Figure 2 that plots scaled flight anomaly score (6) for sequential flights of one of the three aircraft The amplitude of anomaly (the monitoring score) is very large for three flights. Note that the score is above the variation (though below the selected threshold) for several flights around the event. The Corrected AOA and the Indicated AOA at the peak amplitude of the anomaly are included as an insert plot in the figure.

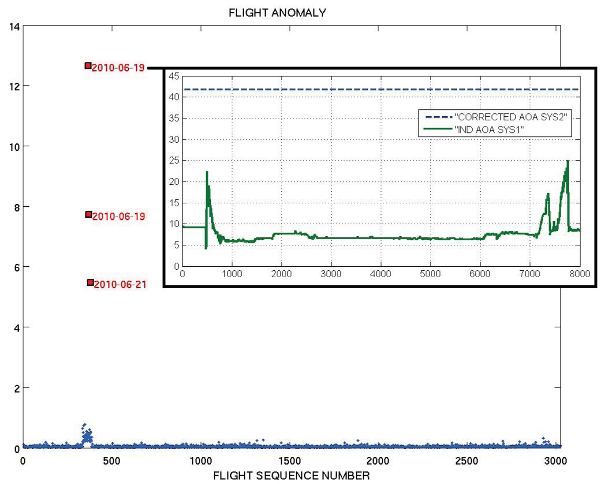

Fig. 2. Corrected AOA anomaly

B. Accelerations

Problems with three accelerometer readings were found for two tails. In one flight, the acceleration signal dropped to an unreasonably low value mid-flight. This is a single event. For another aircraft, the deterioration evolved over 18 flights. Scaled trend anomaly score (7) in Figure 3 shows the amplitude of the anomaly increasing from one flight to another. The three-channel acceleration oscillations at the peak amplitude are included as the insert plot in the figure. It is assumed that after reaching the peak the accelerometer problem was resolved. If the anomalous accelerations were caused by an Inertial Navigation System failure, this could potentially have safety implications. The accelerations influence the aircraft Air Data system through Air Data Inertial Reference System (ADIRS). It is assumed that the oscillating accelerations did not trigger the exceedance event until the amplitude reached its peak. Despite the highly abnormal pattern, the accelerations did not reach unusually high or low values earlier.

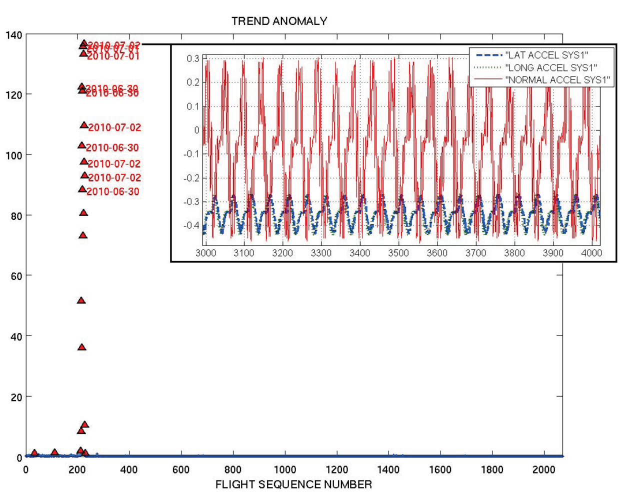

Fig. 3. Acceleration sensor anomaly

C. Aircraft gross weight

In multiple flights, the aircraft gross weight is indicated as 100x larger than normal during first 20-60 min of flight. At the end of this period, the weight suddenly drops to a normal value consistent with the model. For 16 aircraft in the fleet, there are multiple occurrences of this happening. The abnormal gross weight indication is not correlated with any observed changes in the aircraft fuel capacity or other related channels. It does not seem to influence operation of aircraft flight control. This appears to be a problem with FOQA data collection, rather than an actual aircraft related problem.

D. Elevator oscillations

In a single flight of one tail, the left elevator starts oscillating from -26 to 11 deg some 20 min into a flight. No flight attitude disturbance is visible. A possible culprit is an electrical failure in the actuator control circuit that is disconnected from the actuator. In this case, the airline does not have a record of elevator maintenance performed on this tail between the anomalous flight and the next flight when the anomaly disappeared.

E. Elevator bias

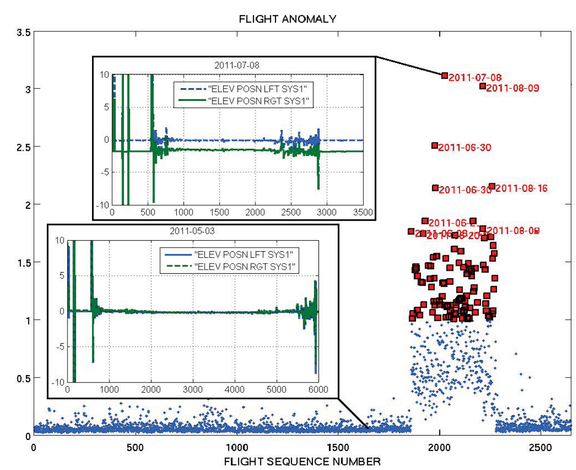

Fig. 4. Elevator bias: persistent bias in the right elevator

For one tail, a 1.5 deg right elevator bias was found. The anomaly persisted for about 400 sequential flights and was absent in the flights before and after that period. This anomaly is illustrated in Figure 4 showing scaled flight anomaly score (6). The upper insert plot in Figure 4 shows two elevator traces in the flight with the largest monitoring score. The lower insert plot shows the elevator traces for normal data, before the 400-flight elevator bias event started.

F. Aileron bias

The anomaly shown in Figure 5 is related to a bias of one of the ailerons. The plot shows a trend anomaly score (7) that increases in a series of 20 flights. After reaching its peak, the anomaly disappears. No maintenance actions related to these anomalies were confirmed by the airline.

The upper insert plot in Figure 5 shows the sum of aileron positions in the flight with one of the largest monitoring scores. The lower insert plot shows the aileron sum for normal data, before the start of the increasing anomaly trend.

An interesting observation is that the ailerons sum is -2 deg in the normal condition, which is the condition maintained in the vast majority of flights for this tail. Near the anomaly peak, the sum of the aileron position is close to zero. This differs from the described normal condition.

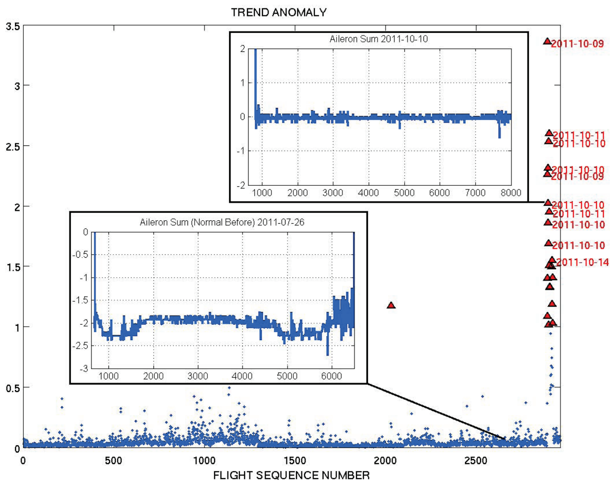

Fig. 5. Aileron bias: trend in the sum of aileron positions

G. Smaller biases

Reducing the anomaly reporting threshold uncovers many additional anomalies that are smaller in magnitude but still well outside of the variation range for most of the data. Many of these anomalies are related to biases in flight actuators or sensors. The limited space of this paper does not allow us to provide a more detailed survey of these smaller anomalies. The actuator bias anomalies might be important because they could lead to changes in aircraft trim and, as a result, might increase the fuel burn compared to an optimized trim flight. They might also indicate an incipient actuator problem.

ACKNOWLEDGMENT

The authors would like to thank Dr. Robert Mah at NASA ARC for supporting an early version of this study within NASA IVHM Architecture NRA project. We are grateful to the airline personnel for providing the data, the discussion of the results of this study, and for the permission to publish the material in this paper. It is the wish of the airline that it is not named in this publication.

REFERENCES

[1] Federal Aviation Administration, *Flight operational quality assurance. (DOT Advisory Circular No. 120-82)*. Washington, DC: U.S. Government Printing Office, 2004.

[2] National Institute of Standards and Technology. (2010, Jun.) NIST/SEMATECH e-handbook of statistical methods. [Online]. Available: http://www.itl.nist.gov/div898/handbook/index.htm

[3] A. N. Srivastava, "Greener Aviation with Virtual Sensors: A Case Study," *Data Mining and Knowledge Discovery*, pp. 1–29, Oct. 2011.

[4] D. L. Iverson, "Inductive system health monitoring," in *Proceedings of The 2004 International Conference on Artificial Intelligence (IC-AI04)*. Las Vegas, Nevada: CSREA Press, Jun. 2004.

[5] S. D. Bay and M. Schwabacher, "Mining Distance-Based Outliers in Near Linear Time with Randomization and a Simple Pruning Rule," in *KDD '03: Proceedings of The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY: ACM Press, 2003, pp. 29–38.

[6] L. Li, M. Gariel, R. Hansman, and R. Palacios, "Anomaly detection in onboard-recorded flight data using cluster analysis," in *Proc. 30th IEEE/AIAA Digital Avionics Systems Conference (DASC)*, Oct. 2011, pp. 4A4–1 – 4A4–11.

[7] E. Chu, D. Gorinevsky, and S. Boyd, "Detecting aircraft performance anomalies from cruise flight data," in *Proc. AIAA Infotech@Aerospace*, Atlanta, GA, 2010.

[8] D. Gorinevsky, "Distributed data mining for aircraft health management," Mitek Analytics LLC, Palo Alto, CA, Tech. Rep. NASA contract NNX11CD04P, Sep. 2011.

[9] H. Goldstein, *Multilevel Statistical Models*. London: Institute of Education, 1999.

[10] J. de Leeuw and E. Meijer, Eds., *Handbook of Multilevel Analysis*, 1st ed. Springer, 2007.

[11] E. Chu, D. Gorinevsky, and S. Boyd, "Scalable statistical monitoring of fleet data," in *World IFAC Congress*, Milano, Italy, 2011.

[12] N. H. McClamroch, *Steady Aircraft Flight and Performance*. Princeton, NJ: Princeton University Press, 2011.