

**T.R.**  
**GEBZE TECHNICAL UNIVERSITY**  
**GRADUATE SCHOOL**

**IDENTIFYING ANOMALIES IN FDR DATA**  
**THROUGH REGRESSION ALGORITHMS**

**MUSTAFA BURAK AKINCILAR**

**A THESIS OF MASTER OF SCIENCE**  
**DEPARTMENT OF ELECTRONICS ENGINEERING**

**ADVISOR: ASSOC. PROF. DR. SERDAR SÜER ERDEM**

**FEBRUARY 2025**

**T.R.**  
**GEBZE TECHNICAL UNIVERSITY**  
**GRADUATE SCHOOL**

**IDENTIFYING ANOMALIES IN FDR DATA**  
**THROUGH REGRESSION ALGORITHMS**

**MUSTAFA BURAK AKINCILAR**

**A THESIS OF MASTER OF SCIENCE**  
**DEPARTMENT OF ELECTRONICS ENGINEERING**

**ADVISOR: ASSOC. PROF. DR. SERDAR SÜER ERDEM**

**FEBRUARY 2025**

**T.C.**  
**GEBZE TEKNİK ÜNİVERSİTESİ**  
**LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

**FDR VERİLERİNDEKİ ANOMALİLERİN REGRESYON**  
**ALGORİTMALARI ARACILIĞIYLA SAPTANMASI**

**MUSTAFA BURAK AKINCILAR**

**YÜKSEK LİSANS TEZİ**  
**ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI**

**DANIŞMANI: DOÇ. DR. SERDAR SÜER ERDEM**

**ŞUBAT 2025**



## **MASTER of SCIENCE JURY APPROVAL FORM**

A thesis submitted by Mustafa Burak AKINCILAR, defended on 20.02.2025 before the jury formed with the 27.01.2025 date and 2025/07 numbered decision of the GTU Graduate Administration Board, has been accepted as a Master of Science thesis in the Department of Electronics Engineering.

### **JURY**

MEMBER

(THESIS ADVISOR) : Assoc. Prof. Dr. Serdar Sür ERDEM

MEMBER : Prof. Dr. Ali TANGEL

MEMBER : Assist. Prof. Dr. İhsan ÇİÇEK

### **APPROVAL**

Gebze Technical University Graduate Administration Board  
...../...../..... date and ...../..... numbered decision.

SIGNATURE/SEAL

## **ABSTRACT**

A new approach has been developed to enhance Flight Operations Quality Assurance (FOQA) in identifying anomalous flights by applying regression analysis techniques. Most of the current methods of FOQA involve the use of thresholds predefined to identify abnormalities. However, this new method adopts the modeling of flight data patterns and detection of deviation from these models, thus allowing a more dynamic and adaptive anomaly detection system.

It fundamentally turns multivariate time series data of various flight parameters into predictive models for each individual parameter. Anomalies will be triggered if the real flight data shows significant deviations from the predicted values that these models establish. The efficacy of this advanced approach was tested using a representative Digital Flight Data Recorder dataset from NASA, considering take-off phases of datasets. The regression-based approach was really effective, as indeed a number of instances of abnormal behavior were identified, including abnormal pitch excursions, unusual flap settings, and high wind conditions. These findings give evidence that regression analysis may be a powerful tool for detecting anomalies in onboard recorded flight data.

The regression analysis based on this first assessment might significantly improve aviation monitoring and safety, hence more precise and reliable quality assurance in flight operations. By transitioning from rigid, threshold-based systems to more flexible, data-driven models, the aviation industry is able to achieve superior levels of operational safety and efficiency.

**Keywords: Regression Analysis, Flight Operations Quality Assurance (FOQA), Flight Data Recorder (FDR), Time Series Analysis, Aviation Safety, Outlier Detection**

## ÖZET

Uçuş Operasyonları Kalite Güvencesini (FOQA) geliştirmek için Regresyon analizi teknikleri uygulanarak anormal uçuşların belirlenmesinde yeni bir yaklaşım geliştirilmiştir. Mevcut FOQA yöntemlerinin çoğu, anormallikleri tanımlamak için önceden tanımlanmış eşik değerlerin kullanılmasını içerir. Ancak bu yeni yöntem, uçuş veri modellerinin modellenmesini ve bu modellerden sapmaların tespit edilmesini sağlayarak daha dinamik ve uyarlanabilir bir anormallik tespit sistemine olanak sağlamaktadır.

Temel olarak çeşitli uçuş parametrelerinin değişkenli zaman serisi verilerini her bir parametre için modellere dönüştürür. Gerçek uçuş verileri bu modellerin oluşturduğu tahmini değerlerden önemli sapmalar gösterirse anormallik belirlenir. Bu tezde, NASA'dan alınan temsili bir Dijital Uçuş Veri Kaydedici veri kümesi kullanılarak, veri kümelerinin kalkış aşamaları dikkate alınarak testler yapılmıştır. Regresyona dayalı yaklaşım gerçekten de etkili olmuş ve anormal yunuslama sapmaları, olağandışı flap ayarları ve yüksek rüzgar koşulları da dahil olmak üzere bir dizi anormal davranış örneği tespit edilmiştir. Bu bulgular, regresyon analizinin uçakta kaydedilen uçuş verilerindeki anormallikleri tespit etmek için güçlü bir araç olabileceğine dair kanıtlar sunmaktadır.

Bu ilk değerlendirmeye dayanan regresyon analizi, havacılık izleme ve güvenliğini artırabilir, dolayısıyla bu, uçuş operasyonlarında daha kesin ve güvenilir kalite güvencesi anlamına gelir. Havacılık endüstrisi daha esnek, veri odaklı modellere geçiş yaparak daha üstün operasyonel güvenlik ve verimlilik seviyelerine ulaşabilir.

**Anahtar Kelimeler: Regresyon Analizi, Uçuş Operasyonları Kalite Güvencesi (FOQA), Uçuş Veri Kaydedici (FDR), Zaman Serisi Analizi, Havacılık Güvenliği, Anomali Tespiti**

## **ACKNOWLEDGEMENTS**

I would like to express my gratitude to Assoc. Prof. Dr. Serdar Süer ERDEM. His classes and assignments were highly instructive. He placed considerable trust in me and set realistic goals and deadlines. I am indebted to him for his patience and encouragement.

I would like to express my thanks to the members of the thesis committee for their time and effort in reviewing my work.

Finally, I would like to give my deepest gratitude to my beloved wife Op. Dr. Gizem YAYLA AKINCILAR, whose love, support, and guidance have been my greatest source of strength, both in life and throughout this thesis journey.

# TABLE OF CONTENTS

|                                                                      | <b><u>Page</u></b> |
|----------------------------------------------------------------------|--------------------|
| ABSTRACT                                                             | v                  |
| ÖZET                                                                 | vi                 |
| ACKNOWLEDGEMENTS                                                     | vii                |
| TABLE OF CONTENTS                                                    | viii               |
| LIST OF SYMBOLS AND ABBREVIATIONS                                    | x                  |
| LIST OF FIGURES                                                      | xi                 |
| LIST OF TABLES                                                       | xiii               |
| <br>                                                                 |                    |
| 1. INTRODUCTION                                                      | 1                  |
| 2. BACKGROUND                                                        | 2                  |
| 2.1 Flight Data Recorder                                             | 2                  |
| 2.1.1 Evolution of FDR Data                                          | 3                  |
| 2.1.2 How FDR Works                                                  | 4                  |
| 2.1.3 Flight Data Monitoring                                         | 4                  |
| 2.1.4 FOQA Program                                                   | 4                  |
| 2.1.4.1 Overview                                                     | 4                  |
| 2.1.4.2 Data Recording                                               | 5                  |
| 2.1.4.3 Analysis Techniques                                          | 5                  |
| 2.2 Aircraft Flight Data Acquisition                                 | 7                  |
| 2.3 Data Analysis Tools                                              | 8                  |
| 2.4 Anomaly Detection                                                | 11                 |
| 2.4.1 Anomaly Detection Categorization                               | 11                 |
| 2.4.1.1 Supervised Anomaly Detection                                 | 11                 |
| 2.4.1.2 Semi-supervised Anomaly Detection                            | 12                 |
| 2.4.1.3 Unsupervised Anomaly Detection                               | 12                 |
| 2.4.2 Anomaly Detection in This Thesis                               | 12                 |
| 2.4.3 Anomaly Detection Approaches in Literature                     | 13                 |
| 2.4.3.1 Classification Approach                                      | 14                 |
| 2.4.3.2 Statistical Approach                                         | 15                 |
| 2.4.3.3 Clustering Approach                                          | 15                 |
| 2.5 Anomaly Detection Methods for Time Series                        | 16                 |
| 2.6 Aviation Anomaly Detection                                       | 18                 |
| 3. RELATED WORK                                                      | 19                 |
| 4. FDR DATA PRE-PROCESSING                                           | 21                 |
| 4.1 FDR Data Anomaly Terminology                                     | 21                 |
| 4.2 Pattern-based Anomalies and Instantaneous Anomalies              | 22                 |
| 4.3 Flight Data Acquisition for This Study                           | 23                 |
| 4.4 Parameter Selection                                              | 24                 |
| 4.4.1 Sampling Rates Synchronization                                 | 27                 |
| 4.4.2 Standardization                                                | 28                 |
| 4.4.3 Correlation Analysis                                           | 30                 |
| 4.5 Align Variables According to Take-off & Touchdown                | 32                 |
| 4.6 Flight Data Pre-Processing Experiment Result                     | 38                 |
| 4.6.1 Common Pattern Analysis                                        | 38                 |
| 4.6.2 Analysis of New Flight Data Using Common Pattern Analysis      | 42                 |
| 4.6.3 Common Pattern Analysis Experimental Results of Flight 3210525 | 44                 |

|                                                                         |    |
|-------------------------------------------------------------------------|----|
| 5. REGRESSION BASED ANOMALY DETECTION                                   | 48 |
| 5.1 Difficulties in FDR Outlier Detection                               | 48 |
| 5.2 Understanding the Importance of Stationarity in Time Series         | 48 |
| 5.2.1 Methodological Perspective on Using Differencing for Stationarity | 50 |
| 5.2.1.1 Time Series Model Assumptions                                   | 52 |
| 5.2.1.2 Simplified Modeling and Forecasting                             | 52 |
| 5.2.1.3 Interpretability of Trends and Patterns                         | 53 |
| 5.2.1.4 Improved diagnostic checks                                      | 53 |
| 5.2.1.5 Improved Model Performance                                      | 53 |
| 5.3 Converting Non-stationary Data to Stationary                        | 53 |
| 5.4 Experimental Results                                                | 56 |
| 5.4.1 High speed Abnormality (Flight Number: 653200103281828)           | 56 |
| 5.4.2 High Wind Take-off Abnormality (Flight Number: 687200108021021)   | 58 |
| 5.4.3 Sensor Fault Abnormality (653200103210525)                        | 61 |
| 5.5 Summary                                                             | 63 |
| 6. CONCLUSIONS AND FUTURE WORK                                          | 64 |
| 6.1 Conclusions                                                         | 64 |
| 6.2 Recommendations for Future Work                                     | 64 |
| REFERENCES                                                              | 66 |
| BIOGRAPHY                                                               | 72 |

## LIST OF SYMBOLS AND ABBREVIATIONS

|       |                                                 |
|-------|-------------------------------------------------|
| ALT   | : Altitude                                      |
| AR    | : Autoregressive                                |
| ARIMA | : Autoregressive integrated moving average      |
| ARMA  | : Autoregressive moving average                 |
| CAS   | : Computed Air Speed                            |
| CASM  | : Maximum Allowable Airspeed                    |
| DA    | : Drift Angle                                   |
| DFDR  | : Digital Flight Data Recording                 |
| DFDRS | : Digital Flight Data Recording System          |
| DMU   | : Data Management Unit                          |
| EGT   | : Exhaust Gas Temperature                       |
| FAA   | : Federal Aviation Administration               |
| FDAMS | : Flight Data and Management Systems            |
| FDAU  | : Flight Data Acquisition Unit                  |
| FDIMU | : Flight Data Interface Management Unit         |
| FDM   | : Flight Data Monitoring                        |
| FDR   | : Flight Data Recorder                          |
| FF    | : Fuel Flow                                     |
| FOQA  | : Flight Operations Quality Assurance           |
| GDRAS | : Ground Data Replay and Analysis System        |
| Hz    | : Hertz                                         |
| IMS   | : Inductive Monitoring System                   |
| IVV   | : Inertial Vertical Velocity                    |
| k-NN  | : K-Nearest Neighbor                            |
| LCS   | : Longest Common Subsequence                    |
| LGDN  | : Landing Gear Down Lock                        |
| LGUP  | : Landing Gear Up Lock                          |
| LREC  | : Local recurrence rates                        |
| MKAD  | : Multiple Kernel Learning                      |
| NASA  | : National Aeronautics and Space Administration |
| PCC   | : Pearson correlation coefficients              |
| PS    | : Static Pressure                               |
| QAR   | : Quick Access Recorder                         |
| TAS   | : True Air Speed                                |
| US    | : United States                                 |
| VSPs  | : Voluntary Safety Programs                     |

## LIST OF FIGURES

|                                                                                                                                                            | <u>Page</u> |
|------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|
| <b>Figure 2.1:</b> A readout from the Flight Data Recorder (FDR) on Lion Air flight 610, a 737 Max which crashed shortly after departure from Jakarta [5]. | 2           |
| <b>Figure 2.2:</b> Example of a distribution evaluation [13].                                                                                              | 7           |
| <b>Figure 2.3:</b> Flight Data Monitoring Process [14]                                                                                                     | 7           |
| <b>Figure 2.4:</b> Boeing Flight Data Analytics – Insight Accelerator [18]                                                                                 | 9           |
| <b>Figure 2.5:</b> Flight Parameters for Exceedance Detection [20]                                                                                         | 10          |
| <b>Figure 2.6:</b> Statistical Reporting Analysis [20]                                                                                                     | 10          |
| <b>Figure 2.7:</b> Outlier Detection Methods [26].                                                                                                         | 14          |
| <b>Figure 2.8:</b> Categorization of Anomaly Detection Methods for Time Series[41]                                                                         | 17          |
| <b>Figure 4.1:</b> FDR Data Process Workflow for Anomaly Detection                                                                                         | 21          |
| <b>Figure 4.2:</b> Pattern-Based Anomaly Example                                                                                                           | 22          |
| <b>Figure 4.3:</b> Instantaneous Anomaly Example                                                                                                           | 23          |
| <b>Figure 4.4:</b> Examples of parameters that include outliers and noise.                                                                                 | 24          |
| <b>Figure 4.5:</b> Examples of parameters that have different value ranges.                                                                                | 24          |
| <b>Figure 4.6:</b> NASA’s FDR Data Example in MATLAB                                                                                                       | 25          |
| <b>Figure 4.7:</b> Different sampling rates of BLAC and ELEV.1                                                                                             | 26          |
| <b>Figure 4.8:</b> The examples of selective parameters of different FDR data.                                                                             | 26          |
| <b>Figure 4.9:</b> Original ALT and PSA parameters                                                                                                         | 27          |
| <b>Figure 4.10:</b> Synchronized ALT and PSA parameters                                                                                                    | 28          |
| <b>Figure 4.11:</b> Original ALT (Altitude) and PT (Total Pressure) time series comparison                                                                 | 29          |
| <b>Figure 4.12:</b> Standardized ALT (Altitude) and PT (Total Pressure) time series comparison                                                             | 30          |
| <b>Figure 4.13:</b> Time series data of correlated parameters.                                                                                             | 32          |
| <b>Figure 4.14:</b> Flight A LGDN and ALT parameter                                                                                                        | 33          |
| <b>Figure 4.15:</b> Flight B LGDN and ALT parameters                                                                                                       | 34          |
| <b>Figure 4.16:</b> Standard Landing Gear Retraction Cycle                                                                                                 | 34          |
| <b>Figure 4.17:</b> Non-Aligned Altitude parameter according to Landing Gear Down Lock parameter                                                           | 36          |
| <b>Figure 4.18:</b> Aligned Altitude parameter according to Landing Gear Down Lock parameter.                                                              | 36          |
| <b>Figure 4.19:</b> Non-Aligned Altitude parameters according to Landing Gear Down Lock parameter in The Analyze Group.                                    | 37          |
| <b>Figure 4.20:</b> Aligned Altitude parameters according to Landing Gear Down Lock parameter in The Analyze Group.                                        | 37          |
| <b>Figure 4.21:</b> ALT: Pressure Altitude Data                                                                                                            | 39          |
| <b>Figure 4.22:</b> TAS: True Airspeed Data                                                                                                                | 39          |
| <b>Figure 4.23:</b> DA: Drift Angle Data                                                                                                                   | 40          |
| <b>Figure 4.24:</b> EGT: Exhaust Gas Temperature                                                                                                           | 40          |
| <b>Figure 4.25:</b> CAS: Computed Airspeed Data                                                                                                            | 41          |
| <b>Figure 4.26:</b> CASM: Max Allowable Airspeed                                                                                                           | 41          |
| <b>Figure 4.27:</b> PS: Static Pressure Data                                                                                                               | 42          |
| <b>Figure 4.28:</b> Example of Flight to Review in ALT parameter.                                                                                          | 43          |
| <b>Figure 4.29:</b> Integrated graph representing the combination of the two datasets.                                                                     | 44          |

|                                                                                                                                                                             |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figure 4.30:</b> Flight 3210525 ALT parameter in Common Pattern                                                                                                          | 45 |
| <b>Figure 4.31:</b> Flight 3210525 TAS parameter in Common Pattern                                                                                                          | 45 |
| <b>Figure 4.32:</b> Flight 3210525 DA parameter in Common Pattern                                                                                                           | 46 |
| <b>Figure 4.33:</b> Flight 3210525 CAS parameter in Common Pattern                                                                                                          | 46 |
| <b>Figure 4.34:</b> Flight 3210525 CASM parameter in Common Pattern                                                                                                         | 47 |
| <b>Figure 4.35:</b> Flight 3210525 PS parameter in Common Pattern                                                                                                           | 47 |
| <b>Figure 5.1:</b> Categorization diagram for some of the most researched transformation methods for coping with non-stationarity in univariate time series prediction [69] | 50 |
| <b>Figure 5.2:</b> Non-Stationary and Stationary ALT parameters of Analyze Group                                                                                            | 54 |
| <b>Figure 5.3:</b> Pattern Anomaly Example                                                                                                                                  | 54 |
| <b>Figure 5.4:</b> Static Pressure Anomaly in Real Data                                                                                                                     | 55 |
| <b>Figure 5.5:</b> Instantaneous Anomaly Example                                                                                                                            | 56 |
| <b>Figure 5.6:</b> Non-Stationary Computed Airspeed and Pitch Angle plot of the Flight Number: 653200103281828                                                              | 57 |
| <b>Figure 5.7:</b> Stationary Common pattern and Reviewed Flight plot of the CAS                                                                                            | 57 |
| <b>Figure 5.8:</b> Angle of Attack Plot (Non-Stationary, Stationary Common Pattern, Stationary Reviewed Flight)                                                             | 58 |
| <b>Figure 5.9:</b> Non-Stationary Computed Airspeed of the Flight Number: 687200108021021                                                                                   | 59 |
| <b>Figure 5.10:</b> Stationary Computed Airspeed of the Flight Number: 687200108021021                                                                                      | 59 |
| <b>Figure 5.11:</b> Non-Stationary Rudder Position of the Flight Number: 687200108021021                                                                                    | 60 |
| <b>Figure 5.12:</b> Non-Stationary Wind Direction and Speed of the Flight Number: 687200108021021                                                                           | 61 |
| <b>Figure 5.13:</b> Non-Stationary and Stationary Static Pressure plot of the Flight Number: 653200103210525                                                                | 62 |

## LIST OF TABLES

|                                                                                                                                                                                    | <b><u>Page</u></b> |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|
| <b>Table 2.1:</b> Evolution of FDR Data                                                                                                                                            | 3                  |
| <b>Table 2.2:</b> Example of exceedance event parameters [12]                                                                                                                      | 6                  |
| <b>Table 4.1:</b> Description of parameter in Figure 4.13                                                                                                                          | 32                 |
| <b>Table 5.1:</b> The nonstationary time series transformation methods selected for the experiment are outlined, along with the data transformation processes they are applied to. | 51                 |

# 1. INTRODUCTION

Improvements in cockpit procedures and technologies have been utilized safety in the airline industry since the 1970s. Initially, safety improvements were predominantly driven by accident investigations, a reactive approach to safety that dates back to the early 1930s. The early safety strategy, summarized by the mantra "fly it, crash it, redesign it, fly it, crash it" [1], has since evolved to incorporate operational investigations and proactive safety measures. This evolution is exemplified by the implementation of Voluntary Safety Programs (VSPs), particularly in the United States. Among these VSPs, the most widely adopted program utilizing quantitative data is Flight Operational Quality Assurance (FOQA). Numerous airlines globally have also adopted FOQA or similar programs to enhance their safety protocols [2].

There is an increasing emphasis on the implementation of proactive safety management systems in order to enhance the high standards of safety in airline operations. These are systems designed to identify and lighten risks before they result in accidents or incidents. [3]. One approach is to analyze operational data that has been archived in modern digital flight data recorders (DFDRs) through flight operations quality assurance (FOQA) programs. The traditional FOQA methods employ predefined criteria to identify anomalies, which may result in the omission of emerging operational or safety issues. This thesis presents a method that employs regression analysis to detect anomalies without the use of predefined criteria. So, this method is offering a more flexible and comprehensive approach.

## 2. BACKGROUND

### 2.1 Flight Data Recorder

The FDR (Flight Data Recorder) system, popularly known as the black box, records hundreds of data during the flight of the aircraft. FDR consists of hundreds of valuable data such as altitude, speed, direction, wing angle, fuel flow depending on time. This data is critical to understand the performance of the aircraft, particularly in the event of an accident. [4]

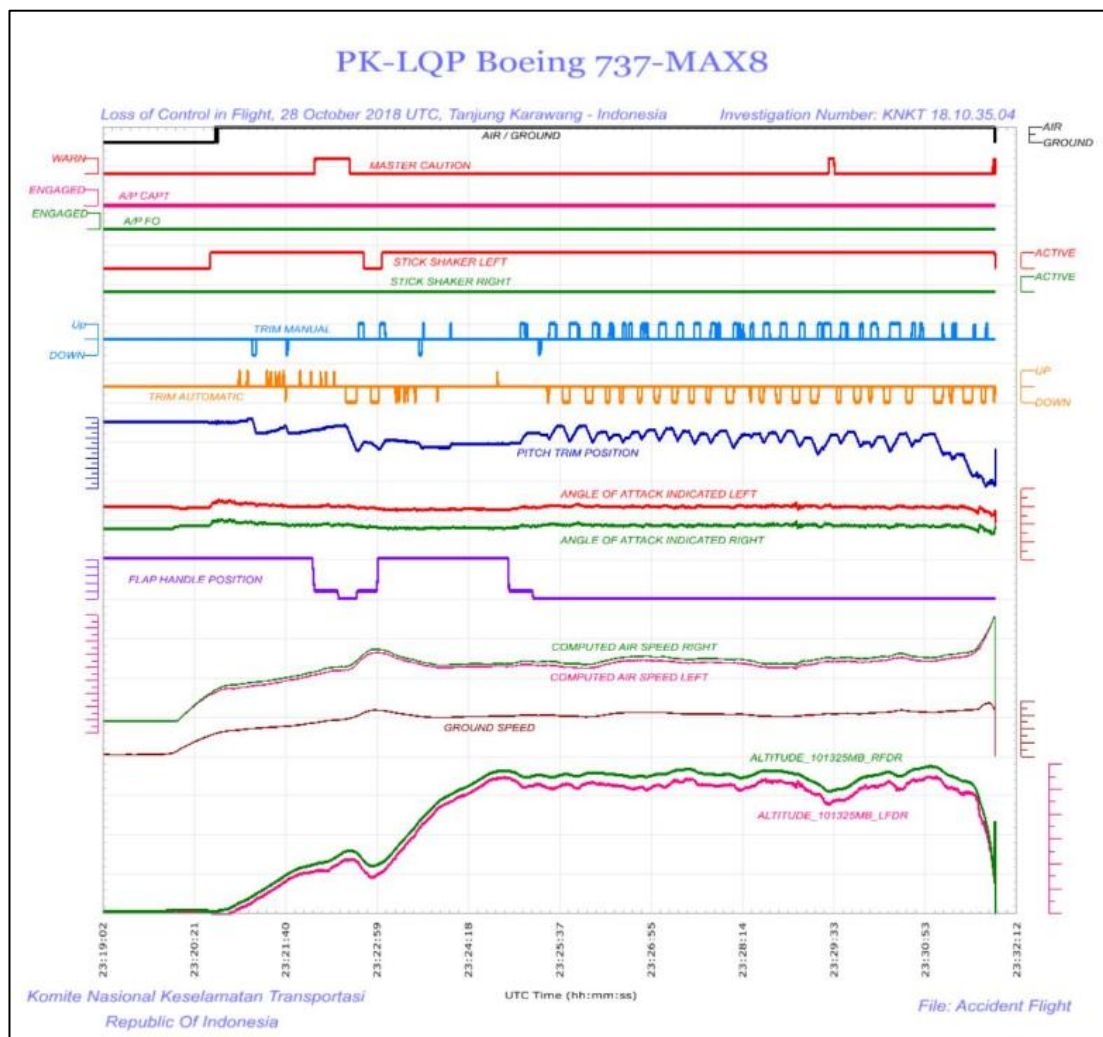


**Figure 2.1:** A readout from the Flight Data Recorder (FDR) on Lion Air flight 610, a 737 Max which crashed shortly after departure from Jakarta [5].

In the past, it was only used for investigation of the accident. Flight Data Recorder records on-board flight parameters for an entire flight [6]. Generally, it contains accelerations, altitude, airspeed, calculated airspeed, static pressure etc. **Figure 2.1** shows the flight parameters recorded during Lion Air Flight 610, before impact in 2018 [5].

### 2.1.1 Evolution of FDR Data

The recording capacity of FDR is considerably increased as shown in **Table 2.1**. This parameter recorded and sampling rates can be varied by the airline requirements. FAA specifies requirements of parameters and sampling rates for US airlines [7]. The FDR can record around 2000 parameters and 50 hours [8]. Sampling rates vary in accordance with the specific nature of the flight parameter under consideration. To illustrate, the evolution of these data collection devices is shown by using the following aircraft types as examples [9] [10]:

**Table 2.1:** Evolution of FDR Data

| Aircraft Type | Year | Data Type                                | Parameters Number | Parameter Capacity                |
|---------------|------|------------------------------------------|-------------------|-----------------------------------|
| B707          | 1958 | Analogue                                 | 5                 | Mechanical limit of ~10 parameter |
| A330          | 1993 | Digital                                  | 280               | 128 wps <sup>1</sup>              |
| Embraer 170   | 2004 | Digital combi-recorder <sup>2</sup>      | 774               | 256 wps                           |
| A380          | 2007 | Digital (solid-state)                    | >1000             | 1024 wps                          |
| B787          | 2009 | Digital (solid-state EAFR <sup>3</sup> ) | >1000             | Ethernet System                   |

---

<sup>1</sup> wps: words per second. An FDR word consists of 12 bits.

<sup>2</sup> Both CVR (Cockpit) and FDR (Flight Data) data is recorder in combi-recorder.

<sup>3</sup> EAFR: Enhanced Airborne Flight Recorder. The EAFR also has the capability of storing video information

The digitalization of FDR data has facilitated easier access to this information. To enable FDR data to be used before an accident, FDR data can be acquired at pre-set intervals during daily operation [9].

### **2.1.2 How FDR Works**

The FDR continuously records data during flight. Modern FDRs can store up to 25 hours of flight data, capturing parameters at specified intervals (e.g., every second). The recorded data is stored in a crash-protected memory unit designed to withstand extreme conditions. After a flight, the recorded data can be downloaded for analysis. In the case of an incident, investigators retrieve the FDR data to reconstruct the flight and understand the sequence of events leading to the occurrence. At last, the recorded data is analyzed to monitor aircraft performance, identify trends, and detect anomalies. This analysis supports Flight Operations Quality Assurance (FOQA) programs aimed at enhancing safety and operational efficiency.

### **2.1.3 Flight Data Monitoring**

Most airlines are routinely collecting and analyzing in-flight data, a process generally known as Flight Data Monitoring (FDM). The FDM process includes processes such as data collection, transmission, storage and analysis. In order to establish a measure for the present methods of flight data analysis, different software for flight data analysis were examined. This research was motivated by the identification of potential improvement opportunities from the literature review.

### **2.1.4 FOQA Program**

#### **2.1.4.1 Overview**

Flight operational quality assurance (FOQA) is a method of capturing, analyzing and/or visualizing the data generated by an aircraft moving from one point to another. The information learned from this analysis can be used to identify new ways to improve flight safety and increase overall operational efficiency. Several airlines and various national air forces have initiated FOQA programs to collect, store and analyze

recorded flight data. The objective of this initiative is to enhance aviation safety, enhance maintenance effectiveness and reduce operational costs [11].

The FOQA analysis process must be developed based on the objective and scope of the intended program. It will focus on evaluating or influencing changes in areas such as operational safety, aircraft performance, aircraft system performance, crew performance, and company procedures [12].

#### **2.1.4.2 Data Recording**

The FOQA analysis is influenced by the available parameters, their sample rates, and recording accuracies. Aircraft with programmable FDAUs or DMUs can exceed minimum parameter requirements and adjust sample rates through reprogramming, depending on storage capacity. Coordination with engineering personnel is essential to identify these parameters [12].

#### **2.1.4.3 Analysis Techniques**

In FOQA data, two types of analysis techniques are used. They are parameter exceedance analysis and statistical analysis.

The Exceedance analysis is an approach detects predetermined operationally unacceptable events. This analysis controls whether certain parameters are out of limits according to predefined values [12]. Safety experts must determine this list of flight parameters and the limiting values of these variables in advance. Typically, the observation pattern is consistent with the airline's standard procedures, including, but not limited to, takeoff pitch, climb rate, and flap retraction time [7]. **Table 2.2** provides some examples [12].

**Table 2.2:** Example of exceedance event parameters [12]

| Event                             | Parameters                                        | Basic Event Definition                               | Description                                                                                                  |
|-----------------------------------|---------------------------------------------------|------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|
| Takeoff<br>Climb<br>Speed<br>High | CAS, Gross<br>Weight, HAT                         | HAT > x feet,<br>HAA < x feet,<br>CAS > V2 + x knots | To detect climb speed higher<br>than desired during the<br>Takeoff Phase                                     |
| Flap Limit<br>Altitude            | Altitude,<br>Flaps<br>Position                    | Altitude > x feet, Flap<br>Position > 0 degrees      | An event to detect when flaps<br>are operated above the<br>maximum allowable altitude<br>for flap operation. |
| Operating<br>Ceiling<br>Exceeded  | Altitude                                          | Altitude > x feet for x<br>seconds                   | An event to detect operation<br>of the aircraft above its<br>certificated maximum<br>operating altitude.     |
| Hard<br>Landing                   | Air/Ground<br>Switch,<br>Vertical<br>Acceleration | Air/Ground = Ground,<br>Vertical Acceleration > x G  | An event that measures<br>excessive Gforce at<br>touchdown, indicating a hard<br>landing.                    |

Statistical analysis is a method that involves examining the patterns of specific flight parameters to analyze a flight operation. Unlike Exceedance Detection, Statistical Analysis provides a more comprehensive view of the activity by considering the distribution of total number of flights. **Figure 2.2** shows an example of “altitude at which landing flap is set” [13].

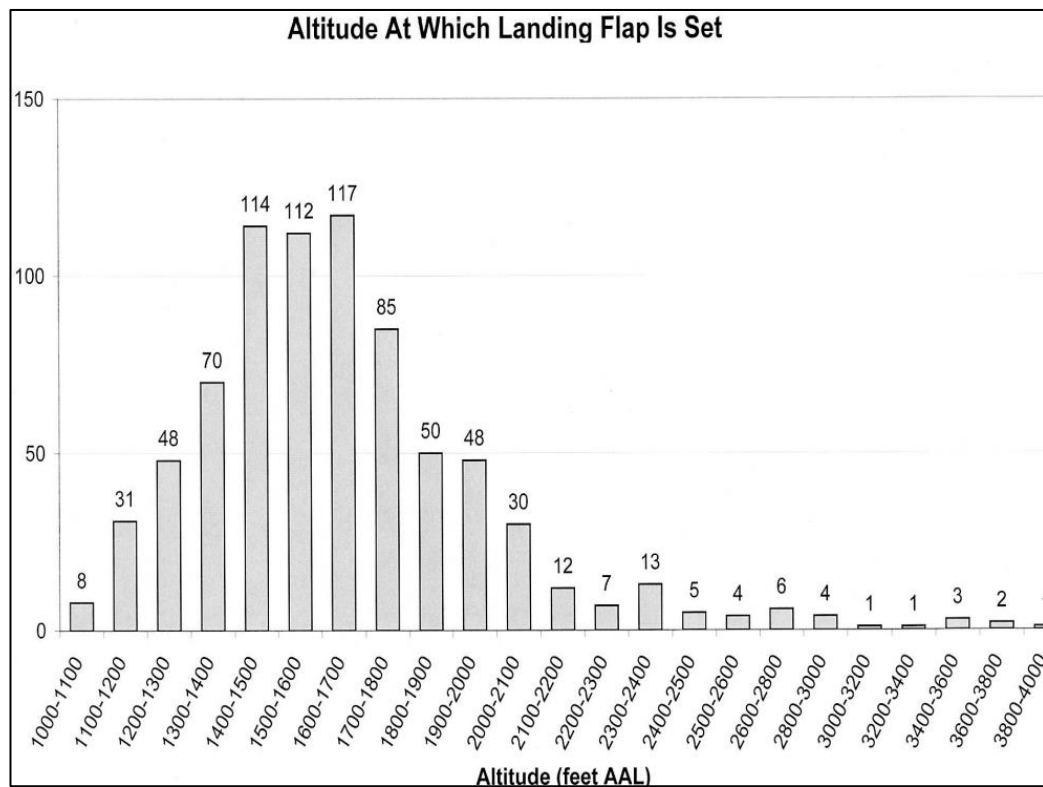


**Figure 2.2:** Example of a distribution evaluation [13].

## 2.2 Aircraft Flight Data Acquisition

The current procedure of FDM comprises several key stages such as recording, downloading, processing. These stages are visualized in **Figure 2.3**. The devices and technologies used to facilitate the individual phases are described in the next section.

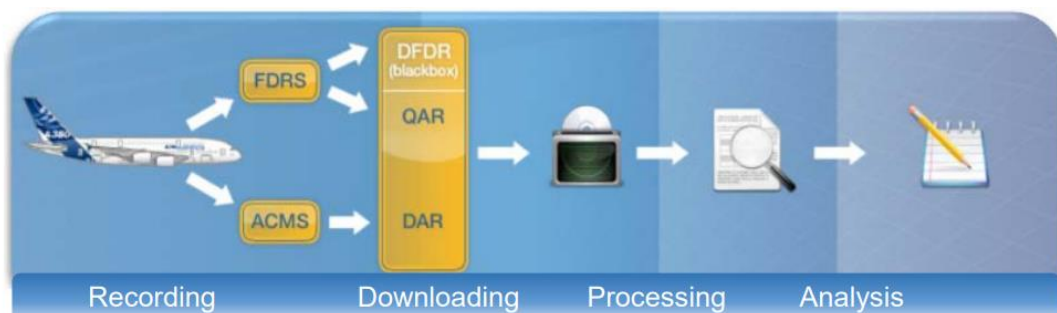


**Figure 2.3:** Flight Data Monitoring Process [14]

Data collection and recording from hundreds of sensors on an aircraft over a complete flight is, in fact, the first stage of parameters generation. Flight Data Interface

Management Unit (FDIMU) collects all these parameters and transforms into a single digital format. It is also known as Flight Data Management Systems, Flight Data Acquisition Unit, etc. [12].

The converted value is transmitted to the FDR and, if installed, also the Quick Access Recorder (QAR)<sup>4</sup>.

The next step is transferring the data: the recorded data in QAR or DFDR are downloaded periodic time and send it to stations. Data retrieval via manual or wireless is possible. The duration of downloading depends on the recording storage capacity of the recorder or the operator's requirement in analyzing flight data. Wireless transmission has the potential to reduce the timeframe of the delivery of data that could range from several days or weeks in the use of manual retrieval of data to a couple of minutes after the aircraft has landed [15].

**Data Analysis:** The process also includes analyzing flight data, generating statistical reports, and creating flight animations. Various analytical tools are used, FOQA or similar initiatives, customized to the specific objectives and scope of the operator's flight data monitoring programs. Once the analysis is complete, the flight data is stored for future reference and further analysis.

## **2.3 Data Analysis Tools**

Data analysis plays a crucial role in flight data monitoring. Specialized commercial software performs a wide range of analyses to support this process, such as GDRAS, FOQA, and other software. Most of the tools provide for Analysis, Animation, and Reporting facilities and employ techniques like Exceedance Detection, Statistical Analysis, and Trend Analysis. The objective is to enable the safety management system of the aviation industry by extracting useful information from recorded data and presenting an easily comprehensible conclusion that can assist decision-making.

Exceedance detection determines whether predefined operational issues have occurred by testing the flight parameters for exceedance against pre-set limits based on certain

---

<sup>4</sup> The QAR (Quick Access Recorder) functions like the FDR (Flight Data Recorder) but doesn't offer the same crash resistance. It's typically installed in the electronic equipment bay below the aircraft for convenient reach. The QAR stores data on formats USB Flash Memory or PCMCIA cards, enabling easy data extraction

conditions. These limits are set by safety specialists according to standard airline procedures, such as pitch upon takeoff or flap-retraction timing.

The profiles of flight, maintenance, or engineering data created by Statistical Analysis permit a broader view over the operation compared to the individual events of exceedance. Also, Trend Analysis helps airlines track their performance, by counting the amount of non-stabile approaches on a monthly basis as a measure of improvement or decline [16], [17].

Animation confirms whether the events identified are hazardous or benign. Using 3-D animation constructed from flight data, analysts review cockpit panel, instruments, and terrain to understand in-flight events [18].



**Figure 2.4:** Boeing Flight Data Analytics – Insight Accelerator [18]

The GDRAS software generates reports on excess detection, statistical analysis, and trend analysis in different formats. Normally, exceedances are grouped together before being reported, but it's also possible to view individual exceedances as shown in **Figure 2.5**. Trend analysis reports include total exceedance information by time. Statistical analysis for TAS outlier at 500 feet by fleet type is in **Figure 2.6** [19].

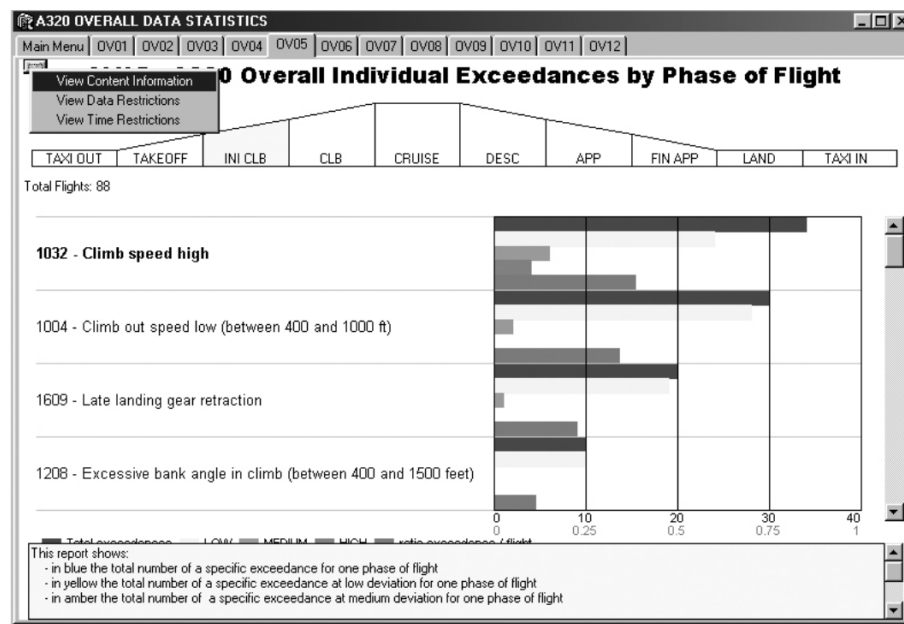


**Figure 2.5: Flight Parameters for Exceedance Detection [20]**

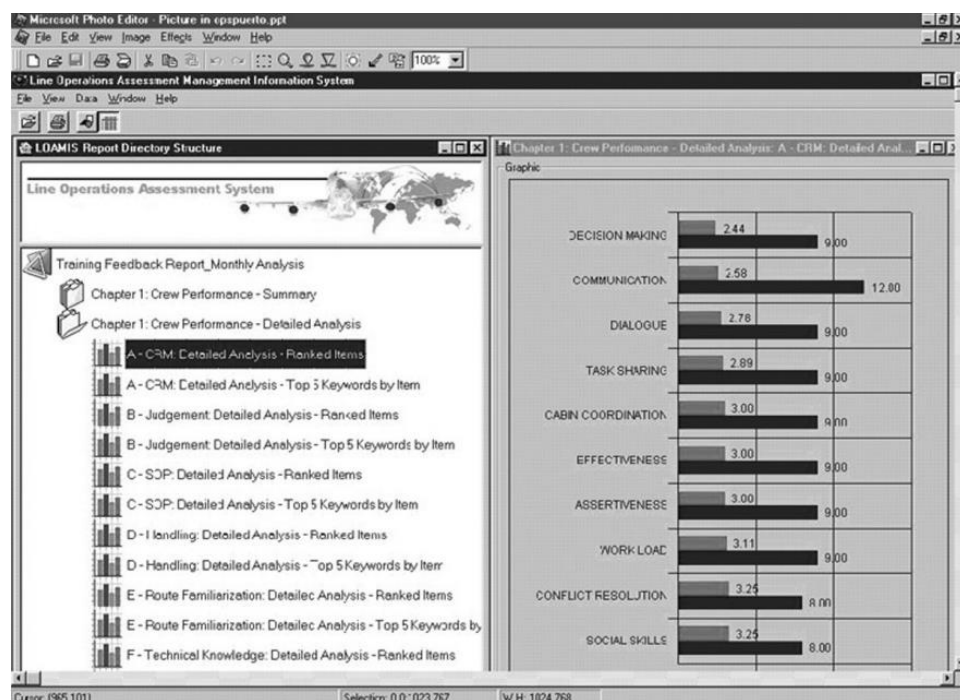


**Figure 2.6: Statistical Reporting Analysis [20]**

In summary, modern technology and advancements enable to watch flight data, analysis, anomaly detection. The collected parameters provides airlines with valuable information, can be downloaded and analyzed on the ground. And this information improves safety and efficiency. Available software tools, which relies on predefined

events and criteria for specific parameters. Information about these predefined events is assessed through exceedance detection, statistical, and trend analyses.

The two most common approaches are event-based analysis and machine learning algorithms. However, event-based analysis has a limitation: it only detects known issues and requires specifying parameter in advance. Unfortunately, many incidents are caused by unexpected risks. This research seeks to create a way to identify these unidentified events without needing to define them beforehand.

## **2.4 Anomaly Detection**

The goal is to develop effective anomaly detection algorithms that can spot flight irregularities in FDR data. Anomaly detection is about finding patterns or observations that stand out as being different from the majority of the data. It's also known by different names, like novelty, fault or deviation identification depending on the field. A lot of techniques for anomaly detection have been developed so far. Some are focused on solving specific issues in certain areas and some versatile and can be applied to a variety of problems. The following sections provide a detailed overview of anomaly detection.

### **2.4.1 Anomaly Detection Categorization**

Anomaly detection comes in many forms, depending on the specific needs and constraints of the application. These differences require the development of different techniques, as the solution often depends on how the problem is defined. In general, anomaly detection can be broken down into three key aspects: the type of input data, the level of monitoring and the type of anomaly [21].

Based on requirement of the training data, anomaly detection algorithms are typically grouped into three types: supervised, semi-supervised and unsupervised techniques.

#### **2.4.1.1 Supervised Anomaly Detection**

The supervised anomaly detection techniques construct a work scope on training data sets that incorporate normal classes and deviation classes.

#### **2.4.1.2 Semi-supervised Anomaly Detection**

This approach requires only a single labeled instance, such as a designated nominal instance or outlier instance, within the dataset. The majority of semi-supervised techniques employ a representative model of nominal behavior, constructed using the nominal instances, with the objective of detecting outliers in instances that do not align with the model.

#### **2.4.1.3 Unsupervised Anomaly Detection**

Unsupervised anomaly detection techniques do not rely on labelled historical data, but instead predict the presence of outliers in data sets by identifying deviations from predefined models or by detecting anomalies as instances of nominal behaviour. It is important to note that the concept of an anomaly varies significantly across diverse fields of application.

The variety of definitions of anomalies gives rise to a range of detection techniques. One of the most elementary types of anomaly is a data instance whose values are incongruent with normal (non-abnormal) instances in the whole data set. In multivariate problems, anomalies based on correlation are often of interest, since the relations among variables can be more significant than individual variable values. For input data with structures, it can frequently be more meaningful to evaluate the values in certain contexts. The occurrence of an anomaly is determined by its temporal and spatial characteristics, and the anomalies of interest may manifest as sudden changes in time series or location-specific deviations in spatial data. Pattern-based outliers are identified through the analysis of trends over time or space, the ordering of sequences, and the frequency of occurrence.

#### **2.4.2 Anomaly Detection in This Thesis**

This study aims to anomaly detection is to identify abnormal flights from the standard flight data with no any trained information.

The input data are characterized as multivariate time series, which are classified as continuous, categorical or discrete, as appropriate, according to the type of flight variable. The nature of the main thesis' objective, namely the unsupervised detection of abnormal flights, necessitates the utilization of unsupervised techniques. The

anomalies to be detected are of a context-based<sup>5</sup> nature, including both pattern-based and instantaneous anomalies. A contemporary aircraft is operated under a multitude of system states. The anomaly can depend on the particular mode in which the aircraft is operating at a given moment. The anomalies in this context can be classified into two types: pattern-based anomalies, where the patterns are abnormal over a certain period, and instantaneous anomalies, which are abnormal observations made at a particular moment.

### 2.4.3 Anomaly Detection Approaches in Literature

A large number of anomaly detection techniques have been formulated to solve a wide variety of problems in many application domains. Traditionally, three major approaches are considered: a classification approach, a statistical approach and a clustering approach. However, it should be noted that none of the categories are exclusive, as certain techniques incorporate concepts from more than one basic approach. The most recent review of techniques for anomaly detection in various domains is the most complete to date [22]. In this section, a review of some common techniques and models of each method will be conducted.

One of the analytical techniques utilized in this study employs data analysis techniques that do not presuppose any a priori knowledge concerning the dataset's distribution. One such strand of the research focuses on the analysis of probability distributions through the use of histograms [23], [24]. Another strand of the research employs kernel-based methodologies, which facilitate the estimation of the data density distribution through the utilization of kernel functions [25]. The various approaches employed for the identification of outliers are illustrated in **Figure 2.7**.

---

<sup>5</sup> The context-based approach posits that an occurrence may be regarded as anomalous or not, contingent on its temporal context.

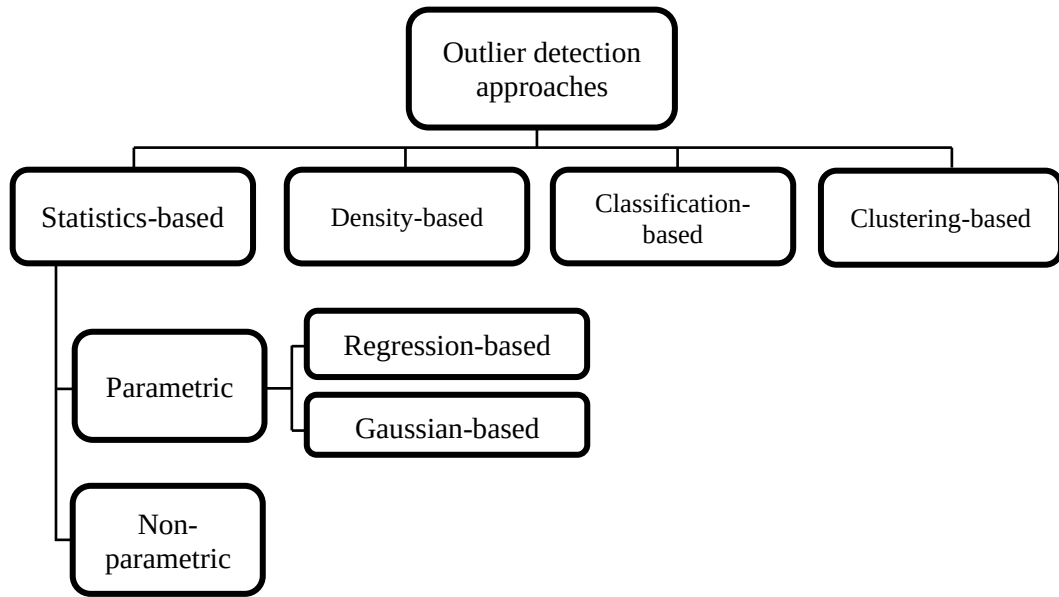


**Figure 2.7:** Outlier Detection Methods [26].

#### 2.4.3.1 Classification Approach

Classification techniques are usually used in supervised learning, where a training dataset with labeled samples—typically showing both "normal" and "abnormal" cases—is required. During training, these examples help establish boundaries between the classes. In the process of evaluation, the learned model utilizes these boundaries to categorize new observations into one of the predefined categories. Classification models for this purpose include neural networks, Bayesian networks [27], Support Vector Machines [28], decision trees [29], and regression models [30].

In many anomaly detection tasks, datasets with labeled examples of both normal and abnormal cases are unavailable. To address this, variations of traditional classification methods have been adapted for semi-supervised learning. Another methodology is one-class classification [31], which identifies one specific class of objects and treats anything outside it as an outlier. In practice, data representing the "normal" class is often used for training, and deviations are identified as parameters that do not fit within this class. In certain instances, the availability of data is limited to examples of unusual cases, and the identification of anomalies is achieved by comparing new data against a learned "anomaly dictionary" [32].

### **2.4.3.2 Statistical Approach**

Statistical methods assume or restore a statistical distribution of data during their learning stage, and an observation's outlier status is contingent on its compatibility with the model. The most widely accepted distribution in this field is arguably the Gaussian model. Maximum Likelihood Estimates are among the popular techniques that are used in the training phase for estimating the parameters of the Gaussian model. Subsequently, the testing phase involves the application of statistical tests for the purpose of verifying whether a given sample belongs to the distribution or not. The most widely used anomaly statistical tests are the Box-Plot rule [33]; the Grubbs test [34], [35] and variants of the Grubb test [36], [37].

When the distribution of data is unknown, regression is used to create models to help analyze it. In the training phase, a regression model is built; then, in the testing phase, new data is tested against this model. In simple regression, the data has two variables, making it easy to visualize on a scatterplot, where outliers can often be spotted by eye. However, as data becomes more complex with multiple variables, detecting outliers becomes much harder, so various specialized regression methods have been developed to address this challenge [26].

A statistical method is comprised of numerous techniques that are founded upon statistical principles. In order to utilize this fundamental statistical approach, a multitude of techniques are employed to transform the original multivariate data representation (e.g. univariate) with a variance that is considered to be "basic".

In this thesis, a regression method will be used to detect outliers after processing the data. Linear regression models have a major advantage: their simplicity. Because they are linear, it is easier to understand and follow the estimation steps in a structured way. The equations themselves are also easier to interpret, making linear models more accessible and understandable compared to more complex, nonlinear models [38].

### **2.4.3.3 Clustering Approach**

Clustering is a method of grouping data instances so that similar instances end up in the same group, while distinct ones are separated into different groups. This way, data is organized into meaningful subsets that reflect the overall structure of the population being analyzed. Clustering is typically represented as a set of these grouped subsets.

In anomaly detection, cluster-based methods identify outliers as instances that don't fit well into any group or belong to unusually small groups, assuming that "normal" data tends to form larger, denser clusters. There are several approaches to clustering, including partition-based, hierarchical, and proximity-based methods, each with a unique way of organizing data [39].

## **2.5 Anomaly Detection Methods for Time Series**

Anomaly detection, defined as the identification of atypical items or events within data, has garnered significant attention from researchers and practitioners and is now recognized as a fundamental task in data mining and quality assurance. The field has advanced substantially across various application domains. Traditional approaches, including linear model-based methods, distance-based techniques, density-based methods, and support vector machines, continue to serve as effective algorithms. However, as systems grow in scale and complexity, these methods encounter notable limitations, particularly in managing high-dimensional data or addressing the scarcity of labeled anomalies.

Anomaly detection within time-series data poses unique challenges, as it requires joint consideration of the sequential order and causal relationships among observations over time. Recently, novel approaches have been developed to address these complexities. For instance, innovative computational method introduced [40] utilizing a recurrence plot (RP)—a matrix indicating instances when a dynamic system returns to similar states. By measuring local recurrence rates (LREC) via a sliding window applied to the RP, this method detects anomalies by analyzing the similarity between LREC curve statistics, thereby providing a robust approach for time-series anomaly detection in complex systems [41].

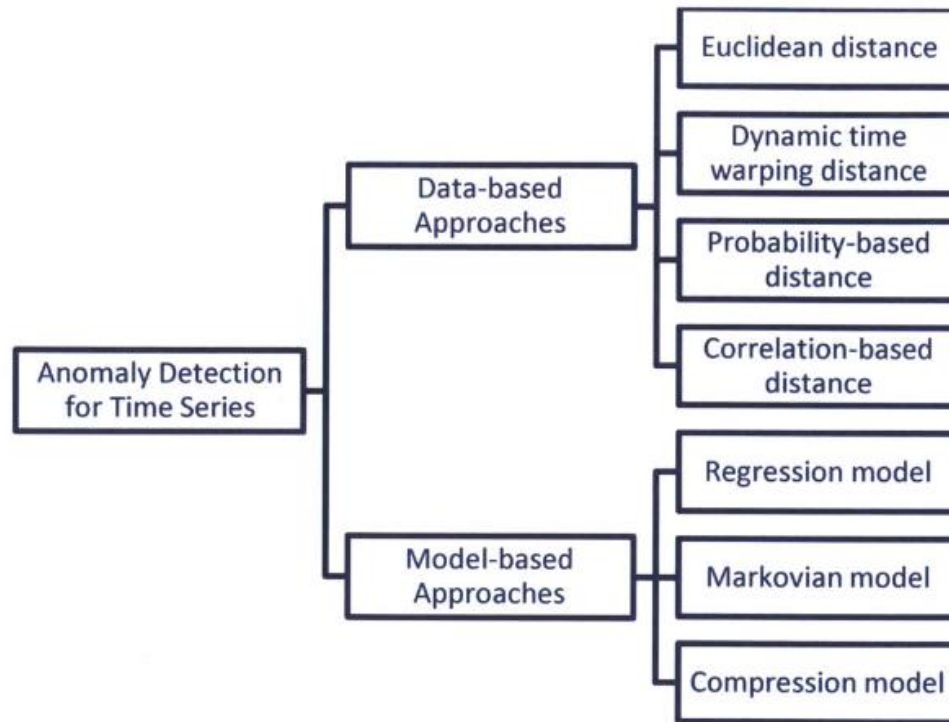


**Figure 2.8:** Categorization of Anomaly Detection Methods for Time Series [41]

Regression-based methods are among the earliest and most applied techniques for modeling time series data. In these models, the difference (or residual) between the model's predicted and the real data observed value serves as an indicator of outliers [42]. provided a detailed overview of using robust regression to build models and identify outliers. Time series were first modeled using an autoregressive (AR) process [43]. Later on, to expand this approach autoregressive moving average (ARMA) models were introduced [44]. This development eventually led to the autoregressive integrated moving average (ARIMA) model, which took things a step further by adapting ARMA to handle more complex patterns in time series data [45].

To better understand time series data and make more accurate predictions, we use different models designed to capture the data's main patterns as closely as possible. For instance, the Autoregressive Moving Average (ARMA) model assumes that the series is stationary—meaning its average value doesn't change over time. By relying on this assumption, ARMA can focus on stable relationships in the data, helping to produce reliable forecasts.

## 2.6 Aviation Anomaly Detection

Traditionally, aviation anomaly detection has targeted the identification of mechanical problems, but much recent research has focused on the monitoring of overall system performance [46], [47]. Software named the "morning report" was created. The methodology adopted by the researchers summarized time-series data with four coefficients and utilized Mahalanobis distance for the identification of anomalies.. Though revolutionary, this had its shortcomings in reliance on just four coefficients and often missed essential details. Mahalanobis' distance alone could not capture the essential structure of flight data.

The algorithm "sequenceMiner" was produced [48] in 2006, which made use of clustering and Bayesian techniques in arriving at anomalies in flight data sequences based on the Longest Common Subsequence (LCS) measure.

Inductive Monitoring System stood out because it defined normal behavior based on nominal datasets and compared real-world data to that [49]; or another anomaly detection algorithms, including support vector machine and inductive system algorithms, which they felt were generally useful on complex systems [50]. However, those algorithms did not model time-based data patterns very well. Semi-supervised approach based on kernel learning, particularly one-class SVM purposed, to model temporal data for sophisticated detection [51].

### 3. RELATED WORK

Previous methods for detecting outliers in flight data include the SequenceMiner algorithm [52], NASA’s Morning Report software, and the Inductive Monitoring System. These methods have limitations, such as not considering temporal patterns or requiring labeled training data. Regression analysis, however, offers a way to model continuous flight parameters over time. Therefore, it is making it possible to detect anomalies based on deviations from expected patterns.

The Inductive Monitoring System (IMS) represents a relatively recent methodology which is used by NASA’s Space Shuttle Columbia (STS-107) mission as a temperature analyzer [53].

IMS is designed with a particular focus on the continuous monitoring of flight parameters, while reducing the need for manual model building through the use of clustering techniques to group consistent system parameter data. In order to gain knowledge of nominal system behaviour, it collects nominal data from the system simultaneously. It uses K-means and density-based clustering techniques to group data into clusters [54].

The Multiple Kernel Learning (MKAD) and Density-Based Algorithm for anomaly detection presents a sophisticated approach utilizing kernel functions to integrate both discrete and continuous data streams [55]. This method, however, uses one-class Support Vector Machine which necessitates labeled training data and operates under the assumption that only a single class of nominal flights is employed for anomaly detection [56], [57].

K-Nearest Neighbor (k-NN) approach detects unusual flight patterns from FDR. This process involves data pre-processing, feature generation classification and anomaly detection steps [58]. The k-NN algorithm is susceptible to issues when confronted with datasets that are inherently imbalanced, where a particular class is underrepresented, potentially leading to the generation of biased predictions.

Literature shows different studies that developed methodologies using both non-stationary and stationary time series data [59], [60], [61], underlining advantages and challenges for each in several domains.

The purpose of this paper is to evaluate some of the advantages of stationary time series analysis in the context of FDR data and highlights how leveraging the intrinsic stationarity of statistical properties over time enhances the reliability and interpretability of aviation analytics. Supervised method for detection of anomalous flights used in this thesis. This method will consider temporal patterns in the data based on continuous flight parameters, as well as multiple nominal patterns.

## 4. FDR DATA PRE-PROCESSING

It is acknowledged that there may be a multitude of patterns present within operative data; however, the vast majority of these are finite in number. Given the highly standardized nature of airline operations, it is evident that a significant proportion of flights will exhibit a small number of common data patterns. Initially, it is necessary to transform the raw FDR data into high-dimensional vectors prior to analysis. The subsequent sections of this chapter will provide detailed descriptions of this method.

In this research study, following the process of data preparation as outlined by FDR, the data will be aligned according to take-off time. Subsequently, a common pattern analysis will be performed. This data will be converted to stationary to use advantage of stationary time-series data.

Finally, regression analysis on vectors approach will be applied to automatically identify the FDR data in which the unusual flight patterns are recorded from a large amount of FDR data. The proposed method as shown in **Figure 4.1** is tested with using a realistic FDR data from the NASA's open database [62].

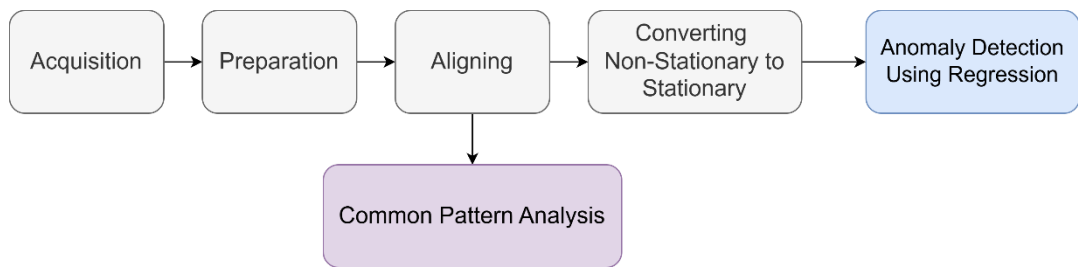


**Figure 4.1:** FDR Data Process Workflow for Anomaly Detection

### 4.1 FDR Data Anomaly Terminology

This segment introduces the fundamental principle of anomaly detection in data analytics, along with the terminologies employed to describe the algorithms involved. These include regression, outlier, anomaly, and abnormal flights.

Parameters of similar data are recognized as common patterns in the data; data not belonging to a pattern are recognized as anomalies, which point to unusual behavior of the data.

Finally, according to the result of regression analysis, anomaly detection is performed. Samples, which are outliers are the anomalies which should be find out. At last, the anomalies are summarized by flight. The term 'abnormal flights' is used to denote flights that exhibit a higher degree of anomalies, whether these are more or more severe.

## 4.2 Pattern-based Anomalies and Instantaneous Anomalies

Anomalies in Flight Data Recorder data can be categorized into two types: pattern-based anomalies and instantaneous anomalies. **Figure 4.2** and **Figure 4.3** illustrate examples of these two types of anomalies, with the anomaly signals depicted in red.

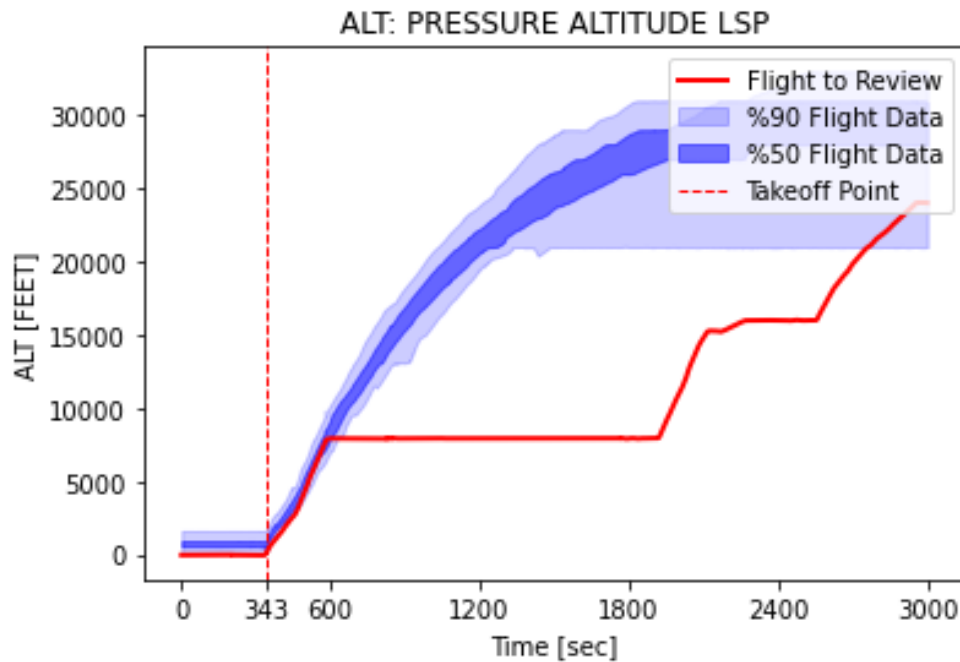


**Figure 4.2:** Pattern-Based Anomaly Example

Pattern-based anomalies are similarly defined as abnormal patterns during certain flight phases. **Figure 4.2** illustrates a pattern-based altitude anomaly, representing a phase that deviates from the common pattern.

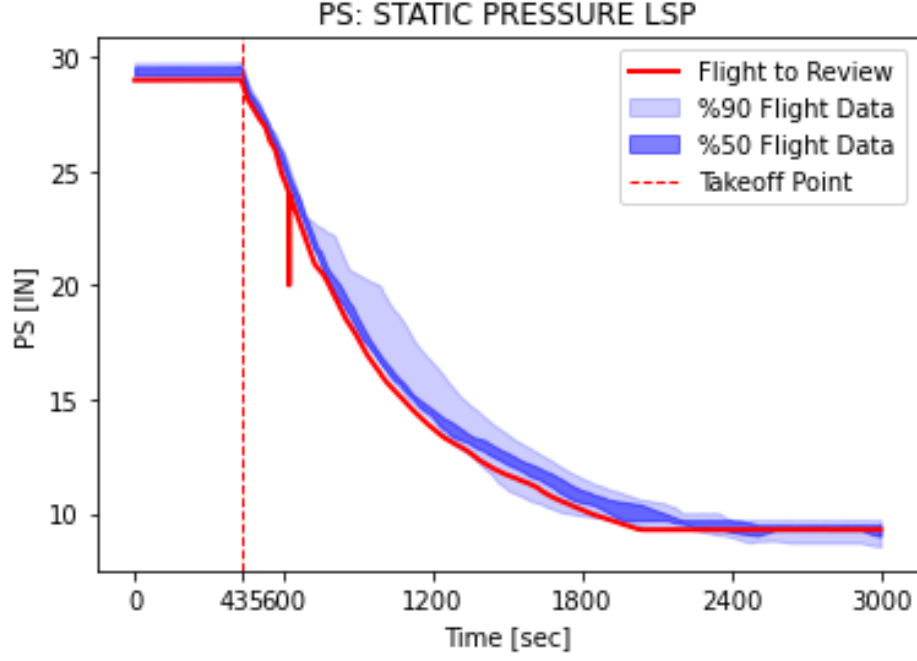


**Figure 4.3:** Instantaneous Anomaly Example

Instantaneous anomalies are abnormal data points happening in an instance. **Figure 4.3** shows the "Static Pressure" parameter anomaly as an instantaneous one.

The goal of anomaly detection encompasses the detection of both kinds of anomalies in this research. As two different data transformation techniques were used in Common Pattern Analysis Sample, it is expected that the two methods would be sensitive to different types of anomalies. Thus, this study also covers a comparison of which kind of anomalies each method detects better.

### 4.3 Flight Data Acquisition for This Study

This thesis is based on the FDR data from the open database of NASA [62]. Each FDR data set contains 186 various parameters including continuous signals, discrete signals, and warning signals. For each parameter, the information that will be extracted is: parameter name, parameter type, parameter sampling rate, parameter unit and parameter descriptions. Each FDR data set contains three types of parameters: continuous, discrete, and warning signals. Furthermore, the discrete category can be further categorized as ordinal or binary. Notice that some parameters have outliers and sensor measures are corrupted by measurement noise as shown in **Figure 4.4**.

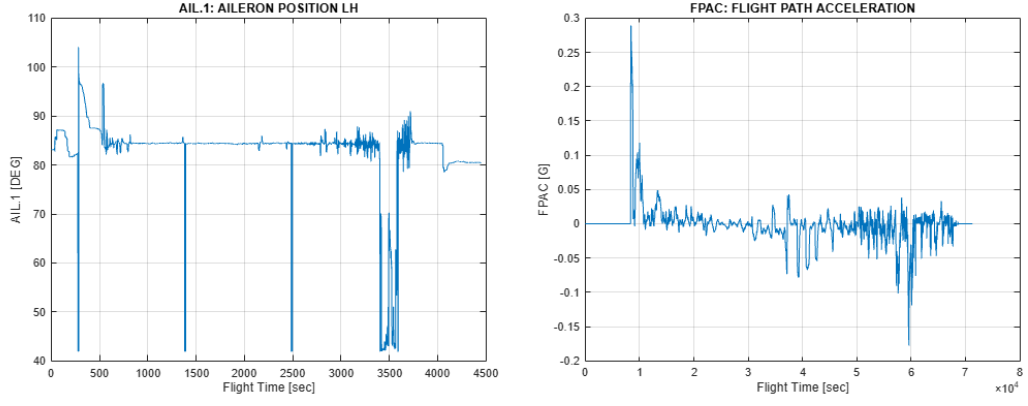


**Figure 4.4:** Examples of parameters that include outliers and noise.

One of the characteristics in the FDR data is that there exist outliers in some parameters, as well as grossly corrupted measurement noise in some sensor measurements as shown in **Figure 4.4**. Another characteristic of the FDR data is that each parameter has a considerably different value range because a physical quantity of each parameter is different, as shown in **Figure 4.5**. The value range of body longitudinal acceleration is about -0.2 to 0.3 and the value range of Elevator Position Left as degree is about -40 to 21 in these examples.

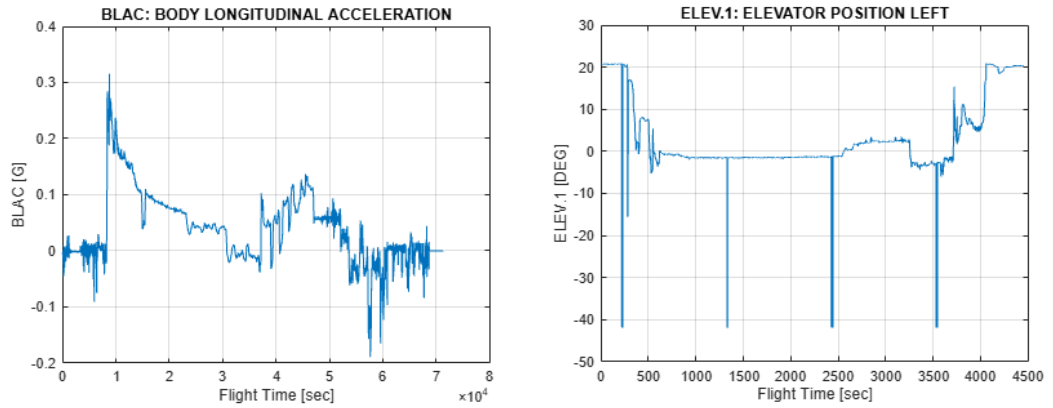
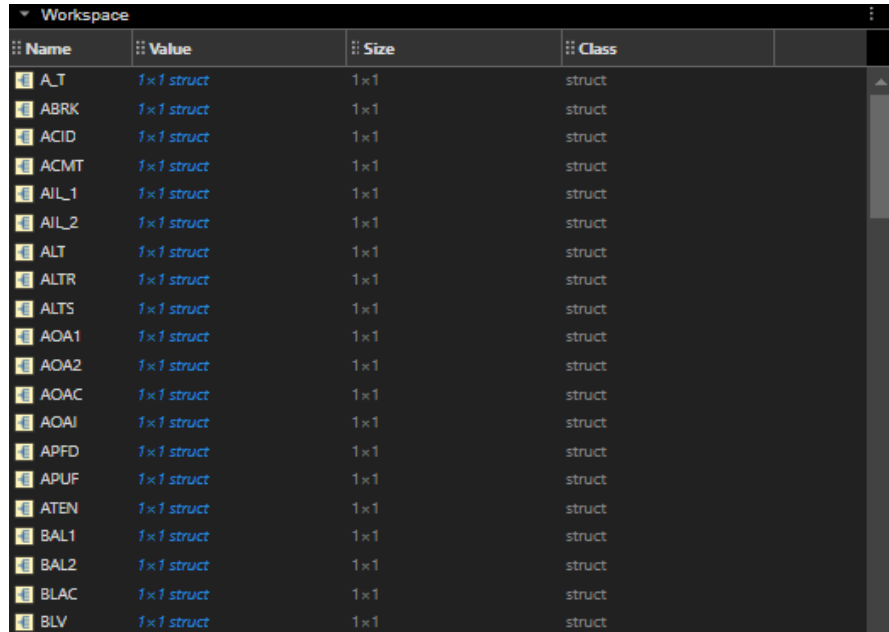


**Figure 4.5:** Examples of parameters that have different value ranges.

## 4.4 Parameter Selection

Correspondingly, each parameter in the FDR data is recorded throughout a flight. Therefore, flight patterns of aircraft are different in terms of their own flight routes; moreover, each parameter is recorded with a different sampling rate. In this regard, time series data in FDRs is not directly comparable. The FDR data should therefore be

mapped to a comparable data space first, in accordance with the time-anchored anchor of a certain event, before getting into the feature generation. In addition, it is also necessary to match up the sampling rate of each parameter. In order to make the FDR data comparable, common flight phases should be selected as the region of interest. That is, given the FDR data, each parameter shall be sampled at a fixed time interval starting from the common anchoring event. Take-off phase is considered as the common anchoring event in this paper because many possible malfunctions occur during this flight phase. **Figure 4.6** shows parameter in FDR data:



| Name  | Value      | Size | Class  |
|-------|------------|------|--------|
| ALT   | 1×1 struct | 1×1  | struct |
| ABRK  | 1×1 struct | 1×1  | struct |
| ACID  | 1×1 struct | 1×1  | struct |
| ACMT  | 1×1 struct | 1×1  | struct |
| AIL_1 | 1×1 struct | 1×1  | struct |
| AIL_2 | 1×1 struct | 1×1  | struct |
| ALT   | 1×1 struct | 1×1  | struct |
| ALTR  | 1×1 struct | 1×1  | struct |
| ALTS  | 1×1 struct | 1×1  | struct |
| AOA1  | 1×1 struct | 1×1  | struct |
| AOA2  | 1×1 struct | 1×1  | struct |
| AOAC  | 1×1 struct | 1×1  | struct |
| AOAI  | 1×1 struct | 1×1  | struct |
| APFD  | 1×1 struct | 1×1  | struct |
| APUF  | 1×1 struct | 1×1  | struct |
| ATEN  | 1×1 struct | 1×1  | struct |
| BAL1  | 1×1 struct | 1×1  | struct |
| BAL2  | 1×1 struct | 1×1  | struct |
| BLAC  | 1×1 struct | 1×1  | struct |
| BLV   | 1×1 struct | 1×1  | struct |

**Figure 4.6:** NASA’s FDR Data Example in MATLAB

The value ranges of each parameter differ significantly, as previously mentioned. Using these parameters together without adjustments can lead to inaccurate results. To address this, all parameters must be transformed to standardize their value ranges.

Preliminary Common Pattern Analysis was performed using a representative DFDR dataset from NASA [55]. The dataset consisted of 655 flights over different aircraft types.

The section below depicts the results obtained when applying the method to this dataset for the anomaly detection during the takeoff phase. All the abnormal flights that were detected have been further analyzed to confirm their abnormality and to characterize the deviations that have been observed. In addition, in those cases where

more than one pattern of nominal behavior was found, differences among such patterns were also analyzed to know what the causes of the variations are.

Sampling rates of each parameter are different as data structure of FDR shown in **Figure 4.7**.

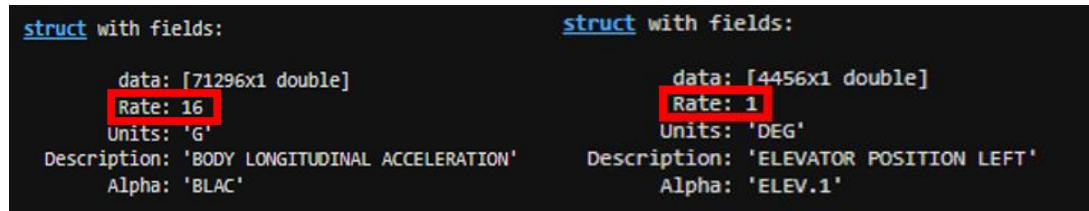


**Figure 4.7:** Different sampling rates of BLAC and ELEV.1

The other oddity of the FDR data in use is that flight patterns in FDR data are different according to the various flight routes of aircraft. For instance, **Figure 4.8** shows the altitude of the FDR data related to two different aircraft, A and B. Here, it is easily observable that the flight pattern of aircraft A is rather different from those of the rest of the aircraft. Hence, time series data obtained from the various FDRs are directly not comparable in nature.

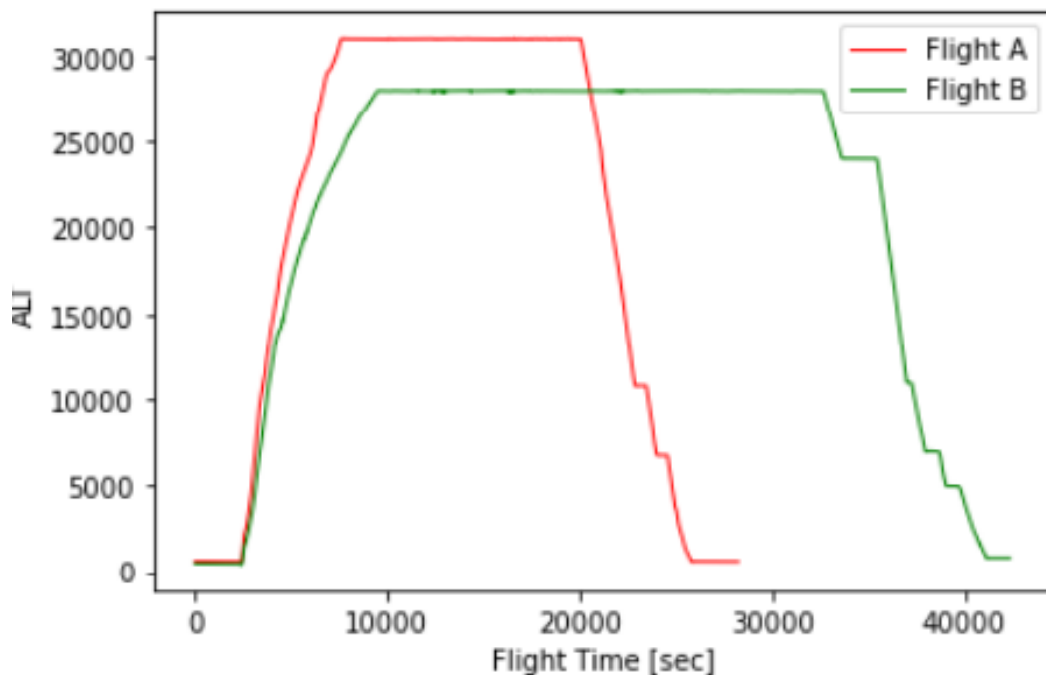


**Figure 4.8:** The examples of selective parameters of different FDR data.

#### 4.4.1 Sampling Rates Synchronization

Where different parameters correlate, acquisition rates may still vary with substantial discrepancies among their sensor sampling frequencies. A few require analysis for standardization and compatibility with all other parameters on the data set. All recorded values in this experiment were thus leveled out for uniformity of sampling intervals to accurately reflect comparisons or analyses that are made.

It can be inferred from the **Figure 4.9** that the "ALT" parameter, which is altitude, contains 40,704 samples, while the "PSA" parameter means pressure altitude and contains 20,352 samples. Therefore, this gives a sampling rate of 4 and 2 samples per second, respectively. This kind of discrepancy in sampling rates complicates direct comparison or joint analysis of these parameters. For this, all parameter values are synchronized to a common temporal resolution of 1-second intervals. This resampling process will ensure that data points are uniformly aligned across the temporal dimension.

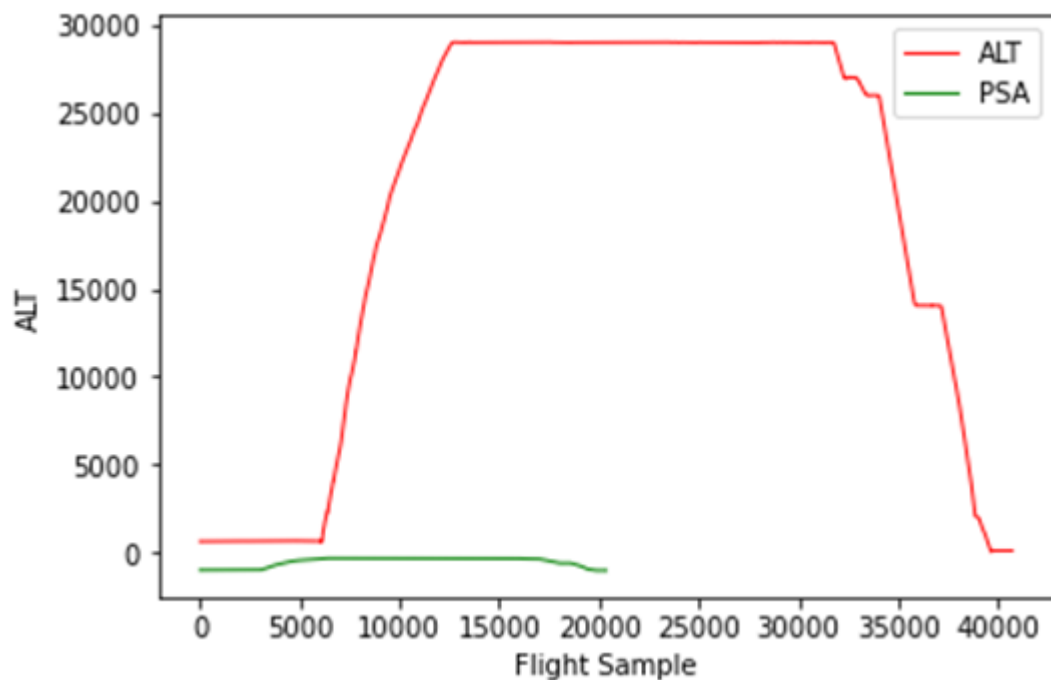


**Figure 4.9:** Original ALT and PSA parameters

Once the synchronization is complete, parameters can be directly compared in terms of time-series length and their respective trends over time. This results in a

standardized dataset, which sets a consistent basis for further analysis, as shown in the **Figure 4.10** below, where the temporal alignment of parameters enables much stronger and more meaningful interpretation of their relationships and behaviors.

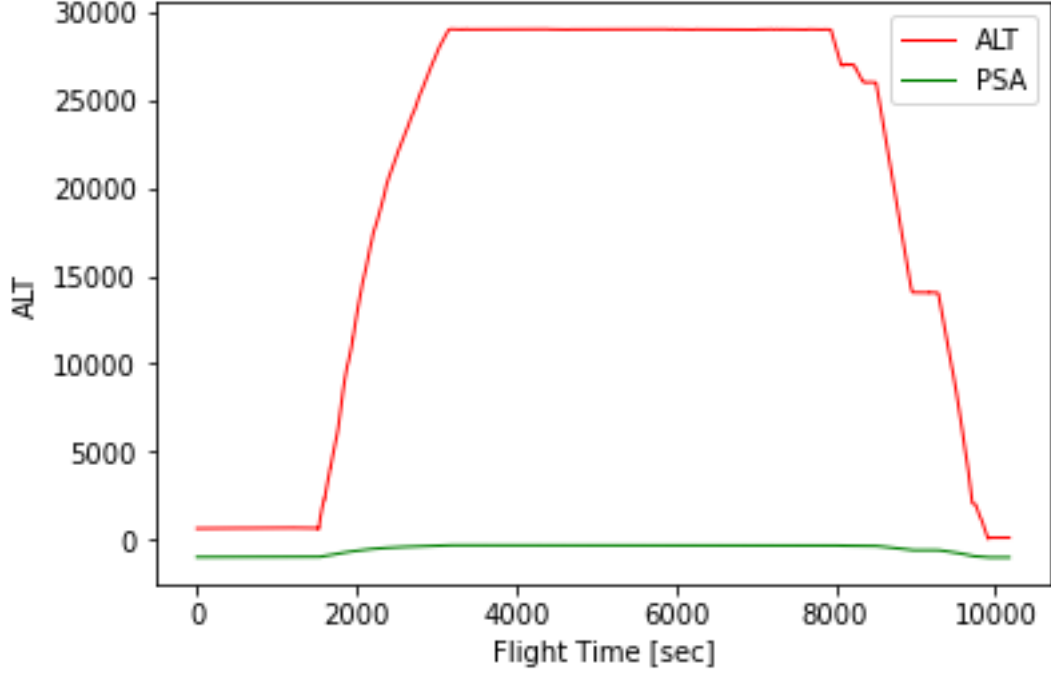


**Figure 4.10:** Synchronized ALT and PSA parameters

#### 4.4.2 Standardization

As mentioned before, the value ranges of each parameter are considerably different. If such parameters are used together, incorrect results are expected. Therefore, all the parameters are needed to be transformed in order to uniformize their value ranges. Standardization, also known as Z-score scaling or zero-mean scaling, is a common method used in data preprocessing to scale and center features in machine learning. This method transforms the data in a way that makes it suitable for algorithms that assume a standard normal distribution. Standardization makes the data more Gaussian-like, which is useful for some machine-learning algorithms. Following formula shows the mathematical for standardization:

$$x'_k = \frac{x_k - \bar{x}}{s_x} \quad (4.1)$$

$$k = 1, \dots, n \quad (4.2)$$

$\bar{x}$  represents the average value of the elements of the time series vector over the sampling interval,  $s_x$  is the variance,  $x_k$  and  $x'_k$  will indicate, respectively the original data point and transformed data point at time  $k$ . One of the key advantages of data transformation is that sometimes it may be able to unmask patterns showing similarities of parameters. For example, although a time series of two parameters can look completely different, as depicted in **Figure 4.11**, their transformations come very close to each other, which actually reveals a relationship among them. As can be seen on the **Figure 4.12** This relationship is achieved because pressure and altitude are inversely proportional physically. The data transformation can extract this physical pattern and it is helpful to the correlation analysis.

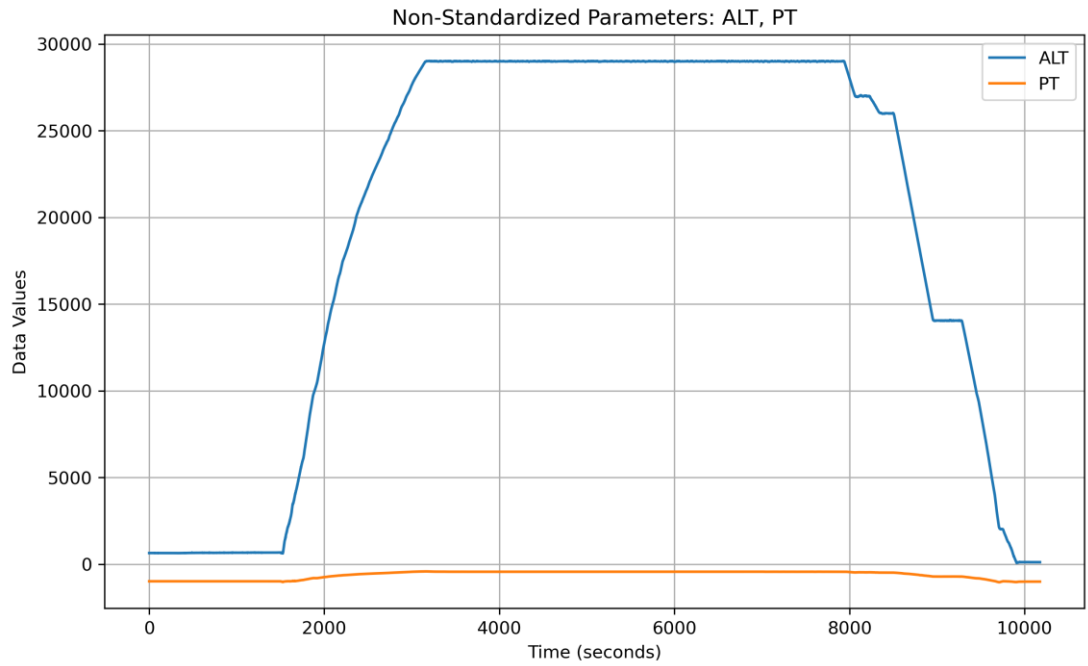


**Figure 4.11:** Original ALT (Altitude) and PT (Total Pressure) time series comparison

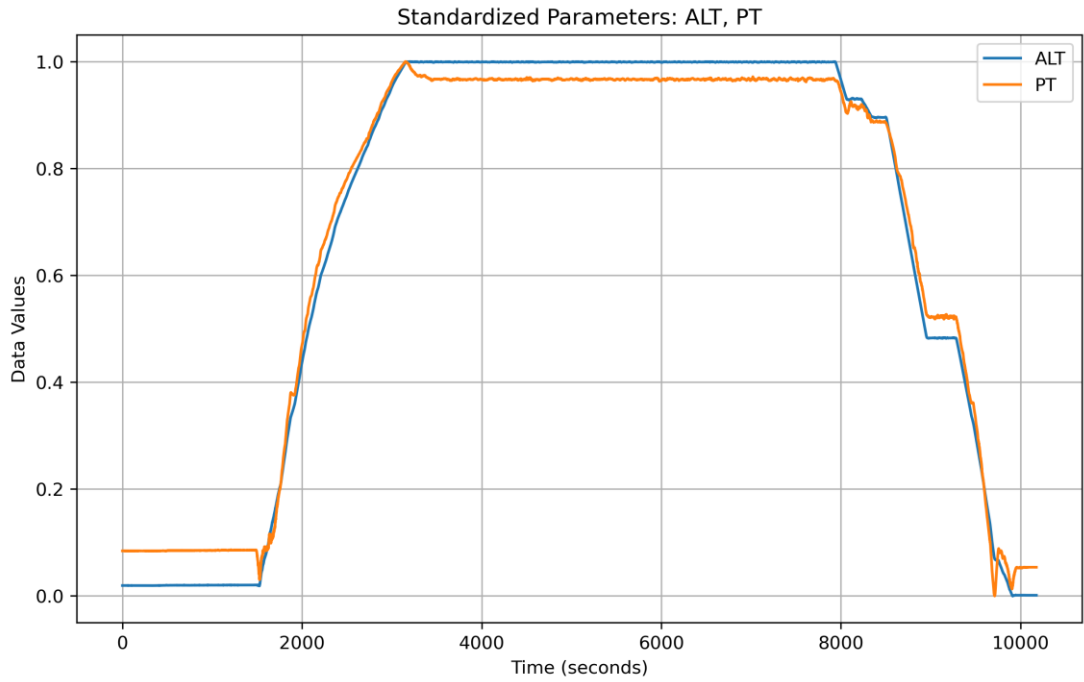


**Figure 4.12:** Standardized ALT (Altitude) and PT (Total Pressure) time series comparison

#### 4.4.3 Correlation Analysis

Following this approach, the feature is defined as a high-dimensional vector constructed by arranging all of the sampled parameters in order. By comparing these feature vectors, the anomalous patterns included in the FDR data can be found. However, the performance might deteriorate due to accuracy among the parameters. This would be avoided by detecting the redundant information in the continuous parameters using Pearson Correlation Coefficients (PCC) in advance.

The approach set out in this thesis mitigates the impact of correlated parameters by providing a method for identifying them and determining two derived measures-the average and maximum differences of all the parameters within each identified set. The average tells the general trend, while the maximum difference is a tell-tale sign that will denote the presence of abnormal patterns. These sets of correlated parameters are then determined by the Pearson correlation coefficients, which refer to one of the oldest measures that were used to determine the strength of linear dependence between two variables. This method is still one of the most important today in statistics [63].

PCC is a statistical metric measuring the strength and direction of the linear interrelationship between two variables [64]. Its application has been wide and varied, covering a number of domains such as data analysis and classification [58], [65], clustering [66], financial analysis [67], and biological research [68] etc.

Suppose that two different time series variable as follows:

$$x_i = x_{i1}, x_{i2}, \dots, x_{ik}, \dots x_{in} \quad (4.3)$$

$$x_j = x_{j1}, x_{j2}, \dots, x_{jk}, \dots x_{jn} \quad (4.4)$$

Where  $x_i$  and  $x_j$  is the values of the  $i$ -th time series data. Hence,  $x_{ik}$  and  $x_{jk}$  is the values at time  $k$ .

The Pearson Correlation Coefficients ( $S_{i,j}$ ) for both series are calculated as follows:

$$S_{i,j} = \frac{\sum_{k=1}^n (x_{ik} - x_i)(x_{jk} - x_j)}{\sqrt{(\sum_{k=1}^n (x_{ik} - x_i)^2)(\sum_{k=1}^n (x_{jk} - x_j)^2)}} \quad (4.5)$$

Some groups of highly correlated parameters were identified by using the correlation analysis. One such group is shown in **Table 4.1** as an example. **Figure 4.13** shows the trends of these parameters, which also justifies that these parameters are highly correlated. The reason for such a correlation is that the physical variables of these parameters naturally depend on each other with respect to altitude. Thus, the previously hidden correlations among these parameters became apparent after data transformation.

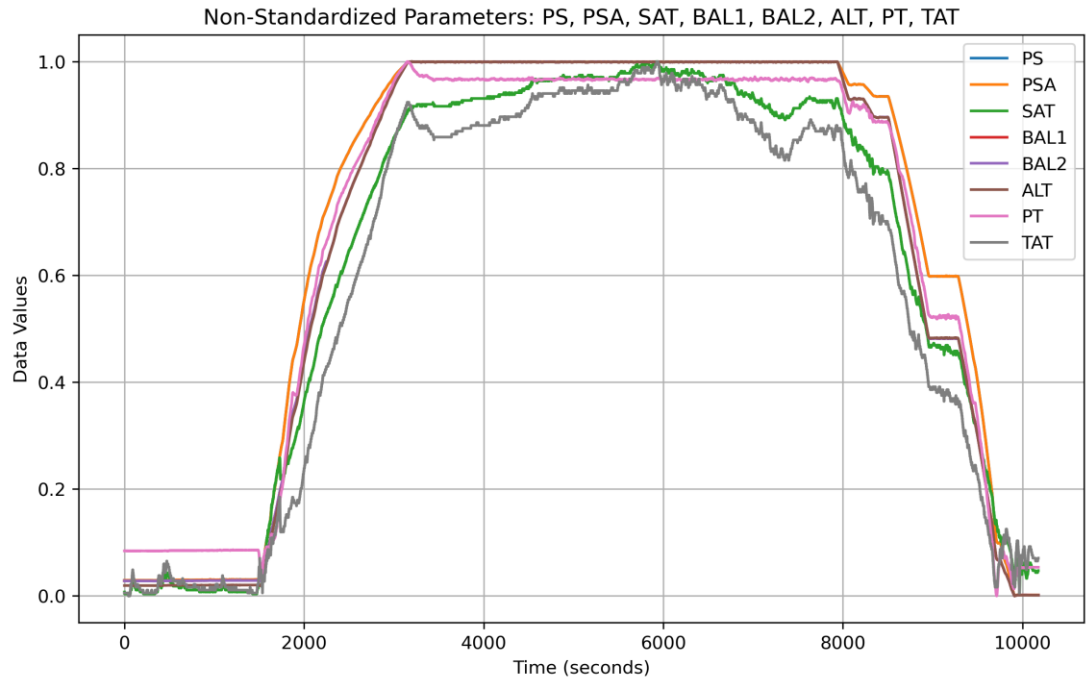


**Figure 4.13:** Time series data of correlated parameters.

**Table 4.1:** Description of parameter in Figure 4.13

| Parameter | Description             |
|-----------|-------------------------|
| PS        | Static Pressure         |
| PSA       | Average Static Pressure |
| SAT       | Static Air Temperature  |
| BAL1      | Baro Correct Altitude 1 |
| BAL2      | Baro Correct Altitude 2 |
| ALT       | Pressure Altitude       |
| PT        | Total Pressure          |
| TAT       | Total Air Temperature   |

## 4.5 Align Variables According to Take-off & Touchdown

Having processed the data and removed unnecessary information, we have now gathered the correlated data. In this section, we will align the parameters accordingly. The proposed algorithm is designed to detect anomalous flights by leveraging patterns

of standard flights. The temporal series of flight parameters are synchronized to particular events, such as the initiation of take-off power at take-off or the occurrence of touchdown in the final approach. Such time anchors are necessary so that meaningful comparisons can be achieved across flights. In this thesis, take-off alignment approach will be used.

The algorithm is designed to function efficiently with continuous flight parameters and is vulnerable to deviations in their magnitudes. We can consider Common Pattern Analysis as one type of exceedance detection. Whereas both methods check deviations in flight parameters, this approach analyses all parameters available and performs its comparisons, using nominal values directly from the data, without relying on predefined thresholds. Another important capability of Common Pattern Analysis is handling multiple standard operation modes in general, and even when the exact number is not known.

For purposes of this analysis, takeoff phase was defined generally as the period of time after takeoff.

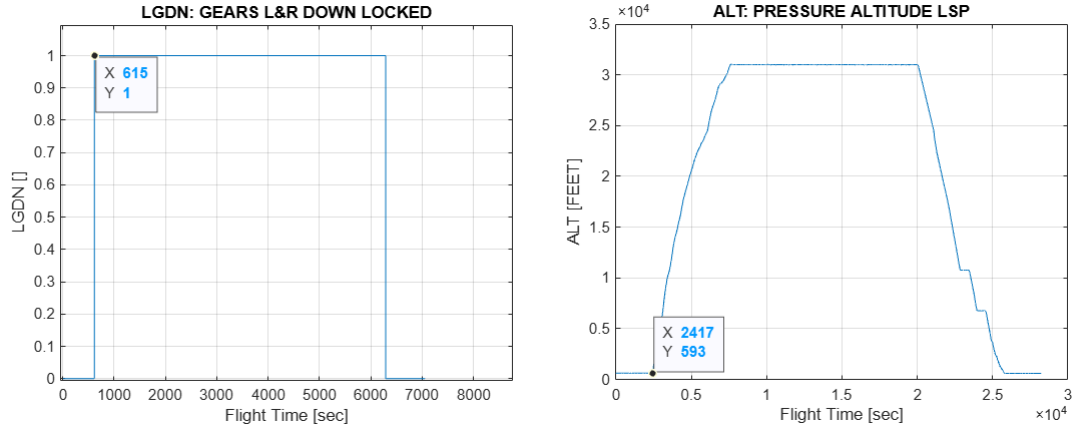


**Figure 4.14:** Flight A LGDN and ALT parameter

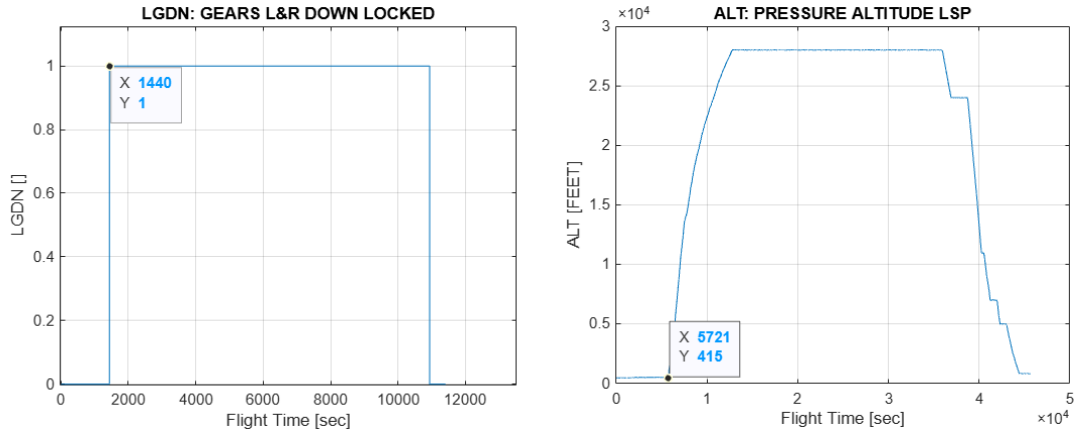


**Figure 4.15:** Flight B LGDN and ALT parameters

The landing gear down lock data can be used for this alignment. This is recorded as LGDN, which is available in the FDR dataset. It can be observed from **Figure 4.14** and **Figure 4.15** that the LGDN value is in Boolean format. In this case, when the value of LGDN is TRUE, that means Down Lock sensor ON as shown in **Figure 4.16**, which designates the instant when the aircraft became airborne. This thesis considers the analysis of flight parameters after taking off into the air; hence, this value will be appropriate for use as an anchor point.

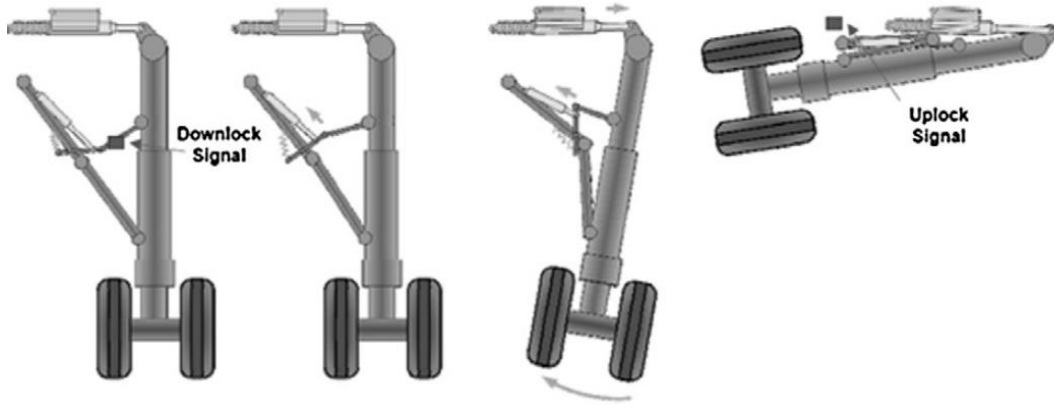


**Figure 4.16:** Standard Landing Gear Retraction Cycle

In the alignment of take-off times, the earliest take-off time is an anchor point to which all the time series data is shifted.

$$x_a = x_{a1}, x_{a2}, \dots, x_{ak_{toff}}, \dots x_{an} \quad (4.6)$$

$$x_b = x_{b1}, x_{b2}, \dots, x_{bk_{toff}}, \dots x_{bn} \quad (4.7)$$

For example, consider the flights A and B that should be aligned, where  $x_{ak_{toff}}$  and  $x_{bk_{toff}}$  denote the take-off for their respective time series data.

$$x_s = x_{k_{toff}} - x_{ap} \quad (4.8)$$

Where  $x_{k_{toff}}$  is take-off time and  $x_{ap}$  is the anchor point of the analyze group.  $x_s$  is calculated as the shifting value for the time series

Earliest take-off sample is selected as an anchor point in the analysis group (the group consists of the flight data to be analyzed). For all other flights in this group, their data are shifted backward by the difference between their takeoff time,  $x_{k_{toff}}$  and the  $x_{ap}$  anchor point. Thus, all flights within the analysis group will be aligned with respect to their takeoff points.

$$x_a = x_{a1} - x_s, x_{a2} - x_s, \dots, x_{ak_{toff}} - x_s, \dots x_{an} - x_s \quad (4.9)$$

This is a very important step in the alignment process, which has to be done to ensure that accurate analysis, such as outlier detection, is performed on the data. **Figure 4.17** illustrates the initial state of Flight A and Flight B, where their data points are misaligned, making direct comparisons and analyses challenging. In contrast, **Figure 4.18** demonstrates these flights after alignment, thereby facilitating effective comparative analysis.

The alignment technique depicted here is essential for analyzing extensive flight datasets. By ensuring that all data points are synchronized, this method allows for precise and meaningful comparisons. It is particularly beneficial when dealing with large groups of flights, as it aligns the entire dataset uniformly.

**Figure 4.19** shows an unaligned group of flights. Due to the lack of synchronization, drawing accurate conclusions from this data is difficult. Conversely, **Figure 4.20** presents the same group of flights post-alignment, illustrating the efficacy of the alignment process in preparing the data for thorough analysis.

In summary, the alignment method is crucial for ensuring the integrity and accuracy of flight data analysis. This process enhances the comparability and reliability of the results, making it an indispensable tool for both small-scale and large-scale datasets.

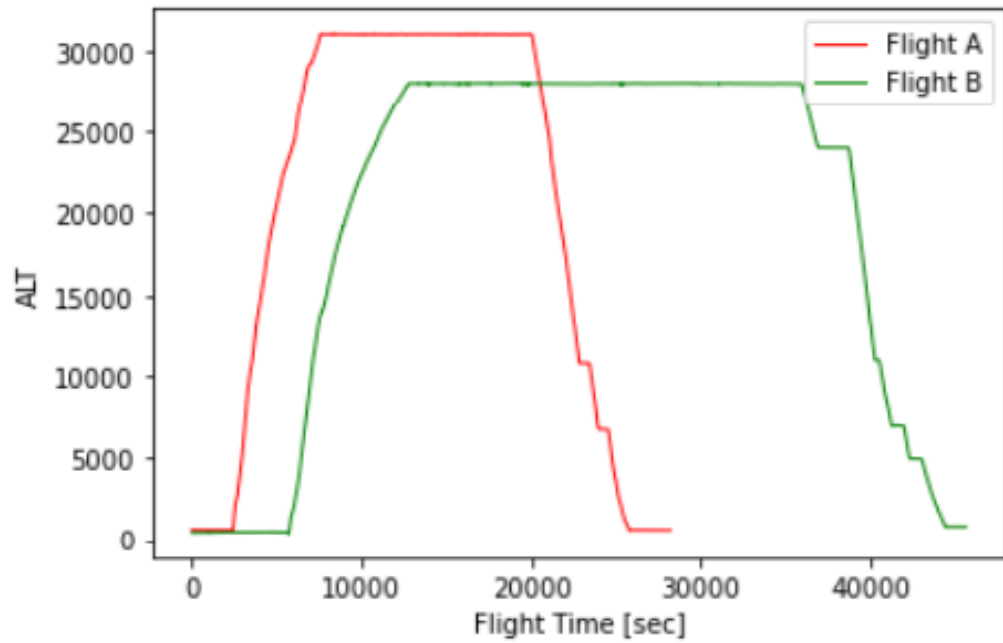


**Figure 4.17:** Non-Aligned Altitude parameter according to Landing Gear Down Lock parameter

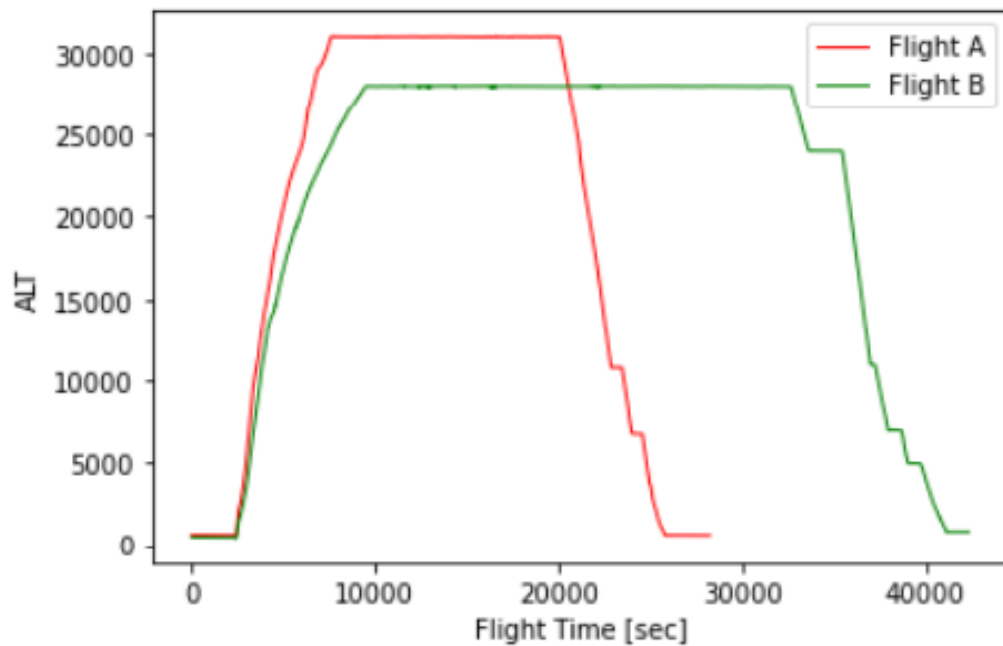


**Figure 4.18:** Aligned Altitude parameter according to Landing Gear Down Lock parameter.

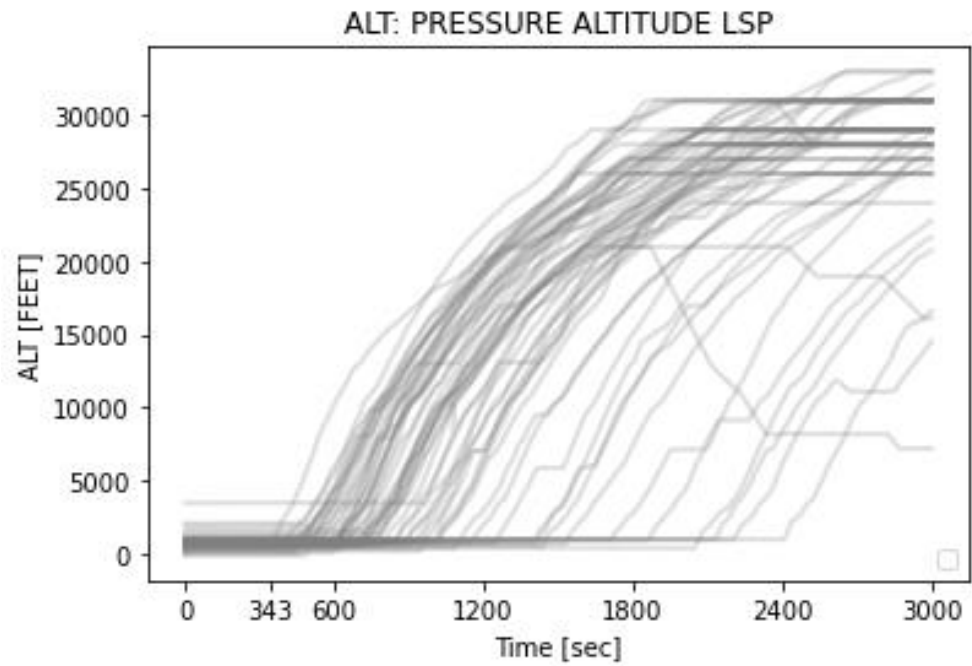


**Figure 4.19:** Non-Aligned Altitude parameters according to Landing Gear Down Lock parameter in The Analyze Group.

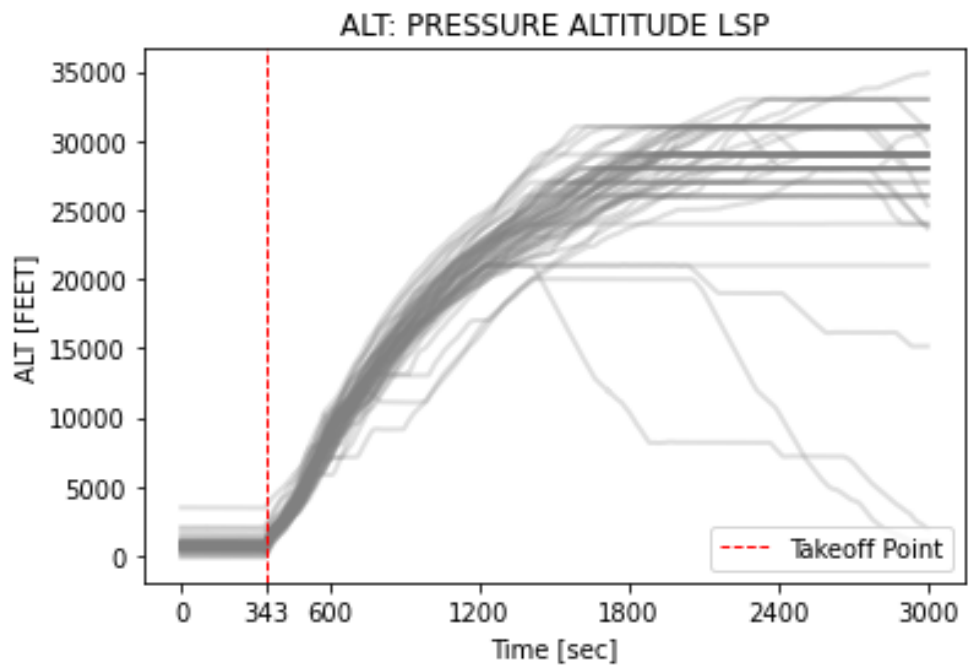


**Figure 4.20:** Aligned Altitude parameters according to Landing Gear Down Lock parameter in The Analyze Group.

The FDR dataset used in the first experiments consisted of 655 flights with varied origins and destinations that one airline operates. As mentioned in Section 4.3

## **4.6 Flight Data Pre-Processing Experiment Result**

### **4.6.1 Common Pattern Analysis**

This section provides examples shown by the following figures of the analyzed group parameters derived from the 655 flights mentioned earlier. The following figures give the general behavior of flights during the takeoff phase. Critical aspects of flight dynamics could be seen from these visual representations. The analysis aims to provide a comprehensive understanding of how flights typically perform during this crucial stage of operation.

Each flight was represented by 186 parameters, which contained the main flight and engine parameters of the position, velocity, temperature, acceleration, attitude, control surface position, wind conditions and environmental pressures.

The data were transformed from time-series to vector, with observations sampled at 1-second intervals. The Analyze Group Data for this research was from pre take-off application to post take-off.

The preprocessed parameters, which have gone through standardization and alignment, will be used in this section to further detect the anomalous flights. And the top 5% of the abnormal flights were analyzed if they behave abnormally by comparing the distributions between their flight parameters all flights. In this section, two detailed examples will be introduced on how to recognize abnormal behavior. For each case, the most characteristic flight parameters are shown in graphs with a standard format. The trends of the majority of flights are visualized by blue strip and dark blue strip.

The dark blue bands on the graph represent the %25th to %75th, while the light blue bands represent the %5th to %95th of all flights. As a result, the dark blue area occupies half of the total area and the blue area constitutes 90 per cent of the total area. The red dashed line indicates the aligned takeoff point across all flights. Examples illustrating the plotting of this calculation is presented as following.

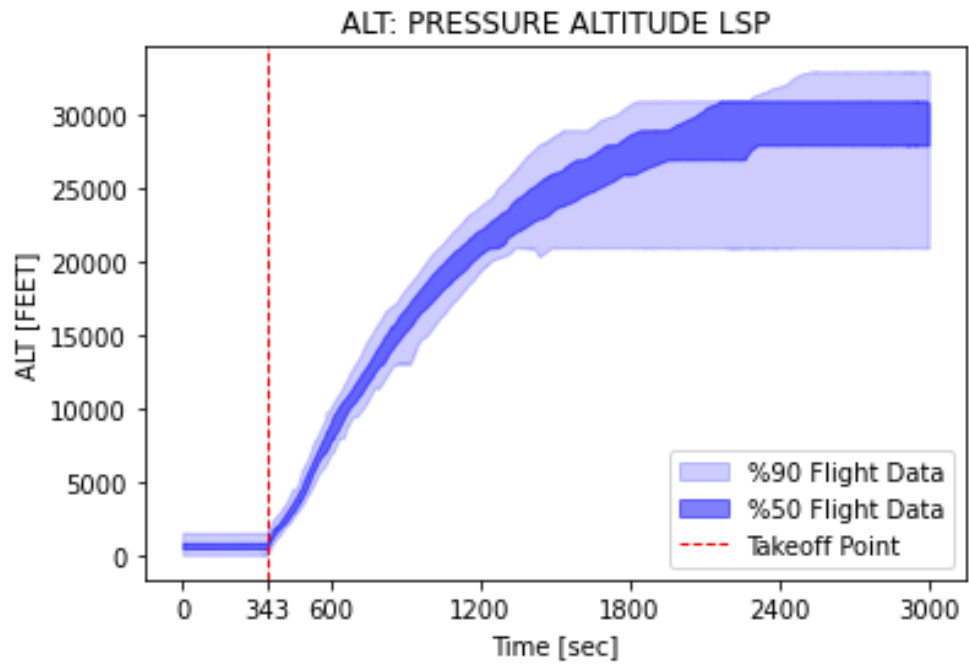


**Figure 4.21:** ALT: Pressure Altitude Data

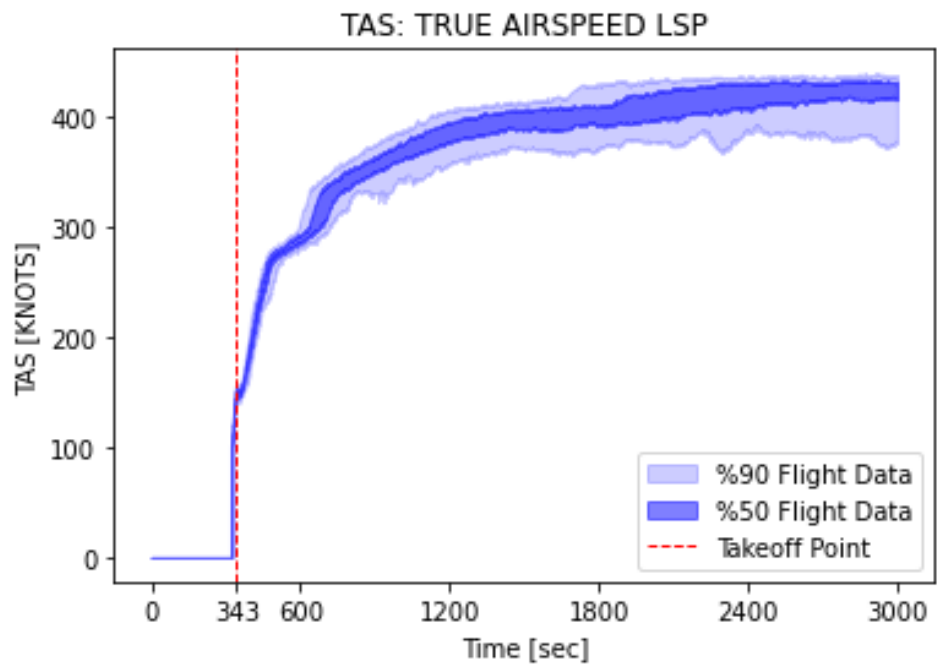


**Figure 4.22:** TAS: True Airspeed Data

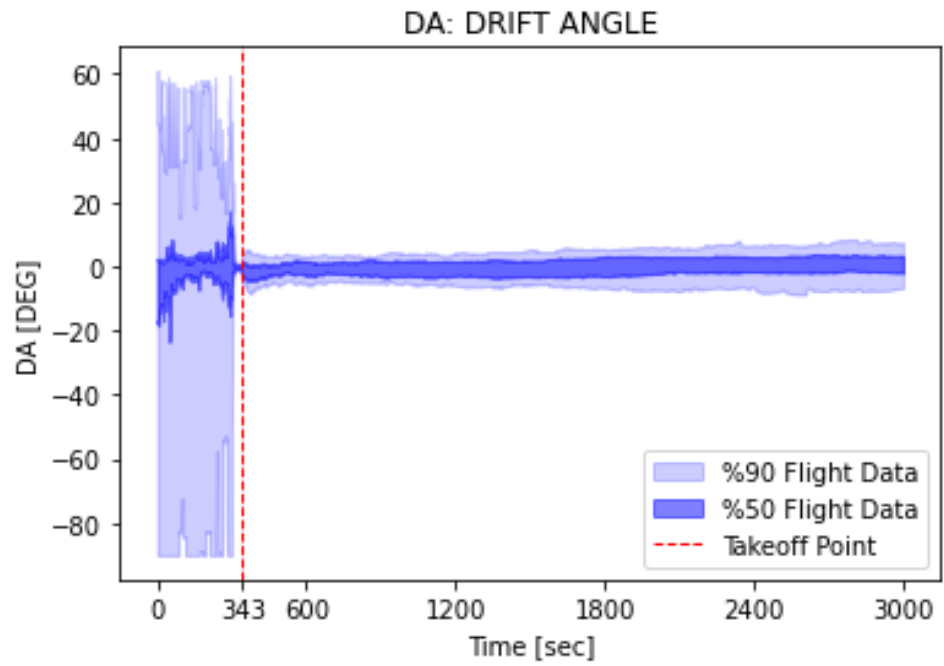


**Figure 4.23:** DA: Drift Angle Data

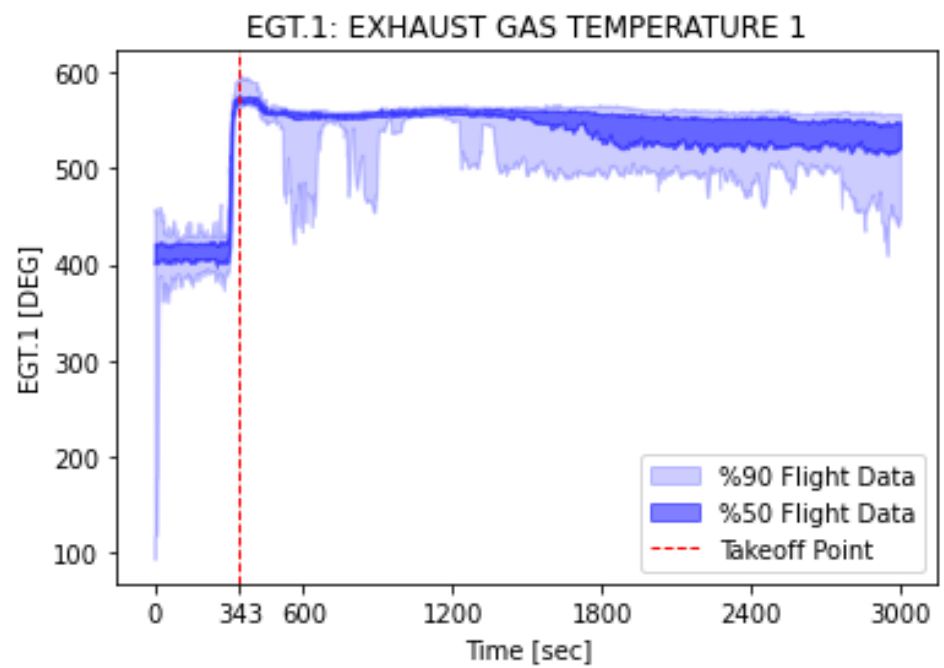


**Figure 4.24:** EGT: Exhaust Gas Temperature

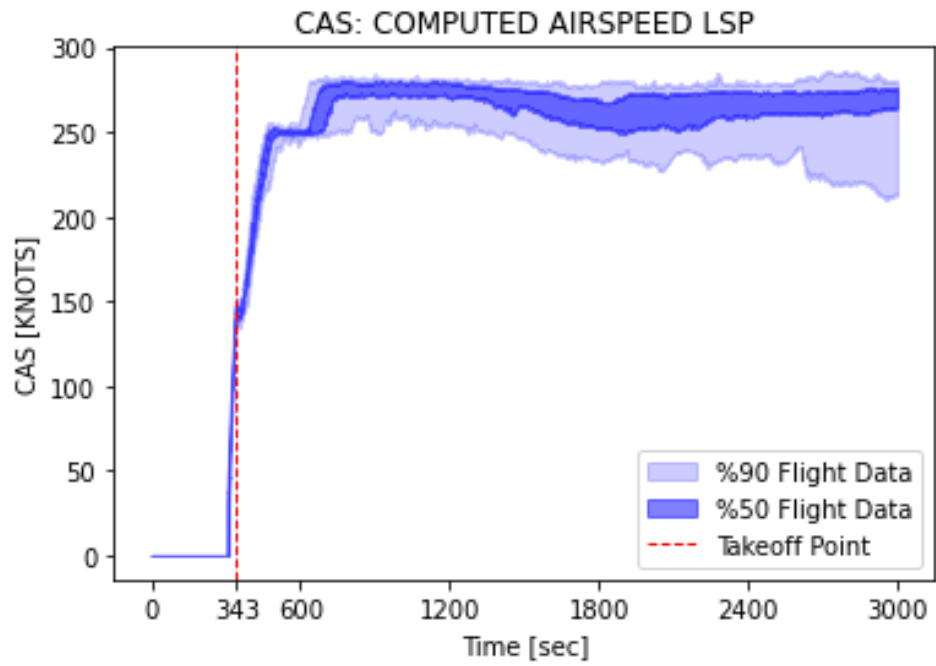


**Figure 4.25:** CAS: Computed Airspeed Data

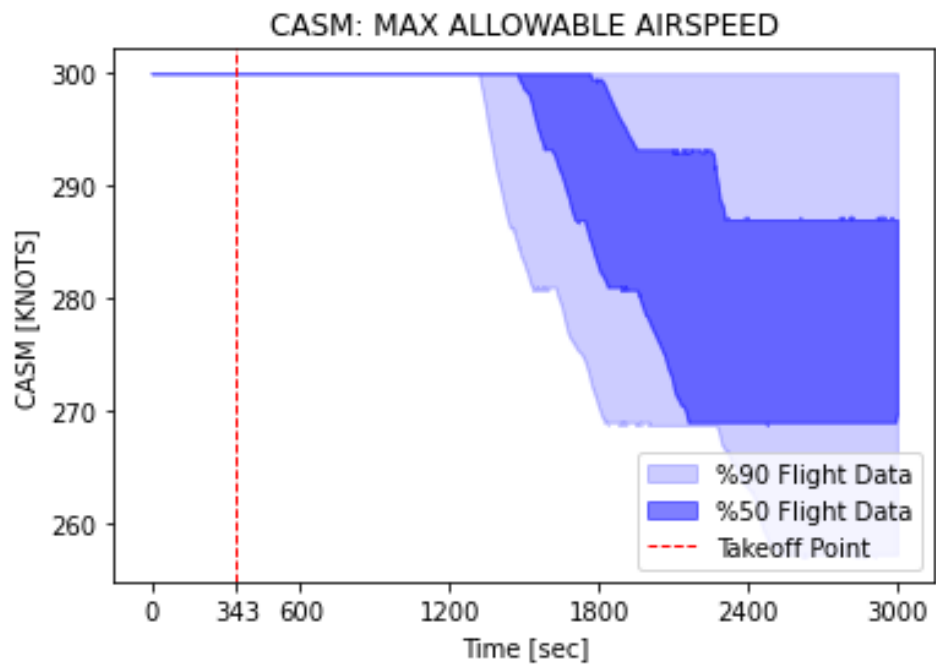


**Figure 4.26:** CASM: Max Allowable Airspeed

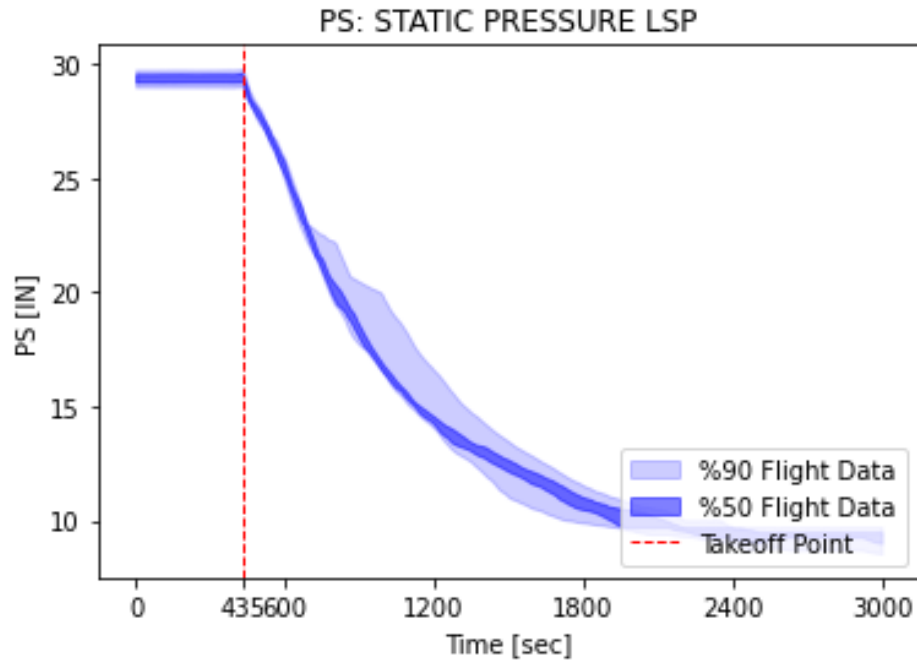


**Figure 4.27:** PS: Static Pressure Data

#### 4.6.2 Analysis of New Flight Data Using Common Pattern Analysis

Creating flight parameter plots requires the generation of percentile statistics derived from raw Flight Data Recorder (FDR) data. Example of an image for a continuous flight parameter is given in **Figure 4.28**. The black dashed circle (if exists) indicates the anomaly in the graph.

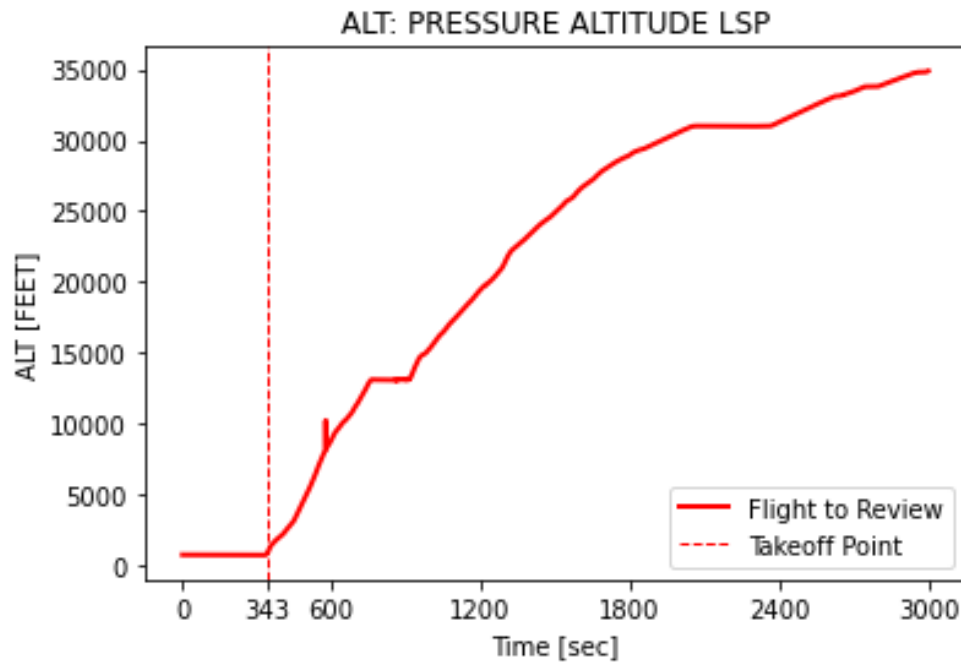


**Figure 4.28:** Example of Flight to Review in ALT parameter.

Here the value of the airspeed for the flight under review is in red. Subsequently, we will integrate two distinct datasets: the parameters derived from the common pattern analysis and those corresponding to the flight under review. This allows the flight under review and common pattern analysis to be directly compared in one graph. A visualization like this makes the identification of clear outliers much easier when flight data comes up against the established patterns. **Figure 4.29** presents the integrated graph representing the combination of the two datasets.

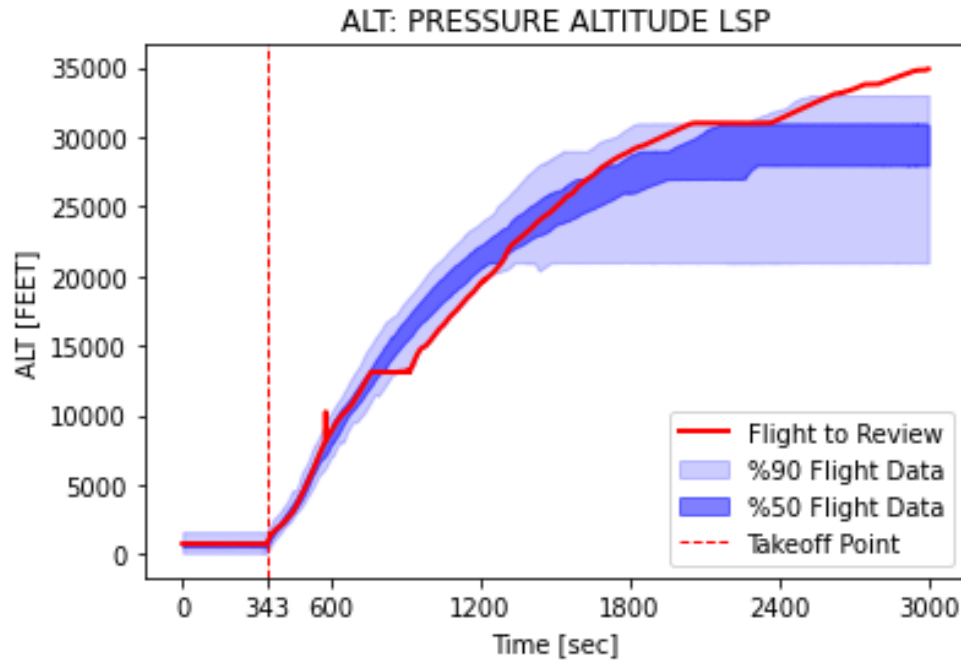


**Figure 4.29:** Integrated graph representing the combination of the two datasets.

In the **Figure 4.29**, anomaly in 600<sup>th</sup> which will be examined in the Section 4.6.3 second is significantly visible.

### 4.6.3 Common Pattern Analysis Experimental Results of Flight 3210525

The ALT value is the altitude of the aircraft for flight number 3210525 and it is illustrated in **Figure 4.30**. Analysis of this graph reveals a minor spike occurring at the 600th second. This anomaly could potentially be attributed to a sensor error or a sudden change in altitude. Therefore, further evaluation is necessary to determine the underlying cause of this spike.

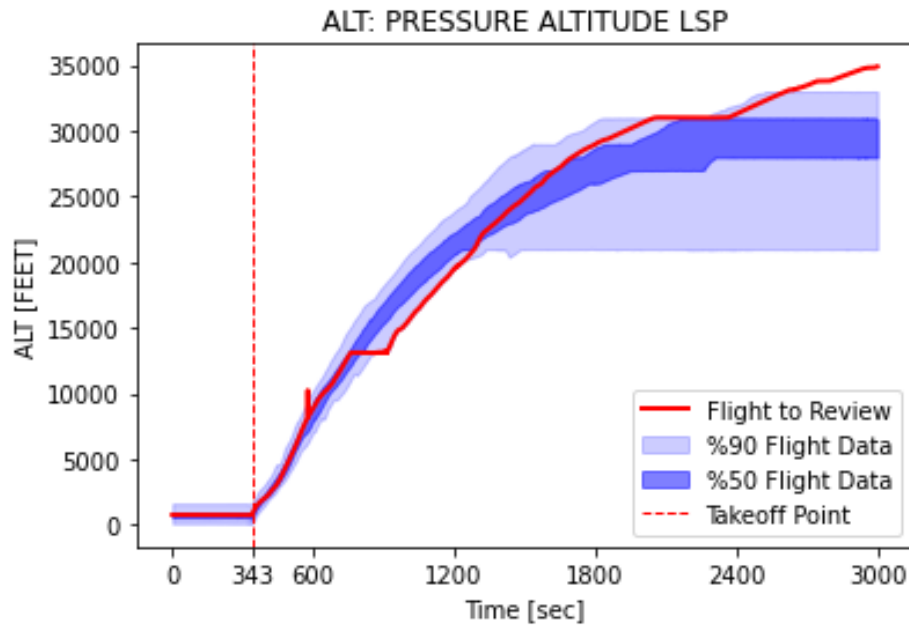


**Figure 4.30:** Flight 3210525 ALT parameter in Common Pattern

The True Airspeed<sup>6</sup> (TAS) value for flight number 3210525 is shown in **Figure 4.31**. It is observed that although the TAS value exhibits some fluctuations, it remains within the 90% confidence interval. Consequently, this behavior can be considered normal.

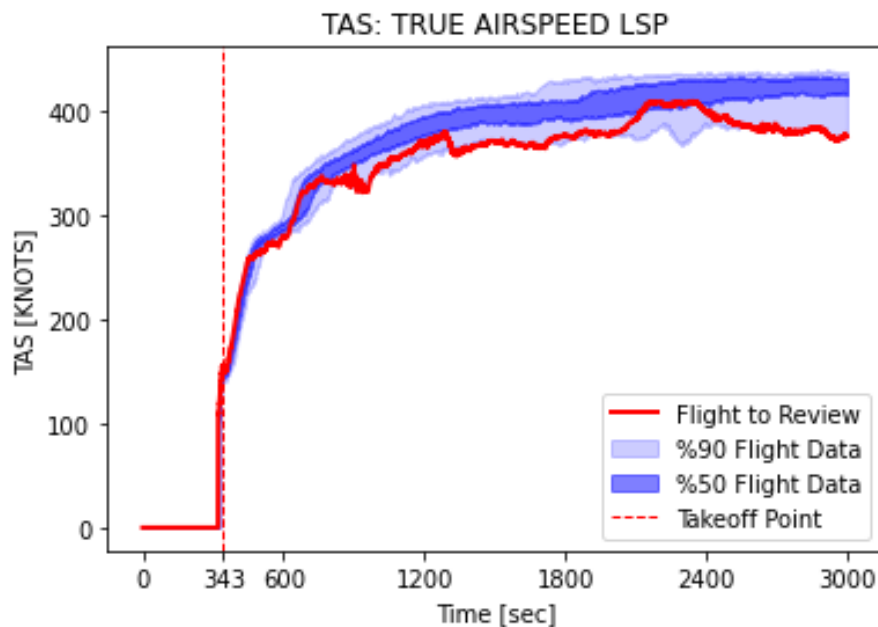


**Figure 4.31:** Flight 3210525 TAS parameter in Common Pattern

---

<sup>6</sup> True airspeed is the speed of your aircraft relative to the air it's flying through. As you climb, true airspeed is higher than your indicated airspeed.

Next, we can observe the values for Drift Angle (DA), Computed Airspeed (CAS), Computed Airspeed Max (CASM), and Pressure Sensor (PS), which fall within the average range for the flights in this analysis group.

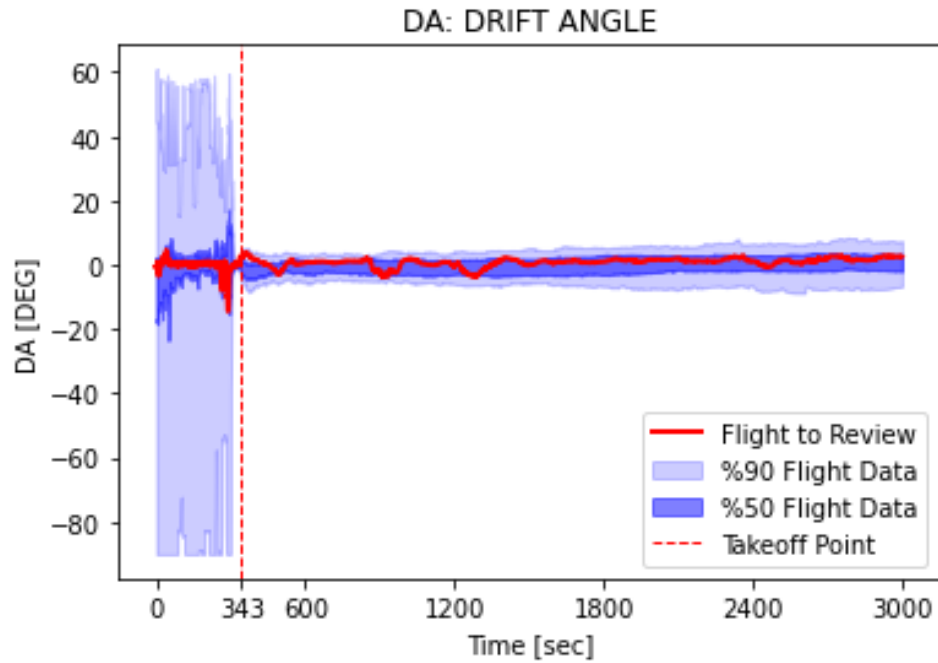


**Figure 4.32:** Flight 3210525 DA parameter in Common Pattern

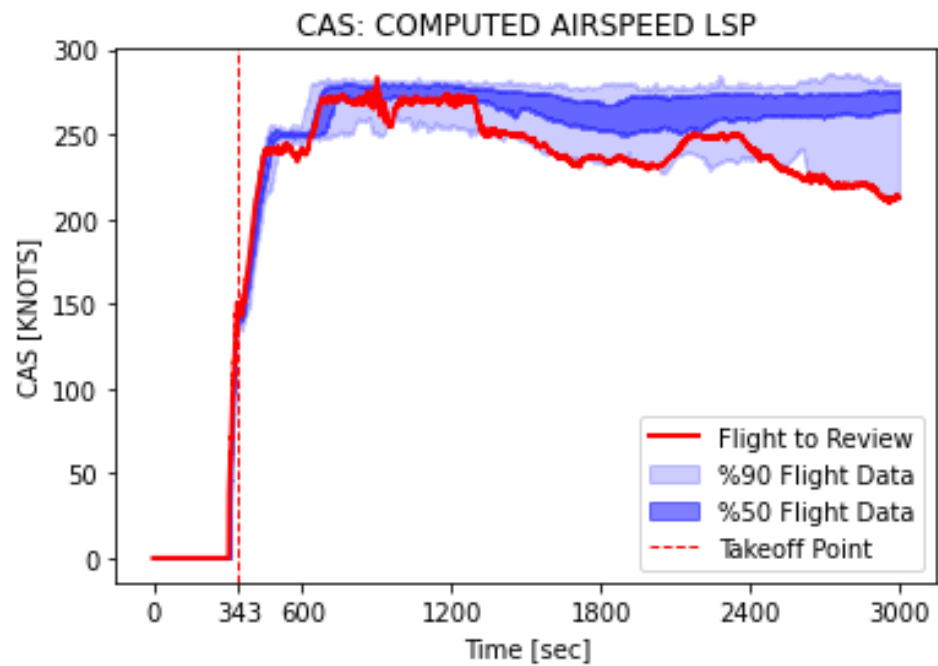


**Figure 4.33:** Flight 3210525 CAS parameter in Common Pattern

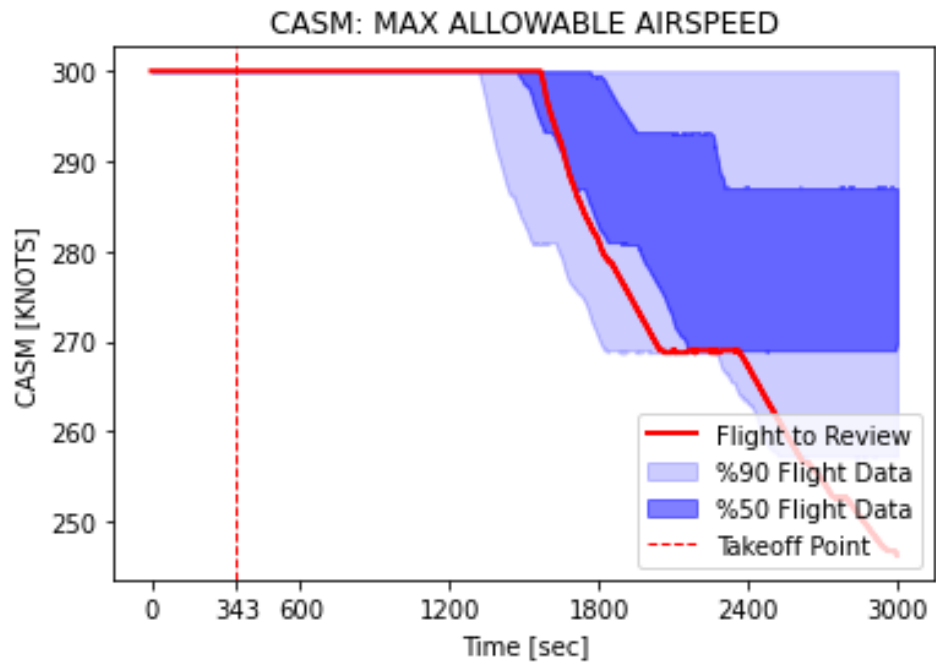


**Figure 4.34:** Flight 3210525 CASM parameter in Common Pattern

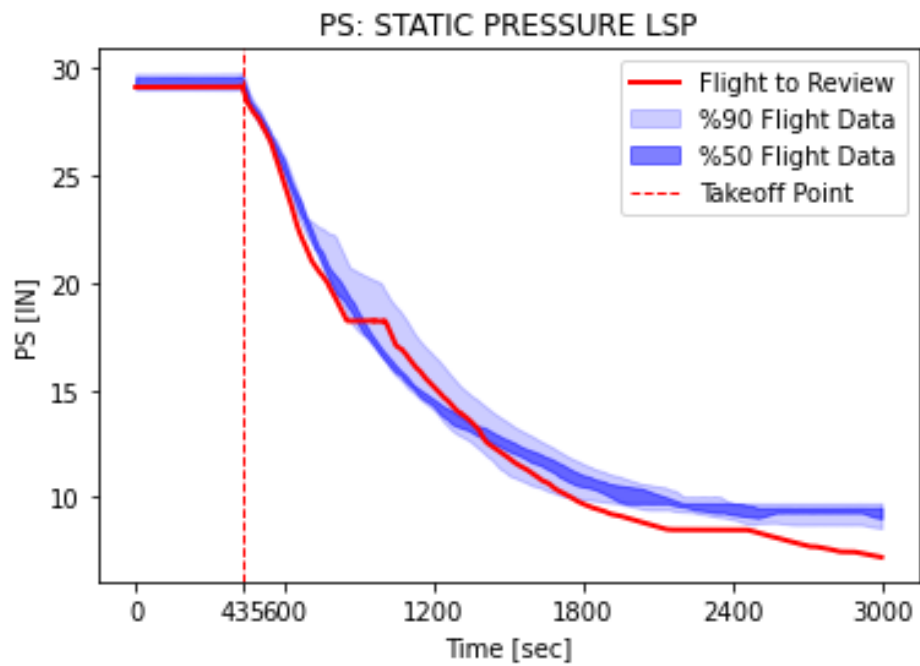


**Figure 4.35:** Flight 3210525 PS parameter in Common Pattern

## **5. REGRESSION BASED ANOMALY DETECTION**

### **5.1 Difficulties in FDR Outlier Detection**

This thesis proposes a new methodology of detecting unknown issues in routine airline flights' FDR data. The nature of this new methodology is quite different from the traditional methodologies, which rely on pre-defined flight parameters and their thresholds to detect the interest flights without any a priori assumptions.

Following are the two major issues that have been tried to develop this anomaly detection method:

No specific target is predefined for interest flights detection in FDR data. There are complex relationships among multiple flight parameters and their time-based patterns which can hardly be captured by simple formulas. The raw FDR data of flight parameters may vary broadly in time and often depend on each other due to dynamics of flight, operational procedures, and environmental conditions. One single parameter may also be affected by the variation of several other parameters. Missing Single Standard Pattern for Normal Data:

A second challenge is the inherent lack of a single standard pattern for normal data. Most existing methods assume one fixed standard and anomalous deviations from it. In fact, operational data often contains many standard patterns because different flight phases and different procedures (airport specific, air traffic control related, company specific, etc.) and different data patterns. Therefore, the anomaly detection method developed here, needs to be capable of having multiples of standard patterns within FDR data.

### **5.2 Understanding the Importance of Stationarity in Time Series**

Data preprocessing is one of the most important steps toward insight extraction and effective learning from data. Among several important steps in preprocessing, data transformation is one of the major steps, especially when one is dealing with time

series prediction. Most of the time series models are based on the assumption of stationarity, which means the statistical properties of the process remain constant over time. Real-world datasets are very rarely similar to this assumption; rather, this assumption is an exception. Addressing non-stationarity requires applying specific transformation methods, the choice of which for any data model and any prediction may be very complicated [69].

Although there is a wide variety of time series prediction methods in the literature [60], most of these and a majority of studies addressing time series operate under the assumption that the data is stationary [70]. In stationary time series, statistical properties such as mean, variance, and covariance remain consistent over time and across any subset of the data [60], [70]. However, in reality, this assumption hardly holds, especially when dealing with real-world domains, time series are mostly nonstationary. In case of detection of non-stationarity within a time series, transformation techniques are usually employed to achieve stationarity in order for established prediction methodologies to be applied [69].

Types of non-stationarity covered in this section, namely, mean and variance non-stationarity, are reflected in the background color of each method's icon in the categorization diagram, showing the transformation techniques that can handle these issues. Some of the reviewed methods also deal with other important time series features, including nonlinearity and seasonality. These properties are graphically portrayed in **Figure 5.1** using distinctive patterns-lines horizontal or vertical-which fill the related icons and visually represent, in an intuitive way, what each method is able to cover.

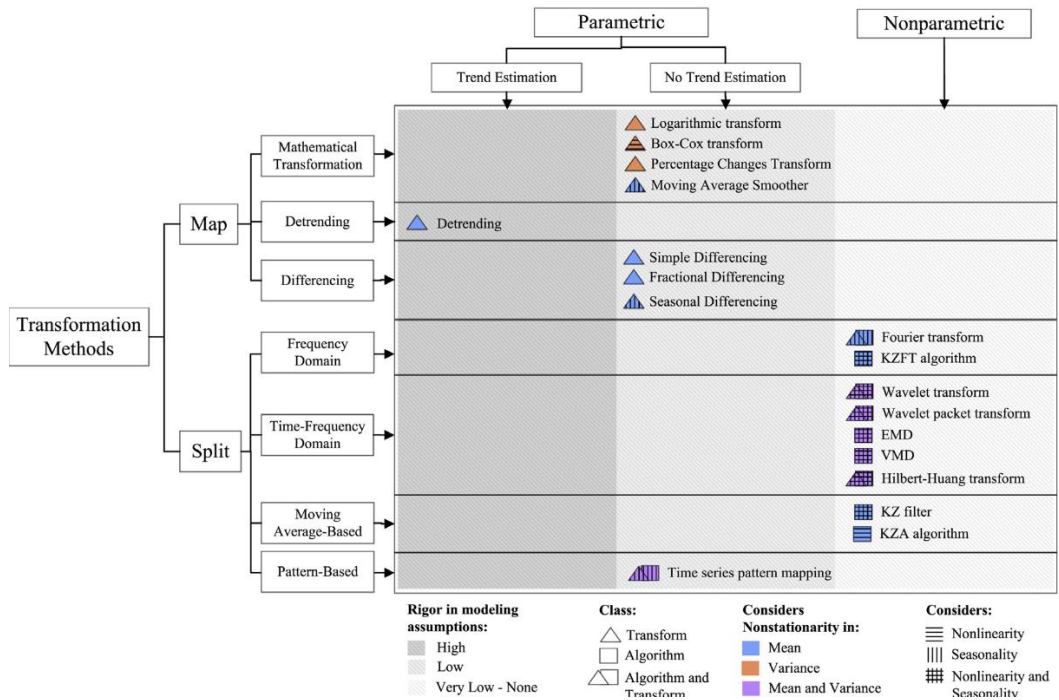


**Figure 5.1:** Categorization diagram for some of the most researched transformation methods for coping with non-stationarity in univariate time series prediction [69]

### 5.2.1 Methodological Perspective on Using Differencing for Stationarity

Stationarity is a very basic and important concept in time series analysis; it denotes a statistical property of the series, whose mean, variance, and autocovariance remain constant through time. This assumption forms one of the basic building blocks for many time series modeling techniques since stationarity allows the relationships within the data to be stable and predictable. Consequently, methods that ensure stationarity, such as detrending and differencing, are an essential part of data preparation while performing time series analysis.

An important aspect of time series analysis is the application of such transformations as detrending and simple differencing, which are often applied together with time series modeling techniques. This variance of transformation and used methods are shown in **Table 5.1**.

**Table 5.1:** The nonstationary time series transformation methods selected for the experiment are outlined, along with the data transformation processes they are applied to.

| Category        | Transformation  | Method                           | Ref   |
|-----------------|-----------------|----------------------------------|-------|
| Mapping-based   | Math. Transform | Logarithmic transform            | LT    |
|                 |                 | Logarithmic transform (base 10)  | LT10  |
|                 |                 | Box–Cox transform                | BCT   |
|                 |                 | Percentage changes transform     | PCT   |
|                 |                 | Moving average smoother          | MAS   |
|                 | Differencing    | Simple differencing              | DIF   |
|                 |                 | Seasonal differencing            | SDIF  |
|                 |                 | Simple and seasonal differencing | DIFs  |
| Splitting-based | Decomposition   | Wavelet transform                | WT    |
|                 |                 | Empirical mode decomposition     | EMD   |
| None            | None            | None                             | Naive |

Detrending DT transformation involves the identification and removal of an inherent trend observed in the behavior of a time series [71]. Traditionally, an estimate of the observed trend is modeled and represented by a deterministic functional form which then can be included as a fixed component in a time series model. It can be represented mathematically and takes the form of Eq. (10): where  $\eta_t$  represents the estimated trend component and  $\hat{x}_t$  represents residual variability series, which is essentially the time series data after removing the trend. While the DT approach is generally adopted, it usually assumes that the deterministic trend is time invariant. This assumption is not always valid, especially in the applications of nonlinear or nonstationary processes

[72]. In this context, recent developments have focused on data-driven and adaptive methods for trend estimation of complex, nonstationary time series [73].

$$\hat{x}_t = x_t - \eta_t \quad (5.1)$$

In several ways, the DIF methodology has some advantages over DT, especially in terms of simplicity and flexibility. Unlike DT, there is no estimation of parameters involved with DIF, which has the added advantage of always producing a stationary time series provided the original series exhibits stationarity around either a deterministic or stochastic trend [71]. DIF transformations can eliminate more complex trends, such as quadratic trends [71]. This sequential process provides a straightforward mechanism for addressing varying levels of non-stationarity [71].

With these considerations, this thesis will be employing the simple differencing method as the transformation technique. This reflects its robustness, ease of implementation, and capability to address stationarity, which is crucial for time series modeling.

Stationarity is one of the foundational concepts in time series analysis; as such, its importance cannot be overemphasized. In this section, some important reasons will be discussed as to why it is important to have stationarity in time series data. Stationary time series offer several significant advantages, particularly in the context of statistical analysis and econometric modeling. These advantages are foundational for researchers and practitioners dealing with time-dependent data and can be summarized as follows:

#### **5.2.1.1 Time Series Model Assumptions**

Most of the time series models, like ARIMA and SARIMA, are designed to use constant statistical features such as mean and variance across a period. The stationarity of series is a major assumption for using different time series models. If a series were to be non-stationary, the output from these models becomes unreliable, and their predictions also become a far cry from reality as these models cannot handle dynamic statistical features.

#### **5.2.1.2 Simplified Modeling and Forecasting**

Maintaining stationarity makes the inherent complexities of time series data simpler to model and forecast, compared to non-stationary time series. Consistent statistical

properties over time allow for the effective use of historical data in developing accurate models and forecasting future values.

#### **5.2.1.3 Interpretability of Trends and Patterns**

Trends and patterns are more interpretable in a stationary time series, since the relationships among data points remain constant. Constancy ensures that an observed trend or pattern does not become a function of random fluctuations; hence, the insights derived from it are more reliable.

#### **5.2.1.4 Improved diagnostic checks**

Improves the robustness of diagnostic checks in time series analyses. By checking for stationarity, an analyst often then finds some problem in the flow of data that may violate this usually necessary assumption and can confidently proceed with the analysis.

#### **5.2.1.5 Improved Model Performance**

Usually, a stationary time series means improved model performance. Since many statistical properties remain constant, models can more precisely learn from the dynamics behind the data and thus provide more accurate predictions.

### **5.3 Converting Non-stationary Data to Stationary**

This section converts the non-stationary time series data obtained from flight operations into a stationary form for better and more meaningful analysis, as described in previous sections. Most of the time series data in its raw form presents a trend, seasonality, or some other form that masks the underlying relationship and inhibits the application of statistical models that assume stationarity. Addressing non-stationarity, therefore, constitutes a very important pre-processing step in time series analysis, especially for an application domain such as aviation analysis, where precision and reliability are essential.

This can be achieved by differencing the technique so that the viewer may see meaningful patterns and correlations, perhaps masked by the presence of non-stationary features. This is important in-flight data analysis, given that insights derived

from stationary time series inform critical operational decisions, enhance predictive maintenance strategies, and generally improve efficiencies in aviation systems.

To understand this better, some examples of stationary parameters have been given below. It is those metrics that may have constant statistical properties on the time axis and may therefore act as the bases for further analysis. These data representations in stationarity also enable higher-order analytics capable of giving deeper insights about flight operation dynamics.

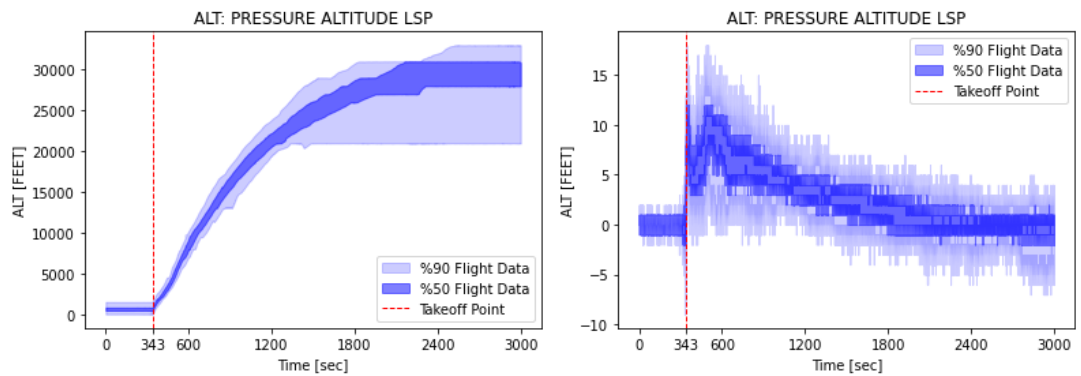


**Figure 5.2:** Non-Stationary and Stationary ALT parameters of Analyze Group

The point is that this example demonstrates the pattern-based anomalies are effectively detectable even in the case of time series data being transformed into a stationary form. However, it gives significantly better results for instantaneous anomalies, which have their nature in sudden deviations of the real behavior from the forecasted one at some points in time. This improvement in performance in detecting instantaneous anomalies could be because the reduced influence makes it more distinct between normal variations and anomaly events.

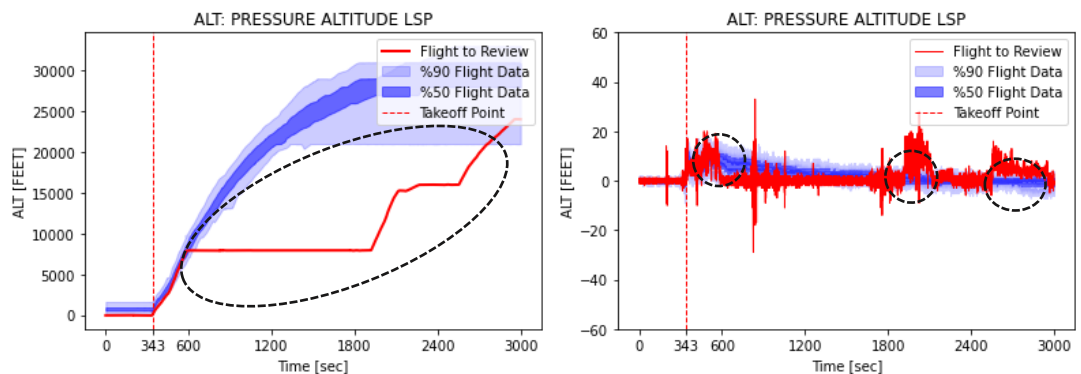


**Figure 5.3:** Pattern Anomaly Example

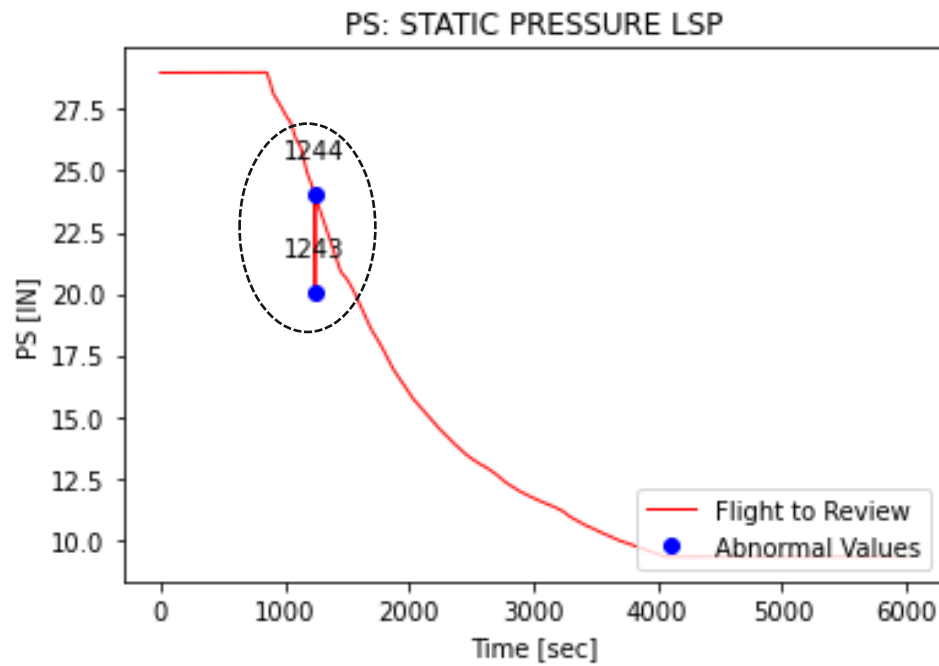


**Figure 5.4:** Static Pressure Anomaly in Real Data

After having confirmed that the data is stationary, a regression model could be devised to explain the behavior of the data over time. Any large discrepancies between the model's predicted values and the actual values may represent anomalies. This will show unusual patterns in the dataset, which might be from system errors, unusual events, or even data entry mistakes.

Performance evaluation of the regression model is a very vital process in anomaly detection. The regression method for anomaly detection allows the analysis of stationary data on how data are behaving with respect to time and what data are anomalies in the whole set. In so doing, it helps increase data quality for better reliable forecasts. Besides this, regression and anomaly detection contribute to insight into data and evidence-based decision-making in various fields of study.

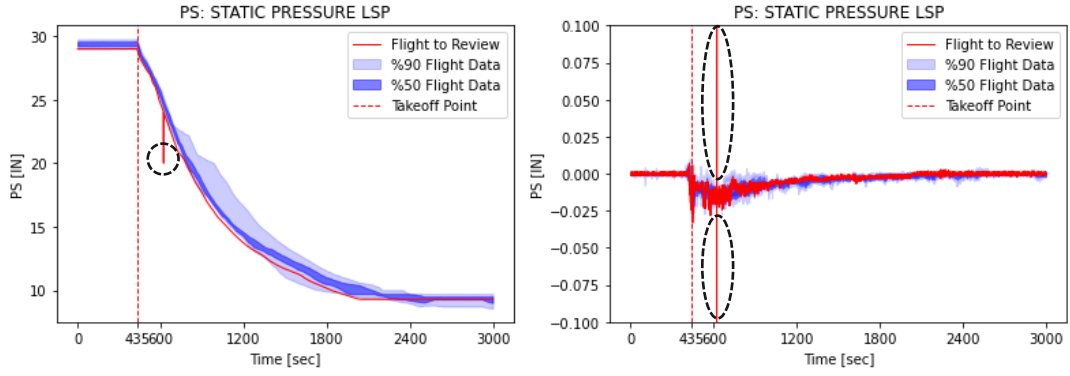


**Figure 5.5:** Instantaneous Anomaly Example

A detailed presentation of experimental results from the application of regression methods on a stationary time series with challenging scenarios of abnormal pitch excursions, high wind conditions, and situations that present sensor errors is done in the following section. These conditions were purposefully selected to assess the robustness and accuracy of the regression approach under real-world complexities. Such analysis investigates model performance under widely varied, stringent conditions and thereby points to its capability for handling irregularities, maintaining predictive accuracy, and adapting to outliers and noise in the data.

## 5.4 Experimental Results

### 5.4.1 High speed Abnormality (Flight Number: 653200103281828)

It can be seen from **Figure 5.6** that the aircraft speed just after take-off has a large deviation from the norm, or the expected operative pattern for similar conditions. This deviation implies that the performance of the aircraft in this stage should be examined to identify the underlying factors leading to the seen abnormality.

By means of such an approach to a deeper investigation, the method of stationary conversion was applied, which allowed an advanced analysis to be performed, at least about speed and angle of attack of flight phases identified as stationary.

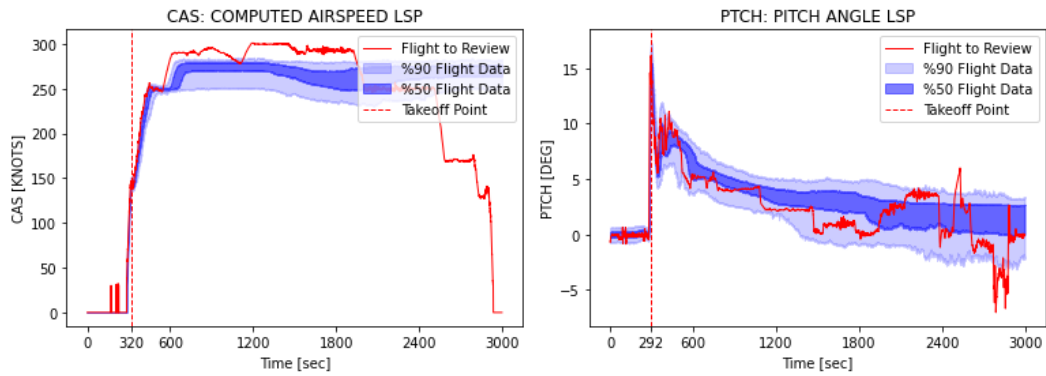


**Figure 5.6:** Non-Stationary Computed Airspeed and Pitch Angle plot of the Flight Number: 653200103281828

Some other outcomes of such a process are described by **Figure 5.7** and **Figure 5.8** and highlight net anomalies in aircraft behavior. Specifically, the stationary speed follows an unphysical oscillation with respect to the expected value of common pattern. The angle of attack is similarly irregular, which, in comparison, controverts the conventional aerodynamic profile. These findings make the suspect aircraft post-takeoff dynamics to be influenced by factors that may not align with established operational norms. Generally, such anomalies, if not addressed, squarely affect both performance and safety. These patterns identified further underline the need for an in-depth investigation into root causes, including design, operational, or environmental factors that might have contributed to such deviations.

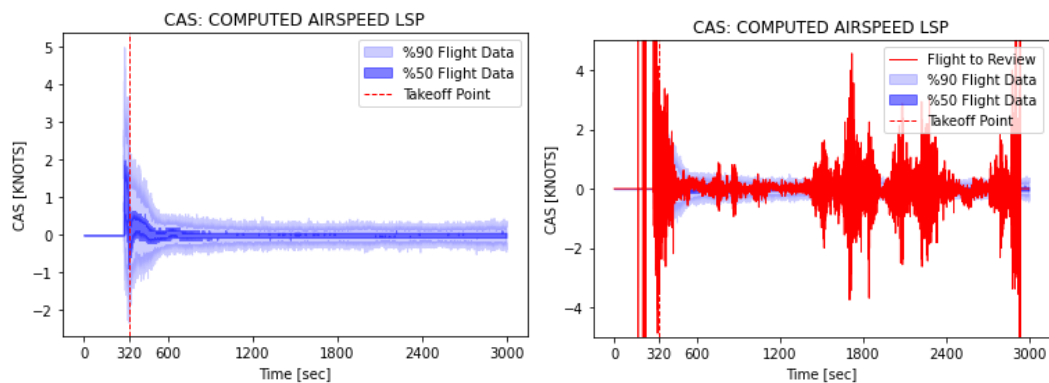


**Figure 5.7:** Stationary Common pattern and Reviewed Flight plot of the CAS

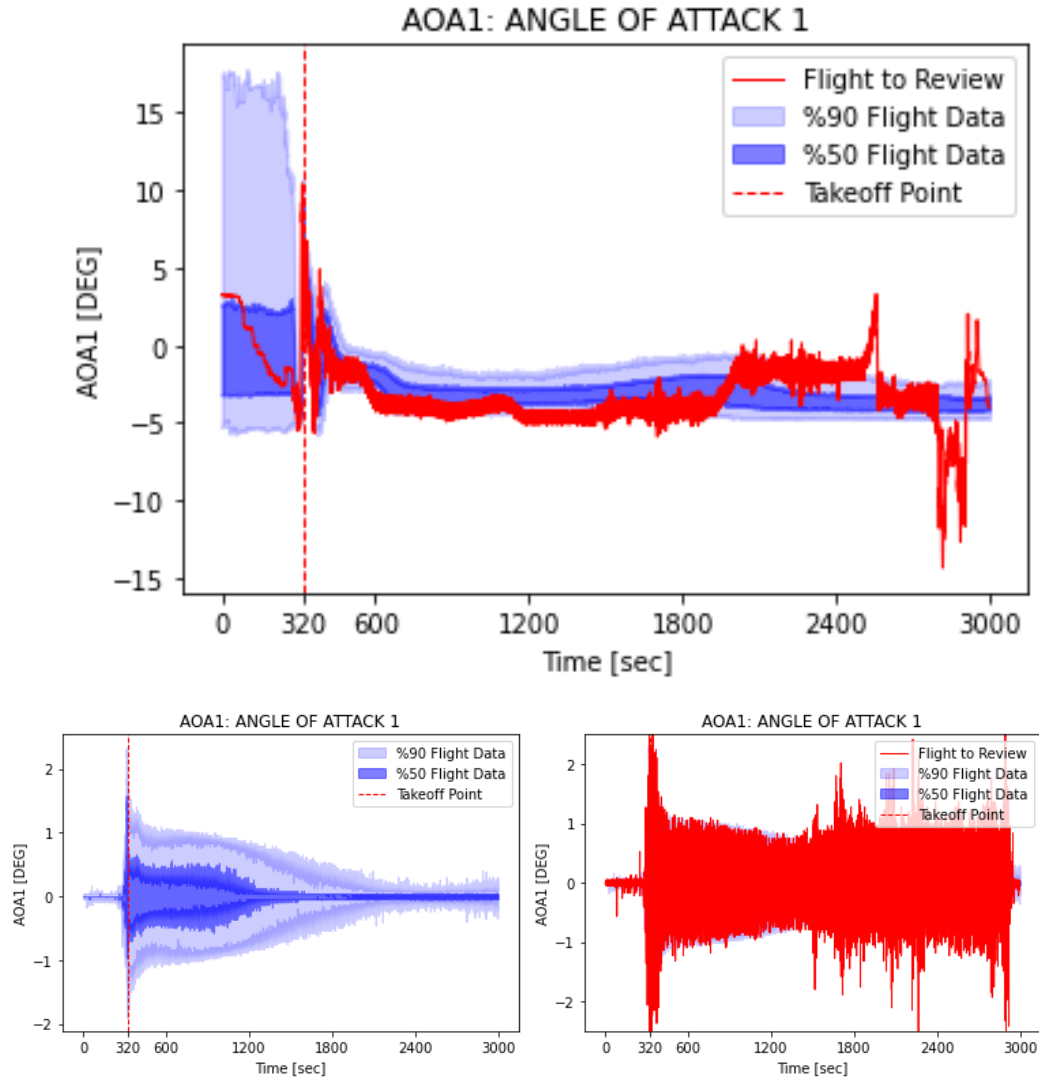


**Figure 5.8:** Angle of Attack Plot (Non-Stationary, Stationary Common Pattern, Stationary Reviewed Flight)

#### 5.4.2 High Wind Take-off Abnormality (Flight Number: 687200108021021)

This section shows another experimental result showing the oscillations in the CAS data as shown in the **Figure 5.10**. From these oscillations, it is possible to realize that the wind conditions during flight 687200108021021 were present. The stationary plot also shows some deviations in the flight data that have been successfully detected in this research study. It has been concluded that the whole process was carried out in highly windy conditions.

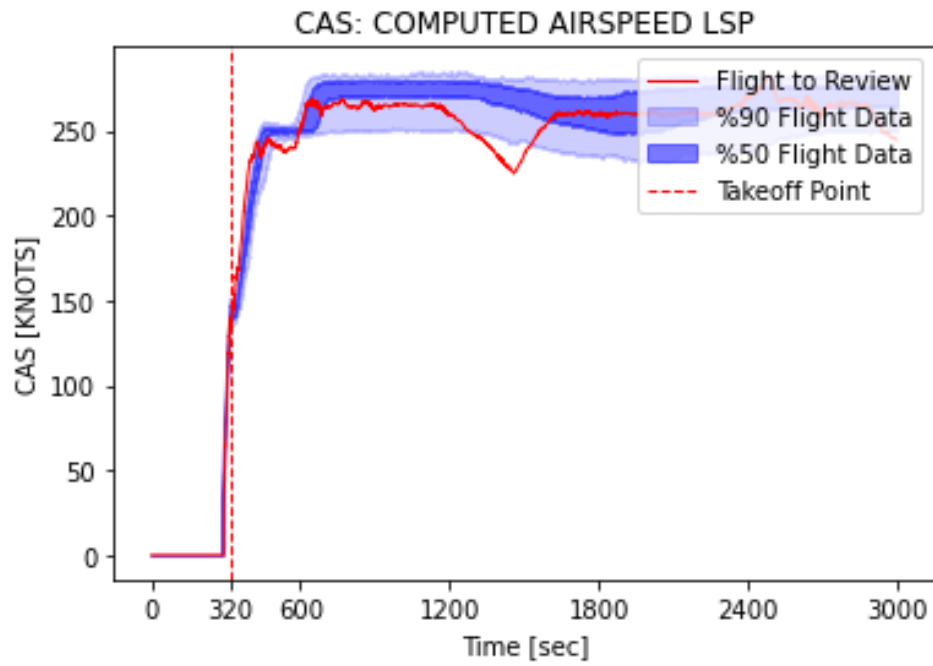


**Figure 5.9:** Non-Stationary Computed Airspeed of the Flight Number: 687200108021021

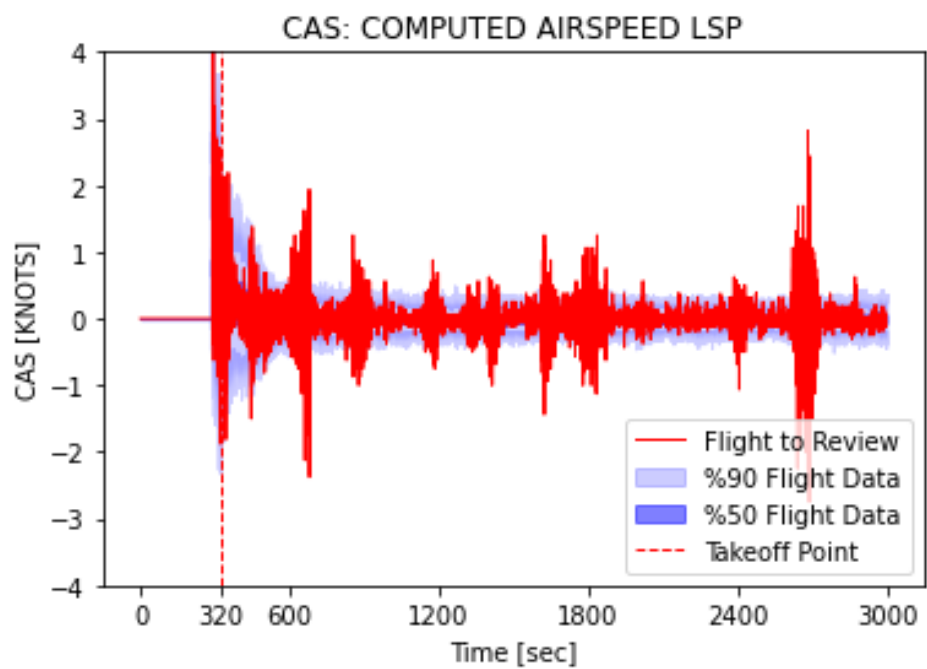


**Figure 5.10:** Stationary Computed Airspeed of the Flight Number: 687200108021021

Although wind conditions were not in focus when designing the analysis, and despite the fact that only the most basic plots, as depicted in **Figure 5.9**, were available for wind, the data turn out to be quite enlightening. The plots explain the anomalous oscillations picked up by most parameters, target airspeed **Figure 5.9**, rudder activity **Figure 5.11**.

This clearly demonstrates that the common pattern analysis-based model was unable to identify any events of interest for this flight. In contrast, the stationary model developed in this thesis effectively detected the anomaly, highlighting its robustness and superior sensitivity in identifying deviations from standard flight parameters.

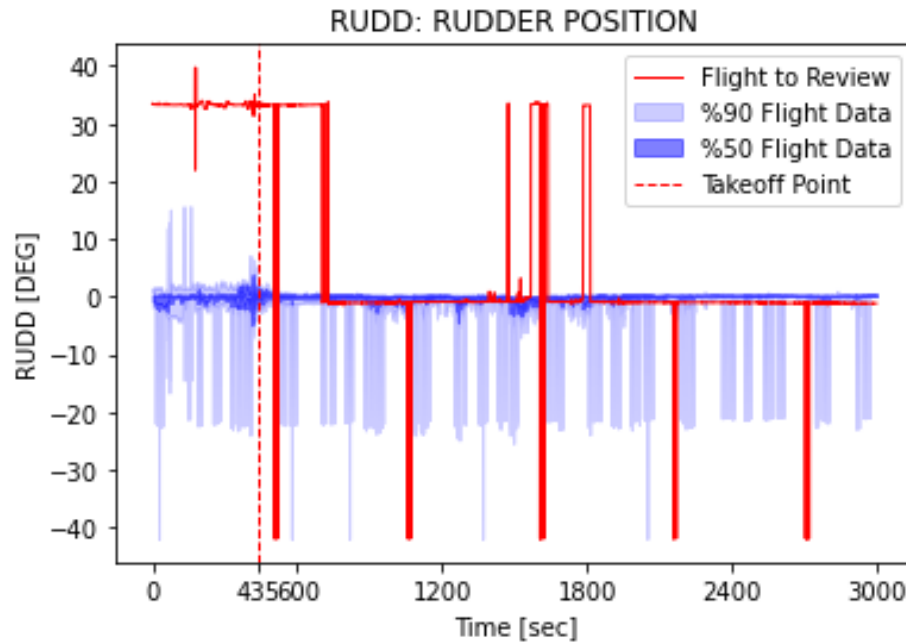


**Figure 5.11:** Non-Stationary Rudder Position of the Flight Number: 687200108021021

Furthermore, the wind direction and speed can be gauged from the following **Figure 5.12** proves the presence of strong winds during the operation. These results indicate that wind effects should not be excluded from anomaly detection models to allow for better detection performance regarding deviations from normal flight parameters as can be seen **Figure 5.10**.

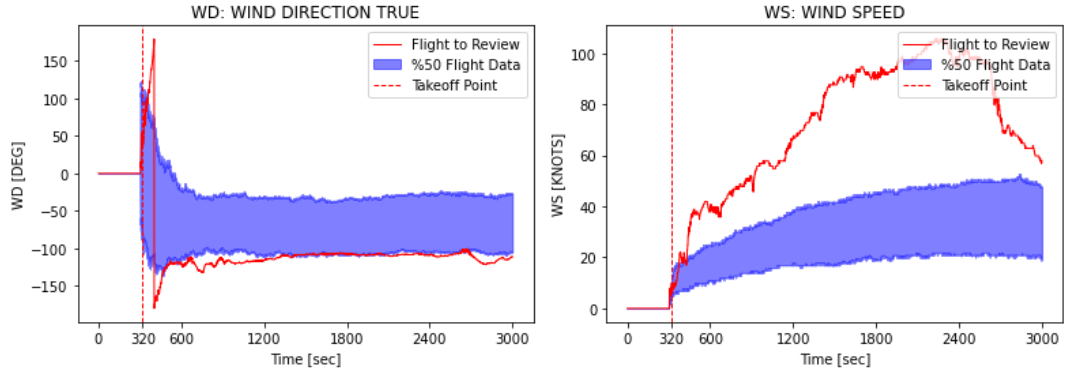


**Figure 5.12:** Non-Stationary Wind Direction and Speed of the Flight Number: 687200108021021

#### 5.4.3 Sensor Fault Abnormality (653200103210525)

The model developed in this thesis successfully identified an anomaly as can be seen in the **Figure 5.13** indicating a static pressure sensor failure. In particular, the deviations observed in the sensor data indicate that the device is not maintaining the expected measurement accuracy. This finding emphasizes that the sensor's performance during flight deviates from the standard behavior and indicates a possible failure. The results emphasize the sensitivity and reliability of the fixed model in detecting sensor anomalies. Based on the data provided in this study, airline companies can proactively implement troubleshooting measures to address increased sensor errors, thereby enhancing operational safety and reliability.

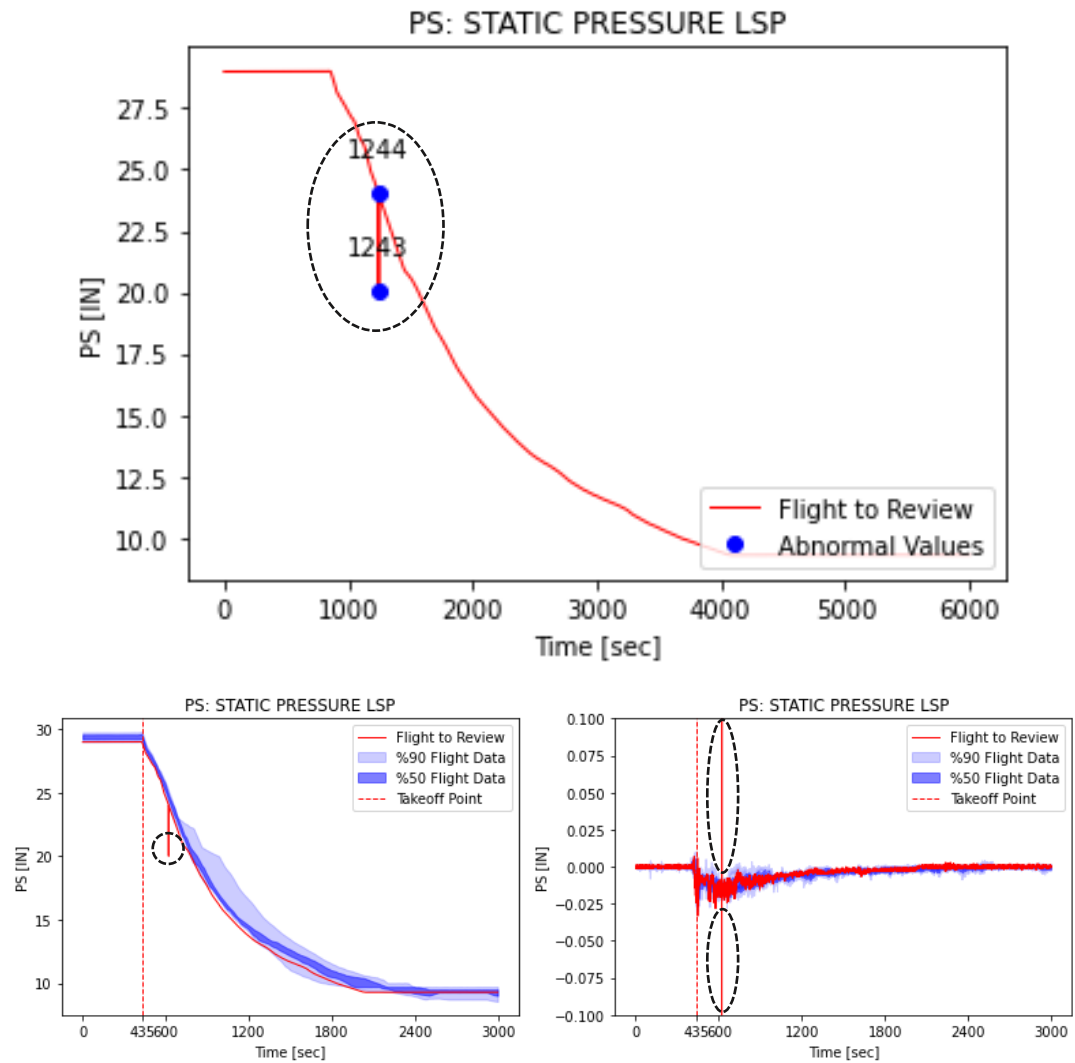


**Figure 5.13:** Non-Stationary and Stationary Static Pressure plot of the Flight Number: 653200103210525

## 5.5 Summary

In summary, the studied method was able to detect different type of functionally abnormal flights. These flights had anomalous parameter sequence and this were more probable to be considered functionally abnormal. There were fluctuations in the time-series of anomalous flight parameters. However, according to the standards these temporary anomalies were not considered as anomaly. In this study, both of anomaly detection method has been conducted, pattern based and instantaneous anomalies.

Two method for anomaly detection were presented to identify abnormal flight patterns without any predefined limits on parameters. These methods use different methods to preprocess raw flight data to a form that can be used for Pattern-Based and Instantaneous anomaly analysis. Prior to the application of these methods, the Common Pattern associated with the specific flight phase was identified and processed.

Common Pattern Analysis transforms real data of flights for given phases into representative patterns. The second approach is Stationary Analysis, which redefines the data in stationary time series, hence enabling statistical analysis. This chapter presented the methodology of the two methods and how to perform various tests using test datasets.

## **6. CONCLUSIONS AND FUTURE WORK**

### **6.1 Conclusions**

This thesis has presented a new technique for detecting anomalies that can help improve proactive safety management by making use of the everyday operational FDR data of airlines. The methodology involves two major steps: identification of flights with abnormal data patterns in a huge dataset of daily flights and subsequent evaluation of these abnormal flights for their operational significance. This research, therefore, demonstrates the potential of this approach in effectively detecting both pattern-based and instantaneous anomalies by focusing on the analysis of FDR data as a stationary time series.

In particular, the transformation of non-stationary FDR data to a stationary format has been found to be a promising method for unmasking small and critical deviations from normal operation conditions. Enabling this approach to catch both periodic abnormal patterns and abrupt deviations opens up a wide spectrum of anomalies, hence turning it into an asset in efforts to improve aviation safety.

This innovative approach entrenches the airlines' potentials as an effective, onetime early detection for any form of safety degradations, latent risk unravelling, dealing proactively with them before anything critical goes out of proportion. These will further be contributive in deploying predictive strategies regarding maintenance, operational decisions, and targeted training programs associated with the airlines. This thesis represents a contribution to the overall objective of improving aviation safety using data-driven approaches, and it forms the groundwork for further research and practical applications in anomaly detection and safety management.

### **6.2 Recommendations for Future Work**

This thesis has elaborated an approach which is not in practical use in current airline operations and requires further investigation before being integrated with existing FOQA programs. For the approach to be feasible in practice, support tool development

will be required. At the time of writing, the method covers only the takeoff phase for which it was tested. Anchoring raw FDR data and standardization of the patterns will be indispensable when its application is extended to other flight phases due to their high diversity in temporal behaviors.

Further studies can also be conducted based on other data sources, such as maintenance logs, pilot reports, and weather data, complementary to or independent of the FDR records. The integration of those sources would raise the anomaly detection method to a diagnostic system. This could identify not just the operational significant anomalies but also their cause. These advancements would improve safety management by enabling more precise detection, predictive maintenance, and targeted operational improvements.

## REFERENCES

- [1] J. M. Walters, “The pilot proficiency audit of knowledge, skills, and abilities: Adding to the air carrier safety toolbox,” *Embry-Riddle Aeronaut. Univ. Dayt. Beach FL*, 2004.
- [2] R. H. Wood and R. W. Sweginnis, “Aircraft accident investigation,” *Endeavor Books*, p. 516, 2006.
- [3] A. J Stolzer, S. Robert L, and J. J Goglia, *Safety Management Systems in Aviation*, 3rd ed. Boca Raton: CRC Press, 2023.
- [4] H. Qin, H. Wang, J. Xie, Y. Tang, and P. Li, “Development of flight data recorder for general aviation aircraft with real time data transmission,” in *2021 IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, Oct. 2021, pp. 162–167. doi: 10.1109/ICCASIT53235.2021.9633449.
- [5] G. Leigh, “What is a black box and how does it work?,” Flightradar24 Blog. Accessed: Aug. 14, 2024. [Online]. Available: <https://www.flightradar24.com/blog/what-is-a-black-box-and-how-does-it-work>
- [6] A. R. Morcom, “Flight Data Recording Systems,” *Aircr. Eng. Aerosp. Technol.*, vol. 42, no. 7, pp. 12–16, Jan. 1970, doi: 10.1108/eb034655.
- [7] “Federal Aviation Administration, 14 CFR Part 121— Operating Requirements: Domestic, Flag and Supplemental Operations.” FAA.
- [8] W. Rosenkrans, “fade-free Memory,” *Aerosafe World January*, pp. 47–48, 2008.
- [9] L. Li, “Anomaly detection in airline routine operations using flight data recorder data,” PhD Thesis, Massachusetts Institute of Technology, 2013. Accessed: Aug. 15, 2024. [Online]. Available: <https://dspace.mit.edu/handle/1721.1/82498>
- [10] N. AH Campbell, “MO3806, The Evolution of Flight Data Analysis.”.
- [11] G. M. L. Tsuruta, “The Analysis of Flight Operational Quality Assurance (FOQA) Data: Exploration of a Proposed List of Improved Safety Parameters”.
- [12] “Federal Aviation Administration, Advisory Circular. 120-82 Flight Operational Quality Assurance , Washington DC.” 2004.
- [13] N. Campbell, “Flight Data Analysis - An Airline Perspective,” Australian and New Zealand Societies of Air Safety Investigators Conference (ANZSASI) Seminar, Jun. 2003.
- [14] P. DUBOIS, “Flight Data Analysis,” ICAO, 2015.

- [15] “GroundLink® Comm+ System.” Accessed: Sep. 11, 2024. [Online]. Available: <https://www.teledynecontrols.com/products/hardware-systems/groundlink-comm+>
- [16] van E. G.W.H., “Advanced Flight Data Analysis,” *14th Eur. Aviat. Saf. Semin. Bp. Hung. March 11–13 2002*.
- [17] S. Jasra, J. Gauci, A. Muscat, G. Valentino, D. Zammit-Mangion, and R. Camilleri, “Literature review of machine learning techniques to analyse flight data,” 2018, Accessed: Nov. 04, 2024. [Online]. Available: <https://www.um.edu.mt/library/oar/handle/123456789/58839>
- [18] “Flight Data Analytics launches Insight Accelerator| Boeing Services.” Accessed: Nov. 04, 2024. [Online]. Available: <https://services.boeing.com/news/flight-data-analytics-insight-accelerator-foqa-fdm>
- [19] *Guide to Methods & Tools for Airline Flight Safety Analysis, Issue 2*. GAIN Working Group B, Analytical Methods and Tools, 2003.
- [20] J.-J. Speyer, “The Flight Operations Monitoring System: A Bundled Approach for Synergistic Safety Management”.
- [21] V. Chandola and V. Kumar, “Outlier Detection : A Survey,” *ACM Comput. Surv.*, vol. 41, Jan. 2009.
- [22] V. Hodge and J. Austin, “A Survey of Outlier Detection Methodologies,” *Artif. Intell. Rev.*, vol. 22, no. 2, pp. 85–126, Oct. 2004, doi: 10.1023/B:AIRE.0000045502.10941.a9.
- [23] S. A. Hofmeyr, S. Forrest, and A. Somayaji, “Intrusion detection using sequences of system calls,” *J. Comput. Secur.*, vol. 6, no. 3, pp. 151–180, Jan. 1998, doi: 10.3233/JCS-980109.
- [24] D. Endler, “Intrusion detection. Applying machine learning to Solaris audit data,” in *Proceedings 14th Annual Computer Security Applications Conference (Cat. No.98EX217)*, Dec. 1998, pp. 268–279. doi: 10.1109/CSAC.1998.738647.
- [25] M. J. Desforges, P. J. Jacob, and J. E. Cooper, “Applications of probability density estimation to the detection of abnormal conditions in engineering Proceedings of the Institution of Mechanical Engineers, Part C,” *J. Mech. Eng. Sci.*, vol. 212, no. 8, pp. 687–703, 1998.
- [26] A. Smiti, “A critical overview of outlier detection methods,” *Comput. Sci. Rev.*, vol. 38, p. 100306, Nov. 2020, doi: 10.1016/j.cosrev.2020.100306.
- [27] B. Tang, H. He, P. M. Baggenstoss, and S. Kay, “A Bayesian Classification Approach Using Class-Specific Features for Text Categorization,” *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 6, pp. 1602–1606, Jun. 2016, doi: 10.1109/TKDE.2016.2522427.

- [28] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.
- [29] M. Ankerst, C. Elsen, M. Ester, and H.-P. Kriegel, "Visual classification: an interactive approach to decision tree construction," in *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, San Diego California USA: ACM, Aug. 1999, pp. 392–396. doi: 10.1145/312129.312298.
- [30] W.-Y. Yoh, "Classification and regression trees WIREs Data Mining and Knowledge Discovery. <https://doi.org/10.1002/widm.8>," 2011, Accessed: Nov. 11, 2024. [Online]. Available: [https://wires.onlinelibrary.wiley.com/doi/full/10.1002/widm.8?saml\\_referrer](https://wires.onlinelibrary.wiley.com/doi/full/10.1002/widm.8?saml_referrer)
- [31] D. M. J. Tax and R. P. W. Duin, "Support vector domain description," *Pattern Recognit. Lett.*, vol. 20, no. 11, pp. 1191–1199, Nov. 1999, doi: 10.1016/S0167-8655(99)00087-2.
- [32] J. B. D. Cabrera, L. Lewis, and R. K. Mehra, "Detection and classification of intrusions and faults using sequences of system calls," *SIGMOD Rec*, vol. 30, no. 4, pp. 25–34, Aralık 2001, doi: 10.1145/604264.604269.
- [33] N. Lavrac, S. Miksch, and B. Kavsekforeword, "5th International Workshop on Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP-2000)," Sep. 2000.
- [34] G. Frank E., "Procedures for Detecting Outlying Observations in Samples: Technometrics," vol. 11, no. 1, pp. 1–21, February.
- [35] W. Stefansky, "Rejecting Outliers in Factorial Designs: Technometrics," vol. 14, no. 2, pp. 469–479.
- [36] B. Rosner, "Percentage Points for a Generalized ESD Many-Outlier Procedure: Technometrics," vol. 25, no. 2, pp. 165–172.
- [37] R. D. Gibbons, D. K. Bhaumik, and S. Aryal, *Statistical Methods for Groundwater Monitoring*. John Wiley & Sons, 2009.
- [38] P. Anandhi and Dr. E. Nathiya, "Application of linear regression with their advantages, disadvantages, assumption and limitations," *Int. J. Stat. Appl. Math.*, vol. 8, no. 6, pp. 133–137, Nov. 2023, doi: 10.22271/math.2023.v8.i6b.1463.
- [39] L. Rokach and O. Maimon, "Clustering Methods," in *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds., Boston, MA: Springer US, 2005, pp. 321–352. doi: 10.1007/0-387-25465-X\_15.
- [40] M. Hu, X. Feng, Z. Ji, K. Yan, and S. Zhou, "A novel computational approach for discord search with local recurrence rates in multivariate time series," *Inf. Sci.*, vol. 477, pp. 220–233, Mar. 2019, doi: 10.1016/j.ins.2018.10.047.

- [41] K. Choi, J. Yi, C. Park, and S. Yoon, “Deep Learning for Anomaly Detection in Time-Series Data: Review, Analysis, and Guidelines,” *IEEE Access*, vol. 9, pp. 120043–120065, 2021, doi: 10.1109/ACCESS.2021.3107975.
- [42] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*. John Wiley & Sons, 2005.
- [43] A. J. Fox, “Outliers in Time Series,” *J. R. Stat. Soc. Ser. B Methodol.*, vol. 34, no. 3, pp. 350–363, Jul. 1972, doi: 10.1111/j.2517-6161.1972.tb00912.x.
- [44] B. Abraham and A. Chuang, “Outlier Detection and Time Series Modeling,” *Technometrics*, vol. 31, no. 2, pp. 241–248, 1989, doi: 10.2307/1268821.
- [45] A. M. Bianco, M. G. Ben, E. J. Martínez, and V. J. Yohai, “Outlier Detection in Regression Models with ARIMA Errors using Robust Estimates. | EBSCOhost.” Accessed: Nov. 11, 2024. [Online]. Available: <https://openurl.ebsco.com/contentitem/doi:10.1002%2Ffor.768?sid=ebsco:plink:crawler&id=ebsco:doi:10.1002%2Ffor.768>
- [46] B. G. Amidan and T. A. Ferryman, “APMS SVD methodology and implementation,” Pacific Northwest National Lab. (PNNL), Richland, WA (United States), PNWD-3026, Apr. 2000. doi: 10.2172/753847.
- [47] B. G. Amidan and T. A. Ferryman, “Atypical event and typical pattern detection within complex systems,” in *2005 IEEE Aerospace Conference*, Mar. 2005, pp. 3620–3631. doi: 10.1109/AERO.2005.1559667.
- [48] S. Budalakoti, A. N. Srivastava, and R. Akella, “Discovering Atypical Flights in Sequences of Discrete Flight Parameters,” in *2006 IEEE Aerospace Conference*, Mar. 2006, pp. 1–8. doi: 10.1109/AERO.2006.1656109.
- [49] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, “Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding,” in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, in KDD '18. New York, NY, USA: Association for Computing Machinery, Temmuz 2018, pp. 387–395. doi: 10.1145/3219819.3219845.
- [50] M. Schwabacher, N. Oza, and B. Matthews, “Unsupervised Anomaly Detection for Liquid-Fueled Rocket Propulsion Health Monitoring,” *J. Aerosp. Comput. Inf. Commun.*, vol. 6, no. 7, pp. 464–482, 2009, doi: 10.2514/1.42783.
- [51] S. Das, B. Matthews, A. Srivastava, and N. Oza, “Multiple Kernel Learning for Heterogeneous Anomaly Detection: Algorithm and Aviation Safety Case Study”.
- [52] S. Budalakoti, A. N. Srivastava, and M. E. Otey, “Anomaly Detection and Diagnosis Algorithms for Discrete Symbol Sequences with Applications to Airline Safety,” *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 39, no. 1, pp. 101–113, Jan. 2009, doi: 10.1109/TSMCC.2008.2007248.

- [53] D. L. Iverson, "INDUCTIVE SYSTEM HEALTH MONITORING WITH STATISTICAL METRICS," Jan. 01, 2005. Accessed: Sep. 06, 2024. [Online]. Available: <https://ntrs.nasa.gov/citations/20050240965>
- [54] D. L. Iverson, "Inductive System Health Monitoring," Jan. 01, 2004. Accessed: Sep. 06, 2024. [Online]. Available: <https://ntrs.nasa.gov/citations/20040068062>
- [55] B. Matthews, "NASA DASH link - Multiple Kernel Learning based Heterogeneous Algorithm (MKAD)." Accessed: Sep. 06, 2024. [Online]. Available: <https://c3.ndc.nasa.gov/dashlink/projects/34/>
- [56] H. J. Shin, D.-H. Eom, and S.-S. Kim, "One-class support vector machines—an application in machine fault detection and classification," *Comput. Ind. Eng.*, vol. 48, no. 2, pp. 395–408, Mar. 2005, doi: 10.1016/j.cie.2005.01.009.
- [57] M. Ester, H.-P. Kriegel, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise".
- [58] C.-H. Lee, H.-S. Shin, A. Tsourdos, and Z. Skaf, "Data analytics development of FDR (Flight Data Recorder) data for airline maintenance operations," in *2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, Daegu: IEEE, Nov. 2017, pp. 289–294. doi: 10.1109/MFI.2017.8170443.
- [59] S. Wang, C. Li, and A. Lim, "A Model for Non-Stationary Time Series and its Applications in Filtering and Anomaly Detection," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021, doi: 10.1109/TIM.2021.3059321.
- [60] C. Cheng *et al.*, "Time series forecasting for nonlinear and non-stationary processes: A review and comparative study," *IIE Trans. Inst. Ind. Eng.*, vol. 47, no. 10, pp. 1053–1071, 2015, doi: 10.1080/0740817X.2014.999180.
- [61] E. R. Rosca, "STATIONARY AND NON-STATIONARY TIME SERIES," *Ann. Stefan Cel Mare Univ. Suceava Fascicle Fac. Econ. Public Adm.*, vol. 10, no. 1(11), pp. 177–186, 2010.
- [62] B. Matthews, "NASA DASH link - Resources." Accessed: Nov. 14, 2024. [Online]. Available: <https://c3.ndc.nasa.gov/dashlink/resources/>
- [63] K. Pearson and O. M. F. E. Henrici, "VII. Mathematical contributions to the theory of evolution.—III. Regression, heredity, and panmixia," *Philos. Trans. R. Soc. Lond. Ser. Contain. Pap. Math. Phys. Character*, vol. 187, pp. 253–318, Jan. 1997, doi: 10.1098/rsta.1896.0007.
- [64] J. L. Rodgers and W. A. Nicewander, "Thirteen Ways to Look at the Correlation Coefficient," *Am. Stat.*, vol. 42, no. 1, pp. 59–66, 1988, doi: 10.2307/2685263.

- [65] S. K. Tyagi, “Correlation coefficient of dual hesitant fuzzy sets and its applications,” *Appl. Math. Model.*, vol. 39, no. 22, pp. 7082–7092, Nov. 2015, doi: 10.1016/j.apm.2015.02.046.
- [66] D. Pavanello, W. Zaaiman, A. Colli, J. Heiser, and S. Smith, “Statistical functions and relevant correlation coefficients of clearness index,” *J. Atmospheric Sol.-Terr. Phys.*, vol. 130–131, pp. 142–150, Aug. 2015, doi: 10.1016/j.jastp.2015.05.012.
- [67] Y. Kim, T.-H. Kim, and T. Ergün, “The instability of the Pearson correlation coefficient in the presence of coincidental outliers,” *Finance Res. Lett.*, vol. 13, pp. 243–257, May 2015, doi: 10.1016/j.frl.2014.12.005.
- [68] M.-T. Puth, M. Neuhäuser, and G. D. Ruxton, “Effective use of Pearson’s product–moment correlation coefficient,” *Anim. Behav.*, vol. 93, pp. 183–189, Jul. 2014, doi: 10.1016/j.anbehav.2014.05.003.
- [69] R. Salles, K. Belloze, F. Porto, P. H. Gonzalez, and E. Ogasawara, “Nonstationary time series transformation methods: An experimental review,” *Knowl.-Based Syst.*, vol. 164, pp. 274–291, Jan. 2019, doi: 10.1016/j.knosys.2018.10.041.
- [70] D. N. Gujarati, D. C. Porter, and S. Gunasekar, *Basics Econometrics*. Higher Education, 2016. Accessed: Dec. 18, 2024. [Online]. Available: <http://macl-ustm.digitallibrary.co.in/handle/123456789/4509>
- [71] R. Shumway and D. Stoffer, *Time Series: A Data Analysis Approach Using R*. New York: Chapman and Hall/CRC, 2019. doi: 10.1201/9780429273285.
- [72] J. D. Cryer and K.-S. Chan, *Time Series Analysis*. in Springer Texts in Statistics. New York, NY: Springer, 2008. doi: 10.1007/978-0-387-75959-3.
- [73] Z. Wu, N. E. Huang, S. R. Long, and C.-K. Peng, “On the trend, detrending, and variability of nonlinear and nonstationary time series,” *Proc. Natl. Acad. Sci.*, vol. 104, no. 38, pp. 14889–14894, Sep. 2007, doi: 10.1073/pnas.0701020104.

## **BIOGRAPHY**

Mustafa Burak AKINCILAR completed his undergraduate education at Yıldız Technical University, graduating from the Faculty of Electrical and Electronics Engineering, Department of Electrical Engineering. Between 2018 and 2020, he worked as a Research and Development Engineer at General Mobile. Since 2020, he has been working as an Avionics Engineer at Turkish Airlines Technic.