A METHODOLOGY FOR QUANTITATIVE DATA-DRIVEN SAFETY ASSESSMENT FOR GENERAL AVIATION

A Thesis
Presented to
The Academic Faculty

by

Tejas Puranik

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Aerospace Engineering

Georgia Institute of Technology
May 2018

A METHODOLOGY FOR QUANTITATIVE DATA-DRIVEN SAFETY ASSESSMENT FOR GENERAL AVIATION

Approved by:

Professor Dimitri Mavris, Advisor
School of Aerospace Engineering
Georgia Institute of Technology

Professor Karen Marais
Department of Aeronautics and Astronautics
Purdue University

Dr. Simon Briceno
School of Aerospace Engineering
Georgia Institute of Technology

Dr. Hossein Eghbali
System Safety Section
Federal Aviation Administration

Professor Daniel Schrage
School of Aerospace Engineering
Georgia Institute of Technology

Date Approved: March 30th 2018

To my family,

ACKNOWLEDGEMENTS

As my journey through graduate school draws to a close, I would like to acknowledge all those who made this a great experience. First of all I would like to thank my adviser Prof. Dimitri Mavris for all that he has done for me throughout graduate school. His constant support, guidance, and advice has enabled me to gain a lot from this wonderful experience. I would like to thank him for the freedom and stability that he has afforded me all these years to pursue my own path. His mentoring and insights have proved valuable to me in all walks of life. I would like to express my profound appreciation for the members of my defense committee: Dr. Simon Briceno, Prof. Karen Marais, Dr. Hossein Eghbali, and Prof. Daniel Schrage for taking the time out of their busy schedules to serve on my committee. Their feedback and comments have proved valuable in steering me in the right direction.

The graduate school experience is not merely about the research but also about the friends and connections you make on this long, arduous journey. Therefore, I would like to thank all my friends from ASDL for some memorable experiences. My mentor, the late Hernando Jimenez who showed me the way from early on and inspired me to constantly push myself. Imon Chakraborty for inspiring me with his passion and scientific rigor and to pursue a pilot's license. Evan Harrison and Sanggyu Min for being great research buddies. Brett Hiller and Etienne Demers-Bouchard for our thesis group and being a sounding board for ideas and feedback. Thanks to my close friends Dushhyanth Rajaram and Mayank Bendarkar for our long endless chats on so many topics and the great times together. And my heartfelt thanks to so many others from the lab and the Aerospace Engineering department with whom I've had positive encounters over the years.

I have had the opportunity to make some wonderful friends outside of graduate school and strengthened old friendships during my time here. Thanks to my friends Neil Karandikar, Achyut Panchal, Pushkar Godbole, and others for all our fun times together. Thanks to Ketan Patwardhan, among my oldest and best friends and to Raunak Bhattacharyya for being a great roommate and close friend through all our graduate school experiences together.

I want to thank all my friends and family back home for their support and encouragement. Whatever I am today is because of my parents Girish and Medha Puranik. I want to thank them for their support and pride in all my achievements. Thanks to my parents-in-law, Pramod and Vandana Joglekar for their constant encouragement. I would like to thank my brother Harshad for being a role model of hard work and dedication. I would also like to thank my sisters-in-law Anoushka and Shivangi for their encouragement. And finally, thanks to the most important person in my life, my wife, Avni. She has been my constant source of happiness and unconditional love. From being a patient listener in times of stress to inspiring me to go above and beyond in all my endeavors. Her presence by my side throughout this process has made it a lot easier.

The work presented in this dissertation was sponsored by the Federal Aviation Administration (FAA) under project number 5 of the FAA Center of Excellence for General Aviation, Partnership to Enhance General Aviation Safety Accessibility and Sustainability, via grant number 12-C-GA-GIT-004 to the Georgia Institute of Technology under the supervision of technical monitor Michael Vu. The information in this research does not constitute FAA Flight Standards or FAA Aircraft Certification policy.

Tejas Puranik

Atlanta, GA

March, 2018

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The safety record of aviation operations has been steadily improving for the past few decades, however, accident rates in General Aviation (GA) have not improved significantly compared to scheduled commercial airline operations. According to the Federal Aviation Administration (FAA), the demand for air travel and traffic is predicted to grow steadily for the next two decades at a rate of approximately 1.8% annually with GA set to receive a much-needed revitalization. However, safety remains a major hurdle and with such a large increase in expected operations, there is an ever-increasing demand for improving safety of GA operations.

Various data-driven safety programs such as Flight Data Monitoring (FDM) that exist in commercial aviation domain have percolated in GA with the aim of improving safety. These programs typically feature a continuous cycle involving data collection from on-board recorders, retrospective analysis of flight data records, identification of operational safety exceedances, design and implementation of corrective measures, and monitoring to assess their effectiveness. While these programs have been shown to be effective in reducing accident rates, there are certain obstacles in their widespread implementation in the GA domain. The variability in recorded parameters in GA flight data recorders (FDR), heterogeneity in GA fleet, different missions flown, etc. are some of the important hurdles. Additionally, existing techniques of analysis such as exceedance detection are designed to identify known unsafe conditions but are potentially blind to safety-critical conditions that may be captured in flight data records but are not present in the set of predefined safety events. With the availability of recorded data in the GA domain, there is an opportunity to improve safety through

the use of more quantitative techniques.

The overarching objective of this dissertation is to develop a methodology that can provide objective metrics for quantifying GA flight safety, enable automatic identification of anomalous operations, and provide predictive capabilities that will complement existing approaches. The use of recorded flight data from GA operations is central to developing algorithms that are robust and applicable to the heterogeneous GA domain. The first objective of this dissertation is to obtain objective metrics for quantifying flight safety in GA operations. The development of metrics is pursued within the constraints imposed by GA FDR and through the examination of past criteria and historical data for identifying important parameters available for defining metrics. The dissertation presents the use of energy-based metrics as objective currency that can be used for quantifying flight safety across the heterogeneous GA fleet. These metrics satisfy some of the important criteria that are desired in metrics - parsimony, safety-relevance, and generalizability. The second objective of the dissertation is enabling automatic identification of anomalous operations. In order to facilitate this, an anomaly detection framework is developed using the defined safety metrics for identifying different types of anomalies (flight-level and instantaneous) in GA operations. The same general framework is adapted to identify both types of anomalous operations and understand their relationship with each other. The third objective of the dissertation is to provide predictive capabilities to improve the quality of the safety assessment task. To that end, models of aerodynamic and propulsion performance are utilized for obtaining unrecorded quantities of interest. A novel technique of calibrating these aircraft performance models starting from a generic GA model is developed. Different options for calibration depending on the type of calibration data available are proposed and tested to be applicable in multiple scenarios.

The completion of the research effort yielded a methodology that encapsulates the

essence of all three research objectives and provides a platform for retrospective safety assessment. The application of this methodology in a case study on real-world flight data enabled obtaining safety insights into the operations from the data set. The ability to define flexible thresholds for identifying abnormal operations from within any given data set reduces the analysis time and provides specific flights or operations for subject-matter-experts to focus on. This accelerates the first step in the retrospective safety analysis process. Thereafter, the review of the identified anomalies by subject-matter experts or safety analysts will complete the safety assessment process and result in conclusions regarding unsafe events. Similarly, the methodology is equally applicable for training purposes to demonstrate student pilots what constitutes significant deviations from nominal operations and how to proactively prevent being in potentially unsafe situations.

CHAPTER I

MOTIVATION

1.1 What is General Aviation?

According to the Federal Aviation Administration [55], General Aviation (GA) is

"*That portion of civil aviation that does not include scheduled or unscheduled air carriers or commercial space operations*".

Similarly, the International Civil Aviation Organization (ICAO) [76] defines GA as - *"all civil aviation operations other than scheduled air services and non-scheduled air transport operations for remuneration or hire"*. GA is thus used as a catch-all term for any operations that do not fall under commercial operations, major military, or cargo operations. However, this encompasses a broad range of diverse applications including but not limited to overnight package delivery, emergency medical evacuation, agricultural applications such as crop dusting, or flying for business, leisure, sports etc. Other than diversity in intended application, GA also exhibits significant diversity in the types of vehicles. The heterogeneity in fleet is evident from the fact that GA includes everything from light sport and experimental aircraft, single and multi-engine fixed wing aircraft, as well as helicopters and other rotorcraft. GA has a sizable impact on the economy as well as the aviation community.

According to recent (2017) statistics the US has one of the largest and most diverse GA communities and flies more than 20 million hours each year. This includes more than 90% of the roughly 220,000 civil aircraft registered in the United States [15]. About 560 U.S airports are certified for scheduled airline service (for airplanes with seating capacity for more than 30 passengers) whereas GA can access all 19,600 public and private landing facilities in the United States. The fleet of GA aircraft

Figure 1: Typical GA aircraft. (Source: http://www.pilotfriend.com/)

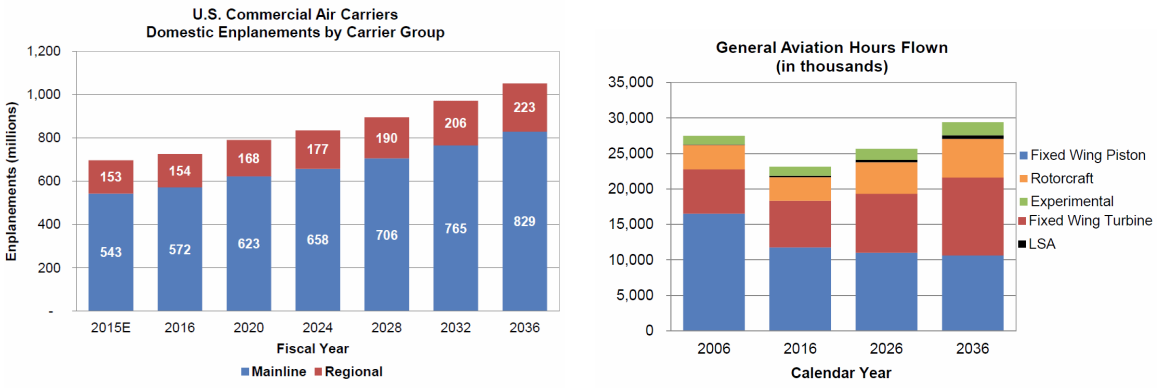
is the mainspring of a $20 billion a year industry. It generates more than $150 billion in economic activity [15]. Thousands of communities benefit because of the positive ripple effects GA airports and operations have on the local economy. All these statistics point to the vital importance of GA to the community as well as the economy.

1.2 Aviation Safety Levels and Trends

The safety record of aviation operations has been steadily improving for the past 50 years [22]. The National Transportation Safety Board (NTSB) statistics indicate that aviation accounted for less than 2% of transportation-related fatalities in 2013 in the U.S. [109]. Advancements in technology, focused efforts of government, manufacturers, and operators through various safety-improvement programs, and increased awareness of safety has resulted in continued improvement in safety in terms of almost all metrics.

According to the Federal Aviation Administration (FAA), the demand for air travel and traffic is predicted to grow steadily through 2036 at a rate of approximately 1.8% annually [52]. As seen in Figure 2, commercial operations are expected to double in the next twenty years. For GA operations, various factors such as the prices of fuel, liability issues, age of pilot population, safety, and others have played a part

in declining numbers for GA hours flown in recent years [141]. Despite declining numbers for GA activity in recent years, the FAA forecasts an optimistic outlook that GA operations are expected to gain a revitalization in the coming years (Figure 2b).



(a) Predicted growth in commercial operations

(b) Predicted growth in GA hours flown

Figure 2: Source: FAA aerospace forecast [52]

Additionally, the density of the airspace is expected to increase due to Unmanned Aircraft Systems (UAS) potentially operating in the already crowded airspace. With such a large increase in expected operations and complexity, there is an ever-increasing demand for maintaining and improving safety of all aviation operations.

1.3 Current Issues in GA

While the overall safety record in aviation has improved, despite the best efforts of regulators and safety practitioners, GA safety continues to lag behind commercial aviation. A significant fraction of accidents that do occur in aviation operations are within the GA domain [109]. As seen from Figure 3, the overall number of GA accidents (both fatal and non-fatal) has decreased steadily over the past few decades as has the accident rate per 100,000 flight hours. However, this improvement has tapered off in recent years with the rate and number of accidents still an order of magnitude higher than commercial operations [109].

Therfore, improving GA safety has been among the top priorities for various regulatory bodies as evidenced by it consistently being featured on the NTSB's Most

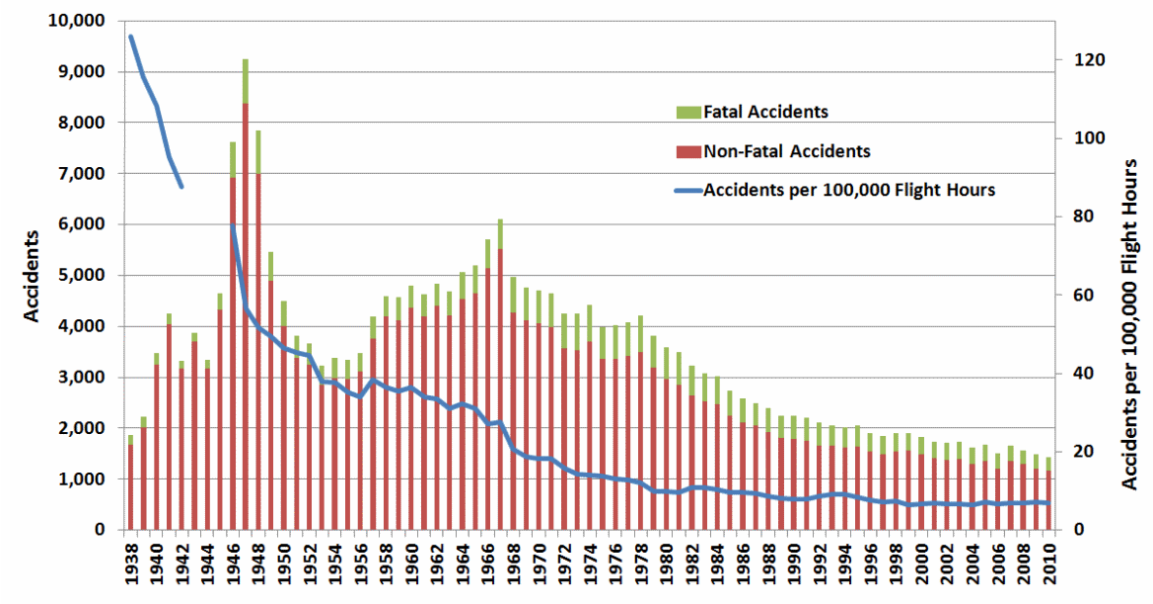


Figure 3: GA accident rates 1980 – 2012 [14]

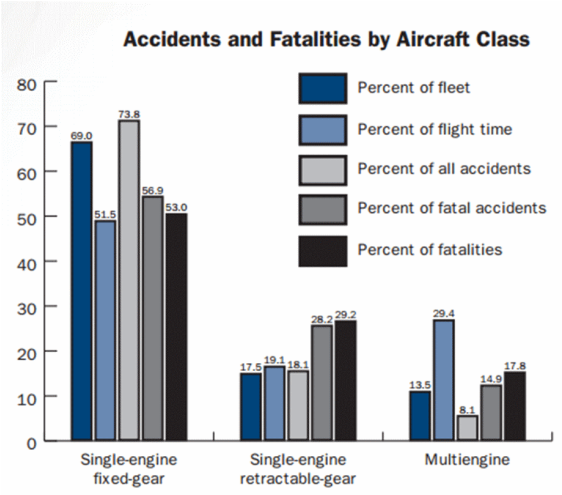
Wanted list of critical improvements to the transportation system over the past several years [110]. The FAA have identified some of the top causes of fatal GA accidents as Loss of Control Inflight, Controlled Flight Into Terrain, System Component Failure Powerplant, Low Altitude Operations, and others.

1.3.1 Category of Aircraft and Operations Considered

What Does General Aviation Fly?

	Air Taxi	General Aviation
Piston single-engine	1,751	145,818
Piston multiengine	1,483	17,854
Turboprop single-engine	582	3,477
Turboprop multiengine	882	4,574
Turbojet	2,681	7,704
Helicopter	1,954	7,613
Experimental	107	23,121
Light Sport	0	6,066
TOTAL	9,440	216,227

(a) Types of aircraft flown in GA



(b) Accidents and fatalities by aircraft class

Figure 4: Source: 2012 Nall Report [14]

While the accident rates for GA flight operations are higher than those in commercial aviation, it is noted that GA contains a heterogeneous fleet and operations. This risk of accident is further accentuated in GA owing to this diversity in operations, pilot experience, fleet composition, and airfields (towered and non-towered). As seen from Figure 4a, GA operations include single and multi-engine piston, turboprop, turbojet powered aircraft as well as helicopters and experimental aircraft. However, accident and incident rates, number of active aircraft, and proportion of hours flown are not uniform across all the aircraft classes within GA. Historically single-engine piston aircraft make up a significant proportion of the entire fleet ($\approx 69\%$) and the number of accidents ($\approx 74\%$) and fatalities ($\approx 53\%$) as seen in Figure 4b. Therefore, it is of particular importance to examine operations of this class of aircraft within GA and to improve its safety record. Most of these aircraft typically belong to the *normal* category under the FAA airplane categories (14 CFR Part 23.3 [2]). The normal category is limited to airplanes that have a seating configuration, excluding pilot seats, of nine or less, a maximum certificated takeoff weight of 12,500 pounds or less, and intended for non-acrobatic operation [2]. Following are some of the important characteristics that distinguish the category of aircraft and operations which form the focus of this thesis:

1. Smaller sized aircraft
2. Less weather-tolerant aircraft
3. Limited or no flight data recording capability
4. Highly variable and heterogeneous mission profiles
5. Variability in pilot certificate and experience level (number of hours flown)
6. Greater variety of airports of operations
7. Operations mostly under Visual Flight Rules (VFR)
8. Large number of instructional flights

All of these factors present important challenges for improving safety of GA operations in this category. Therefore, existing mature methods from commercial operations need to be tailored to this category of GA operations or entirely new methods need to be developed.

1.3.2 Ongoing Efforts and Programs

The problem of improving aviation safety has been acknowledged by many regulatory bodies who have taken strides in the right direction by the initiation of several safety enhancement programs. While majority of these programs are for commercial aviation, recently many of them have also made their way into GA operations. These programs typically include reporting and archiving of incidents, accidents, safety deviations, etc. reported using qualitative or quantitative metrics. Some of them are voluntary while others are official reports of accidents/incidents. Some of these programs and databases are listed here along with their purpose:

1. **Aviation Safety Action Program** (ASAP) is a joint FAA/industry program that allows employees to self-report safety violations to air carriers and FAA while being protected from legal or disciplinary consequences.

2. **Air Traffic Safety Action Program** (ATSAP) is an agreement between the FAA, the National Air Traffic Controllers Association (NATCA), and the National Association of Government Employees (NAGE) that fosters a voluntary, cooperative, non-punitive environment for FAA air traffic employees to openly report safety concerns.

3. **Flight Operational Quality Assurance** (FOQA) alternatively known as Flight Data Monitoring (FDM) involves collection and analysis of routine flight data and has been widely adopted in commercial operations [4].

4. **Aviation Safety Information Analysis and Sharing** (ASIAS) [53] program

partners with the Commercial Aviation Safety Team (CAST) and General Aviation Joint Steering Committee (GAJSC) to monitor known risks, evaluate the effectiveness of deployed mitigation, and detect emerging hazards. ASIAS aims to connect a number of safety databases in order to facilitate integrated queries across multiple safety databases.

5. **Aviation Safety Reporting System** (ASRS) [130] is a system established by the FAA and NASA that enables anonymous (and voluntary) reporting of aviation *incidents* from pilots, air traffic controllers, flight attendants, maintenance staff, and eye witnesses

6. **NTSB Accident Database** [109] is a publicly-available database maintained by the NTSB which investigates and reports findings from accidents that occur in the U.S. The findings from the investigations are disseminated in the form of textual reports (summary and factual reports) and coded information in an accident database.

While many of these and other programs started off for commercial operations, they have trickled into GA operations as well. Specifically within GA, the General Aviation Joint Steering Committee (GAJSC) is a government and industry group that uses the same approach as the CAST on the commercial operations side. The GAJSC analyzes GA safety data to develop intervention strategies to prevent or mitigate problems associated with accident causes, called Safety Enhancements (SE) [63]. Research efforts such as the Partnership to Enhance General Aviation Safety Accessibility and Sustainability (PEGASAS) [54] are also aimed at improving the safety of GA operations. The work presented in this thesis is expected to contribute to the ongoing active research in these areas and offer additional capabilities.

1.4 Summary

Driven by the high accident and incident rates, the GA industry is currently looking for innovative and sustainable methods for assessing and improving the safety record. Safety analysts in the GA domain face significant challenges due to some of the limitations identified in this chapter. While there are many ongoing efforts from government, industry, and academia to address these issues, new methods are sought that can advance the state of practice. These observations lead to the overall research objective:

Research Objective:

Develop a quantitative data-driven methodology that will enhance safety assessment of GA operations by providing metrics for quantifying flight safety, enabling identification of non-standard operations, and providing predictive capabilities that will complement existing approaches

The rest of the document is organized as follows:

- Chapter 2 presents the literature review of the existing methods for quantitative safety assessment from various domains and their applicability to GA
- Chapter 3 presents the identified problems, research questions, hypotheses, and proposed experiments to address these questions
- Chapters 4, 5, and 6 provides details of the developed method including the metrics, algorithms, and techniques developed for quantitative safety assessment
- Chapter 7 outlines the results of a case study conducted using real-world data obtained from GA operations
- Chapter 8 contains the conclusions, contributions, and impact of the work undertaken in this dissertation

CHAPTER II

BACKGROUND AND LITERATURE REVIEW

2.1 Definitions

Prior to presenting the background and literature review of existing methods, the definitions of some important recurring terms in the safety literature are established and their context/application within this thesis is elaborated. While these terms may have different interpretations outside the scope of aviation safety, the following definitions established within the International Civil Aviation Organization (ICAO) Safety Management Manual [76] have been utilized in the present work.

- **Safety** – "*The state in which the possibility of harm to persons or of property damage is reduced to, and maintained at or below, an acceptable level through a continuing process of hazard identification and safety risk management*"
- **Hazard** – "*A hazard is defined as those conditions which could cause or contribute to unsafe operation of aircraft or aviation safety-related equipment, products and services.*"
- **Safety Risk** – "*The projected likelihood and severity of the consequence or outcome from an existing hazard or situation.*"
- **Flight Safety** – "*Flight Safety is defined as the aspect of operational safety associated with the aircraft while it is actually flying.*" Therefore, it excludes operations such as taxiing, ground operations, etc.

While prediction and analysis of risk and hazard identification are important in improving the overall safety record of GA, the focus of this thesis is the flight safety

aspect of GA operations, specifically, methods or metrics for quantifying the flight safety compared to other flights.

2.2 Current Safety Practice

Traditionally, the aviation industry has followed a *reactive* approach to safety assessment and hazard identification. This includes monitoring the number or frequency of accidents, incidents, and fatalities for a specific time period. In the past, incidents, accidents, and/or fatalities have been the primary triggers for identifying problems and developing mitigation strategies [94]. However, such a reactive approach to safety enhancement is not conducive to the early detection of potential issues.

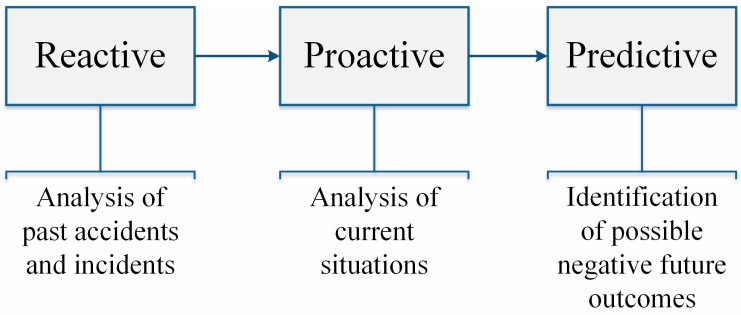


Figure 5: Different types of safety hazard identification methodologies

The industry is now moving towards a more *proactive* and *predictive* approach. In the *proactive* approach, potential unsafe events are identified beforehand and mitigation strategies are implemented in order to prevent accidents and loss of life. This is achieved by the safety assurance function through audits, evaluations, employee reporting, and associated analysis. In the *predictive* approach data obtained from routine operations as well as accident and incident data is monitored to identify possible negative *future* outcomes. The different safety programs outlined earlier in Chapter 2 are geared towards enabling the industry's objective of proactive and predictive safety assessment and hazard identification.

In the past, several researchers and safety organizations have relied on qualitative data such as historical accident or incident data in their efforts to improve or characterize safety. These efforts have ranged from systems level analyses in commercial operations [105], identification of high-risk occurrence chains in helicopter accidents [125], comparison of hazardous states and trigger events in helicopter accidents and incidents [126], weather-related fatal accidents in GA [61], development of pro-active methods of data collection via pilot interviews [139], etc. While valuable information is gained from these studies, the potential availability of additional quantitative data can further enhance the quality of insights by opening up different avenues for analysis.

2.3 Quantitative Safety Assessment

Safety is a concept that may be difficult to quantify, because it is associated more with the absence of something rather than its presence [128]. However it may be approximated using metrics calculated from available data [114]. Previous accident studies (particularly in GA) concluded that *lack of information* is sometimes a major hurdle in performing accident analyses, and methods to obtain richer data from flight operations should be sought [67]. To that end, the capabilities of Digital Flight Data Recorders (DFDRs) have increased tremendously over the past few decades. Quick Access Recorders (QAR) are similar in performance to DFDRs, but they are usually mounted in an easily reachable location on the aircraft and have removable media. Therefore, this leads to increase in data proliferation as it can be easily accessed and downloaded for further analysis.

Figure 6 shows the evolution of the capabilities of DFDRs over the past 60 years. The capability of recording parameters has increased significantly such that modern day transport aircraft can record thousands of parameters at a high frequency. Additionally, transport category airplanes have certain minimum requirements related

Aircraft Type	Year entered into service	Number of parameters	FDR type	FDR data capacity
Boeing 707	1958	5	Analog	Mechanical limit of 5 parameters
Airbus 330	1993	280	Digital	128 wps[1] serial
Embraer 180	2004	774	Digital	256 wps[1] serial
Airbus A380	2007	>1000	Digital	1024 wps[1] serial
Boeing 787	2009	>1000	Digital	Ethernet

Figure 6: Evolution of flight data recorders from Campbell [30]

to DFDRs set by the FAA for U.S carriers [1]. This includes a minimum set of more than 90 parameters related to the state, attitude, control surface deflections, engine information, environmental conditions, Global Positioning System (GPS) information, and others. On the other hand, for the GA aircraft considered in this work, there are no such stringent requirements on DFDR parameters [3]. Therefore, GA flight data recorders (if present) typically record much fewer parameters due to prohibitive costs [73]. More recently, certain GA aircraft have been equipped with glass cockpits displays such as the Garmin G1000 [65] that allow recording of more parameters. Even then, these capabilities fall short of those in commercial operations (missing important parameters such as flap deflection, mixture lever position, weight, etc.). However it is worth exploring the techniques that are currently used on recorded flight data from commercial operations in order to understand which of these may be adapted for GA.

The availability of additional information can augment safety enhancement process in reactive, proactive, as well as predictive approaches. This information can be in the form of more recorded parameters, use of performance models developed offline to estimate or predict unrecorded parameters of interest, estimation of parameters based on the measured values of certain combinations of parameters, weather

data collected by a source other than the aircraft and so on. Methods for quantitative safety assessment within this dissertation are those that utilize data from actual flights in order to assess/quantify the level of flight safety at which the system is operating. This may be expressed in the form of various metrics, safety envelopes, deviations from nominal operations etc. While there are various ways of obtaining, organizing, and analyzing flight data, Flight Operational Quality Assurance (FOQA) is perhaps the most systematic and widespread method currently in existence. The subsequent section explores this in further detail.

2.4 FOQA/ FDM Programs

The most common and widespread programs in existence for quantitative safety assessment are Flight Operational Quality Assurance (FOQA) [4] or Flight Data Monitoring (FDM) [28]. It is defined as:

> "*The systematic pro-active use of digital flight data from routine operations to improve aviation safety within a non-punitive and just safety culture*"

In the United States, FOQA is promoted by the FAA as a voluntary program. Data collected from FOQA efforts by individual fixed-wing commercial airlines have been integrated into ASIAS since 2007. While the specific implementation may differ slightly from operator-to-operator, the process generally involves the same set of typical steps which are - a continuous cycle of data collection from on-board recorders, retrospective analysis of flight data records, identification of operational safety exceedances, design and implementation of corrective measures, and monitoring to assess their effectiveness. FOQA programs typically also contain some ability to replay the flight that happened using the positions, attitudes and accelerations recorded in flight. Figure 7 shows the various steps in a typical FOQA process.

British Airways was one of the first operators to adopt a FOQA program in the 1970s and experienced significant reduction of hull losses since its adoption [57]. Since

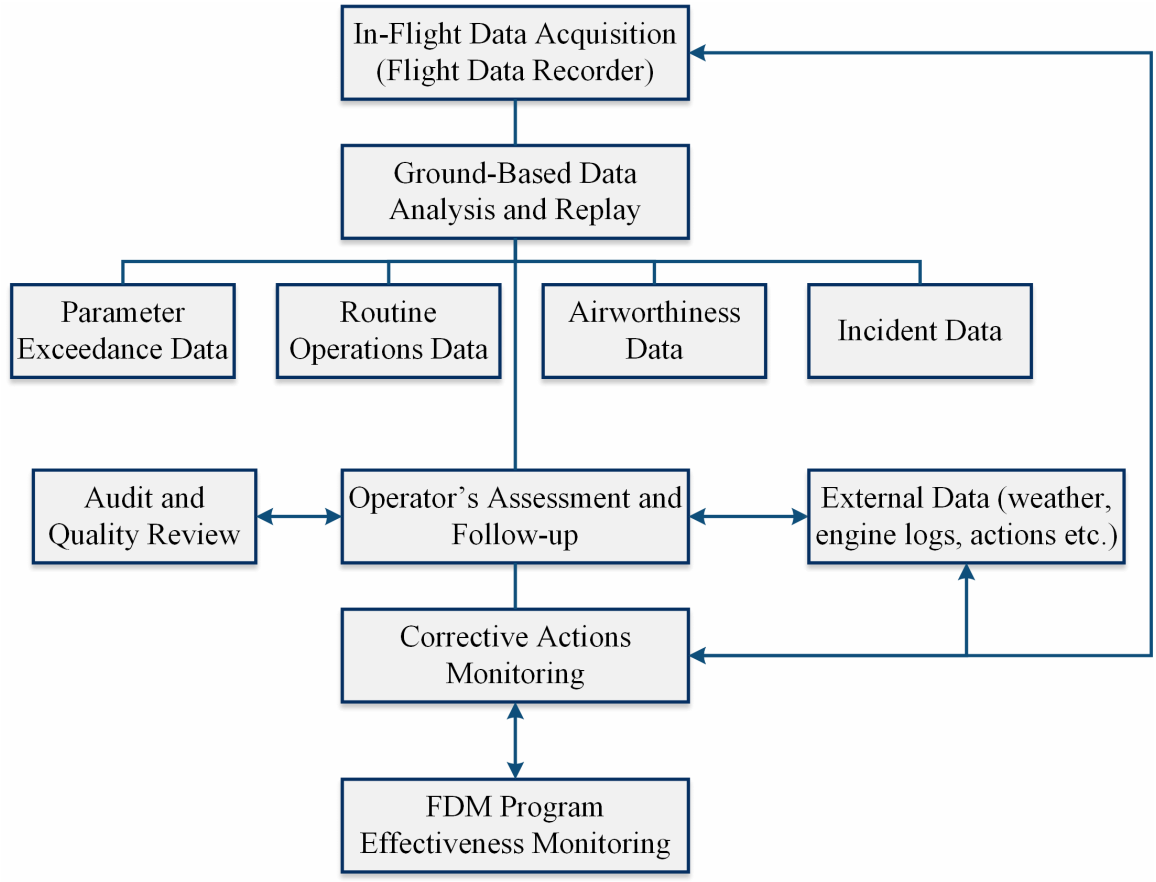


Figure 7: Flight Operational Quality Assurance program components. (reproduced from [28])

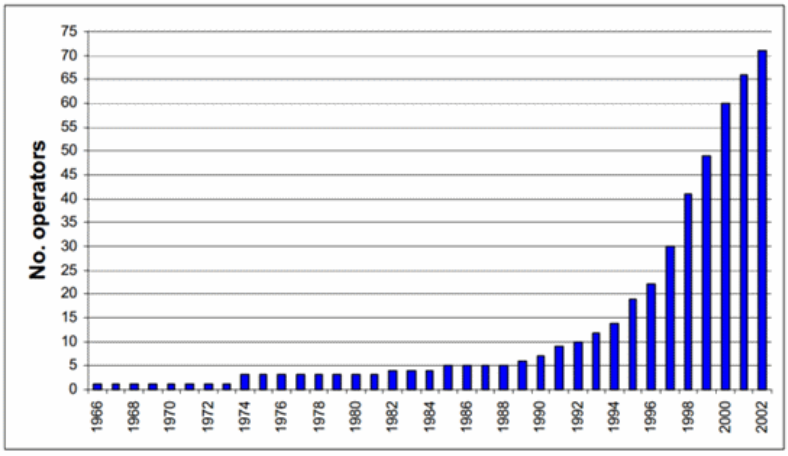


Figure 8: Number of Operators using FOQA from Campbell [30]

then, the number of commercial operators using FOQA has increased tremendously over the past few decades [30]. With the demonstrated effectiveness of FOQA in

fixed-wing commercial airlines for several decades there have been efforts in recent years to promote its adoption among smaller operators, particularly in GA. Through the analysis of GA flight data, significant insight can be gained into the current GA flight safety trends. Mitchell et al. [103] have explored the obstacles that need to be overcome to implement a GA-FOQA program. Kuo et al. [82] conducted studies on the type of recording equipment that could facilitate a FDM implementation for GA.

The data obtained from FOQA programs can be analyzed in various ways. However, prior to introducing the details of analysis methods, it is important to understand the nature of the data obtained from flight data recorders as this has an impact on the techniques that can be used to analyze the data. FDR typically contains different channels that record discrete, continuous, and categorical data at a specified frequency (e.g., once per one second interval). Therefore, in analysis, the data is used as a multi-variate time series for flight records, which are typically of different duration. As noted previously, the number of parameters in this multi-variate time series can be as low as 20 – 30 in GA operations to a number as high as thousands in commercial operations. The data obtained for FOQA implementation from FDR can be analyzed in many different ways. Based on the type of analysis performed, the important methods of analyzing FDR data for safety assessment are grouped as shown in Figure 9. It is noted that this is not the only way in which these methods can be classified, but one of the ways which provides a broad sweep of all techniques relevant to this dissertation. The subsequent subsections elaborate on each of these methods in further detail.

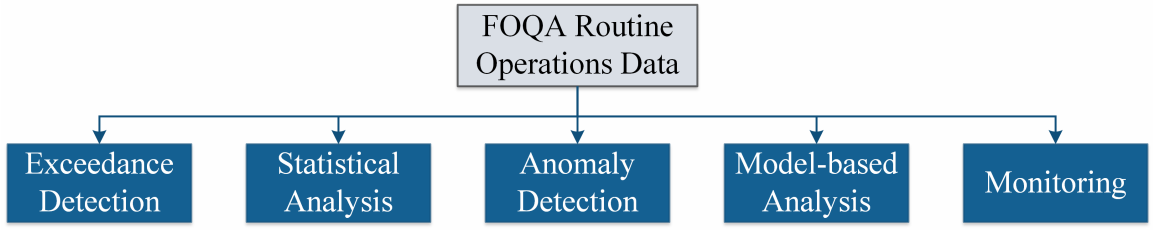


Figure 9: Typical methods of quantitative safety assessment using FOQA data

2.4.1 Exceedance Detection

Exceedance detection is the most common analysis using FOQA data currently in existence. An *exceedance* is the deviation of a single parameter beyond an established threshold. An *event* is defined by one or more parameter exceedances that take place concurrently over a specified period of time. This approach is thus used to monitor for pre-defined events that might be operationally undesirable. The events/exceedances are typically defined by subject matter experts and may correspond to the flight envelope, standard operating procedures, regulations, or previous experience from accident investigations. Individual operators may choose to alter established limits. Typical exceedances and events correspond to significant deviations from Standard Operating Procedures (SOPs). Some examples of exceedances as defined by the FAA Advisory Circular are shown in Figure 10. Hurst et al. [75] recommend that a range of thresholds should be included to generate events when parameters exceed preset values in order to classify events based on their severity. A three-level scheme (low, medium, high) is found in many applications [28, 73].

Observations

This approach performs well on known safety issues but is sometimes blind to safety-critical conditions that may be captured by flight data records but are not in the set of pre-defined safety events. That is because it is designed to provide answers to questions that the domain experts have thought to ask. Setting up the events and exceedances requires extensive subject matter expertise and long periods of fine-tuning, which can result in decreased performance before the system is brought to its full capabilities. Most often, small GA operators do not have the engineering resources available for such a task and this reduces the amount of fine-tuning possible. On the other hand, greater fleet heterogeneity and types of missions in GA can increase chances of missed detection or false alerts without fine-tuning. On the other hand, as shown by Fala and Marais [56], events defined for a particular aircraft may not be

Event Name	Event Description	Parameters and Basic Event Definition	Notes
Takeoff Climb Speed Low	An event to detect climb speed lower than desired during the Takeoff Phase of flight.	CAS, Gross Weight, HAT HAT > x feet, HAA < x feet, CAS < V_2 – x knots	Altitude ranges should be used to accommodate different desired climb speeds in those ranges.
Early Flap Retraction	An event to detect any flap movement from the takeoff position prior to reaching the altitude at which flap retraction should begin.	HAT, Flap Position HAT< x feet, Flap Position < Flap Position in the preceding sample	

Figure 10: Examples of exceedances and events as defined by FAA FOQA advisory circular [4]

directly applicable to other similar aircraft and would therefore, need to be adjusted or tuned to prevent false alerts or missed identifications. In exceedance detection, each flight is monitored as a standalone data point and compared to the defined limits. No information from other flights that might be available gets directly incorporated. This might result in the loss of insights that might be gained from analyzing a larger data set.

2.4.2 Statistical Analysis

Statistical analysis refers to a variety of techniques currently in use and being developed to work with larger data sets. In this approach, information regarding specific parameters/points of interest is aggregated over the entire data set available for analysis. This information is then synthesized and presented in the form of distributions of parameters/events. This analysis enables the operator to gain a more complete picture of the operations based on the distribution of all flights. Statistical analysis enables presenting a high-level summary which the operator can use to evaluate the

overall state of operational performance. Statistical analysis can be performed using different types of data and is not necessarily restricted to quantitative data from FOQA programs. There exists a body of literature in commercial as well as GA operations dedicated to improving safety insights using statistical analysis of collected data.

Campbell [30] has suggested looking at distribution of altitude at which landing flaps are set as shown in Figure 11. Sherry et al. [140] performed a big data analysis of 21 days of surveillance track data to analyze frequency of occurrence of aborted approaches and their safety implications. Cokorilo et al. [39] conducted a statistical study that compares aircraft accidents in relation to the characteristics of the aircraft, environmental conditions, route, and traffic type. Subramanian and Rao [144] have used deep neural networks along with NASA's ASRS database to forecast possibility of go-around events in the NAS. Netjasov and Janic [107] have conducted a review of safety and risk modeling in commercial operations.

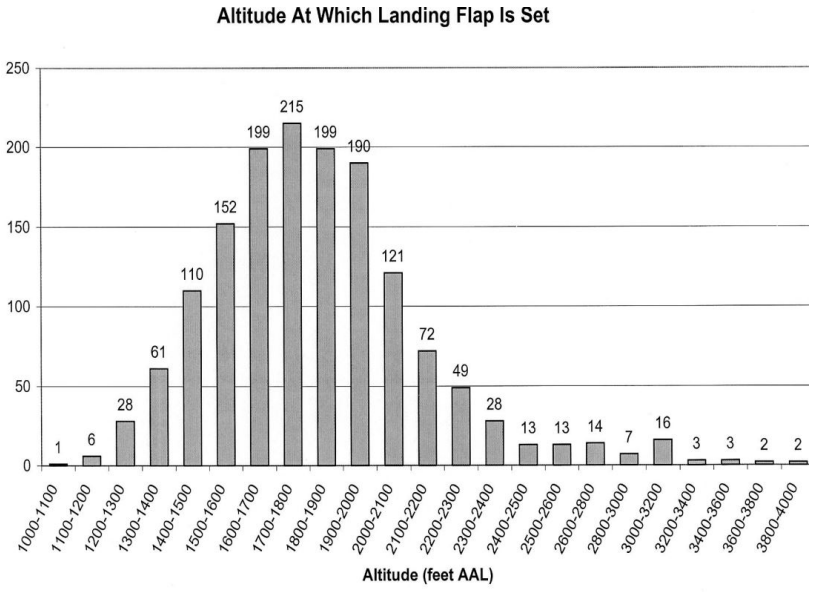


Figure 11: Altitude at which landing flap is set from Campbell [30]

In previous work (Puranik et al. [118]) statistical distributions of various parameters during GA approach and landing operations were analyzed which might be

of interest to the operator – such as the touchdown velocity shown in Figure 12. Archer [17] has looked at a statistical comparison between rounded base turns versus traditional rectangular turns in pattern landings to determine if rounded base turns provide increased safety with respect to angle of attack. Other researchers have used incident, accident data from NTSB and other sources to perform a qualitative statistical analysis. Boyd [23] has conducted a statistical study of GA accidents related to exceeding the weight or center of gravity limits of the aircraft using NTSB accident database. Major et al [97] have investigated perceptions of deficiencies in pilot training based on VFR-into-IMC accident data.

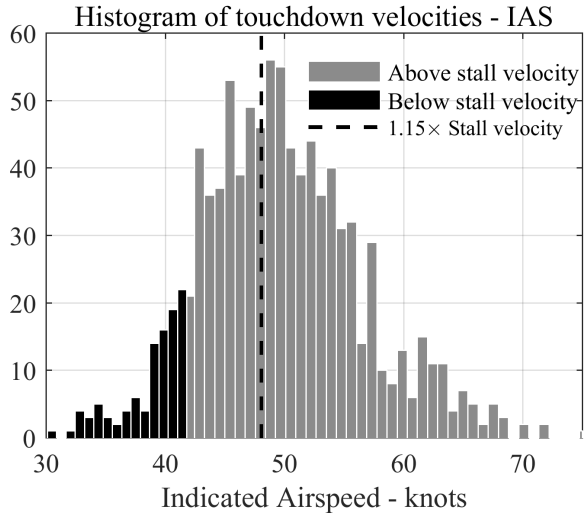


Figure 12: Distribution of touchdown velocity from Puranik et al. [118]

Observations

Statistical analysis, like exceedance detection, requires a watch-list of key parameters and operating conditions. It is very useful for trend analysis of parameters or operations which are known to have a potential impact on operational safety. Comparison of an individual flight record can then be done with respect to the entire data set in order to understand how it fared compared to other flights in similar situations. While this offers an added layer of information over exceedance detection, it typically performs well on known issues but remains blind to emerging risks.

2.4.3 Anomaly Detection

Data mining and anomaly detection techniques for safety analysis, incident examination, and fault detection have garnered increased interest in the aviation community. The objective of anomaly detection techniques is to detect abnormal flights within routine flight data without any prior knowledge of safety events. They involve the use of various machine learning algorithms (clustering, classification etc.) on FDR or accident data to identify emerging risks. Typically, the data obtained from FDR is pre-processed to generate features that can be used in the algorithms. These features are generated from the recorded parameters using the temporal nature of the data as well as distinction in different phases of flight [41, 89, 99]. Anomalous flights obtained (sometimes as ordered lists based on their 'anomalousness') are then further analyzed by experts for potentially dangerous/unsafe conditions. In recent years, due to the availability of data and increased computational resources, there has been a tremendous increase in the application of these techniques in the aerospace domain. Gavrilovski et al. [67] have provided a comprehensive survey of data-mining and anomaly detection techniques applied to flight data. Some of the seminal work in the application of anomaly detection techniques to FDR data as well as upcoming techniques are summarized here.

Majority of the existing literature in this domain is dedicated to data obtained from commercial aviation operations. The literature identifies two main types of anomalies in aviation data – *Flight Level Anomalies* in which the entire flight record or phase of flight considered are off-nominal, and *Instantaneous Anomalies* in which only an instant or small part (a few seconds) of the flight record is off-nominal. Within each type of anomaly considered, three techniques are typically observed – supervised, semi-supervised, and unsupervised techniques. Supervised techniques rely on a labeled training set consisting of typical system behaviors as well as anomalous

behaviors to train algorithms which can then be used on new flight data. Semi-supervised techniques only require a set of training data that contains mostly nominal behaviors. Finally, unsupervised techniques operate under the assumption that the given data set may contain anomalous as well as normal operations in any proportion.

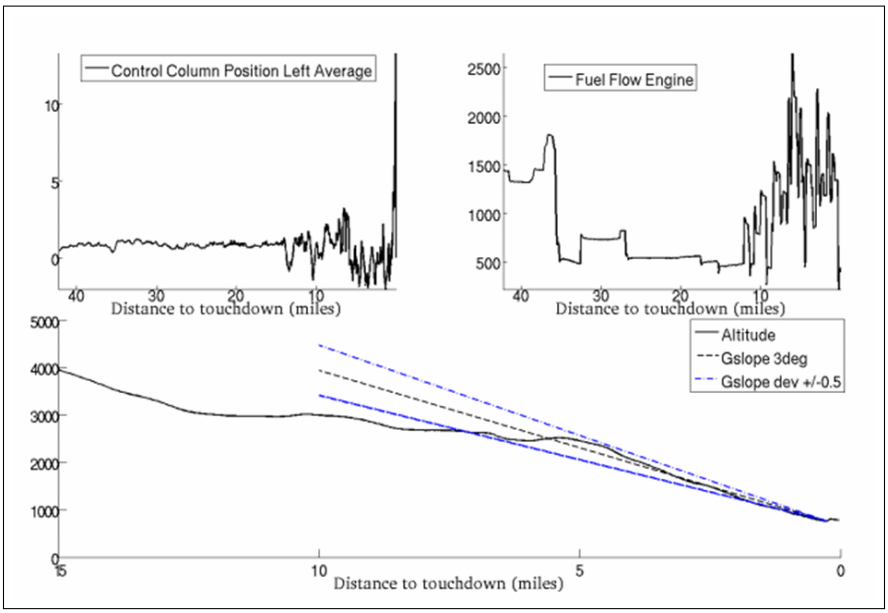


Figure 13: An example of anomaly detected in flight data using MKAD [41]

There have been quite a few studies on identification of flight level anomalies using FDR data. SequenceMiner [27] is a software used to detect anomalies in discrete parameter sequences by learning from a model of normal switching. This technique detects flight-level anomalies but is limited to discrete data. Das et. al. [41] have developed Multiple Kernel Anomaly Detection (MKAD) which applies a one-class support vector machine for anomaly detection. MKAD identifies flight level anomalies well in data that contains discrete and continuous parameters. However, the normalized Longest Common Sub-sequence (nLCS) kernel used may result in the loss of some finer features for continuous data. Smart and Brown [143] used a data-mining approach using a number of one-class classification algorithms to identify anomalies in the descent phase for airliners. They tested their method on a labeled data set containing abnormal flights. Li et al. [89] have developed ClusterAD – an algorithm

that uses density-based clustering for anomaly detection on pre-processed flight data parameters to identify abnormal operations. One of the issues in this method can be that it isolates each sample in each signal as a unique feature, when in fact, the change over that time may be an important factor. Another potential issue is the fact that anomalies identified are purely mathematical artifacts and may need subject matter expertise to validate them. Despite these potential limitations, MKAD and ClusterAD represent the state-of-the-art for anomaly detection in the aviation domain and have shown to be very effective at uncovering a host of anomalies that are missed by traditional approaches. Matthews et al. [99] have discussed and summarized the aviation knowledge discovery pipeline using various state-of-the-art algorithms. While these techniques have primarily been in the commercial aviation domain, there have been recent efforts in the GA domain as well. Clachar [38] has used supervised and unsupervised approaches on GA-FOQA data followed by subject-matter-expert filtering to identify atypical flights.

Typically techniques that are used to identify instantaneous anomalies are different than those used to identify flight-level anomalies. Amidan and Ferryman [11] have utilized Singular Value Decomposition (SVD) to identify instantaneous anomalies. They mapped the five seconds before and after each recorded data point and fit a linear regression model to it. The coefficients of the regression model were then used to create a mathematical signature for each recorded data point which was used to identify outliers. Mugtussidis [106] has used Bayesian classification to distinguish between typical data points, that are present in the majority of flights, and unusual data points that can be only found in a few flights. Orca [18] is a technique that uses scalable k-nearest neighbor approach to detect anomalies in data with continuous and discrete features. Since each data point is a sample in time and treated as independent by the algorithm, Orca struggles to detect anomalies with temporal signatures. Supervised learning methods such as Inductive Monitoring System (IMS) [77] rely

on a training set consisting of typical system behaviors which is compared with real-time data to detect anomalies. Each point is monitored standalone and therefore, the temporal aspect of anomalous sub-sequences is lost when identifying anomalies. Li et al. [90] developed ClusterAD - Data Sample, which is a technique leveraging a mixture of Gaussian models to identify probability of a sample being anomalous during take off, approach, and landing. However, this method also treats each data sample independently and uses additional context to identify whether a particular gaussian component is appropriate at a given time. This could result in missing anomalies with temporal signatures.

Observations

Anomaly detection techniques have been shown to be very effective at uncovering a variety of different abnormalities without any pre-specified limits. They are able to harness the information available in the entire data set in a systematic way. However, in most aviation applications, they have been demonstrated on a fleet of the same aircraft (sometimes at the same airport). Therefore, it is unclear how well it generalizes and allows for comparisons with flights conducted at different locations and with different aircraft. Additionally, most approaches in literature are pure data-mining based in that they only utilize the recorded data and therefore do not use any additional domain expertise. One of the difficulties associated with anomaly detection techniques is in validating the results obtained. Most of the data from routine operations is expected to contain non-anomalous data. There appears to be consensus in the literature that performance of anomaly detection techniques can be improved by including domain-specific knowledge in the data-mining task as this will increase the chances of identifying truly anomalous conditions [66]. Finally, few applications of anomaly detection have been demonstrated on GA flight data.

2.4.4 Model-based Analysis

In model-based analysis, data obtained from routine operations is typically used in one of two ways. It can be used to train models for various purposes [37, 69, 100]. In these type of models, a relationship is assumed between the inputs and outputs of the recorded data and the models are trained by estimating the parameters using common techniques such as linear regression or least squares. This approach is similar to the well known field of System Identification (SysID) [78].

Alternatively, the flight data can be used in conjunction with models developed independently from first principles to identify safety issues. Typically, physics-based approaches are an alternative to data-driven approaches when sufficient data is not available or the available data is insufficient for safety analysis. Physics-based approaches have been previously used for fault analysis [153], monitoring, and helicopter FDM [66].

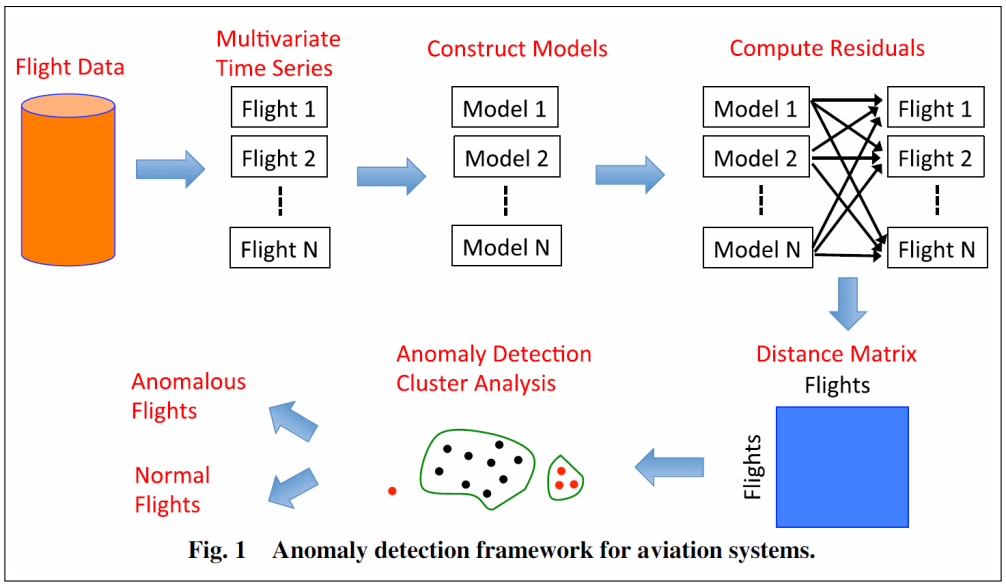


Figure 14: An example of data-driven model based anomaly detection using flight data – from Melnyk et al. [100]

For models trained using flight data, most of the times, their use in safety analysis is closely tied with anomaly detection techniques. Gorinevsky et al. [69] have

described an application of data mining technology called Distributed Fleet Monitoring (DFM) to Flight Operational Quality Assurance data. This application consists of fitting a large scale multi-level regression model to the data set and finding anomalies using these built models. The algorithm is able to identify anomalies within a flight record (instantaneous), abnormal flight-to-flight trends (flight level anomalies) and abnormally performing aircraft. While this framework is capable of identifying instantaneous anomalies, it is limited to models fitted in the clean configuration. Also, most of the anomalies detected are in the estimation of aerodynamic or propulsion parameters or gross weight. Chu et al. [37] have proposed an approach for detecting anomalies from aircraft cruise flight data using a model trained using historical data of a fleet of aircraft. Anomalies are detected as outliers that exceed the scatter caused by turbulence and the modeling error. Melnyk et al. [100] have treated each multivariate time series using a vector auto-regressive exogenous model. Dissimilarity between two flights is measured as the residuals obtained by using the model of one flight on the data of another. Outliers are identified using Local Outlier Factor (LOF) which is a nearest neighbor based anomaly detection method. This method requires that a different model be trained for each flight record being analyzed (as seen in Figure 14) and that inputs-outputs be recorded which may not necessarily be the case for GA. For models generated from first principles, Gavrilovski [66] has used both static performance models and a nonlinear dynamic simulation to improve the detection performance of safety events in helicopters. Harrison et al. [71] have developed a non-linear dynamic model of GA aircraft that can be used in conjunction with flight data records to identify loss-of-control safety margins.

Other than the two types of models described above, another category of model-based analysis consists of using either qualitative or quantitative data to generate models of flight-safety, risk, hazard, or accident chains. There is a wealth of literature that exists on quantifying and assessing risk in aviation. Some of the recent work is

identified here. Netjasov [107] have conducted a review on risk and safety modeling in civil aviation, Rao [124] has proposed a new method for modeling helicopter accidents, Luxhoj and Coit [96] have modeled low probability/high consequence events using an aviation safety risk model, etc.

Observations

For data-driven models, the model is a local approximation with a limited or diminishing capability away from condition at which it is identified. Additionally, many of the models currently developed are for the clean configuration whereas most accidents in GA operations occur in approach and landing which is typically associated with varying levels of flap usage and therefore aerodynamically dirty configurations [109]. Data-driven models typically require a rich data set consisting of inputs-outputs which is usually not available in GA FDR as noted previously. Where data-driven models fall short, physics-based models may be used as they have the ability to relate measurements to unobserved quantities that might be of interest. They are of particular interest for data sets which have limited number of recorded parameters such as those in fixed wing GA. However, predictions made using basic first principles models can suffer from error due to epistemic uncertainty (inadequacy of the model). Therefore, they need to be well-calibrated and the uncertainty associated with predictions from these models needs to be quantified.

2.4.5 Monitoring

This type of analysis usually involves monitoring the mechanical systems for vibration levels and other signals and alert operators to unsafe increases that might signal deterioration of components. Maintenance is scheduled to coincide with actual deterioration of the part and eliminate premature replacement or unsafe usage. Monitoring aims to establish nominal levels of variance in the monitored signals and generate alerts when these are exceeded. The core technique used here is Statistical Process

Control (SPC). The most common type of monitoring in the aviation domain is Health and Usage Monitoring Systems (HUMS) used in the rotor-craft domain. Other similar techniques also exist such as fault diagnosis which deal with monitoring the data from specific components and not FOQA data.

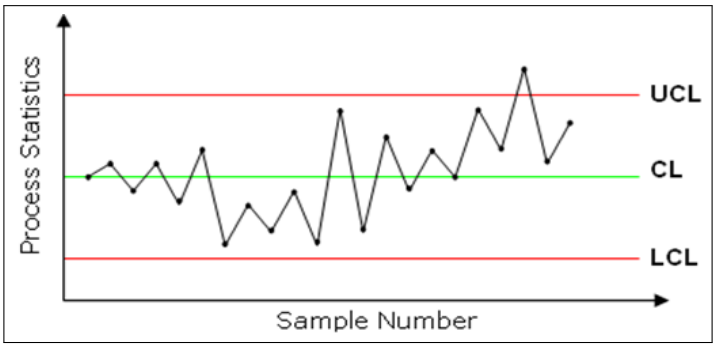


Figure 15: Example of control chart used for monitoring processes (Image Source: `http://www.processma.com/resource/spc.php`)

Observations

Romero et al. [131] studied the feasibility of using HUMS in rotor-craft operations. They concluded that the flight trial program generally demonstrated a high level of reliability in monitoring the rotor system, engines, drive train, and life-limited components. More recently, Delgado et al. [45] conducted a survey of current rotor-craft propulsion health monitoring systems. Monitoring is typically available on advanced systems with on-board capability to monitor and record data from sensors mounted to the dynamic components. There are frequently very well-specified limits on allowable vibration or noise spectrum available to the operators. Monitoring systems are one of the few online applications where alerts might be generated while the system is in operation. However, these systems are primarily used for maintenance scheduling and tracking deterioration of components. Also, they are similar in function to exceedance detection as the set-point about which monitoring is happening and the allowable variances are determined a-priori. In GA, applicability of HUMS would be limited to high-end GA aircraft with the capability of recording required parameters.

Also, applicability of HUMS processes in literature is limited to propulsion system and structural components where vibrations can be easily measured.

2.5 Overarching Observations

1. **Need for data-driven safety environment in GA:**

   FOQA is considered an essential aviation safety tool, with wide adoption and demonstrated safety benefits among airline operators. It was identified earlier, that one of the major hurdles in accident analyses is the lack of recorded information from flight operations. Some issues faced in adoption of FOQA in GA are sociological and cost-related [73] which are not addressed in this research context. However, it is known that GA pilots almost exclusively maintain and improve skills on their own, and their conduct of safe flight depends more on individual abilities and judgment, potentially leaving them unprepared for situations [109]. Therefore, a data-driven approach using data from actual operations will benefit GA pilots and operators by alerting them of potential unsafe behaviors or excursions outside the safety space. Thus there is a need for a data-driven safety assessment approach from the reactive as well as proactive and predictive safety assessment philosophies. No universal approach currently exists for safety assessments of GA operations with associated heterogeneity in mission profiles, data recording capability, and types of aircraft. Therefore, there is a need for an approach that can utilize GA flight data and provide greater level of safety analysis and insights than is currently in existence.

2. **Data sources for GA flight data:**

   It is noted previously that GA aircraft typically have a lower level of sophistication in terms of their flight parameter recording and logging capability. This can range from anywhere between a Personal Electronic Device (PED) such as an iPad to a Garmin G1000 glass cockpit [65]. The number of parameters,

frequency, accuracy, latency, etc. of these devices can vary significantly. Research has been conducted [73, 82] into the type of recording capabilities that would be required to implement a full-scale GA-FOQA program. Therefore, new techniques that might be developed should be cognizant of this variability in data quality and the potential ways of combining information from multiple sources. Care needs to be taken while using data-fusion techniques to account for the additional uncertainty and error that might be introduced due to such operations.

3. **Metrics for safety assessment:**

   In order to enable safety assessments across a heterogeneous fleet, operations, and data recording capabilities, appropriate metrics need to be used that are comparable in these differing scenarios. Safety metrics that are not dependent on the size of the aircraft would be suitable for such a task. Additionally, it is desired that the metrics used for quantifying flight safety be related to those that have had an impact on historical accidents/incidents. Therefore, previous accident databases such as those maintained by the NTSB or existing criteria for safe operations such as the Stabilized Approach Criteria (SAC) [51, 60] by the FAA need to be leveraged.

4. **Existing FDM approaches:**

   Exceedance detection is designed to provide answers to questions that the domain experts have thought to ask. There can be a lack of traceability in event definitions and a need to continuously fine-tune safety events. Moreover, events are defined in limited number of dimensions, which usually correspond to the most critical parameter(s) and are not well established for improbable conditions. Greater fleet heterogeneity and types of missions in GA can increase chances of missed detection or false alerts without fine-tuning [56]. Therefore,

approaches that complement event-based detection must be sought to enhance safety assessments.

Anomaly Detection or discovery-based approaches have proved useful in identifying abnormal flights without prior knowledge of events. However, their applicability has been shown on fleet of similar aircraft at a particular airport. Most of the existing safety approaches use only the raw flight data as features in the anomaly detection task. These algorithms are highly sensitive to the features used. Therefore, appropriate features that correspond to the safety margins of the aircraft need to be selected for anomaly detection. There appears to be a consensus in the literature that including domain-specific knowledge in the anomaly detection task improves the chances of identifying truly anomalous conditions. Different algorithms have been shown to have different strengths. Therefore, rather than using a single algorithm, combining the strengths of different algorithms may provide improved results.

5. **Data-driven versus Model-based safety assessment:**

   Purely data-driven techniques are preferred when a sufficient quantity of rich data is available (e.g. Commercial FOQA), and model-based techniques are preferred when certain aspects are not well represented in the data and/or there is limited data availability (e.g. Helicopter FDM). With regard to the availability of flight records from routine operations, GA lies in between these extremes (as GA flight data records are not as rich as commercial aircraft FDR) and therefore a hybrid approach would be best suited: using a data-driven approach augmented with predictions from basic physics-based performance models. However, the use of first principles models exhibits its own set of limitations or errors because all the input parameters required to accurately predict outputs or metrics of interest might not be recorded in flight. Therefore, calibrating models using real flight data records and/or data in the public domain

literature might provide better accuracy. In the presence of such uncertainties, models that provide reasonable or sufficient estimates of quantities of interest at acceptable computational cost are sought.

6. **Retrospective nature of safety assessments:**

   Even though current safety trends are moving towards proactive and predictive metrics, all existing analysis is still retrospective in nature. While the techniques developed monitor and attempt to proactively prevent future incidents, they are still inherently retrospective in nature. Developing models offline that can provide approximate predictions online would allow assessments to be real-time.

7. **Limits of operations**

   Analyzing routine operations will only yield data that is within the normal operating envelope. Using simulation models to sample the behavior near edges of the operational envelope or at known hazardous conditions can help obtain information about these situations in order to avoid them.

Developing techniques that address some of the gaps from the observations and additional ones made during the course of the implementation and analysis form the foundation upon which the technical approach undertaken in this dissertation is based.

CHAPTER III

PROBLEM FORMULATION

3.1 Research Objective

Given the need for enhancing the safety of GA operations identified earlier in chapter 1 and the challenges, existing work, and opportunities for improvement presented in chapter 2, the overarching research objective of this dissertation is stated as follows:

Research Objective:

Develop a quantitative data-driven methodology that will enhance safety assessment of GA operations by providing metrics for quantifying flight safety, enabling identification of non-standard operations, and providing predictive capabilities that will complement existing approaches

There are certain assumptions and constraints under which the methodologies developed in this dissertation are applicable. These have been identified here.

1. **Flight Data:** The first assumption is the availability of actual or simulated flight data records. In the work presented in this dissertation, flight data from over two thousand training flights conducted on a Cessna C172S and Piper Archer PA28 aircraft equipped with Garmin G1000 [65] glass cockpit displays is utilized. Each data record contains information about aircraft state characteristics such as altitude, true airspeed, indicated airspeed, latitude, longitude, and engine RPM among others, collected at a one second interval. It is noted that all the data used in this disseration is de-identified in accordance with typical FOQA procedures. In addition to data recorded by flight data recorders,

Personal Electronic Devices (PED) such as iPads have become increasingly popular within GA for recording flight data using applications such as the freely available GAARD app developed by MITRE [104] as well as devices (such as the STRATUS [16]) that can be used to display vital information in-flight. It is noted that while the G1000 represents the higher end of instrumented GA aircraft currently around, the dissertation will address potential limitations of the methods developed and solutions in the case of data obtained from sources that might be less accurate (such as PEDs).

2. **Simulated Data:** In addition to this, a flight simulation model capable of simulating the dynamics of the aircraft using a MATLAB/Simulink model which is connected to FlightGear Flight Simulator [58] to allow visual rendering of the motion of the aircraft is available. Further details of the simulation model can be found in Chakraborty et al. [33]. The simulation model has been modified to output a data record in the same format as the Garmin G1000. Thus it can be utilized to generate additional flight data records which can be included in the data set with the data records from actual flights. The simulated flight data can enable exploration of potentially unsafe or abnormal conditions which might not be easily available in routine data.

3. **Performance Models:** Finally, the availability of basic aircraft performance models for modeling the aerodynamic and propulsion characteristics is assumed. The model for the aerodynamics includes a lift and drag polars for an aircraft using empirical build-up methods such as the one presented in Min et al. [102] and can be modified for similar aircraft. The propulsion model on the other hand is one which can provide the thrust and torque curves for a propeller such as the method presented by Harrison et al. [72]. The assumption of availability of these models does not necessarily mean that they are accurate for all the aircraft

being considered. In fact, as demonstrated later calibration of performance to the data available is a necessary step prior to their use in safety analysis as demonstrated later in the dissertation (Ch. 6).

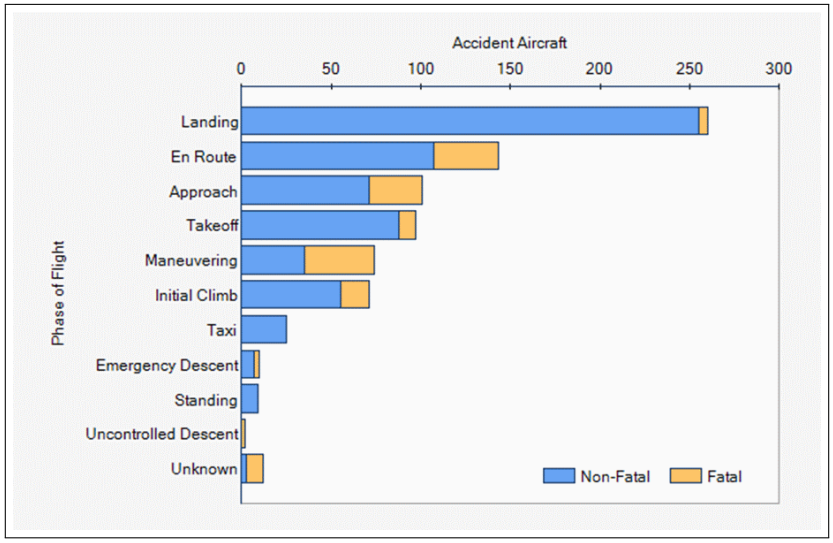


Figure 16: Distribution of accidents by phases of flight in GA

4. **Phases of Flight:** While improving GA safety across all flight regimes is the aim of this and other work in literature, it is noted that not all phases or regimes of flight pose the same amount of risk. For instance take-off and approach/landing are considered among the highest risk phases of flight. For instance, approach and landing account for about 2% of the total flight time, but account for almost 50% of the total accidents [108]. During approach and landing, the aircraft is flying at a reduced airspeed, in a landing configuration, and closer to the ground, compared to its airspeed and altitude during cruise, and therefore requires more precision and focus from the pilot. Errors or deviation from the prescribed flight path during approach and landing are likely to end in accidents or incidents as they have a small safety margin owing to the limited time for a pilot to correct an error or react to a deviation. Similarly take-off is a phase of flight with high number of accidents per time flown. According to a report by AOPA [5] on accidents during instructional flights, more than

50% of all accidents occur during take-off, landing, or go-around. Take-off and landing phases also require clearing obstacles such as tall buildings, windmills etc. Therefore, it is clearly evident that these two phases of flight are among the most critical in terms of targeting safety improvements for GA operations. Due to all these reasons, the methodology developed in this thesis is primarily applied to these two phases of flight with the understanding that it could be extended to other phases as well.

5. **Data Noise:** It is noted that flight data obtained from GA FDR can have noise as well as bias in it. While this dissertation utilizes techniques for data smoothing and filtering such as moving window average or local regression based smoothing, it is acknowledged that more sophisticated techniques may exist for certain types of data. For example, Seimbiring et al. [138] have demonstrated performing data compatibility check might as an explicit way of accounting for systematic biases/ epistemic error. It is not within the scope of this dissertation to explore these techniques.

3.2 Research Questions

The stated research objective may be realized by addressing three major research questions which are identified in the following section. The statement of the research objective is broken down into three constituent parts and each is analyzed here. These three parts are – *(1) metrics for quantifying flight safety*, *(2) identification of non-standard or anomalous operations*, and *(3) providing predictive capabilities*.

3.2.1 Safety Metrics

It was noted earlier in the observations (Sec. 2.5) that there is a need to develop a data-driven safety assessment framework for GA operations. However, due to the heterogeneity of the fleet, operations, and data recording capabilities, there does not

currently exist a universal methodology that is able to take all of these factors into account. The different methods of analyzing FOQA data and their strengths and weaknesses were reviewed in Sec. 2.4. Events purely based on certain parameters and pre-defined limits have a tendency to be inconsistent and aircraft–specific [56]. Therefore, these observations motivate the first major research question:

Research Question 1:

What are metrics or figures of merit corresponding to safety margins and safe operations in GA which are readily estimable from recorded data in GA operations?

In order to identify safety metrics that can be used for quantifying flight safety, it is necessary to look into what have been the causes of incidents and accidents historically, existing criteria for events, and the parameters that are recorded in-flight and available to define different metrics. Using the information from these three sources will enable formulation of useful metrics. This leads to the following research subquestion:

Research Question 1.1:

What are the important parameters that can be used to define potential safety metrics of interest in GA operations?

It is important that metrics used in assessing flight safety of GA operations correspond to actual performance and limits of the aircraft and its systems. Different quantitative safety programs such as FOQA in commercial operations define events or exceedances in their Standard Operating Procedures (SOPs). These events are typically identified using a set of recorded parameters and limits which are based on

historical accidents/incidents as well as subject-matter-expert opinions. While similar literature is limited in GA operations, recent efforts have been made to address this issue [56, 73]. The parameters identified in these efforts and the lessons learnt are incorporated. There have been efforts by the FAA such as the stabilized approach criteria (SAC) [51] which outline the limits on certain parameters for conducting a safe and 'stable' approach and landing. The flight safety foundation (FSF) have identified energy management and associated parameters as among the important causes of unstabilized approaches [59]. On the other hand, various safety databases currently in existence such as the NTSB accident database or the NASA ASRS database contain a wealth of information about deviations in parameters that led to GA accidents or incidents. The parameters and criteria thus observed can then be utilized in defining metrics of interest or reusing existing parameters/metrics which raises the next research sub–question:

Research Question 1.2:

What safety metrics can be defined using typically available important parameters and criteria that correspond to limits of GA aircraft operations and performance?

The parameters whose deviations might have been the cause of accidents or incidents are identified in the previous research sub-question. The information from these parameters as well as past safety criteria need be used in defining safety metrics of interest. The first important consideration in defining metrics is related to the nature of GA operations and limitations of recording capabilities. It is desired to have parsimony in the number of safety metrics that eventually get defined and used. This parsimony is desired not only in terms of how many metrics are used (due to computational cost) but also in terms of what parameters need to be recorded to evaluate these metrics (due to data recording limitations). Secondly, metrics that

have actual relevance from the point of view of flight safety are desirable. Therefore, a mapping of how different metrics correspond to or help in the identification of historically risky situations or specific GA events is desired. Thirdly, safety metrics that have no loss of generality in the characterization of aircraft states and safety boundaries across the GA fleet and operations under consideration are sought. Based on these considerations the following hypothesis for research question 1 is formulated:

Hypothesis 1: *Energy-based metrics such as those quantifying the energy state, rates of change of energy, and their margins and deviations are useful metrics that correlate to safety margins and safe operations in GA and can be readily estimated from recorded data.*

The relevance of the metrics is partially addressed by a literature survey and the usefulness in a retrospective safety analysis setting during actual use of the metrics in the developed methodology. A summary of research question 1, hypotheses, and experiments is shown in Figure 17.

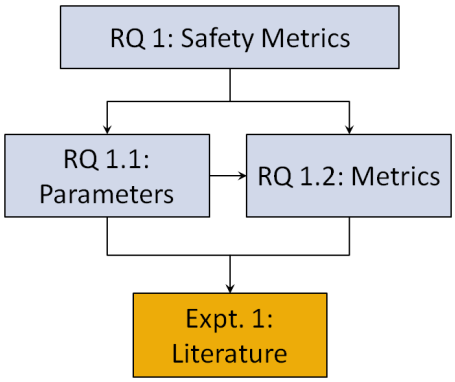


Figure 17: Summary of Research Question 1, Sub-questions, and Experiment

3.2.2 Non-standard/Anomalous Operations

Identification of usable safety metrics is the first step in retrospective safety analysis of GA operations. As previously noted in the literature review, the availability of recorded flight data opens up avenues for quantitative analysis that were earlier unavailable in GA operations. Techniques that are able to leverage large volumes of data and provide insightful results are sought. Similarly, methods that require minimal a priori input for identifying anomalous operations are preferred since they would complement the existing static methods such as exceedance detection. This leads to the statement of the second research question:

Research Question 2:

How can GA flight records be analyzed using energy metrics to automatically identify nonstandard or anomalous operations without a priori knowledge of safety events?

There are typically two types of anomalies associated with multi-variate time-series data such as that encountered in aviation operations – *Flight Level Anomalies* in which the entire flight record or phase of flight under consideration are anomalous, and *Instantaneous Anomalies* in which only a small part (a few seconds) of the flight record is anomalous. In general, different types of techniques need to be used to identify each type of anomaly and it is important to be able to identify each in order to compare them. Existing techniques that have been shown to be effective on flight data can be utilized or tailored to data obtained from GA operations. Due to the desire to obtain non-standard operations without a priori knowledge of safety events, unsupervised techniques are utilized. It is also desired that a quantification of the 'anomalousness' be provided based on information from all the flight data available. All these requirements lead to the following hypothesis for research question 2:

Hypothesis 2: *Using the defined energy metrics as features in a general anomaly detection framework with appropriate techniques will enable the detection of both types of anomalies.*

It is noted that while a general anomaly detection framework is suitable for identifying anomalies, different techniques need to be utilized for each type of anomaly. For flight-level anomalies, an entire flight or phase of flight is identified as anomalous and as such it would work only on well-defined phases of flight. Therefore, as noted earlier, the take-off, approach and landing phases are chosen for the development of this technique. The research sub-question formulated to help identify flight-level anomalies is stated here:

Research Question 2.1:

How can a general anomaly detection framework be modified to identify flight-level anomalies in GA take-off, approach and landing operations using energy metrics?

Techniques used for flight-level anomalies do not necessarily work well for instantaneous anomalies as the unit of analysis is the entire phase of flight rather than an instant or a few seconds. Therefore, within the same general anomaly detection framework, different algorithms and pre-processing need to be used for identifying instantaneous anomalies. This leads to the following research sub-question:

Research Question 2.2:

How can the general anomaly detection framework be modified to identify instantaneous anomalies in GA operations using energy metrics?

Experiment 2.1 and 2.2 are designed to test the detection of different types of anomalies in routine flight data, whereas experiment 2.3 is designed to test and determine the performance of the the method in the presence of limited flight data capturing capability or model generalization. The success of these experiments will depend upon identifying anomalies without specifying what is anomalous a priori. In addition to identifying anomalies from routine flight data, the ability to identify most of the existing exceedances defined in FOQA literature is also investigated. The experiments are described here in brief and details can be found in Chapter 5.

Experiment 2.1: Detection of flight-level anomalies

Using an appropriate combination of machine learning techniques and defined energy metrics, identify anomalous operations from within a given set of almost three thousand actual flight data records as well as some simulated flight data records. Additionally, visualize the obtained anomalous flights in comparison to other flights in terms of their energy metrics as well as flight data parameters. In order to facilitate validation, simulate a normal and an unstabilized final approach and landing and verify whether each gets captured as such using the implemented method. For flight level anomalies, take-off, approach and landing are the two phases of flight considered. These phases are chosen since historically, a majority of accidents in GA domain have occured during these phases [108]. A combination of two popular anomaly detection algorithms developed in the methodology is used in this experiment - density-based clustering (DBSCAN [50]) and one-class support vector machine (SVM). MATLAB is used to analyze the flight data records, evaluate energy metrics, and implement the anomaly detection algorithms because of ready availability of versions of the algorithms in the public domain [98].

Experiment 2.2: Detection of instantaneous anomalies

Using appropriate machine learning techniques and defined energy metrics, identify instantaneous anomalies within a flight record. Information from all flight records will be utilized in training the model used in detection. Visualize instantaneous anomalies using probability of anomaly at every point in the flight record. Additionally, compare instantaneous anomalies to available sample exceedance detection events available from Higgins et al. [73] to check for overlap between developed and existing methods. For instantaneous anomalies all phases of flight can be considered since the techniques do not require well-defined phases however, the demonstration is restricted to the phases most relevant to this dissertation – take-off, approach, and landing for the sake of comparison. A novel sliding window-based technique is utilized in conjunction with Gaussian Mixture Models (GMM) for anomaly detection. MATLAB is used to implement anomaly detection due to the ready availability of GMM algorithm implementation.

Experiment 2.3: Generalization of methodology

In order to test the generalization of the methodology, two scenarios will be tested. These scenarios will enable testing of the methodology for variability in flight data recorder capabilities and the ability to train models to predict anomalies in data not used in training. First, the flight-level anomaly detection algorithms are implemented using different subsets of energy metrics that require fewer number of parameters/models. This will enable understanding of which anomalous flight records are potentially missed due to lack of certain recorded parameters. Second, flight level anomalies will be detected using SVM trained on only a subset of the flight data records. The trained model will be used to identify anomalies within the entire data set. This will enable identification of whether the trained models are able to generalize on new flight data that might come in and will be used in retrospective analysis.

These efforts will enable assessing the generalization and widespread application of the methodology developed in this dissertation.

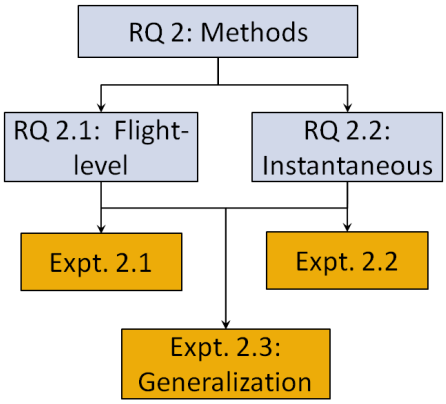


Figure 18: Summary of Research Question 2 and Experiments

A summary of research question 2, hypotheses, and experiments is shown in Figure 18.

3.2.3 Model Calibration

Previous studies have concluded that lack of information is often a hurdle in accident and incident analyses. While the capabilities of flight data recorders have improved significantly over the years, their adoption in GA has still not been widespread mainly due to the associated costs. As recognized previously, using predictions from physics-based models can help estimate quantities of interest that might not be recorded in actual flight data but are of importance in safety analysis. In the context of this dissertation, performance model for an aircraft refers to the steady state point performance of the aircraft in terms of lift and drag characteristics for the airframe as well as thrust and torque characteristics of the propeller for that aircraft which may be queried at specific conditions of interest. Such performance models can be necessary in the evaluation of some energy metrics used in safety analysis as seen later in Ch. 4. Performance models such as those developed by Harrison et al. [72] and Min et al. [102] may be used as a starting point for this purpose. However, such models developed for specific aircraft can suffer from aleatory and epistemic uncertainty when

using in different contexts and therefore need to be calibrated to the specific aircraft in question.

Usually, aside from flight test data (which is very difficult to obtain), the only sources of aircraft performance behavior in the public domain are Pilot Operating Handbooks (POH). While the tables given in the POH are idealized performance that a brand new aircraft can theoretically attain under specified flight conditions, it is nevertheless a good source of initial information to calibrate performance models. For generating performance models of new aircraft that are similar to the one already available, a calibration technique that is repeatable and fast is desired. These considerations motivate the next research question:

Research Question 3:

How can basic empirical models of aerodynamic and propulsive performance of a generic GA aircraft be calibrated to predict energy metrics at conditions of interest for a specific aircraft?

As noted throughout this dissertation, any technique that is to be applied to the GA domain has limitations on the data that is available for developing the technique as well as in its application. Therefore, calibration of performance models of GA aircraft will be different for different types of calibration data available. The eventual aim of aerodynamic and propulsion models is to be able to predict the lift, drag, thrust, torque, and weight of the aircraft with enough accuracy as to allow using them in the evaluation of energy metrics or usage in a flight simulation model.

Two different cases are identified in this dissertation based on typical forms of data that might be available. In the first case, calibration is carried out using data available in the public domain literature, such as a POH. This corresponds to the case where performance models thus calibrated would be used for safety analysis on

flight data recorded using the lower end of the data recording capability spectrum. This case motivates the first research sub–question:

Research Question 3.1:

How can basic empirical models of aerodynamic and propulsive performance of a generic GA aircraft be calibrated using data only available in the public domain to predict energy metrics at conditions of interest for a specific aircraft?

As it is hard to validate performance models without actual flight data, it is assumed that a limited amount of flight data is available to validate the calibrated models from this step. The validation is carried out in terms of a predicted energy metric of interest. The hypothesis for the first sub-question of research question 3 is stated as follows:

Hypothesis 3.1: *Using tables and conditions in the Pilot Operating Handbook to calibrate the performance models results in a process that will yield models for a specific aircraft which can be used for predicting energy metrics of interest.*

Experiment 3.1: Performance model POH calibration

Using performance tables for different phases of flight as well as critical limits of the aircraft (such as stall speeds) identified from the POH, calibrate models using calibration factors to minimize error metric or metrics that represent mismatch between POH performance and predicted performance. Since each phase of flight will yield a different error metric, in general the calibration will be a multi-objective optimization problem. The calibration-optimization is carried out in MATLAB with calibration factors from each discipline as the control variables and the vector of error metrics as the objectives. The well-known Non-Dominated Sorting Genetic Algorithm (NSGA)

II [44] multi-objective optimization algorithm will be utilized for the calibration. Points along the pareto-front thus obtained will be candidate models to be used in the actual application. The choice of one calibrated model versus the other will be dictated by the performance quality of prediction for the intended application.

While calibrating the model to POH does provide performance predictions for that aircraft, it can still suffer from a number of shortcomings. For example – the performance reported in POH is for a brand new aircraft at certain conditions flown by an expert test pilot. The deterioration of the airplane components with use, piloting skill, variations in aircraft model, modifications/maintenance that might have been made to that aircraft, actual gross weight of operations on that particular day, environmental conditions, and noise in recorded parameters are just some of the factors that might cause uncertainty in estimates obtained from a simpler model. While there are some components of these uncertainties that are aleatory (irreducible), others (such as those which are aircraft specific) are epistemic (reducible). The second case of data availability in performance model calibration assumes calibration flights from routine operations with specific parameters are available. The assumptions are based on the higher end of the GA data recording capability spectrum. These observations lead to the next research question:

Research Question 3.2:

How can basic empirical models of aerodynamic and propulsive performance of a generic GA aircraft be calibrated using limited GA flight data to predict energy metrics at conditions of interest for a specific aircraft?

The reducible parts of the uncertainty in the models are those related to the specific make and model of the aircraft. Calibration of performance models to flight data will help reduce this uncertainty and improve the performance predictions. The hypothesis

for the second sub-question of research question 3 is stated as follows:

Hypothesis 3.2: *Using specific total energy rate metric from flight data to calibrate the performance models results in a process that yields models for a specific aircraft which can be used for predicting energy metrics of interest.*

Experiment 3.2: Calibration of performance models to flight data records
In order to calibrate to a flight data record(s), a candidate flight record which has annotated flaps from the available data set is chosen. While the POH calibration contained many performance tables and points for calibration, the actual flight record contains a limited number of points. Therefore, for flight-data-driven calibration, a particular energy metric is chosen as the calibration metric and the discrepancy between the computed and actual values of this energy metric at each point during the flight is evaluated. The aggregation of this residual in the form of root mean square (RMS) error is then minimized.

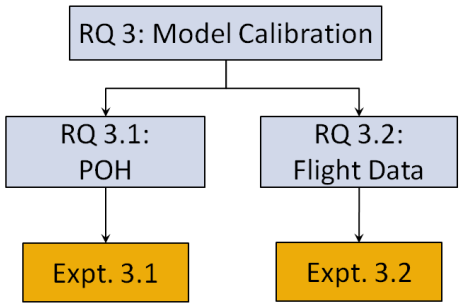


Figure 19: Summary of Research Question 3, and Experiments

In addition to the research questions raised in the previous sections, the entire methodology developed is tested on a large data-set in a case study. The specific questions raised during this application are addressed in the chapter on the results from the case study (Ch. 7). A summary of all research questions, experiments, and their relationship to each other is presented in Figure 20.

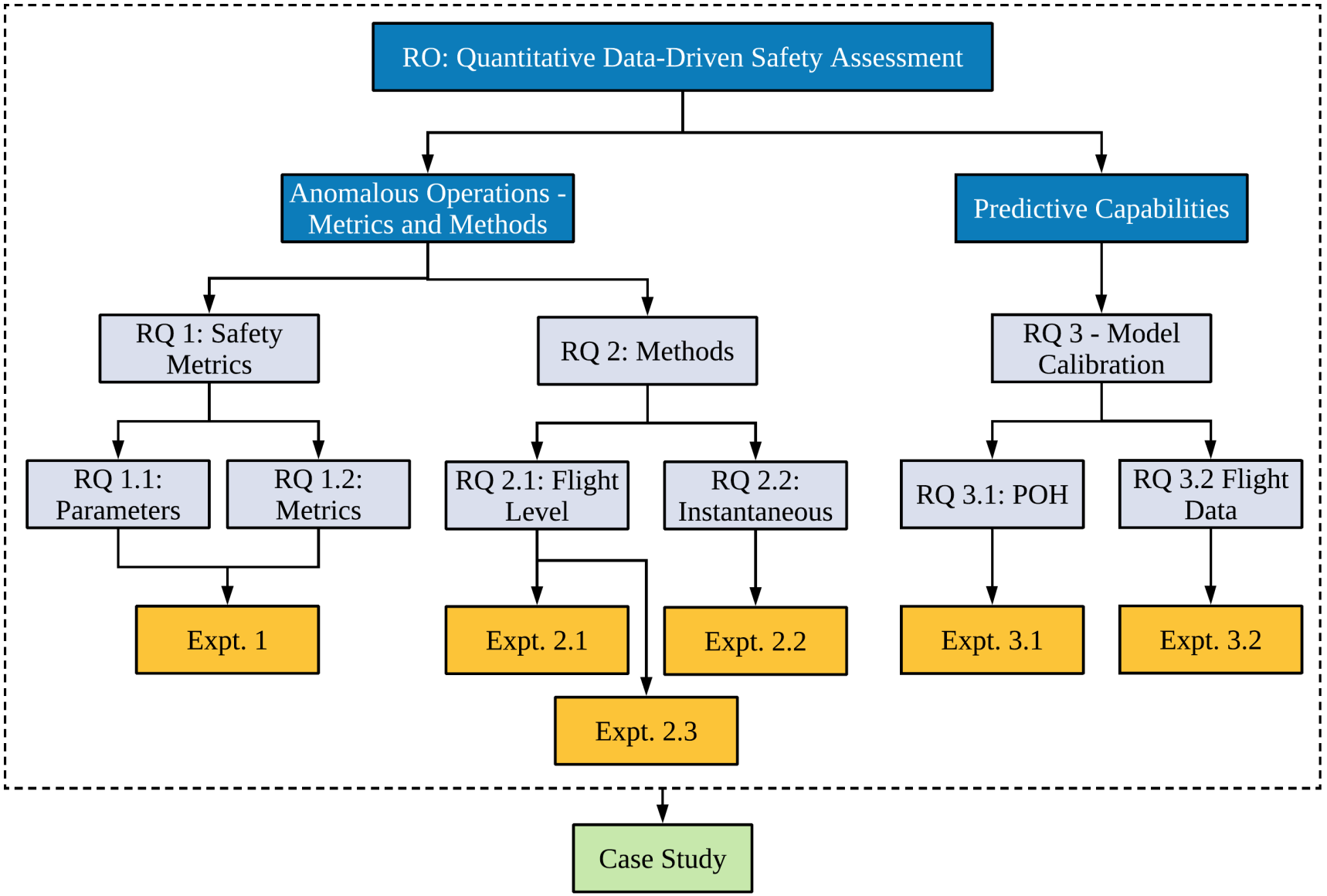


Figure 20: Summary of Research Questions and Experiments

3.3 Developed Methodology

A top-level overview of the quantitative data-driven methodology is depicted in Figure 21. The components highlighted in blue color represent the aspects of the methodology that are developed in this dissertation. On the other hand, the components in white color are external inputs that have been incorporated from previous efforts in literature in the GA domain.

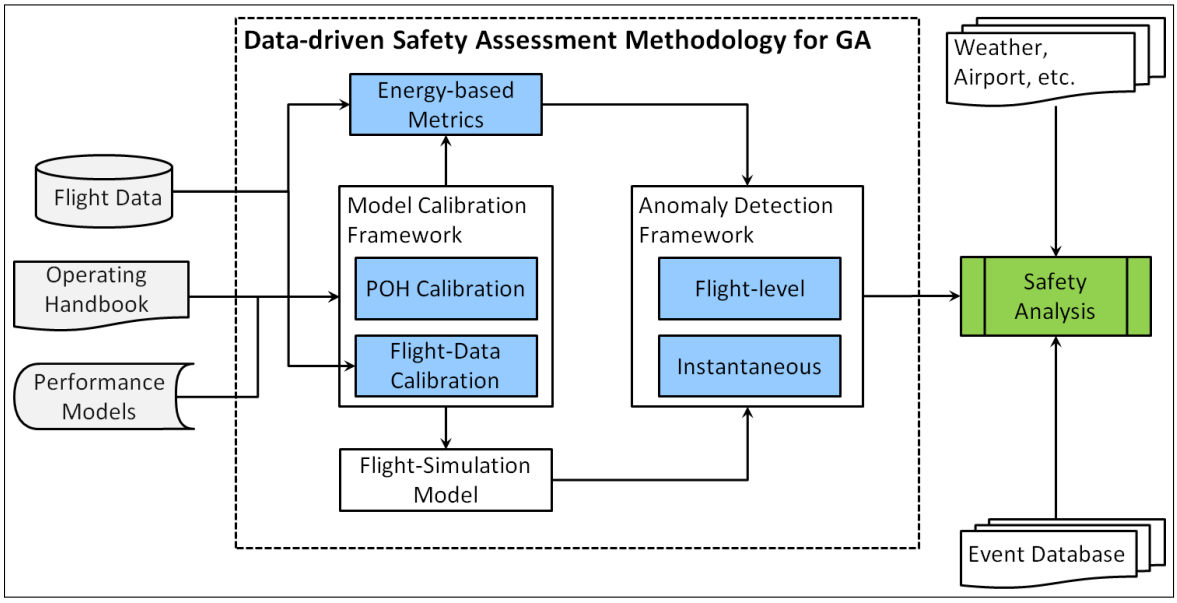


Figure 21: A top level overview of the information flow in the developed methodology

The development of various components of the methodology along with the experiments performed to address the hypotheses are described in further detail in subsequent chapters.

CHAPTER IV

SAFETY METRICS (RESEARCH QUESTION 1)

It is necessary to identify appropriate metrics in order to facilitate quantitative safety assessment. While the concept of safety is inherently subjective and therefore difficult to quantify, it may be approximated using safety metrics. In the context of GA safety assessment, some desirable characteristics of good metrics are identified as follows: First, they must be estimable using available recorded data - this implies that the parameters used for the metrics should be typically recorded using on-board recorders. Second, it is desirable that metrics are usable across different aircraft. Therefore, metrics are invariant with size or weight of the aircraft are preferable. Third, the chosen metrics should correspond to safety or performance margins of the aircraft in some way. This ensures that information relevant to safe operations is embedded or captured when these metrics are used in practice. All these considerations are captured in the statement of the first research question which is reproduced below:

Research Question 1:

What are metrics or figures of merit corresponding to safety margins and safe operations in GA which are readily estimable from recorded data in GA operations?

In order to address the research question, it is broken down into two important components. Before defining the metrics that are going to be used for retrospective safety analysis, it is important to understand what data or parameters are available for defining the metrics. Metrics that provide a good quantification of safety but require the recording of parameters beyond the capabilities of GA FDR would not be useful in this context.

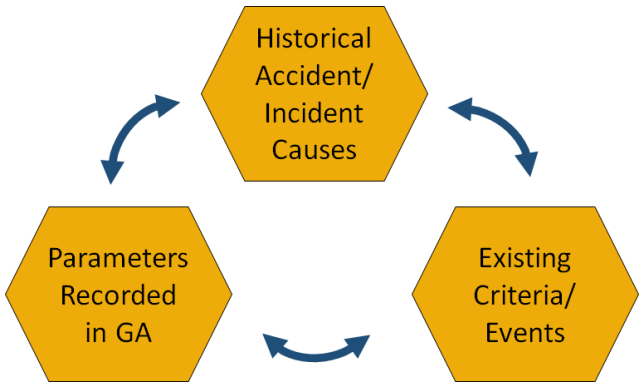


Figure 22: Various sources of information used in determining important parameters

4.1 Parameters – Research Question 1.1

GA FDRs are typically constrained by the number, quality, and type of parameters that they can record. There are different sources from which parameters that are important in defining safety metrics can be understood. These are explored in order to answer the first research sub-question:

Research Question 1.1:
What are the important parameters that can be used to define potential safety metrics of interest in GA operations?

Identifying important parameters to be used in safety analysis can draw upon several different sources of data available in the public domain. Figure 22 provides an outline of the three different criteria used in this dissertation. Each of the choices outlined in the figure are explored further.[1]

[1]The research described in this section is also documented in the following publication:

– Rao A.H., and **Puranik, T.G.**, *Retrospective Analysis of Approach Stability in General Aviation Operations*, accepted for publication in 18th AIAA Aviation Technology, Integration, and Operations Conference, Atlanta, GA, June 2018. [127]

4.1.1 Existing Criteria

There are various criteria currently in existence for defining safe operations in commercial aviation. Some of these criteria are being brought over into GA with the availability of additional flight data. **Stabilized Approach Criteria (SAC)** are well documented in commercial aviation safety literature [51, 60]. The FAA defines a stabilized approach as: " [...] characterized by a constant-angle, constant-rate of descent approach profile ending near the touchdown point, where the landing maneuver begins" [51]. A stabilized approach is viewed as one of the key features of safe approaches and landings in air carrier operations. The specific criteria for classifying an approach as a stabilized approach are presented in Table 4.1.1.

Table 1: Elements of a Stabilized Approach [51, 60]

1	The airplane is on the correct track[a]
2	The airplane is in the proper landing configuration
3	After glide path intercept, or after the final approach fix (FAF), the pilot flying requires no more than normal bracketing corrections[b] to maintain the correct track and desired profile (3° descent angle, nominal) to landing within the touchdown zone
4	The airplane speed is within the acceptable range specified in the approved operating manual used by the pilot (Typically not more than V_{REF}+20 knots indicated airspeed and not less than V_{REF}
5	The rate of descent is no greater than 1000 Feet Per Minute (FPM). If an approach requires greater than 1000 FPM, a special briefing should be conducted
6	Power setting is appropriate for the landing configuration selected, and is within the permissible power range for approach specified in the approved operating manual used by the pilot

[a] A correct track is one in which the correct localizer, radial, or other track guidance has been set, tuned, and identified, and is being followed by the pilot

[b] Normal bracketing corrections relate to bank angle, rate of descent, and power management

Several researchers have analyzed commercial aviation accident data and flight data

records to understand the causes for unstable approaches [88,99,105,140,149]. Payan et al. [115] have classified helicopter approaches into various levels of stability using important parameters. Sherry et al. [140] analyzed 21 days worth of approach and landing trajectory data into Chicago O'Hare (ORD). Their analysis revealed that, on average, there were 7.4 aborted approaches per 1,000 approaches. Further, they reviewed 467 incident reports published in the Aviation Safety Reporting System (ASRS). Issues with the airplane (e.g., on-board malfunctions, failures), traffic separation issues, poor weather, runway issues, and interaction between Air Traffic Controllers (ATC) and pilots were among the top reasons for go-arounds. In related work, Wang et al. [149] analyzed 8,219 landings for Runway 22L at Newark Liberty International Airport (EWR) for a given commercial operator. They evaluated stability criteria (i.e. glide-path, rate of descent, speed change, and acquisition of the runway centerline) at 1,000 ft, 750 ft., and 500 ft. above ground level (AGL). They found that, in general, flights converged towards stable approaches as they approached the runway threshold—66.1% of flights at 1000 ft. AGL, 78.7% of flights at 750 ft. AGL, and 87.2% of flights at 500 ft. AGL satisfied all four stability criteria.

While SAC are not perfect, they have proved to be useful in defining standard operating procedures for airlines. An unstabilized approach may result in a low altitude state at flare or touchdown, a higher ground speed at touchdown, or a high energy state at touchdown, thus leading to landing short, a risk of runway overrun, or a hard landing. While the parameter variations, limits, and some procedures for stabilized approaches as defined for commercial operations may not directly carry over to GA, it is nevertheless valuable to identify the important parameters that are considered in these criteria. Therefore, based on Table 4.1.1, the important parameters utilized in stabilized approach criteria are listed here:

1. Track Angle
2. Configuration (Flap Setting, Landing Gear)

3. Descent Profile (Altitude)
4. Airspeed (V_{ref})
5. Rate of Descent
6. Power Setting

The ability to incorporate these parameters in some form into the eventual metrics developed will prove to be valuable in addressing the relevance of the metrics being developed.

The next set of existing criteria considered are **events or exceedances** currently defined and in use. An **exceedance** is the deviation of a single parameter beyond an established threshold. An **event** is defined by one or more parameter exceedances that take place concurrently over a specified period of time. GA safety events can be hard to standardize because of how GA aircraft and pilots fly, and the subjective nature of the notion of risk and safety. That being said, there have been attempts in GA to define safety events [73] and to gain insights into how these might be useful for improving safety. A complete list of safety events for GA operations identified by Higgins et al. [73] is provided in Appendix B. The following list of parameters is extracted from the event definitions as those being monitored:

1. Airspeed
2. Vertical Speed
3. Bank Angle
4. Pitch Angle
5. Flap Setting
6. Cylinder Head Temperature
7. Power
8. Fuel Quantity
9. Oil Temperature
10. Oil Pressure
11. Centerline Deviation
12. Vertical g-loads

While the criteria explored here are not exhaustive, they represent parameters extracted from various subject-matter-expert drawn limits. Combining these findings

with historical data and availability of parameters in GA will provide valuable insights into which parameters need to be included (implicitly or explicitly) in defining safety metrics.

4.1.2 Historical Data

Other than existing criteria such as stabilized approach, exceedances, events, etc., historical accident and incident data also provide a good source of information about parameters related to safe operations. There are various databases publicly available which contain information from investigations of accidents or voluntary reporting of incidents. Previous studies of accident and incident data have provided important insights into understanding and improving safety. Rao and Marais [125] identified high-risk occurrence chains in helicopter accidents by analyzing NTSB accident data. The current dissertation focuses primarily on quantitative data from routine operations, nevertheless it is of value to explore accident and incident databases for commonly occurring parameters related to accidents. While a causal analysis of accidents and incidents requires a more in-depth analysis, querying the databases for incidents and accidents related to the phases of flight and operations under consideration in this dissertation can provide important insights.

A brief survey of accident and incident data using two publicly available databases is undertaken to identifyg recurring parameters. The first is the NTSB accident database. The NTSB investigates and reports findings from GA accidents that occur in the U.S. The findings from the investigations are disseminated in the form of textual reports (summary and factual reports) and coded information in an accident database. The NTSB maintains a publicly-available aviation accident database, which includes fields such as aircraft type, FAR part, geographical location of the accident, injuries and damage, events in accidents, and narratives. This database is leveraged to find parameters which may frequently result in accidents in take-off or approach

and landing phases. Similarly, in an effort to improve aviation safety, the FAA and NASA established the Aviation Safety Reporting System (ASRS). The system enables anonymous (and voluntary) reporting of aviation incidents from pilots, air traffic controllers, flight attendants, maintenance staff, and eyewitnesses. While the ASRS database can provide valuable insight into situations that did not result in accidents, it is not suitable for statistical analysis. As noted by Bhargava and Marais [21], the tendency for incidents to be under-reported is one of the key reasons for not using incident data for statistical analysis. Both of these databases are queried for accident and incidents in the category and phase of flight of operations considered in this dissertation. The narratives obtained are examined for some of the important parameters identified from criteria provided in the previous section for overlap.

It is observed that among the parameters mentioned earlier, airspeed, altitude or glide path deviations, vertical speed deviations are often important contributors in approach and landing accidents. Additionally, poor energy management, centerline deviations, and unstabilized approach also get frequently cited in the narratives among both accidents as well as incidents.

4.1.3 Parameter Recording Capabilities

As noted earlier in the document, the capabilities of flight data recorders have been improving steadily for the past few decades. While this has been the general trend, GA has typically lagged behind commercial aviation due to constraints on cost. In the development of any new methods of safety analysis for GA domain, the capabilities of flight data recording devices is one of the most important limitations. The spectrum of possible parameters being recorded in GA is fairly wide and heterogeneous. There has been recent research into investigating the use of low-cost PEDs for data collection and safety analysis. While the typical GA aircraft typically lacks an installed avionics system, the use of PEDs such as tablet computers by GA pilots

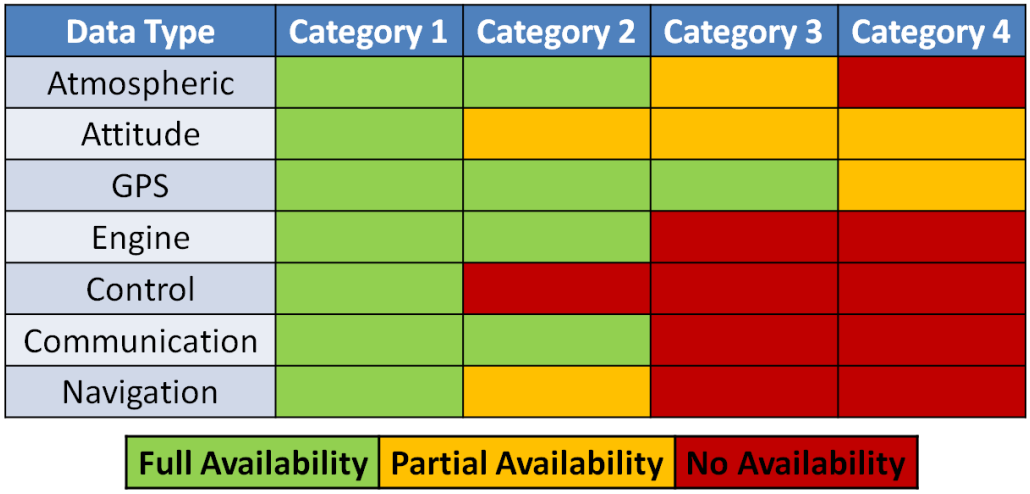


Figure 23: Different categories of data availability in flight data recorders

has become increasingly popular. These can provide access to the GPS position of the aircraft which includes latitude, longitude, and altitude data alongside heading and ground speed. If in addition the pilot utilizes an external attitude and heading reference system (AHRS), such as the commercially available Stratus [16], then additional flight data can be collected by the PED. Similarly, at the higher end of the spectrum are glass cockpit systems such as the Garmin G1000 [65] which are able to log many more parameters accurately. However, they may not record other important factors such as control inputs, weight, flap deflection, etc. Figure 23 sorts the flight data recording capabilities into different categories based on the parameters from various systems that are typically recorded. The capabilities are sorted into four categories from one to four, with decreasing number of recorded parameters and fidelity for higher-numbered categories. The parameters are also divided into different types loosely based on systems in the aircraft. This chart is intended to provide a notional example of each data category - specific recorders may provide more or less capability.

A brief description of the types of parameters in each system is provided here. Atmospheric data refers to data gathered from pitot tubes, barometers, thermometers,

etc. It includes airspeed, wind speeds, pressure altitude, atmospheric temperature, etc. Attitude data refers to roll, pitch, yaw angles and their corresponding rates and accelerations. GPS data contains the latitude, longitude, altitude, and related rates. Engine data contains RPM, Exhaust Gas Temperatures (EGT), Cylinder Head Temperatures (CHT), oil temperature and pressure, fuel flow rates, fuel quantities, etc. Control data contains the deflection of flaps, elevator, aileron, rudder, etc. Communication data includes details about the communication status of the vehicle, such as the comm frequency. Finally, navigation data includes information on any waypoint guidance or autopilot features.

Figure 23 provides a general idea of the spectrum of possible parameters that can be recorded but does not specifically indicate where GA aircraft typically lie. A brief survey of parameters recorded in GA indicates that the actual parameters that can be recorded by typical GA FDRs is rarely at the sophistication of category 1. Commercial aircraft have a certain number of minimum required parameters that need to be recorded as mandated by the FAA [1]. This results in most commercial flight data recorders being in category 1 of the classification provided. There is no such regulatory requirement for GA and therefore, data recorded from GA operations considered in this work fall under categories 2,3,4. The higher end of the GA spectrum such as G1000 can be considered category 2, whereas more common among GA airplanes would be category 3/4 recording capability. It is also noted that some GA aircraft may lack any data recording capability. Therefore, it is noted that while the causes of accidents and events might be related to various sets of parameters, there is a specific constraint on which parameters can be used to define safety metrics in GA operations under consideration in this dissertation.

4.1.4 Summary

Based on the observations from the three sections presented, it is evident that GA FDR are limited in terms of the parameters recorded that can be used for retrospective analysis. Nevertheless, when flight data records are available, some set of recurring parameters are important. The various parameters listed in the previous section are related to different systems of the aircraft.

Some of them relate to the aircraft at the macro level and are a manifestation of the state that the aircraft is in. These are parameters such as ground track, descent profiles (or glide slope), or those related to the energy state of the aircraft such as altitude, airspeed, vertical speed, and associated rates, etc. Others are related to the pilot's inputs or the direct consequences of inputs such as pitch and bank angles, flap setting, RPM (or power setting). Finally, there are some parameters related to the health of the airplane systems such as oil temperatures and pressures, exhaust gas temperatures, fuel quantities, etc.

On the other hand, the different categories of FDR and the type of parameters recorded in those are provided in Figure 23. Metrics that are applicable across the different categories of data recorders yet are able to capture the effects of most of the noted parameters are desired. Deviations in parameters that are overlapping among the different sources investigated here would point to potential safety issues and need to be accounted for by the safety metrics.

4.2 Energy Metrics (Research Question 1.2)

While limited, the parameters recorded in GA can prove to be useful for retrospective safety assessment if utilized in the right manner. Using the parameters identified in the previous subsection various metrics can be defined. The usefulness of the metrics can depend on various factors and motivates the next sub-question which is stated below:

Research Question 1.2:
What safety metrics can be defined using typically available important parameters and criteria that correspond to limits of GA aircraft operations and performance?

In terms of usage in a retrospective safety assessment setting, certain desirable qualities are – parsimony, safety relevance, and generalizability. While there are other qualities that might be important, these three are considered important with the overall objective of the research in mind. This leads to the overarching hypothesis for the first research question:

Hypothesis 1: *Energy-based metrics such as those quantifying the energy state, rates of change of energy, and their margins and deviations are useful metrics that correlate to safety margins and safe operations in GA and can be readily estimated from recorded data.*

4.2.1 Motivation for Energy-based Metrics

This section provides arguments for energy-based metrics to be used for safety assessment based on their appropriateness with respect to the desirable qualities for metrics.

4.2.1.1 Parsimony

The first of the important qualities of metrics is parsimony, i.e., the ability to define the required metrics with a limited number of parameters. This is especially critical in GA safety assessment because of the limitations of flight data recorders. A parsimonious set of parameters used to define metrics as well as a parsimonious set of metrics bring two advantages – (1) it enables flight data collected from various data recorders to be used in the same general framework and (2) it saves on computational time in the case of analysis techniques that do not scale well with higher number of metrics. The energy-based metrics considered in this dissertation are obtained from such a parsimonious set of parameters and others derived from this set. Certain metrics which require extra parameters or information never exceed the requirements of category 2 capability outlined in previous sections. Therefore they are good candidates from the point of view of parsimony.

4.2.1.2 Safety Relevance

Metrics generated from recorded parameters should have relevance to safe operations and safety as identified in the previous sections. Therefore, concepts that correlate well with safety during take-off, approach and landing are preferred since they satisfy the requirement of having relevance. Energy state awareness and energy management are critical concepts in the characterization, detection, and prevention of safety-critical conditions. The FAA has recognized Loss of Control (LoC) and Controlled Flight Into Terrain (CFIT) as the leading causes of fatal accidents in GA [6]. Previous studies have shown that improper or poor energy management and loss of energy state awareness (LESA) are among the top contributors to LoC and CFIT accidents [19]. Ironically, energy state awareness and energy management have been addressed almost exclusively in commercial aviation where the concepts are intrinsic in operational safety and have been the subject of much research but not in GA. The

Commercial Aviation Safety Team (CAST) has identified several safety enhancements that address energy state awareness [32, 134]. The General Aviation Joint Steering Committee (GAJSC) has identified several safety enhancements for new and current GA aircraft intended to improve energy state awareness such as angle of attack systems, stall margin indicators, and stabilized approach indicators [62]. Therefore, it is asserted that energy-based metrics, namely those that characterize the energy state and safety boundary conditions of the aircraft, hold significant potential for improving GA operational safety because they explicitly address poor energy management and state awareness. Similarly, studies from accidents and incidents presented earlier and in literature point to parameters and narratives that further support the relevance of energy-based metrics.

4.2.1.3 Generalizability

The final desirable characteristic of good metrics is that they should be generalizable. Generalizability can be in conflict with parsimony because smaller set of metrics inherently limit how well they are generally applicable or comparable for different aircraft. However, it is asserted that energy-based metrics have no loss of generality in the characterization of aircraft states and safety boundaries across the GA fleet because most of the metrics are normalized by weight and therefore do not depend on the size or weight of the aircraft. Therefore, these metrics may be preferable to flight parameters such as angle of attack, velocity, or rate of descent. Safety boundaries or events expressed with such flight parameters change from one aircraft to another and may need to be modified [56]. This may result in states not being directly comparable. However, energy metrics provide a common and objective currency that is broadly applicable across a heterogeneous GA fleet.

4.2.2 Existing and Newly Defined Metrics

This section provides details of existing and newly developed metrics and the aspect of operations each metric is useful in. The overall argument for the energy metrics based on their appropriateness with respect to the important parameters noted in the previous section is also presented. The literature on energy management outlines two fundamental objectives at the highest level - improving safety and efficiency [101]. Prior to their use in safety analysis, energy metrics were primarily used for analysis of fighter aircraft performance [24, 135, 146, 156]. A classification of the literature on energy metrics based on the aircraft category, phases of flight, and intended purpose of implementation can be found in previous work [117, 119]. Based on the classification it is evident that energy management concepts and some energy metrics have found use in a variety of applications such as development of cockpit displays for alerting pilots [8, 10, 64, 87, 147, 156], control system based on energy management [33, 83–85, 87], trajectory optimization for unmanned vehicles [12], energy management during descent in commercial operations, [43, 59, 142, 147, 148, 150–152] etc.

The work presented in this dissertation is focused mainly on take-off, approach, and landing phases in GA operations and its intended purpose is retrospective safety analysis in a data-driven environment. Such an application of energy-based metrics has not previously been demonstrated. All the energy-based metrics utilized in this dissertation are summarized in Figure 24. Progressively from left to right, the metrics listed in Figure 24 represent increasing information requirement for evaluation.

The metrics in the first column that can be completely evaluated using basic flight data alone. With respect to the categories of flight data introduced earlier (Figure 23), this would refer to any of category 2/3/4 data recording capability. Therefore, this set of metrics is estimable from a broad range of data recorders and satisfies the requirement of generalizability and parsimony for metrics. The metrics in the middle column are contained those metrics for which bulk flight data needs to be available in

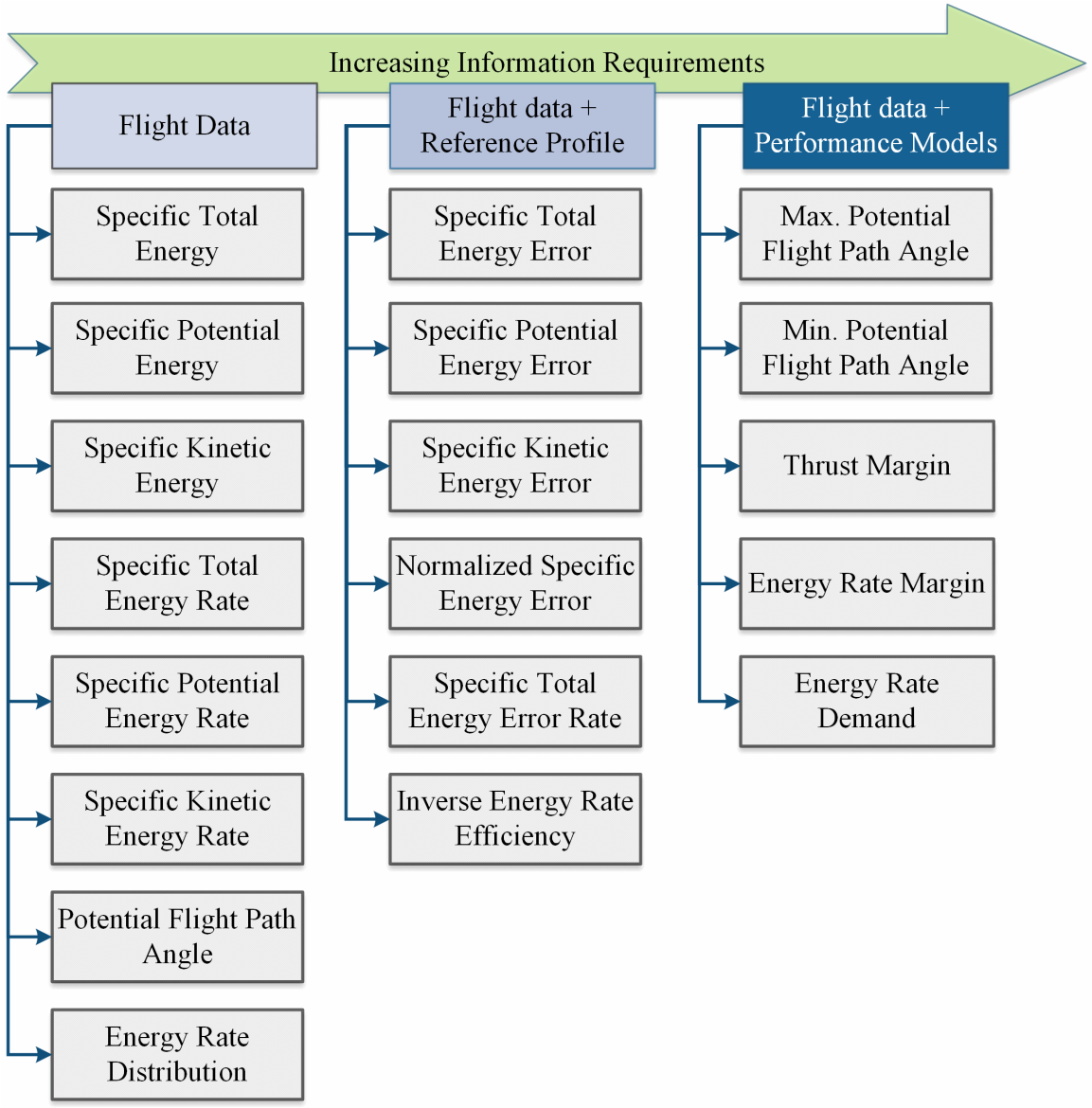


Figure 24: Summary of existing and newly implemented energy metrics utilized in this dissertation

addition to the earlier recording capability of basic flight data. While the parameters recorded might be the same, large amounts of data are required to be able to make statistical inferences and nominal/reference profiles for variation of flight parameters. The final column of metrics on the far right requires the most information in terms of recorded parameters, aircraft performance models, and other derived parameters. This subset is estimable only using category 1/2 flight data recorders (such as the

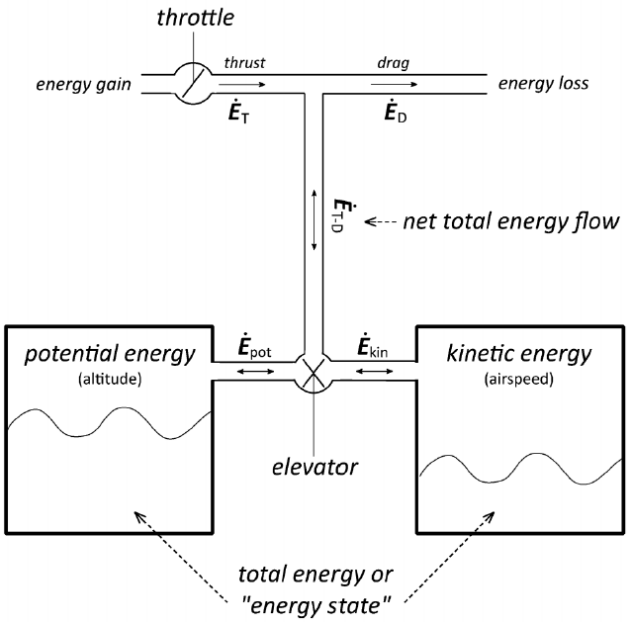


Figure 25: Energy reservoir analogy (adapted from Amelink *et al.* [10])

Garmin G1000).

The metrics listed in Figure 24 are calculated instantaneously during the flight. Therefore, they characterize an *instantaneous* energy state of the aircraft. Other metrics provide measures of energy states, or compliance with nominal states, *aggregated* over a time period. For instance, the mean of the absolute value of energy error measures how well a pilot followed a certain energy profile over a particular phase of flight [64]. For the application considered in this dissertation (which is retrospective safety analysis), instantaneous metrics have been given more importance as the focus is on deviations from safe states. Aggregated metrics are utilized where appropriate but are not defined separately as they are aggregations of instantaneous metrics over a period of time. The remaining part of this section focuses on the formulas and definitions of existing and newly defined energy metrics.

The energy reservoir analogy that was introduced by Amelink et al. [10] is an important concept in the understanding of the origin of many of the energy metrics. The aircraft is visualized as a system comprising of two reservoirs (potential and

kinetic energies). Together these reservoirs represent the total energy state. The energy flow into the system is provided through the throttle (engine) and the energy flow out of the system occurs through the dissipative effects of aerodynamic drag. The net energy flow is the result of the difference between the two, which is then distributed over the kinetic and potential energy flows into and out of the reservoir. The throttle is the energy flow control into the system and the elevator is the valve controlling its distribution. The following metrics are related to the rate, distribution, margins, of various quantities from the energy reservoir. These metrics are manifestations of the energy state of the aircraft as a result of pilot control actions as well as external factors (atmospheric).[2]

1. **Specific Total Energy (STE):**

   Specific total energy is the total mechanical energy of the aircraft per unit weight. It is given by:

$$STE = h + \frac{V^2}{2g} \tag{1}$$

   This is one of the most widely used metrics in literature [20, 24, 83, 101, 156]. This metric has nothing to do with the size or the weight of the airplane (thus, specific), and depends purely on two kinematic quantities. It has been used in a cockpit display concept, in a graphical method to determine the optimum flight profile for an aircraft to reach a certain speed and altitude, in the Energy-Maneuverability Theory to generate "sky-maps" for candidate aircraft.

[2]The research described in this section is also documented in the following publications:

– **Puranik, T.**, Harrison, E., Min, S., Jimenez, H., and Mavris, D., *Energy-Based Metrics for General Aviation Flight Data Record Analysis*, in 16th AIAA Aviation Technology, Integration, and Operations Conference, 2016. Paper No. AIAA 2016-3915, doi:10.2514/6.2016-3915 [117]

– **Puranik, T. G.**, Jimenez, H., and Mavris, D. N., *Energy-Based Metrics for Safety Analysis of General Aviation Operations*, Journal of Aircraft. Vol. 54, No 6. November-December 2017 [119]

2. **Specific Potential Energy (SPE):**

   Specific Potential Energy is defined as the potential energy per unit weight of the aircraft. It is given by:

$$SPE = \frac{mgh}{W} = h \tag{2}$$

3. **Specific Kinetic Energy (SKE):**

   Specific Kinetic Energy is defined as the kinetic energy per unit weight of the aircraft. It is given by:

$$SKE = \frac{\frac{1}{2}mV^2}{W} = \frac{V^2}{2g} \tag{3}$$

4. **Specific Total Energy Rate (STER):**

   Specific Total Energy Rate is defined as the rate of change of Specific Total Energy. It is given by:

$$STER = \frac{d(STE)}{dt} = \dot{h} + \frac{V \times \dot{V}}{g} \tag{4}$$

   It is also called Specific Excess Power [24, 101] and can be alternatively represented as:

$$STER = \frac{(T - D)V}{W} = P_s \tag{5}$$

   Equation 4 is the energy rate from the aircraft kinematic point of view (altitude and velocity), whereas Eq. 5 is from the discipline point of view (Propulsion/Thrust and Aerodynamics/Drag). This metric is perhaps the most important out of all those listed here due to its usage in multiple parts of this dissertation.

5. **Specific Potential Energy Rate (SPER):**

   Specific Potential Energy Rate is defined as the rate of change of potential energy and is given by:

$$SPER = \dot{h} = V \sin\gamma \tag{6}$$

6. **Specific Kinetic Energy Rate (SKER):**

   Specific Kinetic Energy Rate is defined as the rate of change of kinetic energy and is given by:

$$SKER = \frac{d(\frac{V^2}{2g})}{dt} = \frac{V\dot{V}}{g} \tag{7}$$

7. **Potential Flight Path Angle (PFPA):**

   The Potential Flight Path Angle is the theoretical maximum flight path angle that the aircraft can attain at the current throttle setting and configuration. It is given by:

$$PFPA = \gamma_p = \gamma + \frac{\dot{V}}{g} = \frac{T - D}{W} \tag{8}$$

   It is related to the specific total energy rate as:

$$\gamma_p = \frac{STER}{V} \tag{9}$$

   PFPA has been widely used in literature as an energy metric for cockpit displays. Adami et al. [8] and Tadema et al. [145] have used PFPA in their cockpit display concepts. Lambregts et al. [86] have used PFPA and PFPA-max in their work on investigating use of full lateral and vertical control authority for UAV conflict resolution. Lambregts [85] and Kurdjukov et al. [83] have used PFPA in the formulation of the Total Energy Control System. PFPA has been

used under the name *Total Energy Angle* by van den Hoven et al. [147] and Amelink et al. [10] in the description of the *Total Energy Based Perspective Flight Path Display*. PFPA can be considered as an important metric with respect to the recoverability of the aircraft from certain conditions by rapidly exchanging kinetic and potential energies.

8. **Energy Rate Distribution (ERD):**

   Energy Rate Distribution is one of the new metrics defined in this dissertation. It is given by:

$$ERD = sign\left(\frac{SKER}{SPER}\right) \times \exp\left(-|\frac{SKER}{SPER}|\right) \tag{10}$$

   This metric provides an idea of how the energy (incoming or outgoing) is being distributed among kinetic and potential energy reservoirs (refer Figure 25. It can take values between negative and positive one. When the value is negative, it means that one of the energies is increasing and the other is decreasing. A value close to zero indicates kinetic energy rate is much higher than potential energy rate and a value close to unity (positive or negative) vice versa.

9. **Specific Total Energy Error (STEE):**

   Specific Total Energy Error is defined as the difference in specific total energy in actual flight to that of a reference profile. It is given by:

$$STEE = (STE)_{act} - (STE)_{ref} = \underbrace{h_{act} - h_{ref}}_{\text{SPE Error}} + \underbrace{\frac{V_{act}^2 - V_{ref}^2}{2g}}_{\text{SKE Error}} \tag{11}$$

   This metric is very widely used for commercial aircraft, especially in descent and landing where a reference energy profile is available. Williams et al. [152] have used specific energy error to compare various descent trajectories. Jong et al. [43] have used specific energy error as a metric in developing a planning and

guidance concept for optimizing aircraft trajectories during descent. Amelink et al. [10] have used total energy deviation and kinetic energy deviation in the formulation of their *Total Energy Based Perspective Flight Path Display*. Lambregts [87] has used it in his *Energy Management Primary Flight Display* concept. This metric is very useful for defining deviations from a reference trajectory when such a trajectory is available.

10. **Specific Potential Energy Error (SPEE):**

    Specific Potential Energy Error is defined as the difference in specific potential energy in actual flight to that of a reference profile. It is given by:

$$SPEE = (SPE)_{act} - (SPE)_{ref} = h_{act} - h_{ref} \tag{12}$$

11. **Specific Kinetic Energy Error (SKEE):**

    Specific Kinetic Energy Error is defined as the difference in specific kinetic energy in actual flight to that of a reference profile. It is given by:

$$SKEE = (SKE)_{act} - (SKE)_{ref} = \frac{V_{act}^2 - V_{ref}^2}{2g} \tag{13}$$

12. **Normalized Specific Energy Error (NSEE):**

    Normalized Specific Energy Error is the specific total energy error normalized by a tolerance on the specific total energy error. It is given by:

$$NSEE = \frac{(STE)_{act} - (STE)_{ref}}{(STE)_{tol}} \tag{14}$$

    This metric has been used by Gandhi et al. [64] in the development of their "Energy Monitor" display concept. The tolerance in energy error (E_{tol}) is updated dynamically as the aircraft tries to follow a reference approach profile. The authors have developed a crew alerting system which provides various cues when the normalized error exceeds a certain threshold.

13. **Specific Total Energy Error Rate (STEER):**

Specific Total Energy Error Rate (STEER) is one of the existing metrics that has been modified to provide more meaningful insights. It is given by:

$$STEER = sign(STEE) \times \frac{\delta(STEE)}{\delta t} \tag{15}$$

The rate of change of specific total energy error is multiplied by the sign of the error. Therefore, a negative value of STEER is always preferred as this will mean that the error is being driven towards zero. While it is understood that STEER will not always be zero, it is desirable to have this metric within reasonable bounds. It should not be positive for extended periods of time which would indicate that the current energy profile is deviating away from the reference energy profile.

14. **Energy Rate Efficiency (ERE):**

Energy Rate Efficiency is a measure of how closely an aircraft is following the commanded energy profile of an approach trajectory. It is given by:

$$ERE = \eta_{STER} = \frac{STER_{ref}}{STER_{act}} = \frac{V_{ref} W(\gamma_{ref} + \frac{\dot{V_{ref}}}{g})}{V_{act}(T - D)} \tag{16}$$

The Energy Rate Efficiency has been used by van den Hoven et al. [147] to analyze approach trajectories. When this metric is equal to unity, the aircraft is following the commanded trajectory exactly. A value higher than unity indicates a deficit of total energy and value lower than 1 indicates excess total energy than what is required by the approach profile. This metric does not yield meaningful results when there is no ascent/descent or acceleration (such as steady level flight). However, during approach and landing operations in GA, at many points in time the actual total energy rate can be zero. This causes the energy rate efficiency to have sharp peaks and even be undefined at some places (division

by zero). Therefore, in this dissertation, this metric is modified by taking its inverse. It is assumed that the reference profile will never have an actual total energy rate equal to zero during approach and landing.

$$IERE = \frac{STER_{act}}{STER_{ref}} = \frac{V_{act}(T - D)}{V_{red}W(\gamma_{ref} + \frac{\dot{V_{ref}}}{g})} \tag{17}$$

15. **Maximum Potential Flight Path Angle:**

    Maximum Potential Flight Path Angle is the theoretical maximum flight path angle that the aircraft can attain at the max thrust while maintaining current speed and configuration. It is given by:

$$PFPA_{max} = \frac{T_{max} - D}{W} \tag{18}$$

16. **Minimum Potential Flight Path Angle:**

    Minimum Potential Flight Path Angle is the theoretical minimum flight path angle that the aircraft can attain at the idle thrust while maintaining current speed and configuration. It is given by:

$$PFPA_{min} = \frac{T_{idle} - D}{W} \tag{19}$$

17. **Thrust Margin (TM):**

    The thrust margin metric indicates the margin existing between the current thrust magnitude and the theoretical maximum thrust available at that flight condition. It is given by:

$$TM = 1 - \frac{T}{T_{max}} \tag{20}$$

    The thrust margin is an indicator of the amount of energy that can enter the system. Operating at higher margin would mean that the aircraft can escape

possible low-energy scenarios by the aggressive addition of energy. On the other hand, a lower value of this margin would indicate that the capability to add energy is lower. However, it should be noted that high thrust margin does not necessarily indicate a safe flight condition, since it may also be the result of an engine-out situation..

18. **Energy Rate Margin (ERM):**

Energy Rate Margin is another newly developed metric in this dissertation that is defined as the ratio of the actual specific energy rate to the specific energy rate using the theoretical maximum thrust value for the same configuration. It is given by:

$$ERM = \frac{W(\gamma_a + \frac{\dot{V}_a}{g})}{T_{max} - D} \tag{21}$$

During approach and landing, the actual specific energy rate is expected to be negative whereas the maximum specific energy rate will be positive. Therefore, a small negative value (greater than -1) would indicate that the specific energy rate is negative but can be made positive at the current configuration by increasing the thrust. A value less than -1 would indicate that the specific energy rate is negative and the aircraft does not have sufficient margin to make this rate positive. A value greater than zero would indicate that the aircraft has a positive specific energy rate instead of negative. The theoretical upper limit on this metric is +1.

Similarly, during take off, both the actual and maximum specific energy rates are expected to be positive. Therefore, a small positive value of this metric would mean that the aircraft has the ability to increase the energy rate substantially. A larger value (less than 1) would indicate a diminished ability to increase the specific energy rate. A negative value would mean that the aircraft is losing

energy instead of gaining it during the take-off phase.

19. **Energy Rate Demand (ERDm):**

    Energy Rate Demand is the maximum energy dissipation that the aircraft can attain at the current speed and configuration. This is used when the aircraft is descending. It is given by:

$$ERDm = \frac{W(\gamma_c + \frac{\dot{V}_c}{g})}{T_{idle} - D} \qquad \text{Descending flight} \tag{22a}$$

$$ERDm = \frac{W(\gamma_c + \frac{\dot{V}_c}{g})}{T_{max} - D} \qquad \text{Ascending flight} \tag{22b}$$

    When the Energy Rate Demand goes above unity, it indicates that the aircraft, in its current configuration, cannot fly the commanded trajectory. Energy Rate Demand has been used in literature by van den Hoven et al. [147] and Amelink et al. [10]. It has also been used as a constraint by Vormer et al. [148] to ascertain which profile can and cannot be flown by an aircraft during flexible descent trajectory optimization.

All the metrics mentioned in this section are implemented in the current dissertation. In the absence of a reference energy profile, for those metrics that require a reference profile, a data-driven approach can be used to define nominal or reference energy profiles using available data such as the approach presented in Puranik et al. [118] and in the appendix. A summary of all implemented metrics, their formulae, and data required for computation is included in Table 11 (Appendix A).

4.3 Summary

Based on the observations in section 4 and the literature surveyed, the following overarching hypothesis for the first research question is proposed:

Hypothesis 1: *Energy-based metrics such as those quantifying the energy state, rates of change of energy, and their margins and deviations are useful metrics that correlate to safety margins and safe operations in GA and can be readily estimated from recorded data.*

The ability to estimate metrics using varying levels of flight data parameters is quantified in the previous section. Varying levels of richness of data can be used for evaluating different types of metrics which are usable across a heterogeneous fleet of GA aircraft and operations. The relevance of energy metrics to safety is innate as the execution of safe approach and landing as well as take-off is tied closely with managing the aircraft's energy. In addition to this qualitative argument, Table 2 provides a mapping of some common events or maneuvers in GA and their mapping to the relevant energy metrics.

It is hypothesized that deviations in those metrics would be able to capture the corresponding event to which they are mapped and would thus enable quantitative safety assessments. Appendix D provides a table of different metrics and ability to estimate them from recorded parameters. Therefore, the literature search provides the justification and addresses part of the hypothesis and research question related to their relevance. However, it does not, by itself, completely attest to the usefulness in a retrospective safety analysis setting. This can be ascertained by using them in a safety assessment framework and obtaining meaningful conclusions. This is addressed in the next two sections where the metrics are used for identifying non-standard operations in GA and are also used in calibrating GA aircraft performance models for use in

Safety Event/Parameters	**Relevant Metrics**
Unstabilized Approach	SKE, SPE, STE, STER
Excess Vertical Speed	SKE, SKER
Low RPM at Rotation	TM, ERM
Airspeed	SKE
Pitch Attitude	SPER, ERD
Flap Angle at Lift-off	STER
Bank Angle	–
Lateral g-loads	SKER, SPER
Glide slope deviation	SPE, SPER
Centerline Deviation	–
Vertical Speed	SKER
Flap Position	STER
Recoverability and Margins	TM, PFPA-Max, ERD
Adherence to Prescribed Profiles	MTEER, IERE

Table 2: Mapping of some events in take-off, approach, and landing to related energy metrics

safety analysis.

CHAPTER V

NON-STANDARD/ANOMALOUS OPERATIONS (RESEARCH QUESTION 2)

Having identified energy-based metrics as viable candidates for safety analysis metrics in GA operations, the next part of the research focuses on the methods that can be used to identify non-standard or anomalous operations. This is aimed at answering Research Question 2:

Research Question 2:
How can GA flight records be analyzed using energy metrics to automatically identify nonstandard or anomalous operations without a priori knowledge of safety events?

It is noted that in the context of this dissertation, non-standard operations can refer to a variety of things. It could refer to exceedances and events defined by operators or regulatory bodies, deviation from standard operating procedures set by operators, actions leading to incidents or accidents, adverse weather conditions, unique events such as airshows, etc. These non-standard operations are not necessarily unsafe. They simply represent a deviation from nominal sufficiently high enough that further investigation might be warranted. In a data-driven quantitative safety assessment environment, it is important to be able to identify non-standard behavior without any prior knowledge. This enables handing the decision-maker or analyst with the most relevant information for making judgments or proposing solutions.

Methods for identifying non-standard or anomalous operations are abundant in literature within aviation as well as outside. Some of the relevant methods were

surveyed earlier in the literature review section. Gavrilovski et al. [67] provided a taxonomy of techniques from the data-mining and machine learning domain which can and are applied to different problems in aviation safety. The present work focuses on identifying non-standard operations from within routine operations data without a priori information. This is a problem well-suited for the class of unsupervised and semi-supervised techniques encompassed in the *anomaly detection* domain.

Background research revealed that anomaly detection techniques have shown great potential for identifying safety-related issues without any prior knowledge of events. Offline analysis of data offers the ability to analyze and discover something new or off-nominal and gain a holistic understanding of events. Anomaly detection is defined as – *the task of obtaining patterns in data that do not conform to a well defined notion of normal behavior* [35]. Anomaly detection has been widely implemented in a variety of domains generally with a high degree of success [35, 91]. Anomaly detection is important because offline detection of these anomalies allows preventive measures that can save lives whereas online detection of the anomalies is critical for fault diagnosis. Anomalies can stem from various reasons, some of them could be harmless such as faulty calibration and non-standard equipment/sensors whereas others could have serious safety implication such as system malfunctions, incorrect landing configuration etc. Identifying anomalies that are operationally significant and can contribute to the implementation of proactive measures is the main aim of using anomaly detection techniques on aviation data.

Unlike other applications of data mining or anomaly detection, aviation data is typically not labeled. This means that there is no knowledge a priori as to which flight records (if any) are actually anomalous. Also, there is no universal definition for what constitutes an anomaly in this context. Therefore, unsupervised or semi-supervised algorithms need to be used to identify anomalies. Chandola [34] looked at the general anomaly detection problem for symbolic sequences and time series data.

Identifying anomalies as outliers of a clustering algorithm is useful way of anomaly detection as it allows the possibility of multiple standard patterns. Alternatively, one-class learning algorithms can also be used for identifying outliers. There are a number of ways in literature by which the "anomalousness" (or anomaly score) of an outlier can be quantified. Some anomaly detection algorithms directly provide an anomaly score while others use an external metric such as Local Outlier Factor (LOF) [25] to provide the anomalousness. Campos et al. [31] have provided a review of some of these scores using different data sets to quantify their relative performance. In most cases, the performance of an anomaly score is dependent on the type of data.

It was identified here (and also seen in other work such as [89, 91, 99, 100]) that there are two main types of anomalies observed in multi-variate time series data – collective or *flight-level anomalies* and point or *instantaneous anomalies*. In flight-level anomalies, an entire flight or phase of flight is abnormal or deviant. This type of anomaly therefore can be typically defined only for well-defined phases of flight such as take-off, approach, and landing. On the other hand, instantaneous anomalies refer to a small subset or instant(s) within a flight that are abnormal and as such can be defined in any phase of flight. Both types of anomalies are very important as they correspond to different aspects of flight safety.

In general, different types of techniques are more effective in identifying different types of anomalies [35]. Various 'no free lunch' theorems have been proved in literature which state that for any algorithms or techniques, elevated performance over one class of problems is offset by degraded performance over another class [154]. In the GA flight safety domain, techniques that will have elevated performance and generalize well within the GA fleet are desirable and their generalization to other classes of problems is not necessarily of interest, especially since other applications such as commercial aviation already have bespoke techniques and algorithms.

Some of the major hurdles in anomaly detection techniques include availability

of actual data, computational cost, selection of appropriate methods, interpretation of obtained anomalies, etc. In the work presented in this dissertation, availability of flight data is assumed. On the other hand, the use of energy metrics as features in anomaly detection will enable better interpretation of the identified anomalies as they can be related to the energy state of the aircraft. Within the aviation safety literature, several factors have been identified that can affect the choice of anomaly detection technique. Table 5 lists some of the important factors and choices that an analyst has. Depending on the specific problem at hand, these choices get limited prior to the analysis.

Factor	**Choices**
Type of Data	Categorical, Discrete, Continuous
Features Used	Raw Parameters, Derived Metrics
Types of Anomalies Detected	Instantaneous, Flight Level, Contextual
Algorithm Supervision	Supervised, Semi-supervised, Unsupervised
Fleet Heterogeneity	Single aircraft Type, Multiple Aircraft Types

Table 3: Factors affecting choice of anomaly detection technique (those applied in this dissertation are underlined)

Data obtained from GA operations consists of multi-variate time series of varying duration. Because of the variability in data recording capabilities (categories 2–4), the fidelity of data and the number of parameters obtained from different flights may not necessarily be consistent. Almost all parameters recorded correspond to the outputs or effects of different actions (pilot actions, weather, etc.) on the system. Due to the lack of data corresponding to inputs or actions taken in the cockpit, it is typically much harder to attribute causes while analyzing GA flight data.

Based on the observations in literature and the nature of the problem at hand the following over-arching hypothesis is proposed for research question 2:

Hypothesis 2: *Using the defined energy metrics as features in a general anomaly detection framework with appropriate techniques will enable the detection of both types of anomalies.*

In the second hypothesis mentioned above, the word enable is used in a specific context. While the use of energy metrics as features is not a requirement for a general anomaly detection framework, in this particular context, using the metrics results in the possibility of a broad spectrum of data recorders, aircraft, and operations being included. Therefore, the generalizable metrics allow all the heterogeneity to be included in a single framework and thus 'enable' identification of both types of anomalies. A general anomaly detection framework as adapted for the current dissertation is shown in Figure 26. The elements of this framework are consistent with other anomaly detection applications such as those surveyed by Chandola et al. [35]. The overall framework remains the same for identification of both types of anomalies, but the components in each module are modified according to the problem requirements.

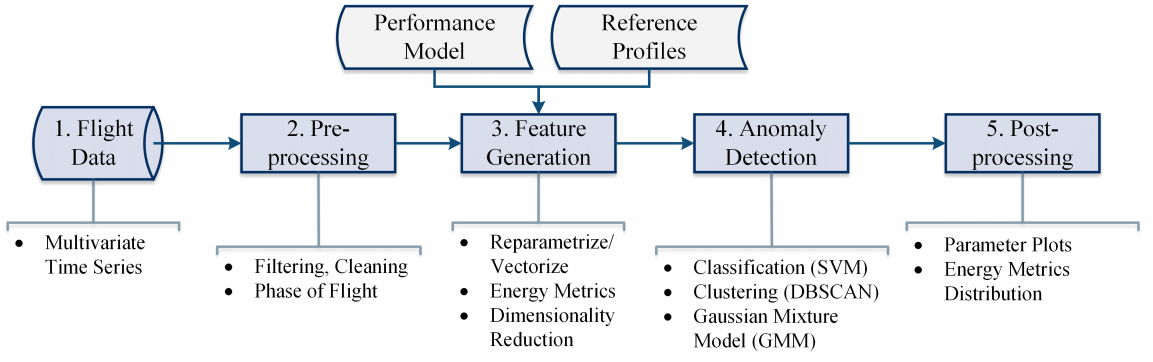


Figure 26: General framework for anomaly detection including the enumeration of specific techniques utilized within this dissertation

There are two external inputs to the framework that are not usually present in the anomaly detection framework from the literature surveyed – *performance models* and *reference profiles*. Further details on performance models are provided in Chapter 6. It was identified previously that obtaining additional information can enhance the

effectiveness of the safety analysis task. Predictions provided by aircraft performance models add this value to the analysis. Reference profiles are nominal variations of energy metrics over certain phases of flight, which can be used to calculate certain energy metrics in those phases of flight. In some cases, these profiles are readily available. For example: the altitude profile that aircraft try to follow during final approach and landing consists of a nominal (3°) glide slope. This corresponds to a certain variation of the specific potential energy profile. On the other hand, when such profiles are not as easily available, they can be calculated using a data-driven approach by averaging over a large number of flight records to get nominal profiles which can be used as a reference. This approach has been elaborated in prior work (Puranik et al. [118]) and the associated calculations and logic for reference profiles and other derived parameters are provided in Appendix C. The following sections will elaborate on appropriate techniques for each of the two types of anomalies and the corresponding analyses performed.

It should be noted that, throughout this dissertation, the terms "anomalous" or "abnormal flights" refer to flights that differ significantly from the majority in terms of certain critical parameters and metrics. This does not necessarily mean that the flights are inherently unsafe. Whereas some of these may be benign flights that simply followed different procedures, others could point to potential safety issues, which are of interest to safety analysts. The identification of these emergent outliers is one of the aims of this work. The main purpose of identifying these anomalies is the retrospective safety analysis of flight data from heterogeneous GA operations. The methodology presented in this research can aid industry experts and GA operators in better understanding and identifying unsafe practices. It could be used to improve flight training and instruction for student pilots, as well as deployed on a large database of flights, such as the National General Aviation Flight Information Database (NGAFID), to understand trends and behaviors.

5.1 Flight-Level Anomalies – Research Question 2.1

The first type of anomaly considered in this dissertation is flight-level anomaly. The associated research sub-question is stated as follows:[1]

Research Question 2.1:
How can a general anomaly detection framework be modified to identify flight-level anomalies in GA take-off, approach and landing operations using energy metrics?

For identifying flight-level anomalies, well-defined phases of flight are considered. Due to the heterogeneity in operations in GA, sometimes the phases of flight are not very easy to define. Goblet et al. [68] have provided definitions for common phases of flight in GA operations and their identification based on recorded parameters. The releveant definitions from this paper have been reproduced in the Appendix C. In GA operations it was noted earlier that a large percentage of the total accidents occurs during take-off, approach and landing phases [6]. These two phases can also be defined in a relatively straightforward manner. Therefore, in this dissertation identification of flight-level anomalies is focused mainly on these two phases of flight. The subsequent sections outline the details of each of the steps of the anomaly detection framework in Figure 26 as applied to identification of flight-level anomalies. The choices made for each step and the modifications or improvements made for GA flight data include the core of the developed methodology for this research sub-question.

[1]The research described in this section is documented in the following publications:

– **Puranik, T. G.**, and Mavris, D. N., *Anomaly Detection in General-Aviation Operations using Energy Metics and Flight-Data Records*, in Journal of Aerospace Information Systems, Vol. 15, No. 1, pp 22–35, January 2018, doi:10.2514/1.I010582. [120]

– **Puranik, T. G.**, Jimenez, H., and Mavris, D., *Utilizing Energy Metrics and Clustering Techniques to Identify Anomalous General Aviation Operations*, in AIAA SciTech Forum, 2017. Paper No. AIAA 2017-0789, doi:10.2514/6.2017-0789 [122]

Step 1 (Flight Data):

The flight data obtained from the DFDR (Digital Flight Data Recorder) are a multivariate time series, whose lengths typically vary between records due to varying duration of flight. The number of parameters can vary from around 20 (in GA operations) to thousands (commercial operations). These parameters are recorded at a specific frequency (e.g., once per one second interval). Typically this contains data from routine operations such as personal flying, pilot training, etc. The experience level of GA pilots is highly variable, ranging from just a few hours of experience to thousands of hours of flying. The aircraft typically have different make/model/tail number but generally belong to the same *normal* category as defined by the FAA [2]. Each data record contains approximately 30 continuous parameters related to the state, attitude, basic engine information, environmental conditions, GPS, and others (categories 2–4 from the previous section). Typically, due to the presence of only continuous data in the set, feature generation and pre-processing are streamlined and made uniform across all parameters, and metrics and special techniques requiring the handling of heterogeneous (binary, continuous, categorical, etc.) data are not required. The parts of the flight data record that correspond to takeoff or approach and landing are extracted for further analysis.

Step 2 (Pre-processing):

In all techniques in literature, the raw data obtained from flights is pre-processed to obtain data suitable for anomaly detection. Noise that may be present in the original data is smoothed using a simple moving average filter to remove spikes that are caused due to noisy sensor readings. Following that, relevant phases of flight are extracted from the flight data – take-off, approach and landing. It is noted that in GA flight data records, the exact touchdown point of the aircraft on the runway is not recorded

(in commercial DFDR it is recorded as a binary parameter Weight-on-Wheels). Previous work done on flight data analysis and extracting derived parameters from flight records (Puranik et al. [118]) addressed this issue by using altitude difference between successive points in the flight data along with combinations of other parameters to obtain the touchdown point on the runway. The process of obtaining the touchdown point along with a demonstration of the implementation is provided in Appendix C.

Step 3 (Feature Generation):

In anomaly detection algorithms for flight-level anomalies, the flight data records or phases of flight are characterized using *features*. These features are typically derived from the data obtained from the flight record and account for the temporal aspect of data. Anomaly detection algorithms then compare features of different flights to each other to find anomalies. There are two important requirements on these feature vectors – Firstly, the feature vector generated must have the same dimensions for all flights. Secondly, the corresponding elements of the vectors of different flights should be comparable. These two requirements give rise to multiple ways of mapping the multivariate time series to a single vector.

Das et al. [41] use the technique of Symbolic Aggregate ApproXimation (SAX) [92] to convert a continuous time series of parameters during approach into discrete symbols by averaging the data over a number of seconds into a single symbol. An advantage of this type of approach is that approach profiles of varying duration can be incorporated as the total number of symbols is fixed. However, approximating continuous time series by discrete symbols usually results in loss of information due to down sampling. Also, this approach works well when the time series associated with the phase of flight is of a large duration (which is not the case for GA approach and landing or take-off). Another approach used by Li et al. [89] is to anchor the flight data record at certain significant points (example – application of take-off power in take-off

and touchdown on runway in approach and landing). From the anchor point, the multivariate time series in take-off are re-sampled at a fixed temporal interval whereas in approach and landing, they are re-sampled based on the distance remaining to touchdown. The reason for re-sampling in approach and landing based on distance rather than time is that many of the procedures in this phase are distance specific. This approach is advantageous because each feature vector thus obtained is of the same dimension without down-sampling or loss of information. A potential disadvantage of this approach is that this requires the ground-track distance of the approach and landing phase to be approximately similar in each case. Other approaches to feature generation include Piecewise Aggregate Approximation (PAA) which is similar to SAX except that the down-sampled series is not converted to a series of symbols. This approach would suffer from similar issues to the SAX approach.

In the current dissertation, the anchor point of take-off phase is the application of take-off power, while for the approach and landing it is the touchdown point on the runway. Since both the touchdown point and ground track distance remaining are not explicitly recorded in GA flight data, they are estimated using different techniques shown in Appendix C. The detailed approach followed in this dissertation for feature generation is shown in Figure 27.

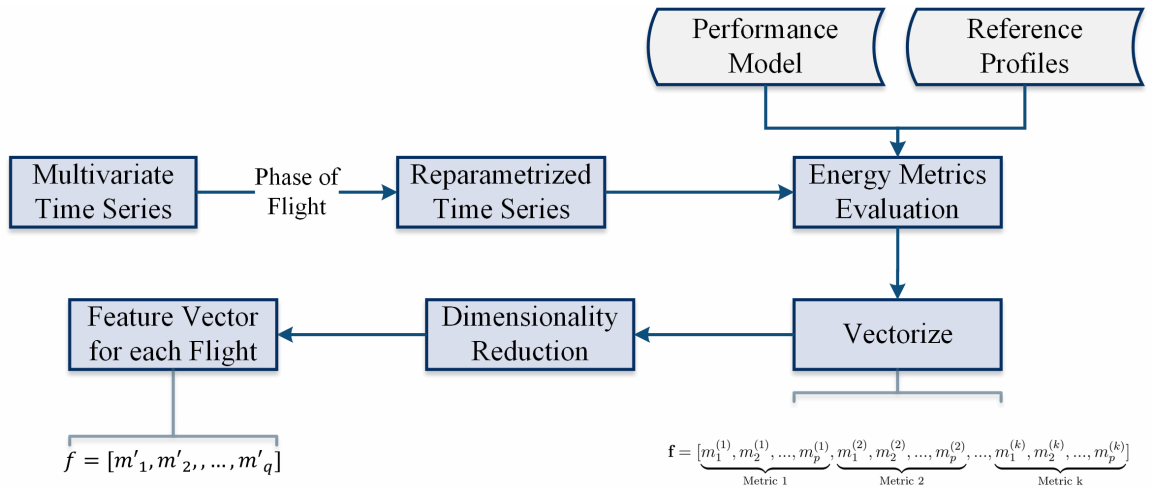


Figure 27: Process of feature vector generation from flight data

The reparametrized time series for the phase of flight in consideration in used to evaluate the energy metrics at each point. Therefore, each flight record has a fixed number of points at which each energy metric is evaluated. This ensures that the dimensionality of each feature vector will be exactly the same. The feature vector for each flight record can be represented by a single vector by concatenating the contribution from each metric as:

$$\mathbf{f} = \left[\underbrace{m_1^{(1)}, m_2^{(1)}, ..., m_p^{(1)}}_{\text{Metric 1}}, \underbrace{m_1^{(2)}, m_2^{(2)}, ..., m_p^{(2)}}_{\text{Metric 2}}, ..., \underbrace{m_1^{(k)}, m_2^{(k)}, ..., m_p^{(k)}}_{\text{Metric k}} \right] \tag{23}$$

where m_j^i is the value of the i^{th} metric at the j^{th} distance-based location or time-stamp in the approach and landing or take-off. Thus, the j^{th} element of each vector is now comparable to every other vector. Even though all the feature vectors have the same length, different metrics have different magnitudes. Therefore, each metric is normalized such that it has a zero mean and unit variance. This is achieved using z-score normalization. Let $m_{j,1}^{(i)}, m_{j,2}^{(i)}, ...m_{j,n}^{(i)}$ be the values of a particular metric i for all n flights at a particular distance j from the runway. The z-score normalized values are given by:

$$m_{j,1}^{(i)} = \frac{m_{j,1}^{(i)} - \overline{m_j^{(i)}}}{\sigma(m_{j,1...n}^{(i)})} \tag{24}$$

where σ denotes the standard deviation of observations. The feature vectors obtained can contain hundreds of dimensions. However, while identifying outliers and clusters, the variability is typically embedded in a smaller number of dimensions. Principal Component Analysis (PCA) is a linear transformation that is used to transform data into a new orthogonal coordinate system [74]. The coordinates in the new system are ranked in descending order of the amount of embedded information (variance) they contain. Dimensionality reduction is achieved by only retaining the first few components that explain majority of the variance (in this method the number is

chosen such that 99% of the variance is captured). If the reduced dimensional vector contains p elements, each flight record can now be represented as:

$$\mathbf{f'} = \left[m'_1, m'_2, ..., m'_p\right] \tag{25}$$

The dimensionality reduction comes at the cost of losing the physical meaning of the components of the vectors thus obtained. Dimensionality reduction techniques such as PCA become intractable when the size of the data set becomes larger [11,89]. The number of energy metrics is typically much smaller than the number of parameters recorded (even in GA flight records). Therefore, this step may be skipped for this approach if the size of the data set becomes intractable. It is noted that, while PCA or other dimensionality reduction techniques may speed up computational time, there should be minimal to no loss of information regarding which flights are identified as anomalous. To that end, a comparison is made of the set of anomalous flights obtained with and without PCA for the same feature vector choice. Details of this comparison are provided in Appendix E.

Step 4 (Anomaly Detection):

The next step in the framework is to use the feature vectors thus generated in an anomaly detection algorithm. A number of anomaly detection algorithms have been proposed in the literature for different types of problems. Various considerations need to be taken into account when selecting the most appropriate anomaly detection techniques. Since the data obtained from routine operations is not labeled, there is no prior knowledge of anomalies in the flight data (if there exist any at all). Therefore, only semi-supervised or unsupervised learning techniques may be utilized. However, even with these techniques, the number of anomalies is expected to be a small fraction of all the flight data records. This information should be utilized in tuning the parameters of the algorithm. In the literature, classification (in the form of one-class

learning) [40, 41, 99] and clustering [18, 27, 89, 99] have been utilized extensively for anomaly detection in flight data records. In the current work, both clustering and classification have been used together to identify anomalous flight data records.

In *clustering-based* approaches, a clustering algorithm is applied on the feature vectors using a function to compare the dissimilarity between feature vectors. One of the most common functions used in clustering literature is the euclidean distance. Let the feature vectors corresponding to two flight records be $f_1 = [m_1, m_2, ..., m_p]$ and $f_2 = [n_1, n_2, ..., n_p]$. Then, the euclidean distance between f_1 and f_2 is given as:

$$D(f_1, f_2) = \sqrt{\sum_{i=1}^{p} (m_i - n_i)^2} \tag{26}$$

Other measures of distance or dissimilarity may also be used. In typical GA operations, the number of clusters present in the data is difficult to predict a priori. Therefore, algorithms which automatically identify the number of clusters (such as density-based clustering) are preferred over those which require specification of the number of clusters (such as k-means, k-medoids etc.) In the current work, the popular density-based clustering algorithm DBSCAN is used [50]. Given a set of points (flight data records), DBSCAN groups together instances that are closely packed together while marking points in low-density regions as outliers. A cluster forms when there are at least a minimum number of points (hereafter called $MinPts$) within a user specified threshold (hereafter called ϵ) of a given point. Clusters grow when additional points satisfy the density criterion specified by the algorithm until all the points have been allotted to a cluster or labeled as outliers.

There are two parameters that need to be supplied to DBSCAN - ϵ and $MinPts$. ϵ depends heavily on the similarity function used, normalization of data, and other factors. In many cases, rather than providing ϵ, its value is varied from the minimum distance observed among all flights in the data set to the maximum distance observed. Instead of ϵ, the user provides the proportion of flights that will be marked as outliers.

This number has a direct correlation with the value of ϵ but is more intuitive to the user of the methodology. $MinPts$ on the other hand, depends on the homogeneity of operations and how similar flights are to each other in terms of the features chosen. Typically, $MinPts$ has a less significant effect on the algorithm than ϵ if it is within a nominal range of values [50, 89].

The main advantage of DBSCAN is that it has the ability to automatically determine the number of clusters and also detect outliers (anomalies) based on a user specified threshold. One of the drawbacks of DBSCAN is that its performance can suffer if there are multiple clusters with varying densities. Another drawback is that the algorithm needs to calculate the pairwise distance between all flights which can be computationally intensive. Also, in order to get a specific proportion of flights as outliers, the algorithm needs to try out different values of ϵ. Despite these limitations, the algorithm's most important quality of automatically determining the number of clusters will be utilized in this work. The algorithm provided by Ester et al. [50] is implemented in MATLAB for use in this work.

For *classification-based* approaches, a one-class classification model is trained assuming the available data set as nominal. One of the most powerful algorithms that is widely used is one-class Support Vector Machines (SVM) [136]. SVMs compare feature vectors using functions known as *kernels*. Kernel functions map pairs of feature vectors to the similarity between those vectors, with a value of 1 indicating maximum similarity and 0 indicating no similarity. Das et al. [41] used normalized Longest Common Subsequence (nLCS) kernel on discrete data and SAX-discretized data (which converts a continuous time series of parameters into discrete symbols by averaging the data over a number of seconds into a single symbol). The limitations of this approach were highlighted previously in the feature generation step. Therefore, in this work, Radial Basis Function (RBF) kernel, one of the most popular kernel functions, is used. Let the feature vectors corresponding to two flight records be

$f_1 = [m_1, m_2, ..., m_p]$ and $f_2 = [n_1, n_2, ..., n_p]$. The RBF kernel is given by:

$$K(f_1, f_2) = \exp\left(-\frac{||(\boldsymbol{f_1} - \boldsymbol{f_2})||}{2\sigma^2}\right) \tag{27}$$

σ is a parameter that defines the scale of the classifier. The one-class SVM then uses this kernel function to find a decision boundary (by solving an optimization problem) to detect outliers. If the decision function predicts a negative label for a given test point z, then it is classified as an outlier. One of the main advantages of training a SVM model is that upon training, the prediction of anomalies is much quicker as it only has to use a fraction of the training points (called support vectors) in the prediction stage. The balance between model complexity and over-fitting can be managed by varying the proportion of points retained as support vectors. Details of the optimization problem and SVM implementation can be found in Schölkopf et al [136]. In the current work, the one-class SVM implementation in MATLAB is utilized and tuned for the GA flight data [98].

Figure 28: Use of anomaly detection algorithms in this dissertation

Figure 28 shows how the algorithms are used in the current work for identifying flight-level anomalies. Compared to clustering methods, model-based methods like classification can be computationally more efficient in the detection stage by not requiring comparison of the new flight data record with each existing record in the database. In the present work, both DBSCAN clustering and one-class SVM are utilized. On an existing set of flight data records, DBSCAN is used to identify the approximate number of clusters in the existing data set. Each flight data record in the data set is then assumed to have a label of the cluster to which it belongs and this information is used to train a support vector machine model for each cluster identified. The trained SVM model(s) are then used to identify outliers from each class. The anomalousness of any

existing or new flight data records can thus be directly obtained using the SVM score. Using both algorithms in this manner provides a consensus and enables combining the strengths of both algorithms while trying to overcome their individual weaknesses.

Step 5 (Post-processing):

Once the anomaly detection algorithm has identified flight level anomalies, the post-processing step involves visualizing them and organizing according to their severity. One of the main advantages of using large amounts of data is that statistical inferences can be drawn more reliably. Visualizing the performance of each flight record compared to every other flight can provide valuable insights for training pilots. Since each energy metric provides information about a different dimension of the safety space, plots of the energy metrics can provide a more holistic view of the safety state. Combinations of energy metrics going outside permissible bounds could be potentially more dangerous than the exceedance of individual metrics/parameters. The anomalousness of each flight data record can be visualized using anomaly scores such as LOF [25] for clustering algorithms, whereas for one-class classification, the support vector machine model provides an anomaly score for each record which can be visualized for this purpose. In the present work, the anomaly score provided by each one-class SVM is utilized for post-processing and visualization.

5.2 Experiment 2.1

The steps outlined in the previous section enable building the methodology for identification of flight-level anomalies in GA operations. In order to demonstrate the implementation of the methodology and address the research sub-question 2.1, the experiment described in this section is performed.

5.2.1 Purpose of Experiment:

The following are identified as the main objectives of this numerical experiment:

1. Demonstrate the usage of energy metrics in a retrospective safety assessment setting, in this case identification of flight-level anomalies

2. Demonstrate automatic identification of flight-level anomalies in take-off, approach and landing phases

3. Demonstrate identification of known anomalous flights inserted into the data-set using a flight simulation model

4. Provide a flight data record anomaly-severity score that can help prioritize critical anomalies compared to benign ones

5. Demonstrate flexibility in developed methodology by allowing identification of anomalies at different levels of significance based on available data-set

In order to facilitate validation efforts listed in point 3, artificial flight data records using a flight simulation model are generated. These simulated records are flown by a private pilot and consist of three approaches and landings corresponding to (1) a nominal approach (2) an unstabilized approach with poor energy management and a (3) high approach. Simulations (2),(3) will enable identification of known anomalies, whereas application of the algorithm to simulation (1) will verify whether it is indeed classified as normal. Similar validation has previously been done by Chu et al. [37]

at NASA using a flight simulator for commercial aircraft operations. This method of using simulated flight data is useful because routine operations are not expected to contain a lot of unsafe situations as all data is from flights that have landed safely. Using simulated data enables performing some amount of validation while working within cost and safety constraints.

5.2.2 Experiment Setup:

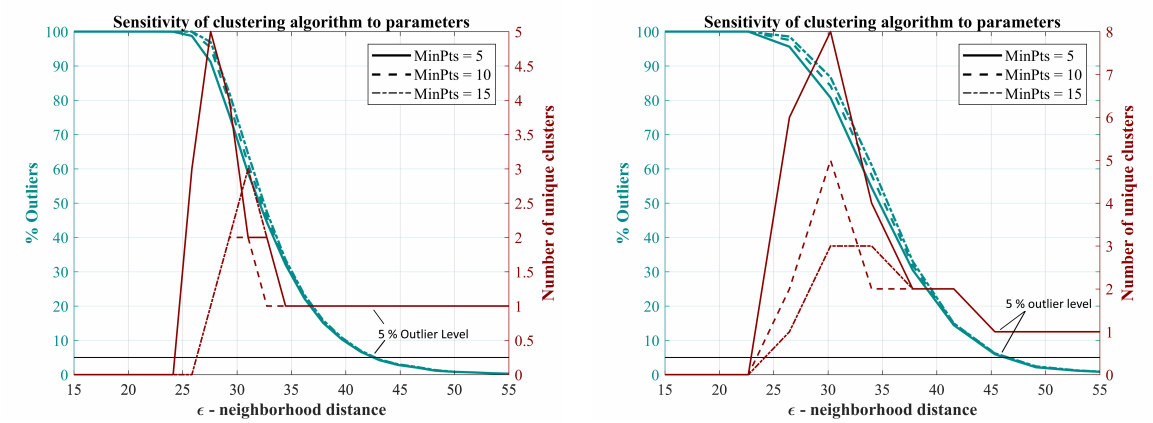
The setup of this experiment consists of around three thousand flight data records obtained from two different flight schools and collected during routine operations. The data is from two different types of aircraft (Cessna 172S and Piper Archer) operated at multiple GA airports. There are multiple tail-numbers of each aircraft type within the data-set. The data is pre-processed using the techniques developed in this methodology and additional derived parameters are obtained using techniques mentioned in Appendix C. The features used in this experiment consist of the complete set of energy metrics surveyed earlier. Anomaly detection using a limited number of metrics and their effect on results is explored in a later experiment.

5.2.3 Results:

The first step in flight-level anomaly detection is identifying the approximate number of clusters that are present in the data set under consideration. One of the reasons for using energy metrics rather than raw parameters is that the metrics are expected to serve as an objective currency across different aircraft. Therefore, in GA operations using similar aircraft such as those under consideration, it is expected that using energy metrics will yield a single main cluster of flights along with outliers.

5.2.3.1 Clustering Results

Figure 29 shows the results obtained from density-based clustering in the approach and landing phases using (a) Energy Metrics and (b) Raw Parameters.



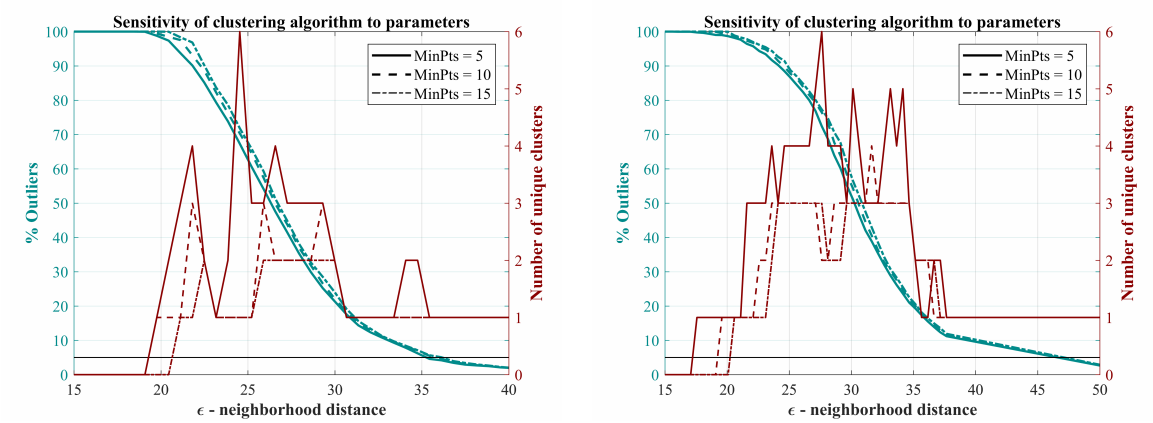
(a) Landing - Energy Metrics

(b) Landing - Raw Parameters

Figure 29: Sensitivity of clustering algorithm to tuning parameters and number of unique clusters obtained

It is observed that, for all values of MinPts, as the value of ϵ increases from the minimum to the maximum value, the proportion of flights that are identified as outliers steadily decreases. This is expected as more points will be included in the clusters when the cluster radius is increased. The curves for different values of MinPts eventually collapse almost into a single curve as ϵ is increased further. For the type of data dealt with in aviation safety, the proportion of anomalous or abnormal flights is expected to be very low (as seen from the low accident and incident rates per million flight hours, even in GA). Therefore, at the values of ϵ that correspond to low outlier percentages, any value of MinPts from the set chosen gives almost equivalent results. This important observation leads to the conclusion that the value of MinPts can be set to a default of ten for the purpose of this work.

The second set of curves (red) from Figure 29 corresponds to the number of unique clusters at each setting of ϵ and MinPts. The trend observed is that, as the value of ϵ increases, the number of clusters settles at one after the initial oscillation. One reason for this oscillation is the nature of the DBSCAN algorithm itself, in particular, the manner in which it starts forming clusters. At lower values of the neighborhood



(a) Take-off - Energy Metrics

(b) Take-off - Raw Parameters

Figure 30: Sensitivity of clustering algorithm to tuning parameters and number of unique clusters obtained

distance, individual points in dense neighborhoods start their own cluster by becoming 'core' points. As the distance increases, clusters consolidate until a new core has enough points around itself to start its own cluster. This process of new cluster formation and consolidation causes the initial oscillations, as seen in Figure 29. However, at the outlier significance levels of interest ($\approx 5\%$ or less), there is only a single cluster present among the current set of flight data records as noted by the annotation in the figures. It is understood that this may change as more flights are added to the set from other aircraft/ airports, etc. Therefore, the clustering algorithm can be periodically rerun to update this and select the appropriate number of clusters. For the one-class SVM, *all* the flights will be given a default label of one prior to fitting the model (since the clustering experiment has indicated that there is roughly a single cluster present among the data). The outliers from this SVM will thus be the anomalies of interest identified for further inspection.

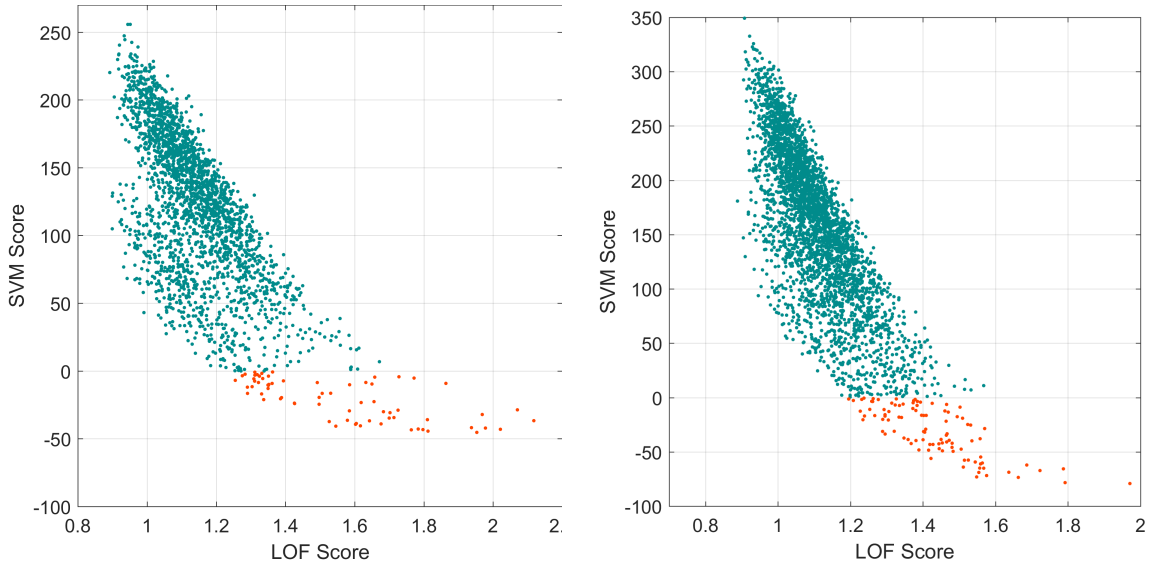
It is noted that in the clustering parameter sensitivity experiment, higher fluctuation is observed in the earlier parts for the raw parameter feature vector usage. This is because, there is expected to be more variability in the raw parameters even for such similar aircraft. However, at the significance levels of interest, this variability

tapers away to form a single major cluster. Similar to approach and landing, the clustering experiment can also be performed for the take-off phase using the same methodology. The results from application of DBSCAN clustering in take-off phase are shown in Figure 30.

The results from the take-off phase follow trends mostly similar to those of approach and landing with the exception that the number of clusters remains more than one for more neighborhood distance options than approach and landing. However, once again, at the significance levels of interest, a single cluster forms when using energy metrics and raw parameters.

5.2.3.2 Score Comparison

Once the approximate number of clusters is identified for both take-off and approach and landing, a one-class SVM can be trained for each cluster identified. Any new flight data that needs to be analyzed can then be evaluated for belonging to any of these defined clusters. SVM for both phases of flight are trained and the outlier scores for each flight are evaluated. The outlier score obtained using support vector machine can be directly used to predict anomalous flights or a different measure can be utilized. LOF [25] is one of the popular alternative method of obtaining the outlier score in a data-set by directly comparing the distance between different flights using a defined measure (e.g. Euclidean distance). Therefore, a comparative assessment is provided between the LOF score and SVM score for quantifying the severity of the anomaly. While the SVM score provides a clear boundary between normal and anomalous (values less than zero are anomalous and greater than zero are normal), the LOF score works in a different manner. A higher LOF score indicates an outlier whereas scores closer to unity are considered normal. Figure 31 provides a comparison of the LOF and SVM anomaly scores obtained for all flights in the take-off and approach and landing phases.



(a) Take-off (b) Approach and Landing

Figure 31: Scatter-plot showing comparison of LOF outlier score and anomaly score obtained from SVM for take-off and approach-and-landing phases

As is evident from the figure, there is a high negative correlation between the two ways of scoring anomalies as is expected. The anomalous flight records (red dots) from SVM also score high on the LOF anomaly score. As the SVM score decreases (more anomalous), the LOF score increases for all flights. However, as the LOF score does not provide a natural boundary for anomalies like the SVM score it introduces subjectivity in choosing the cut-off threshold. The addition of any new flight data to the existing set would require evaluating its distance from each flight which can add to the computational burden as the size of the data set increases. SVM score on the other hand does not require a lot of computational resources in the prediction phase as a linear (or non-linear) model has already been trained. Therefore, while either of these scoring systems may be used, the score directly obtained from the SVM is utilized in this work. For the results of the remainder of the experiment, one section each for take-off and approach-and-landing each are presented.

5.2.3.3 Approach and Landing Phase

After training the SVM model, it can be used for prediction of anomalies within the current data set. The approximate level of outliers can be specified while fitting the model. When new data becomes available, the trained model can be directly used for prediction. The anomaly scores for all flights in the approach and landing phase using SVM are shown in Figure 32. The horizontal dashed line at score of zero represents the boundary between normal and anomalous flights. The scores for a specific set of flights are annotated in the figure. These include two simulated anomalous flights, one simulated nominal flight, two sample anomalous flights from data set with low anomaly score and potential safety implications.

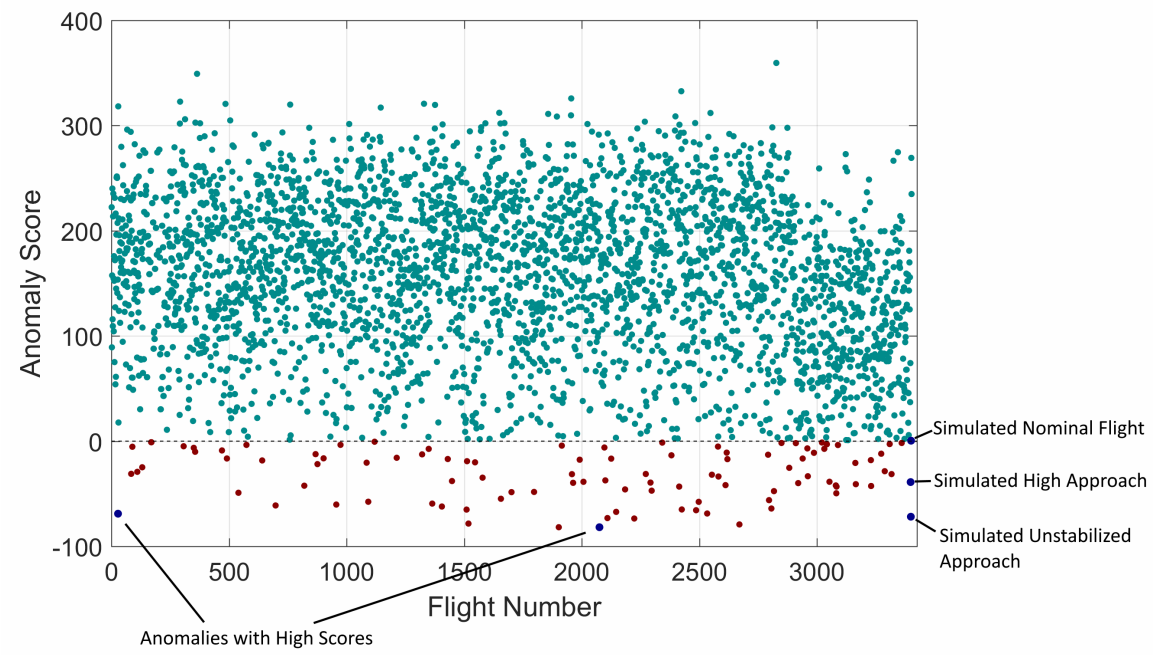


Figure 32: Anomaly scores in approach and landing phase for all the flights in the data set

As is evident from the negative scores, the algorithm is correctly able to identify the two known abnormal simulated flights (unstabilized and high) as anomalous. Similarly, it is also able to put the simulated nominal flight data record into the nominal category. This indicates that both abnormal and normal simulated flights

inserted into the data set are identified as such. In the absence of actual known anomalous flights, this enables verification of the fact that the algorithm is able to identify known anomalies. The scores obtained for each flight enable automatic identification of flight-level anomalies in the approach and landing phases of flight as negative scores indicate anomalous flights. The magnitudes of the scores indicated in Figure 32 are indicative of the severity of the anomaly as compared to all flights in the data set. Assuming that the entire data set is used in training the SVM, the approximate proportion of outliers expected in it can be specified a-priori. This provides a flexibility to the analyst to set outlier significance level and manage the number of anomalies identified. The anomalies identified using the SVM can be post-processed in a number of ways. Plotting the variation of various energy metrics and parameters during the phase of flight can help understand why the flight was considered an anomaly and compare how it performed relative to the entire data set. Two sample anomalies are post-processed in this section to provide further insight and corroboration of their identification as anomalies. For post-processing the same style of plots is presented for energy metrics, parameters, and (in later sections) instantaneous anomaly probability. In each figure, the dark grey bands represent 50^{th} percentile of all flight records and the light grey bands represent the 95^{th} percentile of all flight records. The solid black line represents the specific flight data under consideration. Overlaying it on the percentile bands enables comparative assessment of how a particular flight performed at each instant compared to every other flight in the data set. The distance remaining to the displaced runway threshold in each case is displayed on the x-axis. Since each flight lands at a different point on the runway, the approach phase is not considered beyond the runway threshold. However, the effect of different types of approaches (anomalous or not) on the touchdown performance are important and are explored later in chapter 7.

Case: Anomaly from flight data

One of the anomalous flights from the data set is visualized in Figures 33 and 34. The variation of some critical energy metrics and flight parameters during the approach-and-landing phase for the anomalous flight, along with the bands for nominal variation within the data set, can be seen in the figures. From the energy metrics, it can be seen that the specific potential energy around the 2.5 miles remaining mark starts deviating away from the nominal bands. The specific total energy rate, which should be negative or zero in approach and landing goes over zero for a patch (between 2.5 and 2 miles remaining), so the airplane is gaining energy instead of losing it. The modified total energy error rate is positive in this region indicating that the gain in energy is not required to correct the airplane's energy profile. At the end of this region, a precipitous drop in kinetic energy (along with a high kinetic energy rate magnitude) attempts to correct the flight profile. Due to the deviation of several such metrics from nominal, the flight data record gets identified as anomalous.

Similar to the energy metrics, Figure 34 shows the variation of some raw flight parameters captured by the recorder. The altitude and true airspeed profiles follow similar trends to their normalized versions in the energy metrics (potential and kinetic energy). The pitch and vertical speed profiles indicate deviations from normal variation in the region where the energy metrics also demonstrated deviations. Similarly, because the airplane is too high, the RPM drops (power is reduced) for the end part of the approach. The roll angle variation towards the end of the approach indicates high bank angles at the end which indicate some elements of an unstabilized approach. Refering to the definitions of exceedances in appendix B, the flight has several level 1 (≈ 27) and level 2 (≈ 19) exceedances.

Case: Simulated Anomalous Approach (High and Fast)

Due to the lack of actually unsafe operations expected from routine operations data

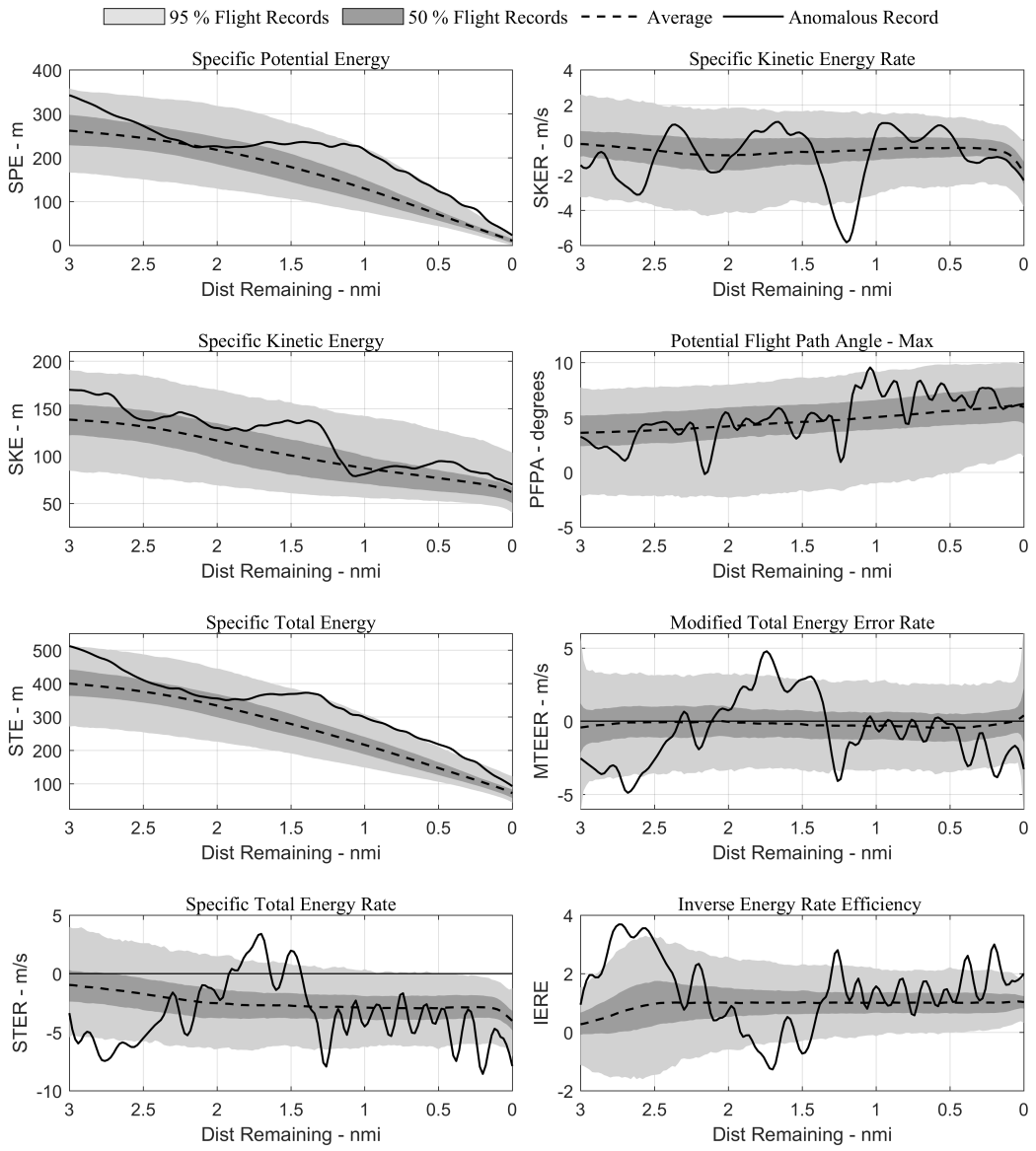


Figure 33: Variation of energy metrics during approach and landing for anomalous flight data record

such as the data used in this work, simulated flight data records with artificial flight-level anomalies were inserted into the data set. One such simulated anomalous approach and landing is visualized in Figures 35 and 36. This flight data was flown by a

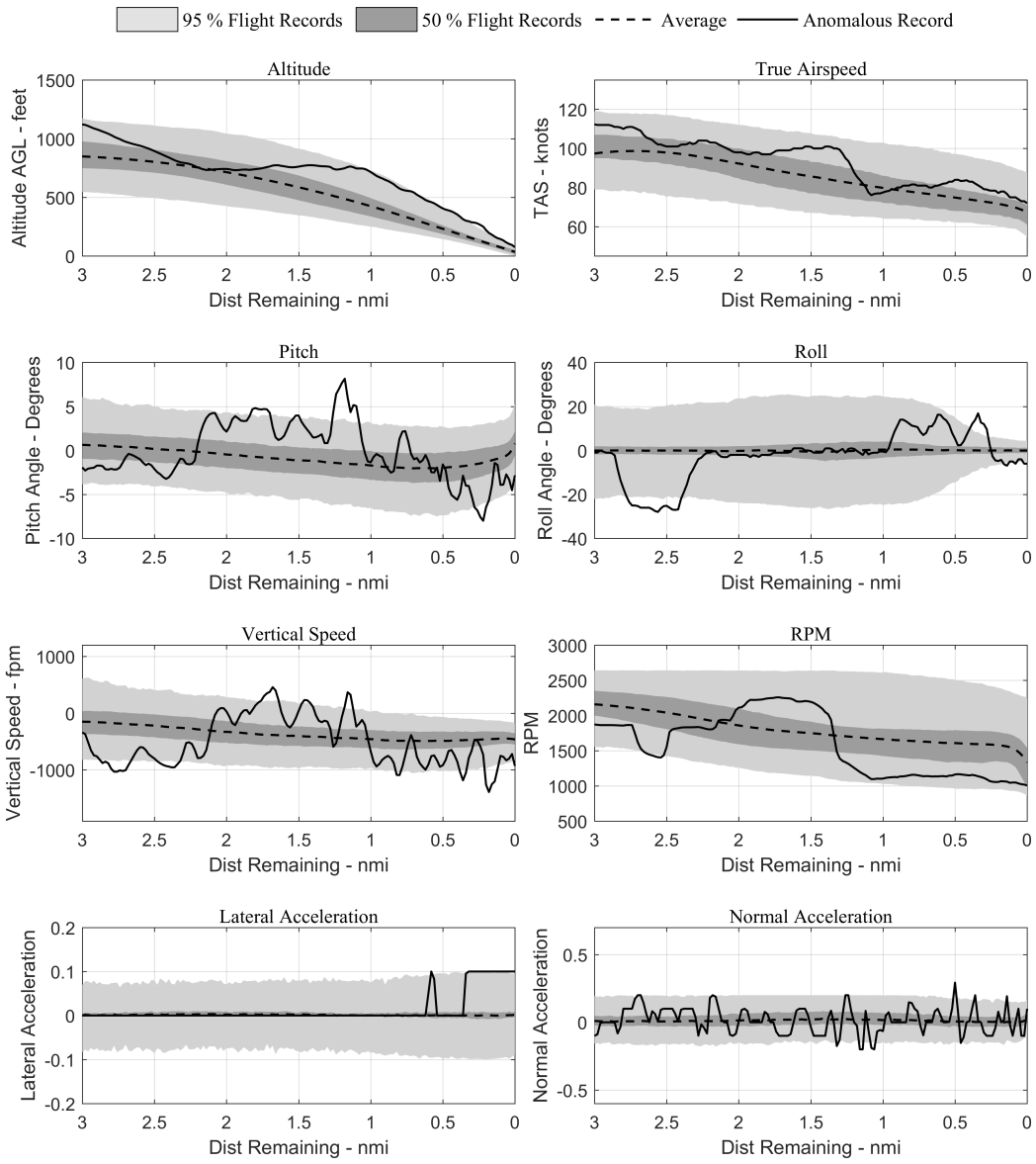


Figure 34: Variation of raw flight parameters during approach and landing for anomalous flight data record

private pilot instructed to fly a high and fast approach. As expected, the score of this flight data record is in the anomalous region. The variation of energy metrics seen in

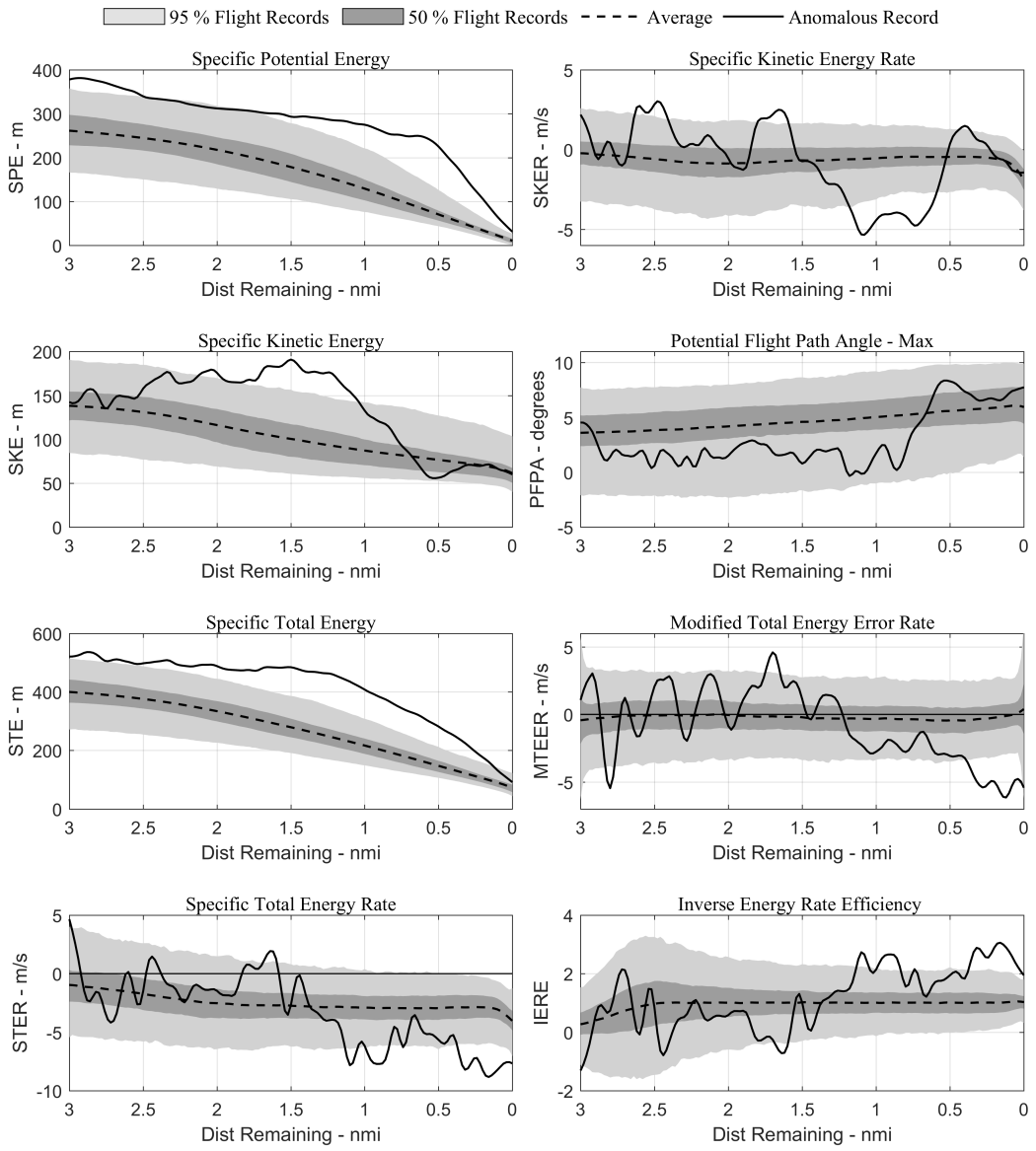


Figure 35: Variation of energy metrics during simulated anomalous approach and landing

Figure 35 indicates the large deviations from normal values for all three energy components (potential, kinetic, and total). Because of the persistent high values (kinetic and potential), abrupt corrections are made towards the end of the approach prior

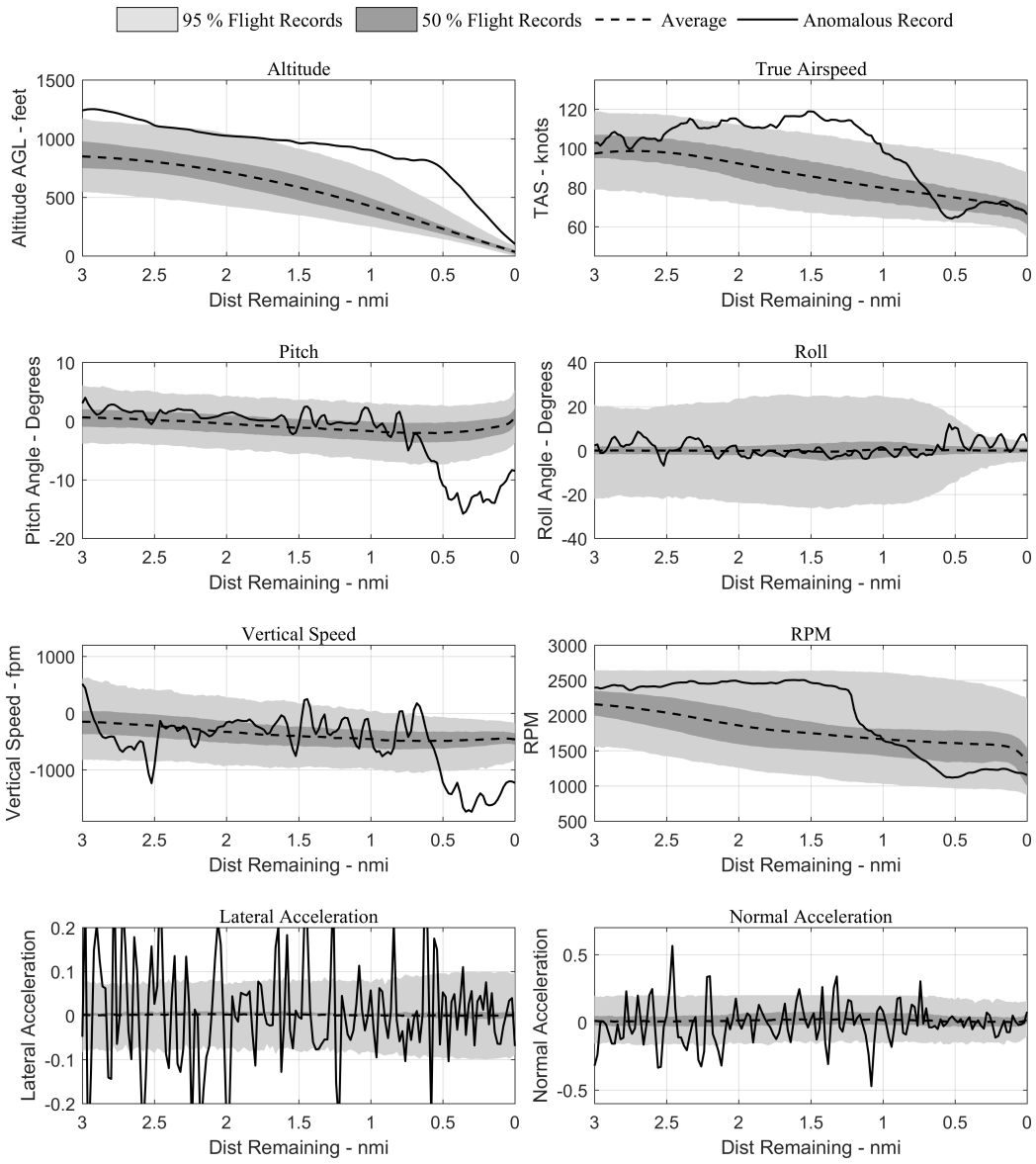


Figure 36: Variation of raw flight parameters during simulated approach and landing

to landing. Therefore large negative values for energy rates result in the end of the approach towards the landing. The pilot is able to correct the kinetic energy profile but still ends up crossing the runway threshold at a higher altitude than normal.

The variation of raw flight data parameters visualized in Figure 36 shows adherence to the normal regions for vertical speed, pitch, and roll for most part of the approach other than the end where the rapid corrections are made prior to touchdown. Similar to the previous flight, RPM drops towards the end (power reduced) to correct for the high altitude profile. This flight data record also contains a some level 1 (≈ 2) and high number of level 2 (≈ 26) exceedances. This case of simulated anomaly is able to demonstrate the ability of the methodology to capture known *unsafe* events successfully in the absence of actual unsafe data.

5.2.3.4 Take-off Phase

For the take-off phase of flight, a separate SVM model is trained and used to predict the anomaly scores for each flight. The scores for all flights in the take-off phase using SVM are shown in Figure 37.

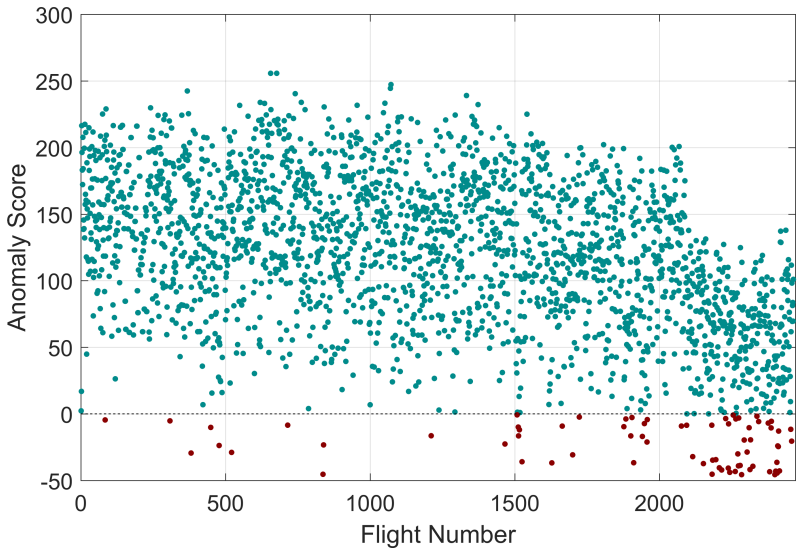


Figure 37: Anomaly scores in take-off phase for all the flights in the data set

The first difference evident for the take-off phase compared to the approach and landing is that the magnitude of the scores for anomalies is typically less severe. This is a trend that is expected because take-off typically has more uniformity in operations

compared to approach. For example, approaches can be straight-in or through a traffic pattern. Higher variation of power settings is expected for approach phase than take-off. Since there isn't as much variation in take-off, the anomalies identified for take-off are not expected to be as obvious or as significantly deviating from nominal as approach phase unless it was an actual anomaly or a training flight with some demonstration. The following two cases describe sample anomalies obtained during the take-off phase along with the visualization of their energy metrics and flight parameters similar to the approach phase. It is noted that the x-axis in this case is time instead of distance because the anchor point in the take-off phase is application of take-off power. Each take-off is compared for the first two minutes after application of take-off power as significant variation in flight profiles is observed beyond that point to draw any meaningful conclusions.

Case: Sample take-off anomaly 1

Figures 38 and 39 show visualization of energy metrics and raw flight parameters for a sample take-off anomaly. The variation of energy metrics indicates deviation in the specific potential energy metric as well as the specific potential energy rate. The take-off is executed at higher rate of climb compared to other flights from the data set. Thus, the kinetic energy in the initial stages of the climb out is lower than usual as a steep climb is executed. The specific potential energy rate is also higher towards the latter part of the climb causing the potential energy profile to deviate further away from the nominal variation.

Turning attention to the raw flight parameters, it is seen that the pitch and vertical speed are higher initially, and the true airspeed is lower than the best climb speed. Vertical speeds exceed nominal variations (sometimes even above 1000 fpm) throughout the climb. The variation of most other parameters, however, is within normal bounds. It is noted that a lightly-loaded airplane with a powerful engine can

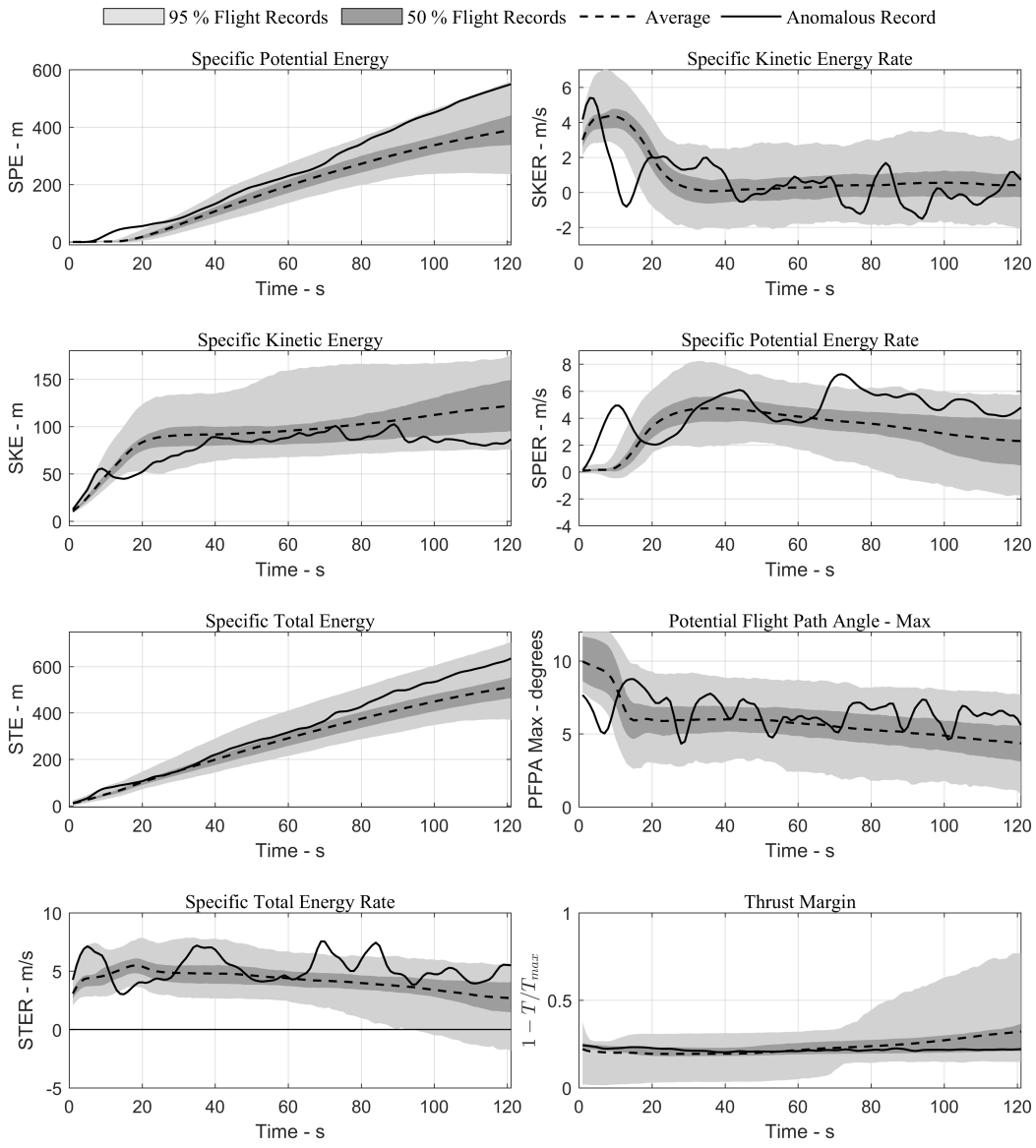


Figure 38: Variation of energy metrics during during take-off for sample anomalous flight data record 1

climb faster and/or sharper, especially on a cold day.

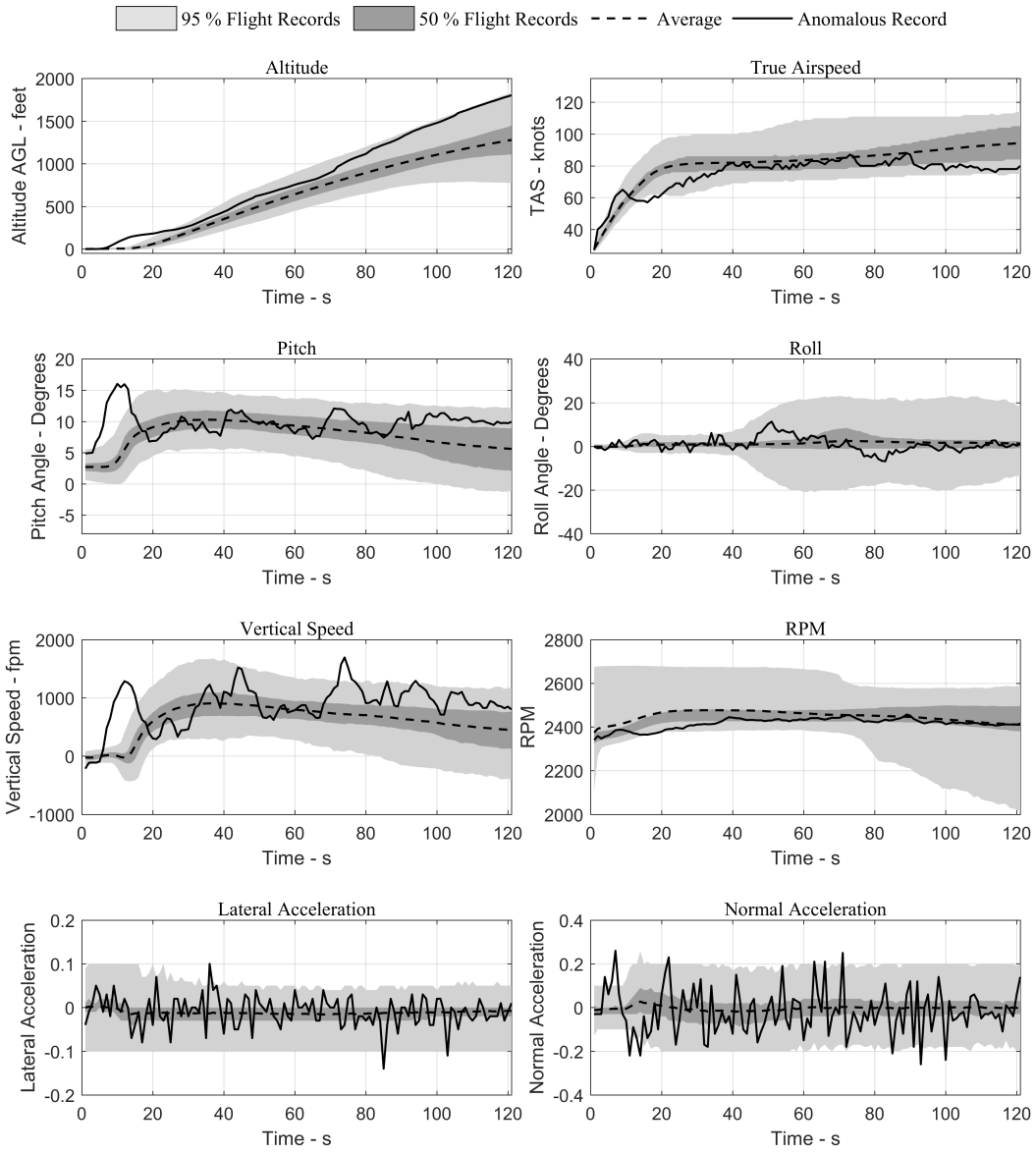


Figure 39: Variation of raw flight parameters during take-off for sample anomalous flight data record 1

Case: Sample take-off anomaly 2

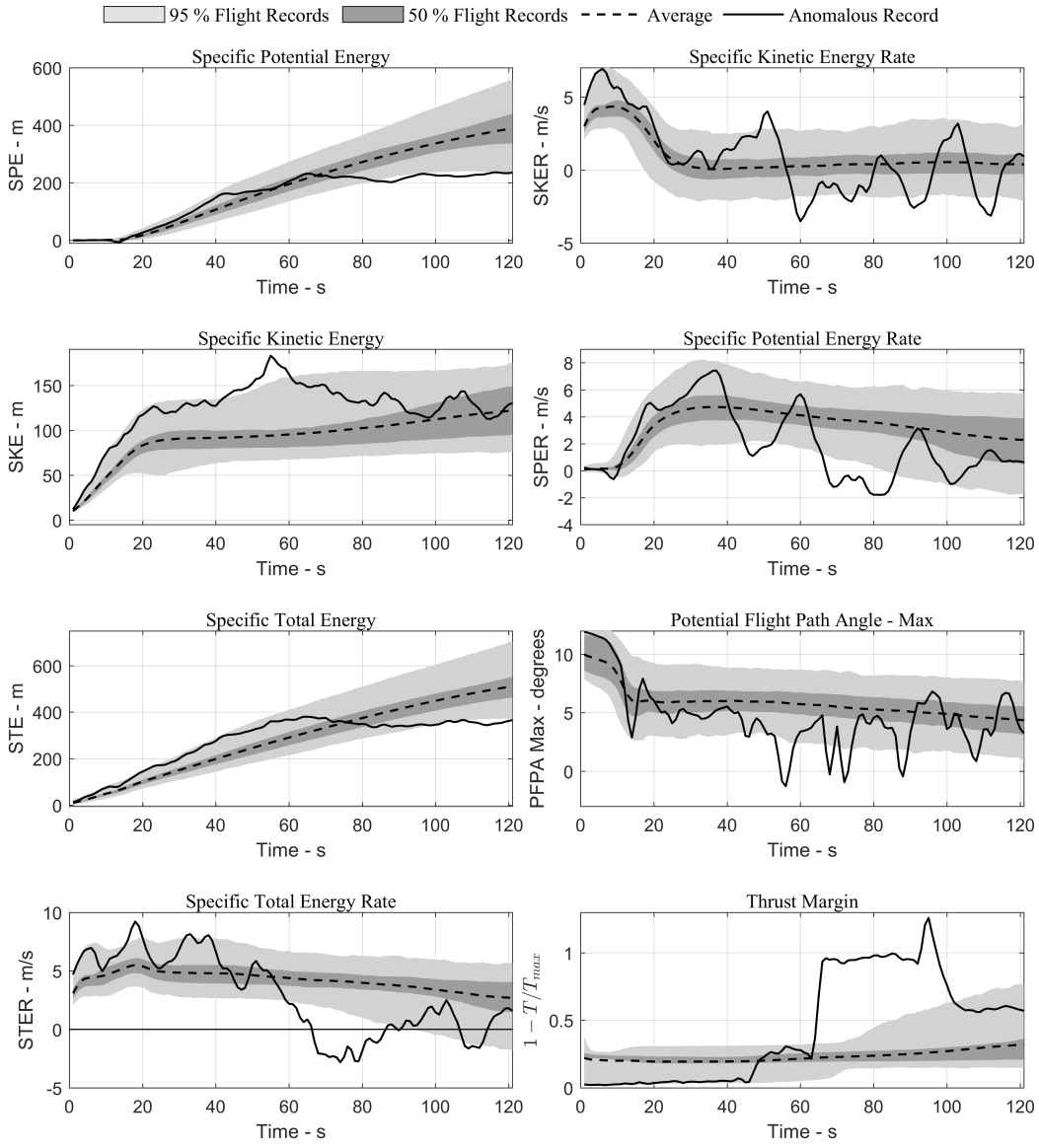


Figure 40: Variation of energy metrics during during take-off for sample anomalous flight data record 2

A different example of a take-off anomaly is visualized in Figures 40 and 41. The

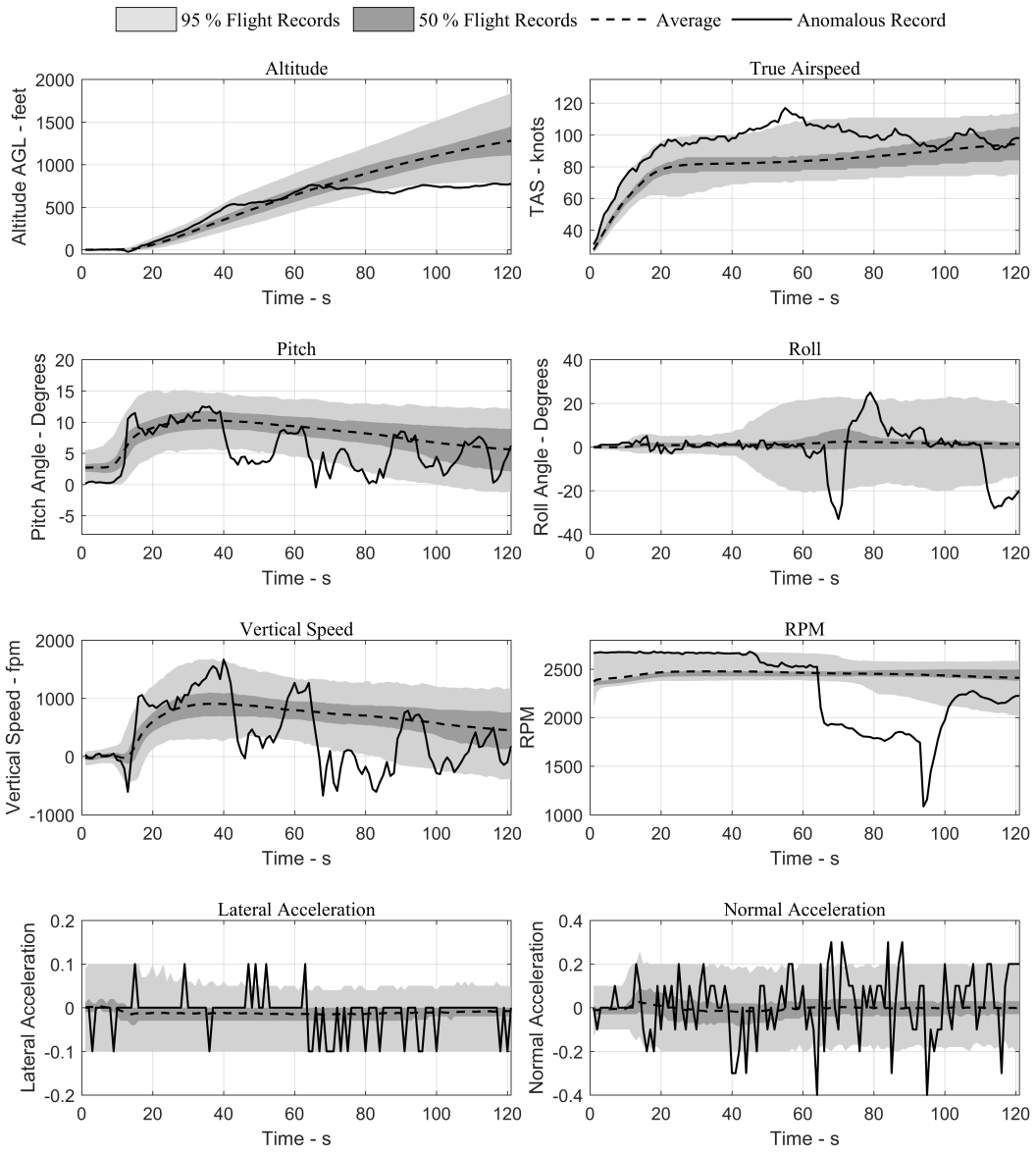


Figure 41: Variation of raw flight parameters during take-off for sample anomalous flight data record 2

variation of energy metrics for this take-off indicates deviations for several parameters. The specific kinetic energy is much higher than most flights in the data set almost throughout the take-off. The specific potential energy on the other hand is within

bounds for half the time and subsequently is much lower than majority of the flights. The specific total energy rate during the latter part of the take-off is lower (and even below zero). This indicates that the flight is losing energy instead of gaining it. A very high thrust margin for large part indicates that the aircraft is generating small proportion of the thrust it could generate in the latter part of the take-off.

The variation of raw flight parameters corroborates the findings from the energy metrics. Vertical speed, pitch, altitude, and true airspeed deviations are observed for the flight. The RPM falls to a very low value before returning to normal towards the end of the take-off indicating power being cut-off (possible for an engine-out simulation).

5.2.4 Summary

This experiment demonstrated the automatic identification of flight-level anomalies in take-off, approach and landing phases. Scores provided by SVM algorithm enable the severity of identified anomalies to be ranked. A natural boundary (at zero) between normal and anomalous flights is provided by the algorithm's score. The identification of known unsafe and anomalous flight data records inserted in the data set using a flight simulation model provided some validation in the absence of actual unsafe operational data. Similarly, a nominal approach inserted using the simulation model is also identified as normal. The demonstration of sample anomalies in the take-off and approach phases enables focusing safety assessment and training efforts on the important flights when a large amount of flight data is available. The reparametrization and alignment of flight records in both phases allows for comparison of different flights at the same (or similar) instants during their respective phases. This allows comparisons of data points that are inherently similar to each other and thus allows easier detection of deviations. Finally, visualization of energy metric and parameter

plots for anomalous flights enable the automatic identification, isolation, and understanding of anomalies and how they deviate from normal operations. The results from this experiment thus fulfill the purpose of the experiment and more broadly, that of research question 2.1.

5.3 Instantaneous Anomalies – Research Question 2.2

The second type of anomaly considered in this dissertation is instantaneous anomaly. The associated research sub-question is stated as follows:[2]

Research Question 2.2:
How can the general anomaly detection framework be modified to identify instantaneous anomalies in GA operations using energy metrics?

The main limitations of flight level anomaly detection techniques is that only well-defined phases of flight are considered. Also, the techniques aim to identify those flight data records that significantly deviate from nominal operations over a long period of time. However, it is not necessary that the whole data record be anomalous – only a few seconds or small part of the data record may be anomalous. Techniques for identifying instantaneous anomalies address some of these issues.

In many of the approaches in literature each point is monitored as a standalone independent sample and therefore, the temporal aspect of anomalous sub-sequences may be lost [77, 90]. This is also the case in FDM event analysis, where parameter values exceeding certain thresholds are flagged as exceedances without necessarily considering their context [4]. Moreover, in many anomaly detection applications for time series (including flight-level anomalies seen in Sec. 5.1) the multivariate time series is converted into a univariate series or a high dimensional vector thereby causing some information to be lost. In some cases of monitoring, the time-series is only compared to reference thresholds (such as in exceedance detection) or data

[2]The research described in this section is documented in the following publication:

– **Puranik, T.G.**, and Mavris, D.N., *Identifying Instantaneous Anomalies in General Aviation Operations*, in 17th AIAA Aviation Technology, Integration, and Operations Conference, 2017. Paper No: AIAA-2017-3779. doi:10.2514/6.2017-3779 [123]

within the same time-series losing out on potential insights that can be gained from similar data available. The existing anomaly detection techniques deal with data in which no relationship is assumed among the data points which is not necessarily true for flight parameters at each point. In some fields other than aviation safety, sliding-window based approaches are used along with SAX approximation to identify anomalous sub-sequences [80]. However the assumption in their approach is that the time series are stationary (i.e. the mean value does not change over time). These assumptions are not necessarily true for most parameters in flight data. Moreover, they are demonstrated only for univariate time series.

The methodology outlined in this subsection aims to address these limitations while continuing to remain within the same general framework outlined earlier in Figure 26. The adaptation of the general framework for instantaneous anomalies is elaborated here.

Step 1 (Flight Data):

The flight data obtained from DFDR is a multivariate time series, usually of a different length. The parameters are recorded at a specific frequency (e.g., once per one second interval). The characteristics and limitations of the flight data used in this step are similar to those for flight-level anomalies mentioned earlier in Sec. 5.1.

Step 2 (Pre-processing):

Sensor noise that may be present in the original data is smoothed using a simple moving average filter. Care is taken to not over-smooth the data as it would result in the loss of finer features in the data. In addition to basic noise removal similar to flight-level anomaly detection, the main pre-processing required for instantaneous anomaly detection is breakdown of the time-series into windows. One of the shortcomings of existing techniques identified earlier was that information about the temporal aspect

of instantaneous anomalies may be lost when data points are treated as stand-alone samples. It is of interest to capture the variation of key metrics before and after the sample under consideration. Therefore, to explicitly address this shortcoming, the present work proposes a window-based pre-processing for the data similar to that undertaken in other domains [35,80]. The multivariate time-series is broken into fixed length sub-sequences or windows. Features at each data point use information from the data within the sliding-window and are treated as a unit of analysis. The window length is the only parameter needed for pre-processing in this type of approach.

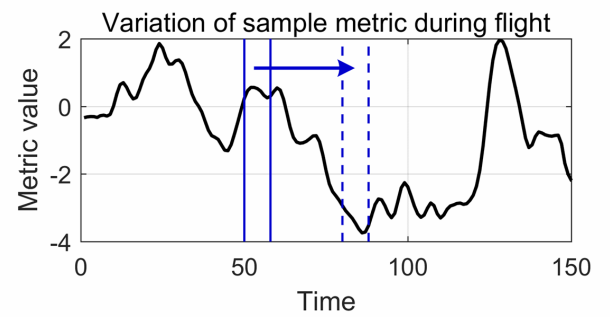


Figure 42: Notional depiction of sliding window across a metric

For each instant, the data contained within the entire window is utilized rather than just the instant under consideration. As noted in literature, the main advantage of window-based techniques is that they need only one intuitive parameter (the window length) as opposed to others that can require 3-7 parameters [80]. There are various factors that can affect the choice of length of the window such as total duration of the time series, typical response time of the system, computational concerns, etc. Due to all these factors, there is no consensus in the literature as to the appropriate length of the window [34]. In the present work, five seconds (two seconds before and after the current point, except at the end of the record) is used as the window length. An important distinction between existing window-based techniques in other domains and the application in this dissertation is that window-based anomaly detection typically tries to find anomalous windows in a time series based on the data

in that time series itself. On the other hand, the current method aims to also leverage the additional information available from the rest of the data set.

Step 3 (Feature Generation):

Similar to Sec. 5.1, features used for anomaly detection are generated from the energy metrics evaluated at each point in the window. The overall process of feature generation is shown in Figure 43. The feature generation step is perhaps the most important step in any anomaly detection application because the success or failure of the application depends upon the use of appropriate features [35]. Usually, these features are derived from the manipulation of the raw data collected. This can be as simple as using the raw data directly, to more complex steps involving calculation of new metrics using a combination of the recorded data and external information.

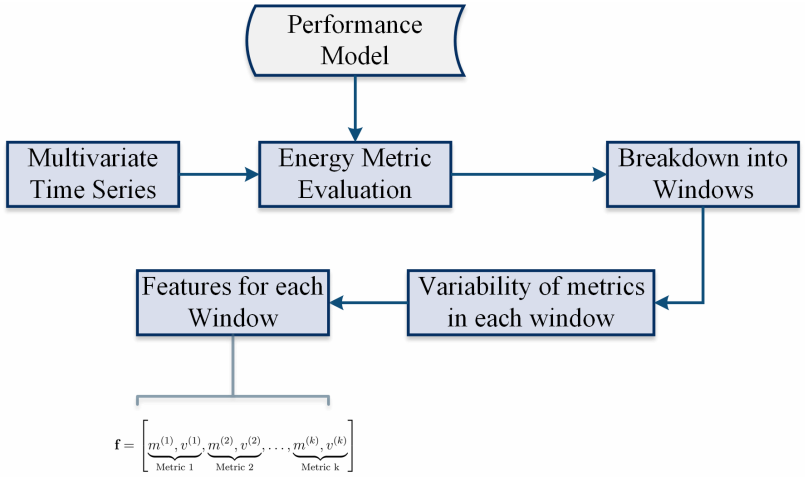


Figure 43: Process of feature vector generation for each window

Since this analysis is not performed for any specific phase of flight, the energy metrics which compare deviation from reference profiles for a particular phase of flight cannot be used in this approach. In order to obtain features of each window, the values of energy metrics at the current point along with their variability within the window is calculated. This allows the temporal aspect of instantaneous anomalies to be preserved while detecting anomalies. A similar window-based approach was used

by Amidan and Ferryman [11], however, their approach consists of fitting a linear regression model to each window and identifying anomalies based on the mathematical signature generated by the coefficients of the regression. In the current approach, the original value of the energy metric is retained as one of the features and additional dimensions corresponding to the variability of the metric within the window are generated. Specifically, the range of the metric values within the small window is used as a measure of its variability. Thus, for example, if the window size is 5 seconds, and there are k metrics being used, then the feature vector for that window will contain $2 \times k$ values (the value of each metric (m) and its variability (v)). Equation 28 shows the feature vector for each window which is then subsequently used in anomaly detection.

$$\mathbf{f} = \left[\underbrace{m^{(1)}, v^{(1)}}_{\text{Metric 1}}, \underbrace{m^{(2)}, v^{(2)}}_{\text{Metric 2}}, \ldots, \underbrace{m^{(k)}, v^{(k)}}_{\text{Metric k}} \right] \tag{28}$$

Each feature vector is thus generated by concatenating the value of all the energy metrics at that point along with its variability within the sliding window. Thus, each point in the time series is transformed into a corresponding feature vector which can then be used for anomaly detection.

Step 4 (Anomaly Detection):

The next step is to use the features generated for each point within an anomaly detection algorithm. It is noted previously that most anomaly detection techniques for instantaneous anomalies did not deal with multivariate data explicitly and converted it into univariate data prior to analysis. However, it is desirable to use algorithms that deal with multivariate data directly. Therefore, a Gaussian Mixture Model (GMM) is used for clustering and anomaly detection. The GMM can cluster normal operations together and help identify anomalies along with their probability based on the data

set. One of the main advantages of using a GMM is that it does operate directly on the multivariate series and does not transform it into a univariate series. The other advantage is that GMM allows multiple standard operations to simultaneously exist. This is very important for algorithms that are not specific to a particular phase of flight as the variation of parameters and energy metrics during two different phases of flight is expected to be quite different.

A GMM is a parametric probability density function represented as a weighted sum of Gaussian component densities. Each component in the mixture is a type of standard observation or behavior of the system (example, one component for each phase of flight). The number of components, k, in the GMM determines number of sub-populations or clusters. The relation between predictor variables (or features) is captured in the form of a covariance matrix Σ. If each member of the population (in this case each feature vector) is an m-dimensional vector, then the GMM with k components and a covariance matrix Σ is given by

$$p(\mathbf{x}|\lambda) = \sum_{i=1}^{k} w_i g(\mathbf{x}|\mu_i, \Sigma_i) \tag{29}$$

where $g(\mathbf{x}|\mu_i, \Sigma_i)$ indicates each of the components of the mixture model which is a multivariate gaussian model and w_i indicates the weighing of the component. The trained GMM is completely defined by the three parameters (w_i, μ_i, Σ_i) and the number of components k. The parameters of the GMM are typically obtained via an Expectation-Maximization (EM) algorithm using the data set available [46]. This technique is utilized in this dissertation. However, there are a few other important decisions that need to be made regarding the nature of the model before the EM algorithm can be used. These are (i) the nature of covariance matrix (full or diagonal), (ii) whether parameters among gaussian components are shared or not, and (iii) the number of components k. Due to the large computational cost of full covariance matrices, diagonal covariance matrix is used and the parameters are not shared among

different gaussian components in order to maximize the goodness of fit of the obtained models. Finally, the number of components can be set based on prior knowledge or it can be obtained using statistical metrics. Information theoretic metrics such as Akaike Information Criterion (AIC) [137] or Bayesian Information Criterion (BIC) have been used to identify the optimal number of components in the past [90]. These information theoretic metrics try to provide a balance between model complexity and overfitting. On the other hand, many internal clustering validation measures (such as Calinski-Harabasz index [29], silhouette criterion, Davies Bouldin index [42], etc.) rely on information within the data to provide a measure of the goodness of a clustering structure. They are typically based on two important criteria – compactness and separation. Objects within same clusters (or components) are closely related or similar to each other (compactness) and how distinct clusters (or components) are from each other [93]. In the present methodology, the Calinski-Harabasz (CH) index has been used to determine the optimal number of components. The number of components is progressively increased and the C-H index is measured. The number of components with the highest value of the index is chosen. While this particular metric is chosen in the current implementation, it is understood that alternative metrics may also be utilized for this purpose.

The advantage of using GMM for clustering is that it can provide statistical inferences about the underlying distributions. Therefore, once the required GMM has been trained using the existing data, it can be used to detect outliers or anomalies among the dataset. Using the values of the parameters obtained for the GMM (w_i, μ_i, Σ_i) the posterior probabilities of any component p for an observation $\mathbf{x}$ can then be calculated as:

$$p(\mathbf{x} \in p) = g(\mathbf{x}|\mu_p, \Sigma_p) \tag{30}$$

The estimated probability density function for each observation is then obtained as

a sum over all components of the component density at that observation times the component probability.

Step 5 (Post-processing):

In the post-processing step for instantaneous anomalies, using the estimated probability density function of each observation, a profile of the probability density over the entire duration of the flight as shown in Figure 44 can then be constructed. Using appropriate thresholds for the probability enables identification of anomalous sub-sequences or instantaneous anomalies. This threshold can be varied to obtain different number of instantaneous anomalies. The safety analyst can then decide this threshold based on the trade-off between workload and missed detection. Once instantaneous anomalies are identified, plots of variation of flight parameters and energy metrics can be used to visualize and understand the reason for identification of this anomaly. This flexibility enables the analyst to focus attention on a limited number of important anomalies as opposed to a large untenable data set. The identified anomalies can also be compared against traditional exceedance events such as those defined in Table 12 in the appendix.

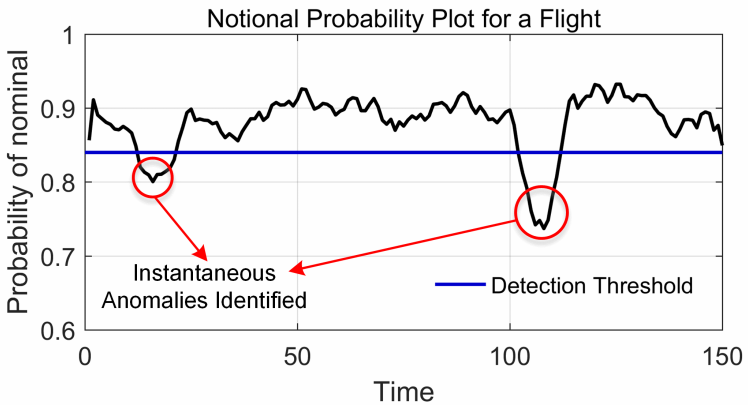


Figure 44: Notional depiction of probability density at each point during a flight record and the detection threshold

5.4 Experiment 2.2

The steps outlined in the previous sub-section enable building the methodology for identification of instantaneous anomalies in GA operations. In order to demonstrate the implementation of the methodology and address the research sub-question 2.2, the numerical experiment described in this section is performed. While instantaneous anomalies can be identified in all phases of flight, it is of interest to focus on take-off and approach-and-landing phases as these phases have maximum proportion of accidents. It is also possible to then compare instantaneous anomalies with flight-level anomalies and exceedance events in these phases to provide a comparative evaluation. Therefore, this section primarily focuses on these two phases with the understanding that the technique can be applied in the exact same manner for other phases.

5.4.1 Purpose of Experiment

The following are identified as the main objectives of this numerical experiment:

1. Demonstrate the usage of energy metrics in a retrospective safety assessment setting, in this case identification of instantaneous anomalies

2. Demonstrate automatic identification of instantaneous anomalies without a priori input

3. Provide a continuous probability of anomaly for any flight record as opposed to existing discrete metrics like exceedances

4. Demonstrate flexibility in identification of instantaneous anomalies with user-defined thresholds for anomalies

The artificial anomalies inserted in the previous section are also included in this section and the instances of unstabilized approach within these flights are examined for instantaneous anomalies. A detailed comparison between instantaneous anomalies and exceedance events is provided later in Chapter 7.

5.4.2 Parameters

One of the important parameters in instantaneous anomaly detection identified in the previous section is the length of the window used for pre-processing. The definition of how small an interval of time qualifies as instantaneous is not always clear in literature. For example, there is a chance that gusty/turbulent conditions can result in a lot of flags for instantaneous anomalies. An airplane caught in a thermal/downdraft/wind-shear, etc. may exhibit rapid changes in airspeed or vertical speed. Instantaneous anomalies should be able to identify these conditions while being insensitive to random (unavoidable) noise in sensor measurements. Introducing the variability of the metric in a window helps avoid some of these false flags due to noise in the data set. Additionally, since the Gaussian Mixture Model used in this dissertation is able to cluster and provide a probability score for each point which is a continuous function over the flight record, sudden gusts/noise will not be able to perturb this measure as easily.

5.4.3 Experiment Setup

The setup of this experiment is similar to the previous experiment. It consists of around three thousand flight data records obtained from two different flight schools collected during routine operations. The data is from two different types of aircraft (Cessna 172S and Piper Archer) operated at multiple GA airports. There are multiple aircraft of each type (different tail numbers) within the data-set. The data is pre-processed using the techniques developed in this methodology and additional derived parameters are obtained using techniques mentioned in Appendix C. The features used in this experiment consist of the complete set of energy metrics surveyed earlier.

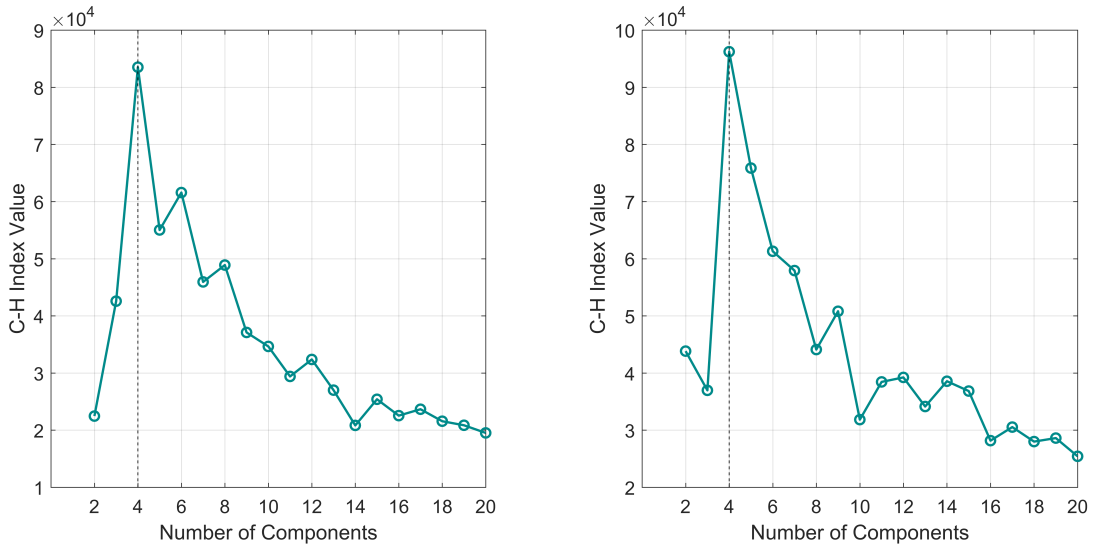
5.4.4 Results

The first step in instantaneous anomaly detection is identifying the approximate number of components that will be used in the Gaussian Mixture Model. As noted in the

previous section, this number is typically obtained using sensitivity analysis based on some information-theoretic indices and depends upon the data set that is available. Identifying the appropriate number of components using sensitivity analysis can be accomplished using different internal clustering criteria. The different internal clustering criteria use the distance between different points (usually Euclidean distance) to provide a measure of quality. This measure is based on various factors like total dispersion, within-cluster similarity, and between-cluster dissimilarity. With internal clustering criteria rugged topography might be more important in decision-making than the magnitude of the criterion itself. The choice of which criterion to use is dependent on the nature of the data set available and its properties. Definitions of various indices and their applicability is provided by Desgraupes [47]. In the current work, different indices are explored and the one with the clearest landscape of the plot is chosen for making the decision on the number of clusters. This criterion is the Callinski-Harabascz index (C-H index). C-H index is most suitable for cases where clusters are more or less spherical and close to normally distributed..

For this step, the GMM is trained with steadily increasing number of components and the C-H index is measured for each trained model. The model with components that gives the highest C-H index is the one displaying the best internal clustering structure and is chosen for this application. This sensitivity analysis is performed for both the take-off and approach-and-landing phases. The results obtained are shown in Figure 45.

Based on this sensitivity analysis, the number of components chosen for the GMM for both takeoff and approach-and-landing cases is four as this number has the highest value of the C-H index. The mean and variance of each gaussian component and the mixing probabilities for the trained GMM with four components are then used in further analysis. These parameters are then utilized to calculate the posterior probability density of being nominal at each point in the appropriate phase of flight



(a) Take-off (b) Approach and Landing

Figure 45: C-H index sensitivity with increasing number of components for different phases of flight

using equation 30. Once a posterior probability is obtained for each point, different thresholds can be set of detecting instantaneous anomalies. Instantaneous anomalies are identified as those points for which the probability falls below the selected threshold. The detection threshold for anomalies can be varied depending on the trade-off between missed detection and excessive analysis workload. Since the data is obtained from routine operations, there is no 'ground truth' available to compare anomalies, however, it is expected that the number of anomalies (if any) will be a small fraction of the total data. The next two sub-sections demonstrate the probability plots, parameter and energy metric plots for a few sample instantaneous anomalies identified in the take-off and approach and landing phases. It is noted that while this technique to identify anomalies is applicable for all phases of flight, these two phases of flight provide ease in visualizing the data of a particular flight compared to all others in the data set as the time-series/distance-series can be aligned in these phases (in addition to the fact that these are the phases with higher accident and incident rates). It also provides the opportunity to compare the two different types of anomalies with

each other (presented later in Chapter 7). The following sub-sections provide specific examples of the anomalies identified in approach and take-off.

5.4.4.1 Approach and Landing

Case: Flight with two instantaneous anomalous windows

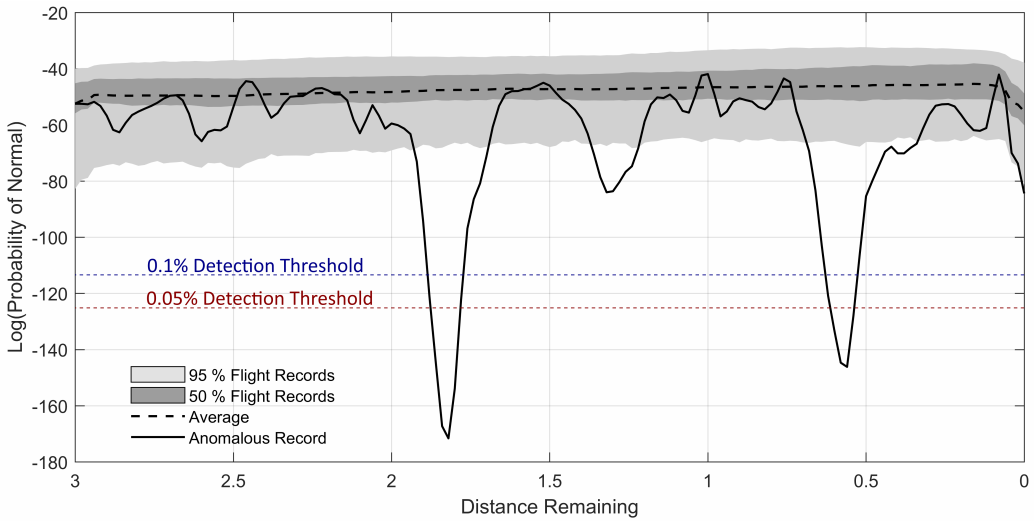


Figure 46: Probability density at each point during approach and landing for a flight record with instantaneous anomaly and the detection thresholds

In this section, an example of an instantaneous anomaly during approach and landing phase identified from the data set is presented. Figure 46 shows the variation of the probability density as a function of the distance remaining to the runway threshold for the flight record. The probability density is depicted as a natural logarithm of the actual density due to the low values typically observed with so many components in the GMM model and the amount of data in the set. Higher values of the probability indicate that the particular point was normal when compared to the entire data set. Lower values indicate abnormalities at that instant or instants. The blue and red colored lines in the figure indicate respectively the different thresholds for anomaly detection (0.05% and 0.1%). A point with a probability below the blue dotted line has a probability lower than 0.1% of all the points in the data set. Similarly, a point below the red dotted line has a probability lower than 0.05 % of all

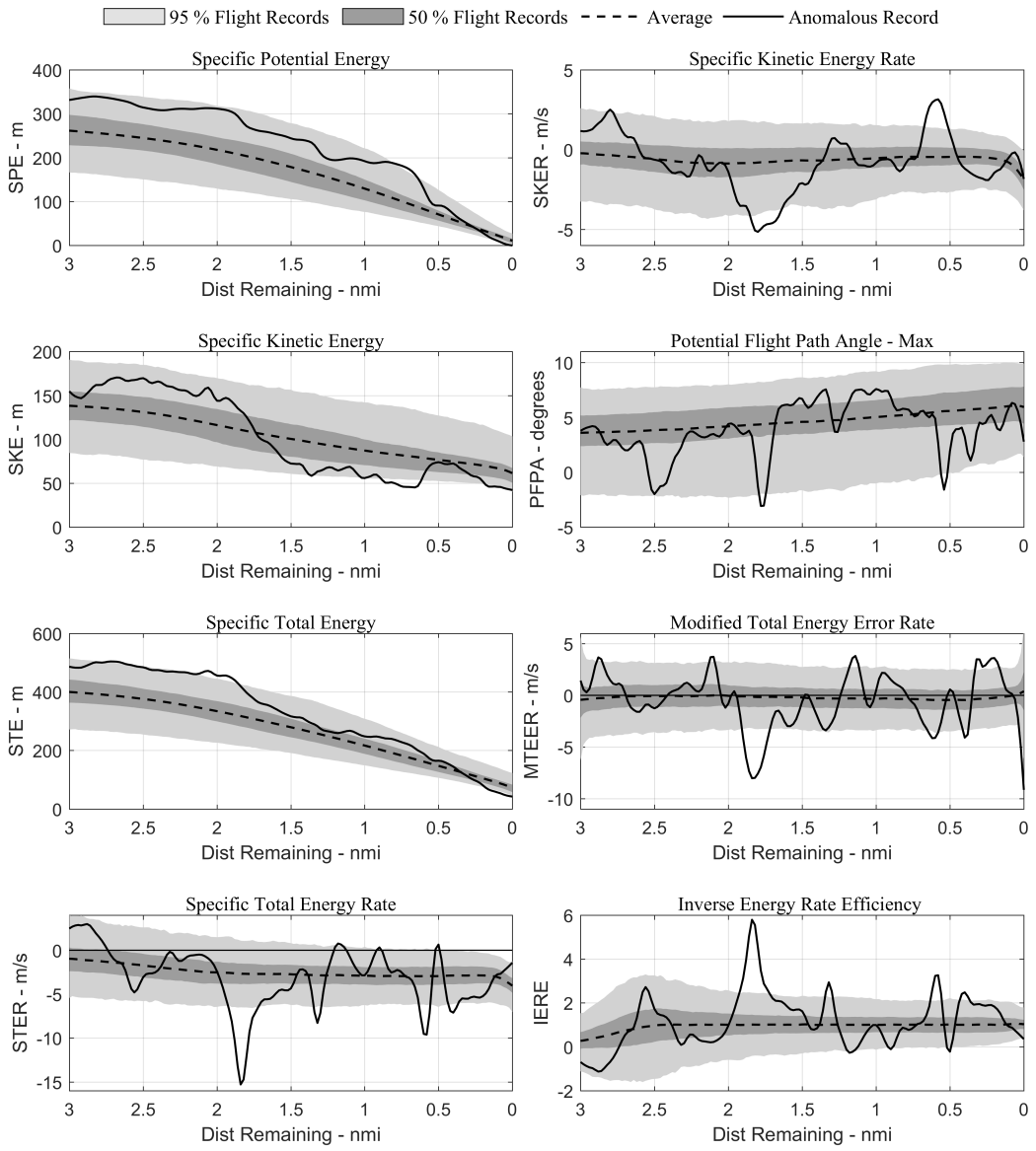


Figure 47: Variation of energy metrics during approach and landing for a flight record with instantaneous anomalies

the points in the data set. The presented instantaneous anomalies in Figure 46 are detected under both thresholds as it has a precipitous drop in probability density in the regions of the anomaly. This representation of the anomaly allows for focusing on the specific section of the flight record where the anomaly occurs and the severity

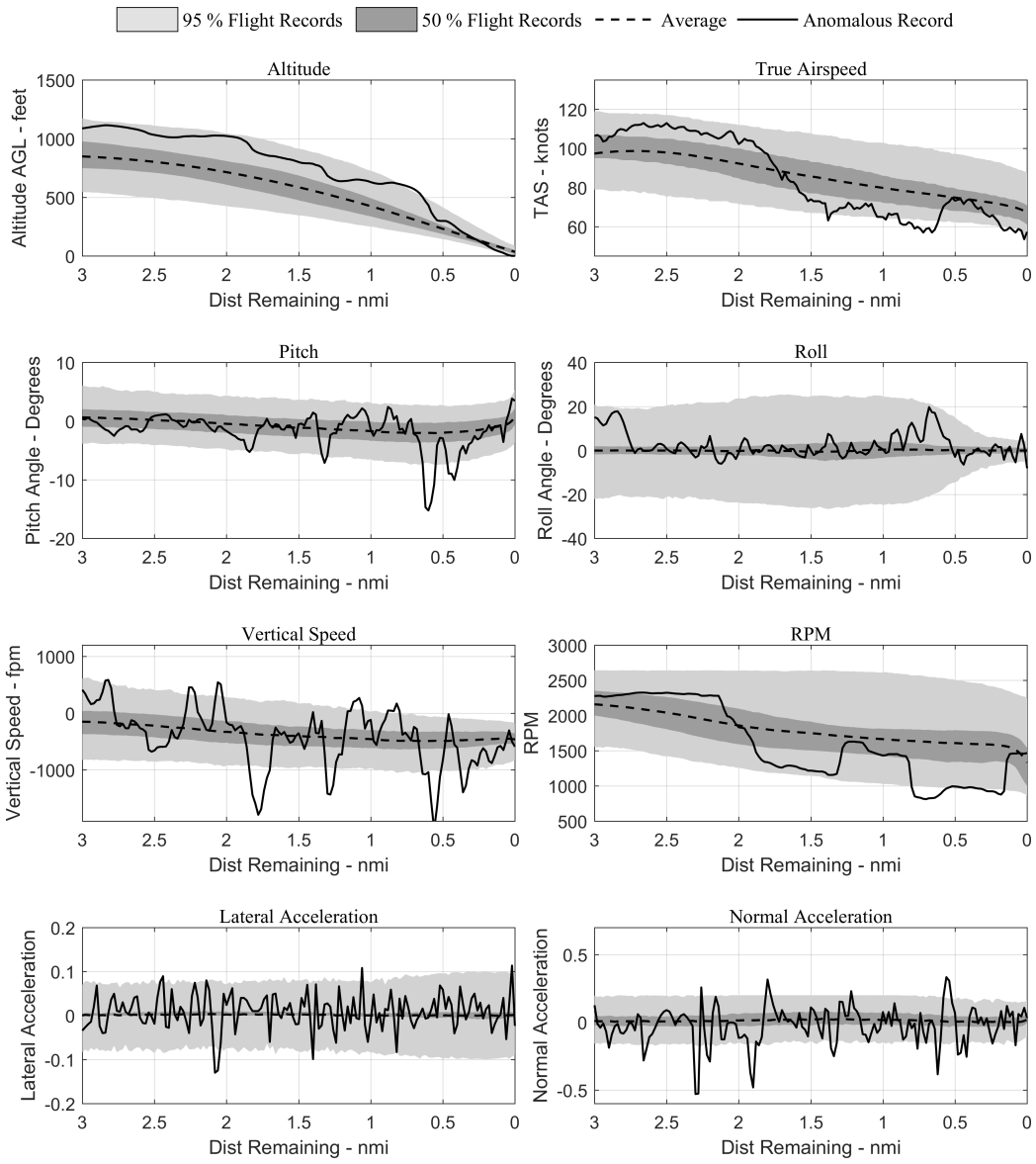


Figure 48: Variation of flight parameters during approach and landing for a flight record with instantaneous anomalies

based on the probability score. It can be further analyzed by visualizing the energy metrics and flight parameter plots in this window and comparing them to all other flight records as well as exceedance events.

Figure 47 shows the variation of the energy metrics for the flight record under

consideration. In the first anomalous window, there is a large negative specific total energy rate where the high and fast approach is sought to be corrected. Large changes are observed over multiple metrics. On the other hand, in the second anomalous window, the kinetic energy has fallen below the nominal variation and is being rapidly corrected at the expense of potential energy. The changes are not as sharp as the previous window, hence the probability value for this window is slightly better than the other one. In terms of flight parameters, the first window has unusual vertical speeds (in excess of 1,000 ft/min rate of descent) and a drop in the RPM and airspeed as well. On the other hand, the second window exhibits fluctuations in true airspeed and altitude, in addition to the roll angle reaching 20 degrees. Looking at the exceedances defined in the appendix and focusing on the regions of instantaneous anomalies, it is observed that the first anomalous window of seven points has one level-1 exceedance and four level-2 exceedances. On the other hand, the second anomalous window of five points has zero level-1 and five level-2 exceedances.

Case: Flight with one instantaneous anomalous window

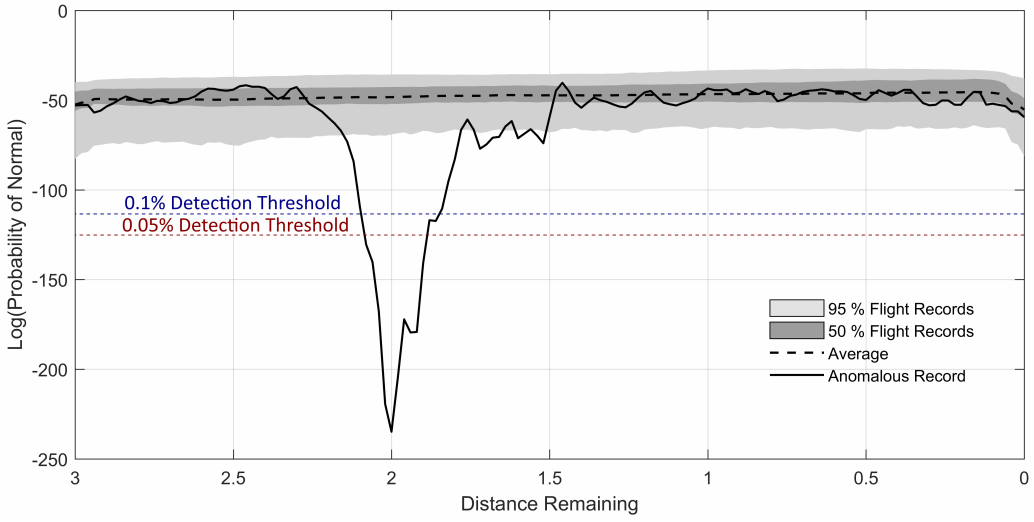


Figure 49: Probability density at each point during approach and landing for a flight record with instantaneous anomaly and the detection thresholds

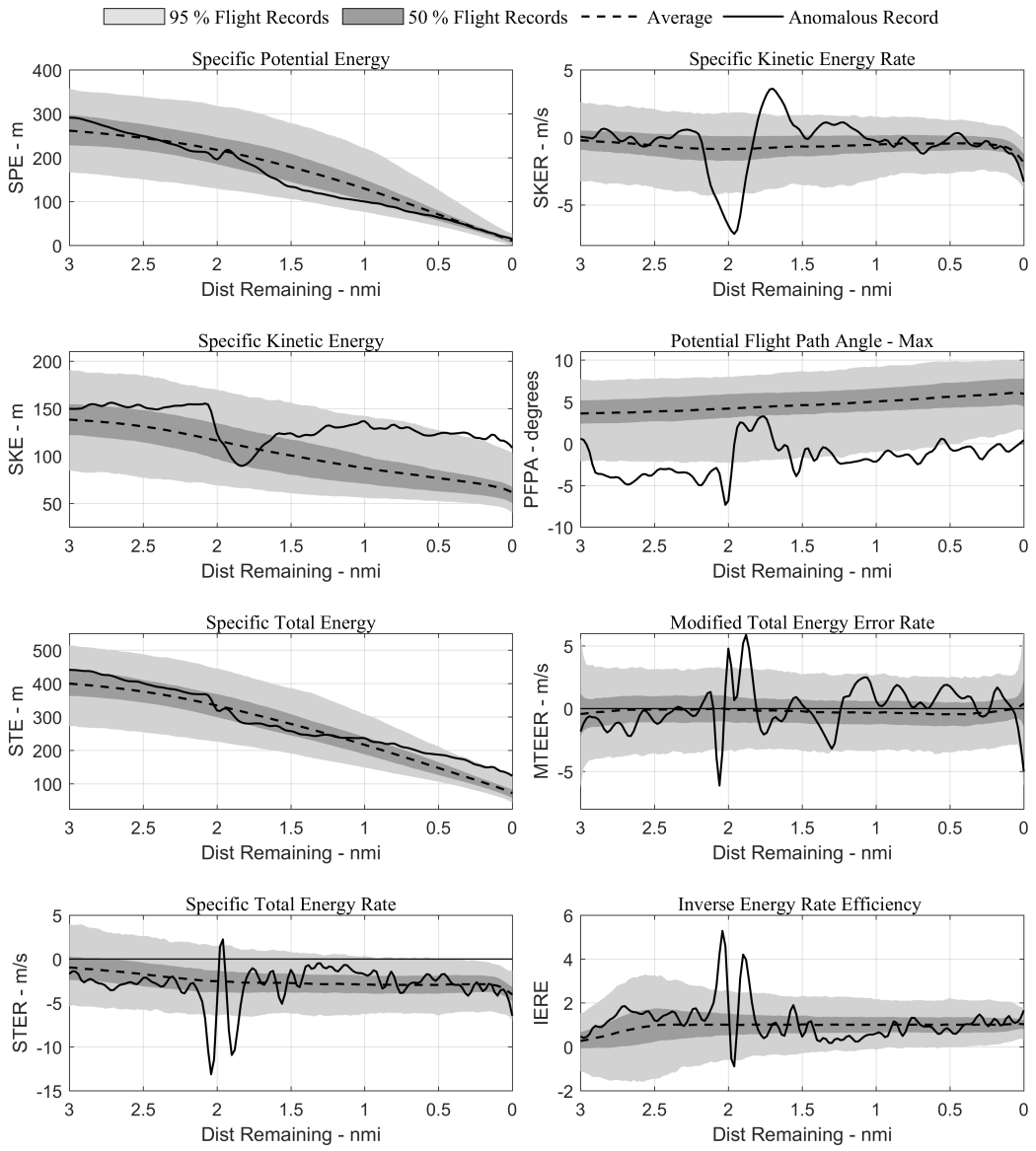


Figure 50: Variation of energy metrics during approach and landing for a flight record with instantaneous anomalies

In this section, a flight data record with a single instantaneous anomaly window is presented. Figure 49 shows the variation of the probability density as a function of the distance remaining to the runway threshold for the flight record. The variation of probability indicates a flight with nominal variation for most parts of the flight other

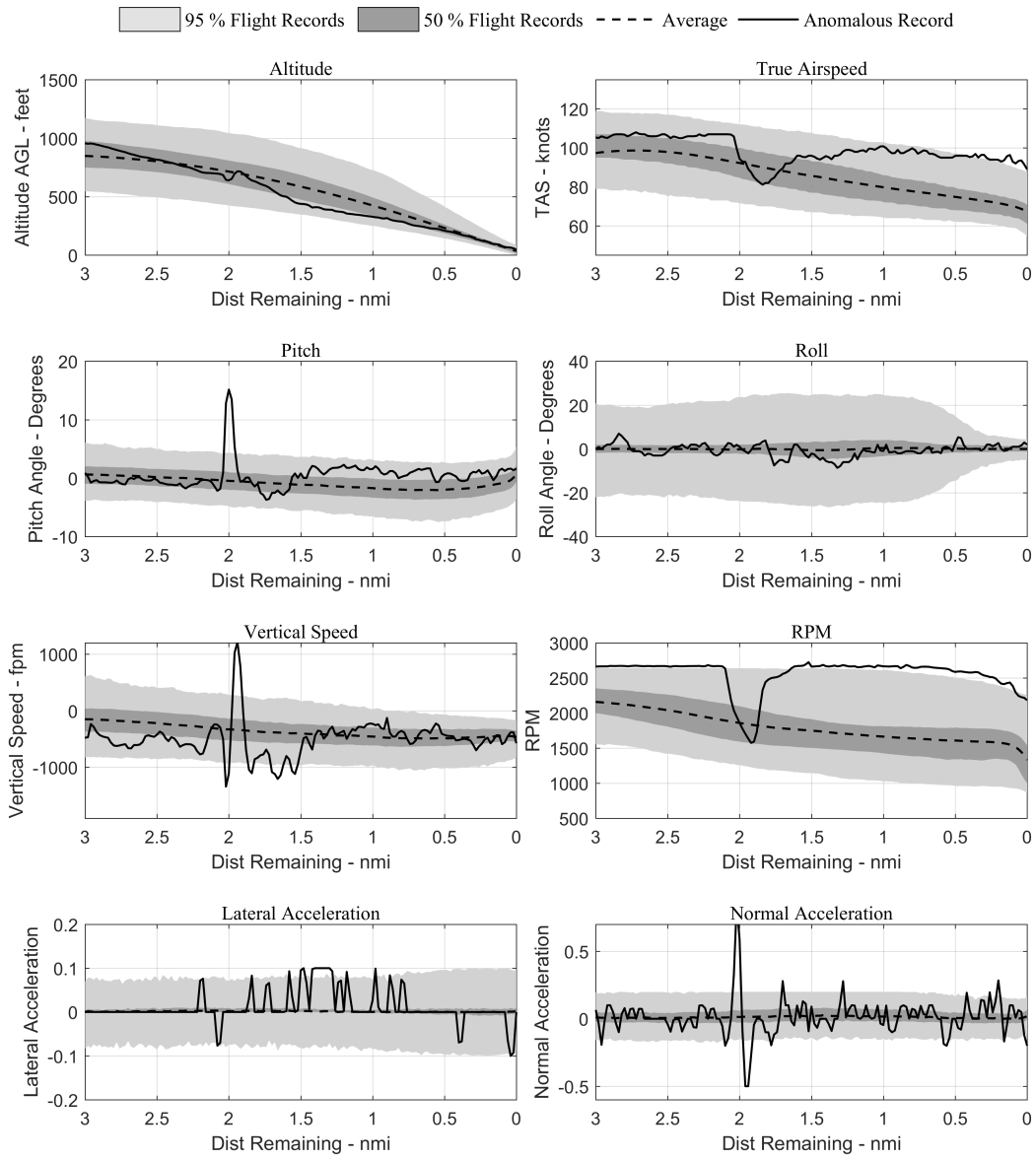


Figure 51: Variation of flight parameters during approach and landing for a flight record with instantaneous anomalies

than the window identified. The anomaly is identified under both detection thresholds mentioned in the previous section. Figure 50 represents the variation of the energy metrics for the flight with a single instantaneous anomalous window. The figure clearly indicates fluctuations in several metrics during the instantaneous anomaly,

notably, potential, kinetic energies, and kinetic energy rate. The profiles indicate a quick recovery to the nominal variations for these and other metrics after the small instantaneous window. Visualizing the raw flight parameters for this flight record it is evident that the RPM suddenly drops to well below its previous value and in order to maintain the altitude profile, the pitch angle is raised to a high value. Vertical speed also becomes greater than zero in this phase which is unusual for approach and landing. On a different note, the high kinetic energy and true airspeed all the way through to the end of approach indicate that the flight landed very fast and may have been a landing with no flaps. For this particular flight record, flap position is recorded in the data set and this hypothesis is confirmed as the flap position remains zero throughout the approach and landing. Upon examining the exceedances for this flight record, it is observed that of the fourteen points in the instantaneous anomaly, there are two level-1 exceedances and nine level-2 exceedances.

Artificial Anomaly

Prior to exploring instantaneous anomalies in the take-off phase, it is of interest to revisit the artificial flight-level anomaly inserted in the data-set earlier. The probability density for this flight record is shown in Figure 52. As is evident from the figure, while the flight record displayed a flight-level anomaly, it does not meet the threshold for instantaneous anomalies and therefore indicates that the corrections applied to correct the flight path were gradual enough to not get captured as instantaneous anomalies. The region where the corrections are applied (near the end of the approach) does display a drop in the probability density below nominal levels, but not precipitous enough to also be an instantaneous anomaly. Detailed analysis of this kind in the relationship between flight-level and instantaneous anomalies is provided subsequently in Chapter 7.

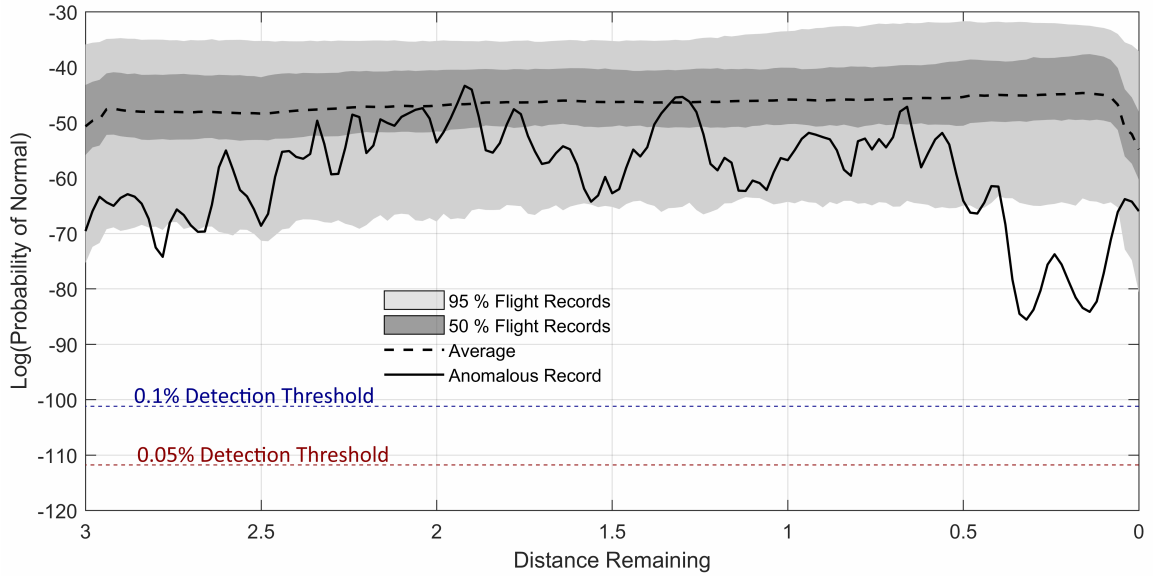


Figure 52: Probability density at each point during approach and landing for a simulated flight record with flight-level anomaly and the detection thresholds

5.4.4.2 Take-off instantaneous anomaly

Instantaneous anomalies identified in the take-off phase can be analyzed and visualized using probability plots and metrics variations. Similar to flight-level anomalies in take-off, the probabilities of anomaly are lower for instantaneous anomalies in take-off compared to approach and landing due to higher homogeneity.

Similarly, as is evident later in Figures 53 and 56 the initial parts of the take-off phase have a higher overall probability than the remaining sections. One of the reasons for this is that this part of the take-off has much higher homogeneity than even the rest of take-off. This is attributed to the fact that all take-off phases in the initial part are characterized by high power, increasing speeds (and kinetic energy), altitude gain (positive potential energy rate), etc. These similar conditions cause the cluster to be tightly knit and subsequent high probabilities.

Case: Flight with one instantaneous anomalous window

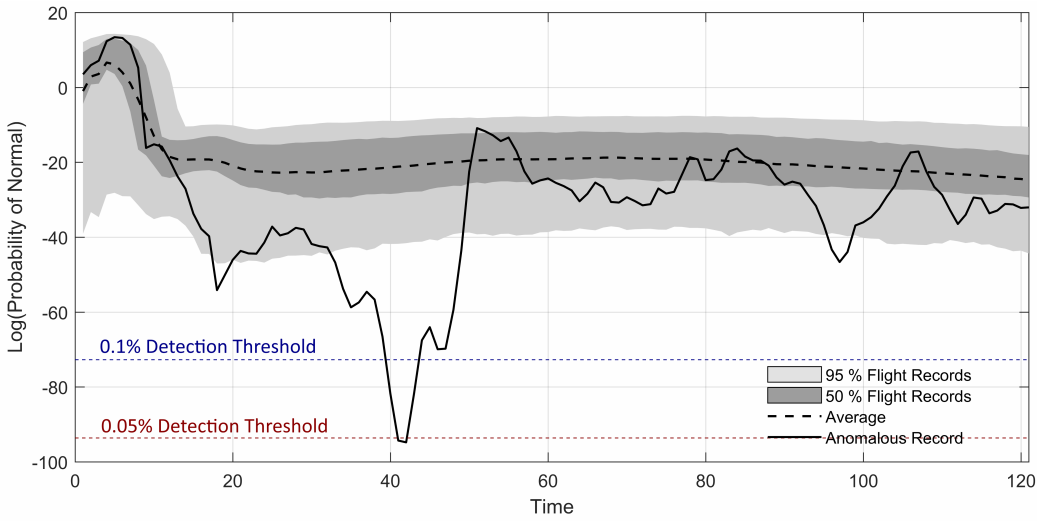


Figure 53: Probability density at each point during take-off for a flight record with instantaneous anomaly and the detection thresholds

Figure 53 shows the variation of probability density at each point during the take-off phase for the flight record under consideration. It is observed that the instantaneous anomaly occurs in the initial part of the take-off phase where there is a large drop in probability density. Exploring the variation of energy metrics during this flight in Figure 54 it is observed that during the anomalous window, the potential and kinetic energy rates display fluctuations out of nominal bounds. The potential energy rises faster than nominal in the initial part which is being sought to be corrected during the window. The variation of raw flight parameters shown in Figure 55 paints a similar picture. The altitude is leveled off by rapid reduction in pitch to gain some airspeed but it results in a negative vertical speed that causes the window to be detected as anomalous.

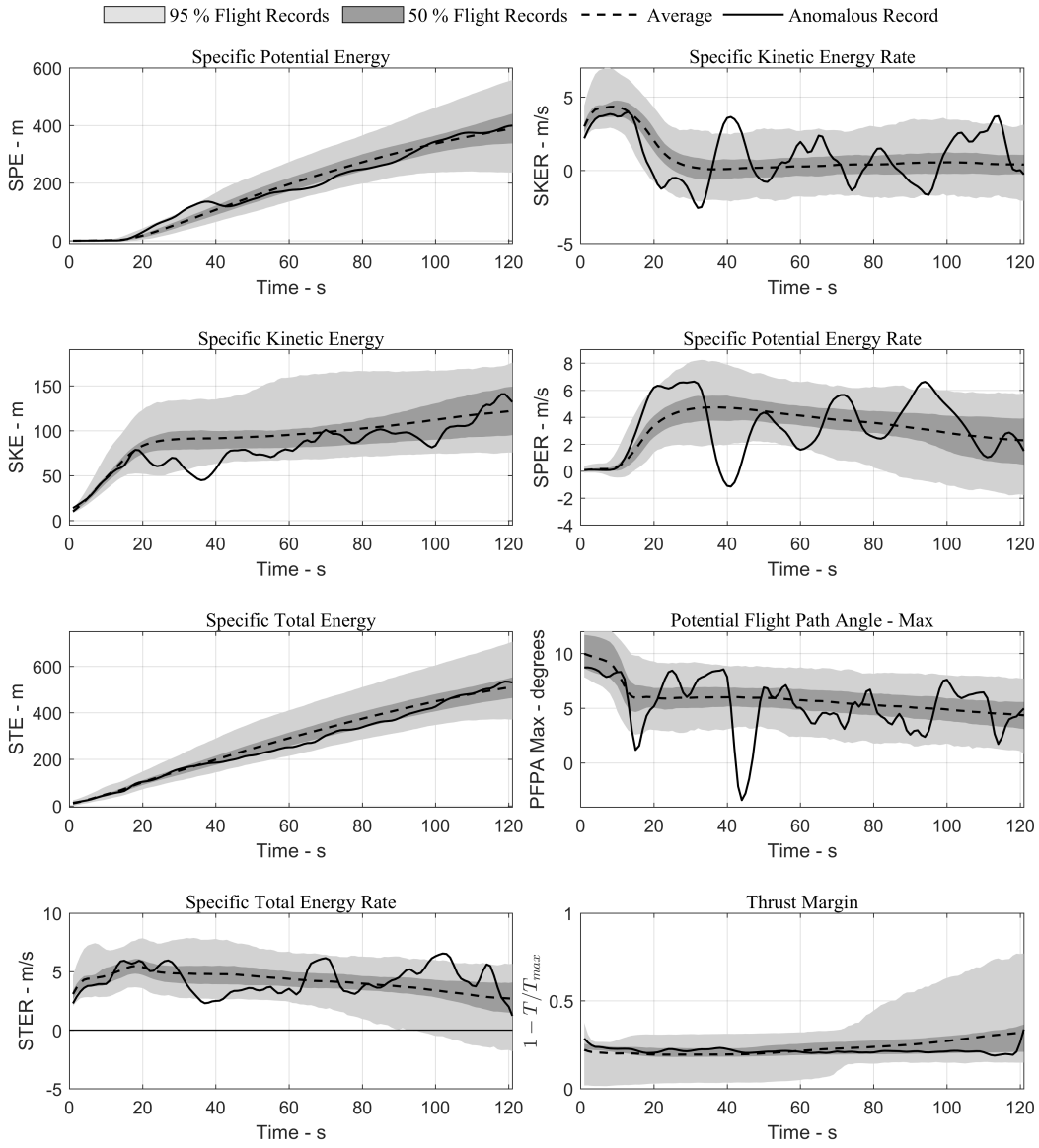


Figure 54: Variation of energy metrics during take-off for a flight record with instantaneous anomaly

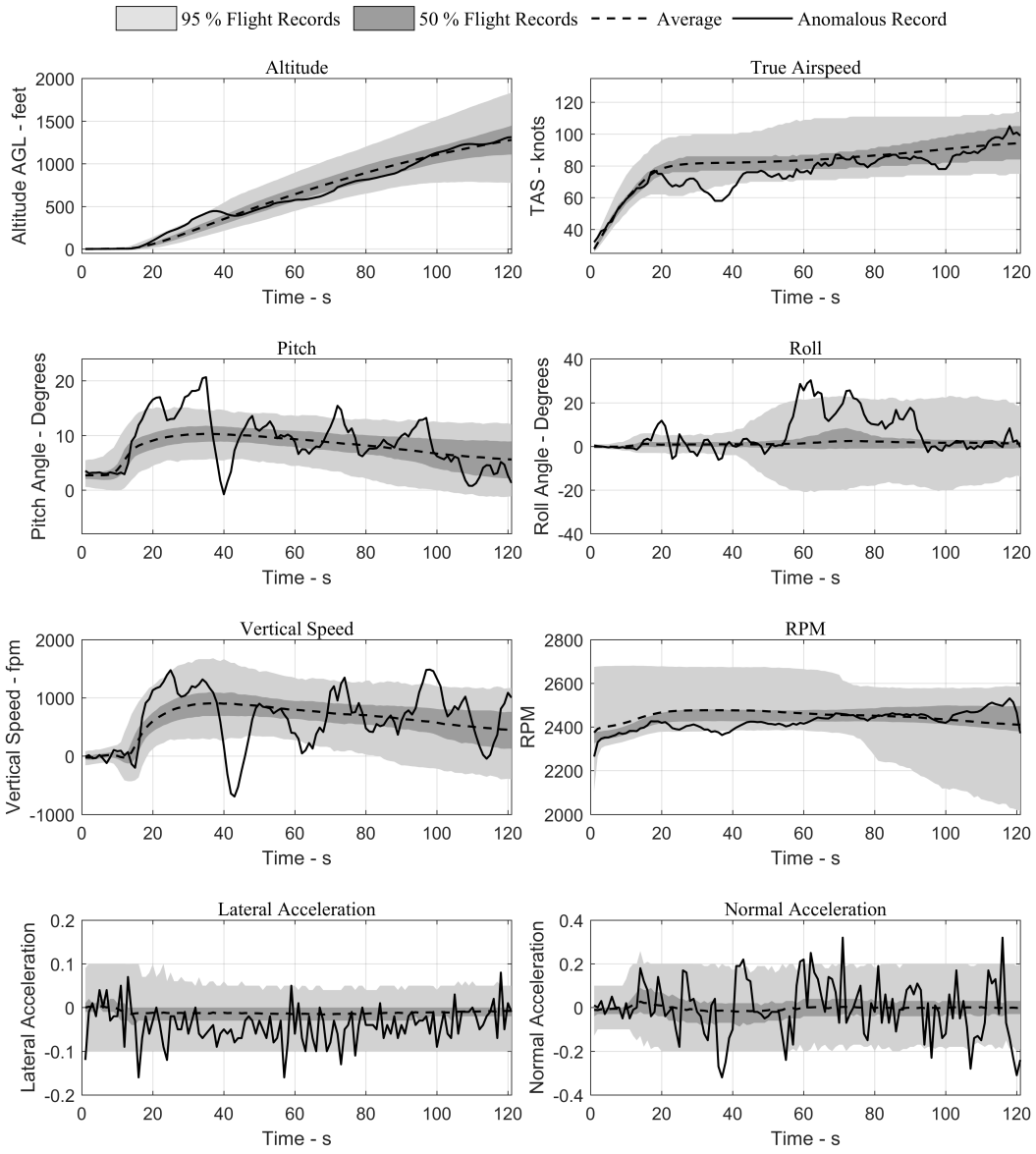


Figure 55: Variation of flight parameters during take-off for a flight record with instantaneous anomaly

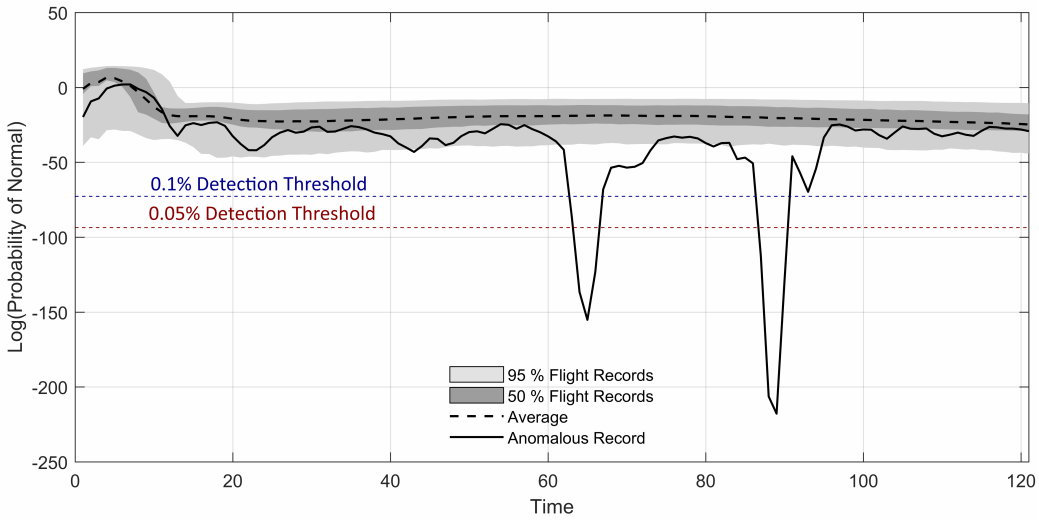


Figure 56: Probability density at each point during take-off for a flight record with instantaneous anomalies and the detection thresholds

Case: Flight with two instantaneous anomalous windows

The second example of instantaneous anomaly in take-off phase is demonstrated in Figure 58. As observed from the figure, this flight contains two anomalous windows in the take-off phase with low probability scores.

The visualization of energy metrics variation is shown in Figure 57. In the first anomalous window, there is a large drop in the kinetic energy and a highly negative kinetic energy rate. The total energy rate is also negative in this part indicating that the airplane is losing energy rather than gaining it. The very high thrust margin jump in this window indicates a decrease in the thrust being produced possibly due to power being cut off. The visualization of raw parameters shown in Figure 58. As expected, there is a precipitous drop in RPM at the beginning of the anomalous window accompanied by a drop in true airspeed. It is interesting to note that other parameters including pitch and vertical speed remain high while the speed is decreasing indicating that the airplane is creeping closer to its stall margin due to the high pitch and reducing airspeed.

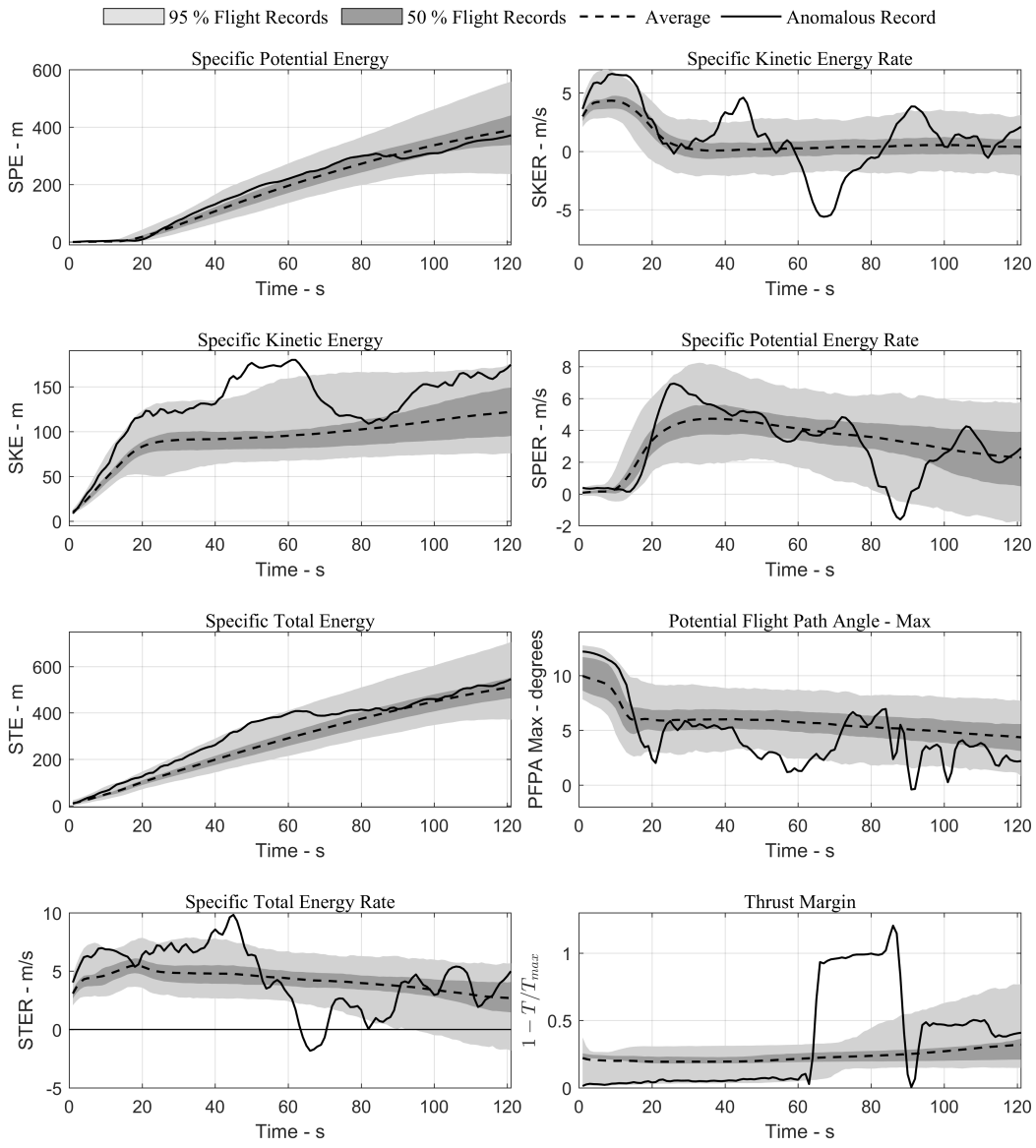


Figure 57: Variation of energy metrics during take-off for a flight record with instantaneous anomalies

In the second anomalous window, the airplane recovers to the nominal variation of energy profiles by trading the potential energy for kinetic energy and reduction of thrust margin (meaning producing closer to maximum thrust). The variation of parameters also provides information about negative vertical speed in this phase in

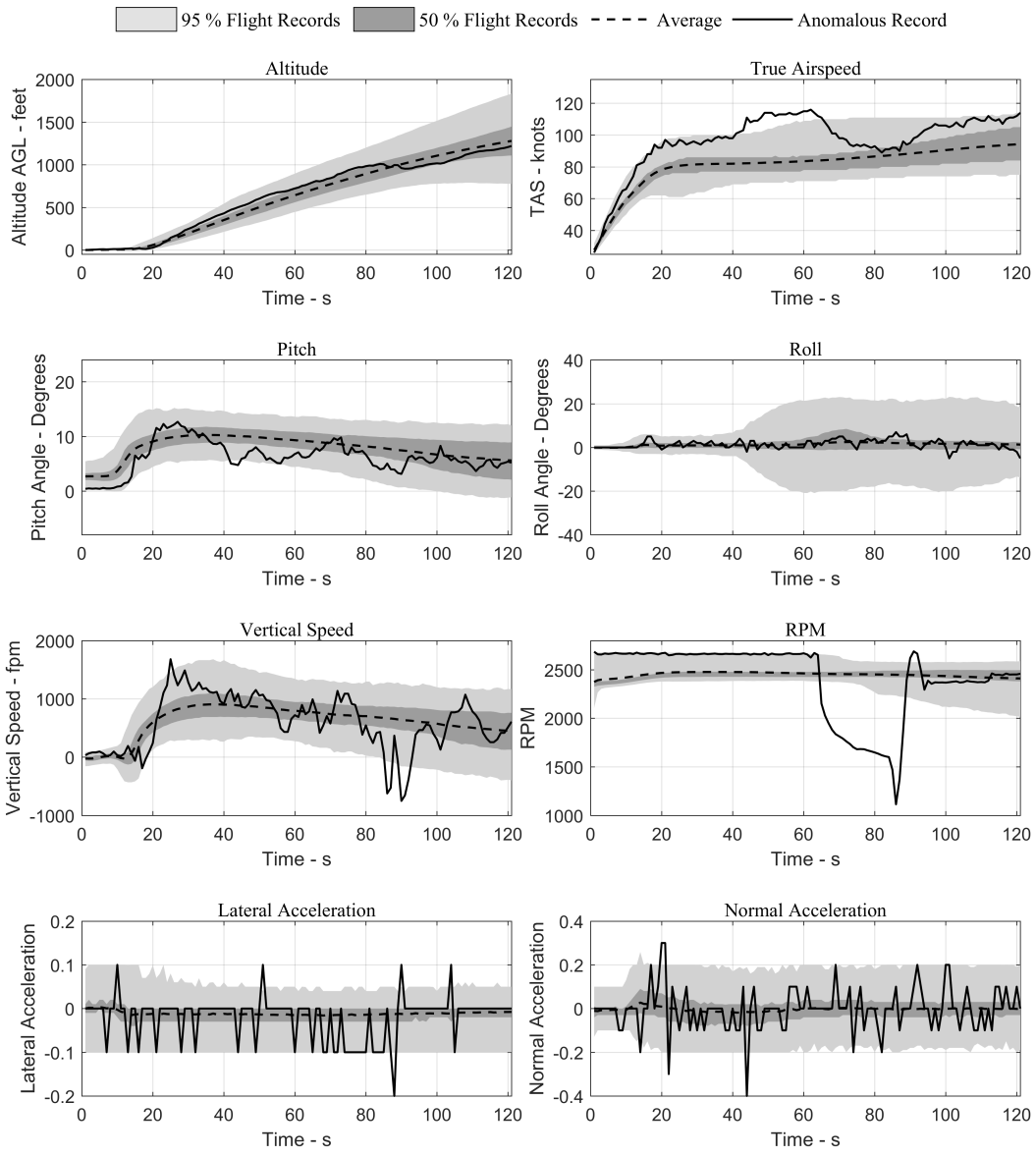


Figure 58: Variation of flight parameters during take-off for a flight record with instantaneous anomalies

order to recover the airspeed. Being the take-off phase, negative vertical speeds fall outside the nominal variations. If the aircraft re-entered the pattern for another landing, negative vertical speeds would be expected, but the variation of the roll angle suggests against this and therefore, this could be a safety issue or a simulated

safety event.

It is noted here in the second example that the second instantaneous anomalous window resulted because the airplane is attempting to recover from the effects seen in the first anomalous window. However, the algorithm does not have the ability to differentiate between the two. The variation of metrics in this window is abnormal for this stage of take-off and therefore it gets identified as an anomaly.

5.4.5 Summary

The experiment presented in this section outlines the application of a novel method for identifying instantaneous anomalies in GA flight data. This method utilized energy-based metrics as features in an anomaly detection framework. A sliding-window based pre-processing technique is formulated to ensure that the temporal aspect of features is captured. A mixture of gaussian models is trained using the available flight data for cluster analysis and outlier detection. Multivariate series are explicitly treated in GMM and it also allows for multiple standard operations which explicitly addresses some of the limitations identified previously.

This experiment demonstrated the application and usefulness of energy metrics for automatically identifying instantaneous anomalies in flight data. The main advantage of using these methods is that the expert review process is cut down due to the specific windows identified. The probability density function provides a convenient metric to distill and focus on points or specific regions of interest within the flight. It also allows dynamic thresholds to be placed by subject-matter-experts which can enable striking a balance between missed detection (too stringent thresholds) and excessively high anomalies (too loose thresholds). The results from this experiment and the demonstration of flights with instantaneous anomalies in both take-off and approach and landing thus fulfill the purpose of the experiment and more broadly that of research question 2.2.

5.5 Experiment 2.3 (Generalization)

In order to further test the capabilities of the methodology, two types of generalization scenarios are explored. These scenarios will enable testing of the flight-level anomaly detection methodology for variability in flight data recorder capabilities and the ability to train models which can predict anomalies in data not used in training. The results demonstrated in this section are for the approach and landing phase as very similar trends are observed for the take-off phase.

5.5.1 FDR Capabilities

Earlier sections have outlined the development of the flight-level anomaly detection methodology as well as the experiment to demonstrate their implementation. In this sub-section, an important question regarding limitation of the methodology due to flight data recorder capabilities is addressed.

5.5.1.1 Purpose of Experiment

Which anomalous flight data records can be identified using the developed methods in the presence of decreasing flight data recorder parameter availability?

5.5.1.2 Experiment Setup

In this scenario, effect of variability in flight data recorder capabilities on the flight-level anomalies detected is measured. GA FDRs can have varying data recording capabilities and it is desired that data obtained from multiple types of recorders be utilized in the same unified environment. In order to facilitate this, the defined energy metrics are divided into three types of feature vectors based on increasingly fewer data recorder requirements. This is achieved by omitting additional parameters in the feature vector options to indicate less-sophisticated recorder options. The exact parameters that are included can be seen in Table 15 in the appendix. The flight level anomaly detection algorithm is then run using each feature vector (in addition

to not using energy metrics at all - Feature Vector 1 in Table 15) to compare resulting anomalous flights obtained. In each case, the top anomalous flights are selected from the entire data set of around three thousand flight records. The total amount of overlap in the anomalous flights is then obtained to identify how many flight records get missed when limited information is available or when energy metrics are not used at all. Since there is no truth value to compare against, it is desired that the set of flights identified as anomalous is largely invariant with respect to the feature vectors employed.

5.5.1.3 Results

The overlap of anomalous flights among the different feature vector options identified by the methodology is shown in Figures 59 and 60. The two figures represent the overlap at two different significance levels (3% and 5%)

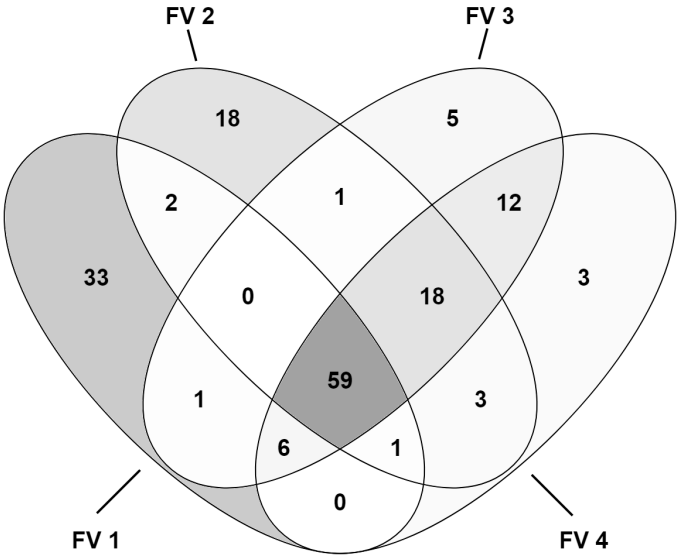


Figure 59: Overlap at 3 % outlier significance level

At the 3% significance level, 102 anomalies are identified among the data whereas

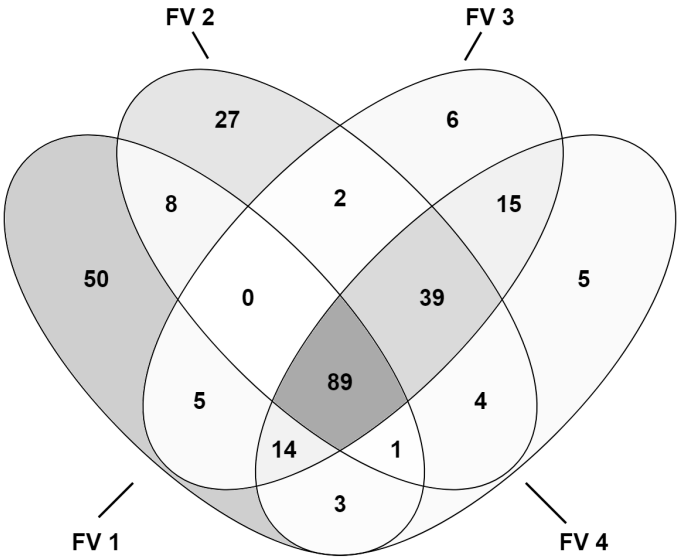


Figure 60: Overlap at 5 % outlier significance level

for the 5% significance level, 170 anomalies are identified. Pre-determining the proportion of anomalies this way ensures that while using each type of feature vector the same number of anomalies are obtained. The overlap is shaded darker for higher number of flights. A few interesting trends are apparent from the figure. Firstly, there is a fair amount of overlap among the different feature vector options for both significance levels. The most significant overlap of anomalies is common to all feature vector approaches ($\approx$ 58% for the 3% significance level and $\approx$ 52% for the 5% significance level). This set is not only observed in all the approaches but is also the set with the worst anomaly scores across all feature vector options. Therefore, the most significant anomalies observed in the existing data set are captured by energy metrics from all levels of sophistication as well as raw flight parameters.

The next observation is related to different feature vectors obtained from energy

metrics (2,3,4). The overlap of anomalous flights obtained from energy-metrics feature vectors is $\approx 75\%$ for both the 3% and 5% significance levels. Therefore, even with limited amount of information available for evaluating energy metrics, the methodology is able to recover 75% of the same anomalies. This implies that even if flight data records with limited amount of parameters are included in the data set, there is a good probability of identifying anomalous ones (especially those with high anomaly scores) among these flights. Another important observation from the figures is that there is a set of flights ($\approx$ 16% in both) that are identified as anomalous by the energy metrics approach alone. These flights are identified as anomalous using the extra capabilities provided by predictive models and corresponding metrics defined.

It is noted that there are some flights that do not get recognized with energy metrics but do get recognized with flight parameters ($\approx$ 30% of the records for both significance levels in FV 1 section). The anomalies detected by FV 1 only correspond to unusual variations of parameters that are not explicitly used in the energy-metric definitions [e.g., roll angle, latitude, longitude, heading, track, fuel quantity, oil temperature, cylinder-head temperature (CHT), etc.]. The anomalies detected by FV 1 only can be of different types: flights that land at a different runway/airport (anomalous due to latitude, longitude, track, etc.); flights that follow different procedures (to enter the traffic pattern for instance); and flights that have abnormalities in other parameters not used in energy metrics (oil temperature, CHT, etc.). While these types of flights differ in terms of their parameter variations, it does not necessarily indicate deviations in terms of their energy metrics and therefore do not get identified as anomalous in the energy metrics based approach. However, even for identifying anomalies that are significant for safety analysis, the recording capability for these additional parameters needs to be available for all the aircraft (which, as argued earlier, it is not). Therefore, if the necessary data is available from the flight data record, it is possible to use FV1 to complement the demonstrated energy metrics based approach,

to further increase the probability of detecting truly anomalous events.

Finally, for each feature vector, there are a few stray flights which get identified as anomalous which can be considered as a consequence of the algorithm forcing a fixed number of total anomalous flights to be identified in each case. The main appeals of the energy-metric approach (FVs 2-4) are the requirement of a parsimonious set of parameters, (relative) invariance to the size (weight) of the aircraft, and applicability to multiple airports/runways.

5.5.2 Model Generalization

Apart from data recorder limitations, it is also important to gauge the generalization capability of the SVM itself on this type of problem. SVMs have been shown to be powerful algorithms for classification, however, as with any other machine learning algorithm, their performance can suffer if the parameters are tuned too much to the available data (resulting in the phenomenon known as 'overfitting' in the machine learning community).

5.5.2.1 Purpose of Experiment

Does a support vector machine model trained on a subset of an existing data set predict the same anomalies that would be predicted had the entire data set been used to train the model?

This situation can arise in deployment of this methodology when the entire data-set may not be available to any operator for generating the model or when the models generated on a data set need to be used for identifying anomalies for a different data-set where the data cannot be shared but the trained model can be. In these situations it is useful to know how well the methodology and models can generalize.

5.5.2.2 Experiment Setup

Typically, cross validation is the technique used in classification problems to test how well a technique will generalize to an independent data set. One of the common implementations of cross validation is k-fold cross validation [129]. In this technique, the data set is partitioned into k distinct equal sized sets. One of the sets is held back for validation and the model is trained on the remaining k-1 sets. This process is repeated k times until each set has been used exactly once for validation. However, cross-validation typically used in classification problems cannot be used in the same way for a one-class learning problem as there are no truth values available to compare the performance of the algorithm on the validation set. Using existing benchmarks such as exceedance detection would limit the algorithms to only identify known anomalies and hinder its potential for discovering anomalies that are not already defined in the exceedance set. Therefore, in this dissertation, a slightly different approach is used.

The ability of the algorithm to identify anomalous flights using only a subset of the total available flight records is obtained. For this purpose, progressively lower numbers of flights are chosen to train the model and the ability of the trained model to identify the same anomalous flights as that identified by the full data set is evaluated. In each case, a random subset of the flights is chosen to train the model and this trained model is used to predict anomalies on the entire data set. Since the smaller subset can be chosen in many different ways, the SVM algorithm is run multiple times at each setting and the results for overlap of anomalous flights are averaged.

5.5.2.3 Results

It can be seen from Figure 61 that when progressively lower number of flights are used to train the model, the overlap of anomalous flights decreases slightly from 100% (full data set used in training) to around 92% (10% flights used in training).

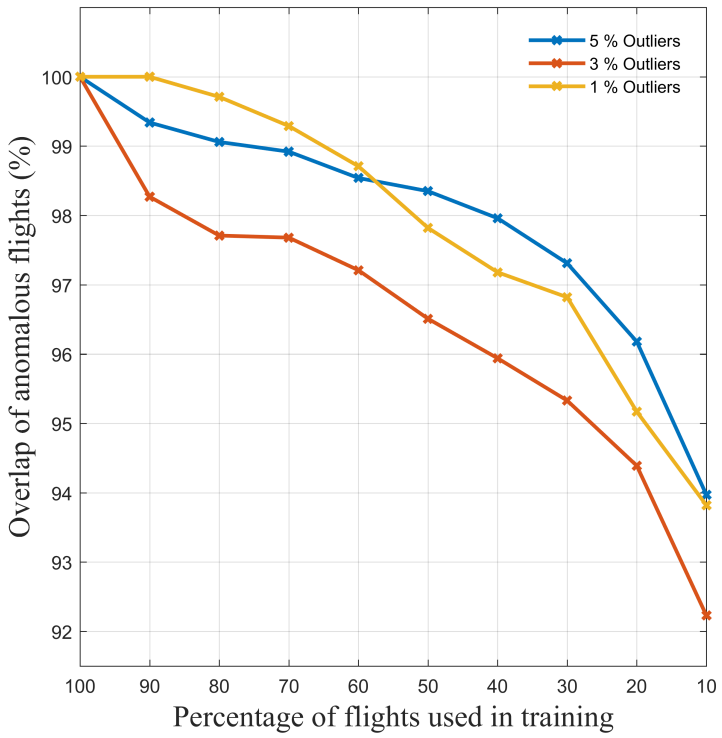


Figure 61: Generalization capability of trained one-class SVM during approach and landing

This trend is seen across different outlier significance levels indicating that whatever the overall proportion of anomalies assumed in the data set, using a smaller set of data to train the model does not significantly reduce the ability to identify anomalies. This shows that the ability of the model to capture important features of anomalous flights is robust and it can generalize well to situations where the trained model is to be deployed on a different set without sharing the flight data itself.

5.5.3 Summary

This experiment has demonstrated the generalization capability of the developed methodology in two different scenarios. While limitations imposed by the flight data recorder capabilities are more restrictive than the limitations of SVM model generalization, in both cases a reasonable level of capability is still retained when the specific

restrictions of each scenario are imposed. The three experiments described in this chapter enable addressing research question 2. It is reiterated that while the anomalies are automatically identified by the methodology developed in this work, they are not necessarily unsafe flights. The goal of the anomaly detection framework is to serve as the first step in the retrospective safety analysis process. Thereafter, the review of the identified anomalies by subject-matter experts or safety analysts will complete the safety assessment process and result in conclusions regarding unsafe events. The goal of this dissertation, however, is to demonstrate techniques pertaining to the former step, and not the latter.

CHAPTER VI

MODEL CALIBRATION (RESEARCH QUESTION 3)

The retrospective safety analysis pursued in this dissertation depends on availability of reasonable performance models. However, using performance models created by fairly simple methods are subject to uncertainty and error. Therefore, a formal technique for calibrating these models to known performance data can improve their predictive value and enable analyses for multiple aircraft.

6.1 Motivation for Model Calibration

The section provides motivation for use of performance models to gain predictive capabilities in retrospective safety assessment, existing models, and limitations.

6.1.1 Performance Model Uses

The development and use of performance models is ubiquitous in a broad range of performance studies. One of the main advantages of using performance models in safety analysis tasks is the ability to use recorded data and predict unrecorded quantities of interest. For example, accurate predictions of thrust and drag utilized with the kinetic and potential energy rates from flight data may enable inference regarding flap usage in flight. This is a very important piece of information that may be utilized in retrospective safety analysis to identify incorrectly configured approach and landings etc. Similarly, accurate estimation of thrust and drag enables the calculation of some of the energy metrics identified earlier in Figure 24. The quality of results obtained from anomaly detection techniques depends on the availability of well-calibrated performance models. Performance models can also be used in a flight simulation model such as that presented in Chakraborty et al. [33] to simulate various conditions for

GA aircraft. Simulated flight data generated using the models calibrated in this work is used earlier in anomaly detection. Use of such simulated data to test unsafe or unusual conditions helps understand the boundaries or limits of safe operations and validate developed algorithms to some extent as seen in previous sections. While these and other uses of performance models can be envisioned, the primary interest in performance models in this dissertation is for the accurate prediction of energy metrics.

6.1.2 GA Performance Models and Uncertainties

Harrison et al. [72] and Min et al. [102] demonstrated the development of performance models for propulsion and aerodynamic characteristics of typical GA aircraft using publicly available data and handbook methods. These models consisted of variations of typical non-dimensional quantities relevant to aircraft performance such as lift and drag coefficients, thrust coefficient, propeller efficiency, etc. Due to lack of actual flight data, these models were individually validated against publicly available data such as engine handbooks, pilot operating handbook, manufacturer's manuals, other published literature etc. Such models can be used along with information of flight conditions to provide estimates of the performance of the aircraft. However, empirical models such as these can suffer from uncertainties and errors.

The uncertainties in the performance models can stem from various sources. There are two main types of uncertainty [133] – aleatory uncertainty due to inherent variation (irreducible) and epistemic uncertainty due to lack of knowledge (reducible). Even when dealing with performance models for the same GA aircraft, there can still be several reasons due to which the performance actually observed can differ from predicted performance. These errors arise due to the cumulative effect of unmodeled factors and random effects on the estimate. In the context of the present models,

some of the important factors that can cause a difference in actual versus predicted performance are:

- Variations in aircraft gross weight between flights and the fact that gross weight is most often unknown (epistemic)
- Degradation in performance of aircraft systems (such as engine etc.) with age (epistemic) or additional drag due to unclean surfaces, insect accretions, etc.
- Changes or modifications made to aircraft that affect its aerodynamic behavior (epistemic)
- Variation in aircraft model (C172S versus C172R - epistemic)
- Model Inadequacy (epistemic)
- Unknown/Inaccurate model parameters (epistemic)
- Piloting skill (epistemic)
- Noise in recorded data (aleatory)
- Environmental conditions (aleatory)

As noted by Kennedy and O'Hagan in their seminal paper [79] – to use a model to make predictions in a specific context it may be necessary to first *calibrate* the model using some observed data. These observations had motivated Research Question 3 which is restated here:[1]

[1]The research described in this chapter is documented in the following publication:

– **Puranik, T.G.**, Harrison, E.H., Min, S.M., Chakraborty, I.C., and Mavris, D.N., *A Framework for General Aviation Aircraft Performance Model Calibration and Validation*, in 18th AIAA Aviation Technology, Integration, and Operations Conference, Atlanta, GA, 2018. Accepted for publication [116]

Research Question 3:

How can basic empirical models of aerodynamic and propulsive performance of a generic GA aircraft be calibrated to predict energy metrics at conditions of interest for a specific aircraft?

Calibration can mean different things according to the context it is used in. Model calibration is defined as "*the process of adjusting numerical or physical modeling parameters in the computational model for the purpose of improving agreement with experimental data*" [9]. The overall process of calibrating computational models has been outlined by Oberkampf et al. [112, 113] and has been reproduced in Figure 62. The process consists of comparing the predictions from computational models with some benchmark data. The comparison is usually made in terms of certain critical output parameters called validation metrics by Oberkampf [112]. The accuracy requirement for models is externally imposed and is generally related to the context in which the model is to be applied for prediction. In the type of calibration undertaken here, the form of the model is assumed to be fixed and the parameters are varied to achieve accuracy of model predictions. In such cases, that the values of context-specific inputs and variables are unknown or uncertain and the observations or benchmark data are used to tune these parameters. There exist other types of model calibration frameworks such as system identification [78] or Bayesian Calibration [79] in which the model form is not assumed to be fixed which are not considered here as they require more detailed input-output data which is typically not available for GA performance and safety studies.

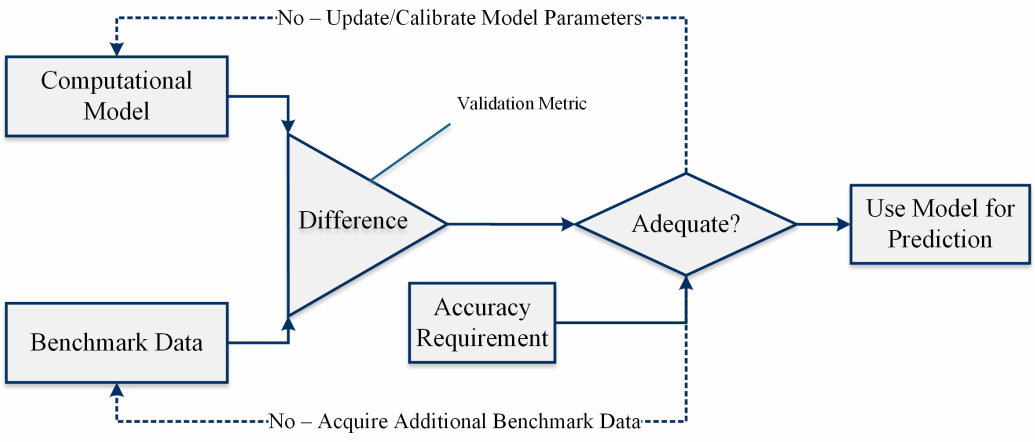


Figure 62: Calibration of computational models and use for prediction (reproduced from Oberkampf and Barone [112])

6.1.3 Existing efforts and limitations in GA

Aircraft performance model calibration literature is relatively sparse and, when available, it is problem-dependent. There have been previous efforts in aircraft performance model calibration in different domains such as trajectory optimization, performance monitoring, etc. Most of them have been performed on commercial aircraft operations. For example, Bronsvoort et al. [26] have performed real-time trajectory prediction-calibration using Base of Aircraft Data (BADA) [111] performance models. Their calibration involves static corrections (bias) as well as dynamic ones (operational changes etc.). The 'kinetic' calibration undertaken results in a calibration factor for each segment thereby establishing a calibration function. The work in this dissertation is aimed at establishing a single calibrated model for an aircraft-type/tail-number and does not propose varying the calibration factors within a flight. However, the need for static calibration corroborates the ideas introduced in the earlier section about errors and uncertainties in model predictions compared to actual performance. Krajcek et al. [81] have performed calibration or performance degradation estimation using fuel-flow calculations. Their approach for calibration allows distilling out contributions of individual models to the total performance degradation

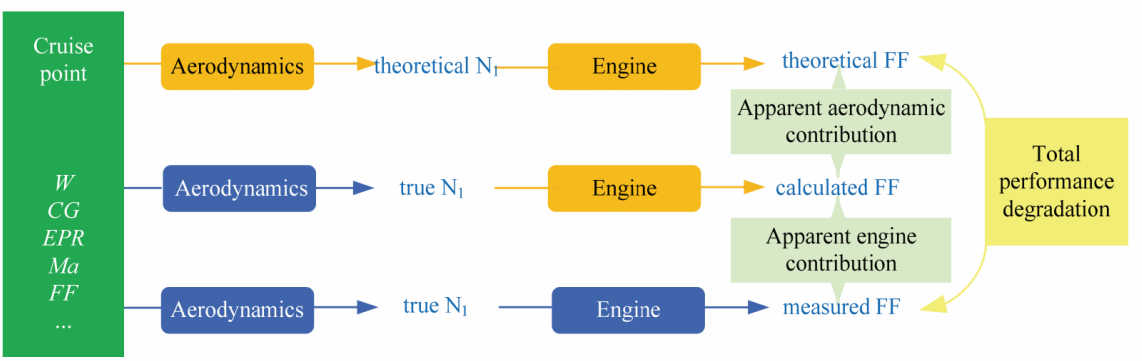


Figure 63: Contributions to total performance degradation [81]

as shown in Figure 63. While this approach requires predictions from engine models beyond the capabilities of GA FDR, it is worthwhile to note the quantification of the differences expected in theoretical and actual aircraft performance. The previous studies indicate the value of performance models as well as the need to calibrate them to be useful for making accurate predictions.

6.2 Model Calibration Framework

The discussions in the preceding sub-section lead to the development of the 2–level model calibration framework developed in this dissertation and presented here. The overview of the framework is presented in Figure 64. As noted previously, many of the decisions made in GA safety analysis depend upon the availability of data. The previous empirical models developed for GA operations ([72, 102]) utilized public domain data due to lack of flight data. These models were independently calibrated against available benchmarks. However, rarely are such performance models used in isolation. Sometimes, metrics that are not estimable using a single model alone can be estimated using predictions from a combination of models. While this could add uncertainty to the prediction, it opens up new avenues to test, calibrate, and fine-tune model parameters. Similarly, in the event of availability of some amount of flight data, a method needs to be developed to utilize that information to improve model predictions. Therefore, two levels of calibration-data availability are defined

(**publicly available data** and **flight data**) and make up the corresponding two levels of the developed model calibration framework.

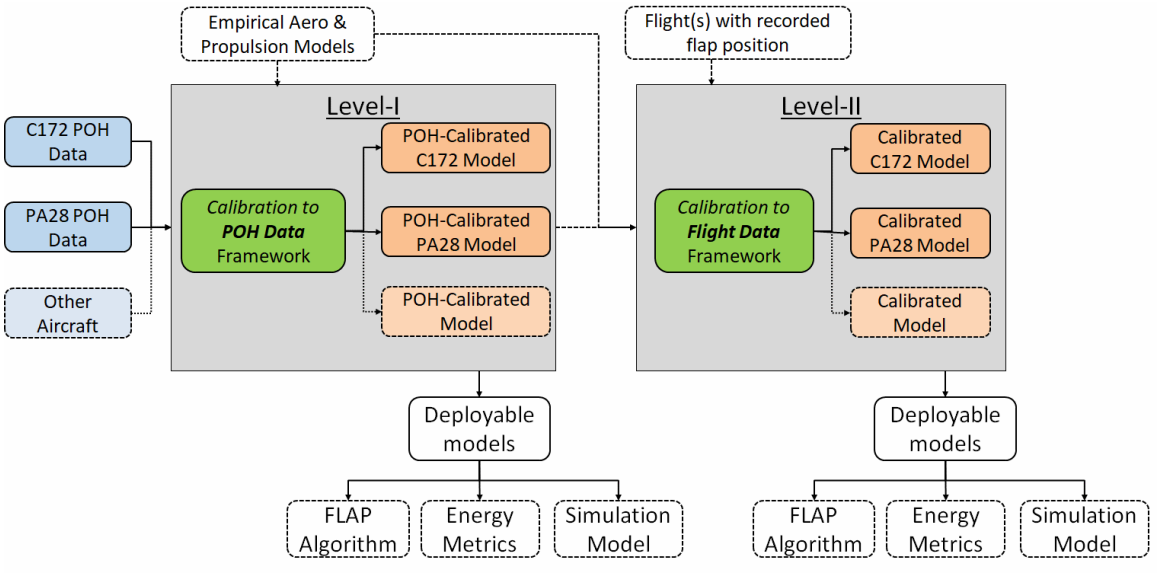


Figure 64: Overview of 2–level model calibration framework developed in this dissertation

In the context of this dissertation, an aircraft performance model consists of two individual disciplinary models - aerodynamics and propulsion. The various inputs and constituents of the framework are described here and the implementation of each level is described in the subsequent sections.

6.2.1 Aircraft Data

Basic data about the geometry of the aircraft along with publicly available performance data for calibration and validation is gathered. Manuals published by the manufacturer of the aircraft such as the Pilot Operating Handbook (POH) are useful in this regard. The POH is a document developed by the airplane manufacturer and approved by the FAA, which lists important information regarding the design, operations, and limitations of the aircraft, as well as its performance characteristics. Although the performance tables listed in the POH are idealized capabilities that a

brand new aircraft can theoretically attain under ideal conditions and expert piloting, it is nevertheless a very valuable source of information for calibrating models. As the POH is also readily available for most aircraft, it would enable a calibration process that is easily repeatable. The data from the performance tables in the POH is collected in comma separated variable (CSV) format and used for calibration.

6.2.2 Empirical Models

Figure 65 shows the flow of information in the individual discipline models used in this work. Given a set of inputs and environmental conditions, it is of interest to predict the aerodynamic and propulsive performance of a GA aircraft in terms of certain non-dimensional coefficients.

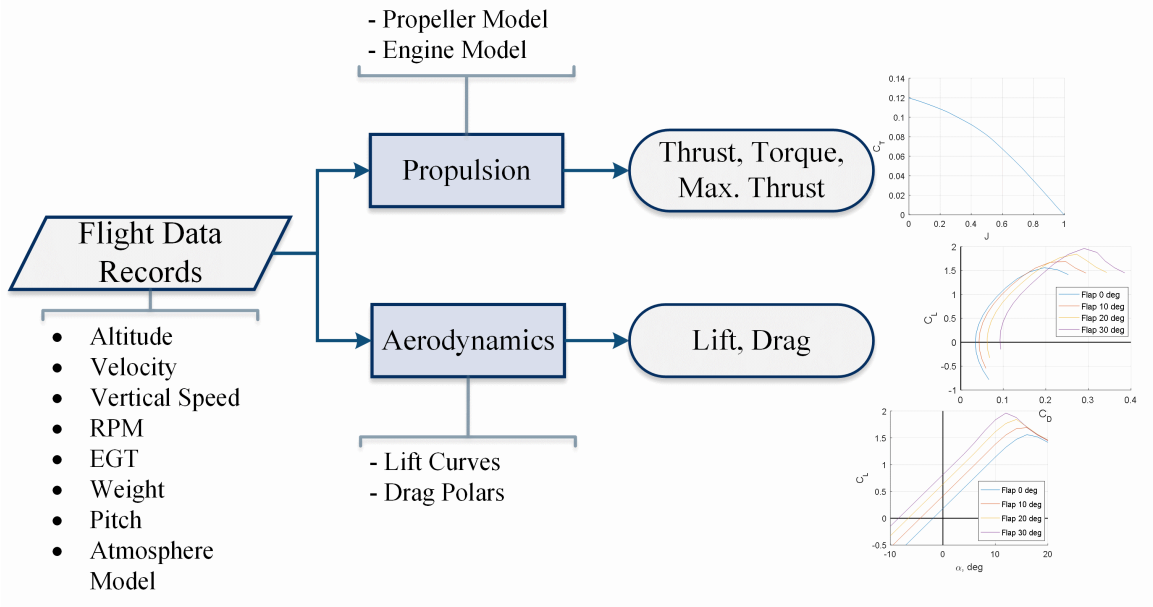


Figure 65: Information flow in aircraft performance models

6.2.2.1 Propulsion

Among currently operated GA vehicles, the most common means of propulsion is a propeller driven by an internal combustion engine. Following this trend, the propulsion model is similarly divided into individual models of internal combustion engine

performance and propeller performance, which combined yield a propulsive model that is appropriate for a large subset of the GA fleet. Many approaches for engine modeling exist within the literature, and a short survey and implementation of some notable methods can be found in Harrison et al. [72] In that work it was seen that a detailed modeling of the engine's cycle produces accurate results, though a large number of input parameters must be specified and a high computational cost is incurred. This expense can however be avoided while maintaining a similar level of accuracy by using published relationships for engine power lapse as functions of local atmospheric density ratio, such as that given by Gudmundsson [70]. This latter approach from engine modeling is chosen. Alongside the engine model, a model of propeller performance was implemented which predicts the efficiency characteristics of a given propeller geometry. As an input, the model requires propeller geometry in the form of blade chord lengths and pitch angles at several intervals along the length of the propeller blade. This geometric data is combined with airfoil aerodynamic data computed using XFOIL [48], an open-source 2-D aerodynamic prediction program. To generate the 3-D performance of the full propeller, the geometric and 2-D aerodynamic properties are passed to a open-source analysis tool for propeller and windmill analysis called QPROP [49]. When combined with the engine model, this propeller model yields a prediction of the propulsive performance of the desired GA aircraft over a wide range of operating conditions.

6.2.2.2 Aerodynamics

A variety of aerodynamic modeling methods for a fixed-wing aircraft exists ranging from first order approximations to high-fidelity computational fluid dynamics tools. Among the existing aerodynamic performance prediction methods, a theoretical empirical-based modeling method is the most appropriate method for developing an aerodynamic model for a fixed-wing general aviation aircraft because it only requires

a minimal set of information that is publicly available and provided by authoritative and reliable sources. Min et al. reviewed, implemented, and comparatively evaluated some of the most well-known physics-based modeling methods for aerodynamic performance prediction as applied to a representative general aviation aircraft [102]. The inputs for the aerodynamic model developed by Min et al. are aircraft geometric information from pilot's operation handbook, publicly available empirical data [7, 132], and typical operating conditions. The outputs provided by the developed model are lift curves and drag polars for flap-deployed configuration as well as clean configuration. Although this aerodynamic model is validated against some reference data published in literature, the main limitation of this modeling method is the difficulty of acquiring reliable reference data for model validation.

6.2.3 Parameterization

The empirical models available are in the form of curves that can be queried for the flight at the conditions of interest. In order to calibrate these models, it is necessary to parameterize the curves using appropriate factors to provide flexibility to the calibration process. In the calibration process, these factors are varied to modify the nature of the curves as desired. These so-called *calibration factors* correspond to certain physical characteristics of the aircraft. Therefore, factors for aerodynamics would be related to the lift-curve slope for airfoil, wing, and aircraft, the drag coefficient increment for different flap settings, scaling of intercepts for different curves, etc. Similarly for the propulsion models, sample parameters could be sectional airfoil characteristics of the propeller, slope shift of the propeller pitch, scaling of propeller chord, etc. The advantage of parameterizing in this manner is that it allows for an easier setup for optimization that needs to be performed for calibration and also allows modifying individual curves without affecting others.

A total of thirty calibration factors are included in the calibration process and a

detailed list of each factor, their description, and range of possible values is included in Appendix F. For each calibration factor, judgment is used in determining whether its effect should be additive or multiplicative on the curve. Similarly, the upper and lower limits also need to be chosen carefully so as not to defy physics-constraints of the problem.

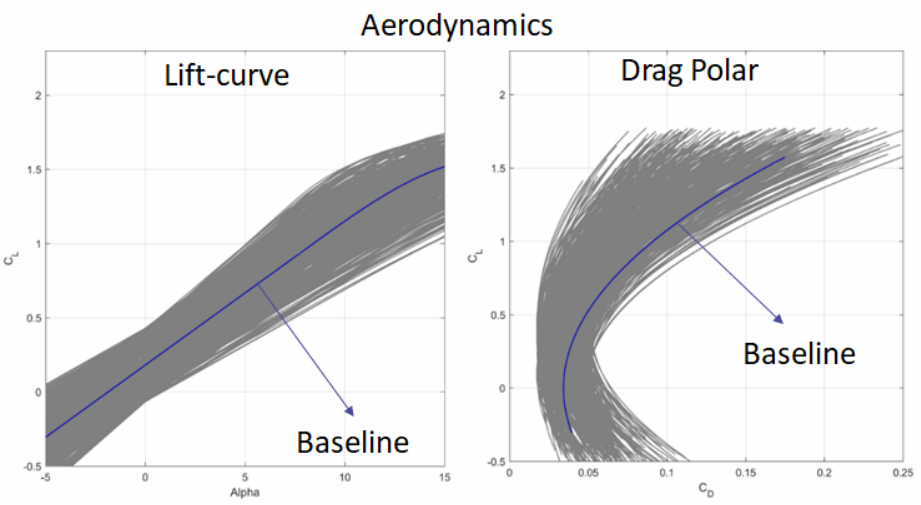


Figure 66: Spread of possible lift curves and drag polars using proposed parameterization

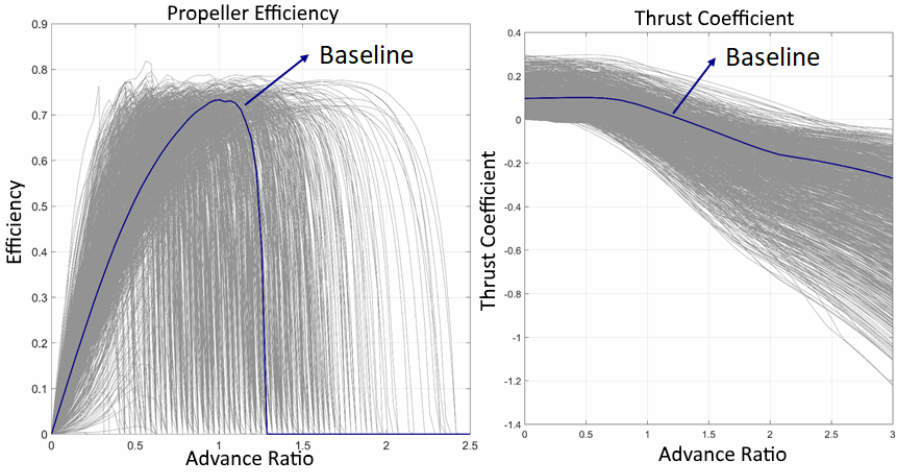


Figure 67: Spread of possible curves for thrust coefficient and efficiency of propeller using proposed parameterization

Figures 66 and 67 provide a visualization of the spread of possible curves that can

be obtained using the proposed parameterization in Appendix F. The baseline curves (in blue color) in each case correspond to the empirical models from Harrison et al. and Min et al. The grey curves represent 10,000 possible curves obtained when the calibration factors are varied randomly within the limits prescribed for them. As is evident from the figures, the parameterization provides a large possible spread for each performance curve and thus enables covering a big portion of the design space. The curves for aerodynamics shown here correspond to clean aerodynamic configuration. For each such clean configuration, there are multiple flapped configurations with corresponding lift and drag characteristics. The spread of these curves is included in the appendix. The propulsion curves are generated assuming a generic GA propeller as the baseline and varying the parameters of the propeller according to the ranges in the appendix. The variation of the thrust coefficient and propeller efficiency with advance ratio is demonstrated in Figure 67.

6.2.4 Calibration Setup

In order to adapt the model calibration framework developed in this dissertation to the standard framework presented in Figure 62, four important components need to decided in each level of the calibration. These components have been enumerated here and described in further detail in the description of each level:

1. Benchmark Data
2. Discrepancy Metric
3. Calibration Factors
4. Optimization

6.3 Calibration to POH Data – Research Question 3.1

The first research sub–question for model calibration based on data available in the public domain can now be stated as:

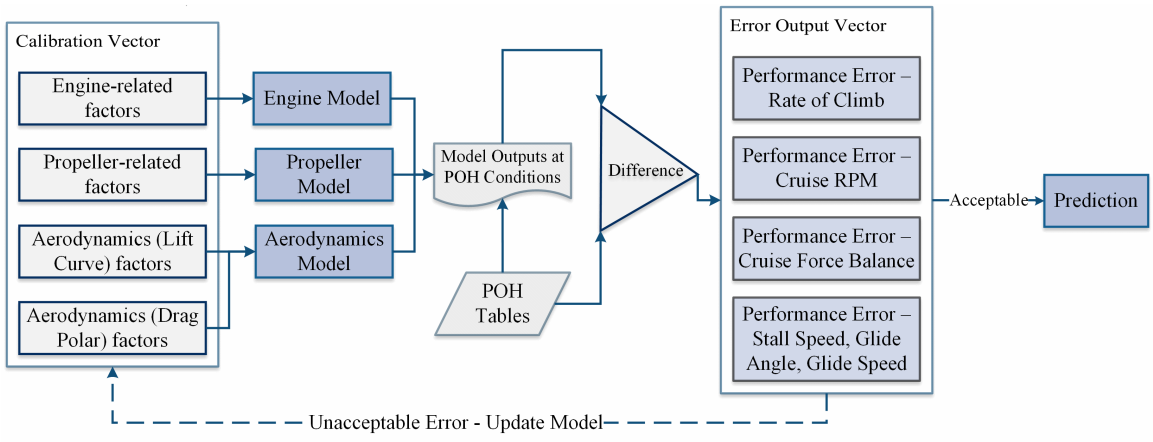


Figure 68: Information flow in calibration of aircraft performance models

Research Question 3.1:

How can basic empirical models of aerodynamic and propulsive performance of a generic GA aircraft be calibrated using data only available in the public domain to predict energy metrics at conditions of interest for a specific aircraft?

Figure 68 shows the overall process of calibration of models to POH data developed in this methodology. The four important components identified earlier are elaborated further in order to develop the hypothesis.

1. Benchmark Data:

As it is of interest to develop a repeatable way of calibrating aircraft performance models, benchmark data for level-1 calibration should be sought from performance data about the aircraft available in the public domain. A good available source of aircraft performance is the Pilot Operating Handbook (POH) of the aircraft. In addition to performance characteristics, the POH can be used to identify or obtain values of certain input parameters such as those relating to the geometry of the aircraft. As the POH is also readily available for most aircraft, it enables a calibration process that is

readily repeatable. As both aerodynamic and propulsion models are being calibrated simultaneously, the benchmark data in the POH that require both estimates can be used in calibration. This leads to the following hypothesis for research sub–question 3.1:

Hypothesis 3.1: *Using tables and conditions in the Pilot Operating Handbook to calibrate the performance models results in a process that will yield models for a specific aircraft which can be used for predicting energy metrics of interest.*

2. Discrepancy Metrics:

The POH contains performance tables that can be used to compare the model output and the expected output at the conditions specified in the tables. While some variation in the format of data presented is expected between different POHs, most POH have some common performance tables – those related to rate of climb, cruise performance, stall speeds etc. In addition, the POH typically also identifies best glide speeds, speed for best rate of climb etc. All of these conditions can be used to calibrate the performance models. For the purpose of this dissertation, the conditions specified in Table 17 have been utilized.

The performance tables/condition in some cases have multiple points for comparison (such as cruise, rate of climb) and in other cases have a single scalar value to compare (glide angle, best glide speed). Therefore, the root mean square (RMS) error metric is computed for conditions having multiple calibration points so that a single value of error metric is obtained for each performance condition. Therefore, as seen from Figure 68, each calibration vector results in multiple error measures as represented in the table. Further details on how each metric is mathematically computed and the data used from POH are provided in Appendix G.

Table 4: Various conditions from POH used in level–1 model calibration

Calibration Factors Utilized From	**Phase of Flight/ Condition**	**Discrepancy Metric**
Aerodynamics, Engine, Propeller	Rate of Climb	$\frac{(ROC_{pred} - ROC_{POH})}{ROC_{POH}}$
Engine, Propeller	Cruise RPM Imbalance	$\frac{(RPM_{pred} - RPM_{POH})}{RPM_{POH}}$
Aerodynamics, Propeller	Cruise Thrust-Drag Imbalance	$\frac{(T-D)}{T}$
Aerodynamics	Stall Speeds	$\frac{(Vs_{pred} - Vs_{POH})}{Vs_{POH}}$
Aerodynamics	Glide Angle	$\frac{(FPA_{pred} - FPA_{POH})}{FPA_{POH}}$
Aerodynamics	Best Glide Speed	$\frac{(Vbg_{pred} - Vbg_{POH})}{Vbg_{POH}}$

3. Calibration Factors:

The calibration vector contains various factors that correspond to the physical aspects of the performance curves whose values may or may not be accessible from literature. There are a total of 30 calibration factors are defined to parameterize the models (Table 17 in Appendix F). Each combination of calibration factors results in a unique performance model which can then be tested at various conditions. However, to effectively calibrate using a factor, there must be benchmark data available. Otherwise, changing the value of the factor will not have any effect on the calibration. For example, if the POH does not contain reference data for flights with flaps set at 10 degrees, the effect of varying the calibration factors which affect the aerodynamic characteristics of that configuration on the discrepancy metrics cannot be ascertained. Therefore, those factors for which there is no benchmark data are not varied in level-1 calibration in order reduce computational time and avoid erroneous results.

4. Optimization:

The discrepancy metrics resulted in a vector of errors for each calibration vector based which quantify the differences between performance as predicted by the performance models and actual published performance figures. The calibration process involves tuning the calibration vectors to minimize the discrepancy. Therefore, it raises the following methodology research question:

Methodology Research Question: Does there exist a unique calibration vector that minimizes all the errors calculated from the POH conditions?

Hypothesis: *There is, in general, no unique calibration vector that simultaneously minimizes each error metric for each flight condition, and therefore, multi-objective rather than single-objective optimization should be pursued in the calibration approach*

While the perfect model would ideally simultaneously minimize the error for all POH conditions, due to the epistemic uncertainty in the models it is expected that different phases of flight will have different errors. One of the most likely reasons for this is the fact that the models themselves have different fidelity in different operating conditions. For example, a drag polar from empirical build-up methods for clean configuration is expected to be more accurate than that for a configuration with flaps down, due to the inherent additional complexities involved (such as non–linearity, flow separation, flap geometry etc.) Therefore, a numerical experiment is conducted to test this hypothesis.

Experiment:

Purpose: The main purpose of this experiment is to ascertain whether the same calibration vector can simultaneously minimize different discrepancy metrics from POH. This will facilitate a decision regarding whether single or multi-objective optimization techniques would need to be used for level-1 calibration.

Setup: Due of the relatively low cost of generating the performance models and computing the POH performance errors for each calibration vector, a Monte Carlo simulation is conducted by varying the components of the calibration vector. The values of each calibration vector is varied around its baseline values of the original individual models. A total of 10,000 distinct combinations of calibration factors (i.e. 10,000 calibration vectors) are evaluated and the performance errors with respect to POH are calculated for each.

Results: The resulting distribution of errors is visualized in Figure 69. Along with the distribution of each error metric, the calibration vector which minimizes each of the other three error metrics is noted and its error for the metric under consideration is marked. Therefore, in each error distribution, the optimum value for the other three errors is plotted with vertical lines. As clearly seen from Figure 69, there is no single calibration vector that can minimize the error across all four error metrics considered. Furthermore, the best performing models for each case are noticeably poor in terms of their performance for some other metrics. Therefore, this experiment confirms the hypothesis that there is, in general, no unique calibration vector that simultaneously minimizes each error metric for each flight condition. Therefore, a multi-objective optimization algorithm is used for calibrating the performance models to the POH performance tables. Suitable calibrated models can then be selected from the pareto frontier resulting from the multi-objective optimization.

Since it is not known a priori which POH discrepancy metric would correspond to

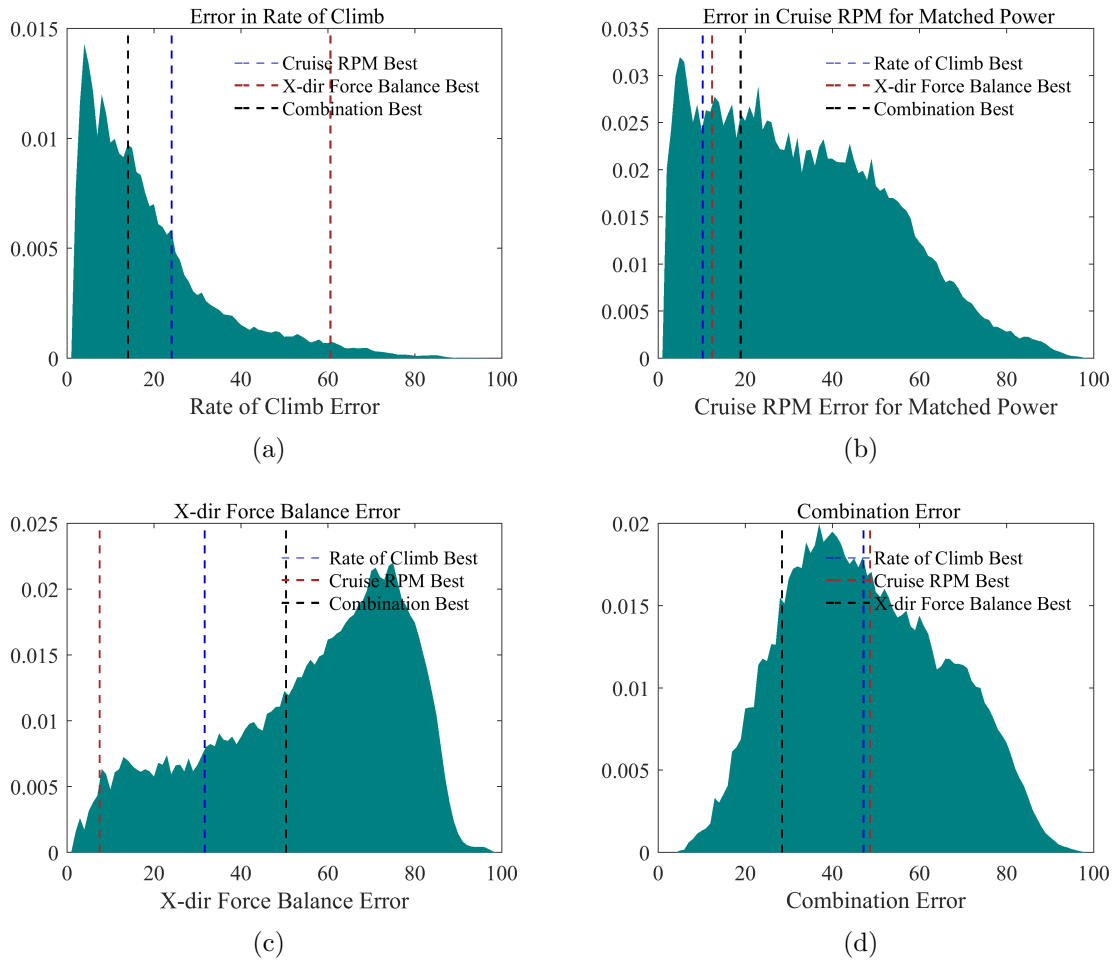


Figure 69: Distribution of calibration errors for different conditions and optimum solutions for other conditions

a model which provides good predictions on flight data, multi-objective optimization enables a pareto-optimal set of models to be obtained. The well known algorithm NSGA-II is utilized for multi-objective optimization [44]. The NSGA-II optimization algorithm is used to drive the MATLAB-based performance model calibration setup to find the pareto front of calibrated models.

6.4 Experiment 3.1

The steps outlined in the previous section provide details on the methodology used for calibrating models to POH data. The methodology experiment enabled identifying that a multi-objective optimization framework is necessary for POH calibration.

This experiment will enable answering research question 3.1 and demonstrating the calibration of empirical models for GA safety analysis.

6.4.1 Purpose of experiment

The following are identified as the main objectives of this numerical experiment:

1. Demonstrate level-1 calibration (using POH data) using the same framework for two common GA aircraft (Cessna 172 and Piper Archer)

2. Test calibrated models using limited flaps-annotated flight data and the specific total energy rate metric

The main reason this experiment is set up in this manner is that it ensures that all information available from a POH gets utilized. It is also a low-cost option that does not require any flights to be flown in order to get reasonable calibrated models. The multi-objective optimization framework provides flexibility and ensures that whichever conditions for calibration are available for a particular aircraft type in the POH can be included in the calibration.

6.4.2 Experiment Setup

The experiment is performed for each aircraft type. The data from Cessna and Piper handbooks is saved in the form of comma separated values in spreadsheets that are read into MATLAB. The multi-objective optimization is set up up in MATLAB using the native NSGA-II algorithm. The discrepancy metrics noted in the previous section are optimized for in the calibration. The embedding of the model parameterization, creation of new models for each unique calibration vector, and evaluation of discrepancy is handled using developed MATLAB functions for each. Those calibration factors for which performance conditions are not available are frozen at their baseline value. For this experiment, one annotated Cessna 172 flight is assumed to be available in addition to a few Piper Archer flights.

6.4.3 Experimental Metrics

In terms of metrics being tracked, the errors with respect to the POH conditions are used in the calibration and also in the visualization of the obtained set of calibrated models at the end of calibration. For choosing the best out of the pareto-optimal set of aircraft, some flaps-annotated flight data is used to evaluate the error between the predicted and actual time histories of the specific total energy rate metric. This metric is useful for basic validation as it can be evaluated using models and flight data independently. More details on the appropriateness of this metric are provided later in level-2 calibration.

6.4.4 Results

The following section contains results from the calibration of performance models to POH data. In the multi-objective optimization, the different error metrics identified earlier are minimized to obtain a set of pareto-optimal aircraft models. Following that, this set of models is tested using limited flight data available to obtain error in the energy metric of interest. Not knowing take-off weight can add extra uncertainty to the model predictions, described further in Appendix C.

6.4.4.1 Cessna 172 Results

Figure 70 shows a visualization of the pareto-optimal set of models obtained after POH-calibration of the Cessna 172 aircraft. Each point in the figure is non-dominated with respect to a discrepancy metric or the other. The metrics in the figure are the POH discrepancy metrics listed in Table 4. As is evident from the figure, there is no obvious choice for the appropriate model to be selected for use in prediction. In such situations, it is difficult to make a choice for the appropriate model and in the absence of further information, multiple models may be used to make predictions. However, it is rarely the case that models are going to be developed as a stand-alone exercise. These models are developed to be used in a data-driven safety assessment

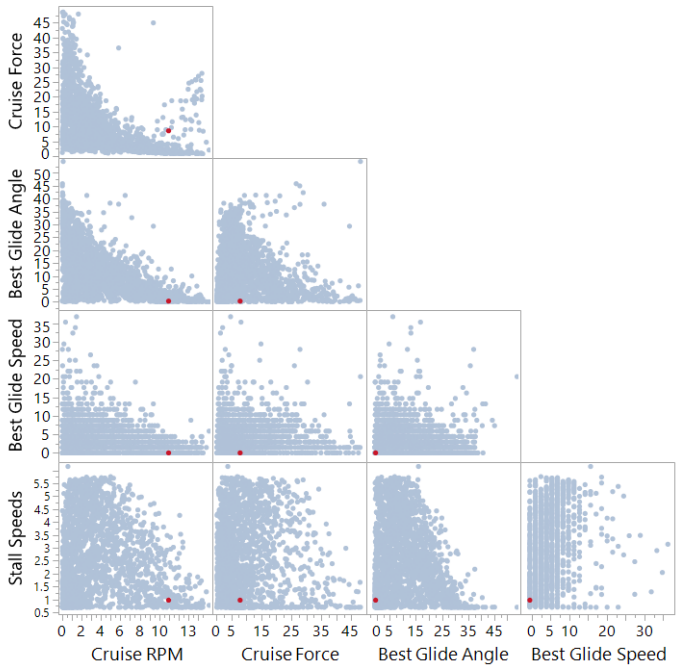


Figure 70: Scatter of pareto-optimal models obtained from level-1 calibration for Cessna 172

framework. Therefore, some amount of additional information is assumed to be available to make the decision. A limited number of flight data records may be assumed to be available for validation and to aid in down-selecting the most appropriate model.

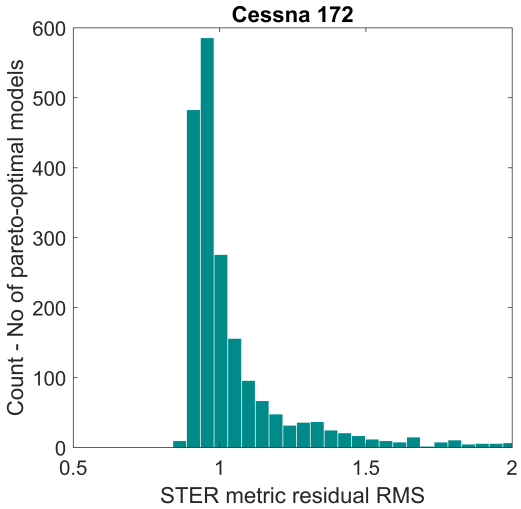


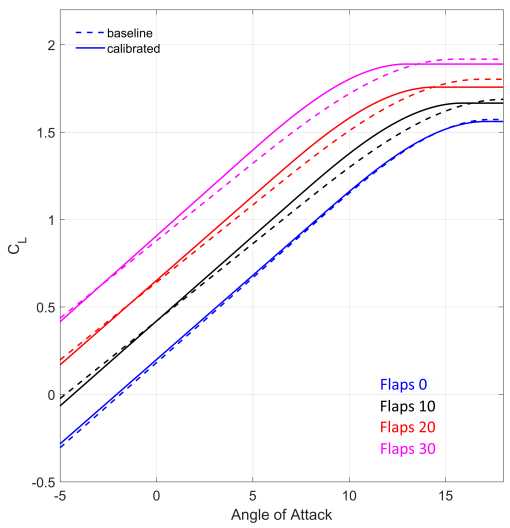
Figure 71: Distribution of STER metric residual for all pareto-optimal models obtained from level-1 calibration

In this section, the most minimal case of a single flight data record from an

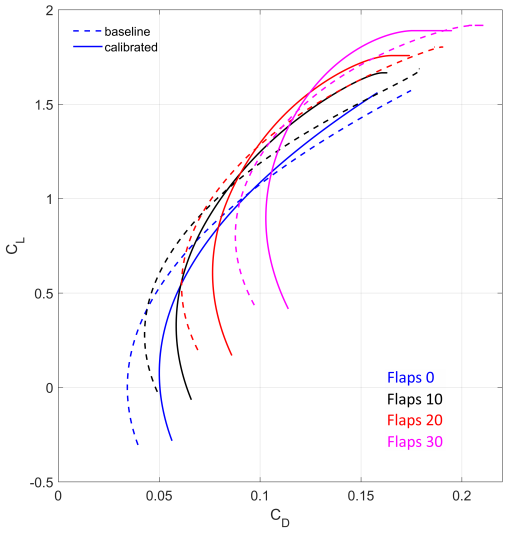
actual Cessna 172 aircraft is used. The next section on the Piper Archer will be demonstrated with additional available data. When GA flight data is available, the specific total energy rate metric used in earlier sections for anomaly detection can be used to validate the model (or in this case, select the most appropriate model). One of the main reasons this type of model selection technique is used is because of the limited number and type of parameters available even from actual flight data. Specific total energy rate is one of the few metrics that can be independently evaluated from models and flight data.

For the current experiment, this metric is evaluated independently using the flight available and each model from the pareto-optimal set. The residual between the metric value is calculated throughout the flight data record and a distribution of RMS of the residual for each pareto-optimal model is obtained. The main reason the residual (and not the relative error) is used directly is to prevent numerical ill-conditioning as the metric value may often be zero. Thus, each pareto-optimal model will now yield a single RMS error for discrepancy of this metric for the flight record available. The histogram of this RMS error for all models from the pareto-optimal set is shown in Figure 71. With the availability of this additional information, the choice can be made for the POH-calibrated model. The model with the lowest RMS error on the actual flight data record is chosen as the calibrated model to be used for prediction. This model is also indicated in Figure 70 as a red dot.

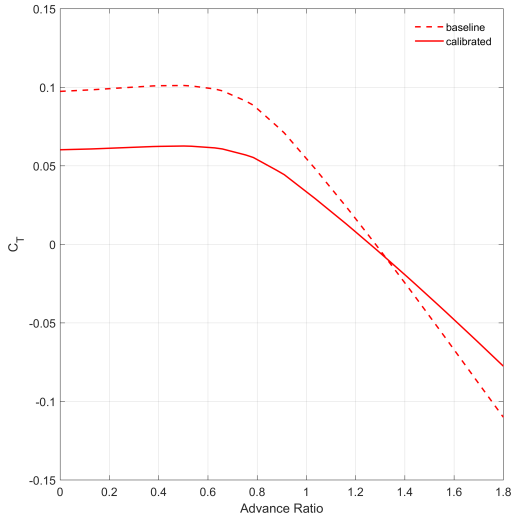
Once the actual model to be used has been selected, it is useful to visualize how the individual aerodynamics and propeller curves for the model changed from the generic GA baseline model. This can be visualized in a series of curves shown in Figure 72. The figures indicate that the chosen calibrated model has a slightly higher lift coefficient for the same angle of attack at all configurations. Similarly, the general trend in the drag polars is that the drag coefficient is higher than the baseline model for the same value of lift coefficient. On the propeller side, the calibrated model has a



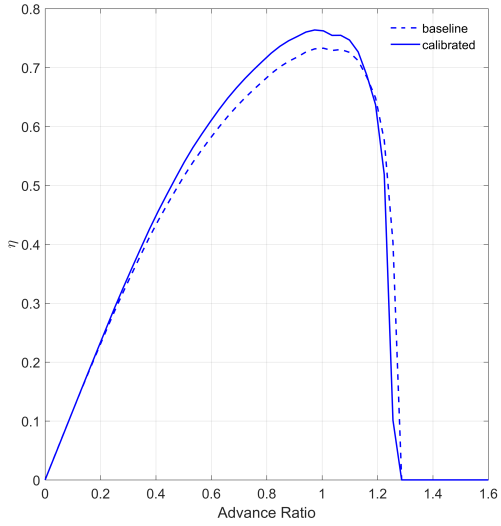
(a) Aerodynamics - Lift



(b) Aerodynamics - Drag



(c) Propeller - Thrust



(d) Propeller - Efficiency

Figure 72: Changes in discipline models before and after calibration

lower thrust coefficient and slightly higher efficiency at the advance ratios of interest.

Since the flight data record is also available it is useful to visualize the data trace of the quantity being compared and used for selection of the model (Specific Total Energy Rate) over the duration of the flight. This can be seen in Figure 73. It is observed from the figure that the calibrated model is able to capture the overall trace of the metric well and has a low RMS of the residual (actual minus predicted) of 0.86

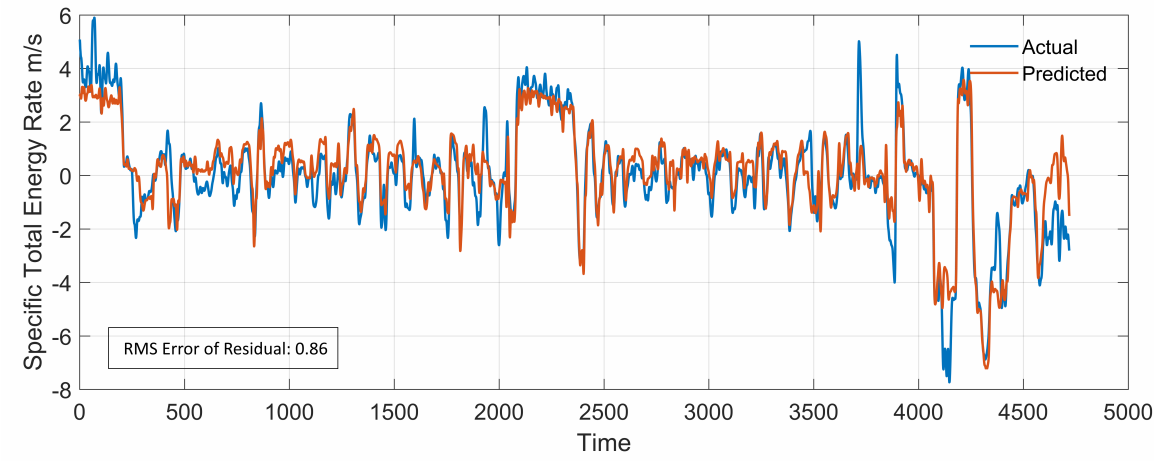


Figure 73: Variation of specific total energy rate metric – actual and predicted (Cessna 172)

for the flight. The residual of the metric is also visualized as a distribution shown in Figure 74. It is observed that the overall residuals are low and approximately normally distributed with a mean close to zero.

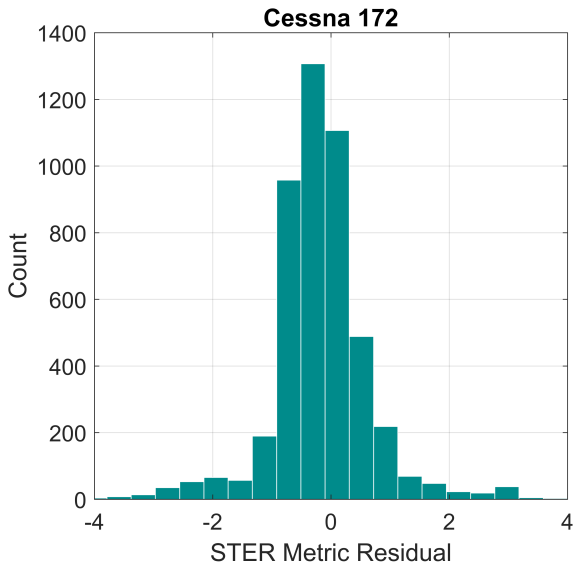


Figure 74: Distribution of STER metric residual for calibration flight (Cessna 172)

6.4.4.2 Piper Archer Results

Similar to the previous sub-section, POH-calibration is also undertaken and demonstrated for the Piper Archer aircraft. Figure 75 shows a visualization of the pareto-optimal set of models obtained after POH-calibration. Like the previous example, there is no obvious candidate model that is better than every other model in terms of all the metrics under consideration. Any of these models can thus be used for prediction of Piper Archer performance. When additional information in the form of a limited number of flight data records is available, it can be used to make the final selection of the model. It is noted that while the flight data is used to guide the decision on which model to select, it is not used in in level-1 calibration for the creation of the calibrated models themselves.

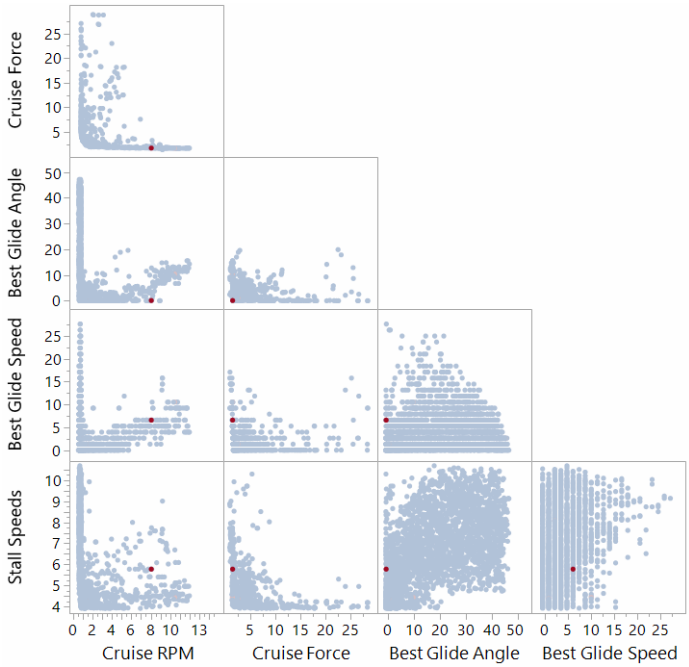


Figure 75: Scatter of pareto-optimal models obtained from level-1 calibration for Piper Archer

While calibration for the Cessna could only be tested on a single flight data record due to limited data availability, more flights were available for the Piper Archer. Therefore, for each model from the pareto-optimal set, the RMS error in metric

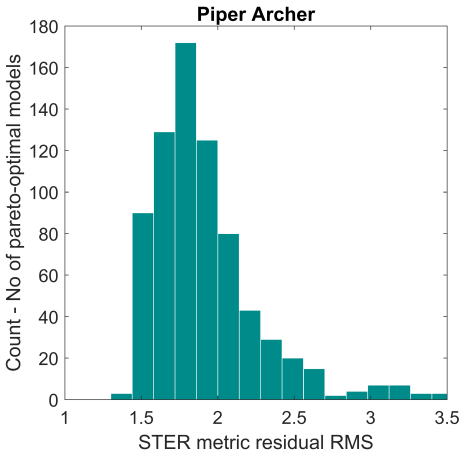
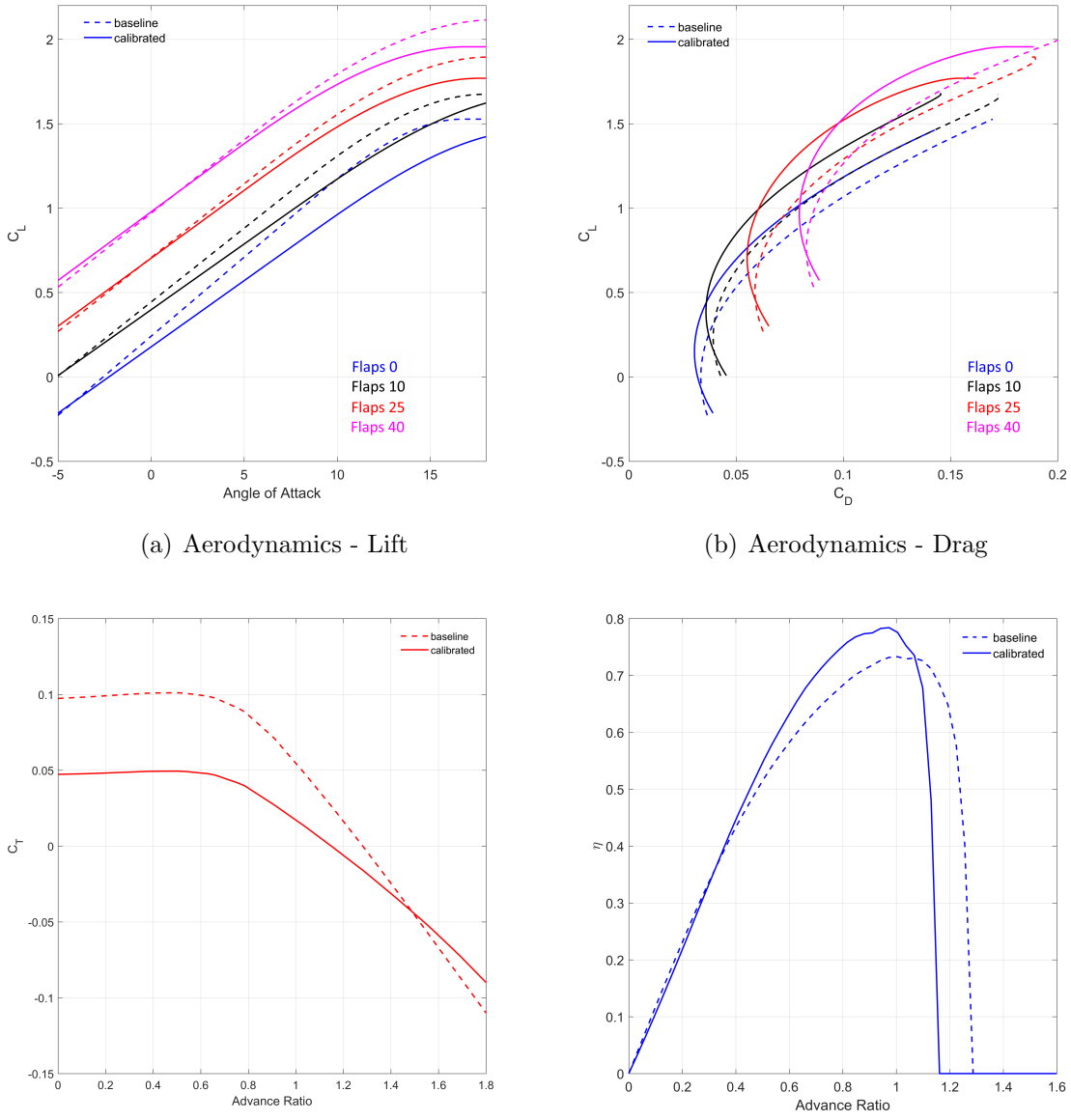


Figure 76: Distribution of STER metric residual for all pareto-optimal models obtained from level-1 calibration

residual is evaluated over the entire set of flight records available for validation. Thus, each pareto-optimal POH model will have a vector of RMS errors – one for each validation flight. The RMS of this vector of errors is calculated yielding the overall RMS error which is shown in Figure 76. The model with the lowest overall RMS error over the entire validation set is then chosen for prediction.

The changes in the actual aerodynamics and propulsion models is seen below in Figure 77. There are downward shifts and slope decreases for each of the lift curves whereas, the drag polar changes are not consistent throughout. At low lift coefficient values, the drag is slightly higher, but it is lower in the higher lift coefficient regions. The calibrated thrust coefficient is lower than that for the generic baseline and the efficiency is lower at low advance ratios and higher close to advance ratios of unity.

The calibrated model can be used to predict the energy metric for all the flights in the validation set. The distribution of RMS errors for all the flights in the validation set using the chosen calibrated model is shown in Figure 78. As seen in the distribution, the overall errors of the chosen model are low compared to the errors generated using other models in the pareto-optimal set. There is a spread observed in the distribution indicating that even the calibrated model is not uniformly effective



(a) Aerodynamics - Lift

(b) Aerodynamics - Drag

(c) Propeller - Thrust

(d) Propeller - Efficiency

Figure 77: Changes in discipline models before and after calibration (Piper Archer)

over all the flight data records available and that it performs better for some flight records than others. This is due to the various epistemic uncertainties mentioned previously related to actual flight data.

As more flight data is available for the Piper Archer, data traces of the predicted and actual energy metric can be visualized for some flight data records. Two sample flights are visualized in the following figures. It is noted that there is some variability

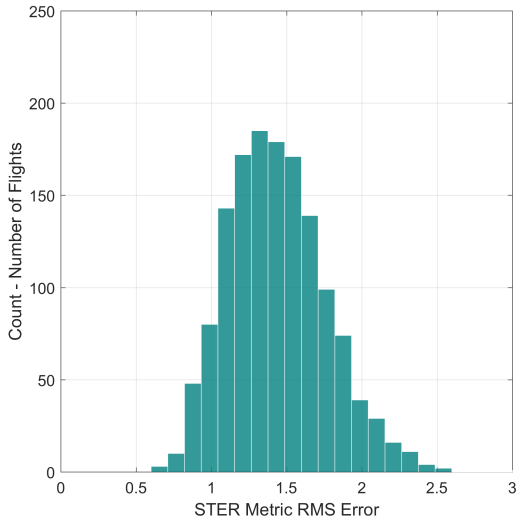


Figure 78: Distribution of metric RMS error for the entire set of flights using chosen calibrated model (Piper Archer)

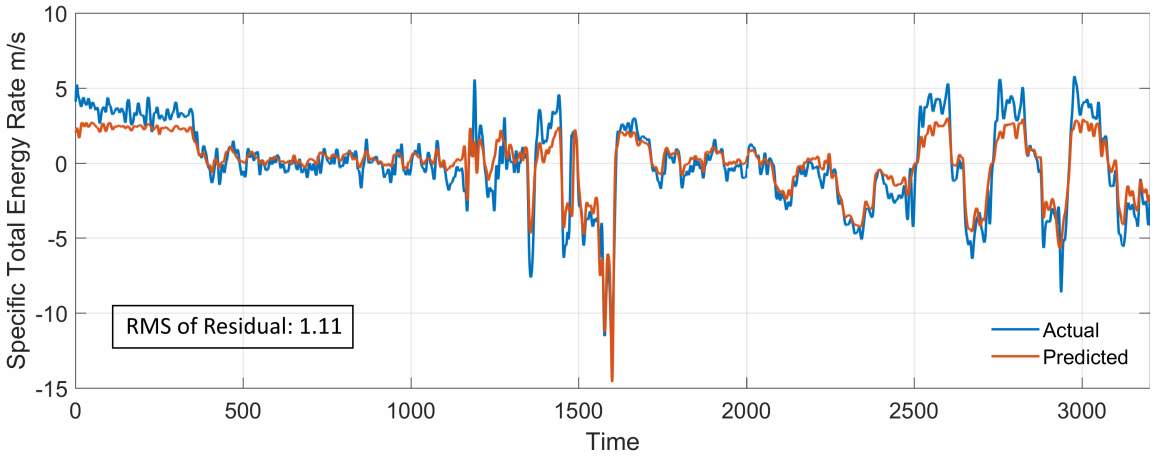


Figure 79: Data traces of STER metric for sample flight record 1 (Piper Archer)

as well as error in the traces for the Piper data, one of the reasons for this is that the actual take-off weight of the aircraft is known for the Cessna flight but is unknown for all the Piper flight data records. Additionally, the epistemic uncertainty due to various sources plays a part in the aircraft-to-aircraft variation. The data traces shown in this section are from representative flights that have different levels RMS error of the residual.

The first flight data record has an overall RMS of residual of 1.11 and is shown

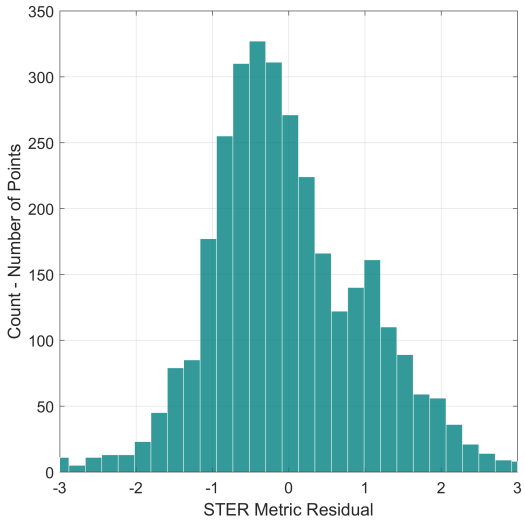


Figure 80: Distribution of STER metric residual for sample flight record 1 (Piper Archer)

in Figure 79. The distribution of the metric residual (actual minus predicted) is visualized in Figure 80. As is evident from the histogram, the calibrated model while predicting close to the actual flight data record tends to under-predict in some regions while it over-predicts in other regions of the flight.

The data traces for a second flight data record with RMS of residual of 0.88 are shown in Figure 81. Along with low residual RMS error for this flight, the distribution of residual shown in Figure 82 indicates that for this flight the calibrated model also under and over-predicts the metric slightly in different regions. However, the RMS as well as the overall residuals are low for both the flight records shown here.

Similar to the Cessna energy metric traces, the level-1 calibrated model is able to capture the overall variation of the metric well throughout the flight for different flights shown here. The data traces of the metric for the two sample flights show varying levels of errors that can be attributed to different types of uncertainty. Overall it is noted that the level-1 calibrated model for the Piper Archer tends to over-predict when the STER metric is close to zero, while it tends to under-predict when the metric's absolute value is higher.

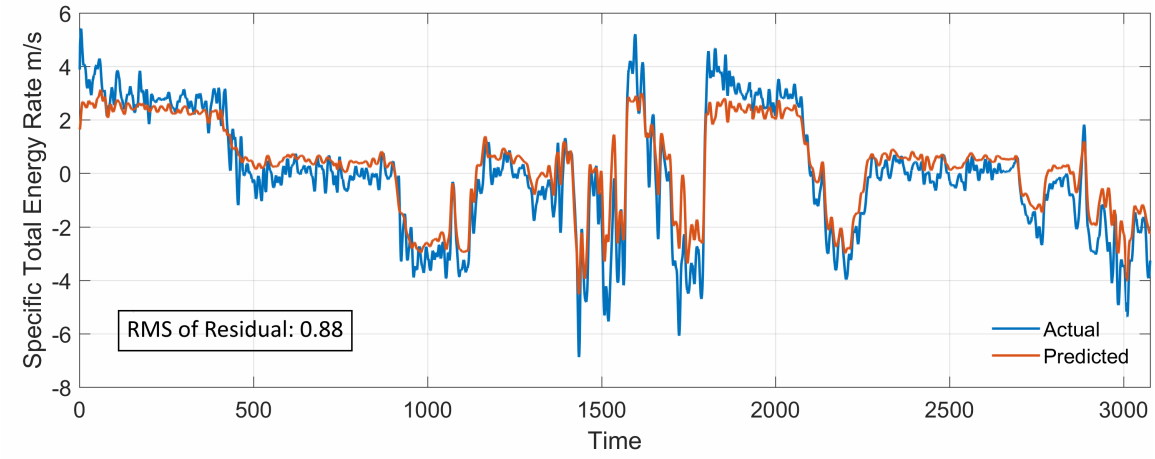


Figure 81: Data traces of STER metric for sample flight record 2 (Piper Archer)

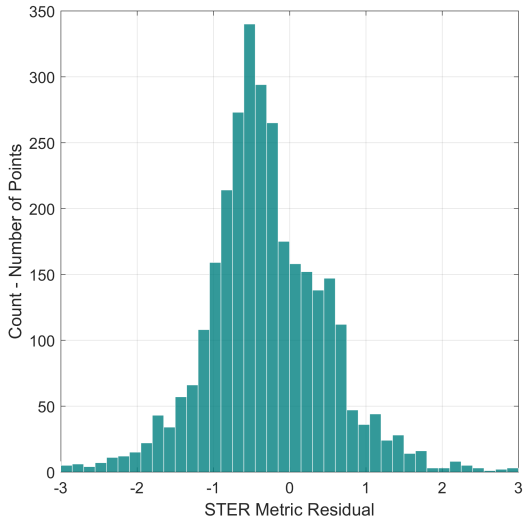


Figure 82: Distribution of STER metric residual for sample flight record 2

6.4.5 Summary

Through this experiment, it is demonstrated that performance data available in the public domain, such as those in a Pilot Operating Handbook can be used to calibrate empirical performance models of a generic GA aircraft to a specific aircraft. This calibration is able to improve accuracy of predictions of unrecorded quantities and provides some predictive capabilities for retrospective safety analysis in GA. The results indicated in this experiment thus address research question 3.1 corresponding to level-1 calibration. The main advantage of using level-1 calibrated models is that

when additional flight data is available that is not as rich in the set of recorded parameters, the calibrated models can be used to provide predictions for energy metrics of interest to be used in retrospective safety analysis. Similarly, in the event that flight data records cannot be shared between entities for various reasons, calibrated models that are validated on such flight records can be transferred between different users for retrospective analysis.

6.5 Calibration to Flight Data – Research Question 3.2

Though the POH calibration results in improved predictions for the energy metric of interest, certain discrepancies are still observed. One of the main reasons for this is that, in the POH calibration, only a limited number of conditions are available for calibration. This does not allow the calibration of all calibration factors but only those for which the POH provides directly relevant performance data. Most notably, POH performance data pertaining to aerodynamically dirty configurations is sparse, and thus calibration factors associated with the aerodynamic characteristics for such configurations cannot be tuned in the calibration process.

Despite calibration to POH, models are expected to suffer from uncertainties. For example - Longmuir et al. [95] have demonstrated that something as simple as unwashed exteriors on commercial aircraft can increase drag up to 1%. GA aircraft are known to have much more deformities, non-flush sitting doors, scratches, etc. that can cause significant increases in drag. Similarly, higher surface roughness can lead to degraded aerodynamic performance. Therefore, the second level of the model calibration framework seeks to improve model predictions using limited flight data. The research sub–question motivated for this is reproduced below:

Research Question 3.2:

How can basic empirical models of aerodynamic and propulsive performance of a generic GA aircraft be calibrated using limited GA flight data to predict energy metrics at conditions of interest for a specific aircraft?

1. Benchmark Data

The benchmark data for level-2 calibration is a limited number of flights with certain key additional parameters recorded. Revisiting the categories of flight data introduced earlier in the dissertation, this refers to category 2 data which is considered higher end of the GA spectrum. This data could be thought of as obtained from a G1000-type recorder. An important piece of information assumed to be available is the position of the flaps at each point during the flight. This type of flight data is not required for all flights on which the models will be deployed but only for a handful of calibration flights. The information about flaps can recorded by other means such as by manually recording the deployment flaps. It should be noted that this is only required for a limited number of flights intended to aid calibration, and allows the development of the novel method presented in this section. The remaining parameters recorded during the calibration flights are assumed to be the same as previously outlined.

2. Discrepancy Metric

The availability of data from actual GA flights presents an interesting question about the discrepancy metric that can be used for calibration. It is noted that even with the highest level of data recording capability typically available on GA aircraft, the main outputs of the performance models – thrust, drag, lift, etc. cannot be directly

measured or recorded. However, as noted in earlier sections, there are certain energy-based metrics which can be estimated using performance models and flight data, among them, the specific total energy rate (STER).

$$STER = \underbrace{\frac{(T-D)\,V}{W}}_{\text{From Performance Models}} = \underbrace{\dot{h} + \frac{V \times \dot{V}}{g}}_{\text{From Flight Data}} \tag{31}$$

Unless the weight of the calibration flight is explicitly recorded, there might still be uncertainty in the weight (W) on the left side of Eq. 31. For cases when it is not explicitly recorded, the methodology will assume a reasonable estimate for the weight that is explained later. The quantification of this uncertainty is demonstrated in Appendix C. These observations motivate the following hypothesis:

Hypothesis 3.2: *Using specific total energy rate metric from flight data to calibrate the performance models results in a process that yields models for a specific aircraft which can be used for predicting energy metrics of interest.*

A brief explanation of the discrepancy metric and the type of calibration is provided here. As seen from Figure 83, the metric can be evaluated using the combined outputs of the aerodynamics and propulsion models. Similarly, the rates of change of altitude and velocity data can also be used to evaluate the metric directly from recorded data. This method of calibration is different than what is done in commercial operations because of the constraints imposed by GA flight data. As noted in the figure, it introduces the possibility of additional uncertainty due to the potential propagation of uncertainty in the original models and/or data. However, it also enables bringing in additional predictive capabilities into the analysis with the limited data available which can prove to be valuable.

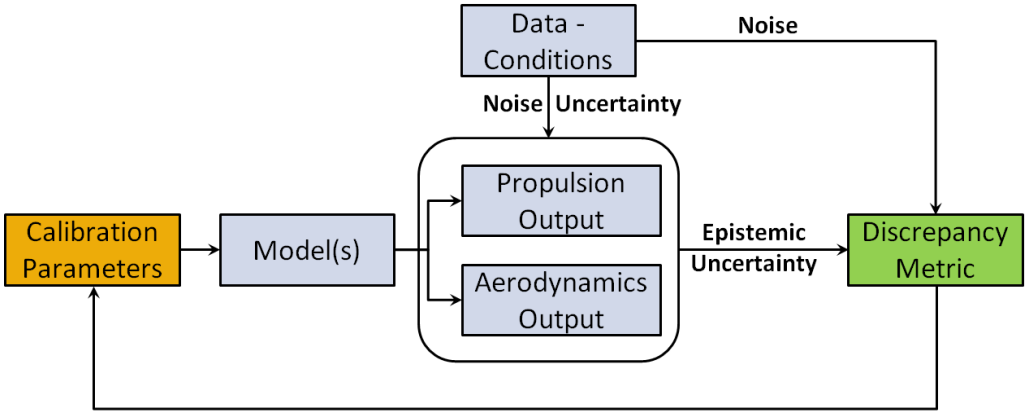


Figure 83: Schematic of indirect calibration setup for level-2 calibration

3. Calibration Factors

Unlike level-1 calibration using POH data, where the calibration factors are restricted to those for which test conditions exist in the manual, level-2 calibration can use the entire set of calibration factors assuming the flaps have been used at least once in the calibration flights. Therefore, all thirty calibration factors can be varied within their range in the level-2 calibration. While this increases the dimensionality of the space being explored, it enables sweeping a larger possible spread of performance curves (details presented in Appendix F).

4. Optimization

Unlike level-1 calibration, the level-2 calibration setup uses only a single discrepancy metric. This metric is calculated at each point in the flight data record(s) available for calibration. Therefore the root mean square of this discrepancy over the entire flight is calculated and used. In case multiple flights are available for calibration, the RMS over all the flights can be used. Since this is an indirect calibration process, an evolutionary optimization algorithm is chosen. Additionally, as the calibration flight will only have limited flight conditions at which to test the discrepancy, there may be multiple local minima. Therefore, a genetic algorithm is used as the optimization algorithm of choice to ensure that the best possible calibration model is obtained.

In terms of implementation details, the initial starting point of the algorithm can be chosen randomly within the ranges of calibration factors or using information from prior efforts. In particular, the pareto-optimal set of calibration vectors obtained from level-1 calibration can be used to 'warm start' optimization in level-2 calibration. This could help the algorithm converge faster and provide better results. The implementation of GA in MATLAB is utilized for this purpose.

6.6 Experiment 3.2

The steps outlined in the previous section provide details on the methodology used for flight data calibration. The experiment outlined in this section enables answering research question 3.2 and demonstrating calibration of empirical models using limited annotated flight data.

6.6.1 Purpose of experiment

The following are identified as the main objectives of this numerical experiment:

1. Demonstrate level-2 calibration using the model calibration framework for two common GA aircraft (Cessna 172 and Piper Archer)

2. Test calibrated models using available flaps-annotated flight data and the specific total energy rate metric

6.6.2 Experiment Setup

Unlike the level-1 calibration, this setup does not have multiple discrepancy metrics and therefore, a single-objective genetic algorithm is used for calibration. The specific total energy rate metric is evaluated over the entire flight data record and the RMS error is calculated. Typically, it is expensive to fly dedicated flights for use in calibration and as such not many flight data records are expected for calibration purposes. On the other hand, data from routine operations which need not be at

the level of parameter-recording sophistication of the calibration flights may be more easily available. It is noted that the information about the take-off weight of the aircraft is only available for the Cessna 172 flight data and the uncertainty that may be introduced due to unknown weight is addressed in Appendix C. Similar to level-1 calibration, this experiment is set up in MATLAB and exploits its native capabilities in terms of optimization algorithms and visualization. All the calibration factors are varied in this level as data is available corresponding to all flapped aerodynamic configurations.

One of the advantages of level-1 (POH) calibration is that it does not require any flight data records. The resulting calibrated aircraft obtained from this step are shown to have good predictive capability. Therefore, for level-2 calibration it is useful to start the optimization from the POH-calibrated model chosen in the previous step. This can accelerate the convergence of the calibration and produce models that have improved performance over the models from level-1 calibration. Starting from a level-1 calibrated model that performs well on the flight data record available ensures that the predictive capability of the model is already good. Therefore, the changes in calibration factors necessary for optimization are less drastic than when starting from any random point in the feasible space. An important difference noted previously is that the quantities being predicted by each model are not individually recorded in the flight data. Rather an energy metric that uses the difference between predictions from the two models is used. Therefore, simultaneously varying both models that are already providing good predictions might cause the optimization to move away from the minimum. Therefore, in level-2 calibration approach starting from level-1 optimum, the aerodynamics and propulsion models are varied one at a time and the optimization is performed in an iterative manner till it converges. The calibration factors are appropriately separated into aerodynamics and propulsion specific factors and inactive factors are held to their level-1 optima or optima from the previous

iteration. This does not apply if level-2 calibration does not start from the level-1 optima, in which case level-2 optimization starts from the baseline model.

6.6.3 Experimental Metrics

For basic validation, similar to level-1 experiment, some flaps-annotated flight data is used to evaluate the error in the same energy metric. For the Cessna model, this is the same flight used for calibration, whereas for the Piper model(s) it would be all other flights available that are not used in calibration. These restrictions arise due to the availability of data and do not affect the manner in which the methodology is deployed. The RMS of the residual of the metric is used as the optimization objective in level-2 calibration.

The level of accuracy for the model calibration is typically dependent on the intended application and quality of available data [79]. The primary use of the models calibrated in this work is for anomaly detection using energy metrics and within the flight simulation model for generating simulated flight data. Therefore, the overall residuals that are acceptable for calibrated models should be lower than the values of the metric that separate nominal and anomalous records. A few examples of the variation of the calibration energy metric for anomalous flights are demonstrated in order to understand the applicability of the level-2 calibrated models for anomaly detection.

6.6.4 Results

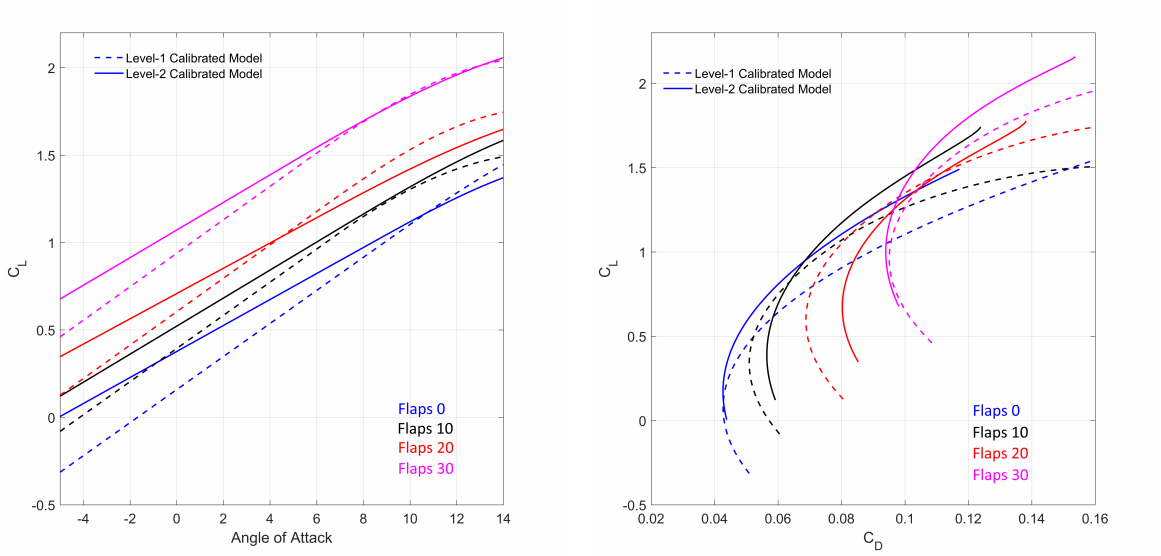
The results from level-2 calibration of performance models using a single Cessna 172 flight data record with known weight and a number of Piper Archer flight data records with unknown weight is demonstrated in this section.

6.6.4.1 Cessna 172 Results

For the Cessna 172 aircraft, the single flight data record available with known take-off weight is used to perform level-2 calibration. The resulting model obtained after convergence of the optimization is shown in Figure 84. Since there is no choice in the data used for calibration, the algorithm outlined in the previous section is applied using the available data. The model obtained from the convergence of the calibration is shown in this section along with its predictions on the same flight data record.

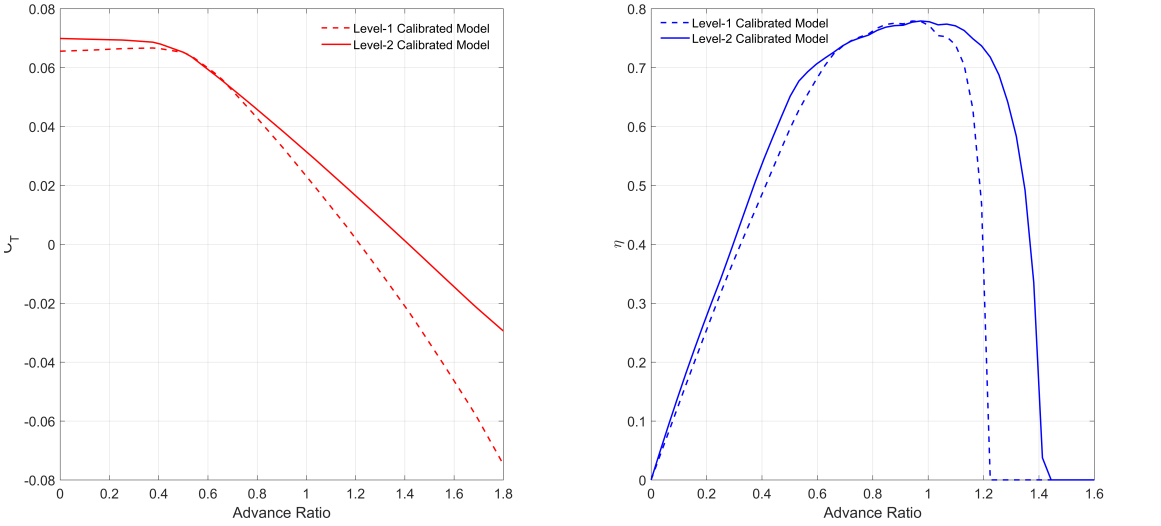
As seen from the figures, the changes in aerodynamics models indicate that the level-2 calibrated model has higher lift at the same angle of attack for all configurations and slightly higher drag for the same lift coefficients for all configurations. The thrust curve for regions below advance ratios of unity is similar, but it is slightly higher for higher advance ratios. A similar trend to the thrust curves is observed in the propeller efficiency curve.

The visualization of the data trace of the STER metric can be seen in Figure 85. The RMS of the residual drops from 0.85 for level-1 model to 0.81 for level-2 calibrated model. The level-2 calibration process corrects for the places in the flight record where the earlier model performs poorly. It is also of interest to visualize the distribution of the actual residual of the metric over the entire flight. Figure 86 contains this residual and it is observed that the residual is symmetrically distributed about zero with the mean approximately at zero. Therefore, the calibrated model is neither over-predicting nor under-predicting and therefore does not display any systematic bias. It is of interest to look at the difference between the residual of the metric at each point in the flight to be able to visually confirm and also quantify the improvement from level-1 to level-2 calibration. Therefore, the difference between the absolute value of the residuals from level-1 and level-2 calibrated models is visualized in Figure 87. It is evident from the skew of the distribution towards the left of zero that the residuals from the level-2 calibration are in general lower than those from



(a) Aerodynamics - Lift

(b) Aerodynamics - Drag



(c) Propeller - Thrust

(d) Propeller - Efficiency

Figure 84: Changes in discipline models between level-1 and level-2 calibration (Cessna 172)

level-1 calibration. Therefore, it is evident that the calibration framework is able to produce a model capable of predicting the energy metric of interest more accurately after level-2 calibration.

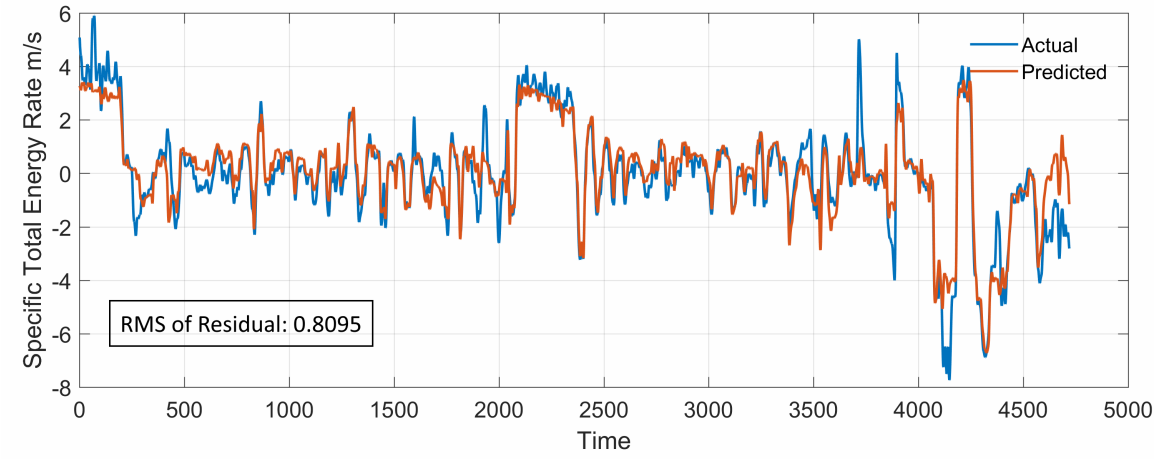


Figure 85: Variation of STER metric – actual and predicted (Cessna 172)

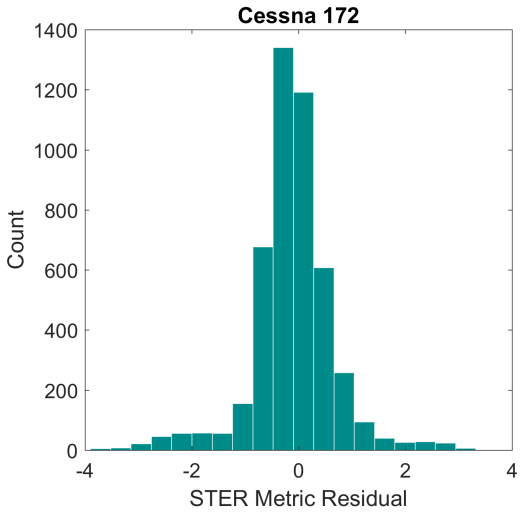


Figure 86: Distribution of STER metric residual for calibration flight (Cessna 172)

6.6.4.2 Piper Archer Results

Level-2 calibration is also undertaken for the Piper Archer aircraft using flight data records available. It should be noted that the flight data records for the Piper Archer do not have take-off weight recorded. Therefore, there is additional epistemic uncertainty for this aircraft. However, using the fuel flow rate data that is recorded and intelligent estimates of take-off weight, level-2 calibration can still be carried out. The uncertainty introduced due to unknown weight is quantified and its effect on the calibration metrics is shown in Appendix C. For the current implementation, there are

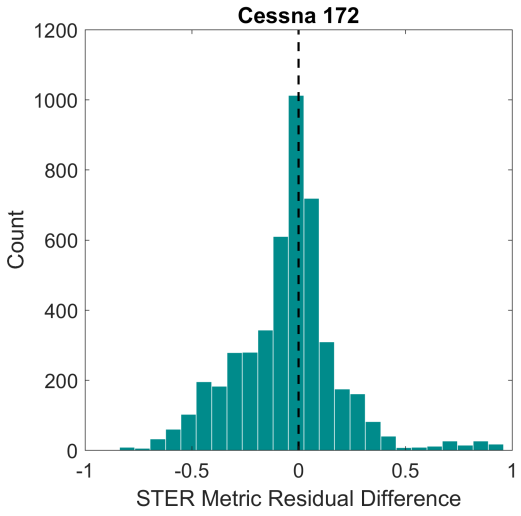


Figure 87: Difference between absolute value of STER metric residual between level-2 and level-1 calibrated models

multiple flight data records available which can potentially be used for calibration.

The requirements for calibration flight data records stem from the nature of the calibration metric and the different curves that are being modified during calibration. It is noted that while thrust and drag are estimated by the models, the weight needs

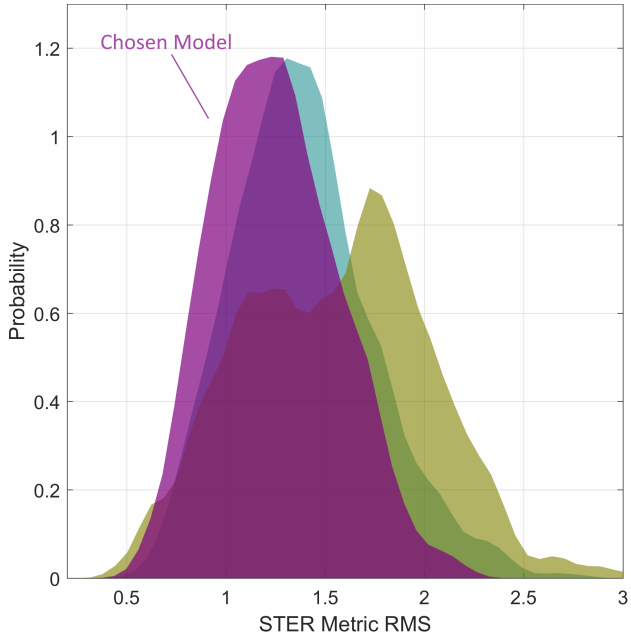


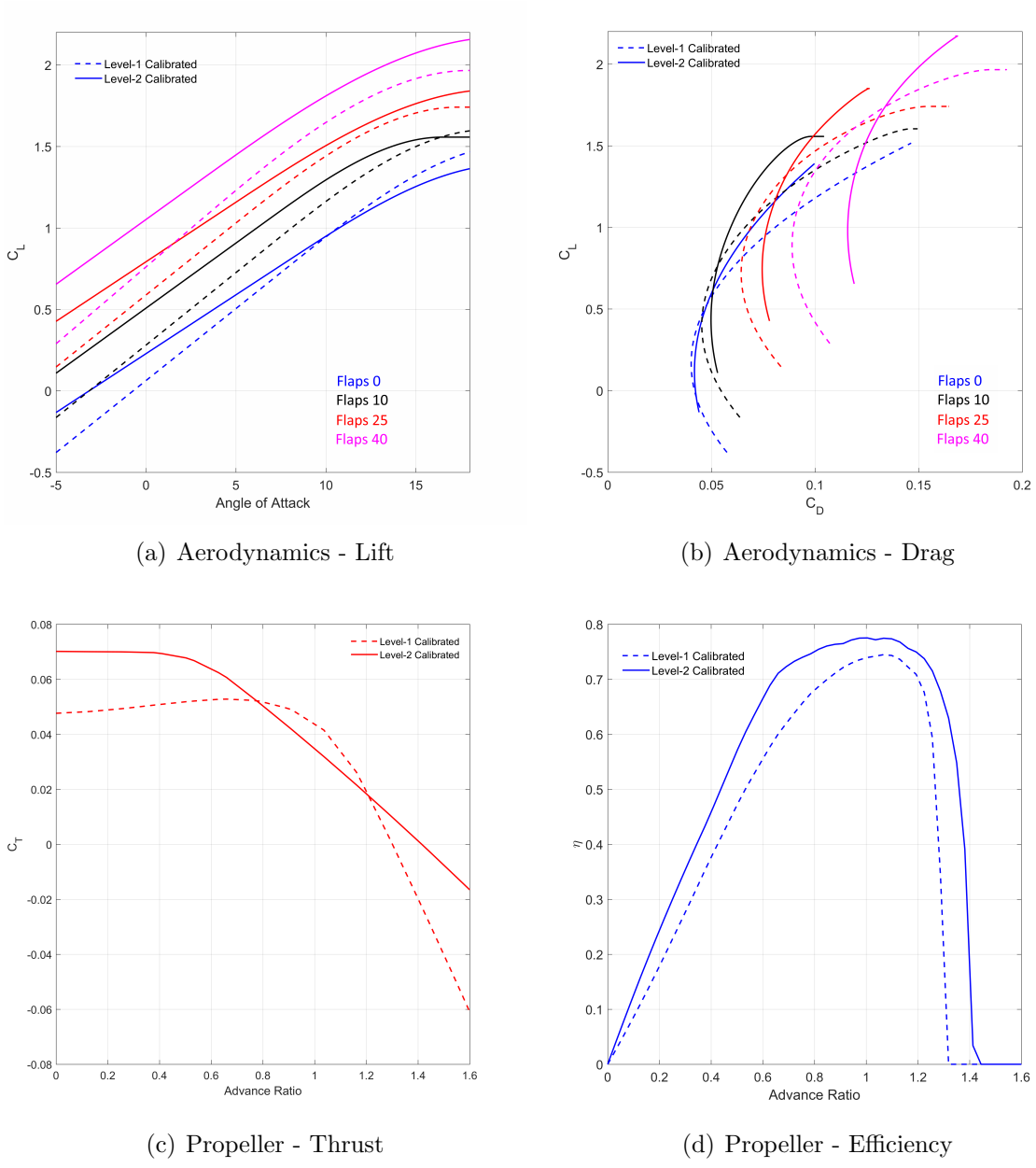
Figure 88: Distribution of STER metric RMS for all flights in the validation set using calibrated models from three different calibration flights

to be recorded. Therefore, ideally, data records for which the take-off weight (TOW) is known should be used. As TOW is not available for the data set being used an estimate based on the POH data and typical occupancy can be made. Secondly, as there are four different aerodynamic curves, one for each flapped configuration, the calibration flight should have points that correspond to all levels of flap usage. This enables exercising all the calibration factors available to the optimization. Thirdly, multiple cycles of flap usage from different ambient conditions within the calibration flight enables choosing calibration factors that are robust to variations not explicitly captured in the modeling process. Finally, data records that have less noisy flight data are preferred over those that have a high amount of noise in the sensor recordings. While the ideal situation is flying a dedicated flight profile containing multiple flap cycles with known take-off weight, this may not always be possible or feasible and therefore, the methodology should be able to work at different levels of data availability.

In the case of the current work, multiple flights are available but all with unknown TOW. The handful of flight data records that satisfy most of the requirements for calibration flights from the previous paragraph are short-listed and the calibration is carried out using each flight individually. Due to the differences from one aircraft to another and other uncertainties, the RMS error of the metric over all the flights in the validation data set varies if calibration is performed using different flights. As an illustration, Figure 88 shows the distribution of RMS errors for the calibrated models using three different flights. As is evident from the figure, using different calibration flights can provide varying levels of RMS error on the validation set. Therefore, the calibrated model with the lowest average RMS error is chosen. The spread in the RMS error values for the validation set is due to the fact that even though the validation set contains data from Piper Archer aircraft, there are more than twenty unique tail numbers which introduces variability. When information regarding the tail number

of a flight data record is available, the calibration can be further fine-tuned based on each tail number. This is demonstrated later in Appendix H.

Similar to earlier results, a visualization of the changes in the actual models is shown in Figure 89. A general trend similar to Cessna level-2 models is observed in that the lift curves have a higher intercept and shallower slope, drag polars have



(a) Aerodynamics - Lift

(b) Aerodynamics - Drag

(c) Propeller - Thrust

(d) Propeller - Efficiency

Figure 89: Changes in discipline models between level-1 and level-2 calibration for Piper Archer

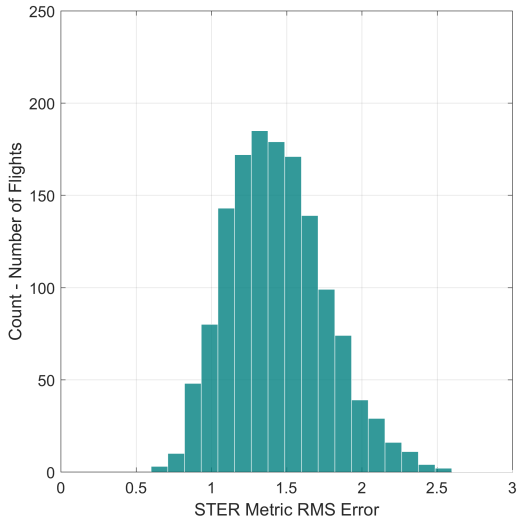


Figure 90: Distribution of metric RMS error for the entire set of flights using chosen level-2 calibrated model (Piper Archer)

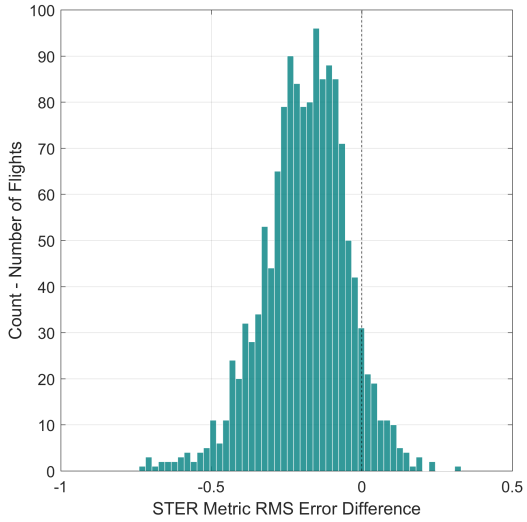


Figure 91: Difference between absolute value of STER metric residual between level-2 and level-1 calibrated models (Piper Archer)

increased drag (especially for flapped configurations), and thrust and efficiency are slightly higher than the level-1 calibrated models (especially at higher advance ratios).

The actual distribution of RMS error for the metric is visualized in Figure 90 along with the difference between level-1 and level-2 calibrated models in Figure 91. As is evident from Figure 91, there is a noticeable improvement from level-1 to level-2 calibrated model performance over the entire validation data set. The majority of the

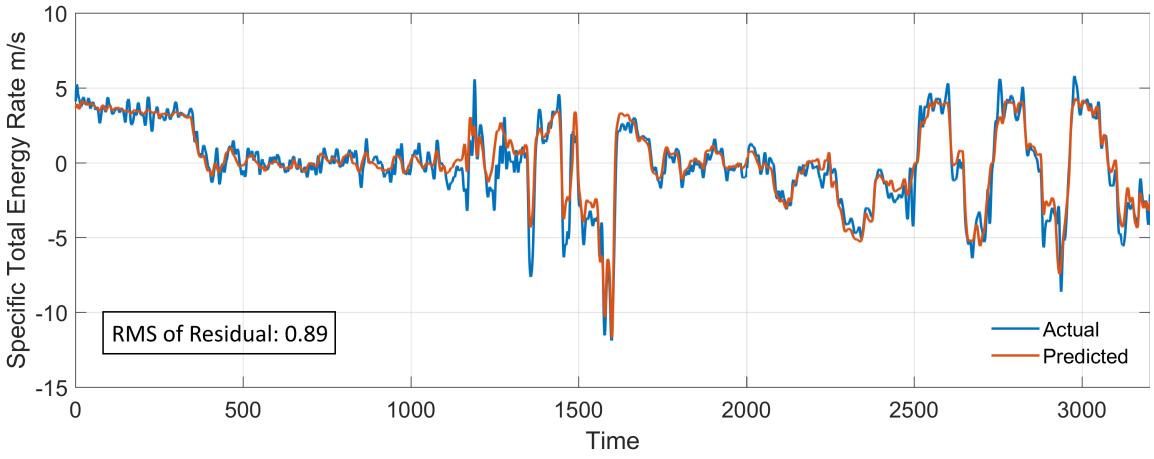


Figure 92: Data traces of STER metric for sample flight record 1 (Piper Archer)

distribution from Figure 91 being to the left of zero indicates that for almost all the flights in the validation set, the level-2 calibrated model provided improved average performance.

Figures 92 and 93 show the data traces and residual distributions for the same sample two flights that were visualized for level-1 calibration of Piper Archer aircraft earlier. For the first flight data record, the RMS error reduces from 1.11 to 0.89. The level-2 calibrated model improves predictions, especially in regions where the metric values are higher. This was the region where level-1 calibrated models were found to be deficient in the earlier experiment and research question. The distribution of the metric residual shown in Figure 93 indicates a more symmetric distribution of the residual throughout the flight with a slight tendency of the model to over-predict.

Figure 94 shows the data traces for the second sample flight data record from the validation data set. For this data record, the RMS residual dropped from 0.88 for level-1 model to 0.75 for level-2 calibrated model. Figure 95 shows the distribution of the residual for this flight data record.

Figure 96 shows the distribution of the residual for all points from all flight data records in the validation set. This figure shows good overall agreement with the

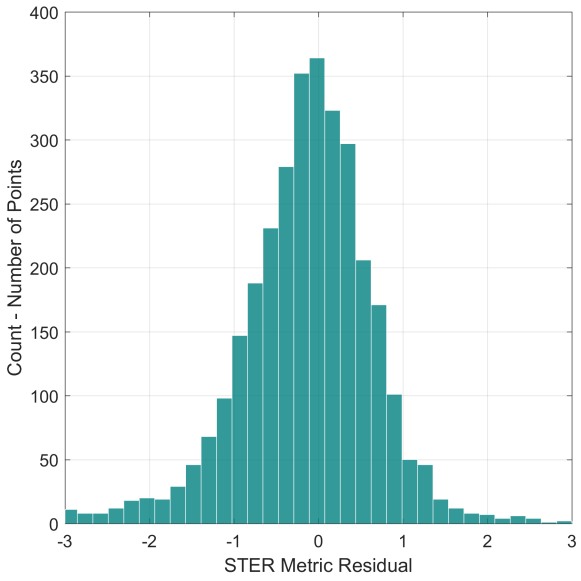


Figure 93: Distribution of STER metric residual for sample flight record 1 (Piper Archer)

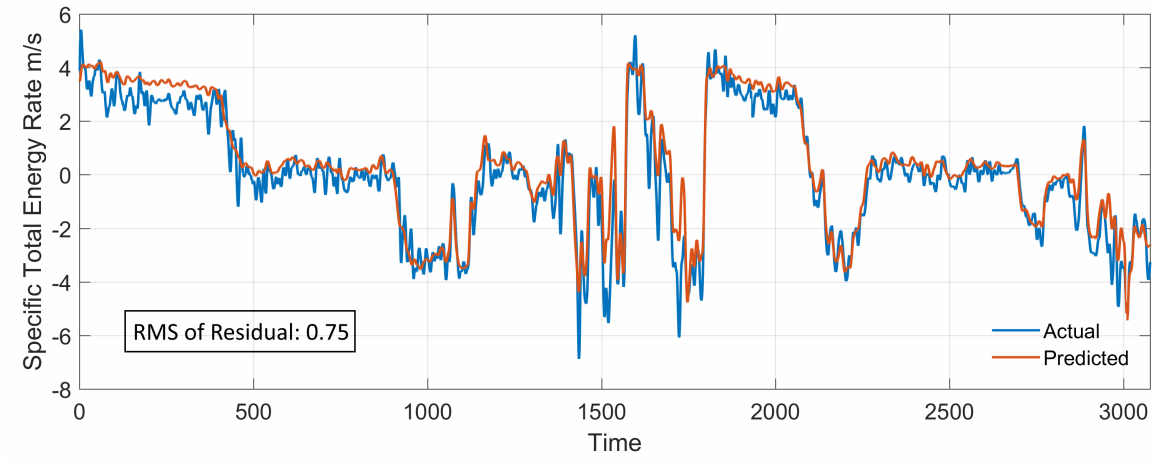


Figure 94: Data traces of STER metric for sample flight record 2 (Piper Archer)

actual trace using the level-2 calibrated model, with a slight tendency towards over-prediction.

6.6.5 Summary

Through this experiment it is demonstrated that using limited flaps-annotated flight data for calibration of performance models using the level-2 framework improves accuracy of predictions of energy metrics of interest. This improvement is demonstrated

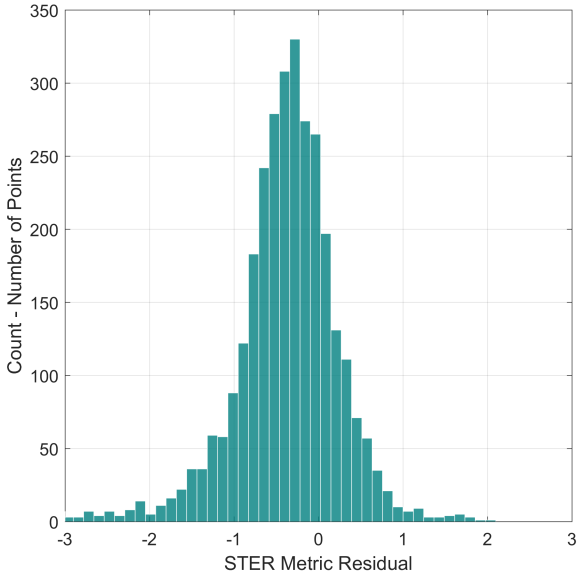


Figure 95: Distribution of STER metric residual for sample flight record 2

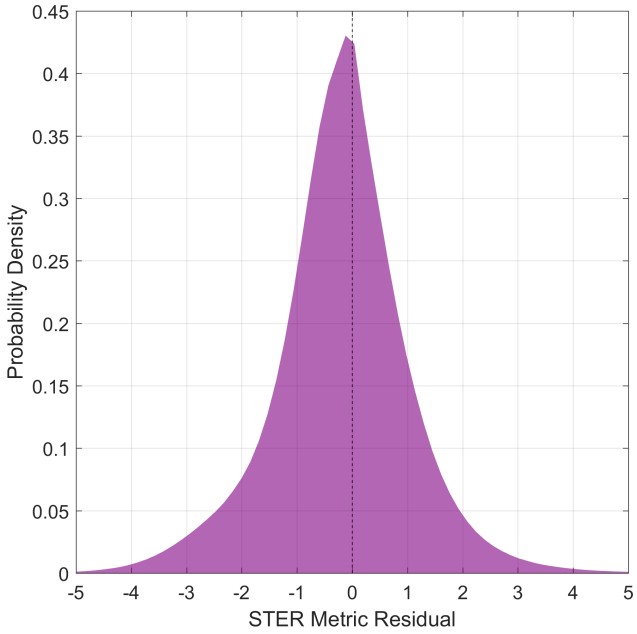


Figure 96: Distribution of STER metric residuals over all points in the validation data set

for both aircraft types considered. Testing the calibrated models using available validation data indicates that calibration yields models that can be used for predicting energy metrics of interest despite the uncertainty present in the analysis due to unknown take-off weight of the aircraft. The results of the experiment thus address the

research question 3.2. The intended use of calibrated models in this dissertation is for two different purposes. The first one is to enable the evaluation of some energy metrics used in the anomaly detection task presented previously in the dissertation. In order to examine the usability of calibrated models for evaluation of energy metrics and anomaly detection, anomalies are separately identified for a particular feature vector option using STER metric evaluated from the data side and the model side. The anomaly score in each case is obtained and the anomaly-status of each flight is compared in the two cases. The results are shown in Figure 97.

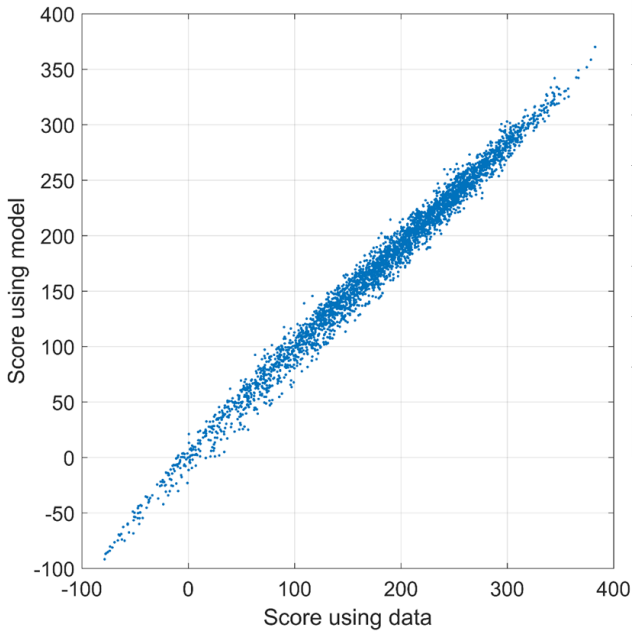


Figure 97: Anomaly score obtained using metric evaluated from data-side and model-side

As is evident in the figure, there is a high positive correlation between the two scores indicating that the calibrated models are able to provide sufficient level of accuracy for energy metric prediction and anomaly detection. The most important observation from this experiment is that the anomaly status of all flights remains the same whether the model is used or the data is used to evaluate the metric.

The second use is in the flight simulation model which can subsequently be used to generate anomalous/unsafe simulated flight data records. For the second use, while

the calibrated models should be for an aircraft similar to the ones under consideration, it need not match existing flight data records exactly as the simulated flights can be considered to be from another aircraft. Since one of the main pillars of the dissertation is use of multiple types of aircraft, this would still fit within the framework. For the first usage of models in energy metrics estimation, it is desired that detection of anomalies should not be affected by the uncertainty in performance models. To that end, the deviations of metric values in anomaly detection for abnormal flights should be larger than the residual in calibration errors. As is observed from the residuals and RMS distribution of errors, the deviations of STER metric in both types of anomalies are typically higher than the maximum values of the residuals and therefore, the calibrated models can be used to predict metrics in anomaly detection.

CHAPTER VII

CASE STUDY

The individual components of the methodology, their corresponding research questions, and the experiments conducted to test various hypotheses were presented earlier in Chapters 4, 5, and 6. In this chapter, the methodology is deployed on a real-world data set consisting of flight records from two distinct aircraft operating at different airports. The insights obtained, comparison to existing benchmarks, and advantages and limitations of the developed methodology are presented. A high-level road-map for implementing the different components of the methodology is shown in Figure 98. Also, from the perspective of any scientific effort, the ability to replicate the results in another setting is important. The different components of the developed method are disparate and a set of guidelines for the order and manner in which to implement them is essential. Therefore, a step-by-step process of the order of implementation of the components of the methodology is presented. This is meant to provide guidelines for repeating the efforts presented here in a different setting or improving the developed techniques. Due to the data-driven nature of each step as well as the requirements for multiple pieces of information in each step, an accompanying example is not presented. However, for each step, the appropriate section of the thesis is referenced to provide context and implementation details.

The first step of the methodology is data collection. The data required and used in this methodology is of different forms. Flight data collected from routine operations using different data recorders is required for analysis. The information contained within the POH for each aircraft in the flight data set is also required. Dedicated calibration flights with the required parameters recorded, if available, are also collected.

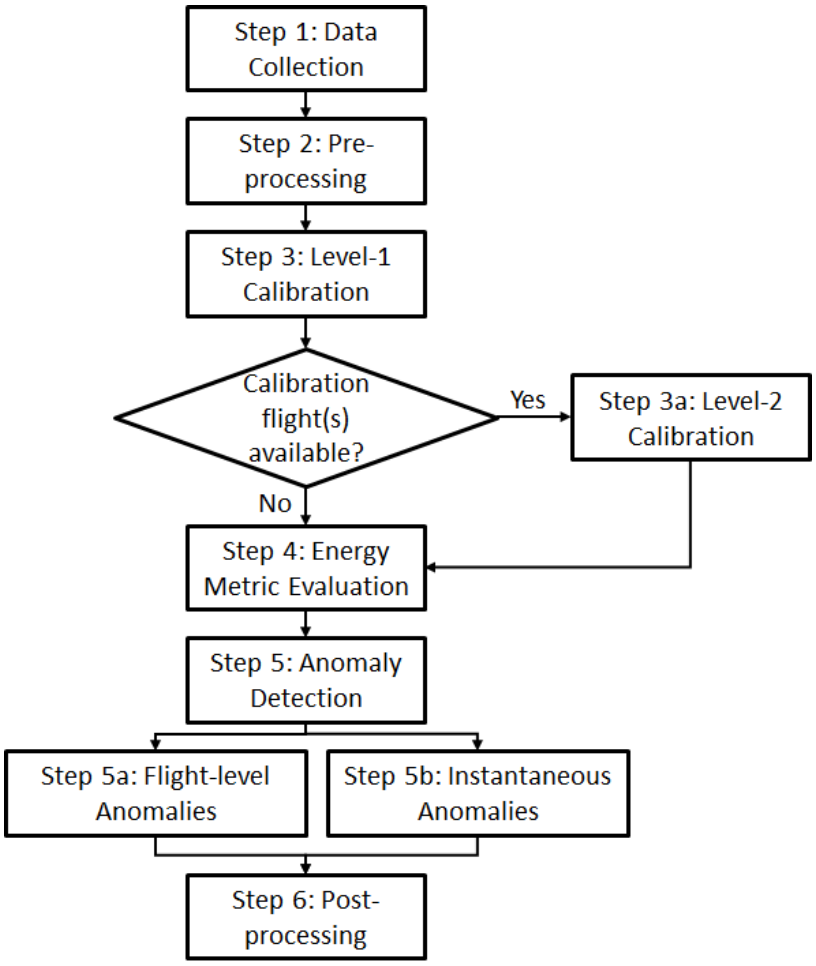


Figure 98: Important steps of the methodology

Following the collection of data, the next step is basic pre-processing operations to filter noise, obtain touchdown point [118], extract phases of flight [68], estimate weight, etc. Once these operations have been performed successfully performed, the third step is level-1 calibration using only publicly available data (from the POH). Sec. 6.3 provides the details of the level-1 model calibration framework demonstrated in this thesis. Following level-1 calibration, if dedicated calibration flights are available, level-2 calibration is pursued to improve the predictive capability of the models. Sec 6.5 provides the details of level-2 calibration process demonstrated in this thesis.

After model calibration is complete, the requisite tools necessary for evaluation of

all possible energy metrics are available. Step 4 of the methodology involves evaluating energy metrics. The formulas and details of each metric are provided in Sec. 4.2.2. Statistical averaging and nominal variations of metrics can also be evaluated at this stage using all the flight data records available. Once all the possible metrics are evaluated, anomaly detection can be implemented. Implementation of each type of anomaly detection is independent and can be pursued concurrently. Sec. 5.1 provides details of flight-level anomaly detection whereas Sec. 5.3 provides the details of instantaneous anomaly detection. The final step of the process is post-processing anomalies. In this step, parameter and energy-metric charts can be visualized along with anomaly scores or probabilities. Sec. 7.2 contains details of the different types of post-processing charts used in this thesis and their interpretations. While the steps presented here are high-level, they are meant to provide a road-map for the implementation of the methodology rather than the details.

Table 5 presents a summary of some of the important facts about the data set used in the case study.

Table 5: Summary of the current data set

Total number of flights	3,402
Number of airports of take-off and landing	2
Number of unique aircraft	2
Total number of unique tail numbers in data set	37
Number of flights with flaps annotated	1,903 ($\approx$ 55 %)
Approximate number of parameters in each data record	30

As seen from Table 5, the data contains a heterogeneous set of flight records with respect to the aircraft, airports, landing patterns, etc. The methodology developed in this dissertation is one of the few quantitative safety applications on such a large and varied real-world data set from GA operations. A variety of meaningful insights can be drawn from application of the methodology to this set which would not have been otherwise possible. The following sections provide details of some of the analyses that

are possible because of the techniques and algorithms developed in the quantitative data-driven methodology developed in this work. The first section contains details of application of anomaly detection using energy metrics (research questions 1 and 2) and comparison to existing benchmarks and the next sections presents the application of model calibration (research question 3) on a real-world problem other than those already explored in this work.

7.1 Anomaly Detection Case Study

This section outlines the results and insights gained from application of the developed methodology on a large data-set ($\approx$ 3,400 flights). Various trends observed during comparison of different types of anomalies to existing benchmarks, segmentation of anomalies by types of operations, effect of anomalous approach on touchdown performance, etc. are highlighted. The case study is a demonstration of the ways in which the methods developed in this dissertation can be applied.

7.1.1 Relation between types of anomalies

In this section, the relationship between flight-level and instantaneous anomalies is examined. Some of the pertinent questions that arise during application of anomaly detection include:

1. What is the number of flights with instantaneous anomalies at different probability thresholds?

2. What is the overlap of flights with instantaneous and flight-level anomalies in take-off and approach and landing at different levels of significance?

3. Do flights with instantaneous anomalies have a higher chance of also being identified as flight-level anomalies or vice versa?

In order to answer the first question, the instantaneous anomaly detection algorithm is run at different probability thresholds and the proportion of flights with

instantaneous anomalies is obtained. The results obtained for both phases of flight are visualized in Figure 99. The x-axis on this plot represents the probability percentile threshold. It means the threshold is based on the percentile value of the probability scores of all instants from all flights put together. Therefore, while a 1 % threshold in flight-level anomalies will imply 1 % of all flights are identified as anomalous, the threshold for instantaneous anomalies does not have to be linear in the same way as a single flight can have multiple anomalies.

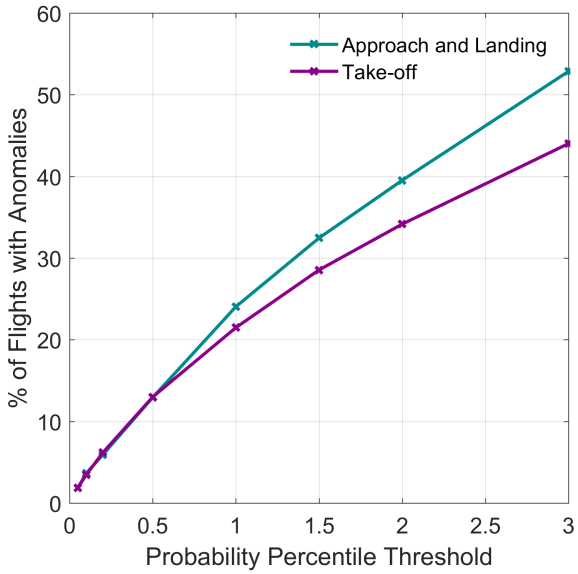


Figure 99: A plot showing proportion of flights with instantaneous anomalies with varying detection threshold

Two important observations can be made from this plot. First, while it is obvious that the proportion of flights with anomalies will increase as the threshold of probability percentile is relaxed (increased), the overall proportion of flights with anomalies even with the highest percentile considered in this work (3 %) the number of flights with anomalies is still just about 50 % of the data set for both phases. This will be contrasted later in the chapter with exceedance detection, where more than 95 % of the data set contains some exceedance or the other. The second observation is with regard to the higher homogeneity of take-off operations noted earlier. The anomaly

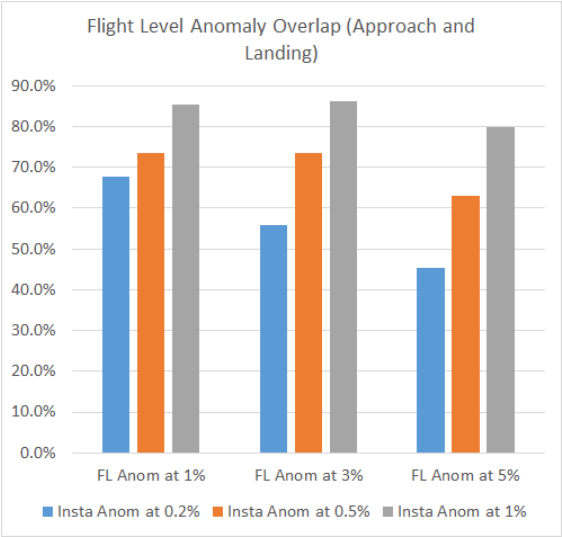
percentage for the curve for take-off instantaneous anomalies is lower than the approach and landing curve indicating that even at relaxed thresholds, more anomalies are being identified in the same flights than in newer flights. Unless otherwise specified, the anomalies considered in this case study are usually between the 0.1 – 0.5 percentile threshold. This corresponds to approximately 1 – 10 % of the flights having instantaneous anomalies. This threshold is however flexible and can be modified as required by the user of the methodology.

The rest of the questions raised earlier in the section can be directed at both the phases of flight under consideration. Therefore, the following sub-sections present the answers for the above questions and others raised during implementation of the methodology for both phases of flight.

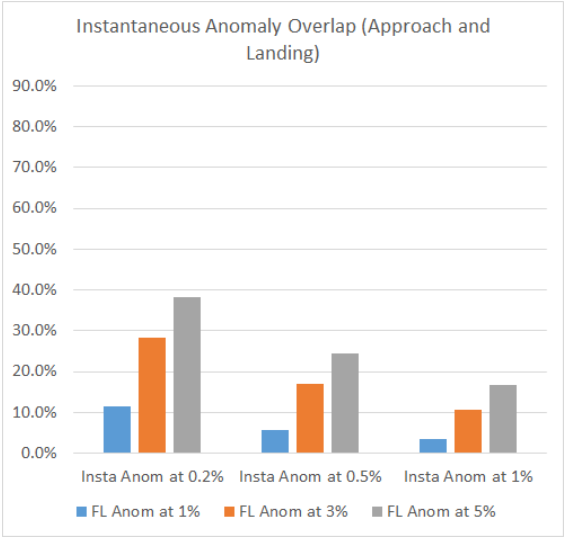
7.1.1.1 Approach and Landing

In the approach and landing phase, flight-level anomalies are first identified at the 1%, 3%, and 5% significance levels. The list of flight data records obtained as anomalous from these experiments are first explored for presence of instantaneous anomalies. Since the threshold for instantaneous anomalies is also flexible, this relationship is explored for multiple threshold levels. For each detection level of the flight-level anomaly detection, the number of flights that also contain instantaneous anomalies at three different detection thresholds are obtained. Figure 100(a) depicts the proportion of flight-level anomalies that also contain instantaneous anomalies. As evident from the figure, a large percentage of flight-level anomalies generally also contain instantaneous anomalies. This overlap increases with increasing instantaneous anomaly threshold for the same flight-level anomaly threshold. On the other hand, Figure 100(b) depicts the reverse relationship, i.e. the proportion of flights with instantaneous anomalies that also contain flight-level anomalies at different thresholds. As is evident from the figure, the trends are similar to the earlier one but the magnitude of overlap is much

smaller. Thus, while the specific flights that contain both type of anomalies are the same number in each case, the instantaneous anomaly set contains many other flights whereas the flight-level anomaly set contains few others.



(a) Proportion of flight-level anomalous flights that also have instantaneous anomalies in approach and landing phase



(b) Proportion of flights with instantaneous anomalies that are also flight-level anomalies in approach and landing phase

Figure 100: Overlap of anomalous flights - approach and landing phase

In order to quantify the correlation between the two types of anomalies, the flight-level anomaly detection score at 3 % significance level and the averaged instantaneous probability score of each flight are plotted in Figure 101. The clear positive correlation indicates what is observed in the earlier figure with regard to flight-level anomalies – that flight-level anomalies on an average have lower instantaneous probability scores which results in most flight-level anomalies also having at least one instantaneous anomalous window.

7.1.1.2 Take-off

Similar to approach and landing, a comparison between anomaly types and scores can be made for the take-off phase. The overall trends observed in the overlap (Figure 102) are similar to those in approach and landing. The main difference is that

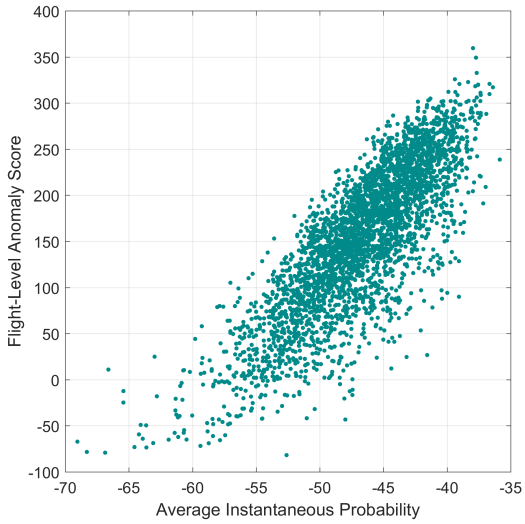


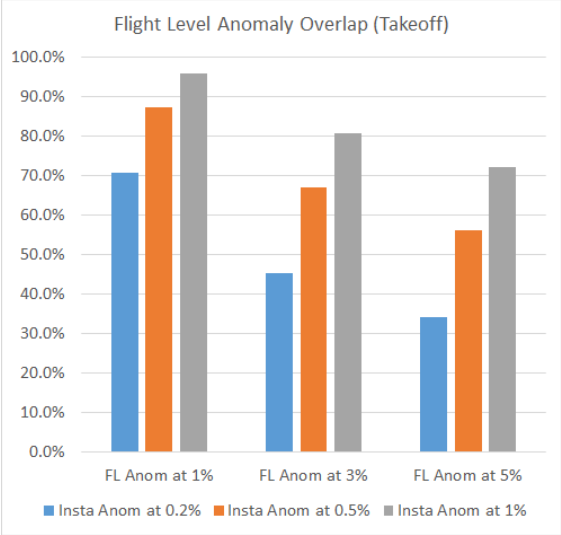
Figure 101: Correlation between average instantaneous anomaly score and flight-level anomaly score in approach and landing

the magnitude of overlap in flight-level anomaly detection drops more rapidly in take-off than in approach and landing. This observation is consistent with others made throughout the dissertation that there seem to be fewer deviations among take-off phase and therefore, at higher percentage thresholds, the algorithm might start identifying flights as anomalous simply because it is forced to select a certain percentage of flights.

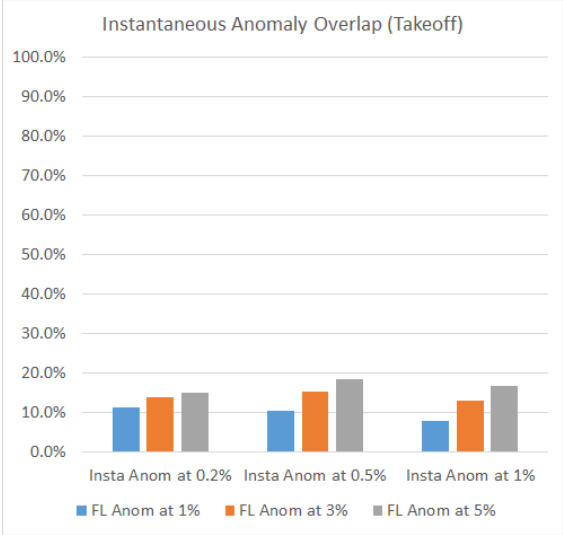
Figure 103 shows the correlation between the average instantaneous anomaly score for each flight and the flight-level anomaly score at the 3 % detection threshold. A distribution similar to approach and landing phase is observed, albeit with a tighter spread among the scores of normal flights, which points to more uniformity.

7.1.1.3 Summary

Through this section, the relationship between flight-level and instantaneous anomalies for both phases of flight is explored. It is observed that flight-level anomalies are expected to usually also have at least one instantaneous anomalous window whereas the relationship is not as strong in the other direction. This confirms the need to



(a) Proportion of flight-level anomalous flights that also have instantaneous anomalies in take-off phase



(b) Proportion of flights with instantaneous anomalies that are also flight-level anomalies in take-off phase

Figure 102: Overlap of anomalous flights - take-off phase

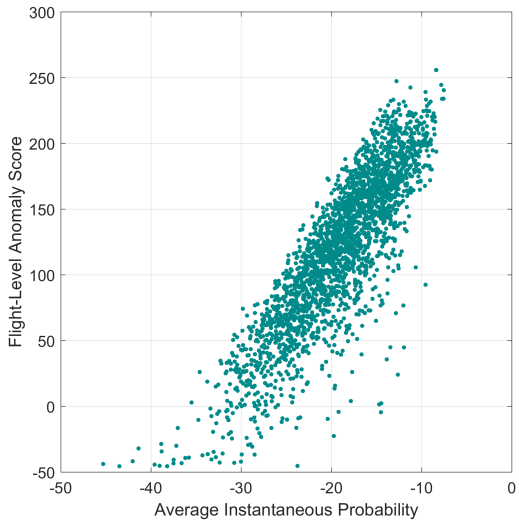


Figure 103: Correlation between average instantaneous anomaly score and flight-level anomaly score in take-off

develop two separate anomaly detection techniques for identifying non-standard operations.

7.1.2 Comparison to Exceedance Detection

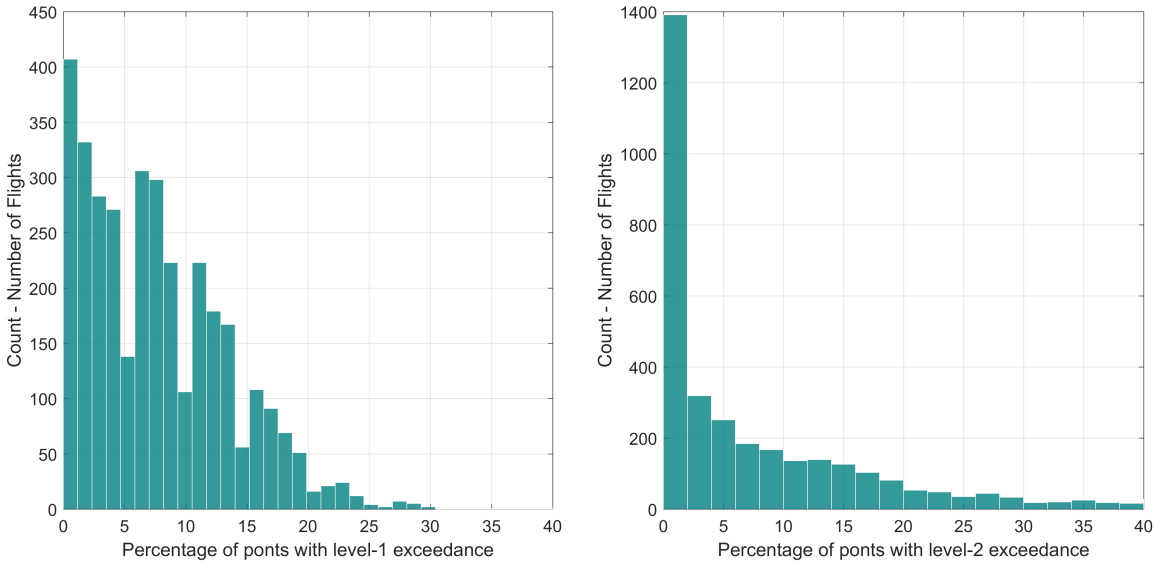
Exceedance detection is identified previously as the current technique most widely used for identifying exceedances and safety events. While this technique has limitations which are also outlined in previous chapters, it is nevertheless of value to compare the flight data records with anomalies against already defined exceedances or events. This can provide insight into the ability of anomaly detection using energy metrics to identify some generally accepted exceedance events. It should be noted that in many cases, the exceedance limits are set arbitrarily or based on gut-feel and are therefore susceptible to error and inaccuracies in identifying conditions that really matter. The main reason exceedances from [73] are used is that it is currently one of the few publicly available set of exceedances. In the present section, comparison with exceedance detection is only provided for approach and landing phase as take-off phase has few documented exceedances as seen from the table in the appendix.

Total types of exceedances considered (From Table 12)	28
Proportion of Flights with at least one level-1 exceedance	92.5 %
Proportion of Flights with at least one level-2 exceedance	76.9 %
Average of number of level-1 exceedances per flight	11.55 (7.65%)
Average of number of level-2 exceedances per flight	16.22 (10.74%)
Median of number of level-1 exceedances per flight	10 (6.62%)
Median of number of level-2 exceedances per flight	6 (3.97)

Table 6: Summary of exceedance events in approach-and-landing phases in the current data set

It is important to understand why exceedance detection is limited in its applicability in addition to the reason than that event definitions might not be transferable between different aircraft. For that purpose, all the flights in the data set are examined for exceedances as defined in Higgins et al. [73]. Each point in the flight data record is examined as a standalone point to identify any level-1 or level-2 exceedance events that are occurring. The summary and some statistics for the total number of events

for the approach and landing phase is presented in Table 6.



(a) Level-1 Exceedance

(b) Level-2 Exceedance

Figure 104: Distribution showing % of points in approach and landing phase with exceedances for each flight

As is evident from the table, a high number of flights from the current data set have at least one level-1 exceedance ($\approx$ 92.5 %) or at least one level-2 exceedance ($\approx$ 76.9 %). This indicates that based on the exceedances defined in Table 12 a high number of flights have at least one exceedance event. Despite the flight data being collected from routine operations and training flights, a high number of flights returning with exceedance events is not ideal. The distribution shown in Figure 104 depicts the distribution of the proportion of points in the approach and landing phase which have exceedances for each flight. Thus, while a high proportion of flights have at least one point with an exceedance, the total proportion of approach and landing points with exceedances is usually not high (on an average around 7.65 % for level-1 and 10.74 % for level-2 exceedances). The number of points with level-1 exceedances on an average are lower than level-2 but the median is higher for level-1 exceedances compared to level-2 exceedances. This implies that in general, flights have more level-1 than level-2 exceedances, but there are certain flights which have a high number of

level-2 exceedances which causes the distribution to be skewed.

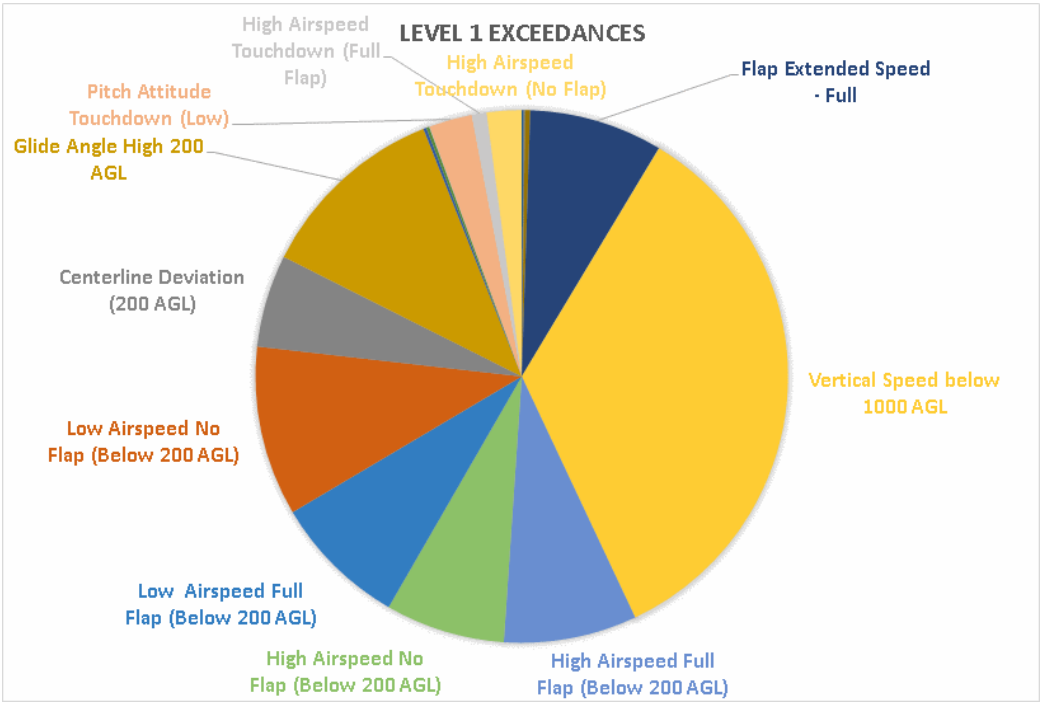


Figure 105: Frequency of different types of level-1 exceedance events in approach and landing

Examining the exceedances more closely, the different type of exceedances and their frequency at level-1 and level-2 is shown in Figures 105 and 106. As is evident from the figures, some types of exceedances occur much more often than others. Exceedances related to vertical speed, high and low airspeed, centerline deviation, gilde angle, and pitch attitude at touchdown are among the most commonly occurring exceedances in the data set. It is noted that the calculation of airspeed exceedances that is also dependent on flap configuration are only evaluated for those flight data records that have flap angle recorded. They are estimable for only two-thirds of the entire data set.

Apart from the high number of flights that are identified as having an exceedance, another issue with this technique is that many of the exceedance definitions require parameters to be recorded that are not usually available in GA flight records (for example: flaps setting). Therefore, for such flights, there is no way to evaluate

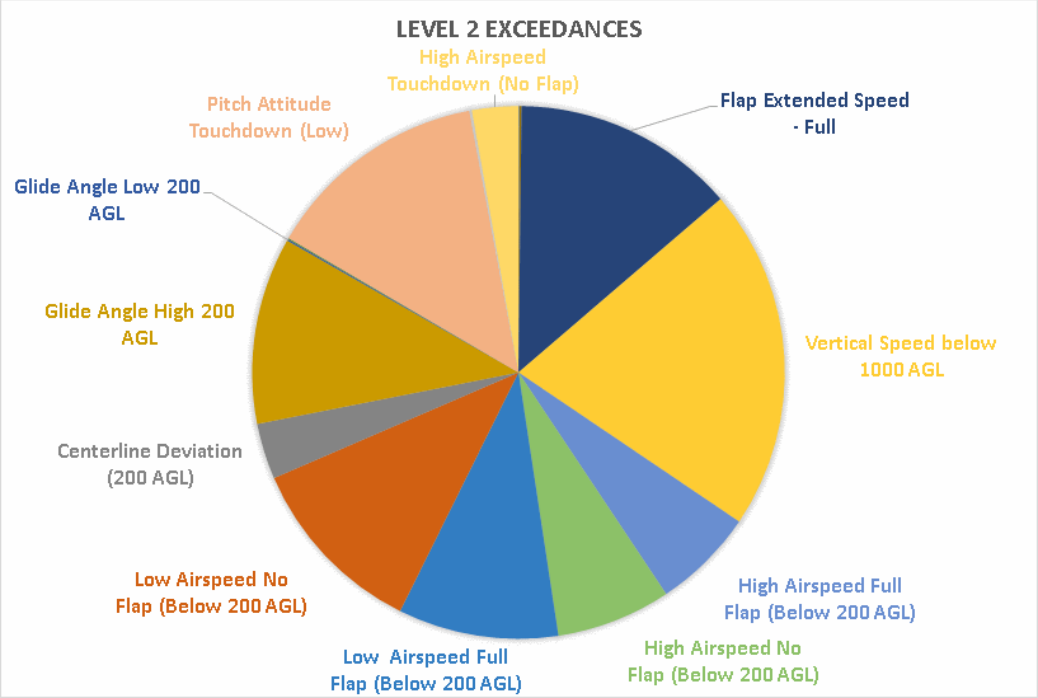


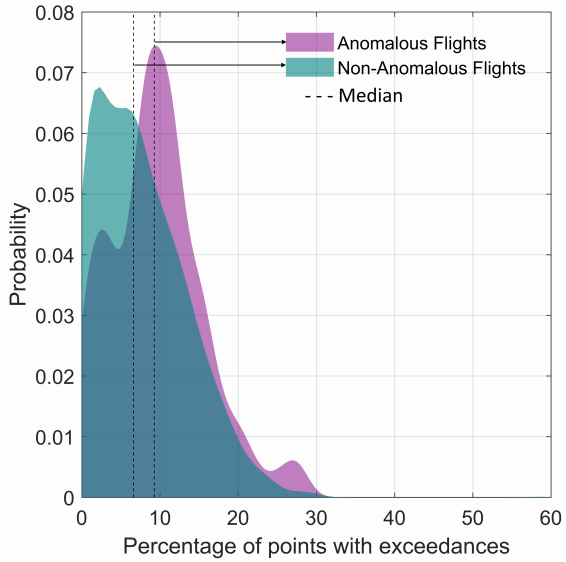
Figure 106: Frequency of different types of level-2 exceedance events in approach and landing

detection of all exceedances despite having a data record. Nevertheless, the overlap of exceedance detection with different types of anomalies is of interest because it helps understand the strengths and weaknesses of the two types of anomaly detection techniques and where each can provide insights for safety. In the next two sections, the overlap between each type of anomaly and number of exceedance events is compared. Since most flights have at least one event, it does not make sense to compare for simple overlap as this would not provide very large agreement between the two but not any meaningful insights.

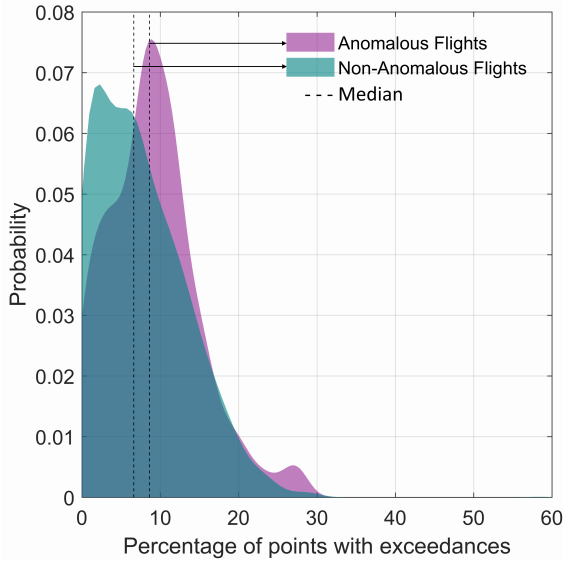
7.1.2.1 Flight-level anomalies

For comparison of exceedance detection with flight-level anomalies, the proportion of points with exceedances in the approach and landing phase is noted for each flight data record and compared with the anomaly status of that flight data record at a particular significance level. Figures 107 and 108 show the distribution of the percentage of

exceedances in anomalous and normal flights at two different significance levels (3 % and 5 %) of flight-level anomaly detection.

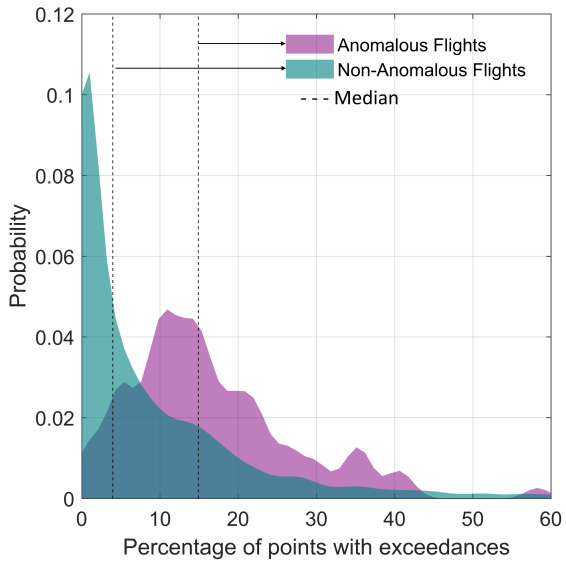


(a) Distribution of number of level-1 exceedances at 3% anomaly significance level

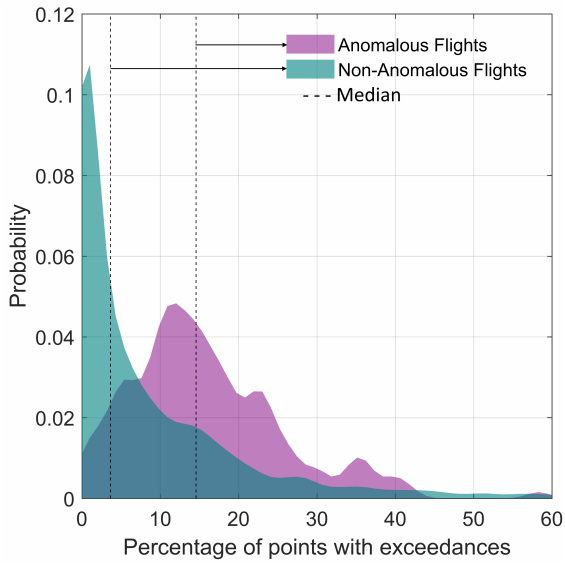


(b) Distribution of number of level-1 exceedances at 5% anomaly significance level

Figure 107: Distribution of percentage of level-1 exceedances in anomalous and non-anomalous flights



(a) Distribution of number of level-2 exceedances at 3% anomaly significance level



(b) Distribution of number of level-2 exceedances at 5% anomaly significance level

Figure 108: Distribution of percentage of level-2 exceedances in anomalous and non-anomalous flights

The figures clearly indicate that for both anomaly significance levels, the percentage of exceedances is higher for flight-level anomalies compared to non-anomalous flights. This distinction is more marked in level-2 anomalies (Figure 108) than in level-1 anomalies as noted by the skewed distribution and the shift of the median of the distribution. This indicates that anomalous flights identified by the algorithm are better differentiated from normal flights in terms of level-2 anomalies than level-1 anomalies. It also suggests that level-1 anomaly definitions in this case might need be more stringent in order to identify actual issues. This conjecture would however, have to be supported by subject-matter-expert reviews. Therefore, while analyzing flights obtained from flight-level anomaly detection, anomalies are more likely to have a higher percentage of level-2 exceedances than normal flights. The percentage of level-1 exceedances is expected to be similar for both normal and anomalous flights. The conclusions drawn here are based on the exceedance definitions currently available from literature. It is noted that operators or users sometimes modify the definitions based on experience and in those cases, the overlap should be evaluated separately. The overall trends, however are expected to be similar as the thresholds for exceedances are usually modified, not the event types.

7.1.2.2 Instantaneous anomalies

Similar to flight-level anomalies, the relationship between instantaneous anomalies and exceedances can be explored using the data set. Initially, the average instantaneous probability for different types of points in the entire data set is obtained as shown in Table 7. It is observed that the average log probability of level-2 exceedance points (-51.08) is lower than level-1 exceedance points (-49.17), which is lower than points with no exceedances (-46.36). Figure 109 shows the distribution of instantaneous anomaly score for level-1 exceedance points, level-2 exceedance points, and those points with no exceedances along with the median for each distribution. Points

with lower instantaneous anomaly score are considered worse in the algorithm.

Type of Point	Average Instantaneous Probability (logarithm)
Points with no exceedance	– 46.36
Points with level-1 exceedance	– 49.17
Points with level-2 exceedance	– 51.08

Table 7: Average instantaneous probabilities for different types of points in flights

These results indicate that there is an overall correlation between the instantaneous anomaly score and the different exceedance levels. Those points with no exceedances typically have better instantaneous anomaly score than those with either type of exceedances. The skew towards the left in the level-2 exceedance score indicates that many points with level-2 exceedances have a markedly lower instantaneous anomaly scores than other points. Despite the shift in the distribution for exceedance points versus non-exceedance points, there is still a significant overlap between all three distributions. The typical thresholds for different levels of instantaneous anomaly detection are also shown in Figure 109. This leads to the following conclusion: a point with an exceedance may not necessarily be an instantaneous anomaly, but a point with instantaneous anomaly likely has at least one exceedance.

7.1.2.3 Summary

In this section, an existing method for quantitative safety analysis (exceedance detection) is compared with the developed methodology. It is observed that in general exceedance detection returns a large proportion of flights with at least one exceedance event, even though this is not expected from routine operations. The number of points that exhibit exceedances per approach phase are however smaller. Upon further examination, the specific exceedance events that get frequently identified for both level-1

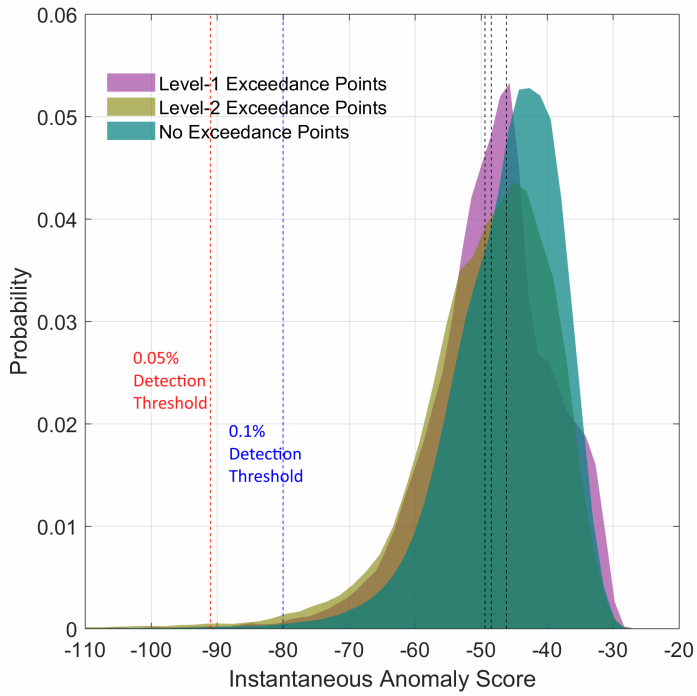


Figure 109: Distribution of probabilities of level-1, level-2, and no-exceedance points

and level-2 exceedances are related to vertical speed, high and low airspeed, center-line deviation, gilde angle, and pitch attitude at touchdown. Exceedance detection is limited in its applicability due to the requirements for some parameters such as flaps to be recorded.

Comparison of exceedance detection to both types of anomaly detection is also presented. It is observed that in general, points or flights with anomalies also tend to have a number of exceedances. This correlation is more pronounced for level-2 rather than level-1 exceedances. There is not an equivalent converse relationship, that is, exceedance points may not necessarily be instantaneous anomalies or be part of a flight-level anomaly. Thus, while overlap is seen in anomalies and exceedance detection, it is not always conclusive. This can be due to several reasons such as vatiations in exceedance definitions from aircraft to aircraft, lack of recorded information for identifying exceedances, inadequate or incomplete exceedances list, etc. However, traditional exceedance detection can definitely be used to complement the

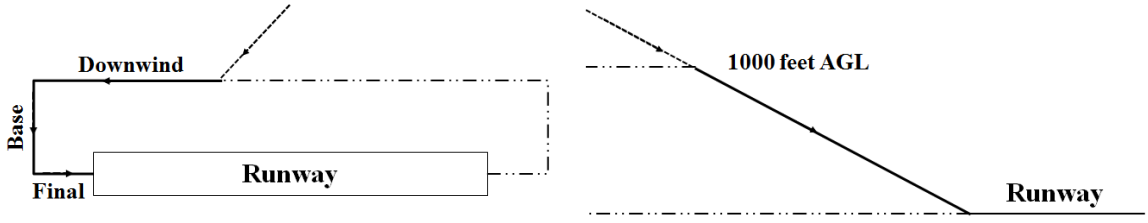
demonstrated energy-metrics based anomaly detection techniques. The main appeal of the anomaly detection approach is its usability in a variety of situations, with different aircraft, with multiple types of data recorders, etc. Anomaly detection is also aimed at providing a smaller, focused list of flights/regions to examine rather than the large set of points that might be obtained from exceedance detection. Thus, anomaly detection as it is undertaken in this dissertation is able to provide incisive insights into the deviating flights and are more usable than the existing benchmark technique of exceedance detection.

7.1.3 Segmentation of Anomalous Flights

While it is interesting and insightful to isolate anomalies and compare them to exceedances, it does not provide the complete picture for safety analysis. The algorithms operate on the recorded metrics and provide anomalies that are mathematical artifacts. The deviations in performance during approach and landing or take-off that are identified using anomaly detection or exceedance detection usually do not take into account operational aspects of the flight data record. This can be an important factor that can provide insight into the reason or causality behind the identification of flight record as anomalous. Within GA flight data records, it is difficult to ascertain the causality due to the fact that a number of important parameters are not recorded even in the more advanced GA recorders (such as the Garmin G1000). Prominent among these parameters are the pilot's control inputs, prevailing weather conditions, flap position, etc. Therefore, it is difficult to pinpoint whether the observed deviations are due to piloting error or unusual/inclement weather conditions or some other cause. However, it is possible to segment the identified exceedances based on different known operational characteristics of the flights to understand whether a particular type of operation has higher chance of exhibiting anomalous behavior. These operational aspects are deemed to be important enough to warrant segmentation of the

data but not significant by themselves (according to the algorithm results) to get separated into different clusters during anomaly detection. Two examples of such operational differences are demonstrated here. It is noted that depending on other data being available, more such scenarios may be explored. The main idea behind these segmentations is that when each subset has sufficient number of flights, the anomaly rates within that subset should roughly equal that of the super-set. If a particular subset of flights that is separated operationally exhibits higher rate of anomalies, then flights in that operation might be worth further examination or separate analysis.

7.1.3.1 Straight-in v/s pattern approaches



(a) Approach entering through traffic pattern (b) Straight-in Approach

Figure 110: Difference between approach profiles for straight-in versus pattern-entry approaches

In GA operations under VFR, there might be differences in the approach profiles for a straight-in approach versus an approach entering through the airport traffic pattern (as seen in Figure 110). Since the two type of approaches differ in the typical tasks undertaken, it could be expected that one or the other type of approach is more difficult from an energy management perspective for student pilots or private pilots with less flying experience. In the present data set of 3,402 approaches, 449 are identified as straight-in approaches and the remaining 2,953 are pattern-entry approaches. Therefore, it is evident that a higher proportion (86 %) of the current data set consists of pattern-entry approaches. During application of the anomaly detection algorithms, no distinction is made between these two type of approaches as the clustering carried out as the first step points to only a single cluster. However,

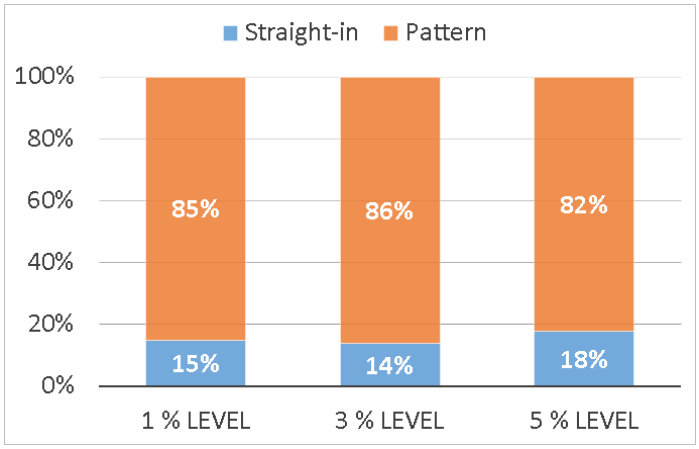


Figure 111: Proportion of different types of approaches within anomalous flights at various significance levels

in the post-processing step, the anomalies from each type of operational scenario are separated and examined.

Figure 111 shows the proportion of straight-in and pattern approaches identified as anomalous at different flight-level anomaly detection thresholds. As seen from the figure, the proportion of straight-in anomalous flights is low ($\approx$ 15 %) at all significance levels. However, this could be due to the fact that the data set being training flights, there are a lot more pattern approaches than straight-in approaches in this data set ($\approx$ 86 % pattern and $\approx$ 14 % straight-in).

Therefore, in Figure 112, the number of anomalous approaches at each significance level is normalized by the total number of approaches of that type in the data set. If a particular type of approach is more susceptible to being anomalous, then the proportion of anomalous flights in Figure 112 would be quite higher than the anomaly significance level in each case. The results indicate that for each type of approach, approximately the same proportion of flights from that approach type are anomalous as the overall signigicance level. This indicates that despite pattern approaches following a slightly different operational routine, there is not a tendency for higher anomalies in this type of approach compared to straight-in approach

For instantaneous anomalies, the probability density for the two different types

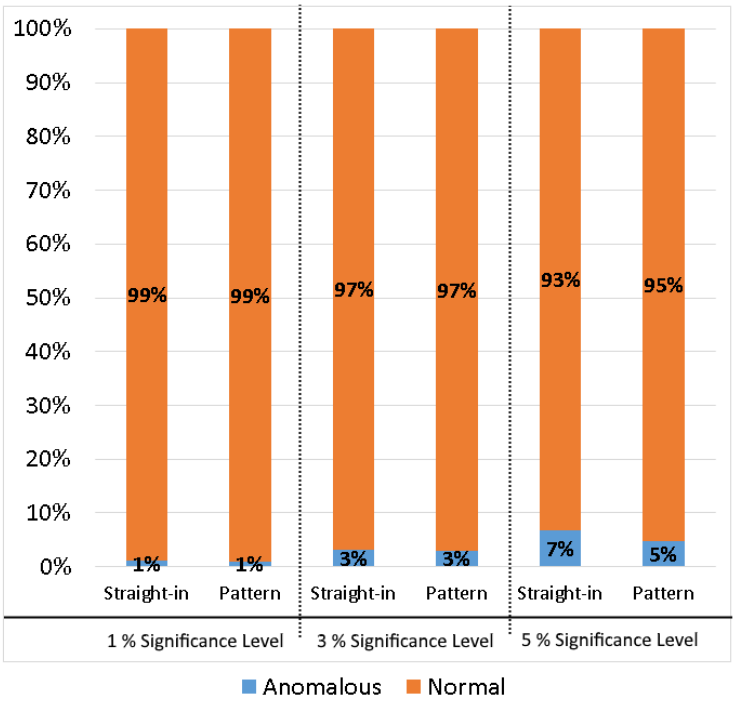


Figure 112: Proportion of anomalous flights within each type of approach at different significance levels

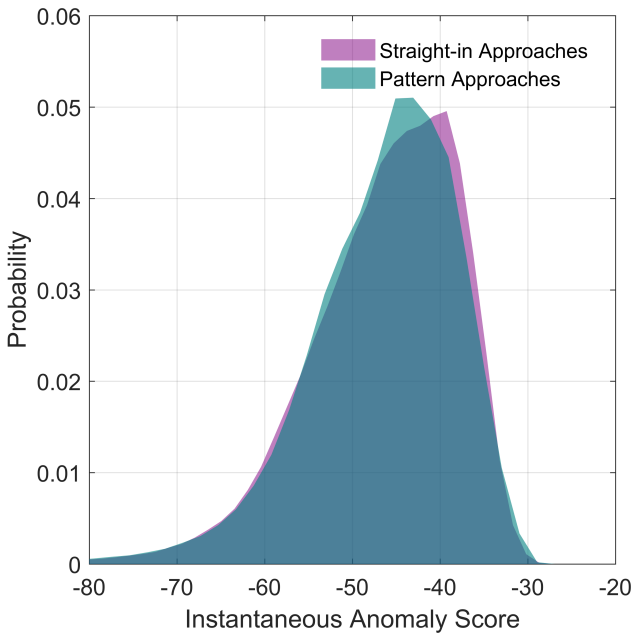


Figure 113: Distribution of probabilities for points from straight-in and pattern approaches

of approaches are also plotted, however, as evidenced from Figure 113, there is little noticeable difference in the overall probability distributions during approach between straight-in and rectangular pattern approaches either in terms of the median, or the spread of the distributions. The difference between the variation of metrics and raw parameters between the two types of approaches can be visualized in the next four figures (Figures 114, 115, 116, 117).

As seen from the variation of the metrics, the main difference between the two types of approaches is in the specific potential and kinetic energies. The spread of potential energy during straight-in approaches is lower than rectangular pattern approaches as is expected due to clear sight of the runway from earlier in the approach. The spread and average of most of the other metrics is quite similar for the two approach types. Similarly, examining the raw flight parameters, the roll angle variation is markedly different due to the turns in pattern approaches. Similarly, as flights in the pattern approaches tend to intercept the glideslope from above, the pitch variation is more negative towards the end of pattern approaches where the flight path is steeper than three degrees. Finally, it is noted that pattern approaches have a larger spread in the RPM values indicating that more power corrections are required in pattern approaches than in straight-in approaches where the aircraft can be configured earlier.

Thus, this subsection demonstrated that while straight-in and pattern approaches differ operationally and have some differences in the distribution of certain metrics, either type of approach does not have higher tendency of being identified as anomalous (either flight-level anomaly or instantaneous anomaly).

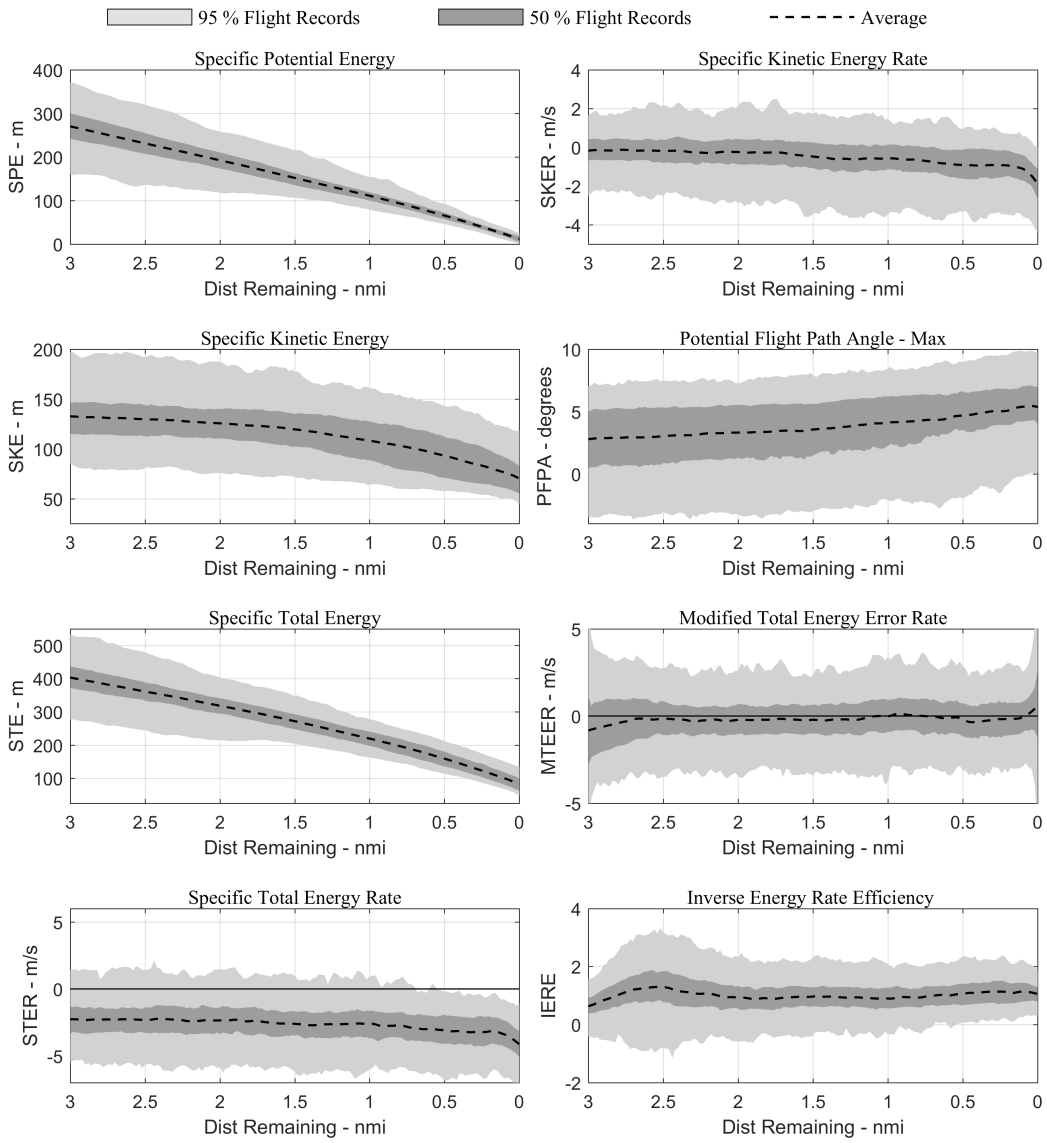


Figure 114: Variation of energy metrics for straight-in approaches

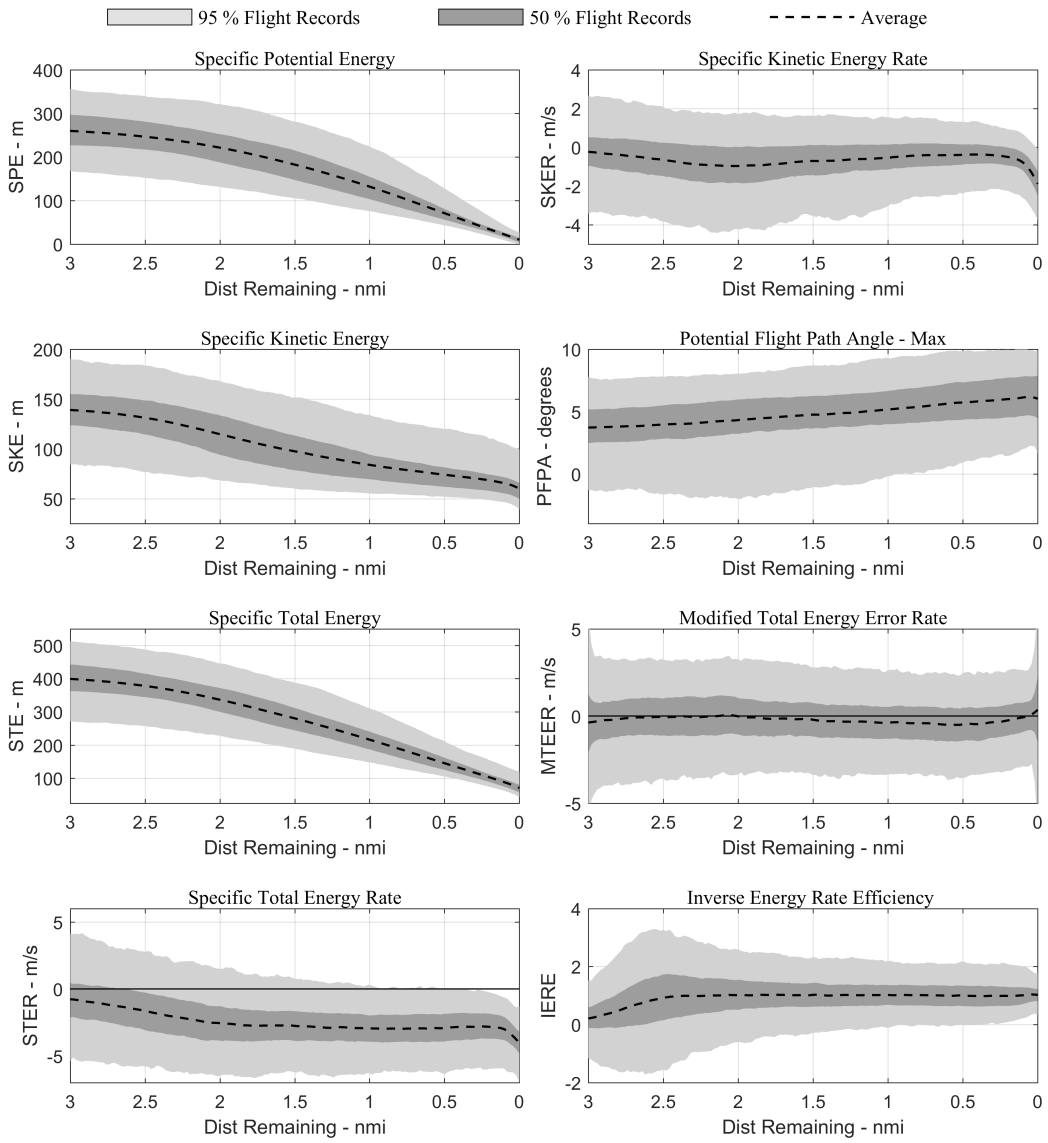


Figure 115: Variation of energy metrics for pattern approaches

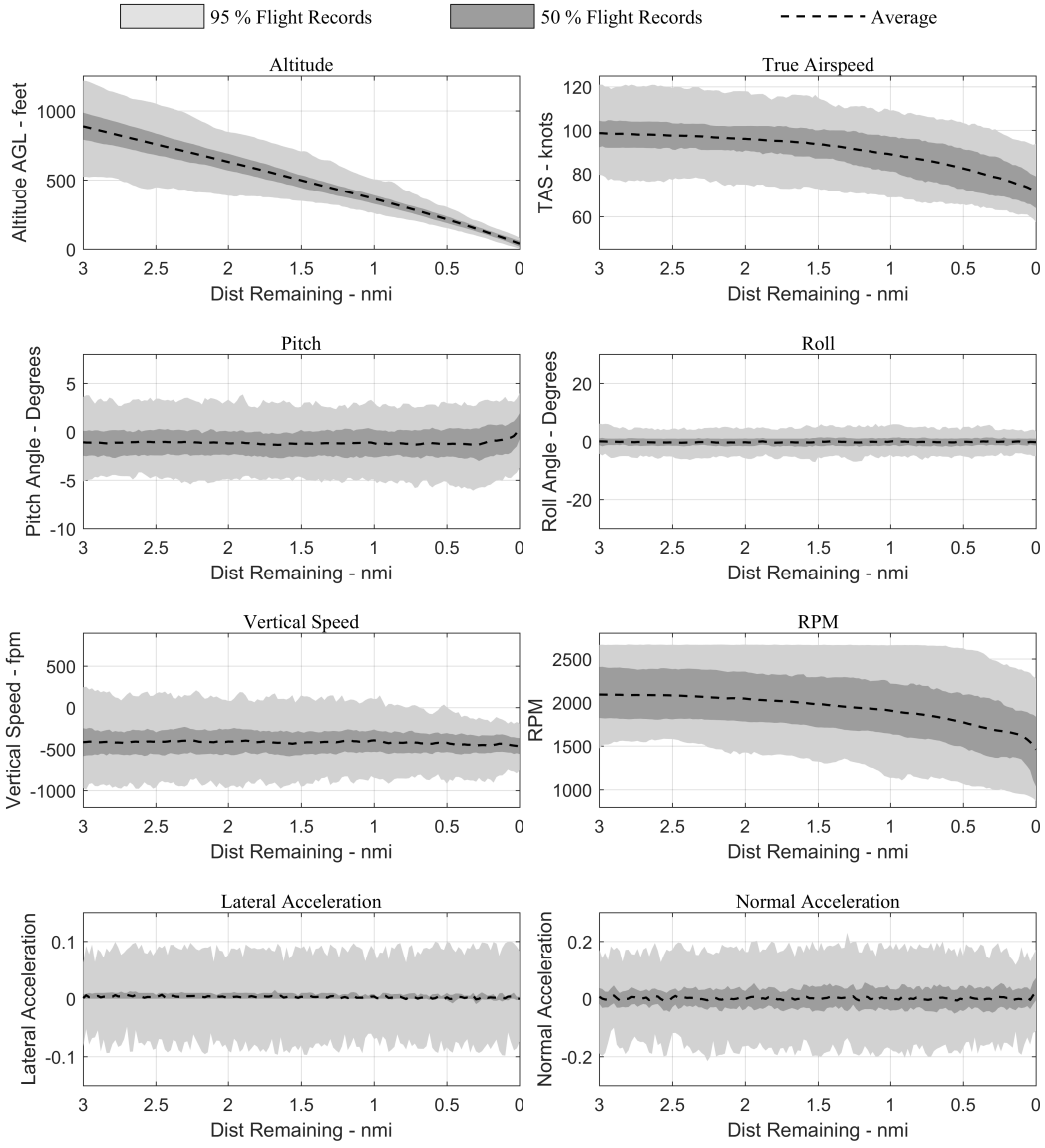


Figure 116: Variation of raw parameters for straight-in approaches

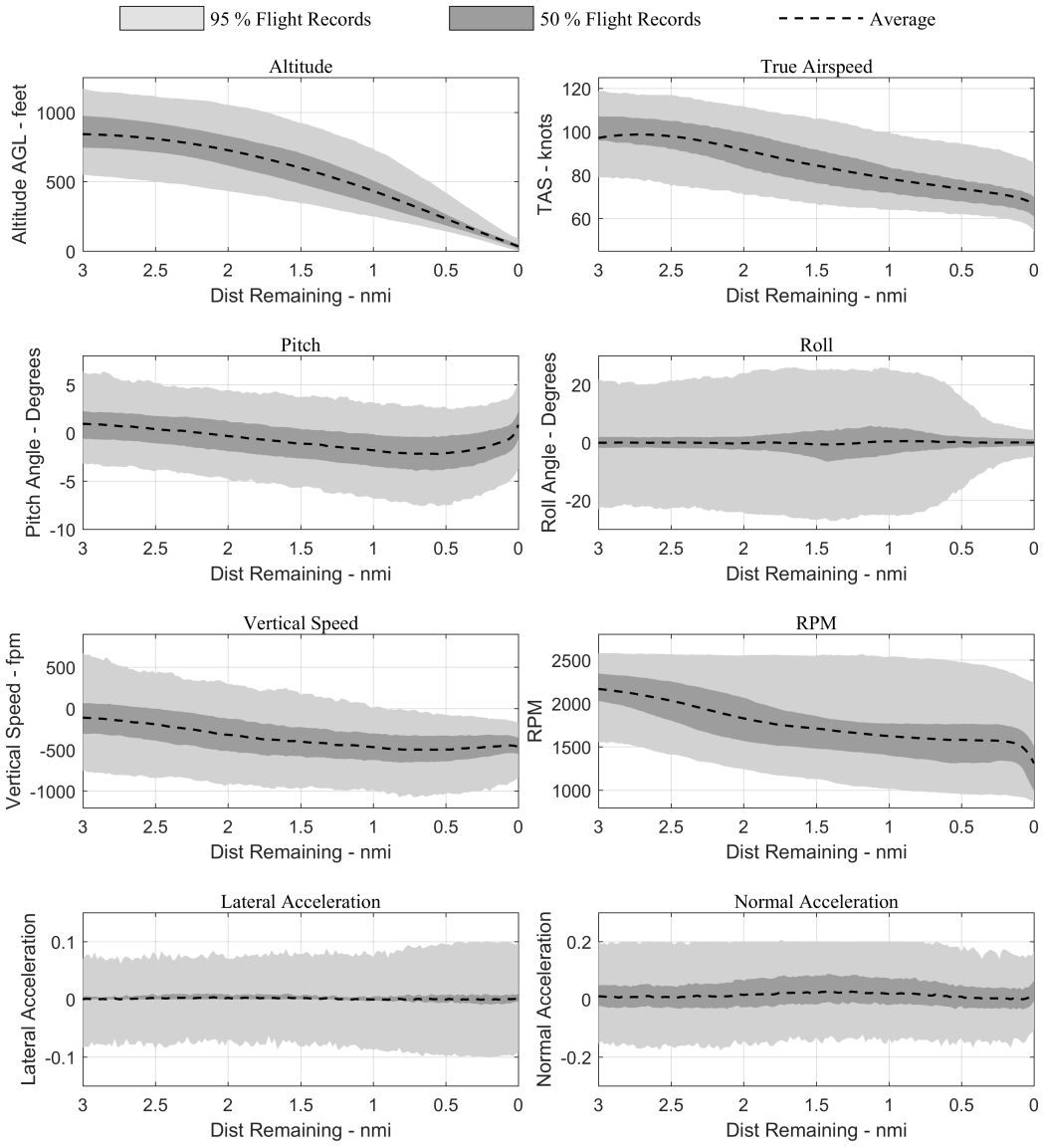


Figure 117: Variation of raw parameters for pattern approaches

7.1.3.2 Flap Configuration

The use of high-lift devices or flaps in final approach and landing can impact the anomaly detection because the energy profiles for flights are expected to be quite different with and without flaps extended. Because of the limitations in the current data set $\approx$ 60% of the flights have flap angle recorded. Therefore, for this subsection, only those flight data records are focused upon to gauge the relationship between anomalies and flap usage in final approach and landing. Table 8 provides a summary of the different approaches with full, partial, and no flaps. In each case, the total number of flights, and the number of anomalous flights from that subset is provided. The term in parenthesis in each case represents what percentage of flights from that subset are anomalous.

	No. of Flights	**No. of flight-level anomalies (1%)**	**No. of flight-level anomalies (3%)**	**No. of flight-level anomalies (5%)**
Full Flaps	662	6 (0.90%)	12 (1.81%)	22 (3.32%)
Partial Flaps	931	12 (1.29%)	36 (3.86%)	60 (6.44%)
No Flaps	395	7 (1.77%)	22 (5.57%)	41 (10.37%)

Table 8: Summary of flights with full, partial, and no flaps during approach and anomalous flights among them

The results from Table 8 indicate that the number of full flaps approaches is smaller than those with partial or no flaps. Comparing the anomalies within the three types of approaches it is observed that the proportion of anomalies within a particular type increases from full to partial to no flaps. This indicates that at the same anomaly detection level, a higher proportion of anomalies are obtained within the subsets that have progressively lower flap usage. This distinction becomes more marked as the overall anomaly significance level changes from 1 % to 5 % (from left-to-right in the table). For example, when the overall significance level is 5 % (last

column of the table), the proportion of full flaps anomalies is just 3.32 % compared to 10.37 % for no flaps. Ideally this number would be close the 5 % for each of the subsets. Therefore, this suggests that the segmentation of flight data set according to flaps usage is a useful operational segregation of the flights as it pin-points to some subsets that have higher rate of anomalies.

In order to understand this further, the variation of different energy metrics for full, partial, and no flaps configurations during approach is visualized in Figures 118, 119, and 120. As is evident from the figures, the variation of specific kinetic and total energy is quite different for the three figures. This is due to the fact that for approaches with partial or no flaps, the speeds have to be higher in order to avoid drifting close to the stall speed. For flights with lower flap setting, the spread is also higher for specific potential energy metric. Flights with full flap setting also have lower kinetic energy rates than the lower flap setting. Thus, overall, it can be concluded that flights with no flap usage or partial flap usage have higher rates of flight-level anomalies and also exhibit some differences in their metric variation.

Comparing instantaneous anomaly scores for approaches with different amount of flap usage is shown in Figure 121. As is evident from the figure, the distribution of no-flap anomaly scores is to the left of partial-flap anomaly score which is further to the left of full-flap anomaly scores. Therefore, on an average, reduced flap setting has a slightly higher likelihood of instantaneous anomaly than higher flap usage. However, the difference between the distributions is not significant enough to draw any conclusions. Understanding the nature of the operational difference between different flap configuration approaches, it is not expected that there will be any significant differences in their instantaneous anomaly score as this score looks for sudden changes rather than a systematic difference such as different flap configuration throughout the approach and landing.

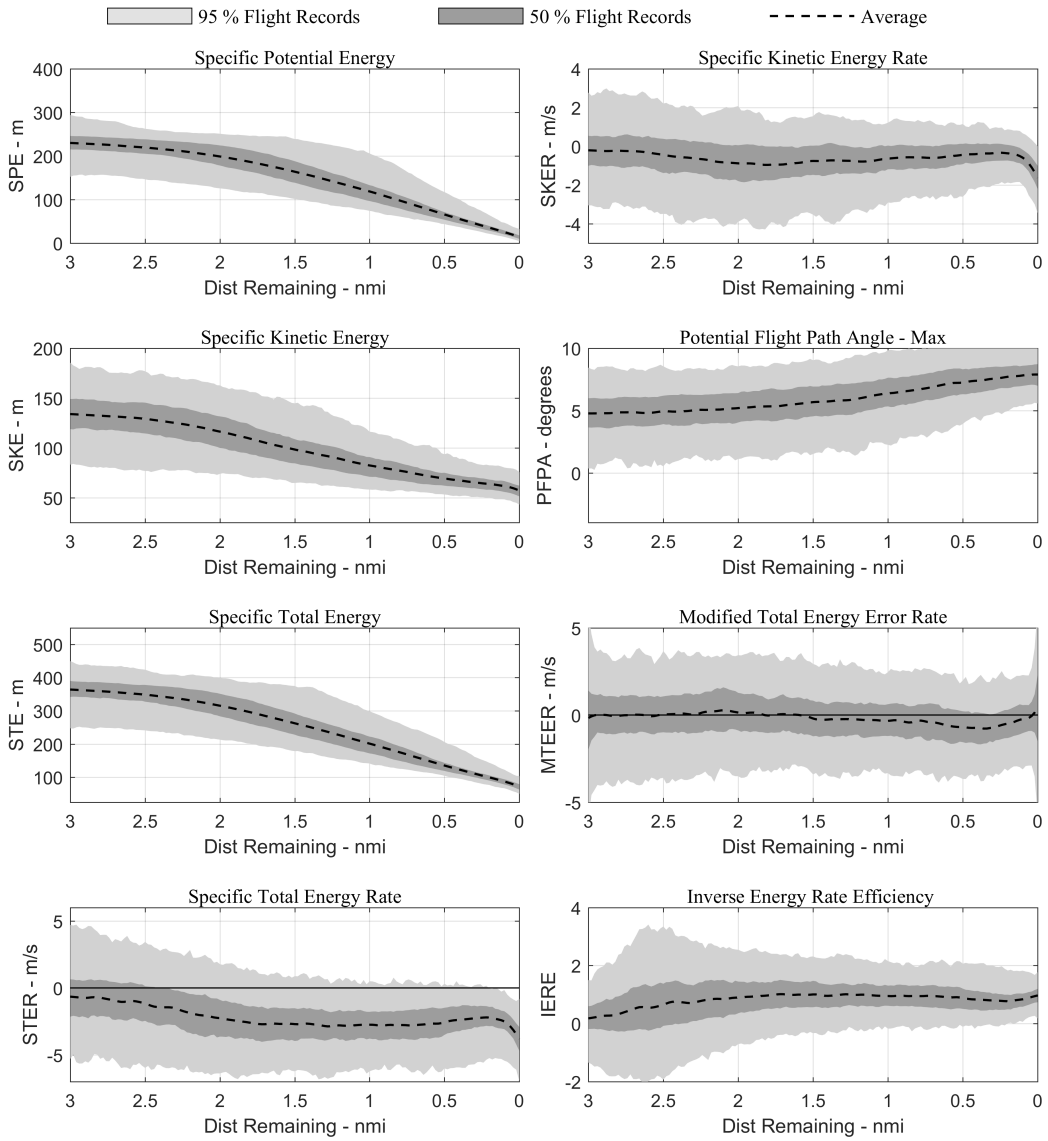


Figure 118: Variation of energy metrics during approach for flights with full flaps usage

7.1.3.3 Summary

In this sub-section the current data set was segmented into different types of subsets based on certain operational characteristics. In each case it is observed that the typical characteristics of that operational difference manifest themselves in the

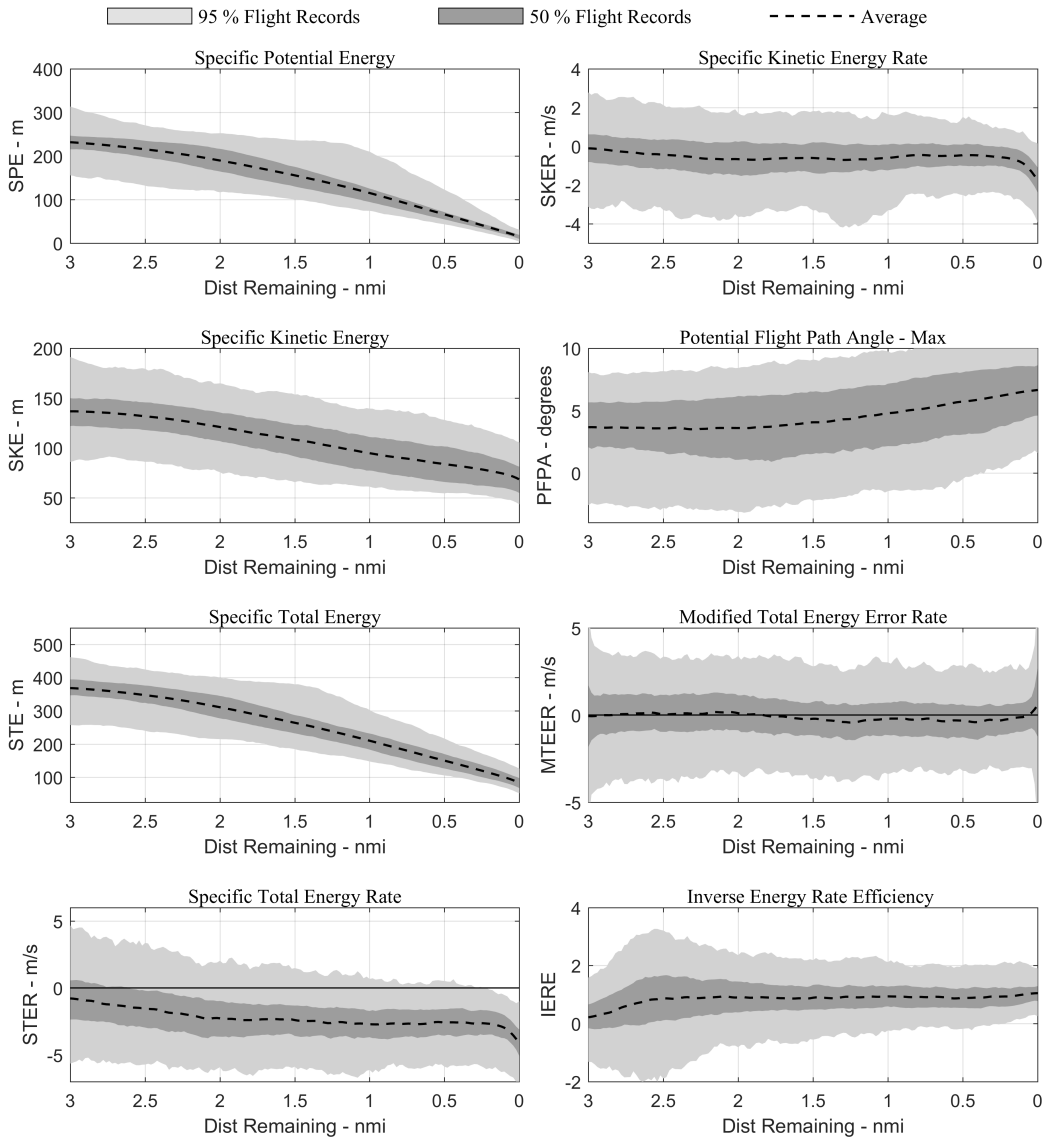


Figure 119: Variation of energy metrics during approach for flights with partial flaps usage

variation of some metric associated with the operational difference (e.g. potential energy profiles for straight-in and pattern or kinetic energy profiles for different flap configuration). While the straight-in and pattern approaches did not exhibit much difference in the anomaly detection rates, segmenting by the usage of full, partial, or

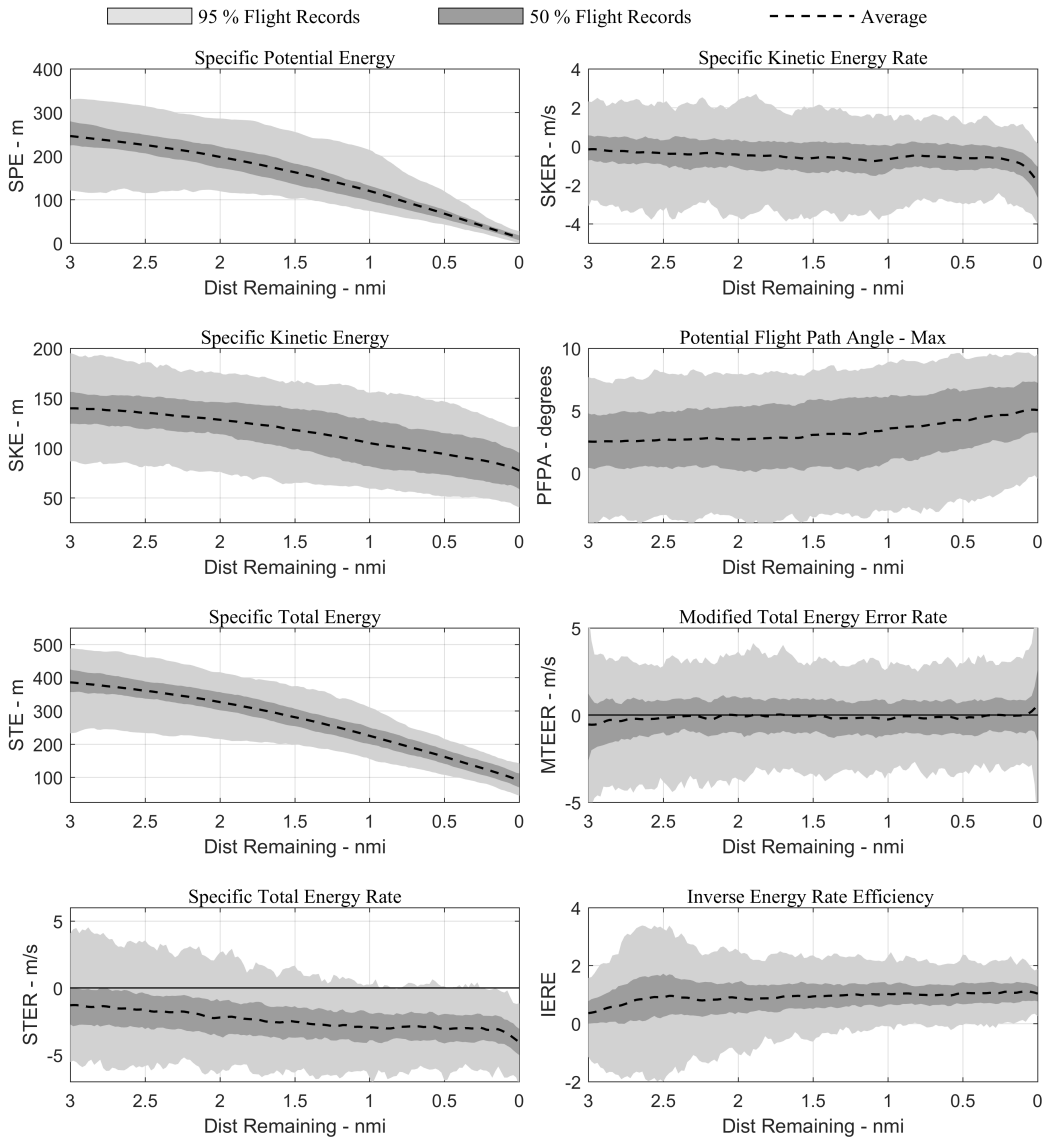


Figure 120: Variation of energy metrics during approach for flights with no flaps usage

no flaps during approach revealed slight differences in the anomaly rates. Thus, this section demonstrated the benefits of using subject-matter knowledge such as type of approach or landing configuration in the anomaly detection post-processing to gain insights.

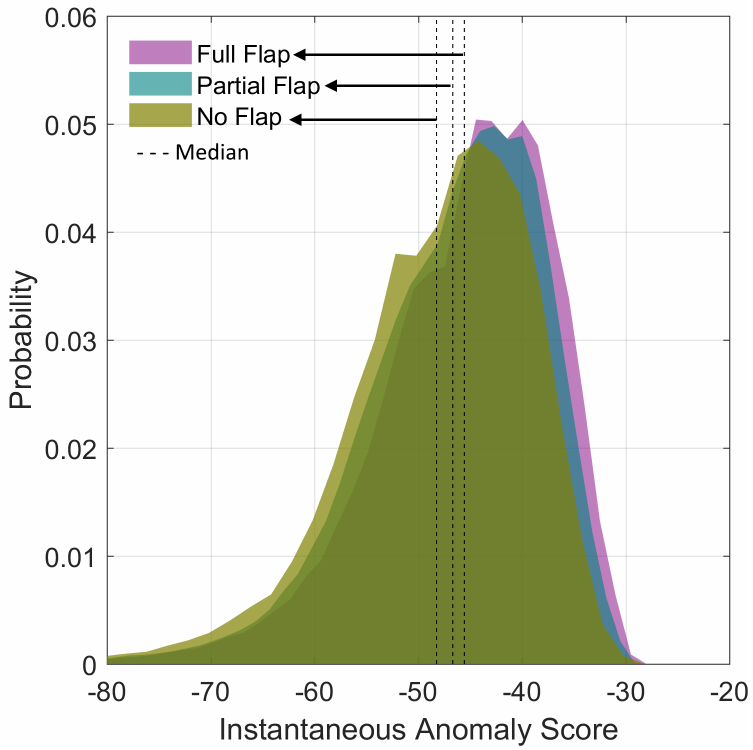


Figure 121: Distribution of probabilities of for points from full-flap, partial-flap, and no-flap approaches

7.1.4 Touchdown Performance

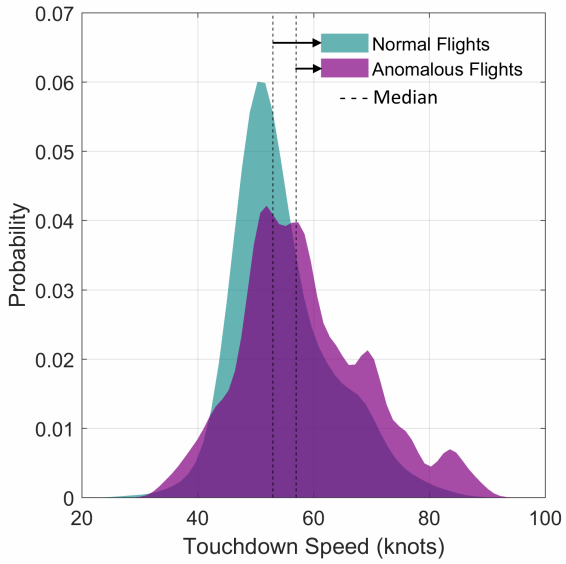
A poorly executed approach is likely to end with a poor landing. An assessment of landing and touchdown operational safety with FDM can be realized by segregating the portions of one (or more) flight data records corresponding to these phases of flight, and comparing each data points against aircraft states of interest. There can be various measures of flight quality and safety during landing and touchdown that can be extracted from flight data records. For example, the touchdown distance is important because landing too close to the runway threshold could lead to a higher risk of a controlled-flight-into-terrain (CFIT) accident, and landing too far down the runway could lead to a runway overrun. Similarly, large deviations from the center line could lead to a runway veer-off. Similarly, rate of descent at touchdown is important as an unstable approach with high descent rate could potentially lead to a hard or bounced landing. A list of potential metrics of interest is enumerated here:[1]

1. Touchdown point on runway
2. Touchdown velocity
3. Touchdown vertical speed
4. Touchdown lateral deviation from centerline
5. Difference between track angle and runway direction at touchdown
6. Touchdown lateral velocity (should be zero)
7. Bank Angle at touchdown
8. Aircraft crab angle at touchdown

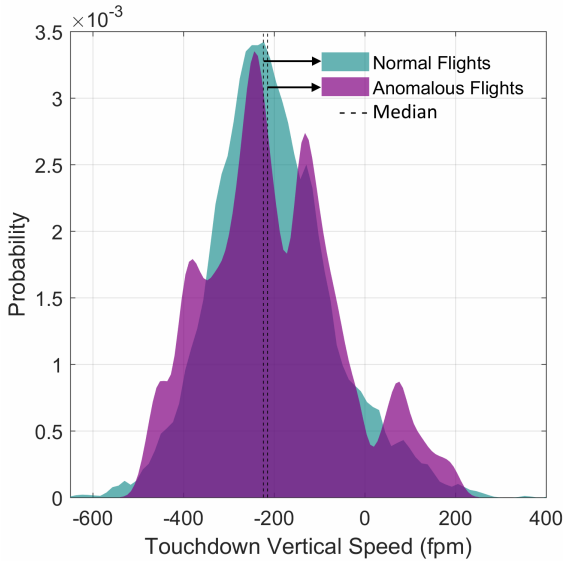
It is noted that in typical GA flight data recorders, due to the scarcity of recorded

[1]The research described in this section is also documented in the following publication:

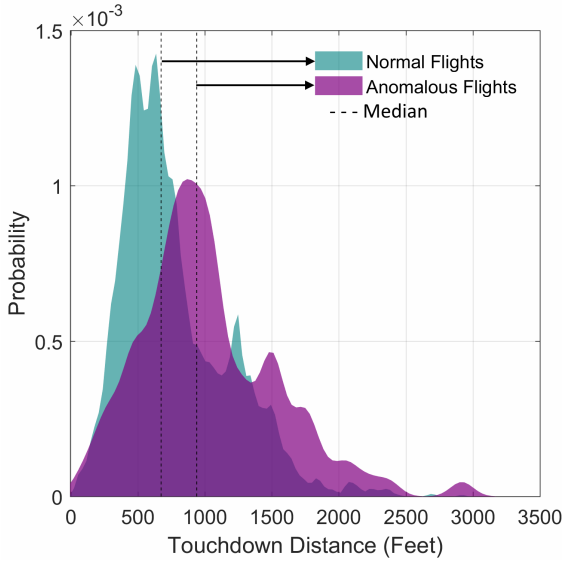
– **Puranik, T.G**, Chakraborty, I.C., Rao, A.H., and Mavris, D., *Safety Analysis of General Aviation Landing and Touchdown using Approach Performance and Weather Data*, in 18th AIAA Aviation Technology, Integration, and Operations Conference, Atlanta, GA (Accepted for Publication) [121]



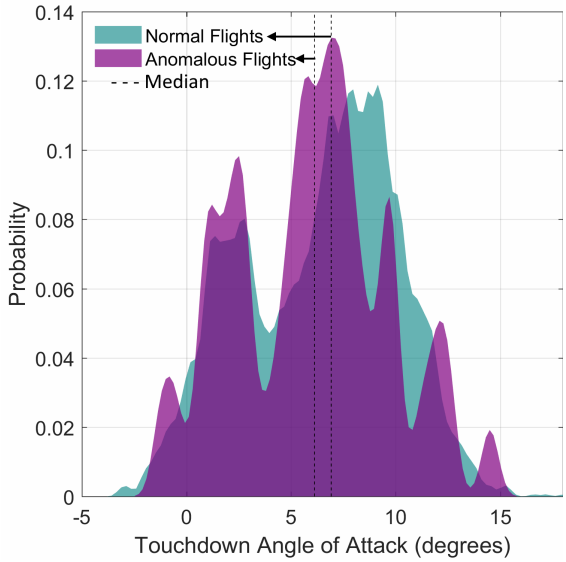
(a) Distribution of touchdown speed for normal and anomalous flights



(b) Distribution of touchdown vertical speed for normal and anomalous flights



(c) Distribution of touchdown distance for normal and anomalous flights



(d) Distribution of touchdown angle-of-attack for normal and anomalous flights

Figure 122: Distributions of various touchdown parameters for normal and anomalous appraoches

parameters, some of the measures of flight quality identified above may not be estimable. A comparison of the correlation between the identified touchdown parameters of interest and anomaly scores during approach will provide additional insights

as to potentially why the touchdown performance was such. For each of the parameters from the list above that are recorded in the data set available, a distribution is obtained for flight-level anomalies and normal flights to understand if flight-level anomalies affect touchdown and landing performance. The quantities at touchdown are evaluated at two points before and after the identified touchdown point and averaged. This is done because the touchdown point itself is not recorded in the flight data record but estimated using the algorithms described in Appendix C.

The results shown in Figures 122, indicate the differences in some of the touchdown parameters of interest between normal and anomalous flights as obtained at the 3 % significance level. The first figure indicates the difference between touchdown speed for normal and anomalous flights. While there is a range of speeds which overlap for the two, anomalous flights tend to have a larger spread of speeds at which they touchdown. There are some flights among the anomalous distribution that land at speeds higher than 20 knots of the normal median. The second figure shows the distribution of vertical speed at touchdown. In this case, the distributions are more similar to each other than the speed distribution. However, a slight rise in the probability is observed for positive vertical speeds for anomalous flights.

The third chart shows the distribution of touchdown distance from the displaced runway threshold. In this figure, there is a clear shift of the median and the overall distribution towards the right for anomalous flights. This indicates that more anomalous flights tend to touchdown long on the runway compared to normal flights. The distribution of the touchdown angle of attack shown in the fourth figure has a bimodal tendency for both normal and anomalous flights indicating two different sets of flights coming in at a higher and lower angle of attack. There is not noticeable difference between the normal and anomalous flights with respect to the touchdown angle of attack.

7.1.4.1 Summary

In this section, the impacts of a normal or anomalous approach on the touchdown performance are examined. Mixed results are observed for different parameters, where an approach anomaly has an effect on some parameters such as airspeed at touchdown and touchdown distance but does not have noticeable difference on other parameters such as vertical speed and angle of attack. It is noted that these results are obtained without true knowledge of the actual touchdown point and are therefore subject to more uncertainty than other results in this chapter.

7.2 Specific Anomaly Examples

The anomalies identified (flight-level and instantaneous) using the methodology can be further analyzed by subject-matter-experts or operators to understand and improve safety. As seen in the case study, not all anomalies necessarily correspond to pre-defined events or exceedances. Similarly there is not necessarily a consensus between the two types of anomalies either. In this chapter, a few specific flights in the take-off and approach-and-landing phases are chosen for visualization. The properties of these flights in terms of their anomaly scores, exceedances, and other important aspects are listed in separate tables to enable further understanding. Three types of flights are presented in each section, the first is a flight with instantaneous anomalies but no flight-level anomaly. The second case is a flight with flight-level anomaly but no instantaneous anomaly. The third is a flight with both types of anomalies. For every flight, the variation of the instantaneous probability density over the flight phase, energy metric plots, and flight parameter plots are provided. Commentary and comparative evaluation is provided for each case. The flight data records presented in this section can be considered as the end-product of the methodology developed in this dissertation. Eventually, the aim of the methodology is to aid in quantitative safety assessment and help regulatory bodies as well as pilots learn and understand

from data. Therefore, the post-processing presented in this section can be considered as an end-user using the methodology to understand deviating behavior. A non-expert on the energy metrics approach can also understand and interpret the results shown in this section and thus speaks to the methodology's appeal beyond academic or scholarly circles.

7.2.1 Approach and Landing

No.	Score (Flight-level Anomaly (Y/N))	No. of Instantaneous Anomalies	No. of L-1 Exceedances	No. of L-2 Exceedances
561	30.0 (N)	4	32	20
2184	- 45.8 (Y)	0	50	61
2007	- 38.6 (Y)	3	65	39

Table 9: List of anomalies visualized in approach and landing

7.2.1.1 Instantaneous Anomaly Only

The first specific type of anomaly considered in approach and landing is represented in Figure 123. This flight contains an instantaneous anomalous window but does not get tagged as anomalous for flight-level anomaly detection.

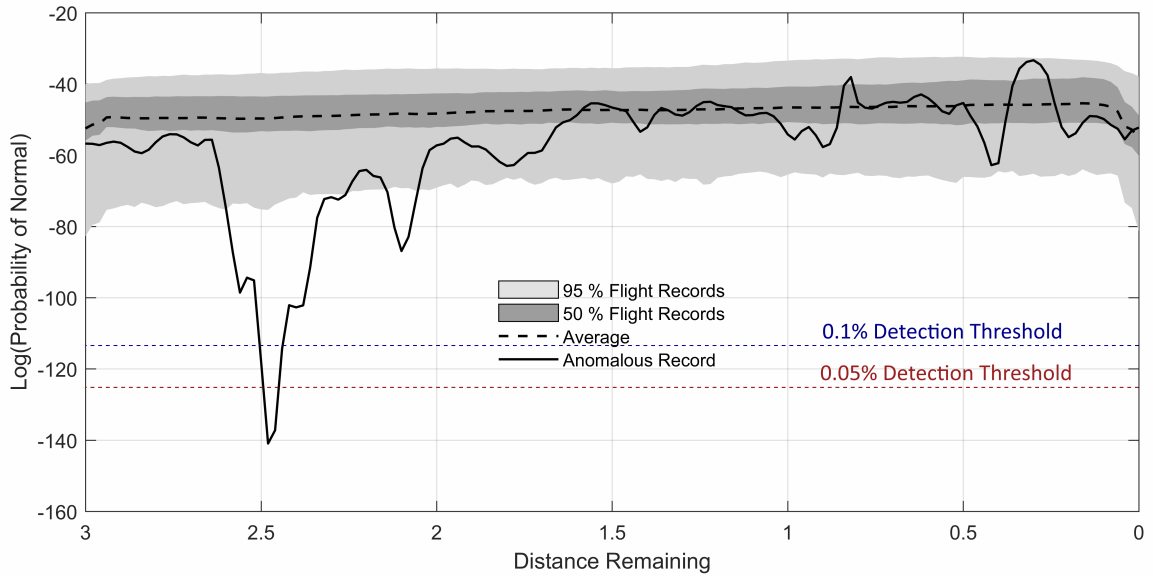


Figure 123: Variation of probability density during approach for flight with instantaneous anomaly only

Exploring the location of the instantaneous anomaly further using the energy metrics plot provided in Figure 124, it is observed that at the location of the instantaneous anomaly there is a rapid decrease in kinetic energy and a simultaneous increase in potential energy. This is possibly because the aircraft comes into the approach with both kinetic and potential energies higher than normal and is attempting to correct it in the early stages of the approach. Visualizing the raw flight parameters for this flight record (Figure 125), the exchange between potential and kinetic energies is confirmed with a higher than usual (and positive) pitch angle and vertical speed. The parameters for the rest of the approach follow nominal variations. Two instances of roll angle varying indicate that it is likely an approach following the traffic pattern. The RPM does go higher towards the end indicating the possibility of the flight going

too low and being corrected for, but it is not unusual enough to warrant a drop in instantaneous probability density.

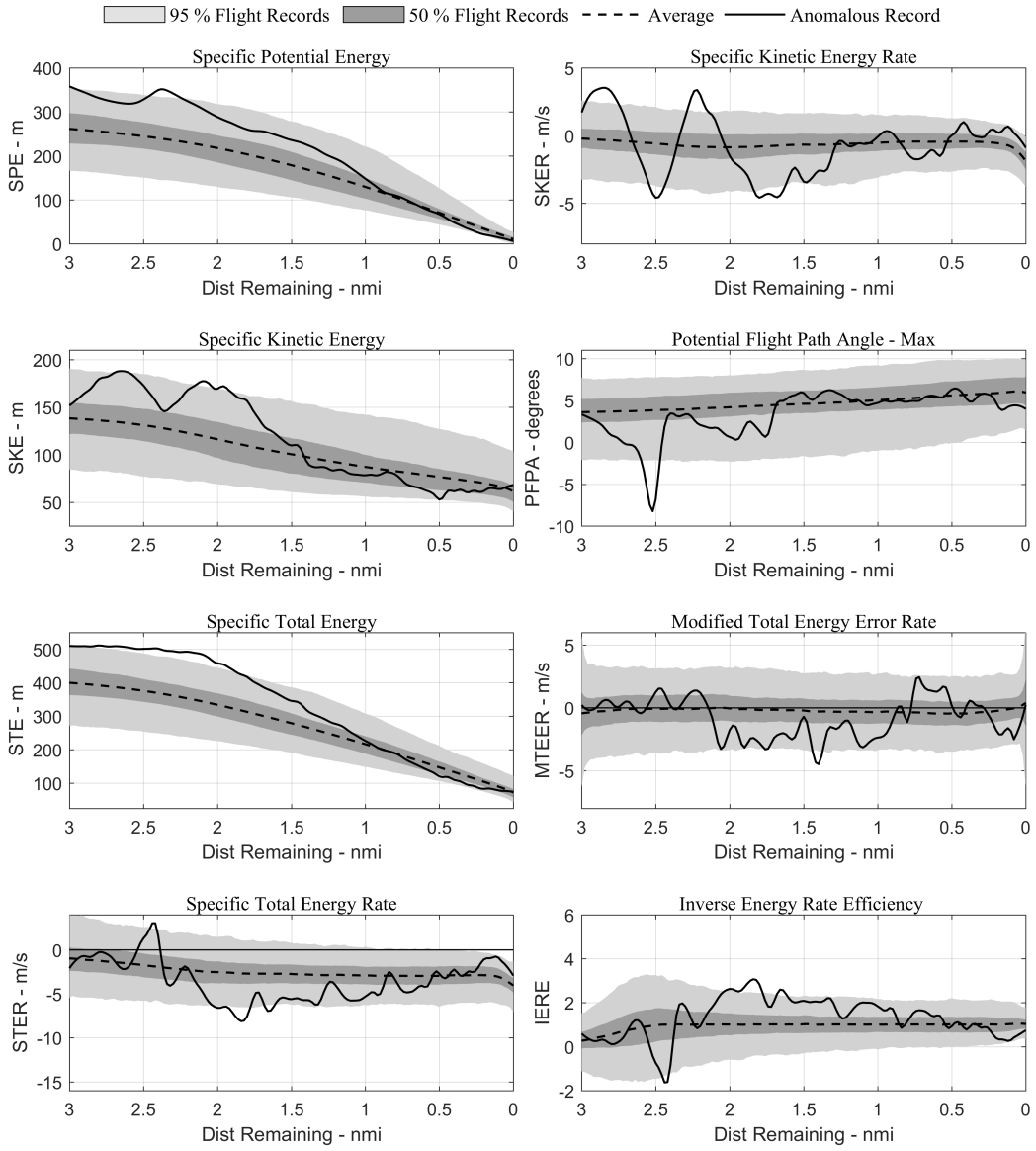


Figure 124: Variation of energy metrics for a flight record with instantaneous anomaly only

Thus for this approach, the corrections made in the early approach set it up for

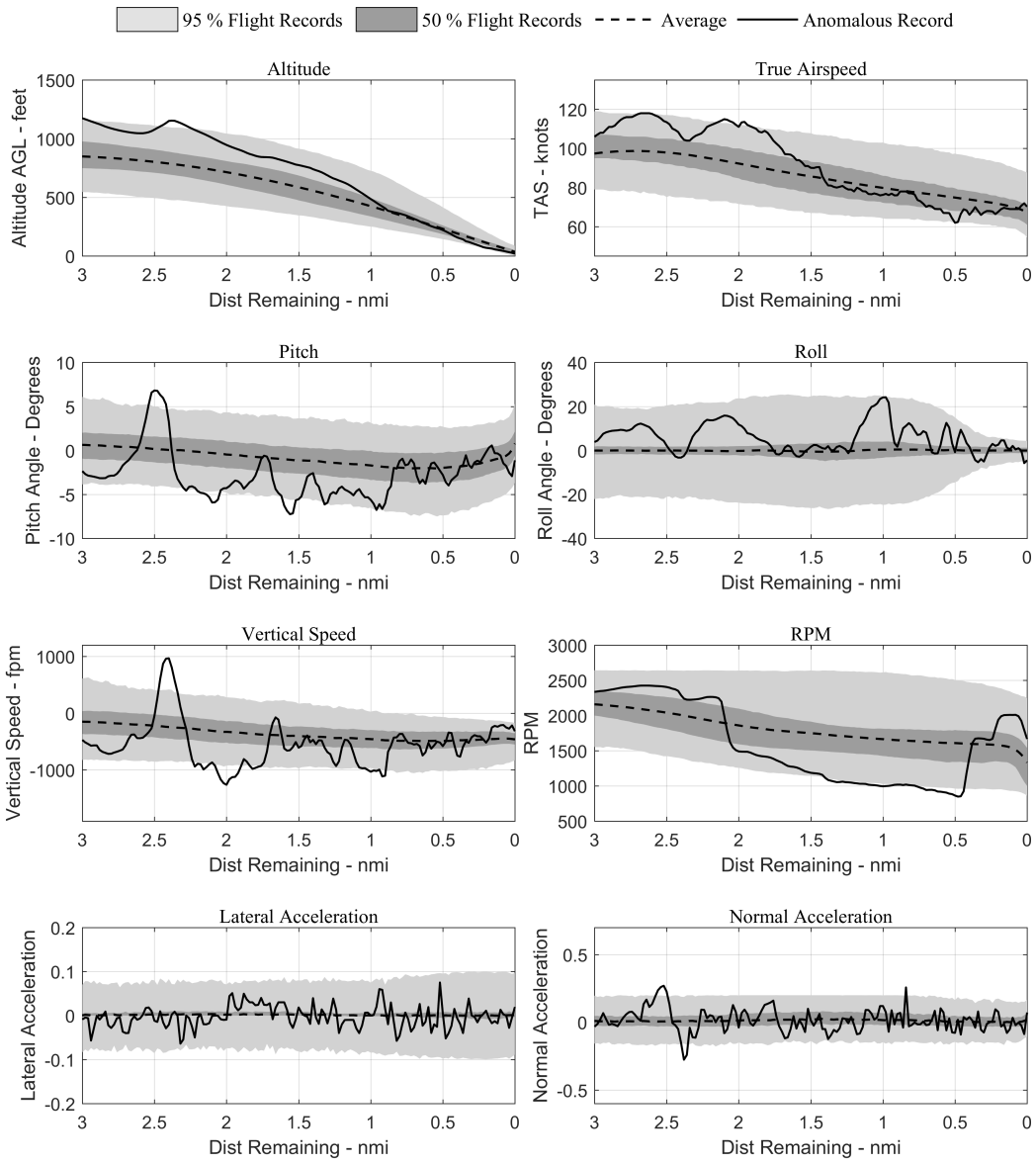


Figure 125: Variation of raw parameters for a flight record with instantaneous anomaly only

nominal variations in all parameters and energy metrics following the instantaneous anomaly and therefore, the flight-level anomaly score returns positive. This flight also has 32 level-1 and 20 level-2 exceedances as listed in Table 9.

7.2.1.2 Flight-level Anomaly Only

The next type of approach anomaly considered here is a flight with flight-level anomaly but no instantaneous anomalies. The instantaneous probability density for this flight is visualized in Figure 126. As seen from the figure, this flight has three drops in instantaneous probability density, however none of them are serious enough to fall below the thresholds set in the case study.

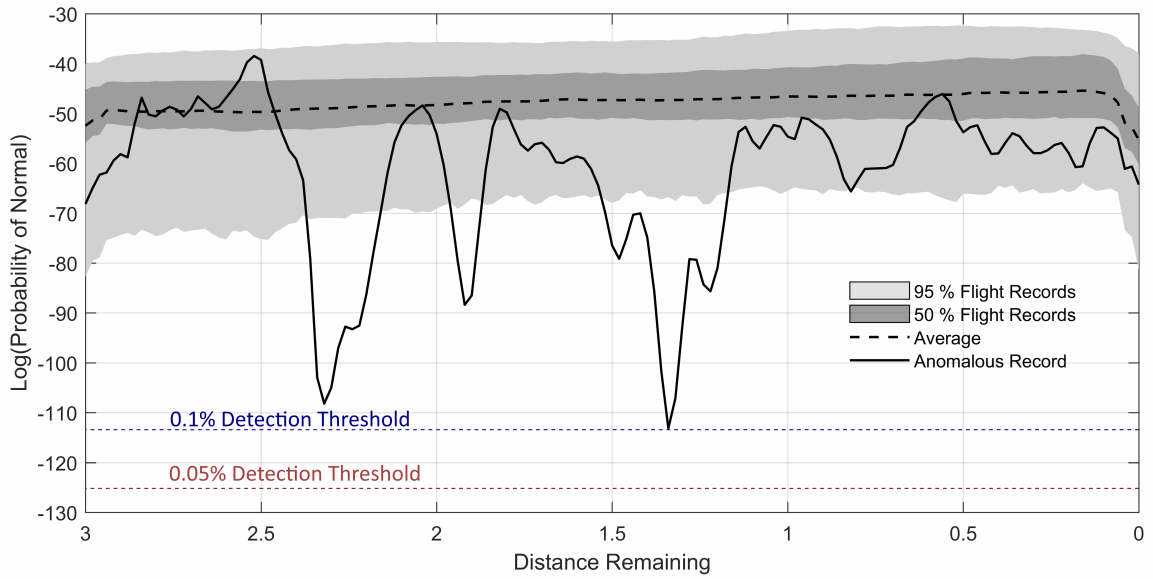


Figure 126: Variation of probability density during approach for flight with flight-level anomaly only

Visualizing the variation of energy metrics in Figure 127, it is evident that the flight starts out with lower potential energy and higher kinetic energy. At the location of the first drop of probability, the flight is aiming to correct the potential and total energy profiles, as seen from the negative modified total energy error rate metric. However, even after intercepting the correct profiles, the aircraft keeps gaining energy and ends up with a very high total energy towards the landing portion near the runway threshold.

The visualization of raw parameters (Figure 128) paints a similar picture indicating RPM drop near the second probability drop where an attempt is made to correct

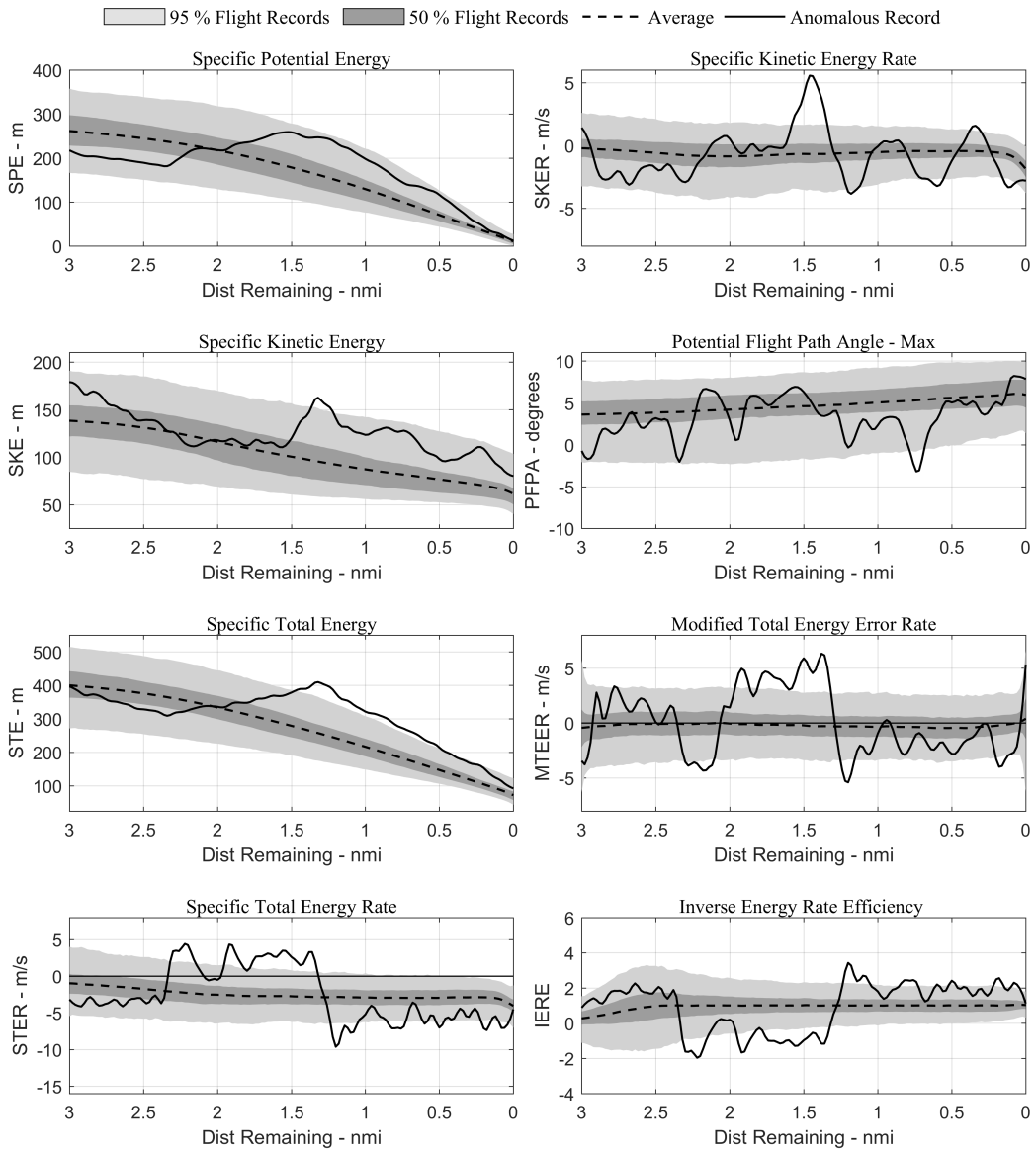


Figure 127: Variation of energy metrics for a flight record with flight-level anomaly only

the overshoot. Vertical speeds for the flight are regularly exceeding nominal bounds in the early and latter stages of the approach. This flight has 50 level-1 and 61 level-2 exceedances as listed in Table 9.

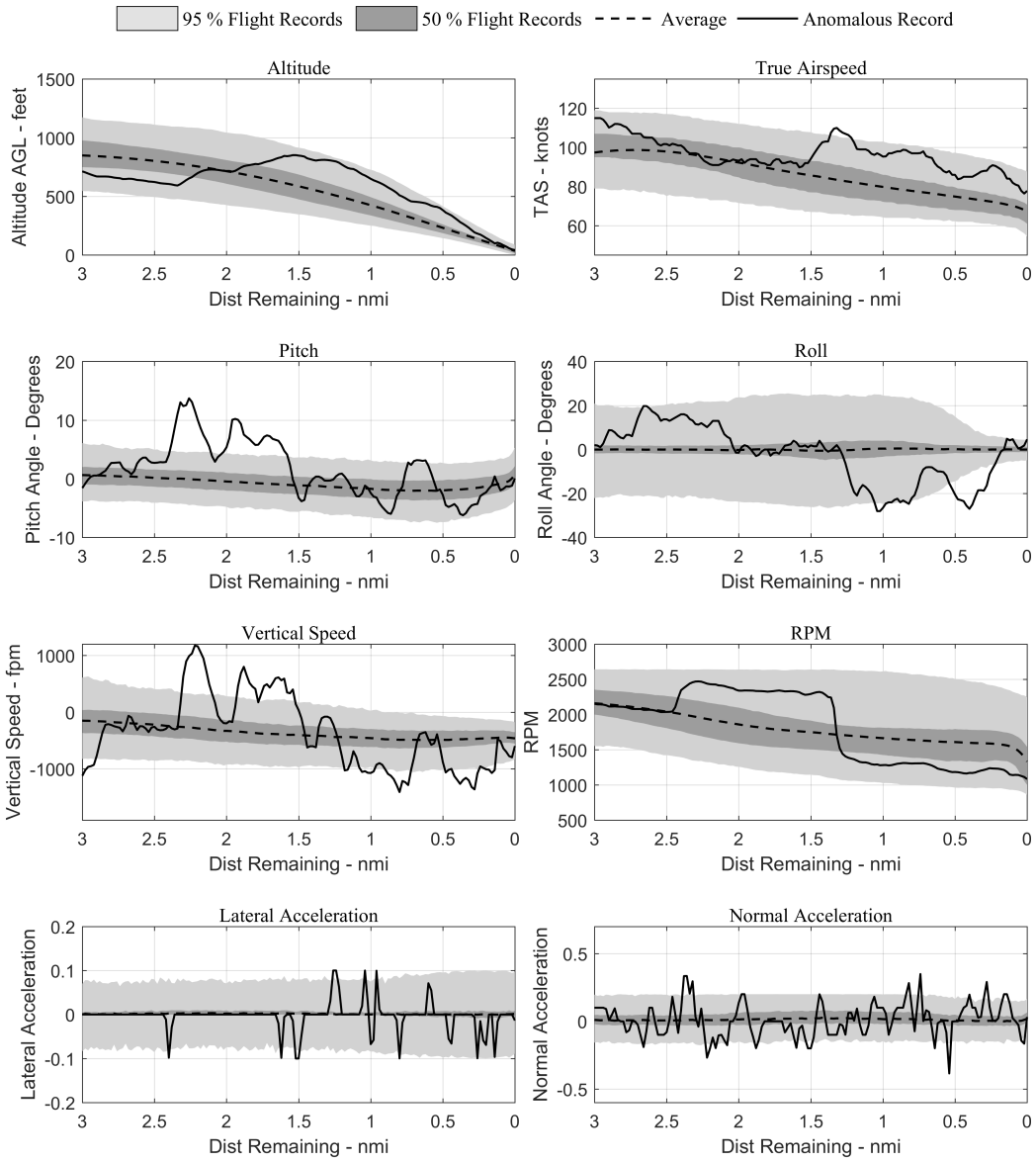


Figure 128: Variation of raw parameters for a flight record with flight-level anomaly only

7.2.1.3 Flight-level and Instantaneous Anomaly

The final anomaly in approach and landing phase is an flight record with instantaneous and flight-level anomalies. The probability density for this flight is depicted in Figure 129. The flight has one instantaneous anomalous window as well as another region where there is a drop in probability but which does not fall below the threshold for instantaneous anomalies. A visualization of energy metrics for this flight record is provided in Figure 130. As seen from the figure, the flight data record starts at a much lower kinetic, potential, and total energy value than nominal flights. The initial part of the approach is spent in recovering towards the nominal variations.

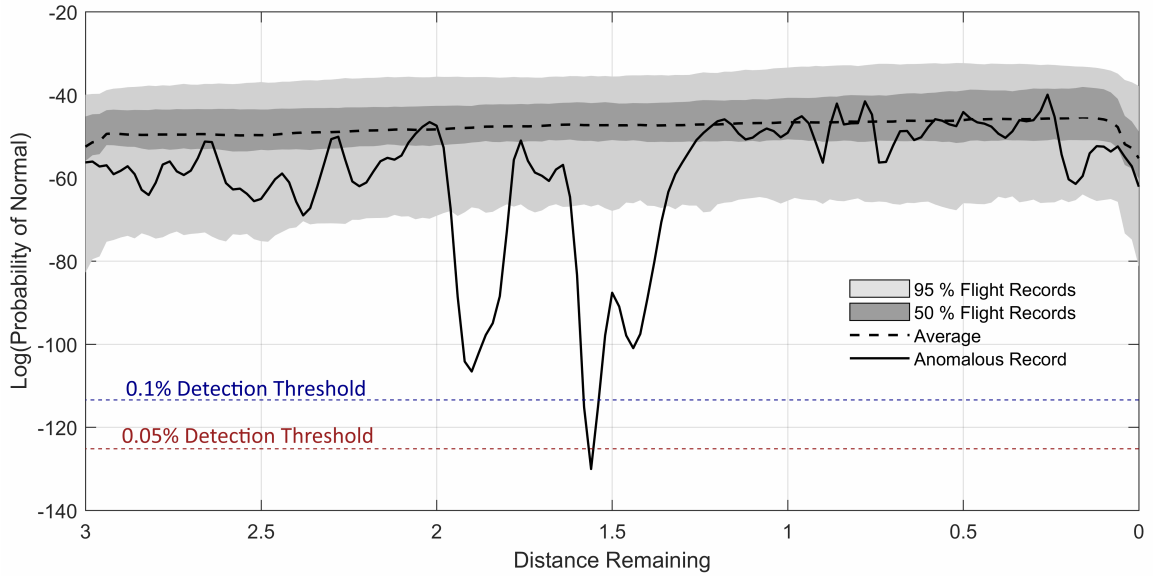


Figure 129: Variation of probability density during approach for a flight with instantaneous and flight-level anomaly

However, this recovery overshoots for all the energy metrics by a large amount as seen from the modified total energy error rate which is positive between 2.5 and 1.5 miles. At around 1.5 miles remaining mark, the deviations have exceeded sufficiently that measures are taken to recover to the nominal. This is where the instantaneous anomaly occurs due to the magnitude of the corrective changes as seen by the energy metrics. Turning attention to the variation of raw parameters seen in Figure 131 it is

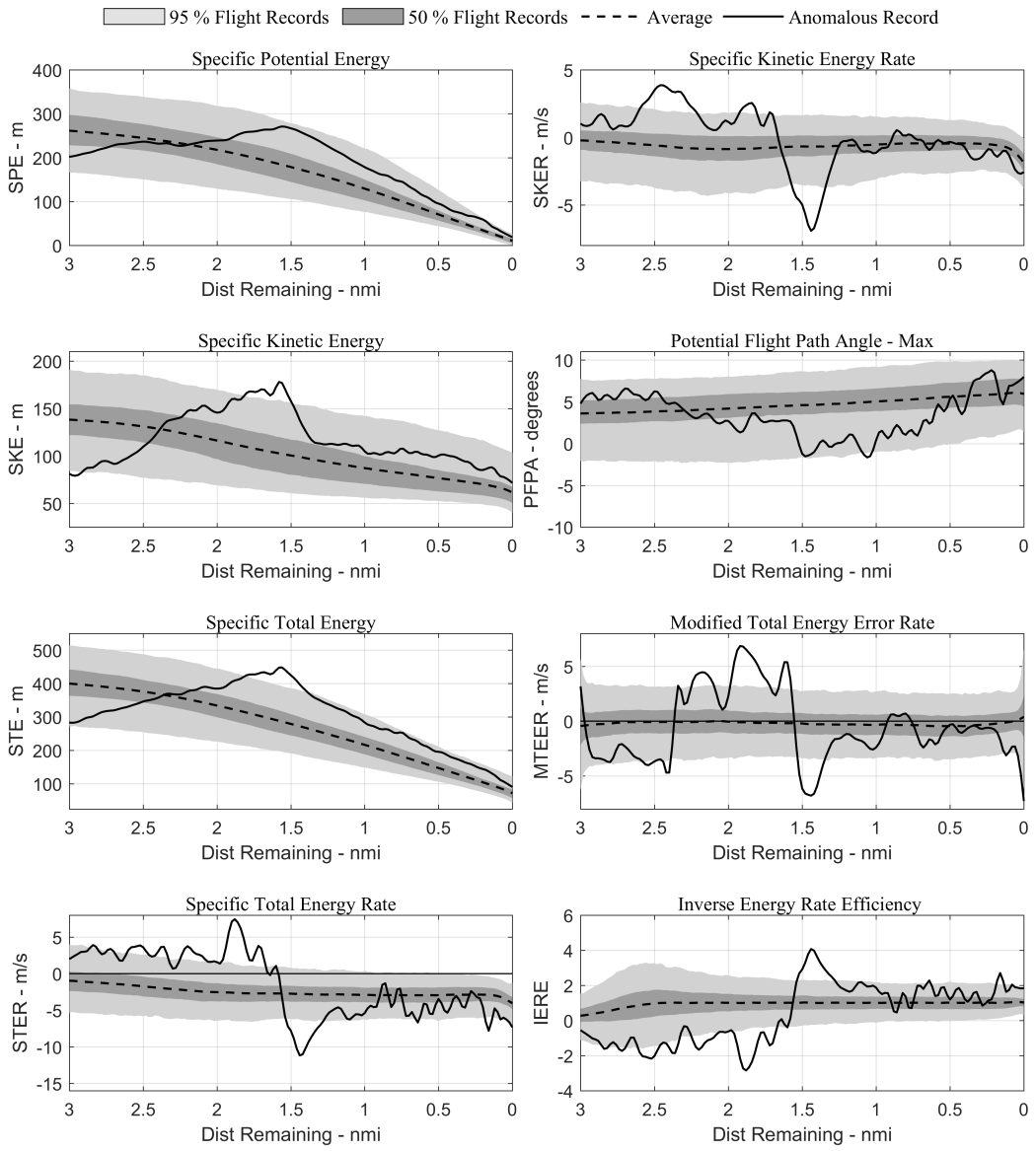


Figure 130: Variation of energy metrics for a flight record with instantaneous and flight-level anomaly

evident that the corrective measure taken at the instantaneous anomaly is the cutting off of the power. The pitch angle is still maintained in order to rapidly lose airspeed which has crept well above 100 knots. Thus, the flight record displays both types of

anomalies and deviations from nominal operations and might be worth understanding in order to improve energy management during approach and landing. This flight has 65 level-1 and 39 level-2 exceedances as listed in Table 9.

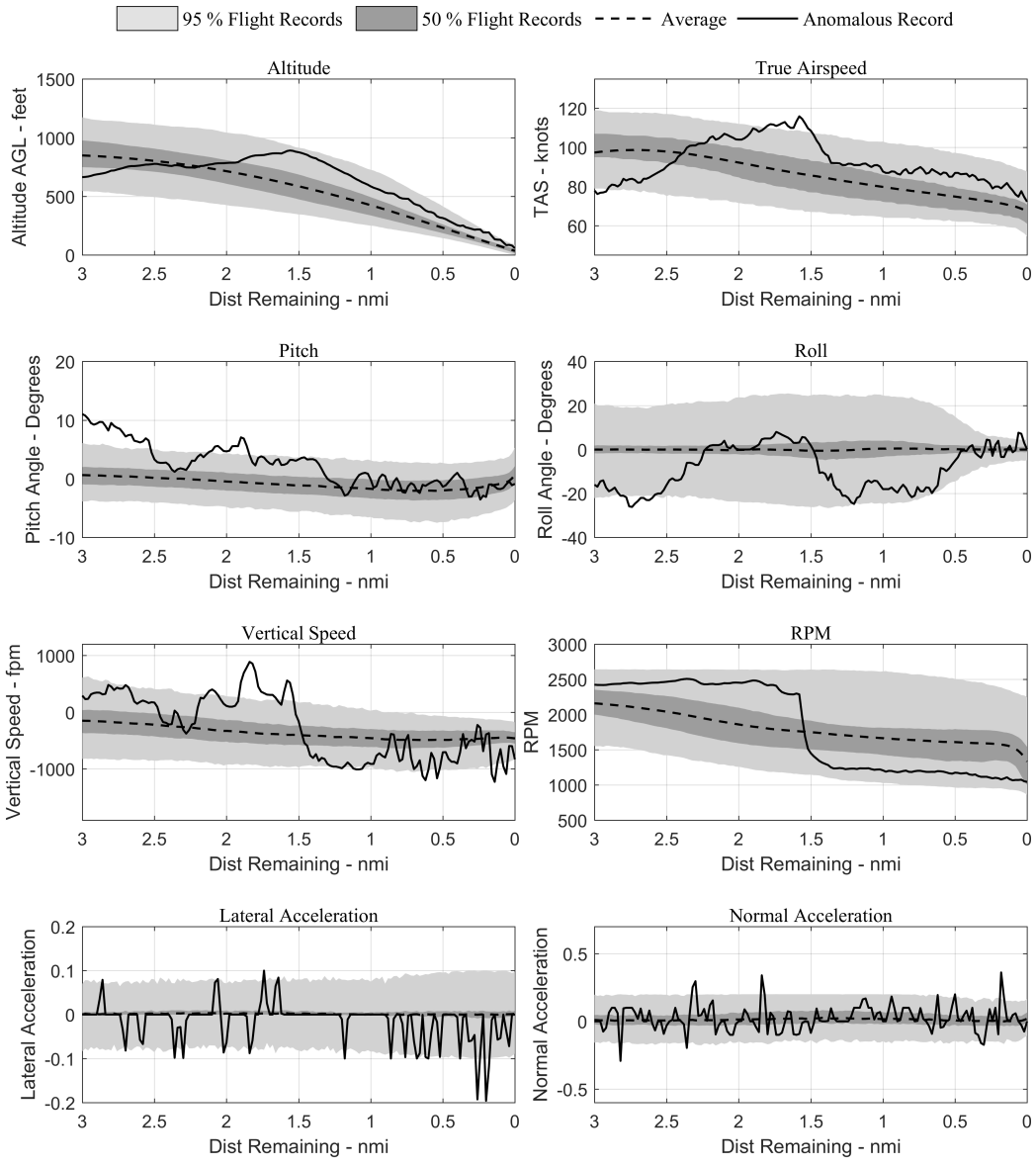


Figure 131: Variation of raw parameters for a flight record with instantaneous and flight-level anomaly

7.2.2 Take-off

For the take-off phase, exceedance detection is not performed due to the limited types of exceedances in literature and requirement for recorded parameters

No.	Score (Flight-level Anomaly/Not)	No. of Instantaneous Anomalies	No. of L-1 Exceedances	No. of L-2 Exceedances
119	26.3 (N)	7	–	–
381	- 29.4 (Y)	0	–	–
2152	- 37.3 (Y)	2	–	–

Table 10: List of anomalies visualized in take-off

7.2.2.1 Instantaneous Anomaly Only

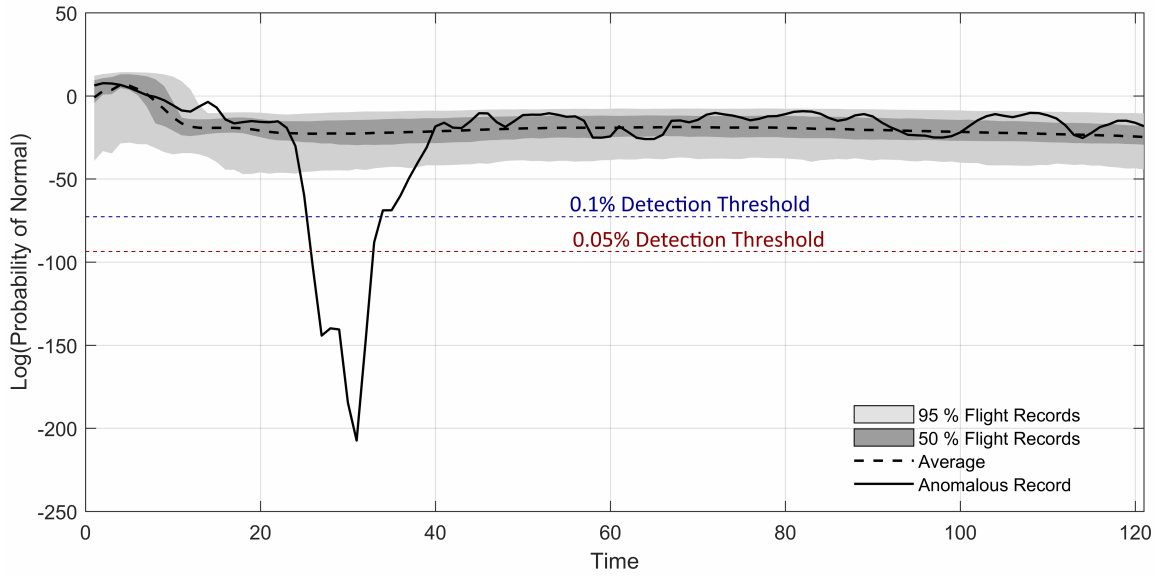


Figure 132: Variation of probability density during take-off for flight with instantaneous anomaly only

In this subsection, a flight with instantaneous anomaly in take-off but no flight-level anomaly is described. The probability density for the flight record in this phase of flight is shown in Figure 132. There is a precipitous drop in the probability in the initial stages of the take-off which is identified as an instantaneous anomaly under both

thresholds shown in the figure. Further investigation of the anomaly can be performed by visualizing energy metric and raw parameter variations in take-off (Figure 133 and 134).

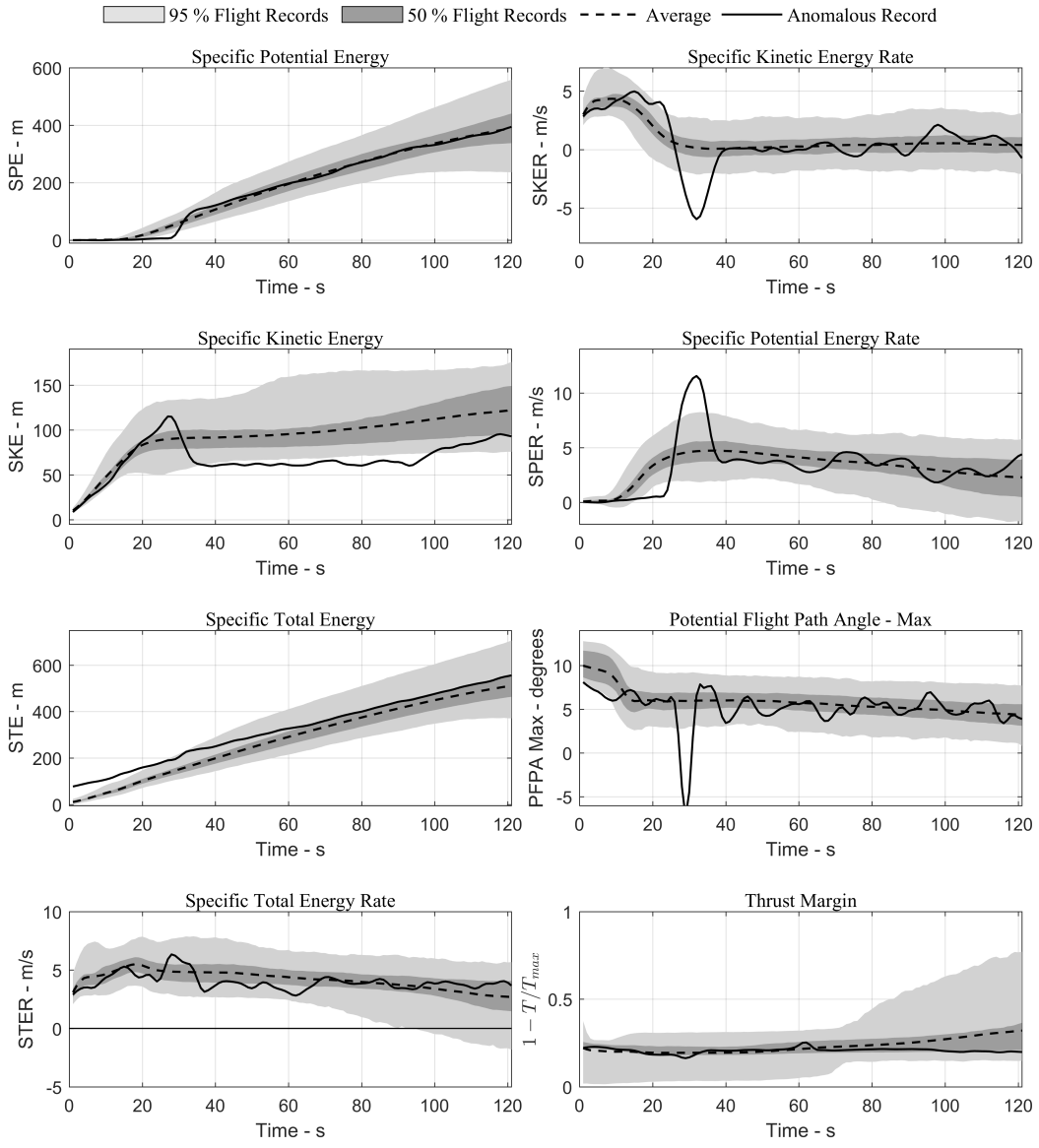


Figure 133: Variation of energy metrics during take-off for flight with instantaneous anomaly only

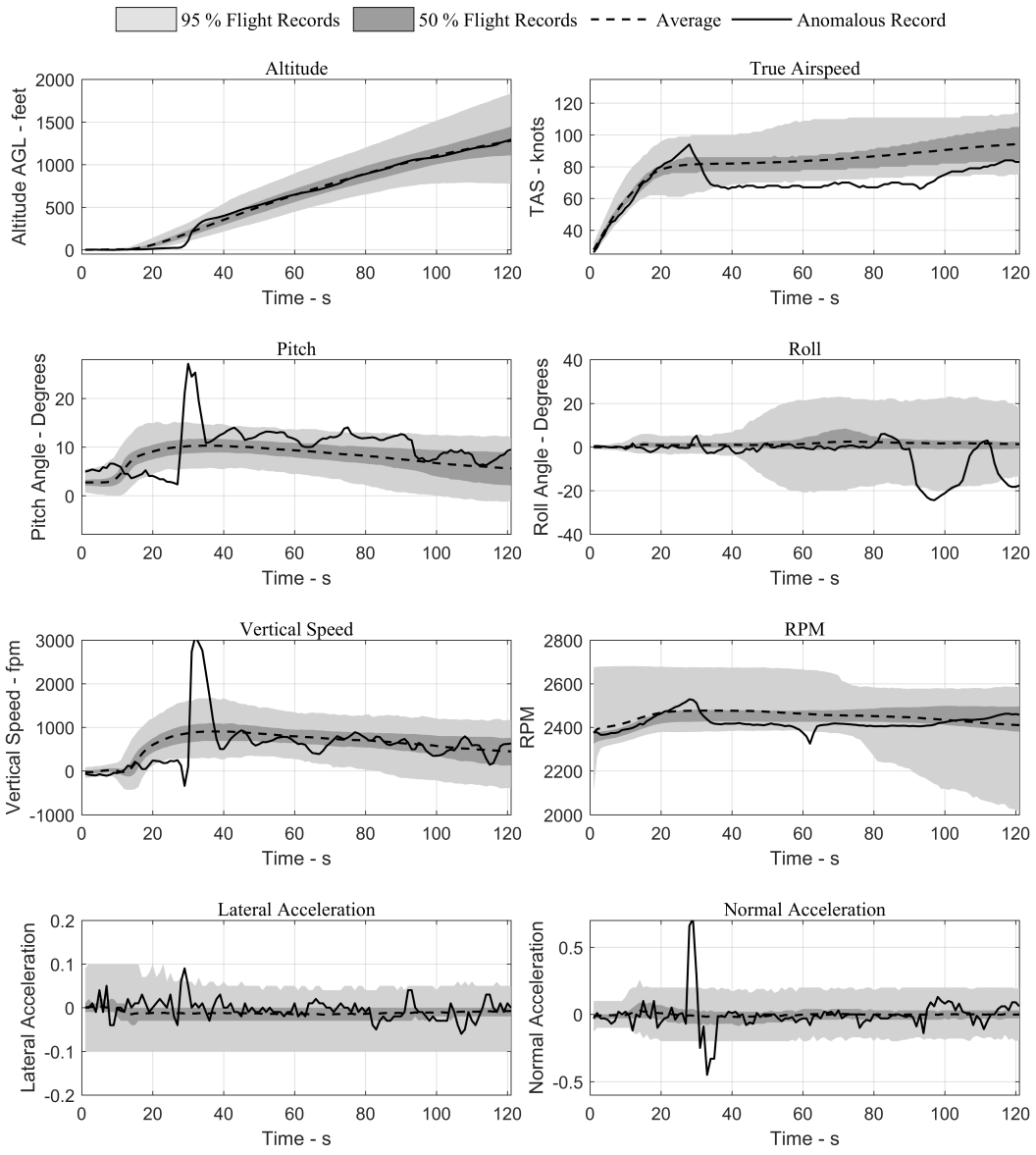


Figure 134: Variation of raw parameters during take-off for flight with instantaneous anomaly only

The energy metric variation indicates that at the location of the anomalous window, there is a steep drop in kinetic energy and a high positive potential energy rate. These magnitudes are unusual for this stage of the take-off where both potential and

kinetic energy rates are usually positive. The variation of other metrics during the approach and the anomalous window are within nominal bounds of the data set. The raw parameter variation indicates a very high pitch attitude and vertical speed as well as normal acceleration well outside normal bounds at the location of the instantaneous anomaly.

7.2.2.2 Flight-level Anomaly Only

The next anomaly considered in take-off phase is a flight with flight-level anomaly only and not any instantaneous anomalous windows. The variation of probability density for this flight can be seen in Figure 135. As is evident from the figure, the instantaneous probability, while lower and outside of nominal bounds for the initial part of the take-off does not fall below the thresholds for this data set at any point. However, the variation of energy metrics in Figure 136 and raw parameters in Figure 137 provides the reasoning for why this flight is still identified as a flight-level anomaly.

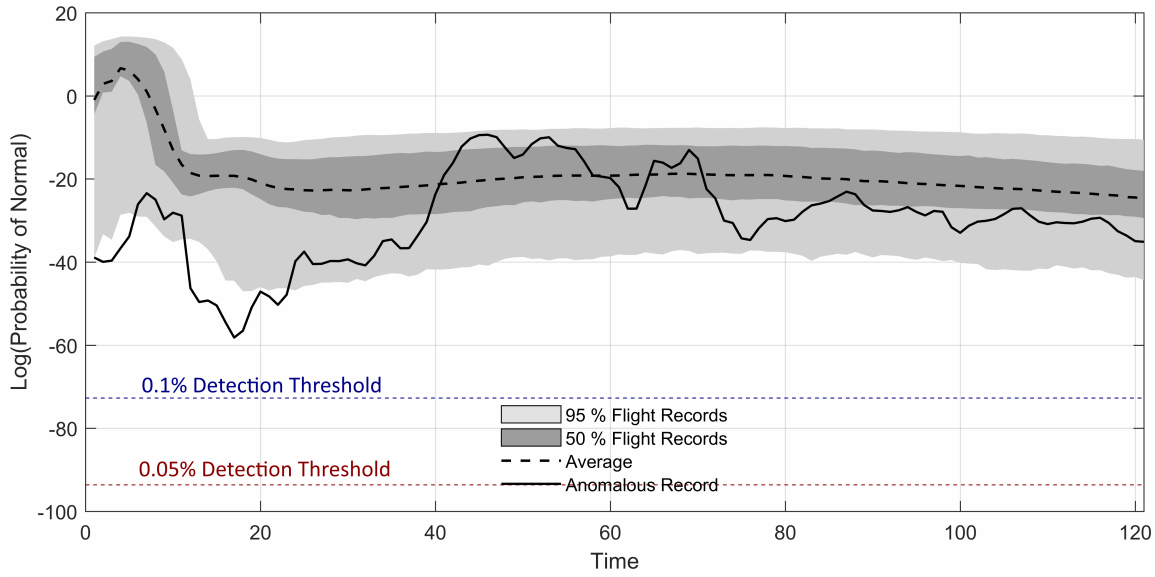


Figure 135: Variation of probability density during take-off for a flight with flight-level anomaly only

The kinetic and potential energies and their rates during the initial parts of the

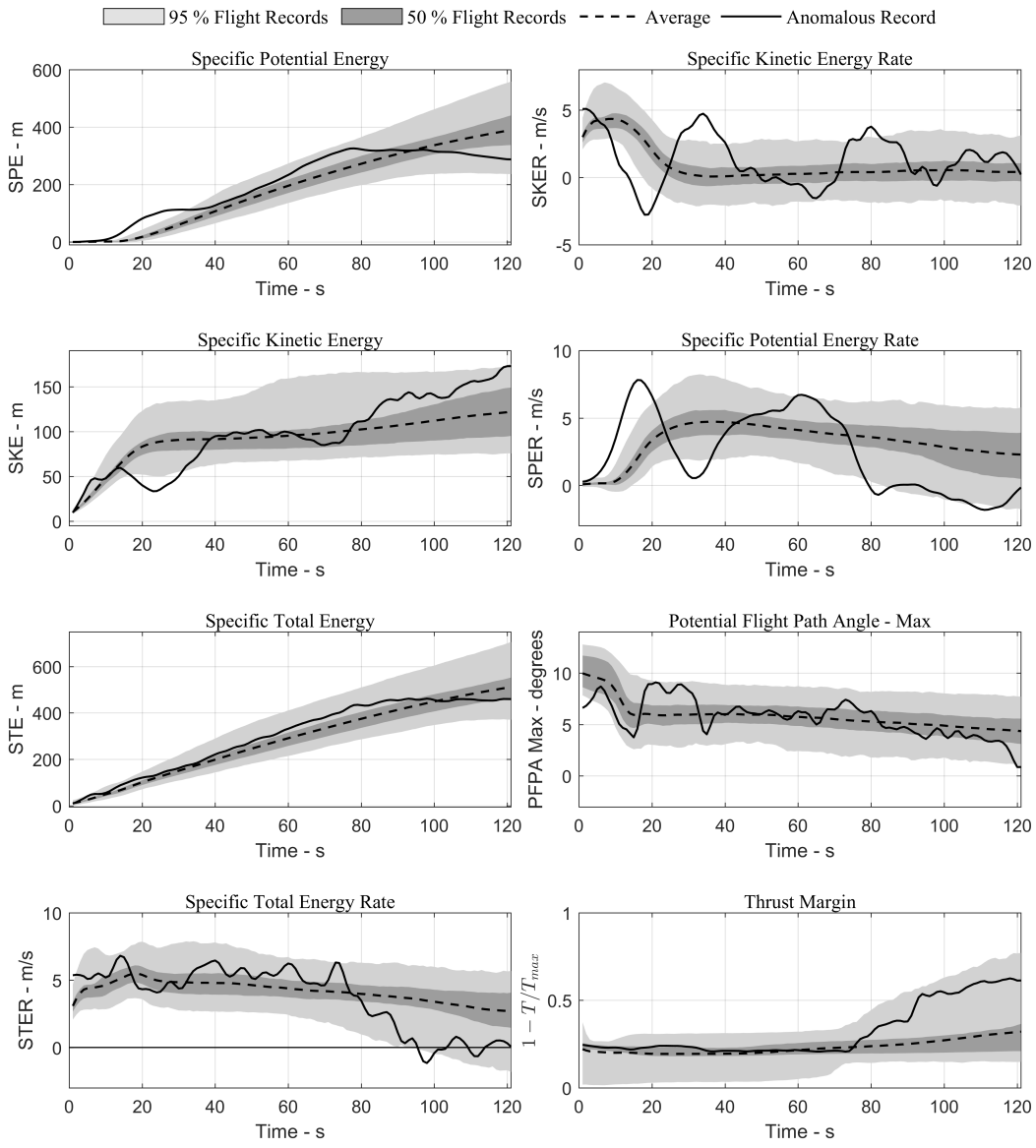


Figure 136: Variation of energy metrics during take-off for a flight with flight-level anomaly only

take-off are outside nominal bounds and keep fluctuating a more than usually observed. This results in an unusual kinetic energy profile for the intitial 40 seconds or so after the application of take-off power while the airplane is executing the take-off.

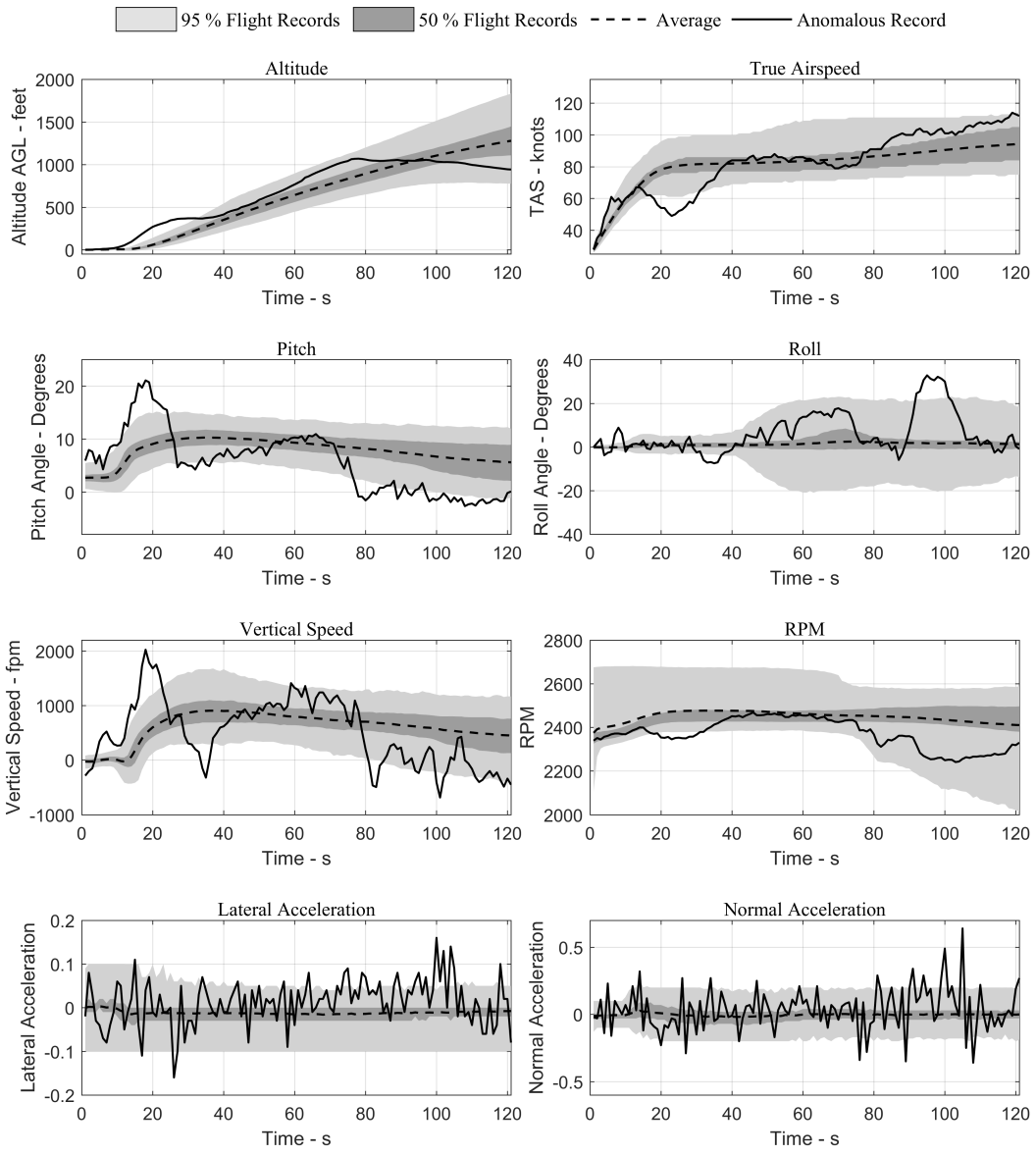


Figure 137: Variation of raw parameters during take-off for a flight with flight-level anomaly only

Variation of pitch angle and vertical speed in this initial segment also deviate from normal bounds. It is of interest to note that both lateral and normal accelerations are noisier and have higher magnitudes for this take-off indicating the possibility of

rough weather conditions which might have caused the flight to deviate.

7.2.2.3 Flight-level and Instantaneous Anomaly

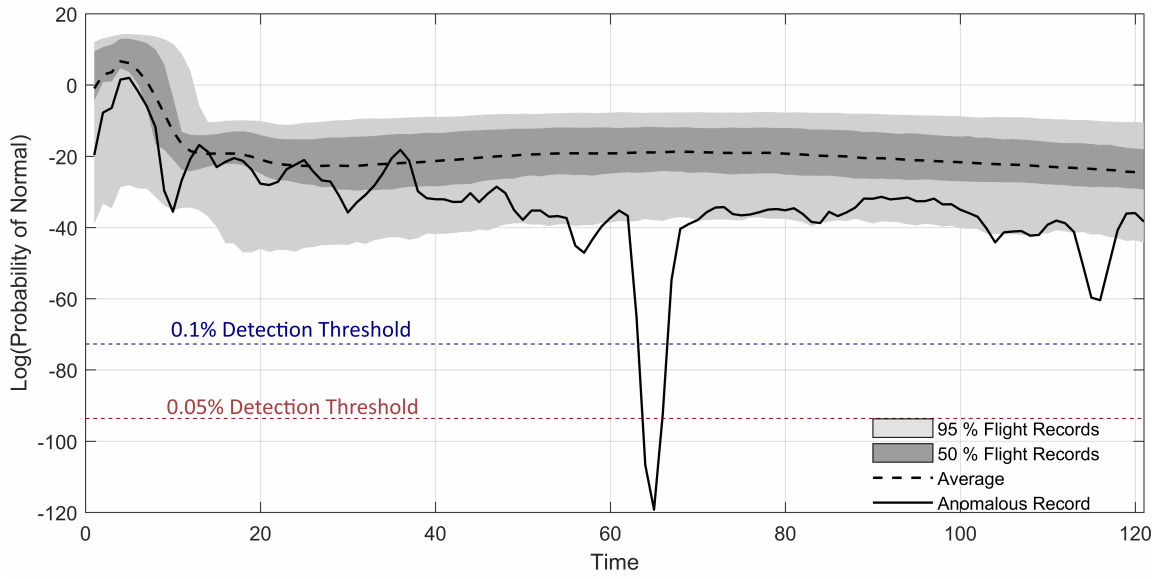


Figure 138: Variation of probability density during take-off for a flight with flight-level and instantaneous anomaly

The final anomaly considered in this subsection is a flight with both flight-level and instantaneous anomalies. The variation of probability density for this flight is demonstrated in Figure 138. The instantaneous anomalous window occurs near 60 seconds into the flight. The variation of energy metrics can be seen in Figure 139. As is evident from the metrics variation, the flight begins to deviate from the nominal take-off at the location of the instantaneous anomaly window and does not return to the nominal region all the way through to the end of take-off which results in it being identified among flight-level anomalies as well. The sudden drop of specific total energy rate at the beginning of the instantaneous anomaly below a value of zero along with a simultaneous high thrust margin indicates that the aircraft is producing much more drag than thrust and the thrust that is produced is nowhere near maximum. Kinetic energy drops more rapidly than potential energy after the event indicating that high pitch might be maintained in order to not lose altitude at such an early

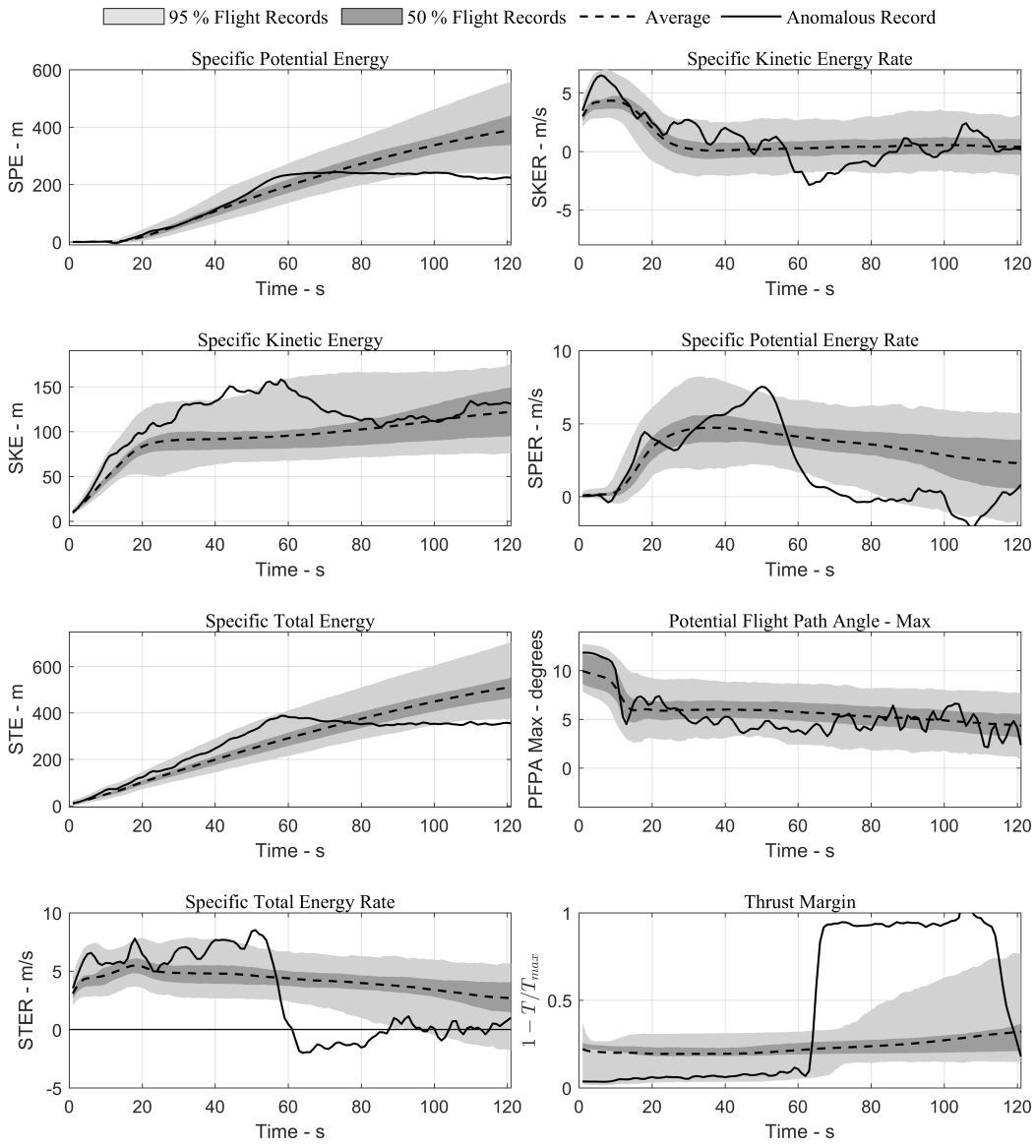


Figure 139: Variation of energy metrics during take-off for a flight with flight-level and instantaneous anomaly

stage in take-off. Upon examination of raw parameters in Figure 140 it is evident that the power is cut-off (possibly to simulate engine failure and capture best glide speed) before it is restored towards the end of the two minutes. While the power

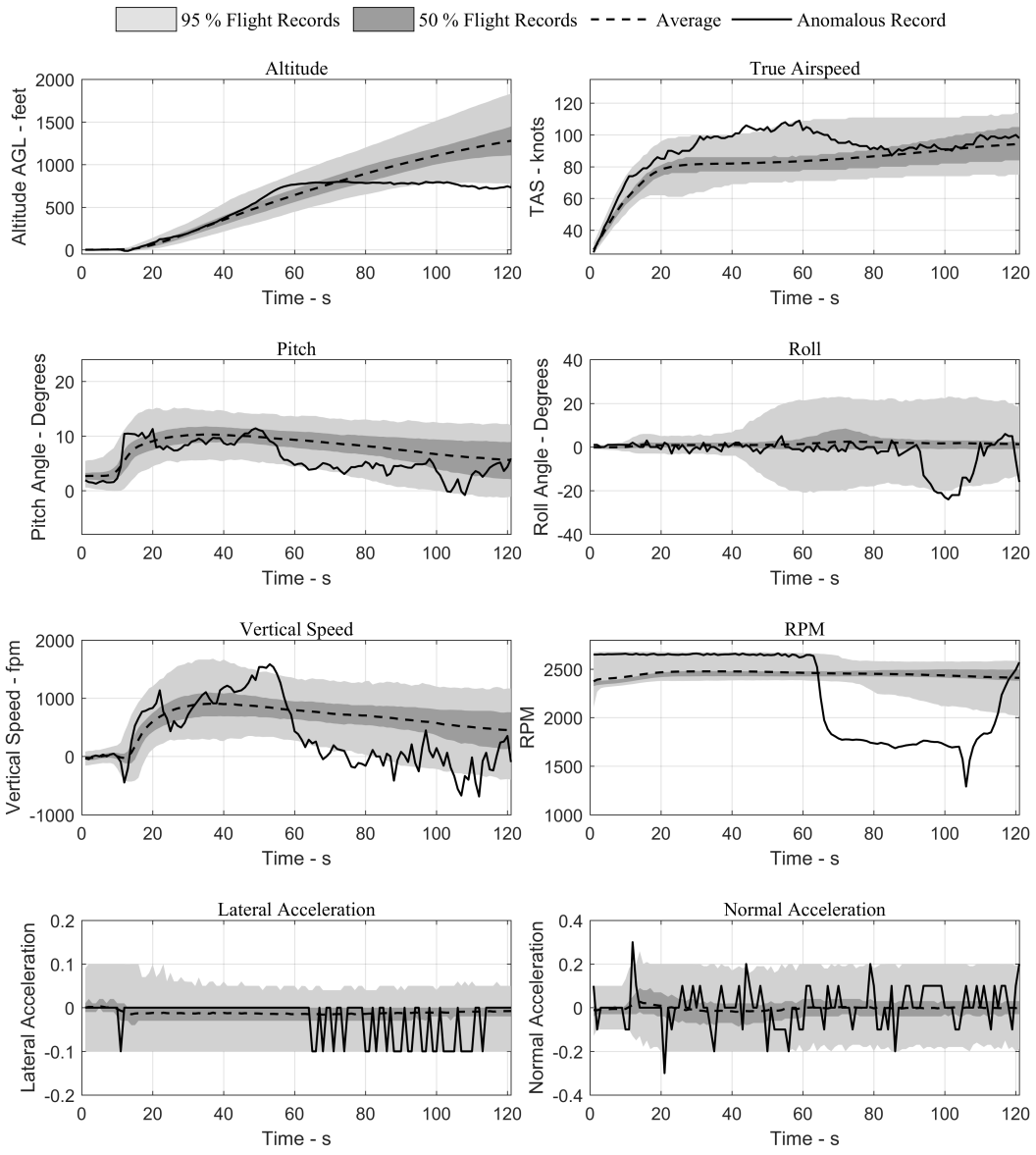


Figure 140: Variation of raw parameters during take-off for a flight with flight-level and instantaneous anomaly

gets restored towards the end, if indeed a training exercise, the fact that a simulated unsafe situation in flight is captured as anomalous with both types of anomalies is encouraging for the applicability of the algorithm.

7.2.3 Summary

This section has presented additional examples of the different types of anomalies and their metric visualizations. Flight records with only one type of anomaly but not the other as well as those with both types of anomalies are presented. While it is not possible to definitely say whether any of the anomalies pose flight safety risks without subject matter expert review, indication of the anomaly severity and probability density definitely provide some guidance. Flight records with both types of anomalies and severe scores definitely warrant being examined before other types when there is no other information available. While some anomalies are possibly benign, some might point to a safety event or simulation of safety event or proximity to unsafe conditions.

As noted earlier in the development of the methodology, take-off phase being more uniform exhibits the possibility of lower number of actual anomalies. Since the algorithm identifies anomalies as a proportion or percentage, there are bound to be some anomalies that are benign. The review by subject matter expert or flight instructors of students can address these issues. It also allows dynamic thresholds to be placed by subject-matter-experts which can enable striking a balance between missed detection (too stringent thresholds) and excessively high anomalies (too loose thresholds).

CHAPTER VIII

CONCLUSION

This chapter concludes the dissertation with a summary of conclusions from the investigations performed, a summary of the key contributions, and identification of possible avenues for future work that can draw upon the work from this dissertation. During the course of the development of the methodology, a series of experiments was performed to test and verify capabilities of the methodology. These experiments helped answer the research questions raised and to bolster the arguments in favor of the methodology. The experiments and salient conclusions drawn are summarized here. The first experiment was conducted in the form of a literature survey on metrics that could be used in a retrospective safety analysis setting. As a part of this experiment, various existing criteria for assessing flight safety, historical incidents and accidents, and parameter recording capabilities in GA were explored. The appropriateness of energy-based metrics as having the necessary qualities of parsimony, safety relevance, and generalizability was demonstrated. Similarly, adaptation of metrics used in other settings as well as definition of new metrics was provided.

As a part of experiment 2.1, the automatic identification of flight-level anomalies in take-off, approach, and landing phase using energy metrics was demonstrated. The reparametrization and alignment of flight records in both phases allows for comparison of different flights at the same (or similar) instants during their respective phases. The SVM model developed provided a natural boundary between normal and anomalous flights. Additionally, a score which indicates the severity of the anomaly was also provided by the SVM model developed. The identification of known unsafe flight data records was verified using flight data generated from a flight simulation model.

The demonstration of sample anomalies in the take-off and approach phases enabled focusing safety assessment and training efforts on the important flights when a large amount of flight data is available. Explanation of specific anomalies identified and their potential relation to unsafe or unusual events helped understand the power of anomaly detection techniques.

As part of experiment 2.2, the automatic identification of instantaneous anomalies in take-off, approach, and landing phase using energy metrics was demonstrated. A novel-sliding window based pre-processing technique was used in conjunction with a gaussian mixture model that allowed multiple standard operations. The probability density function provided a continuous metric for quantifying the normality of each point or window within the data set and also enabled focusing on regions with low probability. Demonstration of a set of instantaneous anomalies in take off and approach phases was made along with the possible reasoning for the behavior of the flight data record. Both the experiments 2.1 and 2.2 were demonstrated using data from multiple aircraft at different airports.

As part of experiment 2.3, the generalizability of the anomaly detection techniques was tested, specifically for flight-level anomaly detection. The first scenario tested was decreasing amount of flight data parameters being available for energy metric evaluation. The results revealed that the algorithms are able to identify majority of the most severe anomalies despite limited data availability. The second scenario tested the ability of the trained SVM model to identify anomalies in data not used for training the model. The results indicated that the model generalized very well and using even a small proportion of data for training is enough to identify majority of the same anomalous flights.

As part of experiment 3.1, the calibration of performance models from a generic GA aircraft to a specific aircraft was demonstrated using only publicly available data

from the POH. The accuracy of the developed models was tested on a set of validation flights and revealed that the models calibrated for both Cessna 172 and Piper Archer were usable in other parts of the dissertation. As part of experiment 3.2, calibration of performance models using limited GA flight data was demonstrated. A novel method of calibration using the specific total energy rate metric was proposed and implemented. The calibration was able to produce models that could predict the energy metric with increased accuracy despite unknown take-off weight of the aircraft. The calibrated models were used in the previous experiments for inserting the simulated anomalous flight data record in the data set for validation.

As part of the case study, different aspects of using the developed methodology in a real-world setting were demonstrated. The relationship between flight-level and instantaneous anomalies is explored in both phases of flight. It was found that flight-level anomalies have a high likelihood of also containing instantaneous anomalies but the converse relationship is not as strong. Evaluation of exceedance detection revealed that with the current defintions, exceedances occur often, even in routine flights. Comparing anomalous flights with exceedance detection revealed that the number (or percentage) of exceedances is typically higher for flight-level anomalies than normal flights, with the difference being more pronounced for level-2 anomalies. Similarly, exploring the relationship between instantaneous anomalies and exceedances revealed that points with exceedances have a lower probability of being normal than points with no exceedances. Segmenting the data set based on operational differences in two different cases presented different results. The distinction between straight-in approaches and rectangular pattern approaches was not significant from the proportion of anomalies in each type of approach. However, segmenting the data set according to flaps usage (full flaps, partial flaps, and no flaps) during final approach and landing revealed that flights with lower flaps usage had a higher proportion of anomalies according to the algorithm. Finally, the performance of normal and

anomalous flights on the touchdown metrics revealed a difference in performance for some metrics (airspeed, touchdown distance) but no significant difference for others (vertical speed, angle of attack).

In summary, these experiments revealed that the developed methodology was capable of providing safety assessment using quantitative data-driven techniques a level of detail deemed acceptable for the specific application in GA, and thus partially or completely supported all hypotheses associated with the three major research questions.

8.1 Contributions

The primary contribution of this dissertation is the creation and demonstration of a quantitative, data-driven methodology for retrospective safety assessment. In particular it allows rapid processing of large amounts of flight data to automatically identify different types of abnormal or anomalous operations within the data. A salient feature of the demonstrated approach is that the required initial input from processing large amounts of data is limited to the thresholds for anomaly detection. This moves away from traditional methods requiring specific event or exceedance definitions for each type of aircraft in the data set. This ensures that the processing time gets reduced significantly as the analyst, instructor, or user now has to focus a specific set of important flights or anomalies rather than the entire data set.

The first among the specific contributions within the developed environment is the usage of energy-based metrics in a retrospective safety analysis setting. These metrics have previously been used for performance assessments and design studies, but their use in a safety setting has not been widespread. The adaptation of existing metrics and definition of some completely new metrics is one of the important contributions of this work. With an array of metrics available at the disposal of the analyst or user, different types of insights can be gained. The usage of these metrics for retrospective

safety analysis explicitly addresses poor energy management and loss of energy-state awareness which are among top contributors to accidents and incidents. Similarly, the defined metrics satisfy the desirable qualities of parsimony, safety-relevance, and generalizability for safety metrics. Applicability of metrics across different aircraft and heterogeneous flight data recorder capabilities ensures their usefulness in a diverse data set.

The second set of contributions from this dissertation is the techniques for automatically identifying and isolating different types of anomalies. One of the limitations of existing approaches identified earlier was that a priori definitions are required to identify exceedances or events. With the availability of large quantities of flight data, the shift away from a priori definitions to data-driven thresholds is possible. However, in order to successfully facilitate this transition, techniques from different domains need to be adapted for GA operational data. For identification of flight-level anomalies, a two-step approach using density-based clustering and one-class support vector machine is developed. For identification of instantaneous anomalies a sliding-window based pre-processing technique is used in conjunction with gaussian mixture model for anomaly detection. Both these approaches are developed within a general anomaly detection framework. Demonstration of anomaly detection using multiple aircraft and airport of operations in the same framework is one of the unique contributions of the current dissertation. The use of performance models to simulate a few known anomalous flight data records is demonstrated, which facilitates validation of the developed methodology. Similarly, through both anomaly detection techniques, the entire data set is utilized in making decisions on outliers. This enables moving away from a priori static definitions of events or exceedances.

The third set of contributions from this dissertation is related to the calibration and usage of performance models in the safety assessment methodology. Literature survey revealed that incorporating additional information in the safety analysis task

can make the analysis richer and more meaningful. Performance models can be used to evaluate energy metrics, simulate flight data records containing anomalies or known unsafe operations, and used for a variety of other purposes. In the current dissertation, starting from a generic GA performance model, calibrated models for specific aircraft are obtained. A novel 2-level framework is demonstrated with each level corresponding to different quality of calibration data. Calibration to publicly available data using the Pilot Operating Handbook is demonstrated in level-1 calibration, whereas level-2 calibration uses limited amount of flight data records and a novel method of calibration using the STER energy metric.

In summary, the demonstrated approach fills a hitherto existing gap by facilitating the rapid investigation, isolation, and visualization of abnormal flight data records from within a large set of heterogeneous operations in a single unified environment. The use of quantitative energy-based metrics along with predictive capabilities from performance models enables gaining insights and comparisons that would not be possible with traditional exceedance detection techniques. The developed framework can be deployed as a retrospective safety assessment tool or as a flight training tool to help pilots understand and improve their piloting techniques during critical phases of flight.

8.2 Recommendations for Future Work

Improvement of safety in GA operations is an ongoing process. Thus, even though the hypotheses tested in the current work have been able to answer the identified research questions and fulfill the research objective outlined at the beginning, there are ways in which the capability of the developed methodology can be extended further. Some of these are identified here. There are two main stream of thought in which the capabilities presented here can be extended. The first is related to the limitations of the work presented in this thesis. These limitations could be in terms

of data recording capability, model calibration framework, anomaly detection, and other areas. The second stream of potential future work is related to extending the developed methodology to problems in other domains. Each of these are explored in this section.

While the data used in this dissertation comes from different aircraft, it is recorded using a similar data recorder (G1000). The generalization experiment addressed how the method would be impacted in the presence of limited parameters from within the same flight logs. However, it is also of interest to investigate how data recorded at different frequencies and latency can be integrated into the methodology. To that end, flight data collected from different devices such as PED, iPad, etc. can be integrated into the data set. This will truly enable addressing the limitations of the developed method. While majority of the accidents pertained to single engine fixed gear aircraft, data from other types of aircraft can also be utilized and integrated in the same manner as the current data set. Model calibration on the propulsion side can be extended from the current version of single engine naturally-aspirated propeller engine to other types in order to capture an even wider spectrum of the heterogeneous fleet. While the aim and scope of this dissertation is limited to retrospective analysis from the point of view of training and safety, extensions of the methodology can be envisioned for real-time assessment where pilots could get feedback when the flight is drifting into an anomalous state. The computational time requirements for the different steps are low and they are so chosen in order to enable rapid analysis. The potential advantages of using the data set available to train an offline model that can be queried online is worth exploring and can directly feed off of the work presented here. This would require additional efforts to reduce computational time as well as store standard operations as a surrogate model that can be quickly queried in-flight.

With regard to extension of this methodology to other domains, it was noted earlier that techniques that generalize well within the GA fleet are sought in this

work. Nevertheless, energy-based metrics are an inherent property of any aircraft and as such, this methodology could include other categories of aircraft. In particular, the same methodology and metrics can be used for commercial aircraft or others. Of particular interest would be how the metrics can be extended for novel concepts such as electric aircraft within the same overall framework. Since aircraft performance is such a deep-rooted part of this methodology, other potential uses of this method in the design phase can be envisioned in the long term. With the recent push by the FAA towards reworking the certification standards for small aircraft to performance-based standards, the safety analysis pursued here can be of additional relevance in the future.

APPENDIX A

IMPLEMENTED ENERGY METRICS

Table 11: Summary of implemented energy metrics and data required for computation

		Can be estimated using		
Metric	**Formula**	**Flight Data (FD)**	**FD + Ref. Profile**	**FD + Perf. Model**
STE	$h + V^2/2g$	✓	✓	✓
SPE	h	✓	✓	✓
SKE	$V^2/2g$	✓	✓	✓
STER	$\dot{h} + V\dot{V}/g = (T - D)V/W$	✓	✓	✓
SPER	$\dot{h} = V \sin\gamma$	✓	✓	✓
SKER	$V\dot{V}/g$	✓	✓	✓
PFPA	$\gamma + \dot{V}/g$	✓	✓	✓
ERD	$sign(\frac{SKER}{SPER}) \times \exp(-\lvert\frac{SKER}{SPER}\rvert)$	✓	✓	✓
STEE	$h_{act} - h_{ref} + (V_{act}^2 - V_{ref}^2)/2g$	×	✓	×
SPEE	$h_{act} - h_{ref}$	×	✓	×
SKEE	$(V_{act}^2 - V_{ref}^2)/2g$	×	✓	×
NSEE	$((STE)_{act} - (STE)_{ref})/(STE)_{tol}$	×	✓	×
STEER	$sign(STEE) \times \delta(STEE)/\delta t$	×	✓	×
IERE	$V_{act}(T - D)/V_{red}W(\gamma_{ref} + \dot{V_{ref}}/g)$	×	✓	×
Max. PFPA	$T_{max} - D/W$	×	×	✓
Min. PFPA	$T_{idle} - D/W$	×	×	✓
TM	$1 - T/T_{max}$	×	×	✓
ERM	$(W(\gamma_a + \dot{V_a}/g))/(T_{max} - D)$	×	×	✓
ERDm	$W(\gamma_c + \dot{V_c}/g)/(T_{idle} - D)$	×	×	✓

APPENDIX B

LIST OF EXCEEDANCE EVENTS

No	Event	Level 1	Level 2
1	V_{NE} (Never Exceed Velocity)	158 knots	163 knots
2	Vertical g Load	3.0	3.8
3	Vertical g Load - Min	-1.0	-1.52
4	Oil Temperature - Max	–	245 F
5	Oil Pressure - Min	–	20 psi
6	Oil Pressure - Max	–	115 psi
7	Max RPM	$\geq$ 2700 1s	$>$ 2700 5s
8	Max Cylinder Head Temperature	–	500 F
9	Fuel Quantity - Min	8 gal.	5 gal.
10	$V_{FE} - 10°$	108 kt	110 kt.
11	$V_{FE} > 10°$	84 kt	85 kt.
12	Bank Angle	60°	$\geq$ 65°
13	Bank Angle (below 1300 AGL)	50°	$\geq$ 55°
14	Pitch Attitude (positive)	30°	$\geq$ 35°
15	Pitch Attitude (negative)	-30°	$\leq$ -30°
16	Vertical Speed Magnitude Below 1000 AGL	$\geq$ 800 fpm	$\geq$ 1000 fpm
17	Airspeed $\leq$ 200 feet AGL High - Full Flaps	66 kt 2s	71 kt 2s
18	Airspeed $\leq$ 200 feet AGL High - Zero Flaps	75 kt 2s	80 kt 2s
19	Airspeed $\leq$200 feet AGL Low - Full Flaps	60 kt 2s	$\leq$56 kt 1s
20	Airspeed $\leq$ 200 feet AGL Low - Zero Flaps	69 kt 2s	$\leq$65 kt 2s
21	Extended Center-line deviation at 200 feet AGL	2°	3°
22	Glide angle High (Too steep) at 200 feet AGL	4°	5°
23	Glide angle Low (Too shallow) at 200 feet AGL	2°	1°
24	Bank Angle at or below 200 feet AGL	20°	25°
25	Pitch Attitude at Touchdown (High)	10.5°	12°
26	Pitch Attitude at Touchdown (Low)	3°	1°
27	Airspeed at Touchdown (High - Full Flap)	55 knots	60 knots
28	Airspeed at Touchdown (High - No Flap)	63 knots	68 knots

Table 12: Exceedances set for Cessna 172 aircraft in approach and landing [73]

The following table lists exceedances for take-off phase. Those that are common for all phases and listed in Table 12 are excluded.

No	Event	Level 1	Level 2
1	Heading Variation at Power Application	10°	20°
2	Low RPM Rotation	2200	2000
3	Airspeed Lift-off (Non-soft)	44 kt	40 kt
4	Pitch at Lift-off	10.5	12
5	Bank Angle (Take-off)	20	25
6	Airspeed between 100 and 500 ft	57 kt	52 kt
7	Bank Angle < 400 ft	30°	45°
8	Altitude Drop < 500 ft	< 0 fpm	4 < -200 fpm

Table 13: Sample exceedances set for a Cessna 172 aircraft in take-off and climb [73]

APPENDIX C

DERIVED PARAMETERS

This appendix contains the derivation of additional parameters which are used throughout the dissertation. Some of the techniques for deriving required parameters are adapted from existing literature whereas some are developed as a part of this dissertation. The section for each parameter contains details for techniques developed in this dissertation and a summary and reference for techniques adapted from existing literature.

No	Derived Quantity
1	Phases of Flight
2	Touchdown Point
3	Nominal/Reference Profiles
4	Glide Slope
5	Angle of Attack
6	Weight
7	Rates
8	Flap Setting

Table 14: List of parameters derived from basic flight data using different techniques

C.1 Phases of Flight

In GA operations, defining phases of flight may not be as straightforward as commercial operations. For example, pilot training involves maneuvers such as a "touch-and-go" for landing practice. In such maneuvers it is difficult to define a distinct cruise phase while the airplane is in the pattern. Similarly, most GA operations involve higher variability in parameters even during the same phase than commercial operations. Some of the reasons for this include less experienced pilots, smaller aircraft

with lower wing-loading being more susceptible to gusts, etc. However, identifying phases of flight can be of importance as it can help understand safety events, hazards, and operations associated with higher risk. Goblet et al. [68] have provided a simple classification of GA flight data into different phases based on commonly-recorded parameters and adapting prior definitions by NTSB and CAST. The definitions provided there for take-off, climb, descent, approach, and landing are used in this dissertation. The parameter values have been reproduced here:

1. **Take-off:** Engine RPM (>2500 RPM) and Altitude (<35 ft above the runway elevation)
2. **Climb:** Change in Altitude (constant increase > 200 feet per minute)
3. **Descent:** Change in Altitude (constant decrease < 200 feet per minute)
4. **Approach:** Altitude (constant decrease < 200 feet per minute) and Altitude (< 1000 ft above runway elevation)
5. **Landing:** Indicated airspeed (< 60 KIAS) and Altitude (constant decrease < 200 feet per minute for one minute)

C.2 Touchdown Point and Reference Profiles

The phases of flight identified in the previous section help to narrow down the focus of the flight data record on the particular phase of interest. However, an important piece of information typically missing in GA flight data records is the approximate touchdown point of the aircraft on the runway during the landing phase. In commercial aircraft, a binary parameter called Weight-on-Wheels (WOW) is indicative of the touchdown point but it is not typically recorded in GA flight records. This parameter is important as flight data records can be anchored at this point and re-sampled to facilitate comparison of different flights with each other as well as to obtain reference

or nominal variations of parameters during the approach and landing phase. This section describes the efforts to obtain these pieces of information using manipulation of recorded parameters and a large data-set.[1]

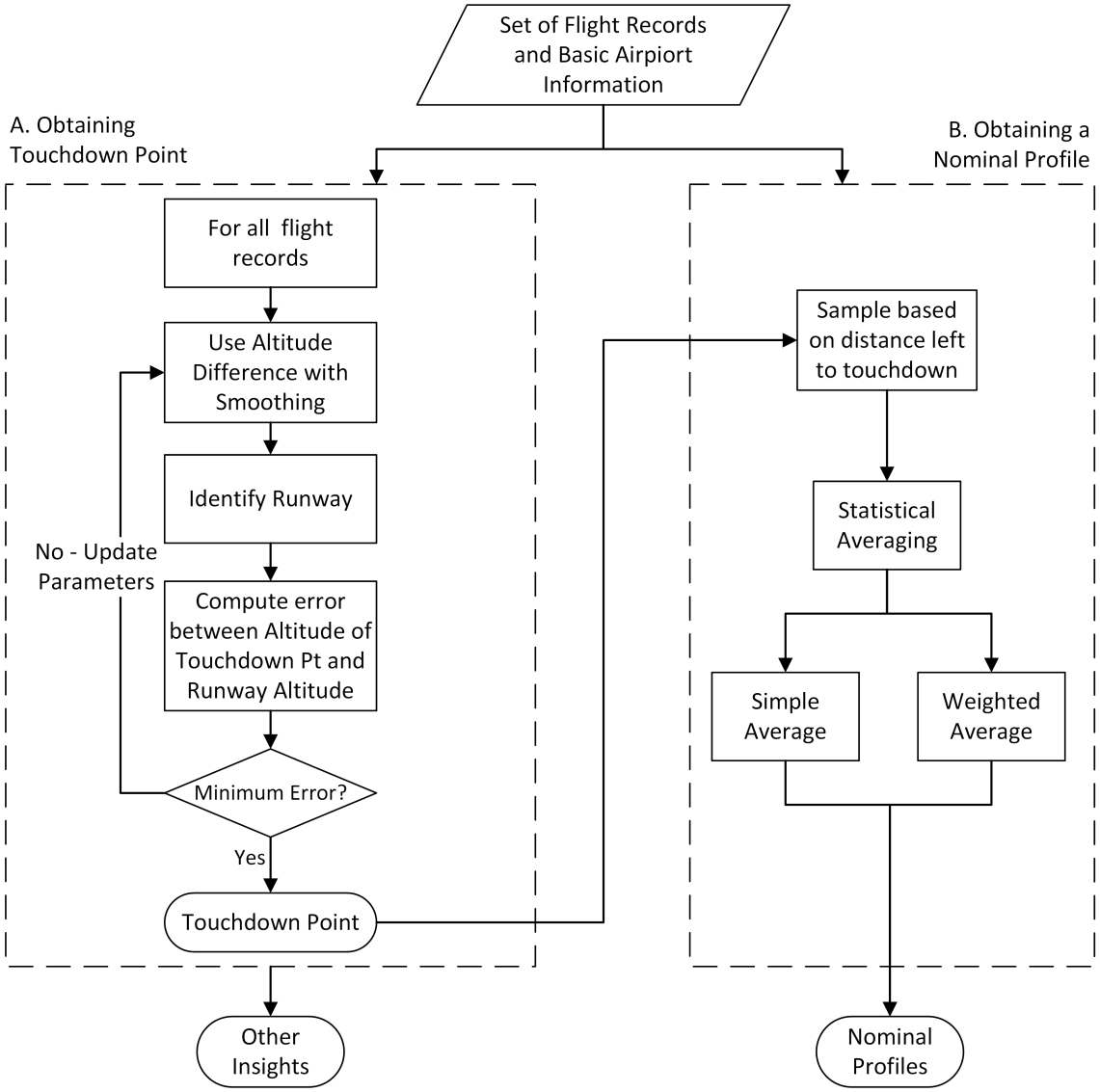


Figure 141: Obtaining touchdown point and reference profiles from flight data

The methodology consists of two key analyses - obtaining the touchdown point

[1]The research described in this section is also documented in the following publication:

– **Puranik, T.G**, Harrison, E., Min, S., Jimenez, H., and Mavris, D., *General Aviation Approach and Landing Analysis using Flight Data Records*, in 16th AIAA Aviation Technology, Integration, and Operations Conference, 2016. Paper No. AIAA 2016-3913, doi:10.2514/6.2016-3913 [118]

and obtaining a nominal profile. The entire process is outlined in Figure 141 which shows the inputs, outputs, and steps involved in each analysis. During analysis, the flight data is typically anchored at a specific event in time and data from all the flights is sampled at a fixed temporal or distance-based intervals from this point. This makes the flight data across a large number of flights comparable to each other. In approach and landing operations, this is typically the touchdown point on the runway. The flight data parameters (including altitude and velocity) can be backtracked from this point at a desired discretization and upto a specific distance. The flight data needs to be smoothed as it may contain noise in the parameters recorded. Obtaining the touchdown point involves calculating the root mean square (RMS) error of the touchdown altitude and reported runway altitude for a large set of flights. The set of parameters that minimize this RMS error over the entire data set are chosen as the final parameters and the touchdown point is evaluated for each flight. The flight data is also smoothed and the runway of landing identified in intermediate steps in this analysis.

The second analysis involves obtaining statistically averaged nominal profiles. For achieving this, each flight is sampled based on the distance remaining to the runway threshold. The nominal profile is then obtained by averaging the altitude and velocity across a large database of flight records at small intervals of distance from the runway threshold. Along with a simple average, weighted averages are also explored in this part. Similar to altitude and velocity, the reference profiles for any recorded or derived parameter can be obtained in the exact same manner.

C.2.1 Obtaining Touchdown Point

The flight data is smoothed using a local regression with weighted linear least squares in MATLAB (called "loess" smoothing). Details of the implementation can be found in the MATLAB documentation [98]. Obtaining the touchdown point accurately

requires two parameters to be tuned. The span or window of smoothing (hereafter called smoothing parameter) is one of the parameters used to reduce RMS error. The touchdown point for the current work is defined as the last point in the final approach beyond which the altitude difference between successive smoothed data points do not exceed a certain threshold (example one foot). This altitude threshold is the other parameter under our control to reduce the RMS error.

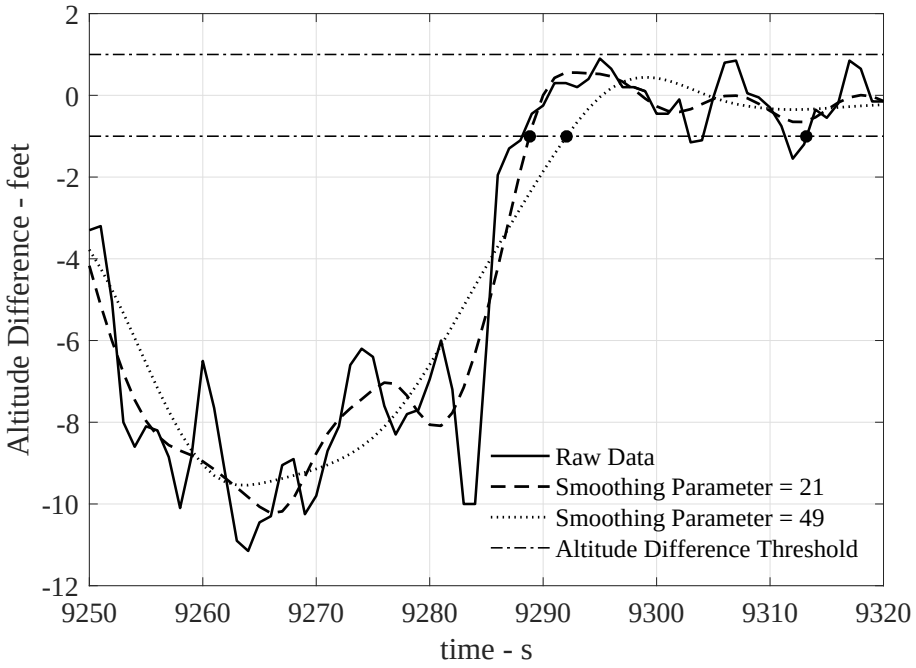


Figure 142: Comparison of different smoothing parameters and their effect on identification of touchdown point

Vertical speed or altitude difference can both be used as the threshold parameter. The reason for using the altitude difference as opposed to the vertical speed is the noise of the vertical speed data. Figure 142 shows the effect of the two tuning parameters on the touchdown point identification algorithm. The horizontal dashed lines represent the altitude threshold parameter. This parameter is used to specify the rule for maximum altitude difference permissible between successive points after the aircraft has touched down. The higher the value of the parameter, the wider is the gap between the two horizontal lines, which would imply detection of touchdown

point possibly earlier than the actual touchdown. If the value is too small, then the touchdown point would be detected too late (or not detected at all due to the noise in the data). Therefore, it is important to choose this parameter carefully. In the current work, this parameter is chosen as 1 ft based on an analysis described later in the section.

For the smoothing parameter, the solid line plots the altitude difference from the raw data collected from the flights. The altitude difference between successive points in the raw data does not strictly go to zero even after the aircraft has touched down on the runway. In some cases, the magnitude of the noise is as high as 7 feet even after the aircraft has landed. Formulating a general rule for hundreds of data records that may contain this kind of noise proves to be difficult and error-prone. Therefore, smoothing using local regression is utilized. The dashed and dotted curves shown in the Figure 142 indicate the effect of changing the smoothing parameter (the smoothing window) from 21 to 39. Increasing the smoothing parameter uses a larger window for smoothing and thereby some of the features are lost. On the other hand, a smaller smoothing parameter does not smooth the raw data sufficiently and may retain some of the noise that we want to eliminate from the original data. Therefore, this factor also plays an important role in implementing a robust strategy that can be used over a large set of flight records without tuning it separately for each flight record.

As seen in Figure 142, the touchdown points identified using the raw data and the different smoothing parameters with an altitude threshold of one foot are highlighted as black circular markers. It is quite clear that using the noisy raw data can lead to erroneous touchdown point identification as illustrated in the figure. Smoothing the data can prevent this error, but excessive smoothing might result in a touchdown point identification after the actual touchdown (Smoothing Parameter 21 versus 49 in the figure).

From a set of N flight data records, if the altitude of runway touchdown for flight number i is h_i and the corresponding runway altitude is $h_{r,i}$, then the touchdown altitude error for flight i is:

$$e_i = h_i - h_{r,i} \tag{32}$$

Using the error for each individual record, the total RMS error is given by Equation 33

$$RMS = \sqrt{\frac{\sum_{n=1}^{N} e_i^2}{N}} \tag{33}$$

To observe the effects of each parameter and to aid in the appropriate selection of particular values for each parameter, an experiment is performed in which each parameter is varied between certain lower and upper bounds and the RMS error is calculated.

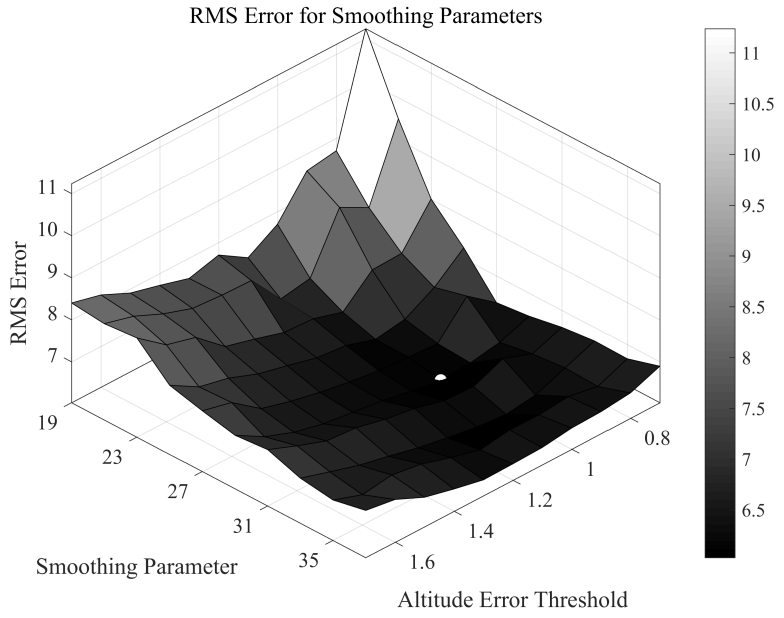


Figure 143: 3-D Surface plot of the RMS error for different smoothing parameters and thresholds

Setting the upper bound of the smoothing parameter too large would smooth the

data too much resulting in lower RMS error but result in touchdown points that are further along the ground roll than at actual touchdown. This experiment is performed for a large set of flight records, and the results are shown in Figure 143.

Figure 143 shows the trend of the RMS error on the z-axis with the two parameters on the x and y-axis. As we can see from the figure, low values of smoothing parameters generally result in higher errors as expected. This goes down as the smoothing is increased. Also, lower values of altitude threshold also tend to increase the RMS error. High values of altitude threshold may result in premature detection of touchdown point resulting in higher individual errors for each record. The chosen point (highlighted in the figure with a white dot), results in the smallest RMS error. This point corresponds to a parameter pair of altitude threshold = 1 and smoothing parameter = 27.

C.2.2 Obtaining Reference Approach Profiles

Once the touchdown point is obtained for all the flight data records, they are resampled according to the distance remaining to the touchdown point. It is important to discretize/sample the flights as a function of distance as this allows the generation of a statistical average (nominal) profile. The segment leading up to the touchdown point is obtained by using a spline interpolation of available data points at a discretization of 0.01 nautical miles. This facilitates efficient visualization of altitude (and other parameters) for each flight for when it was a certain distance away from the runway threshold.

Figure 144 shows this approach and landing data visualized for the current data set. In Figure 144, a reference 3° glide slope line (common practice for approach operations) from the "aim point" on the runway is shown by the solid black line. The dotted line shows the average altitude profile followed by all the flights. The shaded regions show the spread of the flight data records, with the dark grey denoting the

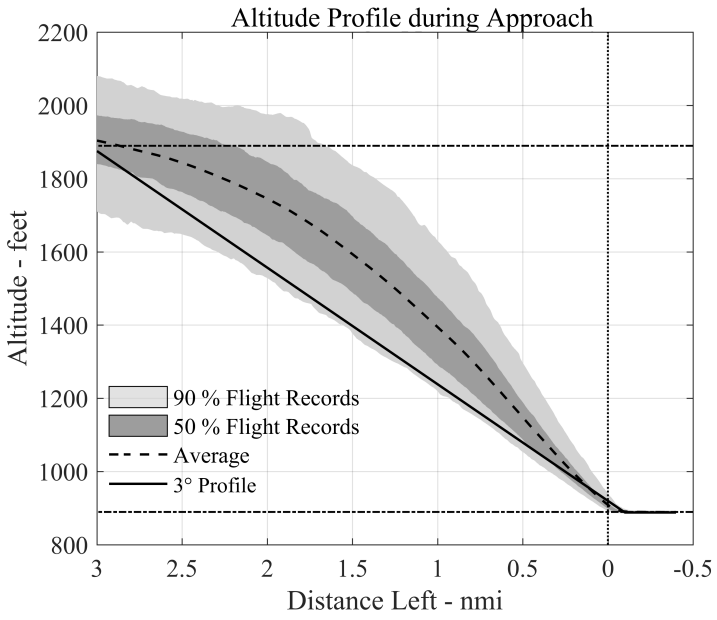


Figure 144: Nominal approach profile - Altitude

50th percentile and the light grey denoting the 90th percentile of data records. The dotted vertical line is the runway threshold and the horizontal line at approx 1900 feet represents the 1000 feet above ground level marker.

It can be seen that most of the flight records tend to intercept a 3° glide slope approximately one thousand feet above the airport/runway altitude (horizontal dashed line) as is typical for GA landing operations. This also corresponds to approximately 3 nautical miles out from the runway threshold. The average approach profile and most of the individual records tend to fly above the reference 3° glide slope line. But we can also see that there is much variation in the altitude profile of different flights as they approach and land at this runway. While some variation is expected, it is interesting to note that almost all (90th percentile) of the flights tend to fly above the 3° glide slope. Therefore, it is evident that in GA operations, the altitude profile followed during approach and landing is not necessarily a simple linear 3° slope line.

Therefore, the data presented here clearly indicates that during actual operations, defining a "reference" trajectory as simply along a 3° glide slope to the touchdown

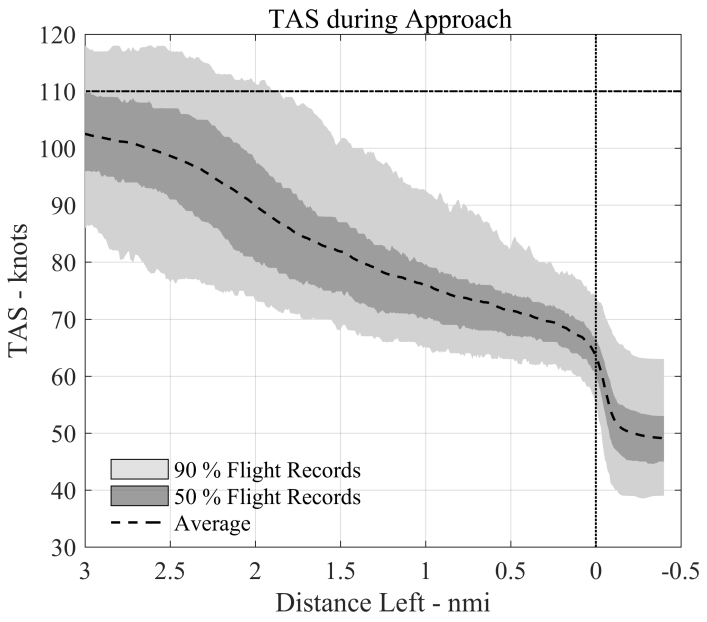


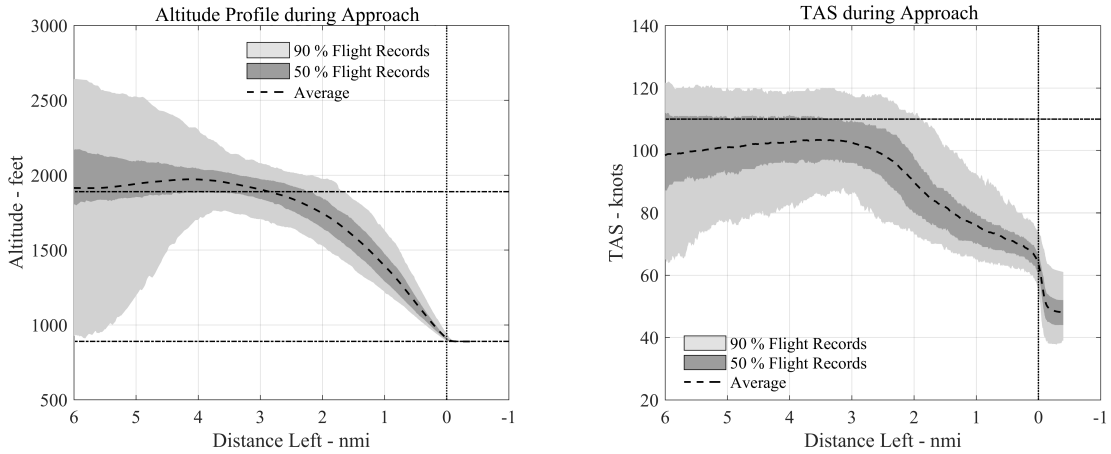
Figure 145: Nominal approach profile - Velocity

point would be an oversimplification of the actual process. For this reason, it is proposed to use or define the statistical average profile developed here as the nominal profile from an operational perspective.

Having already defined a nominal altitude profile, a similar approach can be taken to look at the velocity profile. Using the same set of flight records and the same technique of discretizaion and sampling, an average statistical profile for the velocity is obtained in Figure 145. In this figure, the horizontal line at 110 knots represents the "no flaps speed" for the current aircraft (obtained from Pilot Operating Handbook) [13]. The dotted line is the average and the shaded regions represent the 50 percentile and 90 percentile of the spread. Although there is no crisp reference to compare against such as the 3° profile for altitude, it is again evident that there is a significant drop in the velocity from the time of interception of the 1000 feet above ground level (AGL) line to the actual touchdown. The velocity profile during approach indicates that the average velocity profile drops from around 100 knots to 60 knots from the point of interception of 1000 feet AGL line to the runway threshold.

Also, there are no noticeable steps in the velocity profile as is typically recommended during pattern landings.

Once the nominal profiles for altitude and velocity are obtained, these are then extended beyond the 3 nautical mile distance to further distances to visualize and understand them better. This extension (upto 6 nautical miles distance) can be seen in Figure 146a and 146b. From the altitude profile it can be seen that beyond 3 nautical miles, the spread of flight data records increases quite a bit. This can be attributed to touch-and-go maneuvers that might be executed by the aircraft. Therefore, the nominal or average profile that is being defined only makes sense upto 3 nautical miles for final approach. Similar spread can be seen in the velocity profiles.



(a) Average/Nominal altitude profile

(b) Average/Nominal velocity profile

Figure 146: Visualization of average profiles beyond 3 nautical miles

All the above inferences were purely based on the statistical data of the flight records in use and are therefore more of operational cut-offs. But they are very useful in achieving the purpose of defining the nominal approach profiles. Similar to the altitude and velocity, nominal profiles are obtained for other recorded as well as derived parameters. For the sake of brevity, only altitude and velocity are presented. Similar work has been performed in commercial operations by Wynnyk et al. [155]

The other phase of flight focused on in this dissertation is take-off. However,

the discretization and obtaining of nominal profile is comparatively straightforward for that phase as there is no requirement to calculate an equivalent of a touchdown point. The point of application of take-off power (marked by the rise in RPM) is typically chosen as the anchoring point and flight data sampled from that point onward. For obtaining the reference profiles, the same steps as approach and landing can be utilized.

C.3 Glide Slope

Prediction of the glide-slope from recorded data draws upon two typical parameters – the ground speed and vertical speed. The instantaneous glide slope can be obtained by taking the inverse tangent of the ratio between the vertical speed and ground speed. It is interesting to note the variation of actual glide slope for all flights in the data set and compare it to the prescribed nominal 3° glide slope during approach. Figure 147 shows this visualization for the approach phase.

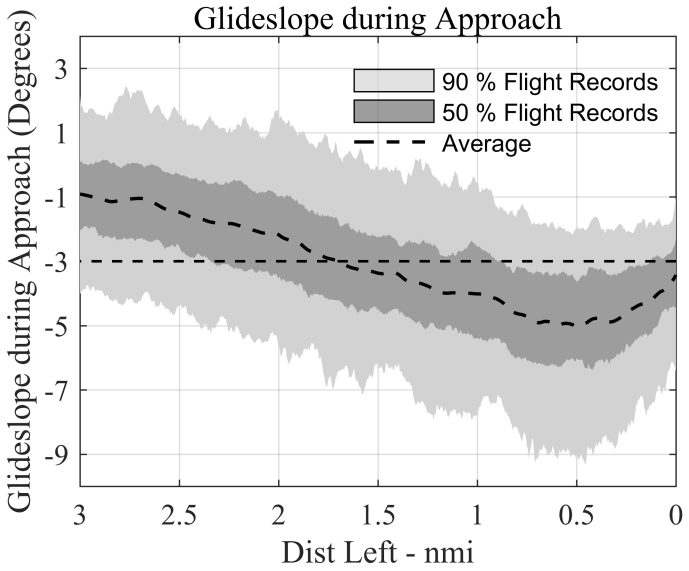


Figure 147: Visualization of glide slope during approach

It is interesting to note that the typical flight data records from the data set tend to intercept the 3 degree line mid-way, indicating a shallower approach in the initial

stages and a steeper slope during the latter stages of the approach.

C.4 Angle of Attack

Obtaining angle of attack is an important step in the use of aerodynamic performance models because it is the central piece of information required to predict both lift and drag. The relationship between the angle of attack (α), flight-path angle (γ), and pitch angle (θ) is as follows:

$$\alpha = \theta - \gamma \tag{34}$$

In a category 2/3 GA flight data recording capability, the pitch angle is typically measured and recorded. The flight path angle can be obtained by taking the inverse tangent of the ratio between vertical speed and true airspeed. Therefore, equation 34 can be used to calculate angle of attack at any point in the flight data record.

C.5 Weight

Accurate knowledge of the weight of the aircraft at any point during the flight is an important piece of information which can affect a number of metrics in safety analysis. Knowledge of the weight can affect calibration of performance models as well as affect the values of evaluated energy metrics. Some of the exceedances are also related to weight. In some cases, when the ground truth data is available, purely statistical models have been used to predict quantities important for aircraft performance. Chati and Balakrishnan [36] have used Gaussian Process Regression on commercial aircraft DFDR archives to estimate take-off weight (TOW). Due to availability of ground truth, their framework uses parameters that are recorded in the DFDR as features for the statistical model and results in high accuracy for prediction of take-off weight. In GA operations, ground truth is may not available even for FOQA-type DFDR data. Whether the take-off weight is available or not, in flight records such as those

utilized in this work, the fuel-flow rate is recorded. This can be used with the take-off weight (or an estimate of it) to obtain the instantaneous weight at every point during the flight. The following two cases of weight-recording sophistication are discussed – take-off weight is available and fuel-flow rate is recorded, take-off weight is not available and fuel-flow rate is recorded. In the first case, the following equation is used, assuming data is collected every one second:

$$W(i) = TOW - \rho \times \sum_{k=1}^{i} (FF(k)) \tag{35}$$

Here TOW represents the take-off weight, FF is the fuel-flow rate, ρ is the density of the fuel (typically 6 lb. per gallon). If the take-off weight is not available, it can be estimated using cruise points identified by phase of flight algorithm [68] and re-arranging the total energy rate metric equation.

$$W(i) = \frac{(T(i) - D(i))\, V(i)}{\dot{h(i)} + \frac{V(i)\dot{V(i)}}{g}} \tag{36}$$

Here i is any cruise point, and all of the terms are evaluated at this cruise point. To get a robust estimate, multiple cruise points can be used and the take-off weight estimate obtained from all of these points can be averaged. To provide an estimate of the uncertainty in important safety metrics due to inaccurate take-off weight information the following figures are presented. In the Figure 149, the specific total energy rate metric is evaluated for a flight with known take-off weight. The same metric is then evaluated using the performance models with three different weight inputs – the actual weight, 80 % of the actual weight and 120 % of the actual weight. The histogram of STER metric relative error is shown in Figure 148 and the actual metric values are plotted over the duration of the flight in Figure 149.

The figures indicate that a variation of the take-off weight has a non-negligible

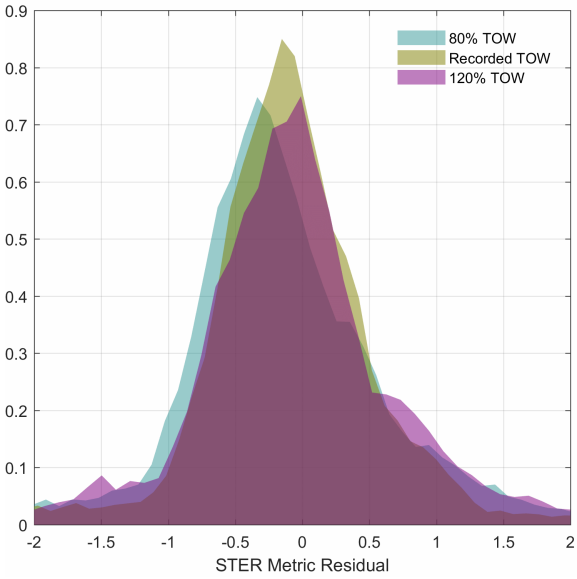


Figure 148: Uncertainty in STER metric with unknown weight - histograms

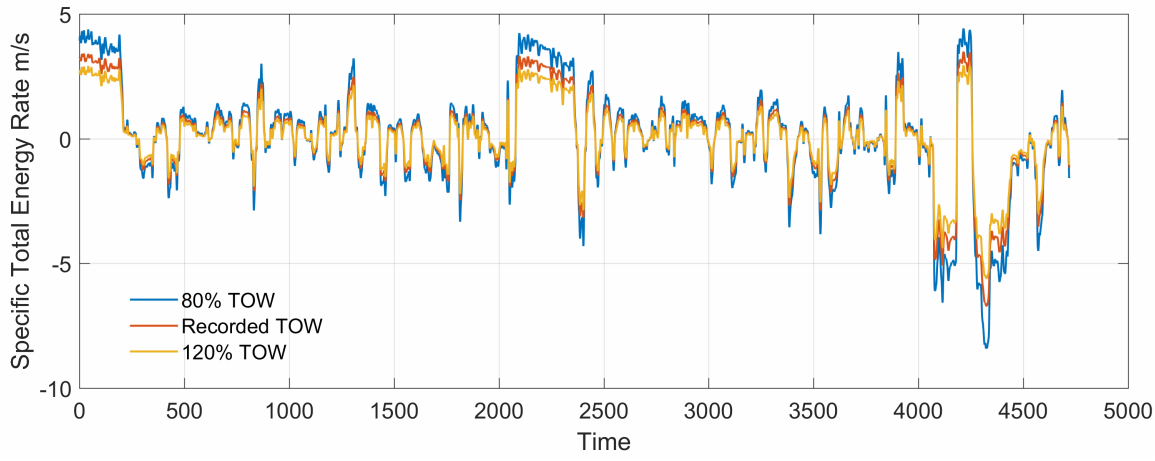


Figure 149: Uncertainty in STER metric with unknown weight

effect on the calculation of the STER metric in certain parts of the flight record. As is evident from the histogram and the data traces though, the spread in error from an uncertainty of 20 % in the take-off weight is not significant and the overall landscape of the metric variation is captured despite take-off weight being uncertain. It is important to note here that the uncertainty considered is in take-off weight only – this means a reliable recording of fuel flow rate is assumed to be available and therefore the effect the take-off weight has in this case is only to shift the curves up

or down.

In case the take-off weight and fuel flow rate are both unknown, in order to use the models to predict this particular metric reasonable assumptions would need to be made about possible take-off weight and typical values of fuel consumption in different phases of flight. Dealing with this uncertainty is however, outside of the scope of this work.

C.6 Rates

For many parameters, the value of the raw parameter is recorded but not its the rate of change. In these situations, simple numerical differentiation can be used to obtain the rate of change of that parameter. If the noise in the data is not large then sufficient accuracy can be obtained using this technique.

C.7 Flap Setting

The position of the high-lift devices (flap position) in a GA aircraft is a relevant factor in the investigation of GA accidents/incidents. However, even for aircraft equipped with state-of-the-art data-logging capability (e.g., Garmin G1000-type systems), the flap position is often not one of the logged parameters. The ability to infer the flap position from flight data records with sufficient accuracy is desirable for this reason. Incorrectly configured aircraft is one of the reasons for unstabilized approaches and also possible accident cause in some cases.

Prior work done by researchers at Georgia Institute of Technology (called the FLAP algorithm) has demonstrated the feasibility of estimating flap position given basic flight data and accurate performance models of the aircraft.

APPENDIX D

DIFFERENT FEATURE VECTOR OPTIONS

Table 15: Summary of parameters required for computation of various feature vectors

No	Parameter/Model	Feature Vector 1	Feature Vector 2	Feature Vector 3	Feature Vector 4
1	Altitude	✓	✓	✓	✓
2	True Airspeed	✓	✓	✓	✓
3	Indicated Speed	✓	✓	✓	✓
4	Vertical Airspeed	✓	✓	✓	✓
5	Outside Air Temperature	✓	✓		
6	Latitude	✓			
7	Longitude	✓			
8	Altitude (GPS)	✓			
9	Ground speed	✓			
10	Pitch	✓	✓		
11	Roll	✓			
12	Lateral Acceleration	✓			
13	Normal Acceleration	✓			
14	Heading	✓			
15	Track	✓			
16	Fuel Quantity (Left)	✓			
17	Fuel Quantity (Right)	✓			
18	Fuel Flow Rate	✓	✓		
19	Oil Temperature	✓			
20	Oil Pressure	✓			
21	RPM	✓	✓		
22	Cylinder Head Temp.	✓			
23	Exhaust Gas Temp.	✓	✓		
24	Flight Path Angle		✓	✓	✓
25	Angle of Attack		✓		
26	Reference Profiles		✓	✓	
27	Aerodynamics Model		✓		
28	Propulsion Model		✓		

APPENDIX E

PCA STUDY

In this appendix, the effect of dimensionality reduction technique PCA (Principal Component Analysis) on the performance of anomaly detection algorithms is tested. It is claimed in Chapter 5 that the value of PCA diminishes if the same or nearly same results **cannot** be obtained with and without PCA on various feature vectors. Therefore, in this appendix, the following numerical experiment is performed: For various feature vector options the overlap between top anomalous flights at various outlier significance levels (1%, 3%, and 5%) is examined. Using different options of feature vectors from energy metrics demonstrated earlier, the list of outliers is obtained with and without using PCA. The two lists are compared with each other for each feature vector option independently at each significance level. Therefore, this experiment indicates the effect of using PCA at a given significance level for a given feature vector. In each case, the PCA is performed such that 99% of the variability of the feature vector is retained while still reducing the size significantly. Ideally, the overlap should be 100% for all the options. Table 16 provides the results obtained.

	Proportion of overlap with and without PCA		
Outlier Significance Level	**FV 2**	**FV 3**	**FV 4**
1%	100.00 %	97.05 %	100.00 %
3%	96.07 %	100.00 %	100.00 %
5%	98.82 %	98.82 %	97.06 %

Table 16: Overlap between sets of anomalous flights obtained using different feature vectors with and without PCA

As seen from the table, in majority of the cases, there is little information lost about the top anomalous flight data records by using PCA for reducing dimensionality as the overlap between anomalous sets is almost 100%. The few cases where it is not exactly 100%, indicate a difference of 1 or 2 anomalous flights with and without PCA. Thus, this experiment demonstrates that for any feature vector option at any significance level under consideration in this dissertation, using PCA does not result in any significant loss of information.

APPENDIX F

LIST OF CALIBRATION FACTORS

Table 17: Description of calibration factors for models

Model	**Factor Description**	**Range**	
Engine	Engine de-rate factor	[0.9 – 1.0]	×
Propeller	Vertical shift of airfoil sectional lift curve	[–1.0 – 1.0]	+
Propeller	Scaling of sectional lift curve slope	[0.1 – 2.0]	×
Propeller	Scaling of sectional minimum drag	[–0.001 – 0.001]	+
Propeller	Scaling of sectional quad. parameters	[–0.001 – 0.001]	+
Propeller	Scaling of sectional lift at min. drag	[–0.2 – 0.2]	+
Propeller	Slope shift of propeller pitch	[0.8 – 1.2]	×
Propeller	Translational shift of propeller pitch	[0 – 15]	+
Propeller	Scaling of propeller chord	[0.5 – 2.0]	×
Aero.	Scaling of maximum lift coefficient (Flap 0)	[–0.2 – 0.2]	+
Aero.	Scaling of maximum lift coefficient (Flap 1)	[–0.2 – 0.2]	+
Aero.	Scaling of maximum lift coefficient (Flap 2)	[–0.3 – 0.2]	+
Aero.	Scaling of maximum lift coefficient (Flap 3)	[–0.3 – 0.3]	+
Aero.	Scaling of parasite drag factor (clean config)	[0.5 – 1.75]	×
Aero.	Scaling of induced drag factor (clean config)	[0.5 – 1.5]	×
Aero.	Drag polar shifting factor	[0.1 – 0.35]	×
Aero.	Scaling of parasite drag increment (Flap 1)	[0.1 – 1.9]	×
Aero.	Scaling of induced drag increment (Flap 1)	[0.1 – 1.9]	×
Aero.	Scaling of parasite drag increment (Flap 2)	[0.2 – 1.8]	×
Aero.	Scaling of induced drag increment (Flap 2)	[0.2 – 1.8]	×
Aero.	Scaling of parasite drag increment (Flap 3)	[0.5 – 1.7]	×
Aero.	Scaling of induced drag increment (Flap 3)	[0.5 – 1.7]	×
Aero.	Scaling of lift–curve slope (Flap 0)	[0.75 – 1.25]	×
Aero.	Scaling of lift–curve slope (Flap 1)	[1 – 1.2]	×
Aero.	Scaling of lift–curve slope (Flap 2)	[1 – 1.2]	×
Aero.	Scaling of lift–curve slope (Flap 3)	[1 – 1.2]	×
Aero.	Scaling of lift–curve intercept (Flap 0)	[–0.25 – 0.25]	+
Aero.	Scaling of lift–curve intercept (Flap 1)	[0.5 – 1.5]	×
Aero.	Scaling of lift–curve intercept (Flap 2)	[0.7 – 1.35]	×
Aero.	Scaling of lift–curve intercept (Flap 3)	[0.85 – 1.2]	×

The last column in Table 17 refers to the manner in which the calibration factor is embedded in the framework. A '+' sign indicates an additive factor and a '×' sign indicates a multiplicative factor. The decision on whether to make a factor additive or multiplicative is based on the amount of freedom required for each factor, the range of possible values, how accurately baseline values are known, etc. Figure 150 shows the variation of curves of different flapped configurations for a specific clean configuration model. The three colors represent curves for three different flap settings. Modifying the ranges of calibration factors will enable making the spread of possible curves wider or narrower as required.

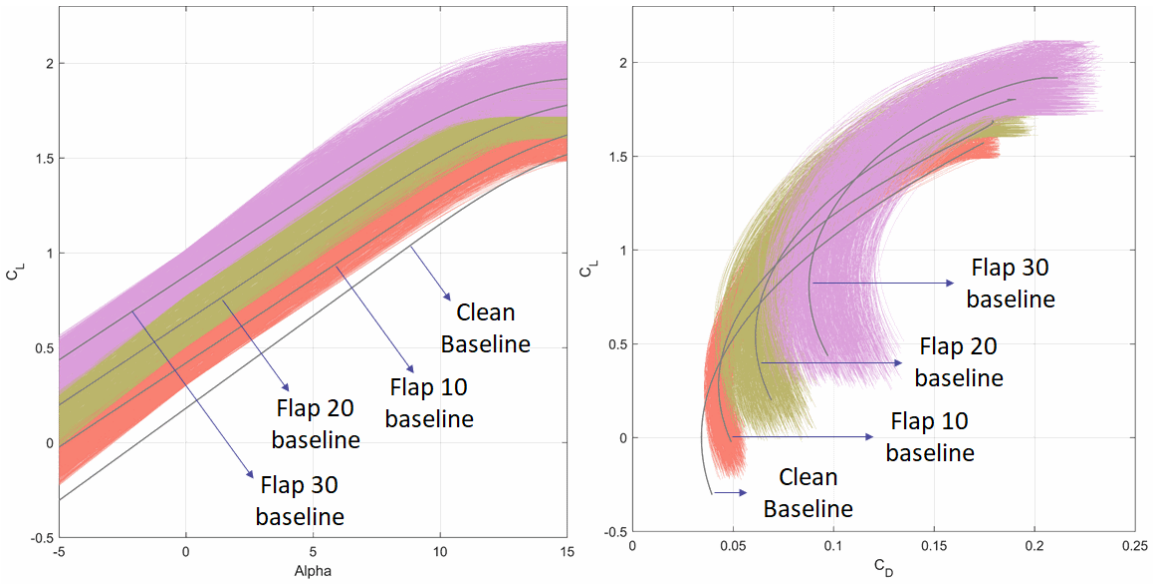


Figure 150: Spread of possible curves using proposed parameterization for flapped configurations

APPENDIX G

POH DISCREPANCY METRICS

This appendix contains details of the discrepancy metrics evaluated for each phase of flight/condition from the POH used in level-1 calibration. Typical tables from a sample POH [13], the logic behind selection of particular metrics, and the method of aggregation is explained for each condition.

G.1 Climb

MAXIMUM RATE-OF-CLIMB AT 2550 POUNDS

CONDITIONS:

Flaps Up
Full Throttle

PRESS ALT FT	CLIMB SPEED KIAS	RATE OF CLIMB - FPM			
		-20°C	0°C	20°C	40°C
S.L.	74	855	785	710	645
2000	73	760	695	625	560
4000	73	685	620	555	495
6000	73	575	515	450	390
8000	72	465	405	345	285
10,000	72	360	300	240	180
12,000	72	255	195	135	- - -

Figure 151: Sample rate-of-climb table

For calculating climb performance, POH tables such as the one in Figure 151 are used. For each condition provided in the table, a particular rate of climb is available. In order to estimate a rate of climb using the performance models, first an RPM is estimated. This is done using the engine model to obtain power supplied by the engine and propeller power curve to obtain power absorbed by the propeller at that advance ratio. The RPM that provides least mismatch is the estimated RPM. Once the RPM is obtained, the lift, drag, and thrust at that condition are estimated and

the rate of climb which minimizes mismatch of specific excess power is the model-predicted rate of climb. This is now compared with the POH table. This discrepancy metric thus exercises input from all three component models – aerodynamics, engine, and propeller.

G.2 Cruise

CRUISE PERFORMANCE

CONDITIONS:
2550 Pounds
Recommended Lean Mixture At All Altitudes (Refer to Section 4, Cruise)

PRESS ALT FT	RPM	20°C BELOW STANDARD TEMP			STANDARD TEMPERATURE			20°C ABOVE STANDARD TEMP		
		% BHP	KTAS	GPH	% BHP	KTAS	GPH	% BHP	KTAS	GPH
2000	2550	83	117	11.1	77	118	10.5	72	117	9.9
	2500	78	115	10.6	73	115	9.9	68	115	9.4
	2400	69	111	9.6	64	110	9.0	60	109	8.5
	2300	61	105	8.6	57	104	8.1	53	102	7.7
	2200	53	99	7.7	50	97	7.3	47	95	6.9
	2100	47	92	6.9	44	90	6.6	42	89	6.3
4000	2600	83	120	11.1	77	120	10.4	72	119	9.8
	2550	79	118	10.6	73	117	9.9	68	117	9.4
	2500	74	115	10.1	69	115	9.5	64	114	8.9
	2400	65	110	9.1	61	109	8.5	57	107	8.1
	2300	58	104	8.2	54	102	7.7	51	101	7.3
	2200	51	98	7.4	48	96	7.0	45	94	6.7
	2100	45	91	6.6	42	89	6.4	40	87	6.1
6000	2650	83	122	11.1	77	122	10.4	72	121	9.8
	2600	78	120	10.6	73	119	9.9	68	118	9.4
	2500	70	115	9.6	65	114	9.0	60	112	8.5
	2400	62	109	8.6	57	108	8.2	54	106	7.7
	2300	54	103	7.8	51	101	7.4	48	99	7.0
	2200	48	96	7.1	45	94	6.7	43	92	6.4

Figure 152: Sample cruise performance table

Two types of cruise metrics are evaluated using typically available cruise performance tables from the POH. The first metric is the RPM mismatch in cruise. This metric utilizes the engine and propeller calibration factors. The steps involved in calculating this discrepancy metric are to first use %-power from POH cruise table to determine engine shaft-power output. Next step is to query propeller curves to determine the RPM at which the propeller absorbs this power. The final step is to

compare this RPM to the RPM listed in the cruise performance table (Figure 152.

The second discrepancy metric involves longitudinal force imbalance during cruise and utilizes the calibration factors from aerodynamics and propeller. The steps to calculate this metric are to first compute predicted thrust from the propeller model at POH flight condition. The next step is to compute predicted drag from aerodynamics model at POH flight condition. Finally, the mismatch between these two quantities is obtained as the discrepancy metric.

G.3 Glide

The POH contains the best glide speed and a glide performance table from which best glide angle can be calculated. The glide metrics calculate the discrepancy in the model prediction for both of these quantities and the values reported in the POH.

G.4 Stall

The POH typically contains a table such as that shown in Figure 153 which lists the speed at which the aircraft at maximum take-off weight is expected to stall for different flap settings and bank angle. The $C_{L,max}$ of the aerodynamics model and that obtained using the conditions in the table can be directly compared to calculate and minimize discrepancy between the two. The stall discrepancy metric also exercises only the aerodynamics model.

STALL SPEEDS AT 2550 POUNDS

Conditions:
Power Off

MOST REARWARD CENTER OF GRAVITY

FLAP SETTING	ANGLE OF BANK							
	0°		30°		45°		60°	
	KIAS	KCAS	KIAS	KCAS	KIAS	KCAS	KIAS	KCAS
UP	48	53	52	57	57	63	68	75
10°	42	50	45	54	50	59	59	71
30°	40	48	43	52	48	57	57	68

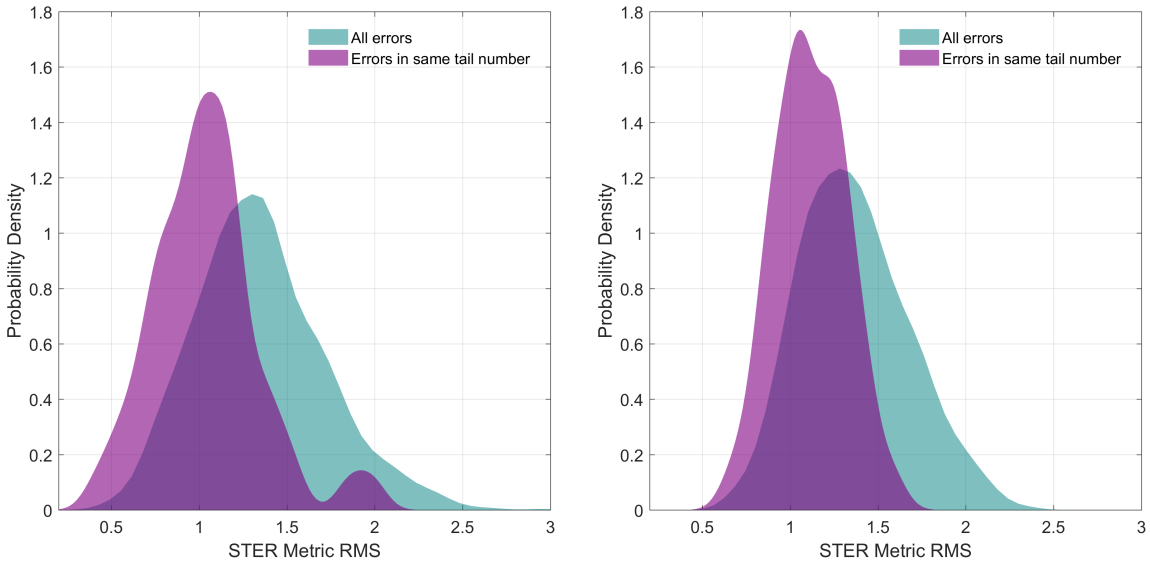
Figure 153: Sample stall speeds table

APPENDIX H

MODEL CALIBRATION CASE STUDY

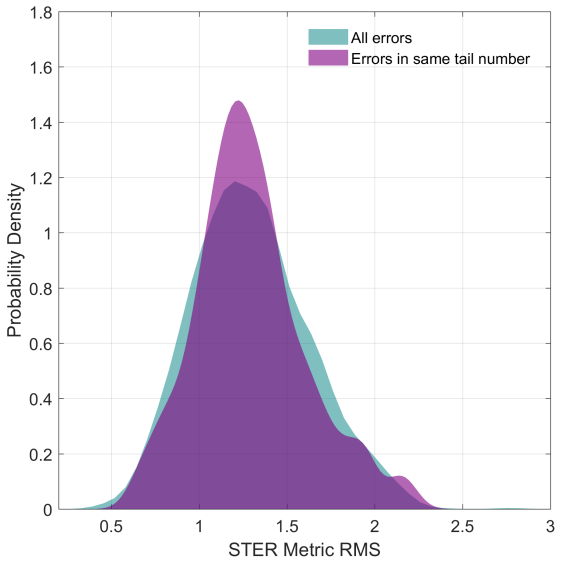
One of the important pieces of information which may not always be available during model calibration is the tail number of the aircraft in the flight data record. This is important because it is expected that the performance of the flight for the same tail number will be similar compared to the performance of a flight on a different tail number at the same conditions. This is due various factors such as the degradation of engine performance with age of the aircraft, aerodynamic performance changes due to modifications/surface roughness, etc. It is hypothesized that when a performance model is calibrated to a particular flight data record, the predictive power of that model would be inherently better on other flight data records collected on the same aircraft. To that end, calibration is carried out using a data record from different tail numbers and the distribution of errors on records from the same flight and errors on the entire data set are obtained. The distributions for four sample calibrations from the data set are presented in Figure 154.

The results reveal that, in general, the performance on the same tail number as that used for calibration is markedly better than the performance on the entire data set. For all the other cases, similar trends are observed with the magnitude of the shift in error RMS distribution different for different aircraft. In some cases, such as flight 3 in the figure, there is no noticeable difference between the specific flight data record and the remaining data set. Therefore, a unique model for each tail number can be calibrated and used when such information is available as it would improve the predictive power. However, in the most general cases, such information may not be available and in those situations, using a model that provides good predictions on

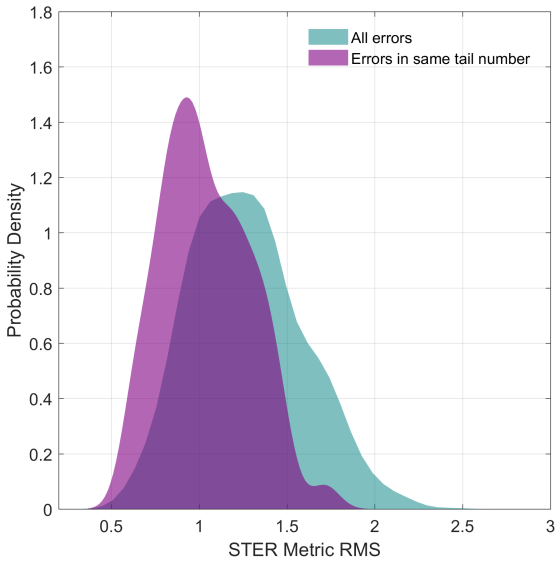


(a) Calibration Flight 1

(b) Calibration Flight 2



(c) Calibration Flight 3



(d) Calibration Flight 4

Figure 154: Differences in STER RMS error when measured on same tail number as calibration flight versus different tail number

the entire data set is preferred.

REFERENCES

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

[2] "Federal Aviation Administration, 14 CFR §23.3 Airplane Categories." url: https://www.ecfr.gov/cgi-bin/text-idx?node=pt14.1.23.

[3] "Federal Aviation Administration, 14 CFR §91.609 appendix E, Flight Data Recorders and Cockpit Voice Recorders." url: https://www.ecfr.gov/cgi-bin/text-idx?node=14:2.0.1.3.10.

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

[5] "Accidents During Flight Instruction: A Review," tech. rep., Air Safety Institute, 2014. Technical Report.

[6] "Federal Aviaition Administration Fact Sheet - General Aviation Safety," 2017. Retrieved - 04/2017. url: https://www.faa.gov/news/fact_sheets/news_story.cfm?newsId=16774.

[7] Abbott, I. H., VonDoenhoff, A. E., and Stivers, L. S., "Summary of Airfoil Data," Tech. Rep. NACA TR-824, National Advisory Committee for Aeronautics (NACA), 1945. url: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19930090976.pdf.

[8] Adami, T. M., Uijt de Haag, M., Theunissen, E., Sizoo, D., and McGuire, R., "An Energy Management Sisplay for General Aviation Safety Enhancements," in *Digital Avionics Systems Conference (DASC), 2014 IEEE/AIAA 33rd*, pp. 2D1–1 – 2D1–11, IEEE, 2014. doi:10.1109/DASC.2014.6979430.

[9] AIAA, *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations, American Institute of Aeronautics and Astronautics, 1998 AIAA G-077-1998*, 1998. doi:10.2514/4.472855.

[10] Amelink, M. H., Mulder, M., Van Paassen, M., and Flach, J., "Theoretical Foundations for a Total Energy-based Perspective Flight-path Display," *The International Journal of Aviation Psychology*, vol. 15, no. 3, pp. 205–231, 2005. doi:10.1207/s15327108ijap1503_1.

[11] Amidan, B. G. and Ferryman, T. A., "APMS SVD Methodology and Implementation," tech. rep., U.S. Department of Energy PNWD-3026, 2000. doi:10.2172/753847.

[12] Anderson, M., Lopez, J., and Evers, J., "A Comparison of Trajectory Determination Approaches for Small UAV's," in *AIAA Guidance, Navigation, and Control and Co-located Conferences*, Aug. 2006. Paper No. AIAA 2006-6644, doi:10.2514/6.2006-6644 .

[13] Anon., *Information Manual: Cessna Skyhawk SP, 1st ed., Pilot Operating Handbook, Model 172S NAV III GFC 700 AFCS, Cessna Aircraft Company, 2007.*, 2007.

[14] AOPA, "24th Joseph T. Nall Report - Aircraft Owners and Pilots Association," 2012. url: `https://www.aopa.org/-/media/files/aopa/home/pilot-resources/safety-and-proficiency/accident-analysis/nall-report/15-fn-0022-1-24th-nall-v6.pdf`.

[15] AOPA, "Aircraft Owners and Pilots Association - What is General Aviation?," 2017. url: `https://www.aopa.org/-/media/files/aopa/home/advocacy/what_ga.pdf`.

[16] Appareo, "Stratus by Appareo," 2017. Retrieved - 04/2017. url: `https://www.foreflight.com/products/stratus/`.

[17] Archer, L. R. J., "Use of Rounded Base Turns in General Aviation Traffic Patterns: A Quantitative Analysis," Master's thesis, University of North Dakota, 2017. url: `http://prx.library.gatech.edu/login?url=https://search.proquest.com/docview/1939999652?accountid=11107`.

[18] Bay, S. D. and Schwabacher, M., "Mining Distance-based Outliers in Near Linear Time with Randomization and a Simple Pruning Rule," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 29–38, ACM, 2003. doi:10.1145/956750.956758.

[19] Belcastro, C. M. and Foster, J. V., "Aircraft Loss-of-Control Accident Analysis," in *AIAA Guidance, Navigation, and Control Conference*, August 2010. Paper No. AIAA-2010-8004, doi:10.2514/6.2010-8004.

[20] Belcastro, C. M. and Jacobson, S. R., "Future Integrated Systems Concept for Preventing Aircraft Loss-of-control Accidents," in *AIAA Guidance, Navigation, and Control Conference*, August 2010. Paper No. AIAA-2010-8142, doi:10.2514/6.2010-8142.

[21] Bhargava, D. and Marais, K., "Archetypal Models of Runway Incursions," in *17th AIAA Aviation Technology, Integration, and Operations Conference*, 2017. Paper Number: AIAA-2017-4392.

[22] Boeing, "Statistical Summary of Commercial Jet Airplane Accidents - Boeing Commercial Airplanes," 2012. url: http://www.boeing.com/resources/boeingdotcom/company/about_bca/pdf/statsum.pdf.

[23] Boyd, D. D., "General Aviation Accidents Related to Exceedance of Airplane Weight/Center of Gravity Limits," *Accident Analysis and Prevention*, vol. 91, pp. 19 – 23, 2016. doi: doi:https://doi.org/10.1016/j.aap.2016.02.019.

[24] Boyd, J. R., Christie, T. P., and Gibson, J. E., "Energy Maneuverability," *Air Proving Ground Center Report (APGC-TR-66-4)*, vol. 1, 1966.

[25] Breunig, M. M., Kriegel, H.-P., Ng, R. T., and Sander, J., "Lof: Identifying Density-based Local Outliers," in *ACM sigmod record*, vol. 29, pp. 93–104, ACM, 2000. doi:10.1145/342009.335388.

[26] Bronsvoort, J., McDonald, G., Paglione, M., Young, C. M., Boucquey, J., Hochwarth, J. K., and Gallo, E., "Real-time Trajectory Predictor Calibration through Extended Projected Profile Down-link," in *Eleventh USA/Europe Air Traffic Management Research and Development Seminar*, 2015.

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

[28] CAA, "Civil Aviation Authority - Flight Data Monitoring CAP 739 Second Edition," June 2013. ISBN 978-0-11792-840-4. url: https://publicapps.caa.co.uk/docs/33/CAP739.pdf.

[29] Caliński, T. and Harabasz, J., "A Dendrite Method for Cluster Analysis," *Communications in Statistics*, vol. 3, pp. 1–27, Jan. 1974. doi:10.1080/03610927408827101.

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

[31] Campos, G. O., Zimek, A., Sander, J., Campello, R. J., Micenkov, B., Schubert, E., Assent, I., and Houle, M. E., "On the Evaluation of Unsupervised Outlier Detection: Measures, Datasets, and an Empirical study," *Data Mining and Knowledge Discovery*, pp. 1–37, 2015. doi:10.1007/s10618-015-0444-8.

[32] CAST, "Mode Awareness and Energy State Management Aspects of Flight Deck Automation Safety Enhancement 30 Revision – 5," tech. rep., Commercial Aviation Safety Team (CAST), 2008. url: http://www.cast-safety.org/pdf/cast_automation_aug08.pdf.

[33] Chakraborty, I., Lozano, B., and Mavris, D. N., "Pilot-friendliness Considerations for Personal Air Vehicle Flight Control Systems," in *AIAA Aviation*, American Institute of Aeronautics and Astronautics, June 2015. Paper No. AIAA 2015-2852, doi:10.2514/6.2015-2852.

[34] Chandola, V., *Anomaly Detection for Symbolic Sequences and Time Series Data*. PhD thesis, University of Minnesota, 2009. url: `http://hdl.handle.net/11299/56597`.

[35] Chandola, V., Banerjee, A., and Kumar, V., "Anomaly Detection: A Survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, 2009. doi:10.1145/1541880.1541882.

[36] Chati, Y. S. and Balakrishnan, H., "Statistical Modeling of Aircraft Takeoff Weight," in *Twelfth USA/Europe Air Traffic Management Research and Development Seminar (ATM 2017)*, 2017. `http://www.atmseminarus.org/seminarContent/seminar12/papers/12th_ATM_RD_Seminar_paper_94.pdf`.

[37] Chu, E., Gorinevsky, D., and Boyd, S., "Detecting Aircraft Performance Anomalies from Cruise Flight Data," in *AIAA Infotech@Aerospace, Atlanta, Georgia.*, 2010. Paper No: AIAA 2010-3307. doi:10.2514/6.2010-3307.

[38] Clachar, S. A., "Identifying and Analyzing Atypical Flights by Using Supervised and Unsupervised Approaches," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2471, pp. 10–18, 2015. doi:10.3141/2471-02.

[39] Cokorilo, O., Luca, M. D., and DellAcqua, G., "Aircraft Safety Analysis using Clustering Algorithms," *Journal of Risk Research*, vol. 17, no. 10, pp. 1325–1340, 2014. doi: doi:10.1080/13669877.2013.879493.

[40] Das, S., Matthews, B. L., and Lawrence, R., "Fleet Level Anomaly Detection of Aviation Safety Data," in *Prognostics and Health Management (PHM), 2011 IEEE Conference on*, pp. 1–10, IEEE, 2011. doi:10.1109/ICPHM.2011.6024356.

[41] Das, S., Matthews, B. L., Srivastava, A. N., and Oza, N. C., "Multiple Kernel Learning for Heterogeneous Anomaly Detection: Algorithm and Aviation Safety Case Study," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 47–56, ACM, 2010. doi:10.1145/1835804.1835813.

[42] Davies, D. L. and Bouldin, D. W., "A Cluster Separation Measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, 1979. doi:10.1109/TPAMI.1979.4766909.

[43] de Jong, P., de Gelder, N., Verhoeven, R., Bussink, F., Kohrs, R., van Paassen, M., and Mulder, M., "Time and Energy Management During

Descent and Approach: Batch Simulation Study," *Journal of Aircraft*, vol. 52, no. 1, pp. 190–203, 2014. doi:10.2514/1.C032668.

[44] Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T., "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II," *IEEE transactions on evolutionary computation*, vol. 6, no. 2, pp. 182–197, 2002. doi:10.1109/4235.996017.

[45] Delgado, I. R., Dempsey, P. J., and Simon, D. L., "A Survey of Current Rotorcraft Propulsion Health Monitoring Technologies," tech. rep., NASA/TM2012-217420, 2012.

[46] Dempster, A. P., Laird, N. M., and Rubin, D. B., "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977. doi:10.2307/2984875.

[47] Desgraupes, B., "Clustering Indices - Clustercrit Package for R," 2013. `ftp://apache.cs.uu.nl/mirror/CRAN/web/packages/clusterCrit/vignettes/clusterCrit.pdf`.

[48] Drela, M., "Xfoil: An Analysis and Ddesign System for Low Reynolds Number Airfoils," in *Low Reynolds number aerodynamics*, pp. 1–12, Springer, 1989.

[49] Drela, M., "Qprop formulation," *Massachusetts Inst. of Technology Aeronautics and Astronautics, Cambridge, MA*, 2006.

[50] Ester, M., Kriegel, H.-P., Sander, J., and Xu, X., "A Density-based Algorithm for Ddiscovering Cclusters in Large Spatial Databases with Noise," in *Kdd*, vol. 96, pp. 226–231, 1996.

[51] FAA, "Federal Aviation Administration Advisory Circular 120–71a," 2003. url: `https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC120-71A.pdf`.

[52] FAA, "Federal Aviaition Administration Aerospace Forecasts Fiscal Years 2016-2036.," 2017. url: `https://www.faa.gov/data_research/aviation/aerospace_forecasts/media/FY2016-36_FAA_Aerospace_Forecast.pdf`.

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

[54] FAA, "Partnership to Enhance General Aviation Safety Accessibility and Sustainability (PEGASAS)," 2017. url: `www.pegasas.aero`.

[55] FAA, "Pilot/Controller Glossary - Federal Aviaition Administration," 2017. Retrieved: 3/ 2017 (url: `https://www.faa.gov/air_traffic/publications/media/pcg.pdf`).

[56] Fala, N. and Marais, K., "Detecting Safety Events During Approach in General Aviation Operations," in *AIAA Aviation*, American Institute of Aeronautics and Astronautics, June 2016. Paper No. AIAA-2016-3914, doi:10.2514/6.2016-3914.

[57] Fernandes, R. V., "An Analysis of the Potential Benefits to Airlines of Flight Data Monitoring Programmes," Master's thesis, Cranfield University, 2002.

[58] FlightGear, "FlightGear Flight Simulator Website," 2017. Retrieved - 04/2017.

[59] FSF, "Flight Safety Foundation ALAR Briefing Note 4.2 - Energy Management." Technical Report, 2000. url: `https://flightsafety.org/files/alar_bn4-2-energymgmt.pdf`.

[60] FSF, "Flight Safety Foundation ALAR Briefing Note 7.1 - Stabilized Approach." Technical Report, August–November 2000. url: `https://flightsafety.org/wp-content/uploads/2016/09/alar_bn7-1stablizedappr.pdf`.

[61] Fultz, A. J. and Ashley, W. S., "Fatal Weather-related General Aviation Accidents in the United States," *Physical Geography*, vol. 37, no. 5, pp. 291–312, 2016. doi: doi:10.1080/02723646.2016.1211854.

[62] GAJSC, "Loss of Control Work Group - Approach and Landing," tech. rep., General Aviaition Joint Steering Committee, 2012. url: `http://www.gajsc.org/loss-of-control/`.

[63] GAJSC, "General Aviaition Joint Steering Committee (GAJSC)," 2017. url: `http://www.gajsc.org/`.

[64] Gandhi, N., Richards, N., Bateman, A., Bolstad, C., and Costello, A., "Pilot-in-the-Loop Demonstration of an Energy Monitor and Crew Alerting System," in *AIAA Guidance, Navigation, and Control and Co-located Conferences*, Aug. 2012. Paper No. AIAA-2012-4819, doi:10.2514/6.2012-4819.

[65] Garmin, *G1000 Integrated Flight Deck Pilot's Guide*, 2017. Retrieved 03/2017. url: `http://static.garmincdn.com/pumac/190-00498-07_0A_Web.pdf`.

[66] Gavrilovski, A., *Predictive Helicopter Flight Data Monitoring for Flight Safety Design*. PhD thesis, Georgia Institute of Technology, 2016.

[67] Gavrilovski, A., Jimenez, H., Mavris, D. N., Rao, A. H., Shin, S., Hwang, I., and Marais, K., "Challenges and Opportunities in Flight Data Mining: A Review of the State of the Art," in *AIAA SciTech Forum*, American Institute of Aeronautics and Astronautics, Jan. 2016. Paper No: AIAA 2016-0923 doi:10.2514/6.2016-0923 .

[68] Goblet, V., Fala, N., and Marais, K., "Identifying Phases of Flight in General Aviation Operations," in *AIAA Aviation*, American Institute of Aeronautics and Astronautics, June 2015. Paper No: AIAA 2015-2851. doi:10.2514/6.2015-2851.

[69] Gorinevsky, D., Matthews, B., and Martin, R., "Aircraft Anomaly Detection using Performance Models Trained on Fleet Data," in *Intelligent Data Understanding (CIDU), 2012 Conference on*, pp. 17–23, IEEE, 2012. doi:10.1109/CIDU.2012.6382196.

[70] Gudmundsson, S., *General Aviation Aircraft Design: Applied Methods and Procedures*. Butterworth-Heinemann, 2013.

[71] Harrison, E., Jimenez, H., and Mavris, D. N., "Investigation and Flight Dynamic Analysis of General Aviation Safety," in *16th AIAA Aviation Technology, Integration, and Operations Conference*, American Institute of Aeronautics and Astronautics, June 2016. Paper Number: AIAA 2016-3108, doi: doi:10.2514/6.2016-3108 .

[72] Harrison, E., Min, S., Jimenez, H., and Mavris, D. N., "Implementataion and Validation of an Internal Combustion Engine and Propeller Model for General Aviation Aircraft Performance Studies," in *15th AIAA Aviation Technology, Integration, and Operations Conference*, American Institute of Aeronautics and Astronautics, June 2015. Paper No. AIAA-2015-2850, doi:10.2514/6.2015-2850.

[73] Higgins, J., Clachar, S., and Hennselek, K., "Flight Data Monitoring (FDM)/Flight Operational Quality Assurance (FOQA) for General Aviation," tech. rep., U.S. Department of Transportation - Federal Aviation Administration, December 2013. Technical Report DOT/FAA/AR-13/21.

[74] Hotelling, H., "Analysis of a Complex of Statistical Variables into Principal Components.," *Journal of educational psychology*, vol. 24, no. 6, pp. 417–441, 1933. doi:10.1037/h0071325.

[75] Hurst, M., Cramp, T., Collins, M., Delanghe, C., Lau, S., Morgan, J., Pilgrim, M., Price, M., Reed, C., and Tietjen, S., "Helicopter Flight Data Monitoring - Industry Best Practices," 2012. url: `http://www.hfdm.org/`.

[76] ICAO, *Safety Management Manual - International Civil Aviation Organization Doc 9859 AN/474*, 2013. url: `http://www.icao.int/safety/SafetyManagement/Documents/Doc.9859.3rd%20Edition.alltext.en.pdf`.

[77] Iverson, D. L., "Inductive System Health Monitoring," tech. rep., National Aeronautics and Space Administration, 2004. url: `https://ntrs.nasa.gov/search.jsp?R=20040068062`.

[78] Jategaonkar, R., *Flight Vehicle System Identification: a Time Domain Methodology*, vol. 216. AIAA, Reston, VA, USA, 2006.

[79] Kennedy, M. C. and O'Hagan, A., "Bayesian Calibration of Computer Models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 3, pp. 425–464, 2001. doi:10.1111/1467-9868.00294.

[80] Keogh, E., Lin, J., and Fu, A., "HOT SAX: Efficiently finding the most unusual time series subsequence," in *Fifth IEEE International Conference on Data Mining (ICDM'05)*, p. 8 pp., IEEE, 2005. doi:10.1109/ICDM.2005.79.

[81] Krajcek, K., Nikolic, D., and Domitrovic, A., "Aircraft Performance Monitoring from Flight Data," *Tehnicki vjesnik/Technical Gazette*, vol. 22, no. 5, 2015. doi:10.17559/TV-20131220145918.

[82] Kuo, B. C., Guan, W.-l., and Chen, P.-c., "In Search of General Aviation Flight Data Monitoring: Lightweight Recording System," in *17th AIAA Aviation Technology, Integration, and Operations Conference*, 2017. Paper Number: AIAA 2017-3439, doi: doi:10.2514/6.2017-3439.

[83] Kurdjukov, A., Natchinkina, G., and Shevtchenko, A., "Energy Approach to Flight Control," in *AIAA Guidance, Navigation, and Control Conference and Exhibit, Boston, MA*, pp. 543–553, 1998. Paper No. AIAA-98-4211, doi:10.2514/6.1998-4211.

[84] Lambregts, A., A., "Vertical Flight Path and Speed Control Autopilot Design Using Total Energy Principles," in *AIAA Guidance, Navigation, and Control and Co-located Conferences*, Aug. 1983. Paper No. AIAA-83-2239, doi:10.2514/6.1983-2239.

[85] Lambregts, A., "Integrated System Design for Flight and Propulsion Control using Total Energy Principles," in *Aircraft Design, Systems and Technology Meeting, Fort Worth, TX*, American Institute of Aeronautics and Astronautics, 1983. Paper No. AIAA-83-2561, doi:10.2514/6.1983-2561.

[86] Lambregts, A., Tadema, J., Rademaker, R., and Theunissen, E., "Defining Maximum Safe Maneuvering Authority in 3D Space Required for Autonomous Integrated Conflict Resolution," in *Digital Avionics Systems Conference, DASC'09. IEEE/AIAA 28th*, pp. 5.C.1–1 – 5.C.1–17, IEEE, 2009. doi:10.1109/DASC.2009.5347465.

[87] Lambregts, T., Rademaker, R., and Theunissen, E., "A New Ecological Primary Flight Display Concept," in *Digital Avionics Systems Conference, 2008. DASC 2008. IEEE/AIAA 27th*, pp. 4.A. 1–1 – 4.A.1–20, IEEE, 2008. doi:10.1109/DASC.2008.4702820.

[88] Lee, Y.-H. and Liu, B.-S., "Inflight Workload Assessment: Comparison of Subjective and Physiological Measurements," *Aviation, Space, and Environmental Medicine*, vol. 74, no. 10, pp. 1078–1084, 2003.

[89] Li, L., Das, S., John Hansman, R., Palacios, R., and Srivastava, A. N., "Analysis of Flight Data Using Clustering Techniques for Detecting Abnormal Operations," *Journal of Aerospace Information Systems*, vol. 12, pp. 587–598, Sept. 2015. doi:10.2514/1.I010329.

[90] Li, L., Hansman, R. J., Palacios, R., and Welsch, R., "Anomaly Detection via a Gaussian Mixture Model for Flight Operation and Safety Monitoring," *Transportation Research Part C: Emerging Technologies*, vol. 64, pp. 45–57, 2016.

[91] Liao, T. W., "Clustering of Time Series Data - A Survey," *Pattern Recognition*, vol. 38, no. 11, pp. 1857 – 1874, 2005. doi:10.1016/j.patcog.2005.01.025.

[92] Lin, J., Keogh, E., Lonardi, S., and Chiu, B., "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, pp. 2–11, ACM, 2003. doi:10.1145/882082.882086.

[93] Liu, Y., Li, Z., Xiong, H., Gao, X., and Wu, J., "Understanding of Internal Clustering Validation Measures," in *2010 IEEE 10th International Conference on Data Mining (ICDM)*, pp. 911–916, IEEE, 2010. doi:10.1109/ICDM.2010.35.

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

[95] Longmuir, M. and Ahmed, N. A., "Commercial Aircraft Exterior Cleaning Optimization," *Journal of Aircraft*, vol. 46, pp. 284–290, Jan. 2009.

[96] Luxhoj, J. T. and Coit, D. W., "Modeling Low probability/high Consequence Events: An Aviation Safety Risk Model," in *RAMS '06. Annual Reliability and Maintainability Symposium, 2006.*, pp. 215–221, Jan 2006. doi: doi:10.1109/RAMS.2006.1677377.

[97] Major, W. L., Carney, T., Keller, J., Xie, A., Price, M., Duncan, J., Brown, L., Whitehurst, G. R., Rantz, W. G., Nicolai, D., and others, "VFR-into-IMC Accident Trends: Perceptions of Deficiencies in Training," *Journal of Aviation Technology and Engineering*, vol. 7, no. 1, p. 4, 2017. doi: doi:10.7771/2159-6670.1153.

[98] Mathworks, "Matlab online documentation," 2017. url: `http://www.mathworks.com/help/matlab/`.

[99] Matthews, B., Das, S., Bhaduri, K., Das, K., Martin, R., and Oza, N., "Discovering Anomalous Aviation Safety Events Using Scalable Data Mining Algorithms," *Journal of Aerospace Information Systems*, vol. 10, pp. 467–475, Oct. 2013. doi:10.2514/1.I010080.

[100] Melnyk, I., Matthews, B., Valizadegan, H., Banerjee, A., and Oza, N., "Vector Autoregressive Model-based Anomaly Detection in Aviation Systems," *Journal of Aerospace Information Systems*, pp. 161–173, 2016. doi:10.2514/1.I010394.

[101] Merkt, J. R., "Flight Energy Management Training: Promoting Safety and Efficiency," *Journal of Aviation Technology and Engineering*, vol. 3, no. 1, pp. 24–36, 2013. doi:10.7771/2159-6670.1072.

[102] Min, S., Harrison, E., Jimenez, H., and Mavris, D. N., "Development of Aerodynamic Modeling and Calibration Methods for General Aviation Aircraft Performance Analysis - a Survey and Comparison of Models," in *AIAA Aviation*, American Institute of Aeronautics and Astronautics, June 2015. Paper No. AIAA 2015-2853, doi:10.2514/6.2015-2853.

[103] Mitchell, K., Sholy, B., and Stolzer, A. J., "General Aviation Aircraft Flight Operations Quality Assurance: Overcoming the Obstacles," *IEEE Aerospace and Electronic Systems Magazine*, vol. 22, no. 6, pp. 9–15, 2007. doi: doi:10.1109/MAES.2007.384075.

[104] MITRE, "General Aviation Pilots Get Their GAARD Up with New App," 2015. Retrieved - 04/2017. url: `https://www.mitre.org/publications/project-stories/general-aviation-pilots-get-their-gaard-up-with-new-app`.

[105] Moriarty, D. and Jarvis, S., "A Systems Perspective on the Unstable Approach in Commercial Aviation," *Reliability Engineering & System Safety*, vol. 131, pp. 197–202, 2014. doi: doi:10.1016/j.ress.2014.06.019.

[106] Mugtussids, I. B., *Flight Data Processing Techniques to Identify Unusual Events*. PhD thesis, Virginia Polytechnic Institute and State University, 2000. url: `http://hdl.handle.net/10919/28095`.

[107] Netjasov, F. and Janic, M., "A Review of Research on Risk and Safety Modelling in Civil Aviation," *Journal of Air Transport Management*, vol. 14, no. 4, pp. 213 – 220, 2008.

[108] NTSB, "National Transportation Safety Board Accident Statistics," 2013. Retrieved 04/2017. url: `https://www.ntsb.gov/investigations/data/Pages/AviationDataStats2014.aspx`.

[109] NTSB, "National Transportation Safety Board Website (retrieved 10/2016)," 2013. url: `http://www.ntsb.gov/investigations/data/Pages/AviationDataStats.aspx`.

[110] NTSB, "NTSB Most Wanted List 2017-2018," 2017. url: `https://www.ntsb.gov/safety/mwl/Documents/2017-18/MWL-Brochure2017-18.pdf`.

[111] Nuic, A., "User Manual for the Base of Aircraft Data (BADA) Revision 3.10," tech. rep., Eurocontrol, 2010. Technical Report EEC Note no. 20/00, https://www.eurocontrol.int/sites/default/files/content/documents/sesar/bada3.11-user-manual.pdf.

[112] Oberkampf, W. L. and Barone, M. F., "Measures of Agreement Between Computation and Experiment: Validation Metrics," *Journal of Computational Physics*, vol. 217, no. 1, pp. 5–36, 2006. doi:10.1016/j.jcp.2006.03.037.

[113] Oberkampf, W. L. and Roy, C. J., *Verification and Validation in Scientific Computing*. Cambridge University Press, 2010. Cambridge Books Online.

[114] Payan, A. P., Gavrilovski, A., Jimenez, H., and Mavris, D. N., "Improvement of Rotorcraft Safety Metrics Using Performance Models and Data Integration," *Journal of Aerospace Information Systems*, vol. 14, no. 1, pp. 1–14, 2016. doi:10.2514/1.I010467.

[115] Payan, A. P., Lin, P.-N., Johnson, C., and Mavris, D. N., "Helicopter Approach Stability Analysis Using Flight Data Records," in *17th AIAA Aviation Technology, Integration, and Operations Conference*, 2017. Paper Number: AIAA 2017-3437, doi:10.2514/6.2017-3437.

[116] Puranik, Tejas. G. Harrison, E. H., Min, S. M., Chakraborty, I., and N., M. D., "A Framework for General Aviation Aircraft Performance Model Calibration and Validation," in *18th AIAA Aviation, Technology, Integration, and Operations Conference, Atlanta, GA. June*, 2018.

[117] Puranik, T., Harrison, E., Min, S., Jimenez, H., and Mavris, D., "Energy-based Metrics for General Aviation Flight Data Record Analysis," in *16th AIAA Aviation Technology, Integration, and Operations Conference*, 2016. Paper No. AIAA 2016-3915, doi:10.2514/6.2016-3915.

[118] Puranik, T., Harrison, E., Min, S., Jimenez, H., and Mavris, D., "General Aviation Approach and Landing Analysis Using Flight Data Records," in *16th AIAA Aviation Technology, Integration, and Operations Conference*, 2016. Paper No. AIAA 2016-3913, doi:10.2514/6.2016-3913.

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

[120] Puranik, T. and Mavris, D., "Anomaly Detection in General-Aviation Operations Using Energy Metrics and Flight-Data Records," *Journal of Aerospace Information Systems*, vol. 15, pp. 22–35, 2017. doi:10.2514/1.I010582.

[121] Puranik, T. G., Chakraborty, I., and Mavris, D., "Analysis of General Aviation Landing and Touchdown Performance using Approach Performance and Weather Data," in *18th AIAA Aviation, Technology, Integration, and Operations Conference, Atlanta, GA. June*, 2018.

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

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

[124] Rao, A. H., *A New Approach to Modeling Aviation Accidents*. PhD thesis, 2016. url:`https://search.proquest.com/docview/1881532512?accountid=11107`.

[125] Rao, A. H. and Marais, K., "Identifying High-Risk Occurrence Chains in Helicopter Operations from Accident Data," in *15th AIAA Aviation Technology, Integration, and Operations Conference*, 2015. Paper No. AIAA 2015-2848, doi:10.2514/6.2015-2848.

[126] Rao, A. H. and Marais, K., "Comparing Hazardous States and Trigger Events in Fatal and Non-fatal Helicopter Accidents," in *16th AIAA Aviation Technology, Integration, and Operations Conference*, 2016. Paper Number: AIAA 2016-3916, doi:10.2514/6.2016-3916 .

[127] Rao, A. H. and Puranik, T. G., "Retrospective Analysis of Approach Stability in General Aviation Operations," in *18th AIAA Aviation, Technology, Integration, and Operations Conference, Atlanta, GA. June*, 2018.

[128] Reason, J., "Safety Paradoxes and Safety Culture," *Injury Control and Safety Promotion*, vol. 7, pp. 3–14, Mar. 2000. doi:10.1076/1566-0974(200003)7:1;1-V;FT003.

[129] Refaeilzadeh, P., Tang, L., and Liu, H., "Cross-validation," in *Encyclopedia of database systems*, pp. 532–538, Springer, 2009. doi:10.1007/978-0-387-39940-9_565.

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

[131] Romero, R., Summers, H., and Cronkhite, J., "Feasibility Study of a Rotorcraft Health and Usage Monitoring System (HUMS): Results of Operator's Evaluation.," tech. rep., National Aeronautics and Space Administration, 1996. DOT/FAA/AR - 95/50.

[132] Roskam, J., *Airplane Design, Part VI: Preliminary Calculation of Aerodynamic, Thrust and Power Characteristics*. Kansas: Roskam Aviation and Engineering corporation, 1987.

[133] Roy, C. J. and Oberkampf, W. L., "A Comprehensive Framework for Verification, Validation, and Uncertainty Quantification in Scientific Computing," *Computer methods in applied mechanics and engineering*, vol. 200, no. 25, pp. 2131–2144, 2011. doi:10.1016/j.cma.2011.03.016.

[134] Russell, P. and Pardee, J., "Joint Safety Analysis Team-CAST Approved Final Report Loss of Control JSAT Results and Analysis," *Commercial Aviation Safety Team, Washington, DC*, 2000.

[135] Rutowski, E. S., "Energy Approach to the General Aircraft Performance Problem," *Journal of the Aeronautical Sciences (Institute of the Aeronautical Sciences)*, vol. 21, no. 3, pp. 187–195, 2012. doi:10.2514/8.2956.

[136] Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., and Williamson, R. C., "Estimating The Support of a High-dimensional Distribution," *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001. doi:10.1162/089976601750264965.

[137] Schwarz, G., "Estimating the Dimension of a Model," *The annals of statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[138] Sembiring, J., Drees, L., and Holzapfel, F., "Extracting Unmeasured Parameters Based on Quick Access Recorder Data Using Parameter-estimation Method," in *AIAA Atmospheric Flight Mechanics (AFM) Conference*, pp. 19–22, 2013.

[139] Shappell, S., Hackworth, C., Holcomb, K., Lanicci, J., Bazargan, M., Baron, J., Iden, R., and Halperin, D., "Developing Proactive Methods for General Aviation Data Collection," tech. rep., Federal Aviation Administration, 2010. DOT/FAA/AM-10/16, url: `http://www.dtic.mil/dtic/tr/fulltext/u2/a534693.pdf`.

[140] Sherry, L., Wang, Z., Kourdali, H. K., and Shortle, J., "Big Data Analysis of Irregular Operations: Aborted Approaches and their Underlying Factors," in *Integrated Communications, Navigation and Surveillance Conference (ICNS), 2013*, pp. 1–10, IEEE, 2013. doi:10.1109/ICNSurv.2013.6548548.

[141] Shetty, K., "Current and Historical Trends in General Aviation in the United States," tech. rep., MIT International Center for Air Transportation (ICAT), 2012. url: `http://hdl.handle.net/1721.1/72392`.

[142] Shish, K. H., Kaneshige, J., Acosta, D. M., Schuet, S., Lombaerts, T., Martin, L., and Madavan, A. N., "Trajectory Prediction and Alerting for Aircraft Mode and Energy State Awareness," in *AIAA SciTech Forum*, Jan. 2015. Paper No. AIAA-2015-1113, doi:10.2514/6.2015-1113.

[143] Smart, E. and Brown, D., "A Two-Phase Method of Detecting Abnormalities in Aircraft Flight Data and Ranking their Impact on Individual Flights," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1253–1265, 2012. doi:10.1109/TITS.2012.2188391.

[144] Subramanian, S. V. and Rao, A. H., "Deep-learning Based Time Series Forecasting of Go-around Incidents in the National Airspace System," in *2018 AIAA Modeling and Simulation Technologies Conference*, 2018. Paper Number: AIAA 2018-0424, doi: doi:10.2514/6.2018-0424.

[145] Tadema, J. and Theunissen, E., "Design of an Integrated Traffic, Terrain and Energy Awareness Display Concept for UAVs," in *AIAA Guidance, Navigation, and Control Conference*, 2009. Paper No. AIAA-2009-5980, doi:10.2514/6.2009-5980.

[146] Takahashi, T., "Aircraft Concept Design Performance Visualization Using an Energy-maneuverability Presentation," in *AIAA Aviation Technology, Integration, and Operations (ATIO) Conferences*, Sept. 2012. Paper No. AIAA-2012-5704, doi:10.2514/6.2012-5704.

[147] Van den Hoven, M., De Jong, P., Borst, C., Mulder, M., and Van Paassen, M., "Investigation of Energy Management during Approach: Evaluating the Total Energy-based Perspective Flight-path Display," in *AIAA Guidance, Navigation, and Control Conference*, August 2010. Paper No. AIAA-2010-8401, doi:10.2514/6.2010-8401.

[148] Vormer, F., Mulder, M., Paassen, R. V., and Mulder, J., "Optimization of flexible approach trajectories using a genetic algorithm," *Journal of aircraft*, vol. 43, no. 4, pp. 941–952, 2006. doi:10.2514/1.13609.

[149] Wang, Z., Sherry, L., and Shortle, J., "Airspace Risk Management using Surveillance Track Data: Stabilized Approaches," in *2015 Integrated Communication, Navigation and Surveillance Conference (ICNS)*, pp. W3–1–W3–14, April 2015.

[150] Williams, D. H., "Flight Deck Merging and Spacing and Advanced FMS Operations," in *The 8th EWG Operations Workshop*, no. 1, 2008. URL: `https://smartech.gatech.edu/handle/1853/34339`.

[151] Williams, D. H., Oseguera-Lohr, R. M., and Lewis, E. T., "Energy Navigation: Simulation Evaluation and Benefit Analysis," tech. rep., NASA Technical Report (NASA/TP-2011-217167), 2011. URL: `https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20110014508.pdf`.

[152] Williams, D., Osequera-Lohr, R., and Lewis, E., "Design and testing of a low noise flight guidance concept," tech. rep., NASA Technical Report (NASA/TM-2004-213516), 2004. URL: `https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20050040769.pdf`.

[153] Witczak, M., *Modelling and Estimation Strategies for Fault Diagnosis of Non-linear Systems: from Analytical to Soft Computing Approaches*, vol. 354. Springer Science & Business Media, 2007.

[154] Wolpert, D. H. and Macready, W. G., "No Free Llunch Theorems for Optimization," *IEEE transactions on evolutionary computation*, vol. 1, no. 1, pp. 67–82, 1997. doi:10.1109/4235.585893.

[155] Wynnyk, L., Lunsford, C. R., Tittsworth, J. A., and Pressley, S., "Development of approach and departure aircraft speed profiles," *Journal of Aircraft*, vol. 54, no. 1, pp. 94 – 103, 2016. doi:10.2514/1.C033708.

[156] Zagalsky, N. R., "Aircraft Energy Management," in *Aerospace Sciences Meeting*, American Institute of Aeronautics and Astronautics, Jan. 1973. Paper No. AIAA-73-228, doi:10.2514/6.1973-228.