AIAA AVIATION Forum
12-16 June 2023, San Diego, CA and Online
AIAA AVIATION 2023 Forum

10.2514/6.2023-3548

Capturing Multivariate Time Series Interactions to Detect High-Risk Instability During Approach

Ezequiel Juarez Garcia*, Markus L. Mulvihill†, and Mark S. Kharab‡
University of Florida, Gainesville, FL, 32611

Chad L. Stephens§
NASA Langley Research Center, Hampton, VA 23681

Nicholas J. Napoli¶‖
University of Florida, Gainesville, FL, 32611

With the projected increase in passenger load factor and rollout of more autonomous systems into the national airspace, the need to detect high-risk events in-time or ahead-of-time is becoming increasingly crucial. New anomaly detection and precursor identification algorithms will need to scale to different airframes, levels of autonomy, and system complexity. While the pervasiveness of deep learning has resulted in the development of performant anomaly detection methods, these sophisticated models currently suffer from low end-user interpretability. Building off our previous work on identifying adverse events in multivariate flight data during descent, we propose a data-driven approach for detecting in-flight adverse events caused by the complex interplay of flight variables. Our approach utilizes ordinal patterns of important aircraft stability variables (e.g., airspeed and descent rate) to capture multivariate flight dynamics that can be used to predict the onset of unstable approaches, a high-risk adverse event that can occur during approach. Through the use of ordinal patterns, we aim to create more interpretable detection models of in-flight adverse events that can be used to gain insight on future autonomous systems and make model translation to different airframes less difficult. Our analysis shows the presence of distinct ordinal pattern distributions that can be used to predict unstable approaches 1 minute ahead-of-time with an accuracy of 0.69 and a recall of 0.73 and 30 seconds ahead with an accuracy of 0.70 and a recall of 0.86.

I. Introduction

Despite efforts by aviation agencies to continue driving down safety metrics beyond target thresholds, high-risk events such as instability during landing approach remain a present danger to passenger safety. Approach instability and other in-flight adverse events, or anomalies, possess hidden signatures that may go undetected until its offset, where the risk to aircraft integrity and passenger safety increases substantially. To identify adverse events ahead-of-time, a shift away from univariate threshold exceedance methods has resulted in the proliferation of more sophisticated anomaly detection and prediction methods [1–4]. While an improvement over univariate techniques, these newer methods tend to lack the interpretability of exceedance-based methods, resulting in poor actionable information and intuition for the end-users, such as pilots and air traffic services. As air traffic continues to increase [5, 6] and more types of aircraft with varying degrees of autonomy are introduced into the national airspace system (NAS) [7–9], the need for not only robust but also interpretable adverse event detection becomes evident. Detection models with high interpretability allow the users or decision makers to associate specific aircraft states to normal or anomalous aircraft operation and make informed decisions. This direct link to the underlying state of the aircraft enhances the generalizability of an

*Graduate Research Assistant and L3Harris Fellow in the Human Informatics and Predictive Performance Optimization (HIPPO) Laboratory, Department of Electrical and Computer Engineering, AIAA Student Member (e-mail: ejuarezgarcia@ufl.edu).

†Undergraduate Researcher in the Human Informatics and Predictive Performance Optimization (HIPPO) Laboratory, Department of Electrical and Computer Engineering (e-mail: markus.mulvihill@ufl.edu).

‡Undergraduate Researcher in the Human Informatics and Predictive Performance Optimization (HIPPO) Laboratory, Department of Electrical and Computer Engineering (e-mail: mark.kharab@ufl.edu).

§Researcher, Crew Systems & Aviation Operations Branch and System-Wide Safety Project (e-mail: chad.l.stephens@nasa.gov).

¶Assistant Professor and Director of the Human Informatics and Predictive Performance Optimization (HIPPO) Laboratory, Department of Electrical and Computer Engineering, AIAA Senior Member (e-mail: n.napoli@ufl.edu).

interpretable paradigm, allowing users to translate their model knowledge to future airframes with reduced retraining and difficulty. To this end, this study focuses on the design of an interpretable method for detecting adverse events ahead-of-time by analyzing the multivariate state of the aircraft, rather than the single-signal analysis of traditional methods. We establish the viability of our method by applying it to instability prediction during descent, where we provide a straightforward prediction modeling approach without relying on complex machine learning (ML) models.

Prior Work. Work in adverse event detection can be grouped into two categories: in-time and ahead-of-time detection. Traditional in-time detection relies on threshold exceedance and simple statistical comparisons of individual variables (univariate) to identify anomalous aircraft conditions [10]. While being highly interpretable, these methods fail to scale to the more elaborate patterns of failure of today's complex aviation systems [11, 12]. Moving away from exceedance-based methods while retaining model interpretability, our work in [4] has shown the applicability of entropic methods, which quantify the regularity or predictability of a signal, in detecting unstable approaches. According to the FAA, an approach is considered unstable when at least one the variables in [13] (e.g., airspeed or vertical speed) linked to stability exceeds a predefined threshold. Failure to maintain a stable approach can result in accidents such as rough landings, runway excursions, and terrain or infrastructure collision [14]. Due to the prevalence of accidents during approach and landing [15], the detection of approach instability is vital for airlines and manufacturers to continue reducing the number of in-flight accidents. Despite the pertinent work in anomaly detection in [4], its design does not provide a discrete, multivariate view of the aircraft's state, which can enhance the interpretability of the flight dynamics associated with an anomaly. The need to track how all relevant variables are interacting and evolving during an unstable approach to improve detection motivates the following question: *How can flight dynamics be captured by multivariate entropic analysis methods and linked to approach instability?*

In ahead-of-time adverse event detection, a growing trend over the last decade has been the identification of precursors from multivariate flight data. Precursor are key events that manifest prior and are associated with the onset of adverse events [16]. A precursor can be the adverse interplay between multiple flight variables, such as high airspeed and incorrect flap position prior to landing, or a single failure event like a sensor malfunction or variable threshold exceedance. The ability to predict an adverse event, like an unstable approach, using precursors can reduce the severity of the adverse event or potentially prevent its onset through corrective action. At the forefront of precursor mining methods are deep learning (DL) models. Recently, work in [17] has demonstrated the effectiveness of using multiple-instance learning, a form of weakly-supervised ML, and deep neural networks to identify precursor events. Similar DL methods such as [16, 18] have also demonstrated progress towards interpretable deep learning models. However, despite their pervasiveness, DL methods lack the level of precursor or feature interpretability commonly found in traditional ML and data analysis. A certain level of interpretability is required to not only inform the user of an imminent risk, but also provide the information and means for the user to take corrective action. Therefore, the multivariate nature of anomalies must be summarized in a way that is presentable and understandable to pilots and air traffic services. The prediction of adverse events using multivariate precursors raises another question: *Can the features captured by multivariate entropic analysis methods be used to create more interpretable prediction models of unstable approaches?*

Challenges: One of the leading challenges with adverse event detection is determining the cause of the adverse event. Due to potentially intricate and multivariate nature of failure patterns in today's aircraft, univariate exceedance-based methods are becoming less and less capable of detecting complex interactions in the flight dynamics. Nowhere is this is more evident than in the Lion Air fatal accident due to a design flaw in the Boeing 737 MAX's Maneuvering Characteristics Augmentation System (MCAS). The MCAS was designed to stabilize the aircraft at high angles of attack by pitching the aircraft down. Human factors aside, the aircraft exhibited series of precursor events, such as sensor readout differences in airspeed and altitude, prior to the catastrophic crash [19, 20]. This type of cascading, multivariate aircraft failure is difficult to frame in terms of threshold-exceedance monitoring. A different detection paradigm is therefore required to monitor the state of multiple aircraft variables and link the state with a high-risk condition of the aircraft. Despite their ability to handle multivariate flight data, DL models suffer from reduced interpretability compared to traditional anomaly detection methods. A high level of model interpretability allows the end-user to link the risk of adverse event to the underlying state of the aircraft in terms of an intuitive physical quantity, rather than something abstract. The low interpretability of DL models may pose a challenge when implementing them in actual safety systems, where model interpretability may be deemed a more important factor than performance. Moreover, despite progress in interpretable DL in aviation applications [21], DL models may not be able to seamlessly translate to other, present and future, aircraft. This is due to their need to be trained on specific patterns of failure of an aircraft using large quantities of training data. Therefore, non-ML models can often lend themselves to not only higher interpretability, but also better generalizability, due to their simpler design and intuition of their inner workings.

Insights. As the complexity of aviation systems continues to grow, a shift away from traditional exceedance-based

methods for in-flight risk identification is necessary. Despite the popularity and performance of DL models, there is still room for improvement, especially in model interpretability [22]. An alternative approach to flight data analysis, one that does not depend on sophisticated ML, comes from the field of entropic time series analysis. Entropic analysis relies on the notion of Shannon entropy [23] to quantify the degree of complexity in a dynamical system. Complexity is the degree of order or regularity present in a signal [24, 25]. Signals that exhibit discernible and often repeating patterns tend to have a relatively low complexity compared to signals with highly irregular patterns. Our previous work on anomaly detection has demonstrated the viability of using entropy to build explanatory models for unstable approaches. To extend our prior work to multivariate problems and capture the complex interactions between multiple flight variables, we can analyze the ordinal patterns present during and prior to the occurrence of an adverse event. Ordinal patterns are a tool from entropic analysis that capture the ordering, or ranking, of multiple signals at any point in time. The illustration of ordinal patterns in Fig. 1 shows the ranking of three signals, which provide a discrete snapshot of the multivariate state of the system. The use of ordinal patterns on flight data can reveal hidden aircraft states that have a high likelihood of appearing prior to an adverse event or during its onset. Importantly, ordinal patterns possess a high level of interpretability due to the plain variable rankings that are summarized within them. These patterns can be used to create simple predictive models that maintain the same level of interpretability of the patterns and allow us to gain insight on high-risk aircraft states. The identification of generalizable high-risk states in today's aircraft has important implications on the design considerations of future aircraft operating with reduced human intervention.

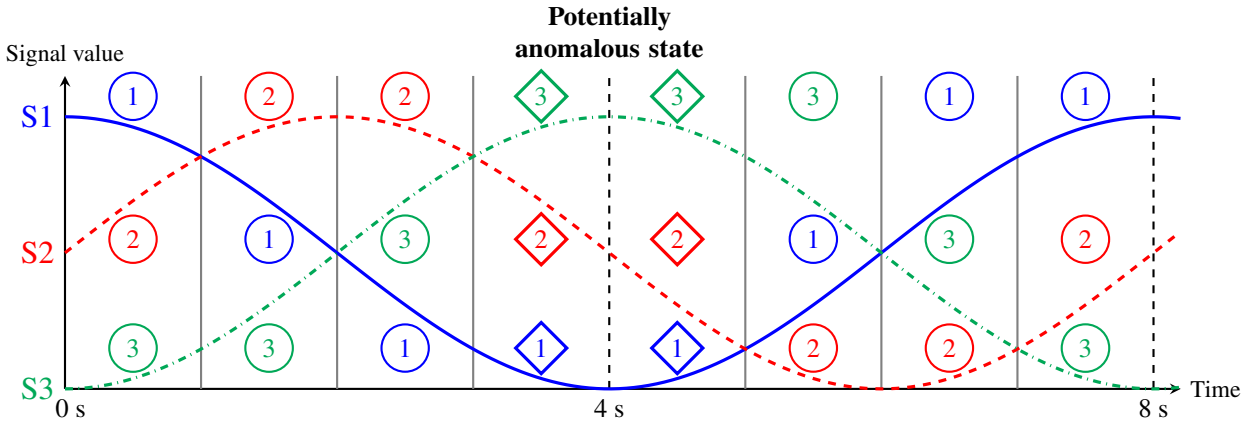


Fig. 1 Explanatory figure on ordinal patterns. Ordinal patterns provide a simple approach to modeling the multivariate interactions of flight variables by looking at how signals are ranked, or ordered, at any point in time. The specific variable interactions between 3 and 5 seconds (i.e., the ordinal pattern (1, 2, 3) when S1 ≤ S2 ≤ S3), may indicate the presence of an adverse event or its precursor. This and other ordinal patterns may manifest prior to and during adverse events. The occurrence of an ordinal pattern may suggest the current or future presence of an adverse event.

Contributions. Using ordinal patterns, our aim is to create simple, highly-interpretable prediction models of unstable approaches. More specifically, our work in this study provides the following contributions to interpretable multivariate analysis and prediction of unstable approaches:

- We provide the means for measuring the distance between two ordinal patterns to quantify the degree of similarity of the patterns. This is used to explore differences in the flight dynamics of stable and unstable approaches.
- We show that ordinal patterns not only provide an interpretable method to characterize the multivariate flight dynamics of an aircraft, but they also serve as precursors of unstable approaches. Precursor patterns reveal aircraft states that have a higher likelihood of occurring in stable approaches.
- Using Bayesian analysis and ordinal patterns, we provide a predictive modeling methodology capable of detecting approach instability up to a minute ahead-of-time, while maintaining a high level of model interpretability. Predictions of future instability can alert or provide decision support to pilots and air traffic services to bring the aircraft back into a stable configuration.

The above findings allow us to gain algorithmic insight into what variable interactions are most important in identifying and preventing adverse events during approach and landing. This analysis can help us better understand high-risk states of current and future autonomous aircraft. Details of our methodology are provided in the following Methods section.

II. Methods

The crux of our methodology is the identification of ordinal patterns (summarized in Fig. 1) in multivariate, heterogeneous flight data during periods during and prior to the occurrence of an adverse event. These patterns are then used to create models capable of predicting aircraft instability during approach. We refer to heterogeneous flight data as a collection of flight signals where each signal has a different set of possible values and value ranges. For example, the range and values of an aircraft's airspeed is different from the values of a landing gear binary switch that is engaged on approach. Extra care and processing must be taken to make meaningful comparisons between the values of the two signals. A time series normalization technique is discussed later in this paper that brings the analyzed flight variables to a common value range. The ordinal patterns are extracted from flight data starting at five minutes prior to landing to when the aircraft makes contact with the runway. As previously mentioned, unstable approaches are a leading adverse event that can occur during the approach and landing phases of an aircraft. Due to the high danger they pose to human life, unstable approaches are the main focus of our analysis.

By characterizing the ordinals that manifest during periods of instability, our aim is to link the presence of certain ordinal patterns to periods of instability during the flight. This characterization will help us build simple and interpretable predictive models that can identify an unstable approach prior to the stability check window that occurs between 500 and 1,000 ft above the runway, as shown in Fig. 2. Before creating these predictive models, however, we must first perform other tasks, such as the categorization of flights as stable or unstable. This is accomplished through the use of traditional threshold exceedance criteria provided by aviation regulatory agencies [13]. This is followed by the extraction of ordinal patterns during periods of instability. Using this, we provide explanatory analysis on the link between ordinal patterns and the unstable dynamics of an aircraft. This analysis aids us in building predictive models to detect instability at various time checkpoints prior to the aircraft's descent below 1,000 ft.

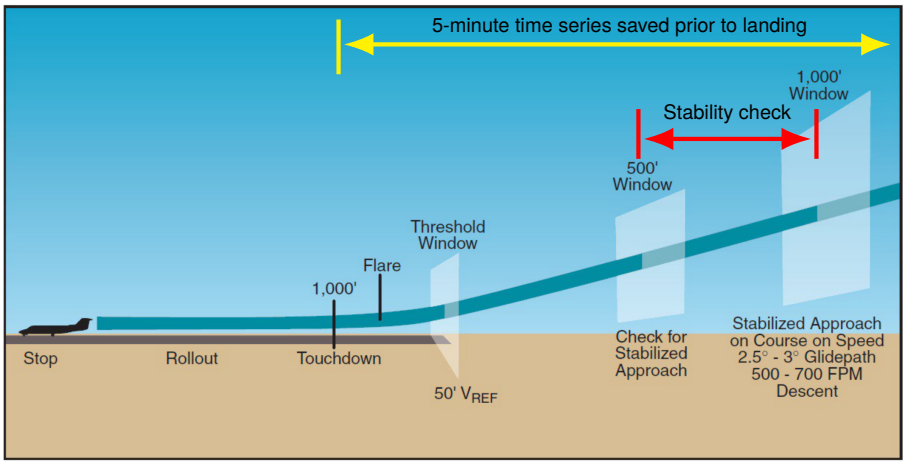


Fig. 2 Descent profile of a stabilized approach as per the FAA [26]. For a flight with successful touchdown, the determination of an unstable approach is made during the *Stability check* flight window marked in red. The ordinal patterns we analyzed were collected over the 5-minute flight window prior to touchdown marked in yellow.

A. Identification of Unstable Approaches

Before explaining the criteria used to identify unstable approaches, we provide an overview of the dataset used and the preprocessing performed on the flight data.

1. Dataset Overview and Preprocessing

The publicly available NASA DASHlink Sample Flight Data [27], hereafter referred to as the *flight data* or, simply, the *dataset*, provides anonymized flight data from a domestic regional airline. The dataset contains flight data from 35 regional jets, also referred to as tails, recorded over the time span of 2001-2004. The manufacturer and model of the aircraft are not provided in the dataset. Each tail comprises approximately 5,000 individual flight files on average, with a grand total of 180,000 flights in the entire dataset. Each flight file is saved in the MATLAB (The MathWorks,

Inc.) MAT binary format and contains the time series of 186 different flight variables such as airspeed, flap position, and flight phase*. With the exception of wind speed and wind shear alerts, no other environmental flight variables are provided in the dataset. The sampling rate of the signals varies from 0.25–16Hz, with a median sampling frequency of 4 Hz. Unless otherwise noted, the flight variables analyzed in this study were resampled at a rate of 4 Hz using previous value interpolation to maintain a consistent sampling rate across all flight variables. After resampling the signals to a common sampling rate, a 10 second (40 samples at 4 Hz) moving median filter was applied to the data. The median filter provided light noise filtering of the data while, more importantly, not altering the possible values of integer-valued signals, such as the weight-on-wheels signal with values of either 0 or 1.

Following the filtering of the flight signals, several condition checks were applied to each flight to determine if the flight data was valid. Invalid flights were discarded from the dataset and valid flights were kept for subsequent processing. This further processing is discussed in later sections. A flow chart of the logic used to determine valid flights is provided in Fig. 3. This process resulted in the elimination of approximately 30,000 flights out of the original 180,000. Following this elimination process, valid flights were searched for their touchdown point or, in the case of a rejected landing, the point of the aircraft's lowest height above ground level (AGL) before commencing a go-around. With this information, we proceeded to the identification of unstable approaches.

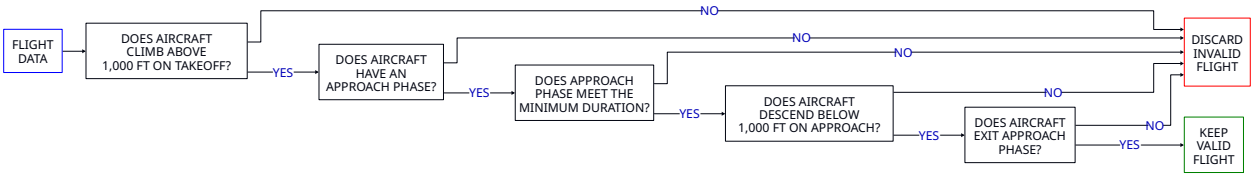


Fig. 3 Flight data validation logic. Approximately 30,000 flights (out of total of 180,000) were discarded using these condition checks.

2. Instability Criteria and Identification of Unstable Approaches

The determination of the stability of a flight during approach was made using FAA guidelines on stabilized approaches [13]. According to the FAA [26], a stabilized approach is "one in which the pilot establishes and maintains a constant angle glidepath towards a predetermined point on the landing runway." Translating the stabilization criteria to threshold exceedances resulted in searching for the exceedance thresholds in our curated flight data. These thresholds were derived from aviation agency recommendations [28, 29] and other literature [30]. The stabilization criteria variables and their threshold values are summarized in Table 1. These variables are computed airspeed (CAS), inertial vertical speed (IVV), localizer deviation (LOC), and glideslope deviation (GLS). Glideslope deviation tracks the vertical misalignment from a 3-degree angle between the flight path and the runway, while localizer deviation tracks horizontal misalignment with the runway centerline. One key omission in the FAA stabilization criteria is the duration that the instability must be sustained for it to register as an exceedance. This causes a problem when values close to a threshold cross over momentarily due to noisy sensor readings. To prevent noisy exceedances in our search for unstable flights, we discarded threshold exceedances sustained for less than 5 seconds. The search for exceedances took place between the aircraft's 1,000 and 500 ft altitude descent crossings. In the case of a rejected landing, the lowest height above ground level before go-around marked the end of the stability check window. This time window for a successful landing is visualized in Fig. 2.

Table 1 Exceedance events analyzed in this study during aircraft descent. An exceedance of any flight variable results classification of the approach as unstable.

Exceedance Event	Flight Variable	Variable Description	Condition
Airspeed exceedance	CAS	Computed airspeed	CAS < 110 or > 150 knots
High descent rate during approach	IVV	Inertial vertical velocity	IVV < -1,000 FPM
Localizer deviation exceedance	LOC	Localizer deviation	Magnitude > 3 SD
Glideslope deviation exceedance	GLS	Glideslope deviation	Magnitude > 3 SD

*More information on the dataset can be found on the dataset's resources page at `https://c3.ndc.nasa.gov/dashlink/resources/901/`.

B. Identification of Ordinal Patterns and Similarity

Ordinal patterns provide information on the state of the aircraft, by encoding the arrangement or ordering of multiple flight signals at any point in time (see Fig. 1). Symbolic representations of the signals, rather than the signal values themselves, are stored in a least-to-greatest sorting in the patterns. A change in the ordering creates a different ordinal pattern. Instead of relying on pure thresholds exceedances that may be statistically but not operationally anomalous, by capturing ordinals patterns we can provide a multivariate view of how all flight variables of interest are evolving with respect to each other. This motivates our use of ordinal patterns as a multivariate feature that maintains the interpretability of the system. The following sections explain our approach to identifying ordinal patterns during descent and landing of stable and unstable flights. Moreover, we discuss the metric used to define similarity between ordinal patterns.

1. Data Normalization

Before identifying the ordinal patterns present during an aircraft approach, we first addressed the lack of a common value range for the signals. That is, each signal is the heterogeneous flight data can take on a different range of values throughout the flight. For example, the typical range of airspeed signal is anywhere between 0 and 500 knots, while the range of the integer-valued landing gear down signal is 0 to 1. Naturally, this leads to the problem where airspeed will be greater than landing gear down over the entire approach phase, not accounting for noise. To address the lack of a common value range, we used a normalization technique where each signal is mapped to a range approximately between 0 and 1. The range is approximate because some signals will be mapped outside this range. This is a consequence of the normalization formula

$$x_{\text{norm}} = (x - x_{\text{min}})/(x_{\text{max}} - x_{\text{min}}). \tag{1}$$

In the equation above, x is a value in the time series of a flight variable, x_{min} is the 1st percentile minimum value of the flight variable, and x_{max} is the 99th percentile maximum of the same flight variable. The values for x_{min} and x_{max} were calculated prior to the normalization using all the available flight data. Therefore, they are not the minimum and maximum values of the time series being normalized. We selected the 99th percentile highs and 1th percentile lows to prevent outlier extremum values from reducing the usable range between 0 and 1. Consequently, these outlier values may lie outside the [0, 1] interval. For example, if we obtained the values $x_{\text{min}} = 0$ and $x_{\text{max}} = 500$ (knots) from the entire ground speed flight data, then a time series with values ranging from 0 to 400 knots would be mapped down to the unitless range [0, 0.8]. This simple normalization formula allows us to map all the signals to a comparable range while decreasing the effects of outliers on the usable range.

2. Time Series Extraction and Pattern Identification

With the normalized flight variables, we extracted a 5-minute time series of all flight variables prior to a successful or rejected landing (see visualization in Fig. 2). At five minutes before landing, an aircraft is approximately 12 nautical miles (NM) away from touchdown and 3,500 ft above ground level (see Fig. 4). Distances greater than 12 NM put the aircraft past the effective range of the airport antennas used in the calculation of glideslope and localizer deviation, two important variables that determine instability. At a uniform sampling frequency of 4 Hz for all flight variables, this resulted in time series with a length of 1200 samples. Using Fig. 1 as a flight data example, the ordinal patterns tuples (a, b, c) are organized as rows in an $m \times n = 8 \times 3$ matrix

$$\mathbf{\Pi} = \begin{bmatrix} 3 & 2 & 1 \\ 3 & 1 & 2 \\ 1 & 3 & 2 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \\ 2 & 1 & 3 \\ 2 & 3 & 1 \\ 3 & 2 & 1 \end{bmatrix},$$

where m is the number time steps and n is the number of flight variables. The ordering of each pattern was calculated by sorting the values of the multivariate signal at each time step and saving the integer value (i.e., the symbols 1, 2, and 3) assigned to each signal. Ambiguity in the order of equal values can be resolved using any tiebreaker method, so long

as the use of the method remains consistent. Our tiebreaker method was assigning a lower rank (i.e., position in the ordering) to the first equal value in a pattern vector, moving from left to right.

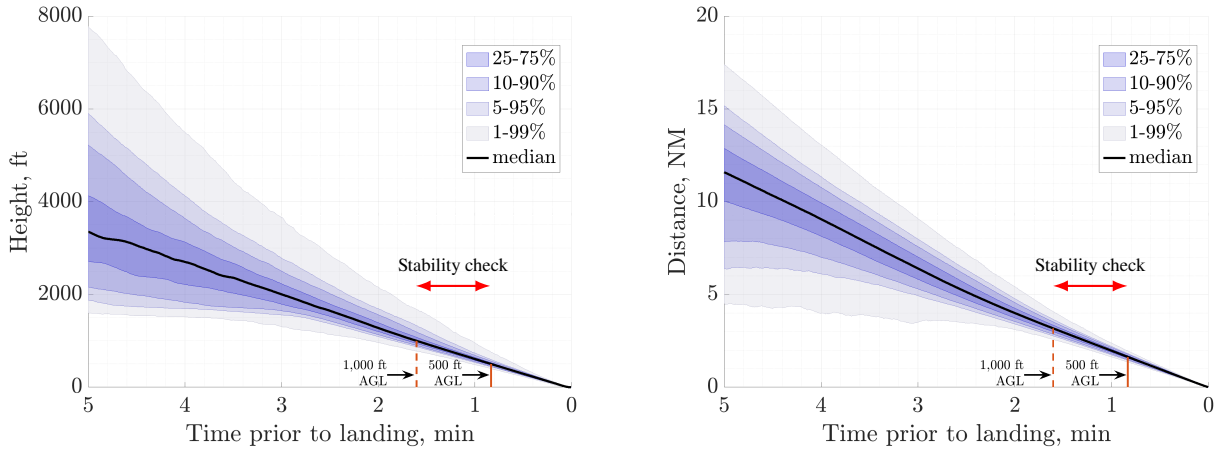
By assigning a symbol π_i to each ordinal pattern, we can rewrite the matrix of ordinal patterns as

$$\mathbf{\Pi} = \begin{bmatrix} \pi_6 & \pi_5 & \pi_2 & \pi_1 & \pi_1 & \pi_3 & \pi_4 & \pi_5 \end{bmatrix}^T.$$

Here, $\pi_1 = (1,2,3)$, $\pi_2 = (1,3,2)$, $\pi_3 = (2,1,3)$, $\pi_4 = (2,3,1)$, $\pi_5 = (3,1,2)$, and $\pi_6 = (3,2,1)$. A valid set of symbols that can be used in computations are $\{\pi_i = i\}_{i=1}^{6}$, so that

$$\mathbf{\Pi} = \begin{bmatrix} 6 & 5 & 2 & 1 & 1 & 3 & 4 & 5 \end{bmatrix}^T.$$

This representation of ordinal patterns is useful for storage in computer memory and visualization, as seen later on in this paper.



(a) (b)

Fig. 4 Percentile distribution of (a) height above ground level (AGL) during approach and (b) distance away from touchdown point. Based on the median, an aircraft crosses the 1,000 ft and 500 ft AGL thresholds at approximately 96 seconds (3.2 NM away) and 50 seconds (1.6 NM away) prior to landing, respectively. This results in a stability check window approximately 45 seconds long.

3. Quantifying Similarity of Ordinal Patterns

Attempting to compare two ordinal patterns leads to the natural question of, *how can we quantify the distance or degree of similarity between two ordinal patterns?* Due to the arbitrary symbolic representation of a signal in an ordinal pattern, integers in this study, a distance metric such as Euclidean distance is not ideal. Motivated by the work in string similarity [31], we define the distance between two equal-length ordinal patterns π_i and π_j using the following recursive implementation

$$\text{dist}(\pi_i, \pi_j) = \begin{cases} 0, & |\pi_i| = |\pi_j| = 0 \\ \text{dist}\left(\text{tail}(\pi_i), \text{tail}(\pi_j)\right), & \pi_i(1) = \pi_j(1) \\ 1 + \text{dist}\left([\text{tail}(\pi_i), \pi_i(1)], \pi_j\right), & \text{otherwise.} \end{cases} \tag{2}$$

In this equation, $\pi(1)$ is the first element of π, $|\pi|$ is the length of π, $[\pi_1, \pi_2]$ is the concatenation of two patterns, and $\text{tail}(\pi)$ is the ordinal pattern excluding the first element. This distance metric computes the number of index shifts required to make π_i equal to π_j. For example, the distance between $\pi_1 = (1,3,4,2)$ and $\pi_2 = (1,2,3,4)$ is 2, due to the two index shifts to the left required to move 2 from the fourth position to the second position in π_1. Normalizing the value of Eq. (2) by its upper bound equal to $\binom{m}{2}$, where $m = |\pi_i| = |\pi_j|$, we obtain the normalized distance metric

$$\overline{\text{dist}}(\pi_i, \pi_j) = \frac{\text{dist}(\pi_i, \pi_j)}{m} \tag{3}$$

Using normalized distance, we define the similarity metric between two ordinal patterns π_i and π_j as

$$\text{sim}(\pi_i, \pi_j) = 1 - \overline{\text{dist}}(\pi_i, \pi_j). \tag{4}$$

Given m-length patterns, similarity can take on $\binom{m}{2} + 1$ values on the interval [0, 1], with a spacing of $1/\binom{m}{2}$ between successive values. Unless otherwise noted, we refer to Eq. (3) when we discuss the distance between ordinal patterns. In the example with $\pi_1 = (1, 3, 4, 2)$ and $\pi_2 = (1, \underline{2, 3}, 4)$, Eq. (4) results in a similarity value of $1 - 2/6 \approx 0.67$. Maximum distance (i.e., minimum similarity) is given by $\overline{\text{dist}}\big((1, 2, 3, 4), (4, 3, 2, 1)\big)$, where 6 index shifts are required to turn the first pattern into the second one, resulting in a similarity of $1 - 6/6 = 0$.

C. Explanatory Analysis and Predictive Modeling

Our explanatory analysis consists of comparing the similarity between ordinal patterns that occur during unstable approaches (unstable patterns) and those that occur during stable approaches (stable patterns). These comparisons are made during the aircraft's stability check window between 1,000 and 500 ft above the runway (refer to Fig. 2 & 4) when an exceedance event is registered. Using Eq. (4) to calculate similarity between stable and unstable patterns, we then compute the mean and standard deviation of the similarity values to determine the degree of similarity. With this information, we can establish whether unique ordinal patterns are present during unstable approaches. In addition to this explanatory analysis, we seek to predict the onset of instability before its actual determination during the flight's stability check window. By comparing the probabilities of a flight being stable or unstable, using Bayesian analysis, we perform a binary classification of the aircraft's stability, where unstable is the positive class and stable is the negative class. These predictions are made at various time checkpoints prior to the aircraft's descent below 1,000 ft. Predictions on future instability can alert the pilot and air traffic control (ATC) ahead-of-time to perform corrective action to stabilize the flight.

More specifically, we leverage the flight stability distributions, either stable or unstable, conditioned on the ordinal patterns. These conditional stability distributions constitute the *likelihoods* in Bayesian analysis and are formally expressed as $P(x|\pi_j)$. In this notation, π_j represents the ordinal pattern assigned the integer j (see Table 3), and $x \in \{\text{stable, unstable}\}$ is the stability of the aircraft. Two likelihoods, one for stable and another for unstable, are generated from a training set consisting of 8,700 stable flights and 8,700 unstable flights. The underpinning equation that takes advantage of these likelihoods is Bayes' rule:

$$P(\pi_j|x) = \frac{P(x|\pi_j)P(\pi_j)}{P(x)}. \tag{5}$$

Informally, we can rewrite Bayes' rule as

$$posterior = \frac{likelihood \times prior}{evidence}. \tag{6}$$

This informal representation allows us to identify terms that we have not calculated yet: the *prior* and *evidence*. The prior provides the probability that we observe a specific ordinal pattern while the evidence is the probability that the aircraft is stable or unstable. The evidence term can be expanded into

$$P(x) = \sum_{i=1}^{24} P(x|\pi_i)P(\pi_i). \tag{7}$$

Our main interest in the Bayesian formulation is the evidence term. This term, however, still requires us to know the prior distribution. To solve this, we approximate the prior over non-overlapping 30-second time windows, prior to the stability check window. The end of these 30-second windows occur at 4.5, 4, 3.5, 3, 2.5, 2, and 1.5 minutes prior to landing. Thirty-second intervals strike a balance between sufficient ordinal patterns to approximate the prior (120 samples at 4 Hz) and our assumption that the likelihoods adequately model the aircraft's stability over the interval. Longer intervals would yield better prior approximations but would allow the likelihoods to change drastically, rendering them useless. At the end of each time window, we use the likelihoods and prior to calculate the probability of the aircraft being stable, $P(\text{stable})$, or unstable, $P(\text{unstable})$. The highest probability determines the stability classification of the 30-second flight window.

III. Results & Discussion

This section provides the interpretation of the results of the following research questions (RQs):

RQA Can the flight dynamics from multivariate sensor interactions be mapped to periods of instability using unique ordinal patterns?

RQB Can we predict unstable approaches using ordinal patterns for more interpretable adverse event detection?

For RQA, we hypothesize the presence of patterns that are unique to unstable approaches. These unique patterns can help us identify the precise state(s) of the aircraft during instability. Thus, we should observe similarity values between unstable and stable patterns close to zero with a tight confidence interval (e.g, standard deviation) to indicate the presence of patterns unique to approach instability. In RQB, we seek to build ahead-of-time instability prediction models with information on the presence, or absence, of ordinal patterns. Even if no unique patterns exist in unstable approaches, we hypothesize the presence of distinct pattern distributions (i.e., probability mass functions) at multiple instances throughout the descent stage that can help us build predictive models to detect unstable approaches ahead of their onset.

A. Linking Ordinal Patterns to Aircraft Instability

Before moving forward with the analysis of ordinal patterns and instability, we first break down the frequency that the exceedance events occurred in unstable flights. In total, of 34,500 out of 150,000 flights met the instability criteria of Table 1. As shown in Table 2, the most occurring exceedance event was vertical velocity (IVV) exceedance, which occurred in 66% of unstable flights. The next most frequent events, in order, were airspeed (CAS) exceedance (29%), localizer deviation (LOC) exceedance (21%), and glideslope deviation (GLS) exceedance (13%). Together, the energy management variables CAS and IVV were responsible for 69.5% of unstable approaches. Exceedance of energy management variables are known to be large contributing factors of unstable approaches [17, 30]. Therefore, our results are in alignment with what is reported in literature.

Similarity Results: To determine the link between ordinals patterns and unstable approaches, we measured the similarity between ordinal patterns in flights that experienced unstable approaches and patterns in flights without unstable approaches. Both types of patterns, stable and unstable, were collected over the stability check window that occurs between 500 and 1000 ft above the runway, as visualized in Fig. 2. In particular, unstable patterns were collected during any of exceedance event windows that could be triggered by the exceedance of a flight variable associated with an unstable approach (listed in Table 1). These variables associated with the stabilization criteria are airspeed (CAS), glideslope deviation (GLS), vertical velocity (IVV), and localizer deviation (LOC). Therefore, a maximum of four exceedance events could occur during a flight. From each exceedance event window in an unstable approach, we randomly selected 10 ordinal patterns. The same number of stable patterns were randomly sampled from a stable flight during each exceedance event window of a corresponding unstable flight. Thus, for an unstable flight that experienced all four exceedance events, 40 random patterns were drawn from a stable flight to make the pairwise similarity comparisons.

To perform the ordinal pattern similarity comparisons using the 34,500 unstable flights, an additional 34,500 stable flights were randomly sampled from the dataset. This resulted in the calculation of 345,000 of similarity values using Eq. (4) (10 values per flight). The similarity values for each exceedance event are shown in Fig. 5. In the figure, we observe the lack of any clear distinction between the similarity of stable and unstable patterns across all exceedance events linked to instability. Our hypothesis that a mean similarity close to 0, with tight one-standard deviation (SD) bounds, would have indicated the presence of unique ordinal patterns in unstable approaches is not supported by the large spread of similarity values in all four exceedance events. However, this large similarity spread suggests the presence of both similar and dissimilar patterns during each exceedance event. Therefore, despite the lack of evidence to support our original hypothesis, we may be able to exploit the presence of similar and dissimilar patterns in the predictive models for RQB. These predictive models can take advantage of the distributions of stable and unstable ordinal patterns, so long as the distributions are distinct enough prior to the onset of the instability.

Most Occurring Patterns: To shed more light on the similarity results and learn more about the ordinal patterns present during approach and landing, we extracted the most common ordinal patterns present throughout the 5-minute time window before landing. That is, at each time step, we found the most occurring (i.e., the mode) ordinal pattern. These ordinal pattern modes can reveal information on why the similarity values in Fig. 5 exhibit a large spread. Because the ordinal pattern modes occur more often then any other ordinal pattern, they will have a non-negligible influence on the similarity mean and standard deviation during the stability check window. The modes were calculated from 2,500 randomly sampled stable and unstable approaches (5,000 flights in total). To visualize these modes, we assigned an integer value to each of the possible ordinal patterns. Given the 4 flight variables related to flight instability, the number

Table 2 Frequency of exceedance events during the unstable approach of 34,500 flights.

Exceedance Event	Frequency	Percentage
Airspeed (CAS)	9,818	28.5%
Glideslope deviation (GLS)	4,400	12.8%
Vertical velocity (IVV)	22,894	66.4%
Localizer deviation (LOC)	7,121	20.7%

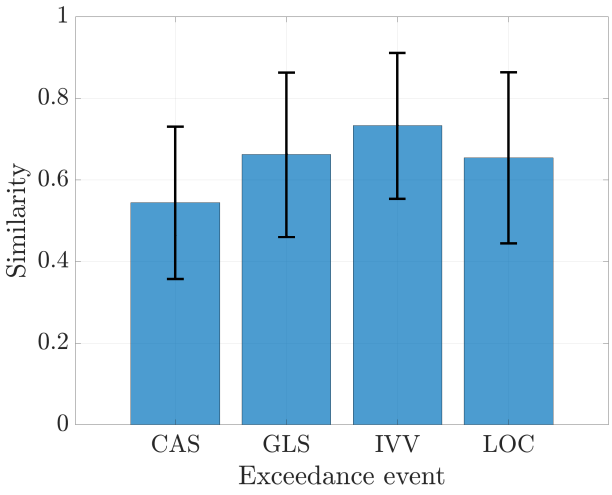


Fig. 5 Mean similarity values (± 1 SD) between unstable and stable patterns across four exceedance events. Unstable patterns are ordinal patterns found in the stability check window (refer to Fig. 2) during the exceedance of a flight variable related to instability (e.g., airspeed in Table 1). Stable patterns are those found during the stability check window of flights that did not experience variable exceedances. The four exceedance events analyzed are airspeed exceedance (CAS), glideslope deviation exceedance (GLS), vertical velocity exceedance (IVV), and localizer deviation exceedance (LOC). Similarity was calculated using the similarity metric in Eq. (4).

of possible patterns is 4! = 24. The symbolic representation of these ordinal patterns using integers is given in Table 3. For example, the normalized flight variables encoded in the pattern (1, 2, 3, 4), which is represented by the integer 1, are ranked as follows: airspeed (1) ≤ glideslope deviation (2) ≤ vertical velocity (3) ≤ localizer deviation (4). With the integer representations of 1 through 24 for the ordinal patterns, we can visualize the evolution of the modes during the five minutes before landing in Fig. 6. Here, we observe differences in the modes starting at around the 4.5-minute mark prior to landing. The differences in the modes are present until the 1-minute mark, were both stable and unstable flights maintain the same ordinal patterns until landing. Given the arbitrary symbolic representation of the ordinal patterns, the modes may be highly similar. For example, pattern (4, 3, 2, 1) could have been assigned the symbol 1 and (3, 4, 2, 1) the symbol 24. This would have suggested a low similarity between the two patterns, despite only differing in the rank of the first two variables. Therefore, we sought to quantify the actual similarity between modes to draw appropriate observations. To make sense of the similarity values, note that given ordinal patterns with a length of 4, the spacing between successive similarity values is $1/\binom{4}{2} \approx 0.167$. The evolution in stable and unstable mode similarity is shown in Fig. 6(b), where we observe a similarity of approximately 0.83 between 3 and 4.5 minutes before landing. This value is due to the change in the rank of vertical velocity (3) and localizer deviation (4) at around the 4.5-minute mark, where the two variables swap ranks. The lowest similarity between modes is found at approximately 2 minutes before landing for a brief 30-second window. This large drop in similarity is explained by airspeed (1) switching from a rank of 4 to a rank of 1, a position shift of 3 that results in similarity of 0.5. More importantly, the similarity of the modes is 0.83 at the start of stability check window and 1 at the end. This change is easily explained by the drop in rank, from 2 to 1, of the normalized airspeed (CAS) signal as shown in Fig. 6(a).

The evolution of ordinal pattern mode similarity provided in Fig. 6(b), corroborated by the changes in ordinal

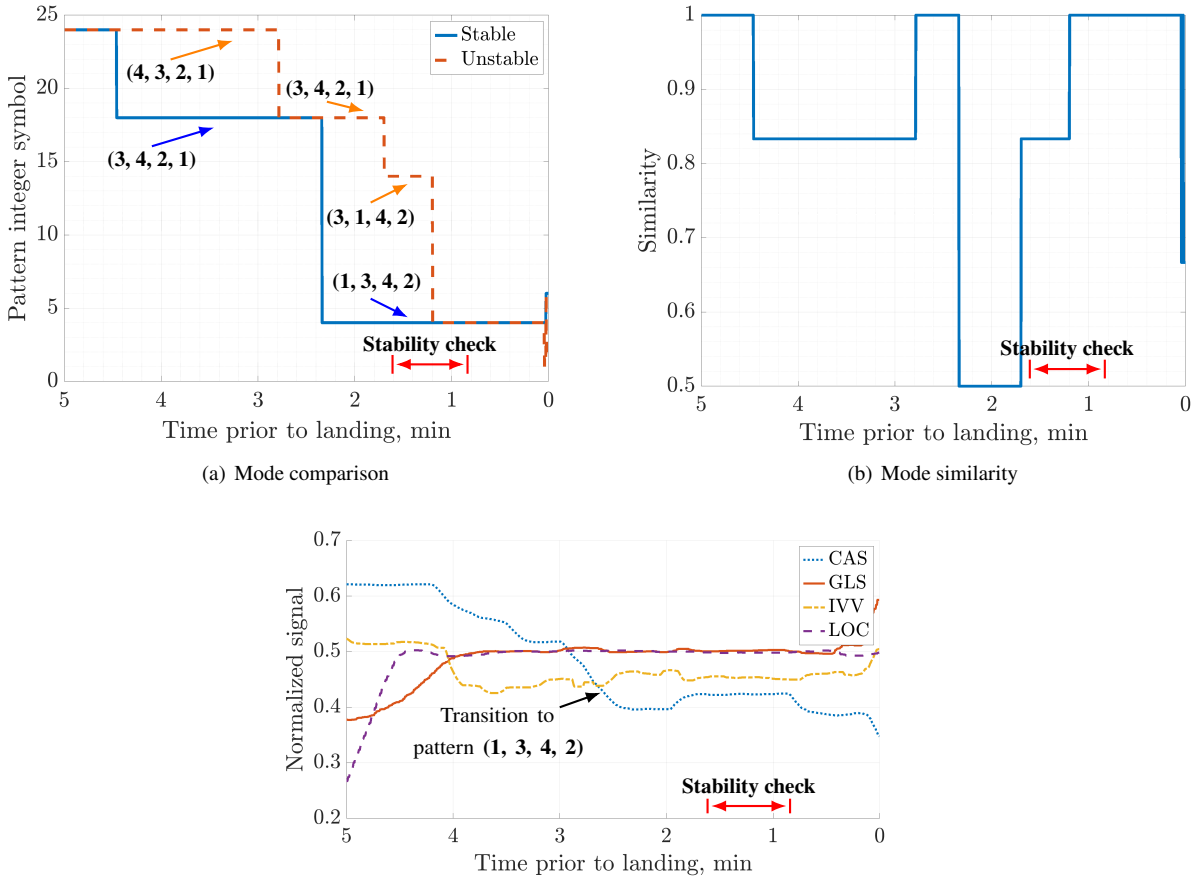
patterns in Fig. 6(a), suggests a highly probable reason as to why the similarity values in Fig. 5 are not close to 0, as we had hypothesized. This is because despite the presence of distinct modes in stable and unstable approaches, their high similarity indicates that the modes are not that different, or dissimilar, from one another. Hence, they yield high similarity values that strongly influence the distributions in Fig. 5 across all exceedance events. At the same time, these results also suggest the existence of certain aircraft states (i.e., patterns) that drive down the similarity between the stable and unstable modes at key moments prior to landing, especially at 2 minutes before landing. These states can be exploited to predict unstable approaches. These observations drive our hypothesis for RQB that the presence of both similar and dissimilar patterns (i.e., their distribution) can be used to predict the onset of an unstable approach. However, before continuing with the discussion for RQB, we further illustrate the interpretability of ordinal patterns by plotting an example of the dynamics of the normalized flight variables for a stable flight in Fig. 6(c). This figure depicts what the ordinal pattern (1, 3, 4, 2), airspeed (1) ≤ vertical velocity (3) ≤ localizer deviation (4) ≤ glideslope deviation (2), would look like during approach. A swap in the rank of two successive variables, say airspeed (1) and vertical velocity (3), would result in a smaller drop in similarity compared to the scenario where airspeed (1) swaps ranks with glideslope deviation (2). Over the same period of time, the latter scenario would suggest a more complex evolution of the flight dynamics.

Table 3 Ordinal pattern symbolic representation as integer values. These symbols are used in Fig. 6. Inside each pattern tuple, 1 stands for airspeed (CAS), 2 for glideslope deviation (GLS), 3 for vertical velocity (IVV), and 4 for localizer deviation (LOC). All variables are normalized using Eq. (1). The rankings represent the order of the variables.

Symbol	Pattern	Ranking	Symbol	Pattern	Ranking
1	(1, 2, 3, 4)	CAS ≤ GLS ≤ IVV ≤ LOC	13	(3, 1, 2, 4)	IVV ≤ CAS ≤ GLS ≤ LOC
2	(1, 2, 4, 3)	CAS ≤ GLS ≤ LOC ≤ IVV	14	(3, 1, 4, 2)	IVV ≤ CAS ≤ LOC ≤ GLS
3	(1, 3, 2, 4)	CAS ≤ IVV ≤ GLS ≤ LOC	15	(3, 2, 1, 4)	IVV ≤ GLS ≤ CAS ≤ LOC
4	(1, 3, 4, 2)	CAS ≤ IVV ≤ LOC ≤ GLS	16	(3, 2, 4, 1)	IVV ≤ GLS ≤ LOC ≤ CAS
5	(1, 4, 2, 3)	CAS ≤ LOC ≤ GLS ≤ IVV	17	(3, 4, 1, 2)	IVV ≤ LOC ≤ CAS ≤ GLS
6	(1, 4, 3, 2)	CAS ≤ LOC ≤ IVV ≤ GLS	18	(3, 4, 2, 1)	IVV ≤ LOC ≤ GLS ≤ CAS
7	(2, 1, 3, 4)	GLS ≤ CAS ≤ IVV ≤ LOC	19	(4, 1, 2, 3)	LOC ≤ CAS ≤ GLS ≤ IVV
8	(2, 1, 4, 3)	GLS ≤ CAS ≤ LOC ≤ IVV	20	(4, 1, 3, 2)	LOC ≤ CAS ≤ IVV ≤ GLS
9	(2, 3, 1, 4)	GLS ≤ IVV ≤ CAS ≤ LOC	21	(4, 2, 1, 3)	LOC ≤ GLS ≤ CAS ≤ IVV
10	(2, 3, 4, 1)	GLS ≤ IVV ≤ LOC ≤ CAS	22	(4, 2, 3, 1)	LOC ≤ GLS ≤ IVV ≤ CAS
11	(2, 4, 1, 3)	GLS ≤ LOC ≤ CAS ≤ IVV	23	(4, 3, 1, 2)	LOC ≤ IVV ≤ CAS ≤ GLS
12	(2, 4, 3, 1)	GLS ≤ LOC ≤ IVV ≤ CAS	24	(4, 3, 2, 1)	LOC ≤ IVV ≤ GLS ≤ CAS

B. Building Predictive Models for Unstable Approaches

Predictive Modeling: Despite the lack of distinct ordinal patterns unique to unstable approaches, our hypothesis for RQB is that we can leverage the distribution of ordinal patterns to predict an unstable approach ahead-of-time. This requires the calculation of the patterns distributions at multiple time instances in the descent phase prior to the stability check window (see Fig. 2). The time instances we selected were 4.5, 4, 3.5, 3, 2.5, 2, and 1.5 minutes before landing. At each time instance, we derived the empirical distributions (i.e., probability mass functions (PMFs)) of the ordinal patterns present in 8,700 stable and 8,700 unstable flights through the calculation of normalized histograms. More specifically, each distribution represents the conditional probability of the aircraft's stability, either stable or unstable, given the observation of an ordinal pattern. The seven distributions we obtained are shown in Fig. 7. We refer to these distributions as the conditional stability distributions or *likelihoods* (see Eq. (6)). Figure 7(a)–(g) track the evolution of the conditional distributions throughout the span of 3 minutes. Despite the inclusion of the likelihood at 1.5 minutes before landing, the further out the instability prediction is from 1.5 minutes, the longer pilots have to stabilize the aircraft. At the 3-minute mark in Fig. 7(d), we begin seeing the emergence of more distinct distributions for stable flights, as the probability mass from higher number patterns 22–24 starts to decrease. For example, closer to the 1.5-minute mark, patterns 3 and 4 occur more frequently in stable flights than in unstable flights. This shift of mass towards lower number patterns as we get closer to landing is not present in unstable flights, where instead we observe higher probabilities for



(a) Mode comparison

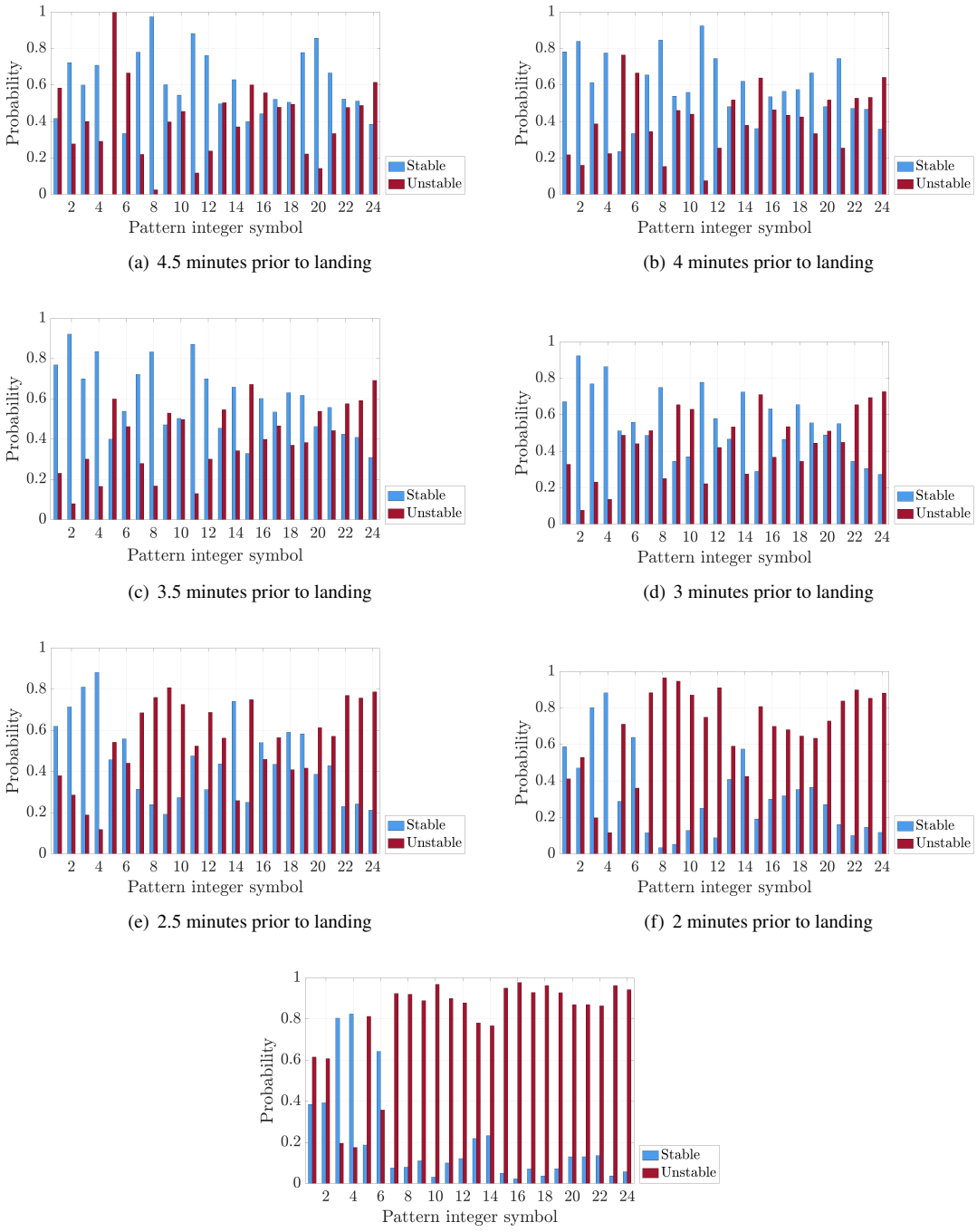
(b) Mode similarity

(c) Example of stable flight dynamics

Fig. 6 At each time step of the 5-minute time window prior to landing, we obtained the most occurring ordinal pattern, the mode, across 2500 stable and 2500 unstable flights. The integer value assigned to each mode is plotted at each time step in (a). In (b), Eq. (4) quantifies the similarity between the stable and unstable modes. The decreased similarity at the 4- and 2-minute marks indicates a reordering of the flight variables. (c) Shows an example of the dynamics of a stable flight, where a change in the ordinal pattern is visualized as a change in the rank (i.e., the ordering) of the signals.

patterns 7–24 compared to stable flights. This suggests that the presence of higher number patterns between 2 and 1.5 minutes before landing may be a good indicator of instability. To learn more about the evolution of ordinal patterns, we plotted the pattern distributions conditioned on the stability of the aircraft in Fig. 8. We refer to these PMFs as conditional pattern distributions or *posteriors*. These posteriors highlight a shift in mass to patterns 3 and 4 near the 2.5-minute mark for stable flights. This shift compared to unstable flights is more evident at the 2- and 1.5-minute marks, where patterns 3 and 4 are predominantly found in stable flights. Looking at Table 3, patterns 1 through 6 indicate that normalized airspeed (CAS) has gone below all other signals. This observation is in agreement with aviation agency recommendations that reducing and managing airspeed before the 1,000 ft and 500 ft altitude crossings can reduce the probability of instability on approach.

Having calculated the likelihoods using thousands of training flights, for each test flight, we approximate the prior over non-overlapping 30-second intervals starting at 5 minutes before landing. This prior is essentially a normalized histogram of the ordinal patterns. At a sampling rate of 4 Hz, the prior is calculated using 120 samples. With the priors and likelihoods at every prediction window (4.5, 4, 3.5, 3, 2.5, 2, and 1.5 minutes before landing), we can calculate the probabilities $P(x = \text{stable})$ and $P(x = \text{unstable})$ and classify the stability flight over those 30-second's worth of ordinal patterns. The highest of the two probabilities determines the classification and, consequently, the prediction of the aircraft's stability.



(a) 4.5 minutes prior to landing

(b) 4 minutes prior to landing

(c) 3.5 minutes prior to landing

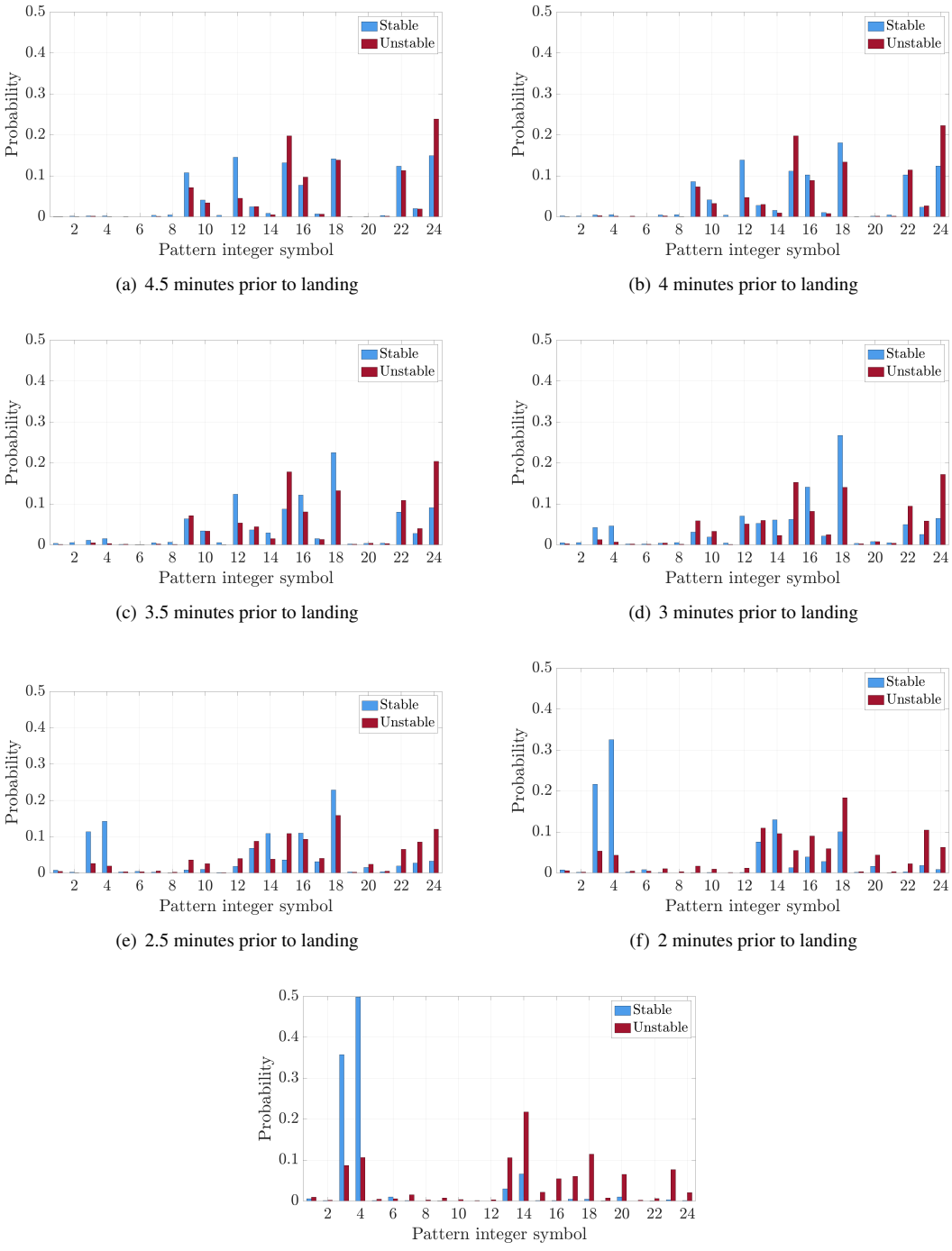
(d) 3 minutes prior to landing

(e) 2.5 minutes prior to landing

(f) 2 minutes prior to landing

(g) 1.5 minutes prior to landing

Fig. 7 Empirically derived probability mass functions (PMFs) of the stability of the aircraft, conditioned on the ordinal patterns, at multiple time instances prior to landing. A total of 17,400 (8,700 stable and 8,700 unstable flights) were used to create the PMFs. These *likelihoods* are used to make a prediction on the stability of a flight using Bayes' theorem at the following time instances: (a) 4.5, (b) 4, (c) 3.5, (d) 3, (e) 2.5, (f) 2, and (g) 1.5 minutes before landing.



(a) 4.5 minutes prior to landing

(b) 4 minutes prior to landing

(c) 3.5 minutes prior to landing

(d) 3 minutes prior to landing

(e) 2.5 minutes prior to landing

(f) 2 minutes prior to landing

(g) 1.5 minutes prior to landing

Fig. 8 Empirically derived probability mass functions (PMFs) of the ordinal patterns, conditioned on the flight stability, at multiple time instances prior to landing. A total of 17,400 (8,700 stable and 8,700 unstable flights) were used to create the PMFs. These *posteriors* are calculated at the following time instances: (a) 4.5, (b) 4, (c) 3.5, (d) 3, (e) 2.5, (f) 2, and (g) 1.5 minutes before landing.

Prediction Results: The classification results of using evidence probability comparisons (Eq. (6)) on 10,350 stable and 10,350 unstable test flights are summarized in the normalized confusion matrices in Fig. 9 and performance metrics in Table 4. The labels for each flight, stable (negative class) or unstable (positive class), were obtained using the stabilization criteria in Table 1 during the stability check window. The predicted class over each 30-second interval acts as a prediction of future instability. The results show a consistent classification performance improvements as we get closer to the stability check window near the 1.5-minute mark, based on the median aircraft trajectory in Fig. 4. A large drop in false positives can be observed between the 3.5- and 2.5-minute marks, Fig. 9(c) & (e). This increased performance is potentially due to the lower occurrence of high-integer ordinal (> 6) in stable flights, as depicted when comparing Fig. 7(c) & (e). This results in a lower number of stable flights being classified as unstable that benefits accuracy and precision, which both see an increase of 0.07 going from 3.5 to 2.5 minutes. These improvements in classification performance allow us to associate certain patterns in Fig. 7 with the stability of the aircraft. In particular, we highlight the overwhelming frequency of patterns 3 and 4 in stable flights compared unstable flights. Using Table 3, we can directly map these patterns to the rankings of CAS $\leq$ IVV $\leq$ GLS $\leq$ LOC for pattern 3 and CAS $\leq$ IVV $\leq$ LOC $\leq$ GLS for pattern 4. This suggests that when normalized airspeed (CAS) and vertical velocity (IVV) are the first two variables in an ordinal pattern, the aircraft is more likely to be in a stable state. Therefore, it is important to carefully manage these two energy variables on approach.

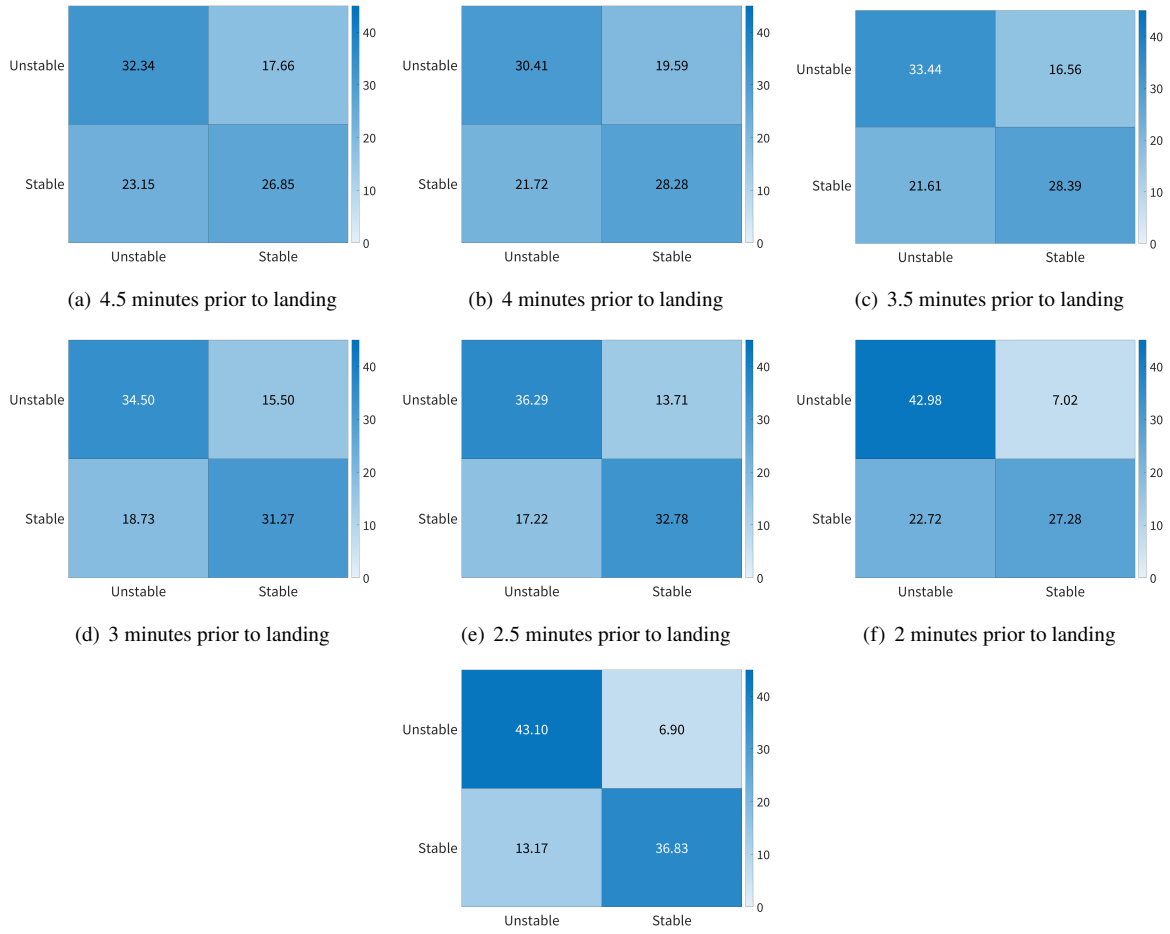
The classification improvements as we get closer to landing are only interrupted at the 2-minute mark as shown in Fig. 9(f) and Table 4. This drop in performance is characterized by an increase in falsely identified unstable approaches (false positives) and decrease in correctly identified stable approaches (true negatives). To explain the misclassification of stable flights, we draw observations from the likelihoods in Fig. 7(f). This figure shows a drop in the probability mass $P(x = \text{stable} \mid \pi_j)$ for higher number ordinal patterns such as $\pi_7 = 7$ through $\pi_{24} = 24$. Despite the shift of mass towards lower number ordinal patterns, lingering mass in the higher ordinal patterns may be a likely cause of the drop in classification performance at the 2-minute mark, resulting in the nearly 50/50 split in the class prediction of stable flights. Intuitively, this performance drop may be attributed to last minute aircraft corrections performed by pilots and automated systems to stabilize the aircraft, leading to unstable approaches that appear to be stable. Despite these results, the decrease in false negatives and increase in true positives at the same 2-minute mark greatly improves the classification of unstable flights, as evidenced by a recall value of 0.86. Arguably, the correct prediction of unstable approaches is more important than the prediction of stable approaches. Moreover, the classification results at 1.5 minutes before landing show that even if the goal is to solely detect instability in-time, our simple approach manages an accuracy of 0.80, precision of 0.77, and recall of 0.86. Importantly, the predictions maintain the interpretability of the ordinal patterns, as the pilots can assess the flight variable rankings that make up the patterns and attempt to bring the aircraft to a stable configuration, such as pattern 3 or 4.

Table 4 Classification performance metrics of using Bayes' theorem on a 2-class problem: stable or unstable flight. The metrics are calculated at non-overlapping 30-second time windows prior to landing. The time window labels aid in the comparison with Fig. 7, 8, & 9.

Time (min)	Window Label	Accuracy	Precision	Recall
4.5	a	0.59	0.58	0.65
4.0	b	0.59	0.58	0.61
3.5	c	0.62	0.61	0.67
3.0	d	0.66	0.65	0.69
2.5	e	0.69	0.68	0.73
2.0	f	0.70	0.65	0.86
1.5	g	0.80	0.77	0.86

IV. Conclusion

Despite recent advances in anomaly detection and precursor identification methods designed to improve aviation safety, many of them are hindered by poor end-user interpretability. This is especially evident in deep learning where the designer may be able to interpret the model to some extent, but this interpretability does not end up translating to the end-user (e.g., pilots, air traffic control). As patterns of failure become more complex due to the rise in



(a) 4.5 minutes prior to landing

(b) 4 minutes prior to landing

(c) 3.5 minutes prior to landing

(d) 3 minutes prior to landing

(e) 2.5 minutes prior to landing

(f) 2 minutes prior to landing

(g) 1.5 minutes prior to landing

Fig. 9 Flight stability classification results at (a) 4.5, (b) 4, (c) 3.5, (d) 3, (e) 2.5, (f) 2, and (g) 1.5 minutes before landing. The true class labels were obtained from the classification of the flight's stability during the stability check window that occurs at approximately 96 seconds prior to landing. The normalized confusion matrices summarize the class prediction performance (in percentage) on 20,700 test flights (10,350 stable and 10,350 unstable).

system complexity, more robust and interpretable risk detection paradigms will need to be designed. The shift away from univariate exceedance detection methods highlights the important trade-off between interpretability and model complexity. Although this shift is necessary to continue improving aviation safety, new algorithms that can generalize to future aircraft still need to maintain a high degree of model interpretability. In this study, we provide a straightforward and highly interpretable multivariate analysis of flight dynamics linked to approach instability. We present the viability of using ordinal patterns to capture the underlying state of the aircraft and simple Bayesian predictive modeling to detect future instability. This allowed us to associate certain patterns with the stability state of the aircraft in a probabilistic sense. This level of feature association can be more complicated and unintuitive when using complex machine learning and deep learning models. Future work will focus on improving classification performance by continuing to use Bayesian analysis or exploring other probabilistic frameworks, such as the fusion framework in [32] and fusion frameworks for time series [33–35]. A formulation of adverse event detection in a sensor fusion framework will allow us to fuse multimodal information of the flight dynamics, air traffic services, and the pilots' physical [36–38] and cognitive state [37, 39, 40]. This enables us to better capture the root cause of an anomaly, improve training, and understand the environmental changes imposed on the human. Thus, this stresses the importance of the development of interpretable ML models where we can take insights from other research human-focused areas that can inform us on environmental

conditions and training applications [32, 41–43]. Moreover, future studies built on the foundational work in this paper will extend our analysis into multi-class anomaly detection, while maintaining the interpretability and generalizability required by the future national airspace.

References

[1] Basora, L., Bry, P., Olive, X., and Freeman, F., "Aircraft Fleet Health Monitoring with Anomaly Detection Techniques," *Aerospace*, Vol. 8, No. 4, 2021, p. 103. https://doi.org/10.3390/aerospace8040103, URL https://www.mdpi.com/2226-4310/8/4/103, number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

[2] Odisho, E. V., and Truong, D., "Applying Machine Learning to Enhance Runway Safety Through Runway Excursion Risk Mitigation," *Proc. Integrated Communications Navigation and Surveillance Conference (ICNS)*, 2021, pp. 1–10. https://doi.org/10.1109/ICNS52807.2021.9441554.

[3] Lee, C.-H., Shin, H.-S., Tsourdos, A., and Skaf, Z., "Data analytics development of FDR (Flight Data Recorder) data for airline maintenance operations," *Proc. IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 2017, pp. 289–294. https://doi.org/10.1109/MFI.2017.8170443.

[4] Juarez Garcia, E., Stephens, C. L., and Napoli, N. J., "Detecting High-Risk Anomalies in Aircraft Dynamics Through Entropic Analysis of Time Series Data," *AIAA AVIATION Forum*, Chicago, IL, 2022.

[5] International Civil Aviation Organization, "Future of Aviation," [Online], 2019. URL https://www.icao.int/Meetings/FutureOfAviation/Pages/default.aspx.

[6] U.S. Department of Transportation, Bureau of Transportation Statistics, "Transportation Statistics Annual Report 2020," [Online], 2020. URL https://rosap.ntl.bts.gov/view/dot/53936.

[7] Federal Aviation Administration (FAA), "Next Generation Air Transportation System (NextGen)," [Online], 2011. URL https://www.faa.gov/nextgen/.

[8] Ellis, K., Krois, P., Koelling, J., Prinzel, L., Davies, M., and Mah, R., "A Concept of Operations (ConOps) and Design Considerations for an In-time Aviation Safety Management System (IASMS) for Advanced Air Mobility (AAM)," *Proc. AIAA SciTech Forum*, 2021.

[9] Hatfield, M., Cahill, C., Webley, P., Garron, J., and Beltran, R., "Integration of Unmanned Aircraft Systems into the National Airspace System-Efforts by the University of Alaska to Support the FAA/NASA UAS Traffic Management Program," *Remote Sensing*, Vol. 12, No. 19, 2020, p. 3112.

[10] Oza, N. C., and Stephens, C., "Data-Driven Safety," *Flight Safety Foundation*, 2021. URL https://flightsafety.org/asw-article/data-driven-safety/.

[11] Oliveira Filho, P. S., "The growing level of aircraft systems complexity and software investigation," [Online], 2018. URL https://www.isasi.org/Documents/library/technical-papers/2018/Wed/The%20Growing%20Level%20of%20Aircraft%20Systems%20Complexity%20and%20Software%20Investigation%20-%20Paulo%20Soares%20Oliveria%20Filho.pdf.

[12] Sheard, S., Konrad, M. D., Weinstock, C. B., and Nichols, W. R., "Definition and Measurement of Complexity in the Context of Safety," report, Carnegie Mellon University, Oct. 2016. https://doi.org/10.1184/R1/6572957.v1, URL https://figshare.com/articles/report/Definition_and_Measurement_of_Complexity_in_the_Context_of_Safety/6572957/1.

[13] Federal Aviation Administration (FAA), "Advisory Circular 91-79A – Mitigating the Risks of a Runway Overrun Upon Landing," [Online], 2018. URL https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_91-79A_Chg_2.pdf.

[14] Flight Safety Foundation, "Reducing the Risk of Runway Excursions: Report of the Runway Safety Initiative," [Online], 2017. URL https://flightsafety.org/files/RERR/fsf-runway-excursions-report.pdf.

[15] International Air Transport Association (IATA), "Unstable Approaches: Risk Mitigation Policies, Procedures and Best Practices," [Online], 2017. URL https://www.iata.org/contentassets/7a5cd514de9c4c63ba0a7ac21547477a/iata-guidance-unstable-approaches.pdf.

[16] Janakiraman, V. M., "Explaining Aviation Safety Incidents Using Deep Temporal Multiple Instance Learning," *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Association for Computing Machinery, New York, NY, USA, 2018, pp. 406–415. https://doi.org/10.1145/3219819.3219871, URL https://doi.org/10.1145/3219819.3219871.

[17] Bleu-Laine, M.-H., Puranik, T. G., Mavris, D. N., and Matthews, B., "Predicting Adverse Events and their Precursors in Aviation Using Multi-Class Multiple-Instance Learning," *Proc. AIAA SciTech Forum*, 2021. https://doi.org/10.2514/6.2021-0776.

[18] Memarzadeh, M., Matthews, B., and Templin, T., "Multi-Class Anomaly Detection in Flight Data Using Semi-Supervised Explainable Deep Learning Model," *AIAA Scitech 2021 Forum*, American Institute of Aeronautics and Astronautics, 2021. https://doi.org/10.2514/6.2021-0774, URL https://arc.aiaa.org/doi/abs/10.2514/6.2021-0774, _eprint: https://arc.aiaa.org/doi/pdf/10.2514/6.2021-0774.

[19] National Transportation Safety Board (NTSB), "Aviation Safety Recommendation Report DCA19RA017/DCA19RA101," [Online], 2019. URL https://www.ntsb.gov/investigations/accidentreports/reports/asr1901.pdf.

[20] Komite Nasional Keselamatan Transportasi (KNKT), "Aircraft Accident Investigation Report KNKT.18.10.35.04," [Online], 2018. URL http://knkt.dephub.go.id/knkt/ntsc_aviation/baru/2018%20-%20035%20-%20PK-LQP%20Final%20Report.pdf.

[21] Basora, L., Olive, X., and Dubot, T., "Recent Advances in Anomaly Detection Methods Applied to Aviation," *Aerospace*, Vol. 6, No. 11, 2019, p. 117. https://doi.org/10.3390/aerospace6110117, URL https://www.mdpi.com/2226-4310/6/11/117, number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

[22] Bleu-Laine, M.-H., Puranik, T. G., Mavris, D. N., and Matthews, B., "Multiclass Multiple-Instance Learning for Predicting Precursors to Aviation Safety Events," *Journal of Aerospace Information Systems*, Vol. 19, No. 1, 2022, pp. 22–36. https://doi.org/10.2514/1.I010971, URL https://doi.org/10.2514/1.I010971, publisher: American Institute of Aeronautics and Astronautics _eprint: https://doi.org/10.2514/1.I010971.

[23] Shannon, C. E., "A mathematical theory of communication." *Bell Syst. Tech. J.*, Vol. 27, No. 3, 1948, pp. 379–423. URL http://dblp.uni-trier.de/db/journals/bstj/bstj27.html#Shannon48.

[24] Zhang, H., and He, S., "Analysis and Comparison of Permutation Entropy, Approximate Entropy and Sample Entropy," *2018 International Symposium on Computer, Consumer and Control (IS3C)*, 2018, pp. 209–212.

[25] Delgado-Bonal, A., and Marshak, A., "Approximate entropy and sample entropy: A comprehensive tutorial," *Entropy*, Vol. 21, No. 6, 2019, p. 541.

[26] Federal Aviation Administration (FAA), "Air Traffic Bulletin 2019-1 – Stabilized Approaches," [Online], 2019. URL https://www.faa.gov/air_traffic/publications/media/atb_april_2019.pdf.

[27] NASA, "DASHlink – Sample Flight Data," [Online], 2012. URL https://c3.ndc.nasa.gov/dashlink/projects/85/.

[28] Flight Safety Foundation, "FSF ALAR Briefing Note 7.1 — Stabilized Approach," [Online], 2000. URL https://flightsafety.org/wp-content/uploads/2016/09/alar_bn7-1stablizedappr.pdf.

[29] Data4Safety, "Guidance for identifying unstable approach with flight data," [Online], 2022. URL https://www.easa.europa.eu/en/downloads/136957/en.

[30] Odisho, E. V., and Truong, D., "Applying Machine Learning to Enhance Runway Safety Through Runway Excursion Risk Mitigation," *2021 Integrated Communications Navigation and Surveillance Conference (ICNS)*, 2021, pp. 1–10. https://doi.org/10.1109/ICNS52807.2021.9441554, iSSN: 2155-4951.

[31] Zhang, S., Hu, Y., and Bian, G., "Research on string similarity algorithm based on Levenshtein Distance," *2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 2017, pp. 2247–2251. https://doi.org/10.1109/IAEAC.2017.8054419.

[32] Napoli, N. J., Stephens, C. L., Kennedy, K. D., Barnes, L. E., Juarez Garcia, E., and Harrivel, A. R., "NAPS Fusion: A framework to overcome experimental data limitations to predict human performance and cognitive task outcomes," *Information Fusion*, Vol. 91, 2023, pp. 15–30. https://doi.org/https://doi.org/10.1016/j.inffus.2022.09.016.

[33] Napoli, N. J., and Barnes, L. E., "A Dempster-Shafer approach for corrupted electrocardiograms signals," *The Twenty-Ninth International Flairs Conference*, 2016, pp. 355–360.

[34] Napoli, N., Leach, K., Barnes, L., and Weimer, W., "A MapReduce framework to improve template matching uncertainty," *2016 International Conference on Big Data and Smart Computing (BigComp)*, 2016, pp. 77–84. https://doi.org/10.1109/BIGCOMP.2016.7425804.

[35] Napoli, N. J., Barnes, L. E., and Premaratne, K., "Correlation coefficient based template matching: Accounting for uncertainty in selecting the winner," *2015 18th International Conference on Information Fusion (Fusion)*, 2015, pp. 311–318.

[36] Stephens, C., Kennedy, K., Napoli, N., Demas, M., Barnes, L., Crook, B., Williams, R., Last, M. C., and Schutte, P., "Effects on task performance and psychophysiological measures of performance during normobaric hypoxia exposure," *19th International Symposium on Aviation Psychology*, 2017, p. 202.

[37] Napoli, N., Harrivel, A., and Raz, A., "Improving Physiological Monitoring Sensor Systems for Pilots," *Aerospace America*, Vol. 12, 2020.

[38] Juarez Garcia, E., Ferguson, D. P., and Napoli, N. J., "Estimating Core Body Temperature Under Extreme Environments Using Kalman Filtering," *AIAA SciTech Forum*, 2022, p. 1271.

[39] Napoli, N. J., Demas, M., Stephens, C. L., Kennedy, K. D., Harrivel, A. R., Barnes, L. E., and Pope, A. T., "Activation complexity: A cognitive impairment tool for characterizing neuro-isolation," *Scientific Reports*, Vol. 10, No. 1, 2020, pp. 1–20.

[40] Napoli, N. J., Adams, S., Harrivel, A. R., Stephens, C., Kennedy, K., Paliwal, M., and Scherer, W., "Exploring cognitive states: Temporal methods for detecting and characterizing physiological fingerprints," *AIAA SciTech Forum*, 2020, p. 1193.

[41] Blanchfield, J. E., Hargroves, M. T., Keith, P. J., Lansing, M. C., Nordin, L. H., Palmer, R. C., St. Louis, S. E., Will, A. J., Scherer, W. T., and Napoli, N. J., "Developing Predictive Athletic Performance Models for Informative Training Regimens," *2019 Systems and Information Engineering Design Symposium (SIEDS)*, 2019, pp. 1–6.

[42] Roe, J., Barnes, L., Napoli, N. J., and Thibodeaux, J., "The Restorative Health Benefits of a Tactical Urban Intervention: An Urban Waterfront Study," *Frontiers in Built Environment*, Vol. 5, 2019.

[43] Napoli, N. J., Demas, M. W., Mendu, S., Stephens, C. L., Kennedy, K. D., Harrivel, A. R., Bailey, R. E., and Barnes, L. E., "Uncertainty in heart rate complexity metrics caused by R-peak perturbations," *Computers in Biology and Medicine*, Vol. 103, 2018, pp. 198–207.