Development of a Reliable Method for General Aviation Flight Phase Identification

Qilei Zhang, John H. Mott, *Senior Member, IEEE*, Mary E. Johnson, and John A. Springer

***Abstract*— Aircraft operations statistics have typically received significant attention from U.S. airport owners and operators and state, local, and federal agencies. Accurate operational data is beneficial in assessing airports' performance efficiency and impact on the environment, but operational statistics at nontowered general aviation airports are, for the most part, limited or not available. However, the increasing availability and economy of capturing and processing Automatic Dependent Surveillance-Broadcast (ADS-B) data shows promise for improving accessibility to a wide variety of information about the aircraft operating in the vicinity of these airports. Using machine learning technology, specific operational details can be decoded from ADS-B data. This paper aims to develop a reliable and economical method for general aviation aircraft flight phase identification, thereby leading to improved noise and emissions models, which are foundational to addressing many public concerns related to airports.**

***Index Terms*— ADS-B, flight identification, TICC, clustering, machine learning.**

I. Introduction

APPROXIMATELY 500 landing facilities in the United States, including 378 major airports, provide scheduled air services. In addition to these, 2,952 landing facilities support aeromedical, aerial firefighting, law enforcement, and disaster relief operations; these, plus another 1,350 noncommercial public-use airports, are referred to as general aviation airports [1]. However, most of these general aviation airports have only limited or nonexistent air traffic control facilities. This suggests that many specific airport operations details cannot be quickly or easily obtained. Currently, in the case of towered airports, aircraft operations are counted by air traffic personnel [2], [3]. The number of aircraft operations at nontowered airports may be estimated based on statistical sampling or other methods [4], implying that many of the details of these operations are unknowable. Without this information, it is difficult to develop reliable noise or emissions models for the related airports. Therefore, it is useful to determine a feasible means of estimating summary information about aircraft operations, especially that related to the duration of each flight phase. In other words, a data-driven approach is needed to more accurately quantify operation estimates for general aviation airports [5], [6]. Flight phase identification, the process of classifying time-series air vehicle flight data into different flight phases, can be utilized to provide these estimates.

A reliable and economic framework is proposed herein to solve this problem of classification and labeling of large quantities of flight data. The classification part of the proposed model performs Toeplitz Inverse Covariance-based Clustering (TICC), an unsupervised machine learning algorithm by leveraging the TICC Python module, to clustering ADS-B data [7]. There are currently many clustering methods that may be utilized to classify flight trajectories. Most of them [8]–[11], however, are appropriate for traffic flow classification and pattern prediction and are not suitable for single trajectory problems such as flight phase identification. Traditional clustering methods such as DBSCAN (Density-Based Spatial Clustering of Application with Noise) [12] examine the absolute value of each dimension of the signal to determine the similarity between the signals. TICC, on the other hand, determines signal similarity by examining the correlation between the dimensions of the signal, decomposing a high-dimensional time series into a clear sequential timeline of a few key states [7]. Hence, TICC is suitable for flight data, with proper considerations.

II. Data Selection

There is a significant amount of extant data related to aviation operations, including data generated by testing, simulation, and operation of actual aircraft. However, a large part of the data is not publicly accessible. Hence, obtaining a publicly available data source is essential to the model development process. According to the FAA NextGen Program, "as of January 1, 2020, ADS-B Out equipment is required to operate in the airspace defined in 14 CFR 91.225" [13]. This suggests that a large portion of general aviation aircraft are now equipped with ADS-B transponder equipment when operating in controlled airspace. Thus, ADS-B is a suitable choice for the research under consideration. In fact, ADS-B has gradually become one of the most popular aviation data sources because of its low cost and high update frequency. Choosing it as a data source therefore meets the two suggested initial goals of reliability and economy.

Manuscript received February 12, 2021; revised June 12, 2021 and August 16, 2021; accepted August 18, 2021. This work was supported by a Grant from Purdue Polytechnic Institute. The Associate Editor for this article was D. Sun. *(Corresponding author: John H. Mott.)*

Qilei Zhang, John H. Mott, and Mary E. Johnson are with the School of Aviation and Transportation Technology, Purdue University, West Lafayette, IN 47907 USA (e-mail: zhan3599@purdue.edu; jhmott@purdue.edu; mejohnson@purdue.edu).

John A. Springer is with the Department of Computer and Information Technology, Purdue University, West Lafayette, IN 47907 USA (e-mail: jaspring@purdue.edu).

Digital Object Identifier 10.1109/TITS.2021.3106774

III. Methods

Figure 1 illustrates the basic process structure of this model. It also provides an overall framework and describes individual model components.

A. Data Preprocessing

The first two sections in the flowchart describe how data is collected, subsetted, cleaned and aggregated. The data package consists of two years of data (2019 to 2020) received by the ADS-B equipment installed at the Purdue University Airport (KLAF). In the national airspace system, Class A airspace consists of the airspace from 18,000 feet mean sea level (MSL) up to and including flight level (FL) 600 [14]. Because these altitudes are typically above those at which most general aviation aircraft operate in phases of flight other than cruise, the related data from these aircraft have been removed from the model. Due to signal and reception anomalies [15], some data or some attributes of data may be missing. Data with incomplete entries need to be discarded. When necessary, geographic information is used to filter to a specific geographic region of interest. Simultaneously, the aircraft registration database may be obtained and merged with the ADS-B data using the hex identification code[1] as a database key to identify the aircraft model, engine type, and other useful information. Before the analysis, the altitude column of the ADS-B data is corrected by obtaining the compiled Meteorological Terminal Air Report (METAR) data. The pressure data may be extracted from the METAR to correct the altitude data [17] at any particular time. After that, the operation statistics table for a single aircraft can be established, and if necessary, the table based on the aircraft type. Table II shows a filled sample form, compiled according to the data collected from all aircraft by the receiving equipment at KLAF over the course of a single day.

B. Data Analysis

An ADS-B dataset is typical of what is considered "big data," characterized by large amounts of data with low information densities. Machine learning is an appropriate method with which to extract the desired information. The data is firstly grouped by the unique hexadecimal identification number of the aircraft and clustered chronologically, i.e., a time series of T observations is extracted as one operation. But if one data point from an aircraft is too "far"[2] from others from the same aircraft, the data is separated into two time blocks to process and considered to be associated with two different operations. Specifically, this situation may occur when the aircraft performs multiple non-continuous takeoff and landing operations at the same airport on the same day.

1) TICC: Next, spatial clustering is performed with TICC. It treats multivariate data $x_i \in \mathbf{R}^n$ as a short subsequence of size $w \ll T$ which ends at t, e.g., $x_{t-w+1}, \ldots, x_t$. This short subsequence gives *temporal consistency*, which is helpful to provide a proper context for each of the observations [7]. The new concatenated $nw-$dimensional vector X_t replaces the original isolated data observation, generating a new sequence from X_1 to X_t. For example, X_1 is the first segmentation incorporated with $x_1, x_2, \ldots, x_w$. As a result, the algorithm clusters these new sequences instead of directly clustering observations.

Each cluster is defined as a Markov Random Field (MRF) to highlight the correlations between various observations in the representative subsequence [18]. A $nw \times nw$ matrix Θ_i is constrained to be block Toeplitz as follows:

$$\Theta_i = \begin{bmatrix} A^{(0)} & \left(A^{(1)}\right)^\top & \left(A^{(2)}\right)^\top & \cdots & \cdots & \left(A^{(w-1)}\right)^\top \\ A^{(1)} & A^{(0)} & \left(A^{(1)}\right)^\top & \ddots & & \vdots \\ A^{(2)} & A^{(1)} & \ddots & \ddots & \cdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \left(A^{(1)}\right)^\top & \left(A^{(2)}\right)^\top \\ \vdots & & \ddots & A^{(1)} & A^{(0)} & \left(A^{(1)}\right)^\top \\ A^{(w-1)} & \cdots & \cdots & A^{(2)} & A^{(1)} & A^{(0)} \end{bmatrix}$$

where $A^{(0)}, A^{(1)}, \ldots, A^{(w-1)} \in \mathbf{R}^{n \times n}$. Sub-block $A^{(0)}$ shows the intra-time partial correlation so that $A^{(0)}_{ij}$ refers to the inter-relationship between concurrent values of different attributes, e.g. altitude and ground speed (GS). The overall objective is to find the K inverse covariances of each cluster $\Theta = \{\Theta_1, \ldots, \Theta_K\}$ and get the assignment results $\mathbf{P} = \{P_1, \ldots, P_K\}$, where $P_i \subset \{1, 2, \ldots, T\}$. The optimization problem then is generalized by the following objective function [19].

$$\underset{\boldsymbol{\Theta} \in \mathcal{T}, \mathbf{P}}{\operatorname{argmin}} \sum_{i=1}^{K} [\overbrace{\|\lambda \circ \Theta_i\|_1}^{\text{sparsity}} + \sum_{X_t \in P_i} (\overbrace{-\ell\ell\left(X_t, \Theta_i\right)}^{\text{log likelihood}} + \overbrace{\beta I\{X_{t-1} \notin P_i\}}^{\text{temporal consistency}})]$$

Here, $\mathcal{T}$ is the set of symmetric block Toeplitz matrices. $\|\lambda \circ \Theta_i\|_1$ is an l_1-norm penalty of the Hadamard product to encourage a sparse Θ_i and prevent overfitting, where $\lambda \in \mathbf{R}^{nw \times nw}$. $-\ell\ell\left(X_t, \Theta_i\right)$ is the negative likelihood that observation X_t belongs to cluster i [20], which is expressed as $\ell\ell\left(X_t, \Theta_i\right) = -\frac{1}{2}\left(X_t - \mu_i\right)^\top \Theta_i \left(X_t - \mu_i\right) + \frac{1}{2}\log\det\Theta_i - \frac{n}{2}\log(2\pi)$, where u_i is the empirical mean of cluster i, and $\beta I\{X_{t-1} \notin P_i\}$ is an indicator function with a penalty parameter to enforce temporal consistency.

2) Parameter: The objective function is then solved using expectation maximization (EM) iteration by alternating between cluster assignment and cluster parameter update [21]. It is worth noting that there are two regularization parameters λ and β, which are chosen by hand. For example, in the choosing of β, TICC encourages neighboring subsequences to belong to the same cluster with no penalty. As the transition parameter β in the model increases, neighboring subsequences are more likely to be assigned to the same cluster. As the parameter approaches infinity, the switching penalty becomes so large that all the points in the time series are grouped into a single cluster. Hence, careful parameter selection is essential. In the validation part of this article, the parameters were set as $\lambda = 0.11, \beta = 100, w = 1$.

Another critical model parameter is the cluster size. Simply choosing this parameter to equal the size of the data block

[1]It is a part of the aircraft's registration process to assign a unique ICAO 24-bit address, which can be represented in hexadecimal digital format [16].

[2]This threshold is set to 10 minutes in the model under consideration.

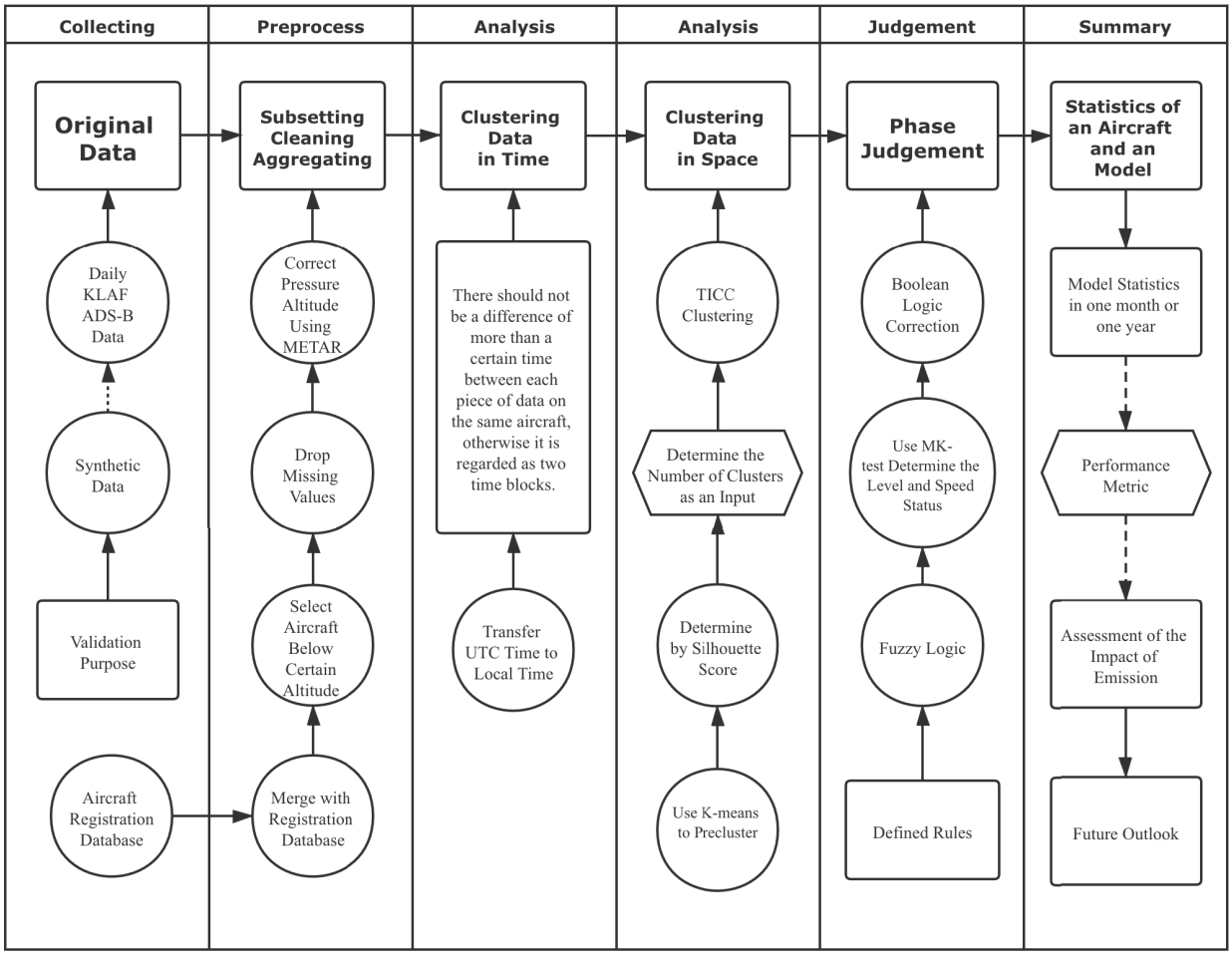


Fig. 1. Flowchart of the methodology.

yields suboptimal results. As the amount of data increases, a larger number of clusters tends to be chosen. A more reasonable approach [7] is to use K-means clustering [22] as a pre-clustering procedure and utilize the silhouette score [23] to evaluate and select this parameter. a is the mean intra-cluster distance, i.e., the mean distance to the other instances in the same cluster. b depicts the distance between a sample and the nearest cluster of which the sample is not a part [19]. As can be seen in Equation (1), the value of the silhouette score ranges from -1 to 1. If the score is close to 1, clusters are distinct and well separated from one another. On the contrary, when the score is close to -1, clusters are assigned improperly [22]. The cluster number assigned the highest silhouette score is most likely to be selected. An example is shown in Figure 2. Consider cluster numbers 3 and 4, which received the highest silhouette scores and are therefore most likely to be chosen. When the scores are similar,[3] a larger cluster number is preferred, as smaller cluster numbers may reduce the excision probabilities in similar consecutive phases (e.g., descent and approach).

$$\textbf{Score} = \frac{b-a}{max(a,b)} \tag{1}$$

[3]Defined as the difference in scores being less than 0.01.

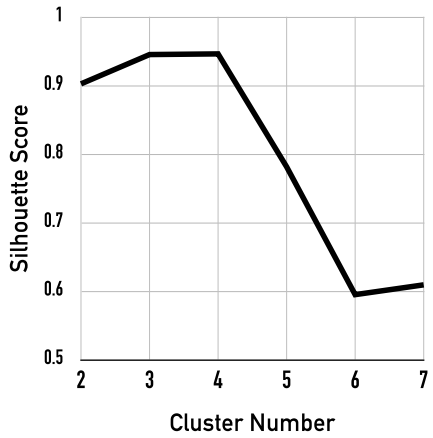


Fig. 2. Example of choosing a silhouette score.

3) Example: When the algorithm is applied, one aircraft operating trajectory will emerge, as shown in Figure 3. This example data comes from the aircraft coded A3E813 on August 17, 2020, 17:34 to 18:32. Each point in the trajectory represents a single ADS-B data series. The pattern of the aircraft's landing and takeoff (LTO) is displayed in the figure, and that figure also depicts the flaws in the ADS-B data.

TABLE I
DEFINITION FOR PHASE [24] AND FUZZY LOGIC RULES

Phase of Flight	Definition	Altitude	GS	Level Status	Poly Control[1]
Taxi (TXI)	The aircraft is moving on the aerodrome surface under its own power prior to takeoff or after landing.	Low	Low	Level	In Poly
Takeoff (TOF)	From the application of takeoff power, through rotation and to an altitude of 35 feet above runway elevation.	Low	Low or Med	Up	In Poly
Climb (CLB)	A controlled climb during any airborne phase leading to a increase in altitude from the end of the Takeoff phase.	Med or High	Med or High	Up	-
Cruise (CRS)	From completion of Initial Climb through cruise altitude and completion of controlled descent to the Initial Approach Fix (IAF).	High or Very High	High	Level	-
Descent (DCT)	A controlled descent during any airborne phase before approach phase leading to a decrease in altitude.	Med or High	Med or High	Down	-
Approach (APR)	From the IAF to the beginning of the landing flare.	Low or Med	Low or Med	Down	-
Unknown (UNK)	Phase of flight is not discernible from the information available.	-	-	-	-

[1] "In poly" means the position items of data are shown within the range of the airport.
[2] "-" stands for no requirement.
[3] Unknown is the situation that does not belong to any of these phases.

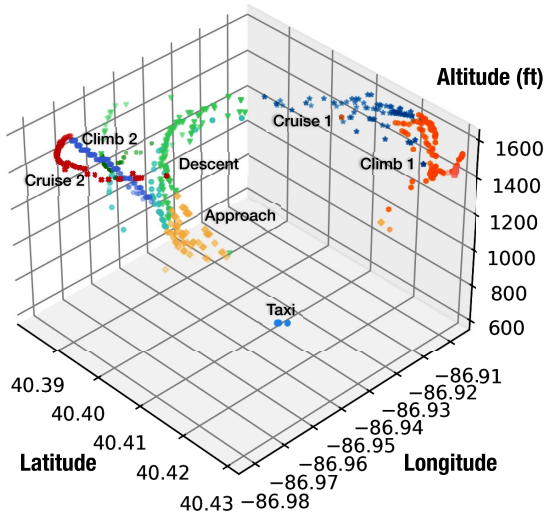


Fig. 3. Flight trajectory example with real ADS-B data.

On one hand, it can be seen that the aircraft's motion is divided into different groups. In the labeling function, the "Cruise 1" and "Cruise 2" clusters will both be classified as cruise phases, and the climb clustering is the same. So, in general, the phases of flight are well grouped. On the other hand, one notices some points with relatively lower altitudes, isolated from others. It is apparent from knowledge of the labeling function that these points belong to the taxi phase, as they are located on the airport's elevation plane, and the latitude and longitude are also within the bounding box surrounding the airport. Other characteristics of this data, such as ground speed, also meet conditions for the taxi phase. Nevertheless, the number of these data points is not large enough, and they are not continuous enough with other data points. As mentioned before, this is due to dipole receiving antenna characteristics that result in some ADS-B data, especially from aircraft operating at very low altitudes, being lost. This will reduce the accuracy of the statistics in the taxi phase.

C. Data Judgement

After the data is clustered, the next step consists of labeling each flight phase. First, the metrics of each data block are summarized; the summary includes the average altitudes, speeds, and tendencies toward variation. The flight phases are then determined based on ICAO definitions [24] using the integrated data. Specific flight phase definitions are found in Table I. Using traditional logic, for example, one would expect that an increase in an aircraft's altitude suggests a climb and a decrease in altitude implies a descent. However, when actual data are considered, this becomes problematic. Because the data are derived by a discrete sampling of a continuous system, it is often difficult to determine a clear boundary between phases of flight, such as between the take-off phase and the climb phase. While more detailed airborne data such as engine power levels would assist in the classification process, these data are unobtainable from ADS-B. Another significant problem is that ADS-B provides ground speed rather than airspeed, suggesting that the velocity data should be corrected for wind effects. Fuzzy logic [25], [26] can be employed to facilitate the solution of classification and decision-making problems such as these.

1) Fuzzy Logic: The process of fuzzification is used to convert the logical input value (average altitude) into membership degrees (low altitude, middle altitude, high altitude) of each set. The first five diagrams in Figure 4 provide an intuitive representation of the relationship between the input value and the membership.[4] One may then define rules to determine which attributes belong to a particular phase. The specific logic list is depicted in Table I. Finally, the results can be de-fuzzified to provide a numerical output. In this particular case, the output (value on the horizontal axis) and corresponding phase label can be determined by referring to each phase interval of the last diagram in Figure 4.

[4]The value of altitude is chosen for KLAF. The specific value should be adjusted according to airport elevation relative to mean sea level.

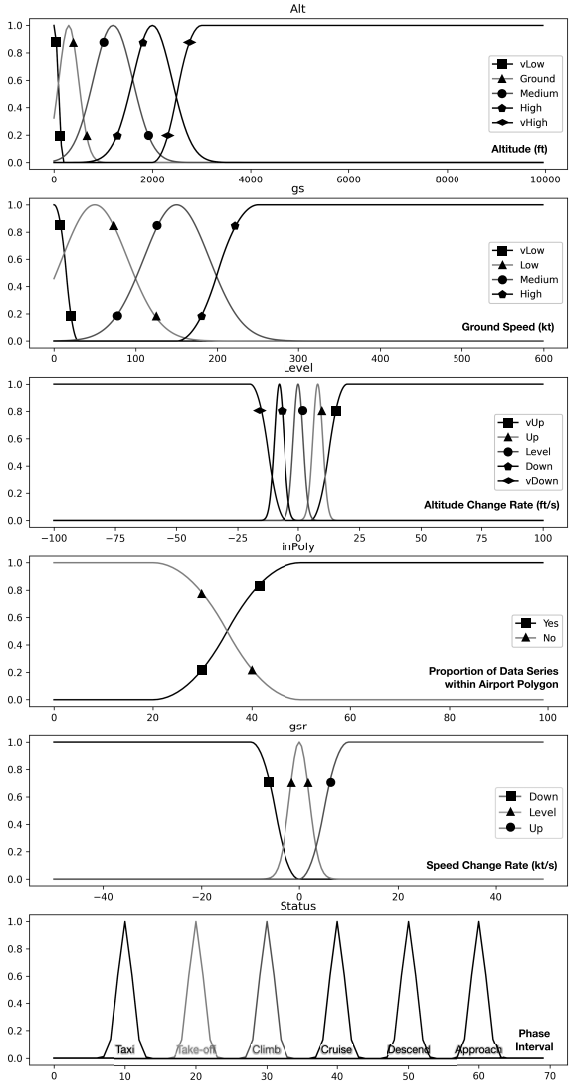


Fig. 4. Fuzzy logic membership.

The advantage of fuzzy logic is that it can resolve defects in the ADS-B data to a certain extent. An important point is that ADS-B data provides only ground speed; hence, implementing fuzzy logic classification can result in a more reasonable judgment using only GS information without introducing one or more weather databases to reconstruct the real-time airspeed. True Airspeed (TAS) is affected by many factors, such as temperature and pressure [27]; however, the probabilistic characteristics of fuzzy logic can compensate for the errors introduced by classifying phases using GS vs. TAS. In this manner, GS may be used as a reference condition instead of an absolute condition threshold, as is the case with traditional Boolean logic.

2) Correction: It is clear that fuzzy logic has several inherent flaws. First, the system is highly dependent on human knowledge and expertise. As a result, many solutions may arise for a particular problem and the system may require significant validation, verification, and regular update [28]. Second, the system has a limited number of logic operations (if . . . then . . .) to respond to non-standard operating applications in the real world. While increasing the number of inferences will improve the reliability of the system, it still will not address all possible situations [29]. It is possible, however, to verify the results of the fuzzy logic classification algorithm using traditional Boolean logic, and if a large deviation occurs, it may be corrected. This situation is especially pronounced in the initial phases of taxiing and take-off and is detrimental to the classification process. Because typical ADS-B signals often contain significant noise, which may lead to deviations or missing data at certain aircraft altitudes or orientations of the aircraft with respect to the ADS-B receiver, these data outliers may cause the single decision algorithm to produce classification errors. Therefore Boolean logic is implemented to test and correct the judgment. It is worth mentioning that in Boolean logic, when it is necessary to judge whether the altitude and airspeed are increasing or decreasing, the Mann-Kendall test (MK test) [30], [31] may be used to evaluate the trend of the data. The advantage of the MK test over linear regression is that it does not require linearity. According to Equation (2) [32], if the result is a positive number, later observations tend to be larger than earlier observations. Here, n is the number of observations in the set, and x_j, x_k are values of data points, where $j > k$.

$$S = \Sigma_{k-1}^{n-1} \Sigma_{j-k+1}^{n} sgn\left(x_j - x_k\right) \tag{2}$$

D. Data Summary

Finally, the data from all time blocks can be counted and summarized. Table II is an example showing a one-day collection for different aircraft models. In this manner, using daily analysis, the annual flight statistics of the airport and its surrounding airspace may be obtained. These statistics can be further utilized to assess the environmental impact of emissions from flights or to perform additional research.

IV. VALIDATION

Because the ADS-B data does not provide any means of ground-truthing for phases of flight [26], one cannot determine the accuracy of the model when used with real data. It is impractical to manually identify the segmentation of each time block and the accuracy of the corresponding label. Therefore, synthetic simulation data can be used as input data and the model's output compared with the initial label. The differences will facilitate an understanding of the model's accuracy and therefore provide a means of model validation.

A. Synthetic Data

Each set of simulation data represents an aircraft, and these are all generated at simulation time t_0. An initial random vector $\vec{x}_0$ is given, the attributes of which include the necessary variables that can be obtained from the ADS-B data, such as altitude, GS, and heading. Initial rates of change can be derived by setting the time interval to be one second. Some other attributes such as position (longitude, latitude) can be calculated based on speed and previous position. For the

TABLE II
EXAMPLE OF DAILY DATA SUMMARY

#	Hex_Ident List	MDL_Code	Take-Off	Climb	Cruise	Descent	Approach	Unknown
1	AA15CC, A66B52. . .	7102807	149.13	9.4	68.45	246.55	2.13	0.00
2	A89925,. . .	2072439	215.87	10.68	177.62	248.98	1.88	0.00
3	ABD536,. . .	7102803	75.15	11.75	181.53	309.15	2.33	0.00
4	A72842,. . .	2072449	65.15	3.83	54.40	364.87	0.33	0.00
5	A86B87,. . .	2072436	59.87	43.20	80.43	248.55	3.55	0.00

[1] The unit of the data in the table is minutes.

subsequent calculations, the following equation (3) may be used:

$$\vec{x}_{i+1} = \vec{x}_i + \frac{\partial \vec{x}_i}{\partial t} \times \Delta t + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2) \tag{3}$$

where $\vec{x}_i \in \mathbb{R}^d$ is a vector that has d attributes and represents a data point at a given time. Δt is the time interval for each neighboring data point, and ϵ is a normally distributed random error, simulated as the signal noise. The standard deviation σ is set randomly in a reasonable range according to the gauge units of the attribution and the specific situation. For example, for altitude, the σ can be set to 10 ft as the maximum reporting error of the equipment[5] for flight levels greater than 2000 ft or to 1 ft when the aircraft is on the ground, since the source of error is primarily from the deviation of barometric measurements due to atmospheric temperature and pressure variation in the higher flight levels, and from instrument error at ground level. When the status label changes, the vector of its change rate will also change significantly. Otherwise, only a certain amount of noise needs to be added to the previous change vector.

After the data item at each instance in time is generated, it is checked to ensure that the corresponding label is reasonable. For example, the speed in the initial climb phase should not be excessive, the flight altitude in the cruise phase should not be too low, and so on. Furthermore, random simulated ADS-B data can be produced with a different combination of phases, such as [taxi, take-off, climb, cruise, descent, approach, taxi] (hereinafter called Combination I), [take-off, climb, cruise, descent, approach] (Combination II), and [climb, cruise, descent] (Combination III). Combination I describes a complete idealized flight process. Combination II is a simpler version of the first, accounting for missing data from an aircraft operating at ground level. Combination III attempts to simulate realistic situations for helicopters and other vertical take-off and landing aircraft [34]. The matrix $\mathbf{X}_{synthetic} \in \mathbb{R}^{d \times n}$, that is $\left[x_1^\top, x_2^\top, \ldots, x_n^\top\right]^\top$. n can be used to represent the entire set of simulated data.

B. Performance Metrics

To determine model performance, a large number of datasets may be generated cyclically and the discrepancy between the initial data with predetermined labels and the result processed by the model recorded. There are many metrics that may be utilized to measure the quality of the model. Root Mean Square Error (RMSE) and Symmetric Mean Absolute Percentage Error (SMAPE) have been used here to evaluate the model's quality.

[5]In practical applications, FAA [33] uses %Fail to assess the performance of ADS-B equipment. The precise definition is "Percentage of flight that the corresponding category element failed performance assessment."

TABLE III
RMSE RESULT FOR COMBINATION II

Phase	RMSE[1]
Take-Off	25.1
Climb	34.1
Cruise	148.9
Descent	168.8
Approach	86.4

[1] Although RMSE is the summary of the synthetic data, which is unitless, the unit can be regarded as seconds equivalently.

RMSE may be employed directly to measure the performance of a single phase, such as climb. In the following formula, t_i is the total climb time that was calculated for every data set, $\hat{t}_i$ is the real total climb time, and N represents the number of tests performed, as well as the number of data sets. Table III lists examples of the RMSE for 500 generated data sets. The RMSE for each phase ranges from 25 seconds to 170 seconds.

$$\mathrm{RMSE_{Climb}} = \sqrt{\frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{t_i} - \hat{\mathrm{t}}_\mathrm{i})^2}{\mathrm{N}}} \tag{4}$$

The total data set measurement error is shown in equation (5). m depends on the number of phases, F_{nm} represents the calculated model time, and A_{nm} represents the defined real phase duration, while ΔT_{nm} is the difference between them. Additionally, the calculated total time F_n and the real total time A_n are equal, and both of them can be represented by T_{ntotal}. Because the range of SMAPE is [0, 1], one minus SMAPE (the "accuracy") can be used to provide a sense of the correct degree.

$$\begin{aligned} \mathrm{SMAPE} &= \frac{100\%}{N} \sum_{n=1}^{N} \frac{\sum_1^m |F_{nm} - A_{nm}|}{|A_n| + |F_n|} \\ &= \frac{100\%}{N} \sum_{n=1}^{N} \frac{\sum_1^m |\Delta T_{nm}|}{2 \cdot T_{ntotal}} \\ \mathrm{Accuracy} &= 1 - SMAPE \end{aligned} \tag{5}$$

The accuracy description is quite different when applied to RMSE. Because the magnitude of the RMSE is affected by the

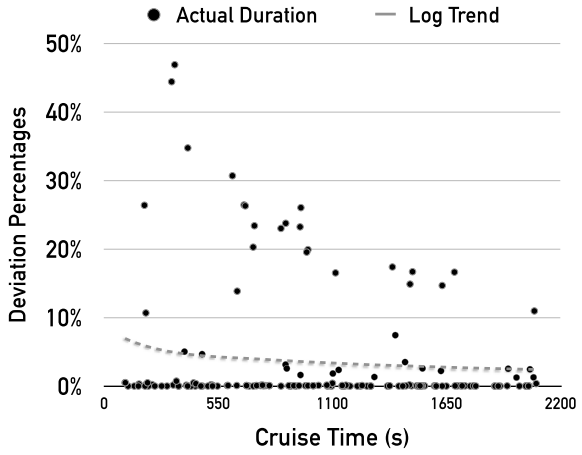
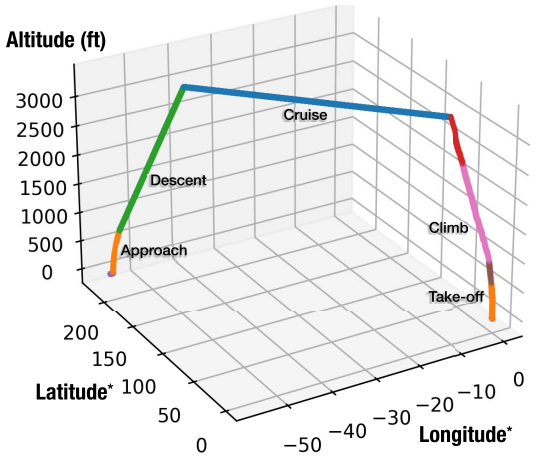
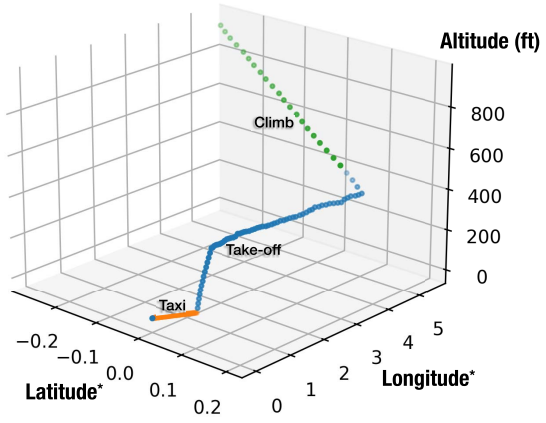


Fig. 5. Relationship between cruise time and deviation.



(a) Overall Situation.



(b) Enlarged Part of Starting Data.

Fig. 6. Example of flight trajectory with synthetic data. The latitude* and longitude* here do not represent the true measurement, but only the relative movement on the plane.

time unit and the duration of the phase itself, in the absence of the original correct phase duration one cannot intuitively understand the degree of error. Hence, this term does not reflect the model's quality well. However, SMAPE can provide a degree of error relative to the original correct time. For example, after several trials, an RMSE of 34.1 seconds for Climb is presented in Table III. It is not apparent whether this number is reasonable until the average total flight time is known to be 1908 seconds. SMAPE, on the other hand, includes the ratio of the deviation to the overall quantity. However, there is a certain inverse relationship between SMAPE and time duration. Specifically, the longer the flight time, the smaller the relative deviation from true value as shown in Figure 5. This trend is depicted by obtaining a log trendline using the least-squares method. The points lie in the bottom line with a deviation percentage of 0% represent that the program accurately provides the duration prediction for these flights without any deviation. To summarize, the two metrics should both be employed in the evaluation of the performance of the model.

Due to the Monte Carlo simulation approach employed here [35], the performance metric becomes more stable as the number of the data sets increases. Phase Combination III in Figure 7 depicts the manner in which the accuracy changes. After more than 500 tests, it stabilizes at roughly 97.3%.

C. Results

Figure 6a is a three-dimensional plot depicting the synthetic data and Figure 6b is a larger display of taxiing, take-off, and partial climb derived from Figure 6a. The two figures show clearly how the classifier model produces distinct flight phase results. The data starts at point (0,0,0). After a complete flight phase, the aircraft returns to the airport elevation plane. Some data series that should belong to different phases, but with similar altitude characteristics, may incorrectly have the same temporary label in the unsupervised learning clustering procedure. However, the data have a particular chronological order due to their timestamps, and this additional information may be used as discussed in Section III-C to ensure that the proper classification label is appended. The resulting total time summaries for the different phases [taxi, take-off, climb, cruise, descend, approach, taxi] are, respectively, [58.0, 95.0, 349.0, 181.0, 591.0, 63.0, 68.0]. Ground truthing provided by the synthetic data yields total times of [58, 67, 378, 180, 590, 64, 68], and it is evident that the differences between the two are small. The largest deviation is from the transition part of take-off and climb, with a smaller difference in cruise and descent, and an accurate result in the approach phase.

Monte Carlo simulation is then employed to evaluate the model's quality. When considering the synthetic Combination III, the corresponding RMSE is [23,138,135], yielding an accuracy of 97.3%. For Combination II, the corresponding RMSE is [25.1,34.1,148.9,168.8,86.3], with accuracy decreasing to 94.0%. When the taxi phase is introduced to the initial and final phases, the RMSE is [111.0, 39.2, 81.2, 160.1, 162.9, 98.4], and the accuracy is 91.6%. This indicates that the model performs well on the take-off, climb, and approach classification, with a somewhat greater deviation on the taxi, cruise, and descent portions, suggesting that the parameters for these phases can be improved.

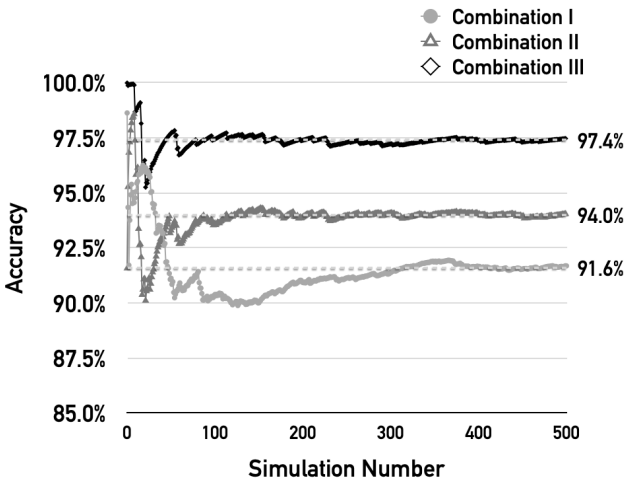


Fig. 7. Accuracy for three phase combinations.

TABLE IV
MODELS ACCURACY PERFORMANCE COMPARISON

Classification	Labeling	Accuracy	95% Confidence Interval
TICC	Fuzzy + Boolean	94.10 %	93.25% - 94.85%
TICC	Fuzzy	92.03 %	91.10% - 92.95%
K-means	Fuzzy + Boolean	78.90%	77.80% - 79.99%
K-means	Fuzzy	78.48%	77.47% - 79.48%
DBSCAN	Fuzzy + Boolean	77.37%	75.74% - 79.00%
K-means	Boolean	53.42%	51.26% - 55.59%

D. Model Comparison

To validate the effectiveness of the proposed model, the performance of different combinations of methodologies is demonstrated in Table IV. All the listed models adopt the same synthetic dataset with 500 flights from Phase Combination II. Results indicate that K-means and DBSCAN have basically identical performances on the unsupervised classification task. There is a slight gap between the "Fuzzy" and "Fuzzy + Boolean", since, as discussed in Section III-C, Boolean logic is applied to correct some extreme instances. Regardless, it is clear that the proposed model yields a substantial improvement relative to other methodologies. In general, the choice of the proposed methodology is effective and well applicable solving the flight phase identification problem.

V. DISCUSSION

A. Methodology Difference

In the previous section, simulated data was utilized to explore the effectiveness of the model with satisfactory results. The proposed model therefore appears to be a reliable, low-cost solution with the potential for widespread application. The methods suggested by Sun [36], Gu [4], and Goblet [37] are somewhat inadequate for several reasons.

First, these researchers did not use a large amount of data in the verification process, but rather conducted a small number of flight experiments, leading to concerns over the universality of the proposed method. Specifically, the selection of parameters may be optimized only for certain datasets [37]. While this is helpful for research using specific data sets, it is problematic for the extraction of daily operational statistics.

Due to the advantages of the TICC method in processing time-series data, the clustering of the data stream is continuous. As a simple example, during cruise, the controller gives the pilot an instruction to descend to a certain altitude. In this quick altitude change process, as long as the appropriate penalty parameters are selected, the entire phase is still considered cruising flight. In the earlier method [26], [37], because a single data point is used for classification statistics, the quick descent during the cruise may be attributed to an approach phase, resulting in an incorrect classification. The TICC method therefore maintains the continuity of the data tags.

Finally, the method proposed by the earlier researchers cannot distinguish more detailed flight phases at a low cost. In short, more detailed classification relies on more data sources [26]. For example, using G1000[6] data or directly acquiring aircraft instrumentation data will result in more accurate classification. However, in the actual application process, the cost of doing so could potentially be quite high. It is necessary to require a G1000 equipment installation in each aircraft and then manually download the resulting data, which is an unrealistic expectation for most general aviation aircraft. The method proposed herein achieves, in the view of the authors, a good balance between accuracy and economy.

B. Model Inaccuracies

Inaccuracies in the model occur principally from three sources. The first is from the lag characteristic of the TICC algorithm. If the transition penalty value parameter is too large, the state transition will show a certain degree of hysteresis. Specifically, when the aircraft state has changed, the algorithm needs extra data to confirm that the state has indeed changed. The result is that the calculated time of the previous phase will be longer, and the time of the later phase will be shortened. This leads to a certain degree of error. However, it is also necessary to realize that for some cases, the division of phases is indeed vague, and there is no accurate standard even for manual judgment. The most typical example is the distinction between descent and approach. In an ideal situation, after Air Traffic Controller (ATC) guides the aircraft to the Initial Approach Fix (IAF), the aircraft could be classified as in the approach phase, but there will be many deviations in actual operation, with a clear dividing moment difficult to determine.

A second portion of the error results from labeling the time block. As mentioned previously, when fuzzy logic encounters extreme values, classification errors often occur. Although the subsequent application of Boolean logic can correct some deviations, there are still special cases that cannot be corrected.

The last part of the error comes from the combination of the previous two. Phases with large amounts of missing data, such as the taxi phase, may result in this sort of error.

[6]The G1000 integrated flight instrument system consists of several integrated components, which can sample and exchange flight information [38].

Because ADS-B data is sporadic from aircraft operating at very low altitudes, TICC cannot accurately segment it, and the attribute value obtained in these time blocks does not fully reflect the state of motion of the aircraft, resulting in erroneous classifications.

In summary, discontinuous, inaccurate, or missing ADS-B signals due to the influence of various propagation factors may cause difficulty in data cleaning and classification. From the perspective of the flight phase itself, because the ADS-B data lacks more detailed power settings and other information, it is difficult to distinguish between take-off and climb, descent and approach when the phase definition is inaccurate. Unlike those engaged in commercial aviation, general aviation aircraft often execute more frequent maneuvers, such as touch-and-go movements during training, increasing the difficulty of discriminating between phases of flight.

C. Future Work

After each flight phase is accurately determined, environmental impacts of general aviation, including noise and emissions, can be accurately modeled and assessed. From an emissions perspective, 100LL[7] is the only remaining transportation fuel in the United States that contains the additive tetraethyl lead (TEL). This gasoline powering small aircraft is now the largest source of lead emissions in the U.S. [39]. It is therefore important to provide effective estimates of the impact of such pollution. From a longer-term perspective, the algorithm developed for general aviation can be applied not only to GA aircraft, but also can be extended to commercial aviation and even the operation of eVTOL in the future. An eVTOL's urban operating environment is particularly sensitive to noise. For manufacturers and aviation operators, it may be possible to use airborne data to calculate and report noise levels, while for regulators, third-party data can be used to provide a better evaluation of environmental noise.

There are potential improvements to the model that may be developed and implemented. Regarding data sources, more diverse data such as surface radar and video surveillance [40] can be considered to mitigate some of the issues caused by inaccurate or missing ADS-B data. Because the model operates using an existing data set, it may need to be augmented to allow the use of real-time data in the subsequent actual deployment so that real-time metrics [41] can be provided and monitored. In addition, many parameters need to be adjusted accordingly to better adapt to the special circumstances of each airport.

VI. Conclusion

This article describes a promising algorithm model that uses ADS-B data to analyze flight phase time. Through the sorting, analysis, judgment, and summary of ADS-B data, large quantities of time-series flight data is divided into marked flight phases. At the methodology's core is the TICC unsupervised learning algorithm for time series data and fuzzy logic judgment. The ability to obtain these specific flight operations' details cost-effectively is essential for many nontowered general aviation airports.

[7]LL is short for low lead. 100LL is considered as the most commonly used grade of aviation gasoline.

References

[1] G. A. Airports, "A national asset," in *FAA, US Department of Transportation*. Washington, DC, USA: U.S. Department of Transportation Federal Aviation Administration, 2012. [Online]. Available: https://www.faa.gov/airports/planning_capacity/ga_study/media/2012 AssetReport.pdf

[2] *Facility Operation and Administration*, document FAA Order JO 7210.3, Aug. 2007, Sec. 7.

[3] J. H. Mott, "Measurement of airport operations using a low-cost transponder data system," *J. Air Transp.*, vol. 26, no. 4, pp. 147–156, Oct. 2018.

[4] M. Johnson and Y. Gu, "Estimating airport operations at general aviation airports using the FAA NPIAS airport categories," *Int. J. Aviation, Aeronaut., Aerosp.*, vol. 4, no. 1, p. 3, 2017.

[5] J. H. Mott and N. A. Sambado, "Evaluation of acoustic devices for measuring airport operations counts," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2673, no. 1, pp. 17–25, Jan. 2019.

[6] C. Yang, J. H. Mott, B. Hardin, S. Zehr, and D. M. Bullock, "Technology assessment to improve operations counts at non-towered airports," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2673, no. 3, pp. 44–50, Mar. 2019.

[7] D. Hallac, S. Vare, S. Boyd, and J. Leskovec, "Toeplitz inverse covariance-based clustering of multivariate time series data," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 215–223.

[8] M. Gariel, A. N. Srivastava, and E. Feron, "Trajectory clustering and an application to airspace monitoring," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1511–1524, Dec. 2011.

[9] M. C. R. Murça and R. J. Hansman, "Identification, characterization, and prediction of traffic flow patterns in multi-airport systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 5, pp. 1683–1696, May 2019.

[10] M. Z. Li and M. S. Ryerson, "Novel terminal arrival airspace robustness metrics via topological density clustering," in *Proc. ICRAT Conf.*, Barcelona, Spain, 2018. [Online]. Available: http://icrat.org/ICRAT/seminarContent/2018/papers/ICRAT_2018_paper_49.pdf

[11] A. M. Churchill and M. Bloem, "Clustering aircraft trajectories on the airport surface," in *Proc. 13th USA/Eur. Air Traffic Manage. Res. Develop. Seminar*, Chicago, IL, USA, 2019, pp. 10–13.

[12] D. Birant and A. Kut, "ST-DBSCAN: An algorithm for clustering spatial–temporal data," *Data Knowl. Eng.*, vol. 60, no. 1, pp. 208–221, Jan. 2007.

[13] FAA. *ADS-B—Frequently Asked Questions*. Accessed: Oct. 5, 2020. [Online]. Available: https://www.faa.gov/nextgen/programs/adsb/faq/

[14] *Procedures for Handling Airspace Matters*, document FAA Order JO 7400.2N, Effective, Feb. 2006.

[15] J. H. Mott, "Estimation of aircraft distances using transponder signal strength information," *Cogent Eng.*, vol. 5, no. 1, Jan. 2018, Art. no. 1466619.

[16] S. Eskilsson, H. Gustafsson, S. Khan, and A. Gurtov, "Demonstrating ADS-B AND CPDLC attacks with software-defined radio," in *Proc. Integr. Commun. Navigat. Surveill. Conf. (ICNS)*, Sep. 2020, pp. 1B2-1–1B2-9.

[17] J. H. Mott, C. Yang, and D. M. Bullock, "Atmospheric pressure calibration to improve accuracy of transponder-based aircraft operations counting technology," *J. Aviation Technol. Eng.*, vol. 9, no. 2, p. 35, Jan. 2021.

[18] J. Li *et al.*, "Identification and classification of construction equipment operators' mental fatigue using wearable eye-tracking technology," *Autom. Construct.*, vol. 109, Jan. 2020, Art. no. 103000.

[19] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011. [Online]. Available: http://jmlr.org/papers/v12/pedregosa11a.html

[20] H. Ouyang, X. Wei, and Q. Wu, "Stock index pattern discovery via Toeplitz inverse covariance-based clustering," *Romanian J. Econ. Forecasting*, vol. 23, no. 2, pp. 58–72, 2020.

[21] J. Gui *et al.*, "An approach to extract state information from multivariate time series," *J. Comput.*, vol. 31, no. 6, pp. 1–11, 2020.

[22] K. R. Shahapure and C. Nicholas, "Cluster quality analysis using silhouette score," in *Proc. IEEE 7th Int. Conf. Data Sci. Adv. Analytics (DSAA)*, Oct. 2020, pp. 747–748.

[23] J. Yadav. (2019). *Selecting Optimal Number of Clusters in Kmeans Algorithm(Silhouette Score)*. Accessed: Oct. 13, 2020. [Online]. Available: https://medium.com/@jyotiyadav99111/selecting-optimal-number-of-cluster%s-in-kmeans-algorithm-silhouette-score-c0d9ebb11308
[24] C. Icao. (2013). *Phase of Flight, Definitions and Usage Notes*. CAST/ICAO Common Taxonomy Team. [Online]. Available: http://www.intlaviationstandards.org/Documents/PhaseofFlightDefinitions.pdf
[25] L. A. Zadeh, "Fuzzy logic," *Computer*, vol. 21, no. 4, pp. 83–93, Apr. 1988.
[26] J. Sun, J. Ellerbroek, and J. Hoekstra, "Large-scale flight phase identification from ADS-B data using machine learning methods," in *Proc. 7th Int. Conf. Res. Air Transp.*, 2016, pp. 1–7.
[27] J. Sun, J. Ellerbroek, and J. Hoekstra, "Modeling aircraft performance parameters with open ADS-B data," in *Proc. 12th USA/Europe Air Traffic Manage. Res. Develop. Seminar*, 2017, pp. 1–10.
[28] K. Goyal. (Mar. 2020). *Fuzzy Logic in Artificial Intelligence: Architecture, Applications, Advantages & Disadvantages*. [Online]. Available: https://www.upgrad.com/blog/fuzzy-login-in-artificial-intelligence/
[29] G. Heald, "Issues with reliability of fuzzy logic," *Int. J. Trend Sci. Res. Develop.*, vol. 2, no. 6, pp. 829–834, 2018, doi: 10.13140/RG.2.2.33328.40968.
[30] H. B. Mann, "Nonparametric tests against trend," *Econometrica, J. Econ. Soc.*, vol. 13, pp. 245–259, Jul. 1945.
[31] M. Kendall, *Rank Correlation Methods*, 4th ed. London, U.K.: Griffin, 1975.
[32] R. O. Gilbert, *Statistical Methods for Environmental Pollution Monitoring*. Hoboken, NJ, USA: Wiley, 1987.
[33] *Public ADS-B Performance Report (PAPR) User's Guide*, FAA, US Department of Transportation, ADS-B Focus Team, Washington, DC, USA, 2020.
[34] B. Wang, X. Peng, and D. Liu, "Airborne sensor data-based unsupervised recursive identification for UAV flight phases," *IEEE Sensors J.*, vol. 20, no. 18, pp. 10733–10743, Sep. 2020.
[35] A. M. Horowitz, "A generalized guided Monte Carlo algorithm," *Phys. Lett. B*, vol. 268, no. 2, pp. 247–252, Oct. 1991.
[36] J. Sun, J. Ellerbroek, and J. Hoekstra, "Flight extraction and phase identification for large automatic dependent surveillance–broadcast datasets," *J. Aerosp. Inf. Syst.*, vol. 14, no. 10, pp. 566–572, Oct. 2017.
[37] V. P. Goblet, "Phase of flight identification in general aviation operations," M.S. thesis, Purdue Univ., West Lafayette, IN, USA, Apr. 2016. [Online]. Available: https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1807&context=open_access_theses
[38] J. Liu, Y. R. Wang, and W. J. Zhao, "Cockpit display system simulation of general aviation aircraft based on VAPS XT," in *Adv. Mater. Res.*, vol. 846, pp. 1893–1898, Nov. 2014.
[39] U. A. T. A. R. Committee *et al.*, *FAA UAT Arc Final Report Part I: Unleaded Avgas Findings & Recommendations*, Federal aviation administration, US Department of Transportation, Unleaded AVGAS Transition Aviation Rulemaking Committee, Washington, DC, USA, Feb. 2012.
[40] C. Piciarelli, G. L. Foresti, and L. Snidaro, "Trajectory clustering and its applications for video surveillance," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill. (AVSS)*, Sep. 2005, pp. 40–45.
[41] K. Karboviak, "Providing metrics-based results to student pilots for critical phases of general aviation flights," M.S. thesis, Univ. North Dakota, Grand Forks, ND, USA, May 2018. [Online]. Available: https://www.proquest.com/docview/2063065115?pq-origsite=gscholar&fromopenview=true

Qilei Zhang received the B.S. degree from Civil Aviation University of China, Tianjin, China, in 2019, and the M.Eng. degree in transportation engineering from the University of California at Berkeley, USA, in 2020. He is currently pursuing the Ph.D. degree with Purdue University, IN, USA. In 2019, he worked as an Intern with the Air Navigation Bureau, International Civil Aviation Organization, Montreal, Canada. His research interest includes air transportation that could improves the safety, reliability or sustainability of the air transportation systems.

John H. Mott (Senior Member, IEEE) was born in Tuscaloosa, Alabama, in 1964. He received the B.S. and M.S. degrees in electrical engineering from the University of Alabama in 1986 and 1988, respectively, and the Ph.D. degree in civil engineering from Purdue University, West Lafayette, IN, USA, in 2017.

He previously served as the Engineering Manager at Command Alkon, Inc. In 2006, he joined the School of Aviation and Transportation Technology, Purdue University, where he is currently an Associate Professor. He serves as the Director with the Advanced Aviation Analytics Institute for Research and the Purdue University Center of Research Excellence. In addition, he is a commercially rated pilot and flight instructor. His research interests are in the areas of signal processing, stochastic modeling of transportation operations, and aerospace applications of Bayesian inference.

Prof. Mott's awards and honors include the 2019 President's Award for Excellence from the University Aviation Association, the 2019 Outstanding Graduate Student Mentor Award from the Purdue Polytechnic Institute, and the 2014 John P. Lisack Early Career Engagement Award from the Purdue College of Technology. He chairs the Central Indiana Section combined chapters of the IEEE Communications and Signal Processing Societies. He was the Founding Editor of the *Journal of Aviation Technology & Engineering*.

Mary E. Johnson received the B.S., M.S., and Ph.D. degrees in industrial engineering from The University of Texas at Arlington. After five years in aerospace manufacturing as an IE, she joined the Automation and Robotics Research Institute, Fort Worth. She was a program manager for applied research programs. She is currently a Professor and the Associate Head for graduate studies with the School of Aviation and Transportation Technology (SATT), Purdue University, West Lafayette, IN, USA. Fourteen years later, she became an Industrial Engineering Assistant Professor at Texas A&M–Commerce, before joining the Department of Aviation Technology, Purdue University, in 2007, as an Associate Professor. She is a Co-PI on the FAA Center of Excellence for general aviation research known as PEGASAS and leads the Graduate Programs in SATT. Her research interests are aviation sustainability, data driven process improvement, and aviation education. Her recent works in PEGASAS have centered on aviation weather technology research for GA.

John A. Springer is currently a Professor with the Department of Computer and Information Technology (CIT), Purdue University. He serves as a Leader of the Purdue Data Laboratory. He currently serves in a variety of leadership roles at Purdue, such as the Graduate Education Committee Chair in CIT, the Software and Data Curricular Subcommittee Chair in CIT, and the Co-Chair of Purdue's Integrative Data Science Initiative (IDSI) Curriculum Committee. His main research interest includes data science with recent emphasis on data-intensive applications in cyber security. He is a member of Purdue's Graduate Council and University Senate.