Journal of Open Aviation Science (2023), Vol.**1**
doi:10.59490/joas.2023.7211

JOAS

POSTER | ***The 11th OpenSky Symposium***

Hidden Markov Models and Flight Phase Identification

Rémi Perrichon,[*,1] Xavier Gendre,[2,3] and Thierry Klein[2,1]

[1]École Nationale de l'Aviation Civile, Université de Toulouse, Toulouse, France
[2]Institut de Mathématiques de Toulouse (UMR5219), Université de Toulouse, Toulouse, France
[3]Pathway.com, Paris, France
*Corresponding author: remi.perrichon@enac.fr

(This poster paper is not peer reviewed.)

Abstract
The use of Hidden Markov Models (HMMs) in segmenting flight phases is a compelling approach with significant implications for aviation and aerospace research. It leverages the temporal sequences of flight data to delineate various phases of an aircraft's journey, making it a valuable tool for enhancing the analysis of flight performance and safety. In this work, we implement a multivariate HMM to identify 6 flight phases: taxi, takeoff, climb, cruise, approach and rollout. We reach a median global accuracy of about 97% over a sample of several thousand flights with a very low number of decoded unlikely transitions. Regarding several performance metrics, our method is competitive with existing methods in the literature, such as fuzzy logic. Additionally, it provides, for each point of the flight, a probability of belonging to each phase. Even in situations where there are missing values in the data, HMMs remain effective, ensuring that no critical information is lost during the segmentation process.

Keywords: Flight Phase; Hidden Markov Model; ADS-B; Air Traffic Management; Time Series

Abbreviations: ADS-B: Automatic Dependent Surveillance–Broadcast, HMM: Hidden Markov Model, RoC : Rate of Climb

1. Introduction

Given some trajectory data, *flight phase identification* aims at segmenting a flight into different phases. More precisely, a segmentation is a partition of data points. This task has been popularized with the increasing availability of large Automatic Dependent Surveillance–Broadcast (ADS-B) datasets, for which flight phases are not labeled [1].

The segmentation of flights has several uses from building aircraft performance models [2] to enhancing the development of reliable noise or emissions models around airports [3].

A key aspect of flight trajectories is the undefined number of segments to uncover due to different flight frequencies and operations. Even within the same phase, aircraft may climb at different rates or fly at different cruise altitudes. Another specificity is the strong correlation in time and space between two consecutive points of a trajectory. Additionally, trajectory data may be noisy and/or have missing values.

These characteristics, along with the variety of air operations, account for the wide diversity of approaches presented in the literature on the subject. As put in [4], two main approaches are employed to identify phases from flight data records: logical rule-based decision-making, and probabilistic-based decision-making.

Given the challenge of specifying universal thresholds [5, 6], the fuzzy logic approach has established itself in the literature as a flexible, simple, and fast method [7, 8, 9, 10]. For each point, it is worth noting that fuzzy logic does not strictly return the probability of belonging to each class. Additionally, it does not consider the temporal nature of the trajectory.

Recently, many contributions have framed the problem of flight phase detection as a machine learning task [11, 12, 13, 14]. To achieve good results, some methods often require a large number of inputs, many steps are required as well as some training data.

Up to our knowledge, Hidden Markov Models (HMMs) have not often been used to segment flight phases. In this framework, the trajectory is modeled as a multivariate time series that explores multiple states (the flight phases). A segmentation is the result of a *decoding* procedure: once the parameters of the HMM are estimated, the most probable sequence of states is retrieved. Unlike threshold-based methods or fuzzy logic, HMMs place the temporal aspect of the trajectory at the core of segmentation by modeling the transition probabilities from one flight phase to another. Using HMMs allows for uncertainty quantification in segmentation, providing the probability of belonging to each class for each point. Unlike supervised methods, HMMs require only a very limited number of inputs and do not need a training phase. HMMs have been used for at least three decades in signal-processing applications, especially in the context of automatic speech recognition, but interest in their theory and application has expanded to other fields (environment, biophysics, ecology etc.) [15].

Given some trajectory data, we develop a multivariate HMM for the detection of the taxi, climb, cruise, approach, and rollout phases. Like [14], we use de-identified aggregate flight recorded data made available by NASA. We focus on data for tail 687. After a few basic data cleaning steps, we are working with 2,868 flights. Each flight is resampled to 1000 points (linear interpolation). Time is scaled so that each flight starts at $t = 0$ and ends at $t = 1$ (each flight is of different duration).

2. Hidden Markov Models

2.1 Theoretical framework

A univariate HMM consists of two parts:

- An unobserved *parameter process* (or *hidden state process*) denoted $\{C_t : t = 1, 2, ...\}$. It is a sequence of discrete random variables valued in $\{1, ..., m\}$. This process is assumed to be a discrete-time Markov chain.
- A *state-dependent process* denoted $\{X_t : t = 1, 2, ...\}$ (the *observation process*). It is a sequence of discrete random variables typically valued in $\mathbb{N}$ or $\mathbb{R}$. The distribution of this process is assumed to depend only on the current state C_t and not on previous states or observations. It is a conditional independence assumption, $\forall\ t \geq 2, \forall\ c_1, ..., c_t \in \{1, ..., m\}$,

$$\mathbb{P}(X_t = x_t \mid X_{t-1} = x_{t-1}, ..., X_1 = x_1, C_t = c_t, ..., C_1 = c_1) = \mathbb{P}(X_t = x_t \mid C_t = c_t). \tag{1}$$

An m-state HMM has m *state-dependent distributions*. Every observation is assumed to have been generated by one of m component distributions. The hidden state process selects which of the distributions is active at any time. The state-dependent distributions are defined as, for $i = 1, 2, ..., m$, $\forall\ t \geq 1$,

$$p_i(x_t) = \mathbb{P}(X_t = x_t \mid C_t = i). \tag{2}$$

That is, p_i is the probability mass or density function of X_t if the Markov chain is in state i at time t. We use p as a general symbol for probability mass or density functions.

To summarize, an HMM is a special type of a dependent mixture model in which a Markov chain selects the component distributions. We first estimate the parameters of the HMM (numerical maximization of the likelihood). Then, we deduce information about the states occupied by the underlying Markov chain. Such inference is known as *decoding*. *Local decoding* of the state at time t refers to the determination of that state which is most likely at that time. More details may be found in [15].

2.2 A model for flight phase identification

Suppose an aircraft is observed at integer times $t = 1, 2, ..., T$. For the moment, we assume that there are no missing values (this assumption can be relaxed). For each time index, we observe q values: it could be the position of the aircraft, its speed, the vertical rate and so on.

We consider the rate of climb (RoC), the ground speed (in knots) and the first differences of the ground speed to identify six flight phases: taxi, takeoff, climb, cruise, approach, rollout. Flight phases may be seen as states. We specify a constrained 6-state multivariate HMM for which the transition graph of the corresponding Markov chain is represented in Figure 1. The first state is a good candi-

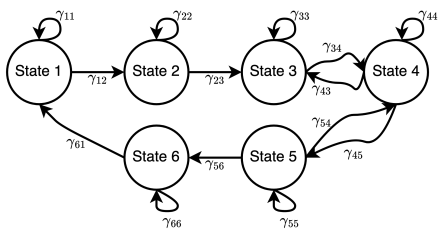

Figure 1. Transition graph of the constrained 6-state Markov chain.

date to represent the taxi phase. To ensure this, the initial distribution is taken to be $(1, 0, 0, 0, 0, 0)$ (it is fixed). State 2 refers to the takeoff, state 3 to the climb, state 4 to the cruise, state 5 to the approach, state 6 to the rollout. We use Gaussian distributions to set up the state-dependent densities of the RoC. The ground speed is transformed into a binary variable (1 if the ground speed is less than 0.05, 0 otherwise). We use Bernoulli distributions as the state-dependent densities of this variable. Finally, a discretized version of first differences of the ground speed is used. A value of 1 is assigned if the first difference at the point is greater than the quantile $q_{0.995}$, -1 if the first difference is less than the quantile $q_{0.05}$, and 0 otherwise. We use multinomial distributions as the state-dependent densities of this variable. A visual result for a typical flight is provided in Figure 2. Results are very good from a visual perspective. The value of several performance metrics over a subsample are presented in Figure 3. Among the 2,868 flights, the subsample corresponds to flights that have at least the 6 flight phases of interest. The median global accuracy is more than 97%.

Fuzzy logic provides a measure of uncertainty that is not perfect: by its nature, the degree of membership in each class is not a probability. This is not the case with HMMs, for which it is possible to obtain a probability of belonging to each class. An illustration is provided for the multivariate model (Figure 4).

Author contributions

- First Author: Conceptualization, Data Curation, Methodology, Software, Visualization, Writing – Original Draft, Writing – Review & Editing
- Second Author: Supervision, Validation, Writing – Review & Editing

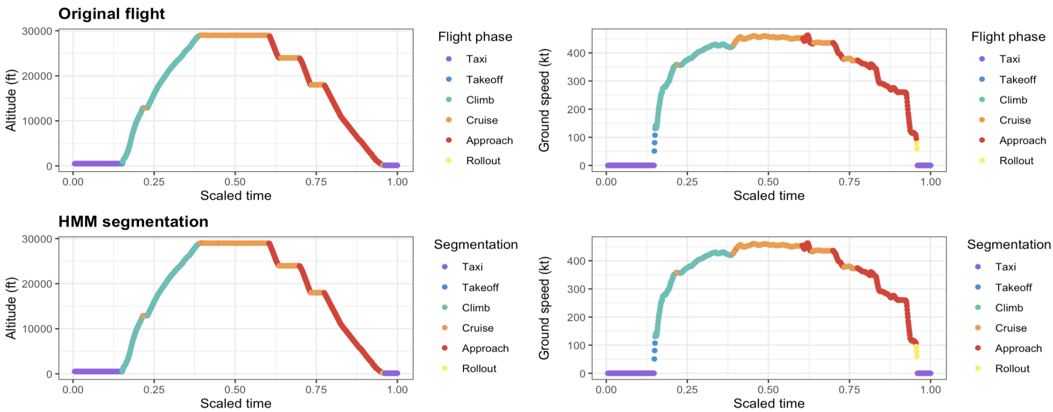

Figure 2. Identification results for a typical flight on the altitude profile and on the ground speed profile.

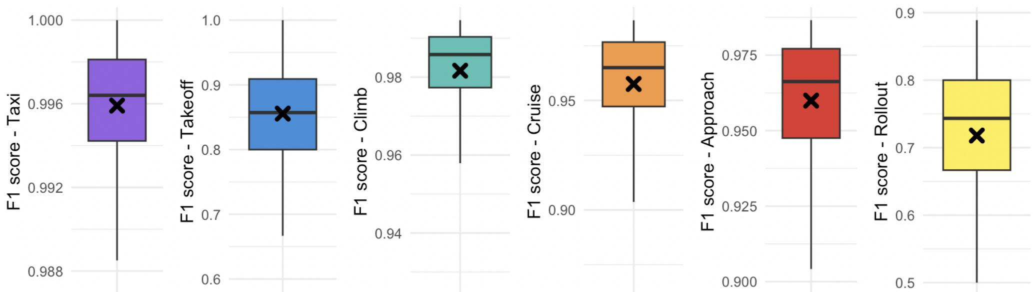

Figure 3. Evaluation of the performance. Box plots of the F-1 scores per state. The crosses correspond to the averages.

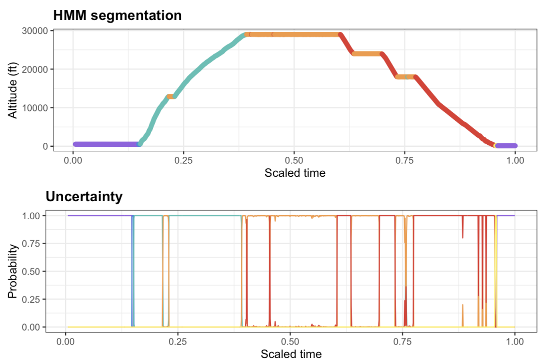

Figure 4. Segmentation of the 6 main phases of the flight using the multivariate HMM and probabilities of belonging to each class.

- Third Author: Supervision, Validation, Writing – Review & Editing

Open data statement

Data are accessible on the NASA DASHlink project page.
https://c3.ndc.nasa.gov/dashlink/projects/85/ (last check: 2023/10/22)

Reproducibility statement

The source code can be accessed at: https://github.com/remiperrichon/hmm_flight

References

[1] Junzi Sun, Luis Basora, Xavier Olive, Martin Strohmeier, Matthias Schafer, Ivan Martinovic, and Vincent Lenders. "OpenSky Report 2022: Evaluating Aviation Emissions Using Crowd-sourced Open Flight Data". In: *41th Digital Avionics Systems Conference (DASC)*. 2022. DOI: 10.1109/DASC55683.2022.9925852.

[2] J. Sun, J. Ellerbroek, and J. M. Hoekstra. "Modeling aircraft performance parameters with open ADS-B data". In: *12th USA/Europe Air Traffic Management Research and Development Seminar*. 2017.

[3] Qilei Zhang, John H. Mott, Mary E. Johnson, and John A. Springer. "Development of a Reliable Method for General Aviation Flight Phase Identification". In: *IEEE Transactions on Intelligent Transportation Systems* 23.8 (Aug. 2022), pp. 11729–11738. ISSN: 1558-0016. DOI: 10.1109/TITS.2021.3106774.

[4] Nicoletta Fala, Georgios Georgalis, and Nastaran Arzamani. "Study on Machine Learning Methods for General Aviation Flight Phase Identification". In: *Journal of Aerospace Information Systems* 20.10 (2023), pp. 636–647. ISSN: 1940-3151. DOI: 10.2514/1.I011246.

[5] Valentine Goblet, Nicoletta Fala, and Karen Marais. "Identifying Phases of Flight in General Aviation Operations". In: *AIAA Aviation Technology, Integration, and Operations*. June 2015. DOI: 10.2514/6.2015-2851.

[6] Mike Paglione and Robert Oaks. "Determination of Horizontal and Vertical Phase of Flight in Recorded Air Traffic Data". In: *AIAA Guidance, Navigation, and Control Conference and Exhibit*. Aug. 2006. DOI: 10.2514/6.2006-6772.

[7] Wallace E. Kelly and John H. Painter. "Flight Segment Identification as a Basis for Pilot Advisory Systems". In: *Journal of Aircraft* 43.6 (Nov. 2006), pp. 1628–1635. ISSN: 0021-8669. DOI: 10.2514/1.20484.

[8] Junzi Sun, Joost Ellerbroek, and Jacco Hoekstra. "Large-Scale Flight Phase Identification from ADS-B Data Using Machine Learning Methods". In: *7th International Conference on Research in Air Transportation*. June 2016.

[9] Junzi Sun, Joost Ellerbroek, and Jacco Hoekstra. "Flight Extraction and Phase Identification for Large Automatic Dependent Surveillance–Broadcast Datasets". In: *Journal of Aerospace Information Systems (online)* 14.10 (2017). DOI: 10.2514/1.I010520.

[10] Junzi Sun, Jacco M. Hoekstra, and Joost Ellerbroek. "OpenAP: An Open-Source Aircraft Performance Model for Air Transportation Studies and Simulations". In: *Aerospace* 7.8 (Aug. 2020), p. 104. DOI: 10.3390/aerospace7080104.

[11] Feng Tian, Xiaoke Cheng, Guanglei Meng, and Yimin Xu. "Research on Flight Phase Division Based on Decision Tree Classifier". In: *2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA)*. Sept. 2017, pp. 372–375. DOI: 10.1109/CIAPP.2017.8167242.

[12] Stephen Kovarik, Liam Doherty, Kiran Korah, Brian Mulligan, Ghulam Rasool, Yusuf Mehta, Parth Bhavsar, and Mike Paglione. "Comparative Analysis of Machine Learning and Statistical Methods for Aircraft Phase of Flight Prediction". In: *International Conference on Research in Air Transportation 2020, 9th International Conference* (Sept. 2020).

[13] Emy Arts, Alexander Kamtsiuris, Hendrik Meyer, Florian Raddatz, Annika Peters, and Stefan Wermter. "Trajectory Based Flight Phase Identification with Machine Learning for Digital Twins". In: *DLRK2021*. Sept. 2021.

[14] Datong Liu, Ning Xiao, Yujie Zhang, and Xiyuan Peng. "Unsupervised Flight Phase Recognition with Flight Data Clustering based on GMM". In: *2020 IEEE International Instrumentation

and Measurement Technology Conference (I2MTC). ISSN: 2642-2077. May 2020, pp. 1–6. DOI: 10.1109/I2MTC43012.2020.9128596.

[15] Walter Zucchini, Iain L. MacDonald, and Roland Langrock. *Hidden Markov Models for Time Series: An Introduction Using R*. 2nd ed. Chapman and Hall/CRC, June 2016. ISBN: 978-1-4822-5383-2.