Hidden Markov Models and Flight Phase Identification

Rémi Perrichon, Xavier Gendre, Thierry Klein

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

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. We show that HMMs work seamlessly with the fine granularity of Flight Data Recorder (FDR) data. HMMs offer remarkable flexibility and adaptability, proving particularly effective when the number or order of phases is unknown or not predetermined, as is often the case with complex flight scenarios such as helicopter flights. This adaptability is crucial for handling the diverse range of flight operations that differ from one aircraft to another. An example is given with the segmentation of an Automatic Dependent Surveillance–Broadcast (ADS-B) helicopter flight operated by the Swedish National Police.

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

From a conceptual point of view, there is no trouble in defining *flight phases*, that is to say different periods within a flight. Common taxonomies are, for instance, provided by the International Civil Aviation Organization (ICAO) [1] or by the International Air Transport Association (IATA) in Annex 1 of [2]. 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. It would be tedious to annotate them manually. A famous example of this rising accessibility of ADS-B data is the development of the non-profit OpenSky Network that has grown to 5,000 registered receivers all around the world, providing a large historical database [3].

The segmentation of flights has several uses. As stated in [4], flight phase segmentation is utilized to build aircraft performance models. In [5], the mass estimation method for ground-based aircraft climb prediction involves a *filtering* of climb segments. In [6], flight phase identification is related to delay analysis and safety. As explained by [7], estimating the duration of each flight phase is also

believed to enhance the development of reliable noise or emissions models around airports.

To be entirely precise, flight phase identification has several meanings. For the majority of applications, the identification of flight phases is a *vertical* segmentation problem (say, the identification of the takeoff, climb, cruise, approach and so on). We naturally visualize the different phases by representing them on the altitude profile. However, there are applications for which *horizontal* flight phases can also be defined. As recently reviewed in [8], this is the case for conflict detection for which we are also interested in detecting turns. In this contribution, we will focus solely on providing a vertical segmentation. We mostly focus on commercial aviation.

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, whether it be on the side of thresholding methods or probabilistic ones. The segmentation methods used in the literature only occasionally take into account the strong temporal correlation that exists between the data points that make up the flight. For example, the widely popular fuzzy logic method developed in [9] would produce an identical segmentation if the observations were permuted in time meaning that each point would have the same label.

Up to our knowledge and despite a well-known plasticity, Hidden Markov Models (HMMs) have not often been used to segment flight phases even though they exhibit very interesting characteristics for this problem. 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. This reduces the number of invalid transitions 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.) [10]. As a result, numerous packages are available for their implementation such as [11].

The contributions of this paper are of various types:

- The development of a univariate HMM for the detection of the three main flight phases (climb, cruise, and approach), as well as a multivariate model for the detection of the taxi, climb, cruise, approach, and rollout phases.
- A comparison of segmentation performances with the fuzzy logic approach for the three main flight phases.
- The calculation of several performance metrics, ranging from global accuracy to the number of invalid transitions on a sample comprising several thousand flights.
- A discussion on the impact of data preprocessing on the quality of flight segmentation.
- A discussion about the feasibility of adapting HMMs for the segmentation of a flight for which the phases to be identified are not specified in advance.

The paper is organized as follows. First, we provide a brief overview of existing methods as well as common performance metrics in Section 2. Second, the data we use is outlined in Section 3. Then, we present the theoretical framework of univariate HMMs in Section 4 as well as a model to detect the three main flight phases. The detection of additional flight phases is discussed in Section 5 and

falls within the framework of multivariate HMMs. The topics of data preprocessing and adapting models when the phases to be identified are not known in advance are addressed in Section 6.

2. Brief review of existing approaches

2.1 The two main approaches

As put in [12], two main approaches are employed to identify phases from flight data records: logical rule-based decision-making, and probabilistic-based decision-making.

Regarding rule-based approaches, several studies have focused on establishing thresholds to segment flight phases [13, 14]. Given the challenge of specifying universal thresholds for flight phase segmentation, the fuzzy logic approach has established itself in the literature as a flexible, simple, and fast method. Early references on the subject include the work of [15]. Several publications [16, 9], and its implementation in OpenAP [17] have now made it a widespread method. 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. Data smoothing is often necessary to achieve good results in practice.

Recently, many contributions have framed the problem of flight phase detection as a machine learning task. The use of decision trees classifiers to segment flight phases has been explored in [18]. Some machine learning methods are compared in [8]. Combined K-means clustering and LSTM neural networks have been combined in [19]. Gaussian Mixture Models have been used in [20]. To achieve good results, some methods often require a large number of inputs, often unavailable in ADS-B data. For instance, the engine fan speed is used in [20]. In any case, many steps seem necessary in the machine learning literature: selection of the parameters, implementation of a decision tree classifier and clustering of the results in [18], transformation of trajectory data into fixed length sequential data before using an LSTM neural network in [19]. The difficulty of obtaining a reliable training dataset leads some authors to use simulated data [19].

HMMs do not suffer from most of the mentioned limitations, as explained in the sequel.

2.2 Performance metrics

The comparison of flight phase identification methods is complex on several levels. One initial challenge relates to the number and types of flight phases selected. These can vary greatly depending on whether one considers commercial aviation or general aviation. A second challenge lies in the lack of consensus on the choice of a performance metric. It appears that the latter can be grouped into three main categories:

- The traditional metrics for classification problems such as the error rate, precision and recall [13, 14, 18, 19, 20]
- Metrics that focus on the total duration of each phase [7]
- Metrics that examine the transitions that are incorrectly predicted between phases as well as the total number of transitions [9]

In all contributions, the results are, of course, initially visualized. Because it is easy to find a degenerate segmentation that would provide an exact value for the duration of each phase while alternating the flight phases very randomly, it seems reasonable to consider that at least two metrics should be used. The use of classification metrics for each flight phase allows for the detection of the model's inability to segment some flight moments correctly, while global metrics provide an overview of the model's average performance. Since certain flight phases last significantly longer than others, the overall accuracy metric must be interpreted with caution. Counting the number of improbable transitions as well as the total number of transitions seems to be crucial in measuring the realism

of a segmentation. From an operational perspective, the aircraft does not spend its time rapidly transitioning between phases. In the following, we systematically consider multiple performance metrics.

For each flight phase, we typically define the usual F-1 score as the harmonic mean of *precision* and *recall*. If we consider the cruise phase, precision would be the amount of correctly predicted cruise points among all the points the model predicted as belonging to the cruise phase. Recall would be the number of cruise points are correctly identified as such among all the cruise points in the reference trajectory. The F-1 score is a metric commonly used in binary classification tasks. It rewards models that can achieve high precision and recall simultaneously. Using the F-1 score avoids to select a method that would label all points of the flight as belonging to a single phase (maximum recall for that phase but very poor precision), or another one that would consist of not labeling much points as belonging to that phase (poor recall but high precision for that phase).

3. Data

Because ADS-B data does not provide a ground-truth regarding the segmentation of flight phases, several other options are possible. Synthetic data have been used in [7] to validate the model. Data from an aircraft simulator are employed in [18] and [19]. Flight Data Recorder (FDR) data are encountered in [20].

Likewise, we have chosen to use de-identified aggregate flight recorded data made available by NASA. As written on the corresponding DASHlink project page, the files contain actual data recorded onboard a single type of regional jet operating in commercial service over a three-year period. While the files contain detailed aircraft dynamics, system performance, and other engineering parameters, they do not provide any information that can be traced to a particular airline or manufacturer. Appropriate parties have allowed NASA to provide the data to the general public for the purpose of evaluating and advancing data mining capabilities that can be used to promote aviation safety.

In this dataset, flight phases are determined based on the Aircraft Condition Monitoring System (ACMS). It is predictive maintenance tool consisting of a high capacity flight data acquisition unit and the associated sensors that sample, monitor, and record, information and flight parameters from significant aircraft systems and components. There are 8 possible flight phases in the dataset: unknown, preflight, taxi, takeoff, climb, cruise, approach, rollout. Note that the sampling frequency of each sensor is different, resulting in unequal data lengths of the parameters.

We focus on data for tail 687 for which there are 5,376 flights. After a few basic data cleaning steps, we are working with 2,868 flights. To be precise, only flights with a duration of more than thirty minutes, for which the main flight phases are documented, are retained for further analysis. Many flights are very short, thus explaining the final size of the sample. Each flight is resampled to 1000 points (linear interpolation). For a given observation time, it ensures that each sensor value is available (it solves the sampling frequency problem). Time is scaled so that each flight starts at $t = 0$ and ends at $t = 1$ (each flight is of different duration). Each flight can be easily visualized, as shown in Figure 1.

4. Univariate Hidden Markov Models

To introduce some definitions and notations, the basic framework of univariate HMMs is presented in Subsection 4.1. A simple model used for flight phase identification is introduced in Subsection 4.2.

4.1 The basic univariate framework

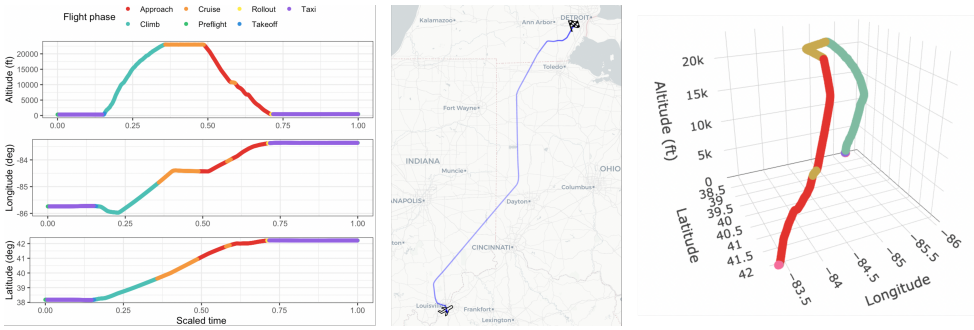

Figure 1. Visualization of a randomly selected flight from the dataset. Altitude, longitude, and latitude profiles [left], flat view [center], and three-dimensional view [right].

4.1.1 Definition

A HMM consists of two parts:

- We first consider 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. It then satisfies the so-called Markov property:

$$\forall\ t \geq 2, \forall\ c_1, ..., c_t, c_{t+1} \in \{1, ..., m\}, \mathbb{P}(C_{t+1} = c_{t+1} \mid C_t = c_t, ..., C_1 = c_1) = \mathbb{P}(C_{t+1} = c_{t+1} \mid C_t = c_t). \quad (1)$$

  That is, conditioning on the history of the process up to time t is equivalent to conditioning only on the most recent value C_t. The Markov property can be regarded as a first relaxation of the assumption of independence. The random variables are dependent in a specific way that is mathematically convenient.
- We then consider 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, $\forall\ t \geq 2, \forall\ c_1, ..., c_t \in \{1, ..., m\}$,

$$\mathbb{P}(X_t = x_t \mid X_{t-1} = x_{t-1}, ..., X_1 = x_1, C_t = c_t, ..., C_1 = c_1) = \mathbb{P}(X_t = x_t \mid C_t = c_t). \quad (2)$$

  As the Markov chain $\{C_t\}$ has m states, $\{X_t\}$ is called an m-state HMM. 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 $i = 1, 2, ..., m$, $\forall\ t \geq 1$,

$$p_i(x_t) = \mathbb{P}(X_t = x_t \mid C_t = i). \quad (3)$$

  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.

4.1.2 Characterization, homogeneity, and stationarity

Due to the Markov property, $\{C_t\}$ is fully characterized by:

- The *initial state distribution* $\boldsymbol{u}(1) = (\mathbb{P}(C_1 = 1), ..., \mathbb{P}(C_1 = m))$
- The one-step *state transition probabilities* denoted $\gamma_{ij}(t) = \mathbb{P}(C_{t+1} = j \mid C_t = i), \forall\ t \geq 1, \forall\ i, j \in \{1, ..., m\}$.

Due to the conditional independence assumption, $\{X_t\}$ is fully characterised by the state-dependent distributions.

Unless stated otherwise, we will always assume that the Markov chains we use are *homogeneous*, that is $\forall t \geq 1, \gamma_{ij}(t) = \gamma_{ij}$. This hypothesis is classic in the literature on HMMs, and its relaxation is not always necessary to improve the model's performance. We denote the *transition probability matrix* as:

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \vdots & \ddots & \vdots \\ \gamma_{m1} & \cdots & \gamma_{mm} \end{pmatrix}. \tag{4}$$

Note that it must be the case that:

- $\forall i, \forall j, \gamma_{ij} \in [0, 1]$,
- $\forall i,\ \sum_{j=1}^{m} \gamma_{ij} = 1$.

A Markov chain with transition probability matrix $\mathbf{\Gamma}$ is said to have stationary distribution $\boldsymbol{\delta}$ (a row vector with non-negative elements) if $\boldsymbol{\delta}\mathbf{\Gamma} = \boldsymbol{\delta}$ and $\sum_{k=1}^{m} \delta_k = 1$. Homogeneity alone is not sufficient to render the Markov chain a stationary process. It is sometimes useful to assume that the homogeneous Markov chain starts from its stationary distribution (in this case, it is said to be a *stationary* Markov chain). In the following, we do not make any stationarity assumption.

4.1.3 Likelihood

Suppose that we have T consecutive observations $x_1, ..., x_T$, assumed to be generated by an m-state HMM. To select the HMM parameters for which the model has the highest chance of having generated the observed data, the *likelihood* should be defined and computed. To be precise, we seek the probability L_T of observing the sequence, as calculated under an m-state HMM which has initial distribution $\boldsymbol{u}(1)$, a transition probability matrix $\mathbf{\Gamma}$, and state-dependent probability functions p_i.

The likelihood is given by:

$$L_T = \boldsymbol{u}(1)\boldsymbol{P}(x_1)\mathbf{\Gamma}\boldsymbol{P}(x_2)\mathbf{\Gamma}\boldsymbol{P}(x_3)\cdots\mathbf{\Gamma}\boldsymbol{P}(x_T)\mathbf{1}^\top,$$

where $\boldsymbol{P}(x_t)$ is defined as the diagonal matrix with i-th diagonal element $p_i(x_t)$. If $\boldsymbol{u}(1)$ is the stationary distribution of the Markov chain, we have $\boldsymbol{u}(1) = \boldsymbol{\delta}$ and $\boldsymbol{\delta}\boldsymbol{P}(x_1) = \boldsymbol{\delta}\mathbf{\Gamma}\boldsymbol{P}(x_1)$ (which makes the likelihood easier to code in some cases).

The matrix expression of the likelihood is computationally very interesting. To evaluate the likelihood, it is much more efficient to use recursive computation rather than relying on brute-force summation over all possible state sequences (we would have a m^T summands). Indeed, the naive approach separately considerate all possible state sequences that might have given rise to the observations which makes many calculations redundant.

A so-called *forward algorithm* is used to compute the likelihood. We first define the vector $\boldsymbol{\alpha}_t$, for $t = 1, 2, ..., T$, as:

$$\boldsymbol{\alpha}_t = \boldsymbol{u}(1)\boldsymbol{P}(x_1)\mathbf{\Gamma}\boldsymbol{P}(x_2)\mathbf{\Gamma}\boldsymbol{P}(x_3)\cdots\mathbf{\Gamma}\boldsymbol{P}(x_T), \tag{5}$$

with the convention that an empty product is the identity matrix. It follows immediately from this definition that:

$$L_T = \boldsymbol{\alpha}_T\mathbf{1}^\top \tag{6}$$

and

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1}\mathbf{\Gamma}\boldsymbol{P}(x_t),\ t \geq 2. \tag{7}$$

The elements of the vector $\boldsymbol{\alpha}_t$ are usually referred to as *forward probabilities*. Computations go:

$$\begin{aligned} \boldsymbol{\alpha}_1 &= \boldsymbol{u}(1)\boldsymbol{P}(x_1); \\ \boldsymbol{\alpha}_t &= \boldsymbol{\alpha}_{t-1}\boldsymbol{\Gamma}\boldsymbol{P}(x_t), \ t = 2, ..., T; \\ L_T &= \boldsymbol{\alpha}_T \mathbf{1}^\top. \end{aligned} \tag{8}$$

In practice, the maximization of the likelihood with respect to the parameters can be made numerically. It leads to well-known problems:

- Numerical underflow (in the discrete case, the likelihood approaches 0 with probability 1 exponentially fast). In this case, it is enough to scale the likelihood computation.
- Unbounded likelihood (when we consider continuous state-dependent distributions, it may happen that the likelihood is unbounded in the vicinity of certain parameter combinations). To tackle this issue, the discrete likelihood is maximized instead of the joint density.
- Constraints on the parameters (for example, row sums of Γ are equal to 1). A common practice is to reparametrize the model.
- Multiple local maxima in the likelihood. A sensible strategy is to use a range of starting values for the maximization, and to see whether the same maximum is identified in each case.

Instead of performing a direct numerical optimization, another popular approach is to treat the states as missing and to employ the EM algorithm in order to find the maximum likelihood estimates of the parameters. Yet, direct optimization is advisable when some constraints are added to the model.

4.1.4 Local decoding

Given the HMM and the observations, one can 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. To perform local decoding, analogous to forward probabilities, we should define *backward probabilities* as the elements of the vector $\boldsymbol{\beta}_t$:

$$\boldsymbol{\beta}_t^\top = \boldsymbol{\Gamma}\boldsymbol{P}(x_{t+1})\boldsymbol{\Gamma}\boldsymbol{P}(x_{t+2}) \cdots \boldsymbol{\Gamma}\boldsymbol{P}(x_T)\mathbf{1}^\top, \tag{9}$$

with the convention that an empty product is the identity matrix.

For each time $t \in \{1, ..., T\}$, one can therefore determine the distribution of the state C_t, given the observations $x_1, ..., x_T$, which for m states is a discrete probability distribution with support $\{1, ..., m\}$. The conditional distribution of C_t given the observations can be obtained, for $i = 1, 2, ..., m$, as

$$\mathbb{P}(C_t = i \mid \boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)}) = \frac{\mathbb{P}(C_t = i, \boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)})}{\mathbb{P}(\boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)})} \tag{10}$$

$$= \frac{\alpha_t(i)\beta_t(i)}{L_T}. \tag{11}$$

For each time $t \in \{1, ..., T\}$, the most probable state given the observations, is defined as:

$$i_t^* = \underset{i=1,...,m}{\operatorname{argmax}} \, \mathbb{P}(C_t = i \mid \boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)}), \tag{12}$$

where $\boldsymbol{X}^{(T)}$ is the the history $(X_1, ..., X_T)$ and $\boldsymbol{x}^{(T)}$ is $(x_1, ..., x_T)$. This approach determines the most likely state separately for each t by maximizing the conditional probability $\mathbb{P}(C_t = i \mid \boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)})$.

Local decoding comes with one crucial advantage: an uncertainty quantification in the decoded state sequence. It is illustrated in Subsection 6.2.

4.1.5 Global decoding

Instead of the most likely state for each separate time t, it is often the case that we are interested in the most likely sequence of hidden states. Instead of maximizing $\mathbb{P}(C_t = i \mid \boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)})$ over i for each t (Equation 12), one seeks that sequence of states $c_1, c_2, ..., c_T$ which maximizes the conditional probability $\mathbb{P}(\boldsymbol{C}^{(T)} = \boldsymbol{c}^{(T)} \mid \boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)})$. This represents a subtly distinct maximization problem compared to local decoding and is referred to as *global decoding*. The outcomes of local and global decoding are frequently quite similar, although not identical. There is a highly efficient algorithm for determining this sequence of states, known as the Viterbi algorithm [21]. Details may be found in [10] (Subsection 5.4.2) and in [22] (Subsection 4.5.2).

4.2 A univariate 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 is relaxed is the following). 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 naturally consider q time series $\left\{\boldsymbol{X}_t = \left(X_{t1}, X_{t2}, X_{t3}, ..., X_{tq}\right) : t = 1, ..., T\right\}$. Each series may represent observations of different type. For example, longitude and latitude are circular-valued whereas the altitude is positive or zero. In this section, we focus on the case $q = 1$. It means that only one time series is used for the segmentation.

For the identification of the three main flight phases (climb, cruise, approach), we begin by presenting an initial model based on the rate of climb (RoC) expressed in ft.min^{-1} in Subsection 4.2.1. We illustrate the case in which there are missing values in Subsection 4.2.2.

4.2.1 Identification of the three main phases with the rate of climb (RoC)

For the first model, we consider the rate of climb (RoC) to identify three flight phases: the climb, the cruise, and the approach. Flight phases may be seen as states. Transitions between the states are very constrained: we should go from the climb to the cruise and from the cruise to the approach. It is very unlikely to jump from the climb directly to the approach. It is impossible to go from the approach to the climb. We naturally specify a constrained 3-state univariate HMM for which the transition graph of the corresponding Markov chain is represented in Figure 2. The transition probability matrix of the first model is:

$$\boldsymbol{\Gamma_1} = \begin{pmatrix} \gamma_{11} & \gamma_{12} & 0 \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ 0 & \gamma_{32} & \gamma_{33} \end{pmatrix}. \tag{13}$$

The first state is a good candidate to represent the climb phase. To ensure this, the initial distribution

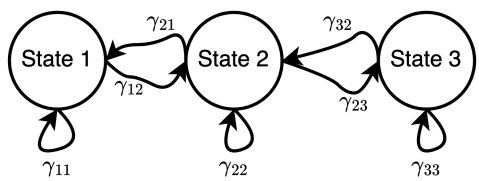

Figure 2. Transition graph of the constrained 3-state Markov chain.

is taken to be $\boldsymbol{u}(1) = (1, 0, 0)$ (it is fixed). State 2 refers to the cruise and state 3 to the approach. We use 20 different starting values to increase the chances of finding the global maximum. The state-dependent density we consider for the RoC is the Gaussian one. Initial values are chosen as follows:

- As the climb is known to last for some time, we draw γ_{11} from the uniform distribution $\mathcal{U}_{[0.8,0.95]}$ and we set $\gamma_{12} = 1 - \gamma_{11}$. Likewise, we draw γ_{21} from $\mathcal{U}_{[0.01,0.04]}$, γ_{22} from $\mathcal{U}_{[0.9,0.95]}$ (the cruise lasts some time), we fix $\gamma_{23} = 1 - \gamma_{22} + \gamma_{23}$ (after the cruise comes the approach), γ_{32} from $\mathcal{U}_{[0.01,0.04]}$ and $\gamma_{33} = 1 - \gamma_{32}$.
- The means of the normal distributions are drawn randomly as well as the standard deviations. Because there are 3 states, there is one mean and one standard deviation per state.

The choice of plausible starting values will avoid numerical instabilities.

Per se, HMM are unsupervised methods. The model does not return a segmentation involving the original data labels (climb, cruise, approach). Indeed, the states of the HMM are fully data-driven and do not have a predefined interpretation. Yet, the a priori meaning of the states has been integrated into the constraints such that there is no ambiguity in assigning the original labels. Results are very satisfactory. We compare them with a naive segmentation based on the following rules:

- If the altitude rate is positive (with some tolerance ε), the phase is said to be the climb.
- If the altitude rate is zero (with some tolerance ε), the phase is said to be the cruise.
- If the altitude rate is negative (with some tolerance ε), the phase is said to be the approach.

The tolerance parameter ε is chosen through trial and error. We also consider a fuzzy logic segmentation with values provided in [9]. A visual result for a typical flight is provided in Figure 3. With

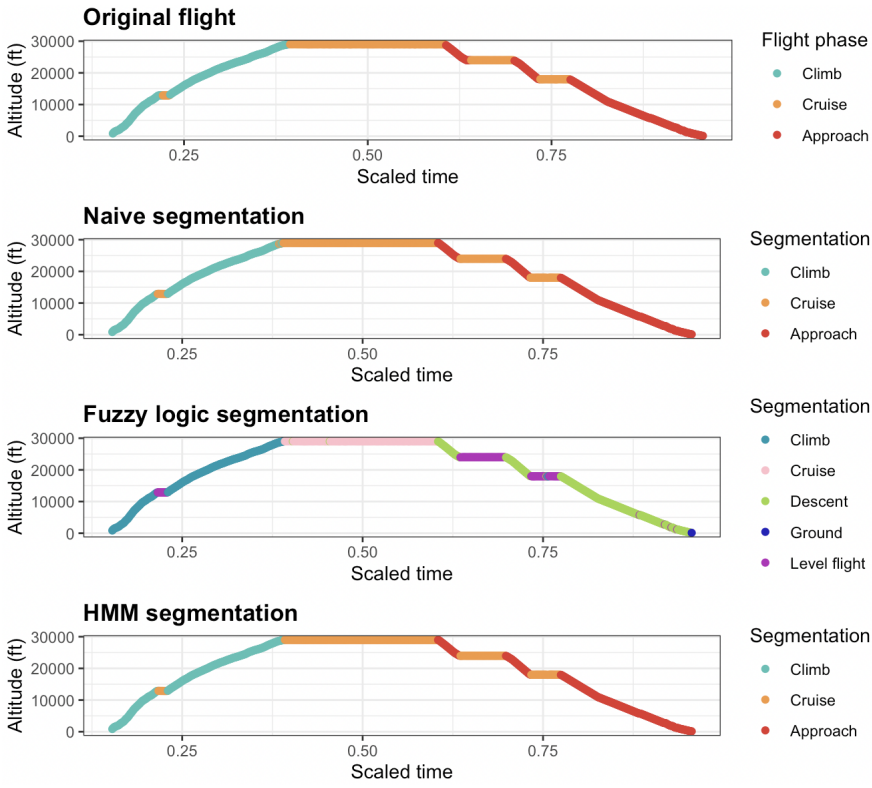

Figure 3. Identification results for a typical flight on the altitude profile.

the naked eye, the obtained segmentations appear very satisfactory. On this particular flight, there is no striking difference.

We examine four performance metrics to assess the quality of the results. First, we use the global accuracy per flight (proportion of points that are correctly labeled). Second, we compute the F-1 score for each phase separately. We also consider the number of unlikely transitions per flight and the number of transitions per flight. In this section, we consider two unlikely transitions: going (directly) from climb to approach and from approach to climb. The empirical distributions of these performance metrics over a subsample of 2,823 flights are presented as several box plots in Figure 4. Among the 2,868 flights, the subsample corresponds to flights that have at least the 3 flight phases of interest. Details on how to calculate the performance metrics using the fuzzy logic of [9] are provided in Appendix 1 (flight phases are not exactly the same). Regarding the global accuracy, it appears

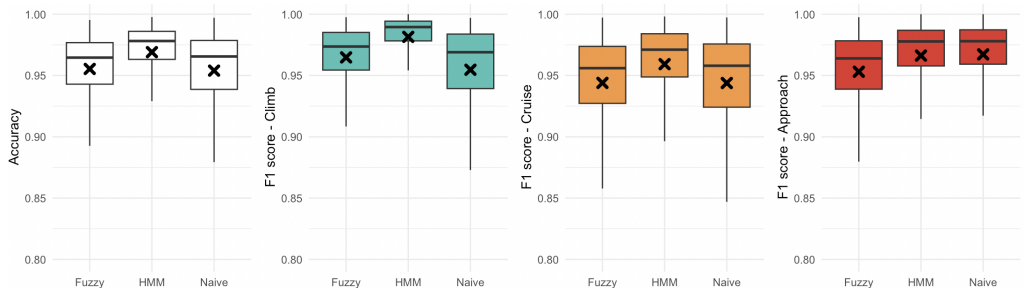

Figure 4. Box plots of the global accuracy [left box plot] and F-1 scores per state. The crosses correspond to the averages.

that our model is very good. The lower performance of the fuzzy logic is surely explained by the absence of any data pre-processing. Discussion on data pre-processing is postponed to Subsection 6.1. Allowing a tolerance ε for the naive method explains its good results. The dependency of fuzzy logic on the erratic nature of FDR data logically results in a large number of transitions. While the median number of transitions is 6 for the entire set of reference flights, it is 21 for fuzzy logic, 14 for the naive method, and 6 for the HMM. Taking into account the temporal dependence of the points helps avoid too frequent alternation between the phases. The inflation in the number of transitions translates into some unlikely transitions. The median number of unlikely transitions across the entire sample is 0, the same as for the naive logic and the HMM. However, with fuzzy logic, if the data is not pre-processed, at least 50% of the flights have 2 unlikely transitions. Unlikely transitions are inherently quite uncommon with HMMs because small transition probabilities make certain sequences very rare. Crucially, it must be highlighted that there is a non-zero proportion of invalid transitions in the reference data. About 8% of the reference flights in the subsample have at least an invalid phase transition. With our method, 91% of the flights have no invalid transitions (74% for the naive method, 21% for the fuzzy logic if no pre-processing is done).

4.2.2 Missing values

In the case of HMM, a simple adjustment needs to be made to the likelihood computation if data are missing. Suppose that the value of the RoC is missing at integer time $t = k$. Let L_T^{-k} be the likelihood of the observations. We have:

$$L_T^{-k} = \boldsymbol{u}(1)\boldsymbol{P}(x_1)\boldsymbol{\Gamma}\boldsymbol{P}(x_2)\boldsymbol{\Gamma}\boldsymbol{P}(x_3)\cdots\boldsymbol{\Gamma}\boldsymbol{P}(x_{k-1})\boldsymbol{\Gamma}^2\boldsymbol{P}(x_{k+1})\cdots\boldsymbol{\Gamma}\boldsymbol{P}(x_T)\mathbf{1}^\top. \tag{14}$$

The diagonal matrix $\boldsymbol{P}(x_k)$ has been replaced by the identity matrix.

If one assumes that missingness is ignorable, the *ignorable likelihood* (Equation 14) is a reasonable basis for estimating parameters. To be more precise, this likelihood may be used if one assumes that data are missing at random as defined by [23]. An important case in which this assumption does not hold is the state-dependent missingness case. Note that it is necessary to include time points with

missing observations to allow the state probabilities to be computed properly (simply ignoring the missing time points is not valid). Missing points may be consecutive or not. Figure 5 shows that the quality of the final segmentation is minimally affected by missing values.

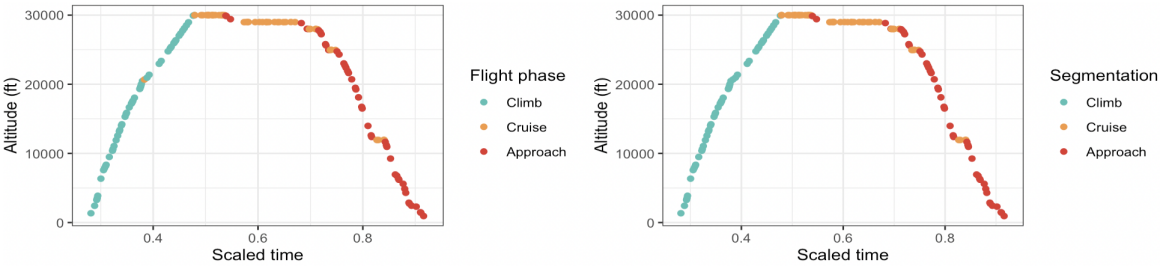

Figure 5. Identification results for a given flight. The initial segmentation [left] can be compared to our one [right]. 500 points are missing in this example (drawn uniformly at random).

5. Multivariate Hidden Markov Models

In order to detect more flight phases, multiple variables are required, and it is necessary to introduce multivariate HMM. The basic framework is presented in Subsection 5.1. The multivariate model for flight phase identification is detailed in Subsection 5.2.

5.1 The multivariate framework

In this section, $s \in \{2, ..., q\}$ series are used for the segmentation (multivariate case). We assume that, conditional on $\boldsymbol{C}^{(T)} = \{C_T, ..., C_1\}$, the random vectors $\boldsymbol{X}_1, ..., \boldsymbol{X}_T$ are mutually independent. This assumption is called *longitudinal conditional independence* and is the multivariate counterpart of the univariate conditional independence assumption (Equation 2).

Regarding the state-dependent distributions, there are two main things to specify:

- The distribution of the random vector $\boldsymbol{X}_t = \left(X_{t1}, X_{t2}, X_{t3}, ..., X_{tq}\right)$ in each of the m states of the parameter process. That is, we should specify:

$$p_i(\boldsymbol{x_t}) = \mathbb{P}(\boldsymbol{X}_t = \boldsymbol{x_t} \mid C_t = i), \forall t \geq 1, \forall i \in \{1, ..., m\}. \tag{15}$$

- The m joint distributions, a task that is far from being easy.

Once the distributions have been selected, the likelihood of a general multivariate HMM is easy to write down. It has the same form as that of the basic model, namely:

$$L_T = \boldsymbol{u}(1)\boldsymbol{P}(\boldsymbol{x_1})\boldsymbol{\Gamma}\boldsymbol{P}(\boldsymbol{x_2})\boldsymbol{\Gamma}\boldsymbol{P}(\boldsymbol{x_3})...\boldsymbol{\Gamma}\boldsymbol{P}(\boldsymbol{x_T})\boldsymbol{1}^\top, \tag{16}$$

where $\boldsymbol{x}_1, ...\boldsymbol{x}_T$ are the observations and $\boldsymbol{P}(\boldsymbol{x_t})$ is again the diagonal matrix with i-th diagonal element $p_i(\boldsymbol{x_t})$.

5.1.1 Contemporaneous conditional independence

The task of finding suitable joint distributions is greatly simplified if one can in addition assume *contemporaneous conditional independence*. The assumption of contemporaneous conditional independence is that the state-dependent joint probability $p_i(\boldsymbol{x}_t)$ is just the product of the corresponding

marginal probabilities. In other words, $\forall\ t \geq 1, \forall\ i \in \{1, ..., m\}$,

$$p_i(\boldsymbol{x}_t) = \prod_{j=1}^{q} \mathbb{P}\left(X_{tj} = x_{tj} \mid C_t = i\right). \tag{17}$$

Note that even when the model is assumed to verify longitudinal conditional independence and contemporaneous conditional independence, the Markov chain induces both serial dependence and cross-dependence in the component series. In the sequel, we will assume longitudinal conditional independence as well as contemporaneous conditional independence.

5.2 A multivariate model for flight phase identification

For the multivariate model, 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. Again, flight phases may be seen as states. We naturally specify a constrained 6-state multivariate HMM for which the transition graph of the corresponding Markov chain is represented in Figure 6. The transition probability matrix of the multivariate model is:

$$\boldsymbol{\Gamma_2} = \begin{pmatrix} \gamma_{11} & \gamma_{12} & 0 & 0 & 0 & 0 \\ 0 & \gamma_{22} & \gamma_{23} & 0 & 0 & 0 \\ 0 & 0 & \gamma_{33} & \gamma_{34} & 0 & 0 \\ 0 & 0 & \gamma_{43} & \gamma_{44} & \gamma_{45} & 0 \\ 0 & 0 & 0 & \gamma_{54} & \gamma_{55} & \gamma_{56} \\ \gamma_{61} & 0 & 0 & 0 & 0 & \gamma_{66} \end{pmatrix}. \tag{18}$$

The first state is a good candidate to represent the taxi phase. To ensure this, the initial distribution

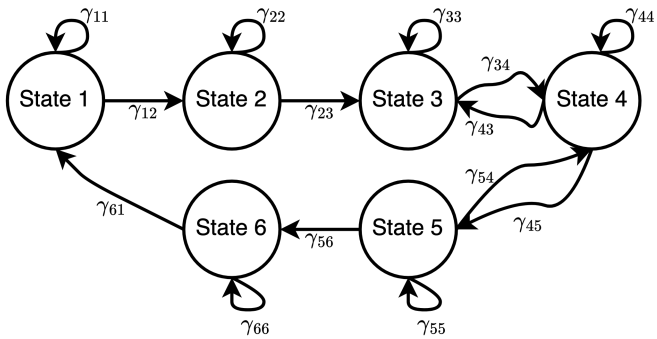

Figure 6. Transition graph of the constrained 6-state Markov chain.

is taken to be $\boldsymbol{u}(1) = (1, 0, 0, 0, 0, 0)$ (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 20 different starting values to increase the chances of finding the global maximum. 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 $\varepsilon_M = 0.05$, 0 otherwise). We use Bernoulli distributions as the state-dependent densities of this variable. Finally, a discretized version of first difference 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. The initial values are chosen in the same way as for the univariate model.

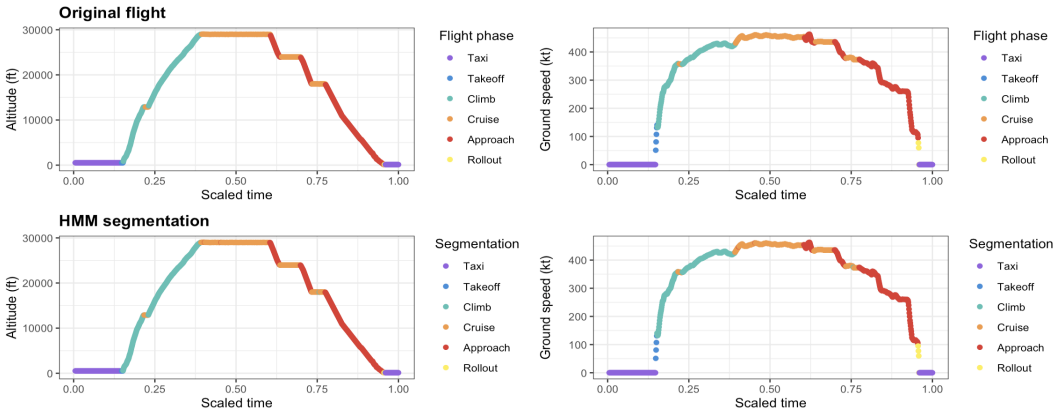

Figure 7. Identification results for a typical flight on the altitude profile and on the ground speed profile.

A visual result for a typical flight is provided in Figure 7. Results are very good from a visual perspective. The value of several performance metrics over a subsample are presented in Figure 8. Among the 2,868 flights, the subsample corresponds to flights that have at least the 6 flight phases of interest. If F-1 scores are very high, we can observe significant disparities among the flight phases.

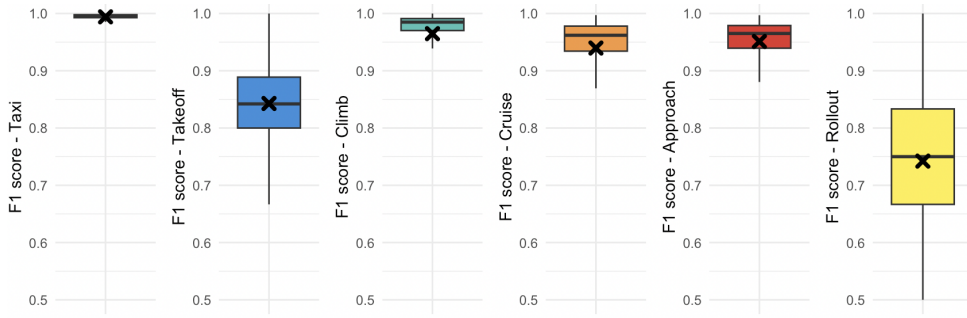

Figure 8. Evaluation of the performance. Box plots of the F-1 scores per state. The crosses correspond to the averages.

Strikingly, the F-1 score is lower for the takeoff and landing phases. Several reasons can explain this. First, these phases represent a very small number of data points across the entire trajectory (on average 6 points out of 1,000 for the takeoff and 4 out of 1,000 for the landing). Second, it may be necessary to include other variables to more accurately identify the takeoff and rollout phases (considering variables such as pitch angle may be interesting).

6. Discussions

6.1 Pre-processing data

Raw data may be erratic. To solve this problem, a common practice is to smooth the input curves with a kernel. The effect of smoothing on the quality of segmentation is quite clear when it comes to the number of transitions. In general, smoothing data tends to result in a lower number of transitions. Using kernel smoothing with a bandwidth value of 0.01 for the RoC and a bandwidth of 0.01 for the ground speed, the median number of transitions drops to 4 with the fuzzy logic. The median number of unlikely transitions drops to zero. To assess the benefits of pre-processing data, we adopt a grid search approach in which we vary the bandwidth values for the RoC and the ground speed.

We observe that the HMM achieves a better overall accuracy with minimal smoothing as illustrated in Figure 9. However, the global accuracy of the naive method and fuzzy logic improves with some smoothing of the RoC. It is not surprising that the bandwidth value of the ground speed doesn't play

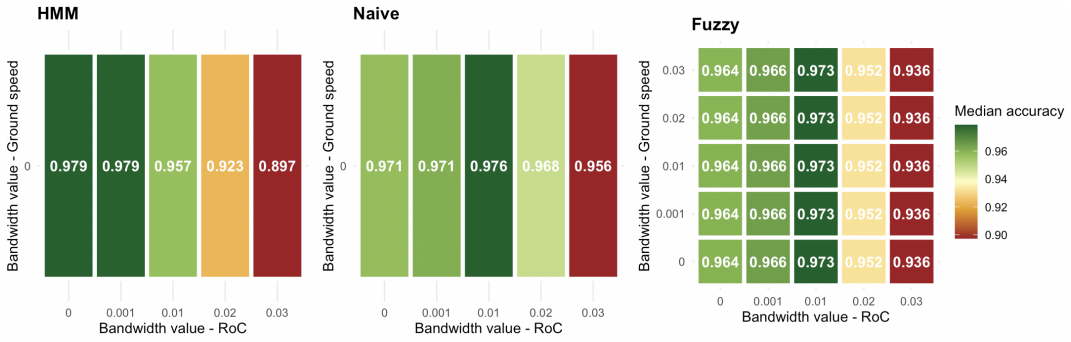

Figure 9. For each method, the median value of the overall accuracy for different bandwidth values. The naive method and the HMM do not use the ground speed as an input. In abuse of notation, a bandwidth value of 0 means that there has been no smoothing. A subsample of several hundred flights was selected to limit the calculation time.

any role in the evolution of the fuzzy logic global accuracy. Indeed, ground speed primarily serves to segment the ground phase, a phase that is not of interest at this stage of the analysis.

When considering the multivariate HMM for detecting 6 flight phases, there is a slow decrease in the median value of the overall accuracy with an increase in the bandwidth value of the RoC (Figure 10). A similar pattern emerges with the distributions of F-1 scores by phase as we observe in Figure 11. The effect of data preprocessing is particularly significant on the decoded number of phases, as observed in Figure 12.

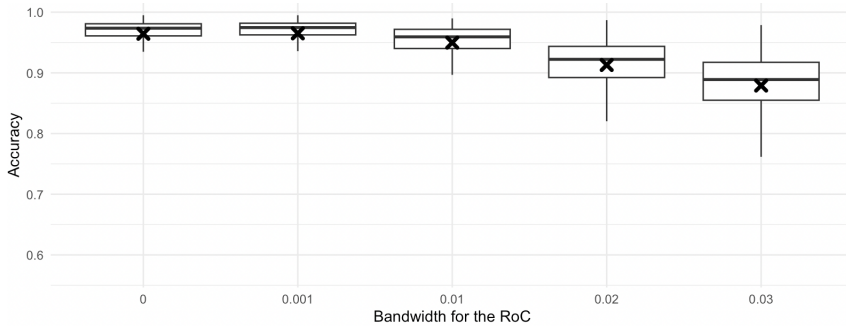

Figure 10. For each bandwidth value of the RoC, box plots of the global accuracy for the multivariate HMM. In abuse of notation, a bandwidth value of 0 means that there has been no smoothing. A sample of several hundred flights was selected to limit the calculation time.

6.2 Uncertainty

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 (Equation 10). For a given point in the flight, the sum of the probabilities of belonging adds up to 1. An illustration is provided for the multivariate model (Figure 13).

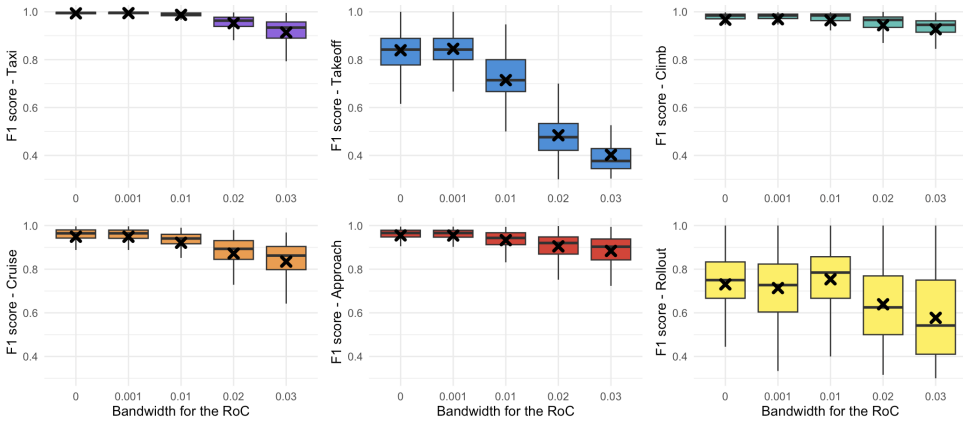

Figure 11. For each bandwidth value of the RoC, box plots of the F-1 scores for the multivariate HMM. In abuse of notation, a bandwidth value of 0 means that there has been no smoothing. A sample of several hundred flights was selected to limit the calculation time.

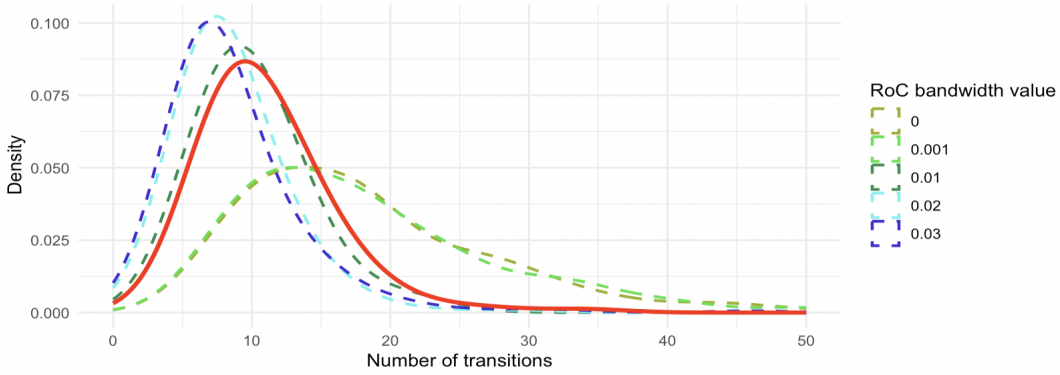

Figure 12. For each bandwidth value of the RoC, density of the number of decoded transitions. The distribution of the number of transitions in the reference data is in red.

6.3 Unknown number of states

Unlike fuzzy logic and supervised methods, where it is necessary to know the number of phases in advance, HMMs can be employed even when flight phases are not known. It is the case when extracting continuous flights from a scattered ADS-B datase. Another interesting application is the maneuver detection problem for rotorcraft and fixed-wing aircraft. The order of maneuvers is not predetermined.

Regarding HMMs, it not possible to propose a model with constraints because the sequence of phases is unknown by assumption. The decoding step (local or global decoding) will provide a segmentation with labels that need to be interpreted afterward. The main difficulty with HMM is the following: it won't be enough to test a different number of states and to choose the best model because in a basic HMM with m states, increasing m always improves the fit of the model (as judged by the likelihood) at the cost of a quadratic increase in the number of parameters. It is common *model selection* problem in statistics. In the frequentist approach, a common criterion is the Akaike Information Criterion (AIC):

$$\mathrm{AIC} = -2 \log L_T + 2p \tag{19}$$

where $\log L_T$ is the log-likelihood of the fitted model and p denotes the number of parameters of

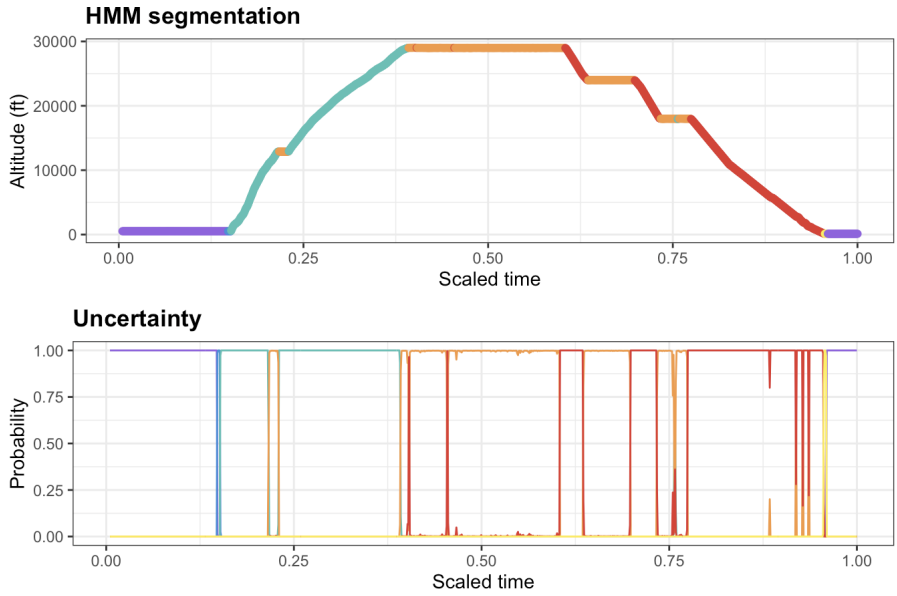

Figure 13. Segmentation of the 6 main phases of the flight using the multivariate HMM and probabilities of belonging to each class.

the model. The first term is a measure of fit, and decreases with increasing number of states m. The second term is a penalty term, and increases with increasing m. In the Bayesian approach, the Bayesian information criterion (BIC) differs from AIC in the penalty term:

$$\text{BIC} = -2 \log L_T + p \log T \tag{20}$$

where $\log L$ and p are as for the AIC, and T is the number of observations. Compared to the AIC, the penalty term of the BIC has more weight for $T > \exp(2)$, which holds in most applications. Thus the BIC often favours models with fewer parameters than does the AIC. More theoretical details can be found in [22] (Subsection 2.6.2).

We illustrate this use of HMMs with the segmentation of a helicopter flight. We have downloaded some ADS-B data from the Opensky Network's Impala shell for the helicopter with registration SE-JPU (ICAO24: 4aaa15) operated by the Swedish National Police. We select a flight from June 7, 2021. The flight has a complicated shape as shown in Figure 14. It goes without saying that the obtained segmentation depends largely on the inputs chosen for the HMM. For instance, the detection of hover manoeuvres can only be done correctly with the ground speed variable.

We consider a multivariate HMM with the longitude first differences, the latitude first differences, the ground speed (m.sec^{-1}), and the vertical rate (m.sec^{-1}). We use 100 different starting values to increase the chances of finding the global maximum. For each iteration and for each number of states, we compute the BIC. The final number of states is the one that has reached the lowest median value of the BIC, that is to say 8 in this case as shown in Figure 15. The final segmentation is shown in Figure 16. It is observed that certain states are easily interpretable. States 1 and 3 correspond to climbing phases at medium (state 1) and high (state 3) speeds. State 2 is characterized by significant oscillations in the first differences of longitude and latitude. In fact, these are rotations (as clearly seen with the track angle). State 5 corresponds to very rapid horizontal movement. The same goes for state 7, but the direction is different. State 6 corresponds to a descent.

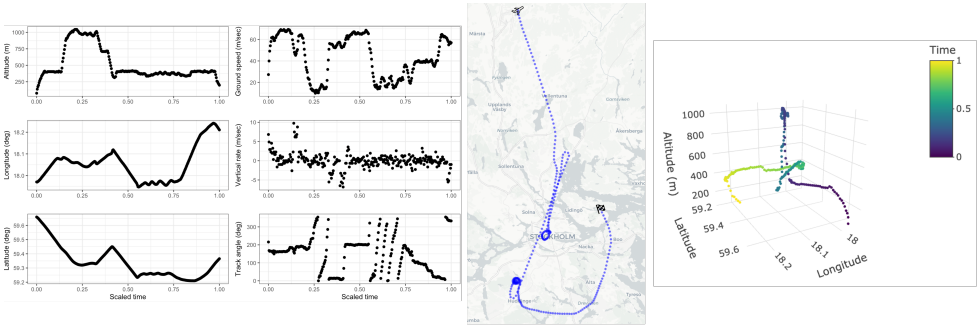

Figure 14. Visualization of the helicopter flight. Altitude, longitude, latitude, ground speed, vertical rate, and track angle profiles [left], flat view [center], and three-dimensional view [right]. Time is scaled so that the flight starts at $t = 0$ and ends at $t = 1$. There are $T = 297$ points. Time resolution is 10 seconds. The track angle is a clockwise angle from the geographic north.

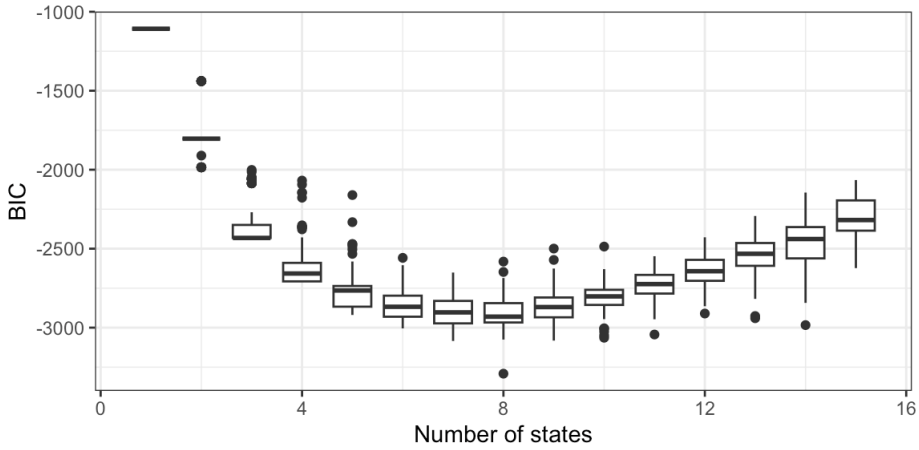

Figure 15. Distribution of the BIC value for each number of states.

7. Conclusion and perspectives

For commercial aviation, and when the number of states is predetermined, the HMM we propose can detect up to 6 flight phases (taxi, takeoff, climb, cruise, approach, rollout). The overall accuracy on nearly 3,000 flights is about 97% (median accuracy). These results are highly competitive with the state-of-the-art literature. When looking at each phase separately, notable differences emerge. While the taxi phase is identified almost perfectly, takeoff and landing appear to be more challenging to detect. We believe that this is primarily explained by the fact that these phases represent, on average, 6 and 4 points, respectively, out of the 1,000 points in the trajectory. F-1 scores associated to these flight phases are still very high. In any case, HMMs seem to adapt well to the fine granularity of FDR data. Missing values in ADS-B data do not pose any issues.

For each point, it is possible to obtain a probability of belonging to each class, which is not the case with most existing methods (namely fuzzy logic). Depending on operational applications, one may focus on points for which the flight phase is decoded with high confidence.

The strength of HMMs lies in their great flexibility. When the number of phases is not known or their sequence is not predetermined (as it is the case with helicopters, for example), HMMs can still be used. We have illustrated this point with a flight example for which the HMM produces interpretable

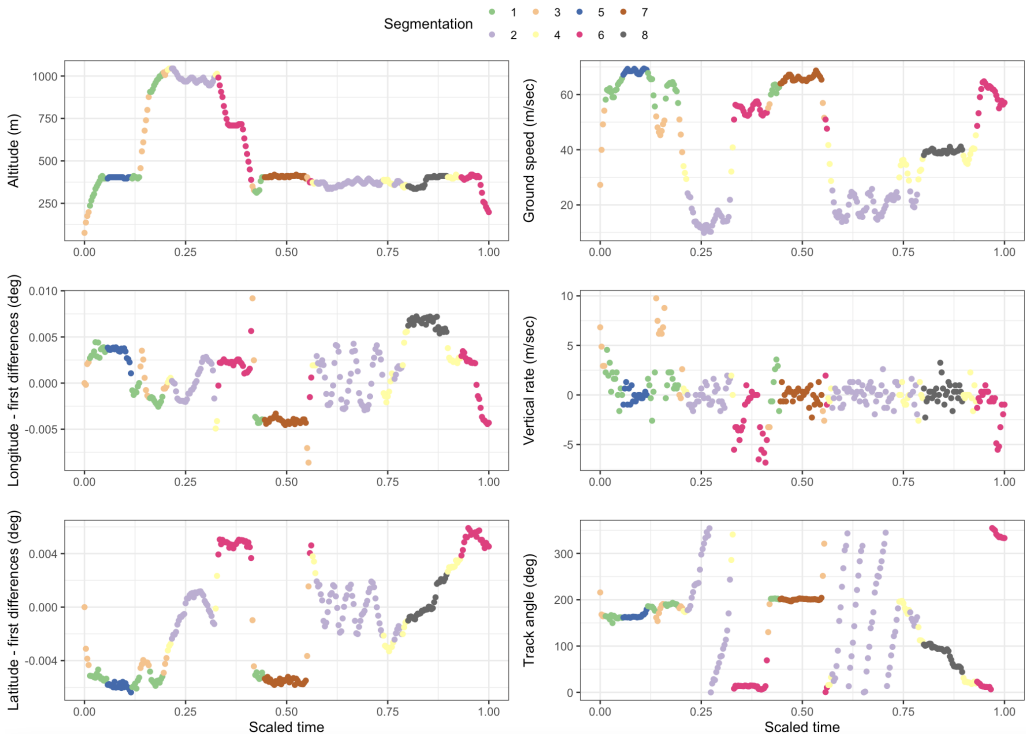

Figure 16. Identification results for the helicopter flight.

phases. By working on the inputs, we believe that it is possible to detect the relevant maneuvers for each application.

Several exciting aspects fall out of the scope of the paper. First, it would be possible to define new performance metrics. Since flight phase segmentation produces a sequence, it would be interesting to compare the obtained sequence to the one of the reference flight. A first step in this direction would be to investigate distances employed for text analysis. Next, it would be interesting to study the evolution of the segmentation quality by integrating covariates into HMMs. In summary, HMMs offer numerous opportunities for aviation research applications.

Appendix 1. Issues with the definition of flight phases

Defining performance metrics is complicated by variations in terminology. This is a classic issue in the literature on flight phase detection. For example, there is no such thing as a level flight subphase in our reference flight phase labels. In order to calculate various performance metrics, the subphase is typically identified using fuzzy logic and then renamed climb, cruise or approach. If half of the flight has already been completed and if the altitude is below 10,000 feet, the level flight subphase is labeled approach. If half of the flight is not done and if the altitude is below 10,000 feet, the level flight subphase is labeled climb. Otherwise, it is labeled cruise. These few adjustments allow for accommodating the definition of flight phases used in the reference data.

Similarly, if a cruise phase is detected below 10,000 feet with the naive method or the HMM, it is renamed to climb or approach depending on whether half of the flight has already been completed or not. This small adjustment is due to the fact that the altitude profile is not used as an input for these models. It would be possible to avoid this last transformation by incorporating a binary variable

into the model, indicating whether the altitude is above or below 10,000 feet at the cost of a less parsimonious model.

Author contributions

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

Open data statement

Data are accessible on the NASA DASHlink project page.
url : https://c3.ndc.nasa.gov/dashlink/projects/85/ (last check: 2023/11/17)

Reproducibility statement

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

References

[1] CAST/ICAO Common Taxonomy Team (CICTT). *Phase of flight*. Tech. rep. International Civil Aviation Organization (ICAO), Apr. 2013. URL: https://www.intlaviationstandards.org/Documents/PhaseofFlightDefinitions.pdf.

[2] International Air Transport Association. *Safety Report 2014*. Tech. rep. Apr. 2015.

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

[4] 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.

[5] Richard Alligier, David Gianazza, and Nicolas Durand. "Machine Learning and Mass Estimation Methods for Ground-Based Aircraft Climb Prediction". In: *IEEE Transactions on Intelligent Transportation Systems* 16.6 (Dec. 2015), pp. 3138–3149. DOI: 10.1109/TITS.2015.2437452.

[6] Nataliia Kuzmenko, Ivan Ostroumov, Yurii Bezkorovainyi, Yuliya Averyanova, Vitalii Larin, Olha Sushchenko, Maksym Zaliskyi, and Oleksandr Solomentsev. "Airplane Flight Phase Identification Using Maximum Posterior Probability Method". In: *IEEE 3rd International Conference on System Analysis & Intelligent Computing (SAIC)*. 2022, pp. 1–5. DOI: 10.1109/SAIC57818.2022.9922913.

[7] 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.

[8] 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).

[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] 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.

[11] Ingmar Visser and Maarten Speekenbrink. "depmixS4: An R Package for Hidden Markov Models". In: *Journal of Statistical Software* 36 (Aug. 2010), pp. 1–21. DOI: 10.18637/jss.v036.i07.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] A. Viterbi. "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm". In: *IEEE Transactions on Information Theory* 13.2 (Apr. 1967), pp. 260–269. DOI: 10.1109/TIT.1967.1054010.

[22] Ingmar Visser and Maarten Speekenbrink. *Mixture and Hidden Markov Models With R*. 1st ed. Springer International Publishing AG, June 2022. ISBN: 978-3-031-01438-3.

[23] Donald B. Rubin. "Inference and Missing Data". In: *Biometrika* 63.3 (1976), pp. 581–592. DOI: 10.2307/2335739.