CHAPTER TEN

Hidden Markov model applications

Aviation prognostics

Thanos Kontogiannis, Mariana Salinas-Camus, and Nick Eleftheroglou

Intelligent Sustainable Prognostics Group, Aerospace Structures and Materials Department, Faculty of Aerospace Engineering, TU Delft, Delft, Netherlands

Chapter points

- The family of hidden Markov models can effectively be used for the prediction of the remaining useful life (RUL) of aviation assets.
- HSMMs present superior prognostic performance when compared to HMMs, mainly due to the relaxed assumptions for the duration probabilities of the degradation process.
- The novel time-dependent prognostic measure provides more accurate RUL predictions than traditional time-invariant measures.

10.1. Introduction

Prognostics and health management (PHM) involves the continuous monitoring of equipment health and the prediction of potential failures before they occur, allowing for predictive maintenance actions for enhanced reliability and cost reduction. By extracting information coming from condition monitoring (CM) techniques, a prognostic model can predict the remaining useful life (RUL). The RUL is defined as the time in which an engineering system can continue to perform its intended function efficiently and effectively before it becomes unreliable or reaches the end of its operational life.

The RUL can be predicted by prognostic models, which are physics-based, data-driven, or hybrid. The latter type combines characteristics of both physics-based and data-driven. For some engineering systems, physics-based prognostic models are not feasible due to the high complexity of their governing physics. For example, aviation structures use composite materials, of which there is no physics-based prognostic model yet developed due to the limited understanding of the governing physics of damage initiation

Stochastic Modeling and Statistical Methods
https://doi.org/10.1016/B978-0-44-331694-4.00015-3

and propagation, especially for dynamic loading conditions and complex stacking sequences.

Therefore data-driven prognostic models are more extensively used since it is not necessary to know the physics of the engineering system. Instead, it is only necessary to have historical CM data of previous run-to-failure experiments. There are different types of data-driven prognostic models. Generalizing, we can categorize them into two big groups: machine learning (ML) models and stochastic models.

Stochastic models, such as hidden Markov models (HMMs) and their variants, have been used successfully in prognostics. HMMs model the degradation process of the engineering system Rabiner (1989). The hidden states represent the levels of degradation or the system's health, whereas the observed data corresponds to CM data. By training the HMM on historical data, it is possible to estimate the degradation and observation processes. The degradation process is defined with the transition probabilities between hidden states. The observation process is defined with the emission probabilities, which describe the likelihood of observing a particular data point given the current hidden state. Therefore once the degradation and observation processes are obtained, the HMM can be utilized to estimate the current hidden state of a new sequence of sensor data, i.e., test data, of the engineering system, using, for example, the Viterbi algorithm. Once the current hidden state is estimated, a prognostic measure is used to predict the RUL. In the literature, different variants of HMMs, training schemes, and prognostic measures have been developed.

HHMs have been used to estimate the system's health and to predict the RUL of turbofan engines in Giantomassi et al. (2011). The results show good performance for the second half of the lifetime of the engines. For the first half of the trajectory, the HMM has lower accuracy, which is caused by the misidentification of the first states. In Le et al. (2014) a multibranch HMM is used to perform prognostics for bearings in different degradation modes. In this case, three branches are used since three different degradation modes are identified. Therefore three different HMMs are trained with data from bearings that experienced a particular degradation mode. For the test data, a diagnostic measure identifies the degradation mode to switch to the corresponding HMM branch.

However, in most cases, HMM is not considered a suitable model for prognostics for more complex systems given its simplifying assumptions. One of these assumptions is that the sojourn time of the hidden states is characterized by an exponential distribution (continuous case) or a geometric distribution (discrete case). Consequently, an extension of the classical

HMM is used: the hidden semi-Markov model (HSMM) Yu (2010), which allows for variable sojourn time of the hidden states. Naturally, due to the increased complexity of the HSMM, there is a higher computational cost in comparison with HMM.

HSMMs are used in Dong and He (2007) for the diagnosis and prognosis of hydraulic pumps. The paper includes a comparison between HMMs and HSMMs, with the latter having better results in terms of RUL accuracy. HSMMs have better performance since the duration of the states is modeled by an explicit Gaussian probability function. This allows for better modeling of the degradation process compared to the exponential distribution modeling of the HMMs. Another important point is the fact that HMMs include information only from the previous state, and HSMMs include information from the previous d states. The increase in information used in HSMMs has a positive influence on the model performance. The prognostic measure used here is an iterative equation, which includes the calculation of the RUL prediction in the next time step and the probabilities of transition between neighbor states.

An HSMM with duration-dependent state transition probabilities (DD-HSMM), is presented in Wang et al. (2014) for prognostics on rolling element bearings. The DD-HSMM has duration-dependent state transition probabilities, unlike HSMM, which has duration-invariant ones. The introduction of duration-dependent state probabilities means that the model can include information about the stages of the engineering system, and its age. The prognostic measure used in this case consists of the sum of the expected duration from state i to state $N-1$. The results show that the DD-HSMM method is more accurate for RUL prediction than HSMM since it has a self-updating capability.

In Liu et al. (2015), an adaptive HSMM (AHSMM) is used for the prognosis of hydraulic pumps. The adaptive training uses maximum likelihood linear regressions to represent the differences in multiple sensors. Therefore the AHSMM can use multisensor information, and the duration distributions of the states are calculated accordingly. The results show an improvement in the four error metrics when comparing it to HSMMs. In this paper, the prognostic measure used is the same as the one presented in Wang et al. (2014).

Moreover, in the current state-of-the-art, more complex and advanced variations of HMMs have emerged. For example, the nonhomogeneous hidden semi-Markov model (NHHSMM) was introduced in Moghaddass and Zuo (2014). Unlike traditional HSMMs, the NHHSMM relaxes one

assumption by incorporating the age of the engineering system. Therefore the sojourn time spent in a hidden state is time-dependent. NHHSMM has been used in multiple case studies, such as turbofan engines (Moghaddass and Zuo, 2014), composite specimens (Eleftheroglou and Loutas, 2016), lithium-polymer batteries (Eleftheroglou et al., 2019), valves (Loutas et al., 2019) and others. Other variants include an adaptive extension of the NHHSMM, called the adaptive nonhomogeneous hidden semi-Markov model (ANHHSMM) (Eleftheroglou et al., 2020). This model has an adaptation mechanism, which changes the model parameters when dealing with unseen testing data. The ANHHSMM was trained on composite specimens, which were subjected to fatigue loading and tested on specimens that were subjected to fatigue and impact loading. The result showed an increase in performance when compared to the NHHSMM and the potential to integrate unseen data into prognostics. Finally, Eleftheroglou et al. (2024) developed the Similarity Learning Hidden Semi-Markov Model (SLHSMM), an extension of the NHHSMM. This model assesses the similarity between a test degradation history and training data to perform a re-estimation process. As a result, it provides more reliable RUL predictions by not only improving accuracy but also reducing the confidence intervals of the RUL estimates.

Nonetheless, this study conducts a comparison between an HMM and an HSMM, given their established status within the field, along with the introduction of a novel prognostic measure. The new prognostic measure includes time-dependent features that improve the accuracy of the RUL predictions in comparison with the time-invariant expression. The paper is organized as follows: Section 10.2 offers a theoretical background of HMMs and HSMMs. Section 10.3 describes the prognostic measure. The case study, including the dataset description, parameters initialization, and results and discussion, is presented in Section 10.4. Finally, the paper is concluded in Section 10.5.

10.2. Degradation modeling

In the current study, an HMM and an HSMM are chosen to model the degradation process of the assets. In Sections 10.2.1 and 10.2.2, the definition, the assumptions, the procedure, and required expressions needed for the three fundamental problems associated with the modeling of the process (likelihood, decoding, and learning), will be provided for the HMM and HSMM, respectively.

10.2.1 Hidden Markov model

An HMM is a stochastic model, which describes a system evolving through time. The state of the system is hidden from the observer and can only be inferred from the observations emitted by the system in a probabilistic manner. An HMM is then defined by the number of states (N), the number of distinct observations (M), transition probabilities between states (A), the probability distribution of observations in each state (B), and the initial state (π). The complete parameter set of the model is denoted as $\lambda = (A, B, \pi)$.

- N: number of states. Individual states are denoted as $S = \{S_1, S_2, ..., S_N\}$, and the state at time t as q_t.
- M: number of distinct observation symbols per state. Individual observations are denoted as $V = \{v_1, v_2, ..., v_M\}$.
- State transition: the state transition probability distribution is denoted as $A = \{a_{ij}\}$, where $a_{ij} = P[q_{t+1} = S_j | q_t = S_i]$.
- Observation distribution: the observation symbol probability distribution in state j, $B = b_j(k)$, where $b_j = P[v_k \text{ at } t | q_t = S_j]$, with $1 \leq j \leq N$ and $1 \leq k \leq M$.
- Initial state: the initial state distribution $\pi = \{\pi_i\}$, where $\pi_i = P[q_1 = S_i]$ with $1 \leq i \leq N$.

To model the evolution of a system, the HMM has assumptions. First, the probability of transitioning to a future state only depends on the current state, which is known as the Markov property. Second, the distribution of the sojourn time in a state corresponds to an exponential distribution (continuous case) or geometric distribution (discrete case). Third, one observation is emitted per state.

To characterize the degradation and observation processes within a system, three fundamental problems must be addressed. The three fundamental problems are inherent to HMMs and are key aspects of understanding and using HMMs to model real-life situations effectively.

- **Likelihood**: Calculate the likelihood $P(O|\lambda)$ given the model λ and an observation sequence O.
- **Decoding**: Estimate the sequence of hidden states Q that best explains the observations, given the model λ and an observation sequence O.
- **Learning**: Estimate the transition and emission matrix that best describes the degradation process of the observation sequence O.

The forward-backward algorithm is used for the first problem, which is the **calculation of the likelihood**. However, to calculate the likelihood, only the forward variable is needed. The forward variable is defined as $\alpha_t(i) = P(O_1 O_2 ... O_t, q_t = S_i | \lambda)$. This is the probability of a partial observation sequence (until time t) and state S_i at time t, given the model λ. The forward variable can be solved inductively as follows:

1. Initialization: Here the forward probabilities are initialized as the joint probability of state S_i and initial observation O_1.

$$\alpha_1 = \pi_i * b_i(O_1), \quad \text{where } 1 \leq i \leq N \tag{10.1}$$

2. Induction:

$$\alpha_{t+1}(j) = \left(\sum_{i=1}^{N} \alpha_t(i) * a_{ij} \right) * b_j(O_{t+1}), \quad \text{where } 1 \leq t \leq T-1 \text{ and } 1 \leq j \leq N \tag{10.2}$$

3. Termination:

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{10.3}$$

The second problem, **decoding**, can be solved with the Viterbi algorithm. For this algorithm, it is necessary to define the variable $\delta_t(i)$, which is the highest probability for a path at time t ending at state q_i, and the variable $\psi_t(i)$, which is the path that has the highest probability until state q_i at time t.

1. Initialization:

$$\delta_1(i) = \pi_i * b_i(O_1), \quad \text{where } 1 \leq i \leq N, \tag{10.4}$$

$$\psi_1(i) = 0 \tag{10.5}$$

2. Induction:

$$\delta_t(i) = \max_{1 \leq i \leq N} \delta_{t-1}(i) * a_{ii} * b_j(O_t), \quad \text{where } 2 \leq t \leq T, 1 \leq j \leq N, \tag{10.6}$$

$$\psi_t(j) = \arg\max_{1 \leq i \leq N} \delta_{t-1}(i) * a_{ii} \tag{10.7}$$

3. Termination:

$$P(Q, O) = \max_{1 \leq i \leq N} \delta_T(i), \tag{10.8}$$

$$\hat{q}_T = \arg\max_{1 \leq i \leq N} \delta_T(i), \tag{10.9}$$

$$\hat{q}_t = \psi_{t+1}(\hat{q}_{t+1}), \quad \text{where } t = T-1, T-2, ..., 1 \tag{10.10}$$

In the termination step, Eq. (10.9) is the best last state, and Eq. (10.10) represents the backtracking used to get the best states in each time step.

Finally, the last problem of the HMM, **the learning problem**, is solved by the expectation-maximization (E-M) algorithm. The E-M algorithm iterates between two steps: the E-step, where the expected values of the hidden states are computed based on the current model parameters using the forward-backward algorithm, and the M-step. During the M-step, the model parameters are updated to maximize the likelihood of the k observation sequences (O) (Eq. (10.11), Eq. (10.12)) by incorporating the expected hidden states obtained from the E-step. This iterative process continues until convergence, refining the parameter estimates and improving the model's fit to the data.

$$L(\lambda, O^{(1:K)}) = \prod_{k=1}^{K} P(O^{(k)}|\lambda) \xrightarrow{L'=log(L)} L'(\lambda, O^{(1:K)}) = \sum_{k=1}^{K} log(P(O^{(k)}|\lambda)), \tag{10.11}$$

$$\lambda^* = arg\max_{\lambda} \left(\sum_{k=1}^{K} log(P(O^{(k)}|\lambda)) \right) \tag{10.12}$$

The E-M algorithm utilizes both the forward and backward variables of the forward-backward algorithm. Therefore before introducing the E-M algorithm, it is necessary to define the backward variable. The backward variable is defined as $\beta_t(i) = P(O_{t+1}O_{t+1}...O_T|q_T = S_i, \lambda)$, so $\beta_t(i)$ is the probability of the future observations from $t+1$ given the current state S_i and the model λ. In the same manner as the forward variable, the backward variable can be solved inductively.

1. Initialization:

$$\beta_T(i) = 1, \quad \text{where } 1 \le i \le N \tag{10.13}$$

2. Induction:

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} * b_j(O_{t+1}) * \beta_{t+1}(j), \quad \text{where } t = T-1, T-2, ..., 1, \text{ and } 1 \le i \le N \tag{10.14}$$

3. Termination:

$$P(O|\lambda) = \sum_{j=1}^{N} \pi_j * b_j(O_1) * \beta_1(j) \tag{10.15}$$

For the E-step, the variables $\gamma_t(j)$, which is the probability of being in a state j at time t, given the observations and the model parameters, and $\xi_t(i,j)$, which is the probability of being in a state i at time t and state j at time $t+1$, given the observations and the model parameters. Both $\gamma_t(j)$ and $\xi_t(i,j)$ can be written in terms of the forward α and backward β probabilities:

$$\gamma_t(j) = P(q_t = j|O,\lambda) = \frac{\alpha_t(j) * \beta_t(j)}{P(O|\lambda)}, \tag{10.16}$$

$$\xi_t(i,j) = P(q_t = i, q_{t+1} = j|O,\lambda) = \frac{\alpha_t(i) * a_{ij} * b_j(O_{t+1}) * \beta_{t+1}(j)}{\sum_{j=1}^{N} \alpha_t(j) * \beta_t(j)} \tag{10.17}$$

In the M-step, the variables $\gamma_t(j)$ and $\xi_t(i,j)$ are used to reestimate the new probabilities for the transition A and emission B matrices:

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \sum_{k=1}^{N} \xi_t(i,k)}, \tag{10.18}$$

$$\hat{b}_j(v_k) = \frac{\sum_{t=1 s.t. O_t = v_k} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}. \tag{10.19}$$

In the particular case of prognostics and this paper, some assumptions are made. First, the last state is not hidden but observable and represents failure. Second, in the failure state, only one observation value is emitted. Third, only left-to-right transitions are allowed, and the transition can occur only to a neighbor's hidden state. Fourth, the initial state is always the first state since it is assumed that the engineering system is as good as new.

10.2.2 Hidden semi-Markov model

For the application of aviation prognostics, a comparison between the HMM and HSMM will be presented. An HSMM is an extension of HMM, which introduces a variable duration for each state, thus allowing the underlying process to be semi-Markov. Therefore the assumption of one emitted observation per state is also relaxed since the number of observations emitted depends on the time spent in each state (duration d). The parameter set for the complete definition of an HSMM follows the one presented in Section 10.2.1, with the following modifications and assumptions:

- Observation distribution: The probability distribution for the observations is assumed to be a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$. Therefore the B parameters are the parameters of the μ and σ^2 observation vectors,

with each row representing the hidden state and the values representing the mean and standard deviation of the $\mathcal{N}$ probability distribution relating to each state producing v observation.

- M: since the observation process is modeled with a continuous probability distribution (Gaussian), the indicator space consists of all the real numbers, so $M \to \infty$ and $V = \{v \in \Re\}$.
- D: number of integer values in the space $\{1, 2, \ldots, D\}$ that the state duration random variable d obtains. Therefore the degradation process is modeled in a nonparametric distribution.

To solve the three fundamental problems of HSMM, namely the likelihood calculation, decoding and learning, the same procedure described in Section 10.2.1 is followed. The same definitions of the calculated variables are followed. However, due to the added duration probability variable, a modification in the expressions of some of the variables to include the duration probabilities is required.

- **Likelihood calculation:** The procedure for calculating α:
  1. Initialization: Eq. (10.1)
  2. Induction:

$$\bar{\alpha}_{t+1}(i, d) = \sum_{j=1, j\neq i}^{N} \left(\alpha_{t+1-d}(j) * a_{ji},\right) * \prod_{\tau=t-d+1}^{t} \left(b_i(O_\tau)\right) * p_i(d), \tag{10.20}$$

$$\alpha_{t+1}(i) = \sum_{d=1}^{D} \bar{\alpha}_{t+1}(i, d), \tag{10.21}$$

$$\text{where } 2 \leq t \leq T-1, \text{ and } 1 \leq i \leq N$$

  3. Termination: Eq. (10.3).
- **Decoding:**
  1. Initialization: Eq. (10.4), Eq. (10.5)
  2. Induction:

$$\delta_t(j, d) = \max_{\substack{1\leq i\leq N,\, i\neq j,\\ 1\leq d'\leq D}} \delta_{t-d}(i, d') * a_{ij} * \prod_{\tau=t-d+1}^{t} b_j(O_\tau), \tag{10.22}$$

$$\psi_t(j, d) = \arg\max_{\substack{1\leq i\leq N,\, i\neq j,\\ 1\leq d'\leq D}} \delta_{t-d}(i, d') * a_{ij} * \prod_{\tau=t-d+1}^{t} b_j(O_\tau), \tag{10.23}$$

$$\text{where } 2 \leq t \leq T,\ 1 \leq j \leq N.$$

3. Termination:

$$P(Q, O) = \max_{\substack{1\le i\le N,\\ 1\le d\le D}} \delta_T(i, d), \tag{10.24}$$

$$\hat{q}_T = \arg\max_{\substack{1\le i\le N,\\ 1\le d\le D}} \delta_T(i, d), \tag{10.25}$$

$$\hat{q}_t = \psi_{t+1}(\hat{q}_{t+1}), \quad \text{where } t = T-1, T-2, ..., 1. \tag{10.26}$$

- **Learning:** The equations for maximizing the likelihood (Eq. (10.11), and Eq. (10.12)) remain the same. For the backward probability β we define the following:
  1. Initialization: Eq. (10.13)
  2. Induction:

$$\beta_t(i) = \sum_{\substack{j=1,\\ j\neq i}}^{N} \left(a_{ij} * \sum_{d=1}^{D} \left(p_i(d) * \beta_{t+d}(i) * \prod_{\tau=t+1}^{t+d} b_j(O_\tau) \right) \right) \tag{10.27}$$

  3. Termination: Eq. (10.15).

  For the E-step, for the γ variable, we follow Eq. (10.16). For the rest, we define the following:

$$\xi_t(i, j) = \frac{\sum_{d=1}^{D} \left(\alpha_t(i) * a_{ij} * \beta_{t+d}(j) * \prod_{\tau=t+1}^{t+d} b_j(O_\tau) \right)}{\sum_{j=1}^{N} \alpha_t(j) * \beta_t(j)}, \tag{10.28}$$

$$\eta_t(j, d) = \bar{\alpha}_t(j, d) * \beta_t(j). \tag{10.29}$$

  For the M-step, for $\hat{a}_{ij}$, we follow Eq. (10.18). For the reestimation of the emission parameters (μ, σ^2) and the duration probability matrix $p_j(d)$, we define the following:

$$\hat{\mu}(j) = \frac{\sum_{t=1}^{T} \gamma_t(j) * O_t}{\sum_{t=1}^{T} \gamma_t(j)}, \tag{10.30}$$

$$\hat{\sigma^2}(j) = \frac{\sum_{t=1}^{T} \gamma_t(j) * (O_t - \mu(j))^2}{\sum_{t=1}^{T} \gamma_t(j)}, \tag{10.31}$$

$$\hat{p}(j, d) = \frac{\sum_{t=1}^{T} \eta_t(j, d)}{\sum_{d=1}^{D} \sum_{t=1}^{T} \eta_t(j, d)}. \tag{10.32}$$

10.3. Prognostics

The prognostic measure is independent of the model and is used to compute the RUL predictions. Once the model is trained, meaning that the degradation and the observation processes are estimated, and the Viterbi algorithm has estimated the most likely sequence of hidden states that explains the observed data, the prognostic measure can be applied. It is worth noting that the Viterbi algorithm does not require the entire observation sequence to estimate the most likely sequence of hidden states, rather, the observation sequence up to time t to estimate the most likely state at time t. Thus it provides the possibility to estimate the sequence in the testing phase, where at each time t, the observations up to t are used.

In this paper, two different definitions are examined: a time-invariant prognostic measure, which is the standard measure used, and a novel time-dependent prognostic measure, which is presented for the first time.

The time-invariant expression is shown in Eq. (10.33). RUL_i^t, is the pdf of the RUL in the state i and time step t. The variables $a_{i,i}$ and $a_{i,i+1}$ correspond to the transition probabilities for remaining in the same state and transitioning into the next state, respectively. The variable $D_i(d)$ represents the pdf (or pmf for the discrete case) evaluated in the probability of transition to the same state i in the case of HMMs. For HSMMs, $D_i(d)$ is the state duration probability function.

$$RUL_i^t = a_{i,i} * (D_i(d) + RUL_{i+1}) + a_{i,i+1} * (RUL_{i+1})$$
$$i = 1, 2, \ldots, N-1 \text{ and } t = 1, 2, \ldots, T. \tag{10.33}$$

In Eq. (10.34), the time-dependent prognostic measure is presented, which is used for both HMM and HSMM. Similar to the previous case, $D_i(d)$ represents the pdf (or pmf for the discrete case) evaluated in the probability of transition to the same state i. The variable τ is the time spent in the current state i. Therefore the term $D_i(d-\tau)$ represents a shift in the pdf making this RUL expression time-dependent. The variables $d_{i,i+1}$ and $d_{i,i}$ for the HMM are derived from the transition matrix, while for the HSMM, are estimated during the learning problem, and are defined as shown in (10.35) and (10.36), respectively:

$$RUL_i^t = d_{i,i} * \left(D_i(d-\tau) + \sum_{k=i+1}^{N-1} D_k(d) + \mathcal{N}(1,\epsilon) \right)$$
$$+ d_{i,i+1} * \left(\sum_{k=i+1}^{N-1} D_k(d) + \mathcal{N}(1,\epsilon) \right), \tag{10.34}$$

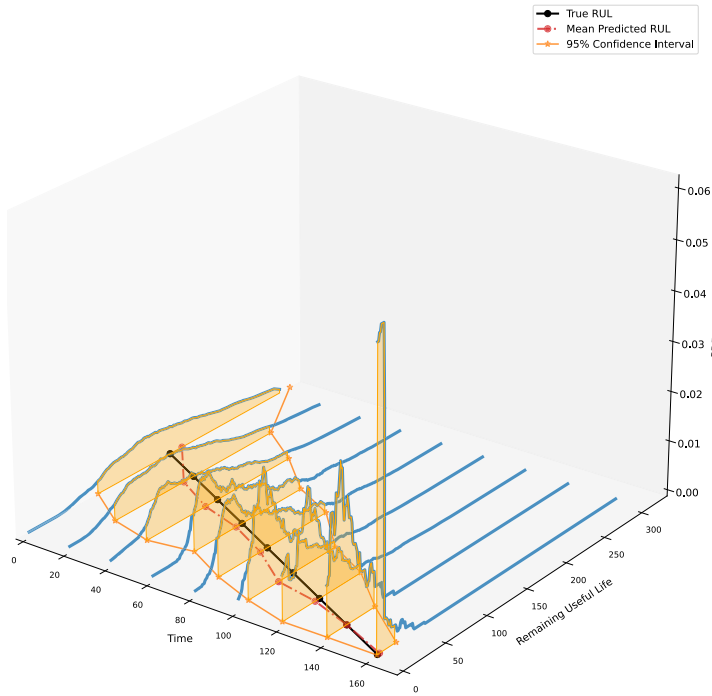


Figure 10.1 3-dimensional plot of the RUL pdfs and the operational time.

$$d_{i,i+1} = P(d \leq \tau | S_t = i), \tag{10.35}$$

$$d_{i,i} = 1 - d_{i,i+1}. \tag{10.36}$$

The result of both prognostic measures is the pdf of RUL per time step. In the following equations, the subscript i will be omitted, as specific information about the states is not relevant for error metrics and related computations. The confidence intervals can easily be obtained by calculating the cumulative density function (CDF) and, later, choosing the confidence level, usually 95%. To visualize how the prognostic measures are calculated, a 3-dimensional plot of the PDFs of the RUL every 20 time steps utilizing the estimated parameters of the HSMM is presented in Fig. 10.1.

10.3.1 Metric for evaluating RUL predictions

In this paper, RUL predictions will be evaluated in terms of accuracy and uncertainty quantification. To evaluate the accuracy, the root mean squared

error (RMSE) is used, as shown in Eq. (10.37), where $\widehat{RUL}^t$ are the RUL predictions, and RUL^t are the ground truth values.

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(RUL^t - \widehat{RUL}^t)^2}. \tag{10.37}$$

To evaluate uncertainty, the weighted spread of uncertainty (WSU) metric is introduced, where a weighted spread of the confidence intervals is calculated, as shown in Eq. (10.38). Variable t_k is the time unit; RUL_{upper} is the RUL value of the upper confidence interval, and RUL_{lower} is the value of the lower confidence interval.

$$WSU = \sum_{k=1}^{T-1}(t_{k+1} - t_k)\left[\left(\frac{RUL_{upper}^{k+1} + RUL_{upper}^{k}}{2}\right) - \left(\frac{RUL_{lower}^{k+1} + RUL_{lower}^{k}}{2}\right)\right]. \tag{10.38}$$

The expression calculates the area between the confidence intervals, while penalizing wider confidence intervals at the end of the lifetime. The penalization is considered, because the longer time that has passed, the more information is available. Therefore it is expected that the predictions hold less uncertainty as time passes.

10.4. Case study

In this part, firstly, the dataset, for which the RUL prognostics will be applied, will be presented, as well as its required preprocessing. Then, the initialization procedure of the Markov models will be shown, followed by their estimated parameters. Finally, the prognostic results are presented and discussed for both models and RUL expressions presented in Section 10.3.

10.4.1 Dataset

With the increasing popularity of prognostic health management (PHM) frameworks, both data-driven and physics-based ones, a bottleneck in their development became apparent; that is, the lack of run-to-failure datasets. This is attributed to the fact that in real-world systems, failure is avoided for safety and cost reasons. To that end, the C-MAPSS (commercial modular aero-propulsion system simulation) (see Frederick et al. (2007)) was utilized to generate synthetic datasets with a variety of operational conditions

and faults. The C-MAPSS tool is coded in the MATLAB-Simulink® environment to simulate the turbofan engine model for the 90.000 lb thrust class. The tool allows for different operating and environmental conditions and closed-loop controllers. It also allows the modification of different efficiency parameters, and thus the simulation of faults. Utilizing this tool, four datasets were created by Saxena et al. (2008). Selected faults were injected at different times for varying operation conditions, and sensor measurements were gathered from different parts of the engines. This led to the generation of diverse and multidimensional run-to-failure datasets for a fleet of engines.

For the current demonstration, the first dataset (FD001) is utilized, which considers a single operational condition (sea-level flight) and a single fault (high-pressure compressor fault). From the available 21 sensor measurements, only one is utilized, namely, sensor 11, which presents a monotonous behavior. These choices were made to avoid multidimensional and highly noisy data since, to utilize such data for PHM tasks, extensive preprocessing and feature extraction would be required, which is beside the scope of this chapter. The aim is to demonstrate the application of different HMMs utilizing the new prognostic measure. Additionally, FD001 contains 100 trajectories, which are run to failure and are typically used for the training phase of data-driven PHM frameworks and 100 trajectories, which are incomplete, and which are used for the RUL prediction in the testing phase. However, since a benchmark for comparing the presented HMM and HSMM performance is required, run-to-failure trajectories are needed to obtain the true RUL values. Thus only the first 100 trajectories (train set) are used by performing an 80/20% random split to obtain the train and test data accordingly for the current demonstration. The sensor values are discretized in 20 values based on the monotonicity criterion defined by the Mann–Kendall test (Yue and Pilon (2004)). Three of the test trajectories can be seen in Fig. 10.2.

10.4.2 Parameter estimation of Markov models

One of the most important parameters for the topology ζ of both HMM and HSMM is the number of hidden states since it affects all of the rest of the parameters of the models. In theory, the choice of the number of hidden states could have been part of the E-M process. However, since a change in the number of states also changes the rest of the topology, it's computationally inefficient to include it in the E-M. Therefore selecting the number of hidden states is typically based on some physical meaning

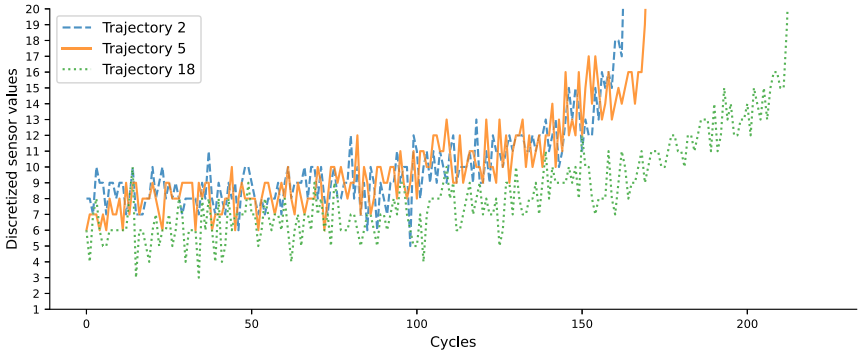


Figure 10.2 Discretized sensor measurements for 3 out of the 20 test trajectories.

related to the damage states, which can be observed in the studied system, or using information criteria, such as the BIC. In this study, to follow an objective selection procedure for their number, the Bayesian information criterion (BIC), as seen in Eq. (10.39), is utilized (see Moghaddass and Zuo (2014))

$$BIC(M_i) = \sum_{k=1}^{K} \log_e \alpha_{T^{(k)}}(i) - \omega * \frac{H_i}{2} * log_e(n). \tag{10.39}$$

In Eq. (10.30), M_i represents the candidate model. Although the expression for the BIC does not explicitly use M_i, the parameters for M_i are included via the forward variable $\alpha_{T^{(k)}}(i)$, and $T^{(k)}$ represents the lifetime of trajectory k. The variable ω is the penalty factor, which was set to 1 for this application. H_i refers to the total number of parameters in the model, and n represents the total number of observations across all trajectories, with $n = \sum_{k=1}^{K} T^{(k)}$. Models are fitted with 3 to 10 states, and the model that maximizes the BIC criterion is selected. The results can be seen in Fig. 10.3. For the HMM, the BIC metric is maximized with 10 states, while for the HSMM, it is maximized with 5 states.

For the HMM, the transition matrix A and the emission matrix B, which are the estimated parameters, are initialized with a uniform probability distribution. For the HSMM, the duration probability distribution matrix is initialized with a uniform probability distribution as well, whereas the μ vector of the emission parameters is initialized by running a K-means clustering algorithm Lloyd (1982) on the longest trajectory. Uniform probability initialization would also work in this case, but the K-means initialization provides faster convergence. Finally, the σ^2 vector is initialized with ones for all of the states.

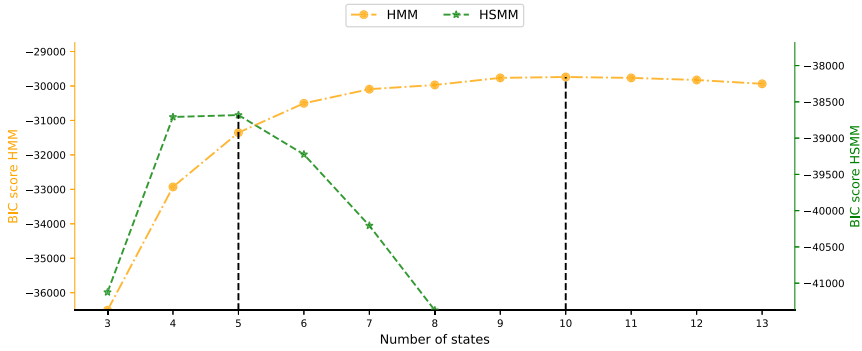


Figure 10.3 BIC criterion for HMM and HSMM.

After the models are initialized, the procedure described in Sections 10.2.1 and 10.2.2 are followed, and the parameters of the models are estimated. The estimated parameters of the HMM and the HSMM are presented in Figs. 10.4 and 10.5, respectively. In Figs. 10.4a and 10.4b, the duration distributions of the HMM are shown. They follow geometric distributions, which is inherent to the definition and assumptions of the HMM, as presented in Section 10.2.1. Contrary to that, as seen in Fig. 10.5a, the duration distributions of the HSMM have an arbitrary shape, which is to be expected since they are modeled in a nonparametric way. It can be observed that there is a high probability associated with the first state having a duration of only 1 time step. This is attributed to the fact that, based on the assumptions made for this specific HSMM, the hidden states always start from the first one. Thus due to the high variability of the initial values of the data (Fig. 10.2), the first value of the trajectory may be higher than the observation values associated with the first hidden state. Therefore, in some cases, the model needs to transit immediately after the first time step to the second state since it cannot directly start from the second hidden state. The estimated emission distributions for both the HMM (Fig. 10.4c) and the HSMM (Fig. 10.5b) have a higher expected value for every state due to the left-to-right assumption that both follow for modeling a degradation process. After the parameters are estimated, to obtain the RUL predictions, the framework explained in Section 10.3 is applied, and the findings are presented in Section 10.4.3.

10.4.3 Prognostic results and discussion

In this section, the prognostic results of the aforementioned case study will be presented. In the first subsection, the difference between the two differ-

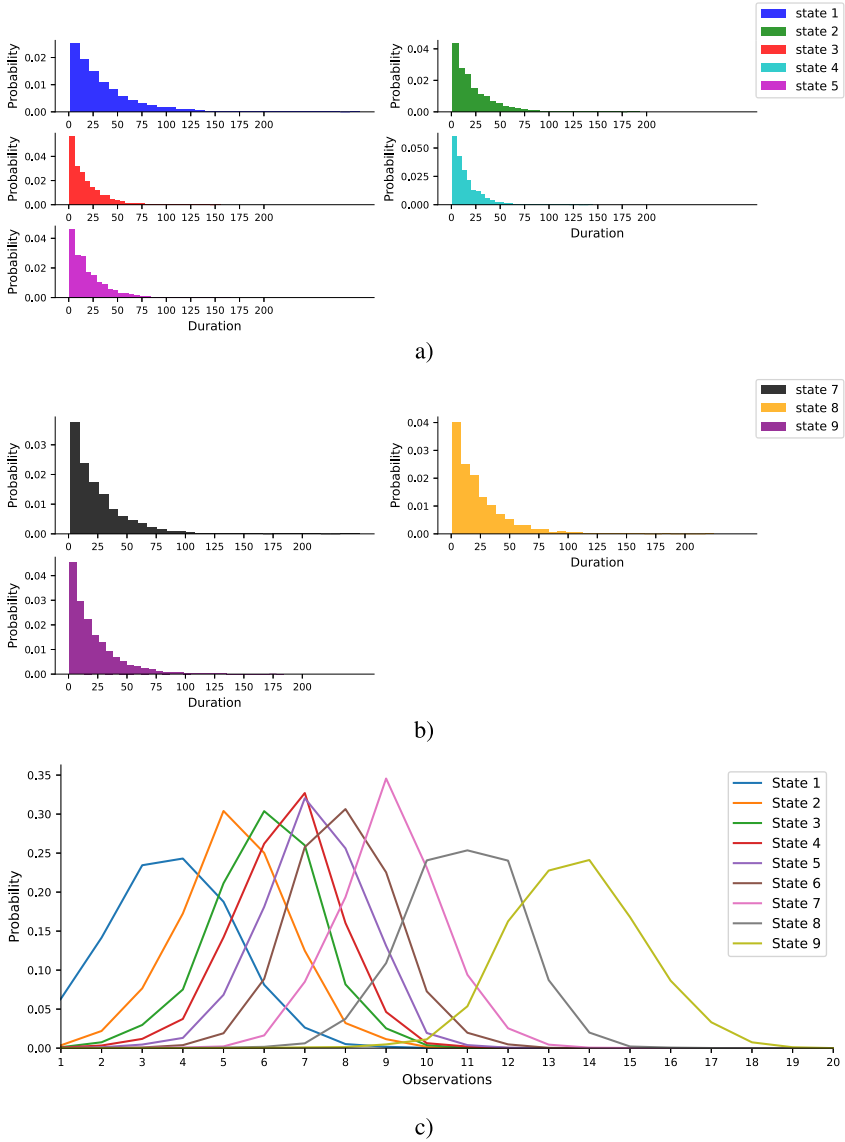


Figure 10.4 Estimated a) and b) duration probabilities and c) emission probabilities of the HMM model.

ent expressions for the RUL prediction presented in Section 10.3, namely the time-invariant (TI) and time-dependent (TD), will be demonstrated. In the next subsection, the difference in aviation prognostic performance of the introduced HMM and HSMM models will be discussed.

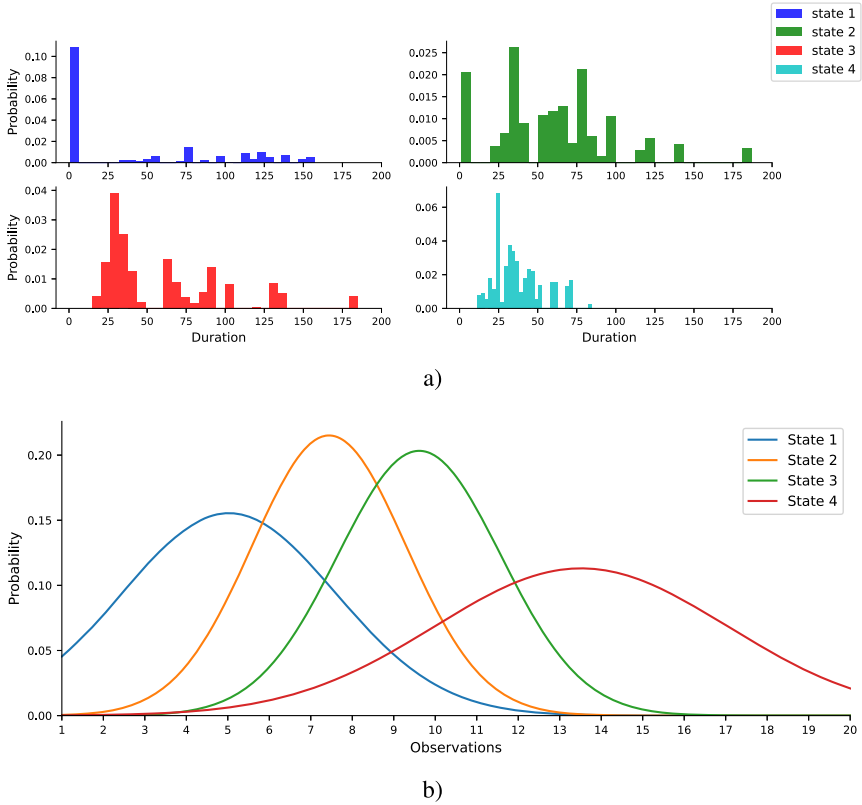
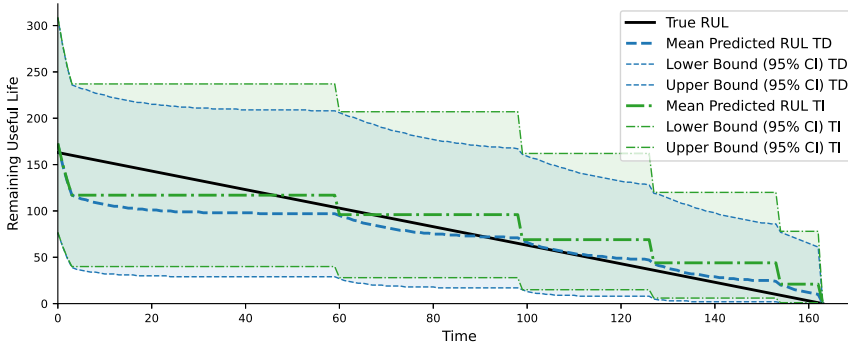


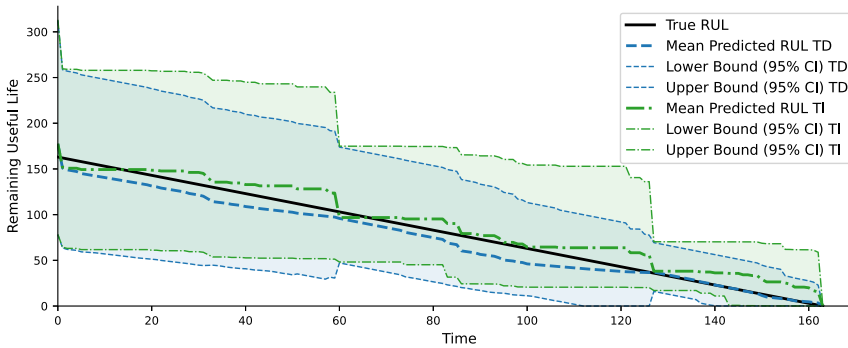
Figure 10.5 Caption: Estimated a) duration probabilities and b) emission probabilities of the HSMM model.

10.4.3.1 Time-invariant and time-dependent prognostics

When the two provided expressions of estimating the RUL (Eq. (10.33) and (10.34)) are observed, the main difference is that in (10.33), the RUL depends only on the current state as estimated by the Viterbi algorithm. Thus it can be seen in both Fig. 10.6a and 10.6b (green dash-dotted) that the RUL prediction, using the TI expression, has a step-wise behavior. This is attributed to the fact that the predicted RUL values stay roughly the same until the model transitions to the next hidden state. This behavior does not only produce higher RMSE values and wider CI bounds, as seen in Table 10.1, but it is also highly suboptimal for prognostics. This can be understood if one considers that for a long time (until the transition to the next hidden state), the prognostics provide no useful information about the remaining useful life of the asset (aircraft engine in this case) since the value of the RUL does not change. Hence the TI expression updates the



a)



b)

Figure 10.6 Comparison of TD and TI expressions for RUL prediction of the a) HMM and b) HSMM models.

prediction of RUL only when a transition occurs. As a result, the RUL prediction is conditional only to the current observation, excluding all the information from the CM dataset, which has been built up to the current time step. Our proposed TD expression effectively handles this limitation when predicting the RUL. This is also demonstrated in Fig. 10.6 (a and b) (blue dashed lines), where the RUL predictions present a decreasing trend, thereby providing more informative predictions for every time step, where new CM data arrive. Adding to that, a decrease in the RMSE values and the width of the CI bounds (quantified by the proposed WSU metric) can also be observed in Table 10.1.

10.4.3.2 HMM and HSMM for prognostics

In the previous subsection, the superiority of the proposed TD expression for RUL prediction was demonstrated. Thus this equation is chosen to

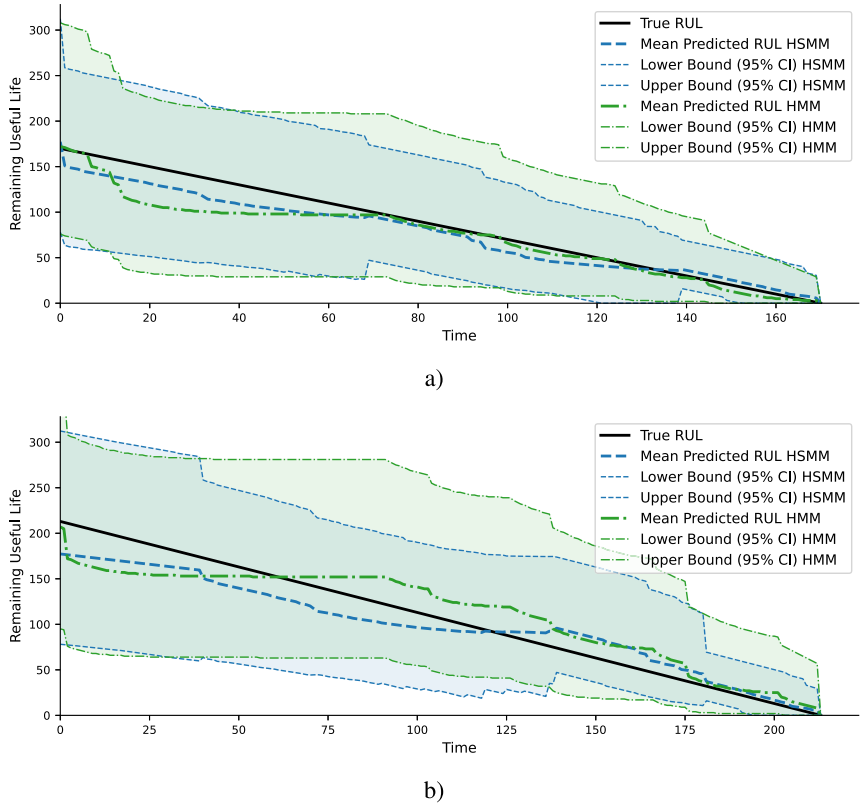


Figure 10.7 Comparison of HMM and HSMM RUL predictions for test trajectory a) 5, b) 18.

Table 10.1 Average values of the test dataset for the prognostic performance metrics considering the TI and TD expressions of RUL for the HMM and HSMM models.

RUL Expression	RMSE	WSU
	HMM	
TI	45.00	3 328 839.60
TD	43.10	2 978 334.12
	HSMM	
TI	42.63	2 621 750.39
TD	42.35	2 358 269.49

compare the prognostic performance of the presented HMM and HSMM models. The results for three of the test engines can be seen in Figs. 10.6a

and 10.6b for test trajectory 1 and in Figs. 10.7a and 10.7b for trajectories 5 and 18, respectively, whereas Table 10.1 presents the average values of the evaluation metrics (Eq. (10.37) and (10.38)) for the entire test dataset. It can be observed that the HSMM presents an overall higher accuracy, but also narrower CI bounds. Narrower CI bounds mean that there is reduced uncertainty associated with the prediction of the RUL, which makes the RUL prediction more valuable to decision-makers. The HMM assumes that the duration probability distributions follow geometric distributions, an assumption that rarely holds in the real degradation behavior of assets. The relaxation of this assumption and the nonparametric probability distribution modeling of the degradation are the main culprits of the enhanced prognostic performance. Contrary to that, due to the simplifying assumptions of the HMM, it comes with likelihood and reestimation expressions of lower complexity, and thus significantly lower computational cost. In our example, the HSMM needs approximately one order of magnitude more time to converge when compared to the HMM, although we consider the HSMM code to be under-optimized.

10.5. Conclusions

In the current chapter, an introduction to HMM and HSMM models with the appropriate modifications for prognostics was presented. Two alternative expressions are provided (the literature-proposed one, which is time-invariant, and a custom time-dependent one) for predicting the RUL of assets. For their evaluation, a custom performance metric is introduced, which calculates the width of the CI bounds of the RUL predictions, while penalizing wider bounds towards the end of the life of the asset. The framework for prognostics is applied to a publicly available dataset for turbofan engines, and the findings of the comparison between the different RUL expressions and the proposed models are discussed. The superior prognostic performance of the TD RUL expression coupled with the HSMM model was expected due to its increased complexity and time-dependent modeling of the degradation process. However, this does not come without a cost. The increased complexity of the HSMM (and thus the number of estimated parameters) is associated with increased computational time. Namely, the computational time of the HSMM can be up to an order of magnitude greater than that of the HMM (depending on the number of iterations required for convergence and the degree of optimization of the code). Based on this, the authors wish to raise awareness when it comes to applying

prognostic models (or any machine learning models for any task) so that the sustainability aspect of the applied models is also considered. Due to the ever-increasing popularity of ML models, their environmental impact can no longer be considered negligible. It is estimated that more than 500 tonnes of CO_2 is emitted for the training of high-end large-language models. To put this into perspective, this equals the amount of CO_2 emitted for the 200 round trips of a 2-and-a-half hour flight for a single passenger. An additional concern is that it is estimated that the amount of electricity used for the training is surpassed in weeks, or even days, by the electricity used by millions of everyday users. Thus we urge researchers to consider if the increased accuracy justifies the increased emissions associated with the higher computational times for their specific application.

References

Dong, M., He, D., 2007. Hidden semi-Markov model-based methodology for multi-sensor equipment health diagnosis and prognosis. European Journal of Operational Research 178, 858–878.

Eleftheroglou, N., Loutas, T., 2016. Fatigue damage diagnostics and prognostics of composites utilizing structural health monitoring data and stochastic processes. Structural Health Monitoring 15, 473–488.

Eleftheroglou, N., Mansouri, S.S., Loutas, T., Karvelis, P., Georgoulas, G., Nikolakopoulos, G., Zarouchas, D., 2019. Intelligent data-driven prognostic methodologies for the real-time remaining useful life until the end-of-discharge estimation of the lithium-polymer batteries of unmanned aerial vehicles with uncertainty quantification. Applied Energy 254, 113677.

Eleftheroglou, N., Zarouchas, D., Benedictus, R., 2020. An adaptive probabilistic data-driven methodology for prognosis of the fatigue life of composite structures. Composite Structures 245, 112386.

Eleftheroglou, N., Galanopoulos, G., Loutas, T., 2024. Similarity learning hidden semi-Markov model for adaptive prognostics of composite structures. Reliability Engineering & System Safety 243, 109808.

Frederick, D.K., DeCastro, J.A., Litt, J.S., 2007. User's guide for the commercial modular aero-propulsion system simulation (C-MAPSS). Technical Report.

Giantomassi, A., Ferracuti, F., Benini, A., Ippoliti, G., Longhi, S., Petrucci, A., 2011. Hidden Markov model for health estimation and prognosis of turbofan engines. In: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, pp. 681–689.

Le, T.T., Chatelain, F., Bérenguer, C., 2014. Hidden Markov models for diagnostics and prognostics of systems under multiple deterioration modes. In: Proceedings of the in European Safety and Reliability Conference-ESREL, pp. 1197–1204.

Liu, Q., Dong, M., Lv, W., Geng, X., Li, Y., 2015. A novel method using adaptive hidden semi-Markov model for multi-sensor monitoring equipment health prognosis. Mechanical Systems and Signal Processing 64, 217–232.

Lloyd, S., 1982. Least squares quantization in pcm. IEEE Transactions on Information Theory 28, 129–137.

Loutas, T., Eleftheroglou, N., Georgoulas, G., Loukopoulos, P., Mba, D., Bennett, I., 2019. Valve failure prognostics in reciprocating compressors utilizing temperature measurements, PCA-based data fusion, and probabilistic algorithms. IEEE Transactions on Industrial Electronics 67 (6), 5022–5029.

Moghaddass, R., Zuo, M.J., 2014. An integrated framework for online diagnostic and prognostic health monitoring using a multistate deterioration process. Reliability Engineering & System Safety 124, 92–104.

Rabiner, L.R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77, 257–286.

Saxena, A., Goebel, K., Simon, D., Eklund, N., 2008. Damage propagation modeling for aircraft engine run-to-failure simulation. In: 2008 International Conference on Prognostics and Health Management. IEEE, pp. 1–9.

Wang, N., Sun, S.d., Cai, Z.q., Zhang, S., Saygin, C., 2014. A hidden semi-Markov model with duration-dependent state transition probabilities for prognostics. Mathematical Problems in Engineering 2014, 632702.

Yu, S.Z., 2010. Hidden semi-Markov models. Artificial Intelligence 174, 215–243.

Yue, S., Pilon, P., 2004. A comparison of the power of the t test, Mann-Kendall and bootstrap tests for trend detection/une comparison de la puissance des tests t de student, de Mann-Kendall et du bootstrap pour la détection de tendance. Hydrological Sciences Journal 49, 21–37.