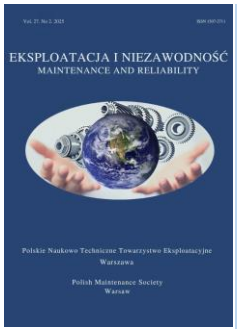

Eksploatacja i Niezawodnosc – Maintenance and Reliability

Volume 27 (2025), Issue 2

journal homepage: http://www.ein.org.pl

Article citation info:
Janczura J, Żuławiński W, Shiri H, Barszcz T, Zimroz R, Wyłomańska A, Prediction of machine state for non-Gaussian degradation model using Hidden Markov Model approach, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2025: 27(2) http://doi.org/10.17531/ein/193898

Prediction of machine state for non-Gaussian degradation model using Hidden Markov Model approach

Indexed by: Web of Science Group

Joanna Janczura[a,*], Wojciech Żuławiński[a], Hamid Shiri[b], Tomasz Barszcz[c], Radosław Zimroz[b], Agnieszka Wyłomańska[a]

[a] Faculty of Pure and Applied Mathematics, Hugo Steinhaus Center, Wroclaw University of Science and Technology, Wroclaw, Poland
[b] Faculty of Geoengineering Mining and Geology, Wroclaw University of Science and Technology, Poland
[c] Faculty of Mechanical Engineering and Robotics, AGH University, Kraków, Poland

Highlights

- On-line prediction method for forthcoming machine state is proposed.
- Degradation model with three machine states (healthy, warning and alarm) is used.
- α-stable HMM method is developed for non-Gaussian (impulsive) HI data.

Abstract

Machinery health management becomes an essential issue in many sectors. The ultimate goal is to predict machinery degradation and accordingly plan maintenance actions. However, prediction becomes much harder if data is noisy. We propose a procedure for on-line prediction of the forthcoming machine state. This procedure is dedicated to the non-Gaussian (impulsive) health index (HI) data. It is based on a simplified degradation model with three machine states, i.e. healthy, warning and alarm, described in terms of a Hidden Markov Model (HMM). Using simulated trajectories we demonstrate that the α-stable HMM dedicated to time series with impulsive behaviour outperforms the classical Gaussian approach and can be an efficient alternative in such a case. In particular, the percentage errors of the predicted alarm state transition points decrease from 20%−45% to 1%−6%, if the α-stable HMM is used instead of the Gaussian one. We illustrate the proposed methodology for two datasets acquired during experiment on the VIBstand test rig and for a benchmark FEMTO dataset.

Keywords
machine health management, prediction, hidden Markov model, α−stable distribution, machine condition monitoring

1. Introduction

In various industries mechanical equipment holds a central role and its dependable operation is a key factor that greatly influences production efficiency. Any harm or breakdown in this equipment can have broad repercussions on the entire production process, resulting in economic losses. In order to achieve optimal operating costs, maintenance actions should be performed during planned outages. Unplanned shutdowns should be avoided, as they cause the largest losses. Therefore, prognostics and health management (PHM) have gained increasing significance in the industrial environment. The PHM commonly involves four key technical processes: data acquisition, health indicator (HI) calculation, health state (HS) estimation, and remaining useful life (RUL) prediction. Initially, data, such as vibration signals, are gathered from sensors to monitor machinery health. Subsequently, HIs are formulated from these data using techniques like signal processing and

(*) Corresponding author.
E-mail addresses: J. Janczura (ORCID: 0000-0001-6110-3394) joanna.janczura@pwr.edu.pl, W. Żuławiński (ORCID: 0000-0002-8666-5868) wojciech.zulawinski@pwr.edu.pl, H. Shiri (ORCID: 0000-0002-1878-4718) hamid.shiri@pwr.edu.pl, T. Barszcz (ORCID: 0000-0002-1656-4930) tbarszcz@agh.edu.pl, R. Zimroz (ORCID: 0000-0003-4781-9972) radoslaw.zimroz@pwr.edu.pl, A. Wyłomańska (ORCID: 0000-0001-9750-1351) agnieszka.wylomanska@pwr.edu.pl

artificial intelligence, representing the machinery's health status. Following this, the entire lifespan of the machinery is segmented into distinct HSs based on the changing degradation trends of HIs. Next, within HSs displaying discernible degradation trends, RUL is predicted by analysing degradation patterns and a predefined failure threshold. It is crucial to note that the accuracy of RUL prediction is significantly influenced by the quality of the HS division. This is the case, for instance, when determining the initiation point for starting time for RUL estimation. Detecting this point prematurely may lead to the underestimation of RUL. Conversely, a delayed detection could result in a failure occurrence before any decision-making is done. Therefore, developing precise approaches for accurate on-line HS division is imperative for effective RUL prediction and decision-making in PHM.

However, analysis and segmentation of HIs are often complicated, as data from real life assets might exhibit non-Gaussian behaviour. In this paper, we consider the HI being statistics constructed based on vibration signals. Their non-Gaussian (impulsive) characteristics are mostly related to the type of machine (as well as the process carried out by the machine) and its operational regime. Examples include compressors, wind turbines, mining equipment (such as sieving screens, crushers, etc.), and engines for both automobiles and airplanes. The second reason why non-Gaussian behaviour can occur in HI data is related to the progressive degradation process of the machine. It was recognised that along with a change of the machine state (healthy-warning-alarm), the distribution of the HI time series may also change. More precisely, the analysis of real HI data shows that at the healthy state the heavy-tailed impulsive behaviour is not as apparent as in the alarm state, see e.g. [1], and this regularity is observed for many machines.

The main objective of this paper is to propose a methodology for on-line machine state prognosis that can be efficiently applied, even if the examined HI data follows non-Gaussian distribution. The second goal is to show that a Gaussian-based approach is useful, provided that the data exhibit behaviour adequate for this distribution. In the case of on-line prognosis of heavy-tailed distributed data, more advanced techniques need to be applied. The methodology presented here complements previous research devoted to the analysis and modelling of non-Gaussian long-term data, see e.g., [2] and their off-line segmentation, see e.g. [3]. In both areas (modelling and segmentation), it was proved that the application of the methods that are dedicated to non-Gaussian data is crucial for their efficiency.

In this paper, we use the hidden Markov model (HMM) approach for predicting machine state based on HI data analysis in a high-noise environment. First, an HMM is trained using data with the full history of the degradation process (i.e., healthy, warning, and alarm states) from a similar asset to the monitored one. Next, it is used to predict the probability of being in a given state at the next time point for new data (i.e., for the monitored asset) without assuming that it represents all states of the degradation process. Based on this, we can predict the forthcoming machine state. The HMM is, therefore, applied in an online mode, which has higher practical importance for condition monitoring systems than post-processing and offline segmentation. To account for possible non-Gaussianity of the data, we apply an HMM with an α-stable distribution. The prediction performance of the α-stable HMM is compared with the corresponding Gaussian HMM, based on simulated data from the adopted degradation model with a possible non-Gaussian distribution. The HMM-based methods for online prognosis are also applied to three real datasets. The first dataset, FEMTO data, is a benchmark time series frequently used for HI data analysis. The other datasets were acquired on a test rig during an experiment where the growth of the shaft unbalance was introduced. To our knowledge, the proposed methodology based on the α-stable HMM is being used for the first time for online prognosis of HI data and tested on real time series.

The rest of the paper is organised as follows. In Section 2 we present state of the art in the area of interest. Next, in Section 3 we describe the real data used in the further analysis as well as the degradation model used in simulations. In Section 4 we describe the methodology for machine state prediction based on two versions of HMM, namely the Gaussian and the α-stable ones. Further, in Section 5 we verify the proposed HMM-based approaches for simulated data from the degradation model, while in Section 6 we demonstrate the usefulness of the proposed methodology for the real HI datasets. Last section concludes the paper.

2. State of the art

One of the most common approaches used in the prediction of machine health state based on HI data utilises an adopted model. In recent years, substantial research has been conducted in this domain, broadly categorised into three classes: data-driven models [7-11], physics-based models [12], and hybrid models [13, 14]. Physics-based models attempt to represent the degradation process using mathematical equations derived from failure and damage mechanics. While this approach can yield precise results, it often necessitates a highly accurate physical model, which may be unavailable or entail significant computational costs. On the contrary, data-driven approaches aim to construct models that describe the degradation process using historical data. This category shows promise for complex applications where the derivation of physics-based models is challenging. Data-driven approaches can be further classified into two main subclasses: machine learning-based (see e.g. [15, 16]) and statistical model-based (see e.g. [17-19]) approaches. Machine learning-based methods are powerful tools for modelling, segmentation, and prediction, commonly used in complex applications. However, they require substantial data for training, which may not be accessible in many real-world scenarios. On the other hand, statistical model-based approaches do not require large amount of training data, but a careful selection of an appropriate model to describe the process remains of paramount importance. The statistical approaches offer probabilistic results, a crucial advantage in dealing with uncertainty in the degradation process. Hence, in this paper we follow this approach. It should be noted that in the literature some hybrid models aiming to harness the strengths of both of the aforementioned classes have been discussed. For further details on the hybrid models and their applications, the interested readers are referred to the following references [13, 14, 20].

For the on-line prediction of the degradation state based on the adopted model one of the most common approaches is to use the predefined fault detection thresholds. The thresholds are often supplied by manufacturers and typically delineate the transition from a healthy state (good condition) to a warning state and from a warning state to an alarm state (called also critical state). Unfortunately, in numerous instances, the limit values and desired lifetimes remain unknown, particularly for unique or customised machinery. Moreover, it is essential to note that these thresholds are typically established by manufacturing industries based on specific operating and environmental standards, which may lose validity when conditions deviate. When machinery operates in harsh environments, where the collected data show impulsive behaviour, this task becomes notably more complex. This problem was discussed e.g. in [21], where authors indicated some possible solutions in this area. It is worth to note that the problem of limit setting did not draw large attention of the research community and some authors highlight the lack of appropriate solutions and address this problem, see [12]. In this paper, we follow another approach and detect changes in the probabilistic behaviour of the HI data as an indicator for the degradation state.

The problem of prediction of machine state based on HI data can be considered as the on-line segmentation of such data. This issue was discussed in the literature. Shiri et al. [4] introduced on-line methods that worked based on the development of a robust switching Kalman filter, tailored to non-Gaussian distribution of the data. Nevertheless, this model relies on deterministic components that may be affected by non-linearity in the HI trends. Furthermore, numerous studies in the literature have attempted to address this issue (on-line detection) by applying one-class classification methods [22-24]. However, this approach is primarily effective for identifying transition points between the healthy and alarm states.

Most of the methods used for HI data modelling are based on the assumption of Gaussian (or Gaussian-like) distribution of the data, see e.g. [25-29]. However, in the literature there are also considered non-Gaussian processes useful for diagnostic features, like Gamma process [30], Generalised Cauchy process [31, 32] or even processes with heavy-tailed distributions, such as fractional Lévy stable motion, see e.g., [33, 34]. See also the positions [34-37], where authors highlight this aspect indicating that in practice, the data related to degradation of industrial equipment obey non-Gaussian distributions. There are also methods based on the extended Kalman filter dedicated for HI data modelling with non-Gaussian characteristics [38-41]. Applications of the Markov jumps systems [42, 43] or state-space models [44] are also used in that context. The alternative approaches are presented e.g., in [45]. For other reliability

models describing the non-Gaussian HI data variation see e.g., [9, 46-49].

The methodology for on-line prognosis proposed in this paper is based on HMMs, which were theoretically introduced in 1960s by Baum [50]. Since then they were used for modelling data in a variety of fields, see [51] for a comprehensive review of the possible applications of HMMs. Since HMMs are suitable for a description of a latent temporal patterns in recorded signals, they are often used for segmentation purposes in fault diagnosis problems. They were used, among others, for condition monitoring of spent nuclear fuel shearing machines [52], diagnosis of rotating components [53], detecting weld defects [54], analysis of vibration signals representing various bearing conditions [55], or diagnosis of fatigue cracks of a helicopter gearbox [56]. For other applications of HMMs in machine fault detection see e.g. [57-60]. As a natural follow-up of these modelling approaches, HMMs were also used for machine fault prediction. The authors of [61] proposed to use a combination of a Gaussian HMM and an expert based system with environmental factors for real-time failure prognosis. A method involving several HMMs for prediction of degradation state was considered by [62, 63], while [64] used log-likelihood profiles for the different trained HMMs in a drill bit failure prediction. In the context of machinery diagnostics HMMs and their extensions were used also for the remaining useful life (RUL) prediction, see e.g. [65]. A survey on the applications of HMMs in machinery fault prediction can be found in [66]. The most common setup of HMMs for failure modelling is a discrete or Gaussian distribution of the observed data.

A three-state HMM approach for non-Gaussian HI data segmentation was proposed in [3]. It was based on the α-stable distribution [6], which generalises the Gaussian distribution and accounts also for heavy-tailed cases. The α-stable distribution was first used for financial applications [67] in 1960. From that time it was applied in various areas of interest, also including condition monitoring, see e.g. [68-71]. It is important to note that the selection of the α-stable distribution is not arbitrary. Firstly, the α-stable distribution, akin to the Gaussian distribution in the finite-variance scenario, is the sole limiting distribution for infinite-variance samples. This phenomenon is described by the generalised central limit theorem (GCLT), an extension of the classical central limit theorem (CLT). According to the GCLT, α-stable distributions attract distributions of sums of random variables with diverging variances, much like the Gaussian law attracts distributions with finite variances (as per the CLT). Thus, this class of distributions can be considered a general one. It was shown in [3] that generalising the Gaussian HMM model to the α-stable one, leads to a significant improvement in the accuracy of the off-line segmentation of HI data. Such a methodology can be applied only when the whole history of the time series (i.e. data corresponding to healthy, warning and alarm states) is already available. However, it has less practical importance for on-line condition monitoring. Thus, in this paper we extend the approach proposed in [3] for the on-line operation mode and forthcoming state prognosis.

3. Description of the considered datasets

3.1. FEMTO dataset

FEMTO stands for Franche-Comté Electronics Mechanics Thermal Science and Optics-Sciences and Technologies institute. FEMTO dataset has been acquired using the PRONOSTIA platform. It was also used at the IEEE International Conference of PHM 2012 during a prognosis challenge. This dataset contains 17 various signals (acceleration and temperature) describing bearing degradation. The experiments have been carried out under stationary load/speed conditions. For the FEMTO dataset it is assumed that a bearing failure occurs when the amplitude of the vibration signal has increased above 20 g [48]. The dataset is widely used in prognostic community (see e.g. [72, 73]). Some researchers used this dataset for fault diagnostics [74, 75]. Islam et al. [76] introduce a data-driven framework for predicting the health of rolling element bearings using a degree-of-defectiveness (DD) metric in the frequency domain, employ Bayesian inference-aided one-class LSSVM for anomaly detection, and apply their proposed approach to the FEMTO dataset. In addition, this dataset is widely used as a benchmark dataset for the application of RUL estimation [77-81]. The analysed HI is constructed as the RMS of the raw time series.

3.2. AMC Tech dataset

The AMC Tech dataset is a new one and it was acquired specifically for this research. It was acquired during experiment

on the VIBstand test rig. The rig design has numerous structural resonances and thus it is very similar to real machines, which have complex designs and generate vibration signals rich in nonlinear components and noise. Many faults can be introduced, namely unbalance, misalignment, looseness as well as bearing and gear faults. For the dataset generation, unbalance was chosen as the introduced fault. Unbalance is a common fault mode for virtually every rotating machine type. It can be caused by variety of reasons, e.g. erosion of a blade, build-up of deposits, thermal effect, assembly errors. During the experiment unbalance was slowly and steadily increased by adding additional weights to the disk on the rotor. Due to a complex structure of the test rig, there was almost no increase of HIs typically associated with unbalance, e.g. peak-peak amplitude or RMS, despite significant increase in added weights. Actually, HIs which were sensitive to the unbalance were narrowband energy HIs. There were two bands analysed in the dataset, the first one has a carrier frequency of 16 Hz, tracking energy of the first harmonic of the shaft, later referred to as the "AMC Tech dataset 1". The second one, further called the "AMC Tech dataset 2′′, has a carrier frequency of 107 Hz, which was one of higher harmonics (no. 6). Note that bands were calculated with 3% bandwidth from the center frequency and the actual shaft rotational speed was 917rpm (equals 15.28 Hz). As mentioned earlier, vibrations were affected by complex kinematic of the test rig and resulted with very rich signals with numerous signal components. Another benefit was gradually increasing level of impulsive disturbances, accurately reproducing real life data. Vibration signals were acquired from MTN2200 accelerometers with a commercial condition monitoring system of AVM4000 type. The sampling frequency was 25kHz and 1s long series of raw vibration data were taken to calculate each HI. There were 2350 signals acquired during the whole experiment, thus there were 2350 data points in each analysed HI. Experiment was stopped when the unbalance caused excessive vibration of the rig and could not be safely continued. The analysed HI datasets are presented in top panels of Figures 8 and 9.

The FEMTO and AMC Tech datasets complement each other. The first is clear and it is easy to distinguish visually between health states. The latter is much more similar to a real life complex data, where distinctions are more ambiguous but still they follow the three-state model. Thus, the three-state degradation model based on the FEMTO dataset is used in a simulation study. It is worth to note that existence of three states is a common case for real-life machinery maintenance in daily basis. Certainly, the three states may differ in duration, magnitude and other parameters. Sometimes when fault initiates, the HI growth is so quick and in that case the alarm state follows healthy state without the warning (or with a very short warning state). Nevertheless, the degradation model is capable of handling such situations, so such a rare cases do not undermine the validity of the general model.

3.3. Degradation model

The degradation model used in this paper for simulating HI data is based on the one introduced in [2]. Let us mention, that the model does not correspond to any specific HI type (like RMS etc.) but it is considered as a general model applicable for various health indicators. We assume that a degradation process $S_t, t = 1, \dots, T$, is represented as

$$S_t = X_t + \eta(t), \tag{1}$$

where X_t and $\eta(t)$ are, respectively, random and trend parts (note the difference in notation for random and deterministic components). Based on the HI characteristics we assume that S_t is given by a three-state model: for $1 \le t < \tau_1$ it is in state 1 (corresponding to the healthy state), for $\tau_1 \le t < \tau_2 -$ in state 2 (warning), and for $\tau_2 \le t \le T$ - in state 3 (alarm). Parameters τ_1 and τ_2 are called change points. The states of the process determine the behaviour of both X_t and $\eta(t)$ components. However, as the proposed procedure will be further applied to detrended data, in the following description of the model we focus only on the random term.

According to the assumptions of the considered model, the random part of the process is a product of two components

$$X_t = SC(t)Z_t, \tag{2}$$

where $SC(t)$ is a state-dependent deterministic function representing a time-varying scale of the process. In state 1, we assume that the scale is constant and equal to $SC(1) = SC(\tau_1) = s_1$. In state 2, the scale grows linearly from $SC(\tau_1) = s_1$ to $SC(\tau_2) = s_2$, and in state 3 - the growth is exponential, from $SC(\tau_2) = s_2$ to $SC(T) = s_3$. In other words, we assume the following formula for the scale

$$SC(t) = \begin{cases} s_1, & 0 \le t < \tau_1, \\ s_1 + (s_2 - s_1)\dfrac{t-\tau_1}{\tau_2-\tau_1}, & \tau_1 \le t < \tau_2, \\ s_2 \ exp\left(\dfrac{ln\, s_3 - ln\, s_2}{T-\tau_2}(t-\tau_2)\right), & \tau_2 \le t \le T. \end{cases}$$

In this paper we assume that the Z_t component is an autoregressive model (AR (p)) with different coefficients in each state. This model of Z_t was identified in [2] where different HI data were analysed. We recall, the AR (p) model is defined as

$$Z_t = \sum_{j=1}^{p} \phi_j^{(i)} Z_{t-j} + \xi_t, \tag{3}$$

where $\phi_1^{(i)}, \dots, \phi_p^{(i)}$ are the AR model coefficients in the i-th state. Let us note that in particular states, the actual AR model orders might be different. In our case we assume that the innovations ξ_t are independent identically distributed (i.i.d.) random variables with zero mean, here assumed to be symmetric α-stable distributed with stability parameter α_i in the i-th state ($0 < \alpha_i \le 2$). Recall that the symmetric α-stable distribution can be defined through the characteristic function $\Phi(t) = \exp(-c^\alpha |t|^\alpha)$ and for $\alpha = 2$ it reduces to the Gaussian distribution with the scale parameter $\sigma = \sqrt{2}c$. It is the only case when the variance is finite. Thus, in the particular case of $\alpha = 2$, the innovations ξ_t in the model (3) are Gaussian distributed. For other α values the α-stable distribution belongs to the so-called heavy-tailed class of distributions for which large observations are more probable than in the Gaussian case, see [82] for more details. Here, we assume $c = \frac{1}{\sqrt{2}}$, what corresponds to the standard normal distribution $\mathcal{N}(0,1)$ in the Gaussian case.

The procedure used for the estimation of the degradation model parameters and the algorithm for simulation of the degradation model trajectories are described in the Appendix A. Sample trajectories of the fitted degradation model are plotted in Figure 2.

4. Methodology for machine state prediction

For machine state prediction we use two HMM methods, coming from a simplification of the degradation model [3]. In this section, first, we describe two proposed HMMs and, then, the methodology used for predicting a machine state. Note that the online prediction methodology complements the approach from [3], which is dedicated to the offline segmentation. The procedure utilising HMM methodology proposed in this paper consists of two crucial steps: training and prediction. In Sections 4.2 and 4.3 we describe them in details. Let us emphasise that the methodology is dedicated to detrended data (i.e. to data without a deterministic trend $\eta(t)$, see Eq. (1)). Thus, in the preliminary step of the analysis the trend is estimated using the moving median technique (see [2] for details) and subtracted from HI data. Sample trajectories of simulated time series S_t, used for the trend estimation, are plotted in Figure 2 (middle panel for the Gaussian case and bottom panel for the α-stable one). A similar approach in the context of the HMM-based segmentation was used in [3], where the authors compared three HMM approaches: designed for original HI data, for detrended HI data and for detrended and normalized HI data. The most accurate results were obtained for detrended data, what indicates that the HMM approach is more suitable for detecting degradation changes in the probabilistic part of the signal. Therefore, here, we only use such an approach. Note that the detrended HI data corresponds to the random component X_t of the degradation model (1).

4.1. Description of HMMs

We assume that the detrended HI data is given by a three-state HMM with states corresponding to the healthy, warning and alarm states of the machine. Let $\{R_t\}_{0 \le t \le T}$ denote the Markov chain, $R_t: \Omega \to \{1,2,3\}$. In this paper, we consider two HMMs, namely with the Gaussian and the α-stable distributions. In the Gaussian HMM method we assume that the detrended HI data is given by

$$X_t \sim \begin{cases} \mathcal{N}(0,\sigma_1), & \text{if } R_t = 1, \\ \mathcal{N}(0,\sigma_2), & \text{if } R_t = 2, \\ \mathcal{N}(0,\sigma_3), & \text{if } R_t = 3, \end{cases} \tag{4}$$

while in the α-stable HMM method we assume

$$X_t \sim \begin{cases} \mathcal{S}(2,0,c_1,0), & \text{if } R_t = 1, \\ \mathcal{S}(\alpha_2,0,c_2,0), & \text{if } R_t = 2, \\ \mathcal{S}(\alpha_3,0,c_3,0), & \text{if } R_t = 3, \end{cases} \tag{5}$$

where $\mathcal{N}(0,\sigma_i)$ is the zero-mean Gaussian distribution with the variance σ_i^2 and $\mathcal{S}(\alpha_i,0,c_i,0)$ is the symmetric α-stable distribution [82], with the characteristic function $\Phi(t) = \exp(-c_i^{\alpha_i}|t|^{\alpha_i})$. Parameters σ_i, c_i, describe the scale, while α_i -

the heaviness of the distribution tails. Note that the location parameter in both models is set to 0, since only detrended HI data is further analysed.

The state process R_t is assumed to be an unobserved Markov chain with the following transition matrix

$$\mathbf{P} = \begin{pmatrix} p_{11} & 1-p_{11} & 0 \\ 0 & p_{22} & 1-p_{22} \\ 0 & 0 & 1 \end{pmatrix}, \quad (6)$$

where $p_{ij} = P(R_{t+1} = j \mid R_t = i)$ is the probability of switching from state i at time t to state j at time $t+1$. Note that the form of the transition matrix is driven by the assumption that the states come successively, i.e. the change from the warning (corresponding to $R_t = 2$) to the normal (corresponding to $R_t = 1$) state as well as from the alarm (corresponding to $R_t = 3$) to the warning or normal state is not possible. Finally, although the moments of switching are not directly observable, we assume that the first observation x_1 in the data comes from the healthy state, i.e. $P(R_1 = 1) = 1$.

Let us note that the proposed HMMs given in Eqs. (4) and (5) can be considered as a simplification of the degradation model described in Section 3.3. In both HMMs we assume the changing scale parameter, similar as in the assumed degradation model. However, in contrast to the degradation model, within a given state the scale parameter is considered as a constant value in both HMM methods. Moreover, the interdependence of the data is not taken into account in contrast to the assumed degradation model, where AR component is included.

4.2. Training

Predicting the future state of HI data requires a knowledge about the process characteristics in each of the three states (healthy, warning and alarm). Thus, at this step we follow a common approach (see e.g. [62, 64, 83]), and assume that the long-term data with full history is available and the training of the HMMs is performed on such time series (after removing the deterministic trend). To this end we use the Expectation-Maximization (EM) algorithm of [84]. It yields estimators of the HMM model parameters $\theta = (\sigma_1, \sigma_2, \sigma_3, \mathbf{P})$ (or $\theta = (\alpha_2, \alpha_3, c_1, c_2, c_3, \mathbf{P})$ in the α-stable HMM) and at the same time also probabilities of the state process values $P(R_t = i \mid x_1, x_2, \dots, x_T)$, based on the observed HI data $(x_1, x_2, \dots, x_T)$. The EM algorithm is an iterative two-step procedure. It starts with an arbitrarily chosen vector of initial parameters $\theta^{(0)}$. Next, the state probabilities $P(R_t = i \mid x_1, \dots, x_T; \theta^{(0)})$ are derived based on the Bayes rule and the initial parameters $\theta^{(0)}$. These probabilities are then used for deriving new maximum likelihood (ML) estimates $\theta^{(1)}$ of the parameter vector. Next, the state probabilities $P(R_t = i \mid x_1, \dots, x_T; \theta^{(1)})$ are updated using new ML estimates $\theta^{(1)}$ and they again are used for deriving new ML estimates of θ. Both steps: i) updating probabilities (called the E-step), and ii) updating ML estimates (called the M-step) are repeated until the (local) maximum of the likelihood function is reached. A detailed description of these two steps of the algorithm is given below, while a schematic diagram of the training procedure is plotted in Figure B.13.

The E-step

Assume that $\theta^{(n)}$ is the parameter vector calculated in the M-step during the previous iteration and $P(R_1 = 1 \mid x_0; \theta^{(n)}) = 1$, according to the model assumption that $P(R_1 = 1) = 1$. Let $\mathbf{x}_t = (x_1, x_2, \dots, x_t)$ be the observed data up to time t. Then [85]:

1. Forward filtering: For $t = 2,3, \dots, T$ iterate on equations:

$$P(R_t = i \mid \mathbf{x}_t; \theta^{(n)}) = \frac{P(R_t=i|\mathbf{x}_{t-1};\theta^{(n)}) f(x_t|R_t=i;\mathbf{x}_{t-1};\theta^{(n)})}{\sum_{j=1}^{3} P(R_t=j|\mathbf{x}_{t-1};\theta^{(n)}) f(x_t|R_t=j;\mathbf{x}_{t-1};\theta^{(n)})}, \quad i = 1,2,3,$$

where $f(x_t \mid R_t = i; \mathbf{x}_{t-1}; \theta^{(n)})$ is the density of the underlying data at time t conditional that the process was in regime i ($i \in \{1,2,3\}$) with parameters equal to $\theta^{(n)}$,

and $P(R_{t+1} = i \mid \mathbf{x}_t; \theta^{(n)}) = \sum_{j=1}^{3} p_{ji}^{(n)} P(R_t = j \mid \mathbf{x}_t; \theta^{(n)})$,

until $P(R_T = i \mid \mathbf{x}_T; \theta^{(n)})$ is calculated. Note that the first formula comes from the Bayes rule, while the second one comes directly from the definition of the transition matrix.

2. Backward smoothing: for $t = T-1, T-2, \dots, 2$ iterate on

$$P(R_t = i \mid \mathbf{x}_T; \theta^{(n)}) = \sum_{j=1}^{3} \frac{P(R_{t+1}=j|\mathbf{x}_T;\theta^{(n)})}{P(R_{t+1}=j|\mathbf{x}_t;\theta^{(n)})} p_{ij}^{(n)} P(R_t = i \mid \mathbf{x}_t; \theta^{(n)}).$$

Note that the smoothing formula is calculated based on $P(R_t = i \mid \mathbf{x}_T; \theta^{(n)}) = \sum_{j=1}^{3} P(R_{t+1} = j, R_t = i \mid \mathbf{x}_T; \theta^{(n)})$ and

$$P\left(R_{t+1}=j,R_t=i\mid \mathbf{x}_T;\theta^{(n)}\right)=P\left(R_{t+1}=j\mid \mathbf{x}_T;\theta^{(n)}\right)P\left(R_t=i\mid R_{t+1}=j,\mathbf{x}_T;\theta^{(n)}\right)$$
$$=P\left(R_{t+1}=j\mid \mathbf{x}_T;\theta^{(n)}\right)P\left(R_t=i\mid R_{t+1}=j,\mathbf{x}_t;\theta^{(n)}\right)$$
$$=\frac{P\left(R_{t+1}=j\mid \mathbf{x}_T;\theta^{(n)}\right)P\left(R_t=i,R_{t+1}=j|\mathbf{x}_t;\theta^{(n)}\right)}{P(R_{t+1}=j\mid \mathbf{x}_t;\theta^{(n)})},$$

where the second equality comes from the fact that for a known R_{t+1} value the observations $x_{t+1},x_{t+2},\dots,x_T$ do not contain more information on R_t than already included in $R_{t+1}$and $\boldsymbol{x_t}$ (see also [85] for a more detailed derivation). Finally, it can be rewritten as

$$P\left(R_{t+1}=j,R_t=i\mid \mathbf{x}_T;\theta^{(n)}\right)=\frac{P\left(R_{t+1}=j\mid \mathbf{x}_T;\theta^{(n)}\right)P\left(R_t=i|\mathbf{x}_t;\theta^{(n)}\right)p_{ij}^{(n)}}{P(R_{t+1}=j\mid \mathbf{x}_t;\theta^{(n)})}. \quad (7)$$

Under the model assumptions (4)-(5), $f\left(x_t\mid R_t=i;\mathbf{x}_{t-1};\theta^{(n)}\right)$ is the α-stable or Gaussian density with parameters α_i,c_i or σ_i, respectively. Since, in general, the α-stable distribution does not have a closed-form representation of the probability density function (pdf), for regimes with unknown α parameter, i.e. $R_t=2$ and $R_t=3$, it is calculated numerically, using the method of Nolan, [86]. In the other cases, the Gaussian pdf is used.

The M-step

The probabilities derived in the E-step $P\left(R_t=i\mid \mathbf{x}_T;\theta^{(n)}\right)$, are then used for the calculation of the new ML estimates $\theta^{(n+1)}$. The parameters of the distributions α_i and c_i (or σ_i in the Gaussian case) are estimated as the maximizers of the expected (due to the hidden states) log-likelihood function, $\theta^{(n+1)}=argmax_\theta L(\theta;\boldsymbol{x_T})$, given by the formula

$$L(\theta;\mathbf{x}_T)=\sum_{t=2}^{T}\sum_{i=1}^{3}P\left(R_t=i\mid \mathbf{x}_T;\theta^{(n)}\right)\ln[f(x_t\mid R_t=i;\theta)], \quad (8)$$

where the inner sum comes from the law of total expectation (see also [85] for more details). The new transition probabilities, $p_{ij}^{(n+1)}$, are estimated as:

$$p_{ij}^{(n+1)}=\frac{\sum_{t=1}^{T-1}P\left(R_{t+1}=j,R_t=i\mid \mathbf{x}_T;\theta^{(n)}\right)}{\sum_{t=1}^{T-1}P(R_t=i\mid \mathbf{x}_T;\theta^{(n)})} \quad (9)$$

and the joint probability $P\left(R_{t+1}=j,R_t=i\mid \mathbf{x}_T;\theta^{(n)}\right)$ can be calculated as in formula (7).

Estimators obtained in the M-step are then used as a new parameter vector $\theta^{(n+1)}$ in the next iteration of the E-step. The algorithm stops if $\left|\theta^{(n+1)}-\theta^{(n)}\right|<\delta$ for some small δ and $\hat{\theta}=\theta^{(n+1)}$ is set as the vector of trained HMM parameters.

4.3. Prediction

The prediction step is performed for new data. Let T' be its length. We assume that the new data exhibit similar characteristics as the trajectory used at the training step. Having the trained HMM parameters $\hat{\theta}$ we can calculate the conditional probabilities about the state process for a new trajectory. Precisely, for $t=2,\dots,T'$ we iterate on equations

$$P\left(R_t=i\mid \mathbf{x}_t;\hat{\theta}\right)=$$
$$\frac{P\left(R_t=i\mid \mathbf{x_{t-1}};\hat{\theta}\right)f\left(x_t\mid R_t=i;\mathbf{x_{t-1}};\hat{\theta}\right)}{\sum_{i=1}^{3}P\left(R_t=i\mid \mathbf{x_{t-1}};\hat{\theta}\right)f\left(x_t\mid R_t=i;\mathbf{x_{t-1}};\hat{\theta}\right)} \quad (10)$$

and

$$P\left(R_{t+1}=i\mid \mathbf{x}_t;\hat{\theta}\right)=\sum_{j=1}^{3}\hat{p}_{ji}P\left(R_t=j\mid \mathbf{x}_t;\hat{\theta}\right), \quad (11)$$

yielding probabilities of states for each of the new observations under the trained HMM parameters. The starting point for the iteration is chosen as $P\left(R_1=1\mid \mathbf{x_0};\hat{\theta}\right)=1$, according to the assumption that $P(R_1=1)=1$. Finally, the trained HMM transition probabilities, $\hat{p}_{ij}$, are updated based on the probabilities of states in the new data analogously to the calculations in the training step (8)

$$p'_{ij}=\frac{\sum_{t=1}^{T'-1}\frac{P\left(R_{t+1}=j\mid \mathbf{x}_{T'};\hat{\theta}\right)}{P\left(R_{t+1}=j\mid \mathbf{x}_t;\hat{\theta}\right)}P\left(R_t=i\mid \mathbf{x}_t;\hat{\theta}\right)\hat{p}_{ij}}{\sum_{t=1}^{T'-1}P\left(R_t=i\mid \mathbf{x}_{T'};\hat{\theta}\right)}. \quad (12)$$

Such update of the transition matrix allows for adjusting the trained model to different proportions of the states' duration in the new data. The inferred state probabilities of the last observation, $P\left(R_{T'}=i\mid \mathbf{x}_{T'};\hat{\theta}\right)$, $i\in\{1,2,3\}$, and the updated transition probabilities p'_{ij} are then used for prediction of the state in the next time point

$$\hat{P}\left(R_{T'+1}=j \mid \mathbf{x}_{T'};\hat{\theta}\right)=\sum_{i=1}^{3} P\left(R_{T'}=i \mid \mathbf{x}_{T'};\hat{\theta}\right)p'_{ij}. \quad (13)$$

Finally, we can use the probabilities (13) to predict the next state. Namely, the predicted state is calculated as the most probable state in the next time point, i.e. the healthy state prediction corresponds to $\hat{P}\left(R_{T'+1}=1 \mid \mathbf{x}_{T'};\hat{\theta}\right)>0.5$, the warning state prediction to $\hat{P}\left(R_{T'+1}=2 \mid \mathbf{x}_{T'};\hat{\theta}\right)>0.5$, while the alarm state prediction to $\hat{P}\left(R_{T'+1}=3 \mid \mathbf{x}_{T'};\hat{\theta}\right)>0.5$. Note that in order to simplify the notation in the rest of the paper, we further omit conditioning on $\mathbf{x}_{T'}$ and $\hat{\theta}$ in the notation for the predicted probabilities, i.e. we use $\hat{P}(R_{T'+1}=i)$. A schematic illustration of the whole procedure is given in Figure 1.

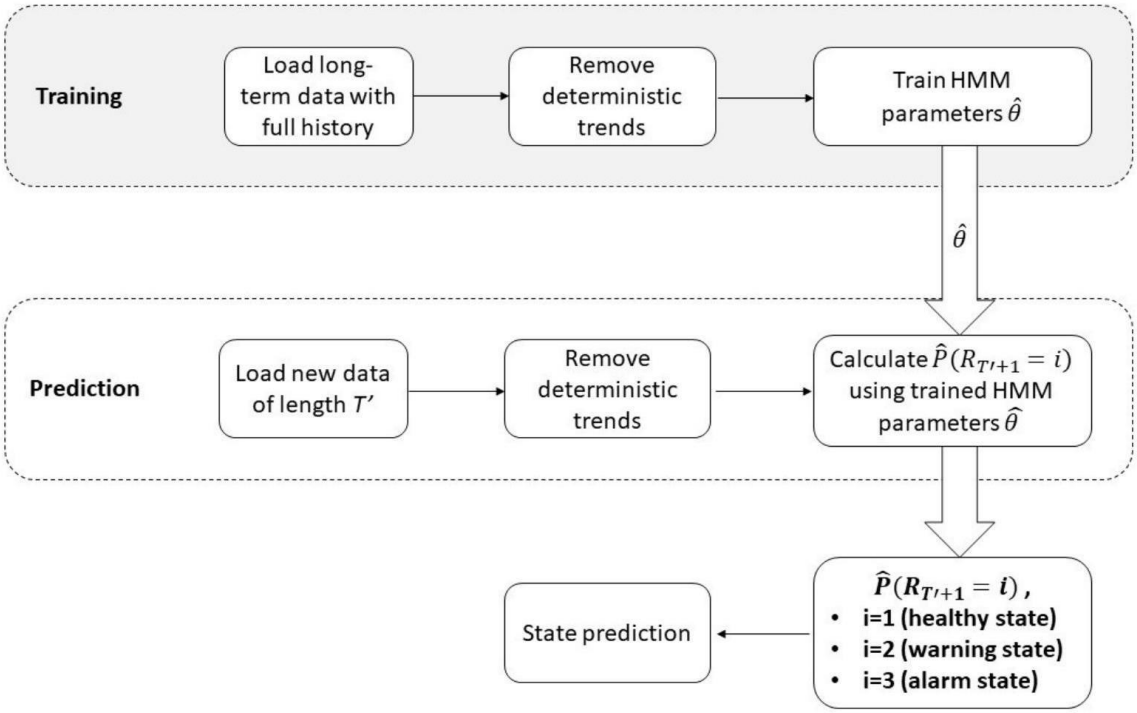

Figure 1. Illustration of the state prediction procedure for the HMM method (see Sections 4.2 and 4.3).

5. Comparison of the predictive performance of Gaussian and non-Gaussian models for simulated datasets

In this section we demonstrate the efficiency of the proposed procedure for simulated data. Here in the training and prediction steps we use the simulated trajectories from the degradation model described in Section 3.3.

5.1. Simulated data description

The parameters used for the simulation are obtained based on the real time series from the FEMTO dataset (see Section 3.1 and Appendix A for more details). The considered data is plotted in Figure 2 (top panel) together with the true change points τ_1 and τ_2 $(\tau_1=1176, \tau_2=2649)$ expressed in the number of observations. The other estimated parameters of the degradation model are (see the notation used in Section 3.3): $s_1=0.0018$, $s_2=0.0235$, $s_3=0.0816; \phi_1^{(1)}=0.1177, \phi_2^{(1)}=0.0427, \phi_3^{(1)}=0.0328$, $\phi_4^{(1)}=0; \phi_1^{(2)}=0.1654, \phi_2^{(2)}=0.0278, \phi_3^{(2)}=0.0054, \phi_4^{(2)}=0.0100$, and $\phi_1^{(3)}=0.0666$, $\phi_2^{(3)}=0.0871, \phi_3^{(3)}=0.3038, \phi_4^{(3)}=0$. They were obtained by applying the procedure described in Appendix A.1 to the full FEMTO data. In the middle panel of Figure 2 we present an exemplary trajectory of the degradation model with the parameters corresponding to the FEMTO data and the Gaussian distribution of the innovations of the AR model (i.e. ξ_t in Eq. (3)), while in the bottom panel we demonstrate an exemplary trajectory of the degradation model with the α-stable distribution of ξ_t with $\alpha=1.65$. The trajectories were simulated according to the algorithm described in Appendix A.2. The length of the trajectories were set to $T=2704$ observations, as in the FEMTO dataset. One can clearly see that the overall pattern of both simulated trajectories corresponds to the FEMTO data (increasing trend, time-dependent scale). However, in the Gaussian case (middle panel of Figure 2) we do not observe any impulsive behaviour of the time series while in

the α-stable case large observations are clearly visible.

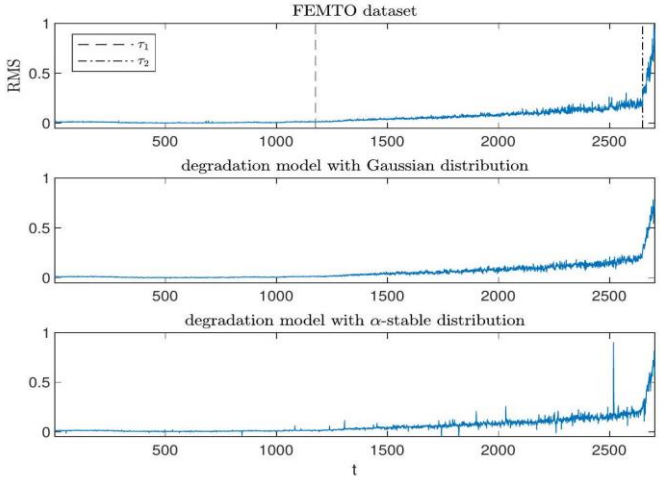

Figure 2. FEMTO dataset (top panel) and the simulated trajectories of the degradation model with the Gaussian (middle panel) or the α-stable distribution (bottom panel). The parameter α was set to 1.65. The true change points τ_1, τ_2 are marked with black, vertical lines in the top panel.

5.2. State prediction with HMM methods

5.2.1. Gaussian HMM

We start with training the HMM using the Gaussian HMM method (4). To this end, we simulate a full HI trajectory according to the procedure described in Appendix A.2 with a Gaussian distribution and remove its deterministic trend using a moving median technique [2]. Next, this trajectory is used in the training step of the procedure (see Figure 1), in which we apply the EM algorithm.

As a result we derive the trained HMM parameters: $\sigma_1 = 0.0020, \sigma_2 = 0.0148, \sigma_3 = 0.1006, p_{11} = 0.9993$ and $p_{22} = 0.9992$. EM algorithm also yields the probabilities of the possible state process values. These probabilities are then used for the change points estimation. Precisely, we assign each observation to the most probable state, i.e. $R_t = i$, if $P\left(R_t = i \mid \mathbf{x}_T; \hat{\theta}\right) > 0.5$. The change points are then simply $\tau_1 = \min\{t: R_t = 2\}$ for the first one and $\tau_2 = \min\{t: R_t = 3\}$ for the second one. The trained change points are equal to $\tau_1 = 1398$ and $\tau_2 = 2633$.

After training the HMM parameters, we simulate next 100 trajectories of the model (1) with the Gaussian distribution and random change points. These points are simulated according to the uniform distribution on the intervals $[\tau_1 - 0.1(\tau_2 - \tau_1), \tau_1 + 0.1(\tau_2 - \tau_1)]$ and $[\tau_2 - 0.1(T - \tau_2), \tau_2 + 0.1(T - \tau_2)]$ for the first and the second change point, respectively. In other words, the possible values are simply chosen from an interval containing the points detected in the training trajectory $\pm 10\%$ of the length of the corresponding state (see also Figure 3 for illustration).

Next, for the simulated trajectories we apply the prediction step of the procedure (see Figure 1). Precisely, for each of the time points $T' = 500,501, \ldots, T-1$ we subtract the moving median and calculate the predicted next state probability $\hat{P}(R_{T'+1} = i)$ based on the data up to this point, i.e. from the interval $[1, T']$. The calculations are done according to formula (13) with the trained HMM parameters. The average of the obtained values for the warning and alarm state, $i = 2,3$, from all simulated trajectories is plotted in Figure 3. The range of the simulated, random change points is also marked in the figure with coloured areas. The mean of the predicted probabilities changes visibly within the range of the simulated change points for the corresponding states, which means that the change of states is predicted close to the simulated change points. At the same time, it is on expected levels (i.e. is close to 0 or 1) outside the change periods.

As mentioned before, based on the predicted state probabilities we also calculate predictions of the change points as the first time point for which the probability of a given state is higher than 0.5, i.e.,

$$\hat{\tau}_1 = \min\left\{T': \hat{P}(R_{T'+1} = 2) > 0.5\right\} + 1$$
$$\text{and}$$
$$\hat{\tau}_2 = \min\left\{T': \hat{P}(R_{T'+1} = 3) > 0.5\right\} + 1. \qquad (14)$$

The obtained values are in the range of $[821,1514]$ or $[2222,2649]$, for the first and the second change point, respectively. The means of the predicted change points $\hat{\tau}_1$ and $\hat{\tau}_2$ are equal to 1251 and 2485. The state prediction shows on average earlier moments of change points than the simulated ones, but the differences are not high.

Gaussian HMM method applied to the simulated trajectories with the Gaussian distribution yields reasonable state predictions. Now, we verify the performance of the Gaussian HMM method, when applied to heavy-tailed trajectories. Again, we simulate 100 trajectories of the model (1), but now with an α-stable distribution. The parameter α is set to 1.65 and, for simplicity, we assume that it is the same for all states. Analogously as for the Gaussian simulations, the change points

are again drawn from a uniform distribution on the intervals $[\tau_1 - 0.1(\tau_2 - \tau_1), \tau_1 + 0.1(\tau_2 - \tau_1)]$ and $[\tau_2 - 0.1(T - \tau_2), \tau_2 + 0.1(T - \tau_2)]$ with $\tau_1 = 1398$ and $\tau_2 = 2633$. Next, we remove the trend using a moving median and apply the Gaussian HMM method with the previously trained Gaussian HMM parameters, i.e. $\sigma_1 = 0.0020, \sigma_2 = 0.0148, \sigma_3 = 0.1006$, $p_{11} = 0.9993$ and $p_{22} = 0.9992$, to predict the next state. We repeat the calculations for each of the time points $T' = 500,501, \ldots, T-1$ using only the values from the interval $[1, T']$.

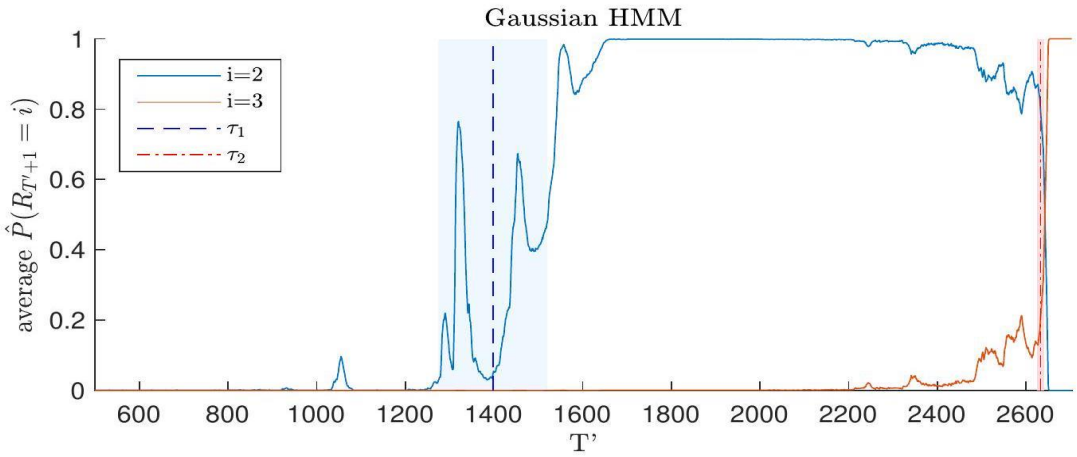

Figure 3. Mean of the predicted next state probabilities for the Gaussian HMM method applied to the simulated Gaussian trajectories of the degradation model (1). The range of the simulated change points is marked with the light blue, or light red area for the first and second change, respectively. The trained change points τ_1 and τ_2 are marked with dashed, vertical lines. The mean was calculated from 100 simulations.

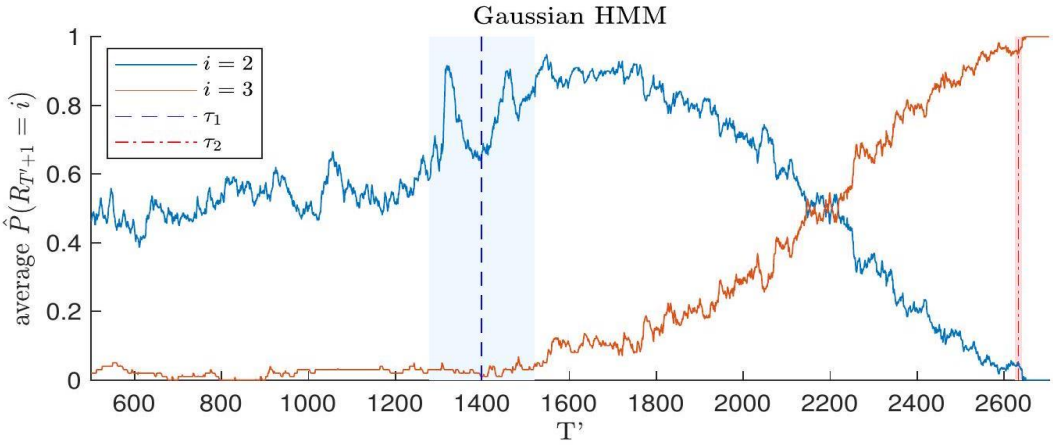

Figure 4. Mean of the predicted next state probabilities for the Gaussian HMM method applied to the simulated α-stable trajectories of the degradation model (1). The range of the simulated change points is marked with the light blue, or light red area for the first and second change, respectively. The trained change points τ_1 and τ_2 are marked with dashed, vertical lines. The mean was calculated from 100 simulations.

The obtained predicted probabilities are plotted in Figure 4. The predicted change points $\hat{\tau}_1$ and $\hat{\tau}_2$ are in the range [501,875] and [501,2212], while their means are equal to 531 and 1443, for the first and the second change point, respectively. As can be observed, the Gaussian HMM method does not work well for the α-stable distributed data. The predicted points are much lower than the expected ones and the spread of the results is much wider than the simulated one. These are the effects of extreme observations, coming from heavy tails, that are being wrongly classified to the next state by the Gaussian HMM.

5.2.2. α-stable HMM

The results obtained in Section 5.2.1 show that the Gaussian HMM method does not work well, if it is applied to data with heavy-tailed distribution (here α-stable). Hence, in this section we consider the α-stable HMM method (5). Again, we start with training the HMM. To this end, we simulate a trajectory of HI data according to the procedure described in Appendix A.2 (see

also Section 5.1 for more details on the parameters). However, now we use the α-stable distribution for the random part ξ_t. We consider different cases of the α parameter values, namely $\alpha = \{1.65, 1.7, \cdots, 2\}$ and, for simplicity, we assume that they do not change with the states. Next, we remove the trend using a moving median technique and separately for each considered α value train the α-stable HMM using the EM algorithm (training step of the procedure illustrated in Figure 1). As a result, in each case, we obtain trained HMM parameters as well as the change points for the training trajectories.

After training the α-stable HMM, we simulate next 100 trajectories of the model (1) with the α-stable distribution and random change points. Again, we consider separately different values of α ($\alpha = \{1.65, 1.7, \cdots, 2\}$) and the change points are drawn from the uniform distribution on the interval $[\tau_1 - 0.1(\tau_2 - \tau_1), \tau_1 + 0.1(\tau_2 - \tau_1)]$ and $[\tau_2 - 0.1(T - \tau_2), \tau_2 + 0.1(T - \tau_2)]$, where τ_1 and τ_2 are the points detected in the training trajectory with the corresponding α. Next, we perform the prediction step of the proposed procedure for the simulated trajectories and each of the time points $T' = 500, 501, \ldots, T - 1$.

In order to evaluate the obtained predictions for different values of α, we calculate the mean absolute percentage errors of the predicted change points

$$\mathrm{MAPE}(\tau_i) = \frac{1}{100} \sum_{n=1}^{100} \frac{|\tau_i(n) - \hat{\tau}_i(n)|}{\tau_i(n)}, \qquad (15)$$

where $\tau_i(n)$ is the simulated change point for n-th trajectory, while $\hat{\tau}_i(n)$ is the corresponding prediction calculated as in (14). Results obtained for different values of α are plotted in Figure 5. In the case of the trained α-stable HMM MAPE (τ_1) is between $11\% - 34\%$ and decreases with increasing parameter α, i.e. the predictions are more accurate for less heavy tails. MAPE (τ_2) is between $1\% - 6\%$ and its level is similar for different α values.

In order to compare both HMM methods, i.e. the Gaussian as well as the α-stable one, we also calculate MAPE for the previously trained Gaussian HMM (see Section 5.2.1) applied to the simulated trajectories with the α-stable distribution. The obtained MAPE values are also plotted in Figure 5. Note that the results obtained for the Gaussian HMM applied to the Gaussian simulations correspond to the case $\alpha = 2$. The errors of the predictions obtained with the Gaussian HMM applied to trajectories with heavy tails are from the range $35\% - 62\%$ and $20\% - 45\%$ for τ_1 and τ_2, respectively. The obtained values are much higher than those calculated using the α-stable HMM method. The errors decrease with increasing α, i.e. with the distribution used for simulations being closer to the trained model.

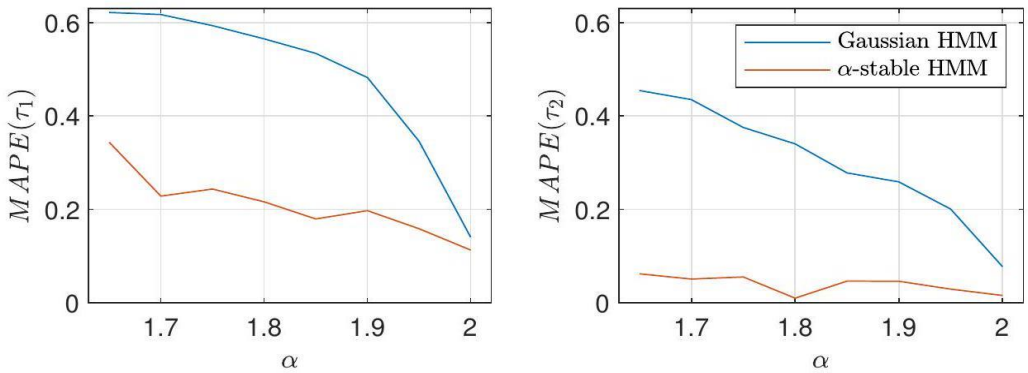

Figure 5. Mean absolute percentage errors of $\hat{\tau}_1$ (left panel) or $\hat{\tau}_2$ (right panel) prediction of trained Gaussian as well as α-stable HMM applied to simulated trajectories of the model (1) with the α-stable distribution. The means were calculated from 100 simulations.

In Figure 6 we additionally illustrate the distribution of the percentage errors. As can be observed, most of the errors are too early predictions, represented by negative values of MAPE. This effect is especially pronounced for the Gaussian HMM applied to the α-stable simulations. For low values of α (heavy tails) all of the predictions too early indicate on the state change. This is, again, the effect of wrongly classifying extreme observations, coming from heavy tails, to the next (worse) state. The errors of the trained α-stable HMM are closer to 0, but for τ_1 predictions their spread, especially for lower values of α, is

higher. Such effect is also visible for τ_2, but the scale of the errors is much lower.

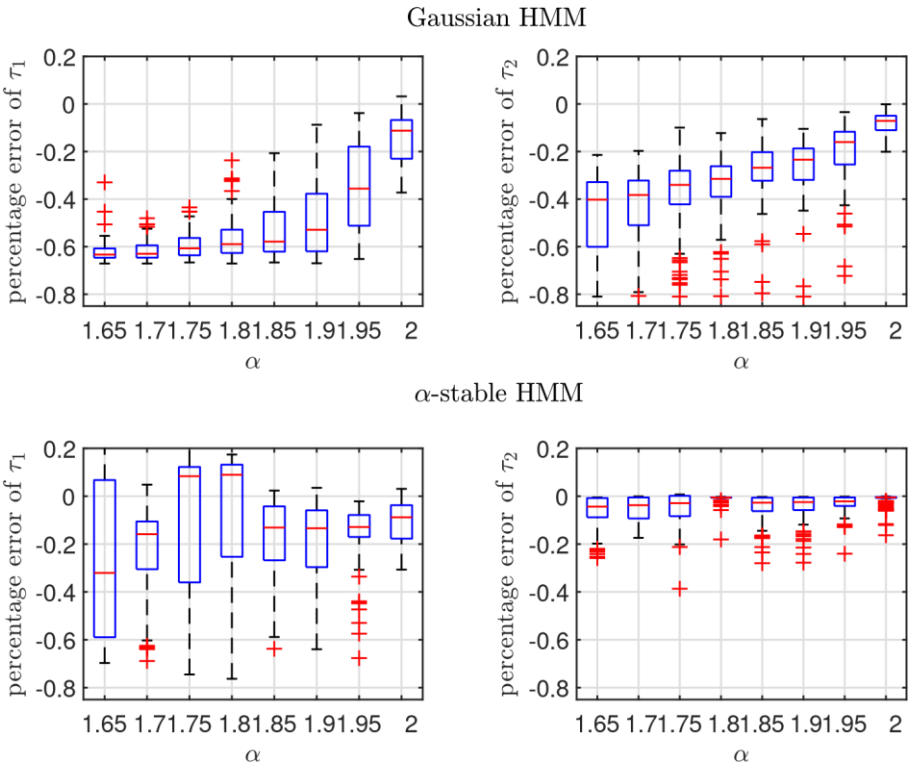

Figure 6. Boxplots of the percentage errors of $\hat{\tau}_1$ (left panels) or $\hat{\tau}_2$ (right panels) prediction calculated from 100 simulated trajectories of the model (1) with the α-stable distribution. Results obtained with the trained Gaussian HMM are plotted in the top panels, while with the α-stable HMM in the bottom panels.

6. Comparison of the predictive performance of Gaussian and non-Gaussian models for real datasets

In this section we demonstrate the efficiency of the proposed methodology for real datasets. In contrast to the simulated data analysis, presented in Section 5, here at the training and prediction steps we use the real HI time series described in Section 3. For each case, both steps are applied for the same real HI trajectories. However, as mentioned above at the training step we analyse the data with full history. In real monitoring systems the proposed methodology can be applied in case when the training step and the prediction steps are applied for different HI trajectories assuming that they exhibit similar characteristics, i.e. when they represent the same HI and come from the same type of machine working under similar working conditions. Because we do not have such data, in this section the efficiency is presented based on exactly the same trajectories at both steps of the procedure.

We start with training the Gaussian (see Eq. (4)) and the α-stable (see Eq. (5)) HMMs using full detrended trajectories of the analysed HI data. Next, we use the trained HMM to predict the next step state from a given point in time based on the past data. We repeat the procedure for all available points, starting from the set of 500 observations, i.e. $T' = 500,501,\ldots,T-1$, where T is the corresponding data length. The predicted states are calculated as the most probable state in the next time point, namely the healthy state prediction corresponds to $\hat{P}(R_{T'+1} = 1) > 0.5$, the warning state prediction to $\hat{P}(R_{T'+1} = 2) > 0.5$, while the alarm state prediction to $\hat{P}(R_{T'+1} = 3) > 0.5$. Note that, here, differently than in the simulation study, we use the same trajectories for training and deriving on-line predictions. The datasets, trained change points τ_1, τ_2 as well as the predicted states are plotted in Figures 7-9 for the FEMTO dataset, AMC Tech dataset 1 and 2, respectively. Note that the values of predicted states are calculated in the online mode. The observed transitions from the warning to the healthy state as well as from the alarm to the warning state come from the fact that at each time point new information (observed HI value) is included in the calculations. It can result in lowering the forecasted warning or alarm state probability. The

corresponding predicted probabilities are plotted in Figures B.10-B.12 in the Appendix. For comparison we also show the predicted probabilities calculated using the switching Kalman filter (SKF) and switching maximum correntropy Kalman filter (SMCKF) methods (see the bottom panels in Figures B.10-B.12) that recently were proposed also for the analysis of HI data with possible non-Gaussian behaviour, see [4] for more details.

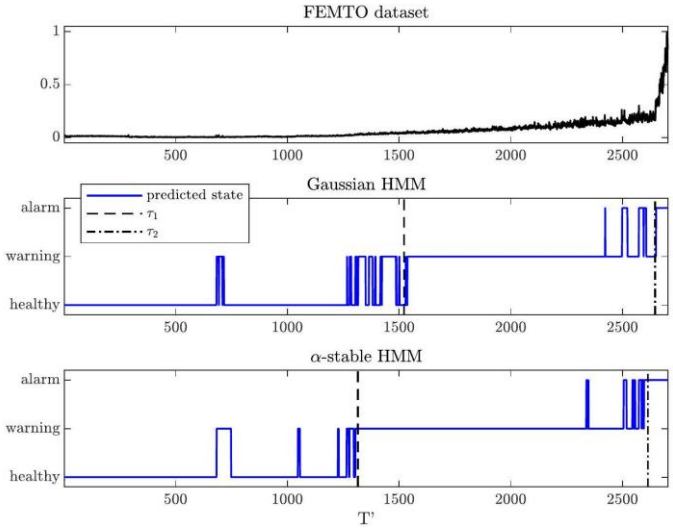

Figure 7. On-line predictions of the next state using the Gaussian (middle panel) as well as the α-stable HMMs (bottom panel) applied to the detrended FEMTO dataset. The original HI data is plotted in the top panel. Change points detected based on the whole data (at the level of the training step of the procedure), are denoted by τ_1 and τ_2.

For the FEMTO dataset (Figures 7 and B.10) the trained change points obtained from the whole data are earlier in the case of the α-stable HMM, especially for τ_1. Both methods show some too early predictions of the warning as well as the alarm state. On-line state prediction is based on smaller samples than the whole dataset, so the methods are more sensitive to short-time changes in the signal behaviour. This effect is more pronounced for the Gaussian HMM. On the other hand, there are no indications on the previous state after the trained change points are achieved for HMM methods. Looking at the corresponding predicted probabilities we observe a similar picture as most of the calculated values are either close to 0 or 1, i.e. indicate on the given state with high certainty. Comparing the predicted HMM-based probabilities with the SKF and MCSKF methods, we observe that the false predictions diminish visibly faster in the former case. The predictions from the α-stable HMM resemble the MCSKF results, while the Gaussian HMM yields similar picture as SKF.

For the AMC Tech dataset 1 (Figures 8 and B.11 in the Appendix) we obtain almost identical results using the Gaussian and α-stable HMMs. Slight differences are visible in the values of the predicted probabilities, but they do not change the qualitative conclusions. Comparing the results with the SKF and MCSKF methods we observe that similar results were obtained with the latter one, with the only exception between points 1560 and 1671. On the other hand, the SKF method predicts the alarm state for almost whole sample, starting from the point $T' = 520$. Such a behaviour was caused by a smoother growth of the HI than in the former dataset. The alarm state not only caused HI increase, but also its strong variability. This was a result of the load type applied in the test rig.

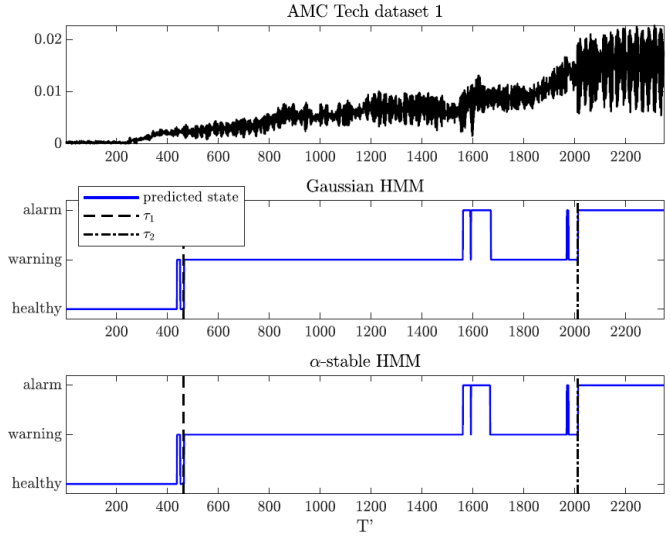

Figure 8. On-line predictions of the next state for the Gaussian (middle panel) as well as the α-stable HMM (bottom panel) applied to the detrended AMC Tech dataset 1. The original HI data is plotted in the top panel. Change points detected based on the whole data (at the level of the training step of the procedure), are denoted by τ_1 and τ_2.

Finally, for the AMC Tech dataset 2 (Figures 9 and B.12 in the Appendix) the trained change points are earlier in the case of the Gaussian than for the α-stable HMM, especially in case of τ_1. Again, some too early predictions of the alarm state are visible for both methods and also of the warning state for the α-stable HMM. However, in the latter case, these predictions are less certain than for the previous datasets, as the values of the probabilities are far from 0 and 1. After the point 1176 all predictions are made with high certainty. On the other hand, the SKF and MCSKF methods do not distinguish between the

healthy and warning state as most of the predicted probabilities are close to 0.5. Similarly as in the previous dataset, it was caused by the model of HI growth. Warning state was correctly detected when HI started to grow steadily. Later, alarm state was detected when HI grew further and increased its variability. This is very similar to what a human expert reaction should be.

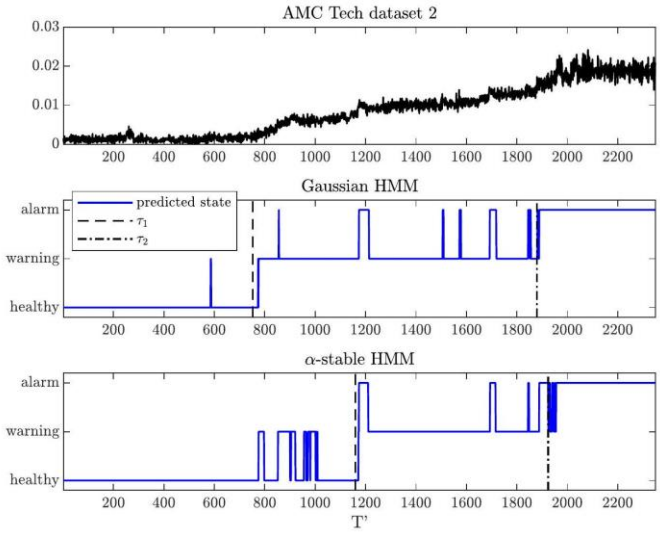

Figure 9. On-line predictions of the next state for the Gaussian (middle panel) as well as the α-stable HMM (bottom panel) applied to the detrended AMC Tech dataset 2. The original HI data is plotted in the top panel. Change points detected based on the whole data (at the level of the training step of the procedure), are denoted by τ_1 and τ_2.

7. Conclusions

In this paper, we discuss the problem of prediction of machine state based on HI data. To this end we propose to use the HMM-based methodology. Due to the fact that the data may exhibit non-Gaussian impulsive behaviour, the classical HMM methodology is extended to the more general case, in which the $\alpha-$stable distribution is assumed as describing the diagnostic features. The $\alpha-$stable distribution is considered as a general class of distributions that captures the heavy- and light-tailed behaviour. The problem of non-Gaussian distribution of diagnostic features analysed in monitoring systems was highlighted by many authors. The non-Gaussianity of HI data may be related to the machine type that is diagnosed but also the level of non-Gaussianity may increase as the damage progresses. Thus, the discussed problem does not appear to be a singular issue, but a general problem which is observed in practical applications.

The problem of machine state prediction can be considered as on-line segmentation of the HI data. This issue was discussed in the literature but most of the methods are dedicated to Gaussian distributed data. Here we are going a step forward. The proposed procedure is based on two steps: training and prediction. In the training step by using the EM algorithm we train the HMM based on benchmark data with full history, while in the prediction step the probability of being in a given state (healthy, warning and alarm) is calculated with every new datapoint. Based on that, one can identify the most probable state of the machine in the next time point.

The efficiency of the $\alpha-$stable HMM approach is verified for the simulated trajectories of the adopted degradation model with Gaussian and non-Gaussian distributions. We have verified that the method dedicated to the HI data with possible non-Gaussian distribution is more effective for on-line segmentation of data with large observations than the classical Gaussian HMM algorithm. In particular the obtained percentage errors (MAPE) of the transition points decreased from 35%-62% to 11%-34% for the warning state and from 20%-45% to 1%-6% for the alarm state, if the $\alpha-$stable HMM approach was applied to non-Gaussian HI data instead of the Gaussian one. Finally, the proposed methodology is applied to three real HI datasets and compared with the Kalman filter-based approach. Models trained and verified in all three cases were very similar to a reaction of a human expert. Therefore, the proposed method can be further developed and possibly applied in an on-line condition monitoring system. With such a new functionality, CMS can reduce human effort required to properly configure anomaly detection procedures. Since in current engineering practice only a few of calculated HIs are actually analysed against anomaly detection, such a progress can significantly reduce unplanned shutdowns, which would improve availability of assets.

The proposed methodology also has some limitations. First, the segmentation algorithm is designed for cases when the random component of the data changes. However, the deterministic components may also change over time, which is often the case in real scenarios. In the article [3], where the HMM-based approach was proposed for offline segmentation, the problem of changing trend was discussed, and a similar methodology could be adopted in our case. Additionally, we

assume the α-stable distribution for the analysed data. As mentioned, this class of distributions can be considered general, as it is a limit for infinite-variance (when $\alpha < 2$) and finite-variance (when $\alpha = 2$) distributions. However, when the amount of data is relatively small, the limiting case may not be a good approximation. In such cases, it is recommended to use more appropriate distributions. Moreover, as can be observed in Figure 6, there are significant differences in the predicted results for different α values. In our methodology the α parameter is estimated during the training part of the procedure based on the distribution of data that is assumed to have similar behaviour to the considered one. However, it is also possible to additionally assess the α parameter before applying the HMM model for state prediction. The procedure might be similar to the one proposed in [2].

Acknowledgements

The work of WŻ, RZ and AW was supported by National Center of Science under Sheng2 project No. UMO2021/40/Q/ST8/00024 "NonGauMech - New methods of processing non-stationary signals (identification, segmentation, extraction, modeling) with non-Gaussian characteristics for the purpose of monitoring complex mechanical structures".

The work of JJ was supported by project No. POIR.01.01.01-00-0350/21 entitled "A universal diagnostic and prognostic module for condition monitoring systems of complex mechanical structures operating in the presence of non-Gaussian disturbances and variable operating conditions" co-financed by the European Union from the European Regional Development Fund under the Intelligent Development Program. The project was carried out as part of the competition of the National Center for Research and Development no: 1/1.1.1/2021 (Szybka Ścieżka).

HS gratefully acknowledges the European Commission for its support of the Marie Sklodowska Curie programme through the ETN MOIRA project (GA 955681).

The work of Tomasz Barszcz was supported by Department of Robotics and Mechatronics, AGH University of Cracow.

References

1. Cempel C. Limit value in the practice of machine vibration diagnostics. Mechanical Systems and Signal Processing 1990; 4(6): 483–493, https://doi.org/10.1016/0888-3270(90)90047-O.
2. Żuławiński W, Maraj-Zygmąt K, Shiri H, Wyłomańska A, Zimroz R. Framework for stochastic modelling of long-term non-homogenous data with non-Gaussian characteristics for machine condition prognosis. Mechanical Systems and Signal Processing 2023; 184: 109677, https://doi.org/10.1016/j.ymssp.2022.109677.
3. Janczura J, Barszcz T, Zimroz R, Wyłomańska A. Machine condition change detection based on data segmentation using a three-regime, α-stable Hidden Markov Model. Measurement 2023; 220: 113399, https://doi.org/10.1016/j.measurement.2023.113399.
4. Shiri H, Zimroz P, Wodecki J, Wyłomańska A, Zimroz R, Szabat K. Using long-term condition monitoring data with non-Gaussian noise for online diagnostics. Mechanical Systems and Signal Processing 2023; 200: 110472, https://doi.org/10.1016/j.ymssp.2023.110472.
5. Nikias CL, Shao M. Signal processing with alpha-stable distributions and applications, USA, Wiley-Interscience: 1995.
6. Lévy P. Calcul des Probabilites. Paris, Gauthier-Villars: 1925.
7. Diez-Olivan A, Del Ser J, Galar D, Sierra B. Data fusion and machine learning for industrial prognosis: trends and perspectives towards industry 4.0. Information Fusion 2019; 50: 92-111, https://doi.org/10.1016/j.inffus.2018.10.005.
8. Yan S, Ma B, Zheng C. Health index extracting methodology for degradation modelling and prognosis of mechanical transmissions. Eksploatacja i Niezawodność – Maintenance and Reliability 2019; 21(1): 137-44, https://doi.org/10.17531/ein.2019.1.15.
9. Si X-S, Wang W, Hu C-H, Zhou D-H. Remaining useful life estimation - a review on the statistical data driven approaches. European Journal of Operational Research 2011; 213(1): 1-14, https://doi.org/10.1016/j.ejor.2010.11.018.
10. Ye Z-S, Xie M. Stochastic modelling and analysis of degradation for highly reliable products. Applied Stochastic Models in Business and Industry 2015; 31(1): 16-32, https://doi.org/10.1002/asmb.2063.
11. Li Q, Yan C, Chen G, Wang H, Li H, Wu L. Remaining useful life prediction of rolling bearings based on risk assessment and degradation state coefficient. ISA Transactions 2022; 129: 413-428, https://doi.org/10.1016/j.isatra.2022.01.031.

12. Heng A, Zhang S, Tan AC, Mathew J. Rotating machinery prognostics: State of the art, challenges and opportunities. Mechanical Systems and Signal Processing 2009; 23(3): 724-739, https://doi.org/10.1016/j.ymssp.2008.06.009.
13. Liao L, Köttig F. Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. IEEE Transactions on Reliability 2014; 63(1): 191-207, https//doi.org/10.1109/TR.2014.2299152.
14. Zhao Z, Wu J, Li T, Sun C, Yan R, Chen X. Challenges and opportunities of AI-enabled monitoring, diagnosis & prognosis: A review. Chinese Journal of Mechanical Engineering 2021; 34(1): 1-29, https://doi.org/10.1186/s10033-021-00570-7.
15. Zhan X, Liu Z, Yan H, Wu Z, Guo C, Jia X. A novel method of health indicator construction and remaining useful life prediction based on deep learning. Eksploatacja i Niezawodność - Maintenance and Reliability 2023; 25(4): 171374, https://doi.org/10.17531/ein/171374.
16. Shi Z, Chehade A. A dual-LSTM framework combining change point detection and remaining useful life prediction. Reliability Engineering & System Safety 2021; 205: 107257, https://doi.org/10.1016/j.ress.2020.10725.
17. Yan B, Ma X, Yang L, Wang H, Wu T. A novel degradation-rate-volatility related effect Wiener process model with its extension to accelerated ageing data analysis. Reliability Engineering & System Safety 2020; 204: 107138, https://doi.org/10.1016/j.ress.2020.107138.
18. Le Son K, Fouladirad M, Barros A. Remaining useful lifetime estimation and noisy gamma deterioration process. Reliability Engineering & System Safety 2016; 149: 76 – 87, https://doi.org/10.1016/j.ress.2015.12.016.
19. Nguyen KT, Fouladirad M, Grall A. Model selection for degradation modeling and prognosis with health monitoring data. Reliability Engineering & System Safety 2018; 169:105 – 116, https://doi.org/10.1016/j.ress.2017.08.004.
20. Ye Y, Yang Q, Zhang J, Meng S, Wang J. A dynamic data driven reliability prognosis method for structural digital twin and experimental validation. Reliability Engineering & System Safety 2023; 240: 109543, https://doi.org/10.1016/j.ress.2023.109543.
21. Jablonski A, Barszcz T, Bielecka M, Breuhaus P. Modeling of probability distribution functions for automatic threshold calculation in condition monitoring systems. Measurement 2013; 46(1): 727-738, https://doi.org/10.1016/j.measurement.2012.09.011.
22. Jin X, Sun Y, Que Z, Wang Y, Chow TW. Anomaly detection and fault prognosis for bearings. IEEE Transactions on Instrumentation and Measurement 2016; 65(9): 2046-2054, https://doi.org/ 10.1109/TIM.2016.2570398.
23. Jiang W, Hong Y, Zhou B, He X, Cheng C. A GAN-based anomaly detection approach for imbalanced industrial time series. IEEE Access 7 2019: 143608-143619, https://doi.org/10.1109/ACCESS.2019.2944689.
24. Vos K, Peng Z, Jenkins C, Shahriar MR, Borghesani P, Wang W. Vibration-based anomaly detection using LSTM/SVM approaches. Mechanical Systems and Signal Processing 2022; 169: 108752, https://doi.org/10.1016/j.ymssp.2021.108752.
25. Yan B, Ma X, Huang G, Zhao Y. Two-stage physics-based Wiener process models for online RUL prediction in field vibration data. Mechanical Systems and Signal Processing 2021; 152: 107378, https://doi.org/10.1016/j.ymssp.2020.107378.
26. Huang M, Yu W, Yang F. Analysis of remaining useful life of slope based on nonlinear Wiener process. Eksploatacja i Niezawodność – Maintenance and Reliability 2024, https://doi.org/10.17531/ein/187160.
27. Huang Z, Xu Z, Ke X, Wang W, Sun Y. Remaining useful life prediction for an adaptive skew-Wiener process model. Mechanical Systems and Signal Processing 2017; 87: 294-306, https://doi.org/10.1016/j.ymssp.2016.10.027.
28. Xi X, Chen M, Zhang H, Zhou D. An improved non-Markovian degradation model with long-term dependency and item-to-item uncertainty. Mechanical Systems and Signal Processing 2018; 105: 467-480, https://doi.org/10.1016/j.ymssp.2017.12.017.
29. Zhang H, Zhou D, Chen M, Shang J. FBM-Based Remaining Useful Life Prediction for Degradation Processes With Long-Range Dependence and Multiple Modes. IEEE Transactions on Reliability 2019; 68(3): 1021-1033, https://doi.org/10.1109/TR.2018.2877643.
30. Ling M, Ng H, Tsui K. Bayesian and likelihood inferences on remaining useful life in two-phase degradation models under gamma process. Reliability Engineering & System Safety 2019; 184: 77-85, https://doi.org/10.1016/j.ress.2017.11.017.
31. Liu H, Song W, Niu Y, Zio E. A generalized Cauchy method for remaining useful life prediction of wind turbine gearboxes. Mechanical Systems and Signal Processing 2021; 153: 107471, https://doi.org/10.1016/j.ymssp.2020.107471.
32. Liu H, Song W, Zhang Y, Kudreyko A. Generalized Cauchy degradation model with long-range dependence and maximum Lyapunov exponent for remaining useful life. IEEE Transactions on Instrumentation and Measurement 2021; 70: 1 – 12 , https://doi.org/10.1109/TIM.2021.3063749.
33. Song W, Liu H, Zio E. Long-range dependence and heavy tail characteristics for remaining useful life prediction in rolling bearing degradation. Applied Mathematical Modelling 2022; 102: 268-284, https://doi.org/10.1016/j.apm.2021.09.041.

34. Duan S, Song W, Zio E, Cattani C, Li M. Product technical life prediction based on multi-modes and fractional Lévy stable motion. Mechanical Systems and Signal Processing 2021; 161: 107974, https://doi.org/10.1016/j.ymssp.2021.107974.
35. Yuan Y, Song W. Degradation prediction of tool based on fractional Levy prediction model. 2022 Global Reliability and Prognostics and Health Management (PHM-Yantai) 2022: 1-4, https://doi.org/10.1109/PHM-Yantai55411.2022.9942217.
36. Li M. Multi-fractional generalized Cauchy process and its application to teletraffic. Physica A: Statistical Mechanics and its Applications 2020; 550: 123982.
37. Song W, Duan S, Zio E, Kudreyko A. Multifractional and long-range dependent characteristics for remaining useful life prediction of cracking gas compressor. Reliability Engineering & System Safety 2022; 225: 108630, https://doi.org/10.1016/j.physa.2019.123982.
38. Lim C, Mba D. Switching Kalman filter for failure prognostic. Mechanical Systems and Signal Processing 2015; 52-53(1): 426 – 435, https://doi.org/10.1016/j.ymssp.2014.08.006.
39. Kordestani M, Samadi MF, Saif M, Khorasani K. A new fault prognosis of MFS system using integrated extended Kalman filter and Bayesian method. IEEE Transactions on Industrial Informatics 2018, https://doi.org/ 10.1109/TII.2018.2815036.
40. Singleton RK, Strangas EG, Aviyente S. Extended Kalman filtering for remaining-useful-life estimation of bearings. IEEE Transactions on Industrial Electronics 2015; 62(3): 1781 – 1790, https://doi.org/10.1109/TIE.2014.2336616.
41. Reuben L, Mba D. Diagnostics and prognostics using switching Kalman filters. Structural Health Monitoring 2014; 13(3): 296 – 306, https://doi.org/10.1177/1475921714522844.
42. Cheng P, Wang H, Stojanovic V, He S, Shi K, Luan X, Liu F, Sun C. Asynchronous fault detection observer for 2-d Markov jump systems. IEEE Transactions on Cybernetics 2021: 1-12, https://doi.org/ 10.1109/TCYB.2021.3112699.
43. Cheng P, Chen M, Stojanovic V, He S. Asynchronous fault detection filtering for piecewise homogenous Markov jump linear systems via a dual hidden Markov model. Mechanical Systems and Signal Processing 2021; 151: 107353, https://doi.org/10.1016/j.ymssp.2020.107353.
44. Sun J, Zuo H, Wang W, Pecht MG. Application of a state space modeling technique to system prognostics based on a health index for condition-based maintenance. Mechanical Systems and Signal Processing 2012; 28: 585 – 596, https://doi.org/10.1016/j.ymssp.2011.09.029.
45. Nedić N, Pršić D, Fragassa C, Stojanović V, Pavlovic A. Simulation of hydraulic check valve for forestry equipment. International Journal of Heavy Vehicle Systems 2017; 24(3): 260-276, https:/doi.org/ 10.1504/IJHVS.2017.084875.
46. Okoh C, Roy R, Mehnen J, Redding L. Overview of remaining useful life prediction techniques in through-life engineering services. Procedia CIRP 2013; 16: 158-163, https://doi.org/10.1016/j.procir.2014.02.006.
47. Zhang Z, Si X, Hu C, Lei Y. Degradation data analysis and remaining useful life estimation: A review on Wiener-process-based methods. European Journal of Operational Research 2018; 271 (3): 775-796, https://doi.org/10.1016/j.ejor.2018.02.033.
48. Lei Y, Li N, Guo L, Li N, Yan T, Lin J. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mechanical Systems and Signal Processing 2018; 104: 799-834, https://doi.org/10.1016/j.ymssp.2017.11.016.
49. Li Y, Zhang X, Ran Y, Zhang G. Reliability modeling and analysis for CNC machine tool based on meta-action. Quality and Reliability Engineering International 2021; 37: 1451–1467, https://doi.org/10.1002/qre.2806.
50. Baum LE, Petrie T. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. The Annals of Mathematical Statistics 1966; 37(6): 1554 – 1563, https://doi.org/10.1214/aoms/1177699147.
51. Mor B, Garhwal S, Kumar A. A systematic review of hidden Markov models and their applications. Archives of Computational Methods in Engineering 2021; 28: 1429-1448, https://doi.org/10.1007/s11831-020-09422-4.
52. Chen J-H, Zou S-L. An intelligent condition monitoring approach for spent nuclear fuel shearing machines based on noise signals. Applied Sciences 2018; 8(5), https://doi.org/10.3390/app8050838.
53. 53. Sadhu A, Prakash G, Narasimhan S. A hybrid hidden Markov model towards fault detection of rotating components. Journal of Vibration and Control 2017; 23(19): 3175-3195, https://doi.org/10.1177/1077546315627934.
54. Hwang KH, Lee JM, Hwang Y. A new machine condition monitoring method based on likelihood change of a stochastic model. Mechanical Systems and Signal Processing 2013; 41(1): 357-365, https://doi.org/10.1016/j.ymssp.2013.08.003.
55. Ocak H, Loparo KA. HMM-Based Fault Detection and Diagnosis Scheme for Rolling Element Bearings. Journal of Vibration and Acoustics 2004; 127(4): 299-306, https://doi.org/10.1115/1.1924636.

56. Fan L, Wang S, Duan H, Ran H. Fatigue crack fault diagnosis and prognosis based on hidden semi-Markov model. The Journal of Engineering 2019; 2019(13): 406-410, https://doi.org/10.1049/joe.2018.8960.
57. Ying J, Kirubarajan T, Pattipati K, Patterson-Hine A. A hidden Markov model-based algorithm for fault diagnosis with partial and imperfect tests. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 2000; 30: 463-473, https://doi.org/10.1109/5326.897073.
58. Pravin RN, Raman KG, Kumar PP, Bharathi KS, Shajeev M. Performance evaluation of HMM and neural network in motorbike fault detection system. 2011 International Conference on Recent Trends in Information Technology (ICRTIT), Chennai, India 2011: 1175-1179, https://doi.org/10.1109/ICRTIT.2011.5972479.
59. Zhao Q. Fault diagnosis method for wind power equipment based on hidden Markov model. Wireless Communications and Mobile Computing 2022; 2022(4): 1-9, http://dx.doi.org/10.1155/2022/6937616.
60. Tobon-Mejia DA, Medjaher K, Zerhouni N, Tripot G. Hidden Markov models for failure diagnostic and prognostic. 2011 Prognostics and System Health Management Conference, IEEE 2011: 1-8, https://doi.org/10.1109/PHM.2011.5939488.
61. Zhou Z-J, Hu C-H, Xu D-L, Chen M-Y, Zhou D-H. A model for real-time failure prognosis based on hidden Markov model and belief rule base. European Journal of Operational Research 2010; 207(1): 269-283, https://doi.org/10.1016/j.ejor.2010.03.032.
62. Soualhi A, Clerc G, Razik H. Fault prognosis based on hidden Markov models. 2015 IEEE Workshop on Electrical Machines Design, Control and Diagnosis (WEMDCD) 2015: 271-278, https://doi.org/10.1109/WEMDCD.2015.7194540.
63. Zhiyong G, Jiwu L, Rongxi W. Prognostics uncertainty reduction by right-time prediction of remaining useful life based on hidden Markov model and proportional hazard model. Eksploatacja i Niezawodność - Maintenance and Reliability 2021; 23(1): 154-164, https://doi.org/10.17531/ein.2021.1.16.
64. Baruah P, Chinnam RB. HMMs for diagnostics and prognostics in machining processes. International Journal of Production Research 2005; 43(6): 1275-1293, https://doi.org/10.1080/00207540412331327727.
65. Eleftheroglou N, Galanopoulos G, Loutas T. Similarity learning hidden semi-Markov model for adaptive prognostics of composite structures. Reliability Engineering & System Safety 2004; 243: 109808, https://doi.org/10.1016/j.ress.2023.109808.
66. Ramezani SB, Killen B, Cummins L, Rahimi S, Amirlatifi A, Seale M. A survey of HMM-based algorithms in machinery fault prediction. 2021 IEEE Symposium Series on Computational Intelligence (SSCI) 2021: 1-9, https://doi.org/10.1109/SSCI50451.2021.9659838.
67. Mandelbrot B. The Pareto-Lévy Law and the distribution of income. International Economic Review 1960; 1(2): 79-106, https://doi.org/10.2307/2525289.
68. Chouri B, Fabrice M, Dandache A, Aroussi MEl, Saadane R. Bearing fault diagnosis based on alpha-stable distribution feature extraction and SVM classifier. 2014 International Conference on Multimedia Computing and Systems (ICMCS) 2014: $1545-1550$, https://doi.org/10.1109/ICMCS.2014.6911199.
69. Gao M, Yu G, Wang T. Impulsive gear fault diagnosis using adaptive Morlet wavelet filter based on alpha-stable distribution and kurtogram. IEEE Access 2019; 7: 72283-72296, https://doi.org/10.1109/ACCESS.2019.2919981.
70. Xiong Q, Zhang W, Xu Y, Peng Y, Deng P. Alpha-stable distribution and multifractal detrended fluctuation analysis-based fault diagnosis method application for axle box bearings. Shock and Vibration 2018; 2018: 1-12, https://doi.org/10.1155/2018/1737219.
71. Gang Y, Changning L, Xu Y, Jianfeng Z. A new statistical modeling and detection method for rolling element bearing faults based on alpha-stable distribution. Mechanical Systems and Signal Processing 2013; 41(1-2): 155-175, https://doi.org/10.1016/j.ymssp.2013.08.015.
72. Li X, Teng W, Peng D, Ma T, Wu X, Liu Y. Feature fusion model based health indicator construction and selfconstraint state-space estimator for remaining useful life prediction of bearings in wind turbines. Reliability Engineering & System Safety 2023; 233:109124, https://doi.org/10.1016/j.ress.2023.109124.
73. Hou W, Peng Y. Adaptive ensemble gaussian process regression-driven degradation prognosis with applications to bearing degradation. Reliability Engineering & System Safety 2023; 239: 109479, https://doi.org/10.1016/j.ress.2023.109479.
74. Saha DK, Hoque ME, Badihi H. Development of intelligent fault diagnosis technique of rotary machine element bearing: a machine learning approach. Sensors 2022; 22(3): 1073, https://doi.org/10.3390/s22031073.
75. Niu G, Liu E, Wang X, Zhang B. A hybrid bearing prognostic method with fault diagnosis and model fusion. IEEE Transactions on Industrial Informatics 2023; 20(1): 864-872, https://doi.org/10.1109/TII.2023.3265532.

76. Islam MM, Prosvirin AE, Kim J-M. Data-driven prognostic scheme for rolling-element bearings using a new health index and variants of least-square support vector machines. Mechanical Systems and Signal Processing 2021; 160: 107853, https://doi.org/10.1016/j.ymssp.2021.107853.
77. Li W, Shang Z, Gao M, Qian S, Feng Z. Remaining useful life prediction based on transfer multi-stage shrinkage attention temporal convolutional network under variable working conditions. Reliability Engineering & System Safety 2022; 226: 108722, https://doi.org/10.1016/j.ress.2022.108722 .
78. Suh S, Lukowicz P, Lee YO. Generalized multiscale feature extraction for remaining useful life prediction of bearings with generative adversarial networks. Knowledge-Based Systems 2022; 237: 107866, https://doi.org/10.1016/j.knosys.2021.10786.
79. Lin T, Song L, Cui L, Wang H. Advancing RUL prediction in mechanical systems: A hybrid deep learning approach utilizing non-full lifecycle data. Advanced Engineering Informatics 2024; 61: 102524, https://doi.org/10.1016/j.aei.2024.102524.
80. Sternharz G, Elhalwagy A, Kalganova T. Data-efficient estimation of remaining useful life for machinery with a limited number of run-to-failure training sequences. IEEE Access 2022; 10: 129443-129464, https://doi.org/10.1109/ACCESS.2022.3226780.
81. Bhavsar K, Vakharia V. Prediction of remaining useful life (RUL) of bearing using exponential degradation model. In Sudarshan TS, Pandey KM, Misra RD, Patowari PK, Bhaumik S (eds.): Recent Advancements in Mechanical Engineering. Lecture Notes in Mechanical Engineering. Singapore, Springer: 2023, https://doi.org/10.1007/978-981-19-3266-3_34.
82. Nolan JP. Univariate Stable Distributions: Models for Heavy Tailed Data, Springer: 2020. https://doi.org/10.1007/978-3-030-52915-4
83. Zaidi SSH, Aviyente S, Salman M, Shin K-K, Strangas EG. Prognosis of gear failures in DC starter motors using hidden Markov models. IEEE Transactions on Industrial Electronics 2011; 58(5): 1695-1706, https://doi.org/10.1109/TIE.2010.2052540.
84. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological) 1977; 39(1): 1-22, https://doi.org/10.1111/j.2517-6161.1977.tb01600.x.
85. Kim CJ. Dynamic linear models with Markov-switching. Journal of Econometrics 1994; 60: 1-22, https://doi.org/10.1016/0304-4076(94)90036-1.
86. Nolan JP. Numerical calculation of stable densities and distribution functions. Communications in Statistics - Stochastic Models 1997; 13: 759-774, https://doi.org/10.1080/15326349708807450.
87. Rousseeuw PJ, Croux C. Alternatives to the median absolute deviation. Journal of the American Statistical Association 1993; 88(424): 1273–1283, https://doi.org/10.2307/2291267.
88. Brockwell PJ, Davis RA. Introduction to time series and forecasting, Springer: 2016. https://doi.org/10.1007/978-3-319-29854-2
89. Ma Y, Genton MG, Highly robust estimation of the autocovariance function, Journal of Time Series Analysis 2000; 21(6): 663–684, https://doi.org/10.1111/1467-9892.00203.
90. Chambers JM, Mallows CL, Stuck BW. A method for simulating stable random variables. Journal of the American Statistical Association 1976; 71(354): 340–344, https://doi.org/10.2307/2285309.

Appendix A. Technical details on the degradation model

Appendix A.1. Degradation model fitting procedure

The procedure of fitting the degradation model for a HI sample $\mathbf{S} = [S_1, \dots, S_T]$ (note that it must contain data from all three states) is as follows. First, we set the values of change points $\hat{\tau}_1$ and $\hat{\tau}_2$. If they are not known a priori, one can estimate them using a selected segmentation method, see e.g. [3, 4]. Then, we identify the trend part $\eta(t)$ using moving median with a given window length and subtract it from the original signal to obtain the realisation of the random part, i.e. $\mathbf{X} = [X_1, \dots, X_T]$. The next step is the identification of the time-varying scale component. For that purpose, we calculate a robust scale estimator Q [87] in a moving window over $\mathbf{X}$ and obtain the series $\mathbf{SC} = [SC(1), \dots, SC(T)]$. For a window $\mathbf{W} = [W_1, \dots, W_{N_w}]$, this scale estimator is the following k-th order statistic times a constant

$$Q_{\mathbf{W}} = 2.2191\{|W_i - W_j|; i < j\}_{(k)},\ k = \binom{c}{2} \approx \binom{N_w}{2}/4,\ c = \left[\frac{N_w}{2}\right] + 1, \tag{A.1}$$

where $[\cdot]$ denotes the integer part.

Next, we standardise the $\mathbf{X}$ sequence to obtain $\mathbf{Z} = [Z_1, \dots, Z_T]$, by setting $Z_t = X_t/SC(t)$. After that, we divide the $\mathbf{SC}$ sequence

state-wise, setting $\mathbf{SC}^{(1)} = [SC(1), \dots, SC(\hat{\tau}_1 - 1)], \mathbf{SC}^{(2)} = [SC(\hat{\tau}_1), \dots, SC(\hat{\tau}_2 - 1)]$ and $\mathbf{SC}^{(3)} = [SC(\hat{\tau}_2), \dots, SC(T)]$. Afterwards, to each of these segments we fit appropriate functions corresponding to the model assumptions on the $SC(t)$ function. The constant function of the state 1 (and, in consequence, the estimator $\hat{s}_1$) is equal to the mean of $\mathbf{SC}^{(1)}$ sequence. For state 2 and $\mathbf{SC}^{(2)}$ series, we fit a linear function $f_2(t) = a_2 t + b_2$ and set $\hat{s}_2 = f_2(\hat{\tau}_2)$. Analogously, in case of state 3 and $\mathbf{SC}^{(3)}$ vector, we fit an exponential function $f_3(t) = a_3 \exp(b_3 t)$ and calculate $\hat{s}_3 = f_3(T)$ using the least squares method.

In a similar manner as above, we divide $\mathbf{Z}$ into $\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}$ and $\mathbf{Z}^{(3)}$ sequences; i.e. we set $\mathbf{Z}^{(1)} = \left[Z_1, \dots, Z_{\hat{\tau}_1 - 1}\right], \mathbf{Z}^{(2)} = \left[Z_{\hat{\tau}_1}, \dots, Z_{\hat{\tau}_2 - 1}\right]$ and $\mathbf{Z}^{(3)} = \left[Z_{\hat{\tau}_2}, \dots, Z_T\right]$. Then, to each of these series, we fit a separate autoregressive model (3) using the robust Yule-Walker method [88, 89]. This procedure for the series $\mathbf{Z}^{(i)}, i = 1,2,3$, is as follows. First, let us define the robust autocovariance estimator for this series [89]

$$\hat{\gamma}(h) = \frac{1}{4}[(Q_{\mathbf{U}+\mathbf{V}})^2 - (Q_{\mathbf{U}-\mathbf{V}})^2], \tag{A.2}$$

where $\mathbf{U}$ is the vector consisting of $N_i - h$ first observations of $\mathbf{Z}^{(i)}$, and $\mathbf{V}$ is the vector consisting of $N_i - h$ last observations of $\mathbf{Z}^{(i)}$. Here, N_i denotes the length of $\mathbf{Z}^{(i)}$ series. One can see that the considered robust autocovariance estimator is based on the robust scale estimator Q (A.1). Then, to obtain estimated coefficients $\hat{\phi}_1^{(i)}, \dots, \hat{\phi}_p^{(i)}$, we solve the following system of equations

$$\hat{\Phi} = \hat{\Gamma}^{-1}\hat{\gamma}, \tag{A.3}$$

where

$$\hat{\Phi} = \left[\hat{\phi}_1^{(i)}, \dots, \hat{\phi}_p^{(i)}\right]', \ \hat{\gamma} = [\hat{\gamma}(1), \dots, \hat{\gamma}(p)]' \tag{A.4}$$

and $\hat{\Gamma}$ is a $p \times p$ matrix, where the element in k-th row and l-th column is defined as $\hat{\Gamma}_{k,l} = \hat{\gamma}(k - l)$. The order p is selected using the method described in [2]. For given $p = 1,2, \dots, p_{\max}$ (considered potential orders), we first fit an autoregressive model, i.e. calculate $\hat{\phi}_1^{(i)}, \dots, \hat{\phi}_p^{(i)}$ from (A.3). Then, we calculate the residuals of the fitted model, that is, we put the estimated $\hat{\phi}_1^{(i)}, \dots, \hat{\phi}_p^{(i)}$ values and the elements of $\mathbf{Z}^{(i)}$ to the model equation (3) to calculate its right-hand (for t equal to each of time indices contained in $\mathbf{Z}^{(i)}$). As a result, for each considered potential order p, we obtain the residual vector $\Xi_p^{(i)}$. Finally, we select p that minimizes the following criterion

$$K(p) = \max_{h=1,\dots,h_{\max}} \left|\hat{\rho}_p(h)\right|^2, \tag{A.5}$$

where $\hat{\rho}_p(h)$ is the robust autocorrelation estimator for the residual series $\Xi_p^{(i)}$ given by [89]

$$\hat{\rho}_p(h) = \frac{(Q_{\mathbf{U}+\mathbf{V}})^2 - (Q_{\mathbf{U}-\mathbf{V}})^2}{(Q_{\mathbf{U}+\mathbf{V}})^2 + (Q_{\mathbf{U}-\mathbf{V}})^2}. \tag{A.6}$$

In (A.6), $\boldsymbol{U}$ is the vector consisting of $N_i - h$ first elements of $\Xi_p^{(i)}$, and $\boldsymbol{V}$ is the vector consisting of $N_i - h$ last elements of $\Xi_p^{(i)}$.

Appendix A.2. Simulation of degradation model trajectories

Using the model (1), we are able to simulate synthetic HI trajectories. First, we set the sample length T and the change points $1 < \tau_1 < \tau_2 < T$. Based on these values, the generation of the signal S_t will be divided into three states with respect to t, according to the convention described in Subsection 3.3. Then, we simulate the innovations sequence ξ_t, from the Gaussian or symmetric α-stable distribution [90] with zero mean. After that, using this sequence, we simulate the AR model Z_t from (3) with different sets of the coefficients for each state $\phi_1^{(i)}, \dots, \phi_p^{(i)}$[88]. Next, the sequence Z_t is multiplied by the $SC(t)$ function with the assumed values s_1, s_2, s_3 to obtain the random component X_t, (2). Finally, we add an assumed trend term $\eta(t)$ (see Eq. (1)) which completes the procedure of S_t signal simulation.

Appendix B. Additional figures

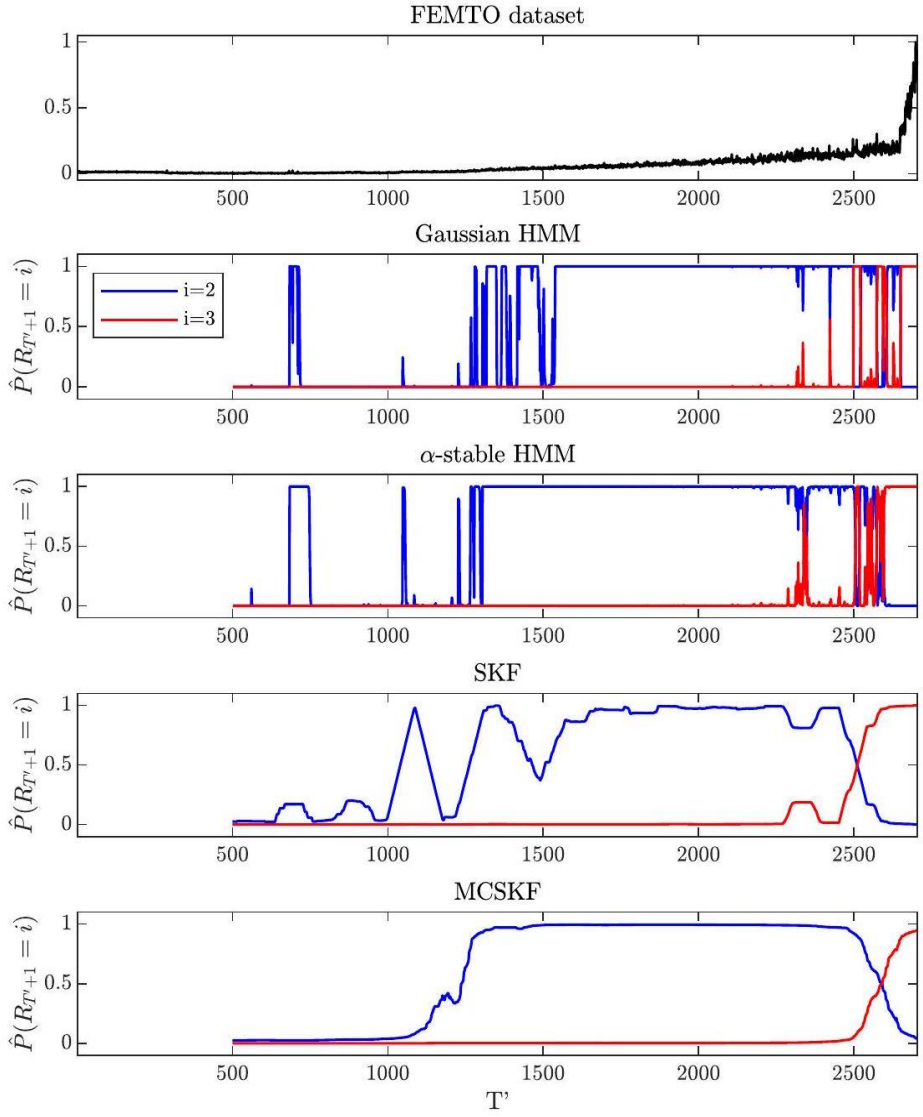

Figure B.10. On-line predictions of the next state probability for the Gaussian as well as the α-stable HMM applied to the detrended FEMTO dataset. The original HI data is plotted in the top panel. For comparison, prediction results of the SKF and MCSKF methods are also plotted in the bottom panels.

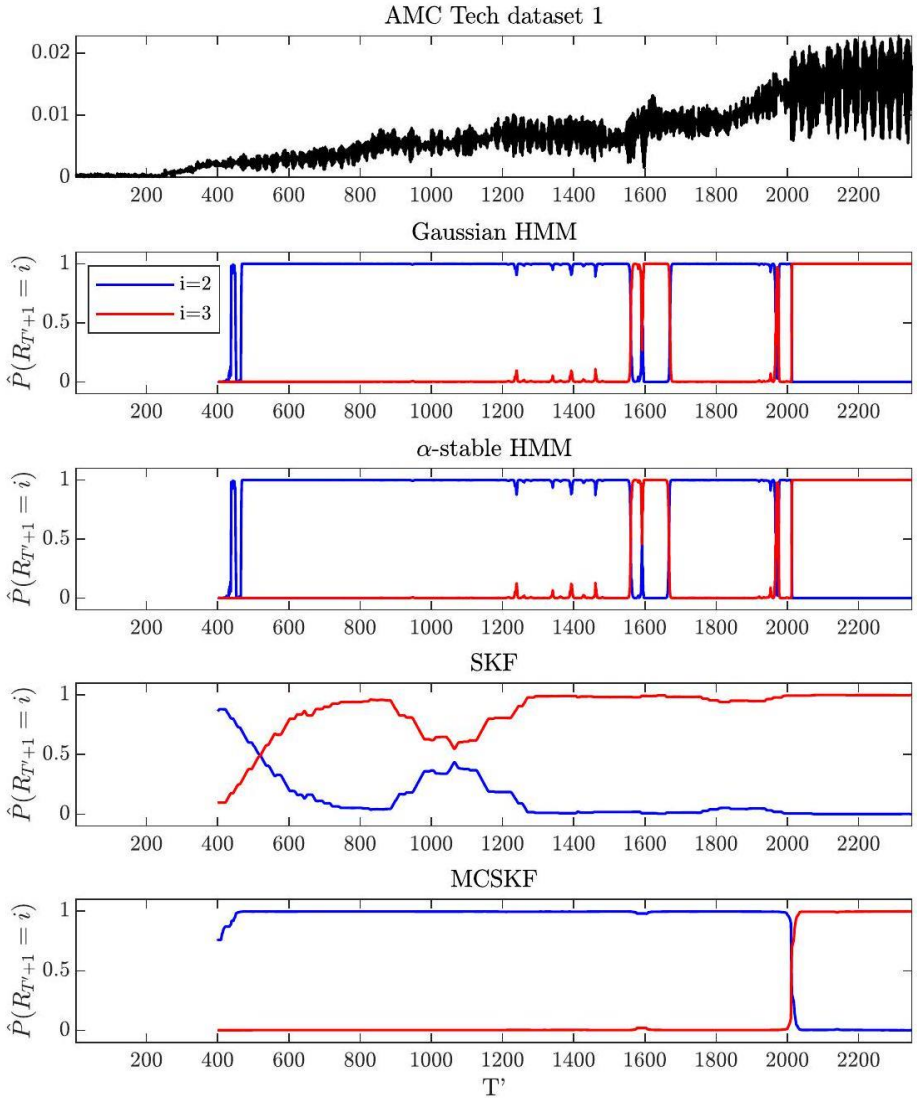

Figure B.11. On-line predictions of the next state probability for the Gaussian as well as the α-stable HMM applied to the detrended AMC Tech dataset 1. The original HI data is plotted in the top panel. For comparison, prediction results of the SKF and MCSKF methods are also plotted in the bottom panels.

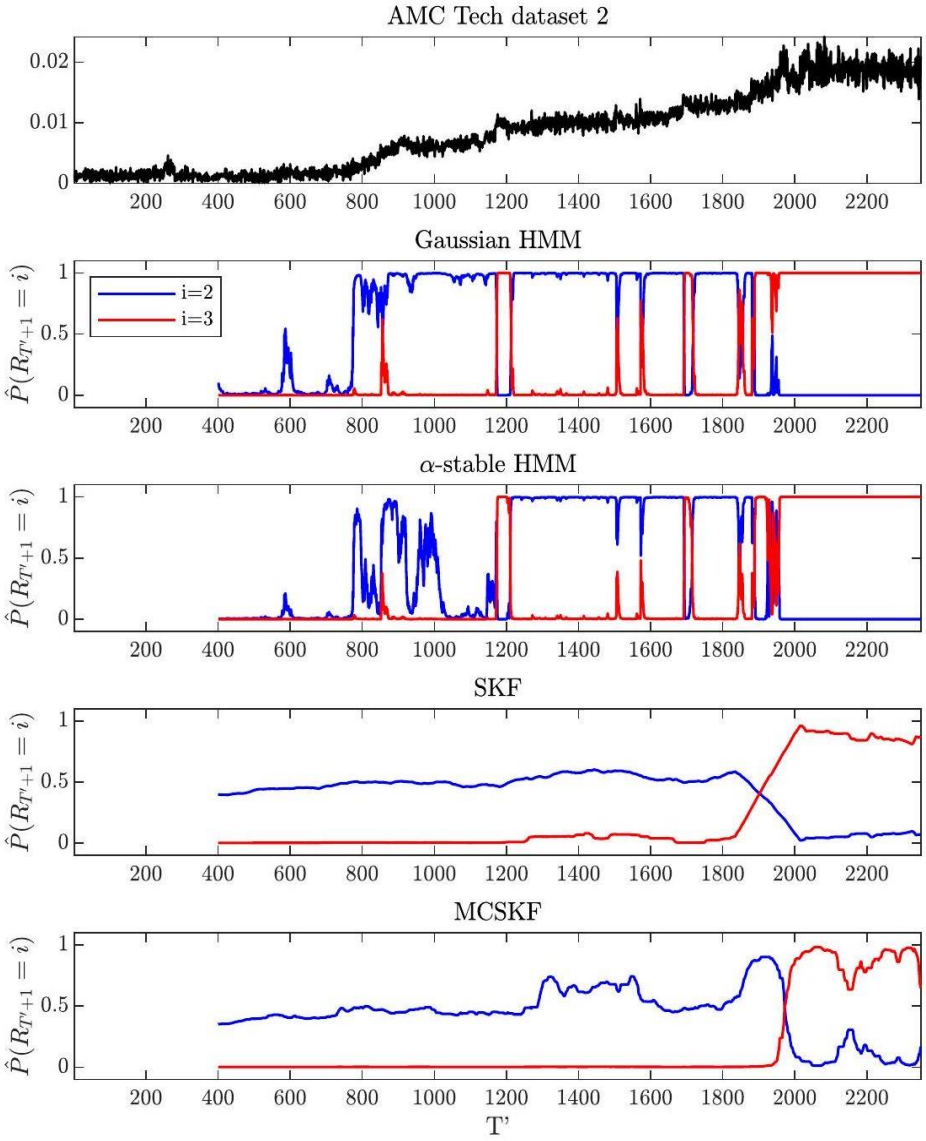

Figure B.12. On-line predictions of the next state probability for the Gaussian as well as the α-stable HMM applied to the detrended AMC Tech dataset 2. The original data HI is plotted in the top panel. For comparison, prediction results of the SKF and MCSKF methods are also plotted in the bottom panels.

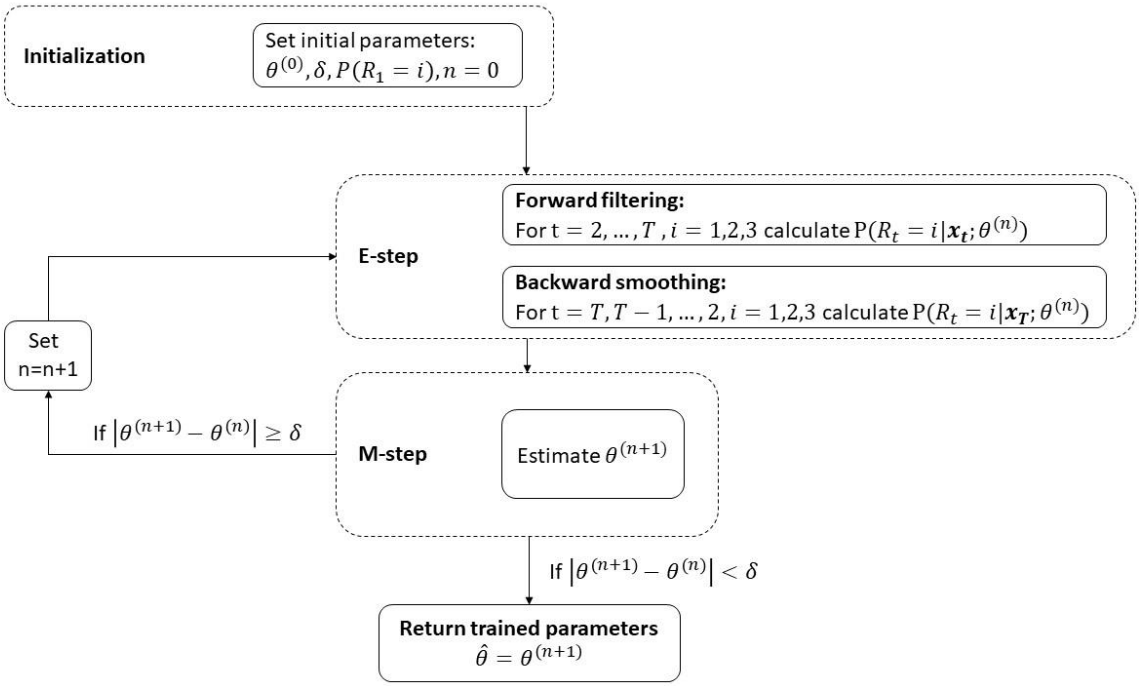

Figure B.13. Scheme of the training procedure for the HMM parameters based on the EM algorithm (see Section 4.2).