Multi-Modal Hidden Markov Model-based Approach for Tool Wear Monitoring

Omid Geramifard, Member, IEEE, Jian-Xin Xu, Fellow, IEEE, Jun-Hong Zhou, Member, IEEE, Xiang Li, Member, IEEE

Abstract—In this paper, a novel multi-modal hidden Markov model-based approach is proposed for tool wear monitoring. The proposed approach improves the performance of pre-existing hidden Markov model-based approach named physically segmented hidden Markov model with continuous output (PSHMCO) by using multiple PSHMCOs in parallel. In this multi-modal approach, each PSHMCO captures and emphasizes on a different tool wear regimen. In this article, three weighting schemes, namely, bounded hindsight, discounted hindsight and semi-nonparametric hindsight are proposed and two switching strategies named soft- and hard-switching are introduced to combine the outputs from multiple modes into one. As an illustrative example, the proposed approach is applied to tool wear monitoring in a computer numerically controlled milling machine. The performance of the multi-modal approach with various weighting schemes and switching strategies is reported and compared with PSHMCO.

Index Terms—Tool Condition Monitoring, Hidden Markov Model, Multi Modal Switching, Diagnostics.

I. INTRODUCTION

In the fast developing and high quality industrial age of today, there is an ever increasing demand to improve the quality of products while reducing the downtime in the industrial machinery. To satisfy this demand, tool condition monitoring approaches are being developed to assess and estimate the condition of tools and pieces used in a machinery system based on non-intrusively acquired signals and without causing unnecessary downtime [1–6].

One of the major issues that would lead to low quality products and cause downtime is tool wear in different types of machinery. As a result, tool wear monitoring (TWM) helps to improve the quality and precision in the products [6], [7].

The idea of tool wear monitoring is to estimate the wearing condition of the tool in the machinery system. At each time step, the wearing condition is a continuous measure based on the past input data. The input data is a set of features extracted from the non-intrusively sensed and captured signals [2], [3]. Signals such as force, vibration, and acoustic can be captured and recorded using various sensors mounted on the machinery system [3], [4]. Estimating the tool wear as a continuous measure instead of setting thresholds and differentiating distinct health states would enable smoother condition based maintenance systems that can incorporate various quality thresholds for different applications to guarantee different qualities in various products [3], [8]. As an example, in a milling machine, the extracted features are used as inputs to predict the continuous wearing metric of the cutter [3], [7–9]. In this work, TWM in a computer numerically controlled (CNC)-milling machine is used as an illustrative example. While formulating the wearing process, gradual degradation is assumed, which holds in most wearing and aging problems (not sudden breakage). Thus, the ideas and approaches developed in this work are applicable to all gradual TWM problems.

As in practice it is infeasible to assess the tool wearing based on physical models, researchers have tended to devise prediction models based on the historical data using data-driven approaches. Consequently, various approaches are proposed to perform tool wear monitoring for different applications. Artificial Neural Network (ANN) is one of the most commonly used approaches in this area. ANNs are used in [10] providing nonlinear projection without the need for prior knowledge. As used in [11], [12], fuzzy inference system-based approaches are another common type of approach used to estimate the tool wear condition, which generally requires a priori knowledge while determining rules and membership functions. The strategy exploiting neuro-fuzzy tools which are data driven and can be regarded as special classes of ANNs, have been also applied to tool wear condition monitoring [13].

Hidden Markov model (HMM) and hidden semi-Markov model are used in [3], [7], [8], [14], [15] to recognize different tool conditions and health states in various applications. In [3], a hidden Markov model-based approach named physically segmented hidden Markov model with continuous output (PSHMCO) is proposed to estimate the continuous tool wear of cutters in a CNC-milling machine, where an explicit relationship between the physical states and the hidden state is given. In [2], the PSHMCO approach outperforms the conventional ANN approaches. However, PSHMCO, that uses a single HMM, adopts a fixed regimen, hence may lose the desired generalization property.

In this work, to improve the performance of PSHMCO, multiple PSHMCOs are used in parallel as multiple modes. In this multi-modal HMM-based (m²HMM) approach, each
PSHMCO captures and emphasizes on a different tool wear regimen. In this article, three weighting schemes, namely, **bounded hindsight, discounted hindsight** and **semi-nonparametric hindsight** are proposed and two switching strategies named **soft- and hard-switching** are introduced to combine the outputs from multiple modes into one. The performance of the multi-modal approach with various weighting schemes and switching strategies is reported and compared with PSHMCO.

The paper is organized as follows. Section II briefly introduces single HMM-based prediction approach called PSHMCO and describes the proposed windowed variant of it. In section III, a multi-modal HMM-based approach is proposed with three weighting schemes and switching strategies along with the windowed variant of m^2HMM. Section IV describes the experimental setup and the extracted features. Then, preliminary experimental results are provided and compared between the single and multi-modal approaches in Section V. Furthermore, the switching strategies, weighting schemes and the windowed variants are compared in section VI. Also, the robustness of the semi-nonparametric weighting scheme with respect to its hyper-parameter value is shown in this section. Finally, the paper is concluded in section VII.

## II. SINGLE HMM-BASED APPROACH

Here, the physically segmented hidden Markov model-based approach with continuous output (PSHMCO) proposed in [3] for tool wear monitoring is briefly reviewed as the single HMM-based approach for further discussions in this article. Moreover, a windowing algorithm is proposed to reduce the computational cost in the PSHMCO.

### A. Review of PSHMCO

Contrary to the conventional use of HMMs which is in classification, in [3] HMM is applied to a continuous problem (regression). In PSHMCO, explicit relationship is provided between the tool conditions (physical health states) and the hidden state values of the HMM. Then, the relationship is further exploited to directly compute the parameters using maximum likelihood method [16] assuming to have a complete training set. Finally, the state estimation is described for different points in time.

In order to use the HMM in the PSHMCO approach, firstly, the output space will be discretized into m ordinal hidden state values or health states \( \{H_1, \ldots, H_m\} \). After that the outputs in the training set would be discretized and assigned to those health states (segmented). Then, the parameters of the hidden Markov model are directly estimated based on the complete training dataset with discretized outputs. Hence, when the new testing data is given to the HMM model, using the inference algorithms and the learned parameters, a probability distribution over real-valued health states (hidden state values) can be computed for each time step. Finally, using the calculated probability distribution and the discretized real-valued labels of health states, an expected real-valued output will be computed for each time step. For more details in the PSHMCO approach refer to [3].

After parameter estimation, the values of HMM state variables at each time step \( t \) may be computed based on given observations with a joint probability \( \gamma_t \). \( \gamma_t(i) \) is the joint probability of observing all input features up to current time \( T \) while being at \( i \)th health state at time step \( t (t \leq T) \). Based on [17], \( \gamma_t \) can be defined and computed as

\[
\gamma_t = P(S_t, O_{1:T} | \lambda) = [\gamma_t(i)]_{m \times 1},
\]

\[
\gamma_t(i) = P(S_t = H_i, O_{1:T} | \lambda) = \alpha_t(i) \beta_t(i),
\]

where \( P(., .) \) is the conditional probability, \( m \) is the number of health states, \( S_t \) is the hidden state variable at time \( t \) and \( H_i \) is the \( i \)th hidden state value, \( O_{1:T} \) is the set of observations from time step 1 up to \( T \). \( \alpha_t \) and \( \beta_t \) are forward-backward variables which may be computed using the forward-backward algorithm [16], [17]. \( \lambda \) is the parameter set of the HMM used in PSHMCO which can be defined as follows [3]

\[
\lambda = \{\pi_0, p_1, \ldots, p_m, \mu_1, \ldots, \mu_m, \Sigma_1, \ldots, \Sigma_m\}
\]

where \( \pi_0 \) is the prior probability distribution of the initial health state, \( p_i \) is the probability of self-transition in \( i \)th health state, \( m \) is the number of health states, \( \mu_i \) and \( \Sigma_i \) are respectively mean and covariance matrix used in the Gaussian distribution to compute the emission probability at time \( t \) given the fact that \( S_t = H_i \).

After computing \( \gamma_t \) based on (1), since the task of TWM in machinery systems is to predict the health state of the tool at the current time \( T \) given all the observations \( (O_{1:T}) \) from time step 1 up to \( T \), the tool wear can be estimated as indicated in [3] as follows

\[
\hat{y}_T = \sum_{i=1}^{m} P(S_T = H_i | O_{1:T}, \lambda) \times H_i,
\]

where \( P(S_T = H_i | O_{1:T}, \lambda) \) is the probability of being at \( i \)th health state at time \( T \) given all observations \( (O_{1:T}) \) and it can be calculated using (1) as follows

\[
P(S_T = H_i | O_{1:T}, \lambda) = \frac{P(O_{1:T}, S_T = H_i | \lambda)}{P(O_{1:T} | \lambda)} = \frac{\gamma_T(i)}{\sum_{j=1}^{m} \gamma_T(j)}.
\]

Finally, the continuous output of PSHMCO, \( \hat{y}_T \), which corresponds to the expected amount of tool wear at the current time step \( T \), can be calculated based on (3) and (4) as follows

\[
\hat{y}_T = \sum_{i=1}^{m} \frac{\gamma_T(i)}{\sum_{j=1}^{m} \gamma_T(j)} \times H_i = \frac{\sum_{i=1}^{m} \gamma_T(i) \times H_i}{\sum_{j=1}^{m} \gamma_T(j)}.
\]

### B. Windowing Algorithm

In order to reduce the computational cost in PSHMCO a windowing algorithm is introduced. In this algorithm, instead of providing the full observation sequence to the HMM up to current time, \( O_{1:T} \), a windowed observation sequence of the recently past observations up to current time, \( O_{T-L_w+1:T} \) is given to HMM where \( L_w \) is the window length. However, to preserve the general trend of the sequence and the path that
HMM is taking through the sequence, the only parameter that has to be updated as the time proceeds is the initial probability. As stated in [3], whenever the prediction starts on a newly given sequence, if no data is available on the initial degradation state, the initial probability is assumed to have a uniform distribution on all possible states. In the proposed windowed approach instead of preserving and using all the past observations and involving all of them into the current predictions, only the windowed observation sequence is used. However, the initial probability distribution at each time step is updated for the next sequence window as follows

\[ \pi_{0,T+1}^T = \gamma_1^T = P(S_1^T, O_{1:T}^T | \lambda_T^T), \]
\[ O_{1:T}^T = O_{T-L_u+1:T} = \{ O_{T-L_u+1}, \ldots , O_T \}, \quad (6) \]
\[ S_1^T = S_{T-L_u+1}, \]

where \( \pi_{0,T+1}^T \) is the updated initial probability at time \( T \) that must be used for the prediction at the next time step, \( T + 1 \). \( \lambda_T^T \) is the updated parameter set for time \( T \). \( \lambda_T^T \) is identical to the parameter set of the given HMM, \( \lambda \), except for its initial probability distribution \( \pi_0 \) that is updated by \( \pi_{0,T}^T \).

At each time step \( T > L_u \), the windowed observations \( O_{1:T}^T \) is given to the HMM with the updated parameter set \( \lambda_T^T \) based on (6).

The proposed windowing algorithm makes the length of observations used in the forward-backward algorithm bounded to a window length \( (L_u) \) smaller than the average full length sequence \( (L_f) \). Consequently, the computation cost for the forward-backward algorithm reduces \( L_f/L_u \) times.

III. Multi Modal HMM-Based Approach

In this work, TWM of a cutter in a CNC-milling machine is used as an illustrative example. While usually all the cutters are made from the same material in one dataset, the geometrical difference of cutters in the training and testing set could deviate and cause inaccurate predictions.

The idea of multi-modal approach is to generalize the prediction model by capturing more possible trends using combinations of distinct experiments for training, and then integrating the prediction results from all the single models in a weighted form. Here, three weighting schemes are proposed along with two switching strategies based on the computed weightages. Moreover, the windowed algorithm introduced in section II is extended for the multi-modal approach.

In the proposed multi-modal approach to improve performance of the single HMM-based approach called PSHMCO in [3], given a set of \( n \) experiments as the training set, \( N_{\text{mode}} \) PSHMCOs are trained. Each PSHMCO is trained on one distinct combination of the experiments to capture the common trend between the combined experiments (mode). It is noteworthy that the number of modes in this approach is less than or equal the number of possible combinations given \( n \) experiments, thus the number of modes is \( N_{\text{mode}} \leq 2^n - 1 \).

After training all the PSHMCOs as in [3], given a new set of observations as the testing set, we would like to assess the importance of each PSHMCO for the final prediction at each time step and reflect that as a weightage. Then, based on the switching strategy, the PSHMCOs’ weightages are used to predict the ultimate output at each time step. Figure 1 schematizes the prediction process in the mHMM approach.

The weightages of the modes can be computed in various ways with each emphasizing on the importance of different factors. The common ground between the three weighting schemes that will be introduced here is that, the weight of each PSHMCO is computed based on the historical observations (hindsight) and the current time, in order to assess the usefulness and similarity of that PSHMCO for prediction at that point of time. However, the factors that are considered to assess the weightages in these three schemes are different.

Two of the weighting schemes are solely based on the parametric models thus named parametric hindsight weighting schemes, namely, bounded hindsight and discounted hindsight. The third weighting scheme uses the parametric models to locate the positions in the corresponding training data used to train them (parametric phase), and then finds the most similar portion to the observations at hand given for prediction and ultimately assigns the similarity score to each model as the weightage (nonparametric phase). As the third weighting scheme has both parametric and nonparametric phases and furthermore its ultimate result is based on the nonparametric phase, it is named Semi-Nonparametric hindsight. Hindsight is the process of reviewing the historical data and calculating the probable health states at the past time steps [16].

A. Most Probable Health States

All the proposed weighting schemes use a dynamic programming method known as Viterbi algorithm. The Viterbi algorithm is used to find the most probable health states (Viterbi-path) taken at the past time steps within each model. As described in [17], the Viterbi-path can be formulated as follows

\[ \delta(i) = \max_{S_{1:t-1}} \{ P(S_{1:t-1}, S_t = H_i, O_{1:t} | \lambda) \}, \quad (7) \]

where \( \delta(i) \) is the highest probability obtained by a single path up to time \( t \) that ends in state \( H_i \). As indicated in [17], by induction a recursive formula can be obtained as follows

\[ \delta_{i+1}(j) = \max_{i_1=1, \ldots , m} \{ \delta(i) \times a_{ij} \} \times P(O_{t+1} | S_{t+1} = H_j, \lambda), \quad (8) \]

where \( a_{ij} \) is the transition probability to go from \( H_i \) to \( H_j \) in the HMM with the \( \lambda \) parameter set. Furthermore because of the specific structure assumed in [3] for PSHMCO based on the gradual process of wearing and degradation, from each health state \( H_i \) only two possible transitions exist i.e. either staying at

![Fig. 1. Schematization of multi-modal HMM-based approach.](image-url)
the same health state or going to the next health state $H_{i+1}$ till the cutter completely worn out and enter the last wearing stage $H_m$. Thus, (8) can be computationally simplified as follows

$$\delta_{t+1}(j) = \max \{ \delta_t(j-1) \times (1-p_{j-1}), \delta_t(j) \times p_{j} \} \times P(O_{t+1}|S_{t+1} = H_{j}, \lambda).$$

(9)

Finally, $V_{1:T}$ that is the sequence of health state indices taken through the Viterbi-path and is required for the weighting schemes, can be computed by performing backtracking on the stored matrix that has kept track of all the arguments which has maximized (8) for every $t$ and $j$. For more details on Viterbi algorithm refer to [17].

In the succeeding subsection, the three weighting schemes are described in details.

B. Weighting Schemes

As shown in Fig. 1, the observations (input features) are fed into all PSHMCOs in parallel. After finding the output of each PSHMCO based on (3), the next issue is to decide about the weightages of PSHMCOs outputs for polling (weighted averaging). Two methods, the parametric and semi-nonparametric hindsight, are adopted and compared for weighted polling. In all weighting schemes, the idea is to re-evaluate the likelihood of preceding time steps (hindsight) along with the current time step and use it as a metric to assess compatibility of each single model with the recent observations.

1) Parametric Hindsight: One of the motivations in using a multi-model approach is to use possible mix of trends captured by various models to improve prediction in new trends. These new trends may be relevant in parts to the modeled trends in multiple models. The aim is to give a higher weightage to the models which are more likely to have a similar trend with the new experiment at each time step. In this section, two schemes are suggested for computation of weightages in parametric form. In both schemes, focus is more on the latest time steps rather than long past ones. This allows a more dynamic switching of influence on the outputs between the multiple models. The basic idea of the proposed parametric hindsight weighting schemes are similar to the sliding window approaches used in the image and signal processing applications such as spatio-temporal visual tracker [18], smoothing as well as enhancing the temporal information. The two parametric schemes are introduced as follows.

Bounded Hindsight: In this scheme, given the observations up to the current time step $T$, the maximum likelihood of the probable health states for $\Delta - 1$ (bounded) preceding time steps (hindsight) along with the current time step in each PSHMCO are calculated. Summation of these $\Delta$ likelihoods, $w_T^i$, is used as an indicator of how relevant each PSHMCO is to the currently on-going experiment at that time. Hence $w_T^i$ is used as the corresponding weightage for the $i$th PSHMCO. An appropriate value for $\Delta$ can be found using cross-validation. The bounded weightage, $w_T^i$, for the $i$th HMM can be computed as follows

$$w_T^i = \sum_{t=T-\Delta+1}^{T} P(S_t = H_{V_t^i}, O_{1:T} | \lambda_i) = \sum_{t=T-\Delta+1}^{T} \gamma_t^i(V_t^i).$$

(10)

where $\Delta$ is the window size that the hindsight is bounded to, $V_t^i$ is the health state index at time $t$ of the Viterbi-path taken by the $i$th PSHMCO. $\lambda_i$ is the $i$th PSHMCO parameter set.

Discounted Hindsight: In this weighting scheme, similar to the bounded hindsight, given the observations up to the current time step $T$, the maximum likelihood of the Viterbi-path health states for preceding time steps (hindsight) in each PSHMCO are calculated. Instead of having a uniform summation on the likelihoods, first, every calculated likelihood is multiplied by a discount factor to increase the importance of the recent preceding time steps comparing to long past ones. Then, the summation of the discounted likelihoods is used as the corresponding weightage for each PSHMCO. In this work, the discount function is set to be a Gaussian distribution over the preceding time steps with its peak value at the current time step. An appropriate standard deviation for this function may be found based on cross-validation. The discounted weightage, $w_T^i$, for the $i$th PSHMCO can be computed as follows

$$w_T^i = \sum_{t=1}^{T} \tau_t \times P(S_t = H_{V_t^i}, O_{1:T} | \lambda_i) = \sum_{t=1}^{T} \tau_t \times \gamma_t^i(V_t^i),$$

$$\tau_t = N(t; T, \sigma^2)$$

(11)

where $\tau_t$ is the value of the Gaussian distribution with mean value of $T$ and standard deviation of $\sigma$ at $t$. Figure 3 illustrates the discounted hindsight weighting scheme.
2) Semi-Nonparametric Hindsight: Here, another weighting scheme is introduced to compute appropriate weightages required in the m²HMM approach to integrate the PSHMCOs outputs into its ultimate output. The basic idea of this weighting scheme is similar to the sequence alignment methods. Sequence alignment is a way of arranging sequences to identify regions of similarity originally developed in bio-informatics for aligning DNA, RNA or protein sequences [19].

Since the proposed scheme has both parametric and non-parametric phases and its final result comes from the non-parametric stage, this weighting scheme is named Semi-Nonparametric Hindsight (SNPH). Given the new observations and the HMMs parameter sets, the procedure of semi-nonparametric hindsight is schematized as follows.

- For each HMM, using its parameter set and the given observations find the most probable path of health states based on Viterbi-path algorithm. Find \( V_{1:T} \) based on \( \lambda_i \) and \( O_{1:T} \).
- Find the area \( A_q \) in the \( q \)th training sequence of the \( i \)th training combination (\( D^{i,q} \)) that corresponds to the most likely path taken in the \( i \)th HMM based on the reference segment of the Viterbi-path \( V_{1:T}^i \) that is denoted by \( V_{T-r+1:T}^i \). \( A_q \) can be identified as a set of (starting, ending) time index pairs as follows

\[
A_q(V_{1:T}^i, D^{i,q}, r) = \bigcup_{h \in \Psi} \{(t_{s}^{i,q,h}, t_{e}^{i,q,h})\},
\]

(12)

\[
\Psi = \{V_{T-r+1:T}^i, V_{T-r+1:T}^i + 1, \ldots, V_{T}^i\}
\]

where \( D^{i,q} = \{O^{i,q}, Y^{i,q}\} \) is the \( q \)th training combination of the \( i \)th training combination that includes observations \( O^{i,q} \) and their actual corresponding tool wear values \( Y^{i,q} \). \( \Psi \) is the set of health states’ indices that have been taken in the reference sequence based on the computed Viterbi-path, \( t_{s}^{i,q,h} \) and \( t_{e}^{i,q,h} \) are respectively the starting and ending time steps of the \( h \)th health state in \( D^{i,q} \).
- Align \( O^{i,q,h} = O^{i,q,h}_{r-1:T} \) and the \( R^h \) which is the corresponding observation segment in the reference sequence that its most likely health state is \( h \) based on Viterbi-path, and is defined in (13). In (13), the starting and ending time steps of the \( h \)th health state in the reference sequence are denoted by \( t_{s}^{h} \) and \( t_{e}^{h} \), respectively.
- Find the Aligned distance between the two matrices as in (14). Length of corresponding segments to the \( h \)th health state in the reference sequence (\( H \)) and training observation sequence \( O^{j,q} \) are denoted in (14) by \( l_{h} \) and \( l_{h}', \) respectively.
- Based on the aligned distance function \( G(., .) \) defined in (14), the total score function for \( q \)th sequence of the \( i \)th PSHMCO can be computed as follows

\[
Score(D^{i,q}, O_{1:T}, V_{1:T}^i, r) = \left[ \frac{1}{\sum_{h=V_{T-r+1:T}^i}^{V_{T}^i} G(O^{i,q,h}, R^h)} \right]^{-1} \sum_{h=V_{T-r+1:T}^i}^{V_{T}^i} \min\{l_{h}, l_{h}'\},
\]

(15)

where \( O_{1:T} \) is the newly given observation sequence from time step 1 up to current time \( T \) and \( V_{1:T}^i \) is the index sequence of the health states taken within the Viterbi-path based on the \( i \)th PSHMCO.
- Finally, maximum of the scores obtained from all included sequences in the \( i \)th PSHMCO training set as the maximum aligned similarity, would be adopted for the weightage of \( i \)th PSHMCO in the final outcome. Thus, the weightage of the \( i \)th PSHMCO can be computed as

\[
\omega_i = \max_q \{Score(D^{i,q}, O_{1:T}, V_{1:T}^i, r)\},
\]

(16)

Figure 4 illustrates the semi-nonparametric hindsight weighting scheme. For illustration purpose, in Fig. 4, it is assumed that \( V_{T-r+1:T}^i \neq V_{T}^j \) meaning that there are more than one distinct health index in the reference segment of the Viterbi-path.

After both parametric and semi-nonparametric hindsight schemes, the computed weightages can be used to either mix the outputs (soft-switching) or perform hard-switching between modes.

C. Switching Strategy

After assigning relevance weightages to all the modes based on any of the weighting schemes that has been used, two strategies are considered for integrating various modes’ outputs into the ultimate multi-modal output. These two strategies can be formulated as follows.
\[ R^h = \{ O_i^{t_h, t_h} | T - r + 1 \leq t \leq t_h, V^h_{i+1} \neq h, V^h_{i-1} \neq h, V^i = h \}. \]

\[
G(O^{i,q,h}, R^h) = \begin{cases}
\min_{i=1, \ldots, l_h} \left\{ \frac{1}{K} \sum_{k=1}^{K} \left( O^{i,q,h}_{k,j} - R^h_{k,j} \right)^2 \right\} & \text{for } t'_h \leq t_h \\
\min_{i=1, \ldots, l_h} \left\{ \frac{1}{K} \sum_{k=1}^{K} \left( O^{i,q,h}_{k,j} - R^h_{k,j+i-1} \right)^2 \right\} & \text{for } t'_h > t_h
\end{cases}
\]

- Soft-Switching,

\[
\hat{y}^{Soft}_T = \frac{\sum_{i=1}^{N_{mode}} w^i_T \times \hat{y}^i_T}{\sum_{i=1}^{N_{mode}} w^i_T}, \tag{17}
\]

where \( \hat{y}^{Soft}_T \) is the ultimate output for time step \( T \) using the soft-switching strategy, \( \hat{y}^i_T \) is the output of the \( i \)th PSHMCO and \( w^i_T \) is its corresponding weightage.

- Hard-Switching,

\[
\hat{y}^{Hard}_T = \arg \max \{ w^i_T \}, \quad i = 1, \ldots, N_{mode}. \tag{18}
\]

As (17) indicates, the soft-switching strategy uses the weighted average of all PSHMCOs as the ultimate output. On the other hand, the hard-switching strategy as shown in (18) assigns the output of the PSHMCO with the highest weightage as the ultimate output.

**D. Windowing Algorithm for \( m^2 \)HMMs**

Similar to the windowing algorithm introduced in Section II, by using the windowed observations instead of full observation sequences, the computational cost can also be reduced drastically in the proposed \( m^2 \)HMMs.

Here, instead of giving full observation sequence up to the current time \( T \), \( O_{1:T} \) to each PSHMCO, a windowed observation sequence \( O_{T-L_w+1:T} \) is given to all PSHMCOs. Thus, similar to the windowing algorithm introduced in Section II, the initial probability will be updated for all the HMMs based on their weightages as follows

\[
\pi^0_{0,T+1} = \frac{\sum_{i=1}^{N_{mode}} w^i_T \times \gamma^i_{1,i}}{\sum_{i=1}^{N_{mode}} w^i_T} = \frac{\sum_{i=1}^{N_{mode}} w^i_T \times P(S^i_{1}, O^i_{1:L_w} | \lambda^i_{1:T})}{\sum_{i=1}^{N_{mode}} w^i_T},
\]

\[
O^i_{1:L_w} = O_{T-L_w+1:T}, \quad S^i_{1} = S_{T-L_w+1}, \tag{19}
\]

where \( \pi^0_{0,T+1} \) is the updated initial probability that must be used for the prediction at next time step, \( T + 1 \). \( \gamma^i_{1,i} \) is the same as parameter set of the \( i \)th HMM, \( \lambda^i \), except that its initial probability distribution \( \pi^0_0 \) is updated by \( \pi^0_{0,T} \).

Similar to the windowing algorithm proposed in Section II, the proposed windowing algorithm for \( m^2 \)HMM approach will reduce the computation cost in the forward-backward algorithm by \( L_f / L_w \) times where \( L_f \) is the average length of the full observation sequence.

**IV. EXPERIMENTAL DATA & FEATURE EXTRACTION**

In this section, the experimental setup and feature extraction for a CNC-milling machine are discussed. The dataset used in this article is obtained through real-time force sensing. The experimental data comprises cutting process of 6 cutters, which are 07BX1, 31PN4, 09BX3, 18SC3, 34PT1 and 33PN6. The cutters are different from one another by the cutter geometry and coating, but all are 6mm Alignment-Tool carbide ball-nose end with three flutes.

In all the cutting processes, Inconel 718, which is used in Jet engines, is set as the work-piece material. During the cutting process, the upper face of the material is cut with horizontal lines from the top edge to the bottom edge. After 320 times of cutting, another cutter starts again at the top edge of the material for another round of experiment. The cutting face is 112.5 mm wide and 78 mm high. Therefore one cutter travels for 112.5 \times 320 = 36000 mm = 36 m. The cuts generated by the CNC-milling machine are 0.25 mm deep and the duration of one cut is approximately 4 seconds. Table I lists the operating conditions and the required components for the experimental setup.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spindle speed</td>
<td>10360 rpm</td>
</tr>
<tr>
<td>Feed rate</td>
<td>1555 m/min</td>
</tr>
<tr>
<td>Width of cuts</td>
<td>125 mm</td>
</tr>
<tr>
<td>Height of cuts</td>
<td>78mm</td>
</tr>
<tr>
<td>Depth of cuts</td>
<td>0.25 mm</td>
</tr>
<tr>
<td>Duration of cut</td>
<td>( \approx 4 ) s</td>
</tr>
<tr>
<td>Number of cuts per experiment</td>
<td>320</td>
</tr>
</tbody>
</table>

**Components**

- 6mm ball nose tungsten carbide cutters, i.e. 07BX1, 09BX3, 18SC3, 31PN4, 33PN6, and 34PT1.
- Inconel 718 workpiece
- Kistler 9265B Quartz 3-component dynamometer (force sensor)
- Kistler 5019A multichannel charge amplifier (force amplifier)
- NI-DAQ PCI 6250 M series
- LEICA MZ12.5 stereomicroscope
- Computer

To capture force signals in three directions, a three-channel dynamometer is mounted on the CNC-milling machine. Moreover, for training and testing purposes, the tool wear data is directly measured and collected using microscope in the
conducted experiments. Figure 5 schematizes the experimental setup used in this study.

In this work, based on the comparative feature selection study in [20], 13 statistical features that are shown to be the most salient features in the conducted study are extracted from the available force signals in 3 directions. Table II lists the extracted features and indicates the direction of the signals that they are extracted from.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Direction</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplitude Ratio</td>
<td>X, Y, Z</td>
<td></td>
</tr>
<tr>
<td>First Order Differencing</td>
<td>X and Y</td>
<td></td>
</tr>
<tr>
<td>Total Harmonic Power</td>
<td>X and Z</td>
<td></td>
</tr>
<tr>
<td>Maximum Force Level</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Total Amplitude of Cutting Force</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Kurtosis</td>
<td></td>
<td>Z</td>
</tr>
<tr>
<td>Average Force</td>
<td></td>
<td>Z</td>
</tr>
<tr>
<td>Skewness</td>
<td></td>
<td>Z</td>
</tr>
</tbody>
</table>

V. PRELIMINARY EXPERIMENTAL RESULTS

In this section, the results of cutter wear diagnostics in a CNC-milling machine obtained by the multi-modal HMM-based approach (m²HMM) with various weighting schemes are compared with the single HMM-based approach proposed in [3] called PSHMCO. The m²HMM is implemented with three weighting schemes and two switching strategies introduced in Section III.

All prediction models are trained using data collected from three experiments, which are conducted using 07BX1, 33PN6, and 34PT1 type cutters. The trained prediction models are then tested on the experimental data acquired from the three remaining distinct cutters 09BX3, 18SC3, and 31PN4. In this study, similar to [3], the number of health states is adopted to be 14 for all the HMMs. Although the number of health states in all HMMs is adopted to be 14, the parameter sets of these HMMs which are computed based on various combinations of training experiments would be different thus leading to various wearing regiments. Moreover, all possible combinations of the three training experiments are used to train different modes in the m²HMM approaches leading to 7 distinct modes. Based on the cross-validation, the hyper-parameters of the three weighting schemes which are the size of bounded hindsight window, the standard deviation of the discounted hindsight, and the size of the reference observation sequence in semi-nonparametric weighting scheme are adopted to be 10, 2.5, and 10, respectively. Figure 6 depicts the resultant relative weigntage values by the three weighting schemes on the three cutters used for testing.

Table III shows the prediction performance of PSHMCO compared with variants of multi HMM-based approach (considering the three suggested weighting schemes and the two switching strategies) in terms of mean squared error (MSE) and mean relative error (MRE).

In this experiment, at each time step (cut), the extracted features (observations) from the beginning of the experiment up to the current time step are given for diagnostics. It is noteworthy that in practice, the precision of a work-piece is ultimately determined by the flute with the maximum wearing value. Thus, during testing, the maximum estimated wearing value among the three flutes at each time step is used as the predicted outcome of the cutter wearing.

It can be seen from Table III that the total performance of m²HMM approach with parametric hindsight (either discounted or bounded) and soft-switching shows about 40% improvement in average precision and about 20% improvement in standard deviation. Furthermore, m²HMM-based approach with semi-nonparametric hindsight and soft-switching strategy (Soft m²HMM-SNPH) achieves the best average total performance. This suggests that semi-nonparametric weighting scheme can be an appropriate way of integrating outputs from various PSHMCOs. Moreover, m²HMM approach variants with hard-switching have been unable to outperform PSHMCO and their counter variants with soft-switching. Although, the results suggests that Soft m²HMM-SNPH outperforms PSHMCO and other variants of m²HMM, these preliminary results may not be convincing as they are reported only on one possible training-testing dataset partition. Thus, in order to assess the significance of improvements made by the m²HMM variants and see whether the Soft m²HMM-SNPH can outperform the other approaches in a statistically significant manner further investigations are conducted.

VI. FURTHER INVESTIGATIONS

In this section, the efficiency of the proposed m²HMM approach with various weighting schemes is further investigated to confirm whether it can statistically outperform PSHMCO. Moreover, performance of the two switching strategies, namely Soft- and Hard- switching, are compared and the effect of reference length in the SNPH weighting scheme is studied. Furthermore, the windowing algorithms proposed in sections II and III are examined to see whether the computational cost can be reduced while maintaining high prediction accuracy in comparison with the original PSHMCO and m²HMM approaches.

For the purpose of having enough samples to assess the statistical significance in performance differences, various trials
are generated by different partitionings of the whole acquired dataset from 6 cutters in section IV based on all possible combinations. Two scenarios are assumed in this section.

- **Easy Scenario**: in which the data from 5 cutters is used for training and then the prediction models are tested on the one remaining cutter. The training-testing (TT) ratio in this case is 5:1.

- **Difficult Scenario**: in which the data from 3 cutters is used for training and then the prediction models are tested on the 3 remaining cutters. TT ratio in this case is 3:3.

The number of possible combinations and thus various trials for the easy scenario is \( C_5^1 = 6 \) and for the difficult scenario is \( C_3^1 = 20 \). In the succeeding subsections the prediction models are trained, tested and compared based on the various trials in the two scenarios. Moreover, every possible combination of experiments in the training partitions are used to train one mode in the \( m^2 \)HMM approaches. Consequently, the \( N_{\text{mode}} \) of the \( m^2 \)HMM approaches in the easy scenario is 31 (\( 2^5 - 1 \)) and for the difficult scenario is 7 (\( 2^3 - 1 \)).

It is noteworthy that in all testing experiments the predictions are made from time step 20 to 320 for the sake of fairness in comparisons and reporting the results while various windowing lengths or hyper-parameter values are considered. In the succeeding subsections, if not indicated differently the values of hyper-parameters are adopted to be identical with the ones in the Section V.

### A. Switching Strategy: Hard Vs. Soft

As mentioned in section III, the resultant weightages from three proposed schemes can be used to perform switching based on two strategies i.e. Soft- and Hard- switching. Table IV and V show the results of the two strategies in the two scenarios based on various weighting schemes in terms of MSE and MRE, respectively. As it can be seen in both Tables, although \( m^2 \)HMM with hard-switching average performances are better than the PSHMCO, the \( m^2 \)HMM with Soft-switching outperforms all their Hard-switching counter-variants and the p-values in the difficult scenario indicate the significance of this matter. However, the p-values in the easy scenario indicate that there is not enough evidence to support or reject the idea in that scenario although the average performance of the soft-switching variants are better. Thus, in general it is recommended to use the soft-switching disregarding the weighting scheme that has been used in the

![Fig. 6](image-url) Resultant weightages for the three cutters using the three weighting schemes. Y-axes show the area value of each weightage and the X-axes show time steps. The weightages are normalized to sum up to one at each time step. The PSHMCO indices in the legend indicate which cutters’ data are included for each mode in a binary manner. The three indices correspond to 07BX1, 33PN6, and 34PT1 respectively from left to right.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Mean Square Error</th>
<th>Mean Relative Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>09BX3</td>
<td>18SC3</td>
</tr>
<tr>
<td>PSHMCO</td>
<td>725.54</td>
<td>30.68</td>
</tr>
<tr>
<td>Soft ( m^2 )HMM Bounded</td>
<td>522.33</td>
<td>148.08</td>
</tr>
<tr>
<td>Soft ( m^2 )HMM Discounted</td>
<td>514.80</td>
<td>163.94</td>
</tr>
<tr>
<td>Soft ( m^2 )HMM Semi-Nonparametric</td>
<td>337.16</td>
<td>156.73</td>
</tr>
<tr>
<td>Hard ( m^2 )HMM Bounded</td>
<td>486.84</td>
<td>167.57</td>
</tr>
<tr>
<td>Hard ( m^2 )HMM Discounted</td>
<td>487.18</td>
<td>188.43</td>
</tr>
<tr>
<td>Hard ( m^2 )HMM Semi-Nonparametric</td>
<td>557.54</td>
<td>303.62</td>
</tr>
</tbody>
</table>

multi-modal approach as it leads to higher average performance without additional computation cost.

In the succeeding subsections soft-switching is adopted for all the m^2HMM approaches.

### B. Overall Performance Comparison

Here, the PSHMCO and the m^2HMMs with various weighting schemes are trained and tested using the two scenario trials that each has partitioned the whole dataset into training and testing sets. Table VI shows the average performance of each approach in the two scenarios in terms of MSE and MRE. Moreover, the p-values from the pair-wise t-test performed between various approaches and PSHMCO as well as m^2HMM-SNPH are reported in Table VI.

From Table VI, it is suggested that all m^2HMM approaches disregarding their weighting schemes have outperformed PSHMCO in terms of both MSE and MRE in a statistically significant manner in both scenarios. Furthermore, the m^2HMM with SNPH weighting scheme significantly outperforms the rest in the difficult scenario. However, the p-values for the easy scenario indicate that there is not enough evidence to support or reject the hypothesis in that scenario.

### C. Full Vs. Windowed Observations

Here, the windowing algorithms proposed in Sections II and III are applied on PSHMCO and the m^2HMM approaches. The results are reported and compared with the original PSHMCO which uses full observations. Table VII shows the average performance of windowed version of all approaches in the two scenarios in terms of MSE and MRE. Pair-wise t-test is conducted between the resultant performances from the original PSHMCO and the windowed version of all mentioned approaches on all trials in the two scenarios. Given the window length $L_w$, the hyper-parameters of bounded and semi-nonparametric hindsight are set equal to $L_w$ and for the discounted hindsight the $\sigma$ is set to $L_w/4$.

As it can be understood from Table VII, the windowed version of the multi-modal approaches disregarding the weighting scheme have significantly outperformed the original PSHMCO. Furthermore, the windowed PSHMCO has achieved a slightly better performance than its original form. Figure 7 and 8 show the average performance of all approaches in the two scenarios in terms of MSE and MRE with respect to various window lengths. Based on Figure 7 and 8, the m^2HMM-SNPH outperforms all the other weighting schemes as well as PSHMCO in the two scenarios. Interestingly, results indicate that the windowing algorithm not only reduces the computational time, but also improves the average performance of PSHMCO by reducing the unnecessary connection to the long past observations if adopted appropriately.

Table VII shows the average computational time required to perform prediction using various approaches at each time step in the two given scenarios. Although the computational time required for m^2HMMs in the two scenarios are higher than PSHMCO, the table indicates feasibility of all approaches as all of them are performed in a fraction of a second. It is also shown that using the windowed observations the required computational time has drastically reduced.

### D. Reference Length Sensitivity Analysis

Here, the effect of reference length hyper-parameter on the performance of the m^2HMM with Semi-Nonparametric Hindsight (m^2HMM-SNPH) is studied. For this purpose, the average performance of the m^2HMM-SNPH approach is measured while varying its hyper-parameter value from 1 to 20 (shown in Fig. 9).

As it can be seen in Fig. 9, the average performance of m^2HMM-SNPH changes are small as its hyper-parameter value varies in the two cases, showing that m^2HMM overall average performance is robust and disregarding its hyper-parameter value better than PSHMCO. Interestingly, the average performance achieved by m^2HMM-SNPH (disregarding
OVERALL PERFORMANCE COMPARISON OF HARD AND SOFT SWITCHING STRATEGIES WITHIN ALL WEIGHTING SCHEMES ALONG WITH THE AVERAGE PSHMCO PERFORMANCE PROVIDED AS THE BENCHMARK. AVERAGE PERFORMANCES ARE COMPARED BASED ON MEAN SQUARED ERROR AND THE P-VALUE SHOWS THE SIGNIFICANCE OF PERFORMANCE DIFFERENCE BETWEEN THE TWO STRATEGIES USING THE SAME WEIGHTING SCHEME.

### TABLE IV

<table>
<thead>
<tr>
<th>Weighting Scheme</th>
<th>Easy (TT ratio 5:1)</th>
<th>Difficult (TT ratio 3:3)</th>
<th>P-Value (Hard vs. Soft)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hard-Switching</td>
<td>Soft-Switching</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean Squared Error</td>
<td>Mean Relative Error</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m²HMM Bounded</td>
<td>0.1603 ± 0.0649</td>
<td>0.1593 ± 0.0617</td>
<td>0.1920 ± 0.0904</td>
</tr>
<tr>
<td></td>
<td>0.0038</td>
<td>0.0037</td>
<td>0.0078</td>
</tr>
<tr>
<td>m²HMM Discounted</td>
<td>0.1631 ± 0.0592</td>
<td>0.1605 ± 0.0605</td>
<td>0.1910 ± 0.0893</td>
</tr>
<tr>
<td></td>
<td>0.0040</td>
<td>0.0063</td>
<td>0.0210</td>
</tr>
<tr>
<td>m²HMM Semi-Nonparametric</td>
<td>0.1617 ± 0.0707</td>
<td>0.1513 ± 0.0539</td>
<td>0.1899 ± 0.0895</td>
</tr>
<tr>
<td></td>
<td>0.0019</td>
<td>0.0071</td>
<td>0.0291</td>
</tr>
<tr>
<td>PSHMCO</td>
<td>0.1890 ± 0.0560</td>
<td>0.1980 ± 0.0673</td>
<td></td>
</tr>
</tbody>
</table>


### TABLE V

<table>
<thead>
<tr>
<th>Weighting Scheme</th>
<th>Easy (TT ratio 5:1)</th>
<th>Difficult (TT ratio 3:3)</th>
<th>P-Value (Hard vs. Soft)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hard-Switching</td>
<td>Soft-Switching</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean Squared Error</td>
<td>Mean Relative Error</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m²HMM Bounded</td>
<td>0.1593 ± 0.0617</td>
<td>0.1605 ± 0.0605</td>
<td>0.1910 ± 0.0893</td>
</tr>
<tr>
<td></td>
<td>0.0037</td>
<td>0.0063</td>
<td>0.0210</td>
</tr>
<tr>
<td>m²HMM Discounted</td>
<td>0.1593 ± 0.0617</td>
<td>0.1605 ± 0.0605</td>
<td>0.1910 ± 0.0893</td>
</tr>
<tr>
<td></td>
<td>0.0037</td>
<td>0.0063</td>
<td>0.0210</td>
</tr>
<tr>
<td>m²HMM Semi-Nonparametric</td>
<td>0.1513 ± 0.0539</td>
<td>0.1681 ± 0.0714</td>
<td>0.1899 ± 0.0895</td>
</tr>
<tr>
<td></td>
<td>0.0071</td>
<td>0.0291</td>
<td>0.0291</td>
</tr>
<tr>
<td>PSHMCO</td>
<td>0.1980 ± 0.0673</td>
<td>0.1980 ± 0.0673</td>
<td></td>
</tr>
</tbody>
</table>

TABLE VI

THE PAIR-WISE T-TEST RESULTS FROM COMPARING THE PERFORMANCE IN ALL THE MULTI-MODAL APPROACHES WITH VARIOUS WEIGHTING SCHEMES AND PSHMCO AS WELL AS M²HMM-SNPH WITH THE REST. EACH P-VALUE SHOWS THE SIGNIFICANCE OF PERFORMANCE DIFFERENCE BETWEEN THE TWO APPROACHES EITHER IN TERMS OF MSE OR MRE BASED ON PAIR-WISE T-TEST.

### TABLE VI

<table>
<thead>
<tr>
<th>Weighting Scheme</th>
<th>Easy (TT ratio 5:1)</th>
<th>Difficult (TT ratio 3:3)</th>
<th>P-Value (Hard vs. Soft)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Squared Error</td>
<td>Mean Relative Error</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m²HMM Bounded</td>
<td>0.1593 ± 0.0617</td>
<td>0.1605 ± 0.0605</td>
<td>0.1910 ± 0.0893</td>
</tr>
<tr>
<td></td>
<td>0.0037</td>
<td>0.0063</td>
<td>0.0210</td>
</tr>
<tr>
<td>m²HMM Discounted</td>
<td>0.1593 ± 0.0617</td>
<td>0.1605 ± 0.0605</td>
<td>0.1910 ± 0.0893</td>
</tr>
<tr>
<td></td>
<td>0.0037</td>
<td>0.0063</td>
<td>0.0210</td>
</tr>
<tr>
<td>m²HMM Semi-Nonparametric</td>
<td>0.1513 ± 0.0539</td>
<td>0.1681 ± 0.0714</td>
<td>0.1899 ± 0.0895</td>
</tr>
<tr>
<td></td>
<td>0.0071</td>
<td>0.0291</td>
<td>0.0291</td>
</tr>
<tr>
<td>PSHMCO</td>
<td>0.1980 ± 0.0673</td>
<td>0.1980 ± 0.0673</td>
<td></td>
</tr>
</tbody>
</table>

THE PERFORMANCE OF WINDOWED VERSION OF THE THREE MULTI-MODAL APPROACHES AND PSHMCO IN TERMS OF MSE AND MRE IN TWO SCENARIOS WHILE Lₜₚ = 13. THE AVERAGE ACCURACIES OF THE ORIGINAL PSHMCO WITH FULL OBSERVATIONS ARE ALSO GIVEN AS THE BASELINE. THE P-VALUES SHOW SIGNIFICANCE OF EACH WINDOWED APPROACH OUTPERFORMING FULL PSHMCO IN TERMS OF MSE IN EACH CASE BASED ON PAIR-WISE T-TEST.

### TABLE VII

<table>
<thead>
<tr>
<th>Windowed Approach</th>
<th>Easy (TT ratio 5:1)</th>
<th>Difficult (TT ratio 3:3)</th>
<th>P-Value (vs. Full PSHMCO)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>MRE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W m²HMM-BH</td>
<td>558.99 ± 453.04</td>
<td>0.1593 ± 0.0626</td>
<td>0.1709 ± 0.0839</td>
</tr>
<tr>
<td></td>
<td>0.0370</td>
<td>0.0049</td>
<td></td>
</tr>
<tr>
<td>W m²HMM-DH</td>
<td>555.04 ± 425.54</td>
<td>0.1605 ± 0.0608</td>
<td>0.1742 ± 0.0850</td>
</tr>
<tr>
<td></td>
<td>0.0163</td>
<td>0.0077</td>
<td></td>
</tr>
<tr>
<td>W m²HMM-SNPH</td>
<td>479.86 ± 467.83</td>
<td>0.1497 ± 0.0530</td>
<td>0.1663 ± 0.0770</td>
</tr>
<tr>
<td></td>
<td>0.0299</td>
<td>0.0006</td>
<td></td>
</tr>
<tr>
<td>W PSHMCO</td>
<td>636.16 ± 466.43</td>
<td>0.1784 ± 0.0624</td>
<td>0.1918 ± 0.0747</td>
</tr>
<tr>
<td></td>
<td>0.3156</td>
<td>0.0778</td>
<td></td>
</tr>
<tr>
<td>Full PSHMCO</td>
<td>720.78 ± 362.45</td>
<td>0.1890 ± 0.0560</td>
<td>0.1980 ± 0.0673</td>
</tr>
<tr>
<td></td>
<td>0.1890</td>
<td>0.0673</td>
<td></td>
</tr>
</tbody>
</table>
TABLE IX
ADVANTAGES AND DISADVANTAGES OF THE DEVELOPED APPROACH IN COMPARISON WITH PSHMCO.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
</table>
| PSHMCO   | - Correspondence between actual physical state and hidden state values.  
- Direct parameter estimation.  
- Capturing multiple trends by utilizing multiple modes.  
- Direct parameter estimation for each mode.  
- Correspondence between actual physical states and the hidden state values within each mode.  
- All m²HMM variants with soft switching outperform PSHMCO as well as their counterparts with hard-switching strategy.  
Bounded Hindsight (BH): lower computational complexity compared to SNPH.  
Discounted Hindsight (DH): lower computational complexity compared to SNPH.  
Semi-Nonparametric Hindsight (SNPH): achieves the highest accuracy compared to PSHMCO and all other m²HMM variants. | - Fixed state-duration distribution and single wear regiment.  
- Computationally is more expensive than PSHMCO Approach. However, the windowed variant of the approach reduces its computational cost drastically.  
- Additional hyper-parameter in all weighting schemes that can be set via cross-validation. |
| m²HMM    | - Direct parameter estimation.  
- Capturing multiple trends by utilizing multiple modes.  
- Correspondence between actual physical states and hidden state values within each mode.  
- All m²HMM variants with soft switching outperform PSHMCO as well as their counterparts with hard-switching strategy.  
Bounded Hindsight (BH): lower computational complexity compared to SNPH.  
Discounted Hindsight (DH): lower computational complexity compared to SNPH.  
Semi-Nonparametric Hindsight (SNPH): achieves the highest accuracy compared to PSHMCO and all other m²HMM variants. | - Additional hyper-parameter in all weighting schemes that can be set via cross-validation. |

Moreover, to reduce the computational burden, a windowing algorithm is proposed for both PSHMCO and m²HMM that dramatically increases the speed. Results indicate that if an appropriate window length is adopted, the windowed version of both approaches can reach the same level of performance as their original and in some cases even outperform them.

In future, the significance of the proposed m²HMM approach can be verified on various real industrial examples. Also further investigations will be carried out on selection of the efficient modes based on various combinations of the training data.

FIG. 9
Reference length sensitivity analysis in m²HMM-SNPH. The average performance of the m²HMM-SNPH is depicted for the two scenarios Easy (TT ratio 5:1) and Difficult (TT ratio 3:3) while its hyper-parameter value varies. The PSHMCO average performance lines in the two cases are also depicted for comparison.

its hyper-parameter value) in the difficult scenario is even outperforming the PSHMCO average performance obtained in the easy scenario. Finally, some of the advantages and disadvantages of the developed approaches are summarized in Table IX.

VII. CONCLUSION
A multi-modal HMM-based approach, which is an extension of existing physically segmented HMM with continuous output (PSHMCO) approach, is proposed for tool condition monitoring. Two switching strategies are introduced to combine the outputs from various modes into one. The relevance of each mode at each time is assessed based on the given observation at that time and its precedings. Three weighting schemes i.e. bounded-, discounted- and semi-nonparametric hindsight (SNPH) are proposed which quantify the relevance of modes as weightages to be used in switching strategies.

As an illustrative example, the proposed multi-modal approach is applied to cutter wear monitoring in the CNC-milling machine. Based on the experimental results, the m²HMM approach with soft switching (disregarding its weighting scheme) outperforms the PSHMCO significantly. Furthermore, the m²HMM with semi-nonparametric weighting and soft switching proves to be the best among all combinations of switching strategies and weighting schemes.

REFERENCES
Omid Geramifard (S’11-M’13) received the Bachelor’s degree in computer engineering from the Isfahan University of Technology, Isfahan, Iran, in 2008, and Ph.D. degree in Machine Learning from National University of Singapore, Singapore, in 2013. After his Ph.D. studies, he has joined Singapore Institute of Manufacturing Technology, where he is currently working as a research scientist. His research interests lie primarily in machine learning and its applications in condition-based maintenance, as well as diagnostics and prognostics in the machinery systems. Dr. Geramifard was awarded Singapore International Graduate Award (SINGA) for Ph.D. studies from 2009 to 2013.

Jian-Xin Xu (SM’98-F’12) received the Bachelor’s degree from Zhejiang University, Hangzhou, China, in 1982, and the Master’s and Ph.D. degrees from the University of Tokyo, Tokyo, Japan, in 1986 and 1989, respectively, all in electrical engineering. He was with Hitachi Research Laboratory as a Research Engineer, Ohio State University as a Visiting Scholar, and Yale University as a Visiting Research Fellow. In 1991, he joined the National University of Singapore, Singapore, where he is currently a Professor in the Department of Electrical and Computer Engineering. He is the author or coauthor of more than 400 peer reviewed journal and conference papers and five edited books on system modeling and control. His research interests include system learning, control, and applications to real-time systems.

Jun-Hong Zhou (M’08) received the B.Eng. degree from the Automation Department, Tsinghua University, Beijing, China, in 1987 and the M.Eng. degree from ECE department, National University of Singapore in 1995, and Ph.D. degree from Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore in 2012. Since 1996, she worked as a Research Fellow at the Singapore Institute of Manufacturing Technology (SIMTech), where she is currently, a principal researcher and the initiative lead in the area of maximizing overall equipment effectiveness. Her current research interests include equipment health monitoring, intelligent condition-based maintenance, fault diagnosis and failure prognosis, multimodal sensing and feature extraction, performance and serviceability optimization, artificial intelligence, and statistical analysis.

Xiang Li (M’09) received the B.Eng. and M.Eng. degrees from Northeastern University, Shenyang, China, in 1982 and 1987, respectively, and the Ph.D. degree from the Nanyang Technological University, Singapore, in 2000. She taught at Northeastern University from 1982 to 1987. From 1987 to 1989, she did research work as a Visiting Research Fellow at Tohoku University, Tohoku, Japan. She joined the Singapore Institute of Manufacturing Technology (SIMTech), Singapore, in 1992, where she is currently a Senior Scientist and team leader of prognostic health management. She has more than 20 years of experience in research on artificial intelligence, data mining, machine learning, and statistical analysis such as neural networks, fuzzy logic systems, unsupervised data clustering, and regression modeling.