A Physically Segmented Hidden Markov Model Approach For Continuous Tool Condition Monitoring: Diagnostics & Prognostics

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

Abstract—In this paper, a temporal probabilistic approach based on hidden Markov model, named, physically segmented hidden Markov model with continuous output, is introduced for continuous tool condition monitoring in machinery systems. The proposed approach has the advantage of providing an explicit relationship between the actual health states and the hidden state values. The provided relationship is further exploited for formulation and parameter estimation in the proposed approach. The introduced approach is tested for continuous tool wear prediction in a CNC-milling machine and compared with two well-established neural network approaches, namely, multi-layer perceptron and Elman network. In the experimental study, the prediction results are provided and compared after adopting appropriate hyper-parameter values for all the approaches by cross-validation. Based on the experimental results, physically segmented hidden Markov Model approach outperforms the neural network approaches. Moreover, the prognosis ability of the proposed approach is studied.

Index Terms—Tool Condition Monitoring, Hidden Markov Model, Feature Selection, Diagnostics, Prognostics.

I. INTRODUCTION

TooL Condition Monitoring (TCM) has become one of the main industrial challenges in the last decade. TCM reduces the amount of unnecessary downtime for maintenance purposes, and consequently reduces the cost of maintenance [1–5]. Moreover, TCM improves the quality and precision of the product.

The idea of continuous tool condition monitoring is to monitor the health condition of the tool at each time step in terms of a continuous metric based on the available input data. In other words, instead of setting thresholds and differentiating distinct health states as various (ordinal) classes, we would like to ultimately monitor the health state of the tool in a continuous form. This task allows us to have smoother decision making systems in the condition based maintenance and it can incorporate different quality thresholds for different applications using the same condition based maintenance system e.g. to guarantee different qualities in various products. The input data in this task, is a set of selected features that are extracted from non-intrusively sensed and captured signals. Signals such as force, vibration and acoustic emission can be captured and recorded using various sensors mounted on the machinery systems.

In this work, continuous TCM in a CNC-milling machine is used as an illustrative example. In a CNC-milling machine, the extracted features are used as inputs to predict the continuous wear metric of the cutter [4–6]. This prediction process is commonly dichotomized into two tasks, namely, diagnosis and prognosis [4], [5], [7–10]. Since a perfect physical model is not available in practice for the tool wearing, researchers focus on developing diagnosis and prognosis approaches based on the historical data. A survey on these approaches can be found in [11], [12].

In [13], trend projection models are used, in which model parameters can be calculated but may overfit the past degradation patterns. Fuzzy inference system (FIS)-based approaches are also extensively used in TCM [14–17], which in general require a priori knowledge to be available when determining the rules and membership functions. The strategy exploiting fuzzy and neuro-fuzzy tools such as ANFIS, have been applied to TCM uses [18–20], which are data-driven, hence can be regarded as special classes of Neural Network methods.

Artificial neural network (NN) is one of the most commonly used approaches in this domain. In [10], [21–28], NNs are used in a time series prediction manner providing nonlinear projection without the need for prior knowledge.

Hidden Markov models (HMM) and hidden semi-Markov models (HSMM) are used [29–35] to distinguish various wearing stages or machinery fault types. In all the existing HMM and HSMM-based approaches, the wearing data is discretized into several stages and then multiple HMMs are used to distinguish between those stages. Each HMM or HSMM is assigned to recognize one specific stage or fault. Training HMMs and HSMMs using expectation-maximization method, however is essentially a black-box approach, which does not provide explicit relationship between the wearing value and the hidden state values in the trained HMMs or HSMMs.

In this paper, a temporal probabilistic approach based on HMM, named physically segmented hidden Markov model with continuous output (PS-HMCO), is proposed to tackle the problem of continuous health assessment of cutters in a CNC-milling machine. The proposed approach depicts the
explicit relationship between the actual physical states and the hidden state values. Furthermore, the provided relationship is exploited for formulation and parameter estimation in PS-HMCO. In addition, the suggested approach ultimately predicts the real-valued health state metric (tool wear) instead of discrete states or stages.

This paper is organized as follows. In Section II, the proposed approach, PS-HMCO, is introduced. Diagnostics and prognostics procedures are described in Section III. Section IV provides information about the experimental setup as well as the extracted and selected features from the acquired signals in the experiments. In Section V, performance of PS-HMCO is compared with MLP and Elman network. The paper is concluded in Section VI.

II. PHYSICALLY SEGMENTED HIDDEN MARKOV MODEL WITH CONTINUOUS OUTPUT

As the name implies, this approach is based on hidden Markov model. Contrary to the conventional use of HMMs which is in classification, here HMM is applied to a continuous problem (regression). In this approach, 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 assuming to have a complete training set. Finally, the state estimation is described for different points in time.

Hidden Markov model is a simple dynamic Bayesian network [36]. This model has only one discrete hidden state variable, and a set of discrete or continuous observation nodes. A first-order temporal Markov model is characterized by the assumption that,

\[ P(S_t = v_i | S_{t-1}, S_{t-2}, \ldots, S_1) = P(S_t = v_i | S_{t-1}), \quad (1) \]

where \( P(\cdot | \cdot) \) is the conditional probability, \( S_t \) is the hidden state variable at time \( t \) and \( v_i \) is the \( i \)th hidden state value. Equation (1) indicates that the conditional probability of any current state, given knowledge of all previous states, is the same as the conditional probability of the current state, given its previous state only [37].

In order to use the HMM in the PS-HMCO approach, firstly, the output space will be discretized into several hidden state values or health states as shown in Fig. 1. 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. In this and the following sections, the PS-HMCO approach for diagnostics and prognostics in TCM is introduced in details.

A. Discretization & Formulation

Using uniform discretization method on the continuous health metric (tool wear), continuous wearing values can be discretized into \( m \) classes or state values with real-valued labels, \{\( H_1, \ldots, H_m \)\}. These state values correspond to \( m \) ordinal wearing stages. As shown in Fig. 1, after discretization and segmenting the tool wear values in the training set, segments are labeled ordinally from \( H_1 \) to \( H_m \) indicating explicit correspondence between the health states (hidden state values) in the HMM and the actual physical wearing metric.

An appropriate value for the number of health states can be adopted based on cross-validation results. Since wearing is a gradual process and sampling rate is relatively high comparing to the number of possible state values, at each time step only two possible transitions are available (excluding the last state which is modeled as an attraction point). The two possible transitions given that \( S_{t-1} = H_i \) are either staying at the same condition \( H_i \) with probability \( p_i \) or going to the next degradation stage \( H_{i+1} \) with probability \( 1 - p_i \), as shown in Fig. 1. Therefore, the transition probability matrix, \( A \), for the HMM used in the TCM can be formulated as,

\[
A = [a_{ij}]_{m \times m} = \begin{bmatrix}
 p_1 & 1 - p_1 & 0 & 0 & \ldots & 0 \\
 0 & p_2 & 1 - p_2 & 0 & \ldots & 0 \\
 \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\
 0 & 0 & 0 & \ldots & p_{m-1} & 1 - p_{m-1} \\
 0 & 0 & 0 & \ldots & 0 & 1
\end{bmatrix}, \quad (2)
\]

where the element at \( i \)th row and \( j \)th column, \( a_{ij} \), is the transition probability of going from \( j \)th health state to \( i \)th health state. \( m \) is the total number of state values that the hidden state can take and \( p_i \) is the probability of self-transition in the \( i \)th health state.

B. Parameter Estimation

Assuming stationarity, the only parameters to be identified are, the initial health state probabilities (prior probabilities),
transition probability matrix in (2) and emission probabilities that connect the hidden states to the observations (input features). These parameters must be estimated for the HMM given the training data. The parameter estimation can be done using either maximum likelihood (ML) or maximum a posteriori (MAP) method. Here, Maximum Likelihood learning method is adopted. Gradient ascent and expectation-maximization (Baum-Welch) are the two conventional algorithms to estimate the parameters based on ML [36]. However, in this work since the data is complete (without missing information) and taking the advantage of explicit relationship between the given actual physical states and the hidden state values, there is no need to use either of the two mentioned algorithms. Using the ML method, parameters can be directly estimated from the discretized health states based on the measured tool wear (as depicted in Fig. 1), and the input feature sequences that are extracted from the dataset.

The ML method calculates the parameters that maximize the likelihood probability of the training dataset. Therefore, in order to find the parameters, the joint probability distribution of the training dataset (likelihood probability) must be derived and then be maximized. Since the $N$ experimental sequences included as the training set $\{D_1, \ldots, D_N\}$ are independent of each other, the joint probability distribution for HMM can be written as

$$P(D_1, D_2, \ldots, D_N|\lambda) = \prod_{j=1}^{N} P(D_j|\lambda),$$

(3)

where $D_j$ is the experimental data sequence collected from the $j$th experiment and $\lambda$ is the set of HMM parameters. In (3) the joint probability distribution for the $j$th experiment given the parameters, $P(D_j|\lambda)$, can be computed as

$$P(D_j|\lambda) = P(S_0^j, O_1^j, \ldots, O_{T_j}^j|\lambda)$$

$$= \prod_{i=1}^{T_j} P(S_i^j|S_{i-1}^j)P(O_i^j|S_i^j),$$

(4)

where $D_j$ includes extracted features $O_i^j$ and tool wear sequence $S_i^j$, and $T_j$ is the length of the $j$th experimental sequence. Assuming that the emission probabilities $P(O_i^j|S_i^j)$ are Gaussian distributions [38] and considering the transition matrix in (2), joint probability distribution of the $j$th experimental data given in (4) can be rewritten as

$$P(D_j|\lambda) = \pi_0 \times \prod_{i=1}^{m-1} P(S_i^j|S_{i-1}^j)$$

$$\times \prod_{i=t_i^j}^{m-1} \frac{1}{(2\pi)^{\frac{\chi}{2}}|\Sigma_i|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(O_i^j-\mu_i)^T\Sigma_i^{-1}(O_i^j-\mu_i)\right),$$

(5)

where $\lambda = \{\pi_0, \mu_1, \ldots, \mu_m, \Sigma_1, \ldots, \Sigma_m\}$, $\pi_0$ is the prior probability of the initial 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 $S_i^j = H_i$. $\chi$ is the dimension of observation vector or in other words the number of features being used for TCM. $t_i^j$ is the starting time step of the $i$th health state and $k_i^j$ is the number of samples belonging to the $i$th health state, both in the $j$th experimental data. From (5), the log likelihood for the implemented HMM with real-valued observations can be computed as

$$L = \sum_{j=1}^{N} L_j = \sum_{j=1}^{N} \log(P(D_j|\lambda)),$$

$$L_j = \log(P(D_j|\lambda)) = \log(\pi_0) - \frac{\chi T_j}{2} \log(2\pi)$$

$$+ \sum_{i=1}^{m-1} (k_i^j - 1) \log(p_i) + \log(1 - p_i) - \frac{k_i^j}{2}(|\Sigma_i|)$$

$$- \sum_{i=1}^{m} \sum_{t=t_i^j}^{t_i^j+1-1} \frac{1}{2}(O_t^j - \mu_i)^T \Sigma_i^{-1}(O_t^j - \mu_i).$$

(6)

To find the parameter set, which maximizes the log likelihood in (6), the partial derivatives of the log likelihood are set to zero. Consequently, the parameters may be computed as follows,

$$\frac{\partial L}{\partial p_i} = 0 \Rightarrow p_i = 1 - \frac{N}{\sum_{j=1}^{N} k_i^j},$$

(7)

$$\frac{\partial L}{\partial \mu_i} = 0 \Rightarrow \mu_i = \frac{\sum_{j=1}^{N} O_t^j}{\sum_{j=1}^{N} k_i^j},$$

(8)

$$\frac{\partial L}{\partial \Sigma_i} = 0 \Rightarrow \Sigma_i = \frac{\sum_{j=1}^{N} (O_t^j - \mu_i)^T (O_t^j - \mu_i)}{\sum_{j=1}^{N} k_i^j}.$$  

(9)

Using the formulas in (7), (8) and (9), all of the required parameters can be estimated. The initial state probability $\pi_0$ is defined as the uniform probability distribution in case there are no prior knowledge about the general initial state value.

C. State Estimation

In order to estimate the state values of the state variables in the implemented HMM at each time step based on the observations, two variables are further defined. $\gamma_t$ is the joint probability of observing all the input features up to the current time $T$ while being at the $i$th health state at time $t$ where $t < T$. $\gamma_t$ is used to find the value of the state variable at $t$ for diagnosis purpose based on all the observations from
beginning of an experiment up to the current time, $T$. From [39], $\gamma_T$ can be defined and computed as follows,
\begin{equation}
\gamma_T(i) \triangleq P(S_T = H_i, O_{1:T}) = P(O_{1:T}, S_T = H_i) = \prod_{t=1}^{T} P(\xi_t | \xi_{t-1}) \times P(\xi_1) \times P(O_1 | S_1) \times P(S_T = H_i | S_1) \times P(O_{t+1:T} | S_T = H_i)
\end{equation}
\begin{equation}
\gamma_T(i) = \alpha_T(i) \times \beta_T(i),
\end{equation}
where $\alpha_T(i)$ and $\beta_T(i)$ are the forward and backward variables which can be computed using the forward-backward algorithm [39].

In order to predict the state value of the state variable in future based on the available observations up to the current time $T$, another variable, $\xi_{T'}$, is defined. Similar to $\gamma_T$, $\xi_{T'}$ is the joint probability of observing all the input features from beginning of an experiment up to the current time, $T$, but being at the $i$th health state at time $t'$ in future ($t' > T$). $\xi_{T'}$ is defined and computed recursively based on transition probabilities and $\gamma_T$ as follows,
\begin{equation}
\xi_{T'}(i) \triangleq P(S_{T'} = H_i, O_{1:T}), \quad t' > T
\end{equation}
\begin{equation}
\xi_{T'}(i) = \sum_{S_{T'-1}} P(S_{T'-1}, O_{1:T}) \times P(S_{T'} = H_i | S_{T'-1}) \times P(O_{1:T} | S_1) \times P(S_T = H_i | S_1)
\end{equation}
\begin{equation}
\xi_{T'}(i) = \xi_{T'-1}(i) p_i + (1 - p_i) \xi_{T'-1}(i) - 1
\end{equation}
with initial condition $\xi_T(i) = \gamma_T(i)$.

In the succeeding section, $\gamma_T$ and $\xi_{T'}$ are used for diagnostics and prognostics.

III. DIAGNOSTICS & PROGNOSTICS

In this section, the TCM approach (diagnostics & prognostics) is provided in a probabilistic manner based on the state estimation variables defined in the previous section and the real valued labels determined in the discretization phase.

Diagnosis is the task of predicting the health state at the current time $T$ given all the observations from time step 1 up to $T$ [12]. In the realm of Bayesian Networks, this task is called filtering or monitoring and it can be written in a probabilistic form as follows,
\begin{equation}
P(S_T = H_i | O_{1:T}, \lambda) = \frac{P(O_{1:T}, S_T = H_i)}{P(O_{1:T})} = \frac{\gamma_T(i)}{P(O_{1:T})}
\end{equation}
\begin{equation}
P(S_T = H_i | O_{1:T}, \lambda) = \frac{\gamma_T(i)}{P(O_{1:T})} \sum_{i=1}^{m} \gamma_T(i)
\end{equation}
\begin{equation}
P(O_{1:T}) = \sum_{i=1}^{m} \gamma_T(i).
\end{equation}

Hence, based on (10) and (13), (12) can be rewritten as
\begin{equation}
P(S_T = H_i | O_{1:T}, \lambda) = \frac{\gamma_T(i)}{\sum_{i=1}^{m} \gamma_T(i)},
\end{equation}
where $\lambda$ is again included to indicate that calculations are based on the parameter set of the trained HMM and $P(O_{1:T})$ is a normalizing factor replaced by $\sum_{i=1}^{m} \xi_{i'}(i)$ similar to (13). In the end, the continuous output of the model, $C_{t'}$, which corresponds to the expected amount of tool wear at time step $t'$ in future, can be computed based on (16) as follows,

$$C_{t'} = \sum_{i=1}^{m} P(S_{t'} = H_i | O_{1:T}, \lambda) \times H_i, \quad (17)$$

IV. DATASET & FEATURES

The experimental data is obtained through real-time sensing on a CNC-milling machine using a force sensor and a vibration sensor both with 3 channels for different directions and an AE sensor. The data comprises cutting process of 6 cutters which are 07BX1, 09BX3, 18SC3, 31PN4, 33PN6 and 34PT1. The cutters are different with one another in terms of cutter geometry and coating but they are all 6mm alignment-tool carbide ball-nose end with three flutes.

In the conducted experiment, a röders TEC vertical milling machine is used as the testbed. For all the cutting processes, Inconel 718, which is used in Jet engines, is adopted 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 each cut, tool snapshots were taken to measure the amount of tool wear. An LECIA MZ12.5 high performance stereomicroscope is used to measure the tool wear of the cutting tool. After 320 cutting times, another cutter process will start at the top edge of the material. Table I shows the operating condition parameters of the CNC-milling machine and the experimental components. According to Table I, each cutter will travel for $112.5 \times 320 = 36000 \text{ mm} = 36 \text{ m}$. Fig. 3 shows 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>1.555 m/min</td>
</tr>
<tr>
<td>Width of cuts</td>
<td>0.125 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>$\approx 4s$</td>
</tr>
<tr>
<td>Number of cuts per experiment</td>
<td>320</td>
</tr>
</tbody>
</table>

**Components**

- röders TEC vertical milling machine
- 6mm ball nose tungsten carbide cutters
- Inconel 718 workpiece
- Kistler 8152B211 Piezotron® (AE sensor)
- Kistler 8636C PiezoBEAM® accelerometers (vibration sensor)
- Kistler 9265B Quartz 3-component dynamometer (force sensor)
- Kistler 5019A multichannel charge amplifier (force amplifier)
- NI-DAQ PCI 6250 M series
- LECIA MZ12.5 stereomicroscope
- Computer

As it is depicted in Fig. 3, the tool wear is measured after each cut and stored on the computer along with the sensed signals which are captured using LabView® software running on the computer. After data acquisition, these signals are used for feature extraction and selection using FDR.

### A. Statistical Features

16 statistical features are extracted from force signal available in each direction (X, Y and Z), resulting in 48 features available in total as a 3-channel dynamometer is mounted on the CNC-milling machine. A list of these features is shown in Table II.

### B. Wavelet Features

Signals that have been captured using the sensors mounted on the milling machine have non-stationary characteristics, therefore wavelet or multi-resolution approaches are used for feature extraction along with the extracted statistical features. In this work, Daubechies wavelet 8 (DB8) is applied to three force signals, and discrete Meyer wavelet is used for three vibrations and AE signals, all wavelets are with 5 decomposition levels. Hence, $62 \times 3 + 2^4 = 62$ coefficients are acquired for each signal, summing up to 434 for all the 7 signals. The average energy of these coefficients are used as the extracted features. According to [29], the average energy can be written as,

$$E_j = \frac{1}{N_j} \sum_{k=1}^{N_j} |d_{j,k}^n(t)|^2$$

where $j$ is the scale, $d_{j,k}^n(t)$ is the wavelet packet coefficient of the signal, $N_j$ is the number of coefficients at each scale and $t$ is the discrete time.

### C. Feature Selection

Constructing the prediction model using only a selected subset of features prevents unnecessary complexity in the model and consequently improves the prediction results. In addition, extracting only the selected features from the newly conducted experiments greatly reduces the feature extraction computation cost for online applications. In this study, the feature selection is performed offline. The features are determined based on

![Experimental setup](image-url)
FDR values on the training set data and are not changed in the simulated online study. In the simulated online monitoring, only the selected features are extracted in order to save the computational effort and also for fast data processing.

The idea of feature selection in a classification domain is to find a subset of features that explicitly discriminate the classes based on the training set. Hence, the features to be selected must have similar values for the samples in one class and distinct values for the samples from different classes. Fisher’s Discriminant Ratio (FDR) shown in (19) is such a metric that shows how discriminative a feature is. It is a ratio of scatter between ($S_b$) and the scatter within ($S_w$). A modified version of FDR introduced in [29] is as follows,

$$FDR(f_i) = \frac{S_b^{f_i}}{S_w^{f_i}} = \frac{\sum_{k=1}^{K} \sum_{j=1}^{K} (\mu_y^{f_i} - \mu_f^{f_i})^2}{\sum_{k=1}^{K} S_{w_k}^{f_i}}, \tag{19}$$

where $K$ is the number of classes, $f_i$ is the $i$th feature (element) in the observation vector $O = [f_1, f_2, \ldots, f_K]^T,$ $\mu_y^{f_i}$ is the mean value of $f_i$ in the $k$th class and $S_{w_k}^{f_i}$ is the scatter within (variance) the $k$th class measured for $f_i$. In order to use the FDR criterion for feature selection in the continuous TCM, samples must be clustered based on their output values. One of the conventional clustering algorithms is called Gaussian Mixture Model (GMM) [40], [41]. In this work, GMM is used for clustering of the samples based on their corresponding outputs. Various number of Gaussian functions are explored for this purpose and finally based on Bayesian information criterion, 3 Gaussian functions are used in the adopted GMM.

After applying the GMM to the output set, the posterior probability $p(m_k|y^j_i)$ can be computed for all samples in the training set, where $p(m_k|y^j_i)$ indicates the probability that the actual tool wear at time $t$ in the $j$th experiment ($y^j_i$), is generated by the $k$th Gaussian function ($m_k$).

After finding the posterior probabilities, $\mu_y^{f_i}$ and $S_{w_k}^{f_i}$ can be computed in a weighted form as follows,

$$\mu_y^{f_i} = \frac{\sum_{j=1}^{T_j} \sum_{i=1}^{T_t} p(m_k|y^j_i) \times f_i^j(t)}{\sum_{j=1}^{T_j} \sum_{i=1}^{T_t} p(m_k|y^j_i)}, \tag{20}$$

$$S_{w_k}^{f_i} = \frac{\sum_{j=1}^{T_j} \sum_{i=1}^{T_t} [p(m_k|y^j_i) \times f_i^j(t) - \mu_y^{f_i}]^2}{\sum_{j=1}^{T_j} \sum_{i=1}^{T_t} p(m_k|y^j_i)} \tag{21},$$

where $f_i^j(t)$ is the value of the $i$th feature at time $t$ in the $j$th experimental sequence.

Then, the computed means and scatter within measures can be used in (18) to find FDR for $i$th feature. Fig. 4 shows the FDR values after being sorted in a descending manner.

The FDR value indicates how discriminant each feature is, hence it can be used to rank the features. As it can be seen in Fig. 4, there is a knee in the curve at $x = 38$. This knee point is used as a rule of thumb to chose the number of features to be selected. Therefore, 38 features with the highest FDR value are chosen for the prediction model.

Table III shows shares of extracted features from each signal in the set of selected features.

### TABLE II

<table>
<thead>
<tr>
<th>No.</th>
<th>Feature</th>
<th>No.</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Residual Error</td>
<td>9</td>
<td>Sum of the Squares of Residual Errors</td>
</tr>
<tr>
<td>2</td>
<td>First Order Difference</td>
<td>10</td>
<td>Peak Rate of Cutting Forces</td>
</tr>
<tr>
<td>3</td>
<td>Second Order Difference</td>
<td>11</td>
<td>Total Harmonic Power</td>
</tr>
<tr>
<td>4</td>
<td>Maximum Force Level</td>
<td>12</td>
<td>Average Force</td>
</tr>
<tr>
<td>5</td>
<td>Total Amplitude of Cutting Force</td>
<td>13</td>
<td>Variable Force</td>
</tr>
<tr>
<td>6</td>
<td>Combined Incremental Force changes</td>
<td>14</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>7</td>
<td>Amplitude Ratio</td>
<td>15</td>
<td>Skewness</td>
</tr>
<tr>
<td>8</td>
<td>Standard Deviation of the Force Components</td>
<td>16</td>
<td>Kurtosis</td>
</tr>
</tbody>
</table>

Table III

<table>
<thead>
<tr>
<th>Signal</th>
<th>$F_x$</th>
<th>$F_y$</th>
<th>$F_z$</th>
<th>$V_x$</th>
<th>$V_y$</th>
<th>$V_z$</th>
<th>AE</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical Features</td>
<td>6</td>
<td>2</td>
<td>6</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>14</td>
</tr>
<tr>
<td>Wavelet Features</td>
<td>11</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>Total Share</td>
<td>17</td>
<td>7</td>
<td>11</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>38</td>
</tr>
</tbody>
</table>

As it can be understood from Table III, features extracted from the force signals are majorly selected as the most dis-

![Fig. 4. FDR values of features sorted in a descending manner.](image-url)
A. Determination of Hyper-parameters

Two modes can be considered for cross-validation in case of condition monitoring for the cutter with multiple flutes i.e. leave one flute out and leave one experiment out. In this work, since the training set is limited to 3 experiments, leave one flute out is conducted. All models are trained on the data from two flutes out of three flutes, and then tested on the data from the excluded flute, all from the training experiments. Different hyper-parameter values for the three models are tested and the results are provided in Fig. 5 to Fig. 7.

As shown in Fig. 5, the MLP structure $10 - 4 - 1$, where 10 and 4 are the number of hidden neurons in the first and second hidden layers respectively, leads to the best cross-validation MSE. Hence, the same structure for MLP is adopted in the succeeding case. Furthermore, in order to prevent over-fitting problem in MLP, Bayesian regulation back-propagation [43], [44] is used as the learning algorithm for the implemented MLP.

Fig. 6 indicates that the Elman network with structure $20 - 3 - 1$ with two hidden layers leads to the best performance in cross-validation comparing to other structures. Hence, the Elman network with structure $20 - 3 - 1$ is adopted to be used in the diagnostics case.

From Fig. 7, it can be seen that 14 may be used as an appropriate number of health states in the PS-HMCO for the given dataset. Hence, the number of hidden state values (health states) is set to 14 in the succeeding cases.

B. Diagnostic Results

Now we compare diagnosis accuracy of the three approaches on the testing set using the implemented programs for each approach in MATLAB2010b. All models are trained using the training set (i.e. data collected from cutters 07BX1, 31PN4 and 34PT1) and then tested for diagnostics on the testing set (i.e. data collected from cutters 09BX3, 18SC3 and 33PN6). In the diagnosis process, at each time step (cut), the input data from the beginning of the diagnosis process up to the current time step is available to all models in order to predict the tool wear of the three flutes at the current time step, $T$.

The required parameters for PS-HMCO are estimated from the training dataset based on equations (7), (8) and (9). Given the observation sequences, $O_{1:T}$, of each flute, $C_T$ is computed

V. Diagnostics & Prognostics Results

In this section, performance of the proposed PS-HMCO approach is compared with two artificial neural network approaches in diagnostics of the cutter. Moreover, the prognostics ability of the PS-HMCO is tested.

As supervised learning algorithms, NNs are used in both regression and classification problems [42]. Among various NNs, multi-layer perceptron (MLP) and Elman network are two of the most commonly used neural networks that can be adopted for continuous tool condition monitoring. In this case, features are regarded as the inputs of the networks and the output is the measured wearing metric [10], [21–23].

First, optimal NNs and PS-HMCO are determined through identifying their hyper-parameters, which are the number of hidden state values in PS-HMCO as well as the structure and number of hidden neurons in both NNs. Here the cross-validation method is applied. Then, performance of all three approaches are compared based on three accuracy indicators, namely, mean relative error (MRE), mean square error (MSE), and coefficient of determination ($R^2$). Finally, the capability of PS-HMCO in prognostics is studied. It is noteworthy that in all three cases, the collected data from the three cutters (07BX1, 31PN4 and 34PT1) that was used in feature selection process is regarded as the training set for all models.

A. Determination of Hyper-parameters

Fig. 5 indicates that the Elman network with structure $X = 4 - 1$ structure which leads to minimum MSE at $X = 10$.

Fig. 6 indicates that the Elman network with structure $X = 3 - 1$ structure which leads to minimum MSE at $X = 20$.
Fig. 7. Cross-Validation results for Case I with different number of discretized health states (hidden state values) while using PS-HMCO approach.

for each flute based on (15). The maximum value of the computed $C_T$ among the three flutes at each time step $T$ is regarded as the ultimate output in all approaches and is compared with the maximum actual wearing value of 3 flutes at $T$. Number of health states in PS-HMCO are set to 14 as suggested in the cross-validation case.

At each time step within each experiment, the maximum wearing value among 3 flutes is used as the desired outcome. The rationale is that, the quality of the ultimate work-piece is determined by the maximum tool wear of the three flutes. Figure 8 depicts the predicted ultimate output of all approaches along with the actual outcome (true output) for one of the testing experiments ($18SC3$). Table IV shows the diagnosis error rate of all three approaches on the testing set in terms of MSE in details as well as the overall diagnosis accuracy of the approaches in terms of MSE, MRE and $R^2$. It is noteworthy that the provided accuracies of the NNs are the averaged accuracies over 10 trials.

As it can be seen in Table IV, Elman network outperforms the conventional MLP approach, which indicates that the underlying temporal information for TCM cannot be ignored. Moreover, the PS-HMCO approach outperforms both MLP and Elman network, which suggests that PS-HMCO approach is stronger in capturing the temporal information for TCM comparing to the Elman network.

It is noteworthy that the approaches are not applied online in this study. The data has collected and stored so that the same data can be provided to test different approaches and conduct fair comparisons. However, the online application is simulated. In order to simulate time progression, data is sequentially provided to the prediction models each time up to a simulated current time step, and then the prediction is made. The average computation time for diagnostics through time by PS-HMCO is measured to be 2.6ms in this study, with 14 health states and 38 input features (observations) at each sampling time. Thus, it suggests that the PS-HMCO approach is computationally feasible for the similar online applications. The issue that has to be addressed for the online application is the buffering of the signals done by LabView® and the real-time connection between LabView® and MATLAB® or any other program that PSHMCO is implemented on.

C. Prognostic Results

In this test, PS-HMCO is unrolled over the time axis for prognostics. Different prediction horizons are explored and the mean of corresponding prediction accuracies over the testing set are provided in Table V in terms of MSE.

Similar to diagnostics, the maximum value among the three flutes at each time step is used as the desired output. Having the prediction horizon, $t_p$, and given the observation sequences from beginning of the experiment up to $T$, the task is to predict the desired output at time $t' = T + t_p$ and this can be computed using (17). Fig. 9 shows the prediction results by PS-HMCO along with the real tool wear on one of the test experiments ($18SC3$) with different prediction horizons, which are 1, 4, 7 and 10 time steps. The starting point is set to 11.

From Table V and Fig. 9, it can be seen that, although the overall accuracy has reduced comparing to the diagnostics (which is expected), interestingly the resultant prognostics accuracy when the prediction horizon is equal to 1 or 2, still outperforms the MLP in diagnostics, which shows the
importance of the temporal information. Although the PS-HMCO’s prognosis accuracy is unable to beat the Elman network diagnostics result, since Elman network also captures temporal information, the results confirms the power of PS-HMCO in capturing the temporal information, and the suitability of PS-HMCO as a predictor.

VI. CONCLUSION

A temporal probabilistic approach is proposed for the continuous tool condition monitoring in machinery systems. The proposed PS-HMCO approach is based on a physically segmented hidden Markov model that can handle continuous output. As an illustrative example, the proposed approach is applied to tool wear prediction in a CNC-milling machine.

The experimental study indicates that PS-HMCO outperforms multi-layer perceptron and the Elman networks in tool condition monitoring. It is also shown that the proposed approach can be used for prognostics by unrolling the model over the time horizon. The PS-HMCO approach is found to be suitable for the applications in which the operating conditions are fixed. The fixed operating conditions property, similar to the conducted experiment, can be seen in applications with high volume of productions such as molding processes.

In future, some of the restrictions that are enforced in the model must be lifted to improve the performance further. Moreover, as completely different regimes would arise when different operating conditions are being considered, multimodal approaches can be studied. Switching models that can each capture a specific regime and increase the resolution within each trend can be applied to the more general cases with different operating conditions.

REFERENCES


### TABLE IV

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean Square Error</th>
<th>Total Testing Error Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>09BX3</td>
<td>18SC3</td>
</tr>
<tr>
<td>PS-HMCO</td>
<td>341.2336</td>
<td>135.3554</td>
</tr>
<tr>
<td>MLP</td>
<td>778.4757</td>
<td>283.4326</td>
</tr>
<tr>
<td>Elm-Net</td>
<td>580.4041</td>
<td>277.9204</td>
</tr>
</tbody>
</table>

#### TABLE V

<table>
<thead>
<tr>
<th>Model</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS-HMCO</td>
<td>422.3071</td>
<td>491.4586</td>
<td>602.7350</td>
<td>680.1204</td>
<td>732.5235</td>
<td>771.1665</td>
<td>802.2608</td>
<td>829.0004</td>
<td>852.7790</td>
<td>874.3407</td>
</tr>
</tbody>
</table>

**REFERENCES**


Omid Geramifard received his Bachelor’s degree in Computer Engineering from Isfahan University of Technology, Isfahan, Iran in 2008. He was awarded Singapore International Graduate Award (SINGA) and joined National University of Singapore as a PhD student in 2009. Currently, he is a PhD candidate in Electrical and Computer Engineering Department. 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.

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. His research interests include system learning, control, and applications to real-time systems. 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.

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.

Since 1996, she worked as a Research Fellow at the Singapore Institute of Manufacturing Technology, A*STAR, Singapore. 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 received her B. Eng. in 1982 and M. Eng. in 1987 from the Northeastern University, China, and PhD from Nanyang Technological University, Singapore in 2000. She taught at the Northeastern University in 1982-87. In 1987-1989, she did research work as Visiting Research Fellow at Tohoku University, Japan. She joined Singapore Institute of Manufacturing Technology (SIMTech) in 1992 and she is currently a Senior Scientist in SIMTech and team leader of Prognostic Health Management.

Dr Li 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 as well as regression modeling. She is a member of the Decision Sciences Institute, USA.