



CNC machine tool's wear diagnostic and prognostic by using dynamic Bayesian networks

D.A. Tobon-Mejia^{a,b}, K. Medjaher^{a,*}, N. Zerhouni^a

^a FEMTO-ST Institute, UMR CNRS 6174 - UFC/ENSMM/UTBM Automatic Control and Micro-Mechatronics Systems Department, 24, rue Alain Savary, 25000 Besançon, France

^b ALSTOM Transport, 7, avenue De Lattre De Tassigny, BP 49, 25290 Ornans, France

ARTICLE INFO

Article history:

Received 30 November 2010

Received in revised form

5 October 2011

Accepted 28 October 2011

Available online 16 December 2011

Keywords:

Diagnostic

Prognostic

Remaining useful life

Condition-based maintenance

Hidden Markov Models

Tool wear

ABSTRACT

The failure of critical components in industrial systems may have negative consequences on the availability, the productivity, the security and the environment. To avoid such situations, the health condition of the physical system, and particularly of its critical components, can be constantly assessed by using the monitoring data to perform on-line system diagnostics and prognostics.

The present paper is a contribution on the assessment of the health condition of a computer numerical control (CNC) tool machine and the estimation of its remaining useful life (RUL). The proposed method relies on two main phases: an off-line phase and an on-line phase. During the first phase, the raw data provided by the sensors are processed to extract reliable features. These latter are used as inputs of learning algorithms in order to generate the models that represent the wear's behavior of the cutting tool. Then, in the second phase, which is an assessment one, the constructed models are exploited to identify the tool's current health state, predict its RUL and the associated confidence bounds. The proposed method is applied on a benchmark of condition monitoring data gathered during several cuts of a CNC tool. Simulation results are obtained and discussed at the end of the paper.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

The maintenance activity plays a major role in industrial systems as it permits to improve the availability, reliability and security, while reducing the life cycle cost. There exist several types of maintenance, which can be classified into two main categories, namely: curative and preventive maintenances [1,2]. In the first case, the interventions are done only after the observation of the failure, whereas in the second case they are realized either systematically or conditionally to the health condition of the system. This type of maintenance is commonly termed as a condition based maintenance (CBM). Indeed, the condition of the industrial system is continuously monitored and inspected by a set of sensors. The data recorded by these latter are then processed in order to extract relevant features that allow to estimate the current health state and to project this one in the future. The estimated and projected states are then used to take appropriate maintenance decisions. Diagnostic aims at assessing the component's current condition and identifying the cause of its failure, whereas prognostic is used to predict its future health state in order to anticipate the failure [3–6].

* Corresponding author. Tel.: +33 3 81 40 27 96.

E-mail address: kamal.medjaher@ens2m.fr (K. Medjaher).

Formally, failure prognostic consists of estimating the time before failure or the remaining useful life (RUL) and the associated confidence value. It can be realized by using three main approaches [7,8], namely: model-based prognostic, experience-based prognostic and data-driven prognostic. Model-based prognostic consists of studying each component or sub-system in order to establish for each one of them a mathematical model of the degradation phenomenon. The derived model is then used to predict the future evolution of the degradation and thus the related RUL value. Experience-based prognostic methods use mainly probabilistic or stochastic models of the degradation phenomenon, or of the life cycle of the components, by taking into account the data and knowledge accumulated by experience during the whole exploitation period of the industrial system. Data-driven prognostic is based on the transformation of the monitoring data into relevant behavioral models permitting to predict the RUL and the associated confidence.

This paper deals with the assessment of the cutting tool's health condition of computer numerical control (CNC) machines through the utilization of dynamic Bayesian networks (DBN). The proposed method belongs to the data-driven prognostic approach and aims at transforming the raw monitoring data provided by the sensors into behavioral models that represent the evolution of the cutting tool's degradation. The obtained models are then used to continuously estimate the current state of the cutting tool and calculate its RUL. The choice of this approach dwells in the fact that in the assessment of the cutting tool's condition of CNC machines, the main problem is that deriving a behavioral model in an analytical form that best fits the dynamic of the tool's wear is not a trivial task. Furthermore, finding experience data for a long period of time is expensive and not easy in practice. This is why the utilization of the data provided by the monitoring sensors may be a trade-off between the model-based prognostic and the experience-based prognostic. Thus, the idea behind this contribution is the transformation of the raw monitoring data into relevant models representing the wear's behavior of the cutting tools of CNC machines.

The proposed method relies on two main phases: a learning phase and an assessment phase, as this is done in the framework of data-driven system health monitoring and prognostic [9,10]. During the first phase, the raw data are used to extract reliable features, which are then used to learn behavioral models representing the dynamic of the degradation in the cutting tool. The modeling of the degradation is done by using a mixture of Gaussians Hidden Markov Model (MoG-HMM) represented by a DBN. This probabilistic graphical model allows to use continuous observations and also to speed up the inference by using the algorithms proposed for DBNs [11]. In the second phase, the learned models are exploited on line to assess the current health state of the cutting tool and to estimate the value of the RUL and its associated confidence value.

The paper is organized as follows: in Section 2 the diagnostic and prognostic paradigms are presented, where some definitions and the related state of the art are given, Section 3 is dedicated to the proposed diagnostic and prognostic method and finally, an application example and simulation results are given in Section 4.

2. Diagnostic and prognostic framework

2.1. Definitions

The term prognostic finds its origin in the Greek word “*prognôstikos*”, which means “to know in advance”. Prognostic is well used in medical domain, where doctors try to make predictions about the health of a patient by taking into account the actual diagnosis of a disease and its evolution compared with other similar observed cases. This reasoning can be transposed into the industrial domain where the patient is a machine, an industrial system or a component.

Several definitions have been given in the literature about industrial prognostic [7,12–14], where three main points are highlighted: the system's actual state, the projection (or extrapolation) of this latter, and the estimation of the remaining time before failure. These definitions are then normalized by the ISO 13381-1 standard [15] in which prognostic is defined as the estimation of the operating time before failure and the risk of future existence or appearance of one or several failure modes. This standard defines the outlines of prognostic, identifies the data needed to perform prognostic and sets the alarm thresholds and the limits of system's reset (total shut-down). The main steps to perform prognostic, as defined in the standard, are summarized in Fig. 1.

The first step consists of monitoring the system by a set of sensors or inspections achieved by operators. The monitored data are then pre-processed to be used by the diagnostic module. The output of this module is an identification of the actual operating mode (more details on failure diagnostic can be found in [3–5]). This mode is then projected in the future, by using adequate tools, in order to predict the system's future state. The intersection point between the value of each projected parameter or feature and its corresponding alarm threshold permits to estimate the RUL (Fig. 2). Finally, appropriate maintenance actions can be taken depending on the estimated RUL. These actions may aim at eliminating the origin of the failure, which can lead the system to evolve to any critical failure mode, delaying the instant of a failure by some maintenance actions or simply stopping the system if this is judged necessary.



Fig. 1. Prognostic steps according to ISO 13381-1 [16].

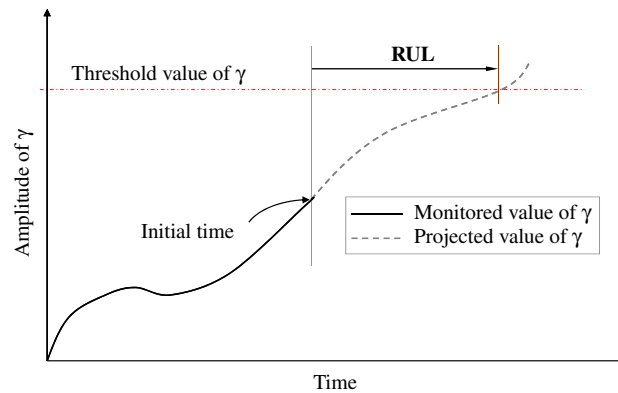


Fig. 2. RUL illustration.

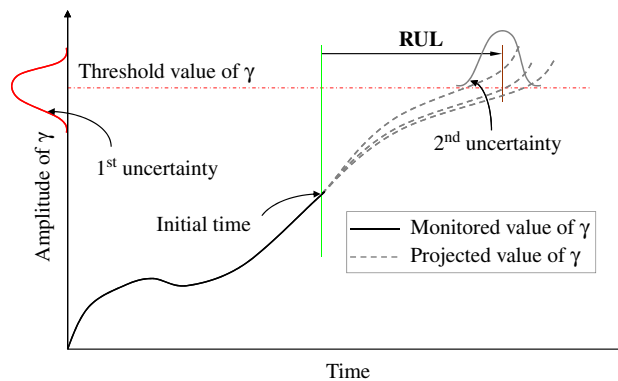


Fig. 3. Uncertainty inherent to RUL prediction.

As in any prediction work, a prediction error should be associated to the estimated value of the RUL (Fig. 3). The sources of the prediction error may be multiple: modeling hypotheses, non-significant data, used prediction tools, uncertainty in the thresholds' values, etc. In addition, uncertainty is intrinsic to any prognostic work [17].

The error associated with a RUL estimation should decrease as the time of the real failure approaches. This is because the predictions are adjusted each time new data are acquired. In addition, a confidence degree should be associated. Indeed, instead of telling an industrial that the machine will fail in x units of time, it would be more realistic to give an estimated RUL with a confidence value.

As mentioned previously, the value of the estimated RUL is the output of some comparison between the projected state of the system and the predetermined threshold values (these values can be determined by using learning algorithms like those of neuro-fuzzy systems [18]). Note that, at the projection step, what is needed is not necessarily a value of a physical parameter but can be a desired performance, an achieved function or the availability of a service, depending on the kind of system on which prognostic is performed.

2.2. Prognostic approaches

Although the novelty of industrial prognostic activity, numerous methods and tools have been proposed and reported in the literature. By using as classification criteria the type of data needed to perform failure prognostics, it is possible to group the different techniques in three main categories [4–6]: model-based prognostic, experience-based prognostic and data-driven prognostic, as shown in Fig. 4a. Note that this classification is not totally binary as it may exist prognostic methods based on the fusion of more than one approach [19], see Fig. 4b.

In model-based prognostic, the physical component or system and its degradation phenomenon are represented by a set of mathematical laws. The obtained behavioral model is then used to predict the future evolution of the degradation [20,21]. Whereas, the methods belonging to the experience-based approach use probabilistic or stochastic models of the degradation phenomenon, or of the life cycle of the components, by taking into account the data and the knowledge accumulated by experience [22,23].

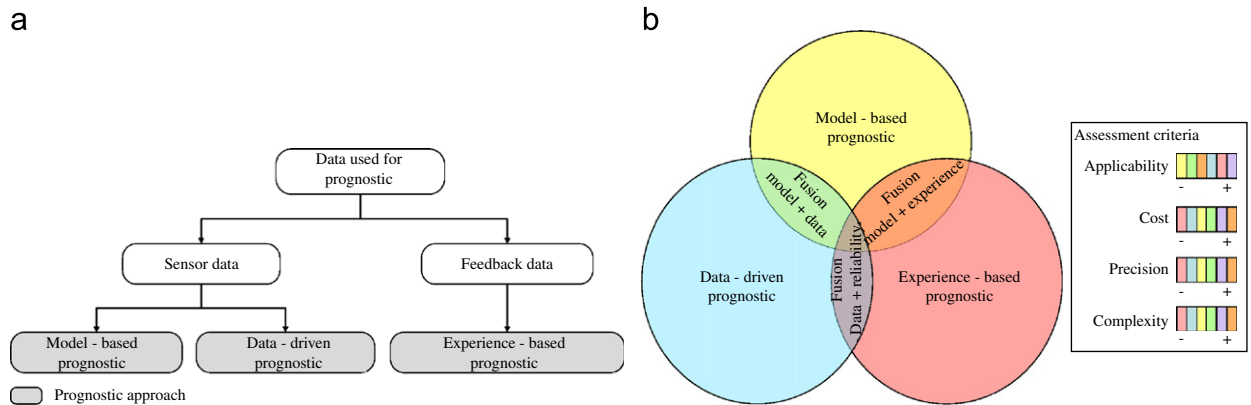


Fig. 4. Prognostic approaches: (a) classification and (b) performance evaluation.

Finally, the data-driven approach aims at transforming the raw monitoring data into relevant behavior models of the system including its degradation. Indeed, the degradation model is derived by using only the data provided by the monitoring system (the sensors mainly), without caring about the system's analytical model neither on its physical parameters (like material properties).

Data-driven methods offer an alternative to the model-based and experience-based ones, especially in cases where obtaining reliable sensor data is easier than constructing analytical behavior models or waiting to obtain sufficient experience data. The use of monitoring data to derive a behavioral model may be considered as a trade-off between complexity and precision. Indeed, in practice most of the degradation phenomena are complex (due to nonlinearities, stochasticity, non-stationarities, etc.) and difficult to model by using analytical models and experience data.

In the data-driven approach several tools are used, most of them are originated from artificial intelligence domain. Among these tools one can cite neuronal networks, dynamic Bayesian networks, Markovian processes and statistical regression methods.

In [24] a prognostic method based on the use of recurrent neural networks to model the crack propagation in a bearing has been proposed. The network structure used by the authors was able to track the time evolution of the crack size and to estimate the value of the RUL. In addition to traditional neural networks, neuro-fuzzy systems were used in failure prognostic. Thus, Wang et al. [13] have proposed a neuro-fuzzy based prognostic method to predict the future health state of a pinion. The fuzzy rules were given by experts whereas the forms of the membership functions were learned by using a neural network. The authors showed that the results obtained by using neuro-fuzzy networks were more relevant than those provided by a simple neural network. The same approach has been applied by Chinnam and Baruah [18] on a vertical machining center. Moreover, posterior simulations conducted by Wang [25] have shown that, compared to neural networks, the use of a feed-forward neuro-fuzzy network can significantly increase the accuracy of the predictions and the accuracy of the estimated value of the RUL. In [26] a self-organizing map (SOM) has been implemented to perform both failure diagnostic and prognostic on bearings by exploiting vibration signals along with a set of historical data. The historical data have been modeled by using independent neural networks, which estimated separate local RULs. The global RUL was then calculated by weighting the local RULs, the weightings were controlled by the degree of representativity of each historical data set. However, the problem of neural networks based methods is that the definition of the network's structure is often complex. Indeed, the generation of the structure is based on a process of test-error estimation, which necessitates long time calculations and learning.

A second tool, which has been used in data-driven failure prognostic is the Kalman filter. In [27] a Kalman filter based prognostic has been proposed in order to track the time evolution of a crack in a tensioned steel band. The Kalman filter is used to model and to estimate the drift of the modal frequency of the steel band as a function of the applied vibrations. More general than Kalman filters, particle filters have also been used to perform nonlinear projections of features. Orchard et al. [28] have proposed a particle filter method used to estimate a crack growth in a turbine's paddle. In their prognostic model, the current size of the crack and its evolution is recursively estimated by using the data provided by sensors. Nevertheless, in the case of Kalman filter and its variants the difficulty to generate a prior model, with the definition of its parameters, needed for filtering and predictions can limit their applicability.

In failure diagnostic and prognostic domains, the hidden semi-Markov models (HSMM), have proved to be a suitable tool as they allow to model the physical component's degradation by using continuous observations provided by the monitoring sensors. They also permit to estimate the stay durations in each health state leading to the prediction of the RUL value [29]. HSMM can be used to represent several failure modes by using historical data for learning. Moreover, the number of observations can be modified depending on the application and the implementation constraints.

The HSMM permit to do failure prognostic for a long time horizon. Indeed, once the current health state is identified and assuming that the stay durations in the states are estimated, the prediction of the RUL is straightforward [30].

Furthermore, contrary to other tools used in the framework of data-driven approach, such as the regression models or neural networks where the structure is not interpretable [31], the states in the HSMM can be interpreted as the health conditions of the component.

In this same context, Chinam and Baruah [30] have used hidden Markov models (HMM) to assess the degradations of bearings and to estimate the underlying RUL. In their method the authors considered the degradation as a stochastic process with several states representing different health states of the physical component. The degradation levels of each bearing are first learned by using vibration data (several HMMs corresponding to each state are obtained during the off-line phase). Then, during the on-line operation of the bearing the processed data are continuously supplied to each HMM in order to calculate a likelihood value, which permits to select the model that best represents the current state of the bearing. Finally, knowing the current state and its corresponding stay duration, it is possible to estimate the value of the RUL.

More recently, dynamic Bayesian networks [11], a tool generalizing the HMMs and the Kalman filter, have been exploited to perform failure prognostic. Prytzula and Choi [32] proposed an integrated DBNs based diagnostic and prognostic method where the uncertainty related to the operating conditions is taken into account. Similarly, Muller et al. [12] proposed a DBNs based procedure integrating both the degradation mechanism and the maintenance actions in the same model. Medjaher et al. [33] published a procedure to estimate the RUL of a work station in a manufacturing system, where maintenance actions on several components were introduced in the DBN model in order to observe the modifications in the estimated RUL. Finally, Dong and Yang [34] implemented a particle filtering algorithm applied to a DBN in order to estimate the RUL of a vertical machining center.

3. Integrated diagnostic and prognostic method

The cutting tool's wear assessment in CNC machines can be done in two different ways: by using mathematical or mechanical methods, such as numerical simulations based on finite elements models, or by exploiting the monitoring data provided by the sensors installed on the machines. The first method is a model-based one and can lead to more precise results at a condition that the derived models represent really the physical phenomena including the wear process. However, this is practically difficult to reach, because the wear phenomena are often nonlinear, non-stationary and not easy to model. To avoid this situation, one can use the monitoring data to build behavioral models valid for the operating conditions in which the machine evolves.

This section presents an integrated diagnostic and prognostic method based on the transformation of the data provided by the monitoring sensors into models. The method is detailed at the end of the section and its objective is to assess the health state of the cutting tool in CNC machines and to estimate its remaining useful life. Before introducing the steps of the method, a brief introduction of some necessary prerequisites would help the reader understanding it. These prerequisites concern data clustering, MoG-HMMs, DBNs and curve fitting.

3.1. Data clustering

Cluster analysis is a technique used to identify groups of individuals or objects that are similar to each other, but different from those in other groups. Clustering techniques are a simple, but can be considered as a powerful tool to understand complex phenomena by using statistics. These methods are principally used by researchers in social sciences and medicine as in [35], where they are used to identify the factors that cause stress in nurses, or also in biology as in [36] where a hierarchical cluster analysis is used to identify interfaces and functional residues in proteins. In the field of clustering two main methods exist [37]:

1. Hierarchical clustering: this is the most straightforward method of clustering, it can be either agglomerative or divisive. It consists of merge the data in different clusters using a greedy algorithm, and of represent their hierarchy using a dendrogram.¹ This technique is suitable for small data sets.
2. Partitional clustering: this method only attempts to directly decompose the data into a set of disjoint clusters. Traditionally, the called k-means algorithm is used, where k is the exact number of clusters to partition the data. Contrary to hierarchical clustering in this technique the quantity of clusters are fixed.

As argued in [38,39], the tool wear evolution is classified into five wear stages: initial wear, slight wear (regular stage of wear), moderate wear (micro-breakage stage of wear), severe wear (fast wear stage) and worn-out (or tool breakage). Thus, the partitional clustering method is suitable to identify in the learning phase the different wear stages. Using the k-means algorithm to partition the amount of wear measured after each tool cut may identify the wear stage at each cut. Then, by knowing the wear stage as a function of tool life, the features bellowing to a particular wear cluster can be used to train their respective "wear model".

¹ Tree diagram frequently used to illustrate the arrangement of the clusters produced by hierarchical clustering.

3.2. Mixture of Gaussians hidden Markov models

The MoG-HMM is primarily a hidden Markov model. This latter is a statistical model used to represent stochastic processes in which the states are not directly observed (hidden), but the outputs (observations) dependent on the hidden state are visible [40]. Each hidden state has a probability distribution over the possible values of the observations.

A discrete HMM is completely defined by the following parameters:

- N : number of states in the model. The individual states are $1, 2, \dots, N$, and the state at time t is defined as s_t .
- K : the number of distinct observations for each state. The individual observation symbols are denoted as $V = v_1, v_2, \dots, v_K$.
- A : the state transition probability distribution, $A = a_{ij}$, where $a_{ij} = P[s_{t+1} = j | s_t = i]$, $1 \leq i, j \leq N$.
- B : the observation probability distribution of a state i , $B = b_i(k)$, where $b_i(k) = P[v_k | s_t = i]$, $1 \leq i \leq N$, $1 \leq k \leq K$.
- π : the initial state distribution $\pi = \pi_i$, where $\pi_i = P[s_1 = i]$, $1 \leq i \leq N$.

For simplicity and clarity of presentation, a compact notation ($\Theta = \pi, A, B$) is used for each HMM. In practice, the HMM and its variants are used to solve three typical problems [40]: identification, decoding and learning. However, the problem with the discrete HMMs is that the observations are discrete symbols chosen from a finite alphabet while in most real applications the observations are continuous signals (or vectors). To overcome this limitation, one can use a MoG-HMM, where each signal can be expressed as a combination of a finite number of mixtures [41], each one of the form

$$b_j(O) = \sum_{m=1}^M C_{jm} \xi(O, \mu_{jm}, U_{jm}), \quad 1 \leq j \leq N \quad (1)$$

In Eq. (1), O stands for the observation vector being modeled, C_{jm} is the mixture coefficient for the m th mixture in state j and ξ is a Gaussian with mean vector μ_{jm} and covariance matrix U_{jm} for the m th mixture component in state j . Similar to an HMM, a MoG-HMM is completely defined by three parameters: the state transition matrix A , the observation matrix B and the initial probability distribution π . Moreover, for a MoG-HMM the observation matrix B is modeled by a Gaussian density with a mean μ , a standard deviation σ and a mixture matrix M . An illustration of a MoG-HMM is given in Fig. 5.

3.3. MoG-HMMs and dynamic Bayesian networks

In the last decade, a new tool namely the dynamic Bayesian networks, derived from the artificial intelligence domain, became popular thanks to its modeling, graphical representation and inference capabilities. A DBN is an extension of the traditional Bayesian network (BN) [11] used to model probability distributions over semi-infinite collections of random variables, Z_1, Z_2, Z_3, \dots, Z . For the HMM case, Z is partitioned into $Z = (S_t, V_t)$ to represent the hidden and the output variables of a state space model. A DBN is defined to be a pair, $(B1, B')$, where $B1$ is a BN that defines the prior probability of all variables known as $P(Z_1)$, and B' is a two-slice temporal Bayes net (2TBN), which defines $P(Z_t | Z_{t-1})$ by means of a directed acyclic graph (DAG) as shown in the following equation:

$$P(Z_t | Z_{t-1}) = \prod_{i=1}^N P(Z_t^i | Pa(Z_t^i)) \quad (2)$$

where Z_t^i is the i th node at time t , which could be a component of S_t, V_t (in regular HMMs) and $Pa(Z_t^i)$ are the parents of Z_t^i in the graph. The nodes in the first slice of a 2TBN do not have any associated parameters, however the nodes in the second slice have associated conditional probability distributions (CPD), which define $P(Z_t^i = | Pa(Z_t^i))$ for all $t > 1$.

A DBN is a powerful tool that allows modeling sequential data [11]. In one hand, DBNs generalize HMMs by representing the state space in a factored form, instead of a single random variable. On the other hand, they model the well known Kalman filters models (KFM) by using arbitrary probability distributions instead of the traditional Gaussian models.

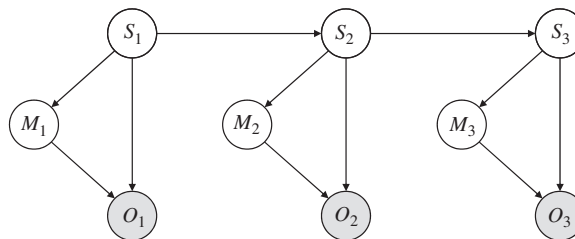


Fig. 5. A mixture of Gaussians HMM [11].

The main advantage in modeling HMMs and its variants by DBNs is that the inference can be done faster. For example, the inference in hierarchical HMMs is performed in $O(T)$ time by using a junction tree algorithm instead of $O(T^3)$ time for the traditional algorithm, where T is the length of the data sequence [11].

By using a DBN, one can easily represent all the variants of HMMs. The only requirement is to build the correct DAG and define the correct type of nodes and CPDs. An example of a MoG-HMM represented by a DBN is shown in Fig. 6.

In this model, S represents the hidden states, M the mixture coefficients and V the observation distributions. For an illustration purpose, the corresponding CPDs are presented below:

$$S_1 = \pi \quad (3)$$

$$M_1 = M_t = C_{jm} \quad (4)$$

$$V_1 = V_t = \xi(O, \mu_{jm}, U_{jm}) \quad (5)$$

$$S_t = a_{ij} = P[S_{t+1} = j | S_t = i] \quad (6)$$

Once the parameters of the model are completely defined, the learning algorithms for DBN [11] can be used to correctly estimate the CPD values.

3.4. Curve fitting

Curve fitting is a numerical process that is used to represent a set of experimentally measured (or estimated) data points by some hypothetical analytical functions [42]. The results of this curve fitting process are the coefficients, or parameters, that are used to define the fitting function. The points are constrained to a supposed polynomial with a fixed order. Let us suppose y the output data and x the input data. By using numerical algorithms, the input and output data will be fitted to the target function, which can be of \mathcal{A} order, as shown in the following equation:

$$y(x, \beta) = \beta_0 + \beta_1 x + \dots + \beta_{\mathcal{A}} x^{\mathcal{A}} = \sum_{\alpha=0}^{\mathcal{A}} \beta_{\alpha} x^{\alpha} \quad (7)$$

where $\beta_{\mathcal{A}}$ stands for the parameters of the supposed polynomial function to be defined.

3.5. Description of the diagnostic and prognostic method

An integrated diagnostic and prognostic method for estimating the current health state of the cutting tool in high-speed CNC machines and predicting its remaining useful life is proposed in the following of the paper. The method relies on the utilization of the monitoring data provided by the sensors installed on the machine to track its tool's condition represented by the magnitude of the wear. The main contribution of the method dwells in the fact that in the generated model the magnitude of wear per cut is learned by using the features extracted from the monitoring data, instead of doing it by using a mathematical model of the wear for which material and environmental coefficients are not trivial to estimate. Furthermore, the proposed method may be used to assess the health state of any physical component (drill bits, milling cutters, tool bits, rolling elements, etc.) at a condition that the necessary features for learning can be extracted.

The principle of the proposed method relies on two main phases: a learning phase and an exploitation phase, as shown in Fig. 7. In the first phase, which is realized off-line, the tool's wear measured after each cut from the new state until the failure is used. This latter is fed as input to the k-means algorithm in order to partition the wear and to identify which information correspond to a specific health state of the cutting tool. In this way, the monitoring data can be classified into different groups, each one corresponding to a particular wear level. Then, the appropriate features are extracted from the raw monitoring signals by using a feature extraction procedure (F.E). In the feature matrix F , each column vector

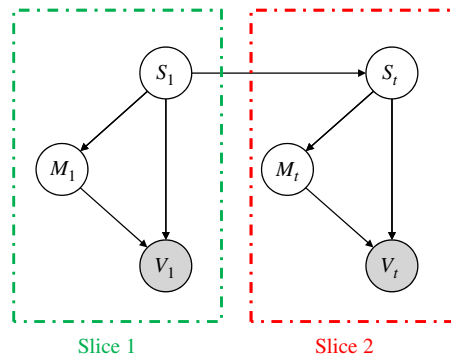


Fig. 6. A MoG-HMM represented by a DBN.

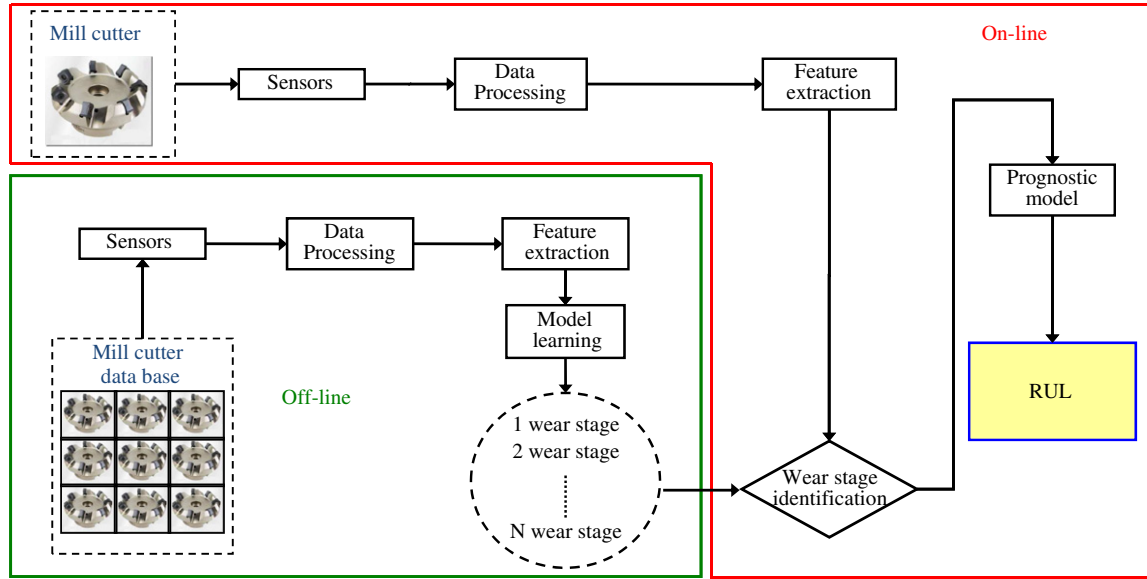


Fig. 7. The two phases of the proposed method.

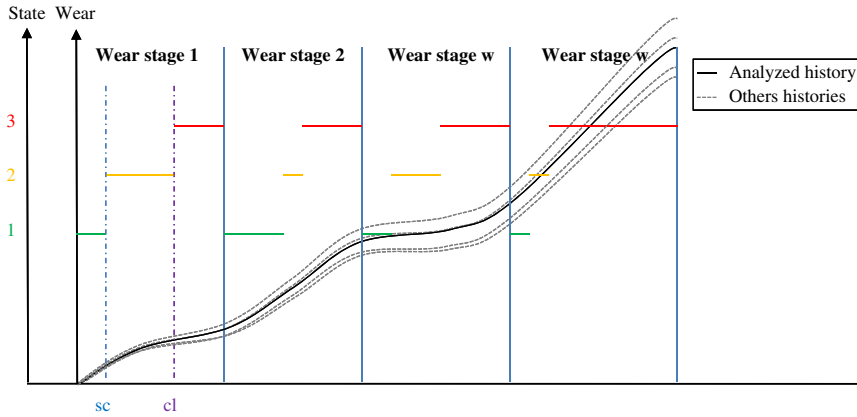


Fig. 8. Viterbi decoding sequence for one history of tool wear evolution.

(of D features at a cut c) corresponds to a snapshot on the raw signal, and each cell f_{dc} represents the feature d at a cut c , as shown in the following equation:

$$\text{Raw signal} \xrightarrow{F,E} F = (f_{1c} f_{2c} \cdots f_{dc})'$$

$$\text{with } 1 \leq c \leq C \text{ and } 1 \leq d \leq D$$

(8)

The matrix of extracted features is finally used to estimate the parameters of different MoG-HMMs represented by a DBN. The advantage of using several features instead of only one is that it can happen that a single feature may not capture all the information related to the behavior of the physical component.

The learned models are divided into two groups. The first group contains the general models per wear level, resuming all the data from all the learning histories and stored in a database, called “global model wear base”. The second group feeds a model database, which contains one model per wear level and learning history, named “local wear model base”. In this way if W are the wear stages and H the learning histories, the global model wear base contains W models (one per wear stage), whereas the local wear model base contains $W \times H$ models (one model per wear stage and history). This latter permits to obtain the state sequence and to compute how much wear is acquired in each model state (hidden states) using its associated wear measures. Also, by using the Viterbi algorithm [43] it is possible to find the cuts belonging to a particular state (Fig. 8). Thus, by assuming that the amount of wear acquired in each wear stage is governed by a Gaussian

distribution, it is possible to estimate the mean “ μ ” and the standard deviation “ σ ” of the amount of wear “ W_r ” and its variation “ ΔW_r ” for each state S_i of the models stored in the local wear model base, as shown in the following equations:

$$\mu(W_r^h(S_i)) = \frac{1}{T_c} \sum_{c=sc}^{cl} wr_w^h(c) \quad (9)$$

$$\mu(\Delta W_r^h(S_i)) = \frac{1}{T_c} \sum_{c=sc+1}^{cl} (wr_w^h(c) - wr_w^h(c-1)) \quad (10)$$

$$\sigma(W_r^h(S_i)) = \sqrt{\frac{1}{T_c} \sum_{c=sc}^{cl} [wr_w^h(c) - \mu(wr_w^h(S_i))]^2} \quad (11)$$

$$\sigma(\Delta W_r^h(S_i)) = \sqrt{\frac{1}{T_c} \sum_{c=sc+1}^{cl} [(wr_w^h(c) - wr_w^h(c-1)) - \mu(\Delta W_r^h(S_i))]^2} \quad (12)$$

In the previous equations wr_w^h stands for the wear associated to the stage $w=1, \dots, W$ and the history $h=1, \dots, H$, i is the state index and c the cut index constrained by the limits found in the cluster analysis (sc =start cut, cl =cut limit and $T_c = cl - sc + 1$, as shown in Fig. 8). A compact representation of the learned model, which can then be used to perform health condition diagnostic and prognostic is given by the following expression:

$$\lambda = DBN_w(\theta), DBN_w^h(\theta), \mu(W_r^h(S_i)), \mu(\Delta W_r^h(S_i)), \sigma(W_r^h(S_i)), \sigma(\Delta W_r^h(S_i)) \quad (13)$$

In the above expression, λ represents the model, $DBN_w(\theta)$ are the parameters of the DBN that capture the behavior of the wear stage w summarizing all the histories H (global model wear base), $DBN_w^h(\theta)$ are the parameters of the DBN that capture the behavior of the wear stage w from the history h (local model wear base), $\mu(W_r^h(S_i))$ and $\mu(\Delta W_r^h(S_i))$ are respectively the mean wear and the mean wear variation of the state i in the stage wear w learned from the history h . In the same way, $\sigma(W_r^h(S_i))$ and $\sigma(\Delta W_r^h(S_i))$ stand for the standard deviation of the wear and the standard deviation of the wear variation in the state i for the wear stage w learned from the history h .

The on-line phase of the proposed method consists of exploiting the learned models to identify the tool's current condition (by using DBN inference algorithms) and to estimate the current tool's wear leading to an estimation of the value of the corresponding RUL. The processed data and the extracted features are thus continuously fed to the learned models in order to identify the actual wear stage. The identification process is based on the evaluation of the likelihood $P(O|\lambda)$ of the models (belonging to the global model wear base) over the observations. By knowing the current wear stage and by using the wear accumulation and the wear variation learned during the off-line phase, the cutting tool's RUL and its associated confidence value can be calculated. The dedicated procedure to estimate the RUL and the associated confidence value during the on-line phase, based on the use of the DBNs “ λ ” are explained through the following steps:

- The first step consists of detecting the appropriate DBN that best fits and represents the on-line observed sequence of features. Indeed, the features are continuously fed to the set of learned models in the global model wear base and a likelihood is calculated in order to select the appropriate model. The selected model is then used to define the current wear stage w (Fig. 8).
- The second step concerns the identification of the DBN that best fits the observations knowing the actual wear stage $P(O|\lambda, w)$. The DBN inference algorithm is thus applied on the local wear model base (belonging to the known wear stage) to compute the likelihood and to choose the best DBN. Then, the Viterbi algorithm is applied on the selected model in order to find the hidden state sequence, which permits to identify the current wear, the wear amount and the wear variation by choosing the most persistent state in the last observation sequence. This state is stored in a global state sequence as $G_w^h(S_i)$ in the cell c , which contains the current and the old states.

$$\text{state sequence} = (G_w^h(S_i)_1, G_w^h(S_i)_2, \dots, G_w^h(S_i)_c),$$

c = current cut

- The current global state sequence and the wear information $\mu(W_r^h(S_i))$, $\mu(\Delta W_r^h(S_i))$, $\sigma(W_r^h(S_i))$ and $\sigma(\Delta W_r^h(S_i))$ are used in the third step to estimate the wear vector. The idea is to estimate at each cut c the amount of mean wear and the bounds using the confidence value. The current state of each cell is compared with the precedent state (precedent cell). If the states are same the mean wear variation $\mu(\Delta W_r^h(S_i))$ is added and the bounds are defined by using the confidence factor n , otherwise the wear and the bounds are updated, as shown in the following equations. Three values

of wear are estimated: the maximum “ $\widehat{W}_{r_u}(c)$ ”, the mean “ $\widehat{W}_{r_m}(c)$ ” and the minimum “ $\widehat{W}_{r_l}(c)$ ” as shown in the following equations.

$$\widehat{W}_{r_u}(c) = \begin{cases} \mu(W_{r_w}^h(S_i)) + n \cdot \sigma(W_{r_w}^h(S_i)), & \text{IF } (c = 1) \text{ OR } ([G_w^h(S_i)]_c \neq [G_w^h(S_i)]_{c-1}) \\ \widehat{W}_{r_m}(c-1) + [\mu(\Delta W_{r_w}^h(S_i)) + n \cdot \sigma(\Delta W_{r_w}^h(S_i))], & \text{IF } [G_w^h(S_i)]_c = [G_w^h(S_i)]_{c-1} \end{cases} \quad (14)$$

$$\widehat{W}_{r_m}(c) = \begin{cases} \mu(W_{r_w}^h(S_i)), & \text{IF } (c = 1) \text{ OR } ([G_w^h(S_i)]_c \neq [G_w^h(S_i)]_{c-1}) \\ \widehat{W}_{r_m}(c-1) + \mu(\Delta W_{r_w}^h(S_i)), & \text{IF } [G_w^h(S_i)]_c = [G_w^h(S_i)]_{c-1} \end{cases} \quad (15)$$

$$\widehat{W}_{r_l}(c) = \begin{cases} \mu(W_{r_w}^h(S_i)) - n \cdot \sigma(W_{r_w}^h(S_i)), & \text{IF } (c = 1) \text{ OR } ([G_w^h(S_i)]_c \neq [G_w^h(S_i)]_{c-1}) \\ \widehat{W}_{r_m}(c-1) + [\mu(\Delta W_{r_w}^h(S_i)) - n \cdot \sigma(\Delta W_{r_w}^h(S_i))], & \text{IF } [G_w^h(S_i)]_c = [G_w^h(S_i)]_{c-1} \end{cases} \quad (16)$$

- Finally, in the fourth step, the estimated wears are used to compute the RUL. This latter is obtained by using the information stored in the three vectors obtained in the previous phase $\widehat{W}_r = [\widehat{W}_{r_u}, \widehat{W}_{r_m}, \widehat{W}_{r_l}]$. Each vector is fitted to a polynomial function $\widehat{W}(\widehat{W}_r, \beta)$ with the same order than the current wear stage ($A = w$) to avoid linear variations and to smooth the data. The RUL is calculated by taking the difference between the wear limit (W_{limit}) and the mean of the three smoothed wear estimations, as shown in the following expression:

$$\widehat{W}(\widehat{W}_r, \beta) = \sum_{\alpha=0}^w \beta_{\alpha} \widehat{W}_r^{\alpha} \\ RUL(c) = \left(W_{limit} - \frac{\widehat{W}(\widehat{W}_{r_u}(c), \beta) + \widehat{W}(\widehat{W}_{r_m}(c), \beta) + \widehat{W}(\widehat{W}_{r_l}(c), \beta)}{3} \right) \quad (17)$$

4. Application and simulation results

The CNC machine tools are widely used in industry for achieving productivity performance goals. Statistically, 20% of the down-time of these machines is attributed to the cutting tool failure, resulting in productivity and economic losses [44]. Thus, the prediction of the amount of wear before reaching the wear limit of the cutters may help improving the reliability, the availability and the safety of CNC machines, while insuring the surface roughness requirements and reducing the maintenance costs.

The wear diagnostic and prognostic method presented in the previous section was tested on the “prognostic data challenge 2010” database [45], which contains several histories of high-speed CNC milling machine 3-flute cutters used until a significant wear stage (Fig. 9). Each cutter made 315 cuts over an identical workpiece for a face milling job. The authors of the experiments recorded the monitoring data from dynamometer, accelerometer and acoustic sensors during the cut process and measured the amount of wear after each cut for three experiments (total of three sets of 315 cut files). These histories were named by the authors as: cutter 1, cutter 4 and cutter 6. The details of each data set are given in Table 1.

The experimental records from these tests were obtained under constant conditions. The cutting parameters were: the spindle speed of the cutter was 10 400 rpm, the feed rate was 1555 mm/min, the Y depth of cut (radial) was 0.125 mm and the Z depth of cut (axial) was 0.2 mm. The data were acquired at 50 kHz/channel. For simulation purposes (learning and on-line RUL estimation), three condition monitoring data were used (two for learning and one for testing), each cutter was considered as worn at the end of its associated history. The learning results discussed hereafter are obtained by choosing the cutters 1 and 4 for learning. However the RUL results of all possible combinations are presented in Fig. 13.

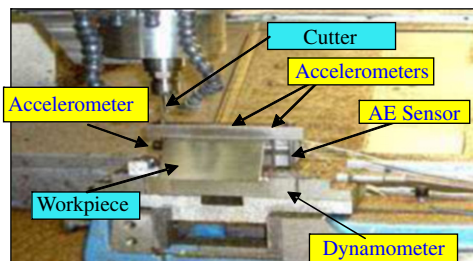


Fig. 9. CNC milling machine testbed (Röders Tech RFM760).

Table 1
Test data set description.

History	No of cuts	Total wear (10^{-3} mm)		
		Flute 1	Flute 2	Flute 3
Cutter 1	315	172.6868	164.6378	158.1924
Cutter 4	315	196.1648	210.9193	202.14954
Cutter 6	315	179.4365	233.6742	234.7159

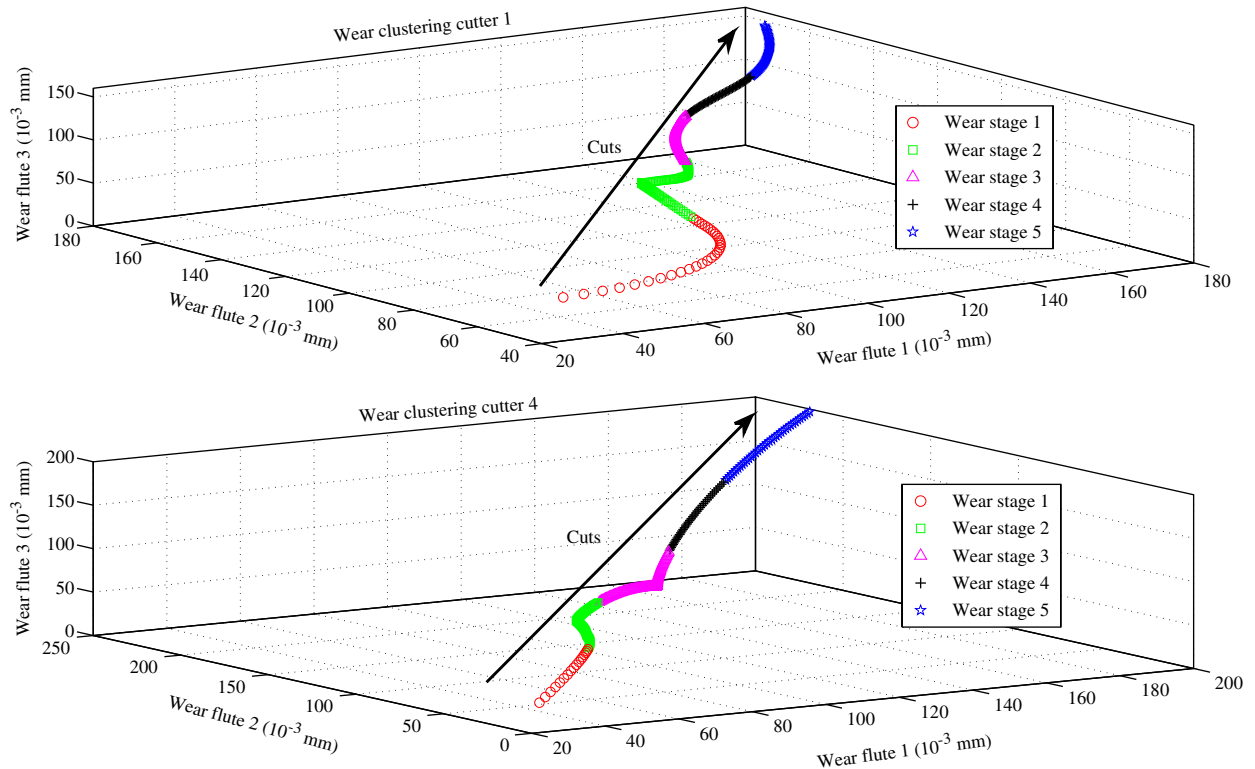


Fig. 10. The clustering results.

During the learning phase, five wear stages were defined to set the size of the global model wear base being consequent by the analysis of the number of tool degradation regions suggested in [38,39]. To classify the different features in the different wear stages, the wear file record of the three flutes of each cutter was classified by using the k-means algorithm. Fig. 10 shows the result of the number of cuts in each cluster for the training sets. For example, for the cutter 1 the features from the first 32 cuts belong to the first wear stage, then the next 126 cuts are classified in the wear stage 2, the 59 following cuts are in the wear stage 3, the next 51 are classified in the wear stage 4 and finally, the last 47 cuts are in the wear stage 5.

For the learning and prognostic phases, different features were extracted from the raw signals provided by the monitoring sensors by following the recommendations stated in [46], where the relevant features used in modeling machining operations are reported. For the dynamometer signals, the retained features were the root mean square (rms), the peak and the standard deviation (std), for the accelerometer signals the rms and the kurtosis were computed and finally, for the acoustic emission sensor the mean and the standard deviation were extracted. Thus, a total of 17 features were extracted for each cut. These features were chosen because they are more suitable for tracking the wear evolution [46].

The parameters of the DBNs (see Section 3.3) in the global and in the local model wear base were first randomly initialized and constrained to be a three states left-to-right MoG-HMM (representing the degradation phenomena) and then, the extracted features were fed to the learning algorithms in order to re-estimate the initialized

parameters (CPDs). The number of mixtures in each MoG-HMM represented by a DBN was set to 10 (each coefficient C_{jm} was uniformly initialized). In order to identify the correct number of mixtures to be used an analysis of representativeness vs number of mixtures was performed. The clustered data from cutters 1 and 4 were used to learn a global model for each wear region (1–5). Then, the corresponding likelihood was recorded after each learning phase by using the learning algorithms for DBN [11]. This learning phase was repeated using different number of mixtures and a plot of likelihood vs the number of mixtures for each global wear model was obtained. In Fig. 11 it can be observed that the likelihood of the models representing the different wear regions has tendency to stabilize after 10 mixtures. In order to ensure good representativeness and to minimize the learning time, 10 mixtures can be a good choice to allow a trade-off between precision and computation time (similar to the results found in [47]).

A total of 15 DBNs (five in the global model base and 10 in the local one) were obtained by using the learning algorithms. The initial and the re-estimated numerical values of the CPDs parameters related to S_1 , S_t and M_t associated to the global DBN for the first wear stage are given in Table 2.

As defined in Eqs. (9)–(12), it is also necessary to estimate the amount of wear obtained and its variation in each state of the models stored in the local wear model base. To give an example of this estimation, the corresponding decoding state sequence of the cuts belonging to the first wear stage (1–32 cuts) from the history associated to the cutter 1 is shown in Fig. 12. Here one can appreciate that the first six cuts belongs to the state 1, the following six to the state 2 and the last 30 to the state 3. Then, combining this information and Eqs. (9)–(12), the mean wear and the mean wear variation and their associated standard deviation were estimated, as presented in Table 3.

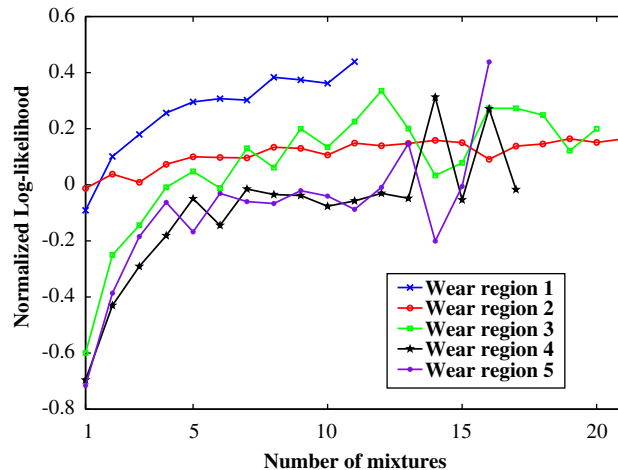


Fig. 11. Normalized likelihood vs number of mixtures for the different wear regions using the cutters 1 and 4 as training sets.

Table 2

Initial (I) and final (F) CPDs of the global DBN model for the first wear stage using the data from cutters 1 and 4 as training sets.

π_1	$\pi_1^I = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$,	$\pi_1^F = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$
S_t	$S_t^I = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \\ 0 & 0 & 1 \end{bmatrix}$,	$S_t^F = \begin{bmatrix} 0.78 & 0.22 & 0 \\ 0 & 0.71 & 0.29 \\ 0 & 0 & 1 \end{bmatrix}$
M_t	$M_t^I = \begin{bmatrix} 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \end{bmatrix}$	$M_t^F = \begin{bmatrix} 0 & 0.08 & 0.18 & 0.01 & 0.15 & 0.18 & 0.3 & 0 & 0.1 & 0.27 \\ 0 & 0.1 & 0 & 0.03 & 0.4 & 0.1 & 0.1 & 0.1 & 0.1 & 0.07 \\ 0.39 & 0.14 & 0.04 & 0.07 & 0.07 & 0.14 & 0.04 & 0 & 0.08 & 0.04 \end{bmatrix}$

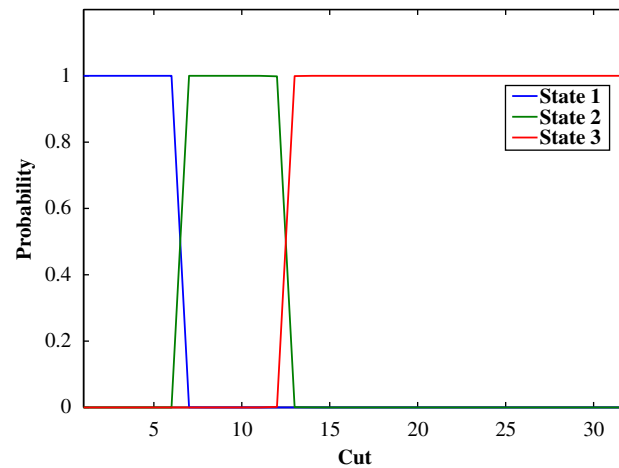


Fig. 12. The decoding sequence.

Table 3
Estimated wear parameters in 10^{-3} mm.

Parameter	State		
	S_1	S_2	S_3
$\mu(Wr_1^1)$	30.9867	43.5107	53.9789
$\sigma(Wr_1^1)$	11.2078	6.5399	4.9714
$\mu(\Delta Wr_1^1)$	2.9175	1.9883	1.1234
$\sigma(\Delta Wr_1^1)$	0.7350	0.3365	0.2054

In order to simulate an on-line wear monitoring case, the diagnostic and prognostic method presented in the previous section was applied on all data histories in order to estimate the wear after each cut and compute the corresponding RUL. The simulation results for all possible combinations of histories used to predict the mean wear after each cut are shown in Fig. 13, where the monitoring data used concern 315 cuts and the true wear at each flute was measured. The results using the data corresponding to the cutter 6 as “test” history are discussed hereafter. For the RUL prediction, it is supposed that the wear threshold is known in advance. In the benchmark used in this application no information was available about the surface finish requirements of the workpiece making impossible the definition of a “logic” wear limit. However, in order to be able to do prognostic and to evaluate the error of the proposed method a high threshold limit of 140×10^{-3} mm was considered, which represents an average tolerance in the surface roughness of the finished part (regular surface quality). In general, the roughness increases almost constantly with the increase of tool flank wear with a tendency to reach a plateau stage towards the end of tool life. In practice, it is recommendable to make a study to find the relation between the tool wear and the surface roughness as the analysis performed in [48] for a hard turning process of 1117 steel hardened up to 62 ± 1 HRC.

In the results presented in Fig. 13, the confidence bound for the RUL's estimation was fixed to 95%. This value was defined using the information presented in Table 4, where the mean absolute percentage error (MAPE) was estimated using the confidence levels defined by the three-sigma rule (68%, 95%, 99.7%) [49]. In this table one can appreciate that the MAPE is minimal when the confidence value is fixed at 95% for the three test histories.

In Fig. 13(f) one can see that the precision of the estimated RUL is near to the real value. However, the prediction of the RUL value is slightly over the real value because the confidence bounds were fixed at 95%. This confidence value gives wide limits leading to a little late prognostic, which can be corrected by introducing a security factor. This can be useful in a sense that it suggests a tool replacement before reaching the wear limit in order to protect the work piece and the CNC machine from damage and breakdowns.

The RUL predictions presented in Fig. 13 were also used to quantify the performance of the method using the accuracy metric (Eq. (18)) for failure prognostics techniques proposed by [4]. This measure is defined to be equal to 1 for the best performance and 0 for the worst. Table 5 summarizes the performance results of the three tests. The mean performance

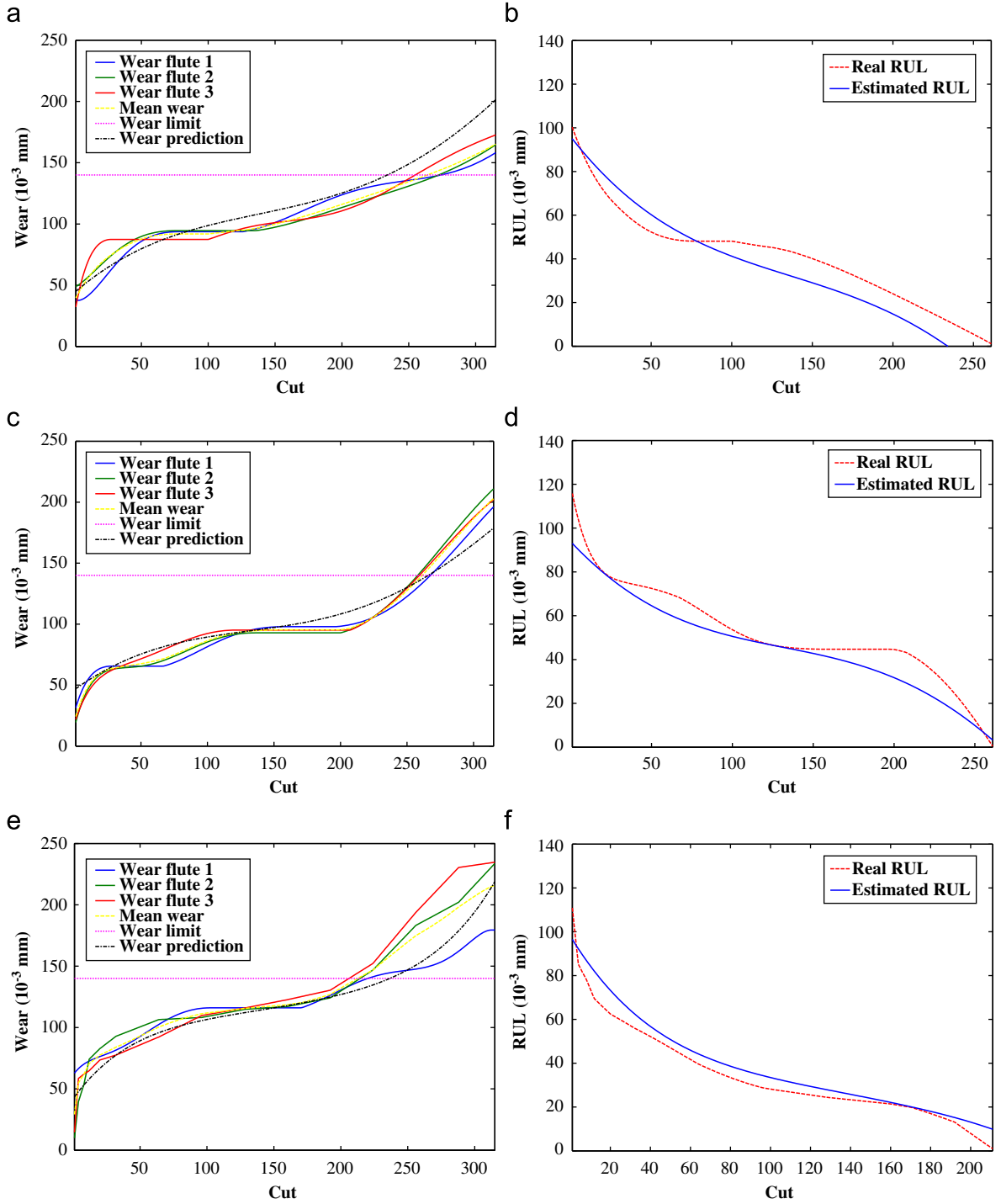


Fig. 13. Simulation results: (a) wear estimation for the cutter 1 using the data from cutters 4 and 6 as training sets; (b) RUL prediction for the cutter 1; (c) wear estimation for the cutter 4 using the data from cutters 1 and 6 as training sets; (d) RUL prediction for the cutter 4; (e) wear estimation for the cutter 6 using the data from cutters 1 and 4 as training sets; and (f) RUL prediction for the cutter 6.

accuracy value is 0.8234 which is near to 1.

$$Accuracy = \frac{1}{C} \sum_{c=1}^C e^{-|RUL_{Real}(c) - RUL_{Estimated}(c)| / RUL_{Real}(c)} \quad (18)$$

Table 4

MAPE comparison using the confidence levels from the three-sigma rule.

History	MAPE at the confidence bound ($n \cdot \sigma$)		
	68% ($n=1$)	95% ($n=2$)	99.7% ($n=3$)
Cutter 1	28.71	24.39	41.67
Cutter 4	20.98	15.71	37.12
Cutter 6	25.47	15.01	40.26

Table 5

Accuracy performance measure for the three tests.

Measure	History		
	Cutter 1	Cutter 4	Cutter 6
Accuracy	0.7286	0.8950	0.8467
Mean	0.8234		

5. Conclusion

A condition monitoring, diagnostic and prognostic data-driven method has been proposed in this paper. The main idea of the work relies on the transformation of the monitoring data into relevant models that capture the degradation's behavior of the machining tools. The choice of a data-driven method instead of a physical model of the degradation dwells in the fact that obtaining this latter may not a trivial task due to the complexity of the physical phenomenon of the wear, which is not easy to model. The utilization of MoG-HMMs allowed thus to model the wear in a stochastic way by taking into account the different stages of the degradation. The identification of the current condition of the tool combined with the models learned in the off-line phase allowed to calculate the RUL and the associated confidence value.

Future works concern the extension of the failure prognostic method by taking into account the variable operating conditions (load, velocity, temperature, etc.) and the integration of maintenance actions for RUL prediction.

References

- [1] X. Jiang, Recent development of damage diagnosis and prognosis in engineering systems, *Recent Patents Eng.* 4 (2) (2010) 1–20.
- [2] CEN/TC 319—Maintenance, Maintenance—maintenance terminology. EN 13306:2010, 2010.
- [3] V. Venkatasubramanian, Prognostic and diagnostic monitoring of complex systems for product lifecycle management: challenges and opportunities, *Comput. Chem. Eng.* 29 (6) (2005) 1253–1263.
- [4] G. Vachtsevanos, F.L. Lewis, M. Roemer, A. Hess, B. Wu, *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*, Wiley, 2006.
- [5] A.K. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, *Mech. Syst. Signal Process.* 20 (7) (2006) 1483–1510.
- [6] A. Heng, S. Zhang, A.C. Tan, J. Mathew, Rotating machinery prognostics: state of the art, challenges and opportunities, *Mech. Syst. Signal Process.* 23 (3) (2009) 724–739.
- [7] M. Lebold, M. Thurston, Open standards for condition-based maintenance and prognostic systems, in: *Maintenance and Reliability Conference (MARCON)*, 2001.
- [8] D.A. Tobon-Mejia, K. Medjaher, N. Zerhouni, G. Tripot, A mixture of Gaussians hidden Markov model for failure diagnostic and prognostic, in: *IEEE Conference on Automation Science and Engineering, CASE'10*, 2010.
- [9] X. Jiang, S. Mahadevan, An intelligent online damage detection system for thermal protection panels with active sensors, in: *Intelligent Sensors and Actuators Symposium at Earth & Space 2008 Conference*, 2008.
- [10] X. Jiang, S. Mahadevan, Wavelet energy-based Bayesian probabilistic approach for damage detection of thermal protection system panel, *Am. Inst. Astronaut. Astronaut. J.* 47 (4) (2009) 942–952.
- [11] K.P. Murphy, *Dynamic Bayesian Networks: Representation, Inference and Learning*, Ph.D. Thesis, University of California, 2002.
- [12] A. Muller, M.-C. Suhner, B. lung, Formalisation of a new prognosis model for supporting proactive maintenance implementation on industrial system, *Reliab. Eng. Syst. Saf.* 93 (2) (2008) 234–253.
- [13] W.Q. Wang, M.F. Golnaraghi, F. Ismail, Prognosis of machine health condition using neuro-fuzzy systems, *Mech. Syst. Signal Process.* 18 (4) (2004) 813–831.
- [14] C.S. Byington, M.J. Roemer, T. Galie, Prognostic enhancements to diagnostic systems for improved condition-based maintenance, *IEEE Aerospace Conference*, vol. 6, 2002, pp. 2815–2824.
- [15] D.A. Tobon-Mejia, K. Medjaher, N. Zerhouni, The ISO 13381-1 standard's failure prognostics process through an example, in: *IEEE — Prognostics & System Health Management Conference*, University of Macau, Macau, China, 2010.
- [16] AFNOR, Condition Monitoring and Diagnostics of Machines—Prognostics—Part 1: General Guidelines. NF ISO 13381-1, 2005.
- [17] G. Provan, Prognosis and condition-based monitoring: an open systems architecture, *IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*, vol. 5, 2003, pp. 57–62.
- [18] R.B. Chinnam, P. Baruah, A neuro-fuzzy approach for estimating mean residual life in condition-based maintenance systems, *Int. J. Mater. Prod. Technol.* 20 (2004) 166–179.
- [19] D.A. Tobon-Mejia, K. Medjaher, N. Zerhouni, G. Tripot, Estimation of the remaining useful life by using wavelet packet decomposition and HMMs, in: *IEEE Aerospace Conference*, 2011.

- [20] J. Luo, M. Namburu, K. Pattipati, L. Qiao, M. Kawamoto, S. Chigusa, Model-based prognostic techniques [maintenance applications], in: IEEE Systems Readiness Technology Conference (AUTOTESTCON), 2003, pp. 330–340.
- [21] D. Chelidze, J. Cusumano, A dynamical systems approach to failure prognosis, *J. Vibr. Acoust.* 126 (2004) 2–8.
- [22] A. Keller, U. Perera, A. Kamath, Reliability analysis of CNC machine tools, *Reliab. Eng.* 3 (6) (1982) 449–473.
- [23] P. Groer, Analysis of time-to-failure with a Weibull model, in: Proceedings of the Maintenance and Reliability Conference, MARCON, 2000.
- [24] G. Vachtsevanos, P. Wang, Fault prognosis using dynamic wavelet neural networks, in: IEEE System Readiness Technology Conference, Autotestcon Proceedings, 2001, pp. 857–870.
- [25] W. Wang, An adaptive predictor for dynamic system forecasting, *Mech. Syst. Signal Process.* 21 (2007) 809–823.
- [26] R. Huang, L. Xi, X. Li, C.R. Liu, H. Qiu, J. Lee, Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods, *Mech. Syst. Signal Process.* 21 (1) (2007) 193–207.
- [27] D.C. Swanson, J.M. Spencer, S.H. Arzoumanian, Prognostic modelling of crack growth in a tensioned steel band, *Mech. Syst. Signal Process.* 14 (1999) 789–803.
- [28] M. Orchard, B. Wu, G. Vachtsevanos, A particle filter framework for failure prognosis, in: Proceedings of the World Tribology Congress, 2005.
- [29] M. Dong, D. He, A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology, *Mech. Syst. Signal Process.* 21 (2007) 2248–2266.
- [30] R.B. Chinnam, P. Baruah, Autonomous diagnostics and prognostics through competitive learning driven HMM-based clustering, Proceedings of the International Joint Conference on Neural Networks, vol. 4, 2003, pp. 2466–2471.
- [31] H. Ocak, K.A. Loparo, F.M. Discenzo, Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: a method for bearing prognostics, *J. Sound Vibr.* 302 (2007) 951–961.
- [32] K. Przytula, A. Choi, An implementation of prognosis with dynamic Bayesian networks, in: Aerospace Conference, 2008, pp. 1–8.
- [33] K. Medjaher, J.-Y. Moya, N. Zerhouni, Failure prognostic by using dynamic Bayesian networks, in: 2nd IFAC Workshop on Dependable Control of Discrete Systems, 2009.
- [34] M. Dong, Z.-b. Yang, Dynamic Bayesian network based prognosis in machining processes, *J. Shanghai Jiaotong Univ. (Science)* 13 (2008) 318–322.
- [35] J.J. Hillhouse, C.M. Adler, Investigating stress effect patterns in hospital staff nurses: results of a cluster analysis, *Soc. Sci. Med.* 45 (12) (1997) 1781–1788.
- [36] R. Landgraf, I. Xenarios, D. Eisenberg, Three-dimensional cluster analysis identifies interfaces and functional residue clusters in proteins, *J. Mol. Biol.* 307 (5) (2001) 1487–1502.
- [37] J. Abonyi, B. Feil, Cluster Analysis for Data Mining and System Identification, Birkhäuser, Basel, 2007.
- [38] H.M. Ertunc, C. Oysu, Drill wear monitoring using cutting force signals, *Mechatronics* 14 (5) (2004) 533–548.
- [39] T. Liu, K. Anantharaman, Intelligent classification and measurement of drill wear, *J. Eng. Ind.* 116 (1994) 392–397.
- [40] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proc. IEEE* 77 (2) (1989) 257–286.
- [41] B.H. Juang, Maximum likelihood estimation for mixture multivariate stochastic observations of Markov chains, *AT&T Tech. J.* 64 (1985) 1235–1249.
- [42] K.D. Vohnout, Curve fitting and evaluation, in: *Mathematical Modeling for System Analysis in Agricultural Research*, Elsevier Science, Amsterdam, 2003, pp. 140–178.
- [43] A. Viterbi, Error bounds for convolutional codes and an asymptotically optimal decoding algorithm, *IEEE Trans. Inf. Theory* 13 (1967) 260–269.
- [44] S. Kurada, C. Bradley, A review of machine vision sensors for tool condition monitoring, *Comput. Ind.* 34 (1997) 55–72.
- [45] PHM Society, PHM data challenge 2010 <<https://www.phmsociety.org/competition/phm/10>>, 2010.
- [46] J.V. Abellan-Nebot, F. Romero Subiron, A review of machining monitoring systems based on artificial intelligence process models, *Int. J. Adv. Manuf. Technol.* 47 (2010) 237–257.
- [47] K. Zhu, Y.S. Wong, G.S. Hong, Multi-category micro-milling tool wear monitoring with continuous hidden Markov models, *Mech. Syst. Signal Process.* 23 (2) (2009) 547–560.
- [48] R. Pavel, I. Marinescu, M. Deis, J. Pillar, Effect of tool wear on surface finish for a case of continuous and interrupted hard turning, *J. Mater. Process. Technol.* 170 (2005) 341–349.
- [49] V. Z. Aladjev, V. N. Haritonov, *General Theory of Statistics*, Fultus, 2004.