The International Journal of Advanced Manufacturing Technology (2019) 101:2861–2872
https://doi.org/10.1007/s00170-018-3157-5

ORIGINAL ARTICLE

Hybrid data-driven physics-based model fusion framework for tool wear prediction

Houman Hanachi[1] · Wennian Yu[2] · Il Yong Kim[2] · Jie Liu[3,4] · Chris K. Mechefske[2]

Received: 7 September 2018 / Accepted: 4 December 2018 / Published online: 12 December 2018

Abstract

An integral part of modern manufacturing process management is to acquire useful information from machining processes to monitor machine and tool condition. Various models have been introduced to detect, classify, and predict tool wear, as a key parameter of the machining process. In more recent developments, sensor-based approaches have been attempted to infer the tool wear condition from real-time processing of the measurement data. Experiments show that the physics-based prediction models can include large uncertainties. Likewise, the measurement-based (or sensor-based) inference techniques are affected by sensor noise and measurement model uncertainties. To manage uncertainties and noise of both methods, a hybrid framework is proposed to fuse together the results of the prediction model and the measurement-based inference data in a stepwise manner. The fusion framework is an extension to the regularized particle filtering technique, used to facilitate updating the state prediction with a numerical inference model, when measurement models alone are not satisfactory. The results show significant improvement in tool wear state estimation, reducing the prediction errors by almost half, compared to the prediction model and sensor-based monitoring method used independently.

Keywords Tool wear · Sensor-based monitoring · Particle filter · Fusion framework

Nomenclature

A, B, C	Prediction model parameters
ANFIS	Adaptive neuro-fuzzy inference system
ANN	Artificial neural network
D	Dimension of state vector
d	Cutting depth
e	Measurement noise
F	State function.
f	Probability density function
f_r	Feed rate
G	Feature model
$g(.)$	Importance density
H_y	Measurement model
h	Kernel bandwidth
$K(.)$	Kernel density
k	Time-step
M	Feature mapping
m	Number of particles
NRMSE	Normalized root mean squared error
PF	Particle filter
RPF	Regularized particle filter
t	Time
t_c	Cutting time
U	Input history
u	Input condition
v	Cutting speed
x	Tool wear state
y	Measurement signal
Z	Feature history
z	Feature vector
α	Time step ratio
Δ	Signal processing time window
Δt	Length of time-step

✉ Wennian Yu
wennian.yu@queensu.ca

[1] Life Prediction Technologies Inc. (LPTi), Ottawa K1J 9J1, Canada

[2] Department of Mechanical and Materials Engineering, Queen's University, Kingston, ON K7L 3N6, Canada

[3] National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, Chongqing 400067, China

[4] Department of Mechanical and Aerospace Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada

$\delta(.)$	Dirac delta function
ε	Feature model error
θ	Momentary time variable
τ	Process noise
Ω	Particle weight
ω	Normalized particle weight

1 Introduction

Metal cutting is a common manufacturing process in industry. It removes excess material from the surface of a workpiece using various kinds of cutting tools in order to bring the workpiece into a specified geometry [1]. However, the unavoidable development of cutting tool wear, especially flank wear, adversely affects the accuracy and surface finish of machined products. Therefore, the development of a robust and reliable online tool condition monitoring system for the prediction of tool wear is crucial in industrial applications. It can facilitate the timely decision for tool change, and yield higher productivity and significant cost savings for manufacturers.

In the literature, a significant amount of research work has been performed dealing with cutting tool wear prediction. Various methods and models have been proposed which can be classified into two categories: physics-based approaches (prediction models) and sensor-based approaches (measurement-based inference models). The former identifies the degradation function of the tool wear though physics-based laws or semi-empirical expressions from which the tool wear is predicted. The latter, on the other hand, use specific optical sensors to directly measure the tool wear, or use multiple types of general sensors to indirectly infer the tool wear based on data-driven models trained using sufficient historical data.

The physics-based approaches, which rely on the mathematical description of the physics of cutting, assumes certain wear mechanisms (abrasion, adhesion, and diffusion) as being responsible. Thus, many researchers have sought to explore the physical process of wear. Different researchers have proposed different tool wear models analytically or empirically, which were summarized by Yen et al. [2] and Palmai [3]. However, it is well accepted that the cutting tool wear originates mainly from two wear mechanisms: mechanically activated processes (abrasion, adhesion) and thermally activated processes (diffusion) [2–7]. They are generally present in combination [8]. The former is dominant at lower cutting speeds, i.e., lower cutting temperatures, and is mainly dependent on the cutting length and time, lubrication conditions, contact loads between the tool and workpiece, and the tool-work material combination. The latter becomes influential as the cutting temperature increases [2, 3]. However, due to the inherent complexity and highly nonlinear nature of machining processes, there is no general agreement in the literature as to an appropriate analytical model for the tool wear process [3, 9].

On the other hand, instead of struggling on deriving accurate analytical tool wear models, many researchers have tried to determine empirical or semi-empirical functions to fit the relationship between the tool wear with several key cutting parameters based on practical experience and measurements [3]. Taylor tool life and its extended versions are well-known semi-empirical functions employed in machining applications which describe the relationship between tool life with cutting parameters like cutting speed and feed rate [2, 10]. Various purely empirical expressions were also reported in the literature, which directly describe the amount of tool wear as a function of the cutting time based on experience of the wear-time curve [11–13]. It was reported by Palmai that the wear-time formula developed by Sipos in [13] was the best match in modeling the measurement results [3].

The sensor-based approaches provide alternative ways to estimate the cutting tool wear. They can be divided up into direct and indirect methods [14]. The former directly measures the amount of tool flank wear using optical sensors, i.e., microscopes or computer vision systems, or micro-isotope sensors. These methods can provide accurate tool wear estimates; however, they are inherently offline, meaning that they cannot be used while cutting. The latter estimates the tool wear state from various process parameters which are indirectly correlated with tool wear, e.g., cutting force, acoustic emission, vibration, and motor current. Obviously, these methods can be used for real-time tool wear determination, and consequently, they have considerable potential to be used in automated machining systems. Therefore, sensor-based approaches for indirect tool wear prediction are becoming popular for online tool wear monitoring. Data-driven models such as support vector regression [15], artificial neural networks [9, 14], and neuro-fuzzy inference systems [16–18] are often employed to model the non-linear dependencies between features extracted from the sensor signals and cutting conditions. Generally, there are three main steps with data-driven approaches for cutting tool wear prediction [14, 19]: (1) use of single or multiple types of sensor to capture process information, i.e., sensor fusion; (2) extraction of sensitive features from the sensor information to reflect the tool condition, i.e., feature extraction; and (3) development of reliable and robust decision-making models using extracted features to predict tool conditions. Recently, the concept of "intelligent machine tools" has become popular, which is described as a machine tool with capabilities of sensing and decision-making to guarantee the optimum machining process [20]. Since 2006, deep learning has been a rapidly growing research area, and is adopted by many researchers as a bridge connecting multi-sensor time-series data and intelligent machine health monitoring [21]. Specifically, Malhotra et al. proposed a recurrent neural network (long short-term memory) based encoder-decoder scheme to obtain

an unsupervised health index from the multi-sensor time-series data collected from a milling machine [22].

Physics-based approaches suffer from the rarity of accurate analytical models to describe tool wear processes due to the inherent complexity of the cutting process and our incomplete understanding of it. Various forms of empirical formulae have been introduced to predict the tool wear in the literature but are still limited in applicability since these prediction models can include large uncertainties. Although sensor-based data-driven approaches for tool wear prediction have attracted considerable attention in the literature and have been demonstrated to yield satisfactory prediction accuracy in different machining applications such as milling, turning, and grinding, data-driven models require sufficient historical data for training. Besides, the prediction accuracy of sensor-based methods is highly affected by the sensor noise and measurement uncertainties.

The physics-based prediction models are associated with modeling uncertainty. Similarly, the measurement-based inference models include modeling errors and are affected by measurement noise. To improve the tool wear estimation results, hybrid techniques fuse together the results from both approaches within a sequential data-driven physics-based model fusion framework (data-model fusing frame work). Compared to the individual approaches, the fusion framework is a relatively unexplored methodology in the field of tool wear prediction. For linear systems with Gaussian noise and modeling uncertainty, Kalman Filter provides the optimum results in a closed form [23]; however, for nonlinear non-Gaussian systems, its application becomes limited. Thanks to powerful computational resources, simulation-based approaches such as particle filter (PF) have been successfully applied on different data-model fusion problems with nonlinear state prediction models and unknown noise structure. For the tool wear estimation problem, PF has been recently used in a few research works, mostly by Wang et al. [24–27]. The shortcoming with the generic PF is that the state prediction and the measurement models do not take the stochastic inputs of the system into the modeling structure [28]. Considering the uncertainties of the measurements on the system inputs, i.e., the machining condition, in this paper, we have proposed a generalized PF with the ability to take in the stochastic system inputs. The framework is an extension to the regularized particle filtering (RPF) to facilitate updating the state prediction with a numerical inference model, when lacking measurement models. The results show significant improvement in tool wear state estimation with prediction errors almost cut in half, compared to the prediction model and sensor-based monitoring method.

The rest of the paper is organized as follows. The tool wear data used in this study are introduced in Section 2. In Section 3, the inference and prediction models are developed and explained. Section 4 lays out the extended structure of PF using inference model to address the stochastic system inputs. The implementation results and the conclusions are presented in Sections 5 and 6, respectively.

2 Milling data set

The experiment was done on a milling machine which describes the tool wear process under various operating conditions [28]. The workpiece material is cast iron. The cutting speed was constant at 200 m/min (826 rev/min). Two different depths of cut (1.5 mm and 0.75 mm), and two feed rates (0.5 mm/rev and 0.25 mm/rev) were investigated. The experimental matrix is 2×2. Hence there are four different cutting conditions, as shown in Table 1. The experiment under each condition was done a second time with a second set of inserts. Therefore, there are eight cases with a variable number of runs per tool life. The number of runs was dependent on the degree of flank wear that was measured between runs at irregular intervals up to a set wear limit. There is a total of 108 data samples (Table 1).

A current sensor was used to acquire the spindle motor current of the milling machine for each run in the experiment. The tool flank wear VB was measured offline using a microscope between runs as a generally accepted parameter for evaluating tool wear [18, 29]. Since the data were collected from a real milling machine instead of simulation models or experimental platforms, the cutting tool wear processes reflected in this dataset are realistic. It provides important information for researchers to study the relationship between health states and measurements [22, 30, 31].

2.1 Tool wear measurement

Figure 1 shows the variations in the amount of tool flank wear over machining time for the four cutting conditions and two sets of inserts arranged sequentially, which has a total of eight segments. It can be noticed from that the amount of tool flank wear generally increases monotonically over the machining time, which is expected. However, at some instances, there are obvious irregularities in the tool wear measurements as shown by the red circles in the figure, which can be attributed

Table 1 Milling dataset [29]

Cutting condition	Depth of cut (d, mm)	Feed (f_r, mm/rev)	Number of runs	
			1st insert	2nd insert
#1	1.5	0.5	17	9
#2	1.5	0.25	7	10
#3	0.75	0.5	14	14
#4	0.75	0.25	14	23

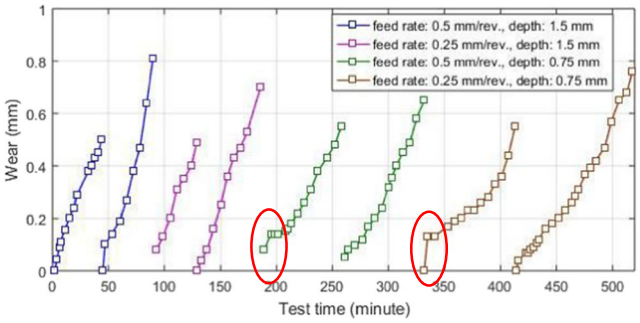

Fig. 1 Tool wear measurement data

to the measurement uncertainty. It can also be noticed that as the machining burden, i.e., the depth of cut or the feed rate decreases, the machining time of the tool generally tends to be longer, meaning that the tool wears out more slowly.

2.2 Spindle motor current and feature extraction

It has been reported by several researchers that the cutting force is the most sensitive parameter to tool condition [19, 32, 33]. It normally increases with wear of tool. However, force-based instrumentation and measurements are costly and comparatively difficult to set up when minimum interference to the cutting process is desired. The power-based measurements are therefore typically used in practice as a cost-effective online monitoring solution [32]. This is expected as the spindle motor current is well correlated with the cutting force, and consequently, it is sensitive to the tool wear.

In this study, we select 10 common signal features from the spindle motor current as listed in Table 2. They include four time-domain dimensional statistical features (peak to peak, root mean square, standard deviation, peak); two time-domain dimensionless features (kurtosis, crest factor); two frequency-domain features (amplitude of the first harmonic, average of the first six harmonic amplitudes); and two statistical features based on the β-distribution (skewness and kurtosis based on the β-distribution). The first eight are classic features typically used in the condition monitoring of mechanical systems. The last two were first introduced by Whitehouse [34], and they were utilized for tool wear monitoring by Kannatey and Dornfeld [35]. Figure 2 shows variations of the 10 features extracted from the spindle current with machining time, corresponding to the tool wear measurements as shown in Fig. 1.

3 Inference and prediction models

There are two approaches for estimating tool wear. Within the first approach, the objective is to predict the future magnitudes of tool wear based on the operating condition and the cutting time. The second approach attempts to infer the tool wear magnitude using the features extracted from the real-time measurement signals from the system.

3.1 Physics-based tool wear prediction

Various analytical models have been suggested for predicting tool wear and wear rate prediction, as surveyed in Section 1. It is known that the abrasive/adhesive effects influence the thermally activated wear process, and at the same time, the wear of the flank leads to a temperature rise on the tool. This phenomenon is a positive feedback process leading to instability of the tool wear state until the end of tool life. In this work, we use the empirical wear-time model proposed by Sipos [13]. At a given cutting speed, feed rate, and cutting depth, the model predicts the tool wear as a function of time.

$$x(t_c) = t_c \exp\left(A + Bt_c + Ct_c^2\right), \tag{1}$$

where x is the flank wear depth, t_c is the cutting time, and A, B, and C are the model parameters.

With the available measurement data, we train the model on the corresponding machining condition and use the trained model to predict the tool wear for an independent part of data with the same machining condition. Figure 3 shows the training and testing data for two different cutting conditions, i.e., cutting conditions #1 and #4 as shown in Table 1. Figure 3a, b correspond to the cutting depth of 1.5 mm and the feed rate of 0.5 mm/rev, whereas Fig. 3c, d correspond to the cutting depth of 0.75 mm and the feed rate of 0.25 mm/rev. The Sipos model was trained on one part and tested on the other part reciprocally. Take Fig. 3a as an instance, the measurement data from the second insert (i.e., training window shown in Fig. 3a) were used to train the Sipos model, and the predictions were made

Table 2 Features extracted from spindle motor current

Index	Feature	Index	Feature
1	Peak to peak (pp)	6	Crest factor (cf)
2	Root mean square (rms)	7	Amplitude of first harmonic (pks1)
3	Standard deviation (std)	8	Average of the first six harmonics (pksa)
4	Peak (max)	9	Skewness based on β-distribution (skwβ)
5	Kurtosis (kur)	10	Kurtosis based on β-distribution (kurβ)

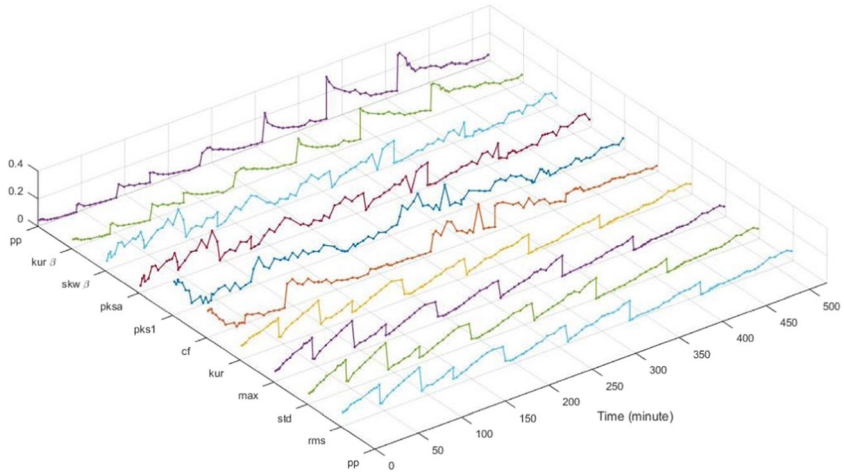

Fig. 2 Measurement signal features

on the tool wear of the first insert (i.e., modeling window shown in Fig. 3a). The same model fitting process was repeated for all eight data segments and the time-based wear prediction models are acquired. It can be found that Sipos model fits well on the training data; however, the prediction results on the test data include certain errors, especially for the tool wear predictions in the end-stage.

3.2 Measurement-based tool wear inference

Tool wear alters the cutting condition and leads to variation of the cutting force and the vibrational signatures. The causality relation between the wear, as a fault of the component and resulting signatures in the extracted signal features is depicted in Fig. 4. With limitation of access to direct measurement during the cutting process, we attempt to infer the magnitude of the wear using the signal features, i.e., the measurable symptoms of the fault. The so-called inference process is shown in Fig. 4, starting from the fault symptoms, and ending to the estimated fault magnitude.

Considering the dimensionality of the selected features and their nonlinear behavior with cutting time, a trainable numerical model can be a good candidate for the instantaneous inference modeling. For this mapping problem, the adaptive neuro-fuzzy inference system (ANFIS) structure [36] is employed for two reasons: we can balance between the degree of function fitting and the expected generalization accuracy by choosing an appropriate number of membership functions, and the high repeatability of the ANFIS model over repeated training trials, unlike artificial neural networks (ANN).

The selected features in Section 2.2 are the inputs and the estimated tool wear at the same time is as the output of the model. Figure 5 shows the structure of the ANFIS model, based on the following procedure:

– In the input layer, the machining condition vector includes the feed rate, the depth of cut, and the cutting

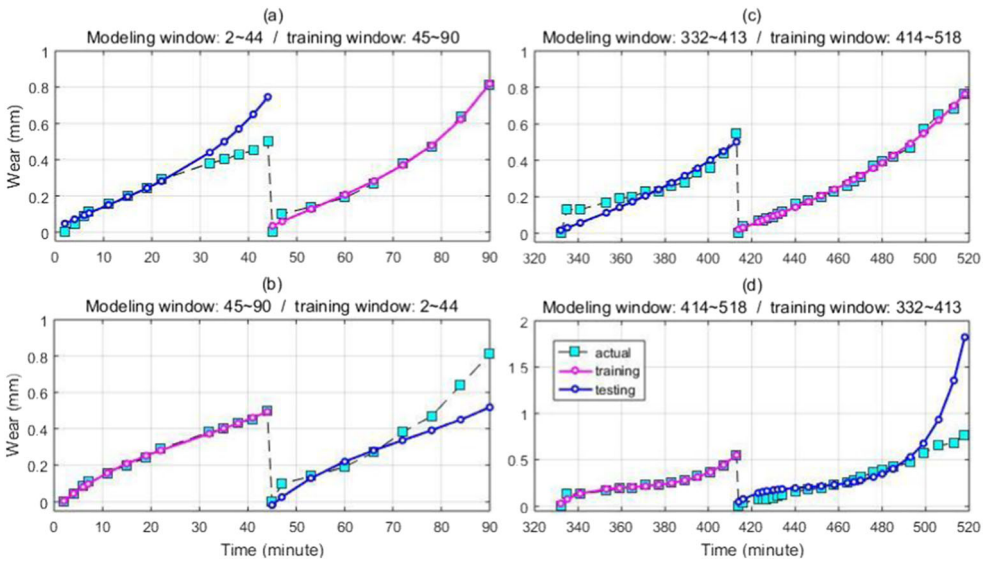

Fig. 3 Tool wear prediction model fitting **a**, **b** with 1.5-mm cutting depth and 0.5 mm/rev feed rate and **c**, **d** with 0.75-mm cutting depth and 0.25 mm/rev feed rate (note: measurements in the training window were used to train the Sipos model, whereas predictions were made on the tool wear in the modeling window using the trained model)

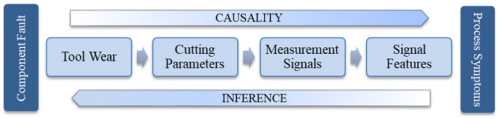

Fig. 4 Tool wear phenomenon leading to observable symptoms

speed: $u = (f_r, d, v)$. The measurement feature vector is z = (pp, rms, std, max , pks1, pksa, skw, kur), as per the selected features in Section 2.2.

- For the fuzzy inference process, a first-order Sugeno model is employed in the structure.
- In the second layer, generalized bell-shaped membership functions are used and two membership functions have been considered for each variable, as the design parameter of the model.
- Linear fuzzy rules are used in the third layer for the possible combination of the membership functions.
- The weights of the rules are generated in layer four, as the results of multiplication of membership function outputs.
- Summation of the fuzzy rule outputs, weighted with the corresponding weights in layer four, yield the inference result on the tool wear in layer five.

The training data are introduced to the model in the training process so that the internal parameters of the model, i.e., the membership functions and the fuzzy rules parameters, are set up. The model can then be used for the tool wear estimation, upon feeding the inputs.

With the available tool wear data and the corresponding signal features, the ANFIS model is trained on seven tool wear data segments and tested on the remaining one segment. Figure 6 shows three cases of training and testing of the model. Take Fig. 6a as an instance, the measurements except the second segment (as shown in Fig. 1) are used to train the ANFIS model, and the tool wear prediction based on the trained model is tested on the second segment. It can be found that the ANFIS model fits well on the training data; however, the inference results on test data include larger errors as especially shown in Fig. 6c.

4 Fusion framework development

Tool wear estimation based on the instantaneous measurements of the sensors includes errors due to the measurement noise and the inference model uncertainty. Similarly, the physics-based tool wear rate prediction results may contain errors due to model uncertainty and the stochastic nature of the machining condition. Based on the two available sources for tool wear prediction, the results of either approach can be fused together within a sequential recursive algorithm.

4.1 State estimation with data-model fusion

During the cutting process, from the dynamic system point of view, the combination of the work-piece and the machine-tool can be considered as the system in this study, where the tool wear level is the hidden state vector x of the system. The machining condition, including the cutting speed and depth and the feed rate, forms the variable input vector u of the system. The sensor(s) signal(s) is the system output measurement vector y and it can be estimated as:

$$y(t) = H_y(x(t), u(t)) + e \tag{2}$$

where, H_y is the measurement model of the system and e is the measurement noise, including the measurement noise and modeling uncertainty.

Feature extraction is a mapping process from the measurement space to the feature space. Feature vector z can be extracted through post-processing the measurement signal over the immediate elemental time window [Δ].

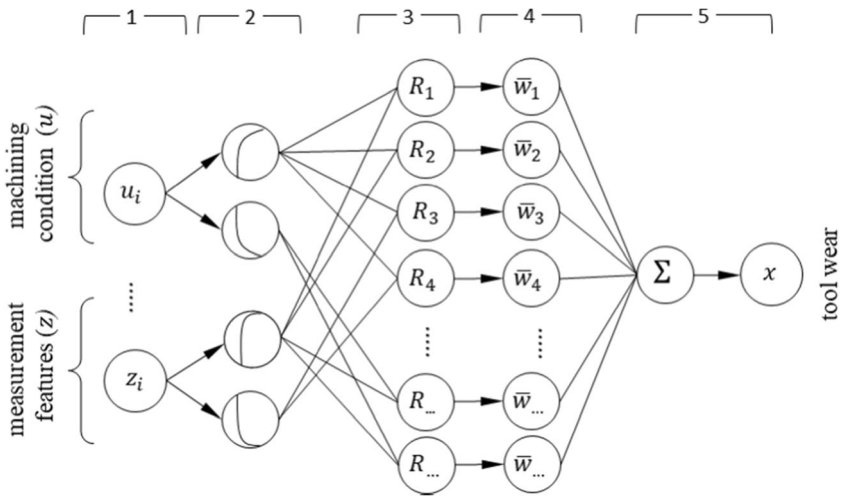

Fig. 5 ANFIS model to estimate the tool wear, given the machining condition and the extracted features from the measurements

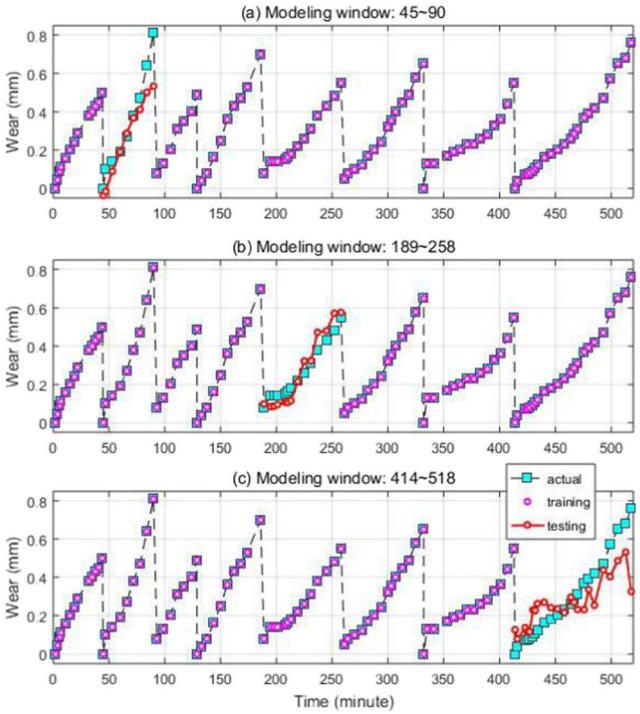

Fig. 6 Tool wear inference with ANFIS model, testing on the **a** second segment, **b** fifth segment, and **c** eighth segment as shown in Fig. 1, and training on the remaining segments

$$z(t) = M\left(y(\theta)|_{\theta=t-\Delta}^{t}\right). \tag{3}$$

From Eqs. 1 and 2, the feature model G is defined as:

$$z(t) = G(x(t), u(t)) + \varepsilon. \tag{4}$$

where ε is the feature model error resulting from the measurement error. For a given material and machine-tool setup, the rate of tool wear depends on the machining condition u, as explained in Section 3.2.

$$\dot{x}(t) = F(x(t), u(t)) + \tau. \tag{5}$$

Tool wear can be assumed as a first-order Markov process, where the rate of wear depends on the current state of the wear and the machining condition. The state space model at a time step k is therefore:

$$z_k = G(x_k, u_k) + \varepsilon_k, \tag{6}$$

$$x_k = F(x_{k-1}, u_k) + \tau_k. \tag{7}$$

In Section 3.1, it was explained that the tool wear can be inferred from the instantaneous measurements of the system. The inverse of the feature model represents the inference model.

$$x_k = G^{-1}(u_k, z_k) + \varepsilon'_k, \tag{8}$$

We assume the past measurements on the machining setup condition $U_k \triangleq \{u_1, u_2, \ldots, u_k\}$ and the output features $Z_k \triangleq \{z_1, z_2, \ldots, z_k\}$ are available from the records. At the same time, the initial tool wear condition and the current machining condition have known distributions as $f_x(x_0)$ and $f_u(u_k)$, respectively. From Bayes' theorem, the current tool wear level x_k, given the updated records of the measurements U_k and Z_k, can be calculated as:

$$f_x(x_k|U_k, Z_k) = \frac{f_{u,z}(u_k, z_k|x_k) f_x(x_k|U_{k-1}, Z_{k-1})}{f_{u,z}(u_k, z_k|U_{k-1}, Z_{k-1})}. \tag{9}$$

where,

$$f_{u,z}(u_k, z_k|x_k) = \frac{f_z(z_k|x_k, u_k)}{f_x(x_k)} \int f_x(x_k|x_{k-1}, u_k) f_x(x_{k-1}) dx_{k-1}, \tag{10}$$

$$f_x(x_k|U_{k-1}, Z_{k-1}) = \int f_x(x_k|x_{k-1}) f_x(x_{k-1}|U_{k-1}, Z_{k-1})\, dx_{k-1}, \tag{11}$$

$$f_{u,z}(u_k, z_k|U_{k-1}, Z_{k-1}) = \int f_{u,z}(u_k, z_k|x_k) f_x(x_k|U_{k-1}, Z_{k-1}) dx_k. \tag{12}$$

As a property of the first order Markov process, the current tool wear level only depends on the state in the previous step and the current system input, as given in Eq. 7. The previous state x_{k-1} is independent from the current input u_k. Therefore, the probability and the marginal probability of $f_x(x_k)$ and $f_x(x_k|x_{k-1})$ in Eqs. 10 and 11 can be found as:

$$f_x(x_k) = \iint f_x(x_k|x_{k-1}, u_k) f_x(x_{k-1}) f_u(u_k)\, \mathrm{d}x_{k-1}\, \mathrm{d}u_k \tag{13}$$

and

$$f_x(x_k|x_{k-1}) = \int f_x(x_k|x_{k-1}, u_k) f_u(u_k)\, \mathrm{d}u_k. \tag{14}$$

Starting from the first step, upon receiving the new measurements, the posterior probability density can be found sequentially for the corresponding time steps by Eqs. 9–14. Figure 7 shows the process of data-model fusion for stepwise tool-wear prediction as laid out above.

For linear state space models with Gaussian noise, the optimal solution for the state probability can be found in closed form by Kalman Filter [23]. For nonlinear non-Gaussian systems, the particle filter (PF) provides a computational platform for state estimation. In generic PF, the state space model lacks the effects of system input. In other words, the state evolution model in Eq. 7 shall not include u_k as an input to the model and consequently, with generic PF, the state evolution model should either be downgraded to an autonomous model with no independent input, or the independent input (u_k) should be treated as a deterministic variable [37]. In a previous work, the authors have extended PF framework to the general class of

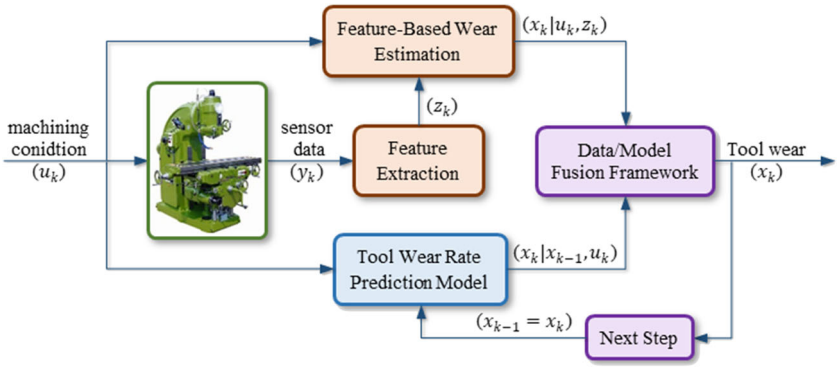

Fig. 7 Data-model fusion framework for tool-wear prediction

non-autonomous dynamic systems in order to address the effects of stochastic inputs [37, 38]. Due to the effect of the machining condition on the tool-wear process, the system input cannot be ignored in the state evolution model, and the above-mentioned extended PF framework for non-autonomous dynamic system is adopted in this work.

4.2 PF on systems with stochastic input

In PF technique, the probability density functions are constructed discretely with weighted state representatives, known as particles. The posterior density in Eq. 9 is therefore presented as:

$$f_x(x_k|U_k,Z_k)\approx\sum_{i=1}^{m}\omega_k^i\delta\left(x_k-x_k^i\right), \tag{15}$$

where $x_k^1,\ldots,x_k^m$ and $\omega_k^1,\ldots,\omega_k^m$ are the particles and their corresponding weights, respectively. For a continuous representation of the posterior density for a D-dimensional state, the rescaled kernel function $K_h(x)=h^{-D}K(x/h)$ can be used, where $h>0$ is the bandwidth [39].

$$f_x(x_k|U_k,Z_k)\approx\sum_{i=1}^{m}\omega_k^i K_h\left(x_k-x_k^i\right). \tag{16}$$

Continuous representation of the density functions for updating the particle weights leads to a revised type of PF, known as regularized particle filter (RPF) [40].

To construct the posterior density, the particles x_k^i are propagated from the previous time step using the state prediction model in Eq. 7, and the corresponding weights are updated upon the new measurements on u_k and z_k. From Bayes' theorem, the weights can be updated as:

$$\omega_k^i\propto\omega_{k-1}^i f_{u,z}\left(u_k,z_k|x_k^i\right)f_x\left(x_k^i|x_{k-1}^i\right)/g_x\left(x_k^i|x_{k-1}^i,u_k,z_k\right). \tag{17}$$

$g_x\left(x_k^i|x_{k-1}^i,u_k,z_k\right)$ is the importance density and it is chosen as a design decision of the filter. In generic PFs, $f_x\left(x_k^i|x_{k-1}^i\right)$ is taken as the importance density, that leads to:

$$\omega_k^i\propto\omega_{k-1}^i f_{u,z}\left(u_k,z_k|x_k^i\right). \tag{18}$$

The likelihood density function $f_{u,z}$ needs to be found over the multi-dimensional space of the system input and output measurements. This is computationally too expensive, even if the joint distributions of all measurements are available. With the objective of deriving a solution with a lower dimensionality for the likelihood density function, the importance density function is suggested as $g_x\left(x_k^i|x_{k-1}^i,u_k,z_k\right)=f_x\left(x_k^i|x_{k-1}^i\right)/f_x\left(x_k^i\right)$. The weight update Eq. 17 is then:

$$\omega_k^i\propto\omega_{k-1}^i f_{u,z}\left(u_k,z_k|x_k^i\right)f_x\left(x_k^i\right). \tag{19}$$

Using Bayes' theorem, for a given particle at time step k, the weight is therefore found as:

$$\omega_k^i\propto\omega_{k-1}^i f_x\left(x_k^i|u_k,z_k\right). \tag{20}$$

Equation 20 enables updating the particle weights upon receiving new measurements on the inputs and outputs of the system, using the marginal probability of the particles. In this research, the state probability density is a one-dimensional function for the tool-wear level. The process of RPF with the proposed modified particle weight updating technique is presented in Table 3.

5 Implementation results and analysis

In this section, the developed fusion framework is applied for tool-wear state estimation, and the results are compared with those of the inference and the prediction models of Section 3.

For the stepwise state estimation process, the discretized form of the state prediction model is derived from Eq. 1:

$$x_k=(1+\alpha_k)\exp\left[B\Delta t_k+C\left(\frac{2}{\alpha_k}+1\right)\Delta t_k{}^2\right]x_{k-1}, \tag{21}$$

where $\Delta t_k=t_k-t_{k-1}$ and $\alpha_k=\Delta t_k/t_{k-1}$ are the time step and the time step ratio in the equation. Eq. 21 predicts the state with multiplication of an updating factor to the previous state. This function is highly sensitive to the errors, as it accumulates the error at each step and carries forward. To mitigate

Table 3 RPF with modified particle weight updating technique

$\{(x_k^i, \omega_k^i)\}_{i=1:m} = RPF\left[\{(x_{k-1}^i, \omega_{k-1}^i)\}_{i=1:m}, u_k, z_k\right]$	
• Draw state prediction error	$\tau_k^i \sim f_\tau(x_k, u_k)$
• Propagate priors (Eq. 6)	$x_k^i = F(x_{k-1}^i, u_k) + \tau_k^i$
• Infer the expected state (Eq. 7)	$\bar{x}_k = G^{-1}(u_k, z_k)$
• Update particle weights	$\Omega_k^i = f_{\varepsilon'}(x_k^i - \bar{x}_k)$
• Normalize the weights	$\omega_k^i = \Omega_k^i / \sum_{i=1}^{m} \Omega_k^i$.
• Regularize posterior density	$f_x(x_k \mid u_k, z_k) \approx \sum_{i=1}^{m} \omega_k^i K_h(x_k - x_k^i)$.
• Resample	$\{(x_k^i, \omega_k^i = 1/m)\}_{i=1:m} \sim f_x(x_k \mid u_k, z_k)$

accumulation of the error, the prediction model in Eq. 1 is expanded by the Taylor series and the first term is utilized as the estimate for the state prediction model. The discretized result, as shown in Eq. 22, is employed for stepwise state prediction in the developed framework.

$$x_k = x_{k-1}\left[1 + \left(1 + Bt_{k-1} + 2C{t_{k-1}}^2\right)\alpha_k\right], \tag{22}$$

The ANFIS model from Section 3.2 uses the features from the latest measurement to infer the expected state as required for Eq. 8:

$$\bar{x}_k = \mathrm{ANFIS}(u_k, z_k), \tag{23}$$

where $u_k = [f_r, d]_k$ and $z_k = [\mathrm{pp}, \mathrm{rms}, \mathrm{std}, \max, \mathrm{skw}, \mathrm{cf}, \mathrm{pks1}, \mathrm{pksa}, \mathrm{skw}\beta, \mathrm{kur}\beta,]_k$ are the machining setup and the selected features at time step k respectively.

The RPF framework is set up as per Table 2 with 500 particles. The framework is then used separately for the eight tool wear cases. Figure 8 shows the results of the prediction

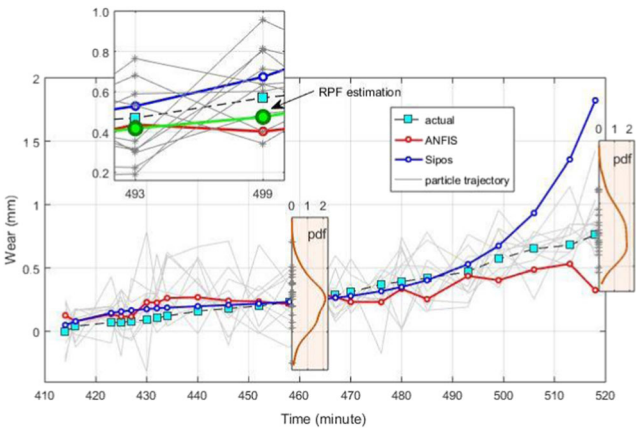

Fig. 8 Trajectories of randomly selected 10 particles and tool-wear estimations on the eighth segment with RPF (note: the RPF-estimated tool wear in the time frame 493–499 min is shown as green pointers. In addition, the probability density of the state and sample representing particles for time step 458 and 518 min are also shown in the figure. The vertical axis has the same scale of the "wear (mm)", representing the span of particle estimated values at that time step, whereas the horizontal axis shows the probability density from 0 to 2 of the particle estimations. The total area under the probability density plot equals 1)

and the inference models, as well as the trajectories of randomly selected 10 particles for the tool wear test #8 (i.e., cutting condition #4 and second insert as shown in Table 1). The particles at each time step represent the distribution of the estimated state. The probability density of the state and sample representing particles for two different times are shown in the figure. The expected value of the state is calculated and assigned as the estimated tool wear magnitude for the corresponding time step. The green pointers in the figure show the estimated values. Figure 9 shows the actual tool wear values and the results for all eight tool wear tests. The estimated results of the fusion framework appear to be the closest to the actual values.

For quantitative comparison of the performance, modeling errors of the methods are calculated and compared. The errors are calculated with three measures: the normalized root mean squared error (RMSE), the normalized mean absolute error (MAE), and the normalized maximum error. Figure 10 shows the normalized root mean squared results for the eight segments. In general, the result shows that the RPF effectively reduces the modeling error. In the segments #3, #4, #6, and #7, the measurement-based inference leads to large errors. Updating the estimation results with the erroneous measurements tends to increase the PRF errors for the corresponding segments. However, the RPF errors do not tangibly increase from the errors of the Sipos prediction model.

In segments #1, #2, and #8, the prediction errors are larger than the errors of the inference model. In segment #8 in particular, the prediction error significantly increases for the last four predictions. This leads to 0.97 RMSE, which is unacceptably large for a prediction model. In the same segment, the RMSE of the measurement-based inference is 0.48. The RPF manages to reduce the error down to 0.26.

Figure 11 compares the overall performance of the methods over the eight experiments. The results show that fusion of the Sipos prediction model and the measurement-based inference leads to smaller errors than either methods used for the fusion. This can be observed for all three metrics of the error as shown in Fig. 11a. A useful metric to compare the modeling performance is the coefficient of determination denoted as R^2. It is the proportion of the variance of the dependent variable that is

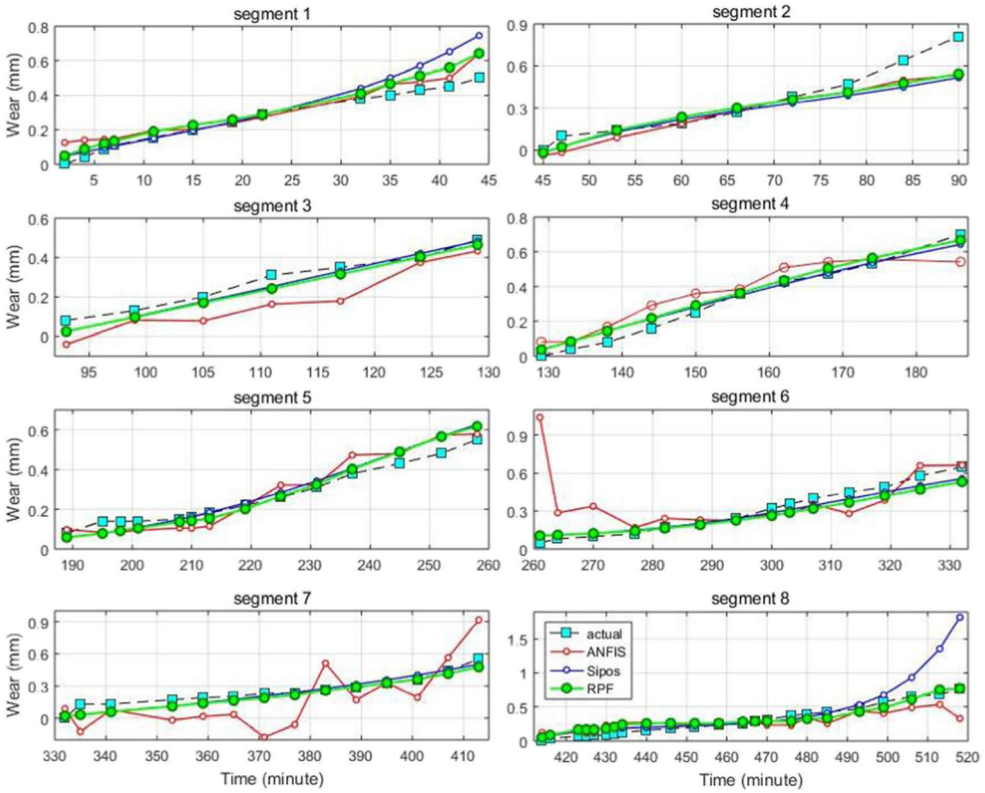

Fig. 9 Tool wear prediction results with: measurement-based inference with ANFIS, Sipos prediction model, and data-model fusion with RPF

predictable from the independent variable. The better the performance of a model at predicting the variable of interest, the more the coefficient of determination approaches to 1. Figure 11b compares the coefficient of determination for the three methods. The coefficient of determination for the RPF is 0.886 whereas this coefficient for the prediction and the inference models are 0.429 and 0.289, respectively.

6 Conclusions

A hybrid data-driven physics-based model fusion framework was developed and applied for estimation of tool wear in this work. This model fuses together the prediction results of an empirical wear-time model (i.e., Sipos model) and a measurement-based inference model (i.e., ANFIS model) in a stepwise manner to manage uncertainties and noise of both methods. The proposed fusion structure is an extension of the PF technique, to which the stochastic characteristic of the system input is taken into account. With this approach, the weights of the particles are updated using the state probability distribution function, instead of the multi-dimensional measurement likelihood that is not available.

A set of tool wear measurement data and the corresponding features extracted from the spindle current measurements were adopted to verify the performance of the developed framework. The results were compared with the results of the individual measurement-based inference model and Sipos tool-wear prediction model. The results show considerable improvement in tool wear estimation results using different modeling error metrics. On average, the RMSE of the fusion framework drops to 0.22, whereas it is 0.42 for the Sipos prediction model and 0.56 for the measurement-based

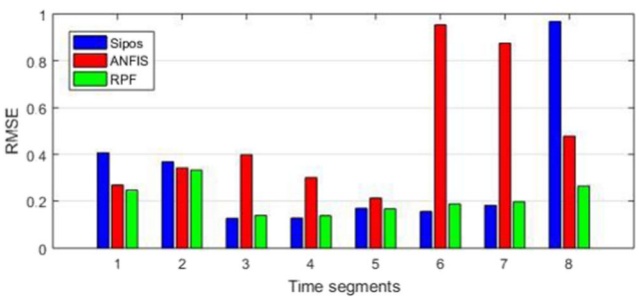

Fig. 10 Normalized root mean squared error of modeling

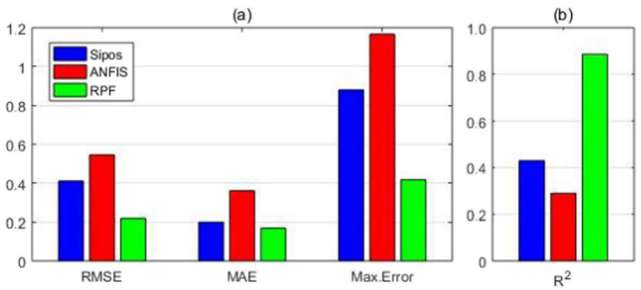

Fig. 11 Overall modeling performance. **a** Normalized errors. **b** Coefficient of determination

inference model. Likewise, MAE and the maximum error metrics show the fusion framework outperforms the individual models with resulting smaller prediction errors. Comparison of the performances using the coefficient of determination shows superior results for the fusion framework with $R^2 = 0.886$, while this metric for the prediction and the inference models are 0.429 and 0.289, respectively. In future research, application of the fusion framework will be extended to tool wear prognostics and estimation of the remaining useful life of the cutting tool.

Acknowledgements This project was financially supported by the Natural Sciences and Engineering Research Council of Canada (Grant number: RGPIN/05922-2014).

References

1. Kalpakjian S, Schmid S (2006) Manufacturing, engineering and technology, 6th edn. Peason, Hoboken
2. Yen YC, Söhner J, Lilly B, Altan T (2004) Estimation of tool wear in orthogonal cutting using the finite element analysis. J Mater Process Technol 146:82–91. https://doi.org/10.1016/S0924-0136(03)00847-1
3. Pálmai Z (2013) Proposal for a new theoretical model of the cutting tool's flank wear. Wear 303:437–445. https://doi.org/10.1016/j.wear.2013.03.025
4. Bhattacharyya A, Ham I (1969) Analysis of tool wear—part I: theoretical models of flank wear. J Eng Ind 91:790. https://doi.org/10.1115/1.3591696
5. Danai K, Ulsoy AG (1987) A dynamic state model for on-line tool wear estimation in turning. J Eng Ind 109(4):396–399
6. Chen XQ, Li HZ (2009) Development of a tool wear observer model for online tool condition monitoring and control in machining nickel-based alloys. Int J Adv Manuf Technol 45:786–800. https://doi.org/10.1007/s00170-009-2003-1
7. Liu J, Shao Y (2018) An improved analytical model for a lubricated roller bearing including a localized defect with different edge shapes. J Vib Control 24(17):3894–3907. https://doi.org/10.1177/1077546317716315
8. Attanasio A, Ceretti E, Rizzuti S, Umbrello D, Micari F (2008) 3D finite element analysis of tool wear in machining. CIRP Ann - Manuf Technol 57:61–64. https://doi.org/10.1016/j.cirp.2008.03.123
9. Colbaugh R, Glass K (1995) Real time tool wear estimation using recurrent neural networks. In: Proc of the 1995 IEEE Int Symp on Intelligent Contr, Monterey, USA, pp 357–362
10. Eker Ö, Camci F, Jennions I (2012) Major challenges in prognostics: study on benchmarking prognostic datasets. In: 1st European conference of the prognostics and health management Society pp 1–8
11. Müller E (1962) Der Verschleiss von Hartmetallwerkzeugen und seine kurzzeitige Ermittlung. Eidgenössischen Technischen Hochschule in Zürich
12. Zorev NN, Granovskij GI, Loladze TN, Tretyakov IP (1967) Razvitie nauki o rezanii metallov. In: Mashinostroenije. Moskow
13. Sipos Z (1986) Investigation of cutting performance of coated HSS tools made in Hungary. NME, Miskoc
14. Sick B (2002) On-line and indirect tool wear monitoring in turning with artificial neural networks: a review of more than a decade of research. Mech Syst Signal Process 16:487–546. https://doi.org/10.1006/MSSP.2001.1460
15. Benkedjouh T, Medjaher K, Zerhouni N, Rechak S (2015) Health assessment and life prediction of cutting tools based on support vector regression. J Intell Manuf 26:213–223. https://doi.org/10.1007/s10845-013-0774-6
16. Li S, Elbestawi MA (1996) Tool condition monitoring in machining by fuzzy neural networks. J Dyn Syst Meas Control 118:665–672. https://doi.org/10.1115/1.2802341
17. Kuo RJ (2000) Multi-sensor integration for on-line tool wear estimation through artificial neural networks and fuzzy neural network. Eng Appl Artif Intell 13:249–261. https://doi.org/10.1016/S0952-1976(00)00008-7
18. Gajate A, Haber R, Del TR et al (2012) Tool wear monitoring using neuro-fuzzy techniques: a comparative study in a turning process. J Intell Manuf 23:869–882. https://doi.org/10.1007/s10845-010-0443-y
19. Dutta RK, Paul S, Chattopadhyay AB (2000) Fuzzy controlled backpropagation neural network for tool condition monitoring in face milling. Int J Prod Res 38:2989–3010. https://doi.org/10.1080/00207540050117404
20. Cao H, Zhang X, Chen X (2017) The concept and progress of intelligent spindles: a review. Int J Mach Tools Manuf 112:21–52. https://doi.org/10.1016/j.ijmachtools.2016.10.005
21. Zhao R, Yan R, Chen Z et al (2016) Deep learning and its applications to machine health monitoring: a survey. Prepr arXiv 161207640:1–14. https://doi.org/10.1016/j.jocs.2017.06.006
22. Malhotra P, TV V, Ramakrishnan A, et al (2016) Multi-sensor prognostics using an unsupervised health index based on LSTM encoder-decoder. 1st ACM SIGKDD work Mach learn Progn heal Manag san Fransisco, CA, USA
23. Simon D (2006) Optimal state estimation: Kalman, H [infinity] and nonlinear approaches. Wiley-Interscience
24. Wang J, Wang P, Gao RX (2013) Tool life prediction for sustainable manufacturing. In: 11th global conference on sustainable manufacturing innovative solutions. Univ.-Verl. der TU, Berlin, pp 230–234
25. Wang J, Wang P, Gao R et al (2015) Enhanced particle filter for tool wear prediction. J Manuf Syst 36:35–45
26. Wang P, Gao RX (2016) Stochastic tool wear prediction for sustainable manufacturing. Procedia CIRP 48:236–241. https://doi.org/10.1016/J.PROCIR.2016.03.101
27. Wang P, Gao RX (2015) Adaptive resampling-based particle filtering for tool life prediction. J Manuf Syst 37:528–534. https://doi.org/10.1016/j.jmsy.2015.04.006
28. Gordon NJ, Salmond DJ, Smith AFM (1989) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In: IEE Proceedings (Radar and Signal Processing). Institution of Electrical Engineers, pp 107–113
29. Agogino A, Goebel K (2007) Milling Data Set. In: BEST lab, UC Berkeley, NASA Ames Progn. Data Repos. https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/
30. Lei Y, Li N, Guo L, Li N, Yan T, Lin J (2018) Machinery health prognostics: a systematic review from data acquisition to RUL prediction. Mech Syst Signal Process 104:799–834. https://doi.org/10.1016/j.ymssp.2017.11.016
31. Rad JS, Zhang Y, Chen C (2014) A novel local time-frequency domain feature extraction method for tool condition monitoring using S-transform and genetic algorithm. IFAC Proc 47(3):3516–3521
32. Ghosh N, Ravi YB, Patra A, Mukhopadhyay S, Paul S, Mohanty AR, Chattopadhyay AB (2007) Estimation of tool wear during CNC milling using neural network-based sensor fusion. Mech

Syst Signal Process 21:466–479. https://doi.org/10.1016/j.ymssp.2005.10.010
33. Jain V, Raj T (2017) Tool life management of unmanned production system based on surface roughness by ANFIS. Int J Syst Assur Eng Manag 8:458–467. https://doi.org/10.1007/s13198-016-0450-2
34. Whitehouse DJ (1978) BETA functions for surface typologie? CIRP Ann 27:491–497
35. Kannatey-Asibu E, Dornfeld DA (1982) A study of tool wear using statistical analysis of metal-cutting acoustic emission. Wear 76: 247–261. https://doi.org/10.1016/0043-1648(82)90009-6
36. Jang JSR (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern 23:665–685. https://doi.org/10.1109/21.256541
37. Hanachi H, Liu J (2016) State estimation for general class of dynamical systems: an extension to particle filters. In: the 3rd international conference on control, dynamic systems, and robotics (CDSR'16). Ottawa, pp 1–9
38. Hanachi H, Liu J, Banerjee A, Chen Y (2016) Sequential state estimation of nonlinear/non-Gaussian systems with stochastic input for turbine degradation estimation. Mech Syst Signal Process 72–73:32–45. https://doi.org/10.1016/j.ymssp.2015.10.022
39. Silverman BW (2018) Density estimation for statistics and data analysis. Routledge
40. Musso C, Oudjane N, Gland F (2001) Improving regularised particle filters. In: Sequential Monte Carlo methods in practice. Springer, New York, pp 247–271