Hybrid heterogeneous prognosis of drill-bit lives through model-based spindle power analysis and direct tool inspection

Luca Bernini[1,2] · **Paolo Albertelli**[1,2] · **Michele Monno**[1,2]

Abstract

In the context of Industry 5.0, manufacturing systems are driven by human-centered production processes, assigning high-level supervisory tasks to operators. This necessitates that machines can perform low-level decision-making actions. This paper presents a novel hybrid heterogeneous prognosis algorithm designed to autonomously inspect the cutting edges of drill-bits and to forecast their Remaining Useful Life along with the associated probability density function. The algorithm leverages specific force coefficients from spindle power and feed axis current measurements, as features correlated with tool wear, to detect tool brittle failures. Additionally, flank wear is automatically measured through a specifically conceived image processing algorithm, using thresholding, convolutional filters, and edge detection techniques. Direct tool wear measurements are analyzed by a hybrid prognosis algorithm, fusing particle filter and multi-layer perceptron, to predict drill-bits' remaining useful lives. The proposed solution offers several advantages. It reduces the need for extensive experimental run-to-failure tests typically required for training standard machine learning algorithms. Instead, it allows for real-time adaptation, even in scenarios involving untested and varying cutting process conditions. Furthermore, it utilizes both indirect wear observations during cutting operations and direct wear observations during setup times (e.g. tool changes, workpiece changes), without interrupting the ongoing process. Exponent of Kronenberg's models for specific force coefficients was found to be sensitive to tool wear. Prognosis could correctly predict the 67% of end-of-lives with an average prognosis horizon of 30%.

1 Introduction

As Industry 4.0 paved the way for advancements in production systems, the emergence of Industry 5.0 emphasizes the integration of societal goals into industrial objectives, necessitating a shift towards human-centric production processes [24]. In this context, operators are tasked with higher-level supervisory responsibilities, while low-level decision-making is delegated to machine intelligence. Consequently, the pursuit of intelligent strategies for monitoring manufacturing processes and aiding maintenance decision-making becomes even more important.

Cutting tools monitoring and prognosis is an active research topic in the field of manufacturing, due to the tools significant impact on machine downtimes [27], production economics [21], and rare material utilization [17]. However, despite ongoing research efforts, effective solutions in this domain are still not achieving widespread industrial adoption. Current approaches often rely on indirect inspection methods utilizing process measurements such as vibrations [9, 25, 26], acoustic emissions [23], cutting forces [9, 14, 20, 25, 26], and cutting power [6] to estimate tool conditions in a real-time fashion. Despite these techniques are widespread, they typically extract features (e.g. mean, standard deviation, harmonic amplitudes and ranges, wavelet decomposition [11]) influenced by process parameters [14], like feed and spindle speed for a drilling application. Further-

✉ Luca Bernini
luca1.bernini@polimi.it

Paolo Albertelli
paolo.albertelli@polimi.it

Michele Monno
michele.monno@polimi.it

1   Department of Mechanical Engineering, Politecnico di Milano, Via La Masa, 1, Milan 20156, Lombardy, Italy

2   MUSP, Macchine Utensili Sistemi di Produzione, strada della Torre della Razza, Piacenza 29122, Emilia-Romagna, Italy

more, these features are generally passed to statistical-based (e.g. Auto-Regressive Moving-Average models) and data-driven models (e.g. Artificial Neural Networks and Support Vector Machines) which need a high quantity and quality of data to estimate whether the cutting tool has failed or not [16]. Both the aspects above, make the conceived solutions less adaptable to varying operational conditions and not ready for an industrial implementation. This is especially true for one-of-a-kind or small-batch production, where cutting is performed under varying cutting conditions, where system retraining is needed and only limited training data are available [14]. Nevertheless, some researchers started to investigate the instantaneous estimation of model-based features, capable of separating the effect of tool wear, from cutting conditions [7, 14]. Conversely, direct inspection techniques, such as profilometric acquisitions [12] or calibrated image analysis [8], are constrained to operate when the cutting tool is not engaged in processing the workpiece.

Both inspection strategies can inform prognosis algorithms aimed at estimating the Remaining Useful Life (RUL) of cutting tools. Existing prognosis techniques fall into four categories, including model-based, statistical-based, data-driven, and experience-based algorithms [16]. Latest research efforts are mostly oriented to data-driven solutions, especially dealing with machine learning and deep learning techniques. However, while effective for mass production scenarios with enough data availability, these approaches often require extensive run-to-failure (RTFs) experiments and parameter combinations to be tested, making them impractical for small-batch production environments. Available commercial monitoring tools suffer from the same limitations. They typically use historical data on repetitive operations (the case of mass production) to tune static or dynamic thresholds. In case the workpiece features change, available thresholds lose their efficacy.

In order to overcome the limitations above, this paper introduces a novel hybrid solution that combines multiple techniques and algorithm categories. At the same time, the method is heterogeneous [27], incorporating both indirect and direct inspection, to monitor and forecast drill-bit wear, respectively. Cutting edge pictures (feeding the direct method) can be taken only when the drill-bit is not engaged in the workpiece. Indeed, depending on the industrial scenario, a picture may be taken after drilling each hole or after a set of holes is drilled. Direct inspection feeds tool prognosis allowing to predict if the drill-bit is able to perform the upcoming set of holes. Indirect monitoring checks the stability of the cutting process and identifies brittle wear phenomena. This allows to stop manufacturing before damaging the part, even if cutting edge pictures are not available. The proposed indirect solution is designed to be independent of cutting parameters and adaptable to unforeseen degradation trends. Moreover, it requires minimal preliminary data for

model calibration and training, making it suitable for small-batch production contexts.

Section 2 of the paper presents the indirect and direct inspection phases (Sect. 2.1), followed by a detailed explanation of indirect monitoring (Sect. 2.2) and direct drill-bit prognosis (Sect. 2.3) using a hybrid approach involving multi-layer perceptron (MLP) and particle filter (PF). The experiments were described at the end of the same section (Sect. 2.4). Results and their discussions are presented in Sect. 3, referring to the process characterization (Sect. 3.1) and RTFs monitoring and prognosis (Sect. 3.2). The paper ends with the presentation of the conclusions in Sect. 4.

2 Materials and methods

The architecture of the devised approach is illustrated in Fig. 1.

2.1 Inspection phase

2.1.1 Indirect inspection

Power and current measurements obtained from the machine undergo pre-processing in order to derive the cutting power and current $\overline{P}_{cut,p,j}$ and $\overline{i}_{cut,p,j}$, respectively. Indeed, the two measured quantities are averaged over the stationary sections of each peck (indexed by p) of a hole (indexed by j). Subsequently, Eq. 1 is applied:

$$
\begin{aligned}
\overline{P}_{cut,j,p} &= \overline{P}_{meas,j,p} - \overline{P}_{air,j,p} \\
\overline{i}_{cut,j,p} &= \overline{i}_{meas,j,p} - \overline{i}_{air,j,p}
\end{aligned}
\tag{1}
$$

Here, $\overline{P}_{cut,j,p}$ and $\overline{i}_{cut,j,p}$ represent the cutting power and current of the p-th peck of hole j, while $\overline{P}_{air,j,p}$ and $\overline{i}_{air,j,p}$ denote the absorption required for moving the spindle and feed axis at their speed set points (when no cutting occurs). $\overline{P}_{air,j,p}$ and $\overline{i}_{air,j,p}$ are functions of the spindle and feed speeds, respectively. To eliminate their influence, two linear models were fitted based on specific tests conducted without cutting. Subsequently, specific force coefficients are derived from the two quantities (Eq. 2):

$$
\begin{aligned}
k_{ct,j,p} &= \overline{P}cut, j, p/MRR_{j,p} \\
\tilde{k}_{ca,j,p} &= k_{ca,j,p}/(\tau_m K_m) = \overline{i}_{cut,j,p}/(2rc_{j,p})
\end{aligned}
\tag{2}
$$

where $k_{ct,p,j}$ and $k_{ca,p,j}$ represent the tangential and axial specific force coefficients, respectively; τ_m is the feed axis motor transmission ratio; K_m denotes the feed axis motor constant; r represents the drill-bit radius; and $c_{j,p}$ is the

Fig. 1 Scheme of the conceived methodology
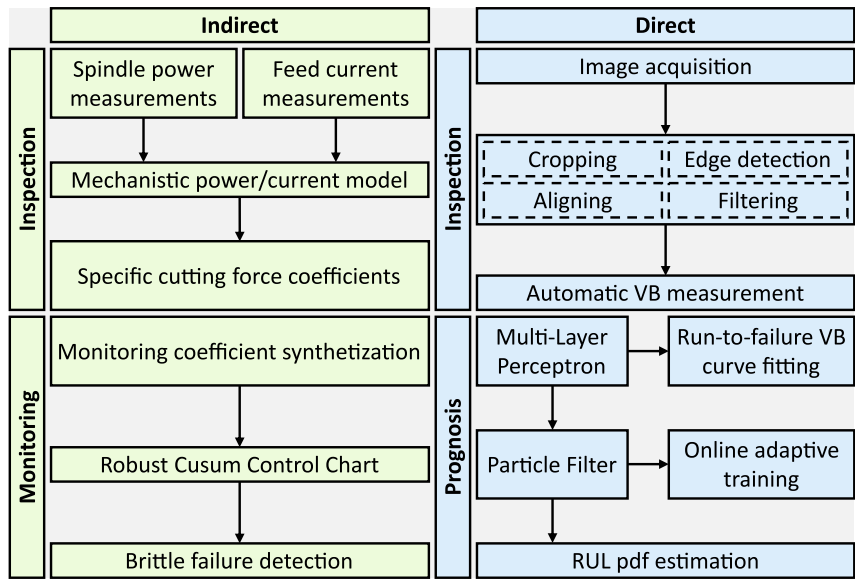
feed per tooth of the peck. Since the motor constant and transmission ratio are unknown, only $\tilde{k}_{ca,p,j}$ was estimated (proportional to the specific axial cutting energy). The specific force coefficients estimated by Eq. 2 are influenced by the undeformed chip thickness $h_{j,p} = c_{j,p} \cdot \sin(\kappa_t)$, where κ_t is half of the point angle of the drill-bit. Hence, an exponential (Kronenberg's) model is employed to mitigate this effect (Eq. 3):

$$
\begin{aligned}
k_{ct,j,p} &= k_{cst} h_{j,p}^{-x_t} \\
\tilde{k}_{ca,j,p} &= \tilde{k}_{csa} h_{j,p}^{-x_a}
\end{aligned} \tag{3}
$$

Here, k_{cst} and x_t, as well as $\tilde{k}_{csa}$ and x_a, represent the models' coefficients in the tangential and axial directions, respectively, depending on the workpiece material-cutting tool pair. Depending on the experiments under analysis, i.e., process characterization (*experiments E1*, see Sect. 2.4.2) and RTFs (*experiments E2*), a mean identification procedure and an instantaneous identification procedure are applied to estimate the four quantities, respectively.

Mean approach In order to apply the mean approach, variable feed per tooth tests are mandatory. Thus, this approach is followed for *experiments E1*. Linear regression is applied on the logarithmic version of Eq. 3, where k_{cst}, x_t, k_{csa}, and x_a are the regression coefficients. Thus, four estimates $\hat{k}_{cst}$, $\hat{x}_t$, $\hat{\tilde{k}}_{csa}$, and $\hat{x}_a$ are obtained through this approach.

Instantaneous approach The instantaneous approach allows to estimate the specific force coefficients when a hole is drilled. Indeed, this approach is useful to track the evolution of a specific force coefficient as the wear of the drill-bit progresses (*experiments E2*). During the drilling of a hole, feed per tooth is constant. Thus, either k_{cst} and k_{csa}, or x_t

and x_a, are set to the value estimated in process characterization. This allows to identify the remaining coefficients by least-square estimation on the relative peck quantities. In this paper, k_{cst} and k_{csa} were set to the values identified during *experiments E1* (i.e. $\hat{k}_{cst}$ and $\hat{\tilde{k}}_{csa}$; the reason of this choice is presented in Sect. 3). Indeed, in Eq. 4, x_t and x_a are estimated for each hole j, obtaining $\hat{x}_{t,j}$ and $\hat{x}_{a,j}$, respectively:

$$
\begin{aligned}
k_{ct,j,p} &= k_{cst} h_{j,p}^{-x_{t,j}} \\
\tilde{k}_{ca,j,p} &= \tilde{k}_{csa} h_{j,p}^{-x_{a,j}}
\end{aligned} \tag{4}
$$

2.1.2 Direct inspection

The direct inspection of the drill-bit begins with flank images, each undergoing four automatic pre-processing steps aimed at centering images on the drill-bit flank surface (pre-processing 1, Fig. 2):

- A grayscale transformation is applied to the image to compress the image feature space [5].
- Binarization is conducted using an intensity threshold of 120 to isolate the cutting edge.
- The image is centered on the center of mass of the masked area.
- The image is cropped to a resolution of 900×400 pixels.

Following this, five additional pre-processing steps (pre-processing 2, Fig. 2) ensure the alignment of drill-bits' reference edges among different images (i.e. at different wear levels):

- Gaussian image blurring is performed using an 11×11 Gaussian kernel to mitigate image noise.

Fig. 2 Direct inspection phase workflow. Red elements represent items identified or used by the current step. Blue elements represent items identified in previous steps
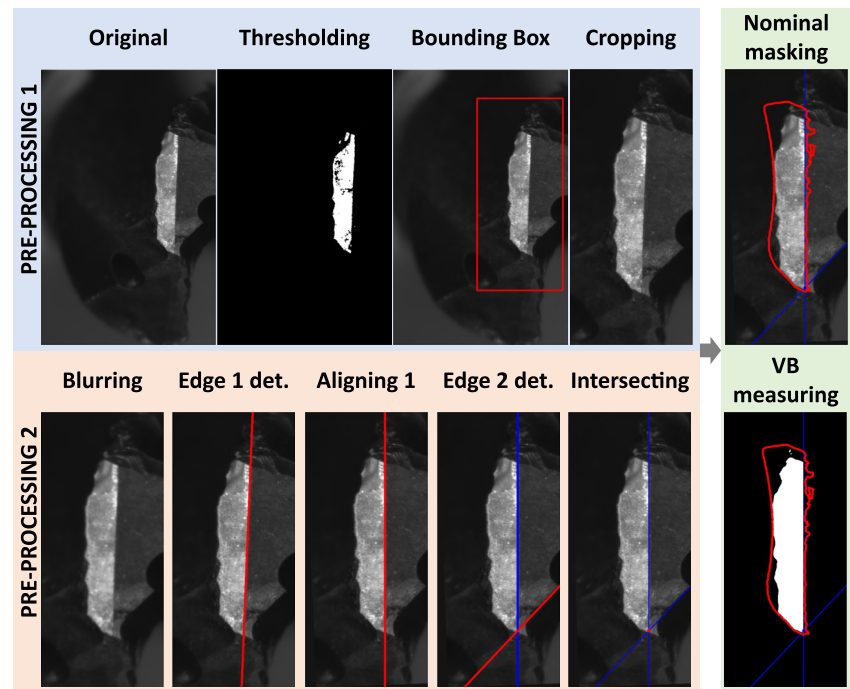
- Edge detection 1 applies a 2D convolutional filter to emphasize the vertical reference edge (specifically, edge 1).
- Alignment 1 rotates all images to maintain edge 1 in a vertical orientation.
- Edge detection 2 utilizes a 2D convolutional filter on the blurred image, followed by Otsu's binarization [15] to highlight the oblique reference edge (edge 2).
- The last step calculates the intersection of the two edges, serving as the reference point for placing the nominal cutting tool flank mask.

Finally, the mean flank wear is calculated. The worn drill-bit flank is masked through binarization with an intensity threshold of 70 (within nominal mask bounds). The mean flank wear (VB) measurement of the cutting edge for each image is determined as the difference between the nominal area and the highlighted area, divided by the flank length. As the drill-bit features two cutting edges, the flank wear measure is computed as the mean of the two at each stop, denoted as VB_m. The kernels of the pre-processing filters described above are listed in Table 1.

Table 1 Convolutional layers' kernels

Layer	Edge 1 detection	Edge 2 detection	
Kernel	$\begin{bmatrix} -4 & 0 & 4.5 \\ -4 & 0 & 4.5 \\ -4 & 0 & 4.5 \end{bmatrix}$	$\begin{bmatrix} -1.5 & \cdots & -1.5 \\ 0 & \cdots & 0 \\ 1.8 & \cdots & 1.8 \end{bmatrix}$	$[3 \times 20]$

2.2 Monitoring phase

The monitoring strategy employs a cusum control chart [13]. Similarly to the approaches outlined in [14] and [1], a summary index X_j is computed for each hole j as follows (Eq. 5):

$$X_j = \sqrt{\left(\frac{\hat{x}_{t,j}}{\hat{x}_{t,1}}\right)^2 + \left(\frac{\hat{x}_{a,j}}{\hat{x}_{a,1}}\right)^2} \tag{5}$$

where X_j represents an equivalent exponent of the Kronenberg's model at hole j, and thus the sensibility of the specific force coefficient to the feed per tooth as a function of tool wear. Moreover, it is noted that brittle wear phenomena tend to accelerate the degradation rate of the specific force coefficient and increase its instability [14]. By having kept fixed the specific force coefficients, this behaviour is expected to affect the summary index X_j. Hence, the absolute value of the moving range of X_j serves as the monitoring variable (Eq. 6):

$$\delta X_j = |X_j - X_{j-1}| \tag{6}$$

where δX_j represents the absolute value of the moving range when hole j is executed, and X_j denotes the summary index estimated during the j-th hole.

A cusum control chart is implemented with standard thresholds following [13], initially tuned on the first 25 holes. This approach ensures that the operator assesses the drill-bit condition to ascertain whether the brittle wear event constitutes a critical failure. The choice of the cusum control chart

is motivated by its rapid responsiveness to minor shifts in the mean of the control variable. When a false positive occurs, the control chart is recalibrated on 25 new observations. Consequently, detecting a second change in the slope of the moving range becomes more challenging.

2.3 Prognostics phase

In this study, a hybrid prognostic framework is formulated, drawing inspiration from methodologies employed in crack growth propagation and Li-ion battery prognostics [4, 19].

2.3.1 Multi-layer perceptron

The MLP is a simple neural network, establishing a mapping between input features and outputs. In our context, the MLP correlates VB_m with the hole number i (which may be different from j, since the MLP can be evaluated at different hole numbers). Consequently, the MLP comprises one input neuron and a single output neuron. To constrain the complexity of the MLP, a single hidden layer with three neurons is introduced. The resultant structure of the MLP is depicted in Fig. 3. The approach followed here introduces the improvements carried by [3] in turning application to the original version applied for drilling in [2]. Thus, the biases of the MLP are set to zero, and the MLP passes through the origin. Furthermore, two mapping functions, $m^{(0)}$ and $m^{(2)}$, are introduced to normalize the input and the output of the MLP (see [3]), respectively. The corresponding equation governing the MLP operation is formulated as follows (Eq. 7):

$$\hat{VB}_m(i) = m^{(2)}\left(\sum_{s=1}^{3}\theta_{s+3}\cdot a\left(\theta_s\cdot m^{(0)}(i)\right)\right) \qquad (7)$$

Here, $\theta_1, \theta_2, \theta_3, \theta_4, \theta_5$, and θ_6 denote the weights of the MLP, respectively; $a(\cdot)$ represents the activation function, defined as per Eq. 8:
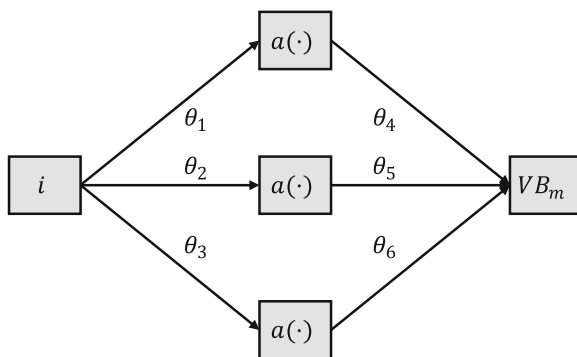
$$a(u) = u/(1 + |u|) \qquad (8)$$

Fig. 3 Adopted multi-layer perceptron architecture

where u signifies a generic input. Consequently, given the input hole number i, the MLP yields the corresponding predicted flank wear $\hat{VB}_m(i)$. Training of the MLP is conducted based on a single RTF test.

2.3.2 Particle filter

The PF is an advancement of the Kalman filter specifically conceived for non-linear dynamical systems and non-Gaussian measurement noise and process disturbances. PF approximates the state PDF of the dynamical system using a set of n_m particles (here, $n_m = 5000$). In our developed solution, PF serves as a means to update the MLP function (i.e. $\hat{VB}_m(i)$) whenever a new flank wear measurement becomes available. Additionally, we assume a flank wear image is accessible every 5 holes (i.e. when $j = f$ and $f = 0, 5, 10, \dots$). Therefore, PF updates the MLP every 5 holes as the sampling frequency. The formulated PF observes the following dynamical system (Eq. 9):

$$\begin{cases} \boldsymbol{\theta}_{f+1} = \boldsymbol{\theta}_f + \boldsymbol{d}_f \\ VB_{m,f}(i) = \theta_{10,f} + \sum_{s=1}^{3}\theta_{s+6,f}a\left(\theta_{s,f}i + \theta_{s+3,f}\right) + n_f \end{cases} \qquad (9)$$

Here, $\boldsymbol{\theta}_f$ represents the state of the dynamical system containing the weights of the MLP observed at hole f; $\boldsymbol{d}_f$ denotes the process disturbances, assumed to be Gaussian with null mean and diagonal covariance, proportional to $\boldsymbol{\theta}_f$ through the disturbance intensity Q (here, $Q = 0.007$). The first equation, known as the process equation, explores the weights space at each PF iteration. The second equation, termed the measurement equation, provides an estimation of the measured quantity (i.e. $\hat{VB}_m$) as a function of the hole number i, with the updated MLP. n_f represents the measurement noise, assumed Gaussian with null mean and standard deviation R (here, $R = 3$).

During each PF cycle, a set of particles is sampled using the process equation. Subsequently, each particle m corresponds to a different state value, resulting in a different MLP. Resampling, the second step of PF, selects the most representative particles by assigning weights proportional to the measurement likelihood to each particle. For details on the resampling stage, refer to [3, 4, 19]. Here, the forgetting factor ξ introduced by [3] was set to 0.9.

To conduct prognostics, at each PF iteration, each resampled MLP (m) is evaluated to identify the hole number corresponding to the threshold intersection (here, $150\mu m$). The intersection hole number is designated as $\hat{RUL}_{m,f}$. The RUL PDF $\hat{RUL}_f$ estimated at the f-th PF iteration is approximated through $\hat{RUL}_{m,f}$, where $m = 1, \dots, n_m$. To evaluate the performance of the prognostics solution, two prognostic metrics are computed: Absolute Prediction Error (APE) and

Prognostics Horizon (PH), determined using the $\alpha - \beta$ criterion with $\alpha = 0.20$ and $\beta = 0.95$ [18]. APE is calculated as a function of the normalized tool life λ (Eq. 10):

$$APE(\lambda) = \left| RUL_{true} - med[R\hat{U}L_f] \right| \tag{10}$$

Here, $\lambda = f/RUL_{true}$; RUL_{true} represents the actual hole number at which the flank wear overcomes the threshold; $med[R\hat{U}L_f]$ is the median of the estimated RUL PDF at the PF iteration f. PH serves as an indicator of how far in advance the algorithm accurately predicts the RUL. Further details on PH computation are available in [18].

## 2.4 Experimental set-up and experiments

2.4.1 Set-up

The experimental configuration comprises a RB50A balancing machine by Balance Systems S.r.l. (Fig. 4a). Throughout the experiments detailed in this section, AISI303 stainless steel (Fig. 4b) is utilized, employing a Balance System customized HSS drill-bit of 5.5mm diameter and a point angle $2\kappa_t = 140°$. The acquisition system facilitates the capture of both direct and indirect measurements (Fig. 4c). Specifically, an external DAQ from National Instruments (NI) DAQ9174 is integrated into the machine tool setup. Two NI acquisi-

tion boards NI9205 are installed on the DAQ to gather a set of five signals, encompassing both analog and digital data (Fig. 4d): machining/approaching (digital), feed speed, feed current, spindle speed, and spindle power (analog). The sampling frequency for all the measured quantities was set at 1kHz. For drill-bit wear characterization during the RTFs, an Optika SZN-T microscope was employed (Fig. 4e), with a 3D-printed support facilitating precise positioning of the drill-bit. In an industrial application context, the microscope can be replaced with cameras featuring macro lenses. In this case, a proper lighting system should be adopted to minimize the effect of ambient lights on the measurements. In order to enhance the robustness and the repeatability of the wear characterization process, cleaning was performed by slightly brushing the drill-bit edge, in order to remove dirt and eventual material adhesion, before taking each picture. This procedure can be automatized.

2.4.2 Experiments

The experimental activity encompassed two distinct phases: the initial phase focused on the characterization of the cutting process (*experiments E1*), while the subsequent phase involved RTFs to develop and validate the drill-bit prognosis methodology (*experiments E2*). Each hole underwent 5 pecks, contributing to a total hole depth of 6mm, conducted under dry conditions. A full-factorial Design of Experiments

Fig. 4 Experimental set-up: **a** balancing machine RB50A; **b** workpiece flange; **c** measurement set-up; **d** DAQ; **e** Optika SZN-T microscope
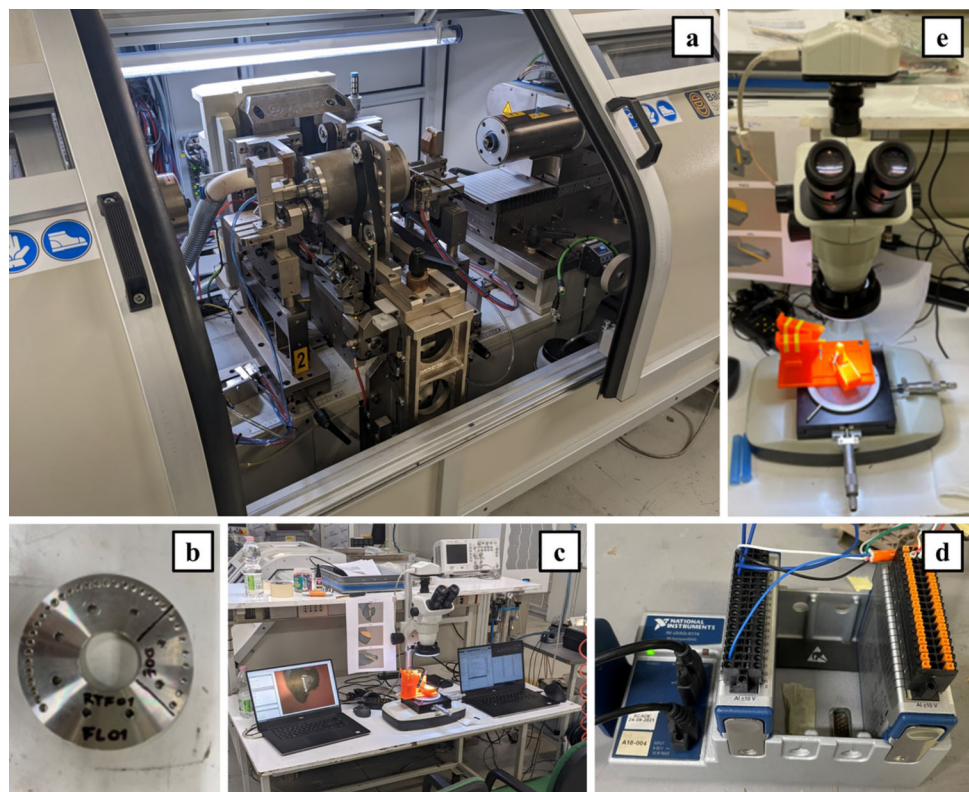
(DoE) was employed for both phases, incorporating central points and validation steps. The parameters for the DoE in *experiments E1*, comprising 25 cutting tests, are outlined in Table 2.

12 RTFs were carried out in *experiments E2 (experiments E2.1)*, and the related cutting parameters are shown in Table 3. The sequencing of tests was randomized to mitigate confounding influences. In order to see the effect of tool wear on the specific force coefficients k_{cs} and the exponential coefficient x, a set of 8 holes, defined through a DoE, was performed at the beginning of the drill-bit life, and once the End-of-Life (EoL) was reached for each RTF test (*experiments E2.2*). *Experiments E2.2* parameters are reported in Table 4. The tests were realized in random order, and the mean identification procedure described in Sect. 2.1.1 was applied for these DoEs.

3 Results and discussion

3.1 Cutting process characterization

The specific force coefficients measured during *experiments E1* were plotted as a function of the uncut chip thickness in Fig. 5. By looking at Fig. 5, the specific force coefficients, both tangential and axial, followed an exponential trend. Thus, the exponential Kronenberg's model coefficients were identified by the mean approach in Sect. 2.1.1. $\hat{k}_{cst}$ was estimated to be 2717 (2328; 3170) MPa, with an exponent coefficient $\hat{x}_t$ of 0.22 (0.17; 0.26). The brackets represent the 95% confidence intervals for the estimated quantity. Furthermore, when looking at the thrust specific force coefficients, $\hat{\bar{k}}_{csa}$ was estimated to be 0.52 (0.42; 0.65) A_{rms}/mm^2, with an exponent coefficient $\hat{x}_a$ of 0.35 (0.30; 0.41). The fitted regression models are shown in Fig. 5. In the figure, the regression models were reported together with the simultaneous prediction intervals at 95% confidence. As can be seen, all the validation points (i.e. ID 6 and 7 from Table 2, indicated by blue crosses, are comprised within these intervals, suggesting a good prediction behaviour of the built models. Furthermore, the models suggest that the influence of the

uncut chip thickness on the thrust specific force coefficient is more relevant than the effect on the tangential specific force coefficient.

3.2 Run-to-failures

During *experiments E2*, the direct and indirect inspection phases were applied. The evolution of the inspected flank wears is shown in Fig. 6. Trends in Fig. 6 were the results of the application of the direct inspection phase. The associated true EoL of each test is reported in Table 5. Figure 6 also demonstrates how the cutting speed and feed rate influence the flank wear evolution of the drill-bits. These relations are well known, and higher cutting speeds, as well as higher feed rates, increase the wear rate of the drill-bit. However, the variability of degradation rates was high, even when the same parameters were used (same colours). This is the main reason why monitoring and prognosis phases are gaining attention in the industrial scenario. Monitoring and prognosis phases must face such variability which is observed not only in the flank wear evolution but also in the signals coming from the machine. In fact, the effect of cutting parameters is typically larger than tool wear effect. Being capable to separate between the two is crucial for a correct application of indirect monitoring. For this reason, as explained in Sect. 2.4.2, tailored DoEs (*Experiments E2.2*) were carried out at the beginning and at the end of the tool lives in order to have a benchmark on the evolution of specific force coefficients through a standard mean identification procedure. The mean identification of the models' coefficients led to Fig. 7. In this figure, it is possible to understand the behaviour of $\hat{k}_{cst}$, $\hat{x}_t$, $\hat{\bar{k}}_{csa}$, and $\hat{x}_a$ when the drill-bit is new or at its EoL.

From Fig. 7, even when the drill-bits are new (blue graphs), the distributions of the specific force coefficients $\hat{k}_{cst}$ and $\hat{\bar{k}}_{csa}$ are really diverse, suggesting that on average, the behaviour of the drill-bits during machining is highly variable. Even if the tools and the material come from the same production batch, the variability on the specific force coefficients is quite large (from about 1000 to 4000 MPa, when referring to $\hat{k}_{cst}$). This behaviour may be associated to the manufacturing repeatability of the drill-bits' angles, which directly

Table 2 *Experiments E1* - DoE for cutting process characterization

ID	c [level]	v_c [level]	$c[mm/tooth]$	$v_c[m/min]$	$v_f[mm/s]$	$n[rpm]$	Replicates
1	Low	Low	0.020	52	2.0	3000	4
2	Low	High	0.020	70	2.7	4051	4
3	High	Low	0.040	52	4.0	3000	4
4	High	High	0.040	70	5.4	4051	4
5	Mid	Mid	0.030	61	3.5	3526	5
6	Val	Val	0.035	66	4.4	3788	2
7	Mid	Low	0.030	52	3.0	3000	2

Table 3 *Experiments E2.1 - DoE for run-to-failure tests*

ID	c [level]	v_c [level]	$c[mm/tooth]$	$v_c[m/min]$	$v_f[mm/s]$	$n[rpm]$	Replicates
1	Low	Low	0.020	52	2.0	3000	2
2	Low	High	0.020	70	2.7	4051	2
3	High	Low	0.040	52	4.0	3000	2
4	High	High	0.040	70	5.4	4051	2
5	Mid	Mid	0.030	61	3.5	3526	2
6	Mid	Low	0.030	52	3.0	3000	2

affect their behaviour during machining. Indeed, the variability associated to $\hat{k}_{cst}$ and $\hat{\bar{k}}_{csa}$ makes the classification between new and worn out drill-bits from these variables unfeasible. The opposite is observed for the exponents representing the effect of the chip thickness on the specific force coefficient generated during the drilling process. By observing the histograms in the second and fourth charts on the diagonal of Fig. 7, an evident distinction between the orange and blue distributions can be observed. This behaviour is associated to the modification of the rake and flank faces of the drill-bits as the wear progresses. In fact, changing the uncut chip thickness influences how much the modified geometry of the lip is involved in the cutting process; while low chip thicknesses involve mainly the damaged part of the lip, higher values of h involve the undamaged part of the rake surface, which preserves the original angles. This behaviour reflects in a change of the specific force coefficient trend with respect to the chip thickness and, consequently, to the model exponents $\hat{x}_t$ and $\hat{x}_a$. Since no appreciable difference can be observed in $\hat{k}_{cst}$ and $\hat{\bar{k}}_{csa}$ when a drill-bit is new or at the EoL, during the instantaneous indirect inspection approach explained in Sect. 2.1.1, they were kept constant and their values were set to their process characterization counterparts.

Another observation must be pointed out with reference to flank wear degradation curves shown in Fig. 6. The degradation phenomenon proceeds with a multi-stage trend, where fast degradation zones are alternated with flatter regions of degradation. This behaviour is related to the wear mechanics of the drill-bit. A cutting lip of the drill-bit starts to wear, leaving the other lip more exposed to the material removal process. This stabilizes the wear on the first lip and causes the second lip to wear out faster. This behaviour is, in gen-

eral, oscillating between flatter and high degradation rates. In Fig. 6, only the mean behaviour of the wear over the two cutting lips can be appreciated. These trends are particularly tough to be predicted and are the main reason for which prognosis algorithms must be robust and adaptable. Static algorithms would fail in this operation.

Two examples of correlations between the flank wear evolution during a RTF and the specific force coefficients exponents estimated by the instantaneous approach in 2.1.1 are shown in Figs. 8 and 9. In these figures, the pictures of the cutting lips are reported next to the measured data and flank wear estimation.

Figure 8 shows the results of the application of the direct and indirect inspection and monitoring phases for RTF 1 (replicate 1). In the top-left graph, the mean flank wear of the cutters is shown. The labels with letters are placed at the hole number at which the cutting lips pictures were taken (stop). Thus, the associated pictures are shown on the right of Fig. 8. By looking at flank wear of Fig. 8, a multi-stage behaviour is observed: in a first phase, up to stop D, the tool starts wearing out gradually near the external zone; in a second phase, before stop E, the wear rate of the tool increases by starting to involve also the central part of the cutting lips; in the third phase, up to stop H, the flank wear is stable and progresses uniformly on the whole lips of the drill-bit; in the fourth phase, up to stop I, chipping involves the external parts of cutters; in the last phase, flank wear is still stable, with limited wear progression on the entire lips. The second and third graphs on the left of Fig. 8 show the estimated exponents of the Kronenberg models for tangential and axial specific force coefficients, respectively. Both the $\hat{x}_t$ and $\hat{x}_a$ exponents show a multi-stage behaviour similar to what was

Table 4 *Experiments E2.2 - DoE executed at the beginning and at the end-of-life of each run-to-failure test*

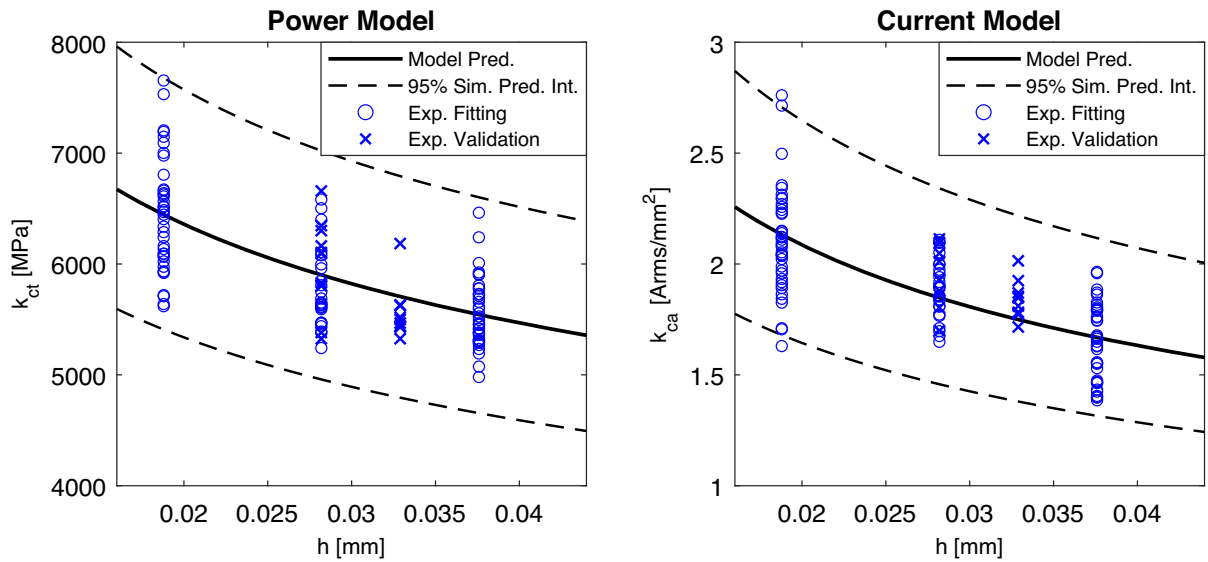| ID	c [level]	v_c [level]	$c[mm/tooth]$	$v_c[m/min]$	$v_f[mm/s]$	$n[rpm]$	Replicates
1	Low	Low	0.020	52	2.0	3000	1
2	Low	High	0.020	70	2.7	4051	1
3	High	Low	0.040	52	4.0	3000	1
4	High	High	0.040	70	5.4	4051	1
5	Mid	Mid	0.030	61	3.5	3526	1
6	Mid	Low	0.030	52	3.0	3000	3

Fig. 5 Process characterization by exponential modelling of specific force coefficients. (Left) model of tangential (cutting) specific force coefficients. (Right) model of axial (thrust) specific force coefficients

seen for the VB_m. Up to stop D, $\hat{x}_t$ faces a regular decreasing trend, whereas $\hat{x}_a$ remains constant. Between D and E stops, both the coefficients progressively increase their derivatives. Between E and H stops, the two coefficients stop their growth and remain constant. Actually, in this region, the coefficient $\hat{x}_a$ faces a little drop, most probably correlated to the microcracks developing on the C1 lip. When the cutters loose their external parts, a sudden drop in both the coefficients is found, followed by a higher growth rate. These two elements started causing the control chart to react at hole 295, finally setting an alarm at hole 331. In the last region, the coefficients remain stable around the values 0.37 and 0.5 for

$\hat{x}_t$ and $\hat{x}_a$, respectively. Since the control chart detected an out-of-control (OOC) condition, in the last region, the parameters of the chart are re-trained, making it more difficult to detect another OOC condition. The stops highlighted with a red box around the cutter pictures represent the first stop after a OOC condition was detected.

A similar behaviour is shown in Fig. 9. With respect to the first RTF analyzed, this RTF was performed at high cutting speed and high feed rate. As expected, this combination led to a faster degradation of the drill-bit that reached the EoL after 142 holes compared to the 388 holes of RTF 1 (1). The mean flank wear during this RTF faced three stages: a
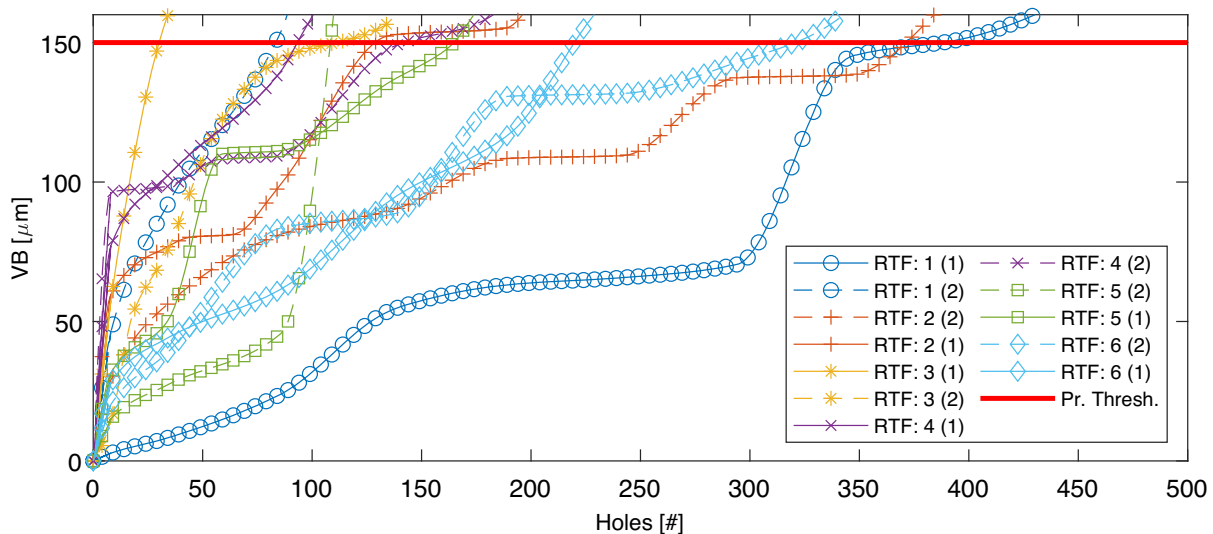

Fig. 6 Evolution of mean flank wear of drill-bits during run-to-failures. The same color suggests the same cutting parameters. In red, the adopted prognosis threshold

Table 5 True end-of-life (EoL) and prognosis horizon (PH) of the 12 drill-bits used during run-to-failures

Test	1 (1)	1 (2)	2 (1)	2 (2)	3 (1)	3 (2)	4 (1)	4 (2)	5 (1)	5 (2)	6 (1)	6 (2)
EoL	388	84	372	129	31	111	94	142	109	165	321	220
PH [%]	13	48	27	31	26	23	0	21	96	74	0	0

first rapid flank wear growth, due to the loss of the external part of cutter 1; a stable region before stop F; and a final faster degradation up the EoL. The trends on the Kronenberg's models exponents have similar behaviours, even if in the second region $\hat{x}_t$ faces a continuous increase in its value, whereas $\hat{x}_a$ is more stable in this phase. The control chart detected four OOC conditions. The first was detected right after stop E, probably related to the effects of the chipping on the C2 lip (already visible at stop E). The other occurrences of OOC conditions were detected right after stop F and between
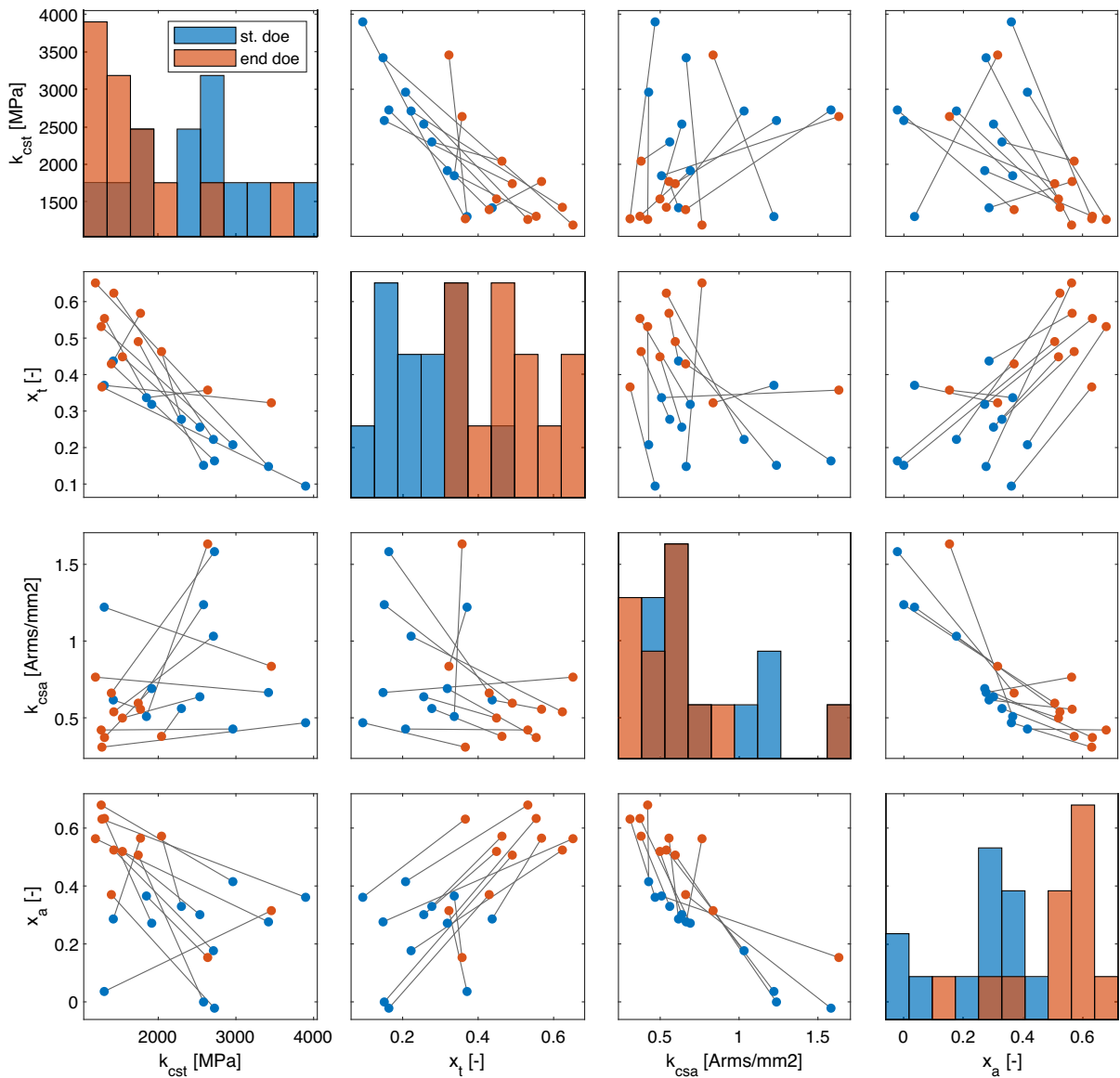
Fig. 7 The graph represents the distributions of the specific force coefficients, for new tools (blue) and end-of-life tools (orange). Main diagonal terms represent the histograms of each variable; off-diagonal plots represent the scatters between the two variable on the associated row/column. The axes of the plots are shared between the off-diagonal graphs. Histograms share the abscissa with off-diagonal charts, whereas their y-axes are just qualitative and do not refer to the reported tick marks. Grey lines connect the data coming from the same RTF
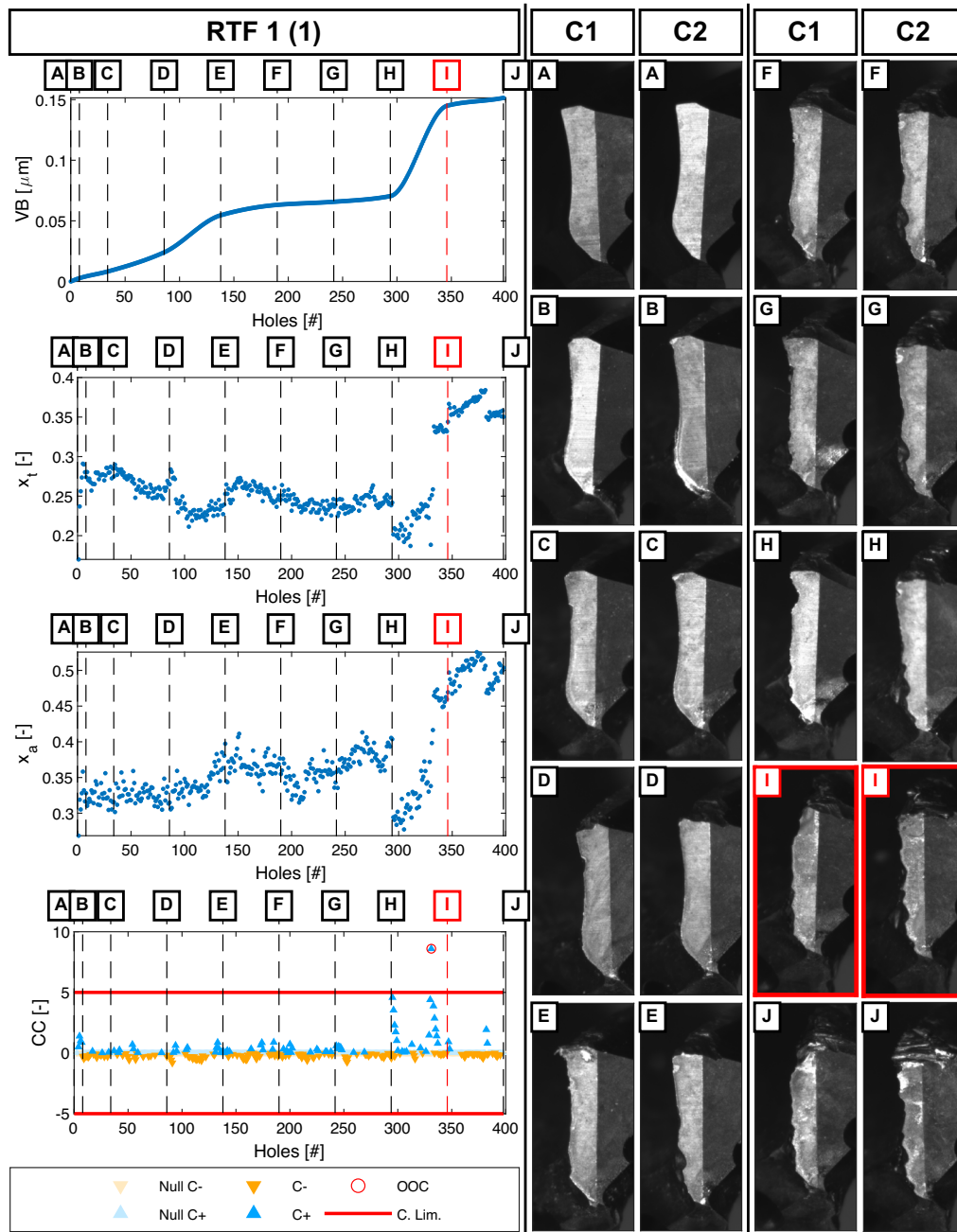
Fig. 8 Inspection and monitoring results of RTF 1 (1). On the left, from top to bottom, VB_m, $\hat{x}_t$, $\hat{x}_a$, and the control chart. Label boxes with capital letters show the stops at which the flank wear was evaluated. Red labels and flank pictures correspond to out-of-control detections

stop G and H, where the cutting lips underwent critical chippings in the external parts. With respect to the control chart related to the RTF 1 (1), the cumulative sums (control chart indicators $C+$ and $C-$) faced higher variability during the whole RTF. A promising result, regarding the use of $\hat{x}_t$ and $\hat{x}_a$ to assist a future indirect prognosis phase, is that the two coefficients assumed comparable values at the beginning and at the EoL for the different RTFs. Indeed, a shared threshold may be applied even if the cutting conditions will change.

In order to predict the evolution of flank wear of drill-bits, prognostics phase was implemented. Prognostics algorithm was trained on one RTF (RTF 4 (2)), performed at high feed and high cutting speed, and tested on all the others, in order to show the adaptivity feature of the conceived solution. It is important to underline that the reported results represent a critical condition for the algorithm, where only one RTF was used for training. In a real application scenario, the used drill-bits can be introduced in the initial state of the PF to
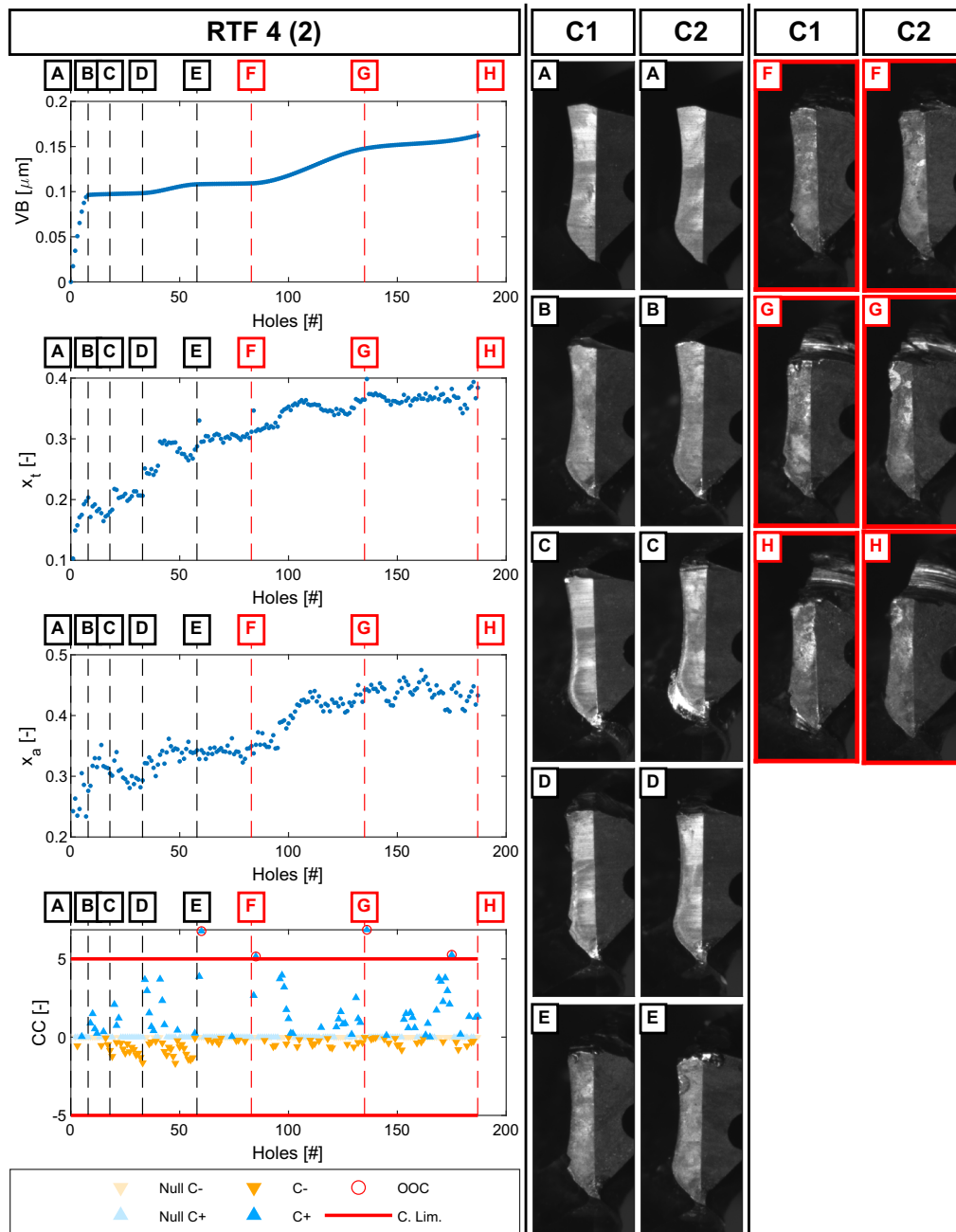
Fig. 9 Inspection and monitoring results of RTF 4 (2). On the left, from top to bottom, VB_m, $\hat{x}_t$, $\hat{x}_a$, and the control chart. Label boxes with capital letters show the stops at which the flank wear was evaluated. Red labels and flank pictures correspond to out-of-control detections

increase the algorithm performances. Furthermore, prognosis is not directly connected to the indirect monitoring system. Nevertheless, the two systems assist each other in different moments of the manufacturing processes or tool life. Prognostics results are shown in Fig. 10, while the associated APEs are shown in Fig. 11. Numerical values assumed by the PH are shown in Table 5. Figure 10 represents the trend of the estimated RUL (ρ: RUL normalized to the true EoL) over the normalized tool life λ. Image with label H represents

the case where the algorithm was tested on the training RTF (RTF 4 (2)). It is important to track how the predicted RUL mean (blue solid line) and RUL 95% confidence intervals (RUL CIs, blue dashed lines) were positioned with respect to the true RUL (red solid line) and the acceptability region (red dashed lines) [18]. Predicted RUL and RUL CIs were included in the acceptability regions for a wide part of the RTF execution. Nevertheless, in the central region ($\lambda = 0.3$ to 0.7), predicted RUL was slightly out of the acceptability
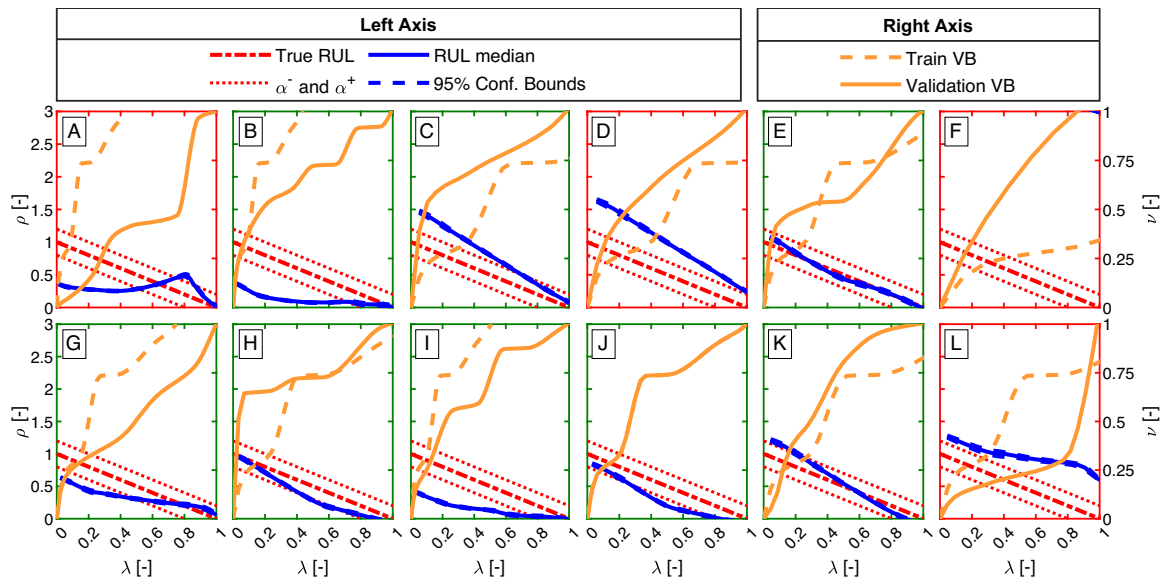
Fig. 10 Prognostic phase applied to the whole RTF set. Training was performed on RTF 2 (2) - Label H. The graph shows the evolution of the normalized RUL (ρ - blue solid line) and the associated 95% confidence bounds (blue dashed line) with respect to the normalized tool life λ [-].

Normalized VB (ν) of train (orange dashed line) and validation (orange solid line) RTFs are shown too. True normalized RUL (red dot-dashed line) and acceptability bounds (red dot lines) are compared to ρ curves. Green boxes highlight RTFs for which PH is above 20%

bounds. In the correspondent chart in Fig. 11, the associated APE (ψ: APE normalized to true EoL) was shown: ψ was almost constant and close to 0 in the first and last part of the tool life, while reaching its maximum at $\lambda = 0.5$. Indeed, for this combination of RTFs, PH reached a value of 21%.

More attention should be given to other tests, which represent unseen cutting tests and parameters. The analysis should

also consider that the results refer to a first application of the prognosis algorithm (which is the most critical condition), and as new cutting tests become available, they should be included in the algorithm training. In each graph, the evolutions of the mean flank wear of validation and training RTFs (ν: VB_m normalized to the prognosis threshold) are shown in orange to graphically compare the curves. RTFs
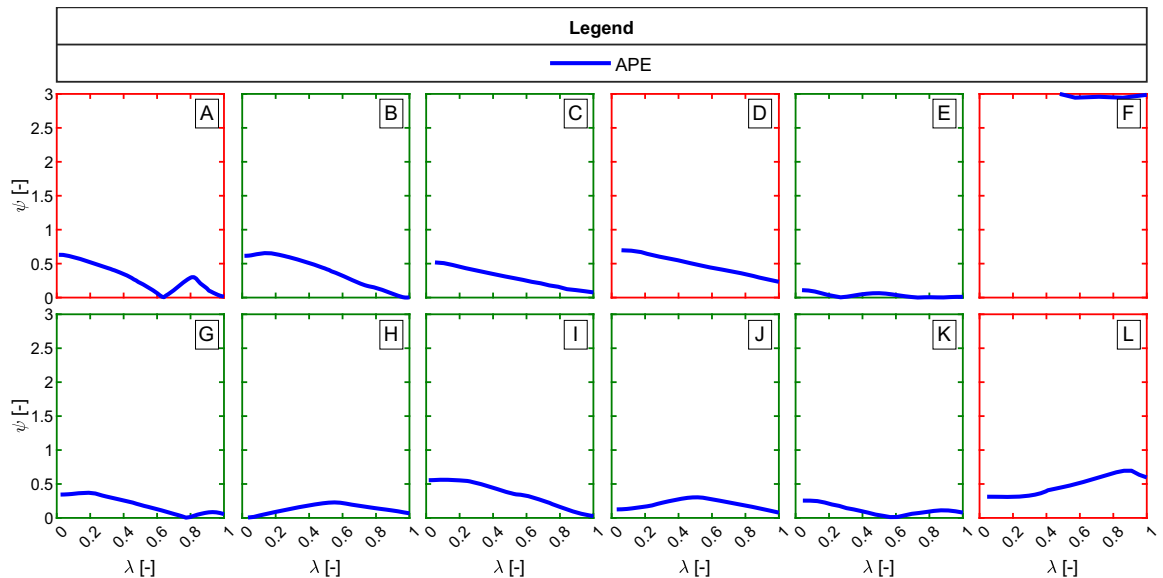

Fig. 11 Prognostic phase applied to the whole RTF set. Training was performed on RTF 2 (2) - Label H. Normalized APE ψ [-] is shown. Green boxes highlight RTFs for which PH is above 20%

can be recognized by considering the row number as the replicate number and the column index as the ID of the test (e.g. chart with letter C was validated upon RTF 3 (1)). Letter K (RTF 5 (2)) showed the largest difference between the VB_m trends of training and validation RTF, bringing to a constant, not adaptive behaviour of the prognosis algorithm. APE is almost constant and around 3 times the true EoL for this test. Furthermore, during tests A, G and L, it was not possible to predict sufficiently in advance the correct EoL of the associated drill-bits. Anyway, these tests behaved differently from test K. In fact, for these cases, an adaptation of the algorithm occurred. For test A, the adaptation led to an overshoot in the estimation of the EoL, causing the RUL estimation to overcome the upper acceptability bound for a small region around $\lambda = 0.8$. Indeed, the PH resulted lower than 20%. On the contrary, during test G, the adaptation of the predictions was not fast enough to enter in the acceptability bounds. Thus, only 4 out of 12 validation conditions (33% of the cases) showed a PH less than 20%.

Tests B, C, F, and I behaved similarly with a stable algorithm performance, allowing to estimate the correct RUL with an acceptable anticipation ($PH > 20\%$). These tests showed a constant slope in the initial phase of the tool lives, indicating no adaptability at the beginning of the drill-bit life. At abound $\lambda = 0.1$, the algorithm starts changing the slope of the RUL predictions by adapting to the new trend of the mean flank wear. This behaviour is clearly visible in the APE plots, where in the initial phase the error remains constant, whereas after 10% of tool life the error starts converging toward 0. Test D followed instead the behaviour of test H (validation on training RTF).

Table 5 reported that PHs were for the 67% of the cases over the 20% of tool lives, meaning that the RUL could be robustly predicted in advance with a minimum of 20% leading time. It must be noted that the PH was expressed as the last hole where the 95% of the estimated RUL PDF was included in the acceptability region (i.e. a really restrictive condition), indicating that the confidence on a correct prediction was really high. Thus, PHs were low in general (30% on average), despite the errors in the prediction mean were still limited (Fig. 11).

Compared to similar adaptive techniques applied in literature, the conceived solution exhibits comparable performances. In [19], the authors used Radial Basis Functions (RBFs) in place of the MLP to make prognosis of Li-ion batteries discharge. A PH of 16% was achieved, while prediction errors oscillated between 15 and 25%. The authors of [19] underlined the advantages of an adaptive approach when prognosis works outside of its training region. Worse results with respect to model-based approaches but adaptivity to changed dynamics of degradation were expected. A similar methodology was adopted by [4], where the authors dealt with structural health monitoring, i.e. crack length propagation. PHs between 55 and 63% were achieved on simulated data under fixed and variable loading conditions and experimental data under variable loading conditions. In [3], prognosis of tool wear data in turning was carried out. Here, the results were reported over 25 cases (representing 5 different training conditions). The authors obtained a PH over 20% (40% average) for the 80% of the cases. In [22], the authors proposed a solution based on particle filter and support vector regression. The study was conducted on a single experimental milling test with fixed cutting conditions. Prognosis was performed on the evolution of the wear of three flutes belonging to the same mill. PH was estimated to be at least 33%, even if the authors did not report it in the text. In [10], the authors applied prognostics to oil pump degradation, feeding indirect measurements to a relevance vector machine within an adaptability framework. PHs resulted to be 45% and 20% for two pumps. Higher PHs found in the reported literature may be relatable to multiple reasons. The measured voltage degradation curves [19], crack propagation data [4], and turning wear data [3] exhibited smoother trends in comparison with the experimental data used in this paper, which showed multiple inflection points and made prognosis more challenging. PH was also overestimated by the use of the deterministic formulation; in the case of [4], such high prediction bounds would have caused a low probabilistic PH according to the $\alpha - \beta$ criterion of [18] used in this paper. Information leakage from training to validation is another source of PH overestimation, like in [22], where prognosis was made on different flutes of the same mill).

4 Conclusions

In this paper, a complete PHM strategy for drill-bits was developed. Indirect inspection and monitoring of cutting power and current was implemented to detect drill-bit chippings, while direct inspection was the input of a hybrid adaptive prognosis solution. The main aspects dealt with in this paper are summarized in the following:

- A mean estimation of Kronenberg's specific force coefficients, set their baseline values for the instantaneous identification approach used on 12 run-to-failure tests.
- The exponents of the Kronenberg's models, representing the effect of the undeformed chip thickness on the specific force coefficients, resulted to be more sensible to tool wear, and insensible to the variability associated to the drill-bits manufacturing accuracy.
- The evolution of the exponents was tracked during the 12 run-to-failures. Monitoring them throughout each run-to-failure allowed to detect critical chippings and brittle damages on the cutting lips of the drill-bits.

- A hybrid prognosis algorithm, with particle filter and multi-layer perceptrons, allowed for adaptively estimating the remaining useful life probability density function online, providing a robust tool to support the maintenance decision-making.
- Prognosis adaptivity allowed to predict drill-bit wear under unseen process parameters, using just one training run-to-failure. A total of 67% of tool end-of-life was correctly estimated with an anticipation greater than 20% (30% on average). The method is applicable for small-batch or one-of-a-kind production scenarios, where worn data are scarce.

Future works will be oriented to the development of a heterogeneous prognosis approach, where specific force coefficients will be included in the prediction of flank wear, providing a cutting parameter independent feature, in order to update the remaining useful life prediction even when a pit-stop is not an option.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Bernini L, Albertelli P, Monno M (2023) Mill condition monitoring based on instantaneous identification of specific force coefficients under variable cutting conditions. Mech Syst Signal Process 185:109820. https://doi.org/10.1016/j.ymssp.2022.109820. https://linkinghub.elsevier.com/retrieve/pii/S0888327022008883
2. Bernini L, Malguzzi U, Albertelli P et al (2024) Hybrid prognosis of drill-bits based on direct inspection. Procedia Comput Sci 232:201–210. https://doi.org/10.1016/j.procs.2024.01.020. https://linkinghub.elsevier.com/retrieve/pii/S1877050924000206
3. Bernini L, Malguzzi U, Albertelli P et al (2024) Hybrid prognostics to estimate cutting inserts remaining useful life based on direct wear observation. Mech Syst Signal Process 210:111163. https://doi.org/10.1016/j.ymssp.2024.111163. https://linkinghub.elsevier.com/retrieve/pii/S088832702400061X
4. Cadini F, Sbarufatti C, Corbetta M et al (2019) Particle filtering-based adaptive training of neural networks for real-time structural damage diagnosis and prognosis. Struct Control Health Monit 26(12):1–19. https://doi.org/10.1002/stc.2451
5. Carbone N, Bernini L, Albertelli P et al (2023) Assessment of milling condition by image processing of the produced surfaces. The International journal of advanced manufacturing technology 124(5–6):1681–1697. https://doi.org/10.1007/s00170-022-10516-5. https://link.springer.com/10.1007/s00170-022-10516-5
6. Drouillet C, Karandikar J, Nath C et al (2016) Tool life predictions in milling using spindle power with the neural network technique. J Manuf Process 22:161–168. https://doi.org/10.1016/j.jmapro.2016.03.010. https://linkinghub.elsevier.com/retrieve/pii/S1526612516300123
7. Farhadmanesh M, Ahmadi K (2021) Online identification of mechanistic milling force models. Mech Syst Signal Process 149:1–18. https://doi.org/10.1016/j.ymssp.2020.107318. publisher: Elsevier Ltd
8. Fernández-Robles L, Sánchez-González L, Díez-González J et al (2021) Use of image processing to monitor tool wear in micro milling. Neurocomputing 452:333–340. https://doi.org/10.1016/j.neucom.2019.12.146. https://linkinghub.elsevier.com/retrieve/pii/S0925231220317501
9. Guo B, Zhang Q, Peng Q et al (2022) Tool health monitoring and prediction via attention-based encoder-decoder with a multi-step mechanism. The International journal of advanced manufacturing technology. https://doi.org/10.1007/s00170-022-09894-7. https://link.springer.com/10.1007/s00170-022-09894-7
10. Hu J, Tse PW (2013) A relevance vector machine-based approach with application to oil sand pump prognostics. Sensors (Switzerland) 13(9):12663–12686. https://doi.org/10.3390/s130912663
11. Jiménez A, Arizmendi M, Sánchez JM (2021) Extraction of tool wear indicators in peck-drilling of Inconel 718. International Journal of Advanced Manufacturing Technology 114(9-10):2711–2720. https://doi.org/10.1007/S00170-021-07058-7/TABLES/4. https://link.springer.com/article/10.1007/s00170-021-07058-7, publisher: Springer Science and Business Media Deutschland GmbH
12. Moldovan O, Dzitac S, Moga I et al (2017) Tool-wear analysis using image processing of the tool flank. Symmetry 9(12):296. https://doi.org/10.3390/sym9120296. http://www.mdpi.com/2073-8994/9/12/296
13. Montgomery DC (2008) Introduction to statistical quality control. Wiley Inc
14. Nouri M, Fussell BK, Ziniti BL et al (2015) Real-time tool wear monitoring in milling using a cutting condition independent method. International Journal of Machine Tools and Manufacture, publisher:Elsevier 89:1–13. https://doi.org/10.1016/j.ijmachtools.2014.10.011
15. Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern 9(1):62–66. https://doi.org/10.1109/TSMC.1979.4310076. http://ieeexplore.ieee.org/document/4310076/
16. Peng Y, Dong M, Zuo MJ (2010) Current status of machine prognostics in condition-based maintenance: a review. Int J Adv Manuf Technol 50(1–4):297–313. https://doi.org/10.1007/s00170-009-2482-0
17. Rizzo A, Goel S, Luisa Grilli M et al (2020) The critical raw materials in cutting tools for machining applications: a review. Materials 13(6):1377. https://doi.org/10.3390/ma13061377. https://www.mdpi.com/1996-1944/13/6/1377
18. Saxena A, Celaya J, Saha B et al (2010) Metrics for offline evaluation of prognostic performance. International Journal of Prognostics and Health Management 1(1):2153–2648. https://doi.org/10.36001/ijphm.2010.v1i1.1336
19. Sbarufatti C, Corbetta M, Giglio M et al (2017) Adaptive prognosis of lithium-ion batteries based on the combination of particle

filters and radial basis function neural networks. J Power Sources 344:128–140. https://doi.org/10.1016/j.jpowsour.2017.01.105

20. Shi D, Gindy NN (2007) Tool wear predictive model based on least squares support vector machines. Mech Syst Signal Process 21(4):1799–1814. https://doi.org/10.1016/j.ymssp.2006.07.016. https://linkinghub.elsevier.com/retrieve/pii/S0888327006001725

21. Stavropoulos P, Papacharalampopoulos A, Vasiliadis E et al (2016) Tool wear predictability estimation in milling based on multi-sensorial data. Int J Adv Manuf Technol 82(1–4):509–521. https://doi.org/10.1007/s00170-015-7317-6

22. Wang J, Zheng Y, Wang P et al (2017) A virtual sensing based augmented particle filter for tool condition prognosis. J Manuf Process 28:472–478. https://doi.org/10.1016/j.jmapro.2017.04.014. https://linkinghub.elsevier.com/retrieve/pii/S1526612517300865

23. Wickramarachchi CT, Rogers TJ, McLeay TE et al (2022) Online damage detection of cutting tools using Dirichlet process mixture models. Mech Syst Signal Process 180:109434. https://doi.org/10.1016/j.ymssp.2022.109434. https://linkinghub.elsevier.com/retrieve/pii/S0888327022005520

24. Xu X, Lu Y, Vogel-Heuser B et al (2021) Industry 4.0 and Industry 5.0—Inception, conception and perception. J Manuf Syst 61:530–535. https://doi.org/10.1016/j.jmsy.2021.10.006. https://linkinghub.elsevier.com/retrieve/pii/S0278612521002119

25. Yan B, Zhu L, Dun Y (2021) Tool wear monitoring of TC4 titanium alloy milling process based on multi-channelsignal and

time-dependent properties by using deep learning. J Manuf Syst 61:495–508. https://doi.org/10.1016/j.jmsy.2021.09.017. https://linkinghub.elsevier.com/retrieve/pii/S027861252100203X

26. Yu J, Liang S, Tang D et al (2017) A weighted hidden Markov model approach for continuous-state tool wear monitoring and tool life prediction. The International journal of advanced manufacturing technology 91(1):201–211. https://doi.org/10.1007/s00170-016-9711-0

27. Zhang X, Wang S, Li W et al (2021) Heterogeneous sensors-based feature optimisation and deep learning for tool wear prediction. Int J Adv Manuf Technol 114(9-10):2651–2675. https://doi.org/10.1007/S00170-021-07021-6/FIGURES/22. https://link.springer.com/article/10.1007/s00170-021-07021-6, publisher: Springer Science and Business Media Deutschland GmbH