Contents lists available at SciVerse ScienceDirect

CIRP Annals Manufacturing Technology

Journal homepage: www.elsevier.com/locate/cirp

Cloud-Enabled Prognosis for Manufacturing

[a]R. Gao (2), [b]L. Wang (2), [c]R. Teti (1), [d]D. Dornfeld (1), [e]S. Kumara (1), [f]M. Mori (1), and [g]M. Helu

[a]Department of Mechanical and Aerospace Engineering, Case Western Reserve University, Cleveland, OH, USA
[b]Department of Production Engineering, KTH Royal Institute of Technology, Stockholm, Sweden
[c]Department of Materials and Production Engineering, University of Naples Federico II, Naples, Italy
[d]Department of Mechanical Engineering, University of California, Berkeley, CA, USA
[e]Department of Industrial & Manufacturing Engineering, The Pennsylvania State University, University Park, PA, USA
[f]DMG Mori Seiki CO., Ltd., Nagoya, Japan
[g]Engineering Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, USA

Advanced manufacturing depends on the timely acquisition, distribution, and utilization of information from machines and processes across spatial boundaries. These activities can improve accuracy and reliability in predicting resource needs and allocation, maintenance scheduling, and remaining service life of equipment. As an emerging infrastructure, cloud computing provides new opportunities to achieve the goals of advanced manufacturing. This paper reviews the historical development of prognosis theories and techniques and projects their future growth enabled by the emerging cloud infrastructure. Techniques for cloud computing are highlighted, as well as the influence of these techniques on the paradigm of cloud-enabled prognosis for manufacturing. Finally, this paper discusses the envisioned architecture and associated challenges of cloud-enabled prognosis for manufacturing.

Predictive Model, Condition Monitoring, Cloud Manufacturing

1. Introduction

Prognosis refers to forecasting the likely outcome of a situation, and typically involves two inherently related steps. First, analytical models are established to summarize the historical evolution of the situation (e.g., variation in stock price, deterioration of machine conditions, or spread of infectious disease) in a quantitative manner. These models are then modified by updated information to predict the future development of the situation. The predicted value is associated with a confidence level, which results from the uncertainty involved in the prediction process.

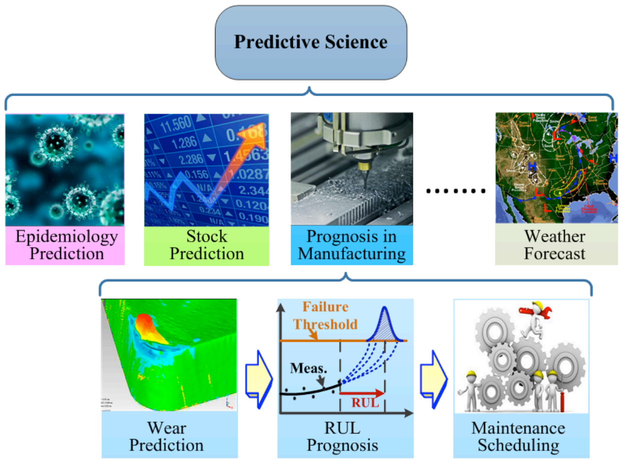

Fig. 1. Predictive science and its application in manufacturing

Prognosis has been investigated for a wide range of applications, including disease [1] and epidemiology prediction [2], weather forecasting [3], and maintenance scheduling [4] (Fig. 1). In the context of manufacturing, prognosis has been used to identify short-term and long-term actions or decisions to estimate the remaining useful life (RUL) of a machine or a system [5-9] based on the conditions monitored and diagnosis obtained [10-12]. It provides a scientific and technological basis for maintenance scheduling, asset management, and more reliable system design [13-14].

1.1. Benefits of prognosis for manufacturing

The operational reliability of industrial machines and assets significantly influences the sustainability of manufacturing [15] and competitiveness of the industry. Because the operational reliability of a machine system decreases as the duration of its operation progresses, ensuring reliability during the designed lifecycle of the machine becomes a critical task for maintenance [16-17]. In traditional time-based maintenance, actions (e.g., machine inspections) are performed periodically at preset intervals independent of a machine's current operation condition [8]. Although such an approach is effective in reducing equipment failures, it generally does not provide information on the RUL of a machine. Furthermore, time-based maintenance can be a major expense with the increasing complexity of machines and equipment in modern manufacturing.

Addressing this challenge, condition-based maintenance (CBM) has been developed as a maintenance strategy that schedules activities based on the result of condition measurements without interrupting normal machine operations [18-19]. Fault (or defect) diagnosis is a critical part of this process that links the identified abnormal behaviors in a machine to possible root causes [20-22]. Maintenance actions may then be performed based on the identified failure type and underlying mechanism [23]. With the advancement of predictive science, prognosis has been increasingly recognized as a valuable complement to CBM in manufacturing. This has led to a more efficient maintenance approach termed intelligent preventive maintenance (IPM), which minimizes the machine down time, maintenance cost, and reliance on human experience for maintenance scheduling.

Failure in a machine progresses through several stages from failure initiation to functional failure. Predictive techniques can help determine how quickly a machine's functional degradation is expected to progress from its current state to its final failure [24-

25]. An important element in devising a preventive maintenance strategy is the trade-off analysis [26]. Fig. 2 illustrates the relationship between maintenance cost and reliability of machines [13]. Preventive maintenance can specifically [27]:

- Increase system safety, improve operational reliability, and extend service life of machines
- Increase maintenance effectiveness and optimization of logistic supply chains
- Reduce maintenance costs created by repair-induced failures or unnecessary replacement of components.

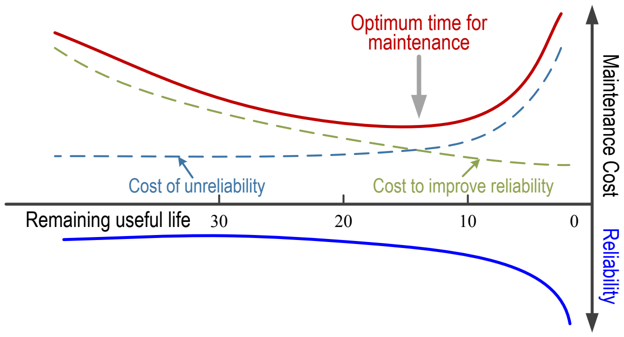

Fig. 2. Relationship between RUL, reliability, and maintenance cost, adapted from [13]

Research on prognostic technologies has grown and provides the basis for prognosis-centered maintenance. Jardine et al. [8] summarized technologies for diagnosis and prognosis that implement CBM. Peng et al. [13] and An et al. [28] reviewed typical prognostic techniques and presented a strengths-and-weaknesses analysis of the candidate techniques. Si [29] discussed statistical approaches. Sikorska et al. [30] compared different modeling options for RUL estimation, from the perspective of industry and business applications. Baraldi [31] investigated the capabilities of prognostic approaches to deal with various sources of uncertainty in RUL prediction, focusing on particle filtering (PF) and bootstrap-centered techniques. Heng et al. [26] and Sun et al. [27] discussed the potential benefits, challenges, and opportunities associated with rotating machinery prognosis.

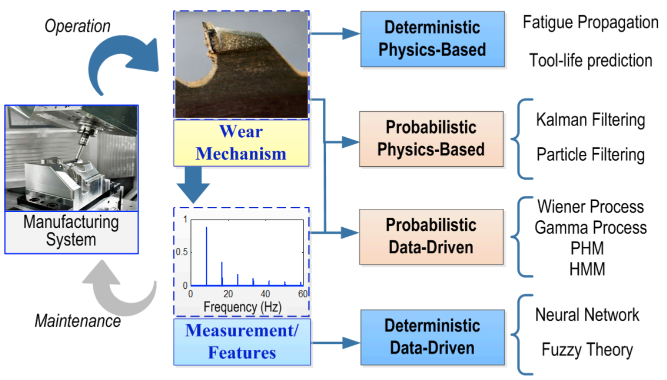

Fig. 3. Classification of prognosis methods (PHM: Proportional Hazard Modeling; HMM: Hidden Markov Model).

In past studies, prognosis techniques have been generally grouped into three categories: data-driven, physics-based, and hybrid techniques. Such classification does not specify the role of uncertainty in the outcome of prognosis. To address this limitation, this paper classifies prevailing prognostic techniques in two categories (Fig. 3): deterministic and probabilistic. Within each category, both physics-based and data-driven models are analyzed, and representative algorithms are comparatively discussed in terms of the strengths and limitations. Specifically, physics-based prognostic models in the probabilistic category, such as Kalman filtering and particle filtering, enable both predicted values and associated confidence levels, which provides insight into the uncertainty involved in the prediction process.

1.2. Cloud-enabled prognosis and cloud manufacturing

Motivated by the potential of cloud computing [32-33] and cloud manufacturing [34-37], cloud-enabled prognosis represents a new type of service-oriented technology to support multiple enterprises in deploying and managing prognostic service over the Internet. The architecture of cloud-enabled prognosis is illustrated in Fig. 4. First, machine condition monitoring realized by sensors and data acquisition systems gather data remotely and dynamically on the shop floor. Based on these measurements, remote data analysis and degradation root-cause diagnosis and prognosis are then performed. For this purpose, collaborative engineering teams can provide expert knowledge in the cloud, which forms the knowledge base that can be referenced by users through the Internet. The result of prognostic service and estimation of time-to-failure form the basis for preventive maintenance planning, which can be remotely and dynamically materialized on the factory floor [38].

The advantages of cloud-enabled prognosis include:

- <u>Collaboration and distribution</u>: The cloud enables a new paradigm where machine prognosis is a remote service instead of a conventional, centralized approach. Currently, most prognosis work is still confined to lab research and study. But, cloud-based information sharing and fusion can help transfer this work to customers who can use the cloud to find the appropriate prognostic models and data.
- <u>Accessibility and promotion of robustness</u>: A cloud-enabled platform as an integrated solution for modular and configurable prognostic services can increase the robustness of existing manufacturing processes. Pay-as-you-go prognostic services and varying maintenance options can be picked from the cloud when necessary or applicable.
- <u>Computation efficiency and data storage</u>: Cloud-enabled computation provides efficient computing cycles for complex prognosis calculation due to its high speed (parallel computing) and low communication overhead.

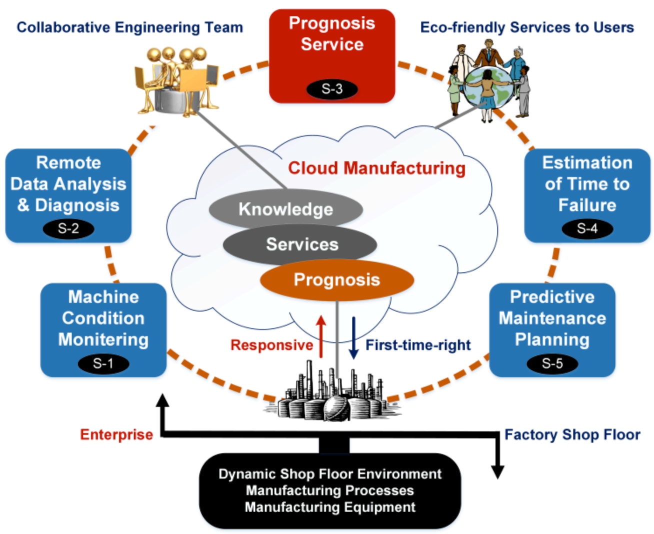

Fig. 4. Architecture of cloud-enabled prognosis.

Similar to the emergence of cloud-enabled prognosis, there is an on-going paradigm shift in manufacturing towards global manufacturing networks, which adopt new computing and Internet-based technology such as cloud computing (CC), to meet new challenges. This development leads to the flexible usage of globally distributed, scalable, and service-oriented manufacturing resources. Sharing resources, knowledge, and information between geographically distributed manufacturing entities can make them more agile and cost-effective with better utilization of resources. The success of many manufacturing firms relies on the distribution of manufacturing capacities around the globe [39].

Li et al. [40] provided one of the initial introductions to Cloud Manufacturing (CM), but the core ideas of the concept originated with research on Manufacturing-as-a-Service (MaaS) [41]. The most prominent and promising feature of CM is the seamless and convenient sharing of a variety of distributed manufacturing resources, which helps realize MaaS. Cloud manufacturing can be regarded as an integrated cyber-physical system that can provide on-demand manufacturing services digitally and physically to best utilize manufacturing resources. As illustrated in Fig. 5, CM encompasses the entire manufacturing process within a cloud-enabled environment, from order placement and product design to machining and facilities maintenance [34, 42] where cloud computing represents the core competence of CM. Based on this concept, more companies in the future may obtain various manufacturing services through the Internet as conveniently as obtaining utilities in daily operations.

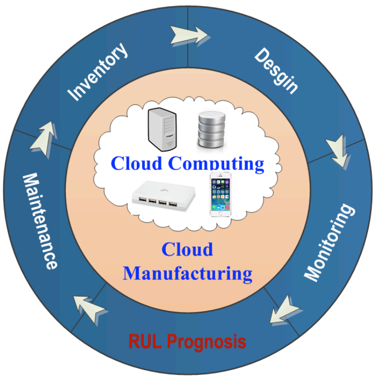

Fig. 5. Cloud manufacturing in view of prognosis.

Within the context of CM, prognosis can support the prediction of resource availability worldwide in addition to predicting machine status and facility performance, which helps determine the most effective and efficient means to manufacture a particular product. Cloud-enabled prognosis shares information from similar machines at different stages of service and utilizes the power of cloud computing for effective and efficient decision making. Challenges in accomplishing this goal include network bandwidth and data transmission speed, security, privacy, reliability, and robustness, which are discussed in this paper.

2. Prognostic methods

Prognosis determines the expected progression of degradation in a machine or its components from its current state to functional failure, and the confidence associated with the prediction. The confidence level quantifies the uncertainty that affects the RUL prediction [43]. In this paper, technologies used for prognosis are classified into two groups: deterministic and probabilistic. Model-associated information (e.g., the required knowledge and data to establish the model or model type) constructs the differences among various methods (Table 1).

Machine-specific data has an essential role in prognosis. Data used in various prognostic models can be categorized into condition monitoring data and event data. Condition monitoring data refer to the data measured by sensors (e.g., force, vibration [44], acoustic emission [45], or temperature) that are reflective of the current health condition or state of the machines [11-12]. Characteristic features can be extracted from the raw data and used as input to establishing analytical models for RUL estimation. Event data includes information on what happened (e.g., installation, breakdown, and overhaul) and what was done (e.g., component change and preventive maintenance) to the machine or component.

Approach	Method	Theory
Deterministic: Physics-based	Taylor model, Paris' law, etc.	Taylor's equation: $VT^n = C$ Paris' law: $da/dN = C\Delta K^m$
Deterministic: Data-driven	Neural networks	$X(t) = \beta_0 + \varepsilon + \sum \beta_{1,h}\varphi\beta_{2,h} + \sum\left[\beta_{1,h}\varphi\sum\beta_{k+2,h}X(t-i)\right]$
	Fuzzy theory	if ($X(t) = A_0$, $X(t\text{-}1) = A_1$ ... then $X(t+1) = f_j$
Probabilistic: Data-driven	Wiener process	$X(t) = \lambda t + \sigma B(t)$
	Gamma process	$X(t) - X(t-1) = D_t \propto Ga(k_t - k_{t-1}, \theta)$
	PHM	$X(t) = X_0(t)\exp\left(\sum \gamma_i x_i(t)\right)$
	HMMs	$\mathrm{argmax}(\mathrm{P}(X_{0:k} \| Z_{0:k}))$,
Probabilistic: Physics-based	KF based models	$P(X\|Z)=P(Z\|X)\mathrm{P}(X)/P(Z)$
	PF based models	$p(X_{0:k} \| Z_{1:k}) = \sum_{i=1}^{N_s} w_k^i \delta(X_{0:k} - X_{0:k}^i)$

Table 1. Classification and theory of prognostic methods.

2.1. *Deterministic models*

Table 2 introduces the general characteristics of deterministic models. Deterministic models only provide prediction values for the next one or few time steps without a confidence evaluation for the prediction.

Table 2. Strength and weakness of deterministic methods.

Method	Strength	Weakness
Taylor model and Paris' law	• Accurate prognosis output • Less training data needed than data-driven methods	• Difficulty in modelling stochastic systems • Limited application due to specific training needs
Neural networks	• Nonlinear and complex system modelling • Prior information not required	• Large amount of training data required • Prognosis output without confidence level
Fuzzy theory	• Modeling in continuous mathematics of fuzzy sets • Processes vague and imprecise input	• Fuzzy logic rules developed manually • Lacks learning capability

2.1.1 Physics-based models

Physics-based approaches provide a reliable and accurate estimate of all modeling options by estimating the RUL using a mathematical representation of the physical behavior of the degradation processes. The difficulty is that this process requires detailed and complete knowledge of the system behavior, which is not readily available for many manufacturing systems. Moreover, the majority of coefficients involved in the physical models need to be determined experimentally, which makes physical models application specific.

A common approach for assessing machining performance is tool life or wear, which directly correlates the tool life to the applied machining parameters (e.g., cutting speed, temperature, and feed rate). Among physical models describing tool life, an important branch is based on Taylor's tool life equation. As described in Mills and Redford [46], Taylor's basic equation relates tool life to cutting speed in a reverse exponential relationship, $VT^n = C$, where the exponential coefficient n is experimentally determined. Hoffman [47] and Niebel et al. [48] introduced an extended Taylor's equation by including the effects of feed rate and cutting depth on tool life. Workpiece hardness has also been considered in the extended Taylor's equation [49].

Lau et al. [50] proposed a relationship between tool geometry, i.e., rake angle and clearance angle, and the Taylor Constant. Also, experiments by Quinto [51] and Oxley [52] indicated that the cutting temperature dominates the tool life. It is straightforward and convenient to predict tool life using Taylor's basic equation and its extended version. However, coefficients in these equations are empirically determined and only work for particular combinations of tools and workpieces.

Tool wear rate models provide information about wear growth rate (volume loss per unit contact area/unit time) due to some wear mechanisms (e.g., abrasive wear, adhesive wear, and diffusion wear), as shown in Fig. 6 [53]. Different mechanisms can derive different dominant equations for the same wear type. Usui et al. [54] derived a wear rate model for carbide tools based on the adhesive wear and discussed the effect of normal stress, cutting temperature, and speed on the wear rate. Usui et al.'s [54] equation is very practical for the implementation of tool wear estimation using the finite element method (FEM). Choudhury and Srinivas [55] developed a mathematical model to estimate the flank wear rate by means of the index of the diffusion coefficient and other cutting parameters, such as rubbing velocity and clearance and rake angles.

It has been indicated experimentally that the cutting velocity and the index of diffusion coefficient have the most significant effect on tool wear rate [56]. The tool wear rate model can be seen as a particular type of crack growth model or fatigue spall progression model. Generally, a crack growth model is characterized by the stress intensity factor at the tip of a crack $K = f(a, \sigma)$, where a is the half crack length and σ is the nominal stress. Theoretically, the crack is assumed to not propagate when K is smaller than a threshold value. After exceeding the value, the crack growth rate will be governed by a power law, such as Paris' law $da/dN = C\Delta K^m$, where C and m are material parameters [57]. However, Paris' law does not account for the mean stress effects and is only valid under conditions with uniaxial loading and "long cracks.".

To improve Paris' law, Pungo et al. [58] proposed a new equation generalizing Paris' law by replacing the intensity factor with a suitable mean stress called the "fracture quantum" and obtaining an appropriate threshold value by interpolating between the Paris and Wöhler regimes. Li et al. [59] studied the rolling element bearing defect growth rate under the existence of information about instantaneous defect size and material constants based on Paris' formula. Aslantas and Tasgetiren [60] extended Paris' formula by mixing mode stress intensity factors to develop an analytical model for two-dimensional rolling sliding contact situations.

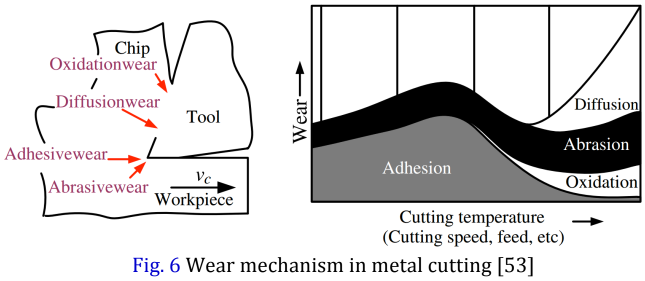

Fig. 6 Wear mechanism in metal cutting [53]

Significant effort has been devoted to devise accurate models for crack propagation studies. Newman et al. [61] selected the crack-tip constraint factors accounting for three-dimensional state-of-stress effects to correlate long crack growth rate data as a function of the effective-stress-intensity factor. Ioannides and Harris [62] proposed a statistical postulation between the probability of survival, the fatigue life, and the stress-related fatigue criterion level for the prediction of fatigue life in rolling bearings. Kotzalas and Harris [63] established a spall progression life mathematical model for ball endurance testing by extending the Ioannides-Harris fatigue life theory. Choi and Liu [64] defined a crack resistance index based on Ioannides-Harris theory and a wear resistance index based on abrasive wear characteristics to model the spall progression of rolling contact. Qiu et al. [65] developed a stiffness-based model to achieve accurate bearing prognosis considering the bearing system as single-degree-of-freedom system with its natural frequency related to the system stiffness. More recently, Fan et al. [66] proposed a mathematical model for the wear analysis of the slide guideway under cutting conditions by revealing the inherent interactions between cutting force, wear, and deformation of the slide guideway, geometric errors, and final accuracy degradation of machine tools. Similar researches on determinations on crack propagation models can be found in [67-69]. Like other physics-based models, these techniques require the empirical estimation of various model parameters.

2.1.2 Artificial intelligence-based data-driven models

In this section, artificial neural network and fuzzy logic are introduced. The common approach for both methods in RUL estimation is to determine the next measurements or extracted feature indices based on the values measured at several preceding time units. A neural network applies historical data to train a model, which is in turn used for prediction. Fuzzy logic compares the transformed input to a series of fuzzy rules to obtain the prediction.

2.1.2.1 Artificial Neural Network

An Artificial Neural Network (ANN) provides an estimated output result for the remaining useful life of a machine or component based on measured condition-monitoring data or event data rather than a physical understanding of the wear or failure mechanism [1.18]. Because ANN is a purely data-based method, it is insensitive to linear or nonlinear characteristics of a studied system and does not require an analytic expression of the system behavior. Its drawbacks include that: a) it requires a comprehensive data set to train the model; b) its performance relies largely on the selected model (network architecture, activation function, etc.); and c) it provides no uncertainty quantification on the estimated output.

The basic components of a neural network are nodes and associated weights, which are connected in the layer format. The ANN learns an unknown function by adjusting its weights with observations of input and output. Based on the input source of each node, the neural network architecture can be classified as a feed-forward network or dynamic network. The inputs to the nodes of a feed-forward network, such as multilayer perceptron (MLP) with back-propagation network (BP) training algorithm, rely only on the output of the preceding layer under the current iteration. For a dynamic network, such as recurrent neural network (RNN), general regression neural network (GRNN), or time delay network [70], the nodes' inputs also depend on information from previous iterations. These networks are supervised learning algorithms, which require the actual outputs for training. Conversely, input stimuli can be automatically clustered without external output information in unsupervised learning, such as self-organizing map (SOM) [71].

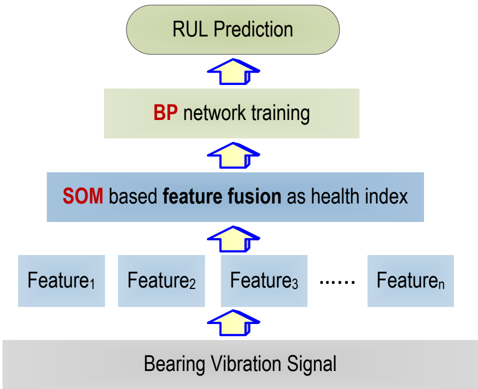

Fig. 7. Flow diagram for bearing RUL prognosis, adapted from [71].

The common approach of ANN in prognosis is to predict the next point of a data series in the fault propagation process and estimate the RUL when the extrapolated data reaches a predefined final failure value. The network takes types of data values at the present and previous inspection points as the inputs and provides values at the next one or several time units as the outputs. The input data can be condition monitoring data and its characteristic features (e.g., the time or frequency features of the vibration and force signal) or mutual information from multiple features (e.g., output of PCA [72] or ANN). Huang et al. [71] generated a new feature through SOM as the bearing life degradation indicator, which was passed to the BP neural network as the input to estimate the RUL of ball bearings as shown in Fig. 7. Ghosh et al. [73] adopted a MLP network to fuse features extracted from a number of machining zone signals to estimate the average flank wear of the main cutting edge of the cutting tool. Age and multiple condition monitoring measurements were fitted for a failure history through a generalized Weibull-FR function by Tian [74], who then used the fitted measurement value as the ANN input.

Chen and Huang [75] applied a traditional MLP network to predict the in-process surface roughness of a cutting tool in end milling operations. A similar approach was adopted by Mahamad et al. [76] to obtain a dependable trend of tool wear curves for optimal utilization of tool life and increase of productivity while preserving the surface integrity of the machined parts. D'Addona et al. [77] used a MLP-combined Levenberg-Marquardt (LM) training algorithm for accurate RUL prediction of bearing failure; the input of the network was time and fitted measurement Weibull hazard rates. Similar findings can be tracked back to Huang and Zhang [78], whose results indicated that the modeling of RUL of laboratory-tested bearings performed better than reliability-based approaches.

Pontes et al. [79] presented a study on the applicability of RBF neural networks for predicting the average roughness in turning processes while the parameters of the network, such as the number of radial units, were determined through Taguchi's orthogonal arrays. Malhi et al. [70] investigated a modified RNN approach to improve the long-time prediction accuracy with application to machine condition monitoring while the dynamic input to the RNN was pre-processed using competitive learning rules to cluster the input data for long-term prediction. Herzog et al. [80] compared the performance between an MLP neural network trained with a LM algorithm and a GRNN neural network with the application to RUL prediction of machines and components. Their results showed that the GRNN had a comparable result to the MLP approach.

It should be noted that one developed neural network generally cannot be extended to other neural network architectures, other kinds of machining operations, or other materials or tools. Also, most of the described works are still laboratory experiments without further evaluation from on-site field tests. One drawback of ANN is its inability to process linguistic and inaccurate input data. To overcome this problem, past research has focused on integrating ANN with other methods, such as expert systems [81] and Bayesian inference [82].

2.1.2.2 Fuzzy system

The fuzzy logic system is a technique for arriving at a definite conclusion using linguistic rules rather than empirically derived *if-then* rules. Compared to traditional expert systems and other estimation techniques, fuzzy systems enable: 1) modeling system behavior in continuous mathematics of fuzzy sets rather than discrete statements (true or false) and offering a reasonable compromise between rigorous analytical modeling and purely qualitative simulation; and 2) qualitative and imprecise reasoning statements to be incorporated with rule bases, which enables these systems to process vague, imprecise, and noisy inputs.

For system behavior and state forecasting, a fuzzy system estimates future system states based on the information collected from previous states. To differentiate the impacts of inputs at different times on the next step value prediction, information weights are added to previous states. However, fuzzy logic systems have major drawbacks: 1) fuzzy rules are always developed by experts, and so fuzzy logic is not considered when there is not sufficient knowledge and experience for one problem; and 2) fuzzy systems lack learning capability, so they need to be integrated with other techniques, such as ANN.

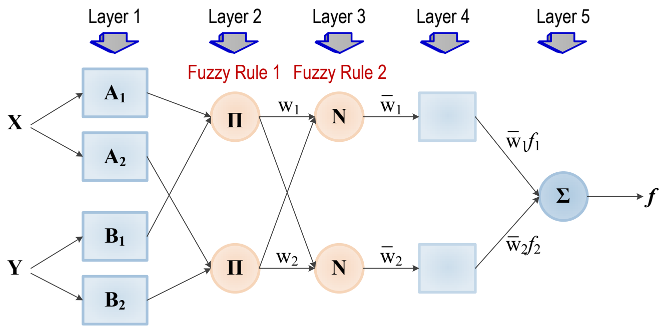

Fig. 8. Neuro-Fuzzy system for prediction, adapted from [85]

There are few examples where a fuzzy logic system has been used as the primary method for RUL prediction. Pan et al. [83] assessed the bearing performance degradation based on fuzzy c-means and lifting wavelet packet decomposition, where the latter provides extracted features as the degradation indicator to the former. This assumed, though, that all data are projected into only two states {normal, final failure}. Zhao et al. [84] attempted to extend the traditional binary state space into a fuzzy state space based on profuse reliability theory to characterize a gradual physical degradation. Wang et al. [85] adopted a neuro-fuzzy (NF) system to develop an on-line machine fault prognosis system; a typical NF structure is shown in Fig. 8. They also indicated that once an NF system is properly trained, it performs better than RNNs in predicting nonlinear and chaotic time series. Liu et al. [86] developed an adaptive multi-step predictor based on a weighted recurrent NF paradigm for system state forecasting and adopted a hybrid training algorithm based on recursive LM and least square estimates to enhance forecasting convergence and accommodate time-varying system conditions. For the same purpose of multi-step ahead prediction of a low methane compressor, Tran et al. [87] adopted an approach based on an adaptive neuro-fuzzy inference system (ANFIS) and indicated that this method was capable of tracking changes in a machine's operating conditions with high accuracy. Other works related to NF-based prognosis have been shown in [88-89].

2.2. Probabilistic models

Probabilistic models assume that system performance degradation or fault deterioration follow a certain distribution, such as Gaussian or Gamma distribution. The benefit of this modeling concept is that both a prediction value and an associated confidence level are provided. As shown in Fig. 2, probabilistic models can be further classified into data-driven (i.e., Wiener and Gamma process, proportional hazard modeling or PHM, and hidden Markov models or HMM) and physics-based (i.e., KF and PF based models) techniques. The strengths and weakness of probabilistic models are summarized in Table 3.

Table 3. Strength and weakness of probabilistic methods.

Method	Strength	Weakness
Wiener process model	• Accumulative degradation modeling • Confidence provided for corresponding FPT	• Limited to Gaussian and linear system • Difficulty in processing monotonic degradation
Gamma process model	• Straightforward degradation modeling • Monotonic system modeling	• Limited to monotonic degradation • Prognosis independent of the historic behavior
PHM	• Integration of different data sources • Quantification of failure probability	• Times to failures assumed to be mutually independent • Comprehensive covariates set required
HMMs	• Degradation with different stages modeling • Underlying physics not required	• Large amount of training data required • Application limited to Gaussian process
KF-based models	• Dynamic process modeling • High computation efficiency	• Limited to linear or weak nonlinear system • Limited to Gaussian process
PF-based models	• Nonlinear and non-Gaussian modeling • Confidence provided with prognosis output	• Large amount of particles required • High computation effort required

2.2.1 Data-driven (statistics-based) models

All statistics-based methods, including the Wiener process, Gamma process, proportional hazard models, and hidden Markov model are essentially regression models that use measurements or extracted features as the indicators and forecast the accumulated degradation state given the current state in the future. Each approach uses different assumptions for the distribution of noise involved in the model, the distribution from the current state to the state in the future, and the data type used in the models.

2.2.2.1 Wiener process

The Wiener process is an advanced stochastic regression model with random noise that can be used for modeling degradation processes and RUL prediction. It was first proposed to model the movement of small particles in fluids and air with small fluctuations. The Wiener process can be used to model the path of degradation processes where successive and accumulative fluctuations in degradation can be observed. A general Wiener process is defined as:

$$X(t) = \lambda t + \sigma B(t)\,, \tag{1}$$

where $X(t)$ is condition-monitoring data, λ denotes the drift coefficient, σ denotes the diffusion coefficient, and $B(t)$ denotes the standard Brownian motion representing the stochastic dynamics of the degradation process. The first term is also the mean of the estimated degradation path, which implies that the Wiener process is a linear process. The second term provides the process covariance $\sigma^2 t$, which also represents the uncertainty of degradation. The RUL can be obtained by estimating the time when $X(t)$ first passes a predefined threshold w, which is called the first passage time (FPT), noted as T = inf $\{t$: $X(t) \geq w \mid x_i\}$, where x_i is the observation at time t_i and T is estimated based on x_i. The probability distribution of T conditioned on x_i is [90]:

$$f_{T|x_i}(t \mid x_i) = \frac{w - x_i}{\sqrt{2\pi (t-t_i)^3 \sigma^2}} \exp\left(-\frac{\left(w - x_i - \lambda (t-t_i)^2\right)}{2(t-t_i)^2 \sigma^2} \right), \tag{2}$$

with mean $t_i + (w\text{-}x_i)/\lambda$ and variance $(w\text{-}x_i)\sigma^2/\lambda^3$. It is directly noted that the RUL is estimated given the current time and tightly related to the current measurement and the selection of two parameters λ and σ in the process.

The Wiener process has the following assumptions or limitations regarding practical application:

- The estimation of degradation uses only the current measurement data. This assumption however can introduce problems as shown in Fig. 9.
- Wiener process was designed to model the non-monotonic motion of small particles. Thus, it is inappropriate to process the monotonic machine degradations.
- The mean representation of modeled degradation, λt, is linear, and thus the application limitation exists when handling nonlinear situations.

Si et al. [91] attempted to incorporate a Wiener process based degradation model with historical data to overcome the limitation shown in Fig. 9 based on a recursive filter. The recursive filter was used to update the drift coefficient while the expectation maximization algorithm was adopted to update all other parameters. Also, the distribution of drift coefficient was taken into account when updating, which led to uncertainty reduction in the estimated RUL. Taking the nonlinearity and product-to-product variability of the degradation into account, Wang et al. [92] presented an adaptive method of RUL estimation based on a generalized Wiener process that subsumes several existing models. Parameters involved in population-based degradation characteristics were obtained by using maximum likelihood algorithm and the parameters that describe the specific degradation model were estimated by using the Markov chain Monte Carlo method. Wang et al. [93] proposed an additive Wiener process-based prognostic model for hybrid deteriorating systems that are made up of both linear and nonlinear parts. Correspondingly, the additive model included both linear and nonlinear degradation; the stochastic parameters in both parts of the model were correlated and estimated by using the condition monitoring observations based on a Bayesian framework.

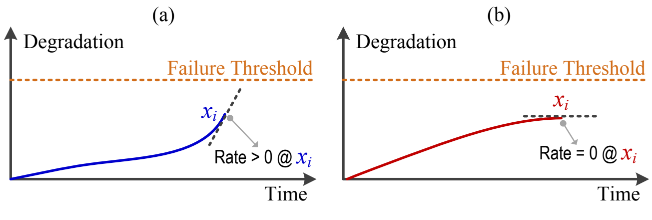

Fig. 9. Different degradation rates at the same current measurement x_i, adapted from [119].

2.2.2.1 Gamma process

In contrast to the Wiener process, which is a non-monotonic process, the Gamma process monotonically models gradual degradation accumulating over time, such as wear, crack growth, and corrosion [94]. The Gamma process assumes that the degradation accumulation follows the Gamma distribution:

$$X_t - X_{t-1} = D_t \propto Ga\left(k_t - k_{t-1}, \theta\right), \quad (3)$$

where Ga(·) denotes the Gamma distribution and k and θ are the shape and scale parameters, respectively, which describe the stochastic distribution. For a Gamma process at time t, the mean is k_t/θ, the variance is k_t/θ^2, and the ratio of the variance and the mean equals $1/\theta$, which is time independent. This also indicates that the degradation increment at a different time interval is independent, which can be seen as a Markov assumption. Under the assumption of modeling the temporal variability in the deterioration with a Gamma process, the increase in the expected deterioration over time remains unanswered. In other words, it is unclear how to estimate the two parameters k and θ [94]. The most common method used is the maximum likelihood, which is applied to estimate the two parameters by maximizing the logarithm of the likelihood function of the increments.

The major advantage of degradation modeling using a Gamma process is that mathematical calculation is straightforward. However, the strict assumptions of the Gamma process limit its applications: 1) the Gamma process is only appropriate to characterize a monotonic degradation process; 2) due to its independent increment property, the estimation of a future state is independent of the historical behavior, which is similar to the assumption of a Wiener process; and 3) the noise involved in the Gamma process that is used to quantify the estimation uncertainty must follow the Gamma distribution.

Park and Padgett [95] presented an accelerated degradation model based on the Gamma process for the situation where both observed failures and degradation measures could be considered for parametric inference of system lifetime. Here, the maximum likelihood algorithm is applied to estimate the parameters in the model. This model is extended to general accelerated test models with several accelerating variables for inference based on observed failure values and degradation measurements [96]. Tseng et al. [97] proposed an optimal step-stress accelerated degradation test for a Gamma degradation process by minimizing the approximate variance of the estimated mean-time-to-failure of the lifetime distribution of the product.

2.2.2.3 Proportional hazards modeling

Different from the Wiener process or other regression models, proportional hazards modeling (PHM) provides a way to estimate the risk of a machine component or set failing over a certain amount of time based on both condition monitoring data and event data. Also, PHM assumes a multiplicative relationship between the probability of failure and the monitored condition data rather than an additive relationship as used in other regression methods:

$$X(t) = X_0(t)\exp\left(\gamma_1 x_1(t) + \gamma_2 x_2(t) + \cdots + \gamma_p x_p(t)\right), \quad (4)$$

where $X(t)$ denotes the conditional probability of failure at time t (also called the hazard function); $X_0(t)$ denotes the baseline hazard function; $x_1(t)$, $x_2(t)$,⋯ $x_p(t)$ are the covariates, which denote the measured data (e.g., force or vibration), extracted features (e.g., RMS or kurtosis), or event data (e.g., minor repair or major repair); and $\gamma 1$, $\gamma 2$,⋯ γ_p denote the corresponding coefficients to each covariate, which represent the degree of influence each covariate has on the hazard function.

The PHM expression indicates that the hazard function is affected by three factors: time or equipment aging, baseline hazard function, and selected covariates that affect the hazard rate exponentially. Usually, the baseline hazard function is assumed to be a Weibull function, and the involved parameters can be estimated through parametric regression methods. It is apparent that covariates with high correlation to failure are assigned high coefficients or weightings and those with little correlation would be assigned low coefficients. The practical application of PHM has the following assumptions and limitations:

- Times to failure are independent, but different failures can occur at the same time in practice; the occurrence of one failure can incur or worsen other failures.
- All influential covariates should be taken into account in the PHM model, but different measurement or features are only sensitive to few failures in practice. Thus it is difficult to obtain a comprehensive covariate set.
- Individual covariates are independent. This assumption sometimes contradicts the second assumption in the situation where features extracted are tightly correlated.

Jardine et al. [98] demonstrated the value of using PHM to assist maintenance professionals to interpret condition data by identifying the key risk factors and their relative influence on the health of equipment in general and wheel motors in particular. Gasmi et al. [99] established a statistical model based on Weibull PHM to capture the potential reduction in failure intensity due to the switching of operating modes and quantify the impacts of performing repair actions on the failure intensities. Tran et al. [100] incorporated a system identification technique, PHM, and a support vector machine to assess the machine health degradation and forecast the machine RUL. An extension of PHM is the proportional intensity model (PIM), which assumes a similar form to the intensity function of the stochastic process and has been widely used for optimizing maintenance decisions. Vlok et al. [101] introduced the use of PIM to estimate residual life for non-repairable systems, such as bearings, utilizing historic failure data and corresponding diagnosis measurements.

2.2.1.3 Hidden Markov Models

A hidden Markov model (HMM) is defined as a combination of two stochastic processes. The underlying stochastic process is a finite-state homogeneous Markov chain that is not observable (i.e., hidden), which affects another stochastic process that produces a sequence of observations [102]. A HMM is characterized by five elements: number of model states; number of distinct observation symbols; an initial state distribution; a state transition probability distribution; and an observation symbol probability distribution. These distributions are either mass functions in the case of discrete observations or specified using a parametric model family-common Gaussian in the case of continuous observations [103]. HMM deals with three basic problems regarding specific applications:

- Computing the probability of an observation sequence given the specific model
- Identifying the most likely state sequence that might produce the observation sequence
- Adjusting the parameters of the model to maximize the likelihood of the given observation sequence

When conducting RUL estimation, the implementation of HMMs includes two stages: training and predicting. Generally, a supervised training scheme is adopted in which measured observations used for training need to be first labelled. Typically, each HMM can only represent two states: normal and failed. Thus, if the entire life of a piece of equipment is segmented into M distinct sequential ranges, M different HMMs should be trained to characterize each range. The presentation of temporally ordered observation sequences from such a process would yield the sorts of log-likelihood trajectories.

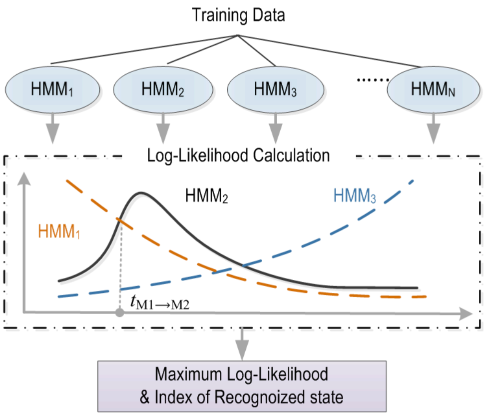

Fig. 10. Log-likelihood for different HMMs, adapted from [104].

If one HMM results in the largest log-likelihood for a given observation sequence acquired within one duration, then this HMM can be declared as the best estimate describing the process during this duration, as shown in Fig 10 [104]. Once parameters in HMMs are determined, RUL prediction is fulfilled by forecasting the progression of health states from the current state (the largest likelihood HMM) to the failure state using transition probability between states and sojourn time in each state (the duration of staying in one state) [105].

As an extension of the HMM, a hidden semi-Markov model (HSMM) assumes that the underlying process is a semi-Markov chain and each state has a variable duration. An important difference between HMM and HSMM is that one observation per state is assumed in HMM, while each state can emit a sequence of observations in HSMM. The number of observations produced while in one state is determined by the length of time spent in this state as shown in Fig 11 [106]. This characteristic makes HSMM better than HMM for RUL estimation because the propagation of health state is always a continuous progress and each stage is a collaboration of continuous values if the whole life is segmented into several stages corresponding to several underlying states in HSMM.

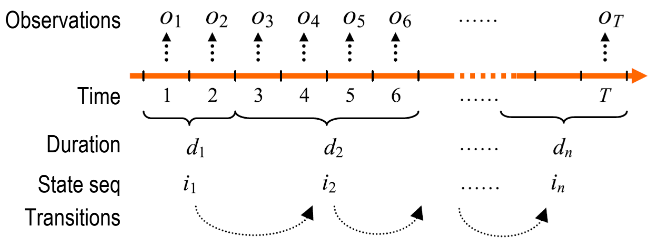

Fig. 11. General HSMM structure [106].

Regular HMMs tend to be limited in their ability to represent complex systems. More importantly, in the absence of labeled state and measurement data, the unsupervised training process is computationally tedious. In addition, regular HMMs do not have intrinsic transition probabilities between underlying states since each HMM represents a distinct health state. Hence, they require additional methods to calculate health-state transition probabilities to be utilized in RUL estimation [103].

Tobon-Mejia et al. [107] introduced the mixture of Gaussian HMM as an approach to estimate the RUL and the associated confidence level of a bearing. This approach using continuous observations derived from the sensors leads to a better representation of the bearings' deterioration. Zaidi et al. [108] predicted the future state of gear fault severity in DC machines based on HMMs using time-frequency features as machine health indicators. They also proposed a way to train HMMs from limited data. Chinnam et al. [103-104] employed HMMs for both machine diagnosis and prognosis. The RUL prediction here was simply the extension of the state identification by adding the forecasting functionally. Zhou et al. [109] attempted to combine HMM with a belief-rule base to predict hidden failure when no accurate mathematical model is available due to changing environmental factors. The belief rule base is adopted to capture the relationships between the environmental factors and the transition probabilities among the hidden states of the system. Dong and He [110] presented an HSMM-based framework and methodology for multi-sensor equipment diagnosis and prognosis. The health states of a component are modeled by the state transition probability matrix and observation probability. Based on their previous work [110], Dong [111] took the machine deterioration into consideration by introducing aging factors (constant aging factor, multiple aging factors, and exponential aging factor) into the traditional health-state-transition matrix.

2.2.2 Physics-based (model-based) approach

Probabilistic physics-based (model-based) approaches construct a relationship between measurement and the underlying physical mechanism by using measurement or extracted features to infer the machine status based on the posterior probability density function (PDF) [112-113]. For RUL prediction, once a conditional reliability function is determined, the RUL is defined as the conditional expected time to failure, given the current state. Compared to other approaches, model-based approaches provide information on all qualification measurements of the estimation: accuracy, precision, and confidence (see Fig.12). The accuracy is a measure of how close a point estimate of failure time is to the actual failure time. Precision is a measure of narrowness of an interval in which the remaining life falls. Confidence is the probability of the actual RUL falling between the bounds defined by the precision [114]. It is also a measure to quantify the estimation uncertainty.

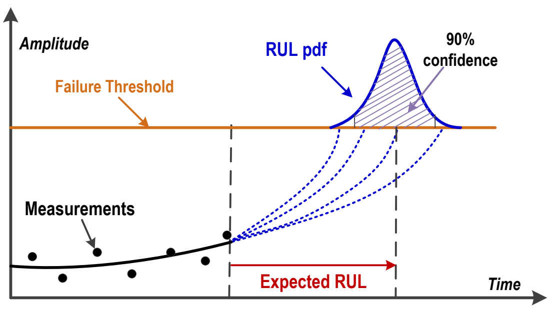

Fig. 12. RUL estimation and associated uncertainty quantification.

A general way to estimate and update the posterior PDF is Bayesian inference based on Bayes' rule:

$$P(\theta / D) = \frac{P(D/\theta)\mathrm{P}(\theta)}{P(D)}. \qquad (5)$$

Bayes' rule converts a priori probability distribution of parameter θ into a posterior distribution based on the observation D. When estimating the unknown parameter θ for the certain measurement, the usual approach is to find the value that maximizes the posterior PDF $P(\theta/D)$. Since the measurement D is a constant data index, the problem becomes maximizing $P(D/\theta)P(\theta)$. In practical applications, the probability distribution $P(\theta)$ of θ is assumed as prior information, which turns the backward PDF estimation problem into a forward estimation of maximizing $P(D/\theta)$, which is just the likelihood of the data.

The concerned uncertainty quantification or confidence calculation in the Bayesian approach is based on the assumption that the measurements will always be contaminated by noise, which should generally be random. In this context one no longer wishes to find an estimate of w, but rather to specify the belief in its value, namely $\theta \sim P(\theta/D, M)$, where M represents the choice of model [115]. Once parameters conditioned on current and previous measurements are determined, the PDF of the degradation state of the next-point is estimated based on the system model [116-117]. The mean of the PDF gives the 'best' estimates for the predictions and the covariance provides the associated confidence levels for estimates.

The whole procedure includes two phases: prediction and update, which corresponds to the time update and measurement update respectively. Given a new measurement input z^*, the predicted values are comprehensive calculation results given the probability of each parameter:

$$P(x^*/z^*, D, M) = \int P(x^*/z^*, \theta, M) P(\theta / D, M) d\theta \, . \tag{6}$$

Based on the assumptions of selected models and noise, RUL prognosis based on Bayesian inference can be implemented by more advanced estimation techniques, such as Kalman and particle filtering methods, which will be discussed in the next subsections. In Fig. 13, various methods for calculating the posteriori distribution are summarized, under the framework of Bayesian inference.

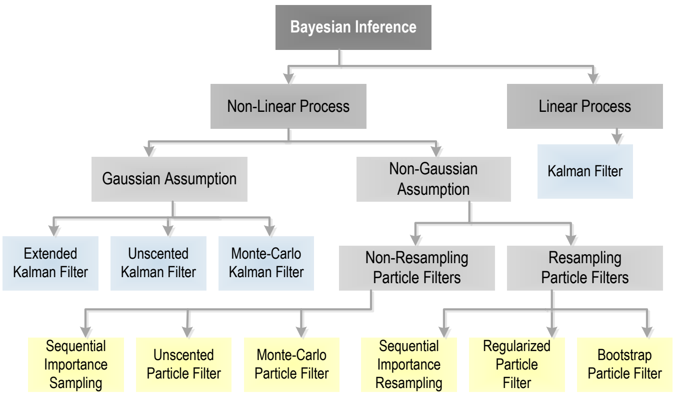

Fig. 13. Overview of methods for calculating posteriori PDF [30].

2.2.2.1 Kalman filter-based model

The Kalman filter (KF) is a computationally efficient recursive data processing technique used to optimally estimate the underlying state of a dynamic system given a set of noisy measurements in the way that minimizes the mean squared error (MSE) of predictions [118]. The general process of Kalman filter includes state and covariance prediction and update as shown in Fig. 14:

It should be noted that the state estimate is just the conditional expectation and the covariance of the estimation error is actually the same as the covariance of the state. KF is based on the Gaussian-Markov process assumption that both process and measurement noise are zero-mean white stochastic processes. Meanwhile, the initial state, process, and measurement noise are assumed to be mutually independent. Under this assumption, the KF is the optimal minimum MSE state estimator [119]. For an observable time-invariant system, the state estimation covariance will be finite and the filter will finally converge to a steady state. However, this introduces another limitation besides Gaussian-Markov assumption that the estimation model for time-variant system degradation can be unstable and its estimations divergent. The time-variant degradation scenario is common in practice, for example, the evolution of degradation of a tool can be categorized into three stages as discussed in Section 2.1.1.

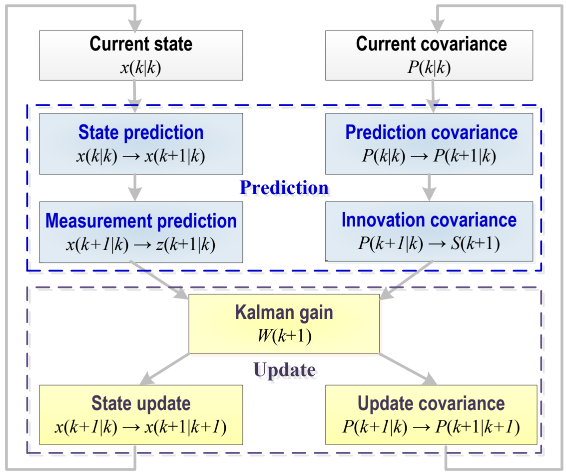

Fig. 14. General flow of Kalman filter process [119].

The utilization of a single degradation model would result in inaccurate state estimates and cause RUL prediction to diverge or fluctuate, depending on whether the degradation estimation is under or over fitted by the model [120]. A possible solution is to increase the process noise level to cover the range of the degradation evolution, but this inversely introduces another problem that high-level noise can largely decrease estimation accuracy. Furthermore, it is difficult to determine when the assumed model can be suitably applied in an online application where the historical data prior to failure is unknown.

The extended Kalman filter (EKF) is a modification of KF that attempts to break up the linearity assumption of the state evolution model and the measurement model. It still conserves the Gaussian assumption of the process noise and measurement noise. The basic idea of EKF is to linearize the nonlinear state or measurement equation through series expansion using partial derivatives around the current state prediction. However, the linearization process also transforms the noise, which may or may not remain Gaussian; this invalidates the original Gaussian assumption. Also, the covariance calculation based on a series expansion is not always accurate. Furthermore, it is very time-consuming to calculate the Jacobians (for the first-order EKF) and Hessians matrix (for the second-order EKF), which replace the transformation matrix in both models in basic KF. Bar-Shalom et al. [119] points out that if the initial error and the noises are not too large, then the EKF performs well.

The unscented Kalman filter (UKF) is an alternative to the EKF for nonlinear estimation. The UKF replaces the assumed Gaussian PDF of the state in the KF and EKF with a probability mass function (PMF) via moment matching based on an unscented transform. Unlike EKF, which approximates the nonlinear state and measurement equations using linearization, UKF uses a set of carefully chosen sample points (also called sigma points) to represent the state distribution [121]. These sample points can completely capture the true mean and covariance of the Gaussian random variables. They can also capture the posterior PDF or conditional mean and covariance accurately to the 3rd order expansion. In other words, the sigma points are symmetric and their nonlinearly transformed values will not be symmetric, which yields a better characterization of the transformed random variable than linearization approach. A simple example comparing UKF to EKF is shown in Fig. 15 for a 2-dimensional system [122]. Another advantage of UKF over EKF is that no explicit calculations of Jacobian or Hessian matrixes are required.

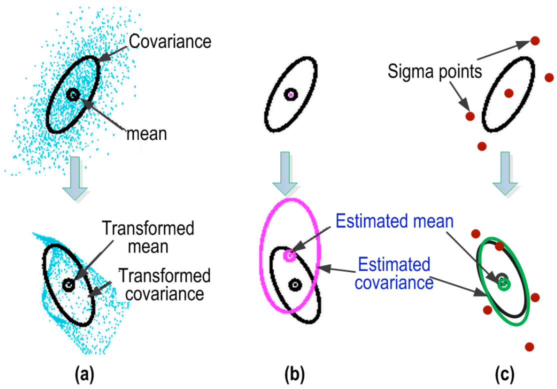

Fig. 15. Example of the UKF for mean and covariance propagation: (a) actual transformation; (b) estimation by first order linearization (EKF); (c) estimation by UKF, adapted from [122].

Another potential solution to the application limitation of KF and EKF in a practical dynamic system where parameters involved in the state evolution equations are time-varying is with multiple models (MM). Models can be in different forms or the same form with different parameters, e.g., the level of the process noise (its covariance while mean is always assumed to be zero), a deterministic input, or other parameters. MM assumes that the system match one of a bank of models most at each sampling point. The conditional PDF of the state at a specific time is obtained using the total probability theorem with respect to the exhaustive set of models. A typical representative in dynamic MM is interacting multiple models (IMM) [119], which computes the state estimate that accounts for each possible current model using a suitable mixing of the previous model-conditioned estimates depending on the current model as shown in Fig. 16. During the switching process, the algorithm undergoes a soft switching according to the latest updated mode probabilities.

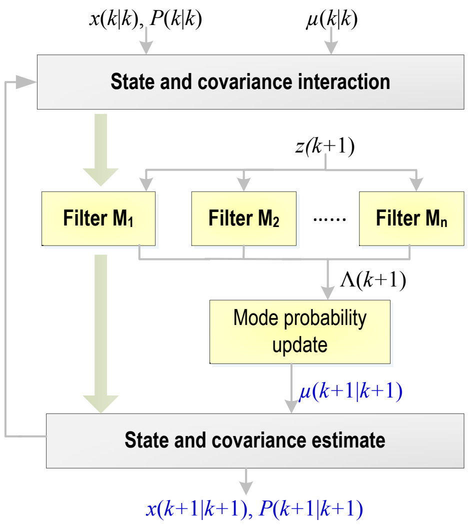

Fig. 16. Illustration of the IMM estimator [119].

Much work has been performed for system estimation and RUL prediction based on KF and its variants. Si et al. [123] attempted to characterize the effect of three sources of variation (temporal, unit-to-unit, and measurement) contributing to the uncertainty of the estimated RUL via a general degradation model. The PDF of the underlying degradation state and random effect parameters were estimated based on KF. Yang [124] proposed a condition-based failure prediction scheme, which links KF to a hybrid Petri-net model coupled with fault-tree analysis for preventive maintenance. Feng et al. [120] applied EKF and the expectation-maximization (EM) algorithm to estimate the deterioration state and the unknown parameters of an age-dependent general nonlinear deterioration model.

Saurabh and Mahendra [121] employed an adaptive UKF to track sudden changes of the stiffness and damping coefficients of structural systems, which can be expanded to track the sudden change of parameters evolved in other dynamic evolution equations. Reuben and Mba [125] presented an approach using the switching KF framework for both diagnosis and prognosis using condition monitoring data under a single framework. It is assumed that the deterioration process evolves over time and multiple filters model the deterioration at different stages. The switching KF would then infer the most probable filter for RUL prediction. Zghal et al. [126] introduced the application of interacting KF on recursively tracking the variations of the model parameters.

2.2.2.2 Particle filter-based model

An alternative to KF under Bayesian inference models and without requiring strict modeling hypotheses such as linearity and Gaussian assumptions, are Particle filters (PF), which is also known as the Monte Carlo (MC) based method. The PF process provides a different approach to estimating the posterior PDF via a set of random samples with associated weights. Similar to KF and other Bayesian inference methods, the PF process contains two steps: 1) prediction: updated posterior PDF of the model parameters at the previous step are used to calculate the system states at the current time through underlying physical models; 2) update: predicted model parameters and system states, (i.e., particles and their weights) are corrected based on the likelihood function combined with condition monitoring data.

Along the evolution history of PF, sequential importance sampling (SIS) forms the basis for other variants. If x denotes the sampled points and w denotes the weights up to time k, the posterior PDF at time k can be represented as:

$$p(\mathrm{x}_{0:k} \mid z_{1:k}) = \sum_{i=1}^{N_s} w_k^i \delta(\mathrm{x}_{0:k} - \mathrm{x}_{0:k}^i), \tag{7}$$

where the weight update is shown in the following relationship:

$$w_k^i \propto w_k^{i-1} \frac{p(\mathrm{z}_k \mid \mathrm{x}_k^i) p(\mathrm{x}_k^i \mid \mathrm{x}_{k-1}^i)}{q(\mathrm{x}_k^i \mid \mathrm{x}_{k-1}^i)}, \tag{8}$$

where $q(\cdot)$ denotes the importance density. The method then recursively updates the point values and weights as each measurement is received sequentially. A common problem with SIS is the degeneracy phenomenon where after a few iterations all but few particles will have negligible weight. This degeneracy implies that a large computational effort is devoted to updating particles whose contribution to the approximation of the posterior PDF is almost zero [127]. The various versions of PF can be regarded as special cases of SIS algorithm. These special cases can be derived from the SIS algorithm by an appropriate choice of importance sampling density and/or with resampling.

A potential solution to the degeneracy problem derives the second representative of PF: sequential importance resampling (SIR). The basic idea is to eliminate particles that have small weights and to concentrate on particles with large weights as shown in Fig 17 (b). After the resampling step, the particles are no longer uniformly generated over the search range, but concentrate on the positions with relatively large possibilities [128]. Besides solving the degeneracy problem, this approach can significantly improve the overall computation efficiency. It is however important to realize that the resampling process can result in many repeated particles: those corresponding to the largest likelihoods. This leads to a loss of diversity among the particles. This problem is much more severe in the case of small process noise such that all particles will collapse to a single point within a few iterations. An illustration describing the degeneracy and sample impoverishment problem is shown in Fig. 17 (b).

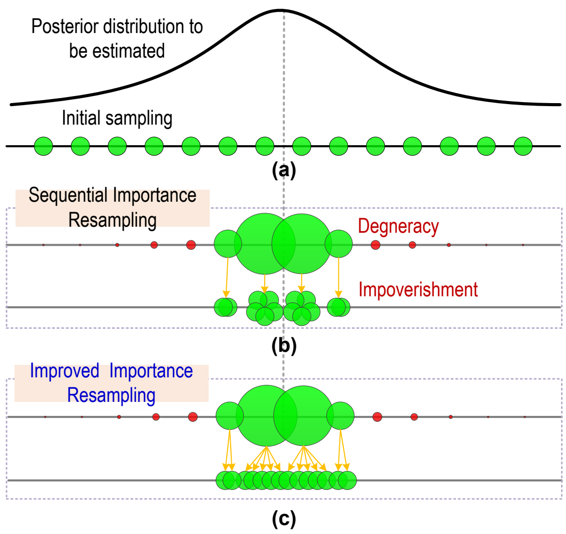

Fig. 17. (a) initial sampling by PF; (b) sequential importance resampling and associated impoverishment problem; (c) improved importance resampling proposed in Ref. [131]

An approach to solving this problem is not to conduct systematic resampling iteratively, but to first decide whether there is a need to apply the resampling procedure through a degeneracy quantification criterion. This introduces another representative variant of PF: regularized particle filtering (RPF). The RPF is identical to the SIR filter, except for the resampling stage. First, it judges the degeneracy degree to determine whether resampling needs to be applied. Second, RPF resamples from a continuous approximation of the posterior PDF rather than a discrete distribution, which reduces the particle collapse problem in SIR.

Besides RPF, different approaches have been proposed recently to improve the resampling strategy. Li et al. [129] described traditional resampling approaches that only consider the weight information disregarding particle spatial distribution information and discarding particles in an uncensored way that reduces diversity. A weight-and-space based resampling method targeting at the described problem is then proposed. Li et al. [130] attempted to combine PF with other methods, such as mean-shift algorithm, artificial intelligence algorithms, and machine learning techniques, to fight the sample degeneracy and impoverishment problem. Wang et al. [131] proposed a local search particle filter, which employs the particles that are intentionally inherited from previous iteration to explore a wide range of prior distributions based on the estimation result from last iteration as depicted in Fig. 17 (c). Hu et al. [138] proposed a particle weight optimization approach to refine the resampling process. Other variants of PF based on SIS do not require resampling, such as MC, Gauss-Hermit, and unscented PF; detailed information can be found in [132].

Compared to KF, which is mature and established, PF is still fast evolving in the field. Wang et al. [131, 133] developed a PF-based framework for precise RUL estimation, which was effective in a prognosis case study on tool wear and RUL prediction; this framework is shown in Fig. 18. Similar work can be found for the fatigue crack dynamic evolution of a mechanical component [134-135]. Sun et al. [128] adopted PF to estimate latent state and parameters jointly for RUL with uncertainty estimation. Wang et al. [136] incorporated regression methods, such as auto regression and support vector regression, into PF for long-term prediction when online measurement is not available.

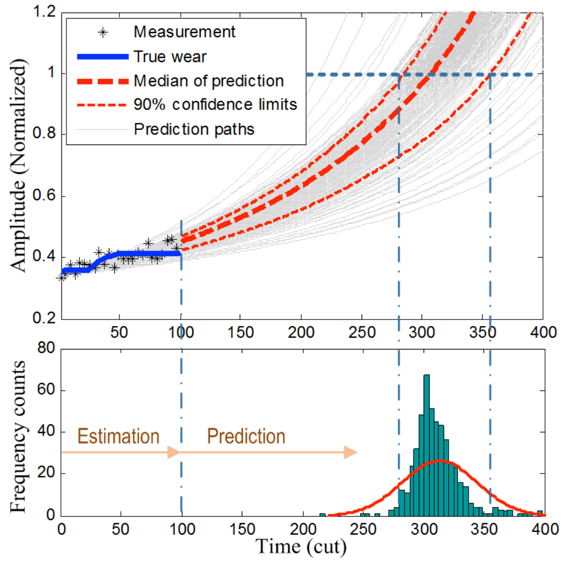

Fig. 18. RUL prediction and uncertainty estimation by PF, adapted from [133]

PF has also been adopted as a preference among Bayesian estimation algorithms. It has also been combined with other methods for its computational efficiency. Liu et al. [137] proposed a multi-step ahead prognosis framework of equipment based on hidden semi-Markov model (HSMM) integrated with the SMC method when the mathematical model or the statistical models of equipment are difficult to obtain. The SMC method was adopted here to decrease the computational and space complexity and describe the probability relationships between multiple health states and measurement. Chen et al. [139] and Baraldi et al. [140] presented an approach based on neuro-fuzzy systems and PF for machine health condition prognosis. The former was employed as a prognostic model to predict the evolution of the machine fault state over time, while PF adopted the predicted data as prior information in combination with online measurements to update the degree of belief in the forecasting estimations. Compare and Zio [141] applied PF to predictive maintenance to identify the optimal time for carrying out the next maintenance action.

2.3. Comparison and evaluation of prognostic models

The strengths and weaknesses of prognosis methods have been summarized in Tables 2 and 3 with a focus on the algorithmic aspect and uncertainty quantification for the prediction results (i.e., predicted failure trend or remaining useful life). Practically, evaluation metrics will be needed both quantitatively and qualitatively for selecting the most appropriate method for a specific application. Saxena et al. [142-143] summarized metrics for prognosis performance evaluation in three groups: algorithm performance, computational performance, and cost benefit. Under algorithm performance, accuracy, robustness, precision, and convergence are included; accuracy quantifies the error between the predicted and true values, and precision describes the convergence of predicted results. Different methods can be applied to quantify accuracy, such as a-λ accuracy (see Fig. 19) and relative accuracy. One indicator of computational performance is the computational complexity, which is especially important for applications where data needs to be monitored in real time to make safety-critical decisions. Other indicators are qualitative in nature, such as robustness within algorithm performance (e.g., practicability of model requirements, sensitivity) and cost benefit.

One major factor that needs to be considered for selecting an appropriate prognostic method is the required information input and assumptions for prognostic models. As described in Section 2.1 and 2.2, both deterministic and probabilistic physics-based prognostic models require good understanding of the physical principles related to machines and the mechanism of fault deterioration. However, the characteristics of and relationships among the various components in a physical system are always too complicated to be modeled effectively [26]. Furthermore, the established models can involve large uncertainty depending on the assumptions used for modeling. For example, even under well-controlled experimental conditions, the rate of tool wear propagation and consequently, parameters involved in Paris' law for identical physical components can be different. A trade-off between prognosis accuracy and computational cost needs to be carefully considered to be practically meaningful and acceptable.

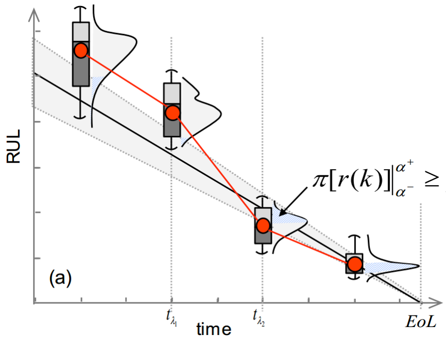

Fig. 19 a-λ accuracy with the accuracy cone shrinking with time on RUL vs. time plot [142]

Another factor for evaluating the prognostic methods is the quantification of uncertainty involved in the prognosis process [144]. The source of this uncertainty can be classified as [31]:

- Modeling error: The failure model that degradation follows should be first determined for prognosis. Various failure models have different triggers to initiate failure and to model failure propagation [145]. Given the complexity of real world scenarios, no one model completely describes the actual degradation. This means that uncertainty in physics-based prognosis approaches comes from assumptions and simplifications of model structures. Incomplete coverage of data for training empirical models introduces additional uncertainty in data-driven approaches [146].
- Data quality: The selection of condition monitoring features can directly determine the performance of a prognosis system and affect [147-148], the nonlinear relationship between features and actual machine health and the sensitivity of features to operating conditions [26].
- Randomness in future degradation of equipment: Novel events, such as changes in operational conditions, maintenance actions, and new failure occurrence, may change the deterioration and progression modes of existing failures [149-150].

Most of the existing prognostic techniques predefine a threshold for the feature to estimate the RUL by assuming the failure takes place at the instant in time when the increased or decreased feature reaches the predetermined threshold. Practical applications of prognosis systems may commonly yield false-negative and false-positive alarms under the effect of uncertainties discussed above. This problem, caused by an insufficient understanding of prognosis, highlights a future direction for research.

3. Cloud-enabled manufacturing and prognosis

Cloud computing (CC) brings new opportunity in accelerating the acceptance of advanced manufacturing technologies. Prognosis, as an integral component of manufacturing, can benefit significantly from CC.

3.1. Characteristics and service orientation of cloud

The "cloud" refers to the Internet as a communication network, for distributed storage and delivery of computational services. Various assets, such as sensor networks, embedded systems, RFID, and GPS, are integrated in cloud manufacturing where manufacturing resources (machines, robots, etc.) can be sensed intelligently and connected to the Internet, as well as monitored, controlled, and managed remotely. This creates the Internet of Things (IoT), which is essential to cloud manufacturing.

The development of CM can be seen as a progression from the sole adoption of CC facilities and functions to the adoption of all manufacturing resources as services realizing the manufacturing version of CC [151-153]. As within CC, procedures involved in the entire manufacturing process, initially from product/machining parameters design to equipment maintenance can benefit from cloud-based services.

Condition monitoring, remote data analysis, degradation/fault root-cause diagnosis, and prognosis all provide supporting information for maintenance decision making as described in Fig. 4. However, massive data analysis is involved in these processes, which requires significant computing resources to perform on-line real-time computation. CC techniques can make these tasks more efficient by leveraging infrastructure-oriented services in the cloud for data storage and analysis, while software-oriented service can be performed in a distributive fashion as web-based programs for interface with manufacturers and consumers.

Small and medium-sized businesses (SMEs) can especially benefit from cloud services since these services provide the ability to use applications and solutions that may be too complex or expensive or designed for use by larger enterprises. The pay-as-you-go model with low cost for usage and maintenance eliminates economic barriers, such as extensive investments in IT systems and manufacturing equipment that are infrequently used and rapidly depreciating. CM also provides other features, such as the best mix and match of resources no matter where they may be physical located, which helps realize concepts, such as DAMA (Design Anywhere, Manufacture Anywhere)[154-155].

A variety of descriptions and definitions of CM exist that have evolved from different perspectives. Many of these definitions emphasize manufacturing services and resource sharing as typical properties of CM [40, 151-152, 156]. They describe collaboration by network-based resources and capability sharing in the form of services between different cloud users (consumers, providers, and operators). In light of IoT through machine connectivity, "cloud manufacturing is an integrated cyber-physical system that can provide on-demand manufacturing services both digitally and physically to best utilize manufacturing resources" [157-158].

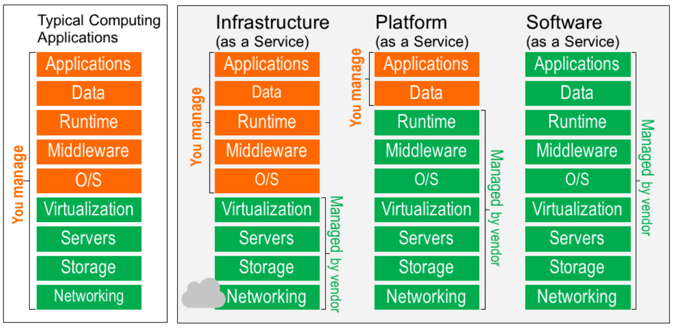

Fig. 20. Comparison of service level and user involvement.

The inclusion of CC as a core enabling technology is one of the major differences between CM and other advanced networked manufacturing systems. It makes it possible to provision manufacturing activities as services in a distributed environment and towards service orientation. A cloud service is differentiated from traditional hosting by three distinct characteristics: it is sold on demand (typically by usage time or subscription); it is elastic (a user can decide how much of a service she wants at any given time); and the provider fully manages the service (the user only needs a computer and access to the Internet). The most common services of CC are defined as: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS) [159]. Fig. 20 compares the services against typical standalone computing applications at service level.

Table 4. Characteristics and benefits of cloud manufacturing

Diversity	All activities in the product lifecycle can be encapsulated as services and be provided to customers as needed.
Dynamic	Services can be changed (add new services or delete obsolete services) according to users' requirements.
Virtualization	Physical resources are virtualized and services are based on these virtual resources.
Elastic	Services can be scaled up and down quickly based on users' demand.
Broad access	Services and available resources are accessible from anywhere at any time.
Fault tolerant	Degree of replication including geo-replication provides fault tolerance with high service availability.
Cost effective	Pay only for what is consumed. Economies of scale allow cost reduction.

Similarly, manufacturing resources and capabilities along the entire product development lifecycle can be realized and offered in the cloud as IaaS, PaaS or SaaS, e.g., Design-as-a-Service (DaaS), Assembly-as-a-Service (AaaS), Monitoring-as-a-Service (MaaS) or Testing-as-a-Service (TaaS). Compared to CC that mainly deals with IT infrastructures and software, implementing and realizing CM is much more demanding since it includes an array of physical manufacturing equipment to be accessible via the cloud. Table 4 summarizes the main characteristics and key benefits of cloud manufacturing from the end users' perspective.

3.2. Supporting technologies, architecture and platforms

The supporting technologies to implement CM or cloud-enabled prognosis successfully include:

- Internet of Things (IoT) – The IoT integrates and connects physical assets (e.g. machines, sensors) into an information network, which enables device interoperability and universal manufacturing resource availability and accessibility [160-161]. The IoT is quickly growing with RFID [151] and sensor technologies, which promotes interconnection between *things*.
- Embedded Systems Technology – The rapid development of embedded systems technology with the IoT enables convenient access to manufacturing resources for status retrieval and control [162].
- Semantic Web – The semantic web facilitates knowledge-based intelligent computation and enables users to search and share data and information easily by allowing data from different sources to be processed directly by machines [163-164]. It provides a common framework for data to be represented and reused across applications and promotes the use of different common formats for data exchange.
- MTConnect – An open and non-proprietary communication standard for machine-to-machine communications and interoperability between existing technologies [165].

Many attempts have been made and reported in the literature to define CM system architecture [39-40, 151, 156, 166-171]. Some proposed architectures have 3-4 layers, while more detailed architectures have up to 12 layers. The naming and content of these layers also differ between architectures. Fig. 21 presents a typical conceptual CM architecture that consists of three layers [168]:

- Manufacturing capability layer: This layer contains the core manufacturing services (computer-aided process planning or CAPP, computer-aided manufacturing or CAM, computer-numerical control or CNC, etc.) in a service application cloud. The services and user data can be safely stored in a storage cloud. Physical manufacturing resources are connected to this layer for on-demand access and service realization.
- Virtual service layer: A central server is placed in this layer for cloud management. Virtual services are matched and mapped to the real services and physical resources based on their availability and capability.
- Application layer: This layer mainly concerns the end users (business users and private users) of the cloud services. Comprehensive user interfaces and convenient access to the cloud is the key. User friendliness, thin-client user interface design and timely information presentation are dealt with at this layer.

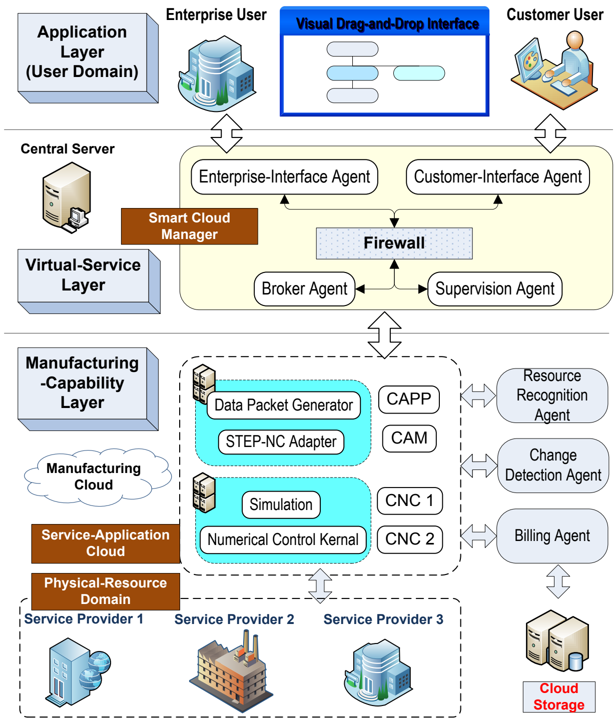

Fig. 21 A 3-layer CM architecture (adapted from [168]).

Despite the difference in architectures, there is an agreement that a CM system has three types of participants:

- Consumer (or end user): Purchases and consumes available manufacturing services in the cloud from providers after supplying engineering requirements to the cloud operator.
- Provider: Provides and sells manufacturing resources and capabilities as services to consumers for their product development. Services supporting the whole lifecycle of the manufacturing processes can be provided.
- Operator: Responsible for the operation and management of a CM system as well as for finding, combining, controlling and coordinating the required services for fulfilling consumer requirements.

The interest of the cloud manufacturing concept and its potential effects is rapidly increasing. Many research initiatives are active among academic and industrial participants in local, national, and international projects of varying size and scope. Some of these initiatives and platforms are summarized below:

- CMfg

CMfg represents an application model of CM that describes CM platform activities in the propagation from user requests to the return of a solution [40, 167]. It utilizes a 5-layer CM architecture design: physical layer, virtualized resource layer, service layer, application layer, and user layer. To demonstrate the feasibility of the CMfg concept, a cloud-based application called Cloud Simulation based on the Cloud Simulation Platform (COSIM-CSP) has been developed where the collaborative work of a virtual flight vehicle prototype is simulated [156] . Related platforms from the same research initiative include: SME-oriented Cloud Manufacturing Service Platform (SME-CMfgSP) with a 12-layer architecture [170] and MfgCloud [172].

- Interoperable Cloud-based Manufacturing System (ICMS)

ICMS [168, 173] adopts a three-layer architecture as shown in Fig. 21. The three layers are associated with the manufacturing cloud at manufacturing capability layer, user cloud at application layer, and smart cloud manager at virtual service layer. A distinction is made between customer users (CUs) and enterprise user (EUs): CUs are defined as customers requesting a self-contained production task, while EUs are organizations or enterprises seeking additional capabilities and support to fulfill bigger and more demanding production tasks in collaboration with temporary partners and their services.

- Cloud-based Design and Manufacturing (CBDM)

CBDM is a prototype system that builds on the concepts of CC with manufacturing resources being available as different services [174]. For the implementation of CBDM, a Distributed Infrastructure with Centralized Interfacing System (DICIS) model was proposed. The centralized interfacing system enables the system to function as a whole. Workflow for distributed and collaborative design and manufacturing in a local and distant user scenario is described, where engineers are able to simultaneously cooperate with each other using computer-aided design (CAD) software in a SaaS mode.

- ManuCloud

ManuCloud is a project funded by the European Commission (EC) [175-176]. The objective of ManuCloud is the development of a service-oriented IT environment to support the transition from mass production to personalized, customer-oriented, and eco-efficient manufacturing. A conceptual architecture with a front-end system and MaaS infrastructure to support cloud-based manufacturing of customized products was proposed. Using Manufacturing Service Descriptions (MSDs), the concept has been proven in some business cases, one with distributed production and customer specification of small series, high-value products.

- CAPP-4-SMEs

CAPP-4-SMEs is another EC-funded project on Collaborative and Adaptive Process Planning [177]. The prototype system extends the distributed process-planning (DPP) [178] concept and a two-stage generic-specific planning method [179]. For better connectivity between the cloud and machines, function blocks [180] are used to bring a process plan to a chosen machine for execution while monitoring its status at the same time. Since machine availability and capability can be monitored in real time through the cloud, resource virtualization has been eliminated.

When implementing a CM system, security is a major concern. Corporate information often contains sensitive data about customers, employees, trade secrets, and intellectual property [181-183]. Securing sensitive data and the ubiquitous availability of requested applications in the cloud are a must for potential users of cloud services. Manifestations of these concerns regularly appear in many existing CC services as a profound unwillingness and anxiety in letting sensitive and important data escape outside the boundaries of the physical company premises [184]. The service models (IasS, PaaS, and SaaS) require different levels of security in a cloud environment. IaaS is the base of all CC and CM services, with PaaS built upon it and SaaS in turn built upon PaaS. Just as capabilities are inherited, so are the cloud security issues and risks [151]. Today, most SaaS business and manufacturing applications that vendors offer are hosted in ISO 27001 and Statement on Auditing (SAS) 70 Type II certified data-centers with service-level agreements offered for applications of 99% and above [185]. More information about CC and CM can be found in [184, 186-189].

3.3. Crowdsourcing for effective prognosis services

Crowdsourcing refers to the outsourcing of functions performed by actors within an organization to an undefined network of people outside of the organization [190-191]. It is used primarily for complex or expensive problems that may be solved efficiently by dividing tasks between different participants who each have a different set of resources or expertise. Crowdsourcing is also used typically to solicit ideas from a large network to leverage multiple perspectives for innovation. These capabilities address an important goal of cloud manufacturing to use a shared pool of resources or knowledge to enable collaboration across all participants in the cloud [40, 151-152, 156, 167-168, 192-193].

Figure 22 highlights some of the potential CM applications enabled by crowdsourcing. Crowdsourcing for CM has typically been suggested to connect service providers to customers for sourcing, resource allocation, and scheduling support [151, 156, 167-168, 190, 192-193]. Similarly, crowdsourcing in conjunction with the supporting technologies described in Section 3.2 (e.g., IoT, embedded systems) may be used to share the computational effort needed for different prognostic models and approaches. It is also especially useful for manufacturing prognosis since knowledge (and by extension the information and data used to generate it) may be shared in the cloud [194]. By leveraging this collective intelligence, crowdsourcing allows for the sharing of experiences that helps identify and implement best practices. These advantages are particularly beneficial in prognosis where vast knowledge holistically collected across a manufacturing system is needed to reach robust, reliable decisions, especially for SMEs who often lack such resources.

One specific application of crowdsourcing for prognosis is in improving data collection and synthesis efforts. For example, there is a need to better correlate machine condition with process and inspection data to provide the context needed to differentiate between process and machine degradation [195]. The difficulty is that diagnostic and prognostic models generally require significant amounts of historical condition monitoring and event data [195-197]. As this data becomes more extensive, the uncertainty of these models decreases [195]. Crowdsourcing can generate extensive, representative historical condition monitoring and event data sets by synthesizing smaller data sets available in the cloud. The appropriate means to synthesize data in this way remains an open research question. Furthermore, there may be intellectual property (IP) concerns involving the ownership of any collectively generated knowledge, information, or data [193].

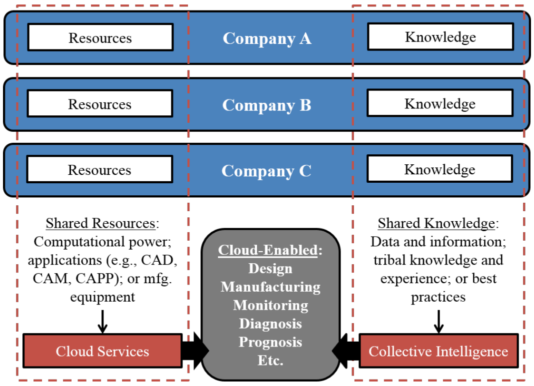

Fig. 22 Potential applications of crowdsourcing in the context of cloud manufacturing.

There are other challenges when implementing crowdsourcing for prognosis. Cybersecurity is perhaps the most significant challenge and refers to protecting intellectual property (IP), sensitive information (e.g., customer information or export control) and the security of devices and assets networked in the IoT [198]. Manufacturing has focused typically on protecting IP and sensitive information [151, 156, 198], which should be the more immediate concern when using crowdsourcing for prognosis since crowdsourcing requires sharing knowledge, information, and data. (The security of networked devices is a more general concern of cloud manufacturing that is discussed in Section 5.) Section 3.2 describes different options to address IP concerns (e.g., ISO27001 and SAS 70 Type II certified data centers). Other options include firewalls, compression, encryption, and virtual local-area networks (LANs) [151]. Test encryption schemas, scheduled backup, service-level agreements, and quantitative analysis of vulnerabilities are all important in ensuring the reliability of many of these protection options [151, 190, 199]. Furthermore, the extent to which the cloud is integrated into an organization's operations can also address IP concerns. For example, Tao et al. [156] suggest using a different cloud manufacturing platform – public, private, community, or hybrid – depending on the types of services and interaction that are desired by the cloud participants.

An alternative approach to address IP concerns is to focus on the knowledge, information, and data being shared in the cloud. Anonymizing sensitive information not needed for useful analysis would help encourage the further development of crowdsourcing for prognosis [151, 192]. Otherwise, the fear is that the data could result in negative publicity and/or inform key competitors of an organization's operations. A major challenge of anonymization efforts, though, is determining what aspects of knowledge, information, or data are needed by the cloud so that crowdsourcing provides positive benefit. In addition, the secure elimination of data is not trivial.

Interoperability is another significant challenge for crowdsourcing [19, 168]. This is especially true in diagnosis and prognosis where sensor fusion is important in providing a holistic perspective of the state of a manufacturing system [196]. Data and information flows from manufacturing systems can appear in a variety of formats, at a variety of sampling rates, and from a variety of sources (see Fig. 23) [19, 200]. There are few standard interfaces and rules that allow for seamless communication of this data and information. This can make aggregating information flows from several data streams that are appropriately contextualized with the manufacturing system an extremely complicated process. Data interoperability standards that bridge multiple levels of the manufacturing hierarchy (i.e., process to enterprise), data and information types, equipment types, and product life cycle stages are critical enablers for data and information synthesis in the cloud [196, 200]. MTConnect is an example standard that can enable crowdsourcing for prognosis. Vogl et al. [197] provide a good summary of the standards landscape for manufacturing prognosis, especially related to health management.

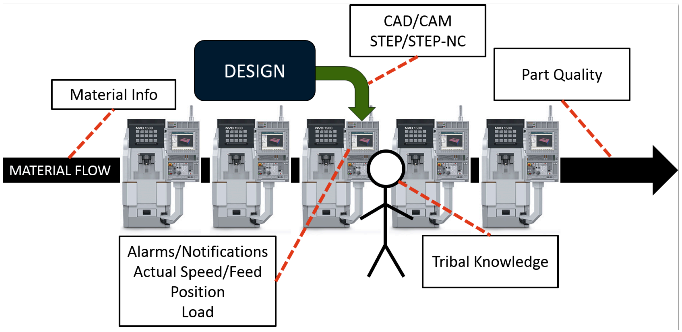

Fig. 23. Potential sources of data on the shop floor.

4. Application Highlights

4.1. Prognosis of machine components, tools, and processes

Tool wear prognosis and remaining tool life prediction are the two most important issues and common applications in manufacturing oriented condition monitoring [201-202]. As described in Section 2.1.1, an important technique to describe tool wear or tool life is physics-based models, namely deterministic equations. Two types of parameters are included in these physical models: measurement parameters (e.g., cutting speed in Taylor's equation) and unknown coefficients (e.g., exponential coefficient in Taylor's equation); the latter is usually determined through a quantity of experiments. For example, Fig. 24 demonstrates a striking difference in wear pattern on the rake face of PCD tools at two different cutting speeds [203], which proves the reverse relationship between cutting speed and tool life as described in Taylor's equation.

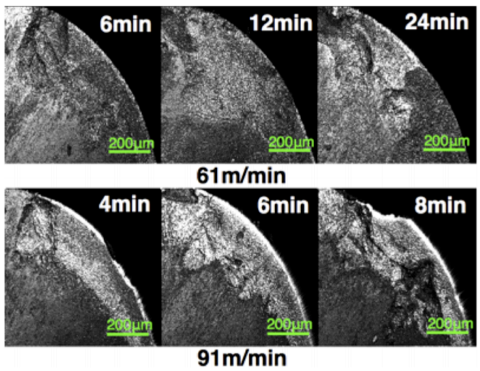

Fig. 24 Tool wear at different cutting speed [203]

There is inherent uncertainty in the empirical constants since they vary with factors, such as tool material, workpiece material, and manufacturing conditions. Hence, tool life is generally considered to be a stochastic process, and the types of estimate methods introduced in Section 2 often estimate unknown parameters. Karandikar et al. [204] applied Bayesian inference to estimate parameters in a Taylor tool life model for milling operations. They extended their work to include cutting feed as well as cutting speed into the Taylor's equation for turning operations and applied the metropolis-Hastings algorithm of the Markov chain Monte Carlo to estimate the parameters in the model [205]. Similar research can be found in [136, 206]. Common approaches for performance degradation assessment and RUL prediction for other machine components have been introduced in Section 2, such as fuzzy theory [87], support vector machine [207], Bayesian inference [139], and particle filters [88].

4.2. Remote monitoring, diagnosis, and prognosis

The IoT essential for cloud manufacturing allows for the collection of real-time information and data from devices and assets networked across the manufacturing enterprise. This information and data provides opportunities for remote monitoring, diagnosis, and prognosis by providing a holistic perspective of the historical and current state of manufacturing equipment and processes on the shop floor. But, the potentially large variety of data types and formats as well as the potentially large amount of data available through the IoT necessitates effective solutions to collect data in context with process-related information to identify physical reasons driving any observed variations in the manufacturing system. This can be accomplished by using open standards for manufacturing data interoperability when collecting sensor and process data.

MTConnect supports data integration (not transmission or use) and is defined to emulate the hierarchy of a conventional machine tool. It enables plug-in architectures that allow for application-focused development by standardizing interfaces between various data sources. It is also extensible and lightweight and supports both legacy and new technologies. To further ease factory integration, subsequent research has focused on incorporating MTConnect with other standards for sensor networks (e.g., IEEE 1451) to address data transmission and use [208].

Figure 25 shows a typical MTConnect architecture deployed on the shop floor for remote monitoring. Using this architecture, data from the machine tool controller (e.g., actual position, actual and commanded feed, actual and commanded speed) and a variety of sensors (e.g., power meters, thermocouples, accelerometers) can be combined for many different remote monitoring, diagnosis, and prognosis applications, such as preventive maintenance [192], process planning verification [200], accurate cycle time estimation, tool position verification [209], and energy monitoring [210] and prediction [211].

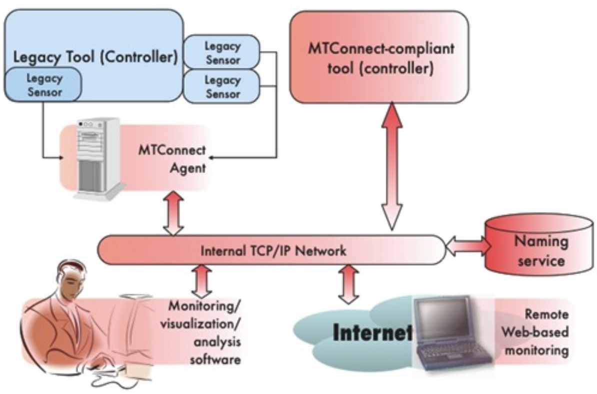

Fig. 25 MTConnect architecture highlighting application for remote, web-based monitoring [200].

Applications of remote monitoring, diagnosis, and prognosis that utilize an infrastructure provided in the cloud can also be found deployed in industry. Teti et al. [19] provided an extensive list of different industrial efforts for remote monitoring and diagnosis. DMG Mori Seiki offers MORI NET for remote monitoring of machine tool status through the internet, including remote alarm support, transmission of alarm information, machine operating status check, and periodic reports [212]. Similarly, Mazak provides Cyber Production Center, which monitors machine operation and job status, tooling data, and machine work load via the internet [213]. Both of these solutions are integrated within specific machine tool systems. Alternatively, third party vendors are offering general software tools, such as VIMANA Core from System Insights, which monitors real-time shop floor data to help organizations manage the productivity of their machine tools [214].

4.3. Event-driven, condition-based maintenance

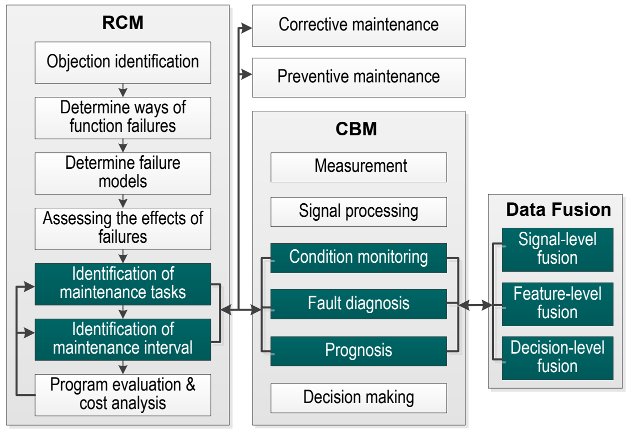

Fig. 26 Condition-based maintenance based on data fusion and reliability-centered maintenance, introduced in [215]

The most important two attributes for maintenance are cost-effectiveness and accuracy, which is a comprehensive factor that includes reliability and probability. As described in Section 1, a significant advantage of prognosis-enabled CBM or preventive maintenance over traditional scheduled maintenance is its effectiveness in reducing maintenance cost. Niu et al. [215] proposed a condition-based maintenance system that employed a reliability-centered maintenance mechanism to optimize maintenance cost, as shown in Fig. 26. However, preventive maintenance requires a better understanding of the nature of maintenance policies in a mathematical way and incorporation of diagnosis and prognosis results into maintenance rules (i.e., the adaption of maintenance polices). The overall objective of formulating or selecting maintenance rules is to minimize the total maintenance cost, including the hidden cost of risk and reliability.

For preventive maintenance, the maintenance decision rules should be incorporated with the information obtained from online measurement, data processing (diagnosis and prognosis), or data fusion, which makes sense especially when equipment works in a complex situation and undergoes a different deterioration rate. An adaptive maintenance model for a gradually deteriorating system is proposed in [216] where the deterioration rate is considered a time-dependent function. The maintenance rules should be adapted after the change points where the transitions of mode of system deterioration are assumed to occur. Grall et al. [217] proposed a maintenance strategy with sequential inspection times taking into account the current system state for the choice of the next inspection, as shown in Fig. 27. The system deterioration is modeled as a Gamma process, and the system is considered failed if its condition jumps above a pre-set failure level. They extended this method to a two-unit series system to obtain an optimized inspection/replacement strategy by minimizing the long-run maintenance cost of the system [218]. To avoid unnecessary investment in CBM equipment, a reliability-centered maintenance enabled CBM model is proposed in [215] by taking the functions of components and their importance into consideration to maximize results with regard to system reliability and outage cost reduction. Also, data fusion techniques at the signal, feature, and decision levels are applied to increase maintenance accuracy.

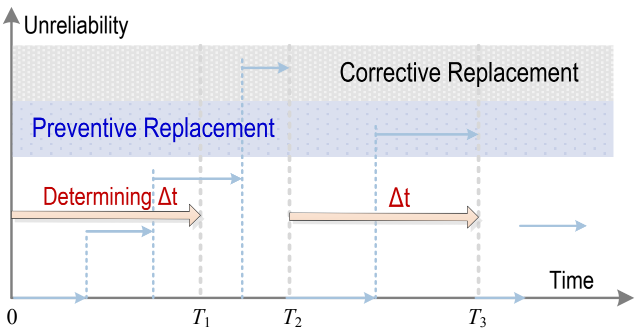

Fig. 27 CBM-based decision making of inspection time adapted from [217]

4.4. Integration of prognosis with cloud for adaptive decision-making

Prognostic models developed based on the deterministic and probabilistic methods described in Section 2 are often application specific and vary with different manufacturing operation conditions, maintenance polices, or environmental conditions. To achieve cloud-enabled prognosis, crowdsourcing is employed as described in Section 3.3 to fully utilize not only parallel computing capability but also the shared information obtained from previous condition monitoring or maintenance experience. To fulfill crowdsourcing, a knowledge sharing method called case-based reasoning (CBR) can be applied. The CBR system continuously adds past experience as cases to the case base. The processes to perform problem solving using CBR involve retrieving similar cases by comparing measurement or post-processed data to cases in the case base, reusing the information in the retrieved cases, revising the suggested solution according to specific conditions in target domain, and retaining a new experience to the case base, as described in Fig. 28 [219].

A new CBR system has been developed in [220] to model infrastructure deterioration satisfying special requirements and to provide government agencies with practical, accurate, and versatile deterioration models. An intelligent CBR with petri-nets-based system has been proposed in [221] for machine fault diagnosis; the petri nets improved the retrieval, addition, deletion, and revision of previous experience-based cases in CBR. Similar studies can be found in [222] – a recommended fault type generated by a classification tool for a specific sensor signal, where the recommendation is based on previously classified cases in a case library. A problem-oriented multi-agent-based e-service system (POMAESS) has been developed in [223] for remote maintenance decision making. The decision-making process in POMAESS seeks technical support from informational exchanges between customers and suppliers as well as cooperation and negotiation based on the sharing of different complementary and contradictory knowledge.

Two problems that need to be addressed when applying CBR to condition monitoring, prognosis or maintenance are efficiency decrease due to incremental cases and feature matching between sensor data and existing cases. As reasoning improves with time, the retrieval speed and efficiency decreases., so case management is important to the overall performance of a CBR system. An adaptation-guided case-base maintenance strategy is proposed to compact a case base by dynamically generating new adaption knowledge to fuse cases [224]. It is inevitable that there is inconsistency between features extracted from new measurements and existing information in the case base, which requires a CBR system to have versatility and extensibility of case and knowledge representation; and fuzziness of retrieval knowledge [220]. To solve this problem, rough set theory is applied to reducing the effect of assigning weights to the case feature attributes [225].

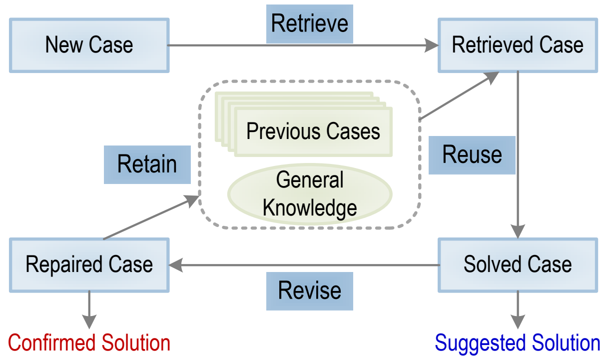

Fig. 28. Architecture of CBR, adapted from [219]

5. Challenges and Limitations

5.1. Bandwidth, speed, distributed storage, and computation

One of the most important characteristics of data processing in manufacturing is that it is real time with three issues involved in this process: data transmission, data storage, and data analysis. Real-time data measured on shop floor are expected to be transmitted to a cloud server over the Internet in a timely fashion, followed by effective data analysis and transmission of the result back to the machine site for operation/process control and/or maintenance. Unlike traditional architectures, a cloud server is an aggregation of distributed computing resources, which may split data files uploaded from clients into several portions to be stored in distributed servers. This poses a challenge for data consistency.

Sensors (e.g., force, vibration) monitoring manufacturing processes that work at high sampling rates can generate a large amount of data within a short time period. The specific application requires high quality cloud service, especially with respect to network and computational performance. Network performance in the environment of CC is determined and affected by the input/output (I/O) virtualization – resources of network links and bandwidth are shared by multiple virtual machines (VMs). Recent research has indicated that the most important issue affecting I/O virtualization performance is communication between VMs and virtual machine monitor (VMM), which is responsible for assigning storage/computational resources and managing transmitted data. When a data packet arrives, the server must determine the VM to which the file should be delivered by analyzing its header and employed protocol and referencing a bridging table. Barham et al. [226] pointed out that 30-40% of execution time for a network to transmit or receive an operation was spent in VMM to remap addresses contained in the transmitted data package. This problem worsens in the context of increased line rate and workloads. It has been demonstrated that the overhead of central processing units (CPUs) and latency increase with the transmitted package rate due to increased communication between the server (VMM) and client (VMs) domains [227].

Various studies and experiments have been conducted on performance measurement and analysis of network I/O applications. For I/O virtualization performed on VMs instantiated on a single physical machine, VM throughput is almost ten times lower than the throughput in the server domain likely because of the costly communication between the server domain and VMs in the client domain [228]. Especially when dealing with small packets but high packet rate, the throughput is even lower since the software stack does not have enough CPU resources to process [229]. For multiple VMs running on multicore platforms, experimental results indicated that the latency for write/read operation does not change much as the number of VMs increases, but does increase exponentially with the package size due to bandwidth saturation. It is considered that, as the link becomes saturated with increased package size, the average bandwidth attained by each VM decreases [230].

The data storage type in the context of CC is known as a distributed file system; Hadoop distributed file system (HDFS) [231] and Google file system (GFS) [232] are representative examples. A significant characteristic of both systems is the partitioning of data and computation across many servers. HDPF stores the file system metadata in NameNode and application data in DataNode separately; NameNode represents files and directories, and record attributes (e.g. permissions, modifications, access times) [231]. An important challenge associated with this kind of file system is the failover of NameNode/hypervisor, which keeps all the namespace and data file locations in memory. The limited size of memory can restrict the capability of NameNode, and easily cause its shutdown. Another challenge in the application scenario of real-time, large-volume data storage and analysis is inefficient execution due to multiple replications of data files because of the need for data locality in maintaining performance.

Similar to the effect of I/O virtualization on network performance, computational capability in CC is affected by the sharing of computing resources, such as CPU, memory, and cache line. When executing multi-tasks based on scientific computing, experiments have indicated that the floating point and double-precision float operations are six to eight times slower than the theoretical maximum, due to the overrun or thrashing of the memory caches [233]. Similar work is found in [234-235]. In [234], several benchmarks have been executed on both native and virtualized systems, and the results showed that the performance overhead for CPU and memory virtualization were up to 5% and 40%, respectively, mainly caused by memory allocation. Kousiouris et al. [235] investigated the effect of a number of critical parameters, such as real-time scheduling decisions, and co-placement of VMs, on the performance of VMs. The performance overhead posed by these parameters has been confirmed to be up to 150%.

An important issue determining the virtualization performance and consequent network and computation performance is dynamic resource management. The most popular option for resource allocation among current cloud-oriented services is to seek trade-off execution quality by the assigned resources via a load balancing mechanism or high availability mechanism. Relevant research has been conducted, such as game theory [236] or k-means clustering [237]. These efforts only address the scaling problem of one resource or a single tier.

5.2. Autonomous communication, security

An important goal of the IoT in cloud manufacturing is to leverage machine-to-machine (M2M) communication to collect and contextualize data from sources across the manufacturing enterprise [200, 208]. Data analytics may then be used to assess the data and generate information to support different goals, such as prognosis [200]. The difficulty is that M2M communication in a manufacturing environment can be challenging due to considerations related to interoperability and cybersecurity.

Section 3.3 highlights significant data integration issues in M2M communication, but these issues are only one set of interoperability considerations in manufacturing. Data collection is also a significant challenge because manufacturing equipment is usually old and low in computational power. Many facilities also use a variety of machine-tool types and each may require an interface to communicate with other machines (see Fig. 29) [238]. Every networked device relies on one of several communications protocols (e.g., Modbus, Fieldbus, or Profibus). These interfaces and protocols can grow rapidly without the appropriate standards that allow for "out-of-the-box" communication [208, 238]. The lack of commonly adopted interfaces and protocols increases the knowledge and resources needed for implementation, which can be substantial given the significant training and setup time required even if expertise is available [20, 238-239].

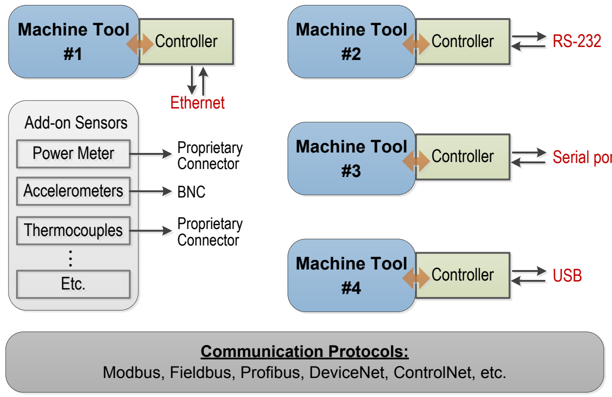

Fig.29 Example of the different communications interfaces and protocols that can be found in a typical manufacturing facility.

The architecture needed for M2M communication must also enable data and information exchange within one and across several levels of the manufacturing hierarchy (i.e., process to enterprise) [210]. It should be scalable for large data volumes and capable of dealing with the time scales (microseconds to days) present in manufacturing data and decision making. While these characteristics increase the complexity of data collection and analysis, they enable automated monitoring that can support autonomous manufacturing systems where machines identify patterns or disturbances using a cumulated set of knowledge and experiences. These machines can then work with other machines to respond to the disturbance and ensure the continued performance of the manufacturing system.

Cybersecurity remains a significant concern hindering cloud manufacturing applications and services. Section 3.3 highlights some of the issues related to the protection of IP and sensitive information, but the threat to the security of networked devices and assets may be the more important concern for cloud manufacturing [198]. Existing infrastructure, such as supervisory control and data acquisition (SCADA) networks, can be a significant vulnerability given its designed function. Stuxnet is one example of a cyberattack that exploited SCADA networks. Developed to target Iranian efforts to enrich uranium, Stuxnet exploits the SCADA's dynamic-link library (DLL), through which SCADA receives information about the system being controlled [240]. Through the DLL, Stuxnet reprograms the programmable logic controller (PLC) so that the system (i.e., Iranian enrichment reactors) operates as the attacker intends.

Through M2M communication, opportunities exist to target any part of the product lifecycle and supply chain if these machines are connected through the cloud [198]. Potential attacks could include altering design files, toolpaths, or quality control. Furthermore, the safety of operators and consumers may also be threatened if an attacker can control these systems. Ultimately, the risks must first be understood and acknowledged so that technologies can be developed to address them. Research is ongoing in this area, such as, Thames [241] who has proposed a distributed, collaborative, and automated infrastructure for the cybersecurity of cloud manufacturing systems. This infrastructure is designed to provide real-time, dynamic, and preemptive protection by using cyberthreat information exchanges, such as the Trusted Automated Exchange of Indicator Information (TAXII) [242], Structured Threat Information Expression (STIX) [243], and CYBEX [244]. Standards are also being developed to support these information exchanges [241].

6. Conclusions

6.1. Summary of state of technology

Cloud-enabled prognosis can leverage advanced manufacturing by using data and information from across the manufacturing hierarchy to improve efficiency, productivity, and profitability. Recent advances in cloud manufacturing have increased the accessibility to many technologies, such as M2M communications, IoT, and semantic web, and now provide an opportunity to transfer prognosis models and techniques from the research lab to industry. Much of the current technological development has focused on providing the infrastructure and architecture to implement prognosis models and techniques. For example, a variety of cloud initiatives and platforms have been suggested to offer different services (e.g., IaaS, PaaS, or SaaS) to manufacturers, and interoperability standards have been proposed for data integration, such as MTConnect. Hardware and software vendors have also started to provide cloud-enabled diagnosis and prognosis solutions to their customers, such as remote monitoring and diagnosis of machine tools and shop floor equipment.

The key challenges for cloud-enabled prognosis will be in data collection and management. Standards will be needed for data interfaces, collection, transmission, and interoperability. Methods to anonymize and remove sensitive information from data and to synthesize data streams from multiple and varied sources will be critical in dealing with the large data volume that may be collected from across the manufacturing hierarchy. Finally, cybersecurity must protect IP sensitive information and the security of networked devices and assets to deploy much of this technology in industry. If these issues are resolved, the potential exists to exploit many aspects of the cloud, such as crowdsourcing, to improve manufacturing efficiently and effectively by providing knowledge and value to actors throughout the product lifecycle, which would drive innovation beyond manufacturing.

6.2. Continued evolution, future research direction

Currently, most of the research activities related to prognosis are confined within controlled laboratory conditions. This is largely due to the fact that prognosis models are application specific. For example, the parameters involved in the Paris' formula for tool wear prediction vary with the type of tools used. Crowdsourcing, if integrated with cloud-based techniques, presents an opportunity for prognosis in an industrial setting. A challenge, as well as an opportunity, in crowdsourcing is the feasibility and interoperability of data for the purpose of fusion given the variety of data (e.g., condition monitoring data and features). Establishing guidelines for designing a prognosis system in a cloud environment, including sensor selection, data transmission, database creation, prognostic method selection, and cooperative, and intelligent decisions, would have a significant impact on advancing the state of prognosis in the context of cloud.

As with cloud-enabled prognosis and its advanced computing capability, dynamic resource allocation can be another research direction, especially in the context of big data analysis. Typically sensor outputs, after preprocessing by local agents, are transmitted to computing resources in the cloud. Challenges and opportunities lie in how to allocate efficiently these data and fuse analysis results to ensure remote yet on-line and real-time manufacturing process and equipment monitoring and prognosis. Also of interest is effective and efficient M2M communication, including data collection, sharing, and transmission, to minimize the bottleneck of current cloud-based techniques and maximize cloud resource utilization.

Cloud-enabled prognosis benefits from both advanced computing capability and information sharing for intelligent decision making. Cloud-enabled prognosis, as well as cloud-enabled design and manufacturing services allocation, is part of cloud manufacturing, which requires an association of distributed industries or manufacturing service providers for information and resource sharing. Significant challenges exist in the creation of mechanisms or standards for information and resource sharing, to maximize the benefit and minimize the potential hazards for industries. Another challenge is effective communication between clients and encapsulated service providers. It is expected that specific service requirements can be intelligently and automatically assigned to one or several industries associated with the cloud with minimal human intervention. Overcoming these challenges will make cloud-enabled prognosis an effective tool for the widespread adoption of cloud-based manufacturing.

Acknowledgment

The authors greatly appreciate contribution to this paper by Peng Wang and graphic support by Zhaoyan Fan. Identification of certain commercial systems in this paper does not imply recommendation or endorsement by NIST. Nor does it imply that the products identified are necessarily the best available for the purpose.

References

[1] Henschke CI, Yankelevitz DF, Libby DM, Pasmantier MW, Smith JP, Miettinen OS (2006) Survivial of patients with stage 1 lung cancer detected on CT screening. *The New England Journal of Medicine* 355(17): 1763- 1771.

[2] Newman ME (2002) Spread of Epidemic Disease on Networks. Physical review E *66*(1): 016128.

[3] Gneiting T, Raftery AE (2005) Weather Forecasting with Ensemble Methods. *Science* 310(5746):248-249.

[4] Fama EF (1998) Market Efficiency, Long-Term Returns and Behavioral Finance. *Journal of Financial Economics* 49(3): 283-306.

[5] Suh JH, Kumara SR, Mysore SP (1999) Machinery Fault Diagnosis and Prognosis: Application of Advanced Signal Processing Techniques. *CIRP Annals-Manufacturing Technology* 48(1): 317-320.

[6] Inman DJ, Farrar CR, Lopes JV, Steffen JV (2005) Damage Prognosis for Aerospace Civil and Mechanical Systems. *John Wiley & Sons*, Ltd., Chichester, UK.

[7] Vachtsevanos G, Lewis F, Roemer M, Hess A, Wu B (2006) intelligent Fault Diagnosis and Prognosis for Engineering Systems. *John Wiley & Sons*, inc., Hoboken, New Jersey.

[8] Jardine AKS, Lin D, Banjevic D (2006) A Review On Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance. *Mechanical Systems and Signal Processing* 20(7): 1483-1510.

[9] Talbott C (1998) Prognosis of Remaining Machine Life Based On Condition. *Sound and Vibration* 32(6): 16-18.

[10] Tonshoff HK, Wulfsberg JP, Kals HJJ, Konig W, van Luttervelt CA (1988) Development and Trends in Monitoring and Control of Machining Process. *CIRP Annals-Manufacturing Technology* 37(2): 611-622.
[11] Byrne G, Dornfeld D, inasaki I, Ketteler G, König W, Teti R (1995) Tool Condition Monitoring (Tcm)—the Status of Research and industrial Application. *CIRP Annals-Manufacturing Technology* 44(2): 541-567.
[12] Jemielniak, K (1999) Commercial Tool Condition Monitoring Systems. *The international Journal of Advanced Manufacturing Technology* 15(10): 711-721.
[13] Peng Y, Dong M, Zuo MJ (2010) Current Status of Machine Prognostics in Condition-Based Maintenance: A Review. *international Journal of Advanced Manufacturing Technology* 50: 297-313.
[14] Elwany A, Gebraeel N (2008) Sensor Driven Prognostic Models for Equipment Replacement and Sparse Parts inventory. *IEEE Transactions on industrial Electronic* 40(7): 629-639.
[15] Marksberry PW, Jawahir IS (2008) A Comprehensive Tool-Wear/Tool-Life Performance Model in the Evaluation of Ndm (Near Dry Machining) for Sustainable Manufacturing. *International Journal of Machine Tools & Manufacture* 48(7-8): 878-886.
[16] Wiendahl HP, ElMaraghy HA, Nyhuis P, Zäh MF, Wiendahl HH, Duffie N, Brieke M (2007) Changeable Manufacturing-Classification Design and Operation. *CIRP Annals-Manufacturing Technology* 56(2): 783-809.
[17] Bi ZM., Lang SYT, Shen W, Wang L (2008) Reconfigurable Manufacturing Systems: The State of The Art. *International Journal of Production Research* 46(4): 967-992.
[18] Kothamasu R, Huang S, VerDuin W (2006) System Health Monitoring and Prognostics-A Review of Current Paradigms and Practices. *International Journal of Advanced Manufacturing Technology* 28: 1012-1024.
[19] Teti R, Jemielniak K, O'Donnell G, Dornfeld D (2010) Advanced Monitoring of Machining Operations. *CIRP Annals-Manufacturing Technology* 59(2): 717-739.
[20] Zhu K, Wong YS, Hong GS (2009) Wavelet Analysis of Sensor Signals for Tool Condition Monitoring: A Review and Some New Results. *International Journal of Machine Tools and Manufacture* 49(7): 537-553.
[21] Dutta S, Pal S K, Mukhopadhyay S, Sen R (2013) Application of Digital Image Processing in Tool Condition Monitoring: A Review. *CIRP Journal of Manufacturing Science and Technology* 6(3): 212-232.
[22] Zhou ZD, Chen YP, Fuh JYH, Nee AYC (2000) integrated Condition Monitoring and Fault Diagnosis for Modern Manufacturing Systems. *CIRP Annals-Manufacturing Technology* 49(1): 387-390.
[23] Mehnen J, Tinsley L, Roy R (2014) Automated In-Service Damage Identification, *CIRP Annals-Manufacturing Technology* 63 (1): 33-36.
[24] Le SK, Fouladirad M, Barros A, Levrat E, Lung B (2013) Remaining Useful Life Estimation based on Stochastic Deterioration Models: A Comparative Study. *Reliability Engineering and System Safety* 112: 165-175.
[25] Gouriveau R, Zerhouni N (2012) Connexionist-Systems-Based Long Term Prediction Approaches for Prognostics. *IEEE Transactions on Reliability* 61(4): 909-920.
[26] Heng A, Zhang S, Tan AC, Mathew J (2009) Rotating Machinery Prognostics: State of The Art, Challenges and Opportunities. *Mechanical Systems and Signal Processing* 23(3): 724-739.
[27] Sun B, Zeng S, Kang R, Pecht MG (2012) Benefits and Challenges of System Prognostics. *IEEE Transactions on Reliability* 61(2): 323-335.
[28] An D, Kim NH, Choi JH (2013) Options for Prognostics Methods: A Review of Data-Driven and Physics-Based Prognostics. *Proceedings of 54th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, Boston, MA, United States, April.
[29] Si XS, Wang W, Hu CH, Zhou DH (2011) Remaining Useful Life Estimation–A Review On The Statistical Data Driven Approaches. *European Journal of Operational Research* 213(1): 1-14.
[30] Sikorska JZ, Hodkiewicz M, Ma L (2011) Prognostic Modelling Options for Remaining Useful Life Estimation By industry. *Mechanical Systems and Signal Processing* 25(5): 1803-1836.
[31] Baraldi P, Mangili F, Zio E (2013) investigation of Uncertainty Treatment Capability of Model-based and Data-driven Prognostic Methods using Simulated Data. *Reliability Engineering and System Safety* 112: 94-108
[32] Zhang Q, Cheng L, Boutaba R (2010) Cloud Computing: State-of-the-Art and Research Challenges. *Journal of Internet Services and Applications*, 1(1): 7-18.
[33] Armbrust M, Fox A, Griffith R, Joseph AD, Katz R, Konwinski A, Zaharia M (2010) A View of Cloud Computing. *Communications of the ACM* 53(4): 50-58.
[34] Ren L, Zhang L, Wang L, Tao F, Chai X (2014) Cloud Manufacturing: Key Characteristics and Applications. *International Journal of Computer Integrated Manufacturing* 1-15.
[35] Lin Q, Xia K, Wang L, Gao L (2013) Research Progress of Cloud Manufacturing in China: A Literature Survey. *Proceedings of the ASME 2013 international Manufacturing Science and Engineering Conference*, Madison, Wisconsin, USA, MSEC2013-1168, June.
[36] Wu D, Greer M.J, Rosen DW, Schaefer D (2013) Cloud Manufacturing: Strategic Vision and State-of-The-Art. *Journal of Manufacturing Systems* 32(4): 564-579.
[37] Venters W, Whitley EA (2012) A Critical Review of Cloud Computing: Researching Desires and Realities. *Journal of information Technology* 27: 179-197.
[38] Yang YJ, Gao RX, Fan ZY, Wang JJ, Wang LH (2014) Cloud-based Prognosis: Perspective and Challenge. *Proceedings of the ASME 2014 International Manufacturing Science and Engineering Conference*, Detroit, Michigan, USA, June.
[39] Valilai OF, Houshmand M (2013) A Collaborative and Integrated Platform to Support Distributed Manufacturing System Using a Service-Oriented Approach Based on Cloud Computing Paradigm. *Robotics and Computer-Integrated Manufacturing* 29:110–127
[40] Li BH, Zhang L, Wang SL, Tao F, Cao JW, Jiang XD, Song X, Chai XD (2010) Cloud Manufacturing: A New Service-Oriented Networked Manufacturing Model. *Computer Integrated Manufacturing Systems* 16(1):1-7.
[41] Goldhar JD, Jelinek M (1990) Manufacturing as a Service Business: CIM in the 21st Century. *Computers in Industry* 14(1):225–245.
[42] Masood T, Roy R, Harrison A, Xu Y, Gregson S, Reeve C (2015) Integrating Through-Life Engineering Service Knowledge With Product Design And Manufacture. *International Journal of Computer Integrated Manufacturing* 28 (1): 59-74.
[43] Tracht K, Goch G, Schuh P, Sorg M, Westerkamp JF (2013) Failure Probability Prediction Based On Condition Monitoring Data of Wind Energy Systems for Spare Parts Supply. *CIRP Annals-Manufacturing Technology* 62(1): 127-130.
[44] Janjarasjitt S, Ocak H, Loparo KA (2008) Bearing Condition Diagnosis and Prognosis Using Applied Nonlinear Dynamical Analysis of Machine Vibration Signal. *Journal of Sound and Vibration* 317: 112-126.
[45] Kamarthi S, Kumara S, Cohen PH (2000) Flank Wear Estimation in Turning through Wavelet Representation of Acoustic Emission Signals. *Journal of Manufacturing Science and Engineering* 122(1): 12-19.
[46] Mills B, Redford A (1983) Machinability of Engineering Materials. *Applied Science*, London.
[47] Hoffman E (1984) Fundamentals of Tool Design. *SME*, Dearborn, MI.
[48] Niebel B, Draper A, Wysk R (1989) Modern Manufacturing Process Engineering. *McGraw-Hill*, New York.
[49] Poulachon G, Moisan A, Jawahir IS (2001) Tool-Wear Mechanisms in Hard Turning with Polycrystalline Cubic Boron Nitride Tools. *Wear* 250(1-2): 576-586.
[50] Lau W, Venuvinod P, Rubenstein C (1980) The Relation Between Tool Geometry and The Taylor Tool-Life Constant. *international Journal of Machine Tool Designing and Research* 20(1): 29–44.
[51] Quinto DT (1988) Mechanical Property and Structure Relationships in Hard Coated Carbide Tools. *Metals Technology* 9: 60–75.
[52] Oxley P (1989) The Mechanics of Machining—An Analytical Approach To Assessing Machinability. *Ellis Horwood*, Chichester, UK.
[53] Boothroyd G, Knight W (2006) Fundamentals of Metal Machining and Machine Tools. *CSC Press*, USA.
[54] Usui E, Hirota A, Masuko M (1997) Analytical Prediction of Three-Dimensional Cutting Process-Part 1: Basic Cutting Model and Energy Approach. *Journal of Engineering for industry* 100(2): 222-228.
[55] Choudhury SK, Srinivas P (2004) Tool Wear Prediction in Turning. *Journal of Materials Processing Technology* 153-154: 276-280.
[56] Kishawy HA, Kannan S, Balazinski M (2005) Analytical Modeling of Tool Wear Progression During Turning Particulate Reinforced Metal Matrix Composites. *CIRP Annals-Manufacturing Technology* 54(1): 55-58.
[57] Paris PC, Gomez MP, Anderson WE (1961) A Rational Analytic Theory of Fatigue. *The Trend in Engineering* 13: 9-14.
[58] Pugno N, Ciavarella M, Cornetti P, Carpinteri A (2006) A Generalized Paris' Law for Fatigue Crack Growth. *Journal of the Mechanics and Physics of Solids* 54: 1333-1349.
[59] Li Y, Billington S, Zhang C, Kurfess T, Danyluk S, Liang S (1999) Adaptive Prognostics for Rolling Element Bearing Condition. *Mechanical Systems and Signal Processing* 13(1): 103-113.
[60] Aslantas K, Tasgetiren S (2004) A Study of Spur Gear Pitting formation and Life Prediction. *Wear* 257: 1167-1175.
[61] Newman JC, Philips EP, Swain MH (1999) Fatigue-life Prediction Methodology using Small-Crack Theory. *International Journal of Fatigue* 21(2): 109-119.
[62] Ioannides E, Harris TA (1985) A New Fatigue Life Model for Rolling Bearings. *Journal of Tribology* 107(3): 367-377.
[63] Kotzalas MN, Harris TA (2001) Fatigue Failure Progression in Ball Bearings. *Journal of Tribology* 123(22): 238-242.
[64] Choi Y, Liu CR (2007) Spall Progression Life Model for Rolling Contact Verified By Finish Hard Machined Surfaces. *Wear* 262(1-2): 24-35.

[65] Qiu J, Seth BB, Liang SY, Zhang C (2002) Damage Mechanics Approach for Bearing Lifetime Prognostics. *Mechanical systems and signal processing* 16(5): 817-829.
[66] Fan KC, Chen HM, Kuo TH (2012) Prediction of Machining Accuracy Degradation of Machine Tools. *Precision Engineering* 36: 288-298.
[67] Abouelatta O, Madl J (2001) Surface Roughness Prediction Based On Cutting Parameters and Tool Vibration in Turning. *Journal of Materials Processing Technology* 118: 269-277.
[68] Bhattacharyya P, Senupta D, Mukhopadhaya S (2007) Cutting force Based Real-Time Estimation of Tool Wear in Face Milling Using A Combination of Signal Process Techniques. *Mechanical Systems and Signal Processing* 21(6): 2665-2683.
[69] Leone C, D'Addona D, Teti R (2011) Tool Wear Modeling Through Regression Analysis and intelligent Methods for Nickel Base Alloy Machining. *CIRP Journal of Manufacturing Science and Technology* 4: 327-331.
[70] Malhi A, Yan R, Gao RX (2011) Prognosis of Defect Propagation Based On Recurrent Neural Networks. *IEEE Transactions on Instrumentation and Measuremenht* 60(3): 703-711.
[71] Huang RQ, Xi LF, Li XL, Liu R, Qiu H, Lee J (2007) Residual Life Prediction for Ball Bearing based on Self-organizing Map and Back Propagation Neural Network Methods. *Mechanical Systems and Signal Processing* 21(1): 193-207.
[72] Halligam G, Jagannathan S (2011) Pca Based Fault Isolation and Prognosis with Application to Pump. *International Journal of Advanced Manufacturing Technology* 55: 699-707.
[73] 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. *Mechanical Systems and Signal Processing* 21: 466-479.
[74] Tian Z (2012) An Artificial Neural Network Method for Remaining Useful Life Prediction of Equipment Subject To Condition Monitoring. *Journal of intelligent Manufacturing* 23(2): 227-237.
[75] Huang B, Chen JC (2003) An in-Process Neural Network-Based Surface Roughness Prediction System Using A Dynamometer in End Milling Operations. *International Journal of Advanced Manufacturing Technology* 21: 339-347.
[76] Mahamad AK, Saon S, Hiyama T (2010) Predicting Remaining Useful Life of Rotating Machinery Based Artificial Neural Network. *Computers and Mathematics with Applications* 60: 1078-1087.
[77] D'Addona D, Segreto T, Simeone A, Teti R (2011) Ann Tool Wear Modeling in The Machining of Nickel Superalloy industry Products. *CIRP Journal of Manufacturing Science and Technology* 4: 33-37.
[78] Huang SH, Zhang HC (1994) Artificial Neural Networks in Manufacturing: Concepts, Application, and Perspectives. *IEEE Transactions on Components, Packaging, and Manufacturing Technology* 17(2): 212-228.
[79] Pontes FJ, Paiva APD, Balestrassi PP, Ferreira JR, Silva MBD (2012) Optimization of Radial Basis Function Neural Network Employed for Prediction of Surface Roughness in Hard Turning Process Using Taguchi's Orthogonal Arrays. *Expert Systems with Applications* 39(9): 7776-7787.
[80] Herzog MA, Marwala T, Heyns PS (2009) Machine and Component Residual Life Estimation through the Application of Neural Networks. *Reliability Engineering and System Safety* 94: 479-489.
[81] DePold HR, Gass FD (1999) The Application of Expert Systems and Neural Networks To Gas Turbine Prognostics and Diagnostics. *Journal of Engineering for Gas Turbines and Power* 121(4): 607-612.
[82] Dong J, Subrahmanyam KVR, Wong YS, Hong GS, Mohanty AR (2006) Bayesian-inference–Based Neural Networks for Tool Wear Estimation. *International Journal of Manufacturing Technology* 30: 797-807.
[83] Pan Y, Chen J, Li X (2010) Bearing Performance Degradation Assessment Based On Lifting Wavelet Packet Decomposition and Fuzzy C-Means. *Mechanical Systems and Signal Processing* 24: 559-566.
[84] Zhao Z, Quan Q, Cai KY (2014) A Profust Reliability based Approach to Prognostics and Health Management. *IEEE Transactions on Reliability* 63(1): 26-41.
[85] Wang WQ, Golnaraghi MF, Ismail F (2004) Prognosis of Machine Health Condition Using Neuro-Fuzzy Systems. *Mechanical Systems and Signal Processing* 18(4): 813-831.
[86] Liu J, Wang W, Golnaraghi F (2009) A Multi-Step Predictor with A Variable input Pattern for System State forecasting. *Mechanical Systems and Signal Processing* 13: 1586-1599.
[87] Tran VT, Yang BS, Tan ACC (2009) Multi-Step Ahead Direct Prediction For The Machine Condition Prognosis Using Regression Tress And Neruo-Fuzzy Systems. *Expert Systems with Applications* 36: 9378-9387.
[88] Chen C, Vachtsevanos G, Orchard ME (2012) Machine Remaining Useful Life Prediction: An integrated Adaptive Neuro-Fuzzy and High-Order Particle Filtering Approach. *Mechanical Systems and Signal Processing* 28: 597-607.
[89] Soualhi A, Razik H, Clerc G, Doan DD (2014) Prognosis of Bearing Failures Using Hidden Markov Models and the Adaptive Neuro-fuzzy inference System. *IEEE Transaction on industrial Electronic* 61(6): 2864-2874.
[90] Cox DR, Miller HD (1977) The Theory of Stochastic Processes. *CRC Press*, USA.
[91] Si XS, Wang W, Hu CH, Chen MY, Zhou DH (2013) A Wiener-Process-Based Degradation Model with A Recursive Filter Algorithm for Remaining Useful Life Estimation. *Mechanical Systems and Signal Processing* 35(1-2): 219-237.
[92] Wang X, Balakrishnan N, Guo B (2014) Residual Life Estimation Based On A Generalized Wiener Degradation Process. *Reliability Engineering and System Safety* 124: 13-23.
[93] Wang ZQ, Hu CH, Wang W, Si XS (2014) An Additive Wiener Process-based Prognostic Model for Hybrid Deteriorating Systems. *IEEE Transactions on Reliability* 63(1): 208-222.
[94] Van-Noortwijk JM (2009) A Survey of The Application of Gamma Processes in Maintenance. *Reliability Engineering and System Safety* 94(1): 2-21.
[95] Park C, Padgett WJ (2005) Accelerated Degradation Models for Failure Based On Geometric Brownian Motion and Gamma Processes. *Lifetime Data Analysis* 11(4): 511–527
[96] Park C, Padgett WJ (2006) Stochastic Degradation Models with Several Accelerating Variables. *IEEE Transactions on Reliability* 55 (2): 379–390.
[97] Tran VT, Thom PH, Yang BS, Tien NT (2012) Machine Performance Degradation Assessment and Remaining Useful Life Prediction Using Proportional Hazard Model and Support Vector Machine. *Mechanical Systems and Signal Processing* 32: 323-330.
[98] Jardine AKS, Banjevic D, Wiseman M, Buck S, Joseph T (2001) Optimizing A Mine Haul Truck Wheel Motors' Condition Monitoring Program: Use of Proportional Hazards Modeling. *Journal of Quality in Maintenance Engineering* 7(4): 286-301.
[99] Gasmi S, Love CE, Kahle W (2003) A General Repair, Proportional-Hazards, Framework To Model Complex Repairable Systems. *IEEE Transactions on Reliability* 52(1): 26-32.
[100] Tran VT, Thom PH, Yang BS, Tien NT (2012) Machine Performance Degradation Assessment and Remaining Useful Life Prediction Using Proportional Hazard Model and Support Vector Machine. *Mechanical Systems and Signal Processing*, 32: 323-330.
[101] Vlok PJ, Wnek M, Zygmunt M (2004) Utilising Statistical Residual Life Estimates of Bearings To Quantify The influence of Preventive Maintenance Actions. *Mechanical Systems and Signal Processing* 18(4): 833-847.
[102] Rabiner LR (1988) A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77(2): 257-286
[103] Camci F, Chinnam RB (2010) Health-State Estimation and Prognostics in Machining Processes. *IEEE Transactions on Automation Science and Engineering* 7(3): 581-597.
[104] Baruah P, Chinnam R (2005) HMMs for Diagnostics and Prognostics in Machining Processes. *International Journal of Production Research* 43(6): 1275-1293.
[105] Yan J, Guo C, Wang X (2011) A Dynamic Multi-Scale Markov Model Based Methodology for Remaining Life Prediction. *Mechanical Systems and Signal Processing* 25(4): 1364-1376.
[106] Yu SZ (2010) Hidden Semi-Markov Models. *Artificial Intelligence*, 174(2): 215-243.
[107] Tobon-Mejia DA, Medjaher K, Zerhouni N, Tripot G (2012) A Data-Driven Failure Prognostics Method Based On Mixture of Gaussians Hidden Markov Models. *IEEE Transactions on Reliability* 61(2): 491-503.
[108] Zaidi SSH, Aviyente S, Salman M, Shin KK, Strangas EG (2011) Prognosis of Gear Failures in Dc Starter Motors Using Hidden Markov Models. *IEEE Transactions on industrial Electronics* 58(5): 1695-1706.
[109] Zhou ZJ, Hu CH, Xu DL, Chen MY, Zhou DH (2010) A Model for Real-time Failure Prognosis Based on Hidden Markov Model and Belief Rule Base. *European Journal of Operational Research* 207(1): 269-283.
[110] Dong M, He David (2007) Hidden Semi-Markov Model-based Methodology for Multi-sensor Equipment Health Diagnosis and Prognosis. *European Journal of Operation Research* 178(3): 858-878.
[111] Peng Y, Dong M (2011) A Prognosis Method using Age-dependent Hidden Semi-Markov Model for Equipment Health Prediction. *Mechanical Systems and Signal Processing* 25(1): 237-252.
[112] Gasperin M, Juricic D, Boskoski P, Vizintin J (2011) Model-Based Prognostics of Gear Health Using Stochastic Dynamic Models. *Mechanical Systems and Signal Processing* 25: 537-548.
[113] Caesarendra W, Widodo A, Thom PH, Yang BS, Setiawan JD (2011) Combined Probability Approach and indirect Data-Driven Method for Bearing Degradation Prognostics. *IEEE Transactions on Reliability* 60(1): 14-20.
[114] Engel SJ, Gilmartin BJ, Bongort K, Hess A (2000) Prognostics, The Real Issues Involved with Predicting Life Remaining. *IEEE Aerospace Conference Proceedings* 6: 457-470.
[115] Worden K, Hensman JJ (2012) Parameter Estimation and Model Selection for A Class of Hysteretic Systems using Bayesian Inference. *Mechanical Systems and Signal Processing* 32: 153-169.
[116] Karandikar JM, Kim NH, Schmitz TL (2012) Prediction of Remaining Useful Life for Fatigue-Damaged Structures Using Bayesian inference. *Engineering Fracture Mechanics* 96: 588-605.

[117] An D, Choi JH, Schmitz TL, Kim NH (2011) in Situ Monitoring and Prediction of Progressive Joint Wear Using Bayesian Statistics. *Wear* 270: 828-838.

[118] Welch G., Bishop G. (1995) An Introduction to the Kalman Filter.

[119] Bar-Shalom Y, Li XR, Kirubarajan T (2001) Estimation with Applications to Tracking and Navigation. *John Wiley & Sons*, NY, USA.

[120] Feng L, Wang H, Si X, Zou H (2013) A State-Space-Based Prognostic Model for Hidden and Age-Dependent Nonlinear Degradation Process. *IEEE Transactions on Automation Science and Engineering* 10(4): 1072-1086.

[121] Saurabh SB, Mahendra PS (2014) An Adaptive Unscented Kalman Filter for Tracking Sudden Stiffness Changes. *Mechanical System and Signal Processing* 49(1-2): 181-195.

[122] Wan EA, Van der Merwe R (2000) The Unscented Kalman Filter for Nonlinear Estimaton. *The IEEE Adaptive Systems for Signal Processing, Communications and Control Symposium*, 153- 158, Lake Louise, Alta, October.

[123] Si XS, Wang W, Hu CH, Zhou DH (2014) Estimating Remaining Useful Life with Three-source Variability in Degradation Modelling. *IEEE Transaction on Reliability* 63(1): 167-190.

[124] Yang SK (2003) A Condition-Based Failure-Prediction and Processing-Scheme for Preventive Maintenance. *IEEE Transactions on Reliability* 52(3): 373-383.

[125] Reuben LCK, Mba D (2014) Diagnostics and Prognostics using Switching Kalman Filters. *Structural Health Monitoring* 13(3): 296-306.

[126] Zghal M, Mevel L, Pierre DM (2014) Modal Parameter Estimation Using Interacting Kalman Filter. *Mechanical Systems and Signal Processing* 47: 139-150.

[127] Arulampalam MS, Maskell S, Gordon N, Clapp T (2002) A Tutorial On Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. *IEEE Transactions on Signal Processing* 50(2): 174-188.

[128] Sun J, Zuo H, Wang W, Pecht MG (2012) Application of a State Space Modeling Technique to System Prognostics Based on a Health index for Condition-Based Maintenance. *Mechanical Systems and Signal Processing* 28: 585-596.

[129] Li TC, Sattar TP, Sun SD (2012) Deterministic Resampling: Unbiased Sampling to Avoid Sample Impoverishment in Particle Filters. Signal *Processing* 92: 1637-1645.

[130] Li TC, Sun SD, Sattar TP, Corchado JM (2014) Fight Sample Degeneracy and Impoverishment in Particle Filters: A Review of Intelligent Approaches. *Expert Systems with Application,* 51: 3944-3954.

[131] Wang P, Gao R, Fan Z (2015) Swithcing Local Search Particle Filtering Applied to Heat Exchanger System, Proceedings of 2015 IEEE International Instrumentation and Measurement Technology Conference, Pisa, Italy, May.

[132] Haug AJ (2005) A Tutorial on Bayesian Estimation and Tracking Techniques Applicable to Nonlinear and Non-Gaussian Processes. *MITRE Technical Report*, MITRE Corporation.

[133] Wang J, Wang P, Gao R (2014) Particle Filter for Tool Wear Prediction, *Proceedings of 42nd North American Manufacturing Research Conference*, Detroit, Michigan, June.

[134] Zio E, Peloni G (2011) Particle Filtering Prognostic Estimation of The Remaining Useful Life of Nonlinear Components. *Reliability Engineering and System Safety* 96(3): 403-409.

[135] Cadini F, Zio E, Avram D (2009) Monte Carlo-Based Filtering for Fatigue Crack Growth Estimation. *Probabilistic Engineering Mechanics* 24(3): 367-373.

[136] Wang J, Wang P, Gao RX (2013) Tool Life Prediction for Sustainable Manufacturing. *Proceedings of 11th Global Conference on Sustainable Manufacturing*, Berlin, Germany, September.

[137] Liu Q, Dong M, Peng Y (2012) A Novel Method for Online Health Prognosis of Equipment Based On Hidden Semi-Markov Model Using Sequential Monte Carlo Methods. *Mechanical Systems and Signal Processing* 32: 331-348.

[138] Hu Z, Liu X, Li J (2014) Two-stage Prediction and Update Particle Filtering Algorithm based on Particle Weight Optimization in Multi-sensor Observation. *High Technology Letter* 20(1): 34-41.

[139] Chen C, Zhang B, Vachtsevanos G (2012) Prediction of Machine Health Condition Using Neuro-Fuzzy and Bayesian Algorithms. *IEEE Transactions on instrumentation and Measurement* 61(2): 297-306.

[140] Baraldi P, Compare M, Sauco S, Zio E (2013) Ensemble Neural Network-based Particle Filtering for Prognosis. *Mechanical Systems and Signal Processing* 41(1-2):288-300.

[141] Compare M, Zio E (2014) Predictive Maintenance by Risk Sensitive Particle Filtering. *IEEE Transactions on Reliability* 63(1): 546-554.

[142] Saxena A, Celaya J, Balaban E, Goebel K, Saha B, Saha S, Schwabacher M (2008) Metrics for Evaluating Performance of Prognostic Techniques. *Proceedings of 2008 International Conference on Prognostics and Health Management,* 1-17, Denver, CO, USA, October.

[143] Saxena A, Celaya J, Saha B, Saha S, Goeble K (2010) Metrics for offline Evaluation of Prognostic Performance. *International Journal of Prognostics and Health Management* 1: 4.

[144] Zhao F, Tian Z, Zeng Y (2013) Uncertainty Quantification in Gear Remaining Useful Life Prediction through an integrated Prognostics Method. *IEEE Transactions on Reliability* 62(1): 146-159.

[145] Son J, Zhou Q, Zhou S, Mao X, Salman M (2013) Evaluation and Comparison of Mixed Effects Model Based Prognosis for Hard Failure. *IEEE Transactions on Reliability* 62(2): 379-394.

[146] Hu C, Youn BD, Wang P, Taek-Yoon J (2012) Ensemble of Data-Driven Prognostic Algorithms for Robust Prediction of Remaining Useful Life. *Reliability Engineering and System Safety* 103: 120-135.

[147] Sun J, Zuo H, Wang W, Pecht MG (2014) Prognostics Uncertainty Reduction by Fusing On-line Monitoring Data based on a State-space-based Degradation Model. *Mechanical Systems and Signal Processing* 45(2): 396-407.

[148] Huynh KT, Barros A, Berenguer C (2012) Maintenance Decision-Making for Systems Operating Under indirect Condition Monitoring: Value of Online information and Impact of Measurement Uncertainty. *IEEE Transactions on Reliability* 61(2): 410-425.

[149] Moghaddass R, Zuo MJ (2014) An integrated Framework for Online Diagnostic and Prognostic Health Monitoring using a Multistate Deterioration Process. *Reliability Engineering and System Safety* 124: 92-104.

[150] Lu C, Tao L, Fan H (2014) An intelligent Approach to Machine Component Health Prognostics by Utilizing only Truncated Histories. *Mechanical Systems and Signal Processing* 42(1-2): 300-313.

[151] Xu X (2012) From Cloud Computing to Cloud Manufacturing. *Robotics and Computer-Integrated Manufacturing* 28:75–86.

[152] Zhou JT, Yang HC, Wang MW, Jing SK, Mo R (2011) GetCM: A Vision of Cloud Manufacturing. In: *Zhang Y* (*ed*) *Advanced Materials Research*, 213: 388-392.

[153] Wu D, Greer MJ, Rosen DW, Schaefer D (2013) Cloud Manufacturing: Strategic Vision and State-of-the-Art. *Journal of Manufacturing Systems* 32(4):564-579.

[154] Heinrichs W, (2005) Do It Anywhere. *IEE Electronics Systems and Software* 3(4):30-33.

[155] Manenti P (2011) Building the Global Cars of the Future. *Managing Automation* 26(1):8-14.

[156] Tao F, Zhang L, Venkatesh VC, Luo Y, Cheng Y (2011) Cloud Manufacturing: A Computing and Service-Oriented Manufacturing Model. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* 225(10):1969-1976.

[157] Wang L, Gao RX, Ragai I (2014) An Integrated Cyber-Physical System for Cloud Manufacturing. *Proceedings of the ASME International Manufacturing Science and Engineering Conference*, MSEC2014-4171, Detroit, MI, USA, June.

[158] Wang L, Wang XV, Gao L, Váncza J (2014) A Cloud-Based Approach for WEEE Remanufacturing. *CIRP Annals – Manufacturing Technology* 63(1):409-412.

[159] Mell P, Grance T (2011) The NIST Definition of Cloud Computing. *NIST Special Publication* 800-145, U.S. Department of Commerce, available online at http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

[160] Xiang F, Hu Y (2012) Cloud Manufacturing Resource Access System Based on Internet of Things. *In: Sun D et al (eds) Applied Mechanics and Materials* 121-126: 2421-2425.

[161] Bandyopadhyay D, Sen J (2011) Internet of Things: Applications and Challenges in Technology and Standardization. *Springer International Journal of Wireless Personal Communications* 58(1):49-69.

[162] Heath S (2003) Embedded Systems Design Second Edition. *Butterworth-Heinemann*, Oxford, UK.

[163] Martin D, Burstein M, Mcdermott D, Mcilraith S, Paolucci M, Sycara K, Mcguinness DL, Sirin E, Srinivasan N (2007) Bringing Semantics to Web Services with Owl-S. *World Wide Web* 10(3):243-277.

[164] Berner-Lee T (1998) Semantic Web Road Map. *Design Issues for the World Wide Web* 1-10.

[165] MTConnect Institute (2014) MTConnect v. 1.3.0, http://www.mtconnect.org/downloads/standard.aspx. Accessed 8 July 2014.

[166] Macia-Perez F, Berna-Martinez JV, Marcos-Jorquera D, Lorenzo-Fonseca I, Ferrandiz-Colmeiro A (2012) A New Paradigm: Cloud Agile Manufacturing. *International Journal of Advanced Science and Technology* 45:47-54.

[167] Zhang L, Luo Y, Tao F, Li BH, Ren L, Zhang X, Guo H, Cheng, Y, Hu A, Liu Y (2012) Cloud Manufacturing: A New Manufacturing Paradigm. *Enterprise Information Systems* 8(2): 167-187.

[168] Wang XV, Xu XW (2013) An Interoperable Solution for Cloud Manufacturing. *Robotics and Computer-Integrated Manufacturing* 29(4):232–247.

[169] Zhou Z, Liu Q, Xu W (2011) From Digital Manufacturing to Cloud Manufacturing. *International Journal of Engineering Innovation and Management* 1(1):1-14.

[170] Huang B, Li C, Yin C, Zhao X (2013) Cloud Manufacturing Service Platform for Small- and Medium-Sized Enterprises. *The International Journal of Advanced Manufacturing Technology* 65(9-12): 1261-1272.

[171] Macia-Perez F, Berna-Martinez JV, Marcos-Jorquera D, Lorenzo-Fonseca I, Ferrandiz-Colmeiro A (2012) Cloud Agile Manufacturing. *IOSR Journal of Engineering* 2(5):1045-1048.

[172] Ren L, Zhang L, Zhao C, Chai X (2013) Cloud Manufacturing Platform: Operating Paradigm, Functional Requirements, and Architecture Design. *Proceedings of the ASME 2013 International Manufacturing Science and Engineering Conference*, MSEC2013-1185, Madison, WI, USA, June.
[173] Wang XV, Xu XW (2013) ICMS: A Cloud-based Manufacturing System. In: *Li WD, Mehnen J (eds) Cloud Manufacturing: Distributed Computing Technologies for Global and Sustainable Manufacturing*, Springer-Verlag, London, 1-22.
[174] Schaefer D, Thames JL, Wellman RD, Wu D, Yim S, Rosen DW (2012) Distributed Collaborative Design and Manufacture in the Cloud – Motivation, Infrastructure, and Education. *American Society for Engineering Education Annual Conference*, AC2012-3017, San Antonio, June.
[175] Meier M, Seidelmann J, Mezgár I (2010) ManuCloud: The Next-Generation Manufacturing as a Service Environment. *European Research Consortium for Informatics and Mathematics (ERCIM) News* 83:33-34.
[176] ManuCloud, http://www.manuCloud-project.eu/index.php?id=233.
[177] CAPP-4-SMEs, http://www.capp-4-smes.eu/.
[178] Wang L, Shen W (2003) DPP: An Agent-Based Approach for Distributed Process Planning. *Journal of Intelligent Manufacturing* 14(5):429-439.
[179] Wang L, Feng HY, Cai N (2003) Architecture Design for Distributed Process Planning. *Journal of Manufacturing Systems* 22(2):99-115.
[180] Wang L, Holm M, Adamson G (2010) Embedding a Process Plan in Function Blocks for Adaptive Machining. *CIRP Annals – Manufacturing Technology* 59(1):433-436.
[181] Wang M, Zhou J, Jing S (2012) Cloud Manufacturing: Needs, Concept and Architecture. *IEEE 16th International Conference on Computer Supported Cooperative Work in Design*, Wuhan, China, May.
[182] Mokhtar A, Houshmand M (2010) Introducing a Roadmap to Implement the Universal Manufacturing Platform Using Axiomatic Design Theory. *International Journal of Manufacturing Research* 5:252–269.
[183] Ryan MD (2011) Viewpoint Cloud Computing Privacy Concerns on Our Doorstep. *Communications of the ACM* 54(1):36-38.
[184] Venters W, Whitley EA (2012) A Critical Review of Cloud Computing: Researching Desires and Realities. *Journal of Information Technology* 27:179-197.
[185] Adiseshan B (2012) Manufacturing in the Cloud. http://www.automation.com/pdf_articles/inkumo/inkumo_white_paper_manufacturing_in_the_cloud.pdf.
[186] Zhang Q, Cheng L, Boutaba R (2010) Cloud Computing: State-of-the-Art and Research Challenges. *Journal of Internet Services and Applications* 1(1):7-18.
[187] Armbrust M, Fox A, Griffith R, Joseph AD, Katz R, Konwinski A, Zaharia M (2010) *A View of Cloud Computing. Communications of the ACM* 53(4):50-58.
[188] Lin Q, Xia K, Wang L, Gao L (2013) Research Progress of Cloud Manufacturing in China: a Literature Survey. *Proceedings of the ASME 2013 International Manufacturing Science and Engineering Conference*, MSEC2013-1168, Madison, WI, USA, June.
[189] Wu D, Greer MJ, Rosen DW, Schaefer D (2013) Cloud Manufacturing: Strategic Vision and State-of-the-Art. *Journal of Manufacturing Systems* 32(4):564-579.
[190] Wu D, Rosen DW, Schaefer D (2014) Modeling and Analyzing the Material Flow of Crowdsourcing Processes in Cloud-Based Manufacturing Systems Using Stochastic Petri Nets. *Proceedings of the ASME 2014 International Manufacturing Science and Engineering Conference*, MSEC2014-3907, Detroit, MI, USA, June.
[191] Howe J (2006) The Rise of Crowdsourcing. *Wired* 14(6):1-4.
[192] Wu D, Rosen DW, Wang L, Schaefer D (2015) Cloud-Based Design and Manufacturing: A New Paradigm in Digital Manufacturing and Design Innovation. *Computer-Aided Design* 59:1-14.
[193] Wu D, Greer MJ, Rosen DW, Schaefer D (2013) Cloud Manufacturing: Drivers, Current Status, and Future Trends. *Proceedings of the ASME 2013 international Manufacturing Science and Engineering Conference*, MSEC2013-1106, Madison, WI, USA, June.
[194] Wu D, Thames JL, Rosen DW, Schaefer D (2012) Towards a Cloud-Based Design and Manufacturing Paradigm: Looking Backward, Looking Forward. *Proceedings of the ASME 2012 International Design Engineering Technical Conference and Computers and Information Engineering Conference*, DETC2012-70780, Chicago, IL, USA, August.
[195] Lee J, Lapira E, Bagheri B, Kao H (2013) Recent Advances and Trends in Predictive Manufacturing Systems in Big Data Environment. *Manufacturing Letters* 1:38-41.
[196] Lee J, Ni J, Djurdjanovic D, Qiu H, Liao H (2006) Intelligent Prognostics Tools and e-Maintenance. *Computers in Industry* 57:476-489.
[197] Vogl GW, Weiss BA, Donmez MA (2014) Standards Related to Prognostics and Health Management (PHM) for Manufacturing. *NIST Internal Report*, NISTIR 8012.
[198] Wells LJ, Camelio JA, Williams CB, White J (2014) Cyber-Physical Security Challenges in Manufacturing Systems. *Manufacturing Letters* 2:74-77.
[199] Wu D, Rosen DW, Schaefer D (2014) Cloud-Based Design and Manufaturing: Status and Promise. In D. Schaefer (ed.) *Cloud-Based Design and Manufacturing (CBDM): A Service-Oriented Product Development Paradigm for the 21st Century*, Springer, London, 1-24.
[200] Vijayaraghavan A, Sobel W, Fox A, Dornfeld D, Warndorf P (2008) Improving Machine Tool Interoperability using Standardized Interface Protocols: MTConnect™. *Proceedings of 2008 International Symposium on Flexible Automation*, Atlanta, GA, USA, June.
[201] Roemer M, Byington CS, Kacprzynski GJ, Vachtsevanos G (2006) An Overview of Selected Prognostic Technologies with Application to Engine Health Management. *Proceedings of 2006 ASME Turbo Expo 2006 Barcelona*, Spain, May.
[202] Möhring HC, Litwinski KM, Gümmer O (2010) Process Monitoring with Sensory Machine Tool Components. *CIRP Annals-Manufacturing Technology* 59(1): 383-386.
[203] Schrock DJ, Kang D, Bieler TR, Kwon P (2014) Phase Dependent Tool Wear in Turning Ti-6Al-4V Using Polycrystalline Diamond and Carbide Inserts. *Journal of Manufacturing Science and Engineering 136*(4): 041018.
[204] Karandikar J, Abbas A, Schmitz T (2014) Tool Life Prediction Using Bayesian updating, Part 2: Turning tool life using a Markov Chain Monte Carlo Approach. *Precision Engineering* 38:9-17.
[205] Karandikar J, Abbas A, Schmitz T (2014) Tool Life Prediction Using Bayesian Updating, Part 1: Milling Tool Life Model Using a Discrete Grid Method. *Precision Engineering* 38:18-27.
[206] Wang GF, Qian L (2012) Online tool wear prediction based on partial least square regression and Monte Carlo cross validation, Proceedings of the institution of Mechanical Engineers, Part B: *Journal of Engineering Manufacture*, 226: 1801-1807.
[207] Widodo A, Yang B (2011) Machine Health Prognostics Using Survival Probability and Support Vector Machine. *Expert systems with applications* 38(7): 8430-8437.
[208] Lee KB, Song EY, Gu PS (2012) Integration of MTConnect and Standard-Based Sensor Networks for Manufacturing Equipment Monitoring. *Proceedings of ASME 2012 International Manufacturing Science and Engineering Conference*, MSEC2012-7398, Notre Dame, IN, USA, June.
[209] Vijayaraghavan A, Huet L, Dornfeld D, Sobel W, Blomquist B, Conley M (2009) Addressing Process Planning and Verification Issues with MTConnect. *Transactions of the North American Manufacturing Research Institution of SME* 37:557-564.
[210] Vijayaraghavan A, Dornfeld D (2010) Automated Energy Monitoring of Machine Tools. *CIRP Annals – Manufacturing Technology* 59(1):21-24.
[211] Bhinge R, Park J, Biswas N, Helu M, Dornfeld D, Law K, Rachuri S (2014) An Intelligent Machine Monitoring System using Gaussian Process Regression for Energy Prediction. *Proceedings of 2014 IEEE International Conference on Big Data*, Washington, DC, USA, October.
[212] DMG Mori (2014) MORI NET Global Edition. http://en.dmgmori.com/products/software/mori-net/
[213] Yamazaki Mazak Europe (2014) Cyber Production Center. http://www.mazak.eu/Cyber-Production-Center/
[214] System Insights, Inc. (2014) VIMANA Core by System Insights. http://www.systeminsights.com/core/
[215] Niu G, Yang BS, Pecht M (2010) Development of an Optimized Condition-Based Maintenance System by Data Fusion and Reliability-Centered Maintenance. *Reliability Engineering & System Safety* 95(7): 786-796.
[216] Fouladirad M, Grall A (2011) Condition-Based Maintenance for a System Subject to a Non-Homogeneous Wear Process with a Wear Rate Transition. *Reliability Engineering & System Safety* 96(6): 611-618.
[217] Dieulle L, Berenguer C, Grall A, roussignol M (2003) Sequential Condition-Based Maintenance Scheduling for a Deteriorating System. *European Journal of Operational Research* 150: 451-461.
[218] Castanier B, Grall A, Berenguer C, roussignol M (2005) A Condition-Based Maintenance Policy with Non-Periodic Inspections for a Two-Unit Series System. *Reliability Engineering & System Safety* 87: 109-120.
[219] Lopez De Mantaras R, McSherry D, Bridge D, Leake D, Smyth B, Craw S, Faltings B, Lou Maher M, Cox M, Forbus K, Keane M, Watson I (2005) Retrieval, Reuse, Revision and Retention in Case-Based Reasoning. *The Knowledge Engineering Review* 20(03): 215-240.
[220] Morcous M, Rivard H, Hanna A (2002) Case-Based Reasoning System for Modelling Infrastructure Deterioration, *Journal of Computing in Civil Engineering* 16(2): 104-114.
[221] Yang BS, Kwon Jeong S, Oh YM, Tan ACC (2004) Case-Based Reasoning System with Petri Nets for Induction Motor Fault Diagnosis. *Expert Systems with Applications* 27(2): 301-311.
[222] Olsson E, Funk P, Xiong N (2004) Fault Diagnosis in Industry using Sensor Readings and Case-Based Reasoning, *Journal of Intelligent and Fuzzy systems* 15(1): 41-46.
[223] Yu R, Iung B, Panetto H (2003) A Multi-Agents Based E-Maintenance System With Case-Based Reasoning Decision Support. *Engineering applications of artificial intelligence* 16(4): 321-333.

[224] Jalali V, Leake D (2014) Adaption-Guded Case Based Maintenance. *Proceedings of 28th AAAI Conference on Artifical Intelligence*, Quebec, Canada, July.
[225] Bahga A, Madisetti V (2012) Analyzing Massive Machine Maintenance Data in a Computing Cloud, *IEEE Transactions on Parallel and Distributed Systems*, 23(10): 1831-1843.
[226] Barham P, Dragovic B, Fraser K, Hand S, Harris T, Ho A, Neugebauer R, Pratt I, Warfield A (2003) Xen and the Art of Virtualization. *ACM SIGOPS Operating Systems Review* 37(5): 164-177.
[227] Mei Y, Liu L, Pu X, Sivathanu S, Dong, X (2013) Performance Analysis of Network I/O Workloads in Virtualized Data Centers. *IEEE Transactions on Services Computing* 6(1): 48-63.
[228] Bourguiba M, Haddadou K, El Korbi I, and Pujolle G (2014) Improving NETWORK I/O VIRTUALIZATION for Cloud Computing. *IEEE Transactions on Parallel and Distributed Systems* 25(3): 673-681.
[229] Guan H, Ma R, Li J (2014) Workload-Aware Credit Scheduler for Improving Network I/O Performance in Virtualization Environment. *IEEE Transactions on Cloud Computing* 2(2): 130-142.
[230] Ranadive A, Kesavan M, Gavrilovska A, Schwan K (2008) Performance Implications of Virtualizing Multicore Cluster Machines. *Proceedings of the 2nd ACM Workshop on System-level Virtualization for High Performance Computing*, Glasgow, Scotland, March.
[231] Shvachko K, Kuang H, Radia S, Chansler R (2010) The Hadoop Distributed File System. *Proceedings of 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies*, Incline Village, NV, USA, May.
[232] Ghemawat S, Gobioff H, Leung ST (2003) The Google File System. *ACM SIGOPS Operating Systems Review*, 37(5): 29-43.
[233] Iosup A, Ostermann S, Yigitbasi MN, Prodan R, Fahringer T, Epema DH (2011) Performance Analysis of Cloud Computing Services for Many-Tasks Scientific Computing. *IEEE Transactions on Parallel and Distributed Systems* 22(6): 931-945.
[234] Huber N, von Quast M, Hauck M, Kounev S (2011) Evaluating and Modeling Virtualization Performance Overhead for Cloud Environments. *CLOSER* 563-573.
[235] Kousiouris G, Cucinotta T, Varvarigou T (2011) The Effects of Scheduling, Workload Type and Consolidation Scenarios on Virtual Machine Performance and Their Prediction Through Optimized Artificial Neural Networks. *Journal of Systems and Software* 84(8): 1270-1291.
[236] Singh R, Sharma U, Cecchet E, Shenoy P (2010) Autonomic Mix-Aware Provisioning for Non-Stationary Data Center Workloads. *Proceedings of the 7th International Conference on Autonomic Computing, Washington*, DC, June.
[237] Sotomayor B, Montero RS, Llorente IM, Foster I (2009) An Open Source Solution for Virtual Infrastructure Management in Private and Hybrid Clouds. *Proceedings of the IEEE international Conference on Internet Computing*, Cancouver, Canada, July.
[238] Dornfeld D, Wright PK, Vijayaraghavan A, Helu M (2009) Enabling Manufacturing Research through Interoperability. *Transactions of the North American Manufacturing Research Institution of SME* 37:443-450.
[239] Liang SY, Hecker RL, Landers RG (2004) Machining Process Monitoring and Control: the State-of-the-Art. *Journal of Manufacturing Science and Engineering* 126(2):297-310.
[240] Langner R (2011) Stuxnet: Dissecting a Cyberwarfare Weapon. *IEEE Security and Privacy* 9(3):49-51.
[241] Thames JL (2014) Distributed, Collaborative and Automated Cybersecurity Infrastructures for Cloud-Based Design and Manufacturing. In D. Schaefer (ed.) *Cloud-Based Design and Manufacturing (CBDM): A Service-Oriented Product Development Paradigm for the 21st Century*, Springer, London, 207-229.
[242] The MITRE Corporation (2014) Trusted Automated eXchange of Indicator Information: Enabling Cyber Threat Information Exchange. http://taxii.mitre.org/
[243] The MITRE Corporation (2014) Structured Threat Information eXpression: A Structured Language for Cyber Threat Intelligence Information. http://stix.mitre.org/
[244] Rutkowski A, Kadobayashi Y, Furey I, Rajnovid D, Martin R, Takahashi T (2010) Cybex the Cybersecurity Information Exchange Framework (X.1500). *ACM SIGCOMM Computer Communication Review* 40(5):59-64.