

REVIEW

Open Access



Remaining Useful Life Model and Assessment of Mechanical Products: A Brief Review and a Note on the State Space Model Method

Yawei Hu¹, Shujie Liu^{1*}, Huitian Lu² and Hongchao Zhang^{1,3}

Abstract

The remaining useful life (RUL) prediction of mechanical products has been widely studied for online system performance reliability, device remanufacturing, and product safety (safety awareness and safety improvement). These studies incorporated many different models, algorithms, and techniques for modeling and assessment. In this paper, methods of RUL assessment are summarized and expounded upon using two major methods: physics model based and data driven based methods. The advantages and disadvantages of each of these methods are deliberated and compared as well. Due to the intricacy of failure mechanism in system, and difficulty in physics degradation observation, RUL assessment based on observations of performance variables turns into a science in evaluating the degradation. A modeling method from control systems, the state space model (SSM), as a first order hidden Markov, is presented. In the context of non-linear and non-Gaussian systems, the SSM methodology is capable of performing remaining life assessment by using Bayesian estimation (sequential Monte Carlo). Being effective for non-linear and non-Gaussian dynamics, the methodology can perform the assessment recursively online for applications in CBM (condition based maintenance), PHM (prognostics and health management), remanufacturing, and system performance reliability. Finally, the discussion raises concerns regarding online sensing data for SSM modeling and assessment of RUL.

Keywords: Remaining useful life, State space model, Online assessment, Bayesian estimation, Particle filter, Remanufacturing

1 Introduction

With the quick development of science and technology, the increasing diversity of demand with growing product complexity comes with many new products safety challenges. Today, global industrial accidents frequently occur and (unfortunately) product safety issues contribute, at least partially, to these accidents. The issues of production safety and product safety not only cause casualties and property damage, but also environmental pollution. All of these issues affect the health of product

life, and are not conducive to either social harmony or stability. According to the research [1], due to safety issues with some mechanical products, many accidents and property damages occur every year. Additionally, the world's machinery accidents account for about 1/3 of the total number of accidents reported. This is especially the case with mechanical products related to civil and defense machinery, such as spaceflight, vehicles, electricity, weaponry, petrochemical, metallurgy. Mechanical product safety and reliability has become a common theme that is of paramount concern around the globe.

Failure prognostics was the definition of “the estimation of the Time to Failure (ETTF) and the risk of existence or later appearance of one or more failure modes” by the International Standard Organization (ISO) [2].

*Correspondence: liushujie@dlut.edu.cn

¹ School of Mechanical Engineering, Dalian University of Technology, Dalian 116024, China

Full list of author information is available at the end of the article

The terminology of RUL was normally reported in the literature instead of ETTF [3]. If the remaining life of mechanical products can be accurately predicted, appropriate measures could be then taken before the product fails, which will prevent accidents. RUL assessment, therefore, is very essential because it can provide not only important information for decision(s) in the condition based maintenance [4, 5], but it will contribute to make decisions on prognostics and health management (PHM). The system contains valuable information of installed equipment about the current health status, the critical components' RUL, and diagnosis of potential failures [6, 7]. For emerging industries, remanufacturing and residual life prediction also play a pivotal role in product remanufacturing feasibility analysis. Determining whether to remanufacture or recycle a product requires an estimate be made of its remaining life. Besides the environmental, economic, and technical effects of this estimate, it is also important to give a comprehensive consideration on retiring the product instead of remanufacturing [8]. The remaining life assessment is necessary for those remanufactured products that are in service to avoid abrupt mechanical failures. RUL prediction studies have been done and the proposals that have resulted from these studies have been introduced quite swiftly to the machinery industry. The RUL and its assessment are applied to online life estimation, forecasting, and performance reliability for critical equipment and machinery.

Machinery system complexities, such as failure mechanisms, extent of failures, and failure growth, when coupled with the dynamics and randomness of a working environment, results in that feature vector extractions are complicated to perform. The relationship between the detected (failure) signals and the remaining lifetime is difficult to determine or model, especially for heavy equipment with complex structure and dynamic loading. As long as the RUL prediction model for mechanical products does meet assessment needs, the operation's security and safety issues of the machinery will remain serious and restrict or limit economic health and sustainable social development. Global information is in a state of rapid advancement and the efficient and accurate residual life and reliability assessments are an urgent focus for academia research.

Residual life assessment for mechanical products, the world research hotspot, is reviewed in this paper. The assessment methods are separate into three categories: physics based method, data driven method, and a hybrid method of these two. The merits and drawbacks of each method are discussed and compared. A newer method based on the state space model (SSM) in assessing remaining life, widely cited in recent literatures, is also presented. By establishing an appropriate state space

model for a dynamic system (RUL assessment from product performance degradation) by state transfer equation and observation equation, the product performance status (degradation) at a different time can be predicted. Using a predetermined condition or threshold, the remaining life is regarded as the duration from current time and then extending out to the time when the state (degradation) meets or exceeds the critical threshold. This paper introduces the state space method in detail for the RUL prediction model, comprehensive analysis of the particle filter methods and the best approximation method today for most systems with nonlinear, non-Gaussian degradation characteristics. The methodology of assessing the remaining life based on particle filter method is provided for researchers' application reference. The paper concludes by raising, deliberating, and dispelling some of the practical issues with online sensor data modeling and RUL evaluation.

2 Basic Methods for RUL Assessment

There are three major methods used in RUL assessment: the physics based method, the data driven method, and a hybrid model of these two methods. The first one is based on physics dynamics and/or engineering principles. The physics based method is usually more flexible in the sense that this model could adapt with greater ease to new or unforeseen cases. If a system state deviates from the expected operation restrict, the model-based techniques continues to update model parameters to describe the new condition. This capability allows the model based techniques to dispense with extensive training and historical information that forms the hybrid model of the model based and data driven methods. The data driven method (usually model-free) works on comparative status assessments of a system under testing in conjunction with other learned occurrences. The model needs to be trained from historical data and under suitable conditions the new comparison is made in order to recognize the fault. The data driven method sets up nonlinear mapping between the input and output by training and the knowledge is stored by a set of parameter values. The task of performance assessment is done through the training of a classification arithmetic. The data driven method can be used in the situations where reliability of the physics model is low and the availability of comparative data is high. The hybrid model method should be employed if the reliability of the physics model and the availability of comparative data are both high.

2.1 Physics model

Physics model is one of the most important research area of life prediction and is based on theory of physics-mechanics and dynamics. There are four common

failure modes of mechanical products: wear, corrosion, deformation, and fracture. Substantially, products failure results from the interactions of materials, mechanics and chemicals [9]. Fatigue life depending on the process of crack initiation, propagation till final fracture, and physics based life prediction method have been considered in the initial stress analysis method, which obtains long-term development by utilizing fracture mechanics and crack mechanics theory. In 1850s and 1860s, WÖHLER first introduced the S–N (stress-rpm) curve and fatigue limits [10] from the fatigue damage tests of rail vehicle axles. Then, mean stress effects on fatigue life attracted wide attention, such as the average stress theory introduced by John Goodmab. Paris [11] raised the laws of crack propagation-Paris law, by crack technique in 1963, which offered a novel approach to estimate RUL. In 1971, Morrow et al. [12] put forward a fatigue life estimation method under the stress-strain (local stress-strain method) on the basis of Manson-Coffin. Subsequently, an amazing breakthrough in fatigue theory, the “damage tolerance design”, was developed. Researchers have since pursued more significant developmental strides based on Paris law. Ray et al. [13] and Mohanty et al. [14] proposed a nonlinear stochastic model of crack growth based on the measured signals, which could estimate the length of crack, the propagation rate of crack and residual life.

China, as a manufacturing industry leader in the world, is working to vigorously strengthen scientific and experimental research. For example, Liu et al. [15] carried out a sprayed dust accelerated wear test on F6L912G diesel engine and established a mathematical model of reliability and wear-resistance (R–W–t). Liu established that as long as the acceleration coefficient can be calculated then wear-resist reliable life of the engine can be estimated. Another example of advancing scientific understanding includes Shang, who proposed a multi-axial fatigue damage parameter in shearing form on the basis of the multi-axial critical damage surface and established a new prediction model [16]. Through fracture mechanics analysis, Ringsberg [17] exploited the fatigue life estimation strategy of rolling contact fatigue crack initiation. And two examples, a twin disc test and a railway wheel-rail rolling contact, are presented to verify the effectiveness of methodology. Fatemi [18] comprehensively reviewed cumulative fatigue damage analysis theory, which is the key to assess component or structure life. Besides, a new technique based on cumulative damage rate was presented [19]. The logarithmic probability density of fatigue life in any stress could be derived and then the fatigue life under fatigue load spectrum would be predicated. Zheng et al. [20] proposed that the load spectrum size can be represented in the load spectrum by overloading

that is caused by a maximum load. Under the interaction of loads, it is possible to predict the cumulative fatigue damage at fatigue crack initiation (FCI) of elements with variable amplitude load by Miner’s rule, which could also be applied to assess the fatigue life and cumulative fatigue damage threshold. Such studies and research were quite popular in China in the late 20th century, and achieved a greater amount of experience and new ideas to advance RUL prediction.

Due to the different mechanism of crack initiation and crack extension, various approaches are required in order to get more accurate results. A research method utilizing strain life methodology was proposed to compute the stress cycle numbers of FCI [21–23]. The Paris law was used to simulate the fatigue crack growth. For high accuracy of prediction, a physics-based approach, integrated observers and life models, was established for crack diagnostics and prognostics [24]. The fracture mechanism based approach assumes that there is a finite crack at the initial stage and that the extent of damage is related to the growth of the crack, whereas the energy-based approach is based on material properties. The strain energy dissipation of a single process cycle of fracture would be identified by measuring the area of monotonic true stress-strain curve and each hysteresis loop during the cycle.

Since energy is scalar that can be cumulative, many scholars have shown interest in the energy based fatigue life prediction. In the fatigue process, the life prediction is calculated by dividing the total monotone strain energy by the strain energy each cycle [25, 26]. Sarihan [27] presented a new damage equation of solder depending on hysteresis energy to assess the damage and the fatigue life. However, the traditional life prediction models based on energy are under the assumptions that one cycle throughout the fatigue life has the same energy dissipation. Some improved energy models were proposed by considering the aging factors in the calculation [28] or observing the experimental strain energy density of per cycle in cyclic strain energy density expression [29].

This paper compared and summed up the advantages and disadvantages of the physics model from six perspectives by comprehensively analyzing the theoretical support and application of physics methods, as shown in Table 1. It shows that the life prediction methods based on the cumulative fatigue damage and fracture mechanics are more mature, and are widely used to do failure analysis testing of mechanical parts. However, the application of the versatile, energy based method is difficult in application. Although its recent developments have been achieved, there is still more information gathering and study required.

Table 1 Comparison of physics based methods

| Approach | Advantage | Disadvantage | Application |
|---|--|---|--|
| Life prediction based on stress [10, 30, 31] | Simple and easy to analyze; Parameters required are few and easy to get; | Low accuracy; Poor adaptability for the parameters depend on geometry and loading forms, etc.; | The initial design estimation; Long-life components such as the spring shaft and gears and other high-strength materials; |
| Life prediction based on strain [32, 33] | Rich data sources with accumulated experience can be obtained. Direct access to the strain parameters by means of measurement; Able to conduct notched fatigue analysis; | Unable to analyze the gap effect without considering crack. Complex calculations; Insufficient gap analysis; | Whole-life analysis combined with linear elastic fracture mechanics. Fatigue test with strain as control condition; |
| Life prediction based on accumulative fatigue damage [33, 34] | Able to judge the impact of the loads order; Able to express the cyclic stress-strain response; Conducive to fatigue-creep mixture analysis. The impact of actual magnitude and the order of loads has been taken into consideration; | Only the crack initiation is considered. | Situation of high temperature, large strain and high stress concentration; Components with less load cycles, large plasticity, such as low-strength structural steel; Whole-life analysis combined with linear elastic fracture mechanics. |
| Life prediction based on fracture mechanics [17, 33, 35] | Matured and widely used. The mechanism of the crack propagation can be physics interpreted since the crack propagation is taken into consideration; Able to control the initial damage, the examination period and working loads, etc. to ensure safety. | Narrow scope of application. Difficult to measure and estimate the initial crack size; no research on the initial crack; Uneasy to calculate the stress intensity factor of the components with complex geometry; | More used in engineering. Large and important parts and structures, such as aircraft, nuclear reactors; |
| Life prediction based on damage mechanics [33, 36] | Good agreement with the fatigue mechanism in experimental observations; Easy to measure the fatigue damage. Unified representation, and strong universality; | Elastic-plastic fracture mechanics is used since linear-elastic fracture mechanics can hardly meet requirements in general cases. Complex calculation and analysis; Insufficient research on damage mechanics of some materials and components. | Metal materials with metallurgical defects in themselves and components with pores, slag and weld defects created in welding. |
| Life prediction based on energy [26, 29, 32] | Allowed to take the mean stress and the impact of multi-directional load into account. | Insufficient research; Few applications. | Only applied for some metal materials, composite materials and concrete materials Composite panels of composite tissue, such as alloy laminate, coating structure; Piezoelectric material and the composites. |

For the past few years, further experimental studies on physics mechanics methods for life prediction have been reported. The improved methods feature a deeper understanding of the knowledge as well as an advanced fatigue theory and integrated modeling method. The experimental data is expected to be more accurate, meaningful for the information of real situations. Special research results are obtained on important structures or components, which has been greatly strengthened the study and achieved many surprising results. However, modeling based on physics mechanics and dynamics becomes more and more intractability because of the highly complexity of the life prediction theory. On one hand, there is a high demand for insight into the dynamic details; on the other hand, the error of model prediction increases with the enhancement of model nonlinearity and complexity. Hence, the life prediction based on physics models and engineering principle is difficult to satisfy the life study for remanufacturing in spatial state.

2.2 Data Driven Based Methods

As social modernization advances so does the complexity and uncertainty of modern machinery. Products that include industrial equipment, manufacturing equipment, and aerospace equipment, are expanding, which makes it difficult to determine the physics model. Chai academician pointed out that complex industrial processes often have integrated and complex features, such as multiple variables, strong coupling/correlation, nonlinearity, frequent boundary-condition change, and dynamic characteristics under variable working conditions [37]. The mathematical models are hardly described by a closed analytic form, thus a data driven based method is considered as a feasible method to figure out the complex system controls, decision making, optimization, etc. This paper summarizes the research within recent years on this aspect. The data driven approach is classified mainly into probabilistic methods, artificial intelligence methods, and stochastic methods.

2.2.1 Probabilistic Methods

Probability statistical mode-based analytical approach on life and reliability is a traditional reliability theory. Life cycle information is taken as the object for statistical analysis. The researchers first obtained failure data from plenty of tests and then use statistical criteria to choose the best fit statistical distribution to get distribution of lifetime. The life of mechanical parts was predicted by probabilistic method [38] in 1924. On the basis of test results, an acceptable life was proposed when the sample has 10% failure rate or 90% survival rate. Probability Miner Cumulative Damage theory was developed on the basis of the original Miner formula and the secure life

by introducing P-S-N curve instead of S-N curve. To improve the analysis accuracy, Goode et al. [38] pointed out that the fault process of many equipments (such as hot rolling mill pump) can be divided into two processes, stationary or non-stationary, which can be distinguished by statistical control method. Later on, autoregressive moving average model (ARMA) and autoregressive integrated moving average model (ARIMA) were extensively used in time series modeling and forecasting.

The performance of the device using the statistical pattern recognition (SPR) method was assessed [39, 40], which predicted the remaining life of the processing equipment's cutting tools using ARMA model. Cui et al. [40] introduced a genetic algorithm (GA) to optimize ARMA model, which resolved the RUL prediction accurately problem of aerogenerator.

However, the above-mentioned methods are directly using the experimental data to statistically predict trends that follow probability distribution under certain conditions. Logistic regression and the proportional hazards model are more acceptable which regarded the obtained information as the conditional probability. Yan et al. [41] suggested to apply logistic regression to build the relationships between characteristic variables of the device and the failure probability. At the same time, it recommended to use ARMA model for forecasting the characteristic variables and the trained logistic model for RUL prediction. Chen et al. [42] presented a reliability assessment method based on logistic model, which was able to accurately estimate the reliability indicators and failure time of running tools. Volk et al. [43, 44] put forward the RUL prediction by proportional intensity model, which could evaluate the effect of preventive maintenance as well. Jardine et al. [45] used the proportional hazards model (PHM) to predict equipment's reliability function and the residual life. Gorgian et al. [46] reviewed two RUL assessment method, PHM and Proportional Covariate Model (PCM), which based covariate hazard models using both condition monitoring data and failure data. From the results showed in this paper, two methods above have achieved considerable prediction precision. Besides, PCM always performed better because the posterior hazard function could be updated with current conditional data.

In early 1990s, Vapnik [47] put forward Support vector machines (SVM) and the concept of VC (Vapnik-Chervonenkis) dimension. These can deal with small sample prediction effectively and are widely adopted in life prediction research [48–50]. But the application of SVM is basically aimed at univariate time series. Chen et al. [51, 52] proposed multivariable support vector machine while taking into account the multiple factors affecting the residual life. The extractions of as much

information as possible that are related to the RUL verified the validity of the method by experiments.

The life prediction method that is based on probabilistic model is simple, but does neither consider the difference of the operating conditions and failure mechanisms, nor does it make full use of relevant operational information before failure. In addition, this probability statistics-based life analysis and prediction method should be supported by large amounts of experimental data, without emphasizing individual nuances. Therefore, the results of life prediction are the “average nature” for a given circumstances. Despite the life analysis and forecasting method can accumulate large amounts of data efficiently to calculate the product’s life distribution, it is difficult to meet the needs of product operational life research in spatial state and dynamical working environment.

2.2.2 Artificial Intelligence Methods

The main purpose of artificial intelligence technology is to make the computer accomplish what only people are able to do (such as reasoning, understanding, decision-making, and learning). As a branch of artificial intelligence, expert systems resort to knowledge and inference procedures to solve complex problems that require a large number of experts to work out. In 1985, the aircraft control system monitor developed by Regenie, the automotive fault diagnostic system (FIXER) developed by Malin in 1987 and the aircraft fault diagnosis expert system (Fault2finder) explored by the Development Research Center of NASA. Langleg, etc. had all reached a level of practical application and put into use. In Pawar’s work [53], fuzzy logic system is adopted to establish composite helicopter rotor mechanical damage and global online prediction of lifetime consumption. This worked well to process noisy data. Neural network with a strong ability of approaching non-linear mapping can effectively present the relationship between real time trends of the actual working conditions of machinery and its state signals. Meantime, in terms of the ability to perform multi-step prediction with multi-factor inputs, artificial neural networks are considered as a quite promising approach in prediction. Subsequently, the application of the artificial neural network in lifetime prediction has rapidly developed. Du et al. built a RUL prediction model of large-scale mechanical systems based on artificial neural networks [35].

In recent years, artificial neural network is in-depth developed towards simulating the human cognition. Combined with fuzzy systems, genetic algorithms, evolutionary mechanism, it has become an important direction of artificial intelligence. From perspective of fault prediction, Zhang et al. [54] explored a combining

prediction model based on neural network, and the best combination of traditional prediction method was given. Vachtsevanos et al. [55] proposed a prediction technique based on dynamic wavelet neural network, and verified the practicability of the method in the rolling bearing case. Fu et al. [56] used an improved neural network method to predict the RUL of coke drum thermal mechanical, designed a corresponding program, and proved its validity in the practical application. Xu et al. [57] used genetic algorithms to train Back-Propagation neural network, developed a learning algorithm of evolutionary neural network, and predicted the residual life of cutting tools with higher accuracy. Study of artificial neural network has received much attention by the developed countries in the world.

It should be noted there are some critical points when applying neural network in engineering applications: (1) the neural network is basically a ‘black box’ method, there is no analytical formulas that can make a full explanation for the neural network learning; (2) because the neural network design is very flexible, there are many variable parameters to determine and there is no general rule to follow in design and parameter selection of a neural network; (3) neural network technology has its featured over-learning or over-fitting problem in the training, which results too much noise components in model and significantly limited the model capability of generalization and prediction; (4) neural network training may require considerable time; and (5) data preparation for a neural network application is a huge and tedious work. There is still not a complete and systematic way in data preparation for neural network applications.

2.2.3 Stochastic Methods

Condition monitoring data-based RUL modeling and forecasting is defined as the time from the current to the time when a system failure occurs or the time of arrival for predefined system failure. Banjevic et al. [58] presented RUL as: $X_t = \{x_t : T - t | T > t, Z(t)\}$, where t is the current time, T represents lifetime, which is a random variable and vector $Z(t)$ are the available observations at time t . Engel et al. [59] used a polynomial model to predict the RUL of the helicopter gearbox and derived characteristic variables. In addition, he indicated that an effective life forecasting method not only needed the expected value of RUL, but uncertainty to be predicted (expressed in confidence intervals). In the turn of the century, Hidden Markov Model (HMM) CBM method was first proposed [60], and study the remaining life estimation, finally applied to verify in a helicopter gearbox. HMM is a model based on Markov processes, mainly for modeling implicit discrete degradation states [61]. In HMM method, the state of the hidden process cannot

be observed directly, but it can infer by the observation sequences, each displays every states by probability density distribution, and each observation vector is generated by the state of the corresponding probability density distribution. Baruah et al. [62, 63] recommended an independent HMMs to be established for per health state in the case of equipment failure, and then the state change points of every sample were built. Finally the condition distribution of state change points were obtained. Camci et al. [64, 65] considered hierarchical HMM as dynamic Bayesian network in diagnosis and forecast method. The methods in the research [62, 64] can be used directly to assess the transition probability for device health status. A direct transition probability estimation method between the health status of the device was propose [45, 64, 65], and hierarchical HMM used for diagnosis and prediction was regarded as a dynamic Bayesian network. But, the calculation of the method is very expensive and only the numerical results of the assessment of RUL are available.

In order to relax the requirements of the state to obey the demands of the Markov property, an improved HMM algorithm—Hidden semi-Markov model (HSMM) was proposed [66], to estimate the remaining life of equipment. The method takes multi-source sensor monitoring data, assuming the equipment health status holding time as Gaussian distribution. Further, Dong et al. [67] pointed out the disadvantages of traditional HMM, which requires clearly portrayed interval length for each state. As the expansion of HMM, HSMM overcome the limitations of a constant probability of changing state in HMM modeling by the presumption of Markov property, has resulted in better modeling and analysis of complexity of the hidden process capabilities than HMM [68–70]. The implementation of HSMM in the pump experiments turned out to obtain better recognition rate than the traditional HMM. On this basis, Peng et al. took a further consideration of correlations between the state interval length and the aging of the device, and they introduced the aging factor into HSMM to characterize interval length distribution of each state at different time. This method proved more accurate in predicting the RUL of the pump [71]. Recently, there have been a number of other extensions [72] to enhance HSMM modeling capabilities for the status of the equipment and improve the accuracy of the estimation [73], for practical issues such as the status of the dwell time, has established a reasonable HSMM and introduced a fast recursive algorithm, highly reducing the computational complexity and storage space. Ma et al. [74] based on the understanding of the continuity of state identification and remaining life prediction, has developed a framework of RUL prediction using HSMM. The frame work has been modeled

and validated by computer simulation. And the technology is approved being a valid methodology on remaining life prediction based on system performance gradation model.

In recent years, with the boom of application of SSM in economic time series analysis, the SSM is widely applied in residual life and reliability assessment of mechanical products [75]. The establishment of the two equations (system transition equation and transformation equation), in dynamic system, SSM provides a consistent modeling framework to fully describes the motion characteristics. SSM with good dynamic analysis, Zheng et al. [76] corrected its matrixes and obtained a steady-state correction model which contributed to solving the SSM test and correction problems of aero-engine. Hong et al. [77] proposed a RUL prediction model for main bearing state of the aircraft engines. The model combined the condition monitoring data and analysis, and the calculation is based on theoretical model, providing one new approach for the assessment of the RUL of the main bearing in aero-engine application. But there is a great lag in the study of state recursive evaluation of SSM and the Kalman filtering method is still in the level of linear model for a considerable period of time, for which the optimal estimation is based on the assumption of linear time-invariant system with Gaussian noise. Particle filtering techniques, a Markov Chain Monte Carlo (MCMC) method, demonstrates the superiority in nonlinear non-Gaussian systems, determining its wide range of applications. But the particle degradation problem in the recursive algorithm has prevented its application and the development till 1990s. Since Gordon proposed the recursive re-sampling process [78], particle filtering theory has been greatly developed. Freitas [79] proposed Rao-Blackwellised particle filter and verified its validity in nonlinear systems fault diagnosis through simulation. The paper [80] put forward fault detection and isolation algorithm according to Sequence Importance Resampling (SIR) particle filter likelihood function. Multiple fault subsets were divided in fault system by the algorithm, each of which was estimated by particle filtering, and all of which were running in parallel. This proposed algorithm is simple and suitable for general nonlinear systems. Recently, Jin et al. [81] proposed a Bayesian framework for RUL prediction of secondary batteries, which utilized history data to model population degradation and on-line data to model individual degradation. For joint estimation of state and model parameter, particle filter was employed to update the posterior degradation model and state iteratively. Wang et al. [82] constructed an SSM for Lithium-Ion batteries capacity and a spherical cubature particle filter was adopted to assess the degradation of battery capacities.

It is the difference from the physics model in the sense that data driven model is not obtained by the mechanism or experiment, but usually through statistical or stochastic models. In the study of life prediction based on data driven methods, the model still cannot leave physics model in essence, thus consideration of the hybrid of model is required. However, currently there are only few researches focusing on this aspect. The comparative analysis of various life prediction methods based on data-driven is shown in Table 2. The pros and cons of each method and the corresponding application range are also listed.

The establishment of life predication model of mechanical products by SSM is a hot research topic in recent years. In the SSM, the existing methods have limitations in state estimation and model parameter estimation (refer to Table 2), which greatly reduce the prediction accuracy. There are still many technical issues to be resolved when applying SSM to RUL evaluation. For nonlinear and non-Gaussian SSM, both the theoretical and applied algorithms are still in their infancy.

3 State Space Method

3.1 State Space Model (SSM)

It is a degradation process of mechanical equipment from normal to fault. Since usually the real states of degradation are unobservable and can only be perceived by the observation variables, so the HMM is used to identify potential degradation states. HMM is developed based on the first order Markov chain property, that is degradation state of future only depends on the current degradation state, as $p(x_{n+1} = x | x_0, x_1, x_2, \dots, x_n) = p(x_{n+1} = x | x_n)$, which is often called memoryless, and the state of the system can be displayed directly by the observation condition monitoring (CM) information. On the basis of above property, SSM, which is introduced in modern control theory, has been widely used in HMM and Bayesian recursive estimation. SSM was first presented by Akaike [85] and further developed by Mehra [98]. SSM provides a consistent analysis framework for dealing with practical problems and adopts a similar model structure.

The ideas of creating SSM are: (1) the concept of state variables without observable is introduced; (2) the model describing the state change, called system state equation (translation equation) is established; (3) the observation equation (transformation equation), contains state transformation from state to observation, is determined. One SSM consists of two equations: the state (translation) equation and the observation (transformation) equation. The state equation represents the relation between the state of next and the current, meanwhile, the observation equation shows the internal relation between the

observed values and the state. The definitions of SSM by mathematical induction definitions of SSM are shown as follows.

Definition 1: state vector \mathbf{x}_t , the state variables at time t , usually random and unobservable, is used to represent the inherent characteristics of the dynamic system. \mathbf{x}_t belongs to an N -dimensional Euclidean space, which means $\mathbf{x}_t \in \mathbf{R}^{N_x}$, where \mathbf{R}^{N_x} is the state space. Normally, the sequence of states at $t=1, 2, \dots, s$ is written by $\mathbf{x}_{1:s}$ and $\mathbf{x}_{1:s} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_s\}$

Definition 2: Observation vector \mathbf{y}_t , which could be observed in dynamic systems, is random also because of noise, $\mathbf{y}_t \in \mathbf{R}^{N_y}$. The observation sequence at $t=1, 2, \dots, n$ is denoted by $\mathbf{y}_{1:n}$ and $\mathbf{y}_{1:n} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$.

Assuming the state of the system be regarded as a first-order Markov process, in which observations are independent of each other. The model of the system (may be a nonlinear and non-Gaussian system) would be expressed as

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}, \mathbf{u}_t, \boldsymbol{\varepsilon}_t, \boldsymbol{\theta}), \quad t \in T, \tag{1}$$

$$\mathbf{y}_t = h(\mathbf{x}_t, \mathbf{u}_t, \boldsymbol{\eta}_t, \boldsymbol{\theta}), \quad t \in T, \tag{2}$$

where Eqs. (1) and (2) are the state and observation equations respectively. $\mathbf{x}_t \in \mathbf{R}^{N_x}$ is the system state vector, $\mathbf{y}_t \in \mathbf{R}^{N_y}$ is observation vector, $\mathbf{u}_t \in \mathbf{R}^{N_x}$ is the input vector, and $\boldsymbol{\theta}$ is the static parameter. $\boldsymbol{\eta}_t (\boldsymbol{\eta}_t \in \mathbf{R}^{N_y})$ and $\boldsymbol{\varepsilon}_t (\boldsymbol{\varepsilon}_t \in \mathbf{R}^{N_x})$ denote the observation noise and states noise respective, they are independent with each other. $f: \mathbf{R}^{N_x} \times \mathbf{R}^{N_\varepsilon} \mapsto \mathbf{R}^{N_x}$ is the state function and $h: \mathbf{R}^{N_x} \times \mathbf{R}^{N_\eta} \mapsto \mathbf{R}^{N_y}$ is the observation function. Assuming these two functions have known and depend on $\boldsymbol{\mu}_t$ (sometime it is omitted for simplicity). Prior distribution of the initial state \mathbf{x}_0 is assumed to be $P(\mathbf{x}_0)$.

3.2 Model Coefficient and State Estimate

3.2.1 Model Coefficient Estimation

Model parameter $\boldsymbol{\theta}$ estimation is needed in establishing the SSM for an application. Because of the asymptotic optimality, maximum likelihood estimation (MLE) and maximum posterior likelihood estimation are extensively used methods for parameter estimation. However, it's hard to obtain the gradient of the likelihood function in some cases, so it is necessary to seek other solutions to meet the requirements. The expectation maximization (EM) algorithm is the most popular method when the potential variables exist. Dempster et al. [99] first raised the EM algorithm, which could transform a more complicated optimization issue of likelihood function

Table 2 Comparison of data-driven stochastic model

| Approach | Advantage | Disadvantage | Application |
|---|---|--|---|
| Life prediction based on probability theory [32, 34] | Simple and easy to analyze; Few parameters required and easy to get; Able to obtain confidence limits. | Too much accumulated experience data required; Only a general estimation without regard to the specific cases of working loads and failure; Inaccuracy prediction in terms of the individual. | The initial design estimation; Less important, small-sized, mass-produced components and parts. |
| Life prediction based on ARMA or ARIMA [81, 83] | Simple to model, methods mature, and widely used. No need for historical data or fully understanding the fault mechanism; High efficiency in computing and allowed to be used online. | Unable to be used for long-term prediction; Unable to synthesize prior knowledge; Sensitive to noise and the initial state; poor effects on the dynamic process of non-stationary. | Suitable for linear-time, invariant systems when the performance of its characteristics are smoothly changing. |
| Life prediction based on Logistic Regression (LR) [84–86] | No need to set the critical threshold; | Not feasible unless normal feature domain description and unpredictable behavior are both available; Unable to do real-time prediction with offline data. | Service performance evaluation for degradation failure equipment like machine tools. |
| Life prediction based on proportional hazards [87, 88] | Able to obtain the time dynamic characteristics without making too much assumptions for equipment failure process and the distribution function. Variety of status information are integrated into account; | Complex computation of parameter estimation involving numerical integration; Assumption that risks change with variables proportionally. | Risk analysis of the joint characteristic variables; Analysis of data distribution, residual distribution unknown or censored data. |
| Life prediction based on Neural Network [32] | The characteristic index able to change with time. No need for fault history data; Strong self-learning ability, non-linear fitting ability and good robustness; Excellent memory ability and nonlinear mapping skill. | No standard method to determine the structure of the neural network; Sufficient sample data required for the parameter calibration; Uncertainty of structure and weight. Unable to model the unknown fault; | Pattern recognition or some certain signal processing. |
| Life prediction based on Bayesian Networks [89, 90] | Able to deal with uncertain problems. Able to avoid data over-fitting. Effective multi-source information fusion and expression; Able to visualize the dependency links among each pair of variables. | Difficulty in computing for the unknown network; Relies on reliability of the prior knowledge; Results are sensitive to the selection of prior distribution. | Used to solve uncertainty and relevance fault of complex equipment; Widely used in the intelligent systems of computer intelligence science, industrial control, medical diagnostics and other fields. |

Table 2 (continued)

| Approach | Advantage | Disadvantage | Application |
|--|---|---|---|
| Life prediction based on Support Vector Machine (SVM) [50, 91] | Concrete realization of the structural risk minimization criterion | Not suitable for large-scale data processing; | Small sample, nonlinear and high dimensional mode recognition, such as predicting porosity and clay content in oil well logging. |
| Life prediction based on Markov Model/Hidden Markov Model (HMM) [60, 61, 92, 93] | Simple structure, good promotion performance and fast learning; Only one minimal in optimization solution. Able to model different stages of degradation; | Theoretical defects in kernel function for nonlinear classification problems; Slow solution speed. Unable to model the unknown fault; | Unable to directly observe the state, but able to observe the vector sequence; Widely used in pattern recognition, speech recognition, behavior recognition, character recognition and fault diagnosis and other fields. |
| State-space prediction based on Kalman Filter [93–95] | No need for prior knowledge of fault mechanism, able to process incomplete data; Predictions with confidence intervals. Able to do the linear unbiased minimum mean square error estimation; | Complex hidden semi-Markov model is required if the fault time is not exponentially distributed; The model calculation increases with the number of states. State equation and measurement equation need to be defined; | Real-time online prediction for linear Gaussian system. |
| State-space prediction based on particle filter [93, 96–98] | Prediction accuracy does not change with the prediction time interval; good robustness; Able to estimate the current state and also predict the future state. By finding a random sample to carry out the approximate calculation for the probability density function to obtain the state minimum variance distribution; High accuracy. | Noise level could influence the performance and stability of the algorithm; Only works with linear system and Gaussian noise. Large amount of sampling are required to avoid degradation; | Real-time online prediction of strong-nonlinear, non-Gaussian noise system, such as radar tracking and robot localization. |
| | | Computation is more complex than Kalman filtering; High-dimensional system and increasing particles complex computation. | |

into a series of relatively simple functions through data enhancement technique. It provides a framework for MLE of unknown parameters with incomplete data (missing data points). EM algorithm is an iterative algorithm and each iteration method includes two steps: the E-step (expectation step) and the M-step (maximization step) [100]. EM algorithm has aroused great interest of statisticians over the last decade and widely used in the field of statistics.

EM algorithm can be summarized as follows.

Initialized θ , and set as θ^0 .

Given the current estimates of the unknown parameter θ^{k-1} , calculate the expectations of logarithmic likelihood function with full data.

$$\begin{aligned} Q(\theta, \theta^{k-1}) &= E_{\theta^{k-1}} [\log p(\mathbf{x}_{1:t}, \mathbf{y}_{1:t} | \theta)] \\ &= \int \log p(\mathbf{x}_{1:t}, \mathbf{y}_{1:t} | \theta) p(\mathbf{x}_{1:t} | \mathbf{y}_{1:t}, \theta^{k-1}) d\mathbf{x}_{1:t} \\ &= \int \log p(\mathbf{x}_{1:t}, \mathbf{y}_{1:t} | \theta) \frac{p(\mathbf{x}_{1:t}, \mathbf{y}_{1:t} | \theta)}{p(\mathbf{y}_{1:t} | \theta^{k-1})} d\mathbf{x}_{1:t}. \end{aligned}$$

Maximize the expectations to update θ :

$$\begin{aligned} \theta^k &= \arg \max_{\theta} Q(\theta, \theta^{k-1}) \\ &= \arg \max_{\theta} \int \log p(\mathbf{x}_{1:t}, \mathbf{y}_{1:t} | \theta) p(\mathbf{x}_{1:t} | \mathbf{y}_{1:t}, \theta^{k-1}) d\mathbf{x}_{1:t} \\ &= \arg \max_{\theta} \int \log p(\mathbf{x}_{1:t}, \mathbf{y}_{1:t} | \theta) \frac{p(\mathbf{x}_{1:t}, \mathbf{y}_{1:t} | \theta)}{p(\mathbf{y}_{1:t} | \theta^{k-1})} d\mathbf{x}_{1:t} \\ &= \arg \max_{\theta} \int \log p(\mathbf{x}_{1:t}, \mathbf{y}_{1:t} | \theta) p(\mathbf{x}_{1:t}, \mathbf{y}_{1:t} | \theta) d\mathbf{x}_{1:t}. \end{aligned}$$

Iterate the above steps until $\|\theta^k - \theta^{k-1}\|$ is sufficiently small and the iteration stops.

Because $p(\mathbf{y}_{1:t} | \theta^{k-1})$ is independent with the parameters being estimated, therefore it can be omitted for simplification without affecting the optimization results.

3.2.2 State Estimation

State estimation is divided into three classes on the basis of the estimation output \mathbf{x}_t and the given data as prediction, filtering, and smoothing. In the state sequence $\mathbf{x}_{1:s}$, s is the current time point. (1) Prediction (when $t > s$), uses available observations to predict the future states. (2) Filtering (when $t = s$), make use of the obtained information to express the current status. (3) Smoothing (when $t < s$), put to use the available information to smooth the past state. The principle of Bayesian filtering is to build the posterior probability density of the state variables by the

known prior statistical information, which belongs to a statistical inference method [101].

The purpose of establishing SSM is that it provides a convenient method to recursively estimate the state \mathbf{x}_t (recursive Bayesian estimation). The estimation of the current state variables \mathbf{x}_t is performed by observed information $\mathbf{y}_{1:t}$ through the state transition equation. The posterior probability density $p(\mathbf{x}_t | \mathbf{y}_{1:t})$ will be recalculated once new observations come, so the calculation will be very expensive. However, the basic estimation theory of state, recursive Bayesian filter, was introduced to update the estimation recursively in dynamic system. The recursive updating will calculate the update by only using one prior step information in each step, avoiding mass storage of historical data and heavy calculations. The state vector of discrete time, time-varying parameters would be updated by recursive Bayesian estimation method in SSM of nonlinear dynamic system, i.e. the estimated current state vector \mathbf{x}_t uses the given current observation \mathbf{y}_t , previous state vectors \mathbf{x}_{t-1} have no connection with all previous states $\mathbf{x}_{1:t-1}$ (Markov property of memoryless). In general, this estimate is carried out in two steps.

Predictions: the prior probability density of current system state is estimated by using the state equation. Due to the influence of noise, there is a certain deviation between the prior probability density and the actual value.

Updated: at time $t + 1$, the new observation information \mathbf{y}_{t+1} was incorporated into the estimation of the state vector, and the prior probability density was corrected. The posterior probability density of the state was obtained by Bayesian formula. The derivation formula of the conditional probability density function is as follows. Initial value:

$$p(\mathbf{x}_0 | \mathbf{y}_0) = p(\mathbf{x}_0).$$

Prediction:

$$p(\mathbf{x}_t | \mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} | \mathbf{y}_{1:t-1}) d\mathbf{x}_{t-1}.$$

Update:

$$\begin{aligned} p(\mathbf{x}_t | \mathbf{y}_{1:t}) &= \frac{p(\mathbf{y}_t, \mathbf{y}_{1:t-1} | \mathbf{x}_t) p(\mathbf{x}_t)}{p(\mathbf{y}_t, \mathbf{y}_{1:t-1})} \\ &= \frac{p(\mathbf{y}_t | \mathbf{y}_{1:t-1}, \mathbf{x}_t) p(\mathbf{y}_{1:t-1} | \mathbf{x}_t) p(\mathbf{x}_t)}{p(\mathbf{y}_t | \mathbf{y}_{1:t-1}) p(\mathbf{y}_{1:t-1})} \quad (3) \\ &= \frac{p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{y}_{1:t-1})}{p(\mathbf{y}_t | \mathbf{y}_{1:t-1})}, \end{aligned}$$

where $p(\mathbf{y}_t | \mathbf{y}_{1:t-1}) = \int p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{y}_{1:t-1}) d\mathbf{x}_t$ is called as normalized factor.

The Chapman–Kolmogorov equation

$$p(\mathbf{x}_t | \mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} | \mathbf{y}_{1:t-1}) d\mathbf{x}_{t-1}$$

offers a relation between the prior density $p(\mathbf{x}_t | \mathbf{y}_{1:t-1})$ and the previous posterior density $p(\mathbf{x}_{t-1} | \mathbf{y}_{1:t-1})$. Therefore that updating Eq. (3) establish the recursive relation for predication. The posterior probability density is presented by $p(\mathbf{x}_{0:t} | \mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_{1:t} | \mathbf{x}_{0:t}) p(\mathbf{x}_{0:t})}{p(\mathbf{y}_{1:t})}$, and with all observations are independent with each other, $p(\mathbf{y}_{1:t} | \mathbf{x}_{0:t}) = \prod_{i=1}^t p(\mathbf{y}_i | \mathbf{x}_{0:t})$. Assume that the observed values are independent with all other states at other moments $p(\mathbf{y}_{1:t} | \mathbf{x}_{0:t}) = \prod_{i=1}^t p(\mathbf{y}_i | \mathbf{x}_i)$, the posterior probability density is gotten as

$$p(\mathbf{x}_{0:t} | \mathbf{y}_{1:t}) = p(\mathbf{x}_0) \frac{\prod_{i=1}^t p(\mathbf{y}_i | \mathbf{x}_i) p(\mathbf{x}_i | \mathbf{x}_{i-1})}{p(\mathbf{y}_{1:t})}, \tag{4}$$

where $p(\mathbf{x}_{0:t}) = p(\mathbf{x}_0) \prod_{i=1}^t p(\mathbf{x}_i | \mathbf{x}_{i-1})$ because the system follows Markov process.

Thus the recursive formula for joint probability density is $p(\mathbf{x}_{0:t} | \mathbf{y}_{1:t}) = p(\mathbf{x}_{0:t-1} | \mathbf{y}_{1:t-1}) \frac{p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{x}_{t-1})}{p(\mathbf{y}_t | \mathbf{y}_{1:t-1})}$.

Figure 1 depicts the recursive estimation of Bayesian posterior density procedure.

As the uncertainty of the probability density functions' form, it is difficult to describe the probability density functions in a closed form of analysis. Therefore, in practical application, it is hard to get an accurate solution by analytical method. Particle filter, as an approximation algorithms of Bayesian filtering, is on the basis of Monte Carlo simulation. The recursive Bayesian filtering is carried out by non-parametric Monte Carlo simulation, and the posterior probability of non-Gaussian, nonlinear and high-dimensional system is computed effectively from the observations. The points of the idea is that utilizing random independent samples, called particles, to approximate the probability density $p(\mathbf{x}_k | \mathbf{y}_k)$ with weights. The mean of the samples can be used for an integral operation to achieve the

minimum variance estimation of the state. Since each sample (particle) on behalf of the possible state, the probability density function of particles is gradually tend to be the real state with the increasing number of particles, which may get the optimal Bayesian estimation results.

Mathematical description of particle filter is that it is assumed the system's posterior probability density at time $k-1$ is $p(\mathbf{x}_{k-1} | \mathbf{y}_{k-1})$ in the stationary random process. According to a certain principle, n random samples are selected from the density. When \mathbf{y}_k at time k is observed, the posterior probability density $p(\mathbf{x}_k | \mathbf{y}_k)$ of n particles will be updated. Two phase, state estimation and correction phase, are classified in the process of particle filtering. In state estimation phase, numerous samples will be selected first according to \mathbf{x}_{t-1} and its probability distribution. These samples distributed in the state space are actually the probability distribution of the state \mathbf{x}_{t-1} . According to the state translation equation in SSM, the state of every particle would be transformed, and the original particles become predicted ones. The weights of particles are updated in correction phase by noise processing and the weight normalization through substituting the predicted particles into the observation equation. The particle weights can be expressed as

$$w_t^{(i)} = \frac{p(\mathbf{y}_{1:t} | \mathbf{x}_{0:t}^{(i)}) p(\mathbf{x}_{0:t}^{(i)})}{q(\mathbf{x}_{0:t}^{(i)} | \mathbf{y}_{1:t})}. \tag{5}$$

And the recursive formula of the particle weights is gotten by

$$\begin{aligned} w_t^{(i)} &= \frac{p(\mathbf{y}_{1:t} | \mathbf{x}_{0:t}^{(i)}) p(\mathbf{x}_{0:t}^{(i)})}{q(\mathbf{x}_t^{(i)} | \mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t}) q(\mathbf{x}_{0:t-1}^{(i)} | \mathbf{y}_{1:t})} \\ &= \frac{p(\mathbf{y}_{1:t} | \mathbf{x}_{0:t}^{(i)}) p(\mathbf{x}_{0:t}^{(i)})}{p(\mathbf{y}_{1:t-1} | \mathbf{x}_{0:t-1}^{(i)}) p(\mathbf{x}_{0:t-1}^{(i)}) q(\mathbf{x}_t^{(i)} | \mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})} w_{t-1}^{(i)} \\ &= \frac{p(\mathbf{y}_t | \mathbf{x}_t^{(i)}) p(\mathbf{x}_t^{(i)} | \mathbf{x}_{t-1}^{(i)})}{q(\mathbf{x}_t^{(i)} | \mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})} w_{t-1}^{(i)}. \end{aligned} \tag{6}$$

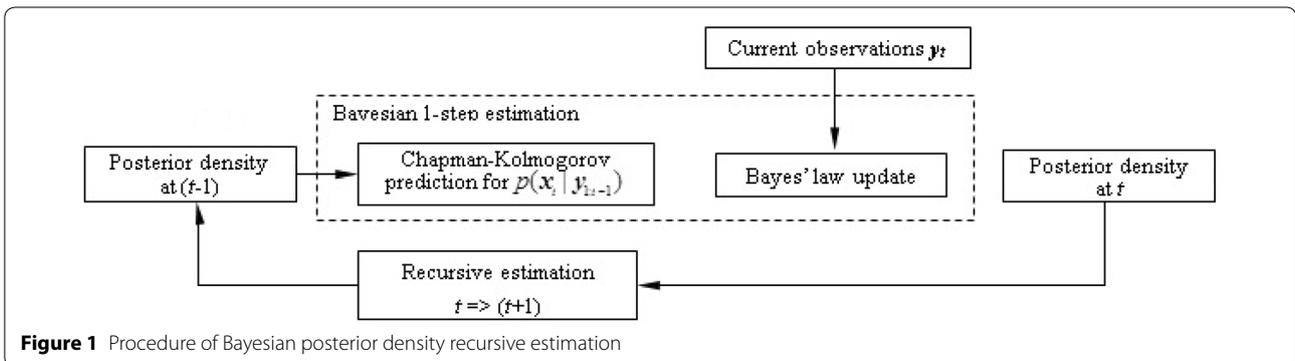


Figure 1 Procedure of Bayesian posterior density recursive estimation

In the 1950s, Zhao et al. [100] proposed sequential importance sampling (SIS) algorithm, which was successfully used in the application of system control in the late 1960s. SIS method stimulates the wide research for particle filtering as a practical application of nonlinear and non-Gaussian filtering.

The SIS algorithm can be explained as follows in general:

1. Initial values at $t=0$. Sampling particles $\{\mathbf{x}_0^{(i)}\}_{i=1}^N$ from prior distribution $p(\mathbf{x}_0)$, and the weight of each particles is set as $w_0^{(i)} = \frac{1}{N}$ for $i = 1, 2, \dots, n$.
2. For $t=1, 2, \dots, T$,
 - i. for $i=1, 2, \dots, n$, extracting the particles $\mathbf{x}_t^{(i)} \sim q(\mathbf{x}_t | \mathbf{x}_{0:t-1}^{(i)} \mathbf{y}_{1:t})$ from importance density function $q(\mathbf{x}_t | \mathbf{x}_{0:t-1}^{(i)} \mathbf{y}_{1:t})$, and assuming $\mathbf{x}_{0:t}^{(i)} = (\mathbf{x}_{0:t-1}^{(i)}, \hat{\mathbf{x}}_t^{(i)})$.
 - ii. for $i=1, 2, \dots, n$, updating the weight by $w_t^{(i)} = \frac{p(\mathbf{y}_t | \mathbf{x}_t^{(i)}) p(\mathbf{x}_t^{(i)} | \mathbf{x}_{t-1}^{(i)})}{q(\mathbf{x}_t^{(i)} | \mathbf{x}_{0:t-1}^{(i)} \mathbf{y}_{1:t})} w_{t-1}^{(i)}$.
 - iii. for $i=1, 2, \dots, n$, normalizing the importance weight by $\tilde{w}_t^{(i)} = w_t^{(i)} / \sum_{i=1}^n w_t^{(i)}$.
3. The updated particle $\{\mathbf{x}_{0:t}^{(i)} : i = 1, 2, \dots, n\}$ is to approximate the posterior probability distributions of the state and the expectations of some functions such as

$$\bar{E}[g_t(\mathbf{x}_{0:t})] = \sum_{i=1}^n g_t(\mathbf{x}_{0:t}^{(i)}) \tilde{w}_t^{(i)}.$$

The variance of the particle weights will increase gradually over time with several algorithm iterations. The large variance of the partial weights will cause a few of particles to have larger weights, while most particles with small values, which will lose their functions of particle degrading. The degradation of the particles with iteration process makes the case that there is no enough particles to express the posterior density distribution after new observation available. Until 1993, Gordon proposed the concept of re-sampling [101], which laid a foundation for the particle filter algorithm and its applications. The concept of effective resampling scale is defined as:

$$N_{eff} = \frac{n}{\sum_{i=1}^n (w_t^{(i)})^2}$$

where $w_t^{(i)} : i = 1, 2, \dots, n$ are the un-normalized particles weights.

The formula is hard to determine the actual calculation results, but the approximation method is generally adopted as

$$N_{eff} = \frac{1}{\sum_{i=1}^n (\tilde{w}_t^{(i)})^2}$$

where $w_t^{(i)} : i = 1, 2, \dots, n$ are normalized weights.

If $N_{eff} < N_{th}$, resampling should be used. In general, $N_{th} = 2N/3$.

Now, the most commonly used re-sampling method is introduced as follows: First, generate n random number $\{\mu_l : l=1, 2, \dots, n\}$ in the $[0, 1]$ uniform distribution, and then apply the search method to find an integer number m that satisfies the following formula $\sum_{j=0}^{m-1} \tilde{w}_j \leq \mu_l \leq \sum_{j=0}^m \tilde{w}_j$. The new samples $\mathbf{x}_k^{(m)}$ are joined in the new set of sampling particles. The interval $[0, 1]$ is divided into n intervals by $\lambda_j = \sum_{i=0}^j \tilde{w}_i (i = 1, 2, \dots, n)$.

When the random number μ_l belongs to the m th interval $I_m = [\lambda_{m-1}, \lambda_m]$, the samples can be copied. Obviously, samples with larger weights would be duplicated repeatedly, while in the case of n sample numbers, samples with smaller weights will be discarded. Now the weights of particles are re-distributed to $1/N$, and the re-sampling process is achieved.

Figure 2 shows a schematic diagram of the standard particle filter algorithm with re-sampling function.

The brief procedures for SIS particle filter with resampling:

- i. Initialization: At $t=0$, particles $\{\mathbf{x}_0^{(i)}\}_{i=1}^n$ from prior distribution $p(\mathbf{x}_0)$ are be chosen, and the weight of each particles is $w_0^{(i)} = \frac{1}{N}$ for $i = 1, 2, \dots, n$.
- ii. Importance sampling: extracting the particles from $q(\mathbf{x}_t | \mathbf{x}_{0:t-1}^{(i)} \mathbf{y}_{1:t})$ for $i=1, 2, \dots, n$, that is $\hat{\mathbf{x}}_t^{(i)} \sim q(\mathbf{x}_t | \mathbf{x}_{0:t-1}^{(i)} \mathbf{y}_{1:t})$. Assuming $\hat{\mathbf{x}}_{0:t}^{(i)} = (\mathbf{x}_{0:t-1}^{(i)}, \hat{\mathbf{x}}_t^{(i)})$ for $i = 1, 2, \dots, n$. The weights by $w_t^{(i)} = \frac{p(\mathbf{y}_t | \hat{\mathbf{x}}_t^{(i)}) p(\hat{\mathbf{x}}_t^{(i)} | \mathbf{x}_{t-1}^{(i)})}{q(\hat{\mathbf{x}}_t^{(i)} | \mathbf{x}_{0:t-1}^{(i)} \mathbf{y}_{1:t})} w_{t-1}^{(i)}$ are updated. The importance weight is normalized by $\tilde{w}_t^{(i)} = w_t^{(i)} / \sum_{i=1}^n w_t^{(i)}$.
- iii. Re-sampling: calculating: If $N_{eff} < N_{th}$, re-sampling particles $\{\mathbf{x}_t^{(i)}, \frac{1}{n}\}_{i=1}^n$ from particle set $\{\hat{\mathbf{x}}_t^{(i)}, \tilde{w}_t^{(i)}\}_{i=1}^n$, else $\{\mathbf{x}_t^{(i)}, w_t^{(i)}\}_{i=1}^n = \{\hat{\mathbf{x}}_t^{(i)}, \tilde{w}_t^{(i)}\}_{i=1}^n$.
- iv. The states posterior distribution $p(\mathbf{x}_t | \mathbf{y}_{1:t}) \approx \sum_{i=1}^n \tilde{w}_t^{(i)} \delta(\mathbf{x}_t - \mathbf{x}_t^{(i)})$.

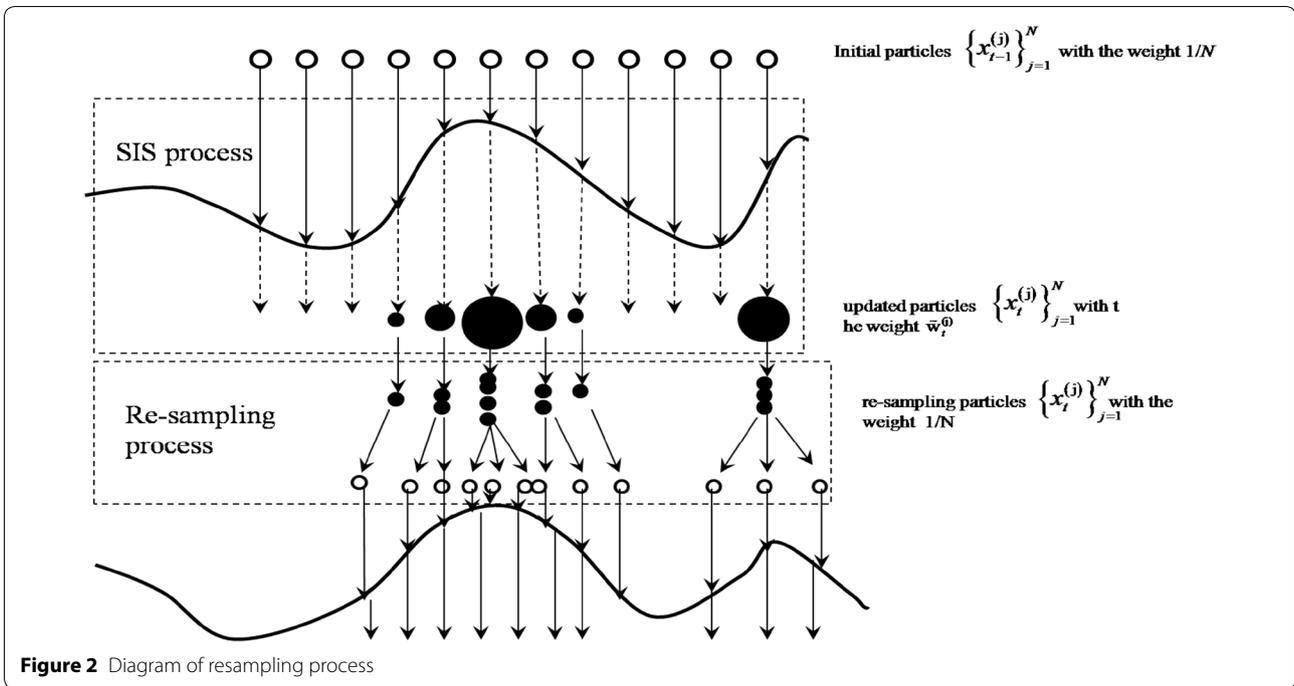


Figure 2 Diagram of resampling process

v. Stop or return for t .

Resampling is not need to be done on every iteration and the condition controlled by if $N_{eff} < N_{th}$.

In spite of particle filter has been extensively used on many occasions, there are still some shortages. In order to further improve the performance and stability of particle filter algorithm, lots of improvements have been made in two aspects, which are selecting more importance function and ensuring the effective particles diversity [102]. There have been many proposed modifications, such as auxiliary particle filter, local linearization methods, evolutionary particle filter algorithm, genetic algorithm operator. The topic of prognostics uncertainty has been highly concerned, and how to reduce the uncertainty is quite a challenging job [103]. In order to make the estimation of states close to the true value, optimal algorithms of particle filter need further study.

3.3 SSM in RUL Assessment—A Note

Much recent research has focused on SSM in RUL assessment, where the products' degradation process is presented using the state in SSM and the degradation modeling is the assessment of the state variables. The prediction of the degradation is the estimation of the future state based on observations available up to current time. The goal of the state estimation in Bayesian estimation is to obtain the posterior distribution of

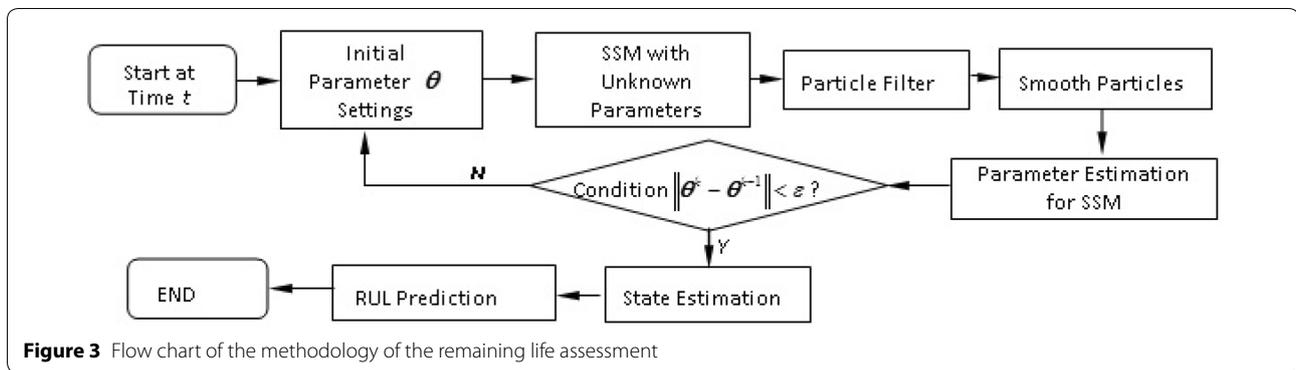
the state, which presents the unobservable degradation process. To be more precisely, the joint posterior distribution $p(\mathbf{x}_{0:t} | \mathbf{y}_{1:t})$ is the key for predicting the degradation and remaining lifetime for the states of the degradation process. On the mechanism of the state estimation above, the particles $\{x_{0:t}^{(i)}\}_{i=1}^n$ are generated in the process of state estimation to predict future degradation. The k -step ahead future state $(\mathbf{x}_{t+1:t+k} | \mathbf{y}_{1:t})$ prediction is given as follows.

For $j = 1, 2, \dots, k$,

For $i = 1, 2, \dots, n$, sampling $\mathbf{x}_{t+k}^{(i)} \sim p(\mathbf{x}_{t+k} | \mathbf{x}_{t+k-1}, \mathbf{y}_{1:t})$ and $\mathbf{x}_{t:t+k}^{(i)} \equiv (\mathbf{x}_{t:t+k-1}, \mathbf{x}_{t+k}^{(i)})$. Finally the sample set $\{\mathbf{x}_{t:t+k}^{(i)}\}_{i=1}^n$ is obtained. The estimation of $p(\mathbf{x}_{t+1:t+k} | \mathbf{y}_{1:t}) = \frac{1}{n} \sum_{i=1}^n \delta_{(\mathbf{x}_{t+1:t+k}^{(i)})}(\mathbf{x}_{t+1:t+k})$, where the $\delta_{(\mathbf{x}_{t+1:t+k}^{(i)})}(\mathbf{x}_{t+1:t+k})$ denotes the Dirac delta mass located at \mathbf{x}_{t+1} .

Once new observed data are obtained, the first is to update the estimation of the state and parameters, then undertake the future degradation prediction. Because of the statistical characteristics of the Bayesian method, the uncertainty of degraded state estimation and prediction is quantified by the probability distribution $p(\mathbf{x}_{t+1:t+k} | \mathbf{y}_{1:t})$. The corresponding predicted degradation is $\hat{\mathbf{x}}_{t+k} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_{t+k}^{(i)}$.

To obtain the prediction of the RUL, the specified threshold value Y , the degradation state should be pre-defined. The remaining lifetime T is determined by the



estimated degradation \hat{x}_{t+k} . In order to derive the distribution of the remaining lifetime, compared Y with each sample path $x_{t,t+k}^{(i)}$, $i = 1, 2, \dots, n$ and obtained n times T_1, \dots, T_n , and the T_i represents the least time at which the sample path $x_{t,t+k}^{(i)}$ equals or exceeds Y [81, 104].

Therefore, the methodology of the RUL assessment using particle filtering method can be summarized as shown in Figure 3.

The state and parameter estimation can be proceed at the same time based on particle filter method [80, 105]. A recent advanced study in joint state and parameter estimation is the method of particle MCMC proposed in Ref. [104], which exploited the complementary strengths of both the MCMC and Sequential Monte Carlo (SMC) algorithms. Particle filtering algorithm of SSM (improved MCMC method) is the most promising method and widely studied for online RUL modeling and evaluation, as expressed the HMM property in SSM and MCMC effectiveness in nonlinear non-Gaussian system. In specific application, SSM of system (the state transition equation and the observation transformation equation) is based on the given degradation dynamics, failure modes and failure mechanisms. System startup conditions can also be resolved by applications. Based on the existing observations and Monte Carlo simulation, the model coefficients are estimated by using EM algorithm. The degradation can be evaluated by the state estimation in space model, and the process of degradation is regarded as the hidden stochastic process. Bayesian recursive estimation of posterior probability density function of system states is carried out by particle filter algorithm with updated observations. This works in a recursive manner of closely simulation. Once the observations are updated, numerical calculation is much expensive at each step. The algorithm accuracy is highly depends on the number of particles. Due to the improved power consumption and calculating speed of the digital computer system, particle filtering in SSM can implement online recursion in RUL applications.

4 RUL Online Assessment through Performance Sensing Data—A Discussion

Online assessment of RUL based on performance observations is an evaluation process to assess the probability distribution of hidden degradation processes approaching predefined threshold (failure). This online estimation, as its first order Markov property, becomes recursive and not only reduces storage space for the data and computing time, but also the model takes into account the system nonlinearity and operating environmental variation. However, there are also many issues about RUL prediction and assessment on the basis of performance observations monitored in real-time.

- (1) Underlying physics degradation of a system or device decides the lifetime is limited. In other words, no concept of RUL exists if there is no non-reversible degradation in aging and usage. In most applications of online RUL assessment, the performance failures are defined by the physics degradation, which reaches or exceeds a predefined threshold. The underlying degradation is a random process which is monotonically increasing (not decreasing) with time. Hence, online RUL assessment must be built on performance degradation. Defining a performance threshold for failure in industrial practice is basically determined through engineers' experience and/or historical maintenance data analysis. Most of the time, the model and assessment are not accurate enough because it assumes that life information of the same products apply to the same working environment. Therefore, defining the degradation threshold for RUL assessment is still an open topic in prognosis and health management (PHM).
- (2) Underlying system degradation and system physics performance are different concepts in RUL assessment engineering. The underlying system degradation is a random process, time-varying, and

undiminished. It is usually material and substantial changing, monotonic, and not reversible. This is the basis for RUL assessment. But the system performance measured using some physics variable(s), which, unlike the degradation could fluctuate showing sometimes high and sometimes low, depending on the devices' working conditions. For example, a secondary battery's performance (charging and discharging and/or output voltage and current) could fluctuate, sensitive to environmental temperature, but the materials consumption and the electrochemical reaction on electrodes and electrolyte are monotonically degrading. In practice, the true physics degradation is difficult to measure on-line. Instead physics variable measures should be used. Almost all observations of sensing data in performance condition monitoring show fluctuations. The rise and fall of the signal are because (1) it is a performance variable; or (2) it is a degradation process with high noise of measurement errors. Usually a signal process will be applied to the signal for noise filtering and/or finding underlying degradation components.

- (3) In industrial practice, to assess observable degradation for RUL, some other measurable physics variables are selected and sensed instead. These variables are highly related to the underlying degradation or sometimes they are the factors causing degradation which directly affect the degrading process. For those factors of the degradation, referred to as covariates of the process, some covariate-based models are proposed. One of the widely reported covariate based models is the PHM, which models hazard rate of a system degradation process in two components: baseline hazard function and a function of covariates. In this model, historical failure and censored data, and CM data are able to be used for RUL assessment without a predefined failure threshold of degradation. For those related online variable measurements, the statistical properties between the degradation and performance observations should be known and able to model. There are at least two stochastic processes recognized and the measurable one will be sensed for observations and modeled to assess the underlying unmeasurable degradation, which leads to a hidden Markov process model.
- (4) A HMM is proposed for degradation modeling. There are two random processes, one is the hidden Markov process, which represents the actual degradation state, and the other is the observable process, which is the performance observation data from monitoring or tests. The state of the system

can only be assessed using updated observations expressed by probability density distributions of observations. Each observation vector is generated by the a sequence of states with a corresponding probability density distribution. The HMM has been extensively studied and applied. However, there is no unified method to determine the direct relationship between observation and state, and there is no uniform way to identify model parameters either. Those are unsolved focuses in RUL research. However, the first order of HMM is naturally connected to a state space model and particle sequence Monte Carlo method.

- (5) A SSM with a state equation (translation function) and an observation equation (transformation function) naturally presents a first order hidden Markov process. It has been studied in depth for a general system, which has high nonlinearity, time-variant parameters, and non-Gaussian noise. A nonlinear SSM with particle filtering (sequence Monte Carlo) method using a recursive estimate of both model parameters and system state will serve as the fundamental methodology for online RUL assessment at a given time. Limitations to linear and constant system dynamics, the optimization, in minimum variance, of Kalman filtering and state estimation are no longer effective in the industrial applications of RUL modeling and assessment. With the rapid growth of computer computation capability, the general SSM plus particle filtering technology might be the key methodology to RUL modeling and online assessment.
- (6) Building a SSM for a specific application in practice is not a trivial thing. There are still no valid and systematic procedures yet. The degradation process is expressed in the state equation and the transformation dynamics from observation to the hidden state is expressed in the observation equation. Currently, the most suggested method is the hybrid one that combines both model based and data driven methods with model parameter online identification.
- (7) Measures in practice always contain measurement errors. This concern exists also in the sensing data observations in RUL assessment. Researchers have understood the concern for a long time, yet it might be the time for them to look for the solution to the problem in the form of SSM since nonlinear stochastic filtering theory and digital computation capabilities have been greatly improved.

In the field of RUL prediction, theories are studied from different aspects, which does not form a system of integration application. As an industry-leading

international business innovation and research and development (R&D) project, IMS (Intelligent Manufacturing Systems) is proposed to develop next-generation manufacturing and processing technologies. Online RUL prediction based on the degradation is especially motivated in emerging applications of automatic prediction system, whose main goal is to evaluate the current and future health states of an individual system in order to make cost-effective decisions about maintenance activities.

5 Conclusions

RUL modeling and assessment is becoming increasingly important in reliability and PHM systems. Management system needs to determine the performance status and the RUL of key systems for maintenance planning, decision making, and system global optimization. This paper reviewed basic methods in RUL modeling and categorized them into three types. They are physics model-based method, data-driven method and a hybrid method of both. Its advantages and disadvantages are also discussed. RUL of all devices is decreasing because of the physics performance degradation. The physics performance degradation is assumed monotonic and nonreversible. And the RUL, a random variable affected by the system aging and its working environment conditions, is modeled and assessed based on the degradation measures. Usually, because of the complexity of the system and failure mechanism, physics degradation is not directly observable. Therefore, RUL evaluation becomes a course for assessing the degradation process based on other available physics performance data. Some variables are causes of the degradation and some variables are indirect measures of the degradation, which are highly related to the underlying degradation. Normally, there are two random processes in RUL evaluation, the hidden degradation process and the observation of measurable processes. The aim is to use the observation process and the data to estimate the degradation process and convert the degradation state into a probability density distribution that forms the mechanism of HMM and Bayesian recursive estimation, which can be best expressed in a state space model.

A SSM with a system equation and an observation equation is actually a format of a (first order) hidden Markov process. Originally, the success of SSM and the Kalman filter in aerospace applications stimulated more industrial application. However, soon enough, the attempts showed the mismatch between the assumption of the Kalman filter and the industrial state estimate problems, which are dynamics with high noise and non-linearity. The SSM, as a first order hidden Markov, is providing the exact format for particle filtering algorithm on

Bayesian posterior distribution estimation. The nonlinear SSM using an online recursive particle sequence Monte Carlo method is a hot topic in online RUL prediction, which covers both online model parameter estimates and state estimates across time. It is a promising methodology, providing online information of RUL for PHM and CBM. However, building an SSM for a particular application is not a trivial thing. There are still no systematic procedures to follow. To model the system equation and the observation equation of the SSM, the suggested way is the combination of the physics model-based method and historical data fitting for parameter initiation—a data driven hybrid method.

Authors' Contributions

HL and HZ was in charge of the whole trial; YH and SL wrote the manuscript; All authors read and approved the final manuscript.

Author Details

¹ School of Mechanical Engineering, Dalian University of Technology, Dalian 116024, China. ² Department of Construction & Operations Management, South Dakota State University, Brookings, SD 57007, USA. ³ Department of Industrial Engineering, Texas Tech University, Lubbock, TX 79409, USA.

Authors' Information

Yawei Hu, born in 1989, is currently a PhD candidate at *Dalian University of Technology, China*.

Shujie Liu, born in 1977, is currently an associate professor at *Dalian University of Technology, China*. She received his PhD degree from the *University of Tokyo, Japan*, in 2007. Her research interests include remaining useful life assessment, precision engineering and sustainable engineering.

Huitian Lu, born in 1948, is currently a professor at *South Dakota State University, USA*. He received his PhD degree from *Texas Tech University, USA*, in 1998. His research interests include reliability estimation, life prediction and dynamic modelling.

Hongchao Zhang, born in 1953, is currently a Thousands of People Plan Distinguished Professor at *Dalian University of Technology, China*, is also professor of the *Department of Industrial Engineering, Texas Tech University, USA*. He received his PhD degree from *Technical University of Denmark, Denmark*, in 1989. His research interests include green manufacturing, remanufacturing and LCA.

Competing Interests

The authors declare no that they have no competing interests.

Funding

Supported by Fundamental Research Funds for the Central Universities of China (Grant No. DUT17GF214).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 13 July 2018 Accepted: 17 December 2018

Published online: 27 February 2019

References

- [1] Yixin Luo. Current status of machinery safety in China and its countermeasure. *China Safety Science Journal*, 2004, 14(5): 92-94.
- [2] ISO 13381-1:2004 Condition monitoring and diagnostics of machines - Prognostics - Part 1: General guidelines. AFNOR, 2005.

- [3] K Medjaaher, D A Tobon-Mejia, N Zerhouni. Remaining useful life estimation of critical components with application to bearings. *IEEE Transactions on Reliability*, 2012, 61(2): 292-302.
- [4] J Lee, J Ni, D Djurdjanovic, et al. Intelligent prognostics tools and e-maintenance. *Computers in Industry*, 2006, 57(6): 476-489.
- [5] Weiyue Wang, Wenjuan Zhang. An asset residual life prediction model based on expert judgments. *European Journal of Operational Research*, 2008, 188(2): 496-505.
- [6] Xiaosheng Si, Wenbin Wang, Changhua Hu, et al. Remaining useful life estimation – A review on the statistical data driven approaches. *European Journal of Operational Research*, 2011, 213(1): 1-14.
- [7] M Gašperin, Đ Juričić, P Bošković, et al. Model-based prognostics of gear health using stochastic dynamical models. *Mechanical Systems and Signal Processing*, 2011, 25(2): 537-548.
- [8] Xugang Zhang, Hua Zhang, Zhigang Jiang, et al. A decision-making approach for end-of-life strategies selection of used parts. *The International Journal of Advanced Manufacturing Technology*, 2013, <https://doi.org/10.1007/s00170-013-5234-0>.
- [9] SUN Weilian, Zailiang Chen, Chengbiao Wang. Failure analyzing thought and some failure cases analysis in mechanical produces failure. *Transactions of Materials and Heat Treatment*, 2004, 25(1): 69-73.
- [10] A Wohler. Versuche über die Festigkeit der Eisenbahnwagenachsen. *Journal of Engineering*, 1860(4): 160-161.
- [11] P C Paris. *The growth of cracks due to variations in load*. PA: Lehigh University, 1962: 1-237.
- [12] J D Morrow, R M Wetzal, T H Topper. Laboratory simulation of structural fatigue behavior. *American Society for Testing and Materials. Symposium on Effects of Environment and Complex Load History on Fatigue Life*, Atlanta, USA, 1970: 74-91.
- [13] A Ray, S Tangirala. A nonlinear stochastic model of fatigue crack dynamics. *Probabilistic Engineering Mechanics*, 1997, 12(1): 33-40.
- [14] S Mohanty, A Chattopadhyay, P Peralta, et al. Bayesian statistic based multivariate gaussian process approach for offline/online fatigue crack growth prediction. *Proceedings of the Society for Experimental Mechanics, Inc*, Bethel, United States, 2011, 51(6): 833-843.
- [15] Yumei Liu, Yaobin Wang, Lianfu Zhang. A study on wear-resisting reliable life of the F6L912G diesel engine. *Transactions of the Chinese Society of Agricultural Machinery*, 1999, 30(5): 100-102. (in Chinese)
- [16] Deguang Shang, Weixing Yao, Dejun Wang, et al. Multiaxial fatigue life prediction model based shear form. *Journal of Mechanical Strength*, 1999, 21(2): 141-144. (in Chinese)
- [17] J W Ringsberg. Life prediction of rolling contact fatigue crack initiation. *International Journal of Fatigue*, 2001, 23(7): 575-586.
- [18] A Fatemi, L Yang. Cumulative fatigue damage and life prediction theories: a survey of the state of the art for homogeneous materials. *International Journal of Fatigue*, 1998, 20(1): 9-34.
- [19] Xiaoyang Zheng, Jilong Xie. Probability failure accumulative method for predicting fatigue life under spectrum loading. *Journal of Mechanical Strength*, 1999, 21(3): 225-227. (in Chinese)
- [20] Xiulin Zheng, Baotong Lv, Tianxie Cui, et al. Fatigue tests and life Prediction of 16Mn steel butt weld without crack-like defect. *International Journal of Fracture*, 1994, 68(3): 275-280.
- [21] S Glode, M Sraml, J Kramberger. A computational model for determination of service life of gears. *International Journal of Fatigue*, 2002, 24(10): 1013-1020.
- [22] M Opalic, K Vuckovic, D Zelizelj. Effect of rotational speed on thin-rim gear bending fatigue crack initiation life. *10th International Conference on Fracture and Damage Mechanics*, Harbin Engineering University, Harbin, September 19-21, 2011.
- [23] D Rial, H Kebir, E Wintrebort, et al. Multiaxial fatigue analysis of a metal flexible pipe. *Materials and Design*, 2014, 54: 796-804.
- [24] C H Oppenheimer, K A Loparo. Physically based diagnosis and prognosis of cracked rotor shafts. *Proceedings of SPIE - The International Society for Optical Engineering. Components and Systems Diagnostics, Prognostics, and Health Management II*, Orlando, FL, USA, April 3-4, 2002: 122-132.
- [25] C E Feltner, J Morrow. Microplastic strain hysteresis energy as a criterion for fatigue fracture. *Journal of Basic Engineering*, 1961, 83(1): 15-22.
- [26] O Scott-Emuakpor, H Shen, et al. An energy-based uniaxial fatigue life prediction method for commonly used gas turbine engine materials. *Journal of Engineering for Gas Turbines and Power*, 2008, 130(6): 062504.
- [27] V Sarihan. Energy based methodology for damage and life prediction of solder joints under thermal cycling. *IEEE Transactions on Components Packaging and Manufacturing Technology Part B*, 1994, 17(4): 626-631.
- [28] H Ozaltun, M Shen, T George, et al. An energy based fatigue life prediction framework for in-service structural components. *Experimental Mechanics*, 2010, 51(5): 707-718.
- [29] O Scott-Emuakpor, T George, et al. Hysteresis-loop representation for strain energy calculation and fatigue assessment. *The Journal of Strain Analysis for Engineering Design*, 2010, 45(4): 275-282.
- [30] Weicheng Cui. A state of the art review on fatigue life prediction methods for metal structures. *Journal of Marine Science and Technology*, 2002, 7(1): 43-56.
- [31] S Curioni. Life prediction analysis based on a local strain approach for shafts - under torsional loading. *International Journal of Materials and Product Technology*, 1987, 2(1): 75-87.
- [32] Xiaoli Zhang, Xuefeng Chen, Bing Li, et al. Review of life prediction for mechanical major equipments. *Journal of Mechanical Engineering*, 2011, 47(11): 100-116. (in Chinese)
- [33] L Yang, A Fatemi. Cumulative fatigue damage mechanisms and quantifying parameters: A literature review. *Journal of Testing and Evaluation*, 1998, 26(2): 89-100.
- [34] J M Ferreira, A H Pereira, C M Branco. A fracture mechanics based fatigue life prediction for welded joints of square tubes. *Thin-Walled Structures*, 1995, 21(2): 107-120.
- [35] Yaowu Du, Weikang Jiang. A predictive model based on artificial neural net for residual life of large machinery. *Journal of Mechanical Strength*, 1997, 19(1): 5-8. (in Chinese)
- [36] J L Chaboche, P M Lesne. A non-linear continuous fatigue damage model. *Fatigue & Fracture of Engineering Materials & Structures*, 1988, 11(1): 1-17.
- [37] A Palmgren. Die Lebensdauer von Kugellagern (The service life of ball bearings). *Zeitschrift des Vereines Deutscher Ingenieure* (Journal of German Engineers), 1924, 68(14): 339-341.
- [38] K B Goode, J Moore, B J Roylance. Plant machinery working life prediction method utilizing reliability and condition-monitoring data. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical, Proceedings of the Institution of Mechanical Engineers*, 2000, 214: 109-122.
- [39] Wenzhu Liao, Ershun Pan, Ying Wang, et al. Research of predicting machine's remaining useful life based on statistical pattern recognition and auto-regressive and moving average model. *Journal of Shanghai Jiaotong University (Science)*, 2011, 45(7): 1000-1005. (in Chinese)
- [40] Jianguo Cui, Yunlong Zhao, Shiliang Dong, et al. Life prognostics for aero-generator based on genetic algorithm and ARMA model. *Acta Aeronautica et Astronautica Sinica*, 2011, 32(8): 1506-1511.
- [41] Jihong Yan, M Koc, J Lee. A prognostic algorithm for machine performance assessment and its application. *Production Planning and Control*, 2004, 15(8): 796-801.
- [42] Baojia Chen, Xuefeng Chen, Bing Li, et al. Reliability estimation for cutting tool based on logistic regression model. *Journal of Mechanical Engineering*, 2011, 47(18): 23-26. (in Chinese)
- [43] P J Volk. *Dynamic residual life estimation of industrial equipment based on failure intensity proportions*. Pretoria: University of Pretoria, 2001.
- [44] P J Volk, M Wnek, M Zygmunt. Utilising statistical residual life estimates of bearings to quantify the influence of preventive maintenance actions. *Mechanical Systems and Signal Processing*, 2004, 18(4): 833-847.
- [45] A K S Jardine, D Lin, D Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 2006, 20(7): 1483-1510.
- [46] N Gorgian, Y Sun, L Ma, et al. Remaining useful life prediction of rotating equipment using covariate-based hazard models—Industry applications. *Australian Journal of Mechanical Engineering*, 2017, 15(1): 36-45.
- [47] V N Vapnik. *Statistical learning theory*. New York: Wiley, 1998.
- [48] L Zhang, G Xiong, H Liu, et al. Fault diagnosis based on optimized node entropy using lifting wavelet packet transform and genetic algorithms. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 2010, 224: 557-573.
- [49] P J Garcia-Nieto, E Garcia-Gonzalo, F Sánchez, et al. Hybrid PSO-SVM-based method for forecasting of the remaining useful life for aircraft

- engines and evaluation of its reliability. *Reliability Engineering and System Safety*, 2015, 138: 219-231.
- [50] Hongzhong Huang, Haikun Wang, Yanfeng Li, et al. Support vector machine based estimation of remaining useful life: current research status and future trends. *Journal of Mechanical Science and Technology*, 2015, 29(1): 151-163.
- [51] Xuefeng Chen, Zhongjie Shen, Zhengjia He, et al. Remaining life prognostics of rolling bearing based on relative features and multivariable support vector machine. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 2013, 227: 2849-2860.
- [52] Zhongjie Shen, Xuefeng Chen, Zhengjia He, et al. Remaining life predictions of rolling bearing based on relative features and multivariable support vector machine. *Journal of Mechanical Engineering*, 2013, 49(2): 183-189. (in Chinese)
- [53] P M Pawar, R Ganguli. Fuzzy-logic-based health monitoring and residual-life prediction for composite helicopter rotor. *Journal of Aircraft*, 2007, 44(3): 981-995.
- [54] Zhiming Zhang, Huitao Cheng, Hong Xu, et al. Neural network based combining prediction model and its application in condition based maintenance of turbo-generator set. *Proceedings of the Chinese Society for Electrical Engineering*, 2003, 23: 204-206.
- [55] G Vachtsevanos, P Wang. Fault prognosis using dynamic wavelet neural networks. *Autotestcon Proceedings, IEEE Systems Readiness Technology*, Valley Forge, USA, August 20-23, 2001: 857-870.
- [56] Jiyang Fu, Fan Wang, Renhuai Liu, et al. Prediction of the residual life of thermo-mechanical fatigue of coke tower based on modified BP neural network. *Pressure Vessel Technology*, 2005, 22(5): 4-7.
- [57] Ling Xu, Dan Yang, Shilong Wang, et al. Prediction of cutting tool life based on evolutionary neural network. *Computer Integrated Manufacturing Systems*, 2008, 14(1): 167-182.
- [58] D Banjevic, A K S Jardine. Calculation of reliability function and remaining useful life for a Markov failure time process. *Journal of Management Mathematics*, 2006, 17(2): 115-130.
- [59] S J Engel, B J Gilmartin, K Bongort, et al. Prognostics, the real issues involved with predicting life remaining. *IEEE Aerospace Conference Proceedings*, Big Sky, MT, USA, March 18-25, 2000: 457-469.
- [60] C Bunks, D A N McCarthy, T Al-Ani. Condition-based maintenance of machines using hidden markov models. *Mechanical Systems and Signal Processing*, 2000, 14(4): 597-612.
- [61] Chen Lu, Tieying Li. Online milling tool wear monitoring based on continuous Hidden Markov Models. *International Conference Vibroengineering*, JVE International, September 17-19, 2013, 2: 65-69.
- [62] [62] P Baruah, R B Chinnam. HMMs for diagnostics and prognostics in machining processes. *International Journal of Production Research*, 2005, 43(6): 1275-1293.
- [63] R B Chinnam, R B Chinnam. Autonomous diagnostics and prognostics through competitive learning driven HMM-based clustering. *Proceedings of the International Joint Conference on Neural Networks*, Detroit, United States, July 20-24, 2003, 4: 2466-2471.
- [64] F Camci, R B Chinnam. Hierarchical HMMs for autonomous diagnostics and prognostics. *International Joint Conference on Neural Networks*, Vancouver, Canada, July 16-21, 2006: 2445-2452.
- [65] F Camci, R B Chinnam. Health-State estimation and prognostics in machining processes. *Automation Science and Engineering*, 2010, 7(3): 581-597.
- [66] Ming Dong, D He. Hidden semi-Markov model-based methodology for multi-sensor equipment health diagnosis and prognosis. *European Journal of Operational Research*, 2007, 178(3): 858-878.
- [67] Ming Dong, D He. A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology. *Mechanical Systems and Signal Processing*, 2007, 21(5): 2248-2266.
- [68] Chen Lin, V Makis. A comparison of Hidden Markov and Semi-Markov modeling for a deterioration system subject to vibration monitoring. *International Journal of Performability Engineering*, 2015, 11(3): 213-228.
- [69] Ning Wang, Shudong Sun, Zhiqiang Cai, et al. A hidden semi-markov model with duration-dependent state transition probabilities for prognostics. *Mathematical Problems in Engineering*, 2014, <https://doi.org/10.1155/2014/632702>.
- [70] Qinming Liu, Ming Dong, Wenyuan Lv, et al. A novel method using adaptive hidden semi-Markov model for multi-sensor monitoring equipment health prognosis. *Mechanical Systems and Signal Processing*, 2015, 64-65: 217-232.
- [71] Ying Peng, Ming Dong. A prognosis method using age-dependent hidden semi-Markov model for equipment health prediction. *Mechanical Systems and Signal Processing*, 2011, 25(1): 237-252.
- [72] Qinming Liu, Ming Dong, Ying Peng. 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*, 2012, 32: 331-348.
- [73] Haifeng Hu, Maochun An, Guojun Qin, et al. Study on fault diagnosis and prognosis methods based on hidden semi-Markov model. *Acta Armamentarii*, 2009, 30(1): 69-75.
- [74] Lun Ma, Jianshe Kang, Qiang Zhao. Implementation of equipment residual life prediction framework based on Hidden Markov Model. *Computer Simulation*, 2010, 27(5): 88-91. (in Chinese)
- [75] H Sunakawa, M Nishimoto, Yamanishin, et al. Reliability and life time evaluation approach of the LE-X engine. *47th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit*, San Diego, CA, United States, July 31-August 3, 2011.
- [76] Tiejun Zheng, Xi Wang, Xiuqin Luo, et al. Modified method of establishing the state space model of aeroengine. *Journal of Propulsion Technology*, 2005, 26(1): 46-49. (in Chinese)
- [77] Jie Hong, Xuewen Miao, Yanhong Ma. Life prediction model of operating state for aeroengine main bearing. *Aeroengine*, 2008, 34(3): 18-21. (in Chinese)
- [78] N J Gordon, D J Salmond, A F M Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Radar and Signal Processing*, 1993, 140(2): 107-113.
- [79] N D Freitas. Rao-Blackwellised particle filtering for fault diagnosis. *IEEE Aerospace and Electronics Systems Society*, Big Sky, MT, United States, March 9-16, 2002, 4: 1767-1772.
- [80] Qinghua Zhang, F Campillo, F Cerou, et al. Nonlinear system fault detection and isolation based on bootstrap particle filters. *44th IEEE Conference on Decision and Control, and the European Control Conference*, Seville, Spain, December 12-15, 2005: 3821-3826.
- [81] Guang Jin, D E Matthews, Zhongbao Zhou. A Bayesian framework for on-line degradation assessment and residual life prediction of secondary batteries in spacecraft. *Reliability Engineering and System Safety*, 2013, 113: 7-20.
- [82] D Wang, F Yang, K L Tsui, et al. Remaining useful life prediction of lithium-ion batteries based on spherical cubature particle filter. *IEEE Transactions on Instrumentation and Measurement*, 2016, 65(6): 1282-1291.
- [83] G X P Box, G M Jenkins. *Time series analysis: Forecasting and control*. San Francisco: The University of Chicago Press, 1976.
- [84] F L Lewis. *Applied optimal control & estimation: digital design & implementation*. Texas: Prentice Hall PTR, 1992.
- [85] H Akaike. Markovian representation of stochastic processes by canonical variables. *SIAM Journal on Control*, 1975, 13(1): 162-173.
- [86] J Yan, J Lee. Degradation assessment and fault modes classification using logistic regression. *Journal of Manufacturing Science and Engineering*, 2005, 127(4): 912-914.
- [87] D Kumar, B Kenkins. Proportional hazards model: a review. *Reliability Engineering and System Safety*, 1994, 44(2): 177-188.
- [88] E A Elsayed, H Zhang. Design of PH-based accelerated life testing plans under multiple-stress-type. *Reliability Engineering & System Safety*, 2007, 92(3): 286-292.
- [89] H Langseth, L Portinale. Bayesian networks in reliability. *Reliability Engineering and System Safety*, 2007, 92(1): 92-108.
- [90] J Z Sikorska, M Hodkiewicz, L Ma. Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing*, 2011, 25(5): 1803-1836.
- [91] C J C Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 1998, 2(2): 121-167.
- [92] L R Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 1989, 77: 257-285.
- [93] J Lee, F J Wu, W Y Zhao, et al. Prognostics and health management design for rotary machinery systems - Reviews, methodology and applications. *Mechanical Systems and Signal Processing*, 2014, 42(1-2): 314-334.

- [94] Zeyi Huang, Zhengguo Xu, Wenhai Wang, et al. Remaining useful life prediction for a nonlinear heterogeneous wiener process model with an adaptive drift. *IEEE Transactions on Reliability*, 2015, <https://doi.org/10.1109/tr.2015.2403433>.
- [95] W L Berendrecht, A W Heemink, F C Vab-Geer, et al. A non-linear state space approach to model groundwater fluctuations. *Advances in Water Resources*, 2006, 29(7): 959–973.
- [96] Peng Wang, R X Gao. Adaptive resampling-based particle filtering for tool life prediction. *Journal of Manufacturing Systems*, 2015, <https://doi.org/10.1016/j.jmsy.2015.04.006>.
- [97] Naipeng Li, Yaguo Lei, Zongyao Liu, et al. A particle filtering-based approach for remaining useful life predication of rolling element bearings. *2014 International Conference on Prognostics and Health Management*, Cheney, WA, United States, June 22-25, 2014.
- [98] R K Mehra. *Kalman filters and their applications to forecasting*. Amsterdam and New York: North-Holland Publishing Co, 1979.
- [99] A P Dempster, N M Laird, D B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 1977, 39(1): 1-38.
- [100] Zhonggai Zhao, Biao Huang, Fei Liu. Parameter estimation in batch process using EM algorithm with particle filter. *Computers & Chemical Engineering*, 2013, 57: 159-172.
- [101] Xuehua Chen. *Theory and algorithms on state space modeling and its applications in financial econometrics*. Jinan: Jinan University, 2007. (in Chinese)
- [102] Lei Feng, Hongli Wang, Xiaosheng Si, et al. A state-space-based prognostic model for hidden and age-dependent nonlinear degradation process. *IEEE Transactions on Automation Science and Engineering*, 2013, 10(4): 1072-1086.
- [103] Jianzhong Sun, Hongfu Zuo, Weibin Wang, et al. Prognostics uncertainty reduction by fusing on-line monitoring data based on a state-space-based degradation model. *Mechanical Systems and Signal Processing*, 2014, 45(2): 396-407.
- [104] Jianzhong Sun, Hongfu Zuo, Weibin Wang, et al. Application of a state space modeling technique to system prognostics based on a health index for condition-based maintenance. *Mechanical Systems and Signal Processing*, 2012, 28: 585-596.
- [105] G Storvik. Particle filters for state-space models with the presence of unknown static parameters. *Signal Processing*, 2002, 50(2): 281-289.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
