

Remaining useful life prediction based on a modified SVR method with health stage division, cluster sampling and similarity matching

Jian-wen Yan

School of Management, Hefei University of Technology

Xiao-hu Zhong

School of Management, Hefei University of Technology <https://orcid.org/0000-0003-3061-1462>

Yu-cheng Cheng

Anhui Sanqing Industry,

Yu Fan (✉ yufan@hfut.edu.cn)

Anhui Province Key Lab of Aerospace Structural Parts Forming Technology and Equipment, Hefei University of Technology <https://orcid.org/0000-0002-0001-2496>

Original Article

Keywords: Remaining useful life, support vector regression, health stage division, cluster sampling, similarity matching

Posted Date: April 6th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-20832/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Remaining useful life prediction based on a modified SVR method with health stage division, cluster sampling and similarity matching

Jian-wen Yan^{1,2}, Xiao-hu Zhong^{1,2,3}, Yu-cheng Cheng⁴ and Yu Fan^{2*}

Author details

¹ School of Management, Hefei University of Technology, Hefei 230009, China

² Anhui Province Key Lab of Aerospace Structural Parts Forming Technology and Equipment, Hefei University of Technology, Hefei 230009, China

³ Anhui Wanwei Group Co., Ltd, Hefei 238002, China

⁴ Anhui Sanqing Industry Co., Ltd, Hefei 230009, China

* Correspondence: yufan@hfut.edu.cn

Email address

Jian-wen Yan: yanjianwen@hfpress.com;

Xiao-hu Zhong: 2017110735@mail.hfut.edu.cn;

Yu-cheng Cheng: cyc@hfpress.com;

Yu Fan: yufan@hfut.edu.cn;

Abstract

Remaining useful life (RUL) prediction is an advanced methodology of prognostics and health management (PHM), and is in advantage of life cycles management of equipment and maintenance cost reduction. Among the data driven methods, support vector regression (SVR) is one of the most suitable methods when there are limited failure history data for analysis. However, many uncertain factors such as individual variation and time varying operating conditions, will lead that the failure time of all equipment statistically dispersive, and with the increase of sample set, such dispersity may inevitably increase and further reflects in the model training. In consequent, the dispersity may cause two drawbacks. On the one hand, the linearity of SVR model will increase with the increase of sample set, and overfitting or underfitting tends to occur. And on the other hand, a single model only performs well in its generalization and robustness, but may lost its effectiveness that it may fail to work well for a new on-service equipment. In order to deal with the two drawbacks, this paper proposes a modified SVR method with health stage division, cluster sampling and similarity matching. Through health stage division and cluster sampling, each state of the whole degradation process can obtain the optimal parameters, then the irrelevant linearity can be reduced. In addition, since that similar input results similar output, the optimal parameters of the most similar testing sample are also suitable for the on-service equipment, and through similarity matching the most similar testing sample can be obtained, thus the drawback of a single model can be avoided. Finally, the effectiveness of the proposed method is verified systematically by a simulated dataset of fatigue crack growth and a real-world degradation dataset of GaAs-based lasers.

Keywords Remaining useful life; support vector regression; health stage division; cluster sampling; similarity matching

1 Introduction

The operating reliability of mechanical equipment significantly influences the sustainability and competitiveness of manufacturing industry. As the reliability of mechanical equipment decreases with the extension of the operating time, prognostics and health management (PHM) techniques attract more and more attentions [1-3]. It determines the optimal maintenance time, inspection interval, spare parts order quantity and other logistics management strategies through efficient and low-cost diagnostic and prognostic activities to achieve the lowest economic cost or equipment failure risk. The key of PHM technology is to predict the equipment's remaining useful life (RUL), i.e. the remaining useful time for a certain part or component to perform its function before its final failure [4]. The accurate results of RUL prediction can help decision makers to make maintenance plans in advance and optimize supply chain management. Compared with traditional reliability-based methods [5,6], existing RUL prediction methods mainly learn the potential degradation law and the failure information through analyzing the condition monitoring data. They can be generally divided into three categories: physics-based methods, data-driven methods and their combinations [7,8]. A physics-based method always requires a broad understanding of the physical mechanism of the equipment, but an excess of physical parameters make it difficult to get an accurate physical model [9]. In comparison, when the explicit physical model is unknown, data-driven methods can directly analyze the condition monitoring data, and predict the RUL.

Existing data-driven methods can be divided into: (1) the degradation-modeling based methods [8], which aims to model the degradation evolution process of equipment; (2) the machine learning methods [10], which attempts to map the relationship between degradation data and its RUL directly [11].

A degradation-modeling based method mainly uses a statistical model to learn the degradation evolution from the condition monitoring data, and then estimates the RUL by comparing the predicted results with the pre-set failure threshold. Examples such as: Wiener process models [12,13], Gamma process models [14,15], inverse Gaussian process models [16,17], hidden (semi) Markov models (HMM/HSMM) [18,19], etc. The inherent drawbacks of the degradation-modeling based methods derive from two strong premises: (1) the degradation process of equipment's performance should follow a certain statistical model, such as the continuous-time Markov chain, the hidden Markov model, the hidden semi-Markov model or the Wiener process, .etc.; (2) the statistical property of the degradation model, for example the transition probability matrix for Markov-based models, should be a priori known or estimated[20]. However, for practical instances, theoretical statistical models such as Markov chain are very hard to be verified, and the transition probability matrix is often hard to estimate or inaccessible. Therefore, the applicability of the degradation-modeling based methods is limited in engineering practice.

In contrast, the machine learning methods such as artificial neural networks (ANN) [21,22], similarity-based RUL prediction (SbRP) method [11,23] and support vector machine (SVM) [24-31], can effectively avoid these issues. Although the machine learning methods don't provide a probability density function (PDF) estimate of the RUL, they are capable of dealing with prognostic issues of complex systems whose degradation processes are difficult to be interrelated by physics-based methods or degradation-modeling based methods, and have been widely studied and applied in recent years.

In practice, considering the validity and expenditure, the main challenge of PHM is how to perform RUL prediction with the limited amount of failure samples. To the best of our knowledge, support vector regression (SVR) as the most common application of SVM algorithm in PHM field, was originally

proposed for small-sample analysis [24], it can effectively estimate the RUL with limited failure samples. Widodo and Yang [25] developed an intelligent system using survival analysis and SVR to predict the survival probability of bearings. Loutas et al. [26] reported the ϵ -SVR, and employed it for estimating the RUL of rolling element bearings. Benkedjough et al. [27,28] used SVR to map the degradation time series into nonlinear regression, and then fitted it into the power model for RUL prediction of the mechanical equipment. Liu et al. [29,30] proposed an improved probabilistic SVR model to predict the RUL of equipment components of a nuclear power plant. Fumeo et al. [31] developed an online SVR model to predict the RUL of bearings by optimizing the tradeoff between accuracy and computational efficiency. Tao et al. [20] considered the dynamic multi-state operating conditions, and trained the corresponding SVR model under different operating conditions, then the RUL of the on-service equipment can be predicted under time varying operating conditions. Some other SVR-based methods were developed in [32-35]. These studies promote the application of SVR-based methods in PHM, but due to the dispersity of samples' failure time, there is an inevitable drawback that the linearity of SVR model will increase with the increase of sample set, and overfitting or underfitting tends to occur [9]. Furthermore, such dispersity also represents the diversity of equipment's degradation processes, thus a single SVR model is not always the best one for a brand-new equipment.

Therefore, in order to deal with the two drawbacks, this paper proposes a modified SVR method with health stage division, cluster sampling and similarity matching. As we shall show, the novelties and contributes of this paper exhibit in the following aspects:

- (1) The reasons of SVR model linearization increasing with data set are analyzed.
- (2) It combines health stage division and cluster sampling in SVR model training, and can obtain the optimal parameters of each health stage, thus the drawback that the linearity of SVR model will increase with the increase of sample set can be solved effectively.
- (3) It applies similarity matching to select the optimal parameters in the online prediction phase. Based on the idea of similar inputs mean similar outputs, the optimal SVR model for the on-service equipment can be obtained from its most similar testing sample, thus the limitation of a single SVR model can be avoided.

The rest of this paper is organized as below. Section 2 introduces several related theories of the proposed method. Section 3 discusses the problem of a traditional SVR-based RUL prediction method. Section 4 describes the modified SVR method for online RUL prediction. Section 5 analyzes systematically the proposed method by two datasets. Finally, Section 6 concludes the whole paper.

2 Related theories

2.1 Support vector regression

SVR is the successful extension of SVM for dealing with regression problems [36]. It aims to find a maximum margin in order to minimize the error of the missed training data. Given a training set $Z = \{x_i, y_i\}_{i=1,2,\dots,n}$ with input data $x_i \in R^N$ and output data $y_i \in R$, the regression function is expressed as follow:

$$y = f(x) = w\phi(x) + b \quad (1)$$

where $\phi(x)$ denotes the feature of inputs, w and b are the coefficients, and can be estimated by minimizing the following regularized risk function:

$$R_{SVR}(C) = \frac{1}{2}\|w\|^2 + C \frac{1}{n} \sum_{i=1}^n \ell_{\varepsilon}(f(x_i), y_i) \quad (2)$$

where the first term $\frac{1}{2}\|w\|^2$ is called the regularized term and used to estimate the flatness of the function, the second term $C \frac{1}{n} \sum_{i=1}^n \ell_{\varepsilon}(f(x_i), y_i)$ is the so-called empirical error or training error. C is the specified criteria chosen to balance the tradeoff between the flatness of the function and the training error. ℓ_{ε} is the ε linear insensitive loss function:

$$\ell_{\varepsilon}(f(x_i), y_i) = \begin{cases} |f(x_i) - y_i| - \varepsilon, & \text{if } |f(x_i) - y_i| \geq \varepsilon \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where ε is the tube size of SVR. If the penalty factor C is larger, it is proved that the training error is large and the sample penalty of ε is larger, otherwise the opposite is true; if ε is smaller, the error of the regression function is smaller.

Then two positive slack variables ξ and ξ^* are introduced to represent the distance from desired values to the corresponding boundary values of the ε -tube. Eq. (2) can be rewritten as:

$$\begin{cases} \text{Minimize } R_{SVR}(w, \xi, \xi^*) = \frac{1}{2}\|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ \text{s. t. } \begin{cases} y_i - w\phi(x_i) - b \leq \varepsilon + \xi_i \\ w\phi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i \geq 0, \xi_i^* \geq 0 \end{cases} \end{cases} \quad i = 1, 2, \dots, n \quad (4)$$

Finally, by introducing Lagrange multipliers and exploiting optimality constrains, Eq. (1) is transformed into the form rewritten as follow:

$$f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x, x_i) + b \quad (5)$$

where α_i, α_i^* are the Lagrange multipliers which satisfy the equality $\alpha_i \alpha_i^* = 0, \alpha_i \geq 0, \alpha_i^* \geq 0$ and obtained by maximizing the dual function of Eq. (4) which Karush-Kuhn-Tucker conditions are applied.

The dual Lagrange form of Eq. (4) is given by:

$$R_{SVR}(\alpha_i, \alpha_i^*) = \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j) \quad (6)$$

with the constrains:

$$\sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0, \quad 0 \leq \alpha_i \leq C, \quad 0 \leq \alpha_i^* \leq C, \quad i = 1, 2, \dots, n \quad (7)$$

where $K(x_i, x_j)$ is the kernel function, and its value is equal to the inner product of two vectors x_i and x_j in the feature space $\phi(x_i)$ and $\phi(x_j)$, i.e. $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$.

Any function satisfied Mercer's condition [24] can be used as the kernel function, such as linear kernel, polynomial kernel, radial basis function (RBF) kernel and sigmoid kernel, etc.

2.2 Cluster sampling

In sampling theory, the coherent sample set composed of several connected basic units is called a cluster. And for cluster sampling, the population is first divided into different sub-clusters according to some criteria, then each sub-cluster is taken as the sampling unit that be extracted integrally for analyzation. Due to the advantages of simplicity, convenience and cost-reduction, cluster sampling is widely used for statistical analysis, and developed a variety of variants for different scenarios, such as adaptive cluster sampling [37], stratified cluster sampling [38], etc. The main disadvantages of cluster sampling are that the sample distribution is not wide and the representation for population is relatively poor. However, it does a good job of preserving the integrity of local information. Thus, the method of cluster sampling is employed in this

paper to improve the SVR-based method.

2.3 Similarity matching

Similarity matching is the key procedure of the SbRP method. As an emerging and universal data-driven method, SbRP has been rapidly studied and developed in recent years. It predicts the RUL of an on-service equipment by a weighted arithmetic mean of those of reference ones [23,39] and is capable to realize long-term estimation and dispense with the degradation modeling [40]. The core idea of the SbRP method is that similar inputs product similar outputs. And based on similarity matching, the most similar trajectories of the reference samples are obtained, then the RUL of the on-service equipment can be calculated by the weighted average RUL of those trajectories [41]. Thus, we consider that the most similar sample's optimal SVR model parameters are also the optimal for the on-service equipment, and employ similarity matching method for parameters selection in the online prediction stage. The most commonly used method of similarity matching is the distance-based algorithms, such as the Euclidian, Manhattan, and Canberra distances.

3 Problem statement

A large number of experiments and engineering cases show that even the same category or the same batch of equipment, their degradation curves are often different with each other due to the difference of internal structure and the variability of operating environment [42], i.e. the failure time of all equipment are statistically dispersive, reference to Fig. 1. With the increase of sample set, such dispersity may inevitably increase and further reflects in model training.

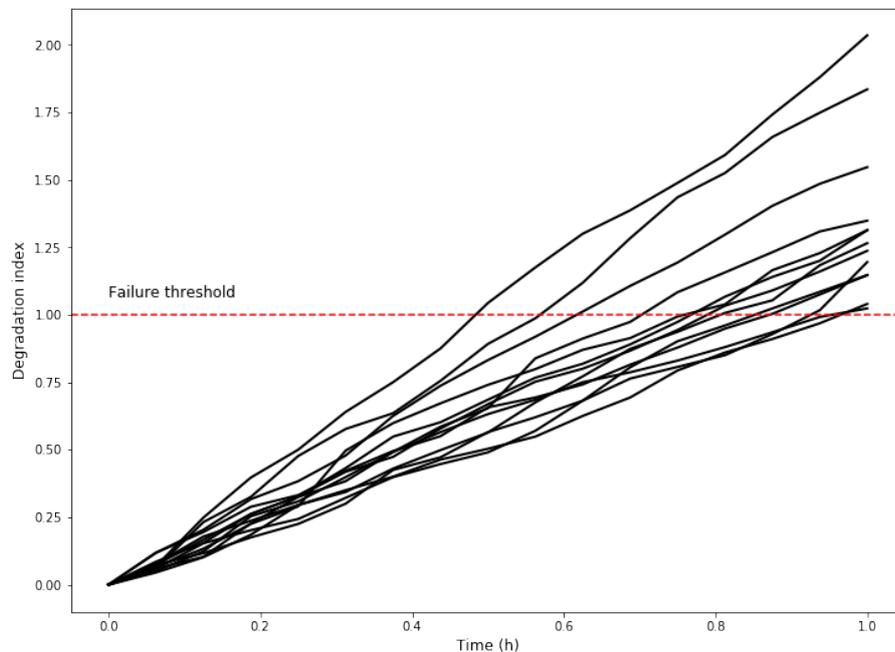


Fig. 1 The degradation curves of the same category of equipment

The goal of SVR is to solve the function $f(x)$ that has at most ε deviation from the obtained targets of y_i for all the training data, errors can be ignored if they are less than ε , namely the loss

function ℓ_ε doesn't calculate the data points between $f(x) + \varepsilon$ and $f(x) - \varepsilon$ (according to Eq. 3). For a trained SVR model, ε is a constant and the ε -tube is unique. However, the uncertainties [43] in the early stages are always more than that in the late stages, thus the dispersity of RUL in the early stages is usually more obvious than that in the late stages, which causes that in the early stages there are more data points outside the ε -tube, and therefore, follow a constant ε may cause that the data points in the late stages have less impact in model training, thus the linearity will increase with the increase of sample set, and overfitting or underfitting tends to occur. For example, Fig. 2 shows the data point distribution of a training data set, they spread out in the early stages and then gradually concentrate, and when using an SVR model to fit them, there is an underfitting for all those data if $x > x_\tau$. Therefore, the ε -tube may be too narrow in the early stages and too wide in the late stages.

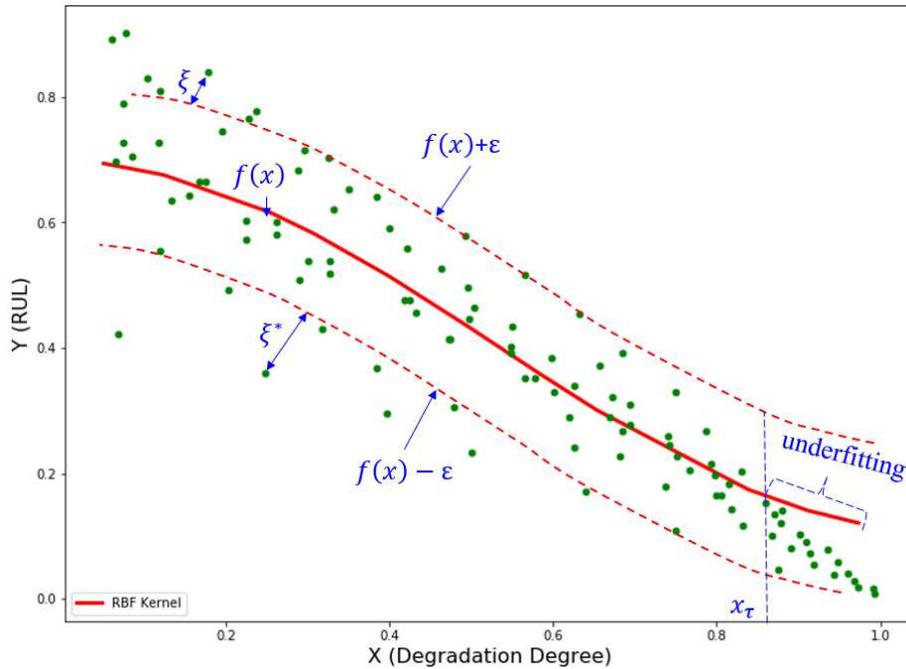


Fig. 2 The problem of RUL prediction by SVR method

In addition, the diversity of equipment's degradation processes may lead that a single SVR model constructed offline fails to work well for a brand-new equipment.

4 Methodology

As the problem discussed in Section 3, the dispersity of equipment's failure time leads the constant ε -tube don't suit for the whole regression process of SVR, health stage division is suit for dealing with this issue. By health stage division, the degradation processes can be divided into several stages according to the different degradation patterns, each stage can be analyzed separately. For instance, Hu et al. [44] divided the degradation processes of generator bearings into four stages based on the change points of confidence levels. Kimotho et al. [45] divided the degradation processes of bearings into five stages by analyzing the changes of frequency amplitudes in the power spectral density. They consider that the degradation process of a product can usually be divided into two or more stages, each stage presents different degradation patterns, and by the multi-stage RUL prediction, the accuracy is improved.

However, many kinds of equipment's degradation processes don't have the obvious change points, and the high-quality failure historical data often less in engineering practice, such issues lead the multi-

stage RUL prediction in challenge. For a multi-stage SVR method, if the training sample set is small, there is an inevitably issue that the corresponding SVR model of two consecutive stages may be overfitting or underfitting in their joining part, for the reason that this part is the predicted part of regression. Reference to Fig. 3, the predicted parts are always full of uncertainty, and may be inaccurate and unsatisfactory. Therefore, cluster sampling is introduced to solve this problem.

The application mode of cluster sampling in the proposed SVR method is shown in Fig. 4. In order to obtain the optimal model parameters of stage k , the corresponding data of stage k of the testing data (such as the data $Z_k = \{z_p, \dots, z_q\}$ shown in Fig. 4) are cluster sampled about M times and added into the original testing data. Then on the one hand, the weight of one of the data points of Z_k increases from $\frac{1}{n}$ to ω_{z_k} , and

$$\omega_{z_k} = \frac{M+1}{n+(q-p+1)M} \quad (8)$$

where $1 \leq q - p + 1 \leq n$ and $\frac{1}{n} = \frac{M+1}{n+nM} \leq \omega_{z_k} \leq \frac{M+1}{n+M}$, thus $\omega_{z_k} \geq \frac{1}{n}$ is always right, and the partial testing data Z_k plays more roles in parameters optimization. On the other hand, other data except Z_k greatly decrease the uncertainty in the process of SVR model construction, even though their weights are relatively lower, if these data outside the ε -tube (far from the optimal regression curve), the loss function ℓ_ε will increase. Thus, by introducing the method of cluster sampling, overfitting or underfitting in the joining part of two consecutive stages can be reduced.

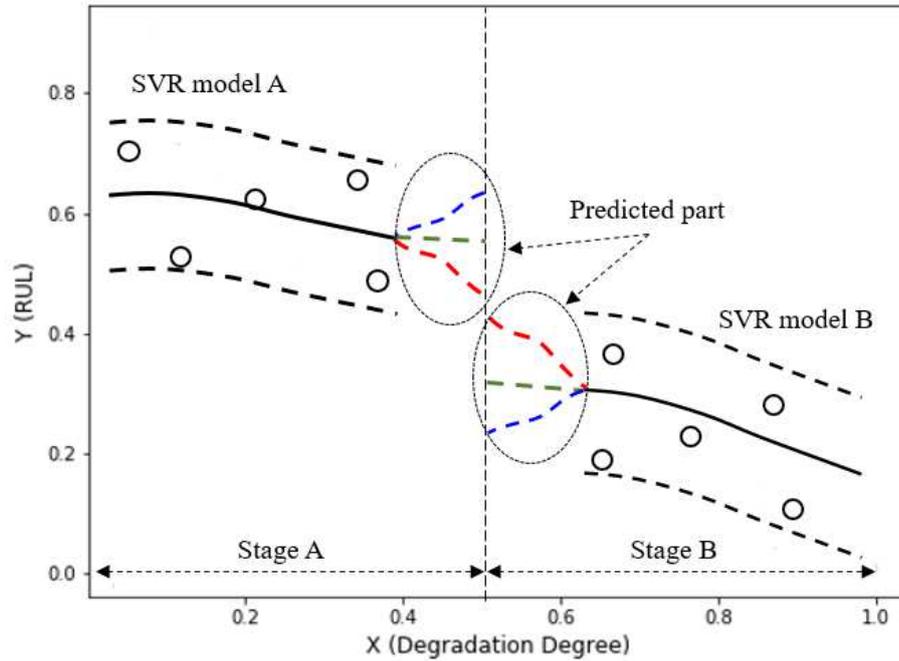


Fig. 3 The issue of just simply divide the degradation into several stages and construct the corresponding SVR models

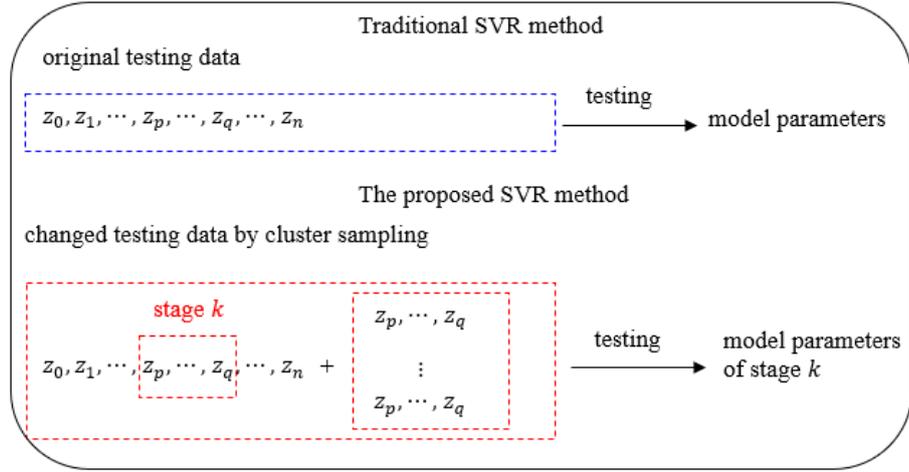


Fig. 4 The difference of model construction between the proposed SVR method and the traditional SVR method

Additionally, in order to refrain from focusing on generalization and neglecting the impact of the diversity of equipment's degradation processes, we introduce similarity matching to select the better parameters in the online prediction stage. The main flowchart of the proposed method is shown in Fig. 5, and described as follows.

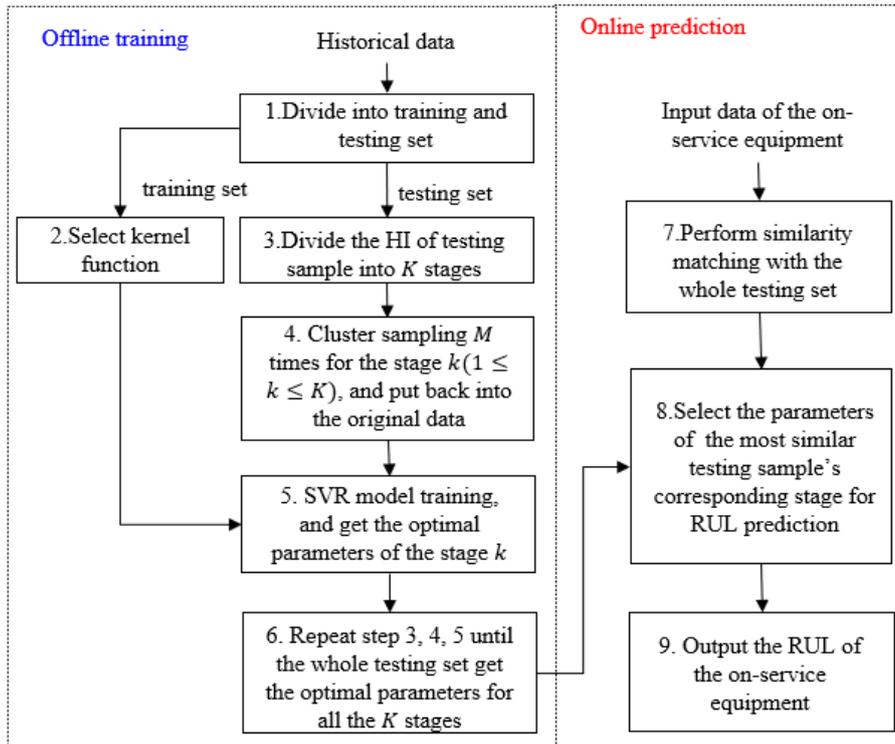


Fig. 5 The main flowchart of the proposed method

Step 1: Divide the original data into training set and testing set. As far as possible, the degradation processes of the testing samples should be significantly different from each other. Thus, each testing sample is of representative.

Step 2: Select the appropriate kernel function. The key of SVR is to select the type of kernel function, which mainly includes linear kernel, polynomial kernel, RBF kernel and sigmoid kernel. The most widely used of these functions should be RBF kernel. Whether the small sample set or large sample

set, high dimensional or low dimensional, the RBF kernel function is applicable. Compared with other functions, it has the following advantages: (1) The RBF kernel can map a sample to a higher dimensional space, and the linear kernel is a special case of RBF, which means that if you consider using RBF, there is no need to consider the linear kernel. (2) Compared with polynomial kernel function, RBF needs to determine fewer parameters, and the number of kernel function parameters directly affects the complexity of the function. In addition, when the order of the polynomial is relatively high, the element value of the kernel matrix will tend to infinity or infinitesimal, while the RBF is above, which will reduce the difficulty of numerical calculation. (3) For some parameters, RBF and sigmoid kernel have similar performance.

Step 3: Divide the HI of testing sample into K stages. Health stage division should follow some criterions, either according to the change points or just evenly divide. Anyway, compared to the entire time series, the internal dispersion of each stage has shrunk.

Step 4: Cluster sampling M times for the stage $k(1 \leq k \leq K)$, and put back into the original data. Reference to Fig. 4, the modified testing sample will be used for training parameters for the stage k .

Step 5: SVR model training, and get the optimal parameters of the stage k . Some algorithms can be used for solve the parameters optimal problem, such as genetic algorithm (GA), particle swarm optimization (PSO), grid search, etc.

Step 6: Repeat step 3, 4, 5 until the whole testing set get the optimal parameters for all the K stages.

Step 7: For the HI of the on-service equipment, perform similarity matching with the whole testing set. Similarity matching in this paper mainly focuses on the similarity of degradation curves.

Step 8: Online prediction, select the parameters of the most similar testing sample's corresponding stage for RUL prediction.

Step 9: Finally, output the RUL of the on-service equipment.

5 Case studies

In this section, we first use a simulated dataset to demonstrate the effectiveness of the proposed method. Then, it is applied to the GaAs-based lasers dataset. The RBF kernel is used in the proposed method and traditional SVR method, and the parameters optimization problem is solved by Grid search method.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (9)$$

where γ is the radius of the RBF kernel function.

In order to demonstrate the performance of the proposed method, we employ mean absolute error (MAE) and root-mean square error (RMSE) as the performance metrics to evaluate the accuracy of the proposed method:

$$e_{MAE} = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{N} \quad (10)$$

$$e_{RMSE} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (11)$$

where N represents the total number of data points in the on-service sample's HI, y_i is the accurate value and \hat{y}_i represents the predicted value. the lower the e_{MAE} or e_{RMSE} , the better the accuracy. The performance of the proposed method is illustrated by comparing with the traditional SVR method and SbRP method.

5.1 Simulation study

In this subsection, the prediction is illustrated and the performance is evaluated through numerical simulations. The degradation curves dataset of 30 units is obtained by an exponential degradation model [46]:

$$L(t) = \theta + \beta t + \varepsilon(t) \quad (12)$$

where $L(t)$ is the log value of the degradation data, $\theta \sim N(\mu_0, \sigma_0^2)$, $\beta \sim N(\mu_1, \sigma_1^2)$ and the noise $\varepsilon(t) \sim N(0, \sigma^2 t)$. In this simulation study, all the initial damages start at 0 inch, and the failure threshold (FT) is set to be 50 inches. The preset values of the model parameters are given in Table 1. Fig. 6 shows the degradation curves of the 30 units. The numerical simulations are all terminated at 21 million of cycles. In this study, we choose 3 units as the on-service samples, 8 units as the testing samples with their failure time have relatively balanced distribution, others as the training samples. Online prediction is performed for the three on-service samples by different methods.

Table 1 Preset parameters of the exponential degradation model

Parameters	μ_0	σ_0^2	μ_1	σ_1^2	σ^2
Preset values	1	0.2	4	0.6	0.06

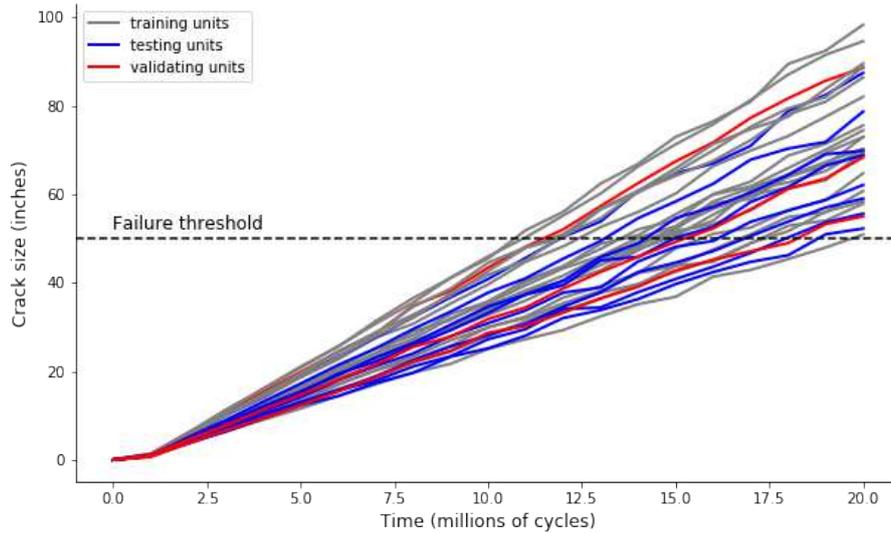


Fig. 6 The degradation curves of the simulation dataset

For the 8 testing units, we divide their respective degradation processes into 5 stages as the interval of each stage is 10 inches, i.e. $K=5$. Then the optimal number of cluster sampling can be analyzed from repeated experiments. When $M=0$, we define the training error of the corresponding stage as a standard unit, i.e. define $e_k(M=0) = 1$, then the variation trends of the training error with the cluster sampling number M are shown clearly in Fig. 7. Obviously, as the M increasing, training errors are non-increasing, and finally converges at a definite level. Therefore, according to Fig. 7, $M=25$ is an appropriate value for cluster sampling in this study.

In the online prediction phase, the prediction results of the three on-service samples are shown in Fig. 8. For the 1st on-service sample, 8 predictions were made in the whole life cycles, the proposed method has 5 prediction errors with the minimum (62.5%), while traditional SVR method has 0 (0%) and SbRP method has 3 (37.5%), results are listed in Table 2. And for the 1st on-service sample, the

MAE and RMSE of the proposed method are all lower than the other two methods. For the 2nd on-service sample, 12 predictions were made in the whole life cycles, the proposed method has 3 prediction errors with the minimum (25%), while traditional SVR method has 1 (8.33%) and SbRP method has 8 (66.67%), results are listed in Table 3. And for the 2nd on-service sample, the MAE and RMSE of the proposed method are all lower than traditional SVR method, but higher than SbRP method. For the 3rd on-service sample, 15 predictions were made in the whole life cycles, the proposed method has 7 prediction errors with the minimum (46.67%), while traditional SVR method has 7 (46.67%) and SbRP method has 5 (33.33%), results are listed in Table 4. And for the 3rd on-service sample, the MAE and RMSE of the proposed method are all lower than traditional SVR method, but the RMSE of the proposed method is higher than SbRP method while its MAE is lower than SbRP method.

Results show that the proposed method is of the better performance than traditional SVR method, but the pros and cons between the proposed method and SbRP method are unclear, and need further analysis in the next subsection.

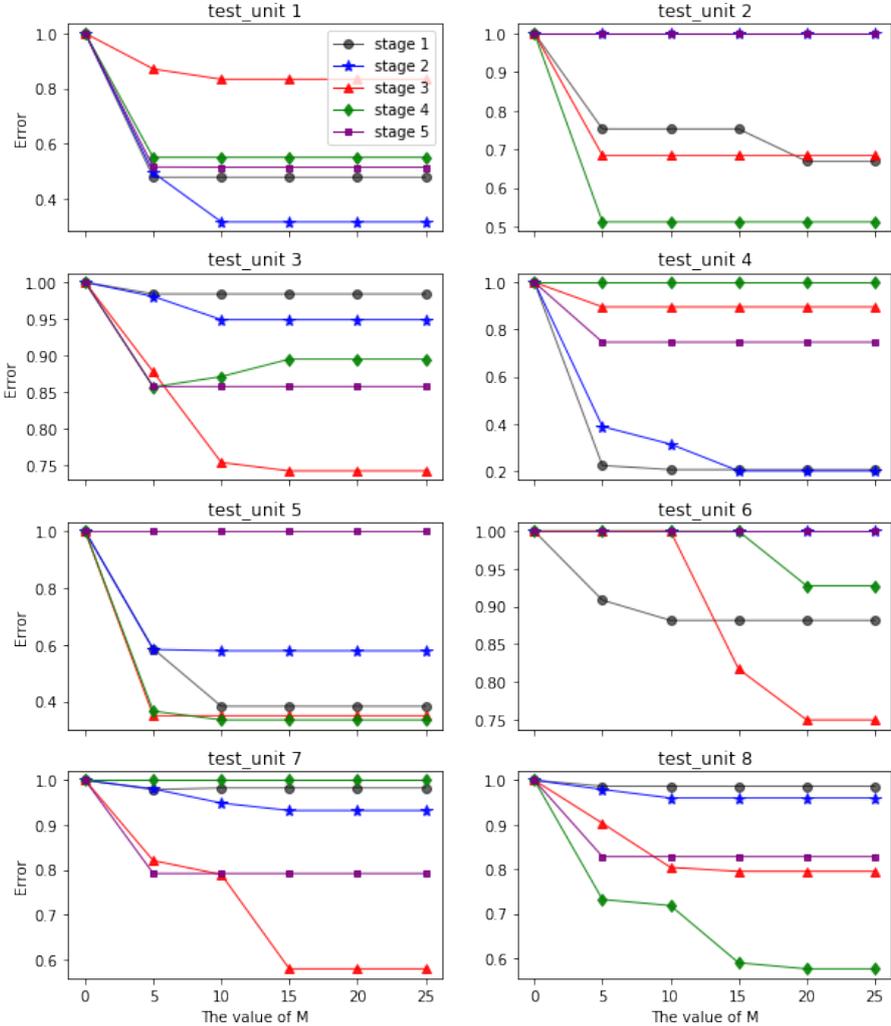


Fig. 7 The variation trend of the error in each stage of the testing samples

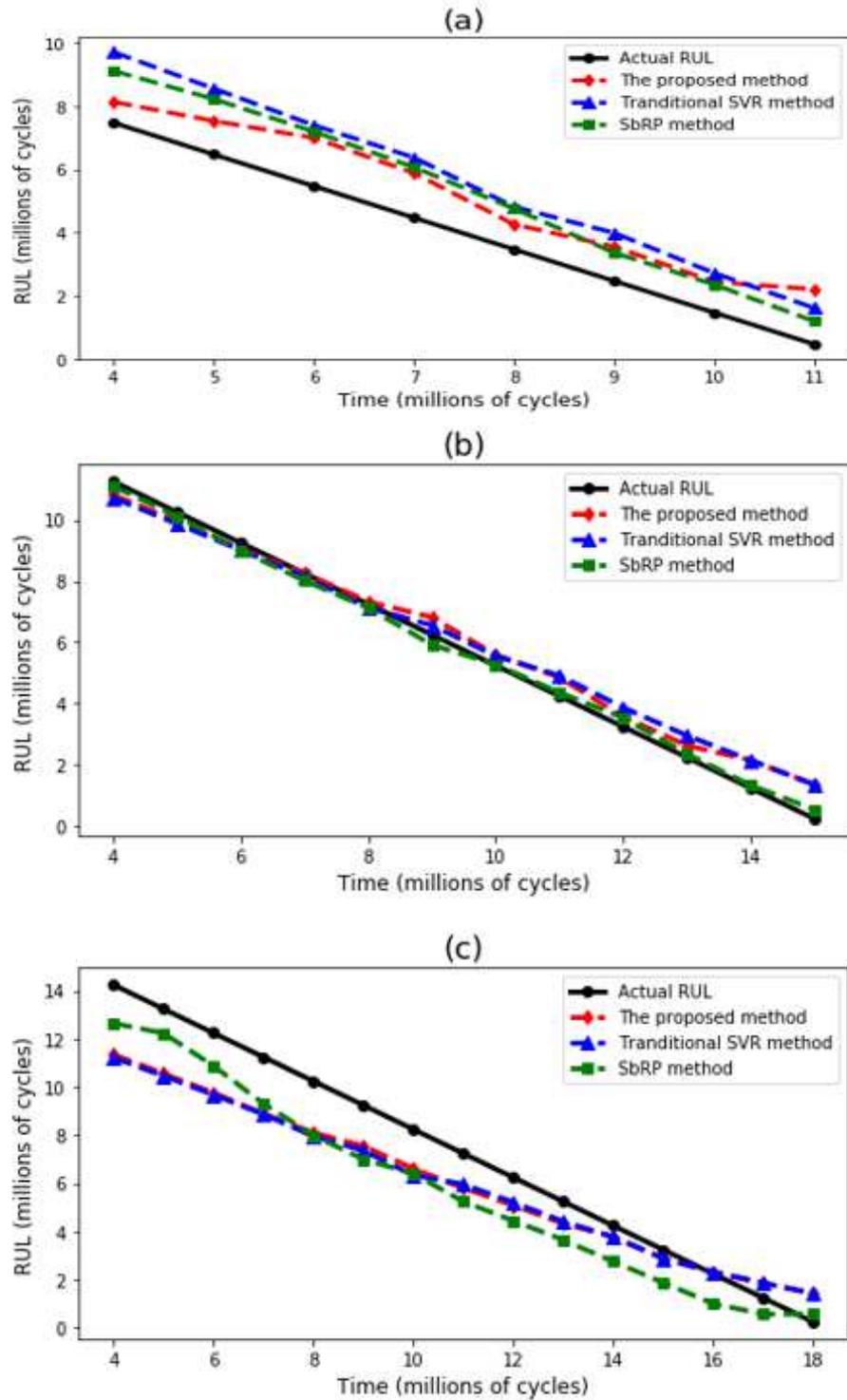


Fig. 8 The prediction results of the on-service samples of the simulation data

Table 2 Predicted results of the 1st on-service sample under different methods

Time	Actual RUL	The proposed method	Error	Traditional SVR method	Error	SbRP method	Error
4	7.475	8.133	0.658	9.708	2.233	9.114	1.639
5	6.475	7.534	1.059	8.557	2.082	8.226	1.751

6	5.475	6.997	1.522	7.397	1.922	7.196	1.721
7	4.475	5.888	1.413	6.376	1.901	6.072	1.597
8	3.475	4.266	0.791	4.819	1.344	4.762	1.287
9	2.475	3.560	1.085	3.994	1.519	3.368	0.893
10	1.475	2.446	0.971	2.724	1.249	2.360	0.885
11	0.475	2.211	1.736	1.618	1.143	1.200	0.725
e_{MAE}		1.154		1.674		1.312	
e_{RMSE}		1.206		1.718		1.371	

Table 3 Predicted results of the 2nd on-service sample under different methods

Time	Actual RUL	The proposed method		Traditional SVR method		SbRP method	
			Error		Error		Error
4	11.245	10.803	0.442	10.710	0.535	11.090	0.155
5	10.245	9.958	0.287	9.866	0.379	10.110	0.135
6	9.245	9.047	0.198	9.011	0.234	8.996	0.249
7	8.245	8.256	0.011	8.144	0.101	8.010	0.235
8	7.245	7.331	0.086	7.125	0.120	7.123	0.122
9	6.245	6.813	0.568	6.556	0.311	5.916	0.329
10	5.245	5.590	0.345	5.556	0.311	5.275	0.030
11	4.245	4.857	0.612	4.912	0.667	4.360	0.115
12	3.245	3.568	0.323	3.869	0.624	3.536	0.291
13	2.245	2.635	0.390	2.958	0.713	2.358	0.113
14	1.245	2.15	0.905	2.15	0.905	1.336	0.091
15	0.245	1.356	1.111	1.356	1.111	0.544	0.299
e_{MAE}		0.440		0.501		0.180	
e_{RMSE}		0.536		0.584		0.202	

Table 4 Predicted results of the 3rd on-service sample under different methods

Time	Actual RUL	The proposed method		Traditional SVR method		SbRP method	
			Error		Error		Error
4	14.237	11.341	2.896	11.212	3.025	12.636	1.601
5	13.237	10.553	2.684	10.454	2.783	12.229	1.008
6	12.237	9.754	2.483	9.670	2.567	10.872	1.365
7	11.237	8.883	2.354	8.861	2.376	9.308	1.929
8	10.237	8.093	2.144	7.953	2.284	7.985	2.252
9	9.237	7.541	1.696	7.356	1.881	6.993	2.244
10	8.237	6.607	1.630	6.33	1.907	6.400	1.837
11	7.237	5.796	1.441	5.949	1.288	5.238	1.999
12	6.237	5.076	1.161	5.206	1.031	4.439	1.798
13	5.237	4.322	0.915	4.407	0.830	3.642	1.595
14	4.237	3.768	0.469	3.768	0.469	2.746	1.491

15	3.237	2.880	0.357	2.880	0.357	1.869	1.368
16	2.237	2.309	0.072	2.309	0.072	0.998	1.239
17	1.237	1.854	0.617	1.854	0.617	0.561	0.676
18	0.237	1.419	1.182	1.419	1.182	0.569	0.332
e_{MAE}		1.473		1.511		1.516	
e_{RMSE}		1.709		1.771		1.604	

5.2 Application to GaAs-based semiconductor lasers

In this subsection, the proposed method is applied to a degradation dataset of GaAs laser devices, which is taken from Meeker and Escobar (1998) [47]. Degradation of the laser device is measured as the operation current, which increases over time. Degradation data from 15 testing units are collected over time with measure frequency 250 h and terminal time 4000 h, their curves are shown in Fig. 9. In this paper, a laser is assumed to have failed if the percentage of its operating current exceeds a predefined threshold level $FT=6\%$. Note that the observation time points are recorded at every 250 h, this leads to the inability to know the true life, i.e. the time arriving the failure threshold. Due to the reasonable degradation process of laser is an approximate linear degradation, the most possible failure time considered as the real one can be calculated by local linear regression method. We choose unit 5, 6, 7 and 13 as the testing samples and unit 1 and 15 as the on-service samples, while others as the training samples. Online prediction is performed for the two on-service samples by different methods.

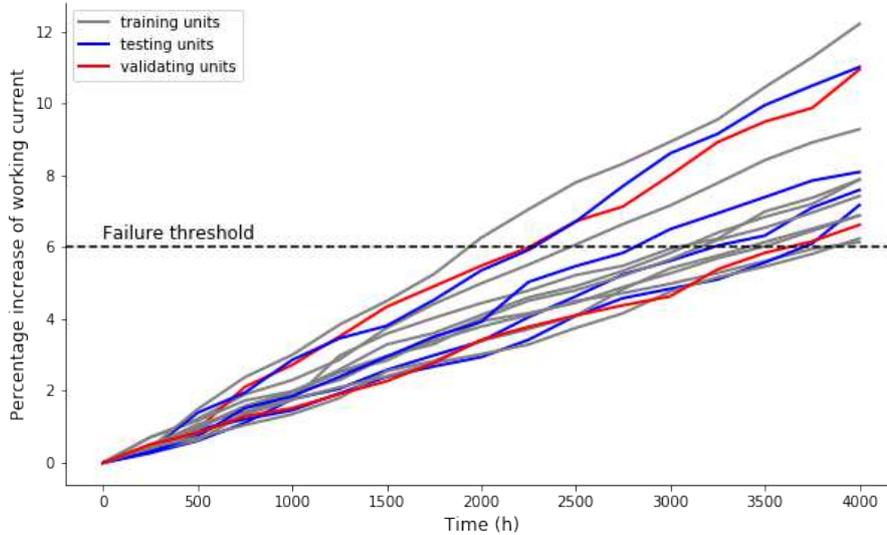


Fig. 9 The degradation curves of GaAs-based laser data

For the 4 testing units, we divide their respective degradation processes into 6 stages equally, i.e. $K=6$. Then the variation trends of the training error with the cluster sampling number M are shown clearly in Fig. 10. Obviously, as the M increasing, training errors are non-increasing, and finally converges to a definite level. Therefore, according to Fig. 10, $M=25$ is an appropriate value for cluster sampling in this study.

In the online prediction phase, the prediction results of the three on-service samples are shown in Fig. 11. For unit 1, 6 predictions were made in the whole life cycles, the proposed method has 4 prediction errors with the minimum (66.67%), while traditional SVR method has 2 (33.33%) and SbrP method has 0 (0%), results are listed in Table 5. And for unit 1, the MAE and RMSE of the proposed

method are all lower than the other two methods. For unit 15, 11 predictions were made in the whole life cycles, the proposed method has 7 prediction errors with the minimum (63.64%), while traditional SVR method has 2 (18.18%) and Sbrp method has 2 (18.18%), results are listed in Table 6. And for unit 15, the MAE and RMSE of the proposed method are all lower than the other two methods.

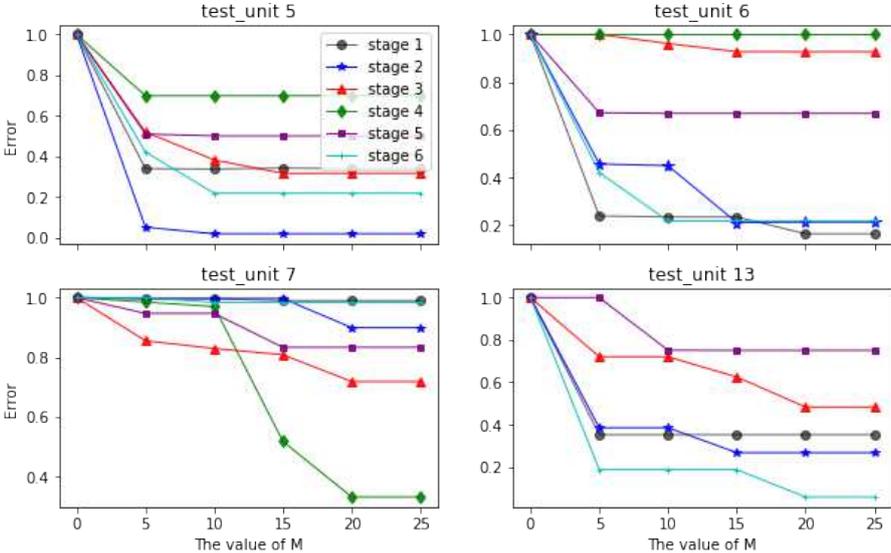


Fig. 10 The variation trend of the error in each stage of the testing samples

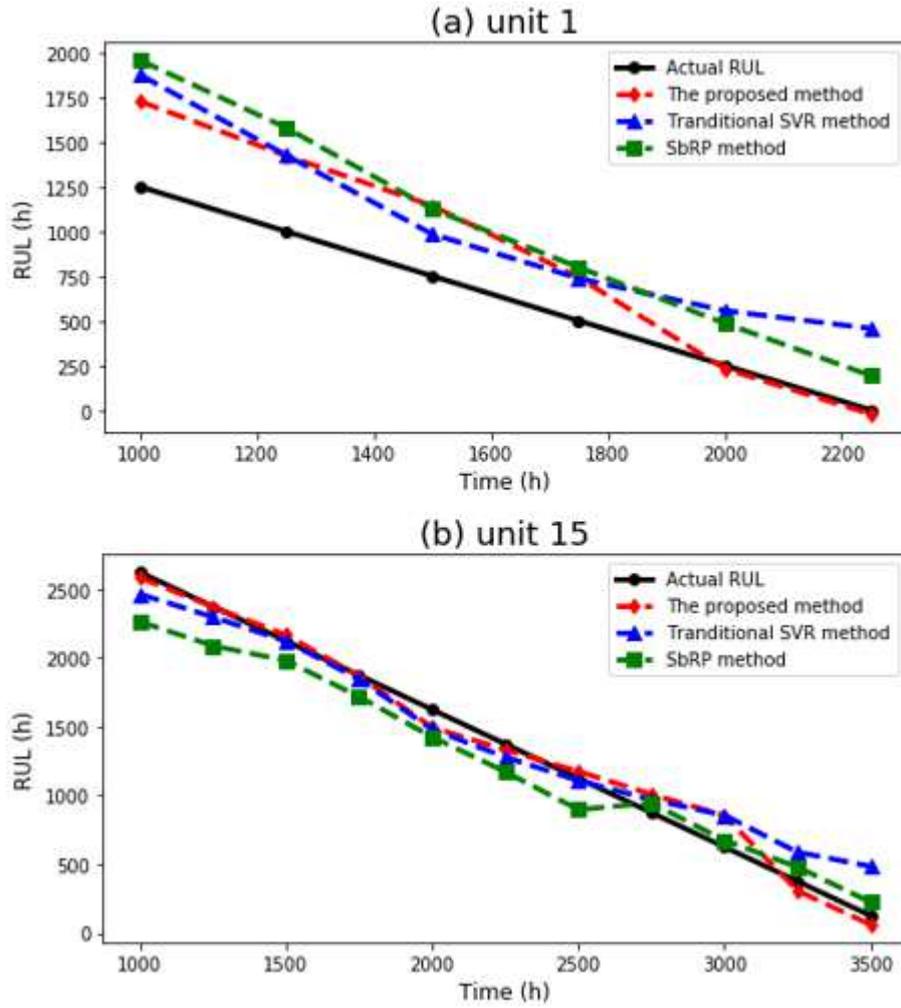


Fig. 11 The prediction results of the on-service samples of GaAs-based lasers data

Table 5 Predicted results of unit 1 under different methods

Time	Actual RUL	The proposed method	Error	Traditional SVR method	Error	SbRP method	Error
1000	1254	1732.047	478.047	1880.404	626.404	1961.817	707.817
1250	1004	1427.825	423.825	1430.111	426.111	1584.092	580.092
1500	754	1147.409	393.409	988.641	234.641	1129.177	375.177
1750	504	746.745	242.745	741.632	237.632	803.040	299.04
2000	254	235.553	18.447	560.118	306.118	489.945	235.945
2250	4	-19.096	23.096	460.363	456.363	196.332	192.332
e_{MAE}		263.262		381.212		398.401	
e_{RMSE}		322.161		405.679		439.764	

Table 6 Predicted results of unit 15 under different methods

Time	Actual RUL	The proposed	Error	Traditional SVR	Error	SbRP method	Error
------	------------	--------------	-------	-----------------	-------	-------------	-------

		method		method			
1000	2625	2585.433	39.567	2465.132	159.868	2264.916	360.084
1250	2375	2372.999	2.001	2298.788	76.212	2089.432	285.568
1500	2125	2167.782	42.782	2122.725	2.275	1981.971	143.029
1750	1875	1868.889	6.111	1846.795	28.205	1718.729	156.271
2000	1625	1494.354	130.646	1481.288	143.712	1425.827	199.173
2250	1375	1332.206	42.794	1279.236	95.764	1172.773	202.227
2500	1125	1174.916	49.916	1108.518	16.482	896.404	228.596
2750	875	1007.784	132.784	969.474	94.474	944.551	69.551
3000	625	854.025	229.025	855.643	230.634	673.313	48.313
3250	375	306.265	68.735	586.742	211.742	478.748	103.748
3500	125	59.504	65.496	483.418	358.418	223.257	98.257
e_{MAE}		73.623		128.890		172.256	
e_{RMSE}		97.200		164.504		194.584	

5.3 Results and discussion

From the above two subsections, the proposed method all performs better than traditional SVR method, the introduction of cluster sampling in SVR method can effectively reduce the training error. But we also have to be aware of that as the number of training samples increase, the proposed method sometimes performs not as good as SbRP method. Nonetheless, when the training samples are limited, the proposed method performs better than SbRP method.

6 Conclusion

This paper proposes a modified SVR method with health stage division, cluster sampling and similarity matching. Through health stage division and cluster sampling, the training error of each stage is non-increasing, and the whole training error is decreasing with the number of cluster sampling, until it finally converges to a definite level. Therefore, the drawback that the linearity of SVR model will increase with the increase of sample set can be solved effectively. Additionally, similarity matching is introduced to select the optimal parameters in the online prediction phase, it solves the limitation of a single SVR model. To demonstrate the advantages of the proposed method, a simulation dataset and the GaAs-based lasers dataset are applied to analyze. Results show that the proposed method is of better effectiveness than the traditional SVR method, and when the training samples are limited, the proposed method performs better than SbRP method. Despite the prediction results by SbRP method are sometimes more accurate than the proposed method, pointwise similarity measure as the key procedure of SbRP method often require extensive computation, and with the training samples increasing, it is much more time consuming than the proposed method. Consequently, the proposed method is of great significant, it extends the studies of SVR methods for RUL prediction, and is of effectiveness in engineering practice.

Authors' contributions

JW-Y designed the study. XH-Z performed the assays and prepared the manuscript. YF conducted the optimization and revised the manuscript. YC-C participated in discussing the results and put forward

some suggestions to improve the study. All authors read and approved the final manuscript.

Availability of data and materials

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This research is supported by Fundamental Research Funds for the Central Universities of China (Grant No. PA2019GDPK0048), and Anhui Province Postdoctoral Science Foundation Funded Project (Grant No. 2018B257).

Acknowledgements

Not applicable.

References

- [1] M G Pecht. Prognostics and health management of electronics. New Jersey: John Wiley, 2008.
- [2] G W Vogl, B A Weiss, M Helu. A review of diagnostic and prognostic capabilities and best practices for manufacturing. *Journal of Intelligent Manufacturing*, 2019, 30: 1-17.
- [3] Fuqiong Zhao, Mingjiang Xie, Zhigang Tian, et al. Integrated equipment health prognosis considering crack initiation time uncertainty and random shock. *Chinese Journal of Mechanical Engineering*, 2017, 30(6): 1383-1395.
- [4] Linxia Liao, F Kottig. Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. *IEEE Transactions on Reliability*, 2014, 63(1): 191-207.
- [5] G Curcuru, G Galante, A Lombardo. A predictive maintenance policy with imperfect monitoring. *Reliability Engineering & System Safety*, 2010, 95(9): 989-997.
- [6] Zhen Zhao, Fuli Wang, Mingxing Jia, et al. Predictive maintenance policy based on process data. *Chemometrics and Intelligent Laboratory Systems*, 2010, 103(2): 137-143.
- [7] Yaguo Lei, Naipeng Li, Liang Guo, et al. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing*, 2018, 104: 799-834.
- [8] 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.
- [9] Hong Pei, Changhua Hu, Xiaosheng Si, et al. Review of machine learning based remaining useful life prediction methods for equipment. *Journal of Mechanical Engineering*, 2019, 55(8): 1-13(in Chinese).
- [10] A Mosallam, K Medjaher, N Zerhouni. Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction. *Journal of Intelligent Manufacturing*, 2016, 27(5): 1037-1048.
- [11] F Cennarile, P Baraldi, E Zio. An evidential similarity-based regression method for the prediction of equipment remaining useful life in presence of incomplete degradation trajectories. *Fuzzy Sets and Systems*, 2018, 367: 36-50.
- [12] Zhengxin Zhang, Xiaosheng Si, Changhua Hu, et al. Degradation data analysis and remaining useful life estimation: a review on Wiener-process-based methods. *European Journal of Operational Research*, 2018, 271(3): 775-796.

- [13] Liang Huang, Junqiang Liu, Yingjie Gong. Multi-phase residual life prediction of engines based on Wiener process. *Journal of Beijing University of Aeronautics and Astronautics*, 2018, 44(5): 1081-1087(in Chinese).
- [14] M Guida, G Pulcini. The inverse gamma process: a family of continuous stochastic models for describing state-dependent deterioration phenomena. *Reliability Engineering & System Safety*, 2013, 120: 72-79.
- [15] Xiaolin Wang, N Balakrishnan, Bo Guo, et al. Residual life estimation based on bivariate non-stationary gamma degradation process. *Journal of Statistical Computation and Simulation*, 2015, 85(2): 405-421.
- [16] Weiwen Peng, Yanfeng Li, Yuanjian Yang, et al. Bayesian degradation analysis with inverse Gaussian process models under time-varying degradation rates. *IEEE Transactions on Reliability*, 2017, 66(1): 84-96.
- [17] Donghui Pan, Jiabao Liu, Jinde Cao. Remaining useful life estimation using an inverse Gaussian degradation model. *Neurocomputing*, 2016, 185: 64-72.
- [18] Daming Lin, V Makis. Recursive filters for a partially observable system subject to random failure. *Advances in Applied Probability*, 2003, 35(1): 207-227.
- [19] 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.
- [20] Tao Tao, E Zio, Wei Zhao. A novel support vector regression method for online reliability prediction under multi-state varying operating conditions. *Reliability Engineering & System Safety*, 2018, 177: 35-49.
- [21] Lei Xiao, Xiaohui Chen, Xinghui Zhang, et al. A novel approach for bearing remaining useful life estimation under neither failure nor suspension histories condition. *Journal of Intelligent Manufacturing*, 2017, 28(8): 1893-1914.
- [22] Zhigang Tian. An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring. *Journal of Intelligent Manufacturing*, 2012, 23(2): 227-237.
- [23] E Zio, F D Maio. A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system. *Reliability Engineering & System Safety*, 2010, 95(1): 49-57.
- [24] V N Vapnik. *The Nature of Statistical Learning Theory*. New York: Springer, 1999.
- [25] A Widodo, B S Yang. Machine health prognostics using survival probability and support vector machine. *Expert Systems with Applications*, 2011, 38: 8430-8437.
- [26] T H Loutas, D Roulias, G Georgoulas. Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic e-support vectors regression. *IEEE Transactions on Reliability*, 2013, 62(4): 821-832.
- [27] T Benkedjouh, K Medjaher, N Zerhouni, et al. Remaining useful life estimation based on nonlinear feature reduction and support vector regression. *Engineering Applications of Artificial Intelligence*, 2013, 26(7): 1751-1760.
- [28] T Benkedjouh, K Medjaher, N Zerhouni, et al. Health assessment and life prediction of cutting tools based on support vector regression. *Journal of Intelligent Manufacturing*, 2015, 26(2): 213-223.
- [29] Jie Liu, V Vitelli, E Zio, et al. A novel dynamic-weighted probabilistic support vector regression-based ensemble for prognostics of time series data. *IEEE Transactions on Reliability*, 2015, 64(4): 1203-1213.
- [30] Jie Liu, E Zio. An adaptive online learning approach for support vector regression: online-SVR-FID. *Mechanical Systems and Signal Processing*, 2016, 76-77: 796-809.
- [31] E Fumeo, L Oneto, D Anguita. Condition based maintenance in railway transportation systems based on big data streaming analysis. *Procedia Computer Science*, 2015, 53: 437-446.
- [32] R Khelif, B Chebel-Morello, S Malinowski, et al. Direct remaining useful life estimation based on support vector regression. *IEEE Transactions on Industrial Electronics*, 2017, 64(3): 2276-2285.
- [33] A Soualhi, K Medjaher, N Zerhouni. Bearing health monitoring based on Hilbert-Huang transform, support vector machine, and regression. *IEEE Transactions on Instrumentation and Measurement*. 2015, 64: 52-62.
- [34] Wei Zhao, Tao Tao, E Zio, et al. A novel hybrid method of parameters tuning in support vector regression for

- reliability prediction: particle swarm optimization combined with analytical selection. *IEEE Transactions on Reliability*, 2016, 65(3): 1393-1405.
- [35] Jingwen Wei, Guangzhong Dong, Zonghai Chen. Remaining useful life prediction and state of health diagnosis for Lithium-ion batteries using particle filter and support vector regression. *IEEE Transactions on Industrial Electronics*, 2018, 65(7): 5634-5643.
- [36] V N Vapnik, S Golowich, A Smola. Support vector machine for function approximation regression estimation and signal processing. *Advances in Neural Information Processing System*, 1996, 9: 281-287.
- [37] A L Dryver, U Netharn, D R Smith. Partial systematic adaptive cluster sampling. *Environmetrics*, 2012, 23(4): 306-316.
- [38] E Tipton. Stratified sampling using cluster analysis: a sample selection strategy for improved generalizations from experiments. *Evaluation Review*, 2013, 37(2): 109-139.
- [39] Tianyi Wang, Jianbo Yu, D Siegel, et al. A similarity-based prognostics approach for remaining useful life estimation of engineered systems. In: 2008 International conference on prognostics and health management (PHM), Denver, CO, 53-58. Oct 06-09
- [40] Mingyi You, Guang Meng. A framework of similarity-based residual life prediction approaches using degradation histories with failure, preventive maintenance, and suspension events. *IEEE Transactions on Reliability*, 2013, 62: 127-134.
- [41] Yingchao Liu, Xiaofeng Hu, Wenjuan Zhang. Remaining useful life prediction based on health index similarity. *Reliability Engineering and System Safety*, 2019, 185: 502-510.
- [42] Xiaosheng Si, Zhengxin Zhang, Changhua Hu, Data-driven remaining useful life prognosis techniques. *Springer Berlin Heidelberg*, 2017.
- [43] S Sankararaman. Significance, interpretation, and quantification of uncertainty in prognostics and remaining useful life prediction. *Mechanical Systems and Signal Processing*, 2015, 52-53: 228-247.
- [44] Yaogang Hu, Hui Li, Xinglin Liao, et al. A probability evaluation method of early deterioration condition for the critical components of wind turbine generator systems. *Mechanical Systems and Signal Processing*, 2016, 76-77: 729-741.
- [45] J K Kimotho, C Sondermann-woelke, T Meyer, et al. Machinery prognostic method based on multi-class support vector machines and hybrid differential evolution–particle swarm optimization. *Chemical Engineering Transactions*, 2013, 33: 619-624.
- [46] Qing Zhang, P W T Tse, Xiang Wan, et al. Remaining useful life estimation for mechanical systems based on similarity of phase space trajectory. *Expert Systems with Applications*, 2015, 42(5): 2353-2360.
- [47] W Meeker, L A Escobar. Statistical methods for reliability data. Probability and Statistics, New York: Wiley, 1998.

Figures

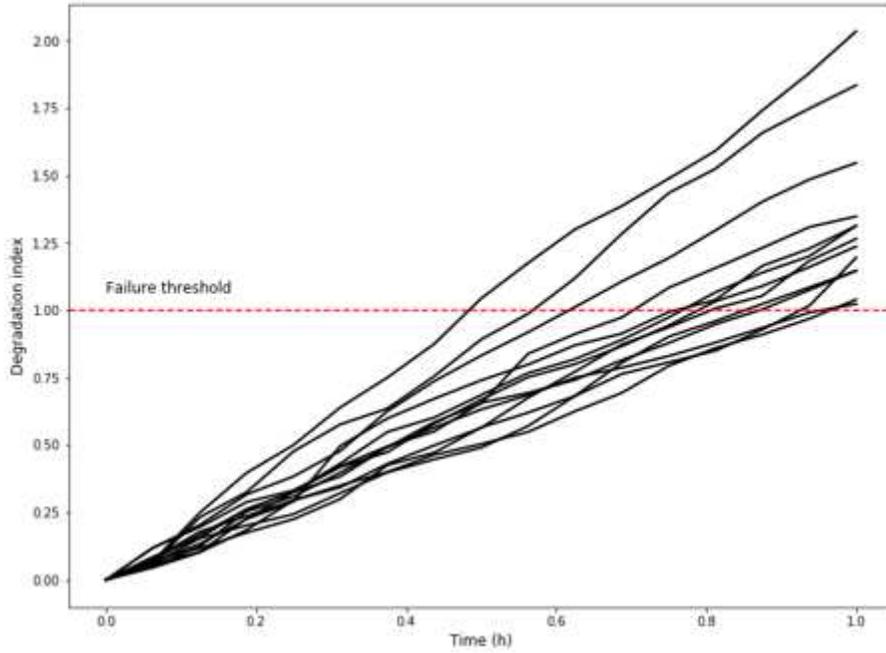


Figure 1

The degradation curves of the same category of equipment

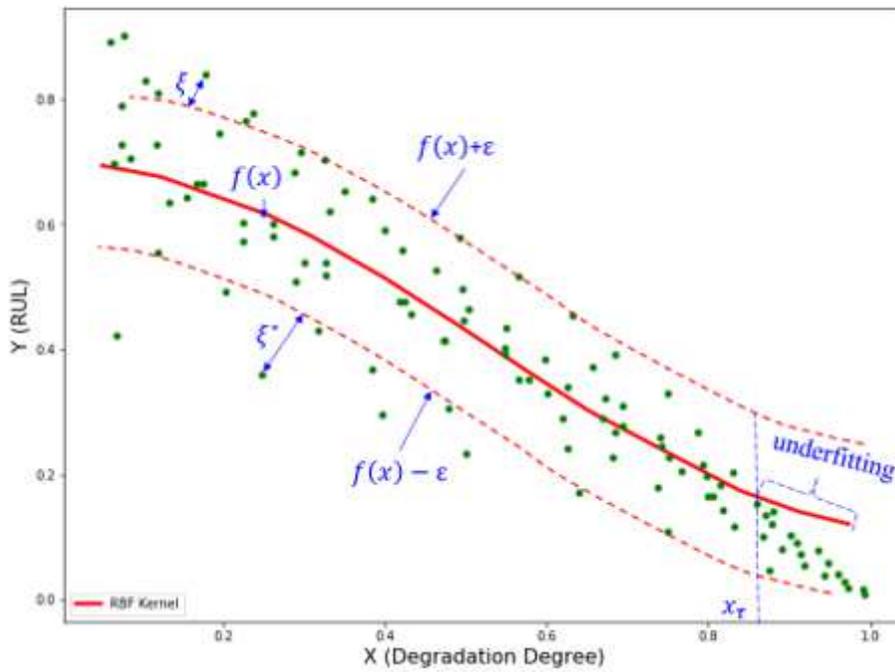


Figure 2

The problem of RUL prediction by SVR method

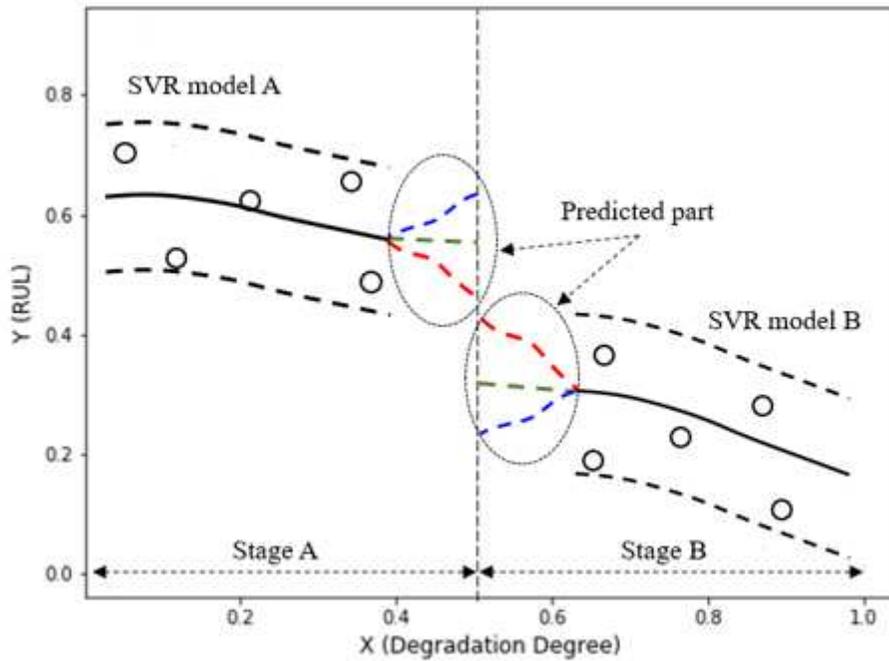


Figure 3

The issue of just simply divide the degradation into several stages and construct the corresponding SVR models

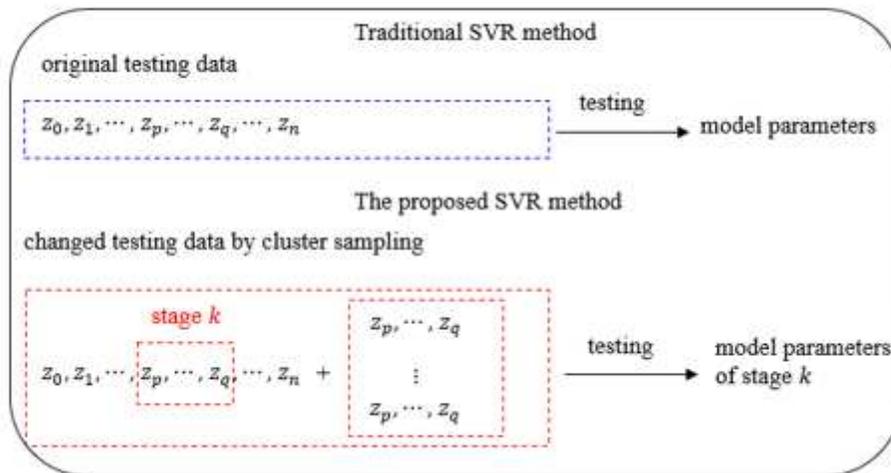


Figure 4

The difference of model construction between the proposed SVR method and the traditional SVR method

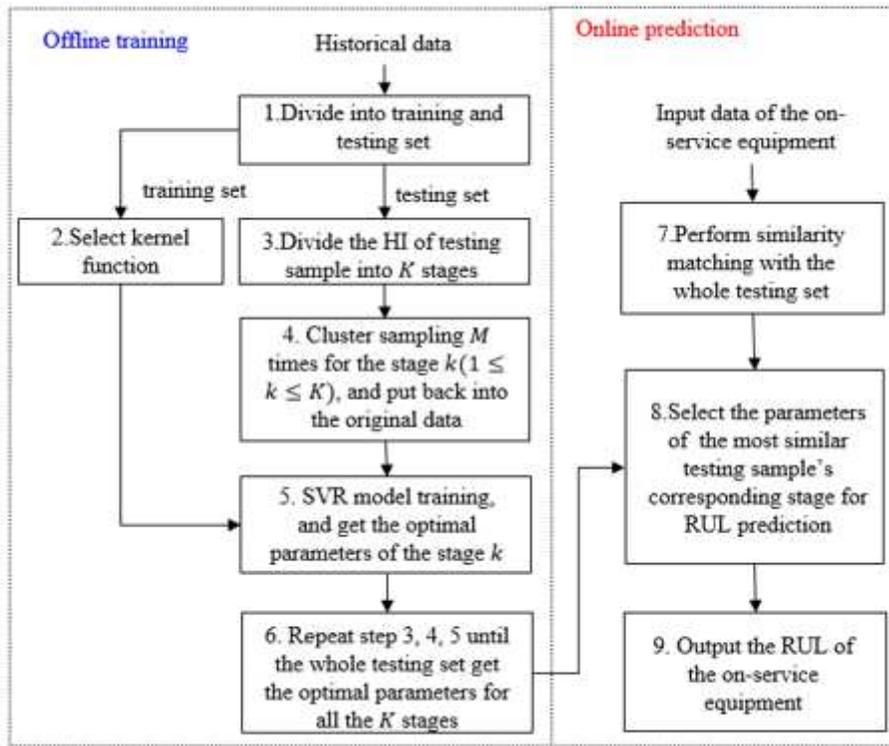


Figure 5

The main flowchart of the proposed method

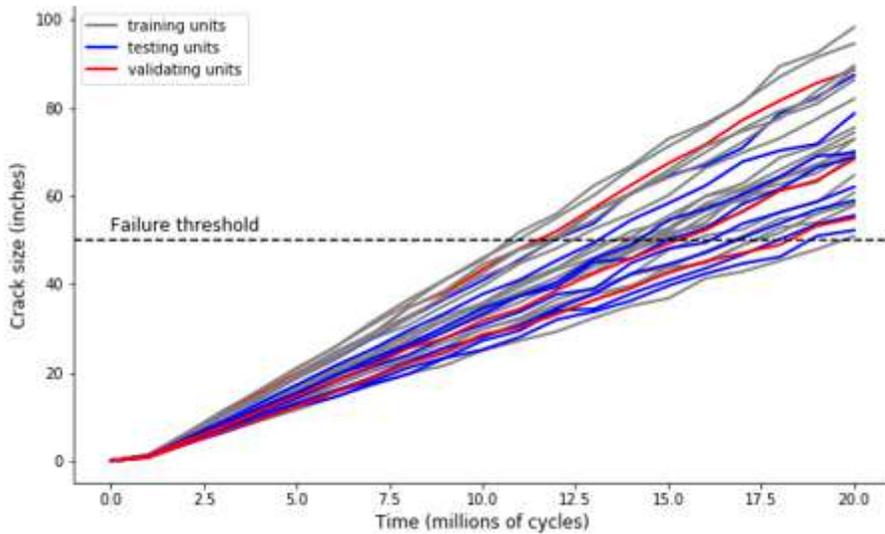


Figure 6

The degradation curves of the simulation dataset

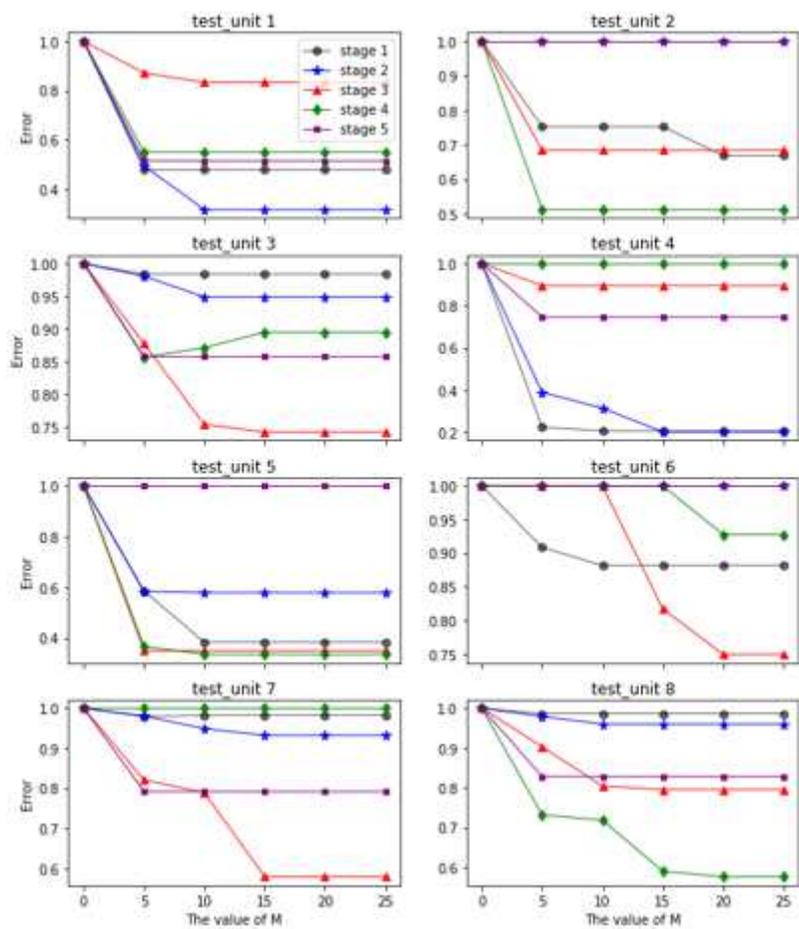


Figure 7

The variation trend of the error in each stage of the testing samples

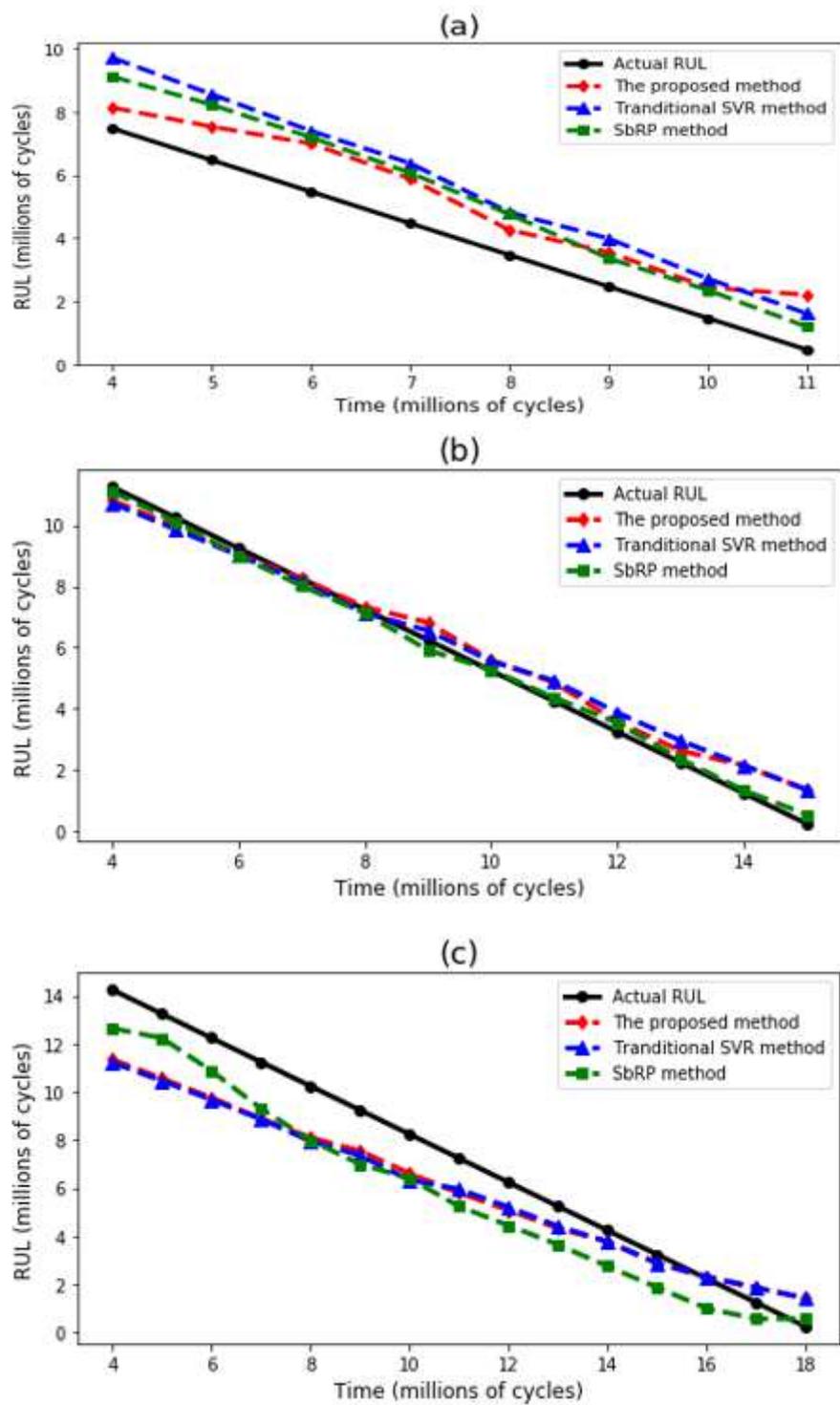


Figure 8

The prediction results of the on-service samples of the simulation data

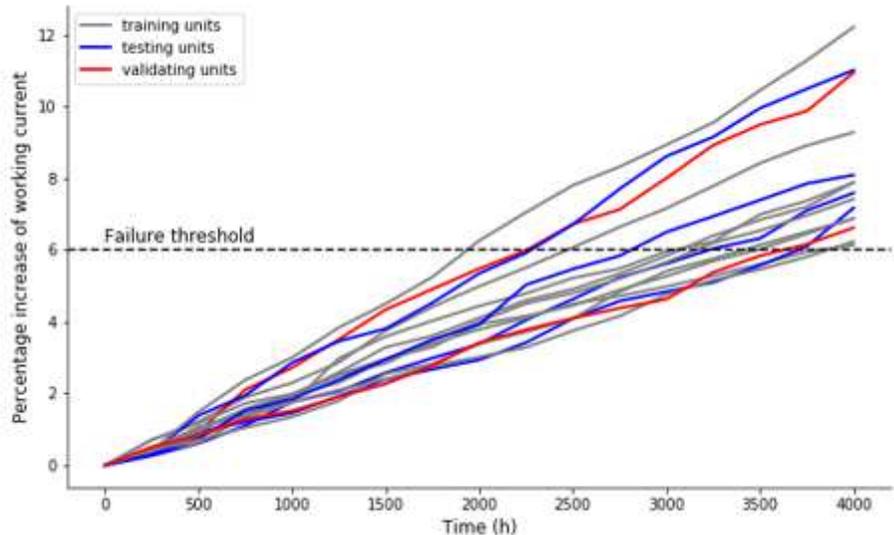


Figure 9

The degradation curves of GaAs-based laser data

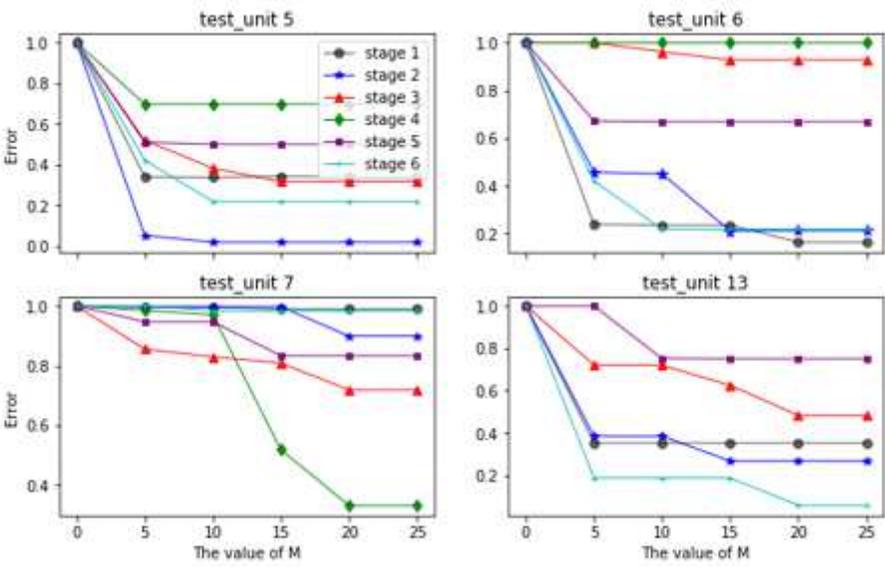


Figure 10

The variation trend of the error in each stage of the testing samples

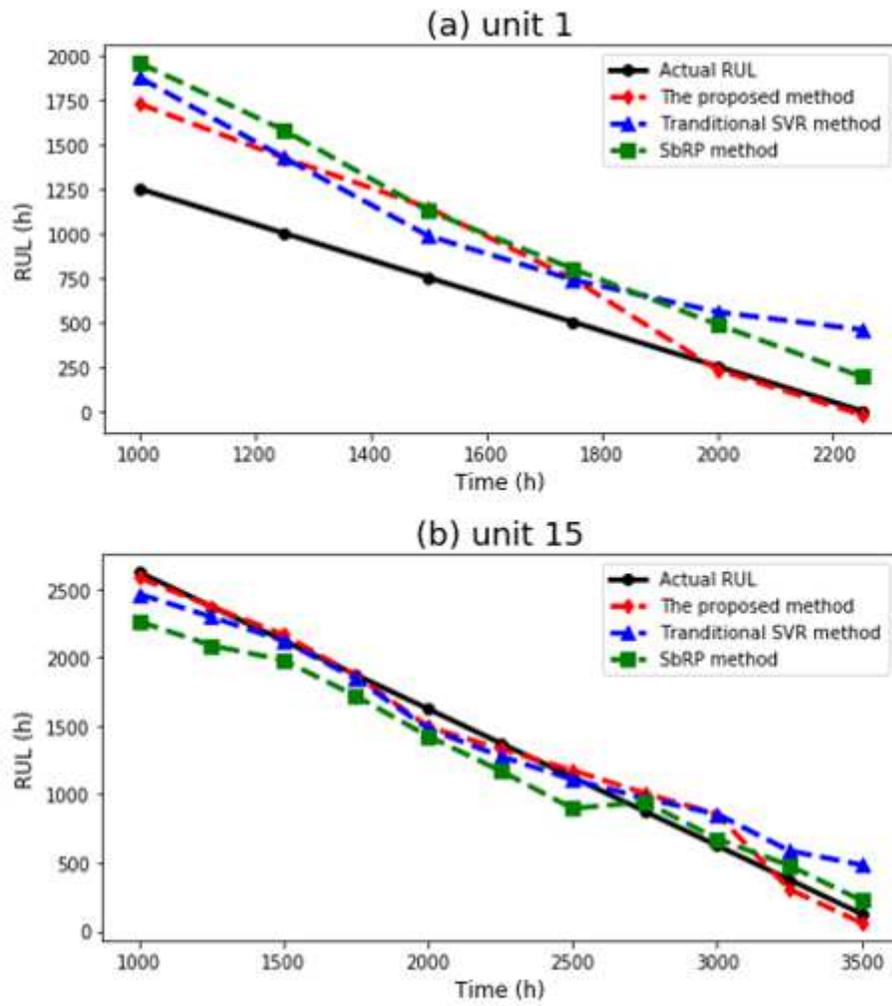


Figure 11

The prediction results of the on-service samples of GaAs-based lasers data