Ensemble Learning for Predicting Degradation under Time-varying Environment

Lizhi Wang[ae], Dawei Lu[b], Xiaohong Wang[c], Rong Pan[d*], Zhuo Wang[c]

a Unmanned System Institute, Beihang University, Beijing 100191, China

b The 52nd Research Institute, CETHIK Group Co.Ltd., Hangzhou 310012, China

c School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China;

d School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ, USA

e Key Laboratory of Advanced Technology of Intelligent Unmanned Flight System of Ministry of Industry and Information Technology, Beijing 100191, China

ABSTRACT: Product lifetime prediction is challenging when the product is subject to a time-varying operational environment. Most of the existing studies use some functions to explicitly specify the relationship between degradation parameters and environmental conditions so as to reveal how the degradation process evolves over time. However, in many applications, the assumptions needed for establishing these functions cannot be validated in engineering practice or they cannot accurately model the entire underlying degradation mechanism. In contrast to previous work, the focus of our study is placed on product degradation prognosis by implementing an ensemble learning method. This method combines the stochastic process modeling approach and the machine learning approach, taking advantage of these approaches to gain a more accurate and stable degradation prediction. The proposed method is demonstrated by some simulation examples and by a case study of lithium-ion battery accelerated degradation test. Both the simulation study and the real case verify the superiority of the proposed method. The case study indicates that the ensemble learning method can further help to effectively manage the energy storage and energy distribution of battery packs.

Keywords: Degradation process, stochastic process, machine learning, ensemble model, Li-ion battery.

* Corresponding author, rong.pan@asu.edu.

1. Introduction

Lifetime prediction is crucial in manufacturing, automotive, aerospace, and many other industries for supporting an effective decision-making process.[1] For a highly reliable product, accelerated degradation tests are frequently employed in this process to reduce the test time and to infer the product's lifetime under its normal operating condition.[2][3][4]

Many studies of degradation-based lifetime prediction had been presented in literature; however, these studies are often based on laboratory testing data obtained under a standard or controlled environment and ignore the complicated time-varying operational environment that products will experience in reality.[5][6][7][8] In practice, many environmental stress variables are indeed random and the product's degradation mechanism could be environmental sensitive. Ignoring the effect of stress variation can lead to significant prediction error. A few studies had tried to utilize field data and real environmental conditions to minimize such deficiency. Kharoufeh et al.[9] presented a method by considering the case of a semi-Markovian environmental model. Gebraeel and Pan[10] introduced a degradation modeling framework with the consideration of time-varying environment. Pan[11] and Wang et al.[12] introduced a calibration factor to the lifetime predictive model to denote the environmental influence. Qu et al.[13] proposed a mission profile-based lifetime predictive method to estimate the lifetime and reliability performance of light emitting diodes in field operations. Thomas [14] adjusted the lifetime prediction model of lithium-ion cell in order to make it more realistic for the dynamic use condition.

One commonality of these existing methods is that the dependence of the degradation processes on environmental conditions is explicitly specified using some presumed functions, which are modeling how the degradation process evolves over time.[15] However, in many applications, such well-defined parametric models cannot be found or they cannot accurately capture the entire underlying degradation mechanism. This problem is particularly true for a product operated under a time-varying environmental condition and/or with multiple failure modes. To tackle these challenges, in this paper we incorporate dynamic environment factors into degradation modeling and treat them as accelerated stress factors. Given the complexity of degradation mechanism, we develop the models for both stress process prediction and for product lifetime prediction.

Generally speaking, existing degradation modeling methods can be summarized into three main categories – the stochastic process modeling approach, the regression modeling approach, and the machine learning approach. Typical stochastic process models include Wiener process model[16][17][18], Gamma process model[19][20], and their variants[21][22]. However, estimating parameters in these models could be a difficult task and estimation results are often unsatisfactory.[23][24] The Bayesian inference approach has been proposed to solve these

problems. It has a considerable advantage over traditional inferential methods because the modern computational Bayesian method can provide more precise results.[25][26] For example, Mukhopadhyay et al.[27] proposed a Bayesian method for reliability estimation using multi-stress accelerated life testing of series systems. Ali et al.[28] combined Gaussian process modeling and Bayesian inference method to estimate the remaining useful life of insulated gate bipolar transistor devices.

The parametric methods mentioned above all require the specification of a product's lifetime-stress acceleration model. However, when the product's degradation mechanism becomes quite complicated, developing an accurate acceleration model is nearly impossible[29]. On the other hand, machine learning methods, which purely rely on observed data, can avoid such restriction. Some popular machine learning methods, such as Support Vector Machine (SVM)[30], Backpropagation Network (BP)[31], and Recurrent Neural Network[32], have been utilized to train accelerated degradation models and to predict a product's operational lifetime.

As discussed above, both Bayesian inference method and machine learning method are useful for lifetime prediction, yet neither of them is perfect. Bayesian method has the merits of uncertainty reduction and being able to use prior information accumulated from field observations, and it can solve small sample problem.[33] However, this method is computationally complicated, and the result could be highly unstable. Machine learning methods are often more versatile and adapted to data, but its use is limited by the large data size requirement and the overfitting problem[34]. And both methods could have a poor generalization property when the training dataset has a great degree of volatility. However, it has been seen that these two methods could be complementary to each other in some way[35], thus by combining them together the deficiency of each method can be somewhat compensated.

Ensemble learning is commonly used for data-based machine learning tasks. Previous studies have shown that an ensemble learning algorithm often outperforms any of the constituent learning algorithms alone.[36] It has applied on the prediction of energy use[37][38], gas turbine engine degradation prediction[39], and the remaining useful life prediction of lithium-ion batteries[40], etc. Therefore, we believe that it is also feasible to synthesize Bayesian inference and machine learning methods to improve the performance of degradation-based lifetime prediction.

In this paper, an ensemble learning-based lifetime predictive approach is proposed with the consideration of a product's time-varying operational environment. The Bayesian inference method and two popular machine learning algorithms (SVM and BP neural network) are chosen to construct the predictive sub-models. The remainder of this paper is organized as follows. In Section 2, the general ensemble process is introduced. A simulated example is

developed in Section 3 to verify the effectiveness of the proposed method. In Section 4, a case study is provided to demonstrate the validity of the proposed method in battery lifetime prediction. Finally, conclusions are drawn in Section 5.

2. Methodology

2.1 Bagging Method

Bagging, which is bootstrap aggregation, is a type of ensemble learning. It usually consists of a class of algorithms which build several instances of a black-box estimator on random subsets of the original training set and aggregate their individual predictions to form the final prediction. Bagging algorithm is often a straightforward way to improve a predictive model without a significant adjustment of the underlying base algorithm.

Data resampling is a necessary step in developing the prediction framework with bagging. The key method of data resampling is bootstrapping, which is a statistical technique operating on measurement data to infer the level of uncertainty on any parametric estimator. It assesses the statistical properties of an estimator, such as its mean or variance, by utilizing repeated random sampling of the same dataset. The core idea of bootstrapping is to generate many replicas of a dataset by randomly selecting *N* observations with replacement, where *N* is the dataset size. Then we train models on these replications to find their predicted responses. After resampling data, there is a certain percentage of observations left from each resampling process. These "out-of-bag" observations can be used as the validation dataset to assess predictive power and feature importance.

Another essential part of bagging is to form a model ensemble, which is the last step in any ensemble-based system. The strategy used in this step depends in part on the type of algorithm used as ensemble member. Usually, for regression, the combining method is to average outputs, and for classification, it is majority voting or weighted majority voting.

It is known that ensemble prediction methods are generally better than using a single prediction method. Breiman[41] presented the following explanation when he proposed the ensemble tree method. Suppose there is a regression problem. Let f denote the ground-truth function and h(x) denote a learner trained from the bootstrap distribution D_{bs}. The aggregated learner generated by Bagging is

$$\mathrm{H(x)} = E_{D_{bs}}[h(x)] \tag{1}$$

With the inequality $(\mathrm{E[X]})^2 \leqslant \mathrm{E[X^2]}$, we have

$$E^2[f(x) - h(x)] = (\mathrm{f(x) - H(x)})^2 \leq E[(f(x) - h(x))^2] \tag{2}$$

We can see that the squared error of the aggregated learner is smaller than the expected squared error of individual learner. The size of difference depends on

$$(E[h(x)])^2 \leq E[h(x)^2] \tag{3}$$

This explains why Bagging is an effective approach to reducing errors associated with unstable learners.

2.2 Predictive Model Ensemble

The basic scheme of the proposed model ensemble will be applied to accelerated degradation data and time-varying environmental data as depicted in Figure 1. It comprises three steps – data processing, sub-model construction, and model ensemble. Essentially, Bayesian regression and machine-learning methods are used to build sub-models. The former method is chosen for its merit of uncertainty reduction and its ability of utilizing the prior information accumulated from field observations; while the latter one is chosen for its ductility in data fitting. At last, the outputs from these sub-models will be combined, according to the ensemble learning theory, to obtain the final prediction.

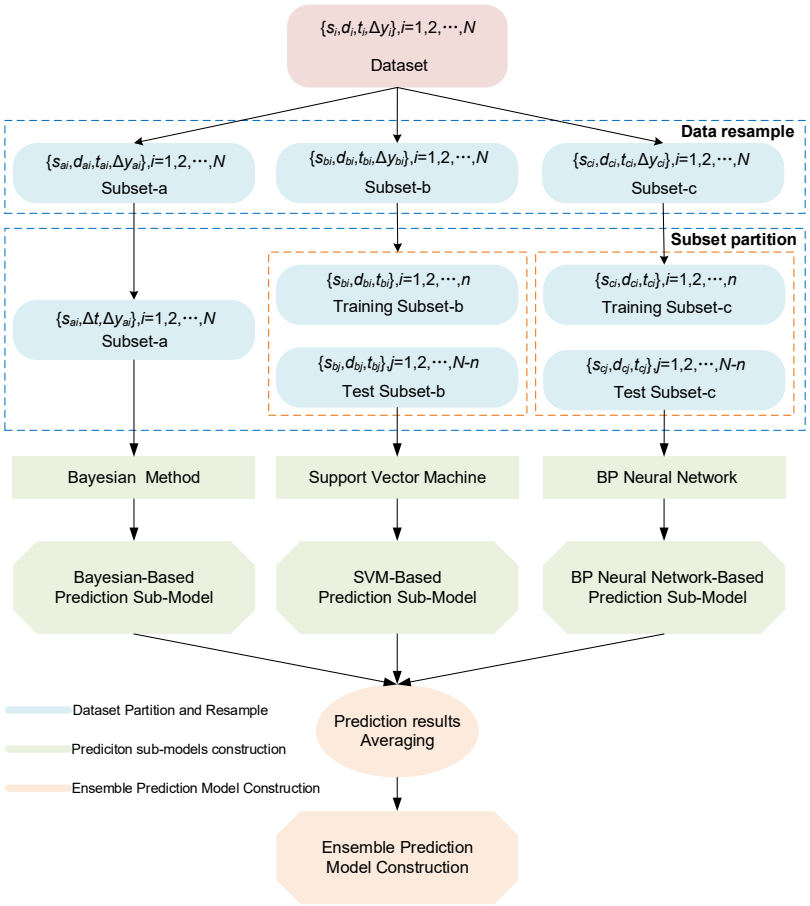

Figure 1 Basic Scheme of Ensemble Predictive Method

2.2.1 Dataset Partition and Resampling

To apply the bagging method, data resampling is used to obtain multiple training and testing datasets. Suppose there are N accelerated degradations and the time-varying

environmental observations are as $\{s_i, t_i, \Delta y_i\}$, i=1, 2,…,N, where s_i=[s_{i1},s_{i2},…s_{im}] are the stress measurements of m stress variables, t_i is the observation time and Δy_i is the degradation increment (Δy_i=y_{i+1} - y_i).

Based on the bootstrap method, we draw N out of N observations with replacement from the dataset of observations three times to obtain the subsets for sub-models (see Figure 1). The subsets are denoted as $\{s_{qi}, t_i, \Delta y_{qi}\}$, i=1, 2,…,N, q= a, b, c, where a, b, and c denote three subsets.

When adopting machine-learning methods, a dataset must be divided into a training set and a test set. The training set is a dataset used to fit a model. The test set, which is independent of the training set but follows the same probability distribution, is used to assess the performance of the trained model. Thus, the last two subsets are further partitioned with a proportion of 75/25 by random sampling, where n is the size of training subset. The full data partitioning and resampling process are shown in Figure 1.

2.2.2 Predictive sub-model construction

(1) Bayesian-based Stochastic Process Sub-model

Suppose a stochastic process model such as the Wiener process model or the Gamma process model is sufficient for describing the degradation process, the model parameters can be estimated by the Bayesian inference method.

Assume that the degradation process follows a Wiener process, then the degradation observations are modeled by

$$Y(t) = \sigma B(t) + d(s)t + y_0 \tag{4}$$

where $Y(t)$ is the performance degradation observation of a product at time t, σ is the diffusion coefficient, $B(t)$ is the standard Brownian motion with mean zero and variance t, y_0 is a known initial value of product performance, and the drift coefficient $d(s)$ represents the degradation rate of the product and it is a function of stress variables s.[12] It is used for incorporating stress-induced degradation acceleration into the degradation model.

Since the stress we consider here is time-varying, Eq. (4) can be re-written as

$$Y(t) = \sigma B(t) + \int_0^t d(s,\tau)d\tau + y_0 \tag{5}$$

Known from the property of Wiener process, degradation increment ΔY during a time interval Δt follows a normal distribution; i.e.,

$$\Delta \mathrm{Y} \sim \mathrm{N}(\int_t^{t+\Delta t} d(s,\tau)d\tau, \sigma^2 \Delta \mathrm{t}) \tag{6}$$

If we divide [0, t] into n intervals with the length of Δt and suppose the s_i remains constant in every interval, we can obtain that

$$\int_0^t d(s,\tau)d\tau \approx \sum_{i=1}^{n} d(s_i)\Delta t \tag{7}$$

Then Eq. (6) becomes

$$\Delta Y \sim N(d(s_i)\Delta t, \sigma^2 \Delta t) \tag{8}$$

Here, since multiple time-varying environmental stress variables are considered, the models that can accumulate multiple stress effects are needed. A general linear model is adopted as the acceleration model of the degradation rate $d(s_i)$. It is given by

$$d(s_i) = \exp[\beta_0 + \sum_{j=1}^{m} \beta_j \varphi(s_{ij})] \tag{9}$$

where $\varphi(s_{ij})$ is a given function of the stress factor s_j at the i-th interval or a transformation of stress factor, the coefficient β_j is the effect of this stress factor, and $d(s_i)$ is an exponential function of the sum of individual stress effects.

We input the stresses s_{ai} and a set of degradation-increment data Δy_{ai} from subset-a. The following regression equations can be obtained from Eqs. (4) and (9):

$$E(\Delta y_{ai}(t)) = exp(\beta_0 + \sum_{j=1}^{m} \beta_j \varphi(s_{aij}))\Delta t \tag{10}$$

$$ln(E(\Delta y_{ai})) = \sum_{j=1}^{m} \beta_j \varphi(s_{aij}) + \beta_0 + ln(\Delta t) \tag{11}$$

The Bayesian method and a Markov chain Monte Carlo (MCMC) algorithm are used to estimate the unknown parameters (β_0, β_1, …, β_m and σ) in the accelerated degradation process model.

By the Bayesian nature, parameters are treated as random variables, and their probabilistic models are obtained by posterior distributions. So, with the integrated system of degradation data, the posterior distribution of unknown parameters can be written as

$$\begin{aligned} &\pi(\Theta|D) \\ &= \pi(\beta_0, \beta_1, \dots, \beta_m, \sigma^2 | \Delta y_a, s_a, \Delta t) \\ &\propto \prod_{i=1}^{n} \left(\frac{1}{\sigma^2 \Delta t}\right)^{\frac{1}{2}} exp\left(-\frac{(\Delta y_{ai} - d(s_{ai}))^2}{2\sigma^2 \Delta t}\right) \pi(\beta_0, \beta_1, \dots, \beta_m, \sigma^2) \end{aligned} \tag{12}$$

The prior distributions of β_0, β_1, …, β_m and σ must be defined. We assume that the prior distribution of β_j is a normal distribution and the prior distribution of σ is an inverse gamma distribution, such as

$$\beta_j \sim N(\mu_{\beta_j}, \varepsilon_j^2),\ \sigma^2 \sim IGa(a, b)$$

The hyperparameters in these prior distributions can be determined from past knowledge or be elicited from domain experts. They can also be chosen according to some principles, such as symmetry or maximizing entropy for some given constraints.[42]

Now, Eq. (12) becomes

$$\begin{aligned} &\pi(\Theta|D) \\ &= \pi(\beta_0, \beta_1, \dots, \beta_m, \sigma^2 | \Delta y_a, s_a, \Delta t) \\ &\propto \prod_{i=1}^{n} \left(\frac{1}{\sigma^2 \Delta t}\right)^{1/2} \exp\left(-\frac{(\Delta y_{ai} - d(s_{ai}))^2}{2\sigma^2 \Delta t}\right) \left(\prod_{j=0}^{m} \phi\left(\frac{\beta_j - \mu_{\beta_j}}{\varepsilon_j}\right)\right) \frac{b^a}{\Gamma(a)} \left(\frac{1}{\sigma^2}\right)^{a+1} \exp\left(\frac{b}{\sigma^2}\right) \end{aligned} \tag{13}$$

Finally, after the posterior distributions of all required parameters are obtained by MCMC, the estimated values of parameters can be calculated. Then the degradation rate $d(s_{ai})$

subjected to specific stresses $s_{i1},\ldots,s_{im}$ can be obtained by Eq. (9) and the future degradation status can be predicted with the degradation model.

(2) SVM-based Predictive Sub-model Construction

The support vector machine (SVM) is a powerful machine learning algorithm introduced by Vapnik[43]. This algorithm is developed based on statistical learning theory and structural risk minimization principle. It is widely used for classification and regression analysis. When using SVM to solve a regression problem, the basic idea is to map the data x into a higher-dimensional feature space, via a nonlinear mapping, and perform linear regression in this space.[44] In this paper, the SVM algorithm is utilized to construct a predictive sub-model.

First, obtain the training and testing dataset, subset-b. Take both the time-varying environmental data, s_{bi}, and time, t, as input vectors, and collect degradation data, d_{bi}, as the output vector of SVM. The input and output data structures are constructed as follows:

$$Input = \begin{bmatrix} s_{b1} & t_{b1} \\ s_{b2} & t_{b2} \\ \cdots & \cdots \\ s_{bn} & t_{bn} \end{bmatrix} = \begin{bmatrix} S_{b11} & S_{b12} & \cdots & S_{b1m} & t_{b1} \\ S_{b21} & S_{b22} & \cdots & S_{b2m} & t_{b2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ S_{bn1} & S_{bn2} & \cdots & S_{bnm} & t_{bn} \end{bmatrix}, Output = \begin{bmatrix} d_{b1} \\ d_{b2} \\ \cdots \\ d_{bn} \end{bmatrix}$$

Then, normalize the input and output data by (14) and (15):

$$\hat{s}_{bi} = \frac{s_{bi} - min_i(s_{bi})}{max(s_{bi}) - min_i(s_{bi})}, \hat{t}_{bi} = \frac{t_{bi} - min_i(t_{bi})}{max(t_{bi}) - min_i(t_{bi})} \quad (14)$$

$$\hat{d}_{bi} = \frac{d_{bi} - min_i(d_{bi})}{max(d_{bi}) - min_i(d_{bi})} \quad (15)$$

where $\hat{s}_{bi}$, $\hat{t}_{bi}$ and $\hat{d}_{bi}$ are, respectively, the i-th normalized input and output data point.

Next, using the principle of SVM, the regression function is expressed as

$$f(s_{bi}, t_{bi}) = \sum_{j=1}^{N} (\alpha_i - \alpha_i^*) K((s_{bi}, t_{bi}), (s_{bj}, t_{bj})) + b \quad (16)$$

where α_i and α_i^* are Lagrange multipliers, and $K((s_{bi}, t_{bi}), (s_{bj}, t_{bj}))$ is the kernel function, which has an important effect on the generalization performance of SVM.

To guarantee effective prediction performance in (16), the Lagrange multipliers and kernel function should be carefully chosen. Here we choose the radial basis function to be kernel function. Based on the Karush-Kuhn-Tucker (KKT) condition, Lagrange multipliers can be defined with a penalty constant (C). So, the penalty constant and kernel parameters (i.e., the variance in the radial basis function) must be initialized.

Then, a parameter-optimization algorithm, such as the grid searching method or Particle Swarm Optimization (PSO), etc., can be utilized to obtain the optimized hyper-parameters (e.g., C), with several iterations of SVM model training and hyper-parameter updating to achieve the desired performance.

Finally, the SVM-based predictive sub-model can be built by training with the optimized hyper-parameters.

(3) BP Neural Network-based Predictive Sub-model Construction

A Back Propagation (BP) neural network is a multilayer feedforward network, consisting of input, hidden, and output layers. The error inverse feedback principle is adapted to continuously revise weight coefficients of each layer to improve its self-learning effectiveness and prediction accuracy. In this paper, BP neural network is built as a predictive sub-model for lifetime prediction.

The structure of BP neural network is shown in Figure 2.

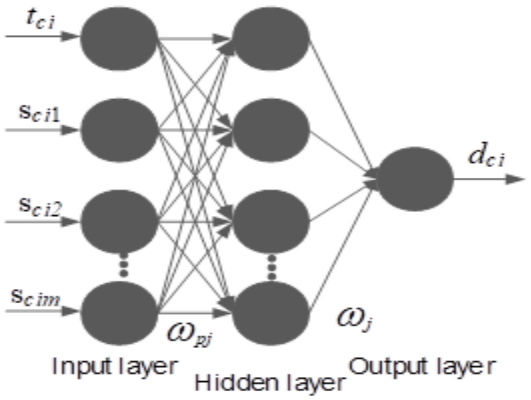

Figure 2 Basic structure of BP neural network

We construct the BP neural network as shown in Figure 2, where s_{ci}'s form a set of accelerated stresses and time-varying environmental data, t_{ci} is observation time, d_{ci} is degradation data, m is the total number of stresses, and ω is weight coefficient. Then the output of hidden layer neurons can be obtained as

$$H_j = f\left(\omega_{pj} t_{ci} + \sum_{p=1}^{m} \omega_{pj}\, s_{ci} - b_j\right),\ i = 1,\ldots,n; j = 1,\ldots,q; p = 0,1,\ldots,m. \quad (17)$$

where H_j is the output of the hidden layer, $f(.)$ is the driving function, ω_{pj} is the weight coefficient between input layer and hidden layer, b_j is the threshold of hidden layer, q is the number of hidden layers, and n is the data size.

With the hidden layer output H, the threshold value k, and the weight coefficient ω_j between hidden layer and output layer, the prediction result P of the output layer can be given as

$$P = f\left(\sum_{i=1}^{q} H_j \cdot \omega_j - k\right). \quad (18)$$

The prediction error ε is calculated as

$$\varepsilon = d_{c_i} - P \quad (19)$$

The feed backward modeling process is conducted along the neural network based on the prediction error ε. This error can be minimized with the updating of weight coefficients and thresholds by the following updating schemes:

$$\omega_{pj} \leftarrow \begin{cases} \omega_{pj} + \xi \cdot t_i \cdot H_j\left(1 - H_j\right) \cdot \omega_j \cdot \varepsilon, & p = 0 \\ \omega_{pj} + \xi \cdot s_{ci} \cdot H_j\left(1 - H_j\right) \cdot \omega_j \cdot \varepsilon, & p = 1,2,\ldots,m \end{cases}$$

$$b_j \leftarrow b_j + \xi \cdot t_i \cdot H_j\left(1 - H_j\right) \cdot \omega_j \cdot \varepsilon, \quad p = 0 \quad (20)$$

$$\mathrm{k} \leftarrow \mathrm{k} + \varepsilon$$

where ξ is the learning rate.

2.2.3 Ensemble Predictive Model Construction and Lifetime Prediction

Of all sub-models, the stochastic process model is an analytical model, while SVM and BP network are the black-box type of predictive models; but, once established, all of them can provide prediction results for any new observations. Therefore, as discussed in section 2.1, the ensemble strategy is to combine these sub-models by averaging their prediction outputs.

Again, the time-varying environmental or working stress data are denoted as $\{s_i\}$, i=1, 2, … n, where n is the data size and s_i=[s_{i1}, s_{i2}, … s_{im}] is the vector of measurements of m stress variables. These time-varying environmental data, for example, can be obtained from historical weather/climate records. Then these data, along with time point t_i, are set to be the input to the three sub-models. The prediction results for a product under its operating condition from these sub-models are denoted as $f_{Bayesian}(s_i, t_i)$, $f_{SVM}(s_i, t_i)$ and $f_{BP}(s_i, t_i)$.

Finally, the prediction result of ensemble model, denoted as $f_{Ensemble}$, is calculated by taking the weighted average of outputs from these sub-models,

$$f_{Ensemble}(s_i, t_i) = w_1 f_{Bayesian}(s_i, t_i) + w_2 f_{SVM}(s_i, t_i) + w_3 f_{BP}(s_i, t_i) \quad (21)$$

where w_1, w_2 and w_3 are the weight value for each method with $w_1 + w_2 + w_3 = 1$ and $f_{Ensemble}(s_i, t_i)$ is the proposed ensemble predictive model.

Let k be the performance threshold and a product fails when $f_{Ensemble}(s_i, t_i) - k < 0$. However, the data may fluctuate within a certain range; in order to eliminate the influence of outliers, we do not use the first passage time as the lifetime prediction. Instead, if there are n consecutive times that the failure criterion, $f_{Ensemble}(s_i, t_i) - k < 0$, is satisfied, the nth passage time is used as the product's predicted failure time under its operating condition.

2.3 Model Performance Estimation Criteria

To evaluate the performance of a predictive model, a metric function (i.e., the function for assessing prediction error) needs to be carefully chosen. In this paper, two metrics -- mean absolute error (MAE) and root mean square error (RMSE) – are used to evaluate the performance of a candidate model. MAE is the average of absolute errors and RMSE is the sample standard deviation of the difference between prediction and observed value. These metrics are widely used for quantifying the quality of a predictive model.

Using the test dataset we obtain the prediction value, $Y_{predicted}(i)$, from $f_{Ensemble}$ and compare it with the observed value, $Y_{real}(i)$. MAE and RMSE are calculated by

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_{real}(i) - y_{predict}(i)| \quad (22)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(y_{real}(i) - y_{predict}(i)\right)^2} \quad (23)$$

3. Simulation Study

In this section, a simulation example is presented to verify the effectiveness of the proposed ensemble predictive model. Moreover, since the proposed model is designed to predict a lifetime accounting for the time-varying operational condition, degradation datasets with multiple random stress variables are generated for this study. The study design is presented in section 3.1, then in section 3.2, we present the construction process of proposed predictive method, and in section 3.3, we compare the model performance with those of benchmark models.

3.1 Data description

Suppose a product is subjected to three different types of stresses that possess stochastic volatility within certain ranges. The variation ranges of these stresses are shown in Table 1. We assume that the product's performance degradation can be described by a linear function and it follows a Wiener process. Then, we choose the general linear model as an acceleration model to denote the relationship between stress and degradation, i.e.,

$$d(s_i) = \exp(\beta_0 + \beta_1\varphi(s_{1i}) + \beta_2\varphi(s_{2i}) + \beta_3\varphi(s_{3i})), i = 1,2,\dots,n \quad (24)$$

where $\varphi(s_{ji})$ (j=1,2,3) is a known function of original stress variable. Moreover, to verify the effect of different variance value on prediction results and to test the stability of the proposed model, we choose three variance values. The initial values of these parameters are listed in Table 2.

Table 1 Ranges of Stress Variable

Stress number	Stress 1	Stress 2	Stress 3
Stress range	20-50	15-70	30-50

Table 2 Initial value of model parameter

Parameter	β_0	β_1	β_2	β_3	σ
Dataset 1	5	-240	-170	-30	0.054
Dataset 2	5	-240	-170	-30	0.254
Dataset 3	5	-240	-170	-30	0.354

Here, we let the time interval Δt to be 1, the number of data points be 1,000 and the initial performance value be 100. A series of accelerated degradation increments can be generated by Monte Carlo using Eq. (24). The accelerated degradation data for different variances can be obtained by subtracting the increment from the initial value. Moreover, to validate the simulated data, we fit the increment data with a normal distribution and compare

the fitted variances with true values. The three simulated datasets used in this study are shown in Figure 3.

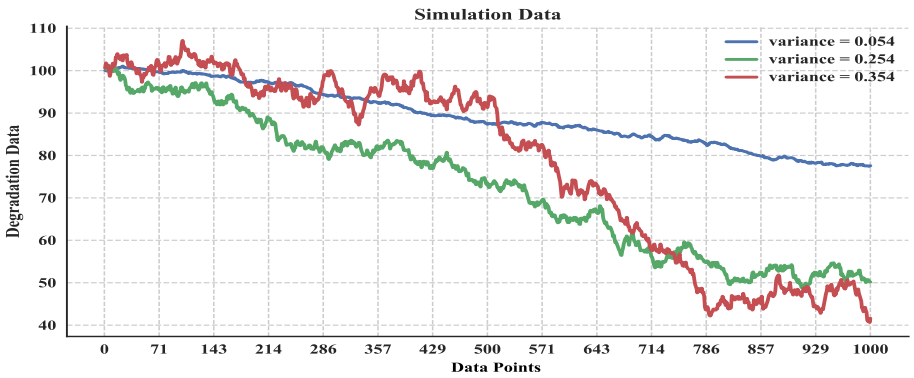

Figure 3 Final simulated accelerated degradation data

3.2 Predictive Model Construction

We now provide the details of how to construct the proposed predictive model for the simulated data. We will evaluate the performance of our proposed ensemble model and compare it with some benchmark models. The model construction process consists of the following steps.

Step 1: Data Partitioning and Resampling. In our study, to test the prediction accuracy of a model, a later part of degradation measurements is removed from the original dataset and preserved for model evaluation. We call this dataset the prediction dataset. We divide the simulated accelerated degradation data into a training and testing dataset and a prediction dataset with the proportion of 75/25. That is, the data with times ranging from 751 to 1,000 are reserved for evaluating the performance of the proposed model. Next, we resample the training and testing dataset and create three subsets by bootstrapping, as explained in Section 2.2.1. These subsets are further divided into training sets and testing sets, as described before.

Step 2: The Bayesian-based Predictive Sub-model Construction. To draw inferences on the unknown parameters with MCMC, the initial values of these parameters must be obtained. Parameters β_0, β_1, β_2, and β_3 can be calculated by Eqs. (10), (11) and (13) with polynomial regression method fitting. As for σ^2, it is known from the properties of Wiener process that the conjugate prior distribution of $1/\sigma^2 \cdot \Delta t$ is a gamma distribution, and so are the parameters a and b. Here, the prior distribution of these parameters are listed in Table 3. Then the parameters for the Bayesian-based predictive sub-model are inferred by MCMC. The results of different variances are shown in Table 4.

Table 3 Prior Distribution of Parameters

Parameter	β_0	β_1	β_2	β_3	σ
Prior Distribution	Normal Distribution	Normal Distribution	Normal Distribution	Normal Distribution	Gamma Distribution
Initial Value	N(4.5, 10^3)	N(-250, 10^3)	N(-180, 10^3)	N(-30, 10^3)	a~G(1,1); b~G(1,0.3)

Table 4 Parameter Inference Results with Different Variance Values

Parameter	β_0	β_1	β_2	β_3	σ
σ=0.054	150.3458	-233.5981	-168.2673	-30.8805	0.0499
σ=0.254	150.5371	-227.9355	-168.4073	-26.9785	0.2292
σ=0.354	149.607	-261.8479	-171.3926	-31.7399	0.3243

Step 3: The SVM-Based Predictive Sub-model Construction. The modeling process described in Section 2.2.2 is used to construct an SVM-based predictive sub-model. The SVM parameters listed in Table 5. Here, we choose the RBF kernel function as the kernel function of SVM and the particle swarm optimization algorithm for parameter optimization. The SVM performance, visualized by the actual and predicted degradation values, are shown in Figure 4.

Table 5 SVM Parameters

Parameter	*Initial Value*
c	[0,100]
g	[0,100]
Evolution generations	100
Population number	20
Cross-validation times	5

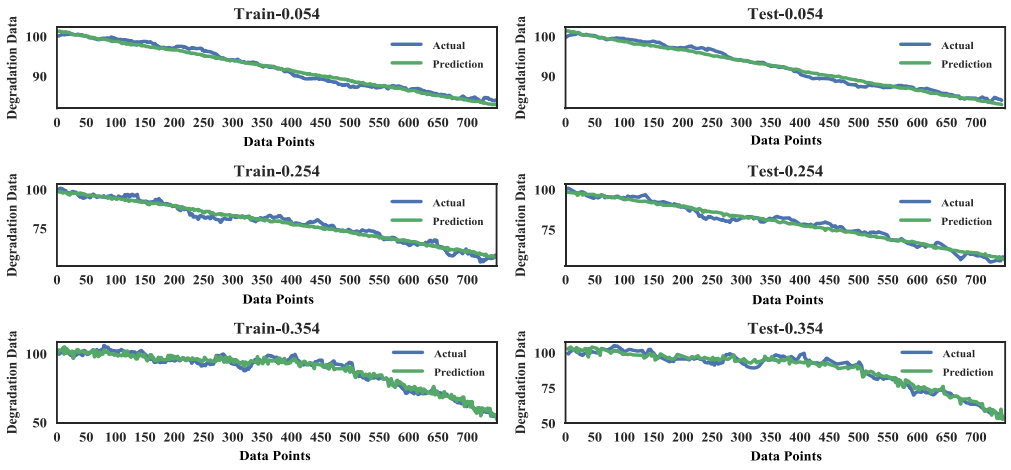

Figure 4 Training and testing process of SVM on three datasets

Step 4: The BP Neural Network-based Predictive Sub-model Construction. Based on the parameters listed in Table 6 and the process described in Section 2.2.2, a neural network with a structure of 4-15-1 (4 input layers, 15 hidden layers, and 1 output layer) is used to construct a BP neural network-based predictive sub-model. The actual and prediction values of three different datasets are shown in Figure 5.

Table 6 BP Neural Network Parameters

Parameter	*Initial Value*	*Parameter*	*Initial Value*
c	[0,100]	***Cross validation times***	5
g	[0,100]	***Maximum iterative number***	1000
Evolution generations	100	***Learning rate***	0.05
Population number	20	***Learning goal***	0.00001

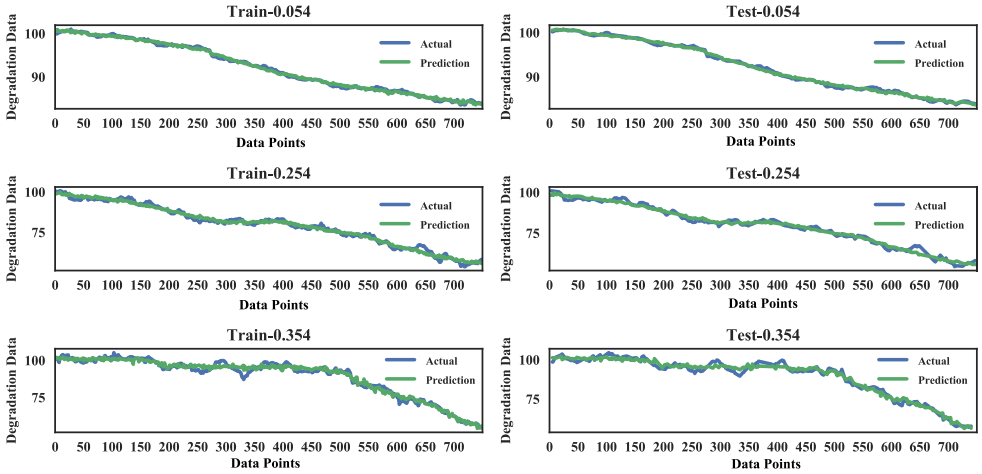

Figure 5 Training and testing process of BP Neural Network on three datasets

Step 5: The Ensemble Predictive Model Construction. After the prediction set is fed into three sub-models, the ensemble predictive model, as described by Eq. (21), with equal weight values $w_1=w_2=w_3=1/3$ is employed. The ensemble predictions on future degradation measurements are visualized in Figure 6.

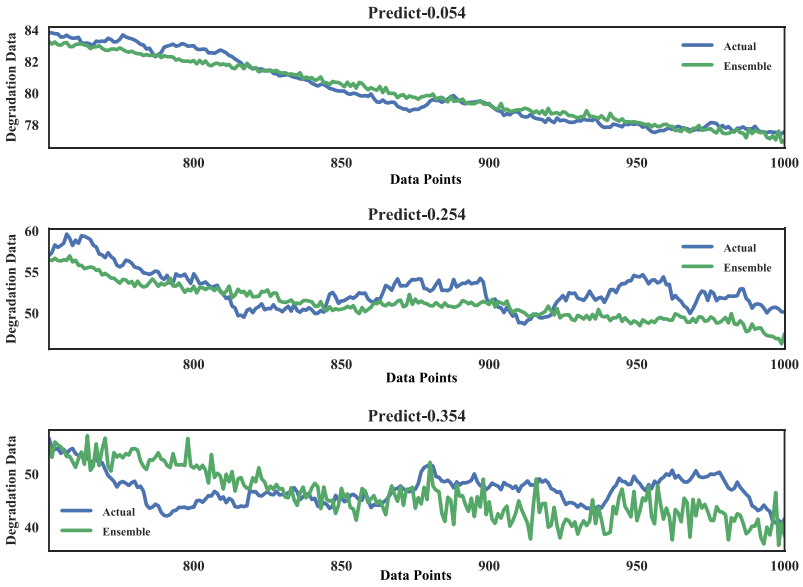

Figure 6 ensemble predictive model performance curves of three datasets

3.3 Comparison and Analysis of Results

The three basic models (SVM, BP neural network) and the stochastic process with Bayesian inference are now presented as benchmark models trained with the original dataset. They all use unresampled data to build lifetime predictive models, and then after the model is well trained, the same prediction dataset as used by the ensemble model will be used to test these benchmark models. The performances of these models are evaluated by RMSE and MAE. Considering the randomness of simulation example, we make the prediction 20 times with the prediction data. The average results of these 20 RMSEs and MAEs are compared with the proposed model's performance to demonstrate the implementation of the proposed method. To further examine the effect of timeframe on a predictive model's performance, five data sets are selected and each of them contains the first 50, 100, 150, 200, and 250 data points of the prediction dataset, respectively. The comparison results are shown in Figure 7, y-axis value represent the error rate of the model.

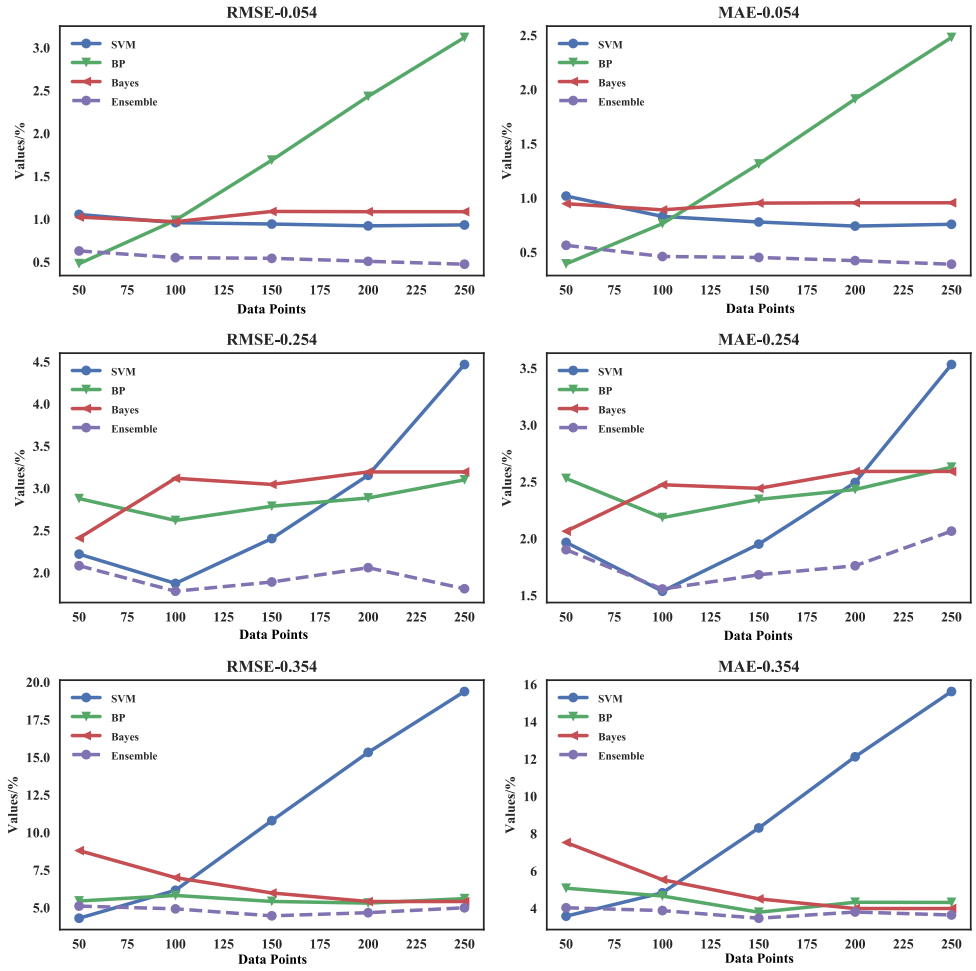

Figure 7 Performance Comparison with Benchmark models

It should be noted that in Figure 7, the ensemble predictive model (purple dotted line) outperforms any individual member algorithm at most data points, as expected. Also, the performance of the proposed model is quite stable and varies little with increased variance or increased prediction timeframe. Thus, we conclude that the proposed ensemble model is effective.

4. Case Study

In this section, we apply the proposed model on an accelerated degradation dataset obtained from a lithium-ion (Li-ion) battery testing program and intend to provide a more effective and stable lifetime prediction.

Due to its high density of specific energy and long lifetime cycle, Li-ion battery is the most popular type of rechargeable battery used in mobile devices, electric vehicles, space systems, and unmanned aerial systems, etc.[45][46] In practice, many Li-ion batteries are operated in severe environments such as extremely high or low temperature, and high

humidity. These factors may affect the energy, power capability, and durability of Li-ion batteries and put applications at risk. Thus, it is meaningful to gain a deeper understanding of the relationship between the battery's degradation process and the foregoing factors to determine an effective energy management strategy.

The test units in this experiment are commercial Samsung 18650 Li-ion batteries. We conducted an ADT program on these batteries. Two electric currents – charge current and discharge current – were selected as the accelerating stresses. Three stress levels were set with a constant battery voltage of 4.2 V. The combination of stresses can be found in Table 7.

Three Li-ion batteries were tested at each level, and the degradation data were recorded. Since the lifetime of Li-ion batteries is affected by temperature and humidity, the environmental data were also recorded. The obtained degradation data and environmental data are shown in Figure 8.

Table 7 Combination of Stress Levels

Stress Level	Voltage	Charge Current	Discharge Current
S1	4.2V	0.8C (1720mA)	1.5C (3225mA)
S2	4.2V	1C (2105mA)	2C (4300mA)
S3	4.2V	1C (2150mA)	2.5C (5375mA)

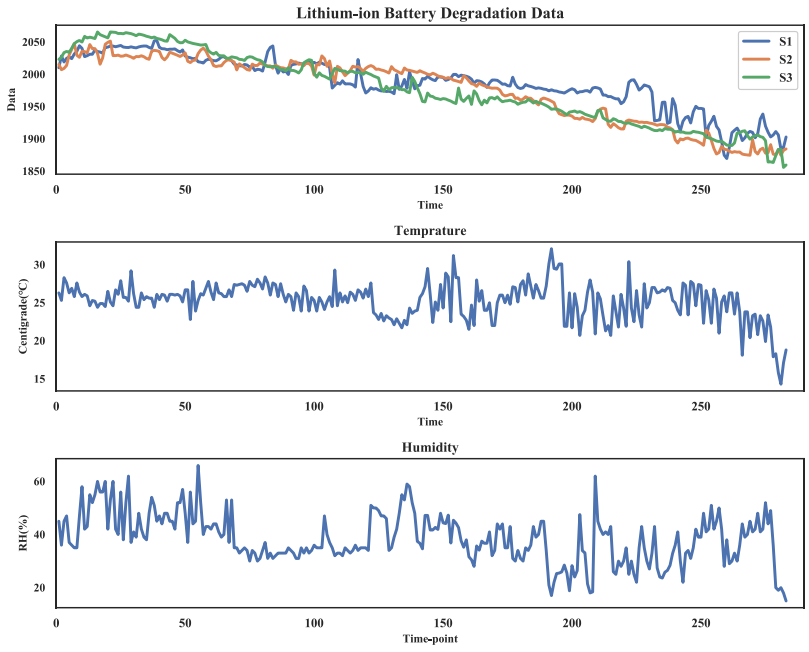

Figure 8 Degradation Data and Environmental Data

To test the proposed model's ability to predict the lifetime under normal operating condition, degradation data from the normal stress level with a voltage of 4.2V, a charge current of 0.5C and a discharge current of 1C were recorded. There were 698 data points at

this stress level. The environmental data (temperature and humidity) were also obtained and shown in Figure 9.

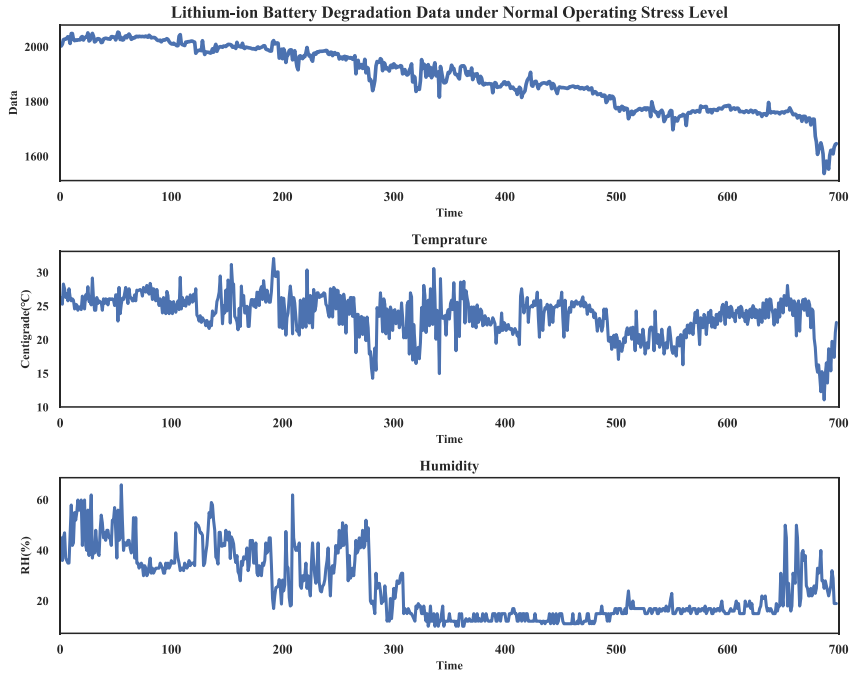

Figure 9 Degradation Data and Environmental Data under Normal Operating Condition

Based on the description in section 2.2, a Bayesian-based stochastic process sub-model was built with a Wiener process and a general linear model. Table 8 lists the inferred parameters of the Bayesian-based predictive sub-model by MCMC.

Table 8 Parameter Inference Results

Parameters	$\widehat{\beta_0}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\hat{\sigma}$
Values	49.9342	295.1899	-193.3770	4.9273	-14.9024

Let s_0 represent the stress in a normal environment; i.e., s_0 represents a voltage of 4.2V, a charge current of 0.5C and a discharge current of 1C.

Then, based on Eq. (9), the predicted degradation rate becomes

$$\mathrm{d}(s_0) = \exp[\widehat{\beta_0} + \sum_{j=1}^{3} \widehat{\beta_j}\, \varphi(s_0)] \tag{25}$$

Eq. (8) becomes

$$\Delta y \sim \mathrm{N}(\mathrm{d}(s_0)\Delta t, \hat{\sigma}^2 \Delta t) \tag{26}$$

The amount of degradation in the normal environment can be obtained by accumulating Δy. Therefore, we can derive degradation prediction at normal environment based on the model built by the collected degradation data at accelerated environment.

The SVM and BP neural network-based sub-models were constructed with the inputs of charging current, discharging current, temperature, and humidity, charge/discharge cycle and the output of capacity-degradation data. Then, we utilized the normal operating stress variables (the voltage, the charge current and the discharge current) and the environmental data obtained from the normal operating condition and normal charge/discharge cycle to establish an ensemble predictive model. The prediction results of this model are shown in Figure 10.

The performance of the proposed model was further estimated by RMSE and MAE, and compared with the three benchmark models at the first 100, 200, 300, 400, 500, 600 and 698 data points. The comparison results are shown in Figure 11 and Figure 12.

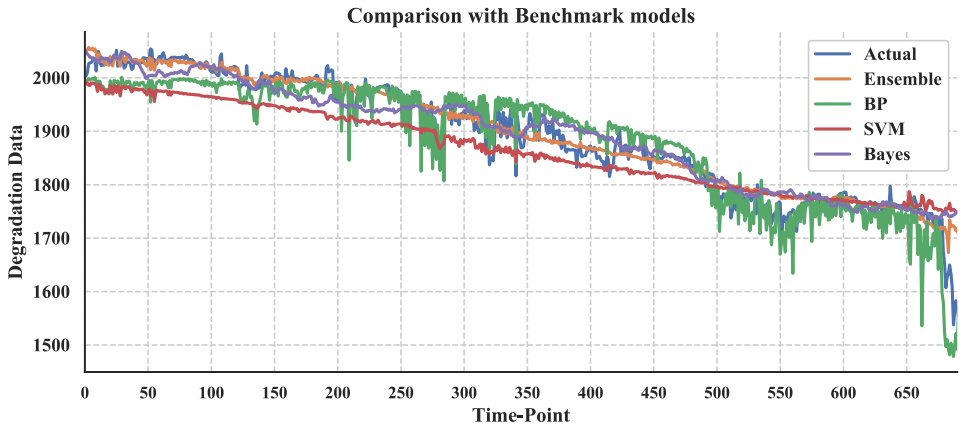

Figure 10 Comparison with Benchmark models

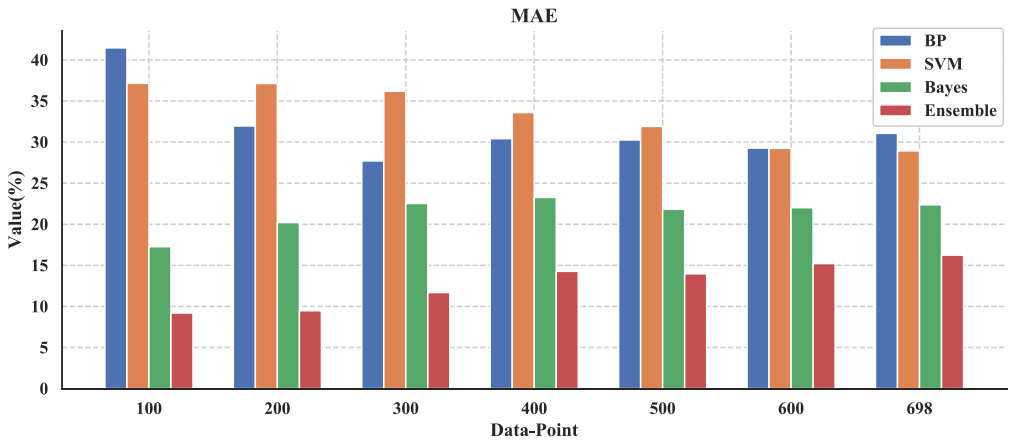

Figure 11 Comparison results of MAE with Benchmark models

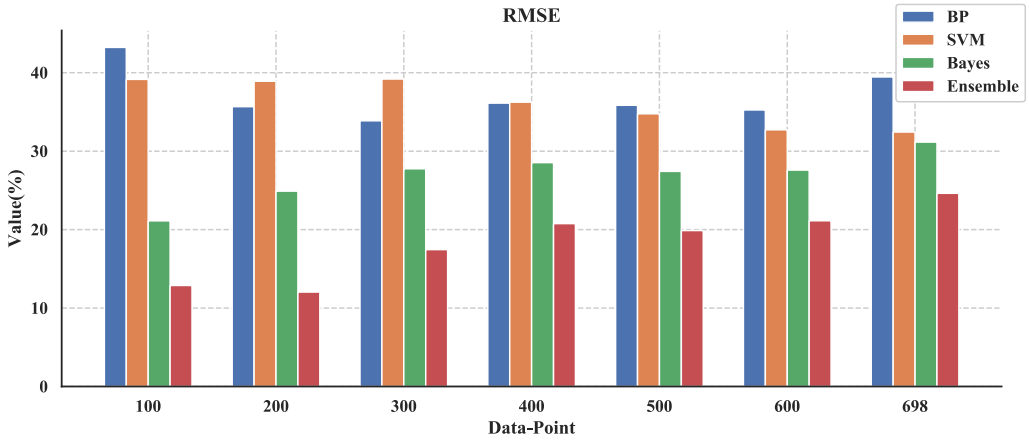

Figure 12 Comparison results of RMSE with Benchmark models

As can be seen in Figure 10, predictions of the proposed ensemble model are consistent with actual data. Also from the results shown in Figure 11 and Figure 12, we can conclude that the ensemble method outperforms the benchmark models at all tested timeframes and the variance of errors remains stable.

To test the performance of the proposed model in lifetime prediction, we let the failure threshold of a test item to be 80% of its capacity, which is 1720 mAH. The battery lifetime is defined as the number of cycles that the battery can be charged and discharged before its capacity reaches the failure threshold. The prediction results and accuracies calculated from the ensemble model, as well as from the benchmark models, are listed in Table 9.

As can be seen in Table 9, the prediction accuracy of the proposed model is the highest compared with others. Therefore, it concludes the effectiveness of the ensemble predictive model for making a lifetime prediction.

Table 9 Comparison of Prediction Result and Accuracy with Benchmark models

Models	SVM	BP neural network	Bayesian model	Ensemble predictive model	Actual Lifetime
Result	670	550	697	686	683
Error Rate	1.9%	19.5%	2.0%	0.4%	/

5. Conclusion

In this paper we tackle the problem of product lifetime prediction in a complicated time-varying operational environment and in order to improve prediction accuracy, we propose an ensemble learning-based degradation predictive model. The Bayesian inference method for building stochastic process models and two other popular machine learning methods – SVM and BP neural network – are chosen to construct predictive sub-models. The former method is

chosen for its merits of uncertainty reduction and being able to use prior information accumulated from field observations, while the latter ones are selected for their strong data adaptability. The simulation example demonstrates that the proposed ensemble method can reduce the prediction error caused by time-varying environmental stresses, as it balances out the shortfalls of individual models and gains a more accurate and stable prediction result. A real case study of the Li-ion battery degradation further verifies its superiority and usefulness in engineering practice. It indicates that the proposed method can further help with effectively managing energy storage and energy distribution of battery pack.

This research aims to understand how to obtain a more accurate prediction result when considering the time-varying environmental data along with the accelerated degradation process. The proposed ensemble learning-based predictive model is built upon the Bayesian inference method and two popular machine learning algorithms. The computation of the proposed method is expensive, which could be alleviated by some optimization methods. Moreover, different machine learning algorithms may have different effects on the accuracy of prediction result and increasing prediction horizon will affect the effectiveness and accuracy of the proposed method too. These aspects would be discussed in our future study.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 51805016) and the Fundamental Research Funds for the Central Universities (Grant No. YWF-19-BJ-J-340). The work of forth author is partially supported by NSF Grant 1726445.

Reference

[1] Zhao Z, Liang B, Wang X, et al. Remaining Useful Life Prediction of Aircraft Engine based on Degradation Pattern Learning. Reliability Engineering & System Safety, 2017, 164.

[2] Liao H, Elsayed E A. Reliability inference for field conditions from accelerated degradation testing. Naval Research Logistics, 2006, 53(6):576–587.

[3] Jiang M, Chen W. Integrated Approach for Field Reliability Prediction Based on Accelerated Life Testing. Quality Engineering, 2015, 27(3):317-328.

[4] Wan F, Tan Y, Jiang Z, et al. Lifetime prediction and reliability estimation methodology for Stirling-type pulse tube refrigerators by gaseous contamination accelerated degradation testing. Cryogenics, 2017, 88: 116-124.

[5] Mehr M Y, Driel W D V, Zhang G Q. Reliability and Lifetime Prediction of Remote Phosphor Plates in Solid-State Lighting Applications Using Accelerated Degradation Testing. Journal of Electronic Materials, 2015, 45(1):1-9.

[6] Liu L, Li X Y, Jiang T M. Integration method for reliability assessment with multi-source incomplete accelerated degradation testing data. Quality Engineering, 2017, 29(3):366-376.

[7] Sun F, Liu J, Li X, et al. Reliability Analysis with Multiple Dependent Features from a Vibration-Based Accelerated Degradation Test. Shock and Vibration, 2016, (2016-12-7), 2016, 2016(5):Article ID 2315916.

[8] You M Y, Meng G. Approaches for component degradation modelling in time-varying environments with application to residual life prediction. Proceedings of the Institution of Mechanical Engineers Part E Journal of Process Mechanical Engineering 1989-1996 (vols, 2012, 226(2):117-131.
[9] Jeffrey P. Kharoufeh, Christopher J. Solo, M. Yasin Ulukus. Semi-Markov models for degradation-based reliability. IISE Transactions, 2010, 42(8):599-612.
[10] Gebraeel N, Pan J. Prognostic Degradation Models for Computing and Updating Residual Life Distributions in a Time-Varying Environment. IEEE Transactions on Reliability, 2008, 57(4):539-550.
[11] Pan R. A Bayes approach to reliability prediction utilizing data from accelerated life tests and field failure observations. Quality and Reliability Engineering International, 2009, 25(2): 229-240.
[12] Wang L, Pan R, Li X, et al. A Bayesian reliability evaluation method with integrated accelerated degradation testing and field information. Reliability Engineering & System Safety, 2013, 112: 38-47.
[13] Qu X, Wang H, Zhan X, et al. A lifetime prediction method for LEDs considering real mission profiles. IEEE Transactions on Power Electronics, 2017, 32(11): 8718-8727.
[14] Thomas E V, Bloom I, Christophersen J P, et al. Statistical methodology for predicting the life of lithium-ion cells via accelerated degradation testing. Journal of Power Sources, 2008, 184(1): 312-317.
[15] Zhou R, Serban N, Gebraeel N. Degradation-based residual life prediction under different environments. Annals of Applied Statistics, 2014, 8(3).
[16] Ye Z and Xie M. Stochastic modelling and analysis of degradation for highly reliable products. Applied Stochastic Models in Business & Industry, 2015, 31(1):16-32.
[17] Liu L, Li X, Sun F, et al. A general accelerated degradation model based on the Wiener process. Materials, 2016, 9(12): 981.
[18] Tang S, Yu C, Wang X, et al. Remaining useful life prediction of lithium-ion batteries based on the Wiener process with measurement error. Energies, 2014, 7(2): 520-547.
[19] Limon S, Yadav O P, Nepal B. Estimation of Product Lifetime Considering Gamma Degradation Process with Multi-Stress Accelerated Test Data. Industrial and Systems Engineering Conference. 2017.
[20] Santini T, Morand S, Fouladirad M, et al. Accelerated degradation data of SiC MOSFETs for lifetime and Remaining Useful Life assessment. Microelectronics Reliability, 2014, 54(9-10): 1718-1723.
[21] Fang G, Rigdon S E, Pan R. Predicting lifetime by degradation tests: A case study of ISO 10995. Quality and Reliability Engineering International, 2018, 34(6): 1228-1237.
[22] Pan R, Crispin T. A hierarchical modeling approach to accelerated degradation testing data analysis: A case study. Quality and Reliability Engineering International, 2011, 27(2): 229-237.
[23] Mukhopadhyay C, Roy S. Bayesian accelerated life testing under competing log-location-scale family of causes of failure. Kluwer Academic Publishers, 2016.
[24] Wang H, Xu T and Mi Q. Lifetime prediction based on Gamma processes from accelerated degradation data. Chinese Journal of Aeronautics, 2015, 28(1):172-9.
[25] Mishra M, Martinsson J, Rantatalo M, et al. Bayesian hierarchical model-based

prognostics for lithium-ion batteries. Reliability Engineering & System Safety, 2018, 172:25-35.
[26] Xu A, Basu S and Tang Y. A Full Bayesian Approach for Masked Data in Step-Stress Accelerated Life Testing. IEEE Transactions on Reliability, 2014, 63(3):798-806.
[27] Mukhopadhyay C, Roy S. Bayesian accelerated life testing under competing log-location-scale family of causes of failure. Computational Statistics, 2016, 31(1):89-119.
[28] Ali S H, Heydarzadeh M, Dusmez S, et al. Lifetime Estimation of Discrete IGBT Devices Based on Gaussian Process. IEEE Transactions on Industry Applications, 2017, PP(99):1-1.
[29] Cheng S, Li B, Yuan Z, et al. Development of a lifetime prediction model for lithium thionyl chloride batteries based on an accelerated degradation test. Microelectronics Reliability, 2016, 65:274-279.
[30] Dong H, Jin X, Lou Y, et al. Lithium-ion battery state of health monitoring and remaining useful life prediction based on support vector regression-particle filter. Journal of power sources, 2014, 271: 114-123.
[31] Freitag S, Beer M, Graf W, et al. Lifetime prediction using accelerated test data and neural networks. Computers & Structures, 2009, 87(19-20): 1187-1194.
[32] Eddahech A, Briat O, Bertrand N, et al. Behavior and state-of-health monitoring of Li-ion batteries using impedance spectroscopy and recurrent neural networks. International Journal of Electrical Power & Energy Systems, 2012, 42(1): 487-494.
[33] Lu Z, Geng X, Chen G. A Bayesian assumption based forecasting probability distribution model for small samples. Computers & Electrical Engineering, 2017.
[34] Zhang W, Li X, Jiang T, et al. Life-prediction of multi-stress accelerated life testing based on BP algorithm of artificial neural network. Acta Aeronautica Et Astronautica Sinica, 2009, 30(9):1691-1696.
[35] Jóhannesson G, Austri R R D, Vincent A C, et al. Bayesian analysis of cosmic-ray propagation: evidence against homogeneous diffusion. Astrophysical Journal, 2016, 824(1):16.
[36] Zhang C, Lim P, Qin A K, et al. Multiobjective Deep Belief Networks Ensemble for Remaining Useful Life Estimation in Prognostics. IEEE Trans Neural Netw Learn Syst, 2016, PP(99):1-13.
[37] Wang Z, Wang Y, Srinivasan R S. A novel ensemble learning approach to support building energy use prediction. Energy & Buildings, 2018, 159:109-122.
[38] Fan C, Xiao F, Wang S. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. Applied Energy, 2014, 127(6):1-10.
[39] Sun J, Zuo H, Yang H, et al. Study of ensemble learning-based fusion prognostics. Prognostics and Health Management Conference. IEEE, 2010:1-7.
[40] Xing Y, Ma E W M, Tsui K L, et al. An ensemble model for predicting the remaining useful performance of lithium-ion batteries[J]. Microelectronics Reliability, 2013, 53(6):811-820.
[41] Breiman, L. (1996). Bagging predictors. Machine learning, 24(2), 123-140.
[42] Bradley P. Carlin, Thomas A. Louis. Bayesian Methods for Data Analysis. Journal of the Royal Statistical Society, 2008, 172(4):935–936.

[43] Tran V T, Hong T P, Yang B S, et al. Machine performance degradation assessment and remaining useful life prediction using proportional hazard model and support vector machine. Mechanical Systems & Signal Processing, 2012, 32(4):320-330.
[44] Patil M A, Tagade P, Hariharan K S, et al. A novel multistage Support Vector Machine based approach for Li ion battery remaining useful life estimation. Applied Energy, 2015, 159:285-297.
[45] Xu X, Chen N. A state-space-based prognostics model for lithium-ion battery degradation. Reliability Engineering & System Safety, 2017, 159:47-57.
[46] Baghdadi I, Briat O, Delétage J Y, et al. Lithium battery aging model based on Dakin's degradation approach. Journal of Power Sources, 2016, 325:273-285.

Bios:

Lizhi Wang received the Ph.D. degree at Beihang University, Beijing, China. He is a lecturer with the Beihang University, Beijing, China. His main research interests include system engineering, ADT, optimal decision, evaluation method, data fusion and lifetime prediction.

Dawei Lu received the master degree at School of Reliability and Systems Engineering, Beihang University, Beijing, China, work at the 52nd Research Institute, CETHIK Group Co.Ltd., Hangzhou, China. His main research interests include UAV swarms, system engineering, system modeling, complex network , lifetime prediction and Electronics reliability.

Xiaohong Wang received the Ph.D. degree at Beihang University, Beijing, China. She is an associate professor with the School of Reliability and Systems Engineering, Beihang University, Beijing, China. Her main research interests include reliability and environment testing, accelerated testing and life prediction.

Rong Pan received his Ph.D. degree Industrial Engineering from the Pennsylvania State University in 2002.He is currently an Associate Professor in the School of Computing, Informatics, and Decision Systems Engineering at Arizona State University. His research interests include quality and reliability engineering, design of experiments, time series analysis, and statistical learning theory. He is a senior member of ASQ, IEEE and IISE, and a lifetime member of SRE.

Zhuo Wang received the master degree at School of Reliability and Systems Engineering, Beihang University, Beijing, China. Her main research interests include batteries dependency analysis, accelerated testing and lifetime prediction.