A Remaining Useful Life Prediction Method in the Early Stage of Stochastic Degradation Process

Yuhan Zhang, Ying Yang, *Senior Member, IEEE*, Xianchao Xiu, He Li, and Ruijie Liu

***Abstract*—In order to achieve more accurate predicted RUL in the early stage of degradation, a novel remaining useful life (RUL) prediction method for the stochastic degradation process is proposed. Technically, modeling the degradation process as a Wiener process (WP) whose drift increment is a weighted sum of kernel functions can flexibly depict the nonlinear degradation trend. Introducing a long short term memory (LSTM) network can capture the long-term dependencies of the offline experimental and online observed degradation data to forecast the future degradation increment. Then, based on the degradation model, a numerical approximate distribution of the RUL is derived to quantify the uncertainty of the predicted RUL. Finally, a practical case study of lithium-ion batteries is provided to demonstrate the high accuracy of the proposed methods for RUL prediction especially in the early stage of degradation.**

***Index Terms*—Remaining useful life (RUL), Wiener process (WP), long short term memory (LSTM) network.**

I. Introduction

AS A KEY technique to evaluate the health status of the equipment in the field of prognostics and health management (PHM), remaining useful life (RUL) prediction can provide effective information for developing maintenance strategies, so as to reduce the economic losses caused by failures or anomalies [1]. Due to the stochastic dynamics, one of the most common features in the actual degradation processes, statistical-data-driven methods [2] have been studied intensively in recent years. As a popular model among the statistical-data-driven methods, Wiener process (WP) is used to model the degradation process to obtain the distribution of the predicted RUL and quantify the uncertainty of the predicted RUL, which is crucial for the analysis of RUL prediction results [3]. WP-based RUL prediction methods have been widely applied to various systems and devices including gyroscopes [4], cutting tools [5], lithium-ion batteries [6], rotating bearings [7], etc. Most of them use specific nonlinear (adaptive [6], [7] or not adaptive [4], [5]) models to describe the degradation trend based on the system's history of the observed data to date and obtain RUL distribution. However, in the early stage of degradation, the historical data are not accumulated sufficiently and the accuracy of RUL prediction is not very satisfactory (see [6, Tab. 1]). Therefore, it motivates this brief to utilize an advanced time series forecasting method to augment the global degradation data based on historical observation data and establish an augmented WP-based model to further improve the RUL prediction accuracy in the early stage of degradation.

The long short term memory (LSTM) network, a deep learning neural network, can memorize information in its internal cell for long periods of time via the introduced gates. It is able to learn and capture the long-term dependencies of the time series data [8], [9]. The application to lithium-ion batteries' [10] and turbofan engine's [11] RUL prediction validates the LSTM network's potential for time series forecasting. Therefore, this brief utilizes LSTM network technique to capture the long-term dependencies of degradation data and forecast the future degradation increment based on the offline experimental and online degradation increment when the data are not sufficient in the early stage of degradation. And a WP-based degradation model is established whose drift increment is a combination of weighted kernel functions. The parameters in the model are calculated based on the online observed increment and the forecasted future degradation increment using a sparse Bayesian algorithm. Then, the global augmented WP model is obtained in the early stage of degradation, and the numerical result of RUL distribution can be derived. Finally, through the case study of lithium-ion batteries, it is proved that in the early stage of stochastic degradation process, compared with the existing two state-of-art WP-based methods, the accuracy of RUL prediction results can be enhanced. And the combination of WP and LSTM network is helpful to improve the accuracy of RUL prediction results.

The main contribution of the proposed method can be summarized as follows:

- The method provides a general and flexible representation of the nonlinear behavior of degradation.
- The offline experimental and online degradation data can be fused to establish a global augmented degradation model by LSTM network.
- A numerical calculation formula of the RUL distribution can be obtained based on the degradation model.
- The case study of lithium-ion batteries demonstrates the practical importance.

Manuscript received October 14, 2020; accepted October 26, 2020. Date of publication October 28, 2020; date of current version May 27, 2021. This work was supported by National Natural Science Foundation (NNSF) of China under Grant 61633001, Grant 12001019, and Grant U1713223. This brief was recommended by Associate Editor Y. Xia. *(Corresponding author: Ying Yang.)*

The authors are with the State Key Laboratory for Turbulence and Complex Systems, Department of Mechanics and Engineering Science, College of Engineering, Peking University, Beijing 100871, China (e-mail: yuhanzhang@pku.edu.cn; yy@pku.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCSII.2020.3034393.

Digital Object Identifier 10.1109/TCSII.2020.3034393

II. Nonlinear Degradation Process Model and RUL Estimation

Consider the general nonlinear degradation model based on the following diffusion process reviewed in [3],

$$dX(t) = \mu(t;\boldsymbol{\theta})dt + \sigma(t;\boldsymbol{\theta})dB(t), \tag{1}$$

where $X(t)$ represents the degradation state at time t, $\boldsymbol{\theta}$ is the parameter vector. $B(t)$ denotes a standard Brownian motion (BM) and $B(t) \sim N(0,t)$. $\sigma(t;\boldsymbol{\theta})$ is the diffusion coefficient function which is mostly fixed as a constant σ to partially depict the dynamics of the degradation process, while $\mu(t;\boldsymbol{\theta})$ is the drift coefficient function.

The degradation increment of (1) with the diffusion coefficient $\sigma(t;\boldsymbol{\theta}) = \sigma$ is given by

$$I_X(t;\tau) = I_{Dr}(t;\tau) + \sigma B(t) - \sigma B(t-\tau), \tag{2}$$

where the diffusion increment $\sigma B(t) - \sigma B(t-\tau) \sim N(0,\sigma^2\tau)$ obeys a normal distribution. τ denotes the increment interval and is set as the integer multiples of the sampling interval of the degradation process in this brief. And it is one of the hyperparameters in the model and should be selected by cross validation. In order to describe the nonlinear trend flexibly, the drift increment function $I_{Dr}(t;\tau) = \int_{t-\tau}^{t} \mu(t;\boldsymbol{\theta})dt$ is formulized as follows when τ is fixed.

$$\begin{aligned} I_{Dr}(t) &= \boldsymbol{\omega}^{\top} \cdot \boldsymbol{\Psi}(t) \\ &= [\boldsymbol{\omega^o}, \boldsymbol{\omega^f}]^{\top} \cdot [\boldsymbol{\Psi^o}(t), \boldsymbol{\Psi^f}(t)], \end{aligned} \tag{3}$$

where $\boldsymbol{\omega}$ represents the vector of weights, $\boldsymbol{\omega^o} = [\omega_0, \omega_1, \ldots, \omega_{N_o}]$, $\boldsymbol{\omega^f} = [\omega_{N_o+1}, \ldots, \omega_{N_o+N_f}]$, $\boldsymbol{\Psi^o}(t) = [1, K(t,t_1), \ldots, K(t,t_{N_o})]$ and $\boldsymbol{\Psi^f}(t) = [K(t,t_{N_o+1}), \ldots, K(t,t_{N_o+N_f})]$ represent the weight and kernel function vector based on the observed and forecasted future degradation increment data, respectively. They are the important part of the proposed augmented model. N_o and N_f are the sample size of the observed and forecasted data set, respectively. The interval is $\tau = t_{i+1} - t_i$, $i = 1, 2, \ldots, N_o + N_f$. For RUL prediction, t_{N_o} denotes the start of prediction (SOP).

Based on the proposed augmented degradation process model, the first hitting time (FHT) is utilized to define the lifetime as $T = \inf\{t : X(t) \geq \xi | X(0) \leq \xi, t > 0\}$. Under the concept of the first hitting time (FHT), the RUL at the SOP t_{N_o} is defined as $L = \inf\{l : X(t_{N_o} + l) \geq \xi | x(t_{N_o}) \leq \xi, l > 0\}$.

Then, the derivation of the RUL distribution is investigated. Based on the WP model (1), the probability density function (PDF) of the FHT for $X(t)$ crossing a constant boundary ξ is equivalent to calculating the PDF of the FHT for a standard BM $\{B(t), t \geq 0\}$ crossing a time-varying boundary $S(t) = \frac{1}{\sigma} \cdot (\xi - \int_0^t \mu(\tau;\boldsymbol{\theta})d\tau)$ as follows [4].

$$P_{X(t)}(\xi,t) \cong \frac{1}{\sqrt{2\pi t}} \cdot \left(\frac{S(t)}{t} + \frac{1}{\sigma} \cdot \mu(t;\boldsymbol{\theta})\right) \cdot \exp\left(-\frac{S^2(t)}{2t}\right). \tag{4}$$

To derive the RUL distribution based on the proposed degradation process model (2) and (3), an assumption needs to be made at first that the SOP t_{N_o}, the FHT T and the RUL L_k satisfy $t_{N_o} = N_o\tau$, $T = N_T\tau$, $N_T \in N^+$ and $L = N_L\tau$, $N_L \in N^+$ respectively. And $N_o + N_f > N_T$.

By applying the central difference to approximate $\mu(t;\boldsymbol{\theta})$ in (4), the PDF value of the FHT of degradation state $X(t)$ crossing ξ can be calculated numerically by

$$\begin{aligned} f_T(N_T\tau) = &\frac{1}{\sqrt{2\pi N_T\tau}} \cdot \exp\left(-\frac{S^2(N_T\tau)}{2N_T\tau}\right) \\ &\times \left(\frac{S(N_T\tau)}{N_T\tau} + \frac{1}{\sigma} \cdot \frac{I_{Dr}(N_T\tau) + I_{Dr}((N_T+1)\tau)}{2\tau}\right), \end{aligned} \tag{5}$$

where $S(N_T\tau) = \frac{1}{\sigma} \cdot (\xi - \sum_{j=1}^{N_T} I_{Dr}(j\tau))$.

After applying a transformation $Y(l) = X(l + t_{N_o}) - X(t_{N_o})$ with $l \geq 0$, the PDF of the RUL is equivalent to the PDF of the FHT of the process $\{Y(l), l \geq 0\}$, crossing the threshold $\xi_{t_{N_o}} = \xi - X(t_{N_o})$. According to (5), the PDF value of the RUL is formulated numerically at the SOP t_{N_o} as follows.

$$\begin{aligned} f_L(N_L\tau) = &\frac{1}{\sqrt{2\pi N_L\tau}} \cdot \exp\left(-\frac{S^2(N_L\tau)}{2N_L\tau}\right) \\ &\times \left(\frac{S(N_L\tau)}{N_L\tau} + \frac{1}{\sigma} \cdot \frac{I_{Dr}(N_L\tau) + I_{Dr}((N_L+1)\tau)}{2\tau}\right), \end{aligned} \tag{6}$$

where $S(N_L\tau) = \frac{1}{\sigma} \cdot (\xi - X(N_o\tau) - \sum_{j=N_o+1}^{N_o+N_L} I_{Dr}(j\tau))$.

In this brief, the mean of the RUL (MRUL) is taken as the predicted RUL at the SOP t_{N_o}. The calculation of the unknown parameters $\{\boldsymbol{\omega}, \sigma\}$ in (6) is introduced in the following sections.

III. Sparse Bayesian Algorithm for Degradation Parameter Calculation

In this section, a sparse Bayesian parameter calculation algorithm [12] is adopted for the proposed degradation model. Given an augmented degradation increment data set of $\{t_k, I_X(t_k)\}_{k=1}^{N_o+N_f}$, $I_X(t_k;\tau) \in R, t_k \in R$, where the observation part $\{I_X(t_k)\}_{k=1}^{N_o}$ is obtained by $\{X(t_k) - X(t_{k-1})\}_{k=1}^{N_o}$, and the augmented future part $\{I_X(t_k)\}_{k=N_o+1}^{N_o+N_f}$ is forecasted by LSTM network using $\{I_X(t_k)\}_{k=1}^{N_o}$ which will be introduced in the next section. Based on the model (1), WP has excellent mathematical properties including the BM has stationary and independent increments. $\{I_X(t_k;\tau)\}_{k=1}^{N_o+N_f}$ are independent and the Gaussian distribution of $I_X(t_k;\tau)$ depends on t_k only through the difference τ.

Following the assumption of independence for $I_X(t_k)$ if τ is fixed, the likelihood of the complete data $\{I_X(t_k)\}_{k=1}^{N_o+N_f}$ can be written as

$$p(\boldsymbol{I_X}|\boldsymbol{\omega}, \sigma^2\tau) = (2\pi\sigma^2\tau)^{-m/2} exp\left(-\frac{\|\boldsymbol{I_X} - \boldsymbol{\Phi\omega}\|^2}{2\sigma^2\tau}\right), \tag{7}$$

where $\boldsymbol{I_X} = [I_X(t_1), \ldots, I_X(t_{N_o+N_f})]^{\top}$, $\boldsymbol{\omega} = [\omega_0, \omega_1, \ldots, \omega_{N_o+N_f}]^T$, and $\boldsymbol{\Phi}$ is an $(N_o+N_f) \times (N_o+N_f+1)$ design matrix with $\boldsymbol{\Phi} = [\boldsymbol{\Psi}(t_1), \boldsymbol{\Psi}(t_2), \ldots, \boldsymbol{\Psi}(t_{N_o+N_f})]^{\top}$ and $\boldsymbol{\Psi}(t_i) = [1, K(t_i,t_1), \ldots, K(t_i,t_{N_o+N_f})]$ as the definition in the model (3).

Because of the same quantity of weights in the model as the size of the data examples $\boldsymbol{I_X}$, the direct maximum likelihood estimation (MLE) of $\boldsymbol{\omega}$ and σ from (7) may result in overfitting [12]. In order to avoid this, additional measurements are required to set additional constraints on the parameters from

a Bayesian point of view by explicitly defining a zero-mean Gaussian prior probability distribution over $\boldsymbol{\omega}$, i.e.,

$$p(\boldsymbol{\omega}|\boldsymbol{\alpha}) = \prod_{i=0}^{N_o+N_f} \frac{\alpha_i}{\sqrt{2\pi}} exp\left(-\frac{\omega_i^2\alpha_i^2}{2}\right), \tag{8}$$

with $\boldsymbol{\alpha}$ as the vector of $N_o + N_f + 1$ hyperparameters, namely $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \ldots, \alpha_{N_o+N_f}]^\top$.

Hence, the posterior distribution over $\boldsymbol{\omega}$ is multivariate Gaussian distribution $p(\boldsymbol{\omega}|\boldsymbol{I_X}, \boldsymbol{\alpha}, \sigma^2\tau) = \boldsymbol{N}(\boldsymbol{\omega}|\boldsymbol{m}, \boldsymbol{\Sigma})$, where the posterior covariance is

$$\boldsymbol{\Sigma} = (\boldsymbol{A} + \sigma^{-2}\tau^{-1}\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}, \tag{9}$$

and $\boldsymbol{A} = diag(\alpha_0, \alpha_1, \ldots, \alpha_{N_o+N_f})$. The posterior mean is

$$\boldsymbol{m} = \sigma^{-2}\tau^{-1}\boldsymbol{\Sigma}\boldsymbol{\Phi}^\top\boldsymbol{I_X}, \tag{10}$$

The optimal values of $\boldsymbol{\alpha}$ and σ are derived using type-2 maximum likelihood to maximize the following marginal likelihood

$$\begin{aligned} p(\boldsymbol{I_X}|\boldsymbol{\alpha}, \sigma^2\tau) &= \int p(\boldsymbol{I_X}|\boldsymbol{\omega}, \sigma^2\tau) \cdot p(\boldsymbol{\omega}|\boldsymbol{\alpha}) d\boldsymbol{\omega} \\ &= (2\pi)^{-\frac{N_o+N_f}{2}} \cdot |\sigma^2\tau\boldsymbol{E} + \boldsymbol{\Phi}\boldsymbol{A}^{-1}\boldsymbol{\Phi}^\top|^{-\frac{1}{2}} \\ &\times exp\left(-\frac{1}{2}\boldsymbol{I_X}^\top(\sigma^2\tau\boldsymbol{E} + \boldsymbol{\Phi}\boldsymbol{A}^{-1}\boldsymbol{\Phi}^\top)^{-1}\boldsymbol{I_X}\right), \end{aligned} \tag{11}$$

where $\boldsymbol{E}$ represents the identity matrix. By simply setting the required derivatives of the marginal likelihood (11) to zero, the optimal solution can be determined by the following re-estimating equations.

$$\alpha_i^{new} = \frac{\gamma_i}{m_i^2}, \tag{12}$$

$$(\sigma^2\tau)^{new} = \frac{\|\boldsymbol{I_X} - \boldsymbol{\Phi}\boldsymbol{m}\|^2}{N_o + N_f - \Sigma_{i=0}^{N_o+N_f}\gamma_i}, \tag{13}$$

where

$$\gamma_i = 1 - \alpha_i\Sigma_{ii}, \tag{14}$$

and m_i is the ith posterior mean weight of $\boldsymbol{m}$, Σ_{ii} refers to the ith diagonal element of $\boldsymbol{\Sigma}$.

Therefore, if (12) and (14) are applied repeatedly, the algorithm can simultaneously update the posterior statistics $\boldsymbol{m}$ and $\boldsymbol{\Sigma}$ from (9) and (10) until the appropriate convergence criteria are satisfied. As a result, the mean weights $\boldsymbol{\omega}$ by (10) and σ by (13) can be estimated. Recall (3), and the numerical result of the RUL distribution can be derived by (6).

IV. LSTM Network for Degradation Increment Forecasting

In this section, the degradation increment time series forecasting using LSTM network is introduced. In order to obtain the augmented future degradation increment in the above section by forecasting with the observed degradation increment data, it is essential to dig out the inherent relations and long-term dependencies from time series observation data which are well suited for LSTM network to process.

The LSTM network gets the correlation information contained in the training time series data and store it in the network connection weights and biases. An LSTM network is usually constructed with multiple hidden layers of LSTM cells. The central idea of LSTM cell lies in that some special mechanisms, namely, three 'gates' (i.e., input, forget, and output gate) can regulate the information flow into and out of the cell. The following equations can describe the mathematical function of a LSTM cell at the current time step t

$$\begin{aligned} h_t &= f(a_t, h_{t-1}; \Theta_{LSTM}) \\ &= \begin{cases} i_t = \sigma(W_i a_t + R_i h_{t-1} + b_i) \\ f_t = \sigma(W_f a_t + R_f h_{t-1} + b_f) \\ \tilde{s}_t = \tanh(W_s a_t + R_s h_{t-1} + b_s) \\ s_t = \tilde{s}_t * i_t + s_{t-1} * f_t \\ o_t = \sigma(W_o h_t + R_o h_{t-1} + b_o) \\ h_t = \tanh(s_t) * o_t \end{cases} \end{aligned} \tag{15}$$

where h_t and a_t represents the output (also called hidden state) and input of the cell at the current time step t, respectively. Θ_{LSTM} is the parameter set of the input weights W_*, the recurrent weights R_* and the bias terms b_*, which is to be learned during the training process. The state activation function is $\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$. The forget gate f_t and the input gate i_t decide whether the cell state of the previous time step s_{t-1} and the current cell state candidate $\tilde{s}_t$ are memorized and stored in the internal state s_t or not. Then the output gate o_t controls whether the current internal state s_t is output as the current hidden state of a LSTM cell h_t or not. These three gates are activated through the sigmoid function $\sigma(z) = \frac{1}{1+e^{-z}}$ by the weighted sum of a_t and the hidden state of the previous time step h_{t-1} and b_*.

In order to forecast the degradation increment sequence of future time steps, a sequence-to-sequence regression LSTM network is trained, where its responses are the training sequence shifted by k_s time step. In other words, at each time step of the input sequence, the LSTM network learns to forecast the value of the next k_s time step. Generally, an LSTM network is compose of k_l-hidden-layer-stacked LSTM cells of $k_{h,j}$ hidden units ($j = 1, .., k_l$) connected sequentially by the hidden state h and the internal state s. The structural parameters of the network k_l and k_h need to be designed according to the specific training sequence in the experiment. The 'LSTM training' module of Fig. 1 shows the architecture of the LSTM network for degradation increment forecasting. For each forecast, the previous k_s forecast value as input to the network is used.

After the LSTM network is trained with the observation part $\{I_X(t_k)\}_{k=1}^{N_o}$, the augmented future part $\{I_X(t_k)\}_{k=N_o+1}^{N_o+N_f}$ can be obtained by the trained network during the forecasting phase.

The steps of the proposed method are summarized in the following with a flowchart shown in Fig. 1.

Step1: Offline experimental and online degradation data are obtained through sensing technology. If some offline experimental degradation data are available to pre-train the LSTM network, it is helpful to update the network with online degradation data for accurate RUL prediction.

Step2: The LSTM network is trained using the offline and online available degradation increment data to forecast the future degradation increment and obtain the augmented increment data.

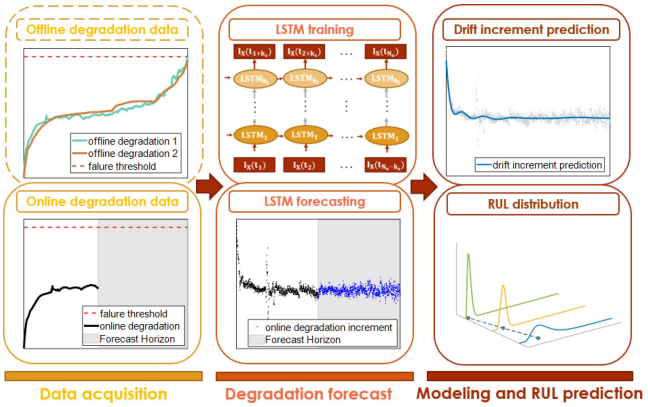

Fig. 1. The framework of the proposed method.

Step3: The sparse Bayesian algorithm is applied to estimate the parameters in (2) and predict the drift increment (3). The WP-based degradation model can be established, and the RUL distribution can be derived numerically by (6).

V. Experiments

In this section, the presented method is applied for the RUL prediction of lithium-ion batteries to demonstrate the effectiveness. For validating the performance of the degradation time series forecasting, LSTM network is compared with support vector regression (SVR) and recurrent neural network (RNN) [10]. And, the superiority of the proposed method is validated by the comparison between it and two existing state-of-art WP-based methods. They are method 1 (the power-law model method in [4]) and method 2 (the adaptive model method in [6]).

A. Data Description and Experiment Preparation

The degradation data sets [13] provided by the NASA Ames Prognostics Center of Excellence are used. A set of four lithium-ion batteries (#5, #6, #7 and #18) were run through 3 different operational profiles, where only the aging progresses of #6 and #18 are under the same operational condition. The capacity of battery is taken as the degradation state $X(t)$ and t denotes cycle of charging and discharging here. The failure threshold is 1.4Ahr. The actual lifetime of #7 can not be determined because of its always higher capacity than 1.4Ahr in this experiment. Therefore, only the other three batteries are conducted the RUL prediction, the lifetime of battery #5, #6 and #18 is $T_{\#5} = 125$ cycles, $T_{\#6} = 109$ cycles and $T_{\#18} = 112$ cycles. To use the proposed method, all the original capacity data of battery #5, #6 and #18 is converted to their initial capacities minus all battery capacities and then their degradation increment data is obtained after preprocessing with $\tau = 1$ cycle. The forecasted trajectories of the batteries capacity is obtained after the inverse transformation. Comparison studies including LSTM network forecasting and RUL prediction are conducted at the wth ($w = 30, 40, \ldots, 90$) percentile of the battery lifetime $T_{\#b}$ ($b = 5, 6, 18$). That is, $N_{o,w,b} = \lfloor w\% \times T_{\#b} \rfloor$, and the SOP of battery #b takes $t_{N_{o,w,b}} = N_{o,w,b}$ cycle for testing.

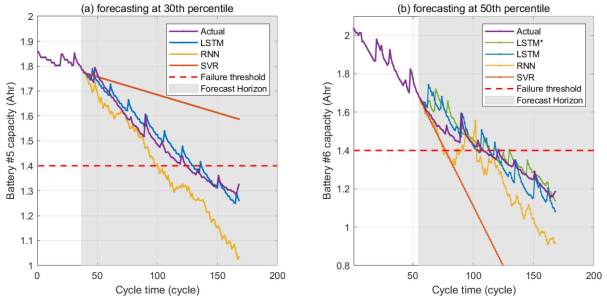

Fig. 2. Forecasted degradation trajectories: (a) forecasting at 30th percentile of Battery #5's lifetime, (b) forecasting at 50th percentile of Battery #6's lifetime.

TABLE I
Average RMSE Comparison Between Different Forecasting Methods

method	Battery		
	#5	#6	#18
SVR	0.074	0.326	0.166
RNN	0.059	0.183	0.094
LSTM	**0.028**	0.065	0.035
LSTM*	-	**0.044**	**0.027**

B. Training and Forecasting of LSTM Network

The LSTM network has the same structure as the RNN used for comparison, with two hidden layers, the first and the second with 50 and 30 hidden units, respectively. And, that is $k_l = 2, k_{h,1} = 50, k_{h,2} = 30$. The training hyper-parameters learning rate, forecasting step k_s and the dropout probability are set as 0.01, 1 and 0.5, respectively. For a better fit when training, standardizing the training observed increment data to have zero mean and unit variance is necessary. LSTM network is compared with RNN and SVR at different SOPs of the degradation process of the three batteries. Fig. 2 gives different methods' forecasted paths at 30th and 50th percentile of battery #5 and #6's lifetime, respectively. Since battery #6 (or #18) is under the same experimental condition as #18 (or #6), the LSTM* for battery #6 (or #18) means further online training and updating the pre-trained LSTM network by several epochs based on the offline data of battery #18 (or #6). As shown in Fig. 2, the forecasted paths using LSTM and LSTM* fit the actual ones of #5 and #6 better than the other methods in the relatively early stage of degradation.

Table I shows the average of root mean square errors (RMSE) at each SOP for battery #b, which is adopted to measure the degradation forecasting precision of the three methods.

$$RMSE_{t_{N_{o,w,b}}} = \sqrt{\frac{1}{N_{e,b} - N_{o,w,b}} \sum_{i=N_{o,w,b}+1}^{N_{e,b}} (X_b(t_i) - \hat{X}_b(t_i))^2}, \quad (16)$$

where $N_{e,b}$ is the total size of degradation data of battery #b. $\hat{X}$ denotes the forecasted degradation value after the inverse transformation. The average RMSEs of LSTM network are smaller than that of SVR and RNN, and LSTM's forecasting of degradation is more accurate. If there is offline data training

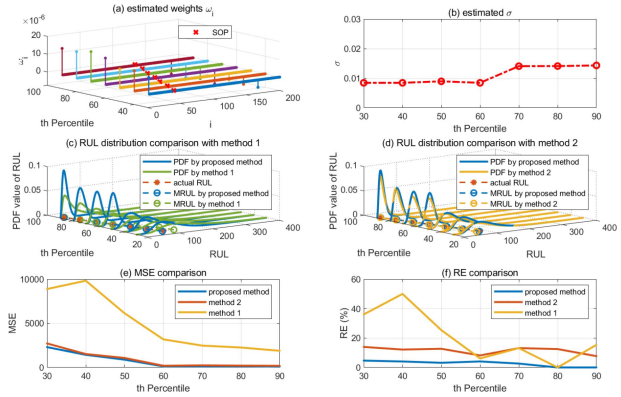

Fig. 3. RUL prediction results of battery #5 at each SOP: (a) weights estimation, (b) σ estimation, (c) RUL distribution comparison with method 1, (d) RUL distribution comparison with method 2, (e) MSE comparison, (f) RE comparison.

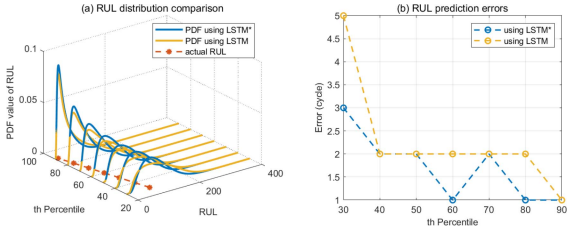

Fig. 4. RUL prediction results of battery #6 at each SOP: (a) RUL distribution comparison between using LSTM and LSTM*, (b) RUL prediction errors comparison.

in advance, the forecasting accuracy is further improved (see the average RMSEs of LSTM* are smaller than LSTM in Table I). It is proved that LSTM forecasting can lay a good foundation for the following RUL prediction step.

C. RUL Prediction and Comparison Study

Based on the augmented degradation data obtained by LSTM and LSTM*, RUL prediction results of Battery #5 and #6 are used to prove the superiority of the proposed method in **Step3**. The kernel function is chosen as $K(t,t_i)=(t-t_i)^p$, where p is 1.2 and 1.3 for battery #5 and #6, respectively.

Fig. 3 shows parameter estimation results of the proposed model (2) and the comparison results of the RUL distribution of battery #5 at each SOP using the proposed model with using method 1 and 2. Notice that sparse weights (see Fig. 3 (a)) consists of observation and augmented future part by LSTM network, which are also the key to the proposed method being more accurate in degradation modeling and RUL prediction than method 1 and 2 with smaller MSEs and REs especially in the early stage of degradation (see Fig. 3 (b) to (d)). The mean squared error (MSE) (see Fig. 3 (e)) and relative error (RE) (see Fig. 3 (f)) are employed to assess the RUL prediction error. The MSE and RE at SOP t_{N_o} are calculated as $MSE_{t_{N_o}} = \int_0^\infty (l-\tilde{l}_{t_{N_o}})^2 f_L(l)dl$, $RE_{t_{N_o}} = \frac{|l_{t_{N_o}}-\tilde{l}_{t_{N_o}}|}{\tilde{l}_{t_{N_o}}}$, where $\tilde{l}_{t_{N_o}} = T_{\#b} - t_{N_o}$ is the actual RUL at SOP t_{N_o}, $f_L(\cdot)$ is the PDF of the RUL and $l_{t_{N_o}}$ is the predicted RUL, i.e., MRUL at t_{N_o}. The average MSE of the proposed method, method 1 and 2 are 429, 877 and 4951, respectively. And the average RE of them are 2.6%, 11.4% and 20.8%, respectively. The RUL prediction accuracy of the proposed method is better than method 1 and 2 (without LSTM network), especially in the early stage of degradation (see comparison results from 30 to 50th percentile).

In addition, to demonstrate the contribution of LSTM* to RUL prediction over LSTM, comparison of RUL prediction for battery #6 using LSTM and LSTM* is shown in Fig. 4 (a). The errors $\tilde{l}_{t_{N_o}} - l_{t_{N_o}}$ of the RUL prediction at each SOP using LSTM* are mostly smaller than using LSTM (see Fig. 4 (b)).

VI. Conclusion

In this brief, a new WP-based LSTM-assisted RUL prediction method is proposed. Compared with the existing WP-based approaches, it incorporates LSTM network and fuse the offline experimental and online degradation data to predict. Compared with the existing LSTM-based approach [10] using Monte Carlo method to quantify RUL prediction uncertainty, it presents an approximation formula for the PDF of the RUL to relieve computation burden. Its practical importance lies in the accuracy improvement for the degraded batteries' RUL prediction and achieves small prediction error in the early stage of the degradation.

References

[1] M. Pecht, "Prognostics and health management of electronics," in *Encyclopedia of Structural Health Monitoring*. Chichester, U.K.: Wiley, 2009.

[2] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation—A review on the statistical data driven approaches," *Eur. J. Oper. Res.*, vol. 213, no. 1, pp. 1–14, 2011.

[3] Z. Zhang, X. Si, C. Hu, and Y. Lei, "Degradation data analysis and remaining useful life estimation: A review on Wiener-process-based methods," *Eur. J. Oper. Res.*, vol. 271, no. 3, pp. 775–796, 2018.

[4] X.-S. Si, W. Wang, C.-H. Hu, D.-H. Zhou, and M. G. Pecht, "Remaining useful life estimation based on a nonlinear diffusion degradation process," *IEEE Trans. Rel.*, vol. 61, no. 1, pp. 50–67, Mar. 2012.

[5] H. Sun, J. Pan, J. Zhang, and D. Cao, "Nonlinear Wiener process-based cutting tool remaining useful life prediction considering measurement variability," *Int. J. Adv. Manuf. Technol.*, vol. 107, nos. 11–12, pp. 4493–4502, 2020.

[6] X.-S. Si, "An adaptive prognostic approach via nonlinear degradation modeling: Application to battery data," *IEEE Trans. Ind. Electron.*, vol. 62, no. 8, pp. 5082–5096, Aug. 2015.

[7] Z. Huang, Z. Xu, W. Wang, and Y. Sun, "Remaining useful life prediction for a nonlinear heterogeneous Wiener process model with an adaptive drift," *IEEE Trans. Rel.*, vol. 64, no. 2, pp. 687–700, Jun. 2015.

[8] K. Khalil, O. Eldash, A. Kumar, and M. Bayoumi, "Economic LSTM approach for recurrent neural networks," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 66, no. 11, pp. 1885–1889, Nov. 2019.

[9] Z. Gao, T. Yuan, X. Zhou, C. Ma, K. Ma, and P. Hui, "A deep learning method for improving the classification accuracy of SSMVEP-based BCI," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, early access, Mar. 26, 2020, doi: 10.1109/TCSII.2020.2983389.

[10] Y. Zhang, R. Xiong, H. He, and M. G. Pecht, "Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 5695–5705, Jul. 2018.

[11] C.-G. Huang, H.-Z. Huang, and Y.-F. Li, "A bidirectional LSTM prognostics method under multiple operational conditions," *IEEE Trans. Ind. Electron.*, vol. 66, no. 11, pp. 8792–8802, Nov. 2019.

[12] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001.

[13] B. Saha and K. Goebel, "Battery data set," in *NASA Ames Prognostics Data Repository*, NASA Ames Res. Center, Moffett Field, CA, USA, 2007. [Online]. Available: http://ti.arc.nasa.gov/project/prognostic-data-repository