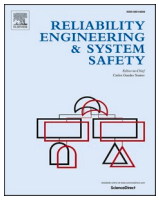
Reliability Engineering and System Safety 222 (2022) 108391

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress



Inspection schedule for prognostics with uncertainty management

Seokgoo Kim [a,b], Joo-Ho Choi [b,*], Nam Ho Kim [a]

[a] Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL, 32611, USA
[b] Department of Aerospace and Mechanical Engineering, Korea Aerospace University, Goyang-si 10540, Republic of Korea

ARTICLE INFO

Keywords:
Uncertainty reduction
Prognostics
Data measurement
Inspection schedule
Epistemic
Aleatory

ABSTRACT

Condition monitoring data is an essential ingredient for prognostics and health management. To minimize unnecessary inspections or measurements, it is crucial to evaluate the value of data to be measured in advance and determine the inspection or data measurement schedule. For this purpose, it is important to predict how much prognostics performance will be improved by adding additional data. Motivated by this objective, this paper proposes a new method that determines the future data measurement schedule which can reduce the uncertainty in prediction to the desired level. The proposed method decomposes the prediction uncertainty into epistemic and aleatory uncertainty, which are caused by the uncertainty of model parameters and the noise in the data, respectively. Then, contributions of these uncertainties to the overall prediction uncertainty in the future are analyzed. The next measurement schedule is determined such that the level of reducible epistemic uncertainty in the prediction is the same as that of aleatory uncertainty. The proposed method is applied to two different prognostics approaches: the model-based and data-driven methods. Two examples showed that the total number of inspections is reduced by about 85% while keeping the same level of prediction uncertainty.

1. Introduction

Prognostics and health management (PHM) estimates the system's health condition and predicts the remaining useful life (RUL) to eliminate unnecessary maintenance and prevent unexpected failure. It is well known that the PHM consists of fault diagnostics and prognostics, and numerous researches have been conducted in these two fields [1–3]. Although there are different aspects for these two functions, both of them use the condition monitoring data obtained via installed sensors or inspection. Therefore, appropriate data measurement is the key ingredient to building successful PHM functions that satisfy the desirable requirements of users [4,5]. Among several issues related to data measurement, this paper focuses on the data measurement schedule; i.e., inspection schedule. In this paper, the terminology of measurement and inspection are used as the same meaning. In the era of big data and industry 4.0, it is not surprising that many industries are trying to install sensors to target systems and gather condition monitoring data. In the view of decision-makers, however, there are several practical problems to be considered. First, the more data are monitored, the bigger data storage is required. Big data storage is directly associated with a high operating cost. Also, too frequent measurement or inspection strategy can result in frequent downtime, which makes PHM inefficient unless measurement does not interrupt system operation. Second, the condition monitoring data that is required for the PHM is degradation data over time. Many mechanical systems do not degrade fast, which means most monitoring data will not show noticeable degradation. In this sense, continuous data measurement may increase the required data storage and does not provide much benefit to PHM quality. Therefore, it would be beneficial to develop a proper methodology that determines an optimum data measurement schedule.

Motivated by the abovementioned reasons, many researchers have aimed to determine the optimum inspection cycle. Inspection or measurement policy can be generally categorized into two groups, i.e., corrective maintenance (CM) and preventive maintenance (PM) [6–8]. Although there is much literature regarding CM-based inspection scheduling, this paper focuses on the PM-based maintenance policy since it is well known that PM brings more benefits than CM in terms of condition-based maintenance. Recently, many researchers have investigated the optimum inspection/maintenance schedule considering the failure process of the machine [9–12], economic benefits [13–21], or system's reliability [22–27]. In addition, some research have focused on joint optimization of production scheduling and maintenance [28–37]. For example, Papakonstantinou et al [10,11]. optimized inspection and maintenance strategies using stochastic models and real-time uncertain

* Corresponding author.
E-mail addresses: seokgoo.kim@ufl.edu (S. Kim), jhchoi@kau.ac.kr (J.-H. Choi), nkim@ufl.edu (N.H. Kim).

https://doi.org/10.1016/j.ress.2022.108391
Received 16 August 2021; Received in revised form 12 February 2022; Accepted 13 February 2022
Available online 15 February 2022

data. For this purpose, they employed the partially observable Markov decision process (POMDP). However, the main drawback is that the model is hard to solve and requires large models which contain various types of health states. Yuan et al [12]. quantified the economic gain from an inspection and preventive maintenance program for managing an aging component population. They proposed a method based on the Bayesian value of information (VoI) with the gamma process of degradation. However, they assumed that degradation data is free of measurement noise. In addition, they did not consider how many unnecessary inspections can be removed. Nguyen et al [15]. proposed a method that improves the inspection/monitoring policy, reducing the maintenance and operating costs. They investigated different inspection policies for multiple degradation models of the system, and proposed prognostic model evaluation criteria. Based on the above information, they aimed to find the optimum inspection policy based on multi-objective optimization. Papakostas et al [16]. addressed the optimal maintenance scheduling based on the cost, RUL, operational risk and flight delay. Liu et al [17]. determined the inspection policy to minimize cost rate per time unit within the POMDP framework. Levitin et al [18]. optimized the inspection and mission policies by minimizing the expected costs related to inspection, mission failure, and failure of the system. Lin et al [19]. proposed the VoI-based inspection for the system that consists of multiple binary components. Wang et al [20,21]. proposed a cost-driven predictive maintenance policy based on the model-based prognostics framework which combines the extended Kalman filter and a first-order perturbation method. Based on the predicted reliability, they selected a group of aircraft panels that should be repaired at a scheduled maintenance stop to minimize the cost. Zhang et al [24]. developed a novel reliability-centered maintenance optimization model for multicomponent reparable systems operating under s-dependent competing risks. Consilvio et al [26]. investigated a rolling horizon approach to suggest optimal predictive maintenance scheduling to reduce the risk of failure of the system. Shi et al [27]. developed a PM strategy optimization model by considering the lifecycle safety that is expressed as the joint time-dependent probability of failure in different time intervals. Lee and Mitici [38] introduced a multi-objective maintenance design method based on reliability and cost-efficiency. Cavalcante et al [39]. modeled inspections by using the delay time concept and focused on what maintenance to schedule at a known time. Mancuso et al [40]. presented a methodology supporting maintenance strategies for PHM-equipped systems. They built a probabilistic network model of the system between components, sensors, maintenance action, and inspection and optimized the target system's inspection and maintenance strategy. Camici [41] presented a method suggesting maintenance scheduling of geographically distributed assets considering the predicted failure risks, travel and maintenance cost. Fanriat and Zio [42] suggested the VoI as a metric to schedule inspections. The next best inspection is determined based on the current condition and the expected gain (i.e., economic benefit) from possible inspections. Wang et al [43]. determined the optimal preventive maintenance thresholds by minimizing the system maintenance cost within the framework of the semi-Markov decision process. Shi et al [44]. developed a condition-based maintenance decision framework for a multi-components system subject to a system reliability requirement. Inspection time is determined when the predictive system reliability is below the reliability requirement.

It is observed from the literature review that the keywords of existing methods are 'optimization' and 'stochastic model (i.e., probabilistic model)'. They focused on building a functional relationship between prognostics information (i.e., RUL or future degradation) and economic benefit and suggested an optimal data measurement schedule. However, most of the studies suggested that a large number of historical data is required to build an effective model, and this problem becomes worse for the probabilistic model that consists of several different health states of the system. In addition, the recent literature shows that existing approaches do not consider the effect of actual degradation data obtained by inspection on uncertainty reduction associated with the probabilistic degradation model [44]. In other words, most of the researches has focused on quantifying the uncertainty to determine the optimal inspection schedule based on economic benefits.

Rather than focusing on economic benefit and uncertainty quantification, this paper aims to suggest the future measurement schedule depending on how much uncertainty due to the lack of knowledge can be reduced by obtaining data in the future. Up to date, none of the studies addresses this aspect when determining the inspection or measurement schedule. In other words, this study is the first attempt to evaluate how much future uncertainty can be reduced and to determine the optimal future measurement time. Uncertainty can be categorized into aleatory and epistemic uncertainty, which are also called irreducible and reducible uncertainty, respectively. Data measurement or inspection can be interpreted as a mechanism to reduce epistemic uncertainty. For this purpose, identifying the source of uncertainty is the first step for appropriate uncertainty management. There have been representative studies that decompose the uncertainty into aleatory and epistemic uncertainty and evaluate their contributions. McFarland [45] presented an uncertainty decomposition method and applied it for sensitivity analysis of models with inputs relating to different uncertainty sources. Sankararaman et al [46]. proposed a methodology that separates the contributions of aleatory and epistemic uncertainty to the overall uncertainty.

Motivated by these methods, this paper extends the contribution analysis into the prediction uncertainty in the future and evaluates the reducibility of the prediction uncertainty. Then, the algorithm determines the proper future measurement time to maintain the level of epistemic uncertainty within that of aleatory uncertainty. Since the prognostics algorithms are generally classified into model-based and data-driven approaches, this paper introduces two different inspection scheduling methods depending on the model-based or data-driven prognostics algorithm. The contributions of this paper are to consider the effect of data measurement on uncertainty reduction on model-based and data-driven prognostics methods, to identify the main contributor to the future prediction uncertainty in the extrapolation region, and to determine the future inspection schedule to reduce the prediction uncertainty to the desired level.

The paper is organized as follows, the theoretical background about uncertainty decomposition is explained in Section 2. The proposed methods of determining the data measurement schedule using model-based and data-driven approaches are introduced in Section 3 using a simple exponential function example. Then, the proposed method is applied to the real prognostics dataset in Section 4. Finally, the paper is concluded with discussions in Section 5.

2. Uncertainty decomposition

In general, uncertainty can be categorized into two groups: aleatory and epistemic uncertainty [47]. The former refers to the randomness of data that cannot be reduced, while the latter represents the uncertainty that is derived from the lack of knowledge, such as the uncertain parameters of the model. The epistemic uncertainty can be reduced by adding more data to estimate the parameters. Different from epistemic uncertainty, aleatory uncertainty is an inherent variation caused by the noise or the variability of data. Therefore, it is also referred to as variability. In many cases, two uncertainties are combined and result in overall uncertainty [48,49]. Since these two uncertainties are different in nature, however, it is important to decompose the overall uncertainty into aleatory and epistemic uncertainty to manage them effectively.

This paper evaluates the contribution of each uncertainty source using the method proposed in the reference [46]. The contributions of aleatory and epistemic uncertainty to the overall uncertainty are analyzed using global sensitivity analysis (GSA). Consider a model with a vector of uncertain input variables $\boldsymbol{X} = \{X^1, X^2, \ldots, X^n\}$ and output Y,

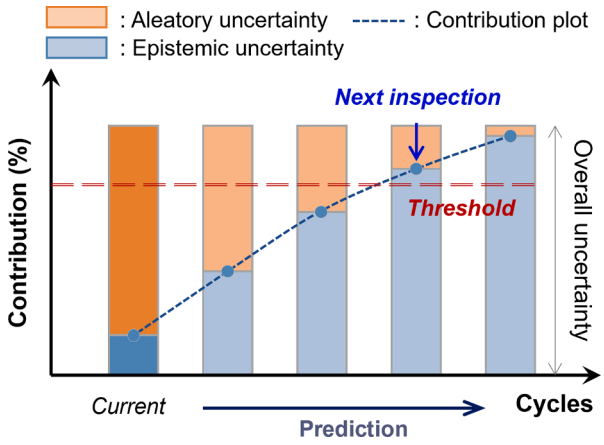


Fig. 1. Contribution plot of epistemic uncertainty to the overall uncertainty.

and each input variable has its own PDF function. In GSA, there are two effects; named the first-order indices and the total indices. The first-order sensitivity index of input variable X^i, which measures the sole contribution of X^i to the variance of the output, is calculated as:

$$S_1^i = \frac{V_{X^i}\left(E_{X^{-i}}\left(Y|X^i\right)\right)}{V(Y)} \tag{1}$$

where X^i is fixed first at a value and the expectation of the output is calculated by considering the variation in all other input variables (denoted as X^{-i}). Then the variance of the expectation of output from different realizations of X^i is calculated and divided by the variance of output. A high value indicates that the variable contributes significantly to the variance of the output. In the same manner, higher-order sensitivity indices which measure the interaction between variables can be calculated by fixing more than one input variable and subtracting individual effects. For example, the second-order sensitivity index, $S_2^{i,j}$ is obtained by

$$S_2^{i,j} = \frac{V_{X^{i,j}}\left(E_{X^{-i,j}}\left(Y|X^{i,j}\right)\right)}{V(Y)} - S_1^i - S_1^j \tag{2}$$

Lastly, the total sensitivity index, S_T^i is expressed as the sum of the first-order and all the interaction effects of X^i. The total sensitivity index can be calculated as

$$S_T^i = 1 - \frac{V_{X^{-i}}\left(E_{X^i}\left(Y|X^{-i}\right)\right)}{V(Y)} \tag{3}$$

A low value of the total sensitivity index indicates that the contribution of the input variable is not significant. Generally, the first-order and total indices are calculated together in GSA. Details about GSA can be found in [46,50,51].

It is noted that GSA cannot be applied for the case when two different sources of uncertainty are present. Sankararaman [46] proposed a method of combining both aleatory and epistemic uncertainties in GSA. To represent the aleatory uncertainty caused by the noise of data, an auxiliary variable, u, is employed. The concept of the auxiliary variable is based on the inverse transform sampling which is used to generate samples from the aleatory variable. A sample of u is drawn from the uniform distribution between 0 and 1. Then, the CDF of the random variable is inverted to find the corresponding samples of the random variable. Once the auxiliary variable is constructed, GSA can be performed. Different from the traditional GSA, sensitivity is decomposed into 'individual effects' and 'overall effects' instead of 'first-order effect' and 'total effect'. The individual (I) and overall (O) contributions of aleatory (U) and epistemic uncertainty (P) to the overall uncertainty in Y can be calculated as [46].

$$S_U^I = \frac{V_{U_Y}(E_{\boldsymbol{X}}(Y|u_Y))}{V(Y)} \tag{4}$$

$$S_U^O = 1 - \frac{V_{\boldsymbol{X}}(E_{U_Y}(Y|\boldsymbol{x}))}{V(Y)} \tag{5}$$

$$S_{\boldsymbol{P}}^I = \frac{V_{\boldsymbol{X}}(E_{U_Y}(Y|\boldsymbol{x}))}{V(Y)} \tag{6}$$

$$S_{\boldsymbol{P}}^O = 1 - \frac{V_{U_Y}(E_{\boldsymbol{X}}(Y|u_Y))}{V(Y)} \tag{7}$$

where $\boldsymbol{X}$ denotes the input variable associated with epistemic uncertainty, and $\boldsymbol{x}$ means a particular realization of $\boldsymbol{X}$. In practical applications, there are multiple sources of epistemic uncertainty. Details about how $S_{\boldsymbol{P}}^I$ and S_U^O are calculated will be explained in the following section.

Fig. 1 illustrates the result of contribution analysis over time, which plots $S_P^{\mathrm{I}} + S_U^{\mathrm{O}}(\%)$. Based on the model parameters estimated at the current time, a future prediction interval can be obtained. It should be noticed that the source of aleatory uncertainty is measurement noise that cannot be reduced even after the uncertainty in standard deviation, σ, is reduced. As mentioned above, the aleatory uncertainty is modeled with auxiliary variable, u, whose uncertainty remains constant over time since it follows uniform distribution between 0 and 1. Since the aleatory uncertainty cannot be reduced and remains the same over time, the epistemic uncertainty will become the main contributor to the overall uncertainty as the prediction interval increases. Therefore, when epistemic uncertainty becomes the main contributor to the overall uncertainty, it is beneficial to measure data.

3. Methodology

3.1. Model-based approach

In the model-based approach, it is assumed that the degradation model exists. To predict the future degradation behavior, the particle filtering (PF) algorithm is chosen in this paper, which is widely used in model-based approaches. It estimates and updates the unknown model parameters in the form of a probability density function (PDF). The PF is based on the Bayesian theorem:

$$p(\boldsymbol{\theta}|\mathbf{z}) \propto L(\mathbf{z}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{8}$$

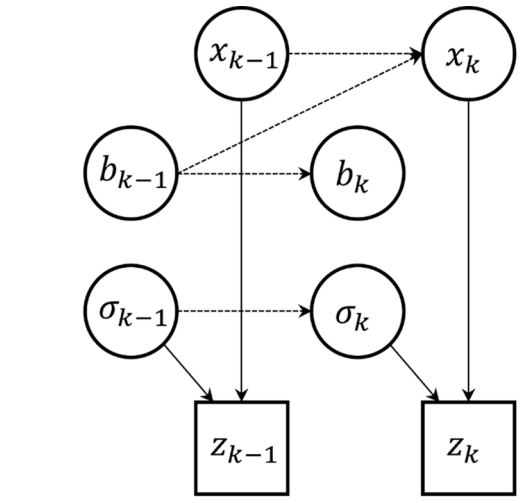
where $\boldsymbol{\theta}$ and $\mathbf{z}$ are the vectors of unknown parameters and observations, respectively, $p(\boldsymbol{\theta})$ is the prior PDF of unknown parameters, $L(\mathbf{z}|\boldsymbol{\theta})$ is the likelihood of observations given $\boldsymbol{\theta}$, and $p(\boldsymbol{\theta}|\mathbf{z})$ represents the posterior PDF of $\boldsymbol{\theta}$ given $\mathbf{z}$. In time series, the posterior PDF at the current time step is used as a prior for the next step and recursively updates the estimation for uncertain parameters. In PF, model parameters are represented by particles (i.e. samples). Generally, PF is expressed by two equations, state model $f(\cdot)$ and measurement model $g(\cdot)$ [52,53], as

$$x_k = f(x_{k-1}, b_{k-1}, v_{k-1}) \tag{9}$$

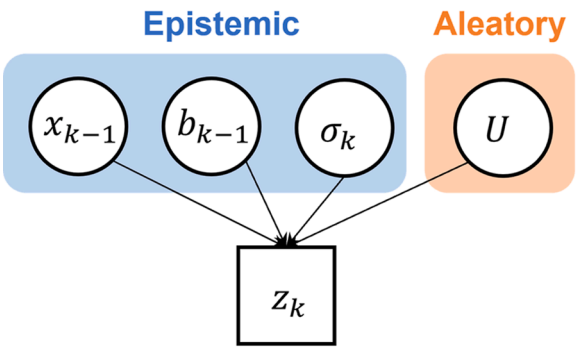
$$z_k = g(x_k, \sigma_k) \tag{10}$$

where x_k, b_k, and z_k represent the state, model parameter, and measurement at k^{th} cycle. The process noise and measurement noise are denoted as v_k and σ_k, respectively. In practical applications, the process noise can be ignored since it can be included in the model parameter's uncertainty [52,53]. In the following explanation, it is assumed that the measurement noise is normally distributed with a zero mean and standard deviation σ_k.

To decompose the contributions of uncertainties, it is necessary to identify the relationship between input variables (uncertain parameters) and the output variable (measurement). For this purpose, the PF process



(a) Dynamic Bayesian network of particle filter



(b) uncertainty sources of the measurement

Fig. 2. Uncertainty source identification.

is visually illustrated using a dynamic Bayesian network in Fig. 2(a). As shown in the figure, z_k is a child node of x_k and σ_k. However, considering the facts that model parameter, b_k, and measurement noise, σ_k, do not change over time and x_k is also a child node of x_{k-1} and b_{k-1}, the relationship between each node can be illustrated in Fig. 2(b). Since the main interest of the proposed method is to identify the contribution of epistemic and aleatory uncertainty to overall prediction uncertainty (prediction interval), contribution analysis is conducted between x_{k-1}, b_{k-1}, σ_k and posterior prediction of z_k. When the posterior prediction of z_k follows a normal distribution, $N(f(x_{k-1}, b_{k-1}), \sigma_k)$, there are epistemic uncertainties in $f(x_{k-1}, b_{k-1})$ and σ_k, and aleatory uncertainty caused by the measurement noise. Epistemic uncertainties are represented with particles (samples) of x_k and σ_k at k^{th} cycle, and they can be reduced by adding more data. In the case of aleatory uncertainty, the auxiliary variable is employed as mentioned in Section 2. When the posterior prediction of z_k is calculated from the normal distribution, $N(f(x_{k-1}, b_{k-1}), \sigma_k)$, the samples of U_k generated from the uniform distribution between 0 and 1 are used to draw samples of prediction of z_k using the inverse CDF sampling method. Since the samples of the auxiliary variable are generated from the same uniform distribution, their relative contribution to the overall prediction uncertainty always remains the same. Therefore, the relative contribution is only changed by the change of epistemic uncertainty. In other words, even if the auxiliary variable does not represent the accurate magnitude of aleatory uncertainty, it is useful to quantify the relative magnitudes of aleatory and epistemic uncertainty.

Table 1
Individual and overall contributions of aleatory and epistemic uncertainty.

Contribution measure	Equation
Individual effect of aleatory (S_U^I)	S_U
Overall effect of aleatory (S_U^O)	$S_U + S_{U,x} + S_{U,b} + S_{U,\sigma} + S_{U,x,b,\sigma}$
Individual effect of epistemic (S_P^I)	$S_x + S_b + S_\sigma + S_{x,b} + S_{b,\sigma} + S_{x,\sigma} + S_{x,b,\sigma}$
Overall effect of epistemic (S_P^O)	$S_x + S_b + S_\sigma + S_{x,b} + S_{b,\sigma} + S_{x,\sigma} + S_{x,b,\sigma} + S_{U,x} + S_{U,b} + S_{U,\sigma} + S_{U,x,b,\sigma}$

Once the sources of uncertainty are identified, individual and overall contributions of aleatory and epistemic uncertainty can be calculated based on Eqs. (4)-7). For example, S_U^I can be obtained by calculating the ratio of variances of posterior predictions with and without fixing auxiliary variable, u, at the particular sample. In Eq (4), $V(Y)$ is the variance of prediction samples, including the effect of aleatory variable u. On the other hand, $E_X(Y|u_Y)$ is the expectation of posterior prediction samples when u is fixed at u_Y. Considering that u has an N_s numbers of samples, the same numbers of $E_X(Y|u_Y)$ will be obtained. Finally, the variance of an N_s numbers of expectation values becomes the numerator in Eq. (4). In the same manner, S_x, S_b, and S_σ can be calculated. Since multiple input variables are associated with the epistemic uncertainty, the individual and overall contributions of the aleatory and epistemic uncertainty are given in Table 1. Among the four effects of uncertainty, the individual effect of epistemic uncertainty, S_P^I, is calculated first, then S_U^O is obtained based on the relationship, $S_P^I + S_U^O = 100\%$. This is because, different from reference [46], this paper does not consider the individual auxiliary variables for the three input variables (x, b and σ). This simplification cannot assess the contribution of individual auxiliary variables but can assess the combined contribution from the three variables, which is enough for our purpose. It is also reasonable to consider a single auxiliary variable because three epistemic parameters (i.e., x, b, σ) share the weights within the PF framework. To resolve this problem, S_P^I which can be calculated accurately is used as a measure of the contribution of epistemic uncertainty. Then, S_U^O is calculated from $S_P^I + S_U^O = 100\%$.

To determine the future inspection schedule, the abovementioned contribution analysis is applied to the future prediction Fig. 3. illustrates the process of the proposed model-based inspection scheduling. Based on the current measurement data, PF estimates the unknown parameters, $\theta_k = \{x_k, b_k, \sigma_k\}$ in the form of samples and predicts future θ_{k_p} by using degradation model $f(\cdot)$. In this step, two points need to be noted. First, the model parameter, b_k, does not change over time and σ_k also does not change because the measurement variability remains the same. Second, in the stage of uncertainty management where the contribution analysis is performed, the posterior prediction samples of x_k and corresponding auxiliary variable U_k are employed to calculate $S_P^I(k)$ and $S_U^O(k)$. This process is made for the pre-defined prediction range. Within the prediction range, the first cycle, k_p, where $S_P^I(k_p)$ becomes larger than the pre-determined uncertainty threshold, s, is determined as the next inspection cycle.

3.2. Data-driven approach

For the data-driven approach, gaussian process regression (GPR) is employed since it is known as a powerful machine learning algorithm that can quantify prediction uncertainty. Different from the model-based method, it is not easy to determine the aleatory and epistemic uncertainty in a data-driven algorithm since there is no physical parameter. However, the hyperparameters in GPR are basically process

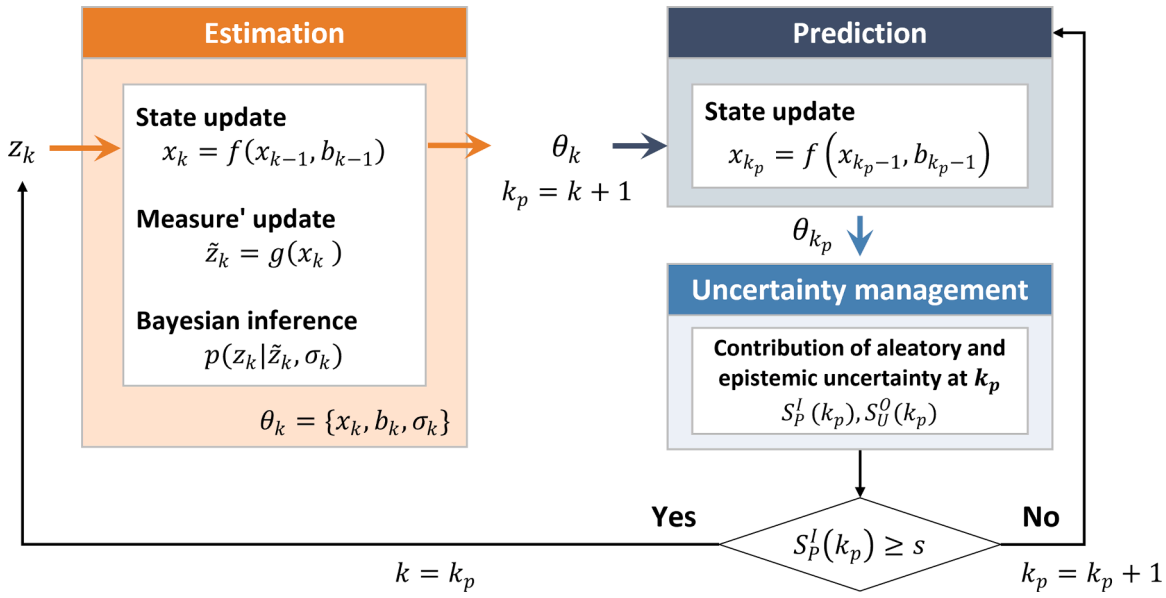


Fig. 3. Measurement scheduling for model-based.

noise and measurement noise which is combined in the form of the covariance matrix. The GPR, $f(\mathbf{x})$ is represented by the global function, $m(\mathbf{x})$, and the covariance function, $K(\mathbf{x}, \mathbf{x})$, as follows [54–57]:

$$f(\mathbf{x}) \sim GP[m(\mathbf{x}), K(\mathbf{x}, \mathbf{x})] \tag{11}$$

where $GP[\cdot]$ represents the Gaussian Process and $\mathbf{x}$ is a input vector. Generally, the covariance function consists of two parts, $K(x_i, x_j) = k_f(x_i, x_j) + k_n(x_i, x_j)$, where $k_f(x_i, x_j)$ is the functional part, while $k_n(x_i, x_j)$ is the noise part,

$$k_f(x_i, x_j) = \sigma_f^2 \exp\left\{-\frac{(x_i - x_j)^2}{2h^2}\right\} \tag{12}$$

$$k_n(x_i, x_j) = \sigma_n^2 \tag{13}$$

where σ_f^2, h, and σ_n^2 are hyperparameters to be estimated. Usually, these hyperparameters are optimized by maximizing the following log-likelihood function, L, given training data set, $\{(x_i, y_i) | i = 1, 2, \ldots, N\}$.

$$L = \log p(\mathbf{y}|\mathbf{x}, \theta) = -\frac{1}{2}\log\left(\det\left(k_f + \sigma_n^2 I\right)\right) - \frac{1}{2}\mathbf{y}^T\left[k_f + \sigma_n^2 I\right]^{-1}\mathbf{y} - \frac{N}{2}\log 2\pi \tag{14}$$

Once the hyperparameters are optimized, the posterior prediction at x' is described by the following equation

$$\bar{f}|\mathbf{x}, \mathbf{y}, x' \sim N(f', cov(f')) \tag{15}$$

where $f' = k_f(x', \mathbf{x})[k_f(\mathbf{x}, \mathbf{x}) + \sigma_n^2 I]^{-1}\mathbf{y}$ is the mean prediction and $cov(f') = k_f(x', x') - k_f(x', \mathbf{x})[k_f(\mathbf{x}, \mathbf{x}) + \sigma_n^2 I]^{-1}k_f(\mathbf{x}, x')$ is the prediction

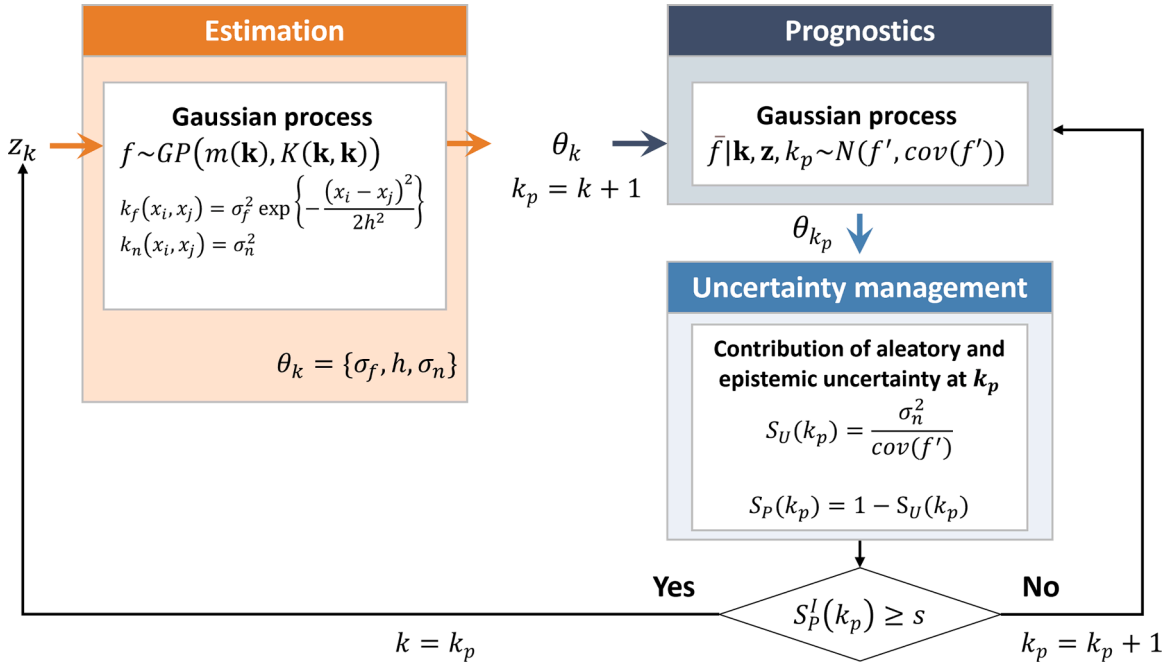


Fig. 4. Measurement scheduling for data-driven.

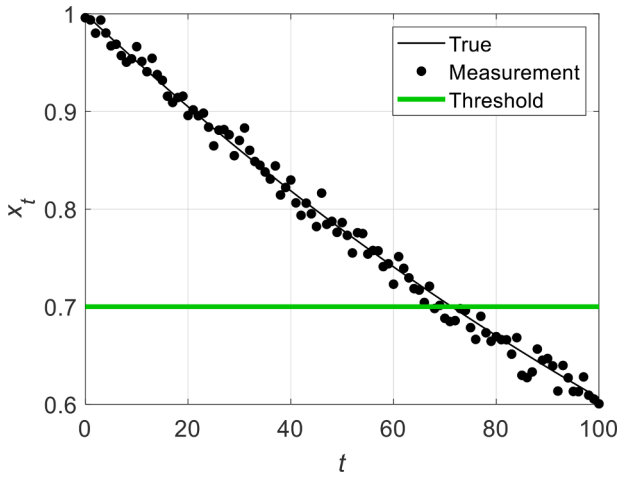


Fig. 5. Synthetic data generated from exponential degradation.

variance.

From the above equation, the uncertainty (i.e. variance) at new point $x^{'}$ has a contribution from the functional part (σ_f^2 and h) and measurement noise term (σ_n^2). In the view of uncertainty management, the measurement noise is regarded as an aleatory uncertainty since it represents the randomness of the signal. In this manner, the ratio of σ_n^2 to the overall uncertainty, $cov(f^{'})$ is defined as the contribution of aleatory uncertainty, S_U in GPR and the corresponding contribution of epistemic uncertainty, S_P can be defined as follows:

$$S_U = \frac{\sigma_n^2}{k_f(x', x') - k_f(x', x)\left[k_f(x, x) + \sigma_n^2 I\right]^{-1} k_f(x, x')} \tag{16}$$

$$S_P = 1 - S_U \tag{17}$$

Fig. 4 shows how to determine the future inspection schedule time, k_p. In the estimation part, hyperparameters, $\theta_k = \{\sigma_f, h, \sigma_n\}$, are estimated whenever the observation z_k comes in. Based on the trained GPR, the future $\bar{f}$ and RUL can be predicted in the prognostics part. Also, the prediction variance (i.e. overall uncertainty) at k_p cycle can be calculated and the contributions of aleatory and epistemic uncertainty, $S_P(k_p)$ and $S_U(k_p)$ are obtained. In the same manner with the model-based method, the k_p cycle when $S_P(k_p)$ becomes larger than the predetermined uncertainty threshold s is suggested as the potential inspection cycle.

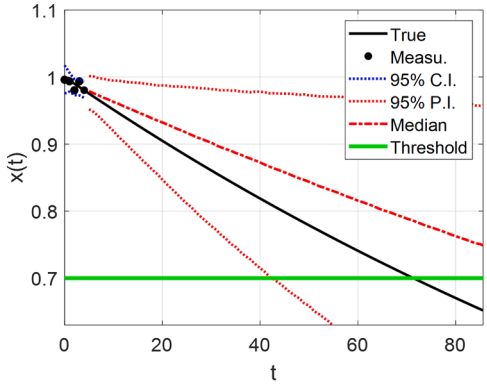
4. Application

4.1. Simulation data

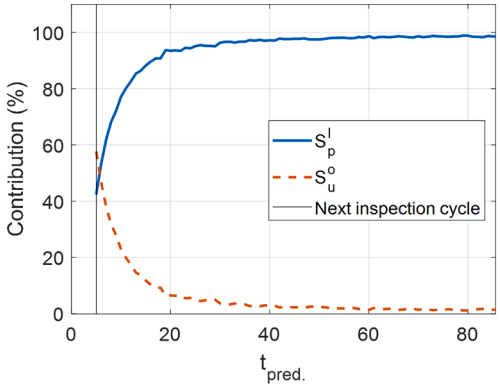
To verify the effectiveness of the proposed method. Synthetic degradation data are generated based on the following exponential function which is widely used in the field of battery degradation [53]:

$$x_k = a \cdot \exp(b \cdot k) \tag{18}$$

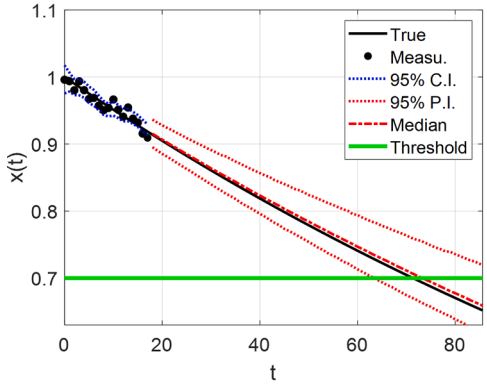
where a and b are model parameters, and x_k is the health condition of the target system at kth cycle. First, the true degradation data is generated with $b_{True} = -0.005$ and $a_{True} = 1$. To simulate the measurement environment with noise, normally distributed noise term,



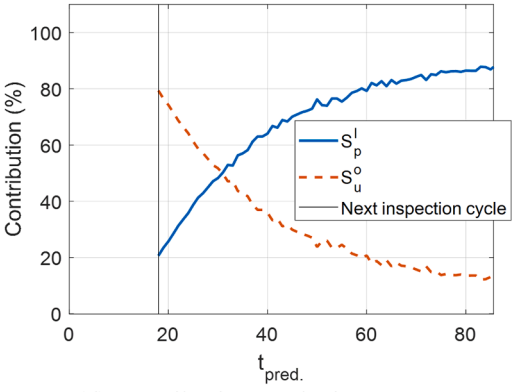
(a) prediction at $t = 4$



(b) contribution analysis at $t = 4$

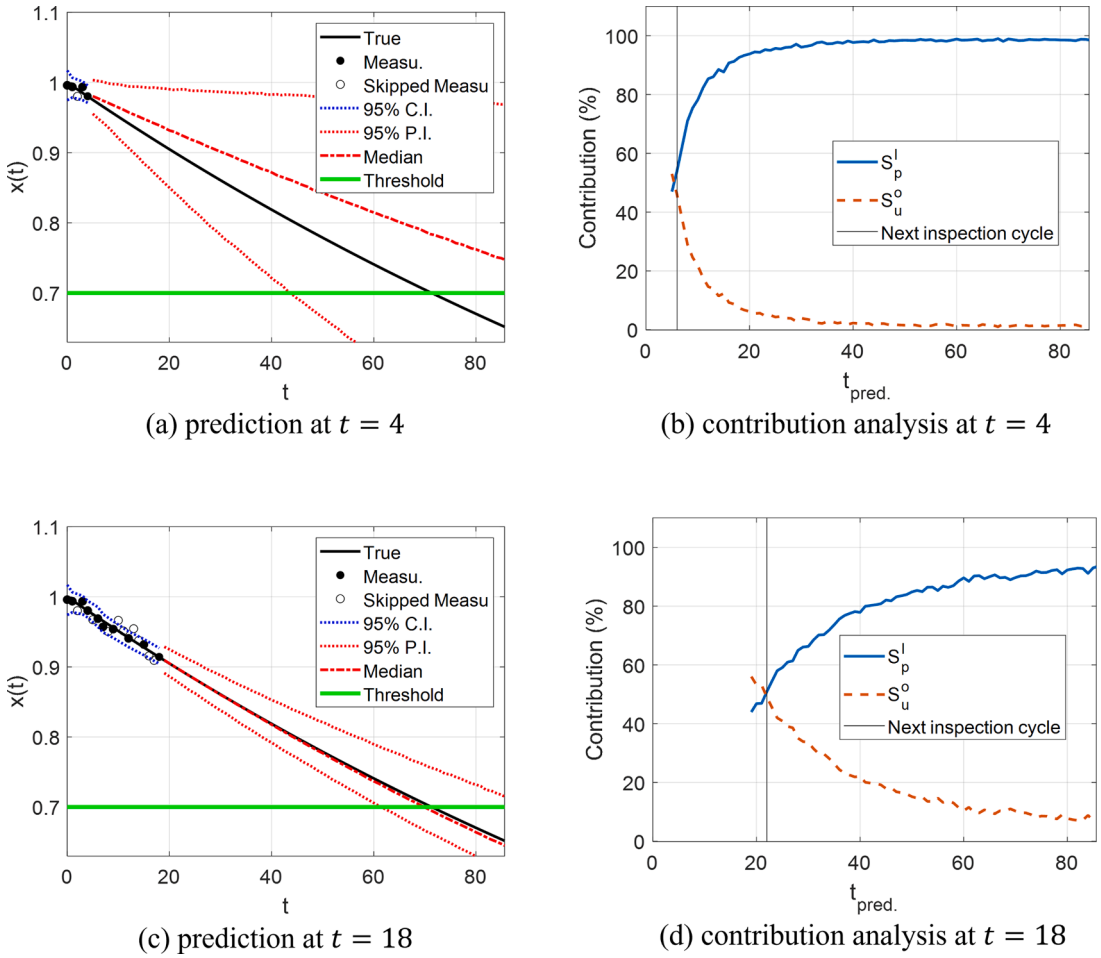


(c) prediction at $t = 17$



(d) contribution analysis at $t = 17$

Fig. 6. Prognostics and contribution analysis using general particle filter.



(a) prediction at $t = 4$ (b) contribution analysis at $t = 4$

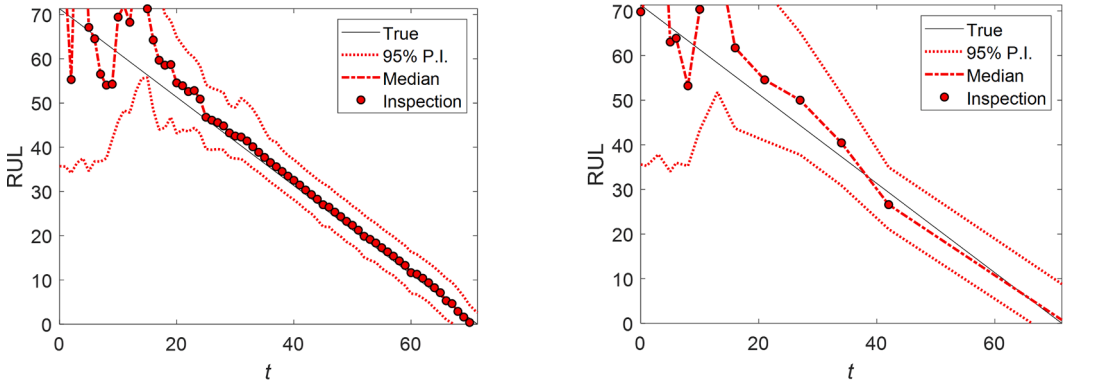
(c) prediction at $t = 18$ (d) contribution analysis at $t = 18$

Fig. 7. Prognostics and contribution analysis of the proposed model-based method (s=0.5).

$N(0, \sigma^2)$ is added to the true degradation data with the magnitude of $\sigma_{True} = 0.01$ as shown in Fig. 5. In the simulation, the nominal condition is assumed to be 1 and the threshold for failure is set as 0.7. It is assumed that data is measured at every cycle.

4.1.1. *Model-based inspection scheduling*

The objective of the model-based method is to identify model parameters, a and b, as well as the noise standard deviation σ using measurement data. The initial distributions of three parameters are assumed as $x_0 \sim U[0.9,\ 1.2]$, $b_0 \sim U[-0.01,\ 0]$, and $\sigma_0 \sim U[0.005,\ 0.015]$. It should be noted that the model parameter a corresponds to the initial health condition, x_0. Then, 10,000 samples of the three parameters are generated as particles in PF Fig. 6. shows the prediction results at $t = 4$ and $t = 17$ when the inspection is conducted at every cycle; i.e., the contribution analysis is not applied. As shown in Fig. 6(a), the confidence interval (C.I) of x becomes narrow as time passes, whereas the predictive interval (P.I) increases as t increases. This is because even if the difference between samples of model parameters is small, their effect on the prediction of $x(t)$ becomes large as time increases Fig. 6.(b) illustrates the contribution analysis at $t = 4$. Based on the P.I. and the sample distributions of b and σ, the contributions of the two uncertainties can be computed for the future. The result indicates that the



(a) RUL prediction without proposed method (b) RUL prediction with the proposed method

Fig. 8. RUL prediction of model-based method.

Table 2
The number of required inspections for different levels of threshold (model-based).

Threshold (s)	0	0.1	0.2	0.3	0.4	0.5
Num. of inspection	70	49	32	25	20	13

epistemic uncertainty will not become dominant in the near future, such as $t = 5$ or 6, whereas after t=20, the epistemic uncertainty will be the main contributor to the prediction uncertainty. Therefore, it makes sense to have a measurement or perform the inspection at the time point to reduce the uncertainty by the chance of 90 % Fig. 6.(c) and (d) show the result at $t = 17$. The figures indicate that 95% C.I. becomes narrow, and the contribution S_p^I is not dominant until $t = 30$. It means even if more data is gathered, the overall uncertainty will not be reduced dramatically before $t = 30$.

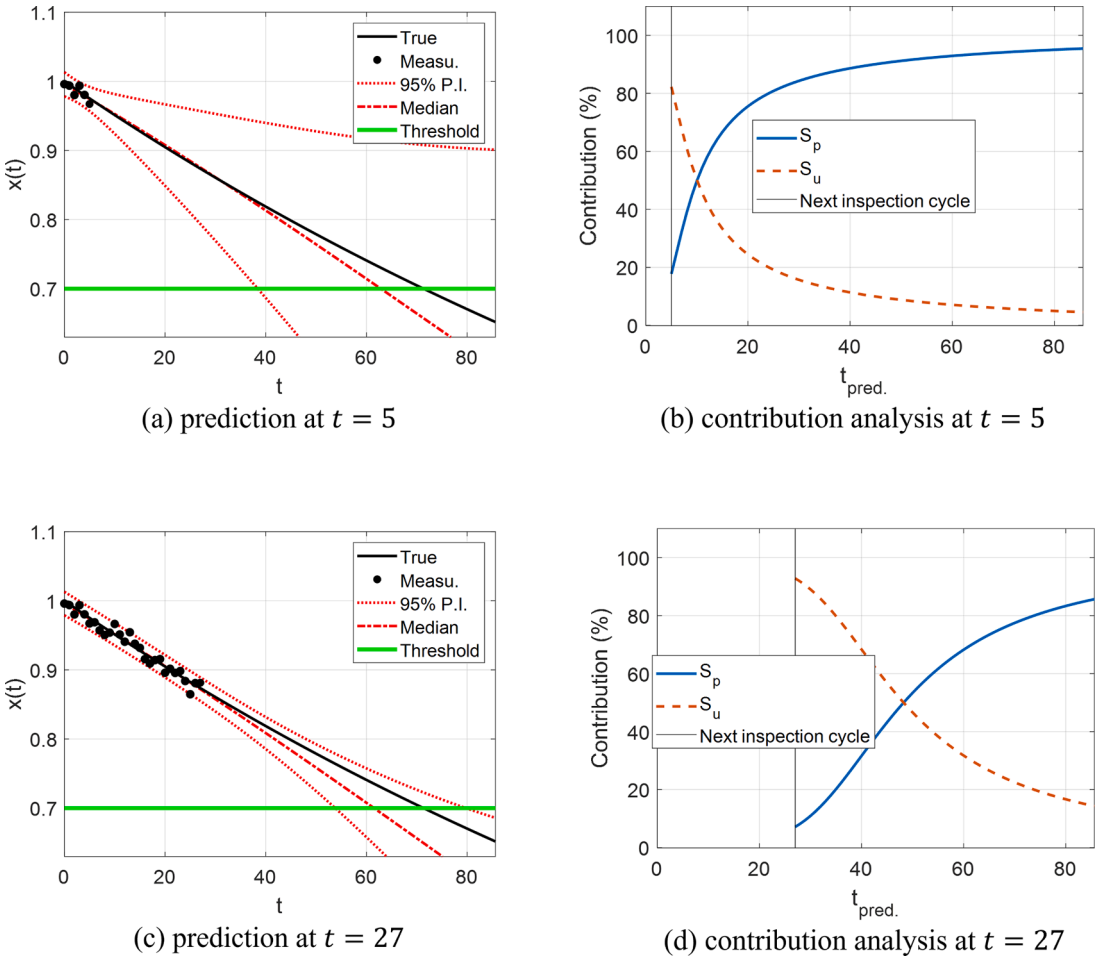
Once the maintenance engineer decides to maintain the effect of epistemic uncertainty less than 50%, the inspection should be made at time t when $S_P^I(t) \geq 0.5$ Fig. 7. shows prediction and contribution analysis after applying the proposed model-based method Fig. 7.(a) and (b) show the result at $t = 4$ when the threshold, $s = 0.5$ Fig. 7.(a) shows that there was an inspection skipped in the past since the epistemic uncertainty was predicted to be smaller than 0.5. Also, the next suggested inspection cycle is set as 6 and depicted as a black solid vertical line in Fig. 7(b). The results at $t = 17$ are shown in Fig. 7(c) and (d). As mentioned above, the inspection at $t = 5$ was skipped and other skipped measurements are illustrated as a blank circle in Fig. 7(c). Although fewer inspections are conducted than in Fig. 6, the 95% P.I. still covers the true degradation trend and shows good performance.

Finally, the RUL prediction performance with/without the proposed method in model-based prognostics is compared in Fig. 8. Red circles indicate the time when the inspection is performed. The number of inspections is dramatically reduced compared to frequent inspection (inspection at every cycle), whereas the prediction uncertainty bound shows a similar range between the two results Table 2. summarizes the relationship between the number of measurements and the threshold. The table shows that as a conservative threshold is used, more data acquisition is required.

4.1.2. Data-driven inspection scheduling

As described in Section 3, GPR is employed as a data-driven method to predict future behavior and analyze the contributions of the aleatory and epistemic uncertainty to the overall prediction uncertainty. To emphasize the pure data-driven method, a constant function with zero is used as a global function and t is used as an input for the GPR Fig. 9. shows the result of prediction and contribution analysis of the exponential degradation function using GPR. As shown in the figure, as more data are available, the contribution of epistemic uncertainty, S_P, is smaller than S_U in the near future. In the same manner with the model-based approach, it means to gathering data in the near future does not highly contribute to uncertainty reduction.

Fig. 10 shows the result of the data-driven method after setting the threshold as 0.5, which means that the contribution of epistemic uncertainty is maintained at less than 50%. As shown in Fig. 10(b) and (d), the proposed method can predict the inspection time when the epistemic uncertainty contributes more than 50%. In the same manner with the



(a) prediction at $t = 5$ (b) contribution analysis at $t = 5$

(c) prediction at $t = 27$ (d) contribution analysis at $t = 27$

Fig. 9. Prognostics and contribution analysis using Gaussian process regression.

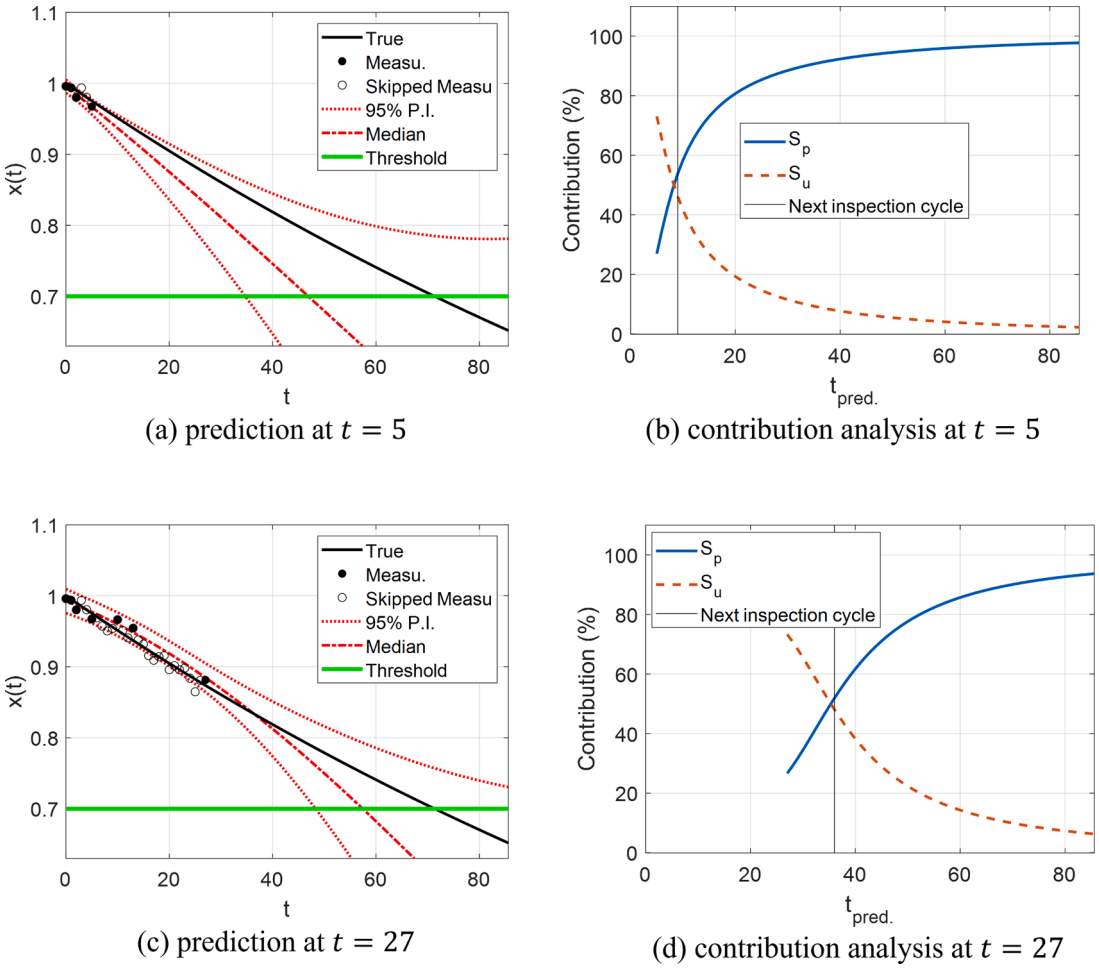


Fig. 10. Prognostics and contribution analysis of the proposed data-based method (s=0.5).

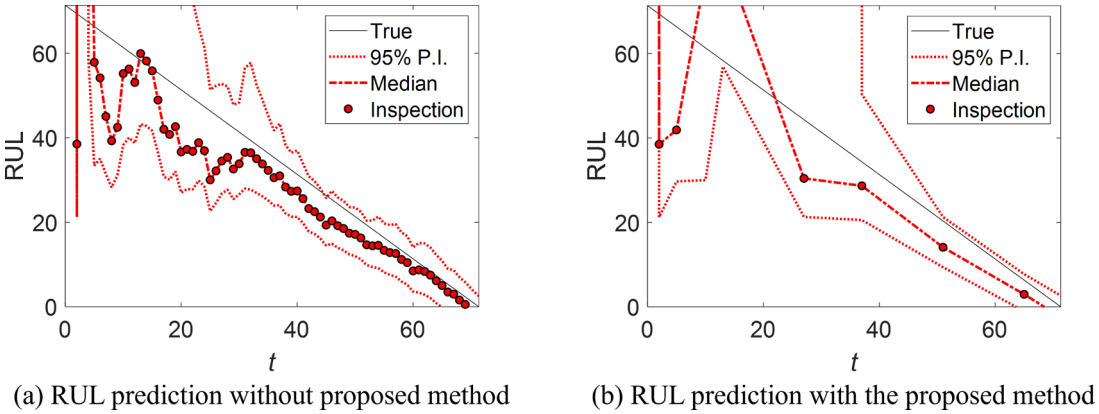


Fig. 11. RUL prediction of the data-driven method.

Table 3
The number of required inspections for different levels of threshold (data-driven).

Threshold (s)	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
Num. of inspection	70	45	22	17	13	10

model-based method, the required number of inspections is reduced compared to Fig. 9. It should be noted that the prediction performance is poor compared to the model-based approach in Fig. 8. This is a natural drawback of a data-driven algorithm that does not have a physical degradation model. Nevertheless, the prediction uncertainty still covers the true future trend.

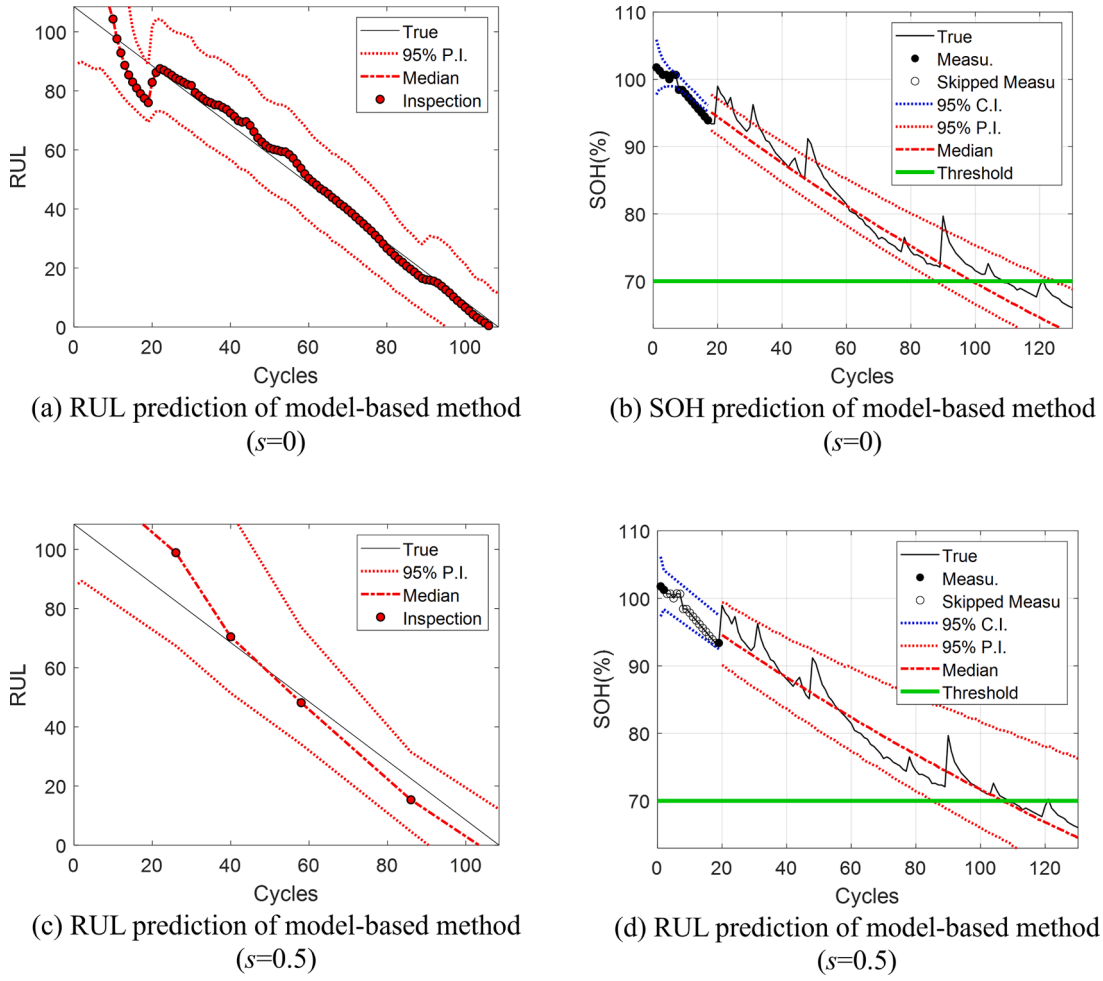


Fig. 12. Application to NASA Battey #0006.

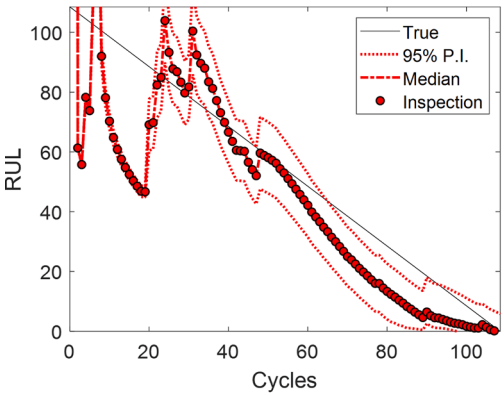
Finally, the RUL prediction results are described in Fig. 11. The result shows that the same uncertainty level can be achieved with fewer measurements when the proposed method is employed. Also, Table 3 shows that as a more conservative threshold is used, more data is required to satisfy the uncertainty requirement. However, there is an important difference between model-based and data-driven methods. In the model-based approach, it was important to obtain samples at an early stage since the samples for model parameters are rapidly converged to the most likely samples. On the contrary, the data-driven method demonstrates that even if the number of inspections can be reduced, it is vital to measure the data at a specific interval to maintain the desired epistemic uncertainty. This is expected because the main source of uncertainty in the data-driven method is due to long-term extrapolation.

4.2. NASA battery data

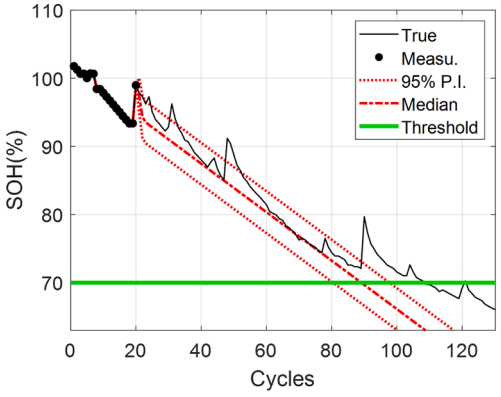
NASA Ames Prognostics Center of Excellence (PCoE) published the lithium-ion battery degradation data sets. The battery was charged with a constant current (CC) mode of 1.5A until the voltage reached 4.2 V and then constant voltage (CV) mode is maintained until the current dropped to 20 mA. Discharge was carried out with CC of 2A until the voltage fell to the pre-defined setting. Battery #6 used in this paper was discharged until the voltage dropped to 2.5V. The degradation test is stopped when the capacity is reduced to 70% of the nominal condition. State of Health (SOH) is a widely used health indicator in the field of battery prognostics. The SOH is defined as [55,58]:

$$SOH = \frac{C_i}{C_0} \times 100(\%) \tag{19}$$

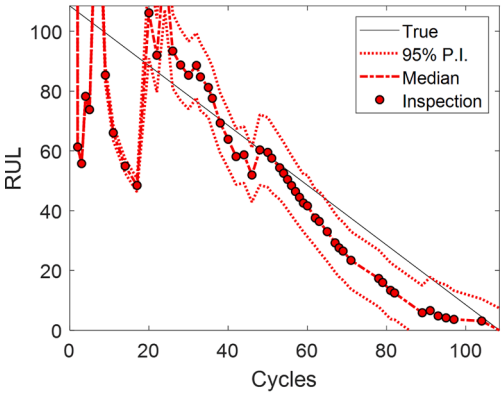
where the capacity at i^{th} cycle and initial state are denoted as C_i and C_0, respectively. Since the nominal capacity is defined as 2Ahr, the capacity is normalized with 2 Fig. 12. shows the results before and after applying the proposed method. For the model-based approach, the same exponential degradation in Section 4.1 is employed. However, there is a subtle modification of the data-driven method. The SOH includes several spike-shaped behaviors which are called capacity regeneration and they make the RUL prediction inaccurate [59,60]. Although there have been several studies to address this problem, it is beyond the scope of this paper. To simply deal with this problem, a linear global function is employed for the GPR in the data-driven method, which is why Fig. 12(f) and (h) show a linear trend in the extrapolation region. Other than that, all of the processes are the same as those in Section 4.1. As a result, the number of required inspections is reduced dramatically from 108 to 5 for the model-based and 54 for the data-driven method, whereas the uncertainty bounds in the RUL plot are similar to the case when the data was measured every cycle.



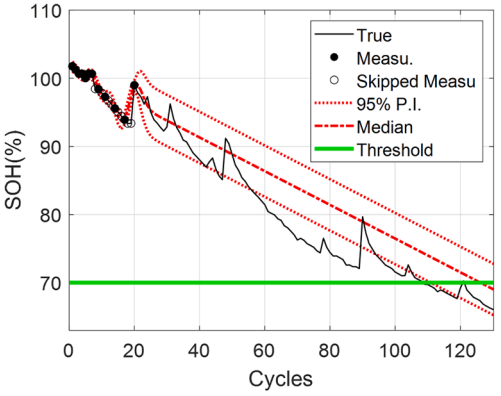
(e) RUL prediction of data-driven method (s=0)



(f) SOH prediction of data-driven method (s=0)



(g) RUL prediction of data-driven method (s=0.5)



(h) SOH prediction of data-driven method (s=0.5)

Fig. 12. (*continued*).

5. Conclusions

The main benefit of the PHM technique is the reduction of the system operating and maintenance costs. Many sources of economic cost, including data measurement, inspection, and data storage, appear in the era of big data. Therefore, it is crucial to evaluate the value of the data and identify how much the prediction performance can be improved by obtaining the data. This paper defines prediction performance as an uncertainty for the future. Global sensitivity analysis is utilized to quantify the contribution of noise in data and uncertainty in model parameters on the prediction uncertainty. Based on the contributions of aleatory and epistemic uncertainty to the overall prediction uncertainty, the proposed method suggests a future inspection schedule to maintain the pre-defined level of the desired epistemic uncertainty. For practical applications, the proposed methods are grouped into model-based and data-driven methods. Both methods showed that the number of data measurements can be significantly reduced while maintaining the same level of prediction uncertainty.

Even if both methods showed a good result, there are still points that can be explored in future work. For the model-based method, measurement scheduling is directly affected by the accuracy of the degradation model. It is still a challenge to obtain an effective degradation model in practice. In the case of the data-driven method, the extrapolation of GPR in the absence of global function becomes useless in long-term prediction. In addition, model-based and data-driven approaches commonly contain model uncertainty which is not yet explored in this research. Future work will include the limitations of the two approaches mentioned above and improve the proposed method to consider the various types of uncertainty including model uncertainty. In addition, the proposed method will be developed to be used for the multiple component systems, which contain multiple sensor data corresponding to various types of components.

CRediT authorship contribution statement

Seokgoo Kim: Conceptualization, Methodology, Software, Writing – original draft. **Joo-Ho Choi:** Supervision, Writing – review & editing. **Nam Ho Kim:** Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1A4A4079904).

References

[1] Singh J, Azamfar M, Li F, Lee J. A systematic review of machine learning algorithms for prognostics and health management of rolling element bearings: fundamentals, concepts and applications. Meas Sci Technol Jan. 01, 2021;32(1). https://doi.org/10.1088/1361-6501/ab8df9.

[2] Rezaeianjouybari B, Shang Y. Deep learning for prognostics and health management: State of the art, challenges, and opportunities. Meas J Int Meas Confed Oct. 2020;163. https://doi.org/10.1016/j.measurement.2020.107929.

[3] Zonta T, da Costa CA, da Rosa Righi R, de Lima MJ, da Trindade ES, Li GP. Predictive maintenance in the Industry 4.0: A systematic literature review. Comput Ind Eng Dec. 2020;150. https://doi.org/10.1016/j.cie.2020.106889.

[4] Si X-S, Chen M-Y, Wang W, Hu C-H, Zhou D-H. Specifying measurement errors for required lifetime estimation performance. Eur J Oper Res 2013;231(3):631–44.

[5] Kim S, Choi J-H, Kim NH. Challenges and opportunities of system-level prognostics. Sensors 2021;21(22):7655. https://doi.org/10.3390/s21227655.

[6] Van Horenbeek A, Pintelon L. A dynamic predictive maintenance policy for complex multi-component systems. Reliab Eng Syst Saf 2013;120:39–50.

[7] Wu S, Gebraeel N, Lawley MA, Yih Y. A neural network integrated decision support system for condition-based optimal predictive maintenance policy. IEEE Trans Syst Man, Cybern A Syst Humans 2007;37(2):226–36.

[8] Curcurù G, Galante G, Lombardo A. A predictive maintenance policy with imperfect monitoring. Reliab Eng Syst Saf 2010;95(9):989–97.

[9] Li H, Zhu W, Dieulle L, Deloux E. Condition-based maintenance strategies for stochastically dependent systems using Nested Lévy copulas. Reliab Eng Syst Saf 2022;217:108038.

[10] Papakonstantinou KG, Shinozuka M. Planning structural inspection and maintenance policies via dynamic programming and Markov processes. Part II: POMDP implementation. Reliab Eng Syst Saf 2014;130:214–24.

[11] Papakonstantinou KG, Shinozuka M. Planning structural inspection and maintenance policies via dynamic programming and Markov processes. Part I: Theory. Reliab Eng Syst Saf 2014;130:202–13. https://doi.org/10.1016/j.ress.2014.04.005.

[12] Yuan X-X, Higo E, Pandey MD. Estimation of the value of an inspection and maintenance program: a Bayesian gamma process model. Reliab Eng Syst Saf 2021;216:107912.

[13] Cheng Y, Wei Y, Liao H. Optimal sampling-based sequential inspection and maintenance plans for a heterogeneous product with competing failure modes. Reliab Eng Syst Saf 2022;218:108181.

[14] Kampitsis D, Panagiotidou S. A Bayesian condition-based maintenance and monitoring policy with variable sampling intervals. Reliab Eng Syst Saf 2022;218:108159.

[15] Nguyen KTP, Fouladirad M, Grall A. New methodology for improving the inspection policies for degradation model selection according to prognostic measures. IEEE Trans Reliab 2018;67(3):1269–80.

[16] Papakostas N, Papachatzakis P, Xanthakis V, Mourtzis D, Chryssolouris G. An approach to operational aircraft maintenance planning. Decis Support Syst Mar. 2010;48(4):604–12. https://doi.org/10.1016/j.dss.2009.11.010.

[17] Liu X, Sun Q, Ye Z-S, Yildirim M. Optimal multi-type inspection policy for systems with imperfect online monitoring. Reliab Eng Syst Saf 2021;207:107335.

[18] Levitin G, Finkelstein M, Xiang Y. Optimal inspections and mission abort policies for multistate systems. Reliab Eng Syst Saf 2021;214:107700.

[19] Lin C, Song J, Pozzi M. Optimal inspection of binary systems via Value of Information analysis. Reliab Eng Syst Saf 2022;217:107944.

[20] Yiwei W, Christian G, Binaud N, Christian BES, Haftka RT. A cost driven predictive maintenance policy for structural airframe maintenance. Chinese J Aeronaut 2017;30(3):1242–57.

[21] Wang Y, Gogu C, Binaud N, Bes C, Haftka RT, Kim N-H. Predictive airframe maintenance strategies using model-based prognostics. Proc Inst Mech Eng Part O J Risk Reliab 2018;232(6):690–709.

[22] Zhang H, Marsh DWR. Managing infrastructure asset: Bayesian networks for inspection and maintenance decisions reasoning and planning. Reliab Eng Syst Saf 2021;207:107328.

[23] Andriotis CP, Papakonstantinou KG. Deep reinforcement learning driven inspection and maintenance planning under incomplete information and constraints. Reliab Eng Syst Saf 2021;212:107551.

[24] Zhang L, Chen X, Khatab A, An Y. Optimizing imperfect preventive maintenance in multi-component repairable systems under s-dependent competing risks. Reliab Eng Syst Saf 2021:108177.

[25] Nguyen K-A, Do P, Grall A. Condition-based maintenance for multi-component systems using importance measure and predictive information. Int J Syst Sci Oper Logist 2014;1(4):228–45.

[26] Consilvio A, Di Febbraro A, Sacco N. A rolling-horizon approach for predictive maintenance planning to reduce the risk of rail service disruptions. IEEE Trans Reliab 2020.

[27] Shi Y, et al. A new preventive maintenance strategy optimization model considering lifecycle safety. Reliab Eng Syst Saf 2022:108325.

[28] Azimpoor S, Taghipour S. Joint Inspection and Product Quality Optimization for a System with Delayed Failure. Reliab Eng Syst Saf 2021:107793.

[29] Gao K, Peng R, Qu L, Wu S. Jointly optimizing lot sizing and maintenance policy for a production system with two failure modes. Reliab Eng Syst Saf 2020;202:106996.

[30] Turan HH, Atmis M, Kosanoglu F, Elsawah S, Ryan MJ. A risk-averse simulation-based approach for a joint optimization of workforce capacity, spare part stocks and scheduling priorities in maintenance planning. Reliab Eng Syst Saf 2020;204:107199.

[31] Azimpoor S, Taghipour S, Farmanesh B, Sharifi M. Joint Planning of Production and Inspection of Parallel Machines with Two-phase of Failure. Reliab Eng Syst Saf 2022;217:108097.

[32] Xiao L, Zhang X, Tang J, Zhou Y. Joint optimization of opportunistic maintenance and production scheduling considering batch production mode and varying operational conditions. Reliab Eng Syst Saf 2020;202:107047.

[33] Cheng G, Li L. Joint optimization of production, quality control and maintenance for serial-parallel multistage production systems. Reliab Eng Syst Saf 2020;204:107146.

[34] Liu G, Chen S, Jin H, Liu S. Optimum opportunistic maintenance schedule incorporating delay time theory with imperfect maintenance. Reliab Eng Syst Saf 2021;213:107668.

[35] Yang H, Li W, Wang B. Joint optimization of preventive maintenance and production scheduling for multi-state production systems based on reinforcement learning. Reliab Eng Syst Saf 2021;214:107713.

[36] An Y, Chen X, Hu J, Zhang L, Li Y, Jiang J. Joint optimization of preventive maintenance and production rescheduling with new machine insertion and processing speed selection. Reliab Eng Syst Saf 2022:108269.

[37] Levitin G, Xing L, Dai Y. Joint optimal mission aborting and replacement and maintenance scheduling in dual-unit standby systems. Reliab Eng Syst Saf 2021;216:107921.

[38] Lee J, Mitici M. Multi-objective design of aircraft maintenance using Gaussian process learning and adaptive sampling. Reliab Eng Syst Saf 2022;218:108123.

[39] Cavalcante CAV, Lopes RS, Scarf PA. Inspection and replacement policy with a fixed periodic schedule. Reliab Eng Syst Saf 2021;208:107402.

[40] Mancuso A, Compare M, Salo A, Zio E. Optimal Prognostics and Health Management-driven inspection and maintenance strategies for industrial systems. Reliab Eng Syst Saf 2021:107536.

[41] Camci F. Maintenance scheduling of geographically distributed assets with prognostics information. Eur J Oper Res 2015;245(2):506–16. https://doi.org/10.1016/j.ejor.2015.03.023.

[42] Fauriat W, Zio E. Optimization of an aperiodic sequential inspection and condition-based maintenance policy driven by value of information. Reliab Eng Syst Saf 2020;204:107133.

[43] Wang J, Qiu Q, Wang H, Lin C. Optimal condition-based preventive maintenance policy for balanced systems. Reliab Eng Syst Saf 2021;211:107606.

[44] Shi Y, Zhu W, Xiang Y, Feng Q. Condition-based maintenance optimization for multi-component systems subject to a system reliability requirement. Reliab Eng Syst Saf 2020;202:107042.

[45] McFarland JM. Variance decomposition for statistical quantities of interest. J Aerosp Inf Syst 2015;12(1):204–18.

[46] Sankararaman S, Mahadevan S. Separating the contributions of variability and parameter uncertainty in probability distributions. Reliab Eng Syst Saf 2013;112:187–99.

[47] Bae S, Kim NH, Jang S. Reliability-based design optimization under sampling uncertainty: shifting design versus shaping uncertainty. Struct Multidiscip Optim 2018;57(5):1845–55.

[48] Sankararaman S, Goebel K. Uncertainty in prognostics: computational methods and practical challenges. In: 2014 IEEE Aerospace Conference; 2014. p. 1–9.

[49] Saha B, Quach CC, Goebel K. Optimizing battery life for electric UAVs using a Bayesian framework. In: 2012 IEEE aerospace conference; 2012. p. 1–7.

[50] Saltelli A, et al. Global sensitivity analysis: the primer. John Wiley & Sons; 2008.

[51] Saltelli A, Tarantola S, Campolongo F, Ratto M. Sensitivity analysis in practice: a guide to assessing scientific models, 1. Wiley Online Library; 2004.

[52] An D, Choi JH, Kim NH. Prognostics 101: A tutorial for particle filter-based prognostics algorithm using Matlab. Reliab Eng Syst Saf 2013. https://doi.org/10.1016/j.ress.2013.02.019.

[53] Kim S, Park HJ, Choi JH, Kwon D. A Novel Prognostics Approach Using Shifting Kernel Particle Filter of Li-Ion Batteries under State Changes. IEEE Trans Ind Electron 2021;68(4):3485–93. https://doi.org/10.1109/TIE.2020.2978688.

[54] Baraldi P, Mangili F, Zio E. A prognostics approach to nuclear component degradation modeling based on Gaussian Process Regression. Prog Nucl Energy 2015;78:141–54.

[55] Liu D, Pang J, Zhou J, Peng Y, Pecht M. Prognostics for state of health estimation of lithium-ion batteries based on combination Gaussian process functional regression. Microelectron Reliab 2013;53(6):832–9.

[56] Zhou D, et al. Prognostics for state of health of lithium-ion batteries based on Gaussian process regression. Math Probl Eng 2018;2018.

[57] Kocijan J. Modelling and control of dynamic systems using Gaussian process models. Springer; 2016.

[58] Kim S, Kim NH, Choi JH. Prediction of remaining useful life by data augmentation technique based on dynamic time warping. Mech Syst Signal Process 2020;136:106486. https://doi.org/10.1016/j.ymssp.2019.106486.

[59] Wang D, Kong J-Z, Zhao Y, Tsui K-L. Piecewise model based intelligent prognostics for state of health prediction of rechargeable batteries with capacity regeneration phenomena. Measurement 2019;147:106836.

[60] Olivares BE, Munoz MAC, Orchard ME, Silva JF. Particle-filtering-based prognosis framework for energy storage devices with a statistical characterization of state-of-health regeneration phenomena. IEEE Trans Instrum Meas 2012;62(2):364–76.