Online Updating With a Probability-Based Prediction Model Using Expectation Maximization Algorithm for Reliability Forecasting

Chang-Hua Hu, Xiao-Sheng Si, Jian-Bo Yang, and Zhi-Jie Zhou

***Abstract*—Recently, a novel prediction model based on the evidential reasoning (ER) approach is developed to forecast reliability in engineering systems. In order to determine the parameters of the ER-based prediction model, some optimization models have been proposed to train the ER-based prediction model. However, these models are implemented in an offline fashion and thus it is very expensive to train and retrain them when new information is available. This correspondence paper is concerned with developing the recursive algorithms for updating the ER-based prediction model from the probability-based point of view. Using the recursive expectation maximization algorithm, two recursive algorithms are proposed for updating the parameters of the ER-based prediction model under judgmental and numerical outputs, respectively. As such, the proposed algorithms can be used to fine tune the ER-based prediction model online once new information becomes available. We verify the proposed method via a realistic example with missile reliability data.**

***Index Terms*—Decision analysis, expectation maximization (EM), forecasting, recursive algorithms, uncertainty.**

I. Introduction

For the improvement of equipment reliability and safety, there is a crucial need for online performance prediction and maintenance scheduling. Traditional preventive maintenance (PM) and newly emerged condition-based maintenance (CBM) have prevented many troubles from happening. However, such shocking disasters as the explosion of the U.S. Space shuttle “Challenger”, the crash of a Chinese early warning aircraft, and the sinking of a Russian submarine still occurred. This may be due to the PMs and the CBMs, in some sense, performed blindly. Maintenance personnel may have no clear picture on the whole system diagnostic performance and no accurate information on internal equipment changes. As a result, predicting the health of engineering systems has been recently an active research area. Until now, several forecasting methods have been proposed in the literature, which can be classified into three categories, i.e., model-based methods, qualitative knowledge based methods, and data-driven models [18], [19].

If the mathematical models of engineering systems are exactly known, the model-based methods including Kalman predictor [37], strong tracking predictor [29], and particle predictor [3] can be used to predict the system behavior reflected by the states or the parameters. If special kind of knowledge on the system is known, the qualitative knowledge-based methods including the expert system [1] and Petri net [34] can be used to predict the future behavior of the system. If the input and output data are available, data-driven learning models can be used to set up forecasting models, including autoregressive integrated moving average model (ARIMA) method [9], grey method [40] and neural networks [32], support vector machines (SVM) [4], and fuzzy-logic-based methods [7], [15], [20]–[22]. The future values of system outputs can be forecasted and further used to predict the future behavior of the system. However, there are some inherent shortcomings existing in the aforementioned methods. The model-based methods are not applicable in the case when the models of complex systems are often difficult to obtain in the engineering practice [16]. Since the qualitative knowledge-based methods need to produce a set of rules with varying confidence and require the users to correctly select the data set, attributes, and decision variables, the combinatorial explosion and the inaccurate prediction often arise. Since data-driven methods only depend on the input/output data, they become increasingly popular in forecasting field. However, data-driven methods can only use either numerical or subjective information [26], not both.

As far as we know, human beings generally hold ultimate responsibilities in most decision situations and their preferences play an irreplaceable role in making final decisions. However, human judgmental information may be incomplete and inaccurate, Hence, it is very important to model and analyze decision problems by combining numerical data and human judgmental information [23]–[25], [31]. In order to handle hybrid information with uncertainty in forecasting, Hu *et al.* proposed a prediction model using the evidential reasoning (ER) approach [33], [35], [36], which was referred as the ER-based prediction model [10]. Compared with the traditional prediction methods such as the ARIMA model, neural network approaches, and SVM model, the ER-based prediction method provides a more informative and realistic scheme for knowledge representation and can capture vagueness, incompleteness, and nonlinear causal relationships. The results in [10] indicate that the ER-based prediction model can generate some satisfactory results. However, the optimization methods in [10] to tune model parameters are implanted in an offline way, the computation can become expensive when a very large set of data is involved.

To overcome the computation burden, it is necessary to develop an online algorithm to update the parameters of the ER-based prediction model in such a way that the parameters can be updated recursively once new information becomes available. Once the prediction model is constructed and its parameters are estimated, the ER-based prediction model can be used to perform forward prediction from the given inputs.

Inspired by the ER-based prediction model and the learning algorithms developed in [10], this correspondence paper proposes two recursive algorithms to update the parameters of the ER-based prediction model with numerical and judgmental outputs, respectively. In order to make use of the recursive expectation maximization (EM) algorithm [5], [6], [27], we assume that 1) the outputs of an ER-based prediction model are independent if its inputs are independent and 2) the actual outputs obey the normal distribution. From the probability-based point of view, the updating algorithms on the basis of the recursive EM algorithm are developed. We demonstrate the usefulness of the proposed method via a realistic example with missile reliability data.

This correspondence paper is organized as follows. In Section II, the ER-based prediction model is briefly reviewed. Section III proposes two recursive algorithms, namely, one for online updating the forecasting model under judgmental distribution outputs and the other under numerical outputs. A practical study is presented in Section IV. The correspondence paper is concluded in Section V.

Manuscript received March 21, 2009; revised May 5, 2010; accepted August 14, 2010. Date of publication May 19, 2011; date of current version October 19, 2011. This work was supported by the National Natural Science Foundation of China under Grant 60736026, the National Science Fund for Distinguished Young Scholars of China under Grant 61025014, and the Program for New Century Excellent Talents in University of China under Grant 07144. This paper was recommended by Associated Editor E. P. Blasch.

C.-H. Hu and Z.-J. Zhou are with Xi’an Institute of Hi-tech, Xi’an710025, China (e-mail: hch6603@263.net; zhouzj04@mails.tsinghua.edu.cn).

X.-S. Si is with Xi’an Institute of Hi-tech, Xi’an 710025, China, and also with Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: sxs09@mails.tsinghua.edu.cn).

J.-B. Yang is with the Manchester Business School, University of Manchester, M15 6PB Manchester, U.K. (e-mail: jian-bo.yang@mbs.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCA.2011.2147304

II. ER-Based Prediction Model

In order to handle hybrid information with uncertainty in the forecasting, a prediction model using the ER approach is presented, termed as the ER-based prediction model [10]. The ER-based prediction method provides a more informative and realistic scheme for knowledge representation inherited from the ER algorithm [36]. In the following, a brief yet complete description of this model is summarized.

A. Prediction Model Structure and Representation

For prediction problems, the inputs used in a prediction model are the past input vector and the lagged observations of the current time, while the outputs are the future values. Each set of input patterns is composed of any moving fixed-length window. The general prediction model can be represented as

$$\hat{y}(t+k-1) = f(x_{t-1}, x_{t-2}, \ldots, x_{t-p}) \tag{1}$$

where $\hat{y}(t+k-1)$ representing the predicted future value at time $t+k-1$, $(x_{t-1}, x_{t-2}, \ldots, x_{t-p})$ is a vector of the lagged input variables, and p represents the embedding dimensions. In this correspondence paper, we apply Cao's method [2] to determine the embedding dimensions p. In the prediction practice, it is reasonable to assume that the current output value $y(t)$ is related to the most recent input x_{t-1} or even extended into the past values $x_{t-2}, \ldots, x_{t-p}$. This is because the next output value is dependent on the current output value to certain extent which is in turn related to the current input. For simplicity, let $\mathbf{X}(t) = (x_{t-1}, x_{t-2}, \ldots, x_{t-p})$ denote the input vector of the prediction model in the following.

Similar to [10], we only consider one-step-ahead prediction. Then, the general prediction model can be given by

$$\hat{y}(t) = f(x_{t-1}, x_{t-2}, \ldots, x_{t-p}). \tag{2}$$

In order to apply the ER approach, we suppose $x_{t-1}, x_{t-2}, \ldots, x_{t-p}$ are p basic attributes associated with system prediction output $\hat{y}(t)$, and then the ER approach attempts to identify the appropriate internal representation between $\mathbf{X}(t)$ and $\hat{y}(t)$. The key for solving this prediction problem is how to approximate the function $f(\cdot)$.

B. Prediction Model Under the ER Framework

In this correspondence paper, the ER approach is applied to deal with the above predication problem. First, we assume that there are p basic attributes $x_{t-1}, x_{t-2}, \ldots, x_{t-p}$ associated with system output y. Define a set of p basic attributes as evidence sources. Then, suppose the weights of the attributes are given by $\mathbf{w} = \{w_1, \cdots w_i, \cdots, w_p\}$ where w_i is the relative weight of the ith basic attribute x_{t-i}, and the weights of the attributes are normalized to satisfy the following constraints:

$$0 \leq w_i \leq 1 \text{ and } \sum_{i=1}^{p} w_i = 1. \tag{3}$$

Define N distinctive evaluation grades as

$$F = \{F_1, \ldots F_n, \ldots, F_N\} \tag{4}$$

where F_n is the nth evaluation grade for the predicted output. It is worth noting that F provides a complete set of standards for assessing attributes referred as the assessment framework.

Using the rule equivalence transformation techniques [35], the p basic attributes $x_{t-1}, x_{t-2}, \ldots, x_{t-p}$ can each be represented by the following structure:

$$S(x_{t-i}) = \{(F_n, \beta_{n,i}(x_{t-i})), n = 1, \ldots, N\}, \quad i = 1, \ldots, p \tag{5}$$

where $\beta_{n,i}$ denotes the matched degree to the grade F_n. The above distributed assessment implies that the attribute x_{t-i} is assessed to the grade F_n with the degree $\beta_{n,i}$. $\beta_{n,i}$ could be generated using various ways, depending on the nature of the attribute and data available such as a quantitative attribute using numerical values or a qualitative attribute using linguistic values [10], [35], [36].

From (5), the ER approach can be directly applied to combine all attributes and generate final conclusions. Using the ER analytical algorithm developed by Wang *et al.* [30], the final prediction result $\hat{\mathbf{y}}(t)$ generated by aggregating all attributes can be obtained by

$$\hat{\mathbf{y}}(t) = \left\{\left(F_n, \hat{\beta}_n(t)\right), \qquad n = 1, \ldots, N\right\} \tag{6}$$

where $\hat{\beta}_n(t)$, $n = 1, \ldots, N$ can be decided by the ER analytical algorithm

$$\hat{\beta}_n(t) = \frac{\prod_{i=1}^{p}(w_i\beta_{n,i}(x_{t-i})+1-w_i) - \prod_{i=1}^{p}(1-w_i)}{D(t)} \tag{7}$$

$$D(t) = \sum_{n=1}^{N}\prod_{i=1}^{p}(w_i\beta_{n,i}(x_{t-i})+1-w_i) - N\prod_{i=1}^{p}(1-w_i). \tag{8}$$

The aggregated assessment $\hat{\mathbf{y}}(t)$ represents the overall assessment of the system. It can tell us which assessment grade is assessed to and what matched degrees are assigned to the defined grade F_n.

C. Utility-Based Construction of the Numerical Outputs

In some cases, since the output value $\hat{\mathbf{y}}(t)$ is a numerical value, it is desirable to generate numerical values equivalent to the distributed assessments (6). The concept of utility can be used to define such values [10], [36], [38]. Suppose u_i, $i = 1, \ldots, N$ are the utilities of the assessment grades F_i, $i = 1, \ldots, N$, respectively. If F_j is preferred to F_i for a prediction problem, then we have utility relation $u_i < u_j$, $i < j$. u_i may be estimated using prior expert knowledge or by constructing optimization models. In this correspondence paper, we use the latter approach.

In order to generate numerical output values equivalent to the distributed assessments (6), the maximum, minimum, and average utilities are introduced [36]. From (6), the prediction outcome $\hat{\mathbf{y}}(t)$ is calculated by [10]

$$\hat{\mathbf{y}}(t) = \sum_{j=1}^{N} \hat{\beta}_j(t) u_j. \tag{9}$$

The logic behind the approach is that, if we set the parameters of the ER-based prediction model accurately, such as w_i, $i = 1, \ldots, p$ or u_j, and $j = 1, \ldots, N$, the predicted output $\hat{\mathbf{y}}(t)$ can be matched to the actual output $\mathbf{y}(t)$ closely.

III. Recursive Algorithms for Updating the ER-Based Prediction Model

Note that the optimal learning process for the ER-based model parameters in [10] is an offline type in essence, and it is time-consuming to train or retrain them once new input-output information is available. As such, there is a clear need to develop an online parameters updating algorithm. The so-called EM algorithm developed by Dempster *et al.* [6] is normally used to achieve this aim. Zhou *et al.* [39] have applied the EM algorithm to update belief-rule-base system developed by Yang *et al.* [38]. Along the line of the works of Zhou *et al.* [39] and Hu *et al.* [10], the recursive algorithms for updating the ER-based model are developed accordingly from the probability-based point of view.

In the recursive algorithms, the observations on system inputs and outputs are required. Assume that a set of observation pairs $(\mathbf{X}(n), \mathbf{y}(n))$ is available, where $\mathbf{X}(n)$ is a given input vector; $\mathbf{y}(n)$ is the observed output vector, either measured using instruments or assessed by experts; and $\hat{\mathbf{y}}(n)$ is the simulated output that is generated by the ER-based prediction model. Therefore, $\mathbf{X}(n)$ and $\mathbf{y}(n)$ can be either numerical or judgmental. Accordingly, we will study the numerical and judgmental cases, respectively. In order to apply the recursive EM algorithm to develop the recursive updating algorithms for the ER-based prediction model, we also assume that the actual outputs $\mathbf{y}(1), \ldots, \mathbf{y}(n)$ can also be independent when the inputs of the ER-based prediction model are independent,. In the following, based on this assumption, two recursive algorithms will be developed to update the ER-based prediction model.

A. Recursive Parameter Estimation Algorithm Based on Judgmental Output

In this case, $\mathbf{y}(n)$ is judgmental output and can be represented as follows:

$$\mathbf{y}(n) = \{(F_j, \beta_j(n)), \qquad j = 1, \ldots, N\} \tag{10}$$

where $\beta_j(n)$ denotes the degree belonging to F_j, assessed at time instant n. This is indeed the default output format of the ER-based prediction model and provides a complete view about the output variations. Let $\mathbf{B}(n) = [\beta_1(n), \ldots, \beta_N(n)]^T$ denote the true output corresponding to the input $\mathbf{X}(n)$ at time instant n. Here, $\beta_j (j = 1, \ldots, N)$ can be considered as a random variable from the probability-based point of view. Then, we assume that the conditional probability density function (PDF) of $\mathbf{B}(n)$ is $f(\mathbf{B}(n)|\mathbf{X}(n), \mathbf{Q})$ given $\mathbf{X}(n)$ and $\mathbf{Q}$, where $\mathbf{Q}$ is the unknown parameters vector. According to the independent assumption, we can obtain directly

$$f(\mathbf{B}(1), \ldots, \mathbf{B}(n)|\mathbf{X}(1), \ldots, \mathbf{X}(n), \mathbf{Q}) = \prod_{\tau=1}^{n} f(\mathbf{B}(\tau)|\mathbf{X}(\tau), \mathbf{Q}). \tag{11}$$

From (11), the expectation of the log-likelihood function at time instant n can be defined as

$$L_{n+1}(\mathbf{Q}) = E\left\{ \sum_{\tau=1}^{n} \log f(\mathbf{B}(\tau)|\mathbf{X}(\tau), \mathbf{Q}) \,\middle|\, \mathbf{X}(1), \ldots, \mathbf{X}(n), \mathbf{Q}(n) \right\} \tag{12}$$

where $E(\cdot|\cdot)$ denotes the conditional expectation operator.

Now consider the recursive formulation and then (12) can be written as

$$L_{n+1}(\mathbf{Q}) = L_n(\mathbf{Q}) + E\{\log f(\mathbf{B}(n)|\mathbf{X}(n), \mathbf{Q})\,|\mathbf{X}(n), \mathbf{Q}(n)\}. \tag{13}$$

To obtain a proper approximation of $L_{n+1}(\mathbf{Q})$, we consider the Taylor expansion of the first term on the right-hand side of (13) and obtain approximately

$$\begin{aligned} L_n(\mathbf{Q}) \approx\ & L_n(\mathbf{Q}(n)) + [\nabla_{\mathbf{Q}} L_n(\mathbf{Q}(n))](\mathbf{Q} - \mathbf{Q}(n)) \\ & + \frac{1}{2}(\mathbf{Q} - \mathbf{Q}(n))^T \left[\nabla_{\mathbf{Q}} \nabla_{\mathbf{Q}}^T L_n(\mathbf{Q}(n))\right](\mathbf{Q} - \mathbf{Q}(n)) \end{aligned} \tag{14}$$

where $\nabla_{\mathbf{Q}}$ is a column gradient operator with respect to $\mathbf{Q}$.

By the definition of $L_n(\mathbf{Q})$, $\nabla_{\mathbf{Q}} \nabla_{\mathbf{Q}}^T L_n(\mathbf{Q}(n))$ is approximately given by [5], [27], [39]

$$\nabla_{\mathbf{Q}} \nabla_{\mathbf{Q}}^T L_n(\mathbf{Q}(n)) \approx -(n-1)\Xi_1(\mathbf{Q}(n)) \tag{15}$$

where $\Xi_1(\mathbf{Q}(n))$ can be calculated as

$$\Xi_1(\mathbf{Q}(n)) = E\left\{-\nabla_{\mathbf{Q}} \nabla_{\mathbf{Q}}^T \log f(\mathbf{B}(n)|\mathbf{X}(n), \mathbf{Q})|\mathbf{X}(n), \mathbf{Q}(n)\right\}. \tag{16}$$

Because $\mathbf{Q} = \mathbf{Q}(n)$ is the maximum point of $L_n(\mathbf{Q})$ in (14), there is

$$\nabla_{\mathbf{Q}} L_n(\mathbf{Q}(n)) = 0. \tag{17}$$

Substituting (16) and (17) into (14), we can obtain

$$\begin{aligned} L_n(\mathbf{Q}) \approx\ & L_n(\mathbf{Q}(n)) - \frac{1}{2}(\mathbf{Q} - \mathbf{Q}(n))^T \\ & \times [(n-1)\Xi_1(\mathbf{Q}(n))](\mathbf{Q} - \mathbf{Q}(n)). \end{aligned} \tag{18}$$

Define

$$\Gamma_1(\mathbf{Q}(n)) = \nabla_{\mathbf{Q}} \log f(\mathbf{B}(n)|\mathbf{X}(n), \mathbf{Q}(n)) \tag{19}$$

where $\nabla_{\mathbf{Q}} \log f(\mathbf{B}(n)|\mathbf{X}(n), \mathbf{Q}(n))$ represents the gradient vector at $\mathbf{Q}(n)$.

Taking Taylor expansion and then formulating conditional expectation of $\log f(\mathbf{B}(n)|\mathbf{X}(n), \mathbf{Q})$ at $\mathbf{Q}(n)$ and $\mathbf{X}(n)$, respectively, we obtain

$$\begin{aligned} L_{n+1}(\mathbf{Q}) =\ & L_n(\mathbf{Q}(n)) \\ & + E\{\log f(\mathbf{B}(n)|\mathbf{X}(n), \mathbf{Q}(n))\,|\mathbf{X}(n), \mathbf{Q}(n)\} \\ & + \Gamma_1(\mathbf{Q}(n))(\mathbf{Q} - \mathbf{Q}(n)) - \frac{n}{2}(\mathbf{Q} - \mathbf{Q}(n))^T \\ & \times [\Xi_1(\mathbf{Q}(n))](\mathbf{Q} - \mathbf{Q}(n)) \end{aligned} \tag{20}$$

with the aid of (13) and (18) after some manipulations.

Since $\mathbf{Q} = \mathbf{Q}(n+1)$ is the maximum point of $L_{n+1}(\mathbf{Q})$ in (20) and the first and second terms of (20) are the constants, there is

$$\nabla_{\mathbf{Q}} L_{n+1}(\mathbf{Q}(n+1)) = 0. \tag{21}$$

So, the recursive estimation of parameter $\mathbf{Q}(n+1)$ can be obtained by

$$\mathbf{Q}(n+1) = \mathbf{Q}(n) + \frac{1}{n}[\Xi_1(\mathbf{Q}(n))]^{-1}\Gamma_1(\mathbf{Q}(n)). \tag{22}$$

From (3), the algorithm (22) should be revised as follows:

$$\mathbf{Q}(n+1) = \prod_{H_1}\left\{\mathbf{Q}(n) + \frac{1}{n}[\Xi_1(\mathbf{Q}(n))]^{-1}\Gamma_1(\mathbf{Q}(n))\right\} \tag{23}$$

where Π_{H_1} is the projection onto the constraint set H_1 composed of the constraints (3).

As to the predicted output $\hat{\mathbf{y}}(n)$, it is desirable that $\hat{\mathbf{y}}(n)$ can be as close to $\mathbf{y}(n)$ as possible for a given input $\mathbf{X}(n)$. In other words, for the observed data $(\mathbf{X}(n), \mathbf{y}(n))$ at time instant n, the ER-based model is updated to minimize the difference between $\beta_j(n)$ and $\hat{\beta}_j(n)$ generated by the ER-based model for each referential term. Here, $\beta_j(n)$ can be considered as a random variable and $\hat{\beta}_j(n)$ as its expectation. Define $\hat{\mathbf{B}}(n) = [\hat{\beta}_1(n), \ldots, \hat{\beta}_N(n)]^T$ as the predicted output at time instant n and assume that $\mathbf{B}(n)$ obeys the following complex normal distribution. We can obtain

$$f(\mathbf{B}(n)|\mathbf{X}(n),\mathbf{Q}) = (2\pi)^{\frac{-N}{2}}|\Sigma|^{\frac{-1}{2}} \times \exp\left\{-\frac{1}{2}\left(\mathbf{B}(n)-\hat{\mathbf{B}}(n)\right)^T \sum{}^{-1}\left(\mathbf{B}(n)-\hat{\mathbf{B}}(n)\right)\right\} \quad (24)$$

where $\mathbf{Q} = [\mathbf{V}^T, \sigma_1, \sigma_2]^T$ is a parameter vector and composed of the parameter vector $\mathbf{V} = [w_i]^T$ and the entries of the covariance matrix $\sum$ which is symmetric positive definite.

Due to the independence between the elements of $\mathbf{V}$ and the entries of $\sum$, $\Gamma_1(\mathbf{Q}(n))$, and $\Xi_1(\mathbf{Q}(n))$ in (23) can be written as

$$\Gamma_1(\mathbf{Q}(n)) = \left[\Gamma_1'(\mathbf{Q}(n))^T, \Gamma_1''(\mathbf{Q}(n))^T\right]^T \quad (25)$$

$$\Xi_1(\mathbf{Q}(n)) = \begin{bmatrix} \Xi_1'(\mathbf{Q}(n)) & \mathbf{0} \\ \mathbf{0} & \Xi_1''(\mathbf{Q}(n)) \end{bmatrix} \quad (26)$$

where $\Gamma_1'(\mathbf{Q}(n))$ and $\Xi_1'(\mathbf{Q}(n))$ are the derivates with respect to $\mathbf{V}$, $\Gamma_1''(\mathbf{Q}(n))$, and $\Xi_1''(\mathbf{Q}(n))$ are the derivates with respect to the entries of $\sum$. Obviously, there is

$$[\Xi_1(\mathbf{Q}(n))]^{-1} = \begin{bmatrix} [\Xi_1'(\mathbf{Q}(n))]^{-1} & \mathbf{0} \\ \mathbf{0} & [\Xi_1''(\mathbf{Q}(n))]^{-1} \end{bmatrix}. \quad (27)$$

When we consider only the parameter vector $\mathbf{V}$, the recursive algorithm (23) can be changed into the following form [from (24) and (27)]:

$$\mathbf{V}(n+1) = \mathbf{V}(n) + \frac{1}{n}[\Xi_1'(\mathbf{Q}(n))]^{-1}\Gamma_1'(\mathbf{Q}(n)). \quad (28)$$

In (28), $\mathbf{V}(n)$ is known. By (16) and (19), the ath element of the gradient vector $\Gamma_1'(\mathbf{Q}(n))$ and the entries of $\Xi_1'(\mathbf{Q}(n))$ at time instant n are

$$[\Gamma_1'(\mathbf{Q}(n))]_a = \frac{\partial\hat{\mathbf{B}}(n)^T}{\partial V_a}\Sigma(n)^{-1}\times\left(\mathbf{B}(n)-\hat{\mathbf{B}}(n)\right)\Big|_{\mathbf{V}=\mathbf{V}(n)} \quad (29)$$

$$[\Xi_1'(\mathbf{Q}(n))]_{a,b} = \frac{\partial\hat{\mathbf{B}}(n)^T}{\partial V_a}\sum(n)^{-1}\frac{\partial\hat{\mathbf{B}}(n)}{\partial V_b}\Bigg|_{\mathbf{V}=\mathbf{V}(n)} \quad (30)$$

where $a = 1,\ldots,p$ and $b = 1,\ldots,p$.

In (29) and (30), the covariance matrix $\sum(n)$ are required. Because $\beta_1(n),\ldots,\beta_N(n)$ should satisfy the equality constraint $\sum_{j=1}^N \beta_j(n) = 1$, they are not independent. In order to simplify the calculation, without loss of generality, we suppose that $\sum = (a_{i,j})_{N\times N}$ satisfies

$$\begin{cases} a_{i,j} = \sigma_1, & i = j \\ a_{i,j} = \sigma_2, & i \neq j. \end{cases} \quad (31)$$

Therefore, $\mathbf{Q} = [\mathbf{V}^T, \sigma_1, \sigma_2]^T$.

When $\mathbf{V}(n)$ is available, $\sigma_i(n)$ can be estimated by

$$\sigma_i(n) = \arg\max_{\sigma_i} \log f\left(\mathbf{B}(n)|\mathbf{X}(n),\mathbf{Q}\right)|_{\mathbf{V}=\mathbf{V}(n)} \quad (32)$$

where $i = 1,\ 2$. This equation can be calculated easily using FSOLVE function in MATLAB.

Because $\mathbf{V}$ should satisfy the constraints (3), the recursive algorithm with the constraints given in (23) should be adopted. First, let $\mathbf{V} = [V_1,\ldots,V_p]^T$ and then the constraints (3) can be represented as

$$h(\mathbf{V}) = h(V_1,\ldots,V_p) = \sum_{j=1}^p V_j - 1 = 0 \quad (33)$$

$$0 \le V_i \le 1, \quad i = 1,\ldots,p. \quad (34)$$

Let $\mathbf{h}(\mathbf{V}) = [h(\mathbf{V})]^T$ and $\mathbf{H}(\mathbf{V})$ denote the Jacobian matrix of $\mathbf{h}(\mathbf{V})$. The recursive algorithm (23) is revised and the following algorithm for dealing with the equality constraints given by (33) can be obtained [12]–[14], [39]

$$\overline{\mathbf{V}}(n+1) = \mathbf{V}(n) + \frac{1}{n}\boldsymbol{\pi}_1\{\mathbf{V}(n)\}[\Xi_1'(\mathbf{Q}(n))]^{-1}\Gamma_1'(\mathbf{Q}(n)) \quad (35)$$

where $\overline{\mathbf{V}}(n+1) = [\overline{V}_1(n+1),\ldots,\overline{V}_L(n+1)]^T$.

Assume that $\mathbf{I}_p$ is the identity matrix with dimension p. Then, there is

$$\boldsymbol{\pi}_1\{\mathbf{V}(n)\} = \mathbf{I}_p - \mathbf{H}(\mathbf{V}(n))^T\left(\mathbf{H}(\mathbf{V}(n))\mathbf{H}(\mathbf{V}(n))^T\right)^{-1} \times \mathbf{H}(\mathbf{V}(n)) = \frac{p-1}{p}\mathbf{I}_p \quad (36)$$

where $\boldsymbol{\pi}_1\{\mathbf{V}(n)\}$ achieves the projection of a vector on the orthogonal complement to the subspace determined by the rows of $\mathbf{H}(\mathbf{V}(n))$.

As given in (34), since the parameter $\overline{V}_j(j = 1,\ldots,p)$ estimated by (35) cannot go beyond the upper and lower bound, and then the projection algorithm is used [39]. Define the projecting operator $\boldsymbol{\pi}_2\{\overline{\mathbf{V}}(n+1)\}$ as

$$\boldsymbol{\pi}_2\left\{\overline{\mathbf{V}}(n+1)\right\} = \sum_{j=1}^p \hat{V}_j(n+1)\mathbf{e}_j \quad (37)$$

where $\mathbf{e}_j$ denotes a vector whose jth element is 1 and the other elements are all 0. $\hat{V}_j(n+1)$ is defined as follows:

First, for $j = 1,\ldots,p$, there is

$$\hat{V}_j(n+1) = \frac{\tilde{V}_j(n+1)}{\sum_{i=1}^N \tilde{V}_j(n+1)} \quad (38)$$

with

$$\tilde{V}_j(n+1) = \begin{cases} 0, & \overline{V}_j(n+1) < 0 \\ 1, & \overline{V}_j(n+1) > 1 \\ \overline{V}_j(n+1), & 0 \le \overline{V}_j(n+1) \le 1. \end{cases} \quad (39)$$

In addition, a group of the attributes in the ER-based model are activated at time instant n, which means that only some parameters can be updated. This may make the matrix $\Xi_1'(\mathbf{Q}(n))$ be singular. Hence, the final recursive algorithm can be given by

$$\mathbf{V}(n+1) = \boldsymbol{\pi}_2\Big\{\mathbf{V}(n) + \frac{\alpha}{n}\boldsymbol{\pi}_1\{\mathbf{V}(n)\} \times [\Xi_1'(\mathbf{Q}(n)) + \gamma\mathbf{I}_p]^{-1}\Gamma_1'(\mathbf{Q}(n))\Big\} \quad (40)$$

where $\alpha > 0$ is the step factor and can change the convergence speed. The matrix $\Xi_1'(\mathbf{Q}(n))$ is amended using $\gamma\mathbf{I}_p$ so that it becomes positive definite. $\gamma \ge 0$ is termed as the revision factor.

As a result of the above discussion, the procedure of the recursive algorithm for updating the ER-based prediction model under the judgmental output can be summarized as follows:

Initialization step: $\mathbf{V}(0), \sum(0), \alpha, \gamma$

Updating step: if $\mathbf{X}(n), \mathbf{y}(n), \mathbf{V}(n)$ are available, then $\sum(n)$ is obtained from (32), and $\mathbf{V}(n+1)$ is calculated from (40)

Prediction step: $\hat{\mathbf{y}}(n+1)$ is obtained by (6)–(8), then go to *Updating step*.

B. Recursive Parameter Estimation Algorithm Based on Numerical Output

In this case, $\mathbf{y}(n)$ is a numerical value. When the inputs of the ER-based model are independent, the true outputs $\mathbf{y}(1), \ldots, \mathbf{y}(n)$ can also be assumed to be independent, so there is

$$f(\mathbf{y}(1), \ldots, \mathbf{y}(n)|\mathbf{X}(1), \ldots, \mathbf{X}(n), \mathbf{Q}) = \prod_{\tau=1}^{n} f(\mathbf{y}(\tau)|\mathbf{X}(\tau), \mathbf{Q}) \tag{41}$$

where $f(\mathbf{y}(\tau)|\mathbf{X}(\tau), \mathbf{Q})$ is the PDF of $\mathbf{y}(\tau)$ at time instant τ.

In addition, for numerical data, the utility of the evaluation grade is a nonnegative number, e.g.,

$$u_i \geq 0, \qquad i = 1, \ldots, N. \tag{42}$$

And if F_j is preferred to F_i, then

$$u_j > u_i. \tag{43}$$

Similar to the deducing process in Section III-A, the following recursive algorithm can be obtained:

$$\mathbf{Q}(n+1) = \prod_{H_2} \left\{ \mathbf{Q}(n) + \frac{1}{n} \left[\Xi_2 (\mathbf{Q}(n))\right]^{-1} \Gamma_2 (\mathbf{Q}(n)) \right\} \tag{44}$$

where $\mathbf{Q}$ also consists of the attribute weights and other possible parameters. H_2 represents the constraint set which is composed of constraints (3), (42), and (43). And there are

$$\Gamma_2 (\mathbf{Q}(n)) = \nabla_{\mathbf{Q}} \log f(\mathbf{y}(n)|\mathbf{X}(n), \mathbf{Q}(n)) \tag{45}$$

$$\Xi_2 (\mathbf{Q}(n)) = E\big\{ -\nabla_{\mathbf{Q}} \nabla_{\mathbf{Q}}^T \log f(\mathbf{y}(n)|\mathbf{X}(n), \mathbf{Q}) \\ |\mathbf{X}(n), \mathbf{Q}(n)\big\}. \tag{46}$$

The output shown in (6) is represented as a distributed structure. In Section II-C, using the method of utility-based construction of the numerical outputs, its average score can be calculated by (9). Similarly, we hope that for a given input $\mathbf{X}(n)$, the ER-based model system can generate an output $\hat{\mathbf{y}}(n)$ as close to $\mathbf{y}(n)$ as possible. Here, $\mathbf{y}(n)$ is considered as a random variable and $\hat{\mathbf{y}}(n)$ can be considered as its expectation. So, we assume that the PDF of $\mathbf{y}(n)$ has the following normal distribution:

$$f(\mathbf{y}(n)|\mathbf{X}(n), \mathbf{Q}) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{ -\frac{(\mathbf{y}(n) - \hat{\mathbf{y}}(n))^2}{2\sigma} \right\} \tag{47}$$

where $\mathbf{Q} = [\mathbf{W}^T, \sigma]^T$ is the parameter vector. $\mathbf{W} = [\mathbf{V}^T, u_1, \ldots, u_N]^T$ denotes the parameters of attribute weights and grade utilities while σ is the variance.

Similar to (25)–(28), due to independence between the elements of $\mathbf{W}$ and σ, the recursive algorithm (44) can also be changed into the following form when we consider only $\mathbf{W}$:

$$\mathbf{W}(n+1) = \prod_{H_2} \left\{ \mathbf{W}(n) + \frac{1}{n} \left[\Xi_2' (\mathbf{Q}(n))\right]^{-1} \Gamma_2' (\mathbf{Q}(n)) \right\} \tag{48}$$

where $\Gamma_2'(\mathbf{Q}(n))$ and $\Xi_2'(\mathbf{Q}(n))$ are the derivates with respect to $\mathbf{W}$.

Let $\mathbf{W} = [W_1, \ldots, W_{p+N}]$. In (48), according to (45) and (46), $\Gamma_2'(\mathbf{Q}(n))$ and $\Xi_2'(\mathbf{Q}(n))$ have the following forms:

1) If $a, b = 1, \ldots, p$, the ath element of the gradient vector $\Gamma_2'(\mathbf{Q}(n))$ and the entries of $\Xi_2'(\mathbf{Q}(n))$ at time instant n are

$$\left[\Gamma_2' (\mathbf{Q}(n))\right]_a = \frac{(\mathbf{y}(n) - \hat{\mathbf{y}}(n))}{\sigma(n)} \sum_{j=1}^{N} u_j(n) \left. \frac{\partial \hat{\beta}_j(n)}{\partial W_a} \right|_{\mathbf{W}=\mathbf{W}(n)} \tag{49}$$

$$\left[\Xi_2' (\mathbf{Q}(n))\right]_{a,b} = \frac{1}{\sigma(n)} \left[\sum_{j=1}^{N} u_j(n) \frac{\partial \hat{\beta}_j(n)}{\partial W_a} \right] \\ \times \left. \left[\sum_{j=1}^{N} u_j(n) \frac{\partial \hat{\beta}_j(n)}{\partial W_b} \right] \right|_{\mathbf{W}=\mathbf{W}(n).} \tag{50}$$

2) If $a, b = p+1, \ldots, p+N$, there are

$$\left[\Gamma_2' (\mathbf{Q}(n))\right]_a = \left. \frac{\hat{\beta}_{a-p}(n) (\mathbf{y}(n) - \hat{\mathbf{y}}(n))}{\sigma(n)} \right|_{\mathbf{W}=\mathbf{W}(n)} \tag{51}$$

$$\left[\Xi_2' (\mathbf{Q}(n))\right]_{a,b} = \left. \frac{\hat{\beta}_{a-L}(n) \hat{\beta}_{b-p}(n)}{\sigma(n)} \right|_{\mathbf{W}=\mathbf{W}(n)} \tag{52}$$

Also, $\sigma(n)$ is required in (49)–(52). When $\mathbf{X}(n)$, $\mathbf{y}(n)$, and $\mathbf{W}(n)$ are available, it can be estimated by

$$\sigma(n) = \arg\max_{\sigma} \log f(\mathbf{y}(n)|\mathbf{X}(n), \mathbf{Q})|_{\mathbf{W}=\mathbf{W}(n)} \\ = (\mathbf{y}(n) - \hat{\mathbf{y}}(n))^2\big|_{\mathbf{W}=\mathbf{W}(n)}. \tag{53}$$

In (48), the constraint set H_2 is composed of the constraints (3), (42), and (43). The constraints (3) can also been written as

$$h(W_1, \ldots, W_p) = \sum_{i=1}^{p} W_i - 1 = 0 \tag{54}$$

$$0 \leq W_i \leq 1, \quad i = 1, \ldots, p. \tag{55}$$

The inequality constraints (43) can be represented as

$$p_g(W_{p+i}, W_{p+j}) = u_i - u_j < 0, \\ i = 1, \ldots, N-1, j = i+1, \ldots, N \tag{56}$$

where $g = (i-1)(N-1) - \sum_{k=1}^{i-2}(i-k-1) + j - i$.

Let $\mathbf{p}(\mathbf{W}) = [p_1(\mathbf{W}), \ldots, p_G(\mathbf{W})]^T$ and $G = (N-1)(N-2) - \sum_{k=1}^{N-3}(N-2-k) + 1$. Assume that $\mathbf{P}(\mathbf{W})$ is the Jacobian matrix of $\mathbf{p}(\mathbf{W})$. The following projecting operator is used to tackle with the inequality constraints [5], [27]

$$\Phi^+(\mathbf{W}(n)) = \mathbf{I}_N - \mathbf{P}(\mathbf{W}(n))^T \left(\mathbf{P}(\mathbf{W}(n))\mathbf{P}(\mathbf{W}(n))^T\right)^{-1} \mathbf{P}(\mathbf{W}(n)) \tag{57}$$

Thus, the following recursive algorithm can be obtained:

$$\mathbf{W}(n+1) = \boldsymbol{\pi}_2^+ \Big\{ \mathbf{W}(n) + \frac{\alpha}{n} \boldsymbol{\pi}_1^+ \{\mathbf{W}(n)\} \\ \times \left[\Xi_2' (\mathbf{Q}(n)) + \gamma \mathbf{I}_{p+N}\right]^{-1} \Gamma_2' (\mathbf{Q}(n)) \Big\} \tag{58}$$

where $\alpha \geq 1$ is the step factor and $\gamma \geq 0$ is used to amend the matrix $\Xi_2'(\mathbf{Q}(n))$. And also, there are

$$\boldsymbol{\pi}_1^+\{\mathbf{W}(n)\} = \begin{bmatrix} \boldsymbol{\pi}_1\{\mathbf{W}(n)\} & \mathbf{0} \\ \mathbf{0} & \Phi^+(\mathbf{W}(n)) \end{bmatrix} \tag{59}$$

$$\boldsymbol{\pi}_2^+\left\{\overline{\mathbf{W}}(n+1)\right\} = \sum_{j=1}^{p+N} \hat{W}_j(n+1)\mathbf{e}_j \tag{60}$$

where $\boldsymbol{\pi}_1\{\mathbf{W}(n)\}$ is defined in a similar way to (36) and used to deal with the equality constraints (54). If $j = 1, \ldots, p$, $\hat{W}_j(n+1)$ is used to deal with inequality constraints (55) and defined in a similar way to (38) and (39).

The procedure for the recursive algorithm for updating the ER-based model based on the numerical output can be summarized as follows:

Initialization step: $\mathbf{W}(0), \sigma(0), \alpha, \gamma$.
Updating step: if $\mathbf{X}(n), \mathbf{y}(n), \mathbf{W}(n)$ are available, then $\sigma(n)$ is obtained from (53), and $\mathbf{W}(n+1)$ is calculated from (58).
Prediction step: $\hat{\mathbf{y}}(n+1)$ is obtained by (9), and then go to *Updating step*.

Due to the proposed algorithms under the assumption of Gaussian distributions with respect to the outputs, they converge according to ergodic assumptions. In addition, the proposed recursive algorithms for updating ER-based prediction model are based on the stochastic approximation algorithm. The convergence theorem of the stochastic approximation algorithm has been proved by Kushner *et al.* [12]–[14], [17]. As such, we no longer give the analysis of convergence due to the limited space. However, it is noted that the proposed algorithms are locally optimal since these algorithms are based on the EM algorithm, which cannot guarantee to converge to the globally optimal point but can guarantee to converge to a locally optimal point.

IV. Practical Example

In this section, a practical example for missile reliability prediction is used to validate the proposed algorithms.

A. Problem Formulation

Missile is a key weapon in national defense. It is of utmost importance to guarantee the missiles' reliability and availability in combat readiness. Complicated in structure, a missile has many on-board systems and subsystems, among which the gyroscope in the inertial navigation system (INS) is most vulnerable to performance degradation caused by long exposure to the natural elements—sharp changing in temperatures and humidity for example. According to our statistics, nearly 70% missile faults result from INS failures. As a result, forecasting the health of this device has become our goal of this case study. Since the reliability of the missile is deteriorating with the increase of gyros drift degradation, the evolution of the missile reliability over time can be used as an indication of the performance of the missile. However, missile's reliability forecasting was overlooked in past several years in practice and there was few method available to achieve this aim. Therefore, there is a clear need to develop reliability forecasting technique for missiles. Particularly, several failures in missiles launching are strongly driving forces for building reliable and cost-effective prediction model that can provide the complete picture of missiles reliability states.

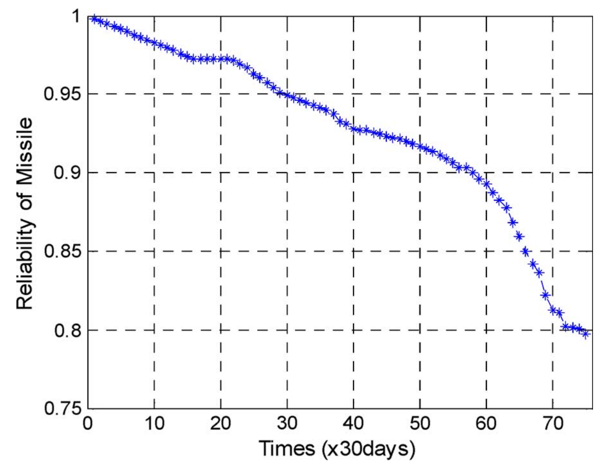

Fig. 1. Entire missile reliability data collected in the trial.

The reliability data for certain missile used in this paper is provided by the Reliability Analysis Center in Hi-Tech Institute of Xi'an, as illustrated in Fig. 1. The data sets include the reliability data for 75 sets of certain missile in the recent seven years. The reliability label on the y-axis denotes the probability that missile can still perform its required function, i.e., reliability, which is a quantity between 0 and 1. These data were collected month by month and obtained through assessing the drift degradation of gyros fixed in the inertial navigation platform and engineers' judgmental information. To protect sensitive and proprietary information, we cannot provide the details of evaluation process for reliability. As a matter of fact, our obtained reliability data is a diagnostic check of the current reliability, which reflects the health status of the missile. But in our practice, conducting such a diagnostic check is time-consuming and cost-expensive. Therefore, our main concern is to construct a prediction model which can reflect the evolution of the missile's reliability. Once this prediction model is constructed, we can use it for reliability prediction. This is valuable in practice since it can cut down the expenditures in reliability testing and perform reliability prediction in a cost-effective way. Fig. 1 shows that the reliability data of this kind of missiles is a time series and experience the decreasing trend in this case study. In the following, we use this data set to demonstrate the proposed method.

B. Referential Points of the Reliability Data

For the missile reliability data, the reliability can be described linguistically due to its easy interpretation. Particularly for the safety-critical equipment, we care about not only the reliability value but also the reliability level. It is common to assess a reliability level by degrees to which it belongs to such linguistic variables as "very high," "high," "average," "low," and "very low" that are referred as reliability expressions. Clearly, such expressions can provide a more informative scheme for knowledge representation. For example, an expert may frequently state that he is 20% sure system reliability is low and 80% sure it is high. As such, it is reasonable to model reliability forecasting using these linguistic variables. In this case study, for simplicity, we only consider a simple case with three reliability levels, "$High$"(F_1), "$Average$"(F_2), and "Low"(F_3). As such, we define the following discernment framework of the missile reliability states as follows:

$$F = \{F_j, j = 1, 2, 3\} = \{High, Average, Low\}$$

All related factors to the missile reliability may then be assessed with reference to this framework using the rule-based information transformation technique. Through analyzing the characteristic of the missile reliability data similar to Yang's method [35], [38], it is

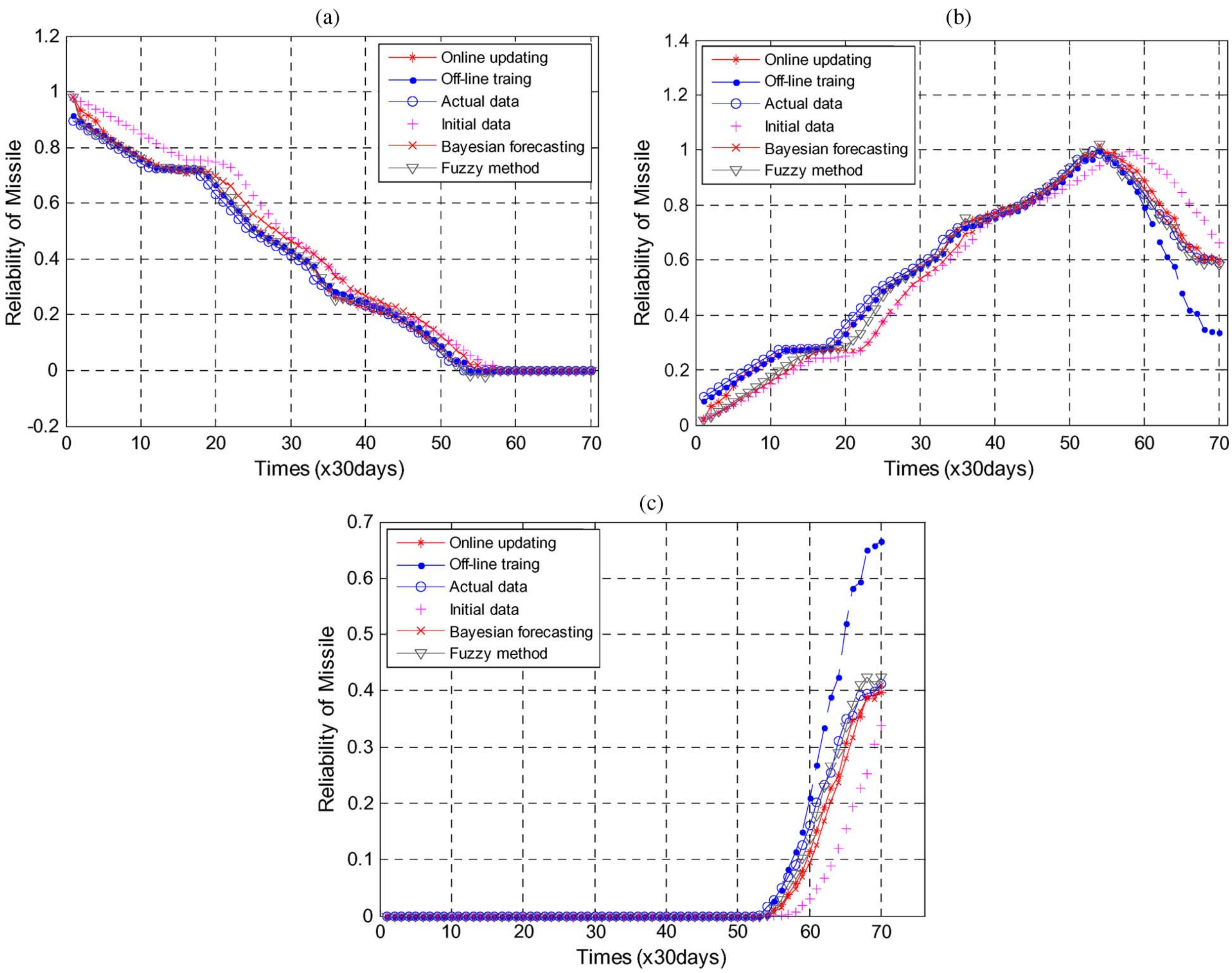

Fig. 2. Predicted output to be (a) “High”, (b) “Average”, (c) “Low”.

assumed that equivalence rules for data transformation can be acquired as follows:

1) If the missile reliability is 1.0, then system reliability state is ranked to be “$High$” with the matched degree 1.0, that is $F_1 = 1.0$.
2) If the missile reliability is 0.8, then system reliability state is ranked to be “$Average$” with the matched degree 1.0, that is $F_2 = 0.8$.
3) If the missile reliability is 0.65, then system reliability state is ranked to be “Low” with the matched degree 1.0, that is $F_3 = 0.65$.

For instance, using these rules, the 35th data with reliability value 0.9412 in Fig. 1 can be transformed into $S(y(35)) = \{(High, 0.706), (Average, 0.294), (Low, 0)\}$ based on rule-based transformation [35]. However, it is worth noting that the referential values used in the above rules are problem-specific. In our case study, it is usually required that the reliability of missiles assessed by drift degradation is no less than 0.8 since missile are a safety-critical object. Through the above rules, all related factors to the missile reliability can be transformed into the discernment framework F. This process is similar to fuzzification step used in fuzzy set theory [21].

C. Model

As illustrated in Fig. 1, the reliability data of the missile is a time series. So, it is reasonable to assume that the current reliability value y_t is related to the most recent y_{t-1} and the past values $y_{t-2}, \ldots, y_{t-p}$. This is because the next reliability value is dependent on the current level of system state to a large extent. As such, the proposed ER-based prediction model can be applied to predict the missile reliability. First, Cao’s method [2] is used to determine the embedding dimension p. In this correspondence paper, a relatively larger number $p = 5$ is obtained for the embedding dimension. Therefore, we can transform 75 observation values to 70 input-output patterns. So, the prediction model can be written as

$$\hat{y}(t) = \hat{y}_t = f(y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, y_{t-5}) \tag{61}$$

In (61), the original input and output data were all provided as numerical numbers, so there is a need to equivalently transform the numerical value into the distributed structures using the rule-based transformation technique [35].

D. Simulation Results Based on Judgmental Output

In this case, along the line of the algorithm summarized at the end of Section III-A, the initial attribute weights of inputs are set to be $w = [0.6, 0.1, 0.1, 0.1, 0.1]$. At the same time, the step factor α and revision factor γ are set to 1500 and 0.015, respectively. The predicted results using these initial values are given in Fig. 2(a)–(c). The results show that the values of the estimated reliability by the initial ER-based prediction model do not match the actual values. This means that the initial ER-based prediction model is indeed rather bad. Therefore, it is necessary to update the ER-based prediction model online. Based on the input-output values and the initial model parameters given in

TABLE I
COMPARISON OF THE PREDICTION RESULTS UNDER JUDGMENTAL OUTPUT

		Initial model	Off-line training [10]	Online updating	Bayesian forecasting [8]	Fuzzy system[21]
MAPE	High	0.1327	0.0408	0.0198	0.1096	0.0363
	Average	0.2096	0.0756	0.0032	0.1549	0.0516
	Low	0.1489	0.1296	0.0659	0.0685	0.0309
RMSE	High	0.0731	0.0174	0.0209	0.0447	0.0237
	Average	0.0964	0.0745	0.0269	0.0651	0.0443
	Low	0.0637	0.0725	0.0169	0.0233	0.0090
Updating/Training time (s)		0.3946	2.4965	0.8041	0.8041	0.3274

initialization step, the *updating step*, and *prediction step* are used to obtain the predicted reliability. The test results are illustrated in Fig. 2(a)–(c), where the comparisons are shown between the actual outputs and the predicted outputs that are generated using the updated ER-based prediction model and the initial one. It is obvious that the predicted outcomes generated by the updated ER-based prediction model can trace the changes of the system reliability well. From Fig. 2(a)–(c), it can be seen that the matched degree assigned to the grade "$High$" decreases from 0.9803 to 0 monotonously. As such, we can say that the probability of system reliability assigned to the grade "$High$" decreases from 0.9803 to 0. At the same time, the probability of system reliability assigned to the grade "$Average$" firstly increases monotonously from 0.0197 to 0.997, then decreases to 0.6025. The probability of system reliability assigned to the grade "Low" increases monotonously from 0 to 0.3975. The above analysis shows that missile reliability is subjected to a degradation process over time and the proposed predicted method can give a complete picture about the state of the missile reliability.

In order to further demonstrate the accuracy of the proposed recursive algorithm, the mean absolute percentage error (MAPE) and the root mean square error (RMSE) are used. After calculation, the MAPE and the RMSE between the observed values and the estimated values to the linguistic term "$High$" generated by the initial ER-based model are 0.1327 and 0.0731, respectively. On the other hand, the MAPE and RMSE for the updated model are 0.0198 and 0.0209, respectively. Obviously, the updated ER-based model can replicate the relationship among $y_{t-1}, y_{t-2}, \ldots, y_{t-5}$ and the actual judgmental output more accurately than the initial one.

It is of interest to compare the results generated by the proposed method with the offline training model [10], classical Bayesian forecasting method [8], [28], and T-S fuzzy rule base [20], [21], since the first one used the same model structure as used in this correspondence paper while the last two are all based on the Gaussian distribution assumption. In the simulation, the parameters of Bayesian forecasting method [8] are updated using Kalman filter, and the parameters in T–S fuzzy rule base are adjusted by recursive least square error algorithm developed by Passino in [21]. The simulated results are illustrated in Fig. 2 and the comparative results are summarized in Table I.

Table I shows that the online updating algorithm outperforms Bayesian forecasting method and fuzzy method with respect to prediction accuracy, but the latter can be implemented more quickly. In the simulated results, it is shown that the proposed updating method as well as Bayesian forecasting and fuzzy method can obtain more satisfactory results than the offline training method in terms of prediction accuracy and speed. The main cause is that the optimization model is multi-objective optimization under judgmental outputs [10]. In reality, to solve multi-objective optimization problem is much more difficult than to solve single-objective optimization problem. Practically, it is very difficult to obtain the general optimal solution for multi-objective optimization problems.

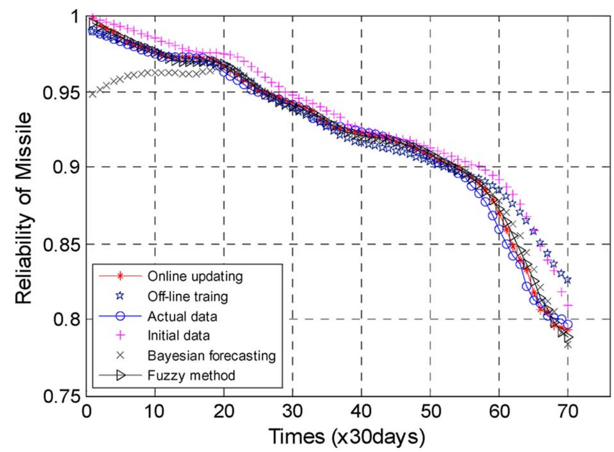

Fig. 3. Predicted results under numerical output.

E. Simulation Results Under Numerical Output

In order to verify the proposed recursive algorithm under numerical output case, we will give the following simulation. According the algorithm summarized at the end of Section III-B, the initial attribute weights of all inputs are set to be equivalent, and the initial grade utilities is set to $[u_1, u_2, u_3] = [1, 0.9, 0.65]$. At the same time, the step factor α and revision factor γ are set to 0.5 and 0.02, respectively. As shown in Fig. 3, it is obvious that the values of the estimated reliability by the initial ER-based prediction model do not match the actual values. Therefore, it is necessary to update the ER-based prediction model online.

After the input values $(y_{t-1}, y_{t-2}, \ldots, y_{t-5})$ are transformed and represented in terms of the referential values, the recursive algorithm summarized in Section III-B is used to update the ER-based model. The updated results are illustrated in Fig. 3. Through calculation, the MAPE and the RMSE between the actual values and the predicted values generated by the initial ER-based model is 1.41% and 0.0167, respectively. On the other hand, the MAPE and RMSE for the updated ER-based model is 0.23% and 0.0027, respectively. Obviously, the updated ER-based prediction model can replicate the relationship among $(y_{t-1}, y_{t-2}, \ldots, y_{t-5})$ and y_t much more accurately than the initial one. Similarly, we compare our method with the offline training method, Bayesian forecasting method, and fuzzy method with respect to the MAPE and the RMSE. The detailed results can be found in Table II.

Similarly, Table II shows clearly the difference in terms of prediction precision among these methods. Obviously, the online updating algorithm can obtain more accurate prediction than other models. On the other hand, the training process of the offline optimization model is time-consuming in this correspondence paper. By contrary, our modeling process, Bayesian method, and fuzzy method are more efficient.

TABLE II
COMPARISON OF THE PREDICTION RESULTS UNDER NUMERICAL OUTPUT

	Initial model	Off-line training [10]	Online updating	Bayesian forecasting [8]	Fuzzy system [21]
MAPE	0.0141	0.0101	0.0023	0.0063	0.0097
RMSE	0.0167	0.0149	0.0027	0.0092	0.0040
Updating/ Training time (s)	0.0175	2.0151	0.0890	0.0335	0.0425

This is very important when there is a high real-time requirement for system reliability forecasting.

From the above numerical studies, we see that the initial forecasting model given by the experts are not accurate and the proposed recursive algorithms can be used to update it no matter the outputs of the ER-based prediction model are numerical or judgmental. The comparative results with Bayesian method and fuzzy method also suggest that our method can generate better prediction accuracy. What's more, the proposed predicted method can give a complete picture about the state of the missile reliability, which can provide very useful information for predictive maintenance of the missile integrity.

V. CONCLUSION

This correspondence paper is concerned with the recursive algorithms to update the ER-based prediction model from the probability-based point of view. The recursive algorithms provide an innovative way to enhance the capability of the recently developed ER-based prediction model. Different from the existing optimization models reported for training ER-based model, the proposed recursive algorithms can be used to fine tune the ER-based prediction model once new information becomes available. A practical study for the missile reliability prediction is examined to demonstrate how the proposed recursive algorithms can be implemented. The experimental and comparative studies show that the developed model with online updating may be widely applied in prediction practice.

The results show that the proposed recursive updating algorithm is an effective method for training the ER-based prediction model. However, in this correspondence paper, few theoretical discussions are directly made to address the performance analysis of the proposed recursive algorithms. Generally, the expected performance of recursive algorithm involves convergence, divergence, self-stabilizing, and parameters' sensitivity, etc. As such, we think it is important to conduct performance analysis of the recursive updating algorithms in further research. First, the detailed analysis will be given to convergence and divergence ability. Second, the influence of parameters in the recursive algorithms will be analyzed fully, such as the step factor and revision factor. This research will facilitate improving the recursive updating algorithm. These requirements pose challenges for future research.

ACKNOWLEDGMENT

The authors would like to thank the Editor-in-Chief and the Associate Editor for giving them the opportunity to publish this correspondence paper. They would like to sincerely thank and acknowledge the wit-outputs and the tremendous work performed by the Associate Editor and the four anonymous reviewers for their insightful suggestions and thorough review, which greatly improve this correspondence paper. The authors are also grateful to Prof. W. Zhang for her help to provide reliability data, Prof. Y.-L. Zhang and Prof. M.-Y. Chen for improving writing quality.

REFERENCES

[1] C. Angeli and A. Chatzinikolaou, "Prediction and diagnosis of fault in hydraulic systems," *Proc. Inst. Mech. Eng.*, vol. 216, no. 2, pp. 293–297, 2002.

[2] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Phys. D, Nonlinear Phenom.*, vol. 110, no. 1/2, pp. 43–50, Dec. 1997.

[3] M. Z. Chen, D. H. Zhou, and G. P. Liu, "A new particle predictor for fault prediction of nonlinear time-varying systems," *Develop. Chem. Eng. Mineral Process.*, vol. 13, no. 3/4, pp. 379–388, 2005.

[4] K. Y. Chen, "Forecasting system's reliability based on support vector regression with genetic algorithms," *Reliab. Eng. Syst. Saf.*, vol. 92, no. 4, pp. 423–432, Apr. 2007.

[5] P. J. Chung and J. F. Bohme, "Recursive EM and SAGE-inspired algorithms with application to DOA estimation," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2664–2677, Aug. 2005.

[6] A. P. Dempster, N. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.

[7] Y. X. Diao and K. M. Passino, "Stable fault-tolerant adaptive fuzzy/neural control for a turbine engine," *IEEE Trans. Control Syst. Technol.*, vol. 9, no. 3, pp. 494–509, May 2001.

[8] P. J. Harrison and C. F. Stevens, "Bayesian forecasting," *J. Roy. Statist. Soc. Ser. B*, vol. 38, no. 3, pp. 205–247, 1976.

[9] S. L. Ho and M. Xie, "The use of ARIMA models for reliability forecasting and analysis," *Comput. Ind. Eng.*, vol. 35, no. 1/2, pp. 213–216, Oct. 1998.

[10] C. H. Hu, X. S. Si, and J. B. Yang, "Systems reliability forecasting based on evidential reasoning algorithm with nonlinear optimization," *Expert Syst Appl.*, vol. 37, no. 3, pp. 2550–2562, Mar. 2010.

[11] K. A. Kaiser and N. Z. Gebraeel, "Predictive maintenance management using sensor-based degradation models," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 39, no. 4, pp. 840–849, Jul. 2009.

[12] H. J. Kushner and M. L. Kelmanson, "Stochastic approximation algorithms of the multiplier type for the sequential Monte Carlo optimization of stochastic systems," *SIAM J. Control Optim.*, vol. 14, no. 5, pp. 827–842, 1976.

[13] H. J. Kushner and S. Lakshmivarahan, "Numerical studies of stochastic approximation procedures for constrained problems," *IEEE Trans. Autom. Control*, vol. AC-22, no. 3, pp. 428–439, Jun. 1977.

[14] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms and Applications*. New York: Springer-Verlag, 1997.

[15] J. R. Layne and K. M. Passino, "A fuzzy dynamic model based state estimator," *Fuzzy Sets Syst.*, vol. 122, no. 1, pp. 45–72, Aug. 2001.

[16] T. Li and L. Guo, "Optimal fault-detection filtering for non-Gaussian systems via output PDFs," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 39, no. 2, pp. 476–481, Mar. 2009.

[17] L. Ljung, "Analysis of recursive stochastic algorithm," *IEEE Trans. Autom. Control*, vol. AC-22, no. 4, pp. 551–575, Aug. 1977.

[18] J. Luo, K. R. Pattipati, L. Qiao, and S. Chigusa, "Model-based prognostic techniques applied to a suspension system," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 38, no. 5, pp. 1156–1167, Sep. 2008.

[19] G. Niu and B. S. Yang, "Dempster-Shafer regression multi-step-ahead time-series prediction toward data-driven machinery prognosis," *Mech. Syst. Signal Process.*, vol. 23, no. 3, pp. 740–751, Apr. 2009.

[20] R. Ordonez, J. T. Spooner, and K. M. Passino, "Experimental studies in nonlinear discrete-time adaptive prediction and control," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 2, pp. 275–286, Apr. 2006.

[21] K. M. Passino and S. Yurkovich, *Fuzzy Control*. Menlo Park, CA: Addison-Wesley, 1998.

[22] W. Pedrycz, *Fuzzy Control and Fuzzy Systems*. Taunton, U.K.: Res. Stud. Press, 1993.

[23] J. Pearl, *Probabilistic Reasoning in Intelligence Systems*. San Mateo, CA: Morgan Kaufmann, 1988.

[24] Y. H. Qian, J. Y. Liang, and C. Y. Dang, "Incomplete multigranulation rough set," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 40, no. 2, pp. 420–431, Mar. 2010.

[25] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton Univ. Press, 1976.

[26] W. Stach, L. Kurgan, and W. Pedrycz, "Numerical and linguistic prediction of time series with the use of fuzzy cognitive maps," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 1, pp. 61–72, Feb. 2008.

[27] D. M. Titterington, "Recursive parameter estimation using incomplete data," *J. Roy. Statist. Soc. Ser. B*, vol. 46, no. 2, pp. 257–267, 1984.

[28] M. West, "Robust sequential approximate Bayesian estimation," *J. Roy. Statist. Soc. Ser. B*, vol. 43, no. 2, pp. 157–166, 1981.

[29] D. Wang, D. H. Zhou, Y. H. Jin, and S. J. Qin, "A strong tracking predictor for nonlinear processes with input time delay," *Comput. Chem. Eng.*, vol. 28, no. 12, pp. 2523–2540, Nov. 2004.

[30] Y. M. Wang, J. B. Yang, and D. L. Xu, "Environmental impact assessment using the evidential reasoning approach," *Eur. J. Oper. Res.*, vol. 174, no. 3, pp. 1885–1913, Nov. 2006.

[31] P. Walley, "Measures of uncertainty in expert system," *Artif. Intell.*, vol. 83, no. 1, pp. 1–58, May 1996.

[32] K. Xu, M. Xie, L. C. Tang, and S. L. Ho, "Application of neural networks in forecasting engine systems reliability," *Appl. Soft Comput.*, vol. 2, no. 4, pp. 255–268, Feb. 2003.

[33] J. B. Yang and M. G. Singh, "An evidential reasoning approach for multiple attribute decision making with uncertainty," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, no. 1, pp. 1–18, Jan. 1994.

[34] S. K. Yang and T. S. Liu, "A petri net approach to early failure detection and isolation for preventive maintenance," *Qual. Reliab. Eng. Int.*, vol. 14, no. 5, pp. 319–330, Sep./Oct. 1998.

[35] J. B. Yang, "Rule and utility based evidential reasoning approach for multi-attribute decision analysis under uncertainties," *Eur. J. Oper. Res.*, vol. 131, no. 1, pp. 31–61, May 2001.

[36] J. B. Yang and D. L. Xu, "On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 32, no. 3, pp. 289–304, May 2002.

[37] S. K. Yang, "An experiment of state estimation for predictive maintenance using Kalman filter on a DC motor," *Reliab. Eng. Syst. Saf.*, vol. 75, no. 1, pp. 103–111, Jan. 2002.

[38] J. B. Yang, J. Liu, D. L. Xu, J. Wang, and H. W. Wang, "Optimal learning method for training belief rule based systems," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 37, no. 4, pp. 569–585, Jul. 2007.

[39] Z. J. Zhou, C. H. Hu, J. B. Yang, D. L. Xu, and D. H. Zhou, "Online updating belief-rule-based systems for pipeline leak detection under expert intervention," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7700–7709, 2009.

[40] P. Zhou, B. M. Ang, and K. L. Poh, "A trigonometric grey prediction approach to forecasting electricity demand," *Energy*, vol. 31, no. 14, pp. 2839–2847, Nov. 2006.