

# Probabilistic Prognosis with Dynamic Bayesian Networks

Gregory Bartram<sup>1</sup>, Sankaran Mahadevan<sup>2</sup>

<sup>1,2</sup>*Vanderbilt University, Nashville, TN, 37235, USA*

*gregory.w.bartram@vanderbilt.edu*  
*sankaran.mahadevan@vanderbilt.edu*

## ABSTRACT

This paper proposes a methodology for probabilistic prognosis of a system using a dynamic Bayesian network (DBN). Dynamic Bayesian networks are suitable for probabilistic prognosis because of their ability to integrate information in a variety of formats from various sources and give a probabilistic representation of the system state. Further, DBNs provide a platform naturally suited for seamless integration of diagnosis, uncertainty quantification, and prediction. In the proposed methodology, a DBN is used for online diagnosis via particle filtering, providing a current estimate of the joint distribution over the system variables. The information available in the state estimate also helps to quantify the uncertainty in diagnosis. Next, based on this probabilistic state estimate, future states of the system are predicted using the DBN and sequential or recursive Monte Carlo sampling. Prediction in this manner provides the necessary information to estimate the distribution of remaining use life (RUL). The prognosis procedure, which is system specific, is validated using a suite of offline hierarchical metrics. The prognosis methodology is demonstrated on a hydraulic actuator subject to a progressive seal wear that results in internal leakage between the chambers of the actuator.

## 1. INTRODUCTION

### 1.1. Background

The rise of complex and costly systems for use in extreme environments has resulted in new challenges in maintenance, planning, decision-making and monitoring for these systems. To reliably execute the missions they were designed for, these systems must be meticulously maintained. Traditional schedule-based maintenance results in unnecessary system downtime and the potential for serious problems to develop between routine maintenance. The alternative, condition-

based maintenance (CBM) (Jardine, Lin, & Banjevic, 2006), monitors systems as they operate, alerting personnel when faults occur. Maintenance is performed on-demand, resulting in less downtime and lower costs. Maintenance and operational decision-making, such as resource allocation, may further benefit from knowing the prognosis of a system.

A system's prognosis is a measure of its fitness to perform a prescribed task. Quantitatively, prognosis is commonly expressed through the remaining useful life (RUL). RUL quantifies the amount of time until a system reaches some failure criterion, e.g. fault magnitude or performance metric crosses a threshold or system is no longer operable. Ideally, the uncertainty in RUL is quantified by estimating the distribution of RUL, resulting in a probabilistic prognosis. Importantly, probabilistic prognosis assesses the outlook for a specific instantiation of a system, or a particular unit under test (UUT). Measurement data updates the belief about the present state and RUL of the particular UUT. In this way, probabilistic prognosis differs from probabilistic reliability analysis, which aggregates data to obtain a probabilistic reliability estimate for a population as opposed to an individual. Such population statistics may be suitable for tasks such as system design, but less helpful for operational and maintenance decisions that focus on individual units, as is the case in CBM. Prognosis, on the other hand, tracks the health of an individual unit.

A prognosis methodology should thus have several important characteristics. It should provide a distribution of RUL as opposed to a point estimate, thus accounting for the uncertainty coming from many sources i.e. natural variability, information uncertainty, and model uncertainty (note: information includes measurements). In most situations, prognosis about the future is based on diagnosis of the current state; therefore it should account for uncertainty in diagnosis (i.e. uncertainty in damage existence, location and size, and sensor performance uncertainty). It should allow easy transition between situations when measurements are available and when they are unavailable. Finally, the methodology should survive rigorous validation.

Prognosis methodologies may be divided into statistical, data-based, model-based, and hybrid approaches (see e.g.

---

Bartram and Mahadevan. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Jardine et al. (Jardine et al., 2006; Tran & Yang, 2009)). Statistical approaches include statistical process control (Goode, Moore, & Roylance, 2000), logistic regression (Yan, Koç, & Lee, 2004), survival models (Banjevic & Jardine, 2006; Vlok, Wnek, & Zygmunt, 2004), and stochastic process models (Lin & Makis, 2004; Wenbin Wang, 2002; W. Wang, Scarf, & Smith, 2000).

Data-based approaches consist of machine learning methods (support vector machines (Farrar & Worden, 2012), relevant vector machines (Tipping, 2001), neural networks (Dong & Yang, 2008; Farrar & Worden, 2012; Vachtsevanos & Wang, 2001; W. Q. Wang, Golnaraghi, & Ismail, 2004; Yam, Tse, Li, & Tu, 2001; Zhang & Ganesan, 1997)), and graphical models such as dynamic Bayesian networks (DBNs), hidden Markov models (HMMs) (Chinnam & Baruah, 2003; Kwan, Zhang, Xu, & Haynes, 2003). Liu et al. (Liu, Saxena, Goebel, Saha, & Wang, 2010) use adaptive recurrent neural networks for the estimation of battery RUL. Goebel et al. (Goebel, Saha, Saxena, Mct, & Riacs, 2008) compare relevance vector machines (RVMs), Gaussian process regression (GPR) and neural network (NN) methods for prognosis. Gebraeel (Gebraeel & Lawley, 2008) uses NNs for degradation modeling and test the methodology on ball bearings. Saha et al. (Bhaskar Saha, Goebel, & Christophersen, 2009) compare relevance vector machines (RVMs, a Bayesian implementation of support vector machines) and particle filtering to estimate RUL distributions for batteries.

In model-based approaches, system models are used to estimate RUL or other relevant metrics. Such methods rely on mathematical models for prediction. These include physics-based failure models (Kacprzynski, Sarlashkar, Roemer, Hess, & Hardman, 2004), filtering models (Orchard & Vachtsevanos, 2009), (Lorton, Fouladirad, & Grall, 2013), (B. Saha, Celaya, Wysocki, & Goebel, 2009) (Khan, Udpa, & Udpa, 2011), and regression and statistical models. Orchard and Vachtsevanos (Orchard & Vachtsevanos, 2009) use state estimation models combined with particle filtering for diagnosis and estimation of the RUL distribution of a planetary gear. Lorton et al. (Lorton et al., 2013) combine the differential equations of a system with system measurements via particle filtering for probabilistic model-based prognosis.

Hybrid methodologies combine multiple approaches, i.e., a combination of data-driven and model-based approaches. E.g. Kozlowski (Kozlowski, 2003) uses Autoregressive moving average (ARMA) models, neural networks, and fuzzy logic for estimation of the state of health, state of charge, and state of life of batteries.

DBNs are probabilistic graphical models with diagnostic and prognostic capabilities. They have shown promise in several recent applications. Bartram and Mahadevan (Gregory Bartram & Mahadevan, 2013) have presented an application of DBNs to prognosis of a hydraulic actuator at the PHM conference. Dong and Yang (Dong & Yang, 2008) use DBNs combined with particle filtering to estimate the RUL

distribution of drill bits in a vertical drilling machine. While very useful, particle filtering is not the only inference method available for prognosis. Jinlin and Zhengdao (Jinlin & Zhengdao, 2012) use DBNs modeling discrete variables and the Boyen-Koller algorithm for prognosis. Tobon-Mejia et al. (Tobon-Mejia, Medjaher, Zerhouni, & Tripot, 2012) use mixtures of Gaussian HMMs (a form of DBN) to estimate the RUL distributions for bearings. The junction tree algorithm is used for exact inference. The prognosis methodology is validated using the hierarchical metrics proposed by Saxena et al. (Saxena, Celaya, Saha, Saha, & Goebel, 2010).

## 1.2. Motivation

While the preceding literature review identifies several prognosis approaches, prognosis is still an emerging research area with room for much additional work. The combination of DBN and particle filtering has many qualities that are attractive for prognosis:

- 1) The graphical representation of a problem provided by DBNs aids understanding of interactions in a system.
- 2) DBNs provide a probabilistic model of the system that accounts for uncertainty due to natural variability, measurement error, and modeling error.
- 3) DBNs can integrate different types of information that may be encountered during prognosis (including expert opinion, reliability data, mathematical models, operational data, and laboratory data) into a unified system model.
- 4) DBNs can update the distributions of all variables in the network when observations are obtained for any one or more variables. This allows the most recent system measurements to be accounted for in prognosis.
- 5) The probabilistic state estimate generated during particle filtering contains information about diagnosis uncertainty that can be used in prognosis and decision making.
- 6) Particle filtering algorithms can take advantage of parallel computing, thus reducing computation time.

Additionally, prognosis methodologies reported in the literature are mostly application-specific. There is a need for a prognosis methodology that can make use of heterogeneous information and be applied to a wide range of problems.

## 1.3. Contributions

This paper proposes a framework for probabilistic prognosis. The methodology advances the use of DBNs in prognosis by integrating heterogeneous information. Further, the DBN-based methodology addresses the need for a general prognosis framework for developing validated prognosis methodologies for any system. The methodology provides a means for quantifying diagnosis uncertainty that can be used in prognosis and decision making.

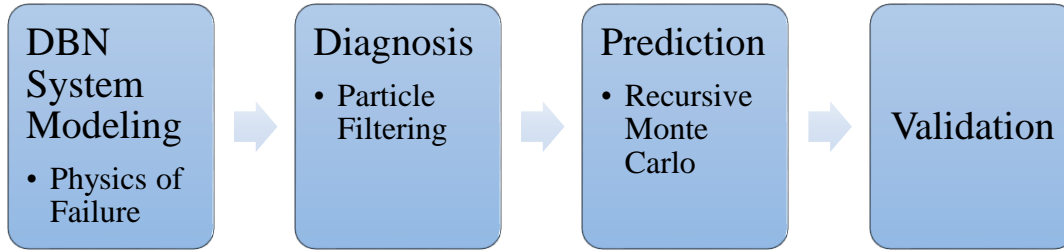


Figure 1. Proposed Prognosis Methodology

The DBN is constructed from prior information, including physics-of-failure models. The DBN is a store of prior information, but also provides the means for integrating current measurements into a probabilistic estimate of the current state of a system. In this paper, particle filter-based inference is used for diagnosis, and forward sampling in the DBN is used for recursive prediction. Diagnosis uncertainty is quantified using the probabilistic state estimate of the system that results from particle filtering. Sample-based estimates of detection and isolation probabilities as well as density estimates of system damage parameters are developed based on marginal distributions in the state distribution. This information can be used in decision making for improvement of the health monitoring system. The probabilistic state estimate also provides a seamless transition from diagnosis to future state prediction using recursive sampling. The ability of the methodology to estimate RUL is validated using metrics from Saxena et al. (Saxena et al., 2010). The methodology is illustrated for a hydraulic actuator with a seal leak.

The remainder of this paper is organized as follows. Section 2 details the proposed prognosis methodology, including system modeling, diagnosis, prediction, and validation. In Section 3, the proposed methodology is demonstrated on a hydraulic actuator system with a progressive internal leak. Section 4 discusses conclusions and future work.

## 2. PROPOSED PROGNOSIS FRAMEWORK

The challenge of prognosis is to minimize the uncertainty in the estimated distribution of RUL given constraints on available information about the system, operating environment and loading conditions, computational resources, and time horizon. In this paper, a DBN-based prognosis framework is proposed. The prognosis framework first constructs a DBN-based system model using heterogeneous information sources. Expert opinion, reliability data, mathematical models, and operational and laboratory data are used in the construction of the DBN model. In particular, inclusion of physics-of-failure models is important in prognosis. The evolution of phenomena such as cracking, wear, and corrosion play a large role in determining the health of a mechanical system. The system model is used for diagnosis to obtain information about the current state of the system. A recursive Monte Carlo sampling then predicts

future system states and estimates the RUL distribution. Finally, the prognosis capability of the resulting system model, diagnostic, and predictive algorithms are validated using a four-step hierarchical procedure. The prognosis procedure is shown in Fig. 1.

### 2.1. Dynamic Bayesian Networks

A dynamic Bayesian network is the temporal extension of a static Bayesian network (BN). A static BN, also referred to as a belief network and directed acyclic graph (DAG), is a probabilistic graphical representation of a set of random variables and their conditional dependencies. Variables are represented by nodes (vertices) and conditional dependence is represented by directed edges. Unconnected nodes are conditionally independent of each other. The acyclic requirement means that no paths exist in the graph where, starting at node  $x_i$ , it is possible to return to node  $x_i$ .

A DBN describes the joint distribution of a set of variables  $\mathbf{x}$  on the time interval  $[0, \infty)$ . This is a complex distribution, but may be simplified using the Markov assumption. The Markov assumption requires only the present state of the variables  $\mathbf{x}^t$  to estimate  $\mathbf{x}^{t+1}$ , i.e.  $p(\mathbf{x}^{t+1} | \mathbf{x}^0, \dots, \mathbf{x}^t) = p(\mathbf{x}^{t+1} | \mathbf{x}^t)$  where  $p$  indicates a probability density function and bold letters indicate a vector quantity. A linear relationship between  $\mathbf{x}^{t+1}$  and  $\mathbf{x}^t$  is implied by this assumption. Additionally, the process is assumed to be stationary, meaning that  $p(\mathbf{x}^{t+1} | \mathbf{x}^t)$  is independent of  $t$ . This approach to modeling DBNs is developed by Friedman et al. (Friedman, Murphy, & Russell, 1998).

DBNs provide a flexible modeling framework, allowing integration of expert opinion, reliability data, mathematical models (including state-space, surrogate, and physics-of-failure models), existing databases of operational and laboratory data, and online measurement information. Bartram and Mahadevan (G. Bartram & Mahadevan, S., 2013) have proposed a methodology for the integration of such heterogeneous information into DBN system models. In the next section, that discussion is extended to consider physics-of-failure models, which are of particular importance in prognosis.

Figure 2 shows a cantilever beam with a potential crack and/or damage at the support. The damage at the support renders a portion of the support ineffective, thus reducing the

stiffness of the system and increasing the magnitude of deflections. The variable amplitude load causes the crack, if present, to grow, resulting in increased downward deflections. Figure 3 shows a DBN of the system. Variables for the DBN are described in Table 1. The DBN encodes four system states (healthy, damaged support, cracked, damaged support and cracked) by using binary variables to indicate the presence of each fault. Gaussian Process (GP) regressions model the stress intensity factor and crack length, and a linear regression models the deflection of the beam. The parameters of the DBN may be inferred offline using synthetic data.

**2.2. Physics-of-Failure Models**

A key distinction between a system model capable of diagnosis and one capable of prognosis is that a prognostic model estimates the evolution of damage in the future while a diagnosis model only needs the ability to infer the current state of damage. Diagnostic procedures based on fault signatures or pattern recognition are examples of this. While they may be able to detect and isolate damage, quantification can be done using least-squares based estimation, they do not necessarily have any ability to model progressive damage mechanisms such as crack growth, wear, and corrosion. One of the challenges of prognosis is to develop accurate and comprehensive physics-of-failure models. These damage mechanisms are complex, varying with system design and dynamics, and can interact in many ways. The cantilever beam DBN includes physics of failure models for the three faulty system states through two mechanisms. First, the linear regression for predicting beam deflection changes

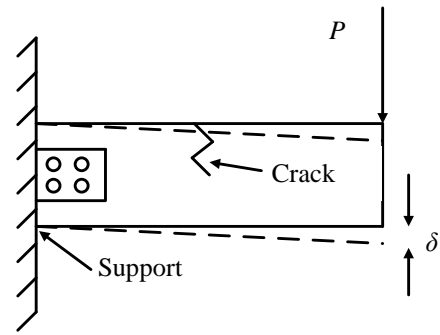


Figure 2. Diagram of cantilever beam.

coefficients depending on the health state of the system. Second, parameters of Paris' law for crack growth as well as crack length are predicted using GP regression. The GP takes the place of direct use of the crack growth law, but the parameters SIF and *a* are still retained within the DBN. The material dependent properties *m* and *C*, while also potential sources of uncertainty, have been taken as constants in this example.

**Diagnosis**

Diagnosis is the process of detecting and isolating damage in a system and quantifying the magnitude of damage. When the probability of a fault occurring crosses the detection threshold, a fault isolation procedure finds fault candidates for further analysis. To isolate candidate faults, statistical

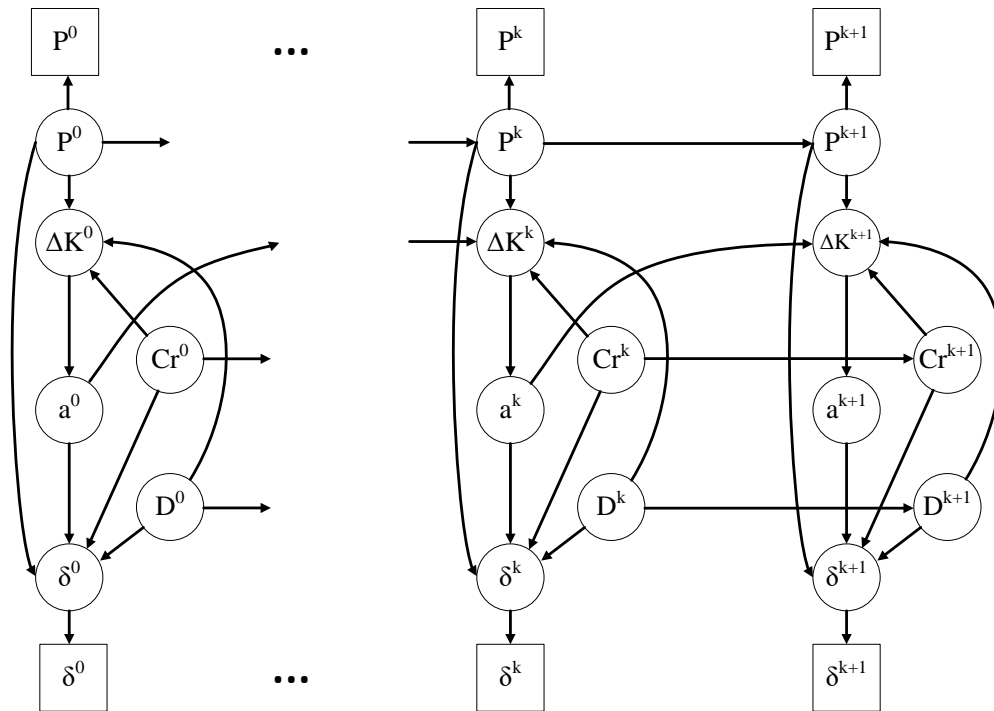


Figure 3. DBN for cantilever beam system. Squares indicate measured values.

Variable	Symbol	Type	Distribution	Role
Damage at support indicator	$D$	Discrete	Binomial (false, true)	Model parameter
Crack indicator	$Cr$	Discrete	Binomial (false, true)	Model parameter
Load	$P$	Continuous	Gaussian	Input (measured)
Stress Intensity Factor	$\Delta K$	Continuous	Gaussian	Model parameter
Crack length	$a$	Continuous	Gaussian	Model parameter
Deflection	$\delta$	Continuous	Gaussian	Output (measured)

Table 1. Cantilever beam DBN (Fig. 3) model variables

inference computes the probability of each fault. The magnitude of the fault may then be estimated.

In the context of prognosis, diagnosis (or more specifically, filtering) provides the initial conditions for prognosis of a UUT. The initial condition for prognosis has a large impact on the accuracy and precision of the RUL distribution. As such, it is important to understand and account for diagnosis uncertainty.

Uncertainty in diagnosis is due to natural variability, measurement error, model error, error due to approximate inference, and any approximations in optimization or least squares procedures used to estimate fault magnitudes. Sankararaman and Mahadevan (Sankararaman & Mahadevan, 2011) propose a methodology for quantifying the uncertainty in diagnosis. This is an integral part of the diagnosis procedure, and it is modified in this paper to accommodate a particle filter (PF) based diagnosis procedure.

Diagnosis can be performed using a DBN model of the system to estimate the state of the system as measurements,  $\mathbf{z}^t$ , become available. The simplest approach is to expand (“unroll”) the two time-slice network (Figs. 3a and 3b) and compute the states of all the unobserved variables in the system,  $\mathbf{x}^t$ , including faults, using a standard inference technique such as the clique tree algorithm (Koller & Friedman, 2009). However, this is generally a computationally intractable problem (Boyen & Koller, 1998). As a result, approximate inference based on Bayesian recursive filtering is pursued.

### 2.2.1. Bayesian Recursive Filtering

The procedure for updating the belief about the system state as new information becomes available is called Bayesian recursive filtering. Bayes’ theorem is the engine for performing the update. Diagnosis of a dynamic system may be achieved by maintaining the joint distribution over the system variables, parameters, and faults and as new noisy

measurements become available via Bayesian recursive filtering. The joint distribution provides the best estimate of whether faults have occurred and what values system parameters and responses may have. This joint distribution is commonly called the belief state  $\sigma^t(\mathbf{x}^t)$ .  $\sigma^t(\mathbf{x}^t) = \mathbf{p}(\mathbf{x}^t | \mathbf{z}^{1:t})$ , where  $\mathbf{p}(\mathbf{x}^t | \mathbf{z}^{1:t})$  is the distribution over the variables  $\mathbf{x}^t$  estimate given all previous measurements  $\mathbf{z}^{1:t}$ . The belief state estimate includes estimates of the states of faults and system parameters, whose states are otherwise unknown. Equation 2, derived from Bayes’ theorem, is the engine for belief state updating:

$$\sigma^{t+1}(\mathbf{x}^{t+1}) = \frac{\mathbf{p}(\mathbf{z}^{t+1} | \mathbf{x}^{t+1}) \mathbf{p}(\mathbf{x}^{t+1} | \mathbf{z}^{1:t})}{\mathbf{p}(\mathbf{z}^{t+1} | \mathbf{z}^{1:t})} \quad (1)$$

where  $\mathbf{p}(\mathbf{z}^{t+1} | \mathbf{x}^{t+1})$  is the likelihood of the measurements,  $\mathbf{p}(\mathbf{x}^{t+1} | \mathbf{z}^{1:t})$  is the prior state estimate at time  $t$ ,  $\mathbf{p}(\mathbf{z}^{t+1} | \mathbf{z}^{1:t})$  is the normalizing constant, and  $\sigma^{t+1}(\mathbf{x}^{t+1})$  is the posterior state estimate at time  $t$ .

Tutorials on Bayesian recursive filtering are available in Koller and Friedman (Koller & Friedman, 2009) and Ristic and Arulampalam (Ristic & Arulampalam, 2004).

### 2.2.2. Particle Filtering

Under certain assumptions, such as when the system is linear and Gaussian, the belief state  $\sigma^{t+1}(\mathbf{x}^{t+1})$  will be of a known parametric form and computationally efficient solutions to the filtering problem (e.g. Kalman filter, extended Kalman filter, unscented Kalman filter) are available. Outside such assumptions, a computationally feasible method for inference in the DBN is found in particle filtering, a form of sequential Monte Carlo based on Bayesian recursive filtering (see e.g. Chen [42]).

Particle filtering is a method for approximating the distribution of the belief state with a set of samples and weights. Common particle filtering methods are based on sequential importance sampling (SIS), which improves upon

the basic sequential MC by weighting point masses (particles) according to their importance sampling density, thus focusing on the samples that affect the posterior belief state the most. A comprehensive tutorial on particle filters is given by Ristic et al. (Ristic & Arulampalam, 2004) and in Koller and Friedman (Koller & Friedman, 2009).

In this paper, an algorithm for systems with multiple operating modes (Andrieu, Davy, & Doucet, 2003) that extends the auxiliary particle filter (APF) (Pitt & Shephard, 1999) is used. This auxiliary particle filter has an auxiliary step where samples  $\tilde{\mathbf{x}}_i^{t+1}$  are drawn based on  $\mathbf{x}_i^t$ . Weights for each particle are then calculated based on  $\mathbf{z}^{t+1}$ . A resampling step, which reproduces particles in proportion to their likelihood (thus eliminating particles of negligible weight), determines a set  $\tilde{\mathbf{x}}_i^t$  that consists of the  $\mathbf{x}_i^t$  that produce the  $\tilde{\mathbf{x}}_i^{t+1}$  with the largest weight.  $\tilde{\mathbf{x}}_i^t$  are then used to sample  $\mathbf{x}_i^{t+1}$ . Resampling may be performed again. Compared to a basic SIS filter with a resampling step, the APF can more closely estimate the true state provided that process noise remains low. Regularization (Ristic & Arulampalam, 2004) is also applied to prevent sample degeneracy.

### 2.2.3. Fault Diagnosis and Diagnosis Uncertainty Quantification

When using a particle filter, the belief state itself provides the information necessary for fault detection, isolation, and damage quantification. The marginal distribution over combinations of the discrete fault indicator variables is a multinomial distribution, whose parameters are easily calculated from the particles representing the current belief state. Given  $m$  fault indicator variables that can take on values of true or false, there are  $n = 2^m$  combinations of faults, including the healthy condition. The  $i^{th}$  combination at the  $t^{th}$  cycle has an expected probability

$$p_i^t = \sum N_i^t w_i^t / N_s \quad (2)$$

The probability of *any* fault (probability of damage) is then

$$p_f^t = \mathbf{1} - p_0^t \quad (3)$$

where  $p_0^t$  is the probability of that no faults occur. When  $p_0^t$  is greater than some threshold, an alert may be issued to a decision maker and a prognosis procedure may be triggered. The remaining  $p_i^t$  ( $i \neq 0$ ) are the isolation probabilities of each fault combination. From the belief state,  $\sigma^{t+1}(\mathbf{x}^{t+1})$ , the marginal distributions over damage parameters may be constructed from the particles and their weights.

The probabilities  $p_i$  that parameterize a multinomial distribution are themselves uncertain and follow a Dirichlet distribution. Based on the Dirichlet distribution, the variance of  $p_k^i$  is

$$\text{Var}[p_t^i] = \frac{N_k^i (N_s - N_k^i)}{N_s^2 (N_s + 1)} \quad (4)$$

The uncertainty in  $p_i$  is directly dependent on the number of samples,  $N_s$ . With the detection and isolation probabilities and their corresponding uncertainties as well as estimates of the distributions of damage parameters known, a decision maker is better able to assess the criticality of damage and the appropriate actions to make to balance safety and cost concerns.

Using a particle filtering algorithm, estimates of the health state of the beam are inferred from measurements of the deflection. These estimates are probabilistic and naturally provide estimates of the SIF and crack length.

### 2.3. Prediction

In probabilistic prognosis, possible future states of the system are generated and the remaining useful life (RUL) distribution,  $r(t)$ , of the particular unit under test (UUT) is estimated. RUL is the amount of time a UUT is usable until corrective action is required and may be measured in hours, minutes, cycles, etc. Measurements are unavailable and the system model is assumed to be valid under future operating conditions. Prediction can be initiated at any time in the life of a system based on the last available state estimate obtained during diagnosis. However, in this paper, the time of prognosis,  $t_p$ , the first time point for which a prognosis estimate is obtained, is after the time of fault detection,  $t_d$ . This ensures that prognosis effort is expended only once a known cause of failure has been identified. Figure 4 illustrates these important prognosis time indices.

One approach to prediction when performing particle filtering on a DBN is a basic sequential Monte Carlo starting with the last available belief state estimate from diagnosis. This belief state estimate is a probabilistic initial condition for prediction that accounts for uncertainty in diagnosis. Starting with the last belief state estimate (with measurements available), particles are recursively sampled through the two time slice DBN until some termination criterion is met, such as  $Pr(r(t) = 0)$  is above some target threshold. Thus, there are  $N_s$  trajectories of the variables of interest beginning at time  $t$ ,  $\{\Phi(t)\}_{i=1}^{N_s}$ . Each trajectory consists of a series of predictions for the variables of interest,  $\Phi(t) = \{\varphi(t|t), \varphi(t+1|t), \dots, \varphi(\text{EoP}|t)\}$ , where the end of prediction (EoP) is the time index of the last prediction before the end of life (EoL) is reached. The EoP depends on frequency and occurs after the end of useful predictions (EoUP), which is the last time index before the required logistics lead time makes corrective active infeasible (and updating of the RUL unnecessary). Particle weights are fixed from the last available measurement, as there is no basis for updating the weights. This results in a particle-based approximation of RUL (similar to the belief state approximation), using the last available set of weights, that

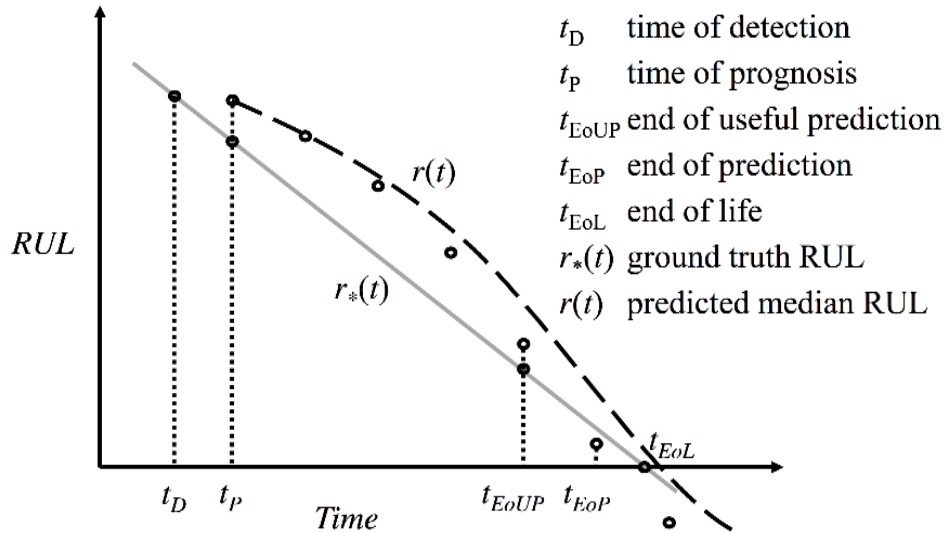


Figure 4. Prognosis time indices:  $r_*(t)$  is the ground truth RUL,  $t_{EoUP}$  is the end of useful prognosis, dashed line depicts median  $r(t)$ .

accounts for uncertainty in diagnosis and prediction. When a new measurement is obtained, a new RUL distribution may be estimated.

To further tailor the prognosis to a particular UUT, the conditional probability distributions in the DBN may be updated as measurements become available. This may be performed via Bayesian updating of the distributions. If a conjugate prior is available, the update can be performed analytically. Otherwise, techniques such as Markov chain Monte Carlo (MCMC) may be required.

The RUL distribution is sensitive to many aspects of the problem. The initial state estimate provided by the diagnosis algorithm must be accurate. As such, the filtering algorithm and number of particles are important algorithmic design decisions. These decisions also involve a tradeoff between accuracy and computational effort, which must be considered. Optimal sensor placement and improved sensor reliability also impact the accuracy of the diagnosis.

The accuracy of predictive models, including those for inputs (loads) and physics-of-failure models, is another large source of uncertainty in the RUL estimate. Because the prediction is recursive with no measurements available to correct the prediction, errors in prediction compound quickly and must be minimized.

Three additional factors that affect the diagnosis (and prognosis) of a system are the *sensitivity* of the system measurements to damage, *noise* in the training and measurement data, and the *sampling rate* i.e. frequency at

which measurements are obtained. If the sensitivity to damage is low and the noise large, diagnosis may be difficult or impossible. Noise in the training and measurement data obscures the true system state and contributes uncertainty to state estimates in diagnosis and prognosis. A low sampling frequency can deteriorate the quality of state estimates in systems that depend on previous states, due to nonlinear dynamic systems as in the actuator system considered in this paper. The state estimates amount to a linearized approximation of the state across the sampling interval (reciprocal of sampling frequency). As sampling rate decreases, the interval becomes wider and the linear approximation worsens. A higher sampling rate decreases the sampling interval and reduces the error due to linearization, thus improving state estimates. In the cantilever beam example, once a crack is diagnosed at  $t_D$ , estimation of the RUL commences. The last filtered system state estimates of the fault states, load, stress intensity factor, crack length, and deflection serve as initial conditions for predictions.

Figure 5 shows a possible distribution of crack length at  $t_D$ . Possible trajectories of these system variables are then simulated, one for each particle, providing in an ensemble of trajectories. Figure 6 shows possible trajectories of the crack length over the next 60 load cycles. A zero value on the horizontal axis indicates the distribution of crack lengths at time  $t_D$  (the distribution of these samples corresponds to Fig. 6). The failure criterion for the beam is its crack length being above a threshold value, indicated by the horizontal red line in Fig. 6. Each particle has a unique crack length and growth trajectory, resulting in a particular number of load cycles until

exceedance of the critical crack length. This results in a distribution of the RUL, depicted as a histogram in Fig. 7, with median RUL also shown.

### 2.3.1. Measurement Gaps

Systems may experience periods of times where measurements are unavailable. This may be a result of the system configuration, availability of measurements, failure of sensing systems, or the desire to have system state estimates at a higher rate than the available measurements. For example, offline inspection data may be available for an aircraft on the ground, while onboard sensing provides a steady stream of information about temperature, altitude, windspeed, pressure, etc. These onboard measurements may only be available for portions of a flight (perhaps during cruising but not takeoff or landing).

Using the same recursive sampling used for RUL estimate, predictions may be produced and used to fill in the information gaps. When a measurement becomes available, the particle filtering algorithm is used to update the last predicted system state. The particle filter update may be performed as long as at least one measurement is available.

### 2.4. Prognosis Validation

Prognosis validation is essential to establish confidence in the RUL estimate. Many sources of uncertainty, including modeling errors, sensor faults, noisy data, and unpredictable loading conditions and operating environments, strongly affect prognosis. Therefore, validation of a prognosis procedure must be done using quantitative performance metrics. These metrics must be carefully chosen, as many issues arise when evaluating prognosis algorithms, such as time scales or the ability to improve accuracy as more measurements are obtained (Saxena et al., 2010). Saxena et al. (Saxena et al., 2010) propose a four-metric hierarchical test to evaluate a prognosis algorithm. This hierarchical test assumes that prognosis will improve as more measurements become available. Together, these four metrics — the prognostic horizon,  $\alpha - \lambda$  accuracy, relative accuracy, and convergence — provide a means for testing and comparing prognostic algorithms.

The first two metrics examine the accuracy of the RUL estimates by determining the probability  $p$  that the RUL estimate is between  $\pm\alpha$  of the ground truth RUL. This probability  $p$  is compared to a threshold value,  $\beta$ . It is desirable for  $p$  to be greater than  $\beta$ . The primary difference between the first two metrics is in how  $\alpha$  is determined, which results in a stricter test for the second metric than the first.

Prognostic horizon (PH) indicates the time at which RUL estimates using a particular prognostic algorithm for a particular system are within acceptable limits. The upper and lower limits are the ground truth RUL plus or minus a

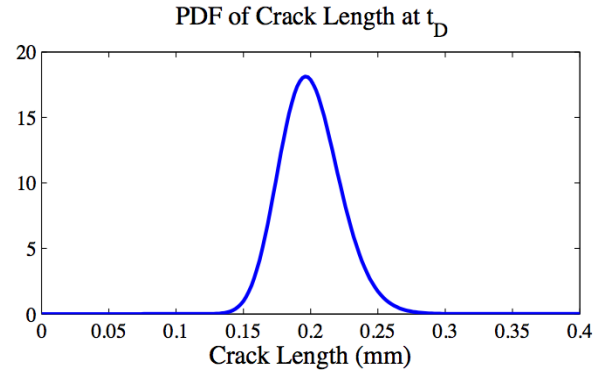


Figure 5. PDF of crack length at time of detection.

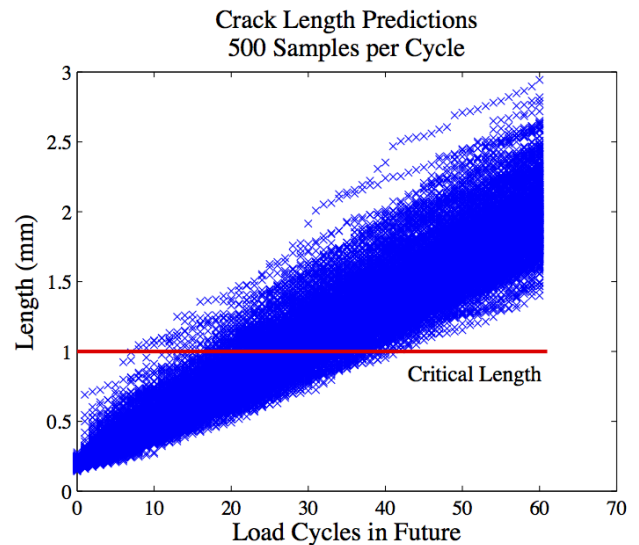


Figure 6. Crack length predictions 60 cycles into the future.

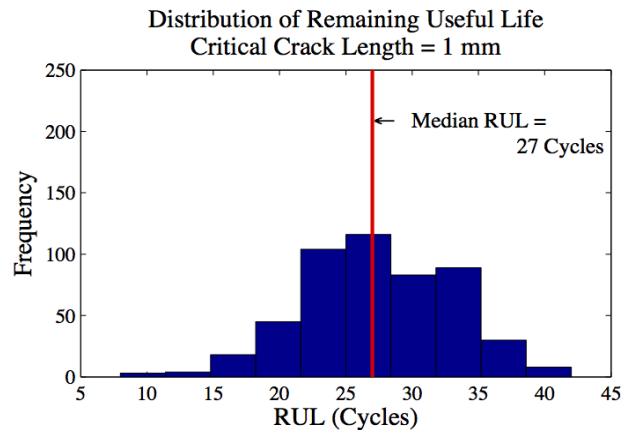


Figure 7. Distribution of remaining useful life.

constant  $\alpha$ , which is some percentage of the EoL value. PH is the difference between the true EoL time and the time when the prognostic algorithm attains this acceptable level of accuracy ( $p > \beta$ ). As this is a validation metric, the true EoL is known. A longer PH is indicative of a better prognostic



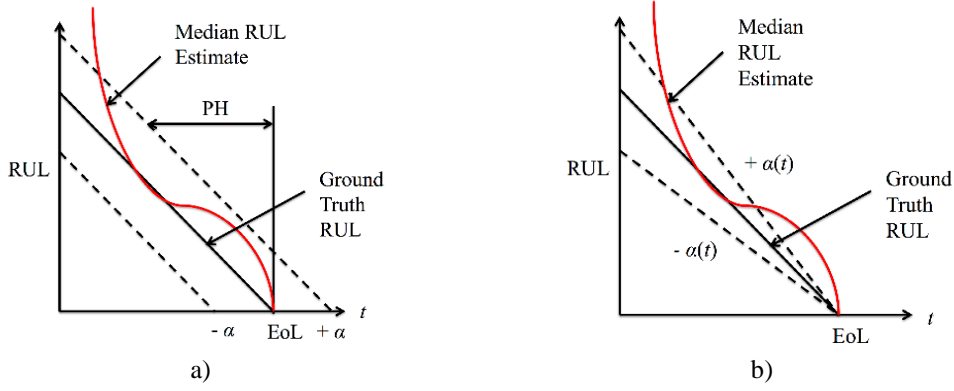


Figure 8. a) Prognostic horizon with  $\pm \alpha$  bounds about the ground truth RUL

b)  $\pm \alpha$  bounds for evaluating  $\alpha$ - $\lambda$  accuracy

algorithm. Figure 8a provides a visual representation of prognostic horizon.

Prognostic horizon maintains upper and lower bounds that are always the same distance from the true RUL. The second validation metric,  $\alpha - \lambda$  accuracy, utilizes a stricter criterion that gradually tightens the limits about the RUL estimate (Figure 8b). Additionally, the accuracy of the RUL is considered at time  $t_\lambda$ , where  $0 \leq \lambda \leq 1$ ,  $t_\lambda = t_p + \lambda(t_{EoL} - t_p)$ , and  $t_p$  is the time at which a prognosis estimate is first obtained. This metric reflects the idea that, as more information is collected about the system, the RUL estimate is expected to improve, and thus the accuracy requirement for the RUL estimate should become more stringent. The  $\alpha - \lambda$  accuracy is equal to 1 when the increasingly stringent accuracy requirements are met, and zero otherwise.

In step three, the relative accuracy (RA) of the prognostic algorithm is calculated. Instead of merely indicating that accuracy requirements have been met, the accuracy of the RUL estimate is quantified. At  $t_\lambda$

$$RA_\lambda = 1 - \frac{|r_*(t_\lambda) - r(t_\lambda)|}{r_*(t_\lambda)} \quad (5)$$

where  $r(t_\lambda)$  is a central tendency point such as the mean or median of the RUL estimate at  $t_\lambda$  and  $r_*(t_\lambda)$  is the ground truth RUL. The RA is computed separately for each time step at which RUL is estimated. RA is a value between 0 and 1, and values closer to 1 indicate better accuracy.

Finally, convergence measures how quickly a metric, such as RA, improves with time. It is determined by the distance from the centroid of the area under the curve in the interval  $[t_p, t_{EoUP}]$  of a prognosis metric ( $M$ ) located at  $(x_c, y_c)$ , to the origin, located at  $(0, t_p)$ . A high rate of convergence is

desirable and leads to a larger PH. The convergence measure  $C_M$  of a prognosis metric  $M$  is defined as

$$C_M = [(x_c - t_p)^2 + y_c^2]^{1/2} \quad (6)$$

where

$$x_c = \frac{\frac{1}{2} \sum_{i=P}^{EoUP} (t_{i+1}^2 - t_i^2) M(t_i)}{\sum_{i=P}^{EoUP} (t_{i+1} - t_i) M(t_i)} \quad (7)$$

and

$$y_c = \frac{\frac{1}{2} \sum_{i=P}^{EoUP} (t_{i+1}^2 - t_i^2) M(t_i)^2}{\sum_{i=P}^{EoUP} (t_{i+1} - t_i) M(t_i)} \quad (8)$$

where  $M(t_i)$  is the non-negative prediction accuracy, EoUP is the end of useful prediction, and  $P$  is the time at which the prognosis algorithm makes its first prediction. End of useful prediction is the time after which corrective action cannot be performed before EoL. A high rate of convergence is better and leads to a larger PH.

Using ground truth data, the four validation metrics indicate the overall suitability for prognosis of the system model represented by the DBN and the Monte Carlo inference algorithm. This subsumes many design decisions for the system model, including which variables to include or exclude and their conditional probability distributions, graph structure of the network, training point selection, frequency of updates, number of samples to use in inference, etc. In particular, the numerical example in Section 3 investigates the sensitivity of the validation metrics to sampling rate.

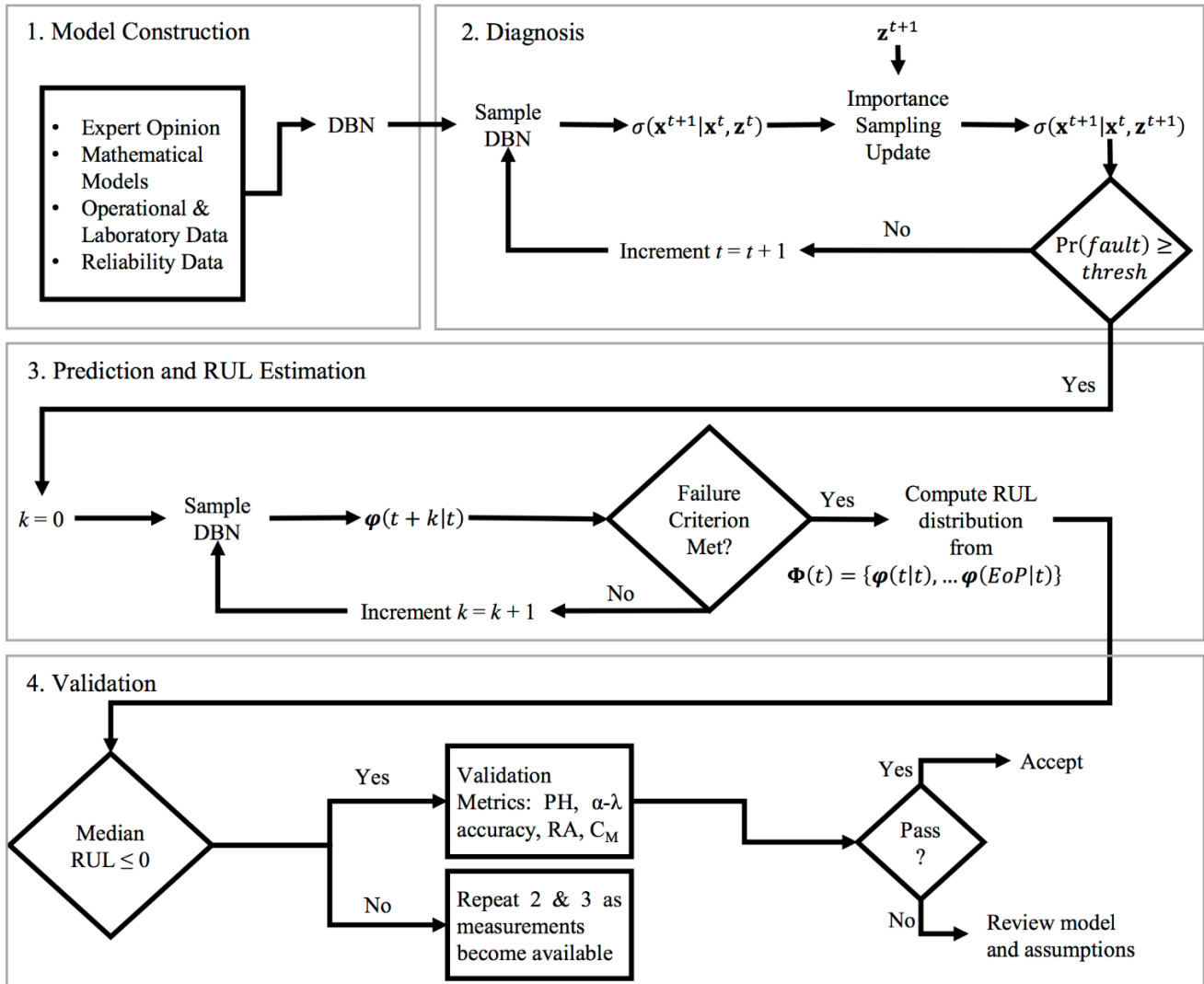


Figure 9. Diagram of proposed prognosis methodology

## 2.5. Summary of Proposed Prognosis Framework

This section presented a framework for probabilistic prognosis, and illustrated portions of the framework with a cantilever beam example. Figure 9 summarizes the framework. DBNs are used as a system modeling paradigm due to their ability to handle uncertainty and to integrate many types of information, both in the offline model construction phase and the online belief state updating phase. For prognosis, it is of particular importance to model complex physics-of-failure phenomena and integrate such models into the DBN. After the DBN model is established, the model is used for tracking the state of a particular UUT. Particle filtering is used to update the belief state as new measurements are obtained. Uncertainty in the state estimate (diagnosis) is quantified, and when a fault is detected,

estimation of RUL via recursive prediction begins. The result is an estimate of the distribution of RUL. Section 2.4 considers the situation when there are gaps in the availability of measurements.

When a prognosis procedure (DBN model of system combined with available measurements and filtering algorithm), is designed for a particular system, it is then validated using the 4 step hierarchical procedure outlined in Section 2.5.

## 3. COMPUTATIONAL EXAMPLE

A hydraulic actuator system was considered to demonstrate the proposed methodology. Such a system is often used to manipulate the control surfaces of aircraft. The system consists primarily of three subsystems: a hydraulic actuator, critical center spool valve, and an axial piston pump (Fig. 10).

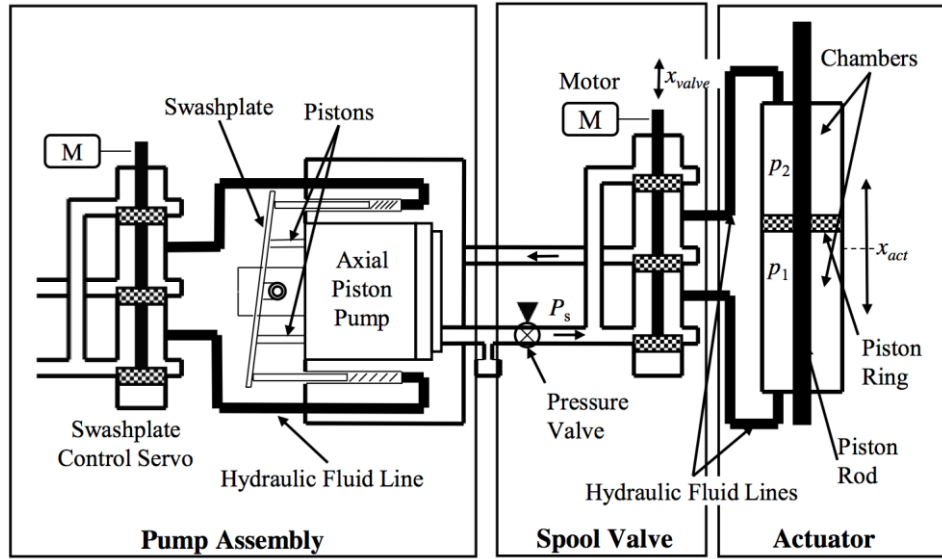


Figure 10. Hydraulic actuator system

The servo-controlled piston pump provides the steady hydraulic pressure to the servo (spool valve) controlled actuator. The position of the spool valve is adjusted via an electric motor. The position of the actuator changes based upon the flow of hydraulic fluid in and out of its two chambers. The control signal to the spool valve motor is typically a result of feedback about the actuator's current displacement versus the desired displacement. Expert opinion, reliability data, mathematical models, operational data, and laboratory data were used to construct a DBN model of the spool valve and hydraulic subsystems.

First, expert opinion is invoked to determine the scope of the problem, variables and faults to model, and establish the DBN structure. Next, reliability data is drawn upon to determine the conditional probabilities for the faults. The mathematical model of the system is used to generate predictions of the system variables. The predictions are treated similar to operational and laboratory data and used to train a regression model for estimating the reduction in seal orifice area, which is equivalent to the seal leakage area.

### 3.1. DBN Model Construction

#### 3.1.1. Expert Opinion

Expert opinion was considered first to define the basic parameters of the problem. Seven state variables and six discrete faults were selected to model the behavior of the system.

A generic initial structure for the DBN is first selected (Fig. 11) based on expert opinion. This generic two time slice

structure consists of the set of faults,  $F$ , model parameters,  $\theta$ , system state,  $x$ , and measurements,  $z$ . In this structure, faults cause changes in system parameters, which then cause changes in system responses, which are observed.  $F$  contains the persistent variables in the DBN – their future values depend upon their present values. The observations,  $z$ , while not connected across time slices, are nonetheless *not* independent across time slices, but correlated via  $\theta$ .

Table 2 lists the faults considered in the actuator system and the parameter affected by that fault (the faults are described further in Section 3.1). For each fault, a binary variable that indicates the presence of the fault is added to the network at time  $t$  and  $t + 1$ . Similarly, a Gaussian variable is added at time  $t$  and  $t + 1$  for each affected parameter. Links are drawn

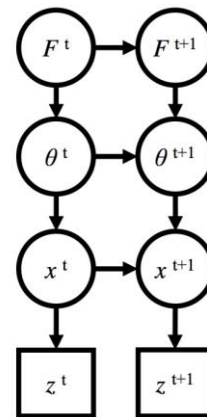


Figure 11. Generic DBN structure

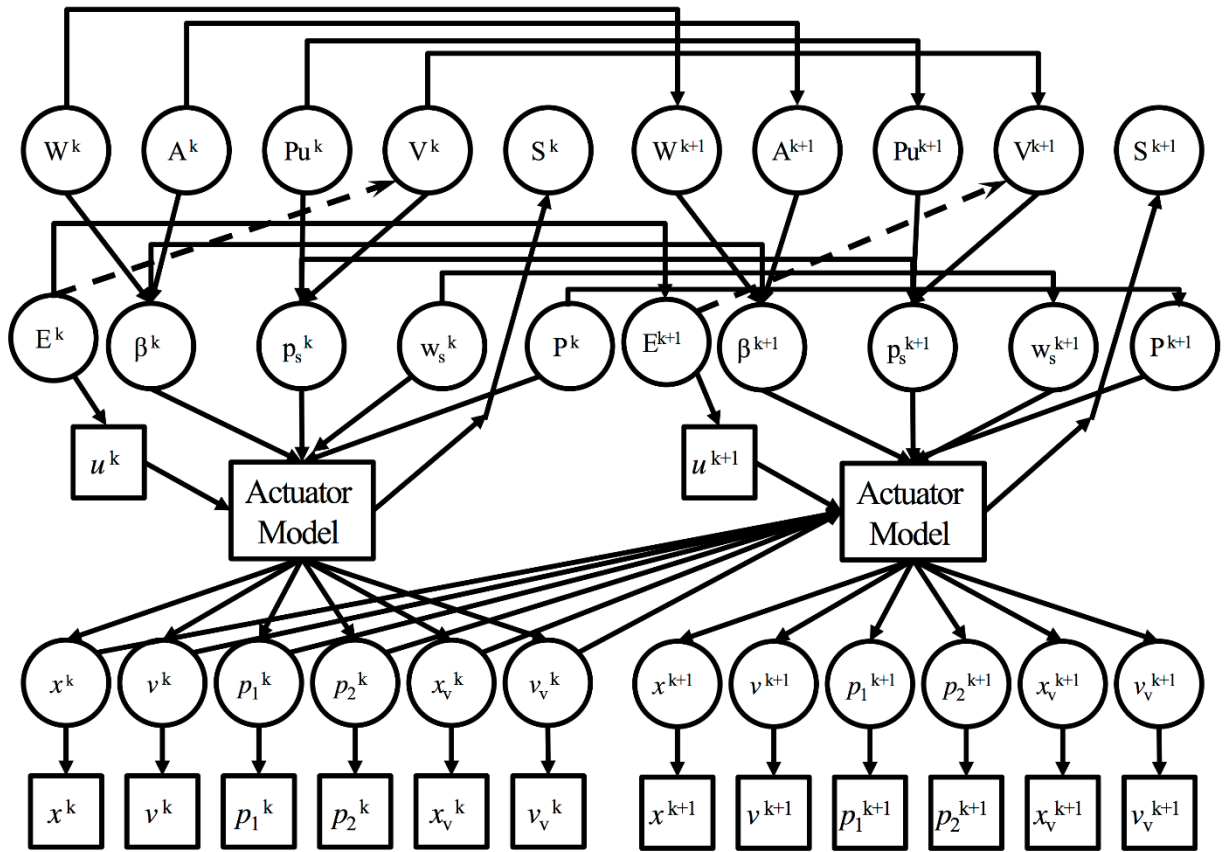


Figure 12. DBN structure (solid lines – based on expert opinion, dotted lines – learned from data)

State Variable	Symbol	Unit	Type	Note
Actuator position	$x_{act}$	m	continuous	measured
Actuator velocity	$v_{act}$	m/s	continuous	measured
Pressure in chamber 1	$P_1$	Pa	continuous	measured
Pressure in chamber 2	$P_2$	Pa	continuous	measured
Valve position	$x_{valve}$	m	continuous	measured
Valve velocity	$v_{valve}$	m/s	continuous	measured
Control signal	$u$	V	continuous	input
Water leak	$W$	-	binary	inferred
Air leak	$A$	-	binary	inferred
Pump sensor fault	$P$	-	binary	inferred
Valve fault	$V$	-	binary	inferred
Seal leak	$S$	-	binary	inferred
Electrical Fault	$E$	-	binary	inferred
Fluid bulk modulus	$\beta$	MPa	continuous	inferred
Supply pressure	$p_s$	MPa	continuous	inferred
Wear rate	$w_s$	$Mm^3/Nm$	continuous	inferred

Table 2. DBN variables

pointing from faults to affected parameters. The parameters are assumed to have Gaussian distributions, whose mean and variance depend on the health state of the system. The leakage area parameter is a special case, as wear and leakage are assumed to be present at all times. The discrete variable for leakage area indicates that the leakage area has increased beyond some threshold value.

Parameters from the current time step and initial conditions from the previous time step are input into a physics-based model of the actuator, which estimates the system responses, assumed to be Gaussian variables. Measurements are then connected to the corresponding system response. Links are also drawn between like faults at time  $t$  and  $t + 1$  and like parameters at time  $t$  and  $t + 1$ . Finally, a Gaussian variable is added at time  $t$  and  $t + 1$  for each measurement available. The resultant DBN is shown in Fig. 12 with parameters described in Table 2.

### 3.1.2. Published Reliability Data

The DBN model of the system should be able to simulate multiple faults and predict system behavior multiple steps into the future for the model to be a useful diagnosis and prognosis tool. The overall failure rate for an actuator may be determined by estimating the base failure rate and making empirical corrections for temperature and fluid contamination (Naval Surface Warfare Center, 2011). The RIAC Databook (*RIAC Automated Databook*, 2006) and the Handbook of Reliability Prediction Procedures for Mechanical Equipment (HRPPME) (Naval Surface Warfare Center, 2011) give failure rates for many mechanical systems. For illustration of the methodology, a handful of the possible faults for the actuator system are considered in this

paper. Table 3 lists the faults that have been considered, the subsystem where they are located, and the parameters affected by those faults.

Failure rates from the literature were then used to calculate the probability of each fault occurring. These probabilities correspond to parameters of the discrete fault indicator variables in the DBN.

### 3.1.3. Mathematical Behavior Models

Several mathematical models are used in this example. A physics-based model of a hydraulic actuator, described by Kulakowski et al. (Kulakowski, Gardner, & Shearer, 2007) and Thompson et al. (Thompson, Prun, & Shukla, 1999) with Eqs. 9-19, is integrated into the DBN through a deterministic conditional probability distribution (Koller & Friedman, 2009). This model has been implemented in the Matlab Simulink environment. The model is run recursively in 0.25 second increments (4 samples per second). Each increment has a constant load value determined by a load model. When an abrupt fault occurs, it is programmed into the model at the beginning of the load interval for which it occurs.

The actuator is shown in Fig. 10. The parameters and variables for this system are described in Table 4. The relationships among these quantities are as follows:

$$\dot{x}_{act} = v_{act} \quad (9)$$

$$\dot{v}_{act} = \frac{1}{m} [(P_1 - P_2)A_{pist} - b_{act}v_{act} \dots - k_{act}x_{act} - F_{ext}] \quad (10)$$

Fault	Parameter Affected	Subsystem	Failure Rate	Data Source
Seal Leak	Leakage Area	Actuator	0.249 failures / $10^6$ hours	RIAC Databook [47], Literature [48]
Water Leak into System	Hydraulic Fluid Bulk Modulus	Piping/Fittings	2.736 failures / $10^6$ hours	RIAC Databook [47], Literature [48]
Air Leak Into System	Hydraulic Fluid Bulk Modulus	Piping/Fittings	0.167 failures / $10^6$ hours	RIAC Databook [47], Literature [48]
Pressure Valve Malfunction	Supply Pressure	Pressure Valve	15.575 failures / $10^6$ hours	RIAC Databook [47], Literature [48]
Pump Pressure Sensor Fault	Supply Pressure	Piston Pump	14.3 failures / $10^6$ hours	HRPPME[38]
Electrical Fault	Control Signal	Electrical	1.510 failures / $10^6$ hours	RIAC Databook [47], HRPPME [38]

Table 3. List of faults and affected parameters and subsystems

Parameter/variable	Symbol	Nominal Value/ Unit
Actuator position	$x_{act}$	m
Actuator velocity	$v_{act}$	m/s
Servovalve position	$x_{valve}$	m
Servovalve velocity	$v_{valve}$	m/s
Pressure in chamber 1	$P_1$	Pa
Pressure in chamber	$P_2$	Pa
Combined mass of actuator and load	$m_{act}$	40 kg
Combined damping of actuator and load	$b_{act}$	800 Ns/m
Combined stiffness of actuator and load	$k_{act}$	$10^6$ N/m
Piston annulus area	$A_{pist}$	0.0075 m <sup>2</sup>
Valve port width	$w_{valve}$	0.0025 m
Spool valve model coefficients	$b_0$	90 m/Vs <sup>2</sup>
	$a_0$	360,000 1/s <sup>2</sup>
	$a_1$	1/s
Hydraulic fluid bulk modulus	$\beta$	1000 MPa
Hydraulic fluid density	$\rho$	847 kg/m <sup>3</sup>
Discharge coefficient	$C_d$	0.7
Supply pressure	$P_{supply}$	20 MPa
Chamber 1 volume	$V_1$	m <sup>3</sup>
Chamber 2 volume	$V_2$	m <sup>3</sup>
Chamber 1 fluid capacitance	$C_{f1}$	m <sup>3</sup> /(kg/s)
Chamber 2 fluid capacitance	$C_{f2}$	m <sup>3</sup> s/(kg/s)
Volumetric flow rate into chamber 1	$Q_1$	m <sup>3</sup> /s
Volumetric flow rate out of chamber 2	$Q_2$	m <sup>3</sup> /s
Externally applied force	$F_{ext}$	0 N
Input voltage	$e_{command}$	Sin(2*pi*t) V
Leakage volumetric flow rate	$Q_{leak}$	0 m <sup>3</sup> /s
Leakage area	$a_{leak}$	0 m <sup>2</sup>

Table 4. Model parameters and variables for a spool valve and hydraulic actuator

$$\dot{P}_1 = \frac{1}{C_{f1}} (Q_1 - A_{pist} Q_2 + Q_{leak}) \quad (11)$$

$$\dot{P}_2 = \frac{1}{C_{f2}} (V_2 \dot{x}_{act} - Q_2 - Q_{leak}) \quad (12)$$

$$\dot{x}_{valve} = v_{valve} \quad (13)$$

$$\dot{v}_{valve} = a_1 v_{valve} + a_0 x_{valve} + b_0 e_{command} \quad (14)$$

$$C_{f1} = \frac{V_1(x_{act})}{\beta} \quad (15)$$

$$C_{f2} = \frac{V_2(x_{act})}{\beta} \quad (16)$$

If  $x_{valve} > 0$ ,

$$\left\{ \begin{array}{l} Q_1 = C_d w_{valve} x_{valve} \text{sign}(P_s - P_1) \sqrt{\frac{2}{\rho} |P_s - P_1|} \\ Q_2 = C_d w_{valve} x_{valve} \sqrt{\frac{2}{\rho} (P_2)} \end{array} \right. \quad (17)$$

If  $x_{valve} < 0$ ,

$$\left\{ \begin{array}{l} Q_1 = C_d w_{valve} x_{valve} \sqrt{\frac{2}{\rho} (P_1)} \\ Q_2 = C_d w_{valve} x_{valve} \text{sign}(P_s - P_2) \sqrt{\frac{2}{\rho} |P_s - P_2|} \end{array} \right. \quad (18)$$

$$Q_{leak} = C_d a_{leak} \sqrt{\frac{2}{\rho} |P_2 - P_1|} \text{sign}(P_2 - P_1) \quad (19)$$

For demonstration of the prognosis methodology, the load is simulated using an ARIMA (autoregressive integrated moving average) model. In reality the load on a flight control actuator is depends on many variables related to the dynamics of the aircraft and the desired flight path (see Mahulkar et al. (Mahulkar, McGinnis, Derriso, & Adams, 2010), Karpenko and Sepeheri (Karpenko & Sepehri, 2003), and McCormick (MacCormick, 1995)).

Finally, the physics-of-failure model for the seal leak is considered. The seal leakage volume is modeled as in Section 2.2. The wear rate is treated as a random variable whose distribution is learned from laboratory data. The volume of material removed is divided by the contact length of the seal (Fig. 13) to arrive at the leakage area required for use in Eqs. 9-19.

Generally, seal leakage is due to wear caused by friction between the seal and piston, which removes seal material and allows fluid to pass between the chambers of the actuator. The cross section of the actuator in Fig. 10 is shown in Fig. 14. The surface area of the seal is  $(r_2^2 - r_1^2)$ , where  $r_1$  and  $r_2$  are the inner and outer radii of the seal, respectively.

The volume of material removed from the seal per cycle depends on the friction force and sliding distance per cycle and may be calculated by

$$V(t) = w_s(t)F \quad (20)$$

where  $w_s$  is the wear rate of the seal in  $\text{mm}^3/\text{N}/\text{m}$ ,  $F$  is the frictional force on the seal,  $d$  is the total sliding distance, and  $t$  refers to the load cycle.

For the seal shown in Fig. 13, where  $L$  is the contact length of the seal and  $P$  is pressure, the leakage area (considered in Eqs. 16-29 in Thompson et al. (Thompson et al., 1999)) based on the volume of material removed is assumed to be  $a_{leak} = V(t)/L$ .

While wear is a continuous process, in this paper the occurrence of wear is modeled as a binary event, where modeling begins when the leakage area has reached a value that has detectable effects. The occurrence is modeled using an empirically derived seal failure rate, which modifies an experimentally determined base failure rate for the seal. Details of deriving the failure rate are available in (Naval Surface Warfare Center, 2011).

The wear rate itself varies with factors such as the age of the seal, temperature, contaminants in the fluid. The load experienced by the seal also varies as does the velocity of the actuator. As a result, the volume of material removed and leakage area are nonlinear functions of seal age, temperature, and contaminant concentration. However, for the sake of demonstration, it is assumed that the wear rate is steady, which is possible outside of the initial wear-in phase and under constant environmental conditions.

### 3.1.4. Operational and Laboratory Data

Operational and laboratory data appear as historical databases and online measurement data. The data is organized in a time-lagged database as in Bartram and Mahadevan (G. Bartram & Mahadevan, S., 2013). This format facilitates structure and parameter learning in the DBN. For the actuator model, the values of system

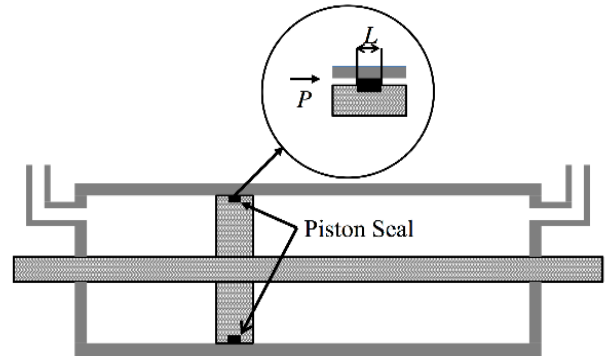


Figure 13. Hydraulic actuator diagram showing dynamic seals

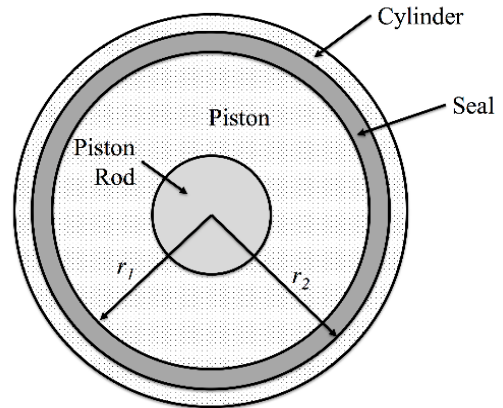


Figure 14. Actuator cross section

parameters  $w_s$ ,  $\beta$ ,  $p_s$  and  $a_{leak}$  are estimated. Because the data is synthetic, 20:1 signal to noise ratio (SNR) Gaussian white noise was added to all system response and load data to mimic measurement error. Further, as operational data becomes available during diagnosis, that data is used to update the distributions of the actuator parameters and responses as well as to estimate the parameters of the ARIMA model used in load history simulation.

### 3.2. Diagnosis

Diagnosis is performed as described in Section 2.3. The actuator was operated for 20 seconds with a leak occurring after 6 seconds. At this point, the system has already reached the steady state. Synthetic measurement data was generated using the Simulink model for two cases. In the first case, the data was not resampled (i.e. the sampling frequency remained

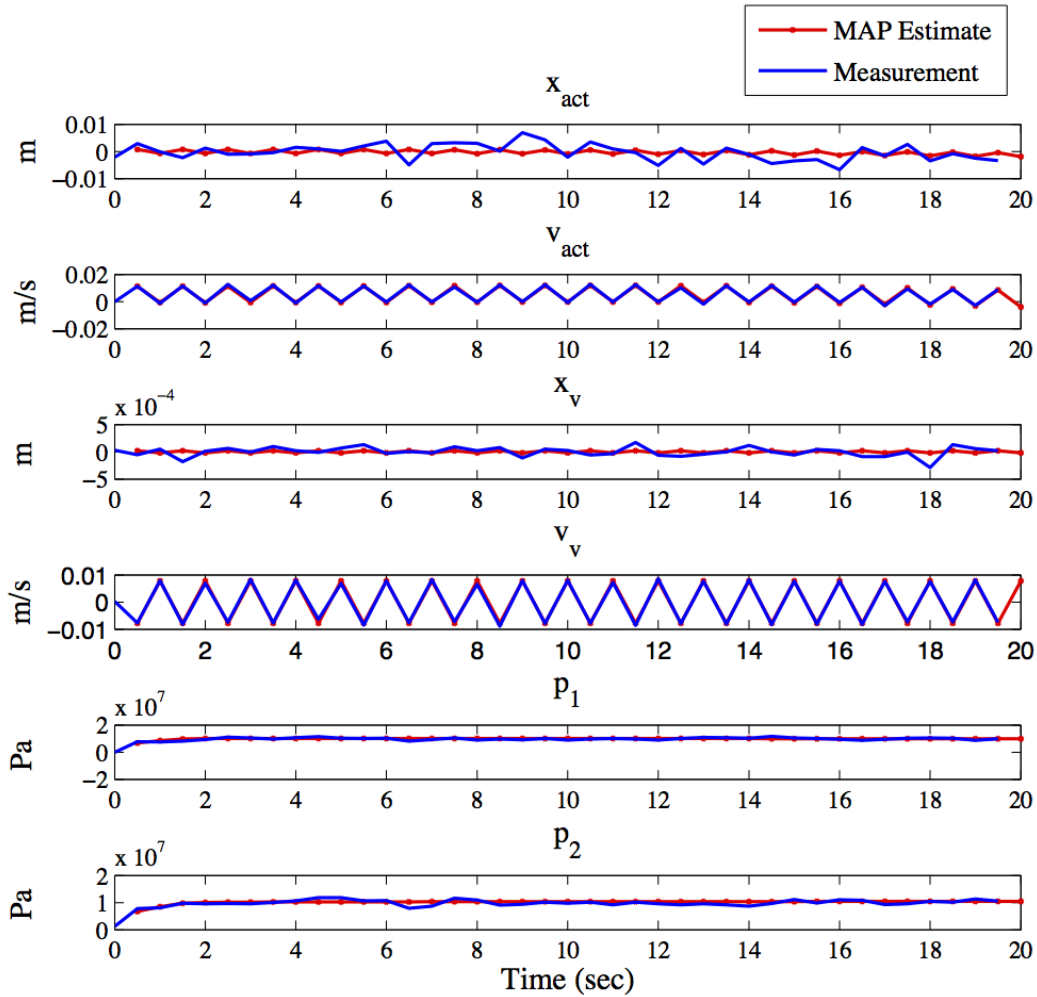


Figure 15. MAP estimates and measured values of actuator position and velocity, servovalve position and velocity, and pressure in each actuator chamber (2 samples/sec)

4 samples per second). In the second case, data was resampled at 2 samples per second. This was done to compare the effect of sampling rate on diagnosis and prognosis. The system responses and load were assumed to be measurable while the system parameters including wear rate and leakage area were assumed to be unobservable. Inference via particle filtering ( $N_s = 250$ ) was performed on the DBN to obtain filtered estimates of the system state.

After obtaining the state estimate at cycle  $t$ , the probability of damage was calculated as in Section 2.2. If the probability of damage exceeded 95%, an alarm was triggered. The fault was then isolated and quantified. Figure 15 shows maximum a posteriori (MAP) estimates of the system responses against their measured values for the 2 samples/sec series (not shown are responses for the 4 samples/sec series that appear similar). It is seen that the MAP system responses track the measured

values closely. Figure 16 shows the MAP estimates of the system parameters for the 2 samples/sec series, including the leakage area, and load against the ground truth values. This figure shows how the leakage area changes with time and how well the filtering procedure can infer the value of the leakage area. The system responses in Fig. 15 are sensitive to changes in the supply pressure and leakage area. However, experimentation with the model revealed that these responses were insensitive to changes in the fluid bulk modulus. Changes in bulk modulus, however, may result in effects such as changes in wear rate that have not been included in the ground truth model. In both Fig. 15 and Fig. 16, the good estimates may be attributed to the use of an accurate physics-based model, but also to the use of synthetic measurement data whose sampling rate matched the internal sampling rate of the generative model. Thus, the performance of filtering is favorably biased.



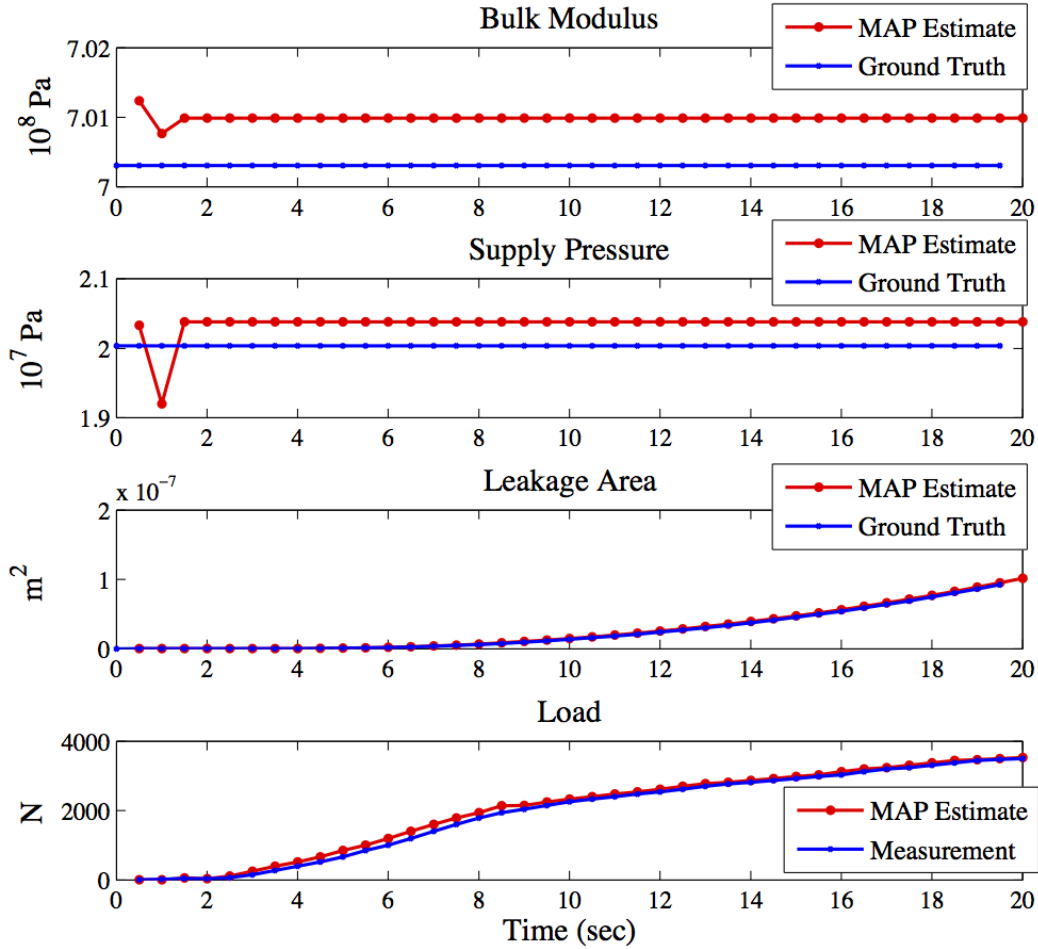


Figure 16. MAP estimates of system parameters and load with ground truth and measured values (2 samples/sec)

### 3.3. Diagnosis Uncertainty

Diagnosis uncertainty was quantified during filtering as in Section 2.3. Figure 17 shows the damage probability as it evolves with time for the 2 samples/sec case. The damage probability passes the alarm threshold 0.5 seconds after the actual occurrence of the fault.

At the time of detection ( $t = 18$  cycles), all particles for the 2 samples/sec series have a value of  $1.038\text{E-}8$  for the leakage area. This is in contrast to a ground truth value of  $9.339\text{E-}9$ . Note that all particles have the same value due to resampling. These leakage area values are the initial condition for prognosis starting at  $t = 18$ . Figure 17 may be used to signal an alarm that a fault has occurred. In some cases, this alarm can be used to initiate RUL estimation or other tasks such as inspection.

### 3.4. Prediction

After alerting the presence of a leak above the threshold of  $7.578\text{E-}9 \text{ m}^2$ , estimation of the RUL distribution was performed as in Section 2.4. The RUL distribution assumes a failure threshold for leakage area of  $4.123\text{E-}8 \text{ m}^2$ . The RUL distribution at the 25<sup>th</sup> cycle is shown in Fig. 18 for the 2 samples/sec series.

### 3.5. Computational Effort

The above diagnosis was performed with 12 Intel Nehalem processor cores working in parallel. Each belief state update in diagnosis takes approximately 30 seconds. Similarly, each prediction step in prognosis takes approximately 30 seconds. The majority of the computational effort is directed towards solving the differential equations for the actuator in the Matlab Simulink environment.

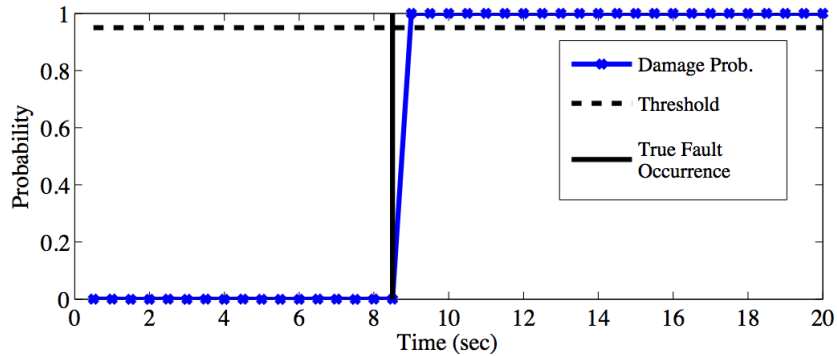
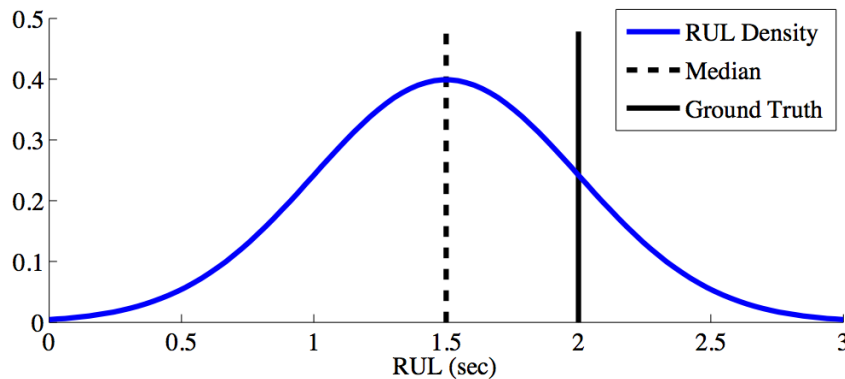


Figure 17. Damage probability with actual fault time (2 samples/sec)

Figure 18. RUL density estimate at  $t = 25$  sec (2 samples/sec)

### 3.6. Prognosis Validation

By continuing to estimate the new RUL distribution as new measurements become available, the performance of the prognostic algorithm may be evaluated. In Fig. 19, median RUL estimates are plotted against the ground truth RUL with  $\pm \alpha$  bounds for the 2 samples/sec series. The  $\pm \alpha$  bounds are selected to be  $\pm 15\%$  of the ground truth EoL about the current ground truth RUL. Figure 20 indicates whether the probability of the RUL estimate being between the  $\pm \alpha$  bounds at a particular time is greater than a threshold value, taken as  $\beta = 0.6$ . From this information, it is also determined that the prognostic horizon (PH) is about 1.25 seconds. To determine the PH, the first time the probability mass of the

RUL distribution lying between the  $\pm \alpha$  bounds exceeds and remains above the threshold of  $\beta = 0.6$  is at  $t = 13.5$  sec is determined. Because the EoL is  $t = 14.75$  sec.,  $PH = 14.75 - 13.5 = 1.25$  seconds. Similarly, Fig. 21 and Fig. 22 show the same plots for the 4 samples/sec series and have a PH of 6.9 seconds. The disparity in PH may be attributed to the sampling rate, as all else is the same.

$\pm \alpha$  bounds that narrow as the EoL approaches are considered for both sampling rates in Fig. 23 and Fig. 24 for  $\lambda = 0.5$  and  $\alpha = 0.20$ .  $\lambda = 0.5$  considers the accuracy of the RUL estimate halfway between the time of prognosis and end of life, termed  $t_\lambda$ . The bottom plots of Fig. 23 and Fig. 24 show the  $\lambda$ - $\alpha$  accuracy, which is a binary value that indicates whether the probability of the RUL estimate being between

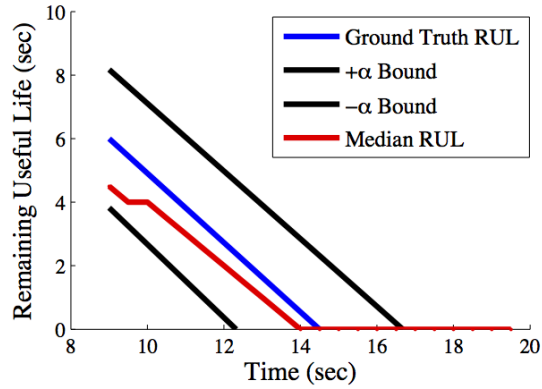


Figure 19. Ground truth RUL, median RUL, and  $\alpha$  bounds with  $\alpha = 0.15$  (2 samples/sec)

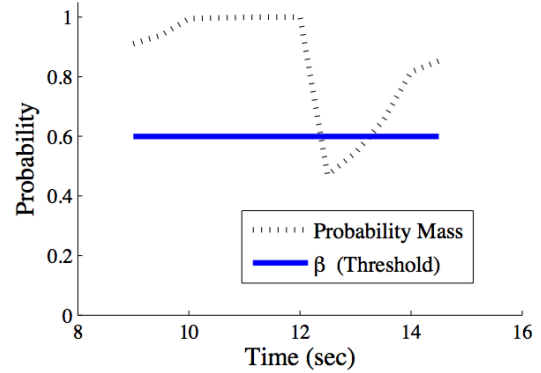


Figure 20. Probability that RUL is within  $\alpha$  bounds with  $\alpha = 0.15$  (2 samples/sec)

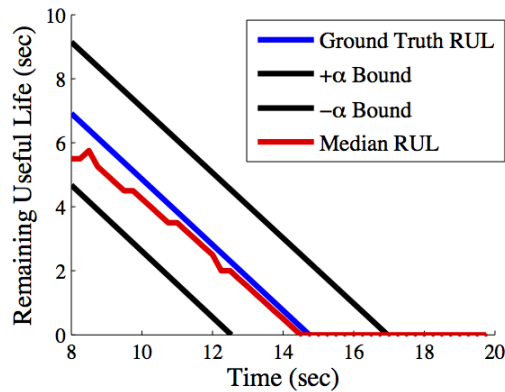


Figure 21. Ground truth RUL, median RUL, and  $\alpha$  bounds with  $\alpha = 0.15$  (4 samples/sec)

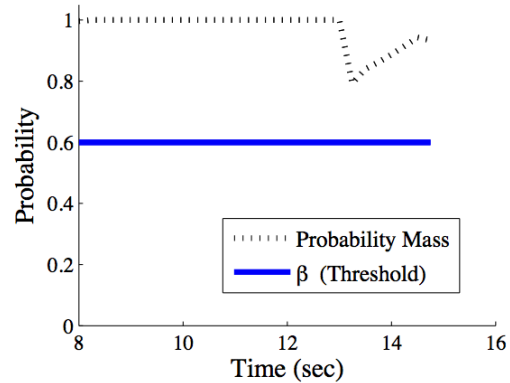


Figure 22. Probability that RUL is within  $\alpha$  bounds with  $\alpha = 0.15$  (4 samples/sec)

the  $\pm \alpha$  bounds at a particular time is greater than a threshold value, taken as 0.8 here. The  $\lambda$ - $\alpha$  accuracy is mostly zero for the 2 sample/sec series, indicating that the RUL estimate is too diffuse to pass this test. However, the 4 sample/sec series performs reasonably well, although it reaches EoL prematurely. The higher sampling rate results in much better performance on this metric. Even at 4 samples/sec, performance is not ideal. This is likely due to the limited number of samples (250) used in particle filtering.

Figure 25 and Fig. 26 show the relative accuracy of the RUL density estimate based on the median RUL value, and shows that the median values are accurate. The 2 sample/sec series again lags the 4 sample/sec series in performance. Based on the relative accuracy and excluding any zero values, the convergence is estimated to be 19.95 for the 4 sample/sec series. Convergence is not calculated for the 2 sample/sec

series because it fails on  $\lambda$ - $\alpha$  accuracy. When comparing prognostic algorithms, larger convergence values are desirable.

### 3.7. Discussion

The DBN-based methodology successfully integrates heterogeneous sources of information to diagnose the system and estimate RUL. Particle-filter based inference provides a seamless method for switching between probabilistic diagnosis and prediction while facilitating uncertainty quantification.

The prognosis validation results indicate that the methodology provides reasonable median estimates of RUL, even as the RUL density estimates are sometimes diffuse. The sampling rate of the measurements is a large factor in whether or not prognosis is successful. Inclusion of

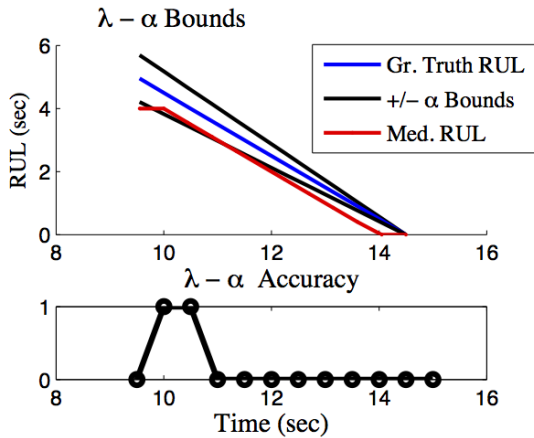


Figure 23. Bounds used for calculating  $\lambda$ - $\alpha$  accuracy with  $\alpha = 0.15$  (top) and  $\lambda$ - $\alpha$  accuracy (bottom) for (2 samples/sec)

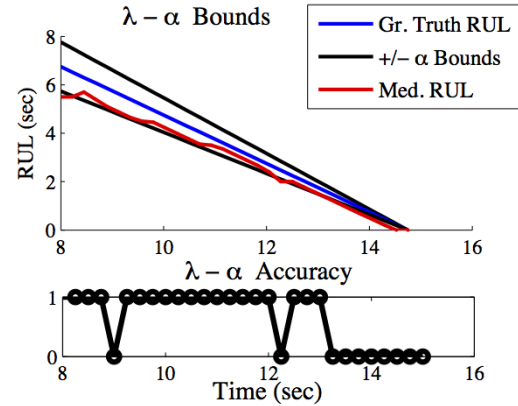


Figure 24. Bounds used for calculating  $\lambda$ - $\alpha$  accuracy with  $\alpha = 0.15$  (top) and  $\lambda$ - $\alpha$  accuracy (bottom) for (4 samples/sec)

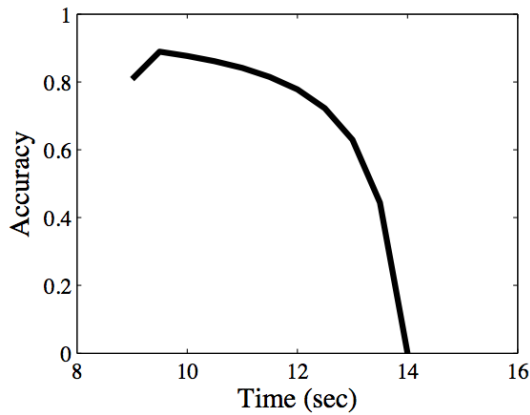


Figure 25. Relative accuracy based on median RUL estimate 2 samples/sec

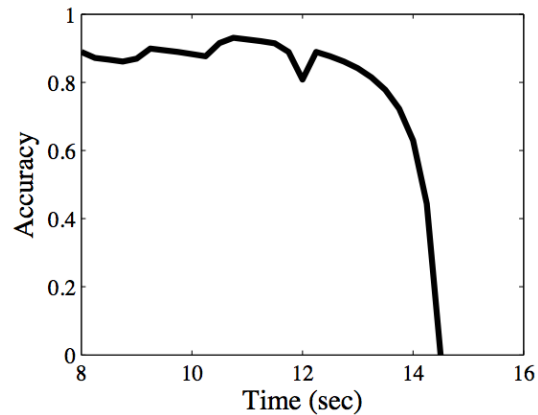


Figure 26. Relative accuracy based on median RUL estimate 4 samples/sec

inspection data may reduce the uncertainty in the leak area estimate and thus the RUL estimate. The accuracy of prognosis, of course, will vary depending on the system, available information, measurement noise, loading conditions, and environmental conditions.

Computational effort is a persistent issue in particle-based methodologies, affected by the complexity of the system, models involved, simplifying assumptions, filtering algorithms, etc. The prognosis methodology described in this paper is flexible with respect to these decisions, so computational effort will vary. Reductions in computational effort may be achieved by using reduced-order models (e.g. linearized model of actuator), feature selection and

dimensional reduction techniques, and improved particle filtering techniques (e.g. Rao-Blackwellized particle filter (Koller & Friedman, 2009)). Additionally, the ability to massively parallelize a particle filter using modern central processing units (CPUs) and graphics processing units (GPUs) provides the potential for greatly decreased computation times. The actuator problem would see immediate decreases in computational time with additional processor cores.

Thus far, the methodology has only been demonstrated using synthetic data, and needs to be tested further using real-world data. Further, more complex physics-of-failure models should be considered.

#### 4. CONCLUSION

A methodology for DBN-based probabilistic prognosis that integrates heterogeneous information sources and diagnosis uncertainty is presented in this paper. First, expert opinion is used to establish the system definition and basic assumptions. Available reliability data is used to calculate conditional probabilities for fault indicator variables for damage at the support and a crack. Operational and laboratory data are organized in a time-lagged database and used for estimating actuator parameters including the wear rate. The system DBN model is used in online diagnosis via particle filter-based inference. Two sampling rates are considered to show the effects of sampling rate on prognosis. The particle-based state distribution also provides information on diagnosis uncertainty (probability of detection, isolation, and quantification). The particles resulting from filtering integrate seamlessly into a sequential Monte Carlo predictive procedure, used for estimating RUL distribution. The prognosis results are validated using a four-step hierarchical procedure. In the future, the methodology needs to be extended to systems of larger dimension and realistic data, thus requiring feature selection, dimensional reduction, more efficient inference, and massive parallelization using multi-core CPUs and GPUs. Further, the full value of diagnosis uncertainty and RUL estimates obtained through the proposed prognosis methodology is realized during risk-informed decision-making for operations, inspection, maintenance, repair, and retirement.

#### ACKNOWLEDGEMENT

This work was partly supported by funds from the U. S. Air Force Research Laboratory, WPAFB through subcontract to General Dynamics Information Technology (Contract No. USAF-0060-43-0001, Project Monitor: Mark Derriso), and by the NASA ARMD/AvSP IVHM project under NRA Award No. NNX09AY54A (Project Manager: Dr. K. Goebel, NASA Ames Research Center; subcontract to Clarkson University, Principal Investigator: Dr. Y. Liu). The support is gratefully acknowledged.

#### REFERENCES

- Andrieu, C., Davy, M., & Doucet, A. (2003). Efficient particle filtering for jump Markov systems. Application to time-varying autoregressions. *IEEE Transactions on Signal Processing*, 51(7), 1762–1770. <http://doi.org/10.1109/TSP.2003.810284>
- Banjevic, D., & Jardine, A. K. S. (2006). Calculation of reliability function and remaining useful life for a Markov failure time process. *IMA Journal of Management Mathematics*, 17(2), 115–130. <http://doi.org/10.1093/imaman/dpi029>
- Bartram, G., & Mahadevan, S. (2013). Dynamic Bayesian Networks for Prognosis. In *Proceedings of the Annual Conference of the Prognostics and Health Management Society 2013*. New Orleans, LA.
- Bartram, G., & Mahadevan, S. (2013). Integration of Heterogeneous Information in SHM Models. *Structural Control and Health Monitoring*, Accepted.
- Boyer, X., & Koller, D. (1998). Tractable inference for complex stochastic processes. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence* (pp. 33–42). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. Retrieved from <http://dl.acm.org/citation.cfm?id=2074094.2074099>
- Chinnam, R. B., & Baruah, P. (2003). Autonomous diagnostics and prognostics through competitive learning driven HMM-based clustering. In *Proceedings of the International Joint Conference on Neural Networks, 2003* (Vol. 4, pp. 2466–2471 vol.4). <http://doi.org/10.1109/IJCNN.2003.1223951>
- Dong, M., & Yang, Z. (2008). Dynamic Bayesian network based prognosis in machining processes. *Journal of Shanghai Jiaotong University (Science)*, 13(3), 318–322. <http://doi.org/10.1007/s12204-008-0318-y>
- Farrar, C. R., & Worden, K. (2012). *Structural Health Monitoring: A Machine Learning Perspective*. John Wiley & Sons.
- Friedman, N., Murphy, K., & Russell, S. (1998). Learning the Structure of Dynamic Probabilistic Networks. In *UAI'98 Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence* (pp. 139–147). Morgan Kaufmann Publishers Inc. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.32.6615>
- Gebraeel, N. Z., & Lawley, M. A. (2008). A neural network degradation model for computing and updating residual life distributions. *Automation Science and Engineering, IEEE Transactions on*, 5(1), 154–163.
- Goebel, K., Saha, B., Saxena, A., Mct, N., & Riacs, N. (2008). A comparison of three data-driven techniques for prognostics. In *62nd Meeting of the Society For Machinery Failure Prevention Technology (MFPT)* (pp. 119–131).
- Goode, K. B., Moore, J., & Roylance, B. J. (2000). Plant machinery working life prediction method utilizing reliability and condition-monitoring data. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*, 214(2), 109–122. <http://doi.org/10.1243/0954408001530146>
- Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7),

- 1483–1510.  
<http://doi.org/10.1016/j.ymsp.2005.09.012>
- Jinlin, Z., & Zhengdao, Z. (2012). Fault prognosis for data incomplete systems: A dynamic Bayesian network approach. In *Control and Decision Conference (CCDC), 2012 24th Chinese* (pp. 2244–2249).  
<http://doi.org/10.1109/CCDC.2012.6244360>
- Kacprzynski, G. J., Sarlashkar, A., Roemer, M. J., Hess, A., & Hardman, B. (2004). Predicting remaining life by fusing the physics of failure modeling with diagnostics. *JOM*, 56(3), 29–35.  
<http://doi.org/10.1007/s11837-004-0029-2>
- Karpenko, M., & Sepehri, N. (2003). Robust Position Control of an Electrohydraulic Actuator With a Faulty Actuator Piston Seal. *Journal of Dynamic Systems, Measurement, and Control*, 125(3), 413–423.  
<http://doi.org/10.1115/1.1592194>
- Khan, T., Udpa, L., & Udpa, S. (2011). Particle filter based prognosis study for predicting remaining useful life of steam generator tubing. In *Prognostics and Health Management (PHM), 2011 IEEE Conference on* (pp. 1–6).  
<http://doi.org/10.1109/ICPHM.2011.6024323>
- Koller, D., & Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT Press.
- Kozlowski, J. D. (2003). Electrochemical cell prognostics using online impedance measurements and model-based data fusion techniques. In *2003 IEEE Aerospace Conference, 2003. Proceedings* (Vol. 7, pp. 3257–3270).  
<http://doi.org/10.1109/AERO.2003.1234169>
- Kulakowski, B. T., Gardner, J. F., & Shearer, J. L. (2007). *Dynamic modeling and control of engineering systems*. Cambridge University Press.
- Kwan, C., Zhang, X., Xu, R., & Haynes, L. (2003). A novel approach to fault diagnostics and prognostics. In *IEEE International Conference on Robotics and Automation, 2003. Proceedings. ICRA '03* (Vol. 1, pp. 604–609 vol.1).  
<http://doi.org/10.1109/ROBOT.2003.1241660>
- Lin, D., & Makis, V. (2004). Filters and parameter estimation for a partially observable system subject to random failure with continuous-range observations. *Advances in Applied Probability*, 36(4), 1212–1230.  
<http://doi.org/10.1239/aap/1103662964>
- Liu, J., Saxena, A., Goebel, K., Saha, B., & Wang, W. (2010). *An Adaptive Recurrent Neural Network for Remaining Useful Life Prediction of Lithium-ion Batteries*. DTIC Document.
- Lorton, A., Fouladirad, M., & Grall, A. (2013). A methodology for probabilistic model-based prognosis. *European Journal of Operational Research*, 225(3), 443–454.  
<http://doi.org/10.1016/j.ejor.2012.10.025>
- MacCormick, B. W. (1995). *Aerodynamics, Aeronautics, and Flight Mechanics*. John Wiley & Sons, Incorporated.
- Mahulkar, V., McGinnis, H., Derriso, M., & Adams, D. E. (2010). *Fault Identification in an Electro-Hydraulic Actuator and Experimental Validation of Prognosis Based Life Extending Control*. DTIC Document.
- Naval Surface Warfare Center. (2011). *Handbook of Reliability Prediction Procedures for Mechanical Equipment*. West Bethesda, Maryland 20817-5700. Retrieved from <http://www.navsea.navy.mil/nswc/carderock/pub/mechrel/products/handbook.aspx>
- Orchard, M. E., & Vachtsevanos, G. J. (2009). A particle-filtering approach for on-line fault diagnosis and failure prognosis. *Transactions of the Institute of Measurement and Control*.  
<http://doi.org/10.1177/0142331208092026>
- Pitt, M. K., & Shephard, N. (1999). Filtering via Simulation: Auxiliary Particle Filters. *Journal of the American Statistical Association*, 94(446), 590–599.  
<http://doi.org/10.1080/01621459.1999.10474153>
- RIAC Automated Databook. (2006). (Version 2.22). Reliability Information Analysis Center. Retrieved from <http://www.theriac.org>
- Ristic, B., & Arulampalam, S. (2004). *Beyond the Kalman filter: particle filters for tracking applications*. Boston, MA: Artech House.
- Saha, B., Celaya, J. R., Wysocki, P. F., & Goebel, K. F. (2009). Towards prognostics for electronics components. In *2009 IEEE Aerospace conference* (pp. 1–7).  
<http://doi.org/10.1109/AERO.2009.4839676>
- Saha, B., Goebel, K., & Christophersen, J. (2009). Comparison of prognostic algorithms for estimating remaining useful life of batteries. *Transactions of the Institute of Measurement and Control*, 31(3-4), 293–308.
- Sankararaman, S., & Mahadevan, S. (2011). Uncertainty quantification in structural damage diagnosis. *Structural Control and Health Monitoring*, 18(8), 807–824. <http://doi.org/10.1002/stc.400>
- Saxena, A., Celaya, J., Saha, B., Saha, S., & Goebel, K. (2010). Metrics for Offline Evaluation of Prognostic Performance. *International Journal of Prognostics and Health Management*, (1). Retrieved from [http://www.phmsociety.org/sites/phmsociety.org/files/phm\\_submission/2010/ijPHM\\_10\\_001.pdf](http://www.phmsociety.org/sites/phmsociety.org/files/phm_submission/2010/ijPHM_10_001.pdf)
- Thompson, D. F., Pruyun, J. S., & Shukla, A. (1999). Feedback design for robust tracking and robust stiffness in flight control actuators using a modified QFT technique. *International Journal of Control*, 72(16), 1480–1497.
- Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. *The Journal of Machine Learning Research*, 1, 211–244.

- Tobon-Mejia, D. A., Medjaher, K., Zerhouni, N., & Tripot, G. (2012). A Data-Driven Failure Prognostics Method Based on Mixture of Gaussians Hidden Markov Models. *Reliability, IEEE Transactions on*, *61*(2), 491–503.
- Tran, V. T., & Yang, B.-S. (2009). Machine Fault Diagnosis and Prognosis: The State of The Art. *The International Journal of Fluid Machinery and Systems (IJFMS)*, *2*(1), 61–71.
- Vachtsevanos, G., & Wang, P. (2001). Fault prognosis using dynamic wavelet neural networks. In *AUTOTESTCON Proceedings, 2001. IEEE Systems Readiness Technology Conference* (pp. 857–870). <http://doi.org/10.1109/AUTEST.2001.949467>
- Vlok, P.-J., Wnek, M., & Zygmunt, M. (2004). Utilising statistical residual life estimates of bearings to quantify the influence of preventive maintenance actions. *Mechanical Systems and Signal Processing*, *18*(4), 833–847. <http://doi.org/10.1016/j.ymsp.2003.09.003>
- Wang, W. (2002). A model to predict the residual life of rolling element bearings given monitored condition information to date. *IMA Journal of Management Mathematics*, *13*(1), 3–16. <http://doi.org/10.1093/imaman/13.1.3>
- Wang, W. Q., Golnaraghi, M. F., & Ismail, F. (2004). Prognosis of machine health condition using neuro-fuzzy systems. *Mechanical Systems and Signal Processing*, *18*(4), 813–831. [http://doi.org/10.1016/S0888-3270\(03\)00079-7](http://doi.org/10.1016/S0888-3270(03)00079-7)
- Wang, W., Scarf, P. A., & Smith, M. a. J. (2000). On the application of a model of condition-based maintenance. *Journal of the Operational Research Society*, *51*(11), 1218–1227. <http://doi.org/10.1057/palgrave.jors.2601042>
- Yam, R. C. M., Tse, P. W., Li, L., & Tu, P. (2001). Intelligent Predictive Decision Support System for Condition-Based Maintenance. *The International Journal of Advanced Manufacturing Technology*, *17*(5), 383–391. <http://doi.org/10.1007/s001700170173>
- Yan, J., Koç, M., & Lee, J. (2004). A prognostic algorithm for machine performance assessment and its application. *Production Planning & Control*, *15*(8), 796–801. <http://doi.org/10.1080/09537280412331309208>
- Zhang, S., & Ganesan, R. (1997). Multivariable Trend Analysis Using Neural Networks for Intelligent Diagnostics of Rotating Machinery. *Journal of Engineering for Gas Turbines and Power*, *119*(2), 378–384. <http://doi.org/10.1115/1.2815585>