Dynamic Bayesian Networks for Prognosis

Gregory Bartram[1], Sankaran Mahadevan[2]

[1,2]*Vanderbilt University, Nashville, TN, 37212, USA*

gregory.w.bartam@vanderbilt.edu
sankaran.mahadevan@vanderbilt.edu

ABSTRACT

In this paper, a methodology for probabilistic prognosis of a system using a dynamic Bayesian network (DBN) is proposed. 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 a system. 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. From this state estimate, future states of the system are predicted using the DBN and sequential Monte Carlo sampling. Prediction in this manner provides the necessary information to estimate the distribution of remaining use life (RUL). The DBN-based recursive prediction procedure may be used to estimate the system state between available measurements, when filtering is not possible. 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. Further, online system measurements may occur on different time scales from one another or only be available in particular system configurations. This necessitates seamless integration of current state estimation and predictive techniques, which are part of a prognosis methodology.

Prognosis is the process of predicting the future state of a system coupled with information about the implications of that estimate of the system health state. The quantitative prognosis of a system 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 probabilistic reliability data 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.

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 (variability, information uncertainty, and model uncertainty). It should track the health of an individual unit. 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. (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; W. Wang, Scarf, & Smith, 2000; Wenbin Wang, 2002).

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. (2010) use adaptive recurrent neural networks for the estimation of battery RUL. Goebel et al. (2008) compare relevance vector machines (RVMs), Gaussian process regression (GPR) and neural network (NN) methods for prognosis. Gebraeel & Lawley (2008) use NNs for degradation modeling and test the methodology on ball bearings. Saha et al. (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 accurate physics-based models for prediction. These include physical 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 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 (2003) uses ARMA (autoregressive moving average) models (Box, Jenkins, & Reinsel, 2008), 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. Dong and 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 (2012) use DBNs modeling discrete variables and the Boyen-Koller algorithm for prognosis. Tobon-Mejia et al. (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. (2010).

1.2. Motivation

While the preceding literature review represents a number of prognosis approaches, prognosis is still an emerging research area with room for much additional work. One promising approach that has received relatively little attention is based on DBNs. DBNs have 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 many 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.

Additionally, many prognosis methodologies are application-specific. There is still a need for prognosis methodologies that can be applied to a wide range of problems.

1.3. Contributions

In this paper, a framework for probabilistic prognosis is proposed. The methodology advances the use of DBNs in prognosis by building upon previous work in system modeling under heterogeneous information (Bartram & Mahadevan, 2013). Further, the DBN-based methodology addresses the need for a general prognosis framework for developing validated prognosis methodologies for any system.

The DBN is constructed from prior information, including physics of failure models — a key function of a prognosis methodology. 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. Particle-filter based inference is used for diagnosis, and forward sampling in the DBN is used for recursive prediction. The particle-based probabilistic state estimate of the system that results from particle filtering is ideal for diagnosis uncertainty quantification, and provides a seamless transition from diagnosis to future state prediction using sequential Monte Carlo sampling. The ability of the methodology to estimate RUL is validated using metrics from Saxena et al. (2010). In the second, online state estimation is desired, but measurements are not available or available periodically. 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 system. The system model is used for diagnosis to obtain information about the current state of the system. A sequential Monte Carlo 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 BN. A static BN, also referred to as a belief network and directed acyclic graph (DAG), is a probabilistic
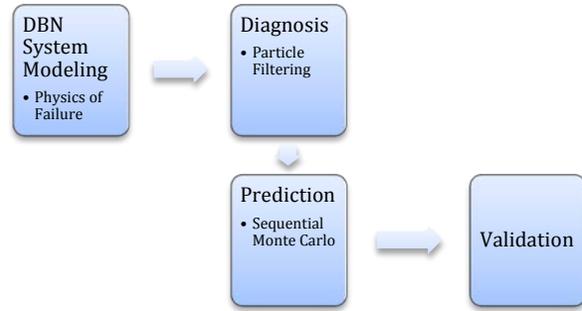

Figure 1. Proposed Prognosis Methodology

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 interval $[0, \infty)$. This is a complex distribution, but may be simplified by 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. $p(\mathbf{x}^{t+1} \mid \mathbf{x}^0, \ldots, \mathbf{x}^t) = p(\mathbf{x}^{t+1} \mid \mathbf{x}^t)$ where p indicates a probability density function and bold letters indicate a vector quantity. Additionally, the process is assumed to be stationary, meaning that $p(\mathbf{x}^{t+1} \mid \mathbf{x}^t)$ is independent of t. This approach to modeling DBNs is developed by Friedman et al. (Friedman, Murphy, & Russell, 1998).

A DBN may be composed of all discrete variables, all continuous variables, or hybrid set of discrete and continuous variables. A conditional probability distribution (CPD) is chosen for each variable, e.g. Gaussian, tabular (multinomial), softmax, deterministic, logic, etc. See Koller and Friedman (2009) for a detailed explanation of CPDs. For modeling systems with faults, it is advantageous to consider a hybrid system, typically with the continuous variables being modeled as continuous and the faults being discrete. Theory for networks with Gaussian continuous variables is developed in Heckerman and Geiger (1995) and Lauritzen (1992).

DBNs provide a flexible modeling framework, allowing integration of expert opinion, reliability data, mathematical models (including system state space, surrogate, and physics of failure models), existing databases of operational and laboratory data, and online measurement information. Bartram and Mahadevan (2013) have proposed a methodology for 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.

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 can estimate 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 developing accurate and comprehensive physics of failure models. These damage mechanisms are complex, varying with system design and dynamics, and can interact in many ways.

For illustration, the example problem in this paper considers a dynamic seal in a hydraulic actuator. Seal failure is discussed in great detail in (Naval Surface Warfare Center, 2011). A dynamic seal prevents leakage when there is relative motion between two surfaces. The seal under consideration prevents leakage between the two chambers of the actuator. Modeling the failure of a seal can become complicated very quickly, as a number of factors influence seal failure, including, material characteristics, amount of seal compression, surface irregularities, seal size, fluid pressure, pressure pulses, temperature, fluid viscosity, fluid contamination, fluid/material compatibility, allowable leakage levels, and assembly and quality control procedures. The failure symptoms include excessive leakage and slow mechanical response. Many mechanisms and causes of these symptoms are described in (Naval Surface Warfare Center, 2011).

In this paper, the wear mechanism is considered for a seal in a hydraulic actuator. 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. There are multiple wear mechanisms including adhesive wear, abrasive wear, surface fatigue, fretting wear, and erosive wear (Jones, 1983). Lancaster (1969) explains many of the complexities of abrasive wear while Briscoe and Sinha (2002) and Briscoe (1981) review wear of polymers. Due to the complexity of the mechanisms of wear, wear is typically modeled through the use of an experimentally determined wear rate.

Nikas (2010) has written an extensive literature review on seal wear in actuators. The leakage area is the result of the removal of seal material — typically a polytetrafluoroethylene (PTFE) polymer — which is a function of load, distance traveled, material properties of the actuator and seal, geometry of the actuator, temperature,

hydraulic fluid viscosity, and contaminants. Experimentally determined wear rates (mm^3/m/N) are available for PTFE composites used in hydraulic actuators e.g. Sawyer et al. (2003) and Khedkar et al. (2002).

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

$$V(t) = w_{seal}(t)F(t)d(t) \qquad (1)$$

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

For the seal shown in Fig. 2, where L is the contact length of the seal and P is pressure, the leakage area (considered in Eqs. 16-29 as in (Thompson, Pruyn, & Shukla, 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. However, for the sake
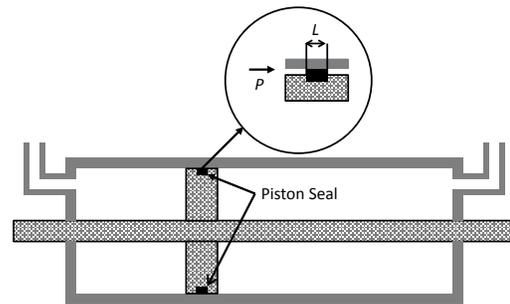

Figure 2. Hydraulic actuator diagram showing dynamic seals

of demonstration, it is assumed that the leakage area and volume of material removed vary linearly. This implies that the wear rate is steady, which is possible outside of the initial wear-in phase and under constant environmental conditions. Additionally, the load and velocity of the actuator are assumed to remain steady.

2.3. 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 set candidates to further analyze. To isolate candidate fault sets, statistical inference computes the probability of each fault set. The magnitude of the fault may then be estimated.

In the context of prognosis, diagnosis (or more generally, 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, hypothesis testing error, error in inference, and any approximations in optimization or least squares procedures used for estimating fault magnitudes. Sankararaman and Mahadevan propose a methodology for quantifying the uncertainty in diagnosis. This is an integral part of the diagnosis procedure, and it is expanded in this paper to accommodate a particle filter (PF) based diagnosis procedure.

Detection and isolation 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 procedure is to "unroll" the two time slice network and compute the states of all the unobserved variables in the system, $\mathbf{x}^t$, including faults, using standard inference techniques such as the clique tree algorithm or Markov chain Monte Carlo (MCMC) (Koller & Friedman, 2009). However, exact inference is generally a computationally intractable problem (Boyen & Koller, 1998). As a result, approximate inference based on Bayesian recursive filtering is pursued.

2.3.1. 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 σ^t. $\sigma^t = p(\mathbf{x}^t|\mathbf{z}^{1:t})$, where $p(\mathbf{x}^t|\mathbf{z}^{1:t})$ 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 (see Appendix 1), is the engine for belief state updating.

$$\sigma^{t+1}(\mathbf{x}^{t+1}) = \frac{p(\mathbf{z}^{t+1}|\mathbf{x}^{t+1})p(\mathbf{x}^{t+1}|\mathbf{z}^{1:t})}{p(\mathbf{z}^{t+1}|\mathbf{z}^{1:t})} \qquad (2)$$

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

Complete tutorials on Bayesian recursive filtering are available in Koller and Friedman (2009) and Ristic and Arulampalam (2004).

2.3.2. Particle Filtering

Under certain assumptions, such as when the system is linear 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 (2003)).

Particle filtering is a method for approximating the distribution of the belief state with a set of samples and weights. Common particle filtering method 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. (2004) and in Koller and Friedman (2009).

A summary of the SIS algorithm for one time step is as follows. A previous (or initial if $t = 1$) set of N_s weights w_t^i and N_s corresponding particles $\mathbf{x}_t^i$ are given initially or known from the previous time step, where i denotes the ith particle. These particles represent an approximation of the belief state by

$$\begin{aligned} \sigma^{t+1}(\mathbf{x}^{t+1}) &= p(\mathbf{x}^{t+1}|\mathbf{z}^{1:t+1}) \\ &\approx \sum_{i=1}^{N_s} \delta(\mathbf{x} - \mathbf{x}_i^{t+1})w_i^{t+1} \end{aligned} \qquad (3)$$

N_s samples are drawn from the importance distribution, $q(\mathbf{x}_i^{t+1}|\mathbf{x}_i^t, \mathbf{z}^{t+1})$, where $\mathbf{z}^{t+1}$ are the measurements at the $t+1$th time step. In a DBN, sampling is performed in the

two-slice template after computing $\mathbf{x}^t$ and $\mathbf{z}^{t+1}$. The values for the remaining $\mathbf{x}^{t+1}$ are then sequentially sampled in topological order (parent then child).

A weight w_i^{t+1} is then computed for each particle $\mathbf{x}_i^{t+1}$ up to a normalizing constant based on the ratio of the belief state to the importance density,

$$w_i^{t+1} \propto \frac{\sigma(\mathbf{x}_i^{1:t+1}|\mathbf{z}^{1:t+1})}{q(\mathbf{x}_i^{1:t+1}|\mathbf{z}^{1:t+1})} \tag{4}$$

using the weight update equation in Eq. (5), which is derived from the ratio of the pdf of the belief state to the pdf of the importance sampling density.

$$w_i^{t+1} \propto \frac{p(\mathbf{z}^{t+1}|\mathbf{x}_i^{t+1})p(\mathbf{x}_i^{t+1}|\mathbf{x}_i^t)}{q(\mathbf{x}_i^{t+1}|\mathbf{x}_i^t, \mathbf{z}^{t+1})} \tag{5}$$

The weights $\mathbf{w}^{t+1}$ are normalized so their sum is equal to 1. The normalized weights and points $\{\mathbf{x}_i^{t+1}\}_{i=1}^{N_2}$ form an approximation to the belief state estimate in Eq. (3).

The basic SIS algorithm suffers from the degeneracy phenomenon, wherein all but a few of the particles have negligible weight after only a few updates. This tends to waste computational effort on particles with practically zero probability. Two techniques to reduce this phenomenon are choosing an optimal importance density $q(\mathbf{x}_i^{t+1}|\mathbf{x}_i^t, \mathbf{z}^{t+1})$ and resampling. The optimal importance density may only be determined analytically when the system variables are discrete with a finite number of possible values or when the system variables are Gaussian. In other cases, suboptimal approximations based on local linearization (Doucet, Godsill, & Andrieu, 2000) or Gaussian approximations using the unscented transform (West & Harrison, 1997) may be used. Often, for convenience the importance density $q(\mathbf{x}^{t+1}|\mathbf{x}_i^t, \mathbf{z}^{t+1})$ is taken as the prior $p(\mathbf{x}_i^{t+1}|\mathbf{x}_i^t)$ or the likelihood $p(\mathbf{z}^{t+1}|\mathbf{x}_i^{t+1})$. If the prior is used, $q(\mathbf{x}^{t+1}|\mathbf{x}_i^t, \mathbf{z}^{t+1}) = p(\mathbf{x}^{t+1}|\mathbf{x}_i^t)$, and the weight update in Eq. (5) simplifies to

$$w_i^{t+1} \propto w_i^t p(\mathbf{z}^{t+1}|\mathbf{x}_i^{t+1}) \tag{6}$$

Resampling focuses the particle filter on the particles with the largest weights. An empirical CDF is constructed based on the weights $\mathbf{w}_t$. Particles are sampled (with replacement), replicating the particles with the largest weights. The result is N_s particles all with weight $1/N_s$. Resampling may be performed after every update or when a measure of degeneracy, the effective sample size, N_{eff}, falls below a threshold. N_{eff} may be taken as (Ristic & Arulampalam, 2004)

$$N_{eff} = \frac{N_s}{\sum_{i=1}^{N_s}(w_i^{t+1})^2} \tag{7}$$

Many variations of the SIS particle filter exist with different importance densities $q(\mathbf{x}^{t+1}|\mathbf{x}_i^t, \mathbf{z}^{t+1})$ and resampling procedures (Ristic & Arulampalam, 2004). In this paper, an algorithm for systems with multiple operating modes (Andrieu, Davy, & Doucet, 2003) that extends the auxiliary particle filter (Pitt & Shephard, 1999) is used.

2.4. 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 ith combination at the tth cycle has an expected probability $p_i^t = \sum N_i^t w_i^t / N_s$, where N_s is the number of samples used in particle filtering, N_i^t is the number of occurrences of the ith fault combination, and w_t^i are the normalized weights for those particles.

The probability of *any* fault (detection probability) is then $p_F^t = 1 - p_0^t$, where p_0^t is the probability of the fault combination where 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 $\mathbf{p}_t$ 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

$$Var[p_t^i] = \frac{N_k^i(N_s - N_k^i)}{N_s^2(N_s + 1)} \tag{8}$$

The uncertainty in $\mathbf{p}_t$ 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 access the criticality of damage and the appropriate actions to make to balance safety and cost concerns.

2.4.1. 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. 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. Figure 3 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 belief state estimate (with measurements available), particles are recursively sampled through the two time slice DBN until some termination criteria 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, $\{\mathbf{\Phi}(t)\}_{i=1}^{N_s}$. Each trajectory consists of a series of predictions for the variables of interest, $\mathbf{\Phi}(t) = \{\boldsymbol{\varphi}(t|t), \boldsymbol{\varphi}(t+1|t), \dots \boldsymbol{\varphi}(EoP|t)\}$, where the end of prediction (EoP) is the time index of the last prediction before the end of life (EoL) is reached. Particle weights are fixed from the last available measurement, as there is no basis for updating the weights (Eq. 5). This results in a particle-based approximation of RUL (similar to the belief state approximation), using the last available set of weights. When a new measurement is obtained, a new RUL distribution is estimated.
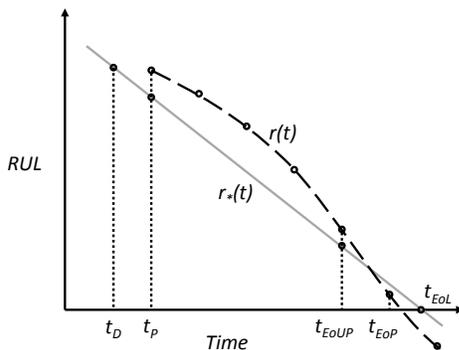
To further tailor the prognosis to a particular UUT, the conditional probability distributions in the DBN may be updated as measurements become available. This may be performed via Bayesian updating of the distributions. If a conjugate prior is available, the update can be performed analytically. Otherwise, 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.

2.4.2. 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 frequency 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. The process is shown in Fig. 4.
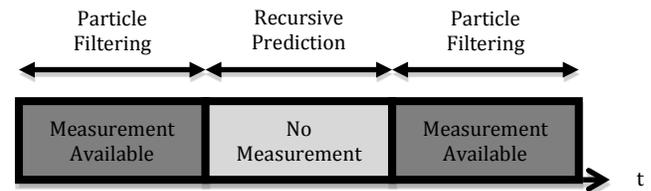

Figure 3. Prognosis time indices: $r_*(t)$ is the ground truth RUL, t_{EoUP} is the end of useful prognosis, dashed line depicts mean $r(t)$.

Figure 4. Handling measurement gaps

2.5. Prognosis Validation

Prognosis validation is essential to establish confidence in the RUL estimate. Many sources of uncertainty, including modeling errors, sensor faults, data noise, and unpredictable loading conditions and operating environments, strongly affect prognosis. Therefore, validation of a prognosis procedure must be done using strong 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. (2010) propose a standard offline four metric hierarchical test to evaluate a prognosis algorithm. This hierarchical test assumes that prognosis improves as more measurements become available. Combined, these four metrics 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, β. It is desirable for p to be greater than β. The primary difference between the first two metrics is in how α is determined, which results in a stricter test for the second metric than the first.

In the first metric, prognostic horizon (PH) is considered. Prognostic horizon 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 constant α, 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 algorithm. Figure 5a 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 (Fig. 5b). Additionally, the accuracy of the RUL is considered at time t_λ , where $0 \le \lambda \le 1$, $t_\lambda = t_P + \lambda(t_{EoL} - t_P)$, 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
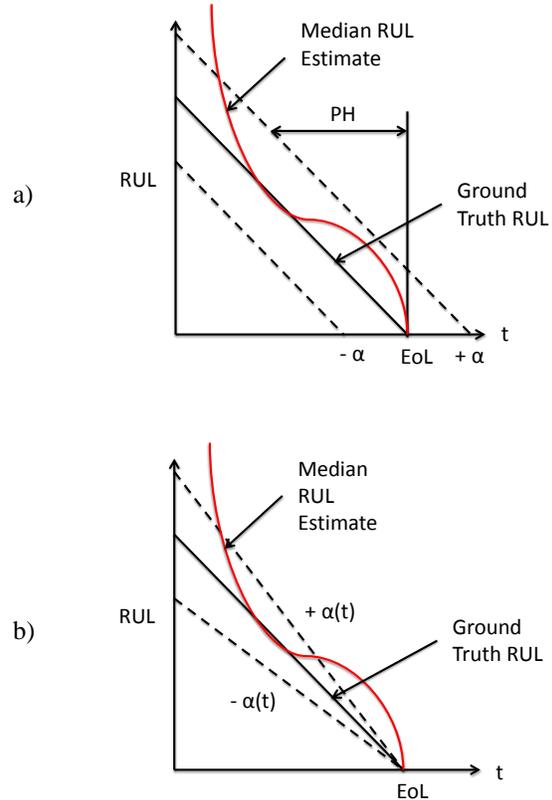

Figure 5. a) Prognostic horizon with +/- α bounds about the ground truth RUL
b) +/- α bounds for evaluating α-λ accuracy

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 estimates are quantified. At t_λ

$$RA_\lambda = 1 - \frac{|r_*(t_\lambda) - r(t_\lambda)|}{r_*(t_\lambda)} \qquad (9)$$

where $r(t_\lambda)$ is a central tendency point such as the mean or median of the RUL estimate at t_λ 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, if the prognostic algorithm satisfies all the previous metrics, a final metric to compute is convergence. Convergence is a measure of how quickly a metric, such as RA, improves with time. A high rate of convergence is desirable and leads to a larger PH. To estimate convergence of a prognosis algorithm based on some metric *M,*

$$C_M = [(x_c - t_P)^2 + y_c^2]^{1/2} \qquad (10)$$

where

$$x_c = \frac{\frac{1}{2}\sum_{i=P}^{EoUP}(t_{i+1}{}^2 - t_i{}^2)M(t_i)}{\sum_{i=P}^{EoUP}(t_{i+1} - t_i)M(t_i)} \qquad (11)$$

and

$$y_c = \frac{\frac{1}{2}\sum_{i=P}^{EoUP}(t_{i+1}{}^2 - t_i{}^2)M(t_i)^2}{\sum_{i=P}^{EoUP}(t_{i+1} - t_i)M(t_i)} \qquad (12)$$

$M(t_i)$ 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.

2.6. Summary of Prognosis Framework

This section presented a framework for probabilistic prognosis. 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. ILLUSTRATIVE 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. 6). The pump moves hydraulic fluid through the servovalve and into the actuator. The servovalve controls the flow of hydraulic fluid into the actuator, thus modulating the position of the actuator. 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. Considering the actuator cross section in Fig. 7, 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.

3.1. DBN Model Construction

3.1.1. Expert Opinion

Expert opinion was considered first to define the basic parameters of the problem. A DBN representation of the system was chosen because heterogeneous information
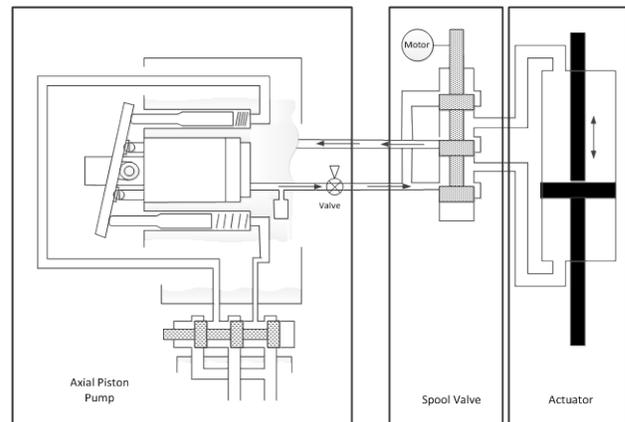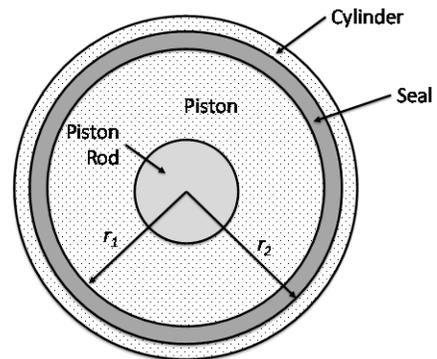

Figure 6. Hydraulic actuator system

Figure 7. Actuator cross section

sources were available, the intended use of the model is diagnosis and prognosis, and the system is dynamic. 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. 8) based on expert opinion. This generic two time slice structure consists of the set of faults, F, model parameters, θ, system state, y, 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 θ.
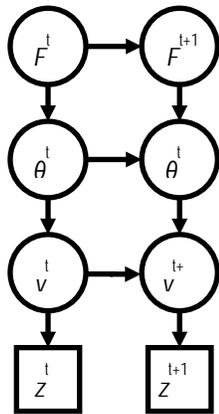

Figure 8. Generic DBN structure.

Table 1. List of faults and affected parameters.

Fault	Parameter Affected
Seal Leak	Leakage Area
Water Leak into System	Hydraulic Fluid Bulk Modulus
Air Leak Into System	Hydraulic Fluid Bulk Modulus
Pressure Valve Malfunction	Supply Pressure
Pump Pressure Sensor Fault	Supply Pressure
Electrical Fault	Control Signal

Table 1 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 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 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 it is zero when no leakage exists. Upon instantiation of a leak, its value is assumed to follow a Gaussian distribution. Thereafter, the leak is assumed to grow according to a polynomial regression model (Section 3.1, Mathematical Models), which is constructed using laboratory data.

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. 9 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 extrapolate 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
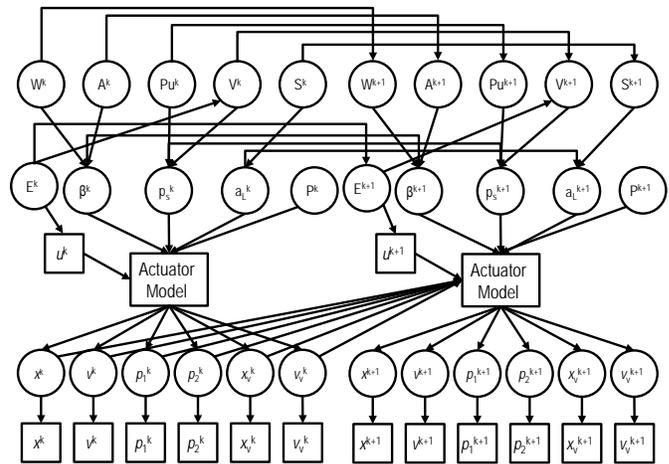

Figure 9. DBN structure as a result of expert opinion

Table 2. DBN variables.

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
Fluid bulk modulus	β	MPa	continuous	inferred
Supply pressure	p_s	MPa	continuous	inferred
Leakage area	a_{leak}	mm^2	continuous	inferred

Table 3. Faults Considered

Fault	Subsystem	Information Source
Seal Leak	Actuator	RIAC Databook (*RIAC Automated Databook*, 2006), Literature (Sepeheri, Karpenko, An, & Karam, 2005)
Water Leak into System	Piping/Fittings	RIAC Databook (*RIAC Automated Databook*, 2006), Literature (Sepeheri et al., 2005)
Air Leak Into System	Piping/Fittings	RIAC Databook (*RIAC Automated Databook*, 2006), Literature (Sepeheri et al., 2005)
Pressure Valve Malfunction	Pressure Valve	RIAC Databook (*RIAC Automated Databook*, 2006), Literature (Sepeheri et al., 2005)
Pressure Sensor	Piston Pump	Mathematical Model (Zeiger & Akers, 1986), Literature (Zeliang, 2005)
Electrical Fault	Electrical	RIAC Databook (*RIAC Automated Databook*, 2006)

RIAC Databook (2006) and the Handbook of Reliability Prediction Procedures for Mechanical Equipment (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 information source for that fault.

The failure rates 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. See Bartram and Mahadevan (Bartram & Mahadevan, 2013) for details.

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. (2007) and Thompson et al. (1999) (see Appendix), is integrated into the DBN as a deterministic conditional probability distribution

(Koller & Friedman, 2009). This model has been implemented in the Matlab Simulink environment.

For demonstration of the prognosis methodology, the load is synthesized using an ARIMA (autoregressive integrated moving average) model, which is treated as a deterministic conditional probability distribution in the DBN. 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 (for e.g. see Mahulkar et al. (2010), Karpenko and Sepeheri (2003), and McCormick (1995)).

Finally, the physics of failure model for the seal leak is considered. The seal leakage area is modeled as in Section 2.2. The leakage area is modeled from laboratory data using a polynomial regression of the form $a_{leak}^t = c_1 + c_2(a_{leak}^{t+1})^2$.

3.1.4. Operational and Laboratory Data

Operational and laboratory data appear as historical databases and online measurement data. Laboratory data are used to train the polynomial regression model to estimate wear rate. Online measurement data (of the load P) are used to estimate the parameters of the ARIMA model used in load estimation.

3.2. Diagnosis

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. Measurements were obtained and updating performed at 0.5 second intervals. 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 detection was calculated as in Section 2.2. If the probability of detection exceeded 95%, an alarm was triggered. The fault was then isolated and quantified. Figure (10) shows maximum a posteriori (MAP) estimates of the system responses against their measured values. It is seen that the MAP system responses track the measured values closely. Figure 11 shows the MAP estimates of the system parameters, 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 Figure 11 are sensitive to changes in the supply pressure and leakage area, but 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

Figure 11. MAP estimate system parameters and load with ground truth and measured values.

included in the ground truth model. In both Fig. 10 and Fig. 11, the good estimates may be attributed to the use of an accurate physics-based model, but also to the use of synthetic measurement data, which may favorably bias the performance of filtering.

3.3. Diagnosis Uncertainty

Diagnosis uncertainty was quantified after performing the diagnostic tasks of detection, isolation, and quantification. Figure 12 shows a kernel density estimate of the seal leak area from the particles at $t = 6.5$ seconds. Figure 13 shows the detection probability as it evolves with time. The detection probability passes the detection threshold soon after the fault occurs.
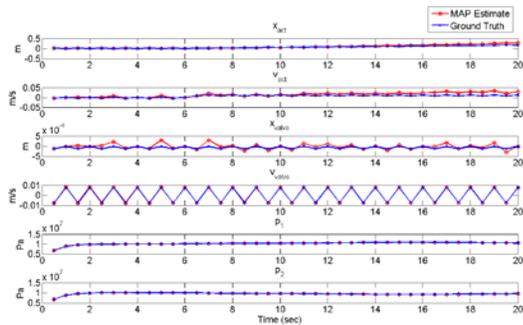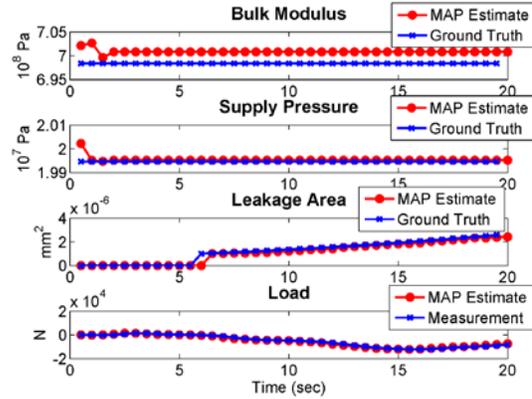

Figure 10. MAP estimates and measured values of actuator position and velocity, servovalve position and velocity, and pressure in each actuator chamber.
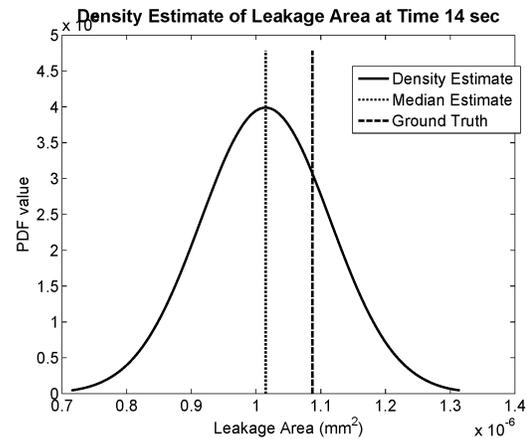

Figure 12. Kernel density estimate of leakage area estimate from particle filtering
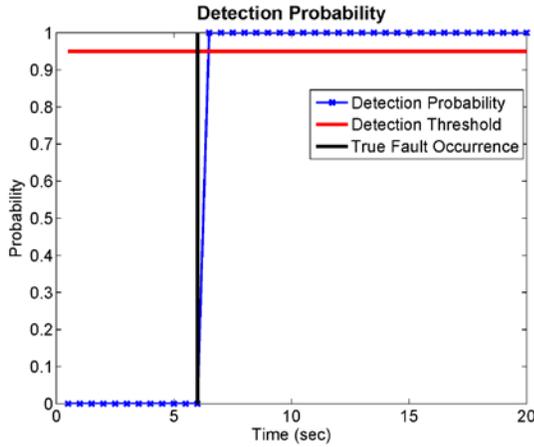
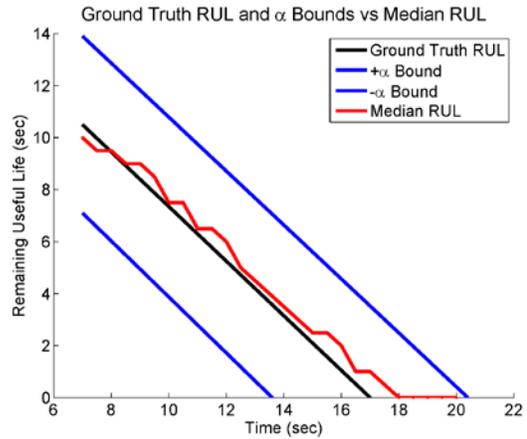Figure 13. Fault detection probability with actual fault time

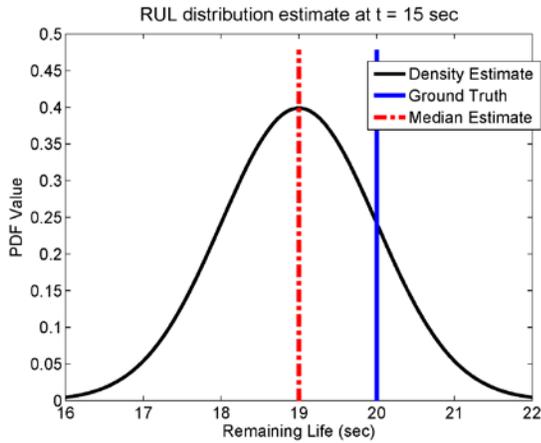Figure 15. Ground truth RUL, median RUL, and α bounds with α = 0.10

Figure 14. RUL density estimate at t = 15 sec

Figure 16. Probability that RUL is within α bounds with α = 0.10

3.4. Prediction

After diagnosing the leak, estimation of the RUL distribution was performed at the time of prognosis, t_P, as per Section 2.3. The RUL distribution assumes a failure threshold for leakage area of 2E-6 m^2. The resulting RUL distribution is shown in Fig. 14.

3.5. 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. 15, median RUL estimates are plotted against the ground truth RUL with +/- α bounds. The +/- α bounds are selected to be +/- 10% of the ground truth EoL about the current ground truth RUL. Figure 16 indicates whether the probability of the RUL estimate being between the +/- α bounds at a particular
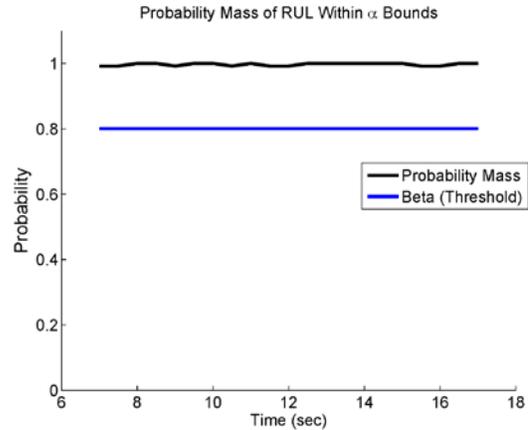
time is greater than a threshold value, taken as 0.8. From this information, it is also determined that the prognostic horizon is 10 seconds (or 20 time steps with a sampling frequency of 2 samples/sec) because the first time that $0.8 \leq \pi[r(t_j)]_{-\alpha}^{+\alpha}$ is at $t = 7$, the EoL is $t = 17$. This is 10 seconds before the EoL.

+/- α bounds that narrow as the EoL approaches are considered in Fig. 17 for λ = 0.5 and α = 0.20. λ = 0.5 considers the accuracy of the RUL estimate halfway between the time of prognosis and end of life, termed $t_λ$. Figure 18 shows the λ-α accuracy, which is a binary value that indicates whether the probability of the RUL estimate being between the +/- α bounds at a particular time is greater than a threshold value, taken as 0.8 here. Although the median RUL estimate appears close to the ground truth
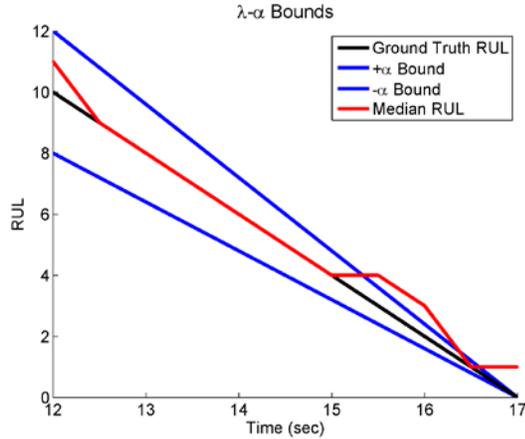
Figure 17. Bounds used for calculating λ-α accuracy with α = 0.20

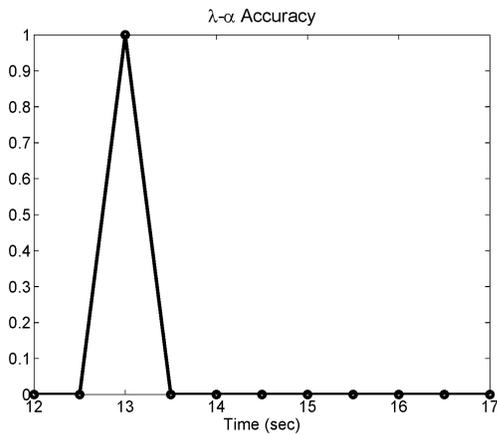Figure 18. Probability RUL is within α bounds with α = 0.10

RUL, the λ-α accuracy is generally zero, indicating that the RUL estimate is too diffuse to pass this test. This indicates that model error in the physics of failure model is the dominant source of error as opposed to other errors that may decrease as the EoL is approached, such as errors in load estimation.

Based on the relative accuracy, the convergence is estimated to be 8.40. When comparing prognostic algorithms, larger convergence values are desirable.

Figure 19 shows the relative accuracy of the RUL density estimate based on the median RUL value, and shows that the median values are accurate.

3.6. Discussion

The DBN-based methodology successfully integrates heterogeneous sources of information to diagnose the system and estimate RUL. Particle-filter based inference
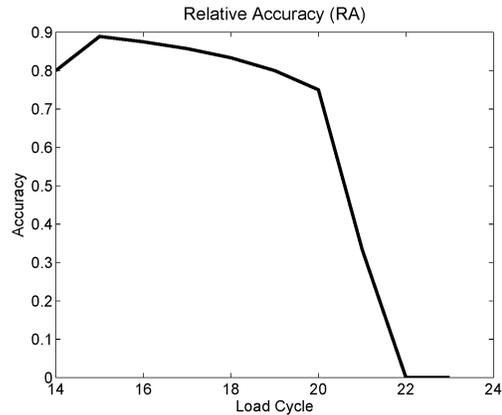

Figure 19. Relative accuracy based on median RUL estimate

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 diffuse. Additional measurements primarily affect the median RUL estimate, not the variance of RUL, primarily due to the simplifying assumptions that remove feedback from the actuator dynamic model into the leakage area model. Inclusion of 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, 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.

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 is presented in this paper, considering heterogeneous information sources and diagnosis uncertainty. First, expert opinion is used to establish the system definition and basic assumptions. 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 database and used for estimating a

polynomial regression model. This system model is used in online diagnosis via particle filter-based inference. 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, thus requiring feature selection, dimensional reduction, and more efficient inference.

ACKNOWLEDGEMENT

APPENDIX 1: BAYESIAN RECURSIVE FILTERING

Given the belief state $\sigma^t(\mathbf{x}^t)$, before obtaining $\mathbf{z}^{t+1}$, the belief state at $t+1$ is

$$\sigma^{t+1}(\mathbf{x}^{t+1}) = p(\mathbf{x}^{t+1}|\mathbf{z}^{1:t}) \tag{13}$$

which may be expanded by summing over the states of $\mathbf{x}^t$ as

$$\sigma^{t+1}(\mathbf{x}^{t+1}) = \sum_{\mathbf{x}^t} p(\mathbf{x}^{t+1}|\mathbf{x}^t, \mathbf{z}^{1:t}) p(\mathbf{x}^t|\mathbf{z}^{1:t}) \tag{14}$$

Using the Markov assumption, which says that the future $\mathbf{x}^{t+1}$ is independent of all else given the previous state $\mathbf{x}^t$, the term $\mathbf{z}^{1:t}$ may be eliminated from $p(\mathbf{x}^{t+1}|\mathbf{x}^t, \mathbf{z}^{1:t})$, resulting in

$$\sigma^{t+1}(\mathbf{x}^{t+1}) = \sum_{\mathbf{x}^t} p(\mathbf{x}^{t+1}|\mathbf{x}^t) p(\mathbf{x}^t|\mathbf{z}^{1:t}) \tag{15}$$

where $p(\mathbf{x}^t|\mathbf{z}^{1:t}) = \sigma^t(\mathbf{x}^t)$. Upon receiving the measurement at time $t+1$, Bayes' rule may be used to update the belief state, and Eq. 3 becomes

$$\sigma^{t+1}(\mathbf{x}^{t+1}) = p(\mathbf{x}^{t+1}|\mathbf{z}^{1:t}, \mathbf{z}^{t+1}) \tag{16}$$

By Bayes' rule expansion of the right hand side of Eq. 4,

$$\sigma^{t+1}(\mathbf{x}^{t+1}) = \frac{p(\mathbf{z}^{t+1}|\mathbf{x}^{t+1}, \mathbf{z}^{1:t}) p(\mathbf{x}^{t+1}|\mathbf{z}^{1:t})}{p(\mathbf{z}^{t+1}|\mathbf{z}^{1:t})} \tag{17}$$

Because the measurements $\mathbf{z}^{t+1}$ and $\mathbf{z}^{1:t}$ are conditionally independent given $\mathbf{x}^{t+1}$ (Section 2.2 Fig. 2a), $p(\mathbf{z}^{t+1}|\mathbf{x}^{t+1}, \mathbf{z}^{1:t}) = p(\mathbf{z}^{t+1}|\mathbf{x}^{t+1})$, resulting in

$$\sigma^{t+1}(\mathbf{x}^{t+1}) = \frac{p(\mathbf{z}^{t+1}|\mathbf{x}^{t+1}) p(\mathbf{x}^{t+1}|\mathbf{z}^{1:t})}{p(\mathbf{z}^{t+1}|\mathbf{z}^{1:t})} \tag{18}$$

where $p(\mathbf{x}^{t+1}|\mathbf{z}^{1:t})$ is equivalent to Eq. 15.

APPENDIX 2: HYDRAULIC ACTUATOR MODEL

Parameters and variables for the system are given in Table 1A.

$$\dot{x}_{act} = v_{act} \tag{19}$$

$$\dot{v}_{act} = \frac{1}{m}\left[(P_1 - P_2)A_{pist} - b_{act}v_{act} - k_{act}x_{act} - F_{ext}\right] \tag{20}$$

$$\dot{P}_1 = \frac{1}{C_{f1}}(Q_1 - A_{pist}Q_2 + Q_{leak}) \tag{21}$$

$$\dot{P}_2 = \frac{1}{C_{f2}}(V_2 x_{act} - Q_2 - Q_{leak}) \tag{22}$$

$$\dot{x}_{valve} = v_{valve} \tag{23}$$

$$\dot{v}_{valve} = a_1 v_{valve} + a_0 x_{valve} + b_0 e_{command} \tag{24}$$

$$C_{f1} = \frac{V_1(x_{act})}{\beta} \tag{25}$$

$$C_{f2} = \frac{V_2(x_{act})}{\beta} \tag{26}$$

If $x_{valve} > 0$,
$$Q_1 = C_d w_{valve} x_{valve} sign(P_s - P_1)\sqrt{\frac{2}{\rho}|P_s - P_1|} \tag{27}$$

$$Q_2 = C_d w_{valve} x_{valve}\sqrt{\frac{2}{\rho}(P_2)} \tag{28}$$

If $x_{valve} < 0$,
$$Q_1 = C_d w_{valve} x_{valve}\sqrt{\frac{2}{\rho}(P_1)} \tag{29}$$

$$Q_2 = C_d w_{valve} x_{valve} sign(P_s - P_2)\sqrt{\frac{2}{\rho}|P_s - P_2|} \tag{30}$$

$$Q_{leak} = C_d a_{leak}\sqrt{\frac{2}{\rho}|P_2 - P_1|} sign(P_2 - P_1) \tag{31}$$

Table 1A. Model parameters and variables for a spool valve and a hydraulic actuator.

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 2	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^2
Valve port width	w_{valve}	0.0025 m
Spool valve model coefficients	b_0	90 m/Vs^2
	a_0	360,000 1/s^2
	a_1	1/s
Hydraulic fluid bulk modulus	β	1000 MPa
Hydraulic fluid density	ρ	847 kg/m^3
Discharge coefficient	C_d	0.7
Supply pressure	P_{supply}	20 MPa
Chamber 1 volume	V_1	m^3
Chamber 2 volume	V_2	m^3
Chamber 1 fluid capacitance	C_{f1}	m^3/(kg/s)
Chamber 2 fluid capacitance	C_{f2}	m^{3}s/(kg/s)
Volumetric flow rate into chamber 1	Q_1	m^3/s
Volumetric flow rate out of chamber 2	Q_2	m^3/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^3/s
Leakage area	a_{leak}	0 m^2

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. doi: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. doi:10.1093/imaman/dpi029

Bartram, G., & Mahadevan, S. (2013). Integration of Heterogeneous Information in SHM Models. *Structural Control and Health Monitoring*, Accepted.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time series analysis: forecasting and control*. Hoboken, N.J.: John Wiley.

Boyen, 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

Briscoe, B. (1981). Wear of polymers: an essay on fundamental aspects. *Tribology International*, *14*(4), 231–243. doi:10.1016/0301-679X(81)90050-5

Briscoe, B. J., & Sinha, S. K. (2002). Wear of polymers. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, *216*(6), 401–413. doi:10.1243/135065002762355325

Chen, Z. (2003). Bayesian filtering: From Kalman filters to particle filters, and beyond. *Statistics*, 1–69.

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). Presented at the Proceedings of the International Joint Conference on Neural Networks, 2003. doi: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. doi:10.1007/s12204-008-0318-y

Doucet, A., Godsill, S., & Andrieu, C. (2000). On Sequential Monte Carlo Sampling Methods for Bayesian Filtering. *STATISTICS AND COMPUTING*, *10*(3), 197–208.

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. doi:10.1243/0954408001530146

Heckerman, D., & Geiger, D. (1995). Learning Bayesian Networks: A unification for discrete and Gaussian domains. *PROCEEDINGS OF ELEVENTH CONFERENCE ON UNCERTAINTY INARTI CIAL INTELLIGENCE*. Retrieved from http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.156.7976

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. doi:10.1016/j.ymssp.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). Presented at the Control and Decision Conference (CCDC), 2012 24th Chinese. doi:10.1109/CCDC.2012.6244360

Jones, M. H. (Ed.). (1983). *Industrial Tribology: The Practical Aspects of Friction, Lubrication and Wear*. North Holland.

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. doi: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. doi: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). doi:10.1109/ICPHM.2011.6024323

Khedkar, J., Negulescu, I., & Meletis, E. I. (2002). Sliding wear behavior of PTFE composites. *Wear*, *252*(5–6), 361–369. doi:10.1016/S0043-1648(01)00859-6

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). Presented at the 2003 IEEE Aerospace Conference, 2003. Proceedings. doi: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). Presented at the IEEE International Conference on Robotics and Automation, 2003. Proceedings. ICRA '03. doi:10.1109/ROBOT.2003.1241660

Lancaster, J. K. (1969). Abrasive wear of polymers. *Wear*, *14*(4), 223–239. doi:10.1016/0043-1648(69)90047-7

Lauritzen, S. L. (1992). Propagation of Probabilities, Means, and Variances in Mixed Graphical Association Models. *Journal of the American Statistical Association*, *87*(420), 1098–1108. doi:10.2307/2290647

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. doi: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. doi: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

Nikas, G. K. (2010). Eighty years of research on hydraulic reciprocating seals: Review of tribological studies and related topics since the 1930s. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, *224*(1), 1–23. doi:10.1243/13506501JET607

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*. doi: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. doi:10.1080/01621459.1999.10474153

RIAC Automated Databook. (2006). 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). Presented at the 2009 IEEE Aerospace conference. doi:10.1109/AERO.2009.4839676

Saha, Bhaskar, 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. (n.d.). Uncertainty quantification in structural damage diagnosis. *Structural Control and Health Monitoring*. doi:10.1002/stc.400

Sawyer, W. G., Freudenberg, K. D., Bhimaraj, P., & Schadler, L. S. (2003). A study on the friction and wear behavior of PTFE filled with alumina nanoparticles. *Wear*, *254*(5–6), 573–580. doi:10.1016/S0043-1648(03)00252-7

Saxena, A., Celaya, J., Saha, B., Saha, S., & Goebel, K. (2010). Metrics for Offline Evaluation of Prognostic Performance. *International Journal of Prognositcs and Health Management*, (1). Retrieved from http://www.phmsociety.org/sites/phmsociety.org/files/phm_submission/2010/ijPHM_10_001.pdf

Sepeheri, N., Karpenko, M., An, L., & Karam, S. (2005). A test rig for experimentation on fault tolerant control and condition monitoring algorithms in fluid power systems: from designing through implementation. *Transactions of the Canadian Society for Mechanical Engineering*, *29*(3), 441–458.

Thompson, D. F., Pruyn, 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). Presented at the AUTOTESTCON Proceedings, 2001. IEEE Systems Readiness Technology Conference. doi: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. doi:10.1016/j.ymssp.2003.09.003

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. doi: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. doi:10.1057/palgrave.jors.2601042

Wang, Wenbin. (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. doi:10.1093/imaman/13.1.3

West, M., & Harrison, J. (1997). Bayesian Forecasting and Dynamic Models (Springer Series in Statistics).

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. doi: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. doi:10.1080/09537280412331309208

Zeiger, G., & Akers, A. (1986). Dynamic analysis of an axial piston pump swashplate control. *ARCHIVE: Proceedings of the Institution of Mechanical Engineers, Part C: Mechanical Engineering Science 1983-1988 (vols 197-202)*, *200*(13), 49–58. doi:10.1243/PIME_PROC_1986_200_093_02

Zeliang, L. (2005, November). *Condition Monitoring of Axial Piston Pump* (Master's Thesis). University of Saskatchewan, Saskatoon, Saskatchewan, Canada. Retrieved from http://library2.usask.ca/theses/available/etd-11252005-202705/unrestricted/EricLithesis2005NovA.pdf

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. doi:10.1115/1.2815585