

# Journal of Intelligent Material Systems and Structures

<http://jim.sagepub.com/>

---

## **An adaptive learning damage estimation method for structural health monitoring**

Debejyo Chakraborty, Narayan Kovvali, Antonia Papandreou-Suppappola and Aditi Chattopadhyay

*Journal of Intelligent Material Systems and Structures* published online 14 April 2014

DOI: 10.1177/1045389X14522531

The online version of this article can be found at:

<http://jim.sagepub.com/content/early/2014/04/14/1045389X14522531>

---

Published by:



<http://www.sagepublications.com>

**Additional services and information for *Journal of Intelligent Material Systems and Structures* can be found at:**

**Email Alerts:** <http://jim.sagepub.com/cgi/alerts>

**Subscriptions:** <http://jim.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations:** <http://jim.sagepub.com/content/early/2014/04/14/1045389X14522531.refs.html>

>> [OnlineFirst Version of Record](#) - Apr 14, 2014

[What is This?](#)

# An adaptive learning damage estimation method for structural health monitoring

Debejyo Chakraborty<sup>1</sup>, Narayan Kovvali<sup>2</sup>,  
Antonia Papandreou-Suppappola<sup>2</sup> and Aditi Chattopadhyay<sup>3</sup>

Journal of Intelligent Material Systems  
and Structures

1–19

© The Author(s) 2014

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1045389X14522531

jim.sagepub.com



## Abstract

Structural health monitoring is an important problem of interest in many civil infrastructure and aerospace applications. In the last few decades, many techniques have been investigated to address the detection, estimation, and classification of damage in structural components. One of the key challenges in the development of real-world damage identification systems, however, is variability due to changing environmental and operational conditions. Conventional statistical methods based on static modeling frameworks can prove to be inadequate in a dynamic and fast changing environment, especially when a sufficient amount of data is not available. In this paper, a novel adaptive learning structural damage estimation method is proposed in which the stochastic models are allowed to perpetually change with the time-varying conditions. The adaptive learning framework is based on the use of Dirichlet process (DP) mixture models, which provide the capability of automatically adjusting to structure within the data. Specifically, time–frequency features are extracted from periodically collected structural data (measured sensor signals), that are responses to ultrasonic excitation of the material. These are then modeled using a DP mixture model that allows for a growing, possibly infinite, number of mixture components or latent clusters. Combined with input from physically based damage growth models, the adaptively identified clusters are used in a state-space setting to effectively estimate damage states within the structure under varying external conditions. Additionally, a data selection methodology is implemented to enable judicious selection of informative measurements for maximum performance. The utility of the proposed algorithm is demonstrated by application to the estimation of fatigue-induced damage in an aluminum compact tension sample subjected to variable-amplitude cyclic loading.

## Keywords

Structural health monitoring, damage estimation, time–frequency analysis, matching pursuit decomposition, probability density function, Kullback–Leibler divergence, active data selection, adaptive learning, Dirichlet process mixture, Markov chain Monte Carlo methods, Bayesian filtering

## 1 Introduction

Structural health monitoring (SHM) (Farrar and Worden, 2007; Farrar and Lieven, 2007) is the procedure of implementing a strategy for detecting, estimating, classifying, and predicting damage in engineering structures. By helping to identify, locate, and quantify structural damage that can adversely affect the current and future performance of a system, SHM enables state-awareness of key structural components and ensures system reliability and user safety (Farrar and Worden, 2007). SHM requires non-destructive evaluation (NDE), *in situ* sensing and learning of characteristic signal features to indicate the presence of structural damage or degradation. Applications of SHM, especially for civil, mechanical and aerospace structures,

have been receiving significant attention in the scientific community in the last few decades (Staszewski et al., 2004; Farrar and Worden, 2007; Farrar and Lieven, 2007).

<sup>1</sup>Global Research and Development, General Motors Corporation, Warren, MI, USA

<sup>2</sup>School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ, USA

<sup>3</sup>School for Engineering of Matter, Transport, and Energy, Arizona State University, Tempe, AZ, USA

### Corresponding author:

Debejyo Chakraborty, Global Research and Development, General Motors Corporation, 30500 Mound Road, mail code: 480-106-RA3, Warren, Michigan 48090, USA.

Email: debejyo.chakraborty@gm.com

In the past, the practice of SHM was confined to the performing of qualitative visual inspections, offline laboratory testing, and schedule based maintenance procedures. That approach is inherently imprecise, and can lead to inefficiency from delays and possible under-utilization of resources or risk of failure from possible over-utilization. Recently, multidisciplinary interest in online SHM has replaced those methods by the use of on-board sensors, advanced signal processing, and statistical anomaly identification techniques, that enable realtime condition based damage quantification and structural residual life estimation. A considerable amount of literature now exists in the fields of mechanical, material, and structural engineering, and signal processing and statistical analysis, on methodologies for determining the presence, type, location, and intensity of structural damage (Staszewski et al., 2004; Farrar and Worden, 2007; Farrar and Lieven, 2007). Some examples include statistical methods (Doebbling and Farrar, 1998; Farrar et al., 1999), time series analysis (Sohn and Farrar, 2001), statistical pattern recognition (Sohn et al., 2001), impedance based methods (Park et al., 2003), Fourier component pair analysis (Gelman et al., 2004b), Bayesian methods (Sohn and Law, 2000; Nguyen et al., 2004), extreme value statistics (Sohn et al., 2005), support vector machines (Das et al., 2007), and other advanced time–frequency based signal processing methodologies (Karasaridis et al., 1997; Staszewski, 1998; Jeong and Jang, 2000; Sun and Chang, 2004; Michaels and Michaels, 2005; Yu and Giurgiutiu, 2005; Taha et al., 2006; Feng and Chu, 2007; Pakrashia et al., 2007; Zhou et al., 2007; Chakraborty et al., 2008a,b; Channels et al., 2008; Chakraborty et al., 2009a; Channels et al., 2009; Zhou et al., 2009a).

Much of the existing work on SHM methods relies on the use of a ‘training-testing’ paradigm for damage identification. In this approach, a set of known damage scenarios is first selected upon which the developed method is to be trained. The training process then requires a sufficient amount of data from each damage scenario to appropriately learn relevant statistical characteristics and avoid model overfitting. In the testing phase, the method generalizes from the learned training data to identify the damage for new test data. The performance of the training-testing approach hinges upon the availability of training data that is representative of the structural damage classes of interest. In particular, a key limitation of this strategy is that it cannot be used to identify previously unseen damage or operate under variable conditions. Variability arising due to changing external conditions, including temperature variation and geometry/configurational changes, affects both damage evolution and sensor data, and poses a serious challenge for the reliable identification of structural damage using conventional statistical methods (Sohn, 2007).

In this paper, we propose a novel adaptive learning structural damage estimation method (Chakraborty et al., 2009b) to address the drawbacks of the traditional training-testing based damage identification schemes. Our approach achieves the desired adaptivity by leveraging the power of non-parametric Bayesian analysis techniques (Jordan, 2010a,b). Non-parametric Bayesian techniques, unlike their parametric counterparts, afford highly flexible stochastic representations by enabling models with unlimited degrees of freedom (an infinite number of parameters). Even though the number of possible parameters may be large for a finite amount of data, with appropriate stochastic process priors the model complexity can be managed effectively in a Bayesian framework for sound inference. In this work, we make use of a non-parametric Bayesian analysis technique known as Dirichlet process (DP) mixture modeling (Ferguson, 1973; Blackwell and MacQueen, 1973; Antoniak, 1974; Sethuraman, 1994; Escobar and West, 1995). The DP induces a prior distribution over parameters and over partitions of parameter space, that can be combined with a data likelihood to obtain a statistical mixture model with a countably infinite number of mixture components, namely the DP mixture model. The DP mixture model essentially provides a stochastic description for data with an unbounded number of mixture components or latent clusters. Due to its remarkably flexible modeling properties, the DP has recently found diverse applications, such as target detection (Ni et al., 2007a,b), music analysis (Qi et al., 2007a,b), protein modeling (Ting et al., 2010), and speaker diarization (Fox et al., 2011). In this paper, we utilize DP Gaussian mixture modeling (GMM) to adaptively learn in an unsupervised manner the latent clusters within periodically collected structural data (guided wave sensor measurements). Specifically, we first extract joint time–frequency (TF) (Cohen, 1994; Mallat, 1998; Papandreou-Suppappola, 2002) features from structural signals to capture dispersive wave physics effects, and then employ the adaptively identified DP-GMM clusters in statistical changes therein to monitor structural damage evolving under variable conditions. In particular, the dynamic clustering in the data is associated with changing damage and/or operating conditions. The DP-GMM is well suited for this problem, since

- it makes no restrictive assumption about a finite and *a priori* known number of clusters;
- the appropriate number of clusters is learned automatically from data;
- the clusters can change over time with the availability of new data;
- the algorithm can be implemented efficiently for online processing.

This adaptive information is then incorporated into a state-space setting along with physically based damage

evolution models, and a Bayesian filter is used to effectively estimate progressive structural damage states online under varying external conditions. The DP-GMM parameters are learned efficiently using Markov chain Monte Carlo (MCMC) techniques (Gilks et al., 1996). To our knowledge, such an adaptive clustering algorithm has never before been integrated into a sequential Bayesian estimation framework with constantly changing data. Also, the Bayesian filter is modified to use a model for the relationship between statistical changes in TF information and changes in damage. In addition, we describe a data selection methodology for intelligently selecting the most informative measurements to use in the learning process (MacKay, 1992). The data selection procedure is formulated to select appropriate examples of signals using TF similarity measure features maintained in a first-in-first-out (FIFO) buffer in order to define a reference for the detection of statistical changes needed in the adaptive learning structural damage estimation framework (Chakraborty et al., 2010). A block diagram of the adaptive learning structural damage estimation method is shown in Figure 1. The proposed adaptive learning method is attractive for real-world SHM applications because no prior training is necessary and damage can be estimated in previously unseen conditions. We demonstrate the algorithm with an application for estimating fatigue crack damage in an aluminum compact tension (CT) specimen under variable-amplitude cyclic loading conditions.

The paper is organized as follows. In Section 2 we discuss in detail the proposed adaptive learning

structural damage estimation method, providing an overview of the theoretical aspects of its various components. Results from application of the method for fatigue crack damage estimation are shown in Section 3. This is followed by concluding remarks in Section 4. Appendices A and B contain details of discretized log-normal and reparameterized negative binomial probability mass functions (pmfs) utilized in the models.

## 2 Adaptive learning structural damage estimation method

### 2.1 TF feature extraction

Guided wave propagation in structures is known to be a dispersive phenomenon and the wave velocity varies as a function of the product of frequency and thickness (Rose, 1999). Since wave propagation velocity translates to time-delay (for a given transmitter and receiver configuration), the measured sensor signals exhibit time-varying frequency content. The appropriate tool for studying such signals is joint TF analysis (Cohen, 1994; Mallat, 1998; Papandreou-Suppappola, 2002). Furthermore, the presence of damage (e.g. cracks) affects the wave propagation by introducing additional scatterers and by changing thickness and causing mode conversion (Soni et al., 2010). By monitoring the interplay of these variables using TF analysis of the sensor signals, structural damage can be detected. Note that damage growth is assumed to occur on a different (slower) time scale. The use of TF based methods for SHM has been investigated in Chakraborty et al.

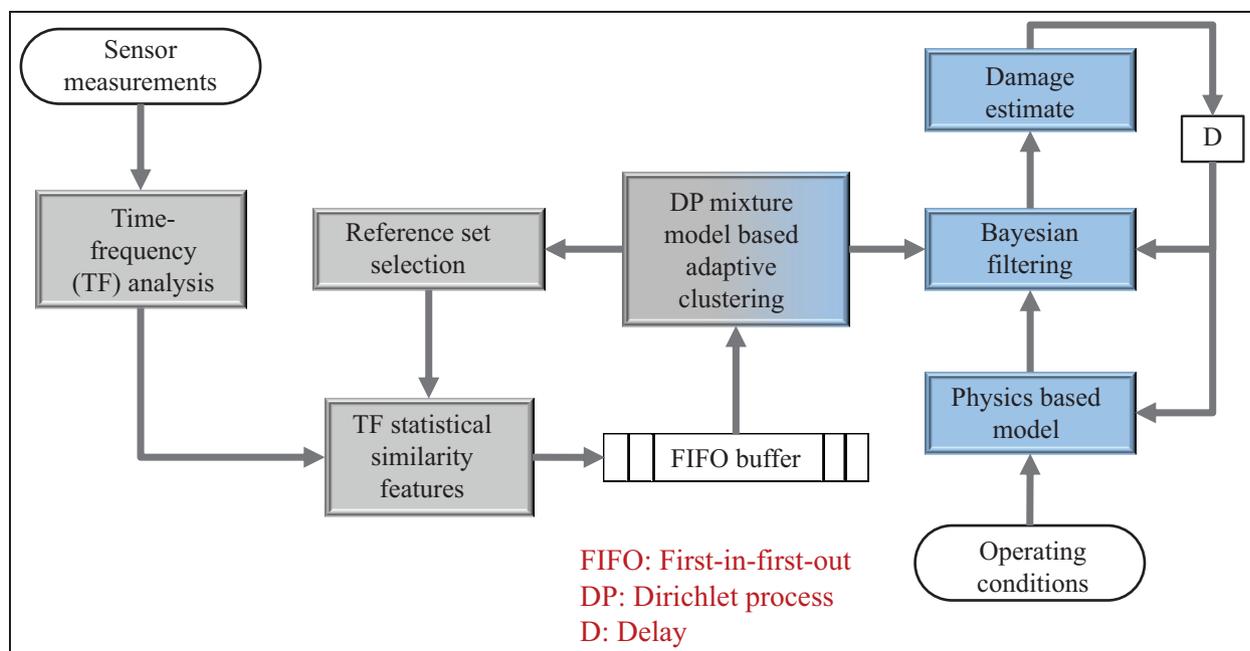


Figure 1. Block diagram of the adaptive learning method for estimating progressive structural damage.

(2008b, 2009a); Zhou et al. (2009a); Xu et al. (2009); Zhou et al. (2009b); Soni et al. (2010).

The matching pursuit decomposition (MPD) (Mallat and Zhang, 1993; Mallat, 1998) is an iterative algorithm that can be employed to efficiently yield approximate representations for signals in terms of a linear combination of basis functions:

$$s(t) \approx \sum_{i=1}^N \alpha_i g_i(t) \quad (1)$$

where  $s(t)$  is a finite-energy time-domain signal ( $t$  denotes time),  $\alpha_i$  are the MPD expansion coefficients,  $g_i(t)$  are the basis functions or ‘atoms’, and  $N$  is the number of terms in the representation. The atoms are selected one-at-a-time in an iterative fashion from an overcomplete dictionary to match components of interest in the signal. The error in the approximation decreases with increasing iterations  $N$ . For specific algorithmic details of the MPD we refer the reader to Mallat and Zhang (1993) and Mallat (1998).

The MPD is very effective for obtaining signal representations in general overcomplete dictionaries. In particular, it is well suited for the representation of signals in dictionaries comprised of TF atoms, and has been used in many TF analysis applications (Papandreou-Suppappola, 2002). In this work, we effect a TF representation of the structural signals by making use of the MPD with a dictionary of real TF shifted and scaled, Gaussian windowed tones (Mallat and Zhang, 1993; Mallat, 1998) of the form:

$$g^{(d)}(t) = \left(\frac{8\kappa_r}{\pi}\right)^{\frac{1}{4}} \exp(-\kappa_r(t - \tau_m)^2) \cos(2\pi\nu_i t) \quad (2)$$

where  $d = (\tau_m, \nu_i, \kappa_r)$  is a parametric representation of an atom, specified by its time shift  $\tau_m$ , frequency shift  $\nu_i$ , and time-scale  $\kappa_r > 0$ . Note that the dictionary atoms are normalized to unit energy:  $\|g^{(d)}(t)\|_2^2 = 1$ . The Gaussian window is chosen because of its desirable TF localization properties (Mallat, 1998), and also because it affords closed-form analytical expressions that simplify analysis and computation. A direct result of applying the MPD with these TF atoms is a parametric representation of the signal  $s(t)$  with a 4-D feature vector of length  $N$ :

$$s(t) \Leftrightarrow \left\{ \begin{bmatrix} \alpha_1 \\ \tau_1 \\ \nu_1 \\ \kappa_1 \end{bmatrix}, \begin{bmatrix} \alpha_2 \\ \tau_2 \\ \nu_2 \\ \kappa_2 \end{bmatrix}, \dots, \begin{bmatrix} \alpha_N \\ \tau_N \\ \nu_N \\ \kappa_N \end{bmatrix} \right\} \quad (3)$$

The MPD of the signal  $s(t)$  also yields a cross-term free time–frequency representation (TFR) (Cohen, 1989; Mallat and Zhang, 1993):

$$\mathcal{E}_s(t, f) \triangleq \sum_{i=1}^N |\alpha_i|^2 \text{WD}_{g_i(t, f)} \quad (4)$$

where  $\text{WD}_{g_i(t, f)}$  denotes the Wigner distribution (Cohen, 1989) of the Gaussian atom  $g_i(t)$ . The Wigner distribution of Gaussian atoms of the form in (2) is given analytically by

$$\text{WD}_{g_i(t, f)} = 2 \exp(-2\kappa_r(t - \tau_m)^2) \exp\left(-\frac{2\pi^2(f - \nu_i)^2}{\kappa_r}\right) \quad (5)$$

The TF correlation between two signals can be computed by correlating their MPD-TFRs.

The Gaussian atom selected from the dictionary at the  $i$ th MPD iteration is  $g_i(t)$ , with parameters  $(\tau_i, \nu_i, \kappa_i)$ , and its Wigner distribution is a 2-D Gaussian function (equation (5)) that can be interpreted as an unnormalized 2-D Gaussian probability density function (pdf) with mean vector  $\begin{bmatrix} \tau_i \\ \nu_i \end{bmatrix}$  and cov-

ariance matrix  $\begin{bmatrix} \frac{1}{4\kappa_i} & 0 \\ 0 & \frac{\kappa_i}{4\pi^2} \end{bmatrix}$ . Based on this interpretation, we define the MPD based pdf (MPD-pdf) for the signal  $s(t)$  as

$$P_s(t, f) = \frac{1}{Z} \sum_{i=1}^N |\alpha_i|^2 \mathcal{N}\left(\begin{bmatrix} \tau_i \\ \nu_i \end{bmatrix}, \begin{bmatrix} \frac{1}{4\kappa_i} & 0 \\ 0 & \frac{\kappa_i}{4\pi^2} \end{bmatrix}\right) \quad (6)$$

where  $\mathcal{N}(\cdot, \cdot)$  denotes a Gaussian pdf and  $Z = \sum_{i=1}^N |\alpha_i|^2$  is the normalizing constant. The MPD-pdf in (6) maps time-domain signals to Gaussian pdfs in TF space, and provides a useful tool for computing TF statistical similarity measures between the wave based sensor measurements obtained from structures.

## 2.2 A statistical similarity measure

The distance between two pdfs can be quantified by well-known statistical measures, such as Kullback–Leibler (KL) divergence (MacKay, 2003), Bhattacharyya distance (Bhattacharyya, 1943; Kailath, 1967), and Hellinger distance (Le Cam and Yang, 2000). The KL divergence is defined for 2-D pdfs  $p(t, f)$  and  $q(t, f)$  as

$$\mathcal{D}^{\text{KL}}(p||q) \triangleq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(t, f) \log \frac{p(t, f)}{q(t, f)} dt df \quad (7)$$

The KL divergence is non-negative, i.e.  $\mathcal{D}^{\text{KL}}(p||q) \geq 0$ , and becomes zero if and only if  $p(t, f) = q(t, f)$ . However, the measure is non-symmetric:  $\mathcal{D}^{\text{KL}}(p||q) \neq \mathcal{D}^{\text{KL}}(q||p)$ , and its computation involves the evaluation of logarithms, which can be expensive.

In this work, we consider a correlation based statistical measure of similarity between pdfs  $p(t, f)$  and  $q(t, f)$ , defined as

$$\mathcal{D}^{\text{corr}}(\mathbf{p}||\mathbf{q}) \triangleq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{p}(t,f)\mathbf{q}(t,f) dt df \quad (8)$$

The correlation based statistical similarity measure has the following properties:

- (a) It is non-negative, and from the Cauchy–Schwarz inequality,  $0 \leq \mathcal{D}^{\text{corr}}(\mathbf{p}||\mathbf{q}) \leq \|\mathbf{p}\|_2 \|\mathbf{q}\|_2$ .
- (b)  $\mathcal{D}^{\text{corr}}(\mathbf{p}||\mathbf{q}) = 0$  if and only if  $\mathbf{p}(t,f)$  and  $\mathbf{q}(t,f)$  are orthogonal.
- (c)  $\mathcal{D}^{\text{corr}}(\mathbf{p}||\mathbf{q}) = \|\mathbf{p}\|_2^2 = \|\mathbf{q}\|_2^2$  if and only if  $\mathbf{p}(t,f) = \mathbf{q}(t,f)$ .
- (d) It is a symmetric measure of similarity:  $\mathcal{D}^{\text{corr}}(\mathbf{p}||\mathbf{q}) = \mathcal{D}^{\text{corr}}(\mathbf{q}||\mathbf{p})$ .
- (e) It is a linear measure of statistical similarity.

The correlation measure  $\mathcal{D}^{\text{corr}}(\mathbf{p}||\mathbf{q})$  can be calculated by using Monte Carlo integration. Let  $\{(t^{(l)}, f^{(l)})\}_{l=1}^L \sim \mathbf{p}(t,f)$  be  $L$  i.i.d. samples drawn from the pdf  $\mathbf{p}(t,f)$ . Then, an unbiased estimator for  $\mathcal{D}^{\text{corr}}(\mathbf{p}||\mathbf{q})$  is given by

$$\hat{\mathcal{D}}^{\text{corr}}(\mathbf{p}||\mathbf{q}) = \frac{1}{L} \sum_{l=1}^L q(t^{(l)}, f^{(l)}) \quad (9)$$

which approaches  $\mathcal{D}^{\text{corr}}(\mathbf{p}||\mathbf{q})$  almost surely as  $L \rightarrow \infty$ . For the Gaussian pdfs and their mixtures used here, closed-form expressions can be derived for (8).

The correlation based measure is used with the MPD-pdfs described in the previous subsection to determine the TF statistical similarity between structural signals. Specifically, given a test signal  $s^{\text{te}}(t)$  and a reference signal set  $\mathbf{S}^{\text{ref}} = \{s_1^{\text{ref}}(t), s_2^{\text{ref}}(t), \dots, s_R^{\text{ref}}(t)\}$  of  $R$  signals, with respective MPD-pdfs  $P_{s^{\text{te}}}(t,f)$  and  $P_{s_r^{\text{ref}}}(t,f)$  for  $r = 1, \dots, R$ , computed first using (6), the average TF similarity  $\mathcal{D}^{\text{avg}}$  of  $P_{s^{\text{te}}}(t,f)$  to the MPD-pdfs of the reference set is computed as

$$\mathcal{D}^{\text{avg}} = \frac{1}{R} \sum_{r=1}^R \mathcal{D}^{\text{corr}}(P_{s^{\text{te}}}||P_{s_r^{\text{ref}}}) = \mathcal{D}^{\text{corr}}(P_{s^{\text{te}}}||P_{\mathbf{S}^{\text{ref}}}), \quad (10)$$

where  $P_{\mathbf{S}^{\text{ref}}}(t,f) = \frac{1}{R} \sum_{r=1}^R P_{s_r^{\text{ref}}}(t,f)$  by linearity of  $\mathcal{D}^{\text{corr}}(\cdot || \cdot)$ . The pdf  $P_{\mathbf{S}^{\text{ref}}}(t,f)$  is the average MPD-pdf that collects the TF statistical variability in the reference set  $\mathbf{S}^{\text{ref}}$ . Note that  $P_{\mathbf{S}^{\text{ref}}}(t,f)$  is a Gaussian mixture because the MPD dictionary is comprised of Gaussian atoms. This scalar TF similarity measure for  $s^{\text{te}}(t)$  with respect to a suitable  $\mathbf{S}^{\text{ref}}$  is employed to efficiently monitor statistical changes in the TF information due to changes in the structural damage condition, and forms the data for the DP mixture model based adaptive clustering algorithm.

### 2.3 Minimum discrepancy reference set selection

Continuous monitoring of measured sensor signal TF similarity to a reference signal set can provide an

indicator for progressive anomalies in structural condition, with the effectiveness in identifying new anomalies depending on the reference set used. In particular, the reference set needs to be constantly updated to include information obtained from new data. The computed scalar TF similarity measure features are maintained in a first-in-first-out (FIFO) buffer, i.e. at each time epoch (new collection of sensor signals) the oldest feature in the buffer is replaced by the most recent one. The data in the buffer is refreshed constantly and there is always a mix of old and newly acquired data. The reference signal set  $\mathbf{S}^{\text{ref}}$  is then defined such that, at each time epoch  $k$ , the resulting TF similarity features are representative of signals observed up to time  $k$ . Thus, if the DP mixture model based adaptive clustering were applied to these features, a single cluster would be identified. Further, based on the TF similarity of the new sensor signals collected at the following epoch  $k+1$  to this reference set, the subsequent DP clustering can correctly identify anomalies if present (via new clusters). The reference signal set  $\mathbf{S}^{\text{ref}}$  is obtained here by selecting uniformly distributed features from the FIFO buffer as described below.

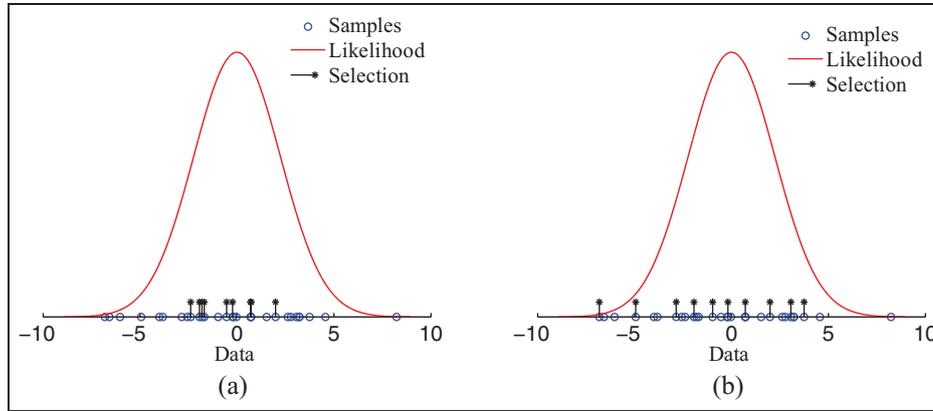
The goal of uniform data selection is to find a subset of uniformly distributed samples from a given set. Figure 2 shows an example for a synthetic dataset comprising Gaussian distributed samples. The selection in Figure 2(a) is clustered near the mean of the Gaussian pdf (from which the data was obtained), whereas that in Figure 2(b) is distributed more uniformly over the data space. Thus, the data selection in Figure 2(b) is more uniform than that in Figure 2(a).

The uniformity of a set  $\mathbf{Y} = \{y_1, \dots, y_T\}$  of size  $T$  in an interval  $[y^{\min}, y^{\max}]$  can be quantified by the discrepancy (Kuipers and Niederreiter, 1974)  $\mathcal{D}_T(\mathbf{Y})$ , defined as

$$\mathcal{D}_T(\mathbf{Y}) \triangleq \sup_{y^{\min} \leq y^l \leq y^h \leq y^{\max}} \frac{|\mathbf{Y} \cap [y^l, y^h]|}{T} - \frac{y^h - y^l}{y^{\max} - y^{\min}} \quad (11)$$

where  $[y^l, y^h]$  is any subinterval of  $[y^{\min}, y^{\max}]$  and  $|\mathbf{Y} \cap [y^l, y^h]|$  is the cardinality of  $\mathbf{Y}$  in  $[y^l, y^h]$ .  $\mathbf{Y}$  is said to be uniformly distributed on  $[y^{\min}, y^{\max}]$  if  $\lim_{T \rightarrow \infty} \mathcal{D}_T(\mathbf{Y}) = 0$  (Kuipers and Niederreiter, 1974; Chazelle, 2002). For a given  $\mathbf{Y}$  in  $[y^{\min}, y^{\max}]$ , the discrepancy  $\mathcal{D}_T(\mathbf{Y})$  can be used as a measure of uniformity. In particular, the discrepancy is minimum for a uniform  $\mathbf{Y}$  in  $[y^{\min}, y^{\max}]$ . In the example shown in Figure 2, the discrepancy of the set of selected samples is greater in (a) than that in (b).

Let  $\mathbf{Y}^{\text{FIFO}} = \{y_1, \dots, y_T\}$  denote the set of scalar TF similarity features in the FIFO buffer (see Section 2.2). We wish to select a uniformly distributed subset of features  $\mathbf{Y}^{\text{ref}}$  from  $\mathbf{Y}^{\text{FIFO}}$ . We define a vector  $\mathbf{z} = \{z_1, \dots, z_T\}$  of associated Boolean variables to indicate the selected features:



**Figure 2.** The data selection in (a) is clustered tightly near the mean of the Gaussian pdf (from which the data was obtained), whereas that in (b) is distributed more uniformly over the data space. Thus, the data selection in (b) is more uniform than that in (a).

$$z_n = \begin{cases} 1, & \text{if } y_n \text{ is selected} \\ 0, & \text{otherwise} \end{cases}, n = 1, \dots, T \quad (12)$$

Given a feature set  $\mathbf{Y}^{\text{FIFO}} = \{y_1, \dots, y_T\}$  of size  $T$  in the interval  $[y^{\min}, y^{\max}]$ , the uniform feature selection problem is then to find the selection subset  $\mathbf{Y}^{\text{ref}}$  of size  $\hat{T}$  which minimizes the discrepancy  $\mathcal{D}_{\hat{T}}(\mathbf{Y}_z^{\text{FIFO}})$ . Here,  $\mathbf{Y}_z^{\text{FIFO}}$  represents a subset of features  $y_n \in \mathbf{Y}^{\text{FIFO}}$  for which the indicator variables  $z_n = 1$ . The desired selection is obtained by solving the optimization problem

$$\begin{aligned} & \text{minimize } \mathcal{D}_{\hat{T}}(\mathbf{Y}_z^{\text{FIFO}}) \\ & \text{subject to } z_n \in \{0, 1\}, n = 1, \dots, T \\ & \|\mathbf{z}\|_0 = \hat{T} \end{aligned} \quad (13)$$

where  $\|\mathbf{z}\|_0$  represents the cardinality of the vector  $\mathbf{z}$ . The problem in (13) is non-convex and generally very hard to solve exactly (Boyd and Vandenberghe, 2004). Direct enumeration is not feasible since the number of possible selections  $\binom{T}{\hat{T}}$  is very large, even for moderate  $T$  and  $\hat{T}$ .

Techniques for evaluating the discrepancy, including those based on sampling, searching, lower and upper bounds, and convex programming (Niederreiter, 1972; Dobkin and Eppstein, 1993; Dobkin et al., 1996; Chazelle, 2002), can also be computationally demanding.

In this paper, an approximate and efficient method is proposed for uniform feature selection by considering the relaxed optimization problem

$$\begin{aligned} & \text{minimize } \tilde{\mathcal{D}}_{\hat{T}}(\mathbf{Y}_z^{\text{FIFO}}) \\ & \text{subject to } z_n \in \{0, 1\}, n = 1, \dots, T \\ & \|\mathbf{z}\|_0 \approx \hat{T} \end{aligned} \quad (14)$$

in which the objective function  $\tilde{\mathcal{D}}_{\hat{T}}(\mathbf{Y}_z^{\text{FIFO}})$  is defined as

$$\tilde{\mathcal{D}}_{\hat{T}}(\mathbf{Y}_z^{\text{FIFO}}) \triangleq \frac{1}{H} \sum_{h=1}^H \left| \frac{|\mathbf{Y}_z^{\text{FIFO}} \cap [y^{\min} + (h-1)\Delta y, y^{\min} + h\Delta y]|}{\hat{T}} - \frac{\Delta y}{y^{\max} - y^{\min}} \right| \quad (15)$$

where  $\Delta y = (y^{\max} - y^{\min})/H$ , with  $H$  the number of equal-length partitions of  $[y^{\min}, y^{\max}]$ . Whereas the discrepancy in (11) is a supremum over all possible subintervals of  $[y^{\min}, y^{\max}]$ , the objective function in (14) is defined as an average over the  $H$  equal-length subintervals partitioning  $[y^{\min}, y^{\max}]$ . The constraint on the precise number of features to be selected has also been relaxed. A simple and efficient algorithm for the solution of (14) for uniform feature selection is given in Algorithm 1. The algorithm attempts to minimize the summation in (15) term-by-term by selecting roughly  $\text{round}(\hat{T}/H)$  features independently in each of the  $H$  equal-length partitions.

## 2.4 Adaptive clustering using DP mixture models

As mentioned earlier, the scalar TF similarity measure features are maintained in a constantly refreshed FIFO buffer containing a mix of old and newly acquired data. Adaptive clustering is then performed on the buffered features via the DP-GMM, with the number of clusters used to assess the extent of variability in the TF statistics of the sensor measurements. Unlike conventional parametric mixture models that assume a fixed number of mixture components or clusters and may lead to over-fitting or under-fitting of the data, DP based mixture modeling provides the capability of automatically determining the correct number of clusters needed for characterizing the data. In particular, if only a single

---

**Algorithm 1** Uniform feature selection via solution of problem (14).

---

*Input:* Feature set  $\{y_1, \dots, y_T\}$  of size  $T$  in  $[y^{\min}, y^{\max}]$ , number of features to be selected  $\hat{T}$ , number of partitions  $H$ .

*Output:* Selection vector  $\mathbf{z}$ .

---

- (1) Initialize selection vector.  
Assign  $z_n = 0$ , for  $n = 1, \dots, T$ .
  - (2) Determine approximate number of features to be selected in each partition.  
Set  $\hat{N}_H = \mathbf{round}(\hat{T}/H)$ .
  - (3) Now select approximately  $\hat{N}_H$  features independently in each partition.  
**for** partition  $h = 1$  to  $H$  **do**  
 Indices of features in partition:  $\mathbf{I} = \{n : y_n \in [y^{\min} + (h-1)\Delta y, y^{\min} + h\Delta y]\}$ .  
 Number of features in partition:  $N_h = |\mathbf{I}|$ .  
 If  $N_h \leq \hat{N}_H$ , select all features in partition:  $\mathbf{z}_\mathbf{I} = \mathbf{1}$ .  
 Otherwise, randomly select  $\hat{N}_H$  features in partition:  
**for**  $j = 1$  to  $\hat{N}_H$  **do**  
   Draw  $u \sim U[0, 1]$   
    $w \leftarrow \lceil u \cdot N_h \rceil$   
    $\mathbf{z}_{\mathbf{I}(w)} = 1$   
    $\mathbf{I} \leftarrow \mathbf{I} \setminus \mathbf{I}(w)$   
    $N_h \leftarrow N_h - 1$   
**end for**  
**end for**
- 

cluster is identified in the data, this implies that the old and new TF similarity features, present concurrently at that epoch in the buffer, are statistically similar, and that the measured signals did not exhibit any changes indicative of a change in the damage state and/or operating conditions. The discovery of more than one cluster, however, indicates the opposite. In this subsection we describe the analytical framework of DP mixture modeling and discuss a Gibbs sampling algorithm for efficient inference in DP-GMMs.

**2.4.1 DP mixture modeling.** Consider a general mixture model of the form

$$P(\mathbf{y}|\mathbf{p}, \Theta^*) = \sum_{m=1}^M p_m f(\mathbf{y}|\theta_m^*) \quad (16)$$

where  $\mathbf{y}$  denotes data,  $\mathbf{p} = \{p_1, \dots, p_M\}$  is a set of positive mixture weights that sum to unity,  $\Theta^* = \{\theta_1^*, \dots, \theta_M^*\}$  is a set of parameters,  $f(\mathbf{y}|\theta_m^*)$  represents a pdf parametrized by  $\theta_m^*$ , and  $M$  is the number of mixture components. Given a data set  $\mathbf{Y} = \{y_1, \dots, y_T\}$  of size  $T$ , the modeling task is to learn the appropriate mixture size  $M$ , the mixture weights  $\mathbf{p}$ , and the mixture parameters  $\Theta^*$ .

If the number of mixture components  $M$  is known, then classical maximum-likelihood (ML) learning techniques can be applied to estimate the parameters of interest (MacKay, 2003; Gelman et al., 2004a). In

particular, the iterative expectation-maximization (EM) algorithm can be used to estimate the parameters by locally maximizing the data likelihood (Dempster et al., 1977; Duda et al., 2001; MacKay, 2003). The EM algorithm introduces latent variables into the model and iterates between inferring the posterior pdf over the latent variables given a parameter setting and computing an updated parameter estimate by maximizing the likelihood using the learned latent variable statistics.

When the number of mixture components  $M$  is not known *a priori*, and in particular needs to be adaptable with the data, Bayesian nonparametrics (Jordan, 2010a,b) provides a versatile mixture modeling approach through the DP (Blackwell and MacQueen, 1973; Ferguson, 1973; Antoniak, 1974; Sethuraman, 1994; Escobar and West, 1995). The DP yields mixture models with a potentially unlimited number of mixture components, the appropriate number of which is automatically determined from given data.

A DP, denoted by  $DP(\iota, G_0)$ , has two parameters: a positive scalar innovation parameter  $\iota$  and a base distribution  $G_0$ . The DP is a distribution over distributions. Let  $G$  be a distribution drawn from a DP,

$$G \sim DP(\iota, G_0) \quad (17)$$

then the expected value of  $G$  is  $G_0$  and the closeness of  $G$  to  $G_0$  is controlled by  $\iota$ . Distributions drawn from a DP are discrete with probability 1, i.e.  $G$  in (17) is

almost surely discrete (even if  $G_0$  is continuous). This property can be used to induce a sharing of parameters as follows. Suppose that  $T$  random variables  $\{\theta_1, \dots, \theta_T\}$  are distributed according to  $G$ :

$$G \sim DP(\iota, G_0) \quad (18a)$$

$$\theta_n | G \sim G, n = 1, \dots, T \quad (18b)$$

The discreteness of  $G$  implies that the variables  $\{\theta_1, \dots, \theta_T\}$  coincide in value with positive probability. It can be shown, by integrating out the distribution  $G$ , that the conditional density function of the variable  $\theta_n$  given all other variables (denoted by  $\Theta^{-n}$ ) is given by the Pólya urn relation (Blackwell and MacQueen, 1973; Ferguson, 1973; Antoniak, 1974; West et al., 1994; Escobar and West, 1995)

$$P(\theta_n | \Theta^{-n}, \iota, G_0) = \frac{1}{\iota + T - 1} \sum_{\substack{n'=1 \\ n' \neq n}}^T \delta_{\theta_n, \theta_{n'}} + \frac{\iota}{\iota + T - 1} G_0(\theta_n) \quad (19)$$

where  $\delta$  is the Kronecker-delta function. (19) can be rewritten as

$$P(\theta_n | \Theta^{-n}, \iota, G_0) = \frac{1}{\iota + T - 1} \sum_{m=1}^M N_m^{-n} \delta_{\theta_n, \theta_m^*} + \frac{\iota}{\iota + T - 1} G_0(\theta_n) \quad (20)$$

where  $\{\theta_1^*, \dots, \theta_M^*\}$  are the distinct values in  $\{\theta_1, \dots, \theta_T\}$  and  $N_m^{-n}$  denotes the number of variables in  $\Theta^{-n}$  equal to  $\theta_m^*$ . Thus, the DP prior implies that each variable  $\theta_n$  either assumes an existing value  $\theta_m^*$  with probability  $N_m^{-n}/(\iota + T - 1)$  or is a fresh draw from  $G_0$  with probability  $\iota/(\iota + T - 1)$ . The innovation parameter  $\iota$  controls the tradeoff between the reuse of existing values and fresh draws. In the limit as  $\iota \rightarrow \infty$ ,  $G$  approaches  $G_0$  so that all draws are i.i.d.  $G_0$ . It can be shown that in this description the number of distinct values  $M$  grows as  $\mathcal{O}(\log T)$ . Also, under the DP prior, the variables  $\{\theta_1, \dots, \theta_T\}$  are exchangeable, i.e. their joint distribution is unaffected by reordering.

In Sethuraman (1994), an explicit characterization of  $G$  is provided via the ‘stick-breaking’ construction:

$$\theta_m^* \sim G_0, m = 1, \dots, \infty \quad (21a)$$

$$v_k \sim \text{Beta}(1, \iota), k = 1, \dots, \infty \quad (21b)$$

$$p_m = v_m \prod_{k=1}^{m-1} (1 - v_k), m = 1, \dots, \infty \quad (21c)$$

$$G(\theta) = \sum_{m=1}^{\infty} p_m \delta_{\theta, \theta_m^*} \quad (21d)$$

This representation shows that  $G$  is discrete. Here,  $p_m^* = \Pr(\theta = \theta_m^*)$  can be viewed as the size of the pieces

of a unit-length stick broken successively an infinite number of times.

Consider a hierarchical Bayesian model in which the DP is used as a non-parametric prior:

$$G \sim DP(\iota, G_0) \quad (22a)$$

$$\theta_n | G \sim G, n = 1, \dots, T \quad (22b)$$

$$y_n | \theta_n \sim f(y_n | \theta_n), n = 1, \dots, T \quad (22c)$$

In this specification, there is a one-to-one association between each observed data point  $y_n$  and latent variable  $\theta_n$ . In particular, the data point  $y_n$  is drawn from a pdf  $f(y_n | \theta_n)$  that is parametrized by the associated latent variable  $\theta_n$ . Owing to the fact that the parameters coincide in values ( $G$  is discrete), a corresponding clustering is induced on the data. In view of the stick-breaking construction (21), (22) can be rewritten as

$$\theta_m^* \sim G_0, m = 1, \dots, \infty \quad (23a)$$

$$v_k \sim \text{Beta}(1, \iota), k = 1, \dots, \infty \quad (23b)$$

$$p_m = v_m \prod_{k=1}^{m-1} (1 - v_k), m = 1, \dots, \infty \quad (23c)$$

$$c_n | \mathbf{p} \sim \text{Categorical}(\mathbf{p}), n = 1, \dots, T \quad (23d)$$

$$y_n | c_n \sim f(y_n | \theta_{c_n}^*), n = 1, \dots, T \quad (23e)$$

Note the introduction of latent variables  $c_n$  that indicate the clusters to which data  $y_n$  are assigned. (23) is equivalent to an infinite mixture model termed the *DP mixture model*:

$$P(\mathbf{y} | \mathbf{p}, \Theta^*) = \sum_{m=1}^{\infty} p_m f(\mathbf{y} | \theta_m^*) \quad (24)$$

It was shown in Ishwaran and James (2001) that the DP mixture model (24) can be truncated to  $K$  terms by setting  $v_K = 1$  and the truncated Dirichlet process (TDP), denoted by  $DP_K(\iota, G_0)$ , provides an approximation for the original DP  $DP(\iota, G_0)$ . The truncation limit  $K$  may be selected based on the error in the marginal density  $Q(\mathbf{Y})$ :

$$\|Q_K - Q_\infty\|_1 \approx 4T \exp(- (K - 1)/\iota) \quad (25)$$

For a fixed  $\iota$ , the error decreases exponentially with  $K$  but increases only linearly with dataset size  $T$ , indicating that  $K$  does not have to be large for accurately modeling even big datasets (Ishwaran and James, 2001). Note that the actual number of mixture components (latent clusters) required in representing a given dataset can be smaller than  $K$  (only a few entries of  $\mathbf{p}$  may be significant).

**2.4.2 Blocked Gibbs sampling for inference in DP Gaussian mixture models.** In Bayesian inference, the goal is to find the posterior pdf over the model parameters given the observed dataset. This posterior pdf can be computed

using MCMC methods (Gilks et al., 1996), in which a Markov chain is constructed in order to generate samples from the pdf of interest. In particular, the collected MCMC samples provide an empirical estimate of the posterior pdf over the parameters, that can be used to approximate various posterior expectations of interest. The MCMC method of Gibbs sampling (Gilks et al., 1996) realizes the Markov chain by iteratively sampling each random variable conditioned on the data and the previously sampled values of all other variables, and is suitable for problems where it is easy to draw from the relevant conditional densities.

An efficient blocked Gibbs sampling method for inference in DP mixture models is described in Ishwaran and James (2001). In the blocked Gibbs sampler of Ishwaran and James (2001), parameters of the TDP are updated in blocks to generate an empirical estimate of the posterior pdf  $P(\Theta^*, \mathbf{c}, \mathbf{p} | \mathbf{Y})$  through its samples. Specifically, samples are iteratively drawn from the conditional pdfs (Ishwaran and James, 2001)

$$\theta_m^{*(l)} \sim P(\theta_m^* | \mathbf{c}^{(l-1)}, \mathbf{Y}), m = 1, \dots, K \tag{26a}$$

$$c_n^{(l)} \sim P(c_n | \theta_m^{*(l)}, \mathbf{p}^{(l-1)}, \mathbf{Y}), n = 1, \dots, T \tag{26b}$$

$$p_m^{(l)} \sim P(p_m | \mathbf{c}^{(l)}), m = 1, \dots, K \tag{26c}$$

where  $l = 1, 2, \dots$  denotes the Gibbs iteration number in the Markov chain:

$$\{\theta^{*(1)}, \mathbf{c}^{(1)}, \mathbf{p}^{(1)}\} \rightarrow \{\theta^{*(2)}, \mathbf{c}^{(2)}, \mathbf{p}^{(2)}\} \rightarrow \dots \rightarrow \{\theta^{*(l)}, \mathbf{c}^{(l)}, \mathbf{p}^{(l)}\} \rightarrow \dots \tag{27}$$

The conditional pdfs in (26) are given by (Ishwaran and James, 2001)

$$P(\theta_m^* | \mathbf{c}, \mathbf{Y}) \propto G_0(\theta_m^*) \prod_{n: c_n = m} f(\mathbf{y}_n | \theta_m^*), m = 1, \dots, K \tag{28a}$$

$$P(c_n | \theta_m^*, \mathbf{p}, \mathbf{Y}) \propto \sum_{m=1}^K \{p_m f(\mathbf{y}_n | \theta_m^*)\} \delta_{c_n, m}, n = 1, \dots, T \tag{28b}$$

$$p_m | \mathbf{c} = v_m^* \prod_{k=1}^{m-1} (1 - v_k^*), m = 1, \dots, K \tag{28c}$$

where  $v_m^* \sim \text{Beta}\left(1 + N_m, \iota + \sum_{j=m+1}^K N_j\right)$  with  $N_m$  denoting the number of  $c_n$  values equal to  $m$ . Note that the update step for the posterior over  $\theta_m^*$  above can be performed very efficiently if the base distribution  $G_0$  is selected to be conjugate to the likelihood term  $f(\cdot | \theta_m^*)$ . Finally, the predictive distribution is approximated as

$$P(\mathbf{y}_{T+1} | \mathbf{Y}, \iota, G_0) = \frac{1}{L} \sum_{l=1}^L \left[ \sum_{m=1}^K p_m^{(l)} f(\mathbf{y}_{T+1} | \theta_m^{*(l)}) \right] \tag{29}$$

where  $L$  is the number of Gibbs samples collected.

For the case of DP-GMM of scalar data  $y$ , the pdf  $f(y | \theta) \equiv \mathcal{N}(y | \mu, \sigma^2)$  is Gaussian with mean  $\mu$  and precision  $1/\sigma^2$ :

$$P(y | \mathbf{p}, \boldsymbol{\mu}^*, 1/\boldsymbol{\sigma}^*) = \sum_{m=1}^{\infty} p_m \mathcal{N}(y | \mu_m^*, \sigma_m^{2*}) \tag{30}$$

In this paper, we utilize a Normal-Gamma form for the base distribution  $G_0$ :

$$G_0(\boldsymbol{\theta}) \triangleq \text{Normal-Gamma}\left(\boldsymbol{\mu}, \frac{1}{\sigma^2} \middle| u_\mu, u_\tau, u_a, u_b\right) = \mathcal{N}\left(\boldsymbol{\mu} \middle| u_\mu, \frac{\sigma^2}{u_\tau}\right) \text{Gamma}\left(\frac{1}{\sigma^2} \middle| u_a, u_b\right) \tag{31}$$

where  $u_\mu, u_\tau, u_a, u_b$  are hyperparameters. The Normal-Gamma prior is conjugate to the scalar Gaussian likelihood with unknown mean and precision, i.e. the posterior pdf over the mean and precision is also Normal-Gamma in form:

$$\begin{aligned} \text{Normal-Gamma}\left(\boldsymbol{\mu}, \frac{1}{\sigma^2} \middle| \tilde{u}_\mu, \tilde{u}_\tau, \tilde{u}_a, \tilde{u}_b\right) &\propto \mathcal{N}(\mathbf{Y} | \boldsymbol{\mu}, \sigma^2) \\ \text{Normal-Gamma}\left(\boldsymbol{\mu}, \frac{1}{\sigma^2} \middle| u_\mu, u_\tau, u_a, u_b\right) &\tag{32} \end{aligned}$$

where  $\mathbf{Y} = \{y_1, \dots, y_T\}$  is an observed dataset of size  $T$  and  $\tilde{u}_\mu, \tilde{u}_\tau, \tilde{u}_a, \tilde{u}_b$  are the posterior hyperparameters, given by Fink, (1995).

$$\tilde{u}_\mu = \frac{u_\tau u_\mu + T \bar{y}}{u_\tau + T} \tag{33a}$$

$$\tilde{u}_\tau = u_\tau + T \tag{33b}$$

$$\tilde{u}_a = u_a + \frac{T}{2} \tag{33c}$$

$$\tilde{u}_b = \left(\frac{1}{u_b} + \frac{T s^2}{2} + \frac{u_\tau T (\bar{y} - u_\mu)^2}{2(u_\tau + T)}\right)^{-1} \tag{33d}$$

Here,  $\bar{y}$  and  $s^2$  are the sample mean and biased sample variance of  $\mathbf{Y}$ , respectively. The conjugacy property reduces the Gibbs sampling update step for the posterior over  $\theta_m^* = \{\mu_m^*, 1/\sigma_m^{2*}\}$  in (28a) to a simple update of the Normal-Gamma hyperparameters. The blocked Gibbs sampler for inference in DP-GMM of scalar data is summarized in Algorithm 2.

We now give a simple example showing the use of the blocked Gibbs sampling algorithm for DP-GMM of a synthetic dataset of size  $T = 500$  generated from a  $M = 3$  component scalar GMM with parameters  $\mathbf{p} = [0.15, 0.5, 0.35]$ ,  $\boldsymbol{\mu}^* = [-2, 4, 8]$ , and  $\boldsymbol{\sigma}^{2*} = [0.05, 0.1, 1]$ . The DP innovation parameter was chosen as  $\iota = 1$ . The Normal-Gamma base distribution hyperparameters were  $u_\mu = 0$ ,  $u_\tau = 1$ ,  $u_a = 1$ , and  $u_b = 1$ . The truncation limit was set to  $K = 50$ . The Gibbs sampler was initialized with parameters corresponding to all data points assigned to a single cluster.  $L_b = 2000$  Gibbs iterations were performed for burn-in, followed

**Algorithm 2** Blocked Gibbs sampling for 1-D DP-GMM.

*Input:* Dataset  $\mathbf{Y} = \{y_1, \dots, y_T\}$ , DP innovation parameter  $\iota$ , Normal-Gamma hyperparameters  $u_\mu, u_\tau, u_a, u_b$ , DP truncation limit  $K$ .

*Output:* Samples  $\{\boldsymbol{\mu}^{*(l)}, 1/\boldsymbol{\sigma}^{2*(l)}, \mathbf{c}^{(l)}, \mathbf{p}^{(l)}\}_{l=1}^L$  from the posterior pdf  $P(\boldsymbol{\mu}^*, 1/\boldsymbol{\sigma}^{2*}, \mathbf{c}, \mathbf{p}|\mathbf{Y})$ .

Initialize Gibbs sampler state  $\{\boldsymbol{\mu}^{*(0)}, 1/\boldsymbol{\sigma}^{2*(0)}, \mathbf{c}^{(0)}, \mathbf{p}^{(0)}\}$ .

For iterations  $l = 1, \dots, L$ , perform the following steps:

- (1) Update for  $\{\boldsymbol{\mu}_m^{*(l)}, 1/\boldsymbol{\sigma}_m^{2*(l)}\} \sim P(\boldsymbol{\mu}_m^*, 1/\boldsymbol{\sigma}_m^{2*} | \mathbf{c}^{(l-1)}, \mathbf{Y})$ ,  $m = 1, \dots, K$ .

Define  $\mathbf{Y}_m \triangleq \{y_n : c_n^{(l-1)} = m\}$  and  $T_m = |\mathbf{Y}_m|$ , for  $m = 1, \dots, K$ .

Compute, for  $m = 1, \dots, K$ ,

$$\bar{y}_m = \frac{1}{T_m} \sum_{n: c_n^{(l-1)} = m} y_n, \quad s_m^2 = \frac{1}{T_m} \sum_{n: c_n^{(l-1)} = m} (y_n - \bar{y}_m)^2$$

$$\tilde{u}_{\mu, m} = \frac{u_\tau u_\mu + T_m \bar{y}_m}{u_\tau + T_m}, \quad \tilde{u}_{\tau, m} = u_\tau + T_m$$

Draw :  $\frac{1}{\sigma_m^{2*(l)}} \sim \text{Gamma}\left(\frac{1}{\sigma_m^{2*}} \left| \tilde{u}_{a, m}, \tilde{u}_{b, m} \right.\right)$ ,  $m = 1, \dots, K$ .

Then draw :  $\boldsymbol{\mu}_m^{*(l)} \sim \mathcal{N}\left(\boldsymbol{\mu}_m^* \left| \tilde{u}_{\mu, m}, \frac{\sigma_m^{2*(l)}}{\tilde{u}_{\tau, m}} \right.\right)$ ,  $m = 1, \dots, K$ .

- (2) Update for  $c_n^{(l)} \sim P(c_n | \boldsymbol{\mu}^{*(l)}, 1/\boldsymbol{\sigma}^{2*(l)}, \mathbf{p}^{(l-1)}, \mathbf{Y})$ ,  $n = 1, \dots, T$ .

Define  $q_{m, n} \triangleq p_m^{(l-1)} \mathcal{N}(y_n | \boldsymbol{\mu}_m^{*(l)}, \sigma_m^{2*(l)})$ , for  $m = 1, \dots, K$ ,  $n = 1, \dots, T$ .

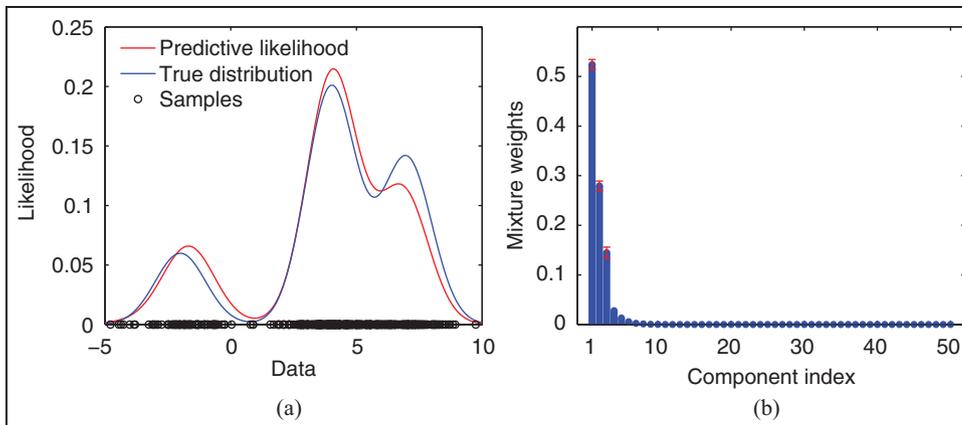
Normalize  $\tilde{q}_{m, n} = \frac{q_{m, n}}{\sum_{m'} q_{m', n}}$ ,  $m = 1, \dots, K$ ,  $n = 1, \dots, T$ .

Draw :  $c_n^{(l)} \sim \sum_{m=1}^K \tilde{q}_{m, n} \delta_{c_n, m}$ ,  $n = 1, \dots, T$ .

- (3) Update for  $p_m^{(l)} \sim P(p_m | \mathbf{c}^{(l)})$ ,  $m = 1, \dots, K$ .

Draw :  $v_m^* \sim \text{Beta}\left(1 + N_m, \iota + \sum_{j=m+1}^K N_j\right)$ ,  $m = 1, \dots, K$ , where  $N_m \triangleq |\{n : c_n^{(l)} = m\}|$ .

Then compute :  $p_m^{(l)} = v_m^* \prod_{k=1}^{m-1} (1 - v_k^*)$ ,  $m = 1, \dots, K$ .



**Figure 3.** Example DP-GMM results for scalar synthetic data.

by  $L_s = 1000$  iterations for collecting samples. Figure 3 shows a comparison of the true and learned GMMs and the estimated mixture weights. It can be seen that the algorithm correctly identified the three clusters in the data. The simulation took about 66 secs to run on a 2 GHz processor.

## 2.5 Bayesian filtering

The adaptive DP clustering results are next combined with information from a physically based progressive damage model in a state-space framework, and the damage state is estimated using a Bayesian filter. Specifically, the state-space framework has two components: (a) a physically based progressive damage model that provides a stochastic description of the damage state evolution, accounting for the structure geometry, loading, and environmental conditions, and (b) a measurement relationship based on the adaptive DP clustering which connects changes in TF statistics of measured sensor signals to changing damage and/or operating conditions. The Bayesian filter integrates predictions from the progressive damage model with data from adaptively learned TF statistics of measured sensor signals and the measurement relation to reliably estimate structural damage states online under variable external conditions.

Let  $x_k$  be the damage state (fatigue crack length) at epoch (fatigue cycle)  $k$ , and  $\phi_k$  and  $M_k$  the environmental or operating condition and the number of adaptively identified DP clusters, respectively. The damage state evolution model is Markov and specifies the conditional pdf  $P(x_k|x_{k-1}, \phi_k)$  of the damage state  $x_k$  given the damage state  $x_{k-1}$  at the previous epoch and the current environmental condition  $\phi_k$ . For crack damage in the aluminum CT specimen under cyclic fatigue loading, the damage evolution model can be obtained from a progressive crack growth model based on fracture mechanics (Yang and Manning, 1990, 1996; Ray and Patankar, 2001), which is of the form:

$$x_k = x_{k-1} + \chi_k \mathcal{H}(x_{k-1}, \phi_k) \quad (34)$$

Here  $\mathcal{H}$  is a non-linear function that depends on material properties, geometry, load, and  $\chi_k$  is a log-normal random variable. In particular, we use the crack growth model derived in Ray and Patankar (2001) and impose a discretized log-normal distribution (see Appendix A) on  $\chi_k$ . The change in damage  $\Delta x_k = x_k - x_{k-1} = \chi_k \mathcal{H}(x_{k-1}, \phi_k)$  then has conditional pdf

$$P(\Delta x_k|x_{k-1}, \phi_k) = \overline{\text{Log-N}}(\mu_{1,k}, \sigma_{1,k}, \Delta_\chi) \quad (35)$$

with  $\Delta_\chi$  a suitable discretization level, and forms the state variable of the Bayesian filter. The distribution parameters  $\mu_{1,k}$  and  $\sigma_{1,k}$  are obtained using equation (42), such that

$$E[\Delta x_k|x_{k-1}, \phi_k] = E[\chi_k] \mathcal{H}(x_{k-1}, \phi_k) \quad (36a)$$

$$\text{Var}[\Delta x_k|x_{k-1}, \phi_k] = \text{Var}[\chi_k] \mathcal{H}(x_{k-1}, \phi_k)^2 \quad (36b)$$

where  $E[\cdot]$  denotes the expected value and  $\text{Var}[\cdot]$  the variance. The variable amplitude loading plays the role of the changing external condition  $\phi_k$ .

The measurement equation of the Bayesian filter is formulated using the relationship between the change in the identified DP clustering  $\Delta M_k = M_k - M_{k-1}$ , the change in damage  $\Delta x_k$ , and environmental condition  $\phi_k$ , quantified with a negative binomial likelihood function:

$$\begin{aligned} \mathcal{L}(\Delta x_k|\Delta M_k, \phi_k) &= P(\Delta M_k|\Delta x_k, \phi_k) \\ &\triangleq \text{Neg-Bin}(\mu_{2,k}, \sigma_{2,k}), \end{aligned} \quad (37)$$

which is discrete, bi-parametric, and can be approximated to a finite support. Note that the negative binomial distribution has been reparameterized here in terms of its mean and standard deviation (see Appendix B). The parameters  $\mu_{2,k}$  and  $\sigma_{2,k}$  are estimated at each epoch based on the change in DP clustering as

$$\mu_{2,k} = E[\hat{\mathbf{p}}_k] - E[\hat{\mathbf{p}}_{k-1}] \quad (38a)$$

$$\sigma_{2,k}^2 = \text{Var}[\hat{\mathbf{p}}_k] + \text{Var}[\hat{\mathbf{p}}_{k-1}] \quad (38b)$$

where  $\hat{\mathbf{p}}_k = \frac{1}{L} \sum_{l=1}^L \mathbf{p}^{(l)}$  are the estimated DP-GMM mixture weights at epoch  $k$ , and  $E[\hat{\mathbf{p}}_k]$  and  $\text{Var}[\hat{\mathbf{p}}_k]$  are the mean and variance of this pmf, used to monitor the change in clustering. Stirling's approximation (Abramowitz and Stegun, 1965) is utilized for stably carrying out computations involving large arguments for the negative binomial distribution.

The damage estimation problem is then to determine the unknown damage state  $x_k$  at epoch  $k$ , given the adaptively learned  $\mathbf{M}_k = \{\Delta M_1, \dots, \Delta M_k\}$  and the environmental conditions  $\Phi_k = \{\phi_1, \dots, \phi_k\}$  up to epoch  $k$ , the state and measurement equations in (35) and (37), and an initial damage state distribution  $P(x_0)$ . Below is a modified Bayesian filter that computes the damage state estimate  $\hat{x}_k$  as the mean of the posterior pdf  $P(x_k|\mathbf{M}_k, \Phi_k)$ , which is updated iteratively for  $k = 1, 2, \dots$ , as:

$$P(\Delta x_k|\mathbf{M}_{k-1}, \Phi_k) = \int P(\Delta x_k|x_{k-1}, \phi_k) P(x_{k-1}|\mathbf{M}_{k-1}, \Phi_{k-1}) dx_{k-1} \quad (39a)$$

$$P(\Delta x_k|\mathbf{M}_k, \Phi_k) \propto P(\Delta M_k|\Delta x_k, \phi_k) \cdot P(\Delta x_k|\mathbf{M}_{k-1}, \Phi_k) \quad (39b)$$

$$P(x_k|\mathbf{M}_k, \Phi_k) \approx P(\Delta x_k + \hat{x}_{k-1}|\mathbf{M}_k, \Phi_k) \quad (39c)$$

Here, the posterior pdf over the damage state  $x_k$  is computed approximately using the distribution over the change in damage  $\Delta x_k$  and the previous damage state estimate  $\hat{x}_{k-1}$ . Algorithm 3 summarizes the steps of the modified Bayesian filter for progressive damage estimation. The damage state variable  $x_k$  has been discretized to a finite alphabet, so that the integration at

**Algorithm 3** Modified Bayesian filter for progressive damage estimation.

*Input:* DP clustering results  $\{\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2, \dots\}$ , damage model parameters:  $E[\chi_k]$ ,  $\text{Var}[\chi_k]$ ,  $\Delta\chi$ ,  $\mathcal{H}(\cdot)$ , environmental conditions  $\{\phi_1, \phi_2, \dots\}$ .

*Output:* Damage state estimates  $\hat{x}_1, \hat{x}_2, \dots$

Initialize damage state distribution  $P(x_0)$ .

For epochs  $k = 1, 2, \dots$ , perform the following steps:

- (1) Predict the distribution over the change in damage  $\Delta x_k$  at epoch  $k$  using the damage evolution model.  
For  $j = 1, 2, \dots$ , compute:

$$\mu_{1,k}^j = \log(E[\chi_k] \mathcal{H}(v_{k-1}^j, \phi_k)) - \frac{1}{2} \log\left(1 + \frac{\text{Var}[\chi_k]}{E[\chi_k]^2}\right), \quad \sigma_{1,k}^j = \sqrt{\log\left(1 + \frac{\text{Var}[\chi_k]}{E[\chi_k]^2}\right)}$$

Then compute, for  $j' = 1, 2, \dots$ ,

$$P(\Delta x_k = w_k^{j'} | \mathbf{M}_{k-1}, \Phi_k) = \sum_j \frac{1}{2} \left( \text{erf}\left[\frac{\mu_{1,k}^j - \log(j'-1)\Delta\chi}{\sigma_{1,k}^j \sqrt{2}}\right] - \text{erf}\left[\frac{\mu_{1,k}^j - \log j' \Delta\chi}{\sigma_{1,k}^j \sqrt{2}}\right] \right) \cdot P(x_{k-1} = v_{k-1}^j | \mathbf{M}_{k-1}, \Phi_{k-1})$$

- (2) Update the distribution over  $\Delta x_k$  based on the likelihood of the observed change in DP clustering  $\Delta M_k$ .  
From the DP-GMM clustering results, calculate:

$$\mu_{\hat{\mathbf{p}}_k} = \sum_{m=1}^K m \hat{p}_{m,k}, \quad \sigma_{\hat{\mathbf{p}}_k}^2 = \sum_{m=1}^K (m - \mu_{\hat{\mathbf{p}}_k})^2 \hat{p}_{m,k}, \quad \mu_{2,k} = \mu_{\hat{\mathbf{p}}_k} - \mu_{\hat{\mathbf{p}}_{k-1}}, \quad \sigma_{2,k}^2 = \sigma_{\hat{\mathbf{p}}_k}^2 + \sigma_{\hat{\mathbf{p}}_{k-1}}^2$$

$$r_k = \frac{\mu_{2,k}^2}{\sigma_{2,k}^2 - \mu_{2,k}}, \quad q_k = 1 - \frac{\mu_{2,k}}{\sigma_{2,k}^2}$$

Then, for  $j' = 1, 2, \dots$ , compute:

$$P(\Delta x_k = w_k^{j'} | \mathbf{M}_k, \Phi_k) = \frac{\frac{\Gamma(j' + r_k)}{\Gamma(j' + 1)} (q_k)^{j'} \cdot P(\Delta x_k = w_k^{j'} | \mathbf{M}_{k-1}, \Phi_k)}{\sum_{j'} \frac{\Gamma(j' + r_k)}{\Gamma(j' + 1)} (q_k)^{j'} \cdot P(\Delta x_k = w_k^{j'} | \mathbf{M}_{k-1}, \Phi_k)}$$

- (3) Compute the posterior pdf over the damage state  $x_k$  at epoch  $k$  using the previous damage state estimate  $\hat{x}_{k-1}$ .

For  $j = 1, 2, \dots$ , assign:  $P(x_k = v_k^j | \mathbf{M}_k, \Phi_k) = P(\Delta x_k = w_k^j | \mathbf{M}_k, \Phi_k)$ , where  $w_k^j = v_k^j - \hat{x}_{k-1}$ .

- (4) Compute damage state estimate  $\hat{x}_k$  at epoch  $k$  as the mean of the posterior pdf.

$$\text{Calculate: } \hat{x}_k \triangleq E[x_k | \mathbf{M}_k, \Phi_k] = \sum_j v_k^j P(x_k = v_k^j | \mathbf{M}_k, \Phi_k).$$

each epoch in (39) reduces to a finite sum and the filter is efficiently implemented.

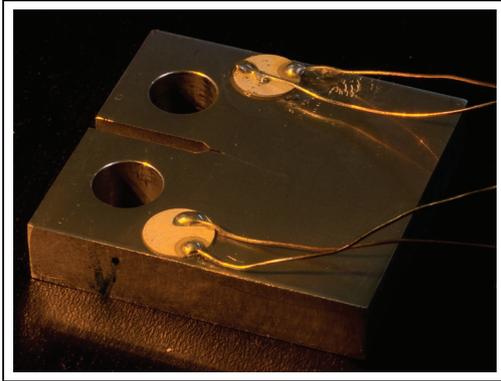
### 3 Estimation of fatigue crack damage in an aluminum CT specimen under variable-amplitude loading

We now discuss an application of the proposed adaptive learning method for fatigue crack damage estimation in a metallic sample subjected to variable loading conditions.

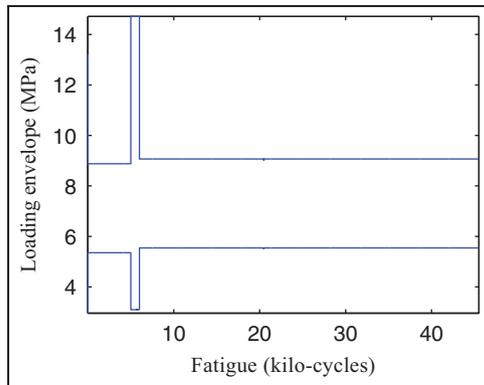
#### 3.1 Experimental setup and data collection

The test sample considered is an aluminum 2024 T3 6.31 mm thick CT specimen shown in Figure 4. The CT specimen was fabricated according to ASTM E647-93, with a width of 25.4 mm from the center of the pin hole to the edge of the specimen. The location of crack initiation was predetermined by making an initial notch of length 5 mm. The fatigue experiments were performed on a Instron 1331 servohydraulic load frame operating at 20 Hz. Real flight conditions were simulated by programming a typical center-wing load

spectrum into the load frame digital controller, with an envelope shown in Figure 5. About 45 kilocycles of the variable-amplitude cyclic loading were applied to induce fatigue crack damage in the CT sample. Data was collected using two surface mounted piezoelectric (PZT) sensors placed symmetrically on the sample as



**Figure 4.** Aluminum 2024 CT specimen used in the fatigue testing experiments.



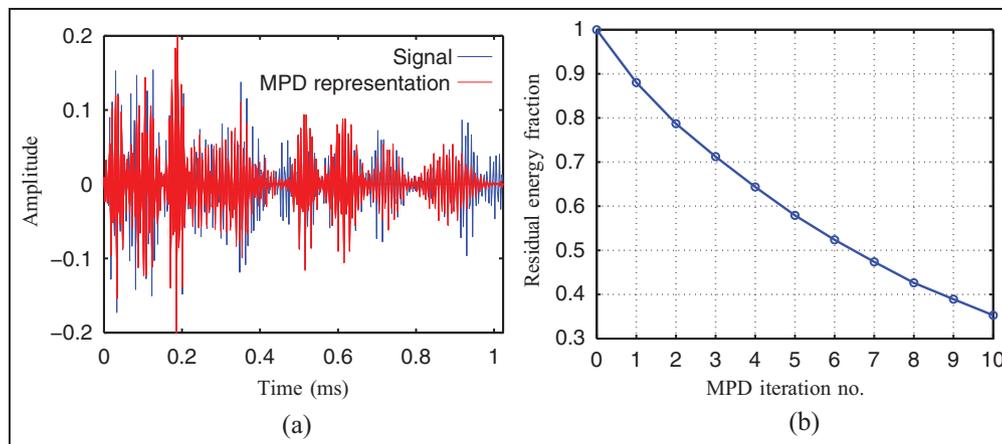
**Figure 5.** Envelope of the variable-amplitude cyclic fatigue loading cycles.

shown in Figure 4, at several stages of the fatigue loading cycles corresponding to various crack lengths. One of the PZT sensors was used as the actuator, and the other used as a receiver to measure the response signals (see Figure 6 for an example). A Gaussian windowed tone burst signal of center frequency 130 kHz was used for excitation. Signals were collected in sets of 210 time-domain waveforms at 4452, 5123, 5720, 20480, 34451, and 45507 cycles of fatigue loading. Crack lengths were measured using scanning electron microscopy (SEM) in an FEI XL-30 operating at 15 kV (Mohanty et al., 2007a) and recorded. Further details of the experimental setup and data collection procedure can be found in Mohanty et al. (2007a,b).

### 3.2 TF statistical processing and DP mixture modeling

The signals were first preprocessed to remove any DC offset, low-pass filtered to eliminate high frequency noise, downsampled to 333 kHz, and normalized to unit energy.  $N = 10$  iterations of MPD were performed on each signal, with a dictionary comprised of 8 million Gaussian TF shifted and scaled atoms, and the corresponding MPD-TFRs were computed (Section 2.1). Figure 6 shows an example response signal for the 6.17 mm crack length (4452 fatigue cycles), and its MPD representation and residual error. Figure 7 shows the MPD-TFRs of two signals from different damage states, measured at 4452 and 5123 fatigue cycles and with measured crack lengths of 6.17 mm and 6.28 mm, respectively. From the plots, we see marked difference in the TF structure of the signals, which is attributed to the change in mode content due to damage growth.

Next, TF statistical similarity features (Section 2.2) are calculated between the current measured signal pdfs and those from a reference set, that is optimized to retain representative measurement sets for the damage states using the uniform feature selection procedure as



**Figure 6.** Example response signal for crack length 6.17 mm and its MPD representation.

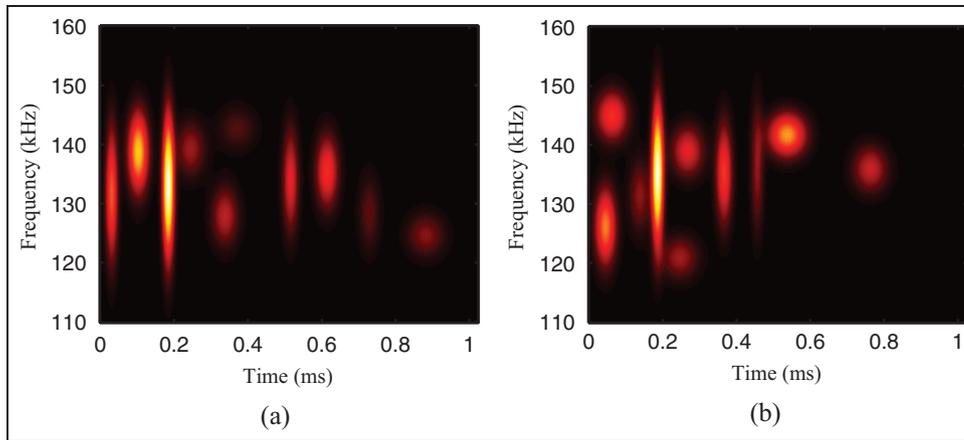


Figure 7. Example MPD-TFRs of two signals for different damage states.

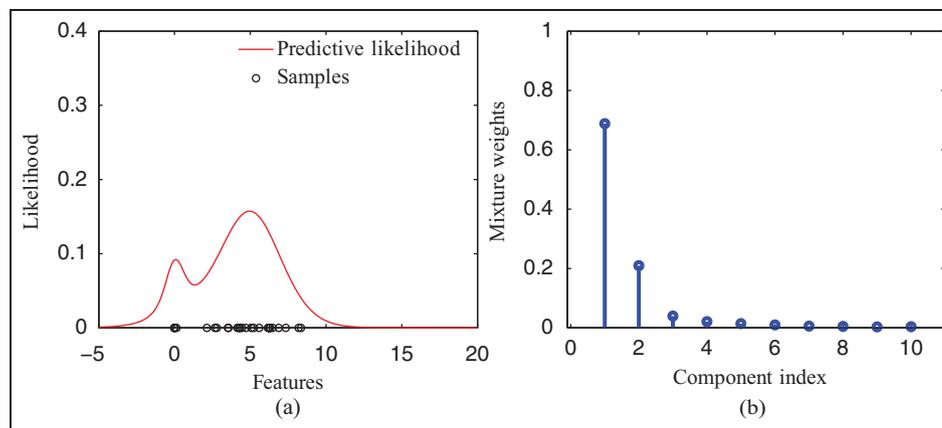


Figure 8. DP-GMM of features from two clusters.

described in Section 2.3. DP Gaussian mixture model based adaptive clustering was then effected on these features as discussed in Section 2.4, and the DP clustering parameters used subsequently for damage estimation. The clustering realized by the DP Gaussian mixture model is shown for a typical scenario in Figure 8. Here, the 1-D features are known to be from two clusters. The base distribution  $G_0$  is Normal-Gamma with hyperparameters  $u_\mu = 0$ ,  $u_\tau = 1$ ,  $u_a = 1$ , and  $u_b = 1$ . The DP innovation parameter was set to  $\iota = 1.5$  and a truncation limit of  $K = 10$  was used. From (25), this choice corresponds to a 1-norm marginal density error of about 0.099. Figure 9 shows a plot of the error as a function of  $K$  and  $\iota$ . A larger value of  $K$  would reduce the error, but at the cost of increased computational complexity. A lower  $\iota$  would similarly yield smaller error, but would also decrease the odds of generating new groups within the given data. For determining the DP model parameters, 100 blocked Gibbs sampling iterations were used for burn-in and sample generation, respectively. From the plots in Figure 8, we see that the GMM learned from the

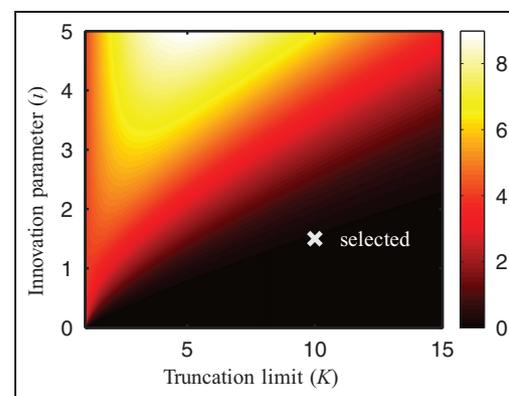
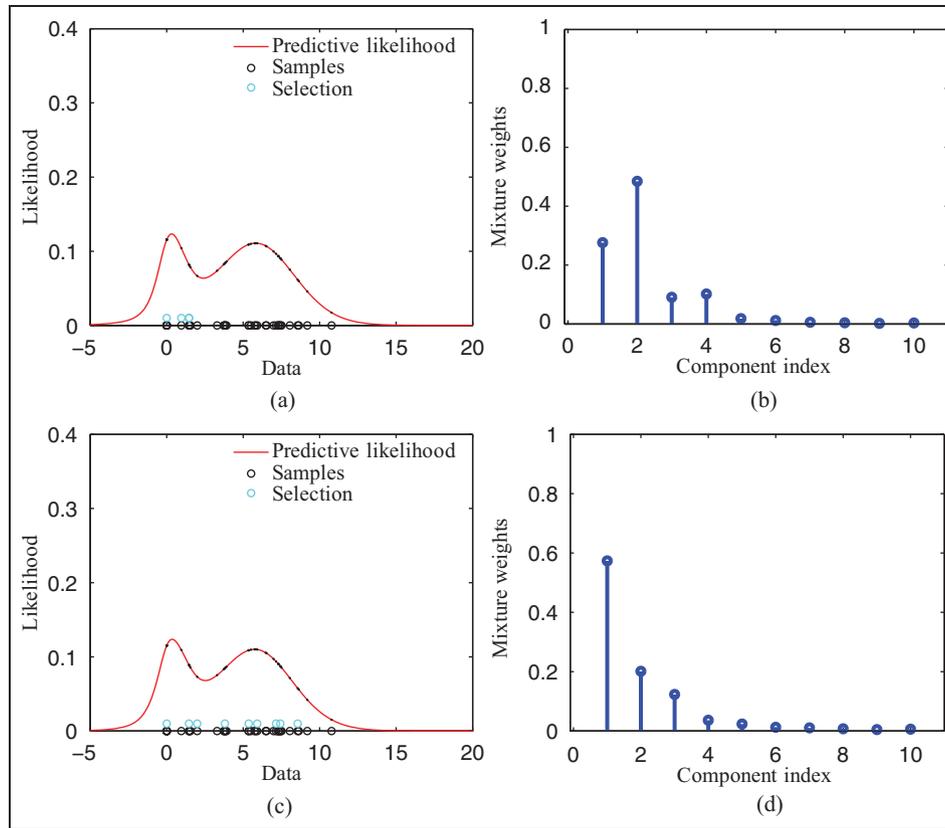


Figure 9. 1-norm marginal density error in the DP mixture model, shown as a function of truncation limit  $K$  and innovation parameter  $\iota$ .

data by the DP mixture modeling is fairly accurate, with the two dominant components identified correctly. The DP mixture learning phase at each epoch requires only about 0.3 s on a 2.8 GHz processor.



**Figure 10.** Data selection for improved adaptive learning.

In the construction of the reference MPD-pdf  $P_{S^{\text{ref}}}(t, f)$ ,  $\hat{T} = 10$  out of  $T = 30$  features were selected. Figure 10 compares the performance of the adaptive clustering applied to features from a single damage state at a typical epoch, with and without the use of feature selection. Figure 10(a) shows the 1-D features, the adaptively learned DP mixture likelihood, and a random selection of features for the reference set  $S^{\text{ref}}$  (Section 2.3). The resulting mixture weights at the following epoch are shown in Figure 10(b). Figure 10(c) shows the features, the adaptively learned DP mixture likelihood, and a uniform selection of features for the reference set. The resulting mixture weights at the following epoch are shown in Figure 10(d). Since all the features were from the same damage state, it is desirable that a single cluster be identified; this occurs with the use of feature selection. Similar results were observed at other time epochs as well.

### 3.3 Damage estimation results

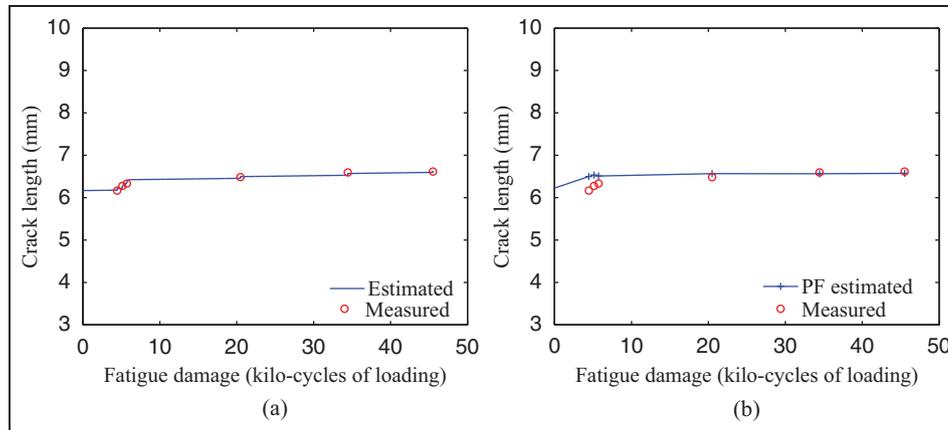
The Bayesian filter was used to optimally combine information from the DP based adaptive feature clustering with the physically driven damage growth model for estimating the probability distribution of the crack length as described in Section 2.5. At the loading cycles where no sensor data was measured, the damage was

predicted solely using the physics based model (recall that the sensor measurements were collected only at 4452, 5123, 5720, 20480, 34451, and 45507 cycles of fatigue loading). The value of the applied variable load was known for every cycle, and was used in the physics based state equation.

Figure 11(a) shows the performance of the proposed adaptive learning structural damage estimation method in estimating the fatigue crack damage in the CT specimen subjected to variable-amplitude loading. It can be seen that the crack length estimates are accurate, with the average error on the order of 0.1 mm. Figure 11(b) is provided to compare with the performance of the particle filtering (PF) algorithm of Zhou et al. (2009a). Observe that the adaptive learning approach performs better in the vicinity of the cycles when the load amplitude changes. This is expected because, unlike the PF method that is based on a fixed model, the adaptive learning algorithm utilizes a flexible model that can change its complexity as needed.

## 4 Conclusion

This paper develops a novel adaptive learning progressive damage estimation method for structural health monitoring under variable conditions. The algorithm achieves adaptation by making use of the flexibility



**Figure 11.** Crack length estimation in the CT specimen as a function of fatigue loading cycles.

offered by Bayesian nonparametric statistics, specifically, DP mixture modeling. The DP is utilized to adaptively learn a potentially unlimited number of mixture components (latent clusters) within joint TF features extracted from periodically measured sensor data, with the DP mixture model parameters determined efficiently using the MCMC technique of Gibbs sampling. The clusters identified in the statistically changing dispersive wave phenomena are associated with structural damage evolving under varying conditions. This adaptive information is then combined with a physically based damage growth model in a state-space formulation and the damage state estimates are computed using a Bayesian filter. A data selection procedure is also incorporated to maximize performance by actively selecting the reference measurements used for monitoring changes. When applied for the estimation of fatigue crack damage in an aluminum CT specimen under variable loading conditions, the proposed method showed promising results with the crack length estimates obtained accurate to within a few millimeters.

In this work, the DP mixture modeling employed Gaussian likelihoods for analytical and computational simplicity. However, other likelihood forms may be more suitable depending on the characteristics of the data, and can be utilized in the DP mixture modeling so long as a conjugate prior is available that is easy to sample from (Escobar and West, 1995; Ishwaran and James, 2001). Regarding sampling, it should be noted that determining convergence of the MCMC iterations used in the learning of the DP mixture model parameters involves separate analysis and calculations (Cowles and Carlin, 1996). Also, the innovation parameter of the DP, which was fixed here for simplicity, can be learned from the data (Escobar and West, 1995). Finally, the DP based adaptive clustering can be used to monitor statistical changes in data, but the method cannot by itself provide information about the physical meaning of the clusters. For that, prior knowledge must be given in the form of models relating the

clustering in feature space to structural damage and external conditions. For example, the state space framework considered here relies on the availability of an accurate stochastic physically based damage evolution model that incorporates the effects of variable environmental and operating conditions. Such models can be non-trivial to obtain for complex structures and certain types of variabilities. These issues are currently under investigation.

### Funding

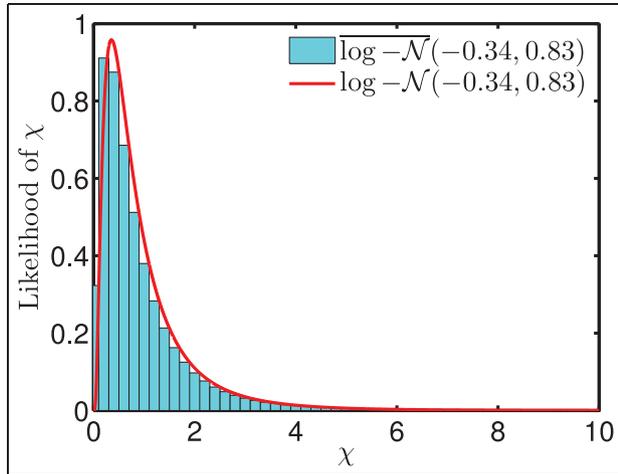
This work was supported by the MURI program in the Air Force Office of Scientific Research, (grant number FA95550-06-1-0309; technical monitor: Dr David Stargel).

### References

- Abramowitz M and Stegun IA (eds) (1965) *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables*. New York: Dover.
- Antoniak CE (1974) Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics* 2: 1152–1174.
- Bhattacharyya A (1943) On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society* 35: 99–109. MR0010358.
- Blackwell D and MacQueen JB (1973) Ferguson distributions via Polyá urn schemes. *The Annals of Statistics* 1: 353–355.
- Boyd S and Vandenberghe L (2004) *Convex Optimization*. New York: Cambridge University Press.
- Chakraborty D, Kovvali N, Papandreou-Suppappola A, et al. (2010) Active learning data selection for adaptive online structural damage estimation. In: *Proceedings of SPIE, smart structures and materials and non-destructive evaluation and health monitoring*. San Diego, California, March 2010, vol. 7649, p. 764915.
- Chakraborty D, Kovvali N, Wei J, et al. (2009a) Damage classification structural health monitoring in bolted structures using time–frequency techniques. *Journal of Intelligent Material Systems and Structures* 20: 1289–1305.

- Chakraborty D, Kovvali N, Zhang J, et al. (2009b) Adaptive learning for damage classification in structural health monitoring. In: *43rd Asilomar conference on signals, systems and computers*, Pacific Grove, California, November 2009, pp. 1678–1682.
- Chakraborty D, Soni S, Wei J, et al. (2008a) Physics based modeling for time–frequency damage classification. In: *Proceedings of SPIE, Smart structures and materials and non-destructive evaluation and health monitoring*, San Diego, California, March 2008, Vol. 6926, pp. 1–12.
- Chakraborty D, Zhou W, Simon D, et al. (2008b) Time–frequency methods for structural health monitoring. In: *Proceedings of Sensor, signal and information processing (SenSIP) workshop, Workshop*, Sedona, Arizona, May 2008.
- Channels L, Chakraborty D, Butrym B, et al. (2009) A comparative study of fatigue damage sensing in aluminum alloys using electrical impedance and laser ultrasonic methods. In: *Proceedings of SPIE, smart structures and materials and non-destructive evaluation and health monitoring*, San Diego, California, March 2009, Vol. 7295, pp. 1–10.
- Channels L, Chakraborty D, Simon D, et al. (2008) Ultrasonic sensing and time–frequency analysis for detecting plastic deformation in an aluminum plate. In: *Proceedings of SPIE, smart structures and materials and non-destructive evaluation and health monitoring*, Vol. 6926, pp. 1–10.
- Chazelle B (2002) *The Discrepancy Method*. Cambridge; New York: Cambridge University Press.
- Cohen L (1989) Time-frequency distribution—a review. *IEEE Proceedings* 77(7): 941–981.
- Cohen L (1994) *Time Frequency Analysis: Theory and Applications*. Englewood Cliffs, NJ: Prentice-Hall.
- Cowles MK and Carlin BP (1996) Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association* 91(434): 883–904.
- Das S, Srivastava AN and Chattopadhyay A (2007) Classification of damage signatures in composite plates using one-class SVMs. In: *Proceedings of 2007 IEEE aerospace conference*, Big Sky, MT, 3–10 March 2007, pp. 1–19.
- Dempster A, Laird N and Rubin D (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* 39: 1–38.
- Dobkin DP and Eppstein D (1993) Computing the discrepancy. In: *Proceedings of the 9th annual symposium on computational geometry*, San Diego, California, 18–21 May, 1993, pp. 47–52.
- Dobkin DP, Eppstein D and Mitchell DP (1996) Computing the discrepancy with applications to supersampling patterns. *ACM Transactions on Graphics (TOG)* 15: 354–376.
- Doebling S and Farrar CR (1998) Statistical damage identification techniques applied to the I-40 bridge over the Rio Grande. In: *Proceedings of international modal analysis conference*, Santa Barbara, CA 1998, pp. 1717–1724.
- Escobar MD and West M (1995) Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* 90: 577–588.
- Farrar C, Nix D, Duffey T, et al. (1999) Damage identification with linear discriminant operators. In: *International Modal Analysis Conference*, Kissimmee, FL 1999, pp. 599–607.
- Farrar CR and Lieven NAJ (2007) Damage prognosis: The future of structural health monitoring. *Royal Society of London Transactions Series A* 365: 623–632.
- Farrar CR and Worden K (2007) An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society, Series A* 365: 303–315.
- Feng Z and Chu F (2007) Nonstationary vibration signal analysis of a hydroturbine based on adaptive chirplet decomposition. *Structural Health Monitoring* 6(4): 265–279.
- Ferguson TS (1973) A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* 1: 209–230.
- Fink D (1995) A compendium of conjugate priors. Technical report, Montana State University.
- Fox EB, Sudderth EB, Jordan MI, et al. (2011) A sticky HDP-HMM with application to speaker diarization. *Annals of Applied Statistics* 5: 1020–1056.
- Gelman A, Carlin JB, Stern HS, et al. (2004a) *Bayesian Data Analysis*. 2nd edition. Boca Raton, FL: CRC Press.
- Gelman L, Giurgiutiu V and Petrunin I (2004b) Advantage of using the Fourier components pair instead of power spectral density for fatigue crack diagnostics. *International Journal of Condition Monitoring and Diagnostic Engineering Management* 7: 18–22.
- Gilks W, Richardson S and Spiegelhalter D (1996) *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall/CRC.
- Ishwaran H and James LF (2001) Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* 96: 161–173.
- Jeong H and Jang Y (2000) Fracture source location in thin plates using the wavelet transform of dispersive waves. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 47: 612–619.
- Jordan MI (2010a) Bayesian nonparametric learning: Expressive priors for intelligent systems. In: R Dechter, H Geffner and J Halpern (eds) *Heuristics, Probability and Causality: A Tribute to Judea Pearl*. London: College Publications.
- Jordan MI (2010b) Hierarchical models, nested models and completely random measures. In: MH Chen, D Dey, P Mueller, et al. (eds) *Frontiers of Statistical Decision Making and Bayesian Analysis: In Honor of James O. Berger*. New York: Springer.
- Kailath T (1967) The divergence and Bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Technology* 15(1): 52–60. DOI: 10.1109/TCOM.1967.1089532.
- Karasaridis A, Maalej M, Pantazopoulou S, et al. (1997) Time–frequency analysis of sensor data for detection of structural damage in instrumented structures. In: *Proceedings of international conference on digital signal processing*, Santorini, 2–4 July 1997, Vol. 2, pp. 817–820.
- Kuipers L and Niederreiter H (1974) *Uniform Distribution of Sequences*. New York: John Wiley and Sons.
- Le Cam L and Yang GL (2000) *Asymptotics in Statistics: Some Basic Concepts*. Berlin: Springer.
- MacKay DJ (1992) Information-based objective functions for active data selection. *Neural Computation* 4(4): 590–604.
- MacKay DJC (2003) *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press.
- Mallat S (1998) *A Wavelet Tour of Signal Processing* (2nd edn). New York: Academic Press.
- Mallat SG and Zhang Z (1993) Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing* 41: 3397–3415.

- Michaels J and Michaels T (2005) Detection of structural damage from the local temporal coherence of diffuse ultrasonic signals. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control* 52(10): 1769–1782.
- Mohanty S, Teale R, Chattopadhyay A, et al. (2007a) Mixed Gaussian process and state-space approach for fatigue crack growth prediction. In: FK Chang (ed) *Structural Health Monitoring*. Lancaster, PA: DEStech Publications.
- Mohanty S, Teale R, Chattopadhyay A, et al. (2007b) Mixed Gaussian process and state-space approach for fatigue crack growth prediction. In: *International workshop on structural health monitoring*, Stanford, California, September 2007, Vol. 2, pp. 1108–1115.
- Nguyen M, Wang X, Su Z, et al. (2004) Damage identification for composite structures with a Bayesian network. In: *Proceedings of intelligent sensors, sensor networks and information processing conference*, 14–17 December 2004, pp. 307–311.
- Ni K, Qi Y and Carin L (2007a) Multi-aspect target classification and detection via the infinite hidden Markov model. In: *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 2, pp. 433–436.
- Ni K, Qi Y and Carin L (2007b) Multiaspect target detection via the infinite hidden Markov model. *Journal of the Acoustical Society of America* 121: 2731–2742.
- Niederreiter H (1972) Discrepancy and convex programming. *Annali di Matematica Pura ed Applicata* 93: 89–97.
- Papandreou-Suppappola A (ed.) (2002) *Applications in Time-Frequency Signal Processing*. Boca Raton, FL: CRC Press.
- Pakrashia V, Basu B and Connora AO (2007) Structural damage detection and calibration using a wavelet-kurtosis technique. *Engineering Structures* 29(9): 2097–2108.
- Park G, Sohn H, Farrar CR and Inman DJ (2003) Overview of piezoelectric impedance-based health monitoring and the path forward. *Shock and Vibration Digest* 35(6): 451–463.
- Qi Y, Paisley JW and Carin L (2007a) Dirichlet process HMM mixture models with application to music analysis. In: *IEEE international conference on acoustics, speech and signal processing*, Honolulu, HI, 15–20 April 2007, Vol. 2, pp. 465–468.
- Qi Y, Paisley JW and Carin L (2007b) Music analysis using hidden Markov mixture models. *IEEE Transactions on Signal Processing* 55: 5209–5224.
- Ray A and Patankar R (2001) Fatigue crack growth under variable-amplitude loading: Part I—Model formulation in state-space setting. *Applied Mathematical Modelling* 25: 979–994.
- Rose JL (1999) *Ultrasonic Waves in Solid Media*. Cambridge: New York: Cambridge University Press.
- Sethuraman J (1994) A constructive definition of Dirichlet priors. *Statistica Sinica* 4: 639–650.
- Sohn H (2007) Effects of environmental and operational variability on structural health monitoring. *Philosophical Transactions of The Royal Society A* 365: 539–560.
- Sohn H, Allen DW, Worden K, et al. (2005) Structural damage classification using extreme value statistics. *Journal of Dynamic Systems, Measurement, and Control* 127: 125–132.
- Sohn H and Farrar CR (2001) Damage diagnosis using time series analysis of vibration signals. *Smart Material Structures* 10: 446–451.
- Sohn H, Farrar CR, Hunter NF, et al. (2001) Structural health monitoring using statistical pattern recognition techniques. *Transactions of the ASME* 123: 706–711.
- Sohn H and Law KH (2000) Bayesian probabilistic damage detection of a reinforced-concrete bridge column. *Earthquake Engineering Structural Dynamics* 29: 1139–1152.
- Soni S, Kim SB and Chattopadhyay A (2010) Reference-free fatigue crack detection, localization and quantification in lug joints. In: *Proceedings of 51st AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics, and materials conference (SDM)*, Orlando, FL, 12–15 April 2010.
- Staszewski WJ (1998) Structural and mechanical damage detection using wavelets. *The Shock and Vibration Digest* 30: 457–472.
- Staszewski WJ, Boller C and Tomlinson G (2004) *Structural Health Monitoring of Aerospace Structures, Smart Sensor Technologies and Signal Processing*. New York: John Wiley and Sons.
- Sun A and Chang CC (2004) Statistical wavelet-based method for structural health monitoring. *Journal of Structural Engineering* 130(7): 1055–1062.
- Taha MMR, Noureldin A, Lucero JL, et al. (2006) Wavelet transform for structural health monitoring: A compendium of uses and features. *Structural Health Monitoring* 5(3): 267–295.
- Ting D, Wang G, Shapovalov M, et al. (2010) Neighbor-dependent Ramachandran probability distributions of amino acids developed from a hierarchical Dirichlet process model. *PLoS Computational Biology* 6: E1000763.
- West M, Muller P and Escobar MD (1994) Hierarchical priors and mixture models, with applications in regression and density estimation. In: PR Freeman and AF Smith (eds) *Aspects of Uncertainty*. New York: John Wiley, pp. 363–386.
- Xu B, Giurgiutiu V and Yu L (2009) Lamb waves decomposition and mode identification using matching pursuit method. In: *Proceedings of SPIE, Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, 30 March 2009, Vol. 7292.
- Yang JN and Manning SD (1990) Stochastic crack growth analysis methodologies for metallic structures. *Engineering Fracture Mechanics* 37: 1105–1124.
- Yang JN and Manning SD (1996) A simple second order approximation for stochastic crack growth analysis. *Engineering Fracture Mechanics* 53(5): 677–686.
- Yu L and Giurgiutiu V (2005) Advanced signal processing for enhanced damage detection with piezoelectric wafer active sensors. *Smart Structures and Systems* 1(2): 185–215.
- Zhou W, Chakraborty D, Kowali N, et al. (2007) Damage classification for structural health monitoring using time-frequency feature extraction and continuous hidden Markov models. In: *Proceedings of conference record of the 41st asilomar conference on signals, systems and computers (ACSSC 2007)*, Pacific Grove, California, November 2007, pp. 848–852.
- Zhou W, Kovvali N, Papandreou-Suppappola A, et al. (2009a) Progressive damage estimation using sequential Monte Carlo techniques. In: *Proceedings of the 7th international workshop on structural health monitoring*, Stanford, California, 2009.
- Zhou W, Kovvali N, Reynolds W, et al. (2009b) On the use of hidden Markov modeling and time-frequency features for damage classification in composite structures. *Journal of Intelligent Material Systems and Structures* 20: 1271–1288.



**Figure 12.** Comparison of log-normal and discretized log-normal distributions, for  $E[\chi] = 1$  and  $\text{Var}[\chi] = 1$ .

**Appendix A: Discretized log-normal distribution**

A log-normal random variable  $\chi \sim \text{Log-N}(\mu_\chi, \sigma_\chi)$  has pdf

$$P(\chi|\mu_\chi, \sigma_\chi) = \frac{1}{\chi\sigma_\chi\sqrt{2\pi}} e^{-\frac{(\log\chi - \mu_\chi)^2}{2\sigma_\chi^2}}, \quad \chi > 0 \quad (40)$$

with mean and variance given by

$$E[\chi] = e^{\mu_\chi + \frac{1}{2}\sigma_\chi^2} \quad (41a)$$

$$\text{Var}[\chi] = (e^{\sigma_\chi^2} - 1)e^{2\mu_\chi + \sigma_\chi^2} \quad (41b)$$

Note that the parameters  $\mu_\chi$  and  $\sigma_\chi$  can be expressed in terms of a desired mean  $E[\chi]$  and variance  $\text{Var}[\chi]$  as

$$\mu_\chi = \log(E[\chi]) - \frac{1}{2}\log\left(1 + \frac{\text{Var}[\chi]}{E[\chi]^2}\right) \quad (42a)$$

$$\sigma_\chi = \sqrt{\log\left(1 + \frac{\text{Var}[\chi]}{E[\chi]^2}\right)} \quad (42b)$$

We define a *discretized* log-normal random variable  $\chi \sim \text{Log-N}(\mu_\chi, \sigma_\chi, \Delta_\chi)$ , with pmf

$$\begin{aligned} \Pr(\chi = j|\mu_\chi, \sigma_\chi, \Delta_\chi) &\triangleq \int_{(j-1)\Delta_\chi}^{j\Delta_\chi} \frac{1}{\chi\sigma_\chi\sqrt{2\pi}} e^{-\frac{(\log\chi - \mu_\chi)^2}{2\sigma_\chi^2}} d\chi \\ &= \frac{1}{2} \left( \text{erf}\left[\frac{\mu_\chi - \log(j-1)\Delta_\chi}{\sigma_\chi\sqrt{2}}\right] - \text{erf}\left[\frac{\mu_\chi - \log j\Delta_\chi}{\sigma_\chi\sqrt{2}}\right] \right) \end{aligned} \quad (43)$$

for  $j = 1, 2, \dots$ , where  $\text{erf}[\cdot]$  is the error function

$$\text{erf}[x] = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (44)$$

A comparison of the log-normal and discretized log-normal distributions is shown in Figure 12.

**Appendix B: Reparameterized negative binomial distribution**

The negative binomial distribution is a discrete probability distribution defined over non-negative integers, with pmf of  $\varpi \sim \text{Neg-Bin}(r, q)$  given by

$$\Pr(\varpi = j|r, q) = \frac{\Gamma(j+r)}{\Gamma(j+1)\Gamma(r)} (1-q)^r q^j \quad (45)$$

for  $j = 0, 1, \dots$ , with parameters  $r > 0$  and  $0 < q < 1$ . The mean and standard deviation of  $\varpi$  are

$$\mu_\varpi = \frac{rq}{1-q} \quad (46a)$$

$$\sigma_\varpi = \sqrt{\frac{rq}{(1-q)^2}} \quad (46b)$$

with  $0 < \mu_\varpi < \sigma_\varpi^2$ . It is convenient to reparameterize the negative binomial distribution in terms of its mean and standard deviation as  $\text{Neg-Bin}(\mu_\varpi, \sigma_\varpi)$ . For a desired mean  $\mu_\varpi$  and standard deviation  $\sigma_\varpi$ , the parameters  $r$  and  $q$  in (45) can be computed by inverting (46):

$$r = \frac{\mu_\varpi^2}{\sigma_\varpi^2 - \mu_\varpi} \quad (47a)$$

$$q = 1 - \frac{\mu_\varpi}{\sigma_\varpi^2} \quad (47b)$$