Adaptive Learning for Damage Classification in Structural Health Monitoring

D. Chakraborty†, N. Kovvali†, J. J. Zhang†, A. Papandreou-Suppappola†, and A. Chattopadhyay‡

†School of Electrical, Computer and Energy Engineering, Arizona State University
‡School of Mechanical, Aerospace, Chemical and Materials Engineering, Arizona State University

***Abstract*—A key challenge in real-world structural health monitoring (SHM) is diversity of damage phenomena and variability in environmental and operational conditions. Conventional learning techniques, while adequate for moderately complex inference tasks, can be limiting in highly complex and rapidly changing environments, especially when insufficient data is available. We present an adaptive learning methodology where stochastic models continuously evolve with the time-varying environment and Dirichlet process mixture models are utilized to self-adapt to structure within the data. Coupled with appropriate physics-based phenomenology, the approach provides an adaptive and effective framework for online SHM. The proposed technique is demonstrated for the detection of progressive fatigue damage in a metallic structure under variable-amplitude loading.**

I. Introduction

Structural health monitoring (SHM) is an important problem encountered in many civil, mechanical, and aerospace applications. In the last few decades, several statistical learning techniques [1] were used to address uncertainty in the damage process, with the ultimate goal of identifying damage in structures of interest. One of the key challenges in the development of real-world damage detection and classification systems is diversity of damage phenomena and variability in environmental and operational conditions. Structural damage appears in a multitude of forms, and information collected via measured sensor data is often strongly influenced, for example, by changes in temperature, geometry or configuration, sensor characteristics, and material variability [2]. Conventional learning methods, while adequate for characterizing the underlying mechanism of damage nucleation and evolution, are of limited use in a highly complex and rapidly changing environment, especially when sufficient amount of data is not available. The main problem is that, even though data may be collected continuously, the stochastic modeling framework remains static. For example, the crack monitoring module of an aircraft wing would need to be adjusted depending on whether the aircraft is maneuvering or is moving into a region of drastically different weather, such as turbulence.

In this paper, we propose an adaptive learning based damage classification methodology where the stochastic models are allowed to continuously evolve from experience with the time-varying environment. The adaptive learning framework is based on the use of Dirichlet process (DP) mixture models [3] to provide the modeling with the machinery needed to self-adapt to structure within the data. Essentially, the DP mixture model provides for a growing, possibly infinite number of mixture components, a finite number of which manifest themselves within the given data. The appropriate number of mixture components and mixture proportions, and the mixture distribution parameters, are learned adaptively from the data. The DP mixture model learning is performed using Markov Chain Monte Carlo (MCMC) techniques [4]. The adaptively identified components or classes can then be traced to different types of damage within a structure or different possible variations in the material or the environment for the same type of damage. The damage state inference is performed using a Bayesian filter that combines the adaptive data model with a physics based progressive damage model [5]. The main advantage of this approach is that no baseline training data is required and signals can be classified on the fly to new (previously unseen) damage classes, yielding an adaptive and effective approach for online SHM.

This work was supported by the AFOSR Grant FA95550-06-1-0309.

The remainder of the paper is organized as follows. In Section II, we discuss the underlying theoretical framework of DP mixture modeling. In Section III, we describe in detail the adaptive learning based damage classification algorithm. We establish the relevance of adaptive learning for the online SHM problem and examine the role of the physics-based damage evolution model. In Section IV, we describe the real experimental data used to demonstrate the utility of the proposed approach, and present results demonstrating the performance of the proposed approach for the classification of progressive fatigue crack damage in a aluminum compact-tension (CT) sample subjected to variable-amplitude loading.

II. Dirichlet Process Mixture Models

In this section, we briefly describe the analytical framework of Dirichlet process mixture models used in the proposed adaptive SHM method. For more details on these topics the reader is referred to the literature [3], [4], [6]–[8].

A. Dirichlet Process Mixture Models

Consider a mixture model of the form

$$p(\mathbf{y}|\mathbf{p}, \mathbf{\Theta}^*, M) = \sum_{m=1}^{M} p_m \, f(\mathbf{y}|\boldsymbol{\theta}_m^*), \tag{1}$$

where $\mathbf{y}$ denotes data, M is the number of mixture components, $\mathbf{p} = \{p_1, \ldots, p_M\}$ is a set of mixing proportions or weights (which must be positive and sum to one), $\mathbf{\Theta}^* =$

$\{\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_M^*\}$ is a set of parameters, and $f(\mathbf{y}|\boldsymbol{\theta}_m^*)$ represents a suitable probability distribution (with parameter $\boldsymbol{\theta}_m^*$). Given a data set $\mathbf{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$ of size N, the modeling task comprises learning the mixture size M, proportions $\mathbf{p}$, and the distribution parameters $\boldsymbol{\Theta}^*$.

If the optimum mixture size M is known, then classical maximum-likelihood (ML) learning techniques can be applied to estimate the parameters of interest [6], [7]. In particular, the iterative expectation-maximization (EM) algorithm can be used to locally maximize the data likelihood [6], [9], [10]. The EM algorithm introduces auxiliary or hidden variables and iterates between inferring the posterior distribution over the hidden variables given a current parameter setting and computing a new parameter estimate by maximizing the likelihood using the learned statistics.

An alternative and much more flexible approach to mixture modeling is provided by the Dirichlet process (DP) [3], [8], [11]–[15]. The DP forms a nonparametric prior distribution for mixture models with an *unbounded* number of components. A DP, denoted by $DP(\alpha, G_0)$, is parametrized by a positive scalar innovation parameter α and a base distribution G_0. Let G be a distribution drawn from a DP, then

$$G \sim DP(\alpha, G_0), \tag{2}$$

where G_0 is the expected value of G and α determines the concentration of the prior for G about G_0.

A key feature of the DP is that, with probability one, distributions drawn from a DP are discrete. This property can be used to automatically determine the number of components in a mixture model. Consider N random variables $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N\}$ distributed according to G:

$$\begin{aligned} G &\sim DP(\alpha, G_0), && \text{(3a)} \\ \boldsymbol{\theta}_n|G &\sim G, \;\; n = 1, \ldots, N. && \text{(3b)} \end{aligned}$$

Since G is discrete, the variables $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N\}$ take on coincident values with positive probability. In particular, by integrating out the distribution G, the following Pólya urn property can be shown for the conditional density function of variable $\boldsymbol{\theta}_n$ given all other variables (denoted $\boldsymbol{\Theta}^{-n}$) [3], [11]–[14]:

$$\begin{aligned} p(\boldsymbol{\theta}_n|\boldsymbol{\Theta}^{-n}, \alpha, G_0) &= \frac{1}{\alpha + N - 1} \sum_{m=1}^{M} n_m^{-n}\, \delta(\boldsymbol{\theta}_m^*, \boldsymbol{\theta}_n) \\ &\quad + \frac{\alpha}{\alpha + N - 1} G_0(\boldsymbol{\theta}_n), \end{aligned} \tag{4}$$

where δ is the Kronecker-delta function, $\{\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_M^*\}$ are the distinct values taken by $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N\}$, and n_m^{-n} is the number of variables in $\boldsymbol{\Theta}^{-n}$ equal to $\boldsymbol{\theta}_m^*$. As a result, the DP prior entails that each variable $\boldsymbol{\theta}_n$ either assumes an existing value $\boldsymbol{\theta}_m^*$ with probability $n_m^{-n}/(\alpha + N - 1)$ or is drawn fresh from G_0 with probability $\alpha/(\alpha + N - 1)$. The innovation parameter α controls the extent to which sharing is encouraged. It can be shown that the joint distribution of $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N\}$ is invariant to permutation—this important property is known as exchangeability. Further, the DP prior is nonparametric and the number of distinct values M roughly grows as $O(\log N)$.

An explicit characterization of G is given by the *stick-breaking construction* [15]

$$\begin{aligned} \boldsymbol{\theta}_m^* &\sim G_0, \;\; m = 1, \ldots, \infty, && \text{(5a)} \\ v_i &\sim \text{Beta}(1, \alpha), \;\; i = 1, \ldots, \infty, && \text{(5b)} \\ p_m &= v_m \prod_{i=1}^{m-1} (1 - v_i), \;\; m = 1, \ldots, \infty, && \text{(5c)} \\ G(\boldsymbol{\theta}) &= \sum_{m=1}^{\infty} p_m\, \delta(\boldsymbol{\theta}, \boldsymbol{\theta}_m^*), && \text{(5d)} \end{aligned}$$

which shows that G is discrete and places its probability mass on a countably infinite subset of the sample space.

The DP can be used as a nonparametric prior in a hierarchical Bayesian model

$$\begin{aligned} G &\sim DP(\alpha, G_0), && \text{(6a)} \\ \boldsymbol{\theta}_n|G &\sim G, \;\; n = 1, \ldots, N, && \text{(6b)} \\ \mathbf{y}_n|\boldsymbol{\theta}_n &\sim f(\mathbf{y}_n|\boldsymbol{\theta}_n), \;\; n = 1, \ldots, N. && \text{(6c)} \end{aligned}$$

In this specification, each variable $\boldsymbol{\theta}_n$ forms the (unobserved) parameter of a probability distribution $f(\mathbf{y}_n|\boldsymbol{\theta}_n)$ from which the data point $\mathbf{y}_n$ is generated. Since the parameters are drawn from G and take on coincident values, the data clusters according to those values. In view of the stick-breaking construction (5), (6) can be rewritten as

$$\begin{aligned} \boldsymbol{\theta}_m^* &\sim G_0, \;\; m = 1, \ldots, \infty, && \text{(7a)} \\ v_i &\sim \text{Beta}(1, \alpha), \;\; i = 1, \ldots, \infty, && \text{(7b)} \\ p_m &= v_m \prod_{i=1}^{m-1} (1 - v_i), \;\; m = 1, \ldots, \infty, && \text{(7c)} \\ z_n|\mathbf{p} &\sim \text{Mult}(\mathbf{p}), \;\; n = 1, \ldots, N, && \text{(7d)} \\ \mathbf{y}_n|z_n &\sim f(\mathbf{y}_n|\boldsymbol{\theta}_{z_n}^*), \;\; n = 1, \ldots, N, && \text{(7e)} \end{aligned}$$

which is equivalent to an infinite mixture model termed the *Dirichlet process mixture model*:

$$p(\mathbf{y}|\mathbf{p}, \boldsymbol{\Theta}^*) = \sum_{m=1}^{\infty} p_m\, f(\mathbf{y}|\boldsymbol{\theta}_m^*). \tag{8}$$

The DP construction above can be truncated to K terms by setting $v_K = 1$. This truncated Dirichlet process (TDP) $DP_K(\alpha, G_0)$ closely approximates the original Dirichlet process $DP(\alpha, G_0)$ when K is large enough relative to N. In particular, it can be shown than the marginal density $q(\mathbf{Y})$ obeys

$$\|q_K - q_\infty\|_1 \approx 4N \exp(-(K-1)/\alpha), \tag{9}$$

which provides a criterion for selecting the appropriate truncation limit K.

B. Learning and Inference via Markov Chain Monte Carlo

The goal of Bayesian inference is to find the posterior distribution over the parameters of interest given the observed data. Markov Chain Monte Carlo (MCMC) methods [4] construct a Markov chain on the parameters, for which the target

(stationary) distribution is the posterior conditioned on the data. In the Gibbs sampler [4], the Markov chain is obtained by iteratively sampling each random variable conditioned on the data and the previously sampled values of the other variables. Samples are then collected from the converged Markov chain and used to construct an empirical estimate of the posterior distribution over the parameters. This estimate can be used to approximate various posterior expectations of interest.

A powerful MCMC approach for learning DP mixture models is provided by the blocked Gibbs sampling algorithm [8]. In this method, blocks of parameters of a TDP are updated to summarize the posterior distribution $p(\mathbf{p}, \mathbf{\Theta}^*, \mathbf{z}|\mathbf{Y})$ directly. Specifically, samples are iteratively drawn from the following conditional distributions [8]:

$$\boldsymbol{\theta}^*_m \sim p(\boldsymbol{\theta}^*_m|\mathbf{z}, \mathbf{Y}), \quad m = 1, \ldots, K, \tag{10a}$$

$$z_n \sim p(z_n|\mathbf{\Theta}^*, \mathbf{p}, \mathbf{Y}), \quad n = 1, \ldots, N, \tag{10b}$$

$$p_m \sim p(p_m|\mathbf{z}), \quad m = 1, \ldots, K. \tag{10c}$$

The predictive distribution is then approximated as

$$p(\mathbf{y}_{N+1}|\mathbf{Y}, \alpha, G_0) = \frac{1}{R}\sum_{r=1}^{R}\left[\sum_{m=1}^{K} p_m^{(r)}\, f(\mathbf{y}_{N+1}|\boldsymbol{\theta}_m^{*(r)})\right],$$

where R is the number of collected samples.

III. Adaptive Learning based Online SHM

We now discuss the adaptive learning based method for online SHM.

A. Time-Frequency Feature Extraction

The objective of feature extraction is to condense, with minimum loss, the information contained in measured data into a form suitable for further analysis and processing. In the present work, features are extracted from measured piezoelectric transducer (PZT) sensor signals based on joint time-frequency (TF) analysis [16], which is known to be well-suited for capturing the dispersive nature of damage wave physics [17]. Furthermore, since we are concerned with statistical changes in the measured signals, the features are defined so as to isolate transient effects.

Feature extraction is performed in a two-step process. In the first step, the matching pursuit decomposition (MPD) algorithm is used to construct cross-term free TF representations (MPD-TFRs) for the signals [18]. For a signal $s(t) \in \mathbf{L}^2(\mathbb{R})$, the L-term MPD representation $s_L(t)$, computed iteratively, is of the form

$$s(t) \approx s_L(t) = \sum_{l=0}^{L-1} \alpha_l\, g_{\gamma_l}(t), \tag{11}$$

where α_l are the expansion coefficients and $g_{\gamma_l}(t)$ are basis functions selected from a dictionary $\mathcal{D}$. In this work, the MPD uses a dictionary of highly localized Gaussian functions that are time-frequency shifted and scaled versions of a basic Gaussian atom $g(t) = Ce^{-t^2/2}$. The Gaussian functions have optimal TF resolution properties and therefore afford representations which are compact yet accurate. The cross-term free MPD-TFR of $s(t)$ is given by [18]

$$E_s(t,f) = \sum_{l=0}^{L-1} |\alpha_l|^2\, \mathrm{WD}_{g_{\gamma_l}}(t,f)\,, \tag{12}$$

where $\mathrm{WD}_{g_{\gamma_l}}(t,f)$ is the Wigner distribution TFR of the Gaussian function $g_{\gamma_l}(t)$ [16].

In the second step of the feature extraction process, correlation-based distance features are computed between the test signals' MPD-TFRs and reference MPD-TFRs (defined using the DP mixture and physics-based information from the previous time period) as

$$\mathbf{y}_s = \iint E_s(t,f) \log \frac{E_s(t,f)}{E_{\mathrm{ref}}(t,f)}\, dt\, df. \tag{13}$$

This correlation-based distance can be interpreted as the Kullback-Leibler (KL) divergence measure [6] between TF probability distributions of $s(t)$ and the reference signal, with the MPD-TFRs acting as probability distributions. The computation of the integral in (13) is carried out using Monte-Carlo integration.

B. Data Model based on Dirichlet Process Gaussian Mixtures

In this work, DP mixture models are used to provide a framework for modeling structural data of varying statistical structure. Specifically, damage-related features extracted from periodically buffered structural data are modeled using a DP Gaussian mixture model of the form

$$p(\mathbf{y}) = \sum_{m=1}^{M} p_m\, f(\mathbf{y}|\boldsymbol{\theta}^*_m), \tag{14}$$

where $\mathbf{y}$ denotes the extracted feature vector, M is the number of mixture components, p_m are mixture proportions, $f(\cdot)$ is a Gaussian probability distribution that matches the properties of all possible damage classes in the structure being tested, and $\boldsymbol{\theta}^*_m = (\mu^*_m, \sigma^{2*}_m)$ is the parameter vector of mean μ^*_m and variance σ^{2*}_m of this distribution.

The DP model parameters, i.e., the mixture proportions and means and variances of the Gaussian density functions, are estimated using the blocked Gibbs sampling algorithm [8] in (10). In particular, a Normal-Gamma prior is used for the base distribution G_0, which is conjugate to the Gaussian likelihood $f(\cdot)$ with unknown mean and variance [4]. This property enables convenient iterative sampling from the posterior $p(\boldsymbol{\theta}^*_m|\mathbf{z}, \mathbf{Y})$ in (10).

The relationship between the change in unknown damage $\Delta x_k = x_k - x_{k-1}$, at time step k, and the change in the number of identified components $\Delta M_k = M_k - M_{k-1}$ and environmental conditions ϕ_k is quantified by defining a probabilistic model $p(\Delta M_k|\Delta x_k, \phi_k)$. Based on empirical experience, a negative binomial form is imposed on this distribution. Computations involving large arguments for the negative binomial distribution are carried out stably by making use of Stirling's approximation [19].

C. Physics-based Damage Evolution Model

The evolution of the unknown damage (crack length x_k at fatigue cycle k) in the CT sample is described by making use of a progressive fatigue crack growth model [5], [20] based on fracture mechanics. In its modified form, the model for fatigue crack growth in a Aluminum alloy 2024-T3 CT sample under variable-amplitude loading is given by

$$x_k = x_{k-1} + \chi_k \, C(\Delta K_k^{\text{eff}})^m, \tag{15}$$

where C and m are material-dependent constants, ΔK^{eff} is the effective stress intensity range which depends on the sample geometry and load, input maximum and minimum stress S_{max} and S_{min}, and crack opening stress S_{o} [5], and χ_k is a log-normal random process that is introduced to compensate for the state modeling errors [20]. The mean and variance of the χ_k are chosen for agreement between the model prediction and experimentally observed crack length values.

Equation (15) can be rewritten as

$$x_k = x_{k-1} + \chi_k \, \mathcal{G}(x_{k-1}, \phi_k), \tag{16}$$

where $\mathcal{G}$ is a nonlinear function and the variable amplitude loading plays the role of environmental conditions ϕ_k. The damage evolution model above results in the conditional distribution $p(\Delta x_k | x_{k-1}, \phi_k)$ assuming a log-normal form.

D. Bayesian Filtering Formulation

With the data and damage evolution models $p(\Delta M_k | \Delta x_k, \phi_k)$ and $p(\Delta x_k | x_{k-1}, \phi_k)$ defined, the damage state estimation problem can be stated formally as follows: Given the adaptively learned $\mathcal{M}_k = \{\Delta M_1, \ldots, \Delta M_k\}$ and the environmental conditions $\Phi_k = \{\phi_1, \ldots, \phi_k\}$ up to time k, we wish to estimate the unknown damage state x_k at time k.

The Bayesian filtering approach iteratively computes the damage state estimate $\hat{x}_k$ as the mean of the posterior distribution $p(x_k | \mathcal{M}_k, \Phi_k)$ for $k = 1, 2, \ldots$ as

$$p(\Delta x_k | \mathcal{M}_{k-1}, \Phi_k) = \int p(\Delta x_k | x_{k-1}, \phi_k) \cdot p(x_{k-1} | \mathcal{M}_{k-1}, \Phi_{k-1}) \, dx_{k-1}, \tag{17a}$$

$$p(\Delta x_k | \mathcal{M}_k, \Phi_k) \propto p(\Delta M_k | \Delta x_k, \phi_k) \, p(\Delta x_k | \mathcal{M}_{k-1}, \Phi_k),$$

$$p(x_k | \mathcal{M}_k, \Phi_k) \approx p(\Delta x_k + \hat{x}_{k-1} | \mathcal{M}_k, \Phi_k), \tag{17b}$$

$$\hat{x}_k = E[x_k | \mathcal{M}_k, \Phi_k]. \tag{17c}$$

Note the approximation used in (17b) for computing the posterior distribution over the damage state x_k from the distribution over the change in damage Δx_k. Also, the initial state distribution $p(x_0 | \mathcal{M}_0, \Phi_0) \equiv p(x_0)$ is assumed known.

For simplicity, in this work we discretize the damage state variable x_k to a finite alphabet. The integral at each time-step in (17a) then reduces to a finite sum, and the implementation of the filter becomes straightforward.

Fig. 1 shows the block diagram of the adaptive learning based damage classification algorithm.

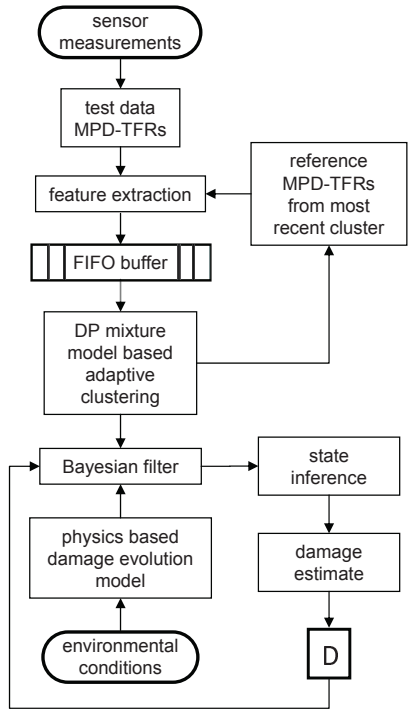


Fig. 1. Block diagram of the adaptive learning based damage classification algorithm.

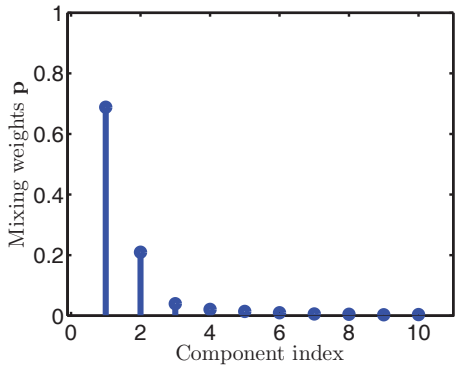
IV. Results

We now demonstrate the adaptive learning technique for the classification of progressive fatigue crack damage in a Aluminum 2024 CT sample of width 25.4 mm subjected to variable-amplitude loading. A surface mounted PZT actuator and sensor were used to collect signals in response to a 70 kHz burst input applied at various fatigue cycles up to 50 kilo-cycles. More details of the experiment and data collection can be found in [21].

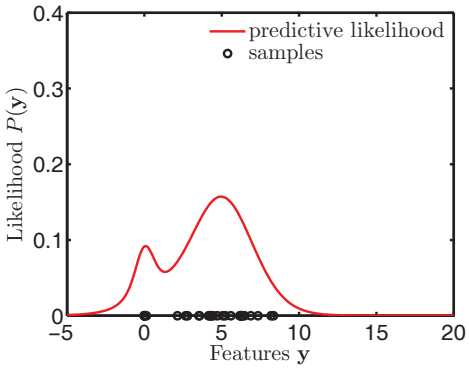
The signals were first preprocessed using low-pass filtering, downsampling, normalization, and mean-removal. MPD was then carried out to $L = 10$ terms with a TF dictionary consisting of about 8 million Gaussian atoms. In the DP Gaussian mixture modeling, the innovation parameter was set to $\alpha = 1.5$ and the DP truncation limit was chosen as $K = 10$. The number of blocked Gibbs sampling iterations performed were $R = 100$, both for the burn-in and burn-out phases. We utilized the convergence diagnostics described in [22] to assess the convergence of the Gibbs sampling algorithm.

Fig. 2(a) shows an example of the number of components identified in the data using the DP mixture modeling approach. Fig. 2(b) shows the corresponding learned DP Gaussian mixture model likelihood function. It can be seen from the plots that two components are found to be predominant in the data at this stage. Fig. 2(c) demonstrates the overall progress of the adaptive learning based damage classification method. We see that the estimated crack length values are very close to the actual crack in the sample measured using scanning electron

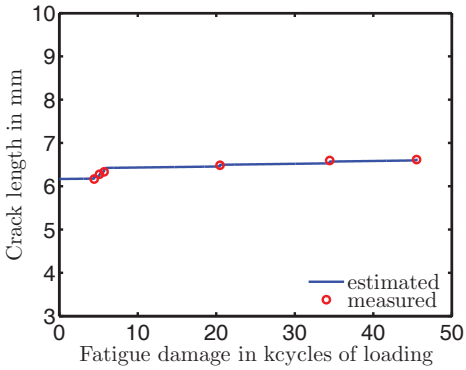
microscopy. The adaptive learning method accurately identifies the crack length in the CT sample under variable loading conditions.



(a) Components identified in data via DP mixture modeling.



(b) Learned DP Gaussian mixture model likelihood function.



(c) Progress of adaptive damage classification method.

Fig. 2. Adaptive learning based damage classification results.

V. Conclusion

In this paper, we have presented an adaptive learning based damage classification methodology in which Dirichlet process mixture models are used to self-adapt to structure within the data. The method combines the adaptive measurement model with available physics-based damage evolution models using a Bayesian filter to estimate damage accurately. Results from applying the proposed method to the detection of progressive fatigue damage in a CT sample under variable-amplitude loading demonstrate good damage classification performance. More tests are underway to ascertain the performance of the algorithm under other types of variability. Also under investigation is the use of active data selection techniques to further improve the damage classification performance.

References

[1] H. Sohn, C. R. Farrar, N. F. Hunter, and K. Worden, "Structural health monitoring using statistical pattern recognition techniques," *Transactions of the ASME*, vol. 123, pp. 706–711, 2001.

[2] C. R. Farrar and K. Worden, "An introduction to structural health monitoring," *Phil. Trans. R. Soc. A*, vol. 365, pp. 303–315, 2007.

[3] M. D. Escobar and M. West, "Bayesian density estimation and inference using mixtures," *Journal of the American Statistical Association*, vol. 90, pp. 577–588, 1995.

[4] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC, 1996.

[5] A. Ray and R. Patankar, "Fatigue crack growth under variable-amplitude loading: Part I - Model formulation in state-space setting," *Applied Mathematical Modelling*, vol. 25, pp. 979–994, 2001.

[6] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[7] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, 2nd ed. CRC Press, 2004.

[8] H. Ishwaran and L. F. James, "Gibbs sampling methods for stick-breaking priors," *Journal of the American Statistical Association*, vol. 96, pp. 161–173, 2001.

[9] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, pp. 1–38, 1977.

[10] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Wiley Interscience, 2001.

[11] T. S. Ferguson, "A Bayesian analysis of some nonparametric problems," *The Annals of Statistics*, vol. 1, pp. 209–230, 1973.

[12] D. Blackwell and J. B. MacQueen, "Ferguson distributions via Polya urn schemes," *The Annals of Statistics*, vol. 1, pp. 353–355, 1973.

[13] C. E. Antoniak, "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems," *The Annals of Statistics*, vol. 2, pp. 1152–1174, 1974.

[14] M. West, P. Muller, and M. D. Escobar, "Hierarchical priors and mixture models, with applications in regression and density estimation," in *Aspects of Uncertainty*, P. R. Freeman and A. F. Smith, Eds. John Wiley, 1994, pp. 363–386.

[15] J. Sethuraman, "A constructive definition of Dirichlet priors," *Statistica Sinica*, vol. 4, pp. 639–650, 1994.

[16] A. Papandreou-Suppappola, Ed., *Applications in Time-Frequency Signal Processing*. Florida: CRC Press, 2002.

[17] D. Chakraborty, W. Zhou, D. Simon, N. Kovvali, A. Papandreou-Suppappola, D. Cochran, and A. Chattopadhyay, "Time-frequency methods for structural health monitoring," in *Sensor, Signal and Information Processing (SenSIP) Workshop*, May 2008.

[18] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3397–3415, 1993.

[19] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables*. New York: Dover, 1965.

[20] J. N. Yang and S. D. Manning, "Stochastic crack growth analysis methodologies for metallic structures," *Engineering Fracture Mechanics*, vol. 37, pp. 1105–1124, 1990.

[21] S. Mohanty, R. Teale, A. Chattopadhyay, P. Peralta, and C. Willhauck, "Mixed Gaussian process and state-space approach for fatigue crack growth prediction," in *International Workshop on Structural Heath Monitoring*, vol. 2, 2007, pp. 1108–1115.

[22] A. E. Raftery and S. M. Lewis, "How many iterations in the Gibbs sampler?" in *Bayesian Statistics 4*, J. B. et al., Ed. Oxford University Press, 1992, pp. 763–773.