6th European Conference on Computational Mechanics (ECCM 6)
7th European Conference on Computational Fluid Dynamics (ECFD 7)
11 – 15 June 2018, Glasgow, UK

PROBABILISTIC DATA MINING FOR AIRCRAFT STRUCTURAL HEALTH MONITORING

SHENFANG YUAN*, JIAN CHEN AND LEI QIU

Research Center of Structural Health Monitoring and Prognosis, the State Key Laboratory of Mechanics and Control of Mechanical Structures, Nanjing University of Aeronautics and Astronautics
29 Yudao Street, Nanjing, China
ysf@nuaa.edu.cn

Key words: Probabilistic Data Mining, Structural Health Monitoring, Probabilistic Damage Diagnosis, Guided Wave.

Abstract. Structural health monitoring for aircraft structures has gradually turned from fundamental research to practical implementations. However, numerous uncertainties arise from practical engineering such as time-varying loads and boundary conditions may have great effects on sensor signals, which make it difficult for reliable evaluation of structural damages. To deal with these uncertainties, probabilistic data mining methods are attracting more and more attention and gradually applied to aircraft structural health monitoring. Probabilistic data mining methods quantify effects of these uncertainties and the damage with probabilistic models, and perform reliable damage evaluation with diagnosis method. This paper aims at discussing probabilistic data mining methods in aircraft structural health monitoring, as well as their applications to aircraft structures in practical engineering taking advantages of guided wave based structural health monitoring.

1 INTRODUCTION

Structural Health Monitoring (SHM) [1] is a key technology used to address the safety and maintenance problems in aircraft engineering. Due to the worldwide attention and advances in maturing SHM technology for aerospace structures in the past two decades, this technology has gradually turned from fundamental research to practical implementations. However, aircraft structures in practical engineering suffer from time-varying conditions such as environment factors, loading, and boundary conditions. These conditions are usually random, which introduce numerous uncertainties to signals acquired from SHM sensors and make it difficult to implement reliable diagnosis of structural damages.

Aiming at reducing influences of these conditions, methods such as environmental parameter compensation and baseline signal dependency reduction [2] are developed. For example, the temperature compensation method is proposed to reduce effects of the temperature on SHM signals with the on-line monitored temperature. However, these methods have their own limitations, since real aircraft structures are complicated and effects of time-varying conditions on SHM signals are strong coupling. On the other hand, probabilistic data mining methods are being paid more and more attention to deal with these uncertainties. This kind of methods quantify uncertainties with probabilistic models, based on which reliable damage evaluation is performed. Different kinds of probabilistic data mining methods are developed for damage diagnosis under time-varying conditions, such as the

Gaussian mixture model (GMM) [3] and particle filtering (PF) [4].

GMM is a promising data-driven statistical model that can be used to describe complicated and unknown probability distributions, which is especially suitable to model uncertainties introduced by complicated time-varying conditions. In the GMM based damage diagnosis, the GMM is used to model the distribution of SHM signal features sequentially. Then the migration of the GMM relative to the baseline GMM is used for reliable damage evaluation. Besides, the PF is a Bayesian filtering method aiming at solving non-linear and non-Gaussian problems, which introduces the prior knowledge of a damage evolution model in damage diagnosis. The posterior estimation of the damage state is evaluated through Bayesian methods. These two probabilistic data mining methods have been studied preliminarily for damage diagnosis under time-varying conditions. Tschope et al. [5] reported the validation of using GMM to classify the damage degree in a plate-like structure. Banerjee et al. [6] used GMM as a classification modeling technique to estimate and quantify the progressive damage with the guided wave based SHM. Corbetta et al. [7] adopted the PF for diagnosis and prognosis of the fatigue crack damage under random loading. Rangaraj et al. [8] incorporated a PF algorithm for crack identification in beams from vibration measurements. However, deep research needs to be further addressed for more complex structures and real service conditions.

This paper discuss the two probabilistic data mining based methods, the GMM and PF, for damage diagnosis under time-varying conditions, taking advantages of the guided wave based SHM. Evaluation of these two methods is carried out on real complicated aircraft structures. The rest of the paper is organized as follows: Section 2 gives a brief discussion about the two probabilistic data mining methods. Section 3 introduces experimental validations of these methods. Section 4 gives conclusion and discussion.

2 PROBABILISTIC DATA MINING METHODS FOR SHM

Structural health monitoring (SHM) is an emerging technology which combines advanced sensors and algorithms to interrogate the state of the structure in real-time or whenever necessary. The basic ideal of SHM is to arrange sensors on the target structure, such as accelerometers, fibre optics, and piezoelectric transducers. With these sensors that are permanently attached to the structure, signals can be acquired at any time. By extracting signal features from SHM signals, the structural damage is identified and evaluated through advanced diagnosis technologies. Fig. 1 illustrates the typical SHM process.

It should be noted that not only the damage but also environment, loading, and boundary conditions may cause changes to the SHM signal. Therefore, probabilistic data mining methods are developed to deal with this problem. The following discuss the basics of two probabilistic data mining methods, the GMM and PF.

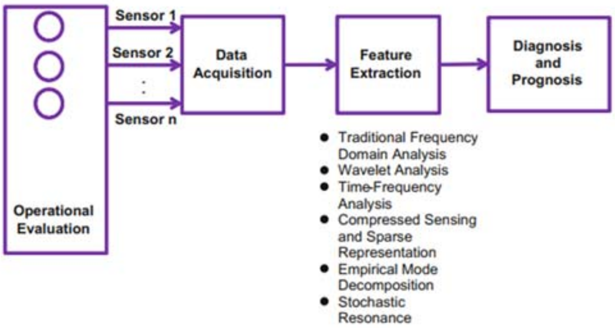


Figure 1: Typical process of SHM [9]

2.1 Gaussian mixture model

Denoting a signal acquired from SHM sensors as $\boldsymbol{s}$, signal features extracted from this signal are expressed as an vector $\boldsymbol{z}$=$[z_1, z_2, ..., z_n]$, where $z_1, z_2, ..., z_n$ are n different features. Initially, a number of baseline signals are collected when the structure is in the healthy state under service conditions. Features of these signals are extracted as $\{\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_K\}$, where K is the number signals. The uncertainty distribution of these signal features can be considered as a mixture effect of the time-varying conditions, which is assumed to be statistically distributed. The GMM is adopted to model this probabilistic distribution of these signal features.

GMM is a probability distribution that is constructed by a weighted sum of a finite number of Gaussian components, as expressed in Eq. (1).

$$\xi(\mathbf{Z}\,|\,\boldsymbol{\theta}) = \sum_{r=1}^{C} \pi_r \xi_r (\mathbf{Z}\,|\,\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r) \tag{1}$$

where $\mathbf{Z}$=$[\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_M]$ is the matrix representation of the signal features set $\{\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_M\}$, π_r is the mixture weight, C is the number of Gaussian components, $\xi_r(\mathbf{Z}\,|\,\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$ is the rth Gaussian component which is expressed as Eq. (2)

$$\xi_r(\mathbf{Z}\,|\,\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r) = \frac{1}{(2\pi)^{D/2}\sqrt{|\boldsymbol{\Sigma}_r|}} \mathrm{e}^{-\frac{1}{2}(z_k-\boldsymbol{\mu}_r)^{\mathrm{T}} \boldsymbol{\Sigma}_i^{-1} (z_k-\boldsymbol{\mu}_r)} \tag{2}$$

where $\boldsymbol{\mu}_r$ and $\boldsymbol{\Sigma}_r$ are the mean and the covariance matrix of the rth Gaussian components.

A GMM is parameterized by the probability distribution parameters $\boldsymbol{\theta}$=$(\pi_1, ..., \pi_C, \boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_C, \boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_C)$. In the GMM theory, unsupervised learning is adopted to obtain these parameters. Given the signal feature set $\mathbf{Z}$, the EM algorithm [3] is performed iteratively to maximum the likelihood function shown in Eq. (3), which gives the optimal parameter vector $\boldsymbol{\theta}_{\mathrm{opt}}$.

$$L(\mathbf{Z}|\xi) = \sum_{k=1}^{K} \log\left[\sum_{i=1}^{C} \pi_i \xi_i \left(\boldsymbol{x}_k \,|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right)\right] \tag{3}$$

At the initial time, signal features $\mathbf{Z}$=$[\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_K]$ extracted from baseline signals are used to train the GMM model, which is called as the baseline GMM. Then, SHM signals are acquired on-line successively. Once a new SHM signal is obtained, the set of signal features $\mathbf{Z}$ is updated as $\mathbf{Z}_{\text{update}}$. Based on the new dataset $\mathbf{Z}_{\text{update}}$, the GMM model is retrained, which is called the called as the monitoring GMM. Since the influence introduced by time-varying conditions is random while the influence of damage is inherent and progressive accompanying with damage propagation, the difference between the monitoring GMM and the baseline GMM represents the damage evaluation. The probability distribution difference measurement such as the Kullback–Leibler (KL) distance can be used to quantify the migration difference, which give rise to the damage evaluation of the structural damage.

2.2 Particle filter

Different from the GMM, the PF introduces the prior knowledge of a damage evolution model in damage diagnosis. A state space model is adopted to represent damage evolution and the SHM monitoring process explicitly, in which the uncertainties introduced by time-varying conditions are modelled as random variables. The state space model composes of a state equation as shown in Eq. (1) and a measurement equation as shown in Eq. (2).

$$\boldsymbol{x}_k = f_k(\boldsymbol{x}_{k-1}, \boldsymbol{\theta}_{k-1}, \boldsymbol{\omega}_{k-1}) \tag{4}$$

$$\boldsymbol{z}_k = g_k(\boldsymbol{x}_k, \boldsymbol{v}_k) \tag{5}$$

where k is the discrete time index, $\boldsymbol{x}_k$ is state vector at time k, $f_k(\cdot)$ is a nonlinear function representing damage evolution from time $k-1$ to time k, $\boldsymbol{\omega}_{k-1}$ is a random variable modelling uncertainties of damage evolution. $\boldsymbol{z}_k$ is the feature vector extracted from the SHM signal, $g_k(\cdot)$ is the mapping between $\boldsymbol{z}_k$ and the damage state $\boldsymbol{x}_k$, $\boldsymbol{v}_k$ is the measurement noise modelling uncertainties during the SHM process.

The basic idea of the PF is to evaluate the posterior probability density (pdf) of the damage state $p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k})$ with the Bayesian filtering framework.

$$p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k-1}) = \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) p(\boldsymbol{x}_{k-1}|\boldsymbol{z}_{1:k-1}) d\boldsymbol{x}_{k-1} \tag{6}$$

$$p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k}) = \frac{p(\boldsymbol{z}_k|\boldsymbol{x}_k) p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k-1})}{p(\boldsymbol{z}_k|\boldsymbol{z}_{1:k-1})} \tag{7}$$

where $p(\boldsymbol{x}_{k-1}|\boldsymbol{z}_{1:k-1})$ is the posterior pdf at time k-1. $p(\boldsymbol{z}_k|\boldsymbol{z}_{1:k-1})$ is a normalization constant, $p(\boldsymbol{z}_k|\boldsymbol{x}_k)$ is the likelihood function.

However, these two equation do not have analytic solutions in most cases. The PF approximates the posterior pdf by means of N_s samples called 'particles' using the importance sampling strategy

$$p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k}) \approx \sum_{i=1}^{N_s} \tilde{w}_k^{(i)} \delta(\boldsymbol{x}_k - \boldsymbol{x}_k^{(i)}) \tag{8}$$

where δ is the Dirac delta function. $\boldsymbol{x}_k^{(i)}$ represents the ith particle, N_s is the number of particles, δ is the Dirac delta function expressed as Equation (6). These particles are sampled from an importance density function $q(\boldsymbol{x}_k|\boldsymbol{z}_{1:k})$. A usually used importance density is the transition pdf $q(\boldsymbol{x}_k|\boldsymbol{z}_{1:k})=p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$, which gives the weight updating equation as Eq. (9).

$$w_k^{(i)} = w_{k-1}^{(i)} p(\boldsymbol{z}_k|\boldsymbol{x}_k^{(i)}) \tag{9}$$

The posterior estimation of the state is evaluated as Eq. (6).

$$\hat{\boldsymbol{x}}_k \approx \sum_{i=1}^{N_s} \tilde{w}_k^{(i)} \boldsymbol{x}_k^{(i)} \tag{10}$$

One of the main problems of this PF is the degeneracy phenomenon. After a few iterations, all but one particle will have negligible weight. To deal with this problem, a resampling procedure is performed, which eliminates particles with small weights, copying which have large weights, and setting all the weights to $1/N_s$.

2.3 Experimental Validation

Experiments using the active guided wave based SHM is are performed to validate these probabilistic data mining methods. Figure 2 illustrates the principle of the active guided wave based SHM. The guided wave is a kind of elastic waves that propagate in wave-guide structures, which can be excited and sensed with the piezoelectric transducers (PZTs). Changes of the guided wave may be caused by the structural damage, based on which damage can be evaluated with diagnosis algorithms.

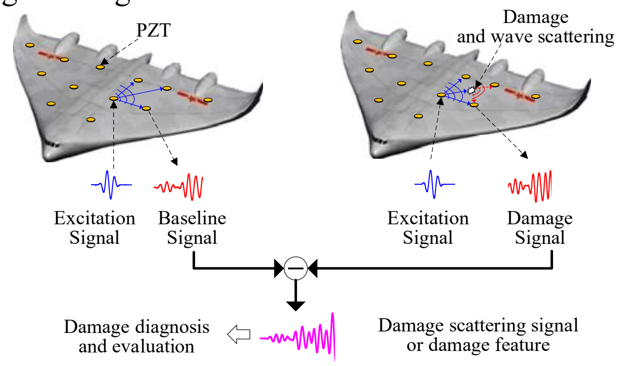


Figure 2: Illustration of the active guided wave based SHM

3.1 Evaluation of the GMM based damage diagnosis

Fig. 3 shows the full-scale aircraft fatigue test, which is performed on an aircraft that have accumulated thousands of flight hours. A fatigue crack is found at the right landing gear spar. Two PZTs are arranged on the surface of the structure to excite and acquire guided wave signals with the SHM system developed by the authors [11, 12]. During the fatigue test, scheduled inspection is carried out with an endoscope.

As illustrated in Fig. 4, the signal features that are also called as damage index (DI) are extracted from the monitored guided wave signal. Two damage indices are involved: the time-domain cross-correlation damage index (DI_1) and the Spectrum magnitude difference damage index (DI_2). It is obvious that the damage indices are strongly affected by the loading and environmental conditions. It is difficult for damage evaluation by directly using these damage indices.

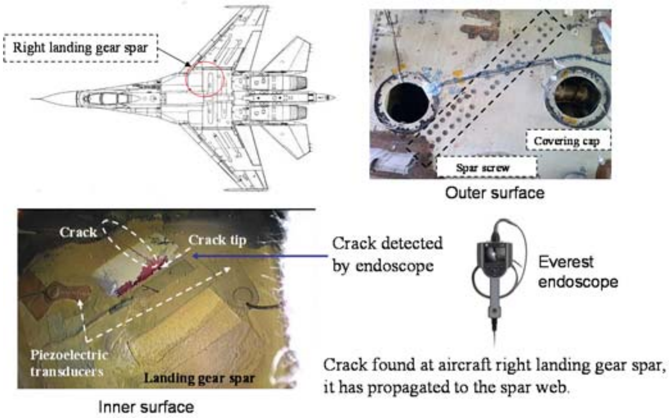


Figure 3: Experimental setup of the full-scale fatigue test

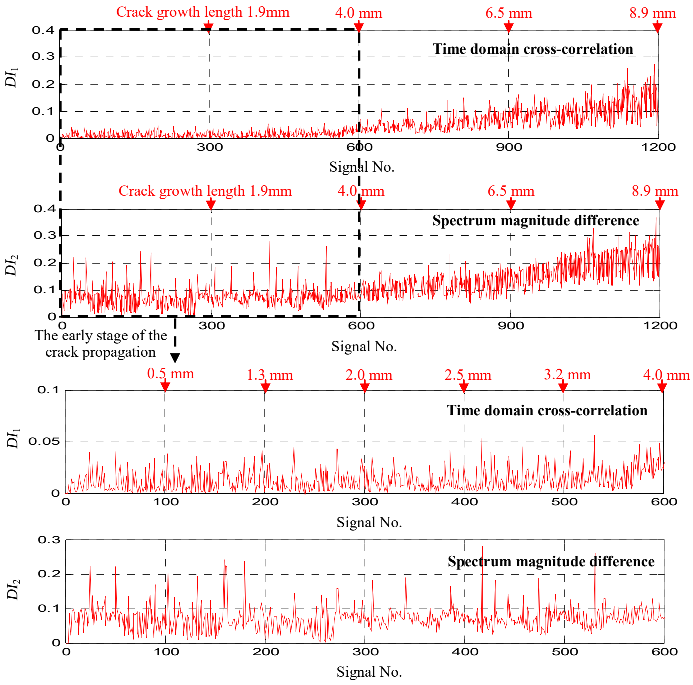
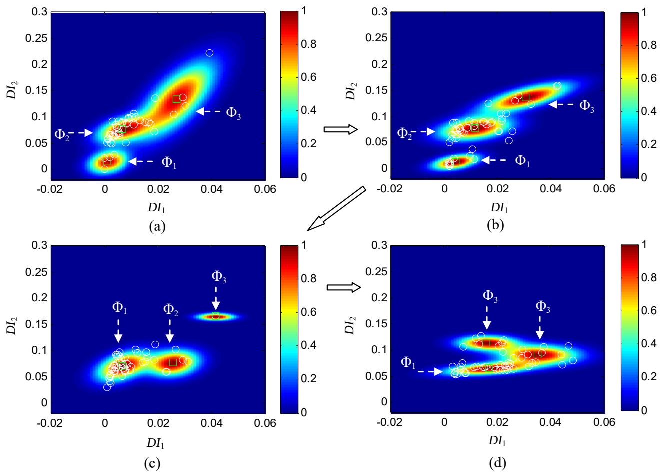
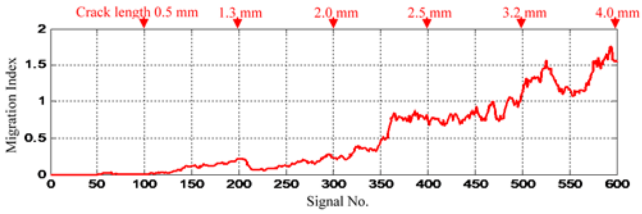


Figure 4: Signal features extracted from guided wave signal

Fig. 5 shows the GMM constructed with the guided wave damage index $z = [DI_1, DI_2]$. Once a new guided wave signal is acquired, the GMM is updated. Fig. 5a illustrates the migration of the GMMs. To evaluate the damage, the KL distance is calculated as shown in Fig. 5b. Comparting with the damage indices, the KL distance increases cumulatively and stably with the crack growth. In addition, the early crack growth can be identified with the KL distance.



(a) Illustration of the migration of GMMs



(b) KL distance of the GMMs

Figure 5: Illustration of the GMM based damage evaluation

3.2 Validation of the PF based damage diagnosis

Fig. 6 shows the fatigue test of the real aircraft attachment lug, which is a kind of important joint type of aircraft structures. This lug is made of 5 mm thick aircraft aluminium. A steel fixture is used to be assembled with this lug for transfering the stretch load. During the fatigue test, the sinusoid load with maximum load 18kN and load ratio 0.1 is applied. PZTs are arranged on the surface of the structure to excite and acquire guided wave signals with the SHM system developed by the authors [11, 12]

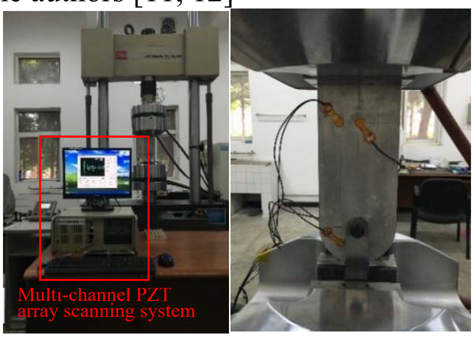
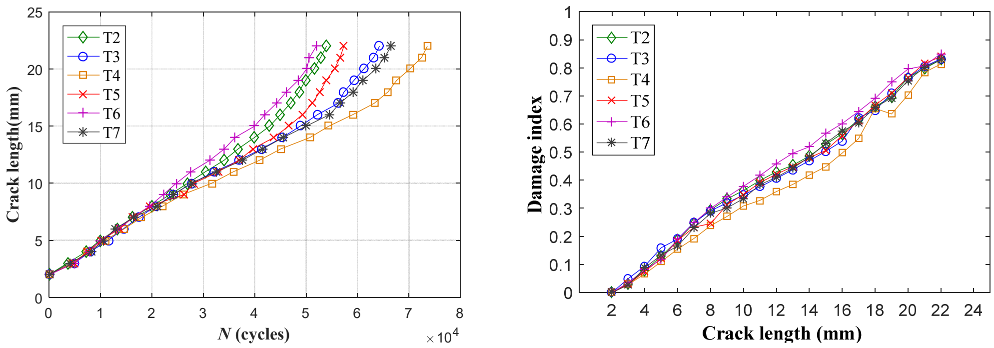


Figure 6: Illustration of the GMM based damage evaluation

As shown in Fig. 7a, experimental crack lengths are recored versus loading cycles. The fatigue crack propagation trajectories of the identity specimens represent obvious dispersion, which is caused by uncertainties arising from the intrinsic material property, specimen machining and the complexity of load transmission. Fig.7b illustrates the normalization correlation moment damage index extracted from guided wave signals. It can also be found that uncertainties introduce difficulties to the evaluation of the crack length with the guided wave SHM.



(a) Crack growth results (b) Damage index

Figure 7: Experimental results

In the validation, specimen T7 is deemed as the target structure to be diagnosed. The state space model of the attchment lug is established based on the Paris model and the data from T2-T7, which is shown in Eq. (11) and (12)

$$x_k = x_{k-1} + \exp(\omega) C_{k-1} (\Delta K)^{4.46} \times 50 \tag{11}$$

$$z_k = 5.491 \times 10^{-5} (x_k - x_0)^3 - 1.521 \times 10^{-3} (x_k - x_0)^2 + 5.120 \times 10^{-2} (x_k - x_0) + v_k \tag{12}$$

where x_k is the crack length, z_k is the guided wave damage index. The random variable ω follows the Gaussian distribution $N(-\sigma_\omega^2/2, \sigma_\omega^2)$, where σ_ω=1.2 . and the measurement noise v_k follows zero-mean Gaussian distribution $N(0, 0.028^2)$.

Once a new guided wave signal is obtained during on-line monitoring, the PF integrated the damage index extracted from guided wave signals to evluated the posterior estiamtion of the crack length. Fig. 8 shows the posterior estimation of the PF based method, which shows the effectiveness of the PF based damage diagnosis.

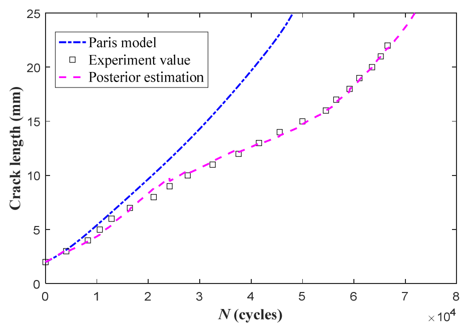


Figure 8: Posterior estimation of the crack length with the PF based method

4 CONCLUSIONS

This paper discusses applications of the probabilistic data mining based methods, the GMM and the PF, on damage diagnosis under time-varying conditions. Fatigue tests of real aircraft structures are carried out to validate these methods, in which the active guided wave based SHM is employed. The validation result shows the effectiveness of these probabilistic data mining methods.

REFERENCES

[1] Yuan, S.F. Ren, Y.Q. Qiu, L. Mei, H.F. A Multi-Response-Based Wireless Impact Monit oring Network for Aircraft Composite Structures, *IEEE T. Ind. Electron* (2016) **63**(12):77 12-7722

[2] Lim, H.J. Sohn, H. DeSimio, M.P. and Brown K. Reference-free fatigue crack detection using nonlinear ultrasonic modulation under various temperature and loading conditions Mech. *Syst. Signal.Process* (2014) **45**(2):468-78

[3] Banfield, J. and Raftery, A. Model-based Gaussian and Non Gaussian Clustering. *Biomet rics* (1993) **4**:803–821.

[4] Gordon, N.J. Salmond, D.J. and Smith, A.F. Novel approach to nonlinearn-Gaussian Bay esian state estimation, *IEEE. Proceedings. F (Radar and Signal Processing)* (1993) **140** (2):107-113.

[5] Tschope.C and Wolff, M. Statistical classifiers for structural health monitoring. *IEEE. Se ns. J* (2009) **9**(11):1567–1576.

[6] Banerjee, S. Qing, X. Beard, S. et al. Prediction of progressive damage state at the hot sp ots using statistical estimation. *J. Intel. Mat. Syst. Str* (2010) **21**(6): 595–605.

[7] Corbetta, M. Sbarufatti, C. Manes, A. et al. Real-time prognosis of random loaded structu res via Bayesian filtering: A preliminary discussion. *Eng. Fract.Mech.* (2015) **145**: 143-1 60.

[8] Rangaraj, R. Pokale, B. Banerjee, A. et al. Investigations on a particle filter algorithm for crack identification in beams from vibration measurements. *Structural. Control. And. He alth. Monitoring* (2015) **22**(8):1049-1067.

[9] Yan, R. Chen, X. Mukhopadhyay, S.C. *Advanced Signal Processing for Structural Healt h Monitoring*.Structural Health Monitoring. Springer, Cham, 2017:1-11.

[10] Qiu, L. Yuan, S.F. Shi, X.L. Huang, T.X. Design of piezoelectric transducer layer with el ectromagnetic shielding and high connection reliability. *Smart. Mater. Struct* (2012) **21** (7):075032

[11] Qiu, L. Yuan, S.F. On development of a multi-channel PZT array scanning systemand its evaluating application on UAV wing box. *Sensor. Actuat. A-Phys* (2009) **151**(2):220-230.