Report # CAIT-UTC-NC3

Bridge Health Monitoring Using a Machine Learning Strategy

FINAL REPORT
January 2017

Submitted by:

Dr. Raimondo Betti
Professor

Columbia University
640 Mudd
New York, NY 10027

External Project Manager
Dyab A. Khazem
Parsons Transportation Group
100 Broadway New York, NY 10027

In cooperation with

Rutgers, The State University of New Jersey
and
U.S. Department of Transportation
Federal Highway Administration

Disclaimer Statement

The Center for Advanced Infrastructure and Transportation (CAIT) is a National UTC Consortium led by Rutgers, The State University. Members of the consortium are the University of Delaware, Utah State University, Columbia University, New Jersey Institute of Technology, Princeton University, University of Texas at El Paso, Virginia Polytechnic Institute, and University of South Florida. The Center is funded by the U.S. Department of Transportation.

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No. CAIT-UTC-NC3	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Bridge Health Monitoring Using a Machine Learning Strategy		5. Report Date January 2017
		6. Performing Organization Code CAIT/Columbia University
7. Author(s) Raimondo Betti		8. Performing Organization Report No. CAIT-UTC-NC3
9. Performing Organization Name and Address Columbia University 640 Mudd New York, NY 10027		10. Work Unit No.
		11. Contract or Grant No. DTRT13-G-UTC28
12. Sponsoring Agency Name and Address Center for Advanced Infrastructure and Transportation Rutgers, The State University of New Jersey 100 Brett Road Piscataway, NJ 08854		13. Type of Report and Period Covered Final Report May 2014–April 2015
		14. Sponsoring Agency Code

15. Supplementary Notes

U.S. Department of Transportation/OST-R
1200 New Jersey Avenue, SE
Washington, DC 20590-0001

16. Abstract

The goal of this project was to cast the SHM problem within a statistical pattern recognition framework. Techniques borrowed from speaker recognition, particularly speaker verification, were used as this discipline deals with problems very similar to those addressed by structural health monitoring. Speaker recognition is the task of verifying whether the speaker is the individual he/she claims to be by analyzing his/her speech signal. Based on the principle that speaker recognition can determine whether it is John or Jane who says the word "mom," it was assumed that it would be possible to find out whether it is the healthy or the damaged bridge that provides that acceleration time history. Inspired by the use of Bayesian Maximum A Posteriori adaptation of the Universal Background Model in speaker recognition, this work proposed a new and computationally efficient method to update the probability distribution function describing the statistical properties of an ensemble of structural parameters through the adaptation of a Gaussian Mixture Model describing the pdf of observed indices parameterizing the structural behavior. Overall the results are very promising, but at its current stage the proposed work must be considered as a proof of concept for the application of the Bayesian MAP adaptation of a training model. The major drawback of the proposed work is its inability to produce estimates of marginal distribution of the structural parameters used as indices of the structure's performance, since the joint distribution of said variables is obtained through an implicit function of the variables of interest.

17. Key Words Structural Health Monitoring, Machine Learning	18. Distribution Statement

19. Security Classif (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No of Pages Total #45	22. Price

Form DOT F 1700.7 (8-69)

Bridge Health Monitoring Using a Machine Learning Strategy

Raimondo Betti

Columbia University

November 30, 2015

Contents

List of Figures

Executive Summary

In the last decade, there has been an substantial shift in the way the structural health of bridges and buildings is assessed. This is due mainly to the advancement of computer and sensor technology that allows computers to learn directly from the recorded data, without going through the identification of a physical model of the structure. In this scenario, taking recourse to the statistical pattern recognition approach to solve the Structural Health Monitoring assignment was recognized as ideal to efficiently account for the uncertainties inherent in the monitoring process of real structures. Central in the formulation of the statistical pattern recognition based SHM approach is the construction of a training model, representative of the probabilistic distribution of the so called damage sensitive features, that are information extracted from the measured response of the structure able to portray the structural condition. In order for the training model to be useful, it is essential to keep it up-to-date by periodically including new information on the system, whenever such information become available. In this work, a method is proposed to accomplish this task. The proposed method updates the probability distribution of the structural properties selected to monitor the structure's integrity by using a Bayesian Maximum A Posteriori *adaptation* technique commonly employed in the field of speaker recognition to verify the speaker identity. The proposed approach does not require reevaluation of the training model when new data are available, but the training model can simply be adapted by recomputing the sufficient statistics parameterizing the distribution select to represent the training model, including the new data observations. The proposed method is tested on simulated data.

1. Introduction

The continued functionality of civil infrastructure systems is a necessity for the development of modern society. Nonetheless, existing infrastructure systems are fast approaching or have already exceeded service life. Replacement of such systems is functionally and economically unfeasible. On the other hand, maintenance and retrofitting operations must be planned wisely so as to be cost-effective, given the limited financial resources. Being able to detect damage at its onset is thus of paramount importance in this scenario and Structural Health Monitoring (SHM) is the discipline able to deliver robust means of dealing with such an assignment.

Structural damage is defined as any change in the structural properties that prevents the system from performing at the desired level of safety and functionality [1]. Inherent in this definition of damage is the concept that damage detection requires a comparison between two states of the system, one of which must be representative of the reference, usually undamaged, conditions of the system. Pattern recognition is concerned with assigning classes of membership to objects, represented by their main traits, called *patterns*. Damage detection can be solved using the techniques of pattern recognition. In structural health monitoring, the patterns to analyze should be representative of the structural behavior, and it is then reasonable to extract them from measurements of the structural response. Many factors influence the structural response, leading to uncertainty in defining the exact value to assign to the aforementioned patterns. For this reason, it is reasonable to consider the patterns as random variables and solve the structural health monitoring problem within a *statistical* pattern recognition framework.

In this work, we consider some aspects of the SHM problem cast within a statistical pattern recognition framework, when using techniques borrowed from speaker recognition. In machine learning, speaker recognition is the task of assessing the speaker identity by analyzing a record of the speaker's voice. In particular, the speaker verification problem is addressed as a hypothesis testing problem, with two hypotheses, namely the null and alternative hypotheses. The null hypothesis is the one that is tested for correctness, and in a speaker verification problem is represented by the hypothesis that the claimed identity of the speaker is the speaker's true identity. The alternative hypothesis is the one complementary to the null one: if we indicate the probability of the null hypothesis as p, the probability of the alternative hypothesis is $1 - p$. Qualitatively, in speaker verification, the alternative hypothesis represents the hypothesis that the speaker's identity is any but the one claimed by the speaker. Hypothesis testing may be solved through machine learning techniques, hence, developed through the *training* and the *testing* phases. During training, a large variety of speech signals is recorded from as many different speakers as possible. The signals are then processed to extract only a reduced number of parameters able to better emphasize the vocal tract characteristics of a speaker. These parameters are the aforementioned *patterns*, often referred to also as *features*. The distribution of the features is then estimated, usually employing the maximum likelihood approach, i.e. by identifying the parameters of an assumed distribution that maximize the plausibility of the recorded data. The function that expresses the plausibility of an observation given a certain value for the distribution parameters is called likelihood, and it explains the name given to the aforementioned estimation process. The resulting model is called Universal Background Model (UBM) and it is used to model the 'alternative hypothesis'. At this point, the models to be used as representative of the null hypothesis are constructed for each speaker enrolled in the speaker verification system. These new training models are said *speaker-dependent* models. According to the guidelines of the former Bell Laboratory [2], in order to increase the efficiency and robustness of the speaker verification system, the speaker-dependent models are *adapted* from the UBM model through a Maximum A Posteriori technique. In this case, the probability of the distribution parameters given the observed data is the one subject to maximization. According to Bayes theorem, such a probability is proportional to the product of the likelihood by the so called prior probability of the parameters, which represents the assumed distribution of the parameters *before* any observation is available. The distribution of the parameters that accounts for the observations and the prior assumption on their distribution is then called *posterior*, so as to emphasize the fact that it represents an estimate of the parameters distribution *after* some observations have been collected. In order to obtain a speaker-dependent model, the features extracted from a speech signal recorded from a specific individual are used as new observations to adapt the UBM, and the resulting model is the sought speaker-dependent model. When the time comes to verify the speaker's identity, the individual to be verified is asked to utter a short sentence or even just a word and to claim his/her identity, the speech signal is then recorded and the same type of features used during training are extracted from the testing speech signal, the speaker-dependent model associated to the claimed identity is then fetched among the available trained speaker-dependent models and the likelihood of the new features observations is evaluated using both the UBM and the speaker-dependent model: if the likelihood of the features evaluated using the speaker-dependent model is larger than that evaluated using the UBM, the claimed speaker identity is verified, otherwise rejected.

The same rationale may be applied to the structural health monitoring problem. In this context, the null hypothesis is that the structure is under *reference* conditions, while the alternative hypothesis is that the structure is *not* under reference conditions. Rejection of the null hypothesis does not necessarily mean that the structure is damaged, but it means that the structure's behavior deviates from the one learnt to be representative of the reference conditions, and further monitoring is required. Deterministic SHM methods are prone to this kind of false alarm, as they do not account for the fluctuations of the structural properties induced by environmental and/or operational conditions or even measurement noise. Also probabilistic approaches can be affected by this kind of error if the training model is not periodically updated to account for the physiological changes affecting the structure's behavior in time. From this observation stems the importance of the current work.

Going back to the analogy between speaker recognition and SHM of bridges and buildings, the 'structure-dependent' model represents the distribution of the structural parameters identified from the structure under normal conditions, while the UBM should represent any possible alternative to the healthy structure. Conventionally, the alternative to the healthy structure is assumed to be the damaged structure. However, if postulated in this fashion, the problem is unsolvable, as it is apparent that the possible damage scenarios a structure can undergo are uncountable for practical purposes, as damage can occur at different locations and be of different severity and types. It will be shown that using the algorithm proposed in this work it will be possible to solve the damage detection problem without the need of modeling the damaged structure. This work, however, goes beyond the mere assessment of the structural conditions, but proposes a method to update the training model, once new measurements of the structural response are available. This task allows to construct a training model that can be used for both short and long-term monitoring. In fact, on one hand, the proposed method gives a way of updating the training model during time, so as to assess whether the evolution of the structural properties proceeds as expected (long-term monitoring problem). On the other hand, it provides an up-to-date training model that may be promptly available for testing, whenever the necessity rises of assessing whether the structure is still under reference conditions (short-term monitoring problem).

2. Theory: A Bayesian Maximum A Posteriori Adaptation of Gaussian Mixture Models

In the next sections, the MAP estimators of the parameters of the so called Gaussian Mixture Models (GMM) are derived according to the procedure proposed by Gauvain and Lee [3]. As acknowledged by the same authors, the proposed procedure follows the same rationale of the Expectation-Maximization (EM) approach, which is usually exploited to get Maximum Likelihood estimators of mixture models parameters. Since the background needed to fully understand the approach described by Gauvain and Lee is fairly different from that available to civil engineers, the explanation of the method is made step by step. The final results of the work of the two authors are given in section 2.4. The closing section of this chapter describes the practical implementation of the work by Gauvain and Lee, as proposed by Reynolds and co-authors [2].

2.1 Gaussian Mixture Models

Let us assume we have a data sample of n independent observations of p-variate random vectors collected in the data sample $D = \{\mathbf{y}_1, ..., \mathbf{y}_n\}$, and that all observations are drawn from an underlying density $f_{\mathbf{Y}}(\mathbf{y})$. If the underlying density $f_{\mathbf{Y}}(\mathbf{y})$ is defined as a finite mixture model with K components, its functional form is given by

$$f_{\mathbf{Y}}(\mathbf{y}) = f(\mathbf{y}|\boldsymbol{\Theta}) = \sum_{k=1}^{K} \omega_k f_k(\mathbf{y}|\theta_k), \tag{2.1}$$

where

- $f_k(\mathbf{y}|\theta_k)$, for $k = 1, ..., K$, are density functions with parameter(s) θ_k, called mixture components. A *Gaussian Mixture Model* is a finite mixture model where each mixture component is represented by a Gaussian density function with parameters $\boldsymbol{\mu}_k$ (mean vector) and $\boldsymbol{\Sigma}_k$ (covariance matrix):

$$f_k(\mathbf{y}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{\exp\{-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{y} - \boldsymbol{\mu}_k)\}}{\sqrt{(2\pi)^p |\boldsymbol{\Sigma}_k|}}, \tag{2.2}$$

- ω_k, for $k = 1, ..., K$, are called mixture weights, and represent the probability that a randomly selected observation $\mathbf{y}$ was generated by the k^{th} mixture component. Mixture weights must satisfy the following condition:

$$\sum_{i=1}^{K} \omega_k = 1. \tag{2.3}$$

The complete set of parameters for a Gaussian mixture model with K components is then given by

$$\Theta = \{\omega_1, ..., \omega_K, \boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_K, \boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_K\}. \tag{2.4}$$

The objective of this chapter is that of obtaining MAP estimators of the GMM parameters. In order to outline the derivation of such estimators, first some auxiliary variables, referred to as mixture indicator variables, are defined. Next, the prior probability functions to use for describing the parameters distributions are introduced. Finally, the function to maximize is derived and optimized to obtain the sought statistics of the GMM parameters.

2.2 Mixture Indicator Variables

Let us assume that the observations in $D = \{\mathbf{y}_1, ..., \mathbf{y}_n\}$ come from a GMM with K mixture components. By making this assumption, we are implying that each one of the n observations in

D comes from only one of the Gaussian mixture components, but we don't know which one. In order to reflect such an uncertainty, we can introduce an indicator variable z_{ik} that is equal to 1 if the i^{th} observation is drawn from the k^{th} mixture component, and 0 otherwise. Through the theorem of total probability, we can estimate the expected value of z_{ik}, given the observation $\mathbf{y}_i$ and an estimate of the mixture model parameters in (2.4), $\hat{\Theta}$:

$$\begin{aligned} E[z_{ik}|\mathbf{y}_i, \hat{\Theta}] &= E[z_{ik} = 1|\mathbf{y}_i, \hat{\Theta}] \cdot P(z_{ik} = 1|\mathbf{y}_i, \hat{\Theta}) + E[z_{ik} = 0|\mathbf{y}_i, \hat{\Theta}] \cdot P(z_{ik} = 0|\mathbf{y}_i, \hat{\Theta}) \\ &= 1 \cdot P(z_{ik} = 1|\mathbf{y}_i, \hat{\Theta}) + 0 \cdot P(z_{ik} = 0|\mathbf{y}_i, \hat{\Theta}) \\ &= P(z_{ik} = 1|\mathbf{y}_i, \hat{\Theta}). \end{aligned} \tag{2.5}$$

By Bayes rule, the probability that the k^{th} mixture component generated the i^{th} observation is given by:

$$P(z_{ik} = 1|\mathbf{y}_i, \hat{\Theta}) = \frac{f(\mathbf{y}_i|z_{ik} = 1, \hat{\Theta}) P(z_{ik} = 1|\hat{\Theta})}{p(\mathbf{y}_i|\hat{\Theta})}. \tag{2.6}$$

The likelihood of observing $\mathbf{y}_i$ when $z_{ik} = 1$, given an estimate of the parameter set, $\hat{\Theta}$, is represented by the k^{th} mixture component:

$$f(\mathbf{y}_i|z_{ik} = 1, \hat{\Theta}) = f_k(\mathbf{y}_i|\hat{\boldsymbol{\mu}}_k, \hat{\mathbf{\Sigma}}_k). \tag{2.7}$$

By definition, the probability that z_{ik} be equal to 1 is equal to the estimated mixture weight $\hat{\omega}_k$:

$$P(z_{ik} = 1|\hat{\Theta}) = \hat{\omega}_k. \tag{2.8}$$

Finally, the distribution of the i^{th} observation, given the estimate of the parameter set, $\hat{\Theta}$, is the mixture model itself:

$$p(\mathbf{y}_i|\hat{\Theta}) = \sum_{m=1}^{K} \hat{\omega}_m f_m(\mathbf{y}_i|\hat{\boldsymbol{\mu}}_m, \hat{\mathbf{\Sigma}}_m). \tag{2.9}$$

By plugging Equations (2.7), (2.8) and (2.9) into Equation (2.6), and then in (2.5), the expected value of the mixture indicator variables is given by

$$E[z_{ik}|\mathbf{y}_i, \hat{\Theta}] = \frac{\hat{\omega}_k f_k(\mathbf{y}_i|\hat{\boldsymbol{\mu}}_k, \hat{\boldsymbol{\Sigma}}_k)}{\sum_{m=1}^{K} \hat{\omega}_m f_m(\mathbf{y}_i|\hat{\boldsymbol{\mu}}_m, \hat{\boldsymbol{\Sigma}}_m)}. \tag{2.10}$$

This result is the founding element of the *Expectation-Maximization* procedure [10], and its use will be clear in section (2.4). The ensemble of GMM parameters Θ and mixture indicator variables z_{ik}, for $i = 1, ..., n$ and $k = 1, ..., K$, is called complete data of the GMM. Since knowledge of the mixture indicator variables is not available at the time of observing the sample D, such variables are often called *hidden variables*.

2.3 Prior distribution for the GMM parameters

Gauvain and Lee [3] suggest that a Gaussian mixture model can be interpreted as the marginal density function of the mixture model parameters expressed as the product of a multinomial density, represented by the mixture weights, and a multivariate Gaussian density, represented by the mixture components.

It can be shown that a suitable candidate to model the prior knowledge about the parameters of a multinomial distribution is the Dirichlet density function. Therefore, if we consider the ensemble of mixture weights to be distributed according to a multinomial distribution, a suitable prior for these parameters is given by

$$p(\omega_1^K|\nu_1^K) \propto \prod_{k=1}^{K} \omega_k^{\nu_k - 1}. \tag{2.11}$$

Moreover, the normal-Wishart distribution can be considered as a suitable prior to model the distribution of mean and precision matrix of a p-variate Gaussian distribution. Hence, for each set

$(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^{-1})$ of parameters of the k^{th} Gaussian mixture, the prior distribution to consider is given by:

$$p(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^{-1}|\tau_k, \mathbf{m}_k, \alpha_k, \mathbf{U}_k) \propto |\boldsymbol{\Sigma}_k^{-1}|^{\frac{\alpha_k - p}{2}} \exp\left\{-\frac{1}{2}\left[\tau_k(\boldsymbol{\mu}_k - \mathbf{m}_k)^T \boldsymbol{\Sigma}_k^{-1}(\boldsymbol{\mu}_k - \mathbf{m}_k) + \mathrm{tr}(\mathbf{U}_k \boldsymbol{\Sigma}_k^{-1})\right]\right\}. \tag{2.12}$$

Finally, by assuming independence between all sets $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^{-1})$, for $k = 1, ..., K$ and the ensemble of mixture weights, the prior distribution for the parameters in Θ is a distribution of parameters $(\nu_k, \tau_k, \mathbf{m}_k, \alpha_k, \mathbf{U}_k)$ given by

$$p(\Theta) = \prod_{k=1}^{K} \omega_k^{\nu_k - 1} |\boldsymbol{\Sigma}_k^{-1}|^{\frac{\alpha_k - p}{2}} \exp\left\{-\frac{1}{2}\left[\tau_k(\boldsymbol{\mu}_k - \mathbf{m}_k)^T \boldsymbol{\Sigma}_k^{-1}(\boldsymbol{\mu}_k - \mathbf{m}_k) + \mathrm{tr}(\mathbf{U}_k \boldsymbol{\Sigma}_k^{-1})\right]\right\}. \tag{2.13}$$

In the next section, it will be shown that the prior in Equation (2.13) is a conjugate prior with the GMM likelihood. In other words, we expect the posterior of GMM parameters to have a functional form as the one in (2.13).

2.4 MAP estimators of the GMM parameters

The Expectation-Maximization algorithm is an iterative procedure for approximating the values of distribution parameters with missing data, as is the case for the GMM, where the mixture membership of each observed data vector is unknown at the time of sampling. Usually, the EM algorithm is employed to obtain approximated ML estimates of the distribution parameters of concern. Here, instead, it is exploited to obtain approximated MAP estimates of the parameters in Θ (Equation (2.4)).

If we knew the mixture component of origin of each observation, obtaining the MAP estimates of the GMM parameters would be simple, as we could use the MAP estimators derived in Equations (**??**), (**??**) and (**??**) to retrieve the MAP estimates of mixture weights, and of mean and covariance matrix of each individual Gaussian component. If we had the mixture indicator variables associated with each of the n independent observations drawn from the GMM, the product of the joint likelihood of the complete data by the prior function for the parameters, given in Equation

(2.13), would yield:

$$q(\mathbf{y},\mathbf{z}|\Theta) = g(\Theta)\prod_{i=1}^{n}\prod_{k=1}^{K}\left[\omega_k f_k(\mathbf{y}^{(i)}|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k^{-1})\right]^{z_{ik}}, \tag{2.14}$$

so that its natural logarithm would give:

$$\ell(\mathbf{y},\mathbf{z}|\Theta) = \log(g(\Theta)) + \sum_{i=1}^{n}\sum_{k=1}^{K} z_{ik}\log\left[\omega_k f_k(\mathbf{y}^{(i)}|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k^{-1})\right], \tag{2.15}$$

where $\mathbf{y}$ is the sample of n independent observations drawn from the GMM, and $\mathbf{z}$ can be considered as the $n \times K$ matrix of mixture indicator variables associated with $\mathbf{y}$. In order to use Equation (2.15), we must get rid of the mixture indicator variables, as in practice we do not know them. To do this, we can evaluate the expected value of $\ell(\mathbf{y},\mathbf{z}|\Theta)$ over the hidden variables, given an estimate, $\hat{\Theta}$, of the parameters of interest; by exploiting the result of Equation (2.10), the sought result is given by

$$\begin{aligned} R(\Theta,\hat{\Theta}) &= E_{\mathbf{z}}[\ell(\mathbf{y},\mathbf{z}|\Theta),\hat{\Theta}] \\ &= \log(g(\Theta)) + \sum_{i=1}^{n}\sum_{k=1}^{K} c_{ik}\log\left[\hat{\omega}_k f_k(\mathbf{y}^{(i)}|\hat{\boldsymbol{\mu}}_k,\hat{\boldsymbol{\Sigma}}_k^{-1})\right] \\ &= \log(g(\Theta)) + \sum_{i=1}^{n}\sum_{k=1}^{K} c_{ik}\log\left[(2\pi)^{-\frac{p}{2}}\hat{\omega}_k|\hat{\boldsymbol{\Sigma}}_k^{-1}|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\mathbf{y}^{(i)}-\hat{\boldsymbol{\mu}}_k)^T\hat{\boldsymbol{\Sigma}}_k^{-1}(\mathbf{y}^{(i)}-\hat{\boldsymbol{\mu}}_k)\}\right], \end{aligned} \tag{2.16}$$

where

$$c_{ik} = \frac{\hat{\omega}_k f_k(\mathbf{y}_i|\hat{\boldsymbol{\mu}}_k,\hat{\boldsymbol{\Sigma}}_k)}{\sum_{m=1}^{K}\hat{\omega}_m f_m(\mathbf{y}_i|\hat{\boldsymbol{\mu}}_m,\hat{\boldsymbol{\Sigma}}_m)}. \tag{2.17}$$

The Expectation-step in the EM algorithm consists in evaluation of c_{ik}.

If we now evaluate the exponential of function $R(\Theta,\hat{\Theta})$, we get the numerator of the posterior distribution function of the complete data GMM

$$\exp\{R(\Theta,\hat{\Theta})\} = g(\Theta)\prod_{i=1}^{n}\prod_{k=1}^{K}\left[(2\pi)^{-\frac{p}{2}}\hat{\omega}_k|\hat{\mathbf{\Sigma}}_k^{-1}|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\mathbf{y}^{(i)}-\hat{\boldsymbol{\mu}}_k)^T\hat{\mathbf{\Sigma}}_k^{-1}(\mathbf{y}^{(i)}-\hat{\boldsymbol{\mu}}_k)\}\right]^{c_{ik}}, \tag{2.18}$$

where

$$f(\mathbf{y},\mathbf{z}|\hat{\Theta}) = \prod_{i=1}^{n}\prod_{k=1}^{K}\left[(2\pi)^{-\frac{p}{2}}\hat{\omega}_k|\hat{\mathbf{\Sigma}}_k^{-1}|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\mathbf{y}^{(i)}-\hat{\boldsymbol{\mu}}_k)^T\hat{\mathbf{\Sigma}}_k^{-1}(\mathbf{y}^{(i)}-\hat{\boldsymbol{\mu}}_k)\}\right]^{c_{ik}} \tag{2.19}$$

is the complete-data likelihood. By maximizing the complete-data likelihood, we obtain the following results:

$$\begin{aligned} n_k &= \sum_{i=1}^{n} c_{ik}; \\ \bar{\mathbf{y}}_k &= \frac{1}{n_k}\sum_{i=1}^{n} c_{ik}\mathbf{y}^{(i)}; \\ \mathbf{S}_k &= \sum_{i=1}^{n} c_{ik}(\mathbf{y}^{(i)}-\bar{\mathbf{y}}_k)(\mathbf{y}^{(i)}-\bar{\mathbf{y}}_k)^T, \end{aligned} \tag{2.20}$$

where n_k/n, $\bar{\mathbf{y}}_k$ and $\mathbf{S}_k/n_k$ represent approximations of the ML estimators of the mixture weights and of the individual mixture mean vector and covariance matrix, while n_k gives a measure of how many observations are expected to be in each mixture component, for the given estimate of the GMM parameters in $\hat{\Theta}$. Equations (2.20) are the formulas employed in the Maximization step of the EM algorithm to retrieve approximations of the ML estimators of a GMM with K components.

By plugging the definitions (2.20) into Equation (2.18), and noting that

$$\sum_{i=1}^{n} c_{ik}(\mathbf{y}^{(i)}-\hat{\boldsymbol{\mu}}_k)^T\hat{\mathbf{\Sigma}}_k^{-1}(\mathbf{y}^{(i)}-\hat{\boldsymbol{\mu}}_k) = n_k(\hat{\boldsymbol{\mu}}_k-\bar{\mathbf{y}}_k)^T\hat{\mathbf{\Sigma}}_k^{-1}(\hat{\boldsymbol{\mu}}_k-\bar{\mathbf{y}}_k)+\mathrm{tr}(\mathbf{S}_k\hat{\mathbf{\Sigma}}_k^{-1}) \tag{2.21}$$

after some algebraic manipulations, the following result is obtained:

$$\exp\{R(\Theta,\hat{\Theta})\} \propto \prod_{k=1}^{K} \omega_k^{\nu_k'-1} |\boldsymbol{\Sigma}_k^{-1}|^{\frac{\alpha_k'-p}{2}} \exp\left\{-\frac{1}{2}\left[\tau_k'(\hat{\boldsymbol{\mu}}_k - \mathbf{m}_k')^T \hat{\boldsymbol{\Sigma}}_k^{-1}(\hat{\boldsymbol{\mu}}_k - \mathbf{m}') + \mathrm{tr}(\mathbf{U}_k' \hat{\boldsymbol{\Sigma}}_k^{-1})\right]\right\}. \tag{2.22}$$

Equation (2.22) shows that the posterior distribution of the GMM parameters belongs to the same family of the prior distribution, with parameters $(\nu_k', \tau_k', \mathbf{m}_k', \alpha_k', \mathbf{U}_k')$ given by:

$$\nu_k' = n_k + \nu_k; \tag{2.23}$$

$$\alpha_k' = n_k + \alpha_k; \tag{2.24}$$

$$\tau_k' = n_k + \tau_k; \tag{2.25}$$

$$\mathbf{m}_k' = \frac{\tau_k \mathbf{m}_k + n_k \bar{\mathbf{y}}_k}{\tau_k + n_k}; \tag{2.26}$$

$$\mathbf{U}_k' = \mathbf{U}_k + \mathbf{S}_k + \frac{\tau_k n_k}{\tau_k + n_k}(\mathbf{m}_k - \bar{\mathbf{y}}_k)(\mathbf{m}_k - \bar{\mathbf{y}}_k)^T. \tag{2.27}$$

Hence, the prior distribution in Equation (2.13) is a conjugate prior for the GMM of the complete data. The MAP estimators of the GMM parameters are obtained by maximizing $\exp\{R(\Theta,\hat{\Theta})\}$.

In section (2.3) the prior distribution of the parameters in Θ was obtained by assuming a Dirichlet distribution for the prior of the ensemble of mixture weights, and a normal-Wishart prior distribution for the mean and precision matrix of each mixture component. Since prior and posterior distributions come from the same distribution family, the MAP estimates of the mixture weights are obtained as the mode of a Dirichlet distribution with parameter ν_k', for $k = 1, ..., K$.

Similarly, the MAP estimates of the k^{th} mixture mean vector and covariance matrix are obtained as:

$$\hat{\boldsymbol{\mu}}_k = \frac{\tau_k \mathbf{m}_k + n_k \bar{\mathbf{y}}}{\tau_k + n_k} \tag{2.28}$$

$$\hat{\boldsymbol{\Sigma}}_k = \frac{\tau_k(\mathbf{m}_k - \hat{\boldsymbol{\mu}}_k)(\mathbf{m}_k - \hat{\boldsymbol{\mu}}_k)^T + \mathbf{U}_k + n_k \mathbf{S}_k}{\alpha_k - p + n_k}. \tag{2.29}$$

2.5 Practical implementation of the MAP estimators into a Bayesian adaptation procedure

The MAP estimators of mixture weights and mixture components parameters will be used in the next chapter to *adapt* a probabilistic model with the information obtained through new observations. This is the rationale followed by the speaker recognition community when adapting the speaker specific model from the Universal Background Model, as discussed in chapter 1.

A simplified implementation of the main formulas previously derived is proposed by Reynolds and co-authors [2]. The simplification aims especially at facilitating the task of estimating the large variety of prior parameters that the rigorous use of the MAP estimates of the GMM parameters would require. Another simplification proposed by Reynolds and co-authors is to use only the diagonal terms of the covariance matrix rather than the entire matrix, i.e. to replace the covariance matrix with the so called variance matrix. In fact, experimental studies have shown that while the computational efficiency benefits immensely by considering only the variance matrix, the model accuracy is only little impaired.

Let us assume we have a sample of n independent new observations that we want to use to adapt a previously defined GMM to account for the new information. The first step in using the Bayesian adaptation approach exploits the formulas previously derived as:

$$c_{ik} = \frac{\omega_k f_k(\mathbf{y}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{m=1}^{K} \omega_m f_m(\mathbf{y}_i|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)} \tag{2.30}$$

$$n_k = \sum_{i=1}^{n} c_{ik}; \tag{2.31}$$

$$\bar{\mathbf{y}}_k = \frac{1}{n_k}\sum_{i=1}^{n} c_{ik}\mathbf{y}^{(i)}; \tag{2.32}$$

$$\mathbf{s}_k^2 = \frac{1}{n_k}\sum_{i=1}^{n} c_{ik}\mathbf{y}^{(i)^2}, \tag{2.33}$$

where ω_k, $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are the mixture weights and mixture components parameters describing the GMM for the model of interest at the time of measurement. Through this step it is "determined the probabilistic alignment of the newly observed realizations generated from the GMM of interest into the original mixture components" [2].

The next step is to update the previous estimates of mixture weights and mixture components parameters to create the adapted parameters for the k^{th} mixture:

$$\hat{\omega}_k = \frac{\alpha_k^w(n_k/n) + (1-\alpha_k^w)\omega_k}{\sum_{m=1}^{K}\alpha_m^w(n_m/n) + (1-\alpha_m^w)\omega_m} \tag{2.34}$$

$$\hat{\boldsymbol{\mu}}_k = \alpha_k^m\bar{\mathbf{y}}_k + (1-\alpha_k^m)\boldsymbol{\mu}_k \tag{2.35}$$

$$\hat{\boldsymbol{\sigma}}_k^2 = \alpha_k^v\mathbf{s}_k^2 + (1-\alpha_k^v)(\boldsymbol{\sigma}_k^2 + \boldsymbol{\mu}_k^2) - \hat{\boldsymbol{\mu}}_k^2. \tag{2.36}$$

In Equations (2.36), the coefficients α_k^w, α_k^m and α_k^v are defined as

$$\alpha_k^w = \frac{n_k}{n_k + r^w} \tag{2.37}$$

$$\alpha_k^m = \frac{n_k}{n_k + r^m} \tag{2.38}$$

$$\alpha_k^v = \frac{n_k}{n_k + r^v}. \tag{2.39}$$

The rigorous definition of α_k^m and α_k^v can be derived from the MAP estimators of the GMM parameters as presented in section 2.4. On the other hand, the formula proposed to update the mixture weights is slightly different from the one that can be obtained from section 2.4 as experimental evidence proved that the use of the first of Equations (2.36) to update the mixture weights gives better results than those obtained through the use of the original formula. For further simplifying the method, Reynolds and co-authors suggest to define each one of the same α_k^w, α_k^m and α_k^s in the same way:

$$\alpha_k^w = \alpha_k^m = \alpha_k^v = \alpha_k = \frac{n_k}{n_k + 8}, \tag{2.40}$$

as through experimental testing it was evidenced that using the actual values led to only minor improvements of the results. In the work presented in this report, the simplified implementation as suggested by Reynolds and co-authors in [2] is adopted.

The only other detail of the implementation left to unfold is the initialization of the GMM parameters. In this work, a hierarchical clustering algorithm, called G-means [11], is used both to identify the number of components best suited to model the initial set of data as well as to estimate initial values of the GMM parameters.

3. Training Model Adaptation Algorithm for Structural Health Monitoring

The aim of SHM is that of monitoring the structure's behavior so as to intervene as soon as it deviates alarmingly from its reference state. Such a reference state could be that corresponding to the structure in its current condition, considered undamaged. To do so, a model representative of the structural reference behavior must be constructed. This model was referred to as *training model* in the introduction. In statistical pattern recognition based SHM, the training model is represented by the probability density function of structural parameters easily relatable to changes in the structural performance due to damage. Using a probability distribution to describe the structure's performance provides a mean to account for the uncertainties inherent in the process of constructing the training model itself at the time of measuring the structural response. To account also for the fluctuations of the structural performance in time, the training model must be periodically updated. In this work, we explore the use of the Bayesian MAP estimation technique detailed in Chapter 2 to update the probability distribution of the indices portraying the structural performance. Often, such indices are difficult to observe directly from the measured structural response, and intermediate properties must be first extracted from the recorded response time histories and then transformed into the structural properties of interest, so as to retrieve their probability density function. In other terms, if we denote as $\mathbf{y}$ the observable properties extracted directly from the structural response, and as $\mathbf{u}$ the structural properties of interest but difficult to retrieve directly from the response data, what we need is the relationship between the distributions of $\mathbf{Y}$ and $\mathbf{U}$. Due to the complexity of the systems usually considered in civil engineering applications, it is rarely the case that such a relationship be available in its analytical form; more often it is necessary to take recourse to numerical models to approximate such a relationship. The problem of dealing with numerical models of the structure under analysis is often encountered in structural health monitoring. The acknowledgement that the numerical models should have been tuned with experimental observations of the structural behavior initiated the field of model updating [12, 13]. Soon, it was evidenced that in order to fully model the structural behavior, the effects of uncertainties stemming from various sources were to be included in the updating procedures. A variety of techniques able to solve the model updating problem within a stochastic framework have since been proposed [6, 8, 14, 15]. In the form proposed in this work, the approach here explored may be also used when addressing model updating problems within a statistical framework.

Figure 3.1 shows some of the sources of uncertainties for the typical case where the parameters selected to represent the structural performance in the numerical model, referred to as *model* structural properties in Figure 3.1, are modified so as for the modal properties of the numerical model to match those identified from the response measured directly on the system, referred to as *identified* modal parameters in Figure 3.1. On one hand, fluctuations in the environmental conditions, such as changes in temperature and humidity, provoke changes in the stiffness and mass properties of the structure, leading to fluctuations in the structural response. At the same time, the structural response is affected by fluctuations in the excitation caused by variations of the operational conditions. Variations of environmental and operational conditions are generated by phenomena for which, at most, the average behavior can be predicted, i.e. changes in environmental and operational conditions stem from *true* random processes, and the uncertainties stemming from these kinds of processes are in fact called *aleatory* uncertainties. The *measured* response is subject to both aleatory uncertainty, stemming from the actual structural response, and epistemic uncertainty, stemming from the measurement error. Sources of epistemic uncertainties stem from lack of knowledge of the process under analysis, and can theoretically be reduced by increasing the number of observations of the process under analysis. Consequently, also the identified modal properties will be subject to both aleatory and epistemic uncertainties, stemming from the

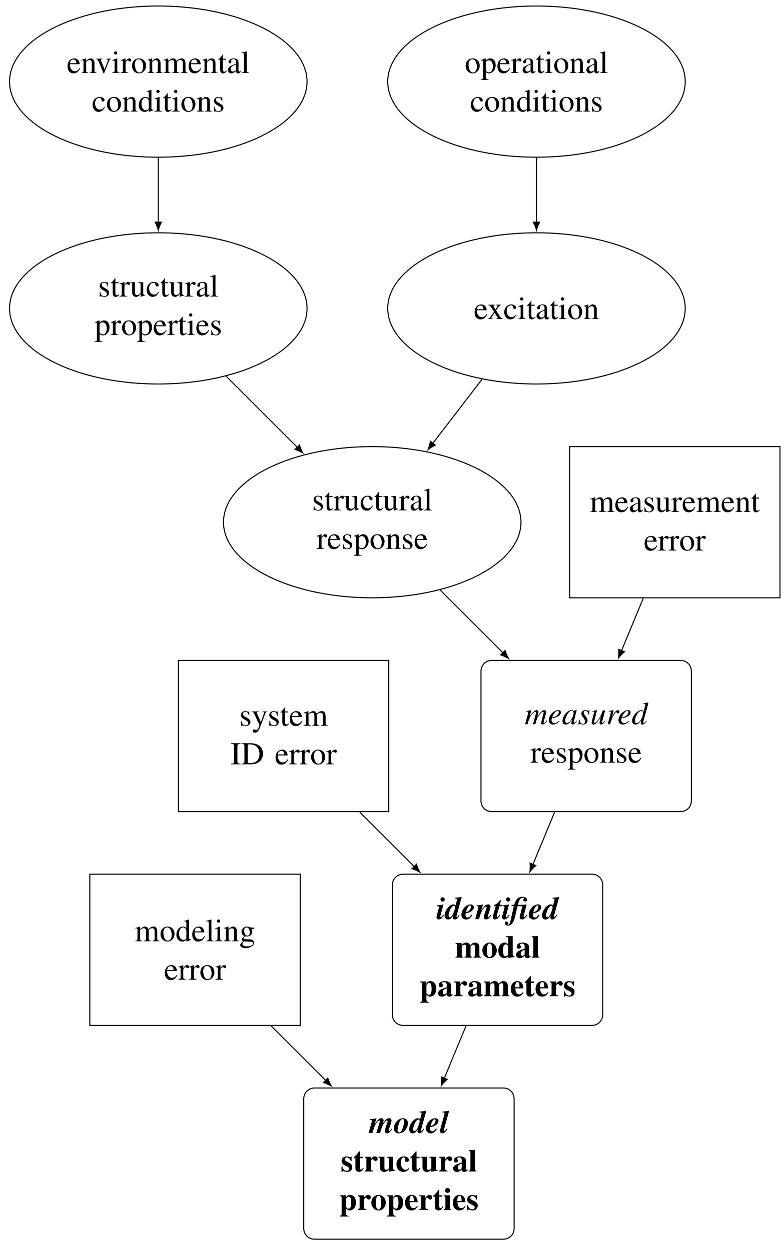


Figure 3.1: Sources of uncertainties in modal properties identified by means of output-only system identification algorithms. In ellipses sources of aleatory uncertainties, in rectangles sources of epistemic uncertainties, in rectangles with rounded edges sources of epistemic and aleatory uncertainties.

measured structural response and the error generated by the assumptions imposed by the specific system identification algorithm in use. Finally, the uncertainties on the *model* structural properties will be due to both the uncertainties coming from the identified modal properties and the error due to the modeling assumptions imposed by the numerical model in use.

Therefore, both model structural parameters and identified modal properties are affected by uncertainties that call the problem to be solved within a statistical framework. The problem we are interested in solving is that of estimating the probability distribution of some structural parameters, whose changes can be easily associated to damage occurrence, and to be able to update the estimates of such parameters, whenever new observations of the structural behavior are available. The problem is complicated by the fact that while a large variety of methods exist capable of extracting the modal properties from the measured response, the same cannot be said for the structural parameters typically used as structural damage indicators. The problem is here approached by implicitly treating the structural damage indicators as functions of the identified modal properties. While modal properties are observable, a functional form of their distribution is hard to define. Hence, the Gaussian Mixture Model is perfectly suited to describe their distribution, as a GMM represents a very general form of distribution, well suited to describe the probability properties of a vast array of random variables. Being this the case, whenever new observations of the modal properties become available, the GMM describing their distribution may be updated using the Bayesian adaptation technique discussed in chapter 2. Once the parameters of the GMM of the modal properties are adapted, we are able to update automatically also the distribution of the structural damage indicators of interest, by making use of the formula for the pdf of function of random variables given in section 2.2.

More formally, let us indicate $\mathbf{U} \in \mathbb{R}^{p\times 1}$ the random vector of structural damage indicators we are interested in and by $\mathbf{Y} \in \mathbb{R}^{m\times 1}$ the random vector of modal properties we identified. Let us assume p be equal to m, i.e. the number of structural properties of interest is equal to the number of identified modal properties. We assume the two random vectors are related through a function $g^{-1}()$:

$$\mathbf{U} = g^{-1}(\mathbf{Y}), \tag{3.1}$$

or, more precisely:

$$\begin{aligned}
\mathbf{U}_1 &= g_1^{-1}(\mathbf{Y}); \\
\mathbf{U}_2 &= g_2^{-1}(\mathbf{Y}); \\
&\vdots \qquad (3.2) \\
\mathbf{U}_p &= g_p^{-1}(\mathbf{Y}). \qquad (3.3)
\end{aligned}$$

Furthermore, we assume we know the joint distribution of $\mathbf{Y}$, that we denote as $f_{\mathbf{Y}}(\mathbf{y})$. Our objective is that of deducing the distribution of $\mathbf{U}$ in terms of $f_{\mathbf{Y}}(\mathbf{y})$, $f_{\mathbf{U}}(\mathbf{u})$, and then update it whenever new realizations of $\mathbf{Y}$ are available. From theory, we know that the distribution of the parameters $\mathbf{U}$ will be given by

$$f_{\mathbf{U}}(\mathbf{u}) = f_{\mathbf{Y}}(g_1(\mathbf{u}), ..., g_p(\mathbf{u}))|J|. \qquad (3.4)$$

While function $g^{-1}(\cdot)$ is difficult to obtain, function $g(\cdot)$ is often available, either through numerical analysis or even analytically, for simple models.

Algorithm 1 summarizes the steps proposed to update the distribution of the structural parameters, collected in a vector $\mathbf{u} \in \mathbb{R}^{p\times 1}$, when the available observations of the structural behavior are represented by properties identified or directly extracted from the measured structural response, collected in vector $\mathbf{y} \in \mathbb{R}^{p\times 1}$, and when the inverse relationship between $\mathbf{u}$ and $\mathbf{y}$ is obtained through numerical analysis. The algorithm is inspired by the approach used in speaker recognition to adapt the Universal Background Model to the speaker specific model. The first step is in fact that of forming a model playing a role similar to that played by the Universal Background Model in speaker recognition. The 'pseudo-UBM' is constructed by first selecting a lower bound $\underline{\mathbf{u}} \in \mathbb{R}^{p\times 1}$ and an upper bound $\overline{\mathbf{u}} \in \mathbb{R}^{p\times 1}$ for the structure's performance indices in $\mathbf{u}$. Then, N_m values of the vector $\mathbf{u}$ are generated, each time picking each of the p elements of $\mathbf{u}$ between the predefined bounds $\underline{\mathbf{u}}$ and $\overline{\mathbf{u}}$. Finally, from the N_m realizations of $\mathbf{u}$, as many numerical analysis are run to obtain the corresponding numerical values of $\mathbf{y}$, each time setting the uncertain values in the numerical model to the i^{th} realization of $\mathbf{u}$, for $i = 1, ..., N_m$. The result of the N_m numerical analyses is contained in the sample $\mathcal{Y} = \{\mathbf{y}^j\}$, for $j = 1, ..., N_m$. The UBM can now be estimated by estimating the distribution of the realizations in $\mathcal{Y}$. In order to exploit the theory discussed in Chapter 2, the UBM must be a Gaussian Mixture Model, where ω_k, $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$, for $k = 1, ..., K$ represent the mixture weights, mixture mean vectors and mixture variance matrices, respectively. In this work, the number of mixtures and the UBM mixture weights and components parameters are estimated through the G-means algorithm discussed in Appendinx **??**.

At this point, as soon as an ensemble of T observations of modal properties, $\mathbf{q}^{(t)}$, for $t = 1, ..., T$, is available, the mixture weights and mixture components parameters are adapted using the simplified procedure discussed in section 2.5. The GMM defined by the adapted parameters evaluated at the points obtained through numerical analysis in the initialization step and multiplied by the Jacobian of function $g(\mathbf{u})$ gives the sought distribution of the structural parameters. For what concerns the evaluation of the Jacobian, for the case where modal frequencies and mode shapes represent the identified modal properties and the structural parameters of interest depend on the stiffness and mass properties, the formulas proposed by Fox and Kapoor in [16] can be used. It is worth noting that the major burden of the approach is led by the creation of the initial UBM, but other than that the procedure is very rapid and can be performed in online mode, i.e. while the structure is operative. This is an extremely important problems in bridge engineering because such structures need to be kept operational during the monitoring process. Moreover, it is worth emphasizing that at each step of the adaptation, only the components parameters obtained at the previous iteration must be retained, making minimal the required amount of information to store. This is an important feature for a training model employed in the context of long-term monitoring applications.

Algorithm 1 MAP ADAPTATION OF TRAINING MODEL

procedure MAP ADAPTATION($\mathcal{U} = [\underline{\mathbf{u}}, \overline{\mathbf{u}}]$, measurements $\mathbf{q}_1^T$)

  Initialization:

  Generate N_m combinations of structural parameters, $\mathbf{u}^{(j)} \in \mathcal{U}$, $j = 1, ..., N_m$

  Evaluate corresponding N_m outputs, $\mathcal{Y} = \{\mathbf{y}^{(j)} : \mathbf{y}^{(j)} = g(\mathbf{u}^{(j)}), \text{for } j = 1, ..., N_m\}$

  Initializze $f_{\mathbf{Y}}(\mathbf{y}) \sim \sum_{k=1}^{K} \omega_k f_k(g(\mathbf{u})|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ using data in $\mathcal{Y}$;

  while new measurements $\mathbf{q}^{(t)}$ available, for $t = 1, ..., T$ **do**

    for $k = 1 \rightarrow K$ **do**

      c_{jk} from Equation (2.17), for $j = 1, ..., T$;

      n_k, $\hat{\mathbf{y}}_k$, and $\mathbf{s}_k^2$ from Equations (2.20);

      $\alpha_k = \frac{n_k}{n_k+8}$

      $\hat{\omega}_k = \alpha_k \frac{n_k}{T} + (1 - \alpha_k)\omega_k$

      $\hat{\boldsymbol{\mu}}_k = \alpha_k \bar{\mathbf{y}}_k + (1 - \alpha_k)\mathbf{m}_k$

      $\hat{\boldsymbol{\sigma}}_k^2 = \alpha_k \mathbf{s}_k^2 + (1 - \alpha_k)(\boldsymbol{\sigma}_k^2 + \boldsymbol{\mu}_k^2) - \hat{\boldsymbol{\mu}}_k^2$

    end for

    $\omega_1^K \leftarrow [\hat{\omega}_1, ..., \hat{\omega}_K] / \sum_{i=1}^{K} \hat{\omega}_i$

    $\boldsymbol{\mu}_1^K \leftarrow [\hat{\boldsymbol{\mu}}_1, ..., \hat{\boldsymbol{\mu}}_K]$

    $\boldsymbol{\Sigma}_1^K \leftarrow \left[\text{diag}\{\hat{\boldsymbol{\sigma}}_1^2\}, ..., \text{diag}\{\hat{\boldsymbol{\sigma}}_K^2\}\right]$

    $f_{\mathbf{U}}(\mathbf{u}) \leftarrow \sum_{k=1}^{K} \omega_k f_k(g(\mathbf{u})|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)|J|$

  end while

end procedure

4. Numerical Example

4.1 Training Model Adaptation

In order to clarify the procedure proposed in this work, as well as to better emphasize some of the details of the approach, the illustrative example discussed in reference [8] is here also considered. The advantage to use such an example is that 1) it has already been studied at great length and that 2) it is simple enough to allow us test the proposed methodology and to check its validity and limits.

A two-story one-bay moment-resisting frame with bay width of 8 m and storey height of 3 m represents the reference system (Figure 4.1). Steel with Young's modulus equal to 2×10^{11} N/m^2 and mass density of 7850 kg/m^3 is assumed to describe the frame's material properties. The four columns have a cross sectional area of 18.8×10^{-3} m^2 and moment of inertia 0.167×10^{-3} m^4, while the two beams have cross sectional area of 10.5×10^{-3} m^2 and moment of inertia 0.562×10^{-3} m^4. Both columns and beams are modeled through plane frame elements, i.e. through 2 nodes elements, where each node has three degrees of freedom:

- dof 1: translational degree of freedom along the horizontal direction;
- dof 2: translational degree of freedom along the vertical direction;
- dof 3: rotational degree of freedom around the axis perpendicular to the frame plane.

Lumped masses of 15×10^3 kg are added to both beams to account for the contribution to the structure's mass of the two floors. The system is assumed to be proportionally damped, with 1%

damping ratio for all 12 modes. The first four modes of the system are plotted in Figure 4.1. In Figure 4.1, the number in the boxes denote the element labels, while the number in circles refer to the node labels. So, for example, the column to the left side of the reader at the first floor is element 1, connecting nodes 1 and 2. Nodes 1 and 4 are fixed so that all three degrees of freedom are constrained. Hence, after imposing the boundary conditions, the frame is represented by a 12 degrees of freedom finite element model.

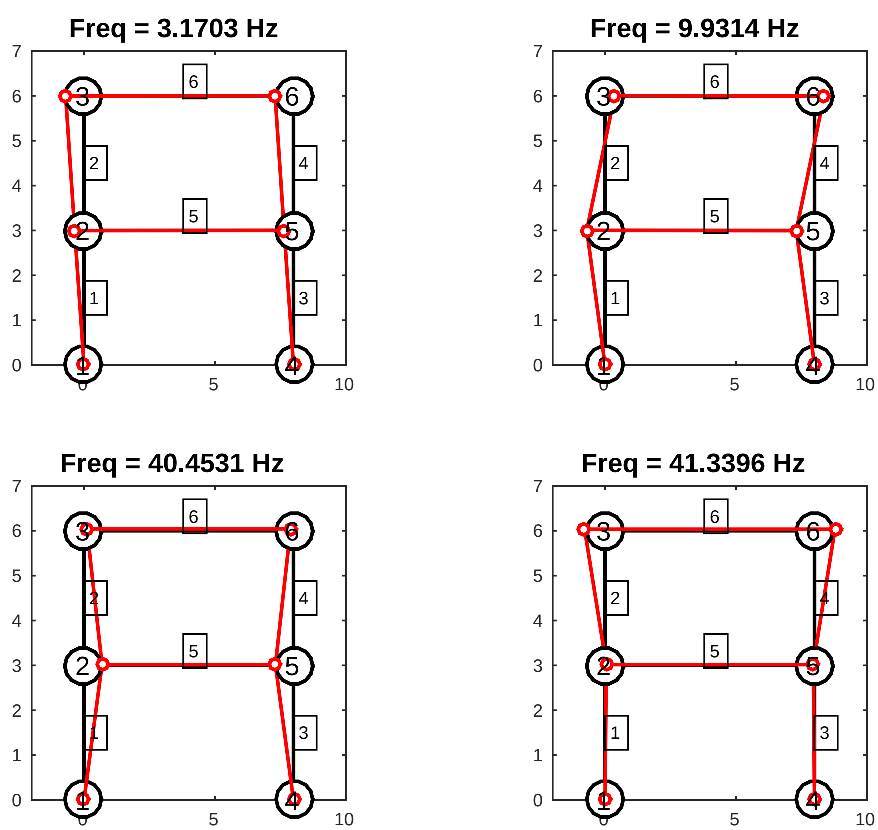


Figure 4.1: First four mode shapes of the reference system used to simulate the structural response. Modal frequencies in Hz.

This system is used to simulate the structural response. Such a simulated response will be treated as obtained from measurements of an actual system to be monitored. In particular, the acceleration response of dof 1 at node 5 is simulated by exciting the structure by means of loads applied in the horizontal direction of nodes 2, 3, 5 and 6. Forty different 'experiments' are simulated. In each experiment, the same realization of a white Gaussian noise process is applied to the four aforementioned nodes and the values of Young's modulus, cross-section area and moment of inertia of both beams and columns, as well as the steel mass density given above are perturbed by a random value uniformly distributed between $\pm$ 0.05. The input time histories and amount of structural parameters perturbation change at every simulation. The set of 40 inputs and outputs is then fed into an input-output system identification algorithm, namely OKID-ERA [17], so as to obtain 40 sets of identified modal properties. The input and output time histories are 60 seconds long and sampled at 0.01 seconds.

As done in reference [8] to introduce modeling error, rather than identifying the 12 dofs system, the reference system is identified as a 2 dofs system. In other words, when prompted to select the order of the state space model associated with the system of interest, we choose an order of 4, and not the true order 24. Being provided with only one sensor at the first floor, only the two modal frequencies can be identified. Moreover, again using the values given in [8], the 2-dofs model of the frame is assumed to have mass matrix given by

$$\mathbf{M} = \begin{bmatrix} 16.5 & 0 \\ 0 & 16.1 \end{bmatrix} \times 10^3 \text{ kg} \tag{4.1}$$

and stiffness matrix given by

$$\mathbf{K} = 29.7 \times 10^6 \begin{bmatrix} u_1 + u_2 & -u_2 \\ -u_2 & u_2 \end{bmatrix} \text{N/m.} \tag{4.2}$$

The objective of this example is that of using the algorithm outlined in Algorithm 1 to update the distribution of the structural parameters in $\mathbf{u} = [u_1, u_2]$ using the first two modal frequencies identified from the inputs and outputs 'measured' from the reference structure. Note that if only the modal frequencies are identified, there are actually two sets of $[u_1, u_2]$ that give the sought modal frequencies and that are obtained by solving for u_1 and u_2 the following system of equations:

$$\begin{cases} \sqrt{900u_1 + \frac{293400u_2}{161} - \frac{900}{161}\sqrt{25921u_1^2 - 1288u_1u_2 + 106276u_2^2}} = 2\pi 3.13; \\ \sqrt{900u_1 + \frac{293400u_2}{161} + \frac{900}{161}\sqrt{25921u_1^2 - 1288u_1u_2 + 106276u_2^2}} = 2\pi 9.83, \end{cases} \tag{4.3}$$

where 3.13 Hz and 9.83 Hz are the first two modal frequencies of the true system, as seen in Figure 4.1. Numerical solution of this system of equations yields:

$$\begin{cases} [u_1, u_2] = [1.8471, 0.2406] \quad \text{or} \\ [u_1, u_2] = [0.4871, 0.9122]. \end{cases} \tag{4.4}$$

It is worth recalling that the first step of the algorithm proposed in this work requires the construction of a model playing a role similar to that played by the UBM in speaker recognition. To construct this 'pseudo-UBM', 100 realizations of the bi-dimensional vector $\mathbf{u} = [u_1, u_2]$ are first generated. These realizations are obtained as samples of a bi-variate uniform distribution with lower limit $[0.1, 0.1]$ and upper limit $[3, 3]$. The 100 values of u_1 and u_2 are then plugged in the definition of the stiffness matrix in Equation (4.2), so as to obtain 100 different stiffness matrices that, combined with the mass matrix in Equation (4.1), are used to solve as many eigenvalue problems to obtain 100 sets of modal frequencies. The distribution of this set of modal frequencies is then estimated using G-means. The result of this estimation is given in Figure 4.2. It can be noted that the pdf is a bi-variate Gaussian distribution, but no prior assumption on the shape of the UBM distribution was made explicitly. This initial distribution can be considered as a representation of the modal frequencies properties for the class of models chosen to represent the frame under analysis. It is noteworthy that, at this point, no actual data have been used yet. The next step is in fact that of adapting the UBM model of the modal frequencies so as to account for the available measurements, and then obtaining an estimate of the distribution of the structural properties u_1 and u_2, exploiting the theory concerning the distribution of function of random variables.

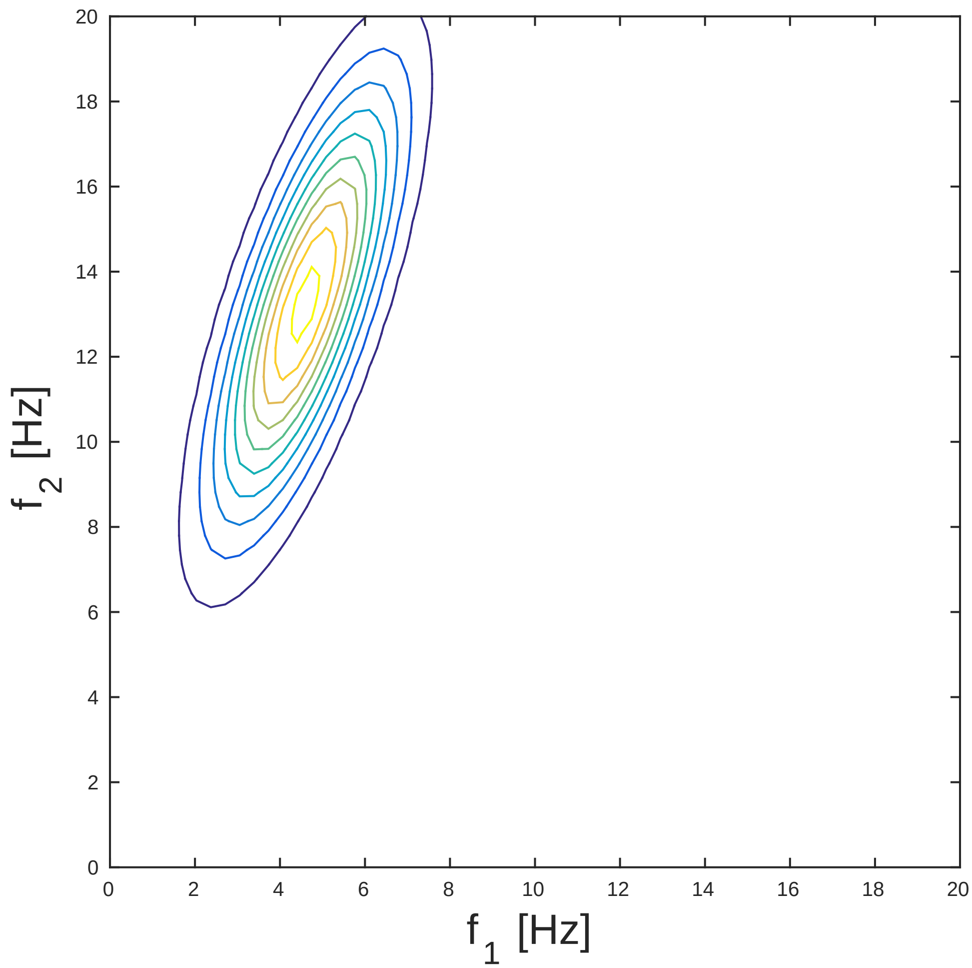
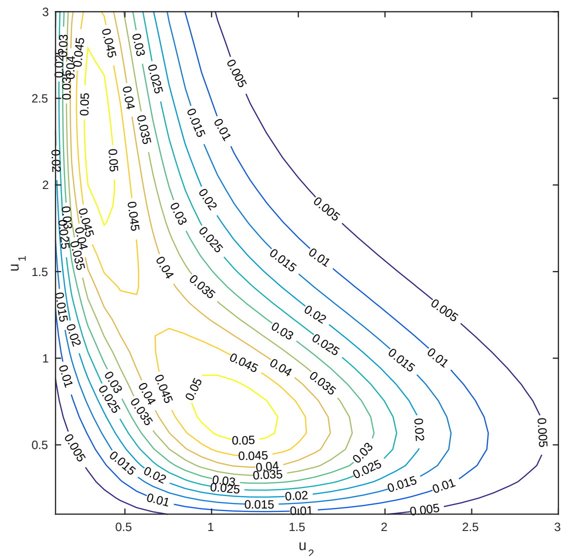
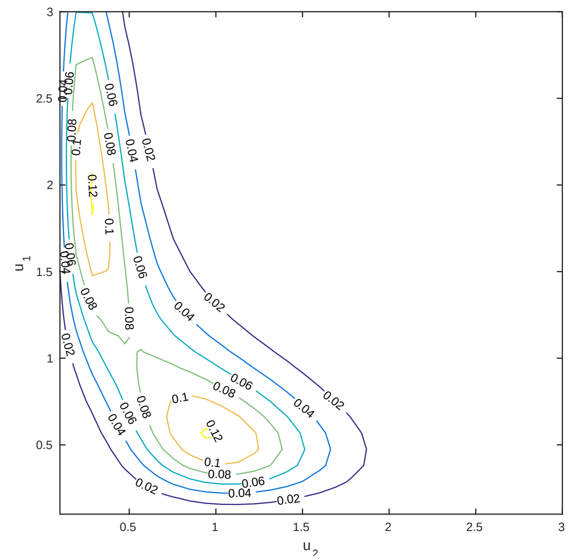


Figure 4.2: UBM for the modal frequencies.

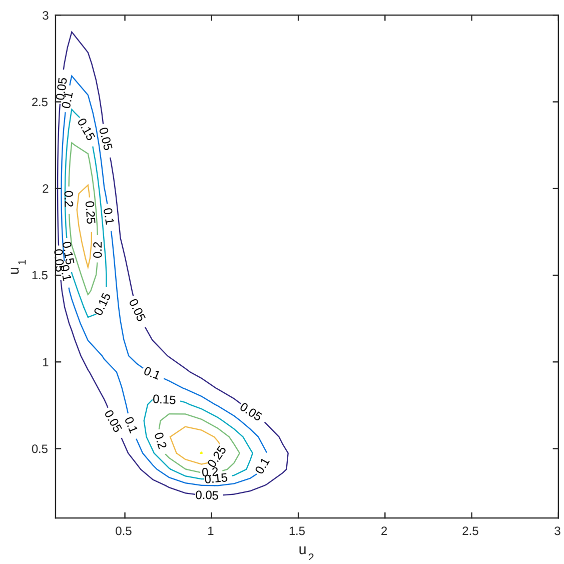
In order to show these results, the ensemble of 40 modal frequencies identified from the reference system is divided into 8 sets of 5 modal frequencies pairs. The 8 sets of modal frequencies are treated as 8 successive monitoring campaigns conducted on the structure. In practice, in each adaptation step, it is assumed that 5 new modal frequencies pairs have been observed, and such observations are used to modify the values of the UBM mixture weight, mean vector and variance matrix. Figure 4.3 shows the results of the adaptation in terms of the distribution of the structural parameters u_1 and u_2 at the second, fourth, sixth and eighth adaptation steps. Note that, at every step, the only parameters to be used are the values of the mixture parameters obtained in the previous step and the 5 modal frequencies pairs observations assumed to be currently available. In other words, only the values of the mixture parameters must be stored at every updating step, and not the data observed until that point.



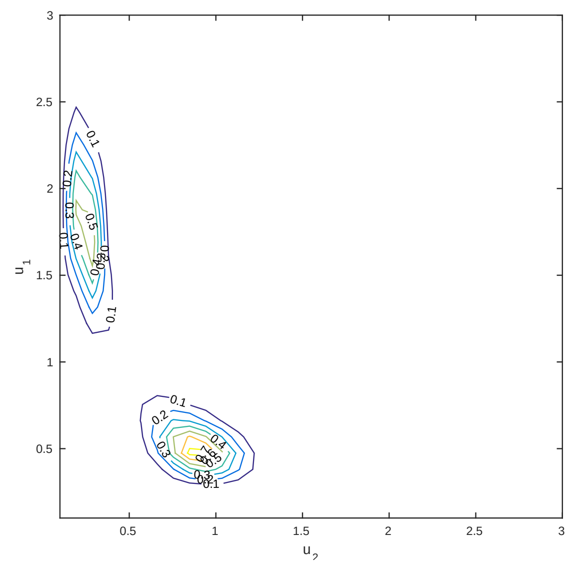
(a) Distribution of the inter-story stiffnesses after 2 sets of new data observations were included.



(b) Distribution of the inter-story stiffnesses after 4 sets of new data observations were included.



(c) Distribution of the inter-story stiffnesses after 6 sets of new data observations were included.



(d) Distribution of the inter-story stiffnesses after 8 sets of new data observations were included.

Figure 4.3: Adaptation of the structure specific model using 5 sets of modal frequencies observations at a time.

Also note that as the number of data used to adapt the model increases, the probability density function bifurcates and becomes more and more sharp around the two combinations of the bi-variate vector $\mathbf{u}$, as it can be appreciated by comparing the values where the pdf is maximum in Figure 4.3 with the values in Equations (4.4). Figures 4.3 are obtained by plotting the GMM describing the modal frequencies evaluated at the 100 sets of modal frequencies simulated to construct the UBM itself. A fundamental drawback of this approach is that the distribution of the parameters is obtained only implicitly. In practice, although the results obtained here compare well with those presented in [8], in the present form the algorithm is not able to be used any further. In fact, while that given in Figure 4.3 represents the joint distribution of the parameters u_1 and u_2, it is often of interest to evaluate the marginal distribution of the individual random variables. In the specific case under analysis, the marginal pdf of u_1 (u_2) would be obtained by integrating over u_2 (u_1) the GMM of the joint distribution of u_1 and u_2. If we had the pdf explicitly expressed in terms of u_1 and u_2 this computation would turn out to be extremely easy, thanks to the choice of using a GMM to represent the joint probability density function of the structural properties. Indeed, the marginal pdf of the i^{th} element of a random vector with joint pdf described by a p-variate GMM with K mixture components is simply a uni-variate GMM with K mixture components and mixture weights, means and variances given by the p-variate mixture weights, ω_k, the i^{th} element of the K mixture mean vectors, m_{ik}, and the $(i, i)^{th}$ elements of the K mixture variance matrices, σ_{ik}^2, for $k = 1, ..., K$, respectively. When specialized for the case under analysis, the marginal pdf of u_1, $f_{U_1}(u_1)$, and u_2, $f_{U_2}(u_2)$, are given by:

$$
\begin{aligned}
f_{U_1}(u_1) &= \int\limits_{u_2 \in \mathcal{U}_2} f_{\mathbf{U}}(\mathbf{u}) d\, u_2 \\
&= \int\limits_{u_2 \in \mathcal{U}_2} \sum_{k=1}^{K} \frac{\omega_k}{(2\pi)\sigma_{1k}\sigma_{2k}} \exp\left\{ -\frac{1}{4} \frac{(u_1 - m_{1k})^2}{\sigma_{1k}^2} \frac{(u_2 - m_{2k})^2}{\sigma_{2k}^2} \right\} d\, u_2 \\
&= \sum_{k=1}^{K} \omega_k \int\limits_{u_2 \in \mathcal{U}_2} \frac{1}{\sqrt{2\pi\sigma_{1k}^2}} \exp\left\{ -\frac{1}{2} \frac{(u_1 - m_{1k})^2}{\sigma_{1k}^2} \right\} \frac{1}{\sqrt{2\pi\sigma_{2k}^2}} \exp\left\{ -\frac{1}{2} \frac{(u_2 - m_{2k})^2}{\sigma_{2k}^2} \right\} d\, u_2 \\
&= \sum_{k=1}^{K} \frac{\omega_k}{\sqrt{2\pi\sigma_{1k}^2}} \exp\left\{ -\frac{1}{2} \frac{(u_1 - m_{1k})^2}{\sigma_{1k}^2} \right\} \int\limits_{u_2 \in \mathcal{U}_2} \underbrace{\frac{1}{\sqrt{2\pi\sigma_{2k}^2}} \exp\left\{ -\frac{1}{2} \frac{(u_2 - m_{2k})^2}{\sigma_{2k}^2} \right\}}_{\mathcal{N}(m_{2k}, \sigma_{2k})} d\, u_2 \\
&= \sum_{k=1}^{K} \frac{\omega_k}{\sqrt{2\pi\sigma_{1k}^2}} \exp\left\{ -\frac{1}{2} \frac{(u_1 - m_{1k})^2}{\sigma_{1k}^2} \right\}; \qquad (4.5) \\
f_{U_2}(u_2) &= \sum_{k=1}^{K} \frac{\omega_k}{\sqrt{2\pi\sigma_{2k}^2}} \exp\left\{ -\frac{1}{2} \frac{(u_2 - m_{2k})^2}{\sigma_{2k}^2} \right\}, \qquad (4.6)
\end{aligned}
$$

where $\mathcal{U}_1$ and $\mathcal{U}_2$ are the spaces where u_1 and u_2 are defined, respectively.

4.2 Structural Health Monitoring of the Reference System

In this example the two-storey one-bay moment resisting frame considered in the previous example is again assumed to represent the reference system. However, in this example, the aim is that of including the training model adaptation technique proposed in this work into a complete structural health monitoring assignment. In an attempt to mimic the conditions that would be encountered in reality, the reference frame structure is subjected to 100 different excitations simulated as realizations of white Gaussian noise random processes to obtain the horizontal acceleration history of node 5. Also in this example, 60 seconds long time histories sampled at 0.01 seconds are generated. At each iteration, the Young's modulus of each element is decreased by 0.02% its original value to simulate ageing effects. At the 71^{st} simulation, in addition to the ageing effects, the cross sectional area of element 1 is decreased by 5% its original value, then of 10% its original value at simulation 81, and of 15% its original value at simulation 91. Similarly, the moment of inertia of the same element is decreased by 14.3% its original value at simulation 71, by 27.1% its original value at simulation 81 and by 38.6% its original value at simulation 91. These additional variations of the structural properties are considered to simulate damage occurrence and consequent structural conditions deterioration beginning from simulation 71. Moreover, in order to simulate environmental effects, at each simulation, the values of Young's modulus, cross-sectional area, moment of inertia and mass density of all elements are further varied by a random quantity uniformly distributed between ± 0.05. Finally, to simulate measurement error, 10% RMS noise is added to both input and output time histories, since, as done for the previous example, the 100 simulated sets of input-output time histories are fed into OKID-ERA to identify the modal frequencies of the system. Figure 4.4 shows the performance of the system as a function of the horizontal component of the stiffness of columns 1 (in red) and 2 (in blue). The circles represent the 100 realizations of said stiffnesses, while the solid curves are obtained by curve fitting the realizations with a parabolic function of the form $ax^2 + b$. The main parabolic trend observed in both curves is due to the effects of the slowly varying value of the Young's modulus, as it could be observed on a real structure as an effect of ageing. The scatter of points around the main trend is due to the random variations between ± 0.05 of each structural properties, mimicking environmental effects. Finally, the sudden drop in performance at the first inter-story at simulation 71 is due to the change in cross-sectional properties of element 1, similarly to what could be observed with damage occurrence.

In order to outline the procedure adopted in this work to solve the SHM problem, let us assume that the pdf of the modal frequencies and, consequently, of the structural parameters, have already been adapted using the measurements collected during t campaigns, through the exact same approach discussed in the previous example. The resulting GMM will be referred to as *training model*. Let us denote as L_1 the likelihood of the new $(t+1)^{th}$ observation and as L_0 the likelihood of the t^{th} observation. Let us assume that each observation comprises T sets of parameters; for instance, in the previous example, T was equal to 5. The value of the likelihood L_0 is obtained as the product of the T values of the training model evaluated at each of the T observations, where the

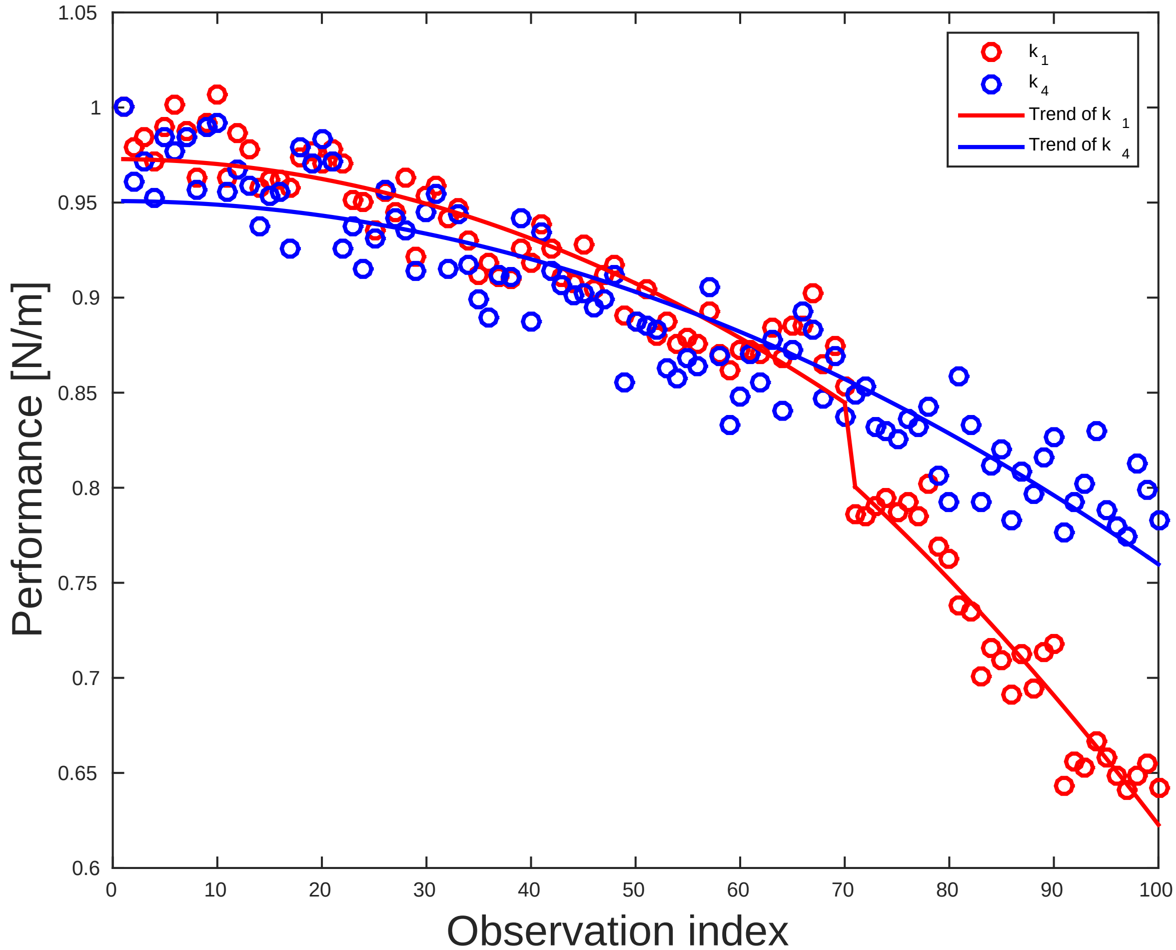


Figure 4.4: Performance of the reference system in terms of inter-storey stiffness.

GMM mixture parameters are adapted to include the observations collected at the t^{th} campaign. Likewise, L_1 is obtained as the product of the T values of the training model evaluated at the each one of the T observations, where the GMM mixture parameters are adapted to include the $(t+1)^{th}$ set of observations. If L_1 is larger than L_0, we can conclude that the new data observed at campaign $(t+1)$ come from the same type of structure we have seen up to campaign t, and we can safely include the observations collected at campaign $(t+1)$ in the training model. If, instead, L_1 is lower than the L_0, the new observations are not included in the training model as they could be associated with a system behaving differently from the one observed up to the current time, and the training model is left invariant, i.e. the observations collected at campaign $(t+1)$ are not used to update the training model.

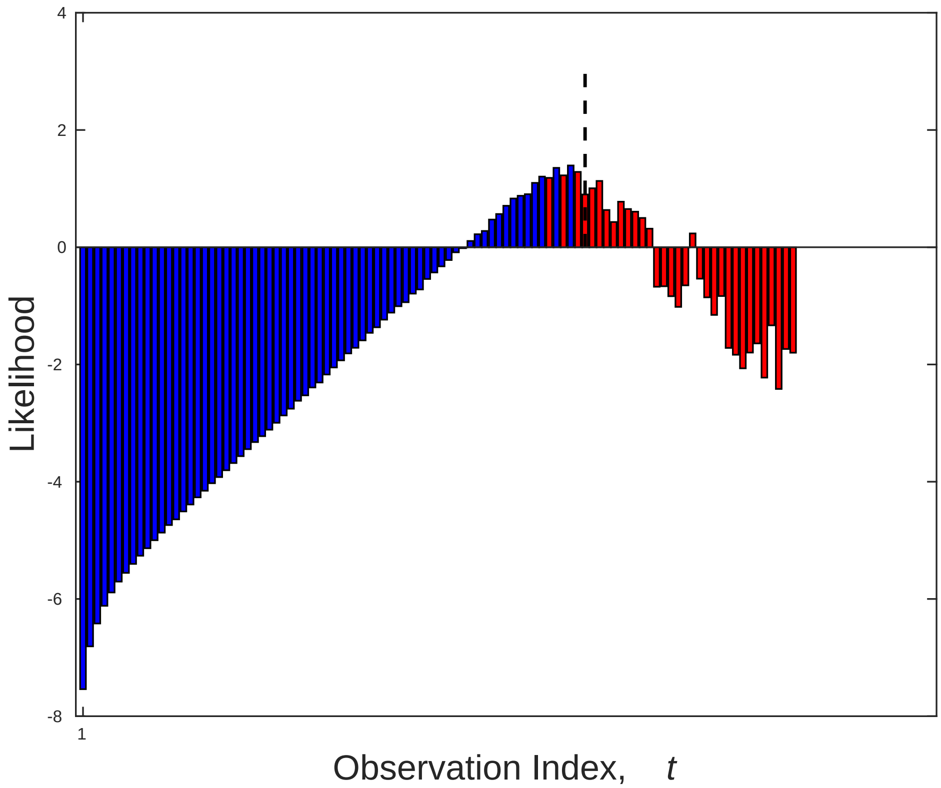


Figure 4.5: Damage detection results.

Figure 4.5 shows the results of this procedure, using each one of the 100 observations individually, i.e. T in this case is equal to 1, and t goes from 1 to 100. A blue bar indicates that the t^{th} new measurement was included in the training model and used to adapt the training model parameters, a red bar indicates that the new measurement was not used to adapt the training model. Note that the structure does not undergo a major damage until the monitoring campaign 71, but, as portrayed in Figure 4.4, its conditions deteriorate steadily up to that time, so that at simulation 65 the likelihood that the identified modal frequencies come from the system associated with the training model built up to that point begins to decrease, and observations from tests 66, 68, and 70 are not used to update the training model. Moreover, the algorithm correctly rejects the inclusion of observations from 71 to 100, as once damage occurs the class of models required to portray the system's behavior becomes different from the one associated with the constructed training model.

The evolution of the adaptation of the probability density function of the structural parameters

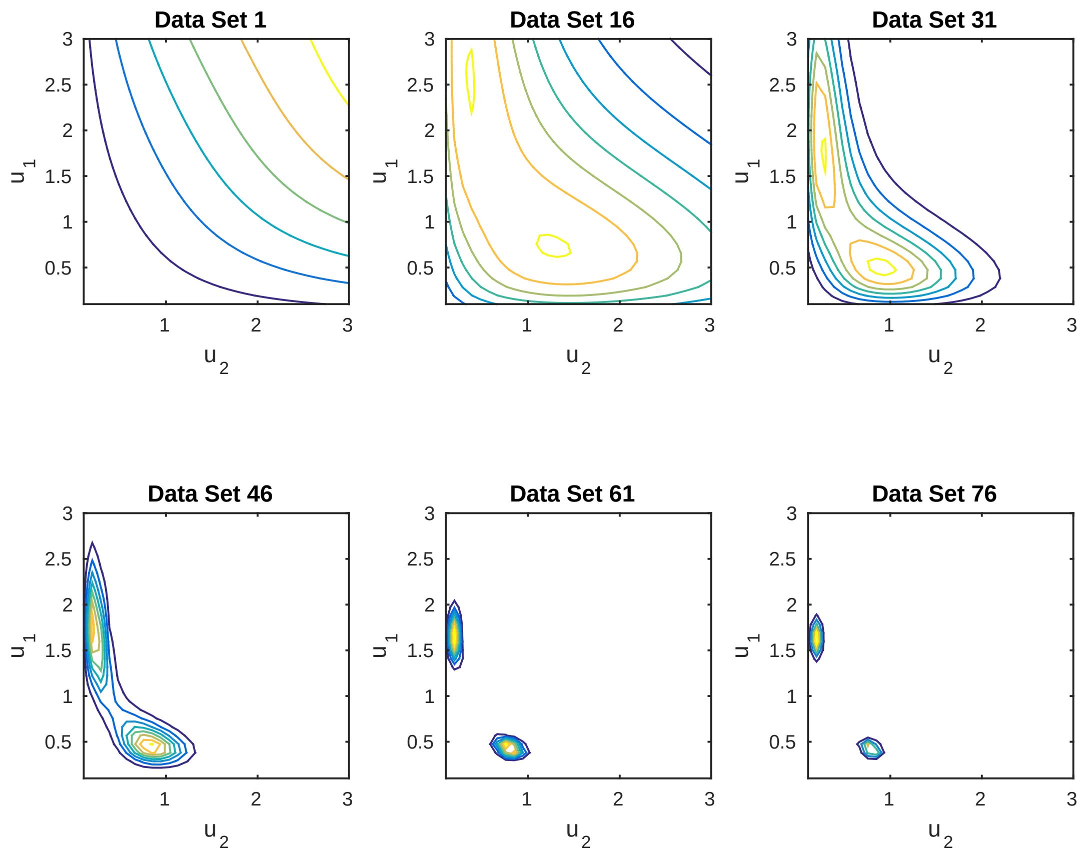


Figure 4.6: Adapted training model for the structural parameters.

is shown in Figure 4.6. Note that measurements 66, 68, 70 and from 71 on were not used to adapt the training model. Since for this example each measurement is used individually to adapt the training model, and since the reference system changes conditions within the different measurements, this time the two sets of structural properties associated with the reference system modal properties begin to be distinguished only after approximately 40 measurements. This suggests that should the algorithm be used in real applications, it is advisable to adapt it with more than one observation at a time to increase the robustness of the training model.

5. Conclusions

Inspired by the use of Bayesian Maximum A Posteriori adaptation of the Universal Background Model in speaker recognition, this work proposes a new and computationally efficient method to update the probability distribution function describing the statistical properties of an ensemble of structural parameters through the adaptation of a Gaussian Mixture Model describing the pdf of observed indices parameterizing the structural behavior. In particular, two examples have been discussed to validate the algorithm performance. In both examples, the reference system is a 12 DOFs two-storeys, one-bay moment-resisting steel frame while the model of the system is assumed to be a 2 DOFs shear-type system. In the first example, the first two modal frequencies identified from the simulated response of the 12 DOFs frame are used as observations to progressively adapt the pdf of a population of modal frequencies generated from the 2 DOFs shear-type system. This first example clarifies what has been defined training model adaptation. In the second example, the performance of the 12 DOFs reference model is made varying to mimic environmental, operational and ageing conditions variability and a criterion to select whether or not the new observations should be used to adapt the training model is also proposed.

Overall the results are very promising, but at its current stage the proposed work must be considered as a proof of concept for the application of the Bayesian MAP adaptation of a training model. The analyzed approach has various features that make it quite interesting for structural health monitoring applications, the most important of which are:

- Use of a Gaussian Mixture Model to represent the statistical distribution properties of the monitored system. The GMM is a versatile tool able to model a very large class of distributions. The fact that the GMM is a weighted sum of Gaussians makes it extremely easily to use to analyze the random variables behavior.

- Minimal amount of information to store. By means of the Bayesian MAP estimation reviewed in this work, the only parameters to store at any time are only the mixture weights and the mixture components parameters, while once the new observations are used to adapt the training model, they can be discarded without loss of information. In practice, in this work, the statistics used to estimate mixture weights, mean vectors and variance matrices are treated as sufficient statistics.

The major drawback of the proposed work is its inability to produce estimates of marginal distribution of the structural parameters used as indices of the structure's performance, since the joint distribution of said variables is obtained through an implicit function of the variables of interest. If such marginal distribution were available, the structural health monitoring approach discussed in the second example of Chapter 4 could be performed analyzing, for example, the reliability of each inter-story, by evaluating the probability that each inter-story be greater than a pre-defined threshold. By proceeding in this fashion, also damage location and severity could be estimated. Given the enormous potentials of the approach, further research is needed to identify ways to express the joint pdf explicitly in terms of the structural properties.

Bibliography

[1] C. R. Farrar and K. Worden, *Structural health monitoring: A machine learning perspective*. John Wiley and Sons, 2013.

[2] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted gaussian mixture models," in *Digital Signal Processing*, pp. 19–41, 2000.

[3] J.-L. Gauvain and C.-H. Lee, "Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains," *IEEE Transactions on Speech and Audio Processing*, vol. 2, pp. 291–298, 1994.

[4] G. Casella and R. L. Berger, *Statistical Inference*. Pacific Grove, CA: Wadsworth and Brooks/Cole, 1990.

[5] L. S. Katafygiotis and L. S. Beck, "Updating models and their uncertainties. ii: Model identifiability," *ASCE Journal of Engineering Mechanics*, vol. 124, pp. 463–467, 1998.

[6] J. L. Beck and L. S. Katafygiotis, "Updating models and their uncertianites i: Bayesian statistical framework," *Journal of Engin*, vol. 124, pp. 455–461, 1998.

[7] C. Papadimitriou, J. L. Beck, and L. S. Katafygiotis, "Updating robust reliability using structural tests data," *Probabilistic Engineering Mechanics*, vol. 16, pp. 103–113, 2001.

[8] J. L. Beck and S.-K. Au, "Beayesian updating of structural models and reliability using markov chain monte carlo simulation," *Journal of Engineering Mechanics*, vol. 128, pp. 380–391, 2002.

[9] S. H. Cheung and J. L. Beck, "Bayesian model updating using hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters," *Journal of Engineering Mechanics*, vol. 135, pp. 243–255, 2009.

[10] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[11] G. Hamerly and C. Elkan, "Learning the k in k-means," in *In Neural Information Processing Systems*, p. 2003, MIT Press, 2003.

[12] J. E. Mottershead and M. Friswell, "Model updating in structural dynamics: a survey," *Journal of Sound and Vibration*, vol. 167, pp. 347–375, 1993.

[13] M. Friswell and J. E. Mottershead, *Finite element model updating in structural dynamics*. Kluwer Academic Publishers, 1995.

[14] C. Mares, J. Mottershead, and M. Friswell, "Stochastic model updating: Part i - theory and simulated example," *Mec*, vol. 20, pp. 1674–1695, 2006.

[15] K.-V. Yuen, *Bayesian methods for structural dynamics and civil engineering*. Wiley, New York, 2010.

[16] R. L. Fox and M. P. Kapoor, "Rates of change of eigenvalues and eigenvectors," *AIAA Journal*, vol. 6, no. 12, pp. 1201–1205, 1968.

[17] H. Luş, R. Betti, and R. W. Longman, "Identification of linear structural systems using earthquake induced vibration data," *Earthquake Engineering and Structural Dynamics*, vol. 28, pp. 1449–1467, 1999.

[18] G. Hamerly and C. Elkan, "Alternatives to the k-means algorithm that find better clusterings," in *Proceedings of the ACM conference on information and knowledge management*, 2002.

[19] M. A. Stephen, "Edf statistics for goodness of fit and some comparisons," *American Statistical Association*, vol. 69, pp. 730–737, 1974.