



ELSEVIER

Contents lists available at ScienceDirect

# Mechanical Systems and Signal Processing

journal homepage: [www.elsevier.com/locate/ymssp](http://www.elsevier.com/locate/ymssp)

## A machine learning approach to nonlinear modal analysis

K. Worden<sup>a,\*</sup>, P.L. Green<sup>b</sup><sup>a</sup> Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, United Kingdom<sup>b</sup> Institute for Risk and Uncertainty, School of Engineering, University of Liverpool, Liverpool L69 3GQ, United Kingdom

### ARTICLE INFO

#### Article history:

Received 8 August 2015  
 Received in revised form  
 13 April 2016  
 Accepted 24 April 2016

#### Keywords:

Nonlinear modal analysis  
 Machine learning  
 Data-based analysis  
 Self-Adaptive Differential Evolution

### ABSTRACT

Although linear modal analysis has proved itself to be the method of choice for the analysis of linear dynamic structures, its extension to nonlinear structures has proved to be a problem. A number of competing viewpoints on nonlinear modal analysis have emerged, each of which preserves a subset of the properties of the original linear theory. From the geometrical point of view, one can argue that the invariant manifold approach of Shaw and Pierre is the most natural generalisation. However, the Shaw–Pierre approach is rather demanding technically, depending as it does on the analytical construction of a mapping between spaces, which maps physical coordinates into invariant manifolds spanned by independent subsets of variables. The objective of the current paper is to demonstrate a data-based approach motivated by Shaw–Pierre method which exploits the idea of statistical independence to optimise a parametric form of the mapping. The approach can also be regarded as a generalisation of the Principal Orthogonal Decomposition (POD). A machine learning approach to inversion of the modal transformation is presented, based on the use of Gaussian processes, and this is equivalent to a nonlinear form of modal superposition. However, it is shown that issues can arise if the forward transformation is a polynomial and can thus have a multi-valued inverse. The overall approach is demonstrated using a number of case studies based on both simulated and experimental data.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Modal Analysis is arguably *the* framework for structural dynamic testing of linear structures. While theoretical precursors existed before, the field really flourished in the 1970s with the advent of laboratory-based digital FFT analysers. The philosophy and main theoretical basis of the framework was encapsulated in the classic book [1] quite early and essentially stands to this day. The overall idea is to characterise structural dynamic systems in terms of a number of structural invariants: natural frequencies, dampings, mode shapes, and FRFs which can be computed from measured excitation and response data from the structure of interest.

While linear modal analysis has arguably reached its final form – although meaningful work remains to be done in terms of parameter estimation etc. – no comprehensive nonlinear theory has yet emerged. Various forms of nonlinear modal analysis have been proposed over the years, but each can only preserve a subset of the desirable properties of the linear. Good

\* Corresponding author.

E-mail address: [k.worden@sheffield.ac.uk](mailto:k.worden@sheffield.ac.uk) (K. Worden).

<http://dx.doi.org/10.1016/j.ymssp.2016.04.029>

0888-3270/© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

surveys of the state of the art at recent waypoints can be found in references [2,3].

Most of the theories of nonlinear modal analysis proposed so far depend on demanding algebra or detailed and intensive numerical computation based on equations of motion. The current paper proposes an approach based only on measured data from the system of interest and adopts a viewpoint, inspired by machine learning, of learning a transformation into 'normal modes' from the data.

The layout of the paper is as follows. Section 2 covers the main features of linear modal analysis, while Section 3 discusses how the approach breaks down for nonlinear systems. Section 4 gives a very condensed survey of some of the main approaches to nonlinear modal analysis taken in the past and takes some of the ideas presented as motivation for a new approach based on measured data. Examples of the new approach are presented in Section 5 via a number of case studies and the paper concludes with some overall discussion.

## 2. Linear systems

One begins here with the standard equation of motion for a Multi-Degree-of-Freedom (MDOF) linear system,

$$[m]\{\ddot{y}\} + [c]\{\dot{y}\} + [k]\{y\} = \{x(t)\} \quad (1)$$

where  $\{x(t)\}$  is an excitation force,  $\{y\}$  is the corresponding displacement response of the system and  $[m]$ ,  $[c]$  and  $[k]$  are, respectively, the mass, damping and stiffness matrices of the system. (Throughout this paper, square brackets will denote matrices, curly brackets will denote vectors and overdots will indicate differentiation with respect to time.)

Linear modal analysis is predicated on the existence of a linear transformation of the responses,

$$[\Psi]\{u\} = \{y\} \quad (2)$$

such that, the equations of motion in the transformed coordinates,

$$[M]\{\ddot{u}\} + [C]\{\dot{u}\} + [K]\{u\} = [\Psi]^T \{x(t)\} = \{p\} \quad (3)$$

have diagonal mass, damping and stiffness matrices:  $[M]$ ,  $[C]$  and  $[K]$ . The matrix  $[\Psi]$  is referred to as the *modal matrix*. The modal matrix encodes patterns of movement or coherent<sup>1</sup> motions of the structure. (It is well-known that matters are actually a little more complicated than this, e.g. complete diagonalisation of the matrices is only obtained if the damping matrix is proportional or of Rayleigh type, but the reader can consult basic texts for the details [1,4].) It immediately follows from Eq. (3) that all the degrees of freedom in the transformed coordinate system are uncoupled and each response  $u_i(t)$  satisfies a Single-Degree-of-Freedom (SDOF) equation of motion,

$$m_i \ddot{u}_i + c_i \dot{u}_i + k_i u_i = p_i \quad (4)$$

where the  $m_i$  are referred to as *modal masses*. Corresponding to each of the new degrees of freedom are the standard natural frequencies and damping ratios,

$$\omega_{ni} = \sqrt{\frac{k_i}{m_i}}, \quad \zeta_i = \frac{c_i}{2\sqrt{m_i k_i}} \quad (5)$$

It is reasonable to refer to quantities like these as *modal invariants* because they are independent of the level of excitation. It is well-known that there is a dual frequency-domain representation of Eq. (4),

$$U_i(\omega) = G_i(\omega)P_i(\omega) \quad (6)$$

where  $U_i(\omega)$  (resp.  $P_i(\omega)$ ) is the Fourier transform of  $u_i(t)$  (resp.  $p_i(t)$ ) and  $G_i(\omega)$  is a modal Frequency Response Function (FRF), defined by,

$$G_i(\omega) = \frac{1}{-m_i \omega^2 + i c_i \omega + k_i} \quad (7)$$

The overall structural FRF matrix  $[H]$  defined by,

$$\{Y(\omega)\} = [H(\omega)]\{X(\omega)\} \quad (8)$$

can be recovered as a linear sum of the modal FRFs, i.e.

$$H_{ij}(\omega) = \sum_{k=1}^N \frac{\psi_{ik} \psi_{jk}}{-m_k \omega^2 + i c_k \omega + k_k} \quad (9)$$

<sup>1</sup> Modes are commonly referred to as periodic motions of the structure; however, this term seems not to capture the spatial synchronisation aspect of the mode shape. For this reason, the term *coherent motion* is preferred here.

where the  $H_{ij}$  are also invariant under changes of excitation level. Eq. (9) is a manifestation of the *Principle of Superposition* for linear systems [4]. From this theory, one can see that modal analysis allows the decomposition of an  $N$ -DOF system into  $N$  independent SDOF systems; the dynamics of the individual SDOF systems then allow the reconstruction of the general dynamics via superposition. If only a single mode, or  $u_i$ , is excited, all the physical degrees of freedom  $y_i$  will be in linear proportion to that mode and the structure will thus execute a coherent motion; this fact can serve as an alternative definition of a mode.

### 3. Nonlinear systems

Unfortunately, most real structures have nonlinear dynamics e.g. nonlinear equations of motion [4]. The effect of structural nonlinearity on modal analysis is very destructive. Generally, all of the quantities previously described as modal *invariants* become dependent on the amplitude of excitation (energy). Furthermore, decoupling of the system into SDOF systems (certainly by linear transformation) is lost. One generally loses superposition.

However, the engineer still needs to characterise the structure and faces essentially three alternatives [5]:

1. Retain the philosophy and basic theory of modal analysis but learn how to characterise nonlinear systems in terms of the particular ways in which amplitude invariance is lost. This alternative encompasses *linearisation*.
2. Retain the philosophy of modal analysis but extend the theory to encompass objects which *are* amplitude invariants of nonlinear systems.
3. Discard the philosophy and seek theories that address the nonlinearity directly.

Although practicality will often drive one to accept option (1), this cannot be regarded as a permanent solution to the issues presented by nonlinearity. It is the belief of the current authors that one needs to proceed by developing a mixture of options (2) and (3). *Nonlinear modal analysis* – the subject of the current paper – falls within the ideas of alternatives (2) and (3).

### 4. Nonlinear modal analysis

Some would argue that *nonlinear modal analysis* does not even make sense as a term; Murdock argues [6]:

“The phrase ‘mode interactions’ is an oxymoron. The original context for the concept of modes was a diagonalizable linear system with pure imaginary eigenvalues. In diagonalized form, such a system reduces to a collection of linear oscillators. Therefore, in the original coordinates the solutions are linear combinations of independent oscillations, or modes, each with its own frequency. The very notion of mode is dependent on the two facts that the system is linear and that the modes do not interact.”

However, this is arguably somewhat pessimistic and based on a limiting semantics. In general, one might argue that engineers are not necessarily thinking of the mathematical basis when they refer to a mode; one might further argue that engineers are generally working with two main ideas regarding what a ‘mode’ is:

- (a) A coherent motion of the structure (it may be global or local).
- (b) A decomposition into lower-dimensional dynamical systems – the motions of which correspond to ‘modes’. Superposition may or may not be possible.

In general, it is not possible to keep all the properties of a linear normal mode (LNM) when passing to a nonlinear theory. Based on the two ideas above of what a mode should be, the origins of nonlinear modal analysis are centred around two main concepts for a nonlinear normal mode (NNM). Adopting definition (a) led to the idea of a *Rosenberg normal mode* [7]; definition (b) led to the idea of a *Shaw/Pierre normal mode* [8].

#### 4.1. Rosenberg normal modes

Based on idea (a) – the coherent motion concept – Rosenberg observed [7]:

- For linear systems, normal solutions are periodic with all coordinate motions sharing the same period.
- The ratios of displacements of given masses are constant for all time i.e.  $u_i = c_i u_1$ , where the  $c_i$  are constants.

Based on a number of assumptions: symmetric systems, conservative systems, no internal resonance, Rosenberg essentially defined a mode as a periodic motion of the system where the second property above was generalised to  $u_i = f_i(u_1)$ . All masses still move with the same period and all pass through equilibrium at the same time. There is no general principle

for reconstruction from these modes although Rosenberg did give special cases where it was possible. He also required that the definition reduced to the correct one for linear systems and provided a simple generalisation. Although the Rosenberg idea has generated a great deal of progress [9–11], the remainder of this paper will concentrate on the Shaw–Pierre concept.

#### 4.2. Shaw–Pierre normal modes

The Shaw–Pierre concept of a nonlinear normal mode is based on idea (b) – decomposition in terms of lower-dimensional dynamical systems [8]. They observed that linear modal analysis can be reformulated in terms of invariant subspaces: if motion is initiated with a subset of LNMs present (even one), only these modes persist. For nonlinear systems, the NNMs are defined in terms of *invariant manifolds*; if motion is initiated on such a manifold it stays there for all time.

There are a number of advantages to the Shaw–Pierre formulation over that of Rosenberg; perhaps foremost is that non-conservative systems are naturally accommodated (although recent work has also attempted to address this issue in the Rosenberg framework [12,13]). In the Shaw–Pierre approach, the equations of motion are cast in first-order form so that displacements and velocities are on equal footing. Suppose  $\{y\} = (y_1, \dots, y_n)^T$  is the vector of displacements and  $\{z\} = (z_1, \dots, z_n)^T$  is the vector of velocities, then the equations of motion are expressed as,

$$\{\dot{y}\} = \{z\} \quad (10)$$

$$\{\dot{z}\} = \{f(\{y\}, \{z\})\} \quad (11)$$

Then the NNM definition generalises the Rosenberg ansatz to say that all coordinates in an NNM are functions of a single displacement/velocity pair  $(u, v)$ ,

$$\begin{pmatrix} y_1 \\ z_1 \\ y_2 \\ z_2 \\ \vdots \\ y_N \\ z_N \end{pmatrix} = \begin{pmatrix} u \\ v \\ Y_2(u, v) \\ Z_2(u, v) \\ \vdots \\ Y_N(u, v) \\ Z_N(u, v) \end{pmatrix} \quad (12)$$

Substituting the ansatz into the equations of motion generates a system of partial differential equations,

$$\begin{aligned} \frac{\partial Y_i(u, v)}{\partial u} v + \frac{\partial Z_i(u, v)}{\partial v} u f_1(u, v, Y_2(u, v), Z_2(u, v), \dots, Y_N(u, v), Z_N(u, v)) \\ = Z_i(u, v) \end{aligned} \quad (13)$$

$$\begin{aligned} \frac{\partial Z_i(u, v)}{\partial u} v + \frac{\partial Y_i(u, v)}{\partial v} u f_1(u, v, Y_2(u, v), Z_2(u, v), \dots, Y_N(u, v), Z_N(u, v)) \\ = f_i(u, v, Y_2(u, v), Z_2(u, v), \dots, Y_N(u, v), Z_N(u, v)) \end{aligned} \quad (14)$$

Unfortunately, these equations are at least as difficult to solve as the original equations of motion; however, one can adopt a power series solution,

$$y_k = Y_k(u, v) = \sum_i \sum_j a_{kij} u^i v^j \quad (15)$$

$$z_k = Z_k(u, v) = \sum_i \sum_j b_{kij} u^i v^j \quad (16)$$

and can then solve a system of algebraic equations for the coefficients  $a_{kij}$  and  $b_{kij}$ . (Clearly, more sophisticated methods of solving the PDEs are also applicable, e.g. reference [14] presents a Galerkin-based procedure.) Shaw and Pierre also showed that the modes can be approximately recombined to the physical motions i.e. approximate superposition. A polynomial expansion of the form above will be one of the main ingredients in the new NNM procedure proposed in this paper; the other major ingredient will come from consideration of the *Principal Orthogonal Decomposition*.

#### 4.3. The principal orthogonal decomposition

All the methods presented so far are based on the equations of motion of the system of interest. However, there also exists a decomposition-motivated approach based on modes adapted to sampled time data – the Principal Orthogonal Decomposition (POD) [15]. The method is essentially Principal Component Analysis (PCA) from the discipline of multivariate statistics [16]. Just as in linear modal analysis, one adopts a linear transformation,

$$\{y\} = [\phi]\{u\} \quad (17)$$

The difference between linear modal analysis and the POD is in the transformation matrix;  $[\phi]$  is constructed as the

matrix that diagonalises the covariance matrix  $[\Sigma_u]$  of the transformed variable  $\{u\}$ , i.e.

$$[\Sigma_u] = E[(\{u\} - \{\bar{u}\})(\{u\} - \{\bar{u}\})^T] \quad (18)$$

where  $E$  denotes the expectation operator and overbars denote mean values. The first mode is then a linear combination of the  $y_i$  with maximal power; the second mode is the orthogonal combination with next greatest power etc. The combinations are termed the *Principal Orthogonal Modes* (POMs). (The usual means of constructing the transformation matrix is via singular value decomposition; in that case, the singular values reflect the power associated with each POM.) The important point for the current paper is that the POMs are *statistically independent*, they do not influence each other in any way; this idea will be adopted in the current paper as a definition of orthogonality or *normality* for NNMs in general. The reason for statistical independence is that the  $u_i$  in the transformed basis are uncorrelated because of the diagonalisation of  $[\Sigma_u]$  i.e.

$$E[(u_i - \bar{u}_i)(u_j - \bar{u}_j)] = 0 \quad i \neq j \quad (19)$$

#### 4.4. A new data-based approach

The relevant background is now in place to allow the proposal of a new approach to NNMs. Like the POD, it is a data-based approach; however, unlike the POD it can be applied to nonlinear systems. The idea is to take a transformation like the Shaw–Pierre transformation – a truncated multinomial – but in this case,

$$u_k = \phi_k^{-1}(y_1, z_1, \dots, y_N, z_N) = \sum_i a_{ki}y_i + \sum_i \sum_j a_{kij}y_iy_j + \sum_i \sum_j b_{kij}y_iy_jz_j + \sum_i \sum_j c_{kij}z_iz_j + \dots \quad (20)$$

so that the obtained  $u_k$  are statistically independent i.e. are unpredictable from each other. Adopting a machine learning approach, one can *learn* the undetermined coefficients from measured data, freeing the decomposition from complicated analysis. Furthermore, one does not need equations of motion. Motivated by the POD example, one can see that the problem can be framed as an optimisation problem. In the case of the POD, the solution can be obtained by varying the transformation coefficients in order to minimise the off-diagonal elements of the covariance matrix  $[\Sigma_u]$  or the second-order correlations as given in Eq. (19). Unfortunately, zeroing the second-order correlations is not sufficient to give independence for nonlinear systems; one would need to remove all higher-order correlations between different  $u_i$  also. Thankfully, this is not an unsurmountable problem in principle, one simply adds the correlations up to a given order in the objective function to be minimised.

As one might infer from the quote from Murdock earlier, the question of terminology concerning NNMs can sometimes evoke spirited discussion. The authors take the viewpoint that words can mean what engineers intend them to mean and are happy to define the objects constructed here as *nonlinear normal modes*. The justification for the term *nonlinear* is obvious, the term *mode* is used in the sense of a decomposition into lower-dimensional systems. Finally, the term *normal* is applied as the independence assumption adopted here can be regarded as an orthogonality condition.

The observant reader may wish to question the novelty of this approach, and they might be correct to do so; some justification is required here. What is proposed here may appear to be no more than a rather brute force approach to *independent component analysis* (ICA) [17]; the novelty here rests in the assertion that *statistical independence* is sufficient to motivate a meaningful definition of an NNM, and in the proposed optimisation framework. In fact, the analysis here represents an aspect of a more ambitious programme of work based on the concept of optimisation-based *simplifying transformations*<sup>2</sup> for nonlinear systems. The idea will be to cast various methods of nonlinear dynamic analysis into a unified machine learning framework based on optimisation; preliminary work showing that normal form analysis is accommodated by the proposed framework has been presented in [18,19]. In fact, it has already been identified that linear modal analysis is, in a sense, equivalent to ICA; reference [20] identified a one-to-one relationship between vibration modes and ICA modes. The correspondence between modal analysis and ICA allowed the development of an output-only modal analysis method based on blind source separation [21].

The optimisation-based process can now be illustrated through a number of examples, both computational and experimental.

## 5. Case studies

The method proposed here will be illustrated on a series of systems of growing complexity; the first two are based on synthetic simulated data and the last is experimental.

Throughout this section, it will prove useful to use the term ‘mode’ in a slightly different sense than in the earlier

<sup>2</sup> For want of a better name.

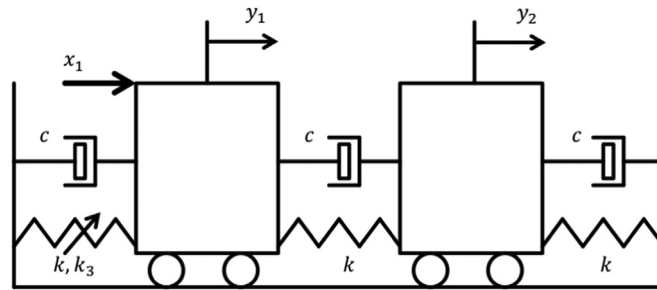


Fig. 1. 2DOF system adopted here for the case study.

discussion. In the linear theory, as discussed in Section 2, each modal DOF contributes a single resonance peak to the system FRF estimated from the physical responses. This gives the engineer or physicist another working definition of ‘mode’ i.e. the motions associated with a resonance. In the results of this section, one of the criteria for assessing the effectiveness of the transformations will be visual i.e. the extent to which the transformed spectra appear to represent a single resonance peak, so the word ‘mode’ will be used to denote a single resonance peak, which in turn will represent the belief that the transformed variable does indeed represent an SDOF system. Of course, one must take care here. If an SDOF nonlinear system is excited with high enough forcing, the FRF will also show peaks associated with harmonics of the fundamental resonance; the FRF of an MDOF system can show peaks at combination frequencies associated with the linear natural frequencies. All of this means that the association ‘peak=mode’ is an oversimplification; however, this association will prove useful in the discussion which follows.

### 5.1. A simulated 2DOF system

The system of interest will be a nonlinear two-DOF lumped parameter system as illustrated in Fig. 1. The equations of motion are,

$$\begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix} \begin{Bmatrix} \ddot{y}_1 \\ \ddot{y}_2 \end{Bmatrix} + \begin{bmatrix} 2c & -c \\ -c & 2c \end{bmatrix} \begin{Bmatrix} \dot{y}_1 \\ \dot{y}_2 \end{Bmatrix} + \begin{bmatrix} 2k & -k \\ -k & 2k \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \end{Bmatrix} + \begin{Bmatrix} k_3 y_1^3 \\ 0 \end{Bmatrix} = \begin{Bmatrix} x_1 \\ 0 \end{Bmatrix} \quad (21)$$

The model parameters adopted here were:  $m=1.0$ ,  $c=0.1$ ,  $k=10$ ,  $k_3 = 1500$ . (Note that this means the damping is

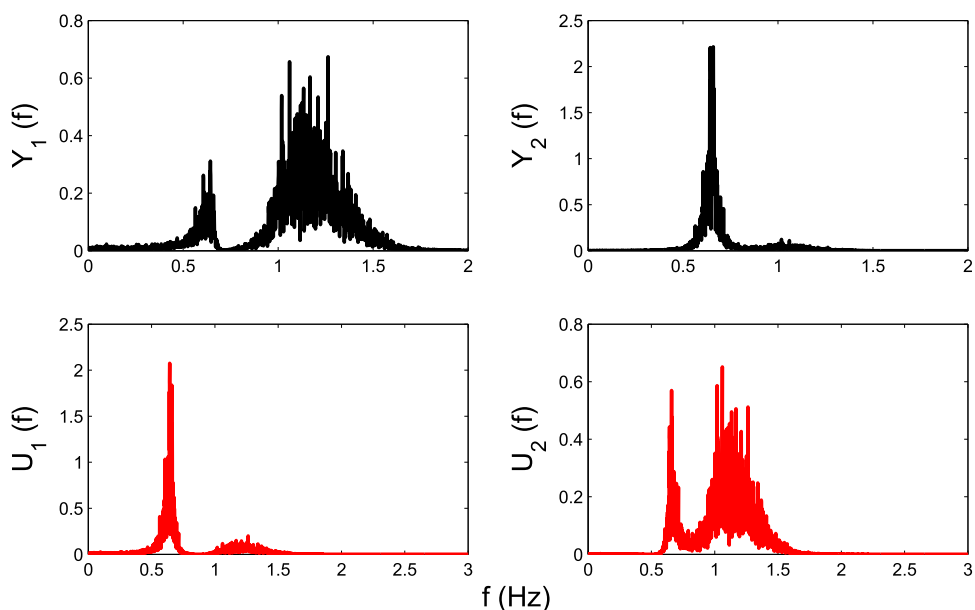


Fig. 2. 2DOF case study, PSDs for physical and transformed variables: standard linear modal analysis.

proportional, so the underlying linear system truly uncouples under the standard linear modal transformation.) With the parameter values stated, the natural frequencies of the underlying linear system are 0.5 Hz and 0.87 Hz. Data were simulated with a sampling frequency of 100 Hz using a fixed-step 4th-order Runge–Kutta algorithm and the excitation  $x_1$  was chosen to be a Gaussian white noise sequence with zero mean and standard deviation 5.0, low-pass filtered onto the interval  $[0, 50]$  Hz. In total, 50 000 points were simulated; these were mainly used in order to estimate the spectral densities shown later (by the Welch method), only 5000 points were used for the optimisation process.

For the first illustration, the  $u_i$  were computed using the standard modal transformation of the underlying linear system. Fig. 2 shows the results. Both modes are present in the power spectral densities (PSDs) for the transformed coordinates; the system is clearly not uncoupled by standard linear modal analysis, as one would expect. The level of distortion of the second ‘mode’ is very clear; the parameters of the system and the level of forcing have been chosen in order to give highly nonlinear response. Further evidence for the large effect of nonlinearity is given by the large shifts in resonance frequencies from the linear natural frequencies.

For the second illustration, the optimisation approach proposed in the last section was adopted. Because the optimisation problem is a continuous one, the Self-Adaptive Differential Evolution (SADE) algorithm was used here, as it has proved powerful for other structural dynamic problems in the recent past [22]. The object of the exercise was to learn the parameters  $a_{ij}$  of the linear transformation,

$$\begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \end{Bmatrix} \quad (22)$$

(writing the transformation in a matrix form turns out to be advantageous in terms of the Matlab [23] computation.)

In order to reproduce the results of the (linear) POD, the objective function was chosen as the sum of second-order correlations between the  $u_j$ . Because the POD is constrained to generate an orthogonal transformation matrix; a penalty term was added to the objective function to enforce the constraint. To improve the conditioning of the problem, the response data were standardised before optimisation. The objective function was thus,

$$J = |\{A_1\} \cdot \{A_2\}| + |\text{Cor}(u_1, u_2)| \quad (23)$$

where  $\{A_i\}$  is the  $i$ th column of the transformation matrix  $[A]$ , and  $\text{Cor}$  denotes the linear coefficient of correlation between the two transformed variables.

Of course, the POD transformation could be estimated using the standard eigenvalue approach; the objective here was to benchmark the SADE approach on a case where the correct transformation could be verified independently. In all the applications of SADE in this paper, 10 runs were made with the coefficients initialised randomly between  $-10$  and  $10$ . This is usual as SADE is a stochastic algorithm; one makes several runs with different initialisations in order to reduce the possibility of hitting local minima. As SADE is an evolutionary algorithm, the probability of hitting a local minimum is already reduced. As a further simplification, the expansion only used the displacement variables  $y_i$ ; this would be sufficient for the POD of a linear system.

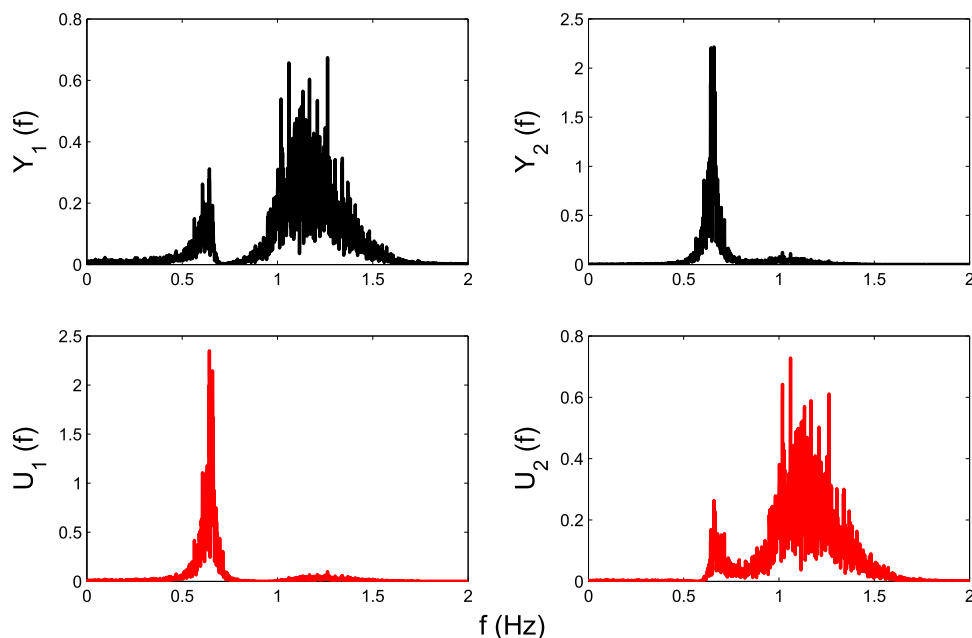


Fig. 3. 2DOF case study, PSDs for physical and transformed variables: SADE with linear transformation.

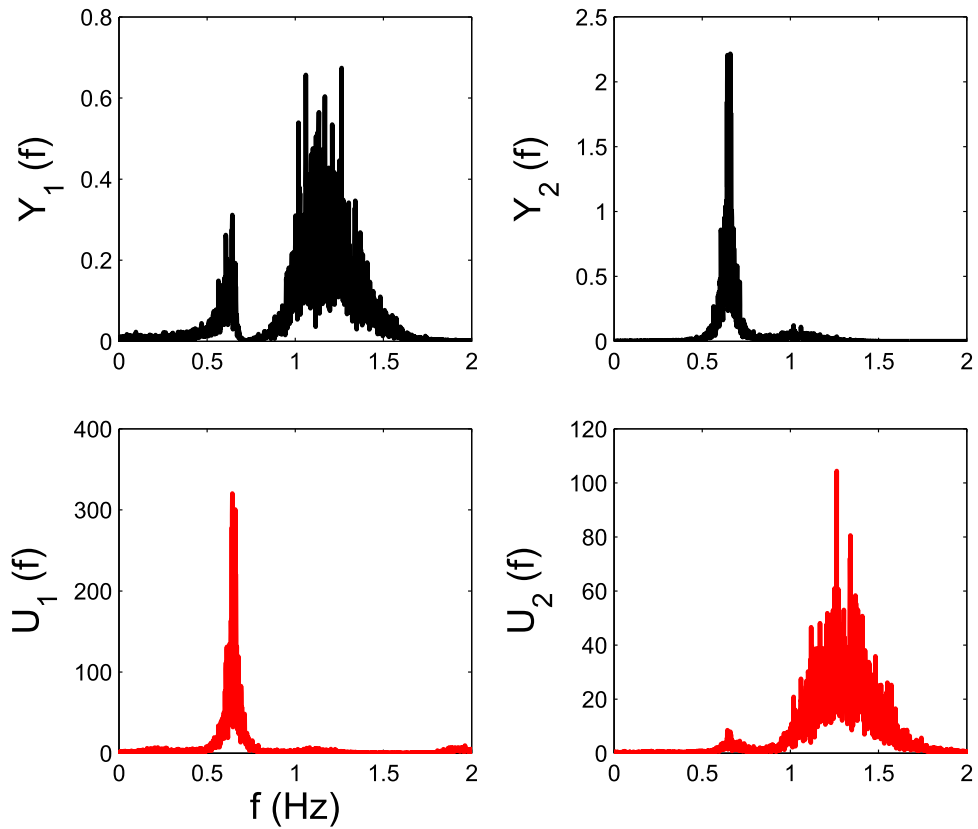


Fig. 4. 2DOF case study, PSDs for physical and transformed variables: SADE with cubic transformation (solution with lowest objective/cost function).

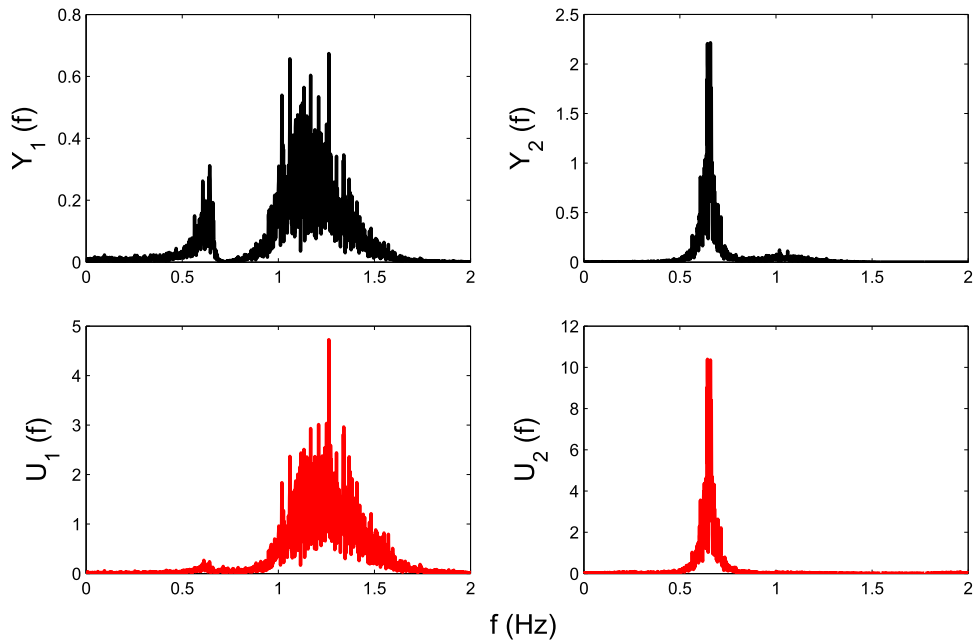


Fig. 5. 2DOF case study, PSDs for physical and transformed variables: SADE with cubic transformation (solution selected 'by eye' to best isolate single peaks).



The results of the POD computation are shown in Fig. 3; the system is not uncoupled. Although the first ‘mode’ is separated out more effectively than when the modal matrix is used, the second ‘mode’ is actually worse. The  $[A]$  matrix estimated corresponded very closely to the linear algebraic estimate via PCA.

The final illustration is based on a SADE optimisation using a cubic transformation,

$$\begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \end{Bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \end{bmatrix} \begin{Bmatrix} y_1^3 \\ y_1^2 y_2 \\ y_1 y_2^2 \\ y_2^3 \end{Bmatrix} \quad (24)$$

and an objective function which minimised all correlations up to third order. Again, to simplify the computation, the terms with velocities were disregarded. As in the last example, a penalty function was used to impose orthogonality on the  $\{A_i\}$  coefficient vectors obtained; however, it is not immediately clear what constraints should be used for the nonlinear problem as the transformation is nonlinear. Orthogonality has the advantage that it reduces to the correct constraint in the linear case. (It is clearly necessary that any credible definition of an NNM should reduce to the definition of an LMN in the limit that a system behaves linearly.) To help the search, the linear transformation coefficients were initialised in a narrow interval around the linear PCA values; the  $b$  coefficients were initialised in the interval  $[-100, 100]$ . The results of the computation are shown in Fig. 4.

The results show a major improvement on the linear results: modal and POD. There seems to be excellent uncoupling into what appear to be individual ‘modes’. This observation is of course based upon the preconception, discussed earlier, that a ‘mode’ would reveal itself as an isolated peak in the spectrum, and is an artefact of linear thinking. However, there is an issue with the current results as follows. Across the 10 SADE runs performed, the transformations varied in their ability to uncouple the DOFs when the results were judged ‘by eye’, based on visual inspection and armed with linear preconceptions. A neighbouring solution with higher cost function actually appeared to achieve better separation ‘by eye’ and the results from this transformation are shown in Fig. 5.

This result was not unexpected; the reason lies with the definition of the objective/cost function. The problem is that the function used in the optimisation does not guarantee statistical independence, it is only designed to remove all correlations between variables up to third order. This means that correlations above third order are not controlled. Suppose that the transformation with lowest cost leaves substantial correlations at fourth order; it is plausible that a less successful transformation on the lower-order correlations may accidentally have better behaviour on the higher-order ones and thus represent an improvement in terms of true statistical independence. There are other potential avenues for improvement on the cost function. Firstly, it is not immediately clear what constraints should be imposed on the coefficient vectors for the transformations; even if orthogonality is one necessary constraint, there may be others. Secondly, the constraint should be imposed using a Lagrange multiplier in the objective function given in Eq. (23); the value of the multiplier was set arbitrarily to unity; in the usual practice of machine learning the value of the weighting should be set in a principled manner like cross-validation on an independent data set. Finally, the transformation here was restricted to use displacement variables only even though the system simulated had damping, so velocities should perhaps have been included. Presumably, velocities would also be needed if the (linear) damping was non-proportional. These matters certainly require further investigation.

## 5.2. A simulated 3DOF system

The second simulated system is simply a three-DOF analogue of the first, a chain system with equations of motion,

$$\begin{bmatrix} m & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & m \end{bmatrix} \begin{Bmatrix} \ddot{y}_1 \\ \ddot{y}_2 \\ \ddot{y}_3 \end{Bmatrix} + \begin{bmatrix} 2c & -c & 0 \\ -c & 2c & -c \\ 0 & -c & 2c \end{bmatrix} \begin{Bmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \end{Bmatrix} + \begin{bmatrix} 2k & -k & 0 \\ -k & 2k & -k \\ 0 & -k & 2k \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \\ y_3 \end{Bmatrix} + \begin{Bmatrix} k_3(y_1)^3 \\ 0 \\ 0 \end{Bmatrix} = \begin{Bmatrix} x_1(t) \\ 0 \\ 0 \end{Bmatrix} \quad (25)$$

with the same mass, damping and stiffness parameters as before. The excitation force and sampling parameters were also the same. The full nonlinear transformation adopted for SADE is given by,

$$\begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \\ y_3 \end{Bmatrix} + \begin{bmatrix} b_{11} & \dots & b_{19} \\ b_{21} & \dots & b_{29} \\ b_{31} & \dots & b_{39} \end{bmatrix} \begin{Bmatrix} y_1^3 \\ y_1^2 y_2 \\ y_1^2 y_3 \\ y_2^3 \\ y_2^2 y_1 \\ y_2^2 y_3 \\ y_3^3 \\ y_3^2 y_1 \\ y_3^2 y_2 \end{Bmatrix} \quad (26)$$

with only the  $a_{ij}$  parameters used for the PCA/POD example. Once again, the notation  $[A] = [\{A_1\} \{A_2\} \{A_3\}]$  is adopted. Third-order correlations were included in the objective function as before, thus it was given by,

$$\begin{aligned} J = & |\{A_1\} \cdot \{A_2\}| + |\{A_1\} \cdot \{A_3\}| + |\{A_2\} \cdot \{A_3\}| \\ & + |\text{Cor}(u_1, u_2)| + |\text{Cor}(u_1, u_3)| + |\text{Cor}(u_2, u_3)| \\ & + |\text{Cor}(u_1^3, u_2)| + |\text{Cor}(u_1^3, u_3)| + |\text{Cor}(u_2^3, u_1)| \\ & + |\text{Cor}(u_2^3, u_3)| + |\text{Cor}(u_3^3, u_1)| + |\text{Cor}(u_3^3, u_2)| \end{aligned} \quad (27)$$

As a basis for comparison, the results of attempting to decouple the system by using the modal matrix of the underlying nonlinear system are shown in Fig. 6. The results are very poor in terms of decoupling, the second and third modes are dominated by the first mode in all the transformed degrees of freedom.

When linear SADE (corresponding to PCA/POD) is applied, the results are somewhat better (Fig. 7); the first mode still dominates; however, the second and third modes are better represented in  $u_2$  and  $u_3$ .

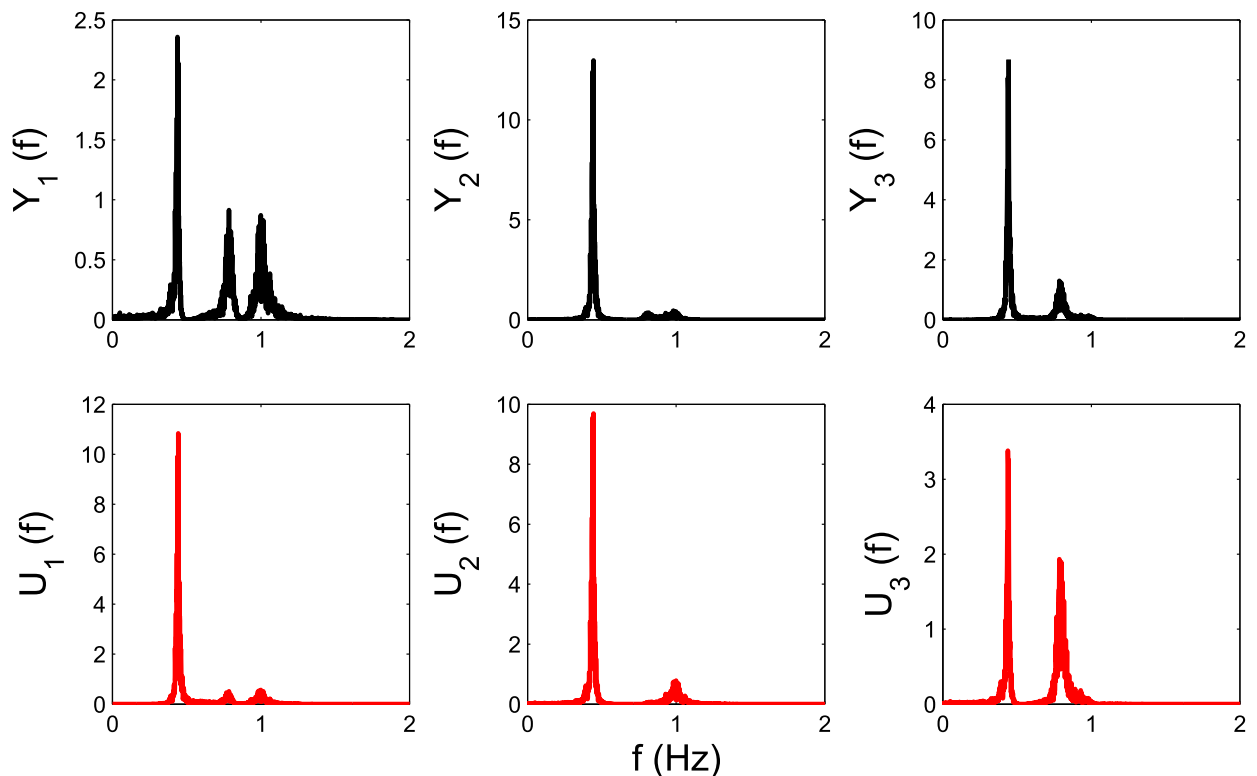


Fig. 6. 3DOF case study, PSDs for physical and transformed variables: standard linear modal analysis.

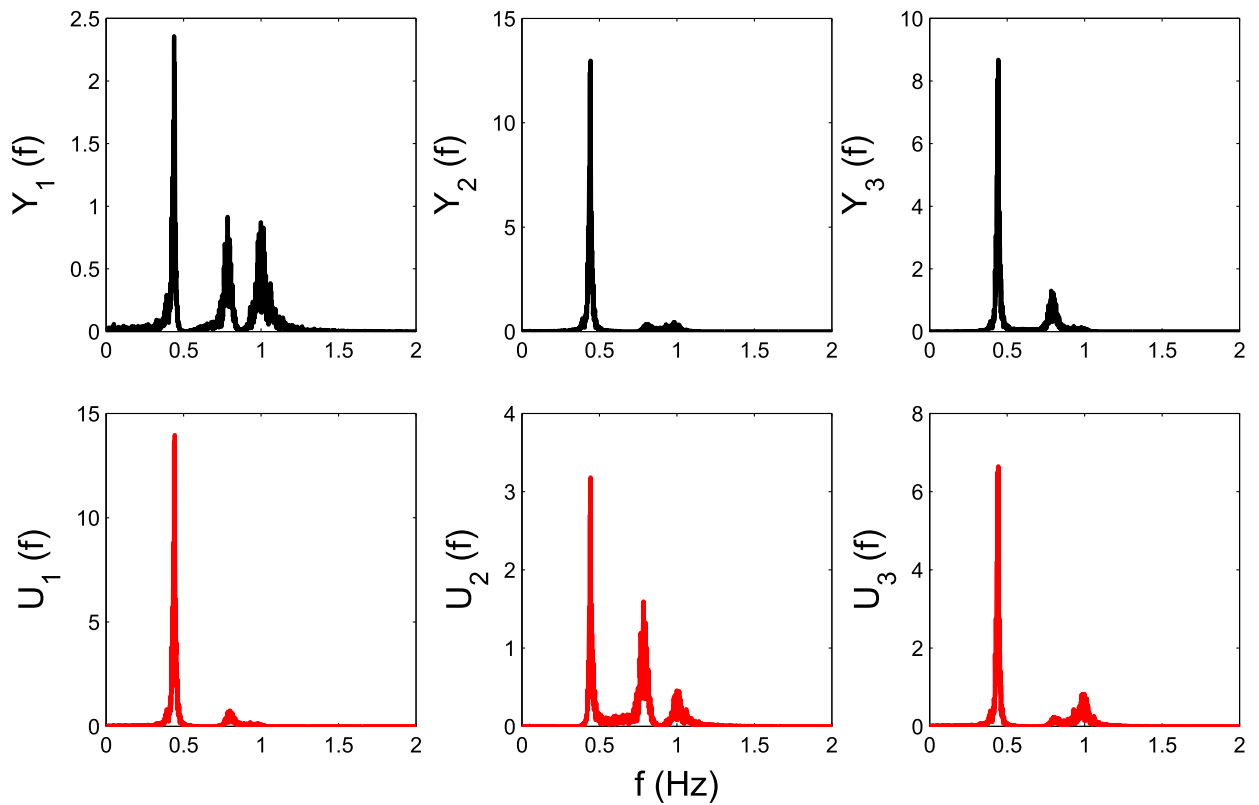


Fig. 7. 3DOF case study, PSDs for physical and transformed variables: SADE with linear transformation.

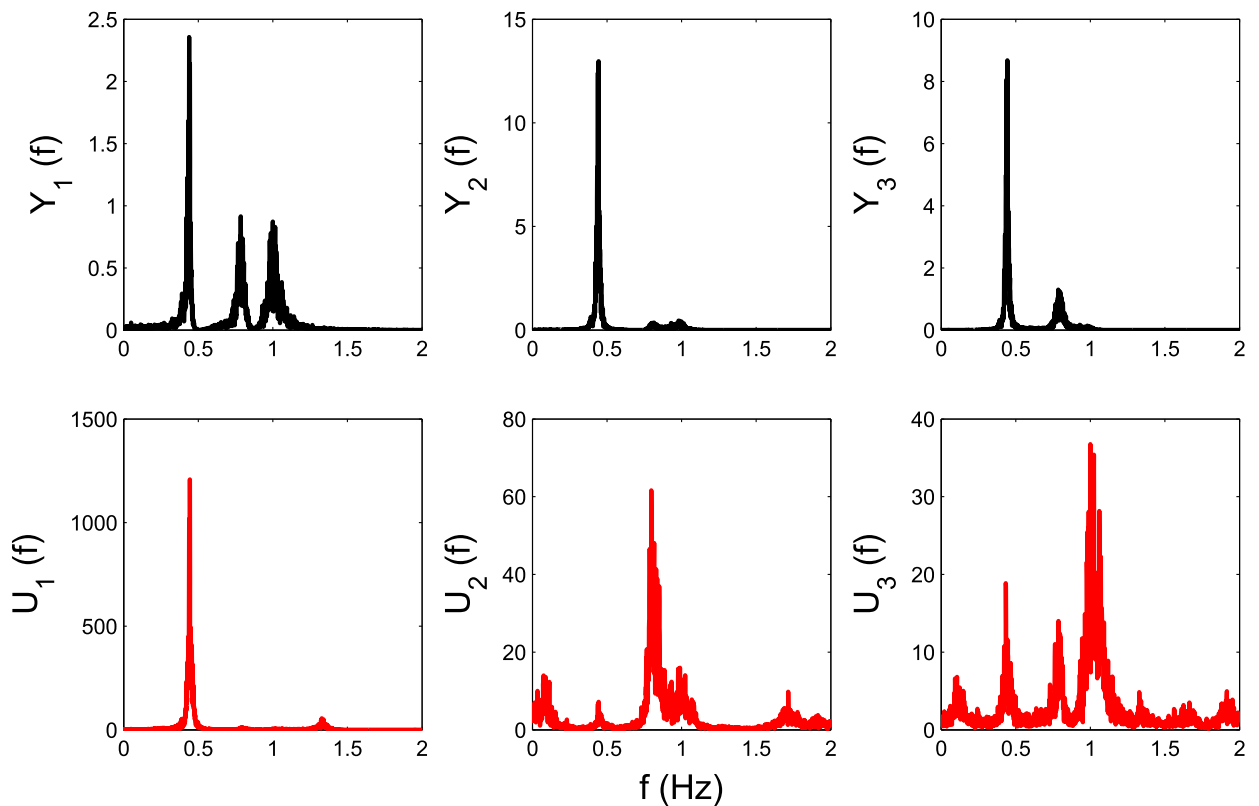


Fig. 8. 3DOF case study, PSDs for physical and transformed variables: SADE with cubic transformation (solution with lowest objective/cost function).

Finally, the nonlinear (cubic) SADE transformation was applied; the results are shown in Fig. 8. The nonlinear transformation is far more effective in separating the modes; the first mode only substantially appears in  $u_1$ , where it is almost

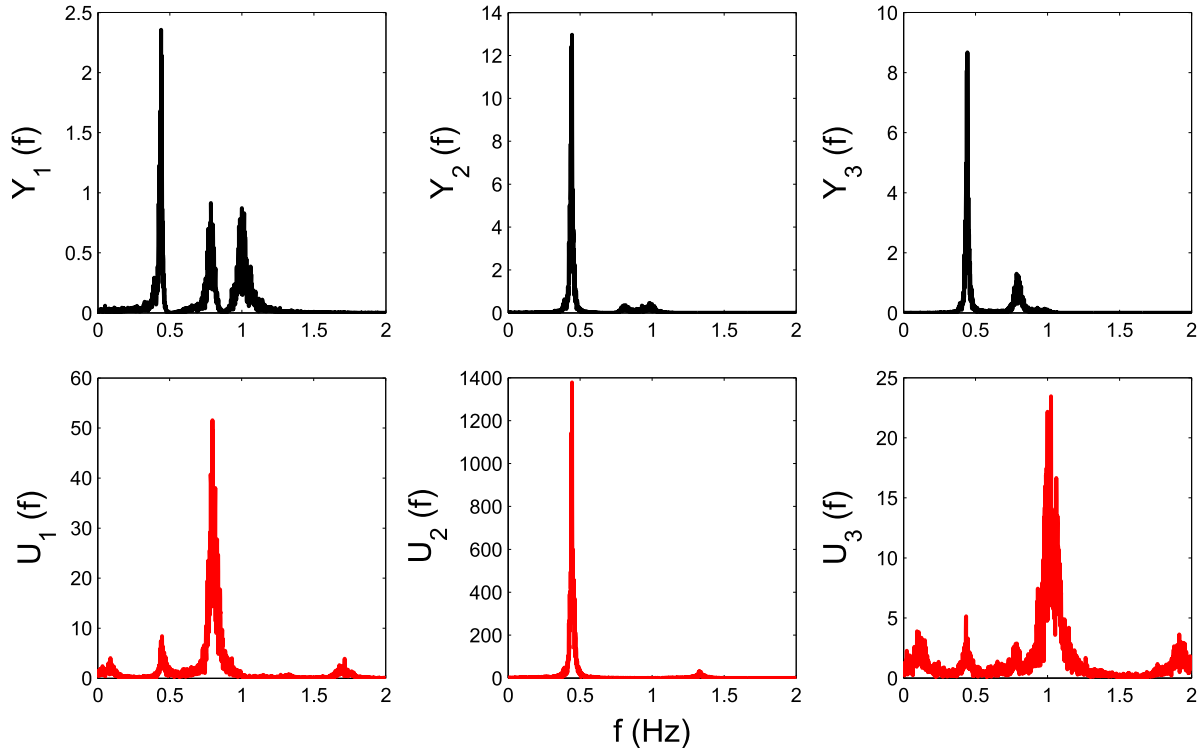


Fig. 9. 3DOF case study, PSDs for physical and transformed variables: SADE with cubic transformation (solution selected 'by eye' to best isolate single peaks).

perfectly isolated. Furthermore, much clearer emphasis on the second and third modes is obtained in  $u_2$  and  $u_3$ . In the lower amplitude  $u_2$  and  $u_3$ , one can see that the nonlinear transformation has induced, or rather amplified, some components at combination frequencies e.g. the second harmonic of the third mode is visible in  $u_3$ . Although the results are excellent, there is some degradation compared with the 2DOF results; this might be expected on the basis that many more coefficients are included in the optimisation problem. As in the 2DOF case, the SADE population produced a transformation with higher cost which appeared to decouple better *by eye*, and this is shown in Fig. 9.

### 5.3. An experimental system

The final case study is just intended as a demonstration on experimental data, rather than on a substantially more complex system. The data were taken from the bookshelf structure developed at Los Alamos National Laboratories within the Engineering Institute [24]. The structure is a three-storey base-excited model of a shear building. Although the overall structure is linear, it is possible to engage a bumper mechanism between floors which introduces a nonlinear contact mechanism. The data studied here correspond to a situation where the bumper is engaged and the dynamics are therefore nonlinear. Fig. 10 shows response spectra (corresponding to a sampling frequency of 320 Hz) for the four floors (including the base) under a broadband base excitation when the bumper was engaged between the top two floors. The spectra only appear to show evidence of three 'modes' as a result of the base excitation, so the SADE algorithm was applied here as if the system had 3DOF and the responses of the floors above 'ground' were taken as the DOFs.

When SADE was applied with a linear transformation (corresponding to the POD/PCA), the results were as shown in Fig. 11; although the second 'mode' is separated out cleanly in  $u_2$ , the other transformed DOFs are clearly not pure. The situation is much improved when a cubic transformation is used with SADE, as illustrated in Fig. 12; each of the transformed DOFs is clearly dominated by one of the modes with only small artefacts showing evidence of remaining interaction. In the case of the experimental data set, the transformation with the lowest cost produced a much less clear extraction of single peaks, so it is the best SADE solution selected 'by eye' which has been presented here.

The different case studies here have all shown that the proposed decomposition shows promise and does support interpretation as a 'modal' decomposition. Before discussing the strengths and weaknesses revealed by this study, it is necessary to consider the other important desirable ingredient of a modal decomposition; i.e. *superposition*.

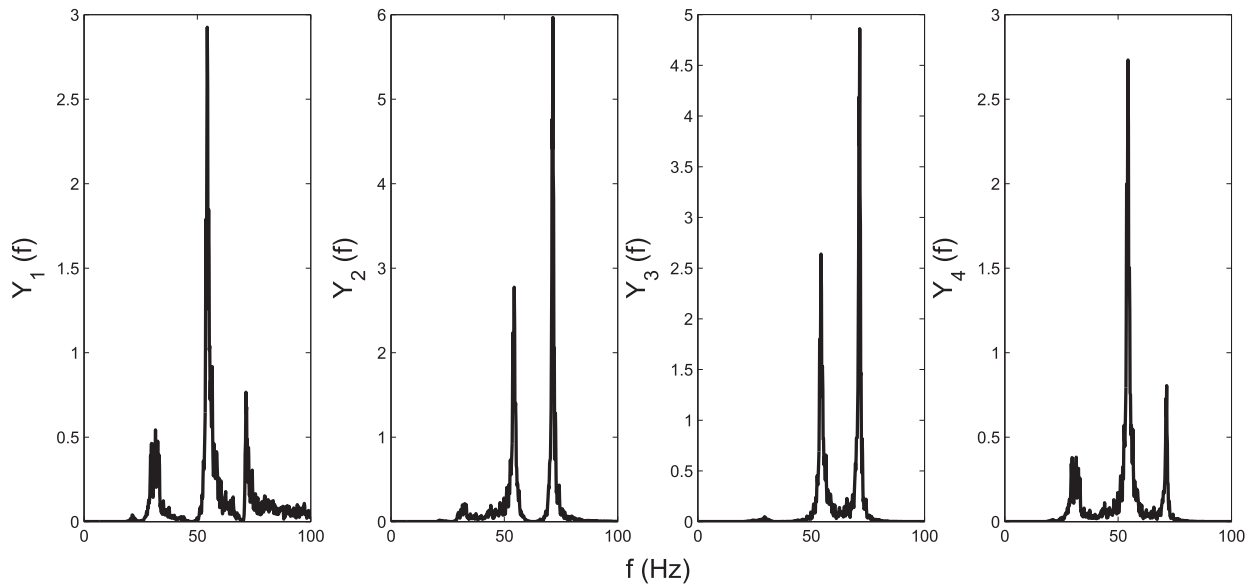


Fig. 10. Bookshelf case study, PSDs for physical variables.

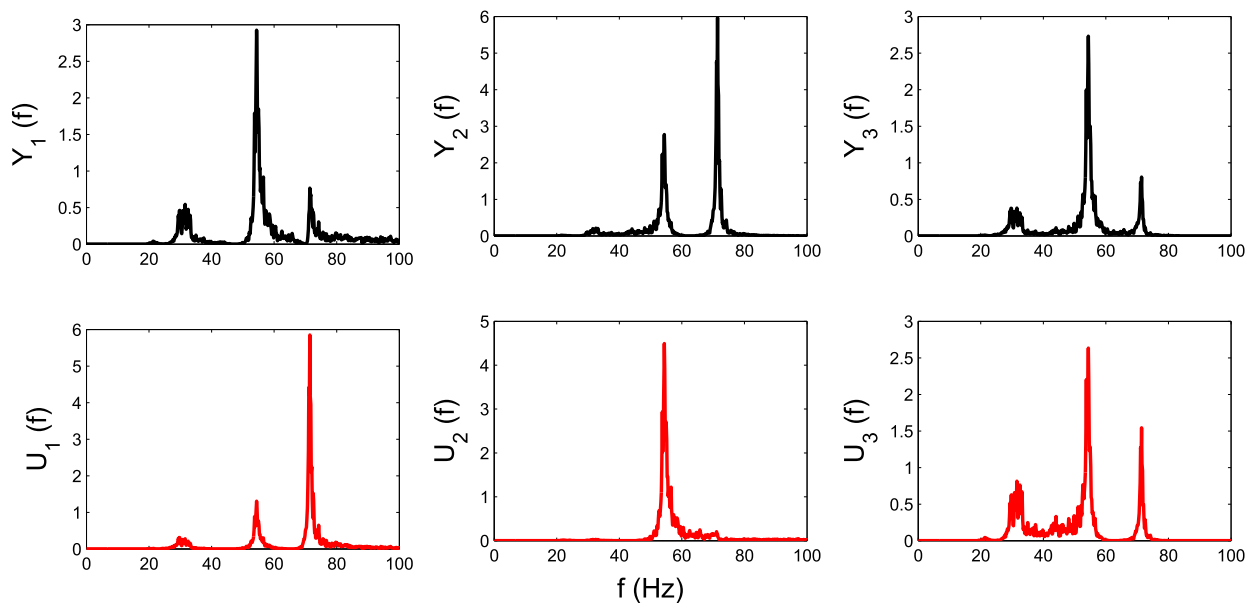
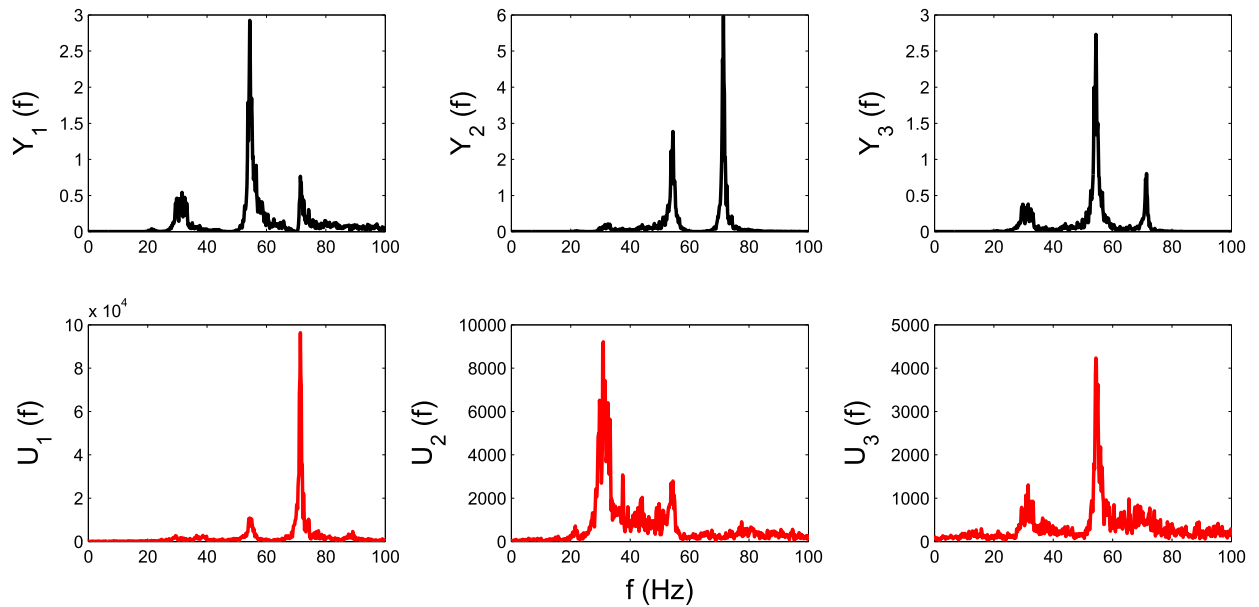


Fig. 11. Bookshelf case study, PSDs for physical and transformed variables: SADE with linear transformation.

## 6. Inversion of the transformation – superposition

As discussed in Section 2, linear modal analysis has arguably two key ingredients; the first is a (linear) transformation into independent ‘modal’ coordinates, the second is an inverse (linear) transformation from the modal coordinates back to physical coordinates. So far in this paper, only the first transformation has been addressed. The forward transformation was found via an optimisation process that did not explicitly specify the targets of the transformation, but rather, demanded a certain condition that those targets (the transformed variables) should satisfy (i.e. statistical independence). This mode of machine learning, without explicit targets is usually referred to as *unsupervised learning* [25]. It has long been recognised that PCA or the POD can be cast as an unsupervised learning problem; in fact, it was shown in the beautiful paper by Roweis and Ghahramani [26] that the POD belongs to a class of linear problems (both static and dynamic) that can be solved within a unifying Expectation Maximisation (EM) framework.

The interesting thing about the inverse transformation required here, is that the targets are now specified i.e. the vector  $\{y\}$  mapping to the modal  $\{u\}$  is known in each case; this means that learning the inverse transformation is a standard



**Fig. 12.** Bookshelf case study, PSDs for physical and transformed variables: SADE with cubic transformation (solution selected 'by eye' to best isolate single peaks).

(nonlinear) regression problem addressable by *supervised learning* [25]. There are any number of learning algorithms which could be employed in the current context; the one chosen here is the *Gaussian Process* (GP) algorithm [25,27]. The GP has been chosen because, apart from providing a principled Bayesian approach to regression, it very naturally provides confidence intervals for its predictions; this property will turn out to be rather interesting. Although GPs are being used more often in a structural dynamic context, they are still not commonplace, so some background theory will be provided here in order to make this paper a little more self-contained.

6.1. *Gaussian processes*

The basic premise of the Gaussian process (GP) method is to perform inference over *functions* directly, as opposed to inference over *parameters* of functions.

For simplicity, the discussion here will assume that the system of interest has a single output variable. Adapting the notation of [27], let  $[X] = [\{x\}_1, \{x\}_2 \dots \{x\}_N]^T$  denote a matrix of multivariate training inputs, and  $\{y\}$  denote the corresponding vector of training outputs. The input vector for a testing point will be denoted by the column vector  $\{x^*\}$  and the corresponding (unknown) output by  $y^*$ .

A Gaussian process prior is formed by assuming a (Gaussian) distribution over *functions*,

$$f(\{x\}) \sim \mathcal{GP}(m(\{x\}), k(\{x\}, \{x\})) \tag{28}$$

where  $m(\{x\})$  is the *mean function* and  $k(\{x\}, \{x^*\})$  is a positive-definite *covariance function*.

One of the defining properties of the GP is that the density of a finite number of outputs from the process is multivariate normal. Together with the known marginalisation properties of the Gaussian density, it is therefore possible to consider the value of this function only at the points of interest: training points and predictions. Allowing  $\{f\}$  to denote the function values at the training points  $[X]$ , and  $f^*$  to denote the predicted function value at a new point  $\{x^*\}$ , one has,

$$\begin{pmatrix} \{f\} \\ f^* \end{pmatrix} \sim \mathcal{N} \left( \{0\}, \begin{bmatrix} K([X], [X]) & K([X], \{x^*\}) \\ K(\{x^*\}, [X]) & K(\{x^*\}, \{x^*\}) \end{bmatrix} \right) \tag{29}$$

where a zero-mean prior has been used for simplicity (see [27] for a discussion), and  $K([X], [X])$  is a matrix whose  $i, j$ th element is equal to  $k(\{x\}_i, \{x\}_j)$ . Similarly,  $K([X], \{x^*\})$  is a column vector whose  $i$ th element is equal to  $k(\{x\}_i, \{x^*\})$ , and  $K(\{x^*\}, [X])$  is the transpose of the same.

In order to relate the observed target data  $\{y\}$  to the function values  $\{f\}$ , a simple Gaussian noise model can be assumed,

$$\{y\} \sim \mathcal{N}(\{f\}, \sigma_n^2 [I]) \tag{30}$$

where  $[I]$  is the identity matrix and  $\sigma_n^2$  constitutes a hyperparameter which can easily be identified by optimisation. Since one is not interested in the variable  $\{f\}$ , it can be marginalised (integrated out) from Eq. (29) [27], as the relevant integral,

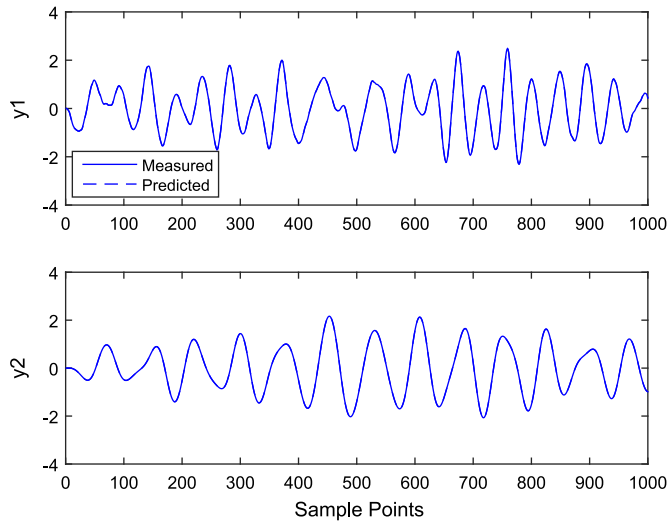


Fig. 13. Inverse modal transformation for 2DOF case study; Gaussian process reconstruction on training data.

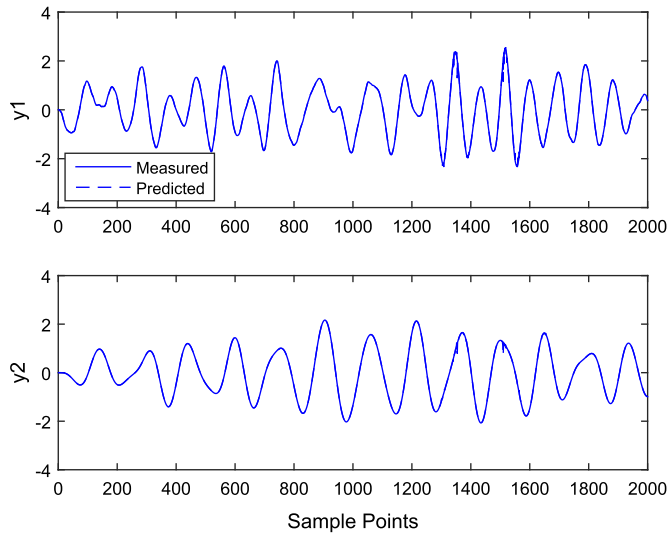


Fig. 14. Inverse modal transformation for 2DOF case study; Gaussian process predictions on training and testing data.

$$p(\{y\}) = \int p(\{y\}|\{f\})p(\{f\})d\{f\} \tag{31}$$

is over a multivariate Gaussian and is therefore analytically tractable. The result is the joint distribution for the training and testing target values,

$$\begin{pmatrix} \{y\} \\ \{y^*\} \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K([X], [X]) + \sigma_n^2[I] & K([X], \{x^*\}) \\ K(\{x^*\}, [X]) & K(\{x^*\}, \{x^*\}) + \sigma_n^2 \end{bmatrix}\right) \tag{32}$$

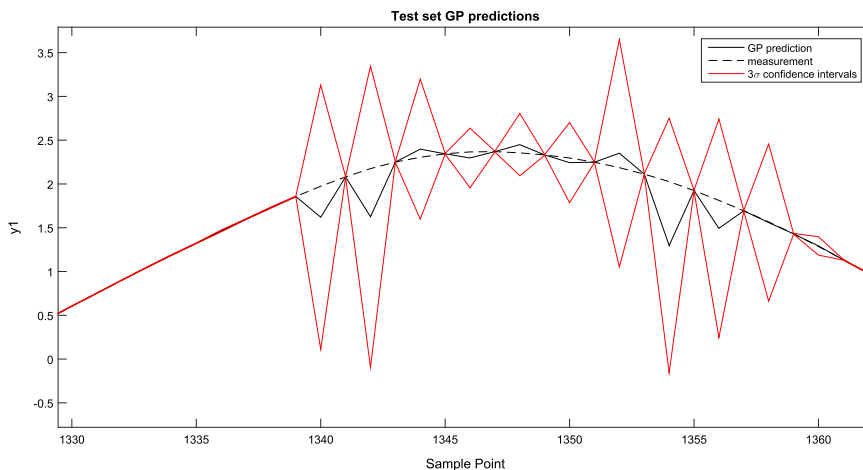
In order to make use of the above, it is necessary to re-arrange the joint distribution  $p(\{y\}, \{y^*\})$  into a conditional distribution  $p(\{y^*\}|\{y\})$ . Using standard results for the conditional properties of a Gaussian reveals [27],

$$\{y^*\} \sim \mathcal{N}(m^*(\{x^*\}), k^*(\{x^*\}, \{x^*\})) \tag{33}$$

where

$$m^*(\{x^*\}) = k(\{x^*\}, [X])[K([X], [X]) + \sigma_n^2[I]]^{-1}\{y\} \tag{34}$$

is the *posterior mean* of the GP and,



**Fig. 15.** Inverse modal transformation for 2DOF case study; Gaussian process predictions on training and testing data. Zoom on a region of poor prediction.

$$k^*({x}^*, {x}') = k({x}^*, {x}') + [X][K([X], [X]) + \sigma_n^2[I]]^{-1}K([X], {x}') - K({x}^*) \quad (35)$$

is the posterior variance.

Thus the GP model provides a posterior distribution for the unknown quantity  $y^*$ . The mean from Eq. (33) can then be used as a 'best estimate' for a regression problem, and the variance can also be used to define confidence intervals.

There does remain the question of the choice of covariance function  $k({x}, {x}')$ . In practice, it is often useful to take a squared-exponential function of the form:

$$k({x}, {x}') = \sigma_f^2 \exp\left(-\frac{1}{2l^2} \|{x} - {x}'\|^2\right) \quad (36)$$

although various other forms are possible (see [27]). Eq. (36) is the form adopted here. The covariance function involves the specification of two *hyperparameters*  $\sigma_f^2$  and  $l$ . The hyperparameters can be optimised using an evidence framework, along with the noise parameter  $\sigma_n^2$  [27]. Denoting the complete set of these parameters as  $\{t\}$ , they can be found by maximising a function,

$$f(\{t\}) = -\frac{1}{2}\{y\}^T [K([X], [X]) + \sigma_n^2[I]]\{y\} - \frac{1}{2} \log |K([X], [X]) + \sigma_n^2[I]| \quad (37)$$

which is equal to the log of the evidence, up to some constant. Since the number of hyperparameters in this case is small, the optimisation can be carried out simply by gradient descent.

## 7. The case studies revisited: superposition

The first illustration of inversion/superposition here will concern the 2DOF case study exhibited earlier. As with all machine learning algorithms, a principled application requires the definition of a training set (on which to learn the mapping) and a testing set (on which to validate the mapping). In order to give the simplest learning problem and to ensure that the GP is trained to interpolate rather than to extrapolate, the training set was taken as the odd-numbered samples from 2000 points of  $\{u, y\}$  data, and the testing set comprised the even-numbered samples.<sup>3</sup> The results from the trained GP when re-presented with the training data are shown in Fig. 13. The GP results on the training data are nearly perfect; the figure does not show the confidence bounds on the predictions as they would be indistinguishable from the prediction line.

When the GP predictions are computed on the training and testing sets, one sees something very interesting. The predictions are just as accurate as they were on the training data, except for some isolated points which show a distinct

<sup>3</sup> It is important to be clear on what is meant by the word *interpolation* here. The GP does *not* interpolate in a temporal sense i.e. the value at a testing point is not obtained by taking temporal neighbours and applying, say, a Lagrangian formula. The GP is a static map and is insensitive to the temporal nature of the training data; in fact, the training data could be randomly re-ordered and the interpolant, which is over the input-space, would not change. In the particular case here, where a squared-exponential function has been chosen for the covariance kernel, the GP is in fact equivalent to a radial-basis function interpolant [28] or RBF neural network [29].



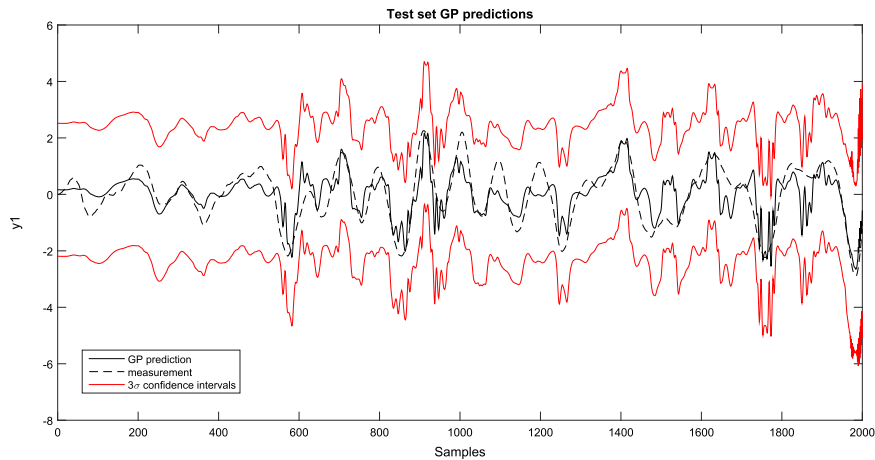


Fig. 16. Inverse modal transformation for 3DOF case study; Gaussian process predictions on training and testing data.

prediction error (Fig. 14). The explanation of the isolated errors requires a little careful thought. Fig. 15 shows a close-up of one of the regions with prediction errors; two important observations can be made. The first observation is that the prediction errors are still (of course) very small on the alternate points that were present in the training data. This shows clearly that the GP errors are not the result of extrapolation as the points with marked prediction errors are directly between training points. Another reason why one might suspect something subtle is that the set of erroneous predictions at alternate points is introducing structure into the predictions with a much shorter correlation length than elsewhere in the (largely very smooth) responses. Clearly the GP has learned a correlation length scale appropriate to the majority of the data and this raises a question as to why short-scale errors should arise. The authors believe that the explanation for the errors lies in the nature of the forward transformation. The forward transformation adopted here was a cubic multinomial in the physical responses; this means that the inverse transformation is necessarily *multi-valued*. The errors in the GP predictions are due to the fact that the GP is a smooth single-valued function and will therefore produce erroneous results if the physical data move (smoothly) between sheets of the true inverse function.

The second interesting observation on the GP inverse transformation is that the 99.7% ( $3\sigma$ ) confidence intervals expand at the erroneous predictions and bound them; this is precisely what is required of confidence intervals. This behaviour is not completely intuitive and requires explanation; however, an analytical explanation does not appear to be trivial. Work is in progress on this matter based on the fact that the factor  $K(\{x\}^*, [X])[K([X], [X]) + \sigma_n^2 [I]]^{-1}$  in the GP predictive mean is also present in the predictive variance, and it thus seems plausible that the factor driving the predictions away from their true values might also help to inflate the confidence bounds at those points. A ‘hand waving’ argument regarding the behaviour of a GP attempting to predict a multi-valued function runs as follows. If a number of training sample pairs have the same  $\{x\}$  input, but distinct outputs  $y$ , the GP can only assume that the  $y$  values are sampled from a single Gaussian; even if the training data are from a noise-free function, the only way the GP can generate different outputs is to inflate the output variance at the point  $\{x\}$ .

The second illustration of learning the inverse is for the 3DOF simulated system in the case studies. In this case, the inversion appears to be much more demanding as a result of the less successful forward transformation. Only the results for the response  $y_1$  will be presented. As before, the GP was trained on alternate (odd-numbered) points from 2000 and tested on all 2000. The GP predictions on the testing set with 99.7% confidence intervals are shown in Fig. 16. Although further study is required, one supposes that the lower accuracy is a result of the fact that truly independent  $u_i$  coordinates were not obtained in the forward transformation and that some information was lost. As always, it is encouraging to see that the confidence intervals do bound the true values.

There is clear degradation in passing from the inversion results for the 2DOF system to the those from the 3DOF system; this will be discussed in the following section. Similar degradation is visible in the results for the bookshelf data; this was not a surprise as it gave similar results on the forward transformation to those from the 3DOF simulated system.

## 8. Discussion and conclusions

The current paper proposes a new approach to nonlinear modal analysis based on a generalisation of the Principal Orthogonal Decomposition (POD) and also motivated by the Shaw-Pierre approach to nonlinear modal analysis. The approach is based on optimising a nonlinear transformation from the physical coordinates to a frame in which the coordinate variables are statistically independent. Independence is used here as the measure of ‘normality’ or orthogonality of the derived ‘modes’. Because independence is central to the approach taken here, it currently only makes sense for random data; however, that issue bears further investigation. An advantage of the approach taken here is that complicated algebraic

analysis is not needed. In fact, even the details of the equations of motion are not needed; this makes the method particularly suited to experimental investigation of nonlinear systems; in principle, only measured responses are needed. Of course, the dependence on data can bring a disadvantage, which is that the transformation learned may only be applicable for the level of excitation experienced by the structure when training data were collected. This is an issue with any machine learning method – the problem of *generalisation*. The solution (in the context of machine learning) is to measure training data which truly capture the operational range of the structure of interest; failing this, one should regard the models obtained as only *effective* models, to be used in similar circumstances to those experienced in training. Of course, generalisation can be an issue with analytical methods also; if an approximate form for the transformation is computed, like the polynomial form used in the original Shaw–Pierre paper on NNMs, then the mapping may also be input-dependent. The issue of generalisation is also going to affect the computation of the inverse modal transformation in the data-based approach proposed. The results presented here are quite preliminary, they arguably raise as many issues as they settle and much remains to be done before a robust and principled methodology can emerge.

The foremost issue to be dealt with concerns the objective function used for optimisation. The main idea of the approach here is to secure statistical independence of the transformed variables or modes and this cannot be accomplished by considering a finite number of correlation statistics. However, other statistics which do measure total independence are available. The most direct assertion of independence for a set of variables  $u_i$  is that their joint density function  $p(u_1, \dots, u_n)$  factors into a product of the marginal densities,

$$p(u_1, \dots, u_n) = p(u_1) \dots p(u_n) \quad (38)$$

where,

$$p(u_1) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(u_1, \dots, u_n) du_2 \dots du_n \quad (39)$$

etc.

Setting aside the difficulty of estimating high-dimensional densities [30,31], one needs to convert the condition (38) into an objective function. This can be accomplished by defining a metric on the space of densities, one of the most well-known quantities of this nature is the Kullback–Leibler (KL) divergence<sup>4</sup> [32,33],

$$D_{KL}[p(x) \| q(x)] = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx \quad (40)$$

(with the obvious generalisation to multivariate densities). The KL divergence is positive semi-definite, and zero only when  $p(x) = q(x)$ . Now, *Total correlation* can be defined as the Kullback–Leibler distance from the joint pdf of a set of variables to the product of the marginal pdfs,

$$C(U_1, \dots, U_n) = D_{KL}[p(u_1, \dots, u_n) \| p(u_1) \dots p(u_n)] \quad (41)$$

This function is only zero if all the  $u_i$  are mutually independent and can therefore serve as an objective function for measuring independence. The quantity can also be understood in terms of entropy [32,33],

$$C(U_1, \dots, U_n) = \left[ \sum_{i=1}^n H(U_i) \right] - H(U_1, \dots, U_n) \quad (42)$$

where  $H$  is the usual (Shannon) entropy,

$$H(x) = - \int_{-\infty}^{\infty} p(x) \log p(x) dx \quad (43)$$

(with the obvious generalisation to multivariate densities). The interpretation of Eq. (42) is straightforward; entropy represents expected information content, the equation says that there is no more information in the joint density than in the sum over the marginal densities – another independence criterion. Unfortunately, although the total correlation is an excellent candidate for the objective function required here, it can be very time-consuming to compute.

Other candidates for the objective function exist; one possibility was motivated by the fact that general correlation measures can be overly sensitive to linear relationships between variables e.g. the Pearson Coefficient. *Distance Correlation* was introduced by Gabor Székely in 2005 [34] as a more refined tool for measuring nonlinear correlations. The definition is lengthy and the reader is referred to the original reference, but like total correlation, distance correlation is only zero for independent variates. Unfortunately, the issue for optimisation is that, like total correlation, distance correlation is much more expensive to compute than the covariance matrix. Fast computation of total independence measures is one of the more important areas of further work concerning the current paper.

These matters concerning the objective function are important because of the issue raised concerning the case studies.

<sup>4</sup> There is an abuse of terminology here; the KL divergence is not truly a metric as it is not symmetric i.e.  $D_{KL}[p(x) \| q(x)] \neq D_{KL}[q(x) \| p(x)]$  and does not satisfy the triangle inequality; it is in fact a *premetric*.

The issue is that those transformations which visibly conform best to our preconceptions of a modal decomposition (clear, isolated, single resonance peaks) are not always those with the lowest objective function values. The current hypothesis of the authors is that this is simply because the objective function used here does not encode true statistical independence.

Another important issue raised in the paper is the nature of the ‘modal’ transformation; here a cubic multinomial has been adopted. This will almost certainly represent a truncation of the ‘true’ transformation; however, increasing the polynomial order of the transformation causes the number of parameters for optimisation to grow very fast, quickly making the problem intractable from a computational viewpoint. A polynomial transformation was also adopted here, because it arguably allows a ‘physical’ interpretation of the coefficients e.g. they could, in principle, be compared with those from the analytical Shaw–Pierre approach. However, there may be advantages in adopting a nonparametric ‘black-box’ form for the transformations. In fact, as observed above, the content of the approach presented here represents a variant of independent component analysis (ICA). Powerful nonparametric approaches to ICA – like kernel-ICA [35] – have been developed recently, and their application to nonlinear modal analysis will be presented in a sister paper to the current one, a (very) preliminary study can be found in [36]. One of the possibilities of adopting the nonparametric approach is that a transformation with a single-valued inverse could be learned, avoiding the issue which arose in this analysis. It is important to mention here that an attempt at nonlinear modal analysis using machine learning has already been presented in [37]. That paper was motivated by the fact that the linear POD is entirely equivalent to PCA, so that a nonlinear ‘modal’ analysis could be implemented by using a nonlinear variant of PCA. Nonlinear PCA was implemented in [37] by using an auto-associative neural network (AANN); this gives a form of unsupervised learning which was identified as a form of PCA in [38]. A five-layer AANN was used in which the central ‘bottleneck’ layer gave a reduced-dimension representation of the measured (physical) variables. The authors of [37] distinguished between ‘modal’ and ‘non-modal’ transformations depending on a single or multiple AANNs respectively. What distinguishes the approach in [37] from the one proposed here is that the variables from the bottleneck layer of an AANN are not proven to be statistically independent (analysis of AANNs is somewhat intractable) and thus are not necessarily ‘modes’ in the sense proposed here. That said, the method proposed in [37] is a perfectly acceptable approach to dimension reduction for nonlinear systems and has been successfully applied in the context of model reduction [39]. Furthermore, the AANN approach in [37] learns the forward and inverse transformations simultaneously in a non-parametric manner and thus neatly avoids any issues associated with multi-valued functions.

It is important to consider other basic restrictions built into the transformations represented here. The first is that the transformations are functions of displacement variables only. This assumption is made here on the basis that the case studies presented were given proportional damping so that a real linear transformation on the displacements was known to suffice in the linear case. Proportional damping was thus assumed in order to try and control the complexity of the optimisation problem; although velocities do not appear to have been needed here, they will almost certainly need to be considered for a general nonlinear non-proportionally damped system. Finally, the examples presented here have been chosen to exclude the possibility of internal resonances; if such resonances are present, the transformations would have to allow the possibility of mapping into higher dimensional manifolds representing more than one ‘modal’ variable at a time. As in the case of non-proportional damping, internal resonances do not represent an obstruction to the validity of the method proposed here, they will simply increase the dimension of the optimisation problem.

Setting aside the issue of whether the approach presented here can be accepted as true nonlinear modal analysis, it is almost certainly a useful tool for model reduction and system identification. The transformed variables  $u_i$  presented in the case studies here clearly resemble SDOF systems in terms of their spectral response. One would therefore hope that identification of the processes  $\{x\} \rightarrow u_i$  would be easier than direct identification of the process  $\{x\} \rightarrow \{y\}$ ; the former identification would then allow the latter via superposition. The approach presented here allows for direct identification of the generative processes  $\{x\} \rightarrow u_i$  via standard methods of nonlinear system identification or via machine learning. Given the likely complexity of the equations of motion for the processes  $\{x\} \rightarrow u_i$  a nonparametric approach like Gaussian process NARX models may be indicated [40].

As a final observation, the paper [26] presented a unified Bayesian framework for linear systems which included the POD as a special case. Work is currently underway to see if the framework can be extended to nonlinear systems and thus provide an alternative approach to nonlinear modal analysis.

## Acknowledgements

The support of the UK Engineering and Physical Sciences Research Council (EPSRC) through grant reference numbers EP/J016942/1 and EP/K003836/2 is gratefully acknowledged. The authors would also like to thank Dr Elizabeth Cross of the University of Sheffield for providing help with Gaussian process software and for her critique of the original manuscript. Finally, the authors are grateful to the members of the Engineering Institute of Los Alamos National Laboratory in the US for providing access to the bookshelf data.

## References

- [1] D.J. Ewins, *Modal Testing: Theory, Practice and Application*, Wiley-Blackwell, 2000.
- [2] A.F. Vakakis, Non-linear normal modes (NNMs) and their applications in vibration theory, *Mech. Syst. Signal Process.* 11 (1997) 3–22.
- [3] Y. Mikhlin, Nonlinear normal modes for vibrating mechanical systems. Review of theoretical developments, *Appl. Mech. Rev.* 63 (2010).
- [4] K. Worden, G.R. Tomlinson, Nonlinearity in Structural Dynamics: Detection, Modelling and Identification, Institute of Physics Press, 2001.
- [5] K. Worden, G.R. Tomlinson, Nonlinearity in experimental modal analysis, *Philos. Trans. R. Soc. Ser. A* 359 (2007) 113–130.
- [6] J. Murdock, *Normal Form Theory and Unfoldings for Local Dynamical Systems*, Springer, 2002.
- [7] R. Rosenberg, The normal modes of nonlinear n-degree-of-freedom systems, *J. Appl. Mech.* 29 (1962) 7–14.
- [8] S. Shaw, C. Pierre, Normal modes for non-linear vibratory systems, *J. Sound Vib.* 164 (1993) 85–124.
- [9] G. Kerschen, M. Peeters, J.-C. Golinval, A. Vakakis, Nonlinear normal modes, part I: A useful framework for the structural dynamicist, *Mech. Syst. Signal Process.* 23 (2009) 170–194.
- [10] M. Peeters, R. Vigiú, G. Sérandour, G. Kerschen, J.-C. Golinval, Nonlinear normal modes, part II: toward a practical computation using numerical continuation techniques, *Mech. Syst. Signal Process.* 23 (2009) 195–216.
- [11] J.-P. Noël, L. Renson, C. Grappasonni, G. Kerschen, Identification of nonlinear normal modes of engineering structures under broadband forcing, *Mech. Syst. Signal Process.* XX (2015) YYY–ZZZ.
- [12] D. Laxalde, F. Thouvez, Complex non-linear modal analysis for mechanical systems: application to turbomachinery bladings with friction, *J. Sound Vib.* 322 (2009) 1009–1025.
- [13] M. Krack, Nonlinear modal analysis of nonconservative systems: extension of the periodic motion concept, *Comput. Struct.* 154 (2015) 59–71.
- [14] E. Pesheck, C. Pierre, S. Shaw, A new Galerkin-based approach for accurate non-linear normal modes through invariant manifolds, *J. Sound Vib.* 249 (2002) 971–993.
- [15] B. Feeny, R. Kappagantu, On the physical interpretation of proper orthogonal modes in vibrations, *J. Sound Vib.* 211 (1998) 607–616.
- [16] S. Sharma, *Applied Multivariate Techniques*, John Wiley and Sons, 1995.
- [17] S. Roberts, S. Everson, *Independent Component Analysis: Principles and Practice*, Cambridge University Press, 2001.
- [18] N. Dervilis, K. Worden, D. Wagg, S. Neild, Simplifying transformations for nonlinear systems: part I, an optimisation-based variant of normal form analysis, in: *Proceedings of the 33rd International Modal Analysis Conference*, Orlando, FL, 2015.
- [19] N. Dervilis, K. Worden, D. Wagg, S. Neild, Simplifying transformations for nonlinear systems: part II, statistical analysis of harmonic cancellation, in: *Proceedings of the 33rd International Modal Analysis Conference*, Orlando, FL, 2015.
- [20] G. Kerschen, F. Poncelet, J.-C. Golinval, Physical interpretation of independent component analysis in structural dynamics, *Mech. Syst. Signal Process.* 21 (2007) 1561–1575.
- [21] F. Poncelet, G. Kerschen, J.-C. Golinval, D. Verhelst, Output-only modal analysis using blind source separation techniques, *Mech. Syst. Signal Process.* 21 (2007) 2335–2358.
- [22] K. Worden, G. Manson, On the identification of hysteretic systems. part I: fitness landscapes and evolutionary identification, *Mech. Syst. Signal Process.* 29 (2012) 201–212.
- [23] *The Matlab Primer*, The Mathworks, 2015.
- [24] E. Figueiredo, E. Flynn, Three-Story Building Structure to Detect Nonlinear Effects, Tech. rep., Los Alamos National Laboratories, 2009.
- [25] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2013.
- [26] S. Roweis, Z. Ghahramani, A unifying review of linear gaussian models, *Neural Comput.* 1 (1999) 305–345.
- [27] C. Rasmussen, C. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, 2006.
- [28] M. Powell, *Radial Basis Functions for Multivariable Interpolation*, Tech. rep., DAMPT 1985/NA12, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, 1985.
- [29] D. Broomhead, D. Lowe, Multivariable functional interpolation and adaptive networks, *Complex Syst.* 2 (1988) 325–355.
- [30] B. Silverman, *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability, Chapman and Hall, 1986.
- [31] D. Scott, *Multivariate Density Estimation: Theory, Practice and Visualization*, John Wiley and Sons, 1992.
- [32] S. Kullback, R. Leibler, On information and sufficiency, *Ann. Math. Stat.* 22 (1951) 79–86.
- [33] S. Kullback, *Information Theory and Statistics*, John Wiley and Sons, 1959.
- [34] G. Székely, M. Rizzo, N. Bakarov, Measuring and testing dependence by correlation of distances, *Ann. Stat.* 35 (2007) 2769–2794.
- [35] F. Bach, M. Jordan, Kernel independent component analysis, *J. Mach. Learn. Res.* 3 (2002) 1–48.
- [36] N. Dervilis, D. Wagg, P. Green, K. Worden, Nonlinear modal analysis using pattern recognition, in: *Proceedings of 26th International Conference on Noise & Vibration Engineering*, Leuven, 2014.
- [37] G. Kerschen, J.-C. Golinval, Feature extraction using auto-associative neural networks, *Smart Mater. Struct.* 13 (2004) 211–219.
- [38] M. Kramer, Nonlinear principal component analysis using auto-associative neural networks, *AIChE J.* 37 (1991) 233–243.
- [39] G. Kerschen, J.-C. Golinval, A model updating strategy of non-linear vibrating structures, *Int. J. Numer. Methods Eng.* 60 (2004) 2147–2164.
- [40] K. Worden, G. Manson, E. Cross, Higher-order frequency response functions from Gaussian process NARX models, in: *Proceedings of 25th International Conference on Noise & Vibration Engineering*, Leuven, 2012.