Mechanical Systems and Signal Processing 205 (2023) 110858

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Sparse Bayesian machine learning for the interpretable identification of nonlinear structural dynamics: Towards the experimental data-driven discovery of a quasi zero stiffness device

Tanmoy Chatterjee [a,b,*], Alexander D. Shaw [b], Michael I. Friswell [b], Hamed Haddad Khodaparast [b]

[a] School of Mechanical Engineering Sciences, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford GU2 7XH, United Kingdom

[b] Faculty of Science and Engineering, Swansea University, Bay Campus, Swansea SA1 8EN, Wales, United Kingdom

ARTICLE INFO

Communicated by J.E. Mottershead

Keywords:
Sparse Bayesian
Interpretable
Machine learning
Nonlinear structural dynamics
Relevance vector machines

ABSTRACT

Data-driven discovery of governing laws for complex nonlinear structural dynamic systems remains a challenging issue of paramount importance. This work addresses the above issue by leveraging the available noisy data and integrating sparse Bayesian machine learning (ML) techniques to discover the governing equations. The problem of discovery is re-cast as the automatic relevance determination of models (model selection) from the library of potential candidate basis terms and their coefficients are determined (parameter identification) using sparse Bayesian linear regression. Two sparsity promoting ML algorithms based on relevance vector machines have been employed. Both these approaches use Bayesian statistics and quantify the uncertainty associated with the model predictions. Results from four representative numerical examples of nonlinear structural dynamics illustrate excellent performance of both proposed approaches. The results have been validated with the true governing equations and time response data. Comparison has also been made with a recent and popular sparse discovery approach. Finally, the proposed framework is applied to real datasets that were generated from an in-house designed experimental setup of a quasi zero stiffness device and good performance has been observed.

1. Introduction

The rapid advancement of computational prowess and technology for data collection have led to significant interest in discovering physics from available data, in order to develop robustly validated models [1]. Moreover, misrepresentation of exact physics of the system due to elusive laws and unknown model form have accelerated the need to discover the governing equations. Determining the underlying equations of motion is of prime importance as they provide the mathematical representation of the system characteristics and hence, useful insights can be derived about the physical phenomena. The equation discovery (ED) problem involves the solution of two sub-problems, identification of the functional form of the equations of motion (i.e., *model selection*) and determination of the unknown parameters of the selected model (i.e., *parameter identification*). The ED problem is also commonly referred to as system identification and their extensive review in dynamical systems can be found in [2–5].

* Corresponding author at: School of Mechanical Engineering Sciences, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford GU2 7XH, United Kingdom.
E-mail address: t.chatterjee@surrey.ac.uk (T. Chatterjee).

https://doi.org/10.1016/j.ymssp.2023.110858
Received 11 June 2023; Received in revised form 28 August 2023; Accepted 8 October 2023
Available online 14 October 2023

Conventional approaches primarily rely on hypothesizing a set of relevant models based on the domain knowledge and expert judgement. The optimal model is selected based on the trade-off between the model accuracy and complexity determined by information-theoretic conditions, for example, Akaike and Bayesian information criteria and minimum description length [6]. However, the effectiveness of these classical techniques may often be restricted in cases of limited prior information and complex functional form of the model, both of which are common scenarios for practical applications. Due to the expeditious progress of data-driven modelling and simulation, recent model selection approaches have depended less on human intuition and domain expertise and more heavily on data. Deep neural networks have been utilized for the identification of non-linear dynamic systems [7–10]. However, the deep learning approaches were employed as black-box models, limiting their use only for forecasting in unseen environments and thus, ensuring *generalizability*. In addition to the desirable predictive ability, the goal of an ED problem is also to determine the governing physics as a function of a small or only relevant number of terms, resulting in better understanding of the physical effects and manifestations and thus, ensuring *interpretability/explainability*. For an excellent discussion on the above foundational principles, the recent review by Kutz and Brunton [5] is recommended.

The seminal work on data-driven ED can be traced back to [11,12], where symbolic regression was employed to search across a dictionary of candidate basis terms. Despite the success of the method, its reliance on genetic programming led to an imbalance between *generalizability* and *interpretability*. A cost-effective computational alternative of the symbolic regression problem was proposed recently by leveraging from sparse regression and compressive sensing [13]. This approach is abbreviated as *SINDy*, which stands for Sparse Identification of Nonlinear Dynamics and its open-source implementation can be found in [14]. Due to SINDy's lucid framework and effectiveness, it has gained wide popularity and has been implemented in variety of applications. They include, but are not limited to, biology [15], chemical reactions [16], plasma dynamics [17], fluid dynamics [18–20], multi-scale systems [21], predictive control [22], structural identification with hysteretic behaviour [23], aerodynamics [24], nonlinear energy sinks [25], sparse selection using integral terms [26], ED using short impulse time response data [27], identification of partial differential equations [28], robust learning of noisy and limited data [29] and stochastic dynamics [30,31], among others. Lately, SINDy has been integrated with deep learning for further enhancement [32–35].

To ensure *interpretability* and promoting sparsity, classical penalization or thresholding approaches through a constrained optimization over the standard least square cost function to identify only the relevant features are deployed [36]. Examples include ridge regression, least absolute shrinkage and selection operator (LASSO), least angle regression, elastic net, matching pursuit, and sequential thresholded least squares (STLS) among other compressive sensing techniques. *SINDy* employs the STLS algorithm due to its ability to scale well in high dimensions. However, the above methods are deterministic in nature and hence, do not quantify the model predictive uncertainty. Bayesian inference [37,38] can be a natural choice to overcome this issue. In addition to yielding the posterior distribution, a Bayesian approach has the following advantages over the deterministic methods, (i) the prior distributions seamlessly apply penalization of the parameters, and, (ii) the penalty parameter can be estimated without additional steps, for example, cross-validation.

Therefore, the proposed approach in this work employs sparse Bayesian learning using *Relevance Vector Machines* (RVM) for the data-driven discovery of nonlinear structural dynamic systems. The RVM was originally developed as a Bayesian treatment of Support Vector Machines for kernel-based learning [39] and followed the principle of ℓ_1 regression for promoting sparsity. The RVM is based on a Student's-t prior distribution which is sharply peaked around zero. A survey of different sparsity promoting priors can be found in [40]. In general, the literature concerning the sparse Bayesian frameworks for ED is found to be scarce and only a handful of very recent investigations are available. In this regard, the spike-and-slab [41], the Student's-t prior [42] and the Gaussian process [43] were employed for the ED of nonlinear structural dynamic systems. The schematic of the proposed framework is presented in Fig. 1. The general problem setup and dictionary/library construction is discussed in Section 2 and the mathematical formulation of sparse Bayesian learning using RVM is discussed in Section 3. Although this work demonstrates two variants of RVM, one of the approaches has an overlap with the above recent work [42] and therefore, the authors' do not claim novelty on the methodology in itself. The real strength of this work lies in the physical insights gained from the numerical examples solved (Section 4) and practical application (Section 5). To further elucidate, the significance of the present work, particularly in relation to the solved problems, is outlined as follows:

- In contrast to considering only single degree of freedom (two state) systems, coupled systems involving the third state parameter have been considered, for example, representation of the rate-independent hysteretic phenomena using Bouc–Wen oscillator. Additionally, as a case study, we assumed that the third state variable was unmeasured, simulating a specific real-life scenario. Thus, the identification is performed on an equivalent 2D system (refer Section 4.3).
- The numerical examples solved include the Valanis model, one of the powerful models for capturing the complex dynamics of nonlinear frictional contact interfaces such as a bolted lap joint. It also consists of the third state as mentioned above and is represented by the nonlinear restoring (frictional) force. Four separate models of the restoring force of varying complexities have been considered, including simulating the switching behaviour (refer Section 4.4.4).
- Besides considering complex nonlinear simulated examples, a unique and highlighting feature of this work is the sparse Bayesian data-driven discovery of a Quasi Zero Stiffness (QZS) device based on the real experimental data. The data has been collected from a recent in-house designed experimental rig (refer Section 5). The ED of QZS devices is new, challenging and useful considering their wide-spread popularity as vibration absorbers and nonlinear energy sinks.

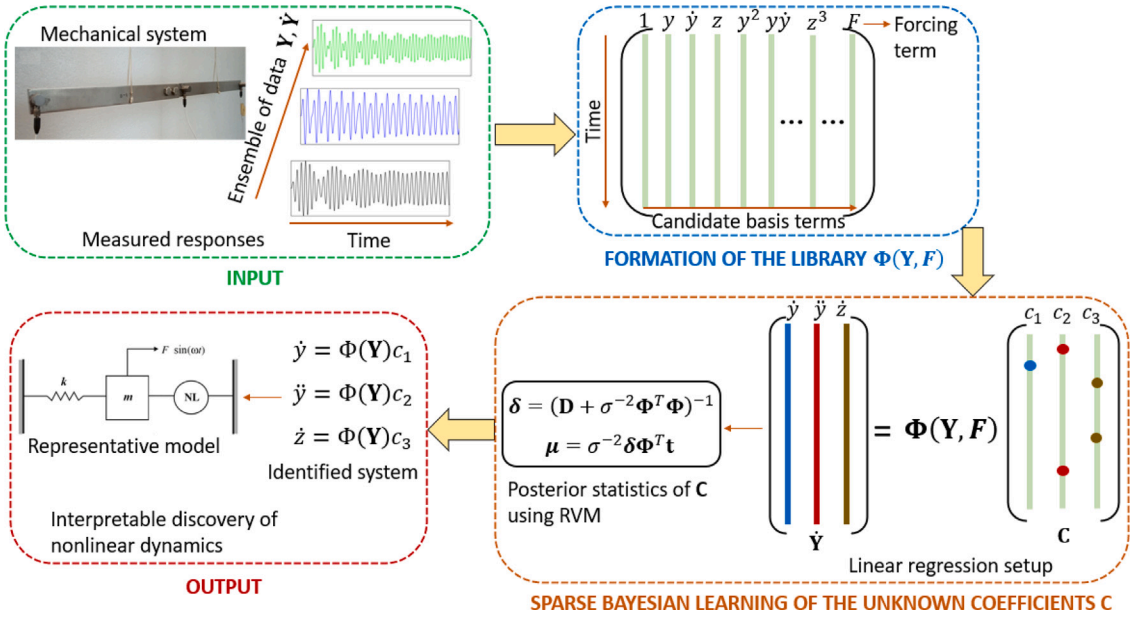


Fig. 1. A schematic flow diagram of the proposed sparse Bayesian framework for the discovery of nonlinear structural dynamic systems. The figure highlights the computational framework comprising of four blocks (each indicated by the dotted rectangles) and inter-connected with arrows. The first block takes in noisy measured time response data as input which is then passed on to the second block for constructing the library of potential candidate bases and the forcing term. The third block performs the sparse Bayesian regression via the Relevance Vector Machines and yields the posterior statistics of the unknown coefficients. Finally, the fourth block uses the unknown coefficients (from the previous block) to obtain the identified set of differential equations, thus leading to the discovery of the governing physics and yielding a representative nonlinear model.

2. Sparse discovery of nonlinear structural dynamics

The discretized mathematical model of a forced continuous dynamical system can be expressed by the following state space representation,

$$\dot{\mathbf{Y}} = g(\mathbf{Y}, \mathbf{F}) \qquad \text{where, } \mathbf{Y} = \begin{bmatrix} \mathbf{y} \\ \dot{\mathbf{y}} \end{bmatrix} \tag{1}$$

where $\mathbf{Y}$ is the state space vector comprising the displacement vector $\mathbf{y}$ and velocity vector $\dot{\mathbf{y}}$, $g(\cdot)$ denotes the generalized functional form of the system dynamics and $\mathbf{F}$ represents the external force vector acting on the structure.

Generally speaking, the sparse discovery of nonlinear structural dynamics aims to discover the governing equations (i.e., determine the function g) based on the state space vector $\mathbf{Y}$ and its first derivative $\dot{\mathbf{Y}}$ comprising the measured (or approximated) time domain data sets $\mathbf{y}(t)$, $\dot{\mathbf{y}}(t)$ and $\ddot{\mathbf{y}}(t)$. To do so, a library $\boldsymbol{\Phi}(\mathbf{Y})$ consisting of the candidate basis functions is constructed. The underlying idea of constructing these bases is to include potential models and nonlinearity information which represent the measurement data in hand. Therefore, for a system with unknown system dynamics g (which is the case in most scenarios), a large pool of candidate terms are included to ensure that the nonlinear data patterns are adequately captured. Typically, $\boldsymbol{\Phi}(\mathbf{Y})$ consists of constant, linear, polynomial, other nonlinear functions (such as, trigonometric, modulus function, etc.) of the state vector and forcing information, for instance, as shown in Eq. (2).

$$\boldsymbol{\Phi}(\mathbf{Y}) = \begin{bmatrix} | & | & | & | & | & | & & | \\ \mathbf{1} & \mathbf{Y} & \mathbf{Y}^{P_2} & \mathbf{Y}^{P_3} & \sin \mathbf{Y} & |\mathbf{Y}| & \dots & \mathbf{F} \\ | & | & | & | & | & | & & | \end{bmatrix} \tag{2}$$

where the constant term $\mathbf{1}$ is included to consider bias in the regression approach. The sampled time data are stored in the following arrangements,

$$\mathbf{y} = \overset{\text{time}}{\updownarrow} \overbrace{\begin{pmatrix} y_1(t_1) & y_2(t_1) & \cdots & y_s(t_1) \\ y_1(t_2) & y_2(t_2) & \cdots & y_s(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ y_1(t_m) & y_2(t_m) & \cdots & y_s(t_m) \end{pmatrix}}^{\text{state}} \quad \dot{\mathbf{y}} = \overset{\text{time}}{\updownarrow} \overbrace{\begin{pmatrix} \dot{y}_1(t_1) & \dot{y}_2(t_1) & \cdots & \dot{y}_s(t_1) \\ \dot{y}_1(t_2) & \dot{y}_2(t_2) & \cdots & \dot{y}_s(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \dot{y}_1(t_m) & \dot{y}_2(t_m) & \cdots & \dot{y}_s(t_m) \end{pmatrix}}^{\text{state}} \tag{3}$$

Using Eq. (3), the linear term $\mathbf{Y}$ can be obtained as,

$$\mathbf{Y} = \begin{bmatrix} y_1(t_1) & y_2(t_1) & \cdots & y_s(t_1) & \dot{y}_1(t_1) & \dot{y}_2(t_1) & \cdots & \dot{y}_s(t_1) \\ y_1(t_2) & y_2(t_2) & \cdots & y_s(t_2) & \dot{y}_1(t_2) & \dot{y}_2(t_2) & \cdots & \dot{y}_s(t_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ y_1(t_m) & y_2(t_m) & \cdots & y_s(t_m) & \dot{y}_1(t_m) & \dot{y}_2(t_m) & \cdots & \dot{y}_s(t_m) \end{bmatrix} \tag{4}$$

$\mathbf{Y}^{P_n}$ in Eq. (2) denotes the nth order polynomial nonlinearity. For example, the quadratic nonlinear terms represented by $\mathbf{Y}^{P_2}$ can be arranged as,

$$\mathbf{Y}^{P_2} = \begin{bmatrix} {y_1}^2(t_1) & y_1(t_1)y_2(t_1) & \cdots & {y_s}^2(t_1) & \dot{y}_1^2(t_1) & \dot{y}_1(t_1)\dot{y}_2(t_1) & \cdots & \dot{y}_s^2(t_1) \\ {y_1}^2(t_2) & y_1(t_2)y_2(t_2) & \cdots & {y_s}^2(t_2) & \dot{y}_1^2(t_2) & \dot{y}_1(t_2)\dot{y}_2(t_2) & \cdots & \dot{y}_s^2(t_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ {y_1}^2(t_m) & y_1(t_m)y_2(t_m) & \cdots & {y_s}^2(t_m) & \dot{y}_1^2(t_m) & \dot{y}_1(t_m)\dot{y}_2(t_m) & \cdots & \dot{y}_s^2(t_m) \end{bmatrix} \tag{5}$$

Thus, the library matrix $\mathbf{\Phi}(\mathbf{Y})$ is constructed from the measured or approximated data sets and each of the columns represents a candidate basis term for mapping the system dynamics (as represented by function g in Eq. (1)). Subsequently, the equation discovery problem is solved by transforming it to a linear regression framework as,

$$\dot{\mathbf{Y}} = \mathbf{\Phi}(\mathbf{Y})\mathbf{C} \tag{6}$$

where the motive is to determine the unknown coefficients $\mathbf{C}$. As previously mentioned, a large number of the candidate features may have to be included in $\mathbf{\Phi}$ due to the unknown nature of the system dynamics. Therefore, to prevent the potential ill-conditioning and ensure the uniqueness of solution, sparse regression by compressive sensing can be leveraged. The underlying assumption is that the nonlinear dynamics can be satisfactorily captured by a handful of basis terms in the library matrix $\mathbf{\Phi}$ so that a sparse set of unknown coefficients (whose majority entries are zero) can be recovered, which has been observed to be true for most physical systems. This is also known as the automatic relevance determination and leads to a low-dimensional or parsimonious representation of the governing dynamics.

The unique sparse solution of $\hat{\mathbf{C}} = [\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_n]$ can be obtained by LASSO and is expressed as follows, for the individual (i^{th}) regression problem,

$$\hat{c}_i = \underset{c_i}{\operatorname{argmin}} \|\dot{\mathbf{Y}}_i - \mathbf{\Phi}(\mathbf{Y})c_i\|_2^2 + \lambda_i \|c_i\|_1 \tag{7}$$

where $\|\cdot\|_2$ and $\|\cdot\|_1$ denote the ℓ_2-norm and ℓ_1-norm, respectively. $\dot{\mathbf{Y}}_i$ represents the ith column of matrix $\dot{\mathbf{Y}}$ and, λ_i is the regularization factor. Note that Eq. (7) corresponds to conventional least squares for the case $\lambda_i = 0$ and $\hat{\mathbf{C}}$ becomes sparser with the increase of λ_i. Hence, with the proper selection of λ_i, over-fitting can be avoided when the number of features in the basis matrix is high, by the trade-off between model accuracy and complexity. Standard information-theoretic criteria are employed for this purpose [6]. Upon solving the optimization problem in Eq. (7), the coefficient matrix $\hat{\mathbf{C}}$ is determined which results in the sparse recovery of the underlying governing equation. This illustrates the overall framework of SINDy. SINDy employs the sequential thresholding least squares approach to solve the LASSO problem in Eq. (7) for computational gains. In the next section, two sparse Bayesian machine learning techniques have been proposed (in contrast to the deterministic thresholding approach described in this section) to solve the data-driven identification problem.

3. Proposed sparse Bayesian discovery of nonlinear structural dynamics

The machine learning technique known as relevance vector machines (RVM) [39], employing a probabilistic sparse kernel, is used to impart the sparse Bayesian feature for the discovery of nonlinear dynamics. In the following sub-sections, sparse Bayesian learning via RVM is summarized along with a step-by-step pseudo-code.

3.1. Model specification

For notational convenience, Eq. (6) can be re-stated as,

$$\mathbf{t} = \mathbf{\Phi}\mathbf{C} + \epsilon \tag{8}$$

where the measurements $\dot{\mathbf{Y}}$ on the left side of Eq. (6) is now represented as the target data $\mathbf{t}$, $\mathbf{\Phi}$ is of dimensions $(m \times b)$ and ϵ denotes the error/noise. The crucial feature in the RVM formulation is the prior distribution of the coefficients which is given as

$$p(\mathbf{C}|\boldsymbol{\alpha}) = (2\pi)^{-b/2} \prod_{i=1}^{b} {\alpha_i}^{1/2} \exp\left(-\frac{\alpha_i {c_i}^2}{2}\right) \tag{9}$$

where $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \ldots, \alpha_b\}$ is a vector of b hyperparameters, which yield essential information on the relevance of the basis terms. It is the prior which is responsible for rendering sparsity to the model. Detailed description of the estimation procedure of hyperparameters $\boldsymbol{\alpha}$, such that the prior distribution $p(\mathbf{C}|\boldsymbol{\alpha})$ is peaked around zero, can be found in [39]. A multivariate Gaussian

likelihood function is considered and is given by

$$p(\mathbf{t}|\mathbf{C},\sigma^2) = (2\pi)^{-m/2}\sigma^{-m}\exp(-\frac{\|\mathbf{t}-\mathbf{\Phi C}\|^2}{2\sigma^2}) \tag{10}$$

3.2. Posterior inference

Subsequently, the posterior distribution over the coefficients can be obtained as the product of the prior in Eq. (9) and likelihood in Eq. (10) according to Bayes theorem as

$$p(\mathbf{C}|\mathbf{t},\boldsymbol{\alpha},\sigma^2) = \frac{p(\mathbf{t}|\mathbf{C},\sigma^2)\ p(\mathbf{C}|\boldsymbol{\alpha})}{p(\mathbf{t}|\boldsymbol{\alpha},\sigma^2)} \tag{11}$$

This results in a Gaussian posterior $\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\delta})$ with the mean and variance are given as, respectively,

$$\boldsymbol{\mu} = \sigma^{-2}\boldsymbol{\delta}\mathbf{\Phi}^T\mathbf{t} \tag{12}$$

$$\boldsymbol{\delta} = \left(\mathbf{D}+\sigma^{-2}\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1} \tag{13}$$

where $\mathbf{D} = \mathrm{diag}(\alpha_1,\ldots,\alpha_b)$. It is worth noting that Eqs. (12) and (13) reflect the classic setting of a Bayesian framework where the prior of the parameters is updated based on the measured/observed data, which in this case is contained in $\mathbf{t}$ and $\mathbf{D}$.

3.3. Maximization of the marginal likelihood

The marginal likelihood can be obtained by integrating out the unknown coefficients $\mathbf{C}$ as,

$$p(\mathbf{t}|\boldsymbol{\alpha},\sigma^2) = \int p(\mathbf{t}|\mathbf{C},\sigma^2)p(\mathbf{C}|\boldsymbol{\alpha})d\mathbf{C} \tag{14}$$

The log marginal likelihood can be obtained from the analytical solution of the above integration as,

$$\log p(\mathbf{t}|\boldsymbol{\alpha},\sigma^2) = -\frac{1}{2}\left[m\log(2\pi)+\log|\mathbf{\Omega}|+\mathbf{t}^T\mathbf{\Omega}^{-1}\mathbf{t}\right] \tag{15}$$

where $\mathbf{\Omega} = \sigma^2\mathbf{I}+\mathbf{\Phi}\mathbf{D}^{-1}\mathbf{\Phi}^T$. Alternatively, Eq. (15) can be expressed in terms of the posterior statistics $\boldsymbol{\mu},\boldsymbol{\delta}$ (Eqs. (12) and (13)) as,

$$\log p(\mathbf{t}|\boldsymbol{\alpha},\sigma^2) = -\frac{1}{2}\left[m\log(2\pi\sigma^2)+\left(\sigma^{-2}\mathbf{t}^T\mathbf{t}-\boldsymbol{\mu}^T\boldsymbol{\delta}^{-1}\boldsymbol{\mu}\right)+\log|\boldsymbol{\delta}|-\sum_{i=1}^{b}\log\alpha_i\right] \tag{16}$$

The objective is to maximize the marginal likelihood, as represented by Eq. (15) or its logarithmic equivalent in Eq. (16) with respect to the hyperparameters $\boldsymbol{\alpha}$ and σ^2. This is known as the type-II maximum likelihood approach and is used to achieve the sparse Bayesian learning. This essentially involves the optimization of the hyperparameters $\boldsymbol{\alpha}$ which controls the sparsity and estimation of the signal noise σ^2 and while updating the posterior statistics $\boldsymbol{\mu},\boldsymbol{\delta}$ in parallel. This is an iterative process until convergence where the hyperparameters $\boldsymbol{\alpha}$ and σ^2 are set to their initial values, the posterior statistics $\boldsymbol{\mu},\boldsymbol{\delta}$ are computed using Eqs. (12) and (13) and subsequently the new estimates for $\boldsymbol{\alpha}$ and σ^2 which maximize the marginal likelihood, are obtained. For a detailed breakdown of the iterative re-estimation procedure, the readers can refer to [39].

However, as pointed out in [44], the aforementioned updating scheme experiences delayed convergence due to the following reasons, (i) the Bayesian learning depends on heuristic re-estimation of the hyperparameters and is not a mathematically rigorous iterative approach, (ii) the guarantee to attain a local maxima of the marginal likelihood is not well discussed and (iii) the algorithm starts with all b basis terms and trims the library matrix $\mathbf{\Phi}$ iteratively, as a result, the initial iterations are computationally expensive. This approach has been used for numerical illustration in this work and will be referred to as slow-RVM (S-RVM) from now onwards.

To address these drawbacks, an accelerated expectation maximization (EM) algorithm was proposed in [44]. A distinctive feature of this algorithm is that the basis terms can be added and deleted sequentially so as to increase the marginal likelihood. This more efficient approach has also been used for numerical illustration and will be referred to as fast-RVM (F-RVM) from now onwards. Specifically, the efficiency is realized by the expression of $\mathbf{\Omega}$ employed in the marginal likelihood function in Eq. (15). As illustrated previously, $\mathbf{\Omega}$ is expressed as $\sigma^2\mathbf{I}+\mathbf{\Phi}\mathbf{D}^{-1}\mathbf{\Phi}^T$ in [39], while in [44], $\mathbf{\Omega}$ is re-written to obtain the explicit contribution of α_i as,

$$\begin{aligned}\mathbf{\Omega} &= \sigma^2\mathbf{I}+\sum_{j\neq i}\alpha_j^{-1}\boldsymbol{\phi}_j\boldsymbol{\phi}_j^T+\alpha_i^{-1}\boldsymbol{\phi}_i\boldsymbol{\phi}_i^T \\ &= \mathbf{\Omega}_{-i}+\alpha_i^{-1}\boldsymbol{\phi}_i\boldsymbol{\phi}_i^T\end{aligned} \tag{17}$$

where $\mathbf{\Omega}_{-i}$ represents $\mathbf{\Omega}$ without the contribution of the ith basis vector. Utilizing matrix determinant and inverse identities and Eq. (17), the determinant and the inverse of $\mathbf{\Omega}$ can be computed, respectively, as

$$|\mathbf{\Omega}| = |\mathbf{\Omega}_{-i}||1+\alpha_i^{-1}\boldsymbol{\phi}_i^T\mathbf{\Omega}_{-i}^{-1}\boldsymbol{\phi}_i| \tag{18}$$

$$\mathbf{\Omega}^{-1} = \mathbf{\Omega}_{-i}^{-1}-\frac{\mathbf{\Omega}_{-i}^{-1}\boldsymbol{\phi}_i\boldsymbol{\phi}_i^T\mathbf{\Omega}_{-i}^{-1}}{\alpha_i+\boldsymbol{\phi}_i^T\mathbf{\Omega}_{-i}^{-1}\boldsymbol{\phi}_i} \tag{19}$$

Eq. (15) can be re-written by substituting Eqs. (18) and (19) as,

$$\log p(\mathbf{t}|\boldsymbol{\alpha},\sigma^2) = -\frac{1}{2}\left[m\log(2\pi) + \log|\boldsymbol{\Omega}_{-i}| + \mathbf{t}^T\boldsymbol{\Omega}_{-i}^{-1}\mathbf{t} - \log\alpha_i + \log(\alpha_i + \boldsymbol{\phi}_i^T\boldsymbol{\Omega}_{-i}^{-1}\boldsymbol{\phi}_i) - \frac{(\boldsymbol{\phi}_i^T\boldsymbol{\Omega}_{-i}^{-1}\mathbf{t})^2}{\alpha_i + \boldsymbol{\phi}_i^T\boldsymbol{\Omega}_{-i}^{-1}\boldsymbol{\phi}_i}\right] \tag{20}$$

$$\log p(\mathbf{t}|\boldsymbol{\alpha},\sigma^2) = \log p(\mathbf{t}|\alpha_{-i},\sigma^2) + \underbrace{\frac{1}{2}\left[\log\alpha_i - \log(\alpha_i + k_i) + \frac{l_i^2}{\alpha_i + k_i}\right]}_{f(\alpha_i)} \tag{21}$$

where $\log p(\mathbf{t}|\alpha_{-i},\sigma^2)$ is the log marginal likelihood without the contribution of the ith basis vector $\boldsymbol{\phi}_i$. The quantities k_i and l_i are,

$$k_i = \boldsymbol{\phi}_i^T\boldsymbol{\Omega}_{-i}^{-1}\boldsymbol{\phi}_i \tag{22}$$

$$l_i = \boldsymbol{\phi}_i^T\boldsymbol{\Omega}_{-i}^{-1}\mathbf{t} \tag{23}$$

Instead of performing the inversion of $\boldsymbol{\Omega}_{-i}^{-1}$ in Eqs. (22) and (23) and save the associated cost, k_i' and l_i' can be determined by using the Woodbury matrix identity as,

$$\begin{aligned} k_i' &= \boldsymbol{\phi}_i^T\boldsymbol{\Omega}^{-1}\boldsymbol{\phi}_i \\ &= \boldsymbol{\phi}_i^T\mathbf{H}\boldsymbol{\phi}_i - \boldsymbol{\phi}_i^T\mathbf{H}\boldsymbol{\Phi}\delta\boldsymbol{\Phi}^T\mathbf{H}\boldsymbol{\phi}_i \end{aligned} \tag{24}$$

$$\begin{aligned} l_i' &= \boldsymbol{\phi}_i^T\boldsymbol{\Omega}^{-1}\mathbf{t} \\ &= \boldsymbol{\phi}_i^T\mathbf{H}\mathbf{t} - \boldsymbol{\phi}_i^T\mathbf{H}\boldsymbol{\Phi}\delta\boldsymbol{\Phi}^T\mathbf{H}\mathbf{t} \end{aligned} \tag{25}$$

and subsequently using Eqs. (24) and (25), k_i and l_i can be computed as follows:

$$k_i = \frac{\alpha_i k_i'}{\alpha_i - k_i'} \tag{26}$$

$$l_i = \frac{\alpha_i l_i'}{\alpha_i - k_i'} \tag{27}$$

where $\mathbf{H} = \sigma^{-2}\mathbf{I}$. It is to be noted that when $\alpha_i = \infty$, $k_i = k_i'$ and $l_i = l_i'$.

The stationary points of the marginal likelihood can be obtained when the derivative

$$\frac{df(\alpha_i)}{\alpha_i} = \frac{\alpha_i^{-1}k_i^2 - (l_i^2 - k_i)}{2(\alpha_i + k_i)^2} \tag{28}$$

is equated to zero. Since α_i is an inverse variance and is always positive, the following two solution case scenarios hold,

$$\alpha_i = \frac{k_i^2}{l_i^2 - k_i}; \quad \text{if } l_i^2 > k_i \tag{29}$$

$$\alpha_i = \infty; \quad \text{if } l_i^2 \le k_i \tag{30}$$

Note that Eqs. (29) and (30) provide an explicit closed-form solution for α_i, depending on whether a particular basis function is included or not and the given hyperparameter values [45]. From Eqs. (29) and (30), it is implied that

- If $\boldsymbol{\phi}_i$ is in the model (i.e., $\alpha_i < \infty$) and $l_i^2 \le k_i$, then $\boldsymbol{\phi}_i$ is to be deleted (which means α_i is set as ∞).
- If $\boldsymbol{\phi}_i$ is not included in the model (i.e., $\alpha_i = \infty$) and $l_i^2 > k_i$, then $\boldsymbol{\phi}_i$ is to be added (which means α_i is set to some optimal finite value).
- If $\boldsymbol{\phi}_i$ is in the model (i.e., $\alpha_i < \infty$) and $l_i^2 > k_i$, then α_i is to be updated using Eq. (29).

Thus, an efficient algorithm can be realized by carrying out these operations in a sequential manner for every basis vector. A detailed pseudo-code for the sequential EM algorithm employed in RVM has been presented in Algorithm 1 to enhance understanding and promote reproducibility.

3.4. Convergence

In this sub-section, the convergence criteria for the EM algorithm is briefly discussed. The EM algorithm aiming to maximize the marginal likelihood commences with an empty basis vector and selects the particular basis vector $\boldsymbol{\phi}_i$ which corresponds to

the maximum change in the marginal likelihood $\mathcal{L}(\boldsymbol{\alpha}, \sigma^2)$ ($\mathcal{L}(\boldsymbol{\alpha}, \sigma^2)$ denotes $\log p(\mathbf{t}|\boldsymbol{\alpha}, \sigma^2)$ of Eq. (15) and is used for notational convenience) within each iteration. For computational efficiency of the EM algorithm, the basis matrix $\boldsymbol{\Phi}$ and the posterior statistics ($\boldsymbol{\delta}$ and $\boldsymbol{\mu}$) consist of only b' (where $b' \leq b$) basis functions which are currently included in the model (i.e., $\boldsymbol{\phi}_i$ corresponding to $\alpha_i < \infty$). Similarly, the diagonal matrix $\mathbf{D}$ contains b' entries of the hyperparameter α_i (i.e., α_i corresponding to $\alpha_i < \infty$). From Eq. (15), the change in the marginal likelihood can be represented as,

$$\begin{aligned} 2\Delta\mathcal{L} &= 2(\mathcal{L}(\widetilde{\boldsymbol{\alpha}}, \sigma^2) - \mathcal{L}(\boldsymbol{\alpha}, \sigma^2)) \\ &= \log(\frac{\boldsymbol{\Omega}}{\widetilde{\boldsymbol{\Omega}}} + \mathbf{t}^{\mathrm{T}}(\boldsymbol{\Omega}^{-1} - \widetilde{\boldsymbol{\Omega}}^{-1})\mathbf{t}) \end{aligned} \tag{31}$$

where $\widetilde{\boldsymbol{\alpha}}$ and $\widetilde{\boldsymbol{\Omega}}$ denote the new updates of $\boldsymbol{\alpha}$ and $\boldsymbol{\Omega}$, respectively. Note that the resulting Eq. (31) will vary depending upon α_i being added, deleted or re-estimated and is illustrated as follows:

- ***Addition:***

$$\begin{aligned} \boldsymbol{\Omega} = \boldsymbol{\Omega}_{-i} \text{ and } \widetilde{\boldsymbol{\Omega}} = \boldsymbol{\Omega}_{-i} + \widetilde{\alpha}_i^{-1}\boldsymbol{\phi}_i\boldsymbol{\phi}_i^{\mathrm{T}}\ ; \\ 2\Delta\mathcal{L}_i = \frac{l_i'^2 - k_i'}{k_i'} + \log\frac{k_i'}{l_i'} \end{aligned} \tag{32}$$

- ***Deletion:***

$$\begin{aligned} \boldsymbol{\Omega} = \boldsymbol{\Omega}_{-i} + \alpha_i^{-1}\boldsymbol{\phi}_i\boldsymbol{\phi}_i^{\mathrm{T}} \text{ and } \widetilde{\boldsymbol{\Omega}} = \boldsymbol{\Omega}_{-i}\ ; \\ 2\Delta\mathcal{L}_i = \frac{l_i'^2}{k_i' - \alpha_i} - \log(1 - \frac{k_i'}{\alpha_i}) \end{aligned} \tag{33}$$

- ***Re-estimation:***

$$\begin{aligned} \boldsymbol{\Omega} = \boldsymbol{\Omega}_{-i} + \alpha_i^{-1}\boldsymbol{\phi}_i\boldsymbol{\phi}_i^{\mathrm{T}} \text{ and } \widetilde{\boldsymbol{\Omega}} = \boldsymbol{\Omega}_{-i} + \widetilde{\alpha}_i^{-1}\boldsymbol{\phi}_i\boldsymbol{\phi}_i^{\mathrm{T}}\ ; \\ 2\Delta\mathcal{L}_i = \frac{l_i'^2}{k_i' + (\widetilde{\alpha}_i^{-1} - \alpha_i^{-1})^{-1}} - \log(1 + k_i'(\widetilde{\alpha}_i^{-1} - \alpha_i^{-1})) \end{aligned} \tag{34}$$

The iterations will continue until either the maximum number of iterations have been reached or the estimates have converged, as illustrated in step 5 of Algorithm 1. To ensure a local maximum of the marginal likelihood has been attained, the convergence criteria include (i) $|\Delta\log\alpha| < 10^{-6}$, where $|\Delta\log\alpha|$ denotes the logarithmic change in α_i and is defined in step 30 of Algorithm 1 and (ii) all other ${l_i}^2 - k_i \leq 0$, which essentially means that any basis vector $\boldsymbol{\phi}_i$ excluded from the model from the current iteration will not be included in the model for the next iteration [44]. The above criteria have been illustrated in steps 30 to 41 of Algorithm 1.

3.5. Prediction step

For making predictions on unseen (test) data $\mathbf{t}^*$ with the posterior model of the coefficients $\mathbf{C}$ given by Eqs. (12) and (13), the distribution $p(\mathbf{t}^*|\mathbf{t}, \boldsymbol{\alpha}, \sigma^2)$ is to be determined. The prediction model results in a multivariate Gaussian as

$$p(\mathbf{t}^*|\mathbf{t}, \boldsymbol{\alpha}, \sigma^2) = \mathcal{N}(\mathbf{t}|\boldsymbol{\mu}^*, \boldsymbol{\Omega}^*) \tag{35}$$

with mean and covariance as, respectively,

$$\boldsymbol{\mu}^* = \boldsymbol{\Phi}\boldsymbol{\mu} \tag{36}$$

$$\boldsymbol{\Omega}^* = \sigma^2\mathbf{I} + \boldsymbol{\Phi}^T\boldsymbol{\delta}\boldsymbol{\Phi} \tag{37}$$

Note that the predictive variance in Eq. (37) is the result of the sum of two terms, one arising from the estimated signal noise σ^2 and the other from $\boldsymbol{\Phi}^T\boldsymbol{\delta}\boldsymbol{\Phi}$ due to the predictive uncertainty associated with the coefficients.

3.6. Implementation

The computations in this work have been performed using MATLAB® R2023a. The MATLAB® codes of RVM implemented in this work are freely available for download from the original author's webpage (Link: <https://www.miketipping.com/sparsebayes.htm>). In addition, for this work, a more recent framework proposed in [46] was also combined with the above for the multi-output extension of RVM. The MATLAB® codes of this implementation can be found at <https://uk.mathworks.com/matlabcentral/fileexchange/49131-fast-multi-output-relevance-vector-regression>. The MATLAB® codes for SINDy used here are available for download from <https://faculty.washington.edu/kutz/page26/>.

The CPU time required by RVM and SINDy reported in the next sections are based on computations performed on a 12th Gen Intel® Core™ i9-12950HX 2.30 GHz CPU.

Algorithm 1: Pseudo-code for the efficient sequential EM algorithm in RVM

Input:

State derivative vector $\dot{\mathbf{Y}}$ and the candidate basis (library) matrix $\boldsymbol{\Phi}$; **Initialization:**

Set an initial value to $\sigma^2 \leftarrow \gamma \text{cov}(\mathbf{t}), 0 < \gamma < 1$;

Set $\alpha_i \leftarrow \infty$ for all basis vectors;

Set iteration number $iter \leftarrow 1$ and number of active basis vectors $b' \leftarrow 0$;

Iteration:

while $iter \leq iter_{max}$ *or* $converged \leftarrow No$ **do**

  /*Maximization*/:

  for $i \leftarrow 1$ *to* b **do**

    update k_i', l_i' using Eqs. (24), (25) and k_i, l_i using Eqs. (26), (27)

    if ${l_i}^2 > k_i$ **then**

      if $\alpha_i < \infty$ **then**

        $task(i) \leftarrow'$ $re-estimation'$

        $\widetilde{\alpha}_i \leftarrow \frac{{k_i}^2}{{l_i}^2 - k_i}$; ▷ From Eq. (29)

        Update $2\Delta\mathcal{L}_i$ using Eq.(34)

      else

        $task(i) \leftarrow'$ $addition'$

        Update $2\Delta\mathcal{L}_i$ using Eq.(32)

      end

    else

      if $\alpha_i < \infty$ **then**

        $task(i) \leftarrow'$ $deletion'$

        Update $2\Delta\mathcal{L}_i$ using Eq.(33)

      end

    end

    if $imag(2\Delta\mathcal{L}_i)$ **then**

      $2\Delta\mathcal{L}_i \leftarrow -\infty$; ▷ If $2\Delta\mathcal{L}_i$ is imaginary; replace with $-\infty$

    end

  end

  $[\sim, i] \leftarrow \max_i(2\Delta\mathcal{L}_i)$; ▷ i gives maximum change of marginal likelihood

  switch $task(i)$ **do**

    case $re-estimation$ **do**

      $\Delta \log \alpha \leftarrow \log \frac{\alpha_i}{\widetilde{\alpha}_i}$

      if $|\Delta \log \alpha| < tol$ **then**

        $converged \leftarrow yes$

        for $i \leftarrow 1 to b$ **do**

          if $\alpha_i \leftarrow \infty$ **then**

            if ${l_i}^2 - k_i > 0$ **then**

              $converged \leftarrow no$

              break for loop

            end

          end

        end

      end

      $\alpha_i \leftarrow \widetilde{\alpha}_i$

    end

    case $addition$ **do**

      $\alpha_i \leftarrow \frac{{k_i}^2}{{l_i}^2 - k_i}$

      $b' \leftarrow b' + 1$

    end

    case $deletion$ **do**

      $\alpha_i \leftarrow \infty$

      $b' \leftarrow b' - 1$

    end

  end

  if $iter \neq 1$ **then**

    update $\sigma^2 \leftarrow \frac{||\mathbf{t} - \boldsymbol{\Phi}\boldsymbol{\mu}||^2}{m - b + \sum_{i=1}^{b} \alpha_i \delta_{ii}}$; ▷ [1] obtained by maximizing Eq. (16) w.r.t σ^2

  end

  /*Expectation*/:

  Sequentially update $\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\Phi}$ (which consist of b' basis vectors that are currently included in the model) and $\mathbf{D}$ (consisting of b' number of hyperparameter α_i, currently included in the model).

  $iter \leftarrow iter + 1$

end

Output:

Posterior statistics: $\boldsymbol{\mu}$ and $\boldsymbol{\delta}$

[1] Setting the derivative of log marginal likelihood in Eq. (16) w.r.t σ^2 to 0, the updated expression for σ^2 [39] is obtained as $\frac{d}{d\sigma^{-2}} \log p(\mathbf{t}, \boldsymbol{\alpha}, \sigma^2) = \frac{1}{2}\left[\frac{m}{\sigma^{-2}} - \|\mathbf{t} - \boldsymbol{\Phi}\boldsymbol{\mu}\|^2 - \text{Tr}\boldsymbol{\delta}\boldsymbol{\Phi}^T\boldsymbol{\Phi}\right] = 0 \implies \sigma^{-2}_{\text{new}} = \frac{m - b + \sum_{i=1}^{b} \alpha_i \delta_{ii}}{\|\mathbf{t} - \boldsymbol{\Phi}\boldsymbol{\mu}\|^2}$ and is used in Step 54 of Algorithm 1.

Table 1
The discovered equation (i.e., the selected models and identified parameters) obtained by the accelerated and conventional relevance vector machines for the Duffing oscillator. The model predictive uncertainty has been shown in the form of the coefficient of variation within brackets below the mean coefficients. The equation discovered by SINDy has been presented for comparison. The true coefficients adopted in the actual equation is also presented for validation.

True	$\begin{bmatrix}\dot{x}\\ \ddot{x}\end{bmatrix} = \begin{bmatrix}\dot{x}\\ 100\cos\omega t - \dot{x} - 25x - 2500x^3\end{bmatrix}$
F-RVM	$\begin{bmatrix}\dot{x}\\ \ddot{x}\end{bmatrix} = \begin{bmatrix}\underset{(0.0028)}{0.9995}\dot{x}\\ \underset{(2.7\times10^{-4})}{99.8353}\cos\omega t - \underset{(0.0028)}{0.9853}\dot{x} - \underset{(0.0042)}{25.3157}x - \underset{(2.1\times10^{-4})}{2493.8}\,x^3\end{bmatrix}$
S-RVM	$\begin{bmatrix}\dot{x}\\ \ddot{x}\end{bmatrix} = \begin{bmatrix}\underset{(7.5\times10^{-4})}{0.9995}\,\dot{x}\\ \underset{(0.0017)}{99.8353}\cos\omega t - \underset{(0.0174)}{0.9853}\dot{x} - \underset{(0.0259)}{25.3157}x - \underset{(0.0013)}{2493.8}x^3\end{bmatrix}$
SINDy	$\begin{bmatrix}\dot{x}\\ \ddot{x}\end{bmatrix} = \begin{bmatrix}0.9995\dot{x}\\ 99.8381\cos\omega t - 0.9857\dot{x} - 25.3271x - 2493.8x^3\end{bmatrix}$

4. Numerical examples

4.1. Duffing oscillator

The governing equation for the Duffing oscillator is given by

$$\ddot{x} + 2\zeta\omega_n\dot{x} + \omega_n^2 x + \epsilon\omega_n^2 x^3 = F_0\cos\omega t \tag{38}$$

where, ω_n represents the undamped linear natural frequency, ζ is the linear damping ratio, ϵ controls the level of nonlinearity, F_0 and ω are the amplitude and frequency of the harmonic forcing, respectively. The initial conditions are considered as $x(t=0)=0$ and $\dot{x}(t=0)=0$. The following parameter values are adopted: $\omega_n = 5$ rad/s, $\zeta = 0.1$, $\epsilon = 100$ and $F_0 = 100$ N.

The state space form of Eq. (38) can be represented as

$$\begin{bmatrix}\dot{x}\\ \ddot{x}\end{bmatrix} = \begin{bmatrix}\dot{x}\\ F_0\cos\omega t - 2\zeta\omega_n\dot{x} - \omega_n^2 x - \epsilon\omega_n^2 x^3\end{bmatrix} \tag{39}$$

Independent simulations are performed at different excitation frequencies by varying the frequency ratio $\nu = \frac{\omega}{\omega_n}$ from 0.7 to 1.3 in steps of 0.1 to construct the dataset. White Gaussian noise with 10% variation, relative to the standard deviation of the observations has been added to the synthetically generated dataset to simulate the measurement error. Time t is varied from 0 to 25 s with step size 0.01, and thus, 2501 points are considered for each forcing frequency. The identification results are presented in Table 1 and Fig. 2.

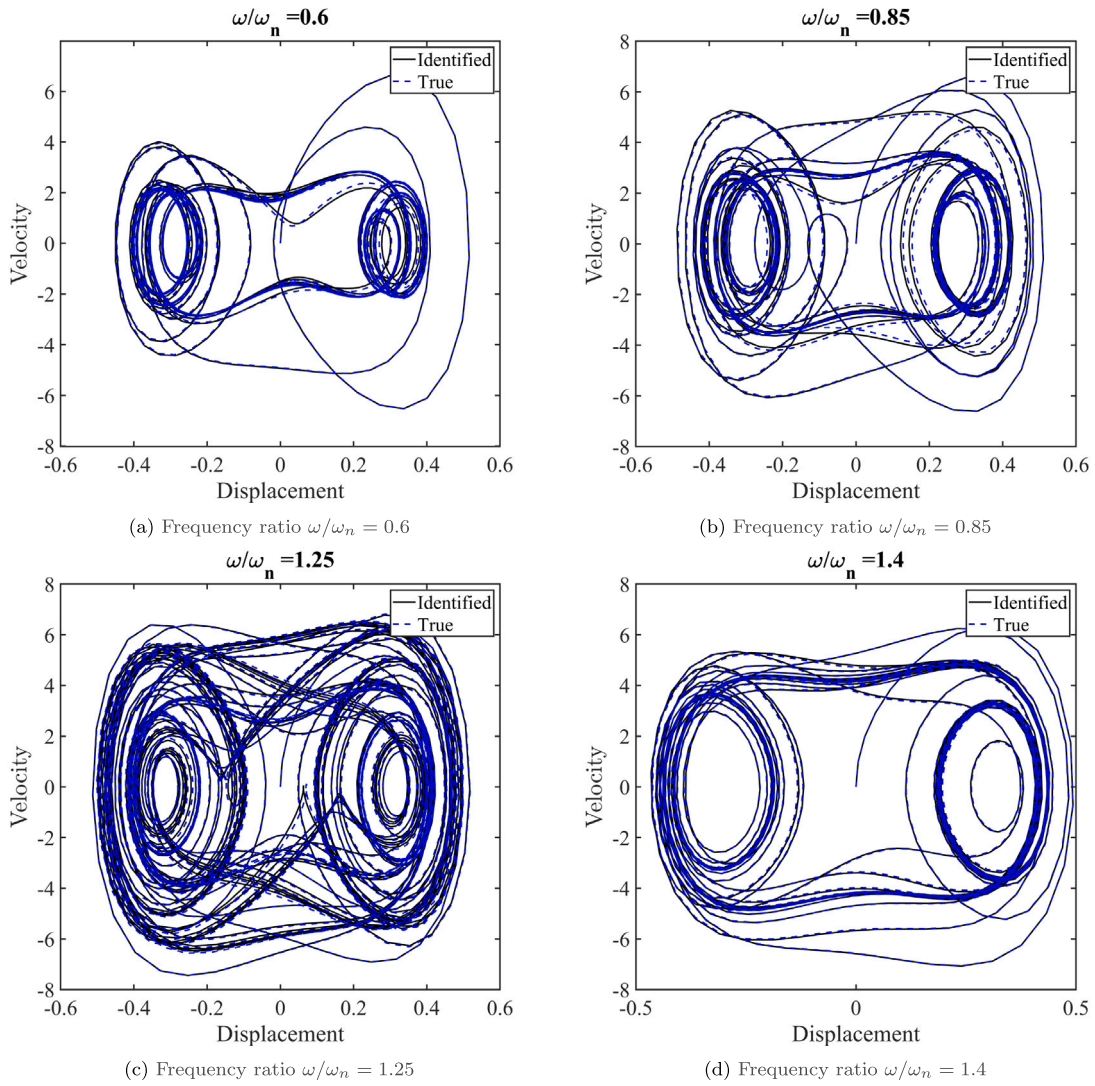
The discovered state space form of the Duffing oscillator in the form of selected models and identified parameters obtained by F-RVM, S-RVM and SINDy have been presented in Table 1. In total, 11 candidate terms were included in the basis library considering up to the third-order polynomials and the forcing term. The close proximity of the identified models with the true coefficients of the actual terms reveal excellent accuracy achieved by all of the three approaches. The sparsity promoting feature in the approaches proved to be effective with only the relevant terms present in the actual equation identified and no false discoveries were observed from Table 1. Since, the RVM approaches are Bayesian in nature, the predictive uncertainty associated with the posteriors are estimated and the coefficient of variation values are represented within parenthesis below their corresponding mean values. The CPU time required for fast RVM, slow RVM and SINDy are 26.1550 s, 212.6206 s and 0.0174 s, respectively. Fig. 2 presents the comparison of true and identified velocity and displacement time response plots obtained by F-RVM corresponding to four frequency ratios (different to that used in the training) for validation. These plots further confirm the accuracy of the equation discovery results in Table 1 and illustrate the effectiveness of the proposed approach on unseen data in mapping the complex nonlinear dynamic behaviour.

4.2. Coulomb friction damping model

The governing equation for the Coulomb friction damping model is given by

$$m\ddot{x} + c\dot{x} + kx + c_F\,\mathrm{sgn}(\dot{x}) = F_0\cos\omega t \tag{40}$$

where, m, c and k are the mass, stiffness and linear damping coefficient, respectively. sgn represents the signum function. c_F is the nonlinear frictional damping coefficient. F_0 and ω are the amplitude and frequency of the harmonic forcing, respectively. The initial conditions are considered as $x(t=0)=0$ and $\dot{x}(t=0)=0$. The following parameter values are adopted: $m = 1$ kg, $c = 2$ N s/m, $k = 1000$ N/m, $c_F = 1$ N and $F_0 = 100$ N.



(a) Frequency ratio $\omega/\omega_n = 0.6$ (b) Frequency ratio $\omega/\omega_n = 0.85$

(c) Frequency ratio $\omega/\omega_n = 1.25$ (d) Frequency ratio $\omega/\omega_n = 1.4$

Fig. 2. Comparison of the true and identified velocity (in m/s) versus displacement (in m) time response for the Duffing oscillator corresponding to frequency ratios ω/ω_n equal to (a) 0.6, (b) 0.85, (c) 1.25 and (d) 1.4 (different to that used in the model training). The identified responses were generated using the mean coefficient values obtained from the accelerated relevance vector machine approach.

The state space form of Eq. (40) can be represented as

$$\begin{bmatrix} \dot{x} \\ \ddot{x} \end{bmatrix} = \begin{bmatrix} \dot{x} \\ \frac{1}{m}\left(F_0 \cos \omega t - c\dot{x} - kx - c_F \operatorname{sgn}(\dot{x})\right) \end{bmatrix} \tag{41}$$

The variation range of the excitation frequencies, time duration and noise level have been considered to be the same as in the previous example (Section 4.1). The results of the discovered system are presented in Table 2 and Fig. 3.

In total, 8 candidate terms were included in the basis library considering up to the second order polynomials, nonlinear frictional damping and the forcing term. The sparsity promoting feature in the approaches proved to be effective with only the 4 terms present in the actual equation identified and no false discoveries were observed from Table 2. However, SINDy was found to be less accurate compared to the RVM based approaches in identifying the nonlinear frictional damping coefficient, which is critical for this example. The CPU time required for fast RVM, slow RVM and SINDy are 21.9345 s, 90.2491 s and 0.0108 s, respectively. The identified displacement and velocity time response plots for various frequency ratios (different to that used in the training) obtained by fast RVM are presented in Fig. 3.

Table 2
The discovered equation (i.e., the selected models and identified parameters) obtained by the accelerated and conventional relevance vector machines for the Coulomb friction damping model. The model predictive uncertainty has been shown in the form of the coefficient of variation within brackets below the mean coefficients. The equation discovered by SINDy has been presented for comparison. The true coefficients adopted in the actual equation is also presented for validation.

True	$\begin{bmatrix}\dot{x}\\ \ddot{x}\end{bmatrix} = \begin{bmatrix}\dot{x}\\ 100\cos\omega t - 2\dot{x} - 1000x - \mathrm{sgn}(\dot{x})\end{bmatrix}$
F-RVM	$\begin{bmatrix}\dot{x}\\ \ddot{x}\end{bmatrix} = \begin{bmatrix}\underset{(5.2\times10^{-4})}{0.9994}\,\dot{x}\\ \underset{(1.2\times10^{-4})}{100.4889}\cos\omega t - \underset{(3.5\times10^{-4})}{2.0206}\,\dot{x} - \underset{(1.6\times10^{-5})}{1000}\,x - \underset{(0.0088)}{1.0515}\,\mathrm{sgn}(\dot{x})\end{bmatrix}$
S-RVM	$\begin{bmatrix}\dot{x}\\ \ddot{x}\end{bmatrix} = \begin{bmatrix}\underset{(7.6\times10^{-4})}{0.9994}\,\dot{x}\\ \underset{(0.0055)}{100.4890}\cos\omega t - \underset{(0.0163)}{2.0206}\dot{x} - \underset{(7.7\times10^{-4})}{1000}\,x - \underset{(0.4070)}{1.0512}\,\mathrm{sgn}(\dot{x})\end{bmatrix}$
SINDy	$\begin{bmatrix}\dot{x}\\ \ddot{x}\end{bmatrix} = \begin{bmatrix}0.9994\dot{x}\\ 100.4983\cos\omega t - 2.0134\dot{x} - 1000x - 1.2209\,\mathrm{sgn}(\dot{x})\end{bmatrix}$

4.3. Bouc–Wen oscillator

The governing equation for the Bouc–Wen oscillator is given by

$$\begin{aligned} m\ddot{x} + akx + (1-a)kz &= F_0\cos\omega t \\ \dot{z} &= \dot{x} - \delta z|z|^{\alpha-1}|\dot{x}| - \gamma|z|^{\alpha}\dot{x} \end{aligned} \tag{42}$$

where, m and k are the mass and stiffness, respectively. The parameters α, δ and γ control the shape of the hysteresis loop and a controls the degree of hysteresis. F_0 and ω are the amplitude and frequency of the harmonic forcing, respectively. The initial conditions are considered as $x(t=0)=0$, $\dot{x}(t=0)=0$ and $z(t=0)=0$. The following parameter values are adopted: $m=1$ kg, $k=50$ N/m, $a=0$, $\alpha=1$, $\delta=0.5$, $\gamma=0.5$ and $F_0=100$ N. As derived in [23], the state space form of Eq. (42) can be represented as

$$\begin{bmatrix}\dot{x}\\ \ddot{x}\\ \dot{a}_s\end{bmatrix} = \begin{bmatrix}\dot{x}\\ \frac{1}{m}\left(F_0\cos\omega t - a_s\right)\\ k\dot{x} - \delta a_s|\dot{x}| - \gamma|a_s|\dot{x}\end{bmatrix} \tag{43}$$

where, $a_s = kz$. Two cases have been simulated for the identification. The first case considers all the three states (in Eq. (43)) have been measured and used in the identification. The second case considers only the first two system states (as given by the first two states). This is to obtain an equivalent model and simulate a real-life scenario where the third state (a_s) is not measured. Hence, the identification is performed on an equivalent 2D system. For both of these individual cases, independent simulations are performed at different excitation frequencies by varying the frequency ratio $\nu = \frac{\omega}{\omega_n}$ from 0.8 to 1.2 in steps of 0.1 to construct the dataset. White Gaussian noise with 10% variation, relative to the standard deviation of the observations has been added to the synthetically generated dataset to simulate the measurement error. Time duration $t = [0\ \ 25]$ s is considered with step size 0.01 s, and thus, 2501 points are considered for each forcing frequency. The results of the discovered system corresponding to the first case (all states measured) are presented in Table 3 and Fig. 4. The results of the second case are presented in Table 4 and Fig. 5.

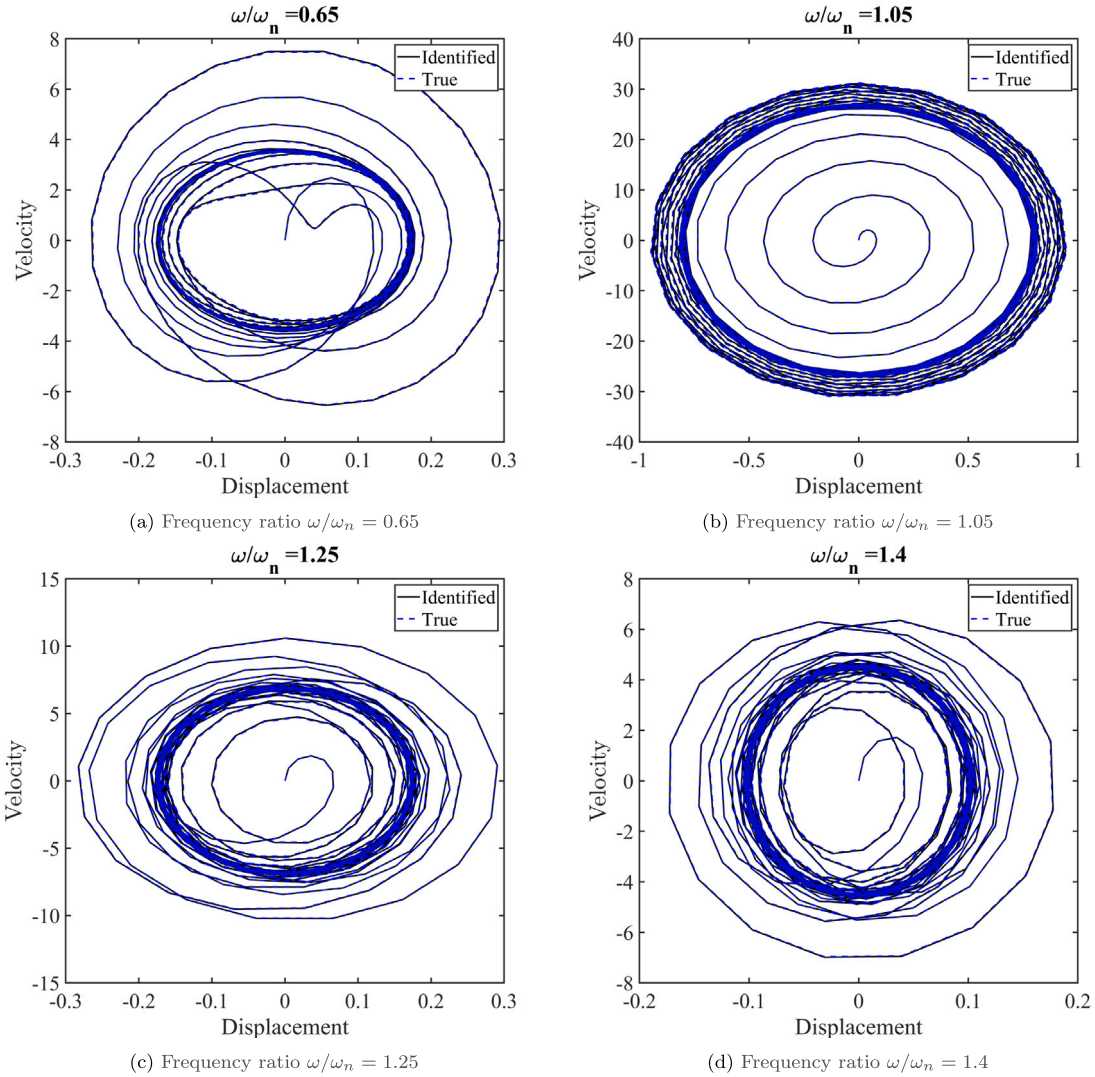
For the first and second case, 13 and 12 candidate terms were included in the basis library, respectively, considering up to the second order polynomials, hysteretic nonlinearity and the forcing term. To obtain an equivalent model for the second case in which the true model is unknown, a few combinations of hysteretic nonlinear terms were included in the library as evident from the discovered equations presented in Table 4. It can be noted from Figs. 4 and 5 that the identified time response by F-RVM is accurate on test datasets with different frequency ratios to that of the training regime. The CPU times required by fast RVM, slow RVM and SINDy for the first case, are 28.7660 s, 122.2755 s and 0.0118 s, respectively. Whereas, for the second case, the CPU times required by fast RVM, slow RVM and SINDy were 10.4534 s, 83.3005 s and 0.0078 s, respectively.

4.4. Valanis model

The governing equation for the Valanis model is given by

$$\begin{aligned} m\ddot{x} + kx + F_{\mathrm{nl}} &= F_0\sin\omega t \\ \dot{F}_{\mathrm{nl}} &= \left(E_1 + \frac{\dot{x}}{|\dot{x}|}\left(P(x,\dot{x}) - E_2F_{\mathrm{nl}}\right)\right)\dot{x} \end{aligned} \tag{44}$$

where m and k are the mass and stiffness coefficient, respectively. F_{nl} represents the nonlinear restoring force. E_1 and E_2 are constants. Four cases to simulate different hysteretic behaviour have been investigated by varying $P(x,\dot{x})$ in the following sub-sections. For details of the models and their physical interpretation, one is referred to [47]. F_0 and ω are the amplitude and frequency



(a) Frequency ratio $\omega/\omega_n = 0.65$ (b) Frequency ratio $\omega/\omega_n = 1.05$

(c) Frequency ratio $\omega/\omega_n = 1.25$ (d) Frequency ratio $\omega/\omega_n = 1.4$

Fig. 3. Comparison of the true and identified velocity (in m/s) versus displacement (in m) time response for the Coulomb friction damping model corresponding to frequency ratios ω/ω_n equal to (a) 0.65, (b) 1.05, (c) 1.25 and (d) 1.4 (different to that used in the model training). The identified responses were generated using the mean coefficient values obtained from the accelerated relevance vector machine approach.

of the harmonic forcing, respectively. The initial conditions are considered as $x(t=0)=0$ and $\dot{x}(t=0)=0$. For the identification, data from all the three system states are employed. Independent simulations are performed at different excitation frequencies by varying the frequency ratio $\nu = \frac{\omega}{\omega_n}$ from 0.7 to 1.3 in steps of 0.1 to construct the dataset. White Gaussian noise with 10% variation, relative to the standard deviation of the observations has been added to the synthetically generated dataset to simulate the measurement error. Time duration $t = [0\ \ 25]$ s is considered with step size 0.01 s, and thus, 2501 points are considered for each forcing frequency. The following parameter values are adopted: $m = 1$ kg, $k = 100$ N/m, and $F_0 = 1$ N. The above details are applicable to all the following four sub-cases considered, unless specified.

4.4.1. Case I: $P(x,\dot{x}) = \alpha x$

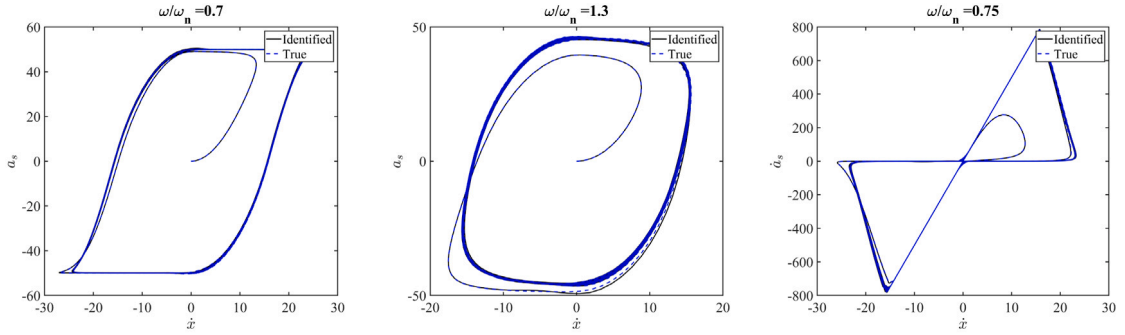
Considering $P(x,\dot{x}) = \alpha x$, the state space form of Eq. (44) can be represented as

$$\begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dot{F}_{\mathrm{nl}} \end{bmatrix} = \begin{bmatrix} \dot{x} \\ \frac{1}{m}\left(F_0 \sin\omega t - kx - F_{\mathrm{nl}}\right) \\ \left(10 + \frac{\dot{x}}{|\dot{x}|}\left(5x - 20F_{\mathrm{nl}}\right)\right)\dot{x} \end{bmatrix} \tag{45}$$

where $E_1 = 10$ N/m, $E_2 = 20$ 1/m and $\alpha = 5$ N/m^2. The results of the discovered system are presented in Table 5 and Fig. 6.

Table 3
The discovered equation (i.e., the selected models and identified parameters) obtained by the accelerated and conventional relevance vector machines for the Bouc–Wen oscillator for the case where all the three system states are considered to be measured and identified. The model predictive uncertainty has been shown in the form of the coefficient of variation within brackets below the mean coefficients. The equation discovered by SINDy has been presented for comparison. The true coefficients adopted in the actual equation is also presented for validation.

True	$\begin{bmatrix}\dot{x}\\ \ddot{x}\\ \dot{a}_s\end{bmatrix} = \begin{bmatrix}\dot{x}\\ 100\cos\omega t - a_s\\ 50\dot{x} - 0.5a_s\lvert\dot{x}\rvert - 0.5\lvert a_s\rvert\dot{x}\end{bmatrix}$
F-RVM	$\begin{bmatrix}\dot{x}\\ \ddot{x}\\ \dot{a}_s\end{bmatrix} = \begin{bmatrix}\underset{(6.4\times10^{-4})}{0.9993}\,\dot{x}\\ \underset{(1.4\times10^{-4})}{100.0320}\cos\omega t - \underset{(2.5\times10^{-4})}{0.9962}\,a_s\\ \underset{(3.8\times10^{-5})}{49.9918}\dot{x} - \underset{(8.2\times10^{-5})}{0.5015}\,a_s\lvert\dot{x}\rvert - \underset{(1.4\times10^{-4})}{0.4997}\,\lvert a_s\rvert\dot{x}\end{bmatrix}$
S-RVM	$\begin{bmatrix}\dot{x}\\ \ddot{x}\\ \dot{a}_s\end{bmatrix} = \begin{bmatrix}\underset{(8.8\times10^{-4})}{0.9993}\,\dot{x}\\ \underset{(0.0014)}{100.0320}\cos\omega t - \underset{(0.0024)}{0.9962}a_s\\ \underset{(0.0012)}{49.9918}\dot{x} - \underset{(0.0026)}{0.5015}a_s\lvert\dot{x}\rvert - \underset{(0.0043)}{0.4997}\lvert a_s\rvert\dot{x}\end{bmatrix}$
SINDy	$\begin{bmatrix}\dot{x}\\ \ddot{x}\\ \dot{a}_s\end{bmatrix} = \begin{bmatrix}0.9993\dot{x}\\ 100.0320\cos\omega t - 0.9962a_s\\ 49.9920\dot{x} - 0.5015a_s\lvert\dot{x}\rvert - 0.4997\lvert a_s\rvert\dot{x}\end{bmatrix}$



(a) Frequency ratio $\omega/\omega_n = 0.7$ (b) Frequency ratio $\omega/\omega_n = 1.3$ (c) Frequency ratio $\omega/\omega_n = 0.75$

Fig. 4. Comparison of the true and identified time response for the Bouc–Wen oscillator corresponding to values of frequency ratio ω/ω_n = 0.7, 1.3 and 0.75. Note that the third state (a_s) is assumed to be measured and the identification considers all three states as per Eq. (43). The identified responses have been generated from the mean coefficient values obtained using the accelerated relevance vector machine approach. The response quantities $\dot{x}$, a_s and, $\dot{a}_s$ depicted in the above plots are expressed in m/s, N and, N/s, respectively.

In total, 13 candidate terms were included in the basis library considering up to the second order polynomials, nonlinear restoring force and the forcing term. Good accuracy has been achieved as evident from Table 5 and Fig. 6. The CPU times required for fast RVM, slow RVM and SINDy are 61.4065 s, 189.9256 s and 0.0206 s, respectively. The identified time response behaviour for various frequency ratios (different to that used in the training dataset) obtained by fast RVM have been presented in Fig. 6.

4.4.2. Case II: $P(x,\dot{x}) = \alpha x + \beta x^2|\dot{x}|$

Considering $P(x,\dot{x}) = \alpha x + \beta x^2|\dot{x}|$, the state space form of Eq. (44) can be represented as

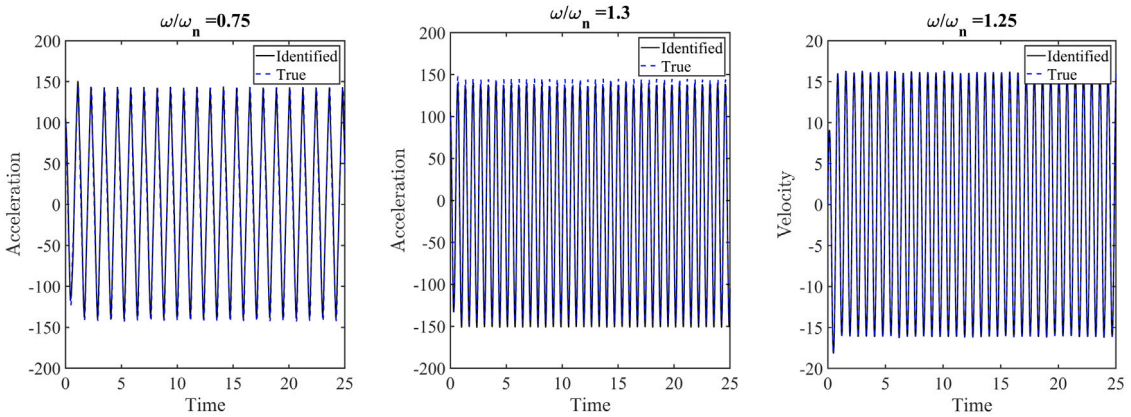
$$\begin{bmatrix}\dot{x}\\ \ddot{x}\\ \dot{F}_{\text{nl}}\end{bmatrix} = \begin{bmatrix}\dot{x}\\ \frac{1}{m}\left(F_0\sin\omega t - kx - F_{\text{nl}}\right)\\ \left(10 + \frac{\dot{x}}{|\dot{x}|}\left(30x + 1400x^2|\dot{x}| - 20F_{\text{nl}}\right)\right)\dot{x}\end{bmatrix} \tag{46}$$

where, $E_1 = 10$ N/m, $E_2 = 20$ 1/m, $\alpha = 30$ N/m^2 and $\beta = 1400$ N s/m^4. The results of the discovered system are presented in Table 6 and Fig. 7.

In total, 14 candidate terms were included in the basis library considering up to the second order polynomials, nonlinear restoring force and the forcing term. Close proximity of the identified and the true models as observed from Table 6 reveal excellent accuracy achieved by all of the three approaches. The CPU times required for fast RVM, slow RVM and SINDy are 89.7748 s, 333.4480 s and

Table 4
The discovered equation (i.e., the selected models and identified parameters) obtained by the accelerated and conventional relevance vector machines for the Bouc–Wen oscillator for the case where the third state is not considered to be measured and therefore, only the first two states are identified. The model predictive uncertainty has been shown in the form of the coefficient of variation within brackets below the mean coefficients. The equation discovered by SINDy has been presented for comparison.

F-RVM	$\begin{bmatrix}\dot{x}\\ \ddot{x}\end{bmatrix} = \begin{bmatrix}\underset{(6.4\times10^{-4})}{1.0012}\,\dot{x}\\ \underset{(0.0025)}{-24.6415} - \underset{(0.0052)}{6.4423}x - \underset{(5.9\times10^{-4})}{2.2892}\,\dot{x} - \underset{(0.0031)}{0.9961}x^2 + \underset{(0.0048)}{8.3936}\lvert x\rvert + \underset{(0.0057)}{1.1851}\lvert\dot{x}\rvert + \ldots\\ \underset{(0.0168)}{0.2205}x\lvert\dot{x}\rvert - \underset{(0.0222)}{0.1725}\lvert x\rvert\lvert\dot{x}\rvert + \underset{(3.4\times10^{-4})}{130.4215}\cos\omega t\end{bmatrix}$
S-RVM	$\begin{bmatrix}\dot{x}\\ \ddot{x}\end{bmatrix} = \begin{bmatrix}\underset{(9.1\times10^{-4})}{1.0012}\,\dot{x}\\ \underset{(0.0310)}{-24.6424} - \underset{(0.0631)}{6.4392}x - \underset{(0.0073)}{2.2892}\dot{x} - \underset{(0.0382)}{0.9961}x^2 + \underset{(0.0586)}{8.3965}\lvert x\rvert + \underset{(0.0694)}{1.1853}\lvert\dot{x}\rvert + \ldots\\ \underset{(0.2051)}{0.2200}x\lvert\dot{x}\rvert - \underset{(0.2701)}{0.1729}\lvert x\rvert\lvert\dot{x}\rvert + \underset{(0.0041)}{130.4222}\cos\omega t\end{bmatrix}$
SINDy	$\begin{bmatrix}\dot{x}\\ \ddot{x}\end{bmatrix} = \begin{bmatrix}1.0012\dot{x}\\ -21.7528 - 6.4999x - 2.3825\dot{x} - 0.9166x^2 + 7.4717\lvert x\rvert + 0.7333\lvert\dot{x}\rvert + \ldots\\ 0.2504x\lvert\dot{x}\rvert - 0.1434\lvert x\rvert\lvert\dot{x}\rvert + 131.5459\cos\omega t\end{bmatrix}$



(a) Test NMSE = 0.0117 (b) Test NMSE = 0.0060 (c) Test NMSE = 1.3580×10^{-6}

Fig. 5. Comparison of the true and identified time response for the Bouc–Wen oscillator corresponding to test datasets with frequency ratio ω/ω_n = 0.75, 1.25 and 1.3. Note that the third state (a_x) is not considered to be measured and therefore, not included in the identification to simulate a practical scenario. The identified responses have been generated from the mean coefficient values obtained using the accelerated relevance vector machine approach. The quantities time, velocity and acceleration depicted in the above plots are expressed in s, m/s and, m/s^2, respectively.

0.0175 s, respectively. The identified time response behaviour corresponding to various values of frequency ratio (different to that used in the model training) obtained by the fast RVM are presented in Fig. 7.

4.4.3. Case III: $P(x, \dot{x}) = \alpha x + \beta e^x \dot{x}$

Considering $P(x, \dot{x}) = \alpha x + \beta e^x \dot{x}$, the state space form of Eq. (44) can be represented as

$$\begin{bmatrix}\dot{x}\\ \ddot{x}\\ \dot{F}_{\mathrm{nl}}\end{bmatrix} = \begin{bmatrix}\dot{x}\\ \frac{1}{m}\left(F_0 \sin\omega t - kx - F_{\mathrm{nl}}\right)\\ \left(30 + \frac{\dot{x}}{|\dot{x}|}\left[(5x + 3e^x\dot{x}) \times 10^4 - 500F_{\mathrm{nl}}\right]\right)\dot{x}\end{bmatrix} \tag{47}$$

where, $E_1 = 30$ N/m, $E_2 = 500$ 1/m, $\alpha = 5 \times 10^4$ N/m^2 and $\beta = 3 \times 10^4$ N s/m^2. The results of the discovered system are presented in Table 7 and Fig. 8.

In total, 14 candidate terms were included in the basis library considering up to the second order polynomials, nonlinear restoring force and the forcing term. The sparsity promoting feature in the RVM based approaches proved to be effective with only identifying the relevant terms present in the actual equation and no false discoveries were observed from Table 7. SINDy was found to be ineffective in capturing the nonlinear dynamics from the noisy data with multiple false discoveries as evident from Table 7. The CPU times required for fast RVM, slow RVM and SINDy are 52.1649 s, 200.3144 s and 0.0210 s, respectively. The identified time response behaviour corresponding to various values of frequency ratio (different to that used in the model training) obtained by fast RVM are presented in Fig. 8.

Table 5
The discovered equation (i.e., the selected models and identified parameters) obtained by the accelerated and conventional relevance vector machines for the Valanis model (Case-I). The model predictive uncertainty has been shown in the form of the coefficient of variation within brackets below the mean coefficients. The equation discovered by SINDy has been presented for comparison. The true coefficients adopted in the actual equation is also presented for validation.

True	$\begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dot{F}_{\mathrm{nl}} \end{bmatrix} = \begin{bmatrix} \dot{x} \\ \sin\omega t - 100x - F_{\mathrm{nl}} \\ 10\dot{x} + 5\frac{\dot{x}^2}{\lvert\dot{x}\rvert}x - 20\frac{\dot{x}^2}{\lvert\dot{x}\rvert}F_{\mathrm{nl}} \end{bmatrix}$
F-RVM	$\begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dot{F}_{\mathrm{nl}} \end{bmatrix} = \begin{bmatrix} \underset{(0.0077)}{1.0006}\dot{x} \\ \underset{(0.0110)}{0.9983}\sin\omega t - \underset{(0.0013)}{99.9523}x - \underset{(0.0361)}{1.0524}F_{\mathrm{nl}} \\ \underset{(0.0027)}{10.0185}\dot{x} + \underset{(0.0092)}{5.0051}\frac{\dot{x}^2}{\lvert\dot{x}\rvert}x - \underset{(0.0030)}{20.0376}\frac{\dot{x}^2}{\lvert\dot{x}\rvert}F_{\mathrm{nl}} \end{bmatrix}$
S-RVM	$\begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dot{F}_{\mathrm{nl}} \end{bmatrix} = \begin{bmatrix} \underset{(7.5\times10^{-4})}{1.0006}\dot{x} \\ \underset{(0.0108)}{0.9983}\sin\omega t - \underset{(0.0013)}{99.9506}x - \underset{(0.0355)}{1.0529}F_{\mathrm{nl}} \\ \underset{(8.2\times10^{-4})}{10.0185}\dot{x} + \underset{(0.0028)}{5.0051}\frac{\dot{x}^2}{\lvert\dot{x}\rvert}x - \underset{(9.1\times10^{-4})}{20.0376}\frac{\dot{x}^2}{\lvert\dot{x}\rvert}F_{\mathrm{nl}} \end{bmatrix}$
SINDy	$\begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dot{F}_{\mathrm{nl}} \end{bmatrix} = \begin{bmatrix} 1.0006\dot{x} \\ 0.9997\sin\omega t - 99.8679x - 1.0605F_{\mathrm{nl}} \\ 10.0186\dot{x} + 5.0054\frac{\dot{x}^2}{\lvert\dot{x}\rvert}x - 20.0378\frac{\dot{x}^2}{\lvert\dot{x}\rvert}F_{\mathrm{nl}} \end{bmatrix}$

Table 6
The discovered equation (i.e., the selected models and identified parameters) obtained by the accelerated and conventional relevance vector machines for the Valanis model (Case-II). The model predictive uncertainty has been shown in the form of the coefficient of variation within brackets below the mean coefficients. The equation discovered by SINDy has been presented for comparison. The true coefficients adopted in the actual equation is also presented for validation.

True	$\begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dot{F}_{\mathrm{nl}} \end{bmatrix} = \begin{bmatrix} \dot{x} \\ \sin\omega t - 100x - F_{\mathrm{nl}} \\ 10\dot{x} + 30\frac{\dot{x}^2}{\lvert\dot{x}\rvert}x + 1400\dot{x}^2x^2 - 20\frac{\dot{x}^2}{\lvert\dot{x}\rvert}F_{\mathrm{nl}} \end{bmatrix}$
F-RVM	$\begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dot{F}_{\mathrm{nl}} \end{bmatrix} = \begin{bmatrix} \underset{(0.0107)}{0.9995}\dot{x} \\ \underset{(0.0107)}{1.0079}\sin\omega t - \underset{(0.0012)}{99.8770}x - \underset{(0.0123)}{1.0046}F_{\mathrm{nl}} \\ \underset{(0.0012)}{9.9942}\dot{x} + \underset{(0.0031)}{30.2019}\frac{\dot{x}^2}{\lvert\dot{x}\rvert}x + \underset{(5\times10^{-4})}{1399.4}\dot{x}^2x^2 - \underset{(5.1\times10^{-4})}{19.9829}\frac{\dot{x}^2}{\lvert\dot{x}\rvert}F_{\mathrm{nl}} \end{bmatrix}$
S-RVM	$\begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dot{F}_{\mathrm{nl}} \end{bmatrix} = \begin{bmatrix} \underset{(7.5\times10^{-4})}{0.9995}\dot{x} \\ \underset{(0.0077)}{1.0079}\sin\omega t - \underset{(8.2\times10^{-4})}{99.8770}x - \underset{(0.0088)}{1.0046}F_{\mathrm{nl}} \\ \underset{(0.0019)}{9.9942}\dot{x} + \underset{(0.0047)}{30.2019}\frac{\dot{x}^2}{\lvert\dot{x}\rvert}x + \underset{(7.7\times10^{-4})}{1399.4}\dot{x}^2x^2 - \underset{(8\times10^{-4})}{19.9829}\frac{\dot{x}^2}{\lvert\dot{x}\rvert}F_{\mathrm{nl}} \end{bmatrix}$
SINDy	$\begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dot{F}_{\mathrm{nl}} \end{bmatrix} = \begin{bmatrix} 0.9995\dot{x} \\ 1.0079\sin\omega t - 99.9048x - 0.9964F_{\mathrm{nl}} \\ 9.9958\dot{x} + 30.1965\frac{\dot{x}^2}{\lvert\dot{x}\rvert}x + 1399.4\dot{x}^2x^2 - 19.9829\frac{\dot{x}^2}{\lvert\dot{x}\rvert}F_{\mathrm{nl}} \end{bmatrix}$

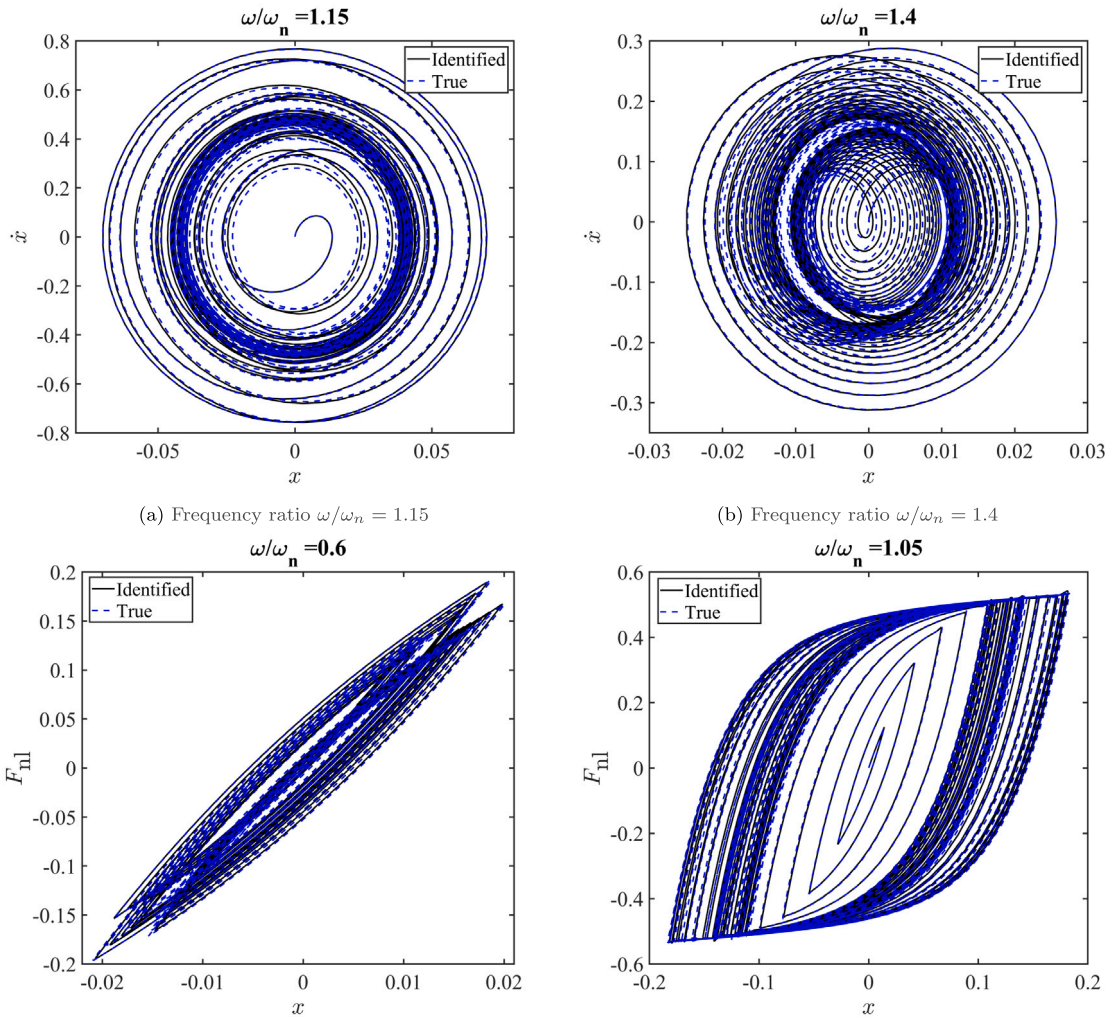
4.4.4. *Case IV:* $P(x,\dot{x}) = \alpha_1 \sin x, \quad \dot{x} \geq 0 \quad$ *and* $\quad P(x,\dot{x}) = \alpha_2 x, \quad \dot{x} < 0$

Considering the above conditional form of $P(x,\dot{x})$, $\dot{F}_{\mathrm{nl}}$ in Eq. (44) can be represented as

$$\begin{aligned} \dot{F}_{\mathrm{nl}} &= \Big(10 + \big(40\sin x - 20F_{\mathrm{nl}}\big)\Big)\dot{x} \quad \dot{x} \geq 0 \\ &= \Big(10 - \big(15x - 20F_{\mathrm{nl}}\big)\Big)\dot{x} \quad \dot{x} < 0 \end{aligned} \tag{48}$$

The results of the discovered system are presented in Table 8 and Fig. 9.

In total, 12 candidate terms were included in the basis library considering up to the second order polynomials, nonlinear restoring force and the forcing term. Close proximity of the identified models with the true coefficients of the actual terms as observed from Table 8 reveal excellent accuracy achieved by both RVM based approaches. The sparsity promoting feature in RVM proved to be effective with only identifying the relevant terms present in the actual equation and no false discoveries were observed. SINDy proved to be ineffective in capturing the nonlinear dynamics from the noisy data with multiple false discoveries as evident from Table 8. The CPU times required for fast RVM, slow RVM and SINDy are 96.1552 s, 280.5071 s and 0.0298 s, respectively. The



(a) Frequency ratio $\omega/\omega_n = 1.15$

(b) Frequency ratio $\omega/\omega_n = 1.4$

(c) Frequency ratio $\omega/\omega_n = 0.6$

(d) Frequency ratio $\omega/\omega_n = 1.05$

Fig. 6. Comparison of the true and identified time response for the Valanis model (Case I) corresponding to various values of frequency ratio (different to that used in the model training). The identified responses were generated using the mean coefficient values obtained from the accelerated relevance vector machine approach. The quantities x, $\dot{x}$ and F_{nl} depicted in the above plots are expressed in m, m/s and, N, respectively.

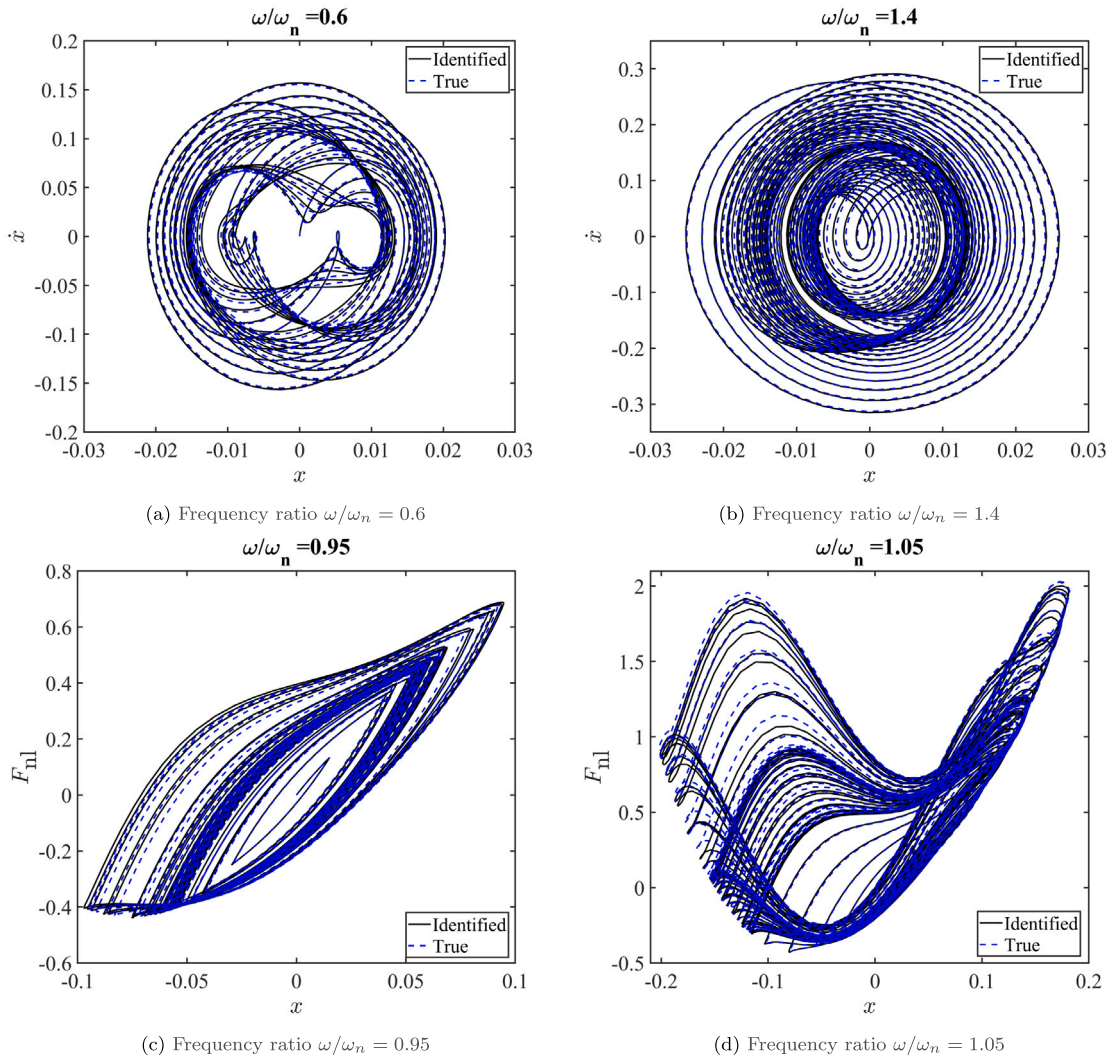
identified time response behaviour corresponding to various frequency ratios (different to that used in the model training) obtained by fast RVM have been presented in Fig. 9. These plots further confirm the accuracy of the equation discovery results in Table 8 and illustrate the effectiveness of the proposed approach on unseen data in mapping the complex nonlinear dynamic behaviour.

5. Application to experimental dataset: Discovery of a quasi zero stiffness device

This section identifies the property of the prototype Quasi Zero Stiffness (QZS) vibration isolator that is first described in [48], with further analysis of the mechanism in [49]. The data used is presented in [50]. Small changes to system parameters can have significant effects on these types of vibration isolator [51], making early detection of these changes through online dynamic measurements advantageous. The device is adjustable, so that it can create an almost zero stiffness at the loaded equilibrium position. Furthermore, it possesses stiffening nonlinearity that means that this does not come at the cost of an excessive static deflection.

5.1. Experimental setup and data generation

Fig. 10 shows the experimental set up; a mass is suspended via a length of low stretch cord from the isolator. The isolator has been manually adjusted to achieve a linear dynamic stiffness as close to zero as possible. The base of the experiment is excited by an APS113 electrodynamic shaker, with a control system that eliminates harmonics to ensure that the base motion is accurately



(a) Frequency ratio $\omega/\omega_n = 0.6$

(b) Frequency ratio $\omega/\omega_n = 1.4$

(c) Frequency ratio $\omega/\omega_n = 0.95$

(d) Frequency ratio $\omega/\omega_n = 1.05$

Fig. 7. Comparison of the true and identified time response for the Valanis model (Case II) corresponding to various values of frequency ratio (different to that used in the model training). The identified responses were generated using the mean coefficent values obtained from the accelerated relevance vector machine approach. The quantities x, $\dot{x}$ and F_{nl} depicted in the above plots are expressed in m, m/s and, N, respectively.

represented by a sinusiodal motion of the desired amplitude. Both the base and payload motions are measured using piezoelectric accelerometers and Omron ZX2-LD100 laser displacement sensors. Despite the isolator being a mult-element mechanism, Fig. 10(d) shows how it can be simply represented as a nonlinear stiffness in parallel with some Coulomb friction representing the internal friction of its joints.

Note that the sensors capture the absolute motions r and y of the base and payload respectively; identification makes use of the relative motion of the payload found by

$$x = y - r\,. \tag{49}$$

Fig. 11 represents results of a stepped sine test of the system; it is a frequency domain representation of the time data used for identification. During the upsweep, at low frequencies there is initially insufficient force to overcome friction, and the response is dominated by regions where the isolator is locked. However as the frequency reaches nearly 1.5 Hz, the response jumps up to a branch of high amplitude resonant responses, where the stiffness nonlinearities in the isolator will be significant. The up sweep must be terminated manually, as this branch of response has no upper bound, and leads to excessive amplitudes that will break the mechanism. The downsweep begins at a frequency well beyond the primary resonance of the isolator, where the motion of the payload is small and the forces transmitted to the payload are dominated by the friction and damping forces within the isolator. As the frequency falls, the amplitude rises until again there is a jump frequency where the response becomes resonant. As the frequency falls further, it drops down from the high amplitude region and soon starts to encounter locking (or sticking) of the isolator. The

Table 7
The discovered equation (i.e., the selected models and identified parameters) obtained by the accelerated and conventional relevance vector machines for the Valanis model (Case-III). The model predictive uncertainty has been shown in the form of the coefficient of variation within brackets below the mean coefficients. The equation discovered by SINDy has been presented for comparison. The true coefficients adopted in the actual equation is also presented for validation.

True	$\begin{bmatrix}\dot{x}\\ \ddot{x}\\ \dot{F}_{\mathrm{nl}}\end{bmatrix}=\begin{bmatrix}\dot{x}\\ \sin\omega t-100x-F_{\mathrm{nl}}\\ 30\dot{x}+5\times10^4\frac{\dot{x}^2}{\lvert\dot{x}\rvert}x+3\times10^4\frac{\dot{x}^3}{\lvert\dot{x}\rvert}\mathrm{e}^x-500\frac{\dot{x}^2}{\lvert\dot{x}\rvert}F_{\mathrm{nl}}\end{bmatrix}$
F-RVM	$\begin{bmatrix}\dot{x}\\ \ddot{x}\\ \dot{F}_{\mathrm{nl}}\end{bmatrix}=\begin{bmatrix}\underset{(0.4680)}{0.9996}\dot{x}\\ \underset{(0.0552)}{0.9991}\sin\omega t-\underset{(0.1462)}{99.7381}x-\underset{(0.05)}{0.9997}F_{\mathrm{nl}}\\ \underset{(0.0595)}{28.8967}\dot{x}+\underset{(0.0233)}{5.1739\times10^4}\frac{\dot{x}^2}{\lvert\dot{x}\rvert}x+\underset{(0.0031)}{3.0114\times10^4}\frac{\dot{x}^3}{\lvert\dot{x}\rvert}\mathrm{e}^x-\underset{(0.0053)}{502.7737}\frac{\dot{x}^2}{\lvert\dot{x}\rvert}F_{\mathrm{nl}}\end{bmatrix}$
S-RVM	$\begin{bmatrix}\dot{x}\\ \ddot{x}\\ \dot{F}_{\mathrm{nl}}\end{bmatrix}=\begin{bmatrix}\underset{(7.6\times10^{-4})}{0.9996}\dot{x}\\ \underset{(0.0013)}{0.9991}\sin\omega t-\underset{(0.0033)}{99.7377}x-\underset{(0.0011)}{0.9997}F_{\mathrm{nl}}\\ \underset{(0.0526)}{28.8967}\dot{x}+\underset{(0.0206)}{5.1740\times10^4}\frac{\dot{x}^2}{\lvert\dot{x}\rvert}x+\underset{(0.0027)}{3.0114\times10^4}\frac{\dot{x}^3}{\lvert\dot{x}\rvert}\mathrm{e}^x-\underset{(0.0047)}{502.7739}\frac{\dot{x}^2}{\lvert\dot{x}\rvert}F_{\mathrm{nl}}\end{bmatrix}$
SINDy	$\begin{bmatrix}\dot{x}\\ \ddot{x}\\ \dot{F}_{\mathrm{nl}}\end{bmatrix}=\begin{bmatrix}0.9996\dot{x}+1.3676x\dot{x}\\ 0.9994\sin\omega t-99.7443x-0.9997F_{\mathrm{nl}}+49.8043x^2-7.5446x\dot{x}-\dots\\ 11.2788\frac{\dot{x}^2}{\lvert\dot{x}\rvert}x\\ 28.9692\dot{x}+5.2060\times10^4\frac{\dot{x}^2}{\lvert\dot{x}\rvert}x+3.0116\times10^4\frac{\dot{x}^3}{\lvert\dot{x}\rvert}\mathrm{e}^x-502.9504\frac{\dot{x}^2}{\lvert\dot{x}\rvert}F_{\mathrm{nl}}-\dots\\ 3.5811x+4.0966\times10^3x^2+226.7527x\dot{x}-12.8410xF_{\mathrm{nl}}+9.9603\dot{x}^2\end{bmatrix}$

Table 8
The discovered equation (i.e., the selected models and identified parameters) obtained by the accelerated and conventional relevance vector machines for the Valanis model (Case-IV). The model predictive uncertainty has been shown in the form of the coefficient of variation within brackets below the mean coefficients. The equation discovered by SINDy has been presented for comparison. The true coefficients adopted in the actual equation is also presented for validation.

True	$\begin{bmatrix}\dot{x}\\ \ddot{x}\\ \dot{F}_{\mathrm{nl}}\end{bmatrix}=\begin{bmatrix}\dot{x}\\ \sin\omega t-100x-F_{\mathrm{nl}}\\ 10\dot{x}-20\dot{x}F_{\mathrm{nl}}+40\dot{x}\sin x\ ;\ \ \dot{x}\geq0\\ 10\dot{x}-15x\dot{x}+20\dot{x}F_{\mathrm{nl}}\ ;\ \ \dot{x}<0\end{bmatrix}$
F-RVM	$\begin{bmatrix}\dot{x}\\ \ddot{x}\\ \dot{F}_{\mathrm{nl}}\end{bmatrix}=\begin{bmatrix}\underset{(0.0187)}{0.9997}\dot{x}\\ \underset{(0.0107)}{1.0033}\sin\omega t-\underset{(0.0038)}{99.9195}x-\underset{(0.0295)}{1.0076}F_{\mathrm{nl}}\\ \underset{(0.0021)}{10.0118}\dot{x}-\underset{(0.0027)}{19.9631}\dot{x}F_{\mathrm{nl}}+\underset{(0.020)}{39.0006}\dot{x}\sin x\ ;\ \ \dot{x}\geq0\\ \underset{(0.0022)}{9.9999}\dot{x}-\underset{(0.0505)}{14.9991}x\dot{x}+\underset{(0.0025)}{20.0309}\dot{x}F_{\mathrm{nl}}\ ;\ \ \dot{x}<0\end{bmatrix}$
S-RVM	$\begin{bmatrix}\dot{x}\\ \ddot{x}\\ \dot{F}_{\mathrm{nl}}\end{bmatrix}=\begin{bmatrix}\underset{(7.5\times10^{-4})}{0.9997}\dot{x}\\ \underset{(0.0046)}{1.0033}\sin\omega t-\underset{(0.0016)}{99.9216}x-\underset{(0.0129)}{1.0074}F_{\mathrm{nl}}\\ \underset{(0.0016)}{10.0118}\dot{x}-\underset{(0.0020)}{19.9631}\dot{x}F_{\mathrm{nl}}+\underset{(0.0153)}{39.0005}\dot{x}\sin x\ ;\ \ \dot{x}\geq0\\ \underset{(0.0017)}{9.9999}\dot{x}-\underset{(0.0402)}{14.9991}x\dot{x}+\underset{(0.0020)}{20.0309}\dot{x}F_{\mathrm{nl}}\ ;\ \ \dot{x}<0\end{bmatrix}$
SINDy	$\begin{bmatrix}\dot{x}\\ \ddot{x}\\ \dot{F}_{\mathrm{nl}}\end{bmatrix}=\begin{bmatrix}0.9997\dot{x}\\ 1.0035\sin\omega t-99.5150x-1.0374F_{\mathrm{nl}}+3.8197x^2\\ 10.0133\dot{x}-19.9589\dot{x}F_{\mathrm{nl}}+232.1757\dot{x}\sin x-193.0051x\dot{x}\ ;\ \ \dot{x}\geq0\\ 9.9997\dot{x}-15.0229x\dot{x}+20.0323\dot{x}F_{\mathrm{nl}}\ ;\ \ \dot{x}<0\end{bmatrix}$

region of possible sticking is omitted from identification, as frictional locking poses some unique problems for an identification that is linear in parameters.

5.2. Results and discussion

The relative displacement and acceleration data used for the identification of the base excited model have been obtained from the laser sensors and accelerometers, respectively. The velocity has been obtained by the spectral derivative of measured displacement. The sampling rate of the measured signals was originally 25.6 kHz; this was downsampled to 25.6 Hz in order to give a manageable

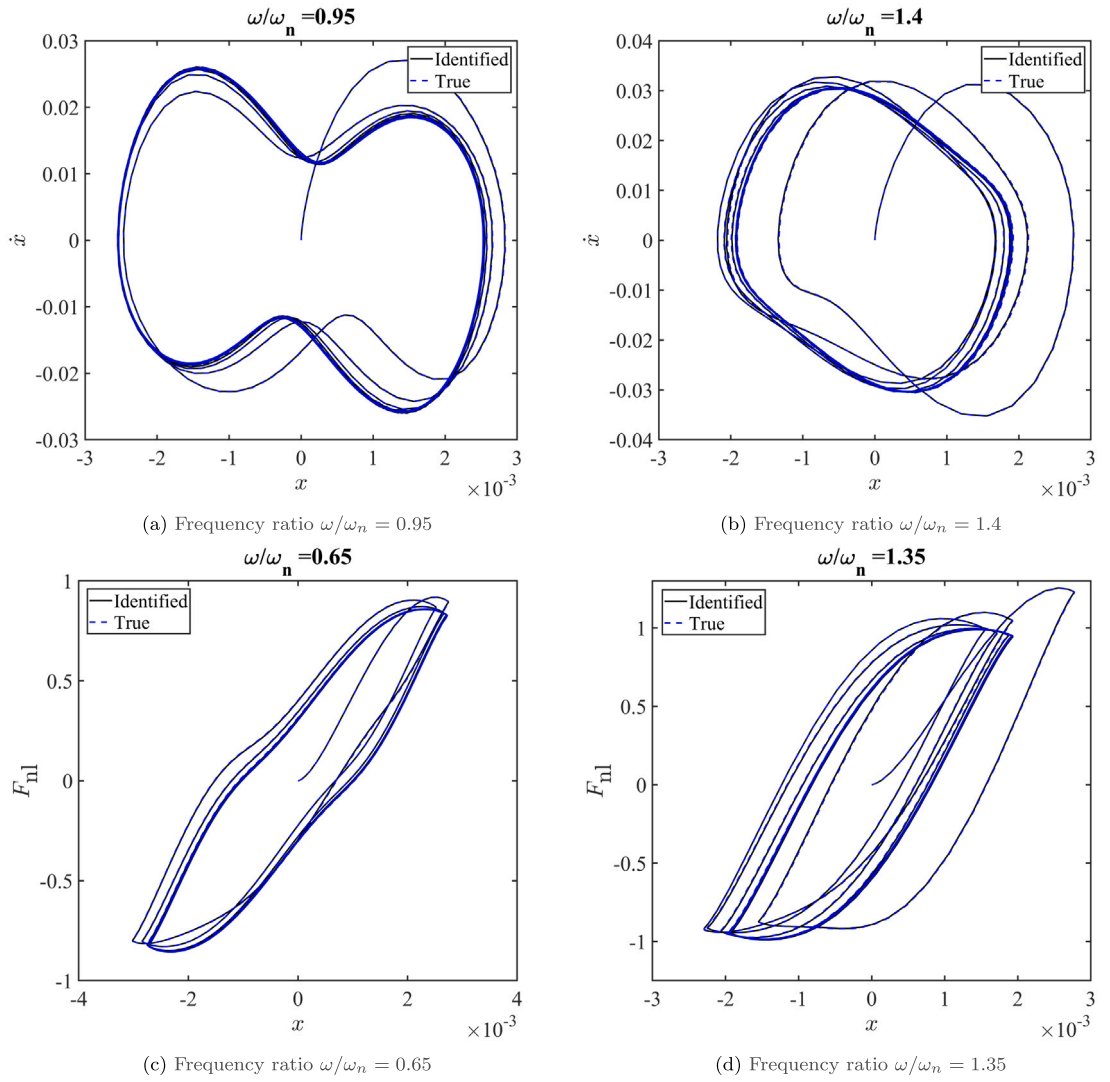


Fig. 8. Comparison of the true and identified time response for the Valanis model (Case III) corresponding to various values of frequency ratio (different to that used in the model training). The identified responses were generated using the mean coefficent values obtained from the accelerated relevance vector machine approach. The quantities x, $\dot{x}$ and F_{nl} depicted in the above plots are expressed in m, m/s and, N, respectively.

data set size, especially in consideration of the slow variant of RVM. Prior to downsampling, a low pass filter was applied to prevent aliasing.

For the identification, the dataset has been segregated in three different groups based on the low, high and mixed amplitude regimes as represented in Fig. 12. This helps verify the discovered models in terms of interpretability and consistency with expected or known physical behaviour in these particular amplitude regimes. The mixed amplitude dataset has been shown to test the generalizability and robustness of the identification algorithms. It should also be noted that Fig. 11 only represents a subset of the time data available. This is because the experiment features a feed-forward algorithm that automatically seeks a pure sinusoidal base motion of the desired amplitude. The algorithm can take many forcing cycles to achieve this condition, and in addition must also wait many cycles for each iteration to settle to a steady state response. While the time data includes all data including non steady state and intermediate iterations to reach the desired base motion, Fig. 11 only includes the end results of this process. It is a substantial advantage of time based identification methods that they make few specific requirements for settling or particular excitation conditions, because these requirements can often mean that nonlinear vibration experiments are very time consuming.

Next, the models are trained up to a certain time of the dataset and tested to predict the future evolution. The construction of training and testing dataset adopted corresponding to each amplitude regime is illustrated as follows. In particular, for the low amplitude dataset, the models are trained using responses from 0 to 500 s and then tested from above 500 s until the end i.e., 830 s. For the high amplitude dataset, the models are trained using responses from 0 to 120 s and then tested from above 120 s until

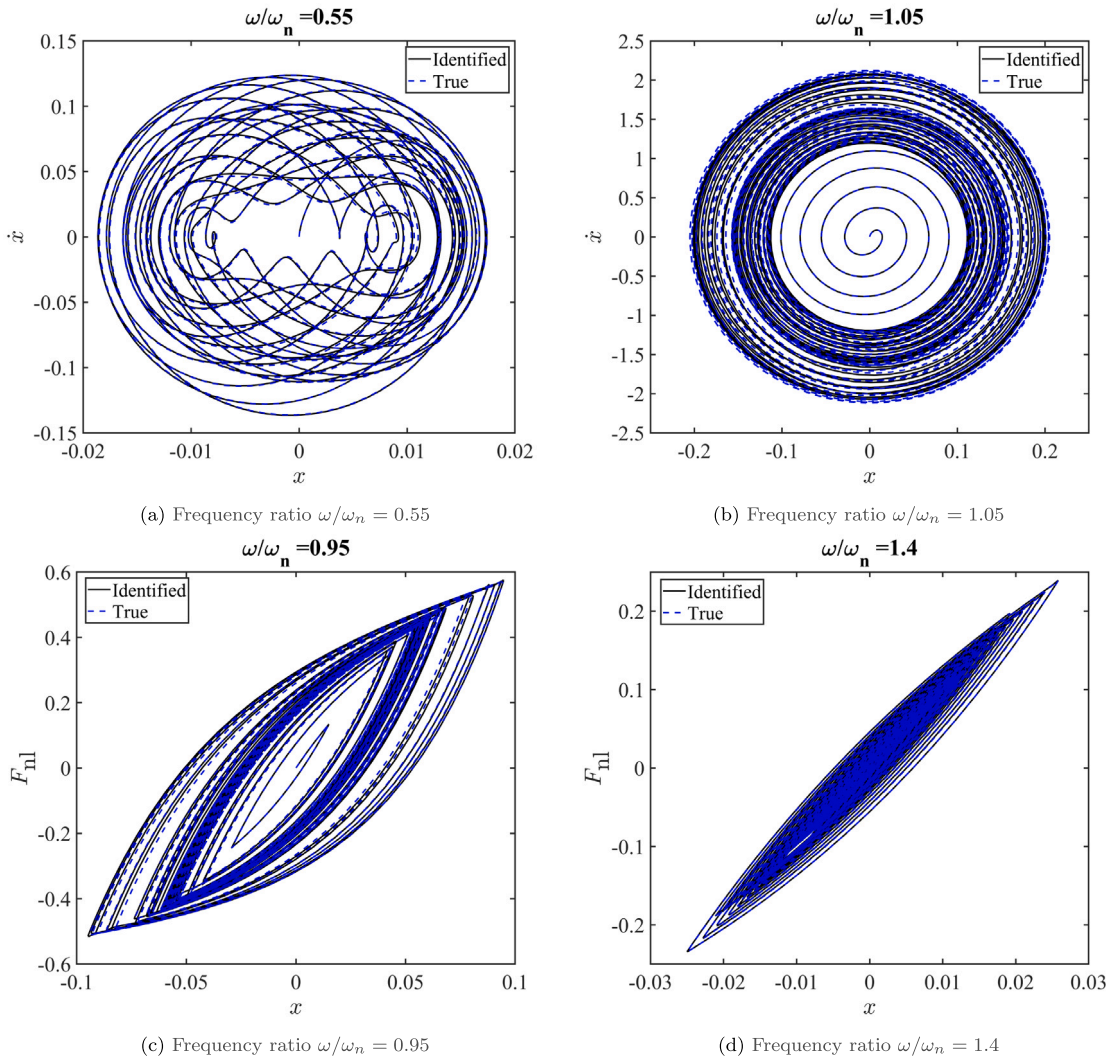


Fig. 9. Comparison of the true and identified time response for the Valanis model (Case IV) corresponding to various values of frequency ratio (different to that used in the model training). The identified responses were generated using the mean coefficent values obtained from the accelerated relevance vector machine approach. The quantities x, $\dot{x}$ and F_{nl} depicted in the above plots are expressed in m, m/s and, N, respectively.

the end i.e., 280 s. For the mixed amplitude dataset, the models are trained from 0 to 1000 s and then tested from above 1000 s until the end i.e., 1720 s. The selection of the model training time duration and train test split for low and high amplitude datasets was primarily based on a parametric study by varying the regularization (sparsity thresholding) parameter λ, the polynomial order of the state terms in the library and to achieve a reasonable test error with the identified models. The selection of the train test split for the mixed amplitude dataset is discussed in the next paragraph.

The identified and true (measured) acceleration time response corresponding to individual amplitude regimes have been presented in Fig. 13. The identified responses have been obtained using F-RVM. The train and test normalized mean squared error (NMSE) obtained using NMSE $= \sum(\hat{\ddot{x}} - \ddot{x})^2 / \sum \ddot{x}^2$ for assessing the model fit on the training and test data based on the resulting predictive models have been reported in the individual figure captions. The NMSE values obtained for the low and high amplitudes range from 1.4–2.4%, which indicates good approximation accuracy. In the mixed amplitude case, it can be observed that the train NMSE is around 2%, however, the test NMSE increases to 9%. This is because the training region mostly consists of low amplitude data whereas the test region consists of higher amplitude data. In spite of this discrepancy, achieving a test accuracy within 10% is worth noting and demonstrates strong generalizability on not only untrained but datasets with very different behaviour. A more detailed comparison of the errors obtained using F-RVM, S-RVM and SINDy have been presented in Table 9. Table 9 illustrates that all three identification algorithms have performed satisfactorily in terms of accuracy. For the mixed amplitude case, as mentioned earlier, the model was required to be trained up to 1000 s so as to cover at least the first manual disturbance (indicated by the sudden amplitude rise during the time period 900–1000 s of the signal in Fig. 13(c)). This vital information from the in-house

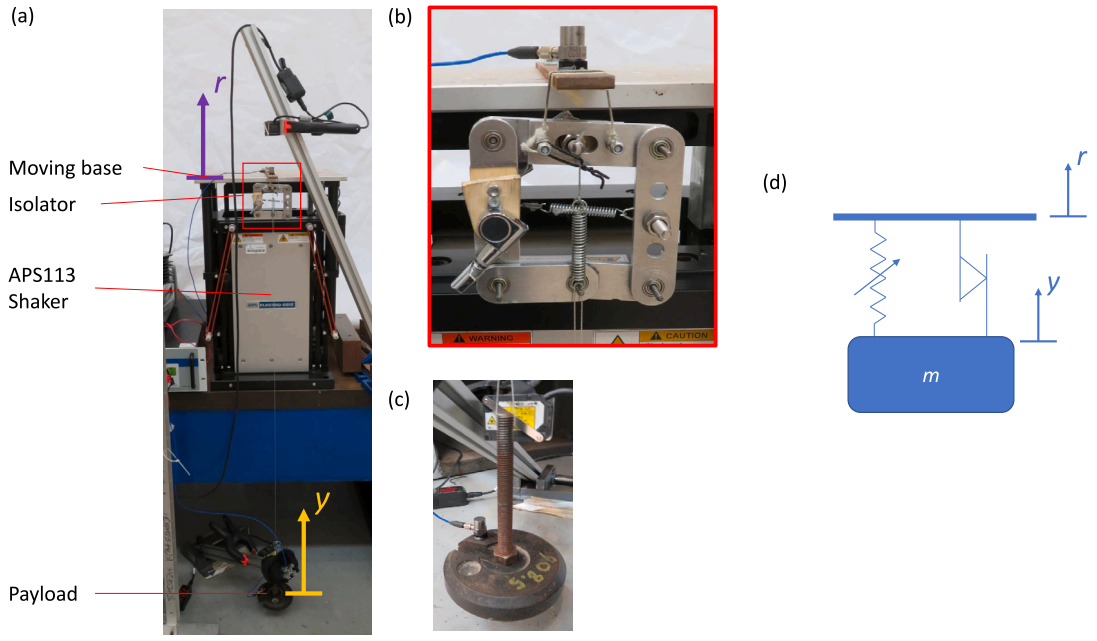


Fig. 10. (a) The experimental set up. (b) Detail of the isolator. (c) Detail of the payload. (d) Assumed SDOF model of a base excited mass suspended by a nonlinear spring in parallel with dry friction.

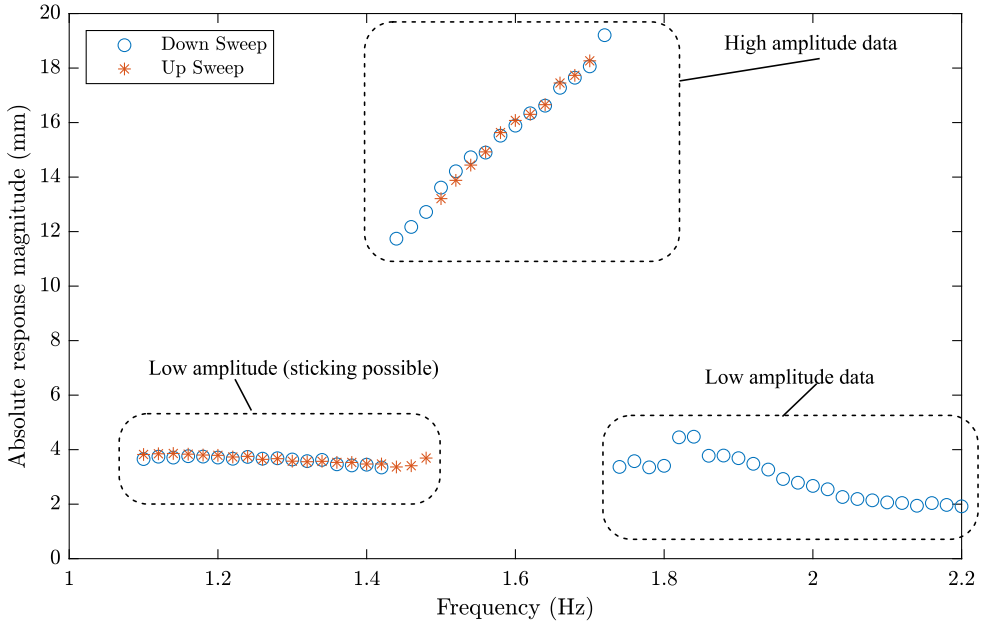


Fig. 11. Response of the suspended mass to sinusoidal base motion of 4.5 mm.

designed experiment helped to facilitate the model training and limit the test NMSE to below 10%. For training the models up to the specified time duration as reported in Table 9, the time required to train F-RVM, S-RVM and SINDy, respectively were 5.3097 s, 20.8211 s and 1.965×10^{-4} for the low amplitude regime, 1.8154 s, 8.9556 s and 0.0051 s for the high amplitude regime, 69.9780 s, 515.4777 s and 0.0087 s for the mixed amplitude regime.

The selected models and the identified parameters obtained by F-RVM, S-RVM and SINDy are presented in Tables 10–12, corresponding to the low, high and mixed amplitude regimes, respectively. As shown in Table 10, the resulting model is comprised of linear state response, base excitation and Coulomb damping $'\,\mathrm{sgn}(\dot{x})'$, which is the expected behaviour of QZS systems in the low-amplitude regime [50]. As an additional exercise, higher order state terms up to the third order were considered in the library to

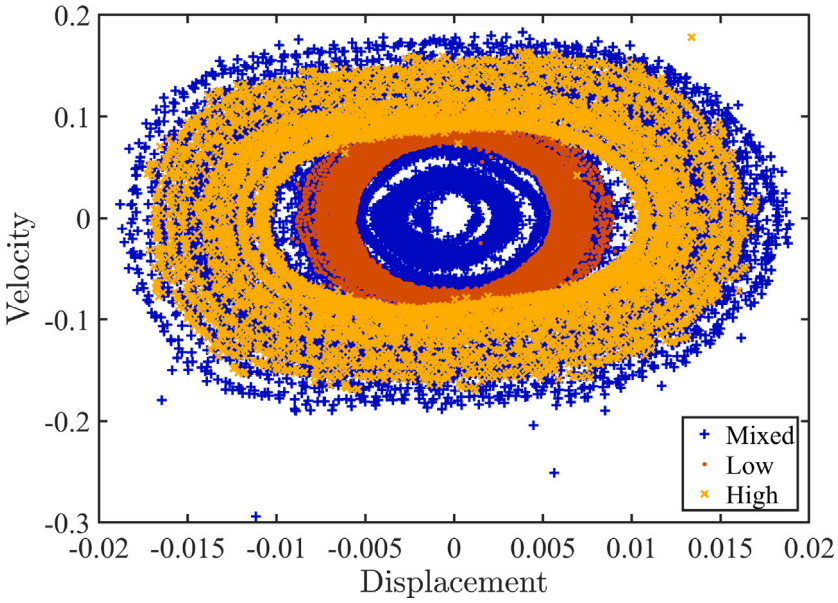


Fig. 12. Measured state data corresponding to different amplitude regimes of the experimental dataset. The displacement and velocity are expressed in m and m/s, respectively. The low, high and mixed amplitude data have been individually used for identification later.

Table 9
The normalized mean squared error (NMSE) corresponding to the mean train and mean test acceleration obtained by accelerated RVM, conventional RVM and SINDy based on predictive models trained up to a certain time and tested to predict the future evolution. The NMSEs are reported as train error/test error.

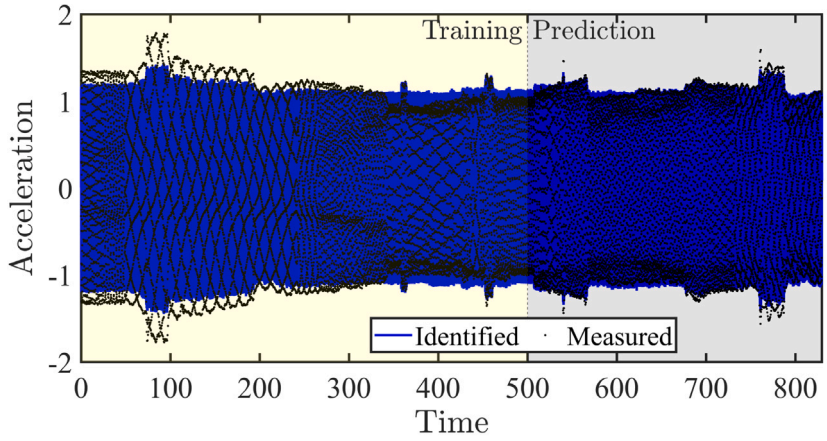
Dataset	Training duration (s)	F-RVM	S-RVM	SINDy
Low amp	500	0.0224/0.0240	0.0224/0.0240	0.0224/0.0240
High amp	120	0.0142/0.0169	0.0142/0.0169	0.0149/0.0167
Mixed amp	1000	0.0190/0.0907	0.0190/0.0907	0.0194/0.0938

Table 10
The discovered expression (i.e., the selected models and identified parameters) for the acceleration response of the quasi zero stiffness device corresponding to the low amplitude regime of the dataset obtained by the accelerated and conventional relevance vector machines. The model predictive uncertainty has been shown in the form of the coefficient of variation within brackets below the mean coefficients. The equation discovered by SINDy has been presented for comparison. Linear polynomial terms with Coulomb damping were considered and satisfactory approximation accuracy was achieved. Note that the candidate basis functions were normalized w.r.t their maximum values to obtain the identified coefficients.

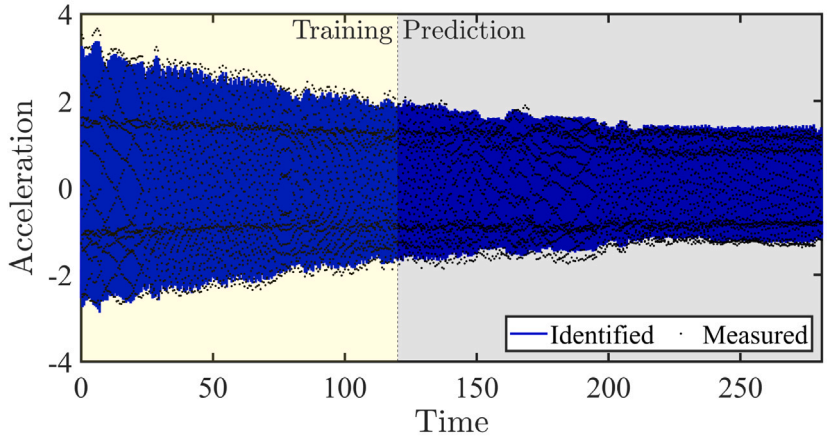
F-RVM	$\ddot{x} = \underset{(0.1740)}{-0.7623}x - \underset{(0.0584)}{0.6331}\dot{x} - \underset{(0.3217)}{0.0622}\,\mathrm{sgn}(\dot{x}) - \underset{(0.2593)}{0.4919}\ddot{r}$
S-RVM	$\ddot{x} = \underset{(0.0214)}{-0.7623}x - \underset{(0.0072)}{0.6331}\dot{x} - \underset{(0.0395)}{0.0622}\,\mathrm{sgn}(\dot{x}) - \underset{(0.0318)}{0.4919}\ddot{r}$
SINDy	$\ddot{x} = -0.7621x - 0.6330\dot{x} - 0.0623\,\mathrm{sgn}(\dot{x}) - 0.4921\ddot{r}$

capture the low amplitude dataset. It was found that most of the nonlinear higher terms did not have any contribution and resulted in close to 5% test NMSE (which is twice the present error obtained with the linear state terms). In contrast to the previous behaviour, Table 11 illustrates the dominance of higher-order nonlinear terms and elimination of Coulomb damping, which is expected in the high-amplitude regime [50]. Thus, such interpretability of the identified models help to gain physical insights on the system behaviour. The results illustrate sparse representation of the nonlinear behaviour of the QZS device (along with the model predictive uncertainty in the case of RVM based approaches). In addition to the present datasets, other datasets including few decay cases were also analysed and similar results were obtained from most, and found to be consistent with those reported. However, further investigation is required to understand cases of complex interaction which pose practical challenges causing significant change in the system behaviour due to small perturbations in damping and friction.

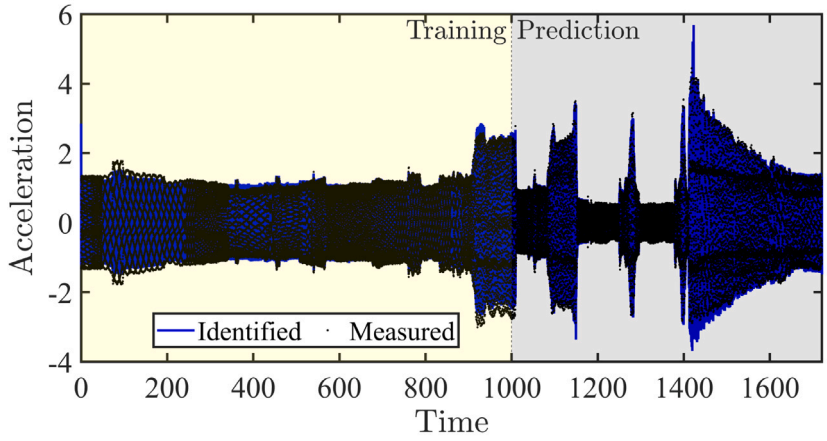
Note that the overestimated predictive variances obtained in some of the examples studied here (in Sections 4 and 5) is consistent with the literature and can be attributed to the fact that the resulting sparse probabilistic model using RVM implies degeneracy of priors over functions. This issue has been illustrated in [52] and has been improved to some extent with the help of an augmented covariance function by adding a new basis function. However, the enhancement comes at the cost of non-sparseness and higher



(a) Low amplitude regime (Train NMSE = 0.0224, test NMSE = 0.0240)



(b) High amplitude regime (Train NMSE = 0.0142, test NMSE = 0.0169)



(c) Mixed (high and low) amplitude (Train NMSE = 0.0190, test NMSE = 0.0907)

Fig. 13. Comparison of the true and identified acceleration response (m/s^2) vs time (s) for the quasi zero stiffness device corresponding to the freqsweep45-down dataset. The identified acceleration response has been generated from the mean coefficent values obtained using the accelerated relevance vector machine approach. The models are trained up to a certain time and tested on the same dataset to predict the future response behaviour. The normalized mean squared error (NMSE) corresponding to the mean test predictions is provided for each sample plots. For a detailed comparison of test NMSE, refer Table 9. The selected models and identified coefficients obtained for generating the plots have been presented in Tables 10–12.

Table 11
The discovered expression (i.e., the selected models and identified parameters) for the acceleration response of the quasi zero stiffness device corresponding to the high amplitude regime of the dataset obtained by the accelerated and conventional relevance vector machines. The model predictive uncertainty has been shown in the form of the coefficient of variation within brackets below the mean coefficients. The equation discovered by SINDy has been presented for comparison. Polynomials up to the third order were considered in the candidate library. Note that the candidate basis functions were normalized w.r.t their maximum values to obtain the identified coefficients.

F-RVM	$\ddot{x} = \underset{(0.3215)}{-0.6049}x - \underset{(0.1833)}{2.1243}\dot{x} + \underset{(0.2892)}{0.1875}x\dot{x} - \underset{(0.7174)}{0.1839}\dot{x}^2 - \underset{(0.1185)}{2.2597}x^3 + \ldots$ $\underset{(0.5347)}{0.3990}x\dot{x}^2 + \underset{(0.3916)}{0.6654}\dot{x}^3 - \underset{(1.0149)}{0.2978}\ddot{r}$
S-RVM	$\ddot{x} = \underset{(0.0560)}{-0.6047}x - \underset{(0.0319)}{2.1255}\dot{x} + \underset{(0.0503)}{0.1875}x\dot{x} - \underset{(0.1247)}{0.1838}\dot{x}^2 - \underset{(0.0206)}{2.2602}x^3 + \ldots$ $\underset{(0.0931)}{0.3988}x\dot{x}^2 + \underset{(0.0681)}{0.6656}\dot{x}^3 - \underset{(0.1763)}{0.2990}\ddot{r}$
SINDy	$\ddot{x} = -0.6321x - 2.1983\dot{x} + 0.1733x\dot{x} - 2.2075x^3 + 0.4248x\dot{x}^2 + \ldots$ $0.7979\dot{x}^3 - 0.2214\ddot{r}$

Table 12
The discovered expression (i.e., the selected models and identified parameters) for the acceleration response of the quasi zero stiffness device corresponding to the mixed (high and low) amplitude regime of the dataset obtained by the accelerated and conventional relevance vector machines. The model predictive uncertainty has been shown in the form of the coefficient of variation within brackets below the mean coefficients. The equation discovered by SINDy has been presented for comparison. Polynomials up to the third order were considered in the candidate library. Note that the candidate basis functions were normalized w.r.t their maximum values to obtain the identified coefficients.

F-RVM	$\ddot{x} = \underset{(0.1321)}{-0.9473}x - \underset{(0.0534)}{1.7042}\dot{x} - \underset{(0.1086)}{1.2727}x^3 - \underset{(0.1506)}{0.3258}x^2\dot{x} + \underset{(1.3339)}{0.1059}x\dot{x}^2 - \ldots$ $\underset{(0.2482)}{1.4196}\dot{x}^3 - \underset{(0.0950)}{0.6296}\ddot{r}$
S-RVM	$\ddot{x} = \underset{(0.0167)}{-0.9473}x - \underset{(0.0068)}{1.7042}\dot{x} - \underset{(0.0138)}{1.2727}x^3 - \underset{(0.0191)}{0.3258}x^2\dot{x} + \underset{(0.1690)}{0.1059}x\dot{x}^2 - \ldots$ $\underset{(0.0315)}{1.4196}\dot{x}^3 - \underset{(0.0120)}{0.6296}\ddot{r}$
SINDy	$\ddot{x} = -0.9617x - 1.7911\dot{x} - 1.2574x^3 - 0.3334x^2\dot{x} + 0.1207x\dot{x}^2 - \ldots$ $1.1905\dot{x}^3 - 0.6231\ddot{r}$

computational complexity. As the present study primarily focuses on an engineering application of nonlinear structural system identification, readers interested in a detailed and insightful discussion on mathematics of the resulting predictive variance by RVM are referred to [53].

6. Summary and conclusions

Considering the relatively new field of deterministic data-driven discovery of physical laws using compressive sensing has gained considerable attention over the past few years, their extension to Bayesian frameworks were found to be handful and very recent (from 2021). To address this gap, two sparse Bayesian frameworks have been investigated for the data-driven discovery of nonlinear structural dynamic systems. For incorporating (i) the sparsity promoting feature and (ii) Bayesian nature of the proposed frameworks, a machine learning approach called, relevance vector machine (RVM) has been employed to determine the governing equation of motion. Out of the two RVM algorithms, the faster version employs an efficient expectation maximization routine and has the provision of sequential addition and deletion of the candidate library terms and thus, is cost-effective compared to the slower version. For comparison, the popular SINDy approach has also been used.

To demonstrate the performance of the proposed approaches, four representative numerical examples of nonlinear structural dynamical systems have been solved. Finally, a major contribution of the present work, the data-driven discovery of a quasi zero stiffness (QZS) device was performed using the time response data generated from an in-house designed experimental setup. The importance of such complex nonlinear system identification lies in its prognostics, as QZS systems have been known to be notoriously difficult to capture. Even slight changes in damping and friction can have a significant effect on their vibration isolation ability. It is worth noting that even under these complex real-life scenarios, the proposed approaches (and SINDy) were able to achieve satisfactory level of accuracy for capturing untrained/unseen behaviour by incorporating potential candidate basis terms. In addition, the model predicted equation of motion yields physically interpretable behaviour. To illustrate the point of interpretability, the response and forcing datasets for the QZS system were divided in different categories based on the amplitude. It was shown that the discovered model could yield physically consistent behaviour such as, identifying linear state response and Coulomb damping in the low amplitude and higher-order nonlinear stiffness and damping terms with no friction in the high amplitude regime. The proposed approaches quantify the model predictive uncertainty due to their Bayesian nature and has been statistically represented by the associated coefficient of variation.

Based on the findings of the study, it would be reasonable to consider the proposed approaches as potential for investigating further complex case scenarios in QZS systems and nonlinear structural dynamics in general. An important future goal will be to reinforce the sparsity criterion for computational tractability, with suitable non-degenerate process models for robust variance prediction in equation discovery applications.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

All the authors gratefully acknowledge the support of the Engineering and Physical Sciences Research Council, United Kingdom through the award of a Programme Grant "Digital Twins for Improved Dynamic Design", grant number EP/R006768.

References

[1] K. Worden, E.J. Cross, R.J. Barthorpe, D.J. Wagg, P. Gardner, On digital twins, mirrors, and virtualizations: Frameworks for model verification and validation, ASCE-ASME J. Risk Uncertain. Eng. Syst. B 6 (3) (2020).
[2] G. Kerschen, K. Worden, A.F. Vakakis, J.-C. Golinval, Past, present and future of nonlinear system identification in structural dynamics, Mech. Syst. Signal Process. 20 (3) (2006) 505–592.
[3] B. Daniels, I. Nemenman, Automated adaptive inference of phenomenological dynamical models, Nature Commun. 6 (8133) (2015).
[4] J. Noël, G. Kerschen, Nonlinear system identification in structural dynamics: 10 more years of progress, Mech. Syst. Signal Process. 83 (2017) 2–35.
[5] J. Kutz, S. Brunton, Parsimony as the ultimate regularizer for physics-informed machine learning, Nonlinear Dynam. 107 (2022) 1801–1817.
[6] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second ed., Springer, 2009.
[7] M. Raissi, Deep hidden physics models: Deep learning of nonlinear partial differential equations, J. Mach. Learn. Res. 19 (2018) 1–24.
[8] Q. Teng, L. Zhang, Data driven nonlinear dynamical systems identification using multi-step CLDNN, AIP Adv. 9 (8) (2019) 085311.
[9] S.H. Rudy, J. Nathan Kutz, S.L. Brunton, Deep learning of dynamics and signal-noise decomposition with time-stepping constraints, J. Comput. Phys. 396 (2019) 483–506.
[10] S. Li, Y. Yang, A recurrent neural network framework with an adaptive training strategy for long-time predictive modeling of nonlinear dynamical systems, J. Sound Vib. 506 (2021) 116167.
[11] J. Bongard, H. Lipson, Automated reverse engineering of nonlinear dynamical systems, Proc. Natl. Acad. Sci. 104 (24) (2007) 9943–9948.
[12] M. Schmidt, H. Lipson, Distilling free-form natural laws from experimental data, Science 324 (5923) (2009) 81–85.
[13] S.L. Brunton, J.L. Proctor, J.N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, Proc. Natl. Acad. Sci. 113 (15) (2016) 3932–3937.
[14] B.M. de Silva, K. Champion, M. Quade, J.-C. Loiseau, J.N. Kutz, S.L. Brunton, PySINDy: A Python package for the sparse identification of nonlinear dynamical systems from data, J. Open Source Softw. 5 (49) (2020) 2104.
[15] N.M. Mangan, S.L. Brunton, J.L. Proctor, J.N. Kutz, Inferring biological networks by sparse identification of nonlinear dynamics, IEEE Trans. Mol. Biol. Multi-Scale Commun. 2 (1) (2016) 52–63.
[16] M. Hoffmann, C. Fröhner, F. Noé, Reactive SINDy: Discovering governing reactions from concentration data, J. Chem. Phys. 150 (2) (2019) 025101.
[17] A.A. Kaptanoglu, K.D. Morgan, C.J. Hansen, S.L. Brunton, Physics-constrained, low-dimensional models for magnetohydrodynamics: First-principles and data-driven approaches, Phys. Rev. E 104 (1) (2021).
[18] J.-C. Loiseau, B.R. Noack, S.L. Brunton, Sparse reduced-order modelling: sensor-based dynamics to full-state estimation, J. Fluid Mech. 844 (2018) 459–490.
[19] S. Beetham, R. Fox, J. Capecelatro, Sparse identification of multiphase turbulence closures for coupled fluid–particle flows, J. Fluid Mech. 914 (2021) A11.
[20] Y. Guan, S.L. Brunton, I. Novosselov, Sparse nonlinear models of chaotic electroconvection, R. Soc. Open Sci. 8 (8) (2021) 202367.
[21] K.P. Champion, S.L. Brunton, J.N. Kutz, Discovery of nonlinear multiscale systems: Sampling strategies and embeddings, SIAM J. Appl. Dyn. Syst. 18 (1) (2019) 312–333.
[22] E. Kaiser, J.N. Kutz, S.L. Brunton, Sparse identification of nonlinear dynamics for model predictive control in the low-data limit, Proc. R. Soc. A 474 (2219) (2018) 20180335.
[23] Z. Lai, S. Nagarajaiah, Sparse structural system identification method for nonlinear dynamic systems with hysteresis/inelastic behavior, Mech. Syst. Signal Process. 117 (2019) 813–842.
[24] S. Li, E. Kaiser, S. Laima, H. Li, S.L. Brunton, J.N. Kutz, Discovering time-varying aerodynamics of a prototype bridge by sparse identification of nonlinear dynamical systems, Phys. Rev. E 100 (2019) 022220.
[25] Q. Liu, J. Cao, Y. Zhang, Z. Zhao, G. Kerschen, X. Jing, Interpretable sparse identification of a bistable nonlinear energy sink, Mech. Syst. Signal Process. 193 (2023) 110254.
[26] H. Schaeffer, S.G. McCalla, Sparse model selection via integral terms, Phys. Rev. E 96 (2017) 023302.
[27] M. Stender, S. Oberst, N. Hoffmann, Recovery of differential equations from impulse response time series data for model identification and feature extraction, Vibration 2 (1) (2019) 25–46.
[28] S.H. Rudy, S.L. Brunton, J.L. Proctor, J.N. Kutz, Data-driven discovery of partial differential equations, Sci. Adv. 3 (4) (2017) e1602614.
[29] P.A.K. Reinbold, L.M. Kageorge, M.F. Schatz, R.O. Grigoriev, Robust learning from noisy, incomplete, high-dimensional experimental data via physically constrained symbolic regression, Nature Commun. 12 (1) (2021) 3219.
[30] L. Boninsegna, F. Nüske, C. Clementi, Sparse learning of stochastic dynamical equations, J. Chem. Phys. 148 (24) (2018) 241723.
[31] J.L. Callaham, J.-C. Loiseau, G. Rigas, S.L. Brunton, Nonlinear stochastic modelling with Langevin regression, Proc. R. Soc. A 477 (2250) (2021) 20210092.
[32] Z. Lai, C. Mylonas, S. Nagarajaiah, E. Chatzi, Structural identification with physics-informed neural ordinary differential equations, J. Sound Vib. 508 (2021) 116196.

[33] K. Champion, B. Lusch, J.N. Kutz, S.L. Brunton, Data-driven discovery of coordinates and governing equations, Proc. Natl. Acad. Sci. 116 (45) (2019) 22445–22451.
[34] G.-J. Both, S. Choudhury, P. Sens, R. Kusters, DeepMoD: Deep learning for model discovery in noisy data, J. Comput. Phys. 428 (2021) 109985.
[35] Z. Chen, Y. Liu, H. Sun, Physics-informed learning of governing equations from scarce data, Nature Commun. 12 (1) (2021) 6136.
[36] T. Hastie, R. Tibshirani, M. Wainwright, Statistical Learning with Sparsity: The Lasso and Generalizations, Chapman & Hall/CRC, 2015.
[37] R. Guimerà, I. Reichardt, A. Aguilar-Mogas, F.A. Massucci, M. Miranda, J. Pallarès, M. Sales-Pardo, A Bayesian machine scientist to aid in the solution of challenging scientific problems, Sci. Adv. 6 (5) (2020) eaav6971.
[38] R. van de Schoot, S. Depaoli, R. King, B. Kramer, K. Märtens, M.G. Tadesse, M. Vannucci, A. Gelman, D. Veen, J. Willemsen, C. Yau, Bayesian statistics and modelling, Nat. Rev. Methods Primers 1 (1) (2021) 1.
[39] M.E. Tipping, Sparse Bayesian learning and the relevance vector machine, J. Mach. Learn. Res. 1 (2001) 211–244.
[40] S. van Erp, D.L. Oberski, J. Mulder, Shrinkage priors for Bayesian penalized regression, J. Math. Psych. 89 (2019) 31–50.
[41] R. Nayek, R. Fuentes, K. Worden, E. Cross, On spike-and-slab priors for Bayesian equation discovery of nonlinear dynamical systems via sparse linear regression, Mech. Syst. Signal Process. 161 (2021) 107986.
[42] R. Fuentes, R. Nayek, P. Gardner, N. Dervilis, T. Rogers, K. Worden, E. Cross, Equation discovery for nonlinear dynamical systems: A Bayesian viewpoint, Mech. Syst. Signal Process. 154 (2021) 107528.
[43] Y.-C. Zhu, P. Gardner, D.J. Wagg, R.J. Barthorpe, E.J. Cross, R. Fuentes, Robust equation discovery considering model discrepancy: A sparse Bayesian and Gaussian process approach, Mech. Syst. Signal Process. 168 (2022) 108717.
[44] M.E. Tipping, A.C. Faul, Fast marginal likelihood maximisation for sparse Bayesian models, in: Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics, Key West, Florida, 2003.
[45] T. Chatterjee, R. Chowdhury, Refined sparse Bayesian learning configuration for stochastic response analysis, Probab. Eng. Mech. 52 (2018) 15–27.
[46] Y. Ha, H. Zhang, Fast multi-output relevance vector regression, Econ. Model. 81 (2019) 217–230.
[47] H. Jalali, N. Jamia, M.I. Friswell, H. Haddad Khodaparast, J. Taghipour, A generalization of the Valanis model for friction modelling, Mech. Syst. Signal Process. 179 (2022) 109339.
[48] A.D. Shaw, G. Gatti, P. Gonçalves, B. Tang, M.J. Brennan, Design and test of an adjustable quasi-zero stiffness device and its use to suspend masses on a multi-modal structure, Mech. Syst. Signal Process. 152 (2021) 107354.
[49] G. Gatti, A.D. Shaw, P. Gonçalves, M.J. Brennan, On the detailed design of a quasi-zero stiffness device to assist in the realisation of a translational Lanchester damper, Mech. Syst. Signal Process. 164 (2022) 108258.
[50] A.D. Shaw, P. Gonçalves, G. Gatti, M.J. Brennan, Frictional phenomena within a quasi zero stiffness vibration absorber, in: Proceedings of ISMA, 2022, pp. 834–848.
[51] A.D. Shaw, S.A. Neild, M.I. Friswell, Relieving the effect of static load errors in nonlinear vibration isolation mounts through stiffness asymmetries, J. Sound Vib. 339 (2015) 84–98.
[52] C. Rasmussen, J.Q. Candela, Healing the relevance vector machine by augmentation, in: Proceedings of the 22nd International Conference on Machine Learning, ACM Press, 2005, pp. 689–696.
[53] J.Q. Candela, Learning with Uncertainty – Gaussian Processes and Relevance Vector Machines (Ph.D. thesis), Technical University of Denmark, 2004.