International Journal of Non–Linear Mechanics 170 (2025) 105002

Contents lists available at ScienceDirect

International Journal of Non-Linear Mechanics

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

Data-driven structural identification of nonlinear assemblies: Uncertainty Quantification

Sina Safari [a,*], Diogo Montalvão [b], Julián M. Londoño Monsalve [c]

[a] School of Electrical, Electronic and Mechanical Engineering, University of Bristol, Bristol BS8 1TR, UK
[b] Department of Design & Engineering, Bournemouth University, Poole BH12 5BB, UK
[c] Faculty of Environment, Science and Economy (ESE), University of Exeter, Exeter EX4 4QF, UK

ARTICLE INFO

MSC:
00-01
99-00

Keywords:
Nonlinear system identification
Uncertainty quantification
Multiple Data sets
Model selection
Vibration testing

ABSTRACT

Nonlinear model identification from vibration data is challenging due to limited measured data collected during the testing campaign and since the identified model should be capable of accounting for the uncertainties arising from the reassembly of the structure, environmental effects, and slight changes in parameters as a result of wear during vibration testing. In this paper, a new technique based on ensembling is proposed for uncertainty quantification during the identification of nonlinear assemblies using multiple data sets. First, an ensemble of parsimonious models is identified using a physics-informed nonlinear model identification method from subsets of measured data. Aggregate model statistics are then employed to calculate inclusion probabilities for the candidate model, which enable uncertainty quantification and a probabilistic estimate of the dynamic response. This results in a robust nonlinear model identification with physical interoperability. An application on a single-degree-of-freedom system idealised for an experimental structure with geometric and friction nonlinearities is presented. The results obtained demonstrate the substantial performance of the proposed technique in selecting accurate nonlinear models that capture the response over a large range of variability and repeatability for real-world data sets.

1. Introduction

There has been rapid growth in developing nonlinear system identification methods for identifying nonlinear mechanical systems in the last decade [1]. Since then, development has continued on different fronts, such as testing techniques [2,3], building mathematical models [4–6], and nonlinear model reduction [7] for identification. One challenge in the identification process for the dynamics of nonlinear structures is that measurements of nonlinear assemblies might exhibit a high degree of uncertainty. This uncertainty, which includes both variability and repeatability [8,9], poses challenges in making accurate predictions of dynamic responses. Variability denotes the inherent differences observed in measurements among nominally identical assemblies, whereas repeatability is defined by the differences in measurements across multiple trials conducted on the same assembly with identical environmental conditions. For example, whenever an aeroturbine undergoes reassembly after maintenance, its dynamic behaviour might undergo significant changes owing to the limited repeatability across assemblies. This would hinder the accurate prediction of the aeroturbine's remaining life, complicating structural health monitoring and rendering optimal maintenance scheduling unfeasible. The diminished repeatability could render much of the previously collected response data ineffective.

Nonlinear system identification methods are typically categorised into parametric and non-parametric methods. Parametric methods need the definition of a model in advance or to select the best model that fits the data. On the other hand, non-parametric methods do not require a model and describe the dynamics in terms of their characteristics, such as instantaneous frequency and damping [10], and frequency response curves [3]. Within parametric methods, several automated model selection algorithms have been developed previously, including the Forward Regression Orthogonal Least Square (FROLS) algorithm [11], Polynomial Nonlinear State–Space (PNLSS) [12], and Sparse Identification of Nonlinear Dynamics (SINDy) [13]. These algorithms have advantages over Neural Network (NN)-based methods [5] due to their interpretability and are commonly used for problems with linear-in-parameter nonlinear models. Moreover, the authors have developed a new physics-informed nonlinear system identification method [14] that

* Corresponding author.
E-mail addresses: sina.safari@bristol.ac.uk (S. Safari), dmontalvao@bournemouth.ac.uk (D. Montalvão), J.Londono-Monsalve@exeter.ac.uk (J.M.L. Monsalve).

https://doi.org/10.1016/j.ijnonlinmec.2024.105002
Received 25 January 2024; Received in revised form 27 October 2024; Accepted 19 December 2024
Available online 26 December 2024
0020-7462/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

handles nonlinear-in-parameter models [15]. This method uses a well-designed optimisation problem for nonlinear identification that allows the use of measured data and modal information about the physical system to constrain the problem so that the identified model is physics-informed and interpretable. This new method has been successfully applied to geometrically nonlinear structures [14], structures with bolted joints where the response of nonlinear elements is not directly measurable [16], and structures with asymmetric nonlinearities [17].

To account for uncertainties, model class selection has been practised using the Bayesian inference approach [18–20]. However, the Bayesian approach requires high computational execution time [21]. As an alternative, ensembling techniques have been introduced within the SINDy method [22]. It works out the most probable model by solving a regression problem multiple times for different data sets generated using the bootstrapping technique. The ensembling technique is compared with the Bayesian approach by Mars Gao et al. [21] and it has been shown to be effective and efficient.

This paper extends the data-driven identification method, developed by the authors [14], by combining it with the ensembling technique so that it can quantify the uncertainties due to variability and repeatability. Therefore, the main contribution of this paper is to embody uncertainties in the identification process using multiple experimentally measured datasets. Further, it attempts to identify and automatically build mathematical models for damping and stiffness nonlinearities that account for large changes in the structure due to reassembly and slight changes in parameters due to wear during the vibration test. It should be noted that the paper demonstrates the application of the proposed uncertainty quantification technique to SDOF systems and studying MDOF systems is out of the scope of this work. Interested readers may refer to [16] for the extension of the identification method to MDOF.

The paper is organised as follows: Section 2 outlines the background theory related to nonlinear structural identification and model selection, including the generation of optimisation problems, with a focus on the algebraic- and simulation-based cost functions. The integration of the ensembling technique with the identification method is described in Section 3. Section 4 presents the results obtained using the proposed method when applied to experimental free decay data sets acquired from a nonlinear structure featuring friction and geometric nonlinearities that are ideally modelled as an SDOF system. This is followed by conclusions in Section 5.

2. Background: nonlinear model identification

Here, a data-driven model discovery method [14] to identify parsimonious nonlinear models from vibration measurement data of engineering structures is discussed. The method described here is based on a formulation for SDOF structures; however, an extension for MDOF structures can be easily implemented based on the authors' work in [14,16,23]. For example, a two-degree-of-freedom system was numerically investigated in [14] to simply show the application of the method on the MDOF systems. In addition, the method has been successfully applied to identify systems with multiple nonlinear elements in [16] where the direct measurement of nonlinear element response is not possible. These two examples show how modal coupling can be considered during the model identification process.

The dynamical system with nonlinear damping and stiffness is represented by the following differential equation:

$$M\ddot{q}(t) + C\dot{q}(t) + Kq(t) + f_{nl}(q(t), \dot{q}(t)) = F(t) \tag{1}$$

In this equation, $\ddot{q}$, $\dot{q}$ and q are respectively the acceleration, velocity, and displacement; M, C, and K are the mass, damping, and stiffness of the structure, and F is the force vector. The function f_{nl} contains all conservative and non-conservative nonlinear forces. To keep the equations short, the time instance t is omitted in the later equations. For the sake of completeness, modal transformation can be applied using $\mathbf{q} = \phi\mathbf{u}$ and the linear part of Eq. (1) can be written in terms of linear natural frequency (ω_n) and damping ratio (ζ).

$$\ddot{\mathbf{u}} + 2\zeta_k\omega_{nk}\dot{\mathbf{u}} + \omega_{nk}^2\mathbf{u} + \boldsymbol{\phi}^T f_{nl}(\boldsymbol{\phi}\mathbf{u}, \boldsymbol{\phi}\dot{\mathbf{u}}) = \boldsymbol{\phi}^T\mathbf{F} \tag{2}$$

It should be noted that $\boldsymbol{\phi}$ is the mode shape that is used for modal transformation in MDOF systems [14] and the underlying linear damping is assumed to be proportional damping. For SDOF systems it can be simply equal to M^{-1}.

We consider here the nonlinear model identification method proposed in [14]. Engineering structures are typically modelled linearly at low vibration amplitudes. This assumption, however, may not hold for certain special cases, such as systems exhibiting pure Coulomb friction nonlinearity. Despite these exceptions, we maintain that a linear model can be derived to approximate the structural behaviour at low amplitudes. It is also assumed that nonlinear effects become increasingly apparent as the vibration amplitude rises. Hence, in this work, the methodology starts with identifying and experimentally validating the characteristics of the underlying linear system, i.e., natural frequencies (ω_{nk}), damping ratios (ζ_k) using one of the standard available methods, e.g., Polymax [24].

Model discovery and parameter estimation for the nonlinear force f_{nl} are carried out based on the cascade of optimisation problems described in the following subsections.

2.1. Algebraic-based cost function

The nonlinear dynamic system described in Eq. (2) can be identified by setting up an optimisation problem. The optimisation problem estimates the parameters of the underlying linear system as well as the nonlinear model using nonlinear algebraic regression. This nonlinear algebraic regression was inspired by numerical solvers of dynamic equations of motion, such as the Newmark method [25], which minimises the residuals of internal and external forces. The optimisation problem uses the mean square error (MSE_a) given by Eq. (3) as the algebraic-based cost function to measure the discrepancy between observed and predicted data:

$$Minimise: MSE_a = \frac{1}{n}\sum_{i=1}^{n}(y_i^* - y_i)^2 \tag{3}$$

where n is the size of the time series data and y^* and y are the observed and predicted data which are equal to the right-hand side and left-hand side of the Eq. (2) respectively. It should be noted that the parameters of the underlying linear system as well as the nonlinear model f_{nl} in Eq. (2) are collected in the parameter vector $\boldsymbol{\theta}$.

The algebraic-based cost function defined in this section requires the availability of the following time-domain data: acceleration, velocity, and displacement responses, as well as the applied excitation forces. Consequently, there is no need to solve the differential equation. The typical response and force acquisition equipment used in the vibration testing of structures are accelerometers and load cells, respectively. In this section, acceleration response and excitation force data are assumed to be measured directly, while displacement and velocity responses are obtained from experimentally measured acceleration data by numerical integration [14,17,26]. This approach does not provide an accurate estimation of displacement for complex nonlinear systems; however, it is still beneficial and has been proven to be effective for rapid model discovery [17].

The type of applied excitation is important. The proposed method could use harmonic, sine sweep and pull-release (free decay) excitation techniques to vibrate the structure near its resonance where the nonlinear feature needing to be identified is activated [14]. In this work, the pull-release technique is used, and the free decay response is used for system identification. It should be noted that Eq. (2) can be directly used for MDOF systems when multiple modal equations exist [16]. However, studying MDOF systems is outside the scope of this work.

Table 1
Library of nonlinear terms used in this paper.

Linear-in-parameters nonlinear terms							
Index	Term	Index	Term	Index	Term	Index	Term
1	$\|q\|q$	12	$\|q\|q\dot{q}$	23	$\|q\|\dot{q}$	34	$\dot{q}^9$
2	q^3	13	$q^3\dot{q}$	24	$\|q\|^{1.5}\dot{q}$		
3	$\|q\|q^3$	14	$\|q\|q^3\dot{q}$	25	$\|q\|^2\dot{q}$		
4	q^5	15	$q^5\dot{q}$	26	$\text{sign}(\dot{q})$ or $\dot{q}/\|\dot{q}\|$		
5	$\|q\|q^5$	16	$\|q\|q^5\dot{q}$	27	$\|\dot{q}\|\dot{q}$		
6	q^7	17	$q^7\dot{q}$	28	$\dot{q}^3$		
7	$\text{sign}(q)\sqrt{\|q\|}$	18	$\text{sign}(q)\sqrt{\|q\|}\dot{q}$	29	$\|\dot{q}\|\dot{q}^3$		
8	$q\sqrt{\|q\|}$	19	$q\sqrt{\|q\|}\dot{q}$	30	$\dot{q}^5$		
9	$q\|q\|\sqrt{\|q\|}$	20	$q\|q\|\sqrt{\|q\|}\dot{q}$	31	$\|\dot{q}\|\dot{q}^5$		
10	$q^3\sqrt{\|q\|}$	21	$q^3\sqrt{\|q\|}\dot{q}$	32	$\dot{q}^7$		
11	$q\dot{q}$	22	$\sqrt{\|q\|}\dot{q}$	33	$\|\dot{q}\|\dot{q}^7$		

Nonlinear-in-parameters nonlinear terms		
Index	Term	Description
35	$\begin{cases} k_t q & \|k_t q\| < f_s \\ f_y \text{sign}(q) & \text{otherwise} \end{cases}$	**Jenkins model** k_t:tangential stiffness, f_y:slip force
36	$f_y((2/(1+e^{(-\sigma_g q)})) - 1) + k_t q$	**Generalised Jenkins Prager model** σ_g:smooth transition parameter
37	$\begin{cases} \text{sign}(q)(k_{d_1}(q - d_1/2))^{c_1} & q < -d_1/2 \\ 0 & -d_1/2 < q < d_2/2 \\ \text{sign}(q)(k_{d_2}(q - d_2/2))^{c_2} & q > d_2/2 \end{cases}$	**Clearance model** k_{d_1}, k_{d_2}: contact stiffness, d_1, d_2:clearance, c_1, c_2: polynomial coefficients
38	$\begin{cases} k_t q - \left(\frac{k_t(\beta + \frac{\chi+1}{\chi+2})}{f_y(1+\beta)}\right)^{1+\chi} \ldots \\ \quad \frac{k_t}{(1+\beta)(\chi+2)} q^{\chi+2} & \|q\| \le \phi_{max} \\ f_y \text{sign}(q) & \|q\| \ge \phi_{max} \end{cases}$	**4 – parameters Iwan model** $\phi_{max} = \frac{f_y(1+\beta)}{k_t(\beta+(\frac{\chi+1}{\chi+2}))}$ β and χ are the shape parameters.
39	$\begin{cases} k_t q - \left(\frac{k_t(\beta + \frac{\chi+1}{\chi+2})}{f_y(1+\beta)}\right)^{1+\chi} \ldots \\ \quad \frac{k_t}{(1+\beta)(\chi+2)} \ldots \\ \quad \left(\frac{\Theta}{\chi+2} - \frac{1}{\chi+1}\right) q^{\chi+2} & \|q\| \le \phi_{max} \\ \Theta f_y & \|q\| \ge \phi_{max} \end{cases}$	**5 – parameters Iwan model** Θ: the ratio of slip force to force required to initiate slip.

2.2. Nonlinear model selection

This section describes the forward–backward (FB) model selection method [14] for discovering nonlinear models using the optimisation problem defined in Section 2.1. The method uses a predefined and comprehensive library of nonlinear terms typically encountered in common engineering structures (Table 1).

The forward selection involves adding each nonlinear term to the structural model one at a time. The benefit of forward selection is that it begins with the underlying linear model and uses estimated parameters of the latest discovered nonlinear model as initial conditions before adding a new term in the subsequent step. This reduces the chance of becoming stuck in local minima in a high-dimensional search space. Then, in order to eliminate terms with a negligible contribution and produce a parsimonious model, a backward elimination technique is used. The key components of the FB method are outlined in [16], and a complete breakdown can be found in [14].

Two stopping criteria, MSE_a and $\Delta MSE_a^{(s)}/\Delta s$ values, are needed in the process of model selection based on the FB method. Here s is the progression number of the model selection method. Primarily, MSE_a is enforced as stopping criteria for model selection when its value drops lower than a user-assigned threshold $< \varepsilon_1$. Besides, when the change of two consecutive MSE_a values is less than ε_2 as in Eq. (4), the algorithm stops and delivers the nonlinear model. Notice that the latter criterion is determinant in the cases when by adding/removing more nonlinear terms, the model prediction does not improve significantly, and so, the extra complexity is not worth it; in such cases, the extra term will be rejected. Only the criterion in Eq. (4) is used for backward elimination with an expected 5% higher error based on the last step of forward selection to provide more flexibility in removing redundant terms.

$$\frac{\Delta MSE_a^{(s)}}{\Delta s} < \varepsilon_2 \tag{4}$$

Once the nonlinear model is selected and parameter estimation is completed based on the algebraic-based cost function, the discovered model can be transferred to the next stage, where parameter tuning is carried out based on simulation responses.

2.3. Simulation-based cost function

This section introduces a parameter tuning (or model updating) method based on matching the characteristics of time-domain data obtained from experimental measurement and simulation. Since the measured time-domain data is acceleration, therefore, the acceleration data obtained from solving the differential equation defined in (2) is used in this section. The characteristics include instantaneous frequency (IF) and instantaneous amplitude (IA). Alternatively, IA can be described as the envelope of the response amplitude. The zero-crossing method [10] is applied to estimate IF and IA. This is applicable when the measured data includes the forced or decay responses near a natural frequency of the system with low nonlinear modal couplings. This means that there is no internal resonance with the mix of multiple distant frequencies in the recorded signal.

The zero-crossing method is based on the detection of the zero-crossing points of the response signal and the use of a standard interpolation algorithm to determine the crossing times [10]. To ensure accurate identification, this resolution is adjusted to match characteristics extracted from simulation and experimental data.

An optimisation problem can now be defined based on the IF and IA values for tuning the parameters of the identified model using the method presented in Sections 2.1 and 2.2. The optimisation problem uses the mean square error (MSE_s) given in Eq. (5) as the simulation-based cost function to measure the discrepancy between measured and simulated data:

$$Minimise: MSE_s = \frac{1}{N}\sum_{j=1}^{N}(IF^m(j) - IF^s(j))^2 + \frac{1}{N}\sum_{j=1}^{N}(IA^m(j) - IA^s(j))^2 \quad (5)$$

where N is the size of the instantaneous characteristic values. The superscripts m and s indicate the characteristics obtained from measured and simulated time-domain data, respectively. Similar to Section 2.1, the parameters of the underlying linear system as well as the nonlinear model f_{nl} in Eq. (2) are collected in the parameter vector θ.

2.4. Optimiser: algorithm, scaling, bounding and initialisation

In many engineering applications, the observed nonlinearities demand using complicated functions for characterisation. These functions are normally nonlinear-in-parameters which means that a nonlinear optimisation method is required to solve the optimisation problem defined above in Eq. (5). Several nonlinear optimisers have been assessed by the authors in [23] for this task; interested readers can refer to it for more details. In this study, the Trust-Region-Reflective (TRR) optimisation algorithm [27] with a multi-start strategy is used to minimise the cost functions defined in Sections 2.1 and 2.3.

Three factors are essential for successfully running a nonlinear optimisation problem and a better chance of reaching a global minimum: scaling, bounding, and initialisation of the parameters in the search space. These facilitate the convergence of optimisation algorithms. The scaling, bounding, and initialisation strategies defined in this section are used when solving both cost functions defined in Sections 2.1 and 2.3. The parameters can be scaled linearly [14] or logarithmically [28]. A combination of linear and logarithmic scaling is used in this study. A new bi-symmetric logarithmic scaling according to Eq. (6) is used to scale the search space of parameters, which handles negative values while maintaining continuity across zero [28]:

$$\theta_s = \text{sign}(\theta)\log_{10}(1 + \frac{|\theta|}{c}) \qquad \theta = \text{sign}(\theta_s)c(-1 + 10^{|p_s|}) \quad (6)$$

where θ is the unscaled (reference) parameter value and θ_s is the corresponding value in the scaled space. Here, c is a vector of scaling factors that controls how much the search space is compressed. For the parameters of the underlying linear system, this scaling factor is set to the values calculated in the linear modal testing. For the parameters of the nonlinear terms, it is set to the maximum physically possible value, which is calculated based on the sum of the estimated maximum absolute linear force f_{lmax} (inertia max($|\mathbf{M}\ddot{\mathbf{q}}|$), damping max($|\mathbf{C}\dot{\mathbf{q}}|$) and stiffness forces max($|\mathbf{C}\mathbf{q}|$)) divided by the estimated maximum quantity of selected nonlinear term. For example, in the case that the nonlinear force is described by $f_{nl}(q) = \theta(4)|q|q + \theta(5)q^3$, the scaling factors are set to $(f_{lmax}/max(||q|q|))$ and $(f_{lmax}/max(|q^3|))$. This is done due to the fact that the contribution of nonlinear force to the total force response should be significant to affect the dynamics of the system. Although it is unlikely that this contribution exceeds the total force of the underlying linear system f_{lmax}, to consider higher variability in the parameters, the bounds can be considered wider. In addition, the maximum relative displacement/velocity in the location of nonlinear elements is used to scale the parameters with displacement/velocity units in the nonlinear model. The employed scaling scheme is suitable for multi-dimensional parameter spaces with different orders of magnitude.

The lower bound (superscript l_b) and upper bound (superscript u_b) are defined for the parameters in the scaled space. The bounds for the linear parameters are defined based on the variations of 5% of natural frequencies and 20% of linear damping ratios obtained from linear modal identification. A low variation for linear frequencies and a high variation for damping ratios are considered, which can be changed by the user.

After applying Eq. (6), the bounds for the scaled parameters of nonlinear model $\theta_s(.)$ are set to [−0.31,0.31]. However, in order to consider higher variability in the parameters, the bounds can be considered wider, e.g., [−1,1]. It should be noted that for some parameters, e.g., the dead-zone (clearance) parameter, a negative value is not physically meaningful; therefore, only a positive search space is defined for them.

Initialisation is carried out differently for model selection based on algebraic-based cost function (Eq. (3)) and parameter tuning steps based on simulation-based cost function (Eq. (5)). In the model selection step, the model parameters are initialised randomly using a multi-start strategy with 50 different values for each parameter within the specified bounds. Nonetheless, these values are more densely distributed near the optimum values calculated in the earlier step. That is, the initialisation of linear parameters is done based on the existing parameters of the underlying linear system from linear modal testing, i.e., damping ratios and the natural frequencies of the system. In addition, during the nonlinear model selection, estimated values for the parameters of a specific nonlinear term are recorded to be used in the following iteration of the model selection (see Section 2.2). It should be emphasised that the initialisation is also done within the bounds of scaled space. Interested readers can refer to [14,23] for more details on strategies for initialisation and defining a physics-informed space of search for the optimisation problem in the context of structural identification of nonlinear assemblies. In the parameter tuning step, the parameters are only initialised from the estimated parameters in the model selection step based on the algebraic-based cost function (Eq. (3)).

3. Uncertainty quantification via ensembling

In this section, we leverage a statistical approach to robustify the identification algorithm presented in Section 2. First, an ensemble of models is identified from subsets of experimentally measured data using the method described in Section 2.2. The ensemble of models includes all terms that are selected for different subsets. The selected model statistics are then used to produce inclusion probabilities for the candidate functions. The inclusion probability for a selected term in the ensemble of models is calculated based on the number of times a term is repeated (or selected) for different subsets over the number of subsets.

This statistical approach enables uncertainty estimates for the discovered model coefficients and probabilistic estimation of dynamic responses, thus connecting to Bayesian model identification techniques. In other words, the identified ensemble of model coefficients can be used to compute parameter probability density functions, which form a posterior distribution $P(\theta|D)$. In terms of estimating the dynamic response, either the mean of the identified parameter can be used or samples can be drawn from multiple identified models to generate ensemble estimates that represent posterior predictive distributions, which provide prediction confidence intervals.

The final model is defined based on the inclusion probability that is assigned by the user and is recommended between 0.5 and 0.8 in this study based on our observation. It should be noted that the higher the inclusion probability the lower the number of nonlinear terms. When the final model is identified for a specific data set based on the inclusion probabilities, the optimisation problem defined in Section 2.3 is employed to refine or tune the distribution of the parameters. In this case, the bounds and initial conditions are defined based on the parameter distributions obtained from the model selection step via

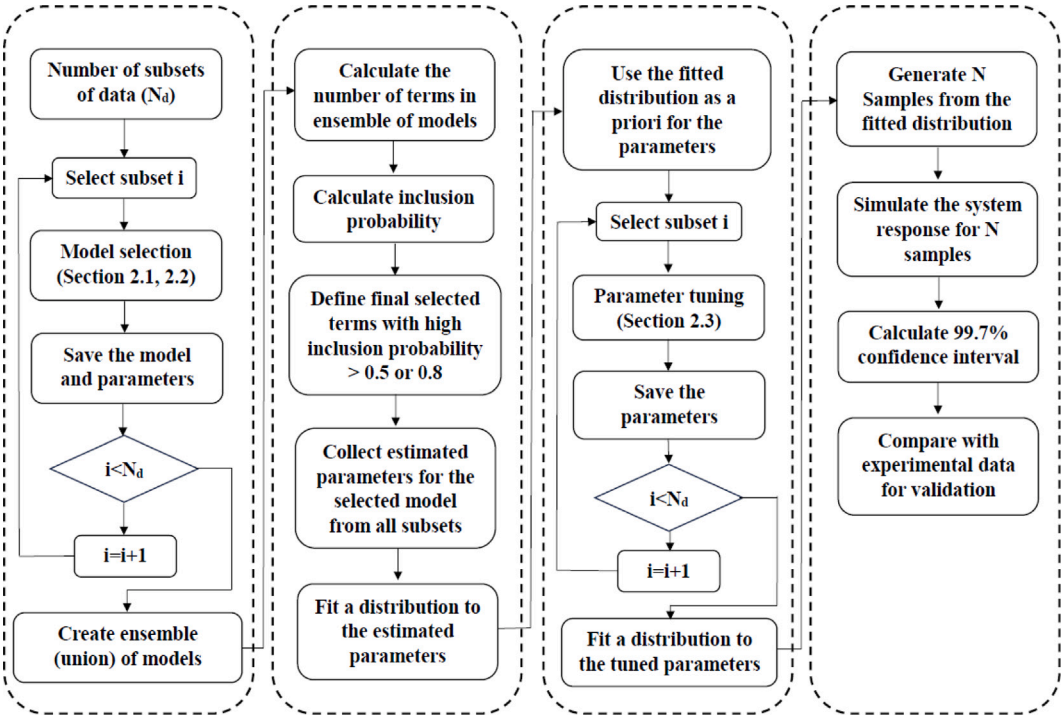

Fig. 1. Flowchart of the proposed method for uncertainty quantification using ensembling technique and multiple data sets.

the algebraic-based cost function (Eq. (3)). Afterwards, the simulation-based cost function (Eq. (3)) is solved for the data set, and a new posterior distribution for the parameters is learned.

Fig. 1 shows the flowchart of the proposed identification using the ensembling technique implemented in this paper.

4. Results

4.1. Experimental setup and data

The SDOF test structure considered in this work consists of a lumped mass supported by four thin plates and fixed as shown in Fig. 2. The thin plates are considered a source of geometric nonlinearity. The mass is connected to the plates using bolted joints with a uniform bolt torque with different magnitudes. Friction and the uneven contact area in the joints can also be considered as an additional source of nonlinearity. The structure is pulled and released in this study so that the free decay data can be measured. Acceleration data is recorded with a piezoelectric accelerometer (PCB M353B18). A displacement laser reading is used to check the accuracy of displacements derived from acceleration.

To study the uncertainties due to variability and repeatability different scenarios are considered. Table 2 shows the test scenarios and their characteristics. It includes the ID number of assembly, bolt pretension, and number of decaying response time series collected for each assembly. It should be noted that the structure was completely dismantled and reassembled for each assembly ID in Table 2. All bolts were uniformly tightened up to 70% of the ultimate value reported in Table 2 and then all of them tightened again up to the ultimate value.

A sample acceleration time series and its instantaneous characteristics belonging to assembly 1 are shown in Fig. 3. IA against IF are plotted in Fig. 3b to obtain the so-called frequency backbone curve. As expected for a bolted assembly, a softening behaviour is noticeable at low amplitudes, transitioning to a hardening behaviour as the vibration amplitude increases due to the geometric nonlinearity of the thin plates.

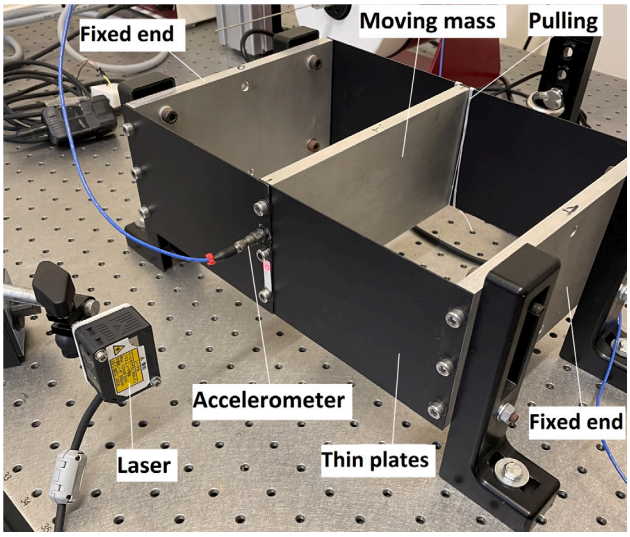

Fig. 2. Experimental setup of the vibrating mass used in this study.

Table 2
Re-assembly and test scenarios.

Assembly No.	Bolt pretension (N m)	Number of time series
1	4.5	8
2	4.5	8
3	4.5	15
4	2.5	15
5	6.5	15
6	10	8

Moreover, the effective damping ratio is calculated using the equation below according to [10]:

$$\zeta(t_i^o) = \frac{1}{2\pi IF(t_i^o)(t_{i+1}^o - t_{i-1}^o)}(ln(IA(t_{i-1}^o)) - ln(IA(t_{i+1}^o))) \quad (7)$$

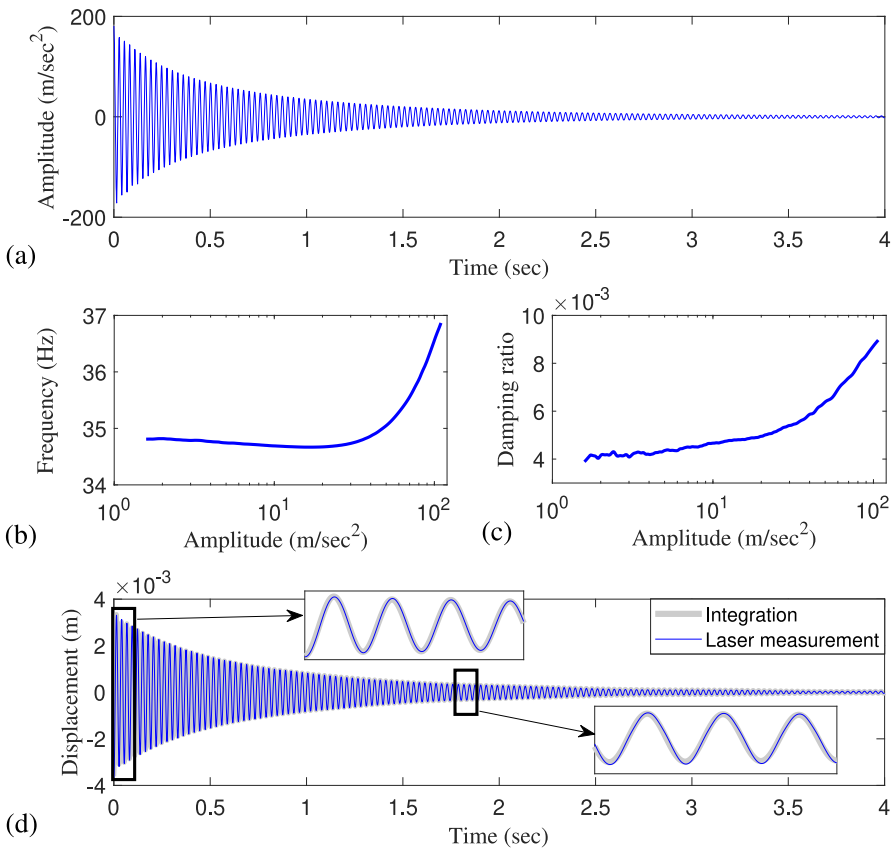

Fig. 3. A sample of experimental data (a) acceleration data, (b) amplitude-dependent frequency, (c) amplitude-dependent damping, and (d) displacement from integration against laser measurement.

where t_i^o is the time instances when IA and IF are recorded. The plot of the effective damping ratio against the IA in Fig. 3c shows an increase in damping with the increase in vibration amplitude which can be related to the friction in the bolted joints. Fig. 3d shows the displacement integrated from acceleration against the displacement measured using the laser. There is a 5% relative discrepancy observed between the displacements. It should be noted that the integration does not capture the tails (both ends of the time series) accurately and therefore, those parts are disregarded.

4.2. Uncertainity quantification

In practice, understanding the uncertainty associated with model parameters is crucial. This helps to estimate the variation in the dynamic response during vibration testing and quantify those variations using mathematical models. This work has developed and demonstrated a robust variant of the nonlinear system identification method for nonlinear assemblies based on the ensembling technique. Unlike the previous works [22] which use bootstrapping to produce multiple data sets from a single data set, this work uses authentic measured vibration data from multiple tests performed. Therefore, real-world uncertainties due to testing environment and assembly variations are considered.

Based on linear modal analysis results, the SDOF system considered for this structure has a mass of 0.6874 and the natural frequency and damping ratio $f_n = 34.95$ Hz, $\zeta = 0.0045$ respectively. These values are used as initial conditions to start the identification process.

When all data sets, including 69 time series are used for the identification, a total of 20 different terms are selected from the library. This is when the stopping criteria for model selection based on Section 2.2 are set to 10^{-3}. Fig. 4 shows the inclusion probability of selected terms. It can be observed that four terms have significant inclusion probabilities higher than 50% including terms: {27,2,8,25} from Table 1.

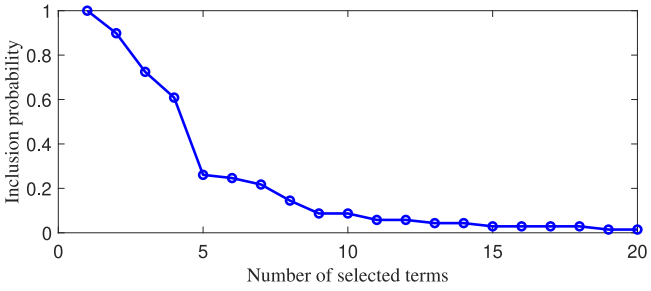

Fig. 4. Inclusion probability of selected terms considering all data sets for ensembling.

The probability of inclusion for the fifth term is almost 20% which is negligible.

In case the selected model includes 4 terms, the posterior distribution of the estimated parameters using all the time series is plotted in Fig. 5. At the stage of model selection, the parameters of the underlying linear system were fixed to the initial condition observed in the linear modal testing. An interesting observation is that the mean value of the coefficient for term 25 is almost zero. This suggests that the term could be also eliminated.

The uncertainty observed in the identified parameters in Fig. 5 is propagated into the dynamic response and is demonstrated based on the instantaneous characteristics of the response. Fig. 6a shows amplitude-dependent frequency and damping which are calculated based on the experimentally measured and estimated responses using the identified model. From the comparison, it can be seen that the estimation of the dynamic response is in good agreement with the experimental response. Nonetheless, there are some discrepancies with narrow confidence intervals in the low amplitude region, which

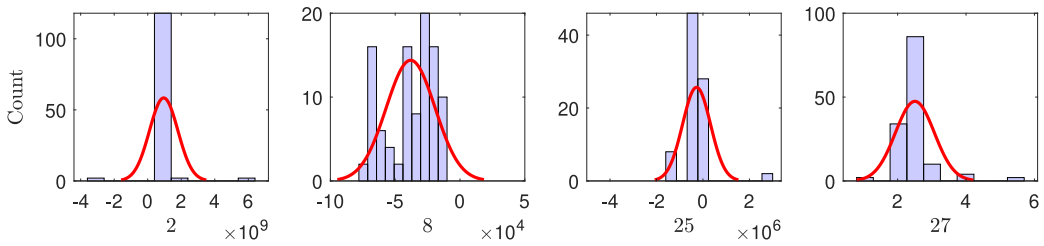

Fig. 5. Estimated posterior distribution of the selected model with 4 terms using the algebraic-based cost function.

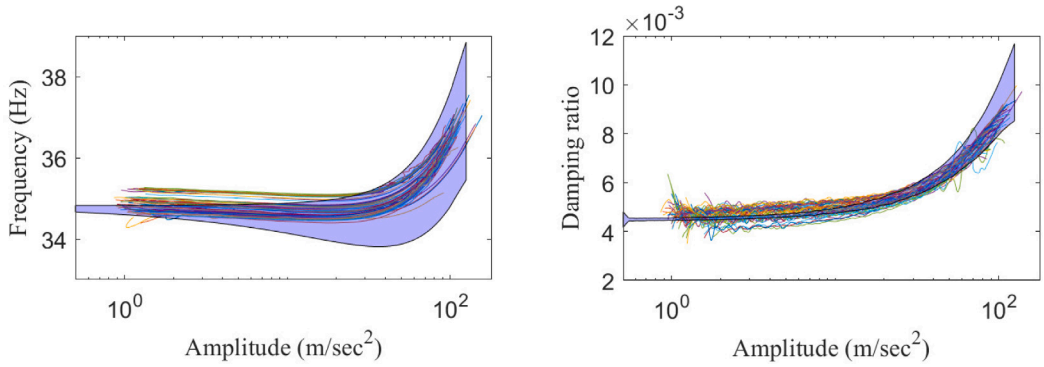

Fig. 6. Estimated uncertainty of the dynamic response using the selected model with 3 terms based on the algebraic-based cost function.

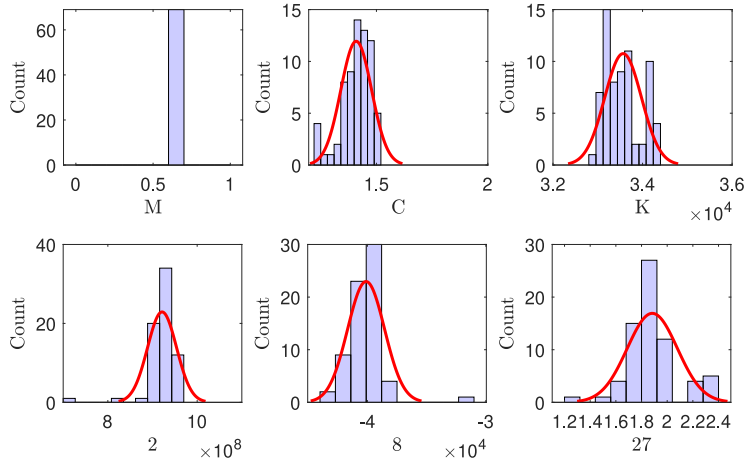

Fig. 7. Estimated posterior distribution of the underlying linear system and nonlinear parameters using the simulation-based cost function for all data sets.

are due to fixing values for the parameters of the underlying linear system when doing the identification using the algebraic-based cost function (Eq. (3)). This suggests that while the accuracy of estimating dynamic response and its uncertainty is acceptable, there is still room for improvement.

After model selection using the ensembling technique, parameter tuning is performed based on the simulation-based cost function introduced in Section 2.3. The prior or initial values for the parameters are considered based on the distributions in Fig. 5. For the parameters of the underlying linear system, i.e., damping ratio and natural frequency, a reasonable bound is considered so that their variation can be estimated at this stage. Fig. 7 shows the estimated uncertainties for the underlying linear system parameters as well as the parameters of the nonlinear model. A normal distribution model is fitted to the estimated parameters so that they can be used for sampling and simulating the probabilistic response of the system.

From the simulations using the parameter distributions in Fig. 7, the 99% confidence interval of the amplitude-dependent frequency and damping ratio responses is shown against the estimation from the experimentally measured data in Fig. 8. It can be observed that tuning the parameters using the simulation-based cost function increases the accuracy of response estimation and its uncertainty. Also, the experimental observation is well within the bounds of the confidence interval. The results show that the proposed approach in this paper is capable of modelling variability due to assembly conditions such as re-assembly or even different bolt pre-tensioning and repeatability during the vibration testing.

To better observe and quantify the experimental effects such as re-assembly and change of bolt pre-tension, the same practice is applied to each individual data set according to Table 2. Remarkably, the same nonlinear model with 3 terms is selected for all data subsets. The mean and standard deviation for the parameters of the identified nonlinear model are reported in Table 3 for each data subset. There is no meaningful pattern in the parameter values observed as expected for a single structure under test. However, the change in the dynamic response calculated using the identified models might be interesting to observe. For this purpose, the probabilistic dynamic responses of each mode in Table 3 are compared in Figs. 9, and 10.

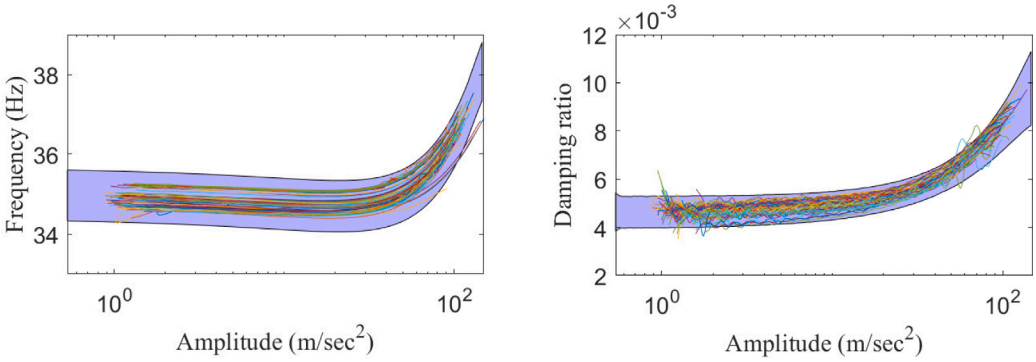

Fig. 8. Estimated uncertainty of the dynamic response using the selected model with 3 terms based on the simulation-based cost function.

Table 3
Estimated parameters of the identified probabilistic model for different data sets.

Assembly No.	Underlying linear system		Nonlinear model		
	$\mu_C(\sigma_C)$	$\mu_K(\sigma_K)$	$\mu_2(\sigma_2)$	$\mu_8(\sigma_8)$	$\mu_{27}(\sigma_{27})$
1	1.31	3.335×10^4	8.94×10^8	-4.266×10^4	2.082
	(0.0886)	(302.38)	(7.37×10^7)	(3.53×10^3)	(0.237)
2	1.44	3.336×10^4	9.54×10^8	-4.77×10^4	1.78
	(0.0166)	(174.26)	(2.11×10^7)	(715.26)	(0.093)
3	1.42	3.34×10^4	8.08×10^8	-2.21×10^4	1.95
	(0.0084)	(157.9)	(1.185×10^7)	(1060)	(0.06)
4	1.49	3.35×10^4	8.8×10^8	-3.56×10^4	1.88
	(0.017)	(222.42)	(1.32×10^7)	(1052)	(0.07)
5	1.32	3.4×10^4	8.5×10^8	-2.6×10^4	2.5
	(0.043)	(83.9)	(2.4×10^7)	(850)	(0.18)
6	1.27	3.3×10^4	8.12×10^8	-2.06×10^4	2.63
	(0.04)	(159.7)	(4.7×10^7)	(1290)	(0.11)

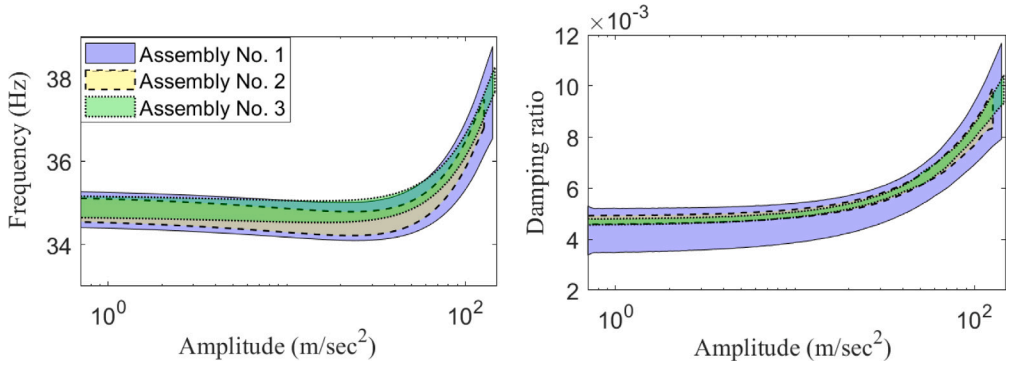

Fig. 9. Comparing the uncertainty of the dynamic response for three re-assemblies with bolt pre-tension 4.5 N.

Fig. 9 shows the propagated uncertainty with a 99% confidence interval for the data subsets with the assembly No. 1 to 3. Different amplitude-dependent response is observed for each data subset which means reassembly changes the dynamic characteristics of the structure under study. It can be observed that the uncertainty of assembly No. 1 is higher than No. 2 and 3. This observation is interesting since the data for subset assembly No. 1 is collected for a single assembly but in multiple days. Whereas, the data for the other two assemblies was collected after the assembly of the structure under a controlled condition for one hour. This highlights the importance of environmental effects such as temperature etc. as well as the routine of performing the experiment.

Fig. 10 shows the amplitude-dependent dynamic characteristics for three different bolt pre-tension during the assembly. A slight increase in the frequency can be observed for the assembly with higher bolt pre-tension. Also, it can observed that the assembly with low bolt pre-tension introduces higher uncertainty in the frequency response. However, there is no meaningful change observed in the amplitude-dependent damping due to the change in the bolt pre-tension.

5. Conclusion

This work has developed and demonstrated a robust variant of the data-driven method for structural identification of nonlinear assemblies based on ensembling technique. The proposed algorithm significantly improves the robustness and accuracy of model discovery. It is also a practical approach for quantifying the uncertainties during the model identification process for engineering structures from vibration data.

The identification method constructs the equation of motion for nonlinear dynamical systems from a single measured vibration data considering physically meaningful constraints. In the presence of multiple data sets, an ensemble of models is generated to aid in pinpointing the most likely model and quantifying uncertainties in its parameters. In other words, aggregate model statistics are used to generate inclusion probabilities of candidate functions from the ensemble of models, which promotes interpretability in model selection and provides probabilistic estimates.

Although this work is presented based on an SDOF system example, it has been shown in other works that it can be extended to systems

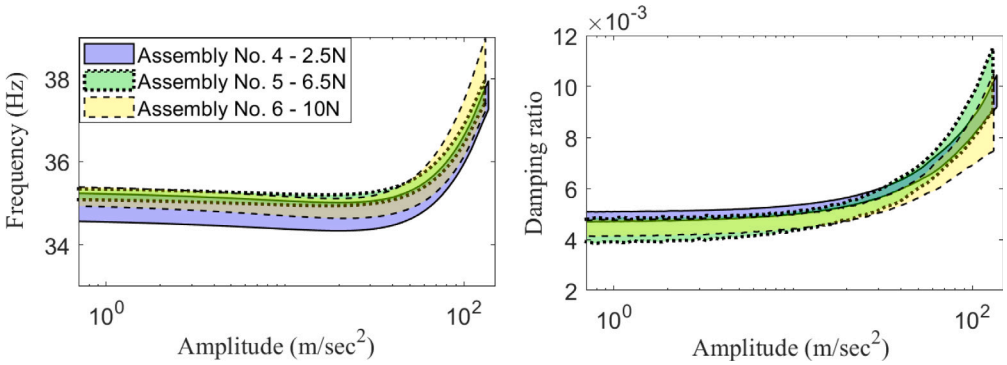

Fig. 10. Comparing the uncertainty of the dynamic response for three re-assemblies with different bolt pre-tension [2.5,6.5,10] N.

with multiple modal couplings and nonlinearities. However, it should be noted that systems with strong nonlinearities induce internal resonances that need special treatment in the data collection, processing, and identification phases which is a promising topic for future works. Besides, the identification of a nonlinear system employed in this paper that runs a nonlinear optimisation algorithm to check the inclusion of each candidate term in the library needs higher computational effort compared to alternative methods that use linear least square optimiser. Therefore, future research should be directed toward a joint algorithm that benefits from the speed of linear least square optimiser and the physics-informed nature of the identification method used in this paper that also allows libraries with nonlinear-in-parameter terms. It is hoped that speeding up the identification process along with the uncertainty quantification scheme proposed in this paper lead to a robust, accurate and online nonlinear model identification.

CRediT authorship contribution statement

Sina Safari: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Diogo Montalvão:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Investigation. **Julián M. Londoño Monsalve:** Writing – review & editing, Supervision, Resources, Methodology, Investigation, Conceptualization.

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.

Acknowledgments

Mr Safari is support by the scholarship from the College of Engineering, Mathematics, and Physical Sciences, University of Exeter which is gratefully acknowledged.

Data availability

Data will be made available on request.

References

[1] J. Noël, G. Kerschen, Nonlinear system identification in structural dynamics: 10 more years of progress, Mech. Syst. Signal Process. 83 (2017) 2–35, http://dx.doi.org/10.1016/j.ymssp.2016.07.020.

[2] L. Renson, A. Gonzalez-Buelga, D. Barton, S. Neild, Robust identification of backbone curves using control-based continuation, J. Sound Vib. 367 (2016) 145–158, http://dx.doi.org/10.1016/j.jsv.2015.12.035.

[3] T. Karaağaçlı, H.N. Özgüven, Experimental modal analysis of nonlinear systems by using response-controlled stepped-sine testing, Mech. Syst. Signal Process. 146 (2021) 107023.

[4] J. Noël, S. Marchesiello, G. Kerschen, Subspace-based identification of a nonlinear spacecraft in the time and frequency domains, Mech. Syst. Signal Process. 43 (2014) 217–236, http://dx.doi.org/10.1016/j.ymssp.2013.10.016.

[5] Z. Lai, C. Mylonas, S. Nagarajaiah, E. Chatzi, Structural identification with physics-informed neural ordinary differential equations, J. Sound Vib. 508 (2021) 116196.

[6] A.B. Abdessalem, N. Dervilis, D. Wagg, K. Worden, Model selection and parameter estimation in structural dynamics using approximate Bayesian computation, Mech. Syst. Signal Process. 99 (2018) 306–325.

[7] D. Shetty, M. Allen, K. Park, A new approach to model a system with both friction and geometric nonlinearity, J. Sound Vib. 552 (2023) 117631.

[8] M. Brake, C. Schwingshackl, P. Reuß, Observations of variability and repeatability in jointed structures, Mech. Syst. Signal Process. 129 (2019) 282–307.

[9] L. Tan, W. Zhang, Z. Wang, B. Hou, W. Sun, Variation in the nonlinear stiffness of bolted joints due to tangential hysteresis behavior, Int. J. Non-Linear Mech. 158 (2024) 104577.

[10] J.M. Londoño, S.A. Neild, J.E. Cooper, Identification of backbone curves of nonlinear systems from resonance decay responses, J. Sound Vib. 348 (2015) 224–238.

[11] S. Billings, Nonlinear System Identification: NARMAX Methods in the Time, Frequency and Spatio-Temporal Domains, Wiley, 2013.

[12] J. Paduart, L. Lauwers, J. Swevers, K. Smolders, J. Schoukens, R. Pintelon, Identification of nonlinear systems using polynomial nonlinear state space models, Automatica 46 (4) (2010) 647–657, http://dx.doi.org/10.1016/j.automatica.2010.01.001.

[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] S. Safari, J.M.L. Monsalve, Direct optimisation based model selection and parameter estimation using time-domain data for identifying localised nonlinearities, J. Sound Vib. 501 (2021) 116056.

[15] H.P. Chintha, A. Chatterjee, Identification and parameter estimation of non-polynomial forms of damping nonlinearity in dynamic systems, Int. J. Non-Linear Mech. 143 (2022) 104017.

[16] S. Safari, J.L. Monsalve, Data-driven structural identification of nonlinear assemblies: Structures with bolted joints, Mech. Syst. Signal Process. 195 (2023) 110296.

[17] S. Safari, J.L. Monsalve, Data-driven structural identification of nonlinear assemblies: asymmetric stiffness and damping nonlinearities, 2024, Under review.

[18] M. Song, L. Renson, B. Moaveni, G. Kerschen, Bayesian model updating and class selection of a wing-engine structure with nonlinear connections using nonlinear normal modes, Mech. Syst. Signal Process. 165 (2022) 108337.

[19] T. Chatterjee, A.D. Shaw, M.I. Friswell, H.H. Khodaparast, Sparse Bayesian machine learning for the interpretable identification of nonlinear structural dynamics: Towards the experimental data-driven discovery of a quasi zero stiffness device, Mech. Syst. Signal Process. 205 (2023) 110858.

[20] R. Nayek, A. Abdessalem, N. Dervilis, E. Cross, K. Worden, Identification of piecewise-linear mechanical oscillators via Bayesian model selection and parameter estimation, Mech. Syst. Signal Process. 196 (2023) 110300.

[21] L. Mars Gao, U. Fasel, S.L. Brunton, J.N. Kutz, Convergence of uncertainty estimates in ensemble and Bayesian sparse model discovery, 2023, arXiv e-prints, arXiv–2301.

[22] U. Fasel, J.N. Kutz, B.W. Brunton, S.L. Brunton, Ensemble-SINDy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control, Proc. R. Soc. A 478 (2260) (2022) 20210904.

[23] S. Safari, J.L. Monsalve, Benchmarking optimisation methods for model selection and parameter estimation of nonlinear systems, Vibration 4 (3) (2021) 648–665.
[24] B. Peeters, H. Van der Auweraer, P. Guillaume, J. Leuridan, The PolyMAX frequency-domain method: a new standard for modal parameter estimation? Shock Vib. 11 (3–4) (2004) 395–409.
[25] M. Géradin, D.J. Rixen, Mechanical Vibrations: Theory and Application to Structural Dynamics, John Wiley & Sons, 2014.
[26] J. Tezcan, C.C. Marin-Artieda, Least-square-support-vector-machine-based approach to obtain displacement from measured acceleration, Adv. Eng. Softw. 115 (2018) 357–362.
[27] R. Isermann, M. Münchhof, Identification of Dynamic Systems: an Introduction with Applications, vol. 85, Springer, 2011.
[28] J.B.W. Webber, A bi-symmetric log transformation for wide-range data, Meas. Sci. Technol. 24 (2) (2012) 027001.