
Article:
Zhang, B. orcid.org/0000-0001-7327-0923, Lang, Z.Q., Billings, S.A. et al. (2 more authors) (2013) System identification methods for metal rubber devices. Mechanical Systems and Signal Processing, 39 (1-2). pp. 207-226. ISSN 0888-3270

https://doi.org/10.1016/j.ymssp.2013.02.002

System Identification Methods for Metal Rubber Devices

B. Zhang[a], Z.Q. Lang[a], S. A. Billings[a, *], G. R. Tomlinson[b], and J. A. Rongong[b]

[a]*Department of Automatic Control and Systems Engineering, The University of Sheffield, Mappin Street, Sheffield S1 3JD, UK*
[b]*Department of Mechanical Engineering, The University of Sheffield, Mappin Street, Sheffield S1 3JD, UK*
E-mail: {b.zhang, z.lang, s.billings, g.tomlinson, j.a.rongong}@sheffield.ac.uk
**Corresponding author*

Abstract

Metal rubber (MR) devices, a new wire mesh material, have been extensively used in recent years due to several unique properties especially in adverse environments. Although many practical studies have been completed, the related theoretical research on metal rubber is still in its infancy. In this paper, a semi-constitutive dynamic model that involves nonlinear elastic stiffness, nonlinear viscous damping and bilinear hysteresis Coulomb damping is adopted to model MR devices. The model is first approximated by representing the bilinear hysteresis damping as Chebyshev polynomials of the first kind and then generalised by taking into account the effects of noises. A very efficient systematic procedure based on the orthogonal least squares (OLS) algorithm, the adjustable prediction error sum of squares (APRESS) criterion and the nonlinear model validity tests is proposed for model structure detection and parameter estimation of MR devices for the first time. The OLS algorithm provides a powerful tool to effectively select the significant model terms step by step, one at a time, by orthogonalising the associated terms and maximizing the error reduction ratio, in a forward stepwise manner. The APRESS statistic regularizes the OLS algorithm to facilitate the determination of the optimal number of model terms that should be included into the model. And whether the final identified dynamic model is adequate and acceptable is determined by the model validity tests. Because of the orthogonal property of the OLS algorithm, the selection of the dynamic model terms and noise model terms are totally decoupled and the approach also leads to a parsimonious model. Numerical ill-conditioning problems which can arise in the conventional least squares algorithm can be avoided as well. The methods of choosing the sampling interval for nonlinear systems are also incorporated into the approach. Finally by utilising the response of a cylindrical MR specimen, it is shown how the model structure can be detected in a practical application.

Keywords: metal rubber; nonlinear damping material; nonlinear system identification; bilinear hysteresis model

1. Introduction

Metal rubber (MR) devices[1], also called wire mesh dampers[2-6], are a new type of material made by coiling thin metal wires with circular or noncircular sections into elastic spirals, which are then stretched, woven, and cold-pressed into various shapes. Subsequent handling procedures such as tempering and vibration stabilization may also be undertaken. The MR element produced by this process has both the rubber-like elasticity and porosity formed by the contacts between adjacent wires. When excited by external forces, the wires will deform, slide and extrude, resulting in vibration energy dissipation[7]. Since the material can be made of special steel wires and can be

processed by some special technologies[1], it can give not only good elastic and high damping capacity but also several other useful characteristics not normally associated with rubber. These include: resistance to corrosion, survival in extreme temperatures, low temperature sensitivity of the stiffness and damping, negligible effects of ageing, non-volatility in vacuum and ability to withstand radiation. Because of these unique and attractive properties, MR devices have already been used in health care and the aerospace industry, to the author's knowledge, since the 1950s, usually in extremely adverse environments, to reduce noise, isolate vibrations, absorb shocks, and even as seals, heat pipe linings, filters, throttle valves and bearing bushes[1, 4, 8-10]. However, despite all these applications, the related theoretical analysis is still in its infancy.

Experiments have showed that the mechanical properties of MR devices, such as the energy dissipation and microscopic deformation mechanisms, depend on a number of factors including the mass, shape, diameter of the spirals, orientational angle of the spirals in the compression direction, diameter of the wires, material of the wires, forming pressure, temperature, and preload[11-13]. This makes the constitutive model of this material, which is usually at a microscopic scale, very difficult to construct. Very few publications on the constitutive model are available. The models already proposed are the model of a bar system[1], the pyramidal model[1], the model of micro spring elements[14], and the model of porous materials[15, 16]. Although these models can reveal some properties of MR devices, all of them are constructed under some assumptions such as static load and have difficulties to be used in real applications.

On the other hand, the dynamic stress-stain relation of MR devices is usually represented in the form of hysteresis loops and thus can be modelled on a macroscopic scale. In the past several decades, much research effort has been made to model and analyse hysteresis phenomena. Mathematical hysteresis is normally defined as a rate independent memory effect, that is, the hysteresis loops are stable with respect to arbitrary changes of the time scale[17]. However, in reality, hysteretic effects are rarely rate independent since hysteresis is coupled with viscous-type effects. Therefore, both rate independent and rate dependent models have been proposed[18]. Of these, the Duhem type models and the Preisach type models are the most used. Variations of these two model types have been studied in different contexts under various names. The ferromagnetic material model, the Bouc-Wen model, the Madelung model, the Dahl friction model, the LuGre friction model, the Maxwell-slip model and the presliding friction model are specialized Duhem models while the Masing model and the Iwan model are special cases of Preisach models. The LuGre friction model is rate dependent while other models are rate independent[19]. More rate dependent models can be found in [20] and [21]. Despite so many models being available, it is surprising to find that very few have been applied to the modelling of MR devices. One application was made by Ulanov and Lazutkin[22], who used Masing's principle and the coordinate transformation and proposed a method to obtain a description of the hysteresis loops in a loading process of MR devices. To obtain a more accurate model in a multi-axial loading process, the influence of the loading history in one axis on that in another was also investigated[23]. Compared with the constitutive models, the hysteretic models do not come, in general, from a detailed analysis of the physical mechanisms but just concentrate on describing the shape of the hysteresis loops. That is to say, they are phenomenological models. But one important aspect of a model is how meaningful the model parameters are. Physically meaningful parameters can give information about the real properties of a system which are important for analysis and design.

Experimental results have demonstrated that the dynamic characteristic of MR devices exhibits nonlinear behaviour which is also dependent on the input frequency[24]. This indicates that the

materials damping also consists of a viscous damping component in addition to the dry friction damping between the wires. Therefore, in this paper, a model that involves nonlinear elastic stiffness, nonlinear viscous damping and bilinear hysteresis Coulomb damping is adopted. This model can not only describe the dynamic restoring force at the macro level but also has a basis on the microstructure of the sliding surfaces between the wires although it is not constructed directly on a microscopic scale. In this sense, it is a semi-constitutive model. In previous studies, an odd ordered polynomial function was used to describe the elastic stiffness and viscous damping characteristics. But how many terms should be included in the model has not been explored. Some researchers used a cubic polynomial[25-27] while others suggest that a quintic polynomial might be needed[28]. In this paper, a new approach based on the Chebyshev polynomial approximation, the orthogonal least squares (OLS) algorithm regularised by an adjustable prediction error sum of squares (APRESS) criterion, and the nonlinear model validity tests will be developed to detect the model structure and then estimate the parameters of the model to provide, for the first time, a systematic procedure for the identification of nonlinear dynamic models of MR devices.

2. Dynamic model

A cylindrical MR specimen is shown in Fig. 1. The model illustrated in Fig. 2 is used to represent the MR device, which consists of a nonlinear elastic spring $f_k(\cdot)$ in parallel with a nonlinear viscous damper $f_c(\cdot)$ and a hysteretic Coulomb damper z. The nonlinear spring and nonlinear damper are only relevant to the current deformation while the hysteretic Coulomb damper has memory characteristics and also depends on the deformation history. The elastic stiffness restoring force is described by an odd ordered polynomial function of current displacement with the highest degree N_1 while the viscous damping restoring force is described by an odd ordered polynomial function of current velocity with the highest degree N_2, such that

$$f_k(\cdot) = \sum_{n=1}^{N_1} k_{2n-1}(\cdot)^{2n-1} \tag{1}$$

$$f_c(\cdot) = \sum_{n=1}^{N_2} c_{2n-1}(\cdot)^{2n-1} \tag{2}$$

where k_{2i-1} $(i = 1, \dots, N_1)$, c_{2i-1} $(i = 1, \dots, N_2)$ are the parameters of the stiffness characteristic and damping characteristic respectively. The MR device is subject to a preload F_0 and a harmonic excitation force of amplitude F_m and frequency Ω. Set the equilibrium position of the MR device under preload as the origin of the displacement. Then the equation of motion for the MR device can be written as

$$\sum_{n=1}^{N_1} k_{2n-1}\left(y(t)+y_0\right)^{2n-1} + \sum_{n=1}^{N_2} c_{2n-1}\dot{y}(t)^{2n-1} + z(t) = F_m\cos(\Omega t) + F_0 \tag{3}$$

where y_0 is the static displacement produced by the preload and $z(t)$ is the bilinear hysteresis restoring force produced by the hysteretic Coulomb damper z.

For convenience of analysis, denote

$$F(t) = F_m\cos(\Omega t) + F_0 \tag{4}$$

Substituting Eq.(4) into Eq.(3) yields

$$\sum_{n=1}^{N_1} k_{2n-1}\left(y(t)+y_0\right)^{2n-1} + \sum_{n=1}^{N_2} c_{2n-1}\dot{y}(t)^{2n-1} + z(t) = F(t) \tag{5}$$

As shown in Fig. 2, the hysteretic Coulomb damper is composed of a hysteretic Coulomb friction model with a serial linear spring, the characteristic of which is described by a bilinear hysteresis model[29, 30] shown in Fig. 3. The incremental representation of this bilinear hysteresis model can be expressed as

$$\mathrm{d}z(t)=\frac{k_s}{2}\left[1+\operatorname{sgn}\left(z_s-\left|z(t)\right|\right)\right]\mathrm{d}y(t) \tag{6}$$

$$k_s=\frac{z_s}{y_s} \tag{7}$$

where k_s is the stiffness of the linear spring, z_s the memorized restoring force when sliding between wires occurs, y_s the elastic deformation limit, and the sign function here is defined as

$$\operatorname{sgn}x=\begin{cases}1 & x>0\\ -1 & x\le 0\end{cases} \tag{8}$$

In Fig. 3, the states from the initial state $0'$ to state 2 represent the static response of the MR device under preload while the states thereafter describe the transient response of the MR device subject to a harmonic excitation. The asymmetric steady state response of the MR device is shown in Fig. 4. In Fig. 3, from state $0'$ to the critical sliding state 1, the restoring force $z(t)$ increases and the excitation energy is stored by the bilinear hysteresis damper. Then from state 1 to state 2, the energy is dissipated by sliding friction and $z(t)$ remains constant. In the subsequent cycles, the energy flows as follows: dissipation (2→3), release (3→4), storage (4→5), dissipation (5→6), release (6→7), storage (7→8), dissipation (8→9), release (9→10), and so on. The energy flow of the steady state response in Fig. 4 repeats the following process: storage (1→2), dissipation (2→3), release (3→4), storage (4→5), dissipation (5→6), and release (6→1). In Fig. 4, y_m is the maximum displacement of the steady state response while $y(t_m)$, $y(t_{m+1})$, etc. in Fig. 3 are the peak displacements of the transient response.

Although Eq. (6) combined with Eq. (5) clearly describes the deformation and energy dissipation mechanism of MR devices, it is not easy to use these expressions to identify the dynamic model. However, by using some series expansions[31, 32], Eq. (6) can be written in a series form and then can be easily used for the model identification. Of these series expansions, Chebyshev polynomials can give a high accuracy by using the least terms and thus is adopted in this paper to approximate the bilinear hysteresis relation.

3. Chebyshev polynomial approximations

Suppose that $y(t_m) > y_s$, where $m \geq M$, m and M are positive integers, and $y(t_m)$ is one of the peak displacements of the MR device's response. That is to say, sliding between the wires happens in each cycle after time t_M. This condition can easily be satisfied by utilising a large amplitude excitation and these cycles will then be collected for the model identification. Without loss of generality, suppose that $M = 1$. The maximum amplitude span is defined as

$$\Delta y=\max_m\left|y\left(t_{m+1}\right)-y\left(t_m\right)\right| \tag{9}$$

3.1. For the condition when $\dot{y}(t) \leq 0$

The bilinear hysteresis restoring force $z(t)$ of one branch with negative velocity, e.g. 3→4→5→6, after continuation can be expressed as

$$z\left(y\left(t\right)\right)=\begin{cases} z_s+k_s\left[y\left(t\right)-y\left(t_m\right)\right], & y\left(t_m\right)-2y_s\le y\left(t\right)\le y\left(t_m\right) \\ -z_s, & y\left(t_m\right)-\Delta y\le y\left(t\right)< y\left(t_m\right)-2y_s \end{cases} \tag{10}$$

where $m=1,3,5,\ \cdots$.

Define

$$\tilde{y}\left(t\right)=\frac{2\left[y\left(t\right)-y\left(t_m\right)\right]}{\Delta y}+1, \qquad y\left(t_m\right)-\Delta y\le y\left(t\right)\le y\left(t_m\right) \tag{11}$$

and then Eq.(10) can be written as

$$z\left(\tilde{y}\left(t\right)\right)=\begin{cases} \dfrac{k_s\Delta y}{2}\left[\tilde{y}\left(t\right)-1+\dfrac{2y_s}{\Delta y}\right], & 1-\dfrac{4y_s}{\Delta y}\le\tilde{y}\left(t\right)\le 1 \\ -k_s y_s, & -1\le\tilde{y}\left(t\right)<1-\dfrac{4y_s}{\Delta y} \end{cases} \tag{12}$$

Now $z\left(\tilde{y}(t)\right)$ becomes a continuous function in $\left[-1,\ 1\right]$ and thus can be approximated by the Chebyshev polynomials of the first kind as [33]

$$z\left(t\right)=\sum_{n=0}^{N_3}{}'a_n\cos\left[n\theta\left(t\right)\right] \tag{13}$$

where the primed summation indicates only half of the first term is included, N_3 is the maximum degree of the truncated Chebyshev polynomials, a_n is the Chebyshev coefficient given by

$$a_n=\frac{2}{\pi}\int_0^{\pi}z\left[\theta\left(t\right)\right]\cos\left[n\theta\left(t\right)\right]d\theta\left(t\right) \tag{14}$$

and

$$\theta\left(t\right)=\arccos\left[\tilde{y}\left(t\right)\right] \tag{15}$$

Denote

$$\varphi=\arccos\left(\frac{4y_s}{\Delta y}-1\right) \tag{16}$$

Considering Eqs.(15) and (16), Eq.(12) now takes the form

$$z\left(\theta\left(t\right)\right)=\begin{cases} \dfrac{k_s\Delta y}{2}\left\{\cos\left[\theta\left(t\right)\right]-\sin^2\left(\dfrac{\varphi}{2}\right)\right\}, & 0\le\theta\left(t\right)\le\pi-\varphi \\ -k_s y_s, & \pi-\varphi<\theta\left(t\right)\le\pi \end{cases} \tag{17}$$

Then substituting Eq.(17) into Eq.(14) yields

$$a_n=\begin{cases}-\dfrac{k_s\Delta y}{\pi}\left(\varphi\cos\varphi-\sin\varphi+\pi\sin^2\dfrac{\varphi}{2}\right), & n=0\\ \dfrac{k_s\Delta y}{2\pi}\left[\pi-\varphi+\dfrac{1}{2}\sin(2\varphi)\right], & n=1\\ (-1)^{n+1}\dfrac{k_s\Delta y}{\pi n\left(n^2-1\right)}\left[n\cos(n\varphi)\sin\varphi-\sin(n\varphi)\cos\varphi\right], & n=2,3,\cdots,N_3\end{cases} \tag{18}$$

The Chebyshev polynomial approximations of $z(t)$ can be obtained by substituting Eqs.(11), (15), and (18) into Eq.(13) as

$$z(t)=\frac{\overline{a}_0}{2}\operatorname{sgn}\left[\dot{y}(t)\right]+\sum_{n=1}^{N_3}\overline{a}_n\operatorname{sgn}^{n+1}\left[\dot{y}(t)\right]\cos\left\{n\arccos\left[\frac{2y(t)-2y(t_m)}{\Delta y}+1\right]\right\} \tag{19}$$

where $t\in[t_m,\ t_{m+1}]$,

$$\overline{a}_n=\begin{cases}\dfrac{k_s\Delta y}{\pi}\left(\varphi\cos\varphi-\sin\varphi+\pi\sin^2\dfrac{\varphi}{2}\right), & n=0\\ \dfrac{k_s\Delta y}{2\pi}\left[\pi-\varphi+\dfrac{1}{2}\sin(2\varphi)\right], & n=1\\ \dfrac{k_s\Delta y}{\pi n\left(n^2-1\right)}\left[n\cos(n\varphi)\sin\varphi-\sin(n\varphi)\cos\varphi\right], & n=2,3,\cdots,N_3\end{cases} \tag{20}$$

and Δy, φ are given by Eqs.(9) and (16) respectively.

Notice that benefiting from the continuation of the displacement to the maximum amplitude span, a uniform expression for $z(t)$, which is convenient for the dynamic model identification, has been arrived at for all of the branches with negative velocity.

3.2. For the condition when $\dot{y}(t)>0$

Similar to the branch with negative velocity, the restoring force $z(t)$ of one branch with positive velocity, e.g. 6→7→8→9, after continuation can be expressed as

$$z\left(y(t)\right)=\begin{cases}-z_s+k_s\left[y(t)-y(t_m)\right], & y(t_m)\le y(t)\le y(t_m)+2y_s\\ z_s, & y(t_m)+2y_s<y(t)\le y(t_m)+\Delta y\end{cases} \tag{21}$$

where $m=2,4,6,\ \cdots$, and Δy is given by Eq.(9).

Define

$$\tilde{y}(t)=\frac{2\left[y(t)-y(t_m)\right]}{\Delta y}-1,\qquad y(t_m)\le y(t)\le y(t_m)+\Delta y \tag{22}$$

and then Eq.(21) can be rewritten as

$$z\left(\tilde{y}\left(t\right)\right)=\begin{cases}\dfrac{k_s\Delta y}{2}\left[\tilde{y}\left(t\right)+1-\dfrac{2y_s}{\Delta y}\right], & -1\le\tilde{y}\left(t\right)\le\dfrac{4y_s}{\Delta y}-1\\ k_s y_s, & \dfrac{4y_s}{\Delta y}-1<\tilde{y}\left(t\right)\le 1\end{cases} \tag{23}$$

Apparently the expression for the transformed displacement $\tilde{y}(t)$ in Eq.(22) is different from that in Eq.(11). This has been confused in [32].

Considering Eqs.(15) and (16), Eq.(23) is expressed as

$$z\left(\theta\left(t\right)\right)=\begin{cases}\dfrac{k_s\Delta y}{2}\left\{\cos\left[\theta\left(t\right)\right]+\sin^2\left(\dfrac{\varphi}{2}\right)\right\}, & \varphi\le\theta\left(t\right)\le\pi\\ -k_s y_s, & 0\le\theta\left(t\right)<\varphi\end{cases} \tag{24}$$

Then substituting Eq.(24) into Eq.(14) gives

$$a_n=\begin{cases}\dfrac{k_s\Delta y}{\pi}\left(\varphi\cos\varphi-\sin\varphi+\pi\sin^2\dfrac{\varphi}{2}\right), & n=0\\ \dfrac{k_s\Delta y}{2\pi}\left[\pi-\varphi+\dfrac{1}{2}\sin\left(2\varphi\right)\right], & n=1\\ \dfrac{k_s\Delta y}{\pi n\left(n^2-1\right)}\left[n\cos\left(n\varphi\right)\sin\varphi-\sin\left(n\varphi\right)\cos\varphi\right], & n=2,3,\cdots,N_3\end{cases} \tag{25}$$

Substituting Eqs. (22), (15), and (25) into Eq.(13) yields

$$z\left(t\right)=\frac{a_0}{2}\operatorname{sgn}\left[\dot{y}\left(t\right)\right]+\sum_{n=1}^{N_3}a_n\operatorname{sgn}^{n+1}\left[\dot{y}\left(t\right)\right]\cos\left\{n\arccos\left[\frac{2y\left(t\right)-2y\left(t_m\right)}{\Delta y}-1\right]\right\} \tag{26}$$

where $t\in\left[t_m,\ t_{m+1}\right]$, a_n, Δy, and φ are given by Eqs.(25), (9) and (16) respectively.

Notice that the Chebyshev coefficient $\bar{a}_n$ given by Eq.(20) and a_n given by Eq.(25) are in the same form. Eqs.(19) and (26) can thus be combined in a uniform expression,

$$z\left(t\right)=\frac{a_0}{2}\operatorname{sgn}\left[\dot{y}\left(t\right)\right]+\sum_{n=1}^{N_3}a_n\operatorname{sgn}^{n+1}\left[\dot{y}\left(t\right)\right]\cos\left\{n\arccos\left[\frac{2y\left(t\right)-2y\left(t_m\right)}{\Delta y}-\operatorname{sgn}\left[\dot{y}\left(t\right)\right]\right]\right\} \tag{27}$$

where $t\in\left[t_m,\ t_{m+1}\right]$, a_n, Δy, and φ are given by Eqs.(25), (9) and (16) respectively.

Substituting Eq.(27) into Eq.(5) gives

$$\begin{aligned}&\sum_{n=1}^{N_1}k_{2n-1}\left(y\left(t\right)+y_0\right)^{2n-1}+\sum_{n=1}^{N_2}c_{2n-1}\dot{y}\left(t\right)^{2n-1}+\frac{a_0}{2}\operatorname{sgn}\left[\dot{y}\left(t\right)\right]\\&+\sum_{n=1}^{N_3}a_n\operatorname{sgn}^{n+1}\left[\dot{y}\left(t\right)\right]\cos\left\{n\arccos\left[\frac{2y\left(t\right)-2y\left(t_m\right)}{\Delta y}-\operatorname{sgn}\left[\dot{y}\left(t\right)\right]\right]\right\}=F\left(t\right)\end{aligned} \tag{28}$$

Eq.(28) shows a relation between the displacement $y(t)$, the velocity $\dot{y}(t)$, and the excitation force $F(t)$. Therefore, if the displacement and the corresponding velocity under various excitation forces

are measured and the higher-order terms are neglected, a model of the MR device can be estimated. For example, if $y(t)$, $\dot{y}(t)$, and $F(t)$ can be measured at N different time instants t_i, $i = 1,2,\cdots,N$, and if the first $\bar{N}_1$ terms are considered for the elastic stiffness force, $\bar{N}_2$ terms for the viscous damping force, $\bar{N}_3 + 1$ terms for the bilinear hysteresis damping force on the left-hand side of Eq.(28), the following equation can be derived,

$$F(t_i) = \sum_{j=1}^{\bar{N}} \theta_j \phi_j(t_i) + \varepsilon_i \tag{29}$$

where

$$\bar{N} = \bar{N}_1 + \bar{N}_2 + \bar{N}_3 + 1 \tag{30}$$

$$\theta_j = \begin{cases} k_{2j-1}, & j = 1,2,\cdots,\bar{N}_1 \\ c_{2(j-\bar{N}_1)-1}, & j = \bar{N}_1+1,\cdots,\bar{N}_1+\bar{N}_2 \\ a_{j-(\bar{N}_1+\bar{N}_2+1)}, & j = \bar{N}_1+\bar{N}_2+1,\cdots,\bar{N} \end{cases} \tag{31}$$

$$\phi_j(t_i) = \begin{cases} \left(y(t_i)+y_0\right)^{2j-1}, & j = 1,2,\cdots,\bar{N}_1 \\ \dot{y}(t_i)^{2j-1}, & j = \bar{N}_1+1,\cdots,\bar{N}_1+\bar{N}_2 \\ 0.5\,\mathrm{sgn}\left[\dot{y}(t_i)\right], & j = \bar{N}_1+\bar{N}_2+1 \\ \mathrm{sgn}^{j+1}\left[\dot{y}(t_i)\right]\cos\left\{ j\arccos\left[\dfrac{2y(t_i)-2y(t_m)}{\Delta y} - \mathrm{sgn}\left[\dot{y}(t_i)\right]\right]\right\}, & j = \bar{N}_1+\bar{N}_2+2,\cdots,\bar{N} \end{cases} \tag{32}$$

and ε_i, $i = 1,2,\cdots,N$, is the model residual.

Eq.(29) can also be written in the matrix form as

$$\mathbf{F} = \mathbf{\Phi\Theta} + \mathbf{\Xi} \tag{33}$$

where

$$\mathbf{F} = \left[F(t_1),\ F(t_2),\ \cdots,\ F(t_N)\right]^T \tag{34}$$

$$\mathbf{\Phi} = \left[\boldsymbol{\varphi}_1,\ \boldsymbol{\varphi}_2,\ \cdots,\ \boldsymbol{\varphi}_{\bar{N}}\right] \tag{35}$$

$$\mathbf{\Theta} = \left[\theta_1,\ \cdots,\ \theta_{\bar{N}}\right]^T = \left[k_1,\ \cdots,\ k_{2\bar{N}_1-1},\ c_1,\ \cdots,\ c_{2\bar{N}_2-1},\ a_0,\ a_1,\ \cdots,\ a_{\bar{N}_3}\right]^T \tag{36}$$

$$\mathbf{\Xi} = \left[\varepsilon_1,\ \cdots,\ \varepsilon_N\right]^T \tag{37}$$

The residual vector $\mathbf{\Xi}$ is assumed to be of zero mean and uncorrelated with $\boldsymbol{\varphi}_j$, $j = 1,2,\cdots,\bar{N}$ and

$$\boldsymbol{\varphi}_j = \left[\phi_j(t_1),\ \cdots,\ \phi_j(t_N)\right]^T \tag{38}$$

where $\phi_j(t_i)$, $j = 1,2,\cdots,\bar{N}$, $i = 1,2,\cdots,N$, is given by Eq.(32).

After the identification of the dynamic model, the parameters of the bilinear hysteresis model can be obtained from Eqs.(25), (16) and (7),

$$\begin{cases} \dfrac{\pi-\varphi+\dfrac{1}{2}\sin(2\varphi)}{\varphi\cos\varphi-\sin\varphi+\pi\sin^2\dfrac{\varphi}{2}}=\dfrac{2\hat{a}_1}{\hat{a}_0}; \\ \hat{y}_s=\dfrac{\Delta y}{4}(1+\cos\varphi); \\ \hat{k}_s=\dfrac{\pi\hat{a}_0}{\Delta y}\left(\varphi\cos\varphi-\sin\varphi+\pi\sin^2\dfrac{\varphi}{2}\right)^{-1}; \\ \hat{z}_s=\hat{k}_s\hat{y}_s. \end{cases} \tag{39}$$

It should be noted that although Eq.(33) holds for both the transient response and the steady state response, a preload is necessary in the steady state case so that the columns of the matrix $\mathbf{\Phi}$ are linearly independent while in the transient case, it is not restricted by this.

The solution of Eq.(33) can be obtained by using the least squares (LS) algorithm as

$$\mathbf{\Theta}=\left(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{F} \tag{40}$$

But in practice, the information matrix $\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}$ is often ill-conditioned. Researchers[34, 35] have indicated that when ill-conditioning is present, the parameter estimation based on the LS approach tends to be biased. In addition, the LS approach needs to make an assumption that the excitation force $F(t)$ in Eq.(28) can be represented by $\overline{N}_1+\overline{N}_2+\overline{N}_3+1$ terms while $\overline{N}_1$, $\overline{N}_2$, and $\overline{N}_3$ are all sufficiently large numbers. However, many of these candidate model terms may be redundant. The inclusion of redundant model terms often makes the model become oversensitive to the training data and is likely to exhibit poor generalisation properties. To overcome these problems, the orthogonal least squares (OLS) algorithm can be used. The OLS method provides a powerful tool to select the significant model terms, determine the optimal number of model terms, and then estimate the model parameters and has already been widely applied in the identification of nonlinear systems[36-43].

4. Orthogonal least squares algorithm

Since the $N\times\overline{N}$ ($\overline{N}\le N$) measured matrix $\mathbf{\Phi}$ has full column rank, it can be uniquely decomposed as

$$\mathbf{\Phi}=\mathbf{QR} \tag{41}$$

where $\mathbf{Q}$ is an $N\times\overline{N}$ unitary matrix and $\mathbf{R}$ is an $\overline{N}\times\overline{N}$ upper triangular matrix with positive diagonal elements r_{11}, r_{22}, $\cdots$, $r_{\overline{N}\overline{N}}$.

Denote $\mathbf{D}=\mathrm{diag}[r_{11},r_{22},\cdots,r_{\overline{N}\overline{N}}]$ and then Eq.(41) can be rewritten as

$$\mathbf{\Phi}=\mathbf{WA} \tag{42}$$

where $\mathbf{A}=\mathbf{D}^{-1}\mathbf{R}$ is an $\overline{N}\times\overline{N}$ upper triangular matrix with unit diagonal elements, that is,

$$\mathbf{A}=\begin{bmatrix} 1 & a_{12} & a_{13} & \cdots & a_{1\bar{N}} \\ 0 & 1 & a_{23} & \cdots & a_{2\bar{N}} \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 & a_{\bar{N}-1\bar{N}} \\ 0 & \cdots & 0 & 0 & 1 \end{bmatrix} \tag{43}$$

and $\mathbf{W}=\mathbf{QD}$ is an $N\times\bar{N}$ matrix with orthogonal columns $\mathbf{w_j}$, $j=1,2,\cdots,\bar{N}$ such that

$$\mathbf{W}^{\mathrm{T}}\mathbf{W}=\mathbf{D}^2=\mathbf{H}=\mathrm{diag}\left[h_1,h_2,\cdots,h_{\bar{N}}\right] \tag{44}$$

where

$$h_j=\left\langle \mathbf{w_j},\mathbf{w_j}\right\rangle,\qquad j=1,2,\cdots,\bar{N} \tag{45}$$

and the symbol $\langle\cdot,\ \cdot\rangle$ denotes the inner product of two vectors.

Substituting Eq.(42) into Eq.(33) gives

$$\mathbf{F}=\mathbf{W}\mathbf{A}\boldsymbol{\Theta}+\boldsymbol{\Xi} \tag{46}$$

Denote

$$\mathbf{A}\boldsymbol{\Theta}=\mathbf{g} \tag{47}$$

and then Eq.(46) can be expressed as

$$\mathbf{F}=\mathbf{W}\mathbf{g}+\boldsymbol{\Xi} \tag{48}$$

or

$$\mathbf{F}=\sum_{j=1}^{\bar{N}} g_j\mathbf{w_j}+\boldsymbol{\Xi} \tag{49}$$

which is an auxiliary model equivalent to Eq.(33) and the space spanned by the orthogonal basis vectors $\mathbf{w_1},\mathbf{w_2},\cdots,\mathbf{w_{\bar{N}}}$ is the same as that spanned by the original model basis $\boldsymbol{\varphi_1},\boldsymbol{\varphi_2},\cdots,\boldsymbol{\varphi_{\bar{N}}}$.

By using the LS algorithm, the auxiliary parameter vector $\mathbf{g}$ can be solved from Eq.(48),

$$\mathbf{g}=\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{F} \tag{50}$$

Substituting Eq.(44) into Eq.(50) gives

$$\mathbf{g}=\mathbf{H}^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{F} \tag{51}$$

or

$$g_j=\frac{\left\langle \mathbf{F},\mathbf{w_j}\right\rangle}{\left\langle \mathbf{w_j},\mathbf{w_j}\right\rangle},\qquad j=1,2,\cdots,\bar{N} \tag{52}$$

Several orthogonalization procedures including classical Gram-Schmidt, modified Gram-Schmidt and Householder transformation[44] can be used to implement the orthogonal decomposition of the measured matrix $\boldsymbol{\Phi}$. Then after obtaining the auxiliary parameter vector $\mathbf{g}$ by Eq.(51), the parameter vector $\boldsymbol{\Theta}$ can be easily solved from Eq.(47) by using backward substitutions. However, our objective is not just to estimate the parameters, but also to detect which terms are significant and should be included within the model. This can be achieved by computing the error reduction ratio(ERR) described below.

Suppose that $F(t_i)$, , $i = 1,2,\cdots,N$ is the output after its mean has been removed. Since $\boldsymbol{\Xi}$ is uncorrelated with $\boldsymbol{\varphi_i}$, $i = 1,2,\cdots,\bar{N}$, the variance of $F(t_i)$ can be expressed as

$$\begin{aligned}\sigma_F^2 &= \frac{1}{N}\mathbf{F}^{\mathrm{T}}\mathbf{F} \\ &= \frac{1}{N}\left[\mathbf{g}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{g}+2\left(\mathbf{W}\mathbf{g}\right)^{\mathrm{T}}\boldsymbol{\Xi}+\boldsymbol{\Xi}^{\mathrm{T}}\boldsymbol{\Xi}\right] \\ &= \frac{1}{N}\left[\mathbf{g}^{\mathrm{T}}\mathbf{H}\mathbf{g}+2\left(\boldsymbol{\Phi}\boldsymbol{\Theta}\right)^{\mathrm{T}}\boldsymbol{\Xi}+\boldsymbol{\Xi}^{\mathrm{T}}\boldsymbol{\Xi}\right] \\ &= \frac{1}{N}\left[\sum_{j=1}^{\bar{N}} g_j^2 h_j + 2\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Xi}+\boldsymbol{\Xi}^{\mathrm{T}}\boldsymbol{\Xi}\right] \\ &= \frac{1}{N}\sum_{j=1}^{\bar{N}} g_j^2\mathbf{w}_{\mathbf{j}}^{\mathrm{T}}\mathbf{w}_{\mathbf{j}} + \frac{1}{N}\boldsymbol{\Xi}^{\mathrm{T}}\boldsymbol{\Xi}\end{aligned} \tag{53}$$

where the first part $\left(\sum_{j=1}^{\bar{N}} g_j^2\mathbf{w}_{\mathbf{j}}^{\mathrm{T}}\mathbf{w}_{\mathbf{j}}\right)/N$, which can be explained by the involved terms, is the desired output variance while the second part $\left(\boldsymbol{\Xi}^{\mathrm{T}}\boldsymbol{\Xi}\right)/N$ represents the unexplained variance. Thus $g_j^2\mathbf{w}_{\mathbf{j}}^{\mathrm{T}}\mathbf{w}_{\mathbf{j}}/N$ is the increment to the explained desired output variance brought by the *j*th term $\mathbf{w_j}$ and the *j*th error reduction ratio introduced by $\mathbf{w_j}$ can be defined as

$$ERR_j = \frac{g_j^2\mathbf{w}_{\mathbf{j}}^{\mathrm{T}}\mathbf{w}_{\mathbf{j}}}{\mathbf{F}^{\mathrm{T}}\mathbf{F}}\times 100\%, \quad j = 1,2,\cdots,\bar{N} \tag{54}$$

Substituting Eq.(52) into Eq.(54) yields

$$ERR_j = \frac{\left\langle \mathbf{F}, \mathbf{w}_{\mathbf{j}}\right\rangle^2}{\left\langle \mathbf{F}, \mathbf{F}\right\rangle\left\langle \mathbf{w}_{\mathbf{j}}, \mathbf{w}_{\mathbf{j}}\right\rangle}\times 100\%, \quad j = 1,2,\cdots,\bar{N} \tag{55}$$

which is also called the squared correlation coefficient between $\mathbf{F}$ and $\mathbf{w_j}$.

From Eq.(53), the residual sum of squares $\|\boldsymbol{\Xi}_{\bar{\mathbf{N}}}\|^2 = \left\|\mathbf{F}-\hat{\mathbf{F}}\right\|^2$, where $\hat{\mathbf{F}}$ is the model prediction produced by the associated $\bar{N}$ terms model, can also be obtained,

$$\left\|\boldsymbol{\Xi}_{\bar{\mathbf{N}}}\right\|^2 = \left\langle \mathbf{F}, \mathbf{F}\right\rangle - \sum_{j=1}^{\bar{N}}\frac{\left\langle \mathbf{F}, \mathbf{w}_{\mathbf{j}}\right\rangle^2}{\left\langle \mathbf{w}_{\mathbf{j}}, \mathbf{w}_{\mathbf{j}}\right\rangle} \tag{56}$$

while the residual vector $\boldsymbol{\Xi}_{\bar{\mathbf{N}}}$ can be expressed from Eq.(49) as

$$\boldsymbol{\Xi}_{\bar{\mathbf{N}}} = \mathbf{F} - \sum_{j=1}^{\bar{N}}\frac{\left\langle \mathbf{F}, \mathbf{w}_{\mathbf{j}}\right\rangle}{\left\langle \mathbf{w}_{\mathbf{j}}, \mathbf{w}_{\mathbf{j}}\right\rangle}\mathbf{w}_{\mathbf{j}} \tag{57}$$

Dividing both sides of Eq.(53) by $\mathbf{F}^{\mathrm{T}}\mathbf{F}/N$ gives

$$1 - \sum_{j=1}^{\bar{N}} ERR_j = \frac{\left(\boldsymbol{\Xi}^{\mathrm{T}}\boldsymbol{\Xi}\right)/N}{\left(\mathbf{F}^{\mathrm{T}}\mathbf{F}\right)/N} = \frac{\sigma_\varepsilon^2}{\sigma_F^2} \tag{58}$$

which clearly indicates that the larger the ERR value associated with a particular term is, the more reduction in the residual variance will be produced if this term is included in the model. Thus the ERR provides a simple but effective means to detect which term is significant and should be

selected. Notice that a term which is introduced at an early stage will have a larger ERR than that would be obtained if it were reordered to enter as a candidate term at a later stage. To overcome the order dependency of ERR, the terms can be selected in a forward stepwise manner. The detailed orthogonalization, for example, using the classical Gram-Schmidt algorithm, and terms selection procedure is described as follows.

□ At the first step, consider all the possible $\boldsymbol{\varphi}_{\mathbf{j}}$, $j = 1,2,\cdots,\bar{N}$ as candidates for $\mathbf{w}_1$, and for $j = 1,2,\cdots,\bar{N}$, compute

$$\begin{cases} \mathbf{w}_1^{(\mathbf{j})} = \boldsymbol{\varphi}_{\mathbf{j}}; \\ ERR_1^{(j)} = \dfrac{\left\langle \mathbf{F}, \mathbf{w}_1^{(\mathbf{j})} \right\rangle^2}{\langle \mathbf{F}, \mathbf{F} \rangle \left\langle \mathbf{w}_1^{(\mathbf{j})}, \mathbf{w}_1^{(\mathbf{j})} \right\rangle} \times 100\%. \end{cases} \tag{59}$$

Find the maximum of $ERR_1^{(j)}$, say $ERR_1^{(j_1)} = \max\left\{ERR_1^{(j)}, 1 \le j \le \bar{N}\right\}$. Then the first term to be included in the model is $\boldsymbol{\varphi}_{\mathbf{j_1}}$. $\mathbf{w}_1 = \mathbf{w}_1^{(\mathbf{j_1})} = \boldsymbol{\varphi}_{\mathbf{j_1}}$ is then selected as the first column of $\mathbf{W}$ together with the first element of the auxiliary parameter vector $\mathbf{g}$, $g_1 = \langle \mathbf{F}, \mathbf{w}_1 \rangle / \langle \mathbf{w}_1, \mathbf{w}_1 \rangle$, the error reduction ratio produced by the first term, $ERR_1 = ERR_1^{(j_1)}$, and the associated sum-squared-error $\|\boldsymbol{\Xi}_1\|^2 = \langle \mathbf{F}, \mathbf{F} \rangle - \langle \mathbf{F}, \mathbf{w}_1 \rangle^2 / \langle \mathbf{w}_1, \mathbf{w}_1 \rangle$. As defined in Eq.(43), the first column of $\mathbf{A}$, $a_{11} = 1$.

□ At the kth step where $k \ge 2$, all the $\boldsymbol{\varphi}_{\mathbf{j}}$, $j = 1,2,\cdots,\bar{N}$, $j \notin \{j_1,\cdots,j_{k-1}\}$ are considered as possible candidates for $\mathbf{w}_k$, and for $j = 1,2,\cdots,\bar{N}$, $j \notin \{j_1,\cdots,j_{k-1}\}$, calculate

$$\begin{cases} \mathbf{w}_k^{(\mathbf{j})} = \boldsymbol{\varphi}_{\mathbf{j}} - \sum_{p=1}^{k-1} \dfrac{\left\langle \boldsymbol{\varphi}_{\mathbf{j}}, \mathbf{w}_p \right\rangle}{\left\langle \mathbf{w}_p, \mathbf{w}_p \right\rangle} \mathbf{w}_p; \\ ERR_k^{(j)} = \dfrac{\left\langle \mathbf{F}, \mathbf{w}_k^{(\mathbf{j})} \right\rangle^2}{\langle \mathbf{F}, \mathbf{F} \rangle \left\langle \mathbf{w}_k^{(\mathbf{j})}, \mathbf{w}_k^{(\mathbf{j})} \right\rangle} \times 100\%. \end{cases} \tag{60}$$

Find the maximum of $ERR_k^{(j)}$, say $ERR_k^{(j_k)} = \max\left\{ERR_k^{(j)}, 1 \le j \le \bar{N}, j \ne j_1, \cdots, j \ne j_{k-1}\right\}$. Then the kth term to be included in the model is $\boldsymbol{\varphi}_{\mathbf{j_k}}$ while the kth column of $\mathbf{W}$, $\mathbf{w}_k = \mathbf{w}_k^{(\mathbf{j_k})}$, the kth element of the auxiliary parameter vector $\mathbf{g}$, $g_k = \langle \mathbf{F}, \mathbf{w}_k \rangle / \langle \mathbf{w}_k, \mathbf{w}_k \rangle$, the kth error reduction ratio $ERR_k = ERR_k^{(j_k)}$, and the kth sum-squared-error $\|\boldsymbol{\Xi}_k\|^2 = \langle \mathbf{F}, \mathbf{F} \rangle - \sum_{j=1}^{k} \langle \mathbf{F}, \mathbf{w}_j \rangle^2 / \langle \mathbf{w}_j, \mathbf{w}_j \rangle$. The elements of the kth column of $\mathbf{A}$ are computed by

$$a_{pk} = \begin{cases} \dfrac{\left\langle \boldsymbol{\varphi}_{\mathbf{j_k}}, \mathbf{w}_p \right\rangle}{\left\langle \mathbf{w}_p, \mathbf{w}_p \right\rangle}, & p = 1,\cdots,k-1 \\ 1, & p = k \end{cases} \tag{61}$$

□ According to Eq.(58), the procedure can be terminated at the $\bar{\bar{N}}$th step ($\bar{\bar{N}} \le \bar{N}$) when

$$1 - \sum_{j=1}^{\bar{\bar{N}}} ERR_j < \rho, \qquad 0 < \rho < 1 \tag{62}$$

where ρ is a chosen error tolerance and in practice, can actually be learnt during the selection procedure.

The criterion (62) concerns only the performance of the model (variance of residuals). Because a more accurate performance is often achieved at the expense of using a more complex model, a trade-off between the performance and complexity of the model is often desired. A number of model selection criteria that provide a compromise between the performance and the number of parameters have been introduced and incorporated into the OLS algorithm over the past few decades. Despite the differences amongst these model selection criteria, they are asymptotically equivalent under general conditions[43]. In this paper, the adjustable prediction error sum of squares (APRESS)[40] is employed to solve the model length determination problem,

$$APRESS(n) = c(n)MSE(n) \tag{63}$$

where

$$c(n) = \left(\frac{1}{1-\alpha n/N}\right)^2 \tag{64}$$

with $\alpha \geq 1$, is the complexity cost function and

$$MSE(n) = \frac{\|\boldsymbol{\Xi}_{\mathbf{n}}\|^2}{N} \tag{65}$$

is the mean squared error corresponding to the model performance.

The model selection procedure is terminated at the $\overline{\overline{N}}$th step when

$$APRESS\left(\overline{\overline{N}}\right) = \min_{1\leq n\leq \overline{N}}\left[APRESS(n)\right] \tag{66}$$

Practically a distinct turning point of the APRESS statistic versus the model length can be easily found, especially when computed by using several adjustable parameters α, and this can then be used to determine the model length.

The final model is thus the linear combination of the $\overline{\overline{N}}$ significant terms $\boldsymbol{\varphi}_{\mathbf{j_1}}, \cdots, \boldsymbol{\varphi}_{\mathbf{j}_{\overline{\overline{N}}}}$ selected from the $\overline{N}$ candidate terms $\boldsymbol{\varphi}_{\mathbf{1}}, \cdots, \boldsymbol{\varphi}_{\overline{\mathbf{N}}}$,

$$F(t) = \sum_{k=1}^{\overline{\overline{N}}} \hat{\theta}_k \phi_{j_k}(t) + \varepsilon(t) \tag{67}$$

where the parameter $\widehat{\boldsymbol{\Theta}} = \left[\hat{\theta}_1, \cdots, \hat{\theta}_{\overline{\overline{N}}}\right]^{\mathrm{T}}$ can easily be computed from Eq.(47) by using backward substitutions,

$$\begin{cases} \hat{\theta}_{\overline{\overline{N}}} = g_{j_{\overline{\overline{N}}}}; \\ \hat{\theta}_k = g_{j_k} - \sum_{p=k+1}^{\overline{\overline{N}}} a_{kp}\hat{\theta}_p \quad for\ k = \overline{\overline{N}}-1, \overline{\overline{N}}-2, \cdots, 1. \end{cases} \tag{68}$$

It should be pointed out that in practice the mean of the output does not need to be removed because adding a constant to the denominator of the ERR (54) will not affect the result of the maximization in this selection procedure. Because of the orthogonal property, this procedure is very efficient and leads to a parsimonious model. Moreover, any numerical ill-conditioning can be avoided by eliminating $\mathbf{w_k}$ if $\mathbf{w_k^T w_k}$ is less than a predetermined threshold. Similar selection procedures can also be derived using the modified Gram-Schmidt algorithm and Householder transformation algorithm.

5. Model identification procedure

5.1 Sampling intervals

In practice, a uniform-rate sampling strategy is normally utilised and the choice of the sampling interval is critical for the quality of the identification. If the sampling interval is chosen too small (the signal is oversampled), the data become redundant. On the other hand, if the sampling interval is chosen too large (the signal is under-sampled), then adjacent data points tend to become independent or irrelevant and the information about the original system will be lost. There are currently two major approaches available for the determination of the sampling interval of nonlinear systems. The first method is based on the autocorrelation functions and the sampling interval T_s can be chosen as[45]

$$\frac{1}{20}\min\left\{T_x, T_{x^2}\right\} \le T_s \le \frac{1}{10}\min\left\{T_x, T_{x^2}\right\} \tag{69}$$

where T_x and T_{x^2} are the time value of the first minimum of the autocorrelation functions $\bar{\gamma}_{x'x'}(T)$ and $\bar{\gamma}_{x^2'x^2'}(T)$ of a signal $x(t)$ respectively and

$$\bar{\gamma}_{x'x'}(T) = E\left[\left(x(t) - E\left[x(t)\right]\right)\left(x(t+T) - E\left[x(t)\right]\right)\right] \tag{70}$$

$$\bar{\gamma}_{x^2'x^2'}(T) = E\left[\left(x^2(t) - E\left[x^2(t)\right]\right)\left(x^2(t+T) - E\left[x^2(t)\right]\right)\right] \tag{71}$$

In some cases, the upper bound on T_s can be somewhat relaxed, say, $\min\{T_x, T_{x^2}\}/5$.

Another method is based on the mutual information which can measure the general dependence of two variables. Assume that the marginal probability distribution functions of two discrete random variables **X** and **Y** are $p_1(x)$ and $p_2(y)$ respectively, and the joint probability distribution function of **X** and **Y** is $p(x, y)$. The mutual information of **X** and **Y** is defined as

$$I(\mathbf{X}, \mathbf{Y}) = \sum_{y\in\mathbf{Y}}\sum_{x\in\mathbf{X}} p(x, y)\log_2\left[\frac{p(x, y)}{p_1(x)\,p_2(y)}\right] \tag{72}$$

In practice, the mutual information can be obtained by either a nonparametric estimator including equal probability distribution or equal distance partitioning of the **XY** plane, histogram based estimator, kernel density estimator, spline estimator, nearest neighbour estimator, and wavelet density estimator, or a parametric estimator such as Bayesian estimator, Edgeworth estimator, maximum likelihood estimator, and least square estimator[46]. The sampling interval T_s can be chosen as the time value of the first minimum in the mutual information[47]

$$\mu_{\mathbf{XX}}(T) = I\left(\mathbf{X}(t), \mathbf{X}(t+T)\right) \tag{73}$$

Note that the time delay calculated based on either the autocorrelation functions method or the mutual information method can only provide some suggestions on choosing sampling intervals. For input-output nonlinear systems, there is no guarantee that the methods will give the exact sampling intervals. Experience has shown that the choice of the sampling interval affects both the model structure selection and the parameter estimation in the identification of a system but in antagonistic ways, which suggests that, given a set of well-sampled data, structure selection can be performed on the decimated data and after choosing a particular structure, the parameters can then be estimated from the original records[45].

Notice that in many practical cases, the system input and output will be corrupted by noise. Noise may be induced from external and/or internal sources as well as from the measuring instruments themselves. This means that the residual $\boldsymbol{\Xi}$ in Eq.(33) is more likely to be a coloured or correlated noise sequence and thus the obtained model may be biased. A biased model will probably predict well over the data used in estimation but may be totally inappropriate for predicting the system response to different data sets or alternative inputs. In other words, it is just a curve fit to one data set and is not a model of the underlying system that produced the data. The bias will only be eliminated if the residual becomes uncorrelated with past measurements. And one way to achieve this is to model the noise.

5.2 Noise model

Taking into account the combined effects of measurement noises, modelling errors and unmeasured disturbances, Eq.(29) can be represented by an NARMAX (Nonlinear Auto Regressive Moving Average with eXogenous input) model[48-50] of the form,

$$F\left(t_i\right)=f\left(y\left(t_i\right),\dot{y}\left(t_i\right),\xi_{i-1},\cdots,\xi_{i-N_\xi}\right)+\xi_i \tag{74}$$

where $N_\xi$is the maximum lag of the system model residual ξ_i and $f(\cdot)$ is an unknown nonlinear mapping and can be constructed using a variety of basis functions including polynomials, rational functions, output-affine functions, radial basis functions, kernel functions, splines, and wavelets, the most popular representation of which is the polynomial model taking the form below,

$$f\left(x_i^1,x_i^2,\cdots,x_i^S\right)=\beta_0+\sum_{l=1}^{L}\sum_{s_1=1}^{S}\cdots\sum_{s_l=s_{l-1}}^{S}\beta_{s_1s_2\cdots s_l}\prod_{k=1}^{l}x_i^{s_k} \tag{75}$$

where

$$S=N_\xi+2 \tag{76}$$

$$x_i^k=\begin{cases}y\left(t_i\right), & k=1\\ \dot{y}\left(t_i\right), & k=2\\ \xi_{i-(k-2)}, & k=3,\cdots,S\end{cases} \tag{77}$$

β_0, $\beta_{d_1d_2\cdots d_l}$ are the parameters and L is the degree of the model which is defined as the highest order among the polynomial terms.

The noise model can then be expressed as

$$\varepsilon_t=f\left(x_i^1,x_i^2,\cdots,x_i^S\right)-f\left(y\left(t_i\right),Dy\left(t_i\right)\right)+\xi_i \tag{78}$$

where D denotes the differential operator and $f\left(y(t_i),Dy(t_i)\right)$ is the process model with a general representation,

$$f\left(y\left(t_i\right),Dy\left(t_i\right)\right)=\beta_0+\sum_{l=1}^{L_1}\sum_{d_1=1}^{2}\cdots\sum_{d_l=d_{l-1}}^{2}\beta_{d_1d_2\cdots d_l}\prod_{k=1}^{l}D^{d_k-1}y\left(t_i\right) \tag{79}$$

while for the MR device, it has already been derived in Section 3 as

$$f\left(y\left(t_i\right),Dy\left(t_i\right)\right)=\sum_{j=1}^{\bar{N}}\theta_j\phi_j\left(t_i\right) \tag{80}$$

where θ_j, $\phi_j(t_i)$, $j=1,2,\cdots,\bar{N}$ are given by Eq.(31) and Eq.(32) respectively.

Specially if an MR device is under accurate displacement control, Eq.(74) is reduced to the following model,

$$F(t_i) = f\left(y(t_i), \dot{y}(t_i)\right) + \varepsilon_i \tag{81}$$

where $f\left(y(t_i), \dot{y}(t_i)\right)$ is the process model or the dynamic model given by Eq.(79) or (80) while the noise model is given by

$$\varepsilon_i = \sum_{l=1}^{L_2} \sum_{d_1=1}^{N_\xi} \cdots \sum_{d_l=d_{l-1}}^{N_\xi} \overline{\beta}_{d_1 d_2 \cdots d_l} \prod_{k=1}^{l} \xi_{i-d_k} + \xi_i \tag{82}$$

and $L = max\{L_1, L_2\}$.

5.3 Model validity tests

If the model structure and parameter estimates obtained by the identification procedure are statistically valid, the system model residual ε_i, $i = 1,2,\cdots,N$, should be uncorrelated with all linear and nonlinear combinations of past inputs and outputs. For nonlinear systems, under some mild assumptions, this can be tested by computing the following normalized correlation functions[51],

$$\begin{cases} \gamma_{\varepsilon\varepsilon}(\tau) = \delta(\tau), & \forall \tau \\ \gamma_{y\varepsilon}(\tau) = 0, & \forall \tau \\ \gamma_{y^{2'}\varepsilon}(\tau) = 0, & \forall \tau \\ \gamma_{y^{2'}\varepsilon^2}(\tau) = 0, & \forall \tau \\ \gamma_{(y\varepsilon)\varepsilon}(\tau) = 0, & \tau \geq 0 \end{cases} \tag{83}$$

where $\delta(\tau)$ is the Kronecker delta function and

$$y^{2'}(t) = y^2(t) - E\left[y^2(t)\right] \tag{84}$$

while the normalized correlation function between two signals $x_1(t)$ and $x_2(t)$ for a time delay τ is defined as

$$\gamma_{x_1 x_2}(\tau) \triangleq \frac{E\left[x_1(t)\, x_2(t+\tau)\right]}{\sqrt{E\left[x_1^2(t)\right] E\left[x_2^2(t)\right]}} \tag{85}$$

In practice, if these correlation functions in Eq.(83) fall within the confidence intervals at a given significance level α ($0 < \alpha < 1$), say $\alpha = 0.05$, which corresponds to the 95% confidence intervals, $\pm 1.96/\sqrt{N}$, the model is regarded as adequate and acceptable.

5.4 Procedure for the model identification

Eq.(82) can also be written by

$$\varepsilon_i = \sum_{j=1}^{\tilde{N}} \tilde{\theta}_j \tilde{\phi}_j \left(\xi_{i-1}, \cdots, \xi_{i-N_\xi}\right) + \xi_i \tag{86}$$

where $\tilde{\phi}_j \left(\xi_{i-1}, \cdots, \xi_{i-N_\xi}\right)$ represents a possible noise term and $\tilde{\theta}_j$ is the corresponding parameter.

Substituting Eq.(80) and Eq.(86) into Eq.(81) yields

$$F(t_i) = \sum_{j=1}^{\bar{N}+\tilde{N}} \lambda_j \psi_{ij} + \xi_i \tag{87}$$

where

$$\lambda_j = \begin{cases} \theta_j, & j = 1, \cdots, \bar{N} \\ \tilde{\theta}_j, & j = \bar{N}+1, \cdots, \bar{N}+\tilde{N} \end{cases} \tag{88}$$

$$\psi_{ij} = \begin{cases} \phi_j(t_i), & j = 1, \cdots, \bar{N} \\ \tilde{\phi}_j\left(\xi_{i-1}, \cdots, \xi_{i-N_\xi}\right), & j = \bar{N}+1, \cdots, \bar{N}+\tilde{N} \end{cases} \tag{89}$$

The matrix form of Eq.(87) is expressed as

$$\mathbf{F} = \mathbf{\Psi}\mathbf{\Lambda} + \Upsilon \tag{90}$$

where

$$\mathbf{\Psi} = \left[\psi_{ij}\right]_{\substack{i=1,2,\cdots,N \\ j=1,2,\cdots,\bar{N}+\tilde{N}}} \tag{91}$$

$$\mathbf{\Lambda} = \left[\lambda_1, \quad \lambda_2, \quad \cdots, \quad \lambda_{\bar{N}+\tilde{N}}\right]^T \tag{92}$$

$$\Upsilon = \left[\xi_1, \quad \xi_2, \quad \cdots, \quad \xi_N\right]^T \tag{93}$$

and **F** is given by Eq.(34).

Notice that delayed noise sequences ξ_{i-N_ξ}, … , ξ_{i-1}have been included within the matrix $\mathbf{\Psi}$ in the system model (90) and they can be estimated using the model residuals. The orthogonal property of the OLS algorithm ensures that the selection of the process and noise model terms can be decoupled. Significant terms in the process model are selected initially. Which terms are included in the process model will not be affected by whatever noise model is produced later because of the orthogonal property. Initial residuals are computed based on the determined process model. The OLS algorithm is then applied to the noise model. Revised residuals are calculated and improved noise model is determined. A few iterations are often enough to get the candidate noise terms convergent. The structure or which terms to include in the noise model can then be selected and the parameters of the process model can be estimated. The final model is validated by performing the model validity tests. The detailed procedure is as follows,

(a) Choose the sampling interval T_s using Eq.(69) or based on the mutual information (73). Select the initial values of $\bar{N}_1$, $\bar{N}_2$, $\bar{N}_3$, N_ξ and L_2 in the system model (90).

(b) Use the OLS algorithm to select terms in the process model. The selection is terminated at the $\bar{\bar{N}}$th step according to the criterion (66). Set $k = 0$ and calculate the initial process model parameter $\theta_j^{(k)}, j = 1,2,\cdots,\bar{\bar{N}}$. Compute the process model residual $\mathbf{\Xi}_{\bar{\bar{N}}}$ using Eq.(57) and set the initial system model residual

$$\hat{\Upsilon}^{(k)} = \mathbf{\Xi}_{\bar{\bar{N}}} \tag{94}$$

where

$$\hat{\Upsilon}^{(k)} = \left[\hat{\xi}_{kN_\xi+1}^{(k)}, \quad \hat{\xi}_{kN_\xi+2}^{(k)}, \quad \cdots, \quad \hat{\xi}_N^{(k)}\right] \tag{95}$$

(c) Set $k = k + 1$. Update the system model (90) using the selected process terms, Eq.(95),

$$\mathbf{F}^{(k)} = \left[F\left(t_{kN_\xi+1}\right), \quad F\left(t_{kN_\xi+2}\right), \quad \cdots, \quad F\left(t_N\right) \right]^T \tag{96}$$

and

$$\hat{\mathbf{\Psi}}^{(k)} = \left[\psi_{ij} \right]_{\substack{i=kN_\xi+1,\, kN_\xi+2,\, \cdots,\, N \\ j=1,\, 2,\, \cdots,\, \bar{\bar{N}}+\tilde{N}}} \tag{97}$$

(d) Apply the OLS algorithm to the updated system model and the selection is terminated when $1 - \sum_{\mathrm{j}=1}^{\bar{\bar{N}}+\tilde{\mathrm{N}}'} ERR_j < \rho_\xi$ ($0 < \rho_\xi < 1$). Compute the process model parameter $\theta_j^{(k)}$, $j = 1,2,\cdots,\bar{\bar{N}}$. If the convergence criterion that

$$\max_{1 \le j \le \bar{\bar{N}}} \left| \frac{\theta_j^{(k)} - \theta_j^{(k-1)}}{\theta_j^{(k)}} \right| < a\ tolerance \tag{98}$$

is satisfied, stop. Otherwise, calculate the system model residual

$$\hat{\Upsilon}^{(k)} = \mathbf{F}^{(k)} - \sum_{j=1}^{\bar{\bar{N}}+\tilde{N}'} \frac{\left\langle \mathbf{F}^{(k)}, \mathbf{w}_\mathbf{j} \right\rangle}{\left\langle \mathbf{w}_\mathbf{j}, \mathbf{w}_\mathbf{j} \right\rangle} \mathbf{w}_\mathbf{j} \tag{99}$$

and go to (c).

(e) Apply the OLS algorithm to the updated system model to select the noise model terms and the selection is terminated at the $\bar{\bar{N}} + \tilde{\bar{N}}$th step according to the criterion (66). Perform the model validity tests. If the criterion (83) is satisfied, compute the process model parameters. Otherwise, go to (a).

In practice, repeating steps (c) and (d) four or five times should be sufficient to achieve convergence. At each iteration, the selection is performed from the original set of candidate noise terms. And to force enough noise terms into the model, a small error tolerance ρ_ξ should be used. Notice that because of the orthogonal property, the selection of the process model terms and the calculation of the process model residuals are both totally decoupled from the noise model updates. However, it should also be pointed out that an unbiased estimation of the process model parameters critically depends on the noise model. This is the reason why the noise model must be appropriately determined although at the end of the identification procedure it is discarded.

Experience has shown that nonlinear terms are often the combinations of significant linear terms in nonlinear models[52]. Therefore in practice, the OLS algorithm may be used to identify the best following linear MA (Moving Average) noise model first,

$$\varepsilon_i = \sum_{d=1}^{N_\xi} \overline{\beta}_d \xi_{i-d} + \xi_i \tag{100}$$

and then to extend it to the nonlinear model (82) to include all nonlinear combinations of the selected linear terms. Another model often utilised as the starting point of the selection is Eq. (82) without cross product terms expressed as follows,

$$\varepsilon_i = \sum_{l=1}^{L_2} \sum_{d=1}^{N_\xi} \overline{\beta}_{ld} \xi_{i-d}^l + \xi_i \tag{101}$$

Such model sets will be significantly smaller than a full nonlinear model set and hence the reduction in the computational burden will be substantial. There is, however, no guarantee that the nonlinear model identified in this way will be as good as that selected from the full model set. If a

simplified model is fitted at the initial stage, the model validity tests can also be used to indicate whether the prediction accuracy of the fitted model could be improved by including more nonlinear terms within the model. If the tests indicate that no significant nonlinear terms remain in the residuals, the selection of the noise model terms would terminate at this stage.

6. Model identification of a metal rubber specimen

The cylindrical MR specimen, shown in Fig. 1, with diameter of 30 mm, height of 30 mm, diameter of the stainless steel (0Cr18Ni9Ti) wires 0.12 mm, diameter of spirals 1.2 mm, relative density (the ratio of MR density to the wire density) 0.24, and forming force 23 kN, was tested on a servohydraulic material testing machine at room temperature. A precompression of 2.2 mm was initially loaded to the MR specimen. Then an amplitude-varying harmonic excitation, which is produced by the testing machine under a displacement control with the frequency of 20 Hz, was applied. The deformation displacement and corresponding force signals were collected by a data acquisition system with fixed sampling interval of 0.1 ms and shown in Fig. 5. The mutual information of both signals was then calculated and shown in Fig. 6, which suggests that they should be downsampled by a sampling interval of 0.3 ms, the time value of the first minimum of the mutual information, to generate the estimation data.

As a starting point, suppose that the term length of the elastic stiffness restoring force, the viscous damping restoring force, and the bilinear hysteresis restoring force $\overline{N}_1, \overline{N}_2, \overline{N}_3$ in Eq.(29) were all 10. By setting that $L_2 = 3$ and $N_\xi = 20$, Eq.(101) was initially utilised to represent the noise model. The initial system model thus involved a total of 31 candidate process model terms and 60 candidate noise model terms. One cycle measurements with the minimum point as the initial value were used for the model identification[53]. Performing the OLS algorithm on the process model indicated that the terms with the order higher than 5 and 13 for the elastic stiffness force and the viscous damping force respectively lead to ill-conditioning of the measured matrix. Therefore they should not be included into the model, that is to say, $\overline{N}_1 = 3$, $\overline{N}_2 = 7$, and $\overline{N}_3 = 10$. Under these assumptions, the OLS algorithm was again used to select and rank the significant process model terms. By setting the adjustable parameter $\alpha = 0, 1, 1.5, \cdots, 3$, the APRESS statistic versus the process model length over the acquisition data, were calculated and shown in Fig. 7, where the bottom line with circles, corresponding to $\alpha = 0$, indicates the mean-squared-errors. It can be seen from Fig. 7 that there is an obvious turning point at the abscissa 11 for various values of the adjustable parameter α. The indexes of the first 11 process model terms selected and ranked in order of the significance by the OLS algorithm, together with the coefficient of each term and its corresponding ERR, are shown in Table 1. Table 1 indicates that the terms with the cubic, quintic and linear stiffness restoring forces, the linear, cubic and ninth order damping restoring forces, and the approximating Chebyshev series of maximum degree 4 should be included within the model.

Reapplying the OLS algorithm to the dynamic model represented by the selected process terms gave the process model residuals, which were then set to be the initial system model residuals. Repeating steps (c) and (d) four times achieved convergence. Perform the OLS algorithm on the system model updated by the 11 process terms and the convergent candidate noise terms once again to select and rank the significant noise model terms. By setting $\alpha = 0, 1, 1.1, \cdots, 1.5$, the APRESS statistic versus the model length were calculated and shown in Fig. 8, which indicates that the number of the total model terms should be 20. The indexes of the first 9 noise model terms selected and ranked in order of the significance by the OLS algorithm, together with the ERR of each term,

are shown in Table 2. Representing the system model using these selected process and noise model terms and performing the OLS algorithm over the measurements once more gave the system model residuals. The associated model validity tests are illustrated in Fig. 9. Notice that the selected noise terms in Table 2 are all linear terms. Fig. 9 indicates that the selected linear noise model is acceptable for this case and the nonlinear analysis doesn't need to be pursued. Then after computing the process model parameters, solving Eq.(39) and discarding the noise model, the final identified dynamic model is obtained,

$$\begin{cases} k_1\left(y(t)+y_0\right)+k_3\left(y(t)+y_0\right)^3+k_5\left(y(t)+y_0\right)^5+c_1\dot{y}(t)+c_3\dot{y}(t)^3+c_9\dot{y}(t)^9+z(t)=F(t) \\ \mathrm{d}z(t)=\dfrac{k_s}{2}\left[1+\mathrm{sgn}\left(z_s-\left|z(t)\right|\right)\right]\mathrm{d}y(t) \end{cases} \tag{102}$$

where $k_1 = 6.9781 \times 10^4$ N m^{-1}, $k_3 = -1.4321 \times 10^{10}$ N m^{-3}, $k_5 = 9.4290 \times 10^{14}$ N m^{-5}, $c_1 = 39.963$ s N m^{-1}, $c_3 = -1.5000 \times 10^4$ s^{-3} N m^{-3}, $c_9 = 6.1047 \times 10^9$ s^{-9} N m^{-9}, $k_s = 8.7072 \times 10^4$ N m^{-1}, $z_s = 53.442$ N and $y_0 = 2.2 \times 10^{-3}$ m, which is corresponding to the preload F_0.

The bilinear hysteresis loop produced by the identified model is shown in Fig. 10, which clearly demonstrates that the elastic deformation limit y_s (0.61377 mm) is smaller than the peak displacements and sliding between wires happened in the collected cycle, just as assumed in Section 2. Note that in Fig. 10, the bilinear hysteresis restoring force is not constant during sliding because it is approximated by the Chebyshev series of maximum degree 4 while the discontinuity at the velocity reversals is caused by the high level noise of the collected data around the peaks. A comparison of the hysteresis loop produced by the identified model and that plotted directly from the experimental measurements (100 cycles later than the estimating data set) is shown in Fig. 11 while a comparison of the corresponding restoring force is shown in Fig. 12. Fig. 11 together with Fig. 12 further confirms the model validity testing conclusion.

Notice that the cubic and ninth order damping term is selected at a later stage by the OLS algorithm and the ERR listed in Table 1 also indicates that they are less significant compared with other selected stiffness and damping terms. In fact as aforementioned, the model without the cubic and ninth order viscous damping force has already been used to model MR devices by some researchers although it is shown here for the first time how the model structure of MR devices can be automatically detected from the experimental data only. In addition, the estimated negative cubic stiffness indicates that MR devices can produce the softening effect and jump phenomenon, which were also reported by Wang et al.[54].

7. Conclusions

A semi-constitutive model that involves nonlinear elastic stiffness, nonlinear viscous damping and bilinear hysteretic Coulomb damping, which was later approximated by Chebyshev polynomials of the first kind, has been adopted to model MR devices. Then after generalising the dynamic model by taking into account the effects of noises, an efficient systematic procedure based on the OLS algorithm, the APRESS criterion and the nonlinear model validity tests was developed for model structure detection and parameter estimation. By utilising the response of a cylindrical MR specimen, it has been shown for the first time how the model structure of MR devices can be automatically detected and then the model parameters can be estimated. In this study, the approach is demonstrated by utilising a harmonic excitation. But for a more general type of excitation,

benefiting from the continuation of the displacement to the maximum amplitude span, the method can still work as long as the excitation satisfies the following conditions,

(a) there is a preload if the excitation is cyclical;
(b) the excitation amplitude is large enough to make sliding between wires happen in each collected cycle;
(c) under some parts of the excitation, the deformation velocity changes but remains positive or negative;
(d) the approximation error produced by the continuation from the extrema to the maxima of the displacement is acceptable.

Of these conditions, (a), (b) and (c) ensure that the measurement matrix is not singular. Another solution to make the method work for general excitations is to represent the hysteresis restoring force by a number of parallel hysteretic Coulomb dampers with one common input and output. This single-input single-output(SISO) multiple degree of freedom(MDOF) representation is often referred to as the Maxwell-slip model[55].

Experiments demonstrated that the mechanical properties of MR devices are so complex that even for a periodic excitation, the hysteretic effects depend on each parameter of the excitation, i.e. preload, amplitude, and frequency. This indicates that a model with preload-dependent and/or excitation-dependent parameters may be necessary to represent the MR device. A comprehensive research of this is far more ambitious and will be reported in a later publication. Although the approach in this paper is demonstrated by using one MR device under one set of conditions, it forms the foundation for the identification of the whole class of MR devices of various shapes, materials, etc. This has never been done before and opens up a principled modelling approach.

Acknowledgement

The authors gratefully acknowledge that this work was supported by the Engineering and Physical Sciences Research Council (EPSRC), UK and the European Research Council. They are grateful to Professor Jie Hong, Beihang University (China), for providing the metal rubber specimen and the continued engagement in the development of metal rubber properties.

References

[1] Д.Е. Чегодаев, О.П. Мулюкин, Е.В. Колтыгин, The Design of Metal Rubber Component, Z. Y. Li, translated in Chinese, ISBN: 9787118021813, National Defence Industry Press, Beijing, 2000.
[2] D.W. Childs, Space-shuttle main engine high-pressure fuel turbopump rotodynamic instability problem, Journal of Engineering for Power, 100 (1978) 48-57.
[3] A. Okayasu, T. Ohta, T. Azuma, T. Fujita, H. Aoki, Vibration problems in the LE-7 LH2 turbopump, Proceedings of the 26th AIAA Joint Propulsion Conference, Orlando, Florida, (1990) 1-5.
[4] M. Zarzour, J. Vance, Experimental evaluation of a metal mesh bearing damper, Journal of Engineering for Gas Turbines and Power, 122 (2000) 326-329.
[5] E.M. Al-Khateeb, Design, Modelling, and Experimental Investigation of Wire Mesh Vibration Dampers, Ph.D. Thesis, Texas A&M University, 2002.
[6] B.H. Ertas, H.G. Luo, Nonlinear dynamic characterization of oil-free wire mesh dampers, Journal of Engineering for Gas Turbines and Power, 130 (2008) 032503-(1-8).

[7] H. Zuo, Y.H. Chen, H.B. Bai, H. Sun, The compression deformation mechanism of a metallic rubber, International Journal of Mechanics and Materials in Design, 2 (2005) 269-277.
[8] X. Wang, Z.G. Zhu, A ringlike metal rubber damper, Journal of Aerospace Power, 12 (1997) 143-145.
[9] H.Y. Jiang, Y.H. Xia, D.Z. Su, A. Belousov, Analysis on infiltrative characteristics of deformable porous metal rubber material and parameters identification, Chinese Journal of Mechanical Engineering (English Edition), 18 (2005) 562-565.
[10] H.R. Ao, H.Y. Jiang, W. Wei, A.M. Ulanov, Study on the damping characteristics of mr damper in flexible supporting of turbo-pump rotor for engine, Proceedings of the 1st International Symposium on Systems and Control in Aerospace and Astronautics, Harbin, China, (2006) 618-622.
[11] Y.H. Ma, B.T. Guo, Z.G. Zhu, Static characteristics of metal rubber, Journal of Aerospace Power, 19 (2004) 326-330.
[12] C.X. Yang, Y. Zhou, H. Zhang, Research on dynamic performance of metal rubber damper, Acta Aeronautica Et Astronautica Siniga, 27 (2006) 536-539.
[13] Y.Y. Li, X.Q. Huang, Influencing factors of damping characteristic for metal rubber, Journal of Vibration, Measurement & Diagnosis, 29 (2009) 23-26.
[14] B.T. Guo, Z.G. Zhu, R.F. Cui, W. Pang, Theoretical model of metal rubber, Journal of Aerospace Power, 19 (2004) 314-319.
[15] Y.Y. Li, X.Q. Huang, W.X. Mao, A theoretical model and experimental investigation of a nonlinear constitutive equation for elastic porous metal rubbers, Mechanics of Composite Materials, 41 (2005) 303-312.
[16] Y.Y. Li, X.Q. Huang, Constitutive relation for metal rubber with different density and shape factor, Acta Aeronautica Et Astronautica Siniga, 29 (2008) 1084-1090.
[17] M. Brokate, J. Sprekels, Hysteresis and Phase Transitions, Springer-Verlag, New York, 1996.
[18] A. Visintin, Differential Models of Hysteresis, Springer-Verlag, New York, 1995.
[19] J.H. Oh, D.S. Bernstein, Semilinear Duhem model for rate-independent and rate-dependent hysteresis, IEEE Transactions on Automatic Control, 50 (2005) 631-645.
[20] L.O. Chua, S.C. Bass, A generalized hysteresis model, IEEE Transactions on Circuit Theory, CT-19 (1972) 36-48.
[21] M.L. Hodgdon, Applications of a theory of ferromagnetic hysteresis, IEEE Transactions on Magnetics, 24 (1988) 218-221.
[22] A.M. Ulanov, G.V. Lazutkin, Description of an arbitrary multi-axial loading process for non-linear vibration isolators, Journal of Sound and Vibration, 203 (1997) 903-907.
[23] H.R. Ao, H.Y. Jiang, A.M. Ulanov, Dry friction damping characteristics of a metallic rubber isolator under two-dimensional loading processes, Modelling and Simulation in Materials Science and Engineering, 13 (2005) 609-620.
[24] J.F. Hou, H.B. Bai, D.W. Li, Damping capacity measurement of elastic porous wire-mesh material in wide temperature range, Journal of Materials Processing Technology, 206 (2008) 412-418.
[25] W. Li, K. Wang, D.M. Zhu, J.L. Wu, Parameter identification of nonlinear hysteretic systems based on genetic algorithm, Journal of Vibration and Shock, 19 (2000) 9-11.
[26] D.W. Li, H.B. Bai, J.C. Yang, Y.J. Liu, Modelling of a nonlinear system with hysteresis characteristics, Chinese Journal of Mechanical Engineering, 41 (2005) 205-214.
[27] Q. Liu, K.X. He, X.Q. Huang, Experimental study on a vibration isolation system with viscous damping and cubic nonlinear stiffness, Journal of Vibration and Shock, 26 (2007) 135-142.

[28] Y.Y. Li, X.Q. Huang, W.X. Mao, Investigation on the computional method of vibration response of five power nonlinear dry friction system for metallic rubber, Journal of Astronaustics, 26 (2005) 620-624.
[29] T.K. Caughey, Sinusoidal excitation of a system with bilinear hysteresis, Journal of Applied Mechanics, 27 (1960) 640-643.
[30] W.D. Iwan, L.D. Lutes, Response of bilinear hysteretic system to stationary random excitation, Journal of the Acoustical Society of America, 43 (1968) 545-552.
[31] M.J.D. Powell, Approximation Theory and Methods, Cambridge University Press, Cambridge, 1981.
[32] H.Y. Hu, Y.F. Li, Parametric identification of nonlinear vibration isolators with memory, Journal of Vibration Engineering, 2 (1989) 17-27.
[33] W.J. Cody, A survey of practical rational and polynomial approximation of functions, Siam Review, 12 (1970) 400-423.
[34] S.D. Silvey, Multicollinearity and imprecise estimation, Journal of the Royal Statistical Society Series B-Statistical Methodology, 31 (1969) 539-552.
[35] Q. Rong, J. Shi, D. Ceglarek, Adjusted least square approach for diagnosis of ill-conditioned compliant assemblies, Journal of Manufacturing Science and Engineering, 123 (2001) 453-461.
[36] S.A. Billings, M.J. Korenberg, S. Chen, Identification of nonlinear output-affine systems using an orthogonal least squares algorithm, International Journal of Systems Science, 19 (1988) 1559-1568.
[37] S.A. Billings, K.M. Tsang, Spectral analysis for nonlinear systems: 1. Parametric nonlinear spectral analysis, Mechanical Systems and Signal Processing, 3 (1989) 319-339.
[38] S. Chen, S.A. Billings, W. Luo, Orthogonal least squares methods and their application to nonlinear system identification, International Journal of Control, 50 (1989) 1873-1896.
[39] H.L. Wei, D.Q. Zhu, S.A. Billings, M.A. Balikhin, Forecasting the geomagnetic activity of the dst index using multiscale radial basis function networks, Advances in Space Research, 40 (2007) 1863-1870.
[40] S.A. Billings, H.L. Wei, An adaptive orthogonal search algorithm for model subset selection and non-linear system identification, International Journal of Control, 81 (2008) 714-724.
[41] X. Hong, C.J. Harris, Nonlinear model structure detection using optimum experimental design and orthogonal least squares, IEEE Transactions on Neural Networks, 12 (2001) 435-439.
[42] X. Hong, P.M. Sharkey, K. Warwick, A robust nonlinear identification algorithm using press statistic and forward regression, IEEE Transactions on Neural Networks, 14 (2003) 454-458.
[43] X. Hong, R.J. Mitchell, S. Chen, C.J. Harris, K. Li, G.W. Irwin, Model selection approaches for non-linear system identification: A review, International Journal of Systems Science, 39 (2008) 925-946.
[44] G.A.F. Seber, Linear Regression Analysis, John Wiley & Sons, New York, 1977.
[45] S.A. Billings, L.A. Aguirre, Effects of the sampling time on the dynamics and identification of nonlinear models, International Journal of Bifurcation and Chaos, 5 (1995) 1541-1556.
[46] J. Walters-Williams, Y. Li, Estimation of mutual information: A survey, Proceedings of the Fourth International Conference on Rough Sets and Knowledge Technology, Gold Coast, Australia, (2009) 389-396.
[47] A.M. Fraser, H.L. Swinney, Independent coordinates for strange attractors from mutual information, Physical Review A, 33 (1986) 1134-1140.
[48] I.J. Leontaritis, S.A. Billings, Input output parametric models for non-linear systems .1. Deterministic non-linear systems, International Journal of Control, 41 (1985) 303-328.

[49] I.J. Leontaritis, S.A. Billings, Input output parametric models for non-linear systems .2. Stochastic non-linear systems, International Journal of Control, 41 (1985) 329-344.
[50] S. Chen, S.A. Billings, Representations of non-linear systems - the NARMAX model, International Journal of Control, 49 (1989) 1013-1032.
[51] S.A. Billings, W.S.F. Voon, Correlation based model validity tests for nonlinear models, International Journal of Control, 44 (1986) 235-244.
[52] S.A. Billings, S. Chen, M.J. Korenberg, Identification of mimo non-linear systems using a forward-regression orthogonal estimator, International Journal of Control, 49 (1989) 2157-2189.
[53] D.D. Rizos, S.D. Fassois, Presliding friction identification based upon the Maxwell slip model structure, Chaos, 14 (2004) 431-445.
[54] H. Wang, J.A. Rongong, G.R. Tomlinson, J. Hong, Nonlinear static and dynamic properties of metal rubber dampers, Proceedings of International Conference on Noise and Vibration Engineering, Leuven, Belgium, (2010) 1301-1315.
[55] B.J. Lazan, Damping of Materials and Members in Structural Mechanics, Pergamon Press, London, 1968.

Table 1
The process model terms selected and ranked in order of the significance by the OLS algorithm together with the coefficient of each term and its corresponding ERR.

index	2	4	3	14	12	1	13
terms	k_3	c_1	k_5	a_3	a_1	k_1	a_2
ERR(%)	97.79	1.7311	0.21183	0.068814	0.020381	0.014412	0.0072245

index	5	11	15	8
terms	c_3	a_0	a_4	c_9
ERR(%)	0.013776	0.047761	0.0058883	0.014322

Table 2
The noise model terms selected and ranked in order of the significance by the OLS algorithm together with the ERR of each term.

index	15	30	24	39	12	27	36
terms	ξ_{i-2}	ξ_{i-7}	ξ_{i-5}	ξ_{i-10}	ξ_{i-1}	ξ_{i-6}	ξ_{i-9}
ERR(%)	2.9146e-4	2.9973e-4	8.2645e-5	4.1408e-5	1.3563e-5	2.4027e-5	9.5013e-6

index	18	51
terms	ξ_{i-3}	ξ_{i-14}
ERR(%)	1.2102e-5	4.7711e-6

Fig. 1. A cylindrical metal rubber (MR) specimen.

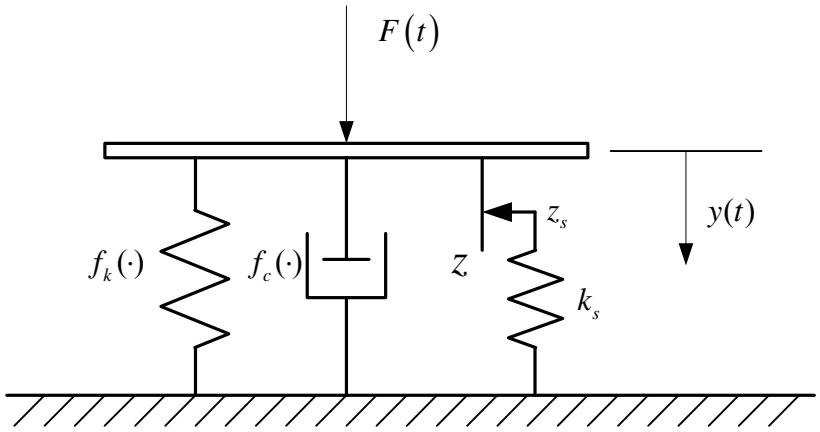

Fig. 2. A semi-constitutive mechanical model of metal rubber.

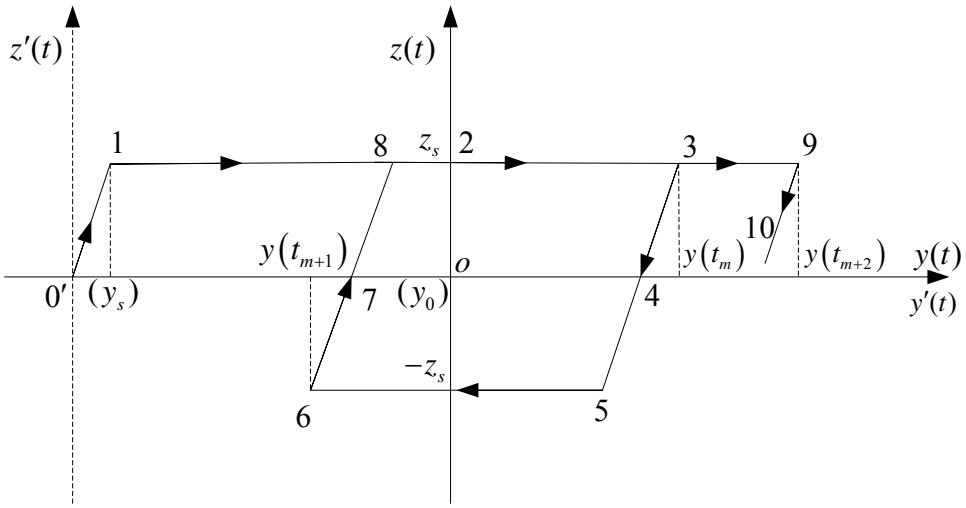

Fig. 3. The transient-state bilinear hysteresis loop of a metal rubber under static preload (from state **0′** to state 2) and then subject to a harmonic excitation (after state 2).

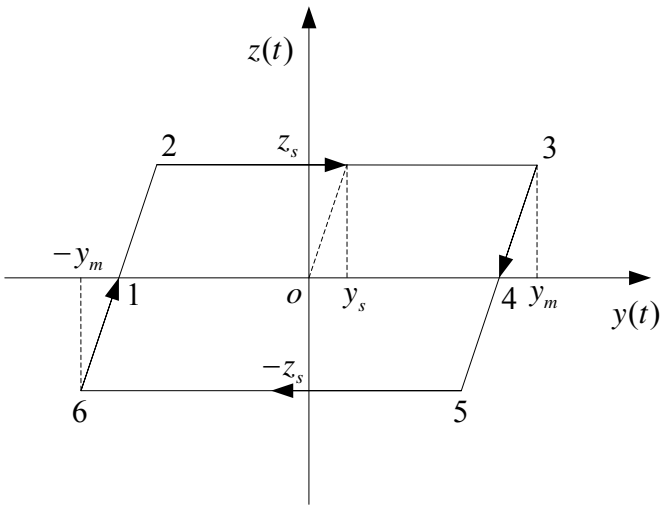

Fig. 4. The steady-state bilinear hysteresis loop of a metal rubber.

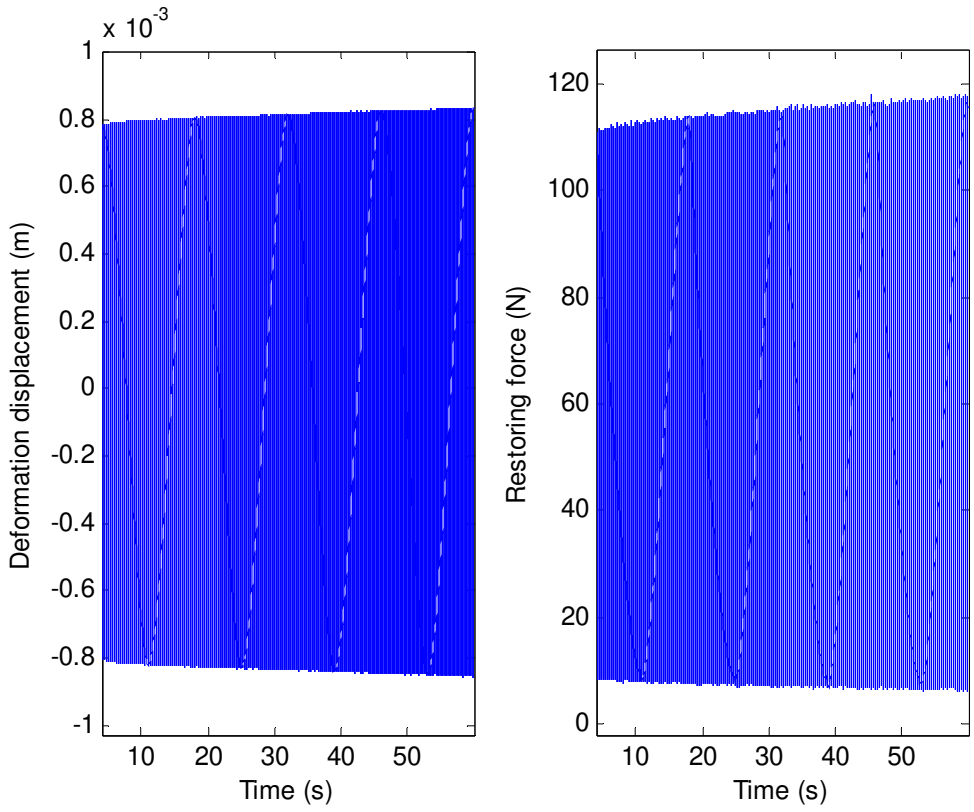

Fig. 5. The displacement excitation and corresponding restoring force signals collected by a data acquisition system.

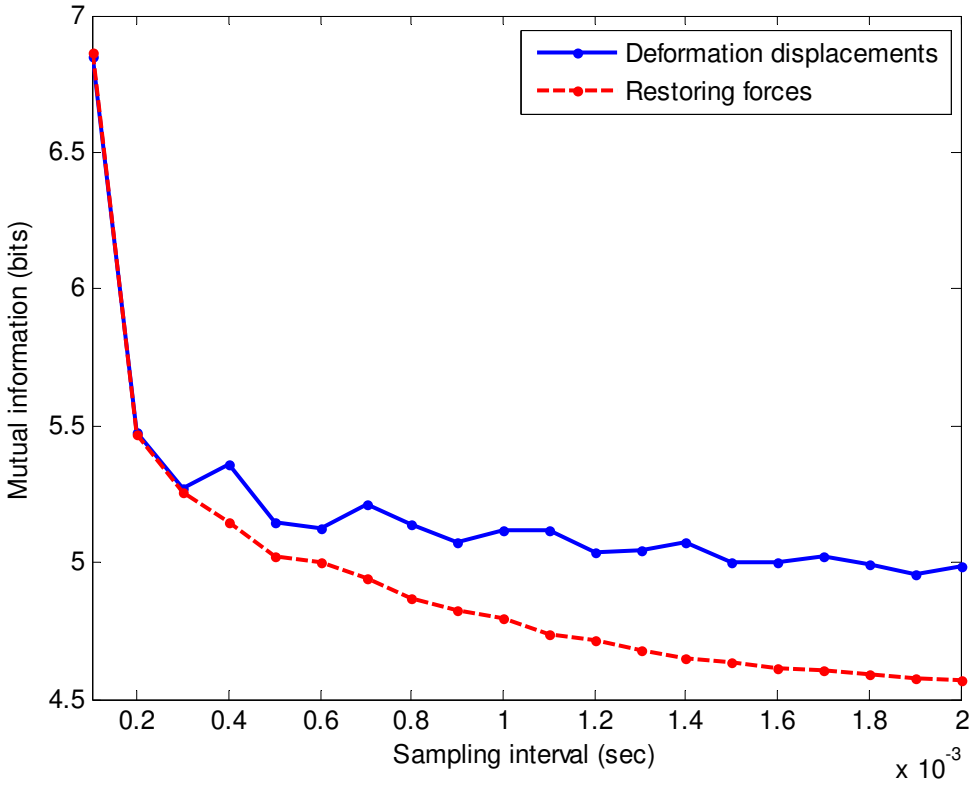

Fig. 6. The mutual information of the displacement and corresponding force signals versus the sampling interval.

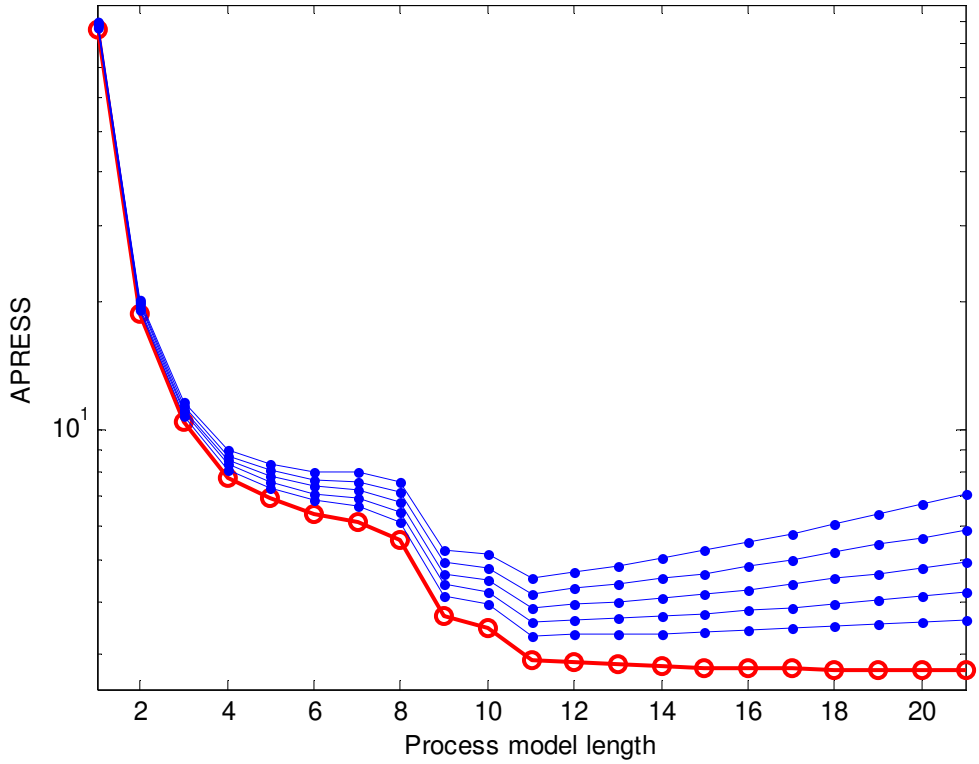

Fig. 7. The APRESS statistic versus the process model length: the lines from bottom to the top correspond to $\boldsymbol{\alpha} = \mathbf{0}, \mathbf{1}, \mathbf{1.5}, \cdots, \mathbf{3}$. The bottom line with circles, corresponding to $\boldsymbol{\alpha} = \mathbf{0}$, indicates the mean-squared-errors (MSE).

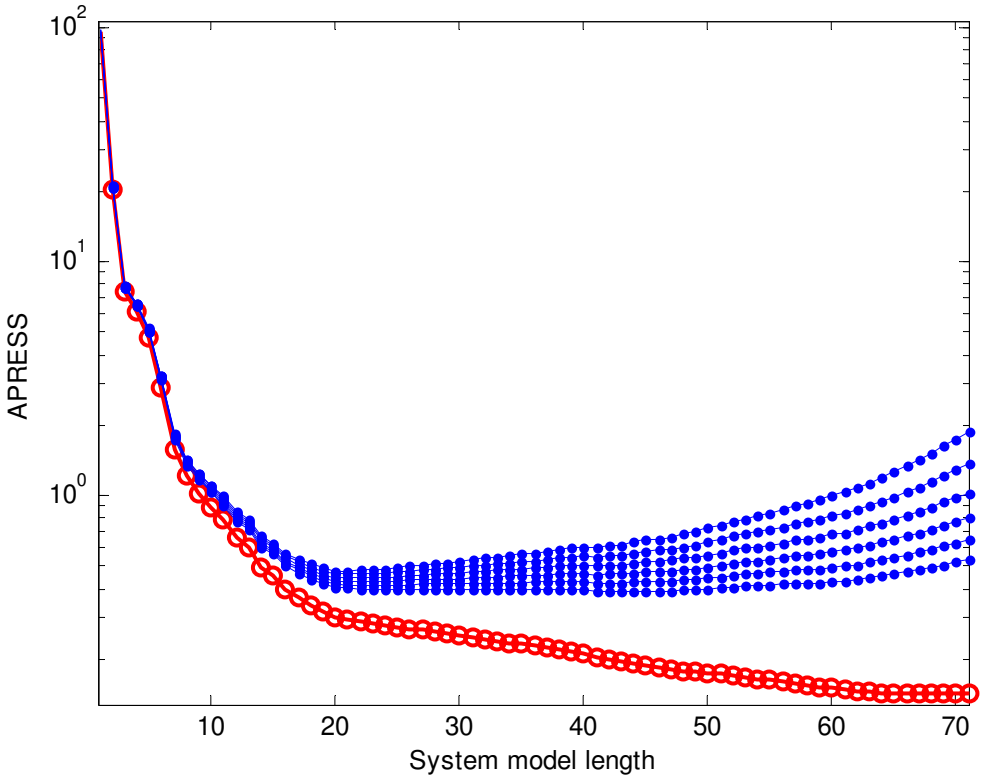

Fig. 8. The APRESS statistic versus the system model length: the lines from bottom to the top correspond to $\boldsymbol{\alpha} = \mathbf{0}, \mathbf{1}, \mathbf{1.1}, \cdots, \mathbf{1.5}$. The bottom line with circles, corresponding to $\boldsymbol{\alpha} = \mathbf{0}$, indicates the mean-squared-errors (MSE).

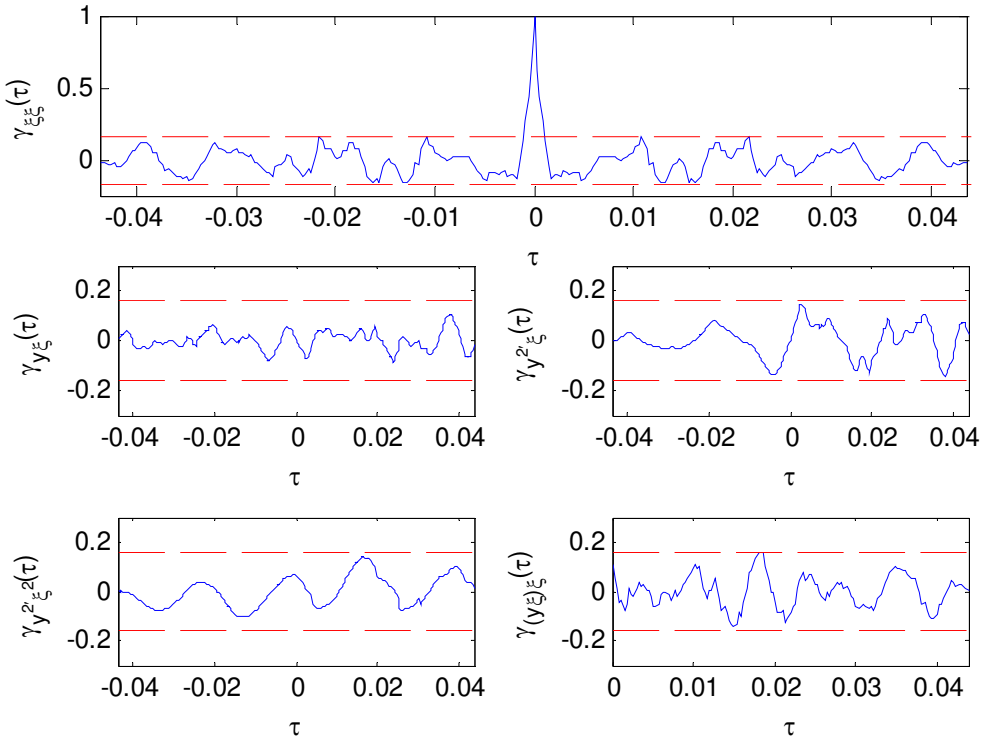

Fig. 9. Model validity tests for the identified model: (a) $\boldsymbol{\gamma_{\xi\xi}(\tau)}$; (b) $\boldsymbol{\gamma_{y\xi}(\tau)}$; (c) $\boldsymbol{\gamma_{y^{2\prime}\xi}(\tau)}$; (d) $\boldsymbol{\gamma_{y^{2\prime}\xi^2}(\tau)}$; (e) $\boldsymbol{\gamma_{(y\xi)\xi}(\tau)}$. The two horizontal lines in each graph indicate the 95% confidence intervals of the associated correlation functions.

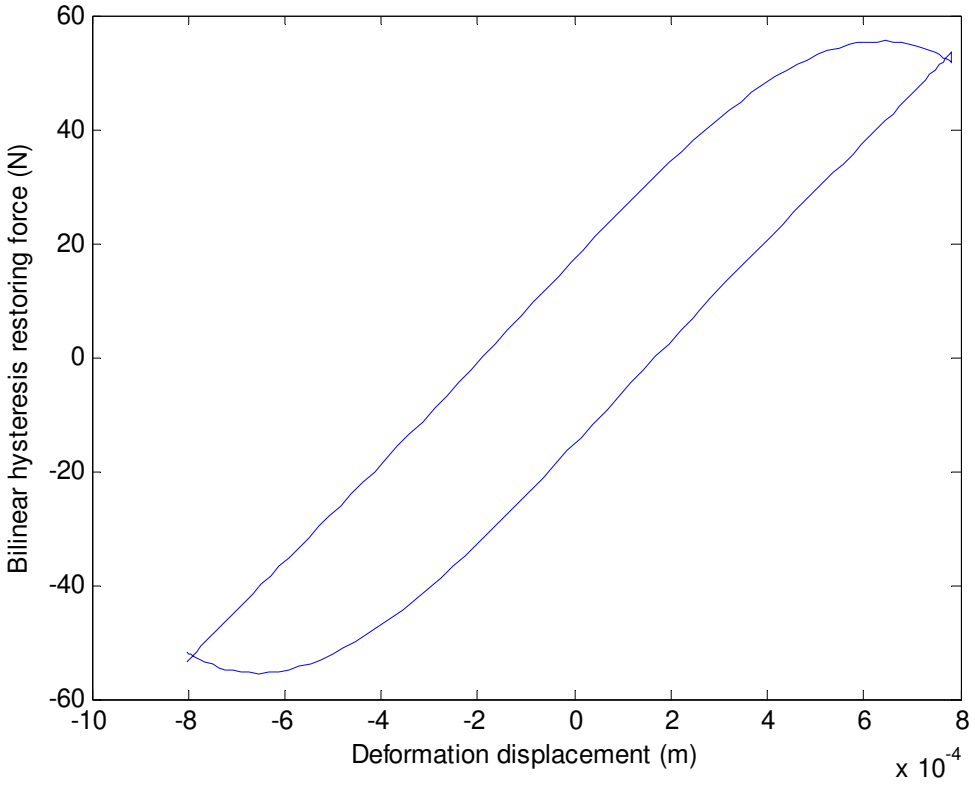

Fig. 10. The bilinear hysteresis loop produced by the identified model of a metal rubber specimen.

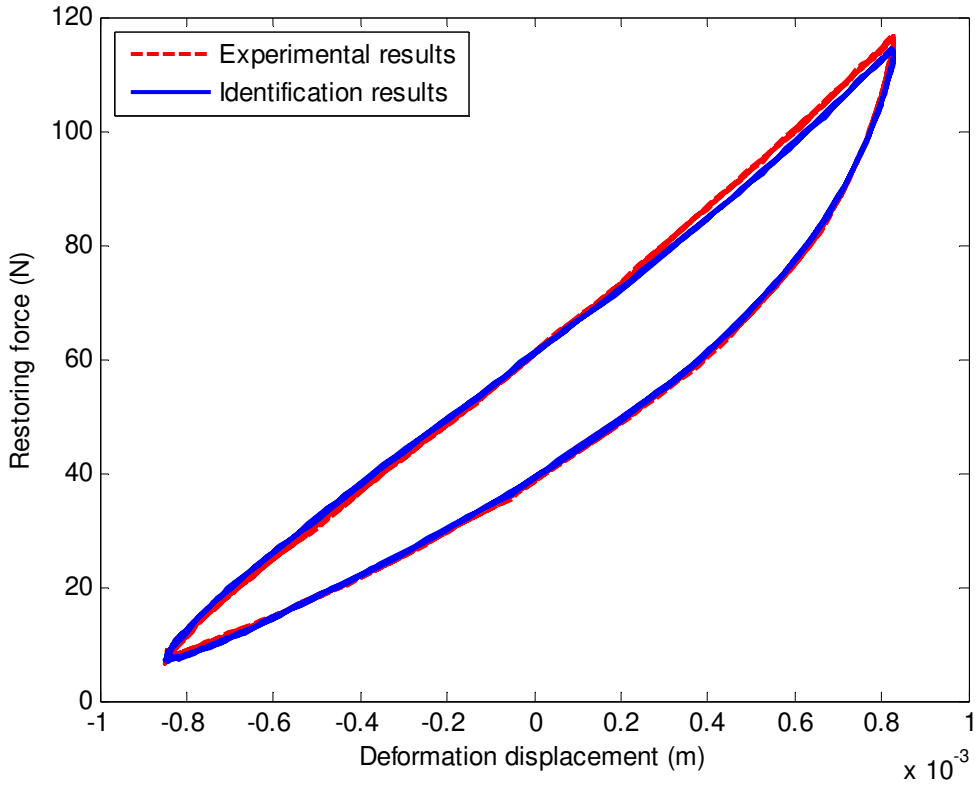

Fig. 11. A comparison of the hysteresis loop predicted by the identified model of a metal rubber specimen with that plotted directly from the corresponding experimental measurements.

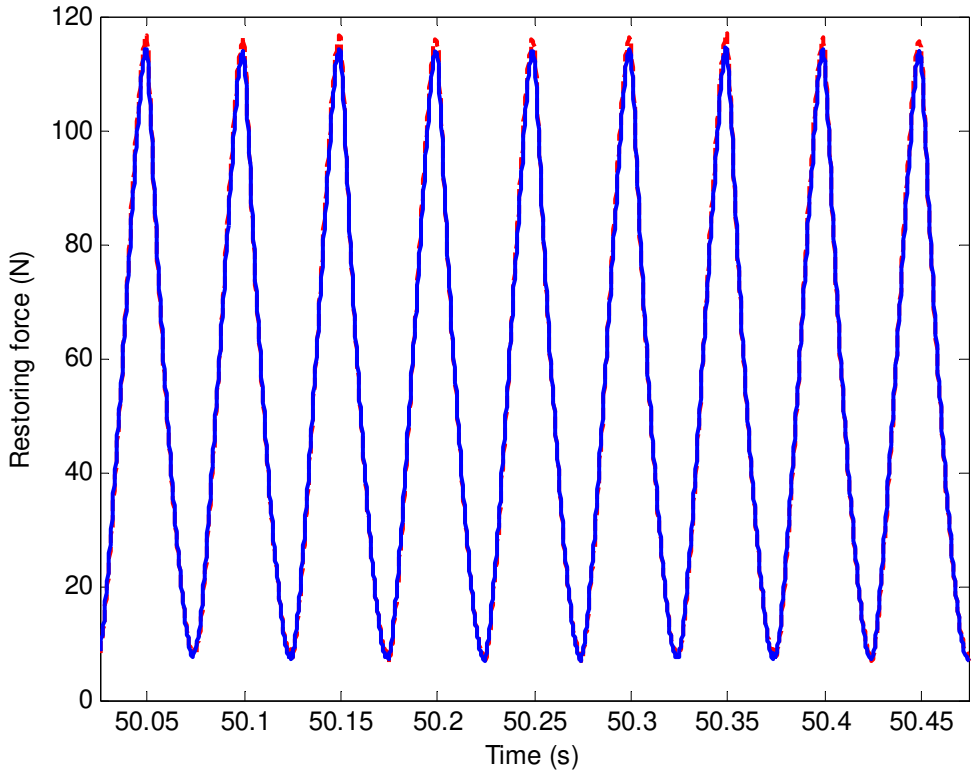

Fig. 12. A comparison of the restoring force predicted by the identified model of a metal rubber specimen with the corresponding experimental measurements. Solid lines: the model predicted ouput; dashed lines: original measurements.