



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/74595/>

Monograph:

Wei, H.L., Lang, Z.Q. and Billings, S.A. (2006) Identification of Nonlinear Parameter-Dependent Common-Structured models to accommodate varying experimental conditions and design parameter properties. Research Report. ACSE Research Report no. 935 . Automatic Control and Systems Engineering, University of Sheffield

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Identification of Nonlinear Parameter-Dependent Common-Structured Models to Accommodate Varying Experimental Conditions and Design Parameter Properties

Hua-Liang Wei, Zi-Qiang Lang and Stephen A. Billings

{ S.Billings@Shef.ac.uk, W.Hualiang@Shef.ac.uk, Z.Lang@Sheffield.ac.uk }



Research Report No. 935

Department of Automatic Control and Systems Engineering
The University of Sheffield
Mappin Street, Sheffield,
S1 3JD, UK

8 August 2006

Identification of Nonlinear Parameter-Dependent Common-Structured Models to Accommodate Varying Experimental Conditions and Design Parameter Properties

Hua-Liang Wei, Zi-Qiang Lang and Stephen A. Billings

Department of Automatic Control and Systems Engineering, University of Sheffield
Mappin Street, Sheffield, S1 3JD, UK

S.Billings@Sheffield.ac.uk, W.Hualiang@Sheffield.ac.uk, Z.Lang@Sheffield.ac.uk

8 August 2006

Abstract: This study considers the identification problem for a class of nonlinear parameter-varying systems associated with the following scenario: the system behaviour depends on some specifically prescribed parameter properties, which are adjustable. To understand the effect of the varying parameters, several different experiments, corresponding to different parameter properties, are carried out and different data sets are collected. The objective is to find, from the available data sets, a common parameter-dependent model structure that best fits the adjustable parameter properties for the underlying system. An efficient common model structure selection (CMSS) algorithm, called the extended forward orthogonal regression (EFOR) algorithm, is proposed to select such a common model structure. Several examples are presented to illustrate the application and the effectiveness of the new identification approach.

Keywords: Forward orthogonal regression, nonlinear system identification, parameter-dependent model.

1. Introduction

The task of system identification is to deduce, from observed data, a model (or a set of models) that can be used for specific purposes such as system analysis, control and prediction. A system model is defined by two properties: the model structure and the associated model parameters. Traditionally, the identification procedure for dynamical systems often merely involves a single (training) data set, corresponding to one specific experimental situation, and the resultant final model is thus experimentally specific; both the model structure and the associated model parameters are fixed. In the real world, however, parameters in a given common model structure for a dynamical system may be required to be changeable to meet varying situations caused by the variation of either internal or exogenous parameters (Billings and Voon 1987). For example, typical mass-spring-damper vibration systems can be described using a common model structure, in the form of second order ordinary differential equation (ODE), where the associated parameters are changeable and determined by the three elements of mass, spring and damper. The ODE model for the vibration system can be viewed as

a special case of internal-parameter-dependent (IPD) models, where the dynamical behaviour of the model is directly affected by changes of the internal parameters.

In terms of system identification, the task for general IPD model identification problems can be summarized as follows. By setting the process internal parameters to be different values, a number of experiments are carried out on the same system, and different data sets are obtained, corresponding to different parameter properties. The objective is to find from the available data a parsimonious common model structure, to accommodate all the different parameter properties by best fitting all the data sets using the common structured model, with varying process internal parameters. This is different from conventional parameter-varying models, where process internal parameters are assumed to be time-varying.

There are many other cases where parameter-dependent models are desirable. Consider the following scenario. In typical normal operating conditions, the dynamical behaviour of an underlying system is often determined by the system model structure and the associated process internal parameters. In many cases, however, several external parameters, for example temperature, pressure intensity, light illumination, geometry shape and size, etc., may also indirectly affect the dynamical behaviour of the system, via the associated process internal parameters. In order to fully understand the mechanisms of the underlying dynamics under different operating conditions, several experiments, with respect to different exogenous parameter properties, may be required. The task of external-parameter-dependent (EPD) model identification is to find a best common model structure based on the available data, to accommodate the effects of all the external parameters, by best fitting all the data sets using the common structured model, with adjustable process internal parameters. This is related to but distinct from the concepts of spatial piecewise linear models and models with single dependent parameters (Billings and Voon 1987).

The objective of this study is to present a unified parameter-dependent common-structured (PDCS) modelling framework for handling the IPD and EPD identification problem, where the selection of the common model structure is the critical stage in the procedure. An efficient common model structure selection (CMSS) algorithm, called the extended forward orthogonal regression (EFOR) algorithm, is developed in this study to select a common model structure based on several data sets collected from different experiments. Once the common model structure has been obtained, relevant model parameters corresponding to each individual experimental condition can then be calculated based on the available individual data sets. The novel study of common model structure identification is very useful for engineering system design and control, where only a fixed common model structure is involved but with adjustable process internal parameters. A PDCS model can be used to analyse the effects of varying parameter properties on the performance of the behaviour of the underlying dynamical systems without carrying out experiments on the real system. This will save time and money spent on real system experiments.

For convenience of description, in the following all non-internal parameters, including different experimental conditions, will be referred to as external or exogenous parameters. Specifically prescribed parameters, either internal or exogenous, will be called experiment parameters or design parameters. This work involves several abbreviations and these are collected in the appendix to facilitate reading of text.

2. The concept of the parameter-dependent commonly structured model

The parameter-dependent common-structured (PDCS) model is defined as below

$$y(t) = f(y(t-1), \dots, y(t-n_y), u(t-1), \dots, u(t-n_u), \boldsymbol{\theta}(\boldsymbol{\xi})) + e(t) \quad (1)$$

where

- the nonlinear mapping f is often unknown and needs to be identified from given observations of the input $u(t)$ and the output $y(t)$; n_u and n_y are the maximum input and output lags; $e(t)$ is the model prediction error, which can often be treated as an independent zero mean noise sequence providing that the function f gives a sufficient description of the system.
- $\boldsymbol{\theta}(\boldsymbol{\xi}) \in \Theta$ represents an internal parameter vector, which is a function of the external parameter set $\boldsymbol{\xi} \in \Omega$, where Θ and Ω are the internal and external parameter sets, respectively. The external parameter set $\boldsymbol{\xi}$ may not explicitly appear in the model but does indirectly affect the dynamical behaviour of the model through the internal parameter $\boldsymbol{\theta}$.

Assume that a total of K experiments, corresponding to K different cases of exogenous parameter properties, $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_K$, have been completed on the same system, model (1) can then be expressed in a more explicit form as

$$y(t) = \begin{cases} f^1(y(t-1), \dots, y(t-n_y), u(t-1), \dots, u(t-n_u), \boldsymbol{\theta}_1), & \text{s. t. } \boldsymbol{\xi}_1 \\ f^2(y(t-1), \dots, y(t-n_y), u(t-1), \dots, u(t-n_u), \boldsymbol{\theta}_2), & \text{s. t. } \boldsymbol{\xi}_2 \\ \vdots & \vdots \\ f^K(y(t-1), \dots, y(t-n_y), u(t-1), \dots, u(t-n_u), \boldsymbol{\theta}_K), & \text{s. t. } \boldsymbol{\xi}_K \end{cases} \quad (2)$$

where $f^i(\cdot)$ ($i=1, \dots, K$) are different linear or nonlinear functions that share a common structure in representation. The symbol ‘s.t. $\boldsymbol{\xi}_i$ ’ means the individual model is subject to the exogenous parameter $\boldsymbol{\xi}_i$. Clearly, if $K=1$, the PDCS model (2) will reduce to the traditional NARX (*Nonlinear AutoRegressive with eXogenous inputs*) model (Leontaritis and Billings 1985, Pearson 1999).

Note that the PDCS model considered here is different from conventional time-varying or parameter-varying models, where process internal parameters are assumed to be time-varying. The PDCS model is also different from the traditional multi-input and multi-output (MIMO) model

structure, where each subsystem model may not need to share the same common model structure, and which often involves one single data set.

3. The identification of the commonly structured model

3.1 The linear-in-the-parameters regression model

The nonlinear mapping f in (1) can be constructed using a variety of local or global basis functions including polynomials, kernel functions, splines, radial basis functions, neural networks and wavelets (Chen and Billings 1990, Kavli 1993, Berger 1994, Wu and Harris 1997, Pearson 1999, Liu 2001, Harris et al. 2002, Chen et al. 2005, Billings and Wei 2005). One of the most popular representations is the polynomial model (Leontaritis and Billings 1985), which takes the form below

$$y(t) = \theta_0 + \sum_{i_1=1}^d f_{i_1}(x_{i_1}(t)) + \sum_{i_1=1}^d \sum_{i_2=i_1}^d f_{i_1 i_2}(x_{i_1}(t), x_{i_2}(t)) + \dots + \sum_{i_1=1}^d \dots \sum_{i_\ell=i_{\ell-1}}^d f_{i_1 i_2 \dots i_\ell}(x_{i_1}(t), x_{i_2}(t), \dots, x_{i_\ell}(t)) + e(t) \quad (3)$$

where $\theta_{i_1 i_2 \dots i_m}$ are parameters, $d = n_y + n_u$ and

$$f_{i_1 i_2 \dots i_m}(x_{i_1}(t), x_{i_2}(t), \dots, x_{i_m}(t)) = \theta_{i_1 i_2 \dots i_m} \prod_{k=1}^m x_{i_k}(t), \quad 1 \leq m \leq \ell \quad (4)$$

$$x_k(t) = \begin{cases} y(t-k) & 1 \leq k \leq n_y \\ u(t-(k-n_y)) & n_y + 1 \leq k \leq d \end{cases} \quad (5)$$

The degree of a multivariate polynomial is defined as the highest order among the terms, for example, the degree of the polynomial $h(x_1, x_2, x_3) = a_1 x_1^4 + a_2 x_2 x_3 + a_3 x_1^2 x_2 x_3^2$ is determined by the term $x_1^2 x_2 x_3^2$ and thus $\ell = 2+1+2=5$. Similarly, a NARX model with a nonlinear degree ℓ means that the order of each term in the model is not higher than ℓ .

A NARX model constructed using basis function expansions can often be expressed using a linear-in-the-parameters form

$$y(t) = \sum_{m=1}^M \theta_m \phi_m(t) + e(t) \quad (6)$$

where M is the total number of candidate regressors, $\phi_m(t) = \phi_m(\mathbf{x}(t))$ ($m=1, 2, \dots, M$) are the model terms generated in some way from the ‘input’ (predictor) vector $\mathbf{x}(t) = [x_1(t), \dots, x_d(t)]^T$ defined by Eq.(5), $\theta_m = \theta_m(\xi)$ are the model parameters, and ξ is a known collection of external parameters.

3.2 The multiple regression model

Assume that a total of K experiments, corresponding to K different cases of the experiment parameter properties, have been carried out on the same system, and K different data sets, with respect to the K experiments, have been obtained. Also, assume that a common model structure, with the form of (6), can best fit all the data sets. Denote the input and the output sequence for the k th experiment by $\{u_k(t)\}_{t=1}^{N_k}$ and $\{y_k(t)\}_{t=1}^{N_k}$, respectively, for $k=1,2,\dots, K$. The k th predictor vector is thus given by $\mathbf{x}_k(t) = [x_{k,1}(t), \dots, x_{k,d}(t)]^T = [y_k(t-1), \dots, y_k(t-n_y), u_k(t-1), \dots, u_k(t-n_u)]^T$. It is assumed that all the K data sets can be represented using a common model structure, with a different parameter set, deduced from the initial candidate regression model below

$$y_k(t) = \sum_{m=1}^M \beta_{k,m} \phi_m(\mathbf{x}_k(t)) + e_k(t) = \sum_{m=1}^M \beta_{k,m} \phi_{k,m}(t) + e_k(t) \quad (7)$$

This can be expressed using a compact matrix form

$$\mathbf{y}_k = \Phi_k \boldsymbol{\beta}_k + \mathbf{e}_k \quad (8)$$

where $\mathbf{y}_k = [y_k(1), \dots, y_k(N_k)]^T$, $\boldsymbol{\beta}_k = [\beta_{k,1}, \dots, \beta_{k,M}]^T$, $\mathbf{e}_k = [e_k(1), \dots, e_k(N_k)]^T$, and $\Phi_k = [\boldsymbol{\phi}_{k,1}, \dots, \boldsymbol{\phi}_{k,M}]$ with $\boldsymbol{\phi}_{k,m} = [\phi_{k,m}(1), \dots, \phi_{k,m}(N_k)]^T$ for $k=1,2, \dots, K$ and $m=1,2,\dots, M$.

For large lags n_y and n_u , the regression model (7) often involves a large number of candidate model terms, even if the nonlinear degree ℓ is not very high, say $\ell=2$ or $\ell=3$. Experience has shown that an initial candidate model with a large number of candidate model terms can often be drastically reduced by including in the final model only the effectively selected significant model terms. Furthermore, a simple concise model is usually desirable for practical applications including system analysis, design, control and prediction. This is one of the motivations of the present study to select significant model terms to form a parsimonious common model structure.

3.3 The extended forward orthogonal regression algorithm

A new common model structure selection (CMSS) algorithm, called the extended forward orthogonal regression (EFOR) algorithm, which is generalized from the standard orthogonal least squares (OLS) algorithm (Billings et al. 1989, Chen et al. 1989) and the recently developed forward orthogonal regression (FOR) algorithm (Wei and Billings 2006, Wei et al. 2006), will be designed for the PDCS identification problem.

Let $I = \{1,2,\dots, M\}$. Denote by $D = \{\phi_m : m \in I\}$ the dictionary of candidate model terms for an initially chosen candidate common model structure that fits to all the K regression models given by (7). For the k th data set, the dictionary D can be used to form a dual dictionary $V_k = \{\boldsymbol{\phi}_{k,m} : m \in I\}$, where the m th candidate basis vector $\boldsymbol{\phi}_{k,m}$ is formed by the m th candidate model term $\phi_m \in D$, in the sense

that $\boldsymbol{\phi}_{k,m} = [\phi_m(\mathbf{x}_k(1)), \dots, \phi_m(\mathbf{x}_k(N_i))]^T$ ($k=1,2, \dots, K$). The common model structure selection problem is equivalent to finding, from I , a subset of indices, $I_n = \{i_m : m = 1,2, \dots, n, i_m \in I\}$ where $n \leq M$, so that \mathbf{y}_i ($i=1,2, \dots, K$) can be approximated using a linear combination of $\boldsymbol{\phi}_{i_1}, \boldsymbol{\phi}_{i_2}, \dots, \boldsymbol{\phi}_{i_n}$ as

$$\mathbf{y}_i = \theta_{i,1} \boldsymbol{\phi}_{i,i_1} + \dots + \theta_{i,n} \boldsymbol{\phi}_{i,i_n} + \mathbf{e}_i \quad (9)$$

Following Billings et al. (1989) and Chen et al. (1989), a squared correlation coefficient will be used to measure the dependency between two associated random vectors. The squared correlation coefficient between two vectors \mathbf{x} and \mathbf{y} of size N is defined as

$$C(\mathbf{x}, \mathbf{y}) = \frac{(\mathbf{x}^T \mathbf{y})^2}{(\mathbf{x}^T \mathbf{x})(\mathbf{y}^T \mathbf{y})} = \frac{(\sum_{i=1}^N x_i y_i)^2}{\sum_{i=1}^N x_i^2 \sum_{i=1}^N y_i^2} \quad (10)$$

The squared correlation coefficient is closely related to the error reduction ratio (ERR) criterion defined in the standard orthogonal least squares (OLS) algorithm for model structure selection. A comprehensive discussion on OLS-ERR algorithm can be found in Billings et al. (1989) and Chen et al. (1989).

Let $\mathbf{r}_{k,0} = \mathbf{y}_k$ ($k=1,2, \dots, K$). For $k=1,2, \dots, K$ and $j=1,2, \dots, M$, calculate $c_1(k, j) = C(\mathbf{y}_k, \boldsymbol{\phi}_{k,j})$, and define

$$s_1 = \arg \max_{1 \leq j \leq M} \left\{ \frac{1}{K} \sum_{k=1}^K c_1(k, j) \right\} \quad (11)$$

The first significant common model term can then be selected as the s_1 th element, ϕ_{s_1} , in the dictionary D . Accordingly, the first significant basis vector for the k th regression model is thus $\boldsymbol{\alpha}_{k,1} = \boldsymbol{\phi}_{k,s_1}$, and the first associated orthogonal basis vector can be chosen as $\mathbf{q}_{k,1} = \boldsymbol{\phi}_{k,s_1}$. The model residual for the k th regression model, related to the first step search, is given as

$$\mathbf{r}_{k,1} = \mathbf{r}_{k,0} - \frac{\mathbf{r}_{k,0}^T \mathbf{q}_{k,1}}{\mathbf{q}_{k,1}^T \mathbf{q}_{k,1}} \mathbf{q}_{k,1} \quad (12)$$

Notice that $c_1(k, s_1)$ can be viewed as the error reduction ratio (ERR) that is introduced by including the first basis vector $\boldsymbol{\alpha}_{k,1} = \boldsymbol{\phi}_{k,s_1}$ into the k th regression model. The criterion (11), by maximizing the sum of the ERR values relative to all the K data sets, guarantees that the variation of the outputs in all the K data sets can be explained by including the model term ϕ_{s_1} , with the highest percentage, compared with selecting any other candidate model term $\phi \in D = \{\phi_m : m \in I\}$. The quantity $\text{AERR}(1) = (1/K) \sum_{k=1}^K c_1(k, s_1)$ is referred to as the first average error reduction ratio (AERR).

In general, the m th significant model term ϕ_{s_m} can be chosen as follows. Assume that at the $(m-1)$ th step, $(m-1)$ significant model terms, $\phi_1, \phi_2, \dots, \phi_{m-1}$, have been selected. Let $\mathbf{a}_{k,1}, \mathbf{a}_{k,2}, \dots, \mathbf{a}_{k,m-1}$ be the associated basis vectors for the k th regression model, and assume that the $(m-1)$ selected bases have been transformed into a new group of orthogonal bases $\mathbf{q}_{k,1}, \mathbf{q}_{k,2}, \dots, \mathbf{q}_{k,m-1}$ via some orthogonal transformation. Let

$$\mathbf{p}_{k,j}^{(m)} = \boldsymbol{\varphi}_{k,j} - \sum_{s=1}^{m-1} \frac{\boldsymbol{\varphi}_{k,j}^T \mathbf{q}_{k,s}}{\mathbf{q}_{k,s}^T \mathbf{q}_{k,s}} \mathbf{q}_{k,s}, \quad j \in J_m \quad (13)$$

where $J_m = \{j : 1 \leq j \leq M, j \neq s_t, 1 \leq t \leq m-1\}$. For $k=1, 2, \dots, K$ and $j \in J_m$, calculate $c_m(k, j) = C(\mathbf{y}_k, \mathbf{p}_{k,j}^{(m)})$, and define

$$s_m = \arg \max_{j \in J_m} \left\{ \frac{1}{K} \sum_{k=1}^K c_m(k, j) \right\} \quad (14)$$

The m th significant common model term can then be selected as the s_m th element, ϕ_{s_m} , in the dictionary D . Accordingly, the m th significant basis vector for the k th regression model is thus $\mathbf{a}_{k,m} = \boldsymbol{\varphi}_{k,s_m}$, and the first associated orthogonal basis vector can be chosen as $\mathbf{q}_{k,m} = \mathbf{p}_{k,s_m}^{(m)}$. The model residual for the k th regression model, related to the m th step search, is given as

$$\mathbf{r}_{k,m} = \mathbf{r}_{k,m-1} - \frac{\mathbf{r}_{k,m-1}^T \mathbf{q}_{k,m}}{\mathbf{q}_{k,m}^T \mathbf{q}_{k,m}} \mathbf{q}_{k,m} \quad (15)$$

Subsequent significant bases can be selected in the same way step by step. Notice that the $(m-1)$ basis vectors, $\mathbf{a}_{k,1}, \mathbf{a}_{k,2}, \dots, \mathbf{a}_{k,m-1}$ (respectively the associated orthogonalized bases, $\mathbf{q}_{k,1}, \mathbf{q}_{k,2}, \dots, \mathbf{q}_{k,m-1}$), by including the m th basis $\mathbf{a}_{k,m} = \boldsymbol{\varphi}_{k,s_m}$ (respectively the orthogonalized basis $\mathbf{q}_{k,m} = \mathbf{p}_{k,s_m}^{(m)}$), can explain the variation in the outputs of the K data sets with a higher percentage than by including any other candidate bases. The quantity $\text{AERR}(m) = (1/K) \sum_{k=1}^K c_m(k, s_m)$ is referred to as the m th average error reduction ratio (AERR).

From (15), the vectors $\mathbf{r}_{k,m}$ and $\mathbf{q}_{k,m}$ are orthogonal, thus

$$\|\mathbf{r}_{k,m}\|^2 = \|\mathbf{r}_{k,m-1}\|^2 - \frac{(\mathbf{r}_{k,m-1}^T \mathbf{q}_{k,m})^2}{\mathbf{q}_{k,m}^T \mathbf{q}_{k,m}} \quad (16)$$

By respectively summing (15) and (16) for m from 1 to n , yields

$$\mathbf{y}_k = \sum_{m=1}^n \frac{\mathbf{r}_{k,m-1}^T \mathbf{q}_{k,m}}{\mathbf{q}_{k,m}^T \mathbf{q}_{k,m}} \mathbf{q}_{k,m} + \mathbf{r}_{k,n} \quad (17)$$

$$\|\mathbf{r}_{k,n}\|^2 = \|\mathbf{r}_{k,n-1}\|^2 - \frac{(\mathbf{r}_{k,n-1}^T \mathbf{q}_{k,n})^2}{\mathbf{q}_{k,n}^T \mathbf{q}_{k,n}} = \|\mathbf{y}_k\|^2 - \sum_{m=1}^n \frac{(\mathbf{r}_{k,m-1}^T \mathbf{q}_{k,m})^2}{\mathbf{q}_{k,m}^T \mathbf{q}_{k,m}} \quad (18)$$

From (17) and (18), the model residual $\mathbf{r}_{k,n}$ can be used to form a criterion for model selection, and the search procedure will be terminated when the norm $\|\mathbf{r}_{k,n}\|^2$ satisfies some specified conditions. In the present study, an approximate minimum description length (AMDL) criterion developed by Saito (1994) and Antoniadis et al. (1997), on the basis of the Rissanen's MDL criterion (Rissanen 1984), will be used to determine the model size. For the case of single regression model, AMDL is defined as

$$\text{AMDL}(n) = 0.5 \log_2[\text{MSE}(n)] + \frac{1.5n \log_2 N}{N} = 0.5 \log_2 \left(\frac{\|\mathbf{r}_n\|^2}{N} \right) + \frac{1.5n \log_2 N}{N} \quad (19)$$

where MSE is the mean-square-error from the associated model, N is the length of the associated training data set, n is the number of model terms, and \mathbf{r}_n is the associated model residual.

The present study uses the following average AMDL as the criterion to determine the number of common model terms

$$\text{AAMD L}(n) = \frac{1}{K} \sum_{k=1}^K \text{AMDL}^{[k]}(n) \quad (20)$$

where $\text{AMDL}^{[k]}(n)$ is value for the AMDL criterion associated to the k th data set.

3.4 Parameter estimation

It is easy to verify that the relationship between the selected original bases $\mathbf{a}_{k,1}, \mathbf{a}_{k,2}, \dots, \mathbf{a}_{k,n}$ and the associated orthogonal bases $\mathbf{q}_{k,1}, \mathbf{q}_{k,2}, \dots, \mathbf{q}_{k,n}$, for the k th data set, is given by

$$\mathbf{A}_{k,n} = \mathbf{Q}_{k,n} \mathbf{R}_{k,n} \quad (21)$$

where $\mathbf{A}_{k,n} = [\mathbf{a}_{k,1}, \dots, \mathbf{a}_{k,n}]$, $\mathbf{Q}_{k,n}$ is an $N_k \times n$ matrix with orthogonal columns $\mathbf{q}_{k,1}, \mathbf{q}_{k,2}, \dots, \mathbf{q}_{k,n}$, and $\mathbf{R}_{k,n}$ is an $n \times n$ unit upper triangular matrix whose entries u_{ij} ($1 \leq i \leq j \leq n$) are calculated during the orthogonalization procedure. The unknown parameter vector, denoted by $\boldsymbol{\theta}_{k,n} = [\theta_{k,1}, \dots, \theta_{k,n}]^T$, for the model with respect to the original bases (similar to (9)), can be calculated from the triangular equation $\mathbf{R}_{k,n} \boldsymbol{\theta}_{k,n} = \mathbf{g}_{k,n}$ with $\mathbf{g}_{k,n} = [g_{k,1}, \dots, g_{k,n}]^T$, where $g_m = (\mathbf{r}_{k,m-1}^T \mathbf{q}_{k,m}) / (\mathbf{q}_{k,m}^T \mathbf{q}_{k,m})$ for $m=1, 2, \dots, n$.

3.5 A general procedure for PDCS model identification

Common model structure selection is a critical step in PDCS identification. Once the common model structure has been identified, relevant model parameters for each individual data set can then be estimated, and the dependency of the model parameters on the associated experiment parameter properties can be deduced finally. The procedure for the identification of PDCS models can briefly be summarized below:

- Collect K different data sets from K experiments.
- Select common model terms using the new EFOR-CMSS algorithm.
- Estimate relevant model parameters for each individual case of experiment parameters.
- Deduce the dependence of the model parameters on the associated experiment parameter set.

4. Applications

Three examples, one for a forced mass-spring-damper system and two for real data sets, are presented to illustrate the application of the new PDCS model identification procedure.

4.1 A forced mass-spring-damper system

The mass-spring-damper system is described by the differential equation model below

$$m \frac{d^2 y}{dt^2} + c \frac{dy}{dt} + ky(t) = u(t) \quad (22)$$

where $u(t)$ is the force imposed on a mass m , $y(t)$ is the displacement of the mass relative to the equilibrium position, c is a damping coefficient, k is a constant relative to the spring stiffness. In this example, the parameters m and k were set to fixed values: $m=1[\text{kg}]$ and $k=100[\text{N/m}]$, and the coefficient c was the design parameter, which was assumed to be adjustable.

Four cases were considered for the design parameter c by respectively setting $c=2, 10, 20$, and 40 . Using these values, the model (22) was simulated four times, at each time the input signal u was chosen to be a random sequence uniformly distributed in $[-100,100]$, and 1000 input-output data points, with a sample interval $\Delta = 0.01$, were collected for each case. The objective was to find a common linear model structure for the four collected data sets with no a priori information other than the measured data. The predictor vector for the common-structured model was chosen to be $\mathbf{x}(t) = [x_1(t), \dots, x_4(t)]^T = [y(t-1), y(t-2), u(t-1), u(t-2)]^T$, and the initial candidate model structure was chosen to be

$$y(t) = \sum_{i=1}^4 \theta_i^0 x_i(t) + \sum_{i=1}^4 \sum_{j=i}^4 \theta_{i,j}^0 x_i(t) x_j(t) + e(t) \quad (23)$$

Note that the candidate model terms $x_i(t)x_j(t)$ were purposely included in the candidate model (23), to check whether the new EFOR-CMSS algorithm can correctly select the significant model terms. The new EFOR-CMSS algorithm was applied to the four data sets, and the associated result is shown in Table 1, where only the first 8 selected model terms are presented. The AAMD index in Table 1 clearly indicates that a common model structure, with 4 model terms, is preferred, and the final common-structured model was thus of the form

$$y(t) = \theta_1(c)y(t-1) + \theta_2(c)y(t-2) + \theta_3(c)u(t-1) + \theta_4(c)u(t-2) + e(t) \quad (24)$$

where the parameters θ_m ($m=1,2,3,4$) were viewed as a function of the adjustable coefficient c . The parameter estimates for the three model terms in (24), about the four data sets, are shown in Table 2. Assume that the parameter θ_m can be fitted using the adjustable coefficient c , with a polynomial of order 3 below

$$\theta_m(c) = \beta_{m,0} + \beta_{m,1}c + \beta_{m,2}c^2 + \beta_{m,3}c^3, \quad m=1,2,3,4, \quad (25)$$

The parameters $\beta_{m,n}$ were directly estimated using the results in Table 2. The PDCS model for the system (22) was

$$\begin{aligned} y(t) = & [1.990002 - 9.945500 \times 10^{-3}c + 4.919167 \times 10^{-5}c^2 - 1.366667 \times 10^{-7}c^3]y(t-1) \\ & + [-0.999993 + 9.995458 \times 10^{-3}c - 4.935854 \times 10^{-5}c^2 + 1.370833 \times 10^{-7}c^3]y(t-2) \\ & + [4.995833 \times 10^{-5} - 1.666667 \times 10^{-7}c + 4.166668 \times 10^{-10}c^2 - 3.070176 \times 10^{-18}c^3]u(t-1) \\ & + [4.995819 \times 10^{-5} - 3.329174 \times 10^{-7}c + 1.252084 \times 10^{-9}c^2 - 4.166670 \times 10^{-12}c^3]u(t-2) \\ & + e(t) \end{aligned} \quad (26)$$

It will be easy to verify that the PDCS model (26) provides a perfect representation for the original continuous model (22), and this model thus can be used to analyse the effect of the design parameter c on the dynamical behaviour of the original system.

Table 1. Identification result for the mass-spring-damper system given by (22), using the EFOR-CMSS algorithm.

Search step	Model term	AERR (%)	AAMD L
1	y(t-1)	99.000423	-6.343372
2	y(t-2)	0.877384	-8.050587
3	u(t-1)	0.068845	-8.613702
4	u(t-2)	0.053349	-52.251190
5	y ² (t-1)	8.273355e-0030	-52.230380
6	y(t-1)y(t-2)	5.605229e-0028	-51.839465
7	y ² (t-2)	9.031232e-0026	-49.679149
8	const	3.103675e-0026	-49.466224
Run time: 0.609sec			

Table 2. Parameter estimates for the selected model terms for the mass-spring-damper system given by (22).

Model term	Parameter estimates			
	c=2	c=10	c=20	c=40
y(t-1)	1.970306	1.895329	1.809675	1.662142
y(t-2)	-0.980199	-0.904838	-0.818731	-0.670375
u(t-1)	4.962667e-005	4.833333e-005	4.679167e-005	4.395833e-005
u(t-2)	4.922973e-005	4.675501e-005	4.376735e-005	3.837817e-005

4.2 Modelling a particle damper system

A particle damper is a device with one or more cavities filled with dry granular particles of diverse shapes and small sizes. The particles can move freely and the frictions and collisions between moving particles or with a container wall will arise under the vibrating motion of the structure. These collisions exchange momentum and thus dissipate kinetic energy due to frictional and in-elastic losses. Particle dampers have the advantage of being simple in geometry, small in volume, and are applicable in extreme temperature environments. More importantly, the interactions between individual grains (and between grains and the container walls) are dissipative because of surface friction and the inelasticity of collisions. An overwhelming advantage of particle dampers is that they can operate in extreme temperature conditions when using metallic, tungsten carbide or ceramic particles. This makes particle dampers extremely applicable in areas such as gas turbines, underwater conditions and other high temperature environments. Comprehensive discussions on particle dampers can be found in the literature say in Liu et al. (2005), and Rongong and Tomlinson (2005).

Several parameters may affect the performance of a particle damper and one crucial parameter is the cavity geometry. This example concerns such a geometry design parameter: the height-to-diameter ratio: $R=H/D$, where H and D are the height and diameter of the particle damper respectively. Five experiments, corresponding to $R=2,4,6,8,10$, have been completed on a particle damper device in the Department of Mechanical Engineering, University of Sheffield, and five different data sets, have been collected. Each data set consists of 2000 data pairs of the input (applied force) and the output (acceleration) observations, sampled with a frequency $f_s=12.8\text{kHz}$. The objective is to identify a PDCS model, with a dependence on the design parameter R , which can be used to analyze the effect of the design parameter R on the performance of the particle damper. Four data sets, corresponding to $R=2,4,6,10$, which are shown in Figure 1, were used for model identification, and one data set, correspond to $R=8$, was used to test the performance of the identified PDCS model.

Denote the system input and the output sequence using $\{u(t)\}_{t=1}^N$ and $\{y(t)\}_{t=1}^N$, respectively, with $N=2000$. The predictor vector for all the common-structured models was chosen to be $\mathbf{x}(t)=[x_1(t), \dots, x_{10}(t)]^T$, where $x_k(t) = y(t-k)$ for $k=1, \dots, 5$, and $x_k(t) = u(t-k+5)$ for $k=6, \dots, 10$. The initial candidate common model structure for all the four data sets was chosen to be a NARX model below

$$y(t) = \theta_0^0 + \sum_{i=1}^{10} \theta_i^0 x_i(t) + \sum_{i=1}^{10} \sum_{j=i}^{10} \theta_{i,j}^0 x_i(t) x_j(t) + e(t) \quad (27)$$

This candidate model involves a total of 66 candidate model terms. Based on the candidate common model structure, the new EFOR-CMSS algorithm was applied to the four training data sets. The AAMD index, shown in Figure 2, suggests that a common model structure, with 11 model terms, is preferred. The 11 selected common model terms, ranked in order of significance (the order that the

terms entered into the model), are shown in Table 3, where results for AERR and AAMD L are also presented. From Table 3, the resultant common model structure is of a simple NARX representation, which only includes linear model terms and a DC term with a small value.

The PDCS model for the particle damper system was chosen to be

$$y(t) = \theta_0(R) + \sum_{m=1}^{10} \theta_m(R)x_m(t) + e(t) \quad (28)$$

where the parameter θ_m ($m=0,1,\dots,10$) depends on the design parameter R . Assume that the parameter θ_m can be fitted using R , with a polynomial function below

$$\theta_m(R) = \beta_{m,0} + \beta_{m,1}R + \beta_{m,2}R^2 + \beta_{m,3}R^3, m=0,1, \dots, 10, \quad (29)$$

The parameters $\beta_{m,n}$ can directly be estimated using the results given in Table 3. The estimated values for $\beta_{m,n}$, for $m=0,1, \dots,10$ and $n=0,1,2,3$, are presented in Table 4.

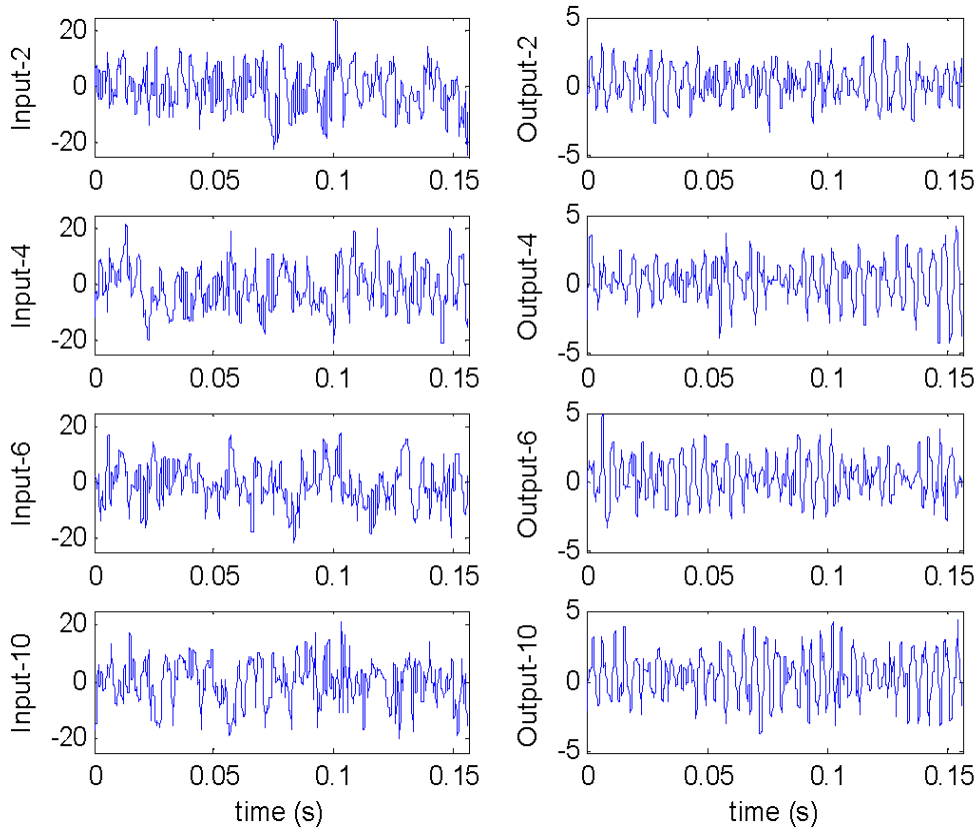


Fig. 1. Input-output data used for the particle damper system identification. Input-2,4,6,10 (Output-2,4,6,10) correspond respectively to the cases $R=2,3,4,6,10$.

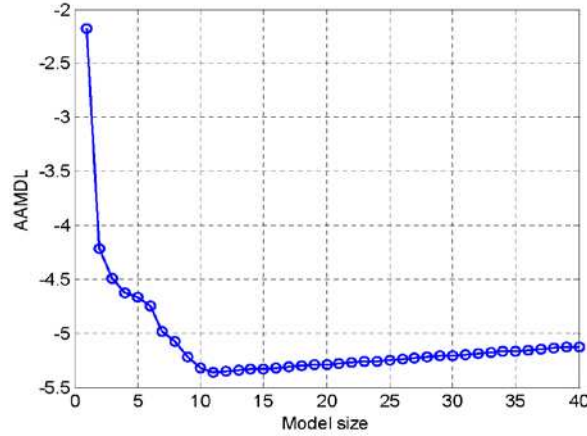


Fig. 2. AAMD L versus the model size of common model structure models, for the four data sets, corresponding to $R=2,4,6,10$, used for the particle damper system identification.

Table 3. Identification result for the particle damper system described in Example 2, using the EFOR-CMSS algorithm.

Search step	Model term	Parameter for different data sets				AERR (%)	AAMD L
		$R=2$	$R=4$	$R=6$	$R=10$		
1	$y(t-1)$	2.1590	1.7173	1.5291	1.2342	97.7609	-2.1776
2	$y(t-2)$	-1.7710	-0.8474	-0.4701	0.1447	2.1065	-4.2127
3	$y(t-5)$	-0.2052	2386	0.3059	0.4939	0.0418	-4.4944
4	$y(t-3)$	0.8049	0.7025	0.5786	0.4173	0.0173	-4.6247
5	$u(t-1)$	-0.3439	-0.5601	-0.6835	-0.6963	0.0046	-4.6630
6	$u(t-5)$	-0.1668	-0.3119	-0.3875	-0.3272	0.0086	-4.7474
7	$u(t-2)$	1.0432	1.8016	2.2228	2.1882	0.0170	-4.9806
8	$u(t-4)$	0.6890	1.3032	1.6290	1.4488	0.0060	-5.2148
9	$y(t-4)$	-0.0065	-0.8290	-0.9637	-1.3123	0.0064	-5.0786
10	$u(t-3)$	-1.2214	-2.2329	-2.7811	-2.6139	0.0041	-5.3216
11	const	0.0051	0.0047	0.0077	0.0083	0.0013	-5.3571
Run time: 2.37sec							

Now consider the performance of the identified PDCS model (28), whose parameters are determined by (29) and Table 4. The data set, corresponding to $R=8$, which has never been used in the identification procedure, was used to test the performance of the identified PDCS model. The PDCS model was simulated using the same input as in the data set corresponding to $R=8$, and the output from the PDCS model was then compared with the corresponding measurements. Figure 3 presents a comparison between the *model predicted output* and the original measurements. Note that the *model predicted output* (MPO) is defined as $\hat{y}(t) = \hat{f}(\hat{y}(t-1), \dots, \hat{y}(t-5), u(t-1), \dots, u(t-5))$, implying that $\hat{y}(t)$ is produced from the identified model iteratively. The mean-square-error was calculated to $MSE=0.1158$. Clearly, the PDCS model provides an almost perfect representation for the test data set.

Table 4. Estimates for the parameters $\beta_{m,n}$ in (29).

Model term	$\beta_{m,n}$				
	m	n			
		0	1	2	3
const	0	0.012800	-0.006325	0.001400	-0.000081
y(t-1)	1	3.023950	-0.566579	0.114194	0.074125
y(t-2)	2	-3.615675	0.775356	-0.161981	0.007808
y(t-3)	3	0.848050	-0.000471	-0.012125	0.000786
y(t-4)	4	2.039450	-1.418113	0.219888	-0.011159
y(t-5)	5	-1.321225	0.485829	-0.120994	0.006161
u(t-1)	6	-0.023800	-0.187875	0.014375	-0.000231
u(t-2)	7	-0.086050	0.662946	-0.050563	0.000701
u(t-3)	8	0.284975	-0.882169	0.065806	-0.000658
u(t-4)	9	-0.221950	0.531054	-0.038138	0.000174
u(t-5)	10	0.047050	-0.123988	0.008500	0.000016

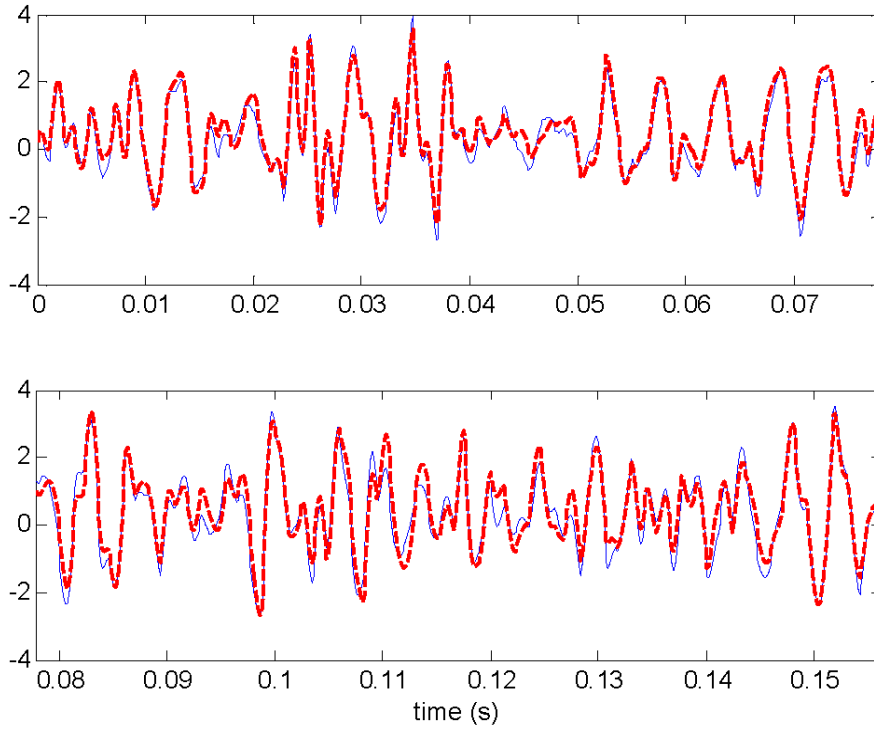


Fig. 3. A comparison between the model predicted output and the corresponding measurements for the particle damper system. The thin solid line indicates the original measurements for the case $R=8$, and the thick dashed line indicates the model predicted output from the identified PDCS model.

4.3 Modelling of thermoplastic auxetic foams

Dynamic tests on a class of auxetic elastomeric foams have been carried out at the Department of Mechanical Engineering, University of Sheffield, and it has been shown from experimental results that the associated foam specimens present nonlinear behaviour that may be applicable to design nonlinear dynamic filters. Several parameters may affect the nonlinear dynamic behaviour of the material and the imposed compression ratio is one crucial factor. This example concerns two design parameters related to the imposed compression ratio: the Axial (A) and the Volume (V) of the associated materials. The objective is to identify a PDCS model, whose parameters depend on the design parameters A and V , and which can be used to analyze the dynamic behaviour of the associated material when the design parameter A and V change.

Ten cases, corresponding to the following values for the design parameter A and V , were considered in this example:

A	2.13	1.88	1.63	1.38	1.13	2.13	1.87	1.60	1.33	1.20
V	5.30	4.67	4.05	3.43	2.80	14.8	12.9	11.1	9.20	8.30

Ten different data sets, symbolized by Data01, Data02, ..., Data10, corresponding to the above 10 cases, have been collected, and each data set consists of 2000 data pairs of observations for the input (displacement: mm) and the output (force: N), sampled with a frequency $f_s = 100\text{Hz}$. Note that all the 10 data sets are with the same input signal, as shown in Figure 4, but with different output signals, as shown in Figure 5, where only part of the observations are plotted for clear visualization. Eight data sets, numbered by 1,2,4,5,6,7,9, and 10, were used for model identification, and the remaining two data sets, numbered by 3 and 8, were used for the performance test of the identified PDCS model.

Denote the system input and the output sequence using $\{u(t)\}_{t=1}^N$ and $\{y(t)\}_{t=1}^N$, respectively, with $N=2000$. The predictor vector for the common model structure was chosen to be $\mathbf{x}(t) = [x_1(t), \dots, x_4(t)]^T$, with $x_k(t) = u(t-k+1)$ for $k=1,2,3,4$. The initial candidate common model structure was chosen to be

$$y(t) = \theta_0^0 + \sum_{i=1}^4 \theta_i^0 x_i(t) + \sum_{i=1}^4 \sum_{j=i}^4 \theta_{i,j}^0 x_i(t) x_j(t) + e(t) \quad (30)$$

This candidate model involves a total of 15 candidate model terms. Based on the candidate common model structure, the new EFOR-CMSS algorithm was applied to the 8 training data sets. The AAMD index, shown in Figure 6, suggests that a common model structure, with 8 model terms, is preferred. The 8 selected common model terms, ranked in order of the significance, are shown in Table 5. The PDCS model for the 8 training data sets was chosen to be

$$y(t) = \theta_1(A, V)u^2(t) + \theta_2(A, V)u(t-1) + \theta_3(A, V)u(t-1)u(t-3) + \theta_4(A, V)u^2(t-3)$$

$$+ \theta_5(A, V)u(t-2) + \theta_6(A, V)u(t) + \theta_7(A, V) + \theta_8(A, V)u(t)u(t-1) + e(t) \quad (31)$$

where the parameter θ_m ($m=1, \dots, 8$) were fitted using the following polynomial function

$$\theta_m(A, V) = \beta_{m,0} + \beta_{m,1}A + \beta_{m,2}V + \beta_{m,3}A^2 + \beta_{m,4}AV + \beta_{m,5}V^2, \quad m=1, \dots, 8, \quad (32)$$

The parameters $\beta_{m,n}$ were directly estimated using the results given in Table 5 and the associated estimates for $\beta_{m,n}$ are shown in Table 6.

To inspect the performance of the identified PDCS model (31), the model was simulated by choosing the same input signal as that in the two test data sets numbered by 3 and 8. The output from the PDCS model was then compared with the relevant measurements. Figures 7 and 8 present comparisons between the model outputs and the associated measurements. Note that only part of the data points are shown in Figures 7 and 8 for a close inspection. The root-mean-square-error (RMSE), defined as the root of the mean-square-error, with respect to two training data sets, was calculated to RMSE=1.71 and 6.44, respectively. Clearly, the PDCS model provides an excellent representation for the two test data sets.

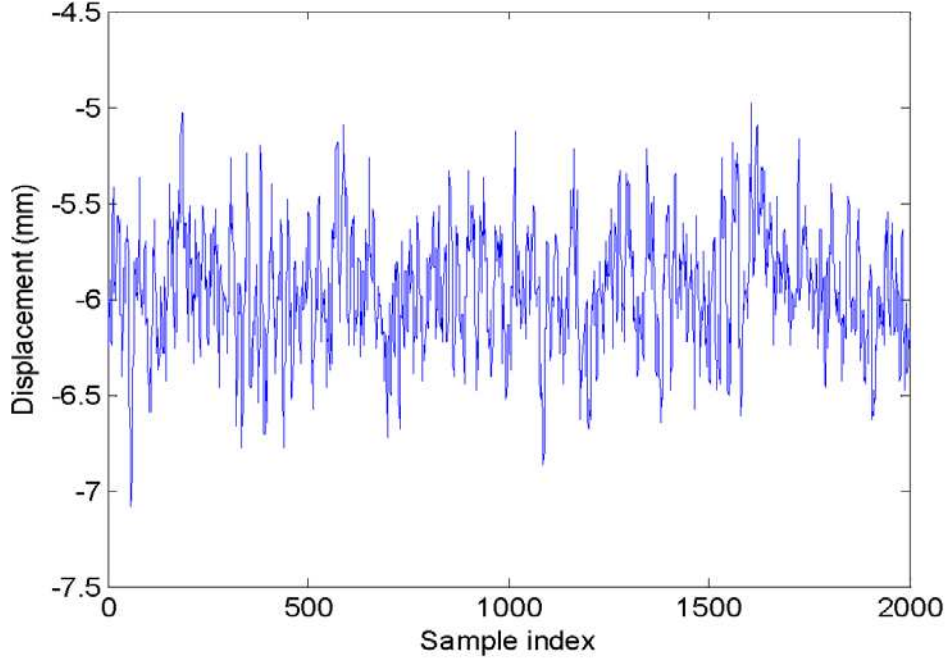


Fig. 4. The input signal used for the modelling of the auxetic elastomeric foams

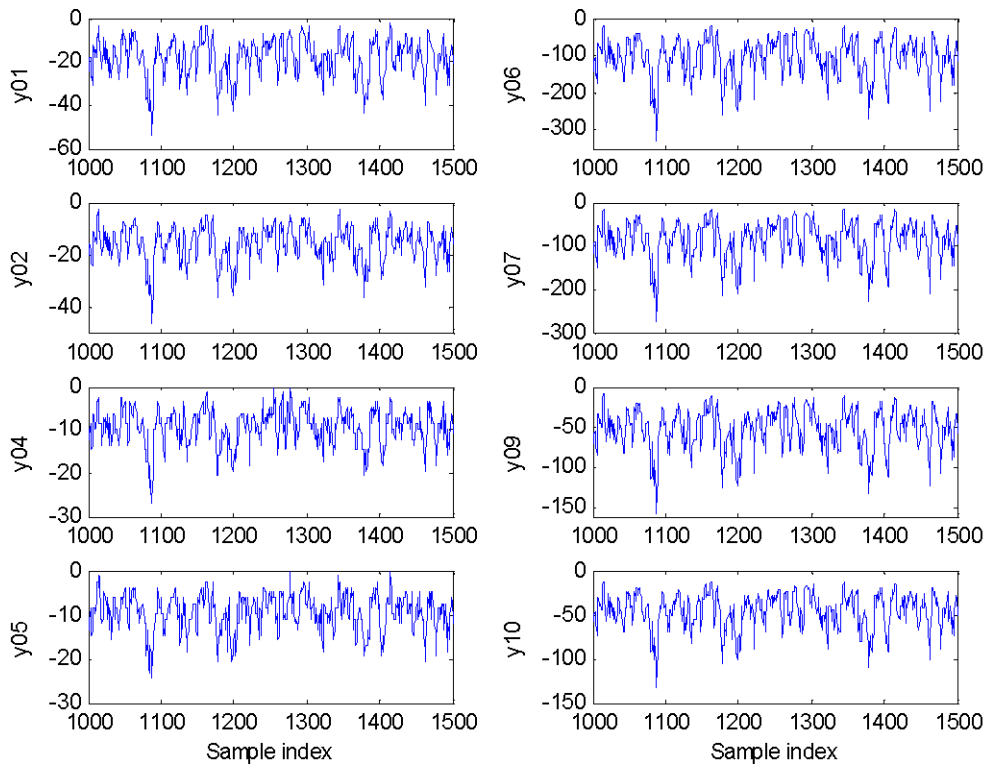


Fig. 5. The output signals in the data sets numbered by 1, 2, 4, 5, 6, 7, 9, 10, for the associated auxetic elastomeric foams.

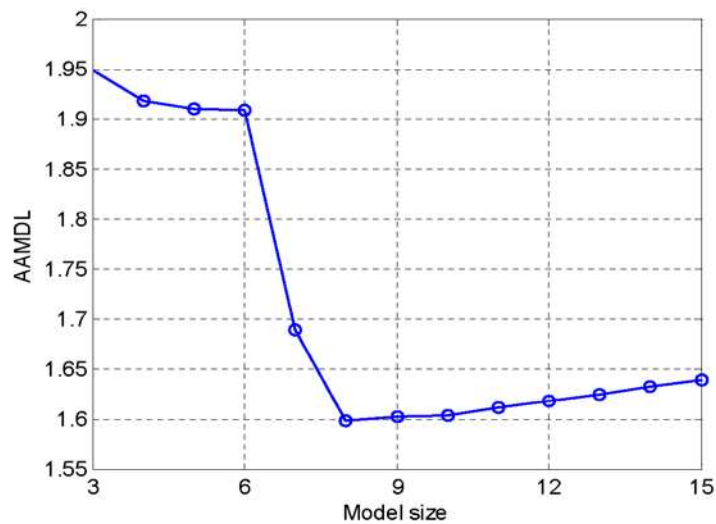


Fig. 6. AAMD L versus model size for common model structure models of the associated auxetic elastomeric foams.

Table 5. Identification result for the associated auxetic elastomeric foams described in Example 3, using the EFOR-CMSS algorithm.

Step	Model term	Parameter for different data sets								AERR (%)
		Data01	Data02	Data04	Data05	Data06	Data07	Data09	Data10	
1	$u^2(t)$	-24.78	-19.10	-10.71	-10.33	-173.65	-148.65	-81.78	-66.62	88.397
2	$u(t-1)$	71.22	51.28	33.09	25.26	477.51	426.92	229.47	196.34	10.042
3	$u(t-1)u(t-3)$	-0.35	0.43	-0.73	0.67	3.53	1.43	1.49	0.66	0.117
4	$u^2(t-3)$	0.52	-0.06	0.35	-0.31	1.80	2.18	0.66	0.70	0.057
5	$u(t-2)$	1.43	-0.50	-2.78	-0.19	37.94	30.22	13.51	8.64	0.025
6	$u(t)$	-168.66	-129.39	-77.67	-67.08	-1187.99	-1016.74	-559.00	-454.19	0.015
7	const	-234.68	-194.87	-118.63	-101.74	-1701.85	-1415.05	-801.24	-632.16	0.025
8	$u(t)u(t-1)$	14.63	10.60	6.23	5.53	100.55	88.73	47.78	40.04	0.083

Run time: 2.53sec

Table 6. Estimates for the parameters $\beta_{m,n}$ in (32).

m	n					
	0	1	2	3	4	5
1	-14.05	20.09	-30.01	292.02	-161.80	16.52
2	47.98	-72.58	10.23	-1172.89	649.59	-67.22
3	1.81	-2.29	0.05	29.88	-16.05	1.72
4	-0.28	0.18	-0.03	-20.62	11.29	-1.19
5	16.35	-22.45	-0.70	-80.71	47.70	-4.83
6	-51.41	78.40	-20.05	1955.65	-1074.11	109.58
7	75.10	-92.45	-27.90	1688.76	-896.05	88.65
8	12.31	-17.74	1.90	-230.03	127.98	-13.21

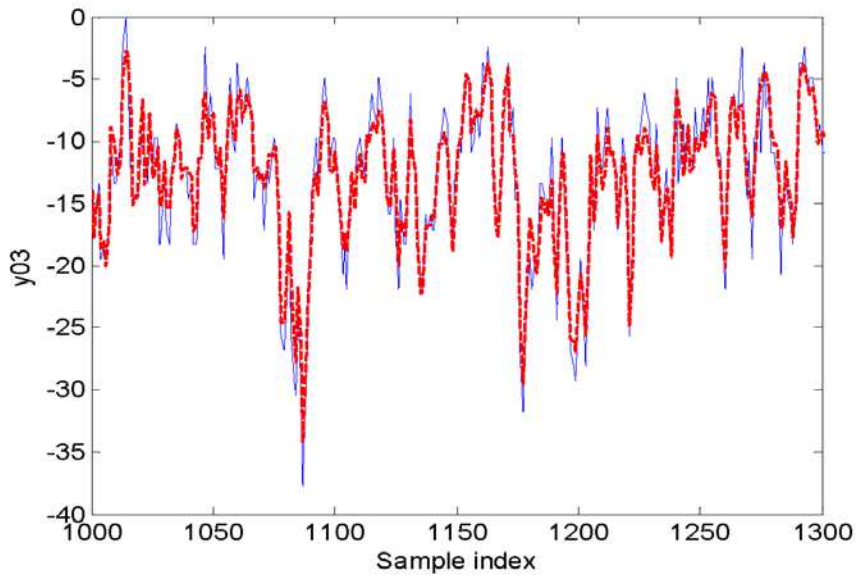


Fig. 7. A comparison between the model predicted output from the identified PDCS model (31) and the corresponding measurements in Data03, for the associated auxetic elastomeric foams. The thin solid line indicates the measurements, and the thick dashed line indicates the model predicted output.

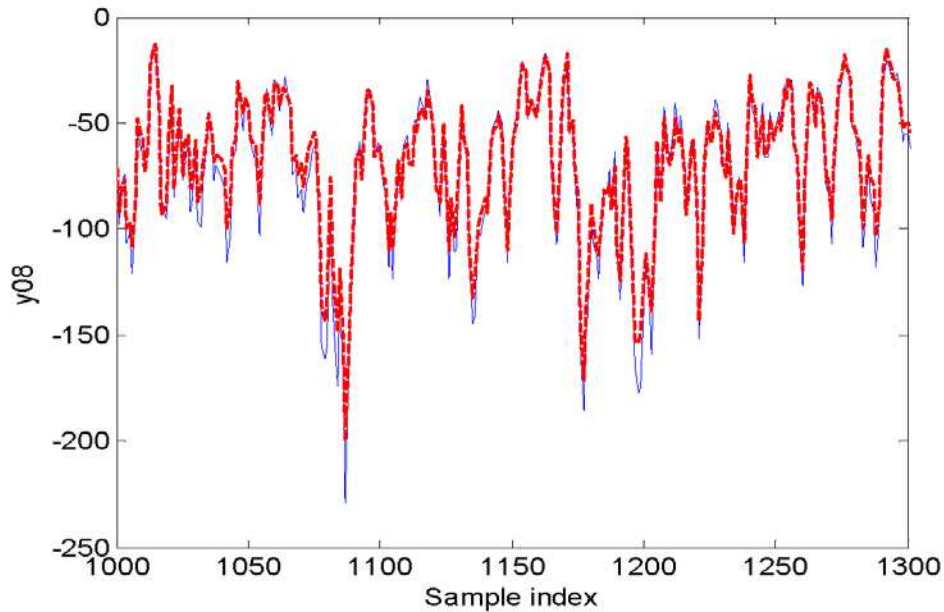


Fig. 8. A comparison between the model predicted output from the identified PDCS model (31) and the corresponding measurements in Data08, for the associated auxetic elastomeric foams. The thin solid line indicates the measurements, and the thick dashed line indicates the model predicted output.

5. Conclusions

Many exogenous parameters may affect the underlying dynamics of a system, where the internal model structure of the system is fixed but the process internal parameters change due to the effects of the variation of the external parameters for example design parameters or experimental conditions. Parameter-dependent common-structured (PDCS) models are thus desirable for system analysis and design. In many cases, the true model structure of the relevant system is unknown but only the input and the output observations, subject to given specific design parameters, are available, and a common model structure is often deduced from the available observations. Common model structure selection (CMSS) is a crucial stage to obtain an effective PDCS model. A new efficient extended forward orthogonal regression (EFOR) algorithm has been designed to solve the CMSS problem. The identification of PDCS models includes two steps. Firstly, the common model structure is selected using the new EFOR algorithm, and individual parameters corresponding to each of the experiments are calculated using this algorithm. Secondly, the individual parameters are linked to the design parameters by fitting some functions where the independent variables are the design parameters. Results from case studies have strongly supported the applicability and effectiveness of the new EFOR algorithm for the CMSS problem.

Acknowledgements

The authors gratefully acknowledge that this work was supported by EPSRC(UK). The authors also gratefully acknowledge the help from J. Rongong and F. Scarpa in the data collection for the particle damper and auxetic foams respectively.

Appendix—Some abbreviations

AAMDL: average approximate minimum description length.

AERR: average error ratio reduction.

AMDL: approximate minimum description length.

CMSS: common model structure selection.

EFOR: extended forward orthogonal regression.

EPD: external-parameter-dependent.

ERR: error ratio reduction.

FOR: forward orthogonal regression.

IPD: internal-parameter-dependent.

MDL: minimum description length.

OLS: orthogonal least squares.

PDCS: parameter-dependent common-structured (model).

References

- A. Antoniadis, I. Gijbels, G. Gregoire, “Model selection using wavelet decomposition and applications,” *Biometrika*, 84(4), pp. 751-763, 1997.
- C. S. Berger, “Linear Splines with Adaptive Mesh Sizes for Modeling Nonlinear Dynamic-Systems,” *IEE Proc.-Control Theory Appl.*, 141(5), pp. 277-284, 1994.
- S. A. Billings and W. S. F. Voon, “Piecewise linear identification of non-linear systems,” *Int. J. Control*, 46(1), pp.215-235, 1987.
- S. A. Billings, S. Chen, and M. J. Korenberg, “Identification of MIMO non-linear systems using a forward regression orthogonal estimator,” *Int. J. Control*, 49(6), pp.2157-2189, 1989.
- S. A. Billings and H. L. Wei, “The wavelet-NARMAX representation: A hybrid model structure combining polynomial models with multiresolution wavelet decompositions,” *Int. J. Syst. Sci.*, 36(3), pp. 137-152, 2005.
- S. Chen and S. A. Billings, “Representations of non-linear systems - the NARMAX model,” *Int. J. Control*, 49(3), pp. 1013-1032, 1989.
- S. Chen, S. A. Billings, and W. Luo, “Orthogonal least squares methods and their application to non-linear system identification,” *Int. J. Control*, 50(5), pp.1873-1896, 1989.

- S. Chen, S. A. Billings, C. F. N. Cowan, and P. M. Grant, "Practical identification of NARMAX models using radial basis functions," *Int. J. Control*, 52(6), pp. 1327-1350, 1990.
- S. Chen, X. Hong, C. J. Harris, and X. X. Wang, "Identification of nonlinear systems using generalized kernel models," *IEEE Trans. Control Syst. Technol.*, 13(3), pp. 401-411, 2005.
- C. J. Harris, X. Hong, and Q. Gan, *Adaptive Modelling, Estimation and Fusion from Data: A Neurofuzzy Approach*. Berlin: Springer-Verlag, 2002.
- J. A. Rongong and G. R. Tomlinson, "Amplitude dependent behaviour in the application of particle dampers to vibrating structures," In the *Proceedings of the 46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics & Materials Conference*, Art No.: AIAA 2005-2327, 18-21 April 2005, Austin, Texas, USA.
- T. Kavli, "Asmod - an algorithm for adaptive spline modeling of observation data," *Int. J. Control*, 58(4), pp. 947-967, 1993.
- I. J. Leontaritis and S. A. Billings, "Input-output parametric models for non-linear systems, part I: deterministic non-linear systems," *Int. J. Control*, 41, pp. 303-344, 1985.
- G. P. Liu, *Nonlinear Identification and Control: A Neural Network Approach*. Berlin: Springer-Verlag, 2001.
- W. Liu, G. R. Tomlinson, and J. A. Rongong, "The dynamic characterisation of disk geometry particle dampers," *J Sound Vibr.*, 280, pp. 849-861, 2005.
- R. K. Pearson, *Discrete-Time Dynamic Models*. Oxford: Oxford University Press, 1999.
- J. Rissanen, "A universal prior for integers and estimation by minimum description length," *Ann. Stat.*, 11, pp. 416-431, 1983.
- N. Saito, "Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion," In *Wavelet in Geophysics*, Foufoula-Georgiou, E. and Kumar, P., Eds, New York: Academic, pp. 299-324, 1994.
- H. L. Wei and S. A. Billings, "Feature subset selection and ranking for data dimensionality reduction," Accepted by *IEEE Trans. Pattern Anal. Machine Intell.*, 2006.
- H. L. Wei, S. A. Billings, and M. A. Balikhin, "Wavelet based nonparametric NARX models for nonlinear input-output system identification," Accepted by *Int. J. Syst. Sci.*, 2006.
- Z. Q. Wu and C. J. Harris, "A neurofuzzy network structure for modelling and state estimation of unknown nonlinear systems," *Int. J. Syst. Sci.*, 28(4), pp. 335-345, 1997.