Gradient-based adaptation of continuous dynamic model structures

William G. La Cava & Kourosh Danai

Taylor & Francis
Taylor & Francis Group

Gradient-based adaptation of continuous dynamic model structures

William G. La Cava and Kourosh Danai*

Department of Mechanical and Industrial Engineering, University of Massachusetts, Amherst, MA, USA

A gradient-based method of symbolic adaptation is introduced for a class of continuous dynamic models. The proposed model structure adaptation method starts with the first-principles model of the system and adapts its structure after adjusting its individual components in symbolic form. A key contribution of this work is its introduction of the model's parameter sensitivity as the measure of symbolic changes to the model. This measure, which is essential to defining the structural sensitivity of the model, not only accommodates algebraic evaluation of candidate models in lieu of more computationally expensive simulation-based evaluation, but also makes possible the implementation of gradient-based optimisation in symbolic adaptation. The proposed method is applied to models of several virtual and real-world systems that demonstrate its potential utility.

Keywords: system identification; gradient-based adaptation; dynamic modelling; symbolic regression; evolutionary programming

1. Introduction

A major goal of science is to characterise analytically the dynamic behaviour of natural phenomena associated with biological, ecological, social, and economic systems, as well as the dynamics of artefacts such as robots, aircraft, and wind turbines. Dynamic behaviours are usually characterised by differential equations which in aggregate represent the dynamic model of the system. Dynamic models, in turn, are the essence of the virtual environments that are used to estimate/predict system behaviour for policy decisions, design, optimisation, control, and/or automation.

Dynamic models are preferably formulated according to first principles, to embody the knowledge of the system. However, first-principles models are usually too crude and incomplete to fully characterise the non-linear dynamics of the system, as represented by process observations. In regress, first-principles models are often abandoned in favour of empirical models such as neural networks (Hertz, Krogh, & Palmer, 1991; Narendra & Parthasarathy, 1990, 1991), fuzzy logic (Zhu, Danai, & McCormick, 1994), or non-linear autoregressive moving average (NARMAX) models (Billings, 2013), among others, which offer the structural flexibility for adaptation to accommodate the measured observations. Although these empirical models provide an effective means of estimation/prediction, they have the major drawback of lacking transparency about the physics of the process (Billings, 2013). This lack of transparency, in turn, obscures the knowledge of the process acquired through adaptation. Ideally, the model of the process

should benefit from the available knowledge of the process, to minimise adaptation, and be intelligible so as to inform the acquired knowledge attained through adaptation.

Attempts have been made to incorporate the available knowledge of the process in fuzzy (Mitra, Konwar, & Pal, 2002) or neural network (Towell & Shavlik, 1994) models. Two examples are the radial basis networks of Gan and Danai (2000, 2001) that are structured according to the system's linear first-principles model and subsequently adapted to improve the accuracy of the model's output relative to observations. Another example is the structure-based neural network of Jammu, Danai, and Lewicki (1998a, 1998b) for fault diagnosis of gearboxes wherein the weights are specified according to the relationship between measured inputs and faults as well as the proximity of sensors to the gearbox components. But these solutions, although they rely on the available knowledge of the process to formulate their initial model, fail to convey the improved knowledge attained through adaptation. Often the distributed form of these models precludes intelligibility of their embedded knowledge.

In order to improve the intelligibility of adapted models, empirical models in the form of symbolic equations can be formulated by symbolic regression (Bongard & Lipson, 2007; Koza, 1992; La Cava, Spector, Danai, & Lackner, 2014; Schmidt & Lipson, 2009). In symbolic regression, the process variables, inputs, and parameters (constants) are treated as symbols and integrated as blocks to form candidate model structures. Free of restrictions on the form

---

*Corresponding author. Email: danai@umass.edu

(structure) of candidate model, the search is conducted by genetic programming for models having best-fit outputs to the measured observations (Koza, 1992). Even though symbolic regression is computationally expensive, requiring anywhere from thousands to billions of evaluations, it offers a viable approach to modelling poorly understood systems that cannot be readily defined by first-principles models. By the same token, the unrestricted structure of symbolic regression renders it unsuitable for application to better understood systems because of the inherent difficulty of seeding it with starting models (Schmidt & Lipson, 2009). In the absence of a presumed model structure, symbolic regression often yields illegible, albeit accurate, models which do not convey any of the physics of the process.

The model structure adaptation method (MSAM) proposed in this research contrasts the unrestricted nature of symbolic regression by considering candidate models closely tied to the starting model that are improved by localised gradient-based adaptation. As such, MSAM is designed to remedy the shortcomings of symbolic regression in application to well understood systems for which first-principles models are available. It achieves this by adjusting the individual components of the original model so as to preserve the model's structural integrity, hence, its intelligibility. A key contribution of MSAM, that enables the implementation of gradient-based adaptation, is its use of the model's parameter sensitivity as the measure of 'model difference magnitude'. This measure is used to scale the structural sensitivities such that they will remain robust to parametric error during adaptation.

2. Problem formulation

The underlying assumption of symbolic regression is that there exists an analytical model of the system that would generate the measured observations $y(t_k)$ at the sample times $t_k = t_1, \ldots, t_N$ under the input, $\mathbf{u}(t)$, as

$$y(t_k, \mathbf{u}) = \hat{y}(t_k, \mathbf{M}^*, \Theta^*, \mathbf{u}) + \nu; \qquad k = 1, \ldots, N \quad (1)$$

where $\hat{y}$ is the modelled output, ν represents measurement noise in y, $\mathbf{M}^*$ denotes the correct model form, and $\Theta^* = [\theta_1^*, \ldots, \theta_Q^*]^T$ is the vector of model parameters. We consider the model to consist of the weighted sum of individual components M_i, as

$$\mathbf{M}_\Theta = \sum_{i=1}^{Q} \theta_i M_i = \Theta^T \mathbf{M} \quad (2)$$

where $\mathbf{M} = [M_1, \ldots, M_Q]^T$ and each component M_i is the product of any combination of state variables, x_i, included in the state vector $\mathbf{x} = [x_1, \ldots, x_n]^T$, and/or inputs, u_i, in the input vector $\mathbf{u} = [u_1, \ldots, u_m]^T$. For instance, consider

the true model of the harmonic oscillator

$$\mathbf{M}_\Theta^* : \ddot{x} = -\frac{c}{m}\dot{x}|\dot{x}| - \frac{k}{m}x^3 + \frac{1}{m}u(t) \quad (3)$$

where x denotes its displacement (that is measured), $\dot{x}$ is its velocity, $\ddot{x}$ is its acceleration, $u(t)$ is its input excitation, and m, c, and k denote its mass, damping coefficient, and spring constant, respectively. This true model consists of three components; i.e., $\mathbf{M}^* = [M_1^*, M_2^*, M_3^*]^T = [\dot{x}|\dot{x}|, x^3, u(t)]^T$ where the true parameter values $\Theta^* = [\theta_1^*, \theta_2^*, \theta_3^*]^T = [-\frac{c^*}{m^*}, -\frac{k^*}{m^*}, \frac{1}{m^*}]^T$. Given that the measured outputs have the quality and breadth to characterise the dynamics of the process, the fidelity of the model can be evaluated by how closely the model outputs match the observations (Dunstan & Bitmead, 2003; Popper, 1959, 1994).

The most common measure of closeness of the model is the magnitude of the prediction error between the process observations, y, and model output, $\hat{y}$, defined as

$$\begin{aligned}\epsilon(t_k) = y(t_k) - \hat{y}(t_k) &= \hat{y}(t_k, \mathbf{M}^*, \Theta^*, \mathbf{u}) \\ &\quad - \hat{y}(t_k, \widehat{\mathbf{M}}, \widetilde{\Theta}, \mathbf{u}) + \nu \end{aligned} \quad (4)$$

where $\widehat{M}$ denotes the candidate model form and $\widetilde{\Theta}$ the vector of nominal parameter values.

In traditional system identification, model formulation and parameter estimation are performed separately. Once the model form is assumed correct; i.e., $\widehat{\mathbf{M}} = \mathbf{M}^*$, the model parameters, ascertained identifiable (Ljung & Glad, 1994), are estimated by minimising a cost function, V, as

$$\widehat{\Theta} = \arg\min_{\Theta} V = \arg\min_{\Theta} \sum_{k=1}^{N} L\left(\epsilon(t_k)\right) \quad (5)$$

where L is a scalar-valued (typically positive) function, such as the square function in non-linear least-squares (NLS). However, when the model form is incorrect (i.e., $\widehat{\mathbf{M}} \neq \mathbf{M}^*$), parameter estimation either fails or leads to erroneous values associated with an inordinate prediction error, indicating the model mismatch and the need for a better model structure. Therefore, structural accuracy of the model transcends its parametric accuracy, hence the focus of adaptation in MSAM.

The common choice for estimating the model output(s) is numerical integration (i.e., simulation) of state variables. When using different model structures in simulation, the candidate models' output is a function of state variables and inputs, so it varies considerably by even minute changes in the model structure. For instance, because in simulation the rate of output change will depend on the previous outputs, the time span of the transients may change, making it difficult to compare simulated outputs of different models. Therefore, simulation-based estimation of model outputs

with different model structures, aside from its computational demand and error propagation tendency, is undesirable for its ambiguity about the model quality.

The alternative to numerical integration is algebraic estimation of candidate model outputs, as commonly performed in symbolic regression (Bongard & Lipson, 2007; Schmidt & Lipson, 2009). In algebraic evaluation of models, states are estimated from the measured output (by differentiation and/or integration together with various smoothing functions) to yield $\tilde{\mathbf{x}} = [\tilde{x}_1, \ldots, \tilde{x}_n]^T$. The estimated output of the candidate model can then be the state represented by the model; e.g., $\hat{x}$ in the harmonic oscillator of Equation (3). In this configuration, the estimated output finds the form

$$\hat{y}(t_k) = \widehat{\mathbf{M}}_\Theta = \Theta^T \mathbf{M}(\tilde{\mathbf{x}}(t_k), \tilde{\mathbf{u}}(t_k)) \qquad (6)$$

and the prediction error will have the form

$$\epsilon(t_k) = y(t_k) - \hat{y}(t_k) = \widetilde{\Theta}^T \mathbf{M}^* - \widetilde{\Theta}^T \widehat{\mathbf{M}} \qquad (7)$$

where $\tilde{\mathbf{u}}$ are the inputs used to produce the measured observations $y(t_k)$. In algebraic evaluation, therefore, model validation is a static test in which the state variables are independent of the model structure and the dependent variable is the modelled variable defined algebraically by the candidate model being evaluated. Referring back to the harmonic oscillator of Equation (3), if the measured variable is the displacement $\tilde{x}(t_k)$, then it can be used to estimate the model velocity $\dot{\tilde{x}}(t_k)$ by numerical differentiation and, say, piece-wise cubic interpolation and LOESS (locally weighted regression) smoothing (Cleveland & Devlin, 1988) to cope with noise and differentiation errors. Now if the candidate model $\widehat{\mathbf{M}} = [\dot{x}, \ x, \ u(t)]^T$ is the linear form of the harmonic oscillator, and the nominal parameter values are considered to be $\widetilde{\Theta} = [-\frac{\tilde{c}}{\tilde{m}}, \ -\frac{\tilde{k}}{\tilde{m}}, \ \frac{1}{\tilde{m}}]^T$, then the model output is estimated as

$$\hat{y}(t_k) = \widehat{\dot{x}}(t_k) = -\frac{\tilde{c}}{\tilde{m}}\dot{\tilde{x}}(t_k) - \frac{\tilde{k}}{\tilde{m}}\tilde{x}(t_k) + \frac{1}{\tilde{m}}\tilde{u}(t_k)$$

and the prediction error is

$$\epsilon(t_k) = -\frac{c^*}{m^*}\dot{\tilde{x}}(t_k)|\dot{\tilde{x}}(t_k)| + \frac{\tilde{c}}{\tilde{m}}\dot{\tilde{x}}(t_k) - \frac{k^*}{m^*}\tilde{x}(t_k)^3 + \frac{\tilde{k}}{\tilde{m}}\tilde{x}(t_k) + \nu$$

Therefore, the prediction error is not only offset by noise, but also by the inaccuracy of the model form; i.e., $\widehat{\mathbf{M}} \neq \mathbf{M}^*$ and the parametric error, $\widetilde{\Delta\Theta} = \Theta^* - \widetilde{\Theta}$.

3. The model structure adaptation method

In MSAM, each component is adapted symbolically with the objective of improving the fitness of the model. To this end, a two-stage adaptation strategy is implemented. In the first stage, a comprehensive set of component adjustments is tested after iterative adaptation to select the 'best candidate model'. In the second stage, this best candidate model is adapted further to improve the fitness of the model. The salient features of MSAM are (1) its unintrusive adjustment scheme which keeps the original model structure intact and conducive to interpretation after adaptation, (2) its use of gradient-based search for improved efficiency over the stochastic search conducted in symbolic regression, (3) its capacity to measure the model change magnitude to accommodate gradient-based search in presence of parametric error, and (4) its ability to find the correct model structure despite parametric inaccuracy.

3.1. Adaptation strategy

In MSAM, the candidate models are formed by adjusting each model component as

$$\widehat{M}_i \implies \widetilde{M}_i \widehat{f}_i(\tilde{\mathbf{x}}, \tilde{\mathbf{u}})^{\gamma_i} \qquad (8)$$

to yield the candidate model

$$\widehat{\mathbf{M}}_{\widetilde{\Theta}} = \sum_{i=1}^{Q} \tilde{\theta}_i \widetilde{M}_i \widehat{f}_i(\tilde{\mathbf{x}}, \tilde{\mathbf{u}})^{\gamma_i} = \widetilde{\Theta}^T \widehat{\mathbf{M}} \qquad (9)$$

where $\widehat{\mathbf{M}} = [\widetilde{M}_1 \widehat{f}_1(\tilde{\mathbf{x}}, \tilde{\mathbf{u}})^{\gamma_1}, \ldots, \widetilde{M}_Q \widehat{f}_Q(\tilde{\mathbf{x}}, \tilde{\mathbf{u}})^{\gamma_Q}]^T$, the $\widehat{f}_i$ are functions of individual state variables or inputs considered to improve the model form, and the γ_i are exponents to achieve two goals: (1) to mitigate the discrete nature of the introduced model change, and (2) to provide a mechanism for calibrating the degree of change to individual model components for higher granularity. For instance, to achieve $\widetilde{M} = \dot{x} \implies \widetilde{M} f^*(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}) = \dot{x}|\dot{x}|$, the adjustment needs to be $\widehat{f}(\mathbf{x}) = |\dot{x}|^{1.0}$. Assuming that the true model can be reached by the introduction of candidate adjustments $\widehat{f}$ to the initial model $\widetilde{\mathbf{M}}$, the true model will have the form $\mathbf{M}^* = [\widetilde{M}_1 f_1^*(\tilde{\mathbf{x}}, \tilde{\mathbf{u}})^{\gamma_1^*}, \ldots, \widetilde{M}_Q f_Q^*(\tilde{\mathbf{x}}, \tilde{\mathbf{u}})^{\gamma_Q^*}]^T$. The adaptation strategy, hence, entails applying adjustments of the form (8) to individual components of the model $\widetilde{M}$ and then adapting the exponents γ_i to fine-tune the model structure. The goal of MSAM is to first find the form $\widehat{\mathbf{f}} = [\widehat{f}_1(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}), \ldots, \widehat{f}_Q(\tilde{\mathbf{x}}, \tilde{\mathbf{u}})]^T$, that will match the correct form $\mathbf{f}^* = [f_1^*(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}), \ldots, f_Q^*(\tilde{\mathbf{x}}, \tilde{\mathbf{u}})]^T$ and then adapt the exponents γ_i, to achieve $\Gamma = [\gamma_1, \ldots, \gamma_Q]^T \implies \Gamma^* = [\gamma_1^*, \ldots, \gamma_Q^*]^T$.

The proposed adaptation strategy, as outlined above, is therefore tailored to starting models with missing couplings. Even though such models may constitute only a subset of non-linear models to be envisioned for a process, they encompass considerable non-linear capacity. Indeed the method could be enhanced in reach by the added

provision of introducing new components to the original model. However, that is beyond the premise of the method which is aimed at refining starting models with adequate components to capture the phenomenological aspects of the process. After all, for a more flexible and expanded model structure one could resort to symbolic regression independent of any structural constraints of the starting model. Another requirement of the proposed method is the set of functions f to be introduced into the starting model. While the method does not pose any limitation to the number of functions to be considered, its computational effort will be increased with the larger number of candidate models produced by the expanded set of functions. Given n functions and Q components, the number of possible candidate models would be Q^n. Fortunately, the adaptation of individual candidate models can be performed independently of each other, making parallel execution possible.

Given that finding the correct structural change $\mathbf{f}^*(\tilde{\mathbf{x}}, \tilde{\mathbf{u}})$ transcends adaptation of the corresponding exponents, the first stage of adaptation in MSAM comprises a round robin competition between candidate models of the form (9), after a limited number of exponent adaptations, according to a fitness function of the prediction error (Equation (7)). For gradient-based adaptation in the round robin stage, the target output $y(t)$ can be defined by its first-order approximation at the nominal parameter values $\tilde{\theta}_i$, and exponents $\widehat{\gamma}_i$, as

$$y(t) \approx \hat{y}(t, \widehat{\mathbf{M}}, \widehat{\Gamma}, \widetilde{\Theta}) + \sum_{i=1}^{Q} \widetilde{\Delta\theta}_i \left(\frac{\partial \hat{y}(t, \widehat{\mathbf{M}}, \widehat{\Gamma}, \widetilde{\Theta})}{\partial \theta_i} \right)$$
$$+ \sum_{i=1}^{Q} \widehat{\Delta\gamma}_i \left(\frac{\partial \hat{y}(t, \widehat{\mathbf{M}}, \widehat{\Gamma}, \widetilde{\Theta})}{\partial \gamma_i} \right) \tag{10}$$

where $\widetilde{\Delta\theta}_i = \theta_i^* - \tilde{\theta}_i$ and $\widehat{\Delta\gamma}_i = \gamma_i^* - \hat{\gamma}_i$. The above approximation holds when the structure of the candidate model provides a close first-order approximation of the target output and the partial derivatives of $\hat{y}$ are reasonably close to the corresponding partial derivatives of y.

Ideally, one would want to adapt both the coefficients θ_i and exponents γ_i for each candidate model $\widehat{\mathbf{M}}$ during the round robin phase with the objective of identifying the correct model form. However, potential collinearity between θ_i/γ_i pairs often hinders their concurrent adaptation, forcing one to adapt the one with the larger influence on the prediction error. As is discussed in the next section, since the exponents (in the absence of bifurcation) have the larger influence on the prediction error, they are the preferred target for adaptation. Therefore, an underlying assumption of MSAM is that a candidate model with the correct function adjustments $\widehat{\mathbf{f}} = \mathbf{f}^*$ will have the best fitness relative to other candidate models after adaptation despite the parametric error $\widetilde{\Delta\Theta}$.
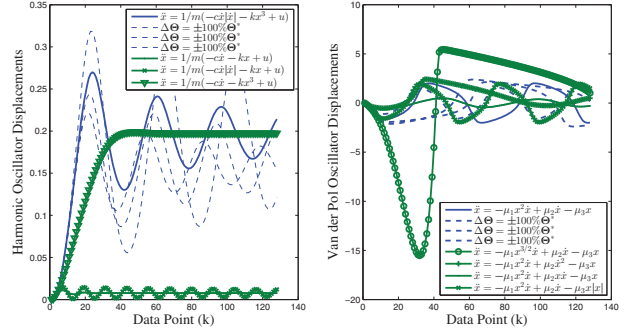

Figure 1. Displacements of the harmonic oscillator (left) and the van der Pol oscillator (right) with different parameter values and structures.

3.2. Precedence of structural error to parametric error

Albeit anecdotal, the prominence of exponents γ_i over the coefficients θ_i in MSAM is shown via two examples in Figure 1. The two models are those of the harmonic oscillator (left plots) and the van der Pol oscillator (right plots). The plots show the displacements of the two models with different structures and parametric error levels, computed as $\Delta\Theta = \sum_{i=1}^{Q} |\theta_i^* - \tilde{\theta}_i|/\theta_i^*$. Even though three parameters are defined for the van der Pol oscillator, to provide breadth for parameter variation, the parameters of the nominal model ($\ddot{x} = -\eta_1 x^2 \dot{x} + \eta_2 \dot{x} - \eta_3 x$) are defined as $\eta_1 = \eta_2$ and $\eta_3 = 1$ to match the standard van der Pol oscillator form ($\ddot{x} = \eta\dot{x}(1 - x^2) - x$). The results in Figure 1 indicate that both oscillators' displacements are similar in shape with even 100% total parameter error, whereas they differ drastically in shape when their structures change. The exception is the displacement of the oscillator on the right from the model ($\ddot{x} = -\eta_1 x^2 \dot{x} + \eta_2 \dot{x} - \eta_3 x|x|$) which is very similar to that of the van der Pol oscillator having erroneous parameter values. At these erroneous parameter values, there will be little distinction between the correct and incorrect structures, making it difficult to identify the correct form.

There are three observations to be made of the results in Figure 1. One is that the models' responses are affected more drastically by the structure than the parameter values. This gives credence to prioritising exponent adaptation over parameter estimation. The second observation is that the shape of the responses is a better indicator of structural differences than their magnitudes. The model response shapes are incorporated in the evaluation of models by including the correlation coefficient between the models' responses and their targets in the fitness function. Consideration of output shapes in model evaluation is shown to improve the capacity for structure search in the presence of erroneous parameters and as a precursor to parameter estimation. The third observation corresponds to the drastic difference of model responses due to the

discrete nature of structural differences in Figure 1. Since such drastic differences would be detrimental to gradient-based adaptation, they are mediated by insertion of exponents into the models so as to regulate the size of adaptation.

3.3. Scaling of model changes

The prominence of structural accuracy to parametric accuracy and the difficulty in concurrent parameter and exponent estimation motivate focusing adaptation on the exponents, rendering the prediction error with the form

$$\epsilon(t) = y(t) - \hat{y}(t, \widehat{\mathbf{M}}, \widetilde{\Theta}) - \epsilon_\theta$$
$$\approx \sum_{i=1}^{Q} \widehat{\Delta \gamma_i} \left(\frac{\partial \hat{y}(t, \widehat{\mathbf{M}}, \widetilde{\Theta})}{\partial \gamma_i} \right) = \epsilon_\gamma = \Phi_\gamma \widehat{\Delta \Gamma} \quad (11)$$

where ϵ_θ denotes the parametric error approximated by its first-order expansion as

$$\epsilon_\theta \approx \sum_{i=1}^{Q} \widetilde{\Delta \theta_i} \left(\frac{\partial \hat{y}(t, \widehat{\mathbf{M}}, \widetilde{\Theta})}{\partial \theta_i} \right) \approx \Phi_\theta \widetilde{\Delta \Theta} \quad (12)$$

and

$$\Phi_\gamma = \begin{bmatrix} \partial \hat{y}(t_1, \widehat{\mathbf{M}}, \widetilde{\Theta})/\partial \gamma_1 \ \dots \ \partial \hat{y}(t_1, \widehat{\mathbf{M}}, \widetilde{\Theta})/\partial \gamma_Q \\ \vdots \ \ddots \ \vdots \\ \partial \hat{y}(t_N, \widehat{\mathbf{M}}, \widetilde{\Theta})/\partial \gamma_1 \ \dots \ \partial \hat{y}(t_N, \widehat{\mathbf{M}}, \widetilde{\Theta})/\partial \gamma_Q \end{bmatrix} \quad (13)$$

Success of adaptation in MSAM, therefore, relies on the quality of Φ_γ in Equation (11). Two factors can degrade the estimation of Φ_γ: (1) the presence of parametric error, ϵ_θ, as a bias in the prediction error, and (2) the non-uniformity of the columns of Φ_γ. As to the first factor, although the parameter error $\widetilde{\Delta \Theta}$ remains constant during adaptation, the parameter sensitivity matrix Φ_θ varies as a function of $\widehat{\Gamma}$. This variation causes the bias due to ϵ_θ to be non-constant during adaptation, hence, a shift in the gradient of the objective function. The second factor, namely the non-uniformity of the columns of Φ_γ, stems from the non-uniformity of structural changes $(f_i(\mathbf{x}, \mathbf{u}))^{\gamma_i}$. Unlike typical parameter perturbations that are applied to non-zero parameter values, the values of γ_i initialise at zero to modulate the introduction of functions. As such, their perturbations can have drastic effects on the outputs of the candidate models.

One possible approach to improving the condition of Φ_γ is to scale the columns of Φ_γ (Bates & Watts, 1988) by the magnitude of model difference caused by the perturbation $\delta \gamma_i$. We quantify the model difference magnitude in terms of parameter sensitivity, according to the following definition.

Definition: The difference between two models of the same structure but a different exponent; i.e., $\mathbf{M}_2 = \widehat{\mathbf{M}}(\Gamma + \Delta \gamma_i)$ and $\mathbf{M}_1 = \widehat{\mathbf{M}}(\Gamma)$ is quantified by the sum of the ℓ^2-norm of their parameter sensitivity difference over time, as

$$\Delta(\mathbf{M}_2, \mathbf{M}_1) = \sum_{k=1}^{N} \left\| \frac{\partial \hat{y}(t_k, \mathbf{M}_2, \widetilde{\Theta})}{\partial \Theta} - \frac{\partial \hat{y}(t_k, \mathbf{M}_1, \widetilde{\Theta})}{\partial \Theta} \right\|_2 \quad (14)$$

where $\Delta(\mathbf{M}_2, \mathbf{M}_1)$ is the model difference magnitude, $\partial \hat{y}/\partial \Theta = \left[\partial \hat{y}/\partial \theta_1, \dots, \partial \hat{y}/\partial \theta_Q \right]$ is the vector of parameter sensitivities at time t_k, and ℓ^2-norm for the vector $\mathbf{v} = [v_1, \dots, v_n]^T \in \mathbb{R}^n$ is defined as $||\mathbf{v}||_2 = \sqrt{\sum_i^n v_i^2} = (\mathbf{v}^T \mathbf{v})^{\frac{1}{2}}$.

For clarification of the above definition, consider the three harmonic oscillator models of increasing complexity below wherein the acceleration of each model is the estimated output obtained algebraically according to the measured displacement $\tilde{x}(t_k)$.

$$\hat{y}_1(t_k) = \widehat{\ddot{x}}(t_k) = -\frac{\tilde{c}}{\tilde{m}} \dot{\tilde{x}}(t_k) - \frac{\tilde{k}}{\tilde{m}} \tilde{x}(t_k) + \frac{1}{\tilde{m}} \tilde{u}(t_k);$$
$$\mathbf{M}_1 = \left[\dot{\tilde{x}}(t), \ \tilde{x}(t), \ \tilde{u}(t) \right]^T$$
$$\hat{y}_2(t_k) = \widehat{\ddot{x}}(t_k) = -\frac{\tilde{c}}{\tilde{m}} \dot{\tilde{x}}(t_k)|\dot{\tilde{x}}(t_k)| - \frac{\tilde{k}}{\tilde{m}} \tilde{x}(t_k) + \frac{1}{\tilde{m}} \tilde{u}(t_k);$$
$$\mathbf{M}_2 = \left[\dot{\tilde{x}}(t)|\dot{\tilde{x}}(t)|, \ \tilde{x}(t), \tilde{u}(t) \right]^T$$
$$\tilde{y}_3(t_k) = \widehat{\ddot{x}}(t_k) = -\frac{\tilde{c}}{\tilde{m}} \dot{\tilde{x}}(t_k)|\dot{\tilde{x}}(t_k)|^2 - \frac{\tilde{k}}{\tilde{m}} \tilde{x}(t_k) + \frac{1}{\tilde{m}} \tilde{u}(t_k);$$
$$\mathbf{M}_3 = \left[\dot{\tilde{x}}(t)|\dot{\tilde{x}}(t)|^2, \ \tilde{x}(t), \ \tilde{u}(t) \right]^T$$

Admittedly, $\mathbf{M}_3$ is more complex (denoted by $\Uparrow$) than $\mathbf{M}_2$, which is more complex than $\mathbf{M}_1$; i.e., $\mathbf{M}_3 \Uparrow \mathbf{M}_2 \Uparrow \mathbf{M}_1$. Then according to the above definition, the model difference magnitude $\Delta(\mathbf{M}_3, \mathbf{M}_1)$ should be as large or larger than $\Delta(\mathbf{M}_2, \mathbf{M}_1)$ as quantified by the norm of their parameter sensitivity differences; i.e.,

$$\Delta(\mathbf{M}_3, \mathbf{M}_1) \geq \Delta(\mathbf{M}_2, \mathbf{M}_1)$$
$$\implies \sum_{k=1}^{N} \left\| \frac{\partial \hat{y}(t_k, \mathbf{M}_3, \widetilde{\Theta})}{\partial \Theta} - \frac{\partial \hat{y}(t_k, \mathbf{M}_1, \widetilde{\Theta})}{\partial \Theta} \right\|_2$$
$$\geq \sum_{k=1}^{N} \left\| \frac{\partial \hat{y}(t_k, \mathbf{M}_2, \widetilde{\Theta})}{\partial \Theta} - \frac{\partial \hat{y}(t_k, \mathbf{M}_1, \widetilde{\Theta})}{\partial \Theta} \right\|_2$$

The above inequality can be confirmed analytically, as

$$\left\| \frac{\partial \hat{y}(\mathbf{M}_3, \widetilde{\Theta})}{\partial \Theta} - \frac{\partial \hat{y}(\mathbf{M}_1, \widetilde{\Theta})}{\partial \Theta} \right\|_2 = \sqrt{\left[\dot{\tilde{x}}(t)|\dot{\tilde{x}}(t)|^2 - \dot{\tilde{x}}(t) \right]^2} \geq$$

$$\left\| \frac{\partial \hat{y}(\mathbf{M}_2, \widetilde{\Theta})}{\partial \Theta} - \frac{\partial \hat{y}(\mathbf{M}_1, \widetilde{\Theta})}{\partial \Theta} \right\|_2 = \sqrt{\left[\dot{\tilde{x}}(t)|\dot{\tilde{x}}(t)| - \dot{\tilde{x}}(t) \right]^2}$$

It should be noted here that the symbolic form of parameter sensitivities, shown here for conceptual verification of the above definition, is not necessary for computation of the model difference magnitude, since it can be readily obtained numerically.

As discussed earlier, the model difference magnitude is defined to measure model changes affected by perturbations to the exponents γ_i in Equation (9). To this end, the model difference magnitude is computed for the perturbed model resulted from an exponential perturbation $\delta\gamma_i$ and normalised to render the 'model perturbation magnitude' δM_i as

$$\delta M_i = \frac{\sum_{k=1}^{N} \left\| \frac{\partial \hat{y}(t_k, \widehat{\Gamma} + \delta\gamma_i, \widetilde{\Theta})}{\partial \Theta} - \frac{\partial \hat{y}(t_k, \widehat{\Gamma}, \widetilde{\Theta})}{\partial \Theta} \right\|_2}{\sum_{k=1}^{N} \left\| \frac{\partial \hat{y}(t_k, \widehat{\Gamma}, \widetilde{\Theta})}{\partial \Theta} \right\|_2} \quad (15)$$

The scaling of structural sensitivity by δM_i then takes the form

$$\frac{\partial \hat{y}(t, \widehat{\Gamma}, \widetilde{\Theta})}{\partial \gamma_i} \approx \frac{\hat{y}(t, \widehat{\Gamma} + \delta\gamma_i, \widetilde{\Theta}) - \hat{y}(t, \widehat{\Gamma}, \widetilde{\Theta})}{\delta M_i} \quad (16)$$

wherein the δM_i are used in place of $\delta\gamma_i$ in the denominator of the finite difference approximation of the output sensitivity. The availability of the Jacobian Φ_γ enables estimation of the exponential errors $\Delta\gamma_i$ according to NLS, as

$$\widehat{\Delta\Gamma} = [\widehat{\Delta\gamma}_1, \ldots, \widehat{\Delta\gamma}_Q]^T = (\Phi_\gamma^T \Phi_\gamma)^{-1} \Phi_\gamma^T \epsilon^N \quad (17)$$

and consequent adaptation of the exponents, as

$$\gamma_i(q+1) = \gamma_i(q) + \mu(q)\Delta\gamma_i(q) \quad (18)$$

where q is the iteration number and $\mu(q)$ is the adaptation step size, determined at each iteration (see Section 4.2).

4. Algorithmic implementation

Given the adjustment strategy in Equation (8), adaptation in MSAM entails finding the adjustments $\mathbf{f}^*$ and their exponents γ_i*. To this end, adaptation is performed in two stages, wherein improvement of the candidate model(s) is attained through adaptation of the exponents γ_i. In the first stage, using a round robin format, all possible candidate models are adapted for their viability of improving the model's accuracy with a nominal number of exponent adaptations.
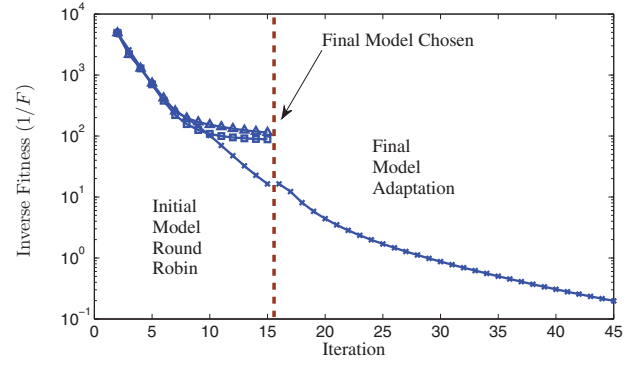
Figure 2. Illustration of candidate model selection by MSAM in the round robin stage, followed by further adaptation of the selected model in the second stage, as represented by the inverse of the fitness value for each model.

Given n adjustments and Q components, the number of candidate models would equal Q^n. The underlying assumption of the round robin stage is that the candidate model with the correct adjustments $\mathbf{f} = \mathbf{f}^*$ will achieve the best fitness in a fixed number of iterations. Therefore, there is always the possibility that MSAM may not find the correct adjustment set because of the limited number of iterations used in the round robin stage. In the second stage, the winner combination is further adapted via its exponents to enhance the model's accuracy.

The adaptation strategy is described in Algorithm 1. Adaptation begins by evaluating the candidate models, composed of unique sets of adjusted components, in a round robin fashion. At the end of the round robin stage, the candidate model associated with the best fitness is chosen for further adaptation in the second stage. For illustration purposes, selection of the best candidate model of the harmonic oscillator in the first stage, followed by its adaptation in the second stage, is shown in Figure 2. The plots in the first stage represent the fitness values of the candidate models during the first 15 iterations of adaptation. The inferior models are discarded for the second stage where adaptation is continued for the best-fit model.

4.1. Fitness function

The fitness function is used mainly to distinguish between the candidate models. Since MSAM emphasises structural adaptation, it requires a fitness criterion that is more sensitive to structural error than parametric error. A significant indicator of the model structure is the shape of the model output, as was observed in Figure 1. Therefore, included in the fitness function is the correlation coefficient between the model's output and its target so as to represent the closeness of the output's shape to its target (Billings & Zhu, 1994; Kommenda, Kronberger, Winkler, Affenzeller, & Wagner, 2013). Accordingly, the fitness function in

---

Algorithm 1 Model structure adaptation algorithm

1: $Q \longleftarrow$ number of model compartments in $\widehat{M}$
2: $n \longleftarrow$ number of states (order of the system)
3: $R \longleftarrow Q^n$
4: $\hat{M} \longleftarrow$ candidate model
5: $\xi \longleftarrow$ round robin iteration
6: $y \longleftarrow$ target output
7: $\hat{y}_\xi \longleftarrow$ current round robin model output
8: $\hat{M}_{\xi*} \longleftarrow$ best round robin model
9: $\hat{M}_{cb} \longleftarrow$ current best model
10: $\hat{M}_{best} \longleftarrow$ final best model
11: $I_1 \longleftarrow$ number of round robin iterations
12: $I_2 \longleftarrow$ numberof winner iterations
13: $F(\hat{y}_{cb}, y) \longleftarrow F(\hat{y}, y)$
14: Estimate $\tilde{\mathbf{x}}$
15: **for** $\xi \in R$ **do**   ▷ Round robin
16:       $\mathbf{f}_\xi(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}) \longleftarrow$ choose set of function perturbations
17:       $\hat{\Gamma}_\xi = 0$
18:       $\hat{M}_\xi \longleftarrow \sum_{i=1}^{Q} \tilde{\theta}_i \tilde{M}_i f_i(\tilde{\mathbf{x}}, \tilde{\mathbf{u}})^{\gamma_i}$ ▷ Perturb model compartments
19:       **for** $j \in I_1$ **do**   ▷ Adapt model $\hat{y}_\xi$
20:             Calculate $\Phi_\gamma(\hat{y}_\xi)$
21:             Update $\hat{\Gamma}_\xi$
22:             Evaluate $\hat{y}_\xi\left(\hat{\Gamma}_\xi\right)$
23:             **if** $F(\hat{y}_\xi, y) > F(\hat{y}_{cb}, y)$ **then**
24:                   $\hat{y}_{cb} \longleftarrow \hat{y}_\xi$
25:                   $\mathbf{f}^*(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}) \longleftarrow \mathbf{f}_\xi(\tilde{\mathbf{x}}, \tilde{\mathbf{u}})$ ▷ Save best set of function perturbations
26:             **end if**
27:       **end for**
28: **end for**
29: $\hat{y}_{\xi*} \longleftarrow \hat{y}_{cb}$
30: $F(\hat{y}_{cb}, y) \longleftarrow F(\hat{y}, y)$
31: **for** $j \in I_2$ **do**   ▷ Adapt winning model $\hat{y}_{\xi*}$
32:       Calculate $\Phi_\gamma(\hat{y}_{\xi*})$
33:       Update $\hat{\Gamma}_{\xi*}$
34:       Evaluate $\hat{y}_{\xi*}\left(\hat{\Gamma}_{\xi*}\right)$
35:       **if** $F(\hat{y}_{\xi*}, y) > F(\hat{y}_{cb}, y)$ **then**
36:             $\hat{M}_{cb} \longleftarrow \hat{M}_{\xi*}$
37:       **end if**
38: **end for**
39: $\hat{M}_{best} \longleftarrow \hat{M}_{cb}$

MSAM is defined as

$$F = \frac{\rho(\hat{y}, y)}{\sum_{k=1}^{N} |\epsilon(t_k)|} \qquad (19)$$

where $\rho(\hat{y}, y)$ denotes the correlation coefficient between the model's output $\hat{y}$ and its target y, computed as $\rho(\hat{y}, y) = C_{\hat{y}y}/\sigma_{\hat{y}}\sigma_y$ where $C_{\hat{y}y}$ is the covariance of $\hat{y}$ and y, and σ denotes standard deviation. The larger the fitness value

the closer the model is to its target, therefore, this fitness function is used primarily to evaluate the fitness of various candidate models in the first stage of adaptation by MSAM.

4.2. Selection of adaptation step size

The adaptation step size μ in Equation (18) specifies the confidence in the estimate of $\widehat{\Delta\Gamma}$ from Equation (17). Since this estimate is based on the approximation of the prediction error in Equation (11), its fidelity can be assessed by the accuracy of the prediction error approximation. As a measure of this accuracy, we use the correlation coefficient between the error and its first-order approximation (i.e., the two sides of Equation (11)) to characterise the closeness of approximation of the error shape by the estimated ϵ_γ. Accordingly, the adaptation step size at the iteration q is computed as

$$\mu(q) = \rho\left(\epsilon^N(q), \hat{\epsilon}_\gamma(q)\right) \qquad (20)$$

where $\epsilon^N = [\epsilon(t_1), \ldots, \epsilon(t_N)]^T$.

5. Application examples

The performance of MSAM is evaluated in two categories: (1) controlled tests, wherein the target model is known and the target data are the simulated output of this model used in lieu of measured observations, and (2) real-world tests, wherein the target model is unknown and the measured observations are obtained experimentally. The first tests are intended to examine whether the true underlying model forms can be attained by adaptation. The second tests demonstrate the applicability of the approach to real systems where there are no true models and the preferred models are those that minimise the prediction error.

5.1. Controlled tests

For the controlled validation of the method, three non-linear models of increasing non-linearity and order are adapted from simpler starting models. The first model is that of the non-linear harmonic oscillator, which has been used thus far to illustrate various aspects of the method. It has two model components ($Q = 2$) and two variables ($n = 2$), generating a round robin size of $Q^n = 2^2 = 4$. The second model is that of the van der Pol oscillator, consisting of three components ($Q = 3$) and two variables ($n = 2$). It not only has a larger round robin size than the harmonic oscillator ($3^2 = 9$) but also a non-linear coupling of velocity and position that needs to be identified. The third model is a third-order state-space model, where each state equation comprises three non-linear components. As such, this model poses a round robin size of 27 with heavily coupled components in each state equation.

Table 1. The three models sought by MSAM in the controlled tests.

	Harmonic oscillator	van der Pol oscillator	Third-order model
Target model	$\ddot{x} = -\frac{c}{m}\dot{x}\lvert\dot{x}\rvert - \frac{k}{m}x^3 + \frac{1}{m}u(t)$	$\ddot{x} + \eta(x^2-1)\dot{x} + x = 0$	$\begin{aligned}\dot{x}_1 &= \theta_1 x_1 x_3 + \theta_2 x_2 x_3 + \theta_3 x_3^2 \\ \dot{x}_2 &= \theta_4 x_1 x_2 + \theta_5 x_1 x_3 + \theta_6 x_2 x_3 \\ \dot{x}_3 &= \theta_7 x_1 x_2 + \theta_8 x_1 x_3 + \theta_9 x_2 x_3\end{aligned}$
Starting model	$\ddot{x} = -\frac{c}{m}\dot{x} - \frac{k}{m}x + \frac{1}{m}u(t)$	$\ddot{x} + \eta_1\dot{x} - \eta_2\dot{x} + \eta_3 x = 0$	$\begin{aligned}\dot{x}_1 &= \tilde{\theta}_1 x_1 + \tilde{\theta}_2 x_2 + \tilde{\theta}_3 x_3 \\ \dot{x}_2 &= \tilde{\theta}_4 x_1 + \tilde{\theta}_5 x_3 + \tilde{\theta}_6 x_2 \\ \dot{x}_3 &= \tilde{\theta}_7 x_1 + \tilde{\theta}_8 x_3 + \tilde{\theta}_9 x_2\end{aligned}$
Parameter values	$\begin{bmatrix} m^* \\ c^* \\ k^* \end{bmatrix} = \begin{bmatrix} 375 \\ 9800 \\ 130,000 \end{bmatrix}$	$\begin{bmatrix} \eta_1^* \\ \eta_2^* \\ \eta_3^* \end{bmatrix} = \begin{bmatrix} 1.5 \\ 1.5 \\ 1 \end{bmatrix}$	$\begin{aligned}{[\theta_1^*, \theta_2^*, \theta_3^*]}^T &= [-3, -2, -3]^T \\ {[\theta_4^*, \theta_5^*, \theta_6^*]}^T &= [-3, 1, -3]^T \\ {[\theta_7^*, \theta_8^*, \theta_9^*]}^T &= [3, 3, -1]^T\end{aligned}$
Excitation	Step input	$\mathbf{x}(0) = [0, -1]$	$\mathbf{x}(0) = [5, 0, 1]$
Perturbance functions	$\{\lvert\dot{\tilde{x}}\rvert, \lvert\tilde{x}\rvert\}$	$\{\lvert\tilde{x}\rvert, \lvert\dot{\tilde{x}}\rvert\}$	$\{x_1, x_2, x_3\}$

The three models sought by MSAM are shown in Table 1 along with the starting models, parameter values, inputs, and perturbation functions. The harmonic oscillator consisted of three components, two of which needed to be adapted to their counterparts in the target model as $\widetilde{M}_1 = \dot{\tilde{x}} \implies M_1^* = \dot{\tilde{x}}\lvert\dot{\tilde{x}}\rvert^{\gamma_1}$ and $\widetilde{M}_2 = \tilde{x} \implies M_2^* = \tilde{x}^{\gamma_2}$. For the van der Pol oscillator, the goal was to adapt its first component to $\eta_1\tilde{x}^2\dot{\tilde{x}}$, leaving the other two components practically untouched. To guarantee real-valued outputs for the third-order system, the perturbations were applied as $\text{sign}(f_i(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}))\lvert f_i(\tilde{\mathbf{x}}, \tilde{\mathbf{u}})\rvert^{\gamma_i}$. For this model, the three state variables were assumed to be accessible.

For algorithmic details, let us consider the adaptation of the harmonic oscillator model, in which the options to be considered for adaptation of the first and second components were $\widehat{M}_1 : \dot{\tilde{x}} \longrightarrow \dot{\tilde{x}}\lvert\tilde{x}\rvert^{\gamma_1}$ and $\dot{\tilde{x}} \longrightarrow \dot{\tilde{x}}\lvert\dot{\tilde{x}}\rvert^{\gamma_1}$ and $\widehat{M}_2 : \tilde{x} \longrightarrow \tilde{x}\lvert\tilde{x}\rvert^{\gamma_2}$ and $\tilde{x} \longrightarrow \tilde{x}\lvert\dot{\tilde{x}}\rvert^{\gamma_2}$. Each of the four models formed from the above options were adapted iteratively by adjusting the exponents γ_1 and γ_2 in 15 iterations. The best model form selected at the end of this first (round robin) stage was further adapted, by improving the exponents γ_1 and γ_2 over 70 more iterations. Adaptation tests were performed with parametric errors ranging from 0% to 200% of the true parameter values. A sample of initial and final outputs before and after structural adaptation is shown in Figure 3. The left plots show the outputs $\hat{\tilde{x}}$ of the starting model at different levels of parameter error together with the target output (note that the increased error in m provides the counterintuitive effect of bringing the starting model output closer to the target). The right plots compare with the target output the outputs of the adapted models at various parametric error levels. The results clearly indicate the effectiveness of MSAM in producing models with outputs that are very close to the target output despite parametric error levels up to 100%.

Even though we use the fitness value as the surrogate, the real goal is the fidelity of the adapted model. To evaluate the reproducibility and accuracy of the model forms achieved by MSAM, 10 adaptation trials were performed
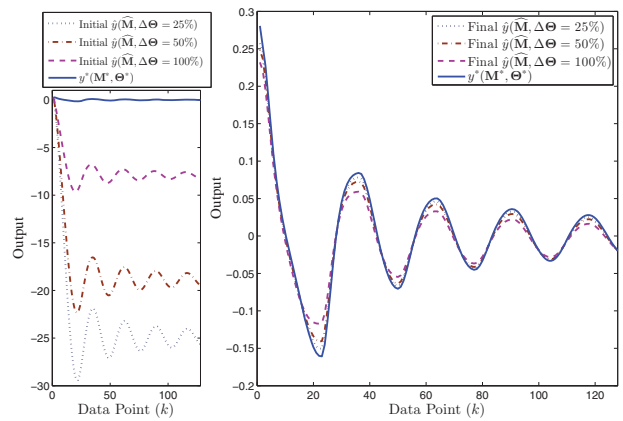

Figure 3. Sample of outputs (i.e., $\hat{\tilde{x}}$) before (left) and after adaptation (right) by MSAM of the non-linear harmonic oscillator at three levels of parametric error. The starting model in all cases was $\ddot{x} = \frac{1}{m}(-\tilde{c}\dot{x} - \tilde{k}x + u)$.

for each of the models with randomly generated parameters at various parametric error levels. The results obtained from these trials at parametric error levels of 0%, 25%, and 50% are shown in Table 2 for the harmonic and van der Pol oscillators, and in Table 3 for the third-order model. These tables show the error minimisation capacity of MSAM, in terms of the prediction error and correlation coefficient between the estimated and target outputs, and the median final model obtained for each error level. The results indicate that even though the fitness value is influenced by parametric error, the correct model form ($\hat{\mathbf{f}} = \mathbf{f}^*$) is achieved in all cases. The main difference between these models is in exponent values of the final models attained, which deviate from their correct values in accordance with the corresponding parametric error level. Noteworthy in the results is the high level of correlation achieved for the final model outputs at all levels of parameter error ($\rho > 0.999$). The fact that this level of success is not shared by the prediction error validates the lower sensitivity of output shape to parametric error and gives credence to the importance of including the

Table 2. Performance of MSAM for the non-linear harmonic oscillator and van der Pol oscillator in terms of the components of the fitness function (i.e., prediction error and correlation coefficient between the estimated and target outputs) and the model forms achieved for different levels of parameter error. The results are from 15 round robin iterations and 70 final choice iterations for the harmonic oscillator and 10 round robin iterations and 75 final choice iterations for the van der Pol oscillator. Ten trials runs were performed at each error level with the parameter values randomly selected. The ± quantities are the standard deviations over the trials.

	Starting		Final		
$\Delta\Theta$	$\sum\|\epsilon(t)\|$	$\rho(\ddot{\tilde{x}}, \hat{x})$	$\sum\|\epsilon(t)\|$	$\rho(\ddot{\tilde{x}}, \hat{x})$	Median final model
			Harmonic oscillator		
	Starting model: $\tilde{m}\ddot{x} + \tilde{c}\dot{x} + \tilde{k}x = u(t)$			Target model: $m^*\ddot{x} + c^*\dot{x}\|\dot{x}\| + k^*x^3 = u(t)$	
0% Θ^*	3780	0.7690	0.0534	0.9999	$\tilde{m}\ddot{x} + \tilde{c}\dot{x}\|\dot{x}\|^{1.00} + \tilde{k}x^{3.00} = u(t)$
25% Θ^*	3685±530	0.769±0.01	0.359±0.25	1.000±0.00	$\tilde{m}\ddot{x} + \tilde{c}\dot{x}\|\dot{x}\|^{0.984} + \tilde{k}x^{2.991} = u(t)$
50% Θ^*	3603±1049	0.772±0.02	0.689±0.43	0.999±0.00	$\tilde{m}\ddot{x} + \tilde{c}\dot{x}\|\dot{x}\|^{0.966} + \tilde{k}x^{2.973} = u(t)$
			van der Pol oscillator		
	Starting model: $\ddot{x} = -\tilde{\eta}_1\dot{x} + \tilde{\eta}_2\dot{x} - \tilde{\eta}_3 x$			Target model: $\ddot{x} = -\eta_1^* x^2\dot{x} + \eta_2^*\dot{x} - \eta_3^* x$	
0% Θ^*	276.12	0.446	5.8256	0.9997	$\ddot{x} = -\eta_1\|x\|^{2.02}\dot{x} + \eta_2\dot{x} - \eta_3 x\|\dot{x}\|^{0.03}$
25% Θ^*	193.5 ±8.393	0.394 ±0.053	25.484 ±5.524	0.995 ±0.003	$\ddot{x} = -\hat{\eta}_1\|x\|^{1.934}\dot{x} + \hat{\eta}_2\dot{x}\|x\|^{0.0226} - \hat{\eta}_3 x\|\dot{x}\|^{0.0634}$
50% Θ^*	205.0 ±21.19	0.250 ±0.0788	37.784 ±13.700	0.9914 ±0.00545	$\ddot{x} = -\hat{\eta}_1\|x\|^{2.355}\dot{x} + \hat{\eta}_2\dot{x}^{1.132} - \hat{\eta}_3 x\|\dot{x}\|^{-0.046}$

Table 3. Performance of MSAM for the three-variable set of ordinary differential equations (SODE) in terms of the components of the fitness function (i.e., prediction error and correlation coefficient between the estimated and target outputs) and the model forms achieved for different levels of parameter error. Ten trials runs were performed at each error level with the parameter values randomly selected. The results are from 15 round robin iterations and 100 final choice iterations for states 1 and 2, and 40 round robin iterations and 60 final choice iterations for state 3.

	Starting		Final			
$\Delta\Theta$	$\sum\|\epsilon(t)\|$	$\rho(\dot{\tilde{x}}_i, \hat{x}_i)$	$\sum\|\epsilon(t)\|$	$\rho(\dot{\tilde{x}}_i, \hat{x}_i)$	Success rate	Median final model
	Starting model: $\dot{x}_1 = \tilde{\theta}_1 x_1 + \tilde{\theta}_2 x_2 + \tilde{\theta}_3 x_3$			Target model: $\dot{x}_1 = \theta_1^* x_1 x_3 + \theta_2^* x_2 x_3 + \theta_3^* x_3^2$		
0% Θ^*	3043.217	0.763	403.833	0.993	100%	$\hat{x}_1 = \tilde{\theta}_1 x_1 x_3^{1.16} + \tilde{\theta}_2 x_2 x_3^{0.26} + \tilde{\theta}_3 x_3^{2.01}$
25% Θ^*	3300.194 ±240.851	0.743 ±0.046	440.222 ±120.129	0.998 ±0.002	100%	$\hat{x}_1 = \tilde{\theta}_1 x_1 x_3^{1.13} + \tilde{\theta}_2 x_2 x_3^{0.57} + \tilde{\theta}_3 x_3^{2.03}$
50% Θ^*	3656.362 ±509.102	0.719 ±0.091	747.273 ±239.304	0.995 ±0.017	80%	$\hat{x}_1 = \tilde{\theta}_1 x_1 x_3^{1.09} + \tilde{\theta}_2 x_2 x_3^{0.35} + \tilde{\theta}_3 x_3^{2.13}$
	Starting model: $\dot{x}_2 = \tilde{\theta}_4 x_1 + \tilde{\theta}_5 x_3 + \tilde{\theta}_6 x_2$			Target model: $\dot{x}_2 = \theta_4^* x_1 x_2 + \theta_5^* x_1 x_3 + \theta_6^* x_2 x_3$		
0% Θ^*	5890.721	0.127	49.201	1.000	100%	$\hat{x}_2 = \tilde{\theta}_4 x_1 x_2^{1.10} + \tilde{\theta}_5 x_1^{0.99} x_3 + \tilde{\theta}_6 x_2 x_3^{1.10}$
25% Θ^*	5966.269 ±362.675	0.12672 ±0.019	121.934 ±47.0814	0.998 ±0.0007	100%	$\hat{x}_2 = \tilde{\theta}_4 x_1 x_2^{1.11} + \tilde{\theta}_5 x_1 x_3^{0.98} + \tilde{\theta}_6 x_2 x_3^{1.01}$
50% Θ^*	6062.185 ±730.500	0.124 ±0.038	230.341 ±107.928	0.994 ±0.015	90%	$\hat{x}_2 = \tilde{\theta}_4 x_1 x_2^{1.17} + \tilde{\theta}_5 x_1 x_3^{0.87} + \tilde{\theta}_6 x_2 x_3^{0.92}$
	Starting model: $\dot{x}_3 = \tilde{\theta}_7 x_1 + \tilde{\theta}_8 x_3 + \tilde{\theta}_9 x_2$			Target model: $\dot{x}_3 = \theta_7^* x_1 x_2 + \theta_8^* x_1 x_3 + \theta_9^* x_2 x_3$		
0% Θ^*	4914.037	0.498	88.991	0.999	100%	$\dot{x}_3 = \tilde{\theta}_7 x_1 x_2^{1.10} + \tilde{\theta}_8 x_1^{1.10} x_3 + \tilde{\theta}_9 x_2 x_3^{0.61}$
25% Θ^*	4795.186 ±375.625	0.508 ±0.033	118.5305 ±29.0678	0.998 ±0.001	100%	$\hat{x}_3 = \tilde{\theta}_7 x_1 x_2^{1.12} + \tilde{\theta}_8 x_1^{1.07} x_3 + \tilde{\theta}_9 x_2 x_3^{0.52}$
50% Θ^*	4683.273 ±730.797	0.518 ±0.063	166.507 ±76.351	0.996 ±0.004	100%	$\hat{x}_3 = \tilde{\theta}_7 x_1 x_2^{1.12} + \tilde{\theta}_8 x_1^{1.04} x_3 + \tilde{\theta}_9 x_2 x_3^{0.48}$

correlation coefficient in the fitness function (see Equation (19)). Noteworthy are the results for the van der Pol oscillator, which indicate that the first component of the final model closely approximates its target ($\eta_1 x^2\dot{x}$), despite the presence of parametric error. The small exponents associated with the other component adjustments render them negligible. Also noteworthy are near unity values of the cor-

relation coefficients in Table 2 between the final model outputs and the target, indicating the lower sensitivity of output shapes to parametric error. Contrary to the correlation coefficient, the prediction error is sensitive to parametric error, as represented by its larger final values at higher parametric error levels. As to the third-order model, the third state required a higher number of round robin iterations (40) to
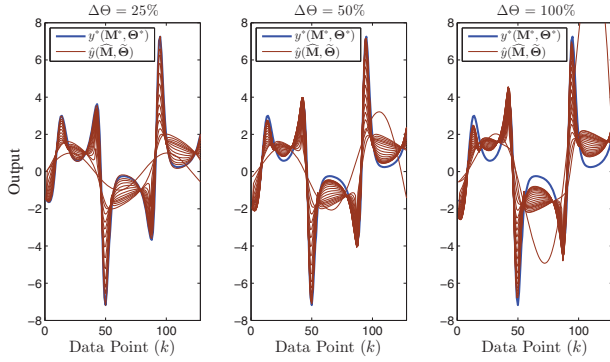
Figure 4. Progression of the outputs towards their van der Pol oscillator output target as the model structure is continually adapted by MSAM. The starting model was $\ddot{x} = -\tilde{\eta}_1\dot{x} + \tilde{\eta}_2\dot{x} - \tilde{\eta}_3 x$ with the parameters randomly selected within the parametric error range.

Figure 5. Condition number of $\Phi_\Gamma{}^T\Phi_\Gamma$ for various levels of parametric error using the structural sensitivities in Equation (16) with and without scaling by δM_i. The condition number is calculated for 30 trials of nine iterations of MSAM at parametric error levels of $\widehat{\Delta\Theta} = 25\%$, 50%, and 100%. The error bars represent standard deviation.

robustly select the correct model structure in the presence of parametric error. With the 15 iterations initially used to adapt this state model, MSAM consistently chose the wrong adjustment for its third component. Conversely, the 40 iterations in the round robin stage were adequate to produce consistently correct models at all parametric error levels.

Crucial to the success of MSAM is selection of the correct model form at the end of the round robin stage. To test this aspect of the performance of MSAM in presence of parametric error, the initial round robin stage was repeated for 50 sets of randomly generated parameters at each level of parametric error up to 200%. Success was declared when the correct model form in terms of adjustments (i.e., $\mathbf{f} = \mathbf{f}^*$) was chosen at the end of the round robin stage. The success rates for different levels of parametric error are shown in Table 4. The results indicate that MSAM is completely successful with parametric error levels of up to 50%, more than 90% successful with up to 100% parametric error, and more than 50% successful with up to 175% parametric error. These results underscore the capacity of MSAM in finding the correct model form despite considerable uncertainty in the parameter values, thus obviating concurrent search of both the structure and model parameters.

Even though secondary to proper model form selection, an important aspect of MSAM is adaptation of the exponents of the candidate model during the second stage. The progression of the selected model outputs towards their target for the van der Pol oscillator during second-stage adaptation of the model structure by MSAM is shown in Figure 4 for different levels of parametric error. The plots clearly indicate the convergence of the outputs toward their
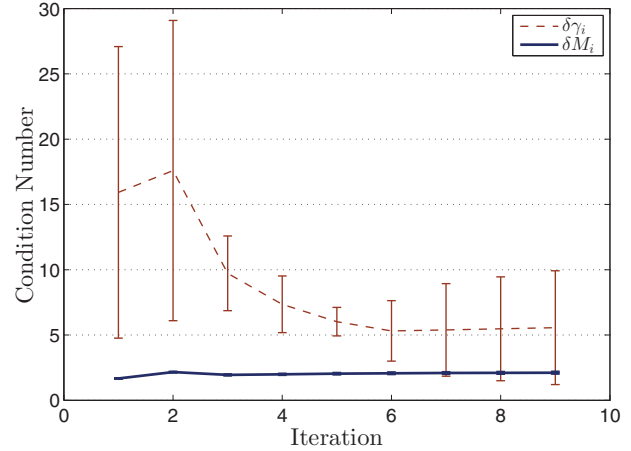
target despite different levels of parametric error. The only distinction at higher parametric errors is the larger distance between the final and target outputs.

Another important feature of MSAM is the use of δM_i from Equation (15) for scaling the columns of Φ_γ. A direct ramification of this scaling is the better quality of Φ_γ, that will result in better estimates of $\widehat{\Delta\Gamma}$ when used in Equation (17). The improved quality of Φ_γ is illustrated through its condition number ($\lambda_{max}/\lambda_{min}$), computed with and without scaling by δM_i at different levels of randomly generated parametric errors ($\widehat{\Delta\Theta}$), as shown in Figure 5. The condition numbers in Figure 5 are much smaller for δM_i-scaled Φ_γ, and given that the closer the condition number is to unity the more separate (less collinear) are the columns of the matrix (Jackson, 1991), the results in Figure 5 clearly indicate the marked improvement in the quality of Φ_γ scaled by δM_i. According to the results in Figure 5, not only are the condition numbers nearly always lower for the scaled Φ_γ, but they are also much less sensitive to the parameter error as evidenced by close to zero standard deviations at different levels of parametric error.

5.2. Real-world tests

To test MSAM's effectiveness in application to real-world cases, target outputs from two sets of experimental data were considered for construction of models. The first set

Table 4. Success rate of MSAM in finding the correct model form for the non-linear harmonic oscillator at different parametric error levels. Results are reported at the end of the round robin stage from 50 trials of randomly generated parameter values within each parametric error level.

$\Delta\Theta$:	0%	25%	50%	75%	100%	125%	150%	175%	200%
Success rate	100%	100%	100%	96%	92%	72%	62%	52%	42%

was experimental data obtained for flow-induced vibration of a slender beam. The second set was economic data from the Federal Reserve Economic Data (FRED) (http://research.stlouisfed.org/fred2/).

5.2.1. Flow-induced vibration

The area of fluid–structure interaction (FSI) offers a pertinent domain for non-linear modelling of coupled states. Briefly, when a flexible or flexibly mounted structure is placed in fluid flow, it can move due to the flow forces. The structure's motion changes the flow forces, which in turn affect the structure's motion, constituting an FSI problem. FSI is observed in wind turbines, offshore structures, novel energy extraction ideas, and biomedical engineering, among others (e.g., Bearman, 1984; Blevins, 1990; Païdoussis, 1998, 2004; Païdoussis, Price, & de Langre, 2004; Sarpkaya, 2004; Williamson & Govardhan, 2004).

For this case study, the experimental data were associated with a uniform cylinder placed in a re-circulating water tunnel, with a test section of 1.27 m × 0.5 m × 0.38 m, a turbulence intensity of less than 1% for up to a flow velocity of $U = 0.08$ m/s, and a velocity uniformity of less than 2% (Seyed-Aghazadeh, Budz, & Modarres-Sadeghi, 2015; Seyed-Aghazadeh, Carlson, & Modarres-Sadeghi, 2015). The set-up used to hold the cylinder in the test section had two air bearings to reduce the damping and constrain the oscillations of the cylinder to one degree of freedom in the crossflow direction. Springs were attached from the supporting plate holding the cylinder to the fixed housing. The cylinder's displacement and the corresponding flow forces were simultaneously measured at a flow velocity of $U = 0.076$ m/s. The experimental data were split into training and validation sets of 25 seconds in length.

The limit cycle characteristics of the dynamics associated with the displacement, x, of the cylinder and the applied force, q, motivate the use of the van der Pol oscillator model for representation of q, in lieu of solving the Navier–Stokes equation. Accordingly, the FSI model is considered to be (Facchinetti, de Langre, & Biolley, 2004)

$$(m_s + 1/4\pi C_M \rho D^2)\ddot{x} + [r_s 2\pi \mathrm{St}(U/D)\rho D^2]\dot{x} + hx = 1/4\rho U^2 D C_{Lo} q \quad (21)$$

$$\ddot{q} + \epsilon[2\pi \mathrm{St}(U/D)](q^2 - 1)\dot{q} + [2\pi \mathrm{St}(U/D)]^2 q = (A/D)\ddot{x} \quad (22)$$

where St is the Strouhal number, m_s is the mass of the structure, ρ is the fluid density, r_s represents viscous dissipation in the support, γ denotes the stall parameter, U is the free steam velocity, D is the cylinder's diameter, C_M is the added mass coefficient, C_{Lo} is the lift coefficient, and ϵ and A are the van der Pol scaling parameters.
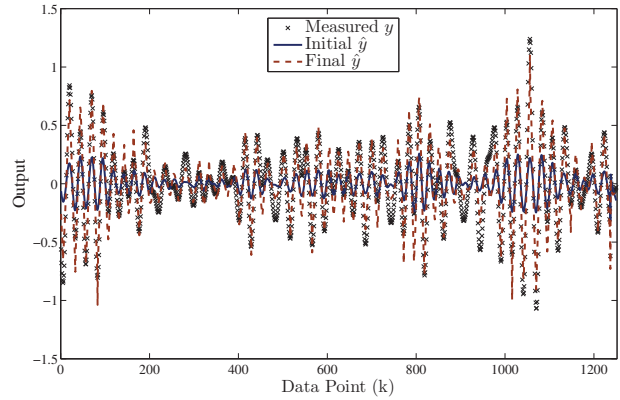

Figure 6. Adaptation of the van der Pol form for modelling vortex-induced vibration. The output of the initial and adapted models is shown against the validation data-set. The initial $\hat{y}$ is the van der Pol form (Equation (22)) with optimised parameters. The form of the final $\hat{y}$ is given in Table 5.

For this study, we focused our analysis on adaptation of the van der Pol oscillator model approximating the vortex force q (Equation (22)). We proceeded first by adapting the model's scaling parameters A and ϵ using NLS and then used the resulting model as the starting model to be adapted by MSAM. For this adaptation case, 20 round robin iterations were used for selection of the best model, choosing structural perturbations from the set $\{ |\ddot{q}|, |\dot{q}|, |q| \}$. The best candidate model was further improved via 20 iterations in the second stage. The resulting model had the form

$$(A/D)\widehat{\ddot{x}} = \ddot{q}|\dot{q}|^{\gamma_1} + \epsilon[2\pi \mathrm{St}(U/D)](\tilde{q}^2|\dot{q}|^{\gamma_2} - |\tilde{q}|^{\gamma_3})\dot{q} + [2\pi \mathrm{St}(U/D)]^2 \tilde{q}|\tilde{q}|^{\gamma_4} \quad (23)$$

with $\gamma_1 = -0.1178$, $\gamma_2 = -0.0023$, $\gamma_3 = 0.1036$, and $\gamma_4 = 3.2329$.

The results are summarised in Table 5 and the initial and final models from MSAM are plotted against the validation data-set in Figure 6. The results indicate that whereas parameter estimation of the original model only marginally improves the accuracy of the original model, the structurally adapted model is about 41.7% improved according to the magnitude of the prediction error.

5.2.2. A macroeconomic model

The second practical case study was the dynamic model of investment savings (IS) for the United States, which was developed based on economic data from the FRED. The nominal form of the IS model (Shone, 2002) is

$$\dot{y}(t) = \alpha(e(t) - y(t)) \qquad \alpha > 0 \quad (24)$$

where e represents total expenditure and y income (gross domestic product (GDP)).

Table 5. Adaptation of the FSI force equation using experimental results for $U = 0.076$ m/s. The exponents are $\gamma_1 = -0.1178$, $\gamma_2 = -0.0023$, $\gamma_3 = 0.1036$, and $\gamma_4 = 3.2329$.

	Training		Validation	
Model	$\sum\|\epsilon(t)\|$	$\rho(\ddot{\tilde{x}}, \hat{\ddot{x}})$	$\sum\|\epsilon(t)\|$	$\rho(\ddot{\tilde{x}}, \hat{\ddot{x}})$
Original model: $(A/D)\widehat{\ddot{x}} = \ddot{\tilde{q}} + \epsilon[2\pi\,\mathrm{St}(U/D)](\tilde{q}^2 - 1)\dot{\tilde{q}} + [2\pi\,\mathrm{St}(U/D)]^2\tilde{q}$	246.53	0.751	243.69	0.769
Parameter-tuned original model	245.45	0.780	242.31	0.803
Adapted model: $(A/D)\widehat{\ddot{x}} = \ddot{\tilde{q}}\|\dot{\tilde{q}}\|^{\gamma_1} + \epsilon[2\pi\,\mathrm{St}(U/D)](\tilde{q}^2\|\dot{\tilde{q}}\|^{\gamma_2} - \|\tilde{q}\|^{\gamma_3})\dot{\tilde{q}} + [2\pi\,\mathrm{St}(U/D)]^2\tilde{q}\|\tilde{q}\|^{\gamma_4}$	136.54	0.890	141.22	0.897
Parameter-tuned adapted model	137.20	0.891	141.33	0.898

Table 6. Adapted IS model by MSAM ($\gamma_1 = 0.7101$ and $\gamma_2 = -0.4419$).

	Training (1959–1989)		Validation (1990–2008)	
Model	$\sum\|\epsilon(t)\|$	$\rho(\dot{\tilde{y}}, \hat{\dot{y}})$	$\sum\|\epsilon(t)\|$	$\rho(\dot{\tilde{y}}, \hat{\dot{y}})$
Parameter-tuned original model: $\widehat{\dot{y}}(t) = \hat{\alpha}(\tilde{e}(t) - \tilde{y}(t))$	4944.38	0.519	6361.00	-0.417
Adapted model: $\widehat{\dot{y}}(t) = \hat{\alpha}(\tilde{e}(t)^{\gamma_1} - \tilde{y}(t)\|\tilde{y}(t)\|^{\gamma_2})$	678.40	0.973	2186.98	0.759

As in the previous case, the above model was parameter-tuned first before being adapted as the starting model by MSAM. The models were trained on historical data from 1959 to 1989, and validated against data from 1990 to 2008. The final model had the form

$$\widehat{\dot{y}}(t) = \hat{\alpha}(\tilde{e}(t)^{\gamma_1} - \tilde{y}(t)\|\tilde{y}(t)\|^{\gamma_2}) \tag{25}$$

where $\tilde{e}$ represented the historical total expenditure data and $\tilde{y}$ the historical income (i.e., GDP). Here, $\gamma_1 = 0.7101$ and $\gamma_2 = -0.4419$.

The fitness values of the model before and after adaptation are shown in Table 6, which again indicate that structural adaptation via MSAM leads to better error minimisation than parameter estimation on the nominal IS model (Equation (24)). In addition, the adapted model extrapolates better to unseen data, although some over-fitting can be seen in both models. Parameter tuning of Equation (25) did not yield any improvement in absolute error or correlation, hence its omission from Table 6.

6. Discussion

The adaptation results presented above demonstrate the effectiveness of MSAM in refining the model topology in presence of parametric uncertainty. The method is found to be effective for parametric errors of up to 50% in the case studies conducted. The error minimisation achieved is significant in all cases, even when the best model topology is not chosen due to large parametric errors. Yet the presented results, while they validate MSAM's adaptation strategy, leave several issues to be addressed in future studies. Such issues include the range of MSAM's capacity in coping with parametric error, its computational scalability,

its ability to cope with structural collinearity, reachability of potential model forms, consistency of model adaptation across different target measurements, the choice of input excitations, and the significance of measurement noise, as briefly discussed below.

- Robustness to parametric error: in the adaptation cases studied so far, MSAM could identify the correct adjustments and adapt the corresponding candidate model to nearly the exact model structure with parametric errors of up to 50%. The robustness of MSAM to parametric uncertainty is contingent upon two factors: (1) that the shape of the candidate model outputs not be substantially affected by parameter error (as illustrated in Figure 1) and (2) the most accurate parameter values be used for the nominal parameters. To enforce the first factor, the correlation coefficient between the model output and its target is included in the fitness function to underscore the significance of output shapes in the model selection process. The second factor can be satisfied by performing parameter estimation on the initial model with the hope of improving the parameter values.

- Scalability: the scalability of MSAM was demonstrated for systems of up to three adjustments applied to three model components. The main scalability issue is the number of candidate models considered during the round robin stage. Given that with n adjustments applied to Q components Q^n candidate models need to be examined during the round robin phase, the selection process can become overwhelming if the models are examined sequentially. However, candidate models can be evaluated separately and independently of each other. As such, the round robin

phase can be run in parallel, reducing the computation time to Q^n/p, with p processors used. Future research could focus on techniques for choosing subsets of round robin models of large-scale problems that cannot be exhaustively searched.

- Collinearity: as a gradient-based method, least-squares estimation depends on the non-singularity of the underlying Jacobian. This condition is not satisfied, for instance, seeking the Lotka–Volterra model of inter-species population dynamics, shown as the right-hand side model, using the starting model shown in the left

$$\dot{x} = \tilde{a}x - \tilde{b}y - \tilde{c}x \implies \dot{x} = ax - bxy - cx^2$$
$$\dot{y} = \tilde{a}y - \tilde{b}x - \tilde{c}y \implies \dot{y} = ay - bxy - cy^2$$

With this starting model, the first and third columns of the structural sensitivity matrix Φ_γ (Equation (13)) will be collinear for both state equations, due to the sole dependence of the first and third components on x and y, respectively. Given that structural adaptation of both components will be impossible in such a case, a possible recourse would be a sequential approach wherein one component is adapted at a time.

- Reachability: in general, MSAM is additive by nature, designed to adapt the potentially inadequate first-principles models of the system by the coupling of functions to individual model components. As such, this method is suited to starting models that are less complex than their targets. Furthermore, MSAM is conducted under the assumption that the starting model comprises adequate components for representing the system dynamics, therefore, it provides a sufficient basis for reaching the true model. Accordingly, MSAM's adaptation is restricted to adjustments made to the components of the starting model. One could, indeed, extend the reach of MSAM by expanding the components of the starting model to allow for higher granularity of adjustments, as was demonstrated in the adaptation of the van der Pol oscillator. However, one should be mindful of the fact that such expansions may lead to violating the original premise of MSAM as an efficient yet constrained alternative to symbolic regression.

- Choice of input excitation: as in all system identification cases, the suitability of the measured output in representing the system is a requisite of search for the true model. As such, the adapted model is only as good as the measurements representing the system. Since in practice one is limited to input excitations that are applicable to the process, the model should be adapted across all measurement sets available from the process to enhance its generality. A focus of our future research is the consistency

of adaptation by MSAM when faced with different observations.

- Measurement noise: measurement noise impedes adaptation by masking the true output of the process (Equation (1)). It also affects algebraic evaluation of candidate models by contaminating the numerical estimates of derivatives of measured observations. Although noise can be addressed to some extent by the application of smoothing and wavelet transforms (McCusker, Currier, & Danai, 2011), its presence can be as inhibiting as in other system identification methods.

7. Conclusion

A gradient-based method of model structure adaptation is introduced for refinement of starting models of the process. This method, which is designed to increase the nonlinearity of the starting model by adjusting its components, uses exponents of the introduced adjustments to reduce their coarseness as well as to make them conducive to gradient-based search. This method relies on parameter sensitivity of the model to quantify the magnitude of model perturbations for scaling the structural sensitivities. This scaling practice is shown to improve the condition number of the structural sensitivity matrix, thus the search for the correct model structure in presence of parametric uncertainty. Moreover, since the proposed scaling method is independent of the excitation input, it can be used with pre-calculated state variables to preclude simulation-based evaluation of candidate models. Algebraic evaluation of candidate models has the added benefits of allowing state equations to be adapted independently, as in the third-order SODE; it is computationally cheaper than numerical integration; and is immune to disruptions of simulation failures. The proposed method is evaluated in controlled tests, wherein the true model is known, as well as in application to real-world problems, wherein the measure of success is solely the improved fitness of the adapted model. The results indicate that MSAM finds the correct form of the model in controlled tests despite parametric errors of up to 50% and that it improves the models' fitness by a wide margin in its real-world applications.

Notes on contributors

William La Cava is a doctoral student at the University of Massachusetts, Amherst. He received his BS and MEng degrees from Cornell University in 2009 and 2010, and went on to work at the National Wind Technology Center in Boulder, Colorado. William is interested in automatic system identification and automatic control design for complex systems. As a National Science Foundation student fellow in the UMass IGERT Offshore Wind Energy Program, he is developing machine-learning methods to create models of offshore wind turbines and associated avian behavior.

Kourosh Danai received all his degrees from the University of Michigan in Ann Arbor, the last one a PhD in 1986. He joined the Mechanical Engineering Department of the University of Massachusetts Amherst in 1987, where he is now a professor. Dr Danai's research has focused on automation solutions that are inspired by artificial intelligence. With his students he has developed solutions for fault diagnosis of helicopter gearboxes and turbojet engines as well as optimisation and control of manufacturing processes. His current research is focused on system identification. He has been the recipient of three innovation awards from NASA and one US patent in system identification. He spent the summer of 1990 at Sikorsky Aircraft Company (working on helicopter track and balance), and the fall of 1994 at the United Technologies Research Center (working on sensor location selection in helicopter gearboxes). Dr Danai is a Fellow of the American Society of Mechanical Engineers.

References

Bates, D.M., & Watts, D.G. (1988). *Nonlinear regression analysis and its applications*. New York, NY: John Wiley & Sons.

Bearman, P.W. (1984). Vortex shedding from oscillating bluff bodies. *Annual Review of Fluid Mechanics, 16*(1), 195–222.

Billings, S.A. (2013). *Nonlinear system identification* (1st ed.). Chichester, UK: John Wiley & Sons.

Billings, S.A., & Zhu, Q.M. (1994). Nonlinear model validation using correlation tests. *International Journal of Control, 60*(6), 1107–1120.

Blevins, R.D. (1990). *Flow-induced vibration*. Malabar, FL: Krieger.

Bongard, J., & Lipson, H. (2007). Automated reverse engineering of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences, 104*(24), 9943–9948.

Cleveland, W.S., & Devlin, S.J. (1988). Locally-weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association, 83*, 596–610.

Dunstan, W.J., & Bitmead, R.R. (2003). *Nonlinear model validation using multiple experiments*. Proceedings of the 42nd IEEE Conference on Decision and Control, Maui, HI.

Facchinetti, M.L., de Langre, E., & Biolley, F. (2004). Coupling of structure and wake oscillators in vortex-induced vibrations. *Journal of Fluids and Structures, 19*(2), 123–140.

Gan, C., & Danai, K. (2000). Model-based recurrent neural network for modeling nonlinear dynamic systems. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, 30*(2), 344–351.

Gan, C., & Danai, K. (2001). Nonlinear state estimation by adaptive embedded RBF modules. *ASME Journal of Dynamic Systems, Measurement, and Control, 123*(1), 44–48.

Hertz, J., Krogh, A., & Palmer, R.G. (1991). *Introduction to the theory of neural computation*. Redwood City, CA: Addison Wesley.

Jackson, J.E. (1991). *A user's guide to principle components*. New York, NY: John Wiley & Sons.

Jammu, V.B., Danai, K., & Lewicki, D.G. (1998a). Experimental evaluation of a structure-based connectionist network for fault diagnosis of helicopter gearboxes. *ASME Journal of Mechanical Design, 120*(1), 106–112.

Jammu, V.B., Danai, K., & Lewicki, D.G. (1998b). Structure-based connectionist network for fault diagnosis of helicopter gearboxes. *ASME Journal of Mechanical Design, 120*(1), 100–105.

Kommenda, M., Kronberger, G., Winkler, S., Affenzeller, M., & Wagner, S. (2013, July). *Effects of constant optimization by nonlinear least squares minimization in symbolic regression*. Proceedings of GECCO'14, Amsterdam, The Netherlands.

Koza, J.R. (1992). *Genetic programming: On the programming of computers by means of natural selection*. Cambridge, MA: MIT Press.

La Cava, W., Spector, L., Danai, K., & Lackner, M.A. (2014, July). *Evolving differential equations with developmental linear genetic programming and epigenetic hill climbing*. Proceedings of GECCO'14, Vancouver, BC, Canada.

Ljung, L., & Glad, T. (1994). On global identifiability for arbitrary model parametrizations. *Automatica, 30*(2), 265–275.

McCusker, J.R., Currier, T., & Danai, K. (2011). Improved parameter estimation by noise compensation in the time-scale domain. *Signal Processing, 91*(1), 72–84.

Mitra, S., Konwar, K.M., & Pal, S.K. (2002). Fuzzy decision tree, linguistic rules and fuzzy knowledge-based network: Generation and evaluation. *IEEE Transactions on Systems, Man, Cybernetics – Part C: Applications and Reviews, 32*(4), 328–339.

Narendra, K.S., & Parthasarathy, K. (1990). Identification and control of dynamical systems using neural networks. *IEEE Transactions on Neural Networks, 1*(1), 4–27.

Narendra, K.S., & Parthasarathy, K. (1991). Gradient methods for the optimization of dynamical systems containing neural networks. *IEEE Transactions on Neural Networks, 2*(2), 252–262.

Païdoussis, M. (2004). *Fluid-structure interactions* (Vol. 2). London: Academic Press.

Païdoussis, M.P. (1998). *Fluid-structure interactions, Vol. 1: Slender structures and axial flow*. San Diego, CA: Academic Press.

Païdoussis, M.P., Price, S.J., & de Langre, E. (2004). *Fluid-structure interactions* (Vol. 1). New York, NY: Cambridge University Press.

Popper, K. (1959). *The logic of scientific discovery* (Vol. 1, 4th ed.). London: Hutchinson.

Popper, K. (1994). *The myth of the framework* (Vol. 1). New York, NY: Routledge.

Sarpkaya, T. (2004). A critical review of the intrinsic nature of vortex-induced vibrations. *Journal of Fluids and Structures, 19*(4), 389–447.

Schmidt, M., & Lipson, H. (2009). Distilling free-form natural laws from experimental data. *Science, 324*(5923), 81–85.

Schmidt, M.D., & Lipson, H. (2009). Incorporating expert knowledge in evolutionary search: A study of seeding methods. In *Proceedings of the 11th Annual conference on Genetic and Evolutionary Computation* (Gecco '09), Montreal, QC, Canada, July 8–12, 2009, New York, NY: ACM, pp. 1091–1098.

Seyed-Aghazadeh, B., Budz, C., & Modarres-Sadeghi, Y. (2015). The influence of higher harmonic flow forces on the response of a curved circular cylinder undergoing vortex-induced vibration. *Journal of Sound and Vibration, 353*, 395–406.

Seyed-Aghazadeh, B., Carlson, D., & Modarres-Sadeghi, Y. (2015). The influence of taper ratio on vortex-induced vibrations of tapered cylinders in the crossflow direction. *Journal of Fluids and Structures, 53*, 84–95.

Shone, R. (2002). *Economic dynamics: Phase diagrams and their economic application*. Cambridge: Cambridge University Press.

Towell, G.G., & Shavlik, J.W. (1994). Knowledge-based artificial neural networks. *Artificial Intelligence, 70*(1–2), 119–165.

Williamson, C., & Govardhan, R. (2004). Vortex-induced vibrations. *Annual Review of Fluid Mechanics, 36*, 413–455.

Zhu, Z., Danai, K., & McCormick, J. (1994). Control design of a high speed robot arm. *SME Transactions on Robotics Research, 3*, 13:1–14.