Grey-box modelling of a friction-affected dynamical system

P.L.Green[1], M.Hendijanizadeh[2], and S.J.Elliott[2]

[1]Institute for Risk and Uncertainty, Centre for Engineering Sustainability , School of Engineering , University of Liverpool , Liverpool, UK
[2]Institute of Sound and Vibration Research, University of Southampton, Southampton, UK

November 3, 2015

Abstract

In this paper a grey-box model of a nonlinear dynamical system is constructed. This involves using a Gaussian process to emulate model error - the error that arises as a result of flaws in one's physical-law based model of the system. The work shows how such an approach can be extended towards *dynamical* systems. Specifically, it is applied to experimental data, obtained from a dynamical system whose response is known to be strongly influenced by friction effects.

1 Introduction

Grey box models are typically described as a combination of a 'white-box model', whose equations of motion have been derived from the underlying physics of the problem of interest, and a 'black-box model', which is purely data-based (examples of black-box models include neural networks and Gaussian processes). In the current paper, the white box model consists of the hypothesised equation of motion of an experimental dynamical system, investigated in [1], whose response is known to be influenced by friction effects. This is augmented with a Gaussian process which is designed to emulated *model error*, thus accounting for the shortcomings of the white-box model.

This concept was given a very general treatment in [2, 3] and, within an engineering context, has been applied for a variety of purposes (see [4, 5, 6, 7] for example). The current paper differs from previous work in that it shows how Gaussian processes can be used to emulate the errors present in *dynamical* models - not just just those that are static.

2 Experiment

A schematic of the nonlinear dynamical system in question is shown in Figure 1. This device, which was originally designed to harvest electrical energy from ambient vibrations, is composed of a sprung mass which is coupled to an electrical generator by a ball-screw (this allows low frequency translational motion to be converted to high frequency rotary motion).

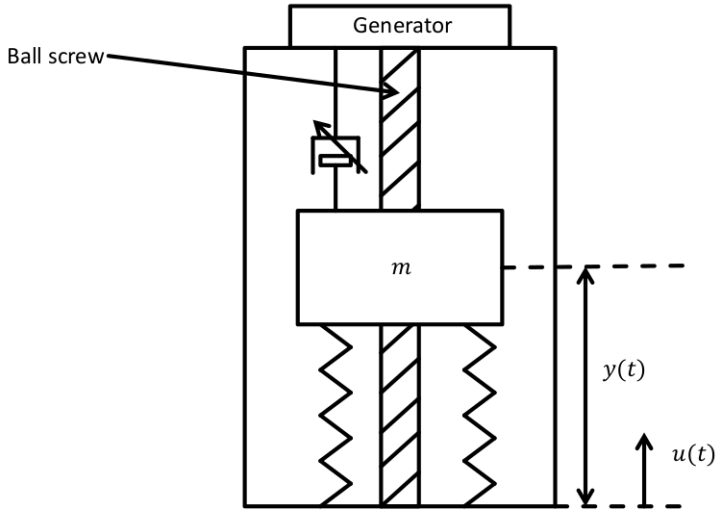

Figure 1: Schematic of friction-affected dynamical system.

Defining l as the ball-screw lead, c_m as mechanical damping, k as spring stiffness, m as the oscillating mass and J as the moment of inertia of the system, it is hypothesised that the system's equation of motion is

$$M\ddot{y} + b_m\dot{y} + ky + f(\dot{y}) = -m\ddot{u} \tag{1}$$

where u is the displacement of the base, y is the relative displacement between the mass and the base,

$$M = m + J\left(\frac{2\pi}{l}\right)^2, \tag{2}$$

$$b_m = \left(\frac{2\pi}{l}\right)^2 c_m \tag{3}$$

and $f(\dot{y})$ is a friction model. Based on the work conducted in [1], it is hypothesised that

$$f(\dot{y}) = F_c \tanh(\beta\dot{y}) \tag{4}$$

where F_c and β are parameters which require estimation. Equation (1) represents the white-box model of the device.

Figure 2 shows the experimental setup used for testing. The device was mounted to an electro-hydraulic shaker and MEMS accelerometers were attached to the oscillating mass and the shaker table. A Gaussian noise base acceleration - filtered by a low-pass filter with cutoff frequency of 15 Hz - was used to perturb the device. Acceleration time histories of the base and the mass were measured in each test. Throughout this work these are denoted by the vectors $\ddot{\boldsymbol{u}} = \{\ddot{u}_1, \ddot{u}_2, ...\}^T$ and $\ddot{\boldsymbol{z}} = \{\ddot{z}_1, \ddot{z}_2, ...\}^T$ respectively. It is important to note that $\ddot{\boldsymbol{z}}$ represents the *measured* relative acceleration while $\ddot{\boldsymbol{y}}$ represents the relative acceleration *predicted by the white-box model.*

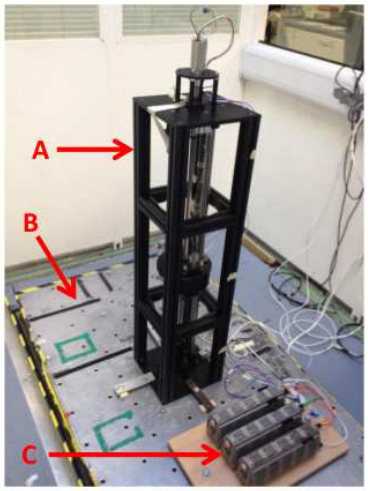

Figure 2: Experimental setup.

3 White-box modelling

Through a combination of the work conducted in [1] and [8], the 'best guess' parameter estimates shown in Table 1 were realised. It should be noted that, utilising a Bayesian framework, probabilistic estimates of those parameters which were difficult to measure directly were realised in [1]. The resulting parameter uncertainties were found to have little influence on the model's ability to make predictions and, as such, only crisp parameter estimates are used in the remainder of this study.

Parameter	Estimate	Unit
M	12.8	kg
m	8	kg
k	250	N/m
c_m	109.5	Ns/m
F_c	11.86	N
β	88.35	s/m

Table 1: Parameters of white-box model.

Using these parameter estimates, the ability of the white-box model (equation (1)) to predict a set of relative acceleration measurements in shown in Figure 3. While the model is able to represent some of the measured behavior there is clearly room for improvement. In the current paper, this is achieved through the addition of a data-based element to the white-box model (details in the next section).

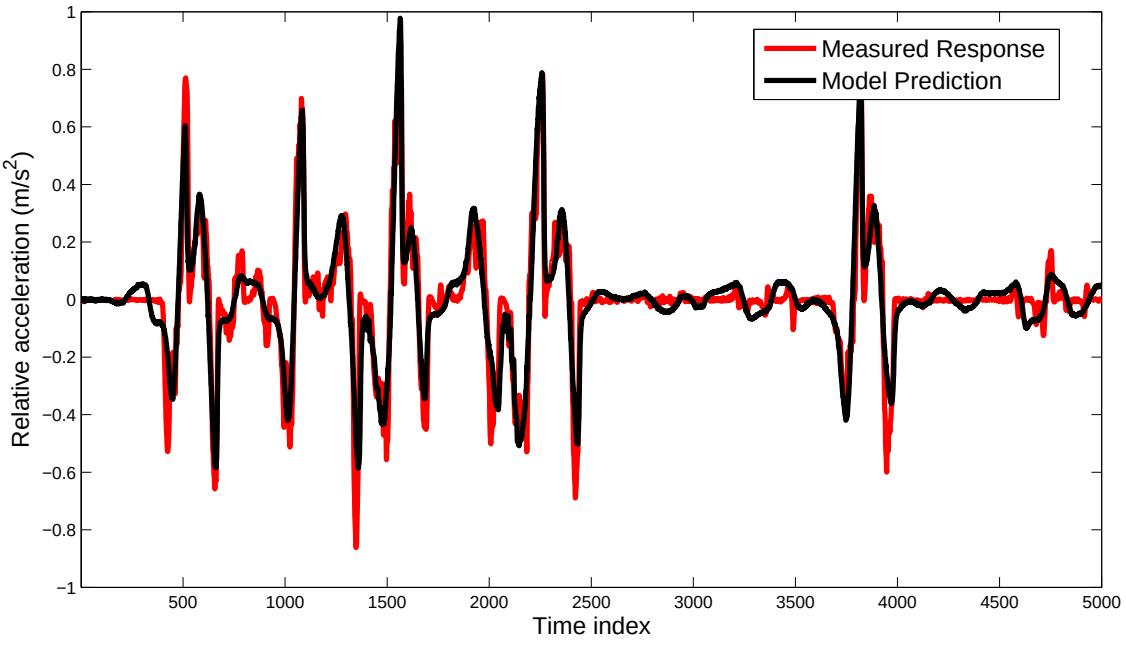

Figure 3: Predictions made by white-box model.

4 Model error

Building on the work[2], it is hypothesised that the measured response of the device is related to the output of the physical-law based model, $\ddot{\boldsymbol{y}}$, by

$$\ddot{z}_i = \ddot{y}_i + \eta_i(\boldsymbol{x}_i) + \epsilon \qquad \epsilon \sim \mathcal{N}(0, \beta^{-1}) \tag{5}$$

where ϵ represents measurement noise, η - the 'discrepancy term' - accounts for the dynamics that cannot realised using the white-box model and $\boldsymbol{x}$ represents an input which will be defined shortly. The aim then, is to develop a data-based model which, given the inputs $\boldsymbol{x}_1, \boldsymbol{x}_2, ...$, can predict the model discrepancies $\eta_1, \eta_2, ...$

Here the vector $\boldsymbol{\eta} = \{\eta_1, \eta_2, ...\}^T$ is modelled using a Gaussian process (GP). This involves first hypothesising a Gaussian prior distribution over the *functions* $\boldsymbol{\eta}$:

$$p(\boldsymbol{\eta}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{A}), \qquad A_{ij} = a(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{6}$$

where $\boldsymbol{A}$ is a user-defined covariance matrix and a is known as the 'kernel function'. This is convenient as, by defining the kernel function, it is possible to introduce correlations between $\eta_i(\boldsymbol{x}_i)$ and $\eta_j(\boldsymbol{x}_j)$ which depend on the closeness of the inputs $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. In the current work this is achieved by defining the kernel function as

$$a(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\alpha}{2}(\boldsymbol{x}_i - \boldsymbol{x}_j)^T(\boldsymbol{x}_i - \boldsymbol{x}_j)\right) \tag{7}$$

where α is a hyperparameter. Defining

$$\hat{\boldsymbol{\eta}} = \boldsymbol{\eta} + \boldsymbol{\epsilon} \tag{8}$$

such that

$$p(\hat{\boldsymbol{\eta}}|\boldsymbol{\eta}) = \mathcal{N}(\boldsymbol{\eta}, \beta^{-1}\boldsymbol{I}). \tag{9}$$

then, from equations (6) and (9), it follows that

$$p(\boldsymbol{t}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{C}) \tag{10}$$

where

$$\boldsymbol{C} = \beta^{-1}\boldsymbol{I} + \boldsymbol{A} \implies C_{ij} = \beta^{-1}\delta_{ij} + a(\boldsymbol{x}_i, \boldsymbol{x}_j). \tag{11}$$

At this point it is useful to recall that the aim here is to develop a data-based model which, given a new input $\boldsymbol{x}^*$, can be used to analyse the probability of the model discrepancy, $\hat{\eta}^*$. Using equation (10), the joint probability distribution of $\{\hat{\boldsymbol{\eta}}, \hat{\eta}^*\}$ is

$$p(\hat{\boldsymbol{\eta}}, \hat{\eta}^*) = \mathcal{N}\left(\boldsymbol{0}, \left[\begin{array}{cc} \boldsymbol{C} & \boldsymbol{a} \\ \boldsymbol{a}^T & c \end{array}\right]\right) \tag{12}$$

where

$$\boldsymbol{a}_n = a(\boldsymbol{x}_n, \boldsymbol{x}^*), \qquad n = 1, ..., N \tag{13}$$

and

$$c = \beta^{-1} + a(\boldsymbol{x}^*, \boldsymbol{x}^*). \tag{14}$$

Using some of the fundamental properties of Gaussian distributions (see [9] for example), it is then possible to show that

$$p(\hat{\eta}^*|\hat{\boldsymbol{\eta}}) = \mathcal{N}(\mu, \sigma^2) \tag{15}$$

where

$$\mu = \boldsymbol{a}^T\boldsymbol{C}^{-1}\hat{\boldsymbol{\eta}}, \qquad \sigma^2 = c - \boldsymbol{a}^T\boldsymbol{C}^{-1}\boldsymbol{a}. \tag{16}$$

GPs are typically presented as a way of modelling a *static* relationship between a set of inputs and a set of outputs. However, by including the past response of a system in the GP inputs, they can also be used to replicate the behaviour of *dynamical* systems. An example of this method - typically referred to as a GP-NARX modelling - can be found in [10]. Here it assumed that there is a dynamical relationship between the response of the system and the model error. To that end, the inputs of the GP are (somewhat arbitrarily) defined such that they include the two previous measurements of model error and the current excitation measurement:

$$\boldsymbol{x}_n = \left\{\begin{array}{c} \ddot{\eta}_{n-2} \\ \ddot{\eta}_{n-1} \\ \ddot{u}_n \end{array}\right\}. \tag{17}$$

Such a formulation allows the GP NARX model to make a 'one step ahead' prediction, η_n^*. To make long term predictions, the inputs to the GP NARX model must contain its own previous estimates such that, for example, an input of the form

$$\boldsymbol{x}_n = \left\{\begin{array}{c} \ddot{\eta}^*_{n-2} \\ \ddot{\eta}^*_{n-1} \\ \ddot{u}_n \end{array}\right\} \tag{18}$$

is required. As η^*_{n-1} and η^*_{n-2} are themselves uncertain, such 'full model' predictions can, in some cases, become associated with high levels of uncertainty.

Before the GP can be implemented, one must estimate the hyperparameters $\boldsymbol{\theta} = \{\alpha, \beta\}$. In the current work this was realiesd through maximisation of the log-likelihood function

$$\log p(\hat{\boldsymbol{\eta}}|\boldsymbol{\theta}) = \frac{1}{2}\left(N \log 2\pi + \log|\boldsymbol{C}| + \boldsymbol{t}^T \boldsymbol{C}^{-1} \boldsymbol{t}\right) \tag{19}$$

(achieved using gradient descent - see [9] for more information).

5 Grey-box modelling

1000 points of time-history were used to train the GP. The grey-box model was then used to predict a further 4000 points of data. Utilising one step ahead predictions, Figure 4 compares the prediction of the white-box model with the mean prediction made by the grey-box model. It is clear that, at both high and low amplitudes, the grey-box model is able to outperform the white-box model. As the variance of the GP predictions is available (equation (16)), it is also possible to quantity the uncertainty in one's predictions of dynamical model error. Figure 5 shows a close-up of the predictions made by the grey-box model, including 3 σ confidence bounds. It is clear that the confidence bounds nicely enclose the measured data. Unfortunately, as predicted in the previous section, full model predictions of model error led to much higher levels of uncertainty. The authors are currently investigating ways of mitigating the uncertainties associated with emulating dynamical systems.

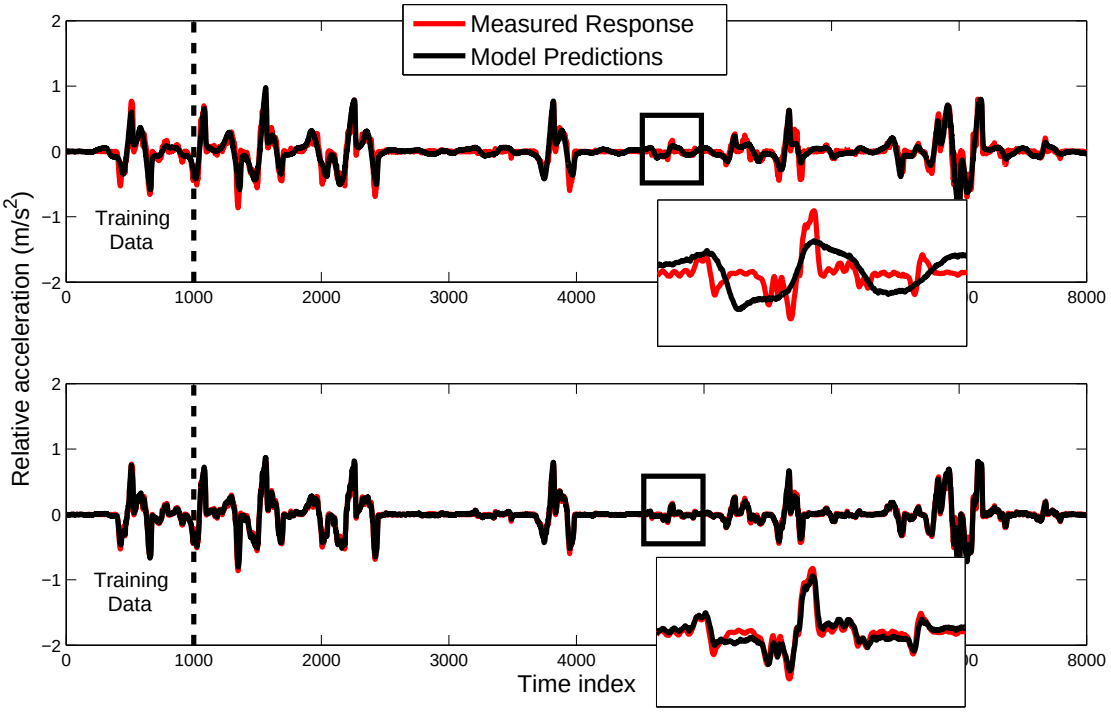

Figure 4: (a) predictions made by the white-box model and (b) average predictions made by the grey-box model.

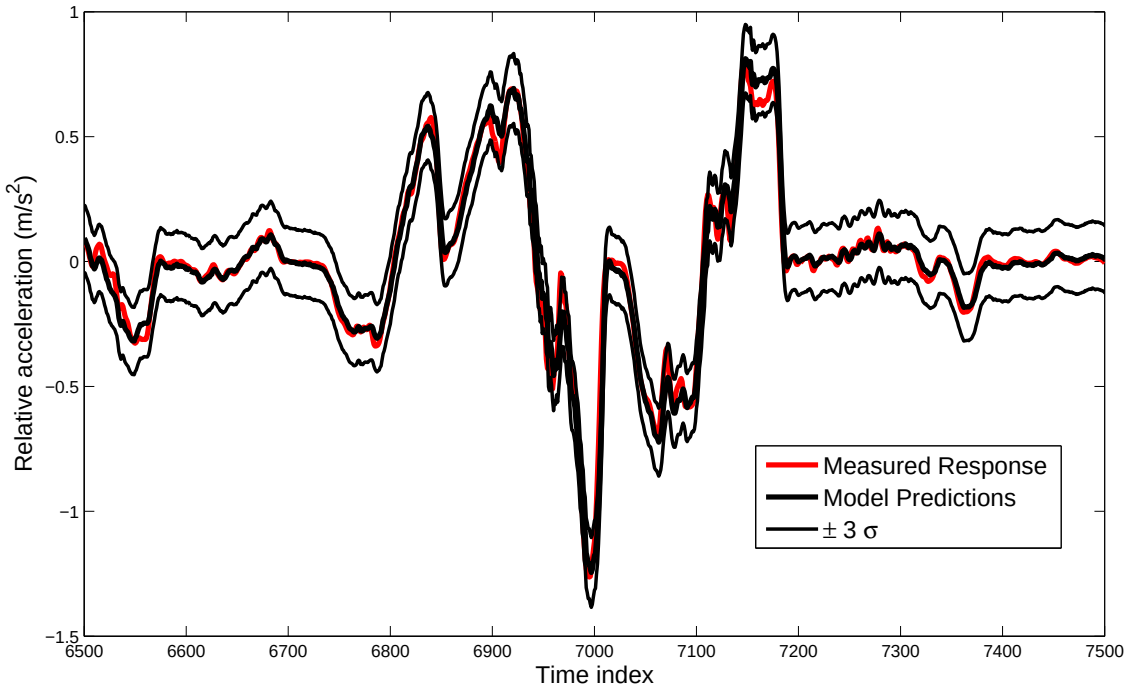

Figure 5: Predictions made by the grey-box model, including 3 σ confidence bounds.

6 Discussion

In this paper, to facilitate the emulation of dynamical model error, it was necessary to include a certain 'lag' in the GP inputs. The amount of lag which is required for the GP to work effectively is not clear at this stage. For future work the authors intend to use Automatic Relevance Determination to assess the relative importance of all the elements contained in the input vector, $\boldsymbol{x}$. It is hoped that this will give a more thorough insight into the amount of lag which should be included in the GP inputs.

It is also worth noting that, in the current study, the influence of parameter uncertainties were not considered - the parameters of the white-box model and the GP hyperparameters ($\boldsymbol{\theta}$ and $\{\alpha, \beta\}$ respectively) were held fixed to their most-probable values. This is because, based on the author's experience, parametric uncertainty tends to be very small relative to the uncertainties involved in predicting model error (an observation also noted in [2]). Typically, if one did want to include parameter uncertainties in the analysis, numerical sampling methods would be need to be used to generate samples from the posterior $p(\boldsymbol{\theta}, \alpha, \beta|\mathcal{D})$ (where $\mathcal{D}$ represents the entire set of experimental data). The authors believe that advanced Markov chain Monte Carlo (MCMC) methods would be well suited to this task (the merits of MCMC within the context of structural dynamics are discussed in [11]).

The current paper only shows the ability of the GP NARX to perform accurate one step ahead predictions of model error. For future work, the authors aim to reduce the uncertainties associated with full model predictions of model error.

References

[1] P.L. Green, M. Hendijanizadeh, L. Simeone, and S.J. Elliott. Probabilistic modelling of a rotational energy harvester. *Journal of Intelligent Material Systems and Structures*, page 1045389X15573343, 2015.

[2] M.C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, pages 425–464, 2001.

[3] J. Brynjarsdóttir and A. O'Hagan. Learning about physical parameters: The importance of model discrepancy. *Inverse Problems*, 30(11):114007, 2014.

[4] F. Hemez, S. Atamturktur, and C. Unal. Defining predictive maturity for validated numerical simulations. *Computers & structures*, 88(7):497–505, 2010.

[5] S. Atamturktur, F. Hemez, B. Williams, C. Tome, and C. Unal. A forecasting metric for predictive modeling. *Computers & Structures*, 89(23):2377–2387, 2011.

[6] D. Higdon, J. Gattiker, B. Williams, and M. Rightley. Computer model calibration using high-dimensional output. *Journal of the American Statistical Association*, 103(482), 2008.

[7] M.J. Bayarri, J.O. Berger, R. Paulo, J. Sacks, J.A. Cafeo, J. Cavendish, C.H. Lin, and J. Tu. A framework for validation of computer models. *Technometrics*, 49(2), 2007.

[8] L. Simeone, M. Ghandchi Tehrani, S.J. Elliott, and M. Hendijanizadeh. Nonlinear damping in an energy harvesting device. *Proceedings of ISMA 2014, International Conference on Noise and Vibration Engineering, Leuven, Belgium*, 2014.

[9] C.M. Bishop. *Pattern recognition and machine learning*. springer, 2006.

[10] K. Worden, G. Manson, and E.J. Cross. On Gaussian Process NARX models and their higher-order frequency response functions. In *Solving Computationally Expensive Engineering Problems*, pages 315–335. Springer, 2014.

[11] P.L. Green and K. Worden. Bayesian and Markov chain Monte Carlo methods for identifying nonlinear systems in the presence of uncertainty. *Phil. Trans. R. Soc. A*, 373(2051):20140405, 2015.