APPLICATION OF A KERNEL METHOD IN MODELING FRICTION DYNAMICS

Yufeng Wan [*], **Chian X. Wong** [**], **Tony J. Dodd** [*], **Robert F. Harrison** [*]

[*] *Department of Automatic Control and Systems Engineering, The University of Sheffield, UK*
[**] *Department of Mechanical Engineering, The University of Sheffield, UK*

Abstract: A kernel method has been developed to model finite degree, finite memory length and infinite degree, finite memory length Volterra series using polynomial and exponential kernels, respectively. Here, the kernel method is extended to identify NARX (Nonlinear AutoRegressive with eXogenous inputs) models. To verify its effectiveness, the proposed approach is used in modeling friction dynamics, which is an important and complex mechanical process. The simulation results are compared with those from a physical model and a specialized neural network, and the advantages and disadvantages of the methods are shown.

Keywords: Intelligent machines, Volterra series, nonlinearity, friction, neural networks, system identification

1. INTRODUCTION

Volterra series (VS) are widely used in nonlinear system identification problems. Nevertheless, their application is frequently restricted by the computational burden of estimating exponentially large number of parameters. A kernel method has been developed to solve this problem and model finite degree, finite memory length and infinite degree, finite memory length Volterra series using polynomial and exponential kernels, respectively (Harrison, 1999; Drezet, 2001; Dodd and Harrison, 2002a; Dodd and Harrison, 2002b; Wan *et al.*, 2003). Meanwhile, NARX (Nonlinear AutoRegressive with eXogenous inputs) models form a more general class of important nonlinear models. By involving the dependent variable, fewer coefficients are required in NARX models than in VS to describe the same system and there is no longer a need for the "fading memory" assumption. However, because of the feed-back, NARX models can be unstable. A kernel method is used in this document to identify finite and infinite degree NARX models based on data from three physical friction processes.

Friction is a very important and complex mechanism in the real world. It is defined as the resistance which any object meets with when sliding over another object. There are various types of friction, all of which are highly nonlinear in displacement and velocity except viscous friction where friction force is proportional to velocity (Parlitz *et al.*, 2004). Since friction is ubiquitous, identification of its behavior is crucial in many systems and control situations. Three types of friction process are discussed in this document including pre-sliding, a regime prior to true sliding, characterized by hysteretic dependence of the friction force on the displacement, true sliding and an intermediate process between these two.

In the next section, the kernel method to identify VS and NARX models is discussed. Introduction of the friction data generator and the identification experiments are proposed in Section 3. Experimental results are also presented in this section. Comparison of the kernel method with a physically derived model, the Dynamic Nonlinear Regression (DNLR) Maxwell slip model (Rizos and Fassois, 2004), and an Acceleration Velocity Displacement (AVD) (Chen and Tomlinson, 1996) neural network (NN) is presented based on their prediction performance in the time domain in Section 4.

2. IDENTIFICATION OF VS AND NARX MODELS USING THE KERNEL METHOD

The identification of VS using the kernel method has been presented in (Dodd and Harrison, 2002a; Dodd and Harrison, 2002b; Wan *et al.*, 2003).

Consider a time-invariant, discrete-time, finite memory length M, finite or infinite degree D, VS

$$y\left(\underline{u}\right)=h_0+\sum_{n=1}^{D}\left\{\sum_{m_1=1}^{M}\cdots\sum_{m_n=1}^{M}h_n\left(m_1,\cdots,m_n\right)\prod_{j=1}^{n}u_{m_j}\right\} \tag{1}$$

where $\underline{u}\triangleq\left[u_1,\cdots,u_M\right]^T$ is the vector of lagged input samples. This series can be written as

$$y=f\left(\underline{u}\right)=\underline{h}^T\underline{\phi}(\underline{u}) \tag{2}$$

in which $\underline{h}$ is a coefficient vector comprising all Volterra coefficients in Eq.(1) and $\underline{\phi}(\underline{u})$ is a vector of all possible distinct polynomial combinations of the inputs, u_i with first element 1.

Construct a reproducing kernel Hilbert space (RKHS) $\mathcal{H}$, equipped with a reproducing kernel k, that contains every possible f,

$$k(\underline{u}_i,\underline{u}_j)=\sum_{l=1}^{L}\lambda_l\phi_l(\underline{u}_i)\phi_l(\underline{u}_j)=\langle\underline{\varphi}(\underline{u}_i),\underline{\varphi}(\underline{u}_j)\rangle_{l_2} \tag{3}$$

where L is typically large, possibly infinite, $\underline{\varphi}(\cdot)$ is a vector with lth element $\varphi_l(\cdot)=\sqrt{\lambda_l}\phi_l(\cdot)$ Define the corresponding inner product

$$\begin{aligned}\langle f,g\rangle_{\mathcal{H}}&=\langle(\underline{h}^f)^T\underline{\phi}^f,(\underline{h}^g)^T\underline{\phi}^g\rangle_{\mathcal{H}}\\&=\sum_{l=1}^{L}\frac{h^f_{l(n,m_1,\cdots,m_n)}h^g_{l(n,m_1,\cdots,m_n)}}{\lambda_l}\end{aligned} \tag{4}$$

in which $\underline{\lambda}=[\lambda_1,\cdots,\lambda_N]$ is a sequence of positive numbers, $\underline{h}^f,\underline{\phi}^f$ are the coefficient and polynomial vectors of the function f, respectively, and $\underline{h}^g,\underline{\phi}^g$ are defined in the same way (Dodd and Harrison, 2002a).

Every function f in $\mathcal{H}$, corresponding to a particular VS, can be expressed as

$$f(\cdot)=\sum_i\alpha_i k(\underline{u}_i,\cdot),\quad\alpha_i\in\mathbb{R} \tag{5}$$

Theoretically, to reveal f, the number of α_i must be the same as the number of independent Volterra kernels in f, giving no computational advantage. However, given a sample of size N, the least squares approximation $\hat{y}$ of f can be evaluated by

$$\hat{y}(\underline{u})=\sum_{i=1}^{N}\alpha_i k(\underline{u}_i,\underline{u}) \tag{6}$$

In the limit, as N approaches the number of independent Volterra kernels in f, $\hat{y}$ tends to f. This is not generally computationally feasible.

Although Eq.(6) is computationally feasible, i.e. a trade-off can be made between accuracy and efficiency of the computation, the evaluation of k involves the inner product of $\underline{\varphi}(\cdot)$, which is again a computational burden.

It can be shown that

$$k_p\left(\underline{u},\underline{v}\right)=\left(1+\langle\underline{u},\underline{v}\rangle\right)^D=\sum_{l=1}^{L}\lambda_l\phi_l(\underline{u})\phi_l(\underline{v}) \tag{7}$$

$$k_e\left(\underline{u},\underline{v}\right)=\exp\left(\frac{\langle\underline{u},\underline{v}\rangle_{l^2}}{p}\right)=\sum_{l=1}^{\infty}\lambda_l\phi_l(\underline{u})\phi_l(\underline{v}) \tag{8}$$

k_p and k_e are the polynomial and exponential kernels, respectively and D, p are positive constants.

By replacing k in Eq.(6) with k_p or k_e in the case of finite or infinite degree VS identification, the calculation can be significantly reduced (Dodd and Harrison, 2002a; Wan *et al.*, 2003).

Now let us consider a polynomial NARX model (Peyton-Jones and Billings, 1989),

$$y(t)=\sum_{n=1}^{D}y_n(t) \tag{9}$$

$$\begin{aligned}y_n(t)=&\sum_{n_y=0}^{n}\sum_{m_1,m_{n_y+n_u}=1}^{M}c_{n_y,n_u}(m_1,\cdots,m_{n_y+n_u})\\&\prod_{j=1}^{n_y}y(t-m_j)\prod_{j=n_y+1}^{n_y+n_u}u(t-m_j)\end{aligned} \tag{10}$$

with $n_y+n_u=n$ and $m_j=1,\cdots,M$, $j=1,\cdots,n$. When $n_y=0$ this describes a VS.

If we introduce a new "general" input vector $\underline{U}$,

$$\underline{U}=[U_1,\cdots,U_{M_y},U_{M_y+1},\cdots,U_{M_y+M_u}] \tag{11}$$

with $U_i=y(t-i),U_{M_y+i}=u(t-i)$, we get

$$y\left(\underline{U}(t)\right)=0+\sum_{n=1}^{D}\left\{\sum_{m_1=1}^{M}\cdots\sum_{m_n=1}^{M}c_n\left(m_1,\cdots,m_n\right)\prod_{j=1}^{n}U_{m_j}\right\} \tag{12}$$

which is of the same form as Eq.(1) with $h_0 = 0$. Eq.(12) can be representesd by the kernel method.

3. FRICTION DATA EXPERIMENTS

The data from three consecutive friction mechanisms are measured, including pre-sliding, true sliding and the intermediate regime. The detailed description of these physical processes is given in (Parlitz *et al.*, 2004). Friction behaviors may be divided into two main regimes: pre-sliding and true or gross sliding. Considering a general friction process, a displacement is always present between two physically contacted objects, given a force tending to cause their relative motion, unless the contact surface is infinitely stiff. Below some force threshold, as long as the force remains constant, the displacement will also remain constant, which is the pre-sliding regime. Once the force exceeds the threshold, the objects will have "obvious" relative motion, which is the true sliding process. In addition, there is an intermediate regime between these two processes, whose physical description is beyond our scope.

What is approximated here is the relationship between the displacement, as input, and the friction force, as output, in three friction regimes based on VS and NARX models, respectively, using the kernel method. The experimental model and the practical device, from which the data are generated, are shown in (Parlitz *et al.*, 2004). The experiment is performed for three levels of increasing excitation. In the lowest level of excitation, the rig is operating mostly in the presliding regime, slightly overlapping into the true sliding regime. The highest level of excitation exhibits predominantly true sliding behaviour, with the medium level of excitation showing reasonable degree of behaviour for both regimes. This is done to ensure that the models identified could work well for the varying levels of different operating regimes. The corresponding three measured time series for testing the different models are shown in Fig.1 [1] .

3.1 VS Identification

The identification of the three friction behaviors is presented based on both finite and infinite degree VS models. To avoid "large" kernel matrix inversion, whose size is $N \times N$, the original data

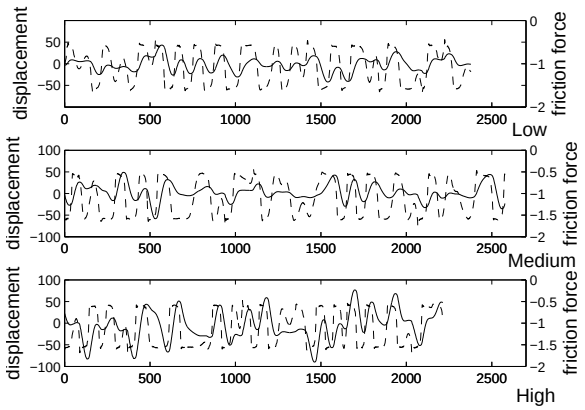

Fig. 1. Measured time series, from the low, medium and high excitation regimes, respectively – displacement (solid), force (broken)

are downsampled by a factor of five to be $N/5$. [2] Cross validation is used to find the optimal regularization parameter [3] ρ and p or D as applicable.

3.1.1. Low Level Excitation Identification The system identification based on a finite degree VS model using the polynomial kernel is discussed at first. There are 2400 sample data for training and testing, respectively. The VS model is of memory length 24 and degree 3, which gives the best multi-step ahead prediction performance in terms of normalized mean square error (NMSE) with $\mathrm{NMSE}(\hat{y}) = \frac{100}{N\sigma_y^2}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2\%$. Regularization is used to avoid numerical problems and to reduce the effects of noise, with the regularization parameter $\rho = 1 \times 10^{-7}$. We then use an infinite degree VS to identify the pre-sliding dynamics. Based on the NMSE, the parameters are set to be $M = 23, p = 9.9, \rho = 1 \times 10^{-7}$. The resulting estimation errors from finite and infinite degree VS models, respectively, are shown in Fig.2 (top).

3.1.2. Medium Level Excitation Identification The medium level excitation regime is first modeled by a finite degree VS. There are 2600 sample data for training and testing, respectively. The simulation parameters are chosen to be $M = 25, D = 3, \rho = 1 \times 10^{-7}$. This regime is then estimated by an infinite degree VS with the same training and testing sample data. The parameters

[1] The apparent anomaly whereby the output (force) amplitude seems hardly to vary even under approximately a doubling of input amplitude is a feature of the transition from pre-sliding to gross sliding regimes. In pre-sliding the effective stiffness is relatively high but reduces substantially when the surfaces are in motion so that the measured force does not change very much in amplitude.

[2] Because of the oversampling from the experimental rig, no significant information loss results from this downsampling.

[3] According to Eqs.(6), (7) and (8), $\underline{\alpha}$ can be solved thus $\underline{\alpha} = K^{-1}\underline{y}$, with $\underline{\alpha} = [\alpha_1, \cdots, \alpha_N]^T, \underline{y} = [y(\underline{u}_1), \cdots, y(\underline{u}_N)]^T, K_{ij} = k_p(\underline{u}_i, \underline{u}_j)$ or $k_e(\underline{u}_i, \underline{u}_j)$. But in "real" situations, where numerical and noise problems are present, $\underline{\alpha}$ is given by $\underline{\alpha} = (K + \rho I)^{-1}\underline{y}$ to provide a biased solution. ρ is called the regularization parameter.

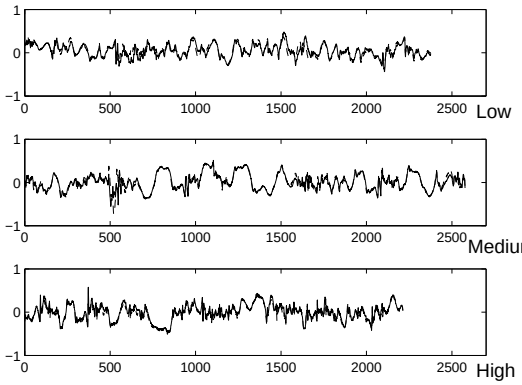

Fig. 2. Multi-step ahead prediction errors in the time domain by the finite (solid) and infinite (broken) degree VS models.

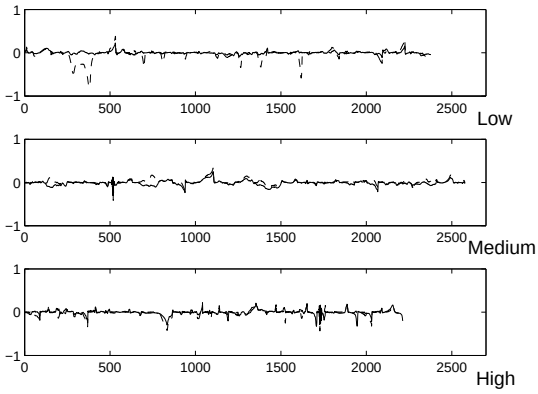

Fig. 3. Multi-step ahead prediction errors in the time domain by the finite (solid) and infinite (broken) degree NARX models.

are $M = 23, p = 9.9$ and $\rho = 1 \times 10^{-8}$. The corresponding estimation errors are shown in Fig.2 (middle).

3.1.3. High Level Excitation Identification Again, a finite degree VS model is used to identify the high level excitation dynamics. The number of both training and testing data is 2234. The parameters are $M = 23, D = 3$ and $\rho = 1 \times 10^{-8}$. An infinite degree VS model is then used to identify the same regime. The same sample data are used and the simulation parameters are $M = 22, p = 9.7$ and $\rho = 1 \times 10^{-8}$. The prediction errors are shown in Fig.2 (bottom).

In each case the infinite and finite degree cases are barely distinguishable.

3.2 NARX Identification

NARX models are used to identify the three friction regimes by using the same downsampled data as those in VS identification. One more parameter is involved to describe the NARX model, which is the number of AR terms, represented by N_{AR}. All the estimation error series are shown in Fig.3.

3.2.1. Low Level Excitation Identification A finite degree NARX model is identified to simulate the low level excitation regime. In terms of NMSE of testing data prediction, the parameters are chosen to be $M = 12, N_{AR} = 3, D = 4$ and $\rho = 1 \times 10^{-5}$. From Fig.3, clearly, the performance is much better than the comparable VS, which means NARX models can capture the dynamics more completely than the VS in this case. We then use an infinite degree NARX model to identify the same dynamics with $M = 11, N_{AR} = 3, p = 1.3, \rho = 1 \times 10^{-4}$.

3.2.2. Medium Level Excitation Identification To identify the medium level excitation process, a finite degree NARX model is computed, $M = 11, N_{AR} = 2, D = 3, \rho = 1 \times 10^{-6}$. An infinite degree NARX model is then identified to model the medium level excitation regime with $M = 11, N_{AR} = 2, p = 1.7$ and $\rho = 1 \times 10^{-6}$.

3.2.3. High Level Excitation Identification A finite degree NARX model is used to simulate the high level excitation dynamics with $M = 11, N_{AR} = 1, D = 3$ and $\rho = 1 \times 10^{-5}$. Then an infinite degree NARX model is computed to identify this same regime. The corresponding parameters are $M = 11, N_{AR} = 1, p = 1.5, \rho = 1 \times 10^{-5}$.

4. COMPARISON OF PREDICTION PERFORMANCES

From the figures in the last section, it is clear that the NARX models are much closer to the friction dynamics than VS. To evaluate the prediction performance further, identification, based on the same downsampled training/testing sample by the DNLR Maxwell slip and the AVD NN models, are demonstrated.

4.1 The DNLR Maxwell Slip Model

The generalized Maxwell slip model is a well-known physics-based model for simulating hysteretic behaviors that is a significant feature in friction systems. By connecting Maxwell slip models in series with a Multiple Input Single Output-Finite Impulse Response filter, the DNLR Maxwell slip model, used here, can be formed.

Based on the same training and testing sample data, the prediction errors on the testing data are shown in Fig.4.

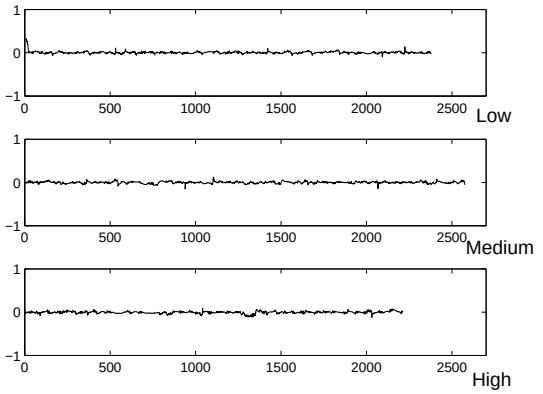

Fig. 4. Multi-step ahead prediction errors in the time domain by the DNLR models.

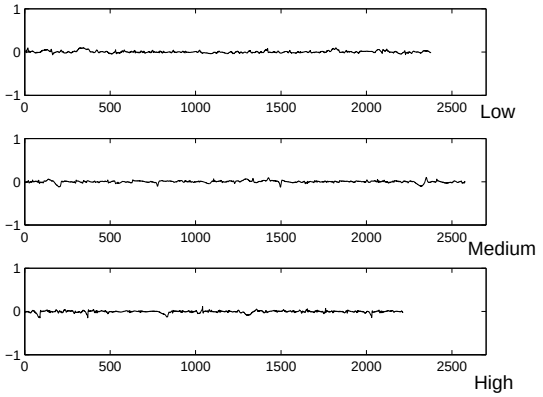

Fig. 5. Multi-step ahead prediction errors in the time domain by the AVD NNs.

4.2 The AVD NN model

It has been argued (Chen and Tomlinson, 1996) that most dynamical engineering systems can be explained by second order differential equations. Therefore, it is worthwhile to "capture" temporal information of systems directly via their 0th, 1st and 2nd derivatives, which are displacement, velocity and acceleration, respectively, in the case of the friction system. To do that, an infinite impulse response (IIR) differentiator/integrator is used (Chen and Tomlinson, 1996).

The prediction errors are shown in Fig.5.

4.3 Comparisons

Visually comparing all the experimental figures in Section 3 with the corresponding ones in this section, except for the prediction results from VS, the others are not far from each other. In terms of the NMSEs, the performances of the physics-based DNLR Maxwell slip models for all three friction behaviors are the best. The AVD NN models' performances are slightly better than RKHS-based finite degree NARX models.

The NMSEs for all the modeling processes are summarized in Table (1).

Table 1. NMSEs for the three friction regimes for all models(%).

	Low	Medium	High
Finite VS	14.09	18.71	18.25
Infinite VS	14.13	20.67	18.40
Finite NARX	0.68	1.49	1.76
Infinite NARX	9.41	1.70	2.95
DNLR Slip	0.41	0.28	0.37
AVD NNs	0.41	0.38	0.66

From these results, some conclusions can be drawn:

(1) From Table (1), clearly, the VS models are relatively much worse than the others in modeling the friction dynamics. This is not surprising because the VS' basic fading memory limitation is violated by these friction processes' hysteric properties.

(2) What we can see from the experimental results in Section 3 is that, for both VS and NARX models, the finite degree models work better than the infinite degree models. Additionally, the degrees of the finite models are not high (3–4). All these suggest that the infinite degree models are over-fitting on this problem, despite the "optimal" choice of ρ.

(3) The difference in the performances of the finite degree NARX, DNLR Maxwell slip and AVD NN models is small, which may come from the models' dynamics, the algorithms and the normalizations [4].

(4) In terms of NMSE, RKHS-based NARX models perform slightly worse than the physically derived models and the AVD NNs. But the kernel-based algorithm makes the evaluation of Volterra and NARX coefficients natural. Reasonably high order coefficients can be extracted efficiently with the practical situation depending on M and N_{AR} (if applicable) [5]. The restriction mainly comes from the inversion of large kernel matrices of dimension $N \times N$. This can be resolved using an iterative algorithm.

(5) The physics-based DNLR Maxwell model is very suited to the identification of these particular systems, which can not only provide accurate multi-step ahead prediction but also reveal the physical nature of the nonlinear dynamics. Its disadvantage is that it is very difficult to extract even low order Volterra

[4] All the sample data are normalized to avoid numerical problems.

[5] According to the authors' experiments, e.g. if the VS evaluated is of infinite degree and $M = 12$ (or $M + N_{AR} = 12$ for a NARX model evaluation), it takes about 2 minutes to compute all the fourth order Volterra kernels using a *Pentium*®4 1.8GHz computer.

coefficients (hence generalized frequency response functions). In addition, it cannot be used as a general model for the control of friction systems with other forms of unknown nonlinearity, e.g. backlash.

(6) The AVD NNs also provide good prediction performances in this case. However, they are beset by the problem of non-convex optimization. It is also hard to extract physical information from them. To acquire low order Volterra kernels is possible, but to acquire high order ones is difficult.

5. CONCLUSION

Friction dynamics are important and complex nonlinear systems. Generally speaking, a physics-based model is the first choice for nonlinear system identification provided it can be revealed. But for many complex nonlinear systems, their physically derived models are extremely difficult to identify. Also, some physical models are difficult to analyze.

VS and NARX models are attractive alternatives to physical models since they contain systems information of interest and, to compute them, one does not need to know much about target systems. In the experiments demonstrated in this paper, VS perform poorly because their basic fading memory requirement is violated by the hysteric properties of typical friction processes.

A common problem with the evaluation of VS and NARX models is the determination of large numbers of parameters. The kernel method addresses this problem by representing VS and NARX models in a compact kernel form and provides a feasible approximation of the models with much less computation.

Note that, in this particular friction identification problem, albeit that finite NARX models perform slightly worse than DNLR Maxwell slip models and AVD NNs, they are still promising, considering they are pure black-box, compared with white-box DNLR Maxwell slip models and grey-box AVD NNs. After all, systems' physical nature, which is the prerequisite for white and grey-box models identification, is not always available.

Whether VS or NARX models should be used in a particular problem depends on the practical situation. NARX models account for a wider range of nonlinear characteristics and involve fewer coefficients to evaluate. But, because of the involvement of previous outputs, prediction errors are always accumulated and fedback to NARX models through inputs. This makes them susceptible to bias and, in addition, they can be unstable. In contrast, VS provide a flexible and stable nonlinear model which is linear in its parameters and which has shown its power in many other applications.

ACKNOWLEDGEMENT

YW would like to thank The University of Sheffield for its financial support. CXW acknowledges the Volkswagenstiftung for financial support under grant no. 1/76938. The authors are also grateful to Dr. F. al-Bender and Mr. T. Tjahjowidodo of Katholieke Universiteit, Leuven for providing details of the rig and the experimental data used in the modeling.

REFERENCES

Chen, Q. and G. R. Tomlinson (1996). Parametric identification of systems with dry friction and nonlinear stiffness using a time series model. *Transactions of the ASME - Journal of Vibration and Acoustic* **118**, 252–263.

Dodd, T. J. and R. F. Harrison (2002a). A new solution to Volterra series estimation. In: *CD-ROM Proceedings of the 2002 IFAC World Congress*. Barcelona.

Dodd, T. J. and R. F. Harrison (2002b). Estimating Volterra filters in Hilbert space. In: *Proceedings of IFAC International Conference on Intelligent Control Systems and Signal Processing*. University of Algarve, Faro. pp. 538–543.

Drezet, P. M. L. (2001). Kernel Methods and Their Application to Systems Identification and Signal Processing. Phd thesis. The University of Sheffield. Sheffield.

Harrison, R. F. (1999). Computable Volterra filters of arbitrary degree. MAE Technical Report 3060. Princeton University. Princeton.

Parlitz, U., A. Hornstein, D. Engster, F. Al-Bender, V. Lampaert, T. Tjahjowidodo, S. D. Fassois, D. Rizos, C. X. Wong, K. Worden and G. Manson (2004). Identification of presliding friction dynamics. *Chaos* **14**, 420–430.

Peyton-Jones, J. C. and S. A. Billings (1989). Recursive algorithm for computing the frequency response of a class of non-linear difference equation models. *International Journal of Control* **50**, 1925–1940.

Rizos, D. D. and S. D. Fassois (2004). Presliding friction identification based upon the maxwell slip model structure. *Chaos* **14**, 431–445.

Wan, Y., T. J. Dodd and R. F. Harrison (2003). Infinite degree Volterra series estimation. In: *CD-ROM Proceedings of the Second International Conference on Computational Intelligence, Robotics and Autonomous Systems*. Singapore.