WIENER-HAMMERSTEIN PARAMETER ESTIMATION USING DIFFERENTIAL EVOLUTION
Application to Limb Reflex Dynamics

Oliver P. Dewhirst, David M. Simpson, Natalia Angarita, Robert Allen
Institute of Sound and Vibration Research, University of Southampton, Southampton, U.K.

Philip L. Newland
School of Biological Sciences, University of Southampton, Southampton, U.K.

Keywords: Wiener-Hammerstein, Differential evolution, Locust, Limb reflex dynamics, Neuromuscular modelling.

Abstract: The nonlinear Wiener-Hammerstein model, which consists of a static (no memory) non-linearity sandwiched between two dynamic (with memory) linear elements, provides a parsimonious and accurate model for representing a number of biological systems. In this study we compare the performance of two Wiener-Hammerstein parameter estimation methods; a commonly used nonlinear local optimisation method and the global optimisation method Differential Evolution. The accuracy and convergence properties of the two methods is tested using experimental data collected from the locusts hind limb reflex control system and using computer simulations.

1 INTRODUCTION

A greater understanding of neuromuscular control of limb movement is vital for optimising the treatment of patients with neuromuscular dysfunction. System identification methods provide a valuable tool for investigating the complex mechanical and neural components of such systems (Bai et al., 2009; Dempsey and Westwick, 2004). Our work uses the locust as a model system because it provides the opportunity to develop new system identification techniques and gain physiological insight into a related but simpler and more accessible system.

Nonlinear Wiener or Volterra series models, a generalisation of the convolution integral for a linear system, have often been used to model the nonlinear responses generated by reflex limb control systems (Newland and Kondoh, 1997) and other biological systems (Song et al., 2009). These series are often truncated to second order because it is difficult to visualise, interpret and calculate, in terms of computational cost and length of data required, the higher order kernels.

Cascade models, a restricted subset of the Volterra series, have been shown to provide a parsimonious and easier to interpret alternative for representing such systems (Dewhirst et al., 2009; Dempsey and Westwick, 2004). Cascade models consist of linear dynamic (with memory) elements and static (no memory) nonlinear elements. Common configurations include the Wiener (linear - nonlinear), Hammerstein (nonlinear - linear) and the Wiener-Hammerstein (linear - nonlinear - linear) model shown in Figure 1. Methods to estimate the parameters of the Wiener-Hammerstein model are currently receiving much attention for control applications (Schoukens et al., 2009) but not for application to biological systems.

A key feature of the locust's hind leg control system is a reflex control loop. This system uses a mechanical loop structure attached to the tibia by an apodeme to move a stretch sensor called the Femoro-tibial Chordotonal Organ (FeCO). Sensory neurons in the FeCO convert mechanical stimuli into electrical signals which are transmitted to motor neurons which activate muscle contraction (Newland and Kondoh, 1997). Previous work (Dewhirst et al., 2009) found that the Wiener-Hammerstein model was the optimum structure for representing the dynamic reflex response of the Fast Extensor Tibia (FETi) motor neuron in the locusts hind leg control system, the system studied in the current work.

As the cascade models are nonlinear in their pa-

P. Dewhirst O., M. Simpson D., Angarita N., Allen R. and L. Newland P. (2010).
WIENER-HAMMERSTEIN PARAMETER ESTIMATION USING DIFFERENTIAL EVOLUTION - Application to Limb Reflex Dynamics.
In *Proceedings of the Third International Conference on Bio-inspired Systems and Signal Processing*, pages 271-276
DOI: 10.5220/0002741902710276

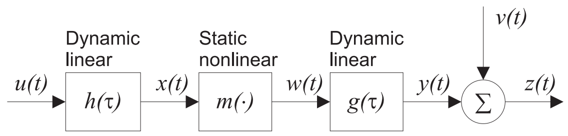

Figure 1: The Wiener-Hammerstein model.

rameters and have a differentiable cost function their parameters can be estimated using a nonlinear local optimisation method. One of the most widely used local optimisation methods for estimating the parameters of the Wiener-Hammerstein models for biological applications was developed by Korenberg and Hunter (Korenberg and Hunter, 1986) (the KH method). As the cost function of the model is multi-modal, such methods require careful parameter initialisation to ensure that they do not converge to a local rather than the global minimum (Schoukens et al., 2008; Dempsey and Westwick, 2004).

An alternative approach is to use a nonlinear global optimisation method such as simulated annealing which has been developed to avoid the problem of local minima. As these methods can suffer from slow convergence and poor local search characteristics their application has often been restricted to finding initial parameter estimates for the local methods (Juusola et al., 2003). One type of evolutionary algorithm, Differential Evolution (DE), however, has been shown to give fast robust convergence and good local search performance. It also requires few control parameters and is easy to use and implement (Storn and Price, 1997). DE has been applied to estimate the parameters of Hammerstein and Wiener models (Al-Duwaish, 2000) but we do not believe to the Wiener-Hammerstein model.

In this study we compare the performance of the DE and KH methods for estimating the parameters of the Wiener-Hammerstein model from experimental data obtained from the locusts leg control system and computer simulations.

2 METHODS

The linear elements of the Wiener-Hammerstein model (Figure 1) have a Finite Impulse Response (FIR) structure. A polynomial function was chosen to represent the static nonlinear element. The noise free output of the Wiener-Hammerstein model is given by

$$y(t) = \sum_{\sigma=1}^{T} g(\sigma) \sum_{q=0}^{Q} c^{(q)} \left(\sum_{\tau=1}^{T} h(\tau) u(t-\sigma-\tau) \right)^{q} \quad (1)$$

Where T is the length of the FIR filter and Q is the polynomial order.

2.1 Differential Evolution

The parameters of the Wiener-Hammerstein model are coded in a vector containing D elements

$$\mathbf{x} = \left(h(1), \ldots, h(T),\ c^{0}, \ldots, c^{Q},\ g(1), \ldots, g(T) \right) \quad (2)$$

Differential Evolution (DE) works with a population, P, of NP D-dimensional parameter vectors (Storn and Price, 1997)

$$P^{(G)} = \begin{bmatrix} x_{1,1}^{(G)} & x_{1,2}^{(G)} & \cdots & x_{1,D}^{(G)} \\ x_{2,1}^{(G)} & x_{2,2}^{(G)} & \cdots & x_{2,D}^{(G)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{NP,1}^{(G)} & x_{NP,2}^{(G)} & \cdots & x_{NP,D}^{(G)} \end{bmatrix} \quad (3)$$

where $x_{1,1}^{(G)} x_{1,2}^{(G)} \cdots x_{1,D}^{(G)}$ are elements of $\mathbf{x}$. Each parameter vector

$$\mathbf{x}_i^{(G)} = \left(x_{i,1}^{(G)}, x_{i,2}^{(G)}, \ldots, x_{i,D}^{(G)} \right) \quad i = 1, 2, \ldots, NP \quad (4)$$

represents a possible solution to the optimisation problem. The population is initialised randomly and evolves in generations (iterations), G. With each generation a mutant vector, v_i, is formed from each of the NP parent vectors, x_i, following

$$\mathbf{v}_i^{(G+1)} = \mathbf{x}_{r_1}^{(G)} + F \left(\mathbf{x}_{r_2}^{(G)} - \mathbf{x}_{r_3}^{(G)} \right) \quad (5)$$

with the multiplication factor F set by the user ($F > 0$) and random, mutually different indexes $r_1, r_2, r_3 \in \{1, 2, \ldots, NP\}$ which are also different from i. The trial vector

$$\mathbf{u}_i^{(G+1)} = \left(u_{i,1}^{(G+1)}, u_{i,2}^{(G+1)}, \ldots, u_{i,D}^{(G+1)} \right) \quad (6)$$

is generated using the crossover operator

$$u_{i,j}^{(G+1)} = \begin{cases} v_{i,j}^{(G+1)} & \text{if } (rnd(j) \leq CR) \\ x_{i,j}^{(G)} & \text{if } (rnd(j) > CR) \end{cases} \quad (7)$$

where $j = 1, 2, \ldots, D$ and $rnd(j)$ is the j^{th} output from a random number generator with a uniform distribution. The cross over constant $CR \in [0, 1]$ is set by the user; $rn(k) \in \{1, 2, \ldots, D\}$ is a randomly chosen index used to ensure that at least one of the parameters from $v_{i,j}$ is transferred to $u_{i,j}$. The widely used Minimum Mean Square Error (MMSE) cost function is used to compare the performance of the trial and parent vectors.

$$\mathrm{J} = \frac{1}{N} \sum_{t=0}^{N-1} \left(z(t) - \hat{y}(t) \right)^2 \quad (8)$$

where N is the signal length in samples. A greedy selection procedure (does not reconsider past generations) is used to determine if the parent vector should be retained for the next generation, or replaced by the trial vector. These steps are repeated either for a specific number of generations or until a certain cost function value is reached. The DE method was set up with $F = 0.5$, $CR = 0.9$, $NP = 10 \times D$ (Storn and Price, 1997) and was run for 900 iterations.

2.2 The Korenberg Hunter Method

Korenberg and Hunter (Korenberg and Hunter, 1986) use an iterative two step approach to estimate the parameters of the Wiener-Hammerstein model. Their method is summarised in Algorithms 1 and 2.

Algorithm 1. KH parameter estimation.

Require: Initial estimate of $\hat{h}(\tau) = (\frac{1}{fs}, 0, 0, \ldots, 0)$

1: Filter $u(t)$ by $\hat{h}(\tau)$ to obtain $\hat{x}(t)$
2: Fit the Hammerstein system between $\hat{x}(t)$ and $z(t)$ using Algorithm 2
3: If first iteration or significant improvement in model accuracy continue, else exit with current parameters
4: Obtain a new estimate of $\hat{h}(\tau)$ using the Levenberg-Marquardt (LM) (Marquardt, 1963) gradient based method to minimise J (Equation 8) given the current estimate of the Hammerstein system. Return to step 2

Algorithm 2. Hammerstein parameter estimation.

1: Fit a linear system between $z(t)$ and $\hat{x}(t)$ to obtain an initial estimate of the inverse of $\hat{g}(\tau)$, $\hat{g}(\tau)^{-1}$
2: Filter $z(t)$ by $\hat{g}(\tau)^{-1}$ to obtain $\hat{w}(t)$
3: Fit a polynomial $\hat{m}(\cdot)$ between $\hat{x}(t)$ and $\hat{w}(t)$
4: Obtain a new estimate of $\hat{w}(t)$ given $\hat{m}(\hat{x}(t))$
5: Re-estimate $\hat{g}(\tau)$ given $\hat{w}(t)$ and $z(t)$
6: Calculate the %MSE (Equation 9) difference between model output $\hat{y}(t)$ and measured output $z(t)$
7: If the %MSE reduction is small compared to last iteration, exit with current parameters
8: Else, compute a new estimate of $\hat{g}(\tau)^{-1}$ using the current $\hat{w}(t)$ and $z(t)$. Return to step 2

2.3 Experimental Data

The performance of the two parameter estimation methods was measured using experimental data collected from the locust (*Schistocerca gregaria*). A brief summary of the experimental methods used by Newland and Kondoh (Newland and Kondoh, 1997) follows. The locusts were mounted in modelling clay; the apodeme of the FeCO was exposed by dissection and was attached by forceps to an electromechanical shaker (LDS, type 101). The FeCO was stimulated by applying a 27Hz low pass filtered Gaussian White Noise (GWN) signal to the shaker to move the apodeme (Figure 2 A). The synaptic inputs to

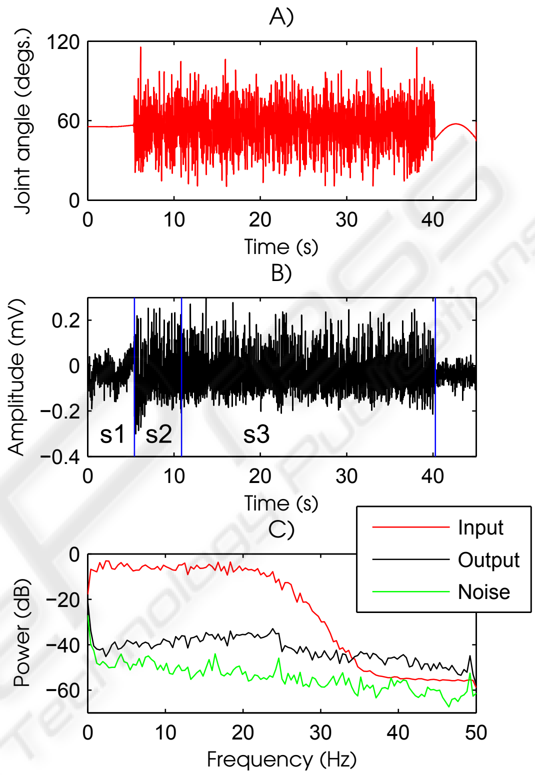

Figure 2: The band limited (0-27Hz) GWN input signal applied to move the apodeme of the FeCO (A). A typical output signal recorded from FETi (B). Noise, the transient adapting response of FETi and its steady state adapted response are marked as s1, s2 and s3 respectively. The spectrum of the input signal, adapted FETi response (s3) and noise signal (s1) are plotted in C.

the FETi motor neuron, caused by stimulation of the apodeme of the FeCO, were measured by inserting a glass microelectrode into its soma. A typical recording from the soma of the FETi motor neuron is shown in Figure 2 B. The high level of measurement and background neural noise, initial transient response and steady state adapted response of FETi can be seen in Figure 2 B sections s1, s2 and s3 respectively. Models were fitted between the first 600 samples (6s with fs=100Hz) of the steady state adapted response of FETi (Figure 2 B s3) and the corresponding samples of input signal. The performance of the parameter estimation methods was determined by calculating the %MSE difference between the predicted output of the model and the measured output of the neuron using 200 samples of validation data.

$$MSE = 100 \times \frac{\frac{1}{N}\sum_{t=1}^{N}(y(t)-\hat{y}(t))^2}{\frac{1}{N}\sum_{t=1}^{N}y(t)^2} \qquad (9)$$

Before the performance of the two parameter estimation methods was compared, the number of model parameters was optimised. As six sets of data were available, six models were estimated using the KH method. The %MSE performance of each of these models with a varying number of parameters was calculated, using validation data. The optimum number of parameters was determined by calculating the mean of the minimum %MSE value. As the gain of the locust system can be arbitrarily assigned to the linear and nonlinear elements of the model these elements have been normalised to allow comparison.

2.4 Simulations

Computer simulations were used to investigate the robustness of the two parameter estimation methods to experimental conditions. The simulated system was based on one of the Wiener Hammerstein models identified from experimental data using the KH method. Measurement noise coupled with the limited bandwidth of the input signal (Figure 2 C) introduces random estimation errors into the linear elements of the model. As the estimation error occurs where there is little input power ($>$ 27Hz) and system dynamics it was removed using a low pass filter (zero phase, 3^{rd} order Butterworth, $f_c = 27$Hz).

"Ideal" and "low pass with measurement noise" data sets were generated using the simulated system. To create the "ideal" data set, a GWN signal was generated and applied to the simulated system. The input signal in the "low pass with measurement noise" data set represents the input signal used in the locust experiments. It was created by filtering the GWN input signal with a low pass filter (zero phase, 5^{th} order Butterworth, $f_c = 27$Hz). The effect of measurement noise on the parameter estimation methods was investigated using a Monte Carlo simulation (100 trials). In each trial measurement noise with a Gaussian distribution was added to the output from the simulated system generated using the same low pass input signal. The noise was added to give a Signal to Noise Ratio of 5dB, a level similar to that found in the locust recordings.

3 RESULTS AND DISCUSSION

The variation in mean %MSE performance of the Wiener-Hammerstein model (parameters estimated using the KH method) with the number of parameters it contains is plotted in Figure 3. This result shows that optimum performance is obtained with 9 parameters in each linear element and a 5^{th} order polynomial, though the minimum is in an almost flat area of the parameter space. The mean %MSE performance of the models, containing the optimum number of parameters, with their parameters estimated using the KH and DE methods is very similar (30.0% and 29.0% respectively, Table 1). Both methods also produce similar mean parameter estimates (Figure 4).

Table 1: %MSE performance of the Wiener-Hammerstein models, with their parameters estimated using the KH and DE methods, on validation data. Models contain the optimum number of parameters.

Animal	KH method	DE method
1	31.0	29.2
2	29.4	32.4
3	17.6	19.6
4	36.0	33.3
5	32.1	28.0
6	34.1	31.3
Mean	30.0	29.0

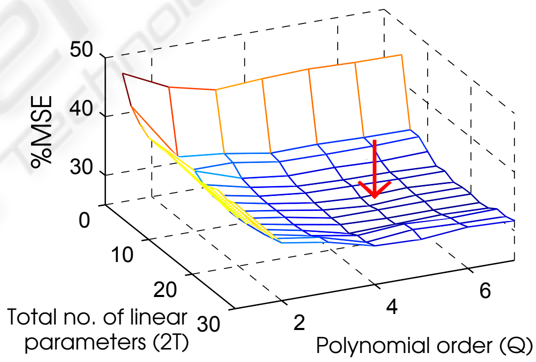

Figure 3: Mean (n=6) %MSE performance variation for validation data with the total number of linear parameters in the model and polynomial order. The minimum mean MSE is 30.0% and occurs with a total of 18 linear parameters and a 5^{th} order polynomial (marked with an arrow).

The non white spectrum of the input signal (Figure 2 C) coupled with the high levels of measurement noise (Figure 2 B s1) results in the parameter estimates containing high frequency estimation error (Figure 4). The rather high %MSE performance of the models, suggests that they provide a poor fit to the system. The recordings prior to the start of stimulation, when the input is constant, show considerable measurement noise (Figure 2 B s1). Analysis has shown that the spectrum of the residual signal (the difference between the model and the measured out-

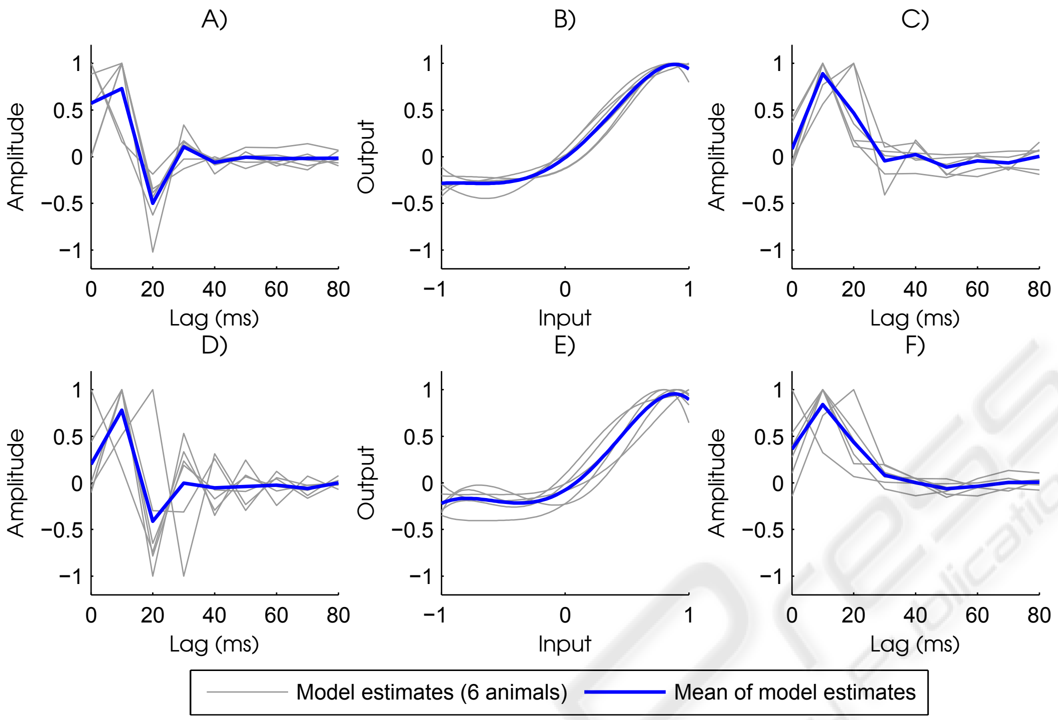

Figure 4: The parameters of the Wiener-Hammerstein models of the FETi system estimated using the KH method (A-C) and the DE method (D-F). The first linear element $h(\tau)$ is plotted in A and D, the second linear element $g(\tau)$ in C and F. The polynomial function is plotted in B and E. Note that the parameters have been normalised to allow comparison.

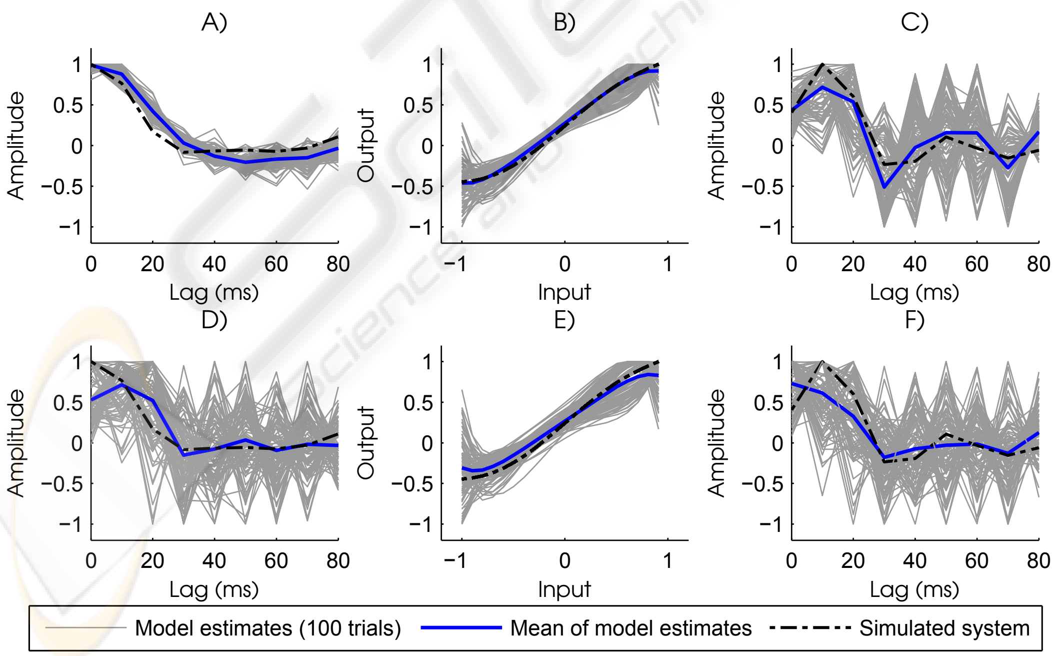

Figure 5: Results of Monte Carlo simulations comparing the performance of the KH (A to C) and the DE (D to F) parameter estimation methods using the "low pass with measurement noise" data set. In each trial measurement noise with a Gaussian distribution was added to the output of the simulated system generated using the same low pass input signal.

put signal) is very similar to the spectrum of the noise signal. This suggests that model fit is good, as one cannot expect the model to be able to predict the random measurement noise.

Computer simulations using the "ideal" data set showed that both parameter estimation methods could accurately identify the simulated system. The normalised parameter bias error for both methods was less than 0.1%. The results of the Monte Carlo simulation using the "low pass with measurement noise" data set are shown in Figure 5. Under these simulated experimental conditions, neither method could retrieve the parameters of the simulated system. It is interesting to note that the DE estimates of the parameters of $h(\tau)$ (Figures 4 and 5 D) contain more high frequency estimation error than the KH estimates (Figures 4 and 5 A) for both experimental and simulated data. This is probably because the LM method (Westwick and Kearney, 2003) used to estimate $h(\tau)$ in the KH algorithm (Algorithm 1 step 4) approximates the Hessian $\hat{\mathbf{H}}$ as

$$\hat{\mathbf{H}} = \frac{1}{N}\mathbf{J}^T\mathbf{J} + \mu_k \mathbf{I_M} \tag{10}$$

and the $\mu_k \mathbf{I_M}$ term improves the condition of $\hat{\mathbf{H}}$.

4 CONCLUSIONS

The accuracy of two Wiener-Hammerstein parameter estimation methods has been compared using experimental data and computer simulations. Our work has shown that DE gives fast robust convergence and good local search performance, providing similar accuracy as the local optimisation method developed by Korenberg and Hunter (Korenberg and Hunter, 1986). The DE method offers the advantage of simplicity and flexibility, for example making it easy to use different cost functions. Also, as it is a global method, it should be less sensitive to parameter initialisation, but additional work is require to verify this feature. Further investigation into the convergence properties of both methods and ways to reduce parameter bias and variance under experimental conditions is required.

ACKNOWLEDGEMENTS

The authors would like to thank the Gerald Kerkut Trust, the LSI Forum, the ISVR Rayleigh Scholarship and the BBSRC for their support.

REFERENCES

Al-Duwaish, H. (2000). A genetic approach to the identification of linear dynamical systems with static nonlinearities. *International Journal of Systems Science*, 31(3):307–313.

Bai, E., Cai, Z., Dudley-Javorosk, S., and Shields, R. K. (2009). Identification of a modified wiener-hammerstein system and its application in electrically stimulated paralyzed skeletal muscle modeling. *Automatica*, 45(3):736 – 743.

Dempsey, E. and Westwick, D. (2004). Identification of hammerstein models with cubic spline nonlinearities. *IEEE Transactions on Biomedical Engineering*, 51(2):237–245.

Dewhirst, O., Simpson, D., Allen, R., and Newland, P. (2009). Neuromuscular reflex control of limb movement - validating models of the locust's hind leg control system using physiological input signals. In *IEEE EMBS 09, 4th International Conference on Neural Engineering*. IEEE EMBS.

Juusola, M., Niven, J., and French, A. (2003). Shaker k+ channels contribute early nonlinear amplification to the light response in drosophila photoreceptors. *J Neurophysiol*, 90(3):2014–2021.

Korenberg, M. and Hunter, I. (1986). The identification of nonlinear biological systems: lnl cascade models. *Biological Cybernetics*, 50(2):125–134.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, 11(2):431–441.

Newland, P. and Kondoh, Y. (1997). Dynamics of neurons controlling movements of a locust hind leg iii: Extensor tibiae motor neurons. *The Journal of Neurophysiology*, 77(5):3297–3310.

Schoukens, J., Pintelon, R., and Enqvist, M. (2008). Study of the lti relations between the outputs of two coupled wiener systems and its application to the generation of initial estimates for wiener-hammerstein systems. *Automatica*, 44(7):1654 – 1665.

Schoukens, J., Suykens, J., and Ljung, L. (2009). Wiener hammerstein benchmark. In *Proc. 15 IFAC Symposium on System Identification*, St Malo, France.

Song, D., Marmarelis, V., and Berger, T. (2009). Parametric and non-parametric modeling of short-term synaptic plasticity. part i: Computational study. *J Comput Neurosci*, 26(1):1 – 19.

Storn, R. and Price, K. (1997). Differential evolution a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11:341–359.

Westwick, D. and Kearney, R. (2003). *Identification of Nonlinear Physiological Systems*. Wiley-IEEE Computer Society Pr, Piscataway, New Jersey, 1st edition.