

DIVIDE & CONQUER IDENTIFICATION USING GAUSSIAN PROCESS PRIORS

D.J. Leith^{1,2}, W.E. Leithead^{1,2}, E. Solak², R. Murray-Smith^{1,3}

¹Hamilton Institute, National University of Ireland, Maynooth, Co. Kildare, Ireland

²Dept. of Electronics & Electrical Engineering, University of Strathclyde, Glasgow, U.K.

³Dept. of Computing Science, University of Glasgow, Glasgow G12 8QQ, U.K.

Abstract

In this paper, we investigate the reconstruction of nonlinear systems from locally identified linear models. It is well known that the equilibrium linearisations of a system do not uniquely specify the global dynamics. Information about the dynamics near to equilibrium provided by the equilibrium linearisations is therefore combined with other information about the dynamics away from equilibrium provided by suitable measured data. That is, a hybrid local/global modelling approach is considered. A non-parametric Gaussian process prior approach is proposed for combining in a consistent manner these two distinct types of data. This approach seems to provide a framework that is both elegant and powerful, and which is potentially in good accord with engineering practice.

1. INTRODUCTION

Nonlinear system identification methods have been extensively studied over the last ten years (e.g. there is a substantial literature on neural/fuzzy approaches and NARMAX methods). The developed methods certainly perform well in appropriate situations. However, they frequently prove less satisfactory in complex engineering applications. With regard to current system identification *practice* it is probably fair to say that, except in specific circumstances where collection of a single data-set covering the entire operating envelope is feasible, the approach almost universally employed is to collect data from a number of identification experiments each of which excites a system in only a relatively small part of the operating envelope at any one time.¹ Factors such as safety

¹ For the sake of clarity, it is worth noting that although blended multiple model representations (e.g. local model networks, Takagi-Sugeno fuzzy models, see the review by Johansen & Murray-Smith, 1997) often appear to be motivated by the need for a divide and conquer approach, identification of such models (even with so-called "local" learning methods) is always carried out in an essentially "global" manner. That is, the nonlinear system is *not* reconstructed from a family of *individually* identified systems. This is not simply because divide and conquer identification methods have not been considered in the context of blended multiple model systems. Rather, recent theoretical results confirm empirical experience that "global"

requirements, operating constraints and minimising disruption to normal operation/production all typically require the adoption of such a practice which leads naturally to a divide and conquer type of identification campaign. Other important advantages of such an approach include (i) relatively well-developed system identification methods for linear systems can be used when performing the local identifications near to equilibrium and (ii) a degree of transparency is maintained which often plays a vital role in achieving successful results when dealing with complex systems. However, each linear model is only valid in the vicinity of a particular equilibrium point and the family of linear models only describes the dynamics near to equilibrium operation. Clearly this can imply a very substantial restriction on the utility of such models in the context of strongly nonlinear systems. Moreover, inference of the characteristics of the underlying nonlinear dynamics from the characteristics of the identified linearisations is typically carried out in a somewhat *ad hoc* manner at present. We wish to investigate the possibility of developing a sound theoretical framework within which to consider the reconstruction of nonlinear systems from locally identified linear models.

In the present paper, attention is focussed on methods which utilise classical equilibrium linearisation models. (Note, linearisation frameworks

identification approaches are necessary in view of a number of characteristics fundamental to the blended multiple model representations typically employed (Shorten *et al.* 1999, Leith & Leithead 1999).

which are more general than this do exist, see for example the VB approach of Leith & Leithhead 1998). It is, of course, well known that the equilibrium linearisations of a system do not uniquely specify the global dynamics. We are therefore interested in combining the information about the dynamics near to equilibrium provided by the equilibrium linearisations with other information about the dynamics away from equilibrium provided by suitably measured data. That is, a hybrid local/global modelling approach is considered. Note that we would like to use the smallest quantity of off-equilibrium data consistent with resolving the ambiguity present in the linearisation data.

Such a hybrid approach necessarily entails combining in a consistent manner two distinct types of data: estimates of the equilibrium linearisations (derivative/tangent map data) and input-output trajectory data measured during transitions between equilibria and operation far from equilibrium. In the special case where only input-output trajectory data is involved, the modelling problem essentially reduces to the usual global nonlinear identification situation studied by many researchers. In the special case where only tangent map data is involved, the modelling task is to simultaneously interpolate/smooth the tangent information while integrating to recover the original nonlinear map. Note that when working in dimensions higher than two, there exist functions which are not integrable. Consequently, it is necessary to impose (either implicitly or explicitly) an integrability constraint on the interpolation/smoothing procedure. This is, of course, a highly non-trivial requirement. Both of these special cases are subsumed within the more general framework considered here.

Full state information is not assumed in this paper and consequently the equilibrium linearisations must be estimated from input-output data alone. This is not unusual: the transfer function of the linearisation is estimated or, equivalently, the state space matrices are estimated up to a non-singular linear state transformation. However, when combining the linearisations at a number of equilibria it is necessary to choose compatible state co-ordinates. This issue is discussed in section 2. In section 3, the consistent combination of derivative and function data is considered and the results illustrated by a simple example. The conclusions are summarised in section 4.

2. EMBEDDING THEORY

Consider a nonlinear system

$$x_{i+1} = f(x_i, r_i), \quad y_i = g(x_i) \quad (1)$$

with state $x \in \mathcal{X}^m$, input $r \in \mathcal{R}^n$ and output $y \in \mathcal{Y}^d$. It is assumed that the input and output are measured but that the state is not. The latter is of course a fundamental issue in empirical modelling. Essentially, we are trying to fit the mapping that lies on the right-hand side of a difference equation without being able to directly measure all of the quantities on which the output of the map depends. Indeed, the subset of the dependent variables associated with the state is not only unmeasurable but also its relationship to the measured input and output is generally completely unknown and must be inferred. Fortunately, Takens embedding theorem, and generalisations thereof, provide a basis for

analysing the time series data generated by a nonlinear dynamic system (Stark *et al.* 1997, 2001, for early work in a control context see also Narendra *et al.* 1996). This family of theorems allows us to reconstruct the unknown dynamics from an observed time series by constructing a new state-space out of successive values of the time series.

Theorem (Takens' theorem for forced systems, Stark *et al.* 1997, 2001). Consider the system (1). Let $x \in M$ and $r \in N$ with M and N compact manifolds of dimension $m \geq 1$ and n respectively. Let the map $f: M \times N \rightarrow M$ be such that $f_y: M \rightarrow M$ is a C^k diffeomorphism of M for any y , where $f_y(x) = f(x, y)$, and let $g: M \rightarrow \mathcal{R}^d$ be C^k with $k \geq 1$. Suppose that $d \geq 2m+1$. Then the delay map $\Phi_{n,d-2,\dots,q}: M \rightarrow \mathcal{R}^d$ defined by

$$\Phi_{n,d-2,\dots,q}(x_i) = \begin{bmatrix} y_i & y_{i+1} & \dots & y_{i+d-1} \\ g(x_i) & g(f_{r_i}(x_i)) & \dots & g(f_{r_{i+d-2}, \dots, r_i}(x_i)) \end{bmatrix}^T$$

is an embedding for almost all (f, g) and inputs, r . Hence, $\Phi_{n,d-2,\dots,q}: M \rightarrow \mathcal{R}^d$ is smooth and invertible. ■

Under the conditions of the theorem, $z_i = \Phi_{n,d-2,\dots,q}(x_i) = [y_i \ y_{i+1} \ \dots \ y_{i+d-1}]^T$ defines an alternative state space for the system and the dynamics may be expressed in terms of readily available input-output quantities as

$$y_{i+d} = F(y_i, y_{i+1}, \dots, y_{i+d-1}, r_i, r_{i+1}, \dots, r_{i+d-1}) \quad (2)$$

which is of course simply a form of NARMAX representation. Hence, at the cost of moving to a non-minimal realisation when $d > m$, embedding theory allows us to address the issue of selecting the state-space realisation.

Remark: Non-minimal realisations of equilibrium linearisations Suppose that we have an estimate of the linearised dynamics about an equilibrium point (obtained for example from local input-output data). More precisely, the input-output map of the equilibrium linearisation has been identified since full state information is not assumed to be measurable, *i.e.* the transfer function of the linearisation has been estimated or, equivalently, the state space matrices are known up to a non-singular linear state transformation. As linear identification techniques generally estimate a minimal model, the identified local linear dynamics will be order m in line with the order of the underlying nonlinear system (note that in the case of linear systems, an embedding dimension of m is sufficient). From Taken's theorem, the nonlinear realisation (2) is, however, generally non-minimal and, consequently, the associated equilibrium linearisations have non-minimal realisation. Nevertheless, we can readily derive the required non-minimal realisation of linearisation from any identified minimal realisation.

3. IDENTIFICATION COMBINING INPUT-OUTPUT DATA & ESTIMATED LINEARISATIONS

The empirical modelling task is to infer the nonlinear mapping F in (2) relating y_{i+d} to the associated vector of delayed outputs and inputs. When input and output time histories have been accurately measured, we can

form the vector of delayed inputs and outputs associated with each output value. The modelling task can then be formulated as a regression problem. In this paper, we consider the situation where this input-output data is augmented with estimates of the equilibrium linearisations of the nonlinear system. The requirement is therefore to derive a consistent model of the nonlinear dynamics which fuses the measured input-output data with the linearisation estimates.

In the recent statistics literature, there has been considerable interest in non-parametric modelling approaches. These offer flexible modelling tools which typically require only relatively weak structural assumptions and provide appropriate regularisation. The latter is essential when working with sparse data, as is the situation in local modelling methods and also when modelling in high dimensional spaces. Conversely, a disadvantage of many non-parametric modelling methods is that the computational burden rapidly increases with the size of the data set. The latter suggests that a fairly clear synergy might exist between a hybrid local/global modelling approach and non-parametric fitting methods, at least in the context of dynamic systems. In dynamic systems, it is inevitably the case that much more data is available in the vicinity of the equilibria than is available far away from equilibrium. This is because the off-equilibrium regions are associated with transient states of the system. The system will rarely linger for long in such regions and consequently measurements are typically sparse. (Note that while there is clearly an experiment design issue to be considered here, safety and operational restrictions also often mitigate against collecting much data in regions far from equilibrium and so the sparsity of data is often unavoidable). Conversely, there is typically a great deal of data available in the vicinity of the equilibria. Consider therefore a hybrid approach whereby the equilibrium data is summarised through local linear models while a non-parametric model is used to incorporate the (sparse) data characterising the dynamics far from equilibrium. In addition to the potential for computational efficiency, as noted previously such an approach provides an elegant and powerful framework which seems in good accord with engineering practice. In this paper, a non-parametric approach based on Gaussian process priors is adopted.

This section is organised as follows. Non-parametric Gaussian process prior models are introduced in section 3.1. In section 3.2, the use of such models to combine derivative and non-derivative observations in a consistent manner is discussed. A simple example illustrating the application of this approach is presented in section 3.3.

3.1 Gaussian Process Prior Models

Consider a stochastic process with output y conditional on input z (note that in this section, and in section 3.2, z denotes the input to a stochastic process rather than delay co-ordinates, but there should be no scope for confusion); specifically, the situation with $y(z)=F(z)+n$ i.e. a smooth scalar function $F(z)$ with additive Gaussian white measurement noise n . Observe that this is a regression formulation and within the context of the

present paper it is assumed the input z is noise free.² Suppose N measurements of input-output pairs, $\{(z_i, y_i)\}_{i=1}^N$, are available and denote them by M . It is of interest here to use this data to learn the mapping $F(z)$ or, more precisely, to determine the probability distribution of y for a new input value z ; that is, $p(y|z, M)$.

Consider initially the situation with standard parametric modelling approaches. For example, say we believe that $F(z)$ has the form $\Psi(z)\theta$ with $\Psi(z) = [\phi_1(z) \dots \phi_p(z)]$ and θ the model parameters (that is, $F(z)$ consists of a weighted combination of fixed basis functions $\phi_i(z), i=1..p$). We have that $p(y|z, M) = \int p(y|z, \theta)p(\theta|M)d\theta$ where $p(\theta|M)$ is the probability distribution over the set of possible models. Bayes' Rule states that $p(\theta|M) = p(M|\theta)p(\theta) / p(M)$ where the likelihood $p(M|\theta)$ embodies the information provided by the measured data, the prior $p(\theta)$ embodies our prior beliefs regarding the process and $p(M)$ is simply a normalising factor which is hereafter ignored. Then we might have

$$p(\theta|M) \propto \exp\left[-\frac{1}{2}(Y - \hat{Y})^T \Lambda_n^{-1}(Y - \hat{Y})\right] \times \exp\left[-\frac{1}{2}(\theta - \bar{\theta})^T \Lambda_\theta^{-1}(\theta - \bar{\theta})\right] \quad (3)$$

where $Y=[y_1 \dots y_N]^T$ is the vector of measured outputs, $\hat{Y}=[\Psi(z_1)\theta \dots \Psi(z_N)\theta]^T$, Λ_n is the measurement noise covariance, $\bar{\theta}$ and Λ_θ the mean and covariance of the prior (with large variance specified when we have little confidence in our prior knowledge). Commonly, it is assumed that the probability distribution $p(y(z)|M)$ is sharply peaked so that

$$p(y|z, M) \approx p(y|z, \theta_{MAP}) \propto \exp\left[-\frac{1}{2}(y - \Psi(z)\theta_{MAP})^T \Lambda_n^{-1}(y - \Psi(z)\theta_{MAP})\right] \quad (4)$$

where θ_{MAP} is the value of the parameter vector θ for which $p(\theta|M)$ is maximal. The mean prediction from this model (i.e. the fit to the function $F(z)$) is therefore $\Psi(z)\theta_{MAP}$, with variance Λ_n .

In non-parametric approaches, the prior belief is simply that the function $F(z)$ is smooth in some appropriate sense. In this way, the imposition of a specific parametric structure is avoided. To achieve this the prior is placed directly on the function space. In particular, in a non-parametric Gaussian Process prior model (see, for example, Juditsky *et al* 1995, Williams 1998) it is assumed that $p(y|z)$ is Gaussian with zero mean (the zero mean is for convenience only and may be relaxed) and covariance function $C(y(z_i), y(z_j)) = \text{cov}(y(z_i), y(z_j))$. The Gaussian assumption may seem strangely restrictive initially, but recall that this is simply a prior on the relevant function space and so places few inherent restrictions on the class of nonlinear functions that can be modelled.

²No attempt to being made here to propagate a Gaussian or other distribution through a nonlinear function.

Indeed, it can be shown that the result is, in fact, a Bayesian form of kernel regression model (Green & Silverman 1994). This model is used to carry out inference as follows. Clearly

$p(y|z, M) = p(y, \{y_i\}_{i=1}^N | z, \{z_i\}_{i=1}^N) / p(M)$ where $p(M)$ is a normalising constant. Hence, with the Gaussian prior assumption

$$p(y|z, M) \propto \exp\left[-\frac{1}{2}[y \ Y]^T \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12}^T & \Lambda_{22} \end{bmatrix}^{-1} \cdot \begin{bmatrix} y \\ Y \end{bmatrix}\right]$$

where Λ_{11} is $C(y(z), y(z))$, the ij th element of the covariance matrix Λ_{22} is $C(y(z_i), y(z_i))$ and the i th element of vector Λ_{12} is $C(y(z), y(z_i))$. Applying the partitioned matrix inversion lemma, it follows that

$$p(y|z, M) \propto \exp\left[-\frac{1}{2}(y - \hat{y}(z))^T \Lambda^{-1}(z)(y - \hat{y}(z))\right] \quad (5)$$

with $\hat{y}(z) = \Lambda_{12}^T \Lambda_{11}^{-1} Y$, $\Lambda(z) = \Lambda_{22} - \Lambda_{12}^T \Lambda_{11}^{-1} \Lambda_{12}$. The mean prediction from this model is therefore $\hat{y}(z)$, with variance $\Lambda(z)$. Note that $\hat{y}(z)$ is simply a weighted linear combination of the measured data points, Y , using weights $\Lambda_{12}^T \Lambda_{11}^{-1}$.

Remark A Gaussian process model is fully specified by its mean and covariance function. Here, for convenience, we assume a prior with zero mean. A common choice of covariance function is a smoothness prior which embodies a belief that outputs associated with nearby inputs should have higher covariance than more widely separated inputs; specifically,

$$C(y(z_i), y(z_j)) = \gamma \exp\left[-\sum_k \left((z_i)_k - (z_j)_k\right)^2 / 2\alpha_k\right] + \delta \quad (6)$$

where $(z_i)_k$ denotes the k th element of vector z_i . The value of α_k characterises the rate of variation of the function in dimension k , thereby, estimating the relative smoothness of different input dimensions. The parameter δ is the variance of the measurement noise, n , on the inputs. The hyperparameters $\{\delta, \alpha_k, \gamma\}$ are typically adapted to maximise the likelihood $p(\{y_i\}_{i=1}^N | \{z_i\}_{i=1}^N, M, (\delta, \alpha_k, \gamma))$.

3.2 Mixed Function and Derivative Observations

The Gaussian process prior modelling framework is readily extended to include situations where derivatives of a function $F(z)$ are observed rather than values of F itself (O'Hagan 1992). It is a standard result that the derivative process associated with the Gaussian stochastic process y is also Gaussian with

$$E\left(\frac{\partial y}{\partial z_i}(z_0), \frac{\partial y}{\partial z_j}(z_1)\right) = \nabla_i^T \nabla_j^T Q(z_0, z_1) \quad (7)$$

where $Q(z_0, z_1)$ denotes $E(y(z_0)y(z_1))$, $\partial y/\partial z_i$ denotes the partial derivative of y with respect to the i th element of z , $\nabla_i^T Q$ denotes the partial derivative of Q with respect to the i th element of its first argument, etc and it is assumed that Q is continuously twice differentiable. In the Gaussian process model, to work with derivative observations we need only replace the covariance function $C(y(z_i), y(z_j))$ by (7). In this case, the Gaussian process model acts to integrate and smooth the noisy derivative observations. The situation with mixed functional and derivative observations is similar. Provided we can define the covariance relating any two data points (namely, $\text{cov}(y(z_0), y(z_1))$, $\text{cov}(y(z_0),$

$\partial y/\partial z_i(z_1)$) and $\text{cov}(\partial y/\partial z_i(z_0), \partial y/\partial z_j(z_1))$), and this covariance function is differentiable, as it is in equation (6), the Gaussian process prior fit to the function $F(z)$ is specified by the analysis in section 3.1.

The useful role that derivative information can play, even in the context of modelling static maps, is illustrated by Figure 1. The dashed line marks the true function which is to be estimated. The values of this function which have been measured at the data points marked by asterisks. Evidently, this is a sparse data set. Using these function measurements, the mean prediction obtained using a GP with covariance function (6) is labelled in Figure 1 as "fit without derivative" and it can be seen that this fit is quite poor as might be expected in view of the few data points measured. Now consider augmenting these measurements with observations of the derivative of the function at the points $x=-2$ and $x=2$. The resulting mean prediction is also shown in Figure 1. It can be seen that the additional derivative information leads to a greatly improved fit.

3.3 Example

Consider the nonlinear dynamics

$$y_{i+1} = \tanh(0.9 \tanh^{-1}(y_i) + 0.5r_i) \quad (8)$$

Thirteen equilibrium points uniformly spanning the operating region of interest are selected. At each equilibrium point a small-scale random Gaussian perturbation input signal with mean 0 and standard deviation 0.02 is applied with the corresponding output signal contaminated with uniform measurement noise in the range [0.005, 0.005]. A linear approximation to the local dynamics at the equilibrium point is identified using the MATLAB algorithm ARX. A global input signal is utilised which drives the system over its operating envelope, with the corresponding output signal corrupted by uniform measurement noise in the range [0.02, 0.02]. A small sparse set of input-output data is selected (only 7 data points are used). A non-parametric Gaussian process prior model is constructed using the following training information:

1. Thirteen equilibrium input-output values spanning the operating region of interest.
2. The set of partial derivatives of the output corresponding to 1. (These are simply the coefficients of the identified linear models).
3. The seven input-output values.

The training data, test data and corresponding outputs are depicted in Figures 2a and 2b. It can be seen that, despite the small number of data points used, the output of the Gaussian process prior model is a good match for the output of the exact system.

4. ELLE SYSTEMS

Any nonlinear dynamics

$$y_{i+d} = F(y_i, y_{i+1}, \dots, y_{i+d-1}, r_i, r_{i+1}, \dots, r_{i+d-1}) \quad (9)$$

may be reformulated as

$$z_{i+1} = Az_i + Br_i + f(\rho), \quad y_i = Cz_i \quad (10)$$

where, $z_i = [y_i \ y_{i+1} \ \dots \ y_{i+d-1}]^T$, $r_i = [r_i \ r_{i+1} \ \dots \ r_{i+d-1}]^T$, A, B, C are appropriately dimensioned constant matrices, $f(\rho)$ is a nonlinear nonlinear functions and $\rho_i = Mz_i + Nr_i$ embodies the nonlinear dependence of the dynamics on the delayed inputs and outputs. Trivially, this reformulation can always be achieved by letting ρ_i

$= [z_i^T \ r_i^T]^T$, in which case the dimension q of ρ_i equals $2d$. However, the nonlinearity of the system may be dependent on only a subset of the delayed inputs and outputs, in which case the dimension, q , of ρ_i is less than $2d$. We have that,

$$\nabla_x z_{i+1} = A + \nabla f(\rho_i)M, \quad \nabla_x z_{i+1} = B + \nabla f(\rho_i)N, \quad \nabla_x y_i = C \quad (11)$$

Observe that the tangents, (11), are parameterised by ρ_i .

$$\text{Let } E = \{z^o, r^o: Az^o + Br^o + f(Mz^o + Nr^o) = z^o\}$$

denote the set of equilibrium points of the system (10), $R_p(E)$ denote the range of ρ_i on E (i.e. $R_p(E) = \{Mz + Nr: (z, r) \in E\}$) and $R_p(\Phi)$ the range of ρ_i on the full operating space of the system, $\Phi = \{(z, r): z \in \mathcal{R}^d, r \in \mathcal{R}^d\}$. Systems, (10), for which

$$R_p(E) = R_p(\Phi) \quad (12)$$

are referred to here as extended local linear equivalence (ELLE) systems (note that a more general form of ELLE system is also considered by Leith & Leithead (1998) in the context of gain-scheduling control design). The condition, (7), simply corresponds to the requirement that ρ_i is parameterised by the equilibrium points. It follows immediately that the tangents (11) are parameterised by the equilibrium points. Consequently, the equilibrium linearisations, together with knowledge of ρ_i , completely defines an ELLE system. In view of the importance of equilibrium information in classical theory (particularly gain-scheduling theory), and the relative ease with which equilibrium dynamics may be identified from measured data, the class of ELLE systems is of considerable interest in its own right. In the context of the present paper, the importance of ELLE systems is self-evident. Note that even if not exactly satisfied, it is often possible to utilise, within a useful operating envelope, an ELLE approximation to a non-ELLE system.

4.1 Example

Using an identical approach to the example in section 3.3, the system

$$y_{i+1} = 0.9y_i + 0.1 \tanh(5r_i) \quad (13)$$

is estimated. The training information is depicted in Figure 3a together with the test input and the corresponding output for the exact system and the output for the Gaussian process prior model are depicted in Figure 3b together with output for each of the equilibrium linearisations. From Figure 3a it can be observed that the training data is once again very sparse. Nevertheless, the output of the Gaussian process prior model matches well the output of the exact system, see Figure 3a.

Remark It is important to stress that the equilibrium linearisation input-output maps alone do not fully specify a nonlinear dynamic system, even when the system belongs to the ELLE class (in the latter case, as noted previously, it is necessary to have information regarding ρ_i in addition to knowledge of the linearisations). In particular, the systems $y_{i+1} = 0.9y_i + 0.1 \tanh(5r_i)$ and $y_{i+1} = \tanh(0.9 \tanh^{-1}(y_i) + 0.5r_i)$ both have linearisations with the same transfer function at every equilibrium point. To distinguish between such systems, additional information is required. The foregoing results nevertheless demonstrate that it can be

enough to include only a small amount of additional information in order to obtain models which are accurate within a large region.

5. CONCLUSIONS

In this paper, we investigate the reconstruction of nonlinear systems from locally identified linear models. It is, of course, well known that the equilibrium linearisations of a system do not uniquely specify the global dynamics. Information about the dynamics near to equilibrium provided by the equilibrium linearisations must therefore combined with other information about the dynamics away from equilibrium provided by suitable measured data. That is, a hybrid local/global modelling approach is considered. A non-parametric Gaussian process prior approach is proposed for combining in a consistent manner these two distinct types of data: namely, estimates of the equilibrium linearisations (derivative/tangent map data) and input-output trajectory data measured during transitions between equilibria and operation far from equilibrium.

ACKNOWLEDGEMENTS

This work was supported by the Royal Society through a personal research fellowship to D.Leith, by Science Foundation Ireland grant 00/PI.1/C067, by the EC through EC TMR grant HPRNCT-1999-00107, EPSRC grants GR/R15863/01 and GR/M76379/01.

REFERENCES

- GREEN, P.J., SILVERMAN, B.W., 1994, *Nonparametric Regression and Generalised Linear Models*. (Chapman & Hall, London).
- JOHANSEN, T.A., MURRAY-SMITH, R., 1997, *The Operating Regime Approach to Nonlinear Modelling and Control. In Multiple Model Approaches to Modelling and Control* (Eds. Murray-Smith, R. Johansen, T.A.) (Taylor & Francis).
- JUDITSKY, A., HJALMARSSON, H., BENVENISTE, A., DEYLON, B., LUNG, L., SJOBERG, J., ZHANG, Q., 1995, Nonlinear black box models in system identification: mathematical foundations. *Automatica*, 31, pp1725-1750.
- LEITH, D.J., LEITHEAD, W.E., 1998, Gain-Scheduled & Nonlinear Systems: Dynamic Analysis by Velocity-Based Linearisation Families. *International Journal of Control*, 70, 289-317.
- LEITH, D.J., LEITHEAD, W.E., 1999, Analytic Framework for Blended Multiple Model Systems Using Linear Local Models. *Int. J. Control*, 72, pp605-619.
- MURRAY-SMITH, R., JOHANSEN, T.A., SHORTEN, R., 1999, On Transient Dynamics, Off-Equilibrium Behaviour and Identification in Blended Multiple Model Structures. *Proc. European Control Conference, Karlsruhe*.
- NARENDRA, K.S., MUKHOPADHYAY, S., XIANG, C., CABRERA, D., 1996, Nonlinear Models for Identification & Control. *Proc. Ninth Yale Workshop on Adaptive & Learning Systems*.
- O'HAGAN, A., 1978, On curve fitting and optimal design for regression. *J. Royal Stat Soc. B*, 40, 1-42.
- O'HAGAN, A., 1992, Some Bayesian Numerical Analysis. *In Bayesian Statistics 4, Ed J.M. Bernardo, J.O. Berger, A.P. Dawid, A.F.M. Smith*, pp 345-363, Oxford University Press.

SHORTEN, R., MURRAY-SMITH, R., BJØRGAN, R., GOLLEE, H., 1999, On the interpretation of local models in blended multiple model structure. *Int. J. Control*, 72.

STARK, J., BROOMHEAD, D.S., DAVIES, M.E., HUKE, J., 1997, Takens Embedding Theorems for Forced & Stochastic Systems. *Nonlinear Analysis*, 30, pp5303-5314.

STARK, J., BROOMHEAD, D.S., DAVIES, M.E., HUKE, J., 2001, Delay Embeddings for Forced Systems. www.ucl.ac.uk/CNDA/papers/Random_Embed.pdf

WILLIAMS, C. K. I., 1998, Prediction with Gaussian Processes: From linear regression to linear prediction and beyond. *In Learning and Inference in Graphical Models* (M. I. Jordan, Ed.), Kluwer.

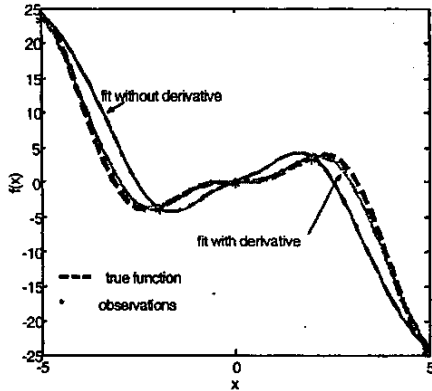


Figure 1 Example illustrating utility of derivative observations.

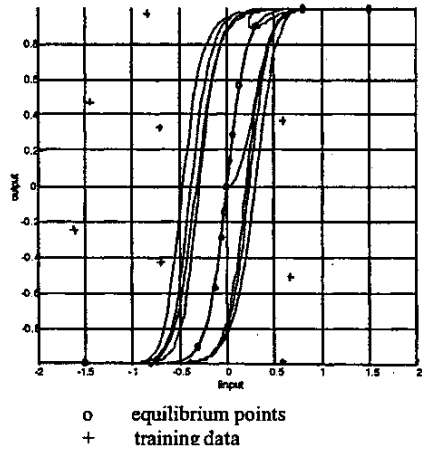


Figure 2a

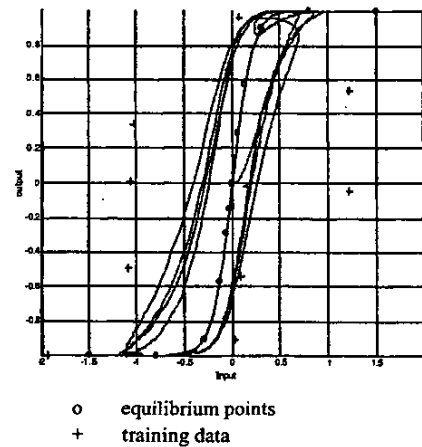


Figure 3a

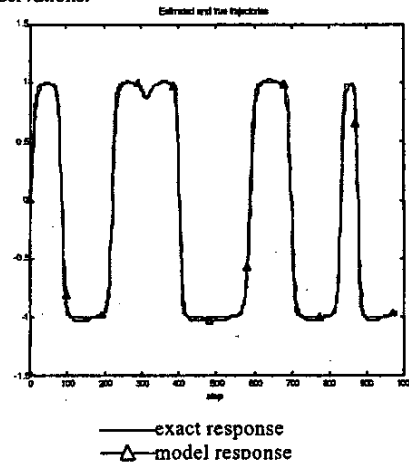


Figure 2b

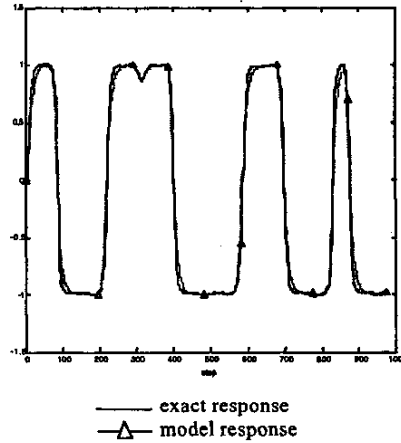


Figure 3b