

paper we describe the basic methodology, review some previous results and present some new applications.

Section 2 describes the modelling procedures we employ. Section 3 gives several examples of applications of this methodology and Section 4 provides some concluding remarks.

2 Modelling

Nonlinear modelling is logically and operationally divided into three important issues: (i) reconstruction, (ii) functional approximation, and (iii) model selection. In Section 2.1 we describe the reconstruction methodology. Section 2.2 describes the basic functional form of radial basis models and Section 2.3 describes the application of minimum description length for model selection. In Section 2.4 we describe the extension of radial basis modelling to cylindrical basis functions. Finally, Section 2.5 describes the operation of time embedding that we employ to model nonstationary data.

2.1 Reconstruction

The much cited Takens' embedding theorem¹³ provides a guarantee that under certain conditions it is possible to reconstruct a finite dimensional dynamical system (Φ, \mathcal{X}) (where $\Phi : \mathcal{X} \mapsto \mathcal{X}$, $x_t \in \mathcal{X}$, $\dim(\mathcal{X}) = d$) from a single scalar time series $\{u_t\}_{t=1}^N$ where $u_t = g(x_t)$. In practice one cannot be sure that the observation function $g : \mathcal{X} \mapsto \mathbf{R}$ is diffeomorphic (i.e. u_t may only observe a subsystem) and that the constraints on noise level and observation length N may not be met. However, a time delay embedding still allows one to reconstruct a noisy attractor with dynamics approximately equivalent to the underlying dynamical system (or a subsystem).

The time delay embedding is defined by

$$v_t = (u_t, u_{t-\tau}, u_{t-2\tau}, \dots, u_{t-(d_e-1)\tau})^T$$

where τ and d_e are the *embedding lag* and *embedding dimension*. Embedding dimension and embedding lag may be selected by one of many criteria including: false nearest neighbours¹⁴, zero of autocorrelation¹⁵, minimum of mutual information¹⁶ or one-quarter the quasi-period⁸.

The *evolution operator* of the underlying dynamical system Φ ($x_{t+1} = \Phi(x_t)$ or the analogous for a continuous system) may then be approximated by a function F ,

$$v_{t+1} = F(v_t) + e_t$$

where $e_t \in \mathbf{R}^{d_e}$ is an i.i.d. noise vector. For simplicity we take $\tau = 1$ and we then have that

$$\begin{aligned} v_{t+1} &= F(v_t) + e_t \\ &\equiv \begin{bmatrix} f(v_t) \\ u_t \\ u_{t-1} \\ u_{t-2} \\ \vdots \\ u_{t-d_e+2} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \end{aligned}$$

The modelling problem becomes one of finding a functional approximation $f : \mathbf{R}^{d_e} \mapsto \mathbf{R}$ such that

$$\sum_{t=d_e+1}^N [f(v_t) - u_{t+1}]^2 \quad (1.1)$$

($\epsilon_t = u_{t+1} - f(v_t)$) is minimised.

2.2 Radial Basis Modelling

Let $x \in \mathbf{R}^{d_e}$ then a radial basis model f is defined as

$$f(x) = \sum_{i=1}^m \alpha_i \phi_i \left(\frac{\|x - c_i\|}{r_i} \right) + \sum_{j=1}^{d_e} \beta_j x_j \quad (1.2)$$

where $\phi_i : [0, \infty) \mapsto \mathbf{R}$ is the *basis function*, $c_i \in \mathbf{R}^{d_e}$ is the *centre* and $r_i \in \mathbf{R}$ is the *radius*. The weights α_i , β_j may be selected by the usual least squares approach¹⁷ (with the caveat that some b_j may be set equal to zero). The parameters c_i and r_i will need to be selected by some nonlinear selection routine and the basis functions are selected from some class of candidates. Typically one may take ϕ_i to be selected from

- $\left\{ \exp\left(\frac{-x^2}{2}\right) \right\}$
- $\{x^3\}$
- $\{\tanh(mx + b) : m \in \mathbf{R}, b \in \mathbf{R}\}$
- $\left\{ \exp\left(\frac{(1-p)x^p}{p}\right) : p > 0 \right\}$

- $\{(2(x/s)^2 - 1) \exp(-(x/s)^2) : s > 0\}$

or some combination of these. The results described in this paper were obtained by selecting

$$\phi_i \in \left\{ \exp\left(\frac{-x^2}{2}\right) \right\} \cup \{1\}.$$

For a given model size m the optimal model of that size is selected according to the algorithm described in Judd and Mees² or Small and Judd⁵ to minimise (1.1). The optimal model size m is selected by computing the description length for all model sizes and selecting the model with *minimum description length*.

2.3 Minimum Description Length

The *description length* of a model is (roughly) the number of bits of information required to describe the data by describing the model of that data, initial conditions and the model prediction errors. If a model is a good model, then the description length should be less than the description length of the raw data (the model provides a compact description of the features observed in the data). If the model is poor then its description length will be larger than that of the data alone (the model is more complex than the data). Description length is described in detail by Rissanen³. The key feature of description length is that the data values, model parameters and model prediction errors need only be specified to some finite accuracy. A model with low description length will have many parameters which need not be specified too precisely.

Rissanen shows that the description length of a parameter λ_i specified to some accuracy δ_i is $\log(\frac{\gamma}{\delta_i})$ ^{2,3}. The constant γ is not critical and is related to the binary representation of floating point numbers². Therefore, the description length of k model parameters $\Lambda = \{\lambda_i\}_{i=1}^k$ is given by

$$L(\Lambda) = \sum_{i=1}^k \log\left(\frac{\gamma}{\delta_i}\right).$$

The description length of the data $\{u_t\}_{t=1}^N$, and the model with parameters Λ , is given by

$$L(u, \Lambda) = L(u|\Lambda) + L(\Lambda) \tag{1.3}$$

where the description length of the data given the model (i.e. the description length of the model prediction errors) $L(u|\Lambda)$ is the negative logarithm of the likelihood of the data under the assumed distribution, $-\ln(P(u|\Lambda))$.

Clearly the *minimum* description length of a given model will depend critically on the optimal selection of the model parameter precisions δ_i . A second order expansion of (1.3) about the maximum likelihood model parameter values indicates that the optimal values of $\delta = (\delta_1, \delta_2, \dots, \delta_k)$ are the solution to

$$(D_{\Lambda\Lambda}L(u|\Lambda) \cdot \delta)_i = \frac{1}{\delta_i}$$

where $D_{\Lambda\Lambda}L(u|\Lambda)$ is the second derivative of the model with respect to the model parameters and $(D_{\Lambda\Lambda}L(u|\Lambda) \cdot \delta)_i$ is the i -th component of the vector $D_{\Lambda\Lambda}L(u|\Lambda) \cdot \delta$. The description length of data with respect to a given model is then

$$L(u|\Lambda) + \left(\frac{1}{2} + \ln(\gamma)\right)k - \sum_{i=1}^k \ln(\delta_i)$$

(the extra $\frac{1}{2}k$ term comes from solving (1.3) for the maximum likelihood values δ_i^2).

2.4 Cylindrical Basis Models

Standard radial basis models are functions of the form (1.2). *Cylindrical basis* models are described by functions of the form

$$f(x) = \sum_{i=1}^m \alpha_i \phi_i \left(\frac{\|P_i(x - c_i)\|}{r_i} \right) + \sum_{j=1}^{d_e} \beta_j x_j$$

where $P_i : \mathbf{R}^{d_e} \mapsto \mathbf{R}^{d_e}$ is a projection onto some subset of the coordinate directions. The inclusion of this additional component effectively removes some of the problems involved in the “correct” selection of embedding parameters τ and d_e and makes embedding an intrinsic part of the modelling process^{5,4}. Moreover, dynamical systems often have increased complexity only in some parts of phase space — the Lorenz system for example is mostly two dimensional (on the wings), three dimensional structure is only significant at the central separatrix⁴. Successful application of the minimum description length principle should mean that the optimal model utilises precisely the dimensionality required in the correct parts of phase space, but no more.

2.5 Embedding Time

Time dependent structure may be incorporated into a radial or cylindrical basis model by considering time as a coordinate along with the scalar observations

u_t . With $\tau = 1$ we embed v_t according to

$$v_t = (u_t, u_{t-1}, u_{t-2}, \dots, u_{t-(d_e-1)}, kt)^T$$

for some constant k .

The evolution operator of the underlying dynamical system Φ may now be approximated by a function F ,

$$v_{t+1} = F(v_t) + e_t$$

$$\equiv \begin{bmatrix} f(v_t) \\ u_t \\ u_{t-1} \\ u_{t-2} \\ \vdots \\ u_{t-d_e+2} \\ k(t+1) \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}.$$

3 Applications

In this section we present several applications of the techniques described in Section 2. In Section 3.1 we exemplify the application of this method to detrend time series. Section 3.2 demonstrates the application of this method to identify a period doubling bifurcation in the logistic equation and bifurcation in the Rössler system. Section 3.3 briefly reviews the results of Judd and Mees¹²: an experimentally vibrating string undergoing a Shil'nikov type bifurcation. In Section 3.4 we present previously unpublished results showing a period doubling bifurcation in infant respiratory recordings prior to onset of periodic breathing. Section 3.5 describes some preliminary results of the application of these techniques to computational simulations of cardiac arrhythmia.

3.1 Nonstationary Simulations

The modelling technique described in this paper is capable of estimating bifurcation diagrams, but also modelling nonstationary nonlinear trends (and also more generally nonstationary data). Figure 1 shows a computational simulation of a nonstationary nonlinear trend added to a random process. The presence of such non-stationarity can be detected with the algorithms described by Yu *et al*^{18,19}. Figure 2 shows the results of nonstationary modelling applied to this data. The trend has been accurately extracted and the noise component has been estimated with a correlation coefficient of 0.9989.