Nonlin. Processes Geophys., 13, 577–584, 2006
www.nonlin-processes-geophys.net/13/577/2006/

Nonlinear Processes in Geophysics

An efficient nonlinear cardinal B-spline model for high tide forecasts at the Venice Lagoon

H. L. Wei and S. A. Billings

Department of Automatic Control and Systems Engineering, University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK

Received: 19 July 2006 – Revised: 16 October 2006 – Accepted: 16 October 2006 – Published: 23 October 2006

Abstract. An efficient class of nonlinear models, constructed using cardinal B-spline (CBS) basis functions, are proposed for high tide forecasts at the Venice lagoon. Accurate short term predictions of high tides in the lagoon can easily be calculated using the proposed CBS models.

1 Introduction

The Venice lagoon is one of the world's most delicate and unstable ecosystems. Since the disastrous flood that occurred in November 1966, the problems of the Venice lagoon have become one of national and international interest. The threatened Venice city has frequently been inundated by high waters formed in the northern Adriatic Sea, where interactions of several astronomical and meteorological phenomena often occur. The end results are the Venice floods due to a combination of astronomical and meteorological effects: the tides induced by the moon and the tides caused by stormy weather arise from low atomospheric pressure combined with winds. To prevent disastrous floods, measures have been taken since 1966, and perhaps the most famous project is the recently endorsed MoSE (*Mo*dulo *S*perimentale *E*lettromeccanico—Experimental Electromechanical Module) project, although the feasibility of this project is still in public debate (Rosenthal, 2005; Salzano, 2005). A parallel and complementary approach to engineering constructions, for example the barrier system as involved in the MoSE, is to build an operational flood warning system, which is used to forecast the main surge, for some time ahead ideally many hours or even several days. The objective of such a flood warning system is to support some necessary actions such as the removal of goods from ground floors, the redirection of the city boat traffic, and the installation of elevated pedestrian walkways (Vieira et al., 1993). The flood warning system is model-based: it utilises both statistical and hydrodynamic models to obtain short term as well as long term forecasts (Vieira et al., 1993). The hydrodynamic modelling usually starts with first principles that require a comprehensive physical insight into the underlying dynamics of the system, whereas the statistical modelling and similar methods often start with observational data, based on which mathematical models that support forecasts of the main surge are deduced.

Several authors have discussed the data-based modelling problem relating to high tide forecasts at the lagoon, by treating the regularly measured water level as a nonlinear time series, with the assumption that no information on the hydrodynamics of the lagoon is involved, but merely observed water level data are available (Zaldivar et al., 2000). Many approaches have been proposed to model the associated nonlinear time series including nonlinear regression models, chaos and embedding methods, neural networks, evolutionary algorithms, and other methods, see Zaldivar et al. (2000) and del Arco-Calderon et al. (2004) and the references therein.

This study aims to present a novel and efficient data-based modelling approach for predicting high tides at the Venice lagoon. In the new modelling approach, it is assumed that no a priori knowledge about the hydrodynamics of the lagoon is available, but merely observed water level data are used. Motivated by the successful applications of wavelet transforms, especially the applications of wavelet multiresolution decompositions, in nonlinear time series analysis and complex dynamical system identification including geophysical and magnetospheric process modelling (Kumar and Foufoula-Georgiou, 1997; Malamud and Turcotte, 1999a, b; Chandre et al., 2003; Maraun and Kurths, 2004; Grinsted et al., 2004; Wei et al., 2004a, b; Wei and Billings, 2004; Kallache et al., 2005), cardinal B-spline multiresolution analysis (MRA) is employed in the present study to construct parsimonious nonlinear models that can be used for high tide forecasting. As will be seen, the resulting CBS models

Correspondence to: H. L. Wei
(w.hualiang@shef.ac.uk)

Published by Copernicus GmbH on behalf of the European Geosciences Union and the American Geophysical Union.

Table 1. Cardinal B-splines of order 1 to 4.

	$N_1(x)$	$N_2(x)$	$2N_3(x)$	$6N_4(x)$
$0 \le x < 1$	1	x	x^2	x^3
$1 \le x < 2$	0	$2-x$	$-2x^2+6x-3$	$-3x^3+12x^2-12x+4$
$2 \le x < 3$	0	0	$(x-3)^2$	$3x^3-24x^2+60x-44$
$3 \le x \le 4$	0	0	0	$-x^3+12x^2-48x+64$
elsewhere	0	0	0	0

provide not only accurate short term forecasts, but also provide good long term predictions for the variation of water levels in the lagoon. Compared with existing data-based methods, the proposed data-based CBS modelling approach can produce more accurate predictions for high tides at the Venice lagoon.

2 Time series forecasting problem

Let $\{y(t)\}_{t=t_0}^{T}$ be a known observed sequence for the underlying dynamical time series. The goal of multi-step-ahead forecasts is to predict the values of $y(t+s)$, with $s \ge 1$, using the information carried by the observed sequence $\{y(t)\}_{t=t_0}^{T}$. To achieve such a goal, a commonly used approach is to learn a model, or a predictor, from the available data. To obtain multi-step-ahead predictions of nonlinear time series, both iterative and direct methods can be employed (Wei and Billings, 2006). In theory, long-term predictions can be obtained from a short-term predictor, for example a one-step-ahead predictor, simply by applying the short predictor many times in an iterative way. This is called iterative prediction. Direct prediction, however, provides a once-completed predictor and multistep forecasts can be obtained directly from the established predictor in a way that is similar to computing one step predictions.

Following Wei and Billings (2006), a direct approach will be considered. Take the case of the s-step-ahead forecasting problem as an example. The task for s-step-ahead forecasts is to find a model that can predict the value of $y(t+s)$ using a set of selected variables $\{y(t), y(t-1), \cdots, y(t-d+1)\}$, in the sense that

$$y(t+s) = f^{(s)}(y(t), \cdots, y(t-d+1)) + e(t) \tag{1}$$

where $f^{(s)}$ with $s \ge 1$ are some nonlinear functions, $e(t)$ is an unpredictable zero mean noise sequence, d is the model order (the maximum lag). For a real system, the nonlinear function $f^{(s)}$ is generally unknown and might be very complex. A class of models that are both flexible, with excellent approximation capabilities, and which can represent a broad class of highly complex systems are therefore required to ensure accurate direct s-step predictions. The model class that uses cardinal B-splines as the basis functions to approximate the s-step predictor $f^{(s)}(\cdot)$ satisfies all these conditions and will therefore be investigated in the present study as a new approach of achieving accurate direct s-step predictions.

3 Cardinal B-spline models

3.1 Cardinal B-splines

The mth order cardinal B-spline function is defined by the following recursive formula (Chui, 1992):

$$N_m(x) = \frac{x}{m-1} N_{m-1}(x) + \frac{m-x}{m-1} N_{m-1}(x-1), \ m \ge 2 \tag{2}$$

where

$$N_1(x) = \chi_{[0,1)}(x) = \begin{cases} 1 \text{ if } x \in [0,1) \\ 0 \text{ otherwise} \end{cases} \tag{3}$$

It can easily be shown that the support of the mth order B-spline function is $\text{supp} N_m = [0, m]$. Compared with other basis functions, the most attractive and distinctive property of B-splines are that they are compactly supported and can be analytically formulated in an explicit form. Most importantly, they form a multiresloution analysis (MRA) (Chui, 1992). B-splines are unique, among many commonly used basis functions, because they simultaneously possess the three remarkable properties, namely compactly supported, analytically formulated and multiresolution analysis oriented, among many popular basis functions. These splendid properties make B-splines remarkably appropriate for nonlinear dynamical system modelling. The most commonly used B-splines are those of orders 1 to 4, which are shown in Table 1.

For the mth order B-spline function $N_m \in L^2(R)$, let $N_{j,k}^m(x) = 2^{j/2} N_m(2^j x - k)$, $D_j^m = \{N_{j,k}^m : k \in Z\}$, where $j, k \in Z$ are called the scale (or dilation) and position (translation) parameters respectively. Following (Chui, 1992), for each $j \in Z$, let V_j^m denote the closure of the linear span of D_j^m, namely, $V_j^m = \text{clos}_{L^2(R)} < D_j^m >$. The following properties (Chui, 1992) possessed by D_j^m and V_j^m form the foundations of the cardinal B-spline multiresolution analysis modelling framework for nonlinear dynamical systems:

i) For any pair of integers m and j, with $m \ge 2$, the family $D_j^m = \{N_{j,k}^m(x) : k \in Z\}$ is a Riesz basis of V_j^m with Riesz bound $A = A_m$ (A_m is a constant related to m) and $B = 1$. Furthermore, these bounds are optimal.

ii) The mth order B-spline function N_m is a scaling function and V_j^m forms a multiresolution analysis (MRA).

From the above discussions, for every function $f \in V_j^m$, there exists a unique sequence $\{c_k^m\}_{k \in Z} \in \ell^2(Z)$ such that

$$f(x) = \sum_{k \in Z} c_k^m 2^{j/2} N_m(2^j x - k) \tag{4}$$

For convenience of description, the symbol ϕwill be introduced to represent the mth order B-spline function N_m and the symbol "m" will be omitted in associated formulas.

3.2 The cardinal B-spline model for high dimensional problems

The result for the 1-D case described above can be extended to high dimensions and several approaches have been proposed for such an extension. Tensor product and radial construction are two commonly used methods (Wei and Billings, 2004; Billings and Wei, 2005). Following the idea in Hastie and Tibshirani (1990) and Kavli (1993), in the present study, a linear additive CBS model structure will be employed to represent a high dimensional nonlinear function. Kavli (1993) suggested a method to successively refine a linear B-spline model for multivariate problems by adding new 1-D submodels step by step.

For a d-dimensional function $f \in L^2(R^d)$, the linear additive representation is given below

$$f(x_1, x_2, \cdots, x_d) = f_1(x_1) + f_2(x_2) + \cdots + f_d(x_d) \quad (5)$$

where $f_r \in L^2(R)$ (r=1,2,...,d) are univariate functions, which can be expressed using the expansion (4) as below

$$f_r(x_r) = \sum_{k \in Z} c^r_{j,k} \phi_{j,k}(x_r) \quad (6)$$

where $\phi_{j,k}(x)=2^{j/2}\phi(2^j x-k)$, and $j, k \in Z$ are the scale and position parameters, respectively.

Now consider the model given by Eq. (1) and let $x_r(t)=y(t-r+1)$ for r=1, 2,..., d. Using Eqs. (5) and (6), model (1) can be expressed as

$$y(t+s) = \sum_{r=1}^{d} f_r^{(s)}(x_r(t)) = \sum_{r=1}^{d} \sum_{k \in Z} c_{j,k}^{(s,r)} \phi_{j,k}(x_r(t)) + e(t) \quad (7)$$

The remaining task is how to deduce, from Eq. (7), a parsimonious model that can be used for s-step-ahead forecasts for a given prediction horizon s. The following problem needs to be solved:

- How to choose the scale and position parameters j and k ?
- In practical modelling problems, the variables $x_r(t)$ (r=1, 2, ..., d), as the lagged versions of $y(t)$, are usually sparsely distributed in the associated space and therefore the problem may be ill-posed. The representation (7) is thus often redundant in the sense that most of the basis functions (or model terms), $\phi_{j,k}(\cdot)$ in Eq. (7), can be removed from the model, and experience shows that only a small number of significant model terms are required for most nonlinear dynamical modelling problems. The question is: how to select the potential significant model terms from a large number of candidate basis functions?

The scale and position determination problem will be discussed in the following section. The model term selection problem has been systematically investigated in Billings et al., 1989; Chen et al., 1989). In the present study, an orthogonal least squares (OLS) algorithm, interfered with by an error reduction ratio (ERR) index (Billings et al., 1989; Chen et al., 1989), and regularized by a Bayesian information criterion (BIC) (Schwarz, 1978; Efron and Tibshirani, 1993), will be used to select significant model terms and to determine the model size (the number of model terms included in the final model). One version of the OLS-ERR type algorithm, called the forward orthogonal regression (FOR) algorithm, is presented in the Appendix.

3.3 Determination of the scale and position parameters

Assume that a d-variate function f of interest is defined in the unit hypercube $[0, 1]^d$. Consider the scale parameter determination problem first. Experience on numerous simulation studies relating to wavelet multiresolution modelling for dynamical nonlinear systems, for example Wei et al. (2004a, b) and Wei and Billings (2004) and the references therein, has shown that the scale parameter j in model (7) should not be chosen too large. A value that is between zero and two or three for j is often adequate for most nonlinear dynamical modelling problems.

For cardinal B-spline functions, the position parameter k is dependent on the corresponding resolution scale j. Indeed, for each fixed point $x \in [0, 1]$, since N_m has compact support, all except a finite number of terms in the expression (4) are zero. Take the 4th-order B-spline function as an example. At a given scale j, the non-zero terms are determined by the position parameter k for $k=-3, -2, -1, \cdots, 2^j-1$. In general, for the B-spline function of order m, whose support is $[0, m]$, the support for the associated function $\phi_{j,k}(x)=2^{j/2}(2^j x-k)$ is $[2^{-j}k, 2^{-j}(m+k)]$, therefore, the position parameter k at a resolution scale j should be chosen as $-(m-1) \le k \le 2^j-1$.

4 Water level modelling and high tide forecasting

4.1 The data

The data used here are the hourly recorded observations of water levels at Punta della Salute, Venice Lagoon, for the period from January 1990 to December 1994. These data were partitioned into 5 data sets, symbolized by "data90", "data91", "data92", "data93", and "data94", corresponding to different year numbers. The number of observations in the data set "data92" was 8784, and was 8760 in the other data sets. In the modelling procedure, the observations in each year were chosen as the training data set and were used to identify a model. The resultant model was then used to predict water levels for the next year. For example, a model identified from the data set "data92" will be used to predict

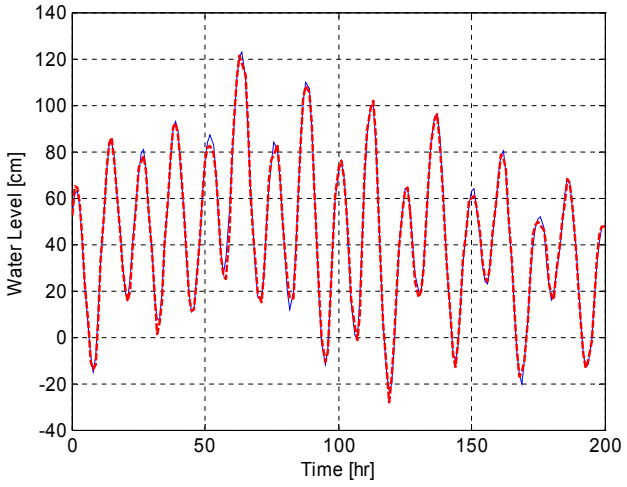

Fig. 1. One-hour-ahead prediction for typical high tides. The thin line with dots indicates the measurements (observed in 1993), and the thick dashed line indicates the prediction values.

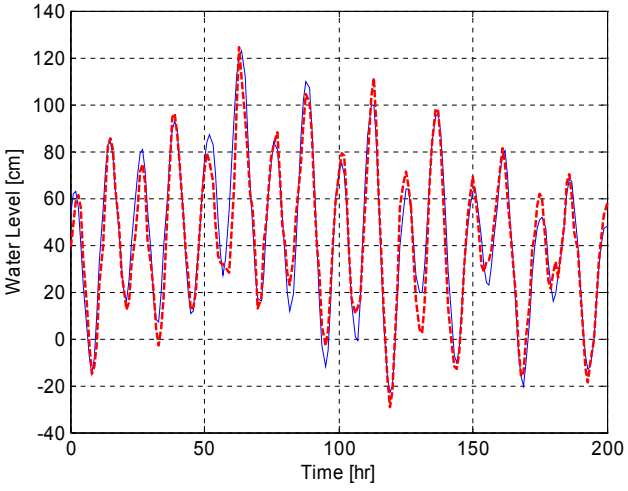

Fig. 2. Four-hour-ahead prediction for typical high tides. The thin solid line indicates the measurements (observed in 1993), and the thick dashed line indicates the prediction values.

water levels in 1993 and the predicted values will then be compared with the real observations.

The maximum lag for the input variables in the initial modelling procedure was chosen to be 24, to cover the range of the maximum oscillation cycle of the related time series. Thus, the variables $y(t), y(t-1), \cdots, y(t-23)$ were used as inputs to form a predictor, whose output was the future behaviour, denoted by $y(t+s)$ $(s \geq 1)$.

Note that the original data were initially normalized to [0,1] via a transform $y(t)=(\tilde{y}(t)-a)/(b-a)$, where $\tilde{y}(t)$ indicate the initial observations, and $a=-100$ and $b=150$. The identification procedure was therefore performed using normalized values $y(t)$. The outputs of an identified model were then recovered to the original measurement space by taking the associated inverse transform.

4.2 The models

Let $x_r(t)=y(t-r+1)$, $r=1, 2, \ldots, 24$. The structure of the initial CBS model was chosen to be

$$y(t+s) = \sum_{r=1}^{24} \sum_{k=-3}^{0} c_{0,k}^{(s,r)} \phi_{0,k}(x_r(t)) + \sum_{r=1}^{24} \sum_{k=-3}^{1} \alpha_{1,k}^{(s,r)} \phi_{1,k}(x_r(t)) \quad (8)$$

where $\phi_{j,k}(x)=2^{j/2}\phi(2^j x-k)$,with $j, k \in Z$, are the 4th-order B-spline functions. Notice that model (8), which involves two scale levels for $j=0$ and $j=1$, is in structure different from model (7), where the model term $\phi_{j,k}(\cdot)$ only involves a single scale level. The reason that the initial model (8) was chosen to be such a structure was to enrich the pool of the model term dictionary, so that basis functions with different scale parameters can be sufficiently utilised. Although a total number of 216 model terms (basis functions) were involved in the initial model (8) for any given s, only a small number of basis functions were required to describe the relationship between $\{y(t), y(t-1), \cdots, y(t-23)\}$ and $y(t+s)$, and significant model terms were efficiently selected by performing a model term detection algorithm,. Also, different values for s usually led to different final models. For each s, an OLS-ERR algorithm (Billings et al., 1989; Chen et al., 1989), regularized by a Bayesian information criterion (BIC) (Schwarz, 1978; Efron and Tibshirani, 1993), was used to determine the number of model terms, and the parameters of the final CBS model was then re-estimated by introducing a linear moving average (MA) model of order 10 (Billings and Wei, 2005; Wei and Billings, 2006).

4.3 Prediction results

Eight cases, corresponding to s=1, 4, 12, 24, 28, 48, 72, and 96, were considered, and eight different CBS models were identified for each of four data sets "data90", "data91", "data92", and "data93". The resultant eight models were applied respectively over four test data sets, "data91", "data92", "data93", and "data94", to calculate s-step-ahead forecasts of water levels. Prediction performance, measured by the root-mean-square-errors (RMSE) as used in Zaldivar et al. (2000) and del Arco-Calderon et al. (2004), over the four test data sets, obtained from the identified CBS models, are shown in Table 2. Compared with the results produced by multilayer neural networks (Zaldivar et al., 2000) and evolutionary algorithms (del Arco-Calderon et al., 2004), the results produced by the proposed CBS models are better, both for short and long term predictions.

To visually illustrate the performance of the identified CBS models for high tide forecasting, both short term and

Table 2. Prediction errors for water levels of the years 1991, 1992, 1993, and 1994, with 8760, 8784, 8760, and 8760 records, respectively.

Prediction horizon	1991		1992		1993		1994	
	Model size	RMSE	Model size	RMSE	Model size	RMSE	Model size	RMSE
1	39	1.521	38	1.538	49	1.519	37	1.489
4	36	5.389	35	5.408	35	5.274	31	5.075
12	24	7.072	21	7.406	18	6.656	20	6.439
24	23	7.325	16	7.353	18	6.858	19	6.584
28	27	9.232	26	9.246	26	8.637	23	8.352
48	22	10.686	18	10.752	23	9.788	20	9.716
72	20	13.036	21	13.352	23	12.019	19	11.940
96	18	14.480	23	15.108	25	13.576	21	13.449

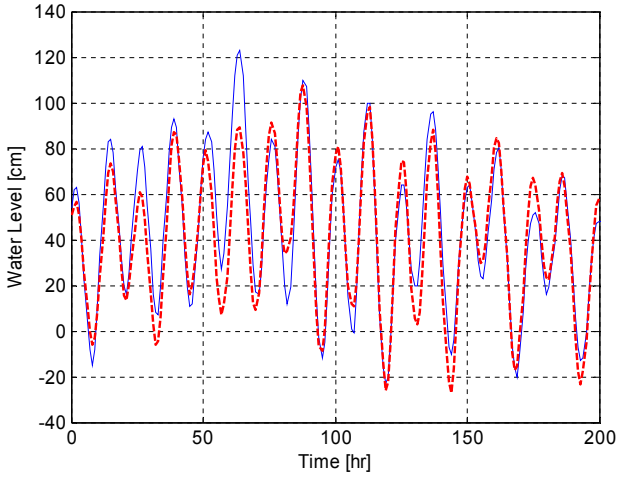

Fig. 3. Twelve-hour-ahead prediction for typical high tides. The thin solid line indicates the measurements (observed in 1993), and the thick dashed line indicates the prediction values.

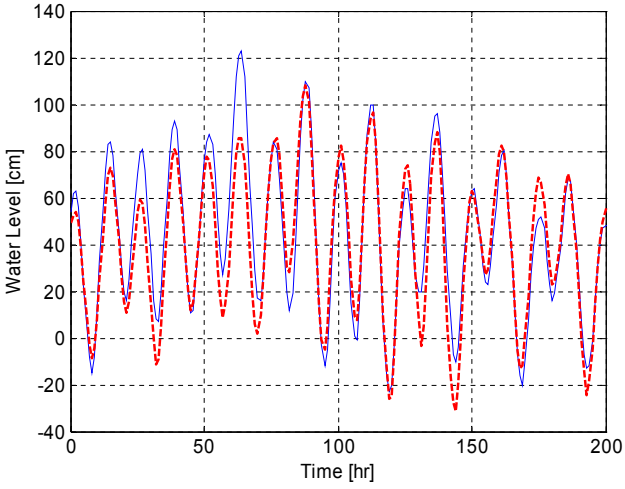

Fig. 4. Twenty-four-hour-ahead prediction for typical normal water level. The thin solid line indicates the measurements (observed in 1993), and the thick dashed line indicates the prediction values.

long term predictions for some high tides were calculated using the identified CBS models. Taking the prediction results for some typical high tides in the year 1993 as an example, the 1-, 4-, 12-, and 24-h-ahead predictions are shown in Figs. 1 to 4, respectively.

It can be seen from Figs. 1 to 4 that albeit the identified CBS models can produce very good short term (4 h-ahead) predictions for typical high tides at the lagoon, the resultant models can not effectively produce long (>12 h-ahead) term predictions for high waters (water level >110 cm). The reason that the models can not provide effective long term predictions may be that the "input" signals, considered here, for the associated dynamical models may not be sufficient to describe the real world dynamical systems, in other words, additional input signals may be required to adequately characterize the underlying dynamical behaviour. One solution to this problem is likely to involve multiple time series, and tides at several places in the Adratic Sea or the use of a 3-D hydrodynamical model with adequate forcing. Following the idea of Cao et al. (1998), the extension of the methodology to multiple time series approach, will be addressed in a future paper.

5 Conclusions

The CBS models are a class of nonlinear representation, where dilated and translated versions of cardinal B-spline functions were chosen to be the basis functions (regressors or model terms). As a special class of linear-in-the-parameters representation, the CBS models are easy to train using some standard model term selection algorithms, and the final identified models usually only include a small number of significant model terms. The proposed CBS models provide an efficient representation for short term forecasts of high tides at the Venice lagoon.

Appendix A The forward orthogonal regression algorithm

The CBS models are based on a prescribed prototype function, and temporal analysis is performed using some dilated and translated versions of the same function. Data analysis can thus be implemented using the corresponding coefficients. The initial CBS model (8), where each basis function (model regressor) is a variant of the same cardinal B-spline function, can easily be converted into a linear-in-the-parameters form

$$y(t) = \sum_{m=1}^{M} \theta_m \psi_m(t) + e(t) \tag{A1}$$

where $\boldsymbol{x}(t)=[x_1(t), x_2(t), \cdots, x_d(t)]^T$, with $x_r(t)=y(t-r+1)$ for r=1,2,...,d, is the "input" (predictor) vector, $\psi_m(t)=\psi_m(\boldsymbol{x}(t))$ are the model regressors, θ_m are the model parameters, and M is the total number of candidate regressors.

The initial regression model (A1) often involves a large number of candidate model terms. Experience suggests that most of the candidate model terms can be removed from the model, and that only a small number of significant model terms are needed to provide a satisfactory representation for most nonlinear dynamical systems. The orthogonal least square (OLS) type algorithms (Billings et al., 1989; Chen et al., 1989) interfered with by an error reduction ratio (ERR) index, can be used to select significant model terms, and a Bayesian information criterion (BIC) (Schwarz, 1978; Efron and Tibshirani, 1993), can be used to aid the determination of the associated model size (Wei et al., 2006).

Consider the term selection problem for the linear-in-the-parameters model (A1). Let $\{(\boldsymbol{x}(t), y(t)):\boldsymbol{x}\in\boldsymbol{R}^d, y\in\boldsymbol{R}\}_{t=1}^N$ be a given training data set and $\boldsymbol{y}=[y(1), \cdots, y(N)]^T$ be the vector of the output. Let $I = \{1, 2, \cdots, M\}$, and denote by $\Omega=\{\psi_m:m\in I\}$ the dictionary of candidate model terms in an initially chosen candidate regression model similar to (A1). The dictionary Ω can be used to form a variant vector dictionary $D=\{\boldsymbol{\phi}_m:m\in I\}$, where the mth candidate basis vector $\boldsymbol{\phi}_m$ is formed by the mth candidate model term $\psi_m\in\Omega$, in the sense that $\boldsymbol{\phi}_m=[\psi_m(\boldsymbol{x}(1)), \cdots, \psi_m(\boldsymbol{x}(N))]^T$. The model term selection problem is equivalent to finding, from I, a subset of indices, $I_n=\{i_m:m=1, 2, \cdots, n, i_m\in I\}$ where $n\leq M$, so that $\boldsymbol{y}$ can be approximated using a linear combination of $\boldsymbol{\alpha}_{i_1}, \boldsymbol{\alpha}_{i_2}, \cdots, \boldsymbol{\alpha}_{i_n}$.

A1 The forward orthogonal regression procedure

A non-centralised squared correlation coefficient will be used to measure the dependency between two associated random vectors. The non-centralised squared correlation coefficient between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ of size N is defined as

$$C(\boldsymbol{x}, \boldsymbol{y}) = \frac{(\boldsymbol{x}^T\boldsymbol{y})^2}{||\boldsymbol{x}||^2||\boldsymbol{y}||^2} = \frac{(\boldsymbol{x}^T\boldsymbol{y})^2}{(\boldsymbol{x}^T\boldsymbol{x})(\boldsymbol{y}^T\boldsymbol{y})} = \frac{(\sum_{i=1}^N x_i y_i)^2}{\sum_{i=1}^N x_i^2 \sum_{i=1}^N y_i^2} \tag{A2}$$

The squared correlation coefficient is closely related to the error reduction ratio (ERR) criterion (a very useful index in respect to the significance of model terms), defined in the standard orthogonal least squares (OLS) algorithm for model structure selection (Billings et al., 1989; Chen et al., 1989).

The model structure selection procedure starts from Eq. (A1). Let $\boldsymbol{r}_0=\boldsymbol{y}$, and

$$\ell_1 = \arg \max_{1\leq j\leq M} \{C(\boldsymbol{y}, \boldsymbol{\phi}_j)\} \tag{A3}$$

where the function $C(\cdot, \cdot)$ is the correlation coefficient defined by (A2). The first significant basis can thus be selected as $\boldsymbol{\alpha}_1=\boldsymbol{\phi}_{\ell_1}$, and the first associated orthogonal basis can be chosen as $\boldsymbol{q}_1=\boldsymbol{\phi}_{\ell_1}$. The model residual, related to the first step search, is given as

$$\boldsymbol{r}_1 = \boldsymbol{r}_0 - \frac{\boldsymbol{y}^T\boldsymbol{q}_1}{\boldsymbol{q}_1^T\boldsymbol{q}_1}\boldsymbol{q}_1 \tag{A4}$$

In general, the mth significant model term can be chosen as follows. Assume that at the (m-1)th step, a subset D_{m-1}, consisting of ($m-1$) significant bases, $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{m-1}$, has been determined, and the ($m-1$) selected bases have been transformed into a new group of orthogonal bases $\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_{m-1}$ via some orthogonal transformation. Let

$$\boldsymbol{q}_j^{(m)} = \boldsymbol{\phi}_j - \sum_{k=1}^{m-1} \frac{\boldsymbol{\phi}_j^T\boldsymbol{q}_k}{\boldsymbol{q}_k^T\boldsymbol{q}_k}\boldsymbol{q}_k \tag{A5}$$

$$\ell_m = \arg \max_{j\neq\ell_k, 1\leq k\leq m-1} \{C(\boldsymbol{y}, \boldsymbol{q}_j^{(m)})\} \tag{A6}$$

where $\boldsymbol{\phi}_j\in D-D_{m-1}$, and $\boldsymbol{r}_{m-1}$ is the residual vector obtained in the ($m-1$)th step. The mth significant basis can then be chosen as $\boldsymbol{\alpha}_m=\boldsymbol{\phi}_{\ell_m}$ and the mth associated orthogonal basis can be chosen as $\boldsymbol{q}_m=\boldsymbol{q}_{\ell_m}^{(m)}$. The residual vector $\boldsymbol{r}_m$ at the mth step is given by

$$\boldsymbol{r}_m = \boldsymbol{r}_{m-1} - \frac{\boldsymbol{y}^T\boldsymbol{q}_m}{\boldsymbol{q}_m^T\boldsymbol{q}_m}\boldsymbol{q}_m \tag{A7}$$

Subsequent significant bases can be selected in the same way step by step. From (A7), the vectors $\boldsymbol{r}_m$ and $\boldsymbol{q}_m$ are orthogonal, thus

$$||\boldsymbol{r}_m||^2 = ||\boldsymbol{r}_{m-1}||^2 - \frac{(\boldsymbol{y}^T\boldsymbol{q}_m)^2}{\boldsymbol{q}_m^T\boldsymbol{q}_m} \tag{A8}$$

By respectively summing (A7) and (A8) for m from 1 to n, yields

$$\boldsymbol{y} = \sum_{m=1}^{n} \frac{\boldsymbol{y}^T\boldsymbol{q}_m}{\boldsymbol{q}_m^T\boldsymbol{q}_m}\boldsymbol{q}_m + \boldsymbol{r}_n \tag{A9}$$

$$||\boldsymbol{r}_n||^2 = ||\boldsymbol{y}||^2 - \sum_{m=1}^{n} \frac{(\boldsymbol{y}^T \boldsymbol{q}_m)^2}{\boldsymbol{q}_m^T \boldsymbol{q}_m} \tag{A10}$$

The model residual $\boldsymbol{r}_n$ will be used to form a criterion for model selection, and the search procedure will be terminated when the norm $||\boldsymbol{r}_n||^2$ satisfies some specified conditions. Note that the quantity $ERR_m = C(\boldsymbol{y}, \boldsymbol{q}_m)$ is just equal to the mth error reduction ratio (Billings et al., 1989; Chen et al., 1989), brought by including the mth basis vector $\boldsymbol{\alpha}_m = \boldsymbol{\phi}_{\ell_m}$ into the model, and that $\sum_{m=1}^{n} C(\boldsymbol{y}, \boldsymbol{q}_m)$ is the increment or total percentage that the desired output variance can be explained by $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$.

In the present study, the following Bayesian information criterion (BIC) (Schwarz, 1978; Efron and Tibshirani, 1993) is used to determine the model size

$$\mathrm{BIC}(n) = \frac{N + n[\ln(N) - 1]}{N - n} \mathrm{MSE}(n) \tag{A11}$$

In the present study, the mean-squared-error (MSE) in (A11) is defined as

$$\mathrm{MSE}(n) = \frac{||\boldsymbol{r}_n||^2}{2N} = \frac{1}{2N} \sum_{t=1}^{N} [y(t) - \hat{y}(t)]^2 \tag{A12}$$

where $\hat{y}(t)$ is the model prediction produced from the associated model of n terms. The model size will be chosen as the value where the index function $\mathrm{BIC}(n)$ is minimized. Note that other popular definitions for MSE in (A11) are also available, for example, $\mathrm{MSE}(n) = ||\boldsymbol{r}_n||^2 / N$ (Wei et al. 2006).

A2 Parameter estimation

It is easy to verify that the relationship between the selected original bases $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$, and the associated orthogonal bases $\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_n$, is given by

$$\boldsymbol{A}_n = \boldsymbol{Q}_n \boldsymbol{R}_n \tag{A13}$$

where $\boldsymbol{A}_n = [\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_n]$, $\boldsymbol{Q}_n$ is an $N \times n$ matrix with orthogonal columns $\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_n$, and $\boldsymbol{R}_n$ is an $n \times n$ unit upper triangular matrix whose entries $u_{ij} (1 \le i \le j \le n)$ are calculated during the orthogonalization procedure. The unknown parameter vector, denoted by $\boldsymbol{\theta}_n = [\theta_1, \theta_2, \cdots, \theta_n]^T$, for the model with respect to the original bases, can be calculated from the triangular equation $\boldsymbol{R}_n \boldsymbol{\theta}_n = \boldsymbol{g}_n$ with $\boldsymbol{g}_n = [g_1, g_2, \cdots, g_n]^T$, where $g_k = (\boldsymbol{y}^T \boldsymbol{q}_k)/(\boldsymbol{q}_k^T \boldsymbol{q}_k)$ for $k = 1, 2, \ldots, n$.

Acknowledgements. The authors gratefully acknowledge that this work was supported by EPSRC (UK). They are grateful to J. M. Zaldivar Comenges and C. L. del Arco-Calderón for the constructive comments and helpful suggestions for improving the manuscript.

Edited by: R. Grimshaw
Reviewed by: C. L. del Arco-Calderón and
J. M. Zaldivar Comenges

References

Billings, S. A., Chen, S., and Korenberg, M. J.: Identification of MIMO non-linear systems suing a forward regression orthogonal estimator, Int. J. Control, 49(6), 2157–2189, 1989.

Billings, S. A. and Wei, H. L.: A new class of wavelet networks for nonlinear system identification, IEEE Trans. Neural Networks, 16(4), 862–874, 2005.

Cao, L. Y., Mees, A., and Judd, K.: Dynamics from multivariate time series, Physica D, 121, 75–88, 1998.

Chandre, C., Wiggins, S., and Uzer, T.: Time-frequency analysis of chaotic systems, Physica D, 181, 171–196, 2003.

Chen, S., Billings, S. A., and Luo, W.: Orthogonal least squares methods and their application to non-linear system identification, Int. J. Control, 50(5), 1873–1896, 1989.

Chui, C. K.: An Introduction to Wavelets, Academic Press, Boston, 1992.

del Arco-Calderon, C. L., Vinuela, P. I., and Castro, J. C. H.: Forecasting time series by means of evolutionary algorithms, Lecture Notes in Computer Science, 3242, 1061–1070, 2004.

Efron, B. and Tibshirani, R. J.: An Introduction to the Bootstrap, Chapman & Hall, New York, 1993.

Grinsted, A., Moore, J. C., and Jevrejeva, S. : Application of the cross wavelet transform and wavelet coherence to geophysical time series, Nonlin. Processes Geophys., 11, 561–566, 2004.

Hastie, T. J. and Tibshirani, R. J. : Generalized additive models, London, Chapman & Hall, 1990.

Kallache, M., Rust, H. W., and Kropp, J.: Trend assessment: applications for hydrology and climate research, Nonlin. Processes Geophys., 12, 201–210, 2005.

Kavli, T.: ASMOD – an algorithm for adaptive spline modeling of observation data, Int. J. Control, 58(4), 947–967, 1993.

Kumar, P. and Foufoula-Georgiou, E.: Wavelet analysis for geophysical applications, Rev. Geophys., 35(4), 385–412, 1997.

Malamud, B. D. and Turcotte, D. L.: Self-affine time series I: generation and analyses, Adv. Geophys., 40, 1–90, 1999a.

Malamud, B. D. and Turcotte, D. L.: Self-affine time series: measures of weak and strong persistence, J. Stat. Plan. Infe., 80, 173–196, 1999b.

Maraun, D. and Kurths, J.: Cross wavelet analysis: significance testing and pitfalls, Nonlin. Processes Geophys., 11, 505–514, 2004.

Rosenthal, E.: Venice turns to future to rescue its past, The New York Times Electronic Edition, online: http://www.nytimes.com/2005/02/22/science/22veni.html, 22 February 2005.

Salzano, E.: The Venice Lagoon: what it is, what they are doing on it, Eddyburg, online: http://www.eddyburg.it/article/articleview/1355/1/122, August 2005.

Schwarz, G.: Estimating the dimension of a model, Ann. Stat., 6(2), 461–464, 1978.

Vieira, J., Fons, J., and Cecconi, G.: Statistical and hydrodynamic models for the operational forecasting of floods in the Venice Lagoon, Coastal Engineering, 21(4), 301–331, 1993.

Wei, H. L., Billings, S. A., and Balikhin, M.: Analysis of the geomagnetic activity of the D-st index and self-affine fractals using wavelet transforms, Nonlin. Processes Geophys., 11, 303–312, 2004a.

Wei, H. L., Billings, S. A., and Balikhin, M.: Prediction of the *Dst* index using multiresolution wavelet models, J. Geophys. Res, 109(A7), A07212, doi:10.1029/2003JA010332, 2004b.

Wei, H. L. and Billings, S. A.: Identification and reconstruction of chaotic systems using multiresolution wavelet decompositions, Int. J. Syst. Sci., 35(9), 511–526, 2004.

Wei, H. L. and Billings, S. A.: Long term prediction of nonlinear time series using multiresolution wavelet models, Int. J. Control, 79(6), 569–580, 2006.

Wei, H. L., Billings, S. A., and Balikhin, M. A. :Wavelet based nonparametric NARX models for nonlinear input-output system identification, Int. J. Syst. Sci., accepted, to appear in 37(15), 2006.

Zaldivar, J. M., Gutiérrez, E., Galván, I. M., Strozzi, F., and Tomasin, A.: Forecasting high waters at Venice Lagoon using chaotic time series analysis and nonlinear neural networks, J. Hydroinformatics, 2(1), 61–84, 2000.