An introduction to models based on Laguerre, Kautz and other related orthonormal functions – Part II: non-linear models

Gustavo H.C. Oliveira*

Department of Electrical Engineering,
Federal University of Paraná (UFPR),
80215-901, Curitiba-PR, Brazil
E-mail: gustavo@eletrica.ufpr.br
*Corresponding author

Alex da Rosa

Department of Electrical Engineering,
University of Brasília (UnB),
70910-900, Brasília-DF, Brazil
E-mail: alex@ene.unb.br

Ricardo J.G.B. Campello

Department of Computer Sciences,
University of São Paulo (USP),
13560-970, São Carlos-SP, Brazil
E-mail: campello@icmc.usp.br

Jeremias B. Machado

Engineering and Information Technology Institute,
Federal University of Itajubá (UNIFEI),
37500-903, Itajubá -MG, Brazil
E-mail: jeremias@unifei.edu.br

Wagner C. Amaral

School of Electrical and Computer Engineering,
University of Campinas (UNICAMP),
13083-852, Campinas-SP, Brazil
E-mail: wagnerg@dca.fee.unicamp.br

Abstract: This paper provides an overview of system identification using orthonormal basis function models, such as those based on Laguerre, Kautz, and generalised orthonormal basis functions. The paper is separated in two parts. The first part of the paper approached issues related with linear models and models with uncertain parameters. Now, the mathematical foundations as well as their advantages and limitations are discussed within the contexts of non-linear system identification. The discussions comprise a broad bibliographical survey of the subject and a comparative analysis involving some specific model realisations, namely, Volterra, fuzzy, and neural models within the orthonormal basis functions framework. Theoretical and practical issues regarding the identification of these non-linear models are presented and illustrated by means of two case studies.

Keywords: modelling; non-linear system identification; robust identification; Volterra models; orthonormal basis functions; fuzzy models; neural networks.

Reference to this paper should be made as follows: Oliveira, G.H.C., da Rosa, A., Campello, R.J.G.B., Machado, J.B. and Amaral, W.C. (2012) 'An introduction to models based on Laguerre, Kautz and other related orthonormal functions – Part II: non-linear models', *Int. J. Modelling, Identification and Control*, Vol. 16, No. 1, pp.1–14.

Biographical notes: Gustavo H.C. Oliveira received his BSc in Electrical Engineering from the Federal University of Juiz de Fora, Brazil in 1988, MSc in Electrical Engineering from the School of Electrical and Computer Engineering of the State University of Campinas in 1992, and PhD in Electrical Engineering from University of Nice – Sophia Antipolis, France, and from State University of Campinas, Brazil in 1997. From 1999 to 2009, he was with the Mechatronic Engineering Department of the Pontifical Catholic University of Paran, Brazil and since 2009, he is with the Electrical Engineering Department of the Federal University of Paran. His research interests include the areas of system identification, adaptive and predictive control, HVAC applications and electrical power systems modelling and simulation.

Alex da Rosa received his BSc in Electrical Engineering from the Federal University of Goiás (UFG) in 2002, and MSc and PhD in Electrical Engineering from the School of Electrical and Computer Engineering of the State University of Campinas (UNICAMP) in 2005 and 2009, respectively. He is currently with the Department of Electrical Engineering of the University of Brasília. His research interests include the areas of modelling and identification of non-linear dynamic systems.

Ricardo J.G.B. Campello received his BSc in Electronics Engineering from the State University of São Paulo (Unesp), Ilha Solteira – SP in 1994, and MSc and PhD in Electrical Engineering from the School of Electrical and Computer Engineering of the State University of Campinas (Unicamp), Campinas – SP in 1997 and 2002, respectively. In 2002, he was a visiting scholar at the Laboratoire D'Informatique, Signaux et Systèmes de Sophia Antipolis (I3S), Université de Nice – Sophia Antipolis (UNSA), France. Since 2007, he is with the Department of Computer Sciences of the University of São Paulo (USP) at São Carlos. His current research interests include dynamic systems identification, computational intelligence, data mining, and machine learning.

Jeremias B. Machado received his BSc in Electrical Engineering from the Federal University of Itajubá (UNIFEI) in 2005, and MSc and PhD in Electrical Engineering from the State University of Campinas (UNICAMP) in 2007 and 2011, respectively. Since 2011, he has been with the Systems Engineering and Information Technology Institute, Federal University of Itajubá (UNIFEI). His research interests include artificial intelligence, modelling, identification and control of non-linear dynamic systems.

Wagner C. Amaral received his BSc, MSc and Doctorate in Electrical Engineering from the State University of Campinas, UNICAMP in 1974, 1976 and 1981, respectively. Since 1991, he has been a Full Professor at the Department of Computer Engineering and Industrial Automation at UNICAMP. From 1995 to 1999, he was the Director of the School of Electrical and Computer Engineering at UNICAMP. His research interests are modelling, identification and predictive control.

1 Introduction

The dynamic of several systems can be well reproduced by using linear models. However, when linear models are not able to represent, with quality, the dynamic of the system under study, it becomes necessary to use non-linear models. These models are usually constructed using the well-known non-linear auto-regressive with exogenous (NARX) inputs structure, in which the system output in a given discrete-time instant is represented in terms of past input and output (I/O) values (Sjöberg et al., 1995; Nelles, 2001), as in polynomial models (Leontaritis and Billings, 1985). As discussed in the first part of this paper, NARX autoregressive aspect generally increases the sensitivity of the model quality regarding the choice of the model order. Non-linear finite impulse response (NFIR) models are models where the system output in a given discrete-time instant is represented only in terms of past samples of the input (Sjöberg et al., 1995; Nelles, 2001). The absence of output recursion in such NFIR models, however, has the drawback of usually requiring a large number of terms in the regression vector, especially when representing slow dynamics. This relation is exponential for most general-purpose non-linear models, as will be seen in subsequent sections. Models without output feedback that circumvent this drawback are the so-called non-linear orthonormal basis functions (NOBF) models (Sjöberg et al., 1995; Nelles, 2001).

This paper is a continuation of the Part I, where this issue is tackled for linear models and models with uncertain parameters. Now, an overview of the state of the art in the identification of dynamic systems using Non-linear OBF models is presented.

The outline of this paper is as follows. Section 2 discusses briefly the problem of identifying linear dynamic systems using an OBF-based framework. Section 3 presents the identification of non-linear systems, providing different possible realisations of NOBF-based models. Section 4 compares the approaches reported in the previous sections. Section 5 discusses the design of OBF that parameterise the corresponding models. Section 6 describes a case study involving an isothermic polymerisation process. Finally, Section 7 addresses the conclusions.

2 Orthonormal basis functions

The use of orthonormal filters to represent signals and systems has a long history, since the pioneering proposals of Takenaka (1925) and Wiener (1958). Discrete-time OBF can be generated by cascading different all-pass filters of order one or two, as follows (Ninness and Gustafsson, 1997; Heuberger et al., 2005):

$$\Psi_i(z) = \frac{z\sqrt{1-|\beta_i|^2}}{z-\beta_i}\prod_{j=1}^{i-1}\left(\frac{1-\overline{\beta}_j z}{z-\beta_j}\right), \quad i = 1, 2, \ldots, \tag{1}$$

where β_i are stable poles of the orthonormal basis ($\beta_i \in \mathbb{C}$: $|\beta_i| < 1$) and $\overline{\beta}_i$ denotes the complex conjugate of β_i. The functions in (1) are the so-called Takenaka-Malmquist functions (Heuberger et al., 2005). The corresponding realisations in the time-domain, $\psi_i(k)$, are given by the inverse Z-transform of (1) and satisfy the orthonormality property. The set $\{\psi_i\}$ is complete on $\ell^2[0, \infty)$ if and only if $\sum_{i=1}^{\infty}\left(1-|\beta_i|\right) = \infty$ (Ninness and Gustafsson, 1997; Heuberger et al., 2005), so any finite energy signal (including absolutely summable functions) can be approximated with any prescribed accuracy by linearly combining a certain finite number of such functions. In general, functions $\psi_i(k)$ are complex-valued, although this is physically unrealistic in system identification problems. It is shown (Ninness and Gustafsson, 1997) that this drawback can be circumvented by constructing a modified basis of functions with real-valued impulse responses consisting of a linear combination of the complex-valued functions generated by (1).

The most popular orthonormal bases are the Laguerre and the Kautz ones (Broome, 1965; Heuberger et al., 2005). Laguerre and Kautz bases are preferred when modelling systems with first- and second-order dominant dynamics, respectively. Systems with more complex dominant dynamics are better represented using models based on GOBF because the mathematical description of such bases involves multiple poles (modes). Although the term generalised orthonormal basis functions is originally due to the formula by Van den Hof et al. (1995) and Heuberger et al. (1995), this term will be used generically hereafter to refer to orthonormal bases of functions with multiple modes.

The OBF are recursive, which means that the i^{th} function can be written in terms of the $(i-1)^{\text{th}}$ one. It is then possible to describe the dynamics of the set of orthonormal functions using a state-space representation. In this case, the OFB linear model can be represented as follows (Oliveira et al., 2011):

$$\mathrm{l}(k+1) = A\mathrm{l}(k) + \mathbf{b}\, u(k), \tag{2}$$

$$\hat{y}(k) = \mathcal{H}\left(\mathrm{l}(k)\right), \tag{3}$$

where the state vector $\mathrm{l}(k) = [l_1(k) \ldots l_n(k)]^T$ is composed by the outputs of the orthonormal filters and $\mathcal{H}$ is the static mapping given by the linear combination of these states, that is:

$$\mathcal{H}\left(\mathrm{l}(k)\right) = \sum_{i=1}^{n} c_i l_i(k). \tag{4}$$

The matrices A and $\mathbf{b}$ in (2) depend solely upon the orthonormal basis. A description of these matrices is found in the literature for Laguerre and Kautz, for instance (Dumont and Fu, 1993).

Once the basis pole is chosen (see Section 5), the model becomes completely determined by the coefficients $c_{(\cdot)}$ of the orthonormal series in (4). If the impulse response of the system is available, these coefficients can be computed analytically. Although this nonparametric approach is simple and mathematically well founded, it may not, however, be effective in practical problems since the impulse response of the system is usually not available. A more efficient approach involves considering the coefficients c_i as parameters to be estimated numerically using I/O data from the system, which can be performed in a simple way using, for example, a least-squares algorithm (Ljung, 1999).

Finally, it is worth remarking that, although no transport delay is explicitly represented in the previous model description, the orthonormal functions are able to represent dynamics with this characteristic (Mäkilä, 1990; Fu and Dumont, 1993). Nevertheless, any information available about the real delay of the system can be explicitly incorporated into the model, which allows reducing the number of functions and filters necessary for modelling the system with any prescribed accuracy. This feature can be obtained by replacing $u(k)$ with $u(k - \tau_d)$ in (2), where τ_d is the estimated delay.

3 Non-linear system identification

The main idea of the OBF-based linear models presented in Section 2, where the static mapping $\mathcal{H}$ in (3) is given by a linear combination of its arguments, is the expansion of the impulse response of the system on an orthonormal basis. One question that arises is whether the linear mapping $\mathcal{H}$ can be substituted by a non-linear mapping so as to obtain a model capable of also representing non-linear dynamics. In this case, the model would be described by linear dynamics that relate the input $u(k)$ to the orthonormal states l_i(i) followed by a non-linear static mapping relating these states and the output $\hat{y}(k)$, that is, a Wiener-type model (Rugh, 1981; Nelles, 2001; Campello and Oliveira, 2007). This section discusses non-linear implementations for the operator $\mathcal{H}$. Such implementations are of particular interest in this context because, among other reasons, they can be interpreted from a mathematical viewpoint.

The shape of the static mapping ($\mathcal{H}$) provides a specific realisation of a model with a (N)OBF structure. All the non-linear realisations considered in this paper can be described such that this mapping is represented in a formulation that is linear in the parameters, as follows:

$$\mathcal{H}(\mathrm{l}(k)) = \lambda(k)^T \zeta, \tag{5}$$

where $\zeta \in \mathbb{R}^{\mu\times 1}$ is the parameter vector to be estimated and $\lambda(k) \in \mathbb{R}^{\mu\times 1}$ is a regression vector that depends solely on the orthonormal states $\mathrm{l}(k)$. The linear model, for example, can be represented using this formulation by setting $\mu = n$, $\zeta = [c_1 \cdots c_n]^T$, and $\lambda(k) = \mathrm{l}(k)$. Alternatively, one can insert an additional term c_0 capable of representing a non-null constant level in the system output by simply setting $\zeta = [c_0\ c_1\ \cdots\ c_n]^T$ and $\lambda(k) = [1\ \mathrm{l}(k)^T]^T$. In this case, the model is called affine and the number of elements of the parameter vector will be $\mu = n + 1$.

3.1 Volterra models

Discrete-time Volterra models assume that the system admits the following polynomial description of order N (Schetzen, 1980; Rugh, 1981; Doyle et al., 2002):

$$y(k) = \sum_{\eta=1}^{N} \sum_{k_1=0}^{\epsilon_\eta} \cdots \sum_{k_\eta=0}^{\epsilon_\eta} h_\eta\left(k_1,\ldots,k_\eta\right) \prod_{j=1}^{\eta} u\left(k-k_j\right), \tag{6}$$

where $u(k)$, $y(k)$, and $h_\eta(k_1, \ldots, k_\eta)$ are the input, the output, and the η^{th}-order kernel, respectively. Moreover, ϵ_η is a truncation limit beyond which the kernel h_η is assumed to be null in each dimension. This representation is a straightforward generalisation of order N from the linear model FIR. In this context, the kernel h_η is an η^{th}-order generalisation of the unit-impulse response function of the FIR model. It is also a specific realisation of the input-output functional $y(k) = \mathcal{G}\left(\{u(\tau)\}_{\tau=-\infty}^{k}\right)$, where $\mathcal{G}$ is a generic non-linear operator. In Boyd and Chua (1985), it is shown that truncated Volterra models, such as those in (6), can approximate with desired accuracy any non-linear system that meets the following requirements:

1. the operator $\mathcal{G}$ is causal, continuous, time-invariant, and has fading memory
2. the input $u(k)$ is (upper and lower) bounded.

These requirements comprise a wide class of real-world systems.

The estimation of Volterra kernels for modelling non-linear systems has been investigated for decades (Eykhoff, 1974; Billings, 1980; Schetzen, 1980; Rugh, 1981; Doyle et al., 2002). The main drawback is that the kernels are, in principle, non-parameterised functions whose measurement is possible only if their individual contributions can be separated from the total system response (Schetzen, 1980). A straightforward approach can be derived if the elements of the Volterra kernels, that is, the coefficients $h_\eta(k_1, \ldots, k_\eta)$, are treated as individual parameters to be estimated. In this case, the Volterra model (6) is linear in these parameters and classical estimation algorithms such as recursive least squares can then be applied. This approach, however, usually makes the model over-parameterised. It is therefore important to reduce its parametric complexity before estimation to improve numerical conditioning while decreasing the variance of the estimated parameters. An interesting approach to dealing with this problem, suggested by Wiener (1958), is to expand the Volterra kernels using OBF (Schetzen, 1980; Doyle et al., 2002). This approach is discussed below.

3.1.1 OBF-Volterra models

The kernels h_η in (6) are assumed to be such that $h_\eta(k_1, \ldots, k_\eta) = 0$ for $k_j > \epsilon_\eta$ ($\forall_j \in \{1, \ldots, \eta\}$), which means that they are absolutely summable on $\ell^2[0, \infty)$ and thus stable. Therefore, the kernels h_η can be represented by means of OBF. For the sake of simplicity and without any loss of generality, the following developments will be presented assuming that all Volterra kernels are expanded using the same orthonormal basis $\{\psi_m\}$. In this case, the η-dimensional expansion of the η^{th}-order kernel is as follows (Schetzen, 1980; Rugh, 1981):

$$h_\eta\left(k_1,\ldots,k_\eta\right) = \sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} c_{i_1,\ldots,i_\eta} \prod_{j=1}^{\eta} \psi_{i_j}\left(k_j\right), \tag{7}$$

where $\psi_i(k)$ is the i^{th} orthonormal function of the basis and $c_{i_1,\ldots,i_\eta}$ are the expansion coefficients given by:

$$c_{i_1,\ldots,i_\eta} = \sum_{k_1=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} h_\eta\left(k_1,\ldots,k_\eta\right) \prod_{j=1}^{\eta} \psi_{i_j}\left(k_j\right). \tag{8}$$

Note that, for the first-order kernel ($\eta = 1$), the previous expansion is equivalent to the expansion of the unit-impulse response of the convolution linear model as presented in Oliveira et al. (to appear).

From equations (6) and (7), provided that $h_\eta(k_1, \ldots, k_\eta) = 0$ for $k_j > \epsilon_\eta$, the Volterra model can be rewritten as:

$$y(k) = \sum_{\eta=1}^{N} \sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} c_{i_1,\ldots,i_\eta} \prod_{j=1}^{\eta} l_{i_j}(k), \tag{9}$$

where l_i is the output of the i^{th} orthonormal filter (i.e., the i^{th} state) given by $l_i(k) = \sum_{\tau=0}^{\infty} \psi_i(\tau) u(k-\tau)$. For practical reasons, second-order models ($N = 2$) are usually adopted in both academic and real-world problems (Billings, 1980;

Dumont and Fu, 1993). Furthermore, equation (7) is, in practice, approximated with a finite number n_η of functions. For instance, considering a Volterra model in which the first- and second-order kernels are truncated using n_1 and n_2 orthonormal functions, respectively, the model is rewritten as:

$$\hat{y}(k) = c_0 + \sum_{i_1=1}^{n_1} c_{i_1} l_{i_1}(k) + \sum_{i_1=1}^{n_2} \sum_{i_2=1}^{i_1} c_{i_1,i_2} l_{i_1}(k) l_{i_2}(k), \tag{10}$$

where c_0 is an additional zeroth-order coefficient inserted only to represent a non-null DC level in the system output. Note that, for computational calculations, the coefficients $c_{(\cdot)}$ can be considered as parameters to be estimated numerically. Hence, since any pair of coefficients $c_{i,j}$ and $c_{j,i}$ multiply the same factor $l_i(k)l_j(k)$ in the second-order term, both coefficients can be represented by a single parameter to be estimated, as suggested in equation (10).

Then, considering the coefficients $c_{(\cdot)}$ as parameters to be estimated, the model (9) can be rewritten as in equations (3) and (5), where the vectors $\lambda(k)$ and ζ in (5) are defined as:

$$\zeta = \big[c_0 \;\; c_1 \cdots c_{n_1} \;\; c_{1,1} \;\; c_{2,1} \;\; c_{2,2} \cdots \;\; c_{n_2,1} \;\; c_{n_2,2} \cdots c_{n_2,n_2}\big]^T, \tag{11}$$

$$\lambda(k) = \big[1 \;\; l_1(k) \cdots l_{n_1}(k) \;\; l_1(k)^2 \;\; l_2(k)l_1(k) \;\; l_2(k)^2 \cdots l_{n_2}(k)l_1(k) \;\; l_{n_2}(k)l_2(k) \cdots \;\; l_{n_2}(k)^2\big]^T, \tag{12}$$

The order of the state-space representation in (2) is $n = \max\{n_1, n_2\}$. In this case, the number of model parameters is given by $\mu = \left(n_2^2 + n_2 + 2n_1 + 2\right)/2$.

3.2 TS fuzzy models

The Takagi-Sugeno (TS) fuzzy models are established by a set of M rules as follows (Takagi and Sugeno, 1985; Yager and Filev, 1994; Babuška, 1998):

$$\begin{aligned} R^i &: \text{IF } x_1 \text{ is } X_1^i \text{ AND } \ldots \text{ AND } x_n \text{ is } X_n^i \\ &\quad \text{THEN } y_i(k) = f_i(x_1, \ldots, x_n), \end{aligned} \tag{13}$$

where R^i is the i^{th} fuzzy rule, $x_j \in \mathbf{X}_j \subset \mathbb{R}$ ($j = 1, \ldots, n$) are the input variables (premise variables), $y_i \in \mathbf{Y} \subset \mathbb{R}$ is the output variable, f_i ($i = 1, \ldots, M$) are the functions that relate the inputs with the output in the model, and X_j^i are the fuzzy sets defined on the universe of discourse $\mathbf{X}_j$ of the respective variables, that is, $X_j^i : \mathbf{X}_j \to [0,1]$.

The inference of the output value $\hat{y}$ from a specific set of input values is calculated as the weighted mean of the individual outputs of each rule, as follows:

$$\hat{y} = \frac{\sum_{i=1}^{M} w_i f_i(x_1, \ldots, x_n)}{\sum_{i=1}^{M} w_i}, \tag{14}$$

where w_i is the firing weight of the i^{th} rule, given by

$$w_i = X_1^i(x_1) X_2^i(x_2) \cdots X_n^i(x_n), \tag{15}$$

Sugeno and his co-workers (Takagi and Sugeno, 1985; Sugeno and Kang, 1986, 1988; Sugeno and Tanaka, 1991) originally proposed the utilisation of affine functions in the rule consequents, that is:

$$f_i(x_1, \ldots, x_n) = \theta_0^i + \sum_{j=1}^{n} \theta_j^i x_j, \tag{16}$$

This choice allows a simple mathematical interpretation of the model as an interpolation of different local affine models and implies that the output in equation (14) is linear on the parameters $\theta_0^i, \ldots, \theta_n^i$ $(i = 1, \ldots, M)$. Hence, these parameters can be estimated using linear estimation algorithms.

Dynamic TS fuzzy models contain exactly the same formulation described previously, except that the local models are dynamic instead of static models. In principle, the local models can hold any structure relating to their dynamic topology. For linear local models with a FIR structure, for example, one can just redefine the premise variables as: $x_1 = u(k-1)$, …, $x_n = u(k-n)$, where $u(k)$ is the input of the dynamic system at the instant k.

3.2.1 OBF-TS fuzzy models

Dynamic NOBF models with static non-linear mapping $\mathcal{H}$ given by fuzzy models were introduced in Oliveira et al. (1999). It can be shown that the proposal in Oliveira et al. (1999) is a particular case of a more general framework given by a TS fuzzy model with linear local models following the OBF dynamic topology (Nelles, 2001; Campello, 2002). Particularly, the inputs of the TS model are given by the outputs of the orthonormal filters, that is, $x_1 = l_1(k)$, …, $x_n = l_n(k)$, which is equivalent to implementing the operator $\mathcal{H}$ in (3) and (5) through a TS model. The model rules take on the following form[1]:

$$\begin{aligned} R^i &: \text{IF } l_1(k) \text{ is } X_1^i \text{ AND } \ldots \text{ AND } l_n(k) \text{ is } X_n^i \\ &\quad \text{THEN } y(k) = \theta_0^i + \theta_1^i l_1(k) + \cdots + \theta_n^i l_n(k), \end{aligned} \tag{17}$$

and the output can be rewritten from (14) as

$$\hat{y}(k) = \frac{\sum_{i=1}^{M} w_i \left(\theta_0^i + \theta_1^i l_1(k) + \cdots + \theta_n^i l_n(k)\right),}{\sum_{i=1}^{M} w_i}, \tag{18}$$

where w_i is rewritten from (15) as

$$w_i = \prod_{j=1}^{\eta} X_j^i\left(l_j(k)\right), \tag{19}$$

The OBF-TS model presented can be easily rewritten in a general form given by equations (3) and (5). For this, one has to define the vectors ζ and $\lambda(k)$ in (5) as:

$$\zeta = \left[\theta_0^1 \;\; \theta_1^1 \;\cdots\; \theta_n^1 \;\cdots\; \theta_0^M \;\; \theta_1^M \;\cdots\; \theta_n^M\right]^T, \tag{20}$$

$$\begin{aligned}\lambda(k) = \gamma(k)\big[w_1 \;\; w_1 l_1(k) \;\cdots\; w_1 l_n(k) \;\cdots \\ w_M \;\; w_M l_1(k) \;\cdots\; w_M l_n(k)\big]^T,\end{aligned} \tag{21}$$

where $\gamma(k) = 1\Big/\sum_{i=1}^{M} w_i$ is the normalisation term in (18). In this case, assuming that fuzzy sets in the rules are predefined [e.g., by fuzzy clustering approaches (Babuška, 1998)], the number of the model parameters is given by $\mu = M(n + 1)$.

Two main questions arise from mapping $\mathcal{H}$ through TS fuzzy models: the mathematical interpretation of the model and its representational power. For the TS fuzzy model described previously, the mathematical interpretation is clear: it is an interpolation of M different linear (affine) OBF models that share the same dynamics of states (e.g., Laguerre functions). The generalised OBF-TS model, a more general model where the local OBF have independent state representations, that is, each model can have a set of orthonormal filters parameterised in different basis functions, has been proposed (Campello, 2002; Campello and Amaral, 2002a). This structure is more flexible in the sense of making complex non-linear dynamic representations possible through a smaller set of parameters. However, the problem of parameterising the orthonormal functions become more complex. Such issue is discussed in details in Section 5.

Concerning the representational capability of the OBF-TS models, there are results (Campello, 2002; Campello et al., 2004) whose starting point is the capability of universal approximation of the fuzzy models (Wang and Mendel, 1992; Zeng and Singh, 1994, 1995; Kosko, 1997) and the representation capability of the Volterra models (Boyd and Chua, 1985) (as discussed in Section 3.1). These results show that, if the input is bounded in a closed interval, the OBF-TS models can approximate with arbitrary precision any causal non-linear dynamic system in discrete time that enables an input/output representation through a continuous operator and with fading memory. It is important to mention that the bounded input hypothesis is essential for the utilisation of any fuzzy or neural model, independently of its dynamic configuration, because this hypothesis is necessary to guarantee a compact input domain for these models. This hypothesis is not restrictive in the engineering context either.

3.3 ANNs models

Artificial neural networks (ANNs) are mathematical models based on the human nervous system structure, composed by simple processing units (artificial neurons) interconnected by a large number of connections (Haykin, 1999). There is a large number of publications on ANN, such as Saraswati and Chand (2009), and this subject is briefly presented here; for more details on this approach, see the classical reference by Rumelhart et al. (1986) or Haykin (1999) and refer to Narendra and Parthasarathy (1990) or Su and McAvoy (1993) for more application-oriented introduction. Different classes of ANN have been presented in the literature. In this work the class of interest is the feedforward networks which implement static maps $\mathcal{H} : \mathbb{R}^m \to \mathbb{R}^n$ between vector spaces of arbitrary dimensions m and n. The two feedforward architectures most widespread are the radial basis functions (RBFs) and the multi-layer perceptron (MLP) networks (Haykin, 1999). In this work there is a special interest in RBF, which are described below.

The basic architecture of an RBF network (Chen et al., 1992; Pottmann and Seborg, 1992; Pearson, 1999) is given by a weighted sum of M RBFs $h_i(\mathbf{x})$ ($i = 1, \ldots, M$) of the input vector $\mathbf{x} = [x_1 \cdots x_m]^T$. One of the most usual function used in the RBF network is the Gaussian function, that is:

$$h_i(\mathbf{x}) = \exp\left(-\left(\mathbf{x}-\mathbf{c}_i\right)^T \Lambda_i^{-1}\left(\mathbf{x}-\mathbf{c}_i\right)^T\right) \tag{22}$$

where $\mathbf{c}_i$ is the vector with the coordinates of the centre of the i^{th} function and Λ_i is a positive-definite matrix $n \times n$ whose eigenvalues represent the variances of the i^{th} function along their characteristic directions (directions of their eigenvalues). If Λ_i is diagonal, then the diagonal elements represent the variances of the function over each axial direction (Nelles, 2001).

3.3.1 OBF RBF model

Let us consider that the consequent of each rule (local model) in (17) is reduced to its constant term only, that is, $y(k) = \theta_0^i$. In this case, the vectors in (20) and (21) can be rewritten as

$$\zeta = \left[\theta_0^1 \;\cdots\; \theta_0^M\right]^T, \tag{23}$$

$$\lambda(k) = \gamma(k)\left[w_1 \;\cdots\; w_M\right]^T. \tag{24}$$

Given, further, that the fuzzy sets in the rules are Gaussian, one has a special case of the TS model that corresponds[2] precisely to the formulation of a particular type of neural network, the RBFs network (Broomhead and Lowe, 1988; Haykin, 1999). This same equivalence is verified for fuzzy models with simplified relational structures (Campello, 2002). More details about the equivalence among different classes of fuzzy models and neural networks can be found in the literature (Hunt et al., 1996; Cho and Wang, 1996).

Since the RBF networks are universal approximators (Haykin, 1999), the approximation capability of the

OBF-RBF model is the same as that of the OBF-TS models, although the latter usually require a smaller amount of parameters to provide the same accuracy. The OBF-RBF models also present a mathematical interpretation that is a particular case of the OBFTS model. On the other hand, OBF neural models proposed in other works (Back and Tsoi, 1996; Sentoni et al., 1996, 1998; Balestrino et al., 1999; Alataris et al., 2000; Vázquez and Agamennoni, 2001; Arto et al., 2001; Campello et al., 2003) are based on the use of other network architectures (e.g., multi layer perceptron), that is, an arbitrary implementation of the static mapping $\mathcal{H}$ through a neural or neuro-fuzzy network.

4 Comparing the (N)OBF approaches

Four alternatives have been mentioned so far for the implementation of the static mapping $\mathcal{H}$ preceding the dynamics of states composed of a set of orthonormal filters in (N)OBF models:

1 linear (or affine) combinations

2 multidimensional polynomials (Volterra)

3 fuzzy TS models

4 neural networks.

Basically, the first approach is limited to the representation of linear dynamic systems, with or without parametric uncertainties; in this latter case, robust identification algorithms are required.

In regard to the non-linear approaches, it is important to highlight some significant structural differences among them. In the Volterra case, for example, the non-linearity of the static mapping is polynomial, so the only way to improve the representational capability of the model is to increase the order of the polynomial. The complexity of the Volterra models, however, depends exponentially on the polynomial order, which usually restricts their application to second-order representations of systems with mild non-linearities (Billings, 1980; Dumont and Fu, 1993). This drawback does not occur with realisations performed by means of universal approximators, such as fuzzy models and neural networks, whose mappings can theoretically take on arbitrary continuous forms over a compact domain. In such cases, although the number of neurons or fuzzy rules needed to obtain a given capability of representation grows exponentially with domain dimension – the curse of dimensionality (Kosko, 1992) – increasing the accuracy of the model becomes possible by means of parametric adjustment of the fuzzy sets or membership functions. On the other hand, this flexibility requires the use of more sophisticated estimation procedures. Although both fuzzy and neural approaches present the same capability of representation, the former has the advantage of mathematical interpretation that, in general, the latter does not.

5 Design of OBF

An overview on the design of linear OBF models is presented in Oliveira et al. (2011).

6 Simulation examples

In this section, the system identification using the techniques described in Sections 3 and 5 is illustrated through two case studies, related with the computational simulation of a continuous stirred tank reactor (CSTR) polymerisation process and the modelling of a levitation magnetic system using an OBF-Volterra model.

6.1 *Example 1: non-linear case*

The identification of a simulated CSTR polymerisation is considered, particularly an isothermal process that uses toluene as solvent (Doyle et al., 1995). The number average molecular weight (NAMW) of the resulting polymer, $y(t)$ (kg/kmol), is controlled by manipulating the rate of flow of the substance initiator, $u(t)$ [m^3/h]. A four-state model of this non-linear process is given by Maner et al. (1996) and Doyle et al. (2002):

$$\begin{cases} \dot{x}_1(t) = 10\left(6 - x_1(t)\right) - 2.4568 x_1(t)\sqrt{x_2(t)} \\ \dot{x}_2(t) = 80u(t) - 10.1022 x_2(t) \\ \dot{x}_3(t) = 0.00241 x_1(t)\sqrt{x_2(t)} + 0.11219 x_2(t) - 10 x_3(t) \\ \dot{x}_4(t) = 245.978 x_1(t)\sqrt{x_2(t)} - 10 x_4(t) \\ y(t) = \dfrac{x_4(t)}{x_3(t)} \end{cases}, \quad (25)$$

The differential equations in (25) are simulated using as initial conditions the nominal operation conditions, given by $x_1(0) = 5.506774$, $x_2(0) = 0.132906$, $x_3(0) = 0.0019752$, $x_4(0) = 49.3818$, $u(0) = 0.016783$, and $y(0) = 25000.5$ (Doyle et al., 1995; Maner et al., 1996). The process is simulated from $t = 0$ up to $t = 32$ hours, with the input u being a sequence of steps of the same duration and random amplitude uniformly distributed within the operational interval [–0.6, 0.8]. The dataset is sampled using a sampling period of $T = 0.03$ hours and normalised in order to avoid numerical problems during the model estimation. This dataset is then split into two parts: one half (16 hours) intended for estimation of the (N)OBF models of this process and the other half intended for validation of the resulting model. The first 1/2 hour of each set is used exclusively to recover the right Laguerre states. These states are unknown (and set equal to zero) at $t = 0$ because, in practice, the values of the input signal anterior to the data available for identification are usually unknown.

This section addresses the identification problem of the CSTR polymerisation process using the different models with NOBF structure presented in Section 3. The models obtained are compared in terms of their accuracy and parametric complexity. They are also compared with OBF linear models in order to emphasise the difference in performance when using the non-linear models in the identification of systems with that characteristic.

The Laguerre pole p has been chosen experimentally based on preliminary tests using an OBF-Volterra model with $n_1 = n_2 = 3$ orthonormal functions in the first-and second-order terms. For the sake of simplicity, the linear search procedure described in Oliveira et al. (2003) has been adopted, where the feasibility interval (–1, 1) has been discretised (sampling of 0.05) and therefore a different model is estimated for each value of the resulting set. The pole providing the model with the lowest error of prediction for data validation is $p = 0.75$. After selecting the pole, various OBF linear and Volterra models containing different numbers of Laguerre functions are estimated via least squares and then validated using the I/O data available. The generation of such data is discussed at the beginning of Section 6. A comparison regarding the mean squared error (MSE) between the measured system output and the output of each model is presented in Tables 1 and 2.

Table 1 Simulation performance of OBF linear models for normalised validation data

No. lag. funct. n	*No. paramet. μ*	*MSE*
1	2	0.00325429
2	3	0.00145855
3	4	0.00145919

Table 2 Simulation performance of second-order OBF-Volterra models for normalised validation data

No. lag. funct.		*No. paramet. μ*	*MSE*
n_1	n_2		
1	1	3	0.00217362
2	1	4	0.00041545
3	1	5	0.00041052
2	2	6	0.00014829
3	2	7	0.00014383
3	3	10	0.00014230

As can be seen in Table 1, the performance of the linear model does not improve when using more than two functions, partially because of the rigorous selection of the Laguerre pole but mainly due to the model's inability to represent the non-linear dynamics of the process. This hypothesis is verified graphically in Figure 1, which shows the actual output measured from the system for data validation jointly with the output predicted by the best obtained model ($n = 2$). It is observed in Figure 1 that the model cannot represent the gains of the process.

Comparing Tables 1 and 2, it becomes clear that the inclusion of the second-order term significantly improves the model's performance. Indeed, comparing models with similar numbers of parameters, particularly the linear model with $n = 2$ or 3 and the Volterra model with $n_1 = 2$ and $n_2 = 1$, one observes that the former model results in an error which is three times the latter. Furthermore, for the same number of Laguerre functions, specifically two or three, the error associated with the Volterra model is approximately ten times lower. The improvement obtained by the non-linear model can also be seen graphically by comparing Figures 1 and 2. The latter shows the output measured from the system for data validation jointly with the output predicted by the best obtained Volterra model ($n_1 = n_2 = 3$).

Figure 1 Output measured from the CSTR (solid line) and predicted output of the linear Laguerre model with $n = 2$ (dotted line) for normalised validation data

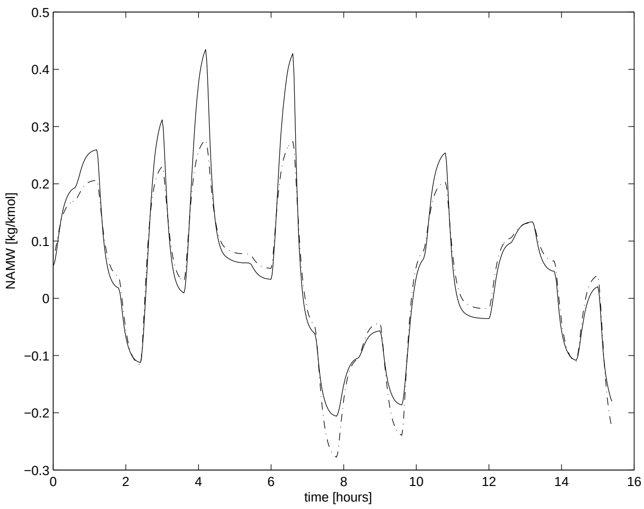
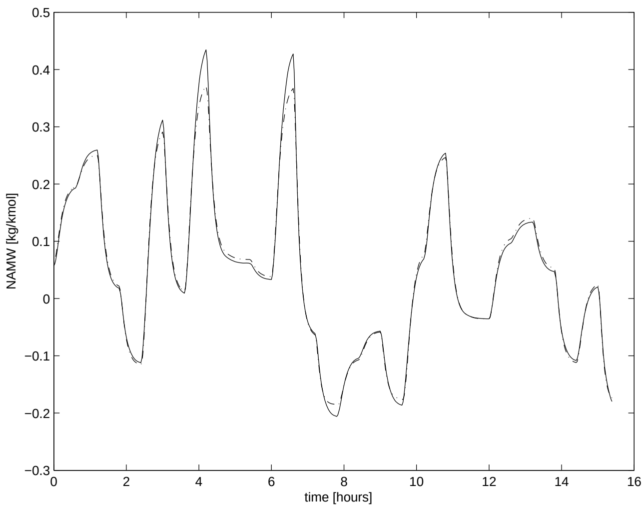


Figure 2 Output measured from the CSTR (solid line) and predicted output of the Laguerre-Volterra model with $n_1 = n_2 = 3$ (dotted line) for normalised validation data



In the case of the OBF-TS and OBF-RBF models, the design parameters (in addition to the already selected pole $p = 0.75$) are the following: the number n of Laguerre functions, the number M of fuzzy rules, the fuzzy sets of the premise variables, and the parameter vector containing the coefficients of the local models (affine or constant). For the sake of simplicity here, let us select a complete fuzzy rule basis associated with a partition in the grid of the input domain (Passino and Yurkovich, 1997). In this approach, a number of reference fuzzy sets are associated with each premise variable (Laguerre states). The corresponding rule basis is called complete because it contains all the possible rules with regard to the combinations of these referential sets. In this case, if α is the number of referential fuzzy sets assigned to each premise variable, the number of rules in the model is $M = \alpha^n$. Let us adopt, also for simplicity, a homogeneous distribution of Gaussian fuzzy sets on the universe of discourse of the corresponding variable, that is, equidistant centres and widths equal to the distance between two consecutive centres. Finally, select $\alpha = 2$ to avoid a large amount of parameters – $\mu = \alpha^n(n + 1)$ in (20) and $\mu = \alpha^n$ in (23) – to be estimated via least squares. This value is sufficient to provide satisfactory results in this application, as will be seen below. It is important to mention that, before, the number of fuzzy sets per premise variable, the shape of these sets, and the rule numbers on the rule basis possibly incomplete, could all be optimised through different strategies, leading to more parsimonious and accurate models (Babuška, 1998; Espinosa et al., 2004). In this case, however, the additional effort undertaken at this stage needs to be considered in the comparative analysis with other models, which is beyond the scope of this work.

From the parametrisation described previously, one now defines the number n of Laguerre functions and then estimates the coefficients of the local models (affine for the TS model and constant for the RBF model) using a least squares algorithm. Models having different numbers of Laguerre functions are estimated and validated using the set of I/O data available. The results regarding the MSE between the measured system output and the output of each model are shown in Tables 3 and 4.

Table 3 Simulation performance of OBF-RBF models for normalised validation data

No. lag. funct. n	*No. paramet. μ*	*MSE*
1	2	0.00375778
2	4	0.00076868
3	8	0.00036230
4	16	0.00038336

Table 4 Simulation performance of OBF-TS models for normalised validation data

No. lag. funct. n	*No. paramet. μ*	*MSE*
1	4	0.00213717
2	12	0.00000520
3	32	0.00000095

Comparing Tables 2 and 3 shows that the RBF model results in a performance similar to that obtained by the Volterra model, but quantitatively worse (to some extent, this is justified by the absence of adjustment of the pre-fixed fuzzy sets). The TS model, in turn, has more flexible local models, thus allowing one to achieve significantly higher accuracy, as shown in Table 4. The improvement obtained with the TS model can also be seen graphically by comparing Figures 3 and 4, which show the measured system output (validation data) jointly with the predicted output of the RBF ($n = 3$) and TS ($n = 2$) models, respectively. The TS model with $n = 2$ is selected because it represents a better trade-off between parsimony and accuracy when compared to that with $n = 3$.

Figure 3 Output measured from the CSTR (solid line) and predicted output of the Laguerre-RBF model with $n = 3$ (dotted line) for normalised validation data

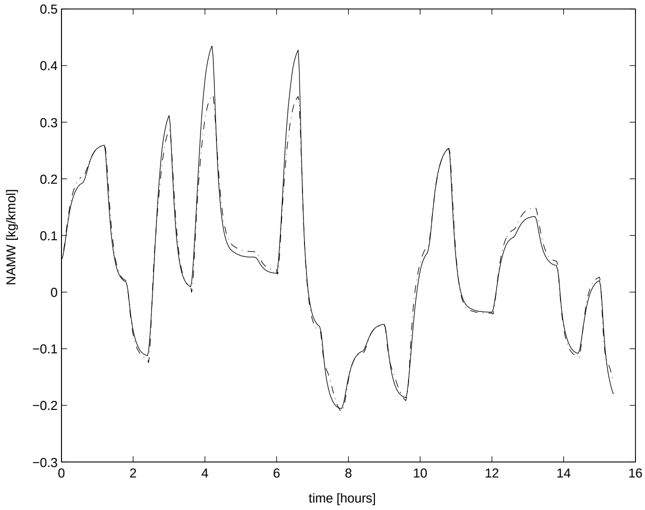
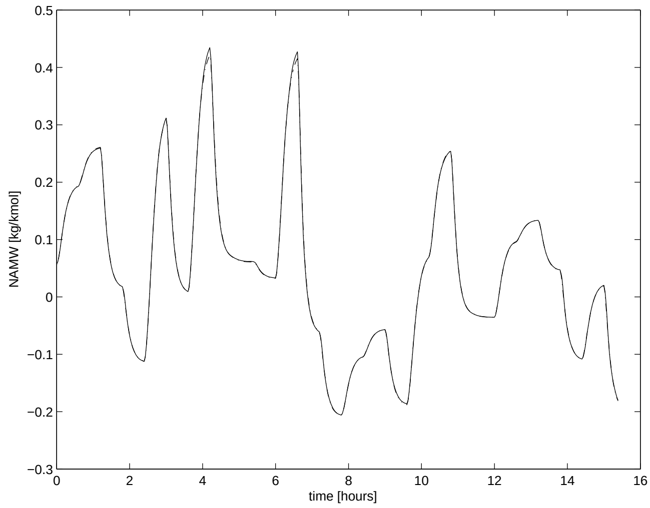


Figure 4 Output measured from the CSTR (solid line) and predicted output of the Laguerre-TS model with $n = 2$ (dotted line) for normalised validation data



Although the exponential growth of the number of parameters due to the number of Laguerre functions for both RBF and TS models is explicit, this drawback is more pronounced for the latter model, as shown in Tables 3 and 4. In order to lessen such a drawback, one approach – an alternative to those already mentioned previously and which does not require abandoning the simplicity of the architecture of the complete rules adopted – consists of using only a subset of the variables in local models as variables in the premise. Specifically for the OBF-TS models, whose rules are given in (17), this means taking the n Laguerre states in the consequent of the rules and then using only the subset of the first $r < n$ states in the antecedents (premise). In other words, it means using only one subset of Laguerre states as decision variables on the operational region of the system, which can result in no loss of performance, depending on the complexity of the dynamics involved. For $r = 1$, for example, the rules are as follows:

$$R^i : \text{IF } l_1(k) \text{ is } X_1^i \quad \text{THEN } y(k) = \theta_0^i + \theta_1^i l_1(k) + \cdots + \theta_n^i l_n(k). \tag{26}$$

For the sake of comparison, Table 5 shows the simulation results for the TS model with the simplified rules in (26). It can be observed, for example, that although the accuracy is significantly lower in this particular case, the model with $n = 2$ represents a trade-off between accuracy and parsimony very close to that obtained by the Volterra model.

Table 5 Simulation results of OBF-TS models with simplified rules (for normalised validation data)

No. lag. funct. n	*No. paramet. μ*	*MSE*
1	4	0.00213717
2	6	0.00015712
3	8	0.00015397

Recent works have presented solutions for the determination of OBF-TS models through genetic algorithms or fuzzy clustering techniques (Medeiros et al., 2006; Machado, 2007). Strategies for automatically determining the optimal number of states in the premise of the rules in OBF-TS models have been investigated (Medeiros et al., 2006) which studied the use of non-linear local models in the consequent of the rules and the efficient estimation of the parameters of these models through local least squares algorithms. A non-stochastic approach is proposed (Machado, 2007) that uses a mixture of clustering validity criteria to automatically determine the number of local models and the membership functions based on product-space fuzzy clustering of I/O data through the well-known fuzzy clustering algorithm by Gustafson and Kessel (1979), which computes a fuzzy partition of the data into fuzzy hyperellipsoidal clusters. The fuzzy sets of each rule can then be obtained by projecting the corresponding fuzzy cluster onto the unidimensional domains of the premise variables (OBF states in the current work) (Babuška and Verbruggen, 1997). Simulations results employing Volterra models are presented in da Rosa et al. (2009).

6.2 Modelling a magnetic levitation system

The magnetic levitation system considered here, schematically shown in Figure 5, consists of upper and lower drive coils that produce a magnetic field in response to a DC current. Two magnets travel along a precision ground glass guide rod. By energising the lower coil, one of the magnets is levitated by a repulsive magnetic force. As current in the coil increases, the field strength also increases and the height of the levitated magnet is also increased. Two laser-based sensors measure the magnet positions. The magnets are of an ultra-high-strength rare earth (NeBFe) type and are designed to provide large levitated displacements (ECP, 1999).

Let m_1 be the mass of the lower magnet, v_1 the viscous friction coefficient between this magnet and air, and g the acceleration of gravity. The movement of the lower magnet (Magnet 1) is governed by $m_1\ddot{y}_1 + v_1\dot{y}_1 + F_{m12} = F_{u11} - F_{u21} - m_1 g$, where y_1 is the position of Magnet 1, F_{u11} is the magnetic force from Coil 1 interacting with Magnet 1, F_{u21} is the magnetic force from Coil 2 interacting with Magnet 1, and F_{m12} is the mutual magnetic force between the two magnets. These forces are described by the following non-linear equations:

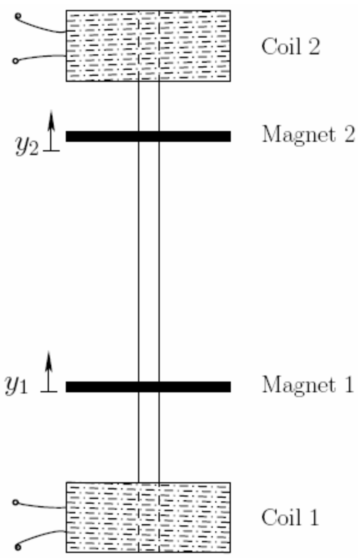
$$F_{m12} = \frac{c}{\left(y_c + y_2 - y_1 + d\right)^4},$$

$$F_{u11} = \frac{i_1}{a\left(k_s y_1 + b\right)^4},$$

$$F_{u21} = \frac{i_2}{a\left(y_c + k_s y_1 + b\right)^4},$$

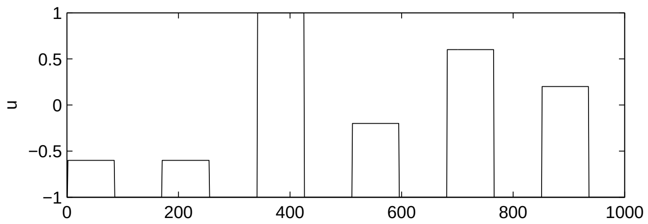
where y_2 is the position of Magnet 2, y_c is the distance between Coils 1 and 2, i_1 and i_2 are the currents through Coils 1 and 2, respectively, and a, b, c, and d are real-valued constants. Some results for modelling a laboratory-scale plant of the magnetic levitation system described above are presented below.

Figure 5 Sketch of the magnetic levitation system

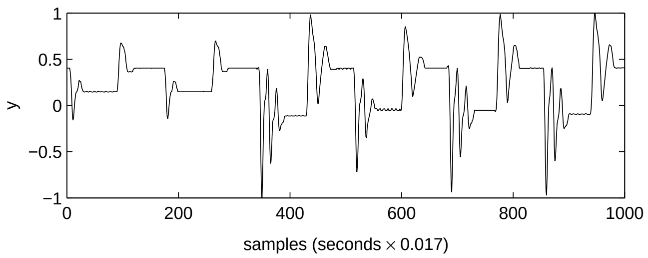


Experimental data have been acquired by keeping constant the DC current applied to Coil 1 of the plant while varying the current through Coil 2. The input signal u (current through Coil 2) has been designed as a sequence of steps with different amplitudes so as to excite different modes of the system. The measured output signal y is taken as the position of Magnet 1 (y_1). Figure 6 shows the I/O data available for estimation of the model. Before estimation, these data are sampled using a sampling period of 0.017 second and are normalised within the interval [–1, 1] to avoid numerical problems. Another similar yet independent set of data has also been acquired and is reserved for further model validation.

Figure 6 (a) Current through Coil 2 (input signal u) (b) Position of Magnet 1 (output signal y)



(a)



(b)

The model adopted here relates the input $u(k)$ and output $y(k)$ in Figure 6 by means of a second-order ($N = 2$) Volterra representation with a symmetric second-order kernel, as is usual in the literature (Billings, 1980; Dumont and Fu, 1993). By describing its first- and second-order terms as expansions on two independent GOBF bases, the model becomes similar to that in (10), i.e.:

$$\hat{y}(k) = \sum_{i_1=1}^{n_1} c_{i_1} l_{1,i_1}(k) + \sum_{i_1=1}^{n_2} \sum_{i_2=1}^{i_1} c_{i_1,i_2} l_{2,i_1}(k) l_{2,i_2}(k), \quad (27)$$

where $l_i(k)$ denotes the result of filtering the input signal $u(k)$ by the orthonormal function with impulse response $\psi_i(k)$, that is, $l_i(k) = \psi_i(q)\, u(k)$.

In model (27), only the I/O signals are known. In this case, no prior information on the model kernels is available. Therefore it is necessary to estimate them from input-output data measured from the levitation system in order to obtain the optimal GOBF poles. This model uses $n_1 = n_2 = 4$ functions which are independent for each kernel order, thus covering two pairs of complex conjugate poles. The parameters of the orthonormal basis (see Oliveira et al., to appear) are set as $\gamma' = \gamma'' = \mu' = \mu'' = 0,$ although their choice has no significant effect on the identification procedure (Ziaei and Wang, 2006). This example has also been simulated by setting $\lambda' = \lambda'' = \rho' = \rho'' = 1$ and similar results were obtained.

The parameters of model (27) are optimised using a numerical approach found in the literature (da Rosa et al., 2009), which consists of simultaneously optimising the pole vector **p** that parameterises the orthonormal basis $\{\psi_i\}$ and the corresponding coefficient vector ζ in (11). This approach involves the analytical computation of gradients that can be used as part of an optimisation method to obtain exact search directions for the OBF poles.

To model the magnetic levitation system using this method, the initial poles are chosen as $\beta_{1,1}^0 = \beta_{1,2}^0 = \beta_{2,1}^0 = \beta_{2,1}^0 = 0.5 + i0.5.$ The final values of the poles are shown in Table 6.

The model of the magnetic levitation system can now be compared to the actual system output with regard to the data samples reserved for model validation. Figure 7 displays the corresponding model output, $\hat{y}(k)$, with the actual output measured from the system, $y(k)$. It can be seen that a nearly exact approximation of this highly non-linear system is obtained. The corresponding approximation error is illustrated in Figure 8.

Table 6 Estimated GOBF poles for the magnetic levitation system

Kernel order (η)	*Optimal poles*
1	$\beta_{1,1} = 0.7592 \pm i0.2042$
	$\beta_{1,2} = 0.7382 \pm i0.3713$
2	$\beta_{2,1} = 0.8698 \pm i0.3242$
	$\beta_{2,2} = 0.7215 \pm i0.3947$

Figure 7 Actual system output (solid line) and predicted output (dotted line) of the model with optimised Kautz poles

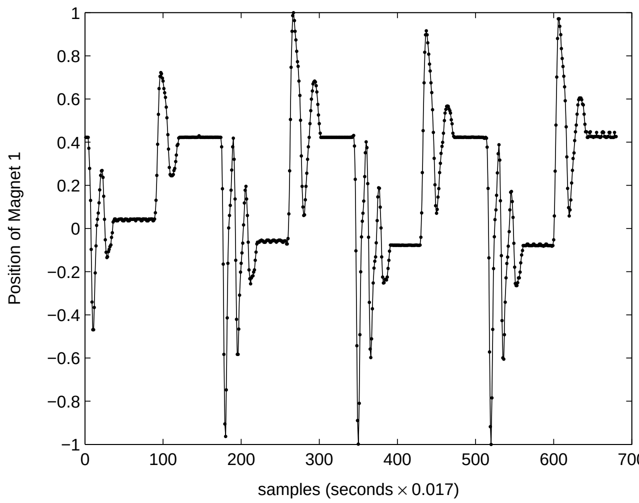
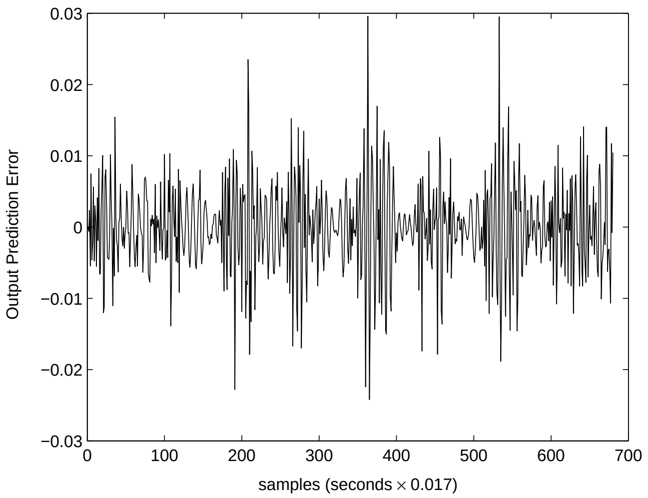


Figure 8 Predicted error between the system output $y(k)$ and the model output $\hat{y}(k)$ for the magnetic levitation system



Finally, Table 7 presents a comparison between different approaches found in the literature that have been used for modelling the magnetic levitation system. All of these cases shown in Table 7 have been simulated using the same number of basis functions (four). In this comparison, the normalised quadratic error (NQE) is defined as:

$$\text{NQE} \triangleq 10 \log \frac{\sum_{k=1}^{N_d} [y(k) - \hat{y}(k)]^2}{\sum_{k=1}^{N_d} [y(k)]^2} \tag{28}$$

where $N_d = 680$ is the number of available I/O data samples reserved for model validation.

Although these models have the same number of parameters (basis functions), it is worth noting that the error obtained with the GOBF model is lower than that obtained with the Kautz models.

Table 7 Comparison showing the errors obtained using different methods for modelling the magnetic levitation system

Method	*NQE (in dB)*
Kautz basis (da Rosa, 2005)	–17.4
multiple Kautz bases (da Rosa et al., 2008)	–22.3
with Kautz basis (da Rosa et al., 2009)	–26.5
with GOBF (da Rosa et al., 2009)	–33.2

7 Conclusions

An overview of the state of the art in the area of identification of non-linear dynamic systems using OBF models has been presented. This so-called NOBF structure presents many advantages when compared to dynamic structures with traditional regressors, such as the absence of output recursion, the orthonormality of the elements of the regression vector, the ability to deal with time delays, tolerance of unmodelled dynamics and others.

The mathematical foundations of the NOBF models have been discussed in the context of system identification based on linear models and non-linear models. In the non-linear case, different approaches have been considered, such as Volterra, fuzzy, and neural models. These approaches have been discussed comparatively in terms of their representational capability, parsimony, complexity of design, and interpretability. In sum, all these approaches provide the same capability of representation for a wide class of non-linear dynamic systems; however, the parametric complexity of these models is greater when compared to the OBF linear models. In the case of Volterra models, for example, the non-linearity is polynomial, so the only way to improve the model's representational capability is to increase the polynomial order, which exponentially increases the number of parameters to be estimated. The fuzzy and neural approaches, in turn, provide more flexibility because they allow one to improve the accuracy of the model by increasing the number of fuzzy sets or activation functions, as well as by means of parametrically adjusting these elements. On the other hand, this flexibility requires the use of more complex estimation procedures. Unlike the linear, Volterra, and fuzzy approaches, in general the neural approach does not present a clear mathematical interpretation.

This paper has presented a widespread bibliographical compilation about the identification of NOBF models, including works comprising all the approaches mentioned previously. Practical aspects of modelling have also been presented and illustrated through a case study involving CSTR polymerisation.

Extensions of the ideas surveyed in this paper towards adaptive, stochastic, and frequency-domain model formulations, for instance, are current subjects of active research by the authors and many other researchers all over the world.

Acknowledgements

The authors would like to acknowledge the financial support from the Brazilian National Council for Scientific and Technological Development (CNPq), from the Research Foundation of the State of São Paulo (Fapesp) and from the Research Foundation of the State of Paraná (Fundação Araucária).

References

Alataris, K., Berger, T.W. and Marmarelis, V.Z. (2000) 'A novel network for nonlinear modeling of neural systems with arbitrary point-process inputs', *Neural Networks*, Vol. 13, No. 2, pp.255–266.

Arto, V., Hannu, P. and Halme, A. (2001) 'Modeling of chromatographic separation process with Wiener-MLP representation', *Journal of Process Control*, Vol. 11, No. 5, pp.443–458.

Babuška, R. (1998) *Fuzzy Modeling for Control*, Kluwer Academic Publishers, Massachusetts, USA.

Babuška, R. and Verbruggen, H.B. (1997) 'Fuzzy set methods for local modelling and identification', in Murray-Smith, R. and Johansen, T.A. (Eds.): *Multiple Model Approaches to Modelling and Control*, Taylor and Francis, London, Chap. 2.

Back, A.D. and Tsoi, A.C. (1996) 'Nonlinear system identification using discrete Laguerre functions', *Journal of Systems Engineering*, Vol. 6, No. 3, pp.194–207.

Balestrino, A., Caiti, A. and Zanobini, G. (1999) 'Identification of Wiener-type nonlinear systems by Laguerre filters and neural networks', *Proc. 14th IFAC World Congress*, Beijing/China, pp.433–438.

Billings, S.A. (1980) 'Identification of nonlinear systems – a survey', *IEE Proc. Pt D*, Vol. 127, No. 6, pp.272–285.

Boyd, S. and Chua, L.O. (1985) 'Fading memory and the problem of approximating nonlinear operators with Volterra series', *IEEE Trans. on Circuits and Systems*, Vol. 32, No. 11, pp.1150–1161.

Broome, P.W. (1965) 'Discrete orthonormal sequences', *Journal of the Association for Computing Machinery*, Vol. 12, No. 2, pp.151–168.

Broomhead, D.S. and Lowe, D. (1988) 'Multivariate functional interpolation and adaptive networks', *Complex Systems*, Vol. 2, No. 3, pp.321–355.

Campello, R.J.G.B. (2002) 'New architectures and methodologies for modeling and control of complex systems combining classical and modern tools', PhD thesis, School of Electrical and Computer Engineering of the State University of Campinas (FEEC/UNICAMP), Campinas-SP, Brazil (in Portuguese).

Campello, R.J.G.B. and Amaral, W.C. (2002a) 'Takagi-Sugeno fuzzy models within orthonormal basis function framework and their application to process control', *Proc. 11th IEEE Internat. Conference on Fuzzy Systems*, Honolulu/USA, pp.1399–1404.

Campello, R.J.G.B. and Oliveira, G.H.C. (2007) 'Modelos não lineares', in L.A. Aguirre, A.P. Alves da Silva, M.F.M. Campos and W.C. Amaral (Eds.): *Enciclopédia de Automática*, Vol. 3 (Cap. 4), Edgard Blücher (in Portuguese).

Campello, R.J.G.B., Meleiro, L.A.C. and Amaral, W.C. (2004) 'Control of a bioprocess using orthonormal basis function fuzzy models', *Proc. 13th IEEE Internat. Conference on Fuzzy Systems*, Budapest/Hungary, pp.801–806.

Campello, R.J.G.B., Von Zuben, F.J., Amaral, W.C., Meleiro, L.A.C. and Filho, R.M. (2003) 'Hierarchical fuzzy models within the framework of orthonormal basis functions and their application to bioprocess control', *Chemical Engineering Science*, Vol. 58, No. 18, pp.4259–4270.

Chen, S., Billings, S. and Grant, P. (1992) 'Recursive hybrid algorithm for non-linear system identification using radial basis function networks', *Int. J. Control*, Vol. 55, No. 5, pp.1051–1070.

Cho, K.B. and Wang, B.H. (1996) 'Radial basis function based adaptive fuzzy systems and their applications to system identification and prediction', *Fuzzy Sets and Systems*, Vol. 83, No. 3, pp.325–339.

da Rosa, A. (2005) 'Expansion of discrete-time Volterra models using Kautz functions', MSc thesis, School of Electrical and Computer Engineering of the State University of Campinas (FEEC/UNICAMP), Campinas- SP, Brazil (in Portuguese).

da Rosa, A., Campello, R.J.G.B. and Amaral, W.C. (2008) 'An optimal expansion of Volterra models using independent Kautz bases for each kernel dimension', *International Journal of Control*, Vol. 81, No. 6, pp.962–975.

da Rosa, A., Campello, R.J.G.B. and Amaral, W.C. (2009) 'Exact search directions for optimization of linear and nonlinear models based on generalized orthonormal functions', *IEEE Transactions on Automatic Control*, Vol. 54, No. 12, pp.2757–2772.

Doyle, F.J., III, Ogunnaike, B.A. and Pearson, R.K. (1995) 'Nonlinear model-based control using second-order Volterra models', *Automatica*, Vol. 31, No. 5, pp.697–714.

Doyle, F.J., III, Pearson, R.K. and Ogunnaike, B.A. (2002) *Identification and Control Using Volterra Models*, Springer-Verlag, London, UK.

Dumont, G.A. and Fu, Y. (1993) 'Non-linear adaptive control via Laguerre expansion of Volterra kernels', *Int. J. Adaptive Control and Signal Processing*, Vol. 7, No. 5, pp.367–382.

Educational Control Products (ECP) (1999) *Manual for Model 730 – Magnetic Levitation System*, California, USA.

Espinosa, J., Vandewalle, J. and Wertz, V. (2004) *Fuzzy Logic, Identification and Predictive Control*, Springer-Verlag, London, UK.

Eykhoff, P. (1974) *System Identification: Parameter and State Estimation*, John Wiley & Sons, UK.

Fu, Y. and Dumont, G.A. (1993) 'An optimum time scale for discrete Laguerre network', *IEEE Trans. on Automatic Control*, Vol. 38, No. 6, pp.934–938.

Gustafson, D.E. and Kessel, W.C. (1979) 'Fuzzy clustering with a fuzzy covariance matrix', *Proc. IEEE CDC*, San Diego, CA, pp.761–766.

Haykin, S. (1999) *Neural Networks: A Comprehensive Foundation*, 2nd ed., Prentice Hall, Upper Saddle River, NJ, USA.

Heuberger, P.S.C., Van den Hof, P.M.J. and Bosgra, O.H. (1995) 'A generalized orthonormal basis for linear dynamical systems', *IEEE Trans. on Automatic Control*, Vol. 40, No. 3, pp.451–465.

Heuberger, P.S.C., Van den Hof, P.M.J. and Wahlberg, B. (2005) *Modelling and Identification with Rational Orthogonal Basis Functions*, Springer-Verlag, London, UK.

Hunt, K.J., Haas, R. and Murray-Smith, R. (1996) 'Extending the functional equivalence of radial basis function networks and fuzzy inference systems', *IEEE Trans. Neural Networks*, Vol. 7, No. 3, pp.776–781.

Kosko, B. (1992) *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*, Prentice Hall, Upper Saddle River, NJ, USA.

Kosko, B. (1997) *Fuzzy Engineering*, Prentice Hall, Upper Saddle River, NJ, USA.

Leontaritis, I. and Billings, S.A. (1985) 'Input-output parametric models for nonlinear systems – Parts I and II', *Int. Journal of Control*, Vol. 41, No. 6, pp.303–344.

Ljung, L. (1999) *System Identification: Theory for the User*, 2nd ed., Prentice Hall, Upper Saddle River, NJ, USA.

Machado, J.B. (2007) 'Design of OBF-TS fuzzy models based on multiple clustering validity criteria', *Proc. IEEE Int. Conf. on Tools with Artificial Intelligence*, Patras/Greece, No. 2, pp.336–339.

Mäkilä, P.M. (1990) 'Approximation of stable systems by Laguerre filters', *Automatica*, Vol. 26, No. 2, pp.333–345.

Maner, B.R., Doyle, F.J. III, Ogunnaike, B.A. and Pearson, R.K. (1996) 'Nonlinear model predictive control of a simulated multivariable polymerization reactor using second-order Volterra models', *Automatica*, Vol. 32, No. 9, pp.1285–1301.

Medeiros, A.V., Amaral, W.C. and Campello, R.J.G.B. (2006) 'GA optimization of generalized OBF-TS fuzzy models with global and local estimation approaches', *Proc. 15th IEEE Internat. Conference on Fuzzy Systems*, Vancouver/Canada, pp.8494–8501.

Narendra, K. and Parthasarathy, K. (1990) 'Identification and control of dynamical systems using neural networks', *IEEE Trans. Neural Networks*, Vol. 1, No. 1, pp.4–26.

Nelles, O. (2001) *Nonlinear System Identification*, Springer-Verlag, London, UK.

Ninness, B. and Gustafsson, F. (1997) 'A unifying construction of orthonormal bases for system identification', *IEEE Trans. on Automatic Control*, Vol. 42, No. 4, pp.515–521.

Oliveira, G.H.C., Amaral, W.C. and Latawiec, K. (2003) 'CRHPC using Volterra models and orthonormal basis functions: an application to CSTR plants', *Proc. IEEE Conference on Control Applications*, Istanbul/Turkey, pp.718–723.

Oliveira, G.H.C., Campello, R.J.G.B. and Amaral, W.C. (1999) 'Fuzzy models within orthonormal basis function framework', *Proc. 8th IEEE Internat. Conference on Fuzzy Systems*, Seoul/Korea, pp.957–962.

Oliveira, G.H.C., da Rosa, A., Campello, R.J.G.B., Machado, J.M. and Amaral, W.C. (2011) 'An introduction to models based on Laguerre, Kautz and other related orthonormal functions – Part I: linear and uncertain models', *International Journal of Modelling, Identification and Control*, Vol. 14, Nos. 1/2.

Passino, K.M. and Yurkovich, S. (1997) *Fuzzy Control*, Addison-Wesley Longman Inc., USA.

Pearson, R.K. (1999) *Discrete-Time Dynamic Models*, Oxford University Press, New York, USA.

Pottmann, M. and Seborg, D. (1992) 'Identification of non-linear process using reciprocal multiquadric functions', *Journal of Process Control*, Vol. 2, No. 4, pp.189–203.

Rugh, W.J. (1981) *Nonlinear System Theory: The Volterra/Wiener Approach*, The Johns Hopkins University Press, Baltimore, USA.

Rumelhart, D., McClelland, J. and PDP Research Group (1986) *Parallel Distributed Processing*, Vol. 1, MIT Press, Cambridge, MA, USA.

Saraswati, S. and Chand, S. (2009) 'Neural network models for multi-step ahead prediction of air-fuel ratio in SI engines', *International Journal of Modelling, Identification and Control*, Vol. 7, No. 3, pp.263–274.

Schetzen, M. (1980) *The Volterra and Wiener Theories of Nonlinear Systems*, Krieger Publishing Company, Malabar, Florida, USA.

Sentoni, G., Agamennoni, O., Desages, A. and Romagnoli, J. (1996) 'Approximate models for nonlinear process control', *AIChE Journal*, Vol. 42, No. 8, pp.2240–2250.

Sentoni, G.B., Biegler, L.T., Guiver, J.B. and Zhao, H. (1998) 'State-space nonlinear process modeling: identification and universality', *AIChE Journal*, Vol. 44, No. 10, pp.2229–2239.

Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P-Y., Hjalmarsson, H. and Juditsky, A. (1995) 'Nonlinear black-box modeling in system identification: a unified overview', *Automatica*, Vol. 31, No. 12, pp.1691–1724.

Su, H-T. and McAvoy, T. (1993) 'Integration of multilayer perceptron networks and linear dynamic models: a Hammerstein modeling approach', *Ind. Eng. Chem. Res.*, Vol. 32, No. 9, pp.1927–1936.

Sugeno, M. and Kang, G.T. (1986) 'Fuzzy modelling and control of multilayer incinerator', *Fuzzy Sets and Systems*, Vol. 18, No. 3, pp.329–346.

Sugeno, M. and Kang, G.T. (1988) 'Structure identification of fuzzy model', *Fuzzy Sets and Systems*, Vol. 28, No. 1, pp.15–33.

Sugeno, M. and Tanaka, K. (1991) 'Successive identification of a fuzzy model and its applications to prediction of a complex system', *Fuzzy Sets and Systems*, Vol. 42, No. 3, pp.315–334.

Takagi, T. and Sugeno, M. (1985) 'Fuzzy identification of systems and its applications to modeling and control', *IEEE Trans. Systems, Man and Cybernetics*, Vol. SMC-15, pp.116–132.

Takenaka, S. (1925) 'On the orthogonal functions and a new formula of interpolation', *Japanese Journal of Mathematics*, Vol. II, pp.129–145.

Van den Hof, P.M.J., Heuberger, P.S.C. and Bokor, J. (1995) 'System identification with generalized orthonormal basis functions', *Automatica*, Vol. 31, No. 12, pp.1821–1834.

Vázquez, M.A. and Agamennoni, O.E. (2001) 'Approximate models for nonlinear dynamical systems and their generalization properties', *Mathematical and Computer Modelling*, Vol. 33, Nos. 8–9, pp.965–986.

Wang, L-X. and Mendel, J.M. (1992) 'Fuzzy basis functions, universal approximation and orthogonal least squares learning', *IEEE Trans. Neural Networks*, Vol. 3, No. 5, pp.807–814.

Wiener, N. (1958) *Nonlinear Problems in Random Theory*, MIT Press, Cambridge, MA, USA.

Yager, R.R. and Filev, D.P. (1994) *Essentials of Fuzzy Modeling and Control*, John Wiley & Sons, USA.

Zeng, X-J. and Singh, M.G. (1994) 'Approximation theory of fuzzy systems – SISO case', *IEEE Trans. Fuzzy Systems*, Vol. 2, No. 2, pp.162–176.

Zeng, X-J. and Singh, M.G. (1995) 'Approximation theory of fuzzy systems – MIMO case', *IEEE Trans. Fuzzy Systems*, Vol. 3, No. 2, pp.219–235.

Ziaei, K. and Wang, D.W.L. (2006) 'Application of orthonormal basis functions for identification of flexible-link manipulators, *Control Engineering Practice*, Vol. 14, No. 2, pp.99–106.

Notes

1 For the sake of simplicity, it is assumed that the local models in (17) are linear (affine) and each one has the same order n, but both hypotheses can be relaxed in a more general case.

2 Except for the normalization term $\gamma(k)$, which does not change the analysis, because it can be incorporated into the parameter vector to be estimated.