Time series prediction by Perturbed Fuzzy Model

Paulo Salgado
UTAD, Vila Real, Portugal
psal@utad.pt

Getúlio Igrejas
IPB, Bragança, Portugal
igrejas@ipb.pt

Fernando Gouveia
UTAD, Vila Real, Portugal
fgouveia@utad.pt

Abstract

This paper presents a fuzzy system approach to the prediction of nonlinear time series and dynamical systems based on a fuzzy model that includes its derivative information. The underlying mechanism governing the time series, expressed as a set of IF–THEN rules, is discovered by a modified structure of fuzzy system in order to capture the temporal series and its temporal derivative information. The task of predicting the future is carried out by a fuzzy predictor on the basis of the extracted rules and by the Taylor ODE solver method. We have applied the approach to the benchmark Mackey-Glass chaotic time series.

Keywords: Derivative approximation, Times Series, Fuzzy modelling.

1 Introduction

Time series is widely observed in many aspects of our lives. Daily temperature, stock market and so forth are examples of time series. A time series is a continuous $x(t)$ or a discrete sequence of measured quantities $x_1, x_2, \cdots, x_n$ taken from human activity data or some physical system. Basically, there are three main goals in time series analysis: prediction, modelling and characterization [6]. The goal of prediction is to accurately forecast the short-term evolution of the system, modelling aims to precisely capture the features of the system's long-term behaviour, and the purpose of system characterization is to determine some underlying fundamental properties of the system and the nature of observations. Forecasting refers to a process by which the future behaviour of a dynamical system or data series is estimated based on our understanding and characterization of the system and the laws of its nature.

Much effort has been devoted over the past several decades to develop and improve the time-series forecasting models. This task is most complex and hard due to multiple reasons, such as: high sensibility to initial conditions in the unstable dynamical system [4]; difficulty on the determination of trends or on the recognition of patterns in presence of stochastic noise on observable sequence data; or just the natural uncertainty, vagueness and incompleteness of data. However, in the absence of something better, there are some statistical or empirical solutions to make reasonable predictions [11]. Most of these linear approaches, such as the well-known Box–Jenkins method, have shortcomings [3][12]: all have a natural default in that they lack the ability to directly incorporate the natural linguistic information in their modelling or in their strategies, even to extract relevant linguistic information from the data series.

More recently, neural networks and fuzzy logic modelling have been applied to the problem of forecasting complex time series. The main advantage of these methodologies is that we do not need to specify a priori the structure of a model, which is clearly needed in the classical regression analysis [14][11]. Also, both models are nonlinear in nature and they can more easily approximate complex nonlinear systems than simple linear statistical models. Of course, there are also disadvantages to statistical regressions models, where we can use the information given by their parameters to understand the process. This problem can be greatly reduced if alternatively the relationships of the process are expressed by linguistic relationships, which are transparent and easily read by an expert, through the use of a fuzzy modelling.

Fuzzy Systems (FS) have been successfully applied to a number of scientific and engineering fields in recent years, but their performance is highly dependent on their structure (hierarchical or flat structure, number of inputs, the partition of the input and output spaces by membership functions, and their shape), their inference mechanism, their aggregation operations, and their fuzzification and defuzzification methodology. Various schemes of fuzzy modelling, with specific algorithms, have been proposed to capture some specificity of the problem at hand and thus contribute to the efficiency of the overall identification schemes [18], if necessary by taking advantage of numeric experimental data and of neurofuzzy techniques [8][15]. Various fuzzy predictors (FP's) of time series have been introduced and have demonstrated their success in accurate predictions [2][7][16][19], whose main research effort was made in the creation of appropriate fuzzy rules and memberships through a learning process.

However, this effort generally does not guarantee a model with reasonable dynamical information, i.e. the derivative information of the time series. So, we propose a new fuzzy system structure that is capable of approximating regular functions as well as their derivatives on compact domains with linguistic information. Here, the linguistic information is associated to the translation process of fuzzy sets within the fuzzy relationships, that when modelled is able to describe local trends of the fuzzy models (temporal or positional derivatives). With derivative models of the temporal series, for the attractive regions of work, it is possible to make a Taylor series that can approximate a solution of ordinary differential equations or of a temporal series in distinct regions of the space. With this result it is possible to use the traditional ODE method to solve dynamical equations or to model temporal series.

This paper is organized as follows. Section 2 describes the perturbed fuzzy system. Section 3 presents theoretical aspects of the approximation of a function and its derivatives by the perturbed fuzzy system. Section 4 presents an ODE Fuzzy solver method based on the perturbed fuzzy system. The proposed algorithm will be used to benchmark prediction problems of the Mackey-Glass chaotic time series (Section 5). Finally, a conclusion is drawn.

2 The Perturbed Fuzzy System

Fuzzy systems modelling [9] provides a framework for modelling complex nonlinear relations, using a rule-based methodology. Consider a system $y = f(\boldsymbol{x})$; y is the output (or consequent) variable and $\boldsymbol{x} = (x_1, \cdots, x_n)^T \in \mathbb{R}^n$ is the input vector (or antecedent) variable. Let $\boldsymbol{U} = U_1 \times \cdots \times U_n$ be the domain of the input vector $\boldsymbol{x} \in \mathbb{R}^n$ and V the output space.

A linguistic model relating variables x and y can be written as a collection of rules that link terms $A_{i,j_i} \in U_i$, $j_i = 1, \ldots, N_i$, $i = 1, \ldots, n$, and $B_j \in V$, $j = (j_1, \ldots, j_n)$, where $A_{i,j_i}(x_i)$ and $B_j(y)$ represent the descriptor sets associated, respectively, to variables x_i, $i = 1, \ldots, n$ and y. In fuzzy systems modelling, this relationship is represented by a collection $\Im$ of fuzzy IF–THEN rules:

$$\begin{aligned} R_j &: \text{IF } x_1 \text{ is } A_{1,j_1} \text{ and ... and } x_n \text{ is } A_{n,j_n} \\ &\quad \text{THEN } y \text{ is } B_j \end{aligned} \tag{1}$$

where $j = (j_1, \ldots, j_n)$ is the index of rule, which belongs to the index set:

$$J = \{ j_1, \cdots, j_n \mid j_i = 1, \cdots, N_i; i = 1, \cdots, n \}.$$

The input space $\boldsymbol{U} = U_1 \times \cdots \times U_n$ and the output space V are being partitioned in $N = \prod_{i=1}^{n} N_i$ fuzzy regions, in which it is possible to define N fuzzy rules of the form of (1). The rule base can be represented by the fuzzy relation defined on the Cartesian product $A \times B$ [18]. If each input space U_j (for $j = 1, \ldots, n$) is completely partitioned by N_j fuzzy sets, then there is always at least one active rule. Given values for the input variables $\boldsymbol{x} = \boldsymbol{x}^*$, the value of y is calculated as a fuzzy subset G by a fuzzy inference process:

1. For each rule j, find the firing level of the rule:

$$A_j(\boldsymbol{x}) = A_{j_1}(x_1) \star A_{j_2}(x_2) \star \cdots \star A_{j_n}(x_n) \tag{2}$$

With the linguistic connective "and" of the antecedent of rule (1) defined as T-norm operation, "$\star$", A_j can be viewed as the fuzzy set $\times_{i=1}^{n} A_{i,j_i}$ with membership functions $A_j(\boldsymbol{x})$.

2. The fuzzy implication of each rule R_j: $A_j \mapsto B_j$ is a fuzzy set in $U \times V$ that is defined as $R_{j:A\mapsto B}(\boldsymbol{x}, y) = A_j(\boldsymbol{x}) \otimes B_j(y)$, where "$\otimes$" is an operator rule of fuzzy implication. For each rule j, calculate the effective output value G_j, based of sup-star composition:

$$G'_j(y) = \sup_{\boldsymbol{x} \in \boldsymbol{U}} \left[A'(\boldsymbol{x}) \star R_{j:A\mapsto B}(\boldsymbol{x}, y) \right]$$

3. Combine (unite) the individual outputs of the activated rules to find the overall system output $G = \bigcup_{j=1}^{N} G_j$.

Generally, $A'(\boldsymbol{x})$ is considered as a singleton set. For the arithmetic inference process, the output of each rule j is given by $G_j(y) = A_j(\boldsymbol{x}) \cdot B_j(y)$. In many situations, e.g. in series prediction and modelling applications, it is desirable to have a crisp value y^* for the output of a fuzzy system, instead of a fuzzy value $G(y)$. The defuzzifier maps the fuzzy sets in V to crisp points in V. In this paper, a centre-average defuzzifier [17] is used, and the output expression of the Fuzzy System can be written as

$$g(\boldsymbol{x}) = \sum_{j=1}^{N} A_j(\boldsymbol{x}) \cdot \theta_j \Bigg/ \sum_{j=1}^{N} A_j(\boldsymbol{x}) \tag{3}$$

where θ_j is the centroid point in V for which the membership function $B_j(y)$ achieves its maximum value, being assumed that $B_j(y)$ is a normal fuzzy set, i.e., $B_j(\theta_j) = 1$.

Fuzzy identification systems are able to integrate information from different sources, namely from human experts and from experimental observation. However, the process of translating this knowledge to linguistic IF–THEN rules is made as a sequence of static or instantaneous pictures of the modelled process, and so the dynamical information is discarded. However, the state variables of a dynamical process or temporal series are not static: in each instant they possess an instantaneous value *and* a trend of evolution. This trend, which contains information of the derivatives, should also be modelled by the diffuse system. A simple way to do this is to ensure that each input and output fuzzy set captures the trend, obtaining what we call a perturbed fuzzy set. These perturbed fuzzy sets are traditional fuzzy sets, characterized by their static position and shape, added with *potential* velocity, acceleration, etc. In the context of this paper, we are concerned with a special type of perturbed fuzzy sets, due both to a translation process and an additive process.

Definition 1: The nonlinear translation of a fuzzy set A of U by $h \in U$, denoted A_h, is the fuzzy subset of U defined as $A_h(x) = A(x - \sigma(h))$, where $\sigma(h)$ is a nonlinear homogenous translation function of the perturbation variable h, i.e., $\lim_{h \to 0} \sigma(h) = 0$.

Perturbation h moves fuzzy set A from its natural position to another position in the neighbourhood. As a special and known case we have $\sigma(h) = h$.

Definition 2: Let $\rho(h, x)$ be an additive perturbed function resulting from the product of a translation function weight $\sigma(h, x)$ by the membership function $A(x)$:

$$\rho(h, x) = \sigma(h, x) A(x)$$

where $\sigma(h, x)$ is a homogenous, nonlinear function of variable x and perturbation h, i.e., $\lim_{h \to 0} \sigma(h, x) = 0$ and $\sigma(h, x) = 0$. The additive perturbed fuzzy set of A is $A_h(x) = \varphi(x, h) A(x)$, where $\varphi(x, h) = 1 + \sigma(h, x)$.

For convenience of representation, consider that $\sigma_x(h) = \sigma(h, x)$. Both previously defined perturbed fuzzy sets obey the following axiom.

Axiom 1: Let $1 \le p \le \infty$ and $f \in L^P(\mathbb{R}^n)$. For $h \in \mathbb{R}^n$, let $A_h(x)$, the perturbed function of Definitions 1 and 2, be continuous with respect to variable h. Then $\lim_{h \to 0} \| A_h(x) - A(x) \| = 0$.

Let $A_{h,j}(\boldsymbol{x}) = A_{h,j_1}(x_1) \star \cdots \star A_{h,j_n}(x_n)$ be the aggregation of perturbed membership functions. For the product T-norm operation and perturbations of the additive type, we have $A_{\boldsymbol{h},j}(\boldsymbol{x}) = A_{h,j}(\boldsymbol{x}) \varphi_j(\boldsymbol{x}, \boldsymbol{h})$, where $\varphi_j(\boldsymbol{x}, \boldsymbol{h}) = \varphi_{h_1, j_1}(x_1, h_1) \cdots \varphi_{h_n, j_n}(x_n, h_n)$ and $\boldsymbol{h} = (h_1, \cdots, h_n)^T$ is the perturbation vector.

The fuzzy relationships that involve fuzzy set A are consequently also perturbed, and that reflects into the fuzzy system. The result of the perturbed fuzzy sets is a perturbed fuzzy system that is equal to the static fuzzy system when the perturbation variables h are null.

Definition 3: A perturbed fuzzy system, PFS, results from the perturbation of input and output fuzzy sets of fuzzy system (3). Let the input fuzzy sets of rules be of the additive type, i.e., $A_{h,j}(x) = \varphi_j(x, h) A_j(x)$, and let the output fuzzy sets be of the nonlinear translation type, $B_{\boldsymbol{h},j}(x) = B_j(x - \sigma_j(\boldsymbol{h}))$. The perturbed version of fuzzy system (3) is:

$$g(\boldsymbol{x}, \boldsymbol{h}) = \frac{\sum_{j=1}^{N} A_{\boldsymbol{h},j}(\boldsymbol{x}) \cdot (\theta_j + \sigma_j(\boldsymbol{h}))}{\sum_{j=1}^{N} A_{\boldsymbol{h},j}(\boldsymbol{x})} \tag{4}$$

Remark: If the input fuzzy sets of the PFS (4) are not perturbed and the perturbation vector h is obtained by difference between vector x and a constant vector $\boldsymbol{x}_0$ (for example, the null vector), $\boldsymbol{h}=\boldsymbol{x}-\boldsymbol{x}_0$, then the PFS given by (4) is the well-known TSK model. Moreover, if all input fuzzy sets of the PFS (4) are perturbed by the same additive function, i.e., $\varphi_j(x,h)=\varphi_i(x,h)$, $\forall i,j\in J$, then the PFS of equation (4) is also the well-known TSK model.

3 The Derivative Approximation

In this section we present the sufficient condition for the perturbed fuzzy systems as universal approximators, in order to answer the questions: "Given a real, continuous and differentiable function in $\mathbb{C}^v$, is there a perturbed fuzzy system that can approximate it up to its v^{th} derivative? How to perturb the membership functions and how many fuzzy sets (or fuzzy rules) are needed to ensure the desired approximation accuracy?"

The design of the static fuzzy system $g(\boldsymbol{x})$ is made by choosing the appropriate partition of the input space, the shape of the membership function and its position in the input space $\boldsymbol{U}$ and output space $\boldsymbol{V}$. These structural learning settings are of great importance to approximate the zero-order function $f(\boldsymbol{x})$. The derivative information could be included in the fuzzy modelling by associating it to the potential perturbation of its membership function. Without lost of generality, we will consider that the additive perturbed function $\varphi_j(x,h)=\varphi_j(h)$ is independent of variable $\boldsymbol{x}$. Furthermore, we assume that the perturbed functions are approximated by multivariate polynomials of the multivariate variable perturbation $\boldsymbol{h}$.

Definition 4: Let the perturbed functions $\sigma_j(\boldsymbol{h})$ and $\varphi_j(\boldsymbol{x},\boldsymbol{h})$, presented on Definition 1 and 2, be multivariate polynomials of degree r and s, respectively, defined on compact set $U\subset\mathbb{R}^n$, i.e.:

$$Q_{s,j}(\boldsymbol{h})=\varphi_j(\boldsymbol{h})=\sum_{d_1=0}^{s_1}\cdots\sum_{d_n=0}^{s_n}a^j_{d_1\cdots d_n}h_1^{d_1}\cdots h_n^{d_n} \tag{5}$$

$$P_{r,j}(\boldsymbol{h})=\theta_j+\sigma(\boldsymbol{h})=\sum_{d_1=0}^{r_1}\cdots\sum_{d_n=0}^{r_n}b_{d_1\cdots d_n}h_1^{d_1}\cdots h_n^{d_n} \tag{6}$$

where $\sum_{i=1}^{n}r_i=r$, $\sum_{i=1}^{n}s_i=s$, v=r+s, $b^j_{0,\cdots,0}=\theta_j$ and $a^j_{0,\cdots,0}=1$, for $j=1,\cdots,N$.

The perturbed fuzzy system, PFS, is now a rational function of polynomials of variable h:

$$g(\boldsymbol{x},\boldsymbol{h})=\frac{\sum_{j=1}^{N}A_j(\boldsymbol{x})Q_{s,j}(\boldsymbol{h})P_{r,j}(\boldsymbol{h})}{\sum_{j=1}^{N}A(\boldsymbol{x})Q_{s,j}(\boldsymbol{h})} \tag{7}$$

Two new theorems show us that fuzzy system (7) can: first, approximate a v^{th}-order polynomial to any degree of approximation; second, extrapolate the last result to approximate any nonlinear function.

Theorem 1: Consider the perturbed fuzzy system of equation (7). It can approximate any N distinct polynomials of order v, $T_{v,\boldsymbol{x}_j}(\boldsymbol{h})=\sum_{|\boldsymbol{\alpha}|=1}^{v}c_{\boldsymbol{\alpha}}(\boldsymbol{x}_j)\boldsymbol{h}^{\boldsymbol{\alpha}}$, in N distinct nodes $\boldsymbol{x}_j\in S$. Also, the i^{th} derivative of $g(\boldsymbol{x}_j,\boldsymbol{h})$ in respect to $\boldsymbol{h}$ can approximate the i^{th} derivative of $T_{v,\boldsymbol{x}_j}(\boldsymbol{h})$, for $i=1,\cdots,v$; i.e:

$$\left|E_{\boldsymbol{h}}^{(i)}(\boldsymbol{x})\right|=\left|\frac{\partial x^{(i)}T_{r,x}(\boldsymbol{h})}{\partial\boldsymbol{h}^{(i)}}-\frac{\partial^{(i)}g(\boldsymbol{x},\boldsymbol{h})}{\partial\boldsymbol{h}^{(i)}}\right|\le\left|\frac{\partial x^{(i+1)}T_{r,x}(\boldsymbol{h})}{\partial\boldsymbol{h}^{(i+1)}}\right|_{\infty}\boldsymbol{h}$$

and $E_{\boldsymbol{0}}^{(i)}(\bar{\boldsymbol{x}}_j)=\lim_{\boldsymbol{h}\to\boldsymbol{0}}E_{\boldsymbol{h}}^{(i)}(\bar{\boldsymbol{x}}_j)=0$.

Theorem 2: Suppose that N_j overlapped and equally distributed fuzzy sets are assigned to each input variable of Fuzzy System (7). Then, for any given real, continuous and differentiable function $f(\boldsymbol{x})$ defined in $\mathbb{C}^v$ and approximation error bound $\varepsilon>0$, there is a perturbed fuzzy system (7) with perturbed functions of Definition 4 that guarantees:

i) $g(x)=\lim_{h\to 0}g(x,h)$

ii) $\sup_{x\in\sup(f)}\left(\lim_{h\to 0}\left\|f(x)-g(x,h)\right\|\right)<\varepsilon$

iii) $\sup_{x\in\sup(f)}\lim_{h\to 0}\left\|\frac{\partial^{(i)}f(x)}{\partial x^{(i)}}-\frac{\partial^{(i)}g(x,h)}{\partial h^{(i)}}\right\|<\varepsilon$ for $i=1,..,v$

4 The Fuzzy ODE Taylor Series Method

A continuous, autonomous, stationary, nonlinear dynamical system can be described by a set of ordinary differential equations, ODE, $dx(t)/dt=F(x(t))$, where $x(t)$ is the vector of system states and F is the system vector field.

By expanding the solution of ODE to the initial value problem, $x(t_0)=x_0$, in a Taylor series about t_0, one obtains a local solution that is valid within its radius of convergence, R_0. If the series is evaluated at t_1, where $t_1<R_0$, we obtain an

approximation to $x(t_1)=x_1$, and the solution may then be expanded in a new series about t_1. The solution may of course then be extended to point t_2, and so forth, so that by a process of "analytical continuation" one obtains a piecewise polynomial solution to the ODE problem.

Perhaps the simplest one-step methods of order P are based on Taylor series expansion of the solution $x(t)$. If $x^{(p+1)}(t)$ is continuous on $[a,b]$, then Taylor's formula gives

$$x_{k+1}=x_k+f(t_k,x_k)h+\cdots+f^{(p-1)}(t_k,x_k)\frac{h^p}{p!}+O\left(h^{p+1}\right) \quad (8)$$

where $O\left(h^{p+1}\right)=f^{(p+1)}\left(\xi_k,x(\xi_k)\right)\frac{h^{(p+1)}}{p!}$, with $t_k\le\xi_k\le t_{k+1}$, and the total derivatives of f are defined recursively by:

$$f^{(i)}(t,x)=f_t^{(i-1)}(t,x)+f_y^{(i-1)}(t,x)\,f(t,x),\ i=1,2,\cdots.$$

To solve the ODE problem with (8), in each observable sample point x_k, it is necessary to (analytically or numerically) estimate the value of the dynamical system's derivatives. With a perturbed fuzzy system, a multi-derivative modelling can be created in order to be used in solving the prediction problem. The result is a fuzzy system (based on linguistic representation structure) that describes the temporal series as well as its derivatives.

With a set of $N_p=v\times N$ points of the temporal series, with N and v as in **Definition 4**, where we now use the values of the local Taylor series terms (up to the k^{th}-order continuous derivative), a multi-variate Padé approximation is used to identify the coefficients of the polynomials.

Definition 5: Consider a function $f(\boldsymbol{x})$, through its series expansion at a certain point $\boldsymbol{x}_k$ in U,

$$f(\boldsymbol{x}_k+\boldsymbol{h})=\sum_{|\boldsymbol{\iota}|=0}^{\infty}c_{\boldsymbol{\iota}}(\boldsymbol{x}_k)\boldsymbol{h}^{\boldsymbol{\iota}} \quad (9)$$

The Fuzzy Padé approximant, $g(\boldsymbol{x},\boldsymbol{h})$, is that rational fuzzy function of degree $v=r+s$ in the numerator and s in the denominator (polynomials Q's, $Q_j^s(\boldsymbol{h})$), whose power series expansion agrees with a given power series to the highest needed degree of f. If the rational function is of the type of (7), then $g(\boldsymbol{x},\boldsymbol{h})$ is said to be a Fuzzy Padé approximation to series (9), which in those points satisfies the conditions:

$$g(\boldsymbol{x},0)=f(\boldsymbol{x}) \quad (10)$$

and

$$\left.\frac{\partial^{(\boldsymbol{i})}g(\boldsymbol{x}+\boldsymbol{h})}{\partial\boldsymbol{h}^{(\boldsymbol{i})}}\right|_{\boldsymbol{h}=\boldsymbol{0}}=\left.\frac{\partial^{(\boldsymbol{i})}f(\boldsymbol{x}+\boldsymbol{h})}{\partial\boldsymbol{x}^{(\boldsymbol{i})}}\right|_{\boldsymbol{h}=\boldsymbol{0}},\quad 0\le|\boldsymbol{i}|\le v \quad (11)$$

These last two equations provide $v+1$ algebraic equations that involve $(v+1)\times N$ unknown parameters. This approach can be applied in a set of N points of space U, as much as the number of fuzzy rules of the PFS. The resulting equations allow us to solve the problem of finding the parameters. In this way, coefficients $a^j_{d_1,\cdots,d_n}$ of polynomials $Q_j^s(\boldsymbol{h})$ and coefficients $b_{d_1,\cdots,d_n}$ of $P_j^r(\boldsymbol{h})$ will be determined. The Padé approximant is the "best" approximation of a function by a rational function of the given order [1]. A Padé approximant often yields better approximation of the function than truncating its Taylor series and it may still work where the Taylor series does not converge.

5 Numerical Example

The PFS approach has been evaluated for a non-linear system identification problem, the Mackey-Glass chaotic time-series prediction problem.

The Mackey-Glass time series has been widely used as a standard benchmark for prediction algorithms (Crowder [5], Lapedes and Farber [10], Moody and Darken [13], …). The time series is generated by integrating the delay differential equation,

$$x'(t)=f(\boldsymbol{y}(t)) \quad (12)$$

where $x'(t)=dx(t)/dt$, $\boldsymbol{y}(t)=\left(x(t),x(t-\tau)\right)^T$ and $f(\boldsymbol{y}(t))=a\,x(t-\tau)/\left(1+x^c(t-\tau)\right)-b\,x(t)$. With $a=0.2$, $b=0.1$, $c=10$ and $\tau=17$, the time series is chaotic, exhibiting a cyclic but non-periodic behaviour. The upper order of temporal derivatives of state variable $x(t)$ can be recursively defined by:

$$x''(t)=f'_{x(t)}\cdot x'(t)+f'_{x(t-\tau)}\cdot x'(t-\tau)$$

$$x'''(t)=f'_{x(t)}\cdot x''(t)+f''_{x(t-\tau)}\cdot x'(t-\tau)^2+f'_{x(t-\tau)}\cdot x''(t-\tau)$$

They will be used for validating the FPS derivative approximation of the series.

For the calculation of the 3rd derivative of $x(t)$ a number of $3\times\tau$ past samples are necessary to save.

The task's goal is to use known values of the time series up to the point $x=t$ to predict the values at some point in the future $x=t+h$. The standard method for this type of prediction is to create a mapping from *D* points of the time series spaced *h* apart, i.e., vector $\boldsymbol{y}(t)=\left(x\left(t-\left(n_d-1\right)h\right),\ldots,x(t)\right)$, to a predicted future value $x(t+1)$. To allow comparison with earlier work [5][10][13], we used n_d = 4 and h = 6. All other simulation settings in this example were purposely arranged to be as close as possible to those reported in [8]. The numerical solution of equation (12) is obtained by the fourth-order Runge-Kutta with time step 0.1, initial condition $x(0)=1.2$, and we assume $x(t)=0$ for $t<0$. The time series thus generated consisted of 3000 data values, 2700 of which were used as training patterns and the other 300 as test data. The domain space is partitioned by a grid of triangular memberships. The total number of rules created, after excluding all unfired rules, was *N*=137.

Next, the PFS was used for simultaneously modelling the time series $x(t)$ and the derivative time series $x'(t)$, $x''(t)$ and $x'''(t)$. The fuzzy model has consequent polynomials of order 3 and antecedents of order zero. With these models the time series $x(t)$ can be predicted by Taylor ODE solver:

$$x(k+1)=g(\boldsymbol{y})+g'(\boldsymbol{y})T+g''(\boldsymbol{y})\frac{T^2}{2}+g'''(\boldsymbol{y})\frac{T^3}{6} \quad (13)$$

Note that, in the predicted process the past system output terms $x(k)$, $x(k-6)$, $x(k-12)$, and $x(k-18)$ were replaced by the respective model predictions. This free-running fuzzy model tests the stability of the obtained model.

Figure 1 shows the approximation of Mackey-Glass time series done by the ODE method (13), where $g^{(i)}(x)=\lim_{h\to 0}G^{(i)}(x,h)$, for $i=0,1,2,3$. In all cases, the resulting time series (from zero to third order) are practically coincident with the correspondent analytical time series. The mean square errors of this time series approximation and its derivatives are shown in Table 1.

6 Conclusions

The ability of the Perturbed Fuzzy Systems, PFS, to approximate any sufficiently smooth function, reproducing its derivatives up to any order, has been demonstrated. The PFS proved its ability to simultaneously estimate functions and their derivatives using information contained in finite numerical samples extracted from the data series, as well as its use in solving ODE problems.

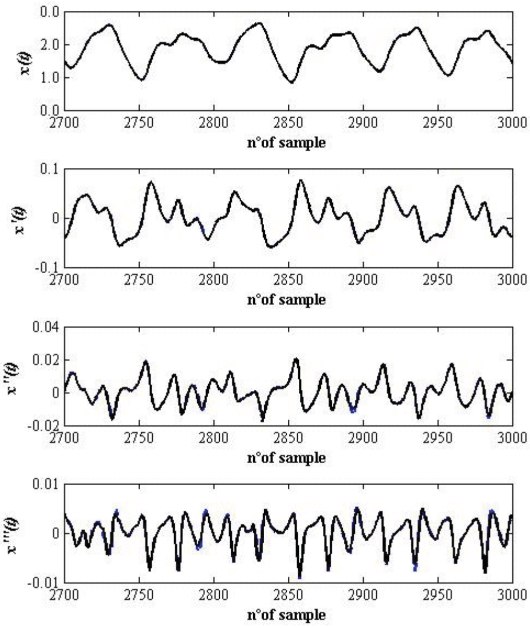

Figure 1: Modelling of the Mackey-Glass time series and its temporal derivative series. The exact time series (solid line) and the fuzzy approximation by PFS (dashed line) are practically coincident.

Table 1: Mean Square Error of the Approximation

ORDER OF DERIVATIVE	MACKEY-GLASS TIME SERIES
0	1.095×10^{-3}
1	9.808×10^{-7}
2	4.538×10^{-7}
3	2.044×10^{-7}

Acknowledgment

This work was supported by the Portuguese Fundação para a Ciência e a Tecnologia (FCT) under grant POSI/SRI/41975/2001.

References

[1] G. A. Baker, Jr. and P. Graves-Morris, *Padé Approximants*, Cambridge University Press, New York, NY, 1996.

[2] D. Benachenhou, Smart trading with (FRET), in: G. J. Deboeck (Ed.), *Trading on the Edge, Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets*, Wiley, New York, NY, 1994, pp. 215–242.

[3] G. E. P. Box and G. M. Jenkins, *Time Series Analysis, Forecasting, and Control*, Holden-Day, San Francisco, CA, 1970.

[4] W. A. Brock, D. A. Hsieh, and B. LeBaron, *Nonlinear Dynamics, Chaos and Instability*, MIT Press, Cambridge, MA, 1991.

[5] R. S. Crowder III, Predicting the Mackey-Glass time series with cascade-correlation learning, in: D. Touretzky, G. Hinton, and T. Sejnowski (Eds.), *Proceedings of the 1990 Connectionist Models Summer School*, Carnegie-Mellon University, Pittsburgh, PA, 1990, pp. 117–123.

[6] James D. Hamilton, *Time series analysis*, Princeton University Press, Princeton, NJ, 1994.

[7] J.-S. R. Jang and C.-T. Sun, Predicting chaotic time series with fuzzy IF–THEN rules, in: *Proceedings of the 2nd IEEE International Conference on Fuzzy Systems* (FUZZ-IEEE'93), San Francisco, CA, 1993, Vol. 2, pp. 1079–1084.

[8] J.-S. R. Jang, C.-T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing*, Prentice-Hall, Englewood Cliffs, NJ, 1997.

[9] George J. Klir and Bo Yuan, *Fuzzy Sets and Fuzzy Logic, Theory and Applications*, Prentice Hall PTR, Upper Saddle River, NJ, 1995.

[10] A. S. Lapedes and R. Farber, Nonlinear signal processing using neural networks: prediction and system modelling, in: *Technical Report LA-UR-87-2662*, Los Alamos National Laboratory, Los Alamos, NM, 1987.

[11] G. S. Maddala, *Introduction to Econometrics*, Prentice-Hall, Englewood Cliffs, NJ, 1996.

[12] M. Magdon-Ismail, A. Nicholson, and Y. S. Abu-Mostafa, Financial markets: very noisy information processing, in: *Proceedings of the IEEE*, Vol. 86, No. 11, Nov. 1998, pp. 2184–2195.

[13] J. Moody and C. J. Darken, Fast learning in networks of locally-tune d processing units, in: *Neural Computation*, Vol. 1, 1989, pp. 281–294.

[14] A.-P. N. Refenes, A. N. Burgess, and Y. Bentz, Neural networks in financial engineering: a study in methodology, in: *IEEE Transactions on Neural Networks*, Vol. 8, No. 6, Nov. 1997, pp. 1222–1267.

[15] L. H. Tsoukalas and R. E. Uhrig, *Fuzzy and Neural Approaches in Engineering*, Wiley, New York, NY, 1997.

[16] L.-X. Wang and J. M. Mendel, Generating fuzzy rules by learning from examples, in: *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 22, Nov. 1992, pp. 1414–1427.

[17] Li-Xin Wang, *A course in fuzzy systems and control*, Prentice-Hall PTR, Upper Saddle River, NJ, 1997.

[18] R. R. Yager and D. Filev, *Essentials of Fuzzy Modeling and Control*, Wiley, New York, NY, 1994.

[19] Z. Ye and L. Gu, A fuzzy system for trading the Shanghai stock market, in: G. J. Deboeck (Ed.), *Trading on the Edge, Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets*, Wiley, New York, NY, 1994, pp. 207–214.