

Equivalences Between Neural-Autoregressive Time Series Models and Fuzzy Systems

José Luis Aznarte and José Manuel Benítez, *Member, IEEE*

Abstract—Soft computing (SC) emerged as an integrating framework for a number of techniques that could complement one another quite well (artificial neural networks, fuzzy systems, evolutionary algorithms, probabilistic reasoning). Since its inception, a distinctive goal has been to dig out the deep relationships among their components. This paper considers two wide families of SC models. On the one hand, the regime-switching autoregressive paradigm is a recent development in statistical time series modeling, and it includes a set of models closely related to artificial neural networks. On the other hand, we consider fuzzy rule-based systems in the framework of time series analysis. This paper discloses original results establishing functional equivalences between models of these two classes, and hence opens the door to a productive line of research where results and techniques from one area can be applied in the other. As a consequence of the equivalences presented in this paper, we prove the asymptotic stationarity of a class of fuzzy rule-based systems. Simulations based on information criteria show the importance of the selection of the proper membership function.

Index Terms—Autoregression, functional equivalence, fuzzy rule-based models, regime-switching, time series.

I. INTRODUCTION

SOFT computing (SC) is a term coined by L. A. Zadeh [38] to refer to a collection of methodologies that aim at exploiting the tolerance for imprecision and uncertainty to achieve tractability, robustness, and low-solution cost. Its principal constituents are fuzzy logic, neurocomputing, evolutionary algorithms and probabilistic reasoning. Moreover, the original inspiring philosophy still encourages the combination of those techniques with new ones, e.g., support vector machines.

While successful engineering applications have led to a speedy growth in interest in SC, the research into the foundations of its components has also concentrated a serious effort in the area. In particular, the study of the connections between them has been a main concern. The relevance of this topic is remarked by the availability of an extensive literature on it.

Manuscript received May 8, 2009; revised October 14, 2009; accepted June 2, 2010. Date of publication August 17, 2010; date of current version September 1, 2010. This work was supported in part by the Ministerio de Ciencia e Innovación of the Spanish Government, under Grant TIN2009-14575.

J. L. Aznarte is with the Renewable Energy Research Group, Center for Energy and Processes of MINES ParisTech, 06904 Paris, France (e-mail: jlaznarte@decsai.ugr.es).

J. M. Benítez is with the Department of Computer Science and Artificial Intelligence, CITIC, University of Granada, 18071 Granada, Spain (e-mail: j.m.benitez@decsai.ugr.es).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TNN.2010.2060209

Some examples are [2], [5], [8], [11], [13], [15], [18], [19], [23], [24], and [35].

The list of benefits attained from those equivalence results quickly grows: interpretations of one kind of models in terms of others, procedures for knowledge extraction or injection, transfer of properties or algorithms between techniques, to name a few. In this paper, we focus on the connections between fuzzy rule-based systems and neural-autoregressive models used for time series analysis and forecasting.

Time series analysis is an application problem which has always attracted the attention of SC researchers. Forecasting future values of a series is usually a very complex task, and many SC methods and models have been used to tackle it, including artificial neural networks (ANNs) and fuzzy rule-based systems (FRBSs) in their various formulations [30]. Notwithstanding, a common characteristic of those approaches is that they usually consider time series as just another dataset which requires some small adaptations to be cast into the regression or classification format for which most SC models were created.

However, time series analysis is a prominent field in Econometrics, where it has been widely studied under a statistical perspective during the last centuries. The works by Bachelier [4], Kolmogorov [20], Khinchin [17], and Yule [37], all from the first third of the 20th century, are some classic milestones for the statistical time series theory. Recently, in 1970, the old idea of forecasting future values of a time series as a combination of its past values received a strong impulse after [6]. In that work, Box and Jenkins proposed a modeling cycle for the autoregressive integrated moving average model, of which the autoregressive (AR) model is a popular special case.

Of course this linearity assumption implies certain limitations, and in the last decades much statistical research has been devoted to nonlinear models. Nonlinear and nonstationary models are more flexible in capturing the characteristics of data and, in some cases, are better in terms of estimation and forecasting.

For some reason, SC researchers do not usually go deep into time series analysis, disregarding all the knowledge gathered through the years in the statistical field. In this paper, we take a step forward in the quest for an SC-based time series research which integrates methods and models coming from the econometric perspective, introducing some equivalence relationships between models from both areas.

The main contribution of this paper deals with the functional equivalences which link the family of regime switching AR

models and FRBSs. As evidenced by the influence of previous papers on equivalence results, these propositions and theorems have a high relevance by themselves in order to establish sound foundations for the field. Furthermore, considering the number of applied results which can be derived from them, they are also of an undeniable practical importance. We illustrate this by proving the asymptotic stationarity of additive FRBSs.

This paper is structured as follows. Regime switching AR models are reviewed in Section II. Then, the original (equivalence) results are provided in a constructive way along Section III. Some immediate consequences, examples and implications of the results are disclosed in Section IV, whereas the results of a Monte Carlo simulation are shown in Section V. The paper is finished with conclusions gathered in Section VI.

II. REGIME SWITCHING AUTOREGRESSIVE MODELS

As stated above, in statistical time series modeling, one of the oldest and most successful concepts is to forecast future values of a time series as a combination of its past values. This is a quite natural idea that we apply on every day life, and it was popularised in 1970 after [6]. In that work, Box and Jenkins formalised the use of the AR model, which assumes that future values of a time series can be expressed as a linear combination of its past values.

An AR model of order $p \geq 1$ is defined as

$$y_t = a_0 + a_1 y_{t-1} + \dots + a_p y_{t-p} + \varepsilon_t \quad (1)$$

where $\{\varepsilon_t\} \sim N(0, \sigma^2)$, usually known as Gaussian *white noise* (equivalent to a random signal with a flat power spectral density). For this model, we write $\{y_t\} \sim \text{AR}(p)$, and the time series $\{y_t\}$ generated from this model is called the $\text{AR}(p)$ process.

Such a simple model proved to be extremely useful and suited to series which, at first sight, seemed to be too complex as to be linear. Applications of the Box and Jenkins methodology spread in the following decades, covering various scientific areas such as biology, astronomy, or econometrics.

However, there were still many problems which could not be addressed using linear models. In 1978, taking a step toward nonlinearity, Tong [34] proposed a *piecewise linear* model: the threshold autoregressive (TAR) model. The success of this model in Econometrics gave birth to a new family of models, the AR regime switching models, which are based on the idea of partitioning the state-space into several subspaces, each of which is to be modeled by an AR model.

A general AR regime switching model with k ($k \geq 2$) regimes can be defined as

$$y_t = \sum_{i=1}^k \mathbf{a}'_i \mathbf{x}_t \cdot \Phi_i(\mathbf{z}_t; \boldsymbol{\psi}_i) + \varepsilon_t \quad (2)$$

where $\mathbf{x}_t = (1, y_{t-1}, y_{t-2}, \dots, y_{t-p})$ is an input vector containing p lagged values of the series and \mathbf{a}_i defines the local AR model i [note that $\mathbf{a}'_i \mathbf{x}_t$ encodes the skeleton of the AR model defined by (1)]. The variable controlling the transition is \mathbf{z}_t and normally is composed of a subset of the elements of \mathbf{x}_t (hence, $\mathbf{z}_t \in \mathbb{R}^q$ with $q \leq p$). The vector of parameters $\boldsymbol{\psi}_i$

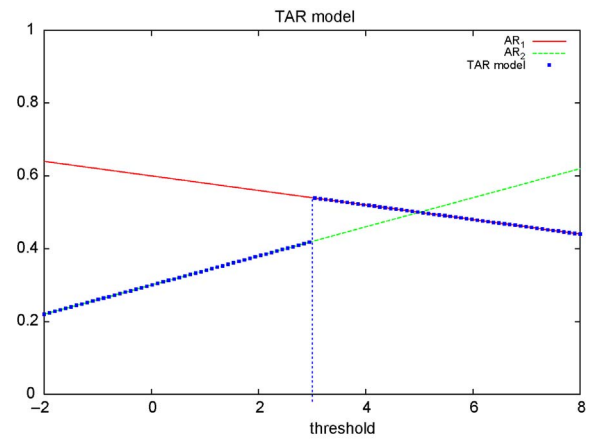


Fig. 1. Example of TAR model.

defines the location and shape of the transition functions Φ_i , whose functional form is one of the main differences among the models of the family.

A. Threshold Autoregressive Model (TAR)

As mentioned above, the TAR is the seminal regime switching model, and is characterised by using the indicator function as transition function, i.e., $\Phi_i = \mathbf{I}_{A_i}$. This function, described below in detail, marks the sharp changes from one linear model to another through a set of thresholds defined on one of the variables involved. This variable can be an exogenous variable associated to the process being modeled or one of the lagged values of the series, in which case the model is called self-exciting.

A self exciting threshold autoregressive (SETAR) model is defined as

$$y_t = \sum_{i=1}^k \mathbf{a}'_i \mathbf{x}_t \cdot \mathbf{I}_{A_i}(y_{t-d}) + \varepsilon_t \quad (3)$$

where y_{t-d} is the value of the series at time $t-d$ and is usually known as the threshold variable, \mathbf{I}_{A_i} is an indicator (or *step*) function (which takes the value zero below the threshold and one above it) and $\{A_i\}$ forms a partition of $(-\infty, \infty)$, with $\bigcup_{i=1}^k A_i = (-\infty, \infty)$ and $A_i \cap A_j = \emptyset$, $\forall i \neq j$.

Usually, we define the interval $A_i = (c_{i-1}, c_i)$, with $-\infty = c_0 < c_1 < \dots < c_k = \infty$, where the c_i 's are called thresholds. The ordering of the thresholds is required in order to guarantee the identifiability of the model. Fig. 1 shows the graphical representation of a two regimes SETAR model.

B. Smooth Transition Autoregressive Model (STAR)

A key feature of TAR models is the discontinuous nature of the AR relationship as the threshold is passed. Taking into account that nature is generally continuous, in 1994 an alternative model called STAR was proposed by Teräsvirta [33]. In STAR models there is a smooth continuous transition from one linear AR to another, rather than a sudden jump.

In this model and variants, the indicator function is substituted by a smooth function with sigmoid characteristics. The

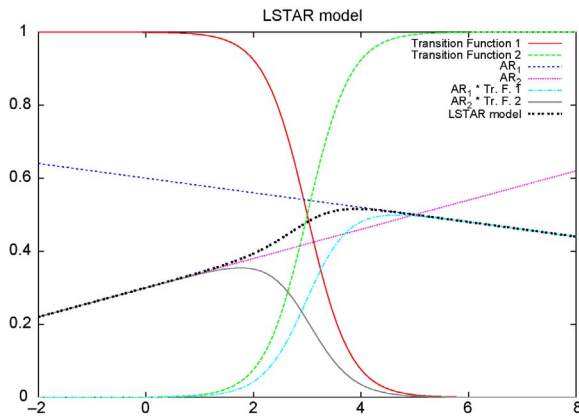


Fig. 2. Example of two regime STAR model using logistic transition function.

STAR model is defined as

$$y_t = \sum_{i=1}^k \mathbf{a}'_i \mathbf{x}_t \cdot \Phi_i(y_{t-d}; \boldsymbol{\psi}_i) + \varepsilon_t. \quad (4)$$

The transition functions $\Phi_i(y_{t-d}; \boldsymbol{\psi}_i)$ are continuous functions, bounded between 0 and 1, with parameters $\boldsymbol{\psi}_i$. The regime that occurs at time t is determined by the observable lagged variable y_{t-d} and the associated value of $\Phi_i(y_{t-d}; \boldsymbol{\psi}_i)$. Different choices for the transition functions give rise to different types of regime-switching behavior. A popular choice is when $\Phi_1 = 1$ (the function constantly equal to 1) and $\Phi_2 = \dots = \Phi_k = f$, where f is the first-order logistic function with parameters $\boldsymbol{\psi}_i = (\gamma_i, c_i)$ for regime i

$$f(y_{t-d}; \boldsymbol{\psi}_i) = (1 + \exp(-\gamma_i(y_{t-d} - c_i)))^{-1}. \quad (5)$$

The resultant model is called the Logistic STAR (LSTAR). Fig. 2 shows a STAR model with two regimes for which $\Phi_1 = 1 - f$ and $\Phi_2 = f$.

The parameters c_i in (5) can be interpreted as the threshold between two regimes, in the sense that the logistic function changes monotonically from 0 to 1 as y_{t-d} increases and $f(c_i; \gamma_i, c_i) = 0.5$.

The parameter γ_i determines the smoothness of the transition from one regime to another. As γ_i becomes very large, the logistic function approaches an indicator function and hence, the change of $f(y_{t-d}; \gamma_i, c_i)$ from 0 to 1 becomes instantaneous at $y_{t-d} = c_i$. Consequently, the LSTAR nests TAR models as a special case. Furthermore, when $\gamma \rightarrow 0$ the LSTAR model reduces to a linear AR model.

In the LSTAR model, the regime switches are associated with small and large values of the transition variable y_{t-d} relative to c_i . In certain applications it may be more appropriate to specify a transition function such that the regimes are associated with small and large absolute values of y_{t-d} (again relative to c_i). This can be achieved by using, for example, the exponential function, in which case the model may be named ESTAR.

As it is the case for the TAR model, symmetries of the parameter space cause unidentifiability of the STAR i.e., it cannot be uniquely identified. Enforcing the ordering of the regimes

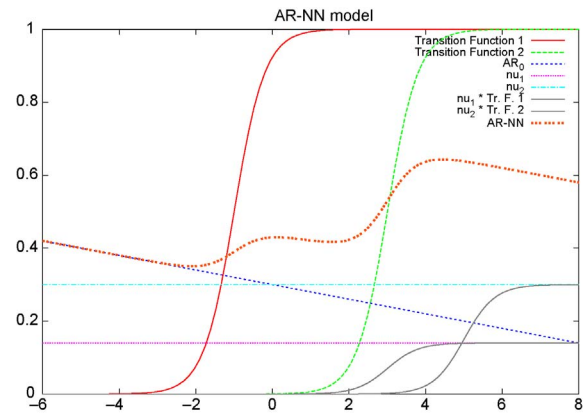


Fig. 3. Example of an AR-NN with two hidden units.

($c_1 < c_2 < \dots < c_k$) partially solves this problem. Notwithstanding, the logistic activation function (which verifies that $f(x) = 1 - f(-x)$), is another source for unidentifiability, so the restriction $\gamma_i > 0$ must also be respected for every i .

C. Autoregressive Neural Network Model (AR-NN)

After the success of ANNs in so many fields including Time Series Analysis, some researchers [25], [32] considered them as statistical nonlinear models and applied statistical inference to the problem of their specification. They devised a “bottom-up” strategy which allowed for proper statistical inference, as well as an in-sample evaluation of the estimated model.

The autoregressive single hidden layer neural network (AR-NN) model [25] is defined as

$$y_t = \mathbf{a}'_0 \mathbf{x}_t + \sum_{i=1}^k \alpha_i \Phi_i(\mathbf{b}'_i \mathbf{z}_t; \boldsymbol{\psi}_i) + \varepsilon_t \quad (6)$$

being α_i the connection weights and \mathbf{b}_i a vector of real valued parameters defining a linear transformation on \mathbf{z}_t . For this AR regime-switching model, the functions Φ_i are assumed to be logistic in this paper, $\Phi_1 = \dots = \Phi_k = f$, as defined in (5). Although in the SC field it is frequent to take $\mathbf{a}_0 = 0$, the original formulation of the AR-NN included this “linear unit.”

The geometric interpretation of this model considers that the AR-NN divides the p -dimensional Euclidean space with hyperplanes (defined by $\mathbf{b}'_i \mathbf{z}_t$) resulting in several polyhedral regions. It computes the output as the sum of the contribution of each hyper-region modulated by the smoothing function f . Fig. 3 shows an example of the shape of the function generated by an AR-NN with two hidden units.

Following [25], an AR-NN can be either interpreted as a semi-parametric approximation to any Borel-measurable function or as an extension of the LSTAR model where the transition variable can be a linear combination of stochastic variables.

Three characteristics of the model imply nonidentifiability. The first one is the interchangeability property of the elements of the AR-NN model. The value in the likelihood function of the model remains unchanged if we permute the hidden units. This results in $k!$ different models that are indistinguishable

from one another and in $k!$ equal local maxima of the log-likelihood function. The second characteristic is that, for the transition function, $f(x) = 1 - f(-x)$. This yields two observationally equivalent parametrisations for each hidden unit. Finally the presence of irrelevant hidden units is also a problem. If (6) has hidden units such that $\alpha_i = 0$ for at least one i , the parameters \mathbf{b}_i remain unidentifiable. Conversely, if $\mathbf{b}_i = 0$ then α_i can take any value without the likelihood function being affected.

The approach devised by [25] overcomes these limitations by imposing some restrictions on the parameters. The first problem is solved by enforcing $\alpha_1 > \dots > \alpha_k$ or $b_{10} < \dots < b_{k0}$. The second problem is solved by enforcing $b_{i1} > 0$ for every i . Finally, the third problem is dealt with by applying statistical inference in the model specification.

D. Linear Local Global Neural Network (L^2GNN)

Another member of the regime switching family, and a recent statistical approach to ANNs, is the local global neural network (LGNN) model [31]. The central idea of LGNN is to express the input-output mapping by a piecewise structure. The model output is constituted by a combination of several pairs, each of those composed by an approximation function and by an activation-level function. The activation-level functions are equivalent to the transition function of the general AR regime switching model, and define the role of an associated approximation function for each subset of the domain. Partial superposition of activation-level functions is allowed. In this way, the problem of approximation functions is faced through the specialisation of neurons in each of the sectors of the domain. In other words, the neurons are formed by pairs of activation-level and approximation functions that emulate the generator function in different parts of the domain.

The LGNN is defined as

$$y_t = \sum_{i=1}^k L(\mathbf{x}_t; \boldsymbol{\chi}_i) \Phi_i(\mathbf{z}_t; \boldsymbol{\psi}_i) + \varepsilon_t \quad (7)$$

where the functions L and Φ_i are the approximation and transition functions, respectively.

In the original formulation [31], $\Phi_i(\mathbf{z}_t; \boldsymbol{\psi}_i)$ is noted as $B(\mathbf{z}_t; \boldsymbol{\psi}_i)$ and is defined as the difference between two opposed logistic functions

$$B(\mathbf{z}_t; \boldsymbol{\psi}_i) = -\left(f(\mathbf{b}'_i \mathbf{z}_t; \gamma_i, c_i^{(1)}) - f(\mathbf{b}'_i \mathbf{z}_t; \gamma_i, c_i^{(2)}) \right) \quad (8)$$

where $\boldsymbol{\psi}_i = (\mathbf{b}_i, \gamma_i, c_i^{(1)}, c_i^{(2)})$, representing a linear transformation of \mathbf{z}_t encoded by \mathbf{b}_i , a steepness parameter γ_i and two location parameters $(c_i^{(1)}, c_i^{(2)})$.

This model is closely related to the *mixture-of-experts* approach [14] and offers a great deal of flexibility in the functional form of the approximation function $L(\mathbf{x}_t; \boldsymbol{\chi}_i)$. This flexibility has not been fully explored so far, but there have been attempts to combine in the same model linear approximators with nonlinear ones [12], for example.

A special case of the LGNN model is the L^2GNN [31]. In this case, the approximation functions are linear, that is, $\boldsymbol{\chi}_i = \mathbf{a}_i$ is a vector of linear parameters and $L(\mathbf{x}_t; \boldsymbol{\chi}_i) = \mathbf{a}'_i \mathbf{x}_t$.

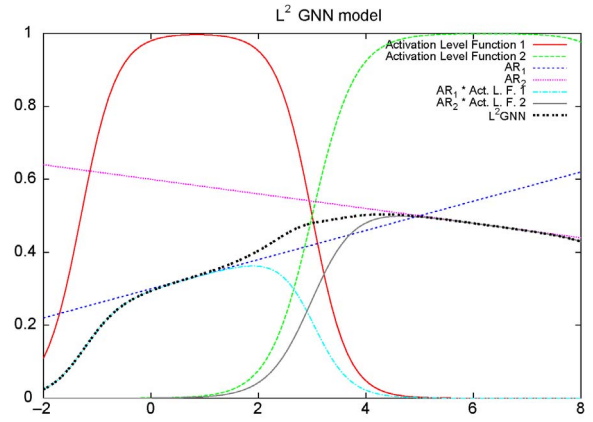


Fig. 4. Example of a two regime L^2GNN .

Hence, the L^2GNN is closely related with the general AR regime switching model of (2), and is defined as

$$y_t = \sum_{i=1}^k \mathbf{a}'_i \mathbf{x}_t B(\mathbf{z}_t; \boldsymbol{\psi}_i) + \varepsilon_t. \quad (9)$$

It is worth noting that, as in the previous models, this model is neither locally nor globally identifiable. In [31] the restrictions which ensure identifiability are stated: for $i = 1, \dots, k$ the ordering of the thresholds is given by $c_i^{(1)} < c_{i+1}^{(1)}$ and $c_i^{(2)} < c_{i+1}^{(2)}$ together with $c_i^{(1)} < c_i^{(2)}$, whereas the identifiability problems posed by the symmetry of the transition function are solved by enforcing $\gamma_i > 0$ and $b_{i1} > 0$. Fig. 4 shows a simplified L^2GNN model with two regimes.

E. Neuro-Coefficient Smooth Transition Autoregressive Model (NCSTAR)

One of the latest developments in threshold-based models is the Neuro-Coefficient STAR [26]. This model is a generalisation of some of the previously described models and can handle multiple regimes and multiple transition variables. This model can be seen as a linear model whose parameters change through time and are determined dynamically by a single hidden layer feed-forward neural network.

Consider a linear model with time-varying coefficients expressed as in (1) and let the coefficients vary through time $a_0(t), a_1(t), \dots, a_p(t)$. The time evolution of such coefficients is given by the output of a single hidden layer neural network with k hidden units

$$a_j(t) = \sum_{i=1}^k \alpha_{ij} f(\mathbf{b}'_i \mathbf{z}_t; \gamma_i, c_i) - \alpha_{j0} \quad (10)$$

where $j = 0, \dots, p$, α_{ji} , and α_{j0} are real coefficients (connection weights) and f is a logistic function as defined in expression (5).

Substituting the p realizations of (10) in the linear model, we obtain the general form of the NCSTAR model

$$y_t = \boldsymbol{\alpha}'_0 \mathbf{x}_t + \sum_{i=1}^k \boldsymbol{\alpha}'_i \mathbf{x}_t f(\mathbf{b}'_i \mathbf{z}_t) + \varepsilon_t. \quad (11)$$

Similarly to \mathbf{a}_i in the previous models, α_i represents a vector of real coefficients, called *linear parameters*. In this model, the value of the slope parameter γ_i is taken to be the norm of \mathbf{b}_i . In the limit, when the slope parameter approaches infinity, the logistic function becomes a step function.

As happened with previous models, this model is neither locally nor globally identifiable, and this is due to the special characteristics of its functional form that cause nonidentifiability. In order to guarantee identifiability, we need to impose some restrictions, namely $c_i < c_{i+1}$ and $b_{i1} > 0$. Also, it is important to ensure that no irrelevant units are included, which can be achieved by using the incremental building procedure proposed in [28].

The choice of the elements of \mathbf{z}_t , which determines the dynamics of the process allows a number of special cases. An important one is when $\mathbf{z}_t = y_{t-d}$. In this case, (11) becomes a LSTAR model with k regimes. It should be noticed as well that this model also nests the SETAR model. When $\gamma_i \rightarrow \infty \forall i$, the LSTAR model becomes a SETAR model with k regimes.

Another interesting case is when $\alpha'_i = (\alpha_{i0}, 0, \dots, 0), \forall i > 0$. Then, the model becomes an AR-NN model with k hidden units. Finally, this model is related to the functional coefficient autoregressive (FAR) model [9], to the single-index coefficient regression model [36], and to fuzzy rule-based models, as we shall see below.

III. RELATIONS WITH FUZZY RULE-BASED MODELS

As stated above, establishing the equivalence of different models has been important in the neural networks field since its establishment (see [5], [8], [11], [15], [18], [23], [24]). These results imply some useful consequences as the possibility of interpreting one family of models in terms of the others or the transfer of properties and algorithms. Concerning neural networks, these results allowed to overcome the “black-box” characteristic as they led to knowledge extraction methods.

In [3] we explored the links existing between an AR model and a fuzzy rule used in the time series framework. As well, we proved that STAR models can be seen as a particular case of a fuzzy rule-based system. Here, we will extend those results to the neural AR models listed in Section II. For the sake of clarity, let us first note the expression of the FRBS considered here.

When dealing with time series problems the Takagi-Sugeno-Kang paradigm is preferred over other variants of FRBSs. When applied to model or forecast a univariate time series $\{y_t\}$, the rules of a TSK FRBS are expressed as

$$\begin{aligned} \text{IF } y_{t-1} \text{ IS } B_1 \text{ AND } \dots \text{ AND } y_{t-p} \text{ IS } B_p \\ \text{THEN } y_t = a_0 + a_1 y_{t-1} + \dots + a_p y_{t-p}. \end{aligned} \quad (12)$$

In this rule, all the input variables are lagged values of the time series $\{y_t\}$.

Concerning the fuzzy reasoning mechanism for TSK rules, the *firing strength* of the i th rule is obtained as the t -norm (usually, multiplication operator) of the membership values of

the premise part linguistic labels

$$w_i(\mathbf{x}_t) = \prod_{j=1}^p \mu_{B_j^i}(y_{t-j}) \quad (13)$$

where the shape of the membership function of the linguistic terms $\mu_{B_j^i}$ can be chosen from a wide range of functions. One of the most common is the Gaussian bell, although it can also be a logistic function as in (5) or some nonderivable functions as the triangular or trapezoidal functions.

The overall output is computed as a weighted average or weighted sum of the rule's output. In the case of the weighted sum, the output expression for an FRBS with k rules is

$$y_t = \sum_{i=1}^k \mathbf{a}'_i \mathbf{x}_t \cdot w_i(\mathbf{z}_t) \quad (14)$$

where \mathbf{a}_i are the so-called *consequent parameters*. While many TSK FRBS perform a weighted average to compute the output, additive FRBS are also a common choice. They have been used in a large number of applications (for example, [7], [10], [16], [21]).

A. AR Model and the TSK Fuzzy Rules

Fuzzy rules are the core element of fuzzy systems. When applied to time series, as seen in (12), fuzzy rules can describe the relationship between the lagged variables in some parts of the state-space. A close look into this equation suggested the following.

Proposition 1: When used for time series modelling, a TSK fuzzy rule can be seen as a local AR model, applied on the state-space subset defined by the rule antecedent.

This connection between the two models opened the possibility of an exchange of knowledge from one field to another, enabling us to apply what we know about AR models to fuzzy rules and vice versa. From the point of view of Box–Jenkins models, each of these rules represents a local AR model which is applied only when some conditions hold. These conditions are defined by the terms in the rule antecedent. The output of the AR system is modulated by the membership degree of the lagged variables to some fuzzy sets describing parts of the state-space domain. This scheme is closely related to the structure of the threshold autoregressive family of models, as shown below.

B. STAR Model and Fuzzy Rule-Based Models

After the previous result, we were able to go further in the exploration of the relationships between threshold models and fuzzy logic-based models. On the one hand, we have seen that AR models are good linear models applicable to prediction problems. As well, we know that a TAR model is basically a set of local AR models, and that it allows for some nonlinearity in its computations. On the other hand, we have seen how a fuzzy rule relates to an AR model, in Proposition 1. Knowing that fuzzy rule-based models contain sets of fuzzy rules, we were interested in considering the relationship existing between threshold models and fuzzy rule-based models.

It is rather clear that there is some parallelism between the two aforementioned families of models. At a high level, models from both sides are composed of a set of elements (AR—fuzzy rules) which happen to be closely related, as stated above. On a lower level, both families of models rely on building a hyper-surface on the state-space which tries to model the relationship between the lagged variables of a time series. Moreover, both define this hyper-surface as the composition of hyper-planes which apply only in certain parts of the state-space.

Indeed, the following was proved. For a deeper discussion on these basic facts, refer to [3].

Proposition 2: The STAR model is functionally equivalent to an Additive TSK FRBS with only one term in the rule antecedents.

C. Neural Autoregressive Models and Fuzzy Rule-Based Systems

As described in Section II, some of the most recent developments of the threshold AR family of models include AR-NN, LGNN, and NCSTAR models. We will now explore the consequences of Proposition 1 regarding those models.

1) *Autoregressive Neural Network (AR-NN):* Recalling (6), it is clear that the AR-NN is composed of an AR linear term plus a neural network. The neural network term is a regular multilayered perceptron, and, as such, is interpretable as a fuzzy additive system, in the way shown in [5]. This paper states as well that, by using the interactive-or operator, it is possible to view ANNs as Mamdani-type fuzzy rule-based models.

Furthermore, under the FRBS paradigm, the AR term of the AR-NN can be considered as a *generic* rule, that is, a rule which applies on the whole domain of the problem. Such generic rules, which fire unconditionally, produce a default answer which is added to the values of the fired rules on those areas covered by them. This type of rules has been used previously by researchers and practitioners to encode knowledge which is domain-wide applicable.

Thus, we can prove the following.

Proposition 3: The AR-NN model is functionally equivalent to a TSK FRBS with a default rule.

Proof: Using the result in [5], which states that a neural network is functionally equivalent to an FRBS, and considering the AR term as a rule of type

$$\text{IF true THEN } y_t = \mathbf{a}'_0 \mathbf{x}_t \quad (15)$$

the proof is trivial. ■

Viewing the AR-NN as a combination of an AR model and a fuzzy inference system allows for linguistic interpretation of the system. In addition, this let us include *a priori* expert knowledge into the model. Other advantages of this equivalence relationship will be addressed in Sections IV-A and IV-B.

2) *Local Global Neural Network:* The more general approach of LGNN models, closely related to mixtures of experts model, satisfies the following.

Proposition 4: Local global neural networks are a generalization of Additive TSK FRBS.

Proof: It is straightforward after considering the expression of TSK rules (12), and the expression for the LGNN (7). Since $L(\mathbf{x}_t; \chi_i)$ can take any form, it can also be a linear function of the inputs, which is exactly a TSK rule. As the aggregation rule for LGNN is additive, we can conclude that the LGNN model is a generalization of Additive TSK FRBS. ■

For the same reason that after setting the general original formulation of the LGNN, researchers straightforwardly focused on linear approximation functions (linear LGNN models), TSK rules are basically used with linear consequents. It is generally preferred to keep the consequents linear and to encode all the nonlinearity in the antecedents.

If linear consequents were used (i.e., in the L^2 GNN model), though, the relationship with Additive TSK FRBS is immediate.

Proposition 5: L^2 GNN models are functionally equivalent to Additive TSK FRBS using $B(\mathbf{z}_t; \boldsymbol{\psi}_i)$ as membership function.

3) *Neuro-Coefficient Smooth Transition Autoregressive Models:* This kind of systems introduces time varying coefficients to combine AR models. Their mathematical formulation is quite similar to the L^2 GNN model, varying only the form of their activation level functions (which has a smaller number of parameters).

Hence, when studying the links of the NCSTAR to FRBSs, we find similar results to those obtained for the previous statistical models. They can be expressed in terms of the following.

Proposition 6: NCSTAR models are functionally equivalent to Additive TSK FRBS with logistic membership function.

Proof: We must recall equations (14) (Additive TSK FRBS) and (11) (NCSTAR model). Considering that the multidimensional logistic function is obtained as the product of unidimensional logistic functions, it is easy to see that the firing degree of a rule is equivalent to the transition function of a hidden unit of the NCSTAR, and hence, that both models are functionally equivalent. ■

Finally, the following Theorem condenses the results drawn above.

Theorem 1: The TSK FRBS is a generalization of the regime switching AR models TAR, STAR, AR-NN, L^2 GNN, and NCSTAR.

Proof: Trivial in the light of Propositions 1, 2, 3, 5, and 6. ■

IV. CONSEQUENCES AND IMPLICATIONS

Theorem 1 entails important implications that may affect the way threshold models and FRBSs are understood and used. Since each of these threshold models can be expressed as a fuzzy rule based model, all the properties and tools of this SC technique are directly applicable to it. The opposite is also true: tools and properties of threshold models are valid for fuzzy inference systems.

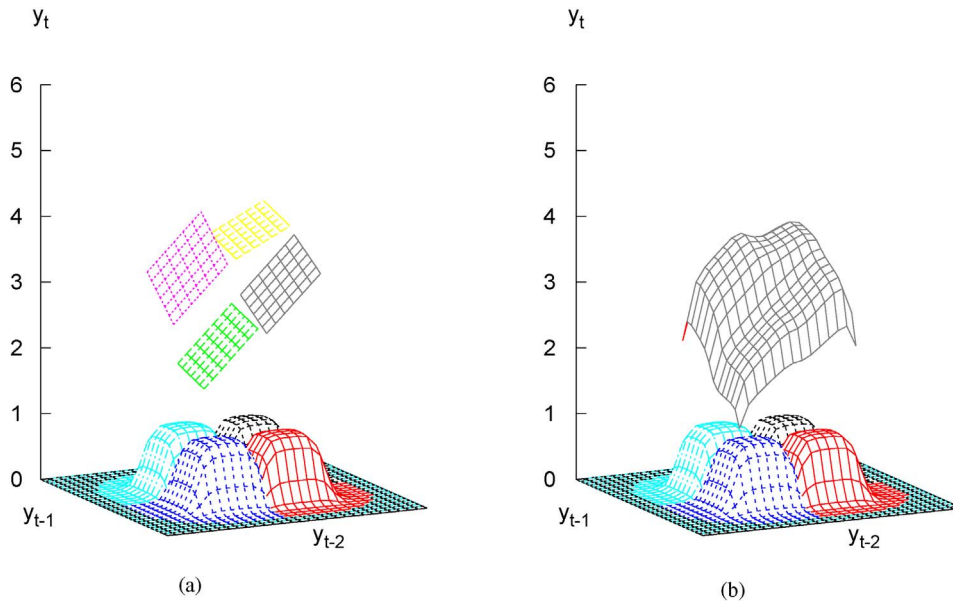


Fig. 5. (a) Four local AR models (or fuzzy rules). (b) L^2 GNN model (or the fuzzy inference system) derived from them.

A. SC Implications

One of the major criticism of the SC models has historically been the lack of mathematical proofs for their statistical properties. This situation is starting to change nowadays. In [25], a coherent modeling strategy which relies on statistical inference is presented to build ANNs for time series. After our results, a similar strategy can be applied to fuzzy systems.

The aforementioned modeling strategy, for example, uses a “bottom-up” strategy to build the model, consisting of the three stages usually applied in statistical modeling: *specification*, *estimation*, and *evaluation*. The three stages rely on well-known statistical procedures: in the specification stage, a variable selection is performed based on assuming the model as linear and applying statistical techniques to choose the variables. Estimation of the parameters, i.e., the number of hidden units, is done by maximum likelihood. This procedure makes it possible to obtain an idea of the uncertainty in the parameter estimates through (asymptotic) standard deviation estimates. This is not possible in most of the empirical approaches adopted in SC. Finally, evaluation of the model is performed through two in-sample misspecification tests: the first one tests for the instability of the parameters and the second one tests the assumption of no serial correlation in the errors.

For the NCSTAR model, an equivalent three-stage procedure is given (the evaluation stage is explained in [27]). This is again directly applicable to fuzzy additive systems based on Proposition 6. In this case, specification is performed through a sequence of Lagrange multiplier tests, which are also used to evaluate the model’s parameter constancy, serial independence and constant error variance.

Another example is given by the main property of the L^2 GNN model: it is asymptotically stationary under mild conditions (see Theorem 1 of [31]). After our Proposition 5, the same can be said about fuzzy additive models. We further develop this result in the remainder of this section.

1) *Asymptotic Stationarity of the Model*: Stationarity of a random process is related to the mean value and variance of the observation data, both of which should be constant over time, and the covariance between the observations x_t and x_{t-d} should only depend on the distance d between the two observations and should not change over time. A time series is *weakly stationary* if $E(x_t) = \mu$ and $cov(x_t, x_{t+h}) = \kappa_h$, $\forall t$, i.e., means and covariances do not depend on time t . A stronger criterion is that the whole distribution (and not only mean and covariance) of the process does not depend on time, and in this case it is called *strictly stationary*. Strong stationarity implies weak stationarity if the second moments of the series exist [22].

If $\{x_t\}$ is strictly stationary, then $P(x_t \in A) = \pi(A)$, $\forall t$, and $\pi(\cdot)$ is called the *stationary distribution* of the series. Obviously the series can only be stationary from the beginning if it is started with the stationary distribution such that $x_0 \sim \pi$. If it is not started with π , e.g., because x_0 is a constant, then we call the series *asymptotically stationary* if it converges to its stationary distribution, that is

$$\lim_{t \rightarrow \infty} P(x_t \in A) = \pi(A). \quad (16)$$

In order to study the asymptotic properties of the threshold AR family of models, the concept of *characteristic equation* was introduced. The characteristic equation of an FRBS can be defined as

$$\lambda^p - c_1 \lambda^{p-1} - c_2 \lambda^{p-2} - \dots - c_p = 0 \quad (17)$$

with

$$c_j = \sum_{i=1}^r \|b_{ij}\| \quad j = 1, \dots, d \quad (18)$$

being b_{ij} the j th coefficient of the i th linear model and r the number of linear models (the number of rules).

It is easy to verify that, when using Gaussian membership functions, model (14) has a finite number of limiting linear models of the form

$$y_t = c_0^{(k)} + c_1^{(k)} y_{t-1} + \dots + c_p^{(k)} y_{t-p} + \varepsilon_t. \quad (19)$$

Obviously, when all the limiting linear models of a model are asymptotically stationary (their roots are inside the unit circle), the model cannot be but asymptotically stationary itself. On the other hand, if one or more of the limiting linear models have roots outside the unit circle, or have unit roots, we must study the model carefully, as its stationarity depends on the membership functions used.

As explained in [31], if the membership functions are “large,” being “active” in half the space, then an explosive limiting regime will lead to asymptotically nonstationary model with probability strictly greater than 0. This is the case with sigmoid functions, and that is the reason why an FRBS using the sigmoid function (5) as membership functions is not guaranteed to be asymptotically stationary.

The problem is different, though, if Gaussian membership functions, defined as

$$\mu_G(\mathbf{x}_i; \boldsymbol{\psi}) = \prod_i \exp\left(-\frac{(x_i - c_i)^2}{2\sigma^2}\right) \quad (20)$$

are used, as these are “small” in the sense that they cover a small fraction of any sufficiently large hypersphere. This is similar to the case for the L^2 GNN, which uses as membership function the difference of two sigmoid functions. The main difference is that the L^2 GNNs membership functions are active in the infinite space left between two parallel hyperplanes, whilst the Gaussian functions are active only inside the limited space of a hypersphere.

Theorem 2: The additive TSK FRBS is asymptotically stationary if it uses Gaussian membership functions with $\frac{1}{\sigma} \neq 0$.

Proof: Immediate after Proposition 5 and Theorem 1 of [31]. With respect to the L^2 GNN, the conditions for an explosive or unit-root limiting linear model to escape to infinity are much simpler: the membership function must have value 1 ($\frac{1}{\sigma} = 0$) or it will always return close to the origin. ■

The effective application of these modeling strategies to fuzzy systems may help overcome the traditional distrust affecting some scientific areas with respect to SC.

B. Statistical Implications

The expression of a threshold model as a set of fuzzy rules has an immediate advantage: the model may be interpretable in terms of human language. Two consequences of this fact allow for an improved use of threshold models.

- 1) There exists the possibility of extracting linguistic knowledge from an already estimated model, in order to contrast it with the knowledge of a human expert. The advantages of this are clear: the human expert could learn from the model and improve her or his knowledge of the problem.
- 2) There exists the possibility of incorporating linguistic knowledge to the models. This allows for a human expert to *teach* the model about specific parts of the

problem which could be hard to capture for the building procedure. As well, it is possible to give initial values to the modeling algorithm based on the expert’s knowledge instead of using other criteria.

As an example, let us revisit the example proposed in [3]. Take a STAR model built to represent the consumer’s expenditure in the U.K. sampled quarterly from 1955 to 1994 [29]. It is composed of three regimes

$$y_t = 0.005 + 2.68y_{t-1} + 0.35y_{t-2} + (-2.59y_{t-3}) \times \Phi_1(y_{t-3}; 1.49, -0.014) + (-1.50y_{t-1} - 0.98y_{t-2}) \times \Phi_2(y_{t-3}; 0.16, 0.016) + \varepsilon_t. \quad (21)$$

After Proposition 2, this STAR model can be seen as an FRBS with three rules. The shape of the membership functions, shown in Fig. 6, is such that it could be understood as a fuzzy logic negation of the most commonly used Gaussian membership function. If we assume that the membership functions ($\Phi(y_{t-3}; 1.495, -0.0136)$ and $\Phi(y_{t-3}; 0.165, 0.0165)$) mean *not low* and *not high*, respectively, the rule base might be written as

$$\begin{aligned} \text{IF } y_{t-3} \text{ IS } \textit{not low} \text{ THEN } y_t &= -2.59y_{t-3} \\ \text{IF } y_{t-3} \text{ IS } \textit{not high} \text{ THEN } y_t &= -1.50y_{t-1} - 0.98y_{t-2} \\ \text{IN ANY CASE } y_t &= 0.005 + 2.68y_{t-1} + 0.35y_{t-2}. \end{aligned}$$

Or, according to the econometric semantics attributed by the authors to the regimes of STAR

$$\begin{aligned} \text{IF } y_{t-3} \text{ IS } \textit{recessive} \text{ THEN } y_t &= -2.59y_{t-3} \\ \text{IF } y_{t-3} \text{ IS } \textit{expansive} \text{ THEN } y_t &= -1.50y_{t-1} - 0.98y_{t-2} \\ \text{IN ANY CASE } y_t &= 0.005 + 2.68y_{t-1} + 0.35y_{t-2}. \end{aligned}$$

Another consequence of this contribution is that the building strategy for threshold models could use the SC advances in automatic model specification and estimation. For example, clustering techniques could be used to fix the number and boundaries of the local regimes of threshold models. As well, a myriad of SC optimization techniques are at hand to be applied in fine-tuning the threshold model parameters.

V. MONTE CARLO EXPERIMENT

From the discussion in Section III, it is clear that the main difference between the set of models considered in this paper is the functional form of the transition or membership function selected, noted Φ in (2). In FRBSs, selecting the shape of Φ is an important problem for which there is no definitive solution, and which is usually related to efficiency and interpretability considerations.

As a further consequence of the equivalence links presented in this paper, we might consider the selection of the membership function of a FRBS from a statistical point of view. In the framework of regime switching models, Monte Carlo simulations are common to evaluate in finite samples if a given method is valid or not, and in this section, we will use them together with information criteria to assess if a chosen model is correctly selected or not.

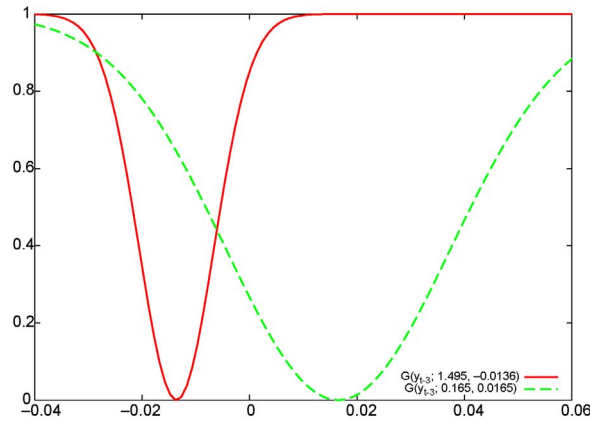


Fig. 6. Exponential transition functions proposed by [29] for the STAR model describing the consumer's expenditure in the U.K.

Let us consider the following general AR regime switching model with two regimes, where $\mathbf{x}_t = (1, y_{t-1}, y_{t-2}, y_{t-3})'$:

$$y_t = (-0.7, 0.2, -0.6, 0.18)' \mathbf{x}_t \cdot \Phi_1(\mathbf{z}_t; \boldsymbol{\psi}_1) + (-0.3, -0.08, 0.5, 0.4)' \mathbf{x}_t \cdot \Phi_2(\mathbf{z}_t; \boldsymbol{\psi}_2) + \varepsilon_t. \quad (22)$$

Depending on the functional form assumed for Φ_1 and Φ_2 , this model can be one of the models studied in this paper. For example, if $\Phi_1 = \Phi_2$ and both are equal to a constant, model (22) becomes an AR model, while if they are defined as in (13), it could be seen as an FRBS.

For example, we can set

$$\begin{aligned} \Phi_1(\mathbf{z}_t; \boldsymbol{\psi}_1) &= \mathbf{I}_{A_1}(y_{t-d}) \\ \Phi_2(\mathbf{z}_t; \boldsymbol{\psi}_2) &= \mathbf{I}_{A_2}(y_{t-d}) \end{aligned} \quad (23)$$

where $A_1 = (-\infty, c]$, $A_2 = (c, \infty)$, and the threshold is fixed at $\boldsymbol{\psi}_1 = \boldsymbol{\psi}_2 = c = -1.0607$. This is a self-exciting TAR model with two regimes.

Then, given a series for which we know that its data generating process is (22) with the transition functions defined by (23), we are interested in knowing empirically which is the functional form of Φ that is most suited to model it. In other words, we are going to test empirically which one is the best model for these series.

We proceed by simulating 100 series from this TAR model. Each of the synthetic series is set to have a length of 500 and we discarded a first portion of the generated data to avoid initialization effects.

These 100 artificially generated series are to be modeled by models described in Section II, i.e., a linear AR model, a TAR, a STAR, an AR-NN, and a FRBS or NCSTAR. In order to compare amongst them, we will use the root mean squared error (RMSE) and the akaike information criterion (AIC) [1] which takes into account not only the goodness of fit but also the number of degrees of freedom of the models. The AIC has no meaning by itself, but it is a way to compare different models, where low values for the AIC correspond to more parsimonious and accurate models and high values correspond to complex and inaccurate models.

Table I shows the values obtained from this Monte Carlo experiment. As we can see, the lowest AIC corresponds to the TAR model, whereas the highest belongs to the linear model.

TABLE I
NUMBER OF PARAMETERS, AVERAGE RMSE, AIC, AND FREQUENCY OF A MODEL ACHIEVING THE SMALLEST AIC OBTAINED BY THE CONSIDERED MODELS OVER 100 SYNTHETIC SERIES

	#Param.	RMSE	AIC	Freq. Min.
AR	4	27.895	227.963	0
TAR	9	22.029	2.607	99
STAR	10	22.047	5.429	1
AR-NN	51	21.632	68.296	0
FRBS	18	22.009	9.710	0

Even though the lowest RMSE corresponds to the AR-NN, the high number of parameters required by this model to give good results makes it one of the worst in terms of AIC.

The fact that the TAR gets the lowest value for the AIC clearly indicates that, for the simulated series, a TAR model is the most effective alternative. This is also supported by the frequency of obtaining the smallest value of the AIC, which reaches the 99% of the instances for the TAR model.

Apart from demonstrating the importance of correctly selecting the shape of the transition functions when using the family of models studied in this paper, this simulation shows that statistical tools such as information criteria can be used to properly identify, on average, the correct regime switching model for a given time series. From the point of view of SC models, this represents still another consequence of the equivalence relationships established above. It points to a better understanding of the classical statistic time series analysis (which in turn might lead to better neural or fuzzy-based time series models).

VI. CONCLUSION

In this paper, we have shown how some neural-based regime switching models, coming from the econometric approach to time series analysis, are closely related to FRBSs. In fact, we proved that a fuzzy rule based model can be seen as a generalization of the TAR, STAR, AR-NN, L^2 GNN, and NCSTAR models.

These results are another step in a continuing effort toward the merger of the statistical-based time series analysis and the SC methods and techniques devoted to this problem. Practical advances stemming from these theoretical results are being developed and are expected to produce improvements in the

resolution of the time series analysis and forecasting problem. Some immediate consequences were demonstrated through the formal proof of the asymptotic stationarity of the additive TSK FRBSs and through Monte Carlo simulations concerning the selection of the transition function.

ACKNOWLEDGMENT

The research included in this paper is part of the first author's Ph.D. dissertation at the Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain.

REFERENCES

- [1] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974.
- [2] H. C. Anderson, A. Lotfi, L. C. Westphal, and J. R. Jang, "Comments on 'functional equivalence between radial basis function networks and fuzzy inference systems' and reply," *IEEE Trans. Neural Netw.*, vol. 9, no. 6, pp. 1529–1532, Nov. 1998.
- [3] J. L. Aznarte, J. M. Benítez, and J. L. Castro, "Smooth transition autoregressive models and fuzzy rule-based systems: Functional equivalence and consequences," *Fuzzy Sets Syst.*, vol. 158, no. 24, pp. 2734–2745, 2007.
- [4] L. Bachelier, *Calcul Des Probabilités*. Paris, France: Garthier-Villars, 1912.
- [5] J. M. Benítez, J. L. Castro, and I. Requena, "Are artificial neural networks black boxes?" *IEEE Trans. Neural Netw.*, vol. 8, no. 5, pp. 1156–1164, Sep. 1997.
- [6] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day, 1970.
- [7] H. S. Byun and K. H. Lee, "A decision support system for the selection of a rapid prototyping process using the modified topsis method," *Int. J. Adv. Manuf. Technol.*, vol. 26, nos. 11–12, pp. 1338–1347, 2005.
- [8] J. L. Castro, C. J. Mantas, and J. M. Benítez, "Interpretation of artificial neural networks by means of fuzzy rules," *IEEE Trans. Neural Netw.*, vol. 13, no. 1, pp. 101–116, Jan. 2002.
- [9] R. Chen and R. S. Tsay, "Functional-coefficient autoregressive models," *J. Am. Statistical Assoc.*, vol. 88, pp. 298–308, Mar. 1993.
- [10] F. V. Coito, L. B. Palma, and R. N. da Silva, "Robust fault diagnosis approach using analytical and knowledge based techniques applied to a water tank system," *Int. J. Eng. Intell. Syst. Electr. Eng. Commun.*, vol. 13, no. 4, pp. 237–244, 2005.
- [11] M. C. P. de Souto, T. B. Ludermir, and W. R. de Oliveira, "Equivalence between ram-based neural networks and probabilistic automata," *IEEE Trans. Neural Netw.*, vol. 16, no. 4, pp. 996–999, Jul. 2005.
- [12] M. Fariñas and C. E. Pedreira, "Mixture of experts and local-global neural networks," in *Proc. ESANN*, 2003, pp. 331–336.
- [13] K. J. Hunt, R. Haas, and R. Murray-Smith, "Extending the functional equivalence of radial basis function networks and fuzzy inference systems," *IEEE Trans. Neural Netw.*, vol. 7, no. 3, pp. 776–781, May 1996.
- [14] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive mixtures of local experts," *Neural Comput.*, vol. 3, no. 1, pp. 79–87, 1991.
- [15] J.-S. R. Jang and C.-T. Sun, "Functional equivalence between radial basis function networks and fuzzy inference systems," *IEEE Trans. Neural Netw.*, vol. 4, no. 1, pp. 156–159, Jan. 1993.
- [16] R. I. John and P. R. Innocent, "Modeling uncertainty in clinical diagnosis using fuzzy logic," *IEEE Trans. Syst., Man, Cybern. B*, vol. 35, no. 6, pp. 1340–1350, Dec. 2005.
- [17] A. Y. Khinchin, "Korrelationstheorie der stationären stochastischen prozesse," *Math. Ann.*, vol. 109, no. 1, pp. 604–615, 1934.
- [18] E. Kolman and M. Margaliot, "Are artificial neural networks white boxes?" *IEEE Trans. Neural Netw.*, vol. 16, no. 4, pp. 844–852, Jul. 2005.
- [19] E. Kolman and M. Margaliot, "Extracting symbolic knowledge from recurrent neural networks: A fuzzy logic approach," *Fuzzy Sets Syst.*, vol. 160, no. 2, pp. 145–161, Jan. 2009.
- [20] A. N. Kolmogorov, "Über die analytischen methoden in der wahrscheinlichkeitsrechnung," *Math. Ann.*, vol. 104, no. 604, pp. 415–458, 1931.
- [21] I. Lee, B. Kosko, and W. F. Anderson, "Modeling gunshot bruises in soft body armor with an adaptive fuzzy system," *IEEE Trans. Syst., Man, Cybern. B*, vol. 35, no. 6, pp. 1374–1390, Dec. 2005.
- [22] F. Leisch, A. Trapletti, and K. Hornik, "Stationarity and stability of autoregressive neural network processes," in *Proc. 1998 Conf. Adv. Neural Inform. Process. Syst. II*, 1999, pp. 267–273.
- [23] H.-X. Li and C. L. P. Chen, "The equivalence between fuzzy logic systems and feedforward neural networks," *IEEE Trans. Neural Netw.*, vol. 11, no. 2, pp. 356–365, Mar. 2000.
- [24] X.-B. Liang, "Equivalence between local exponential stability of the unique equilibrium point and global stability for Hopfield-type neural networks with two neurons," *IEEE Trans. Neural Netw.*, vol. 11, no. 5, pp. 1194–1196, Sep. 2000.
- [25] M. C. Medeiros, T. Teräsvirta, and G. Rech, "Building neural network models for time series: A statistical approach," *J. Forecasting*, vol. 25, no. 1, pp. 49–75, 2006.
- [26] M. C. Medeiros and A. Veiga, "A hybrid linear-neural model for time series forecasting," *IEEE Trans. Neural Netw.*, vol. 11, no. 6, pp. 1402–1412, Nov. 2000.
- [27] M. C. Medeiros and A. Veiga, "Diagnostic checking in a flexible nonlinear time series model," *J. Time Ser. Anal.*, vol. 24, no. 4, pp. 461–482, 2003.
- [28] M. C. Medeiros and A. Veiga, "A flexible coefficient smooth transition time series model," *IEEE Trans. Neural Netw.*, vol. 16, no. 1, pp. 97–113, Jan. 2005.
- [29] N. Ocal and D. R. Osborn, "Business cycle nonlinearities in U.K. consumption and production," *J. Appl. Econometrics*, vol. 15, no. 1, pp. 27–43, 2000.
- [30] A. K. Palit and D. Popovic, *Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications (Advances in Industrial Control)*. New York: Springer-Verlag, 2005.
- [31] M. Suarez-Farinas, C. E. Pedreira, and M. C. Medeiros, "Local global neural networks: A new approach for nonlinear time series modeling," *J. Am. Stat. Assoc.*, vol. 99, pp. 1092–1107, Dec. 2004.
- [32] T. Teräsvirta, D. van Dijk, and M. C. Medeiros, "Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A reexamination," *Int. J. Forecasting*, vol. 21, no. 4, pp. 755–774, 2005.
- [33] T. Teräsvirta, "Specification, estimation and evaluation of smooth transition autoregressive models," *J. Am. Stat. Assoc.*, vol. 89, no. 425, pp. 208–218, 1994.
- [34] H. Tong, "On a threshold model," in *Pattern Recognition and Signal Processing*. Amsterdam, The Netherlands: Sijthoff and Noordhoff, 1978, pp. 575–586.
- [35] C.-H. Wang and J.-S. Wen, "On the equivalence of a table lookup (TL) technique and fuzzy neural network (FNN) with block pulse membership functions (BPMFs) and its application to water injection control of an automobile," *IEEE Trans. Syst., Man, Cybern. C, Applicat. Rev.*, vol. 38, no. 4, pp. 574–580, Jul. 2008.
- [36] Y. Xia and W. K. Li, "On single-index coefficient regression models," *J. Am. Stat. Assoc.*, vol. 94, no. 448, pp. 1275–1285, 1999.
- [37] G. U. Yule, "On the method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers," *Philosophical Trans. A.*, vol. 226, pp. 267–298, Jul. 1927.
- [38] L. A. Zadeh, "Soft computing and fuzzy logic," *IEEE Softw.*, vol. 11, no. 6, pp. 48–56, Nov. 1994.

José Luis Aznarte received the Ph.D. degree from the University of Granada, Granada, Spain, with a thesis entitled "Modelling time series through fuzzy rule-based models: A statistical approach," in 2008. While pursuing Ph.D., he completed research visits with the Department of Economics, PUC-Rio, Rio de Janeiro, Brazil, and the Department of Econometrics, Erasmus University, Rotterdam, The Netherlands.

He is currently a Post-Doctoral Researcher with the Renewable Energy Research Group, Center for Energy and Processes of the MINES ParisTech Engineering School, Paris, France, where he is engaged in the European projects ANEMOS.Plus (FP6) and Safewind (FP7).

Dr. Aznarte is a member of the Soft Computing and Intelligent Information Systems Research Group. In the past, he participated in the Spanish Antarctic Program where he coordinated the installation of permanent seismometric stations as a member of the Andalusian Geophysics Institute, Granada.



José Manuel Benitez (M'98) received the M.S. and Ph.D. degrees in computer science, both from the Universidad de Granada, Granada, Spain.

He is currently an Associate Professor with the Department of Computer Science and Artificial Intelligence, CITIC, Universidad de Granada. He is a member of the research group "Soft Computing and Intelligent Information Systems" and the head of the "Distributed Computational Intelligence and Time Series" research laboratory. He is an active Researcher in the computational intelligence field

where his work covers the whole spectrum from foundations to applications in a number of engineering and scientific areas. His current research interests include time series analysis and modeling, distributed/parallel computational intelligence, data mining, and statistical learning theory.

Dr. Benitez serves on the board of several international journals and is the Editor-in-Chief of the *International Journal of Computational Intelligence Research*. He is a member of a number of scientific associations, including the IEEE Computational Intelligence Society, EUSFLAT, and ISAI.