

Modelación de Series Temporales usando ANFIS

Juan D. Velásquez, Isaac Dyner R

Escuela de Sistemas, Universidad Nacional de Colombia, Sede Medellín
 {jdvelasq, idyner}@unalmed.edu.co

Reinaldo C. Souza

Departamento de Engenharia Elétrica, Pontificia Universidade Católica, RJ
 reinaldo@ele.puc-rio.br

Resumen

En este artículo, se examina la modelación de series temporales no lineales usando ANFIS. Se discute en primer lugar, su relación con otros modelos estadísticos, y se presenta su formulación con una generalización de estos; posteriormente, se analiza una aproximación metodológica para su especificación inicial basada en criterios estadísticos; y finalmente se presenta el caso del pronóstico del precio de la electricidad en el Brasil en el mercado de corto plazo, como un ejemplo de aplicación. Como resultado del estudio, se sugiere que la metodología propuesta sea parte integral de la modelación de series temporales usando ANFIS.

Palabras clave: Sistemas Difusos, Series temporales no lineales, ANFIS.

1. Introducción

En la modelación estadística, una serie temporal es definida como un conjunto de observaciones ordenadas en el tiempo:

$$\dots, y_{-1}, y_0, y_1, \dots, y_T, y_{T+1}, \dots \quad (1)$$

de las cuales sólo se dispone de la realización $\{y_t\}_1^T$. Dichas observaciones son el resultado de la aplicación de un mecanismo generador a un conjunto de variables causales que explican el comportamiento de y_t .

La hipótesis fundamental de la modelación, es que dicho mecanismo puede ser representado por una función computable y desconocida $f(\cdot)$ que es aplicada al conjunto de regresores $\mathbf{x}_t = \{x_t^{(1)}, \dots, x_t^{(U)}\}$:

$$y_t = f(\mathbf{x}_t) + \nu_t \quad (2)$$

donde el término ν_t representa una serie de ruidos aleatorios independientes e idénticamente distribuidos, que siguen una distribución con media cero y varianza σ_ν^2 . Los regresores \mathbf{x}_t en (2), están formados por un subconjunto de las variables causales visibles y cuantificables, que influyen el mecanismo generador de las observaciones. La serie de ruidos estocásticos $\{\nu_t\}_1^T$, representa el efecto de todas aquellas variables causales, cuantitativas o cualitativas, visibles o no visibles, no incluidas en el conjunto de regresores.

Partiendo de la premisa fundamental de que $f(\cdot)$ en (2) existe y es computable, el objetivo fundamental de la modelación de series temporales es construir una función $\hat{f}(\cdot)$ que la aproxime con algún grado de precisión, valiéndose de los datos disponibles $\{y_t, \mathbf{x}_t\}_1^T$, tal que:

$$y_t = \hat{f}(\mathbf{x}_t) + \hat{\nu}_t \quad (3)$$

Ello implica que el proceso generador de la serie temporal es estable, y que la serie de observaciones $\{y_t, \mathbf{x}_t\}_1^T$ contiene suficiente información para la construcción de $\hat{f}(\cdot)$.

En la literatura se han presentado diferentes modelos paramétricos para especificar $\hat{f}(\cdot)$, los cuales se basan en asumir formas preestablecidas de no linealidad, cuya utilidad depende del grado de ajuste entre el modelo usado y el proceso generador de la serie temporal. No obstante, su mayor crítica ha sido asumir la forma de la no linealidad, ya que en la práctica es común que no existan leyes físicas o económicas que guíen completamente la escogencia de un determinado modelo paramétrico. De ahí que los modelos no paramétricos y no lineales de regresión surjan como una alternativa viable para la representación del proceso generador de series temporales con estas características.

De dichas metodologías, los modelos de regresión comúnmente usados en Inteligencia Computacional (sistemas de inferencia borrosa y las redes neuronales artificiales), han venido siendo utilizados en la modelación de series temporales no lineales, con resultados exitosos en los casos de aplicación. No obstante, el volumen de artículos publicados en distintas revistas especializadas muestra que se cuenta con mucha más experiencia en el uso de redes neuronales artificiales en la solución de problemas de pronóstico de series temporales, que en el uso de los sistemas de inferencia borrosa. Igualmente, parece que la complejidad en el proceso de preparación de modelos de redes neuronales ha desviado la atención de la mayor parte de los investigadores y practicantes hacia los aspectos propios de este problema, descuidándose en alguna medida otros aspectos de igual importancia, relacionados en buena parte con la modelación estadística de series temporales.

De acuerdo con nuestro conocimiento, la construcción de modelos de series temporales usando sistemas de inferencia difusa se aborda con la misma metodología para la construcción de modelos de regresión, ignorándose muchos de los desarrollos de la modelación estadística; más aún, no hemos encontrado evidencias de la existencia de trabajos relacionados con la interpretación de los sistemas de inferencia difusa en el contexto de la modelación estadística de series temporales.

En consecuencia de lo anterior, este artículo se centra en la construcción de sistemas adaptativos

de inferencia neurodifusas [ANFIS] [15] interpretados como modelos de series temporales no lineales dentro de un marco estadístico.

Para lograr este objetivo, la sección 2 presenta una introducción muy rápida a los sistemas de inferencia borrosa; posteriormente en la sección 3, se presentan algunos modelos paramétricos no lineales, y como ellos pueden ser derivados como casos particulares de ANFIS; en la sección 4, se discuten como la formulación de esta clase de modelos centrada en un marco estadístico permite su mejor entendimiento, y una determinación clara de sus ventajas y desventajas, bajo la forma de una aproximación metodológica para la formulación preliminar de estos modelos basada en herramientas estadísticas; seguidamente en la sección 5, se realiza la aplicación de la aproximación metodológica al caso real del pronóstico del precio de electricidad en el Brasil. Finalmente en la sección 6, se presenten las principales conclusiones obtenidas en este estudio.

2. Sistemas adaptativos de inferencia neurodifusa (ANFIS)

En la teoría clásica de conjuntos, un subconjunto S del universo U puede ser definido como una función que relaciona cada elemento x del universo, con un elemento del conjunto discreto $\{0, 1\}$, esto es:

$$S : U \rightarrow \{0, 1\} \quad (4)$$

tal que el cero es usado para indicar que el elemento no pertenece con certeza a S , y la unidad para indicar certeza de la pertenencia.

En el caso de los conjuntos difusos [32], S es una función que relaciona cada elemento de U con un punto en el intervalo continuo $[0, 1]$, tal que el cero indica la certeza de la no pertenencia del elemento x al conjunto S , la unidad indica la certeza absoluta de dicha pertenencia; y los valores entre cero y uno, representan grados parciales de pertenencia. En este caso S recibe el nombre de conjunto borroso.

La función que vincula cada elemento del universo en un valor en el intervalo $[0, 1]$, $\mu_S(x)$, recibe el nombre de función de pertenencia, y en el caso de universos que corresponden a variables continuas, puede ser representada usando diferentes funciones matemáticas; entre ellas se encuentran

la función S o sigmoidea:

$$S(u; \gamma, c) = \frac{1}{1 + \exp[-\gamma(u - c)]} \quad (5)$$

la función Z :

$$Z(u; \gamma, c) = 1 - S(u; \gamma, c) \quad (6)$$

o la función Φ o campana generalizada:

$$\Phi(x; \alpha, \beta, \gamma) = \frac{1}{1 + \left| \frac{u-\gamma}{\alpha} \right|^{2|\beta|}} \quad (7)$$

Un sistema difuso está conformado por variables de entrada y salida, reglas difusas, y un método de inferencia que permita obtener los valores de las variables de salida, cuando son conocidos los valores de las variables de entrada.

Un sistema difuso conformado por reglas de Takagi–Sugeno [27, 28, 29] puede ser interpretado como una red neurodifusa conocida como ANFIS [15]. La Figura 1 presenta la arquitectura de ANFIS para un sistema de inferencia de dos variables difusas x y y :

$$\begin{aligned} \text{if } x \in A_1 \wedge y \in B_1 &\Rightarrow z = p_1x + q_1y + r_1 \\ \text{if } x \in A_1 \wedge y \in B_2 &\Rightarrow z = p_2x + q_2y + r_2 \\ \text{if } x \in A_2 \wedge y \in B_1 &\Rightarrow z = p_3x + q_3y + r_3 \\ \text{if } x \in A_2 \wedge y \in B_2 &\Rightarrow z = p_4x + q_4y + r_4 \end{aligned} \quad (8)$$

donde el dominio de x está cubierto por los conjuntos borrosos A_1 y A_2 , y el dominio de y por B_1 y B_2 . El proceso de inferencia para una entrada (x, y) implica: calcular $\mu_{A_1}(x)$, $\mu_{A_2}(x)$, $\mu_{B_1}(y)$ y $\mu_{B_2}(y)$ usando (7); estimar la inferencia Producto para cada regla $w_j = \mu_{A_j}(x) \cdot \mu_{B_j}(y)$; establecer el porcentaje que cada regla aporta a la solución final $\bar{w}_j = \frac{w_j}{\sum_{i=1}^N w_i}$; y finalmente, calcular el resultado del sistema como: $\sum_i \bar{w}_i f_i$.

Nótese que en (8), las variables que conforman los antecedentes de las reglas son las mismas que conforman sus consecuentes; y que el dominio de cada variable está cubierto por al menos dos conjuntos borrosos, de acuerdo con la definición de reglas de Takagi–Sugeno. Usualmente, la función especificada para el consecuente es la combinación lineal de las entradas, pero en el caso general es posible usar funciones más complejas. La ecuación de la campana generalizada (7) es una elección común para modelar las funciones de pertenencia con que se especifican los conjuntos borrosos.

El sistema formado por (8) puede incluir información proveniente del experto en la forma de

reglas difusas, así como aprendida a partir de información histórica de la serie temporal que se está modelando. Esto se logra a través del afinamiento de la cantidad y las funciones de pertenencia asociadas a cada una de las variables explicativas.

La capacidad de ANFIS para aproximar funciones no lineales, está determinada por el número de conjuntos borrosos asignado a cada variable independiente; esto trae como desventaja que problemas con un número moderado de variables independientes, requieran una gran cantidad de información para realizar el entrenamiento.

Un problema fundamental en la aplicación de la metodología es la selección de las variables que conformarán las reglas difusas, y el número de conjuntos borrosos asignados a cada variable. Jang [16] propone el uso del algoritmo CART [7] para determinar la estructura inicial de ANFIS y posteriormente afinar los parámetros del modelo usando un algoritmo de optimización.

3. ANFIS como un modelo de series temporales

3.1. Modelos lineales de Box y Jenkins

La formulación de ANFIS, analizada en el contexto del modelamiento de series de tiempo, permite que este sea interpretado como una generalización no lineal de los modelos autoregresivos. En el caso lineal, $\hat{f}(\cdot)$ definida en (3) es la ecuación paramétrica lineal:

$$\hat{f}(\mathbf{x}_t) = g(\mathbf{x}_t | \{\hat{a}_j\}_0^U) = \hat{a}_0 + \sum_{j=1}^U \hat{a}_j \cdot x_t^{(j)} \quad (9)$$

donde los coeficientes $\{\hat{a}_j\}_0^U$ son estimados usando mínimos cuadrados:

$$\{\hat{a}_j\}_0^U = \arg \min_{\{a_j\}_0^U} \sum_{t=1}^T [y_t - g(\mathbf{x}_t | \{a_j\}_0^U)]^2 \quad (10)$$

En la familia de modelos de Box y Jenkins[5], el vector de regresores \mathbf{x}_t (9) toma diferentes formas, de acuerdo con el tipo de modelo considerado. En el caso general, \mathbf{x}_t está formada por valores rezagados de y_t , $\{y_{t-j}\}_{j=1}^P$, los términos aleatorios $\{\nu_{t-j}\}_{j=1}^Q$, y los rezagos de otras variables

causales $\{u_{t-j}^{(i)}\}_{i=1, j=1}^{C, P_i}$. El uso de esta clase de modelos requiere que $\{y_t\}_1^T$ sea estacionaria en el sentido homogéneo, esto es, que su valor esperado y su varianza sean constantes en el tiempo. Ello implica que toda la serie pertenece a la misma distribución de probabilidades. En algunas situaciones una transformación de la serie temporal puede ser adecuada. Mills [23, p. 32] indica que para series con varianza variable, la transformación de potencias de Box y Cox:

$$y_t = g(z_t, \lambda) = \begin{cases} \log z_t & , \lambda = 0, \\ \frac{z_t^\lambda - 1}{\lambda} & , \lambda \neq 0 \end{cases} \quad (11)$$

puede corregir parcialmente este problema. En (11), λ es estimado de tal forma que maximice la función de verosimilitud:

$$L(\lambda) = -\frac{T}{2} \log \left\{ \frac{\sum_{t=1}^T (y_t^\lambda - (\bar{y})^{2\lambda})}{T} + (\lambda - 1, 0) \sum_{t=1}^T \ln y_t \right\} \quad (12)$$

con

$$\bar{y}^\lambda = \sum_{t=1}^T \frac{y_t^\lambda}{T}$$

Esta familia de modelos sólo es exitosa en la medida de que $g(\mathbf{x}_t | \{\hat{a}_j\}_0^U)$ se aproxime a la función real $f(\cdot)$ definida en (2). Ello implica que en la función aproximada $f(\cdot)$ las relaciones entre los regresores \mathbf{x}_t y y_t son lineales, y segundo, que sus relaciones son constantes en el tiempo, ya que los coeficientes $\{\hat{a}_j\}_0^U$ son independientes de t .

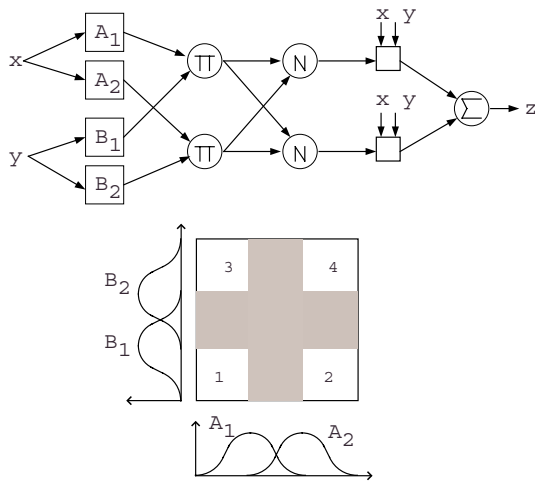


Figura 1. Arquitectura Típica para ANFIS con reglas de Takagi-Sugeno y partición inducida sobre el espacio de entrada

3.2. Modelos SETAR, SETMA y SETARMA

Una generalización de (9) al caso no lineal [31, Sección 3.3.1.1] es dividir el dominio de una variable de transición z_t en $\{R_k\}_1^K$ regiones disyuntas, tal que $f(\cdot)$ puede ser aproximada por:

$$\text{if } z_t \in R_k \Rightarrow y_t = g_k(\mathbf{x}_t | \{\hat{a}_j^{(k)}\}_0^U) + \nu_t^{(k)} \quad (13)$$

La determinación de la cantidad de regiones K , sus límites, y los valores de los parámetros se hace a partir de los datos $\{\mathbf{x}_t, y_t\}_1^T$. El principio fundamental de (13) es dividir la serie temporal generada por un sistema estocástico complejo en subsistemas más simples. Toman su nombre de acuerdo al tipo de modelo de Box y Jenkins con que se especifique el consecuente de las reglas. Cada una de las regiones R_k representa diferentes regímenes de comportamiento lineal dentro del dominio de la variable de decisión z_t . La continuidad en y_t es obtenida a través de la variable aleatoria ν_t . Este modelo (13) puede tomar la forma del modelo paramétrico lineal:

$$y_t = \sum_{k=1}^K I_k(z_t) \cdot \left(g_k(\mathbf{x}_t | \{\hat{a}_j^{(k)}\}_0^U) + \nu_t^{(k)} \right) \quad (14)$$

donde I_k es una función indicadora tal que:

$$I_k(z) = \begin{cases} 1 & \text{Si } z \in R_k \\ 0 & \text{en cualquier otro caso} \end{cases} \quad (15)$$

donde R_k es la k -ésima región en que se encuentra particionada z_t . En este tipo de modelos, la transición entre uno y otro modelo autoregresivo se realiza de forma brusca, sin permitir que existan regiones de transición entre un régimen y otro.

3.3. Modelos LSTAR y ESTAR

Baccon y Watts [2] y Maddala [22] reconocen el hecho de que el cambio de régimen en (14) obedece más a un cambio gradual que a una transición brusca, por lo que incorporan dentro del modelo, una función de transición F (un tratamiento formal es presentado por Tong [31, sec. 3.3.3] y Granger y Teräsvirta [11]). Para el caso del modelo con dos regímenes:

$$y_t = [1 - F(z_t)] \cdot \left(a_0^{(1)} + \sum_{j=1}^{P_1} a_j^{(1)} x_t^{(j)} \right) + F(z_t) \cdot \left(a_0^{(2)} + \sum_{j=1}^{P_2} a_j^{(2)} x_t^{(j)} \right) + \nu_t \quad (16)$$

Para la formulación de $F(\cdot)$ se asume que ella sólo puede tomar valores en el intervalo $[0, 1]$. El modelo representado por (16) es conocido como STAR (smooth transition regression model). Cuando $F(\cdot)$ es especificada como la función logística o S definida en (5), el modelo se conoce como LSTAR (logistic smooth transition regression) Cuando γ tiende a infinito, la función de transición converge a la función indicadora (15), tal que el modelo se reduce a un SETAR donde el parámetro c es el límite entre las dos regiones en que se divide el dominio de la variable de transición z . Si $F(\cdot)$ es especificada como la función exponencial:

$$F(u; \gamma, c) = 1 - \exp[-\gamma(u - c)^2] \quad (17)$$

el modelo resultante es llamado ESTAR (exponential smooth transition regression) Un problema de esta función es que converge a un valor constante cuando γ tiende a cero; cuando tiende a infinito se reduce a un modelo SETAR.

Granger [12] sugiere que la especificación de modelos no lineales debe seguir un procedimiento que vaya de lo general a lo específico, iniciando con un modelo simple o restringido que gana complejidad mientras las pruebas de diagnóstico indiquen que el modelo actual es inadecuado para representar el proceso generador de la serie temporal. Teräsvirta[30] sugiere el siguiente procedimiento empírico de especificación de modelos STAR:

1. Se especifica un modelo autoregresivo de orden p para la serie temporal estudiada
2. Se realiza una prueba estadística donde se confronta la hipótesis de linealidad del modelo contra la alternativa de usar un modelo STAR. Si la linealidad es rechazada, se selecciona la variable de transición z_t y la función de transición $F(\cdot)$.
3. Se estiman los parámetros del modelo.
4. Se ejecutan las pruebas de diagnóstico
5. Se plantean modelos alternativos si es necesario.

3.4. Interpretación de ANFIS

Una generalización directa de la partición realizada sobre el dominio de z_t en (13), es hacer que dichas regiones se traslapen, permitiendo que sobre la intersección tengan influencia los modelos

autoregresivos de cada una de ellas. Un mecanismo para lograr esto, es representar cada región por un conjunto borroso, el cual puede ser modelado usando (7). De esta forma, un modelo con dos regiones puede escribirse como:

$$\begin{aligned} \text{if } z_t \in A_1 &\Rightarrow y_t = g(\mathbf{x}_t | \{\hat{a}_j^{(1)}\}_0^U) + \nu_t \\ \text{if } z_t \in A_2 &\Rightarrow y_t = g(\mathbf{x}_t | \{\hat{a}_j^{(2)}\}_0^U) + \nu_t \end{aligned} \quad (18)$$

con $g(\cdot)$ definida en (9). El sistema de reglas formado por (18) es un caso particular de (8).

La variable de transición z_t puede ser especificada en el caso más simple, como una de las componentes del vector \mathbf{x}_t , pero funciones más complejas pueden ser fácilmente incorporadas dentro del modelo, tales como una combinación lineal de algunas componentes de \mathbf{x}_t . En el caso general, los antecedentes de (18) pueden contener más de una variable de transición, e inducen una partición compleja del dominio de las variables de transición; diversos esquemas han sido presentados en la literatura con el fin de lograr particiones más eficientes del espacio del dominio, en los cuales a cada una de ellos se les asignan dos o más conjuntos borrosos, que pueden ser simétricos o asimétricos, induciendo así diferentes esquemas de partición.

La formulación presentada en (18) difiere de (8) en que se incluye en el consecuente de cada regla a ν_t , el cual es el mismo término para todas reglas; igualmente (18) permite generalizar la formulación de ANFIS, al caso en que los conjuntos de variables que conforman los antecedentes y los consecuentes pueden ser diferentes. La interpretación desarrollada vincula esta clase de modelos de inferencia borrosa con otros modelos comúnmente usados en la modelación estadística; en consecuencia, es posible usar los desarrollos teóricos realizados para estos modelos estadísticos en la especificación y validación de ANFIS.

El proceso de cálculo de y_t en (18), permite que dicho sistema sea transformado a la ecuación paramétrica lineal definida en (9), tal que los coeficientes $\{\hat{a}_j\}_0^U$ son calculados como:

$$\hat{a}_j = \phi_j(z_t) = k_1 \times \hat{a}_j^{(1)} + k_2 \times \hat{a}_j^{(2)} \quad (19)$$

donde:

$$k_j = \frac{m(z_t; \alpha_j, \beta_j, \gamma_j)}{m(z_t; \alpha_1, \beta_1, \gamma_1) + m(z_t; \alpha_2, \beta_2, \gamma_2)}, \quad j = 1, 2$$

De esta forma, (18) es equivalente al modelo autoregresivo:

$$y_t = \phi_0(z_t) + \sum_{j=1}^P \phi_j(z_t) \times x_t^{(j)} \quad (20)$$

cuyo comportamiento depende de la variable de transición z_t , y para el cual los coeficientes $\{a_k\}_0^K$ definidos en (9) son una función continua de la variable de transición z_t . En consecuencia, existe un modelo autoregresivo lineal por cada punto del dominio de z_t , mientras que para los modelos SETAR, SETMA, SETARMA, LSTAR y ESTAR existe un modelo por cada región en que se particiona el conjunto de regresores; y en el caso de los modelos de Box y Jenkins existe un único modelo autoregresivo que se ajusta a la totalidad de los datos. El modelo definido en (20) es atribuido a Prestley [25].

Si en (18), el conjunto borroso A_1 es modelado usando la función $Z(u; \gamma_1, c_1)$ definida en (6) y el conjunto borroso A_2 usando la función $S(u; \gamma_2, c_2)$ definida en (5), y_t es calculado como:

$$y_t = k_1 \left(a_0^{(1)} + \sum_{j=1}^P a_j^{(1)} x_t^{(j)} \right) + k_2 \left(a_0^{(2)} + \sum_{j=1}^P a_j^{(2)} x_t^{(j)} \right) + \nu_t \quad (21)$$

con:

$$k_1 = \frac{Z(z_t; \gamma_1, c_1)}{Z(z_t; \gamma_1, c_1) + S(z_t; \gamma_2, c_2)}$$

$$k_2 = \frac{S(z_t; \gamma_2, c_2)}{Z(z_t; \gamma_1, c_1) + S(z_t; \gamma_2, c_2)}$$

Si se imponen las restricciones $c_1 = c_2 = c$ y $\gamma_1 = \gamma_2 = \gamma$ a los parámetros de $S(\cdot)$ y $Z(\cdot)$, es fácilmente demostrable que:

$$S(z_t; \gamma, c) + Z(z_t; \gamma, c) = 1 \quad (22)$$

de tal forma que (21) se puede reescribir como:

$$y_t = [1 - S(z_t; \gamma, c)] \left(a_0^{(1)} + \sum_{j=1}^P a_j^{(1)} x_t^{(j)} \right) + S(z_t; \gamma, c) \left(a_0^{(2)} + \sum_{j=1}^P a_j^{(2)} x_t^{(j)} \right) + \nu_t \quad (23)$$

la cual corresponde al modelo LSTAR definido en (16). Si γ se hace cero, la función de transición se hace igual a 0,5 de tal forma que (23) converge al modelo lineal:

$$y_t = \frac{1}{2} [a_0^{(1)} + a_0^{(2)}] + \frac{1}{2} \sum_{j=1}^P [a_j^{(1)} + a_j^{(2)}] x_t^{(j)} + \nu_t \quad (24)$$

Si γ tiende a infinito, (23) convergerá a un modelo SETAR.

Es posible obtener funciones de transición más complejas al liberar las restricciones impuestas a los parámetros γ y c , o al usar la función de campana generalizada (7) para modelar los conjuntos borrosos en (18).

4. Estrategia general para la formulación de modelos

En el contexto de la modelación estadística de series temporales, la estrategia general para la especificación y aceptación de modelos fue formulada originalmente por Box y Jenkins [5], la cual ha sido ajustada por diferentes autores (principalmente Tong [31], Granger y Terävirta [11] y Harvey [13]) para ser aplicada en la construcción de modelos no lineales. Los pasos fundamentales son presentados en la Tabla 1.

4.1. Paso 1: Identificación

Durante la fase de identificación, se realiza un análisis exploratorio de la serie que tiene como objetivos la determinación de los regresores \mathbf{x}_t con los que se construirá el modelo de inferencia difusa, la realización de una transformación de los datos que permita al modelo capturar sus propiedades estadísticas, y finalmente, la determinación de las principales propiedades estadísticas de la serie temporal. Como resultado final del proceso de identificación, se especifican varios modelos difusos alternativos para representar el proceso generador de la serie temporal.

4.1.1. Determinación de los regresores

Uno de los principales problemas en la modelación de $\{y_t\}_1^T$ es la selección correcta del conjunto de regresores \mathbf{x}_t , ya que su elección inicial y los resultados del análisis exploratorio preliminar que permite descartar variables irrelevantes, se basan profundamente en las decisiones subjetivas del modelador y su apreciación subjetiva del ambiente donde se genera la serie temporal [8]. En consecuencia, este proceso tiene una alta componente de juicio experto que puede ser viciado por las desviaciones y errores sistemáticos derivados de no realizar un proceso formal de razonamiento, ya que se encuentra sometido a las limitaciones propias de la mente humana (este tópico

es ampliamente discutido por Hogarth [14]). Formalmente, el pronosticador establece un conjunto de regresores \mathbf{x}_t^* , potencialmente explicativos del comportamiento de y_t , basado en el análisis del proceso generador de la serie temporal, y apoyado en leyes físicas o teoría económica:

$$\mathbf{x}_t^* = ((x_t^*)^{(1)}, \dots, (x_t^*)^{(D)}) \quad (25)$$

posteriormente el conjunto de regresores \mathbf{x}_t es obtenido como la solución de un problema de optimización combinatoria, en el que se pretende obtener el menor subconjunto \mathbf{w}_t de \mathbf{x}_t^* que maximiza un criterio $C(\cdot)$ que mide la cantidad de información entre \mathbf{w}_t y y_t :

$$\mathbf{x}_t = \arg \max_{\mathbf{w}_t \subset \mathbf{x}_t^*} C(\mathbf{w}_t, y) \quad (26)$$

La solución de (26) mediante un proceso de búsqueda exhaustiva haciendo un recorrido or-

denado de todos los subconjuntos de \mathbf{x}_t^* , es el único que garantiza encontrar el conjunto óptimo de regresores \mathbf{x}_t ; sin embargo, esto sólo es factible cuando C es calculable rápidamente y el tamaño del conjunto \mathbf{x}_t^* es moderado, ya que puede caer fácilmente en el curso de la dimensionalidad. No obstante, diferentes métodos heurísticos han sido propuestos para la reducción del espacio de búsqueda de los subconjuntos de \mathbf{x}_t^* . Selecciones comunes son algoritmos de ramificación o acotación, la selección de las mejores variables individuales, la selección secuencial hacia delante, la eliminación secuencial hacia atrás o el método stepwise. Sin embargo, todos ellos, realizan un recorrido no óptimo del espacio de búsqueda, y no usan una heurística inteligente que permita descartar regiones con base en las características particulares de $\{\mathbf{x}_t, y_t\}_1^T$.

Tabla 1. Estrategia para la Especificación de Modelos

<i>Paso 1:</i>	Identificación
	– Selección de los Regresores
	– Propiedades estadísticas de la serie temporal
	– Transformación de la serie temporal
	– Especificación de $\hat{f}(\cdot)$
<i>Paso 2:</i>	Estimación de Parámetros
<i>Paso 3:</i>	Diagnóstico y Análisis de la Dinámica de la Serie
<i>Paso 4:</i>	Metadiagnóstico
<i>Paso 5:</i>	Construcción de Pronósticos e Intervalos de Confianza
<i>Paso 6:</i>	Validación del Pronóstico

4.1.2. Especificación de $C(\cdot)$ como un criterio independiente de ANFIS

En las técnicas independientes del modelo difuso, C es especificado como una medida de la cantidad de información que contiene cada uno de los subconjuntos de \mathbf{x}_t^* para pronosticar y_t ; sin embargo, no puede asegurarse que la especificación posterior de ANFIS capture completamente las relaciones entre el conjunto final de regresores seleccionado \mathbf{x}_t y y_t . El exámen de los autocorrelogramas, autocorrelogramas parciales, correlogramas cruzados y correlogramas cruzados parciales de la serie temporal, podría entregar información de utilidad. La autocorrelación parcial es una medida de la dependencia lineal entre y_t y y_{t+k} , después de que la influencia de rezagos interme-

dios ha sido parcialmente eliminada. No obstante, Granger y Terävirta [11] muestran que el uso de correlaciones y autocorrelaciones como medidas de dependencia aplicadas a procesos no lineales o no estacionarios puede ser peligroso. Igualmente sugieren que otras medidas útiles de relación entre $\mathbf{x}_t^{(i)}$ y y_t son obtenidas al analizar sus relaciones cuando son aplicadas diferentes transformaciones, paramétricas y no paramétricas, a cada uno de ellos. Transformaciones comúnmente usadas son $\log |u|$, u , u^2 o $\exp u$. Estas medidas son el coeficiente máximo de correlación entre $\gamma(y_t)$ y $\eta(\mathbf{x}_t^{(i)})$, donde las funciones $\gamma(\cdot)$ y $\eta(\cdot)$ son escogidas para maximizar el coeficiente de correlación cruzada. La máxima correlación media entre y_t y $\eta(\mathbf{x}_t^{(i)})$, que se obtiene escogiendo $\eta(\cdot)$, tal que maximice este criterio. Finalmente, el coeficiente

máximo de regresión, que se obtiene a partir de $y_t = \eta(\mathbf{x}_t^{(i)}) + \nu_t$.

Una medida más general de dependencia es la Información Mutua [4, 3, 18], la cual permite detectar relaciones lineales y no lineales entre variables. Si x y y son variables aleatorias, con distribuciones de probabilidad $f_{\mathbf{X}}$ y $f_{\mathbf{Y}}$ respectivamente, y distribución conjunta $f_{\mathbf{X},\mathbf{Y}}$, la Información Mutua de X respecto a Y , $I(\mathbf{x}; y)$ es:

$$I(\mathbf{x}, y) = \int_{\mathcal{R}} f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, y) \log \frac{f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, y)}{f_{\mathbf{X}}(\mathbf{x})f_{\mathbf{Y}}(y)} d\mathbf{x} dy \quad (27)$$

la cual puede interpretarse como la medida de Kullback–Leiber entre la distribución actual de los datos, $f_{\mathbf{X},\mathbf{Y}}$, y la distribución que ellos tendrían si fuesen estadísticamente independientes, $f_{\mathbf{X}} \cdot f_{\mathbf{Y}}$. La correlación sombra:

$$R^2(\mathbf{x}_t^{(i)}, y_t) = 1 - \exp(-2 \cdot I(\mathbf{x}_t^{(i)}, y_t)) \quad (28)$$

está íntimamente relacionada con la IM y puede ser usada potencialmente para sugerir los rezagos apropiados usados en el modelo.

4.1.3. Especificación de $C(\cdot)$ como un criterio dependiente de ANFIS

En las técnicas dependientes, C se especifica como una función del error de ajuste entre ANFIS y y_t ; esto implica que la selección de \mathbf{x}_t , la especificación de ANFIS y la estimación de sus parámetros se realiza de forma conjunta, dificultando el proceso de construcción del modelo, y haciendo que dichas técnicas requieran recursos computacionales intensivos.

Jang [17] propone que $C(\cdot)$ sea especificada como el error cuadrático medio entre las observaciones reales de la serie temporal y ANFIS para el cual los parámetros de los antecedentes han sido establecidos de forma heurística, y los consecuentes mediante mínimos cuadrados, sin realizar ningún tipo de entrenamiento del modelo.

4.1.4. Determinación de las propiedades estadísticas de $\{y_t\}_1^T$

Cuando hay sospechas de no linealidad, Tong [31] indica que los diagramas de dispersión entre $(y_t, x_t^{(j)})$, los histogramas bivariados, o los estimados no paramétricos de $E[y_t|x_t^{(j)}]$ y $V[y_t|x_t^{(j)}]$

pueden ser útiles. Su principal debilidad es la interpretación subjetiva por parte del experto, y que ellas pueden indicar la posibilidad de existencia de relaciones, pero no pueden usarse como un criterio final en este proceso.

El análisis del espectro de la serie temporal permite la detección de otras características importantes relacionadas con sus patrones estacionales y ciclos de largo plazo. El espectro de energía es comúnmente investigado, pero otras medidas tales como el espectro cruzado, el biespectro o el biespectro cruzado podrían dar otras interpretaciones de la serie temporal.

Como un complemento a los métodos gráficos informales, varias pruebas de linealidad aplicadas en el dominio de la frecuencia o el tiempo pueden ser aplicadas, tanto a los datos originales como a la serie temporal transformada, usualmente con la función $f(u) = u^2$. Una prueba común para verificar la linealidad de los datos es calcular los cuadrados de la autocorrelación de la serie temporal, y la autocorrelación del cuadrado de la serie temporal; estos valores deben ser muy similares cuando la serie es lineal, y su divergencia es una evidencia de no linealidad.

4.1.5. Transformación de la serie temporal

Otra parte importante del análisis exploratorio de datos es seleccionar una transformación de la serie temporal que minimice el ruido e indique relaciones importantes, y prepare los datos en una forma que facilite la estimación del modelo. La mayoría de los modelos son desarrollados bajo la suposición de homogeneidad en el sentido estadístico, lo que implica que la serie temporal debe tener una varianza estable, y que la tendencia, el patrón estacional y los ciclos de largo plazo han sido eliminados antes del modelamiento. Para hacer la varianza estable se usan transformaciones como $\log u$, \sqrt{u} o $\frac{1}{u}$, pero Mills [23, p. 32] indica que ellas son un caso especial de la familia de transformaciones de potencias de Box y Cox [6]. La eliminación de componentes estructurales se realiza a través de diferenciaciones simples y estacionales; sin embargo, este procedimiento asume que dichas componentes son fijas y no son influenciadas por los regresores.

4.1.6. Especificación de ANFIS

Como ya se demostró en las secciones anteriores, los modelos LSTAR son, bajo algunas restricciones, un caso particular de ANFIS, de tal forma que los procedimientos empíricos para su especificación, pueden ser potencialmente útiles en la formulación de esta clase de modelos difusos. En el caso de ANFIS, la formulación de un modelo lineal autoregresivo permite especificar la forma funcional y los regresores usados en los consecuentes de las reglas. No obstante, se parte de la hipótesis de que en los diferentes regímenes tienen influencia las mismas variables, y cambia solamente el valor del coeficiente en cada una de las reglas. Esto no es necesariamente cierto, y podría usarse un proceso de optimización combinatoria que determine que regresores debe contener cada regla. En el caso de los modelos STAR, el rechazo de la hipótesis de linealidad es realizado con base en pruebas estadísticas; sin embargo, no existen desarrollos de este tipo para ANFIS, por lo que debería recurrirse a pruebas indirectas de la calidad de representación y generalización del modelo, tales como la validación cruzada. La selección de la variable de transición debe realizarse de forma empírica, contrastando diferentes modelos que difieren en el regresor utilizado para su especificación.

4.2. Paso 2: Estimación de Parámetros

Jang [15] muestra como los consecuentes de las reglas pueden ser estimados por mínimos cuadrados, para valores conocidos de los parámetros de los conjuntos difusos. Este procedimiento permite reducir el espacio de búsqueda para realizar la optimización. Un procedimiento similar ha sido propuesto por otros autores para la estimación de los parámetros STAR.

4.3. Paso 3: Diagnóstico

El proceso de diagnóstico consiste en verificar el cumplimiento de las hipótesis del modelo una vez ha sido ajustado a la serie de datos. Este es realizado con base en dos tipos de pruebas fundamentales: la emisión de juicios informados basados en la inspección visual de ajuste de ANFIS a la serie, y el análisis estadístico de los resultados, que permitirá corroborar los juicios y detec-

tar inconsistencias que no son fácilmente visibles. El análisis de los errores puede dar indicaciones sobre problemas en la especificación del modelo difuso o la selección de los regresores que conforman x_t . Sin embargo, este proceso contiene una alta cantidad de juicio subjetivo para determinar si el modelo se ajusta apropiadamente a los datos. Este es un proceso comúnmente realizado en el modelado estadístico de series temporales, pero aparentemente no es una práctica usual en el modelamiento usando herramientas de Inteligencia Computacional.

Durante esta fase se examinan los residuos del modelo, $\{\hat{v}_t\}_1^T$, para juzgar si ellos son ruido blanco; el proceso de análisis consta de dos partes, una inspección visual de diferentes gráficos que permiten visualizar el comportamiento de los residuos a lo largo de la serie de tiempo; y de la realización de pruebas estadísticas para verificar que sean efectivamente ruido blanco.

El procedimiento básico para examinar el comportamiento de los residuos es graficarlos contra el tiempo. Sin embargo, una gráfica de los residuos estandarizados puede ser más útil. Otros procedimientos de utilidad es examinar la gráfica de la suma acumulada de residuos estandarizados, y la gráfica de la suma acumulada del cuadrado de los residuos estandarizados. En su conjunto estas tres gráficas pueden ayudar a detectar cambios estructurales en la serie temporal; regiones con errores persistentemente positivos o negativos, o muy altos o bajos en magnitud respecto al resto del gráfico son de interés.

Los modelos descritos en secciones anteriores parten de la hipótesis de normalidad de los errores. Su falta de cumplimiento es un indicativo de que ANFIS no ha capturado adecuadamente las propiedades de $\{y_t\}_1^T$. El procedimiento básico para verificar la normalidad es la construcción del histograma de los errores estandarizados, y compararlo contra la distribución normal estándar.

Varias pruebas estadísticas pueden ser realizadas para verificar especificaciones incorrectas de ANFIS, basadas en la detección de componentes estructurales dentro de los residuos. Ljung y Box [19] proponen el uso del estadístico Q^* para determinar si las primeras P correlaciones no son significativamente diferentes de cero. Cuando hay sospechas de no linealidad, el estadístico anterior puede ser usado sobre diferentes funciones de los residuos; en particular, cuando es aplicado sobre las autocorrelaciones de los cuadrados de los residuos, puede ser indicativo de algunos efectos

no lineales. Igualmente el estadístico de Durbin-Watson puede ser usado para verificar si la autocorrelación para el primer rezago es cero. Un diagnóstico de heterocedasticidad de los residuos [13] puede ser realizado verificando el estadístico:

$$H(h) = \frac{\sum_{t=T-h+1}^T \hat{v}_t^2}{\sum_{t=1}^{1+h} \hat{v}_t^2} \quad (29)$$

donde h es el entero más cercano a $T/3$; $hH(h)$ sigue una distribución χ_h^2 . Para verificar la normalidad de los residuos, se calculan sus coeficiente de asimetría, b_1 , y curtosis, b_2 , los cuales se comportan asintóticamente normales, con distribuciones $\mathcal{N}[0, 6/T]$ y $\mathcal{N}[3, 24/T]$ respectivamente. La no normalidad puede ser verificada por el estadístico:

$$N = (T/6)b_1^2 + (T/24)(b_2 - 3)^2 \quad (30)$$

la cual sigue una distribución χ_2^2 . La regresión de \hat{v}_t respecto a t , a los regresores, $x_t^{(j)}$, y y_t puede indicar una estructura de dependencia en los errores.

Durante la fase de diagnóstico, la presencia de componentes estructurales en la serie de residuos del modelo, revela una inconsistencia dentro de su especificación, debido posiblemente a la necesidad de incluir otros rezagos de la variable pronosticada, nuevas variables causales o nuevos rezagos de las variables causales ya incorporadas.

En el proceso de análisis de la dinámica de la serie, el objetivo fundamental es encontrar regiones o puntos en la serie de tiempo, con muy baja probabilidad de ocurrencia de acuerdo con el modelo ajustado. Dichas regiones pueden corresponder a cambios en la dinámica de la serie o intervenciones puntuales sobre el proceso generador. Una vez detectados, es posible realizar un proceso de indagación que busque los motivos de este comportamiento atípico y reformular el modelo para que se tenga esta consideración en cuenta.

4.4. Paso 4: Metadiagnóstico

Este se basa en dos grupos de pruebas: el primer grupo tiene por objeto determinar la calidad del ajuste del modelo en términos de precisión; los estadísticos comúnmente aceptados son el porcentaje de la varianza explicada por el modelo:

$$R^2 = 1,0 - \sum_{t=1}^T \frac{\hat{v}_t^2}{y_t^2} \quad (31)$$

la varianza del predictor:

$$\sigma_{\hat{v}_t}^2 \sim \frac{1}{T} \sum_{t=1}^T \hat{v}_t \quad (32)$$

y la desviación media absoluta (MAD):

$$\text{MAD} = \frac{1}{T} \sum_{t=1}^T |\hat{v}_t| \quad (33)$$

El segundo grupo tiene como objetivo descartar otras especificaciones que podrían ajustarse a los datos; se realiza eliminando o adicionando términos, y demostrando que ellos son necesarios para el caso de la eliminación, o son estadísticamente no significativos en el caso de su adición.

Una vez que ANFIS a sido especificado y sus parámetros han sido calculados, deben realizarse pruebas de aceptación que midan su bondad en la representación de los datos. Usualmente en los modelos de corte matemático, este proceso está definido por la ejecución de una serie de pruebas que validan su aceptación. Sin embargo, existen otros criterios de validación que tienen que ver mucho más con la representación mental que realiza el experto, y que a pesar de no ser metodologías matemáticas si permiten que se gane confianza en el uso de la herramienta de pronóstico. Este tipo de validaciones es usado comúnmente en Dinámica de Sistemas.

4.5. Paso 5: Construcción de Pronósticos e Intervalos de Confianza

Cuando se ha finalizado el proceso de especificación y validación de ANFIS, se procede a la construcción de pronósticos e intervalos de confianza. El pronóstico para un periodo adelante, y_{T+1} , sigue usualmente una distribución normal, y es calculado de forma directa usando (3). Cuando \mathbf{x}_t contiene uno o más valores rezagados de y_t , la construcción del pronóstico j períodos adelante, requiere el uso de simulación de Montecarlo [10], ya que ANFIS realiza una transformación no lineal de los datos de entrada.

Un factor crítico de este proceso es la incorporación de información contextual o cualitativa en el proceso de pronóstico. Una vez construidos los pronósticos, el experto emite un juicio sobre su validez basado en su apreciación subjetiva del

proceso generador de la serie temporal. Si el experto considera que el pronóstico no refleja adecuadamente la realidad, se realiza una modificación de los resultados entregados por el modelo. Existen dos fuentes posibles de error: la primera es que el experto realice una corrección que no es necesaria, y la segunda que no la introduzca, cuando ella es requerida.

Adicionalmente a las correcciones de los pronósticos, puede hacerse necesario que se revalúe la especificación de ANFIS desde alguna de las fases previas.

Otro problema fundamental en la emisión de pronósticos es cuando el vector de regresores contiene otras variables causales diferentes a valores rezagados de y_t . Durante el proceso de emisión de pronósticos, es necesario conocer los valores de x_t durante todo el horizonte de pronóstico, de tal forma, que es requerido conocer los valores de dichas variables causales, por lo que sus valores deben ser entregados al modelo ya sea como pronósticos o como escenarios de evolución. En el modelamiento de series temporales con regresores no lineales, las variables causales pueden influenciarse parcialmente y de forma no lineal, de tal forma que ya sea sus pronósticos o sus escenarios de evolución no pueden ser desarrollados independientemente.

Es también importante, la preparación de una memoria de especificación, ya que pueden revisarse y revalidarse muchos de los supuestos de ANFIS cuando fue especificado originalmente, haciéndose vital un proceso de documentación ordenado y minucioso.

4.6. Paso 6: Validación del Pronóstico

En la fase de validación del pronóstico, pueden presentarse observaciones muy alejadas de las predicciones realizadas. Esto puede ser debido a que se produjo un valor con una probabilidad muy baja de ocurrencia y es simplemente un valor extremo, o a que se está gestando un cambio en la dinámica del proceso generador de las observaciones. Cuando esta situación ocurre, el pronosticador debe juzgar el motivo de ocurrencia para tomar los correctivos necesarios; sin embargo, una sola observación no es suficiente, y es necesario tener que esperar para recolectar evidencias que permitan juzgar si existe un cambio en la dinámica del proceso generador. Existen también cam-

bios más sutiles, y que no son fácilmente detectables, los cuales ameritan un ajuste del modelo, como por ejemplo, que el pronóstico sea consistentemente superior o inferior a las observaciones.

5. Aplicación al pronóstico del precio de electricidad en Brasil

Las últimas décadas se han caracterizado por procesos de desregulación y liberalización (véase a Steiner [26] para una revisión general) que buscan fundamentalmente [1, 9] la creación de mercados de electricidad teóricamente eficiente, en los cuales las empresas racionalicen su operación, y produzcan beneficios económicos representados por la disminución de las tarifas para el consumidor final. En el caso particular de Brasil, la reestructuración de su sector eléctrico dio como resultado la creación del Mercado Atacadista de Energía (MAE) que tiene dentro de sus objetivos fundamentales, el establecimiento de un precio de la electricidad en el mercado de corto plazo (o precio spot) que refleje el costo marginal de operación (CMO) del sistema en el corto plazo, que permita a los productores y consumidores realizar decisiones más eficientes, y que a su vez sirva como una señal económica para el establecimiento de los precios de la electricidad en los contratos de largo plazo y las decisiones de inversión de los agentes privados. En la actualidad, el precio de electricidad es semanal y es calculado para cada uno de los submercados en que se encuentra dividido el sector eléctrico Brasileño. Ellos son: Norte, Nordeste, Sureste/Centro, Oeste y Sur.

El proceso de cálculo del precio spot, se basa en la planificación centralizada del sistema, donde este precio se asocia al costo marginal de la demanda (la formulación matemática general es presentada por Pereira [24], y las características detalladas de la metodología de solución y representación del sistema son presentadas en [20] y [21]); no obstante, el Gobierno tiene la facultad de intervenir el precio de la electricidad para modificar la señal enviada a los agentes del sistema.

Por simplicidad de la exposición, sólo se considera el caso del modelado univariado de la serie de precios.

5.1. Identificación

En esta primera fase, se analiza el comportamiento de la serie de precios spot y se determinarán sus propiedades estadísticas, para posteriormente realizar la selección de regresores y la formulación preliminar del modelo.

5.1.1. Serie de precios de la electricidad

La inspección visual de la serie presentada en la Figura 2, revela varias características interesante. Desde ene/96 hasta dic/98 la serie presenta un comportamiento caracterizado por la presencia de precios bajos de la electricidad, específicamente en feb/96, dic/96 a feb/97, may/97, ene/98 y nov/98; en ene/99 se produce aparentemente un cambio de nivel, causando que el resto del tiempo, el precio no baje de 44 \$R/kWh. Durante la primera mitad de 1996, el precio aumenta, para luego caer en la segunda mitad, un comportamiento aparentemente asociado a la variabilidad hidrológica. 1997 se caracteriza por tener precio medio más bajo que el año anterior, causado posiblemente por una mayor abundancia de agua para la generación hidráulica. Durante 1998, el precio se hace menos volátil que en años anteriores, y presenta un nivel más estable, con un valor medio similar al de 1997. El precio de la electricidad durante el primer semestre de 1999 tiene poca variabilidad, y presenta una ligera tendencia a la baja; durante la segunda mitad del mismo año, se produce un rápido ascenso del precio spot hasta alcanzar un valor de 285 \$R/kWh en ene/00; los meses de feb/00 a abr/00 presentan un descenso rápido del precio spot, de los 285.5 \$R/kWh registrados en ene/00 a 56.7 \$R/kWh en abr/00. A partir de este periodo se presenta un ascenso y posterior descenso del precio, similar al ocurrido en 1996 pero con un nivel mucho más alto, que coincide con el agotamiento de las reservas hidráulicas, y posterior almacenamiento de agua; en ene/2001 se inicia un ascenso acelerado del precio spot, desde 56.9 \$R/kWh hasta 684 \$R/kWh registrados en jun/01. A medida que la energía afluyente va aumentando, el precio spot empieza a disminuir, pasando de su nivel máximo en sep/00 a 336 \$R/kWh en oct/200, valor en el que permanece estable durante el resto del año. La media de la muestra de datos es 119.68, su varianza es 27375.68, su coeficiente de asimetría es 2.44, y su curtosis de 5.66.

5.1.2. Selección preliminar de regresores

En el caso analizado sólo se consideran los valores pasados del precio de la electricidad, de tal forma que \mathbf{x}_t^* contiene como elementos a todos los valores retrasados en el tiempo de y_t respecto al momento actual t . En el caso en que se incluyeran variables causales, \mathbf{x}_t^* podría incluir la demanda de electricidad, los aportes hidrológicos, o la energía almacenada en el embalse agregado. Si ellas son consideradas en la formulación del modelo, aquí debería justificarse su inclusión.

5.1.3. Transformación de la Serie de Precios

La serie de precios spot no cumple con las características de homogeneidad en el sentido estadístico, ya que la media y la varianza de la primera mitad de la serie es diferente a la de la segunda mitad; una transformación de los datos puede resultar beneficiosa en este aspecto; sin embargo, la decisión del tipo de transformación esta basada en el juicio experto del modelador. En este caso se usa la transformación de Box y Cox con $\lambda = 0,065$. Igualmente esta transformación hacen que la distribución estadística de los datos $\{y_t\}_1^T$ se asemeja más a una normal. La serie transformada de precios spot se presenta en la Figura 3. En ella, se puede visualizar con mayor detalle las características ya discutidas.

5.1.4. Propiedades estadísticas de las series

Los análisis se realizan para la serie del precio spot transformado. El diagrama presentado en la Figura 4, muestra una relación de no linealidad entre el precio spot y el precio spot rezagado una o dos unidades de tiempo; para rezagos mayores la gráfica se va haciendo cada vez más dispersa. El valor esperado precio spot y su varianza condicionales al precio spot rezagado i períodos son presentados en las Figuras 5 y 6. Ellas indican que para los valores medios del precio spot, se espera un crecimiento gradual del precio de orden lineal; mientras que en sus valores extremos, se presenta una relación no lineal. La varianza del precio spot dado el precio rezagado, es variable y decreciente en el tiempo; los precios bajos presentan varianzas mucho más altas que los precios altos; este hallazgo puede confirmarse con la inspección de la Figura 3.

Como una medida de la relación existente entre cada elemento de \mathbf{x}_t^* y y_t , se utiliza la autocorrelación y la autocorrelación parcial tal como se indicó en la metodología. Los resultados del análisis de los autocorrelogramas y autocorrelogramas parciales indican la existencia de una relación causal entre los precios con rezagos uno y dos y el precio actual. Se sospecha de la existencia de una relación no lineal para estos mismos regresores ya que la varianza no es constante.

5.2. Formulación del modelo y diagnóstico

En esta fase se ajustará primero un modelo autoregresivo y luego se procederá con la especificación de ANFIS, usando los valores rezagados de la serie temporal. Las observaciones de las series entre ene/96 y jun/01 son usadas para la estimación de parámetros y diagnóstico del modelo, mientras que las observaciones entre jul/01 y

dic/01 son usados para las pruebas de pronóstico fuera de la muestra.

5.2.1. Modelo autoregresivo

De acuerdo con los resultados del análisis del autocorrelograma y el autocorrelograma parcial, el vector de regresores potenciales \mathbf{x}_t^* se reduce a los rezagos uno y dos de la serie de precios spot. De esta forma, se ajustará el modelo:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \hat{v}_t$$

Sin embargo, la selección final se basa en los resultados de los criterios de Akaike (AIC) y Bayes (BIC) para modelos alternativos donde se considera que entra uno solo de los regresores. Los resultados se presentan en la Tabla 2. Todos los coeficientes caen en los límites de invertibilidad. Los criterios AIC y BIC favorecen el modelo que incluye los dos rezagos del precio spot.

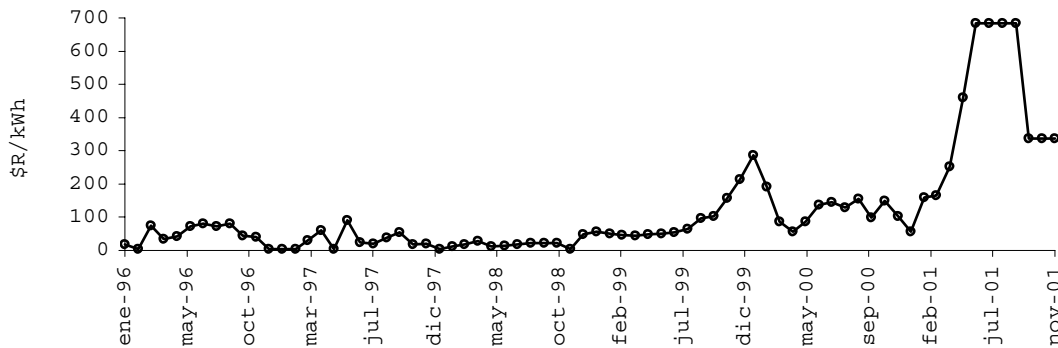


Figura 2. Precio de la electricidad

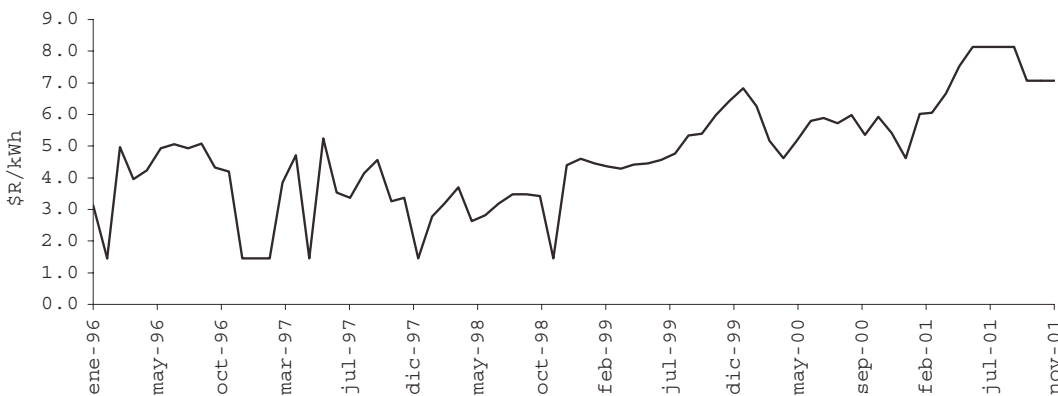


Figura 3. Transformación de Box y Cox del precio de la electricidad. $\lambda = 0.065$

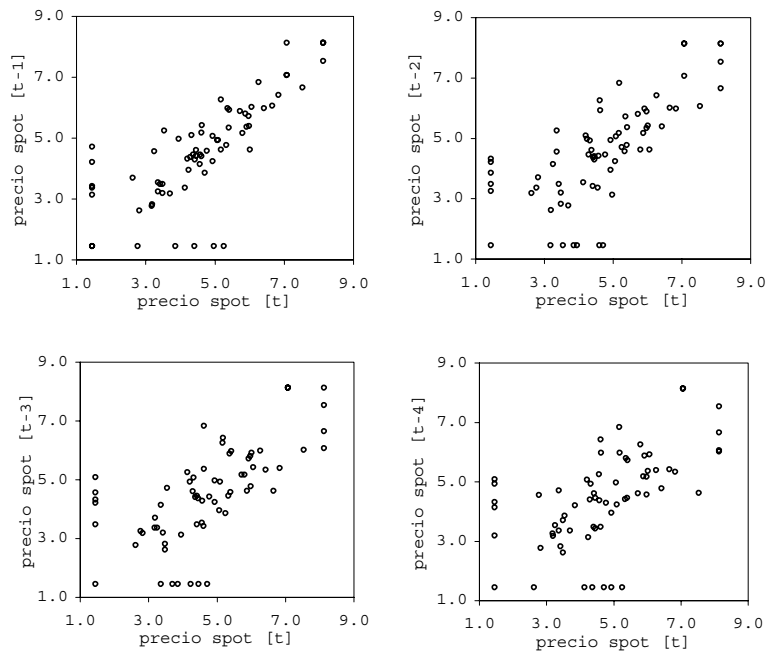


Figura 4. Diagrama de dispersión.

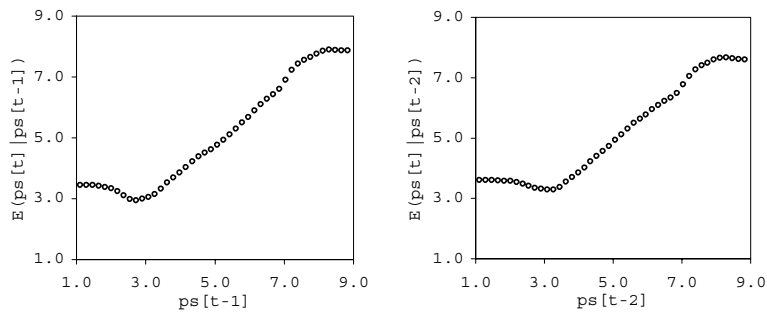


Figura 5. Precio esperado.

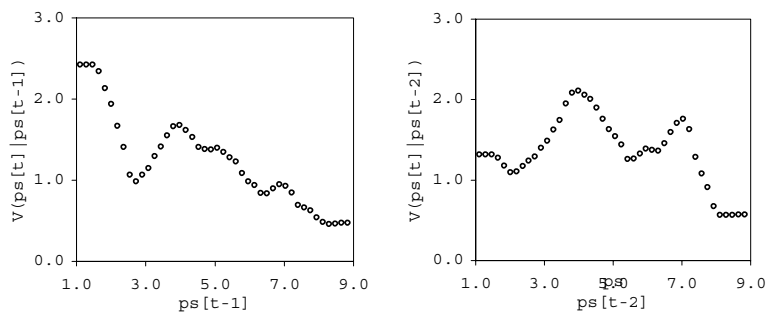


Figura 6. Varianza esperada.

Tabla 2. Selección del modelo autoregresivo

Modelo	ϕ_1	ϕ_2	$\sigma_{\hat{v}_i}^2$	AIC	BIC
i=1	0.71700		1.3122	0.3020	0.3352
i=2		0.63381	1.6249	0.5158	0.5489
i=1,2	0.55558	0.25638	1.2327	0.2698	0.3362

5.2.2. ANFIS

En este caso se especifica un modelo ANFIS para el caso univariado. Los modelos son estimados usando como variable de transición el precio spot con rezago uno o dos; los consecuentes de las reglas son modelos autoregresivos de orden dos. La selección del mejor modelo se realiza, considerando todas las posibles combinaciones de variable de decisión, y posibles modelos autoregresivos en el consecuente de las reglas. El mejor modelo encontrado usa como variable de decisión el precio spot con rezago uno; el consecuente de la primera regla difusa es un modelo autoregresivo cuyo regresor es el precio spot con rezago uno, mientras que para la segunda regla, el consecuente es un modelo autoregresivo de orden uno cuyo regresor es el precio spot con rezago dos. Se usa la función de campana generalizada para representar los conjuntos difusos.

La varianza del predictor es 0,8680, mientras que la desviación media absoluta es 24,68 %. Los residuos cumplen con la hipótesis de normalidad, el coeficiente de asimetría es -0,94, mientras que la curtosis es 3,92. No hay evidencias de heterocedasticidad en los residuos. Las correlaciones seriales para los primeros ocho rezagos de los residuos no son significativamente diferentes de cero. Este análisis preliminar indica que los residuos siguen una distribución normal. Sin embargo, este no es concluyente ya que no garantiza que el modelo represente adecuadamente el comportamiento de la serie de precios. Por ello, es necesario revisar detalladamente los pronósticos, buscando en que observaciones o regiones el modelo puede resultar insatisfactorio: en el pronóstico un periodo adelante [véase las Figuras 7a y 7b] los períodos donde el precio spot toma el valor de 4 \$/kWh, quedan por fuera de los intervalos de confianza del pronóstico, excepto para la observación de nov/98 para la cual ANFIS pronóstica adecuadamente su valor. Durante 1999, ANFIS pronostica valores consistentemente por debajo de las obser-

vaciones. La misma situación se presenta entre feb/01 y jun/01. Igualmente las Figuras 7c, 7d, 7e y 7f, muestran que el error en la primera mitad de la serie es en magnitud superior al de la segunda mitad; igualmente, que hay cierta alternación del signo de los errores en la primera mitad de la gráfica, mientras que para la segunda estos son persistentemente positivos.

5.3. Metadiagnóstico

En el pronóstico extrapolativo (véase la Figura 8), se realiza una predicción ciega de los próximos seis meses, donde cada pronóstico se convierte en un regresor para obtener la predicción del mes siguiente. Ya que el pronóstico y_{T+i} no es un valor sino una distribución de probabilidades, es necesario utilizar simulación de Montecarlo para obtener y_{T+i+1} . Los resultados obtenidos muestran que ANFIS es consistente con la disminución del precio spot del segundo semestre de 2001; en este caso la varianza del predictor es 0,6779, y la desviación media absoluta es 10,35 %. En la Tabla que aparece en la Figura 8, se presentan tanto los pronósticos como los intervalos de confianza para el percentil del 95 %. Es importante notar como dichos intervalos no son simétricos alrededor del pronóstico debido a la transformación no lineal que realiza ANFIS.

5.4. Construcción de pronósticos e intervalos de confianza

Ya que se ha concluido que el modelo es adecuado para representar la serie temporal, se procede refinamiento de los parámetros con la información usada para el metadiagnóstico, y se procede a la construcción de predicciones e intervalos de confianza de igual forma que en el paso anterior.

6. Conclusiones

Se ha mostrado como los modelos ANFIS pueden ser derivados como una generalización al caso no lineal de los modelos autoregresivos, en el contexto de las series temporales. Esta derivación permite especificar de forma práctica, una clase más general de modelos ANFIS, en la cual, los conjuntos de variables que conforman los antecedentes de las reglas son diferentes. Esto posibilita la aplicación de esta metodología a series relativamente cortas pero que requieren un número relativamente alto de variables causales para explicar su comportamiento. En consecuencia, se obtienen modelos con un número inferior de reglas, y por

consiguiente menos parámetros, a los obtenidos al derivar la formulación usando metodologías tradicionales como CART. Su interpretación como un modelo de series temporales, también posibilita el uso de herramientas estadísticas en el proceso de selección de variables, diagnóstico y metadiagnóstico del modelo, enriqueciendo las metodologías tradicionales de Inteligencia Computacional cuando son aplicadas a esta clase de problemas.

La aproximación metodológica propuesta, no sólo da información en términos de precisión numérica sobre la capacidad del pronóstico, sino que también permite realizar un análisis detallado de la calidad del modelo.

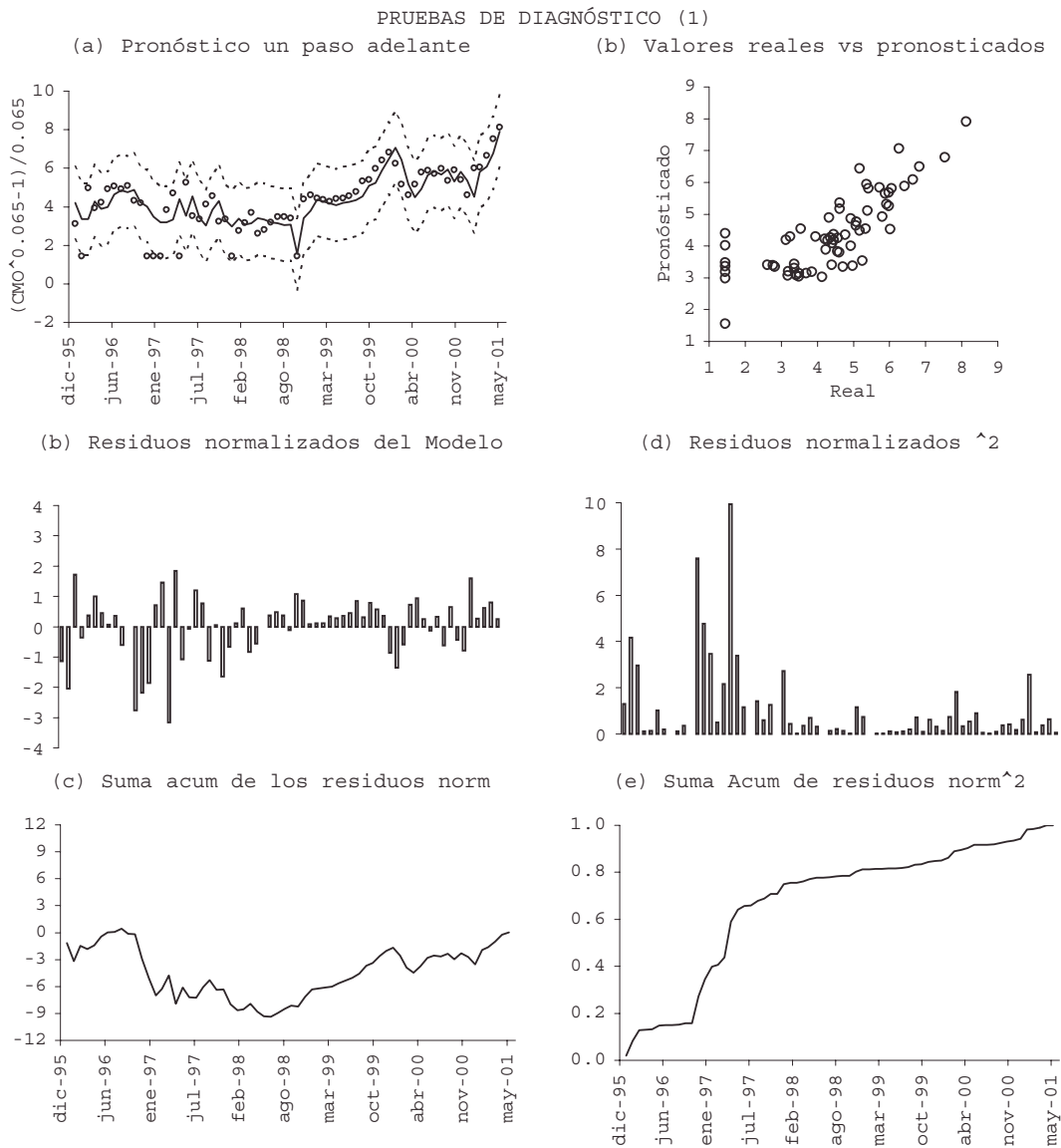


Figura 7. Pruebas de diagnóstico.

PRONÓSTICO HACIA DELANTE

Fecha	Predicción			Obs.
	$\mu-2\sigma$	μ	$\mu+2\sigma$	Real
jul-01	5.72	7.56	9.47	8.13
ago-01	4.06	7.33	9.55	8.13
sep-01	3.27	6.89	9.39	8.13
oct-01	3.09	6.55	9.23	7.07
nov-01	2.78	6.36	9.18	7.07
dic-01	2.40	6.18	9.09	7.07

Estadísticos	
Varianza del Predictor	0.6779
Desviación Media Absoluta	0.1035
Correlación obs. vs pronósticos	0.8958
Coefficiente de Determinación R^2	0.8024

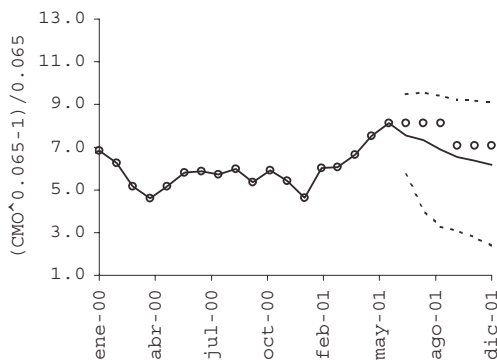


Figura 8. Pronóstico extrapolativo.

La varianza estimada del error para el modelo difuso sin variables explicativas, es aproximadamente el 50% del mismo término para el modelo autoregresivo, mientras que la desviación media absoluta es reducida en un 30% aproximadamente. Estos resultados incentivan la investigación de esta metodología en el modelamiento y predicción de series temporales no lineales, pero aún esta lejos de ser completa. En un próximo estado, es necesario investigar sobre la interpretabilidad de estos modelos en el contexto de la Inteligencia Computacional, y en como usar procedimientos estadísticos más rigurosos que el aquí empleado, tal como es común en otros modelos econométricos.

Referencias

[1] M. Armstrong, S. Cowan, and J. Vickers. *Regulatory Reform: Economic Analysis and British Experience*. The MIT Press, Cambridge, 1994.

[2] D.W. Baccon and D.G. Watts. Estimating the transition between two intersecting straight lines. *Biometrika*, 58:525–534, 1971.

[3] R. Battiti. Using mutual information for selecting features in supervised neural net learning. *IEEE Transactions on Neural Networks*, 5(4):537–550, 1994.

[4] B. Bonnländer. *Nonparametric selection of input variables for connectionist learning*. PhD thesis, University of Colorado. Department of Computer Sciences, 1996.

[5] G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day Inc., 1970.

[6] G.E.P. Box and D.R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, (26):211–243, 1964.

[7] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and regression trees*. Technical report, Wadsworth, Inc, Belmont, California, 1984.

[8] D. Bunn and G. Wright. Interaction of judgemental and statistical forecasting methods: Issues and analysis. *Management Science*, 37(5), May 1991.

[9] D.W. Bunn. Reflections on the progress of electricity re-structuring, privatisation and regulation in the uk during 1988–1998. In *XXXIII Reunión de Altos Ejecutivos C.I.E.R.*, 1998.

[10] George S. Fishman. *Monte Carlo. Concepts, Algorithms and Applications*. Springer Series in Operational Research. Springer, 1996.

[11] Clive W. J. Granger and Timo Teräsvirta. *Modeling Nonlinear Economic Relationships*. Oxford University Press, 1993.

[12] C.W.J. Granger. Strategies for modelling nonlinear time-series relationships. *The Economic Record*, (69):233–238, 1993.

[13] Andrew C. Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, 1989.

- [14] Robin Hogarth. *Judgement and Choice*. John Wiley and Sons, second edition, 1994.
- [15] J-S R. Jang. Anfis: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, 23:665–684, 1993.
- [16] J.-S. R. Jang. Structure determination in fuzzy modeling: A fuzzy cart approach. In *Proc. of IEEE international conference on fuzzy systems, (Orlando, Florida)*, June 1994.
- [17] J-S. R. Jang. Input selection for anfis learning. In *Proceedings of the IEEE Int. Conf. On Fuzzy Systems*, 1996.
- [18] Nojun Kwak and Chong-Ho Choi. Input feature selection for classification problems. *IEEE Transactions on Neural Networks*, 13(1):143–159, January 2002.
- [19] L. Ljung and G.E.P. Box. On a measure of lack of fit in time series models. *Biometrika*, (66):67–72, 1978.
- [20] M.E.P. Maceira. Programação dinâmica dual estocástica aplicada ao planejamento da operação energética de sistemas hidrotérmicos com representação do processo estocástico de afluências por modelos autorregressivos periódicos. Technical report, Relatório Técnico CEPEL DPP/PEL 237/93, 1993.
- [21] M.E.P. Maceira, C.B. Mercio, B.G. Gorenstin, S.H.F. Cunha, C. Suanno, M.C. Sacramento, and A.S. Kligerman. Energy evaluation of north/northeastern and south/southeastern interconnection with newave model. In *IV Symposium of Specialist in Electric Operational and Expansion Planning - SESOPE, Salvador, Brazil*, May 1998.
- [22] G. S. Maddala. *Econometrics*. McGraw-Hill, 1977.
- [23] T. C. Mills. *The econometric modelling of financial time series*. Cambridge Press University, 1993.
- [24] M. Pereira. Modelo de despacho hidrotérmico. Technical report, PSRI, 1997.
- [25] M.B. Priestley. State-dependent models: a general approach to non-linear time series analysis. *Time Series Analysis*, (1), 1980.
- [26] F. Steiner. Regulation, industry structure and performance in the electricity supply industry. *OECD Economic Studies*, (32), 2000.
- [27] M. Sugeno and G. T. Kang. Structure identification of fuzzy model. *Fuzzy Sets and Systems*, 28:15–33, 1988.
- [28] T. Takagi and M. Sugeno. Derivation of fuzzy control rules from human operator's control actions. In *Proceedings of the IFAC Symposium on Fuzzy Information, Knowledge Representation and Decision Analysis*, pages 55–60, July 1983.
- [29] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15:116–132, 1985.
- [30] T. Teräsvirta. Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, (89):208–218, 1994.
- [31] Howell Tong. *Non-linear Time Series, a dynamical system approach*. Oxford Statistical Science Series. Clarendon Press Oxford, 1990.
- [32] L.A. Zadeh. Fuzzy sets. *Information and Control*, (8):338–53, 1965.