

ENERGETICA-MODELAMIENTO ESTADÍSTICO DEL PRECIO SPOT BRASILEÑO USANDO ANFIS

Juan David Velásquez¹, Isaac Dyner Resonsew², Reinaldo Castro Souza³

Resumen

En este artículo, se examina como la especificación de modelos ANFIS para el modelado de series de tiempo financieras puede ser guiado por una estrategia estadística. En primer lugar, se muestra como ANFIS puede ser interpretado como un modelo autoregresivo no lineal de series de tiempo; posteriormente se analiza una estrategia de análisis de series de tiempo no lineales para la especificación inicial de modelos ANFIS. Finalmente, se presenta el caso del pronóstico del precio spot de la electricidad en el Brasil, como un ejemplo de aplicación. Como resultado del estudio, se sugiere que la metodología modificada sea parte integral del modelamiento de series de tiempo usando ANFIS

Palabras Claves: Series de Tiempo, ANFIS, Modelos no lineales, Pronóstico, Metodología de Modelamiento

Abstract

In this article, we examine how the specification of ANFIS for financial time series modeling must be guided for a statistical strategy. In first time, we show how ANFIS must be interpreted as a nonlinear autoregressive model for time series; later, we analyze a nonlinear time series analysis strategy for initial ANFIS specification. Finally, Brazilian electricity spot price forecasting case is presented as application example. As result of study, it's suggested that this modified strategy should become an integral part of modeling methodology of times series using ANFIS.

Key Words: Time Series, Fuzzy Systems, ANFIS, Nonlinear Models, Forecasting, Modeling Methodology

¹ Juan David Velásquez. Escuela de sistemas. Universidad Nacional de Colombia, sede Medellín. Email: jdvelasq@unalmed.edu.co

² Isaac Dyner Resonsew. Escuela de sistemas. Universidad Nacional de Colombia, sede Medellín. E-mail: idyner@unalmed.edu.co

³ Reinaldo Castro Souza. Departamento de Engenharia Elétrica – Pontificia Universidade Católica – RJ. E-mail: reinaldo@ele.puc-rio.br

1 Introducción

En el modelamiento estadístico, se considera que una serie de tiempo $\{y_t\}_1^T$, puede ser representada como el resultado de aplicar una función computable y desconocida $f(\cdot)$ al conjunto de regresores $\mathbf{x}_t = \{x_t^{(1)}, \dots, x_t^{(l)}\}$, tal que:

$$y_t = f(x_t) + \mathbf{n}_t \quad (1)$$

donde la serie de ruidos estocásticos independientes e idénticamente distribuidos, $\{\mathbf{n}_t\}_1^T$, sigue una distribución con media cero y varianza σ_v^2 . El objetivo fundamental del modelamiento de series de tiempo es construir una función $g(\cdot)$ que aproxime (1) con algún grado de precisión, valiéndose de los datos disponibles $\{y_t, \mathbf{x}_t\}_1^T$, tal que:

$$y_t = g(x_t) + \mathbf{v}_t \quad (2)$$

usualmente con el fin de estimar pronósticos de los valores futuros de la serie, $\{y_t\}_T^{T+l}$, para un horizonte de l períodos hacia delante, aplicando $g(\cdot)$ al grupo de regresores $\{x_t\}_T^{T+l}$.

Ya que usualmente la forma de $g(\cdot)$ es desconocida, diversos modelos estadísticos paramétricos han sido propuestos para su aproximación [véase Holt (1957), Winters (1960), Brown (1963), Box y Jenkins (1970), Pole *et al* (1994), Harvey (1989)]. Sin embargo, cuando la serie presenta no linealidades importantes cuya forma funcional es desconocida, los métodos no paramétricos de regresión para aproximación de funciones no lineales se hacen atractivos para esta tarea [véase por ejemplo a Tang y Fishwick (1993), Masters (1993), Masters (1995), Kaastra y Boyd (1996) en el caso de modelos de redes neuronales artificiales].

ANFIS (Jang, 1993) interpretado en el contexto estadístico, como un modelo no paramétrico de regresión no lineal, puede ser usado para aproximar $f(\cdot)$, ya que es un aproximador universal de funciones. A diferencia de otras metodologías, permite incorporar conocimiento experto subjetivo sobre la evolución de la serie de tiempo e incluir simultáneamente información extraída de $\{y_t, x_t\}_1^T$. 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 (1994) propone el uso del algoritmo CART desarrollado por Breiman *et al* (1984) para determinar la estructura inicial de ANFIS y posteriormente afinar los parámetros del modelo usando un algoritmo de optimización. No obstante, esta estrategia de especificación no contempla muchos de los problemas específicos del modelamiento de series de tiempo. Por otro lado, la estrategia general para la especificación y aceptación de modelos de series de tiempo, formulada originalmente por Box y Jenkins (1970), y ajustada por diferentes autores [principalmente Tong (1990), Granger y Teräsvirta (1993), y Harvey (1989)], ignora las características particulares de ANFIS.

En consecuencia el objetivo de este artículo, es realizar una revisión paso a paso de una estrategia que permita la especificación formal de ANFIS en el modelamiento de series de tiempo, basada en una estrategia estadística que tenga en cuenta sus características no lineales. En primer lugar, se discute la interpretación de ANFIS como un modelo de series de tiempo; posteriormente, se propone un procedimiento para su especificación.

2 ANFIS como un modelo de series de tiempo

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. Un análisis que permite llegar a esta conclusión es presentado a continuación. Diversas especificaciones para $g(\cdot)$ han sido presentadas en la literatura. En la familia de modelos de Box y Jenkins (1970), $g(\cdot)$ definida en (2) es la ecuación paramétrica lineal:

$$g(x_t) = g(x_t | \{\hat{a}_j\}_0^U) = \hat{a}_0 + S \hat{a}_j \cdot x_t^{(j)} \quad (3)$$

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

$$\{\hat{a}_j\}_0^U \equiv \arg \min_{\{a_j\}_0^U} \sum_{t=1}^T [y_t - g(x_t | \{a_j\}_0^U)]^2 \quad (4)$$

En (3), el vector de regresores x_t toma diferentes formas, de acuerdo con el tipo de modelo considerado. En el caso general, x_t está formada por valores rezagados de y_t , $\{y_{t-j}\}_{j=1}^p$, los términos aleatorios $\{v_{t-j}\}_{j=1}^p$ y los rezagos de otras variables causales $\{u_{t-j}\}_{j=1}^p$. Esta familia de modelos sólo es exitosa en la medida de que $g(x_t|\{\hat{a}_j\}_0^U)$ se aproxime a la función real $f(\cdot)$ definida en (1). Ello implica que en la función aproximada $f(\cdot)$ las relaciones entre los regresores x_t y y_t son lineales, y segundo, y que sus relaciones son constantes en el tiempo, ya que los coeficientes $\{\hat{a}_j\}_0^U$ son independientes de t .

Una generalización de los modelos de Box y Jenkins (1970), es permitir que el dominio del regresor $x_t^{(i)}$ sea particionado en K regiones disyuntas, $\{R_k\}_{k=1}^K$, dentro de las cuales la serie de tiempo sigue un comportamiento lineal, y asignar un modelo lineal a cada una de ellas [véase Tong (1990, Sección 3.3.1.1)]:

$$\text{if } x_t^{(i)} \in R_k \rightarrow y_t = g_k(x_t | \{\hat{a}_j^{(k)}\}_0^U) + v_t^{(k)} \quad (5)$$

Reciben el nombre de SETAR o SETARMA de acuerdo con el tipo de modelo de Box y Jenkins con que se implemente el consecuente de la regla.

Una generalización directa de (5) es permitir que las regiones $\{R_k\}_{k=1}^K$ 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.

Si se toma como variable de transición a la i -ésima variable del conjunto de regresores, $x_t^{(i)}$, un modelo con dos regiones puede escribirse como:

$$\begin{aligned} \text{if } x_t^{(i)} \in A_1 &\rightarrow y_t = g_k(x_t | \{\hat{a}_j^{(1)}\}_0^U) + v_t \\ \text{if } x_t^{(i)} \in A_2 &\rightarrow y_t = g_k(x_t | \{\hat{a}_j^{(2)}\}_0^U) + v_t \end{aligned} \quad (6)$$

con $g(\cdot)$ definida en (3). El sistema de reglas formado por (6) es un caso particular del sistema adaptativo de inferencia neurodifusa [ANFIS por sus siglas en inglés] desarrollado por Jang (1993), el cual equivale a un sistema difuso con reglas de Takagi-Sugeno. En el caso general, los antecedentes de las reglas pueden contener más de una variable, e inducen una partición del dominio de la serie de tiempo. La formulación

presentada en (6) difiere del modelo original propuesto por Jang (1993) en que se incluye en el consecuente de cada regla a v_t [el cual es el mismo término para todas reglas]; igualmente (6) permite generalizar la formulación de ANFIS, al caso en que los conjuntos de variables que conforman los antecedentes y los consecuentes pueden ser diferentes. El proceso de cálculo de y_t en (6), permite que dicho sistema sea transformado a la ecuación paramétrica lineal definida en (3), tal que los coeficientes $\{\hat{a}_j\}_0^U$ son calculados como:

$$a_j = \frac{m(x_t^{(i)}; \mathbf{a}_1, \mathbf{b}_1, \mathbf{g}_1) \times a_j^{(1)}}{m(x_t^{(i)}; \mathbf{a}_1, \mathbf{b}_1, \mathbf{g}_1) + m(x_t^{(i)}; \mathbf{a}_2, \mathbf{b}_2, \mathbf{g}_2)} + \frac{m(x_t^{(i)}; \mathbf{a}_2, \mathbf{b}_2, \mathbf{g}_2) \times a_j^{(2)}}{m(x_t^{(i)}; \mathbf{a}_1, \mathbf{b}_1, \mathbf{g}_1) + m(x_t^{(i)}; \mathbf{a}_2, \mathbf{b}_2, \mathbf{g}_2)} \quad (7)$$

De esta forma, (7) representa un modelo autoregresivo cuyos coeficientes $\{\hat{a}_j\}_0^U$ son una función continua de la variable de decisión $x_t^{(i)}$.

Esta interpretación implica que existe un modelo autoregresivo lineal por cada punto del dominio de la variable de decisión $x_t^{(i)}$ usada en los antecedentes de las reglas; mientras que para los modelos SETAR y SETARMA, 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. Esta es una clara ventaja en el modelamiento de sistemas no lineales.

3 Estrategia para la especificación de ANFIS

El método presentado para la especificación y aceptación de modelos, se basa fundamentalmente en la estrategia general desarrollada por Box y Jenkins (1970), la cual ha sido ajustada por diferentes autores [principalmente Tong (1990), Granger y Teräsvirta (1993), Harvey (1989) Masters (1993), Masters (1995) y Kaastra y Boyd (1996)] para ser aplicada en la construcción de modelos no lineales. Los pasos fundamentales son presentados en la Tabla 1.

Tabla 1. Estrategia para la Especificación de Modelos.

Paso 1:	Identificación
	- Selección Inicial de Regresores
	- Análisis Exploratorio Preliminar
	- Selección de Regresores
	- Especificación de $g(\cdot)$
	- Estimación de parámetros
Paso 2:	Diagnóstico y Análisis de la Dinámica de la Serie
Paso 3:	Metadiagnóstico
Paso 4:	Construcción de Pronósticos e Intervalos de Confianza

3.1 Paso 1: Identificación

La identificación es un proceso repetitivo destinado a identificar las propiedades de la serie de tiempo, sus regresores y la forma del modelo.

3.1.1 Selección de Variables Potencialmente Explicativas

Un problema fundamental es la selección correcta del conjunto de regresores x_t . Durante la primera fase, el pronosticador establece un conjunto de regresores potenciales:

$$x_t^* = ((x_t^{*(1)}, \dots, (x_t^{*(D)})) \quad (8)$$

que a juicio del pronosticador (Bunn y Wright, 1991), influyan sobre y_t . Esta selección es basada en un modelo mental, que simplifica el proceso generador de la serie de tiempo, debido a las limitaciones propias de la mente humana (Hogarth, 1994); en consecuencia, pueden incluirse variables explicativas que presentan correlaciones espúreas con la variable pronosticada, sin que exista una justificación física o económica que valide su presencia. Bunn y Wright (1991) indican que el proceso de construcción de x_t^* , debe ser realizado de forma racional garantizando la objetividad y correctitud en la selección, y acompañado por un mecanismo de documentación que permita la posterior reconstrucción

del proceso de selección con fines de reevaluación y aprendizaje. Los autores han encontrado que las técnicas de Análisis Decisional, tal como los Diagramas de Influencia, son apropiadas para realizar un proceso formal y documentado de selección de las variables que conformarán x_t^* , permitiendo su defensa posterior durante una evaluación expost.

3.1.2 Análisis Exploratorio Preliminar

Las técnicas de análisis exploratorio permiten recopilar información sobre las características de la serie de tiempo, y pueden ayudar al experto a corroborar cuales de los regresores potenciales inicialmente seleccionados poseen realmente una relación funcional con y_t . Tong (1990) indica que los diagramas de dispersión entre $(y_t, x_t^{(j)})$, los histogramas bivariados, o los estimados de $E[y_t/x_t^{(j)}]$ y $V[y_t/x_t^{(j)}]$ basados en métodos no paramétricos para aproximar la función de densidad de probabilidad [véase trabajos de Izenman (1991) y Katkovnik y Shmulevich (2002)], pueden ser útiles ante la sospecha de relaciones no lineales. 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 de selección.

El examen de los autocorrelogramas, autocorrelogramas parciales, correlogramas cruzados y correlogramas cruzados parciales de la serie de tiempo, podría entregar información de utilidad; no obstante, su uso como medidas de dependencia puede ocasionar que se descarten variables de influencia altamente no lineal. Otro estadístico que puede ser usado como un criterio de medición es la Información Mutua [véase Bonnlander (1996), Battiti (1994), Kwak y Choi (2002), Pompe (1993) y Pompe (1996)].

3.1.3 Selección de los Regresores

Su objetivo es conformar el vector final de regresores x_t , el cual se define con todas aquellas variables de x_t^* , para las cuales se ha probado que existe una relación de causalidad con y_t , ya sea mediante una medida independiente del modelo, tal como la correlación o la Información Mutua; o por una medida dependiente de él, como el error cuadrático medio.

3.1.4 Especificación de $g(\cdot)$ y Estimación de Parámetros

Con base en los resultados del análisis exploratorio de los datos, y los regresores seleccionados, se pueden especificar diferentes modelos alternativos, cuya coherencia será validada a través del diagnóstico. La estimación de parámetros es realizada a través de técnicas convencionales comúnmente aceptadas.

3.2 Paso 2: Diagnóstico y Análisis de la Dinámica de la Serie

La derivación de ANFIS como un modelo de series de tiempo, parte de la hipótesis de normalidad de los errores. Su falta de cumplimiento es un indicativo de que $g(\cdot)$ no ha capturado adecuadamente las propiedades de $\{y_t\}_1^T$. El proceso de diagnóstico consiste en verificar el cumplimiento de dicha hipótesis una vez el modelo 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 del modelo a la serie, y el análisis estadístico de los resultados, que permitirá corroborar los juicios y detectar inconsistencias que no son fácilmente visibles. 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. En consecuencia, el análisis de los errores puede dar indicaciones sobre problemas en la especificación de $g(\cdot)$ a la selección de los regresores que conforman x_t . Sin embargo, contiene una alta cantidad de juicio subjetivo para determinar si el modelo se ajusta apropiadamente a los datos.

Durante la fase de diagnóstico, la presencia de componentes estructurales en la serie de los residuos, revelan una inconsistencia dentro de la especificación del modelo, 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. Estos efectos se consideran conjuntamente como un ruido no visible presente en la serie de tiempo (McCleary y Hay, 1980, pág. 227), y es el causante de muchos de los problemas de modelamiento.

Una vez han sido detectadas inconsistencias en el modelo, 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.

3.3 Paso 3: Metadiagnóstico

Una vez se ha ajustado el modelo, usualmente se realizan dos pruebas de aceptación que miden su bondad en la representación de los datos; la primera 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, la varianza del predictor y la desviación media absoluta (MAD).

La segunda tiene por objetivo descartar otros modelos que podrían ajustarse a los datos; ella se realiza eliminando o adicionando términos del modelo y demostrando que ellos son necesarios para el caso de la eliminación, o son estadísticamente no significativos en el caso de su adición. Medidas como el Criterio de Información de Akaike, Bayes o Schwartz son comúnmente usadas en el caso lineal.

Adicionalmente, 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. Criterios (Harvey, 1989) como parsimonia, coherencia con los datos, consistencia con el conocimiento previo, admisibilidad de los datos, estabilidad estructural o encompassamiento pueden ser de utilidad.

3.4 Paso 4: Construcción de Pronósticos e Intervalos de Confianza

Cuando se ha finalizado el proceso de especificación y validación del modelo, se procede a la construcción de pronósticos e intervalos de confianza para ellos. El pronóstico para un periodo adelante, y_{T+1} , sigue usualmente una distribución normal, y es calculado de forma directa usando (2). Cuando x_t contiene uno o más valores rezagados de y_t , la construcción del pronóstico j periodos adelante, requiere el uso de simulación de Montecarlo (Fishman, 1996), ya que ANFIS realiza una

transformación no lineal de los datos de entrada. Un 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 de tiempo 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 del modelo, ya que pueden revisarse y revalidarse muchos de los supuestos del modelo cuando fue especificado originalmente, haciéndose vital un proceso de documentación ordenado y minucioso.

4 Aplicación al pronóstico del precio de electricidad en Brasil

Las últimas décadas se han caracterizado por procesos de desregulación y liberalización de diferentes sectores eléctricos en el mundo [véase a Steiner (2000) para una revisión general]. 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 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. Estos submercados 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 (1997), y las características detalladas de la

metodología de solución y representación del sistema son presentadas por Maceira (1993) y Maceira *et al.* (1998)].

La inspección visual de la serie de precios spot, presentada en la Figura 1, revela varias características interesantes. 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 evolución del embalse agregado durante el mismo periodo de tiempo. 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 este mismo periodo de tiempo, el embalse agregado tiene niveles más altos que el periodo anterior, y la gráfica de energía afluente muestra un volumen de aportes superior respecto al año anterior.

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; durante este mismo año, el embalse agregado tiene un comportamiento similar al ocurrido a finales de 1996 y principios 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; durante este mismo año, el nivel del embalse agregado a caído rápidamente, hasta alcanzar el valor más bajo de la muestra de datos en dic/99, con un energía almacenada del 18% de la media multianual de largo plazo (MMLP); este comportamiento es congruente con la reducción de los aportes hidrológicos, los cuales presentan una clara tendencia descendente hasta dic/99; 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, los cuales coinciden con el periodo de recuperación del embalse agregado y el aumento en los aportes.

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. El año 2001 se caracteriza por tener aportes hidrológicos muy por debajo de los valores normales, aunque de tendencia creciente, causando que el embalse agregado no pueda recuperar su nivel; en $\text{ene}/2001$ se inicia un ascenso acelerado del precio spot, desde 56.9 $\text{\$/kWh}$ hasta 684 $\text{\$/kWh}$ registrados en $\text{jun}/01$. A medida que la energía

afluente va aumentando, el precio spot empieza a disminuir, pasando de su nivel máximo en $\text{sep}/00$ a 336 $\text{\$/kWh}$ en $\text{oct}/200$, 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.

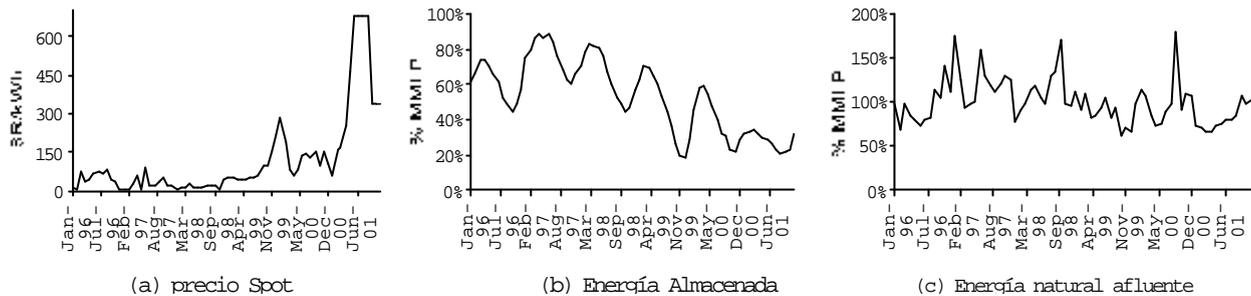


Figura 1. Evolución mensual del precio Spot, la energía almacenada, y la energía natural afluente.

4.1 Identificación

El proceso de selección inicial de los regresores basado en la emisión de juicios informados, es realizado utilizando la metodología de Diagramas de Influencia; el diagrama final obtenido es presentado en la Figura 2. De dichas variables, solo se encuentran disponibles los aportes hidrológicos al embalse agregado, medidos bajo la variable energía natural afluente, expresada como el porcentaje de la media multianual de largo plazo; y la energía almacenada en el embalse agregado, como el porcentaje de la media de largo plazo. La evolución del precio spot y ambas variables causales se presentan en la Figura 1. Una inspección visual rápida permite corroborar las relaciones causales planteadas en la construcción del diagrama causal.

El análisis exploratorio preliminar de la serie de tiempo, permite construir un conjunto de evidencia que soporten y confirmen los juicios realizados por el experto en la construcción del Diagrama de Influencia. La serie de precios spot no cumple con las características de homogeneidad en el sentido estadístico; una transformación de los datos puede resultar beneficiosa en este aspecto; sin embargo, la decisión del tipo de transformación está basada en el juicio experto del modelador. En este caso se usa la transformación de Box - Cox con $\lambda = 0.065$. La serie transformada de precios spot se presenta en la Figura 3. En ella, se puede

visualizar con mayor detalle las características ya discutidas de la serie de tiempo.

El periodograma de la serie de precios [véase la Figura 4a], indica la predominancia de componentes cíclicas con periodos superiores al año, y solo una componente importante en el ciclo intranual con un periodo de 10.3 meses. Las componentes cíclicas de largo plazo, se explican por los valores altos del precio spot durante el último semestre de 2001.

El autocorrelograma y el autocorrelograma parcial de la serie de precios spot indican la existencia de una relación causal entre los precios con rezagos uno y dos y el precio actual. Los diagramas de dispersión entre el precio spot actual y el rezagado, presentados en la Figura 5, indican la existencia de una relación no lineal para los dos primeros rezagos del precio spot; 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 periodos son presentados en las Figuras 6 y 7; ellos son estimados usando una aproximación no paramétrica de la función de densidad de probabilidad basada en el kernel de Epanechnikov. Estas Figuras 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.

El análisis de las correlaciones entre el precio spot con la energía natural afluyente, y con la energía almacenada indican una relación negativa de estos regresores con el precio spot, y positiva entre ellos, tal como se indicó durante la construcción del diagrama causal. Los diagramas de dispersión presentados en las Figuras 8 y 9, corroboran dicha afirmación.

La inspección visual de la serie de energía almacenada (EA) [véase la Figura 1b] muestra una clara componente estacional asociada al ciclo anual de las estaciones, y una componente cíclica de largo plazo, posiblemente asociada a condiciones macroclimáticas [que debería corroborarse posteriormente]; el patrón de la componente estacional intra-anual tiene una amplitud dinámica que varía aparentemente en respuesta a los aportes hidrológicos. La inspección del periodograma [véase la Figura 4b], confirma los hallazgos de la inspección visual de la serie de energía almacenada: se presentan tres componentes de frecuencia con una alta concentración de energía, que corresponden a los periodos de 72 meses ($f = 0.0139$), 18 meses ($f = 0.0556$) y 12 meses ($f = 0.0833$); las dos primeras se asocian al ciclo de largo plazo, mientras que la última corresponde al ciclo intra-anual. La media de la muestra de datos es 0.53, su varianza es 0.04, su coeficiente de asimetría es -0.06, y su curtosis es -1.18.

La inspección visual de la serie de energía natural afluyente (ENA) [véase la Figura 1c] muestra aparentemente una componente cíclica de largo plazo, relacionada con la misma componente en la serie de EA. Igualmente presenta componentes cíclicas importantes dentro del ciclo anual, asociados a la rápida variación de los aportes de un mes a otro. El periodograma presentado en la Figura 4c indica que hay varios ciclos intra-anales, así como otros ciclos de periodo superior al año; los ciclos intra-anales tienen periodos de 3, 4.8, 8 y 12 meses, y que los de largo plazo tienen periodos de

18 y 36 meses respectivamente. La media de la muestra de datos es 1.00, su varianza es 0.06, su coeficiente de asimetría es 1.12, y su curtosis es 1.53.

4.2 Modelo Univariado

Como primer paso, se desarrolla un modelo univariado, y posteriormente, se incorporaran variables causales para contrastar los resultados. 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.

4.2.1 Especificación del Modelo y Estimación de Parámetros

En este caso se especifica un modelo ANFIS con dos reglas difusas, el cual ya fue presentado en (6). Tanto la variable de decisión, como las variables de los consecuentes, se seleccionaron de acuerdo con los resultados del análisis exploratorio preliminar. La variable de decisión $x_t^{(i)}$ corresponde al precio spot rezagado una o dos unidades de tiempo.; los consecuentes de las reglas en (6) son modelos autoregresivos de orden dos. La selección del modelo final, se realiza estimando los parámetros de los modelos obtenidos al considerar todas las posibles combinaciones de antecedentes y consecuentes en (6). 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.

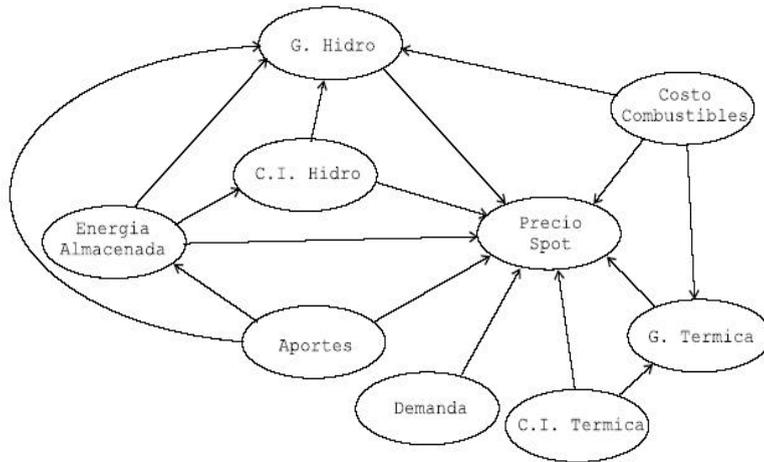


Figura 2. Diagrama de Influencia para el precio Spot.

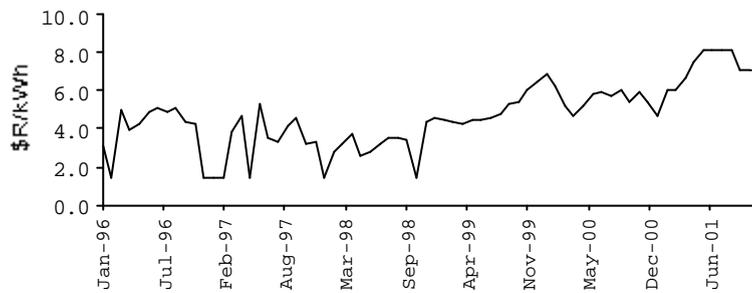


Figura 3. Transformación de Box.Cox del precio Spot. Caso Brasileño.

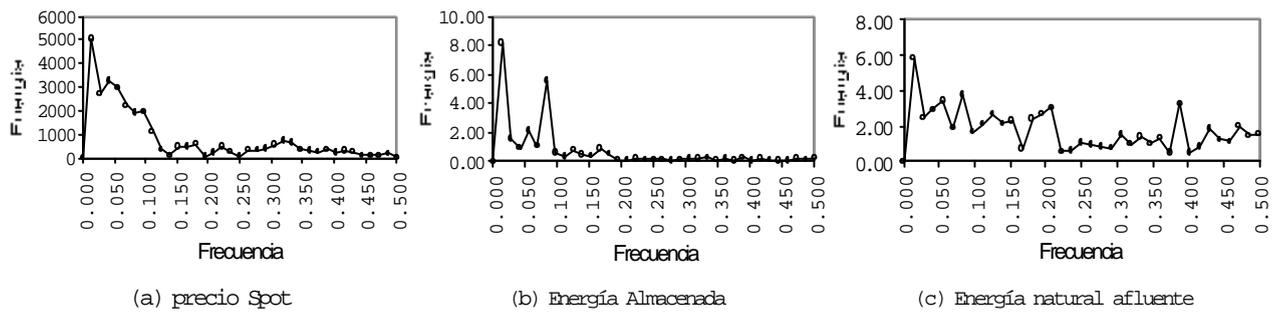


Figura 4. Espectro de Energía. Caso Brasileño

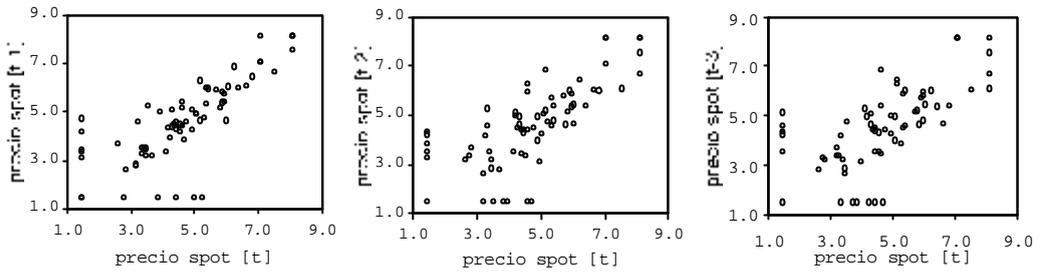


Figura 5. Diagramas de Precio Spot [t] vs Precio Spot [t-i].

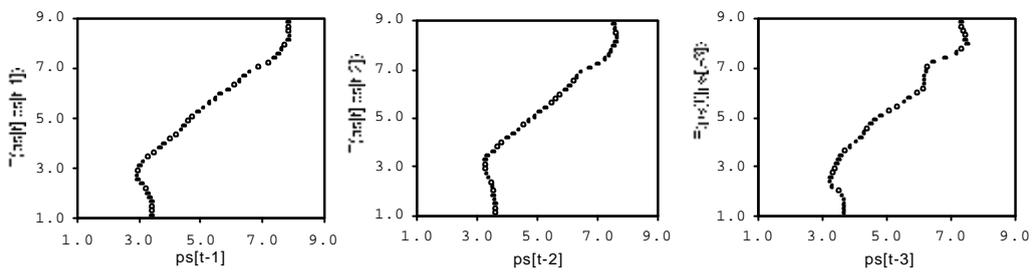


Figura 6. Estimado de $E(\text{precio}[t] | \text{precio}[t-i])$.

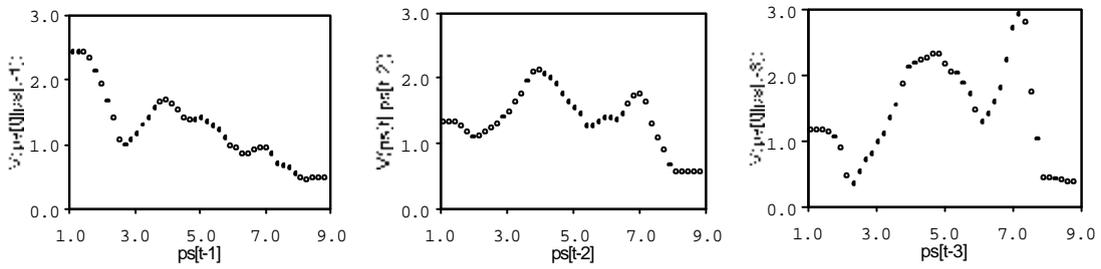


Figura 7. Estimado de $V(\text{precio}[t] | \text{precio}[t-i])$.

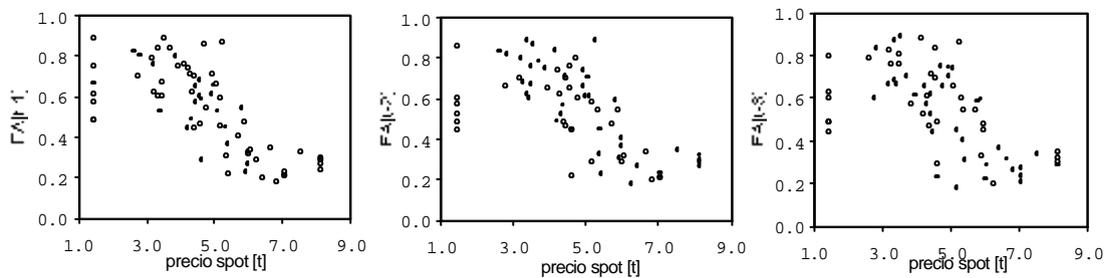


Figura 8. Diagramas de Precio Spot [t] vs Energía Almacenada [t-i]

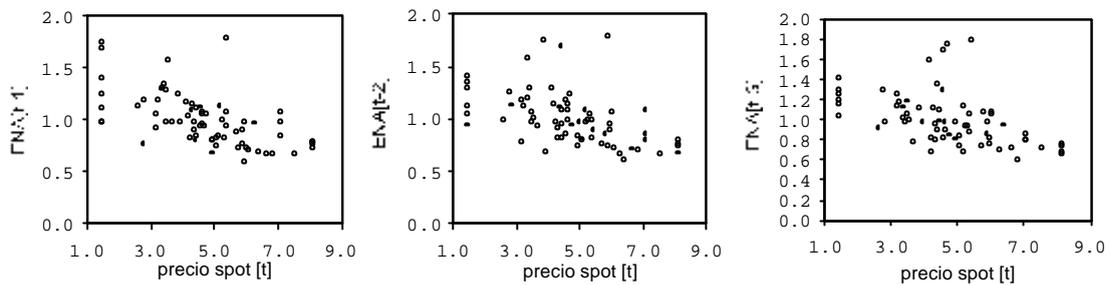


Figura 9. Diagramas de Precio Spot [t] vs Energía Natural Afluyente [t-i].

4.2.2 Diagnóstico

Para el mejor modelo encontrado, la varianza del predictor es 0.8680, mientras que la desviación media absoluta es 24.68%. Se encontró una correlación de 0.7871 entre los pronósticos un paso adelante y la serie real. El coeficiente de determinación es 0.6196. 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.

En la Figura 10a, se presenta el pronóstico un paso adelante [línea continua] y sus correspondientes intervalos de confianza [línea punteada] estimados a dos desviaciones estándar del error, así como la serie real [círculos]. En ella puede observarse, como el modelo tiene problemas para pronosticar los puntos donde el precio spot toma su valor mínimo de 4 \$/kWh, excepto para nov/98. Igualmente, puede verse como el modelo pronostica con más frecuencia valores por debajo de la serie real. Así mismo, la magnitud de los errores es mucho más alta durante la primera mitad del periodo de calibración respecto a la segunda mitad.

4.2.3 Metadiagnóstico

En el pronóstico extrapolativo [véase la Figura 10b], el modelo difuso 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%. Debe notarse la asimetría de

los intervalos de incertidumbre ubicados a dos desviaciones estándar del error, debido a la transformación no lineal del modelo.

4.3 Modelo Univariado con variables causales

La inclusión de los términos asociados a la energía almacenada, y la energía natural afluyente en los consecuentes de las reglas, busca permitir una mejor adaptación del modelo a la serie de tiempo, y mejorar varias deficiencias del modelo detectadas en el diagnóstico.

4.3.1 Diagnóstico

Al igual que en el modelo univariado, se estiman los parámetros del modelo para diferentes variables de decisión, esto es, el precio spot, la energía almacenada, y la energía natural afluyente rezagados una o dos unidades de tiempo; y se escoge el mejor de ellos. El mejor modelo encontrado usa como variable de decisión el precio spot con rezago dos. La predicción un paso adelante es presentada en la Figura 11a. La varianza del predictor es 0.3550, y la desviación media absoluta es del 13.42%; la correlación entre las observaciones y el pronóstico un paso adelante es 0.9189, y el coeficiente de determinación 0.8444. Los estadísticos indican que los residuos siguen una distribución normal. El estadístico N es 1.34, el coeficiente de asimetría es -0.21 y la curtosis es 3.56. No hay indicios de heterocedasticidad en los residuos. La correlación serial indica dependencia para los rezagos 5 y 6. La inspección visual de la gráfica de pronóstico un paso adelante, muestra una reducción pronunciada de los intervalos de confianza respecto al caso univariado. Presentan una probabilidad muy baja de ocurrencia las observaciones en mar/96, jun/97, may/97, oct/97, ene/98, nov/98,

ene/01 y jun/01. El periodo entre feb/99 y may/99 presenta pronósticos por debajo de las observaciones reales.

4.3.2 Metadiagnóstico

El pronóstico por fuera de la muestra de calibración es presentado en la Figura 11b; es este caso, la desviación media absoluta es de 4.99%, y la varianza del predictor es 0.1771, valores cualitativamente superiores a los encontrados para el modelo sin variables causales.

5 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 de tiempo. 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 de tiempo 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 de tiempo, 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 problemas de series de tiempo.

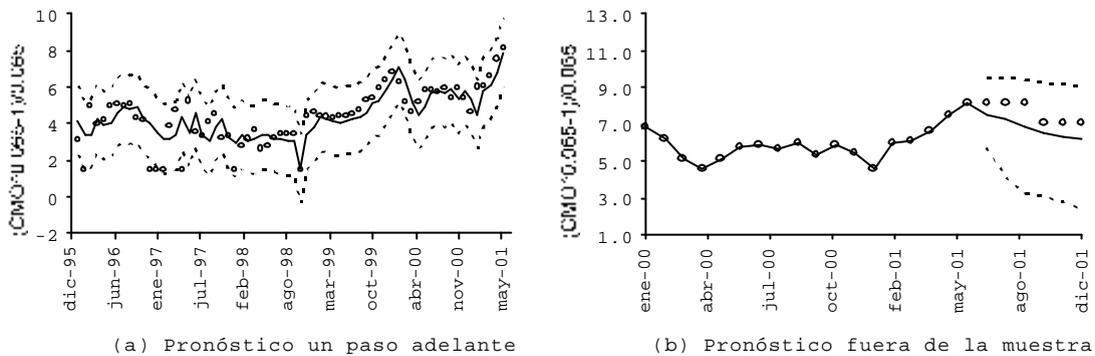


Figura 10. Pronóstico en Modelo Univariado.

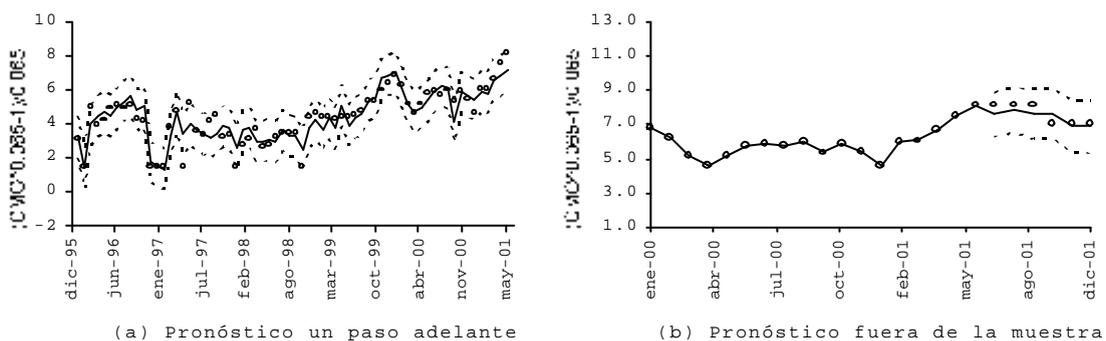


Figura 11. Pronóstico en Modelo Univariado con Variables Causales.

REFERENCIAS

- BATTITI, R. Using mutual information for selecting features in supervised neural net learning. En: *IEEE Transactions on Neural Networks*, Vol. 5, no. 4 (1994), p. 537-550.
- BONNLANDER, B. Nonparametric selection of input variables for connectionist learning. PhD thesis, University of Colorado. Department of Computer Sciences. (1996).
- BOX G.E.P. and JENKINS, G.M. Time Series Analysis: Forecasting and Control. Holden-Day Inc. 1970.
- BREIMAN, L. FRIEDMAN, J.H. OLSHEN, R.A. and STONE, C.J. Classification and Regression Trees. *Technical Report*. Wadsworth, Inc. Belmont, California. 1984.
- BROWN, G.R. Smoothing, Forecasting and Prediction. Englewood Cliffs. Prentice Hall. 1963.
- BUNN, D. and WRIGHT, G. Interaction of judgemental and statistical forecasting methods: Issues and analysis. En: *Management Science* Vol. 37, no. 5 (1991).
- FISHMAN, G.S. Monte Carlo. Concepts, Algorithms and Applications. Springer Series in Operational Research, Springer. 1996.
- GRANGER, C.W.J. and TERÄSVIRTA, T. Modelling Nonlinear Economic Relationships. Oxford University Press. 1993.
- HARVEY, A.C. Forecasting, structural time series models and the Kalman Filter. Cambridge University Press. 1989.
- HOLT, C.C. Forecasting seasonal and trends by exponentially weighted moving averages. En: *ONR Research Memorandum*, Carnegie Institute of Technology, Pittsburg, Pennsylvania, no 52, (1957).
- HOGARTH, R. Judgement and Choice. Second Ed John Wiley and Sons. 1994.
- IZENMAN, A.J. Recent developments in nonparametric density estimation. *Journal of the American Statistical Association*, no. 86 (1991), p. 205-224.
- JANG, J.S.R. ANFIS: Adaptive network based fuzzy inference system. En: *IEEE Transactions on Systems, Man and Cybernetics*, no. 23 (1993), p. 665-684.
- JANG, J.S.R. Structure determination in fuzzy modeling: a fuzzy cart approach. En: *IEEE International Conference of Fuzzy Systems*. Orlando, Florida. (1994).
- KAASTRA, I. and BOYD, M. Design a neural network for forecasting financial and economic series. *Neurocomputing*, no. 10 (1996), p. 215-236.
- KATKOVNIK, V. and SHMULEVICH, I. Kernel density estimation with adaptive varying window size. *Pattern Recognition Letters*, (2002), p. 1641-1648.
- KWAK, N. and CHOI, C.H. Input feature selection for classification problems. En: *IEEE Transactions on Neural Networks*, Vol. 13, no. 1 (2002), p. 143-159.
- MACEIRA, M. Programao dinmica dual estocstica aplicada ao planejamento da operao energtica de sistemas hidrotmicos com representao do processo estocstico de auncias por modelos auto-regresivos peridicos. *Technical report*, (1993), CEPEL DPP/PEL 237.
- MACEIRA, M. MERCIO, C. GORENSTIN, B. CUNHA, S. SUANNO, C. SACRAMENTO, M. and KLIGERMAN, A. Energy evaluation of north/northeastern and south/southeastern interconnection with newave model. En: *IV Symposium of Specialist in Electric Operational and Expansion Planning*, SESOPE, Salvador, Brazil. (1998).
- MASTERS, T. Practical Neural Network Recipes in C++. First Ed. Academic Press, Inc. 1993.
- MASTERS, T. Neural, Novel and Hybrid Algorithms for Time Series Prediction. First Ed. John Wiley and sons, Inc. 1995.
- MCCLEARY, R. and HAY, R. Applied time series analysis for the social sciences. Sage Publications, Inc. 1980.
- PEREIRA, M. Modelo de despacho hidrotmico, *Technical report*, PSRI. (1997).
- POLE, A. WEST, M. and HARRISON, J. Applied Bayesian Forecasting and Time Series Analysis. Chapman & Hall. 1994.
- POMPE, B. Measuring statistical dependences in a time series. En: *Journal of Statistical Phys.*, no. 73, (1993), p. 587.
- POMPE, B. Ranking and entropy estimation in nonlinear time series analysis. (1996).
- STEINER, F. Regulation, industry structure and performance in the electricity supply industry. En: *OECD Economic Studies*, no. 32 (2000).
- TANG, Z. and FISHWICK, P.A. Feedforward neural nets as models for time series forecasting. *Technival Report*, Department of Decision and Information Sciences. University of Florida. (1993).
- TONG, H. Non-linear Time Series, a dynamical system approach. Oxford Statistical Science Series. Clarendon Press Oxford. 1990.
- WINTERS, P.R. Forecasting sales by exponentially weighted moving averages. En: *Management Sciences*, no. 6 (1960), 324-342.