

Predicción de series temporales no lineales con el modelo NARMAX

(Aplicación a la predicción de la demanda de energía eléctrica)

La predicción de series temporales a corto plazo, también llamada identificación de sistemas por los expertos en Automática, es una disciplina que encuentra cada día más aplicaciones en áreas de planificación, gestión, producción, mantenimiento y control de procesos industriales.

Este artículo presenta un modelo de predicción de series temporales no lineales basado en Redes Neuronales Artificiales (RNA). La utilización de RNA como aproximadores funcionales asegura la capacidad de representación y adaptación del modelo propuesto. Estas características son requisitos fundamentales para reproducir la complejidad de los procesos actuales y poder adaptarse a la continua evolución de los mismos. La aplicación del modelo y de la metodología de identificación propuestos a la predicción de la curva de demanda de energía eléctrica pretende tan sólo servir de ejemplo al amplio abanico de aplicaciones industriales que pueden ser tratadas con estos métodos.



Antonio Muñoz San Roque

Doctor Ingeniero del ICAI (1991). En 1992 se incorporó al Instituto de Investigación Tecnológica (IIT) donde desarrolla su actividad investigadora en monitorización y diagnóstico de procesos industriales con técnicas de Inteligencia Artificial.

Thomas Czernichow

Ingeniero Informático por la Escuela de Informática y Matemática Aplicada de Toulouse (1992). El mismo año obtuvo su M.Sc. en paralelismo y supercomputación en la Universidad Paul Sabatier (Toulouse). En 1996 se doctoró en "Inteligencia artificial y reconocimiento de formas" por la Universidad Pierre et Marie Curie (París). Experto del grupo de trabajo de la CIGRE especializado en la predicción de la demanda con redes neuronales



1. Introducción

Los modelos clásicos de series temporales, bajo la bandera de los modelos ARIMA de [Box and Jenkins, 1976], basan su predicción en la hipótesis de existencia de una relación de tipo lineal entre los regresores o entradas del modelo (como pueden ser valores pasados de la salida del modelo, distintos retardos de variables exógenas, ...) y la variable tomada como salida del mismo. Esta hipótesis de linealidad limita en muchos casos la precisión de la estimación, debido a la existencia de relaciones no lineales en el proceso subyacente que se quiere modelar.

Este artículo presenta un modelo de predicción de series temporales que resulta de la extensión de los modelos clásicos al campo no lineal. Esta extensión se realiza mediante la sustitución de la transformación lineal (regresores \rightarrow salida) por un aproximador funcional no lineal. Como aproximadores funcionales se utilizarán en este caso Redes Neuronales Artificiales (cuya descripción queda fuera de los objetivos de este escrito), pero el modelo presentado admite cualquier otro tipo de aproximador.

Tras presentar la estructura del modelo NARMAX se deducirán las ecuaciones necesarias para el ajuste de sus parámetros. Posteriormente se analizará la selección de los regresores mediante el análisis estadístico de sensibilidades y se presentarán dos ejemplos de aplicación a

series temporales sintéticas. Finalmente se aplicará el modelo propuesto a un problema real como es la predicción de la curva de demanda de energía eléctrica.

2. Procesos autoregresivos de media móvil (ARMA)

Es frecuente en la predicción de series temporales con modelos no lineales que el análisis de correlaciones residuales muestre un alto contenido de información en la serie del error de predicción. Esta información, enmascarada bajo forma de ruido coloreado, puede ser realimentada al estimador de tal forma que su utilización mejore considerablemente la calidad de la estimación.

Esta presencia de información en los residuos puede ser debida a una elección indebida de la estructura del modelo o a una estimación incorrecta de los parámetros del mismo. La estructura del modelo puede ser a su vez inadecuada en dos sentidos: el primer caso se presenta cuando habiendo escogido un conjunto correcto de entradas, el modelo no tiene la capacidad de aproximación suficiente para realizar la transformación de entrada/salida requerida. El segundo caso aparece cuando el conjunto de entradas del modelo es incompleto. Si una variable que tiene una influencia relevante en la salida no es suministrada como entrada al modelo, la calidad de la estimación se verá gravemente afectada. En este último caso nos vamos a centrar.

En muchos procesos del mundo real, la conjunción de diversos fenómenos puede tener una influencia aleatoria impredecible en el valor instantáneo de la salida del proceso, pero predecible en valores futuros. Podemos formular matemáticamente esta afirmación de la forma:

$$\mathbf{y}(k) = f(\mathbf{x}(k), \dots, \mathbf{x}(k-r), \epsilon(k-1), \dots, \epsilon(k-s)) + \epsilon(k)$$

Ecuación 1

donde $\mathbf{y}(k) \in \mathfrak{R}^m$ es la salida del proceso en el instante k , $\mathbf{x}(k) \in \mathfrak{R}^n$ es la entrada externa del proceso en el instante k (si el proceso tiene propiedades autorregresivas, algunas de las componentes

de \mathbf{x} podrían corresponder a valores retrasados de las salidas, por ejemplo, $x_1(k) = y_1(k-1)$) y $\epsilon(k) \in \mathfrak{R}^m$ es un proceso de ruido blanco (la contribución de $\epsilon(k)$ a $y(k)$ es impredecible).

El modelado de este tipo de procesos requiere la realimentación de las señales de error para poder estimar valores pasados del ruido. En el caso lineal se utilizan modelos autorregresivos de media móvil (ARMA) para tratar series temporales estacionarias y modelos autorregresivos integrados de media móvil (ARIMA) para el caso no estacionario ([Box and Jenkins, 1976], [Wei, 1990]). La forma general de un proceso ARMA monovariable viene dada por:

$$y(k) = \sum_{i=1}^r \alpha_i y(k-i) + \sum_{j=1}^r \beta_j \epsilon(k-j) + \epsilon(k)$$

Ecuación 2

donde $\epsilon(k)$ es un proceso de ruido blanco estacionario. Los procesos autoregresivos (AR) son un caso particular del proceso ARMA con $\beta_1 = \beta_2 = \dots = \beta_s = 0$. Los procesos de media móvil (MA) son un caso particular del proceso ARMA con $\alpha_1 = \alpha_2 = \dots = \alpha_r = 0$.

3. Estructura del modelo NARMAX

La extensión de los modelos lineales enunciados anteriormente al campo no lineal puede ser llevada a cabo mediante la aproximación de la función f (ver Ecuación 1) por un aproximador funcional no lineal f_W , de parámetros W . La estructura resultante, que será llamada modelo NARMAX, se muestra en la Figura 1, donde se ha utilizado una red neuronal del tipo RBFN ([Moody & Darken, 1988-1989], [Muñoz, 1996], [Czernichow, 1996]) como aproximador funcional. Cualquier otro modelo no lineal podría ser utilizado en lugar de la red neuronal, y los resultados que ahora se presentan siguen siendo válidos en el caso general.

El modelo NARMAX produce la estimación vectorial de la salida $\mathbf{y}(k) \in \mathfrak{R}^m$ en el instante k , tomando como entradas la serie temporal multivariable $\mathbf{x} \in \mathfrak{R}^n$ y valores pasados de la salida deseada $d \in \mathfrak{R}^n$.

Se han incluido módulos lineales de escalado a la entrada y a la salida del modelo para estandarizar las señales presentadas al aproximador funcional. Esta transformación, que trata de igualar las varianzas de las señales tratadas, mejorará de forma significativa la convergencia de la fase de aprendizaje y jugará un papel sumamente importante en el caso de utilizar algoritmos de agrupamiento o “clustering” en el entrenamiento del aproximador funcional ([Kaufman, Rousseeuw, 90]).

La función de transferencia del modelo NARMAX viene dada por:

$$y_i(k) = f_{w_i} [x_1(k), \dots, x_1(k-r_1), \dots, x_n(k), \dots, x_n(k-r_n), e_1(k-1), \dots, e_1(k-s_1), \dots, e_m(k-1), \dots, e_m(k-s_m)]$$

Ecuación 3

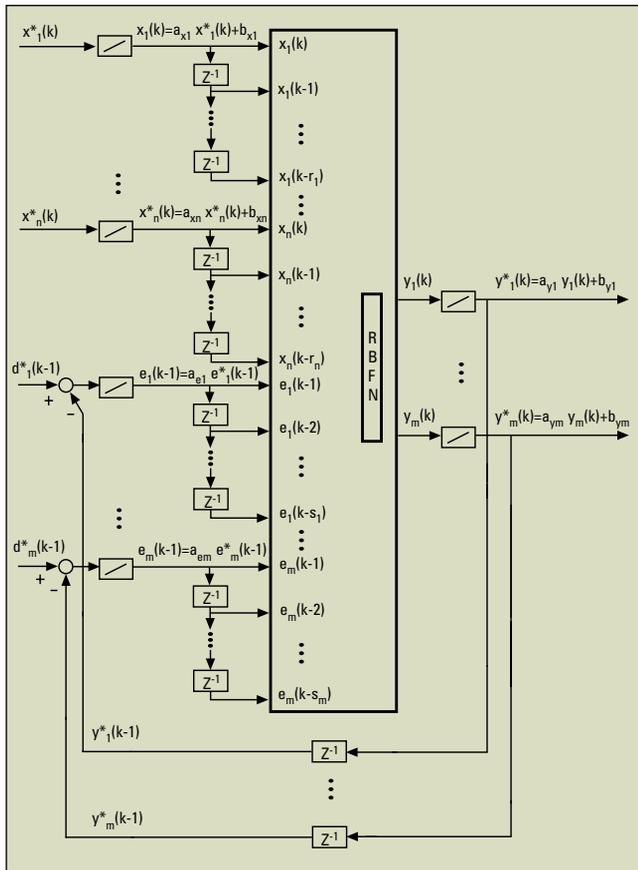


Figura 1. Estructura del modelo NARMAX

donde x , e , y y son las señales internas estandarizadas que se obtienen a partir de las señales externas (notadas con el superíndice ‘*’) según:

$$x_i(k) = a_{x_i} x_i^*(k) + b_{x_i}$$

$$e_i(k) = a_{e_i} (d_i^*(k) - y_i^*(k))$$

$$y_i^*(k) = a_{y_i} y_i(k) + b_{y_i}$$

Ecuación 4

El objetivo de estas transformaciones es obtener señales internas con media cero y desviación típica unidad.

4. Ajuste de los parámetros de un modelo NARMAX (fase de aprendizaje)

El punto de partida del aprendizaje será la obtención de tres conjuntos de datos formados por pares de entrada y salida deseada: el conjunto de entrenamiento, el conjunto de test y el conjunto de validación. Al tratar con modelos recurrentes, estos conjuntos de datos estarán compuestos por series temporales de la forma:

$$D = \{(\mathbf{x}(k), \mathbf{d}(k)), k=1, \dots, K\}$$

Ecuación 5

Plantearemos el problema del aprendizaje como un problema de optimización clásica en el que iremos minimizando de forma iterativa, respecto del vector de parámetros libres \mathbf{W} del aproximador funcional, la función de error definida sobre el conjunto de entrenamiento.

Utilizaremos el conjunto de test para detener la optimización cuando se alcance un mínimo en la función de error definida sobre este conjunto (para evitar el “sobre-entrenamiento”) y el conjunto de validación para dar por bueno el modelo finalmente ajustado.

La función de error utilizada será el error cuadrático medio definido por:

$$F(\mathbf{D}, \mathbf{W}) = \frac{1}{K} \sum_{k=1}^K \sum_{i=1}^m (d_i(k) - y_i(k))^2$$

Ecuación 6

Al utilizar métodos de optimización de primer orden (ya sean métodos quasi-Newton de baja memoria o del descenso del gradiente), será necesario calcular el gradiente de la función de error respecto de los parámetros del aproximador funcional. Este gradiente puede ser puesto de la forma:

$$\nabla_{\mathbf{w}} F = \left[\frac{\partial F}{\partial \mathbf{w}} \right] = \left[\frac{-1}{K} \sum_{k=1}^K \sum_{i=1}^m (d_i(k) - y_i(k)) \frac{\partial y_i(k)}{\partial \mathbf{w}} \right]$$

Ecuación 7

por lo que los únicos términos que faltan por calcular son las derivadas de las salidas respecto de cada parámetro w . La existencia de conexiones recurrentes en la estructura NARMAX complica el cálculo de estas derivadas. La idea general es calcular las derivadas suponiendo constantes las entradas del aproximador (como en el caso “feed-forward”), e incluir posteriormente el efecto de la realimentación calculando la derivada de cada salida respecto de cada entrada recurrente y multiplicando estos valores por las derivadas de las entradas recurrentes respecto de los parámetros (estas últimas derivadas habrán sido calculadas en pasos anteriores). Este procedimiento queda formalizado en la Ecuación 8:

$$\frac{\partial y_i(k)}{\partial \mathbf{w}} = \frac{\partial y_i(k)}{\partial \mathbf{w}} \Bigg|_e^x + \sum_{j=1}^m \sum_{h=1}^{s_j} \frac{\partial y_i(k)}{\partial e_j(k-h)} \Bigg|_w^x \frac{\partial e_j(k-h)}{\partial \mathbf{w}}$$

Ecuación 8

Los términos $\frac{\partial y_i(k)}{\partial \mathbf{w}} \Bigg|_e^x$ son las derivadas de las salidas del aproximador funcional respecto de sus

parámetros manteniendo constantes las entradas, mientras que $\frac{\partial y_i(k)}{\partial e_j(k-h)} \Bigg|_e^x$ son las derivadas de las salidas respecto de las entradas recurrentes manteniendo constantes las entradas externas y los parámetros. Sustituyendo la expresión del error resulta:

$$\begin{aligned} \frac{\partial e_j(k-h)}{\partial \mathbf{w}} &= \frac{\partial [a_{e_j}(d_j^*(k-h) - y_j^*(k-h))]}{\partial \mathbf{w}} \\ &= \frac{\partial [a_{e_j}(d_j^*(k-h) - (a_{y_j} y_j(k-h) + b_{y_j}))]}{\partial \mathbf{w}} \\ &= -a_{e_j} a_{y_j} \frac{\partial y_j(k-h)}{\partial \mathbf{w}} \end{aligned}$$

Ecuación 9

y sustituyendo en la Ecuación 8 queda finalmente:

$$\sum_{j=1}^m \sum_{h=1}^{s_j} \frac{\partial y_i(k)}{\partial e_j(k-h)} \Bigg|_w^x a_{e_j} a_{y_j} \frac{\partial y_j(k-h)}{\partial \mathbf{w}}$$

Ecuación 10

Esta expresión ofrece un algoritmo recurrente para la evaluación del gradiente del error del modelo NARMAX. Los primeros términos de la serie de derivadas, $\partial y_i(k)/\partial \mathbf{w}$, se inicializan a cero para $k=-1, \dots, -s_j$ para que el cálculo del gradiente parta siempre del mismo punto inicial.

Algunos autores entrenan modelos recurrentes utilizando una versión simplificada del gradiente, donde desprecian el efecto que conlleva una variación de los parámetros en futuros valores de las entradas recurrentes (esto equivale a despreciar el segundo término de la Ecuación 8). Esta simplificación puede tener graves consecuencias durante el aprendizaje. Por un lado el algoritmo de aprendizaje puede quedarse estancado en mínimos locales artificiales donde el primer término de la Ecuación 8 es

nulo, pero no el segundo. Por otro lado esta práctica imposibilita la aplicación de eficientes algoritmos de optimización que son muy sensibles a la precisión del gradiente, como los métodos quasi-Newton y del gradiente conjugado ([Luenberger, 1984]).

5. Análisis estadístico de sensibilidades con el modelo NARMAX

Una vez que un modelo NARMAX ha sido convenientemente entrenado de tal forma que es capaz de reproducir la dinámica del sistema que se está modelando de una forma aceptable, es posible analizar la influencia de las variables de entrada del aproximador para determinar qué retardos de las variables de entrada son realmente relevantes. Este estudio permitirá podar aquellas conexiones a entradas no influyentes en la salida, aumentando la capacidad de generalización del modelo finalmente ajustado.

El conjunto de variables de entrada del aproximador funcional de un modelo NARMAX está compuesto por dos tipos de variables estructuralmente distintas. En primer lugar están las entradas externas no recurrentes $x_i(k-j)$ (con $i=1, \dots, n$ y $j=0, 1, \dots, r_i$) que no dependen ni del vector de parámetros \mathbf{W} del aproximador, ni del resto de entradas en instantes anteriores. En segundo lugar aparecen como entradas el conjunto de errores de salida realimentados $e_i(k-j)$ (con $i=1, \dots, m$ y $j=0, 1, \dots, s_i$) que dependen de valores pasados de las salidas, que a su vez son funciones del vector de parámetros \mathbf{W} y de valores pasados de cada una de las entradas del aproximador.

Esta influencia de las salidas en algunas de las entradas genera interdependencias entre las entradas de tal forma que la influencia de una entrada en futuros valores de las salidas puede involucrar a otras entradas. El cálculo de las derivadas de las salidas respecto de las entradas teniendo en cuenta estas dependencias requiere nuevamente un algoritmo recurrente.

Sin embargo, para analizar las variables de entrada que en el instante k está utilizando el aproxima-

dor para estimar el vector de salidas en el mismo instante k , hemos de analizar las sensibilidades de las salidas respecto de cada una de las entradas, manteniendo constantes el resto de las entradas (derivadas en lazo abierto). Por ejemplo, dado el modelo NARMAX:

$$y(k) = f_w(d(k-1), \dots, d(k-5), e(k-1), \dots, e(k-5)),$$

con $e(k) = d(k) - y(k)$

Ecuación 11

ajustado para predecir el valor de salida deseado $d(k)$, las derivadas que han de ser calculadas para llevar a cabo el análisis estadístico de sensibilidades son:

$$\frac{\partial y(k)}{\partial d(k-1)} \bigg|_{\substack{d(k-2), \dots, d(k-5) \\ e(k-1), \dots, e(k-5)}} \dots, \frac{\partial y(k)}{\partial d(k-5)} \bigg|_{\substack{d(k-1), \dots, d(k-4) \\ e(k-1), \dots, e(k-5)}}$$

$$\frac{\partial y(k)}{\partial e(k-1)} \bigg|_{\substack{d(k-1), \dots, d(k-5) \\ e(k-2), \dots, e(k-5)}} \dots, \frac{\partial y(k)}{\partial e(k-5)} \bigg|_{\substack{d(k-1), \dots, d(k-5) \\ e(k-2), \dots, e(k-4)}}$$

Ecuación 12

que corresponden a las derivadas del aproximador funcional f_w respecto a cada uno de sus argumentos, como en el caso no recurrente.

Los ejemplos siguientes ilustran la capacidad de aproximación de la estructura NARMAX y la aplicación del análisis estadístico de sensibilidades a esta arquitectura.

6. Ejemplos de aplicación

6.1. Ejemplo nº 1

Este primer ejemplo es una serie temporal lineal ARMA dada por la expresión:

$$y(t) = 0.5 y(t-3) - 0.4 y(t-4) + 0.6 \epsilon(t-1) - 0.3 \epsilon(t-2) + \epsilon(t)$$

Ecuación 13

donde $\epsilon(t)$ es un proceso de ruido blanco estacionario distribuido según una $N(0,1)$.

Se generó una serie $d(t)$ de 5000 muestras utilizando la Ecuación 13. Las 3000 primeras fueron utilizadas como conjunto de entrenamiento, las 1000 siguientes como conjunto de test, y las 1000 últimas como conjunto de validación. Se entrenó un modelo NARMAX como la descrita por la Ecuación 11 (5 entradas autoregresivas y 5 entradas de media móvil) utilizando como aproximador funcional una red RBFN con 10 entradas y 5 unidades ocultas.

El primer objetivo de este experimento es evaluar la capacidad de representación de este modelo no lineal en un caso lineal. La mejor estimación de $d(t)$ que se puede obtener es aquella que produce como error de estimación la componente impredecible $\epsilon(t)$. Un buen método para evaluar la calidad de la estimación obtenida es pues comparar el error de estimación del modelo NARMAX, $e(t)$, con el ruido original de la serie, $\epsilon(t)$. El segundo objetivo del experimento es evaluar la utilidad del análisis estadístico de sensibilidades como herramienta de selección de variables de entrada. En este caso este análisis debiera identificar como variables irrelevantes las entradas autorregresivas $y(t-1)$, $y(t-2)$ e $y(t-5)$ y las de media móvil $e(t-3)$, $e(t-4)$ y $e(t-5)$. La Figura 2 recoge gráficamente los resultados así obtenidos.

La Figura 2.a es una muestra de la estimación obtenida sobre el conjunto de test. El error de estimación correspondiente queda reflejado en la Figura 2.b, donde se ha superpuesto el ruido original de la serie ($\epsilon(t)$). El modelo NARMAX ha sido capaz de reproducir el ruido original de la serie, probando de este modo sus capacidades de aproximación en este ejemplo lineal.

La Figura 2.c y la Figura 2.d muestran las distribuciones de las sensibilidades de la salida frente a variaciones en cada una de las entradas (AR y MA respectivamente). Las distribuciones de las derivadas asociadas a variables de entrada irrelevantes ($y(t-1)$, $y(t-2)$, $y(t-5)$, $e(t-3)$, $e(t-4)$, y $e(t-5)$ en este caso) están centradas en cero y tienen una desviación típica muy reducida (ver Figura 2.f). Por otro lado, las distribuciones asociadas a

variables de entrada influyentes en la salida del proceso han quedado centradas sobre las derivadas del modelo original (que coinciden con los coeficientes de regresión), transformadas por los módulos de estandarización. Esta discrepancia entre las distribuciones de las sensibilidades de la salida del proceso, frente a variaciones en entradas irrelevantes e influyentes, facilita la identificación del subconjunto de variables de entrada irrelevantes, mediante el análisis de los indicadores mostrados en la Figura 2.e y Figura 2.f.

6.2. Ejemplo nº 2

La serie temporal:

$$y(t) = \beta \epsilon(t-1) \times \epsilon(t-2) + \epsilon(t)$$

Ecuación 14

analizada por Priestley ([Priestley, 1978]) es un buen ejemplo para ilustrar las capacidades no lineales del modelo NARMAX con redes RBFN.

Si $\epsilon(t)$ es un ruido blanco estacionario con media nula, entonces las autocorrelaciones $E[y(t)y(t-k)]$ son nulas para todo valor de k no nulo. Esto significa que, desde un punto de vista de análisis lineal, el proceso $y(t)$ es un proceso de ruido blanco y que por lo tanto la mejor estimación es $y(t)=0$ para todo t .

Se ha aplicado el mismo procedimiento que en el ejemplo anterior a una serie de datos generada mediante la Ecuación 14, con $\beta=0.7$. El modelo NARMAX ajustado contiene una RBFN como aproximador funcional, con 5 entradas AR, 5 entradas MA y 5 unidades ocultas. Los resultados obtenidos se muestran en la Figura 3.

Es importante resaltar cómo esta estructura ha sido capaz de extraer toda la información contenida en valores pasados de la serie, reproduciendo el ruido de la serie original (Figura 3.b), y llevando a cabo la transformación no lineal necesaria para obtener una estimación de la componente impredecible de la salida (Figura 3).

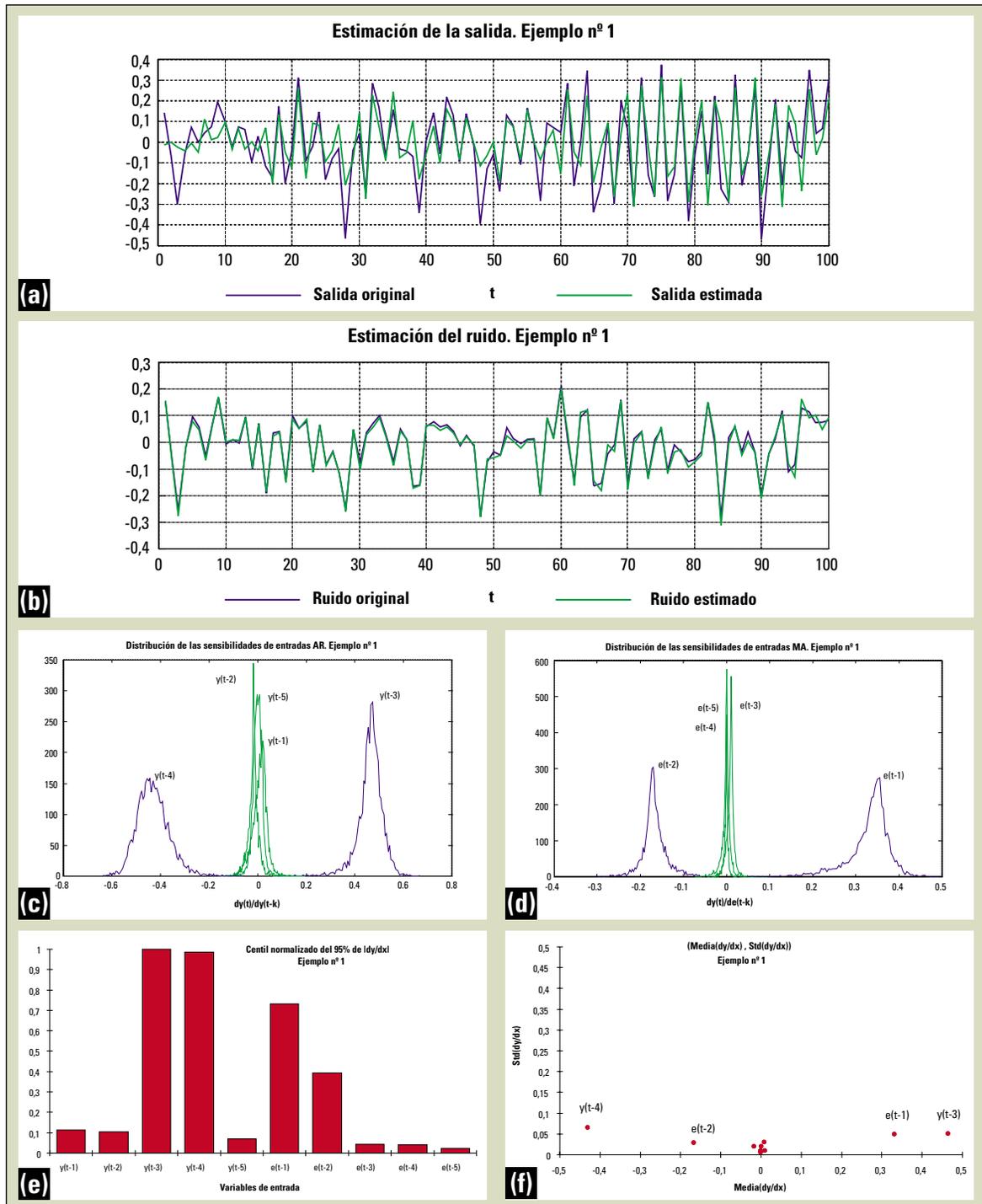


Figura 2: Aplicación del modelo NARMAX con RBF(5 entradas AR, 5 entradas MA, 5 unidades ocultas) a la serie temporal:
 $y(t) = 0.5 y(t-3) - 0.4 y(t-4) + 0.6 \epsilon(t-1) - 0.3 \epsilon(t-2) + \epsilon(t)$ con $\epsilon(t) \sim N(0,0.1)$

(a) Estimación de la salida (b) Estimación del ruido (c) Distribución de sensibilidades de entradas AR (d) Distribución de sensibilidades de entradas MA (e) Centil del 95% normalizado del valor absoluto de las sensibilidades (f) Localización de las sensibilidades en el plano (Media, Desviación Típica)

La Figura 3.e y la Figura 3.f identifican claramente el subconjunto de variables de entrada del aproximador que no influyen en la salida del mismo. Las sensibilidades asociadas con las entradas AR y con los últimos retardos de los errores han quedado distribuidas en un estrecho intervalo centrado en el origen (ver Figura 3.c y Figura 3.d), quedando caracterizadas por valores reducidos de media, varianza y centil del 95%.

7. Aplicación a la predicción de la demanda de energía eléctrica

Para ilustrar la aplicación del modelo y de la metodología de selección de variables propuestos a un caso real de predicción en el ámbito industrial, se presenta a continuación la predicción de la demanda total de energía eléctrica en Francia. La predicción de la demanda de energía eléctrica juega un papel fundamental en la explotación de los sistemas de energía eléctrica, ya que permite programar adecuadamente la estrategia de generación.

La demanda de energía eléctrica es una variable que presenta periodicidad diaria y semanal y que se ve afectada de forma significativa por las condiciones meteorológicas. Por estas razones, además de la propia serie de demanda de energía eléctrica (P) medida cada media hora, se utilizarán dos series adicionales de datos como variables explicativas: la temperatura media (T) y la nebulosidad media (N). La relación entre estas variables explicativas y la demanda de energía no es una relación de tipo lineal, lo que justifica la utilización de un modelo no lineal como el NARMAX.

El período de muestreo de las variables meteorológicas es de tres horas, por lo que será necesario realizar una interpolación para sincronizarlas con la serie de la demanda. La base de datos cubre un período de tres años, desde comienzos del 1990 hasta finales de 1992. El primer año será utilizado como conjunto de entrenamiento, el siguiente como conjunto de test, y el último para la validación.

El procedimiento de ajuste del modelo de predicción consta básicamente de tres pasos: selección de las variables explicativas, selección de la estructura del modelo, ajuste de sus parámetros y validación del modelo finalmente ajustado. En nuestro caso el modelo utilizado será el modelo NARMAX utilizando una red neuronal RBFN como aproximador funcional. La selección de variables deberá proporcionar los retardos de las componentes autorregresivas, de media móvil y exógenas que realmente son relevantes para la predicción.

La selección de variables se realizará en dos pasos, analizando en primer lugar y de forma conjunta las componentes autorregresivas y exógenas, y en segundo lugar las componentes de media móvil.

7.1. Análisis de las componentes autorregresivas y exógenas

Para realizar la selección de variables del modelo NARMAX se propone un procedimiento iterativo en el que en cada iteración se aplica el análisis estadístico de sensibilidades a un modelo ajustado con las variables explicativas (retardos de P, T y N) seleccionadas en la iteración anterior, junto con un nuevo grupo de variables explicativas compuesto por los siguientes V retardos de P, T y N (el entero V recibe el nombre de ancho de ven-

SERIES TEMPORALES	RETARDOS DE LAS VARIABLES EXPLICATIVAS DE P(t) QUE SE AÑADEN A LOS YA SELECCIONADOS EN CADA ITERACIÓN		
	1ª Iteración	2ª Iteración	3ª Iteración
Demanda	t-1, t-2, ..., t-10	t-11, t-12, ..., t-20	t-21, t-22, ..., t-30
Temperatura (T)	t-1, t-2, ..., t-10	t-11, t-12, ..., t-20	t-21, t-22, ..., t-30
Nebulosidad (N)	t-1, t-2, ..., t-10	t-11, t-12, ..., t-20	t-21, t-22, ..., t-30

Tabla I: Procedimiento iterativo de selección de variables (V=10).

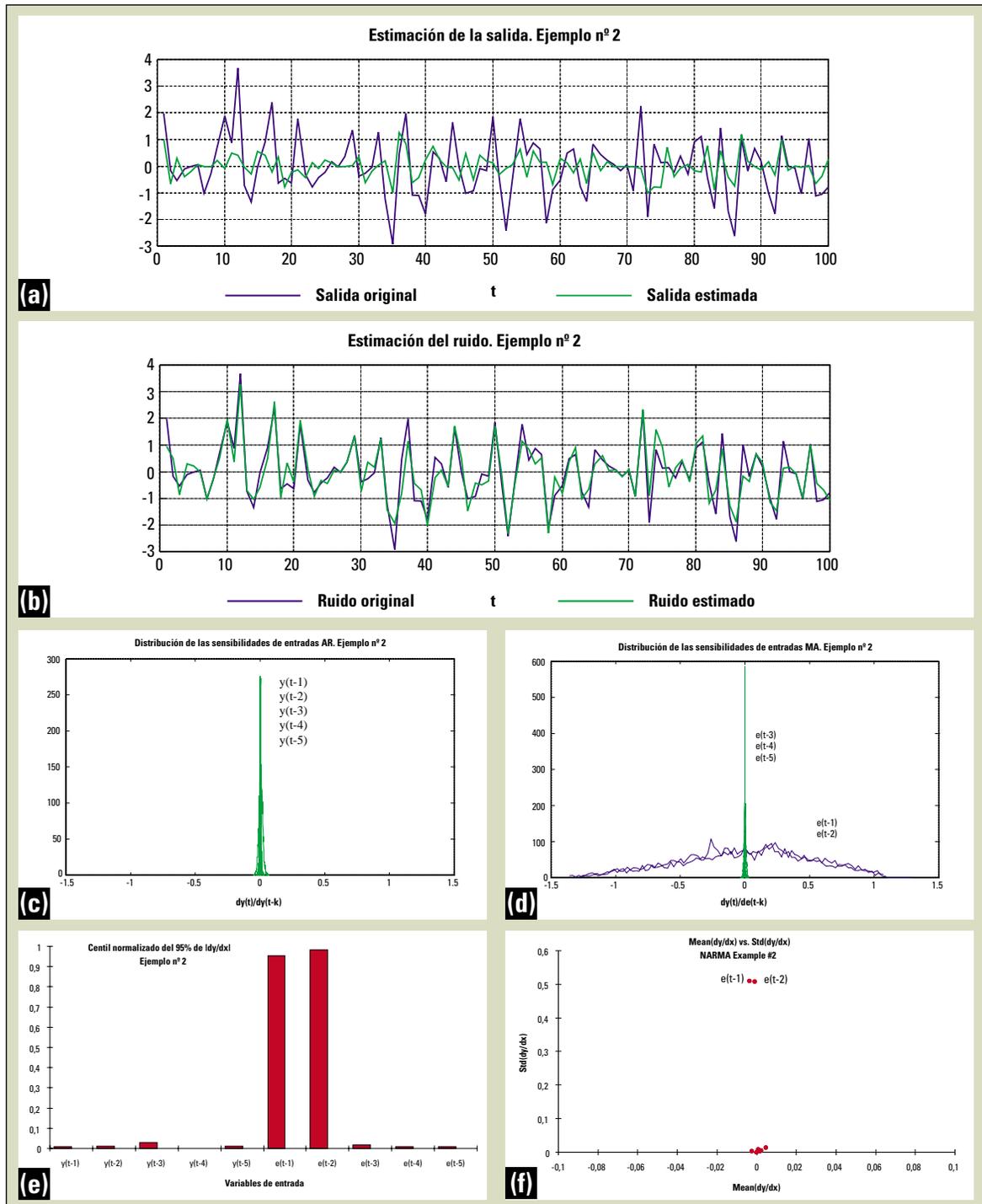


Figura 3: Aplicación del modelo NARMAX con RBF(5 entradas AR, 5 entradas MA, 5 unidades ocultas) a la serie temporal: $y(t) = 0.7 \epsilon(t-1) \epsilon(t-2) + \epsilon(t)$ con $\epsilon(t) \sim N(0,1)$

(a) Estimación de la salida (b) Estimación del ruido (c) Distribución de sensibilidades de entradas AR (d) Distribución de sensibilidades de entradas MA (e) Centil del 95% normalizado del valor absoluto de las sensibilidades (f) Localización de las sensibilidades en el plano (Media, Desviación Típica)

SERIES TEMPORALES	RETARDOS DE LAS COMPONENTES AUTORREGRESIVAS Y EXÓGENAS SELECCIONADOS PARA EXPLICAR P(T)
Demanda (P)	t-1, t-47, t-48, t-49, t-335, t-336, t-337
Temperatura (T)	t, t-30, t-48, t-200, t-336
Nebulosidad (N)	t

Tabla 2: Componentes autorregresivas y exógenas seleccionadas mediante AES

tana). La Tabla 1 ilustra este procedimiento de selección de variables que se repite hasta no observar una mejora significativa del modelo.

Los retardos de la demanda, temperatura y nebulosidad seleccionados como variables explicativas de la demanda en el instante t quedan recogidos en la Tabla 2.

Todas las variables explicativas que han sido seleccionadas pueden ser justificadas por las siguientes consideraciones :

- las componentes autorregresivas seleccionadas corresponden a la demanda inmediatamente anterior (t-1) y a las demandas del día y de la semana anterior (t-47, t-48, t-49 y t-335, t-336, t-337). Esta selección confirma el carácter autorregresivo de la serie y su periodicidad diaria y semanal
- las medidas de temperatura utilizadas por el modelo corresponden de forma similar a la temperatura esperada en el mismo instante de la predicción (t) y a las temperaturas observadas el día y la semana anterior (t-48 y t-336). A estos retardos se suman las temperaturas en t-30 y t-200. Estos dos últimos retardos son también utilizados por EDF para modelar el efecto de la temperatura en la demanda
- en el caso de la nebulosidad, el modelo sólo utiliza el valor esperado en el mismo instante (t) de la predicción, sin requerir la información de valores observados en días anteriores

Una vez identificadas las componentes autorregresivas y exógenas del modelo, queda por analizar sus componentes de media móvil.

7.2. Análisis de las componentes de media móvil

El procedimiento de identificación de las componentes de media móvil del modelo es análogo al descrito en el apartado anterior, de tal forma que en cada iteración se aplica el AES al modelo ajustado con el conjunto de variables explicativas seleccionadas en el paso anterior y una nueva ventana de retardos de los errores de estimación cometidos en el pasado (componentes MA). El punto de partida de este análisis será el conjunto de variables autorregresivas y exógenas seleccionadas anteriormente.

Para evaluar la mejora del modelo durante el procedimiento de identificación de sus componentes de media móvil se utilizará el error porcentual medio (MAPE) con horizonte de predicción h definido por:

$$MAPE = \frac{1}{N} \sum_{t=1}^N \sum_{i=1}^h \frac{|\hat{y}_{t+i|t} - y_t|}{y_t}$$

Ecuación 15

Como modelo de referencia se tomará el modelo ajustado con las variables autorregresivas y exógenas seleccionadas en el apartado anterior, que tiene un MAPE del 6%.

La Tabla 3 ilustra el procedimiento de selección de las componentes de media móvil (con V=48), presentando la evolución del MAPE con un horizonte de predicción de 48 muestras (correspondiente a 24 horas) y los retardos de MA seleccionados.

El modelo finalmente ajustado tiene un MAPE del 1.8%, frente al MAPE del 6% del modelo au-

ITERACIÓN	COMPONENTES DE MEDIA MÓVIL SELECCIONADAS MEDIANTE AES	MAPE
1	t-1	4%
2	t-1, t-48	3%
3	t-1, t-48, t-96	2,9%
4	t-1, t-48, t-96, t-144	2,8%
5	t-1, t-48, t-96, t-144, t-192	2,8%
6	t-1, t-48, t-96, t-144, t-192, t-240	2,9%
7	t-1, t-48, t-96, t-144, t-192, t-240, t-288	2,7%
8	t-1, t-48, t-96, t-144, t-192, t-240, t-288, t-336	1,8%

Tabla 3 : Evolución del MAPE ("Mean Average Percentage Error") durante el procedimiento de identificación de las componentes de media móvil

torregresivo obtenido en el apartado anterior. La componente MA del modelo introduce una mejora significativa debido al efecto estabilizador de la realimentación. En la Figura 4 se representa un ejemplo de predicción de la curva de demanda diaria con el modelo finalmente ajustado. Esta

aplicación de la metodología propuesta a un caso real de predicción ha puesto de manifiesto la capacidad de representación del modelo NARMAX, y la utilidad del Análisis Estadístico de Sensibilidades como herramienta de selección de variables explicativas para la identificación del mismo.

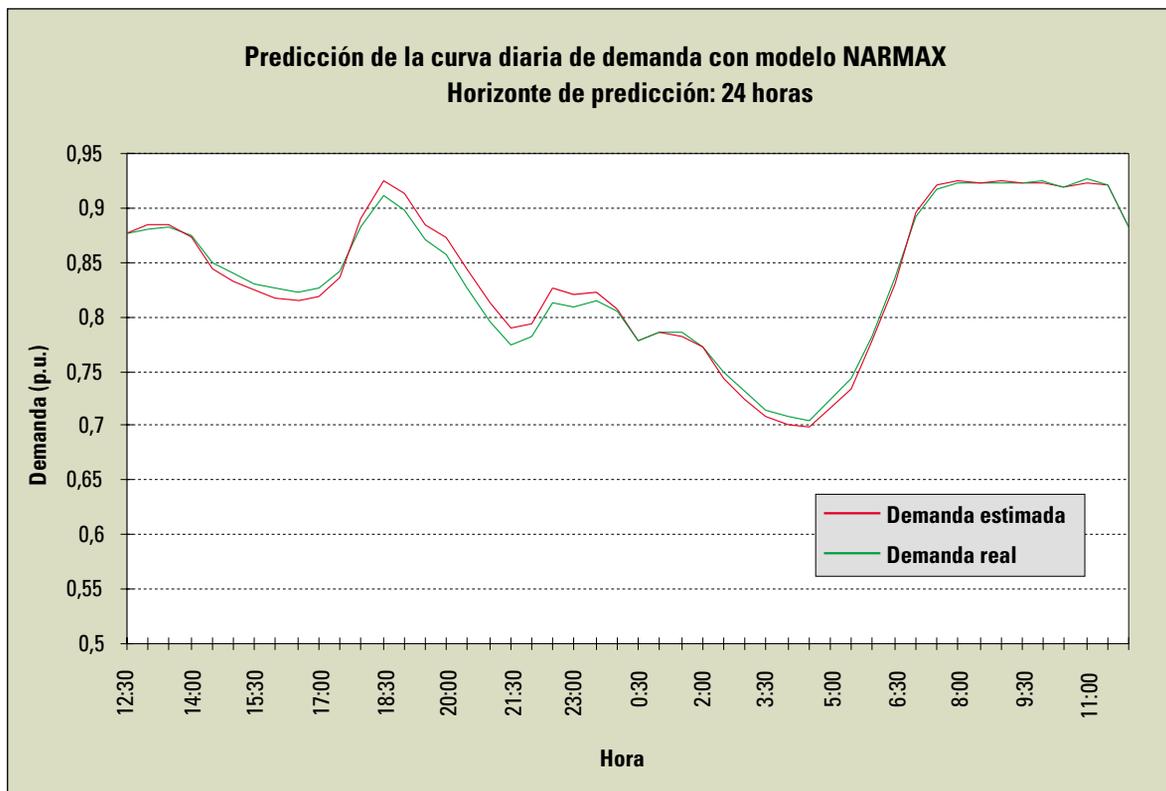


Figura 4 : Ejemplo de predicción de demanda con modelo NARMAX y horizonte de 24 horas

8. Conclusiones

En este artículo se ha presentado la estructura y los métodos de ajuste del modelo NARMAX, especialmente indicado para la predicción de series temporales no lineales. Este modelo es una extensión del modelo lineal autorregresivo de media móvil (ARMA), en el que se ha sustituido la transformación lineal de los regresores por un aproximador funcional no lineal.

Asimismo se ha propuesto un método de selección de las variables de entrada al aproximador funcional (regresores), basado en el análisis estadístico de las sensibilidades del modelo. Este método permite identificar los órdenes autorregresivos (AR) y de media móvil (MA) del proceso que se pretende modelar, de tal forma que la eliminación de entradas irrelevantes del modelo simplifica su estructura y mejora sus prestaciones.

Finalmente se han presentado tres ejemplos de predicción de series temporales, que ilustran las capacidades del modelo propuesto y la aplicación de las técnicas desarrolladas. Estos ejemplos muestran cómo el modelo es capaz de extraer el ruido de la serie original, y cómo el análisis estadístico de sus sensibilidades identifica los retardos de las distintas variables que realmente influyen en la salida del modelo.

El último ejemplo presentado corresponde a un caso real de predicción de la curva de demanda de energía eléctrica. Esta aplicación es tan sólo un caso más dentro del amplio abanico de situaciones en las que la disponibilidad de un buen sistema de predicción permite optimizar la planificación y el control de los sistemas de producción.

La creciente complejidad de los procesos industriales ha generalizado la instalación de sistemas de seguimiento en continuo dando lugar a la necesidad de desarrollar sistemas automáticos de tratamiento de los enormes volúmenes de información que se generan. Estos sistemas ([Muñoz et al., 1995],[Sanz et al., 1996]) requieren nuevas herramientas de modelado y/o predicción capaces de adaptarse a la continua evolución de los procesos subyacentes, por lo que están siendo en gran medida los responsables de los avances realizados en área de identificación de sistemas no lineales con Redes Neuronales Artificiales. 

Bibliografía

[Box & Jenkins, 1976] "Time series analysis: forecasting and control". George E.P. Box, Gwilym M. Jenkins. Holden-Day Inc., CA, (Revised Edition) 1976.

[Czernichow, 1996] "Apport des réseaux récurrents à la prévision de séries temporelles, application à la prévision de consommation d'électricité". Thomas Czernichow. Thèse de Doctorat de L'Université Pierre et Marie Curie. Paris 1996.

[Kaufman & Rousseeuw, 1990] "Finding groups in data: An introduction to cluster analysis". L.Kaufman, P.J.Rousseeuw. John Wiley & Sons, Inc. 1990.

[Luenberger, 1984] "Linear and Nonlinear Programming". David G. Luenberger. Addison-Wesley, 2ª Edición, 1984.

[Moody & Darken, 1988] "Learning with localized receptive fields". J. Moody, J.Darken. Proc. of the 1988 Connectionist Models Summer School, (G.Hinton, T. Sejnowski, and D.Touretzky, eds.), pp.133-143, Palo Alto, CA, 1988.

[Moody & Darken, 1989] "Fast learning in networks of locally-tuned processing units". J. Moody, J.Darken. Neural Computation 1, pp.281-294, 1989.

[Muñoz et al., 1995] "Neural network approach to the diagnosis of the boiler combustion in a coal power plant". Antonio Muñoz, José Villar, Miguel Angel Sanz Bobi, Agustín Gimeno, Luis Zarauza. International Conference on Artificial Neural Networks. ICANN'95, Industrial Conference, Paris, 1995.

[Muñoz, 1996] "Aplicación de técnicas de redes neuronales artificiales al diagnóstico de procesos industriales". Antonio Muñoz. Tesis Doctoral. Escuela Técnica Superior de Ingenieros Industriales. Universidad Pontificia Comillas. Madrid 1996.

[Priestley, 1978] "Nonlinear Models in Time Series Analysis". M.B.Priestley. The Statistician, vol.27, pp.159-176, 1978.

[Sanz et al., 1996] "Condition monitoring of water chemical properties in the cycle of a coal power plant usin neural networks". Miguel Ángel Sanz Bobi, Antonio Muñoz, José Álvarez. Power- Gen Europe, Budapest, 1996.

[Wei, 1990] "Time series analysis: univariate and multivariate methods". William W. S. Wei. Addison-Wesley, 1990.