

ESTIMACIÓN DE LA DEMANDA DE AGUA PARA RIEGO: REGRESIONES LINEALES VERSUS APROXIMACIONES NEURONALES

Inmaculada Pulido-Calvo¹, Pilar Montesinos², José Roldán² y Francisco J. Ruiz-Navarro²

1. Departamento de Ciencias Agroforestales, Universidad de Huelva, España.

2. Departamento de Agronomía, Universidad de Córdoba, España.

ipulido@uhu.es

Recibido para evaluación: 23 de Mayo de 2005 / Aceptación: 25 de Agosto de 2005 / Recibida versión final: 30 de Agosto de 2005

RESUMEN

La demanda de agua constituye una pieza básica de información para diseñar políticas que afecten al consumo del agua en los regadíos. También es la magnitud de referencia en el diseño, modernización y explotación de los sistemas de distribución. En este trabajo, se propone caracterizar la demanda de agua de riego mediante la aplicación de regresiones lineales y de Redes Neuronales Computacionales (RNCs). Se dispone de datos de consumo diario de agua de la zona regable del Genil-Cabra (Córdoba, España) registrados por un sistema automático de telemetría, durante las campañas de riegos 2001/02 y 2002/03. Los modelos se establecen considerando la relación de los datos presentes y pasados de la demanda, aunque también se analiza la influencia de datos climáticos. La utilización de regresiones múltiples y RNCs ha proporcionado una capacidad aceptable de estimación de la demanda de agua en zonas regables cuando se hace un filtrado previo de las series de datos registradas. Los mejores resultados se obtienen cuando se consideran como variables de entrada o independientes las demandas de los dos días anteriores al de estimación.

PALABRAS CLAVES: Demanda de Agua, Zonas Regables, Redes Neuronales, Regresión Múltiple, España .

ABSTRACT

Water demand is a basic information for water consumption management in irrigation districts. Forecasting water demand is one of the main problems for designers and managers of water distribution systems. This paper examines methodologies for irrigation demand modelling. Approaches based on linear multiple regressions and computational neural networks (CNNs) are developed. The models are established using actual data recorded from an irrigation water distribution system in southern Spain during two irrigation seasons (2001/02, 2002/03). The input variables used in CNN and multiple regression models are: water demands from previous days; and water demands and climatic data simultaneously. Good results were obtained when original data were modified to reduce the noise by a smoothing process. The best results were obtained when water demand recorded during the two previous days were used as input data.

KEY WORDS: Water Demand, Irrigation Districts, Neural Networks, Multiple Regression Models, Spain.

1. INTRODUCCIÓN

El recurso agua está sometido a la presión de una demanda cada vez más exigente en cantidad y calidad, condicionada por aspectos sociales, políticos y ambientales. La creciente dificultad para garantizar la satisfacción de las demandas se traduce, en una mayor competencia entre los consumidores tradicionales de agua –agricultura, industria y ciudades- por los recursos hídricos disponibles. Esta competencia ya está restringiendo las actividades de desarrollo de muchos países. Por otro lado, el aumento de las necesidades hídricas sin que los recursos disponibles se incrementen, al menos simultáneamente, está originando una mayor competencia entre regiones o países por el acceso a los mismos (FAO, 1993; Ohlsson, 1995; Sumpsi *et al.*, 1998).

Además, la actual preocupación por el deterioro ambiental ha creado un nuevo factor que afecta a la competencia por el agua. Ciertos usos no consuntivos como los recreativos, ecológicos y paisajísticos están siendo tomados en consideración a la hora de asignar agua para usos consuntivos. No solamente aumenta la competencia por la cantidad de agua, sino también por la calidad.

Como consecuencia de los distintos factores que influyen en la competencia por los recursos hídricos, el agua está considerándose cada vez más como un recurso escaso y valioso que demanda rigor en su manejo y cuidado extremo para evitar su dilapidación. Una de las claves para la resolución de estos problemas está en el sector agrícola, ya que conviene tener en cuenta que la agricultura de regadío es el sector más demandante de agua a escala mundial pues supone un 87% de los usos consuntivos (ONU, 1997; Sumpsi *et al.*, 1998).

El objetivo de este trabajo es analizar la demanda de agua para regadío, ya que disponiendo de ésta se podrá examinar con rigor las posibilidades que hay para modificarla, racionándola al objeto de lograr una política de gestión eficiente. La demanda de agua de las zonas regables constituye una pieza básica de información para diseñar políticas que afecten al consumo de agua en los regadíos. Puede aportarnos información sobre el valor marginal del agua y el grado de respuesta a distintas tarifas del agua de riego. También es la magnitud de referencia en el diseño, modernización y explotación de los sistemas de distribución.

Uno de los principales problemas en la gestión del recurso agua en zonas regables es la falta de información fiable y comparable sobre sus consumos. Por esto, con este trabajo se ha implantado una red de observación, como condición indispensable para la gestión global e integrada de los recursos hídricos en zonas regables. Esta red consta de un sistema de telemetría con el correspondiente del software para su control y la gestión y tratamiento de la información adquirida, lo que permite la monitorización de sistemas de distribución de agua para riego. En concreto, se han registrado los volúmenes de agua consumidos cada hora en distintas parcelas ubicadas en diversos sectores de la zona regable del Genil-Cabra en la provincia de Córdoba (España), durante las campañas de riego 2001/02 y 2002/03.

En este trabajo se caracteriza la demanda de agua de riego mediante la aplicación de regresiones lineales, métodos tradicionales de predicción, y de Redes Neuronales Computacionales (RNCs), modelos heurísticos incluidos en el amplio campo de conocimiento denominado Inteligencia Artificial. Algunas aplicaciones de RNCs en el manejo y la gestión de los recursos hídricos incluyen la caracterización del proceso lluvia-escurrimiento (Hsu *et al.*, 1995; Lorrai y Sechi, 1995; Mason *et al.*, 1996; Abrahart *et al.*, 1999; Tokar y Johnson, 1999; Thirumalaiah y Deo, 2000; Tokar y Markus, 2000; Chiang *et al.*, 2004; Moradkhani *et al.*, 2004; Anttil y Rat, 2005), estimación a corto plazo del estado de ríos (Thirumalaiah y Deo, 1998, 2000; Abrahart and See, 2000, 2002; See y Openshaw, 2000; Cameron *et al.*, 2002), predicción de lluvias (French *et al.*, 1992; Zhang *et al.*, 1997; Kuligowski y Barros, 1998), modelación de aguas subterráneas (Roger y Dowla, 1994; Yang *et al.*, 1997), análisis de sequías (Shin y Salas, 2000), etc.

Trabajos previos sobre la predicción de la demanda de agua tanto en abastecimientos urbanos como en zonas regables (Griñó, 1992; Pulido-Calvo *et al.*, 2002, 2003) muestran que la utilización de RNCs proporciona resultados muy satisfactorios e incluso superiores a los métodos clásicos de regresión.

2. MATERIAL Y MÉTODOS

2.1. Área de estudio

Con objeto de contrastar la metodología desarrollada en este trabajo, ésta se ha aplicado a la zona regable del

Genil-Cabra, con aproximadamente unas 15.000 ha regables en la actualidad y pertenecientes a la cuenca del Guadalquivir en las provincias de Córdoba y Sevilla (España) (Figura 1). El agua para el regadío de la zona proviene del pantano de Iznájar, cuya capacidad máxima de almacenamiento es de 978 hm³. A través del río Genil el agua llega hasta el embalse de Cordobilla, con una capacidad de 34 hm³, donde se encuentra la estación principal de bombeo que eleva el agua hasta el canal principal, que es el que la distribuye hasta cada una de

las estaciones de bombeo sectoriales o de puesta en carga de la red principal de tuberías. Esta última, puede ser de presión o de gravedad en función de la altura de los puntos donde se pretende llevar el agua. Seguidamente, el agua llega hasta las agrupaciones en donde la red secundaria se encarga de distribuirla a una o varias parcelas. Este sistema de distribución de agua permite a los agricultores disponer de agua para riego durante todo el año las 24 horas del día, siempre que no sobrepasen un valor prefijado de caudal máximo instantáneo.

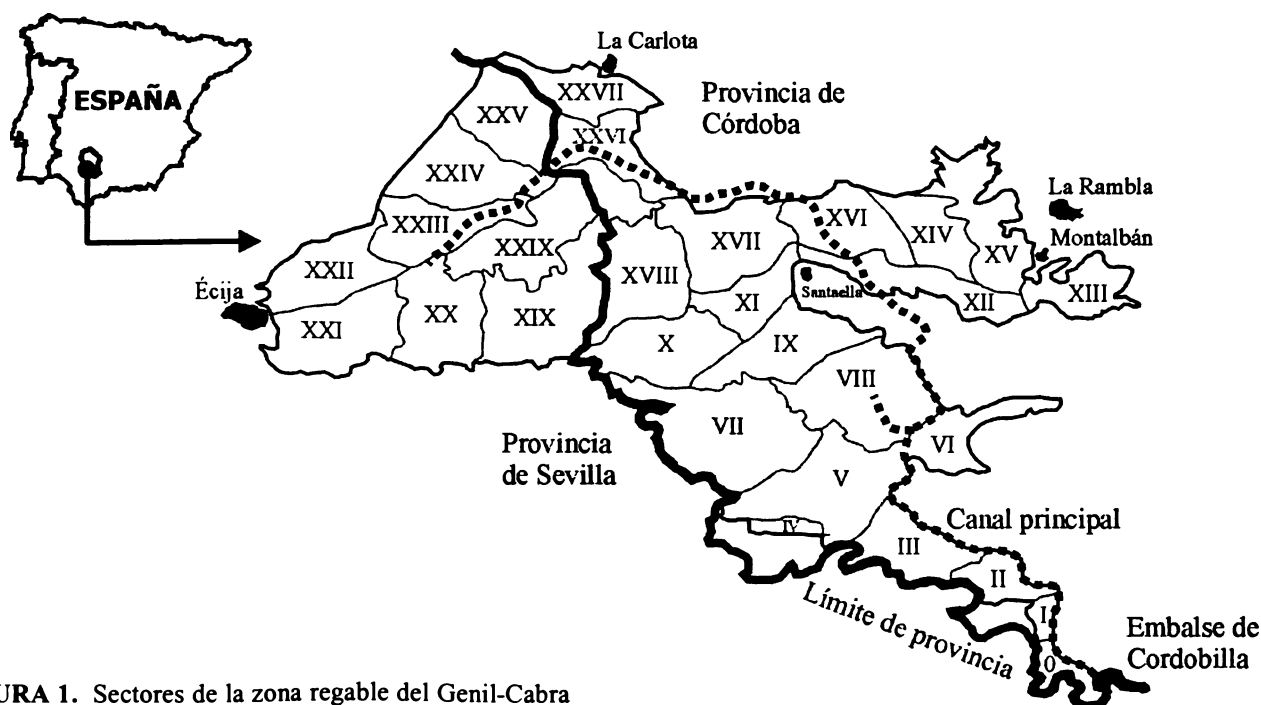


FIGURA 1. Sectores de la zona regable del Genil-Cabra

Los cultivos más abundantes son el olivar, el algodón y los cereales de invierno. Cultivos como el girasol que en el pasado era mayoritario ocupando durante la campaña 1992/93 el 45% de la superficie, actualmente lo han dejado de ser debido a su situación de mercado y a la política agraria europea, ocupando durante la campaña 2002/03 menos del 5%. Frente al girasol, el maíz es el cultivo que está tomando un mayor auge ocupando durante la campaña 2002/03 casi el 10% de la superficie, mientras que en la campaña 1992/93 no alcanzaba ni el 0,5%. Con respecto al olivar, durante la campaña 1992/93 ocupaba casi el 16% de la superficie y hoy es el cultivo que más porcentaje de superficie ocupa con un 34%. Se ha instalado, en 28 tomas de parcela de los sectores II al VII de la red de riego, un sistema comercial de telemetría de volúmenes y presiones en tiempo real y

transmisión de datos vía cable y radio denominado SIGA para monitorizar el funcionamiento de las tomas de agua (ver apartado II.2). La instalación se realizó en dos etapas: en 16 parcelas durante la campaña de riegos 2001/02 y en 12 más en la campaña de riegos 2002/03. Estas parcelas han sido elegidas por su representatividad en cuanto a distribución de cultivos (14 de olivar y 14 dedicadas a cultivos herbáceos), de tipo de suelos (villar y vertisol) y de métodos de riego (aspersión y goteo).

2.2. Sistema de Telemetría

Para conocer los consumos reales de agua en la zona regable se ha instalado un sistema de telecontrol en las 28 tomas seleccionadas. Se trata del sistema conocido comercialmente como SIGA (Sistema Integral del Gestión

del Agua) desarrollado por la Empresa Pública Transformaciones Agrarias S.A. (TRAGSA) y que consiste en un sistema automático que permite registrar, almacenar, enviar, procesar y visualizar de forma ordenada la información captada por los sensores instalados en las tomas de parcela. En este caso, la información que se registra son volúmenes de agua y presiones en el momento de la lectura, así como el tiempo. La estructura arborescente del sistema de captación de datos permite el control de puntos de medida distribuidos irregularmente en una superficie superior a 8.000 ha, la cual se remite a un centro de control, donde mediante el software comercial SCADA se realiza la adquisición, almacenamiento y visualización de los datos.

Los elementos que componen el sistema son (Figura 2):

El *nodo de riego* es el elemento encargado de recoger las señales de los contadores volumétricos y medidores de presión instalados en las tomas de riego. Estos datos se transmiten por radio a las estaciones concentradoras mediante el protocolo de comunicación *LonWorks*. El suministro de energía se realiza mediante un pequeño panel solar de alto rendimiento y además dispone del apoyo de una batería de 12 V, que proporciona una autonomía de 6 días.

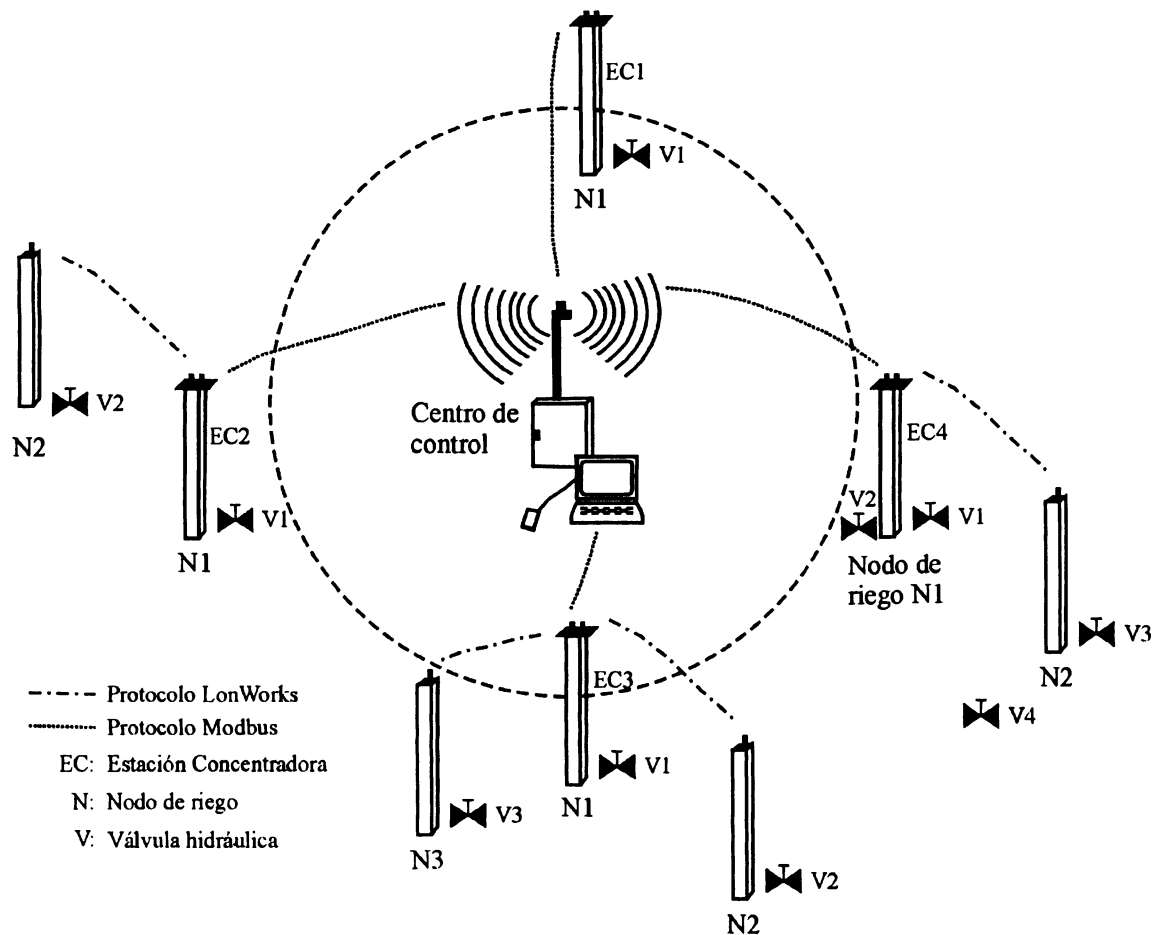


FIGURA 2. Sistema de telemetría y SCADA de toma de datos.

La *estación concentradora* es el elemento del sistema responsable de la adquisición de señales y de la ejecución de la lógica de control. Almacena las señales de campo que le llegan desde los nodos de riego con el fin de transferirlas posteriormente al centro de control. La estación concentradora, es un nodo de riego con capacidad de controlar varios nodos de riego y transmitir la información recogida a mayor distancia. Requiere dos antenas de onda media para comunicar con el centro de control y con los nodos de riego asociados, respectivamente.

El *centro de control* tiene un ordenador con el software de control *Scada Factory Link*, para visualizar y almacenar la información registrada en las estaciones concentradoras y sus nodos asociados. El centro de control se comunica con las estaciones concentradoras a través de un protocolo *Modbus* por medio de un radiomodem con un alcance de hasta 40 km.

La frecuencia de lectura de los datos se ha fijado en 1 hora, ya que 24 registros diarios de cada hidrante permiten identificar perfectamente inicio, duración del riego, volumen consumido y usos del agricultor. El sistema permite modificar la frecuencia de lecturas según las necesidades del usuario.

De este modo, se dispone de datos de la demanda horaria de agua (que se transformará en demanda diaria) en las parcelas seleccionadas, junto con datos climáticos diarios (temperatura máxima, temperatura media, temperatura mínima, precipitación, humedad relativa y velocidad del viento) registrados en la estación meteorológica de Santaella (Córdoba) situada en la zona regable, durante las campañas de riegos 2001/02 y 2002/03, a partir de los cuales se reproducirá el consumo diario de agua de riego en parcela.

2.3. Predicción de la demanda diaria de agua mediante la aplicación de Redes Neuronales Computacionales (RNCs)

Con la información disponible se ha procedido a la caracterización de la demanda de agua en la zona regable mediante el uso de Redes Neuronales Artificiales o Computacionales. Una red neuronal computacional (RNC) es un modelo heurístico que copia la estructura de un sistema neuronal biológico con el fin de alcanzar una funcionalidad similar. El procesador elemental o

neurona es un dispositivo simple de cálculo que, a partir de un vector de entrada del exterior o de otras neuronas, proporciona una única respuesta o salida. En general, una red neuronal consiste en el establecimiento de un conjunto de nodos o neuronas agrupadas en varias capas e interconectadas entre sí (capa de entrada, capas ocultas o intermedias y capa de salida). Las conexiones que unen a las neuronas tiene asociado un peso numérico, y el aprendizaje se realiza con la actualización de tales pesos.

Para la calibración y validación de los modelos de RNCs se ha utilizado el módulo correspondiente del software STATISTICA 6.0 (Statsoft, Inc., 1984-2002). Para las neuronas de la capa de salida se propuso como función de transferencia de la información entre neuronas la función lineal (*I*) y para cada neurona de las capas intermedias la función de transferencia logística, también conocida como sigmoide (*s*). El algoritmo de aprendizaje seleccionado para la calibración y posterior validación de los modelos es el algoritmo supervisado de segundo orden de Levenberg-Marquardt (LM) (Shepherd, 1997).

El algoritmo de aprendizaje de retropropagación estándar o *backpropagation* (Rumelhart *et al.*, 1986) es el algoritmo supervisado más empleado en sistemas neuronales y su funcionamiento se basa en modificar los valores de los pesos proporcionalmente al gradiente de la función de error con objeto de alcanzar un mínimo local (algoritmo de descenso por el gradiente). Ya que la velocidad de convergencia de este algoritmo es lenta, se han propuesto pasos adicionales para acelerar esta convergencia haciendo uso de información de segundo orden de la función de error, esto es, de sus derivadas segundas o, en forma matricial, de su Hessiano (algoritmos de segundo orden). En el caso particular de una función de error a minimizar cuadrática, el Hessiano puede ser aproximado usando sólo las derivadas primeras de las salidas del sistema neuronal, como hace el algoritmo Gauss-Newton. Ya que este algoritmo puede presentar inestabilidad cuando la aproximación no es definida positiva, el algoritmo LM resuelve este problema introduciendo un término adaptativo, por lo que es recomendado como primera opción a ensayar por muchos autores (Tan y van Cauwenberghe, 1999; Martín-del-Brío y Sanz-Molina, 2001; Anctil y Rat, 2005).

El aspecto clave del algoritmo LM es que la determinación de la nueva dirección de búsqueda del mínimo de la función de error sea obtenida mediante el compromiso

entre el salto l y la variación del error con respecto a los pesos W . Por tanto, el cambio en el coeficiente de pesos se expresa como:

$$\Delta W = -(Z^T Z + \lambda I)^{-1} Z^T \varepsilon \quad (1)$$

donde ε es el vector de errores, Z es la matriz de derivadas parciales de estos errores con respecto a los pesos W e I es la matriz identidad.

Un aspecto fundamental de una red neuronal es su capacidad de generalizar a través de ejemplos. Por generalización se entiende la capacidad de la red de proporcionar una respuesta correcta ante patrones de comportamiento que no han sido empleados en su entrenamiento. De este modo, dada una arquitectura de red, ésta deberá entrenarse hasta alcanzar el punto óptimo en el que el error de generalización es mínimo. En este trabajo, se ha utilizado un aprendizaje controlado mediante el método de la validación interna (20% de los datos de calibración para evaluar el error al final de cada época) (Tsoukalas y Uhrig, 1997; Gutiérrez-Estrada *et al.*, 2004). Se denomina época al período de tiempo que comprende todas las iteraciones realizadas al presentar todos los patrones una sola vez. La actualización de los pesos se realiza al final de cada época. Los pesos que se seleccionan son los correspondientes al número de épocas que presentan el menor error en la validación interna.

Aunque una red neuronal de una capa intermedia u oculta puede representar cualquier función continua (Cybenko, 1989), en este trabajo se utilizan redes con dos capas intermedias (aproximador universal de cualquier función: Cybenko, 1989) ya que a priori no se conoce el tipo de relación entre las variables. Asimismo, puede ocurrir que con una sola capa intermedia el número de neuronas ocultas para alcanzar una cierta cota de error sea tan elevado que su aplicación resulte inabordable en la práctica. Se han probado de 2 a 18 nodos en cada una de las capas ocultas y se selecciona la que tiene el mejor funcionamiento en el período de validación.

2.4. Comparación con Análisis de Regresión Múltiple

El análisis de regresión múltiple pretende obtener una ecuación lineal que permita estimar la variable dependiente o criterio e , conocidos los valores que presentan las q variables independientes o predictivas x_1, \dots, x_q :

$$e = b_0 + b_1 x_1 + \dots + b_q x_q \quad (2)$$

donde los parámetros b_0, b_1, \dots, b_q representan las contribuciones de cada variable independiente a la predicción de la variable dependiente.

Con los análisis de regresiones múltiples y de redes neuronales se someten los datos a un proceso semejante: a) Calibración del modelo con los datos de la campaña de riegos 2001/02; b) Predicción de la variable dependiente (demanda diaria de agua) usando los datos de la campaña 2002/03 (validación del modelo), para asegurar que los resultados del modelo son generalizables a la población y que no son específicos de la muestra utilizada en la calibración.

2.5. Identificación de Modelos

Se han establecido dos tipos de modelos de redes neuronales y regresiones múltiples en función del número de variables de entrada o independientes consideradas: a) Demanda diaria de agua de varios días anteriores; b) Demanda diaria de agua y datos climáticos (precipitación, temperatura máxima, temperatura media, temperatura mínima, humedad relativa media y velocidad del viento) de varios días anteriores. En todos los casos la salida o variable dependiente es la predicción de la demanda de agua del día próximo a los empleados para ella, expresada en $m^3/ha/día$.

Estos modelos han sido aplicados considerando las parcelas dedicadas al monocultivo del olivar y, por otro lado, las parcelas dedicadas a los cultivos de algodón y maíz que se están controlando con el sistema SIGA. Los patrones utilizados en el entrenamiento o calibración de los modelos se corresponden con los datos de la campaña 2001/02, y en la generalización o validación con los de la campaña 2002/03.

La selección del número de días previos a considerar como datos de entrada se ha hecho mediante el análisis de la función de autocorrelación parcial de las series de demanda diaria de agua. Esta función muestra que la serie de demandas diarias es un proceso autoregresivo de orden dos y, por tanto, la demanda a estimar es función de sus propios valores pasados en los dos días previos. La fiabilidad de la predicción depende en gran medida de la calidad y cantidad de los datos disponibles. En el caso en estudio, se detectó en la serie de datos de los consumos diarios de agua imprecisiones y ruido como consecuencia de errores en la toma de datos. Para corregir

dentro de lo posible este problema, se ha realizado un proceso de pretratamiento de la información contenida en la serie, consistente en un filtrado o alisamiento de tipo triangular o media desplazante ponderada (Coulbeck, 1988; De Vries y Principe, 1991; Gutiérrez-Estrada *et al.*, 2004):

$$S_j = \frac{x_{j-2} + 2x_{j-1} + 3x_j + 2x_{j+1} + x_{j+2}}{9} \quad \forall j = 3 \text{ a } n-2 \quad (3)$$

donde S_j es el punto j -ésimo de la serie de datos suavizada, x_j es el punto j -ésimo en la serie de datos original y n es el número total de puntos en la serie de datos.

2.6. Medidas de Evaluación de los Modelos (fase de validación)

Se han evaluado diversas medidas de error para asegurar el funcionamiento de las redes neuronales y las regresiones múltiples durante la fase de validación y así identificar el mejor modelo (Legates y McCabe Jr., 1999; Abrahart y See, 2000). El coeficiente de determinación (R^2) describe la proporción de la varianza total en los datos observados que pueden ser explicados por el modelo. Por otro lado, medidas de varianzas son el porcentaje de error estándar de predicción (percent standard error of prediction, %SEP) (Ventura *et al.*, 1995), el coeficiente de eficiencia (E) (Nash y Sutcliffe, 1970; Kitanidis y Bras, 1980) y la varianza relativa media (average relative variance, ARV) (Griño, 1992). Estos estimadores no están influenciados por el rango de variación de sus elementos pero son empleados para ver hasta donde el modelo es capaz de explicar el total de la variación de los datos.

Asimismo, es apropiado cuantificar el error en términos de las unidades de la variable a estimar. Estas medidas, o medidas de errores absolutos, incluyen la raíz cuadrada del error cuadrático medio (square root of mean square error, RMSE), dado por:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Q_i - \hat{Q}_i)^2}{N}} \quad (4)$$

donde Q_i es el caudal observado en el tiempo t ; \hat{Q}_i es el caudal estimado en el mismo tiempo t ; y N es el número total de observaciones del conjunto de validación.

El porcentaje de error estándar de predicción, %SEP, se define como:

$$\%SEP = \frac{100}{\bar{Q}} RMSE \quad (5)$$

donde \bar{Q} es la media de los caudales observados en el conjunto de validación. El coeficiente de eficiencia E y la varianza relativa media ARV se calculan como:

$$E = 1.0 - \frac{\sum_{i=1}^N (Q_i - \hat{Q}_i)^2}{\sum_{i=1}^N (Q_i - \bar{Q})^2} \quad ARV = \frac{\sum_{i=1}^N (Q_i - \hat{Q}_i)^2}{\sum_{i=1}^N (Q_i - \bar{Q})^2} = 1.0 - E \quad (6)$$

Para admitir como aceptable la bondad del ajuste, los valores de R^2 y E deben aproximarse a 1 y los valores de %SEP y ARV a 0. Un valor de cero para E indica que el valor medio observado \bar{Q} es un buen estimador como el modelo, mientras valores negativos indican que la media observada es un mejor estimador que el modelo (Legates y McCabe Jr., 1999).

3. RESULTADOS

Sin un pretratamiento de la serie de datos de los usos registrados de agua, las magnitudes de evaluación de los modelos no ofrecen buenos resultados (Tabla 1). Los mejores valores para las magnitudes de evaluación, tanto para los modelos neuronales como para las regresiones múltiples, se obtienen en las parcelas dedicadas al monocultivo del olivar considerando como variables de entrada o independientes las demandas de agua de los dos días anteriores al de estimación. En todos los modelos considerados, las redes neuronales aportan mejores estimaciones que las regresiones múltiples. En la Figura 3 se muestra los resultados de la estimación de las demandas de agua en parcelas de olivar en el periodo de validación de la mejor red neuronal (2:9s:9s:1f) (2 neuronas en la capa de entrada, 9 neuronas en cada una de las capas intermedias y una neurona en la capa de salida).

TABLA 1. Bondad de ajuste de los modelos de regresión múltiple y redes neuronales considerando los datos originales registrados por SIGA

Cultivo	Variables independientes	Modelo	R^2	RMSE (m ³ /ha/día)	%SEP	E	ARV
Olivar	Demandas dos días anteriores	RNC(2:9:9:1)	0,40	22,76	51,69	0,38	0,62
		Regresión	0,28	35,33	88,32	0,18	0,82
	Demandas y datos climáticos dos días anteriores	RNC(14:11:11:1)	0,36	28,38	60,09	0,30	0,70
		Regresión	0,25	37,22	89,24	0,16	0,94
Algodón	Demandas dos días anteriores	RNC(2:7:7:1)	0,38	24,96	54,09	0,28	0,72
		Regresión	0,25	36,22	90,02	0,15	0,85
	Demandas y datos climáticos dos días anteriores	RNC(14:12:12:1)	0,33	30,09	63,19	0,29	0,71
		Regresión	0,22	40,61	91,14	0,13	0,87
Maíz	Demandas dos días anteriores	RNC(2:7:7:1)	0,36	25,23	55,34	0,22	0,78
		Regresión	0,27	36,10	89,44	0,16	0,84
	Demandas y datos climáticos dos días anteriores	RNC(14:13:13:1)	0,31	32,39	65,90	0,32	0,68
		Regresión	0,23	42,11	91,89	0,14	0,86

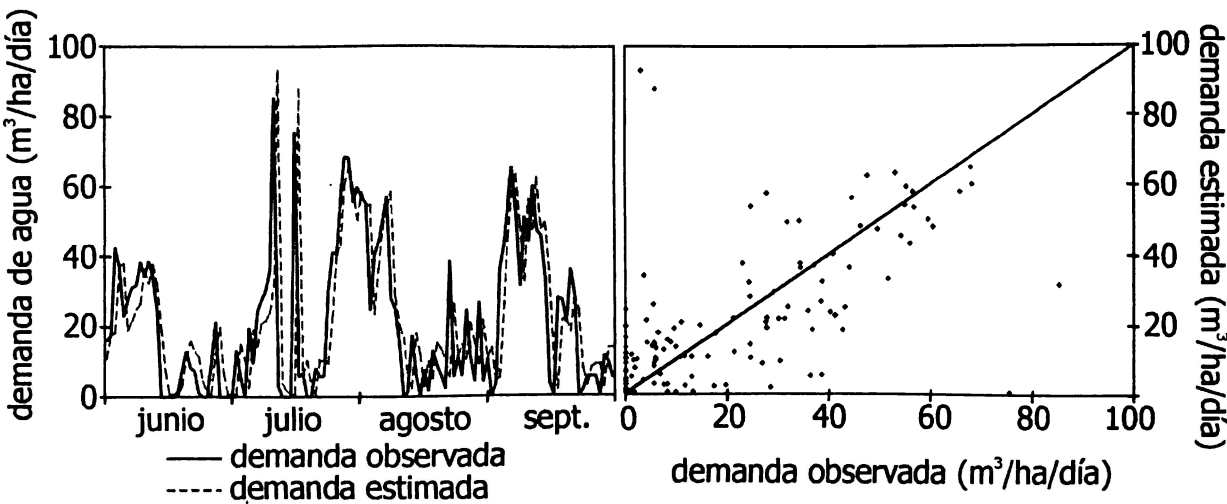


FIGURA 3. Demandas de agua diarias observadas y estimadas en validación por la red neuronal (2:9s:9s:11) con entradas la demanda de agua de los dos días anteriores en las parcelas de olivar

Sometiendo a la serie de datos a un alisamiento triangular, la mejora en los resultados es espectacular consiguiendo en casi todos los modelos R^2 superiores a 0,7 y %SEP menores a 30% (Tabla 2). Los mejores resultados se obtienen considerando como variables de entrada o independientes las demandas de agua de los dos días anteriores al de estimación. En este caso, los modelos de regresión múltiple aportan unas magnitudes de evaluación algo mejores que las RNCs, siendo las diferencias poco significativas. Sin embargo, las redes neuronales son

superiores a las regresiones múltiples cuando se consideran como variables de entrada las demandas de agua y datos climáticos de los dos días anteriores. En la Figura 4 se muestra los resultados de la estimación de las demandas de agua en parcelas de olivar en el periodo de validación de la mejor red neuronal (2:6s:6s:1l) (2 neuronas en la capa de entrada, 6 neuronas en cada una de las capas intermedias y una neurona en la capa de salida).

TABLA 2. Bondad de ajuste de los modelos de regresión múltiple y redes neuronales considerando los datos registrados por SIGA sometidos a un proceso de filtrado

Cultivo	Variables independientes	Modelo	R^2	RMSE (m ³ /ha/día)	%SEP	E	ARV
Olivar	Demandas dos días anteriores	RNC(2:6:6:1)	0,95	3,17	19,19	0,94	0,06
		Regresión	0,96	3,03	17,10	0,95	0,05
	Demandas y datos climáticos dos días anteriores	RNC(14:7:7:1)	0,78	7,29	24,11	0,67	0,33
		Regresión	0,75	8,33	47,10	0,45	0,55
Algodón	Demandas dos días anteriores	RNC(2:8:8:1)	0,93	4,36	19,40	0,92	0,08
		Regresión	0,95	3,71	15,49	0,95	0,05
	Demandas y datos climáticos dos días anteriores	RNC(14:7:7:1)	0,78	7,32	25,00	0,67	0,32
		Regresión	0,73	9,07	48,89	0,47	0,53
Maíz	Demandas dos días anteriores	RNC(2:7:7:1)	0,92	4,89	19,87	0,91	0,09
		Regresión	0,93	4,78	19,02	0,92	0,08
	Demandas y datos climáticos dos días anteriores	RNC(14:11:11:1)	0,79	6,76	24,03	0,69	0,31
		Regresión	0,73	9,33	48,09	0,46	0,54

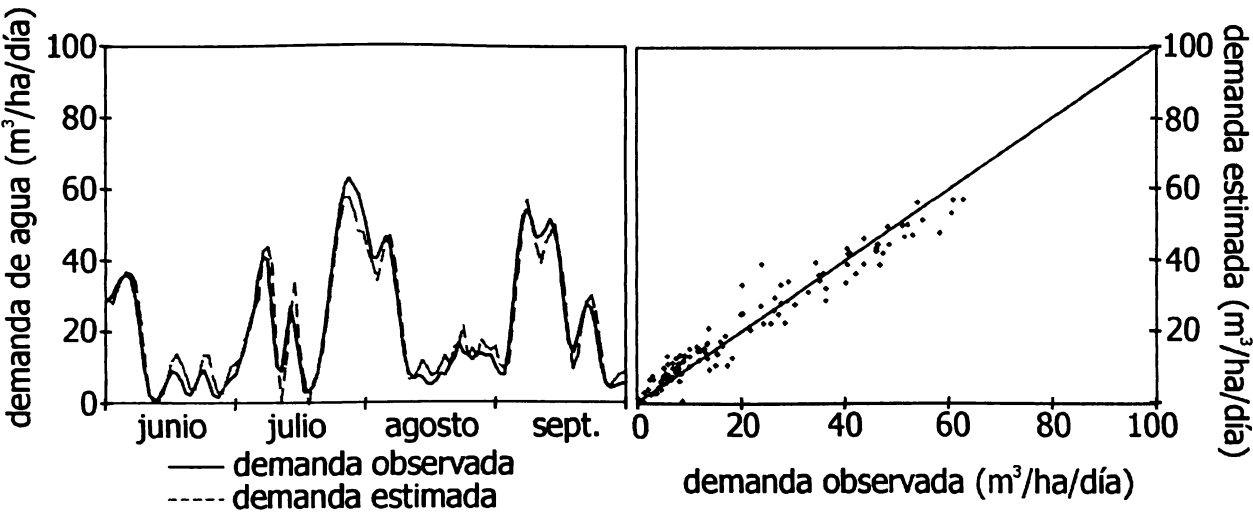


FIGURA 4. Demandas de agua diarias observadas y estimadas en validación por la red neuronal (2:6s:6s:1l) con entradas la demanda de agua de los dos días anteriores en las parcelas de olivar (series filtradas)

4. DISCUSIÓN

La caracterización de la demanda de agua de riego a partir de los datos originales registrados por el sistema SIGA, ha supuesto niveles de error muy alejados de los rangos estadísticos de aceptación para los modelos empleados de redes neuronales y regresiones múltiples. Esta falta de bondad del ajuste puede deberse a que se han considerado los datos registrados durante el proceso de implementación del sistema de captación de datos en la zona regable, período en el que hubo diversos problemas de robustez y fiabilidad del sistema detectándose en la serie de datos imprecisiones, errores y ruido.

La estimación de la demanda de agua a partir de la serie original filtrada aporta estimaciones correctas y términos de error aceptables. Con el proceso de filtrado definido se lleva a cabo un cambio de escala en la serie original de usos de agua, aumentando amplitudes y disminuyendo frecuencias. De este modo, se consigue espaciar temporalmente eventos de elevación o disminución brusca del parámetro a estimar, que es el origen del aumento de error. En definitiva, se consigue un efecto de suavizado gracias al cual se elimina el ruido e imprecisiones presentes en la serie original. Otra razón por la que se obtienen buenas estimaciones es el cambio en la naturaleza de la serie. Se pasa de una serie no lineal a una serie filtrada lineal (Blommfield, 1976). Esto explica el buen funcionamiento de los modelos lineales de regresiones múltiples y de las redes neuronales, aportando resultados similares. Sin embargo, la mayor capacidad de generalización de las RNCs queda patente cuando las variables de entrada tienen alta multicolinealidad entre ellas.

Los mejores resultados obtenidos se consiguen con los modelos que tienen como variables de entrada o independientes las demandas de los dos días anteriores al de estimación, cuando *a priori* parecería bastante probable que la precipitación condicionaría el uso de agua de riego. Esto puede deberse a la escasa precipitación de las campañas de riego estudiadas, y por tanto, podría ser interesante analizar en futuros trabajos campañas de riego u otras zonas regables con períodos de lluvia más abundantes para así evaluar la sensibilidad de la variable precipitación. Por otra parte, los resultados obtenidos pueden indicar que la inclusión del factor precipitación, así como de otras variables climáticas, se contempla de forma implícita en las observaciones de la demanda de agua (Saporta y Muñoz, 1994).

En los mejores modelos propuestos (con variables independientes las demandas de los dos días anteriores) se tienen las curvas estimadas desplazadas con respecto a las curvas observadas. Este desplazamiento se debe a que la variable demanda de agua del día anterior es la que muestra la mayor significación en los análisis de la varianza. En consecuencia se obtienen resultados en la validación con las series filtradas muy buenos según las magnitudes de evaluación R^2 , E y ARV, ya que éstas caracterizan la variación de los valores estimados frente a los observados. Sin embargo, la magnitud de %SEP resulta peor ya que ésta hace una evaluación punto a punto entre los valores observados y estimados. Esto mismo se muestra en trabajos realizados por otros autores en las estimaciones de otros parámetros utilizando regresiones múltiples, modelos ARIMA y/o redes neuronales (Park, 1998; Abrahart y See, 2000; Gutiérrez-Estrada *et al.*, 2005). Estos trabajos indican que este retraso temporal puede ser debido a factores no incluidos en los modelos. En el caso de la estimación de la demanda de agua, esta última hipótesis puede ser válida ya que en este trabajo no se aprecia el mencionado desplazamiento en los modelos que además consideran los datos climáticos de días anteriores.

El método de entrenamiento propuesto con validación interna se ha mostrado muy eficiente durante la generalización, ya que las redes han proporcionado una respuesta correcta ante patrones que no han sido empleados en el entrenamiento, es decir, no se incurre en un sobreajuste de los ejemplos presentados durante la fase de calibración.

El principal problema asociado a las RNCs desarrolladas en este trabajo ha sido la identificación de la arquitectura que implica un menor error de generalización. Esto ha supuesto un mayor tiempo de procesamiento de los datos en comparación con las regresiones múltiples. Siendo los sistemas neuronales modelos heurísticos, no se pueden dar reglas concretas sobre el control de la convergencia, del diseño de la red, y de la inicialización y cambio de pesos para resolver un problema concreto. En la bibliografía especializada lo único que se encuentra son unas cuantas pautas generales deducidas de la experiencia de numerosos autores. Por ello, es necesario conocer de antemano la naturaleza de la situación problema, condición no necesaria en otros muchas técnicas estadísticas de predicción. Por esto, en este trabajo se inician los análisis calculando el máximo retraso de las

variables de entrada usando la metodología de Box-Jenkins al igual que hace Griño (1992). En otros trabajos se realizan análisis de componentes principales para el preprocesamiento de los datos (Ventura *et al.*, 1997) o regresiones múltiples para identificar las variables independientes que influyen significativamente sobre la variable dependiente (Pulido-Calvo *et al.*, 2002, 2003).

5. CONCLUSIONES

La utilización de regresiones múltiples y redes neuronales computacionales ha proporcionado una capacidad aceptable de estimación de la demanda de agua en zonas regables cuando se hace un filtrado previo de las series de datos registradas por el sistema de telemetría. El filtrado previo de los datos originales de demanda de agua normaliza la serie favoreciendo, en general, un mejor ajuste de las regresiones múltiples y de las RNCs. Los mejores resultados se obtienen cuando se consideran como variables de entrada o independientes las demandas de los dos días anteriores al de estimación.

Esta caracterización de la demanda a corto plazo se puede utilizar como entrada en métodos y/o programas de gestión en tiempo real de los sistemas de distribución de agua, pero asimismo, permite un mejor ajuste de los volúmenes a bombear con la demanda real de la red de distribución y, por consiguiente, una utilización más racional de los recursos de agua y de energía.

Sería de interés ampliar la metodología desarrollada en este trabajo, e implementarla en otras zonas regables para establecer un modelo general para la gestión eficiente del agua en zonas regables. Este modelo comprenderá una primera parte de análisis y estimación de las demandas de agua, que servirá como ayuda a la toma de decisiones en una política de gestión eficiente del recurso.

6. BIBLIOGRAFIA

- Abrahart, R.J., See, L. y Kneale, P.E., 1999. Using pruning algorithms and genetic algorithms to optimise network architectures and forecasting inputs in a neural network rainfall-runoff model, *J. Hydroinformatics*, 1(2), pp. 103-114.
- Abrahart, R.J., y See, L., 2000. Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments, *Hydrol. Process.*, 14, pp. 2157-2172.
- Abrahart, R.J., y See, L., 2002. Multi-model data fusion for river flow forecasting: an evaluation of six alternative methods based on two contrasting catchments, *Hydrol. Earth Syst. Sci.*, 6(4), pp. 655-670.
- Anctil, F., y Rat, A., 2005. Evaluation of neural network streamflow forecasting on 47 watersheds, *J. Hydrol. Eng.*, 10(1), pp. 85-88.
- Bloomfield, P., 1976. Fourier analysis of time series: an introduction, John Wiley & Sons, Nueva York.
- Cameron, D., Kneale, P. y See, L., 2002. An evaluation of a traditional and a neural net modelling approach to flood forecasting for an upland catchment, *Hydrol. Process.*, 16, pp. 1033-1046.
- Chiang, Y.-M., Chang, L.-C. y Chang, F.-J., 2004. Comparison of static-feedforward and dynamic-feedback neural networks for rainfall-runoff modeling, *J. Hydrol.*, 290, pp. 297-311.
- Coulbeck, B., 1988. Computer control of water supply, Leicester Polytechnic, Leicester.
- Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function, *Math. Controls, Signals, and Systems*, 2: pp. 303-314.
- De Vries, B., y Principe, J.C. 1991. A theory for neural networks with time delays, *Advances in neural information processing systems* 3, Morgan Kaufmann Publishers, California.
- FAO, 1993. Las políticas de recursos hídricos y la agricultura, El estado mundial de la agricultura y la alimentación, Organización de las Naciones Unidas para la Agricultura, Roma.
- French, M.N., Krajewski, W.F., y Cuykendall, R.R., 1992. Rainfall forecasting in space and time using a neural network, *J. Hydrol.*, pp. 137, 1-31.
- Griño, R., 1992. Neural networks for univariate time series forecasting and their application to water demand prediction, *Neural Network World*, 2(5): pp. 437-450.
- Gutiérrez-Estrada, J.C., De Pedro-Sanz, E., López-

- Luque, R., y Pulido-Calvo, I., 2004. Comparison between traditional methods and artificial neural networks for ammonia concentration forecasting in an eel (*Anguilla anguilla* L.) intensive rearing system, *J. Aquacult. Eng.*, 31, pp. 183-203.
- Gutiérrez-Estrada, J.C., De Pedro-Sanz, E., López-Luque, R., y Pulido-Calvo, I., 2005. Estimación a corto plazo de la temperatura del agua. Aplicación en sistemas de producción en medio acuático, *Ing. Agua* (En prensa).
- Hsu, K., Gupta, H.V., y Sorooshian, S., 1995. Artificial neural network modeling of the rainfall-runoff process, *Water Resour. Res.*, 31(10): pp. 2517-2530.
- Kitanidis, P.K., y Bras, R.L., 1980. Real time forecasting with a conceptual hydrological model.2: Applications and results, *Water Resour. Res.*, 16(6): pp. 1034-1044.
- Kuligowski, R.J., y Barros, A.P., 1998. Experiments in short-term precipitation forecasting using artificial neural networks, *Mon. Wea. Rev.*, 126(2), pp. 470-482.
- Legates, D.R., y McCabe JR., G.J., 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation, *Water Resour. Res.*, 35(1), pp. 233-241.
- Lorrai, M., y Sechi, G.M., 1995. Neural nets for modelling rainfall-runoff transformations, *Water Resour. Manage.*, 9, pp. 299-313.
- Martín-Del-Brío, B., y Sanz-Molina, A., 2001. Redes neuronales y sistemas borrosos, Ra-Ma, Madrid.
- Mason, J.C., Tem'ne, A., y Price, R.K., 1996. A neural network model of rainfall-runoff using radial basis functions, *J. Recherches Hydrauliques*, 34(4), pp. 537-548.
- Moradkhani, H., Hsu, K., Gupta, H.V., y Sorooshian, S., 2004. Improved streamflow forecasting using self-organizing radial basis function artificial neural networks, *J. Hydrol.*, 295, pp. 246-262.
- Nash, J.E., y Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. I: A discussion of principles, *J. Hydrol.*, 10, pp. 282-290.
- Ohlsson, L., 1995. *Hydropolitics-conflicts over water as a development constraint*, Zed Books y University Press, Londres.
- ONU, 1997. *Comprehensive assessment of the freshwater resources of the world*, United Nations Department for Policy Coordination and Sustainable Development (DPCSD), Commission on Sustainable Development.
- Park, H.-H., 1998. Analysis and prediction of walleye pollock (*Theragra chalcogramma*) landings in Korea by time series analysis, *Fisheries Res.*, 38, pp. 1-7.
- Pulido-Calvo, I., Roldán, J., López-Luque, R., y Gutiérrez-Estrada, J.C., 2002. Técnicas de predicción a corto plazo de la demanda de agua. Aplicación al uso agrícola, *Ing. Agua*, 9(3): pp. 319-331.
- Pulido-Calvo, I., Roldán, J., López-Luque, R., y Gutiérrez-Estrada, J.C., 2003. Demand forecasting for irrigation water distribution systems, *J. Irrig. and Drain. Eng.*, 129(6): pp. 422-431.
- Roger, L.L., y Dowla, F.U., 1994. Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling, *Water Resour. Res.*, 30(2), pp. 457-481.
- Rumelhart, D.E., Hinton, G.E., y Williams, R.J., 1986. Learning representations by backpropagation errors, *Nature*, 323(9): 5pp. 33-536.
- Saporta, D., y Muñoz, M., 1994. El consumo en redes de distribución. Predicción diaria de la demanda, Mejora del rendimiento y de la fiabilidad en sistemas de distribución de agua, Cabrera y Vela (eds.), Universidad Politécnica de Valencia, Valencia.
- See, L., y Openshaw, S., 2000. A hybrid multi-model approach to river level forecasting, *Hydrol. Sci. J.*, 45(4), pp. 523-536.
- Shepherd, A. J., 1997. *Second-Order Methods for Neural Networks*, Springer, Nueva York.
- Shin, H.-S., y Salas, J.D., 2000. Regional drought analysis based on neural networks, *J. Hydrol. Eng.*, 5(2), pp. 145-155.

- Sumpsi, J.M., Garrido, A., Blanco, M., Varela, C., y Iglesias, E., 1998. Economía y política de gestión del agua en la agricultura, Ministerio de Agricultura, Pesca y Alimentación y Ed. Mundi-Prensa, Madrid.
- Tan, Y., y Van Cauwenberghe, A., 1999. Neural-network-based d-step-ahead predictors for nonlinear systems with time delay, *Eng. Applic. Artif. Intell.*, 12(1), pp. 21-25.
- Thirumalaiah, K., y Deo, M.C., 1998. River stage forecasting using artificial neural networks, *J. Hydrol. Eng.*, 3(1), pp. 26-32.
- Thirumalaiah, K., y Deo, M.C., 2000. Hydrological forecasting using neural networks, *J. Hydrol. Eng.*, 5(2), 180-189.
- Tokar, A.S., y Johnson, P.A., 1999. Rainfall-runoff modeling using artificial neural networks, *J. Hydrol. Eng.*, 4(3), pp. 232-239.
- Tokar, A.S., y Markus, M. 2000. Precipitation-runoff modeling using artificial neural networks and conceptual models, *J. Hydrol. Eng.*, 5(2), pp. 156-161.
- Tsoukalas, L.H., y Uhrig, R.E., 1997. Fuzzy and neural approaches in engineering, Wiley Interscience, Nueva York.
- Ventura, S., Silva, M., Pérez-Bendito, D., Hervás, Y.C., 1995. Artificial neural networks for estimation of kinetic analytical parameters, *Anal. Chem.*, 67(9): pp. 1521-1525.
- Ventura, S., Silva, M., Pérez-Bendito, D., Hervás, Y.C., 1997. Computational neural networks in conjunction with principal component analysis for resolving highly nonlinear kinetics, *J. Chem. Inf. Comput. Sci.*, 37(2): pp. 287-291.
- Yang, C.C., Prasher, S.O., Lacroix, R., Sreekanth, S., Patni, N.K., y Masse, L., 1997. Artificial neural network model for subsurface-drained farmland, *J. Irrig. Drain. Eng.*, 123(4), pp. 285-292.
- Zhang, M., Fulcher, J. y Scofield, R.A. 1997. Rainfall estimation using artificial neural network group, *Neurocomputing*, 16, pp. 97-115.

