

MODELADO DE SERIES CLIMATOLÓGICAS MEDIANTE UNA RED NEURONAL ARTIFICIAL

Jesús Palazón González¹ y Adela García Guzmán²

Resumen:

Se ha desarrollado un modelo de red neuronal para caracterizar series meteorológicas que son difíciles de modelar con los métodos clásicos de inferencia estadística. Concretamente, se ha utilizado la red neuronal para cuantificar la relación intensidad – duración de la lluvia, variables que se encuentran interrelacionadas de una forma muy imprecisa. El modelo contiene funciones de transferencia no lineales e incluye términos de naturaleza estadística en la función de error. Para estimar los parámetros de la red neuronal se ha desarrollado un algoritmo de aprendizaje adaptado a funciones de error no derivables.

Palabras clave: Intensidad de la lluvia, dependencia de la intensidad y duración de la lluvia, red neuronal artificial, términos de error en una red neuronal artificial.

INTRODUCCIÓN

La finalidad principal de un estudio hidrológico es la obtención de datos básicos que permitan evaluar los recursos hídricos de una cuenca. En la actualidad, la mejor forma de evaluar estos recursos se basa en la implantación de modelos que caractericen y simulen la forma en la que las precipitaciones se transforman en escorrentía. En este sentido, la variable climática más importante es la lluvia, ya que de ella dependerán la distribución y los valores máximos de los recursos disponibles.

El modelado de la lluvia plantea ciertas dificultades debido a la dependencia temporal de las observaciones y a la dependencia cruzada entre las variables asociadas a la ocurrencia de lluvia, como pueden ser la duración, la intensidad media o la variación de la intensidad dentro de cada episodio lluvioso. Todas estas relaciones de dependencia adquieren especial importancia en modelos que generan datos agregados en intervalos cortos de tiempo.

En este trabajo se propone un modelo de red neuronal artificial (RNA) que relaciona la duración y la intensidad media de una lluvia. Se han elegido estas variables porque son componentes importantes de los modelos pluviométricos y porque sus relaciones son difíciles de cuantificar, especialmente en datos con escala de agregación pequeña.

Las RNAs constituyen una alternativa a la generación de modelos empíricos y han sido utilizadas con éxito en tareas de modelado de sistemas y ajuste de funciones no lineales. Su flexibilidad para el ajuste de datos con gran dispersión numérica, junto con la ventaja que supone no tener que precisar el enunciado de una relación matemática exacta (Alvarez y Bolado, 1996), hacen interesante su uso para caracterizar procesos hidrológicos, en los que resulta deseable una descripción precisa de la variabilidad.

Departamento de Estadística e Investigación Operativa. Escuela Técnica Superior de Ingenieros Agrónomos y Montes de Córdoba
Apdo. 3048. 14080 Córdoba. Tlf. 957218478
E-mail: ¹ ma2pagoj@uco.es ² ma1gagua@uco.es

Artículo recibido el 8 de abril de 2002, recibido en forma revisada el 19 de noviembre de 2002 y el 14 de marzo de 2003 y aceptado para su publicación el 8 de abril de 2003. Pueden ser remitidas discusiones sobre el artículo hasta seis meses después de la publicación del mismo siguiendo lo indicado en las "Instrucciones para autores". En el caso de ser aceptadas, éstas serán publicadas conjuntamente con la respuesta de los autores.

Entre dichos procesos hidrológicos cabe citar la predicción de la lluvia en el espacio y en el tiempo (French et al., 1992), transporte de solutos (Rogers y Dowla, 1994), modelado del proceso lluvia – escorrentía (Hsu et al., 1995; Mason et al., 1996), retención del agua en el suelo (Schaap y Bouten, 1996), determinación de la red de drenaje (Kao, 1996), estimación de parámetros en movimiento no forzado (Morshed y Kaluarachchi, 1998), predicción de avenidas (Whitley y Hromadka II, 1999), predicción de la salinidad (Maier y Dandy, 1999), estimación de variables físicas a partir de imágenes tomadas con satélite (Hsu et al., 1999), control de redes hidráulicas (Dibike et al., 1999), gestión de embalses (Cancelliere et al., 2002), etc. Una revisión más completa puede verse en ASCE (2000) y en Govindaraju y Ramachandra (2000).

REDES NEURONALES ARTIFICIALES

Una RNA es un modelo matemático que trata de emular a los sistemas neuronales biológicos en el procesado de la información. En la figura 1 se muestra una RNA estructurada en capas, dentro de las cuales se integran las neuronas, representadas mediante funciones matemáticas que procesan y envían información de unas capas a otras. La información que se transmite a través de las conexiones de la red está ponderada por pesos, los cuales modulan la intensidad de la relación entre neuronas.

El fundamento matemático de las RNAs se encuentra en el teorema de Kolmogorov (1957) sobre la representación exacta de una función continua de d variables mediante la suma de una composición de $2d+1$ funciones continuas univariantes. Hecht–Nielsen (1987 y 1990), aplicando el teorema de Kolmogorov, demostró que una arquitectura de características similares al perceptrón multicapa, con d neuronas de entrada y una única capa oculta de $2d+1$ neuronas, resultaba ser un aproximador universal de funciones continuas si se encuentran funciones de transferencia que cumplieren con las hipótesis de Kolmogorov. Asimismo, Leshno et al. (1993) basándose, entre otros, en los trabajos de Hecht–Nielsen (1987 y 1990), Cybenko (1989), Funahashi (1989), Hornik et al. (1990) y Kurkova (1991), demostraron que un perceptrón multicapa convencional, con una única capa oculta, bajo un conjunto de condiciones para las funciones de transferencia, constituye un aproximador universal de funciones. No obstante, Hecht–Nielsen (1990) indica que, para ciertos problemas,

podría ser más operativo emplear dos o más capas ocultas porque permite flexibilizar la elección de las funciones de transferencia o disminuir el número de neuronas.

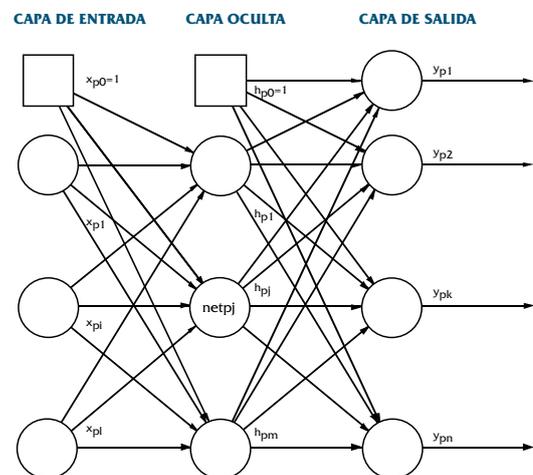


Figura 1. Arquitectura de una red con una capa oculta y conexiones hacia adelante

Al construir una red, las características fundamentales que deben definirse son el número de capas, el número de neuronas por capas, el tipo de conexiones entre neuronas, la función de transferencia o respuesta de las neuronas y el mecanismo de aprendizaje.

El tamaño de las capas de entrada y de salida está fijado por la propia naturaleza de la aplicación, y sólo pueden escogerse el número de neuronas ocultas que componen las capas intermedias y cuántas de estas capas habrá. Para determinar el número de neuronas ocultas es mejor iniciar la red con un número bajo, y si los resultados no son satisfactorios se va aumentando progresivamente el número de éstas hasta que el error sea aceptable. En lo que se refiere al número de capas ocultas, una o dos suelen ser suficientes en la mayoría de los problemas.

Con respecto al tipo de conexiones, las más usadas son las conexiones hacia adelante o *feedforward*, es decir, las neuronas de una capa reciben señales de entrada de otra capa anterior, más cercana a la entrada de la red, y envían las señales de salida a una capa posterior, más cercana a la salida de la red. La red utilizada en este trabajo tiene conexiones hacia adelante y una capa oculta (figura 1).

Como funciones de transferencia, las más empleadas son las funciones sigmoideal e identidad.

Aunque puede utilizarse cualquier función, es necesario que ésta sea continua y derivable si se va a emplear un método de optimización que haga uso de derivadas. Según Fine (1999), si las funciones de transferencia admiten sucesivas derivadas no nulas, son necesarios menos parámetros y una única capa oculta para conseguir un mismo error.

El proceso de ajuste de los pesos a los valores correctos, el aprendizaje, se lleva a cabo mediante un conjunto de ejemplos de entradas – salidas objetivo, llamados patrones de entrenamiento, de forma que se minimice la diferencia entre la respuesta estimada por la red y la salida deseada. Una vez ajustados los pesos, el modelo debe ser capaz de dar respuestas correctas para valores de entrada que no hayan sido utilizados en la fase de entrenamiento. Con respecto al mecanismo de aprendizaje (Hecht–Nielsen, 1990; Freeman y Skapura, 1993; Hiler y Martínez, 1995; Swingler, 1996; Martín del Brío y Sanz Molina, 1997), uno de los más usados es el de la regla delta generalizada. Este método es muy simple de cálculo y fácil de implementar, pero tiene el inconveniente de verse afectado por el orden en el que son presentados los patrones a la red y de converger en los múltiples mínimos locales existentes. Para evitar estos inconvenientes, Johansson et al. (1992), Rogers y Dowla (1994), y Whitley y Hromadka II (1999) utilizan un método de gradiente conjugado. Otros autores (Maier y Dandy, 1999) emplean métodos como el quickprop, basados en el método de Newton que requiere derivadas segundas, y Hagan y Menhaj (1994) utilizan el método de Levenberg–Marquardt. Algunos trabajos relativamente recientes en redes neuronales utilizan métodos estocásticos de optimización, los cuales no requieren derivadas y tienen la ventaja de que pueden eludir mínimos locales. Entre estas estrategias cabe citar el método de *Nelder y Mead* combinado con mínimos cuadrados lineales (Hsu et al., 1995), el temple simulado (Bárdossy, 1998) y los algoritmos genéticos (Morshed y Kaluarachchi, 1998). La abundancia de trabajos sobre el ajuste de pesos indica que no hay un procedimiento comúnmente aceptado y que, para cada RNA, el éxito del mismo depende de la eficiencia del método de optimización y de su capacidad para eludir los mínimos locales.

RELACIÓN INTENSIDAD – DURACIÓN

En los modelos de tiempo continuo, una forma muy simple de describir la cantidad de lluvia es expresándola como el producto de la duración por una intensidad aleatoria constante, independiente de la duración, y, generalmente, con distribución exponencial. Este esquema se utiliza en el modelo de Eagleson (1978) y en la serie de modelos cluster estudiados por Rodríguez–Iturbe et al. (1987) y Entakhabi et al. (1989). En otros modelos, la descripción de la cantidad se hace más compleja al introducirse relaciones de dependencia entre la intensidad y otros elementos del modelo. Tal es el caso de Istok y Boersma (1989), que introducen la dependencia entre la intensidad y la duración de la lluvia en sucesivos intervalos de tiempo a partir de un proceso autorregresivo y no estacionario. Hutchinson (1990) mantiene la distribución exponencial, pero introduce dicha dependencia a través de un proceso de cadena de Markov con tres estados, y Goel et al. (2000) incorporan la correlación entre la duración y la intensidad de la lluvia en un modelo de distribución de frecuencias de avenidas.

Siguiendo estos trabajos, se ha supuesto, y comprobado, que existe una cierta dependencia entre la duración total y la intensidad media de la lluvia. Los datos disponibles para desarrollar el modelo son duraciones de la lluvia, en minutos, y cantidades de lluvia, en centésimas de pulgada, registradas en un observatorio de Coshocton (Ohio), entre los años 1938 y 1989, a excepción de 1973 y 1974, de los que no se dispone de datos. La ocurrencia de lluvia está registrada en forma continua, es decir, se registra la hora exacta de inicio, de cada instante en que cambia la intensidad y de finalización de la lluvia, así como la cantidad recogida en cada uno de estos intervalos. En lo que sigue, las cantidades de lluvia se expresan en milímetros, y los datos a utilizar se reducen a la duración total de cada lluvia y a la intensidad media de la misma.

La intensidad media y la duración de las lluvias varían a lo largo del año, por lo que se decidió analizar los datos, separándolos en dos periodos: uno húmedo, de octubre a mayo con 4886 pares de datos, y otro seco, de junio a septiembre con 2681 pares de datos.

Tabla 1. Medias mensuales de la duración e intensidad media de la lluvia

	Oct.	Nov.	Dic.	Ene.	Feb.	Mar.	Abr.	May	Jun.	Jul.	Ago.	Sep.
Duración	94.06	111.49	129.11	136.95	128.38	101.77	95.02	82.52	64.04	64.09	70.03	88.68
Intensidad	3.24	3.20	3.15	3.59	3.48	3.42	3.53	3.61	4.76	5.67	5.24	5.05

La selección de los meses a incluir en cada grupo se hizo con el método de clasificación de k-medias, para datos tipificados. Los datos utilizados en la clasificación se muestran la tabla 1 y en ella podemos comprobar que los dos grupos están claramente diferenciados.

La hipótesis de independencia entre la duración y la intensidad se rechazó, en ambos periodos, con un test χ^2 realizado con tablas de frecuencias 3x3. Los resultados de este test se resumen en la tabla 2 junto con un test de igualdad de distribución en los dos periodos. Las frecuencias de los distintos intervalos de duración - intensidad se muestran en la figura 2 donde puede observarse la relación entre ambas variables ya que con las duraciones cortas (menores o iguales a 30 minutos) hay una mayor frecuencia de intensidades medias y altas (más de 2 mm h⁻¹), y con duraciones largas (mayores a 90 minutos) hay una mayor frecuencia de intensidades bajas (menores o iguales a 2 mm h⁻¹).

También puede verse en la figura 2 que la relación duración - intensidad, aunque sigue una pauta decreciente común a ambos periodos, presenta características específicas de cada uno de ellos, predominando las lluvias largas de baja intensidad en el periodo húmedo, y las lluvias cortas e intensas en el seco. Estas diferencias se confirman con el test estadístico de la tabla 2, por el que se rechaza la igualdad de la distribución bivalente en los dos periodos.

Otro aspecto importante que muestra la figura 2 es que la relación duración - intensidad no puede describirse por una simple fórmula matemática, porque para una duración dada puede esperarse cualquier intensidad, a veces casi con igual probabilidad, como sucede con las duraciones cortas del periodo húmedo o las duraciones medias del periodo seco.

Esto nos ha llevado a seleccionar un modelo de RNA con entradas múltiples, las duraciones, e igual número de salidas, las intensidades, lo que permite asociar entradas de igual valor con salidas diferentes.

Figura 2. Relación entre la duración y la intensidad en el periodo húmedo (PH) y seco (PS).

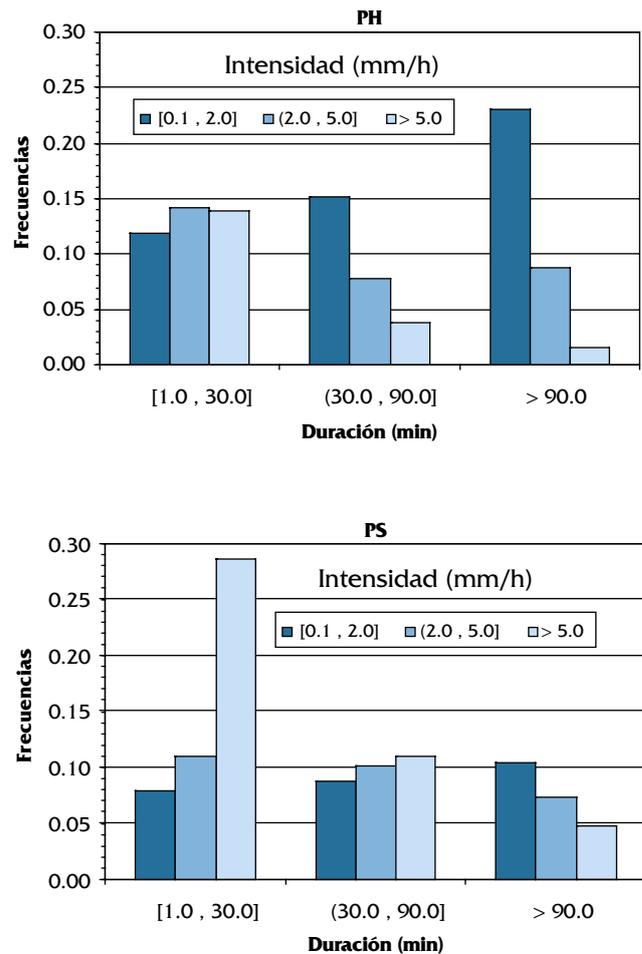


Tabla 2. Análisis estadístico previo: test y región de la hipótesis nula con $\alpha=0.05$

HIPÓTESIS NULA	Periodo húmedo	Periodo seco
Independencia	Test $\chi^2 = 750.86$ $C_0 = (0, 9.49)$ Grados de libertad = 4	Test $\chi^2 = 311.44$ $C_0 = (0, 9.49)$ Grados de libertad = 4
Igualdad de Distribución	Común a los dos periodos Test $\chi^2 = 656.06$ $C_0 = (0, 15.51)$; Grados de libertad = 8	

FORMULACIÓN DEL MODELO

Muchas de las aplicaciones (Freeman y Skapura, 1993) de las RNAs se han orientado hacia modelos de respuesta a estímulos, diagnóstico de enfermedades, reconocimiento de formas o corrección de fallos, que con frecuencia, requieren datos de salida binarios. Esta especificidad ha condicionado, en cierta medida, la elección de las funciones matemáticas para procesar la información de la RNA y el desarrollo de los procedimientos de estimación de pesos. Este es el caso de la función sigmoideal (1), que reúne muchas de las cualidades deseables en una función de transferencia, ya que admite sucesivas derivadas no nulas y toma valores en (0, 1), lo que facilita el redondeo para representar datos binarios; pero tiene el inconveniente de ser muy poco sensible para valores de x fuera del intervalo (-5, 5), ya que se hace prácticamente nula para valores de entrada inferiores a -5 y prácticamente la unidad para valores superiores a 5.

$$f(x) = (1 + e^{-x})^{-1} \quad (1)$$

Como alternativa a la anterior función, en este trabajo se propone la función seno cuadrado:

$$f(x) = (\text{sen } x)^2 \quad (2)$$

para la capa oculta. Esta función también admite sucesivas derivadas y toma valores en [0, 1] pero no se hace constante para valores grandes de x y, al ser periódica, puede emular la posible repetición de los valores de salida para entradas muy diferentes.

Para la capa de salida se propone la función coseno hiperbólico, corregida con el coeficiente 0.9, para poder reproducir la mínima intensidad de 0.1 mm h⁻¹:

$$f(x) = \cosh x - 0.9 = (e^x + e^{-x})/2 - 0.9 \quad (3)$$

Esta función se ha preferido a la función identidad, de uso más frecuente, porque admite sucesivas derivadas y, cabe esperar, que mantenga el error aceptablemente bajo con sólo una capa oculta y un número reducido de neuronas en ella. Otra ventaja de la función (3), frente a la función identidad, es que no se hace negativa y produce salidas exactamente en el intervalo deseado [0.1, ∞).

Las funciones (2) y (3) se introducen en la red propuesta, con una capa oculta, P patrones de entrenamiento, l neuronas de entrada, m neuronas ocultas

y n neuronas de salida, para producir las salidas y_{pk} (intensidades) como respuesta a las entradas $\ln(x_{pi})$ (logaritmos neperianos de las duraciones):

$$y_{pk} = \cosh \left(\sum_{j=0}^{l+m} w_{jk} \cdot h_{pj} \right) - 0.9 \quad ; \quad k = 1, \dots, n \quad (4)$$

$$h_{pj} = \begin{cases} x_{p0} = 1 & ; \quad j = 0 \\ \left[\text{sen} \left(\sum_{i=0}^{l-1} w_{ij} \cdot \ln(x_{pi}) \right) \right]^2 & ; \quad j = 1, \dots, m \end{cases} \quad (5)$$

donde:

w_{ij} , w_{jk} son los pesos o parámetros de la red
 x_{p0} , h_{p0} son las neuronas ficticias o de sesgo
 h_{pj} es la salida de la neurona oculta j para el patrón p
 $n=l$, lo que proporciona un total de $2nm+m+n$ pesos

La evaluación del aprendizaje de la red se mide comparando la salida estimada por la red, y_{pk} , con el valor de salida objetivo o deseado, d_{pk} , para todos los patrones de entrenamiento. La medida más frecuente del error de un modelo es la suma de los errores al cuadrado:

$$e = \sum_{p=1}^{P} \sum_{k=1}^{n} (d_{pk} - y_{pk})^2 \quad (6)$$

Esta función depende de la escala de medida de las variables, por lo que muchos autores aconsejan transformar los datos dividiéndolos por el máximo valor observado. Las variables de este trabajo no tienen un claro límite superior, sino tan sólo valores históricos que en algún momento pueden superarse, lo que nos ha llevado a buscar otro tipo de transformaciones. Concretamente, se han suavizado los datos de entrada mediante el logaritmo neperiano, y se plantea la función de error (7), que permite trabajar con datos muy heterogéneos sin la necesidad de normalizar o escalar.

$$e = \sum_{p=1}^{P} \sum_{k=1}^{n} \left[\ln \left(\frac{d_{pk}}{y_{pk}} \right) \right]^2 \quad (7)$$

Otra característica que se ha querido incorporar al modelo de RNA es la capacidad de reproducir en los datos de salida la estructura aleatoria que se observa en los datos originales.

Sin aumentar el número de neuronas, o el de capas, se puede alcanzar este objetivo incorporando a la función de error algunos términos adicionales. El uso de términos de penalización se recomienda para evitar resultados anómalos, tales como valores excesivamente grandes o pequeños de los pesos, o salidas inverosímiles de la red. En este trabajo se proponen los términos (8) y (9) asociados a la distribución de frecuencias de los datos de salida.

Para tener en cuenta la distribución univariante de las salidas se utilizan las diferencias entre la media, \bar{d} , y varianza, S_d^2 , observadas con las estimadas con la red, \bar{y} y S_y^2 :

$$B = \sqrt{\frac{(\bar{d} - \bar{y})^2 + (S_d^2 - S_y^2)^2}{2}} \quad (8)$$

Y para tener en cuenta la distribución bivalente de las entradas – salidas de la red se ha empleado el coeficiente C , formalmente igual al coeficiente de Cramer, el cual toma valores en el intervalo [0,1]:

$$T = \sum_{i=1}^{i=f} \sum_{j=1}^{j=c} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad ; \quad C = \frac{T}{N \cdot (q-1)} \quad ; \quad (9)$$

$$N = \sum_{i=1}^{i=f} \sum_{j=1}^{j=c} n_{ij} = P \cdot l = P \cdot n$$

siendo n_{ij} la frecuencia absoluta observada en los pares de datos de entrenamiento duración – intensidad, e_{ij} la frecuencia absoluta dada por las salidas de la red, f y c el número de filas y columnas, respectivamente, de la tabla bidimensional de frecuencias, N el número total de pares de datos de entrenamiento y $q = \text{mínimo}(f, c)$.

Estos términos de naturaleza estadística se incorporan a la función de error (7), resultando:

$$e = A + B + C = \sqrt{\frac{\sum_{p=1}^{p=P} \sum_{k=1}^{k=n} \left[\ln \left(\frac{d_{pk}}{y_{pk}} \right) \right]^2}{N}} + \sqrt{\frac{(\bar{d} - \bar{y})^2 + (S_d^2 - S_y^2)^2}{2}} + \frac{T}{N \cdot (q-1)} \quad (10)$$

MÉTODO DE AJUSTE

El término C de la ecuación (10) no es derivable con respecto a los pesos, w_{ij} y w_{jk} , lo que hace necesario emplear un método de optimización que no haga uso de las derivadas. En este trabajo se ha empleado el método de Rosenbrock (Himmelblau, 1972), que es muy eficiente para hallar el mínimo, sin restricciones, de una función multidimensional no derivable, por medio de búsquedas unidimensionales a lo largo de un conjunto de direcciones ortogonales generadas durante el proceso iterativo. Sin embargo, la búsqueda con el método de Rosenbrock también puede detenerse en algún mínimo local, \vec{X}^* , por lo que deben buscarse un vector unitario, \vec{S} , y un escalar, λ , tales que:

$$\vec{X} = \vec{X}^* + \lambda \cdot \vec{S} \quad ; \quad f(\vec{X}) < f(\vec{X}^*) \quad (11)$$

El conjunto de vectores \vec{S} es infinito, por lo que es necesario definir una estrategia para seleccionarlos. Concretamente, con W variables, se definen W^2 nuevas direcciones de búsqueda, formadas por W vectores unitarios, que proporcionan cambios en una variable, $W(W-1)/2$ vectores que proporcionan cambios del mismo signo en dos variables y $W(W-1)/2$ vectores que proporcionan cambios de signo contrario en dos variables. A partir del mínimo local previamente obtenido, \vec{X}^* , y utilizando las nuevas direcciones, se inicia la búsqueda de un nuevo mínimo. Si la búsqueda resulta fructífera se reinicia el método de Rosenbrock a partir de la mejor solución obtenida y, en caso contrario, el último mínimo localizado constituye la solución buscada.

Para iniciar el proceso de optimización hay que preparar los patrones de entrenamiento. Para este fin se seleccionan aleatoriamente y sin reposición N pares de datos (x_{pi}, d_{pi}) del total de la muestra disponible, que formarán, también al azar, P patrones con n pares de datos cada uno, $N=P \cdot n$. Las duraciones constituyen las entradas a la red, x_{pi} , y las intensidades medias las salidas deseadas de la misma, d_{pk} , $k=i$. Se ordenan dichos pares de datos en orden creciente de duración y, en caso de tener igual duración, por orden creciente de intensidad. La ordenación de los pares de datos es específica de este modelo porque, como puede verse en (4) y (5), cualquier cambio en el orden de las entradas $(x_{p1}, x_{p2}, \dots, x_{pl})$ afecta a todas las salidas, lo que haría inconsistente el posterior emparejamiento (x_{pi}, d_{pi}) de la duración – intensidad.

La ordenación de los pares de igual duración por orden creciente de intensidad, se requiere sólo para la estimación de los pesos del modelo (aprendizaje), ya que en la fase de validación, o para la simulación, los valores observados de intensidad no se usan, o no se conocen.

El entrenamiento de esta red puede resumirse en los siguientes pasos:

- 1) Inicializar los pesos de la red, w_{ij} y w_{jk} , con valores aleatorios uniformes elegidos al azar en el intervalo (-0.5 , 0.5).
- 2) Presentar un patrón de entrenamiento con entradas ($x_{p1}, x_{p2}, \dots, x_{pl}$) y salidas objetivo o deseadas ($d_{p1}, d_{p2}, \dots, d_{pn}$), siendo: $l = n$.
- 3) Calcular la salida estimada por la red neuronal ($y_{p1}, y_{p2}, \dots, y_{pn}$), para el patrón que se acaba de presentar, según la expresión (4).
- 4) Calcular el error cometido mediante la expresión (10) después de haber presentado todos los patrones disponibles para entrenar la red.
- 5) Actualizar los pesos w_{ij} y w_{jk} mediante el método de Rosenbrock y los refinamientos propuestos, de tal modo que disminuya el error.
- 6) Repetir los pasos 2 a 5 hasta que se cumpla alguno de los siguientes supuestos: el término de error resulte aceptablemente pequeño ($< 10^{-6}$), se hayan probado sin éxito las W^2 direcciones para eludir el último mínimo local encontrado, o se haya superado el número máximo de iteraciones permitidas, que en este caso ha sido de 1000.

El valor óptimo de los pesos así obtenidos proporciona la relación existente entre la intensidad media (salidas) y la duración de la lluvia (entradas). Estos pesos se podrán emplear posteriormente en la fase de validación o para la simulación de la intensidad media.

RESULTADOS Y DISCUSIÓN

A partir de los datos de duración e intensidad de la lluvia disponibles se procedió, de forma heurística, a buscar la arquitectura más simple posible compatible con un resultado aceptable. Después de algunas pruebas iniciales, se adoptó una red con 20 entradas y 20 salidas buscando conciliar dos objetivos contrapuestos: minimizar el número de parámetros del modelo y maximizar la información, sobre la estructura aleatoria de las entradas, que se pretende introducir en el mismo. La limitación en el número de entradas, y en el total de parámetros, viene impuesta por el número de patrones de entrenamiento disponibles, ya que con pocos patrones y muchos pesos, se obtiene un modelo sobredimensionado que no siempre produce resultados consistentes en la fase de validación. Esta limitación ha resultado especialmente restrictiva en este trabajo porque en el periodo seco sólo hay 2681 observaciones.

Para entrenar todos los modelos de RNA se seleccionaron aleatoriamente y sin reposición 2000 pares de datos duración – intensidad del total de la muestra disponible, así como otros 680 pares de datos, diferentes de los anteriores, para validar el modelo. De este modo resultaron, para 20 entradas, 100 y 34 patrones para entrenar y validar, respectivamente.

Previamente se había comprobado que con una RNA clásica, es decir, con funciones de transferencia sigmoidea e identidad para la capa oculta y de salida, respectivamente, función de error según ecuación (7) y entrenada con el método del gradiente conjugado, eran necesarias 20 entradas, 20 salidas y 25 neuronas ocultas, con lo cual resultaban 1045 pesos. Los resultados de esta prueba se muestran en la tabla 3.

Los términos B y C del error no aparecen en la fase de entrenamiento porque no existen en este modelo, sin embargo se han calculado en la fase de validación para mostrar la necesidad de incluirlos en la función de error si se quieren reproducir algunas características de la distribución de frecuencias de las salidas.

Tabla 3. Términos de error de la red BPN clásica con 25 neuronas ocultas

ERROR	PERIODO HUMEDO		PERIODO SECO	
	Entrenamiento	Validación	Entrenamiento	Validación
A	1.07	1.32	1.31	1.52
B	---	13.06	---	15.56
C	---	0.17	---	0.14

Si se compara el término de error A con el obtenido en las tablas siguientes, tablas 4 a 7, se comprueba que es del mismo orden, tanto en la fase de entrenamiento como en la de validación, pero se consigue con 1045 pesos frente a los 102 que utiliza el modelo propuesto. El excesivo número de pesos, con relación a los datos disponibles, nos llevó a desechar la RNA clásica.

En las tablas 4 y 5 se presentan los diferentes términos de error de la ecuación (10) para el periodo húmedo, y en las tablas 6 y 7 para el seco. El proceso de selección de patrones, estimación de pesos y vali-

dación se repitió tres veces para confirmar la consistencia de los resultados obtenidos y se eligió la RNA con 2 neuronas ocultas porque producía menor error de validación. En las tablas 5 y 7 se muestran las tres pruebas realizadas para el modelo propuesto.

En todos los casos, salvo en el modelo de 1 neurona para el periodo húmedo, se produce un mayor error en el periodo seco, debido, quizás, a la mayor variabilidad de la intensidad en este periodo. En general, el error A predomina en la fase de entrenamiento, ya que este término incluye la suma de 2000 términos.

Tabla 4. Términos de error del periodo húmedo para 1, 3, 4 y 5 neuronas ocultas

ERROR	1 neurona oculta		3 neuronas ocultas		4 neuronas ocultas		5 neuronas ocultas	
	Entrenamiento	Validación	Entrenamiento	Validación	Entrenamiento	Validación	Entrenamiento	Validación
A	1.40	1.40	1.31	1.36	1.28	1.33	1.23	1.29
B	6.68	14.30	0.00	6.59	0.00	4.65	0.00	4.53
C	0.03	0.10	0.03	0.09	0.02	0.09	0.03	0.10
TOTAL	8.11	15.80	1.34	8.04	1.30	6.07	1.26	5.91

Tabla 5. Términos de error del periodo húmedo. Red propuesta con 2 neuronas oculta

ERROR	PRUEBA 1		PRUEBA 2		PRUEBA 3	
	Entrenamiento	Validación	Entrenamiento	Validación	Entrenamiento	Validación
A	1.33	1.25	1.31	1.35	1.28	1.29
B	0.03	1.98	0.00	1.99	0.00	0.24
C	0.03	0.09	0.02	0.08	0.01	0.08
TOTAL	1.39	3.32	1.33	3.42	1.29	1.61

Tabla 6. Términos de error del periodo seco para 1, 3, 4 y 5 neuronas ocultas

ERROR	1 neurona oculta		3 neuronas ocultas		4 neuronas ocultas		5 neuronas ocultas	
	Entrenamiento	Validación	Entrenamiento	Validación	Entrenamiento	Validación	Entrenamiento	Validación
A	1.59	1.56	1.56	1.54	1.55	1.53	1.52	1.58
B	0.00	4.37	0.00	9.93	0.00	9.09	0.00	4.14
C	0.02	0.06	0.03	0.12	0.03	0.07	0.02	0.09
TOTAL	1.61	5.99	1.59	11.59	1.58	10.69	1.54	5.80

Tabla 7. Términos de error del periodo seco. Red propuesta con 2 neuronas ocultas

ERROR	PRUEBA 1		PRUEBA 2		PRUEBA 3	
	Entrenamiento	Validación	Entrenamiento	Validación	Entrenamiento	Validación
A	1.59	1.69	1.58	1.61	1.56	1.67
B	0.00	3.20	0.00	2.96	0.00	3.31
C	0.02	0.08	0.02	0.05	0.03	0.11
TOTAL	1.61	4.97	1.60	4.62	1.59	5.09

También se observa que los errores A y C en la fase de validación se mantienen del mismo orden que los conseguidos en la fase de entrenamiento, mientras que el error de tipo B, que es prácticamente nulo en la fase de entrenamiento, aparenta un mayor crecimiento. Esta discrepancia de error B no es estadísticamente significativa, al menos en la RNA de dos neuronas ocultas, ya que la media y varianza de los datos generados por la red no difieren de la media y varianza históricas, como se muestra en las tablas 8 y 9 y en el resumen de pruebas estadísticas de la tabla 10.

En las tablas 8 y 9 se pone de manifiesto la capacidad del modelo para reproducir diferentes parámetros de la distribución univariante de la intensidad, tales como la media aritmética (\bar{X}) y la varianza (S^2), que definen el error B, y los valores mínimo y máximo, el coeficiente de asimetría (K_1) y el de curtosis (K_2), que no se utilizaron como entrenamiento. En estas tablas, la mayor discrepancia se observa en la máxima intensidad histórica del periodo húmedo, 104.14 mm h⁻¹, pero este dato es excepcional, ya que los siguientes datos observados en este periodo, en orden decreciente de magnitud, son 73.66, 71.12 y 68.58 mm h⁻¹, respectivamente.

Otro aspecto de la validación del modelo es el estudio de la distribución de frecuencias de los datos estimados por la red. En la figura 3 se muestran las frecuencias acumuladas de los valores

históricos de la intensidad y de los datos de salida en las tres pruebas de validación (val_1, val_2 y val_3). Se observa que las diferencias entre las series de validación y la histórica son pequeñas, aunque algo mayores en el periodo seco, debido a la mayor varianza de los datos históricos, que produce mas variabilidad en las series simuladas. Este resultado es importante para validar la red, porque la distribución de frecuencias univariante no se incluye con tal grado de detalle en la función de error y, aunque alguna serie no pase el test de igualdad de distribución (tabla 10), las diferencias entre la distribución histórica y las generadas por la red, en el caso más desfavorable, son inferiores al 7%.

También se ha comprobado si la red reproduce de forma adecuada la distribución bivariante de frecuencias (figura 4). Comparando cada gráfica de la figura 4 con la correspondiente de la figura 2, de datos históricos, vemos que son muy similares y, de acuerdo con los resultados de la tabla 10, no se rechazaría la igualdad de la distribución bivariante de todas las series de validación y las reales. Este resultado justifica la utilidad del término específico para las frecuencias bivariantes en la función de error.

Aunque ya se han comentado los resultados de los análisis estadísticos recogidos en la tabla 10, conviene puntualizar algunos de los métodos utilizados.

Tabla 8. Parámetros estadísticos de la intensidad media (mm h⁻¹) para el periodo húmedo

	HISTORICOS	PRUEBA 1		PRUEBA 2		PRUEBA 3	
		Entrenamiento	Validación	Entrenamiento	Validación	Entrenamiento	Validación
Mínimo	0.10	0.10	0.10	0.10	0.10	0.10	0.10
Máximo	104.14	71.93	72.64	74.94	72.32	70.94	71.17
\bar{X}	3.75	3.72	3.70	3.75	3.52	3.75	3.58
S^2	35.94	35.94	33.14	35.94	33.13	35.94	36.23
K_1	5.80	5.82	6.14	5.68	5.89	5.88	5.95
K_2	50.42	45.42	52.52	46.29	50.50	48.40	49.31

Tabla 9. Parámetros estadísticos de la intensidad media (mm h⁻¹) para el periodo seco

	HISTORICOS	PRUEBA 1		PRUEBA 2		PRUEBA 3	
		Entrenamiento	Validación	Entrenamiento	Validación	Entrenamiento	Validación
Mínimo	0.13	0.10	0.10	0.10	0.10	0.10	0.10
Máximo	92.61	90.74	79.28	84.43	82.79	97.53	86.47
\bar{X}	8.24	8.24	8.56	8.24	8.21	8.24	8.05
S^2	122.43	122.43	126.94	122.43	118.25	122.43	127.10
K_1	2.96	3.04	2.79	3.07	3.09	3.26	3.09
K_2	11.32	12.34	10.32	11.98	12.24	15.41	12.63

Tabla 10. Análisis estadístico de la validación: test y región de la hipótesis nula con $\alpha=0.05$

HIPÓTESIS nula	Periodo húmedo			Periodo seco		
	val_1	Val_2	Val_3	val_1	val_2	val_3
Igual media $C_0 = (-1.96, 1.96)$	$Z = -0.23$	$Z = -1.02$	$Z = -0.74$	$Z = 0.74$	$Z = -0.08$	$Z = -0.44$
Igual varianza $C_0 = (-1.96, 1.96)$	$Z = -0.28$	$Z = -0.28$	$Z = 0.03$	$Z = 0.27$	$Z = -0.23$	$Z = 0.28$
Igual distribución univariante $C_0 = (0, 16.92)$; 9 grados de libertad	$\chi^2 = 5.44$	$\chi^2 = 17.25$	$\chi^2 = 32.13$	$\chi^2 = 36.91$	$\chi^2 = 7.01$	$\chi^2 = 12.98$
Igual distribución bivariante $C_0 = (0, 15.51)$; 8 grados de libertad	$\chi^2 = 8.29$	$\chi^2 = 3.33$	$\chi^2 = 3.32$	$\chi^2 = 5.90$	$\chi^2 = 4.20$	$\chi^2 = 3.19$

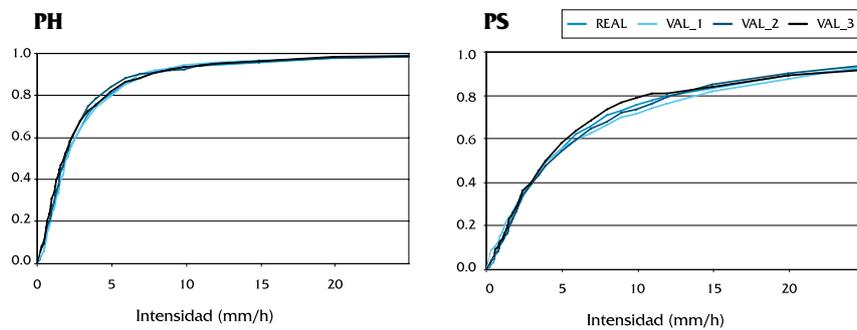


Figura 3. Distribución de frecuencias de la intensidad en el periodo húmedo (PH) y seco (PS)

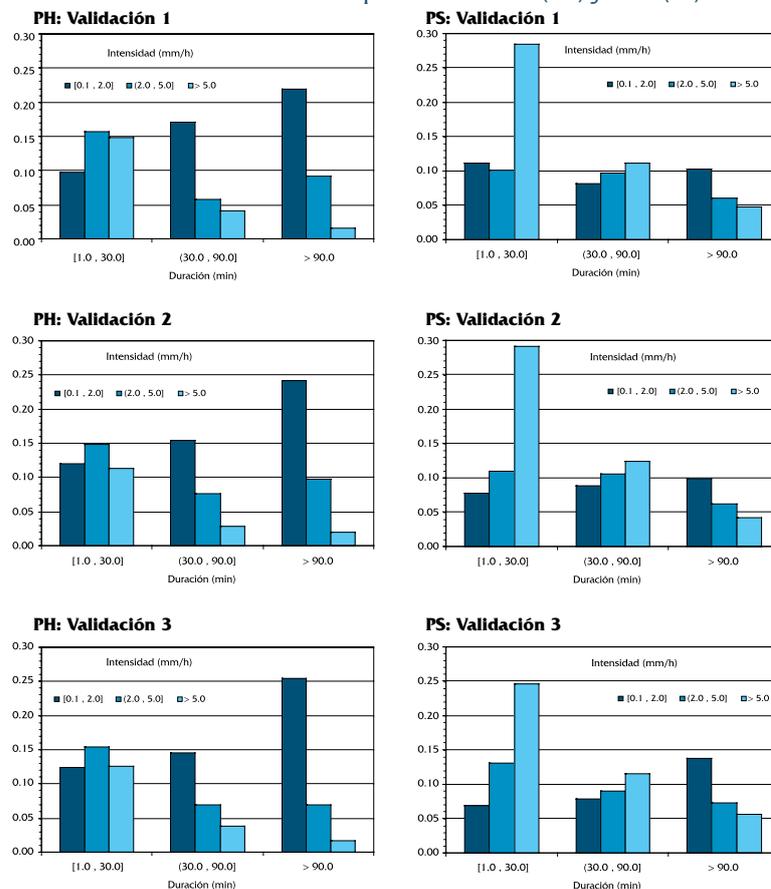


Figura 4. Relación intensidad – duración en el periodo húmedo (PH) y seco (PS)

La comparación de medias y varianzas se ha hecho frente a los correspondientes valores históricos, y se ha utilizado la distribución Normal porque se trata de muestras muy grandes, con 680 datos en cada serie de validación. Por el contrario, los tests de igualdad de distribución se han hecho considerando dos muestras de igual tamaño, la formada por los datos producidos por la red y la correspondiente a los datos reales, a partir de los cuales se calcula el error de validación.

CONCLUSIONES

Los resultados obtenidos al modelar la relación intensidad – duración de la lluvia, nos llevan a proponer el uso de redes neuronales como una herramienta alternativa cuando no se dispone de un modelo estadístico para la función de densidad de probabilidad conjunta de dos series climatológicas relacionadas.

En modelos como el propuesto, con 20 o más entradas, el inconveniente que representa el cálculo de un mayor número de coeficientes en la RNA, queda ampliamente compensado por la generalidad de su uso y por la posibilidad de incorporar directamente, además de las relaciones entre los datos de entrada y los de salida, la estructura aleatoria de los datos de entrada, sin especificar un modelo de probabilidad para los mismos.

Otro aspecto que cabe destacar del proceso de ajuste de pesos propuesto es la facilidad con la que pueden incorporarse a la función de error los términos correspondientes a ciertos coeficientes estadísticos, tales como el de Cramer, el de variación o cualquier otro que pueda derivarse de los momentos estadísticos o de algún test específico. Todos estos términos contribuyen a que los datos producidos por la red, con muy pocas neuronas ocultas, tengan unas propiedades estadísticas previamente especificadas.

Aunque este trabajo no pretende ser un estudio exhaustivo de cada elemento de una RNA, sí queremos señalar algunos de los aspectos que nos parecen más interesantes para conseguir un modelo parsimonioso: a) la elección de las funciones de transferencia, que deben adaptarse a los datos que trata de reproducir la RNA; b) la elección de la función de error, que debe facilitar el que los datos de salida tengan unas determinadas propiedades; c) la elección del método de ajuste de pesos, que debe ser eficiente, y adecuado, para la función de error y funciones de transferencia seleccionadas.

LISTA DE SÍMBOLOS

- A:** error logarítmico
- B:** diferencia entre la media y varianzas reales y estimadas
- C:** coeficiente de Cramer
- d_{pk} :** salida objetivo o deseada de la red para la neurona k y el patrón p
- e :** valor de la función de error
- h_{pj} :** salida de la neurona oculta j para el patrón p
- l :** número de neuronas de entrada
- m :** número de neuronas ocultas
- n :** número de neuronas de salida
- N :** número total de pares de datos de entrenamiento
- P :** número de patrones de entrenamiento
- W :** número total de pesos o parámetros de la red
- w_{ij} :** pesos o parámetros de la red de la capa de entrada a la capa oculta
- w_{jk} :** pesos o parámetros de la red de la capa oculta a la capa de salida
- x_{pi} :** entrada a la red para la neurona i y el patrón p
- y_{pk} :** salida de la red para la neurona k y el patrón p

BIBLIOGRAFÍA

- Alvarez, J., y S. Bolado (1996). Descripción de los procesos de infiltración mediante redes neuronales artificiales. *Ing. Agua*. 3: 39 – 46.
- ASCE task committee on application of artificial neural networks in hydrology (2000). Artificial neural networks in hydrology, I: preliminary concepts. *J. Hydrol. Engng*. 5: 115 – 123.
- Bárdossy, A. (1998). Generating precipitation time series using simulated annealing. *Water Resour. Res.* 34: 1737 – 1744.
- Cancelliere, A., G. Giuliano, A. Ancarani y G. Rossi (2002). A neural networks approach for deriving irrigation reservoir operating rules. *Water Resour. Mgmt.* 16: 71 – 88.
- Cybenko, G. (1989). Approximations by superpositions of a sigmoidal function. *Math. Control, Signals Syst.* 2: 303 – 314.
- Dibike, Y. B., D. Solomatine y M. B. Abbott (1999). On the encapsulation of numerical hydraulic models in artificial neural network. *J. Hydr. Res.* 37: 147 – 161.
- Eagleson, P. S. (1978). Climate, soil and vegetation (2): The distribution of annual precipitation derived from storm sequences. *Water Resour. Res.* 14: 713 – 721.
- Entakhabi, D., I. Rodríguez Iturbe y P. S. Eagleson (1989). Representation of the temporal rainfall process by a modified Neyman Scott rectangular pulse model: Parameter estimation and validation. *Water Resour. Res.* 25: 295 – 302.
- Fine, T. L. (1999). Feedforward neural network methodology. *Ed. Springer – Verlag*.

- Freeman, J. A., y D. M. Skapura (1993). Redes neuronales: algoritmos, aplicaciones y técnicas de programación. Ed. Addison – Wesley.
- French, M. N., W. F. Krajewski y R. R. Cuykendall (1992). Rainfall forecasting in space and time using a neural network. *J. Hydrol.* 137: 1 – 31.
- Funahashi, K. I. (1989). On the approximate realization of continuous mappings by neural networks. *Neural Networks* 2: 183 – 192.
- Goel, N. K., R. S. Kurothe, B. S. Mathur y R. M. Vogel (2000). A derived flood frequency distribution for correlated rainfall intensity and duration. *J. Hydrol.* 228: 56 – 67.
- Govindaraju, R. S., y A. Ramachandra Rao (2000). Artificial neural networks in hydrology. Ed. Kluwer Academic Publishers.
- Hagan, M., y M. Menhaj (1994). Training feedforward networks with the Marquardt algorithm. *IEEE Trans. On Neural Networks* 5: 989 – 993.
- Hecht – Nielsen, R. (1987). Kolmogorov's mapping neural network existence theorem. *Proc. Int. Conf. on Neural Networks* 3: 11 – 13, IEEE Press.
- Hecht – Nielsen, R. (1990). Neurocomputing. Ed. Addison – Wesley.
- Hilera González, J. R., y V. J. Martínez Hernando (1995). Redes neuronales artificiales. Fundamentos, modelos y aplicaciones. Ed. Ra – Ma.
- Himmelblau, D. M. (1972). Applied nonlinear programming. Ed. McGraw – Hill.
- Hornik, K., M. Stinchcombe y H. White (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks* 3: 551 – 560.
- Hsu, K. L., H. V. Gupta y S. Sorooshian (1995). Artificial neural network modeling of the rainfall – runoff process. *Water Resour. Res.* 31: 2517 – 2530.
- Hsu, K. L., H. V. Gupta, X. Gao y S. Sorooshian (1999). Estimation of physical variables from multichannel remotely sensed imagery using a neural network: application to rainfall estimation. *Water Resour. Res.* 35: 1605 – 1618.
- Hutchinson, M. F. (1990). A point rainfall model based on a three state continuous Markov occurrence process. *J. Hydrol.* 114: 125 – 148.
- Istok, J. D. y L. Boersma (1989). A stochastic cluster model for hourly precipitation data. *J. Hydrol.* 106: 257 – 285.
- Johansson, E. M., F. U. Dowla y D. M. Goodman (1992). Backpropagation learning for multilayer feedforward neural networks using the conjugate gradient method. *Int. J. Neural Syst.* 2: 291 – 301.
- Kao, J. J. (1996). Neural net for determining DEM based model drainage pattern. *J. Irrigation and Drainage Engng.* 122: 112 – 121.
- Kolmogorov, A. N. (1957). On the representations of continuous functions of many variables by superpositions of continuous functions of one variable and addition. *Dokl. Akad. Nauk USSR* 114: 953 – 956.
- Kurkova, V. (1991). Kolmogorov's theorem is relevant. *Neural Computation* 3: 617 – 622.
- Leshno, M., V. Lin, A. Pinkus y S. Schocken (1993). Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks* 6: 861 – 867.
- Maier, H. R., y G. C. Dandy (1999). Empirical comparison of various methods for training feedforward neural networks for salinity forecasting. *Water Resour. Res.* 35: 2591 – 2596.
- Martín del Brío, B. y A. Sanz Molina. (1997) Redes neuronales y sistemas borrosos. Ed. Ra – Ma.
- Mason, J. C., R. K. Price y A. Tem' me (1996). A neural network model of rainfall runoff using radial basis functions. *J. Hydr. Res.* 34: 537 – 548.
- Morshed, J., y J. J. Kaluarachchi (1998). Parameter estimation using artificial neural network and genetic algorithm for free product migration and recovery. *Water Resour. Res.* 34: 1101 – 1113.
- Rodríguez Iturbe, I., B. Febres de Powder y J. B. Valdés (1987). Rectangular pulses point processes for rainfall: Analysis of empirical data. *J. Geophysical Res.* 92: 9645 – 9656.
- Rogers, L. L., y F. U. Dowla (1994). Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling. *Water Resour. Res.* 30: 457 – 481.
- Schaap, M.G., y W. Bouten (1996). Modeling water retention curves of sandy soils using neural networks. *Water Resour. Res.* 32: 3033 – 3040.
- Swingler, K. (1996). Applying neural networks: a practical guide. Ed. Morgan Kaufman Publishers, Inc..
- Whitley, R., y T. V. Hromadka II (1999). Approximate confidence intervals for design floods for a single site using a neural network. *Water Resour. Res.* 35: 203 – 209.

AGRADECIMIENTOS

Los autores agradecen al Dr. D. A. Woolhiser por proporcionarles los datos, así como por sus interesantes comentarios sobre el modelado de la lluvia. También agradecen los comentarios de los revisores que han permitido mejorar la presentación de este trabajo.