

4 Método propuesto basado en el Perceptrón Multicapa

En este capítulo se presenta el nuevo enfoque que se propone en esta tesis para la generación de series sintéticas de radiación solar. Está basado en la red neuronal artificial (RNA) perceptrón multicapa (PMC), por lo que inicialmente se realizará una breve descripción de este tipo de red neuronal artificial. En una primera sección se resumen las principales características del PMC, sin particularizarlas para la aplicación que se propone en esta tesis. Posteriormente, en la siguiente sección se hace hincapié en la estructura específica utilizada para generar las series de radiación. Como en prácticamente todos los trabajos de investigación, inicialmente se realizó un determinado prototipo que poco a poco fue evolucionando hasta alcanzar la forma final más elaborada. En este sentido es interesante resumir muy brevemente cómo fue el proceso de gestación de todo este trabajo. Inicialmente se partió de una serie de datos de la localidad de Madrid y sobre estos datos se creó la primera estructura neuronal artificial que generó valores sintéticos horarios de radiación. En la sección 4.2 se describe la preparación de dicha estructura y todo lo relacionado a las primeras pruebas y resultados. La primera conclusión destacable fue que el perceptrón creado era capaz de proporcionar resultados válidos para la componente de tendencia de las series de radiación solar horaria en Madrid.

El siguiente paso fue tratar de aplicar dicho perceptrón a otras zonas con climas diferentes dentro de España y tratar de comprobar si el perceptrón era capaz de universalizar los resultados. Para ello se añadieron nuevos datos a los ya existentes para Madrid, utilizando en este segundo estadio, además, las localidades de Oviedo y Málaga. Como se detallará en la sección 4.3, el PMC presenta una serie de capacidades de universalización que lo hacen especialmente interesante y ventajoso frente a los métodos clásicos.

Puesto que ya se tenía demostrado que el PMC se podía utilizar en cualquier localidad, el siguiente paso fue preparar una red neuronal única para gran cantidad de localidades españolas. En este punto surgió el problema de que la masa de datos con los que trabajar era demasiado grande, lo que resultaba bastante complejo de manejar. Una nueva idea surgió y fue la utilización del año meteorológico típico (AMT), que será descrito con detalle más adelante, con el fin de concentrar los datos a manejar. De

nuevo el PMC proporcionó una mejora en el entrenamiento y unos resultados totalmente satisfactorios que serán detallados en la sección 4.4.

Sin embargo, a pesar de los excelentes resultados hasta aquí obtenidos había un pequeño inconveniente y era el hecho de que únicamente se obtenía la componente de tendencia de las series. A continuación se realizó un estudio exhaustivo de la componente aleatoria, el cual condujo a la incorporación de nuevas redes neuronales para el cómputo de esta componente, que también fueron diseñadas por medio del PMC, con resultados acordes a los de los métodos clásicos. Este proceso será expuesto en la sección 4.5.

Por último, quedaba pendiente la generación de series solares a escala diaria, problema que fue resuelto y se describe en la sección 4.6.

4.1.- El Perceptrón multicapa

De entre los diferentes modelos de RNA que existen, en esta tesis se ha utilizado el llamado perceptrón multicapa (PMC), por lo que en el presente capítulo se describe dicha red neuronal artificial.

4.1.1.- El combinador lineal adaptativo

El combinador lineal adaptativo o “neurona artificial” consiste en un elemento cuya salida es una combinación lineal de sus entradas (figura 4.1). En cada instante k , la salida lineal se calcula como suma ponderada de las entradas $x_k = [x_0, x_{1k}, x_{2k}, \dots, x_{nk}]^T$, es decir, la salida es igual al producto escalar $s_k = x_k^T w_k$. El conjunto de coeficientes $w_k = [w_{0k}, w_{1k}, w_{2k}, \dots, w_{nk}]^T$ se denomina *vector de pesos*. Las componentes del vector de entrada, pueden ser valores analógicos continuos o valores binarios. Los pesos son variables reales que pueden tomar valores tanto positivos como negativos.

Durante el proceso de entrenamiento (aprendizaje), se presentan al combinador lineal los patrones de entrada, x_k , y las respuestas deseadas correspondientes, d_k . El algoritmo de aprendizaje ajusta los pesos de forma que las salidas se aproximen tanto como sea posible a los valores deseados. Un método muy empleado para adaptar los pesos es el algoritmo de mínimos cuadrados (α -LMS, “Least Mean Square”) también denominado regla delta de Widrow-Hoff.

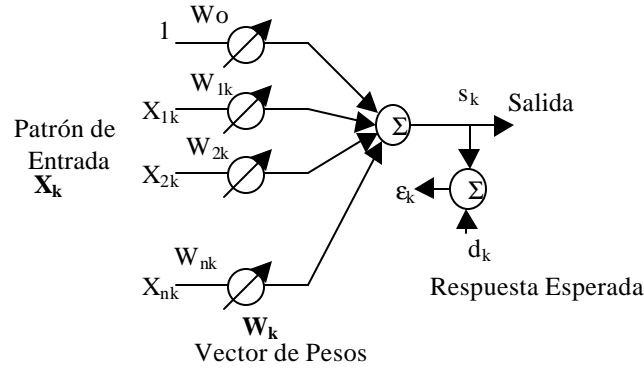


Figura 4.1. Combinador Lineal Adaptativo

Este algoritmo minimiza la suma de los cuadrados de los errores lineales del conjunto de entrenamiento. El error lineal ϵ_k se define como la diferencia entre la respuesta deseada y la salida en el instante k , o sea $\epsilon_k = d_k - s_k$. Esta señal de error es necesaria para adaptar los pesos.

4.1.2.- El perceptrón

El elemento constitutivo básico utilizado en muchas redes neuronales, es el *Perceptrón* [1]. Consiste en un combinador lineal adaptativo seguido de un elemento no lineal (figura 4.2). En esa figura el error lineal es ϵ_k y el error en la salida no lineal es ϵ_k^* , siendo sus expresiones:

$$\epsilon_k = d_k - s_k, \quad \epsilon_k^* = d_k - y_k.$$

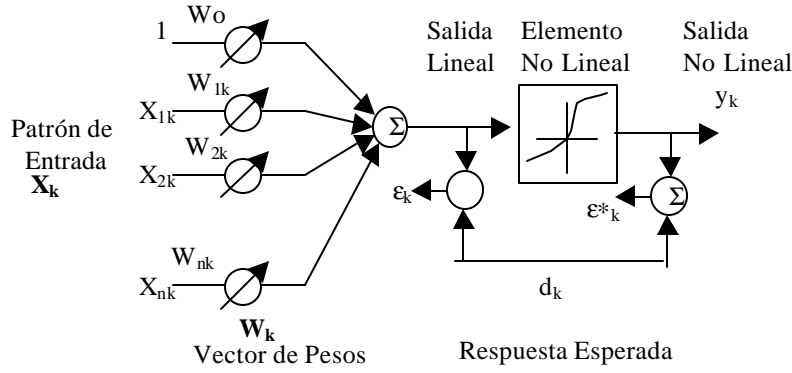
Existen tres tipos comunes de no linealidades: los cuantizadores (función signo), los elementos con umbrales de saturación y las no linealidades de tipo sigmoideal (figura. 4.2). El peso w_o , conectado a una entrada constante $x_o = 1$, controla el nivel de umbral del elemento no lineal.

Separación Lineal: Con n entradas binarias y una salida, un perceptrón como el de la figura 4.2 es capaz, tan solo, de realizar un subconjunto de todas las posibles funciones lógicas: las funciones linealmente separables. El perceptrón separa el espacio de patrones de entrada en dos categorías, dependiendo dicha separación de los valores de los pesos. Para un perceptrón de dos entradas, la condición umbral crítica se produce cuando la salida lineal s se iguala a cero:

$$s = x_1 w_1 + x_2 w_2 + w_o = 0. \quad (4.1)$$

Por tanto:

$$x_2 = w_1 / w_2 \cdot x_1 - w_o / w_2. \quad (4.2)$$



Tipos de no linealidades

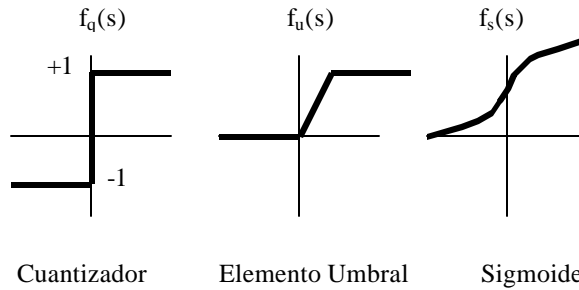


Figura. 4.2 Perceptrón y no-linealidades

En la figura 4.3. se muestra esta relación lineal para un perceptrón con un elemento no lineal de tipo signo. Los tres pesos determinan la pendiente, y la ordenada en el origen, y los semiplanos, a ambos lados de la línea de separación, corresponden a salidas positivas y negativas. Para un perceptrón con tres entradas, la frontera de separación es un plano; con más de tres entradas la frontera es un hiperplano. Si el peso de polarización es cero, el hiperplano es homogéneo, es decir, pasará por el origen del espacio de patrones de entrada. Como se muestra en la figura 4.3, los patrones de entrada binarios se clasifican como sigue: $(+1, +1) \rightarrow +1$, $(+1, -1) \rightarrow +1$, $(-1, -1) \rightarrow +1$, y $(-1, +1) \rightarrow -1$. Esta función constituye un ejemplo de función linealmente separable.

Un ejemplo de función que no es linealmente separable es la función exclusiva (*xor*); no existe ninguna línea recta que pueda lograr la separación de los patrones de entrada, de forma que un simple perceptrón no puede realizar la función XOR.

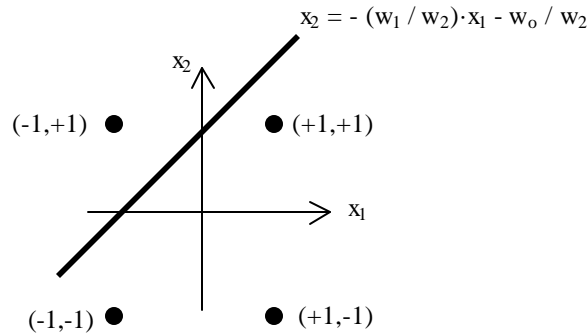


Figura. 4.3 Línea de separación en el espacio de los patrones

Los pesos de las conexiones y el umbral (w_0) del perceptrón pueden adaptarse utilizando diversos algoritmos. El algoritmo de convergencia del perceptrón, o regla del perceptrón, debido a Rosenblatt [2], utiliza el error de cuantización, definido como la diferencia entre la respuesta deseada y la salida del elemento no lineal, $\varepsilon_k^* = d_k - s_k$. Los pesos sólo se adaptan cuando el error es distinto de cero. Si la salida es diferente de la deseada, se efectúa la adaptación de los pesos, de forma que

$$w_{k+1} = w_k + \mathbf{h}(d_k - y_k) x_k. \quad (4.3)$$

Rosenblatt demostró que si el espacio de patrones de entrada es linealmente separable en dos clases, la regla del perceptrón converge, y sitúa el hiperplano de decisión entre ambas clases.

Entre las diversas variantes propuestas a la regla del perceptrón, la más extendida introduce un segundo término en la ecuación de actualización de los pesos, de forma que

$$w_{k+1} = w_k + \varepsilon_k^* \mathbf{h}(d_k - y_k) x_k + \eta (w_k - w_{k-1}) \quad (4.4)$$

Al término dado por $\eta (w_k - w_{k-1})$, se le denomina término de *momento*. Intuitivamente, si el cambio de los pesos en la iteración previa fue grande, al añadir una fracción de esta cantidad al peso actual, se acelerará el proceso de descenso del algoritmo hacia el punto de convergencia.

También puede emplearse el algoritmo α -LMS o regla delta para el entrenamiento del perceptrón, sustituyendo en las ecuaciones (4.3) ó (4.4) el error de cuantización ε_k^* por el error lineal ε_k . No existe, sin embargo, una garantía de

convergencia para este algoritmo, si bien en el caso de clases no separables linealmente, suele conducir a soluciones con menor error. Por el contrario, el algoritmo del perceptrón continúa indefinidamente si los patrones no son linealmente separables, y no produce, necesariamente, soluciones con errores pequeños, aún en el caso de que existan.

4.1.3.- El perceptrón multicapa

Minsky y Papert (1969) [3], realizaron un estudio detallado de los tipos de representaciones posibles con el perceptrón, y demostraron que, en un gran número de casos, estos tipos de redes son incapaces de resolver los problemas de clasificación. Por otro lado, también demostraron que un perceptrón constituido por varias capas, puede realizar cualquier aplicación desde las capas de entrada a las de salida. El inconveniente que se tenía en estos primeros estadios del manejo del perceptrón, era el desconocimiento de un algoritmo de entrenamiento eficiente para redes con más de una capa.

El problema de la separación lineal limita la capacidad de representación del perceptrón, ya que, como se muestra en la tabla 4.1, la probabilidad de que una función lógica sea linealmente separable es muy pequeña, incluso para un número moderado de variables. Por esta razón el perceptrón de una sola capa oculta está limitado, en la práctica, a problemas muy sencillos.

n	Nº Funciones	Funciones linealmente separables
1	4	4
2	16	14
3	256	104
4	65.536	1.882
5	$4.3 \cdot 10^9$	94.572
6	$1.8 \cdot 10^{19}$	5.028.134

Tabla 4.1. Funciones linealmente separables.

El perceptrón multicapa es una red acíclica de propagación directa con una o más capas de nodos entre las entradas y las salidas. Estas capas adicionales contienen

nodos denominados *ocultos*, ya que no son visibles directamente ni desde las entradas, ni desde las salidas.

La capacidad de computación del perceptrón multicapa procede de las funciones de activación no lineales empleadas en los nodos. La salida de una capa lineal consiste en el producto del vector de entrada por el vector de pesos entrantes en esa capa, es decir, $y_I = W_I x_I$. Si a continuación se conecta en cascada otra capa también lineal, cuyas entradas sean las salidas de la capa anterior, la salida de esta segunda capa será $y_2 = W_2 y_I = W_2 W_I x_I = W x_I$. Por tanto, una red de dos (o más) capas lineales, equivale a otra de una sola capa, cuya matriz de pesos sea igual al producto de las dos (o más) matrices de pesos de partida. Como se ha visto en el apartado anterior, las redes de una sola capa están muy limitadas en su capacidad de computación; por tanto, las funciones de activación no lineales son imprescindibles para extender la capacidad de las redes.

Las posibilidades del perceptrón multicapa pueden considerarse desde tres perspectivas diferentes. La primera está relacionada con su capacidad para realizar expresiones booleanas, la segunda, con su capacidad para dividir el espacio de patrones en los problemas de clasificación, y la tercera con su capacidad para realizar transformaciones no lineales en problemas de aproximaciones funcionales.

Un perceptrón puede considerarse como una unidad básica de lógica binaria que es capaz de realizar todas las funciones booleanas excepto las funciones xor y xnor. Conectando en cascada varios perceptrones de forma adecuada, pueden llevarse a cabo expresiones booleanas arbitrarias. En el campo de las redes neuronales el interés se centra no tanto en realizar expresiones booleanas conocidas como en aprender dichas expresiones partiendo de ejemplos de entradas/salidas. Dado que son suficientes dos niveles de puertas lógicas para expresar funciones booleanas arbitrarias, no son necesarias más que dos capas de perceptrones para, potencialmente, poder aprender dichas funciones.

También ha sido demostrado que un perceptrón de dos capas con funciones de activación de tipo sigmoideal, es suficiente para realizar un mapa no lineal continuo arbitrario. A pesar de estos resultados, para algunos problemas, una red de tres capas puede requerir una cantidad mucho menor de nodos y pesos que una red con sólo una capa oculta.

4.1.3.1.- El algoritmo de aprendizaje. Retropropagación

Una de las principales limitaciones de la formulación original de Rosenblatt, fue la falta de un algoritmo de aprendizaje adecuado para el perceptrón multicapa. La publicación de la técnica de Retropropagación (“*Backpropagation*”) por Werbos [4], perfeccionada posteriormente por Rumerhalt [5], supuso un hito clave en el resurgimiento de las redes neuronales, sumidas en una fase de letargo casi total desde la publicación del trabajo de Minsky y Papert.

La dificultad principal para la deducción del algoritmo de aprendizaje del PMC estriba en que el gradiente del error es cero prácticamente en todo el espacio, cuando se utiliza como función de activación no lineal el cuantizador. Dado que la investigación inicial en el campo de las redes neuronales trataba casi exclusivamente con este tipo de no linealidad, la idea de reemplazarla no surgió durante casi dos décadas. La no linealidad debe, no obstante, estar presente, pues como se ha visto, sin ella el PMC no realizaría más que una transformación lineal en cada capa, en cuyo caso podría reducirse a una red equivalente de una sola capa. La solución consiste en utilizar una función de activación diferenciable, con la que es posible realizar una búsqueda de descenso del gradiente en el espacio de los pesos.

Las funciones de activación más utilizadas son las denominadas semilineales. Una función de activación semilineal es una función no decreciente y diferenciable aplicada sobre la salida lineal del nodo, es decir

$$y_{pj} = f(s_{pj}) \quad (4.5)$$

en donde

$$s_{pj} = \sum w_{ji} y_{pi} \quad (4.6)$$

es la salida lineal del nodo j , para el patrón de entrada p (si el nodo i es una entrada de la red, entonces $y_{pi} = x_{pi}$).

Las funciones de activación semilineales empleadas con mayor frecuencia son las funciones sigmoideal y tangente hiperbólica:

$$\text{sgm}(x) = 1 / (1 + e^{-x}), \quad (4.7a)$$

$$\tanh(x) = (1 - e^{-2x}) / (1 + e^{-2x}). \quad (4.7b)$$

Ambas funciones tienen la misma forma, pero difieren en sus valores de saturación de salida, y en la pendiente y amplitud del intervalo en el cual toman valores “quasi-lineales”.

El algoritmo de retropropagación es una generalización de la regla delta, que ajusta los pesos en la dirección opuesta al gradiente del error de cada patrón de entrenamiento. El cambio en los pesos, para cada presentación de un patrón de entrada viene dado por la expresión

$$\Delta_p w_{ji} \propto -\partial E_p / \partial w_{ji}, \quad (4.8)$$

en donde E es la función de error, igual a la suma de los cuadrados de las diferencias entre los valores deseados y los valores de las neuronas de la capa de salida, para un patrón p :

$$E_p = 1 / 2 \sum (d_{pj} - y_{pj})^2. \quad (4.9)$$

En el cálculo de la derivada, se aplica la regla de la cadena para expresarla como producto de dos términos: uno que refleja el cambio en el error, en función de las variaciones en la salida lineal de los nodos, y otro que representa el efecto de las variaciones de los pesos sobre dicha salida lineal. Por tanto, se puede escribir:

$$\partial E_p / \partial w_{ji} = (\partial E_p / \partial s_{pj}) \cdot (\partial s_{pj} / \partial w_{ji}). \quad (4.10)$$

De la ecuación (4.5) se obtiene que el segundo factor vale

$$\partial s_{pj} / \partial w_{ji} = \partial (\sum w_{jk} y_{pk}) / \partial w_{ji} = y_{pk}. \quad (4.11)$$

Definiendo

$$\delta_{pj} = -\partial E_p / \partial s_{pj}. \quad (4.12)$$

Se puede escribir la ecuación (4.11) como

$$-\partial E_p / \partial w_{ji} = \delta_{pj} y_{pi}. \quad (4.13)$$

Lo que indica que para realizar un descenso del gradiente en E , se deben actualizar los pesos de acuerdo con

$$\Delta_p w_{ji} = \eta \delta_{pj} y_{pi}. \quad (4.14)$$

Tan solo falta por calcular el valor de δ_{pj} para cada unidad de la red. El resultado de Rumelhart et al., consiste en la obtención de una regla recurrente para el cálculo de este factor, que puede realizarse propagando las señales de error, desde la salida hacia la entrada, a través de la red. Para el cálculo de (4.12), se aplica nuevamente la regla de la cadena, escribiendo la derivada parcial como producto de dos factores. El primero expresa las variaciones en el error en función de las variaciones en la salida del nodo. El segundo, el cambio en la salida no lineal, en función de los incrementos o decrementos de valor de la salida lineal de dicho nodo. Por tanto se tendrá.

$$\delta_{pj} = -\partial E_p / \partial s_{pj} = -(\partial E_p / \partial y_{pj}) \cdot (\partial y_{pj} / \partial s_{pj}). \quad (4.15)$$

El segundo factor, por la ecuación (4.6), vale

$$\partial y_{pj} / \partial s_{pj} = f'(s_{pj}). \quad (4.16)$$

que es simplemente la derivada de la función semilineal para la unidad j , evaluada en el valor de salida lineal s_{pj} de dicha unidad. Para calcular el primer factor, se consideran dos casos. En primer lugar, supongamos que la unidad j es un nodo de salida de la red. En este caso, de la definición de E_p se tiene que

$$\partial E_p / \partial y_{pj} = -(d_{pj} - y_{pj}), \quad (4.17)$$

y sustituyendo en la ecuación (4.15) se obtiene

$$\delta_{pj} = f'(s_{pj}) (d_{pj} - y_{pj}), \quad (4.18)$$

para cualquier neurona j de salida. Si el nodo j no es de salida, utilizando la regla de la cadena se tiene

$$\sum (\partial E_p / \partial s_{pk}) (\partial s_{pk} / \partial y_{pj}) = -\sum \delta_{pk} w_{kj}. \quad (4.19)$$

En este caso, sustituyendo nuevamente en la ecuación (4.15) se obtiene

$$\delta_{pj} = f'(s_{pj}) \sum \delta_{pk} w_{kj}, \quad (4.20)$$

cuando j no es una unidad de salida. Las ecuaciones (4.18) y (4.20) dan un procedimiento recurrente para el cálculo de δ_{pj} para cada una de las neuronas (j) de la red, con cada patrón de entrenamiento p . Estos términos se utilizan a su vez para el cálculo de los nuevos pesos, de acuerdo con la ecuación (4.14).

Considerando el hecho de que

$$\partial E / \partial w_{ji} = \sum (\partial E_p / \partial w_{ji}) \quad (4.21)$$

se concluye que el cambio neto en w_{ji} después de un ciclo completo de presentación de los patrones, denominado *época*, es proporcional a esta derivada y por tanto el método de retropropagación realiza un descenso del gradiente en E . De hecho, esto es estrictamente cierto solamente si los valores de los pesos no se cambian durante este ciclo. Si se actualizan los pesos después de la presentación de cada patrón, nos apartamos en cierta medida del descenso por el gradiente en E . No obstante, supuesto que la tasa de aprendizaje (es decir la constante de proporcionalidad η) sea suficientemente pequeña, la desviación es despreciable. Este esquema de actualización de los pesos se conoce como *modo secuencial*, mientras que el esquema clásico del descenso del gradiente se denomina *modo "batch"*.

La aplicación del método de retropropagación supone dos fases. Durante la primera fase, se presenta en las entradas de la red un patrón y se propaga hacia adelante

a través de la red, computando la salida de cada unidad y_{pj} . Esta salida se compara con los valores deseados, resultando una señal de error para cada unidad de salida. La segunda fase supone una propagación hacia atrás a través de la red (análoga a la propagación directa inicial), durante la que se propaga la señal de error y se modifican los pesos. La propagación inversa tiene la misma complejidad computacional que la directa. En esencia, el método de retropropagación (en modo *batch*) consiste por tanto en una regla eficiente para el cálculo del gradiente de la función de error.

4.1.3.2.- Capacidades y limitaciones del perceptrón multicapa

El PMC es capaz de realizar mapas no lineales arbitrarios; además, dado un conjunto de ejemplos, el algoritmo de retropropagación es capaz de aprender a representar dichos mapas en los puntos de ejemplo. Existen sin embargo otros aspectos de tipo práctico que deben considerarse. Uno de dichos aspectos es la complejidad del aprendizaje, esto es, la posibilidad de aprender el mapa deseado en una cantidad razonable de tiempo.

Se debe considerar también la capacidad de la red para generalizar, es decir, la habilidad para producir respuestas correctas para patrones fuera del conjunto de entrenamiento. En los siguientes párrafos se consideran estas y otras cuestiones.

En primer lugar, los resultados teóricos indican que el PMC es capaz de representar mapas no lineales arbitrarios, pero esto sólo es cierto si el tamaño de la red puede crecer arbitrariamente. Normalmente, no se conoce cuál es el tamaño adecuado de la red para un problema determinado, si bien no suelen ser necesarias más de tres capas (dos capas ocultas). Seleccionar el número de nodos en cada capa no es tampoco una tarea trivial, y requiere un conocimiento específico del problema, a pesar de que se han establecido ciertas reglas para clases de problemas específicos.

Otra cuestión fundamental es la elección de los parámetros para el algoritmo de aprendizaje. Como se ha visto en el apartado anterior, uno de los parámetros clave es el paso de aprendizaje, cuyo valor óptimo depende del problema en consideración. No obstante, las mayores tasas de convergencia suelen lograrse para valores grandes tanto de η (siempre que el aumento en dicho valor no produzca inestabilidades) como el momento μ . La elección de los pesos iniciales es también importante, ya que la función de error para el PMC puede tener varios mínimos locales, y el punto final de

convergencia depende de las condiciones iniciales. Típicamente, dicho problema se resuelve aplicando el algoritmo de retropropagación varias veces con diferentes pesos iniciales hasta que se encuentra una solución adecuada.

En caso de que sea posible determinar el tamaño óptimo de la red y los parámetros de aprendizaje, el encontrar los pesos de la red que minimizan el error, sigue siendo un problema inherentemente difícil. El problema de encontrar el conjunto de pesos que representa el mapa buscado de forma correcta para un conjunto de entrenamiento, es un problema *NP-completo* [6]. Esto implica que si el problema es muy grande no es probable que sea resuelto en un tiempo razonable.

Por otro lado, a medida que se entrena la red, los pesos pueden tomar valores muy grandes, lo que puede provocar que la mayoría de las neuronas operen en sus regiones de saturación, donde las derivadas de la función de activación son muy pequeñas. Dado que el error retropropagado para el aprendizaje es proporcional a esta derivada, el proceso de entrenamiento puede detenerse virtualmente. Este fenómeno se conoce con el nombre de *parálisis* de la red.

Una última consideración respecto al PMC es que normalmente son necesarios una gran cantidad de patrones de entrenamiento para que la solución aprendida generalice bien nuevos datos. Parece que el número de patrones de entrenamiento necesarios para una buena generalización es linealmente proporcional al número de pesos (aproximadamente unas diez veces mayor). Dependiendo de las dimensiones del problema y el tamaño de la red, el número de pesos puede ser muy grande, lo que puede hacer aumentar en gran medida el tiempo necesario para aprender la solución.

A pesar de las limitaciones descritas en los párrafos procedentes, el perceptrón multicapa ha sido aplicado con éxito a cierto número de problemas reales. Como se detalló en un capítulo anterior, éstos incluyen procesamiento de señales sonar, control, reconocimiento de caracteres manuscritos, medicina, finanzas, etc. El perceptrón multicapa y sus extensiones proporcionan una potente herramienta para la resolución de problemas de procesamiento de información compleja.

4.2.-Aplicación del PMC en la generación de series sintéticas

Los métodos clásicos para generar secuencias artificiales de radiación solar reciben principalmente dos críticas: la elevada carga computacional que se requiere para

el cálculo de cada valor de la secuencia o serie y la dependencia del método con el lugar donde se han tomado los registros históricos para realizar el modelo. Sin embargo con la utilización del modelo de red neuronal artificial perceptrón multicapa (PMC) se subsanan esas deficiencias que tienen los métodos clásicos, como se verá a continuación.

La característica principal que hace a este modelo de red particularmente atractivo y útil para solventar el problema de la generación series temporales de radiación solar es su capacidad de aproximar cualquier tipo de función continua definida en un conjunto compacto, dentro de un margen de error preestablecido.

Otra característica importante del PMC es que la forma de utilizarlo es altamente independiente del problema bajo consideración. Ello permite desarrollar metodologías sencillas de diseño con aplicabilidad general. En el campo de la predicción de series temporales ha sido aplicado con resultados satisfactorios debido a su universalidad para la aproximación de funciones.

Esta sección se divide en tres subsecciones en las que se describe la estructura del PMC utilizado junto con la descripción de las diferentes entradas de que consta, como se verá en los apartados 4.2.1 y 4.2.2. A su vez, en el apartado 4.2.3, se expone cómo se llegó a esa estructura por medio de la inclusión de la información sucesivamente en diferentes etapas.

4.2.1.-Estructura utilizada. Predicción por evolución de la red

El método desarrollado para la generación de series horarias de radiación solar se basa en estudios previos de aplicación de las redes neuronales en la predicción de series temporales (PST) [7,8,9,10], que han sido utilizados satisfactoriamente. Generalmente, dada una serie temporal con n términos $\{s_n\}$, entre ellos existirán ciertas dependencias. Por ejemplo, puede existir una función G que relaciona cada valor de la serie con los p valores precedentes:

$$s_{n+1} = G [s_{n-p+1}, \dots, s_n].$$

La PST mediante PMC consiste en dada esa serie $\{s_n\}$ asumir la existencia de la función G , y tratar de aproximarla:

$$s_{n+1} = G [s_{n-p+1}, \dots, s_n] \cong \text{PMC}[s_{n-p+1}, \dots, s_n].$$

Si se dispone de conjunto de datos suficientemente representativo (los registros históricos de radiación solar de la localidad) se podrá entrenar un PMC de p entradas y 1 salida, y éste encontrará, caso de que exista, la relación buscada, aproximando la función G . Una vez que el PMC esté entrenado, para calcular futuros valores de la serie, es decir, generar uno o varios años sintéticos o artificiales, basta con realimentar las predicciones según se van produciendo. Este método se denomina predicción por evolución de la red, y su estructura puede verse en la figura 4.4.

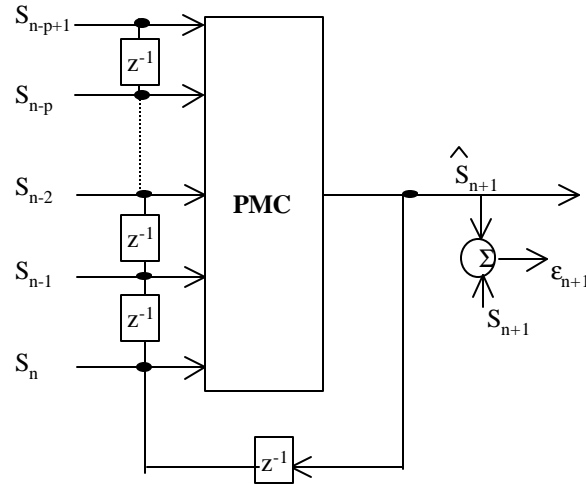


Figura 4.4. Predicción por evolución de la red.

Un método de generación de series de radiación basado en el PMC tiene la cualidad de que la mayor parte de la carga computacional recae en el entrenamiento y no en la generación de valores de las series. Además, una vez desarrollado un método a partir de datos históricos de una localidad determinada se podrá aplicar en distintas localidades sin más que repetir el entrenamiento con los datos históricos correspondiente a la nueva localidad.

Para el estudio inicial sobre el comportamiento de este tipo de arquitectura se dispuso de un conjunto de valores horarios de radiación (en particular el índice de claridad horario k_{th}) medidos en Madrid entre 1978 y 1986. Dichos datos corresponden a un total de 9 años x 365 (días por año) x 16 (horas medidas por día, las 4 primeras horas de cada día, a partir de las 12 de la noche y las 4 últimas siempre son de radiación 0, por lo que no se tienen en cuenta) y sus respectivos 9 x 365 valores del índice diario K_T . Esos datos se emplean de la siguiente manera: los 8 primeros años de datos se utilizan para entrenar el PMC y el último año se usa para validar el método. Como

medida de validez se utilizó el parámetro Varianza Relativa Media (VRM) que cuantifica el error relativo y se emplea con asiduidad en el campo del tratamiento de señal. La VRM se define como el cociente entre la potencia de la señal de error de predicción y la potencia de alterna de la señal a predecir:

$$VRM = \sum_h (k_{th} - k_{thp})^2 / \sum_h (k_{th} - k_{thm})^2$$

siendo:

- k_{th} el valor del índice de claridad real,
- k_{thp} el índice de claridad predicho (generado por el PMC) y
- k_{thm} el valor medio del índice de claridad.

Un valor cercano a 0 indicaría que la serie predicha se acercaría en gran medida a los valores reales de partida. Valores altos indicarían lo contrario, una mala aproximación.

4.2.2.- Descripción de las entradas

Aunque será en la próxima sección donde se describa el porqué de la estructura utilizada, se adelanta en esta sección la forma del PMC utilizado. Dicha estructura del PMC consta de una primera capa o capa de entrada con 7 entradas, una capa oculta con 15 nodos o neuronas y una última capa de salida con un único nodo de salida. En la figura 4.5 puede verse dicha estructura:

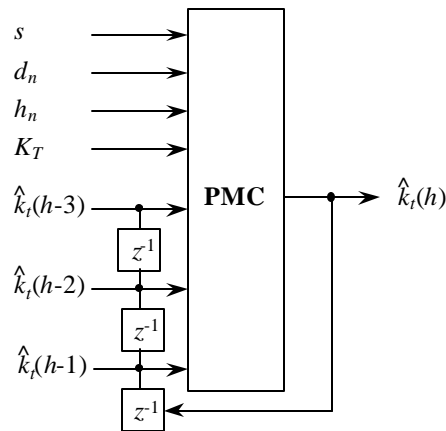


Figura 4.5. PMC Arquitectura para la predicción de los índices de claridad horarios.

Las 7 entradas del PMC son:

d_n : distancia en días entre el día correspondiente al valor a generar y el día con valor máximo en la distribución anual de $\{k_{th}\}$ Está normalizado en el rango $[0,1]$:

$$d_n = 1 - |Nd-183| / 183$$

siendo Nd el orden del día en el año (1 para el día 1 de enero y 365 para el día 31 de diciembre).

h_n : indica el numero de orden de la hora del valor de k_{th} . Normalizado entre $[0,1]$:

$$h_n = (\text{hora}-p) / (16-p)$$

siendo hora un valor que varía entre 4 (primera hora a generar) y 16 (última hora a generar), y p el tamaño de predicción, es decir, el número de valores precedentes que se muestran, 3 en este caso.

K_T : índice de claridad diario.

$k_{th-1}, k_{th-2}, k_{th-3}$: índices de claridad horario de las tres horas precedentes al valor a generar. El porqué de la elección de las tres horas previas se explicará más adelante.

s : indica si la hora está entre la salida del sol y la puesta del mismo. Toma solamente los valores 0 ó 1, para indicar si ha salido ya el sol (1) ó no ha salido todavía (0).

El entrenamiento del PMC se realiza usando el algoritmo de retropropagación.

4.2.3.- Inclusión de la información paso a paso

Como ya se ha adelantado en apartados anteriores, una de las grandes ventajas que conlleva el uso de redes neuronales es que, aunque a priori no se conozcan explícitamente las relaciones entre las variables del sistema, se le puede ir añadiendo a la red información asociada al problema en estudio paulatinamente y de una manera muy sencilla, y ésta cada vez va aproximando mejor sus resultados a la solución que se busque. Como se describirá a continuación, éste fue el proceso seguido hasta alcanzar la estructura que se acaba de presentar en el apartado 4.2.2. Cada una de las nuevas informaciones que va recibiendo el PMC son consideradas como etapas sucesivas en las que se van mejorando los resultados. Estas etapas se describen en los apartados siguientes.

Secuencia numérica

En un primer estadio se consideró toda la serie de datos (9x365x16 valores) como una secuencia numérica, mostrada aleatoriamente, pero sin añadir ningún tipo más de información. Como ya ha sido mencionado, se utilizaron los 8 primeros años

para entrenar la red y el noveno para contrastar con los valores predichos. En esta primera etapa la topología del PMC fue de 20-4-1, 20 neuronas de entrada, 4 en la capa oculta y 1 neurona de salida. Así mismo, se probaron dos versiones diferentes en cuanto a la función utilizada en la última capa. Una versión con función de transferencia lineal y otra versión con función sigmoideal. Los resultados en este caso no fueron satisfactorios, principalmente debido a las siguientes causas:

- La utilización del método de evolución de la red durante un número muy elevado de valores provoca una acumulación de errores, debida a la realimentación. Ya se verá más adelante que en las versiones finales del PMC propuesto en esta tesis, se ha disminuido este número tan elevado de valores, utilizando el llamado Año Meteorológico Típico.
- El método carece de ciertas informaciones esenciales como son el conocimiento del día del año en el que se está produciendo la generación, así como sobre la hora del día correspondiente al valor que se va a generar, por lo que no es posible generar una serie que siga las tendencias estacionales mensuales y horarias de un año real.

Índices de claridad previos. Información diaria

Para tratar de solventar las deficiencias vistas en la etapa anterior, se decidió incluir como información algunos valores precedentes del índice de claridad horario al valor a generar. El hecho de utilizar algunos valores del índice de claridad horario de horas previas al valor a generar, se debe a la propiedad de persistencia radiativa, ya explicada en el capítulo 3. Gracias a muchos estudios realizados por diferentes investigadores, se sabe que existe cierta dependencia entre la radiación solar global recibida durante una determinada hora, y varias horas previas. La dependencia es mayor con la hora anterior, y comienza a disminuir conforme las horas previas a la hora en estudio están más alejadas en el tiempo. También según estudios de autocorrelación de las series de radiación solar, que se mostrarán en el capítulo 5, se decidió incluir los tres valores del índice de claridad horario previos al valor a generar. Esto en principio presentaba un pequeño inconveniente y es que a la hora de la generación, era necesario siempre usar los índices de claridad horarios de las tres primeras horas del día en cuestión. Sin embargo este inconveniente es mínimo, puesto que en esos tres primeros

valores suelen ser cero o próximos a cero (todavía no ha salido o acaba de salir el sol para esas horas), por lo que tomando esas tres entradas como nulas en la generación no modificaban en gran medida los resultados finales (como posteriormente se mostrará).

También, para que la serie generada tuviera en cuenta la estacionalidad mensual de las series reales de radiación, se añadió una nueva entrada que indica la distancia en días entre el valor a generar y el día correspondiente al máximo de la curva de distribución anual de la serie de índices de claridad. Esa entrada recibe el nombre de día normalizado y su expresión es la indicada en el apartado 4.2.2.

Con el fin de ir mostrando los resultados que se obtienen en la generación de series sintéticas de radiación utilizando las diferentes estructuras neuronales, y observar el grado de similitud con las series reales, se muestra a continuación una figura en la que se ve la evolución del índice de claridad horario para varios días reales.

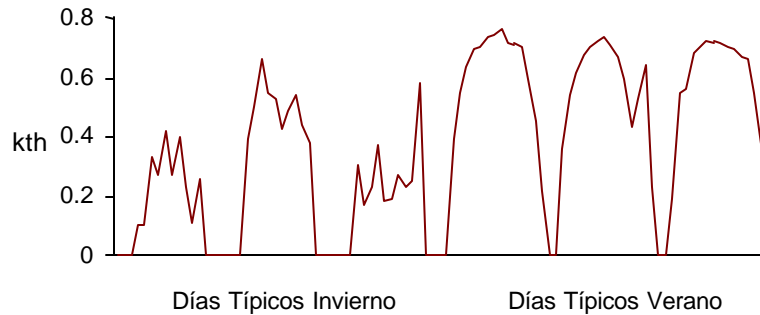


Figura 4.6. Días típicos reales de invierno y verano.

En este caso los resultados fueron bastante mejores que los obtenidos con la primera arquitectura, como se puede observar en la figura 4.7., pero se siguen observando ciertos defectos, como:

- La serie sintética presenta una simetría con respecto a un valor central que no es estrictamente cierta en las series de valores horarios de índice de claridad k_{th} .
- Al asumir la dependencia con el día normalizado, el nivel medio de los días generados crece de forma uniforme hasta el día central, disminuyendo de nuevo a partir de él hasta final de año. La forma de la distribución que se produce no es totalmente igual a la distribución real.

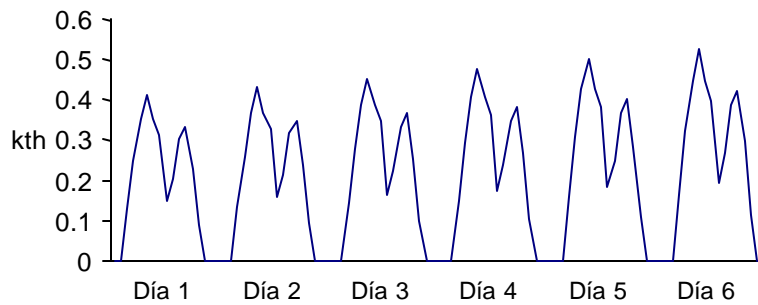


Figura 4.7. Serie generada al incorporar la información diaria.

Índice de claridad diario

La nueva entrada que se añadió fue la inclusión del índice de claridad diario. Para ello es necesario disponer de esos valores del año artificial que se quiera generar. Esta nueva información hace disminuir el valor de VRM que se utiliza como factor de calidad. En concreto se obtuvo un valor de VRM de 0.78 para el año noveno utilizando una capa oculta de tan solo 3 neuronas, frente a valores de VRM del orden de 1.5, obtenidos en las etapas anteriores. En la figura 4.8 se presentan los 5 primeros días de una serie generada con esta topología de red.

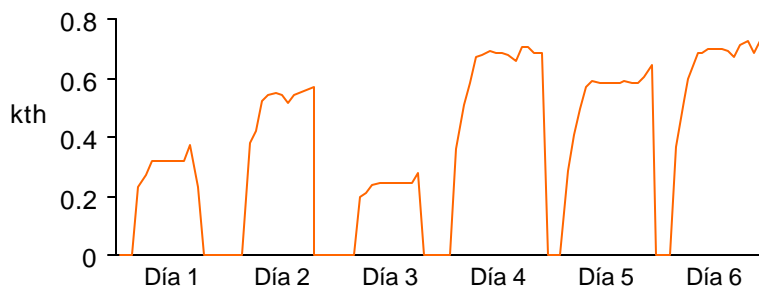


Figura 4.8. Serie generada tras incorporar el índice de claridad diario.

Sin embargo de nuevo se presentaron algunos defectos:

- La evolución de los valores horarios a lo largo de un día es casi perfecta durante la mayor parte de las horas, pero no es capaz de descender a valores bajos para las últimas horas del día como así ocurre en las series reales.
- En las curvas diarias no existe el rizado aleatorio de los valores reales, sino que sigue la tendencia que explican Graham y Aguiar en sus métodos.

Información de la hora del día

La última entrada que se incorporó al PMC fue la relativa a la hora del día correspondiente al valor de k_{th} a generar. Se normalizó según la expresión presentada en el apartado 4.2.2. De nuevo en este caso, la forma de las curvas del índice de claridad horario sigue adaptándose correctamente a la tendencia de la curva, pero no presenta el rizado de la componente estocástica. Sin embargo, con esta nueva entrada si se solventó el primer problema presentado anteriormente sobre la caída de las curvas a las últimas horas del día. Con esta nueva entrada la topología del PMC incluía una capa oculta compuesta de 15 neuronas con función sigmoideal. En este caso el valor del VRM tras 100 épocas de entrenamiento fue de 0.16. Junto a esta entrada se incorporó también una nueva entrada que indica si ha salido o no a ha salido el sol, de modo que no se presenten situaciones absurdas como la posibilidad de que el valor de k_{th} sea distinto de cero cuando no haya salido el sol o ya se haya puesto.

En la siguiente figura puede verse la forma de las curvas de algunos días de ejemplo tras la incorporación de esta nueva entrada. Se han incluido datos del índice de claridad horario reales para mostrar cómo los valores del índice de claridad horario sintético generados por el PMC siguen perfectamente la tendencia de las series, aunque no presentan el rizado aleatorio característico de las mismas.

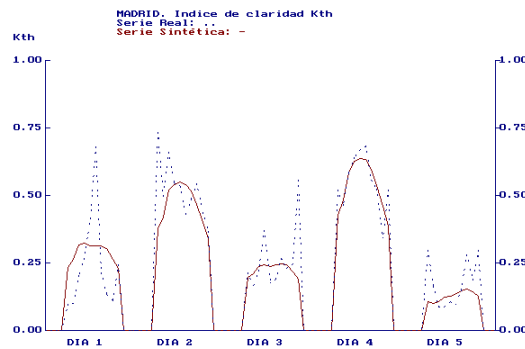


Figura 4.9. Serie real y serie generada con la estructura neuronal final.

Hay que resaltar que la incorporación de nuevas entradas a la red se realiza de una manera muy sencilla, como se ha visto a lo largo de este capítulo, por lo que no supone ninguna complicación el realizar modificaciones sobre el PMC según los propósitos que se persigan.

Resumen de parámetros utilizados

Para finalizar este apartado se presenta un pequeño resumen de los parámetros característicos utilizados en la estructura final del PMC.

a) Capa de entrada

La capa de entrada consta de 7 entradas que fueron descritas anteriormente. (Apartado 4.2.2).

b) Capa de salida

Sólo tiene una salida que es el valor del índice de claridad a predecir.

c) Capa oculta

Finalmente consta de 15 nodos. La forma de obtener esos 15 nodos fue realizando diferentes pruebas. Para valores por encima de 15, además de que el entrenamiento empezaba a ralentizarse, aparecía el problema de la saturación en las neuronas lo que conducía a resultados totalmente erróneos. Con menos neuronas en la capa oculta, menos de 10, la convergencia era bastante rápida pero los resultados eran peores que para valores mayores de 10 neuronas. Con valores entre 10 y 15 los resultados eran muy similares, aunque los mejores se producían con 15, por lo que este fue el número final adoptado. En la tabla que se presenta a continuación se muestran los tiempos de entrenamiento tras 10 épocas, junto con el valor de VRM, según el número de nodos en la capa oculta.

Nº Nodos	1	3	5	8	10	12	14	15	17	19	20	25
Tiempo (s)	373	457	578	798	1103	1215	1376	1476	2987	3787	4191	5673
VRM	2.13	1.98	1.77	0.94	0.23	0.19	0.17	0.16	0.97	1.58	7.77	9.82

Tabla 4.2. Comparación tiempos de entrenamiento versus nº nodos en al capa oculta.

d) Tasa de aprendizaje η

Con este parámetro también se realizaron diferentes pruebas llegándose a la conclusión de que el valor mejor era 0.5. Para valores de la tasa de aprendizaje menores de 0.5, la convergencia era muy lenta. Para valores de este parámetro, η , mayores, aunque disminuían los tiempos de convergencia los resultados de VRM no eran mejores que con 0.5.

e) Término de momento α

Tras las diferentes pruebas realizadas, se comprobó que el término de momento no mejora sustancialmente los resultados, por lo que en la versión final no se utiliza término de momento, ya que complica el algoritmo de entrenamiento.

f) Funciones sigmoidales

Las funciones sigmoidales son aplicadas a todos y cada uno de los nodos de que consta el PMC utilizado.

g) Tipo de entrenamiento

El entrenamiento se realiza siguiendo el siguiente esquema. En primer lugar, del conjunto de datos de entrenamiento (recordar localidad Madrid con 8 años de radiación solar global) se elige un día aleatoriamente, del que se dispondrá de los valores de índice de claridad diario (K_T), hora de salida del sol, orden del día en la secuencia anual (d_n), etc. Se muestran a continuación a la red (figura 4.5), los 16 valores del índice de claridad horario (k_t) de forma secuencial y ordenada, y se va acumulando el error entre el valor esperado y el valor producido por la red, pero sin actualizar los pesos. Cuando se han mostrado todos los valores del k_t correspondientes a ese día en particular se realiza la actualización de los pesos antes de repetir el proceso con otro día elegido aleatoriamente, siempre que el error producido en este caso sea menor que el error obtenido en el proceso similar anterior. Cuando se han mostrado todos los valores correspondientes a todos los días del conjunto de entrenamiento, se dice que se ha completado una época.

4.3. Universalización del método propuesto

Como ya se ha adelantado en la sección anterior, se realizó un primer estudio que trató de demostrar las capacidades de universalización que posee el PMC. Para ello se seleccionaron 3 localidades de España de las que se disponían de registros históricos de radiación solar horaria. Estas otras localidades, junto con algunas características se presentan en la siguiente tabla:

Localidad	Clima	Latitud	Altitud (m)	Años Registrados
Oviedo	Atlántico	43.35 °	348	1.977 - 1.984
Madrid	Continental	40.45 °	664	1.978 - 1.986
Málaga	Mediterráneo	36.66 °	7	1.977 - 1.984

Tabla 4.3. Localidades auxiliares para el entrenamiento.

Se observa que se eligieron localidades situadas a diferentes altitudes sobre el nivel del mar, diferentes latitudes y diferentes climas. El proceso seguido fue el siguiente. Para cada uno de estas localidades se entrenó un PMC que se denominó PMC propio, siguiendo la forma descrita en apartados anteriores, es decir, se tomaron los primeros años para el entrenamiento y se reservó el último para la comparación y validación de resultados. Posteriormente, y con el fin de tratar de probar la universalidad se generaron con la red PMC propia de Madrid años artificiales para Oviedo y Málaga. También se generaron series artificiales de radiación solar horaria con los métodos de Graham y Hollands (GH) y Aguiar y Collares-Pereira (AC). Resumiendo, los años artificiales de series de radiación que se obtuvieron para comparar con los datos reales fueron:

- a) 3 años de radiación solar horaria sintéticos a partir de la propia red de cada localidad: uno para Madrid (red PMC propia de Madrid), otro para Oviedo (con la red entrenada para Oviedo) y otro para Málaga (con el PMC propio entrenado para tal fin con datos de Málaga)
- b) 2 años de valores de radiación solar horaria generados con la red que había sido entrenada para los datos de Madrid (PMC de Madrid) pero aplicada en Oviedo y Málaga (es decir, utilizando el PMC de Madrid, pero con los datos de entrada, índice de claridad diario, correspondientes al año de prueba de Oviedo y de Málaga)
- c) 3 años sintéticos horarios generados según GH para cada una de las tres localidades en estudio.
- d) 3 años sintéticos horarios generados según AC, de nuevo para cada localidad.

En las siguientes figuras se muestran los resultados obtenidos por los diferentes métodos comparados con los datos reales. Como se observa, los tres métodos presentan valores similares de las series para la componente de la tendencia, los cuales se ajustan bastante bien a los valores reales, aunque en un estudio que se presenta a continuación

se ve que el PMC es el mejor método de los tres, al superar los resultados de los métodos clásicos. Se han presentado tanto días de invierno como días de verano. En invierno, al haber mayor probabilidad de presencia de nubes, los datos reales presentan el rizado aleatorio característico debido a la variabilidad de la radiación global, y los valores generados sintéticamente no se ajustan tan bien a los reales. En verano sin embargo, al haber mayor número de días despejados, las curvas reales y sintéticas se ajustan con gran exactitud, siendo prácticamente indistinguibles entre sí.

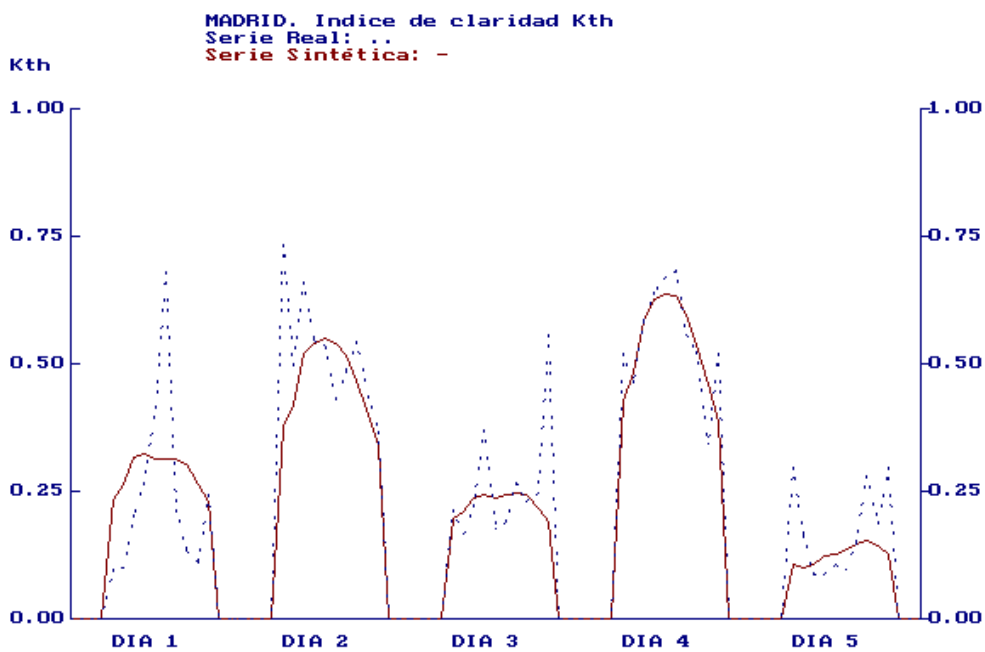


Figura 4.10. Serie real frente a serie generada por el PMC. Madrid. 1.986. Invierno.

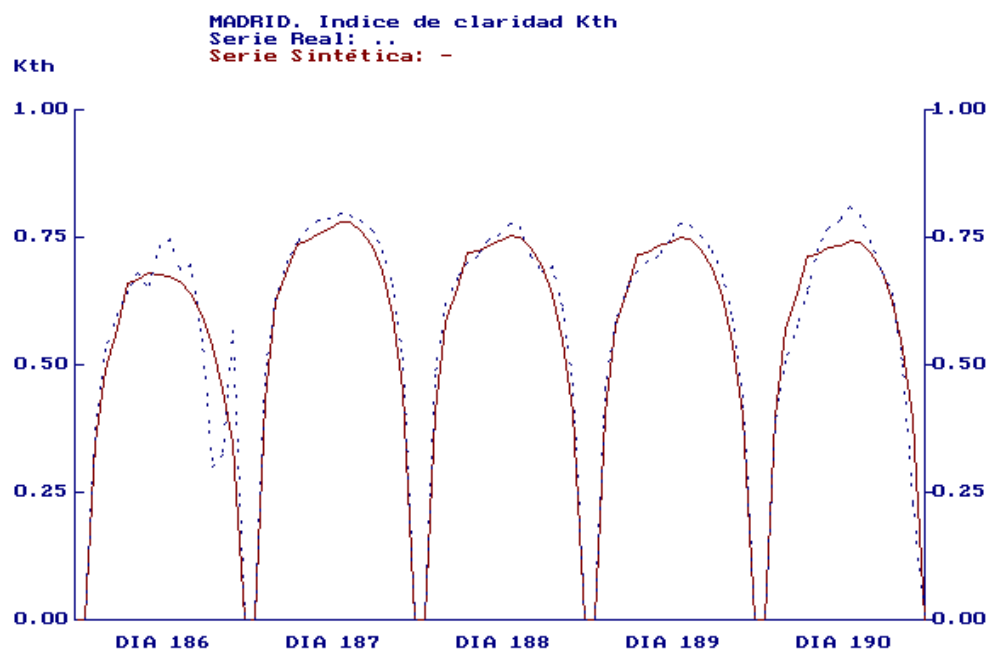


Figura 4.11. Serie real frente a serie generada por el PMC. Madrid. 1.986. Verano.

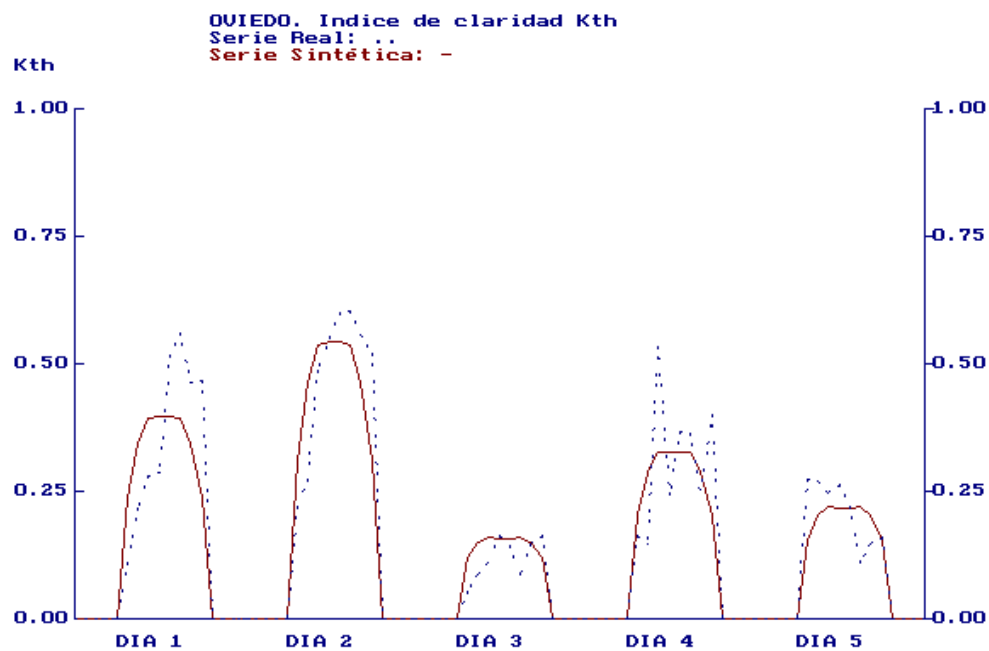


Figura 4.12. Serie real frente a serie generada por el PMC. Oviedo. 1.984. Invierno.

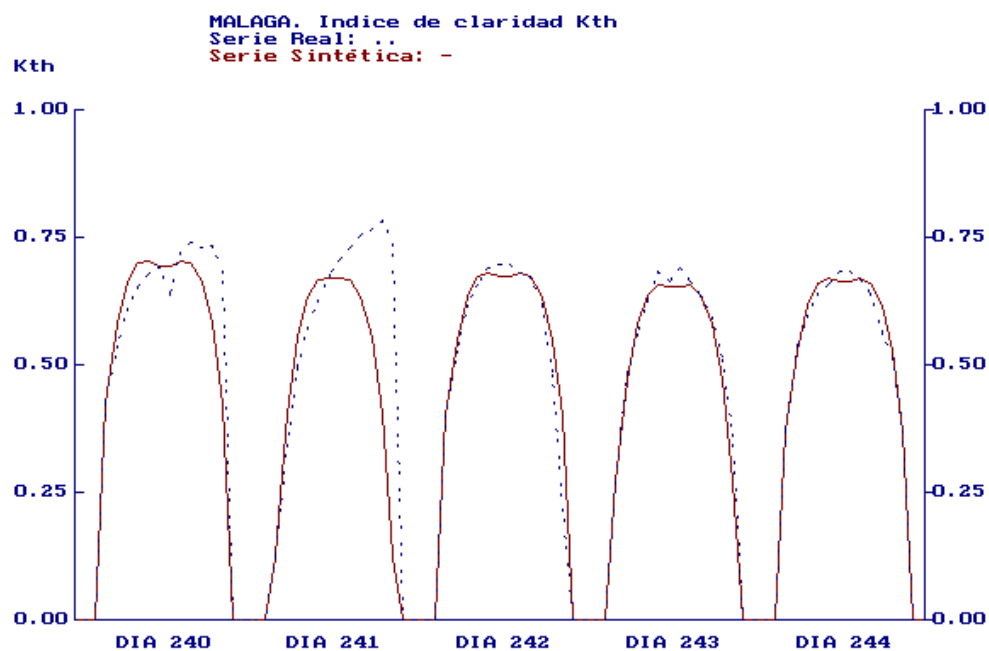


Figura 4.13. Serie real frente a serie generada por el PMC. Málaga. 1.984. Verano.

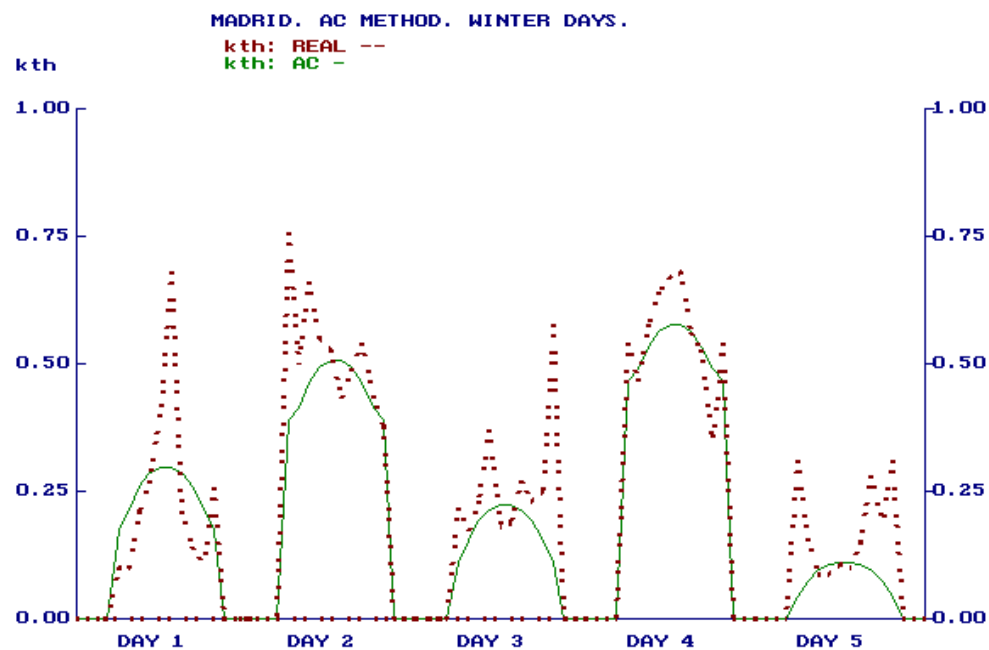


Figura 4.14. Serie real frente a serie generada por AC. Madrid. 1.986. Invierno.

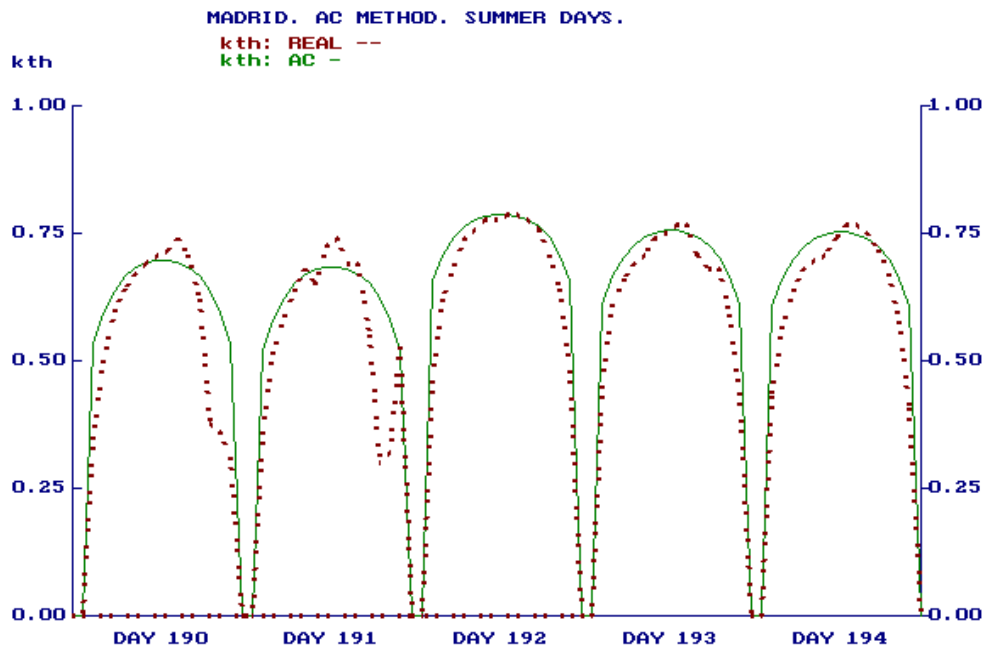


Figura 4.15. Serie real frente a serie generada por AC. Madrid. 1.986. Verano.

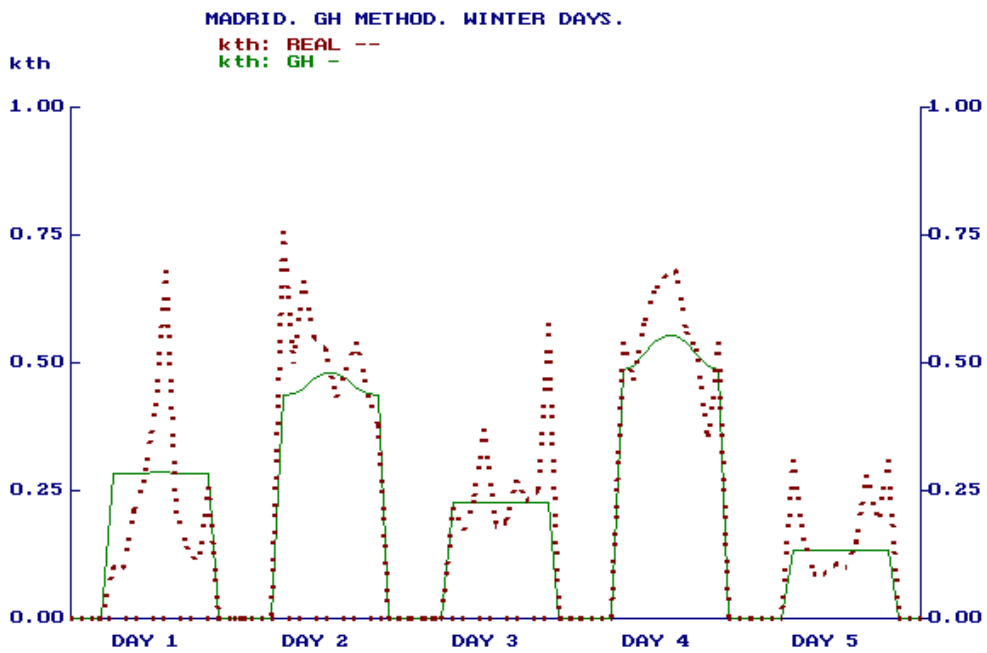


Figura 4.16. Serie real frente a serie generada por GH. Madrid. 1.986. Invierno.

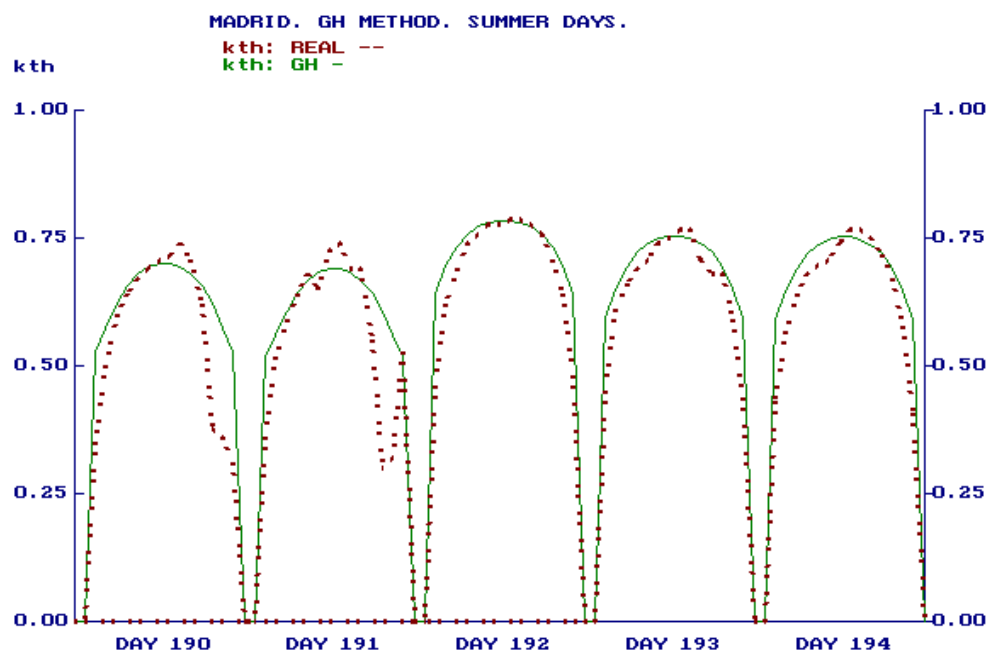


Figura 4.17. Serie real frente a serie generada por GH. Madrid. 1.986. Verano.

Gráficamente, aunque se intuye por la forma de las curvas, no se puede saber si los resultados obtenidos por el método neuronal se ajustan mejor a los datos reales que los resultados obtenidos por los otros dos métodos. Es necesario realizar un estudio más detallado utilizando otros parámetros como son el valor VRM, que ya ha sido definido.

Todos los años sintéticos obtenidos fueron comparados con datos reales de cada localidad, aplicando en este primer estudio el parámetro VRM. Los resultados se muestran en la siguiente tabla:

Método	Localidad	VRM
PMC (Red propia)	Madrid	0.1022
GH	Madrid	0.1611
AC	Madrid	0.1574
PMC (Red propia)	Oviedo	0.1723
PMC (Red entrenada para Madrid)	Oviedo	0.1761
GH	Oviedo	0.2481
AC	Oviedo	0.2567
PMC (Red propia)	Málaga	0.1255
PMC (Red entrenada para Madrid)	Málaga	0.1364
GH	Málaga	0.2418
AC	Málaga	0.2384

Tabla 4.4. VRM. Series reales frente a series predichas por PMC, AC y GH. (Se incluyen los resultados de la red entrenada en Madrid pero utilizada en Oviedo y Málaga.)

En la medida del valor VRM de los tres métodos frente a los datos reales, el resultado de dicho valor con los datos del PMC es sensiblemente mejor que los métodos clásicos, como se muestra en la tabla 4.4. Los mayores valores de VRM se presentan para Oviedo, como era lógico, puesto que en Oviedo, al existir un clima atlántico bastante lluvioso, la mayoría de los días presentan bastante rizado, por lo que la tendencia de las curvas obtenidas por los tres métodos difiere en mayor medida de los valores reales.

Como también se observa, los valores del parámetro VRM son mejores para PMC que los obtenidos siguiendo los métodos de Graham y Aguiar, incluso para la red entrenada en Madrid pero aplicada en Oviedo y Málaga. Aunque el valor de VRM empeora un poco con respecto a los valores que se obtienen cuando se aplica la red propia, los resultados siguen siendo mejores que los obtenidos por AC y GH. Con este estudio se puede concluir que una red entrenada en cierta localidad podrá ser utilizada para generar datos en otras localidades, lo que demuestra las capacidades de universalización del PMC.

4.4 Uso del Año Meteorológico Típico. Mejora del Entrenamiento

A la vista de las capacidades de universalización del PMC, el siguiente paso fue entrenar una sola red neuronal artificial para un conjunto más amplio de localidades españolas. En este caso las localidades en estudio fueron las siguientes:

Localidad	Clima	Latitud	Altitud	Años Registrados
Oviedo	Atlántico	43.35 °	348	1.977 - 1.984 (8)
Logroño	Continental	42.46 °	372	1.981 - 1.984 (4)
Madrid	Continental Extremo	40.45 °	664	1.977 - 1.984 (8)
Palma de Mallorca	Mediterráneo Insular	39.33 °	32	1.977 - 1.984 (8)
Badajoz	Continental Atenuado	38.89 °	185	1.976 - 1.983 (8)
Murcia	Mediterráneo Árido	38.00 °	69	1.977 - 1.984 (8)
Málaga	Mediterráneo	36.66 °	7	1.977 - 1.984 (8)

Tabla 4.5. Localidades en estudio para obtener el AMT.

Sin embargo, apareció el problema de la gran cantidad de datos de radiación solar global horaria con los que se tenía que operar (52 años x 365 días x 16 horas = 303.680 datos). Esto suponía un serio problema para los entrenamientos. Con el fin de reducir el número de datos a manejar se pensó en utilizar en vez de todos los años de cada localidad, el Año Meteorológico Típico (AMT) [11,12] de la misma. El AMT es un año obtenido a partir de los datos reales en el que se resumen las características meteorológicas más importantes de los diferentes años registrados. Es un año con valores reales, no con valores artificiales, puesto que se obtiene a partir de los datos reales, pero habiendo seleccionado algunos de ellos según unos criterios que se detallan a continuación.

En primer lugar, para la construcción del AMT, se pueden utilizar diferentes periodos base, aunque es conveniente que ese periodo sea el mes, es decir, utilizar para cada mes genérico que compondrá el AMT todos los datos de un mes único de la localidad en cuestión. Así, el AMT representará tanto la variación de las medias mensuales a lo largo de todo el año como la distribución de los valores diarios y horarios dentro de cada mes. Se utiliza el mes como periodo base, porque si se tomasen los datos de radiación solar de un único año como año típico, en este caso no se tendría la distribución correcta de las secuencias de radiación en ese periodo, y si se utilizase el

día como periodo base, se tendría una sucesión de días de índice de claridad casi uniforme.

Para la selección de los meses que constituirán el AMT de cada localidad en estudio, se han seguido los criterios habituales [11,12], que son principalmente:

- a) Los valores medios mensuales de la radiación diaria
- b) La distribución mensual de los valores del índice de claridad.

Resumiendo, para la construcción del AMT en una localidad se realiza lo siguiente. Supóngase que se cuenta con los datos de radiación solar de cierta localidad medidos durante 20 años. Es decir, se dispondrá de $20 \times 12 = 240$ meses de datos de radiación. Habrá así, lo que se podría definir como 20 meses “enero”, 20 meses “febrero”, etc. Con los 20 meses “enero” se calculan a continuación las medias mensuales de radiación solar correspondientes a cada mes y a su vez la media de estas medias mensuales obtenidas. Se tendrán entonces 21 valores: 20 medias mensuales más la media de dichas medias mensuales. Posteriormente se comparan las 20 medias mensuales con el valor correspondiente a la media de las medias, y el valor que más se aproxime a dicha media, será elegido, en esta primera etapa como mes constituyente del AMT para enero. A continuación, en una segunda etapa se hace un estudio similar pero para la distribución mensual de los valores del índice de claridad. Se dispondrán de 20 distribuciones mensuales correspondientes a cada mes, más una distribución mensual media genérica. La distribución mensual del mes que más se aproxime a la distribución media, indicará el mes elegido para constituir el AMT para enero, en esta segunda y definitiva etapa. Este proceso se repite para el resto de los meses, completando de esta forma el AMT.

En un caso real, por ejemplo un AMT para cierta localidad cuyos datos registrados abarquen desde el año 1976 a 1983 podría estar formado por el mes de enero de 1977, el mes de febrero de 1982, el mes de marzo de 1980, etc. Se ha realizado el cálculo del AMT para las localidades en estudio presentadas en la tabla 4.5 y los meses que constituyen el AMT de esas localidades son:

Localidad	Oviedo	Logroño	Madrid	P.Mallorca	Badajoz	Murcia	Málaga
Mes	Año Seleccionado						
Enero	1980	1984	1984	1977	1980	1980	1978
Febrero	1984	1983	1981	1984	1983	1979	1982
Marzo	1977	1983	1978	1981	983	1981	1983
Abril	1979	1983	1981	1978	1979	1977	1980
Mayo	1982	1983	1983	1980	1983	1979	1982
Junio	1984	1983	1977	1983	1983	1984	1984
Julio	1979	1981	1984	1980	1982	1982	1977
Agosto	1982	1984	1978	1982	1982	1982	1979
Septiembre	1982	1981	1978	1977	1983	1977	1978
Octubre	1983	1981	1982	1978	1983	1983	1978
Noviembre	1982	1983	1978	1977	1983	1980	1980
Diciembre	1977	1981	1981	1977	1983	1979	1979

Tabla 4.6. Meses constituyentes del AMT para las localidades en estudio.

Una vez obtenido el AMT de cada localidad se entrenó una única red, con la misma estructura que la explicada en apartados anteriores (7x15x1), pero utilizando para el entrenamiento en vez de los datos de todos los años de cada localidad, sólo los datos del AMT. Tras el entrenamiento se generaron años sintéticos correspondientes a todos los años de cada localidad y se compararon los resultados obtenidos con los datos reales y con los datos que se obtienen empleando la red propia de cada localidad. Como ejemplo, se muestra a continuación los valores de VRM para la localidad de Málaga. En ella se presentan los resultados de proporcionados por la red única (la que utiliza para el entrenamiento los AMTs de todas las localidades en estudio) y la red propia (la que utiliza para el entrenamiento sólo los datos de la localidad en cuestión).

Málaga		
VRM		
Año	PMC Propio	PMC Único
1977	0.1034	0.0977
1978	0.1087	0.1021
1979	0.1124	0.1015
1980	0.1024	0.0880
1981	0.1111	0.0916
1982	0.0943	0.1055
1983	0.0965	0.0818
1984	0.1200	0.1123

Tabla 4.7. Valores de VRM para Málaga con dos entrenamientos diferentes.

En los anexos se muestran las restantes tablas para las diferentes localidades. Lo importante es destacar que con la utilización del AMT se han obtenido ventajas importantes:

- a) En primer lugar, se trabaja con una cantidad menor de datos. Se pasa de utilizar 52 años x 365 días x 16 horas = 303.680 datos, a utilizar sólo 40.880.
- b) Se obtiene una única red válida para cualquier localidad (española).
- c) Disminuyen los tiempos de entrenamiento, principalmente como consecuencia de tratar con un número menor de datos. En este caso el entrenamiento dura sólo alrededor de 375 segundos.

4.5. Componente aleatoria

Siguiendo la técnica que AC y GH proponen en el estudio de las series de k_{th} , de dividir dichas series en dos componentes, una que sigue la tendencia de la serie y otra que corresponde a pequeñas variaciones aleatorias sobre dicha tendencia, en este trabajo se hizo algo similar. A la vista de los resultados obtenidos hasta este momento, se comprendió que por medio del entrenamiento con la estructura de red neuronal utilizada hasta aquí, el PMC sólo era capaz de proporcionar la componente de tendencia y no la aleatoria. Se decidió, en primer lugar, realizar un estudio comparando las series reales y las series obtenidas por medio del PMC y comprobar si los residuos obtenidos como diferencia de ambas correspondían a una serie aleatoria. Es decir, se restan ambas series

(la real y la sintética) y a la nueva serie obtenida se le aplica algún tipo de test para comprobar su aleatoriedad.

Existen diferentes tests para comprobar la aleatoriedad de una serie numérica. Un test no paramétrico a tal fin lo proporciona la *teoría de las rachas* [13].

Para entender qué son las rachas (o escalones) considérese una secuencia con dos símbolos: a y b, tal como:

a a | b b b | a | b b | a a a a | b b b | a a a a |.

Al tirar una moneda, por ejemplo, a sería “cara” y b sería “cruz”. Una racha se define como un conjunto de símbolos idénticos (o relacionados) contenido entre dos símbolos diferentes o uno sólo si se está al comienzo o la final de la secuencia. Leyendo de izquierda a derecha en la secuencia anterior, la primera racha, indicada por una barra vertical, consiste en dos *aes*, la segunda de tres *bes*, la tercera una *a*, etc. En este ejemplo hay **siete** rachas en total.

Parece claro que existe relación entre aleatoriedad y el número de rachas. Así para la secuencia:

a | b | a | b | a | b | a | b | a | b | a | b | a | b | a |

hay un *esquema cíclico*, en el que se va pasando siempre de *a* a *b* y viceversa, que difícilmente puede ser aleatorio. En este caso hay *demasiadas rachas* (de hecho, hay el máximo posible con ese número de letras *a* y *b*).

Por otra parte, para la secuencia: a a a a a | b b b b | a a a a | b b b |

parece haber un *esquema de tendencia o de inercia*, en el que las *aes* y las *bes* están agrupadas. En este caso hay *muy pocas rachas*, y no se consideraría tampoco aleatoria a esa secuencia.

Así pues, una secuencia se considera **no aleatoria** si hay **demasiadas** o **muy pocas rachas**, y aleatoria en los demás casos. Para cuantificar esa idea, supongamos que se forman todas las posibles secuencias con N_1 *aes* y N_2 *bes*, para un total de N símbolos ($N = N_1 + N_2$). La colección de todas esas secuencias conduce a una distribución muestral. Cada secuencia tiene asociado un número de rachas, denotado por V . De este modo se forma una distribución muestral del estadístico V . Se demuestra [10] que esa distribución muestral tiene de media y varianza las siguientes expresiones:

$$\mu_v = 2 \cdot N_1 \cdot N_2 / (N + 1)$$

$$\sigma_v = 2 \cdot N_1 \cdot N_2 \cdot (2 \cdot N_1 \cdot N_2 - N) / N^2 \cdot (N-1)$$

Mediante esas expresiones se puede contrastar la hipótesis de aleatoriedad a niveles de significación apropiados. Resulta que si N_1 y N_2 son ambos al menos iguales a 8, entonces la distribución muestral de V es muy próxima a una distribución normal. Luego:

$$z = (V - \mu_v) / \sigma_v$$

está normalmente distribuido con media 0 y varianza 1.

Una forma de aplicar el test de aleatoriedad es el llamado *test sobre y bajo mediana*. Para determinar si unos datos numéricos son aleatorios, se calcula la mediana de la serie y, posteriormente, colocados en el mismo orden en que fueron tomados, se calcula la diferencia entre cada valor numérico y la mediana. Se asigna a continuación la letra *a* o *b* según el valor esté sobre o bajo la mediana. Si un valor coincide con la mediana se suprime. La muestra será aleatoria según lo sea la secuencia de *aes* y *bes* así obtenida.

Aplicando estas ideas a este trabajo, en este caso se cuenta con la serie de la componente α (parte supuestamente aleatoria) obtenidas como diferencia entre el k_{th} real y el k_{th} calculado por la red neuronal (tendencia). Esa serie tiene valores comprendidos entre (-1,1). El proceso seguido ha sido el siguiente:

- 1.- Se ha tomado cada serie de 365 días x 16 valores horarios (5840 datos) y se ha calculado la mediana. (Para ello se han de disponer en orden creciente y la mediana es el valor intermedio.) Se han tomado las series de datos de Madrid, Oviedo y Málaga, obteniéndose como **mediana el valor 0**.
- 2.- Se forma una nueva serie como la diferencia entre el valor de la componente α y la mediana (en ese orden). Si el resultado es mayor que cero se le asigna la letra *a* y si el valor es menor que cero se le asigna la letra *b*.
- 3.- Se cuentan el número de *aes* (N_1) el número de *bes* (N_2) y el número de Rachas (V).
- 4.- Se calculan los parámetros estadísticos media muestral (μ_v), varianza muestral (σ_v) y valor de Zeta.
- 5.- Se aplica el criterio de contraste de hipótesis de aleatoriedad para un nivel de significación de 0.05 (es decir, que se cumpla: $1.96 \geq z \geq -1.96$). Si lo cumple, la hipótesis de que la muestra es aleatoria es cierta con un nivel de confianza del 95 %.

En las tablas siguientes se muestran los resultados:

Madrid						
Año	N_1	N_2	V	μ_v	σ_v	z
1	2066	1927	1313	1995	996	-0.685
2	2114	1887	1351	1995	994	-0.648
3	2181	1830	1352	1991	987	-0.647
4	2103	1910	1350	2003	998	-0.654
5	2027	1984	1380	2006	1002	-0.625
6	2058	1956	1299	2007	1002	-0.706
7	2049	1963	1304	2006	1002	-0.701
8	2136	1872	1369	1996	993	-0.632

Tabla 4.8. Estadísticos del test de aleatoriedad para la componente **a** Madrid.

Málaga						
Año	N_1	N_2	V	μ_v	σ_v	z
1	1965	2041	1332	2003	1001	-0.671
2	1924	2084	1352	2002	999	-0.651
3	1773	2217	1251	1971	973	-0.741
4	2000	1991	1278	1996	997	-0.720
5	2115	1884	1329	1994	993	-0.670
6	2052	1931	1270	1991	994	-0.725
7	1916	2075	1291	1993	994	-0.706
8	1866	2059	1204	1959	976	-0.773

Tabla 4.9. Estadísticos del test de aleatoriedad para la componente **a** Málaga.

Oviedo						
Año	N_1	N_2	V	μ_v	σ_v	z
1	2041	1905	1254	1972	984	-0.729
2	2098	1841	1150	1962	976	-0.832
3	1972	1953	1261	1935	981	-0.716
4	2012	1914	1320	1963	980	-0.656
5	2077	1867	1328	1967	980	-0.652
6	2046	1918	1260	1981	989	-0.729
7	2168	1803	1282	1969	976	-0.705
8	2128	1844	1269	1977	983	-0.720

Tabla 4.10. Estadísticos del test de aleatoriedad para la componente **a** Oviedo.

Según estos resultados se observa que las series de componente aleatoria son efectivamente aleatorias con un nivel de significación de 0.05 (es decir un 95 % de confianza).

Una vez que se comprobó que las series de la componente aleatoria α son efectivamente aleatorias, con el fin de aplicar de alguna forma el PMC para su generación, se decidió utilizar un algoritmo similar al que propuso AC, pero con la diferencia de que todos los valores de parámetros que AC proponen tras un estudio estadístico de los diferentes datos de radiación de que disponían, en este trabajo se obtuvieron aplicando diferentes estructuras neuronales con PMC.

Por lo tanto el algoritmo aplicado para calcular la parte aleatoria de las series, y completar así el trabajo es el siguiente:

1. Cálculo de la autocorrelación ϕ_1 y de la desviación estándar σ_α por medio de estos dos PMCs:



Figura 4.18. PMC para la desviación estándar σ_α y coeficiente de autocorrelación ϕ

La estructura de estos PMC es $2 \times 5 \times 1$. La forma de elegir 5 nodos en la capa oculta es igual a la explicada en el apartado 4.2.3. Igualmente, la forma de entrenar estos dos nuevos PMCs es similar a la que se ha venido utilizando en todo este trabajo, es decir, se utilizan datos reales para el entrenamiento y se reservan algunos datos para la validación del mismo. En la siguiente figura se muestra la evolución de la desviación estándar de la componente aleatoria real (en función del índice de claridad diario K_T) y la obtenida según AC y según el PMC.

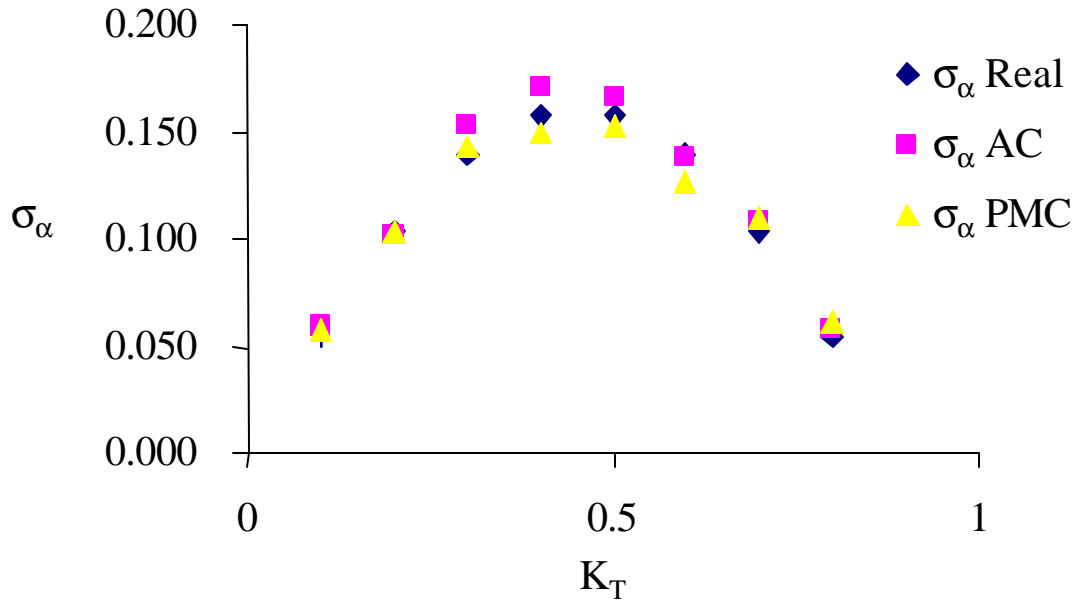


Fig. 4.19. Desviación estándar σ_a frente al índice de claridad K_T .

De nuevo los resultados obtenidos por el PMC son muy similares a los obtenidos por otros métodos y prácticamente indistinguibles de los reales.

2. Cálculo de la variable normalizada “y” para cada hora “h” por medio de la siguiente ecuación:

$$y(h) = \phi_1 y(h-1) + r.$$

El primer valor para construir esa serie puede ser un número aleatorio pequeño.

r es un número aleatorio gaussiano con media nula y desviación estándar σ^* dada por:

$$\sigma^* = \sigma (1 - \phi_1^2)^{0.5}.$$

3. Finalmente conocidos la serie “y” y la desviación estándar σ , la componente aleatoria se construye como producto de ambas:

$$\mathbf{a} = \mathbf{S}_a \mathbf{y}.$$

4.6. Series diarias

Con el fin de completar el estudio sobre la generación de series de radiación solar, quedaba resolver el problema de la generación de series diarias. Puesto que todo el estudio importante para la generación de series utilizando el PMC se había realizado para la escala horaria, se pensó aplicar todo el conocimiento adquirido para este caso, pero para la escala diaria. Se preparó un nuevo PMC de estructura similar al utilizado

para las series horarias. Las entradas utilizadas en este caso, tal y como puede verse en la figura 4.20, se detallan a continuación:

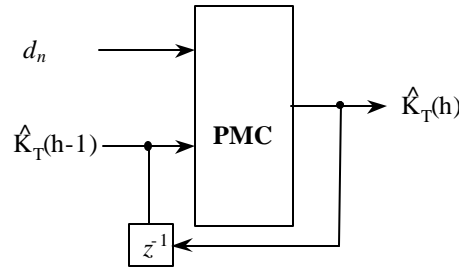


Figura 4.20. PMC Arquitectura para la predicción de los índices de claridad diarios.

Capa de entrada

En este caso sólo se usan dos entradas, una que proporciona la información del orden del día, que es exactamente igual al caso horario (d_n) y otra que es el índice de claridad diario del día anterior al día a generar. Sólo se emplea el índice de claridad diario del día anterior puesto que como han demostrado muchos investigadores (ver [14]), y de nuevo recordando la propiedad de persistencia radiativa, la influencia de la radiación solar es importante entre días consecutivos, pero disminuyen en gran medida para distancias temporales mayores de un día.

Capa oculta

En este caso, tras el estudio de las distintas opciones se llegó a la conclusión que con una capa oculta de 7 nodos se obtenían los mejores resultados.

Capa de salida

En la última capa se tiene el índice de claridad diario que se quiere obtener.

Todo lo referente al entrenamiento es similar al caso horario. Se utiliza una única red entrenada con los AMT de las localidades en estudio. Una diferencia con el caso horario es que se consiguió ampliar la base de datos en estudio, puesto que hay más estaciones que miden radiación global diaria. Por tanto, en este caso, las localidades en estudio, junto con los correspondientes años fueron las presentadas en la tabla siguiente.

Localidad	Años registrados	Localidad	Años registrados
Albacete	1976-1983	Lérida	1977-1983
Alicante	1976-1982	Logroño	1975-1993
Almería	1976-1983	Lugo	1976-1984
Badajoz	1976-1983	Madrid	1973-1986
Barcelona	1975-1984	Málaga	1976-1993
Bilbao	1976-1982	Murcia	1975-1993
Burgos	1976-1982	Oviedo	1973-1993
Cádiz	1976-1984	Palma de Mallorca	1975-1993
Castellón	1979-1982	Salamanca	1976-1984
Cáceres	1983-1993	Santander	1975-1993
Ciudad Real	1976-1983	Santiago	1975-1982
Córdoba	1976-1984	Sevilla	1975-1988
Gran Canaria	1976-1983	Toledo	1979-1991
Huelva	1976-1983	Valladolid	1975-1982
Lanzarote	1976-1983	Valencia	1975-1980
La Palma	1976-1983	Vigo	1977-1984
León	1976-1983	Zaragoza	1971-1983

Tabla 4.11. Localidades en estudio. Años registrados con radiación global diaria.

A la vista de los resultados, en este caso las series sintéticas obtenidas presentaban una aleatoriedad similar a la que tienen los datos reales. Como se verá en el siguiente capítulo, los resultados del estudio comparativo con otras series sintéticas producidas por otros métodos demuestran que las series artificiales generadas por medio del PMC son indistinguibles de las reales. Se presenta en la tabla siguiente los valores de VRM obtenidos para algunas localidades estudiadas:

Localidades			
Año	Cáceres	Santander	Murcia
1983	1.0994	1.6526	1.3126
1984	0.9872	1.5316	1.0968
1985	0.9346	1.5241	1.0899
1986	1.2105	1.5249	1.3421
1987	1.0838	1.6001	1.2707
1988	0.8918	1.7938	1.1493
1989	1.1238	1.6068	1.2306
1990	1.1189	1.5210	1.0623
1991	1.0982	1.5831	1.0231
1992	1.0751	1.6233	1.0319
1993	1.1827	1.6097	1.1102

Tabla 4.12. Valores de VRM para el caso de diarias en algunas localidades.

En la siguiente figura se muestra para un año genérico la evolución del índice de claridad diario real y el generado por medio del PMC:

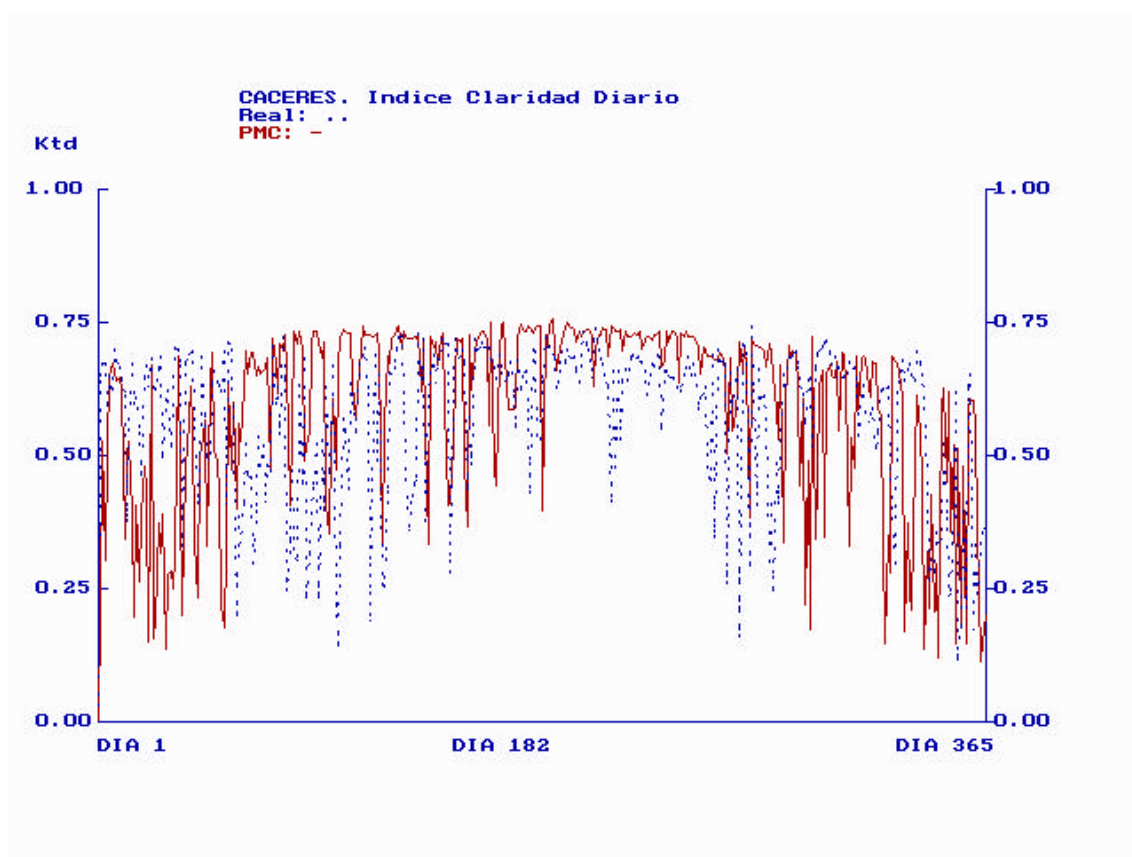


Figura 4.21. K_T real frente a K_T generado por PMC. Cáceres.

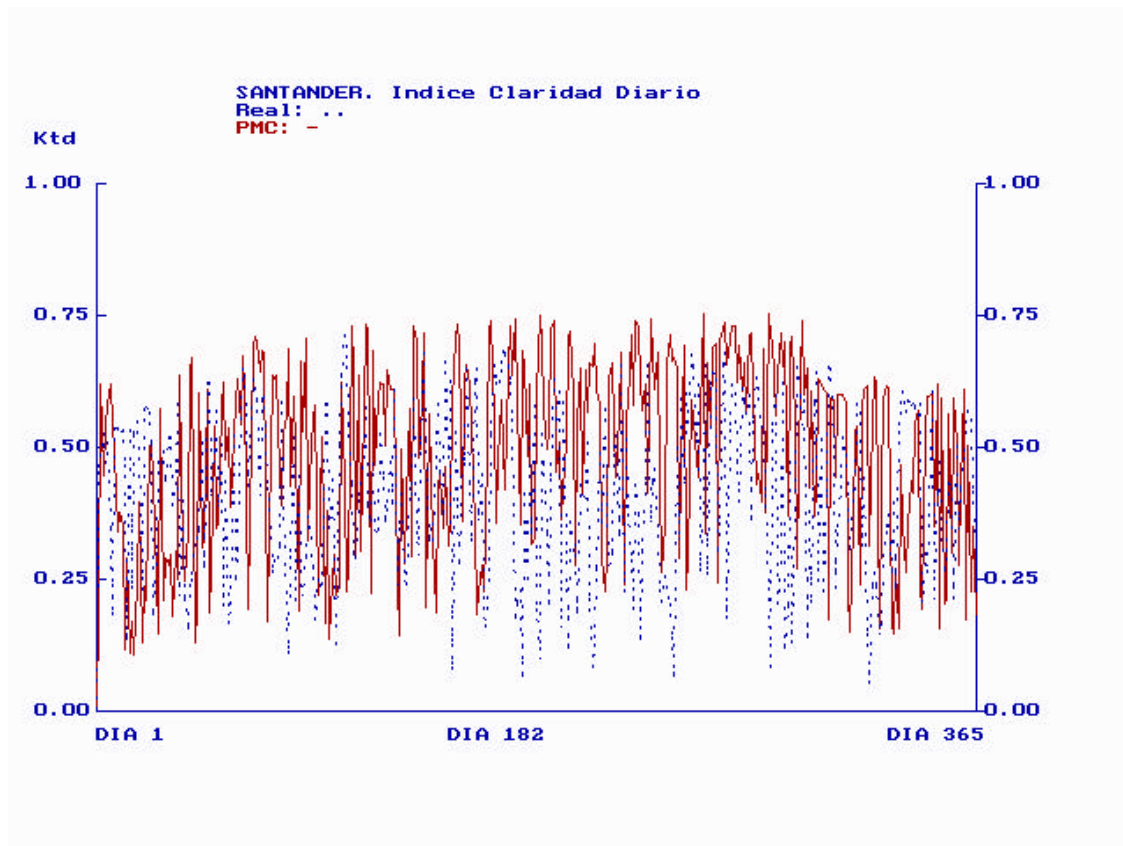


Figura 4.22. K_T real frente a K_T generado por PMC. Santander.

En los anexos se presentan más gráficas para otras localidades.

Referencias

- [1] McCulloch WS., & Pitts W. *A logical calculus of the ideas immanent in nervous activity*. Bulletin of Mathematical Biophysics. Vol. 5. pp 115-133. 1943.
- [2] Rosenblatt R. *Principles of Neurodynamis*. Spartan Books, New York. 1959.
- [3] Minsky M & Papert S. *Perceptrons: an introduction to computational geometry*. The MIT Press. Cambridge, MA. 1969.
- [4] P. Werbos. *Beyond Regression: New Tools for Predicting and Analysis in the Behavioral Sciences*, Ph. D. thesis, Harvard University. 1974.
- [5] Rumelhart, D. Hinton G. & Williams R. *Learning internal representations by error propagation*. Parallel distributed processing: Explorations in the microstructure of cognition. vol. 1. Foundations. The MIT Press. Cambridge. Rumelhart and McClelland (eds.). 1986.
- [6] Aho A. V., Hopcroft, J.E. Ullman J.D. *Data structures and algorithms*. Addison-Wesley, 1983.
- [7] A.S. Lapedes y R. M. Faber. *Non linear signal processing using neural networks: prediction and system modelling*. Technical Report. Los Álamos National Laboratory. 1987.
- [8] A. S. Weigend, D. E. Rumelhart y B.A. Huberman. *Back-Propagation, weight elimination and Time Series Prediction*. Capítulo de los Proceedings of the 1990 Connectionist models Summer School. Morgan Kaufman, 1990.
- [9] A. Vázquez. *Identificación de sistemas mediante Redes Neuronales para control de robots*. ETSI Telecomunicación, Madrid. 1992.
- [10] J. Riesco Prieto. *Optimización global: aplicación en el entrenamiento de redes neuronales*. Proyecto Fin de Carrera. UPM Madrid. ETSI Telecomunicación. 1993.
- [11] M.H. Macagnan. *Caracterización de la radiación solar para aplicaciones fotovoltaicas en el caso de Madrid*. Tesis doctoral. ETSI Telecomunicación Madrid. 1993.
- [12] W. Marion y K. Urban. *User's Manual for TMY2s (Typical Meteorological Years)*. NREL. Golden Colorado. Junio 1995.
- [13] M.R. Spiegel. *Estadística*. McGraw Hill. Colombia. 1998.

