

Centro de Investigación Científica y de Educación Superior de Ensenada



**PREDICCIÓN DE SERIES DE TIEMPO APLICANDO
REDES NEURONALES ARTIFICIALES**

**TESIS
MAESTRIA EN CIENCIAS**

MIGUEL ANGEL PEREZ CHAVARRIA

Ensenada, Baja Cfa., Mexico.

Abril de 1999.

TESIS DEFENDIDA POR
Miguel Angel Pérez Chavarría
Y APROBADA POR EL SIGUIENTE COMITÉ



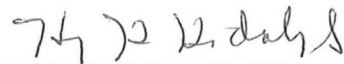
Dr. Francisco Javier Ocampo Torres

Director del Comité



Dr. Edgar Gerardo Pavía López

Miembro del Comité



M.C. Hugo Homero Hidalgo Silva

Miembro del Comité



Dr. Silvio Guido Marinone Moschetto

*Jefe del Departamento de
Oceanografía Física*



Dr. Federico Graef Ziehl

Director de Estudios de Posgrado

16 de abril de 1999

CENTRO DE INVESTIGACIÓN CIENTÍFICA Y DE EDUCACIÓN SUPERIOR
DE ENSENDA.

DIVISIÓN DE OCEANOLOGÍA.
DEPARTAMENTO DE OCEANOGRAFÍA FÍSICA.

PREDICCIÓN DE SERIES DE TIEMPO APLICANDO REDES NEURONALES
ARTIFICIALES

TESIS
que para cubrir parcialmente los requisitos necesarios para obtener
el grado de MAESTRO EN CIENCIAS presenta:

MIGUEL ANGEL PÉREZ CHAVARRÍA.

Ensenada, Baja California, México, abril de 1999.

RESUMEN de la tesis de Miguel Angel Pérez Chavarría, presentada como requisito parcial para la obtención del grado de MAESTRO EN CIENCIAS en OCEANOGRAFÍA FÍSICA. Ensenada, Baja California, México, abril de 1999.

PREDICCIÓN DE SERIES DE TIEMPO APLICANDO REDES NEURONALES ARTIFICIALES.

Resumen aprobado por:



Dr. Francisco J. Ocampo Torres
Director de tesis.

En este trabajo se aplicaron algoritmos basados en redes neuronales artificiales (RNA) para la predicción de series temporales de las variables oceanográficas altura significativa del oleaje (H_S), índice de oscilación del sur (IOS) y anomalía de temperatura superficial del mar ($ATSM$).

Se probaron tres tipos diferentes de redes, las RNRT, las RNR y las RNRFI, encontrando que las últimas fueron las que produjeron las mejores predicciones. Para determinar qué arquitectura de la RNRFI tenía el mejor comportamiento, se hicieron experimentos variando los diferentes parámetros de la red, y se encontró que para series de H_S la red debe conformarse de una capa de entrada con una neurona y un filtro de orden 34, dos capas ocultas cada una con 12 neuronas con filtros de orden 8 y una capa de salida con una neurona. Para el caso de series temporales de IOS y $ATSM$ la red se construyó con una capa de entrada con una neurona y un filtro de orden 10, dos capas ocultas con 8 neuronas en cada una de ellas y filtros de orden 5 y una capa de salida con una neurona. Todas las redes fueron entrenadas con el algoritmo de aprendizaje de retropropagación, usando la sigmoide como función de activación y una razón de aprendizaje de 0.001.

De acuerdo a los experimentos realizados se puede observar que en la mayoría de los casos se obtuvo un error cuadrático medio normalizado, de alrededor de 0.4 ± 0.1 y un coeficiente de correlación entre la serie original y la predicha mayor que 0.8. Estos resultados son válidos en la predicción hasta de veinte datos, cuya utilidad dependerá de la escala temporal del fenómeno. Por ejemplo, para el caso de los fenómenos que se estudiaron en esta tesis, predecir veinte datos de H_S tendrá sentido para algunas aplicaciones, pues equivale a pronosticar uno o dos días. Para el fenómeno de El Niño, las series de tiempo analizadas tienen un intervalo de muestreo mensual, esto significa que veinte datos equivalen a una predicción de año y medio. De la comparación que se hace con otros métodos para la predicción de la $ATSM$ se puede observar que las RNA tuvieron un mejor comportamiento, considerando predicciones a corto plazo.

Palabras clave: predicción, serie-temporal, red-neuronal

ABSTRACT of the thesis of Miguel Angel Pérez Chavarría, presented as partial requirement to obtain the degree of MASTER OF SCIENCE in PHYSICAL OCEANOGRAPHY. Ensenada, Baja California, México, april 1999.

TIME SERIES PREDICTION BY ARTIFICIAL NEURAL NETWORKS.

In this work artificial neural network (ANN) algorithms are used to predict time series of the oceanographic variables significant wave height (H_S), southern oscillation index (SOI) and sea surface temperature anomaly ($SSTA$).

Three types of neural networks were applied, namely TDNN, RNN and FIRNN; it was found the latter produced the best results. In order to determine the most efficient FIRNN architecture some experiments were made varying different parameters. It was found that for H_S series, the network has to be built as follows: one input neuron with an 34^{th} order filter, two hidden layers with twelve neurons each, with 8^{th} order filters and one output neuron. For the SOI and the $SSTA$ the network was build with one input neuron with an 10^{th} order filter, two hidden layers with 8 neruons each, with 5^{th} order filters and one output neuron. All the networks were trained using the backpropagation learning algorithm, a sigmoide as activation function and a learning rate of 0.001 were chosen.

In most experiments a normalized mean square error of 0.4 ± 0.1 and a correlation coefficient between the original and the predicted series greater than 0.8 were found. These results are valid just for the first twenty forecasted data and their usefulness depends on the specific phenomenon time scale. For instance, for the parameters studied in this work, the prediction of twenty H_S data would make sense for some applications, since it means to forecast one or two days. For the ENSO, the studied time series have a monthly sampling interval, which means that twenty data correspond to a year and a half forecast. From the comparison with other $SSTA$ prediction methods it is concluded that the neural networks had better results in case of short term forecastings.

Keywords: forecasting, time-series, neural-network

DEDICATORIA

A mi familia

A mis amigos

AGRADECIMIENTOS

A mi amigo y director de tesis, Dr. Francisco Javier Ocampo Torres, por su apoyo incondicional durante el desarrollo de este trabajo.

A mi amigo y miembro del comité de tesis M.C. Hugo Homero Hidalgo Silva, por su desinteresada ayuda en el desarrollo y conclusión de este proyecto de tesis.

A los miembros del comité de tesis, Drs. Julio Sheinbaum y Edgar G. Pavía, por sus acertados comentarios.

Al Instituto de Investigaciones Oceanológicas de la Universidad Autónoma de Baja California

Al Consejo Nacional de Ciencia y Tecnología

Al Centro de Investigación Científica y de Educación Superior de Ensenada.

Índice General

I INTRODUCCIÓN	1
I.1 Presentación del problema	1
I.2 Objetivos	3
I.3 Organización del trabajo	3
II PREDICCIÓN DE SERIES DE TIEMPO	4
II.1 Modelos de predicción de series de tiempo	4
II.2 Modelos lineales	4
II.2.1 Modelos de promedio corrido (MPC).	5
II.2.2 Modelos Autoregresivos (MAR)	6
II.3 Modelos no-lineales	8
II.3.1 Espacios de estado	8
II.3.2 Redes neuronales artificiales	8
III REDES NEURONALES ARTIFICIALES	10
III.1 Introducción	10
III.2 Modelo biológico	11
III.2.1 El sistema nervioso central	11
III.3 Modelo cibernético	13
III.3.1 Entrenamiento y retropropagación de errores	16
III.3.2 Arquitecturas	20
III.4 Redes Neuronales del tipo RFI	22
IV EXPERIMENTOS Y RESULTADOS	24
IV.1 Predicción de series de tiempo	24

IV.1.1	Introducción	24
IV.1.2	Predicción de H_S	24
IV.1.3	Predicción del IOS	31
IV.1.4	Predicción de $ATSM$	36
V	CONCLUSIONES	38
	REFERENCIAS	40

Índice de Figuras

1	La neurona y sus partes principales	11
2	Esquema de una Red Neuronal Artificial. Las a son las señales de activación procedentes de otras neuronas, las w son las ponderaciones de los lazos entre neuronas, las x son las señales de entrada y la y la salida. Cada línea representa una conexión sináptica.	14
3	Arquitectura tipo RFI. En el modelo espacio-temporal de la neurona, las señales de entrada $a(k)$ se pasan a través de filtros sinápticos w . Todas las conexiones se modelan con filtros de orden k tipo RFI	23
4	a) Localización de la estación San Clemente en donde se obtuvieron los datos que se utilizan en este trabajo. b) Serie de altura significativa del oleaje medido durante 1995-1996. De la serie completa se utilizaron 1000 datos en el entrenamiento de la RNA para predecir 100 (línea realzada). c) Subconjunto de datos de la serie original y suavizada (línea realzada)	25
5	Altura significativa reportada en San Clemente CA. y su predicción (línea continua) en forma no iterada.	27
6	Altura significativa estimada a partir de mediciones en San Clemente CA.(línea continua) y predicciones iteradas obtenidas con una red con una neurona en la entrada y un filtro de orden 34, dos capas ocultas cada una con 12 neuronas y filtros de orden 8 y una neurona en la salida. La red se entrenó con los primeros mil datos de la serie y se predijeron los siguientes cien,empleando a) 6000 épocas de entrenamiento y b) 8000 épocas de entrenamiento.	29
7	Altura significativa estimada a partir de mediciones en San Clemente CA.(línea continua) y predicciones iteradas obtenidas con una red con una neurona en la entrada y un filtro de orden 34, dos capas ocultas cada una con 12 neuronas y filtros de orden 8 y una neurona en la salida.La red se entrenó con los segundos mil datos de la serie y se predijeron los siguientes cien, empleando a) 6000 épocas de entrenamiento y b) 8000 épocas de entrenamiento.	30

8	Serie de tiempo de la diferencia de presión entre Darwin y Tahití, de 1950 a 1998. La parte realizada corresponde al intervalo que se predice, entrenando la red con todos los datos anteriores	31
9	Predicción del <i>IOS</i> no iterada, con datos de 1951 a 1996 inclusive. La línea gruesa corresponde a la predicción y la fina a la señal real.	32
10	Predicción de una serie de tiempo del <i>IOS</i> con datos de 1951 a 1996 inclusive. Se predice el 97 y el 98. La predicción es la línea punteada. La diferencia entre las partes superior e inferior de la figura, es la cantidad de neuronas de la red.	34
11	Predicción de una serie de tiempo del <i>IOS</i> con datos de 1951 a 06/1997 inclusive. Se predice la segunda mitad del 97 y el 98. La predicción es la línea punteada. La diferencia entre las partes superior e inferior de la figura, es la cantidad de neuronas de la red.	35
12	Comparación de la predicción de la <i>ATSM</i> para 1998 a partir de una serie de tiempo por tres métodos diferentes. Para las RNA el coef. de correlación resulta de 0.979, para el GCM de 0.832 y para el CMF de 0.841.	37

Índice de Tablas

I	Parámetros de la arquitectura de la red y archivos de entrenamiento/prueba que se variaron en los experimentos E1-E15, para predecir H_S	27
II	Parámetros de la arquitectura de la red y archivos de entrenamiento/prueba que se variaron en los experimentos E1-E11, para predecir el IOS	32

PREDICCIÓN DE SERIES DE TIEMPO APLICANDO REDES NEURONALES ARTIFICIALES

I INTRODUCCIÓN

I.1 Presentación del problema

El deseo de predecir el futuro y conocer el pasado nos motiva a investigar las leyes que explican el comportamiento de ciertos fenómenos. Si el sistema que se desea comprender es determinista no-caótico y se conoce la ecuación que lo describe, no hay problema ya que ésta se puede resolver para obtener una predicción, conociendo las condiciones iniciales. Sin embargo, si no se conoce la ecuación, entonces se deben encontrar las reglas que gobiernan la evolución del sistema y su estado actual. En este trabajo nos enfocaremos en problemas cuya ecuación no existe, por lo tanto las leyes que determinan la evolución del sistema deben ser inferidas a partir de regularidades en el mismo.

Entre los fenómenos físicos más sorprendentes y de mayor belleza en la naturaleza, se encuentra el oleaje, que en forma sencilla se puede definir como ondas en la superficie del mar, con periodos que van desde 1 a 10 seg. y longitudes de onda desde algunos metros hasta cientos de ellos. El agente generador de las olas es el viento y la fuerza restauradora es la asociada a la gravedad. En presencia del viento, el oleaje se conoce como local y cuando se propaga a través de grandes distancias, se le llama mar de leva.

El régimen de oleaje ha sido asociado a diversos aspectos de interés. La ocurrencia de tormentas intensas posteriores al evento de El Niño en 1982, influyó en la presencia de condiciones extremas del oleaje en las costas de California y Baja California provocando grandes daños (Seymour, Strange, Cayan & Nathan 1984). Asimismo, el oleaje que se genera por dichas tormentas extremas, puede producir velocidades orbitales altas que aunadas al rompimiento de olas muy energéticas, influyen en forma decisiva en la permanencia o destrucción de bosques de alga café (*Macrocystis*), como en el caso de

la región del sur de California (Seymour, Tegner, Dayton & Pamell 1989). Otro aspecto importante que se observó a partir de mediciones de oleaje en el Atlántico Norte, fué una tendencia de crecimiento gradual a largo plazo, de la altura típica del oleaje, que se puede asociar al cambio climático global (Bacon & Carter 1991).

Así pues, la predicción del oleaje es de gran importancia tanto para el estudio profundo de los fenómenos anteriores, como en la seguridad marítima en regiones costeras y actividad pesquera, para el diseño de estructuras de ingeniería, evaluación y prevención de la erosión de playas, transporte de sedimento, etc. Uno de los parámetros más importantes del oleaje, en todas estas aplicaciones, es la altura significativa H_S que se define como el valor promedio de las alturas de la tercera parte de las olas más altas medidas durante un cierto periodo. Es por esto que surgió el interés de intentar encontrar una metodología adecuada para el pronóstico de H_S .

Otro parámetro, cuya predicción ha tomado gran importancia en la última década, es el índice de oscilación del sur (*IOS*), que representa la diferencia de presión al nivel del mar entre Darwin (Australia) y Tahití. Se sabe que esta oscilación está asociada con cambios en la circulación atmosférica en varias partes del mundo. Como consecuencia, existen cambios en los patrones de lluvia y campos de viento en los océanos Pacífico e Índico tropical que se correlaciona con fluctuaciones meteorológicas en otras partes del globo (Philander 1990). Este parámetro nos permite estimar de alguna manera, la existencia del fenómeno de El Niño Oscilación Sur (*ENOS*), que se presenta cuando una alta presión superficial sobre el poniente y una baja sobre el sureste del Pacífico tropical, coincide con lluvia abundante, temperaturas superficiales tibias poco usuales y vientos relajados en el centro y oriente del Pacífico tropical. Dado que esta condición atmosférica se correlaciona fuertemente con la lluvia en diferentes lugares del mundo y con variaciones en la temperatura global que pueden afectar la pesca y la agricultura, es importante determinar dicha condición, pues esto permitiría prevenir desastres globales. Por lo tanto es importante predecir tanto el *IOS* como las

anomalías en la temperatura superficial del mar (*ATSM*) para conocer la probabilidad de que se pueda dar la condición de El Niño.

I.2 Objetivos

El objetivo de este trabajo consiste en aplicar las Redes Neuronales Artificiales (RNA) a la predicción de series de tiempo, en particular de altura significativa de oleaje H_S , del índice de oscilación del sur *IOS* y de las anomalías de temperatura superficial del mar *ATSM*.

I.3 Organización del trabajo

El trabajo está organizado de la siguiente forma. En el capítulo II se describen en forma breve los modelos de predicción de series de tiempo más comunes, haciendo distinción entre los modelos lineales y los no lineales. En el capítulo III se presenta una descripción de las Redes Neuronales Artificiales (RNA), comparando el modelo biológico con el cibernético, intentando lograr una mejor comprensión del tema. Posteriormente se explica en qué consiste el entrenamiento de una RNA y se presentan diferentes arquitecturas, enfatizando en la que corresponde al tipo de respuesta finita al impulso (RFI), que es la que se aplica en este trabajo. En el capítulo IV se muestran los experimentos que se llevaron a cabo con las series de tiempo de H_S , del *IOS* y de *ATSM* y se presentan los resultados que se obtuvieron en las predicciones con una breve discusión al final de cada tema. Finalmente, en el capítulo V se describen las conclusiones más importantes del trabajo.

II PREDICCIÓN DE SERIES DE TIEMPO

II.1 Modelos de predicción de series de tiempo

El análisis de series de tiempo tiene tres metas principales, la predicción, el modelado y la caracterización. La predicción consiste en pronosticar la evolución del sistema a corto plazo, el modelado en encontrar una descripción que capture las características principales del comportamiento del sistema a largo plazo y la caracterización en determinar en base a un conocimiento *a priori* las propiedades fundamentales del sistema.

Antes de 1920 la predicción se hacía sólo extrapolando la serie en el tiempo. Posteriormente se propuso una técnica autoregresiva (Yule 1927), que predecía los valores siguientes en función de una suma ponderada de los valores anteriores. El avance en la predicción de series de tiempo se dió en los años ochenta con el desarrollo de dos técnicas. La reconstrucción de espacios de estado mediante el anidado de retardos de tiempo, que provee una forma de reconocer cuando una serie de tiempo fué generada por un sistema determinista y si así fue, entender la estructura geométrica que envuelve el comportamiento observado, dicha técnica se basa en la topología diferencial y en sistemas dinámicos. La segunda, que surge en la Inteligencia Artificial con los sistemas de aprendizaje, en particular, las Redes Neuronales Artificiales, que exploran en forma adaptable un gran espacio de modelos potenciales.

A continuación revisaremos en forma breve los modelos de análisis para series de tiempo.

II.2 Modelos lineales

Las ventajas principales de estos modelos son que se pueden comprender con gran detalle y que se pueden aplicar con relativa facilidad. Su desventaja es que no son capaces de describir un sistema moderadamente complicado. Supondremos que los sistemas que revisaremos son lineales y estacionarios. Básicamente existen dos tipos

de modelos, aquellos que son función solo de entradas externas y los que dependen de la memoria interna (autoregresivos).

II.2.1 Modelos de promedio corrido (MPC).

Consideremos un sistema cuya entrada es una serie de tiempo e_t y se desea generar otra serie a la salida x_t , si suponemos linealidad y causalidad, es decir que la salida depende del presente y de N valores anteriores, la relación entrada-salida está dada por

$$x_t = \sum_{n=0}^N b_n e_{t-n} = b_0 e_t + b_1 e_{t-1} + \dots + b_N e_{t-N} \quad (1)$$

donde la nueva serie se genera por un filtro de N -ésimo orden con coeficientes $b_0, b_1, \dots, b_N$ a partir de la serie e_t . Los estadísticos conocen este modelo como un promedio corrido (*moving average*) de N -ésimo orden, cuyo comportamiento es semejante al de un filtro que suaviza los últimos valores de e_t . En ingeniería esto se conoce como un filtro de respuesta finita al impulso (FRFI) pues la salida convergerá a cero N pasos después de que la entrada sea cero.

Los MPC presentan características determinadas en el dominio del tiempo, de la frecuencia (espectro) y en términos de sus coeficientes de autocorrelación. En el primer caso se supone que la entrada es diferente de cero solo en un tiempo t_0 y es cero para cualquier otro tiempo. La respuesta a este impulso está dada por los coeficientes. Para cada paso de tiempo, el impulso se mueve hacia el siguiente coeficiente, hasta que después de N pasos la respuesta se hace cero. La serie $b_N, b_{N-1}, b_{N-2}, \dots, b_1, b_0$ es la respuesta del sistema al impulso. Por lo tanto, la respuesta a una entrada arbitraria se calcula superponiendo las respuestas a retardos apropiados ponderados por sus respectivos valores de entrada (convolución). De esta manera, la función de transferencia describe un sistema lineal, ya que satisface el principio de superposición.

En el dominio de la frecuencia, la representación es más simple ya que una convolución equivale a un producto. Si la entrada a un MPC es un impulso (espectro plano) la transformada discreta de Fourier de la salida está dada por

$$\sum_{n=0}^N b_n e^{-i2\pi n f} \quad (2)$$

y el espectro de potencia está dado por la magnitud del cuadrado de la ecuación anterior. Los coeficientes de autocorrelación ρ_τ se definen en función de la media $\mu = \langle x_t \rangle$ y la varianza $\sigma^2 = \langle (x_t - \mu)^2 \rangle$ mediante

$$\rho_\tau = \frac{1}{\sigma^2} \langle (x_t - \mu)(x_{t-\tau} - \mu) \rangle \quad (3)$$

en donde $\langle \rangle$ denota esperanza. Los coeficientes de autocorrelación describen en promedio, el grado de similitud entre valores separados τ intervalos de tiempo. Si la entrada al sistema es un proceso estocástico con valores de entrada a tiempos diferentes, no correlacionados $\langle e_i e_j \rangle = 0$ para $i \neq j$, entonces todos los términos cruzados desaparecen de la ecuación (3) y los coeficientes de autocorrelación están dados por

$$\rho_t = \begin{cases} \frac{1}{\sum_{n=0}^N b_n^2} \sum_{n=\tau}^N b_n b_{n-|\tau|} & |\tau| \leq N \\ 0 & |\tau| > N \end{cases} \quad (4)$$

II.2.2 Modelos Autoregresivos (MAR)

La diferencia principal con respecto a los filtros RFI, es que éstos operan sin retroalimentación, mientras que los modelos autoregresivos contienen lazos de retroalimentación, lo cual es equivalente a tener memoria que es lo que les permite generar la dinámica interna de un proceso, así pues

$$x_t = \sum_{m=1}^M a_m x_{t-m} + e_t \quad (5)$$

se conoce como modelo autoregresivo de emésimo orden o como un filtro de respuesta infinita al impulso (FRII) donde la respuesta puede continuar aunque la entrada cese. Generalmente e_t es una entrada controlada. Para este tipo de filtros, encontrar los coeficientes de la autocorrelación implica resolver un sistema de ecuaciones para cada uno de ellos, debido a los lazos de retroalimentación de este modelo. Los coeficientes se encuentran mediante la ecuación Yule-Walker

$$\rho_\tau = \sum_{m=1}^M a_m \rho_{\tau-m} \quad \tau > 0 \quad (6)$$

En forma alternativa, los coeficientes se pueden estimar como un problema regresivo, expresando el valor siguiente de la serie como una función ponderada de los M valores anteriores. Esto se puede hacer de tal forma que la suma de los cuadrados de las diferencias entre la respuesta del modelo y el valor observado sobre todos los pasos de tiempo, sea mínima.

La combinación lineal de los MPC y los MAR se conoce como modelo autoregresivo de promedio corrido (MARPC). Estos modelos por ser más completos, han dominado todas las áreas en el análisis de series de tiempo y en el procesamiento de señales discreto, por ejemplo en el reconocimiento y síntesis de voz, codificación predictiva, compresión de datos, etc.

Con respecto al orden del modelo, no existe ninguna técnica para determinar cual es el orden más adecuado para el ajuste de una serie de tiempo, aparentemente mientras mayor es el orden, mejor es el ajuste. Sin embargo, en un punto, el error en el ajuste aumenta debido a que el modelo empieza a ajustar ruido en el sistema. Existen algunas técnicas heurísticas para aproximar el orden del modelo, aunque la forma más común de estimar el orden, es mediante prueba y error.

Ahora bien, un modelo de este tipo será una buena elección si el espectro de potencia es una representación útil de alguna característica relevante de la serie de tiempo, en caso contrario, los modelos lineales pueden fallar aún para sistemas con pequeñas no-linealidades. Por lo tanto se hacen necesarios los modelos no-lineales.

II.3 Modelos no-lineales

II.3.1 Espacios de estado

Entre las técnicas más comunes para modelar sistemas no-lineales, se encuentra la reconstrucción de espacios de estado, que se basa en el teorema de Takens (Takens 1981), quien demostró que a partir de un vector suficientemente largo, construido con valores anteriores de una serie de tiempo, es posible reconstruir la estructura de la dinámica del sistema que produjo la serie temporal. Sin embargo, si consideramos un gran número de valores, esto nos acarrea un gran número de grados de libertad. Afortunadamente no es necesario que un sistema se analice en el espacio completo ya que se puede analizar en un subespacio para simplificar el problema.

II.3.2 Redes neuronales artificiales

Las redes neuronales artificiales (RNA) representan una técnica más reciente. Si se desea aplicar esta nueva técnica para modelar un sistema que varía en el tiempo, es necesario saber o al menos sospechar, que el sistema presenta no-linealidades. Se dice que un sistema tiene un comportamiento no-lineal, si diverge a partir de dos puntos cercanos, con una rapidez no constante. El sistema puede ser no-lineal sin ser caótico pero se convierte en caótico cuando las no-linealidades son tales que el sistema diverge a partir de dos estados próximos a una razón exponencial expresada como 2^λ donde λ es el exponente de Lyapunov del sistema (Parker & Chua 1989).

Para aplicar las redes a la predicción de una serie de tiempo, se puede partir como en el caso anterior, de una representación del sistema en función de un vector generado

con valores anteriores de la serie $[x(k-1), x(k-2), \dots, x(k-N)]$, a este vector se le aplica una función no lineal g para encontrar el siguiente valor de la serie

$$y(k) = g[x(k-1), x(k-2), \dots, x(k-N)]$$

En general la función no lineal g puede ser modelada por una red neuronal. En el siguiente capítulo se describen las bases y algunas propiedades importantes de las RNA.

III REDES NEURONALES ARTIFICIALES

III.1 Introducción

Las redes neuronales artificiales son modelos estadísticos que realizan una transformación entrada/salida mediante el ajuste de un conjunto de parámetros llamados pesos. Entre las aplicaciones más comunes de las RNA, están: el reconocimiento de patrones, donde una colección de características se le muestran a la red y su tarea consiste en asignar una característica de entrada, a una o más clases; regresiones no lineales, que consisten en encontrar una interpolación suavizada entre puntos; procesamiento de señales, por ejemplo, extracción de características, desarrollo de filtros, reconocimiento de voz, compresión de datos, etc.; modelos financieros, aproximación de funciones, control, predicción de series de tiempo, etc.

Una de las ventajas más relevantes que presentan las RNA sobre los modelos estadísticos tradicionales radica en que el tipo de restricciones que imponen las redes, son menos rigurosas. Por ejemplo, consideremos una regresión polinomial cuyas componentes del vector de entrada $(x_1, x_2, x_3, \dots, x_d)$ se pueden tomar en pares, triadas, etc. así como en potencias más altas. El gran número de combinaciones posibles, puede aproximar cualquier superficie de salida deseada, esto podría interpretarse como que las redes son capaces de representar el espacio de una función más grande con un número menor de parámetros. Sin embargo esto no es cierto, se ha demostrado (Cover 1965) que la capacidad de ambos, el desarrollo polinomial y las RNA, es proporcional al número de parámetros. La ventaja principal de un modelo con respecto al otro consiste en el tipo de restricciones que imponen las redes. Para el caso polinomial, el número posible de términos crece rápidamente con la dimensión de la entrada, lo que a veces hace imposible usar los términos de segundo orden o las interacciones de dos o más componentes. Una red con capas ocultas, limita solo el número total de interacciones, más no el orden de ellas, además aprende a seleccionar una combinación

adecuada de entradas (Weigend & Gershenfeld 1994).

III.2 Modelo biológico

III.2.1 El sistema nervioso central

Las Redes Neuronales Artificiales (RNA), surgen como una idea basada en el comportamiento de las redes neuronales biológicas. A continuación se hará una breve revisión de la estructura y el funcionamiento del Sistema Nervioso Central (SNC) y en particular del proceso de transferencia de información entre neuronas.

El SNC se forma aproximadamente de 10^{11} células llamadas neuronas (Figura 1). Hace alrededor de 50 años, se descubrió que todas las neuronas se constituyen de tres partes básicas. La parte central, se llama cuerpo o soma, de donde se generan algunas extensiones llamadas dendritas y una fibra tubular única llamada axón que se ramifica en su extremo. Las dendritas funcionan como receptores de señales de neuronas adyacentes y el axón como el transmisor de la señal generada de acuerdo a su actividad, hacia otra neurona o hacia un músculo. La longitud total de una neurona varía enormemente, desde 0.01mm -neuronas en el interior del cerebro- hasta 1m -neuronas en los músculos-.

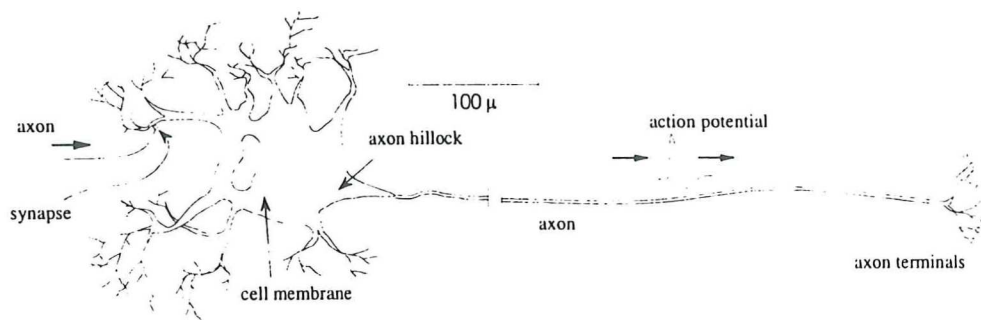


Figura 1: La neurona y sus partes principales

Uno de los procesos más importantes en el SNC es la comunicación entre neuronas. Las señales entre ellas se transmiten en forma eléctrica y mediante una reacción química. La transmisión eléctrica se da en el interior de la neurona, mientras que la interacción entre ellas se efectúa mediante mecanismos químicos llamados sinapsis. La transmisión eléctrica se lleva a cabo mediante una descarga eléctrica que se inicia en el cuerpo de la célula y viaja por el axón. Estas descargas se producen por el intercambio de cargas que se establece a través de la membrana de la neurona entre el interior de la misma y el líquido que la rodea. Cuando la neurona está en reposo, tiene un potencial de -70mV con respecto al exterior, que se mantiene debido a que la membrana es impermeable al flujo de iones de Na^+ . Cuando la neurona empieza a recibir señales debido a conexiones sinápticas, el potencial comienza a disminuir y cuando llega a -60mV , repentinamente la membrana se hace permeable a los iones de Na^+ . Estos iones fluyen hacia el interior, neutralizan la diferencia de potencial y la hacen ligeramente positiva con respecto a su exterior. Gradualmente la membrana recupera sus propiedades originales y regenera su potencial de reposo en el interior, esto sucede en un periodo de algunos milisegundos. Durante este periodo, la neurona es incapaz de excitarse nuevamente, debido a esto, resulta evidente que el pulso eléctrico se propaga siempre en una sola dirección que se aleja del cuerpo.

De lo anterior se podría concluir que la transmisión de la señal en el SNC es del tipo digital binaria, ya que la neurona puede estar o no activa. Sin embargo, esto es erróneo, pues características de las señales nerviosas, como el hecho de que estén codificadas en frecuencia, o que el intervalo entre dos picos pueda tomar cualquier valor (mayor que el periodo de regeneración), determina que la transmisión se puede considerar analógica.

Una vez que la señal eléctrica llega al extremo del axón, se lleva a cabo la sinapsis. Este proceso químico se inicia al liberar ciertas sustancias especiales llamadas neurotransmisores, que llegan hasta la neurona postsináptica y modifican la conductancia de la membrana para ciertos iones (Na^+ , K^+ , Cl^- , etc.) los cuales fluyen hacia o desde

dentro de la neurona, causando un cambio ΔU en el potencial postsináptico. Si ΔU es positivo, el potencial se reduce, y se dice que la sinapsis fue excitante, si sucede lo contrario la sinapsis es inhibitoria.

Sin embargo, el que una sinapsis sea excitante no significa que esto sea suficiente para activar la neurona, depende también de la intensidad de esta sinapsis, de la localización donde se da ésta con respecto al cuerpo y de la cantidad de señales que llegan. Se sabe además que la intensidad de la sinapsis no es constante, sino que cambia y estos cambios están directamente relacionados con el aprendizaje.

III.3 Modelo cibernético

Existen diversas definiciones de lo que actualmente se concibe como una RNA, en esta sección se intenta presentar una que sea completa y simple.

Una Red Neuronal Artificial es un sistema de procesamiento de información distribuida en paralelo. Consiste de elementos de procesamiento (ep) simples y adaptables, interconectados entre sí. Su objetivo es interactuar con los objetos del mundo real en forma semejante como lo hace el cerebro.

En forma rigurosa y desde un punto de vista matemático, una RNA es un grafo dirigido donde cada nodo i corresponde a un elemento de procesamiento (ep), al que se le asocia una variable de estado n_i y un valor real de polarización ϕ_i . Los lazos entre los nodos i, k se llaman conexiones, a cada una de ellas se le asigna un peso w_{ik} de valor real. Se define una función de transferencia $f_i(n_i, w_{ik}, \phi_i, k \neq i)$ para cada ep, que determina el estado del ep como una función de su polarización, de los pesos de las conexiones de entrada y de los estados de otros ep's conectados. En terminología biológica los nodos se llaman neuronas, las conexiones sinapsis y las polarizaciones umbrales de activación.

Esquemáticamente, las RNA se pueden representar como se muestra en la figura 2.

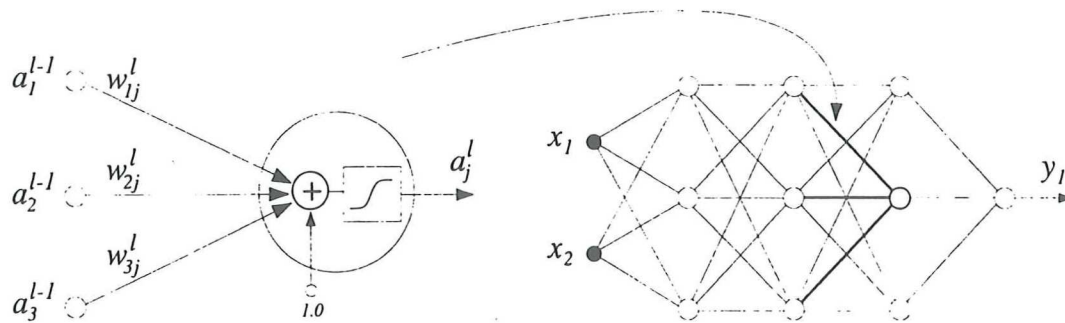


Figura 2: Esquema de una Red Neuronal Artificial. Las a son las señales de activación procedentes de otras neuronas, las w son las ponderaciones de los lazos entre neuronas, las x son las señales de entrada y la y la salida. Cada línea representa una conexión sináptica.

Para interpretar mejor la simulación de una neurona, se divide el proceso de comunicación de señales en cuatro subprocesos; **propagación de la señal** en el axón, **sinapsis**, **integración espacio-temporal** en la dendrita y **generación del potencial de activación**.

La comunicación se inicia en el cuerpo de la neurona, donde se genera una serie de impulsos eléctricos que viajan por el axón. Esta serie de impulsos no tiene ni frecuencia ni periodo constante. La conducción se lleva a cabo gracias al rápido flujo de iones de sodio y potasio a través de la membrana lo cual inicia la **propagación de la señal** por el axón, que actúa como un elemento resistivo propagando una señal pasiva (Wan 1993).

Una vez que la señal llega a las dendritas la neurona presináptica libera sustancias químicas llamadas neurotransmisores, que viajan a través de la separación entre las neuronas presináptica y postsináptica, la **sinapsis** se puede considerar como un modulador de conductancia que inhibe o excita y que genera un potencial postsináptico. Esto se puede modelar como un simple circuito que consista de una trayectoria conductiva, dependiente del tiempo e independiente del voltaje. En estado estacionario, el

potencial postináptico es proporcional (en forma burda) al estímulo de entrada, lo que nos da una idea de ponderación. Como respuesta a la sinapsis, se genera un impulso que continúa su propagación a través de las dendritas de la neurona receptora. Este proceso de nuevo se puede simular como la propagación de una señal pasiva a través de una red de cables. El cálculo que se efectúa en el cuerpo de la neurona receptora, equivale a una **integración espacio-temporal** de todas las señales que llegan, lo cual podría considerarse como un proceso lineal ya que la propagación de una señal pasiva es inherentemente lineal. Finalmente, si el potencial que está presente en la parte inicial del axón excede un cierto umbral, genera el **potencial de acción**, que viajará por el axón hacia otra conexión sináptica. Este proceso se puede explicar con canales con voltajes conmutados en los cuales ocurre una conmutación rápida de permeabilidades iónicas y matemáticamente se puede simular con un conjunto de ecuaciones acopladas no-lineales.

Obviamente, si intentamos incluir todas las ecuaciones que modelan cada uno de los procesos anteriores, resultaría un sistema muy complicado, es por eso que se prefiere presentar un modelo lo más simple posible. Así pues, la respuesta s de una sola neurona se puede simular como una función no-lineal de la suma ponderada de las entradas

$$s = \sum_{i=1}^N w_i x_i + w_b \quad (7)$$

donde las x_i son las entradas y los w_i son los pesos.

Luego

$$y = f(s) \quad (8)$$

es la función no-lineal, que generalmente es una sigmoide;

$$f(s) = \frac{1}{e^{-s} + 1} \quad (9)$$

Esta aproximación es bastante burda con respecto al modelo biológico, todo el proceso biológico, no es para nosotros mas que una simple multiplicación wx y la dinámica para generar el potencial de activación, se reduce a una función sin memoria $y = f(s)$. De hecho, no existe una dinámica interna asociada con el modelo ingenieril, es por eso que este modelo solo captura las propiedades distribuidas espacialmente, las relaciones temporales son despreciadas.

III.3.1 Entrenamiento y retropropagación de errores

Ahora bien, una sola neurona no tiene ninguna utilidad, es por eso que se construyen redes, que representan neuronas interconectadas. De acuerdo a la forma de propagación de la información, la red más conocida es la *feedforward* o de alimentación en avance (AA), su nombre se debe a que la información se propaga en un solo sentido, ya que la salida de una neurona en una cierta capa, actúa como entrada para otras neuronas en la siguiente capa y además, no existen lazos de retroalimentación.

La red AA realiza una transformación compleja $y = N(\mathbf{w}, \mathbf{x})$ del vector de entrada $\mathbf{x}$ en la primera capa al vector de salida y de la última capa, ponderada por los pesos $\mathbf{w}$, esta transformación queda determinada por los valores del vector $\mathbf{w}$. En general, el entrenamiento de una red consiste en encontrar el conjunto de pesos que producen la mejor transformación entrada/salida, en términos de una función de error ε

$$\varepsilon = \sum_i (d_i - y_i)^2$$

Donde d_i es la salida deseada y y_i es la salida de la red.

Para su entrenamiento, la red se alimenta con un conjunto de vectores de entrada,

cada uno asociado con un vector de salida deseada, $\{(\mathbf{x}_1, \mathbf{d}_1), (\mathbf{x}_2, \mathbf{d}_2), \dots, (\mathbf{x}_P, \mathbf{d}_P)\}$. El error $\mathbf{e}$ se calcula como la diferencia entre la salida deseada y la salida real de la red

$$\mathbf{e} = \mathbf{d} - \mathbf{y} \quad (10)$$

El error total cuadrático sobre todo el conjunto de patrones P , está dado por

$$J = \sum_{p=1}^P \mathbf{e}_p^T \mathbf{e}_p \quad (11)$$

El objetivo del entrenamiento consiste en determinar el conjunto de pesos $\mathbf{w}$, que minimizan la función de costo $J(\mathbf{w})$. Para esto se aplica el algoritmo de gradiente descendente (Hertz et al, 1991) que consiste en encontrar un conjunto de pesos $\mathbf{w}$ mejorado. Esto se puede lograr, visualizando la función de error como una superficie n -dimensional definida por $\mathbf{w}$ y deslizándose sobre ella hacia un mínimo, con una velocidad proporcional a la pendiente de la superficie. Es decir, cada peso se incrementa por una cantidad $\Delta \mathbf{w}$ proporcional al gradiente del error ∇ , mediante

$$\Delta \mathbf{w} = -\mu \nabla \quad (12)$$

donde $\nabla = \frac{\partial \mathbf{e}^T \mathbf{e}}{\partial \mathbf{w}}$ es el gradiente del error instantáneo que corresponde al patrón de entrada actual y μ es la razón de aprendizaje. Ahora, para determinar los valores de los pesos se utiliza el algoritmo de retropropagación, el cual se describe a continuación. Por claridad, consideremos una sola neurona lineal, descrita completamente por

$$y = \sum_{i=0}^N w_i x_i = \mathbf{w}^T \mathbf{x} \quad (13)$$

Donde $\mathbf{w} = [w_0, w_1, \dots, w_N]$, $\mathbf{x} = [x_0, x_1, \dots, x_N]$ y

$$\begin{aligned}
\nabla &= \frac{\partial e^2}{\partial \mathbf{w}} \\
&= 2e \frac{\partial (d - y)}{\partial \mathbf{w}} \\
&= -2ex
\end{aligned} \tag{14}$$

Por lo tanto, la actualización de los pesos está dada por

$$\Delta \mathbf{w} = 2\mu ex \tag{15}$$

o en forma escalar $\Delta w_i = 2\mu ex_i$.

Para el caso de una red con l capas, para calcular el gradiente con respecto al peso w_{ij}^l , se aplica la regla de la cadena

$$\frac{\partial \mathbf{e}^T \mathbf{e}}{\partial w_{ij}^l} = \frac{\partial \mathbf{e}^T \mathbf{e}}{\partial s_j^l} \frac{\partial s_j^l}{\partial w_{ij}^l} = \delta_j^l a_i^{l-1} \tag{16}$$

considerando a s_j^l como el punto de suma, a_i^{l-1} el valor de activación que actúa como una entrada al peso w y donde se define

$$\delta_j^l \triangleq \frac{\partial \mathbf{e}^T \mathbf{e}}{\partial s_j^l} \tag{17}$$

Por lo tanto, la ecuación de actualización está dada por

$$\Delta w_{ij}^l = -\mu \delta_j^l a_i^{l-1} \tag{18}$$

que es válido para cualquier peso en la red y además hay una δ asociada a cada neurona, cuyos valores se encuentran en forma recursiva, empezando en la capa de salida

$$\begin{aligned}
\delta_j^L &\triangleq \frac{\partial \mathbf{e}^T \mathbf{e}}{\partial s_j^L} \\
&= \frac{\partial \mathbf{e}^T \mathbf{e}}{\partial a_j^L} \frac{\partial a_j^L}{\partial s_j^L} \\
&= \frac{\partial \mathbf{e}^T \mathbf{e}}{\partial y_j} f'(s_i^L)
\end{aligned} \tag{19}$$

donde $f'(s_i^L)$ es la derivada de la función sigmoide evaluada en el punto de suma y se substituyó $y_j = a_j^L$. Luego

$$\begin{aligned}
\frac{\partial \mathbf{e}^T \mathbf{e}}{\partial y_j} &= \frac{\partial e_j^2}{\partial y_j} \\
&= 2e_j \frac{\partial (d_j - y_j)}{\partial y_j} \\
&= -2e_j
\end{aligned} \tag{20}$$

Substituyendo en la ecuación anterior se obtiene:

$$\delta_j^L = -2e_j f'(s_j^L) \tag{21}$$

En resumen, los términos de actualización (deltas) que se obtienen por el algoritmo de retropropagación son:

$$\Delta w_{ij}^l = -\mu \delta_j^l a_i^{l-1} \tag{22}$$

$$\delta_i^l = \begin{cases} -2e_i f'(s_i^L) & l = L \\ f'(s_i^l) \sum_j \delta_j^{l+1} w_{ij}^{l+1} & 1 \leq l \leq L-1 \end{cases} \tag{23}$$

Este proceso de entrenamiento consiste en tomar un patrón $\mathbf{x}$ del conjunto de entrenamiento y ponerlo a la entrada de la red, la entrada causa una distribución de

valores de activación a través de la red y un vector de salida y . La salida se compara con la respuesta deseada d para formar el vector de error e . El gradiente del error cuadrático se retropropaga recursivamente a través de la red y en seguida los pesos se actualizan de acuerdo al algoritmo de gradiente descendente. Este proceso se repite para todos los patrones del conjunto de entrenamiento hasta que se alcanza un error aceptable (Hertz, Krogh & Palmer 1991).

III.3.2 Arquitecturas

Existe un gran número de arquitecturas de redes neuronales que se aplican a diferentes problemas, pero no hay nada que establezca con precisión qué tipo de arquitectura es adecuada para resolver tal o cual problema. Sin embargo, se sabe por referencias anteriores que para resolver problemas de tipo temporal, las arquitecturas más usadas son la Recurrente (RNR), la de Retardo en el Tiempo (RNRT), (Kevin 1996) y más recientemente la de Respuesta Finita al Impulso (RNRFI)(Weigend & Gershenfeld 1994).

Las redes neuronales con retardo en el tiempo, son un grupo de redes que tienen una topología especial, generalmente se usan para el reconocimiento de características independientes dentro de un patrón de gran tamaño. La activación de una unidad, normalmente se calcula aplicando la función de activación -usualmente una sigmoide- a la suma ponderada de sus entradas. Sin embargo para el caso de las RNRT's este procedimiento se modifica mediante la introducción de retardos. Ahora, cada una de las entradas de una unidad de procesamiento consiste de un vector con valores anteriores de la serie de tiempo, que se puede generar usando una ventana temporal que se desplaza sobre toda la serie o alimentando la señal a una línea de retardo. En realidad todo este proceso es equivalente a convertir una serie temporal en un patrón espacial. Por lo tanto la construcción de una RNRT completa, puede relacionar entradas en diferentes puntos tanto en el espacio como en el tiempo.

El entrenamiento de este tipo de redes se realiza mediante un algoritmo muy similar al de retropropagación, con algunas modificaciones que consideran la estructura de los lazos acoplados. Para realizar este proceso, se le presentan a la capa de entrada una sucesión de vectores desplazados en el tiempo, dentro de los patrones. Además, ya que cada una de las unidades de respaldo se duplica para cada desplazamiento en el tiempo, se dispone de la historia completa de las activaciones en la red.

Las redes recurrentes, tienen algunas características que las distinguen de las AA, la principal es que se pueden usar para representar el tiempo implícitamente. Entre las arquitecturas recurrentes más conocidas se encuentra la de Elman, que supone que la red opera en pasos discretos en el tiempo (Elman 1991). La salida de las unidades ocultas en un tiempo t es retroalimentada como entrada adicional a la red, en un tiempo $t + 1$ para almacenar la salida de las unidades ocultas, Elman introdujo el concepto de unidades de contexto que representan una especie de memoria a corto plazo. El algoritmo de aprendizaje de las RNR es la retropropagación, con algunas pequeñas variantes. Este tipo de redes generalmente se aplica para procesar datos que varían en el tiempo, para predecir valores futuros, clasificar series de tiempo, predecir el comportamiento de un sistema, etc.

Finalmente, las RNRFI son redes muy semejantes a las RNRT, ya que cada uno de los filtros representan retardos de la señal de entrada, dependiendo del orden del filtro, estas dos arquitecturas son totalmente equivalentes cuando las RNRFI se desdoblán en el tiempo, eliminando de esta manera la dependencia temporal. A continuación daremos una descripción más completa de este tipo de redes, ya que estas fueron las que dieron mejores resultados para la predicción de las series de tiempo que se revisaron en este trabajo.

III.4 Redes Neuronales del tipo RFI

Regresando al modelo biológico, recordemos que la señal que se propaga a través del axón se modula por una sinapsis y luego se disipa a través de las dendritas y la membrana. Este proceso corresponde a una transmisión lineal, por lo tanto la ecuación del potencial somático se representa por

$$s(t) = \sum_i w_i(t)x_i(t) + w_b \quad (24)$$

donde la suma se realiza sobre todas las señales que alimentan la neurona, $s(t)$ sigue siendo como la unión de suma y w_b se puede interpretar como el potencial de reposo. En la sumatoria, $x_i(t)$ corresponde al potencial de acción de entrada, modificado por $w_i(t)$. Para modelar este proceso en una forma simple, se ignoran las dimensiones físicas, las diferencias en los retardos que corresponden a los diferentes tamaños de axón, localización de las sinapsis, etc. En el modelo tradicional, el punto de unión, corresponde a una simple suma ponderada estática, proporcionando solo la representación espacial de las señales que llegan de diferentes neuronas, al agregar el filtrado, se está incluyendo la dimensión temporal fundamental, pues cada filtro corresponde a una línea de retardo.

Para adaptar los filtros a la red, se supone que cada uno de ellos se comporta como una respuesta finita al impulso y discreta, es decir

$$s(k) = \sum_i \sum_{n=0}^M w_i(n)x_i(k-n) + w_b = \sum_i \mathbf{w}_i \cdot \mathbf{x}_i(k) + w_b \quad (25)$$

donde k es el índice discreto de tiempo para alguna razón de muestreo Δt y $w_i(n)$ corresponde a los coeficientes de los filtros RFI. En forma compacta

$$\mathbf{w}_i = [w_i(0), w_i(1), \dots, w_i(M)] \quad (26)$$

y el estado retardado

$$\mathbf{x}_i(k) = [x_i(k), x_i(k-1), \dots, x_i(k-M)] \quad (27)$$

El filtrado se realiza mediante el producto punto $\mathbf{w}_i \cdot \mathbf{x}_i(k)$ donde la relación temporal está implícita en la notación. La figura 3 muestra una representación típica de un filtro RFI como una línea de retardo.

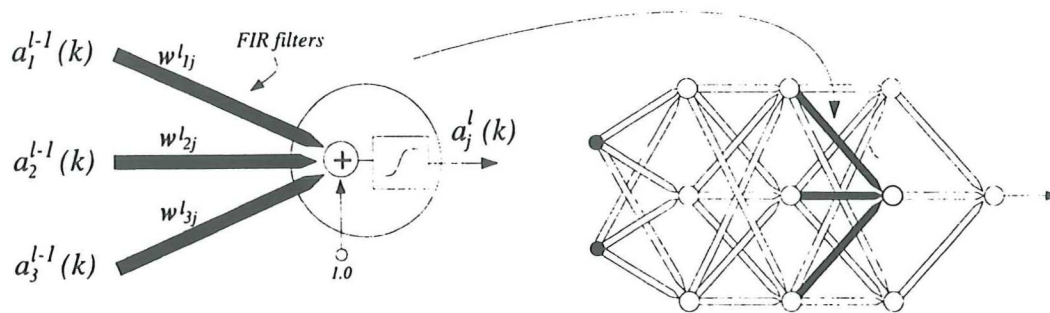


Figura 3: Arquitectura tipo RFI. En el modelo espacio-temporal de la neurona, las señales de entrada $a(k)$ se pasan a través de filtros sinápticos $\mathbf{w}$. Todas las conexiones se modelan con filtros de orden k tipo RFI

La contraparte biológica de este filtro sería la sinapsis. Para completar el modelo, tenemos que la salida de la neurona es:

$$y(k) = f(s(k)) \quad (28)$$

donde, generalmente $f() = \tanh()$, y solo se ha agregado el índice temporal k . Ahora, la salida $y(k)$ representa un promedio a corto plazo del potencial de acción y no el potencial de acción en sí.

IV EXPERIMENTOS Y RESULTADOS

IV.1 Predicción de series de tiempo

IV.1.1 Introducción

Recientemente se llevó a cabo una comparación de la predicción de diferentes series temporales con diversas técnicas (Weigend & Gershenfeld 1994). En general, la metodología con más éxito fue la que utiliza Redes Neuronales Artificiales. Esto no significa que dicho método sea el mejor para cualquier tipo de serie, sino que las RNA pueden ser mejores que las otras técnicas para ciertas series temporales. Este hecho fue una de las motivaciones para intentar aplicar las RNA a la predicción de series de tiempo oceanográficas, en particular la altura significativa del oleaje (H_S), el índice de oscilación del sur (IOS) y la anomalía de temperatura superficial del mar ($ATSM$).

IV.1.2 Predicción de H_S

Se utilizó una serie temporal de estimaciones de la altura significativa (H_S), que se obtuvo a partir de las mediciones de oleaje realizadas en San Clemente CA.(Fig. 4a), con un arreglo de alambres, correspondiente a los años de 1995 y 1996 (Scl-95 n.d.). En la Fig. 4b, se muestra una serie completa en donde los fragmentos realzados son los cien datos que se predicen una vez que la red se entrena con los mil datos anteriores. Para eliminar las altas frecuencias y facilitar el aprendizaje de la red, se filtró la serie de tiempo. Como ejemplo se muestra una sección de la series original y filtrada en la Fig. 4c.

El ejercicio consistió en probar las arquitecturas que de acuerdo a la literatura, son las más adecuadas para la predicción de series temporales, las RNRT y las RNR. Para cada uno de estos modelos se variaron algunos de los parámetros que determinan la eficiencia del proceso de aprendizaje. Entre otros, se varió el número de capas ocultas (1 ó 2), el número de neuronas en las diferentes capas de la red, el número de épocas de entrenamiento, etc. Una época es una unidad de entrenamiento en donde todos

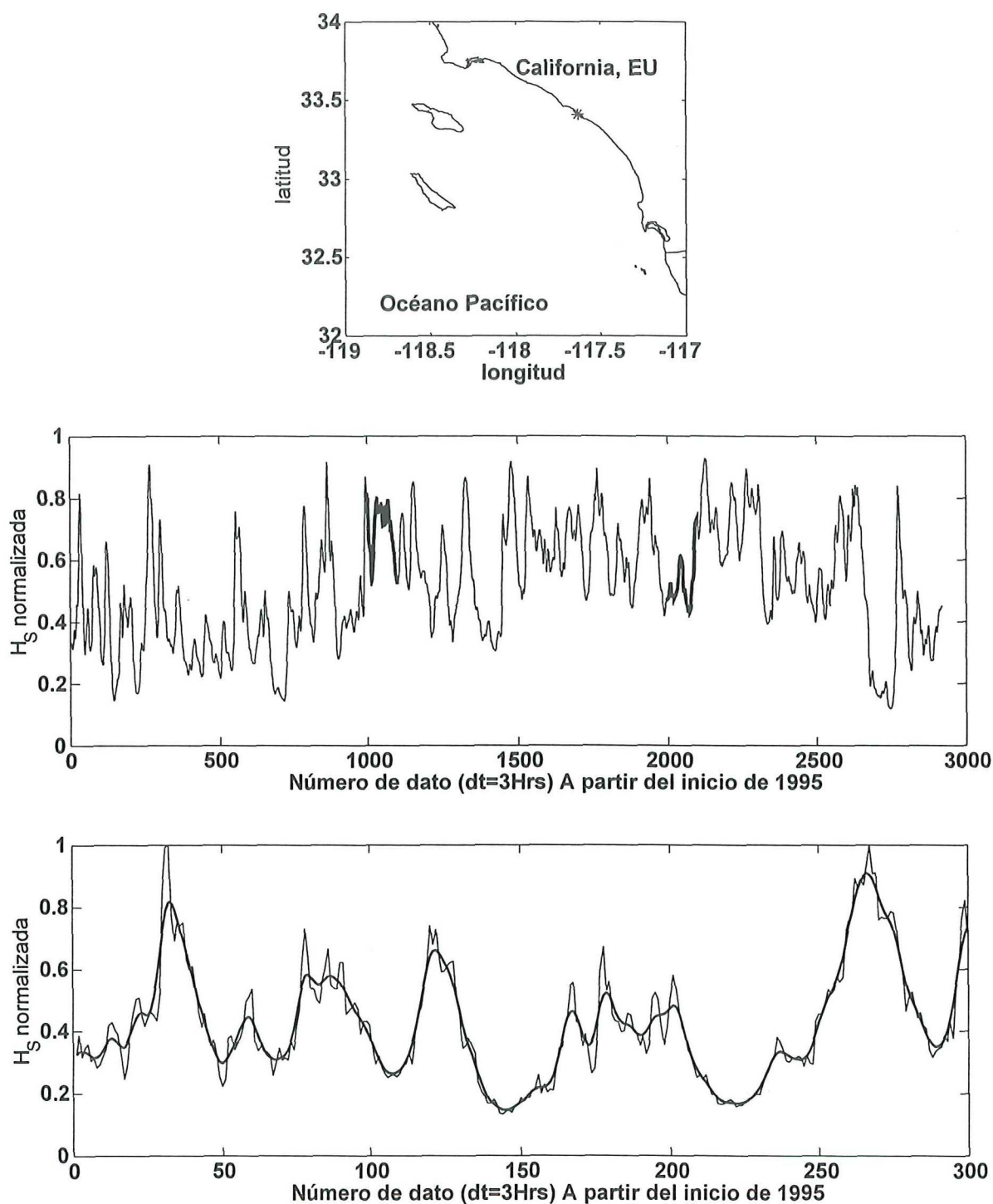


Figura 4: a) Localización de la estación San Clemente en donde se obtuvieron los datos que se utilizan en este trabajo. b) Serie de altura significativa del oleaje medido durante 1995-1996. De la serie completa se utilizaron 1000 datos en el entrenamiento de la RNA para predecir 100 (línea realzada). c) Subconjunto de datos de la serie original y suavizada (línea realzada)

los patrones de un archivo de entrenamiento son leídos por la red. Las variaciones se hicieron alrededor de una arquitectura típica, una neurona en la entrada, dos capas ocultas cada una con ocho neuronas y una neurona en la salida. Estos ejercicios se hicieron con los dos tipos de arquitecturas, las RNR y las RNRT. El entrenamiento se realizó hasta que el error fué del orden de 10^{-3} , que indica que aparentemente la red “aprendió”. Sin embargo, el error cuadrático medio estimado a partir de la predicción resultó del orden de 10^0 .

Posteriormente se optó por probar con otro tipo de arquitectura, las RNRFI, que comúnmente proporcionan buenos resultados en la predicción de series de tiempo generadas por sistemas no-lineales (Wan, 1994). El primer experimento consistió en probar con una red de una sola capa oculta. Los resultados, aunque alentadores, no fueron tan buenos. Después se decidió probar con una red que consistió de una neurona en la entrada, dos capas ocultas y una neurona en la salida. Se variaron tanto el número de neuronas en las capas ocultas, como el orden de los filtros en la neurona de entrada y de las capas ocultas.

En la Tabla I se presentan los parámetros utilizados en algunos experimentos realizados. La nomenclatura es la siguiente; unidades en la capa oculta (uo), orden del filtro de entrada (ofe), orden del filtro en la capa oculta (ofo), archivo de entrenamiento (aren), archivo de prueba (arpr), número de corrida (nc), épocas de entrenamiento (een), coeficiente de correlación (ccor), error cuadrático medio normalizado (ecmn). Se puede observar que la configuración que dio mejores resultados fue con una neurona en la entrada, dos capas ocultas cada una con 12 neuronas y una neurona en la salida ([1 12 12 1]). El orden del filtro en la entrada fué de 34 y de 8 en cada una de las capas ocultas ([34 8 8]). El entrenamiento consistió de 6000 épocas. Cabe aclarar que el ecmn se calculó solo para los primeros 20 datos predichos.

En la Fig. 5 se muestra una predicción no iterada de cien datos de la serie de H_S , es decir para predecir el siguiente dato, todos los anteriores usados son observaciones.

Tabla I: Parámetros de la arquitectura de la red y archivos de entrenamiento/prueba que se variaron en los experimentos E1-E15, para predecir H_S .

	uo1	uo2	ofe	ofo1	ofo2	aren	arpr	nc	een	ccor	ecmn
E1	12	12	34	8	8	scl1	sclpr1	4	500	0.873	0.221
E2	12	12	34	8	8	scl1	sclpr1	4	3000	0.910	0.145
E3	12	12	34	8	8	scl1	sclpr1	5	6000	0.948	0.094
E4	12	12	34	8	8	scl1	sclpr1	4	8000	0.927	0.148
E5	12	12	34	8	8	scl2	sclpr2	1	6000	0.695	0.051
E6	12	12	34	8	8	scl2	sclpr2	1	8000	0.730	0.130
E7	12	12	34	8	8	scl3	sclpr2	2	5000	0.038	0.948
E8	8	8	34	8	8	scl1	sclpr1	2	500	0.907	0.189
E9	8	8	34	8	8	scl1	sclpr1	3	5000	0.953	0.117
E10	8	1	34	8	8	scl1	sclpr1	1	500	0.901	0.182
E11	8	1	34	8	8	scl1	sclpr1	3	5000	0.947	0.120
E12	8	8	25	5	5	scl1	sclpr1	3	500	0.929	0.291
E13	8	8	25	5	5	scl1	sclpr1	3	5000	0.976	0.256
E14	8	8	7	5	5	scl1	sclpr1	1	500	0.930	0.172
E15	8	8	7	5	5	scl1	sclpr1	1	5000	0.887	0.273

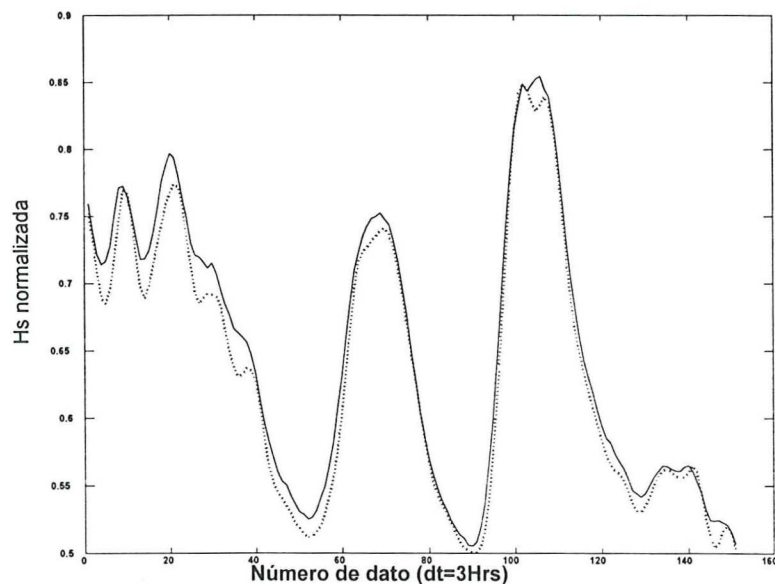


Figura 5: Altura significativa reportada en San Clemente CA. y su predicción (línea continua) en forma no iterada.

En la Fig. 6 se muestran cien puntos de la señal original y la señal predicha. Para obtener este resultado la red se entrenó con el patrón de entrenamiento H_S1 formado con los primeros mil datos de la serie completa (del dato 1 al 1000). Se puede observar, que los primeros treinta datos se predicen con un ecmmn aceptable, notando que a partir de este valor la diferencia entre las series crece. La Fig. 7 muestra la predicción de los cien valores posteriores a los segundos mil datos de la serie original, después de entrenar la red con el patrón H_S2 formado con los segundos mil datos (del dato 1001 al 2000). El resultado es similar al de la Fig. 6.

Estos resultados no se pueden comparar con ningún otro trabajo, ya que no existen predicciones de H_S a partir de una serie de tiempo. Por lo tanto la bondad de la predicción iterada dependerá de la aplicación. Por ejemplo, para el caso de aplicaciones relacionadas con maniobras marítimas y portuarias, arrastre de barcasas u otro tipo de plataformas y para la navegación segura es indispensable contar con información de las condiciones del oleaje. Con respecto a la cantidad de datos predichos en forma iterada, es difícil encontrar un porqué del hecho de que sean aproximadamente veinte y no más o menos. Se podría suponer que esto se debe al alto grado de no-linealidad del fenómeno del oleaje. Además la respuesta no lineal que presentan las RNA en sus capas ocultas aún son inexplicables (Hsieh & Tang 1998), por lo que no es fácil determinar que tan larga o corta podrá ser una predicción en el tiempo.

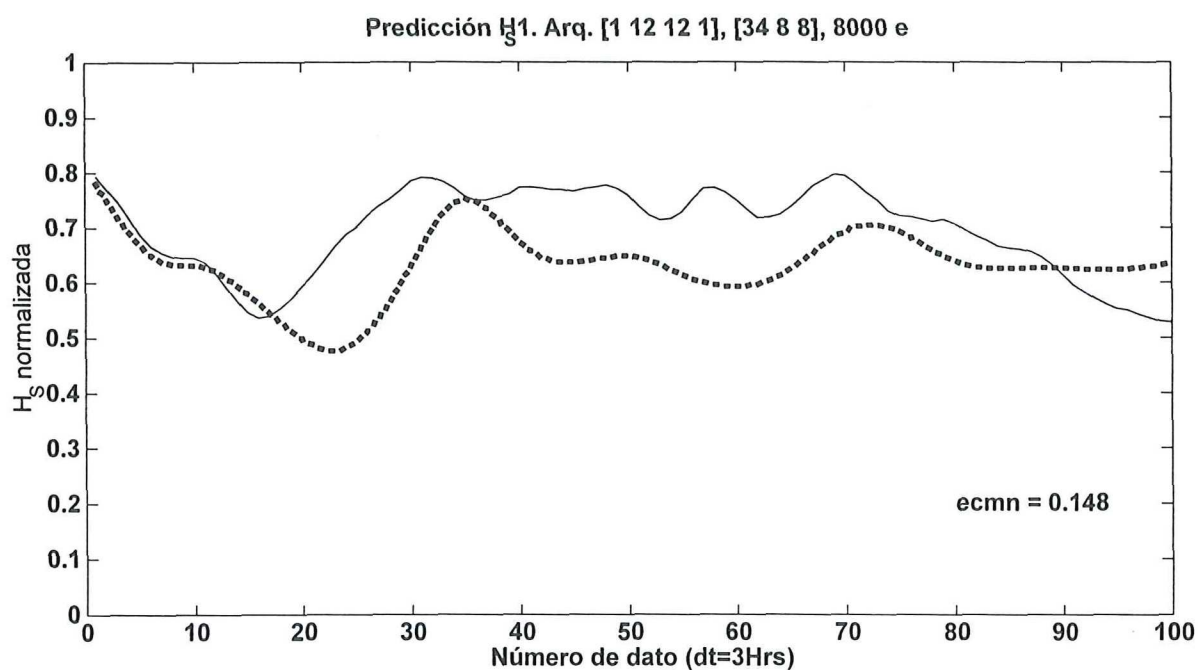
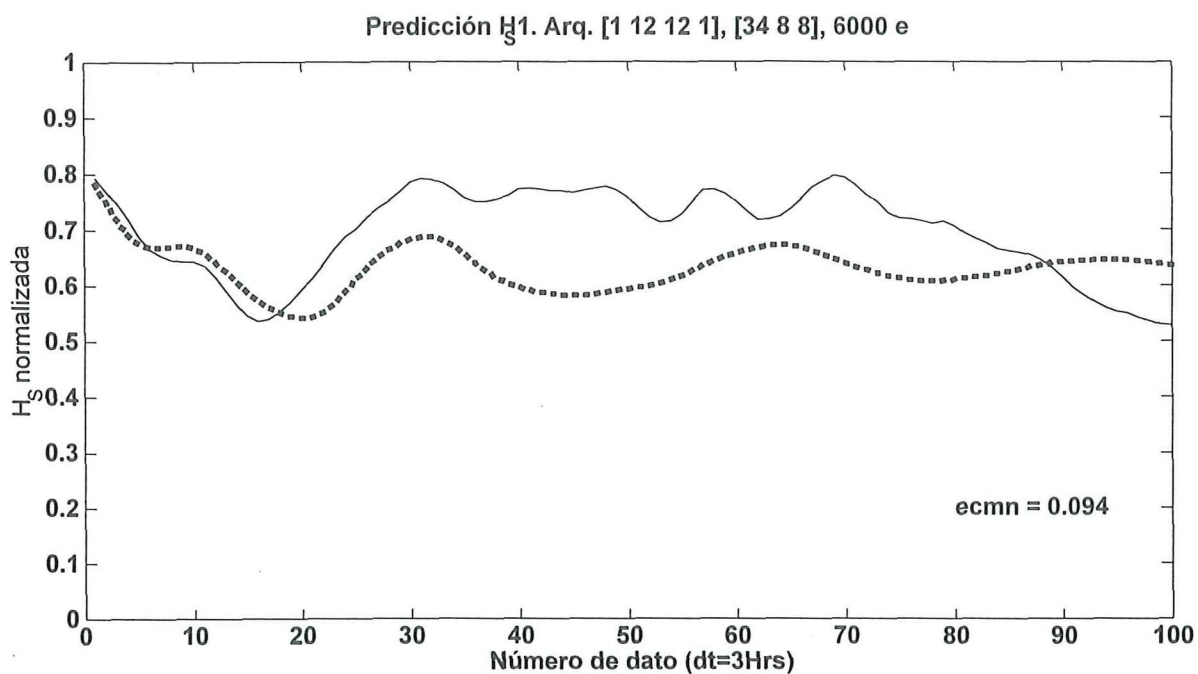


Figura 6: Altura significativa estimada a partir de mediciones en San Clemente CA. (línea continua) y predicciones iteradas obtenidas con una red con una neurona en la entrada y un filtro de orden 34, dos capas ocultas cada una con 12 neuronas y filtros de orden 8 y una neurona en la salida. La red se entrenó con los primeros mil datos de la serie y se predijeron los siguientes cien, empleando a) 6000 épocas de entrenamiento y b) 8000 épocas de entrenamiento.

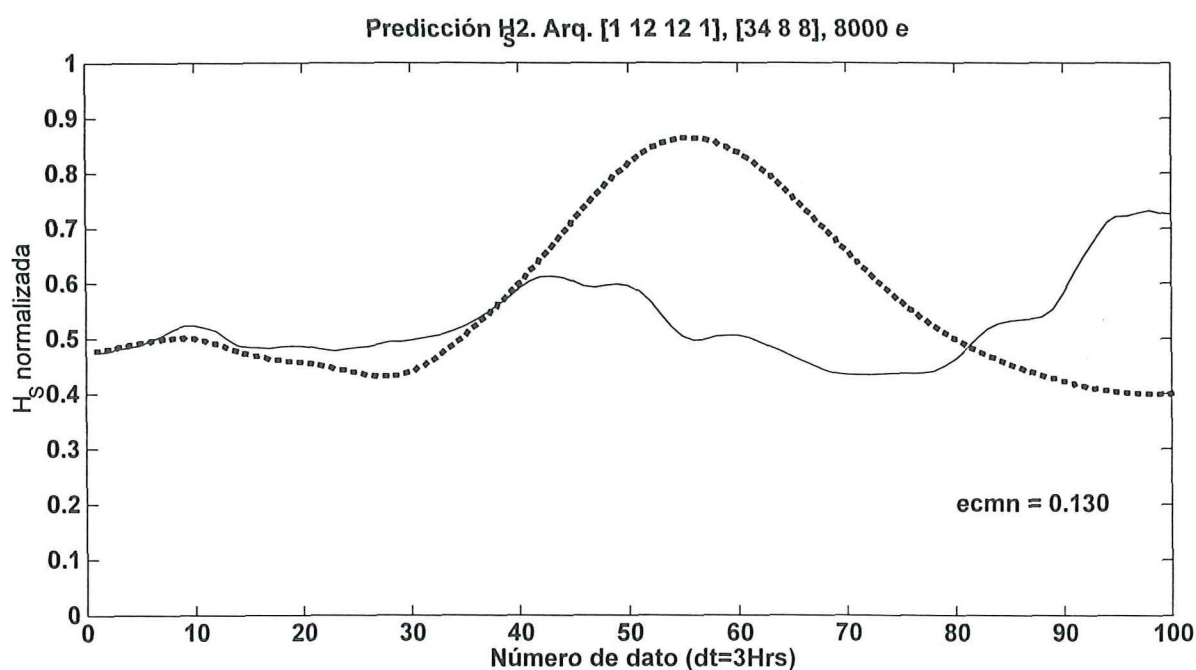
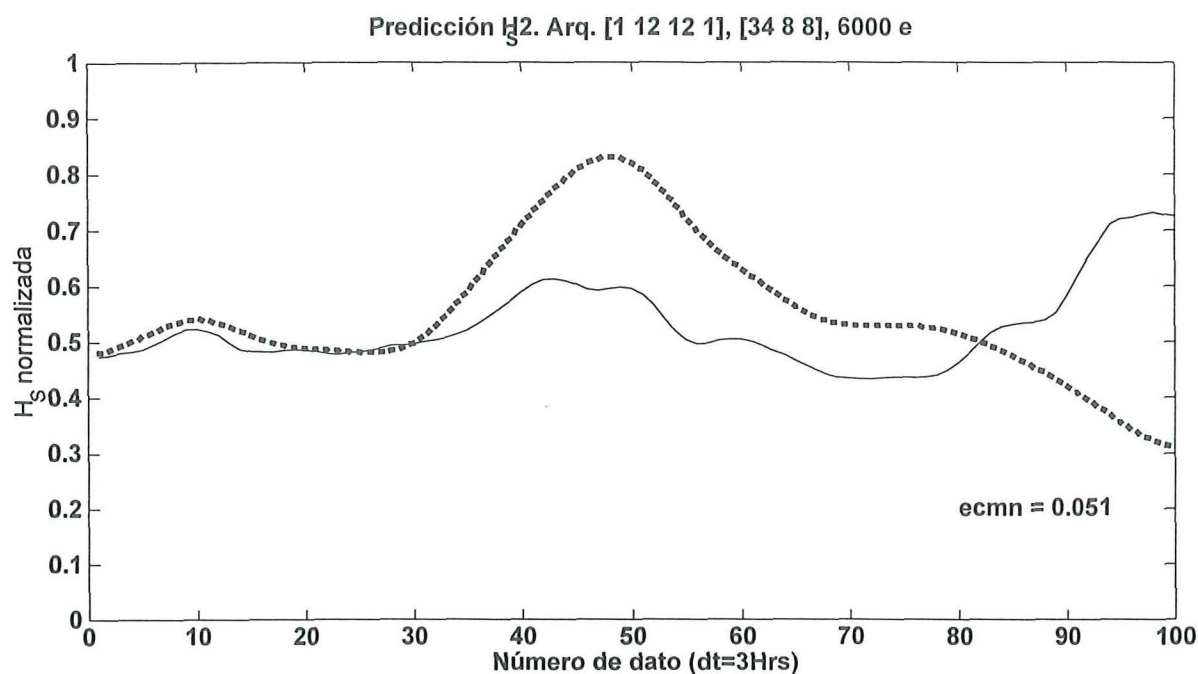


Figura 7: Altura significativa estimada a partir de mediciones en San Clemente CA.(línea continua) y predicciones iteradas obtenidas con una red con una neurona en la entrada y un filtro de orden 34, dos capas ocultas cada una con 12 neuronas y filtros de orden 8 y una neurona en la salida. La red se entrenó con los segundos mil datos de la serie y se predijeron los siguientes cien, empleando a) 6000 épocas de entrenamiento y b) 8000 épocas de entrenamiento.

IV.1.3 Predicción del IOS

Para la predicción del índice de oscilación sur, los datos fueron tomados también de Internet (Sci-95 n.d.). La serie de tiempo que se muestra en la Fig. 8, comprende desde el año de 1951 hasta 1998, muestreados con un intervalo de tiempo de un mes. Estos datos no fueron filtrados, pero sí se normalizaron escalando a promedio cero y varianza unitaria.

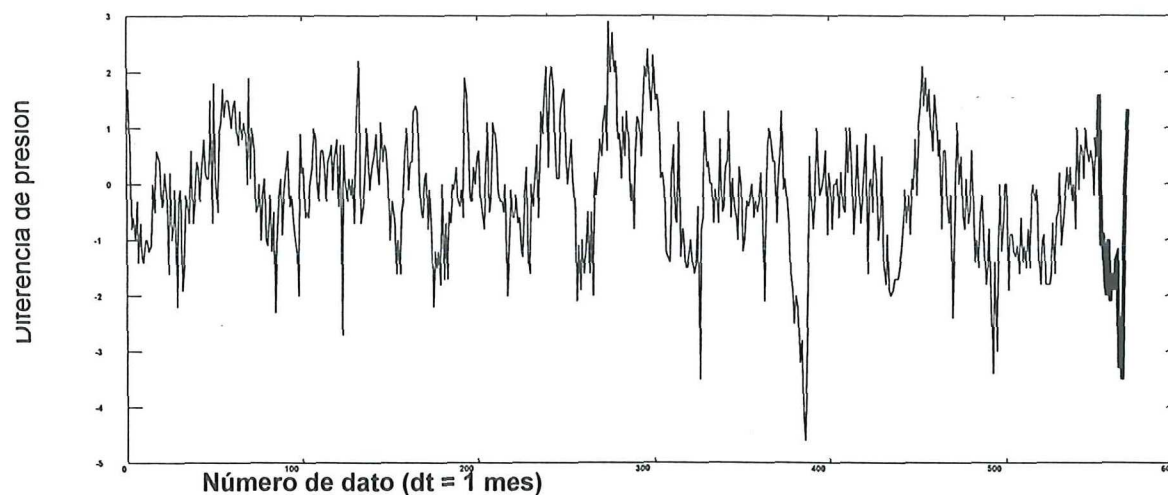


Figura 8: Serie de tiempo de la diferencia de presión entre Darwin y Tahití, de 1950 a 1998. La parte realizada corresponde al intervalo que se predice, entrenando la red con todos los datos anteriores

En la tabla II se muestran algunos de los experimentos que se efectuaron y en las gráficas se presentan las mejores predicciones.

Para este caso la arquitectura que dio mejores resultados fue [1 8 8 1], [10 5 5] y cuando solo se desea conocer la tendencia de la serie, se reduce el número de neuronas y se obtienen buenos resultados. En la Fig. 10 se presentan resultados al entrenar la red con datos de 1951 a 1996 inclusive y se predijeron 24 datos, es decir los que corresponden a 1997 y 1998. En la Fig. 11 se muestran resultados al entrenar la red con una serie de 1951 a 06/1997 inclusive y se predice la segunda mitad de 1997 y 1998 completo.

Tabla II: Parámetros de la arquitectura de la red y archivos de entrenamiento/prueba que se variaron en los experimentos E1-E11, para predecir el *IOS*.

	uo1	uo2	ofe	ofo1	ofo2	aren	arpr	nc	een	ccor	ecmn
E1	8	8	10	5	5	se96	sp96	2	500	0.770	0.478
E2	8	8	10	5	5	se96	sp96	1	1000	0.600	0.642
E3	8	1	10	5	5	se96	sp96	2	500	0.866	0.366
E4	8	1	10	5	5	se96	sp96	3	1000	0.625	0.741
E5	8	1	25	8	8	se96	sp96	4	500	0.778	0.389
E6	8	1	25	8	8	se96	sp96	4	1000	0.521	0.715
E7	8	8	10	5	5	se90	sp90	2	500	0.900	0.310
E8	8	1	10	5	5	se90	sp90	3	500	0.895	0.345
E9	12	12	10	5	5	se97	sp98	5	500	0.617	0.704
E10	8	8	10	5	5	se97	sp98	6	500	0.869	0.407
E11	8	1	10	5	5	se97	sp98	5	500	0.890	0.427

En la figura 9 se muestra una predicción no iterada del *IOS*. Se observa que la señal predicha casi se empalma sobre la real.

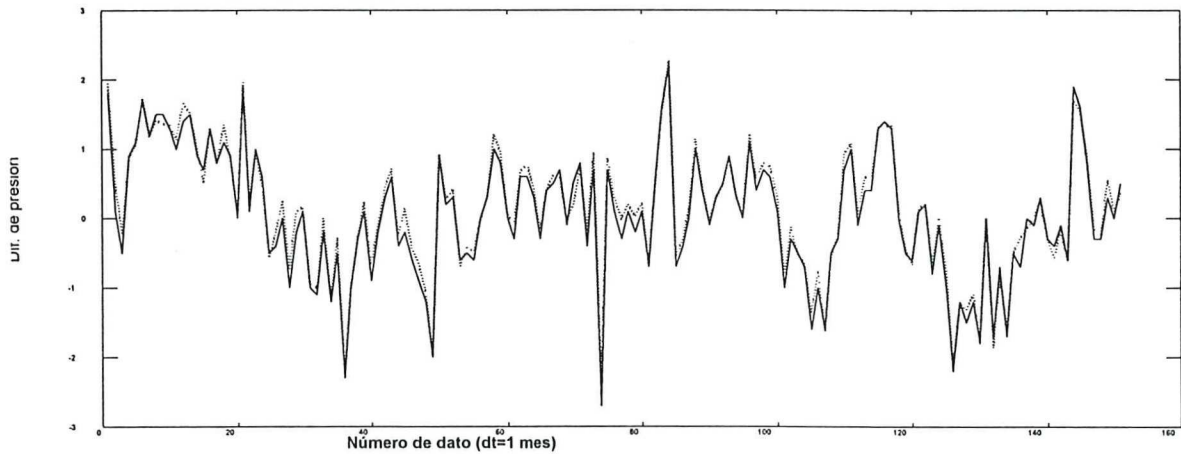


Figura 9: Predicción del *IOS* no iterada, con datos de 1951 a 1996 inclusive. La línea gruesa corresponde a la predicción y la fina a la señal real.

Para este parámetro, al igual que H_S no es posible comparar los resultados con otras predicciones, ya que en el fenómeno de El Niño, la variable que más comúnmente se correlaciona, es la anomalía de la temperatura superficial del mar *ATSM* y no el *IOS*. Con respecto a la cantidad de datos predichos se puede observar de la Tabla

II, que el menor error y la mayor correlación se presentan en E7, que corresponde a una predicción de año y medio, es decir dieciocho datos. Sin embargo esto no significa que los fenómenos sean iguales, pero al menos tienen en común el alto grado de no-linealidad, que es lo que se supone que determina que el pronóstico sea relativamente a corto plazo.

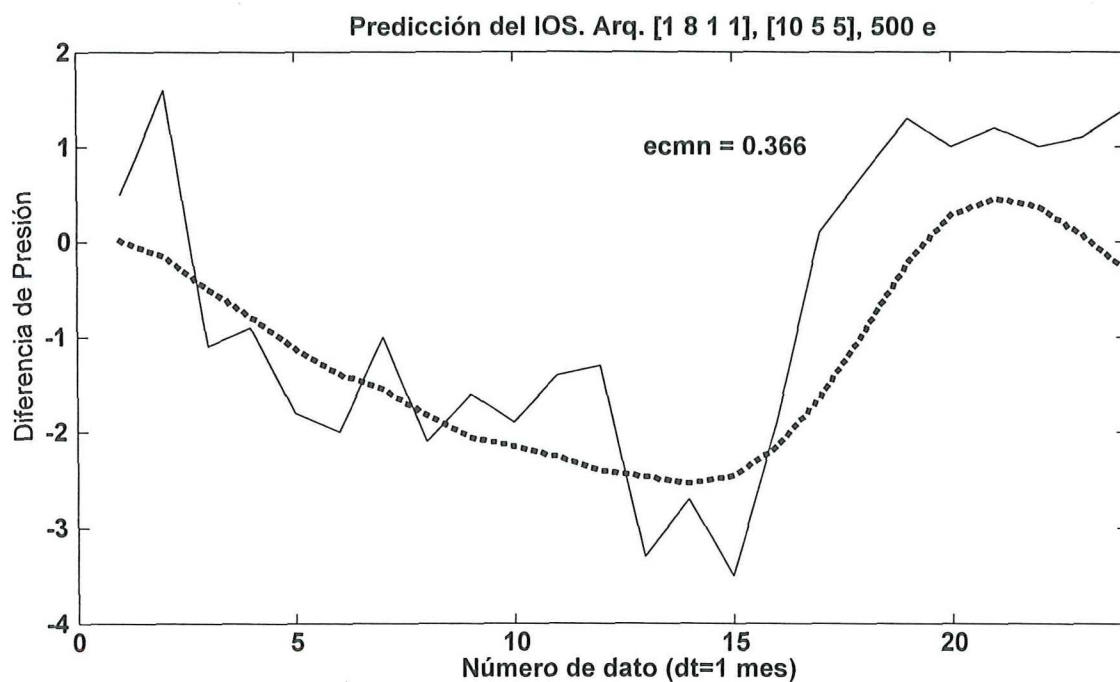
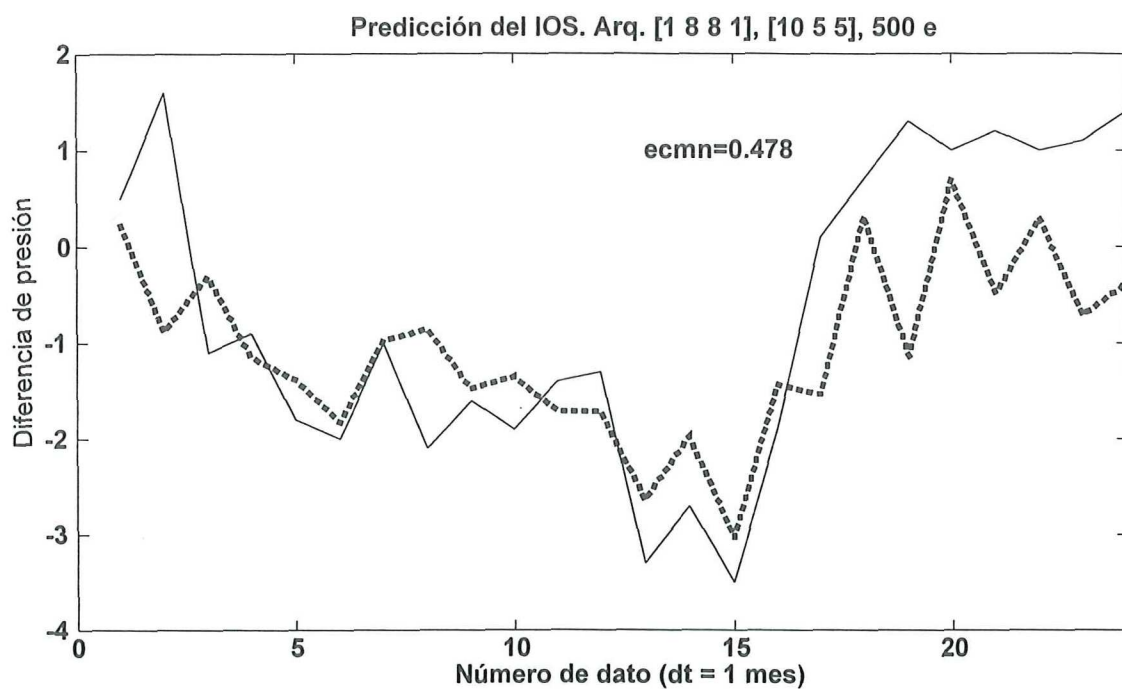


Figura 10: Predicción de una serie de tiempo del *IOS* con datos de 1951 a 1996 inclusive. Se predice el 97 y el 98. La predicción es la línea punteada. La diferencia entre las partes superior e inferior de la figura, es la cantidad de neuronas de la red.

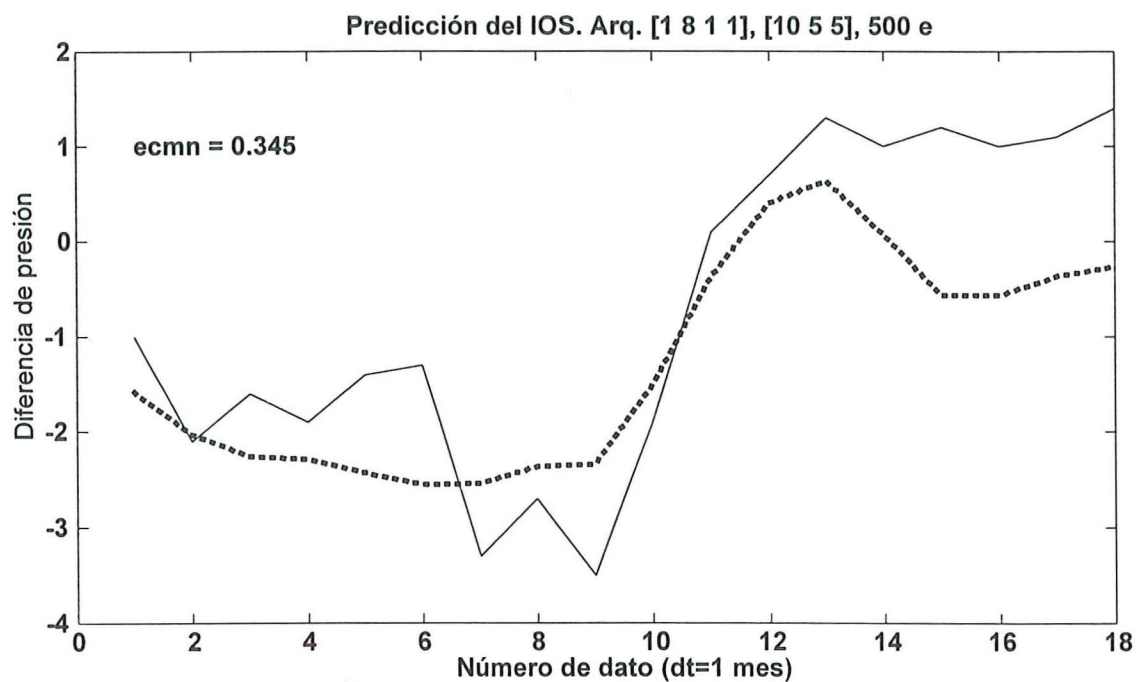
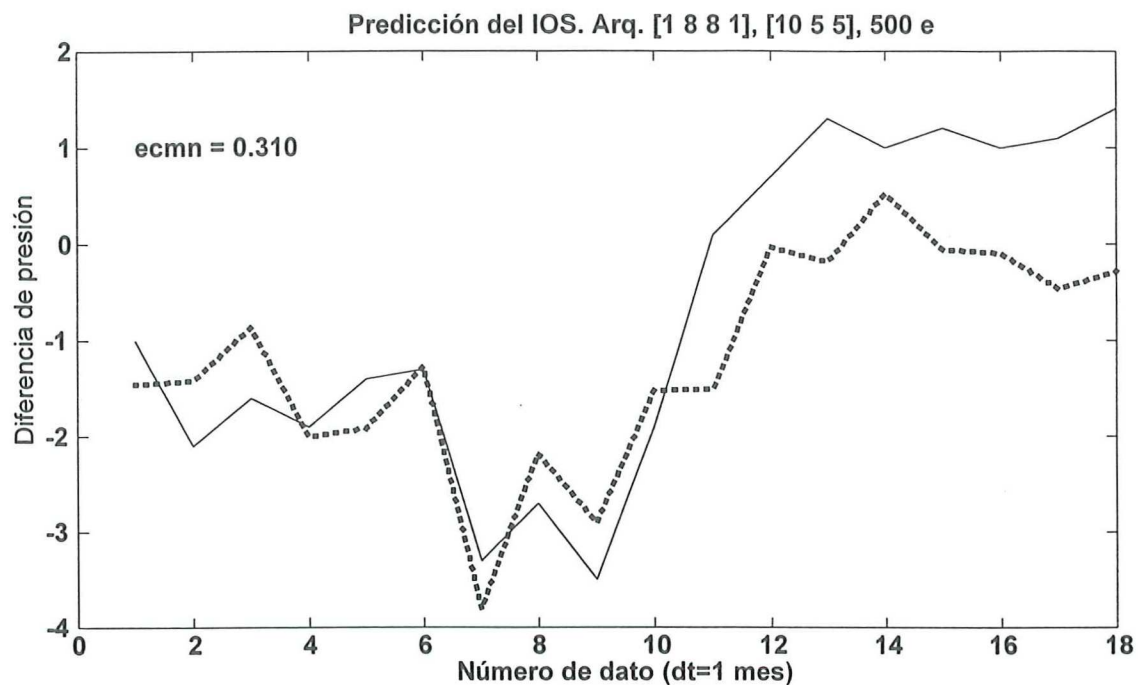


Figura 11: Predicción de una serie de tiempo del *IOS* con datos de 1951 a 06/1997 inclusive. Se predice la segunda mitad del 97 y el 98. La predicción es la línea punteada. La diferencia entre las partes superior e inferior de la figura, es la cantidad de neuronas de la red.

IV.1.4 Predicción de *ATSM*

Otro parámetro que comúnmente se usa en el estudio del fenómeno de El Niño es la anomalía de la temperatura superficial (*ATSM*). Para efectos de comparación entre pronósticos hechos con RNA y con otros modelos, se consideró una serie de tiempo de *ATSM*. Los datos fueron tomados de Internet (Sci-95 n.d.), la serie comprende los años de 1950 a 1998.

A continuación se comparan tres métodos de predicción del parámetro *ATSM*. El primero es un modelo acoplado océano-atmósfera con circulación global (GCM) que incorpora mediciones subsuperficiales en el océano en las condiciones iniciales (Huang et al, 1998). Este modelo se aplica en el COLA (Center for Ocean-Land-Atmosphere). El segundo es un modelo dinámico acoplado océano-atmósfera (CMF) de bajo orden (Kleeman, 1997), que se usa en el BMRC (Bureau of Meteorology Research Center). El tercero, en el que se aplican las RNA, utiliza una arquitectura muy semejante a la que se usó con el *IOS*, una neurona en la entrada, con un filtro de orden 10, ocho neuronas en la pimer capa oculta con un filtro de orden 5, una neurona en la segunda capa oculta con un filtro de orden 5 y una neurona en la capa de salida. El entrenamiento consistió de 500 épocas con el conjunto de datos del periodo de 1950-1997 y se predijo la *ATSM* correspondiente a 1998. Se calcularon los coeficientes de correlación (cc) para cada caso. Para el modelo GCM el cc estimado fue de 0.832, para CMF de 0.848 y para las RNA de 0.977. La Fig. 12 muestra la curva real y las tres predicciones, en ella se puede observar que las RNA generan la mejor predicción con el ccor más alto.

Aunque la predicción sigue siendo buena solo a corto plazo, en este caso (El Niño) el poder pronosticar un año y medio ya es una gran ventaja, dada la escala temporal del fenómeno.

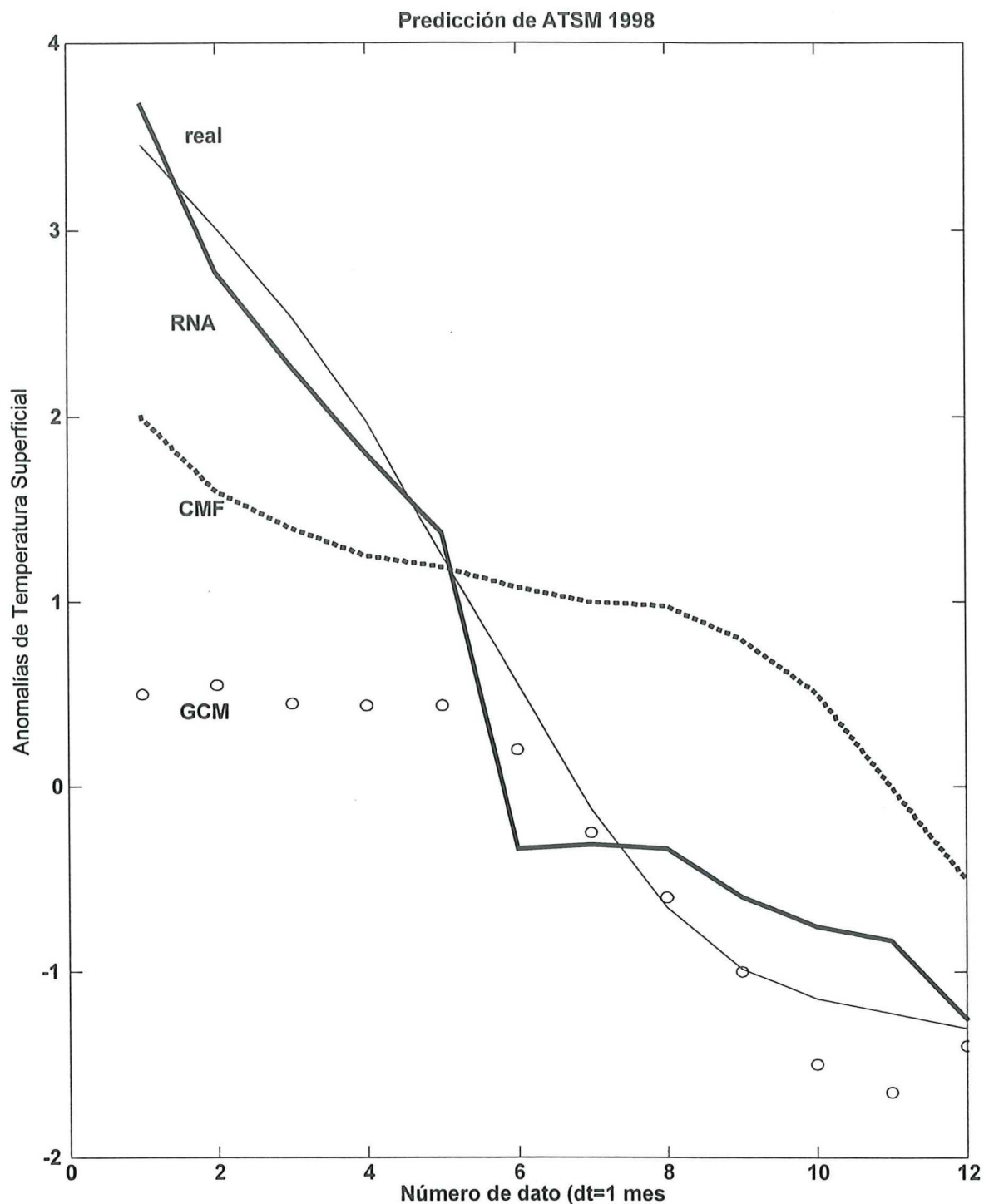


Figura 12: Comparación de la predicción de la *ATSM* para 1998 a partir de una serie de tiempo por tres métodos diferentes. Para las RNA el coef. de correlación resulta de 0.979, para el GCM de 0.832 y para el CMF de 0.841.

V CONCLUSIONES

Se aplicaron algoritmos basados en RNA para la predicción de series de tiempo de las variables oceanográficas H_S , IOS y $ATSM$ a partir solo de las series de tiempo, es decir no se usa ningún modelo complicado que tenga que ser inicializado con condiciones de frontera o alguna otra variable relacionada con el fenómeno físico al que pertenece la serie de tiempo que se predice.

Se probaron tres tipos diferentes de redes, las RNRT, las RNR y las RNRFI, encontrando que las últimas fueron las que produjeron las mejores predicciones. Para determinar que arquitectura de la RNRFI tenía el mejor comportamiento, se hicieron experimentos variando los diferentes parámetros de la red, y se encontró que para series de H_S la red debe conformarse de una capa de entrada con una neurona y un filtro de orden 34, dos capas ocultas cada una con 12 neuronas con filtros de orden 8 y una capa de salida con una neurona. Para el caso de series temporales de IOS y $SSTA$ la red se construyó con una capa de entrada con una neurona y un filtro de orden 10, dos capas ocultas con 8 neuronas en cada una de ellas y filtros de orden 5 y una capa de salida con una neurona. Todas las redes fueron entrenadas con el algoritmo de aprendizaje de retropropagación, usando la sigmoide como función de activación y una razón de aprendizaje de 0.001

De acuerdo a los experimentos realizados se puede observar que en la mayoría de los casos se obtuvo un error cuadrático medio normalizado, de alrededor de 0.4 ± 0.1 y un coeficiente de correlación entre la serie original y la predicha mayor que 0.8. Estos resultados son válidos en la predicción hasta de veinte datos, cuya utilidad dependerá de la escala temporal del fenómeno. Por ejemplo para el caso de los fenómenos que se estudiaron en esta tesis, predecir veinte datos de H_S solo tendrá sentido para algunas aplicaciones, pues equivale a pronosticar uno o dos días. El periodo total de predicción dependerá del intervalo de muestreo de los datos. Para el fenómeno de El Niño, las

series de tiempo analizadas tienen un intervalo de muestreo mensual, esto significa que veinte datos equivalen a una predicción de año y medio.

Con respecto a la cantidad de datos predecibles y de acuerdo a la teoría de Caos, se puede tener una idea de la dinámica que describe el sistema que generó la serie de tiempo. Es decir, es posible conocer en forma aproximada que tan determinista o estocástico es el sistema físico que genera la serie. Esto se hace con la ayuda de la técnica de reconstrucción de espacios de estado (Weigened, 1991). En este caso particular se podría decir que los sistemas físicos que generan ambas series son no-lineales y caóticos, es por eso que solo se pueden hacer predicciones a corto plazo (aproximadamente 20 datos) con técnicas deterministas como son las RNA.

Entre las ventajas principales que presenta la técnica de predicción aplicando RNA, está el hecho de que lo único que se necesita para alimentar el modelo, es la serie de tiempo. Además, el tiempo de máquina para realizar un pronóstico, es del orden de minutos, una vez que la red ha sido entrenada. Otra ventaja es que el usuario no requiere tener conocimientos de redes, ya que el modelo funciona como una caja negra.

Las desventajas más fuertes que tiene este método son; La selección de parámetros necesarios para conformar la arquitectura de una red, se realiza mediante prueba y error. Para el caso de series generadas por sistemas no-lineales, la predicción que se puede lograr es solo a corto plazo. Las RNA no aportan un avance en el entendimiento de la física del fenómeno que origina la serie temporal.

Sin embargo, a pesar de todas estas desventajas, las RNA funcionan, y funcionan bien, esto se puede observar de la comparación que se hace con otros métodos para la predicción de la *ATSM*.

Referencias

- Bacon, S. & Carter, D. (1991). Wave climate changes in the north atlantic and north sea, *International Journal of Climatology* **11**: 545–558.
- Cover, T. M. (1965). Geometrical and statistical properties of systems of linear inequalities with application in pattern recognition, *IEEE Trans. Electronics* **14**: 326–334.
- Elman, J. (1991). Finding structure in time, *Cognitive Science* **14**: 179–211.
- Hertz, J., Krogh, A. & Palmer, G. (1991). *Introduction to the Theory of Neural Computation*, Addison Wesley.
- Hsieh, W. & Tang, B. (1998). Applying neural network models to prediction and data analysis in meteorology and oceanography, *Bulletin of the American Meteorological Society* **79**: 1855–1870.
- Kevin, S. (1996). *Applying Neural Networks*, Academic Press.
- Parker, T. S. & Chua, L. O. (1989). *Practical Numerical Algorithms for Chaotic Systems*, Springer-Verlag.
- Philander, S. G. (1990). *El Niño, La Niña, and the Southern Oscillation*, Academic Press.
- Scl-95 (n.d.). [Http://nic.fb4.noaa.gov/data/cddb](http://nic.fb4.noaa.gov/data/cddb).
- Seymour, R., Strange, R. R., Cayan, D. R. & Nathan, R. A. (1984). *Coastal Engineering*.
- Seymour, R., Tegner, M., Dayton, P. & Pamell, P. (1989). Storm wave induced mortality of giant kelp, *macrocystis pyrifera*, in Southern California, *Estuarine, Coastal and Shelf Science* **28**: 277–292.

- Takens, F. (1981). *Detecting Strange Attractors in Fluid Turbulence*, Springer-Verlag.
- Wan, E. A. (1993). *Finite Impulse Response Neural Networks with Applications in Time Series Prediction*, PhD thesis.
- Weigend, A. S. & Gershenfeld, N. A. (1994). *Time Series Prediction, Forecasting the Future and Understanding the Past*, Addison Wesley.
- Yule, G. (1927). On a method of investigating periodicity in distributed series with special reference to wolfer's sunspot numbers, *Phil. Trans. Roy. Soc. London A* **226**: 267–298.