



# UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

ANALISIS DE SERIES TEMPORALES CAOTICAS PARA LA  
DETECCION DE DINAMICA NO - LINEAL: SISTEMAS  
TIEMPO /CLIMA

**T E S I S**

QUE PARA OBTENER EL TITULO DE:

F I S I C O

P R E S E N T A :

ADRIAN TOMAS CASILLAS LAVIN



DIRECTOR DE TESIS: DR. TOMAS MORALES ACOLTZI



FACULTAD DE CIENCIAS  
SECCION ESCOLAR  
2004



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

ESTA TESIS NO SALE  
DE LA BIBLIOTECA

UNIVERSIDAD NACIONAL AUTÓNOMA DE  
MÉXICO

FACULTAD DE CIENCIAS

*Análisis de Series Temporales Caóticas para la  
Detección de Dinámica No-lineal: Sistemas  
Tiempo/Clima*

Adrián T. Casillas Lavín

2004

Autorizo a la Dirección General de Bibliotecas de la  
UNAM a difundir en formato electrónico e impreso el  
contenido de mi trabajo recepcional.

NOMBRE: Adrián Tomás  
CASILLAS LAVÍN

FECHA: 20-09-04

FIRMA: 



**ACT. MAURICIO AGUILAR GONZÁLEZ**  
**Jefe de la División de Estudios Profesionales de la**  
**Facultad de Ciencias**  
**Presente**

Comunicamos a usted que hemos revisado el trabajo escrito:  
"ANÁLISIS DE SERIES TEMPORALES CAÓTICAS PARA LA DETECCIÓN DE DINÁMICA NO-LINEAL:  
SISTEMAS TIEMPO/CLIMA"

realizado por Adrián Tomás Casillas Lavín

con número de cuenta 7400664-9, quien cubrió los créditos de la carrera de: Física.

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis  
Propietario

Propietario Dr. Tomás Morales Acoltzi

Dr. Ramón López Peña

Propietario

Dr. Carlos Alberto Reyes García

Suplente

Dr. Fernando García García

Suplente

Dr. Javier Vitela Escamilla

Consejo Departamental de Física

DRA. PATRICIA GOLSTEIN-MENACHE  
Coordinadora de Licenciatura

FACULTAD DE CIENCIAS  
DEPARTAMENTO DE FÍSICA

In memoriam

María Elena Lavín Espinosa

Dedico este trabajo a Bertha, Tomás y Priscila por su gran cariño.

Mi gratitud al Dr. Tomás Morales Acoltzi por su franca amistad, dedicación y excepcional apoyo.

Mi reconocimiento y aprecio a los doctores: Fernando García García, Ramón López Peña, Carlos Alberto Reyes García y Javier Vitela Escamilla por sus valiosas contribuciones.

Agradezco al Centro de Ciencias de la Atmósfera de la UNAM, él haberme permitido hacer uso de sus instalaciones y equipo de computo.

Agradezco también al Programa de Apoyos a Proyectos de Investigación e Innovación Tecnológica (PAPIIT), proyecto IN111402.

# Índice General

<b>1</b>	<b>INTRODUCCIÓN</b>	<b>5</b>
<b>2</b>	<b>CONCEPTOS BÁSICOS</b>	<b>11</b>
2.1	Series de tiempo . . . . .	11
2.2	Análisis de Fourier . . . . .	12
2.3	Ondeletas . . . . .	14
2.3.1	Análisis de Multiresolución . . . . .	15
2.4	Sistemas Dinámicos . . . . .	19
2.5	Atractores . . . . .	19
2.6	Dimensión Fractal . . . . .	23
2.7	Dinámica Cuantitativa . . . . .	28
2.8	Búsqueda de Atractores en TIEMPO/CLIMA . . . . .	30

<i>ÍNDICE GENERAL</i>	2
2.9 Envoltentes . . . . .	33
2.10 Método de retrasos . . . . .	35
2.11 Coeficientes de Lyapunov . . . . .	38
2.12 Dimensión de Atractores para una dimensión envolvente fija $m$ . . . . .	40
<b>3 REDES NEURONALES</b>	<b>42</b>
3.1 Introducción . . . . .	42
3.2 Memoria Asociativa . . . . .	44
3.3 Límites de Confiabilidad . . . . .	48
3.4 Perceptrones y algoritmo de aprendizaje . . . . .	50
3.5 Prueba de convergencia para el algoritmo de aprendizaje . . . . .	55
3.6 Aprendizaje de Gradiente Descendiente . . . . .	57
3.7 Retropropagación . . . . .	58
<b>4 REDES DE RESPUESTA DE IMPULSO FINITO (FIR)</b>	<b>66</b>
4.1 Representación Espacio Temporal . . . . .	66
4.2 Equivalencia entre las arquitecturas de las redes . . . . .	67
4.3 Retropropagación en el tiempo. Funcionalidad de las redes FIR . . . . .	69

**5 EVALUANDO PROCEDIMIENTOS 73**

5.1 Caos y Ruido . . . . . 73

    5.1.1 Espectro de Potencias . . . . . 73

    5.1.2 Generando y analizando imágenes de Ondeletas . . . . . 78

5.2 Método de Información Mútua . . . . . 82

    5.2.1 Ecuaciones de Rössler - Método de Runge Kutta . . . . . 90

    5.2.2 Estimación de la Dimensión del Atractor . . . . . 93

    5.2.3 Límite de Predictibilidad . . . . . 96

**6 EXPLOTANDO LA RED NEURONAL 97**

6.1 Normalizando la Serie . . . . . 97

6.2 Seleccionado la arquitectura . . . . . 100

6.3 Condiciones iniciales - ensamble . . . . . 102

6.4 Desarrollo en Armónicos . . . . . 103

    6.4.1 Pronóstico Extendido . . . . . 104

    6.4.2 Pronóstico Secuencial . . . . . 120

    6.4.3 Pronóstico Tradicional . . . . . 123

6.4.4 Dinámica de la Habilidad de Pronóstico . . . . . 125

**7 DISCUSIÓN DE RESULTADOS . . . . . 127**

7.1 Habilidad de Pronóstico . . . . . 127

7.2 Extensión de la Calidad de Pronóstico en el tiempo . . . . . 130

7.3 Cantidad anual de lluvia . . . . . 131

7.4 Pronóstico año 2004 . . . . . 131

7.5 Modelo Determinístico Europeo . . . . . 134

7.6 Aportaciones . . . . . 135

7.7 Trabajo Futuro . . . . . 136

# Capítulo 1

## INTRODUCCIÓN

El identificar la dinámica determinística en una serie de tiempo, relacionada con el clima; como la precipitación, conduce a una forma más profunda de comprender los procesos físicos involucrados, nos permite encontrar un umbral de validez a nuestras predicciones y establecer mecanismos para generar modelos del comportamiento del fenómeno en estudio.

Para mucha gente, el tiempo/clima parece ser un proceso aleatorio y caótico. A pesar de que existen ciclos diurnos y estacionales; el tiempo/clima continúa, día a día, mostrándose completamente impredecible. No obstante, una investigación cuidadosa de cartas de tiempo/clima y de fotografías de satélite muestra que hay en efecto sistemas bien definidos y coherentes que son responsables en buena parte del tiempo/clima. La existencia de un sistema coherente embebido en un fluido caótico, es indicativa de la presencia de una inestabilidad dinámica en el fluido.

Esta inestabilidad es responsable de una bifurcación en la cual el fluido cambia de un régimen dinámico a otro; por lo que se hace necesario identificar y entender los procesos dinámicos responsables para estas bifurcaciones.

La atmósfera terrestre es un sistema físico de complejidad extraordinaria que muestra una rica diversidad de comportamiento fenomenológico; para ser comprendida incide en el dominio de los físicos, pero también se traslapa fuertemente con el campo de las Matemáticas aplicadas y los campos interdisciplinarios de Oceanografía y Ciencias de la Atmósfera.

El realizar predicciones es uno de los principios básicos de la ciencia. Dado que conocemos el desempeño de un sistema, ¿cómo podemos predecir su comportamiento futuro?. El primer paso consiste en elaborar un modelo, basado en consideraciones teóricas y, posteriormente, utilizar mediciones experimentales como datos iniciales.

Tomando como base un fuerte conocimiento teórico de los principios que intervienen en la meteorología, los modelos numéricos de predicción del tiempo/clima resuelven un conjunto de ecuaciones diferenciales que describen el comportamiento de un fluido embebido en una esfera que gira, no obstante, dado que las mediciones se realizan sobre una malla de estaciones con un arreglo no-homogéneo, las condiciones iniciales para modelar consisten de conjuntos de medidas discretas.

Implícitamente nuestro conjunto de ecuaciones diferenciales continuas opera sobre un conjunto discreto de mediciones; debido a esta limitante, surge la necesidad de generar modelos directamente de los datos

con que contamos (series de datos temporales).

La motivación principal para la elaboración de este trabajo, consiste en comprender y analizar algunos de los métodos que se están explotando en la teoría de los Sistemas Dinámicos, para proporcionar información sobre la física del tiempo/clima. Estos métodos permiten extraer información cualitativa sobre los Sistemas Dinámicos a partir de series de datos experimentales.

Cada día es más claro que, en la atmósfera, por encima de los fenómenos que muestran persistencia y comportamiento cíclico o estacionario sobre una región determinada, existe una fuerte relación incluso entre sistemas climatológicos que se consideraban ajenos, a través de teleconexiones; por ejemplo, el sistema conocido como 'EL NIÑO - LA NIÑA ', que corresponde a un calentamiento o enfriamiento del océano Pacífico Ecuatorial, respectivamente, y su impacto en la circulación general de océanos atmósfera en regiones bien localizadas [45].

La atmósfera es un sistema dinámico donde se desarrollan diferentes tipos de desplazamiento de fluidos, en una variedad que va desde los elementales transportes laminares hasta los más complejos desplazamientos con una gran turbulencia.

Los fenómenos de la naturaleza que son más interesantes requieren modelos que conlleven ecuaciones diferenciales de orden superior, que generalmente no son solubles en forma explícita, que involucran procesos iterativos, que son recursivas (autorepetibles), en los que existe intermitencia (se comportan en forma suave sólo en intervalos discontinuos) y en los que una pequeña variación en las condiciones

iniciales puede desencadenar diferencias sustantivas en la evolución del modelo.

Considerando adicionalmente los errores comprendidos en la medición de las variables observadas, la gran diferencia en el comportamiento atmosférico que se encuentra para pequeñas variaciones en las condiciones iniciales, el número de variables involucradas y el aparentemente irregular desempeño de cada una de ellas; pareciera muy lejana la posibilidad de poder describir el comportamiento del tiempo/clima y realizar predicciones, incluso a nivel local. Además, existen errores de redondeo asociados al sistema de cómputo aplicado.

Los modelos elaborados ignorando variables cuya influencia en el comportamiento global del sistema pudiera considerarse de segundo orden, alcanzan una simplicidad tal que no permiten avanzar en la comprensión del problema.

¿Cómo podemos generar modelos que nos permitan la descripción del comportamiento de la naturaleza caracterizando el mosaico de su enorme complejidad?

¿Cuántas veces nos hemos encontrado frente a una serie de datos sin ningún orden aparente, con la intuición, o al menos el deseo, de que detrás de ella exista alguna forma de describir ese comportamiento complejo?

Desarrollos recientes en la Teoría de Sistemas Dinámicos aplicados a problemas meteorológicos; consis-

ten en estimar el número de variables independientes (grados de libertad), necesarios para elaborar un modelo que permita describir el comportamiento del tiempo/clima y su evolución en el tiempo (predicción).

'Los científicos que desean estudiar los sistemas dinámicos han recurrido a otro enfoque de medición, la matemática cualitativa. En la vieja matemática cuantitativa la medición de un sistema se concentra en indagar cómo la forma del movimiento del sistema afecta la cantidad de movimiento de otras partes. En cambio, en la medición cualitativa se trata de mostrar la forma del movimiento del sistema como totalidad' [28].

El análisis considerado para identificar la naturaleza de la dinámica de las series temporales de este estudio, corresponde al procedimiento descrito por Grassberger and Proccacia [2], el método consiste en reconstruir el espacio de fase del sistema en estudio a partir directamente de la serie de tiempo de alguna de las variables que interactúan dentro de éste.

Esto implica esencialmente desde el punto de vista físico, que toda la información respecto al comportamiento del sistema y la evolución de las variables que forman parte de éste, se encuentra presente de algún modo en la evolución de cualquiera de ellas.

En el año de 1997, se firma un convenio de colaboración en el que intervienen productores agrícolas del estado de Tlaxcala, la Universidad Autónoma del Estado de Tlaxcala y la Universidad Nacional Autónoma de México, con la finalidad de generar un pronóstico de precipitación; que evaluado por

los productores, sirva como una herramienta de apoyo a su labor, permitiendo, además intercambiar conocimientos [44].

En este trabajo en particular se utilizaron dos series de datos de precipitación mensual: la primera se verifica en la población de Apizaco, Tlaxcala desde Enero de 1942 a Diciembre de 2002 y la segunda serie corresponde a la población de Balboa, Panamá desde Enero de 1906 a Diciembre de 2000. En ambos casos, a partir de una serie de datos temporal (discreta); para poder plantear un sistema de ecuaciones diferenciales, se considera que: el tomar corrimientos (desplazamientos en el tiempo de longitud finita  $\tau$ ), es análogo al proceso de diferenciación sobre una variable continua en el espacio de fases.

A fin de determinar el tiempo de retraso  $\tau$  que permita la reconstrucción más completa del sistema en estudio se aplica el procedimiento de análisis de información mutua propuesto por Fraser y Swinney [20].

Una vez que se identificaron los elementos que caracterizan al atractor que representa dentro del espacio de fases al sistema en estudio; se procedió a la construcción de una red neuronal, que nos permita reproducir la dinámica del mencionado atractor, a fin de poder efectuar predicciones dentro de los límites establecidos por los métodos anteriores. Finalmente, se analizaron los espectros de potencia de las series de datos mediante el método de Wavelets (Ondeletas) y de autocorrelación, a fin de contrastar los resultados, en lo que se refiere a definir propiedades de la no-estacionalidad de la Serie de Tiempo.

# Capítulo 2

## CONCEPTOS BÁSICOS

### 2.1 Series de tiempo

En muchas áreas, datos experimentales están constituidos por una secuencia de medidas realizadas en intervalos de tiempo iguales, esto es lo que se denomina serie de tiempo.

Una serie de tiempo estocástica es aquella que muestra una evolución en el tiempo de acuerdo a las leyes de probabilidad; una serie de tiempo estocástica se describe por sus momentos estadísticos: (media, varianza y función de autocorrelación).

Dentro de las series de tiempo estocásticas encontramos las series de tiempo estacionarias, en ellas la media y la varianza son constantes finitas (Independientes del tiempo).

Consideremos una serie de tiempo  $X_t$  que contiene una componente periódica a una frecuencia conocida  $w_m$ , esto es, si nuestra serie de tiempo esta constituida por  $n$  mediciones de algún proceso separadas por un intervalo constante de tiempo  $\Delta t$

$$w_m = \frac{2\pi m}{n\Delta t} \quad (2.1)$$

Podemos modelar nuestra serie mediante:

$$X_t = R \cos(w_m t + \theta) + Z_t \quad (2.2)$$

donde  $R$  es la amplitud,  $\theta$  es la fase y  $Z_t$  una serie aleatoria estacionaria (ruido blanco), [38].

En la práctica las variaciones en una serie de tiempo se producen a diferentes frecuencias, generalizando:

$$X_t = \sum_{j=1}^k R_j \cos(w_j t + \theta_j) + Z_t \quad (2.3)$$

donde  $R_j$  es la amplitud a la frecuencia  $w_j$ .

Como  $\cos(wt + \theta) = \cos(wt)\cos(\theta) - \text{sen}(wt)\text{sen}(\theta)$ , la ecuación anterior se puede expresar como

$$X_t = \sum_{j=1}^k a_j \cos(w_j t) + b_j \text{sen}(w_j t) + Z_t \quad (2.4)$$

donde  $a_j = R_j \cos(\theta_j)$  y  $b_j = -R_j \text{sen}(\theta_j)$

## 2.2 Análisis de Fourier

Una de las herramientas fundamentales para analizar series de tiempo es el análisis espectral, éste nos permite identificar si la serie de tiempo en cuestión es periódica y los valores asociados de periodicidad,

en su caso.

El análisis espectral es una variante del análisis de Fourier, que consiste básicamente en aproximar una función mediante una suma de términos en senos y cosenos. Sea  $f(t)$  una función definida en  $(-\Pi, \Pi]$ , integrable dentro de este dominio, con un número finito de discontinuidades y un número finito de máximos y mínimos (condiciones de Dirichlet), entonces  $f(t)$  puede aproximarse por la Serie de Fourier:

$$f(t) \approx \frac{a_o}{2} + \sum_{r=1}^k (a_r \cos(rt) + b_r \sen(rt)) \quad (2.5)$$

$$a_o = \frac{1}{\Pi} \int_{-\Pi}^{\Pi} f(t) dt$$

$$a_r = \frac{1}{\Pi} \int_{-\Pi}^{\Pi} f(t) \cos(rt) dt$$

$$b_r = \frac{1}{\Pi} \int_{-\Pi}^{\Pi} f(t) \sen(rt) dt$$

$$(r = 1, 2, \dots)$$

Para poder aplicar la Serie de Fourier a series de tiempo discretas, basta con considerar la serie cuando  $f(t)$  es definida sobre los enteros  $1, 2, \dots, n$ . El efecto del análisis de Fourier sobre los datos, consiste en particionar la variabilidad de la serie en componentes a las frecuencias  $\frac{2\Pi}{n}, \frac{4\Pi}{n}, \dots, \Pi$ , la componente a la frecuencia  $w_r = \frac{2\Pi r}{n}$  es llamado el  $r$ -ésimo armónico, para  $r \neq \frac{n}{2}$ . La amplitud del  $r$ -ésimo armónico, se expresa como

$$R_r = (a_r^2 + b_r^2)^{\frac{1}{2}} \quad (2.6)$$

De esta manera,  $\frac{R_r^2}{2}$  es la contribución del  $r$ -ésimo armónico a la varianza, si definimos  $I(w_r) = \frac{nR_r^2}{4\Pi}$ ; la gráfica  $I(w)$  vs.  $w$  se conoce como espectro de potencias, a pesar de que  $I(w)$  es una función de la frecuencia. El área total bajo la curva es igual a la varianza de la serie de tiempo. La técnica soló

aplica para series de tiempo estacionarias.

### 2.3 Ondeletas

Con una metodología similar a la aplicación de la Serie de Fourier, surge la descomposición de una serie de tiempo en Ondeletas o Wavelets como una herramienta que nos permite no solo identificar las componentes principales de la serie de datos en términos de la frecuencia, sino también en el tiempo.

Una ondeleta es una función  $\Psi(\bullet)$  definida sobre los reales  $(-\infty, \infty)$  que cumple con:

La integral de  $\psi(\bullet)$  es 0.

$$\int_{-\infty}^{\infty} \psi(\mu) \partial\mu = 0 \quad (2.7)$$

La integral de  $\psi(\bullet)^2$  es 1.

$$\|\psi^2\| \equiv \int_{-\infty}^{\infty} \psi^2(\mu) \partial\mu = 1 \quad (2.8)$$

Luego entonces, las ondeletas están contenidas en  $L^2(\mathfrak{R})$ ; (el espacio de todas las funciones con norma finita), [39].

### 2.3.1 Análisis de Multiresolución

El Análisis de Multiresolución involucra las operaciones de traslación y variación de escala:

Dada una función  $\phi(\bullet)$  definida sobre  $\mathfrak{R}$ , la aplicación de traslación unitaria  $\tau$  a la función  $\phi(\bullet)$  produce una nueva función  $\phi_{0,1}(\bullet)$ .

$$\phi_{0,1}(t) \equiv \phi(t - 1)$$

$$-\infty < t < \infty$$

análogamente:

$$\phi_{0,n}(t) \equiv \phi(t - n)$$

$$n \in \mathbb{Z};$$

En particular,  $n = -1$  corresponde a la traslación inversa unitaria  $\tau^{-1}$ , esto es, la función  $\phi_{0,n}(\bullet)$  se transforma en  $\phi_{0,n-1}(\bullet)$ .

$$\phi_{0,0}(\bullet) = \phi(\bullet).$$

También podemos definir una operación de variación de escala unitaria en potencia de dos  $\kappa$ . La aplicación de  $\kappa$  a la función  $\phi(\bullet)$  produce una nueva función  $\phi_{1,0}(\bullet)$ .

$$\phi_{1,0}(t) \equiv \frac{\phi(\frac{t}{2})}{\sqrt{2}}$$

$$-\infty < t < \infty$$

aplicando sucesivamente  $\kappa$ :

$$\phi_{m,0}(t) \equiv \frac{\phi(\frac{t}{2^m})}{\sqrt{2^m}}$$

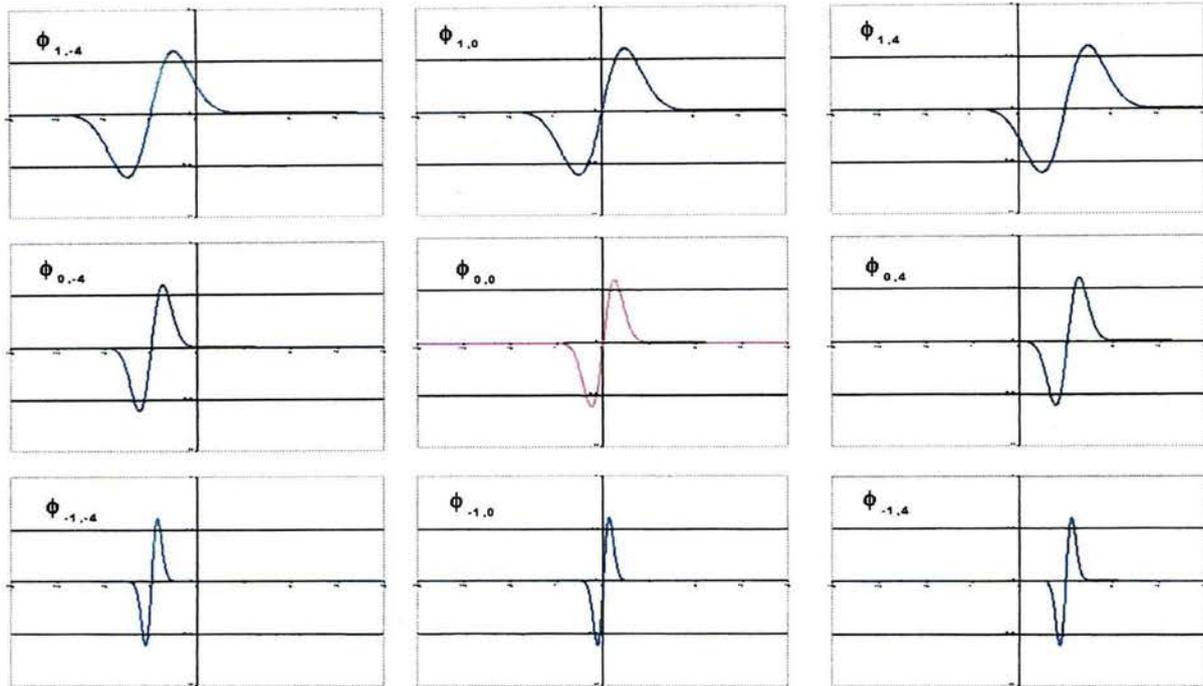
$$m \in \mathbb{Z};$$

Nuevamente  $m = -1$  define la operación inversa  $\kappa^{-1}$ . Finalmente, la composición de operaciones, (aplicando primeramente la traslación  $n$  y luego el escalamiento  $m$ , no son operaciones conmutativas),

esta determinada por:

$$\phi_{m,n}(t) \equiv \frac{\phi(\frac{t}{2^m} - n)}{\sqrt{2^m}}$$

$$-\infty < t < \infty$$



**Fig. 2.1** Traslación y escalamiento de la función  $\gamma = t \exp(-t^2/2)$

La fila central de la figura, muestra de izquierda a derecha:  $\phi_{0,-4}(\bullet)$ ,  $\phi(\bullet)$ ,  $\phi_{0,4}(\bullet)$ , para la función  $\phi(t) = t e^{(-\frac{t^2}{2})}$ , notar que  $n > 0$  traslada una función hacia la derecha, mientras que  $n < 0$  hace lo contrario. En la columna central, se muestra de arriba a abajo:  $\phi_{1,0}(\bullet)$ ,  $\phi(\bullet)$ ,  $\phi_{-1,0}(\bullet)$ . Notar que  $m > 0$  expande la gráfica de la función, en tanto que  $m < 0$  la contrae.

Sea  $f(\bullet) \in L^2(\mathfrak{R})$ . Sea  $\phi(\bullet) \in L^2(\mathfrak{R})$ , de forma que las traslaciones  $\{\phi_{0,n}(\bullet) : n \in \mathbb{Z}\}$  constituyan una base ortonormal para un subespacio cerrado  $V_0 \subset L^2(\mathfrak{R})$  donde:

$V_0 \equiv \{ \text{conjunto de todas las combinaciones lineales } \phi_{0,n}(\bullet) : n \in \mathbb{Z} \}$ . Entonces

$$\int_{-\infty}^{\infty} \phi_{0,n}(t)\phi_{0,p}(t)\partial t = \int_{-\infty}^{\infty} \phi(t-n)\phi(t-p)\partial t = \delta_{m,p} \quad (2.9)$$

$$\delta_{m,p} = \begin{cases} 1 & \text{sí } m = p \\ 0 & \text{de otra manera} \end{cases}$$

La función  $\phi(\bullet)$  se conoce como función de escalamiento, en tanto que el espacio  $V_0$  se denomina espacio de aproximación. Podemos analizar cualquier función  $f(\bullet)$  con respecto a  $V_0$  proyectando la función  $f(\bullet)$  en términos de las funciones base de  $V_0$ . Sí hacemos

$$\langle x(\bullet), y(\bullet) \rangle \equiv \int_{-\infty}^{\infty} x(t)y(t)\partial t$$

Representa el producto interno entre dos funciones  $x(\bullet)$  y  $y(\bullet)$ . La expresión para el análisis de la función  $f(\bullet)$  es:

$$v_{0,n} \equiv \langle x(\bullet), \phi_{0,k}(\bullet) \rangle = \int_{-\infty}^{\infty} f(t)\phi(t-n)\partial t \quad (2.10)$$

Los coeficientes  $v_{0,n}$  se denominan coeficientes de escala de la función  $f(\bullet)$ . Sí  $f(\bullet) \in V_0$ ,  $f(\bullet)$  puede expresarse en términos de sus coeficientes:

$$f(t) = \sum_{-\infty}^{\infty} v_{0,n}\phi_{0,n}(t) = \sum_{-\infty}^{\infty} v_{0,n}\phi(t-n) \quad (2.11)$$

Sí  $f(\bullet) \notin V_0$ , la expresión anterior constituye una aproximación a  $f(\bullet)$ , (proyección sobre  $V_0$ ). Realizando un cambio de variable en la ecuación (2.9), tenemos:

$$\int_{-\infty}^{\infty} \phi_{0,n}(t)\phi_{0,p}(t)\partial t = \int_{-\infty}^{\infty} \frac{\phi_{0,n}(\frac{t}{2^m})}{\sqrt{2^m}} \frac{\phi_{0,p}(\frac{t}{2^m})}{\sqrt{2^m}} \partial t = \int_{-\infty}^{\infty} \phi_{m,n}(t)\phi_{m,p}(t)\partial t \text{ donde}$$

$$\phi_{m,n}(t) \equiv \frac{\phi_{0,n}(\frac{t}{2^m})}{\sqrt{2^m}} = \frac{\phi(\frac{t}{2^m}-n)}{\sqrt{2^m}}$$

Esto es, dado que  $\{\phi_{0,n}(\bullet) : n \in \mathbb{Z}\}$  forma una base ortonormal para  $V_0$ ;  $\{\phi_{m,n}(\bullet) : n \in \mathbb{Z}\}$  constituye una base ortonormal para  $V_m$ , donde:

$V_m \equiv$  [ conjunto de todas las combinaciones lineales  $\{\phi_{m,n}(\bullet) : n \in Z\}$  es el espacio de aproximación con escala  $2^m$  ]. Sí tenemos que:

$$\bigcap_{m \in Z} V_m = \{0\} \text{ y } \overline{\bigcup_{m \in Z} V_m} \quad (2.12)$$

Podemos asegurar que una señal, se puede aproximar por una serie de subespacios anidados,  $V_j$ , [40].

Un análisis de multiresolución es por definición una secuencia de subespacios cerrados  $V_m$  de  $L^2(\mathfrak{R})$ ,  $m \in Z$ , de forma que:

- 1)  $V_m \subset V_{m-1}$  para toda  $m \in Z$ ;
- 2)  $f(\bullet) \in V_0 \Rightarrow f_{0,n}(\bullet) \in V_0$ , donde  $f_{0,n}(x) = f(x - n)$  para  $n \in Z$ ;
- 3)  $f(\bullet) \in V_0 \Rightarrow f_{m,0}(\bullet) \in V_0$ , donde  $f_{m,0}(x) = \frac{f(\frac{x}{2^m})}{\sqrt{2^m}}$  para  $m \in Z$ ;
- 4)  $\bigcap_m V_m = \{0\}$  y  $\overline{\bigcup_m V_m} = L^2(\mathfrak{R})$ ;
- 5) Existe una función de escalamiento  $\phi(x) \in V_0$ ; tal que  $\{\phi_{0,n}(\bullet) : n \in Z\}$  forman una base ortonormal de  $V_0$ , donde  $\phi_{0,n}(t) = \phi(t - n)$

Podemos definir la transformada de Ondeletas continua para  $f(\bullet) \in L^2(\mathfrak{R})$  mediante:

$$(W_\phi f)_{m,n} = \langle f(\bullet) \phi_{m,n}(\bullet) \rangle = \int_{-\infty}^{\infty} f(t) \phi_{m,n}(t) \partial t \quad (2.13)$$

La transformada es simplemente una combinación lineal de las funciones de escalamiento a diferentes niveles, esta descomposición en el tiempo permite conocer el desempeño de la función a diferentes escalas y su contribución a la función total. Cualquier función puede expresarse como la suma de  $m$  diferentes niveles.

$$f_m(t) = \sum_n x_n^m \phi_{m,n}(t) \quad (2.14)$$

La transformada inversa se define por:

$$f(t) = \sum_{m,n} x_n^m \phi_{m,n}(t) \quad (2.15)$$

## 2.4 Sistemas Dinámicos

De acuerdo con Tsonis y Elsner [15], un sistema dinámico es un sistema para el cual la evolución en el tiempo a partir de un estado inicial, que conocemos; se puede describir por un conjunto de ecuaciones matemáticas.

De esta manera, un sistema dinámico lo constituye un número de ecuaciones o reglas de comportamiento en el tiempo, que involucran a las  $n$  variables que influyen en el sistema. En nuestro caso, el sistema dinámico del tiempo/clima es simulado por un conjunto de ecuaciones diferenciales ordinarias de la forma:

$$\frac{dy}{dt} = F(\bar{y}) \quad (2.16)$$

donde cada  $\bar{y} = (y_1, \dots, y_n)$  representa un estado del sistema dentro del espacio de estados  $S$ .

## 2.5 Atractores

La dinámica del sistema descrito por el conjunto de ecuaciones (2.16), puede representarse, sin pérdida de información, dentro del llamado espacio fase. Un atractor  $A$  dentro del campo vectorial  $\Theta_t$  de un espacio fase; es un conjunto compacto con las siguientes características:

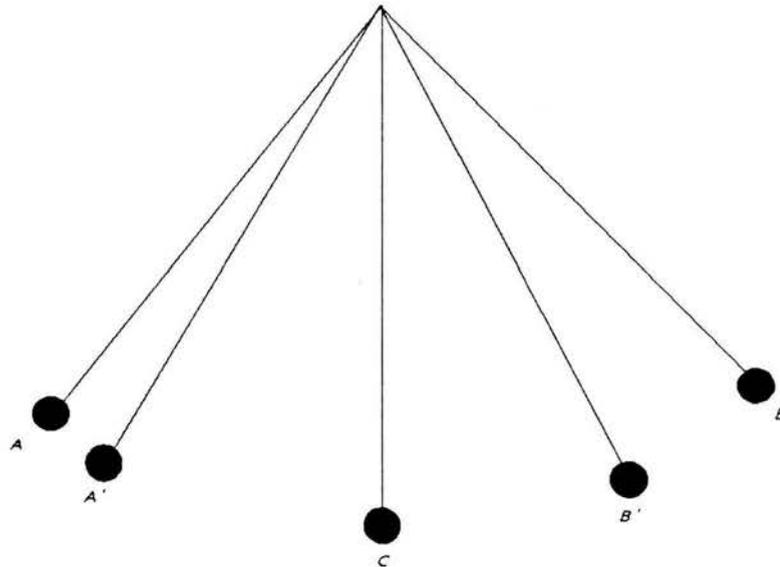
- (1.) Para todo tiempo  $t$ , el atractor  $A$  es invariante bajo la influencia del campo vectorial  $\Theta_t$ , ( $\Theta_t(A) = A$ ).
- (2.) El atractor  $A$  tiene una vecindad abierta (cuenca de atracción) que se contrae bajo la influencia

del campo vectorial  $\Theta_t$  sobre  $A$ .

(3.) No existe un subconjunto de  $A$ , que satisfaga las condiciones anteriores [23].

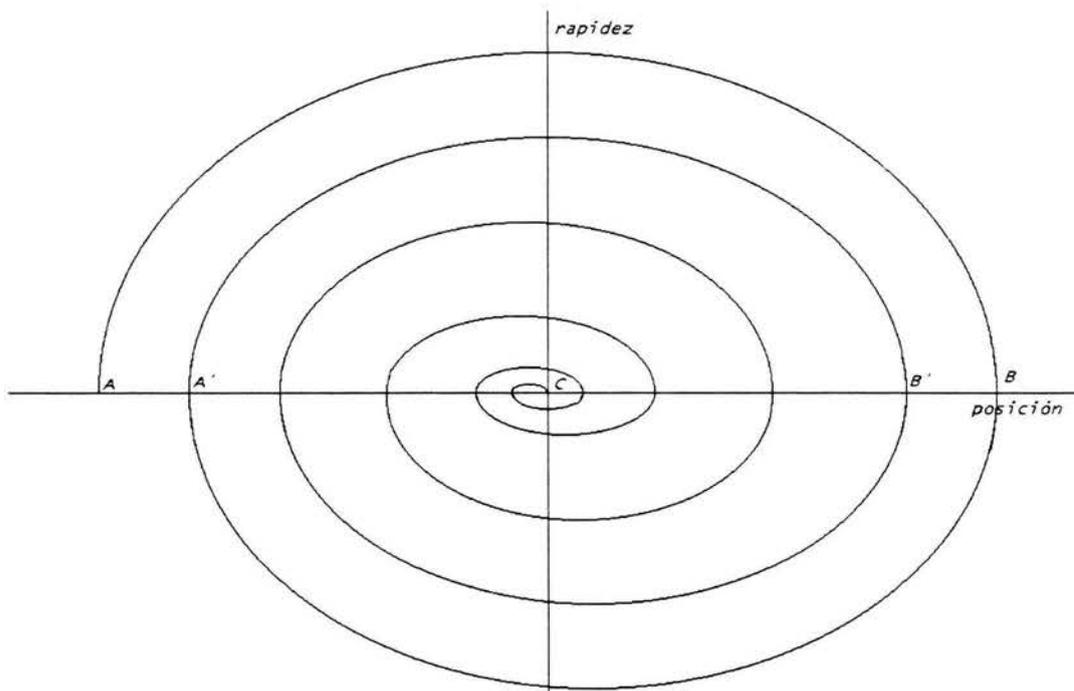
**Ejemplo:**

**Fig. 2.2**  
El péndulo en el espacio de estado



Consideremos el sistema dinámico simple de un oscilador armónico, como un péndulo bajo el efecto de fuerzas disipativas (fricción, resistencia del aire, etc.). El espacio fase estará compuesto por las variables: rapidez (velocidad) y su desplazamiento o posición (ángulo con respecto a la vertical). La posición  $A$ , corresponde al punto donde se inicia el movimiento (desplazamiento máximo), en este punto su rapidez es nula; al soltar el péndulo gracias a la fuerza de gravedad adquiere mayor velocidad hasta

llegar al punto C, donde su rapidez es máxima y su desplazamiento es nulo; al continuar su movimiento va perdiendo velocidad hasta llegar al punto B, en donde su rapidez se anula nuevamente; por efecto de la fricción el desplazamiento es inferior al logrado en A, el péndulo nuevamente se acelera hasta el punto C, donde su desplazamiento es nulo y su velocidad es inferior a la alcanzada anteriormente, de nuevo, el péndulo gana desplazamiento y pierde rapidez hasta alcanzar la posición A'; y sucesivamente, hasta que pierde toda velocidad y se detiene en la posición de equilibrio. Al graficar la rapidez en función del desplazamiento, cada punto en la gráfica representa la situación del sistema en un momento dado. La evolución del sistema corresponde a la trayectoria que une a todos los puntos. Como se muestra en la (Fig. 2.3), la línea converge al punto C.



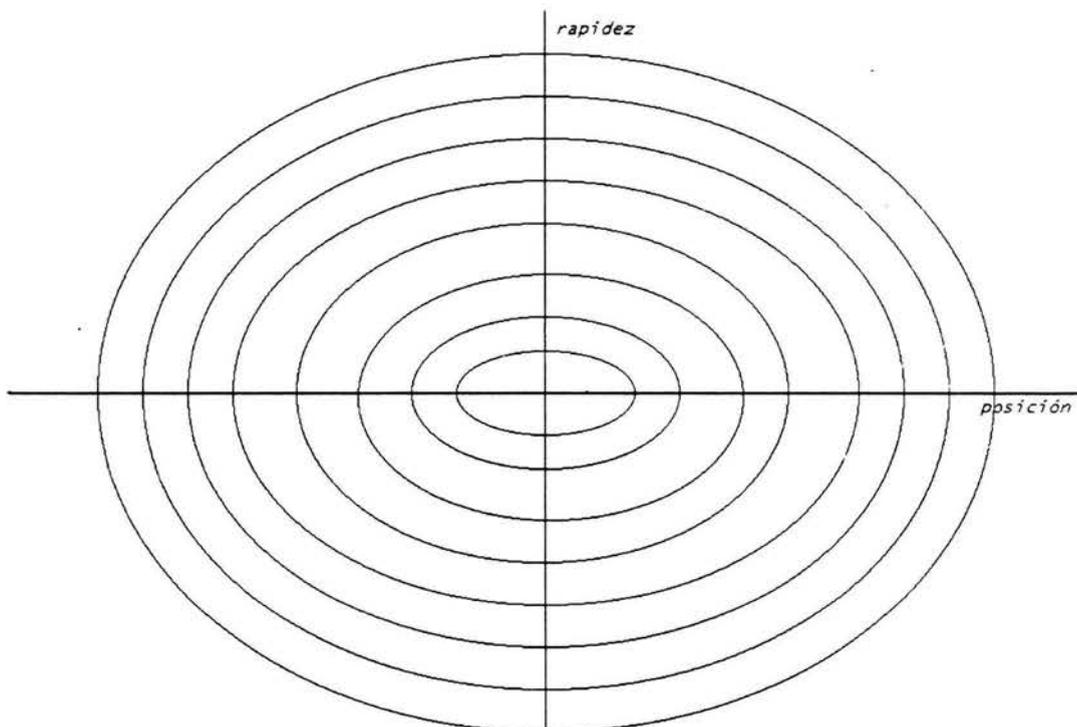
**Fig. 2.3** Representación de la trayectoria del péndulo en el espacio de fases bajo la acción de Fuerzas disipativas: (fricción, resistencia del aire, etc.)

El punto central en el espacio fase representa un péndulo en reposo. Como a este punto concurren to-

das las trayectorias independientemente de las condiciones iniciales (rapidez y elongación), se le conoce como punto atractor o atractor de punto fijo.

En el caso de que sobre el péndulo en estudio se aplique una fuerza que contrarresta el efecto de las fuerzas disipativas, el resultado es que el péndulo oscila regularmente sin pérdida de Energía.

En el espacio de fases el movimiento total del péndulo corresponde a un círculo o ciclo límite, (Fig. 2.4).



**Fig. 2.4** Atractor de ciclo límite. "Péndulo sin la acción de fuerzas disipativas"

El siguiente atractor, en complejidad, es el atractor de toro, que tiene la forma de un neumático; y

sobre su superficie se encuentran atrapadas todas las trayectorias del espacio fase. El atractor de toro, el atractor de ciclo límite y el atractor de punto se caracterizan por tener una dimensión entera; por lo tanto, el comportamiento a largo plazo no se ve afectado por las condiciones iniciales; esto es, dos trayectorias que inicialmente se encuentran cercanas, al evolucionar en el tiempo se mantienen sobre la superficie (superficie, línea, punto) del atractor, lo que garantiza predictibilidad a largo plazo, [15].

## 2.6 Dimensión Fractal

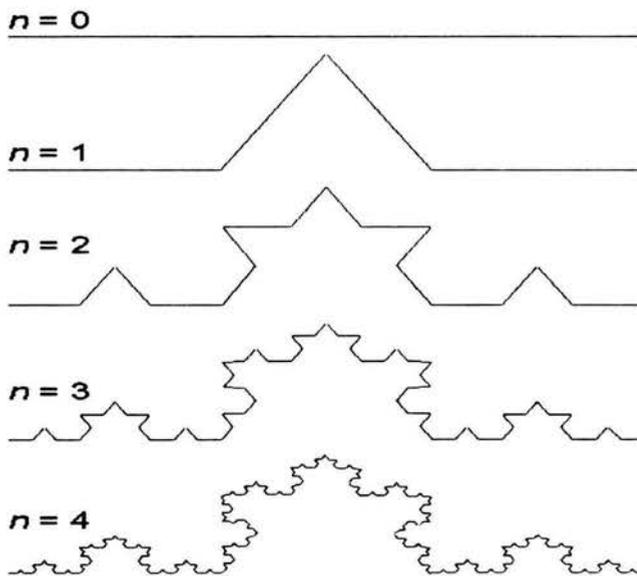
En la naturaleza existen sistemas dinámicos que cuentan con un atractor cuya dimensión no es un número entero y cuya forma geométrica permanece inalterada ante cambios de escala. Estos atractores se conocen como atractores fractales o extraños. En ellos el comportamiento a largo plazo se ve sensiblemente afectado por las condiciones iniciales, [1].

Nuestra concepción del mundo está regida por la geometría euclidiana, en la cual las dimensiones espaciales corresponden a números enteros (el punto tiene dimensión 0, una línea tiene dimensión 1, un plano tiene dimensión 2, etc.). Parafraseando a Benoit Mandelbrot: *'creemos que las nubes son esferas, las montañas son conos, las líneas costeras son círculos'*, [28].

Cada vez más, la geometría fractal se presenta ya no como un conjunto de *'curvas monstruo'* hechas a propósito, que rompen con nuestra idea de continuidad - diferenciabilidad, (función de Weierstrass, conjunto de Cantor, curva de Koch, curva de Besicovitch, etc.); sino como una característica inherente de

la naturaleza (la estructura fractal de la forma de las nubes es asociada con fluctuaciones de parámetros meteorológicos en todas las escalas del tiempo/clima), y en las leyes de la física (la ley de la fuerza de atracción gravitacional, magnética, eléctrica, entre dos objetos; que es inversamente proporcional al cuadrado de la distancia entre éstos), [32].

Veamos la curva de Koch, (Fig. 2.5). Ésta se genera a partir de un segmento de línea de longitud unitaria que es remplazada por un triángulo equilátero; de esta manera tenemos el primer paso de construcción, una curva formada por cuatro segmentos lineales de longitud  $\frac{1}{3}$ . La longitud total de la curva es de  $L(\frac{4}{3}) = \frac{4}{3}$ , proseguimos añadiendo en cada uno de sus lados un nuevo triángulo equilátero tres veces más pequeño que el original. Ahora tenemos una curva que consiste de  $N = 4^2 = 16$  segmentos de longitud  $d = 3^{(-2)} = \frac{1}{9}$ . La longitud de la curva es de  $L(\frac{1}{9}) = (\frac{4}{3})^2 = \frac{16}{9}$ .



**Fig. 2.5**  
Construcción de la curva de Koch.

Para la  $n$ -ésima iteración tenemos que la longitud de la curva es de  $L(d) = (4/3)^n$ . La longitud de

cada segmento es  $\delta = 3^{-n}$ . Como el número de iteraciones se puede expresar como  $n = -\frac{\ln \delta}{\ln 3}$ , la longitud de la curva puede expresarse como:  $L(\delta) = (\frac{4}{3})^n = \exp(-\frac{\ln \delta [\ln 4 - \ln 3]}{\ln 3}) = \delta^{-D}$

Esta ecuación es de la forma:

$$L(\delta) = a * \delta^{-D} \tag{2.17}$$

con  $D = \frac{\ln 4}{\ln 3} \cong 1.2628$ .

El número de segmentos es  $N(\delta) = 4^n = 4^{-\frac{\ln \delta}{\ln 3}}$ .

Designaremos a  $D$  como la dimensión fractal de la curva de Koch.

Notemos que cuando  $n$  crece,  $\delta$  tiende a cero y la longitud de la curva diverge.

Nos encontramos con una objeto que a pesar de estar definido sobre una región finita del espacio posee una frontera de extensión ilimitada. Claramente no es posible dibujar una tangente en algún punto a lo largo de su perímetro. Es más que una recta, pero no llega a formar un plano.

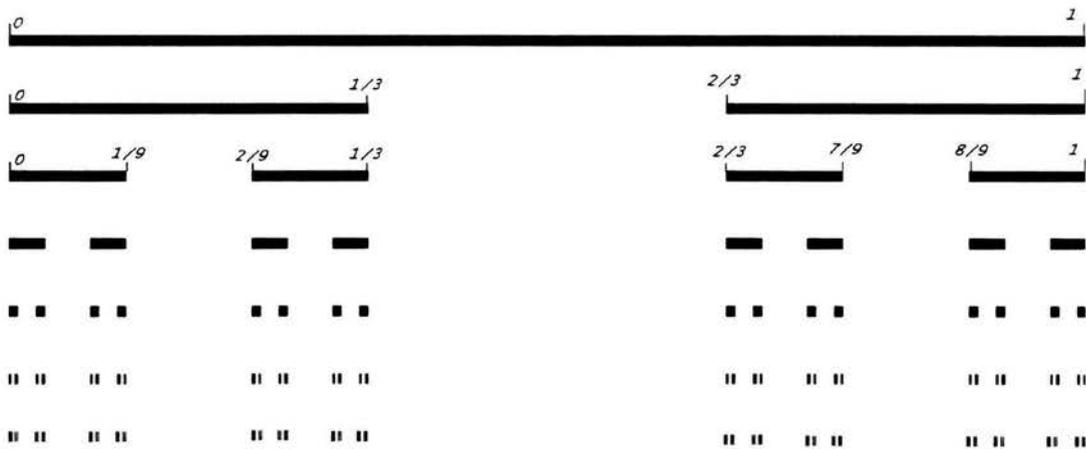


Fig. 2.6 Construcción del conjunto simétrico de Cantor

Ahora veamos el conjunto de Cantor simétrico (Fig. 2.6). Éste se obtiene a partir de un segmento de línea de longitud igual a la unidad, que se divide en tercios y se elimina el tercio medio sin incluir los extremos (conjunto abierto). Al siguiente paso cada uno de los dos segmentos restantes se separa a su vez en tercios y se suprime el tercio medio de cada uno de ellos. Al siguiente paso cada uno de los cuatro segmentos restantes es dividido a su vez en tercios, eliminando el tercio medio para cada uno. Y así, sucesivamente, hasta infinito; conceptualmente, hasta un número de iteraciones 'muy grande'. El conjunto de Cantor, o 'polvo de Cantor', lo forman los puntos que permanecen contenidos en el segmento original; la longitud total de los intervalos que se han eliminado  $L_e$  es 1, para mostrar esta afirmación, de la (Fig. 2.6).

$$L_e = 1\left(\frac{1}{3}\right) + 2\left(\frac{1}{9}\right) + 4\left(\frac{1}{27}\right) + \dots + 2^{(n-1)}\left(\frac{1}{3^n}\right) = \frac{1}{3} \sum_{n=0}^{\infty} \left(\frac{2}{3}\right)^n = \frac{1}{3} \left(\frac{1}{1 - \frac{2}{3}}\right). \quad (2.18)$$

Y, por lo tanto, dado que la longitud inicial es 1, la longitud del conjunto de Cantor es 0, aunque el número de puntos es infinito, [31].

Tabla 2.1 Dimensión del conjunto simétrico de Cantor

paso (n)	número de segmentos N(l)	longitud segmentos (l)
1	$2=2^1$	$1/3=(1/3)^1$
2	$4=2^2$	$1/9=(1/3)^2$
3	$8=2^3$	$1/27=(1/3)^3$
4	$16=2^4$	$1/81=(1/3)^4$
•	•	•
•	•	•
•	•	•
k	$2^k$	$(1/3)^k$

Para calcular la dimensión del conjunto de Cantor consideremos el siguiente proceso: construimos cubos (segmentos) de longitud cada vez más pequeña, determinando a cada iteración cuántos de estos cubos (segmentos); requerimos para cubrir el conjunto de Cantor, ver tabla anterior, (Tabla 2.1).

De esta manera, la dimensión  $D$  se determina por:

$$D = \lim_{l \rightarrow 0} \frac{\log N(l)}{\log(\frac{1}{l})} = \frac{\log 2^k}{\log(\frac{1}{(\frac{1}{3})^k})} = \frac{\log 2^k}{\log 3^k} = 0.630929753 \quad (2.19)$$

Nuevamente nos encontramos con una expresión de la forma (2.17), [35].

Ahora tenemos un conjunto que es más que un punto, pero no alcanza a formar una recta.

El Conjunto simétrico de Cantor no podría concebirse dentro de la simple geometría euclidiana, donde la dimensión 0 se asocia a un punto y la dimensión 1 corresponde a una línea; no obstante, en la naturaleza se han encontrado innumerables comportamientos que son fácilmente modelados utilizando el polvo de Cantor.

Por ejemplo: La distribución temporal de desbordamientos en Venecia, [8]. El espaciamiento de los anillos de Saturno, [9]. La forma geométrica de las fracturas en la corteza terrestre [10], ver también [30], [33].

## 2.7 Dinámica Cuantitativa

Retomemos la ecuación (2.16):

$$\frac{d\mathbf{y}}{dt} = \mathbf{F}(\bar{\mathbf{y}})$$

Donde cada  $\mathbf{y} = (y_1, y_2, \dots, y_n)$  representa un punto en el espacio de estados  $S$ ; la dimensión de  $S$  se encuentra asociada con el número de grados de libertad del sistema.

El campo vectorial  $\mathbf{F}(\bar{\mathbf{y}})$  es un operador no-lineal; define, bajo ciertas condiciones de diferenciabilidad, un problema de valores iniciales, y podemos escribir la solución al tiempo  $t$ , dado un valor inicial  $\mathbf{y}_0$ , como  $\mathbf{y}(t) = \phi_t(\mathbf{y}_0) \bullet \phi_t$ , representando una familia paramétrica de curvas solución  $\phi_t(S)$ .

En principio, la dimensión de  $\phi_t(S)$  sería la misma que el espacio  $S$ . Al evolucionar el sistema en el caso de sistemas disipativos existe una contracción en conjuntos (atractores) de menor dimensión.

En el atractor, el sistema tiene un número menor de grados de libertad y, consecuentemente, requiere menos información para especificar su estado.

La solución del sistema (2.16) es equivalente a determinar completamente a la familia de soluciones  $\phi_t(S)$ , en caso de lograr describir geoméricamente la trayectoria de las soluciones en el espacio fase podremos extraer información cualitativa que nos permita comparar y clasificar diferentes sistemas.

Dos campos vectoriales de dimensión  $C^n$  (orden  $n$ ),  $F$  y  $G$ , son  $C^m$  equivalentes ( $m < n$ ), si existe un difeomorfismo (cambio de coordenadas),  $\Theta$  que tome órbitas  $\phi_t(\mathbf{y})$  de  $F$ , y las transforme en órbitas  $\Phi_t(\Theta(\mathbf{y}))$  de  $G$ , conservando orientación, (Fig 2.7).

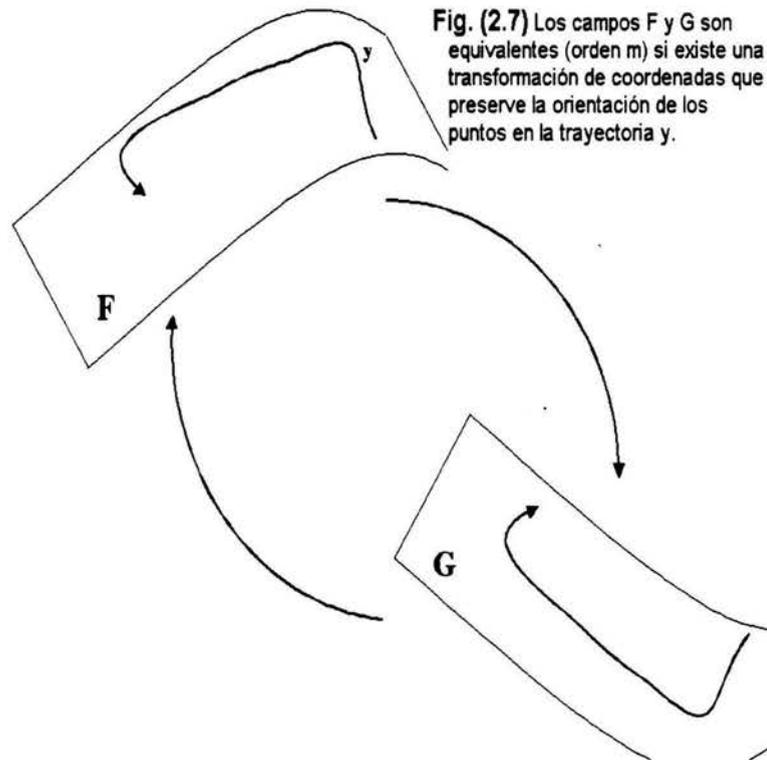


Fig. (2.7) Los campos  $F$  y  $G$  son equivalentes (orden  $m$ ) si existe una transformación de coordenadas que preserve la orientación de los puntos en la trayectoria  $y$ .

Bajo la equivalencia, tenemos:

- I. El punto  $\mathbf{y}$  es un punto singular del campo  $F$ , si  $\phi(\mathbf{y})$  es una singularidad del campo  $G$ .
- II. Una órbita cerrada de  $F$  es una órbita cerrada de  $G$ .
- III. La frontera de  $\phi(\mathbf{y})$  en  $G$ , corresponde a la frontera de  $F$ .

Ésto nos garantiza que cualquier objeto se preserva a través de la relación de equivalencia. Podemos afirmar que bajo la equivalencia, ambos estados tienen la misma dinámica cualitativa.

## 2.8 Búsqueda de Atractores en TIEMPO/CLIMA

La dinámica del sistema climático es descrita en el espacio de estado por  $n$  ecuaciones diferenciales parciales que describen el comportamiento físico del sistema (análogamente con (2.16)).

$$\dot{x}_1 = f_1(x_1, x_2, \dots, x_n)$$

$$\dot{x}_2 = f_2(x_1, x_2, \dots, x_n)$$

$$\vdots$$

$$\dot{x}_n = f_n(x_1, x_2, \dots, x_n)$$

donde  $\dot{x}$  representa la derivada de  $x$  con respecto al tiempo. En este caso, las coordenadas dentro del espacio de estado son  $(x_1, x_2, \dots, x_n)$  (vector espacio de estado).

El sistema de ecuaciones puede, mediante derivación sucesiva de una de las ecuaciones que describen el sistema; convertirse en una ecuación diferencial de orden  $n$ -ésimo de la forma (por ejemplo, diferenciando sucesivamente  $x_1$ ):

$$x_1^n = f(x_1, \dot{x}_1, \dots, x_1^{(n-1)}), \quad (2.20)$$

reemplazando las coordenadas del espacio de estado con  $(x_1, \dot{x}_1, \dots, x_1^{(n-1)})$  (vector espacio de fase).

La nueva trayectoria  $(x_1, \dot{x}_1, \dots, x_1^{(n-1)})$  del sistema 'transformado' en el espacio fase, describe la misma dinámica del espacio de estados original, sin pérdida de información. En principio, cualquier variable pudo ser escogida [15].

Sorprendentemente, toda la evolución de un sistema puede obtenerse a partir de la evolución de una sola de sus variables.

Consideremos como ejemplo el sistema propuesto en 1976 por Otto E. Rössler, el cual constituye la muestra más elemental de caos dentro de sistemas continuos [11].

sí (' significa derivada con respecto al tiempo), las ecuaciones de Rössler son:

$$\mathbf{x}' = -z - \mathbf{y} \quad (2.21)$$

$$\mathbf{y}' = x + a\mathbf{y} \quad (2.22)$$

$$\mathbf{z}' = b + z(x - c) \quad (2.23)$$

Tomando la ecuación (2.22) y derivando sucesivamente

$$y'' = x' + ay' \quad (2.24)$$

$$y''' = x'' + ay'' \quad (2.25)$$

De (2.21) se obtiene

$$x'' = -z' - y' \quad (2.26)$$

Sustituyendo en (2.25),

$$y''' = -z' - y' + ay'' \quad (2.27)$$

Empleando la ecuación (2.23),

$$y''' = -[b + z(x - c)] - y' + ay'' \quad (2.28)$$

Despejando (2.24), resulta

$$x = y' - ay \quad (2.29)$$

$$y''' = -[b + z(y' - ay - c)] - y' + ay'' \quad (2.30)$$

Utilizando (2.21) y derivando nuevamente (2.29), llegamos a

$$x' = y'' - ay' \implies z = -(y'' + ay') - y = -y'' + ay' - y \quad (2.31)$$

$$y''' = -[b + (-y'' + ay' - y)(y' - ay - c)] - y' + ay'' \quad (2.32)$$

Luego entonces, tenemos que el sistema de ecuaciones de Rössler es descrito, sin pérdida de información, por una sola ecuación de orden mayor; en términos de una sola variable dentro del espacio de fase [7].

$$y''' = F(a, b, c, y, y', y'') \quad (2.33)$$

## 2.9 Envoltentes

Diferentes sistemas dinámicos pueden tener una equivalencia de clase en el espacio fase, al observar que estos sistemas presentan atractores de dimensión semejante. Ésto nos lleva a la posibilidad de que fenómenos físicamente diferentes puedan ser descritos mediante una misma dinámica cualitativa; y que evolucionen mediante una clase equivalente de modelo canónico.

Un envolvente es un mapeo, digamos  $\Phi$ , del espacio  $M$  a un subespacio  $U$  de forma tal que la imagen de  $\Phi(M) \subset U$  es un subconjunto 'suave' de  $U$ , y  $\Phi$  constituye un difeomorfismo entre  $M$  y  $\Phi(M)$

El teorema importante relacionado con envoltentes fue demostrado por Takens en 1981, estableciendo:

Sea  $M$  un espacio compacto de dimensión  $m$ . Dados  $(F, v)$ ,  $F$  campo vectorial 'suave' (i.e.  $C^2$ ) y

$v$  una función 'suave' definida en  $M$ , es una propiedad genérica que  $\phi_{(F,v)}(\mathbf{y}) : M \rightarrow R^{(2m+1)}$ , definida por:

$$\Phi_{(F,v)}(\mathbf{y}) = (v(\mathbf{y}), v(\phi_1(\mathbf{y})), \dots, v(\phi_{2m}(\mathbf{y})))^T$$

Es una envolvente, donde  $\phi_t$  es el flujo del campo vectorial  $F$ ;  $v(\mathbf{y})$  corresponde al valor de una medida de la variable de estado  $\mathbf{y} \in M$ , [23].

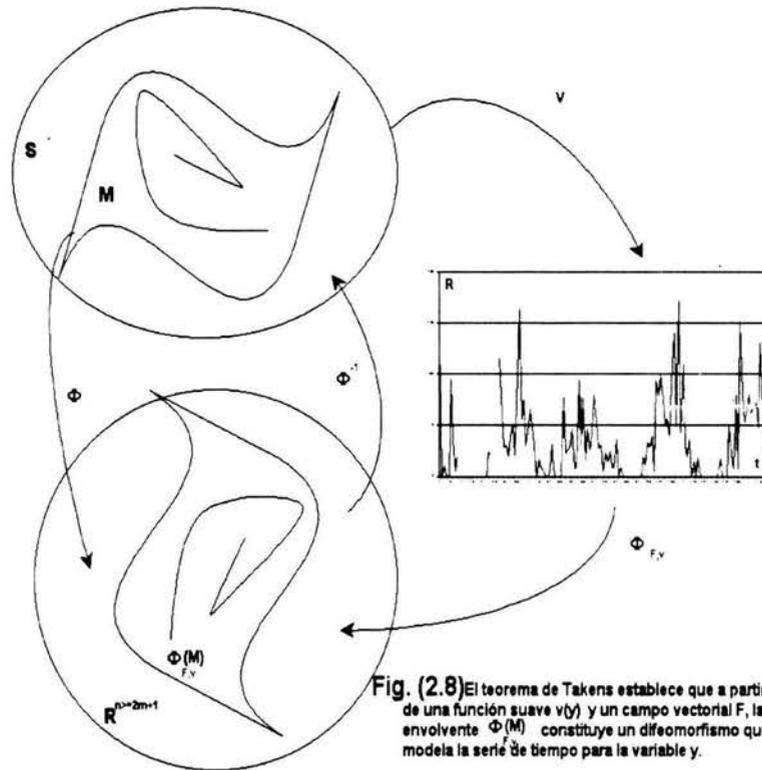


Fig. (2.8) El teorema de Takens establece que a partir de una función suave  $v(y)$  y un campo vectorial  $F$ , la envolvente  $\Phi(M)$  constituye un difeomorfismo que modela la serie de tiempo para la variable  $y$ .

Los atractores forman superficies (sábanas) invariantes de los sistemas dinámicos, que regularmente tienen una dimensión  $d$  menor a la del espacio fase  $n$ ,  $d < n$ . Esto quiere decir que estos atractores (superficies) con dimensión  $d$ , pueden estar cubiertos (envueltos) en un espacio  $n$  dimensional. Según Fraedrich, 1987 [17] y otros,  $n \geq 2d + 1$ .

### 2.10 Método de retrasos

En la práctica, a partir del teorema de Takens, es claro que para poder determinar la dimensión envolvente del sistema en estudio, necesitamos asociar una serie de medidas en intervalos de tiempo idénticos, respecto a una misma variable, (serie de tiempo):  $x_1, x_2, \dots, x_i, x_{i+1}, \dots$ , donde  $x_i \equiv x(\varphi_i(y))$ .

En principio, el intervalo de tiempo entre dos valores sucesivos de la serie de tiempo no necesariamente coincide con el intervalo de tiempo (que desconocemos) entre dos mapeos sucesivos  $\phi_i, \phi_{i+1}$ .

Para comprender el método de retrasos, utilizaremos el concepto de ventana  $(n, j)$  para referirnos a un conjunto formado por  $n$  elementos, separados  $j$  intervalos de tiempo en nuestra serie de tiempo original.

(a) ventana (5,1)

$$v_1 \quad v_2 \quad \cdot \quad \cdot \quad \cdot \quad v_{i-1} \quad \boxed{v_i} \quad \boxed{v_{i+1}} \quad \boxed{v_{i+2}} \quad \boxed{v_{i+3}} \quad \boxed{v_{i+4}} \quad v_{i+5} \quad \cdot \quad \cdot \quad \cdot$$

$$X_i = (v_i, v_{i+1}, v_{i+2}, v_{i+3}, v_{i+4})^T$$

(b) ventana (5,3)

$$v_1 \quad v_2 \quad \cdot \quad \cdot \quad \cdot \quad \boxed{v_i} \quad \cdot \quad \cdot \quad \boxed{v_{i+3}} \quad \cdot \quad \cdot \quad \boxed{v_{i+6}} \quad \cdot \quad \cdot \quad \boxed{v_{i+9}} \quad \cdot \quad \cdot \quad \boxed{v_{i+12}} \quad \cdot \quad \cdot \quad \cdot$$

$$X_i = (v_i, v_{i+3}, v_{i+6}, v_{i+9}, v_{i+12})^T$$

Fig. (2.9) Ejemplos del uso del método de retardos:

(a) construcción de un vector en  $R^5$  a partir de una serie de tiempo de la variable  $v$  utilizando "saltos" de longitud igual a un intervalo de tiempo

(b) Igual que en (a), pero con "saltos" de longitud igual a tres intervalos de tiempo.

Al subespacio del espacio de fase que contiene la imagen  $\Phi_{F,v}$  lo denominaremos como espacio envolvente y su dimensión corresponde a la dimensión envolvente  $n$ . Ver figura (2.8). De esta manera los elementos de la ventana  $(n, j)$  constituyen las componentes de un vector en el espacio envolvente,  $\mathbf{R}^n$ , esto es:  $\mathbf{x}_i = \Phi_{F,v}(\varphi_i(\mathbf{y})) = (v_i, v_{(i+J)}, \dots, v_{i+(n-1)J})^T$ .

Geoméricamente, y enlazando las secciones anteriores, en lugar de una variable continua  $\mathbf{x}(t)$  y sus  $(n - 1)$  derivadas  $\dot{\mathbf{x}}, \dots, \mathbf{x}^{(n-1)}$ , tenemos una serie de datos discreta  $\mathbf{x}(t), \mathbf{x}(t + 1), \dots, \mathbf{x}(t + n)$  y, análogamente a la derivada, consideraremos saltos de longitud  $\tau(J)$ . Nuestro vector de espacio de fase se define como

$$\mathbf{x} = [\mathbf{x}(t), \mathbf{x}(t + \tau), \dots, \mathbf{x}(t + (n - 1)\tau)].$$

Seleccionamos  $\tau$  de forma que nos garantice independencia entre  $\mathbf{x}(t)$  y  $\mathbf{x}(t + \tau)$ ; de otra manera,  $\mathbf{x}(t + \tau) \cong \mathbf{x}(t)$ , y todos los puntos tienden a encontrarse sobre una diagonal. En el presente trabajo aplicaremos el método propuesto por Fraser y Swinney, [20].

En este punto, tenemos que el atractor de un sistema dinámico está contenido en un espacio de fase que se construye a partir de una serie de datos y sus retrasos, sin afectar las propiedades topológicas del sistema en estudio. Para un espacio de fase  $n$  dimensional se observa una nube de puntos. Para determinar la dimensión envolvente, como en el caso del conjunto simétrico de Cantor, se utiliza un número  $N(l)$  de elementos de volumen (cubos), de longitud  $l$  y dimensión  $d$ , para cubrir al volumen  $V$  que contiene todos los puntos asociados a un estado del sistema.

$$V = \sum_{N(l)} l^d = N(l)l^d. \tag{2.34}$$

$$N(l) = V l^{-d} \implies l^d = \frac{V}{N(l)}. \quad (2.35)$$

Introduciendo  $V_o$  y  $l_o$ , unidades escalares

$$d = \frac{\ln(N(l))}{\ln(\frac{l_o}{l})} + \frac{\ln(\frac{V}{V_o})}{\ln(\frac{l}{l_o})}. \quad (2.36)$$

Claramente, el segundo término se anula cuando  $l \rightarrow 0$

$$d = \lim_{l \rightarrow 0} \frac{\ln(N(l)) - \ln(v)}{\ln(\frac{l}{l_o})}. \quad (2.37)$$

Finalmente,

$$N(l) = \left(\frac{l}{l_o}\right)^{-d}. \quad (2.38)$$

A partir de la ecuación anterior, la dimensión  $d$  se puede encontrar mediante una gráfica de  $\ln(N)$  vs.  $\ln(\frac{l}{l_o})$ .

En lugar de contar elementos de volumen o cajas, contaremos pares de puntos

$$\mathbf{x}_m(t_i) = (\mathbf{x}(t_i), \dots, \mathbf{x}(t_i + (m-1)\tau)),$$

$$\mathbf{x}_m(t_j) = (\mathbf{x}(t_j), \dots, \mathbf{x}(t_j + (m-1)\tau));$$

que se encuentran separados a lo más por una distancia fija  $l$ , mediante una función de acumulación

$N_m(l)$ , dada por:

$$N_m(l) = \sum_{i,j=1}^N \Theta(l - r_{ij}(m)) \quad (2.39)$$

con  $r_{ij}(m) = |\mathbf{x}_m(t_i) - \mathbf{x}_m(t_j)|$  y  $|t_i - t_j| > \tau$ , donde  $\Theta$  es la función Heaviside:

$$\theta(a) = \begin{cases} 1 & \text{sí } a \leq 0 \\ 0 & \text{de otra manera} \end{cases}$$

La función de acumulación  $C_m(l)$  está normalizada por el total de pares de puntos  $N^2$

$$C_m(l) = \frac{N_m(l)}{N^2}. \quad (2.40)$$

donde  $N$  es el número total de puntos

Esto quiere decir que la dimensión de un objeto geométrico, como un atractor en el espacio fase; puede determinarse mediante la función de acumulación de distancias de pares de puntos situados en la trayectoria del sistema dinámico, y obteniendo el valor de la pendiente de la gráfica de  $\ln(N(l))$  contra  $\ln(\frac{l}{l_0})$ :

$$d = \frac{\ln(N(l))}{\ln(\frac{l}{l_0})} \quad (2.41)$$

## 2.11 Coeficientes de Lyapunov

Los coeficientes de Lyapunov son un promedio de la razón exponencial de convergencia o divergencia de trayectorias que se encontraban inicialmente dentro de un elemento diferencial de volumen.

Cualquier sistema que cuente con al menos un exponente positivo de Lyapunov se define como caótico y la magnitud del coeficiente indica la escala de tiempo a la cual el sistema se vuelve impredecible [3].

Intuitivamente, se toma un elemento diferencial de volumen, digamos una esfera, en nuestro espacio de fases  $n$  dimensional, y se sigue su evolución en el tiempo; nuestra esfera se convierte en un elipsoide debido a deformaciones naturales en la evolución del sistema. Para una dimensión ( $n = 1$ ),

$$\frac{dx}{dy} = f(x), \quad (2.42)$$

al evolucionar en forma no-lineal en el tiempo, se presentan pequeñas desviaciones  $\delta(x)$  que obedecen a:

$$\frac{d}{dt}(\delta x) = \left. \frac{df}{dx} \right|_{x_0} \delta x \quad (2.43)$$

Estas desviaciones crecen exponencialmente en el tiempo  $t = t_0 + m\tau$ , incrementándose por  $m$  pasos de tiempo de duración  $\tau$ :

$$\delta x = \delta x_0 e^{(\lambda m \tau)} \quad (2.44)$$

El exponente característico (eigenvalor)  $\lambda$  se define por el operador Jacobiano  $\lambda = \left. \frac{df}{dx} \right|_{x_0}$ .

Despejando  $\lambda$  y sustituyendo en la expresión anterior, si tomamos el límite  $m \rightarrow \infty$ , obtenemos el promedio de crecimiento exponencial (divergencia):

$$\lambda(x_0, \delta x_0) = \lim_{m \rightarrow \infty} \frac{1}{m\tau} \ln(|\delta x|) \quad (2.45)$$

Para generalizar a  $n$  dimensiones, tomamos  $D = |\delta(x)|$ :

$$\lambda(x_0, \delta(x_0)) = \lim_{m \rightarrow \infty} \frac{1}{m\tau} \ln |D(t)| \quad (2.46)$$

$\lambda$  toma el máximo de los  $n$  valores posibles:  $\lambda_1, \lambda_2, \dots, \lambda_n$ . Estos exponentes característicos definen la razón de expansión o contracción del elemento diferencial de volumen.

Los coeficientes de Lyapunov proporcionan información sobre la evolución de las condiciones iniciales del sistema en estudio. En el caso de atractores de punto fijo (puntos de equilibrio)  $\lambda_i < 0$  para toda

$i, i = (1, \dots, n)$ . Para un atractor de ciclo límite  $\lambda_1 = 0$  y  $\lambda_i < 0$  para  $i = (2, \dots, n)$ . En un atractor en forma de toro  $\lambda_1 = 0, \lambda_2 = 0$  y  $\lambda_i < 0$  para  $i = (3, \dots, n)$ . En caso de que exista al menos un coeficiente positivo, tendremos un sistema caótico.

Cada coeficiente positivo describe una dirección en la que el sistema se estira, generando divergencia entre órbitas que inicialmente se encontraban cercanas, contribuyendo a la divergencia o expansión de un elemento de volumen centrado inicialmente en el estado  $x_0$  en el espacio fase; la suma de todos los coeficientes positivos nos proporciona la razón exponencial de crecimiento  $h$ :

$$h(x_0) = \sum_{\lambda_i > 0} \lambda_i \quad (2.47)$$

El inverso  $\frac{1}{h}$ , expresa una medida de tiempo sobre la cual es posible la predictibilidad.

## 2.12 Dimensión de Atractores para una dimensión envolvente fija $m$

Consideremos una distribución homogénea de puntos en el espacio fase, el número de puntos dentro de una esfera de radio  $l$ , crece en forma proporcional a la dimensión que se considere ( $l, l^2, l^3$ ). En la presencia de un atractor contenido dentro de una dimensión  $m$  suficientemente grande, el número de puntos converge, conforme a la función de distancia acumulativa  $C_m(l)$ , y es proporcional al valor  $l^{d_\infty}$ , esto es:

$$C_m(l) \approx l^{d_\infty} \quad (2.48)$$

$d_\infty$  es el valor al que converge la dimensión del atractor para cuando  $m \rightarrow \infty$

Un punto  $\mathbf{x}(t_0) = (\mathbf{x}(t_0), \mathbf{x}(t_0 + \tau), \dots, \mathbf{x}(t_0 + (m-1)\tau))$  en el espacio fase  $m$ -dimensional define una sección de trayectoria de tiempo de los  $(m-1)\tau$  intervalos. La función de acumulación  $C_m(l)$ , define el número medio de puntos que permanecen juntos dentro de una esfera de radio  $l$ , durante el lapso de tiempo  $(m-1)\tau$ ; como las trayectorias no iniciaron al tiempo  $t_0$ , es necesario promediar sobre cada uno de los puntos (estados) del sistema.

Al considerar el segmento de tiempo siguiente  $(m)\tau$ , la nueva función acumulativa  $C_m(l)$  proporciona un número diferente de puntos (inferior) dentro de la esfera de radio  $l$ ; esto es, existen varias trayectorias que ya no inciden dentro de la esfera de radio  $l$ . Este cambio nos proporciona una medida de la razón de divergencia de las trayectorias que inicialmente se encontraban a una distancia menor que  $l$  sobre el atractor, lo que es equivalente a los coeficientes de Lyapunov.

Consideremos un par de trayectorias que han permanecido dentro de la esfera de radio  $l$  durante  $(m-1)\tau$  intervalos de tiempo; la probabilidad de que para el siguiente intervalo se mantengan dentro del esferoide es proporcional a  $e^{-\lambda m \tau}$ ; únicamente los términos con  $\lambda > 0$  contribuyen; de esta forma, el número promedio de puntos dentro del esferoide de radio  $l$  decrece en forma proporcional a  $e^{-m\tau h}$

$$C_m \approx e^{-m\tau h} \quad (2.49)$$

De esta forma, la predictibilidad  $h$  puede interpretarse como el promedio sobre todos los puntos del atractor, de la razón exponencial de divergencia de trayectorias en el tiempo  $mt$  [16].

# Capítulo 3

## REDES NEURONALES

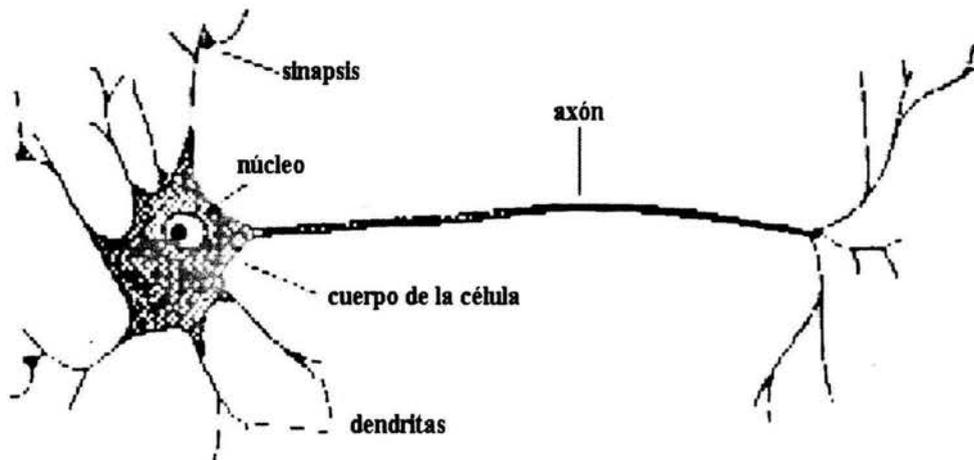
### 3.1 Introducción

La manifiesta superioridad del cerebro humano sobre la más sofisticada supercomputadora ha dado origen, desde los años cuarentas, a una prolífera y sorprendente área del conocimiento humano.

En particular, las redes neuronales artificiales se encuentran inspiradas en el conocimiento que brinda la Biología respecto al comportamiento del cerebro humano y cómo se procesan los mensajes que recibimos del mundo externo.

El elemento fundamental del cerebro lo constituyen las células nerviosas o neuronas (el cerebro humano está compuesto aproximadamente por  $10^{11}$  neuronas). Un esquema básico de una neurona nos muestra una red de fibras nerviosas (dendritas) que se conectan al cuerpo de la célula (soma), donde se ubica el núcleo de la célula; de este cuerpo se desprende una fibra nerviosa (axón) que subsecuentemente se ramifica en filamentos y subfilamentos; al final de éstos se encuentran las uniones sinápticas o sinapsis

con otras neuronas. El axón usualmente se conecta mediante varios cientos de sinapsis con otras neuronas; las terminales receptoras de estas uniones se localizan en el soma y en las dendritas.



**Fig. (3.1) Representación esquemática de una neurona**

En 1943, McCulloch y Pitts propusieron un poderoso, pero simple, modelo computacional de una neurona como una unidad binaria de conexión que realiza una suma de los valores de las conexiones que llegan a ella y emite una salida 1 o 0 conforme al valor que encuentra,

$$n_i(t + 1) = \theta\left[\sum_j w_{ij}n_j(t) - \mu_i\right] \quad (3.1)$$

$n_i$  puede valer 1 o 0, correspondiendo al estado activo o inactivo de la neurona, el tiempo es una variable discreta y  $\theta(x)$  es la función de Heaviside:

$$\theta(x) = \begin{cases} 1 & \text{sí } x \geq 0 \\ 0 & \text{de otra manera} \end{cases}$$

El peso  $w_{ij}$  representa el valor de la sinapsis que enlaza a la neurona  $i$  con la neurona  $j$ , puede ser positivo o negativo; en caso de valer 0 no existe sinapsis entre las neuronas  $i$  y  $j$ ; el valor  $\mu_i$  es el valor de activación del nodo  $i$ .

Una de las razones que ha hecho atractivo el desarrollo y la implementación de las redes neuronales artificiales es la habilidad que tienen éstas para generalizar hacia nuevas situaciones. Después de haber sido entrenada con un número de ejemplos de una relación, la red puede inducir un comportamiento que interpola y extrapola la secuencia aprendida en una forma aproximada.

### 3.2 Memoria Asociativa

El primer modelo que analizaremos consiste en establecer una red, que como a un bebé, al que le mostramos un objeto y simultáneamente repetimos el nombre del objeto hasta que desarrolle la capacidad de asociar el objeto con el nombre, y viceversa. La red debe ser capaz de que una vez que ha aprendido un patrón o una imagen, la muestre como respuesta en la presencia de una imagen o patrón similar, e incluso ante la presencia de imágenes incompletas o difusas del mismo objeto.

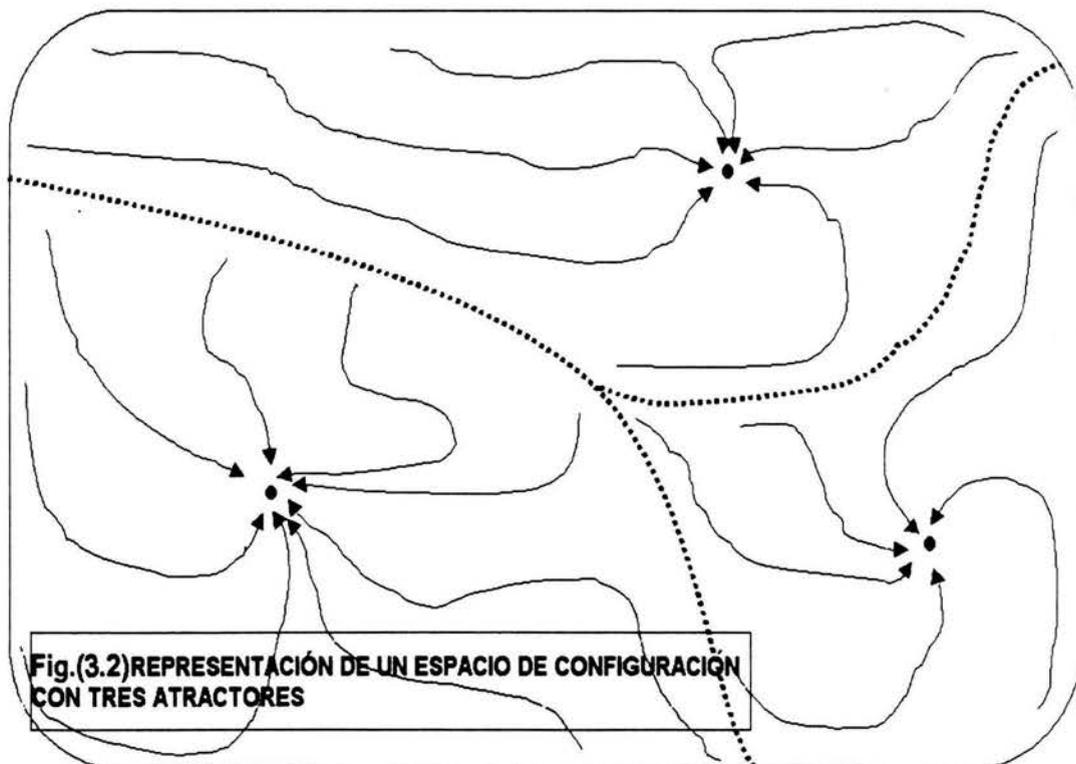
Planteado de otra manera, el problema consiste en establecer un conjunto de  $p$  patrones  $\xi_i^n$  de forma tal que en la presencia de un nuevo patrón  $\xi_i$ , la red regrese aquel patrón del conjunto establecido que más se asemeje al patrón  $\xi_i$ .

Parecería viable almacenar un conjunto lo suficientemente grande de patrones respuesta posibles y crear un programa que nos permita encontrar en una forma rápida cuál de ellos se encuentra a una distancia mínima de nuestro patrón objetivo  $\xi_i$ . Es claro que si ésta fuera la forma de solución al complicar las imágenes o patrones de entrada, agotaríamos cualquier cantidad de memoria almacenando patrones

posibles, y siempre existiría la posibilidad de encontrar uno que no se ajustara ni remotamente a alguno de los anteriores.

Luego entonces, necesitamos un sistema con la capacidad de autorganizarse e involucrar procesos iterativos, que sea autorrecursivo. Las redes neuronales son de hecho sistemas dinámicos. Es natural pensar que la mejor manera de realizar predicciones sobre la evolución de un sistema dinámico no-lineal es la utilización de otro sistema con estas mismas características.

En la siguiente figura se muestra el espacio de todos los posibles estados de nuestra red (espacio de configuración). Dentro de este espacio se encuentran estados que constituyen atractores,  $\xi_i^{\mu}$ . La dinámica del sistema arrastra cualquier estado inicial hacia alguno de los atractores; todo el espacio se encuentra dividido en cuencas de atracción.



De forma que nuestra red debe tener la capacidad de que a partir de una configuración inicial  $n_i = \xi_i$

determine el conjunto de funciones de peso  $w_{ij}$ , que coloque a la red en el estado  $n_i = \xi_i^{\mu_o}$  de forma que el patrón respuesta  $\mu_o$  efectivamente se encuentre a una distancia mínima del patrón objetivo  $\xi_i$ .

Por conveniencia realizamos una transformación donde los posibles valores de activación de las unidades son  $+1$  o  $-1$ , en lugar de 1 y 0. La ecuación (3.1) ahora se expresa como

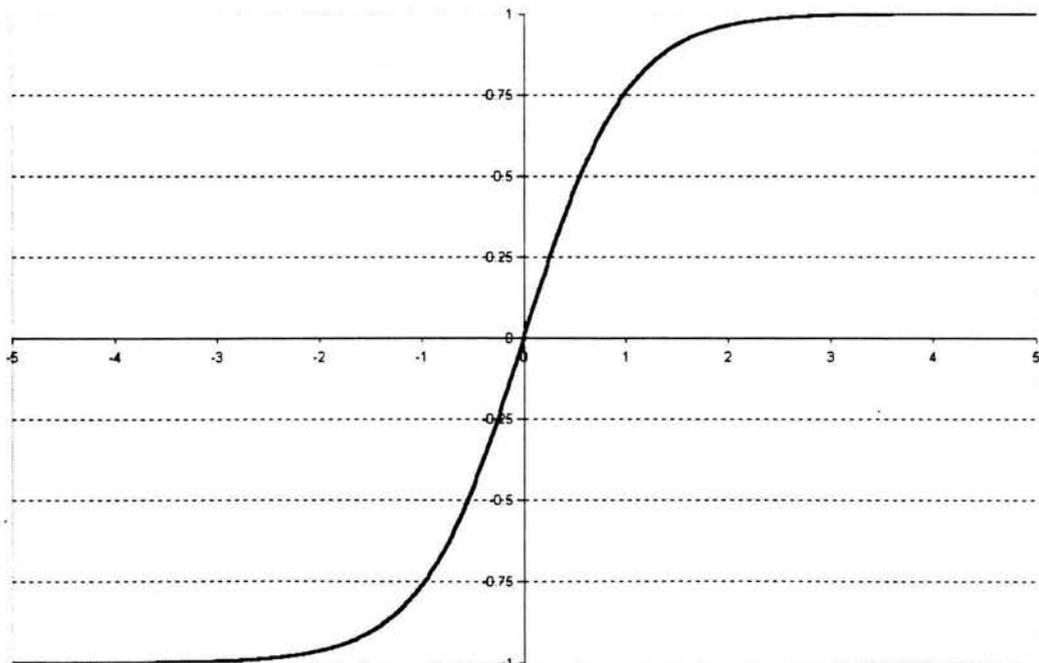
$$S_i = sgm\left(\sum_j w_{ij} S_j - \theta_i\right) \quad (3.2)$$

donde  $sgm$  corresponde a la función sigmoide:

$$sgm(x) = \tanh(\beta x) \quad (3.3)$$

$\beta$  es un parámetro de amplitud de la función sigmoide y puede valer 1 para valores de activación:  $+1$  o  $-1$ , o bien  $\frac{1}{2}$  para valores de activación: 0 o  $+1$ .

Fig. (3.3) FUNCION SIGMOIDE(X)



La condición para que el patrón  $\xi_i$  sea estable es

$$\text{sgm}\left(\sum_j w_{ij}\xi_j\right) = \xi_i \quad \forall i \quad (3.4)$$

Dado que  $\xi_j^2 = 1$ , si consideramos

$$w_{ij} \propto \xi_i \xi_j \quad (3.5)$$

e introducimos una constante de proporcionalidad  $\frac{1}{N}$ , donde  $N$  corresponde al número de unidades en la red, llegamos a

$$w_{ij} = \frac{1}{N} \xi_i \xi_j \quad (3.6)$$

Es claro que si partimos de un patrón de entrada con un número de bits 'erróneos' inferior a la mitad, éstos serán subvaluados al considerar la suma sobre todo el patrón

$$h_i = \sum_j w_{ij} S_j. \quad (3.7)$$

De forma que  $\text{sgm}(h_i)$  regrese  $\xi_i$ ; ésto significa que la red corregirá errores como se desea y podemos afirmar que  $\xi_i$  constituye un atractor. De hecho existen dos atractores en este caso:  $-\xi_i$  también constituye un atractor como se puede observar si invertimos el signo en el desarrollo anterior.

La forma natural de extender la capacidad de nuestra red para que funcione para más de un patrón es constituir  $w_{ij}$  como una superposición de términos, uno por cada patrón que se desee  $\xi_i^\mu$ :

$$w_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu. \quad (3.8)$$

$p$  es el número de patrones.

Una red de neuronas que opera bajo la regla anterior con unidades binarias se conoce como modelo de Hopfield [29]. Verifiquemos el comportamiento de la red. Para un patrón  $\xi_i^\mu$ , la condición de estabilidad se generaliza como

$$\text{sgn}(h_i^\nu) = \xi_i^\nu \quad \forall i \quad (3.9)$$

$h_i^\nu$  constituye la entrada a la unidad  $i$ . Para el patrón  $\nu$  está dada por

$$h_i^\nu \equiv \sum_j w_{ij} \xi_j^\nu = \frac{1}{N} \sum_j \sum_\mu \xi_i^\mu \xi_j^\mu \xi_j^\nu. \quad (3.10)$$

Reagrupando

$$h_i^\nu = \xi_i^\nu + \frac{1}{N} \sum_j \sum_{\mu \neq \nu} \xi_i^\mu \xi_j^\mu \xi_j^\nu. \quad (3.11)$$

Claramente, cuando el segundo término es 0 o de magnitud inferior a 1, no afecta el signo de  $h_i^\nu$  y el sistema funciona como se espera.

### 3.3 Límites de Confiabilidad

Es necesario determinar bajo qué condiciones nuestro sistema tendrá un funcionamiento confiable.

Consideremos el término  $C_i^\nu$ , expresado como

$$C_i^\nu = -\xi_i^\nu \frac{1}{N} \sum_j \sum_{\mu \neq \nu} \xi_i^\mu \xi_j^\mu \xi_j^\nu. \quad (3.12)$$

La expresión anterior depende solamente de los patrones  $\xi_j^\nu$  que queremos alcanzar. Al considerar patrones que sean aleatorios, con la misma probabilidad para  $\xi_j^\nu = 1$  o  $\xi_j^\nu = -1$  para cada  $j$  y  $\nu$ , entonces podemos expresar la probabilidad  $P_{error}$  de que algún elemento del patrón sea inestable

$$P_{error} = Prob(\xi_j^\nu \geq 1). \quad (3.13)$$

Para  $p$  patrones aleatorios y  $N$  unidades tenemos una distribución binomial con varianza  $\sigma^2 = \frac{p}{N}$ ; entonces

$$P_{error} = \frac{1}{\sqrt{2\pi\sigma}} \int_1^\infty \exp\left\{-\frac{x^2}{2\sigma^2}\right\} dx = \frac{1}{2} \left[1 - erf\left(\frac{1}{\sqrt{2\sigma^2}}\right)\right] = \frac{1}{2} \left[1 - erf\left(\sqrt{\frac{N}{2p}}\right)\right] \quad (3.14)$$

la función error  $erf(x)$  se define como

$$erf(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x \exp(-u^2) du \quad (3.15)$$

La tabla siguiente muestra los valores de  $\frac{p}{N}$  relacionados con los correspondientes valores de  $P_{error}$ . Determina que para  $p > 0.138N$  nuestro sistema pierde confiabilidad.

Tabla (3.1) Capacidad de la red. Valores para  $P_{error}$  para diferentes valores de  $p$ .

$P_{error}$	$P_{max}/N$
0.0010	0.105
0.0036	0.138
0.0100	0.185
0.0500	0.370
0.1000	0.610

### 3.4 Perceptrones y algoritmo de aprendizaje

Las redes neuronales denominadas perceptrones constan de un conjunto de terminales de entrada con la única función de alimentar al resto de la red. Después de esta capa pueden venir una o varias capas (capas ocultas), y al final se tiene una capa de salida.

Denominaremos a las entradas a la red como  $\xi_k$  y a las salidas como  $O_i$ . Su desempeño se describe mediante

$$O_i = g(h_i) = g\left(\sum_k w_{ik} \xi_k\right). \quad (3.16)$$

$g(h)$  es la llamada función de activación; usualmente se utiliza la función sigmoide. La tarea de asociación de la red consiste en que a partir de un patrón de entrada  $\xi_k^\mu$ , obtengamos un patrón deseado  $\zeta_i^\mu$ ; esto es, deseamos que el patrón de salida  $O_i^\mu$  sea igual al patrón deseado  $\zeta_i^\mu$ , para cada  $i$  y cada  $\mu$ ;

$$O_i^\mu = \zeta_i^\mu \quad (3.17)$$

En términos del patrón,

$$O_i^\mu = g(h_i^\mu) = g\left(\sum_k w_{ik} \xi_k^\mu\right) \quad (3.18)$$

Gráficamente, el elemento determinante consiste en buscar que coincidan los signos del patrón de entrada con la respuesta deseada  $\xi_k^\mu$ ; en esta perspectiva, los pesos  $w_{ik}$  constituyen un vector de pesos  $\vec{w} = (w_1, w_2, \dots, w_N)$ ; el patrón de entrada forma parte del mismo espacio.

El vector de pesos debe ser perpendicular al plano que divide los patrones de entrada positivos de los negativos, (*Fig.3.4(a)*). La ecuación (3.17) queda como

$$\text{sgn}(\vec{w} \cdot \vec{\xi}^\mu) = \zeta^\mu. \quad (3.19)$$

Estableciendo el vector de pesos  $\vec{w}$  de forma que la proyección de  $\xi^\mu$  sobre  $\vec{w}$  tenga el mismo signo que el patrón deseado  $\zeta^\mu$ ; definiendo  $[\xi_i^\mu \equiv \zeta_i^\mu \xi_i^\mu$  o  $\vec{x}^\mu \equiv \zeta_i^\mu \vec{\xi}_i^\mu]$ , ver (Fig.3.4(b)), la ecuación (3.18), se expresa como

$$\vec{w} \cdot \vec{\xi}^\mu \geq 0 \tag{3.20}$$

Para cada  $\mu$ .

Lo que nos indica que el vector  $\vec{w}$  debe alinearse en la dirección de los vectores  $\vec{x}$  que dependen del producto de pares de vectores [entrada - salida].

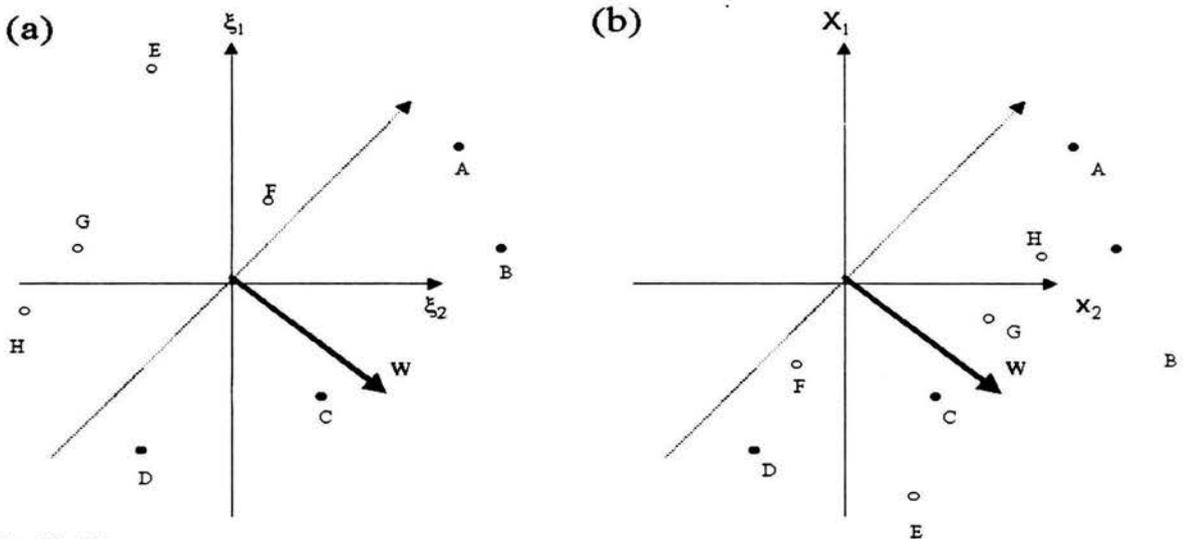


Fig.(3.4)

Representación gráfica de las condiciones (4.18) y (4.19), existen dos valores de entrada  $\xi_1$  y  $\xi_2$ , y ocho patrones ( $\mu = 1 \dots 8$ ) etiquetados con las letras A-H. Sólo se considera una salida. Los puntos sólidos representan patrones con  $\zeta^\mu = +1$ , en tanto que los puntos huecos  $\zeta^\mu = -1$ . En (a) los ejes coinciden con las entradas  $\xi_1$  y  $\xi_2$ , en (b) los ejes son  $X_1$  y  $X_2$ , con  $X_i^\mu = \zeta_i^\mu \xi_i^\mu$ . La condición de la respuesta correcta es que el plano perpendicular al vector de pesos  $w$  separa los puntos positivos de los negativos, como en (a) y a la vez incide en un sólo lado de los puntos "transformados" en (b).

¿Qué sucede si no existe este plano?, entonces el problema no tiene solución, de otra forma; el perceptrón simple sólo puede resolver problemas que sean linealmente separables, o superficies para las

que se pueda encontrar un plano tangencial a cada punto.

El mecanismo de aprendizaje consiste en introducir un factor de corrección;

$$w_{ik}^{nuevo} = w_{ik}^{ant} + \Delta w_{ik} \quad (3.21)$$

sí  $\zeta_i^\mu \neq O_i^\mu$

donde

$$\Delta w_{ik} = \begin{cases} 2\eta \zeta_i^\mu \xi_k^\mu & \text{sí } \zeta_i^\mu \neq O_i^\mu \\ 0 & \text{de otra manera} \end{cases}$$

o

$$\Delta w_{ik} = \eta(1 - \zeta_i^\mu O_i^\mu) \zeta_i^\mu \xi_k^\mu \quad (3.22)$$

$$\Delta w_{ik} = \eta(\zeta_i^\mu - O_i^\mu) \xi_k^\mu \quad (3.23)$$

sí en lugar de pedir que coincidan los signos de los parámetros de entrada  $h_i^\mu$ , requerimos que exista una cota mínima para la diferencia, tendremos

$$\zeta_i^\mu h_i^\mu \equiv \zeta_i^\mu \sum_k w_{ik} \xi_k^\mu \geq N_k. \quad (3.24)$$

Reagrupando

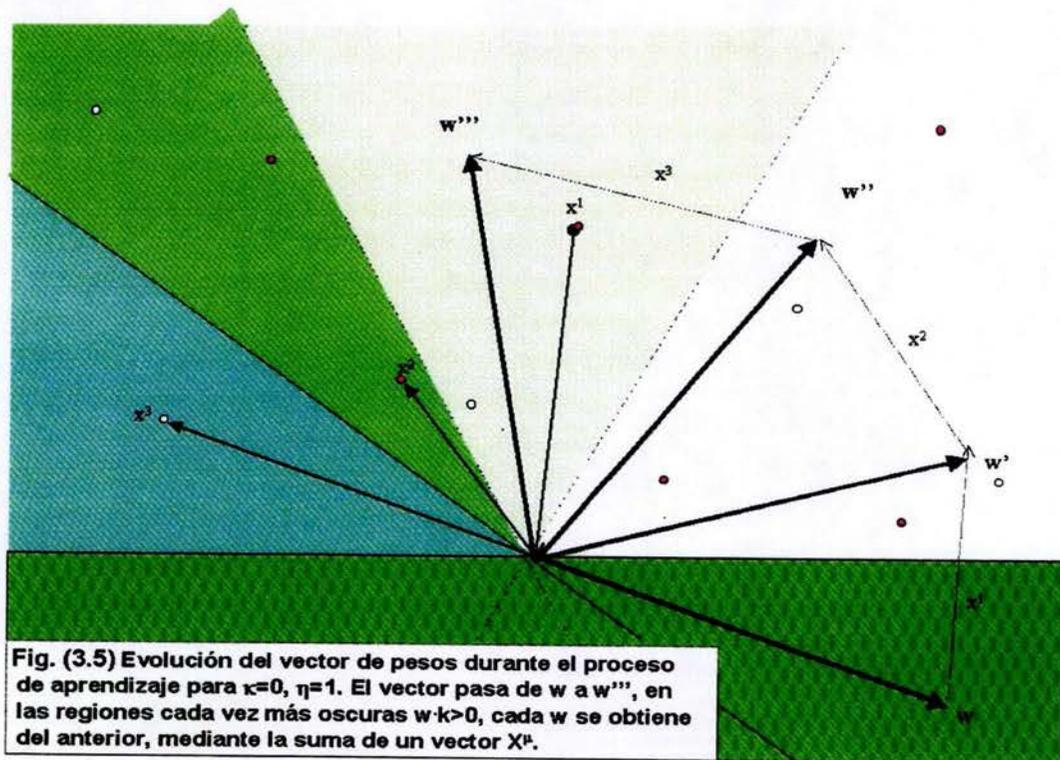
$$\Delta w_{ik} = \eta \Theta(N_k - \zeta_i^\mu h_i^\mu) \zeta_i^\mu x_{ik}^\mu. \quad (3.25)$$

Vectorialmente,

$$\Delta w_{ik} = \eta \Theta(N_k - \vec{w} \cdot \vec{\xi}^\mu) \vec{x}^\mu. \quad (3.26)$$

Esta expresión nos indica que para cada iteración, el vector  $\vec{w}$  sufre un cambio en la dirección  $\vec{x}^\mu$ , si la proyección  $\vec{w} \cdot \vec{x}^\mu$  es menor que  $\frac{N_k}{|\vec{w}|}$ .

La figura (3.5) muestra gráficamente el proceso de aprendizaje, la solución final se encuentra después de los pasos de corrección:  $w \rightarrow w' \rightarrow w'' \rightarrow w'''$ .



Dependiendo de los vectores  $\vec{x}^\mu$ , puede presentarse una región amplia de direcciones correctas o un cono estrecho, o ninguna solución. El valor

$$D(\vec{w}) = \frac{1}{|\vec{w}|} \min_{\mu} (\vec{w} \cdot \vec{x}^{\mu}). \quad (3.27)$$

Que corresponde a la 'peor' de las proyecciones, representa la distancia de la 'peor'  $\vec{x}^{\mu}$  al plano perpendicular a  $\vec{w}$ ; si  $D(\vec{w}) \leq 0$ , todos los puntos del patrón están situados correctamente y se puede encontrar una solución para  $|\vec{w}|$ .

### 3.5 Prueba de convergencia para el algoritmo de aprendizaje

Para cada iteración se procesa un patrón y los pesos son ajustados únicamente si no se satisface la condición (4.7). Para que se dé la convergencia, basta con mostrar que el número de pasos es finito. Sea  $M^{\mu}$  el número de correcciones realizadas en algún paso del proceso.

$$\vec{w} = \eta \sum_{\mu} M^{\mu} \vec{x}^{\mu}. \quad (3.28)$$

La idea es lograr acotar superiormente a  $|\vec{w}|$  y a la proyección  $\vec{w} \cdot \vec{w}^*$ , donde  $\vec{w}^*$  es un vector elegido en la forma adecuada; esto es,  $D(\vec{w}^*) \geq 0$

$$\vec{w} \cdot \vec{w}^* = \eta \sum_{\mu} M^{\mu} \vec{x}^{\mu} \cdot \vec{w}^* \leq \eta M \min_{\mu} \vec{x}^{\mu} \cdot \vec{w}^* = \eta M D(\vec{w}^*) |\vec{w}^*|. \quad (3.29)$$

Entonces  $\vec{w} \cdot \vec{w}^*$  es proporcional a  $M$ . Por otro lado para  $|\vec{w}|$ , calculamos el cambio en  $\vec{w}$  para una corrección por un patrón  $\alpha$ :

$$\Delta |\vec{w}|^2 = (\vec{w} + \eta \vec{x}^\alpha)^2 - \vec{w}^2 = \eta^2 (\vec{x}^\alpha)^2 + 2\mu \vec{w} \cdot \vec{x}^\alpha \leq \eta^2 N + 2\eta Nk = N\eta(\eta + 2k). \quad (3.30)$$

Directamente de  $Nk \leq \vec{w} \cdot \vec{x}^\alpha$  y  $x_k^\alpha = \pm 1$ , considerando los  $M$  incrementos, tenemos

$$|\vec{w}|^2 \leq MN\eta(\eta + 2k). \quad (3.31)$$

Finalmente, si sustituimos en

$$\phi = \frac{(\vec{w} \cdot \vec{w}^*)^2}{|\vec{w}|^2 |\vec{w}^*|^2} \quad (3.32)$$

que corresponde al coseno cuadrado del ángulo que forman  $\vec{w}$  y  $\vec{w}^*$ , los valores de las ecuaciones anteriores, encontramos

$$1 \geq \phi \geq \frac{MD(\vec{w}^*)^2 \eta}{N(\eta + 2k)}. \quad (3.33)$$

De donde  $M$  está acotado por

$$M \leq \frac{N(1 + \frac{2k}{\eta})}{D_{max}^2}. \quad (3.34)$$

Debemos notar que la cota es proporcional al número de unidades de entrada  $N$ .

### 3.6 Aprendizaje de Gradiente Descendiente

Definimos una medida de error o función de costo:

$$E[\vec{w}] = \frac{1}{2} \sum_{i\mu} (\zeta_i^\mu - O_i^\mu)^2 = \frac{1}{2} \sum_{i\mu} (\zeta_i^\mu - \sum_k w_{ik} \xi_k^\mu)^2. \quad (3.35)$$

Esta ecuación sólo depende de los pesos  $w_{ik}$  y los patrones involucrados. Cada  $w_{ik}$  se modifica en una cantidad  $\Delta w_{ik}$  proporcional al gradiente  $E$ :

$$\Delta w_{ik} = -\eta \frac{\partial E}{\partial w_{ik}} = \eta \sum_{\mu} (\zeta_i^\mu - O_i^\mu) x_i^\mu. \quad (3.36)$$

Para cada patrón de entrada se tiene

$$\Delta w_{ik} = \eta (\zeta_i^\mu - O_i^\mu) x_i^\mu = \eta \delta_i^\mu \xi_k^\mu. \quad (3.37)$$

Donde

$$\delta_i^\mu = \zeta_i^\mu - O_i^\mu. \quad (3.38)$$

Este resultado se conoce como regla de Widrow-Hoff o regla LMS (least mean square).

La función de costo (3.35), es una expresión cuadrática para los diferentes vectores de peso, en el subespacio de los patrones forma una parábola, cuando los vectores son línealmente independientes, y existe una solución, el mínimo de la función de costo ocurre cuando  $E = 0$ .

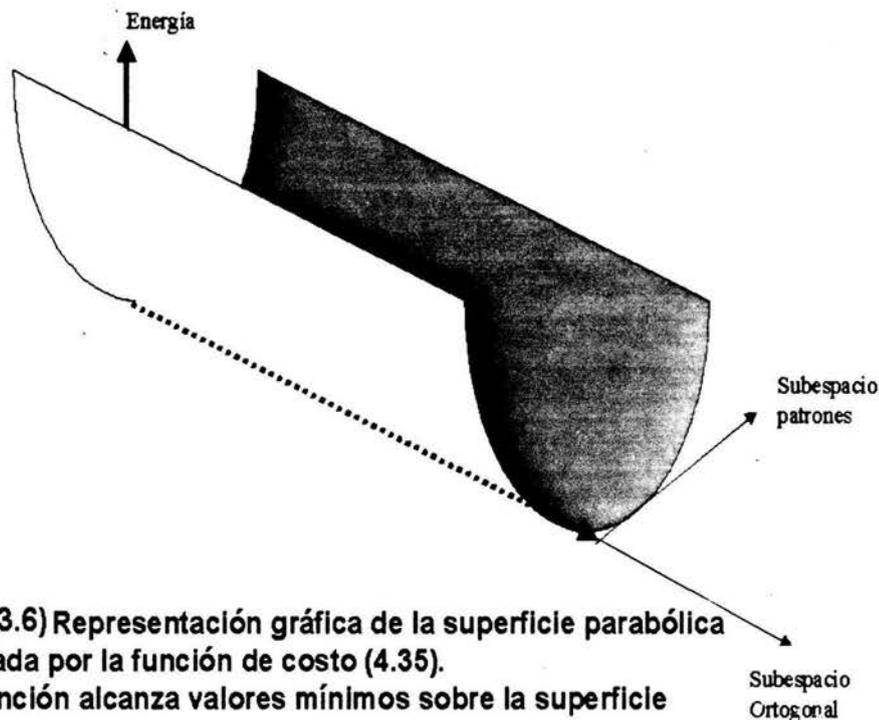


Fig. (3.6) Representación gráfica de la superficie parabólica formada por la función de costo (4.35). La función alcanza valores mínimos sobre la superficie ortogonal a los patrones  $\xi^\mu$ .

### 3.7 Retropropagación

El algoritmo de retropropagación es esencial para comprender el proceso de aprendizaje de las redes neuronales; proporciona un método para variar los pesos  $w_{pq}$  en cualquier red a partir de un conjunto de patrones de entrada salida  $\{\xi_k^\mu, \zeta_k^\mu\}$ .

Consideremos la red mostrada en la figura siguiente:

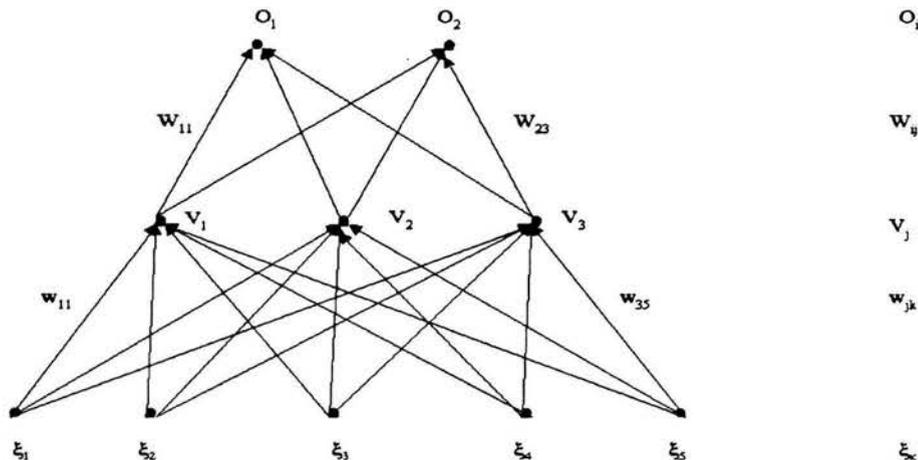


Fig. (3.7) Red de retropropagación de dos capas

Dado un patrón  $\mu$ , las unidades ocultas  $j$  reciben

$$h_j^\mu = \sum_k w_{jk} \xi_k^\mu, \tag{3.39}$$

y producen una salida

$$V_j^\mu = g(h_j^\mu) = g\left(\sum_k w_{jk} \xi_k^\mu\right). \tag{3.40}$$

La unidad de salida  $i$  recibe

$$h_i^\mu = \sum_j W_{ij} V_j^\mu = \sum_j W_{ij} g\left(\sum_k w_{jk} \xi_k^\mu\right), \quad (3.41)$$

y produce como respuesta de la red

$$O_i^\mu = g(h_i^\mu) = g\left(\sum_j W_{ij} V_j^\mu\right) = g\left(\sum_j W_{ij} g\left(\sum_k w_{jk} \xi_k^\mu\right)\right). \quad (3.42)$$

La expresión para la función de error queda

$$E[\vec{w}] = \frac{1}{2} \sum_{\mu i} [\zeta_i^\mu - O_i^\mu]^2 = \frac{1}{2} \sum_{\mu i} [\zeta_i^\mu - g\left(\sum_j W_{ij} g\left(\sum_k w_{jk} \xi_k^\mu\right)\right)]^2. \quad (3.43)$$

Para las conexiones de la capa oculta a la salida, la regla del gradiente descendiente nos dá

$$\Delta W_{ij} = -\eta \frac{\partial E}{\partial W_{ij}} = \eta \sum_{\mu} [\zeta_i^\mu - O_i^\mu] \dot{g}(h_i^\mu) V_j^\mu = \eta \sum_{\mu} \delta_i^\mu V_j^\mu. \quad (3.44)$$

Donde

$$\delta_i^\mu = \dot{g}(h_i^\mu) [\zeta_i^\mu - O_i^\mu]. \quad (3.45)$$

Este resultado es totalmente análogo al encontrado con anterioridad (3.35) y (3.36). Para las conexiones entre la capa de entrada y las capas ocultas  $\Delta w_{jk}$  debemos diferenciar con respecto a  $w_{jk}$ , mediante la regla de la cadena

$$\Delta w_{jk} = -\eta \frac{\partial E}{\partial w_{jk}} = \eta \sum_{\mu} \frac{\partial E}{\partial V_j^\mu} \frac{\partial V_j^\mu}{\partial w_{jk}} = \eta \sum_{\mu i} [\zeta_i^\mu - O_i^\mu] \dot{g}(h_i^\mu) W_{ij} \dot{g}(h_j^\mu) \xi_k^\mu = \eta \sum_{\mu i} \delta_i^\mu W_{ij} \dot{g}(h_j^\mu) \xi_k^\mu = \eta \sum_{\mu} \delta_j^\mu \xi_k^\mu. \quad (3.46)$$

Con

$$\delta_j^\mu = \dot{g}(h_j^\mu) \sum_i W_{ij} \delta_i^\mu. \quad (3.47)$$

Este resultado se puede extender para un número arbitrario de capas. Por inducción, siendo  $m$  el número de capas ocultas, para  $m=1$  tenemos la ecuación (3.44). Consideramos la estructura siguiente con  $M$  capas ocultas:

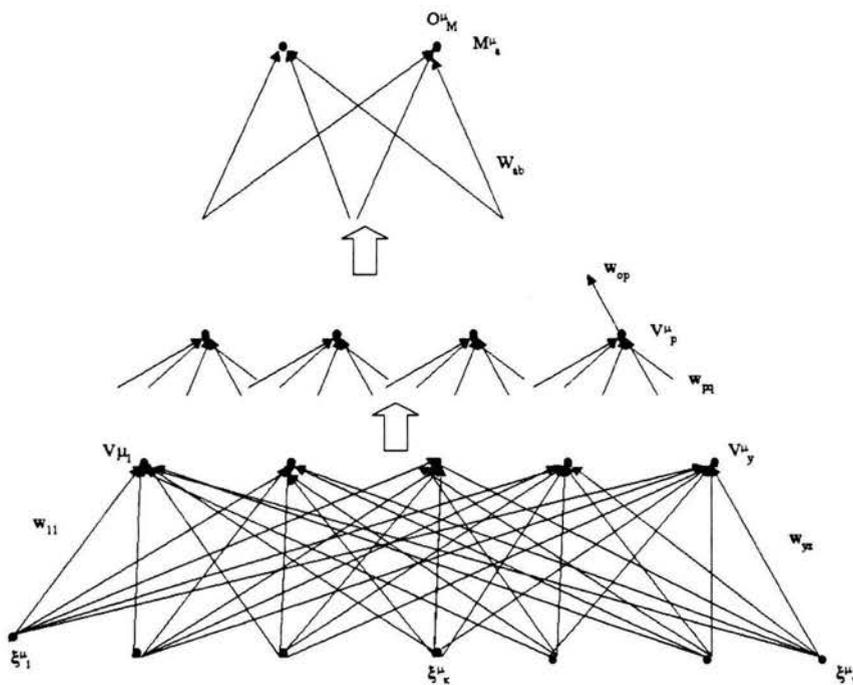


Fig. (3.8) Red de retropropagación con  $M$  capas ocultas

Dado el patrón de entrada  $\xi_z^\mu$ , la primera capa oculta (índice  $y$ ) recibe

$$h_y^\mu = \sum_z w_{yz} \xi_z^\mu, \quad (3.48)$$

y produce como salida

$$V_y^\mu = g(h_y^\mu) = g\left(\sum_z w_{yz} \xi_z^\mu\right). \quad (3.49)$$

La capa p-ésima recibe

$$h_p^\mu = \sum_q w_{pq} V_p^\mu, \quad (3.50)$$

y produce como salida

$$V_p^\mu = g(h_p^\mu) = g\left(\sum_q w_{pq} V_p^\mu\right). \quad (3.51)$$

La capa M (salida) recibe

$$h_M^\mu = \sum_b w_{ab} V_b^\mu = \sum_b w_{ab} g\left(\sum_c w_{bc} V_c^\mu\right), \quad (3.52)$$

y produce la salida final

$$O_M^\mu = g(h_M^\mu) = g\left(\sum_b W_{ab} V_b^\mu\right) = g\left(\sum_b W_{ab} g\left(\sum_c w_{bc} V_c^\mu\right)\right). \quad (3.53)$$

Para la capa M

$$E[\vec{w}] = \frac{1}{2} \sum_{\mu a} [\zeta_a^\mu - O_a^\mu]^2 = \frac{1}{2} \sum_{\mu a} [\zeta_a^\mu - g\left(\sum_b W_{ab} g\left(\sum_c w_{bc} V_c^\mu\right)\right)]^2. \quad (3.54)$$

$$\Delta W_{ab} = -\eta \frac{\partial E}{\partial W_{ab}} = \eta \sum_{\mu} [\zeta_M^\mu - O_M^\mu] \dot{g}(h_M^\mu) V_b^\mu = \eta \sum_{\mu} \delta_M^\mu V_b^\mu. \quad (3.55)$$

donde

$$\delta_M^\mu = \dot{g}(h_M^\mu) [\zeta_M^\mu - O_M^\mu]. \quad (3.56)$$

$$\Delta W_{ab} = \eta \sum_{\mu} \delta_M^\mu \cdot V_b^\mu. \quad (3.57)$$

para la capa M-1; ahora descenderemos en la estructura de la red, para la capa m-1:

$$\Delta w_{bc} = -\eta \frac{\partial E}{\partial w_{bc}} = \eta \sum_{\mu a} [\zeta_a^\mu - g(h_M^\mu)] \dot{g}(h_M^\mu) W_{ab}^\mu V_c^\mu = \eta \sum_{\mu a} \delta_M^\mu \dot{g}(h_M^\mu) V_b^\mu. \quad (3.58)$$

$$\delta_{M-1}^\mu = \dot{g}(h_M^\mu \sum_a W_{ab} \delta_M^\mu). \quad (3.59)$$

$$\Delta w_{bc} = \eta \sum_\mu \delta_{M-1}^\mu V_b^\mu. \quad (3.60)$$

lo suponemos cierto para la capa  $p$ , esto es

$$\Delta w_{op} = \eta \sum_\mu \delta_{M-p}^\mu V_q^\mu. \quad (3.61)$$

para la siguiente capa  $q$ , tenemos

$$\Delta w_{pq} = -\eta \frac{\partial E}{\partial V_q^\mu} \frac{\partial V_q^\mu}{\partial w_{pq}^\mu}. \quad (3.62)$$

partiendo de la expresión de la función de costo (3.43) y aplicando reiteradamente la regla de la cadena

$$E[\vec{w}] = \frac{1}{2} \sum_{\mu a} [\zeta_a^\mu - g(\sum_b w_{ab} g(\sum_c w_{bc} g(\sum_d w_{cd} \dots g(\sum_q w_{pq} V_q^\mu) \dots)))]^2. \quad (3.63)$$

$$\Delta w_{pq} = \eta \sum_{\mu a, b, c, \dots, p} [\zeta_a^\mu - g(h_M^\mu)] \dot{g}(h_M^\mu) W_{ab}^\mu \dot{g}(h_{M-1}^\mu) W_{bc}^\mu \dot{g}(h_{M-2}^\mu)_{cd}, \dots, \dot{g}(h_p^\mu) W_{op}^\mu \frac{\partial V_q^\mu}{\partial w_{pq}^\mu}. \quad (3.64)$$

introduciendo la igualdad adoptada para la capa  $p$ , (3.61):

$$\Delta w_{pq} = \eta \sum_\mu \delta_{M-p}^\mu \frac{\partial V_q^\mu}{\partial w_{pq}^\mu} = \dot{g}(h_q^\mu) \sum_q w_{pq} V_r^\mu = \eta \sum_\mu \delta_{M-q}^\mu V_r^\mu. \quad (3.65)$$

generalizando:  $\implies$

$$\Delta w_{pq} = \eta \sum_\mu \delta_{salida} \times V_{entrada} \quad (3.66)$$

donde *entrada* y *salida* se refieren a las conexiones involucradas, la forma de  $\delta$  depende de la capa involucrada.

Resumiendo el procedimiento de retropropagación consiste en:

▷ Inicializar los pesos en forma aleatoria.

▷ ▷ Proporcionar un patrón de entrada  $\xi_k^\mu$

$$V_k^0 = \xi_k^\mu \quad (3.67)$$

∀  $k$

▷ ▷ ▷ Propagar la señal ascendentemente a través de la red utilizando:

$$V_i^m = g(h_i^m) = g\left(\sum_j w_{ij}^m V_j^{m-1}\right). \quad (3.68)$$

▷ ▷ ▷ ▷ Calcular las deltas para la capa de salida, comparando con el patrón de respuesta deseado  $\zeta_i^\mu$  las salidas actuales  $V_i^M$

$$\delta_i^M = \dot{g}(h_i^M)[\zeta_i^\mu - V_i^M]. \quad (3.69)$$

▷ ▷ ▷ ▷ ▷ Calcular las deltas para las capas anteriores, propagando los errores en forma descendente

$$\delta_i^{m-1} = \dot{g}(h_i^{m-1}) \sum_j w_{ji}^m \delta_j^m. \quad (3.70)$$

para  $m = M, M - 1, \dots, 2$ .

▷ ▷ ▷ ▷ ▷ ▷ Corregir el valor de peso para cada conexión

$$\Delta w_{ij}^m = \eta \delta_i^m V_j^{m-1}. \quad (3.71)$$

utilizando el algoritmo de gradiente descendente

$$w_{ij}^{\text{nuevo}} = w_{ij}^{\text{anterior}} + \Delta w_{ij}. \quad (3.72)$$

▷ ▷ ▷ ▷ ▷ ▷ ▷ Regresar al paso ▷ ▷ considerando el siguiente patrón de entrada.

# Capítulo 4

## REDES DE RESPUESTA DE IMPULSO FINITO (FIR)

### 4.1 Representación Espacio Temporal

Para el caso de las redes neuronales que hemos analizado, la respuesta de la red es función de una suma de valores de peso, proporcionando una representación 'espacial' de las señales que se reciben de diferentes neuronas. Al sustituir las neuronas por filtros de respuesta que operativamente almacenan memoria de la evolución del sistema en un intervalo de tiempo, (nominalmente, sustituimos escalares por vectores y la multiplicaciones por productos vectoriales), introducimos una dimensión 'temporal'.

Para cada filtro tenemos:

$$s(k) = \sum_i \sum_{n=0}^M w_i(n) x_i(k-n) + w_b = \sum_i \vec{w}_i \cdot \vec{x}_i(k) + w_b. \quad (4.1)$$

Donde  $k$  es un índice que recorre el tiempo,  $w_i(n)$  son los coeficientes del filtro para cada tiempo,

representados en forma vectorial como

$$\vec{w}_i = [w_i(0), w_i(1), \dots, w_i(n)]. \quad (4.2)$$

y el vector de retraso corresponde a

$$\vec{x}_k = [x_i(k), x_i(k-1), \dots, x_i(k-M)]. \quad (4.3)$$

La función de operación se representa por el producto vectorial  $\vec{w}_i \cdot \vec{x}_i(k)$

Analógicamente al caso anterior la función de respuesta de potencial del filtro se representa como

$$y(k) = f(s(k)). \quad (4.4)$$

donde  $f() = \tanh()$ , y solo hemos añadido el índice  $k$ .

## 4.2 Equivalencia entre las arquitecturas de las redes

Una red de retropropagación corresponde a un caso particular de una red FIR cuando la longitud del vector de retraso  $M$  es igual a 0. No obstante, con la interpretación adecuada, se puede mostrar que una red FIR es un caso particular de una red de retropropagación.

La estrategia consiste en desarrollar la red FIR en una estructura equivalente de un tamaño mayor.

La red consiste de tres capas con un solo nodo de salida, dos nodos en cada capa oculta y dos nodos de entrada. Todos los filtros tienen un vector de tres elementos (sinapsis de segundo orden); ésto es, cuando aparentemente existen sólo 10 conexiones en la red, considerando los vectores temporales tenemos en realidad 30 conexiones.

Para expandir la red, cada nodo empezando por la última capa, se sustituye por la réplica del vector de retraso, el número de veces que sea necesario, ver figura (4.1).

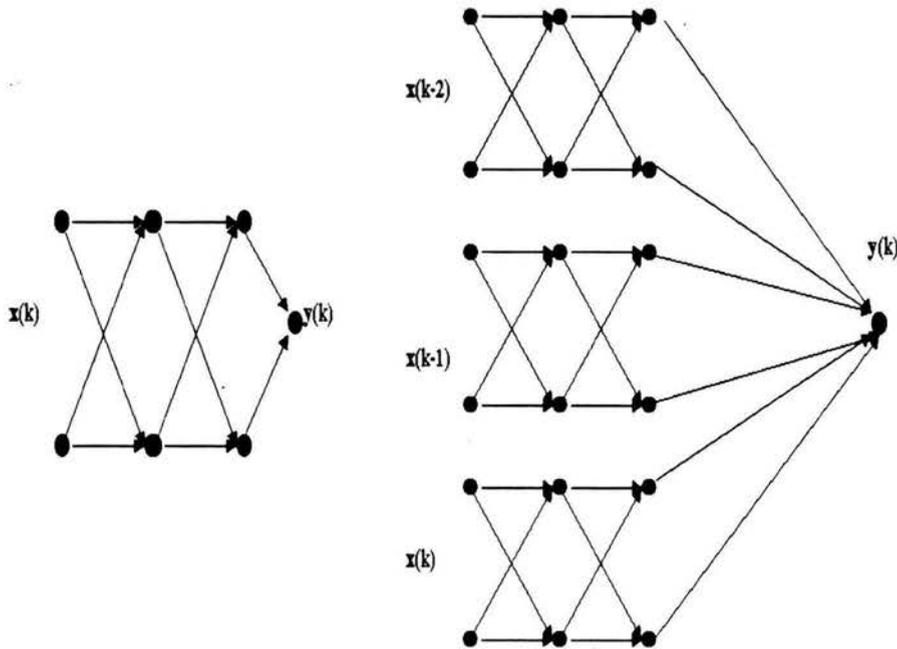
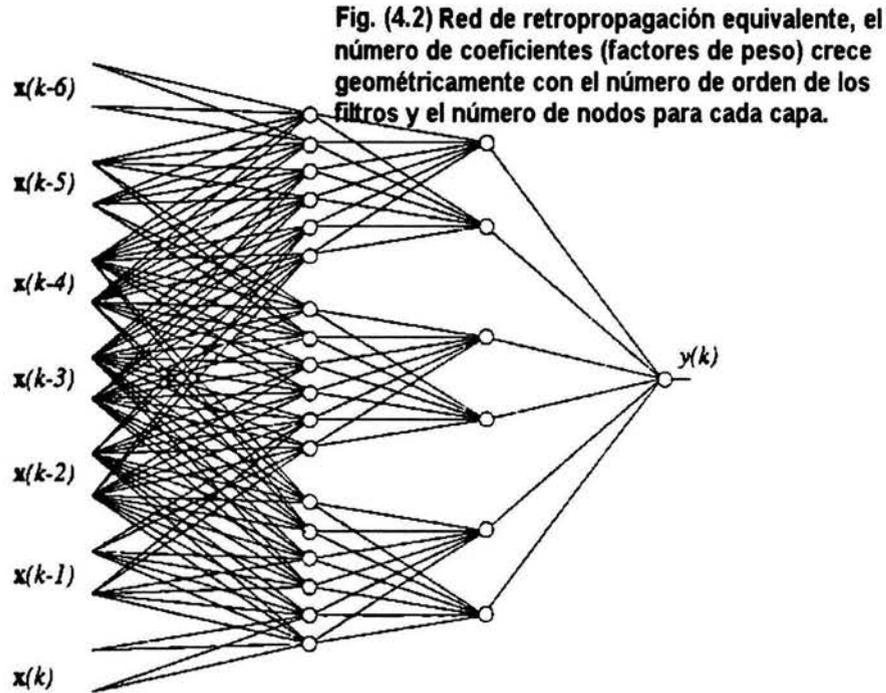


Fig. (4.1) Una red FIR con 3 capas: (2 nodos en la primera, 2 nodos en la segunda y un nodo final como salida) y conexiones de segundo orden (3 retrasos), se expande en una estructura estática equivalente, cada coeficiente es sustituido por el número de nodos correspondiente (retrasos).

Continuando este proceso hasta eliminar todos los retrasos arribamos a la siguiente estructura: (*figura 4.2*)



Notar que donde antes teníamos 30 coeficientes con componentes vectoriales de tiempo, ahora tenemos 150 coeficientes escalares.

### 4.3 Retropropagación en el tiempo. Funcionalidad de las redes FIR

Trasladando la ecuación (3.35), y considerando que tenemos  $k$  términos para cada nodo:

$$E[\vec{w}] = \sum_k \left[ \frac{1}{2} \sum_{\mu i} [\zeta_i(k)^\mu - O_i(k)^\mu]^2 \right]. \tag{4.5}$$

Análogamente, para obtener el gradiente descendiente (3.44)

$$\Delta W_{ij} = -\eta \frac{\partial E_k(\vec{w})}{\partial W_{ij}^l}. \tag{4.6}$$

Denominando a  $s_j^l$  como la función de acumulación de las sinapsis que contribuyen al nodo  $j$  de la capa  $l$ :

$$\frac{\partial E_k(\vec{w})}{\partial W_{ij}^l} = \sum_k \frac{\partial E_k(\vec{w})}{\partial s_j^l(k)} \frac{\partial s_j^l(k)}{\partial W_{ij}^l}. \quad (4.7)$$

De esta manera

$$\Delta W_{ij}^l = -\eta \frac{\partial E_k(\vec{w})}{\partial s_j^l(k)} \frac{\partial s_j^l(k)}{\partial W_{ij}^l}. \quad (4.8)$$

Definimos  $s_j^l(k)$  como

$$s_j^l(k) = \sum_i \vec{W}_{ij}^l \cdot \vec{a}_i^{l-1}(k) + w_b^l. \quad (4.9)$$

Donde  $w_b^l$  corresponde al valor de activación de la red en su estado de reposo. Por lo tanto

$$\frac{\partial s_j^l(k)}{\partial W_{ij}^l} = \frac{\partial \vec{W}_{ij}^l \cdot \vec{a}_i^{l-1}(k)}{\partial W_{ij}^l} = \vec{a}_i^{l-1}(k). \quad (4.10)$$

Definiendo

$$\frac{\partial E_k(\vec{w})}{\partial s_j^l(k)} \cong \delta_j^l. \quad (4.11)$$

La ecuación (3.44), se reescribe como:

$$\vec{W}_{ij}^l(k+1) = \vec{W}_{ij}^l(k) - \mu \delta_j^l(k) \cdot \vec{a}_i^{l-1}(k). \quad (4.12)$$

Buscaremos una fórmula recursiva para  $\delta_j^l(k)$ . Para la capa de salida la función  $s_j^L(k)$  tiene solo un término que corresponde al error del nodo de salida  $e_j(k)$ .

$$\delta_j^L(k) \cong \frac{\partial \sum_k \vec{e}^T(k) \vec{e}(k)}{\partial s_j^L(k)} = \frac{\partial e_j^L(k)}{\partial a_j^L} \frac{\partial a_j^L}{\partial s_j^L(k)}. \quad (4.13)$$

$$\delta_j^L(k) = -2e_j(k) f'(s_j^L(k)). \quad (4.14)$$

Para una capa oculta, debido a los retrasos,  $s_j^l(k)$  afecta al error a través del tiempo, aplicando la regla de la cadena:

$$\delta_j^l(k) = \sum_m \sum_t \frac{\partial E_k(\vec{w})}{\partial s_j^{l+1}(k)} \frac{s_m^{l+1}(k)}{s_j^l(k)}. \quad (4.15)$$

(Notar que el índice  $m$  se refiere a la siguiente capa). Por (4.11)

$$\frac{\partial E_k(\vec{w})}{\partial s_j^{l+1}(t)} = \delta_m^{l+1}(t). \quad (4.16)$$

Para el segundo término

$$\frac{\partial(s_m^{l+1}(k))}{\partial(s_j^l(k))} = \frac{\partial(s_m^{l+1}(k))}{\partial(a_j^l(k))} \partial(a_j^l(k)) \partial(s_j^l(k)). \quad (4.17)$$

$$\frac{\partial(s_m^{l+1}(k))}{\partial(s_j^l(k))} = \frac{\partial(s_m^{l+1}(k))}{\partial(a_j^l(k))} f'(s_j^l(k)). \quad (4.18)$$

Reescribiendo (4.9) en términos de los retrasos, y notando que  $a_j^l(k)$  únicamente contribuye en  $s_m^{l+1}(k)$  por la sinapsis que une la unidad  $j$  en la capa  $l$  con la unidad  $m$  en la capa  $l+1$ :

$$s_{jm}^{l+1}(t) = \sum_{p=0}^{M^{l+1}} w_{jm}^{l+1}(p) a_j^l(t-p). \quad (4.19)$$

Entonces

$$\frac{\partial(s_m^{l+1}(k))}{\partial(a_j^l(k))} = w_{jm}^{l+1}(p). \quad (4.20)$$

Para  $t-p=k$ .

$$\frac{\partial(s_m^{l+1}(k))}{\partial(a_j^l(k))} = \begin{cases} w_{jm}^{l+1}(t-k) & \text{si } 0 \leq (t-k) \leq M^{l+1} \\ 0 & \text{de otra manera} \end{cases}$$

Regresando a (4.12),

$$\delta_j^l(k) = f'(s_j^l(k)) \sum_m \sum_{t=k}^{M^{l+1}+k} \delta_j^l(t) w_{jm}^{l+1}(t-k) = f'(s_j^l(k)) \sum_m \sum_{n=0}^{M^{l+1}} \delta_m^{l+1}(k+n) w_{jm}^{l+1}(n) \quad (4.21)$$

$$\delta_j^l(\mathbf{k}) = f'(s_j^l(\mathbf{k})) \sum_m \bar{\delta}_m^{l+1}(\mathbf{k}) \cdot \bar{w}_{jm}^{l+1}. \quad (4.22)$$

donde hemos definido:

$$\bar{\delta}_m^{l+1}(\mathbf{k}) = [\delta_m^l(\mathbf{k}), \delta_m^l(\mathbf{k} + 1), \dots, \delta_m^l(\mathbf{k} + M^l)]. \quad (4.23)$$

El algoritmo de retropropagación temporal se resume en los siguientes pasos:

$$\Delta \bar{w}_{ij}^l(\mathbf{k}) = -\mu \delta_j^{l+1}(\mathbf{k}) \cdot \bar{a}_i^l(\mathbf{k}). \quad (4.24)$$

$$\delta_j^l(\mathbf{k}) = \begin{cases} -2e_j(\mathbf{k}) f'(s_j^L(\mathbf{k})) & \text{sí } l = L \\ f'(s_j^l(\mathbf{k})) \sum_m \bar{\delta}_m^{l+1}(\mathbf{k}) \cdot \bar{w}_{jm}^{l+1} & 1 \leq l \leq L - 1 \end{cases}$$

Notar que si reemplazamos los vectores  $\bar{\mathbf{a}}$ ,  $\bar{\mathbf{w}}$  y  $\bar{\delta}$ , por escalares obtenemos el algoritmo de retropropagación para redes estáticas, al introducir la relación temporal, las  $\delta$  no se obtienen simplemente sumando pesos, sino mediante retroalimentación de los filtros. Para cada nueva entrada,  $\bar{\mathbf{x}}$  y su correspondiente respuesta deseada  $\bar{\mathbf{d}}(\mathbf{x})$ , los filtros de las capas anteriores se incrementan un paso de tiempo y producen la respuesta  $\bar{\mathbf{y}}(\mathbf{k})$  con su correspondiente error  $e(\mathbf{k})$ , para cada uno de los retrasos ( $M$ ).

# Capítulo 5

## EVALUANDO PROCEDIMIENTOS

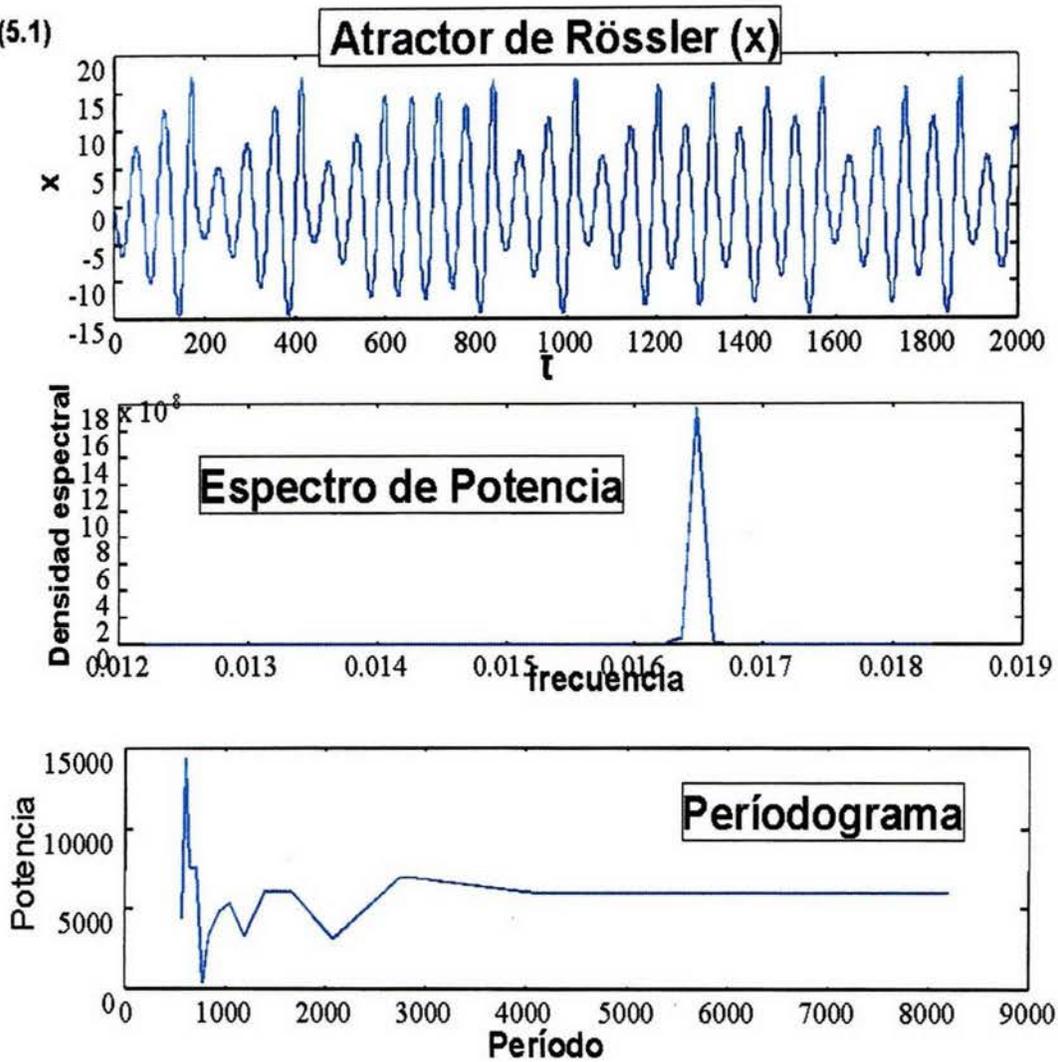
### 5.1 Caos y Ruido

#### 5.1.1 Espectro de Potencias

La primer tarea a la que nos enfrentamos es identificar sí la serie de tiempo que deseamos analizar es efectivamente caótica, para ello utilizamos en principio el espectro de potencias, de acuerdo con [43]; el espectro de potencias permite distinguir entre regímenes periódicos, semiperiódicos y caóticos, el espectro de potencias para una serie periódica o semiperiódica, muestra algunos picos aislados, (funciones delta).

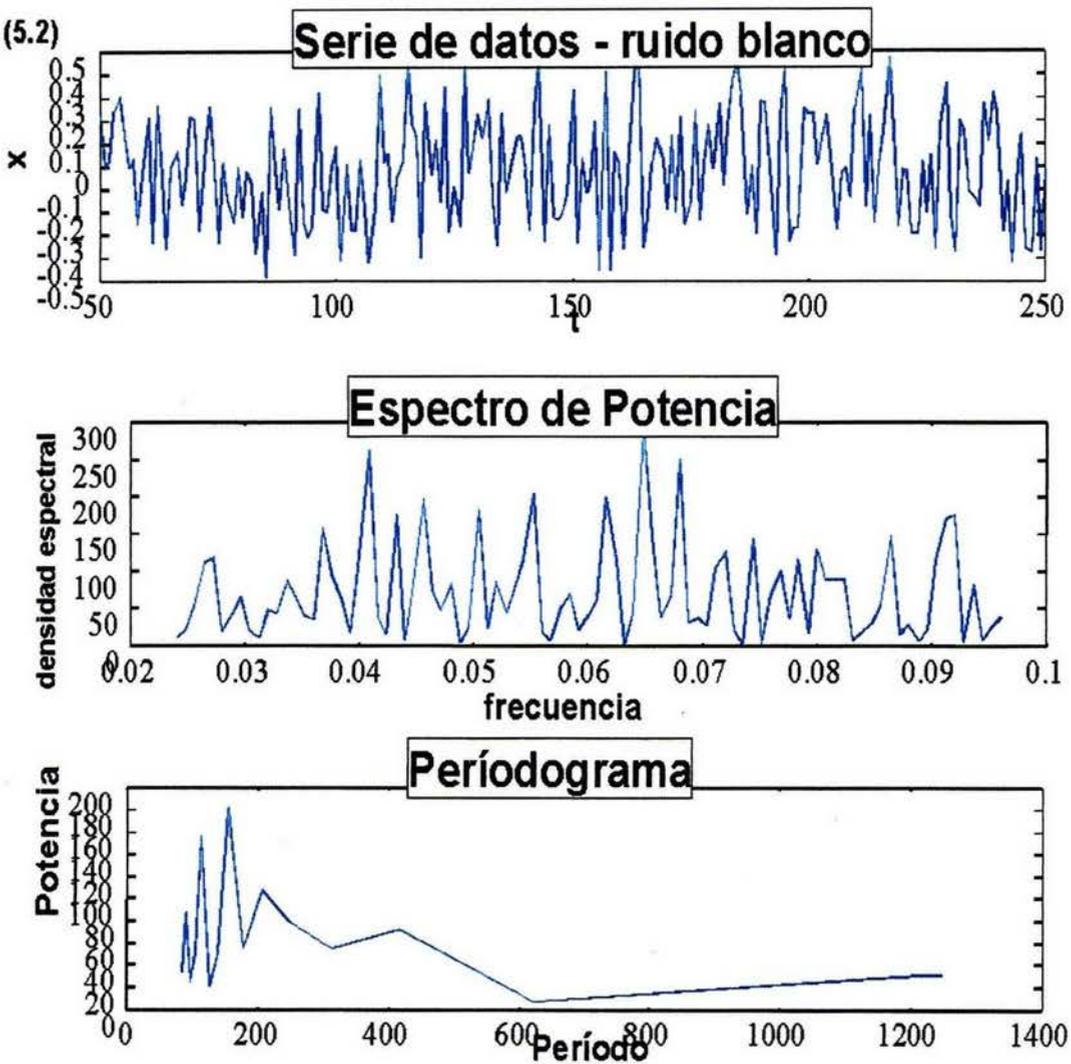
La serie de datos del atractor de Rössler, muestra un máximo de densidad espectral en la frecuencia  $f_c = 0.0167$  que corresponde a un período  $p = \frac{\pi}{f_c} \approx 190.6$  ciclos, cabe señalar que Fraser y Swinney [20], reportan un valor de  $p = 193.3$  ciclos, que fué reproducido en el presente trabajo, (sección 5.2.1).

Fig. (5.1)



La no-periodicidad se caracteriza por un espectro de banda ancha, según [41], en el caso del ruido se caracteriza por funciones delta distribuidas en todas la frecuencias, (espectro de banda ancha). Algunos atractores tienen un espectro de potencias que es una superposición de un espectro de banda ancha y funciones delta.

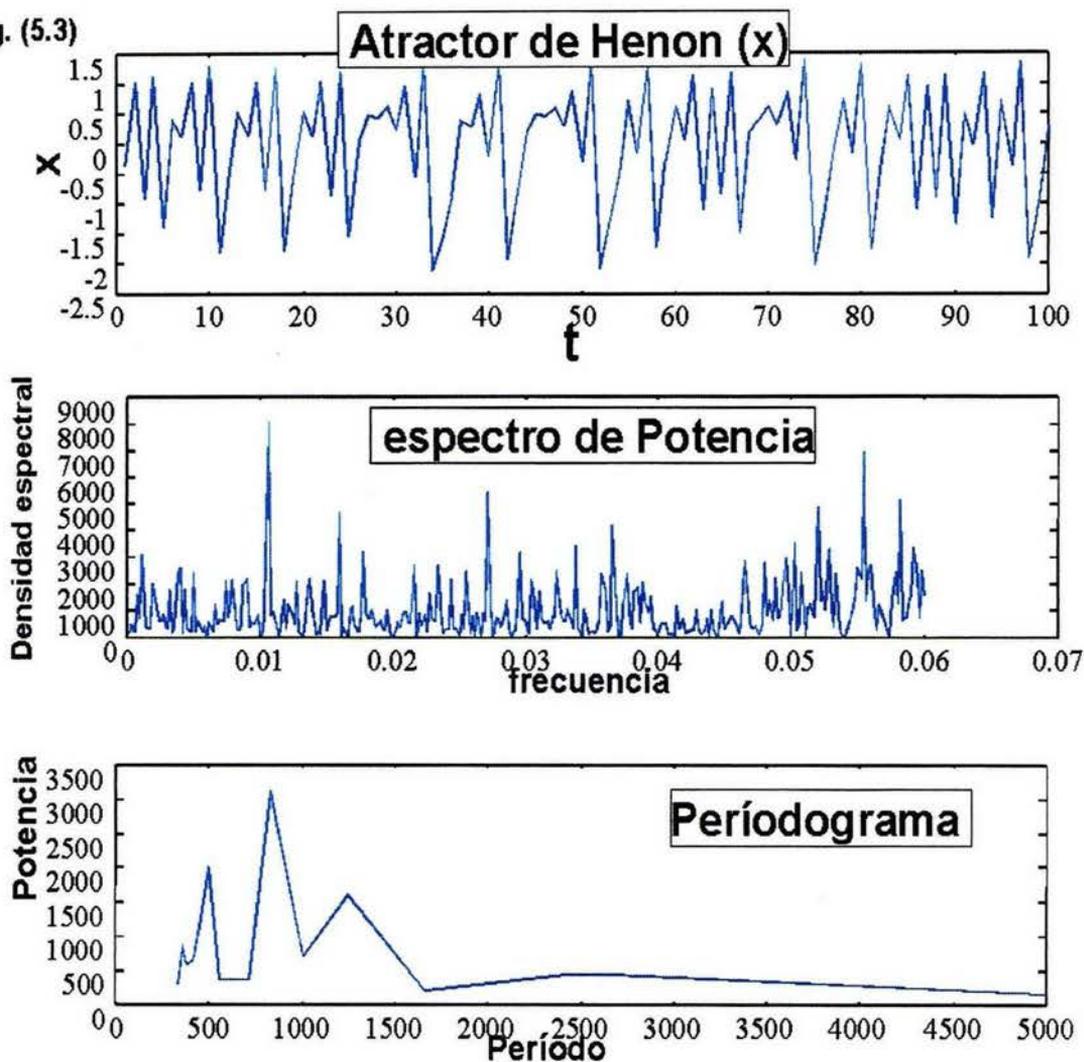
Fig. (5.2)



El simple espectro de potencias, que para efectos de este trabajo, lo definimos como el gráfico de

la norma al cuadrado de la transformada discreta de Fourier de una serie de tiempo en función de la frecuencia (capítulo 1), no es suficiente para determinar la fractalidad de una serie, al graficar en función del período, obtenemos el (períodograma). Si nuestra serie de tiempo es fractal (con autosimilitud), entonces su períodograma deberá mostrar un comportamiento análogo a la curva  $1/f^\beta$ , donde  $f$  es la frecuencia (es una condición necesaria pero no suficiente).

Fig. (5.3)



En el caso de la serie de tiempo de precipitación acumulada mensual en Apizaco, Tlaxcala. El espectro de

potencias nos muestra que es una serie periódica al ubicar un máximo en la frecuencia  $f = 0.0848$ , que corresponde a un período de 11.8 meses que asociamos al ciclo anual. El periodograma correspondiente muestra que es una serie fractal.

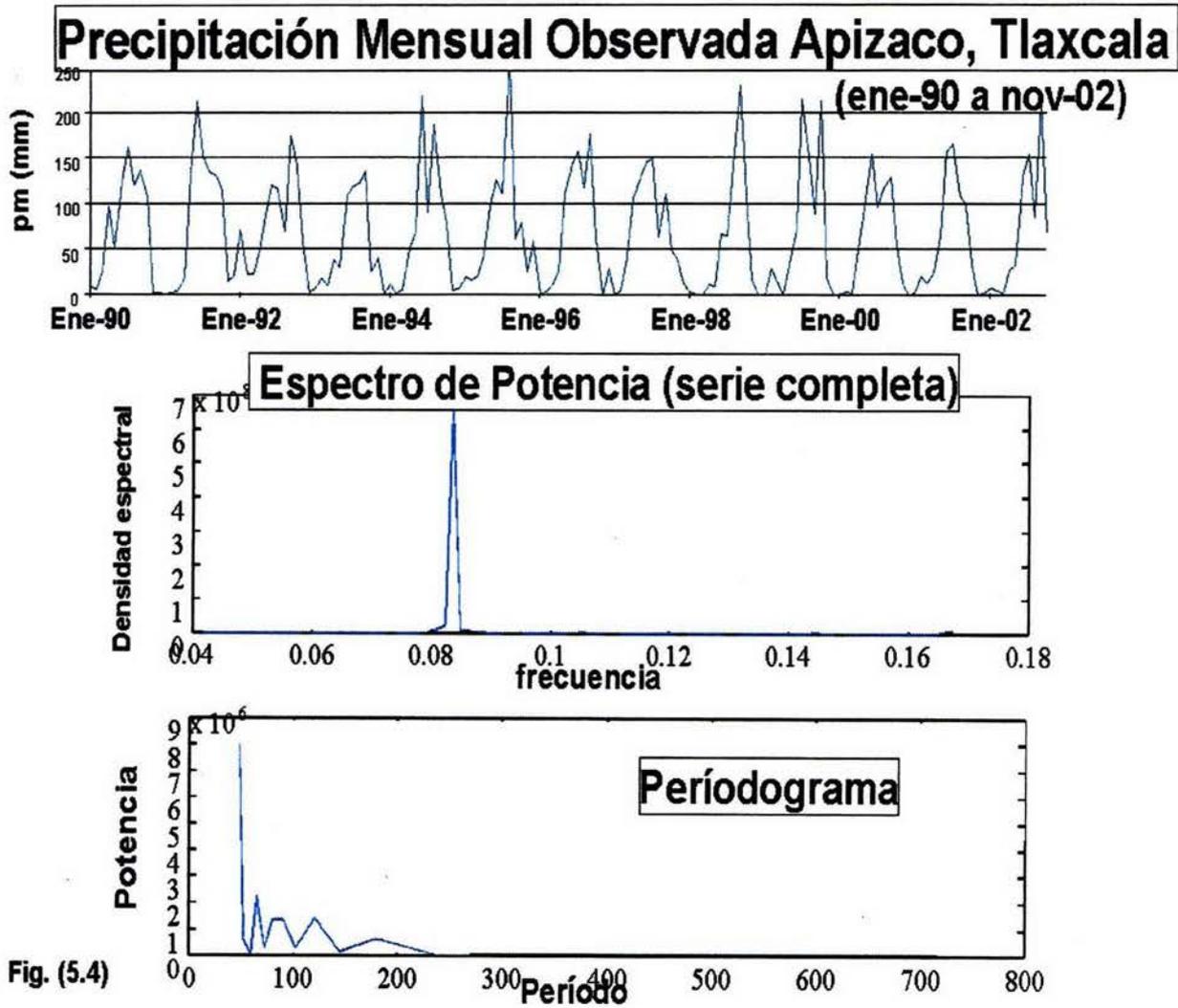
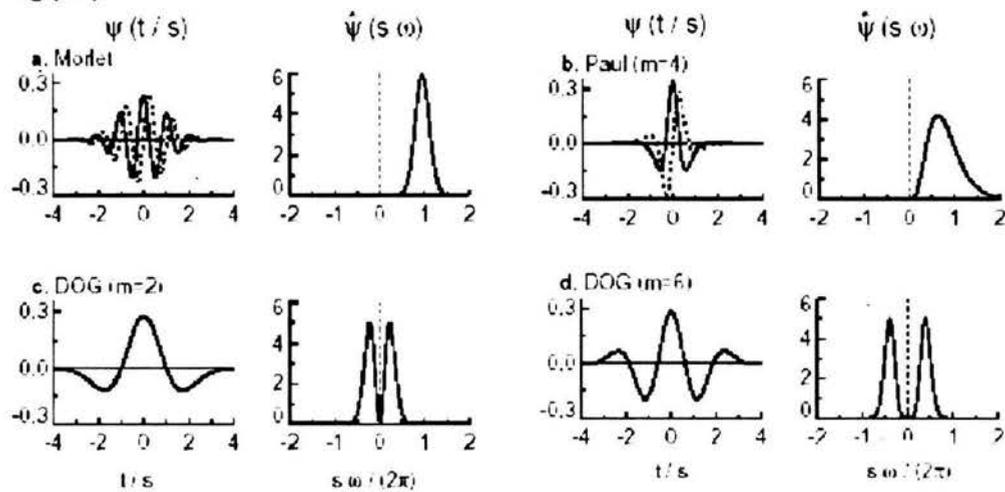


Fig. (5.4)

### 5.1.2 Generando y analizando imágenes de Ondeletas

Existe en la red un sitio disponible para poder generar imágenes de ondeletas sobre nuestra serie de tiempo, auspiciado por la Universidad de Colorado [27].

**Fig.(5.5)**



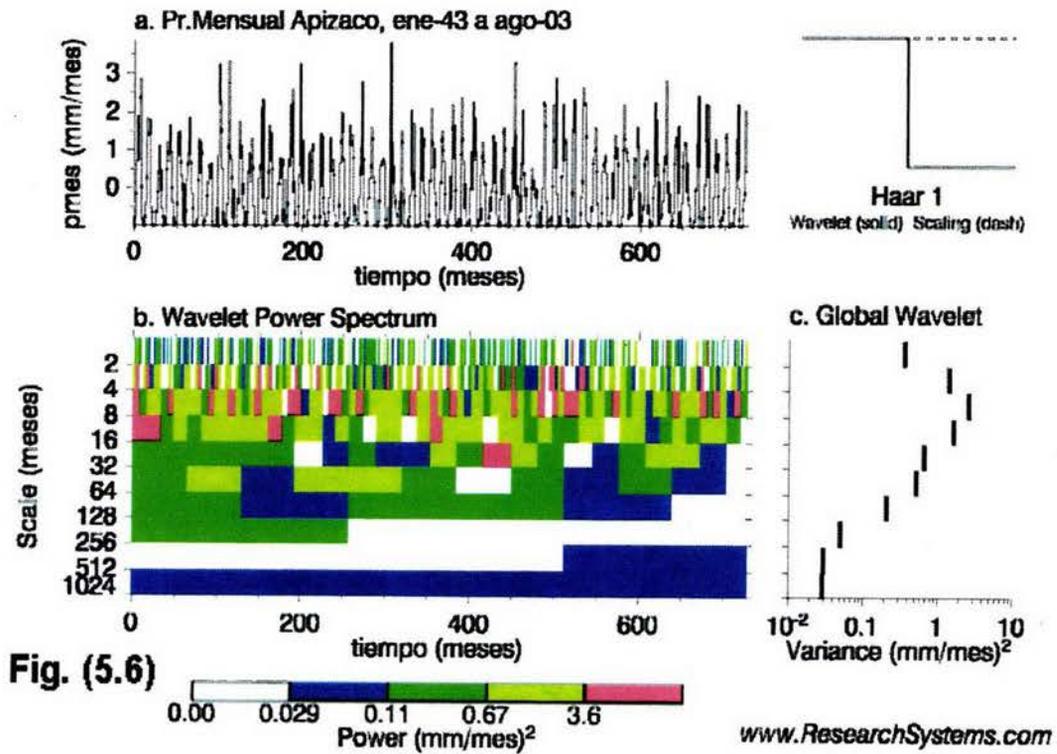
Diferentes funciones base de ondeletas: las gráficas a la izquierda muestran la componente real (trazo continuo) e imaginaria (trazo punteado).

Las gráficas en la derecha para cada caso muestran la ondeleta correspondiente en el dominio de las frecuencias.

Parece sustancial en el análisis de una serie de tiempo, el escoger una base de funciones de ondeleta óptima conforme al conocimiento previo del comportamiento de nuestra serie y al objetivo de nuestro estudio, considerando los siguientes aspectos:

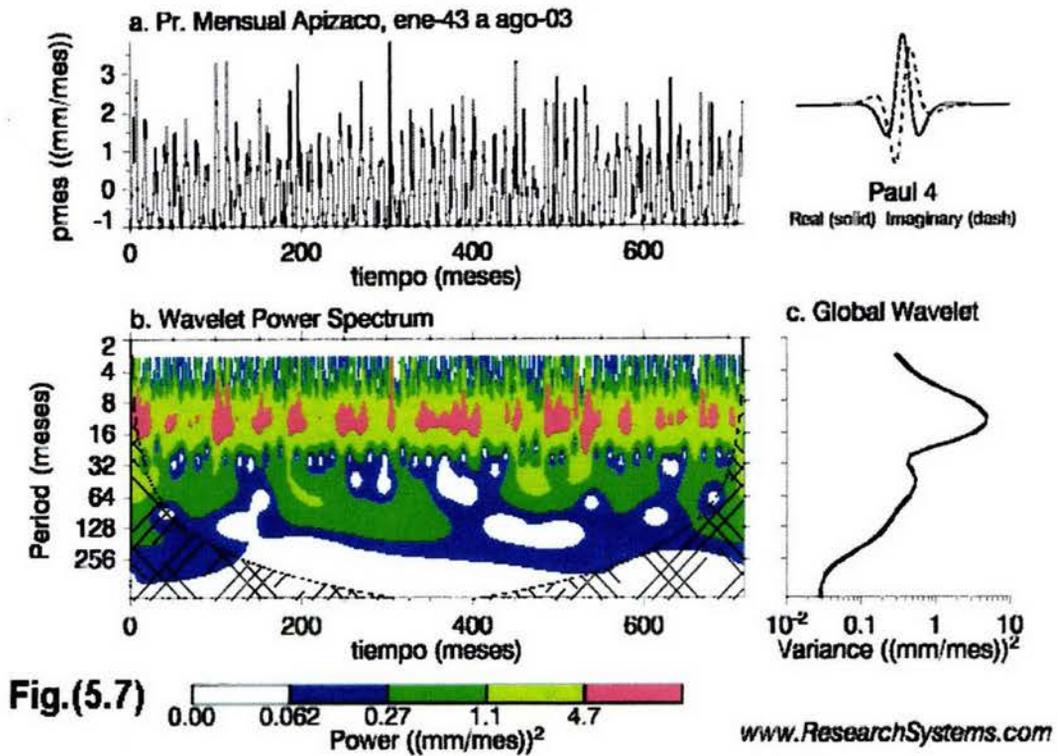
1) **Ortogonalidad.** Bajo una base ortogonal de ondeletas, el número de transformaciones es proporcional al ancho de la función ondeleta para cada escala, bajo una base ortogonal se produce un espectro

de ondeletas constituido por rectángulos, siendo útil para procesamiento de señales. El aplicar una base de ondeletas no ortogonal es adecuado para análisis de series de tiempo donde se esperan variaciones continuas en la amplitud de la ondeleta.



ESTA TESIS NO SALE  
DE LA BIBLIOTECA

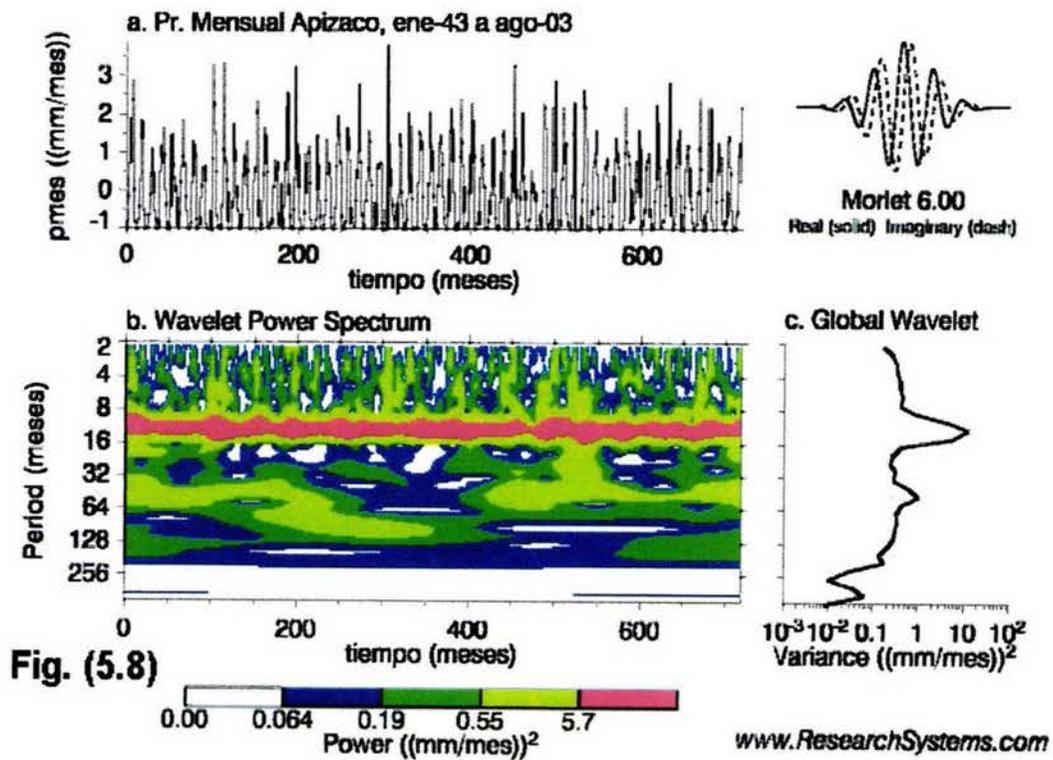
2) **Real o compleja.** Una base de ondeletas compleja produce información sobre la amplitud y la fase; es aplicable para observar comportamiento oscilatorio, una base real produce solo una componente y puede aplicarse para identificar discontinuidades o funciones delta.



3) **Ancho.** La definición en nuestro análisis de ondeletas está ajustada precisamente por el ancho que seleccionemos para nuestra función base, un ancho muy angosto produce una buena definición en la escala temporal y una muy pobre definición para la frecuencia; si escogemos un ancho muy amplio tendremos una mejor resolución en la frecuencia, sacrificando la resolución temporal.

4) **Forma.** Es deseable que la función seleccionada sea conforme al comportamiento mostrado por nuestra serie de tiempo, para una serie de tiempo que muestre discontinuidades o saltos, uno debiera

escoger una función tipo Haar (escalón), en tanto que una serie de tiempo que muestre un comportamiento continuo (suave), puede ser analizada bajo una función como la de Morlet. No obstante las consideraciones anteriores, sí estamos interesados en el espectro de potencias, cualquier función que eligamos proporcionará los mismos resultados cualitativos.



**Fig. (5.8)**

De la imagen anterior, para la serie de tiempo motivo de este estudio, podemos observar claramente un máximo de potencia (franja), centrado en el período de los 12 meses, que podemos asociar a la variabilidad interanual en la precipitación, igualmente se encuentra un máximo relativo en el período de 48 meses, (gráfico a la derecha, parte inferior).

## 5.2 Método de Información Mútua

Fraser y Swinney en 1986 propusieron el criterio de Información Mútua, como un mecanismo para determinar el valor más apropiado del valor de retraso  $\tau$ , a fin de reconstruir un atractor a partir de una Serie de Tiempo,[20].

El método consiste en medir la dependencia general de dos variables, (en este caso  $x(t)$  y  $x(t + \tau t)$ ). Para un atractor bidimensional, en caso de que no garantizemos que exista independencia lineal entre las variables  $x(t)$  y  $x(t + \tau t)$ , todas las trayectorias coincidirán en la recta  $x(t + \tau t) = mx(t)$ ; luego entonces  $\tau$  debe seleccionarse de forma que garantice una dependencia mínima entre  $x(t)$  y  $x(t + \tau t)$ .

Consideremos un proceso en el que una señal se envía a un receptor. Sea  $S$  un sistema que consiste de un conjunto de posibles señales:  $s_1, s_2, \dots, s_n$  y sus probabilidades asociadas:  $P_s(s_1), P_s(s_2), \dots, P_s(s_n)$ .  $P_s$  mapea señales con sus probabilidades.

Sea  $P_s$  una función de probabilidad que se distribuye en  $k$  valores discretos, ¿De cuántas formas se pueden distribuir  $n$  objetos idénticos en  $k$  cajas?

Número de maneras para seleccionar  $n$  objetos:  $n!$ ; cada caja tiene  $n_i!$  configuraciones, entonces el número de posibles arreglos es:

$$W = \frac{n!}{\prod_i n_i!} \quad (5.1)$$

donde  $\sum_i n_i = n$

La cantidad de sorpresa que alguna distribución pueda causarnos esta dada por la entropía  $S$ :

$$S = -\frac{1}{n} \ln(W) = -\frac{1}{n} \ln(n!) - \sum_i \ln(n_i!). \quad (5.2)$$

sí  $n \rightarrow \infty$ :

$$\ln(n!) \approx n \ln(n) - n. \quad (5.3)$$

De esta manera:

$$S = -\lim_{n \rightarrow \infty} \sum \left( \frac{n_i}{n} \ln \left( \frac{n_i}{n} \right) \right) \quad (5.4)$$

$$S = -\sum_i p_i(s_i) \ln(p_i(s_i)) \quad (5.5)$$

Estamos interesados en conocer la dependencia de  $x(t + \tau)$  con respecto a  $x(t)$ , designaremos al sistema  $[s, q]$  como  $[x(t), x(t + \tau)]$ , en estos términos deseamos conocer la incertidumbre sobre un valor  $q$ , una vez que hemos encontrado un valor  $s_i$ .

$$S = -\sum_i p_i(s_i) \ln(p_i(s_i)) \quad (5.6)$$

$$H(Q | s_i) = -\sum_j P_{q|s}(q_j | s_i) \log[P_{q|s}(q_j | s_i)] \quad (5.7)$$

$$H(Q | s_i) = \sum_j \left[ \frac{P_{sq}(s_i, q_j)}{P_s(s_i)} \right] \times \log \left[ \frac{P_{sq}(s_i, q_j)}{P_s(s_i)} \right]. \quad (5.8)$$

Donde  $P_{sq}(s_i, q_j)$  corresponde a la probabilidad de que dado  $s$  se presente  $q$ , esto es, el número de parejas  $(s_i, q_j)$  entre el número de parejas con origen  $s_i$ .

Sí promediamos sobre  $s_i$  obtenemos la incertidumbre de  $x(t + \tau)$  dado  $x(t)$ :

$$H(Q | S) = - \sum_i P_s(s_i) H(Q | s_i) = - \sum_{i,j} P_{sq}(s_i, q_j) \log \left[ \frac{P_{sq}(s_i, q_j)}{P_s(s_i)} \right] = H(S, Q) - H(S). \quad (5.9)$$

Donde

$$H(S, Q) = - \sum_{i,j} P_{sq}(s_i, q_j) \log [P_{sq}(s_i, q_j)]. \quad (5.10)$$

Dado que  $H(Q)$  es la incertidumbre de  $q$  y  $H(Q | S)$  es la incertidumbre sobre  $q$  dado  $s$ , la información mútua (I.M.), corresponde a la cantidad de información que ganamos respecto a  $q$  dado  $s$ .

$$I(Q, S) = H(Q) - H(Q | S) = H(Q) + H(S) - H(S, Q) = I(S, Q). \quad (5.11)$$

La expresión para el valor de (I.M.)  $M$ , es

$$M = \sum_{i=1}^{(N-(m-1))} \tau P(x_i, x_{i+\tau}, x_{i+2\tau}, \dots, x_{i+(m-1)\tau}) \times \log \left[ \frac{P(x_i, x_{i+\tau}, x_{i+2\tau}, \dots, x_{i+(m-1)\tau})}{P(x_i) P(x_{i+\tau}) P(x_{i+2\tau}) \dots P(x_{i+(m-1)\tau})} \right] \quad (5.12)$$

donde  $P(x_i)$  es la probabilidad de ocurrencia de la variable  $x_i$  y  $P(x_i, x_{i+\tau}, x_{i+2\tau}, \dots, x_{i+(m-1)\tau})$  es la probabilidad de ocurrencia de que dado  $x_i$ , la coordenada independiente esté dada por la serie  $x_i, x_{i+\tau}, x_{i+2\tau}, \dots, x_{i+(m-1)\tau}$ .

¿Cómo se obtiene el valor de (I.M.)?. El primer paso consiste en cubrir el espacio  $(s, q)$  con una r cula de  $N_r \times N_c$  celdas, este paso nos enfrenta al problema de seleccionar el tama o  ptimo para cada celda, s  seleccionamos un tama o de celda muy 'grande' obtenemos 'muchos' puntos dentro de cada celda y un valor para  $P_{sq}$  'sobreevaluado' y como consecuencia 'subestimamos' el valor de I. M., s  por el contrario elegimos un tama o de celda muy 'peque o' el valor de  $P_{sq}$  sera 'subvaluado' y el valor de I. M. sera 'sobreestimado', el algoritmo de Fraser y Swinney resuelve este paradigma seleccionando un tama o que garantiza el valor exacto para cada regi n:

1) Partición de cada eje (s,q) en  $2^m$  segmentos equiprobables. De la ecuación (5.10), tenemos:

$$I(Q, S) = H(Q) + H(S) - H(S, Q). \quad (5.13)$$

Desarrollando cada término:

$$H(S) = - \sum_{i=1}^m P_s(s_i) \log(P_s(s_i)) = - \sum_{i=1}^m \left(\frac{1}{2}\right)^m \log\left[\left(\frac{1}{2}\right)^m\right] = H(Q). \quad (5.14)$$

$$H(S) + H(Q) = -2 \left[ \sum_{i=1}^m \left(\frac{1}{2}\right)^m \log\left[\left(\frac{1}{2}\right)^m\right] \right]. \quad (5.15)$$

Como

$$\sum_{i=1}^m \left(\frac{1}{2}\right)^m \log\left[\left(\frac{1}{2}\right)^m\right] = \sum_{i=1}^m \left(\frac{1}{2}\right)^m m [\log(1) - \log(2)]. \quad (5.16)$$

y  $\sum_{i=1}^m \left(\frac{1}{2}\right)^m = 1 - \left(\frac{1}{2}\right)^m$  cuando  $m \rightarrow \infty$

$$H(Q) + H(S) = m \log(4) \quad (5.17)$$

sustituyendo en (5.10),

$$I_m(Q, S) = m \log(4) - H(S, Q). \quad (5.18)$$

2) Generar una función autorrecursiva que nos permita, de acuerdo a la distribución de los puntos para cada región, identificar el tamaño exacto para obtener el valor de I.M. Definimos una secuencia de particiones del plano (s, q),  $G_0, G_1, G_2, \dots, G_m, \dots$  cada partición constituye una red de  $4^m$  elementos que se obtienen al dividir el plano en  $2^m$  celdas equiprobables, designamos a cada una de estas celdas (regiones) como  $R_m(K_m)$ ,  $K_m$  toma todos los valores entre 1 y  $4^m$ , de esta manera la ecuación (5.13) se expresa como:

$$i_m = m \log(4) + \sum_{i,j=0} P_{sq}(R_m(i, j)) \log[P_{sq}(R_m(i, j))] \quad (5.19)$$

Notemos que para obtener  $G_1$  a partir de  $G_0$ ,  $R_0(K_0)$  se divide en 4 subelementos:  $R_1(0)$ ,  $R_1(1)$ ,  $R_1(2)$ ,  $R_1(3)$ . Para obtener  $G_2$  cada elemento de  $G_1$  se divide en 4 subelementos:  $R_2(k_1, k_2)$ , donde  $k_1, k_2 \in 0, 1, 2, 3$ .

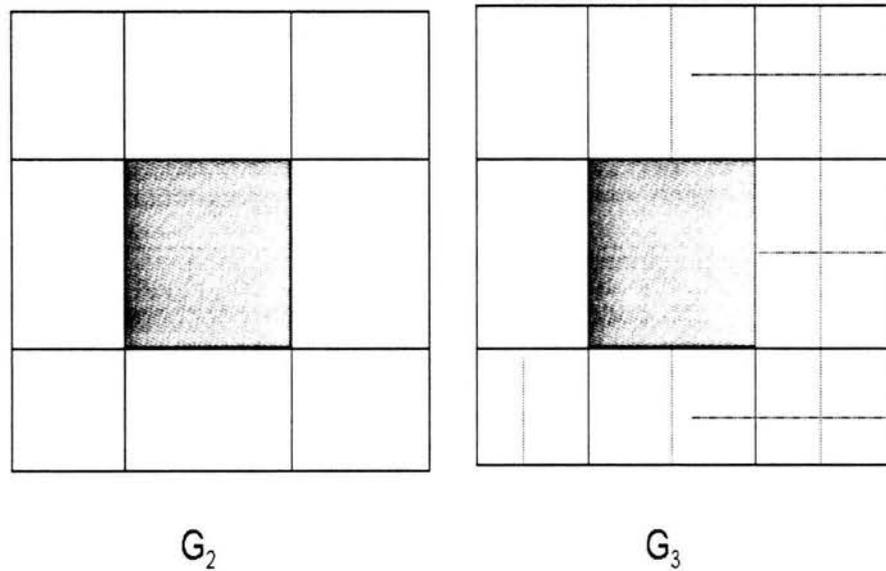


Fig. (5.9) Dos pasos consecutivos en la secuencia de particiones. La región sombreada en  $G_2$ , se subdivide en 4 subespacios en la partición  $G_3$ .

Veamos que sucede al elemento  $R_m(K_m)$  de  $G_m$  al pasar a la siguiente partición  $m + 1$ , designamos como  $C_T$  a la contribución al valor de  $I_m$  de  $R_m(K_m)$  y  $C_j$  a la contribución al valor de  $I_{m+1}$  del subelemento  $R_{m+1}(K_m, j)$ , entonces

$$C_T = P_T m \log(4) + P_T \log(P_T). \quad (5.20)$$

donde  $P_T = P_{sq}(R_m(K_m))$  y

$$C_j = P_j(m+1)\log(4) + P_j\log(P_j). \quad (5.21)$$

donde  $P_j = P_{sq}(R_{m+1}(K_m, j))$ . La contribución de la región cubierta por  $R_m(K_m)$  a  $i_{m+1}$  es

$$\sum_{j=0}^3 C_j = P_T(m+1)\log(4) + \sum_{j=0}^3 P_j\log(P_j). \quad (5.22)$$

Comparando (5.15) y (5.16) sólo hemos incrementado un término  $P_T\log(4)$  y el término  $P_T\log(P_T)$  se sustituye con  $\sum_{j=0}^3 P_j\log(P_j)$ , si  $P_{sq}$  es homogénea en  $R_m(K_m)$ , entonces  $P_0 = P_1 = P_2 = P_3 = \frac{P_T}{4}$  y  $C_T = C_0 + C_1 + C_2 + C_3$  y no ganamos información al subdividir el elemento.

De esta manera, podemos expresar  $I(S, Q)$  empleando una función recursiva  $F$ , como

$$I(S, Q) = F(R_0(K_0)) \quad (5.23)$$

Donde si  $P_{sq}$  tiene una distribución homogénea sobre  $R_m(K_m)$

$$F(R_m(K_m)) = P_{sq}(R_m(K_m))\log[P_{sq}(R_m(K_m))]. \quad (5.24)$$

y si  $P_{sq}$  **no** tiene una distribución homogénea sobre  $R_m(K_m)$

$$F(R_m(K_m)) = P_{sq}(R_m(K_m))\log(4) + \sum_{j=0}^3 F(R_{m+1}(K_{m+1})). \quad (5.25)$$

Es así como la función recursiva  $F$  continúa subdividiendo la partición en las regiones donde se necesita.

La implementación del algoritmo utilizado se ilustra en la figura (5.10).

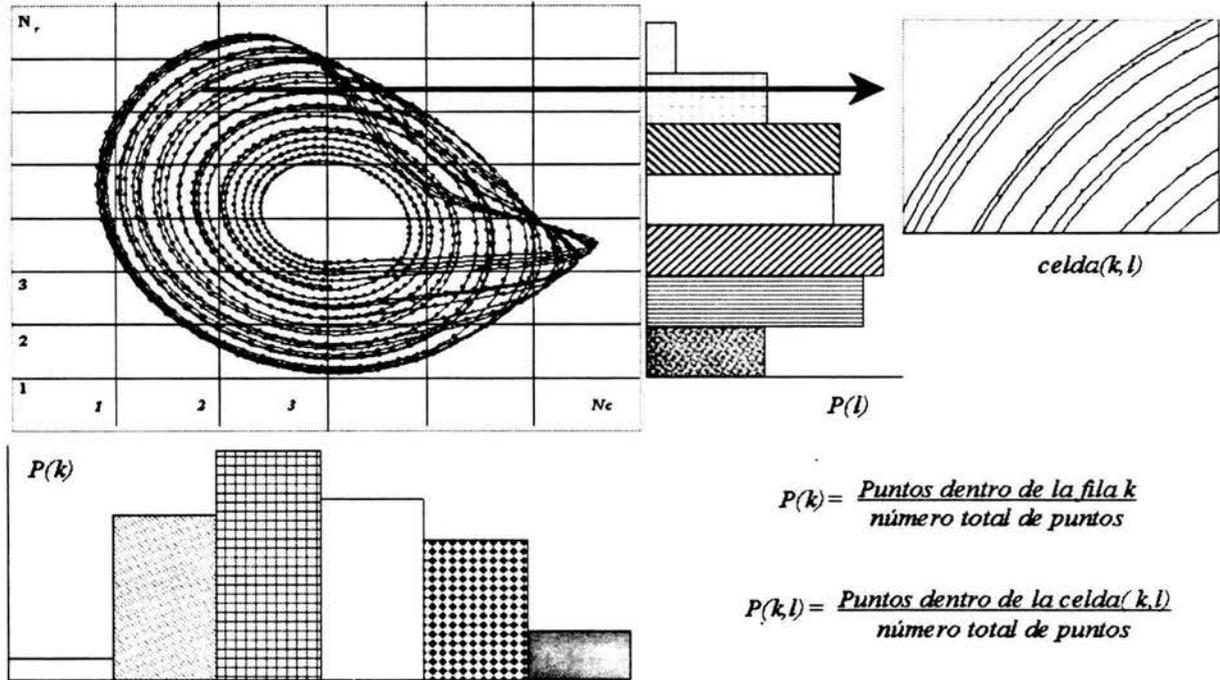


Fig. (5.10) Procedimiento para encontrar el valor de Información Mútua.

Se realiza una reconstrucción bidimensional del atractor en el plano  $x(t), x(t + \tau)$ . El plano se divide en una retícula de  $N_c$  columnas y  $N_r$  renglones; las funciones de probabilidad para  $x_i$  y  $x_{i+\tau}$  se generan contando los puntos que inciden en cada una de las celdas y dividiendo por el número de puntos de la serie.  $P(k)$  y  $P(l)$  corresponden a las probabilidades de ocurrencia de los puntos del atractor en la columna  $k$  y el renglón  $l$ , y la probabilidad de la unión  $P(k, l)$ , para cada región.

El valor de (I.M.) se obtiene mediante la suma

$$M = \sum_{k=1}^{N_c} \sum_{l=1}^{N_r} P(k, l) \log \left[ \frac{P(k, l)}{P(k)P(l)} \right] \quad (5.26)$$

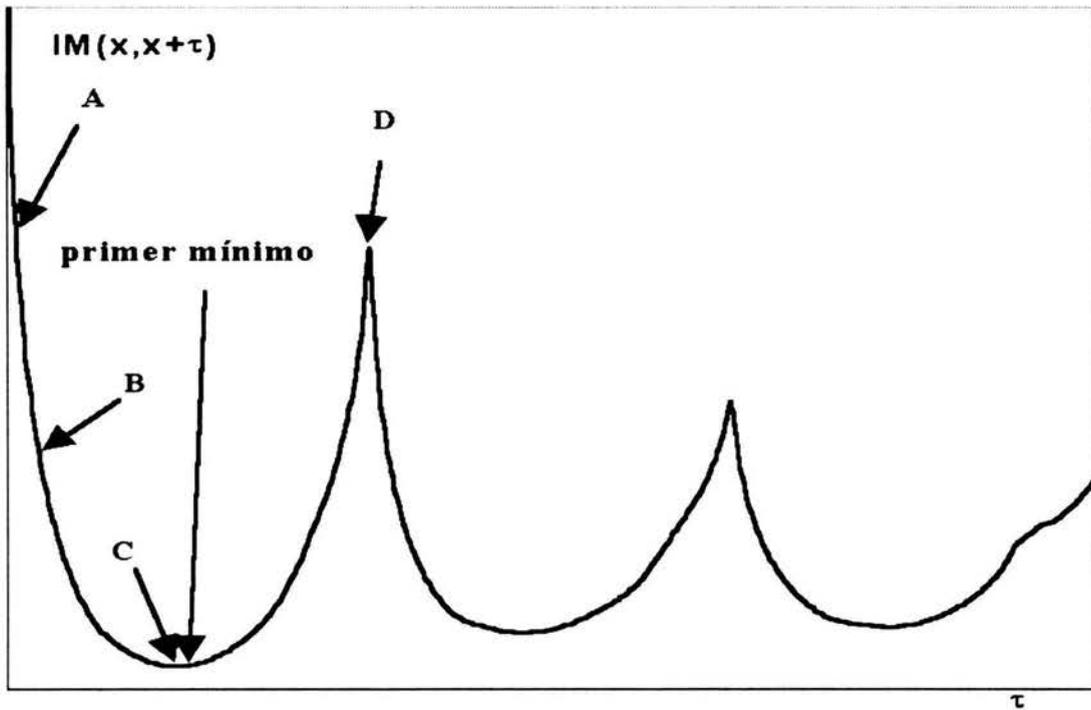
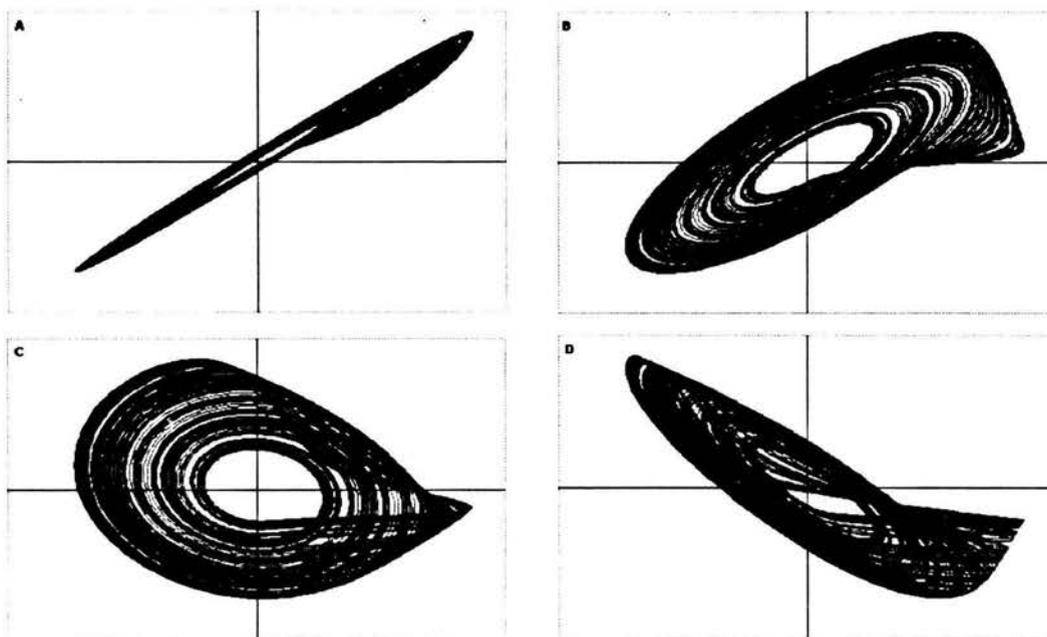


Fig. (5.11) Gráfica del valor de información mútua en función de  $\tau$

Este proceso se realiza para diferentes valores de  $\tau$ ; aquel valor de  $\tau$  que proporcione un mínimo en  $M$ , es seleccionado como el valor óptimo de retraso,[30].



**Fig. (5.12) Reconstrucción del atractor correspondiente en el espacio de fases, para diferentes valores de  $\tau$ , en (A),  $\tau$  es muy pequeño  $x \cong x+\tau$  y el atractor se alinea a  $45^\circ$ , al aumentar  $\tau$ , el atractor se expande hasta (C), (Primer mínimo).**

### 5.2.1 Ecuaciones de Rössler - Método de Runge Kutta

A fin de contar con una serie de datos sin ruido, se resolvió como se sugiere en el artículo de Fraser y Swinney, el sistema de Ecuaciones de Rössler (ver capítulo 2 sección 8),[11],[7].

$$x' = -z - y$$

$$y' = x + ay$$

$$z' = b + z(x - c)$$

Con los parámetros  $a = 0.15$ ,  $b = 0.20$  y  $c = 10.0$  y las condiciones iniciales  $(10, 0, 0)$ , utilizando el Método de Runge Kutta.

Se encontró en forma exacta el valor para el período planteado por Fraser y Swinney de 193.3.

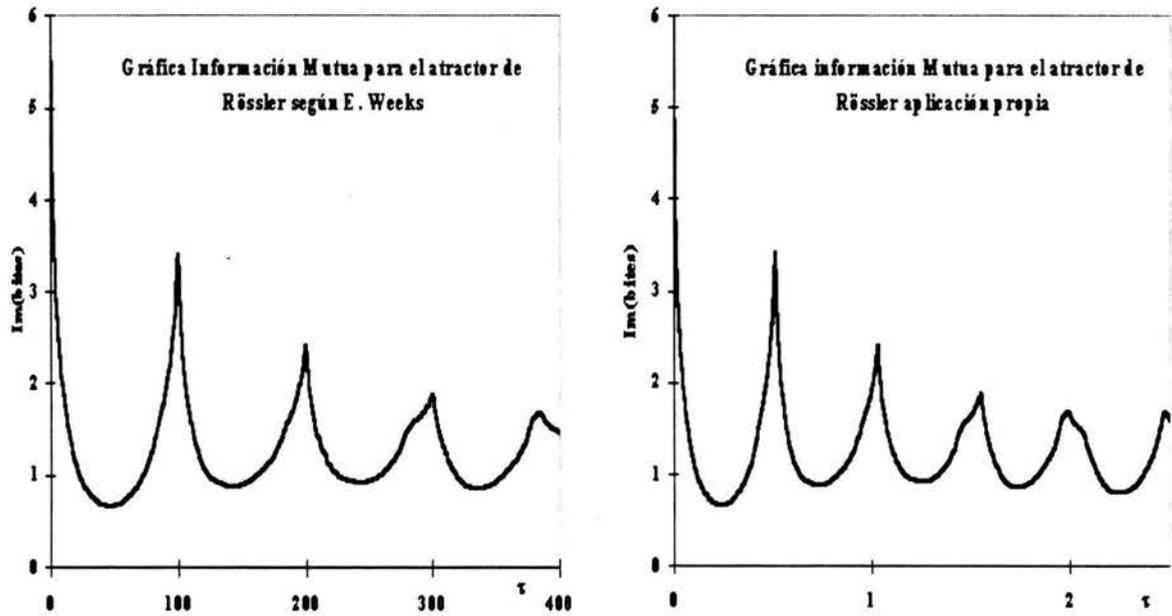
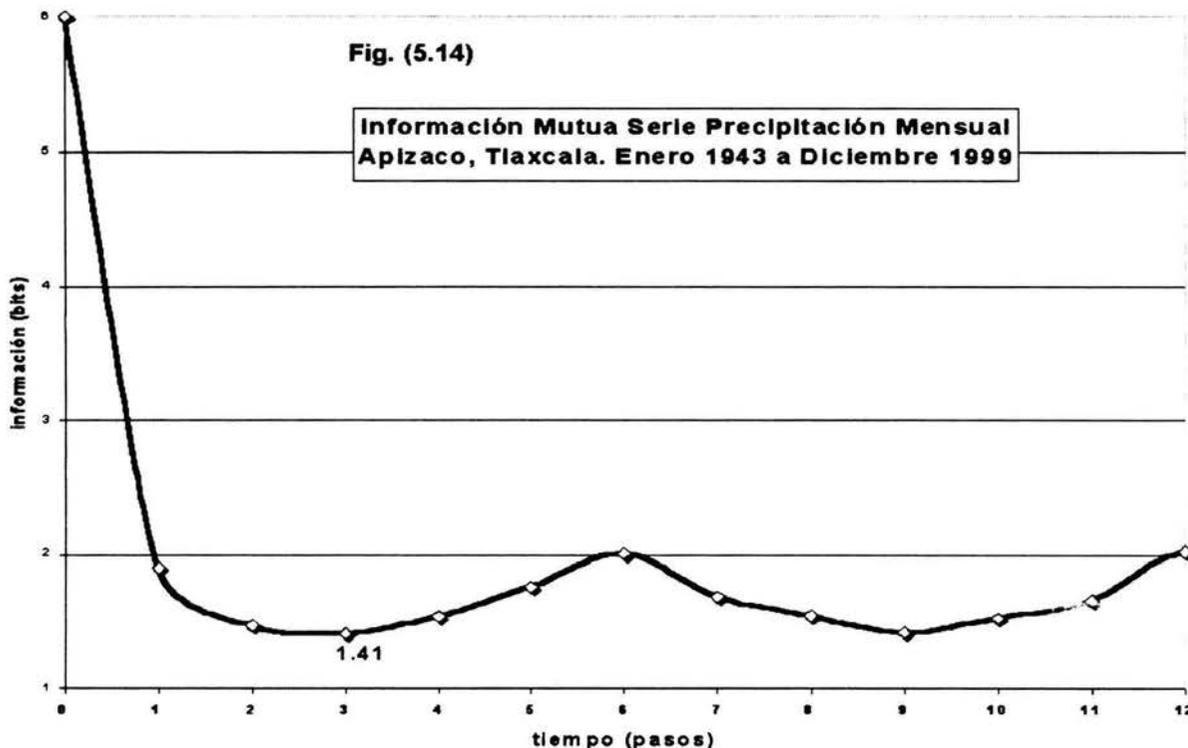


Fig. (5.13) Comparando resultados publicados y propios.

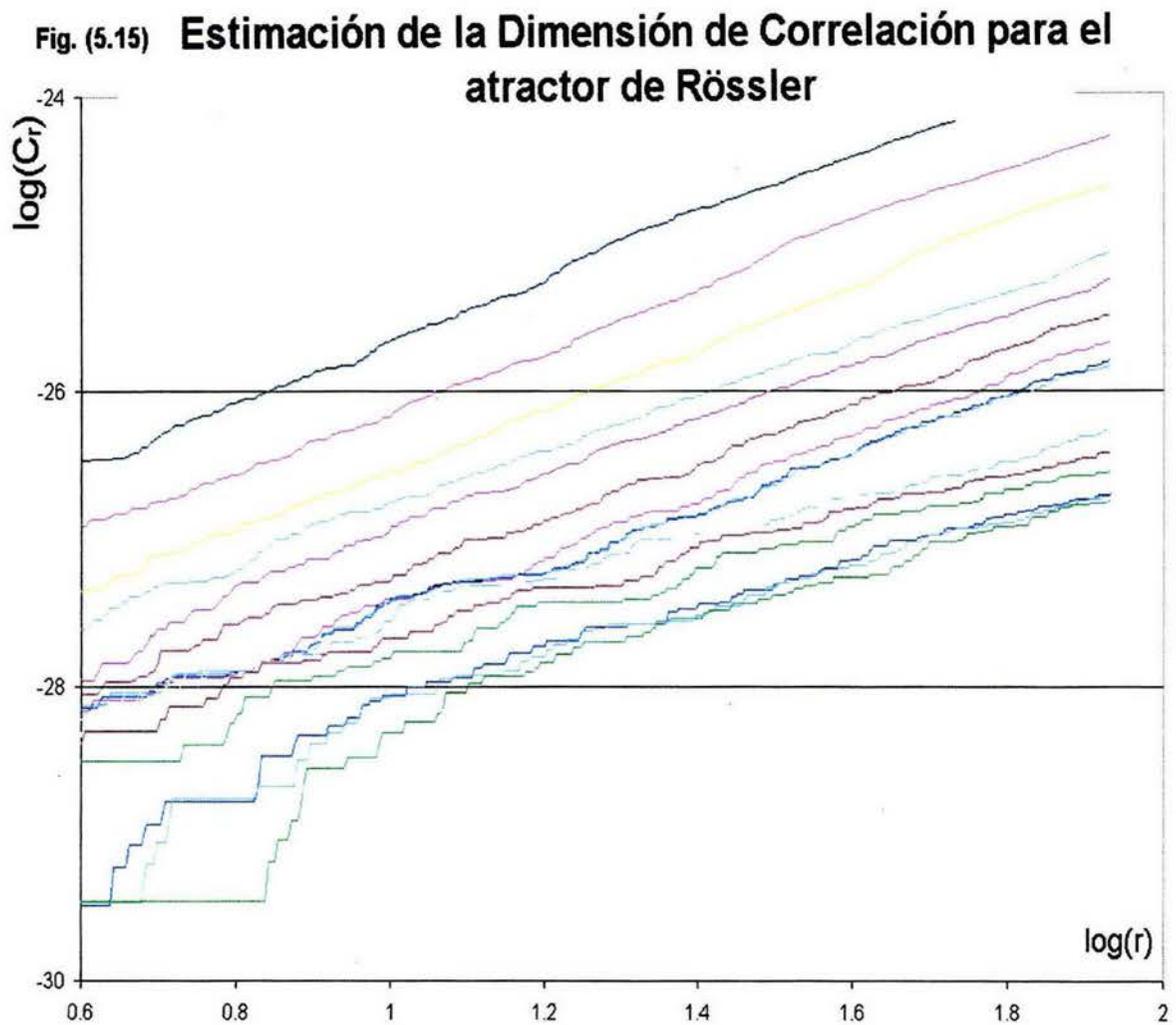
El algoritmo utilizado para obtener el valor de (I.M.) fue implementado por E. Weeks, [26], reproduciendo para la serie de datos del atractor de Rössler los resultados descritos por Fraser y Swinney,

permitiéndonos garantizar la bondad del algoritmo. Una vez asegurado el buen funcionamiento del programa se aplicó el método de (I.M.) a la serie de datos de precipitación mensual en la estación de Apizaco, Tlaxcala desde Enero de 1943 a Diciembre de 1999, resultando un valor para  $\tau$  de 3 meses.



### 5.2.2 Estimación de la Dimensión del Atractor

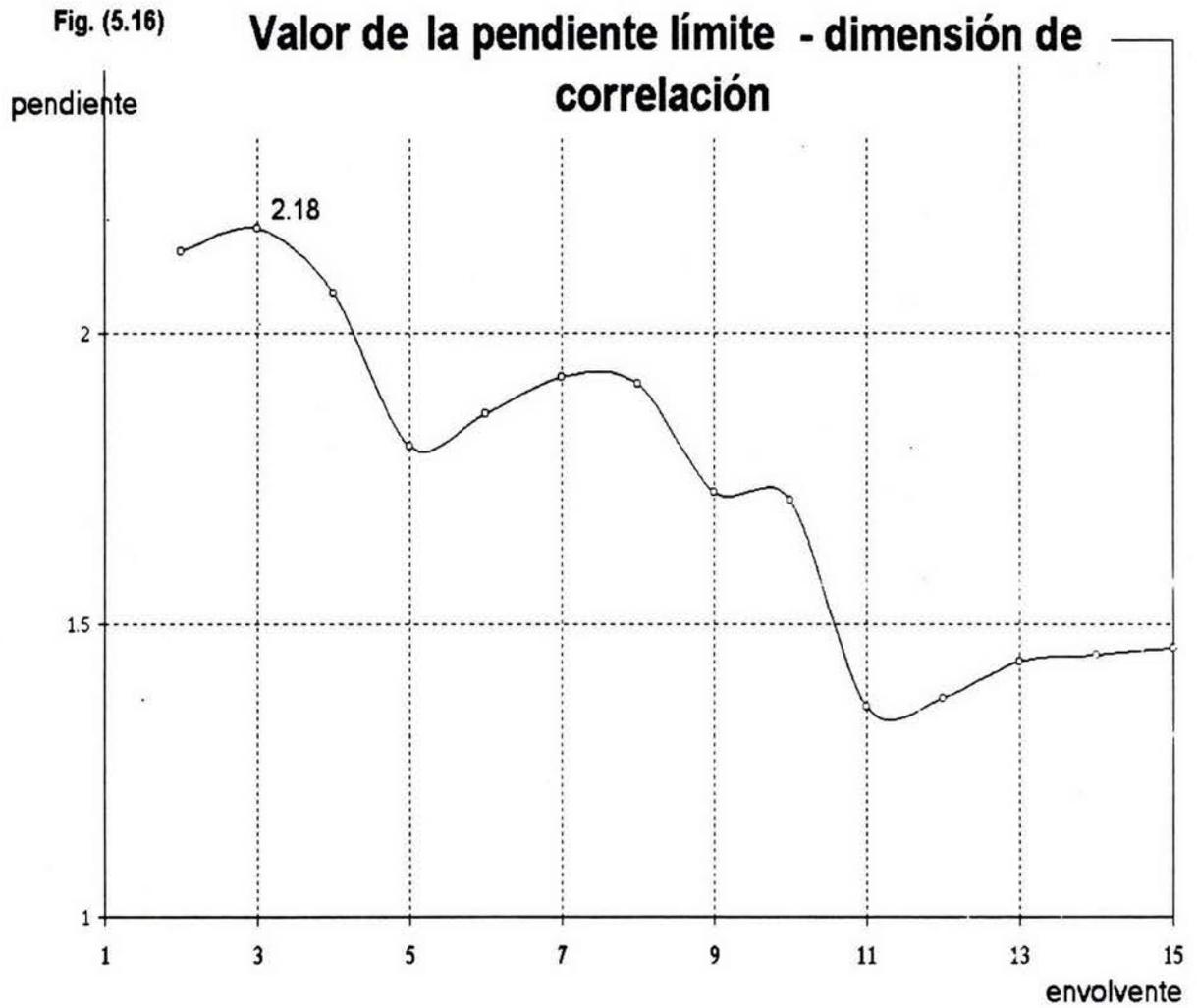
De acuerdo con el proceso descrito en la sección 2.7, se implementó un programa en Lenguaje Fortran 77 para estimar la Dimensión del Atractor asociada a una serie de tiempo; encontrando que para la serie de datos Rössler se obtiene una dimensión envolvente de 3, conforme a lo esperado.[30].



Recordar que:

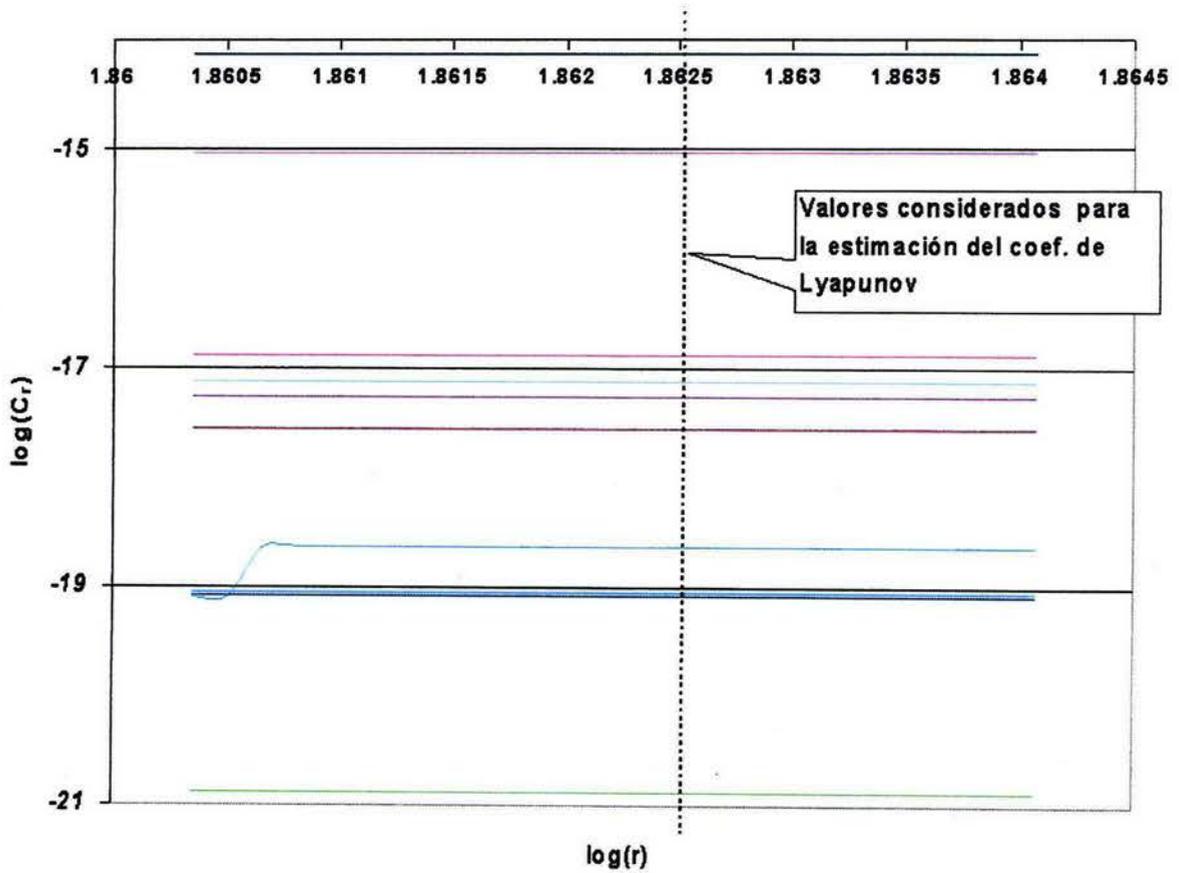
$$m > D_c$$

(5.27)



Para el caso de la serie precipitación mensual se obtiene una dimensión envolvente de 7 que nos indica el nivel de complejidad extrema de esta serie.

**Fig. (5.17) Estimación de la Dimensión de Correlación para Precipitación Mensual (Apizaco, Tlax.)**



5.2.3 Límite de Predictibilidad

Encontramos un coeficiente de Lyapunov de 0.22 lo que nos permite asociar un límite de predictibilidad de 4.48 meses para la serie de tiempo en estudio.

**Esquema del proceso realizado para la estimación del Coeficiente de Lyapunov**

pendiente (i)	log(Cr)	$\log(Cr)_i - \log(Cr)_{i+1}$	abs ( $\log(Cr)_i - \log(Cr)_{i+1}$ )	promedio	1 / promedio
1	-14.13	0.897	0.897	0.223	4.477
2	-15.02	1.850	1.850		
3	-16.87	0.239	0.239		
4	-17.11	0.143	0.143		
5	-17.25	0.288	0.288		
6	-17.54				

Tabla (5.1)

# Capítulo 6

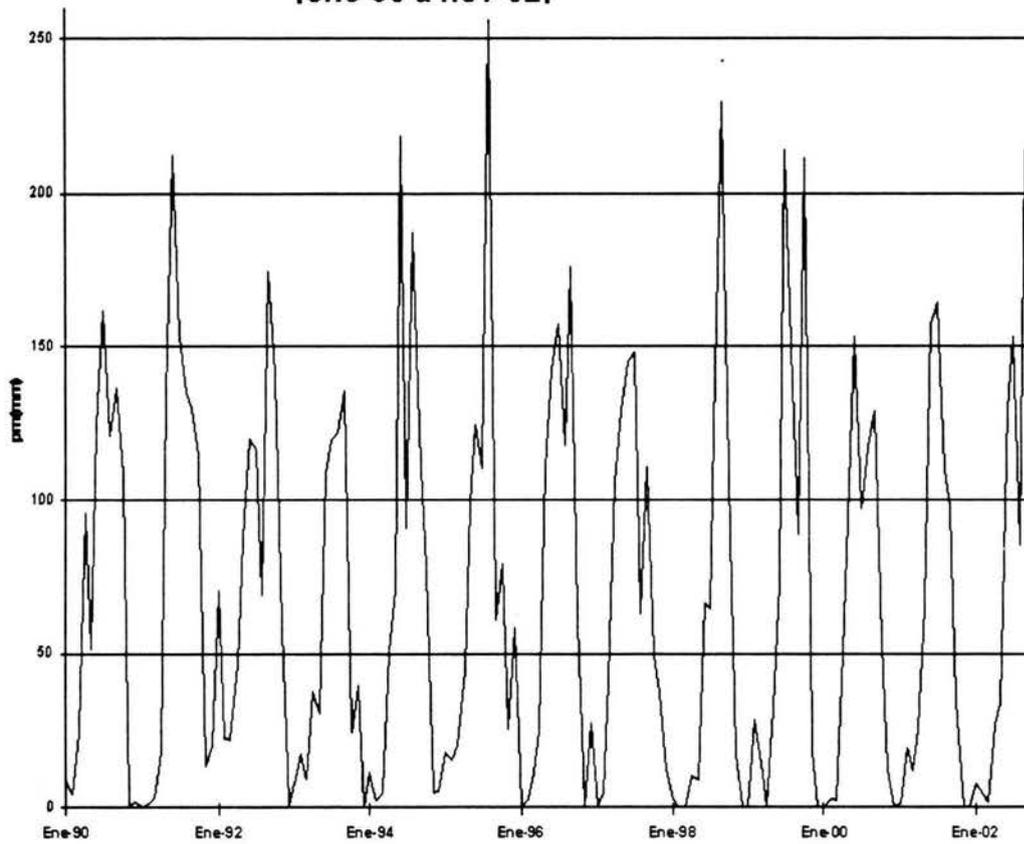
## EXPLOTANDO LA RED NEURONAL

### 6.1 Normalizando la Serie

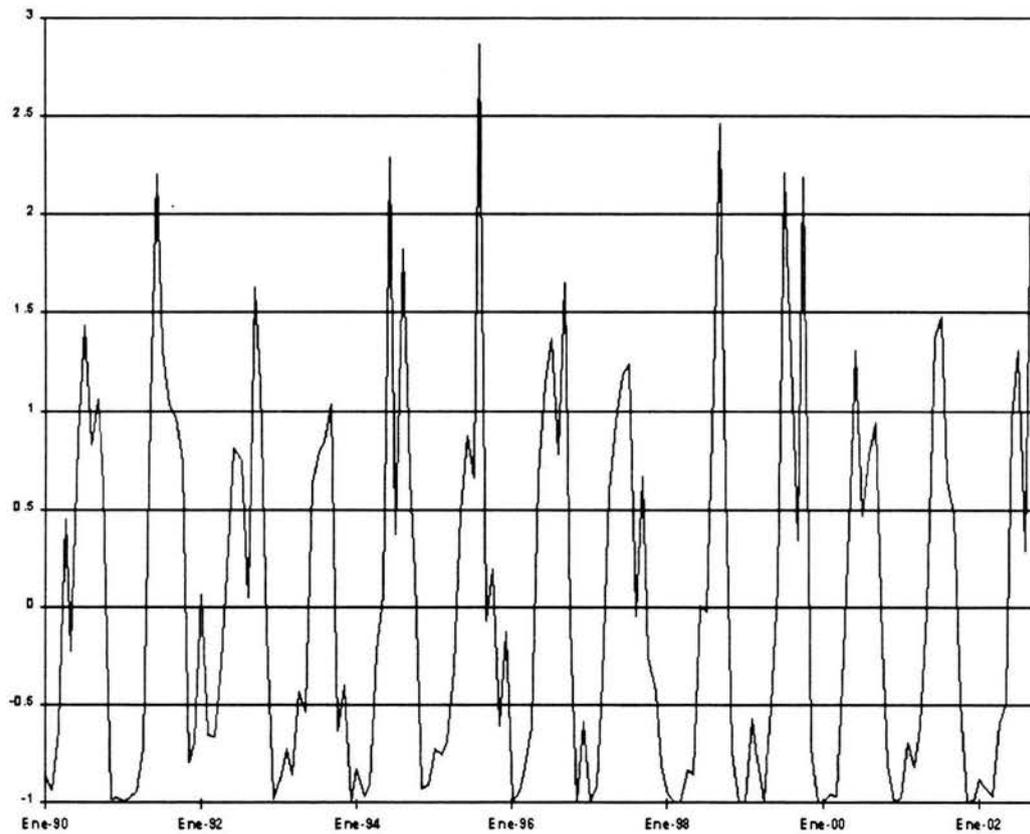
Una vez que conocemos la información relativa a la dinámica de la Serie de tiempo, podemos abocarnos a la tarea de reconstrucción en el espacio fase a fin de explotar la serie generando pronóstico; utilizando el software FIRNET [37], desarrollado en g++. Requerimos aplicar una transformación lineal a la serie de tiempo de modo que para alimentar la red, la serie transformada cuente con media 0 y varianza 1.

Serie de datos real, para cada mes se integra la cantidad de lluvia (mm) acumulada.

**Fig. (6.1) Serie de datos precipitación mensual acumulada, Apizaco, Tlax. (ene-90 a nov-02)**



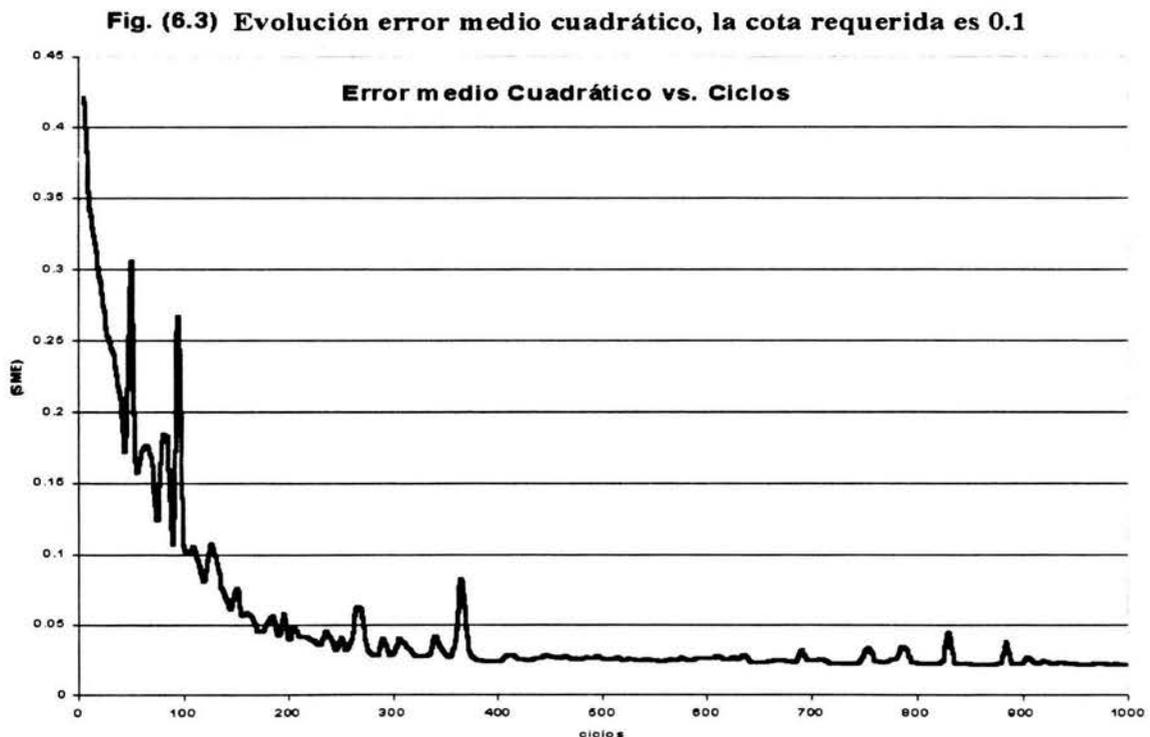
Serie de datos transformada para alimentar a la red, conforme a los requisitos establecidos por la herramienta: (media=0, varianza=1).



**Fig. (6.2) Serie Precipitación Mensual Acumulada Ene-90 a Oct-92 transformada para alimentar a la red.**

## 6.2 Seleccionado la arquitectura

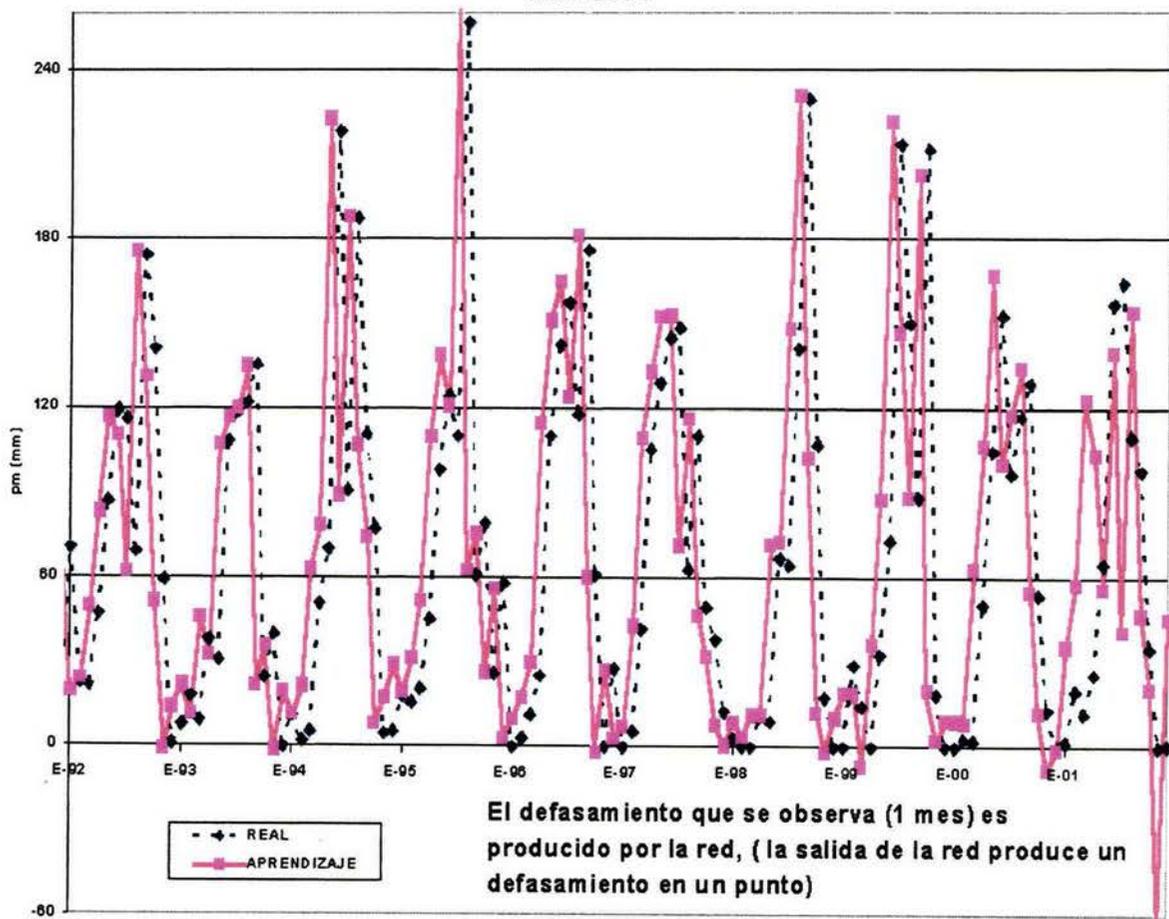
La red nos arroja como parámetro el valor del Error Medio Cuadrático (SME), para cada ciclo, tenemos que buscar una arquitectura que permita a la red converger rápidamente (menor número de ciclos o épocas) a valores inferiores a una cota ( $SME < 0.1$ ); por otro lado Wang nos indica que prácticamente cualquier sistema converge con una arquitectura 25:5:5, (25 nodos en la capa externa, 5 nodos en la capa oculta y 5 nodo en la capa de salida), heurísticamente y después de un buen número de intentos seleccionamos 15:25:1, para la serie de de datos de Apizaco, Tlaxcala.



La curva típica de evolución del Error Cuadrático Medio(SME), es similar a la mostrada en la gráfica 6.3.

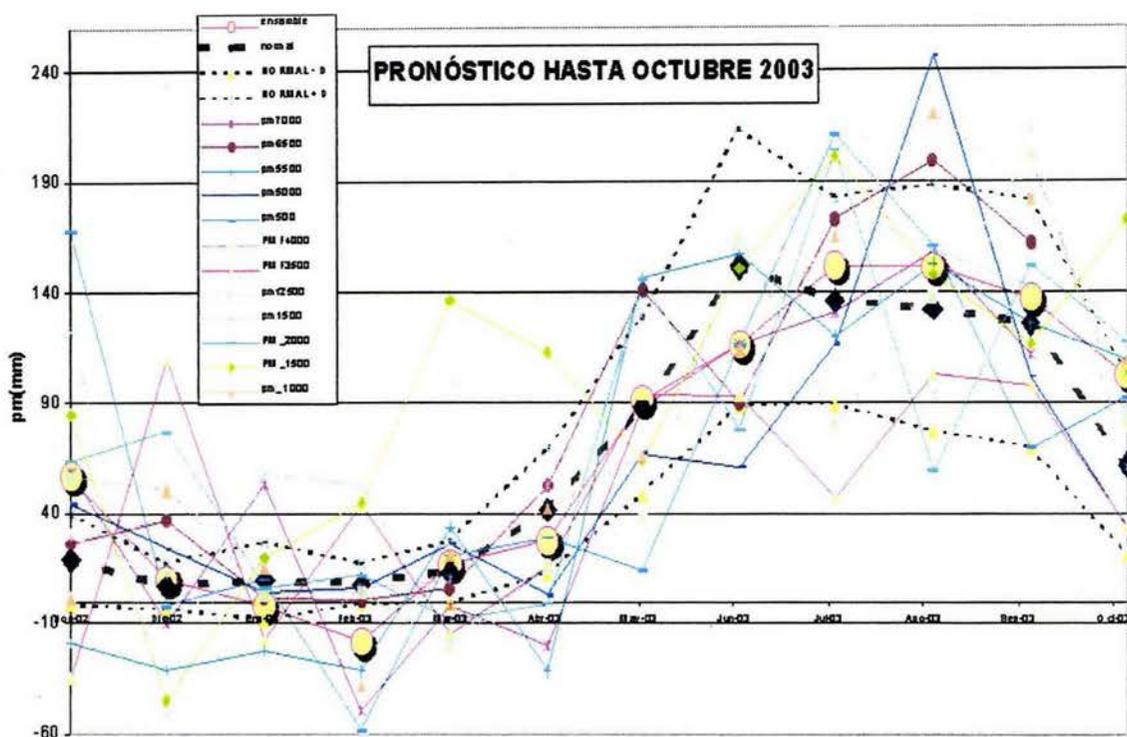
Durante la etapa de aprendizaje la red ajusta sus parámetros de acuerdo a los algoritmos descritos (capítulo 5), generando una curva que comparada con nuestra serie, minimiza el valor de error (SME).

**Fig. (6.4) Curva generada por la red (aprendizaje), datos reales hasta Dic. 2001**



### 6.3 Condiciones iniciales - ensamble

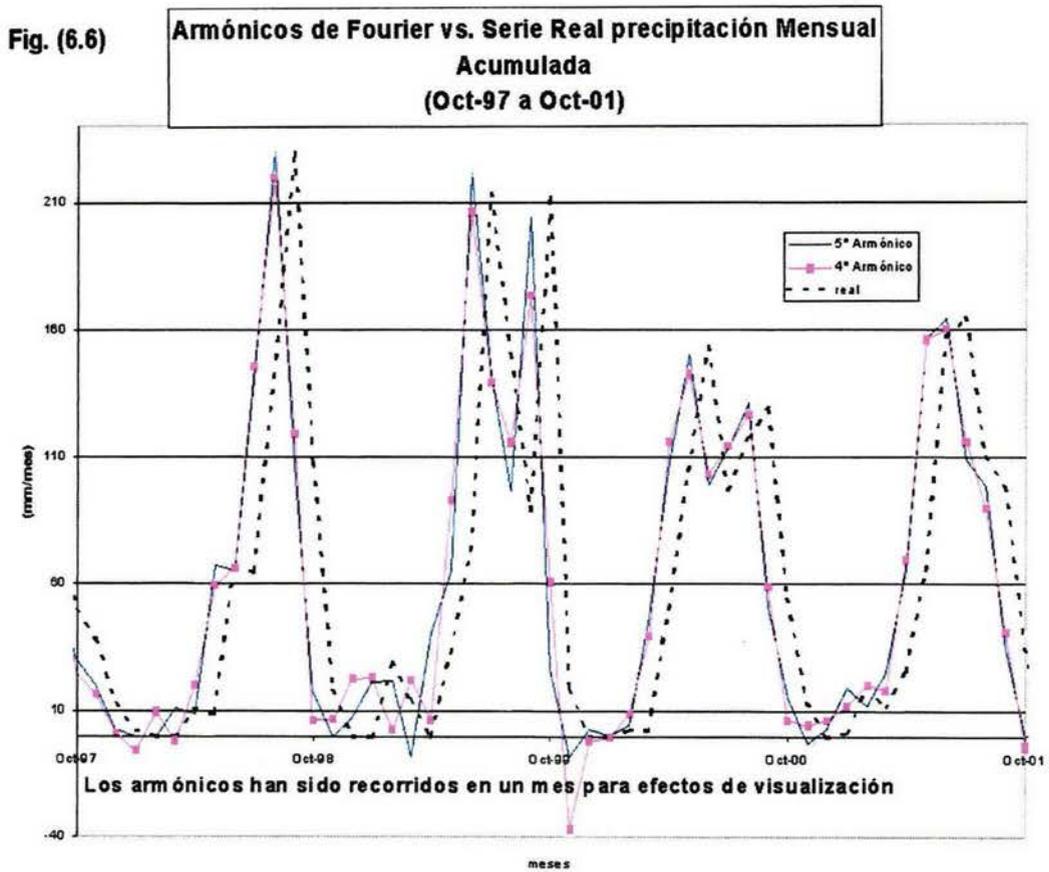
Se modificó el programa de Wang [37], de forma que nos permitiera variar la semilla aleatoria inicial y así poder originar diferentes condiciones de inicio para nuestro sistema, generando para cada condición una curva, se seleccionaron 20 valores entre -7000 y 7000 para la semilla aleatoria inicial, de esa manera se obtiene un ensamble de curvas solución como el mostrado.



**Fig. (6.5) Ensamble de curvas solución, cada curva es generada con un valor de semilla aleatoria diferente, la red se alimenta con datos hasta Octubre 2002, la red genera pronóstico hasta Octubre 2003. La línea azul es la observada. La línea con círculos corresponde a la solución resultante, la línea punteada gruesa representa el promedio histórico (normal).**

### 6.4 Desarrollo en Armónicos

Considerando la complejidad de la serie, extrajimos las series de Fourier en Armónicos para cada año, hasta el cuarto y quinto armónicos, alimentando la red también con estas series, con la esperanza de mejorar el desempeño de la red.(fig. 6.6).

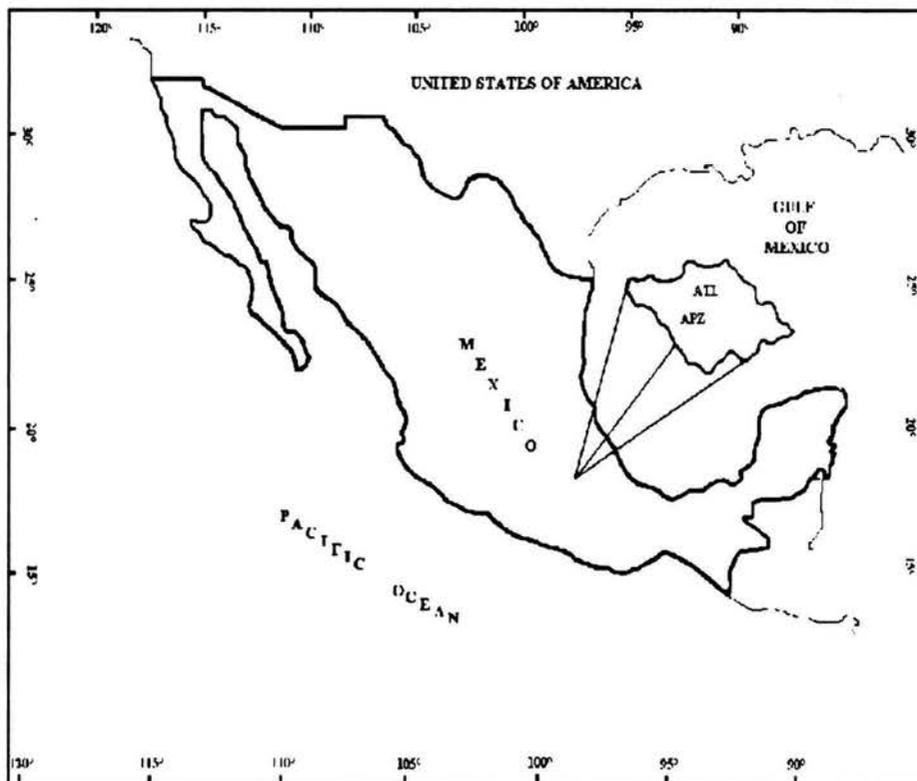


### 6.4.1 Pronóstico Extendido

#### Series de tiempo estudiadas

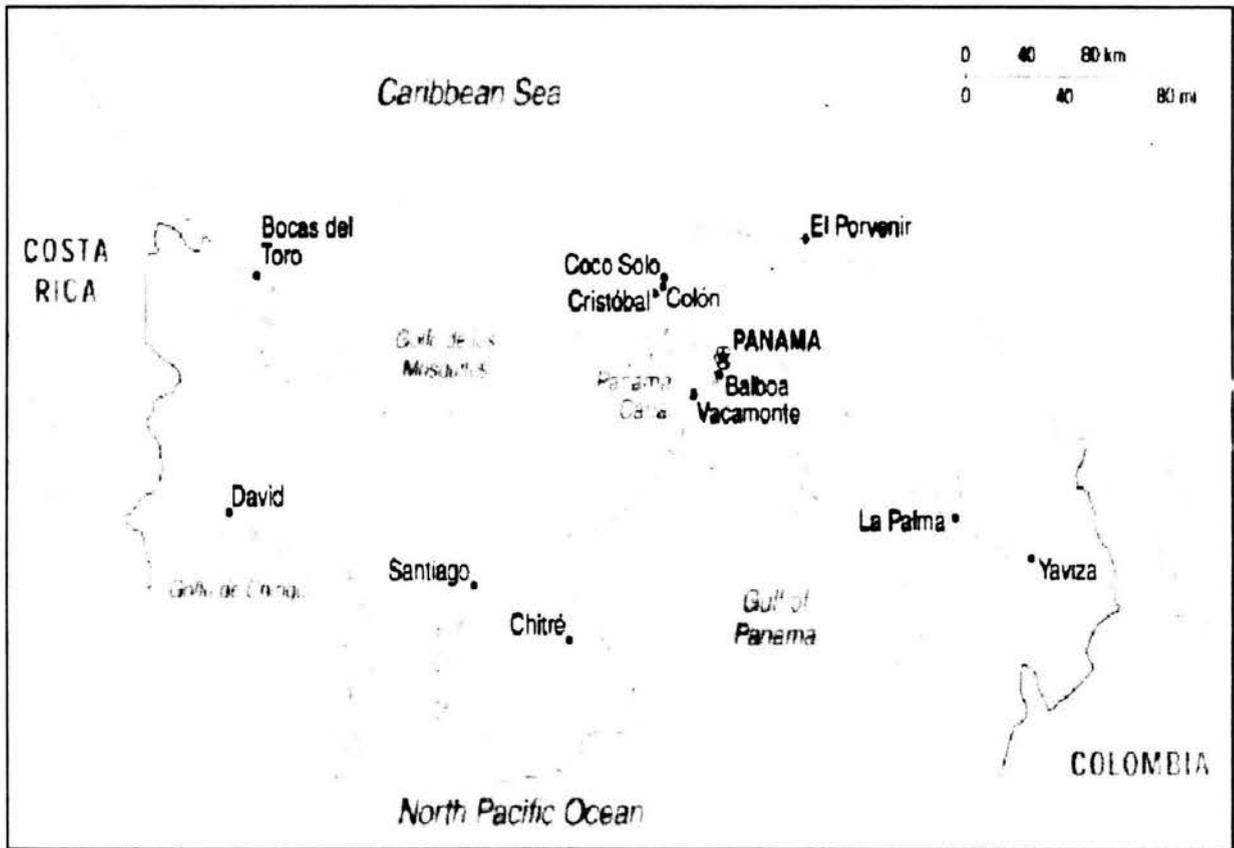
A fin de contrastar resultados, se aplicó la metodología a dos series de tiempo: Apizaco, Tlaxcala y Balboa, Panamá. Las condiciones geográficas son completamente diferentes para ambas poblaciones, Apizaco se encuentra rodeado de montañas a una altura aproximada de 2,000 metros sobre el nivel del mar, a 19° de latitud norte y 97° de longitud oeste.

Fig. (6.7) Ubicación geográfica Apizaco, Tlaxcala



En tanto que Balboa se encuentra al nivel del mar, sin ningún cuerpo montañoso cercano, a  $8^{\circ}$  de latitud norte y  $78^{\circ}$  longitud oeste.

Fig. (6.8) **Ubicación geográfica Balboa, Panamá**



La tabla (6.1), nos presenta las diferencias descriptivas que presentan ambas series.

**Tabla (6.1) Comparativo características series en estudio**

	serie de datos	
	Apizaco	Balboa
período de tiempo	Enero 1942 a Diciembre 02	Enero 1906 a Diciembre 00
número de datos	720 puntos	1,140 puntos
rango precipitación acumulada mensual	0 a 323 mm/mes	0 a 23 mm/mes
media	67 mm/mes	6 mm/mes
varianza	4,443 mm/mes	19 mm/mes
desviación estándar	176 mm/mes	4 mm/mes

Una vez que se aplican todos los procedimientos descritos (capítulo 5) a las series en estudio, la tabla (6.2) nos muestra los resultados obtenidos.

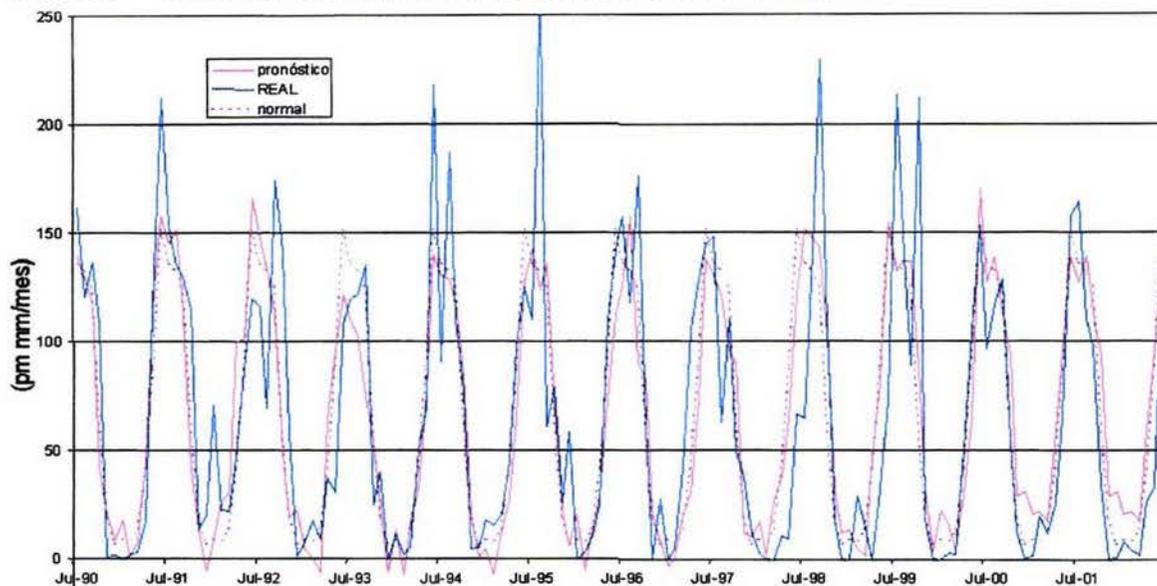
**Tabla (6.2) Comparativo resultados al aplicar la metodología**

	característica	serie de datos		unidades
		Apizaco	Balboa	
<b>Espectro de Potencias (análisis de Fourier)</b>	componentes principales periodo asociado (meses)	12, 3, 6, 48, 4	12, 6, 3, 4, 18	meses
<b>Ondeletas (wavelets)</b>	Máxima varianza asociada	12, 48	12, 6	meses
<b>Información Mutua</b>	tiempo óptimo de retraso	3	3	meses
	dimensión de Correlación	3	3	
	dimensión envolvente	7	7	
	límite predictibilidad	4	14	meses
<b>red neuronal</b>	pronóstico extendido	6	12	meses
	capas	3	3	
	nodos x capa	15-25-1	11-22-1	
	vectores entrada	40-40-1	9-10-1	
	ventana análisis	Jul-90 a Jun-02	Ene-90 a Dic-00	

## Serie de tiempo Precipitación Mensual acumulada Apizaco, Tlaxcala 1943-2002

Se utilizó una ventana de tiempo de 12 años (Julio 1990 a Junio 2002), para validar el desempeño de la red, la mecánica es la siguiente: se corta la serie hasta un cierto punto, (por ejemplo Dic. 1990), la red se alimenta con la serie cortada y una vez alcanzada la cota (SME) (aprendizaje), se le pide un pronóstico para los siguientes doce puntos (Enero 1991 a Dic. 1991), generando un ensamble de curvas solución (aproximadamente 20 soluciones con diferentes condiciones iniciales).

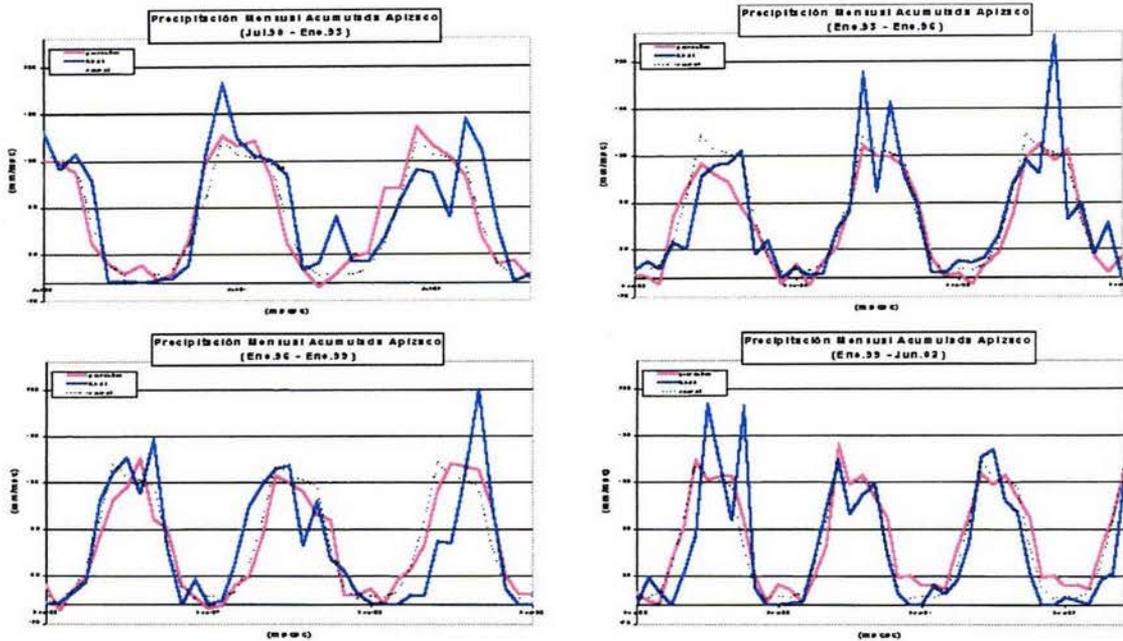
**Fig. (6.9) Serie reconstruida vs. Serie real (observada) Jul-90 a Jun-02**



**Desempeño de la red neuronal en la ventana de prueba, la línea azul, corresponde a los valores observados, la línea roja corresponde al pronóstico creado por el ensamble de curvas solución, la línea punteada es la climatología (valores promedio ene-43 a dic. 2002)**

Para facilitar su visualización seccionamos la serie en 4 partes.

**Fig. (6.9.a) Serie reconstruida vs. serie real, Apizaco.**



(Se ha agrupado en intervalos de 3 años aproximadamente, para facilitar su visualización)

El procedimiento para alimentar la red neuronal, se realizó cortando la serie cada 6 meses: (diciembre 90, junio 91, diciembre 91 hasta diciembre 2002) comparando el pronóstico obtenido contra la serie real, (promedio mensual desde Enero 1942 hasta Diciembre 2002), considerando solamente los primeros 6 meses para cada pronóstico.

Normalmente las redes efectúan pronóstico extendiendo solamente un punto, se incorpora este punto a la serie y se realiza pronóstico para el siguiente. En nuestro caso tenemos por una parte el límite de predictibilidad obtenido para la serie de precipitación en Apizaco: 4.48, ver sección (5.2.3); y por otro el productor agrícola demanda información preferentemente del ciclo completo (12 puntos) para que le

sea redituable.

¿Qué tan buenos son estos resultados?. Cualitativamente la serie reconstruida reproduce correctamente la intermitencia observada, recupera, además, en forma aproximada las fecha de inicio y término de la temporada de lluvia.

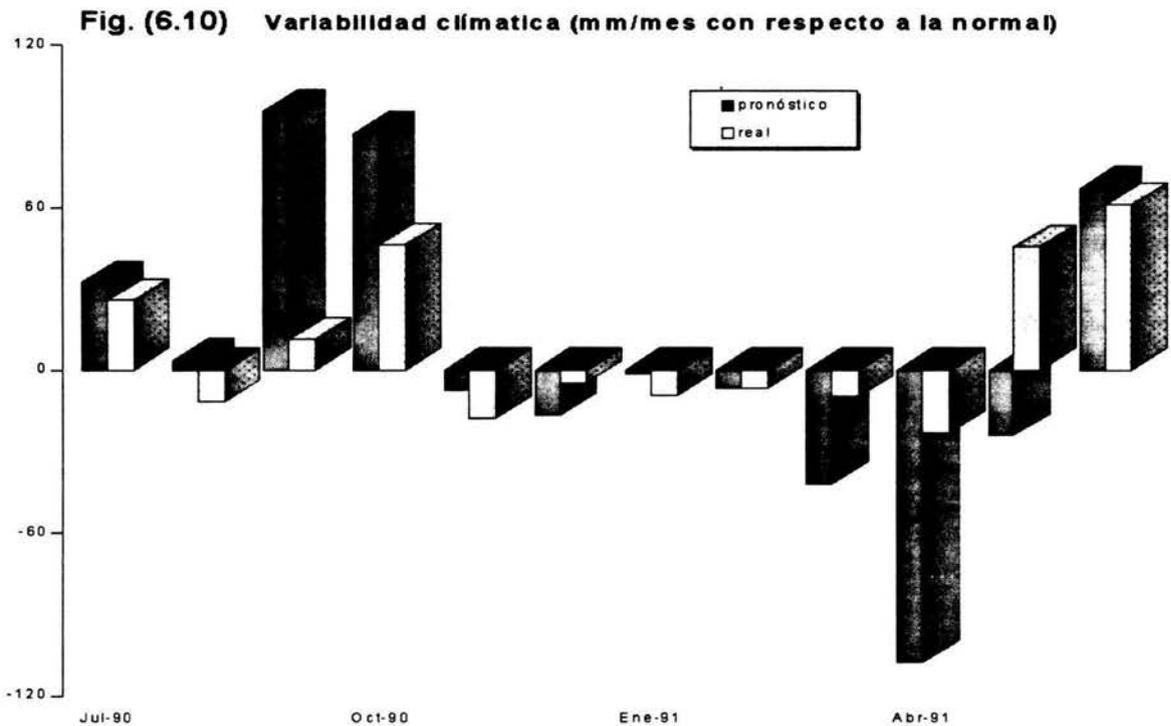
No obstante sí deseamos evaluar a la luz de una medida estrictamente cuantitativa como sería la fórmula de error estimado:

$$E = \left| \frac{(X_{real} - X_{obs})}{X_{real}} \right| \quad (6.1)$$

Por un lado tenemos el inconveniente de no poder evaluar aquellos puntos donde el valor real sea muy cercano a 0 y por otro lado obtenemos un rango de errores que va desde el 1% hasta valores superiores al 5,000% para algunos meses.

Al productor le es más necesario anticipar el signo de las anomalías de la precipitación, que cantidades aproximadas de precipitación esperada, esto es, el productor requiere conocer sí va a contar con más o menos cantidad de lluvia con respecto a la normal, (mes húmedo o mes seco).

Lo anterior nos da la pauta para evaluar nuestros resultados en términos de la variabilidad climática, para ello, en cada mes dentro del intervalo de tiempo, se considera la variabilidad climática para la serie reconstruida (valor pronóstico - valor normal) y la serie observada (valor real - valor normal).



Consideramos como *acierto* el evento en el que coinciden los signos de la variabilidad climática mensual (por ejemplo, Abr-91, en la gráfica anterior) y como *error* el caso contrario (ej. Mar-91).

Finalmente, si agrupamos los aciertos y errores para cada mes dentro de la serie reconstruida, tenemos:

**Tabla (6.3) Habilidad de pronóstico extendido (Apizaco, Tlaxcala).**

meses	Datos			
	Pronóstico	Aciertos	Errores	
Enero	12	9	3	75.0%
Febrero	12	11	1	91.7%
Marzo	12	9	3	75.0%
Abril	12	8	4	66.7%
Mayo	12	7	5	58.3%
Junio	12	10	2	83.3%
Julio	12	11	1	91.7%
Agosto	12	7	5	58.3%
Septiembre	12	9	3	75.0%
Octubre	12	8	4	66.7%
Noviembre	12	8	4	66.7%
Diciembre	12	10	2	83.3%
<b>Total general</b>	<b>144</b>	<b>107</b>	<b>37</b>	<b>74.3%</b>

Evaluando el procedimiento para los armónicos cuarto y quinto obtenemos unas tasas de acierto del 73% y 76%, respectivamente.

La probabilidad de obtener 107 aciertos en 144 intentos para una distribución normal resulta ser altamente significativa. Efectivamente, utilizando para la media  $m = np$  la desviación estandar  $\sigma = \sqrt{npq}$ , y  $z = \frac{X-m}{\sigma}$ ; con  $X = 107$ ,  $n = 144$ ,  $p = q = \frac{1}{2}$ , tenemos:  $m = (144)(\frac{1}{2}) = 72$

$$\sigma = \sqrt{(144)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)} = 6$$

$$z = \frac{107-72}{6} = 5.83$$

Como  $5.83 > 3.84$  (3.84 corresponde al valor de  $\chi^2$  al 5%), entonces  $P(X) < 0.05$ .

Tabla (6.4) Tabla de Contingencia en función de la calidad de mes para la serie reconstruida. (Pronóstico extendido - Apizaco, Tlaxcala).

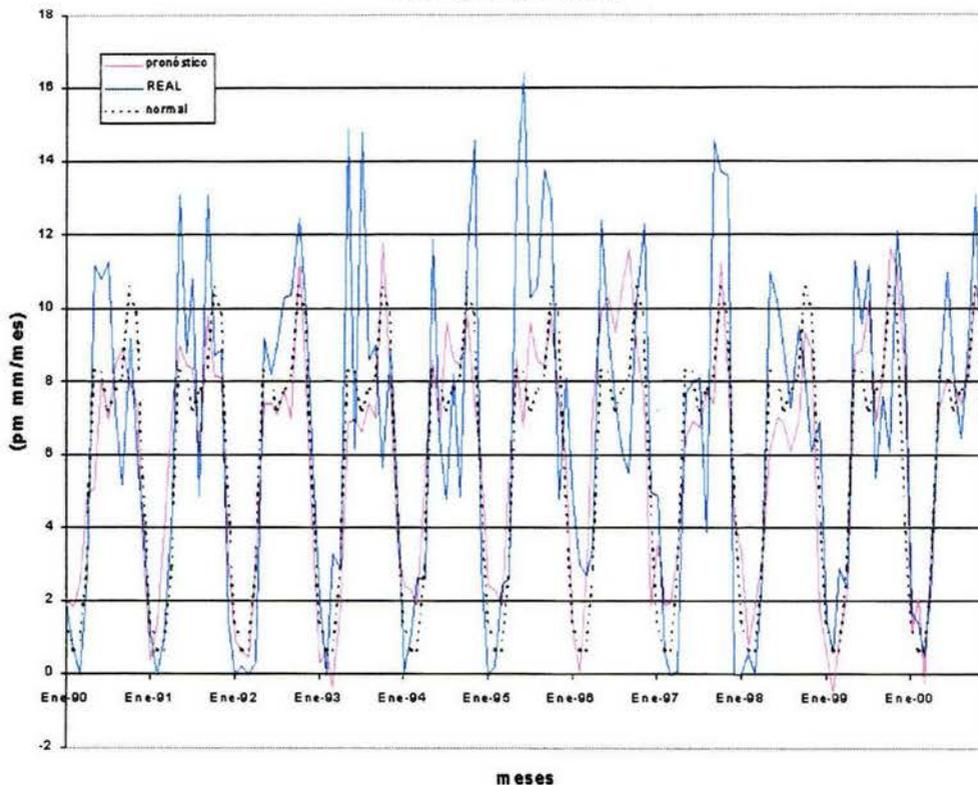
<i>mes</i>	<i>real</i>	<i>pronóstico (aciertos)</i>
<i>húmedo</i>	59	44
<i>seco</i>	85	63
	144	107

Aplicando la prueba  $\chi^2$  a la tabla de contingencia mostrada, concluimos que la habilidad del pronóstico es independiente de la calidad del mes con un nivel de significancia del 5%.

## Serie de tiempo Precipitación Mensual acumulada Balboa, Panamá 1906-2000

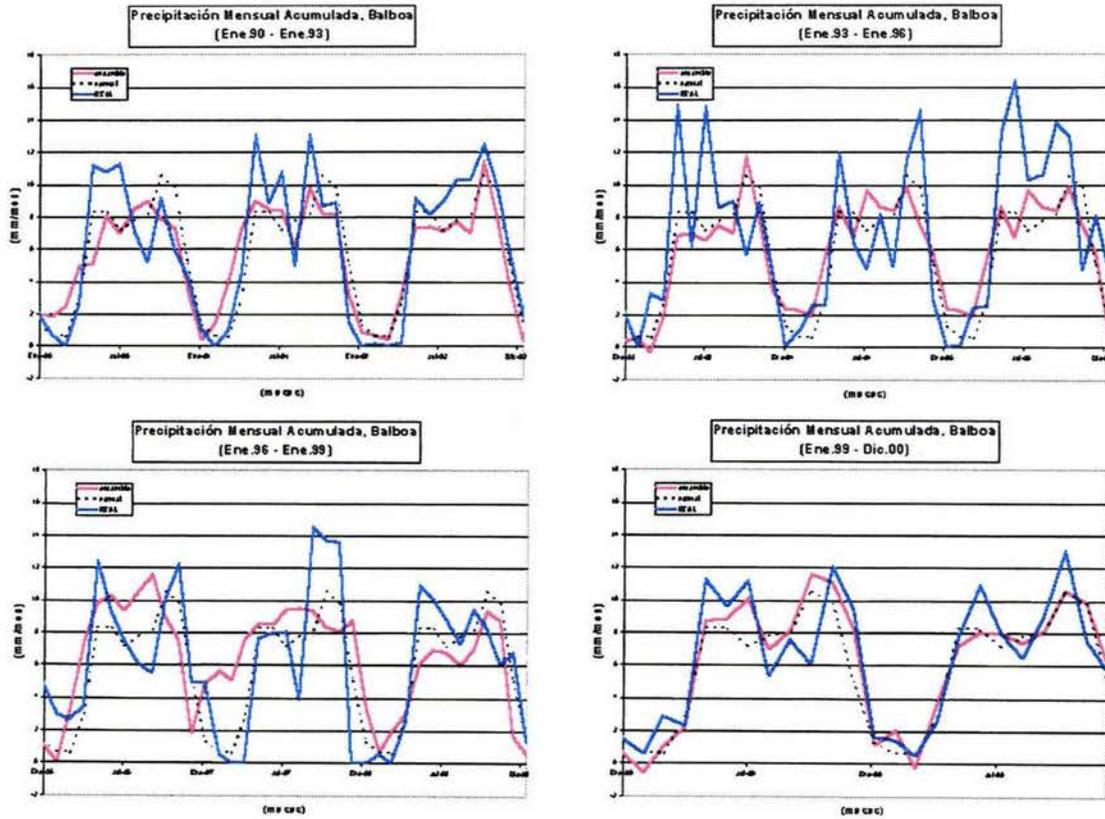
En este caso, se estudió una ventana de tiempo de 11 años, (Enero 1990 a Diciembre 2000); para validar el desempeño de la red, el estudio se realizó cortando la serie cada 12 meses: (diciembre 90, diciembre 91, diciembre 92 hasta diciembre 2000) comparando el pronóstico obtenido contra la serie real, gracias al límite de predictibilidad para esta serie (14 puntos), extendimos el pronóstico considerando 12 meses.

**Fig. (6.11) Serie reconstruida vs. Serie real (observada). Ene-90 a Dic-00. Balboa, Panama**



Para facilitar su visualización seccionamos la serie en 4 partes.

Fig. (6.11.a) Serie reconstruida vs. serie real, Balboa.



En términos de la variabilidad climática, si agrupamos los aciertos y errores para cada mes dentro de la serie reconstruida, tenemos:

**Tabla (6.5) Habilidad de pronóstico extendido (Balboa, Panamá).**

Datos

meses	Pronóstico	Aciertos	Errores	
Enero	11	9	2	81.8%
Febrero	11	7	4	63.6%
Marzo	11	8	3	72.7%
Abril	11	7	4	63.6%
Mayo	11	9	2	81.8%
Junio	11	8	3	72.7%
Julio	11	10	1	90.9%
Agosto	11	9	2	81.8%
Septiembre	11	11		100.0%
Octubre	11	8	3	72.7%
Noviembre	11	9	2	81.8%
Diciembre	11	7	4	63.6%
<b>Total general</b>	<b>132</b>	<b>102</b>	<b>30</b>	<b>77.3%</b>

Nuevamente obtenemos una probabilidad altamente significativa, bajo el supuesto de tener una distribución normal, además, también se verifica que la habilidad del pronóstico es independiente de la calidad del mes con un nivel de significancia del 5%.

## **Ensamble de Análogos**

El proyecto de investigación: **Utilización de Pronósticos Climáticos para Actividades Agrícolas en el estado de Tlaxcala**, fue implementado en 1997, con el objetivo de producir pronósticos con un enfoque práctico para la toma de decisiones, para ello se contó con estrechos contactos con los productores, el método de ensamble de análogos fue aplicado a datos históricos, aquí reproducimos los resultados presentados para la región de Apizaco, comparando el período coincidente [44].

Cabe señalar que el pronóstico de ensambles coincidentemente se expresa en términos de la variabilidad climática.

Tabla (6.6)

**COMPARATIVO PRONÓSTICO DE PRECIPITACIÓN (SIGNO)**

año	Ensamble Análogos		Red Neuronal	
	1er. Sem	2do. Sem	1er. Sem	2do. Sem
1998	75%	62%	67%	83%
1999	66%	42%	100%	83%
2000	83%	75%	67%	100%
2001	50%	41%	83%	100%
2002	100%	88%	83%	100%

<b>75%</b>	<b>62%</b>	<b>80%</b>	<b>93%</b>
------------	------------	------------	------------

**\* Porcentaje de aciertos durante el período**

### Cantidad de lluvia anual

Sí observamos los gráficos de desempeño figuras (6.9 y 6.11), año con año, podemos pensar que los meses en los que existe un déficit de lluvia al comparar la curva pronosticada con la curva real, se compensan en con aquellos que tienen un superávit, para verificar esto, integramos la cantidad de lluvia por año y obtenemos la siguiente tabla:

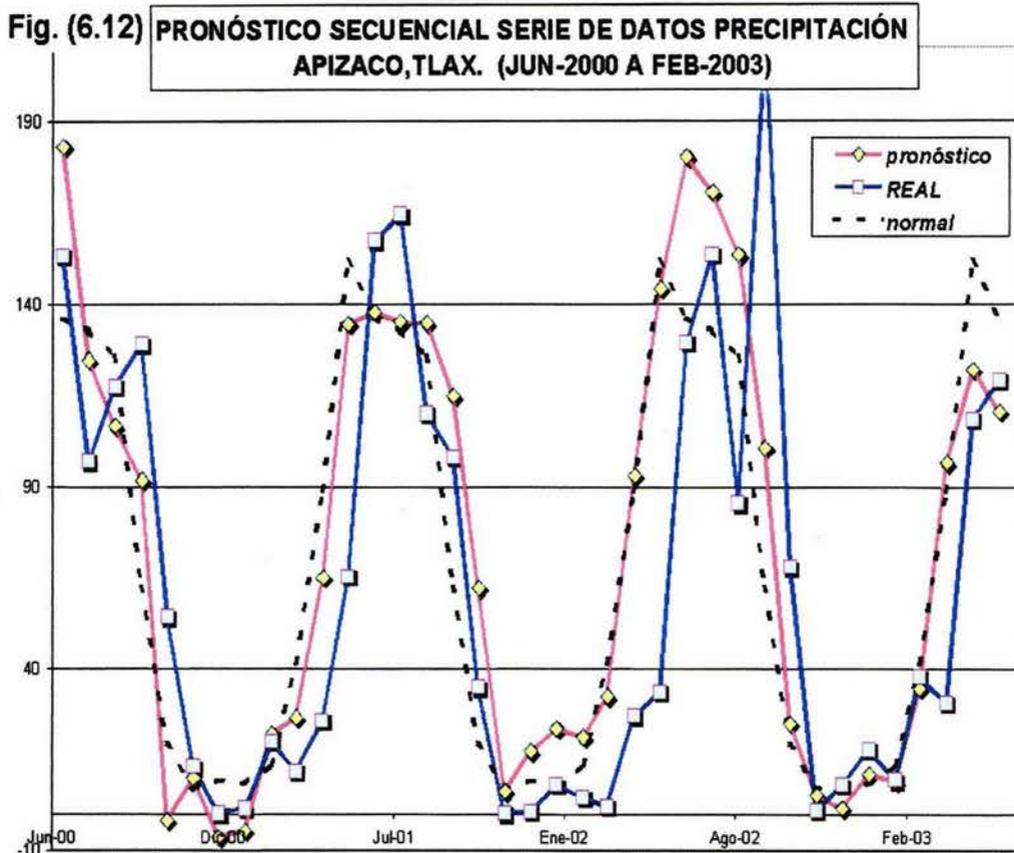
**Tabla (6.7) Comparativo cantidad de lluvia anual pronosticada contra la cantidad de lluvia anual real para ambas series.**

<b>Apizaco, Tlaxcala</b>			
<b>precipitación acumulada anual (mm/año)</b>			
<b>año</b>	<b>pronóstico</b>	<b>real</b>	<b>error</b>
1991	817.0	936.4	12.7%
1992	919.3	930.6	1.2%
1993	644.3	653.2	1.4%
1994	760.6	835.2	8.9%
1995	731.9	913.9	19.9%
1996	731.9	832.4	12.1%
1997	720.5	848.4	15.1%
1998	860.2	649.2	32.5%
1999	836.5	831.4	0.6%
2000	840.6	723.1	16.2%
2001	888.3	689.13	28.9%
<b>Error promedio:</b>			<b>13.6%</b>

<b>Balboa, Panamá</b>			
<b>precipitación acumulada anual (mm/año)</b>			
<b>año</b>	<b>pronóstico</b>	<b>real</b>	<b>error</b>
1990	67.6	70.3	3.9%
1991	74.5	76.5	2.7%
1992	65.1	75.7	13.9%
1993	60.7	81.1	25.1%
1994	77.6	71.9	7.9%
1995	77.6	95.5	18.8%
1996	81.9	82.3	0.5%
1997	94.2	75.2	25.2%
1998	62.4	71.4	12.6%
1999	77.3	80	3.4%
2000	72.6	75.2	3.5%
<b>Error promedio:</b>			<b>10.7%</b>

6.4.2 Pronóstico Secuencial

En otro ejercicio se generaron series mensuales desde Junio de 1996 hasta diciembre de 2002, obteniendo un pronóstico para cada serie de los seis meses subsecuentes, el pronóstico para cada mes resulta una media de los pronósticos obtenidos para cada mes en los seis meses anteriores (para Diciembre 96, promediamos: el sexto punto pronosticado con la serie cortada a junio de 96, el quinto punto pronosticado con la serie cortada a julio de 96, el cuarto punto pronosticado con la serie correspondiente a agosto de 96 y así hasta el primer punto de la serie cortada a Noviembre 96).



Considerando nuevamente la variabilidad climática, como parámetro de evaluación, en este experimento se alcanzó una habilidad de pronóstico del 76%.

**Tabla (6.8) HABILIDAD DE PRONÓSTICO SECUENCIAL**

meses	Datos			
	pronóstico	aciertos	errores	
Enero	3	2	1	<b>67%</b>
Febrero	3	2	1	<b>67%</b>
Marzo	3	3	0	<b>100%</b>
Abril	3	1	2	<b>33%</b>
Mayo	3	3	0	<b>100%</b>
Junio	4	3	1	<b>75%</b>
Julio	3	3	0	<b>100%</b>
Agosto	3	1	2	<b>33%</b>
Septiembre	3	3	0	<b>100%</b>
Octubre	3	2	1	<b>67%</b>
Noviembre	3	3	0	<b>100%</b>
Diciembre	3	2	1	<b>67%</b>
<b>Total general</b>	<b>37</b>	<b>28</b>	<b>9</b>	<b>76%</b>

Este resultado también es altamente significativo, atendiendo a la calidad del mes tenemos la siguiente tabla de contingencia:

Tabla (6.9)

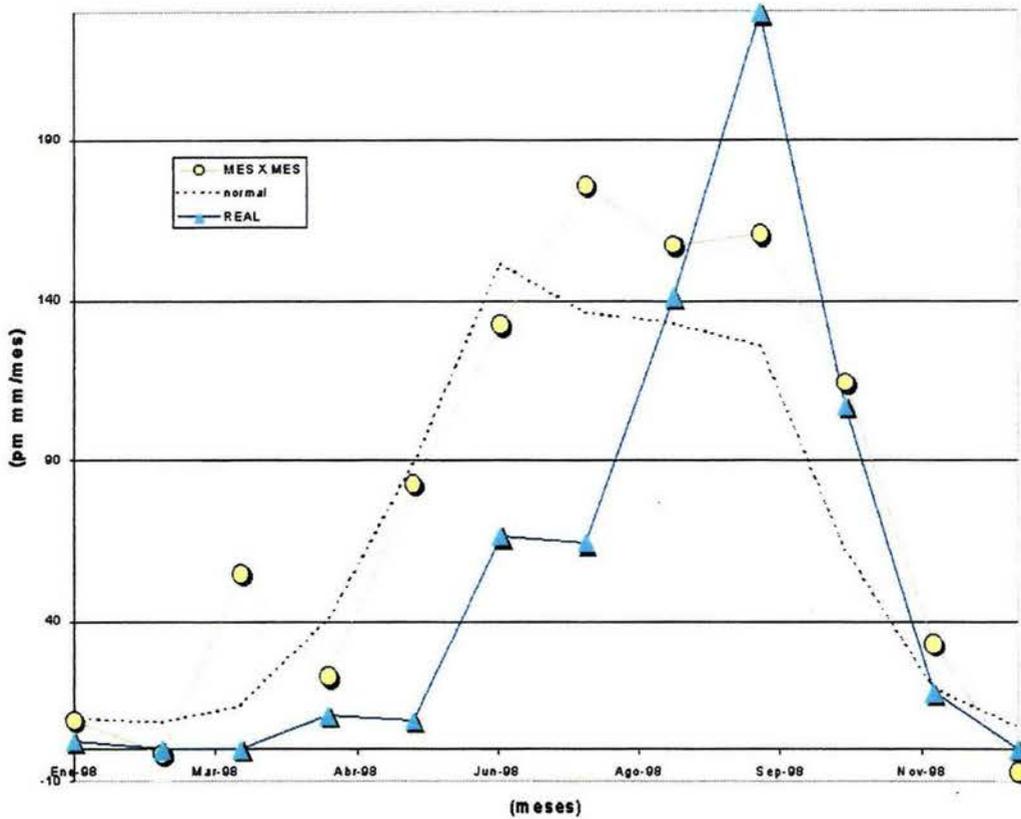
**Tabla de contingencia en función de la calidad de mes para la serie reconstruida (Pronóstico Secuencial)**

<b>mes</b>	<b>real</b>	<b>pronóstico (aciertos)</b>
<b>húmedo</b>	24	16
<b>seco</b>	13	12
	<b>37</b>	<b>28</b>

6.4.3 Pronóstico Tradicional

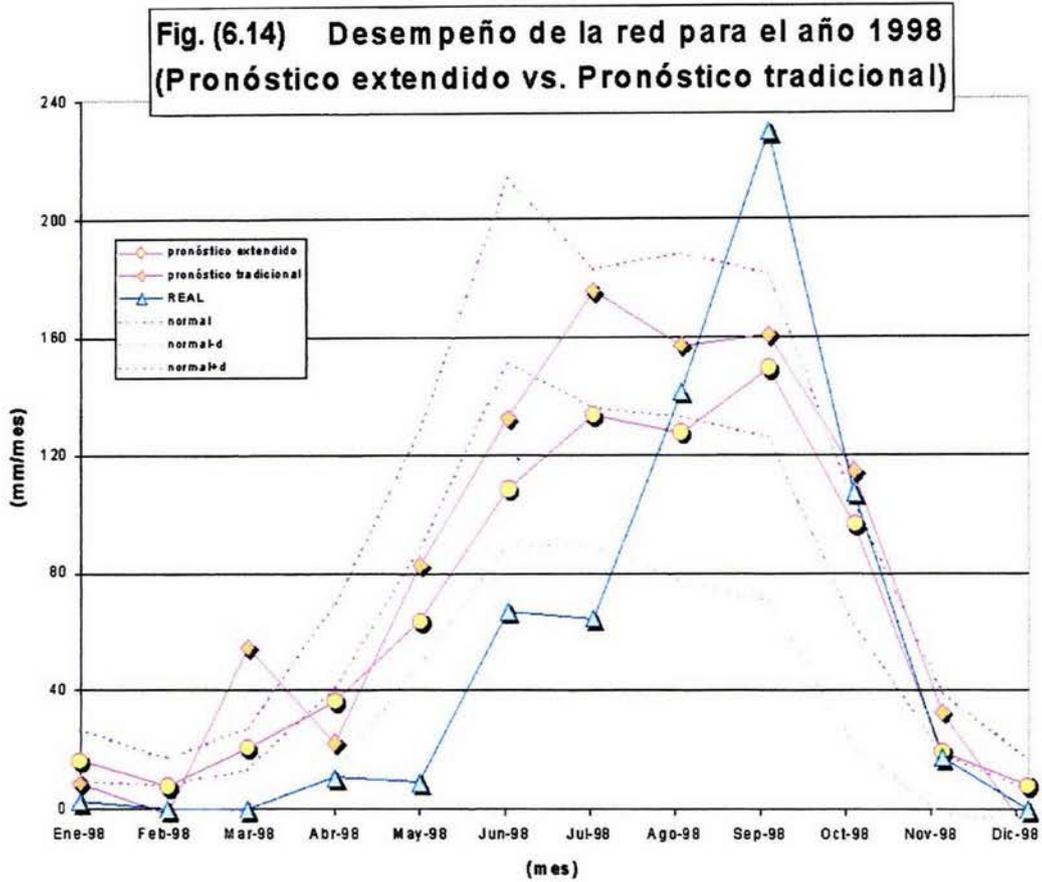
Se realizó un experimento para el año de 1998, toda vez que en el mes de Septiembre, se presenta un máximo que no fué identificado en forma clara por las experimentos anteriores, el experimento consistió en alimentar la red con las series de datos recortadas para cada mes, (dic. 97 a nov. 98) y pedirle a la red sólo el mes siguiente para cada caso.

Fig. (6.13) Pronóstico Tradicional - 1998



En términos de la variabilidad climática, obtenemos 9 aciertos en 12 intentos, esto es, un 75%.

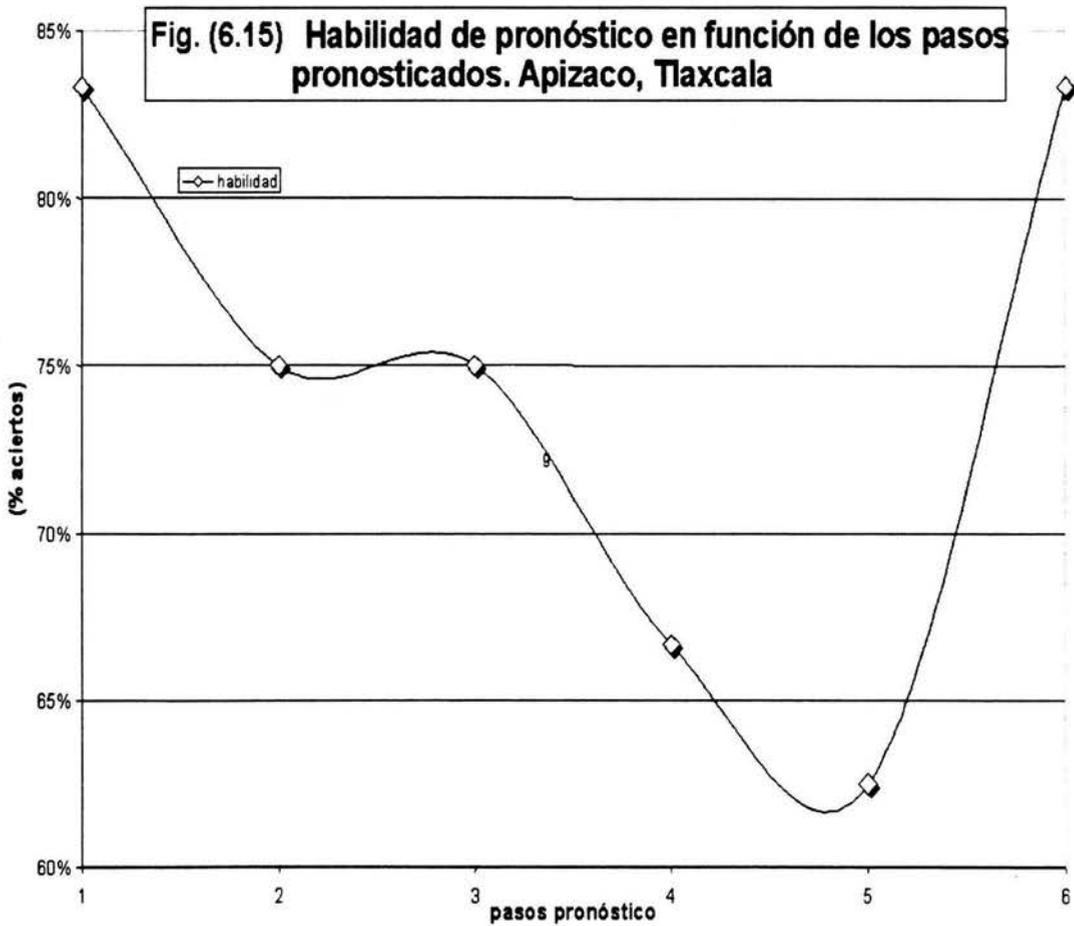
Comparando los diferentes pronósticos que hemos obtenido para el año 1998.



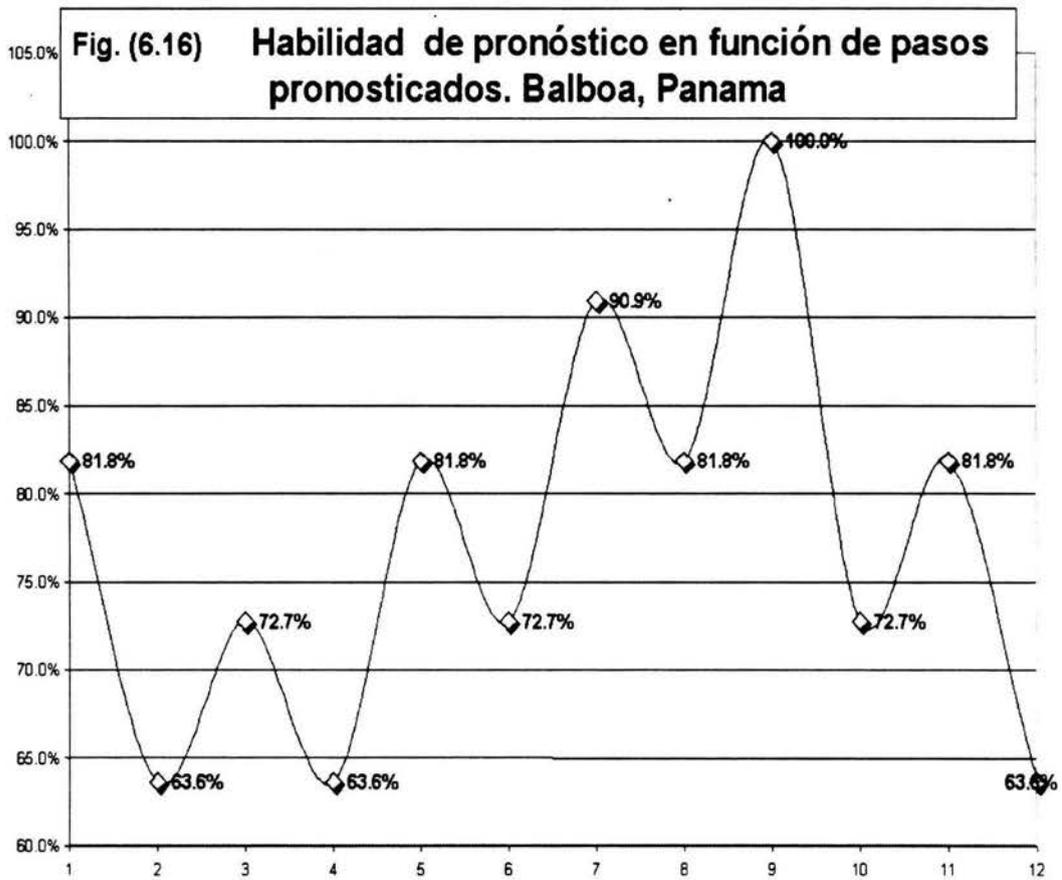
Como se puede apreciar en la figura (6.14), el año de 1998 fue atípico (basta con observar los puntos que no inciden dentro de la franja de lo "posible"), el pronóstico tradicional se aproxima mejor a la curva real en los puntos extremos: Abr-98 y Sep-98.

### 6.4.4 Dinámica de la Habilidad de Pronóstico

Dado que estamos extendiendo el pronóstico de la red a más de un punto, nos interesa conocer como evoluciona la habilidad de la red para pronosticar, para ello, en el caso de Apizaco, Tlaxcala, donde en realidad estamos expandiendo el pronóstico en 6 meses, esto es, cortamos las series en los Diciembres, para pronosticar el primer semestre de cada año, y cortamos las series en los Junios, para pronosticar el segundo semestre de cada año, podemos asociar a cada mes el número de pasos de pronóstico asociado, y generar así una curva de desempeño en función del número de puntos de pronóstico.



Para la serie de Balboa, dado que solicitamos pronóstico extendido por 12 meses, la gráfica de habilidad en función de puntos de pronóstico se obtiene directamente de la tabla (6.5).



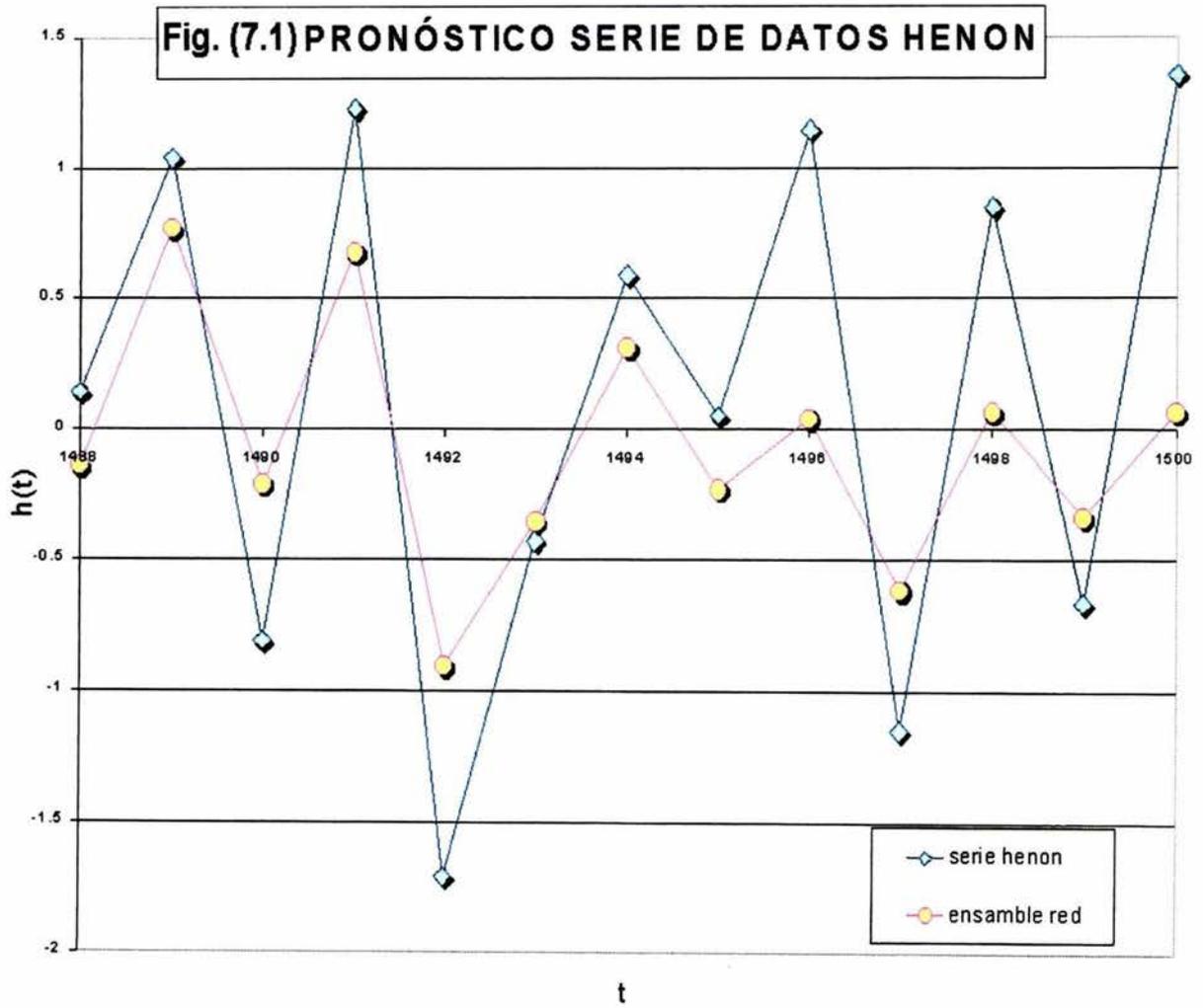
# Capítulo 7

## DISCUSIÓN DE RESULTADOS

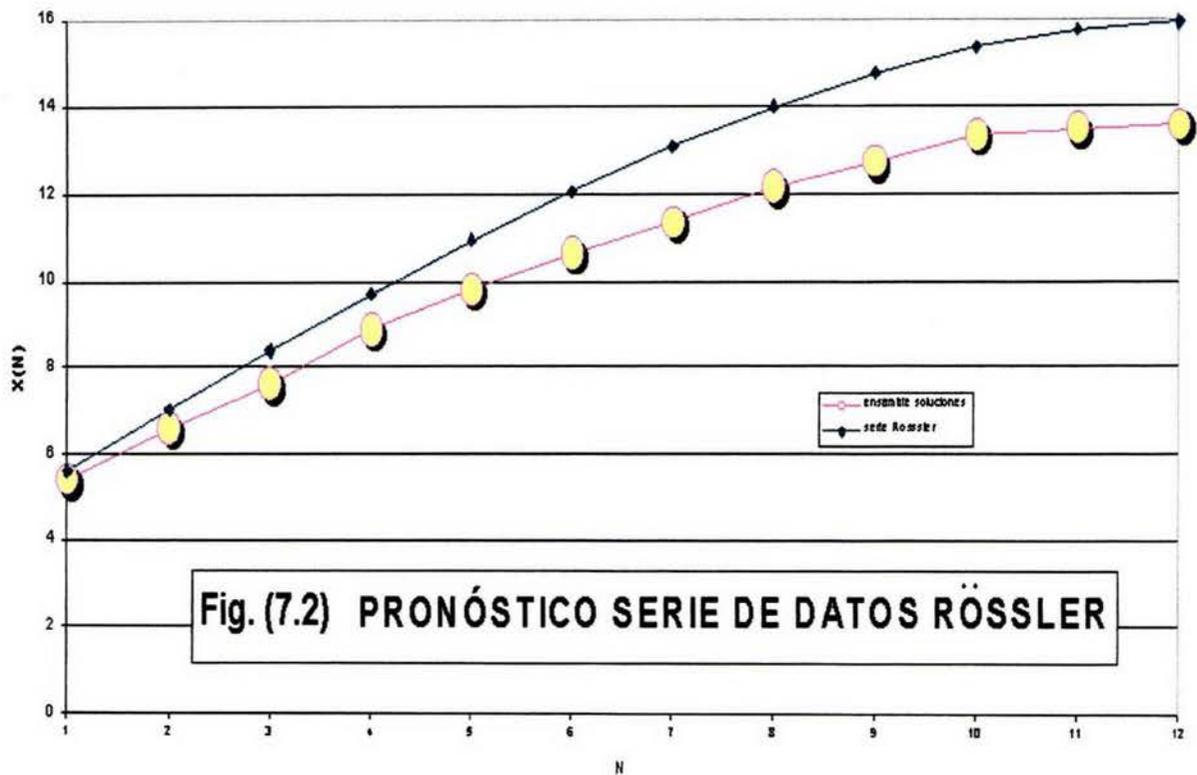
### 7.1 Habilidad de Pronóstico

Es claro a partir de las diferentes pruebas realizadas, que existe un límite para la habilidad de pronóstico alrededor del 75%. ¿Este tope a la capacidad de predicción es intrínseco a la red neuronal o tiene su origen en la serie temporal en estudio?, al parecer, dado que estudiamos dos series temporales sustancialmente diferentes (Apizaco y Balboa), podríamos aducir que es inherente a la red; ¿Qué sucede si realizamos pronóstico sobre una serie pura?, tomamos dos series de tiempo puras: la serie de Henon y la serie de Rössler, repetimos el proceso desarrollado con las series de precipitación, esto es, cortamos en un momento dado, alimentamos la red para diferentes semillas con las series transformadas correspondientes y una vez cubierta la etapa de aprendizaje, solicitamos a la red pronóstico sobre los siguientes doce puntos realizando un ensamble de curvas solución.

Resultado Serie de datos Henon



## Resultado Serie de datos Rössler



En ambos casos encontramos, realizando una prueba  $\chi^2$  para el pronóstico directo, que las diferencias no son significativas. Luego entonces, debemos concluir que las series de datos de precipitación son más complejas y más difíciles de capturar por la red neuronal; podríamos incluso suponer que existen periodicidades involucradas en el tiempo - clima, que por la extensión de la serie de datos en estudio, no logran ser observadas por la red, (ciclicidades que involucren períodos de más

de 100 años, por ejemplo).

## 7.2 Extensión de la Calidad de Pronóstico en el tiempo

El comportamiento observado para las series puras, apoya la idea de poder extender la habilidad de pronóstico más allá de un punto; en el caso de las series de tiempo de Apizaco y Balboa, el análisis debe efectuarse considerando las tablas 6.3 y 6.5, y las gráficas 6.15 y 6.16; de acuerdo con la tabla 6.3, existen dos puntos máximos en la habilidad de pronóstico (91%), asociados a Febrero y Julio, si recordamos el procedimiento descrito para Apizaco, se realizó pronóstico cada 6 meses, cortando la serie en Diciembre y Junio; esto es, Febrero corresponde al segundo punto pronosticado para las secuencias (Enero - Junio), además Febrero es el mes en el que, para esa latitud, se alcanza el punto de equilibrio entre el Océano y la Atmósfera, estos argumentos pueden explicar el buen desempeño mostrado para Febrero, el caso de Julio se explica si notamos que corresponde al primer punto pronosticado para las secuencias (Julio - Diciembre). Este efecto no es tan claro en la gráfica correspondiente (6.15), ya que en ella, por ejemplo, el valor asociado al segundo punto de pronóstico, corresponde al promedio de la habilidad observada para Febrero y Agosto, es sorprendente la habilidad que se muestra en la gráfica para el sexto paso (83.3%); que corresponde a Junio y Diciembre, dado que el valor es el mismo para ambos meses, consideramos que por un lado, tenemos una muy buena habilidad gracias a que se trata del primer punto de pronóstico, (Junios), y por el otro, tenemos un muy buena habilidad para los diciembres, dado que se caracterizan por tener una variabilidad muy baja.

Para el caso de Balboa, tenemos un comportamiento similar, el ejercicio se realizó cortando la serie en Diciembre y obteniendo la secuencia (Enero - Diciembre) para el siguiente año, si observamos la tabla (6.5) el máximo en la habilidad (100%), se alcanza para Septiembre, que coincidentemente es el mes

en el que, para la latitud geográfica de Balboa, se verifica el punto de equilibrio Océano - Atmósfera, en este caso, la gráfica correspondiente, (6.16) es elocuente. En términos generales podemos concluir que para la red seleccionada, la calidad de pronóstico en función de los pasos, no se demerita muy rápidamente e incluso permite identificar aspectos relevantes de la variabilidad climática asociada a la Serie de Tiempo en estudio.

### 7.3 Cantidad anual de lluvia

Sí consideramos la cantidad anual de lluvia pronosticada, (ver tabla 6.7), encontramos un error absoluto promedio del 13.6% para Apizaco y un 10.7% para Balboa; es importante resaltar la similitud en los resultados, no obstante las diferencias en los rangos de precipitación que existen entre las series. (ver tabla 6.1).

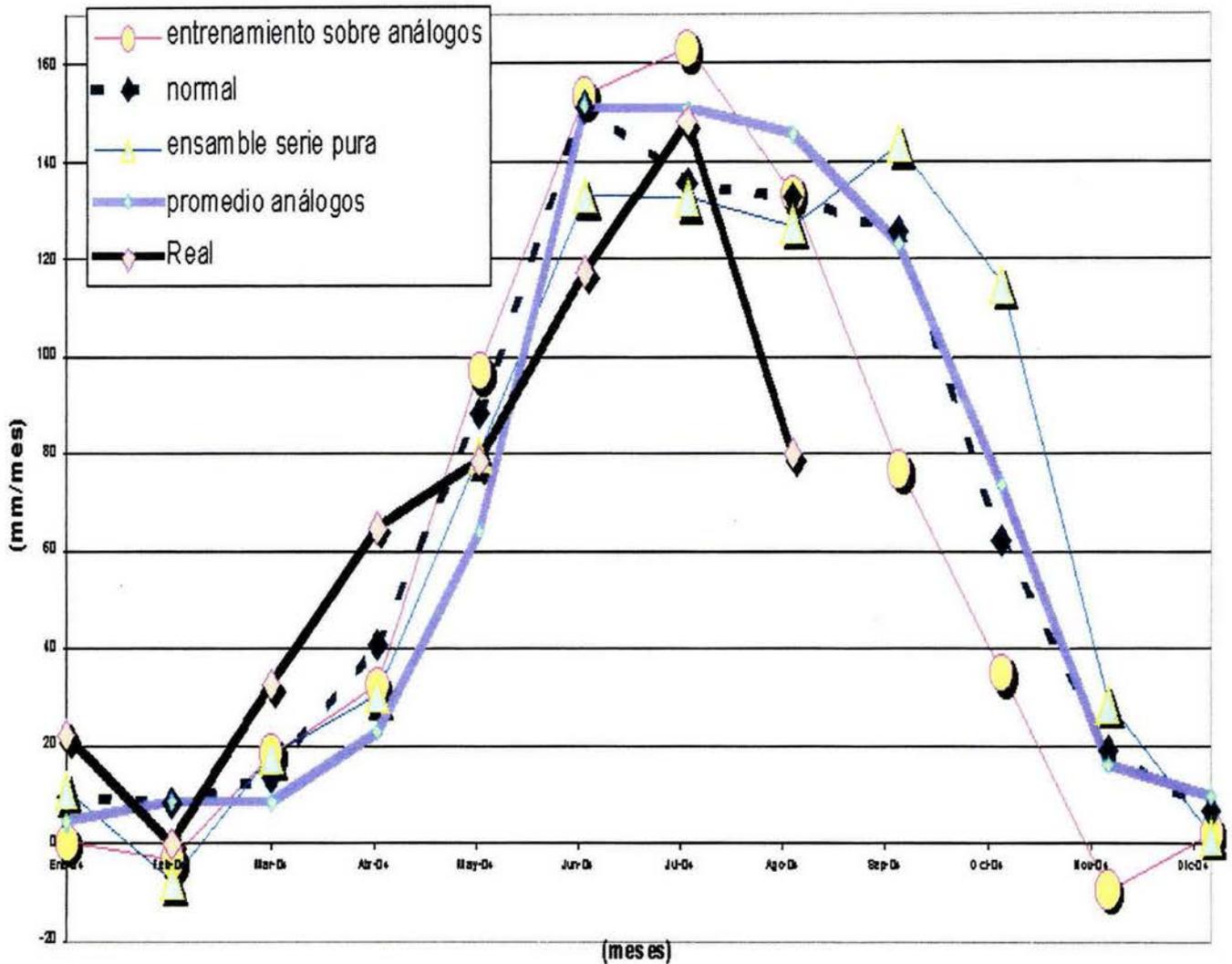
### 7.4 Pronóstico año 2004

En esta ocasión se aplicó la metodología descrita en la sección 6.4.1 a la serie de tiempo de Apizaco Tlaxcala, hasta Diciembre de 2003, solicitándole a la red pronóstico extendido para los siguientes 12 puntos, (para efectos de pronóstico son válidos solamente los primeros 6 puntos), sin embargo, consideramos necesario presentar el año completo.

Adicionalmente, se construyó una serie de tiempo integrada con análogos (aquellos años que se considera contienen condiciones similares a las esperadas para 2004): 1955, 1956, 1963, 1968, 1969, 1971, 1999 y 2000.

La gráfica 7.3 muestra, además de los dos ensambles de soluciones generados por la red, el promedio de los años análogos (pronóstico empírico), y la precipitación presentada (curva real) hasta el mes próximo pasado.

**Fig. (7.1) Pronóstico para el año 2004 y observado (real) hasta el mes de Agosto.**



Siguiendo con la mecánica de trabajo presentada, evaluando el pronóstico en términos de variabilidad (diferencia del valor presentado contra el valor normal), obtenemos la siguiente tabla:

**Tabla (7.1) Evaluación del pronóstico para el año 2004.**

	entrenamiento sobre análogos	real	ensamble serie pura
<b>Ene-04</b>	Seco	Húmedo	Húmedo
<b>Feb-04</b>	Seco	Seco	Seco
<b>Mar-04</b>	Húmedo	Húmedo	Húmedo
<b>Abr-04</b>	Seco	Húmedo	Seco
<b>May-04</b>	Húmedo	Seco	Seco
<b>Jun-04</b>	Húmedo	Seco	Seco
<b>Jul-04</b>	Húmedo	Húmedo	Seco
<b>Ago-04</b>	Húmedo	Seco	Seco

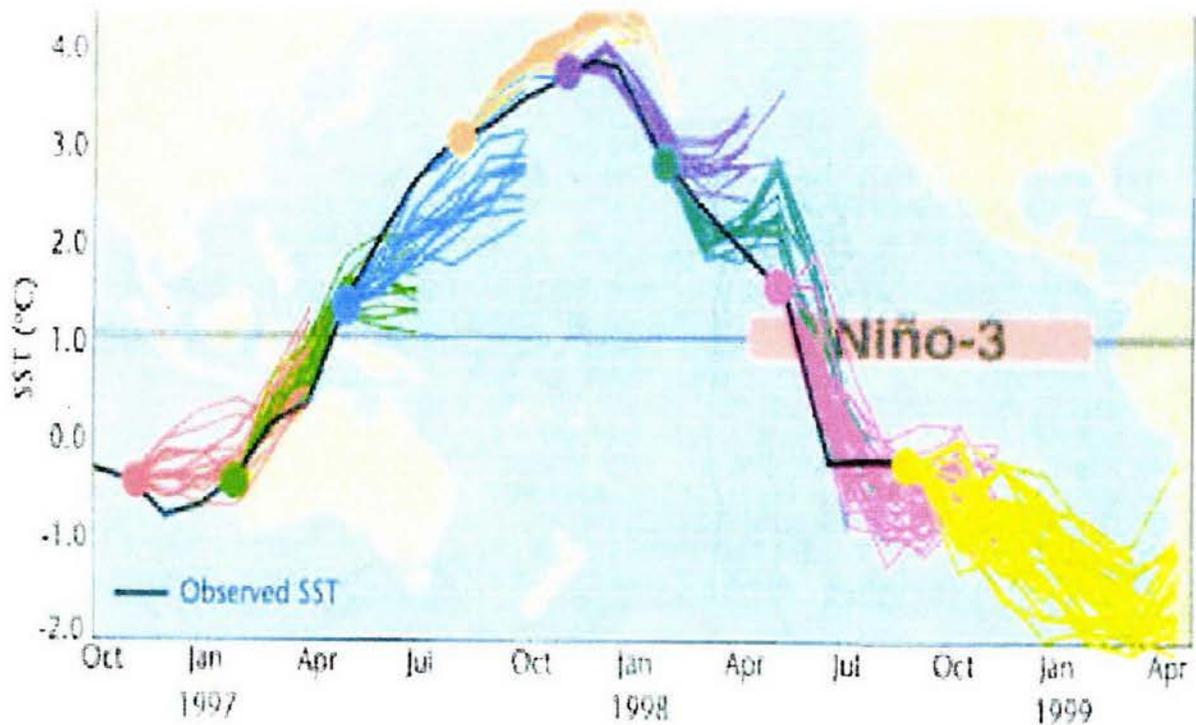
No obstante el bajo desempeño mostrado por el pronóstico generando al alimentar la red con los años análogos, (37.5%); al observar la gráfica 7.3, no podemos considerarlo como un mal pronóstico, ya que prácticamente predice la forma del perfil de precipitación mensual, anticipa la canícula seca que se ha presentado.

Contradictoriamente, el pronóstico sobre la serie pura, apesar de tener una mejor acertabilidad (75.0%), no brinda una información tan rica en términos de la forma.

### 7.5 Modelo Determinístico Europeo

**Fig. (7.2) Pronóstico generado por un modelo determinístico. (European Center Medium Weather Forecast).**

ECMWF Niño-3 SST Forecasts for 6 Month Periods Starting Every 3 Months



La gráfica 7.3 ilustra el pronóstico generado en el ECMWF, empleando un modelo Determinístico, para diferentes fechas de inicio del modelo, y nos permite ponderar varias características que presenta nuestro trabajo:

- 1) El pronóstico se basa en un ensamble de posibles soluciones.
- 2) Extensión del pronóstico en 3 meses, en nuestro caso, hemos extendido el pronóstico en 6 meses para Apizaco y 12 meses para Balboa.
- 3) Desempeño diferenciado de la red de acuerdo al mes de inicio, (los pronósticos generados en Enero y Junio, desarrollan una mejor habilidad que aquellos que se generan en Abril y Octubre).

## **7.6 Aportaciones**

Consideramos conveniente resaltar algunos de los logros obtenidos:

- 1) La generación de una herramienta (laboratorio de análisis de series caóticas), que nos permite extraer información sobre la dinámica de un sistema físico real a partir de una serie de tiempo y realizar predicción conforme a las limitaciones implícitas.
- 2) Se validó el desempeño de la herramienta dentro de un período de 12 años (1990 – 2002), para dos series de tiempo de precipitación, Apizaco Tlaxcala y Balboa Panamá, obteniendo una tasa de efectividad en el pronóstico para la variabilidad climática del orden del 75%.
- 3) Al compararse con un método empírico aplicado en la región y dentro de el período de prueba (1997 – 2002) en Apizaco Tlaxcala, nuestra herramienta logra una mejora sustancial en la habilidad de pronóstico (ver tabla 6.6).

- 4) Explotar la red neuronal ampliando el intervalo de pronóstico, en el caso de Apizaco se obtuvieron 6 meses (pasos) y en el caso de Balboa 12 pasos (una red neuronal sólo garantiza predicción para un paso en el tiempo). Encontrando que este proceso permite recuperar información relevante sobre la dinámica de las series de datos.
- 5) Establecer parámetros en función de la variabilidad, para evaluar el desempeño de la herramienta en el caso particular de la precipitación, considerando el beneficio directo que se puede proporcionar a los productores agrícolas.
- 6) Generar un documento que brinde a las personas interesadas independientemente de su formación, un panorama comprensible de la herramienta y sus alcances.

## 7.7 Trabajo Futuro

Proponemos integrar a nuestro estudio series de tiempo con un mayor número de puntos y a diferentes latitudes, a fin de verificar si se presenta una mejora del pronóstico asociado con la longitud temporal de la serie, y por otro lado, validar si la red captura el efecto del acoplamiento Océano - Atmósfera.

Podría alimentarse a la red con series de datos de estaciones meteorológicas adyacentes, con la finalidad de tratar de obtener un pronóstico sobre una región más extensa, buscando la integración a un pronóstico generado por sistemas de mesoescala.

Por la habilidad de pronóstico mostrada por la Red Neuronal Artificial, es posible pensar aplicar los pronósticos generados por ésta, para *rellenar* datos faltantes en las series de tiempo, inclusive de cualquier área del conocimiento.

Interpretar el pronóstico desde un punto de vista probabilístico, dado que la RNA provee, en forma natural, un ensamble de soluciones.

# Bibliografía

- [1] A. A. TSONIS, J. B. ELSNER, K. P. GEORGAKAKOS *Estimating the Dimension of Weather and Climate Attractors: Important Issues about the Procedure and Interpretation.* JOURNAL OF THE ATMOSPHERIC SCIENCES, 1993, 50, 15, 2249 - 2555.
- [2] GRASSBERGER, P. AND PROCACCIA, I. *Characterization of strange attractors.* PHYSIC REVIEW LETTERS, 1983, 50, 346 - 349.
- [3] J. P. ECKMAN AND D. RUELLE *Fundamental limitations for estimating dimensions and Lyapunov exponents.* PHYSICA D, 1992, 56, 185 - 187.
- [4] MARTIN CASDAGLI *Nonlinear Prediction of chaotic time series.* PHYSICA D, 1989, 35, 335 - 356.
- [5] ROBERT J. KULIOWSKY AND ANA P. BARROS *Experiments in Short-Term Precipitation Forecasting Using Artificial Neural Networks.* MONTHLY WEATHER REVIEW, 1998, 126, 470 - 482.
- [6] A.D. BACK AND E. WAN AND STEVE LAWRENCE AND A.C. TSOI *A Unifying View of Some Training Algorithms for Multilayer Perceptrons with (FIR) Filter Synapses* Neural Networks for Signal Processing 4, J. Vlontzos and J. Hwang and E. Wilson, IEEE Press, 146-154.

1995, email back@elec.uq.edu.au, url //www.neci.nj.nec.com/homepages/lawrence/papers/grad-nnsp94.

- [7] N. H. PACKARD, J. P. CRUTCHFIELD, J. D. FARMER AND R. S. SHAW *Geometry from a Time Series* PHYSICAL REVIEW LETTERS, 1980, 45, 9, 712 - 716.
- [8] A. MAZARELLA *The Time Clustering of Floodings in Venice and the Cantor Dust Method* THEORETICAL AND APPLIED CLIMATOLOGY, 1998, 59, 147 - 150 url "http://link.springer-ny.com/link/service/journals/00704/bibs/80590003/80590147.htm".
- [9] J. E. AVRON, B. SIMON *Almost periodic Hill's equations and the rings of Saturn* PHYSICAL REVIEW LETTERS, 1981, 46, 1166 - 1168.
- [10] W. C. BELFIELD *Multifractal characteristics of nature fracture apertures* GEOPHYSICAL RESEARCH LETTERS, 1994, 21, 24, 2641 - 2644.
- [11] O. E. RÖSSLER *An Equation for continuous chaos* PHYSICS LETTERS, 1976, 57A, 5, 397 - 398.
- [12] J. DOYNE FARMER AND JOHN J. SIDOROWICH *Predicting Chaotic Time Series*, PHYSICAL REVIEW LETTERS, 1987, 59, 8, 845 - 848.
- [13] KLAUS FRAEDRICH AND CHRIS LARNDER *Scaling regimes of composite rainfall time series*, TELLUS, 1993, 45A, 289 - 298.
- [14] J. B. ELSNER AND A. A. TSONIS *Nonlinear Prediction, Chaos and Noise*, BULLETIN AMERICAN METEOROLOGICAL SOCIETY, 1992, 73, 1, 49 - 59.
- [15] A. A. TSONIS AND J. B. ELSNER *Chaos, Strange Attractors and Weather*, BULLETIN AMERICAN METEOROLOGICAL SOCIETY, 1989, 70, 1, 14 - 23.

- [16] ALAN WOLF, JACK B. SWIFT, HARRY L. SWINNEY AND JOHN A. VASTANO *Determining Lyapunov Exponents from a Time Series*, PHYSICA D, 1985, 16, 285 - 317.
- [17] KLAUS FRAEDRICH *Estimating Weather and Climate Predictability on Attractors*, JOURNAL OF THE ATMOSPHERIC SCIENCES, 1987, 44, 4, 722 - 728.
- [18] KLAUS FRAEDRICH *Estimating the Dimensions of Weather and Climate Attractors*, JOURNAL OF THE ATMOSPHERIC SCIENCES, 1986, 43, 5, 419 - 432.
- [19] PETER GRASSBERGER *Do climatic attractors exist?*, LETTERS TO NATURE, 1986, 323, 609 - 612.
- [20] ANDREW M. FRASER AND HARRY L. SWINNEY *Independent coordinates for strange attractors from mutual information*, PHYSICAL REVIEW A, 1986, 33, 2, 1134 - 1140.
- [21] SHAFIQU L. ISLAM, RAFAEL L. BRASS AND IGNACIO RODRIGUEZ-ITURBE *A Possible Explanation for Low Correlation Dimension Estimates for the Atmosphere*, JOURNAL OF APPLIED METEOROLOGY, 1993, 32, 3, 203 - 208.
- [22] GEOFFREY E. HINTON *How Neural Networks Learn from Experience*, SCIENTIFIC AMERICAN, 1992, 145 - 151.
- [23] D. S. BROOMHEAD AND GREGORY P. KING *Extracting Qualitative Dynamics from Experimental Data*, PHYSICA D, 1986, 20, 217 - 236.
- [24] ANDREW M. FRASER *Reconstructing Attractors from Scalar Time Series: a comparison of Singular System and Redundancy Criteria*, PHYSICA D, 1989, 34, 391 - 404.

- [25] LÓPEZ-PEÑA R., WELBROECK H.G., MORALES T. A. AND ZERTUCHE F. G. *Prediction of Tropical Rainfall by Local Phase Space Reconstruction*, JOURNAL OF THE ATMOSPHERIC SCIENCES, 1994, 51, 22, 3360 - 3364.
- [26] ERICK R. WEEKS *Software for calculating mutual information* email "weeks@physics.emory.edu" url "http://www.physics.emory.edu/weeks", Agosto 2004.
- [27] PAOS *Program in Atmospheric and Oceanic Sciences* url "http://paos.colorado.edu/research/wavelets", Agosto 2004.
- [28] J. BRIGGS Y F. D. PEAT *Espejo y Reflejo: del caos al orden*, EDITORIAL GEDISA, 1994.
- [29] JOHN HERTZ, ANDERS KROGH AND RICHARD G. PALMER *Introduction to the theory of Neural Computation*. Lecture Notes SANTA FE INSTITUTE, Addison-Wesley, 1997.
- [30] PAUL S ADDISON *Fractals and Chaos: an illustrated course.*, Institute of Physics Publishing, 1997.
- [31] JOHN ARGYRIS, GUNTER FAUST, MARIA HAASE *An Exploration of Chaos Volume VII Text on Computational Mechanics.*, NORTH-HOLLAND, 1994.
- [32] VICENTE TALANQUER *Fractus, Fracta, Fractal Fractales de laberintos y espejos* FONDO DE CULTURA ECONOMICA, La ciencia desde México, 147, 1996.
- [33] ISAAC SCHIFTER *La Ciencia del Caos* FONDO DE CULTURA ECONOMICA, La ciencia desde México, 142, 1996.
- [34] JACQUES DUBOIS *Non-linear Dynamics in Geophysics* JOHN WILEY and SONS, WILEY-PRAXIS SERIES IN GEOPHYSICS, 142, 1998.

- [35] A. MARY SELVAM AND SUVARNA FADNAVIS *Signatures of a Universal Spectrum for Atmospheric Interannual Variability in Some Disparate Climatic Regimes*, 1998, (sin publicar).
- [36] ROBERTO HERNANDEZ ARRIAGA *Determinación de la dimensión del atractor y el límite de predictibilidad de una serie de tempo de horas calor*, Facultad de Ciencias - Universidad Nacional Autónoma de México, Tesis 1996, Físico.
- [37] ERIC A. WAN *Finite Impulse Response Neural Networks with applications in Time Series Prediction*, Stanford University, PhD Dissertation, 1993.
- [38] C. CHATFIELD *The Analysis of time Series: An Introduction*, EDITORIAL CHAPMAN AND HALL, 1980.
- [39] R. TODD OGDEN *Essential Wavelets for Statistical Applications and Data Analysis* BIRKHÄUSER BOSTON, 1995.
- [40] DONALD B. PERCIVAL Y ANDREW T. WALDEN *Wavelets Methods for Time Series Analysis*, CAMBRIDGE UNIVERSITY PRESS NEW YORK, 2000.
- [41] J.P. CRUTCHFIELD, J.D. FARMER, H. FROEHLING, N.H. PACHARD AND R.S. SHAW *Power Spectra and Mixing Properties of Strange Attractors.*, ANNALS OF THE NEW YORK ACADEMY OF SCIENCES 357, 1980; 453.
- [42] CHRISTOPHER TORRENCE, GILBERT P. COMPO *Bulletin of the American Meteorological Society.*, Vol.79, No. 1, January 1998.
- [43] HARRY L. SWINNEY *Observations of order and chaos in Nonlinear Systems.*, Physica 7D, 1983; 3-15.

- [44] GAY GARCIA C., HERNÁNDEZ VÁZQUEZ M., JIMÉNEZ LÓPEZ J., MAGAÑA RUEDA V. O., MORALES ACOLTZI T., Y OROZCO FLORES S. *Evaluation of climatic Forecasts of Rainfall for the Tlaxcala State (Mexico): 1998-2002.* , ATMOSFERA, 2004; 127-150.
- [45] HALLET T.B., T. COULSON, J. G. PILKINGTON, T. H. CLUTTON-BROCK, J.M. PEMBERTON AND B.T. GRENFELL. *Why large-scale climate indices seen to predict ecological processes better than local weather.* , NATURE 430, 2004; 71-75.