



**INSTITUTO POLITÉCNICO NACIONAL**

**CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN**

**“MEDICIÓN DE LA PREDICTIBILIDAD DE SERIES  
DE TIEMPO: UN ESTUDIO EXPERIMENTAL “**

Tesis que para obtener el grado de:

**Doctor en Ciencias de la Computación**

presenta el:

**M. en C. Ernesto Francisco Bautista Thompson**

Director de Tesis

**Dr. Jesús Guillermo Figueroa Nazuno**

**México, D. F., Abril 2005**





# INSTITUTO POLITECNICO NACIONAL COORDINACION GENERAL DE POSGRADO E INVESTIGACION

## ACTA DE REVISIÓN DE TESIS

En la Ciudad de México, D.F. siendo las 13:00 horas del día 29 del mes de Noviembre del 2004 se reunieron los miembros de la Comisión Revisora de Tesis designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:

### **Centro de Investigación en Computación**

para examinar la tesis de grado titulada:

**“MEDICIÓN DE LA PREDICTIBILIDAD DE SERIES DE TIEMPO:**

**UN ESTUDIO EXPERIMENTAL”**

Presentada por el alumno:

**BAUTISTA**

**THOMPSON**

**ERNESTO FRANCISCO**

Apellido paterno

materno

nombre(s)

Con registro: B 

0	1	1	3	9	7
---	---	---	---	---	---

aspirante al grado de: **DOCTOR EN CIENCIAS DE LA COMPUTACIÓN**

Después de intercambiar opiniones los miembros de la Comisión manifestaron **SU APROBACION DE LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

### LA COMISIÓN REVISORA

Presidente

Dr. Sergio Suárez Guerra

Secretario

Dr. José Germán González Santos

Primer Vocal (Director de Tesis)

Dr. Jesús Guillermo Figueroa Nazuno

Segundo Vocal

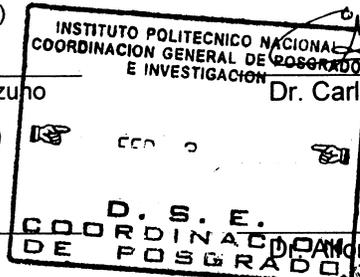
Dr. Carlos Fernando Aguilar Ibáñez

Tercer Vocal

Dr. Eduardo Gómez Ramírez

Suplente

Dr. Monso Gutiérrez Aldana



EL PRESIDENTE DEL COLEGIO

Dr. Juan Luis Díaz de León Santiago

INSTITUTO POLITECNICO NACIONAL  
CENTRO DE INVESTIGACION  
EN COMPUTACION  
DIRECCION

## **Dedicatoria**

Le dedico esta tesis a mis Padres Ernesto y Rosita, primero por todos los buenos valores que me inculcaron y me prepararon para salir adelante en la vida, y segundo por todo el sacrificio y esfuerzo en darme una educación y permitirme seguir mi búsqueda de conocimientos apoyándome incondicionalmente en los buenos y malos ratos del difícil camino de un estudiante de postgrado. Sin su apoyo, esta tesis que es la culminación de mi larga jornada de estudios no hubiese sido posible. Ahora iniciare una nueva jornada, donde tengo la seguridad que les seguiré dando satisfacciones y alegrías, gracias por todo su amor.

A mi amor y luz Fabiola, gracias por todo tu amor y apoyo incondicional, has traído un nuevo sentido a mi vida, gracias por ser como eres.

A la familia Arias-Thompson, a la tía Aída por apoyarme de muchas formas durante tantos años en esta enorme Ciudad de México, va también por el tío Santiago (q. e. p. d.) siempre recuerdo con afecto sus sabrosas platicas de sobremesa.

A mis amigos de siempre: Adrián Arturo, Luis Eduardo, Carlos, Manuel, Marco, Martín y Víctor, por escucharme siempre en los buenos y malos ratos de mis estudios de postgrado, y por todos los buenos momentos que hemos pasado juntos.

## **Agradecimientos**

Agradezco a mi Asesor el Dr. Jesús Figueroa Nazuno por estos cuatro años de fructífera colaboración, el compartir sus conocimientos conmigo y sobre todo su amistad.

A todos los compañeros del grupo de investigación presentes y pasados, de cada uno aprendí algo valioso.

A todos los compañeros del grupo con los que colabore y cuyas contribuciones ayudaron a construir la presente tesis, muchas gracias!

Agradezco en forma especial a mi amigo y colega Hugo Jiménez Hernández por brindarme su amistad y apoyo incondicional, gracias por todo!

A los miembros de mi comité de tesis por todas sus valiosas sugerencias y comentarios que enriquecieron la presente tesis, gracias!

Agradezco al Centro de Investigación en Computación del Instituto Politécnico Nacional, por el apoyo recibido en todos los aspectos y en especial de parte de todo su personal que facilitan a los estudiantes el diario trabajo, y por su esfuerzo en la formación de nuevos investigadores en computación que nuestra querida patria tanto necesita.

Finalmente, le agradezco al Instituto Politécnico Nacional el apoyo recibido para la realización de mis estudios, tanto en lo académico como en lo económico mediante la Beca Institucional del IPN.

## Resumen

En este trabajo se presenta un estudio experimental sobre la predictibilidad de series de tiempo. Se caracterizan series de tiempo de orígenes diversos y se evalúan con una metodología estándar diferentes modelos de predicción y también de modelado. Se construyen métricas para medir la predictibilidad de series de tiempo y la capacidad de los modelos para la predicción o el modelado. Se investiga la relación de la predictibilidad con respecto al comportamiento dinámico de las series de tiempo, su relación con las estructuras que dicho comportamiento genera en las series de tiempo y la relación de los parámetros dinámicos con los que se calcula la predictibilidad, con los modelos para predicción o modelado. Los resultados muestran que hay una relación no lineal entre la predictibilidad y el comportamiento dinámico de las series de tiempo; y que existe una relación entre la estructura de las series de tiempo y su predictibilidad, expresada por medio de patrones de estructura básicos. En cuanto a los modelos, no se encontraron patrones comunes de respuesta con respecto a los diferentes parámetros que caracterizan la dinámica de las series. Finalmente, explotando las diferencias entre los modelos para predecir a las series de tiempo, se desarrollaron y evaluaron con éxito dos nuevas técnicas para la combinación de modelos de predicción: *GABoost* y *CombFEC*, que permiten la mejora de los resultados de la predicción de series de tiempo.

## Abstract

In this work an experimental study about the time series predictability is presented. Time Series from different origins are characterized, and different models for forecasting and modeling are evaluated with a standard methodology. Metrics for the measurement of time series predictability, and for the measurement of model capability for forecasting and modeling are constructed. An investigation is done about the relationship between the predictability and the dynamical behavior of time series, the relationship between the structures generated by this behavior and the predictability, and the relationship between the dynamical parameters used to calculate the predictability and the models used in forecasting or modeling. The results show that there is a non linear relationship between the predictability and the dynamical behavior of the time series, and there is a relationship between the structure of time series and its predictability, expressed through basic structural patterns. About the models, no common patterns of response related with the different parameters that characterize a time series were found. Finally, exploiting the differences between the models for time series forecasting, two new techniques for model combination: *GABOOST* and *CombFEC*, were developed and evaluated successfully, these techniques allow the improvement of the time series forecasting results.

# Contenido

<b>I</b>	<b>Conceptos Fundamentales</b>	<b>8</b>
<b>1</b>	<b>Introducción</b>	<b>9</b>
1.1	Motivación . . . . .	9
1.2	Objetivos . . . . .	12
1.3	Contribución Original de la Tesis . . . . .	14
1.4	Que no se Estudia en este Trabajo (Acotación de la Tesis Doctoral) . . . . .	14
1.5	Desarrollo de la Tesis . . . . .	15
<b>2</b>	<b>Marco Conceptual de la Predicción de Series de Tiempo</b>	<b>18</b>
2.1	Introducción . . . . .	18
2.2	Sistemas de Origen Natural y Artificial . . . . .	20
2.3	Series de Tiempo . . . . .	21
2.4	El Problema de Predicción de Series de Tiempo . . . . .	24
2.5	El Modelo de Predicción . . . . .	25
2.6	Limitaciones de los Modelos y del Proceso de Predicción . . . . .	25
2.7	Predictibilidad de Series de Tiempo . . . . .	26
2.8	Patrones de Predictibilidad en Series de Tiempo . . . . .	27
<b>3</b>	<b>La Predictibilidad de Series de Tiempo: Estado del Arte</b>	<b>29</b>
3.1	Antecedentes . . . . .	29
3.2	Estado del Arte . . . . .	31
3.3	Discusión . . . . .	34

<b>4</b>	<b>Técnicas para el Análisis de Series de Tiempo</b>	<b>38</b>
4.1	Introducción . . . . .	38
4.2	Estadísticas . . . . .	39
4.2.1	Análisis Estadístico Clásico . . . . .	39
4.2.2	Análisis Estadístico Moderno . . . . .	41
4.3	Análisis de Fourier . . . . .	44
4.3.1	Frecuencia Dominante . . . . .	44
4.4	Análisis Espectral Singular (SSA) . . . . .	45
4.4.1	Análisis de Componentes Principales . . . . .	46
4.5	Teoría de Sistemas Dinámicos No Lineales . . . . .	48
4.5.1	Exponentes . . . . .	49
4.5.2	Dimensiones . . . . .	50
4.5.3	Análisis de Mapas de Recurrencia (VRA) . . . . .	52
4.6	Teoría de la Información . . . . .	57
4.6.1	Entropías . . . . .	57
4.7	Teoría de la Computación . . . . .	58
4.7.1	Complejidad Relativa LZ . . . . .	58
4.7.2	Análisis de Gramáticas (Reglas de Producción) . . . . .	59
<b>5</b>	<b>Técnicas de Predicción y de Modelado de Series de Tiempo</b>	<b>61</b>
5.1	Introducción . . . . .	61
5.2	Técnicas de Predicción . . . . .	62
5.2.1	Técnicas Estadísticas . . . . .	62
5.2.2	Técnicas de Inteligencia Artificial . . . . .	63
5.2.3	Técnicas de Teoría de Sistemas Dinámicos No Lineales . . . . .	68
5.3	Técnicas de Modelado . . . . .	70
5.3.1	Técnicas Estadísticas . . . . .	70
5.3.2	Técnicas de Inteligencia Artificial . . . . .	72
5.3.3	Técnicas de Teoría de Sistemas Dinámicos No Lineales . . . . .	79

<b>6</b>	<b>Técnicas para el Análisis de Conjuntos de Datos Multivariados</b>	<b>81</b>
6.1	Introducción . . . . .	81
6.2	Mapas Auto Organizados (SOM y GHSOM) . . . . .	82
6.2.1	SOM . . . . .	82
6.2.2	GHSOM . . . . .	84
6.3	Análisis con Escalamiento Multidimensional (MDS) . . . . .	86
6.4	Análisis de Correlación Bivariada . . . . .	91
<b>II</b>	<b>Predictibilidad de Series de Tiempo</b>	<b>94</b>
<b>7</b>	<b>Cálculo de Parámetros de las Series de Tiempo</b>	<b>95</b>
7.1	Introducción . . . . .	95
7.2	Descripción del Conjunto Experimental de Series de Tiempo . . . . .	95
7.3	Parámetros de las Series de Tiempo . . . . .	101
7.4	Datos Experimentales . . . . .	121
7.5	Discusión . . . . .	122
<b>8</b>	<b>Evaluación de Técnicas de Predicción y de Modelado de Series de Tiempo</b>	<b>123</b>
8.1	Introducción . . . . .	123
8.2	Método Experimental . . . . .	124
8.3	Resultados Experimentales . . . . .	130
8.3.1	Técnicas de Predicción . . . . .	130
8.3.2	Técnicas de Modelado . . . . .	137
8.4	Discusión . . . . .	144
<b>9</b>	<b>Métricas de la Predictibilidad de Series de Tiempo y Coeficientes de Capacidad para Predicción y Modelado</b>	<b>149</b>
9.1	Introducción . . . . .	149
9.2	Métricas de la Predictibilidad de Series de Tiempo . . . . .	150
9.2.1	Ortogonalidad de los Parámetros . . . . .	150
9.2.2	Métrica Coeficiente de Complejidad de Predicción (CCOP) . . . . .	156

9.2.3	Métrica de Predictibilidad con 9 Parámetros (CDP1) . . . . .	158
9.2.4	Métrica de Predictibilidad con 14 Parámetros (CDP2) . . . . .	163
9.3	Comparación de las Métricas de Predictibilidad . . . . .	166
9.3.1	Índice de Dificultad de Modelado (IDM) . . . . .	166
9.3.2	Comportamiento de las Métricas . . . . .	170
9.3.3	Análisis de Correlación de las Métricas . . . . .	176
9.4	Coefficientes de Capacidad de Predicción y de Modelado . . . . .	178
9.5	Discusión . . . . .	182
<b>10</b>	<b>Relaciones de la Predictibilidad de Series de Tiempo</b>	<b>183</b>
10.1	Introducción . . . . .	183
10.2	Agrupamiento de Series de Tiempo en base a Parámetros de Predictibilidad . . .	184
10.3	Relación Predictibilidad-Estructura de Series de Tiempo . . . . .	187
10.4	Relación entre las Características de los Modelos con su Capacidad de Predicción o de Modelado . . . . .	200
10.5	Relación Parámetros de Predictibilidad-Modelos de Predicción y Modelado . . .	204
10.6	Discusión . . . . .	219
<b>III</b>	<b>Combinación de Predictores</b>	<b>223</b>
<b>11</b>	<b>Combinación de Predictores: Mejorando la Predicción de las Series de Tiempo</b>	<b>224</b>
11.1	Introducción . . . . .	224
11.2	Técnicas Clásicas de Combinación de Predictores . . . . .	225
11.3	GABOost: Búsqueda de Pesos con Algoritmo Genético Canónico para Mini- mización del Error de Predicción RMSE . . . . .	227
11.3.1	El Algoritmo Boosting . . . . .	227
11.3.2	Descripción del Algoritmo de la Técnica GABOost . . . . .	229
11.3.3	Técnicas y Series para la Evaluación de GABOost . . . . .	230
11.3.4	Resultados Experimentales . . . . .	232
11.4	CombFEC: Combinación de Predictores con una Función de Pesos de Error y Correlación . . . . .	249

11.4.1	La Función de Pesos de Error y Correlación . . . . .	249
11.4.2	Descripción del Algoritmo de la Técnica CombFEC . . . . .	250
11.4.3	Resultados Experimentales . . . . .	250
11.5	Discusión . . . . .	264
<b>IV</b>	<b>Conclusiones</b>	<b>269</b>
<b>12</b>	<b>Conclusiones y Líneas de Trabajo Futuro</b>	<b>270</b>
12.1	Conclusiones . . . . .	270
12.2	Líneas de Trabajo Futuro . . . . .	274
<b>13</b>	<b>Apéndices</b>	<b>276</b>
13.1	Trabajos Derivados de la Tesis (Artículos, Ponencias, Posters, Seminarios) . . . . .	276
13.1.1	Listado de Artículos, Ponencias y Posters . . . . .	276
13.1.2	Seminarios por Invitación . . . . .	286
13.2	Guía sobre las Metodologías Experimentales para la Caracterización de las Series de Tiempo y de las Técnicas de Predicción y Modelado . . . . .	286
13.2.1	Caracterización de la Predictibilidad de Series de Tiempo . . . . .	289
13.2.2	Evaluación de la Capacidad de Técnicas para la Predicción o el Modelado	293
13.3	Análisis de Gramáticas (Pendiente de la Curva como una Medida de Predictibilidad) . . . . .	296
13.4	Visualizando los Parámetros Dinámicos de las Series de Tiempo . . . . .	300
13.5	Sobre la No Universalidad de los Modelos de Predicción . . . . .	302
13.6	Comentario sobre los Teoremas No Free Lunch . . . . .	309

## Parte I

# Conceptos Fundamentales

# Capítulo 1

## Introducción

### 1.1 Motivación

Desde los tiempos de Isaac Newton cuando logro formalizar matemáticamente el conocimiento empírico producto de las observaciones experimentales de fenómenos astronómicos, algunas de ellas sintetizadas en las leyes postuladas por Johannes Kepler, y construir su obra cumbre *Philosophiae Naturalis Principia Mathematica*, la predicción de la dinámica de los sistemas basándose en la construcción de modelos matemáticos, ha sido el camino a seguir en la búsqueda de entender a los sistemas de la más diversa naturaleza. Sin embargo, desde el inicio de este triunfo del ingenio humano se encontraron dificultades tales como la imposibilidad de resolver el problema de tres cuerpos en interacción en forma analítica, es decir, se tenía la expresión matemática pero no la posibilidad de resolverla en forma exacta, a lo largo de la historia al tratar de estudiar sistemas cada vez de mayor complejidad, nuevas limitantes han surgido a los métodos tradicionales para construir y modelar los sistemas, y nuestra capacidad de predecir la dinámica de ellos se ha visto limitada, lo que ha motivado el surgimiento de *métodos no tradicionales* para modelar y predecir la dinámica de un sistema, como son los *modelos algorítmicos* los cuales permiten el análisis de la dinámica a partir de datos experimentales del sistema, sin necesidad de construir y resolver un modelo matemático sobre dicho sistema [1, 2].

El modelado y predicción de la dinámica de un sistema (de origen natural o artificial) se encuentra limitado en muchos casos por la dificultad experimental para obtener información de la totalidad de los grados de libertad (variables independientes) que describen dicha dinámica,

así como por la dificultad para construir y resolver un sistema de ecuaciones diferenciales no lineales acopladas que la modelen [3]. En consecuencia, para el estudio de la dinámica del sistema es necesario recurrir a la búsqueda de metodologías que permitan a partir del análisis de información parcial, la reconstrucción de su dinámica sin recurrir a modelos fenomenológicos basados en la construcción de sistemas de ecuaciones diferenciales. Las técnicas de análisis que estudian los conjuntos de datos conocidos (*Series de Tiempo*) sobre una o varias variables de un sistema, son el primer paso para la determinación de las características de la dinámica del mismo y para la selección o desarrollo del modelo de predicción de la serie de tiempo. Hay diferentes técnicas de caracterización de las series de tiempo, por ejemplo: el análisis de Fourier y el análisis estadístico, cada una de ellas proporciona información sobre un aspecto del comportamiento de las series de tiempo. En la literatura se encuentran pocos trabajos en los cuales se realicen estudios sistemáticos de *conjuntos de series de tiempo* y además se exploten *diferentes técnicas de análisis* que se complementan para caracterizar a las series [4, 5, 6, 7, 8], es posible que en conjunto los parámetros encontrados, no solo proporcionen una mejor caracterización sobre el comportamiento de las series sino también sobre la *predictibilidad* de las mismas [9, 10]. En la presente tesis se utilizan los modelos algorítmicos para el estudio de *la predictibilidad de series de tiempo*, y de esta forma identificar que información hay en las series de tiempo, y que capacidad se tiene para extraer dicha información e inferir relaciones a partir de la misma. *El medir y caracterizar la predictibilidad de las series de tiempo contribuye a entender el proceso de predicción de la dinámica de los sistemas.*

Los modelos de predicción (*predictores*), se basan en asumir una relación entre el dato predicho y los datos anteriores conocidos (*Auto Regresión*), aún cuando la forma de construir el modelo puede ser diferente: Estadística [11, 12], Teoría de Sistemas Dinámicos No Lineales (Reconstrucción en el Espacio Fase) [3, 13, 14], Inteligencia Artificial (Redes Neuronales, Support Vector Machines, etc.) [15, 16, 17, 18, 19, 20]. Existen reportadas en la literatura especializada el desarrollo de diferentes técnicas de predicción y modelado [17, 21, 22, 23, 24, 25, 26], donde se aplican diversas métricas para cuantificar el error de la predicción o del modelado y se utilizan diferentes series de tiempo en la evaluación de las técnicas, dichas series son aquellas para las que generalmente el modelo fue desarrollado ex-profeso [14, 24, 27, 28]. En la literatura no se reportan los fracasos de las técnicas de predicción, los cuales también son importantes para

entender mejor las limitaciones de las técnicas evaluadas. Aún cuando hay diversos trabajos de investigación y reportes de competencias, donde se presenta la comparación de diferentes modelos de predicción [14, 24, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42], en dichos trabajos se encuentra que presentan una o más de las siguientes limitaciones:

- Uso de un grupo pequeño de series de tiempo para evaluar los modelos o incluso de una sola serie de tiempo.
- Uso de diferentes métricas de error de predicción.
- Muchas de las series usadas son específicas de un sistema en particular, por ejemplo de tipo económico o financiero.
- Comparación de modelos del mismo origen.
- Se busca evaluar únicamente el rendimiento (i.e. el error de predicción) de los modelos, pero no su capacidad de predicción.
- No se considera la existencia de relaciones entre las características de las series de tiempo y el tipo de modelo de predicción.

Todo lo anterior, indica la falta de *metodologías estándares para la evaluación y comparación de las técnicas de predicción y modelado*.

Aún cuando se han desarrollado diferentes técnicas de predicción, se conoce muy poco sobre las *relaciones entre los modelos de predicción de series de tiempo y las características de las series*, lo anterior es consecuencia del *avance del conocimiento sobre predicción de series de tiempo en extensión pero no en profundidad*.

Al intentar predecir el comportamiento de una variable del sistema con datos conocidos sobre su dinámica (*Serie de Tiempo*) se tiene la limitante de la confiabilidad de la predicción, es decir hasta que rango de tiempo (*Horizonte de Predicción*) el error de la predicción es aceptable para el tipo de información que se está prediciendo. El error en la predicción o en el modelado, no puede ser ignorado, y tampoco es posible eliminarlo del todo, pero si es posible tratar de *minimizar* su efecto en los resultados de la predicción. Una de las estrategias más exitosas para mejorar el resultado de una predicción es la *combinación de predictores*, en la literatura

existen numerosos trabajos sobre este tema [29, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54], la estrategia de combinación consiste en la optimización de la relación entre el bias y la varianza del conjunto de datos predichos [11]. A pesar de los muchos trabajos sobre combinación de modelos de predicción, es importante continuar con el desarrollo de *nuevas técnicas*, ya que ningún predictor individual o combinado es capaz de predecir toda la variedad de series de tiempo, es decir *no hay modelos de predicción universales* [55, 56]. Al explotar por medio de la combinación de los predictores, las diferentes capacidades de los modelos, es posible extender el horizonte de predicción para las series de tiempo al mejorar la predicción resultante.

En la Figura 1-1, se presentan en forma esquemática los problemas mencionados en esta sección y que son motivo de estudio de esta tesis doctoral.

## 1.2 Objetivos

Los motivos enunciados en la sección anterior, son la base para el desarrollo de los objetivos con los cuales esta tesis doctoral se propone contribuir al estudio del problema de predicción de series de tiempo en dos temas básicos: **la predictibilidad de las series de tiempo** y **la mejora de la predicción basándose en la combinación de modelos de predicción**. A continuación se enuncian los objetivos de este trabajo:

- Medir *la predictibilidad de las series de tiempo* construyendo métricas basadas en un conjunto de parámetros dinámicos de las series.
- Desarrollo de una *metodología experimental* para la evaluación y comparación de la capacidad de predicción (o de modelado) de diferentes técnicas.
- Los dos objetivos anteriores son la base para el desarrollo de *un estudio experimental* sobre la predictibilidad de series de tiempo.
- Desarrollo de nuevos métodos para mejorar la predicción de una serie de tiempo, por medio de la combinación de predictores, haciendo uso de analogías de técnicas de aprendizaje de máquina (Machine Learning) tales como: Bagging y Boosting, además de las basadas en fundamentos estadísticos básicos.

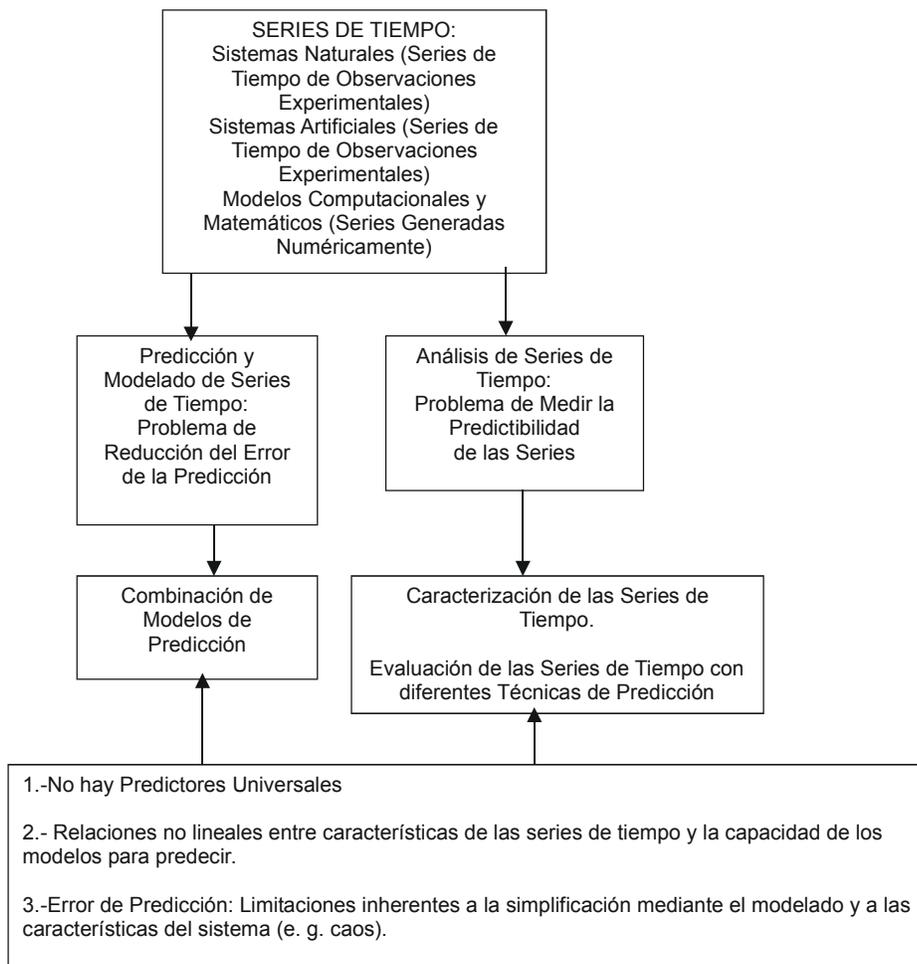


Figura 1-1: Problemas desarrollados en la tesis doctoral

### 1.3 Contribución Original de la Tesis

Las contribuciones originales al estudio de la predicción de series de tiempo son las siguientes:

- Estudio en forma global de un conjunto de series de tiempo de diferentes orígenes.
- Identificación de los parámetros más significativos para la caracterización de la predictibilidad de series de tiempo.
- Desarrollo de una (s) métrica (s) de predictibilidad en base no a una sola característica (de origen computacional, de teoría de sistemas dinámicos no lineales, etc.) sino considerando a todo un conjunto de parámetros de las series de tiempo.
- Estudios comparativos para técnicas de modelado y para técnicas de predicción basados en modelos y teorías diferentes por medio de una metodología experimental estandarizada.
- Desarrollo de métricas de capacidad de predicción y modelado.
- Identificación de la forma en que se agrupan las series de tiempo con relación a los parámetros con los que se determina su predictibilidad.
- Estudio de las relaciones entre la estructura de las series de tiempo y su predictibilidad.
- Estudio de las relaciones entre la dinámica de las series de tiempo y su predictibilidad.
- Estudio de las relaciones entre los parámetros de las series de tiempo y los modelos.
- Desarrollo de técnicas para mejorar la predicción en series de tiempo, utilizando ideas de Machine Learning (Boosting), Estadística y Algoritmo Genético; para la combinación de predictores no lineales basados en Redes Neuronales y Teoría de Sistemas Dinámicos No Lineales.

### 1.4 Que no se Estudia en este Trabajo (Acotación de la Tesis Doctoral)

Se presenta a continuación una serie de puntos que *no se estudian* en la presente tesis doctoral, y que permiten *acotar* los temas de estudio de la misma:

- Análisis de series de tiempo con diferentes regímenes de comportamiento (no estacionarias).
- Análisis de series multivariadas.
- Desarrollo de un modelo formal estricto (analítico) del problema de predicción de series de tiempo.
- Análisis de semejanzas morfológicas entre series de tiempo para su aplicación en la extracción de reglas de modelado.
- Análisis fino de los parámetros que caracterizan a las series de tiempo.
- Análisis (caracterización y parametrización) de casos particulares de series de tiempo y su relación con su sistema de origen.
- Métodos de extracción de reglas para el modelado de una serie de tiempo.
- Desarrollo de técnicas de predicción y modelado.

## 1.5 Desarrollo de la Tesis

En la Figura 1-2, se presenta el diagrama del contenido de la tesis y como se relacionan los diferentes capítulos, los cuales pertenecen a alguna de las cuatro partes temáticas en que se divide la presente tesis. El capítulo 2, presenta el marco conceptual a partir del cuál se desarrolla este trabajo. El capítulo 3, presenta una revisión del estado del arte sobre la predictibilidad de series de tiempo. El capítulo 4, se describen las técnicas de análisis de series de tiempo empleadas en este trabajo. El capítulo 5, describe las técnicas de predicción y modelado que fueron evaluadas. El capítulo 6, describe las técnicas de análisis de datos usadas en el estudio de los resultados experimentales. Los capítulos 7 y 8 presentan respectivamente, los resultados experimentales de caracterizar las series de tiempo y de evaluar las técnicas de predicción y modelado. El capítulo 9, presenta el desarrollo y evaluación de las métricas de predictibilidad y capacidad de predicción y modelado. El capítulo 10, presenta los resultados del estudio de las relaciones de la predictibilidad de series de tiempo. El capítulo 11, presenta el desarrollo y evaluación de dos técnicas para la combinación de predictores. Finalmente el capítulo 12,

presenta las conclusiones de esta tesis y posibles líneas de trabajo futuro. Adicionalmente, se tiene el capítulo 13 que presenta los siguientes apéndices: "trabajos derivados de la tesis (artículos, ponencias, posters, seminarios)", "guía sobre las metodologías experimentales para la caracterización de las series de tiempo y de las técnicas de predicción y modelado", "análisis de gramáticas (pendiente de la curva como una medida de predictibilidad)", "visualizando los parámetros dinámicos de las series de tiempo", "sobre la no universalidad de los modelos de predicción" y "comentario sobre los teoremas no free lunch".

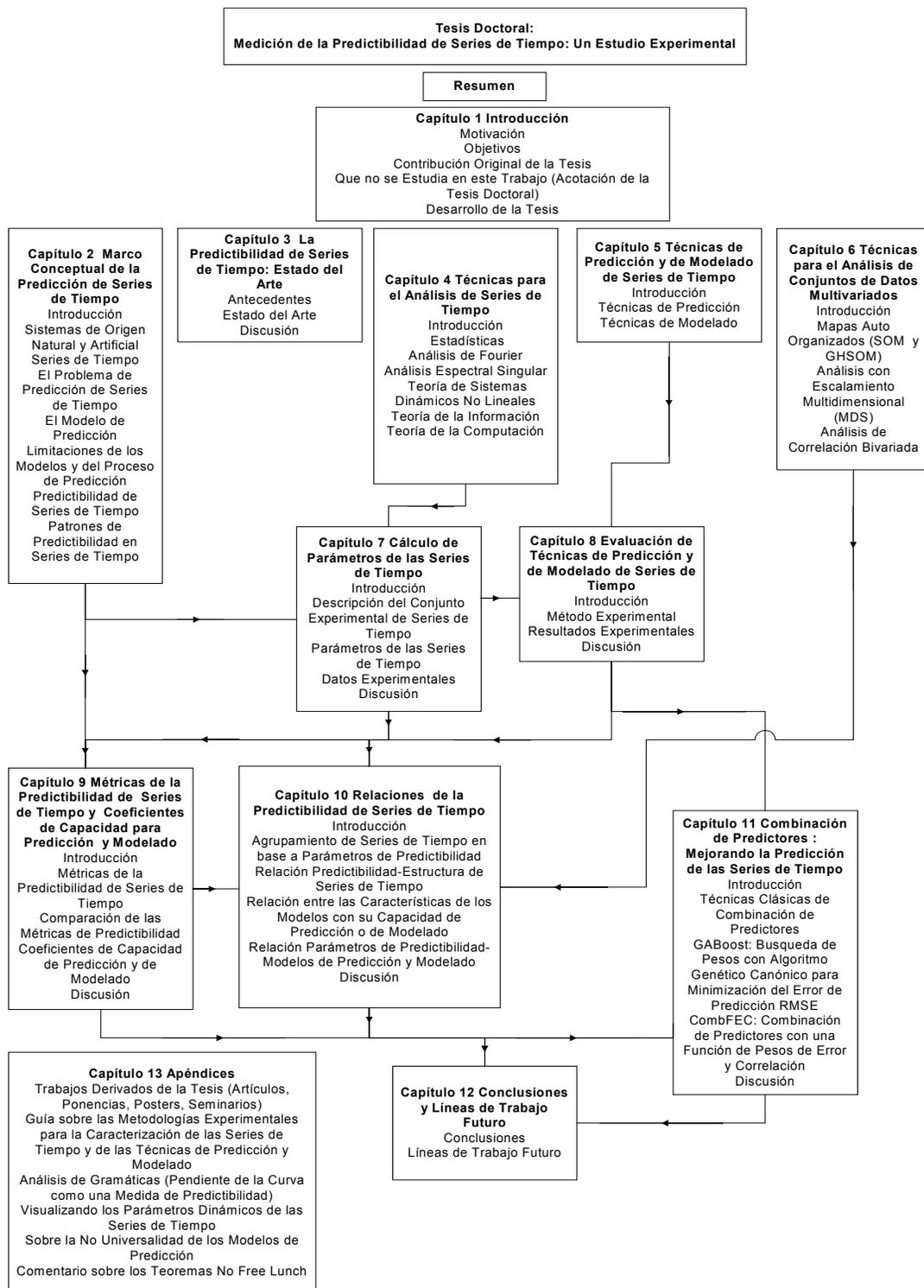


Figura 1-2: Diagrama del contenido de la tesis

## Capítulo 2

# Marco Conceptual de la Predicción de Series de Tiempo

### 2.1 Introducción

La dinámica de un sistema (de origen natural o artificial) puede estudiarse con series de tiempo, una serie de tiempo es un conjunto de datos obtenidos a partir de una observación experimental del sistema, o mediante el cálculo numérico de las ecuaciones de movimiento (en el caso particular que sea posible calcular su solución). Las series de tiempo contienen información sobre las variables independientes de un sistema (grados de libertad) que determinan su dinámica, la extracción de esta información es un problema que se estudia mediante el análisis (caracterización), predicción y modelado de las series de tiempo. La caracterización de las series de tiempo tiene como objetivo la extracción de la mayor cantidad de información sobre el comportamiento de la serie es decir, sobre la dinámica del sistema representado, y así identificar o construir la mejor o mejores técnicas que puedan predecir o en su caso modelar la dinámica del sistema expresada por la serie de tiempo. Las técnicas de predicción y modelado buscan extraer la mayor cantidad de información (proceso de aprendizaje) sobre la dinámica de la serie de tiempo, para construir modelos que reproduzcan (modelado) o extrapolen (predicción) el comportamiento de la serie.

En este trabajo doctoral se busca generar conocimiento sobre el problema de predicción de series de tiempo, no con la construcción de técnicas, algoritmos o métodos particulares

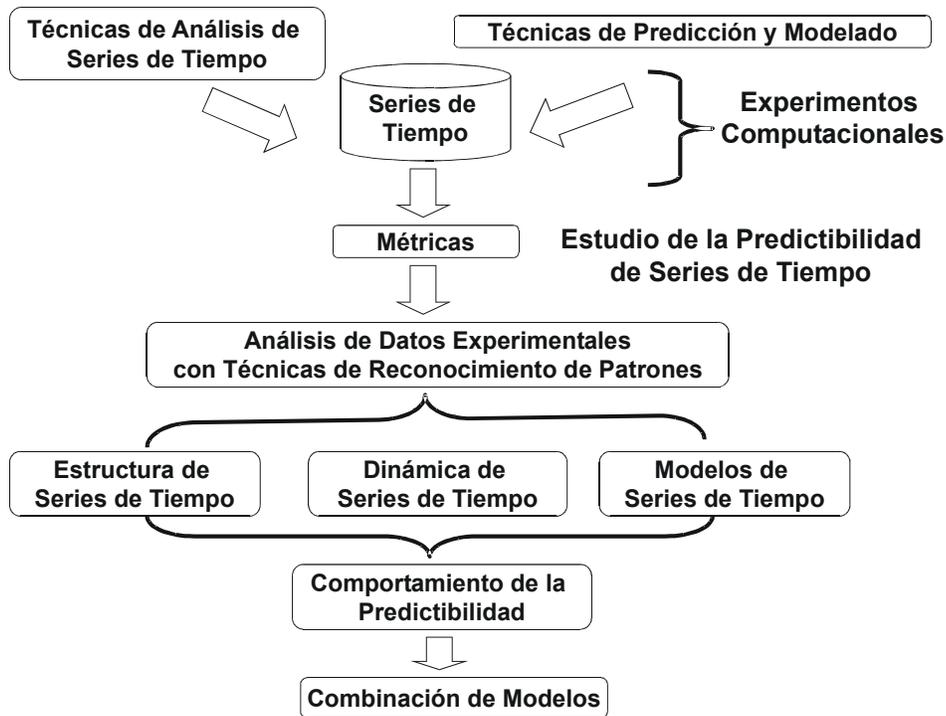


Figura 2-1: Metodología experimental para el estudio de la predictibilidad de series de tiempo

para la predicción de algunas series específicas, sino mediante *un estudio mas general sobre la predictibilidad de las series de tiempo y las relaciones entre series con diferentes características y diferentes técnicas de predicción y modelado*. Para realizar dicho estudio, se hace uso de una metodología que incluye *experimentos computacionales*, el desarrollo de una métrica para medir la predictibilidad y el análisis mediante la comparación, clasificación, jerarquización, etc. del conjunto de datos generados por los experimentos; con la información obtenida de esta forma se construye una base de conocimiento sobre la predictibilidad de series de tiempo en relación con la dinámica y estructura de las series de tiempo y los modelos de predicción, además de desarrollarse métodos para combinar modelos de predicción y así mejorar la predicción de series de tiempo. La metodología general descrita arriba para el estudio de la predictibilidad de series de tiempo se ilustra en la Figura 2-1.

El *marco conceptual* que proporciona las bases para el desarrollo del estudio sobre la predictibilidad de series de tiempo, descrito anteriormente, se define en las siguientes secciones.

## 2.2 Sistemas de Origen Natural y Artificial

En la naturaleza existen sistemas que llamamos de *origen natural* tales como los de tipo biológico por ejemplo: sistemas de un organismo multicelular (sistema nervioso, sistema celular, sistema inmunológico), sistemas de tipo bioquímico (el metabolismo), sistemas de tipo ecológico (ecosistemas, cadenas alimenticias). Otros sistemas de origen natural son de tipo físico: dinámica atmosférica, dinámica de los océanos, etc. Por otro lado, existen sistemas de *origen artificial* es decir creados por el hombre por ejemplo: sistemas de tipo socioeconómico (la guerra, el comercio, la economía de un país), sistemas de tipo cultural (Internet) [57]. Los sistemas anteriores son estudiados de forma experimental, los experimentos sobre estos sistemas generan gran cantidad de información expresada como secuencias de datos que llamaremos series de tiempo (las cuales se definen formalmente más adelante), a las series de tiempo generadas mediante mediciones experimentales las clasificamos como *series naturales*.

Otra forma de generar series de tiempo, es a partir del estudio de la dinámica de sistemas generados a partir de experimentos computacionales como por ejemplo autómatas celulares [58], o partículas bajo la influencia de un potencial [14]. También es posible obtener series de tiempo a partir de la solución numérica de expresiones matemáticas de modelos físicos (como en el caso del sistema de ecuaciones diferenciales de Lorenz) [59] o de construcciones matemáticas abstractas como el Problema de Collatz en Teoría de Números [60, 61]. A las series de tiempo cuyo origen son sistemas artificiales generados con simulaciones por computadora o sobre la base de la solución de expresiones matemáticas las clasificamos como *series de tiempo sintéticas*.

La Tabla 2.1 presenta ejemplos de sistemas que generan series de tiempo naturales y sintéticas.

(2.1)

Origen del Sistema	Tipo de Sistema	Sistema
Natural	Biológico	Nervioso
Natural	Biológico	Inmunológico
Natural	Biológico	Metabolismo
Natural	Físico	El clima
Natural	Físico	El Oceano
Natural	Químico	Reacciones Químicas
Artificial	Cultural	Internet
Artificial	Social	La guerra
Artificial	Social	El comercio
Artificial	Matemático	Mapas Discretos
Artificial	Computacional	Automatas Celulares

En esta tesis doctoral se estudia a las *Series de Tiempo*, las cuales representan el comportamiento dinámico (*en general en el tiempo*) de variables experimentales de sistemas de origen natural o artificial.

### 2.3 Series de Tiempo

Al estudiar la dinámica de un sistema mediante un experimento se seleccionan un subconjunto de variables del sistema las cuales son observadas y dichas observaciones se representan por conjuntos de datos en el tiempo (*Series de Tiempo*). Una serie de tiempo es una proyección en  $R^1$  de la trayectoria de la dinámica del sistema en el espacio fase de  $m$  dimensiones  $R^m$  (grados de libertad del sistema).

La serie de tiempo es una estructura de información a partir de la cual es posible recuperar el conocimiento necesario para reconstruir la trayectoria dinámica del sistema (*al menos en forma parcial*), lo anterior fue demostrado matemáticamente por Whitney (1936), quien demostró la existencia de una función invertible  $\Phi$  que posibilita la transformación de la serie de tiempo de un espacio euclidiano  $R^1$  a una trayectoria en el espacio real euclidiano  $R^p$  (con  $p \leq m$ )

dicha trayectoria tiene una correspondencia biunívoca con la trayectoria original de la dinámica del sistema en el espacio  $R^m$  de  $m$  dimensiones [62]. Posteriormente Takens (1980), propuso y demostró un teorema para la reconstrucción de la trayectoria del atractor en el espacio fase a partir del embebido de la serie de tiempo [63]. La Figura 2-2 ilustra el proceso de la reconstrucción de la trayectoria de un atractor, la cual representa la dinámica de un sistema, a partir de una serie de tiempo. En dicha figura, se ilustra como a la trayectoria original de la dinámica del sistema en el espacio  $R^m$  y que es representada por el funcional  $F$ , le corresponde la serie de tiempo en el espacio  $R^2$  (considerando en forma explícita a la variable tiempo  $t$ ) donde ahora la dinámica del sistema se representa por una aproximación al funcional original denominado  $F'$ , dado que por el teorema debido a Whitney, se sabe que existe la función invertible  $\Phi$  que permite la conexión entre la trayectoria original y la reconstruida a partir de la serie de tiempo, entonces es posible reconstruir, aplicando la idea del embebido a la serie original (teorema de Takens), la trayectoria original del sistema en un espacio de dimensión  $R^p$  y de esta forma recuperar la información sobre la dinámica del sistema que se encuentra oculta dentro de la serie de tiempo.

A partir de las ideas anteriores se construye la siguiente definición formal para las series de tiempo:

**Definición 1** Sea  $ST = \{x(1), x(2), x(3), x(4), \dots, x(t), \dots, x(N)\}$  una secuencia de datos experimentales para un intervalo de tiempo  $T = N$ , de una variable  $u$  observable  $x(t)$  de un sistema de origen natural o artificial  $S$ , al conjunto  $ST$  se le denomina Serie de Tiempo.

En la presente tesis doctoral, se considera el estudio de un conjunto de series de tiempo  $\{ST_i\}$  cuyos orígenes son observables de diferentes sistemas (naturales y artificiales), para estudiar dicho conjunto se hace *abstracción* de sus sistemas de origen, en el sentido de que no son los sistemas que las originaron lo que se estudia en esta tesis, sino las series de tiempo como *estructuras de información*. Dichas estructuras, poseen un conjunto de características comunes cuantificables (*parámetros*) que permiten su estudio y comparación.

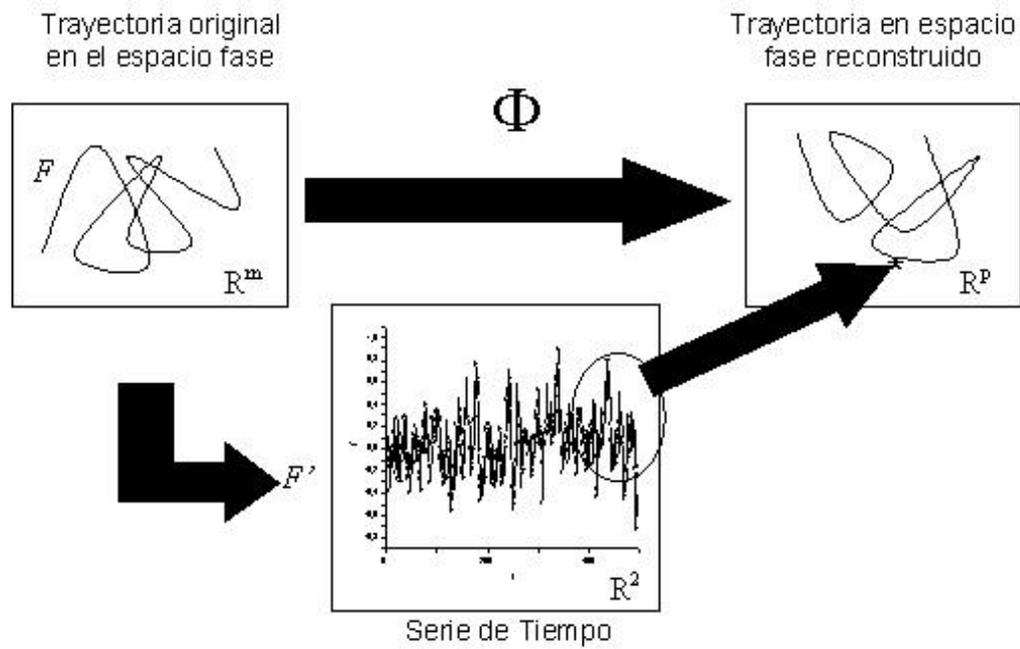


Figura 2-2: Reconstrucción de la trayectoria dinámica de un sistema a partir de una serie de tiempo

## 2.4 El Problema de Predicción de Series de Tiempo

Al estudiar un sistema se busca modelar la dinámica del mismo y como fin último realizar predicciones sobre dicha dinámica, para lograr lo anterior se desarrollan experimentos con los que se obtiene información sobre el comportamiento de algunas de las variables del sistema y con dicha información se trata de inferir las relaciones causales entre dichas variables que permitan desarrollar un modelo aproximado del sistema. Si dicho modelo es soluble analíticamente o numéricamente entonces es posible modelar al sistema y proponer predicciones sobre su dinámica. Sin embargo, las limitaciones experimentales para obtener la información sobre las variables de un sistema y / o las limitaciones de las matemáticas para resolver los modelos que se proponen, no permiten tener la suficiente información para desarrollar modelos solubles de la dinámica del sistema en todos los casos. En el caso de que se tenga la observación experimental de una o más variables del sistema, pero no exista un modelo soluble del sistema, es posible construir otro tipo de modelo el cual reproduzca el comportamiento del observable u observables (*representado por una o varias series de tiempo*) de tal forma que se pueda realizar una predicción sobre la dinámica observada. Es importante aclarar que estos modelos no tienen una relación directa con la dinámica del sistema de donde se obtuvo experimentalmente la serie de tiempo, de hecho el relacionar el modelo de comportamiento de una serie de tiempo con la dinámica del sistema es un *problema abierto* [3]. A continuación se presenta la definición formal del problema de predicción de series de tiempo.

**Definición 2** Sea  $F(x_1(t), x_2(t), \dots, x_k(t))$  el funcional que describe en forma completa la dinámica del sistema  $S$ , dicho funcional no es conocido en forma explícita.

**Definición 3** Sea  $F^A$  una aproximación al funcional  $F$ , que modela la dinámica del sistema  $S$  pero cuya forma analítica o no es conocida en forma completa o no es soluble en forma numérica.

**Definición 4** Dado un sistema  $S$  para el cual se tiene una aproximación  $F^A$  al funcional  $F$ , cuya forma analítica no se conoce en forma completa o no es soluble numéricamente, pero se tiene una serie de tiempo  $ST$  para un intervalo de tiempo  $T$ , sobre la variable  $x(t)$  del sistema  $S$ , entonces es posible construir un modelo  $P$  que permita predecir el comportamiento de dicha variable para un tiempo  $t > T$ .

## 2.5 El Modelo de Predicción

La construcción del modelo de predicción es la parte central del estudio de series de tiempo, la mayoría de los modelos de predicción se basan en la Auto Regresión propuesta por Yule en 1927 [15], representada por la siguiente ecuación,

$$\hat{y}(k) = \sum_{i=1}^p a_i y(k-i) \pm e(k) \quad (2.2)$$

donde  $\hat{y}(k)$  es la predicción del valor que corresponde al tiempo  $k$ , a partir de la combinación lineal de los  $p$  valores anteriores de la forma  $y(k-i)$ , mas un termino de error  $e(k)$ ; con modificaciones este modelo ha sido base de técnicas de predicción desarrolladas en Estadística [11, 12], Teoría de Sistemas Dinámicos No Lineales [3, 13, 14] y Redes Neuronales [15, 16, 17, 18]; en menor cantidad se han desarrollado otros métodos para resolver el problema de predicción por ejemplo con Estadística Bayesiana [64].

En la literatura se habla de la construcción de modelos, métodos, algoritmos o técnicas de predicción, y muchas veces no se separa con claridad a las técnicas que estrictamente *modelan* una serie de tiempo (interpolación de puntos de la serie) y a las técnicas que además de modelar la serie la *predicen* (extrapolación de puntos), en esta tesis doctoral, se hace *una clara diferencia* entre las técnicas que modelan y aquellas que predicen, ya que el verdadero reto es predecir el comportamiento de la serie de tiempo considerando que se posee información incompleta sobre el sistema, y existen limitaciones inherentes a los modelos y al proceso de predicción mismo (ver siguiente sección).

## 2.6 Limitaciones de los Modelos y del Proceso de Predicción

Los modelos actuales de predicción de series de tiempo poseen la siguiente limitante en su capacidad de predecir: son modelos construidos específicamente para series de tiempo particulares y entonces *no son generalizables* o en otras palabras *no existen predictores universales*. Los modelos *son aproximados* lo cual introduce un error en la predicción de la serie de tiempo, el cual se define en este trabajo como *Error de Modelado*.

Por otra parte las características de la dinámica de las series de tiempo también influyen

en la dificultad para ser predichas por las técnicas o métodos de predicción (*predictibilidad de la serie*), algunas series de tiempo poseen cierto grado de caos, el cual introduce un error que crece en forma exponencial, por otra parte si poseen aleatoriedad la suposición de ciertas técnicas basadas en auto regresión de que hay memoria entre los datos, no es valida, este tipo de error lo definimos como un error debido a las características de las series, es un error de tipo *sistemático* sobre el cual no se tiene control [65], a este error se le define como *Error Intrínseco* en el presente trabajo.

Por último el error experimental (en el caso de series de tiempo derivadas de observaciones experimentales), introduce un error en forma de ruido [3, 66, 67], que se define en este trabajo como *Error de Medición*.

Este conjunto de limitaciones, introduce un *Error de la Predicción* definido como:

$$E_{predicción} = E_{modelado} + E_{intrínseco} + E_{medición} \quad (2.3)$$

el cual no es posible eliminar en su totalidad pero si disminuirlo. El error de la predicción y la forma en que éste se expande limita el alcance de la misma (*horizonte de la predicción*) al disminuir su intervalo de confianza.

## 2.7 Predictibilidad de Series de Tiempo

Una serie de tiempo contiene información sobre la dinámica de un sistema, la cual en principio es posible recuperar dada la existencia de una función invertible  $\Phi$  como se mencionó en la sección 2.3, sin embargo la existencia de dicha función no garantiza a priori que sea posible modelar la dinámica de una serie de tiempo y en consecuencia predecir su comportamiento futuro, ya que los modelos de predicción poseen limitaciones para modelar dicha dinámica, como se mencionó en la sección 2.6. Para lograr una mejor comprensión del problema de predicción de series de tiempo, es necesario caracterizar dicho proceso, para ello se estudia *la predictibilidad de las series de tiempo* usando diferentes parámetros que relacionan la dinámica de las series de tiempo con la predictibilidad de las mismas, los parámetros proporcionan información sobre las series de tiempo desde *diferentes facetas* (i. e. *multiplican la información sobre su dinámica*): Estadística, Sistemas Dinámicos No Lineales, Teoría de la Información y Teoría

de la Computación; para cuantificar la predictibilidad se combinan los diferentes parámetros para determinar mediante un valor escalar la predictibilidad de las series de tiempo, de esta forma es posible: *ordenar, comparar, discriminar, clasificar y jerarquizar* la información sobre la predictibilidad y así generar estructuras conceptuales que aporten nuevo conocimiento [68, 69, 70, 71]; entonces como parte del marco conceptual de este trabajo se asume la siguiente hipótesis: es posible construir una métrica a la que se le denomina *predictibilidad* o en el caso de su relación inversa *dificultad de predicción*, a partir de un conjunto de parámetros experimentales que caracterizan a las series de tiempo; dicha métrica permite medir la dificultad que la serie de tiempo posee para ser predicha por un modelo. Esta métrica es independiente de la existencia o no, de un modelo capaz de predecir a una serie de tiempo. Solamente depende de un conjunto de parámetros que caracterizan a la serie de tiempo y en consecuencia de la dinámica que dicha serie representa. A partir de los conceptos anteriores se define formalmente a la métrica de predictibilidad:

**Definición 5** Sea  $C_{ST} = \{c_1, c_2, \dots, c_q\}$  un conjunto de parámetros que caracterizan al conjunto de las series de tiempo  $\{ST_i\}$ , entonces es posible construir una métrica de predictibilidad  $M_{pred}$  para las series de tiempo de la forma:  $M_{pred} = \mathcal{F}(C_{st})$ , donde  $\mathcal{F} : R^q \rightarrow R$ .

## 2.8 Patrones de Predictibilidad en Series de Tiempo

Las limitaciones de los modelos de predicción para las series de tiempo, motivan la búsqueda e identificación de *Patrones de Predictibilidad* que ayuden a un mejor entendimiento de la dinámica del proceso de predicción de series de tiempo. Estos patrones se expresan en forma de clases y relaciones derivadas del análisis de las series de tiempo y de las técnicas (o modelos) de predicción. Un patrón es un *concepto local*, el cuál es válido para un subconjunto de variables o datos o ambos [72]. La definición formal de las ideas anteriores es la siguiente:

**Definición 6** Sea  $D_{ST}$  el conjunto de datos que contienen información sobre un conjunto  $\{ST_i\}$  de series de tiempo.

**Definición 7** Sea  $P_j$  el modelo de predicción  $j$ -ésimo, entonces se define a  $CP = \{P_j(ST_i)\}$  como un conjunto de modelos de predicción para las series de tiempo.

**Definición 8** Sea  $D_P$  el conjunto de datos que contienen información sobre un conjunto  $CP$  de modelos de predicción.

**Definición 9** Un Patrón de Predictibilidad es una Regla Local de la forma  $r \equiv T(D_{ST} \cap D_P)$  donde  $T$  es una transformación que aplica sobre el conjunto intersección de los conjuntos de datos  $D_{ST}$  y  $D_P$ .

## Capítulo 3

# La Predictibilidad de Series de Tiempo: Estado del Arte

### 3.1 Antecedentes

Las bases formales del análisis de series de tiempo se deben a los trabajos de Euler (1707-1783), d'Alembert (1717-1783), Lagrange (1736-1813) y Fourier (1768-1830) que desarrollaron la idea de la descomposición de una señal en componentes armónicas [73]. Los primeros análisis de series de tiempo, se atribuyen a la comunidad astronómica la cual aplicó el análisis de Fourier a señales de origen astronómico, el objetivo era la búsqueda de periodicidades en los datos (datos de actividad solar: manchas solares, datos de índices de luminosidad en estrellas variables). Los primeros métodos datan de 1772 y 1778, cuando Lagrange sugirió métodos para identificar periodicidades ocultas, Buijs-Ballot propuso procedimientos computacionales para el análisis estadístico de datos astronómicos en 1847. Sin embargo, es Sir Arthur Schuster a quien se puede considerar el iniciador de los métodos modernos de análisis de series de tiempo en el dominio de la frecuencia, con su propuesta en 1889 del análisis de periodogramas, de la mano de los desarrollos analíticos en los 1890's, la búsqueda de mecanismos de cálculo automatizados llevo a la construcción de dispositivos tales como los analizadores armónicos de Henrici-Conradi y de Michelson-Stratton, que aún cuando tuvieron un éxito limitado, nos recuerdan la importancia del desarrollo de la computadora para el cálculo numérico intensivo requerido en el análisis de señales [73]. El siguiente gran desarrollo en el estudio de las series de tiempo, tuvo lugar en

el dominio del tiempo, por medio de dos artículos, uno en 1927 por Yule [74], y otro en 1937 por Slutsky [75]. En dichos trabajos, se rechaza la hipótesis determinista con componentes armónicas para analizar y modelar las series de tiempo, en favor de modelos basados en causalidad aleatoria, el modelo de auto regresión y el modelo de promedio móvil, se originan respectivamente a partir de dichos trabajos. Los trabajos de Kolmogorov (1941) y Wiener (1930 y 1950) completaron los fundamentos de la construcción de los predictores lineales para señales estadísticamente estacionarias iniciados por Yule [76, 77, 78]. En 1965 el estudio de series de tiempo basado en el análisis de Fourier (dominio de la frecuencia) alcanza su madurez, con el desarrollo por parte de Cooley y Tukey del algoritmo de la transformada rápida de Fourier (FFT), que redujo considerablemente el esfuerzo para calcular el periodograma de una señal [79]. Por otra parte, en el dominio del tiempo, Box y Jenkins, publicaron en 1976 el libro titulado: *Time-series Analysis, Forecasting and Control*, en el cual desarrollaron una metodología en el dominio de tiempo mediante la combinación de diferentes temáticas y presentando aplicaciones en predicción y control, por primera vez mostraron al gran público en una forma unificada el alcance del estudio de series de tiempo, presentando casos como la predicción del número de pasajeros en una aerolínea o el análisis del proceso de combustión en una caldera de gas. Por último, adaptaron la metodología de análisis para su uso mediante la computadora [80]. En la década de 1980's nuevos desarrollos matemáticos, en algoritmos y en poder de cómputo abren la posibilidad de estudiar series de tiempo que presentan comportamiento caótico, aún cuando el estudio de sistemas caóticos se remonta a Poincaré a fines del siglo XIX [81], se tuvo que esperar al desarrollo de las computadoras y el desarrollo de algoritmos para estudiar los sistemas caóticos a detalle en forma no sólo analítica sino también experimental. Los trabajos de Whitney (1936) y posteriormente de Takens (1980) dieron las bases formales de la reconstrucción en el espacio fase de la dinámica de un sistema a partir de una serie de tiempo (observable de dicha dinámica) [62, 63].

Los avances anteriores han permitido el desarrollo tanto de técnicas de análisis de series de tiempo como de técnicas de predicción, sin embargo muy poco se ha avanzado para entender los mecanismos del proceso de predicción de series de tiempo, en particular en la comprensión de *la predictibilidad de las series de tiempo*, en la siguiente sección se hace una revisión de los trabajos publicados al respecto en la literatura especializada.

## 3.2 Estado del Arte

En la literatura especializada hay pocos trabajos relacionados con la predictibilidad de series de tiempo, no se observa un gran interés en medir y entender la predictibilidad [82], a pesar de la enorme cantidad de trabajos relacionados con el desarrollo de técnicas de predicción de series de tiempo [17, 21, 22, 23, 24, 25, 26]. El concepto de predictibilidad de series de tiempo (predictability, forecastability o accuracy en la literatura de habla inglesa [82, 83, 84, 85]) que se estudia en los trabajos reportados se define como las varianzas relativas del error de predicción y de las observaciones o también como el valor de la información condicional en la reducción del valor esperado de pérdida de una predicción [86], en otras palabras la predictibilidad es el error que la predicción de una serie de tiempo tiene debido al modelo de predicción [87, 88]. En las áreas de Estadística y Econometría es donde se han desarrollado múltiples trabajos sobre la predictibilidad, entendida como el error de predicción, para evaluar modelos de predicción financieros: Stock Returns, Stock Prices, International Equity Returns, etc. [89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99]. Prácticamente, todo trabajo que involucra el desarrollo y evaluación de técnicas para la predicción o modelado de una serie de tiempo toma en cuenta el error de la predicción y en consecuencia la predictibilidad así definida.

A pesar de encontrarse un acuerdo en cuanto a la idea de predictibilidad por parte de los trabajos sobre esta temática, no existe una definición operacional estándar y hay diferentes propuestas de medidas de predictibilidad, por ejemplo la búsqueda de correlaciones canónicas entre valores pasados y futuros de una serie de tiempo [100, 101], la comparación de la varianza de innovación y la varianza incondicional de series estacionarias [102], otros como Granger y Newbold (1986) proponen una medida de predictibilidad basada en la covarianza de series estacionarias bajo la pérdida del error cuadrático, esta medida es de la forma

$$G = \frac{\text{var}(\hat{y}_{t+j,t})}{\text{var}(y_{t+j})} \quad (3.1)$$

donde  $\hat{y}_{t+j,t}$  es la predicción óptima (i.e. media condicional) para el tiempo  $t+j$  realizada al tiempo  $t$  y  $y_{t+j}$  es el valor original para el tiempo  $t+j$  [103]. Diebold y Killian (1998) proponen la siguiente definición de predictibilidad basada en el cociente de los errores de predicción para una predicción a corto plazo (numerador) y una de largo plazo (denominador), de tal forma

que permite determinar la predictibilidad basada en los horizontes de predicción

$$P(s, k) = 1 - \frac{E(L(\varepsilon_{t+s|t}))}{E(L(\varepsilon_{t+k|t}))} \quad (3.2)$$

donde  $\varepsilon_{t+s|t}$  es el error para la predicción en el  $s$ -período de tiempo adelante, hecho al tiempo  $t$  (indicando  $s$  una predicción de corto plazo),  $L(\cdot)$  es la función de pérdida, entonces  $E(L(\varepsilon_{t+s|t}))$  es la pérdida esperada de una predicción de corto plazo, y respectivamente  $E(L(\varepsilon_{t+k|t}))$  es la pérdida esperada de una predicción de largo plazo, tal que  $k \gg s$  y  $k$  esta fija cuando  $s$  varia dentro de algún rango. La definición para ser operativa requiere la selección de una función de pérdida particular y de un modelo de predicción particular, por ejemplo se puede seleccionar como función de pérdida el error medio cuadrado ( $MSE$ ) y como modelo el auto regresivo  $AR(p)$  [82], esta definición es operacionalmente similar a la medida estadística  $U$  propuesta por Theil (1966) [104]. Otros ejemplos de métricas de predictibilidad en el área de Econometría son las desarrolladas por Kaboudan (1998) y una variante desarrollada por Duan y Povinelli (2001) las cuales se basan en definir la predictibilidad de series de tiempo a partir de la capacidad de un algoritmo de Programación Genética de generar un modelo de la serie, entonces se dice que la serie es GP-predecible [105, 106, 107, 108, 109], la métrica de Kaboudan llamada métrica  $\eta$  mide la probabilidad de que una serie de tiempo sea GP-predecible, se basa en comparar dos salidas: el mejor modelo generado de un conjunto simple de datos antes de ser alterado mediante la mezcla de los datos en forma aleatoria (*bootstrap*), con respecto al mejor modelo generado con el conjunto de datos alterado [107, 108, 109]. Específicamente, las variaciones no explicadas, las cuales son medidas mediante la suma del error cuadrado ( $SSE$ ) antes y después de la modificación de las series  $\{Y_t\}$ ,  $t = 1, 2, \dots, N$ , son comparadas. La variación no explicada en  $\{Y_t\}$  antes de la mezcla es

$$SSE_\gamma = \sum_{t=1}^T (Y_t - \hat{Y}_t)^2 \quad (3.3)$$

donde  $\hat{Y}_t$  es la predicción de  $Y_t$ . El mezclado incrementa la variación no explicada en  $\{Y_t\}$  a un máximo, este máximo es

$$SSE_S = \sum_{t=1}^T (S_t - \hat{S}_t)^2 \quad (3.4)$$

donde  $\{S_t\}$  es la mezcla de  $\{Y_t\}$ . Finalmente la medida de predictibilidad se define como:

$$\eta = 1 - \frac{SSE_\gamma}{SSE_S} \quad (3.5)$$

Una medida más de predictibilidad reportada en la literatura, es la propuesta por Galbraith (2003) denominada contenido de predicción, la cual se define como la reducción en el error medio cuadrado ( $MSE$ ), disponible de usar un modelo de predicción dado, en vez de la predicción debida a la media incondicional [110]. La medida de predictibilidad al tiempo  $s$  se describe por la función de contenido

$$C(s) = 1 - \frac{MSE_{\hat{y}(s)}}{MSE_{\bar{y}(s)}} \quad s = 1, \dots, S \quad (3.6)$$

donde  $MSE_{\hat{y}(s)}$  es el error medio cuadrado esperado del modelo para la predicción al tiempo  $s$ , y  $MSE_{\bar{y}(s)}$  es el error medio cuadrado correspondiente a la media incondicional considerada como una predicción al tiempo  $s$ .

Además de calcular la predictibilidad, en la literatura se discute como mejorar la estimación de la predictibilidad en series de tipo económico al considerar diferentes factores que influyen en la toma de decisiones de los inversionistas [111, 112]. También se desarrollan trabajos para la identificación de regiones de predictibilidad a escala local dentro de las series de tiempo utilizando métricas como la Entropía de Información de Shannon [113, 114, 115], o el análisis de gramáticas para la identificación de la dinámica de generación de reglas de producción [116, 117]. Las medidas de predictibilidad desarrolladas en Econometría y Estadística consideran a la predictibilidad como dependiente del modelo de predicción y no atribuible a las características propias de las series de tiempo.

Existe otra área de investigación donde se desarrollan trabajos sobre la predictibilidad de series de tiempo, dicha área es el estudio de Sistemas Dinámicos No Lineales, la idea detrás del concepto de predictibilidad tiene dos vertientes en esta área, la primera es similar a la manejada en las áreas de Econometría y Estadística: el error de la predicción, sin embargo la forma en que

se mide es diferente, para cuantificarlo ahora se considera el horizonte de predicción es decir la expansión del error de predicción conforme avanza el tiempo, esta expansión se determina usando el Exponente de Lyapunov de una serie de tiempo [3, 84, 118, 119, 120, 121, 122, 123, 124], su área de mayor aplicación corresponde a las series que presentan comportamiento caótico, el cual causa que pequeñas perturbaciones en las condiciones iniciales del sistema generen grandes cambios en los estados siguientes del sistema, es decir, aún cuando la serie es determinista (i. e. su dinámica es predecible) su naturaleza caótica limita la capacidad de los modelos de predicción para predecir con un error pequeño, más allá del horizonte de predicción definido con el Exponente de Lyapunov. A diferencia de los trabajos desarrollados en Econometría y Estadística, los trabajos en el área de Sistema Dinámicos No Lineales, consideran a la predictibilidad como dependiente tanto de la serie de tiempo como del modelo de predicción [84, 124]. La predictibilidad es además tratada como una medida que junto con otras métricas forma parte de un amplio conjunto de parámetros para caracterizar a las series de tiempo entendidas como observaciones de la dinámica de fenómenos físicos [6, 7, 113, 125], esta caracterización busca contribuir a una mejor comprensión de los llamados sistemas complejos [57, 126], una temática que no se aborda en este trabajo. La segunda forma de entender a la predictibilidad es a través de diferentes definiciones derivadas de la Teoría de Información: Entropía de Shannon, Entropía Condicional, Información Mutua y Funcionales de Información [7, 87, 88, 113, 114, 115, 127, 128, 129, 130].

Para finalizar, en ninguna de las tres áreas: Econometría, Estadística y Sistemas Dinámicos No Lineales; donde se estudia la predictibilidad de series de tiempo, se encuentran trabajos que usen para dicho estudio conjuntos de series de tiempo de diferente origen y en consecuencia predictibilidad [9, 10, 117, 131, 132], y que además exploten un conjunto de diferentes parámetros que caracterizan a las series de tiempo para cuantificar dicha predictibilidad [9, 10], lo que si va en el espíritu de la presente tesis y de diversos trabajos desarrollados en nuestro grupo.

### **3.3 Discusión**

En la sección anterior se presentó el avance y las tendencias en el estudio de la predictibilidad en series de tiempo (time series predictability). A partir de la información recabada se llega a las siguientes conclusiones:

- La predictibilidad de series de tiempo es un tipo de métrica o medida particular dentro del conjunto de métricas usadas para caracterizar un sistema o fenómeno físico, dichos sistemas son denominados por muchos autores como sistemas complejos.
- Se han propuesto diferentes métricas de predictibilidad basadas en los errores de predicción de las series de tiempo, el exponente de Lyapunov y algunas definiciones de entropía de información. También se han propuesto métricas más generales usadas para la evaluación de la complejidad de una señal (no necesariamente en el área de predicción de series de tiempo) que expresan la dinámica de un fenómeno, estas métricas se basan en algún parámetro individual de la señal: exponente de Lyapunov, diferentes definiciones de entropías de Información y medidas de complejidad de tipo algorítmico como la de Kolmogorov.
- En las áreas de Econometría y Estadística el concepto de predictibilidad de series de tiempo se entiende como: la medida del error probable de predicción por parte de un modelo. Es una medida de confianza en el error de predicción, permite estimar la exactitud (accuracy) de la predicción por parte de un modelo, por consiguiente es dependiente del modelo evaluado.
- En el área de Sistemas Dinámicos No Lineales, la predictibilidad se entiende como una medida del horizonte de predicción por medio de la estimación del coeficiente de Lyapunov, y entonces depende de la serie de tiempo (sensibilidad a las condiciones iniciales del sistema). También, se estudia el efecto en la predictibilidad debido a las limitaciones del modelo de predicción. El concepto de predictibilidad es similar al manejado en Econometría y Estadística, que la predictibilidad es el error de predicción que se espera obtener, pero en lugar de estimar el error promedio, se busca estimar la expansión del error de predicción o incertidumbre (uncertainty) conforme avanza la predicción en el tiempo, en este caso la predictibilidad puede depender tanto de la serie de tiempo como de las limitaciones del modelo o de ambos.
- De los puntos anteriores se concluye que la predictibilidad se entiende como el error probable resultado de la predicción de una serie de tiempo por parte de un modelo. La predictibilidad, unos autores la manejan dependiente del modelo de predicción, otros

como una propiedad intrínseca de la serie de tiempo medida por medio de un parámetro y algunos más como dependiente tanto de la serie como del modelo de predicción.

- Al evaluar las métricas no hay trabajos que utilicen conjuntos de series de tiempo de diferente origen en dicha evaluación, ni comparaciones con otras métricas.
- Los principales objetivos de los trabajos que se han desarrollado sobre predictibilidad son: cuantificar el error de predicción o incertidumbre debida al modelo de predicción e identificar las zonas de alta predictibilidad dentro de una serie de tiempo.
- No hay en el estudio de las series de tiempo un marco conceptual teórico unificado, básicamente el problema de la predicción de series de tiempo ha sido resuelto a partir de un conjunto de avances en diferentes áreas como son: la Estadística, la Econometría, el Análisis de Fourier, la Teoría de Sistemas Dinámicos No Lineales, etc.; donde estas áreas son usadas para la formalización matemática y el desarrollo de técnicas de análisis, modelado y predicción de las series de tiempo. También la predicción de series de tiempo se ha beneficiado de la Computación tanto para la implementación de las técnicas y algoritmos desarrollados en las otras áreas, como para la aplicación y el desarrollo de nuevas soluciones de análisis y predicción de las series.

Finalmente, los resultados de la revisión del estado del arte sobre la predictibilidad de series de tiempo, dan soporte a la investigación desarrollada en este trabajo de tesis en la cual:

- Se han construido y evaluado una serie de métricas de predictibilidad que: dependen de las series de tiempo, las métricas consideran un conjunto de diferentes parámetros dinámicos de las series para calcular la predictibilidad, las métricas se pueden calcular en forma experimental.
- Las métricas se evaluaron utilizando un conjunto amplio de series de tiempo de diferente origen y además se compararon con métricas adicionales.
- Se estudiaron diferentes aspectos del comportamiento de la predictibilidad como son: la relación entre la predictibilidad y el comportamiento dinámico de las series de tiempo, la relación entre la predictibilidad y la estructura de las series de tiempo, la relación entre la

predictibilidad y la predicción con diferentes modelos y la relación entre la predictibilidad y la modelación con diferentes modelos.

## Capítulo 4

# Técnicas para el Análisis de Series de Tiempo

### 4.1 Introducción

El análisis de series de tiempo permite obtener información sobre el comportamiento de los datos que conforman a una serie. La información es utilizada para caracterizar, clasificar o extraer reglas, y predecir o construir modelos sobre la serie de tiempo. En el presente capítulo se describen las bases de las diferentes técnicas de análisis de series de tiempo utilizadas en este trabajo, así como la descripción de los parámetros (*de tipo escalar*) calculados con estas técnicas y que caracterizan a las series de tiempo, a partir de este conjunto de parámetros posteriormente serán seleccionados aquellos que sean más útiles en la caracterización de la predictibilidad de las series de tiempo. La sección 4.2 describe los parámetros de origen Estadístico, la sección 4.3 el Análisis de Fourier, la sección 4.4 el Análisis Espectral Singular, la sección 4.5 los parámetros basados en la Teoría de Sistemas Dinámicos No Lineales, la sección 4.6 los parámetros cuyo origen proviene de Teoría de la Información, finalmente la sección 4.7 presenta los parámetros basados en Teoría de la Computación.

## 4.2 Estadísticas

Los métodos de análisis de series de tiempo basados en estadística clásica y moderna han sido ampliamente utilizados en especial para el estudio de series de tiempo de tipo económico y social.

### 4.2.1 Análisis Estadístico Clásico

A partir del análisis estadístico clásico se obtienen un conjunto de parámetros que caracterizan la distribución estadística de un conjunto de datos, en particular estos pueden pertenecer a una serie de tiempo, a continuación se describen los parámetros estadísticos que fueron evaluados con el objetivo de determinar su utilidad en la caracterización de la predictibilidad de series de tiempo:

- 1 Valor promedio. Conocido también como media aritmética o media, se define como la suma del conjunto  $N$  de datos de la muestra dividida entre el valor  $N$  [133]. Su expresión matemática es:

$$\langle x \rangle = \frac{x_1 + x_2 + x_3 + \dots + x_N}{N} = \frac{1}{N} \sum_{k=1}^N x_k \quad (4.1)$$

- 2 Valor máximo. Es el dato con el valor más grande en la muestra de datos [133].
- 3 Valor mínimo. Es el dato con el valor más pequeño en la muestra de datos [133].
- 4 Valor medio. También llamado Mediana, se define como el valor del dato central (si el número de datos  $N$  es impar) o la media aritmética (promedio) de los dos datos centrales (si el número de datos  $N$  es par) [133].
- 5 Moda (máxima probabilidad). Es una medida de posición para la muestra de datos, se define como el dato (o los datos) que tienen la máxima frecuencia (es decir la máxima ocurrencia de los mismos) [133].
- 6 Rango de datos. Es la diferencia entre el menor valor y el mayor valor de la muestra de datos [134].

7 Desviación promedio. Es una medida de precisión (o imprecisión) también llamada desviación media. Es el valor promedio de los valores absolutos de las desviaciones de la media [134]. Su expresión matemática es:

$$DesviaciónPromedio = \frac{\sum_{k=1}^N |x_k - \langle x \rangle|}{N} \quad (4.2)$$

8 Desviación Estándar. Es una medida de la dispersión alrededor de la media. En una distribución normal el 68% de los casos caen dentro de una desviación estándar de la media y el 95% de los casos caen dentro de dos desviaciones estándar [135]. Su expresión matemática es:

$$\sigma = \sqrt{E[(x - E(x))^2]} \quad (4.3)$$

donde  $E(x)$  es el valor esperado de una variable definido como:

$$E(x) = \sum_k x_k p_k \quad (4.4)$$

donde  $p_k$  es la probabilidad de la variable  $x_k$ . Al cuadrado de la desviación estándar se le conoce como Varianza.

9 Correlación de Pearson (Auto correlación). La correlación mide como las variables o un rango de órdenes están relacionados, en particular el coeficiente de Pearson es una medida de asociación lineal [135].

10 Skewness. Mide el grado de asimetría de una distribución de datos respecto a su valor central. La distribución normal es simétrica y su Skewness es cero, una distribución de datos con un valor positivo de Skewness tiene una cola larga a la derecha. Una distribución de datos con un valor negativo de Skewness tiene una cola larga a la izquierda. Un Skewness con valor mayor a 1 generalmente indica una distribución que difiere significativamente de la distribución normal [134].

- 11 Kurtosis. Es una medida sobre la extensión de un conjunto de datos que se agrupan alrededor de un valor central. Para una distribución normal, el valor de la kurtosis es cero. Una kurtosis positiva indica que los datos se agrupan más y la distribución tiene colas más largas que las correspondientes a una distribución normal, una kurtosis negativa indica que los datos se agrupan menos y la distribución posee colas más cortas que una distribución normal [134].
- 12 Cuartil inferior. Es el dato cuyo valor marca la división del grupo de datos ordenados que corresponde al 25% de todos los datos de la muestra [134].
- 13 Cuartil superior. Es el dato cuyo valor marca la división del grupo de datos ordenados que corresponde al 75% de todos los datos de la muestra [134].
- 14 Resolución. Es el promedio de la diferencia entre los valores de datos consecutivos de la muestra.
- 15 Tiempo de Correlación. Es el intervalo de tiempo correspondiente a la correlación entre una variable y otra variable o consigo misma (auto correlación). Por ejemplo, el tiempo de auto correlación se obtiene de graficar la función de correlación de una misma variable para tiempos diferentes  $x(t)$  y  $x(t-\tau)$  variando  $\tau$ , el tiempo de autocorrelación es donde el valor de  $\tau$  corresponde a la función de correlación cuando decae a  $1/e$  de su valor original.

## 4.2.2 Análisis Estadístico Moderno

### Análisis R / S (Exponente de Hurst)

H. E. Hurst, hidrólogo británico, publicó en 1951 un trabajo titulado “*The Long-Term Storage Capacity of Reservoirs*”. Este trabajo trata sobre el diseño de modelos de reservorios, pero Hurst extendió su trabajo a muchos sistemas naturales y desarrolló una nueva metodología estadística para distinguir sistemas aleatorios y no aleatorios, su método se llama rango reescalado o análisis R / S [136].

Hurst encontró que a diferencia de la hipótesis asumida por la mayoría de los hidrólogos de su tiempo en la cual para calcular la capacidad de almacenamiento de agua de reservorios se estimaba el flujo de agua entrante como un proceso aleatorio, el flujo de agua en el río Nilo posee

un patrón de comportamiento el cual no es aleatorio. Hurst conocía el trabajo de Einstein sobre el movimiento browniano el cual nos dice que la distancia que cubre una partícula aleatoria se incrementa como la raíz cuadrada del tiempo:

$$R = T^{\frac{1}{2}} \quad (4.5)$$

donde  $R$  = a la distancia cubierta y  $T$  = al índice del tiempo. Esta ecuación se le conoce como la regla de  $T$  a la un medio, y es usada comúnmente en estadística.

Ahora si se tiene una serie de tiempo  $\{x(t)\}$  que representa  $n$  valores consecutivos. En el caso que desarrolló Hurst, la serie correspondía a las descargas anuales del río Nilo. Para el caso de mercados financieros, puede ser los cambios diarios en el precio del índice de acciones. El valor medio,  $\bar{x}$  de la serie de tiempo se define como:

$$\bar{x} = \frac{(x(1) + \dots + x(n))}{n} \quad (4.6)$$

La desviación estándar,  $s_n$  se estima como:

$$s_n = n^{-\frac{1}{2}} \sqrt{\sum_{t=1}^n (x(t) - \bar{x})^2} \quad (4.7)$$

El rango reescalado se calcula mediante la normalización de los datos sustrayendo la media muestreada:

$$Z(t) = (x(t) - \bar{x}) \text{ con } t = 1, \dots, n \quad (4.8)$$

La serie resultante posee una media de cero. Ahora se crea una serie acumulativa  $X$  con valores de la forma:

$$X_1 = (Z_1 + Z_r) \quad r = 2, \dots, n \quad (4.9)$$

Nótese que por definición, el último valor de  $X$  (que es  $X_n$ ) siempre será cero ya que  $Z$  tiene una media de cero. El rango ajustado,  $R_n$ , es el máximo menos el valor mínimo de  $X_r$ :

$$R_n = \max(Y_1, \dots, Y_n) - \min(Y_1, \dots, Y_n) \quad (4.10)$$

El subíndice  $n$  para  $R_n$  ahora indica que éste es el rango ajustado para  $x_1, \dots, x_n$ . Dado que  $X$  ha sido ajustado para tener media cero, el valor máximo de  $X$  siempre es mayor o igual a cero, y el mínimo siempre será menor o igual a cero. Luego, el rango ajustado  $R_n$  es siempre no negativo.

El rango ajustado  $R_n$  es la distancia que el sistema viaja para el índice de tiempo  $n$ . Si se hace  $n = T$ , se puede aplicar la ecuación 4.5, considerando que la serie de tiempo  $x$  es independiente para valores crecientes de  $n$ . Sin embargo, la ecuación 4.5, se aplica sólo a series de tiempo que poseen las características de movimiento Browniano: media cero y varianza igual a uno. Para extender este concepto a otros tipos de series, es necesario generalizar la ecuación 4.5 para tomar en cuenta sistemas que no son independientes. Hurst encontró que la forma más general de la ecuación es:

$$(R/S)_n = c * n^H \quad (4.11)$$

El subíndice  $n$  de  $(R/S)_n$  se refiere al valor  $R/S$  para  $x_1, \dots, x_n$ ; con  $c =$  constante.

El valor  $R/S$  se conoce como el rango reescalado ya que posee media cero y se expresa en términos de la desviación local estándar  $S$ . En general el valor  $R/S$  escala conforme aumenta el incremento de tiempo  $n$  como una ley de potencia con un valor  $H$ , el cual corresponde al exponente de Hurst. El exponente de Hurst se encuentra graficando al  $\log(R/S)$  vs.  $\log(n)$  y resolviendo la pendiente por medio de una regresión de mínimos cuadrados, en resumen resolviendo la ecuación:

$$\log(R/S)_n = \log(c) + H * \log(n) \quad (4.12)$$

Este exponente permite determinar si el comportamiento representado por la serie de tiempo presenta correlaciones de largo alcance (es decir memoria y persistencia de largo alcance de su comportamiento). En el caso de que la serie de tiempo posea un comportamiento con persistencia de tendencia positiva entonces el exponente de Hurst es mayor a 0.5, si no hay comportamiento predecible el exponente es igual a cero, por último si hay un comportamiento

con persistencia de tendencia negativa entonces el exponente es mayor que cero y menor a 0.5.

### **4.3 Análisis de Fourier**

El análisis de Fourier como se mencionó en el capítulo 3 fue la primera metodología utilizada en el análisis de series de tiempo. Aún cuando la fuerza del mismo radica en el estudio del comportamiento de señales de tipo periódico, es interesante el explorar la capacidad de algunos de los parámetros derivados de este análisis para evaluar su utilidad en la caracterización de series que poseen comportamientos no periódicos. El análisis de Fourier trata con el estudio de las series de tiempo en el dominio de las frecuencias, basados en la idea de que sobre un intervalo finito, cualquier función analítica puede ser aproximada con el grado de exactitud deseado, por medio de una suma pesada de funciones seno o coseno de frecuencias crecientes armónicamente.

Del conjunto de métodos del análisis de Fourier se seleccionó la búsqueda de frecuencias dominantes, ya que proporciona a partir del análisis del espectro de potencia, no sólo información de tipo gráfico sobre el espectro de las frecuencias de la serie de tiempo sino además, un parámetro que corresponde a la frecuencia máxima (que llamamos frecuencia dominante), la cuál es una medida escalar útil junto con otros parámetros para la búsqueda de un conjunto o conjuntos de parámetros que permitan caracterizar la predictibilidad en las series de tiempo [73].

#### **4.3.1 Frecuencia Dominante**

La frecuencia dominante se obtiene a partir del cálculo del espectro de potencias ya sea por medio de la transformada rápida de Fourier o el método de máxima entropía, permite determinar si existe alguna frecuencia característica de la señal, las series de carácter aleatorio poseen espectros de potencia amplios sin ninguna frecuencia dominante, lo mismo se aplica en cierto grado para las series caóticas. En el caso de las series periódicas y cuasi periódicas éstas poseen espectros con picos bien definidos representativos de sus frecuencias características.

## 4.4 Análisis Espectral Singular (SSA)

El Análisis Espectral Singular (*SSA*) es una técnica para el análisis de series de tiempo que incorpora elementos del análisis clásico de series de tiempo (i.e. Análisis de Fourier, Análisis Estadístico) así como elementos modernos (i.e. Teoría de Sistemas Dinámicos). Esta técnica permite la identificación y extracción de componentes oscilatorias de una serie de tiempo lo que es de utilidad para el estudio de su *estructura*. La descomposición de la serie de tiempo consiste en componentes aditivas que se interpretan como: componentes de tendencias (*trends*) es decir partes suaves y de variación lenta de la serie, componentes oscilatorias (quizás con diferentes amplitudes) y componentes de ruido (pero no en el sentido estocástico ya que sólo se trata con una trayectoria simple de la serie y no una muestra de trayectorias). El origen de la técnica se asocia a la publicación de una serie de artículos por Broomhead y King (1986) y Broomhead et. al. (1987) [137]. La técnica SSA es utilizada en el análisis de series climáticas, meteorológicas y geofísicas. En particular, en el estudio sobre la predictibilidad de las series de tiempo, es de interés explorar la técnica SSA en combinación con el análisis de componentes principales (que se describe en la subsección 4.4.1), con el objetivo de identificar si existen relaciones entre la *estructura* de las componentes de las series de tiempo y la predictibilidad. La versión básica de SSA consiste de 4 pasos que se describen a continuación.

Primer paso (etapa de descomposición), sea  $X = (x(0), x(1), \dots, x(N-1))$  una serie de tiempo de longitud  $N$ , y  $L$  sea un entero, el cual se le llama longitud de ventana. Sea  $K = N-L+1$  y se definen los vectores de retardo (*embebido*)  $X_j = (x_{j-1}, \dots, x_{j+L-2})^T$ ,  $j = 1, 2, \dots, K$  y la matriz de trayectoria

$$\mathbf{X} = (\mathbf{x}_{i+j-2})_{i,j=1}^{L,K} = [X_1, \dots, X_K] \quad (4.13)$$

La matriz  $\mathbf{X}$  es una matriz de Hankel, lo que significa que todos los elementos a lo largo de la diagonal  $i + j = \text{constante}$ , son iguales. La construcción de la matriz de trayectoria es el primer paso del algoritmo de SSA.

El segundo paso (etapa de descomposición), es la descomposición de valor singular (SVD) de la matriz  $\mathbf{X}$ , la cual se logra por medio de los eigenvalores y eigenvectores de la matriz  $\mathbf{S} = \mathbf{X}\mathbf{X}^T$  de tamaño  $L \times L$ . Esto proporciona una colección de  $L$  valores singulares, los cuales

son las raíces cuadradas de los eigenvalores de la matriz  $\mathbf{S}$ , y los correspondientes vectores singulares izquierdo y derecho. Los vectores singulares izquierdo de  $\mathbf{X}$  son los eigenvectores ortonormales de  $\mathbf{S}$  llamados también *funciones empíricas ortogonales*, los vectores singulares derecho son los eigenvectores de la matriz  $\mathbf{X}^T \mathbf{X}$ . Se obtiene entonces una representación de  $\mathbf{X}$  como una suma de rango uno de matrices biortogonales  $\mathbf{X}_i$  con  $i = 1, \dots, d$ , donde  $d \leq L$  es el número de valores singulares de  $\mathbf{X}$  diferente de cero.

Como el tercer paso (etapa de reconstrucción), se separa el conjunto de índices  $I = \{1, \dots, d\}$  en varios grupos  $I_1, \dots, I_m$  y se suman las matrices  $\mathbf{X}_i$  dentro de cada grupo. El resultado de este paso es la representación:

$$\mathbf{X} = \sum_{k=1}^m \mathbf{X}_{I_k}, \quad \text{donde} \quad \mathbf{X}_{I_k} = \sum_{i \in I_k} \mathbf{X}_i. \quad (4.14)$$

El cuarto paso ( etapa de reconstrucción), consiste en promediar sobre las diagonales  $i + j = \text{constante}$ , de las matrices  $\mathbf{X}_{I_k}$ . Esto proporciona la descomposición SSA, esto es, la descomposición de la serie original  $X$  en una suma de series:

$$x_n = \sum_{k=1}^m x_n^{(k)} \quad \text{con} \quad n = 0, \dots, N - 1, \quad (4.15)$$

donde para cada  $k$  la serie  $x_n^{(k)}$  es el resultado del promediado sobre las diagonales de la matriz  $\mathbf{X}_{I_k}$ .

#### 4.4.1 Análisis de Componentes Principales

Relacionado con la descomposición de la serie de tiempo por parte de la técnica SSA, se encuentra el análisis de componentes principales (*PCA*), una vez que la serie a sido separada en las diferentes *series componentes* se les aplica el PCA como una herramienta de análisis complementaria, a continuación se presenta la descripción de dicho análisis. El análisis de componentes principales, consiste en la búsqueda de las componentes linealmente independientes que proyectan a un conjunto de datos multivariado (i. e. multidimensional) en diferentes direcciones en un espacio de dos dimensiones (aunque no está limitado a este número de dimensiones), estas componentes deben cumplir las siguientes condiciones: las combinaciones lineales de las componentes proporcionan la máxima varianza de la muestra de datos, las combinaciones

lineales deben estar no correlacionadas entre sí. En resumen el objetivo del análisis de componentes principales es la reducción de la dimensionalidad del conjunto de datos para facilitar su interpretación [72].

Para calcular las componentes principales se procede de la siguiente forma, sea  $\mathbf{X}$  una matriz de datos  $n \times p$  en la cual las filas representan casos (cada fila es un vector de datos  $\mathbf{x}(i)$ ) y las columnas representan las variables. Estrictamente la  $i$  – esima fila de esta matriz es la transpuesta  $\mathbf{x}^T$  del  $i$  – esimo vector de datos  $\mathbf{x}(i)$ , dado que la convención que se toma es considerar vectores de datos como vectores columnas  $p \times 1$  en lugar de vectores filas  $1 \times p$ . Además se asume que  $\mathbf{X}$  es centrada a la media, de forma que el valor de cada variable es relativo a la media muestreada para esa variable (i. e. la media estimada ha sido sustraída para cada columna).

Sea  $\mathbf{a}$  el vector columna  $p \times 1$  de los pesos de proyección (no conocidos hasta ahora) que resultan en la mayor varianza cuando los datos  $\mathbf{X}$  son proyectados a lo largo de  $\mathbf{a}$ . La proyección de cualquier vector de datos particular  $\mathbf{x}$  es la combinación lineal  $\mathbf{a}^T \mathbf{x} = \sum_{j=1}^p a_j x_j$ . Nótese, que se puede expresar a los valores proyectados sobre  $\mathbf{a}$  de todos los otros vectores de datos en  $\mathbf{X}$  como  $\mathbf{X}\mathbf{a}$  ( $n \times p$  por  $p \times 1$  obteniéndose un vector columna de valores proyectados  $n \times 1$ ). La varianza a lo largo de  $\mathbf{a}$  se define como

$$\sigma_{\mathbf{a}}^2 = (\mathbf{X}\mathbf{a})^T (\mathbf{X}\mathbf{a}) = \mathbf{a}^T \mathbf{X}^T \mathbf{X} \mathbf{a} = \mathbf{a}^T \mathbf{V} \mathbf{a} \quad (4.16)$$

donde  $V = \mathbf{X}^T \mathbf{X}$  es la matriz  $p \times p$  de covarianza de los datos (dado que  $\mathbf{X}$  tiene media cero). La varianza de los datos proyectados  $\sigma_{\mathbf{a}}^2$  (escalar a maximizar) se puede expresar como una función de  $\mathbf{a}$  y de la matriz de covarianza de los datos  $\mathbf{V}$ .

Para maximizar  $\sigma_{\mathbf{a}}^2$  es necesario aplicar una restricción de normalización sobre los vectores  $\mathbf{a}$  la cual es:  $\mathbf{a}^T \mathbf{a} = 1$ .

Entonces el problema de optimización se describe como la maximización de la cantidad

$$u = \mathbf{a}^T \mathbf{V} \mathbf{a} - \lambda (\mathbf{a}^T \mathbf{a} - 1) \quad (4.17)$$

donde  $\lambda$  es un multiplicador de Lagrange. Diferenciando con respecto a  $\mathbf{a}$  se tiene

$$\frac{\partial u}{\partial \mathbf{a}} = 2\mathbf{V}\mathbf{a} - 2\lambda\mathbf{a} = 0 \quad (4.18)$$

que se reduce a la forma familiar de eigenvalores

$$(\mathbf{V} - \lambda\mathbf{I})\mathbf{a} = 0 \quad (4.19)$$

Entonces el primer componente principal  $\mathbf{a}$  es el eigenvector asociado con el eigenvalor más grande de la matriz de covarianza  $\mathbf{V}$ . La segunda componente principal (la dirección ortogonal a la primera componente que tiene la varianza proyectada más grande) es el eigenvector correspondiente al segundo eigenvalor más grande de  $\mathbf{V}$ , y así sucesivamente (el eigenvector para el  $k$ -ésimo eigenvalor más grande corresponde a la  $k$ -ésima dirección componente principal). Una propiedad básica de este esquema de proyección es que si los datos están proyectados en los primeros  $k$  eigenvectores, la varianza de los datos proyectados se expresa como  $\sum_{j=1}^k \lambda_j$ , donde  $\lambda_j$  es el  $j$ -ésimo eigenvalor.

Una característica que se observa es que para conjuntos de datos de alta dimensionalidad, en los cuales las variables están frecuentemente bien correlacionadas, es común encontrar que para un número relativamente pequeño de componentes principales (por decir 5 a 10) se captura el 90% o más de la varianza en los datos.

## 4.5 Teoría de Sistemas Dinámicos No Lineales

A diferencia de los métodos lineales (basados en teoría estadística, análisis de Fourier) los cuales interpretan las estructuras regulares en los datos como correlaciones lineales, es decir que la dinámica del sistema se basa en lo siguiente: ante pequeñas perturbaciones la respuesta del sistema es pequeña, y que el comportamiento irregular (diferente a oscilaciones periódicas o a un crecimiento exponencial) del sistema se atribuye a una perturbación externa de carácter aleatorio; las técnicas de análisis no lineal (Teoría de Sistemas Dinámicos No Lineales o Teoría de Caos) toman en consideración que sistemas no lineales caóticos pueden producir comportamiento irregular de los datos a partir de un sistema de ecuaciones deterministas. Este comportamiento irregular es interpretado como ruido por los métodos de análisis lineal. Las

técnicas de análisis no lineal se pueden agrupar en las siguientes categorías: Reconstrucción del Espacio Fase, Dimensiones, Exponentes de Lyapunov, Entropías, de Datos Sustitutos (Surrogate Data) y Reducción del Ruido [3, 138, 139].

### 4.5.1 Exponentes

#### Exponente de Lyapunov

Los exponentes de Lyapunov son medidas de la deformación en el espacio fase representada por los movimientos relativos de trayectorias cercanas de un atractor (donde la serie de tiempo es una proyección en 1-D de la dinámica del mismo), miden la evolución de trayectorias vecinas en el espacio fase. En otras palabras miden la inestabilidad de la dinámica del sistema debido a cambios en sus condiciones iniciales. El número de exponentes corresponde al número de dimensiones del atractor. El exponente principal de Lyapunov, que corresponde al exponente de mayor magnitud se considera una medida promedio de la sensibilidad del sistema a cambios en las condiciones iniciales y por ende a las desviaciones de las trayectorias del atractor, es una medida de la predictibilidad del sistema [138]. El cálculo de este exponente se basa en construir una vecindad en el espacio fase de tipo esferoidal que posee un conjunto de ejes principales  $p_i$ . Esta vecindad contiene a un conjunto de puntos de diferentes trayectorias del atractor del sistema, la vecindad se mueve siguiendo la evolución de las trayectorias, para seguir esta evolución la vecindad sufre deformaciones las cuales producen cambios en las magnitudes de los ejes que la definen, los exponentes de Lyapunov se determinan por el ritmo de cambio de la magnitud de los ejes principales de la vecindad, se les llama exponentes porque el ritmo de cambio sigue un comportamiento de tipo exponencial. Su expresión matemática es de la siguiente forma:

$$b^{t\lambda_i(t)} = \frac{p_i(t)}{p_i(0)} \quad (4.20)$$

donde  $b$  es una constante arbitraria pero determinada. Tomando logaritmos se obtiene:

$$\lambda_i(t) = \frac{1}{t} \text{Log}_b \frac{p_i(t)}{p_i(0)} \quad (4.21)$$

El valor límite de esta expresión cuando  $t$  crece es el exponente de Lyapunov correspondiente

al  $i$  – *esimo* eje.

## 4.5.2 Dimensiones

### Dimensiones de tipo exponente

Una dimensión exponente frecuentemente permanece constante sobre un rango de escalas de tamaños, lo cual la hace una medida cuantitativa invariante. Este tipo de medida proporciona información sobre que tan complejo es un sistema. Es una medida cuantitativa de evaluar o comparar la complejidad geométrica de objetos o sistemas de diferente tamaño, forma, y estructura. Hay muchos tipos de dimensiones expresadas como un exponente de escalamiento. No existe un consenso sobre cuales de estas dimensiones son las mas útiles para el estudio de las series de tiempo [139].

**Dimensión de Correlación** En 1983 Grassberger y Procaccia, introdujeron una dimensión basada en el comportamiento de la suma o integral de correlación. Esta dimensión se le llama Dimensión de Correlación y es ampliamente usada en la caracterización de atractores caóticos. Para definirla se toma una trayectoria (de un atractor) que evolucione por un largo tiempo, y se toman los valores de  $N$  puntos de la trayectoria. Luego para cada punto  $i$  sobre la trayectoria, se busca el número de puntos de trayectoria dentro de una distancia  $R$  del punto  $i$  (excluido el mismo), a este número lo llamamos  $N_i(R)$ . Luego se define  $p_i(R)$  como el número relativo de puntos dentro de la distancia  $R$  al punto  $i$ :  $p_i(R) = \frac{N_i}{(N-1)}$  (se divide por  $N - 1$ , pues es el número de puntos que como máximo se encuentran en la vecindad quitando a  $i$ ). Finalmente se calcula la llamada suma de correlación

$$C(R) = \frac{1}{N} \sum_{i=1}^N p_i(R) \quad (4.22)$$

Esta definición cumple que  $C(R) = 1$  si todos los puntos caen dentro de la distancia  $R$  respecto de cada uno de ellos. Si  $R$  es menor que la distancia más pequeña entre puntos de la trayectoria, entonces  $p_i = 0$  para todo  $i$  y  $C(R) = 0$ .

Aplicando la función de paso Heaviside:

$$\begin{aligned}\Theta(x) &= 0 \quad \text{si } x < 0 \\ \Theta(x) &= 1 \quad \text{si } x \geq 0\end{aligned}\tag{4.23}$$

se reescribe al número relativo  $p_i(R)$  como

$$p_i(R) = \frac{1}{N-1} \sum_{j=1, j \neq i}^N \Theta(R - |x_i - x_j|)\tag{4.24}$$

entonces la suma de correlación toma la forma

$$C(R) = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1, j \neq i}^N \Theta(R - |x_i - x_j|)\tag{4.25}$$

Para asegurar que se caracteriza al atractor entero se toma  $N \rightarrow \infty$ , la dimensión de correlación se define entonces como el número que satisface

$$C(R) = \lim_{R \rightarrow 0} kR^{D_c}\tag{4.26}$$

o tomando logaritmos

$$D_c = \lim_{R \rightarrow 0} \frac{\log C(R)}{\log R}\tag{4.27}$$

La dimensión de correlación cuantifica la correlación espacial local entre puntos de la trayectoria del atractor en el espacio fase [140].

**Dimensión de Capacidad** Para estimar la dimensión de objetos irregulares, por ejemplo fractales, es posible acercarse a la medida exacta de los mismos utilizando métricas que escalan con el comportamiento de estos objetos, para indicar esta posibilidad es que se utilizan relaciones de la forma

$$D_{cap} = \lim_{\varepsilon \rightarrow 0} \frac{\log N}{\log(1/\varepsilon)}\tag{4.28}$$

donde  $N$  es el número de subsecciones en las que se particiona la longitud del sistema y  $\varepsilon$  es la regla con la que se mide a dicha longitud. A esta dimensión se le denomina la capacidad [141], también se le llama la dimensión de capacidad o la capacidad límite. Técnicamente, la

capacidad se define en el rango de longitudes extremadamente pequeñas  $\varepsilon$ , específicamente en el límite  $\varepsilon \rightarrow 0$ .

La dimensión de capacidad mide el grado de auto-similitud del sistema (comportamiento invariante ante cambios de escala espacial), permite cuantificar el grado de heterogeneidad de la señal a diferentes escalas [139].

**Dimensión Fractal** No existe un consenso en la literatura sobre una sola dimensión fractal, sino mas bien que el conjunto de dimensiones de la forma de exponente forman una familia de dimensiones. Para la caracterización de las series de tiempo, en este trabajo se utilizó la siguiente definición de dimensión fractal denominada en la literatura como  $D_2$

$$D_2 = \lim_{R \rightarrow 0} \frac{\log N}{\log R} \quad (4.29)$$

donde la dimensión fractal mide el número de estados  $N$  del sistema presentes en un volumen de radio  $R$  [139, 140].

### Dimensión Embebida

Aunque la dimensión embebida no corresponde a la familia de dimensiones de tipo exponencial como las definidas en los párrafos anteriores, se sitúa de igual forma en la categoría más general de dimensión. La dimensión embebida, corresponde a la dimensión necesaria para la reconstrucción en el espacio fase del atractor, que forman las trayectorias de los estados de la dinámica del sistema. Dicha dimensión se obtiene a partir de una serie de tiempo del sistema, mediante la aplicación de las ideas establecidas por el teorema de Takens. El valor de la dimensión embebida corresponde de igual forma con el número de grados de libertad del sistema dinámico [138].

### 4.5.3 Análisis de Mapas de Recurrencia (VRA)

Los mapas de recurrencia fueron descritos por primera vez por J. P. Eckmann, S. O. Kamphorst y D. Ruelle en “*Recurrence plots of dynamical systems*” en 1987 [8, 142]. Este análisis cualitativo (vía la visualización del mapa) y cuantitativo (vía el cálculo de parámetros derivados del mapa) permite la detección de *patrones ocultos y cambios estructurales* de los datos que

conforman una serie de tiempo, así como la *identificación de similitudes* entre diferentes series de tiempo bajo estudio, más adelante en el capítulo 10, se explota la información visual de los mapas de recurrencia para el análisis de la relación entre estructura y predictibilidad de las series de tiempo. La idea básica que sustenta a los mapas de recurrencia es que una serie de tiempo es el producto de un proceso dinámico donde interaccionan las variables relevantes del mismo en el tiempo. Y que es posible recuperar la información sobre este proceso multivariado a partir de la serie de tiempo de una variable medida experimentalmente [63]. En el caso del análisis VRA, una serie de tiempo se expande a un espacio multidimensional, en la cual la dinámica del sistema tiene lugar. Para ello se aplica el embebido de coordenadas de retardo, con el cual se reproduce el espacio fase del sistema dinámico bajo estudio. Para realizar la expansión de la señal al espacio fase  $M$ -dimensional, cada observación de la señal original  $X(t)$  se sustituye con un vector

$$y(1) = \{x(i), x(i - d), x(i - 2d), \dots, x(i - (m - 1)d)\} \quad (4.30)$$

donde  $i$  corresponde al índice del tiempo,  $m$  es la dimensión embebida y  $d$  es el tiempo de retardo.

Como resultado se obtiene una serie formada por vectores:

$$Y = \{y(1), y(2), y(3), \dots, y(N - (m - 1)d)\} \quad (4.31)$$

donde  $N$  es la longitud de la series original.

La idea de esta reconstrucción es el capturar los estados del sistema original en cada instante de tiempo. Cada estado  $S(t)$  a un tiempo  $t$  es aproximado por un vector de coordenadas de retardo como el de la ecuación 4.30.

Una vez reconstruida la dinámica del sistema, un mapa de recurrencia permite mostrar que vectores en el espacio fase reconstruido están más cercanos o lejanos entre sí. En forma específica, se calcula la distancia Euclidiana entre todos los pares de vectores y se codifican como colores. Esencialmente el mapa de recurrencia es una matriz de códigos de colores, donde cada entrada  $(i, j)$  corresponde a la distancia entre los vectores  $y(i)$  e  $y(j)$ . Esta distancia se asocia con un código de colores predefinido para su despliegue gráfico en la posición asociada

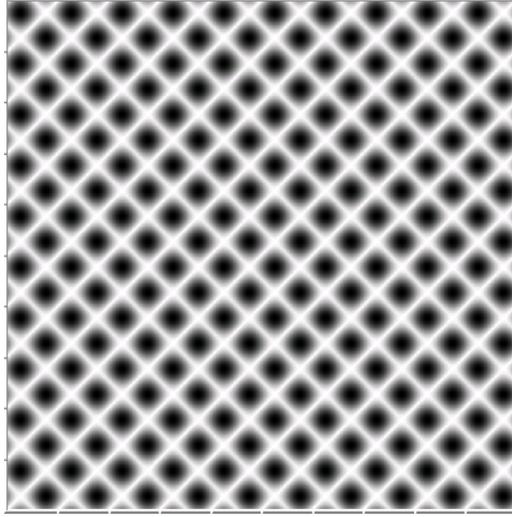


Figura 4-1: Mapa de recurrencia de la serie de tiempo Sine

de carácter *temporal*  $(i, j)$ . Los colores calientes (correspondientes al amarillo, rojo, y naranja o tonos de gris claro) indican distancias pequeñas entre los vectores, mientras que los colores fríos (azul o negro) indican distancias grandes entre los vectores. Esta codificación permite la visualización y estudio del movimiento de las trayectorias que forman los estados del sistema en el espacio fase, para inferir algunas de las características dinámicas que generan a la serie de tiempo original. Un mapa de recurrencia puede interpretarse como una representación gráfica de una integral de correlación. La diferencia importante (y su ventaja) es que a diferencia de las integrales de correlación, los mapas preservan la dependencia temporal en las series de tiempo, además de la dependencia espacial.

A continuación, se presentan tres ejemplos de la forma en que los mapas de recurrencia proporcionan información visual para la detección de patrones, en la Figura 4-1, se presenta el mapa de recurrencia de la serie periódica Sine, la cual es una señal que presenta una estructura definida en el mapa mostrando su carácter periódico.

En la Figura 4-2, se muestra el mapa de recurrencia de la serie de tiempo Rossler que presenta una dinámica caótica, la estructura del mapa de recurrencia es definida mostrando el carácter determinista de la señal.

La Figura 4-3, muestra el mapa de recurrencia para la serie de tiempo White Noise, esta

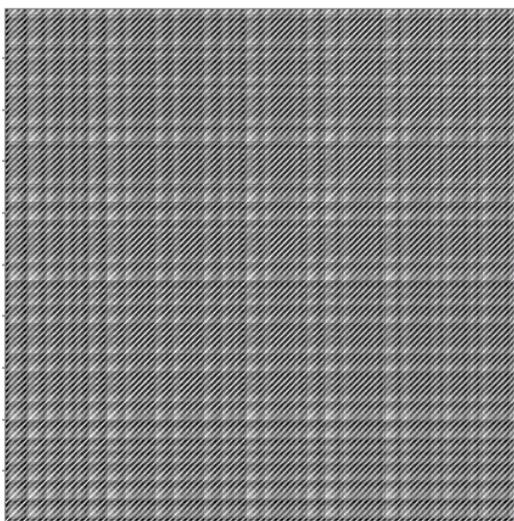


Figura 4-2: Mapa de recurrencia de la serie de tiempo Rossler

señal es aleatoria, la distribución de tonos de gris es uniforme en el mapa no observándose una estructura definida.

A partir de un mapa de recurrencia es posible calcular algunos parámetros para la caracterización de una serie de tiempo, los parámetros (utilizados en este trabajo) se describen a continuación.

### **Entropía Espacio-Temporal**

La entropía espacio-temporal, cuantifica lo estructurado del mapa de recurrencia en los dominios espacial y temporal. Esencialmente, compara la distribución global de colores sobre el mapa completo con respecto a la distribución de colores sobre cada diagonal del mapa. A mayor diferencia entre las distribuciones global y sobre las diagonales individuales, mayor es la estructura del mapa. Físicamente, esta cantidad compara la distribución de distancias entre todos los pares de vectores en el espacio fase reconstruido con las distancias entre las diferentes orbitas que evolucionan en el tiempo. El resultado es normalizado y se presenta como un porcentaje de máxima entropía (aleatoriedad). Una entropía del 100% significa ausencia de cualquier estructura, mientras que una entropía de 0% implica una estructura perfecta.

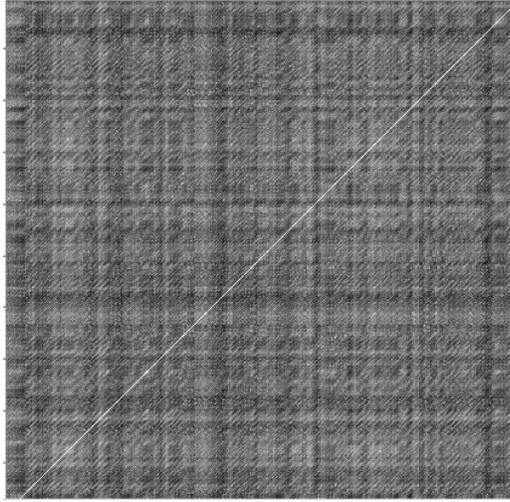


Figura 4-3: Mapa de recurrencia de la serie de tiempo White Noise

### Recurrencia

A partir del mapa de recurrencia se puede realizar el análisis de cuantificación de recurrencia o RQA [143, 144], el cual permite medir el porcentaje de recurrencia (periodicidad y estructura) en el mapa (y en consecuencia en la serie de tiempo original), que a su vez es indicativo de patrones repetitivos en la serie de tiempo, para calcular este porcentaje se toma el área triangular superior de la matriz de recurrencia, excluyendo las entradas que corresponden a distancias de separación de valor cero a lo largo de la diagonal central ( $i = j$ ). El mapa es simétrico a través de esta diagonal ( $\text{distancia}[i, j] = \text{distancia}[j, i]$ ). A continuación, se calcula el número total de puntos recurrentes en esta área ( $\#RECURS$ ) finalmente el porcentaje de recurrencia se obtiene con la expresión,

$$\%Recurrencia = \frac{\#RECURS}{\text{Área Triangular}} \quad (4.32)$$

### Determinismo

Este parámetro se obtiene también a partir del análisis de cuantificación de recurrencia, permite medir el porcentaje de determinismo en el sistema, se calcula con la siguiente fórmula,

$$\%Determinismo = \frac{\#PRecDiag}{\#RECURS} \quad (4.33)$$

donde  $\#PRecDiag$  corresponde al número de puntos recurrentes que forman líneas diagonales hacia arriba en el mapa de recurrencia, y el denominador corresponde a la misma definición usada en el cálculo de la recurrencia. En el caso de que el  $\#RECURS = 0$ , se asigna el valor de -1 al determinismo.

## 4.6 Teoría de la Información

### 4.6.1 Entropías

#### Entropía de Información (Shannon)

Desde un punto de vista operacional, la información ocurre como un mensaje o medida con una distribución de probabilidad *a priori*  $\{p_i\}$  sobre su forma o estructura. Se define entonces el contenido de información o entropía de Shannon de una distribución  $\{p_i\}$  como

$$H \equiv - \sum_i p_i \log_2 p_i \quad (4.34)$$

En el contexto de sistemas dinámicos,  $p_i$  es la probabilidad de observar al sistema en el estado  $i$ . Si la medida es *a priori* cierta, luego  $H = 0$ ; entonces no se aprende nada al observar al sistema. Por otra parte, si una medida tiene *a priori* la misma probabilidad de caer donde sea, por decir el intervalo  $[0, 1]$  y el intervalo es dividido en  $n$  partes medibles iguales, luego la probabilidad  $p_i = \frac{1}{n}$ , y el contenido de información de una medición es

$$H = \sum_i \frac{1}{n} \log_2 n = \log_2 n \quad (4.35)$$

Dado que se asume una distribución de probabilidades uniforme, la ecuación 4.35 es una frontera superior para la entropía. Conforme  $n \rightarrow \infty$ , la resolución crece, y el contenido de información de una medida se incrementa, pero sólo hasta el nivel definido por la exactitud de la medición. En resumen, la entropía de Shannon es una medida de la cantidad de información que se obtiene al observar un sistema y tomar una medida para especificar su estado [138, 145].

## Información Mutua Promedio

Dos sistemas que no tienen relación mutua no poseen información uno del otro. En forma recíproca, si poseen alguna conexión, cada uno posee información sobre el otro. Por ejemplo, una medición de  $x$  ayuda a estimar  $y$  (o reduce la incertidumbre sobre  $y$ ). Para determinar la cantidad por la cual una medida de  $x$  reduce la incertidumbre sobre el valor  $y$ , se debe tomar en cuenta lo siguiente, en primer lugar  $y$  por si misma tiene una incertidumbre que se mide como la auto entropía  $H_y$ , en segundo lugar, hay una incertidumbre de  $y$  dada por una medida de  $x$ , medida por la entropía condicional  $H_{y|x}$ . La entropía condicional  $H_{y|x}$  es un número que representa una cantidad de información sobre  $y$ . En particular, la incertidumbre básica  $H_y$  es disminuida por una cantidad igual a  $H_{y|x}$ . Lo anterior, significa que el decremento en la incertidumbre es  $H_y - H_{y|x}$ . La incertidumbre reducida se le conoce como información mutua,

$$I_{y;x} = H_y - H_{y|x} \quad (4.36)$$

la expresión detallada es

$$I_{y;x} = \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} P(x_i, y_j) \log_2 \frac{P(x_i, y_j)}{P(x_i)P(y_j)} \quad (4.37)$$

donde  $N_s$  corresponde al número de probabilidades diferentes a cero,  $P(x_i, y_j)$  es la probabilidad de ocurrencia de  $x_i$  dado  $y_j$ ,  $P(x_i)$  y  $P(y_j)$  son las probabilidades para  $x_i$  e  $y_j$ .

En el caso particular de las series de tiempo, la información mutua mide el promedio de información que poseen en común una medición en el instante  $t_1$  respecto a una medición en el instante  $t_2$  [138, 139].

## 4.7 Teoría de la Computación

### 4.7.1 Complejidad Relativa LZ

Es una medida de la complejidad algorítmica que mide el número de nuevas subcadenas descubiertas conforme una secuencia evoluciona de izquierda a derecha. Cada nueva subcadena incrementa en 1 la complejidad. Esta medida esencialmente toma en cuenta las repeticiones de patrones a todos los niveles estructurales, capturando no sólo la complejidad de la dinámica de

la cadena sino también su complejidad jerárquica [126, 146]. El cálculo de este parámetro se describe a continuación, la fuente de la secuencia  $S = s_1 s_2 s_3 \dots$  es revisada y rescrita como una concatenación  $w_1 w_2 w_3 \dots$  de palabras  $w_k$  seleccionadas de tal forma que  $w_1 = s_1$  y  $w_{k+1}$  es la palabra más corta que no ha aparecido previamente. Por ejemplo, la cadena 0011101001011011 es transformada como 0.01.1.10.100.101.1011. Una vez que una nueva palabra ha sido identificada, la señal es nuevamente inspeccionada, con la adición de un símbolo más, si el bloque actual ésta ya presente en la lista. De esta forma, cada palabra  $w_{k+1}$  es la extensión de alguna palabra  $w_j$  en la lista, donde  $0 \leq j \leq k$  y, por convención,  $w_0$  es la palabra vacía  $\epsilon$ . Dado que  $w_{k+1} = w_j s$ , con  $s$  el símbolo de entrada actual,  $w_{k+1}$  se codifica como el par  $(j, s)$ . La especificación del entero  $j$  requiere a lo mas un número  $\log_b N_w(N)$  del alfabeto de símbolos, donde  $N_w(N) \leq N$  es el número total de palabras código y  $N$  es la longitud de la secuencia de entrada. De aquí, la longitud total  $L(N)$  de la secuencia codificada crece como

$$L(N) \sim N_w(N)(\log_b N_w(N) + 1) \quad (4.38)$$

para  $N$  creciente. Mientras  $N$  sea pequeña,  $L(N)$  puede exceder a  $N$  (como en el ejemplo de arriba). Cuando  $N \gg 1$ , sin embargo, las palabras  $w_k$  se vuelven tan largas que ellas son más eficientemente identificadas por medio de sus número de orden  $j$  (la posición del prefijo) en la lista y por el sufijo  $s$  más que por una descripción directa. El número  $N_w(N)$  de palabras en una revisión diferente (i.e., consistente de palabras diferentes) de una secuencia  $S = s_1 s_2 \dots s_N$  satisface

$$N_w(N) \sim c \frac{N}{\ln_b N} \quad (4.39)$$

para  $N \rightarrow \infty$ , con  $c \leq 1$ . Finalmente, la complejidad Lempel-Ziv  $C_{LZ}$ , queda definida como

$$C_{LZ} = \limsup_{N \rightarrow \infty} \frac{L(N)}{N} \quad (4.40)$$

### 4.7.2 Análisis de Gramáticas (Reglas de Producción)

El análisis de gramáticas tiene su base en la dinámica simbólica, es útil en la identificación de patrones dentro de una serie de tiempo, los cuales se expresan como un conjunto de reglas que se

denomina reglas de producción de una gramática para la serie de tiempo [116, 117]. El número de reglas de producción para una serie de tiempo es una medida de complejidad (de origen computacional), en la cual a mayor número de reglas de producción necesarias para generar la gramática de la serie corresponde con una mayor dificultad de predicción o también un menor predictibilidad de la serie de tiempo [10]. Partiendo de una gramática libre de contexto que se define como la tupla  $G = (V, T, P, S)$  donde:

- $V$  = conjunto finito de variables.
- $T$  = conjunto finito de terminales.
- $P$  = conjunto finito de producciones de la forma  $A \rightarrow \alpha$  donde  $A$  es una variable y  $\alpha$  es una cadena de símbolos tomados de  $(V \cup T)^*$ .
- $S$  = variable inicial.

se construye la gramática para la serie de tiempo. El conjunto de los símbolos terminales representan la diferencia entre valores contiguos de la serie, de esta forma para el mismo comportamiento de la serie en diferentes secciones de la misma, dicho comportamiento queda codificado con un solo símbolo de la gramática. Esta variable a su vez, puede formar parte de la definición de otra variable y así recursivamente.

El algoritmo para obtener la gramática consiste de las siguientes etapas:

- Se cuantifican los valores de la serie de tiempo en valores discretos de 0 a 100.
- Se codifica la serie discretizada mediante la técnica de Modulación por Codificación de Pulsos Diferencial (DPCM) [147]. De manera que la serie ahora está representada por las diferencia de los valores contiguos de la misma.
- Se genera la gramática de la señal codificada, mediante el algoritmo de Sequitur [148].

Una vez generada la gramática se obtiene el número total de reglas de producción (RP), parámetro que se aplica en este trabajo para la caracterización de la predictibilidad en las series de tiempo.

## Capítulo 5

# Técnicas de Predicción y de Modelado de Series de Tiempo

### 5.1 Introducción

Al estudiar un sistema en general se busca desarrollar un modelo mediante un sistema de ecuaciones que expliquen su dinámica y que además permitan predecirla. El conocimiento previo que tenemos del sistema lo obtenemos a partir de observaciones experimentales, estos modelos son llamados fenomenológicos, en algunos casos como en mecánica clásica es posible derivar dichos modelos a partir de primeros principios, pero aún estos deben ser consistentes con las observaciones experimentales. Para desarrollar un modelo de esta naturaleza, se requiere tener suficiente información sobre las interacciones y relaciones entre las variables del sistema. Sin embargo, no siempre se tiene toda la información necesaria para inferir las relaciones y variables relevantes del sistema que nos permitan construir un modelo que reproduzca en forma aproximada su dinámica. En estos casos, en los cuales se tiene una secuencia de observaciones (*Serie de Tiempo*) de una o de unas cuantas variables, pero que además el comportamiento de dichas variables no permite inferir relaciones causales directas, es necesario recurrir a construir los modelos en base al comportamiento estadístico observado en los datos de la serie. Los modelos así construidos no poseen en general una relación directa con la dinámica del sistema [3]. Sin embargo, permiten modelar el comportamiento de la variable observada y en algunos casos es posible predecir el comportamiento futuro de la variable con cierto grado de error. El

desarrollo de técnicas de predicción en el sentido estricto de la palabra ha sido limitado, la gran mayoría de las técnicas reportadas en la literatura se relacionan con el modelado de series de tiempo, sin pasar a la etapa de predicción [10]. La predicción de una serie de tiempo requiere de la construcción de un modelo *ad hoc* para dicha serie cuando esto sea posible y debido a esto, no es posible *generalizar* la capacidad de predicción de un modelo para que se adapte a series con comportamientos muy disímiles. El estudio de diferentes tipos de modelos que tengan la capacidad ya sea solo de modelar o de modelar y predecir, puede ayudar a un mejor entendimiento de las relaciones entre los modelos y su adaptabilidad para modelar y / o predecir a diferentes series de tiempo (i.e. series con diferente predictibilidad), a esta adaptabilidad de los modelos la llamaremos capacidad de predicción (o de modelado). A continuación se presentan las bases y algoritmos de las diferentes técnicas de predicción y de modelado, evaluadas en la presente tesis. Las técnicas tienen como base de sus modelos a la Inteligencia Artificial, Estadística y Teoría de Sistemas Dinámicos No Lineales.

## 5.2 Técnicas de Predicción

### 5.2.1 Técnicas Estadísticas

#### ARIMA

Box y Jenkins desarrollaron su modelo ARMA (ver una descripción más detallada en la sección 5.3), el cual consiste en la unión de dos modelos primitivos, el primero de ellos es el *Auto Regresivo* (*AR*), en estadística se tiene la técnica de regresión que consiste en que una variable puede ser modelada o predicha utilizando una combinación lineal de una o más variables distintas. En el caso de la auto regresión la variable es modelada mediante la combinación lineal de los valores previos de ella misma. Un modelo auto regresivo asume que el valor actual de la serie es igual al promedio pesado de un número finito de valores previos, más una constante de offset, más un término de ruido aleatorio, estos términos aleatorios se les llama *shocks*. El segundo modelo primitivo que juega un papel importante en la aproximación de Box-Jenkins es el *promedio móvil* (*Moving Average, MA*). Este modelo nos dice que el valor actual de una serie de tiempo es igual a la suma pesada de un número finito de shocks previos, más una constante de offset y el shock actual [80]. Box y Jenkins encontraron que era posible diferenciar en forma implícita

a la serie de tiempo, de tal forma que se logrará la condición estacionaria que es necesaria para el uso adecuado del modelo ARMA, luego aplicar dicho modelo y posteriormente integrar el resultado para regresar al dominio original, a este proceso se le denomina ARIMA. La notación para un modelo ARIMA es una extensión de la correspondiente a un modelo ARMA:  $ARIMA(p, d, q)$ , donde  $d$  es el número de operaciones de diferenciación que tienen lugar antes de aplicar el modelo ARMA y también es el número de integraciones necesarias para regresar al dominio original de la serie de tiempo. El número de términos del modelo AR se define como  $p$  y el número de términos del modelo MA como  $q$  [16, 73].

## 5.2.2 Técnicas de Inteligencia Artificial

### Red Neuronal con Filtros FIR (FIRNet)

La red FIRNet posee una arquitectura de red neuronal multicapa hacia adelante (*feedforward*) en la cual cada peso sináptico está formado por un filtro lineal FIR (*Respuesta de Impulso Finito*) [149]. FIR implica que para una excitación de una entrada de duración finita, la salida del filtro también es de duración finita. Para este filtro, la salida  $y(k)$  corresponde a una suma pesada de los valores pasados retardados  $x(k - n)$  de la entrada:

$$y(k) = \sum_{n=0}^T w(n)x(k - n) \quad (5.1)$$

Nótese que la ecuación anterior corresponde a una componente de promedio móvil de un modelo de promedio móvil auto regresivo (ARMA). El filtro FIR es uno de los primeros elementos adaptables que se estudio en la literatura. El algoritmo de aprendizaje es una modificación del algoritmo de retro propagación (*backpropagation*), la diferencia consiste en las relaciones temporales implícitas y las operaciones de filtrado, por ello se le denomina retro propagación temporal.

Ahora se describe la aplicación de esta arquitectura para la predicción de series de tiempo, se desea modelar una serie  $y(k)$  para cada paso de tiempo la entrada a la red FIR es el valor conocido  $y(k - 1)$ , y la salida es  $\hat{y}(k) = N_q [y(k - 1)]$  es la estimación de valor original de la serie. Luego el modelo es de la forma:

$$y(k) = N_q [y(k-1)] + e(k) \quad (5.2)$$

La red FIR  $N_q$  corresponde a una red restringida que actúa sobre una ventana finita de la entrada, entonces la expresión anterior se describe como:

$$y(k) = N_c [y(k-1), y(k-2), \dots, y(k-T)] + e(k) \quad (5.3)$$

que enfatiza la auto regresión no lineal.

Durante el entrenamiento, el error cuadrado  $e(k)^2 = (y(k) - \hat{y}(k))^2$  se minimiza utilizando la retro propagación temporal para adaptar la red ( $y(k)$  actúa como la respuesta deseada). Nótese que se realiza una adaptación de lazo abierto; ambas la entrada y la respuesta deseada son proporcionadas por la serie de entrenamiento conocida. La salida actual de la red no es retro alimentada como entrada durante el entrenamiento (adaptación de ecuación de error). La expectativa del error cuadrado a minimizar es de la forma:

$$E [e^2(k)] = E [y(k) - \hat{y}(k)]^2 = E [y(k) - N_c [\mathbf{y}_1^T(k)]]^2 \quad (5.4)$$

donde  $\mathbf{y}_1^T(k) = [y(k-1), y(k-2), \dots, y(k-T)]$  especifica el regresor.

Una vez que la red es entrenada, la predicción iterativa de largo plazo se logra tomando la estimación  $\hat{y}(k)$  y alimentando esta de regreso a la entrada de la red (lazo cerrado):

$$\hat{y}(k) = N_q [\hat{y}(k-1)] \quad (5.5)$$

### **Red Neuronal Probabilística (con función Gaussiana PNNGauss y con función recíproca PNNReciprocal)**

La primera red de este tipo fue descrita por Meisel (1972) y el primer algoritmo operacional fue propuesto por Specht (1990), esta red tiene sus bases formales en la teoría de la probabilidad, la red PNN es básicamente un clasificador, fue diseñada originalmente como un algoritmo de clasificación que se entrena con miembros de una o más clases, para posteriormente asignar nuevos elementos a las clases conocidas, sin embargo también puede realizar tareas de regresión [16, 150]. El modelo estadístico de esta red se basa en la estadística de Bayes, lo cual

requiere conocer la función de densidad de probabilidad, para inferirla a partir de los datos de entrenamiento se aplica el método de estimación de la densidad de Parzen, el cual estima la función de densidad univariada a partir de una muestra aleatoria, el estimador converge en forma asintótica a la densidad verdadera conforme la muestra de datos se incrementa. Este método utiliza una función de peso  $W(d)$  llamada kernel o función potencial, la cual tiene su valor más grande en  $d = 0$ , este valor decrece rápidamente conforme el valor absoluto de  $d$  se incrementa. Estas funciones de peso están centradas en cada dato de entrenamiento, y el valor de cada función de la muestra de datos está determinado por su distancia  $d$  respecto del dato muestra. Luego la función de densidad de probabilidades es una suma escalada para todos los datos muestra. Matemáticamente la función de densidad para una muestra de  $n$  datos se expresa como

$$g(x) = \frac{1}{n\sigma} \sum_{i=1}^n W\left(\frac{x - x_i}{\sigma}\right) \quad (5.6)$$

El parámetro de escalamiento  $\sigma$  define el ancho de la curva de campana que rodea a cada dato de la muestra. La función de peso  $W$  más usada es la función Gaussiana, que corresponde al modelo PNNGauss evaluado en este trabajo, nótese que esto no tiene nada que ver con condiciones de normalidad, es solo una cuestión de conveniencia, esta función es bien comportada y fácil de calcular. Con la función Gaussiana, la función de densidad toma la forma

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\left(\frac{-x^2}{2\sigma^2}\right)} \quad (5.7)$$

Otra función de peso que se utilizó en este trabajo fue la correspondiente a la función recíproca

$$W(d) = \frac{1}{1 + d^2} \quad (5.8)$$

con esta función se construyó el modelo de predicción denominado PNNReciprocal.

Finalmente la arquitectura de la red, consiste de una capa de entrada, una capa de patrones, la capa de suma de las funciones de densidad de probabilidad y la capa de salida.

### Red Neuronal Multicapa hacia Adelante (MLFFN)

Este modelo neuronal llamado Perceptrón, fue introducido por Rosenblatt a finales de los años cincuenta [151]. La estructura de este modelo se inspira en las primeras etapas del procesamiento de los sistemas sensoriales de los animales (por ejemplo, la visión), en los cuales la información va atravesando sucesivas capas de neuronas, que realizan un procesamiento progresivamente de más alto nivel. Es un modelo unidireccional, compuesto en su forma básica (perceptrón simple) por dos capas de neuronas, una sensorial o de entradas, y la otra de salida [151, 152]. La operación de una red de este tipo de modelo, con  $n$  neuronas de entrada y  $m$  de salida, se puede expresar como

$$y_i(t) = f\left(\sum_{j=1}^n w_{ij}x_j - \theta_i\right), \quad \forall i, \quad 1 \leq i \leq m \quad (5.9)$$

Las neuronas de entrada no realizan ningún cómputo, únicamente envían la información (en principio se considera como señales discretas  $\{0, 1\}$ ) a las neuronas de salida. La función de activación de las neuronas de la capa de salida puede ser de tipo escalón, o como en el modelo evaluado en este trabajo sigmoideal binaria. Si se añaden capas intermedias (ocultas) a un perceptrón simple, se obtiene un perceptrón multicapa (*Multi-Layer FeedForward Network*). En particular el modelo de predicción evaluado, consiste de un perceptrón simple (cero capas ocultas), con las salidas de tipo lineal y el método de aprendizaje es de regresión con refinamiento con gradiente conjugado. El error a minimizar fue el valor medio absoluto de los errores de predicción [16].

### Máquina de Vectores de Soporte (MySVM Multipunto con kernel radial y dot)

Las SVM son sistemas de aprendizaje introducidos por Vapnik [153], que utilizan un espacio de hipótesis de funciones lineales en un amplio espacio de características, las cuales son entrenadas con un algoritmo de optimización que implementa una tendencia de aprendizaje estadístico [154]. Dicha teoría de aprendizaje trata el problema de encontrar una función de una clase de funciones  $(f_\alpha)_\alpha \in \Lambda$  del tal forma que se minimice el riesgo esperado  $R[f]$ ,

$$R[f] = \iint L(y, f(x))dP(y|x)dP(x) \quad (5.10)$$

respecto a una función de pérdida  $L$  cuando la distribución de probabilidad de los ejemplos  $P(x)$  y sus clasificaciones  $P(y|x)$  no son conocidos, y tienen que ser estimados de un número finito de ejemplos  $(x_i, y_i)_i \in I$ . El algoritmo de las SVM resuelve éste problema minimizando el riesgo regularizado  $R_{reg}[f]$ :

$$R_{reg}[f] = R_{emp}[f] + \lambda \|w\|^2 \quad (5.11)$$

el cual es la suma ponderada del riesgo empírico  $R_{emp}[f]$  con respecto a los datos  $(x_i, y_i)_i = 1, 2, \dots, n$  y a un término de complejidad  $\|w\|^2$ . En su formulación básica, la SVM encuentra una función de decisión  $f(x) = \text{sign}(\omega x + b)$  que minimiza el error de predicción en el conjunto de entrenamiento y procura el mejor desempeño en la generalización.

Una de las principales características de las SVM es el uso de funciones de kernel para extender la clase de funciones de decisión al caso no lineal [154]. Esto se hace mapeando los datos desde el espacio de entrada  $X$  a un amplio espacio de características  $\chi$  mediante una función  $\Phi$ , y resolviendo el problema de aprendizaje lineal en  $\chi$ .

$$\Phi : X \rightarrow \chi \quad (5.12)$$

La función real  $\Phi$  no necesita ser conocida, es suficiente tener una función de kernel  $k$  que calcule el producto interno en el espacio de características.

$$k(x, y) = \Phi(x) \bullet \Phi(y) \quad (5.13)$$

El kernel lineal o dot  $k(x, y) = x \bullet y$  es la función de kernel más simple. La función de decisión toma la forma  $f(x) = \omega x + b$ . Cuando se usa este kernel en el análisis de series de tiempo, el modelo resultante es un modelo auto regresivo de orden  $p$  (AR( $p$ )).

Las funciones de kernel con base radial (FBR) toman la forma  $k_\gamma(x, y) = \exp(-\gamma \|x - y\|^2)$ , siendo  $\gamma$  una constante de proporcionalidad cuyo rango de valores útiles debe ser estimado para cada aplicación en particular. Para las series de tiempo, se ha mostrado que la información dentro de una ventana de una serie puede ser obtenida de otras ventanas que son similares en términos de la distancia euclidiana entre ellas, lo que hace útil al kernel con base radial para la predicción [20].

Las SVM pueden ser aplicadas a resolver el problema de predicción de series de tiempo mediante la siguiente representación, si tenemos una serie de tiempo  $ST = \{x_1, x_2, \dots, x_N\}$  podemos separarla en ventanas  $w = (x_i, \dots, x_{i+p-1})$  de tamaño  $p$ . Entonces hay que encontrar una función  $f : R^p \rightarrow R$ , tal que  $f(x_i, \dots, x_{i+p-1}) = x_{i+p}$  para cada  $i \in \{0, N - p\}$  [20]. La técnica de SVM evaluada en este trabajo se llama MySVM [19], y fue modificada para utilizar una representación de ventana deslizante en el conjunto de entrenamiento del algoritmo SVM y realizar predicciones de múltiples puntos [20]. Se construyeron dos modelos uno que utiliza el kernel de tipo radial y otro con el kernel de tipo lineal.

### 5.2.3 Técnicas de Teoría de Sistemas Dinámicos No Lineales

#### Modelo Lineal Local en el Espacio Fase (Nstep)

Este modelo de predicción consiste en asumir que es posible ajustar un modelo lineal de carácter local para cada punto del sistema en el espacio fase [3], para de esta forma resolver la siguiente ecuación:

$$s_{n+1} = f(s_n) \quad (5.14)$$

donde  $f(s_n)$  es una función suave no conocida que depende del punto  $s_n$ , realizando la aproximación local para  $f(s_n)$  mediante una expansión de Taylor es posible encontrar la solución a la expresión anterior. La condición que se pide es la minimización de la varianza del conjunto de puntos:

$$\sigma^2 = \sum_{s_j \in U_n} (s_{j+1} - a_n s_j - b_n)^2 \quad (5.15)$$

con respecto a los coeficientes  $a_n$  y  $b_n$ , donde  $U_n$  es el  $\epsilon$ -vecindario de  $s_n$ , excluyendo al mismo. Entonces la predicción es:

$$\widehat{s_{n+1}} = a_n s_n + b_n \quad (5.16)$$

El problema de minimización se resuelve mediante un sistema de ecuaciones lineales acopladas.

## Modelo de Funciones Polinomiales en el Espacio Fase (Polynomp)

En este modelo se considera la construcción de un modelo de carácter global que ajuste los puntos del sistema mapeados en el espacio fase. La expresión a resolver es:

$$\sigma^2 = \sum_n (s_{n+1} - f_p(s_n))^2 \quad (5.17)$$

donde  $f_p$  es una función no lineal en forma cerrada con  $p$  parámetros, con respecto a los cuales la expresión anterior debe ser minimizada. Es posible utilizar polinomios (como en esta técnica), funciones de base radial, redes neuronales, polinomios ortogonales, etc.. Los resultados dependen de que el *ansatz* (aproximación o estimación) seleccionado  $f_p$  sea adecuado para modelar la función no lineal desconocida y de que tan deterministas son los datos a modelar [3].

## K-Vecinos Cercanos (K-Nearest-Neighbours)

La idea principal detrás de este método es el predecir el valor objetivo de una nueva observación a partir de observaciones conocidas realizadas en el pasado (base de casos). La nueva observación es comparada con todos los elementos de la base de casos. Las  $k$  observaciones pasadas más similares son luego seleccionadas como referencias para el nuevo candidato. La medida de similitud se define frecuentemente como la distancia entre las nuevas y las viejas observaciones. Los valores objetivo de las  $k$  referencias son combinadas por ejemplo, con un promedio simple, para obtener el valor objetivo de la nueva observación. El método de  $k$  vecinos cercanos imita la habilidad humana conocida como reacción a una nueva situación con la ayuda de la experiencia pasada. El método de  $k$  vecinos cercanos pertenece a la clase de algoritmos de aprendizaje basados en instancias en el campo de aprendizaje de máquina (*Machine Learning*), ya que emplea instancias en lugar de modelos. Este método pertenece a la clase de métodos no paramétricos. No son necesarias suposiciones sobre la distribución de los datos analizados [155]. En el caso de la predicción de series de tiempo, la aplicación de este método consiste en la reconstrucción del atractor en el espacio fase, para posteriormente identificar los  $k$  vecinos cercanos que corresponden a estados similares del sistema y que pertenecen a trayectorias cercanas a la trayectoria donde se predice el estado siguiente no conocido [3, 156], la Figura 5-1

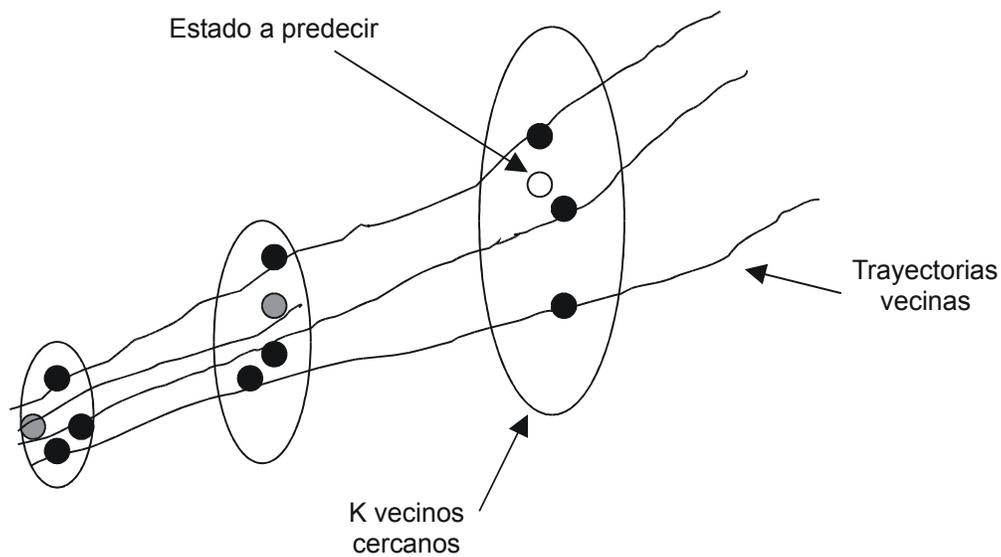


Figura 5-1: Predicción con el método de  $K$  vecinos cercanos

ilustra la explicación anterior. Una vez identificados los  $k$  vecinos cercanos se calcula el estado a predecir mediante un promedio simple o una variante de este como puede ser un promedio pesado, donde se asigna mayor peso a las muestras más recientes.

## 5.3 Técnicas de Modelado

### 5.3.1 Técnicas Estadísticas

#### ARMA

George Box y Gwilym Jenkins en 1976 propusieron que muchas series podrían ser explicadas mediante un modelo relativamente simple, este modelo está basado en dos modelos primitivos [80]. El primero de ellos es el *Auto Regresivo* ( $AR$ ), en estadística se tiene la técnica de regresión que consiste en que una variable puede ser modelada o predicha utilizando una combinación lineal de una o más variables distintas. En el caso de la auto regresión la variable es modelada mediante la combinación lineal de los valores previos de ella misma. Un modelo auto regresivo asume que el valor actual de la serie es igual al promedio pesado de un número finito de valores previos, más una constante de offset, más un término de ruido aleatorio. Estos términos

aleatorios se les llama *shocks*, por definición estos términos son independientes entre sí y con respecto a la serie de tiempo, además están idénticamente distribuidos y se asume poseen media cero y una varianza finita, estrictamente deberían seguir una distribución normal. El modelo primitivo de AR se expresa como:

$$z_t = \phi_0 + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + a_t \quad (5.18)$$

La correlación entre los parámetros es alta lo que produce inestabilidades en su cálculo cuando se buscan muchos de ellos, el offset es  $\phi_0$  y no es la media de la serie, el término shock es  $a_t$ .

El segundo modelo primitivo que juega un papel importante en la aproximación de Box-Jenkins es el *promedio móvil* (*Moving Average, MA*). Este modelo nos dice que el valor actual de una serie de tiempo es igual a la suma pesada (con pesos  $\theta_1, \theta_2, \dots$ ) de un número finito de shocks previos ( $a_{t-1}, a_{t-2}, \dots$ ), más una constante de offset  $\phi_0$  y el shock actual  $a_t$ . Lo anterior se expresa con la siguiente ecuación,

$$z_t = \phi_0 + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + a_t \quad (5.19)$$

El nombre de promedio móvil no es precisamente el mejor. Desde el punto de vista estadístico la expresión anterior implica que los pesos son positivos y su suma igual a la unidad, pero no es necesario que se cumpla estrictamente lo anterior. En este caso la constante de offset si corresponde a la media de la serie de tiempo, esto debido a que por definición los términos shock poseen media cero.

En apariencia ambos modelos parecen ser muy similares pero no son idénticos, de hecho se complementan. Muchos procesos físicos se describen mejor con modelos cuyos términos a pesar de ser simples en forma, se extienden hacia atrás en el tiempo considerablemente, lo cual como se mencionó anteriormente dificulta el cálculo de los parámetros del modelo cuando estos son demasiados. Aquí es donde una propiedad importante de los modelos AR y MA es de utilidad, se puede mostrar que un modelo finito de un tipo es matemáticamente equivalente a un modelo infinito de otro tipo. Por ejemplo, la siguiente expresión muestra la equivalencia de un término simple de un modelo AR y un modelo MA con número de términos infinito,

$$z_t = \phi z_{t-1} + a_t = a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \phi^3 a_{t-3} + \dots \quad (5.20)$$

No es necesario limitar el modelo de elección a un AR o MA puros. Se pueden combinar ambos dando lugar al modelo ARMA:

$$z_t = \phi_0 + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + a_t \quad (5.21)$$

Este modelo sigue siendo popular ya que funciona bien para una gran cantidad de problemas reales. En la literatura se acostumbra definir al número de términos del modelo AR como  $p$  y al número de términos del modelo MA como  $q$ , entonces se dice que se tiene un modelo ARMA( $p, q$ ) [16, 73].

### 5.3.2 Técnicas de Inteligencia Artificial

#### Red Neuronal hacia Adelante con aprendizaje de retro propagación (FFNBP)

Este modelo de red neuronal es similar al utilizado por la técnica *MLFFN* descrita en la subsección 5.2.2, la diferencia principal consiste en que la red neuronal en vez de predecir, ahora solo modela las series de tiempo, en particular la red neuronal evaluada consistió de una capa de entrada, una capa oculta y una de salida [157]. El algoritmo de entrenamiento utilizado es de retro propagación (*backpropagation*) [151]. El cuál busca minimizar el error cuadrático medio

$$E(w_{ji}, \theta_j, w'_{kj}, \theta'_k) = \frac{1}{2} \sum_{\mu} \sum_k \left[ t_k^{\mu} - f \left( \sum_j w'_{kj} y_j^{\mu} - \theta'_k \right) \right]^2 \quad (5.22)$$

mediante el descenso por gradiente, con un gradiente respecto a los pesos de la capa de salida y otro respecto de los de la capa oculta

$$\delta w'_{kj} = -\varepsilon \frac{\partial E}{\partial w'_{kj}} \quad \delta w_{ji} = -\varepsilon \frac{\partial E}{\partial w_{ji}} \quad (5.23)$$

Las expresiones de actualización de los pesos son de la forma

$$\begin{aligned}\delta w'_{kj} &= \varepsilon \sum_{\mu} \Delta'_k{}^{\mu} y_j^{\mu}, \quad \text{con} \quad \Delta'_k{}^{\mu} = \left[ t_k^{\mu} - f(v'_k{}^{\mu}) \right] \frac{\partial f(v'_k{}^{\mu})}{\partial v'_k{}^{\mu}} \\ \delta w_{ji} &= \varepsilon \sum_{\mu} \Delta_j^{\mu} x_i^{\mu}, \quad \text{con} \quad \Delta_j^{\mu} = \left( \sum_k \Delta'_k{}^{\mu} w'_{kj} \right) \frac{\partial f(v'_j{}^{\mu})}{\partial v'_j{}^{\mu}}\end{aligned}\tag{5.24}$$

En estas expresiones está implícito el concepto de propagación hacia atrás de los errores que da nombre al algoritmo, obsérvese como se propaga hacia atrás la señal de error  $\Delta'_k{}^{\mu}$  para calcular la señal de error  $\Delta_j^{\mu}$ .

### **Red Neuronal de Funciones Radiales con aprendizaje de retro propagación (RBFNBP)**

Este modelo es una red unidireccional para aproximación funcional que puede incorporar aprendizaje supervisado y no supervisado. La arquitectura de una red de funciones de base radial (RBF) cuenta con tres capas de neuronas: de entrada, oculta y de salida. Las neuronas de entrada, simplemente envían la información del exterior hacia las neuronas de la capa oculta. Las neuronas de la capa de salida son lineales, esencialmente calculan la suma ponderada de las salidas que proporciona la capa oculta. La diferencia fundamental entre la arquitectura de este modelo y el de un perceptrón multicapa, se centra en la operación de las neuronas ocultas. Éstas, en vez de computar la suma ponderada de las entradas y aplicarle una función de tipo sigmoideo, operan en base a la distancia que separa el vector de entradas respecto del vector sináptico que cada una almacena (denominado centroide), cantidad a la que aplican una función radial con forma Gaussiana. Es decir, así como en un perceptrón multicapa las neuronas ocultas poseen una respuesta de rango infinito (cualquier vector de entrada, con independencia del lugar del espacio de entrada de donde proceda puede causar que la neurona se active), en la red RBF las neuronas son de respuesta localizada, pues sólo responden con una intensidad apreciable cuando el vector de entradas presentado y el centroide de la neurona pertenecen a una zona próxima en el espacio de entradas. Sea  $x_i$  las entradas de la red,  $y_j$  las salidas de la capa oculta, y  $z_k$  las salidas de la capa final (y globales de la red). Cada neurona  $j$  de la capa oculta almacena un vector  $c_{ji}$ , el centroide; a cada una de estas neuronas se le calcula la distancia euclidiana  $r_j$  que separa el vector de entradas  $x_i$  de su centroide

$$r_j^2 = \|\mathbf{x} - \mathbf{c}_j\|^2 = \sum_i (x_i - c_{ji})^2 \quad (5.25)$$

La salida de la neurona  $y_j$  se calcula a partir de una función de activación denominada función radial  $\phi(r)$ . Una de las más típicas es la función Gaussiana

$$\phi(r) = e^{-r^2/2\sigma^2} \quad (5.26)$$

El término función de base radial procede precisamente de la simetría radial de estas funciones (el nodo da una salida idéntica para aquellos patrones que distan lo mismo del centroide). La salida de la neurona oculta  $j$  es

$$y_j = e^{-r_j^2/2\sigma_j^2} = e^{-\sum_i (x_i - c_{ji})^2/2\sigma_j^2} \quad (5.27)$$

Así, si el vector de entradas coincide con el centroide de la neurona  $j$  ( $\mathbf{x} = \mathbf{c}_j$ ), ésta responde con máxima salida (la unidad). Es decir, cuando el vector de entradas se sitúa en una región próxima al centroide de una neurona, ésta se activa, indicando que reconoce el patrón de entrada; si el patrón de entrada es muy diferente del centroide, la respuesta tiende a cero.

Las salidas de las neuronas ocultas son a su vez las entradas de las neuronas de salida, las cuales calculan su respuesta  $z_k$  de la forma

$$z_k = \sum_j w_{kj} y_j + \theta_k = \sum_j w_{kj} \phi(r_j) + \theta_k \quad (5.28)$$

siendo  $w_{kj}$  el peso que conecta la neurona oculta  $j$  con la salida  $k$ , y  $\theta_k$  un parámetro adicional de la neurona  $k$ , que por similitud con el perceptrón llamaremos umbrales (bias). Esta expresión es similar a la correspondiente al perceptrón con una capa oculta y neuronas de salida lineales. El aprendizaje de la red se realiza en dos partes, primero se entrena a las neuronas ocultas Gaussianas, para ello se determinan los centroides  $c_{ji}$  por medio de el algoritmo k-medias, luego se procede al cálculo de los parámetros de escala  $\sigma_j$  de cada neurona, para lo cual suele hacerse uso de criterios heurísticos. La segunda parte del entrenamiento corresponde a las neuronas de salida donde se calculan los pesos  $w_{kj}$  y los umbrales  $\theta_k$ , en este caso se utilizó el algoritmo de retro propagación (*backpropagation*) descrito anteriormente para

la técnica FFNBP [151, 152, 157].

### **Red Neuro-Difusa (ANFIS)**

ANFIS (*Adaptive-Network-based Fuzzy Inference System*) es un sistema de inferencia difusa implementado en el marco de redes adaptables. Mediante el uso de un procedimiento de aprendizaje híbrido, ANFIS puede construir mapeos de entrada-salida basados en conocimiento humano (en la forma de reglas difusas si-entonces) o como en el caso de modelado de series de tiempo basados en pares de datos de entrada-salida asignados [158, 159].

Las reglas difusas si-entonces están caracterizadas por funciones de membresía, en este caso particular del tipo campana Gaussiana. Estas reglas son empleadas para capturar los modos de razonamiento imprecisos que juegan un rol esencial en la habilidad humana para tomar decisiones en un ambiente de incertidumbre e imprecisión. El modelo utilizado en la red ANFIS es el de Takagi y Sugano, donde los conjuntos difusos sólo están presentes en la parte de la premisa, la parte correspondiente a la consecuencia es descrita por una ecuación no difusa.

La arquitectura de la red adaptable es el superconjunto al que pertenecen las redes neuronales hacia adelante con aprendizaje supervisado. En el caso particular de ANFIS, la arquitectura corresponde a una red hacia adelante multicapa con el aprendizaje en dos partes: primero con estimación de mínimos cuadrados en el paso hacia adelante (*feedforward*) y en la segunda parte con gradiente descendente en el paso hacia atrás (*backpropagation*). La arquitectura de la red implementa al sistema difuso de la siguiente forma, durante el proceso de entrenamiento se parte de unas funciones de membresía iniciales, en el paso hacia adelante la estimación de mínimos cuadrados calcula los parámetros de la parte de consecuencia del enunciado difuso, luego en el paso hacia atrás los errores se propagan y los parámetros de la parte de premisa del enunciado son actualizados por el algoritmo de gradiente descendente. Finalmente se obtienen las funciones de membresía ajustadas para el conjunto de datos a modelar durante la etapa de prueba.

### **Red Neuronal Polinomial (RNAP)**

RNAP es un modelo híbrido de red neuronal que define una función de transición en términos de un sistema no lineal de ecuaciones polinomiales, el cual es optimizado en la fase de entre-

namiento por medio de una implementación de Algoritmo Genético con el fin de obtener la mejor arquitectura de la red.

El modelo RNAP se define matemáticamente como:

$$\hat{y}_k = [\phi(x_{1,k}, x_{2,k}, \dots, x_{n_i,k}, x_{1,k-1}, x_{2,k-1}, \dots, x_{n_i,k-n_l}, y_{k-1}, y_{k-2}, \dots, y_{k-n_2})]_{\phi_{\min}}^{\phi_{\max}} \quad (5.29)$$

donde  $\hat{y}_k \in R$ , es la salida de la red,  $\phi(x, y) \in R$  es una función no lineal,  $x_i \in X$ , son las entradas,  $i = 1, \dots, n$ ;  $n_i$  es el número de entradas,  $y_{k,j} \in Y$  son los valores anteriores de la salida  $j = 1, \dots, n_2$ ;  $n_l$  es el número de retardos de la entrada,  $n_2$  es el número de retardos de la salida y  $X, Y$ , son conjuntos compactos de  $R$ .

La función no lineal  $\phi(z)$  está dada por:

$$[\phi(z)]_{\phi_{\min}}^{\phi_{\max}} = \begin{cases} \phi_{\max}, & \phi(z) \geq \phi_{\max} \\ \phi(z), & \phi_{\min} < \phi(z) < \phi_{\max} \\ \phi_{\min}, & \phi(z) \leq \phi_{\min} \end{cases} \quad (5.30)$$

donde  $\phi_{\max}$  y  $\phi_{\min}$  son los límites máximos y mínimos respectivamente.

Para simplificar la notación se reescribe la ecuación 5.30 con el siguiente cambio de variable:

$$z = \{z_1, z_2, \dots, z_{n_v}\} \quad (5.31)$$

Donde  $n_v$  es el número total de elementos de la descripción  $z$ , es decir, el número total de entradas, valores anteriores y valores de salida.

Dado el cambio de variable, la función  $\phi(z_k) \in \Phi_p$  conforma una familia de polinomios que pueden ser representados como:

$$\Phi_p = (z_1, z_2, \dots, z_n) = \left\{ \begin{array}{l} \phi(z) : \phi(z) = a_0(z_1, z_2, \dots, z_{n_v}) + a_1(z_1, z_2, \dots, z_{n_v}) + \\ \dots + a_p(z_1, z_2, \dots, z_{n_v}) \end{array} \right\} \quad (5.32)$$

donde los términos  $a_i(z_1, z_2, \dots, z_i)$  son polinomios homogéneos de grado total  $i$ , para  $i = 0, \dots, p$ .

Para describir el proceso de aprendizaje de la red RNAP se definen los siguientes términos:

- Error de aproximación de RNAP

$$err_n(y^n, \phi(z)) = \frac{1}{n} \sum_{k=1}^n (y_k - \phi(z_k))^2 = \frac{1}{n} \sum_{k=1}^n (y_k - \hat{y}_k)^2, y^n = (y_1, y_2, \dots, y_n) \quad (5.33)$$

Donde  $\phi(z_k) \in \Phi_p$  y  $n$  es el número de puntos.

- Error óptimo:

$$opterr_n(y^n, \phi(z)) = \min \{err_n(y^n, \phi(z))\} = err_n(y^n, \phi_n^*(z)) \quad (5.34)$$

donde  $\phi_n^*(z) \in \Phi_p$  es la estimación óptima de  $y^n$ .

La RNAP  $\phi(z) \in \Phi_p$  aprende uniformemente la salida deseada con la precisión  $\varepsilon$  si

$$err_n(y^n, \phi(z)) - err_n(y^n, \phi_n^*(z)) \geq \varepsilon \quad (5.35)$$

con  $\varepsilon \geq 0$ .

Una vez definidos estos conceptos el problema de aprendizaje de RNAP consiste en encontrar la estructura de  $\phi \in \Phi_p(z)$  que verifica la desigualdad descrita por 5.35.

El problema de aprendizaje de RNAP para una arquitectura específica puede ser representado como un problema de optimización con los siguientes pasos:

$$\min_{W_b} \left\{ \min_{W \in R} \{err_n(y^n, \phi(z))|_{W_b}\} \right\} \quad (5.36)$$

donde  $err_n(y^n, \phi(z))|_{W_b}$  es el error definido en 5.33 para un valor determinado del vector binario  $W_b$ .

El objetivo del algoritmo genético es encontrar el valor de  $W_b$  para el cual el vector de pesos  $W$  hace que se cumpla 5.35. Este arreglo puede ser obtenido utilizando el método de mínimos cuadrados y queda expresado como:

$$\begin{aligned}
W_{W_b} &= \arg \min \{err_n(y^n, \phi(z)|_{W_b})\} \\
W_{W_b} &= \Gamma^N \sum_{k=1}^n y_k(M_b(z_k))|_{W_b}
\end{aligned} \tag{5.37}$$

donde:

$$\begin{aligned}
M_b(z_k) &= M \cdot * W_B^T \\
\Gamma^N &= \left( \sum_{k=1}^n M_b(z_k)^T \right)^{-1}
\end{aligned} \tag{5.38}$$

Los pasos que sigue el Algoritmo Genético para obtener la estructura óptima de RNAP se pueden resumir de la siguiente manera:

- 1 Se obtiene la población inicial de forma aleatoria.
- 2 Se selecciona al mejor conjunto de individuos de la población inicial utilizando como función objetivo el error de aproximación de RNAP definido en 5.33.
- 3 Se obtiene una nueva población  $A_g$  utilizando los operadores de recombinación y mutación propios de Algoritmo Genético.
- 4 Se reevalúa la función objetivo utilizando el error de aproximación de RNAP.
- 5 Se regresa al paso 3 hasta que se alcance el máximo número de generaciones o hasta que alguno de los individuos extraídos del proceso de selección satisfaga la desigualdad descrita en 5.35.

Dado que RNAP asigna dinámicamente el número de neuronas y de generaciones necesarias en el entrenamiento, los únicos parámetros manipulables exteriormente son:

*Pot*, representa la máxima potencia a alcanzar en la aproximación polinomial.

*Eta*, representa la dispersión del ruido presente en los datos de entrada.

*Phiest*, define la estructura de la matriz de aprendizaje.

Esta técnica como se verá posteriormente posee una excelente capacidad para modelar a series de tiempo de diferentes características [160, 161, 162].

### 5.3.3 Técnicas de Teoría de Sistemas Dinámicos No Lineales

#### Predict

Esta técnica de modelado se basa en la idea de aproximación con modelos locales mencionada en la descripción de la técnica de predicción *Nstep* de la sección 5.2, en este caso se realiza una aproximación local manteniendo el orden cero de la misma, esto es, se aproxima la dinámica localmente por una constante [3]. En el espacio fase reconstruido, todos los vecinos del punto  $s_n$  dentro de una vecindad  $U_n$  son buscados, para entonces calcular el siguiente punto usando la expresión,

$$\widehat{s_{n+k}} = \frac{1}{|U_n|} \sum_{s_j \in U_n} s_{j+k} \quad (5.39)$$

#### Polynom

Esta técnica de modelado se basa en el mismo algoritmo de la técnica de predicción *Polynomp* descrita en la sección 5.2, la función  $f_p$  para construir el modelo corresponde a un polinomio, los parámetros  $p$  del polinomio ocurren en forma lineal en la función  $f$  y pueden ser encontrados con álgebra lineal y la solución es única. Esta ventaja se pierde si la función a construir posee parámetros que ocurren en forma no lineal [3].

#### rbf

En esta técnica, se aplica nuevamente el algoritmo de construcción de modelos globales descrito para la técnica *Polynomp*, pero ahora la función  $f_p$  corresponde a funciones de base radial [3]. Se define una función escalar  $\Phi(r)$  que tiene solo argumentos positivos  $r$ . Adicionalmente, se seleccionan  $k$  centros  $\mathbf{y}_i$  en el atractor. Incluyendo la función constante como la función base de orden cero, se obtiene

$$f_p(\mathbf{x}) = \alpha_0 + \sum_{i=1}^k \alpha_i \Phi(\|\mathbf{x} - \mathbf{y}_i\|) \quad (5.40)$$

Funciones base típicas son las de forma de campana, con un máximo a  $r = 0$  y un decaimiento rápido hacia cero conforme  $r$  crece. Pero también es posible usar funciones crecientes

o incluso singulares. La función  $f_p$  es modelada ajustando los coeficientes  $\alpha_i$  de las funciones  $\Phi$ . La determinación de los coeficientes  $\alpha_i$  es un problema lineal (se puede resolver con mínimos cuadrados).

### **Sugimay**

Es una técnica de modelado que se basa en la búsqueda del valor a modelar a partir de vecinos cercanos, primeramente se selecciona la dimensión embebida  $m$  de la serie de tiempo y se reconstruyen los puntos de la serie de tiempo en el espacio fase, para ello se utilizan las coordenadas con retraso para representar las secuencias de datos en dicho espacio de dimensión  $m$ :  $\{x_t, x_{t-\tau}, x_{t-2\tau}, \dots, x_{t-(m-1)\tau}\}$ . La modelación llamada predictee se encuentra en el espacio de dimensión  $m$ , para construirla es necesario localizar los puntos cercanos en este espacio fase, para ello se selecciona un vecindario mínimo definido tal que el predictee está contenido dentro del simplex más pequeño (con diámetro mínimo) formado por los  $m + 1$  vecinos cercanos. La modelación se obtiene proyectando el dominio del simplex en su rango, esto es mediante el seguimiento de donde los puntos del simplex terminan a  $p$  pasos de tiempo. Para obtener el valor modelado se calcula donde el predictee original se ha movido dentro del rango del simplex, dando un peso exponencial a sus distancias originales respecto de los vecinos relevantes. Este método es no paramétrico por lo que no utiliza información a priori sobre el modelo usado para generar la serie de tiempo, solamente la información de los datos de la serie [163].

## Capítulo 6

# Técnicas para el Análisis de Conjuntos de Datos Multivariados

### 6.1 Introducción

El amplio conjunto de datos experimentales generados en esta tesis, requirió del uso de técnicas de análisis de estadística multivariada y reconocimiento de patrones, que ayudasen en la extracción, reconocimiento y clasificación de los *patrones* ocultos en los datos, *mediante la reducción de las dimensiones de los conjuntos multidimensionales de datos experimentales* [72, 164]; para la búsqueda de las posibles relaciones entre características de las series de tiempo, su predictibilidad y los modelos de predicción. Para estudiar los diferentes conjuntos de datos, se seleccionaron tres técnicas: Mapas Auto Organizados (GHSOM), Análisis con Escalamiento Multidimensional (MDS) y Análisis de Correlación Bivariada. Los Mapas Auto Organizados permiten estudiar el agrupamiento de los datos en base a una métrica de similitud que además es de carácter jerárquico, por su parte el Análisis con Escalamiento Multidimensional permite observar como se agrupan los datos usando una métrica de tipo Euclidiano. El Análisis de Correlación Bivariada, permite estudiar la correlación que existe entre diferentes variables experimentales y de esta forma inferir posibles relaciones de dependencia entre ellas. La información generada con las técnicas se complementa entre sí, y permitió extraer conocimiento sobre la predictibilidad de las series de tiempo y sus relaciones. En el presente capítulo se describen a detalle las técnicas mencionadas, la sección 6.2 describe la técnica de Mapas Auto

Organizados y su derivación los Mapas Auto Organizados Jerárquicos. La sección 6.3, describe la técnica de Análisis con Escalamiento Multidimensional. Finalmente, la sección 6.4 describe el Análisis de Correlación Bivariada.

## 6.2 Mapas Auto Organizados (SOM y GHSOM)

### 6.2.1 SOM

Un Mapa Auto Organizado (*SOM*) es una red neuronal de alimentación hacia adelante (*feed-forward*) que utiliza un algoritmo de entrenamiento no supervisado, y mediante un proceso llamado auto organización, configura las unidades de salida en una representación topológica de los datos originales. SOM pertenece a una clase general de métodos de redes neuronales, las cuales son técnicas de regresión no lineal que pueden ser entrenadas para aprender o encontrar relaciones entre las salidas y entradas u organizar los datos de tal forma que es posible identificar *patrones* o estructuras existentes entre los datos. SOM reduce datos multidimensionales a un mapa de menor dimensión o malla de neuronas.

El algoritmo SOM se basa en el aprendizaje competitivo no supervisado. Este algoritmo proporciona un mapeo que preserva la topología del espacio de mayor dimensión en el espacio de unidades del mapa (*neuronas*). Las unidades de mapa generalmente forman una malla de dos dimensiones, entonces el mapeo corresponde de una alta dimensionalidad a un plano. La propiedad de preservación de la topología significa que un Mapa Auto Organizado agrupa vectores de datos de entrada similares en neuronas: es decir puntos que están cerca uno del otro en el espacio de entrada son mapeados a unidades de mapa cercanas en el Mapa Auto Organizado. SOM puede ser de utilidad como una herramienta de agrupamiento (*clustering*) así como para visualizar datos de alta dimensionalidad.

El proceso de crear un Mapa Auto Organizado requiere de dos capas de unidades de procesamiento: la primera es la capa de entrada que contiene unidades de procesamiento para cada elemento en el vector de entrada, la segunda es una capa de salida o malla de unidades de procesamiento que están conectadas en forma total con las unidades de la capa de entrada. El número de unidades de procesamiento en la capa de salida se define en base a la forma y tamaño inicial del mapa que se desea construir. A diferencia de otras redes neuronales no hay

capas ocultas o unidades de procesamiento ocultas.

Cuando un patrón de entrada se presenta a la red, las unidades en la capa de salida compiten entre sí por el derecho de ser declaradas ganadoras. El ganador será la unidad de salida cuyos pesos de la conexión de entrada sean los mas cercanos al patrón de entrada en términos de una distancia Euclidiana. Entonces la entrada se presenta y cada unidad de salida compite para ajustarse al patrón de entrada. La salida que es mas cercana al patrón de entrada es declarada ganadora. Los pesos de conexión de la unidad ganadora son ajustados, es decir, movidos en la dirección del patrón de entrada por un factor determinado por la razón de aprendizaje. Esta es la naturaleza básica de las redes neuronales competitivas. La Figura 6-1 muestra un esquema de la arquitectura de un Mapa Auto Organizado.

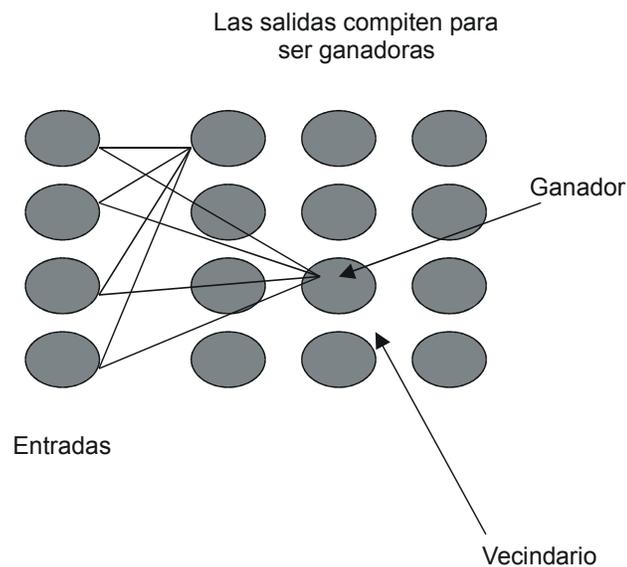


Figura 6-1: Esquema de una red neuronal SOM

SOM crea un mapeo topológico mediante el ajuste no solamente de los pesos ganadores, sino también el de los pesos de las unidades de salida adyacentes que pertenecen a la vecindad cercana del ganador. Entonces no solo el ganador es ajustado sino que todo el vecindario de unidades de salida es movido para ajustarse al patrón de entrada. Cuando se parte de valores aleatorios para los pesos, las unidades de salida lentamente se alinean de tal forma que cuando un patrón de entrada es presentado, una vecindad de unidades responde al patrón de entrada.

Conforme el entrenamiento progresa, el tamaño del vecindario alrededor de la unidad ganadora decrece. Inicialmente, grandes números de unidades de salida serán actualizadas, pero conforme el entrenamiento avanza cada vez un menor número de unidades son actualizadas hasta que finalmente solo la unidad ganadora se ajusta. De forma similar, la razón de aprendizaje decrece conforme el entrenamiento progresa, y en algunas aplicaciones, la razón de aprendizaje decaerá con la distancia de las unidades de salida respecto de las unidades ganadoras.

El resultado son pesos entre los vectores de entrada y las neuronas de salida que representan un prototipo del patrón de entrada para el subconjunto de entradas que caen en un agrupamiento particular (*cluster*). El proceso de tomar un conjunto de datos de alta dimensionalidad y reducirlos a un conjunto de clusters se llama segmentación. El espacio de entrada de alta dimensionalidad se reduce a un mapa de dos dimensiones. Si el índice de la salida ganadora es usada, éste esencialmente particiona los patrones de entrada en un conjunto de categorías o clusters.

Un Mapa Auto Organizado tiene la capacidad de generalizar. Esto implica que la red puede reconocer o caracterizar entradas que nunca había encontrado antes. Una nueva entrada es asimilada con la unidad de mapa que la mapea a ella. Incluso, vectores con datos perdidos pueden ser utilizados y el mapa es capaz de predecir o completar los valores perdidos en base al mapa de entrenamiento [165, 166].

### 6.2.2 GHSOM

En la sección anterior se explicaron los fundamentos de los Mapas Auto Organizados (SOM), los cuales permiten visualizar datos de alta dimensionalidad en un espacio de dos dimensiones, de tal forma, que la similitud entre los datos de entrada es preservada. Sin embargo, estos mapas poseen dos limitantes la primera consiste en la necesidad de conocer a priori la arquitectura de la red en términos del número y arreglo de los elementos neuronales de procesamiento, la segunda es que no permiten la visualización de las posibles relaciones jerárquicas entre los datos agrupados por sus similitud [167]. Lo anterior motiva, el uso de los Mapas Auto Organizados Jerárquicos (*Growing Hierarchical Self-Organizing Maps, GHSOM*) para estudiar las relaciones de agrupamiento y jerarquías de los diferentes conjuntos de datos generados en esta tesis.

La base de los mapas GHSOM, es utilizar una estructura de jerarquías de múltiples capas

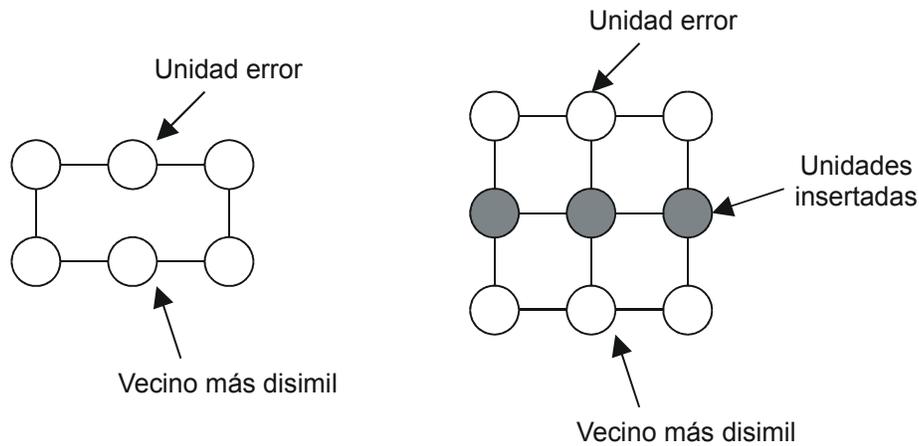


Figura 6-2: Inserción de unidades en un Mapa Auto Organizado Jerárquico

donde cada capa consiste de un número independiente de Mapas Auto Organizados (SOM). SOM se utiliza en la primera capa de la jerarquía, pero además para cada unidad del mapa de esta capa un nuevo mapa SOM se añade formando una segunda capa de la jerarquía. Este principio se repite con una tercera capa, etc. El mapa GHSOM utiliza una arquitectura de red que se incrementa, el tamaño de la red se define durante el proceso de entrenamiento no supervisado. El proceso inicia con una capa virtual 0, la cual consiste de una sola unidad. El vector de pesos de esta unidad se inicializa con el promedio de los datos de entrada. El proceso de entrenamiento empieza con un pequeño mapa, por ejemplo, de tamaño  $2 \times 2$  unidades en la capa 1, que se auto organiza de acuerdo al algoritmo de entrenamiento de SOM descrito en la sección anterior. La diferencia con SOM, consiste en que después de  $\lambda$  iteraciones se localiza a la mayor diferencia dentro del mapa entre un vector de pesos y el conjunto de vectores de entrada que representa, a este vector de pesos se le llama unidad de error. Entre la unidad de error y sus vecinos más disímiles en términos del espacio de entrada, se agrega una columna o fila de nuevas unidades. Los vectores de peso de estas nuevas entradas se inicializan como el promedio de sus vecinos. La Figura 6-2, muestra un ejemplo del proceso de ajuste del mapa GHSOM.

La segunda característica del mapa consiste en la construcción de jerarquías, a partir del mapa SOM original, este se expande a otra capa, en caso de que exista un mapeo de datos de

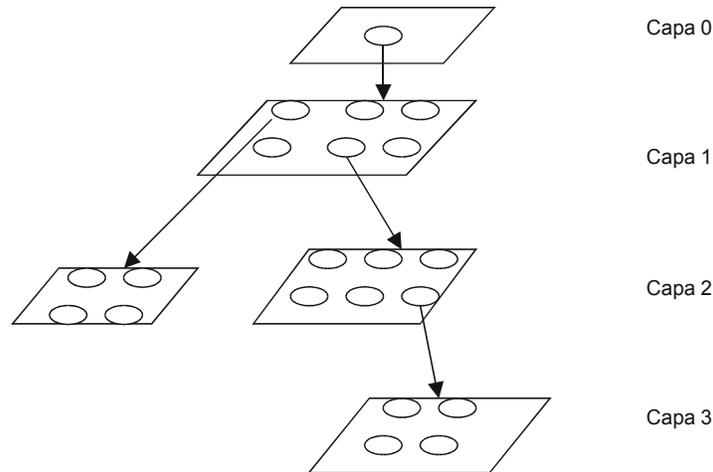


Figura 6-3: Arquitectura de un Mapa Auto-Organizado Jerárquico

entrada disímiles sobre una unidad del mapa SOM original. Estas unidades se identifican por un error, llamado de cuantización, que es mayor a un umbral dado. En la nueva capa, los datos de entrada disímiles se organizan en un nuevo mapa SOM, que se ajusta nuevamente y puede dar lugar o no a nuevas capas, dependiendo del error de cuantización. La profundidad de la jerarquía refleja la no uniformidad que se puede esperar en conjuntos de datos del mundo real [167, 168]. La Figura 6-3, muestra la arquitectura de un mapa GHSOM.

### 6.3 Análisis con Escalamiento Multidimensional (MDS)

El Escalamiento Multidimensional (*MDS*) es un método que representa medidas de *similitud* o *no similitud* entre pares de objetos como distancias entre puntos de *un espacio multidimensional a uno de menor dimensión* (2-dimensional o 3-dimensional) [169]. Los datos, por ejemplo, pueden ser correlaciones entre pruebas de inteligencia, y la representación MDS es un plano que muestra las pruebas como puntos que están cercanos entre sí mientras mas positivamente correlacionados estén las pruebas. La representación gráfica de las correlaciones permite al experto observar y explorar la estructura de los datos, en busca de regularidades ocultas cuando únicamente se analizan arreglos de valores numéricos [170]. En la Figura 6-4 se muestra un ejemplo de un análisis MDS, el cual corresponde a la agrupación de un conjunto de países

en función de tres características: su nivel de desarrollo en ciencia y tecnología, su nivel de desarrollo económico y su pertenencia o no a la OCDE (organización que agrupa a las llamadas economías “más desarrolladas”, a la cual pertenece México). Obsérvese que el análisis MDS permite no solo agrupar y clasificar, sino además, identificar dimensiones relevantes para el conjunto de objetos bajo estudio.

El método MDS tiene su origen en la psicofísica, en la cual se desarrollan modelos que relacionan estímulos con propiedades físicas bien conocidas y sus representaciones perceptuales y cognitivas. Para la psicofísica la noción de escalamiento multidimensional tiene un significado directo en el sentido de que se estudia, cómo dimensiones físicas conocidas son representadas psicológicamente. Su uso se extiende a diferentes ramas de las ciencias sociales (sociología, ciencias políticas, etc.).

Los cuatro propósitos del análisis MDS son:

- Un método para representar similitud o no similitud entre datos mediante distancias en un espacio de baja dimensionalidad, para hacer accesible la *exploración* de los datos en busca de *estructuras ocultas* en las representaciones de arreglos datos.
- Permite la prueba de como un criterio del experto humano para distinguir entre diferentes objetos de interés se refleja en una correspondencia de diferencias empíricas de estos objetos.
- Aproximación analítica de datos que permite el *descubrimiento de dimensiones* que representan juicios de similitud o no similitud.
- Modelo psicológico que explica juicios de no similitud en términos de una regla que mimetiza un tipo particular de función de distancia.

Es importante notar que aunque casi siempre el método MDS hace uso de una geometría euclidiana para el cálculo de las distancias de similitud, no es exclusivo de dicha geometría. La construcción del modelo que utiliza el método MDS, se describirá con fines ilustrativos utilizando como datos de ejemplo al conjunto de parámetros calculados para las series de tiempo [132]. De modo general, se puede decir que el MDS toma como información de entrada una matriz de proximidades,  $\Delta \in M_{n \times n}$ , donde  $n$  es el número de series de tiempo. Cada elemento

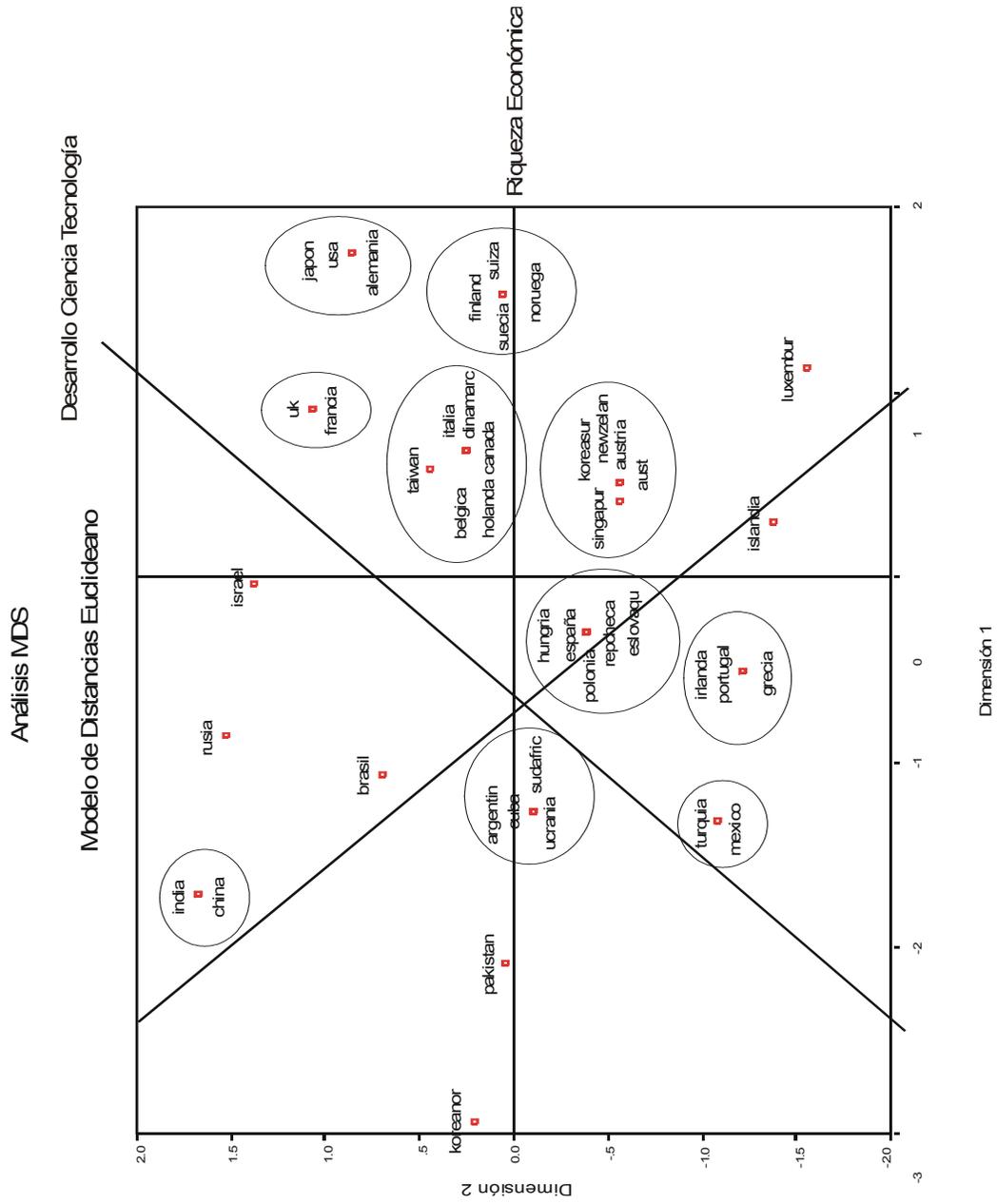


Figura 6-4: Gráfica de MDS para un conjunto de países agrupados por su desarrollo tecnológico y económico

$\delta_{ij}$  de  $\Delta$  representa la proximidad entre los parámetros para las series de tiempo  $ST_i$  y  $ST_j$  donde  $\delta_{ij} = (c_i - c_j)^2$ , siendo  $c_k$  un parámetro de la serie de tiempo  $k$ -ésima. La matriz de proximidades es entonces de la forma:

$$\Delta = \begin{bmatrix} \delta_{11} & \delta_{12} & \dots & \delta_{1n} \\ \delta_{21} & \delta_{22} & \dots & \delta_{2n} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \delta_{n1} & \delta_{n2} & \dots & \delta_{nn} \end{bmatrix} \quad (6.1)$$

El algoritmo inicia con una matriz aleatoria  $X \in M_{n \times m}$ , donde  $n$ , al igual que antes, es el número de series de tiempo, y  $m$  es el número de dimensiones. En este ejemplo se considera  $m = 2$ . Cada valor  $x_{ij}$  representa la coordenada de la  $ST_i$  en la dimensión  $j$ . Así, la matriz  $X$  toma la forma

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix} \quad (6.2)$$

A partir de esta matriz  $X$  se puede calcular la distancia existente entre dos series de tiempo cualesquiera  $i$  y  $j$ , simplemente aplicando la fórmula general de la distancia de Minkowski:

$$d_{ij} = \left[ \sum_{t=1}^m (x_{it} - x_{jt})^p \right]^{\frac{1}{p}} \quad (6.3)$$

donde  $p$  puede ser un valor entre 1 e  $\infty$ , en este caso se toma  $p = 2$ . A partir de estas distancias se puede obtener una matriz de distancias que se denomina  $D \in M_{n \times n}$ ,

$$D = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{bmatrix} \quad (6.4)$$

La solución proporcionada por el MDS debe ser tal que haya la máxima correspondencia entre la matriz de proximidades inicial  $\Delta$  y la matriz de distancias obtenidas  $D$ , lo cuál se logra modificando la matriz  $X$  de forma iterativa de la forma

$$X = \frac{BX}{2n} \quad (6.5)$$

en donde  $B$  tiene como elementos:

$$b_{ij} = \frac{-2\delta_{ij}}{\delta_{ij}} \quad \text{si } i = j \quad (6.6)$$

$$b_{ii} = \sum \sum \frac{2\delta_{ik}}{d_{ik}} \quad \text{si } i = j \quad (6.7)$$

$$b_{ij} = 0 \quad \text{si } d_{ij} = 0 \quad (6.8)$$

Existen dos modelos básicos de MDS que son: el modelo de escalamiento métrico y el modelo de escalamiento no métrico. En el primero de ellos se considera que los datos están medidos en escala de razón o en escala de intervalo y en el segundo que los datos están medidos en escala ordinal. Para el caso de las series de tiempo se aplica el modelo de escalamiento no métrico, ya que es éste el que permite formar clases o agrupaciones, a diferencia del escalamiento métrico que arroja información de tipo cuantitativo [132]. El modelo de escalamiento no métrico no presupone una relación lineal entre las proximidades y las distancias, sino que establece una relación monótona creciente entre ambas, es decir, si  $\delta_{ij} < \delta_{kl} \Rightarrow d_{ij} \leq d_{kl}$ . Para conocer la precisión del modelo se construye una medida de precisión, se sabe que las distancias son una función de las similitudes, es decir,

$$f : \delta_{ij}(x) \rightarrow d_{ij}(x) \quad (6.9)$$

de esta forma se tiene que  $d_{ij} = f(\delta_{ij})$ . Esto no deja ningún margen de error. Sin embargo, en las proximidades empíricas es difícil que se tenga la igualdad, por lo que generalmente se tiene que  $d_{ij} \approx f(\delta_{ij})$  con  $f(\delta_{ij}) = a + b\delta_{ij}$  donde  $a$  y  $b$  son constantes a determinar. A las transformaciones de las proximidades por  $f$  se les llama disparidades. Ahora se puede definir el error cuadrático como

$$e_{ij}^2 = (f(\delta_{ij}) - d_{ij})^2 \quad (6.10)$$

Y la medida de precisión del modelo denominada *S-Stress* se define entonces como

$$S - Stress = \sqrt{\frac{\sum_{i,j} (f(\delta_{ij})^2 - d_{ij}^2)^2}{\sum_{i,j} (d_{ij}^2)^2}} \quad (6.11)$$

Finalmente el proceso completo de MDS consiste de los siguientes pasos [132]:

- 1 Obtención de una matriz  $X \in M_{n \times m}$  de coordenadas aleatorias, que corresponde a las distancias entre las series de tiempo.
- 2 Comparación de las proximidades de  $\Delta$  con las distancias en  $X$ , para obtener las disparidades.
- 3 Cálculo del *S-Stress*.
- 4 Modificación de la matriz  $X$  con la ecuación 6.5, con el objetivo de minimizar el *S-Stress*.
- 5 Regresar al paso 2 hasta que el *S-Stress* alcance el valor deseado.

## 6.4 Análisis de Correlación Bivariada

Cuando medimos objetos sobre dos variables que representan propiedades o parámetros de los mismos, no solo es de interés el medir la tendencia central y la variación de dichas variables,

sino también su *asociación*, si la hay, entre las dos variables. La asociación entre dos variables puede a su vez ser de dos tipos: correlacional y experimental. En relaciones experimentales, el experimentador controla los valores de una de las variables asignándolos en forma aleatoria y observa los cambios provocados en la otra variable que se mide. En relaciones correlacionales, no se tiene control sobre los valores de las variables de los objetos bajo estudio. Únicamente se observa como las dos variables covarian en su ambiente natural. Estas son variables aleatorias en las cuales cualquier objeto tiene una probabilidad de poseer un valor dado de estas variables que no están bajo control del experimentador. La distinción básica entre las dos relaciones yace en si el experimentador asigna valores a la variable de los objetos en una forma aleatoria sin bias o si la asignación es histórica por naturaleza por medios desconocidos. Finalmente llegamos a la definición de *análisis de correlación*, el estudio de las relaciones que existen entre variables aleatorias, incluyendo la identificación y síntesis de tales relaciones. En estudios experimentales controlados, interpretaciones causales pueden ser hechas sobre las relaciones observadas. Por otra parte, en investigaciones correlacionales no existen interpretaciones causales que pueden ser hechas con seguridad.

Ya sea que las relaciones sean correlacionales o experimentales, se pueden etiquetar con varias expresiones que identifican la forma del *patrón* de las asociaciones existentes entre los valores de las variables involucradas. Por ejemplo, puede darse el caso que ambas variables posean valores con tendencia creciente o tendencia decreciente en dicho caso se dice que están relacionadas en forma positiva o directa. Pero también puede darse que cuando los valores de una variable crecen los de la otra decrecen, en cuyo caso existe una relación negativa o inversa. También puede darse el caso que en un rango de valores las variables presenten una relación directa y en otro rango de valores presenten una relación inversa, a este tipo de relaciones se les denomina *no monótonas*. Además las relaciones pueden ser de tipo lineal o *no lineal*, también puede ocurrir que las relaciones sean cíclicas. Por último cuando se encuentra que no existe una asociación sistemática entre los valores de dos variables se dice que las variables no están relacionadas, son *no correlacionadas*, son *ortogonales* o *independientes* entre sí [171].

Existen diferentes métodos de correlación bivariada, en todos ellos se busca calcular un coeficiente de correlación que cuantifique la relación entre dos variables (su grado de asociación), este coeficiente de correlación se identifica por la letra  $r$ . Su expresión más conocida esta dada

por

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{ns_x s_y} \quad (6.12)$$

donde  $(x_i - \bar{x})$  es la desviación del valor individual de un objeto sobre la variable  $x$  con respecto de la media de dicha variable,  $(y_i - \bar{y})$  es la desviación del valor de un objeto sobre la variable  $y$  con respecto de la media de dicha variable,  $s_x$  y  $s_y$  son las desviaciones estándar de las variables  $x$  e  $y$ , por último  $n$  es el número de pares de observaciones. El rango de valores de correlación se encuentra entre -1 y 1, el signo del coeficiente indica la tendencia de la relación (directa o inversa), y su magnitud indica la fuerza de la misma, por ejemplo valores cercanos a uno corresponden a una asociación fuerte entre las variables.

Entre los coeficientes más utilizados se encuentra el coeficiente de correlación de Pearson que mide asociaciones de tipo *lineal* entre dos variables. A pesar de la limitante del coeficiente de ser una medida de asociación lineal, el análisis de correlación es de mucha utilidad para explorar las relaciones entre los diferentes parámetros de las series de tiempo, la predictibilidad de las mismas y los modelos de predicción.

## Parte II

# Predictibilidad de Series de Tiempo

## Capítulo 7

# Cálculo de Parámetros de las Series de Tiempo

### 7.1 Introducción

En este capítulo, se describe al conjunto de series de tiempo utilizadas en los experimentos desarrollados a lo largo de esta tesis, así como al conjunto de datos experimentales obtenidos a partir del cálculo de los parámetros de las series de tiempo. En la sección 7.2 se presenta la descripción del conjunto experimental de series de tiempo utilizado en este trabajo, la sección 7.3 presenta una sinopsis de los diferentes parámetros seleccionados para caracterizar a las series de tiempo, la sección 7.4 presenta una breve descripción de los conjuntos de datos experimentales derivados del cálculo de los parámetros para cada serie de tiempo, por último la sección 7.5 presenta la discusión sobre este capítulo.

### 7.2 Descripción del Conjunto Experimental de Series de Tiempo

Para realizar el análisis de series de tiempo se seleccionaron un conjunto de 29 series, el criterio de selección fue el siguiente: que fueran una muestra representativa de las series de tiempo utilizadas en el análisis y evaluación de técnicas de predicción que se han reportado en la literatura especializada [10], y que fueran representativas de una clasificación básica de referencia, en este caso, en base a su comportamiento dinámico (periódico, cuasi periódico, caótico, complejo y

estocástico), esta forma de clasificar a las series de tiempo fue un primer intento de entender la predictibilidad de las mismas, usando como base la idea de que el comportamiento dinámico observado indicaba mayor predictibilidad en el caso de las series periódicas y en un conjunto discreto de clases de comportamiento dinámico, una menor predictibilidad correspondería a las series con comportamiento estocástico o aleatorio, dicha clasificación fue propuesta originalmente en forma independiente por Figueroa-Nazuno et. al. y Sprott en diversas publicaciones [8, 131, 172, 173].

Este conjunto de series de tiempo por sus características dinámicas son series con estructuras muy variadas desde las más simples como las de tipo periódico hasta series complejas como las correspondientes a fenómenos naturales tales como un sismo, las series son utilizadas ampliamente en los trabajos reportados en la literatura especializada, por lo que son una muestra representativa del estudio experimental sobre la predicción de series de tiempo, las series seleccionadas a diferencia de las series de los sistemas comúnmente analizados en aplicaciones como demografía, econometría, comercio, etc., son series que llevan a las técnicas de predicción al límite de sus capacidades de modelado. A continuación se describe a cada una de las series de tiempo estudiadas en esta tesis.

1 *Sine*. Serie periódica generada para 10 ciclos generada por la función

$$f(x) = \text{Seno}(x) \quad (7.1)$$

2 *Vanderpol*. Serie periódica generada por una ecuación diferencial, que es modelo del paso de la carga eléctrica (denotada por  $y$ ) a través de un circuito oscilador de un tubo de vacío (*triode*) [59]. Su ecuación es:

$$\frac{d^2y}{dt^2} + (y^2 - \eta)\frac{dy}{dt} + \omega^2y = 0 \quad (7.2)$$

3 *Qperiodic2*. Serie cuasi periódica obtenida de la medición de una variable de velocidad (cm/s) en un experimento anular para reproducir un Flujo de Couette (Reología), las mediciones se obtuvieron con un muestreo cada 0.1 segundos. Nota, las condiciones experimentales no se proporcionan en la literatura.

4 *Qperiodic3*. Serie cuasi periódica obtenida de la medición de una variable de velocidad (cm/s) en un experimento anular para reproducir un Flujo de Couette (Reología), las mediciones se obtuvieron con un muestreo cada 0.4 segundos. Nota, las condiciones experimentales no se proporcionan en la literatura.

5 *Mackey-Glass*. Serie caótica generada por una ecuación diferencial de retardo temporal: modelo de formación de células sanguíneas blancas (linfocitos) [156, 174]. La ecuación es de la forma:

$$\frac{dx}{dt} = -bx(t) + \frac{ax(t-\tau)}{1 + [x(t-\tau)]^{10}} \quad (7.3)$$

donde  $a = 0.2$ ,  $b = 0.1$  y  $\tau = 3000$ .

6 *Logistic*. Serie caótica generada por un mapa: este mapa se puede pensar como un modelo ecológico de las variaciones anuales de poblaciones de insectos [59]. Su expresión matemática es:

$$x_{n+1} = rx_n(1 - x_n) \quad (7.4)$$

donde  $n = \text{año}$ ,  $x = \text{Número de insectos que nacen}$ ,  $r = \text{Número de huevos puestos por cada insecto en promedio que eclosionan al año } n+1$ .

7 *Lorenz*. Serie caótica generada por un sistema de ecuaciones diferenciales: modelo de convección de fluidos (convección de Rayleigh-Benard) la cual se presenta en la atmósfera terrestre [59, 140, 175]. El sistema de ecuaciones es de la forma:

$$\begin{aligned} \frac{dX}{dt} &= -\tilde{\sigma}X + \tilde{\sigma}Y \\ \frac{dY}{dt} &= -XZ + \tilde{r}X - Y \\ \frac{dZ}{dt} &= XY - \tilde{b}Z \end{aligned} \quad (7.5)$$

donde  $\tilde{\sigma} = 10$ ,  $\tilde{r} = 28$  y  $\tilde{b} = 8/3$  son parámetros adimensionales,  $X$  es proporcional a la velocidad del flujo de fluido circulatorio,  $Y$  caracteriza la diferencia de temperatura entre

regiones de fluido ascendentes y descendentes y  $Z$  caracteriza la distorsión del perfil de temperatura vertical con respecto de su variación de equilibrio.

- 8 *Rossler*. Serie caótica generada por un sistema de ecuaciones diferenciales: modelo simplificado de Lorenz [156, 176]. El sistema de ecuaciones tiene la forma:

$$\begin{aligned}\dot{x} &= -(y + z) \\ \dot{y} &= x + 0.2y \\ \dot{z} &= 0.4 + xz - 5.7z\end{aligned}\tag{7.6}$$

- 9 *Ikeda*. Serie caótica generada a partir de un mapa construido en el plano complejo: modelo de la dinámica de pulsos de luz que viajan a través de un medio no lineal [59]. La expresión matemática es de la forma:

$$z(n+1) = a + R * \exp(i(\phi - \frac{p}{(1 + |z^2(n)|)}))\tag{7.7}$$

donde  $z(n)$  representa al pulso que viaja a través de dicho medio. Los parámetros tienen los valores  $a = 1$ ,  $R = 0.9$ ,  $\phi = 0.4$  y  $p = 6$ .

- 10 *Henon*. Serie caótica generada a partir de un mapa: modelo simplificado del mapa de Poincaré para el modelo de Lorenz [59, 177]. La expresión matemática es de la forma:

$$\begin{aligned}a &= 1.4; \\ b &= 0.3; \\ x(n+1) &= 1 - a * x^2(n) + y(n); \\ y(n+1) &= b * x(n);\end{aligned}\tag{7.8}$$

- 11 *Cantor*. Serie caótica generada por el conjunto de Cantor (teoría de conjuntos), el cual es un conjunto cerrado que consiste enteramente de puntos de frontera cada uno de los cuales es un punto límite de dicho conjunto [59].

- 12 *Tent*. Serie caótica generada por un mapa de tipo lineal por partes [59]. Su expresión matemática es de la forma:

$$x_{n+1} = 1 - 2 \left| x_n - \frac{1}{2} \right| \quad (7.9)$$

13 *A1 (Concurso Santa Fe)*. Serie compleja obtenida a partir de mediciones experimentales de la intensidad de un láser NH3 Infrarrojo Lejano, las características del láser son las siguientes: láser de onda continua en el infrarrojo lejano, longitud de onda de 81.5 micras, el láser es bombeado ópticamente por un láser de N2O [14].

14 *D1 (Concurso Santa Fe)*. Serie compleja generada por un modelo de la dinámica de una partícula amortiguada en un potencial de interacción [14]. El potencial de Interacción es de la forma:

$$V = a_4(x_1^2 + x_2^2 + x_3^2 + x_4^2)^2 - a_2(x_1^2 * x_2^2)^{\frac{1}{2}} - a_1 * x_1 \quad (7.10)$$

La fuerza se expresa como:

$$F = A * Sen(\omega * t) \quad (7.11)$$

en la dirección  $x_3$ , y la disipación es igual a:

*Disipación* =  $-\gamma * Velocidad$ . El valor de  $a_1$  tiene un pequeño desplazamiento producido por la integración de una variable aleatoria Gaussiana. El observable que se obtiene es:

$$\sqrt{(x_1 - 0.3)^2 + (x_2 - 0.3)^2 + x_3^2 + x_4^2} \quad (7.12)$$

15 *Laser (Concurso Santa Fe)*. Serie compleja obtenida a partir de mediciones experimentales de la intensidad de pulsos de láser NH3 Infrarrojo Lejano, condiciones de la frecuencia: *frecuencia serie Laser*  $\geq 3 * frecuencia serie A1$  [14].

16 *Dow Jones*. Serie compleja obtenida a partir del Índice Industrial Dow Jones del NYSE (New York Stock Exchange), la serie corresponde al promedio semanal de los precios al cierre de operaciones para el periodo del 01/07/1900 al 03/08/1996.

- 17 *Kobe*. Serie compleja obtenida a partir de un acelerograma del sismo de Kobe del 16 de enero de 1995. La serie corresponde a la aceleración vertical, las mediciones fueron realizadas en Hobart, Australia por la Universidad de Tasmania, a partir de las 20:56:51 (GMT) durante 51 minutos a intervalos de 1 segundo.
- 18 *EEG*. Serie compleja obtenida a partir de un electroencefalograma humano [3].
- 19 *ASCIITXT*. Serie compleja generada a partir de código de texto ASCII.
- 20 *El niño*. Serie compleja obtenida a partir de la medición experimental de la dinámica de una variable del fenómeno climático el niño.
- 21 *HIV DNA*. Serie compleja obtenida a partir del código del DNA del Virus de Inmunodeficiencia Humana HIV (1 = A, 2 = C, 3 = G, 4 = T).
- 22 *Human DNA*. Serie compleja obtenida a partir del código del DNA Humano.
- 23 *Lovaina (Concurso Universidad de Lovaina)*. Serie compleja generada a partir del modelo generalizado del circuito de Chua [27]. La expresión matemática es de la forma:

$$\begin{aligned}
 \dot{x}_1 &= \alpha [x_2 - h(x_1)] \\
 \dot{x}_2 &= x_1 - x_2 + x_3 \\
 \dot{x}_3 &= -\beta x_2
 \end{aligned} \tag{7.13}$$

donde

$$h(x_1) = m_5 x_1 + \frac{1}{2} \sum_{i=1}^5 (m_{i-1} - m_i) (|x_1 + c_i| - |x_1 - c_i|) \tag{7.14}$$

y los valores de los parámetros son:  $\alpha = 9$ ,  $\beta = 14.286$  y de los vectores:

$$m = [0.9/7, -3/7, 3.5/7, -2.7/7, 4/7, -2.4/7] \text{ y } c = [1, 2.15, 3.6, 6.2, 9].$$

- 24 *Plasma*. Serie compleja obtenida a partir de la medición de una variable de un experimento con plasma.
- 25 *Primos*. Serie compleja generada a partir de un conjunto de números primos.

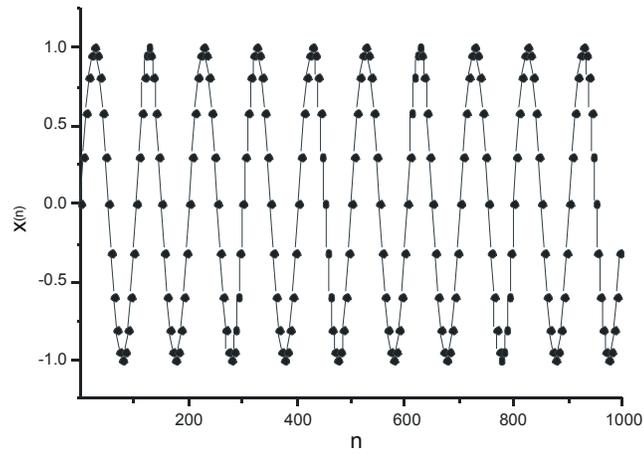
- 26 *SP500*. Serie compleja obtenida a partir del Índice Financiero de Standard & Pool para las 500 empresas mas importantes de la bolsa de valores de Nueva York.
- 27 *Star*. Serie compleja obtenida a partir de la medición de la intensidad luminosa de una estrella variable.
- 28 *Brownian Motion*. Serie estocástica generada a partir del modelado del movimiento browniano (proceso de ruido blanco integrado).
- 29 *White Noise*. Serie estocástica generada a partir del modelado de proceso de ruido blanco (ruido aleatorio uniforme).

A continuación en las Figuras 7-1 a 7-15, se presentan las gráficas de cada una de las series de tiempo, los datos usados para graficar a cada serie de tiempo corresponden a los primeros 1000 datos, para tener una mejor visualización en las gráficas presentadas de la dinámica de las series, los datos fueron muestreados con una razón de 1/5, la razón de muestreo conserva la información relevante de la estructura de las series de tiempo al graficar un total de 200 datos [178].

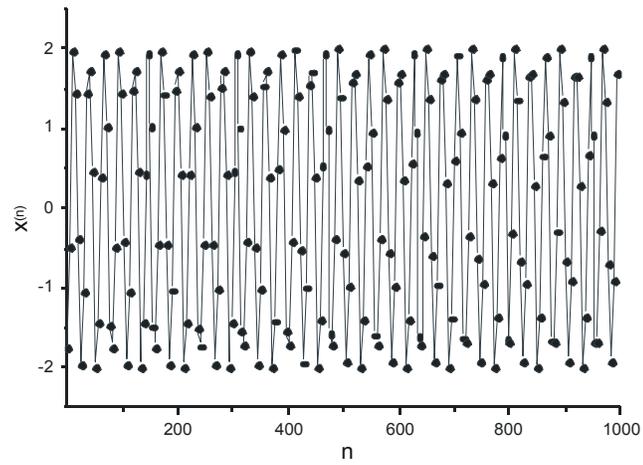
La Figura 7-16, muestra la tabla con el tipo de comportamiento dinámico de cada serie y si ésta es medida experimentalmente (serie natural) o generada artificialmente (serie sintética). El conjunto de series seleccionado presenta los diferentes tipos de dinámicas correspondientes a la clasificación de Figueroa-Nazuno et. al. [8, 131, 172].

### 7.3 Parámetros de las Series de Tiempo

Para la caracterización de las series de tiempo se seleccionaron un conjunto de técnicas de análisis (descritas en el capítulo 4), las cuales permiten el cálculo de diferentes parámetros de las mismas. A su vez del universo de parámetros se eligieron a aquellos que representan una característica particular de forma única, es decir que en el conjunto de características de una serie de tiempo no se tenga redundancia en cuanto a la información que se proporciona sobre la misma, es decir los parámetros deben ser *ortogonales* entre sí, la condición de ortogonalidad se determina en base a un análisis de correlación de los parámetros dicho análisis se presenta posteriormente en el capítulo 9.

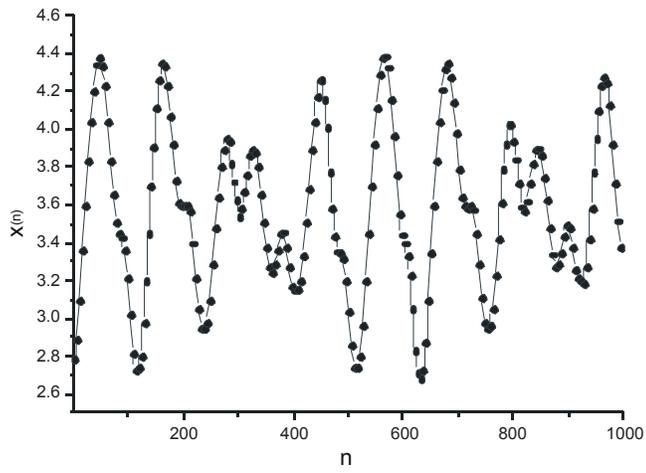


Serie de Tiempo Sine

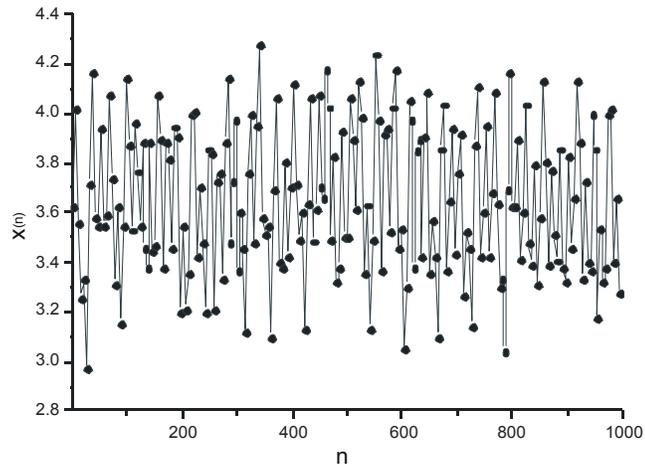


Serie de Tiempo Vanderpol

Figura 7-1

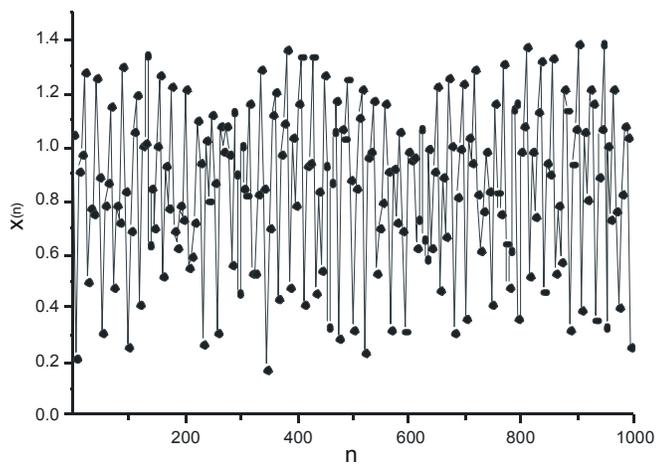


Serie de Tiempo Qperiodic2

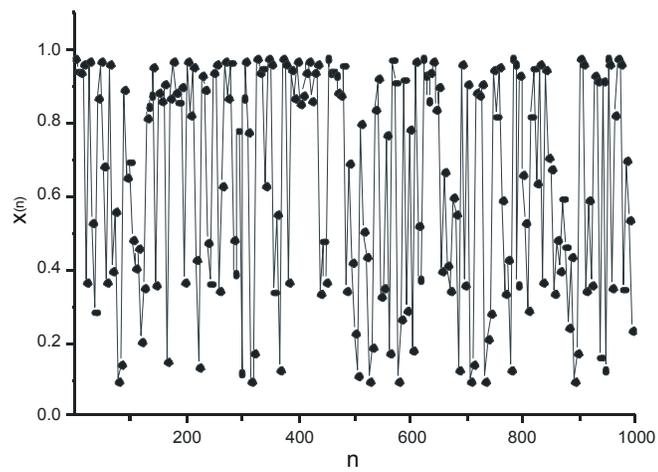


Serie de Tiempo Qperiodic3

Figura 7-2:

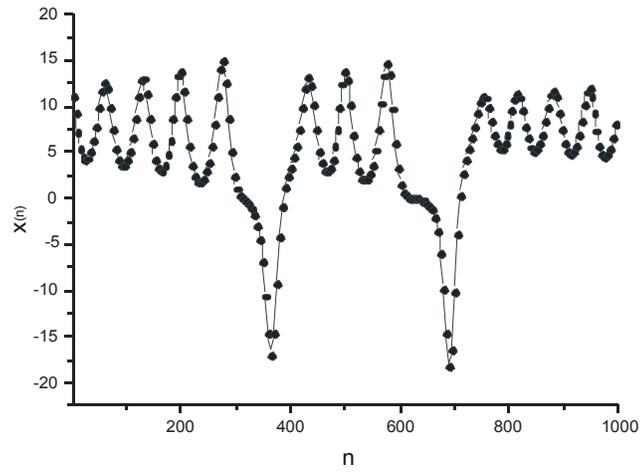


Serie de Tiempo Mackey-Glass

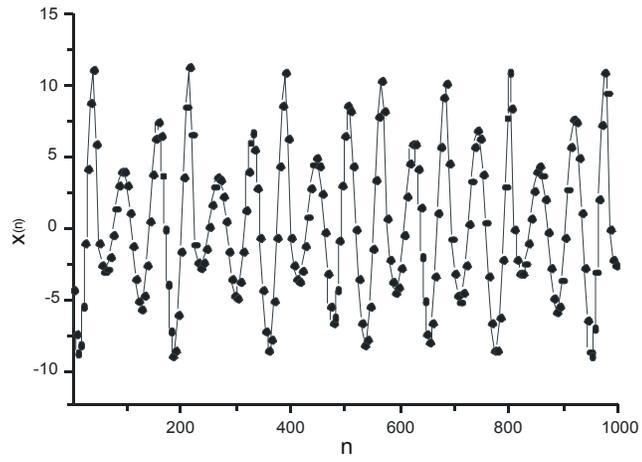


Serie de Tiempo Logistic

Figura 7-3:

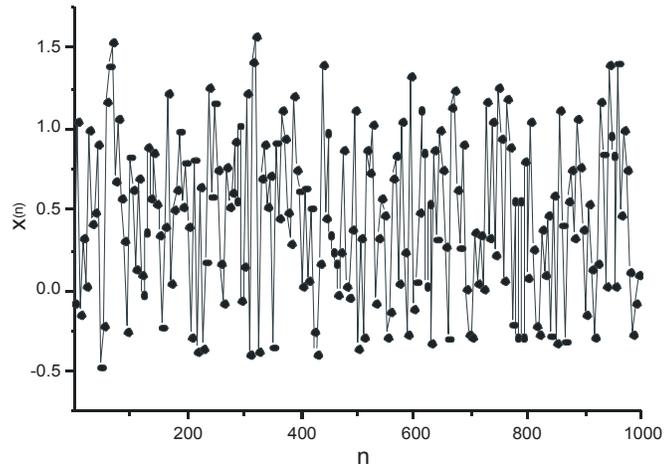


Serie de Tiempo Lorenz

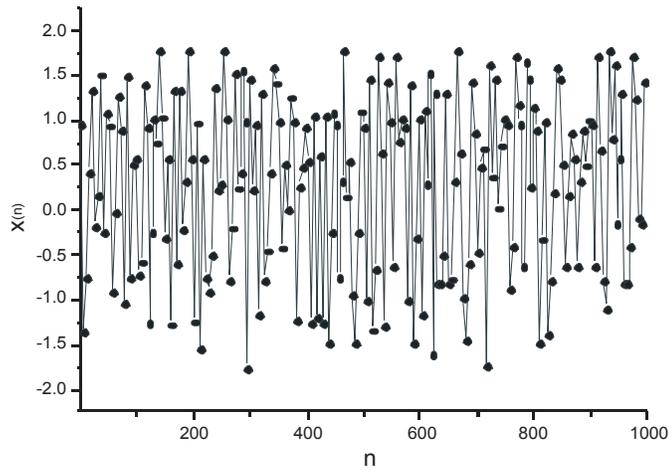


Serie de Tiempo Rossler

Figura 7-4

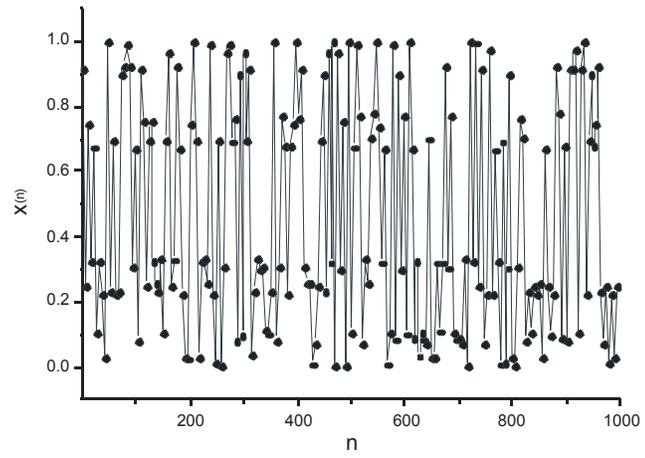


Serie de Tiempo Ikeda

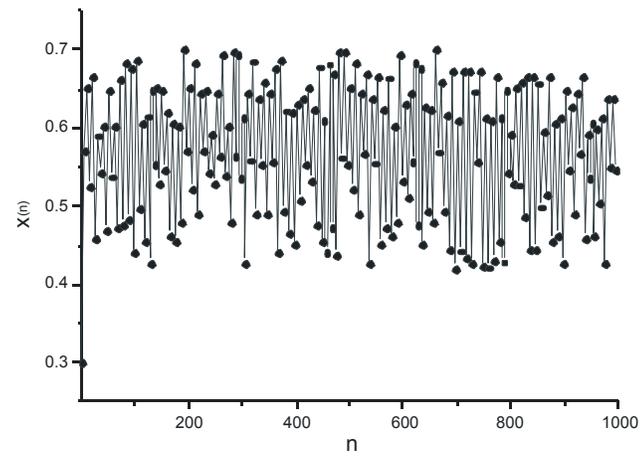


Serie de Tiempo Henon

Figura 7-5:

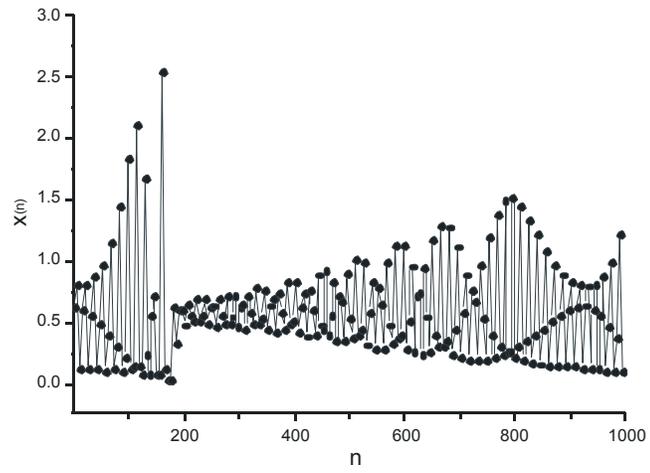


Serie de Tiempo Cantor

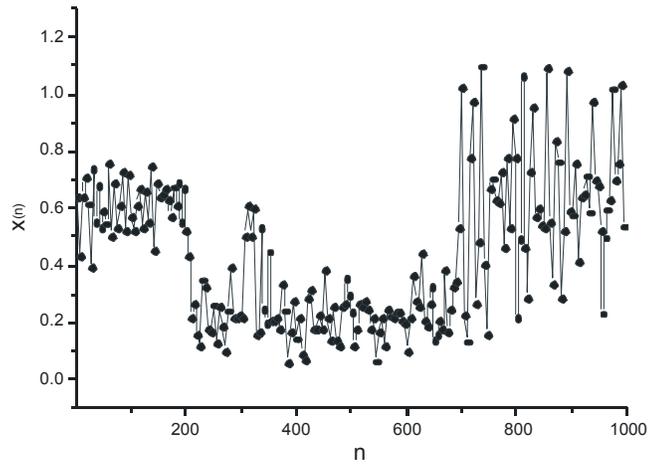


Serie de Tiempo Tent

Figura 7-6:

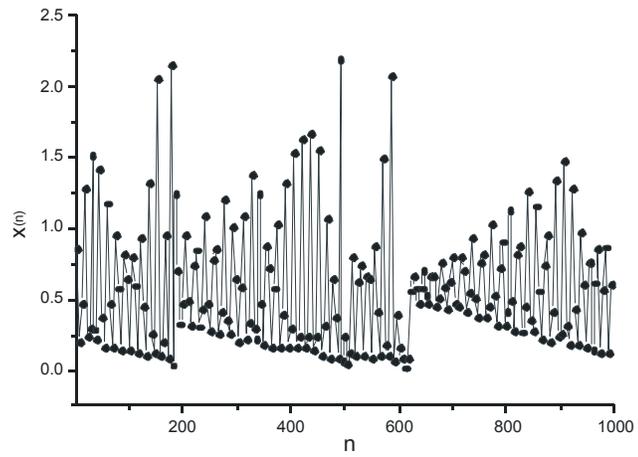


Serie de Tiempo A1

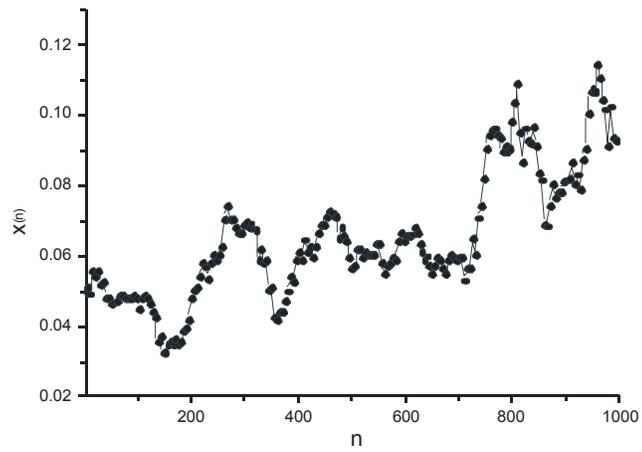


Serie de Tiempo D1

Figura 7-7

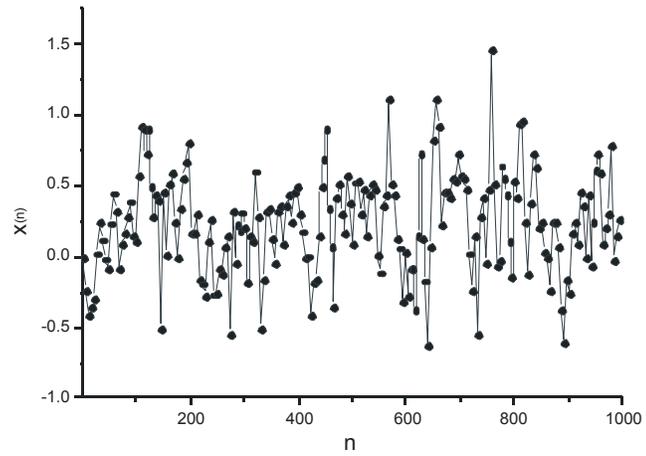


Serie de Tiempo Laser

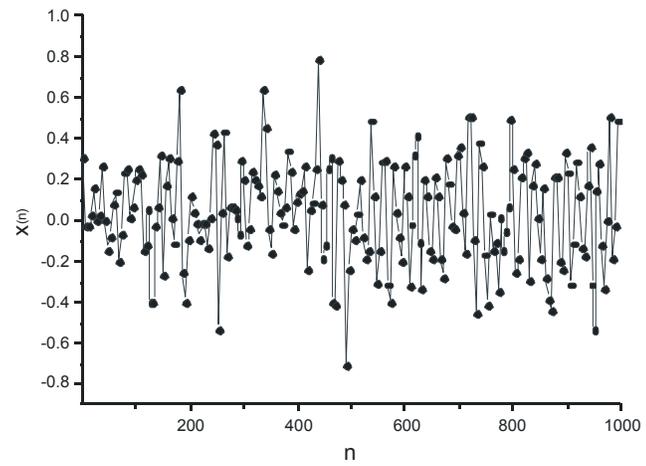


Serie de Tiempo Dow Jones

Figura 7-8:

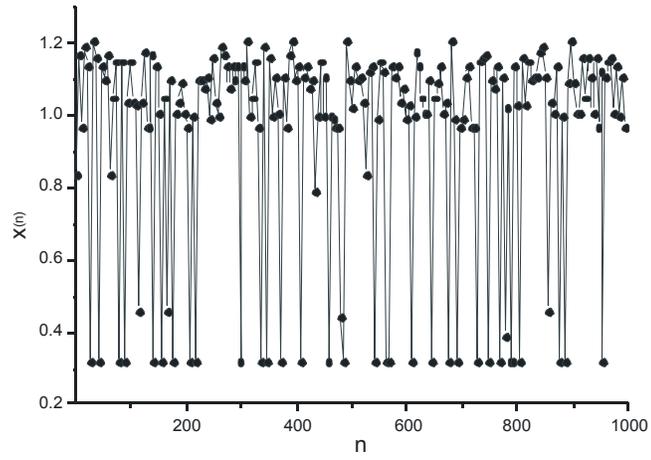


Serie de Tiempo Kobe

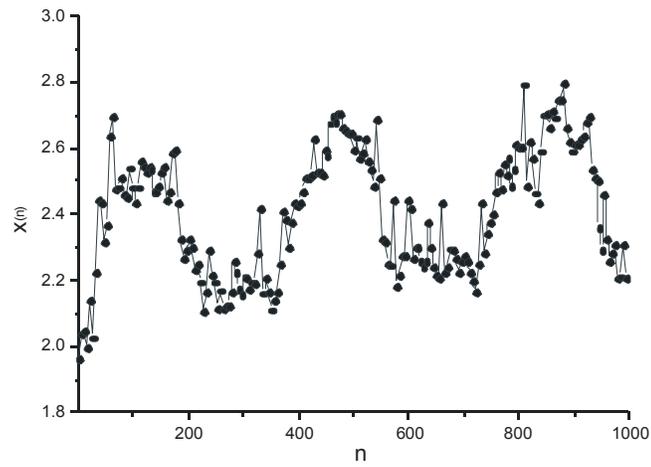


Serie de Tiempo EEG

Figura 7-9:

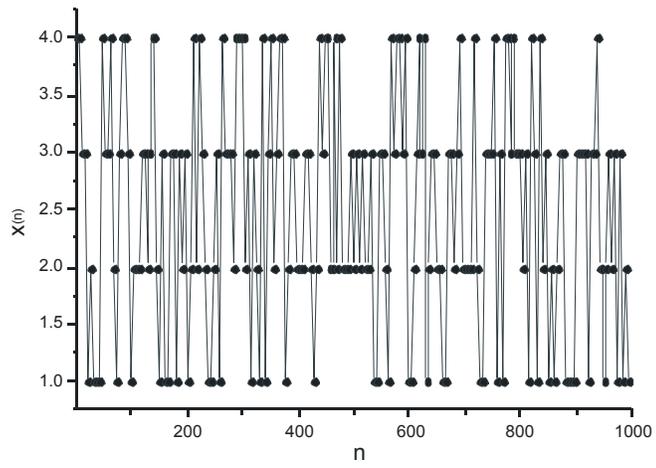


Serie de Tiempo ASCII TXT

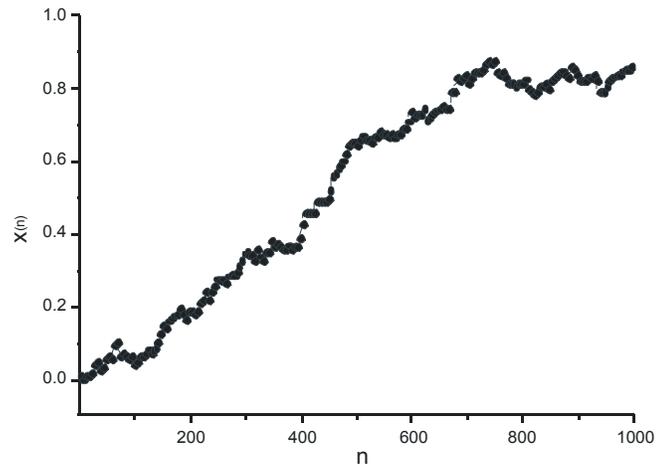


Serie de Tiempo El niño

Figura 7-10:

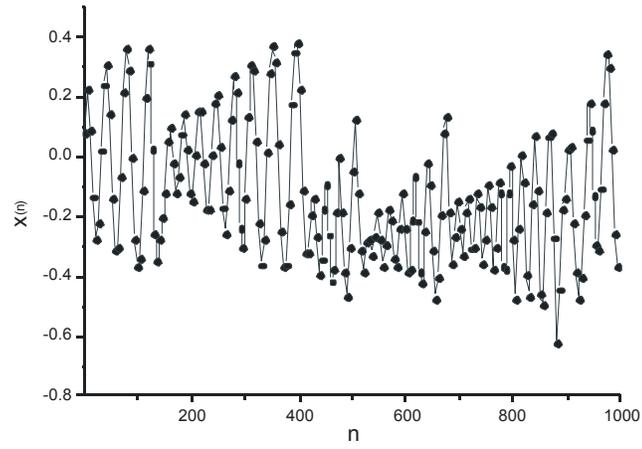


Serie de Tiempo HIV DNA

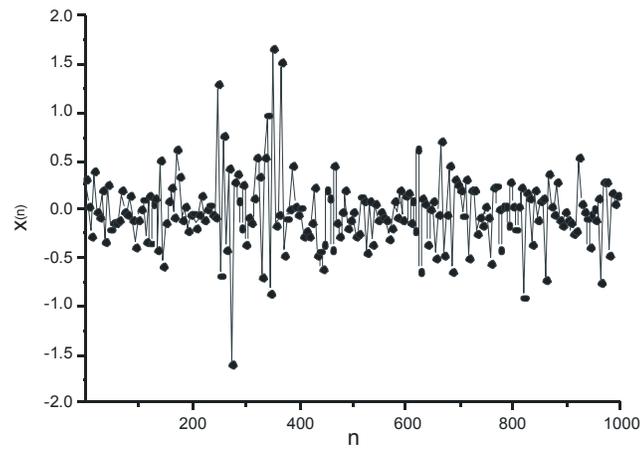


Serie de Tiempo Human DNA

Figura[7-11:

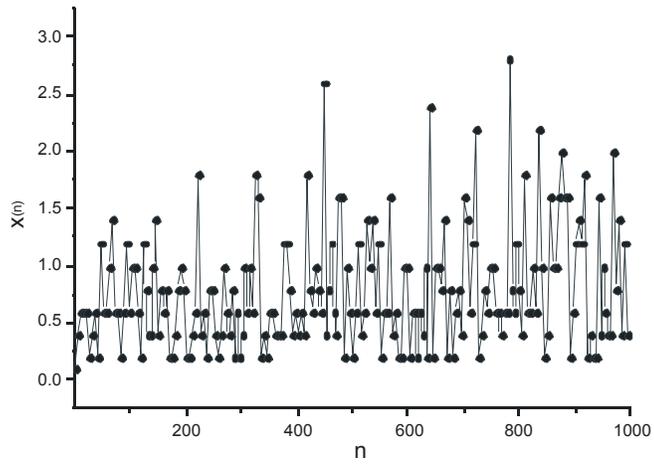


Serie de Tiempo Lovaina

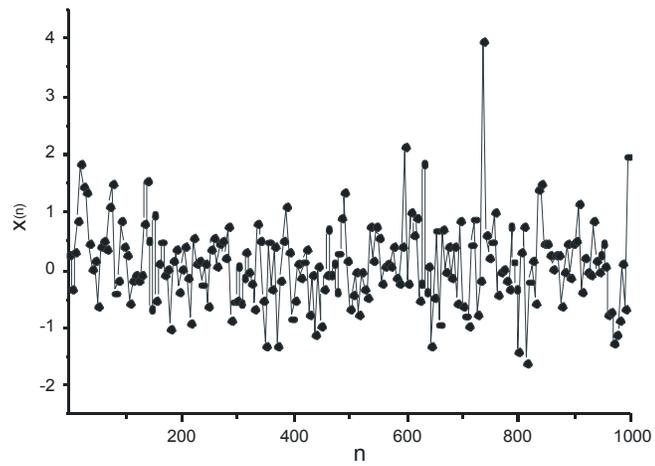


Serie de Tiempo Plasma

Figura 7-12:

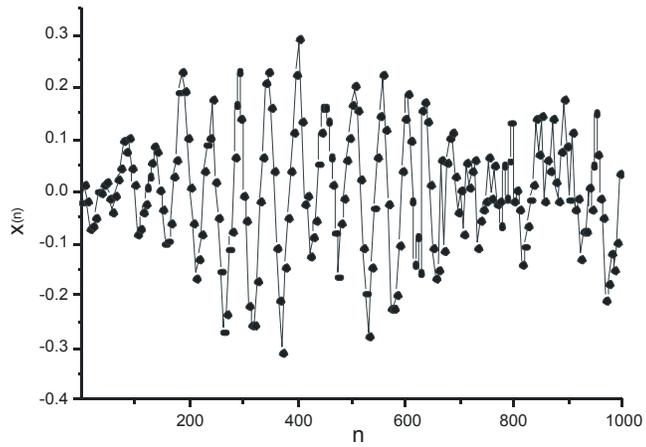


Serie de Tiempo Primos

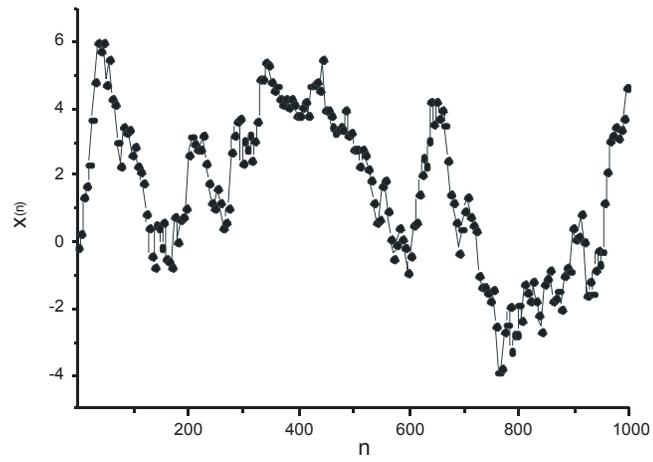


Serie de Tiempo SP500

Figura[7-13:

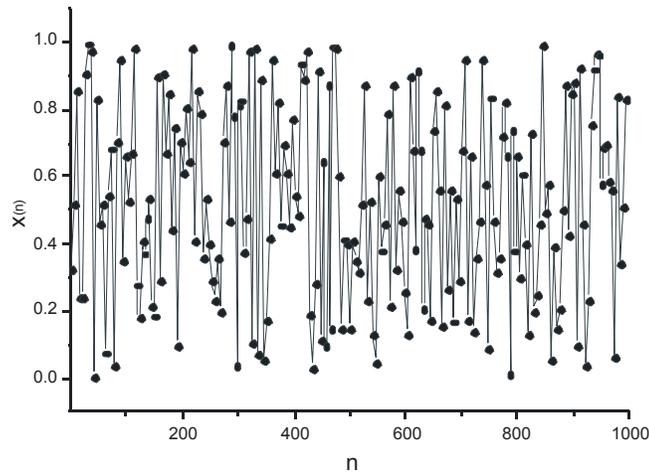


Serie de Tiempo Star



Serie de Tiempo Brownian Motion

Figura 7-14:



Serie de Tiempo White Noise

Figura 7-15:

Las diferentes formas de analizar una serie de tiempo proporcionan una gran cantidad de información expresada en forma de *valores escalares* que representan un parámetro calculado a partir de los datos de la serie de tiempo (por ejemplo, el exponente de Lyapunov) o también en forma de gráficas que visualizan el comportamiento de la serie de tiempo bajo una transformación dada (por ejemplo, el espectro de potencia). Ya que entre los objetivos de este trabajo está el desarrollo de métricas que cuantifiquen la predictibilidad, así como, el estudio de las relaciones entre parámetros de las series, su predictibilidad y los modelos de predicción; el requisito de que un valor escalar exprese una característica de la serie fue otro criterio, además de la ortogonalidad, para seleccionar a los parámetros evaluados en el presente trabajo.

Las Figuras 7-17, 7-18 y 7-19, presentan tablas con una sinopsis de las características de los parámetros calculados: su base teórica, su clasificación por tipo, su alcance, *que información proporciona*, el espacio donde se calcula, la representación de la serie sobre la que opera, el tipo de transformación que realiza sobre los datos y si es un *indicador de predictibilidad*.

Serie de Tiempo	Comportamiento Dinámico	Tipo de Origen
Sine	Periódico	Artificial
Vanderpol	Periódico	Artificial
Qperiodic2	Cuasiperiódico	Natural
Qperiodic3	Cuasiperiódico	Natural
Mackey-Glass	Caótico	Artificial
Logistic	Caótico	Artificial
Lorenz	Caótico	Artificial
Rossler	Caótico	Artificial
Ikeda	Caótico	Artificial
Henon	Caótico	Artificial
Cantor	Caótico	Artificial
Tent	Caótico	Artificial
A1	Complejo	Natural
D1	Complejo	Artificial
Laser	Complejo	Natural
Dow Jones	Complejo	Natural
Kobe	Complejo	Natural
EEG	Complejo	Natural
ASCIITXT	Complejo	Artificial
El niño	Complejo	Natural
HIV DNA	Complejo	Natural
Human DNA	Complejo	Natural
Lovaina	Complejo	Artificial
Plasma	Complejo	Natural
Primos	Complejo	Artificial
SP500	Complejo	Natural
Star	Complejo	Natural
Brownian Motion	Estocástico	Artificial
White Noise	Estocástico	Artificial

Figura 7-16: Tabla de dinámica y origen de las series de tiempo

Parámetros	Base Teórica	Tipo de Parámetro	Alcance de Parámetro	Tipo de Información que proporciona	Espacio en donde se calcula	Representación de la Serie de Tiempo	Transformación de la Información	Indican Predictibilidad: 1. No Indican Predictibilidad: 0
Rango de Datos	Estadística Clásica	Estadístico	Global	Rango de valores que toma la serie de tiempo	Espacio Euclideo	Serie de Tiempo	Sin cambio	0
Resolución	Estadística Clásica	Estadístico	Global	Granularidad de la serie de tiempo	Espacio Euclideo	Serie de Tiempo	Sin cambio	0
Desviación Promedio	Estadística Clásica	Estadístico	Global	Medida de dispersión de la distribución de datos de la serie	Espacio Euclideo	Serie de Tiempo	Sin cambio	0
Desviación Standard	Estadística Clásica	Estadístico	Global	Medida de dispersión de la distribución de datos de la serie	Espacio Euclideo	Serie de Tiempo	Sin cambio	0
Skewness	Estadística Clásica	Estadístico	Global	Grado de asimetría de la distribución de datos de la serie	Espacio Euclideo	Serie de Tiempo	Sin cambio	0
Kurtosis	Estadística Clásica	Estadístico	Global	Medida de dispersión de la distribución de datos de la serie	Espacio Euclideo	Serie de Tiempo	Sin cambio	0
Moda	Estadística Clásica	Estadístico	Global	Valor más frecuente de la distribución de valores	Espacio Euclideo	Serie de Tiempo	Sin cambio	0
Correlación de Pearson	Estadística Clásica	Estadístico	Global	Grado de correlación entre dos valores de la serie	Espacio Euclideo	Serie de Tiempo	Sin cambio	1
Tiempo de Correlación	Estadística Clásica	Estadístico	Global	Intervalo máximo de correlación entre dos valores en la serie	Espacio Euclideo	Serie de Tiempo	Sin cambio	1

Figura 7-17: Tabla de características de los parámetros de las series de tiempo

Parámetros	Base Teórica	Tipo de Parámetro	Alcance de Parámetro	Tipo de Información que proporciona	Espacio en donde se calcula	Representación de la Serie de Tiempo	Transformación de la Información	Indican Predictibilidad: 1. No Indican Predictibilidad: 0
Valor Promedio	Estadística Clásica	Estadístico	Global	Valor promedio de la distribución de valores de la serie	Espacio Euclideo	Serie de Tiempo	Sin cambio	0
Valor Medio (Mediana)	Estadística Clásica	Estadístico	Global	Valor medio de la distribución de valores de la serie	Espacio Euclideo	Serie de Tiempo	Sin cambio	0
Valor Mínimo	Estadística Clásica	Estadístico	Global	Mínimo valor encontrado en la distribución de datos	Espacio Euclideo	Serie de Tiempo	Sin cambio	0
Valor Máximo	Estadística Clásica	Estadístico	Global	Máximo valor encontrado en la distribución de datos	Espacio Euclideo	Serie de Tiempo	Sin cambio	0
Cuartil Mínimo	Estadística Clásica	Estadístico	Global	Valor que indica el 25% de datos de la distribución	Espacio Euclideo	Serie de Tiempo	Sin cambio	0
Cuartil Máximo	Estadística Clásica	Estadístico	Global	Valor que indica el 75% de datos de la distribución	Espacio Euclideo	Serie de Tiempo	Sin cambio	0
Exponente de Hurst	Estadística Moderna	Estadístico	Global	Correlación de largo alcance (memoria)	Espacio Euclideo	Serie de Tiempo	Sin cambio	1
Componentes Principales (5 primeras)	Análisis de Componentes Estructurado	Espacial/Temporal	Global	Información contenida en cada subserie componente sobre la serie de tiempo original	Espacio Euclideo	Componentes de Serie de Tiempo	Separación de la Información	1

Figura 7-18: Tabla de características de los parámetros de las series de tiempo

Parámetros	Base Teórica	Tipo de Parámetro	Alcance de Parámetro	Tipo de Información que proporciona	Espacio en donde se calcula	Representación de la Serie de Tiempo	Transformación de la Información	Indican Predictibilidad: 1. No Indican Predictibilidad: 0
Exponente de Lyapunov	Teoría de Sistemas Dinámicos No Lineales	Topológico	Global	Horizonte de Predicción	Espacio Fase	Atractor	Multiplicación de la Información	1
Dimensión de Correlación	Teoría de Sistemas Dinámicos No Lineales	Topológico	Local	Correlación espacial local	Espacio Fase	Atractor	Multiplicación de la Información	1
Dimensión de Capacidad	Teoría de Sistemas Dinámicos No Lineales	Espacial	Local	Grado de autosimilitud del sistema	Espacio Fase	Atractor	Multiplicación de la Información	1
Dimensión Fractal	Teoría de Sistemas Dinámicos No Lineales	Espacial	Global	Dimensión promedio en una vecindad de tamaño epsilon sobre el atractor	Espacio Fase	Atractor	Multiplicación de la Información	1
Dimensión Embebida	Teoría de Sistemas Dinámicos No Lineales	Topológico	Global	Estimado del número grados de libertad del sistema	Espacio Fase	Atractor	Multiplicación de la Información	1
%Entropía Espacio Temporal	Teoría de Sistemas Dinámicos No Lineales	Espacial/Temporal	Global/Local	Grado de no correlación espacial	Espacio Euclideo	Mapa de Recurrencia	Multiplicación de la Información	1
% Recurrencia	Teoría de Sistemas Dinámicos No Lineales	Espacial/Temporal	Global/Local	Periodicidad y estructuras definidas	Espacio Euclideo	Mapa de Recurrencia	Multiplicación de la Información	1
% Determinismo	Teoría de Sistemas Dinámicos No Lineales	Espacial/Temporal	Global/Local	Determinismo en base a patrón de recurrencia	Espacio Euclideo	Mapa de Recurrencia	Multiplicación de la Información	1
Entropía de Shannon	Teoría de Información	Probabilístico	Global/Local	Cantidad de Información que puede ser extraída al medir el estado del sistema	Espacio Euclideo	Mapa de Recurrencia	Multiplicación de la Información	1
Información Mutua Promedio	Teoría de Información	Probabilístico	Global	Información mutua contenida en una variable para dos instantes de tiempo diferentes	Espacio Euclideo	Mapa de Recurrencia	Multiplicación de la Información	1
Complejidad LZ Relativa	Teoría de Computación	Computacional	Global/Local	Estructuras y jerarquía de cadenas de datos	Espacio Discretizado	Serie de tiempo Codificada Simbólicamente	Reducción de la Información	1
Reglas de Producción	Teoría de Computación	Computacional	Global/Local	Número de reglas gramaticales para modelar la serie (complejidad computacional)	Espacio Discretizado	Serie de tiempo Codificada Simbólicamente	Reducción de la Información	1

Figura 7-19: Tabla de las características de los parámetros de las series de tiempo

## 7.4 Datos Experimentales

Para calcular los diferentes parámetros el conjunto de series de tiempo fue *estandarizado* en el número de datos para cada una de ellas, se consideró a los primeros 1000 datos, ya que algunos de los parámetros pueden tener la característica de ser *extensivos*, es decir, ser dependientes de la cantidad de datos de las series de tiempo. Las tablas de datos generadas corresponden al cálculo de las siguientes clases de parámetros:

- Parámetros estadísticos que proporcionan información sobre la distribución estadística de los datos de cada serie de tiempo, este tipo de información además es útil en la construcción de modelos de predicción de series de tiempo [11, 179], los parámetros son: Rango de Datos, Resolución, Desviación Promedio, Desviación Estandar, Skewness, Kurtosis, Moda, Auto correlación, Tiempo de Correlación, Valor Promedio, Valor Medio, Valor Mínimo, Valor Máximo, Cuartil Mínimo y Cuartil Máximo.
- El parámetro de estadística moderna exponente de Hurst, descrito en el capítulo 4, que permite cuantificar la persistencia y memoria en el comportamiento de una serie.
- El parámetro de frecuencia dominante que permite determinar si hay una frecuencia característica de la serie de tiempo.
- El parámetro de porcentaje de varianza en las componentes principales, obtenido del análisis de componentes principales (PCA) para las componentes de las series de tiempo. Estas componentes son el producto de descomponer las series con el análisis espectral singular (SSA). El porcentaje de varianza se calculó para cada una de las primeras 5 componentes de las series de tiempo. Este parámetro se interpreta como la cantidad de información sobre la serie de tiempo que cada componente posee, producto de la descomposición de la señal. Más adelante, en el capítulo 10, se presenta un análisis sobre la relación entre predictibilidad de una serie y la información de sus componentes, como parte del estudio correspondiente a la relación entre parámetros de tipo estructural y la predictibilidad de las series de tiempo.
- El conjunto de parámetros no lineales basados en la teoría de sistemas dinámicos, los cuales son indispensables en la caracterización de la dinámica de las series de tiempo,

estos parámetros son: Exponente de Lyapunov, Dimensión de Correlación, Dimensión de Capacidad, Dimensión Fractal, Dimensión Embebida, Entropía Espacio Temporal, Recurrencia y Determinismo.

- El conjunto de parámetros basados en la teoría de información: Entropía de Información (Shannon) e Información Mutua Promedio.
- Finalmente, el conjunto de parámetros basados en la teoría de la computación: Complejidad Relativa LZ (Lempel-Ziv) y Número de Reglas de Producción.

## 7.5 Discusión

En este capítulo se ha presentado el *conjunto experimental* de series de tiempo, así como una breve descripción del amplio conjunto de datos experimentales generados con las diferentes técnicas de análisis de series de tiempo (presentadas en el capítulo 4), estos datos son utilizados posteriormente en el desarrollo de métricas de predictibilidad de series de tiempo. El conjunto experimental de series de tiempo cubre un *espectro* de comportamientos dinámicos diversos, además de ser utilizadas ampliamente en la literatura para el estudio de la predicción y el modelado en series de tiempo. En el siguiente capítulo, este conjunto de series será utilizado en la evaluación de diferentes técnicas de predicción y modelado.

## Capítulo 8

# Evaluación de Técnicas de Predicción y de Modelado de Series de Tiempo

### 8.1 Introducción

En este capítulo se presentan los resultados de un estudio sobre la capacidad de un conjunto de técnicas de predicción y un conjunto de técnicas de modelado, de predecir y modelar respectivamente a las series de tiempo. Las técnicas seleccionadas descritas en el capítulo 5, pertenecen a diversos orígenes: Estadística, Inteligencia Artificial y Sistemas Dinámicos No Lineales. Los resultados de la evaluación de estas técnicas serán de utilidad en la construcción de un coeficiente de capacidad de predicción (o modelado) que se presenta posteriormente en el capítulo 9. Para desarrollar este estudio comparativo, se diseñó una metodología experimental que permite estandarizar la evaluación de las técnicas y el análisis de los resultados obtenidos.

En la sección 8.2 se presenta el método experimental utilizado para la evaluación de las técnicas de predicción y de modelado, la sección 8.3 presenta los resultados de dicha evaluación, finalmente la sección 8.4 presenta la discusión de este capítulo.

## 8.2 Método Experimental

El desarrollo de los experimentos para evaluar el éxito o fracaso en la predicción o modelado de una serie de tiempo por parte de una técnica dada, está sujeto a una serie de restricciones metodológicas [10], que son justificadas a continuación:

- Se asume que la serie de tiempo posee un comportamiento estacionario en el conjunto de datos utilizados para la predicción (o modelado), lo anterior implica que es posible tomar cualquier subconjunto de datos para evaluar a la técnica en cuestión.
- El número máximo de experimentos por cada serie y técnica evaluada está acotado. Lo anterior implica que no se realiza un *análisis fino paramétrico* de cada técnica, en particular para las técnicas de red neuronal las cuales poseen varios parámetros sujetos a optimización. La justificación que hay detrás de no realizar un conjunto de experimentos más detallado son los teoremas de *No Free Lunch* para búsqueda y para optimización, los cuáles establecen que todo algoritmo para búsqueda o para optimización tiene la misma eficiencia cuando se promedia sobre todas las posibles optimizaciones o búsquedas realizadas. Es decir, si el algoritmo *A* es mejor que el algoritmo *B* en una clase de problemas, entonces existen otra clase de problemas en los que el algoritmo *B* es mejor que el algoritmo *A* (en el capítulo 13 de apéndices se comenta con mayor extensión sobre dichos teoremas). En resumen, *no existe un algoritmo para búsqueda o para optimización de soluciones que sea universal* [55, 56].
- La identificación de una serie predicha de forma exitosa depende de dos criterios: el comparar la serie original y la predicha, y el error obtenido. El éxito o fracaso de una predicción es una apreciación sujeta al criterio del *experto* u *observador* pero restringido por la siguiente regla: si la serie predicha sigue la tendencia (*trend*) de la serie original, así como su amplitud, aún cuando no exista una mejor predicción a una menor escala (mayor detalle), se considera que la predicción es buena y se anota como éxito.

Para realizar los experimentos de evaluación de la capacidad de predicción y la capacidad de modelado de las diferentes técnicas, se seleccionaron dos conjuntos de técnicas para predicción y modelado respectivamente bajo el criterio de tomar en cuenta modelos de diversos orígenes

(Estadística, Inteligencia Artificial y Sistemas Dinámicos No Lineales). Las técnicas ya fueron descritas en el capítulo 5. Las Figuras 8-1 y 8-2, muestran respectivamente las tablas para ambos conjuntos de técnicas así como un resumen de sus características.

La metodología experimental consistió en la evaluación de cada técnica con el conjunto de 29 series de tiempo previamente seleccionadas (ver capítulo 7), el número de experimentos de cada técnica para cada serie de tiempo fue variable, de 1 hasta 10 experimentos como máximo dependiendo de las características de cada técnica, por ejemplo en el caso de la predicción o el modelado con técnicas de redes neuronales, se exploró el posible espacio de soluciones en forma limitada dado que explorarlo en forma exhaustiva requiere de una cantidad de experimentos extensa (el número de posibilidades es combinatorial en base al número de parámetros que pueden ser modificados). A pesar de las limitaciones impuestas en el número de experimentos para intentar predecir o modelar una serie, el total de experimentos realizados fue del orden de 2000. Cada experimento consistió en separar a la serie de tiempo elegida, en un conjunto de datos de entrenamiento (de 1 a  $N$  datos), donde el número de datos de entrenamiento  $N$  fue en la mayoría de los experimentos de 900 o 1900 datos dependiendo de la serie de tiempo; y un conjunto de 50 datos de prueba (de  $N + 1$  a  $N + 50$ ), usados posteriormente para evaluar los resultados de predicción o modelado de la técnica seleccionada.

La comparación entre los datos originales de prueba y los predichos se hizo aplicando los siguientes criterios:

- Graficando la serie original y la predicha, para evaluar si la serie predicha reproduce la morfología de la serie original (predicción o modelado exitoso).
- Si el primer criterio es validado, se evalúan el Error Medio Raíz Cuadrada ( $RMSE$ ), el Error Bias ( $BE$ ) y el Error Promedio Local ( $MAE$ ).

En la literatura hay reportadas un gran número de fórmulas para cuantificar el error en la predicción o modelado de series de tiempo, se reportan desde las mas tradicionales como: el Error Medio Raíz Cuadrada ( $RMSE$ ), el Error Bias ( $BE$ ), hasta medidas de error especializadas definidas en función del tipo de técnica utilizada para predecir o modelar series de tiempo, como puede ser el caso de una arquitectura particular de red neuronal. A continuación se presentan las definiciones de los diferentes errores utilizados en este trabajo:

Técnica de Predicción	Descripción	Fundamentación Teórica
FIRNet	Red Neuronal con Filtros FIR	Inteligencia Artificial (Aprendizaje de Máquina)
PNNGauss	Red Neuronal Probabilística con función de aprendizaje gaussiana	Inteligencia Artificial (Aprendizaje de Máquina)
PNNReciprocal	Red Neuronal Probabilística con función de aprendizaje recíproca	Inteligencia Artificial (Aprendizaje de Máquina)
MLFFN	Red Neuronal Multicapa hacia delante (Perceptron Multicapa caso básico sin capa oculta)	Inteligencia Artificial (Aprendizaje de Máquina)
MySVM Multipunto (kernel dot)	Máquina de Vectores de Soporte (Kernel de Funciones: de tipo producto punto)	Inteligencia Artificial (Aprendizaje de Máquina)
MySVM Multipunto (kernel radial)	Máquina de Vectores de Soporte (Kernel de Funciones: de base radial)	Inteligencia Artificial (Aprendizaje de Máquina)
Polynomp	Modelo de funciones polinomiales en el espacio fase	Teoría de Sistemas Dinámicos No Lineales
Nstep	Modelo de aproximación por funciones locales en el espacio fase	Teoría de Sistemas Dinámicos No Lineales
K-Nearest-Neighbours	Modelo de aproximación por promedio de vecinos cercanos en el espacio fase	Teoría de Sistemas Dinámicos No Lineales y Aprendizaje de Máquina
ARIMA	Modelo de Auto Regresión y Promedio Móvil Integrado	Estadística

Figura 8-1: Tabla de técnicas de predicción

Técnica de Modelado	Descripción	Fundamentación Teórica
FFNBP	Red Neuronal hacia adelante con retropropagación (perceptron)	Inteligencia Artificial (Aprendizaje de Máquina)
RBFNBP	Red Neuronal con Funciones de Base Radial con retropropagación	Inteligencia Artificial (Aprendizaje de Máquina)
ANFIS	Red Neuronal y Lógica Difusa (híbrido)	Inteligencia Artificial (Aprendizaje de Máquina)
RNAP	Red Neuronal Artificial Polinomial y Algoritmo Genético (híbrido)	Inteligencia Artificial (Aprendizaje de Máquina)
Predict	Modelo de aproximación por funciones locales en el espacio fase: función constante	Teoria de Sistemas Dinamicos No Lineales
Polynom	Modelo de funciones polinomiales en el espacio fase	Teoria de Sistemas Dinamicos No Lineales
rbf	Modelo de funciones de base radial en el espacio fase	Teoria de Sistemas Dinamicos No Lineales
Sugimay	Modelo de aproximación por promedio pesado de vecinos cercanos en el espacio fase	Teoria de Sistemas Dinamicos No Lineales
ARMA	Modelo de Auto Regresion y Promedio Movil	Estadistica

Figura 8-2: Tabla de técnicas de modelado

- *RMSE*: El Error Medio Raíz Cuadrada, representa la medida típica de error de predicción, su expresión matemática es la siguiente:

$$RMSE = \left[ \frac{1}{N} \sum_{i=1}^N (x_i^p - x_i^o)^2 \right]^{\frac{1}{2}} \quad (8.1)$$

donde  $N$  es el número de datos predichos,  $x_i^p$  corresponde al  $i$  –ésimo dato predicho y  $x_i^o$  corresponde al  $i$  –ésimo dato original. Un valor cercano a cero indica una buena predicción, el término de la diferencia al cuadrado le da mayor peso a las grandes discrepancias entre valores originales y los predichos. Es una medida de error global pues realiza un promedio sobre el conjunto de datos predichos [11, 12].

- *BE*: El Error Bias, es una medida de la tendencia a subpredecir o sobrepredecir un conjunto de datos, su expresión matemática es la siguiente:

$$BE = \frac{1}{N} \sum_{i=1}^N (x_i^p - x_i^o) \quad (8.2)$$

donde  $N$  es el número de datos predichos,  $x_i^p$  corresponde al  $i$  –ésimo dato predicho y  $x_i^o$  corresponde al  $i$  –ésimo dato original. Es muy útil para identificar cuando un modelo (por ejemplo, aquel generado a partir del entrenamiento de una red neuronal) se ha sub o sobreentrenado, ambos casos no son adecuados para lograr una predicción o modelado óptimo. Un Error Bias positivo indica una tendencia a sobremodelar y un Error Bias negativo a submodelar a la serie de tiempo. Como el anterior, es una medida de error global pues realiza un promedio sobre el conjunto de datos predichos [11, 12].

- *LRSE*: El Error Local, calcula el error para cada punto predicho de una serie de tiempo  $ST$ , su expresión es:

$$LRSE = |x_i^p - x_i^o| \quad (8.3)$$

- *MAE*: El Error Local Promedio, promedia el error local de cada punto predicho para una serie de tiempo  $ST$ , su expresión matemática es:

$$MAE = \frac{1}{N} \sum_{i=1}^N |x_i^p - x_i^o| \quad (8.4)$$

donde  $N$  es el número de datos predichos,  $x_i^p$  corresponde al  $i$  – *ésimo* dato predicho y  $x_i^o$  corresponde al  $i$  – *ésimo* dato original [11, 12].

- $\langle\langle RMSE_j \rangle_k\rangle$ : Este es un promedio de los promedios del error RMSE para el conjunto de técnicas (ya sea de predicción o de modelado), donde el error  $RMSE_j$  (ver ecuación 8.1) corresponde a la  $j$  – *ésima* serie de tiempo que fue predicha (o fue modelada) con éxito por la  $k$  – *ésima* técnica. Para su cálculo se obtiene primero el promedio del error RMSE para todos los  $L$  casos exitosos de una técnica, a este promedio lo identificamos como el promedio RMSE de la técnica  $k$  – *ésima*:

$$\langle RMSE_j \rangle_k = \frac{1}{L} \sum_{j=1}^L RMSE_j \quad (8.5)$$

el segundo paso consiste en el cálculo del promedio para el conjunto de promedios RMSE obtenidos a partir de las  $M$  técnicas evaluadas:

$$\langle\langle RMSE_j \rangle_k\rangle = \frac{1}{M} \sum_{k=1}^M \langle RMSE_j \rangle_k \quad (8.6)$$

- $\langle\langle BE_j \rangle_k\rangle$ : Este es el promedio de los promedios del error BE para el conjunto de técnicas, y la construcción de su definición tiene la misma estructura del promedio de promedios del error RMSE, así que sólo se presenta la expresión final:

$$\langle\langle BE_j \rangle_k\rangle = \frac{1}{M} \sum_{k=1}^M \langle BE_j \rangle_k \quad (8.7)$$

- $\langle\langle MAE_j \rangle_k\rangle$ : De la misma forma se define el promedio de los promedios del error MAE para el conjunto de técnicas, su expresión es:

$$\langle\langle MAE_j \rangle_k\rangle = \frac{1}{M} \sum_{k=1}^M \langle MAE \rangle_k \quad (8.8)$$

## 8.3 Resultados Experimentales

A continuación se presentan los resultados de la evaluación de las técnicas de predicción y de modelado.

### 8.3.1 Técnicas de Predicción

Las siguientes figuras presentan gráficas que son una muestra de los diferentes experimentos de predicción realizados con éxito. La Figura 8-3, ejemplifica la predicción ideal es decir, con un mínimo error y un excelente seguimiento de la dinámica de la serie. La Figura 8-4, corresponde al caso en cual las características de la serie (en este caso la dinámica caótica) impiden una predicción mas allá de un número pequeño de valores futuros. El siguiente ejemplo, que corresponde a la Figura 8-5, presenta el caso de la predicción en la cual la tendencia y frecuencia de la dinámica se reproducen, sin embargo, las variaciones de la amplitud no logran ser predichas. Por último, la Figura 8-6 muestra el caso en el cuál se sigue la tendencia de la serie aunque presentando un error no despreciable.

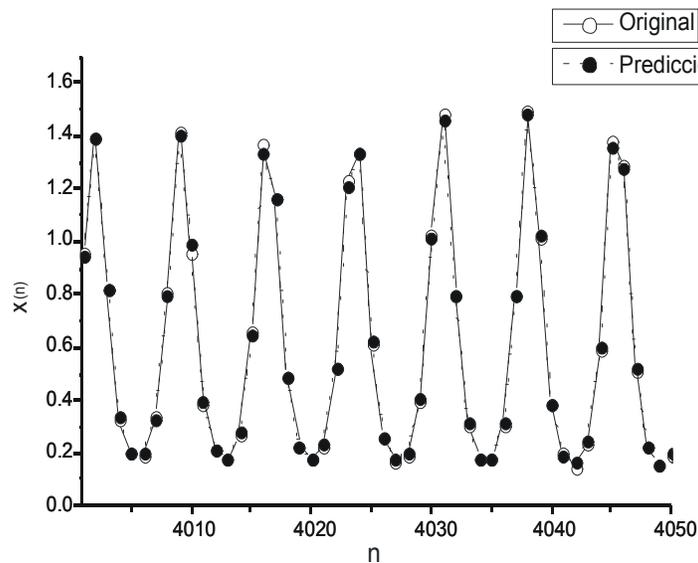


Figura 8-3: Serie Laser predicha con técnica Nstep

La Figura 8-7, presenta la tabla *Matriz de Predicción* de las series de tiempo con las diferentes técnicas evaluadas. En esta matriz se codifica con los valores de: 1 = *Predicción*, 0 =

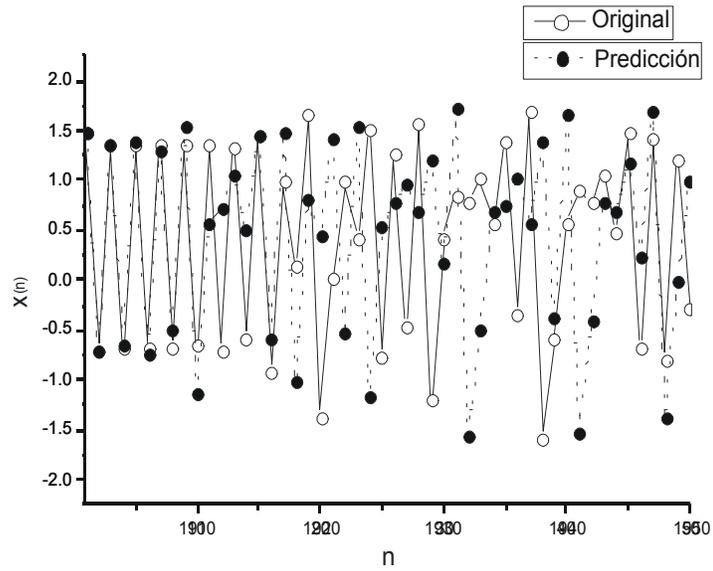


Figura 8-4: Serie Henon predicha con técnica K Vecinos Cercanos

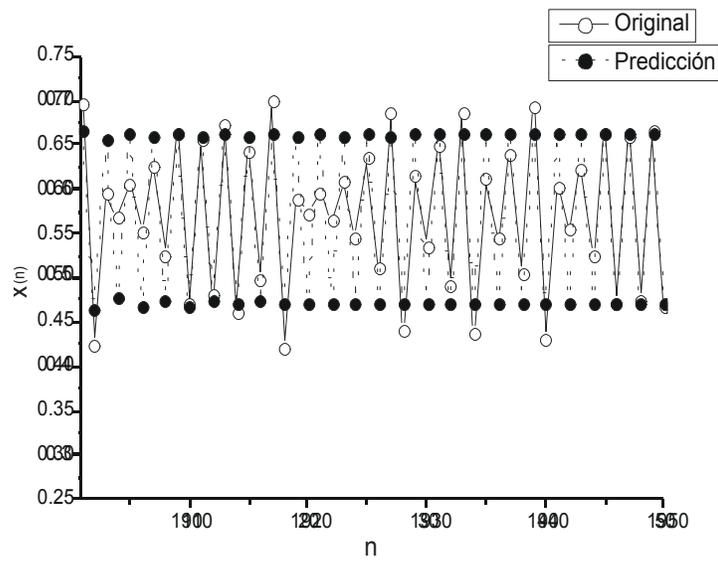


Figura 8-5: Serie Tent predicha con técnica Polynomp

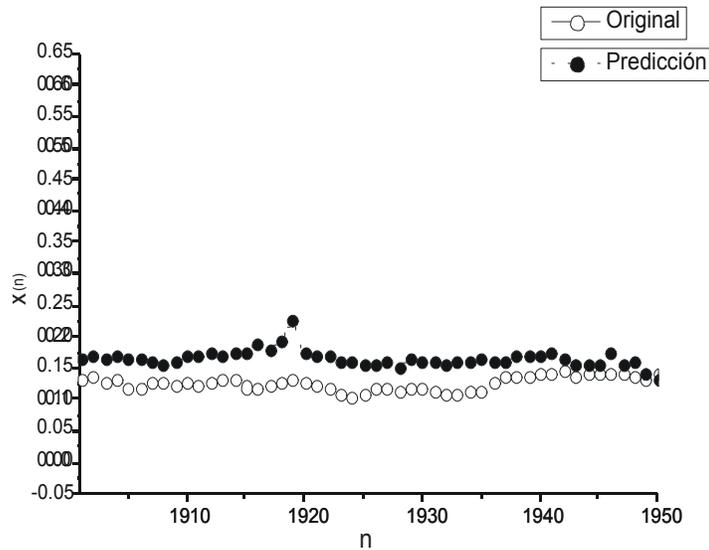


Figura 8-6: Serie Dow Jones predicha con técnica PNNReciprocal

*NoPredicción*, los resultados de la totalidad de los experimentos realizados para evaluar al conjunto de técnicas. Se observa que las técnicas con más resultados exitosos de predicción corresponden a K-Vecinos-Cercanos (con 17 éxitos) y Nstep (con 16 éxitos), basadas en Teoría de Sistemas Dinámicos No Lineales; le siguen en número de éxitos las técnicas MySVM Multipunto con el kernel radial (con 13 éxitos) y PNNGauss (con 12 éxitos) cuya base proviene de Inteligencia Artificial. El resto de las técnicas (FIRNet, PNNReciprocal, MLFFN, Polynomp y ARIMA) que corresponden a las tres familias de modelos presentados en este trabajo, no presentan entre sí una gran diferencia en la capacidad de predicción siendo su media de 9 éxitos, por último la técnica MySVM Multipunto con el kernel dot presenta la más baja capacidad de predicción (solo 3 éxitos), lo que indica que dicho kernel no es el óptimo para el proceso de aprendizaje de la tarea de predicción de series de tiempo.

La evaluación del error de predicción genero tres *Matrices de Error de Predicción*, las cuales no solo permiten el análisis de la propagación del error de la predicción para cada una de las técnicas, sino también utilizar esta información como se muestra en el siguiente capítulo, para construir una métrica para cuantificar la capacidad de predicción de las técnicas.

Para analizar de forma global las matrices de error, las Figuras 8-8, 8-9 y 8-10, presentan gráficas de superficie de los errores de predicción (RMSE, BE y MAE) para el conjunto de téc-

Serie de Tiempo	Técnica de Predicción										Total de Técnicas que Predicen a cada Serie de Tiempo	
	Inteligencia Artificial					Teoría de Sistemas Dinámicos No Lineales						Estadística
	FIRNet	PNNGauss	PNNReciprocal	MLFFN	MySVM Multipunto (kernel dot)	MySVM Multipunto (kernel radial)	Polynomp	Nstep	K-Nearest-Neighbours	ARIMA		
Sine	1	1	1	1	1	1	1	1	1	1	1	10
Vanderpol	1	1	1	1	1	1	1	1	1	1	1	8
Operiodic2	0	1	0	1	0	1	0	0	1	0	0	4
Operiodic3	1	1	1	1	1	1	1	1	1	1	0	9
Mackey-Glass	1	0	0	0	0	1	1	1	1	1	0	5
Logistic	0	0	0	0	0	1	1	1	1	1	0	4
Lorenz	1	0	1	1	0	1	1	1	0	0	0	6
Rossler	0	1	1	1	0	1	0	1	1	0	0	6
Ikeda	0	0	0	0	0	0	0	0	1	0	0	1
Henon	1	1	0	0	0	0	1	1	1	0	0	5
Cantor	0	0	0	0	0	0	0	0	0	0	0	0
Tent	0	1	1	1	0	1	1	1	1	1	1	8
A1	1	1	1	1	0	1	1	1	1	0	0	8
D1	0	0	0	0	0	0	0	0	0	0	0	0
Laser	1	1	1	0	0	1	1	1	1	1	1	8
Dow Jones	0	1	1	0	0	0	0	1	1	1	1	5
Kobe	0	0	0	0	0	1	1	0	1	0	0	3
EEG	0	0	0	0	0	0	0	0	0	1	1	2
ASCII TXT	0	0	0	0	0	0	0	0	0	0	0	0
El niño	0	1	0	1	0	0	0	0	1	0	0	3
HIV DNA	0	0	0	0	0	0	0	0	0	0	0	0
Human DNA	1	1	0	1	0	0	0	1	1	1	1	6
Lovaina	0	0	0	0	0	1	0	1	1	0	0	3
Plasma	0	0	0	0	0	0	1	1	0	0	0	2
Primos	0	0	0	0	0	0	0	0	0	0	0	0
SP500	0	0	0	0	0	0	0	0	0	0	0	0
Star	0	0	0	0	0	0	0	0	0	1	1	1
Brownian Motion	0	0	0	0	0	0	0	0	0	0	0	0
White Noise	0	0	0	0	0	0	0	0	0	0	0	0
Total de Series Predichas por cada Técnica	9	12	9	10	3	13	11	16	17	7	7	

Figure 8-7: Matriz de predicción de series de tiempo

nicas evaluadas. Al construir estas gráficas se quería no solo que mostrarán el comportamiento del error de predicción sino también, que fuese posible una comparación en magnitud del error, ello no es posible sin una normalización previa a su evaluación de las diferentes series de tiempo, para que las magnitudes resultantes de los errores sean del mismo orden. Las series normalizadas para que la magnitud de sus datos estuviese en el rango  $[10^{-2}, 10^1]$  como el resto del conjunto de series fueron: A1, Laser, Dow Jones, Kobe, ASCIITXT, El niño, Human DNA y Primos. Es importante aclarar también, que todos los casos graficados que poseen un valor cero en los errores, corresponden a los fracasos en la predicción de las series correspondientes.

La gráfica de la Figura 8-8 sobre el error RMSE, muestra que salvo dos casos extremos, *el error RMSE es pequeño para la mayoría de las técnicas de predicción*, el promedio de este error para todas las técnicas de predicción es de  $\langle \langle RMSE_j \rangle_k \rangle = 0.1717$ . Los casos de error con valor máximo corresponden a las técnicas MLFFN y MySVM Multipunto con el kernel radial, estos valores máximos indican el castigo a discrepancias grandes entre los datos predichos y los originales para estas series de tiempo en particular. La serie de tiempo cuya predicción generó un mayor error con cada técnica que la predijo fue Henon. Los datos obtenidos sobre el error RMSE como una medida de tipo *global* (i. e. promedio sobre el conjunto de datos predichos para cada serie), nos indican, no solo que en general el RMSE presenta un valor pequeño (i. e. *hay buenas predicciones*) para las diferentes técnicas, sino también que este comportamiento es *independiente* del número de series predichas por cada técnica.

La gráfica de la Figura 8-9, muestra el comportamiento del error BE, nuevamente los casos extremos corresponden a las técnicas mencionadas en el análisis del error RMSE, se observa que *el ajuste de los modelos de predicción a las series de tiempo es cercano al óptimo*, ya que el BE es cercano a cero en la mayoría de los casos. En promedio el error BE para todas las técnicas de predicción es de  $\langle \langle BE_j \rangle_k \rangle = -0.0097$ , lo que indica que es cercano al óptimo aunque hubo una *ligera tendencia al subaprendizaje* por parte de la mayoría de las técnicas. Por último, el comportamiento del error es *independiente* del número de series predichas por cada técnica.

La gráfica de la Figura 8-10, muestra el comportamiento del error MAE, con excepción de los casos extremos mencionados anteriormente, el error MAE que cuantifica el error local promedio de una predicción, se observa es *pequeño* para el conjunto de técnicas de predicción, en promedio el error MAE es de  $\langle \langle MAE_j \rangle_k \rangle = 0.1317$ . Este valor promedio es indicador de

**RMSE Técnicas de Predicción**

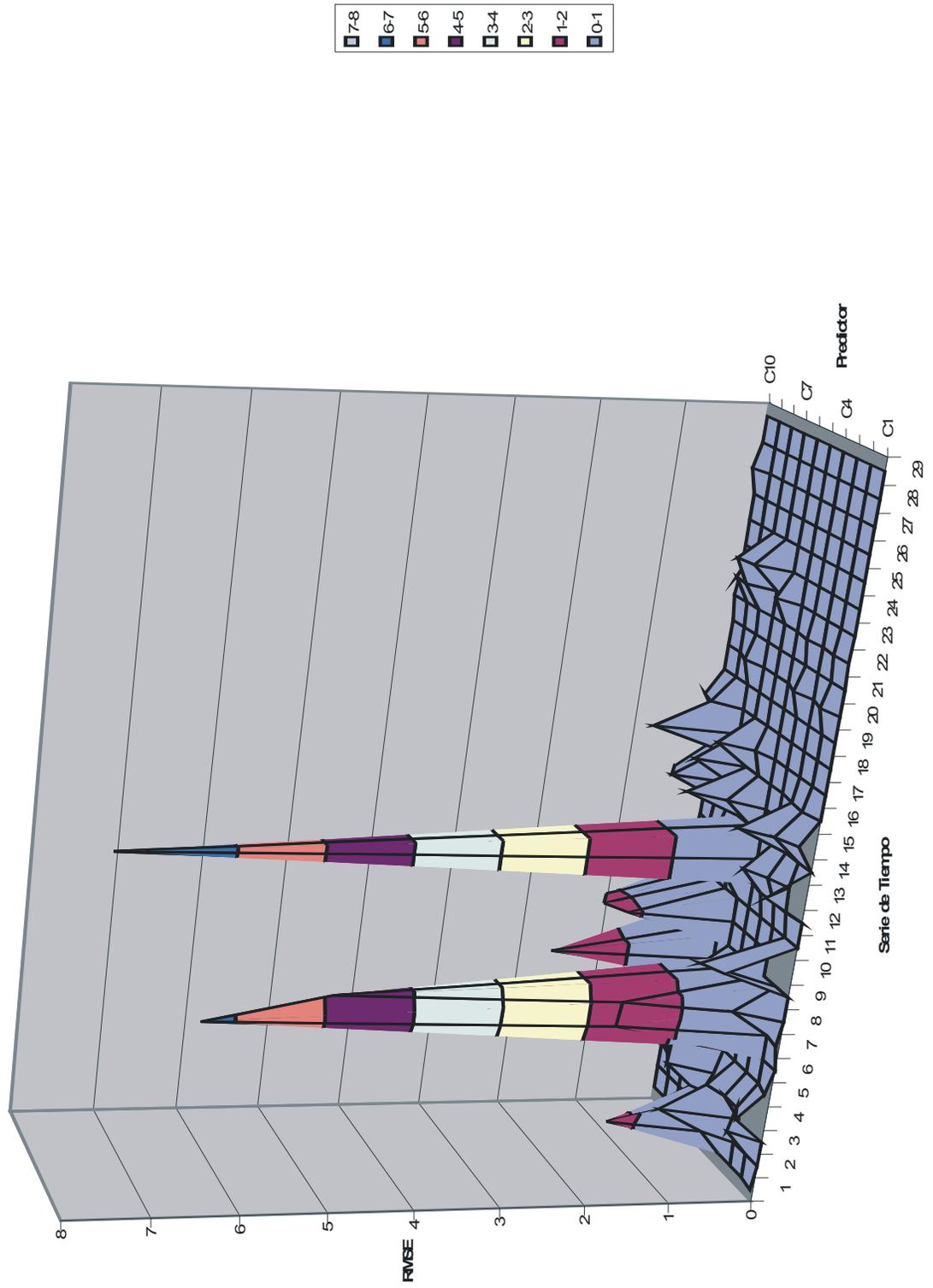


Figura 8-8: Gráfica de superficie del error RMSE para la predicción de series de tiempo

### EE Técnicas de Predicción

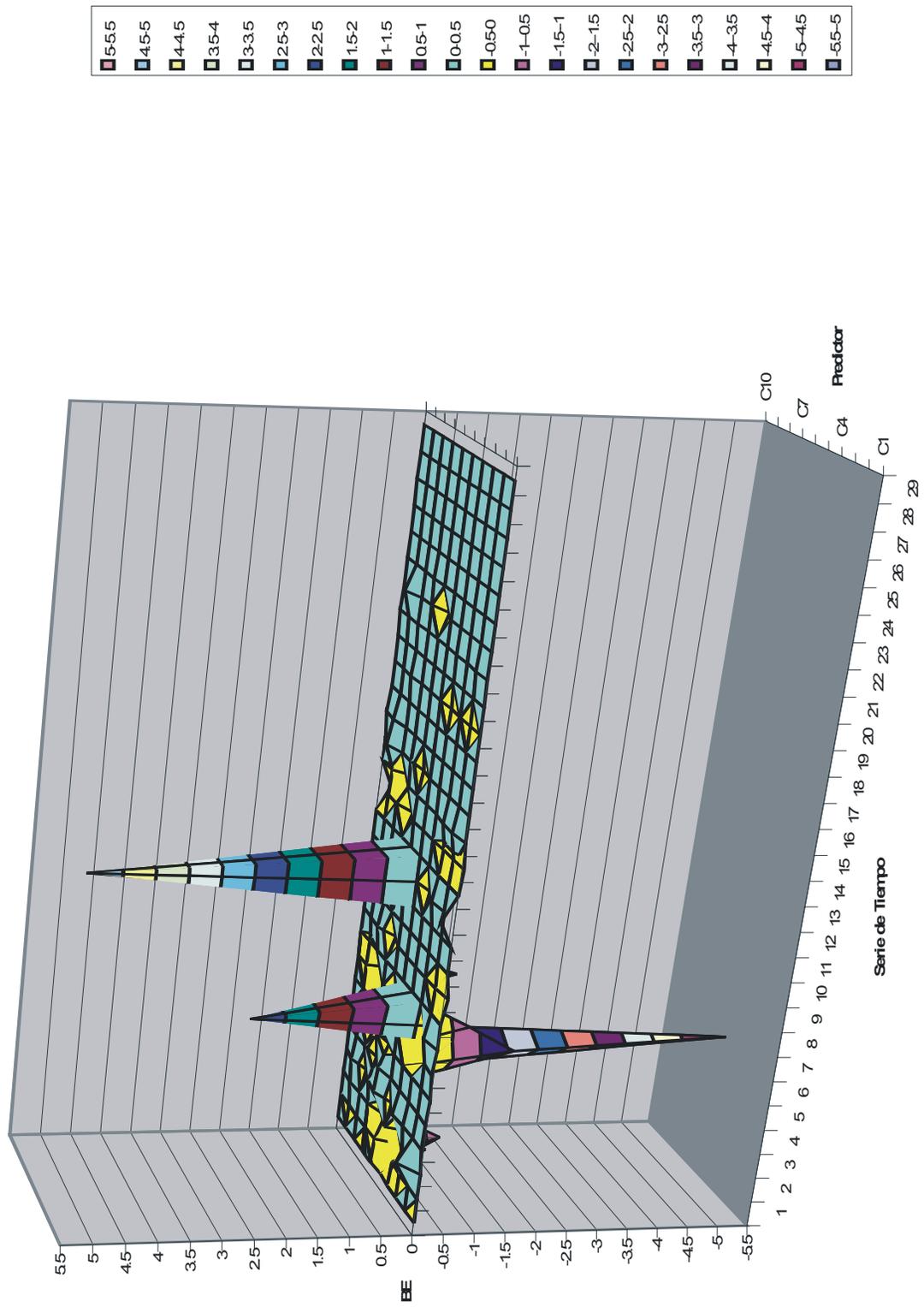


Figura 8-9: Gráfica de superficie del error BE para la predicción de series de tiempo

que el error local generado por las predicciones *no crece* en forma apreciable para el intervalo de predicción de  $N = 50$  unidades. Igual que para los otros dos errores se observa, que su comportamiento es *independiente* del número de series predichas por cada técnica.

### 8.3.2 Técnicas de Modelado

Las siguientes Figuras presentan gráficas que son una muestra de diferentes experimentos de modelado realizados con éxito. La Figura 8-11, muestra el ejemplo de una modelación con un ajuste a la serie original muy cercano al óptimo. La Figura 8-12, presenta un modelo en el cuál para los primeros puntos de la serie se logra reproducir la dinámica de la misma, pero debido a su dinámica caótica, posteriormente el modelo no logra ajustarse a la serie en forma óptima. La Figura 8-13, muestra un modelo que aunque no puede realizar un ajuste fino a la dinámica de la serie, es capaz de modelar la tendencia dominante de la serie de tiempo. Por último, la Figura 8-14 muestra el caso en el cual el modelo reproduce la tendencia de la serie pero falla en ajustarse a la amplitud de la misma.

La Figura 8-15, presenta la tabla *Matriz de Modelado* de las series de tiempo con las diferentes técnicas evaluadas. En esta matriz, como en la correspondiente a técnicas de predicción, se codifica con los valores de:  $1 = Modelado$ ,  $0 = NoModelado$ , los resultados de la totalidad de los experimentos realizados para evaluar al conjunto de técnicas. La primera característica sobre las técnicas de modelado es que el número de éxitos de varias de ellas, sobrepasa al número de éxitos del mejor predictor, lo que indica la *diferencia en la dificultad de resolver un problema de predicción y uno de modelado*. Por ejemplo, dos series que no fueron predichas con éxito son D1 y Brownian Motion, las cuales lograron ser modeladas por varias de las técnicas; también algunas series predichas por solo una técnica de predicción como son las series Ikeda y Star, lograron ser modeladas por un mayor número de técnicas.

Se observa que no hay una técnica de modelado que presente una *superioridad clara* con respecto a las demás para modelar a las series de tiempo, las dos técnicas con mayores éxitos son RNAP (23 éxitos) y Sugimay (22 éxitos), seguidas por las técnicas ANFIS (20 éxitos) y ARMA (19 éxitos). Las técnicas RNAP y ANFIS están construidas en base a algoritmos híbridos y su origen es la Inteligencia Artificial, la técnica Sugimay se basa en la Teoría de Sistemas Dinámicos No Lineales y la Técnica ARMA se basa en Estadística. También se observa que las

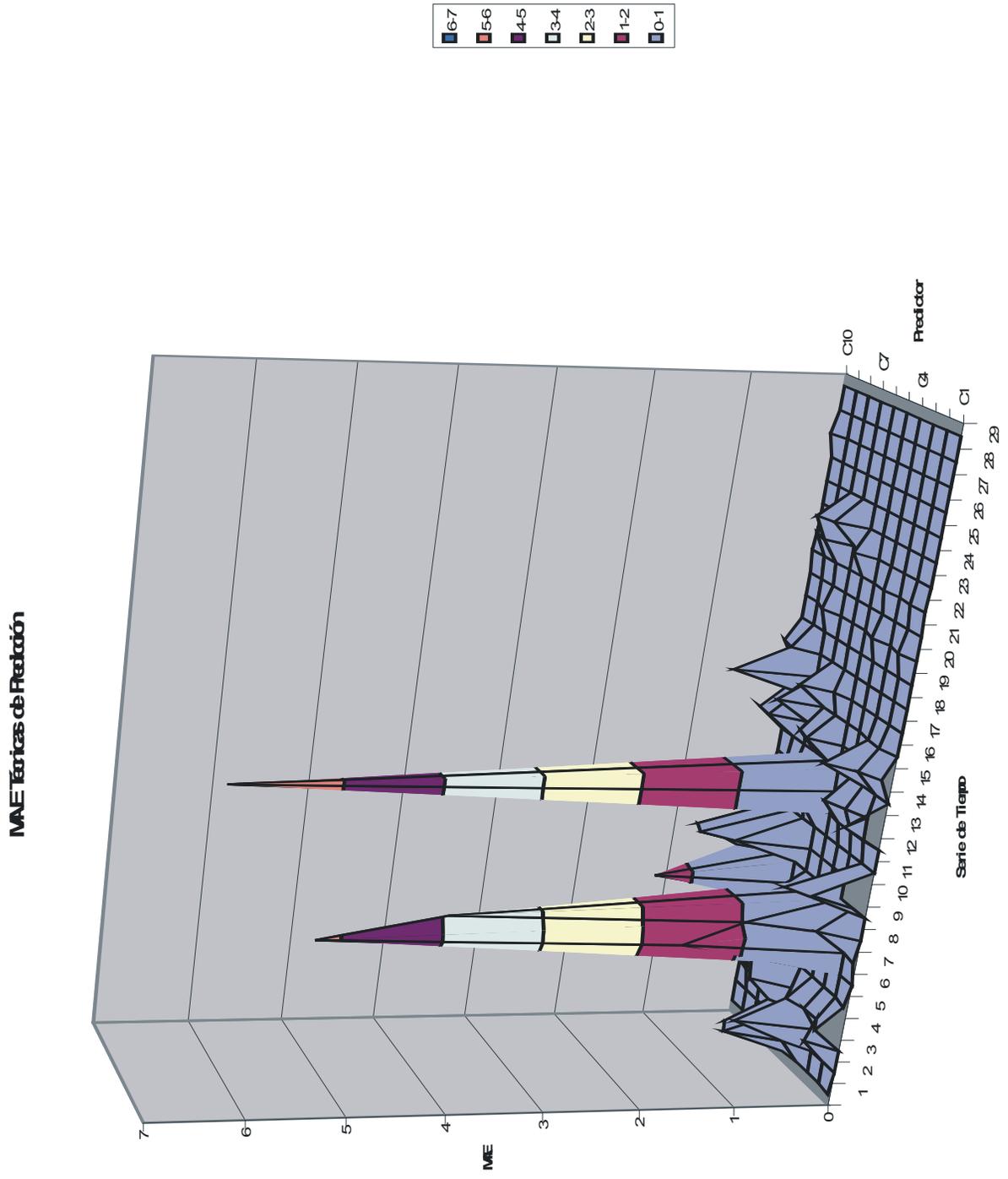


Figura 8-10: Gráfica de superficie del error MAE para la predicción de series de tiempo

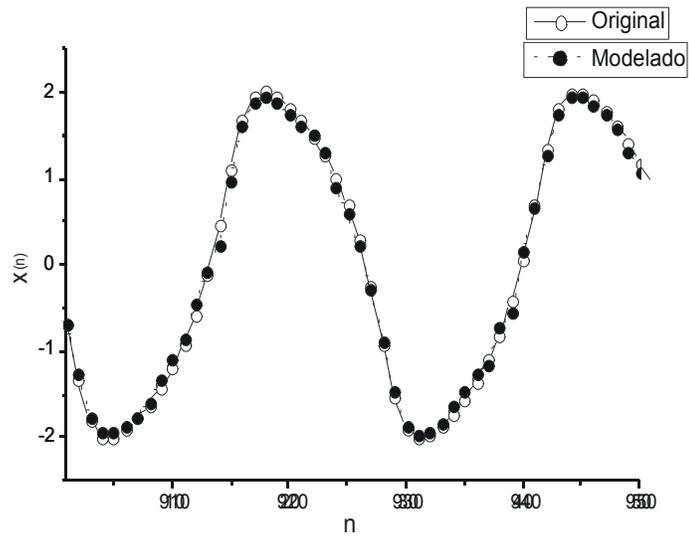


Figura 8-11: Serie Vanderpol modelada con técnica ANFIS

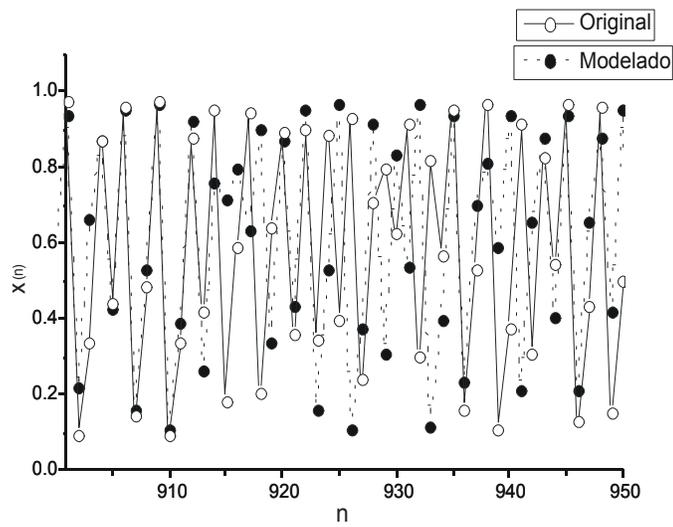


Figura 8-12: Serie Logistic modelada con técnica Polynom

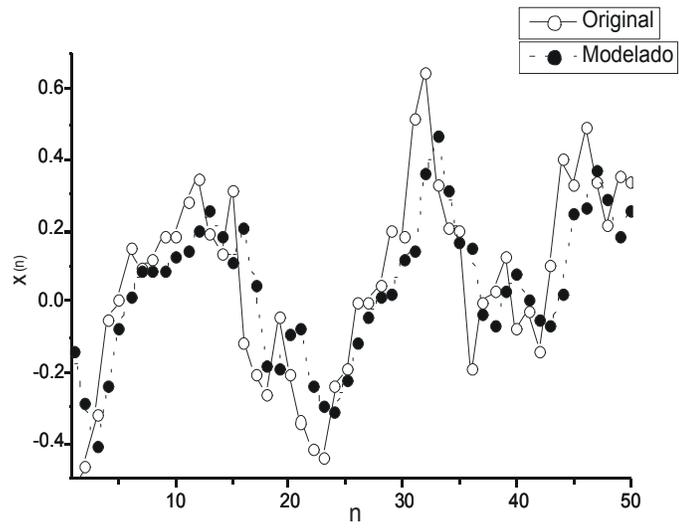


Figura 8-13: Serie EEG modelada con técnica RNAP

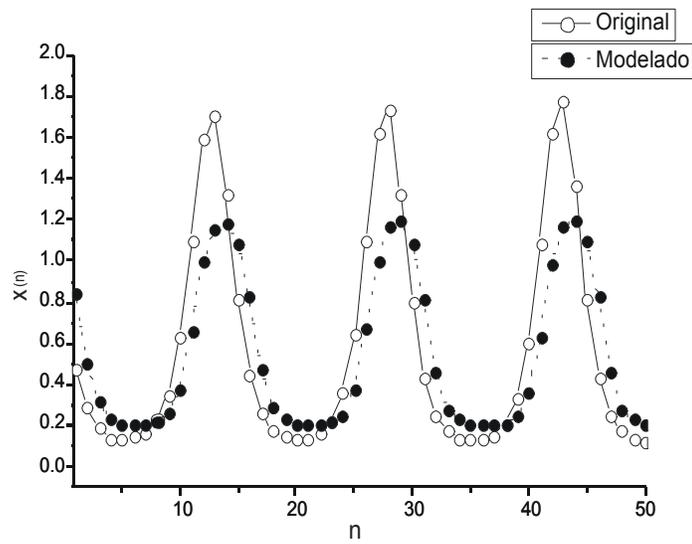


Figura 8-14: Serie A1 modelada con técnica FNNBP

técnicas cuyo origen es la Teoría de Sistemas Dinámicos No Lineales son las de menor capacidad de modelado de las series de tiempo: Predict, Polynom y rbf. En general, las técnicas basadas en *Inteligencia Artificial* poseen una capacidad de modelado superior, siendo las que aplican *métodos híbridos* las mejores de todas.

La evaluación del error de modelado genero tres *Matrices de Error de Modelado*, las cuales no solo permiten el análisis de la propagación del error del modelado para cada una de las técnicas, sino también utilizar esta información para construir una métrica para cuantificar la capacidad de modelado de las mismas.

Nuevamente se presentan las gráficas de superficie para las matrices de errores RMSE, BE y MAE correspondientes al modelado de las series de tiempo. La Figura 8-16, muestra que el comportamiento del error RMSE presenta en general valores pequeños (i.e. *hay buenos modelos*), siendo el valor promedio para el conjunto de técnicas de  $\langle\langle RMSE_j \rangle_k\rangle = 0.1743$ . La excepción es la técnica Sugimay, la cual presenta varios casos extremos de errores mayores a uno, la técnica a pesar de tener capacidad de modelar gran número de series de tiempo, presenta un rendimiento deficiente para el aprendizaje de la dinámica de las series de tiempo, lo que se refleja en el ajuste de los modelos a las series, esto corrobora lo observado en la tabla de la Figura 8-15, que muestra el *bajo rendimiento* de las técnicas derivadas de Teoría de Sistemas Dinámicos No Lineales para *modelar* series de tiempo. Si se excluye la técnica Sugimay para el cálculo del promedio del error RMSE, se obtiene  $\langle\langle RMSE_j \rangle_k\rangle = 0.0830$  que es un valor más representativo del comportamiento del error RMSE para las técnicas de modelado. En forma similar al caso de las técnicas de predicción, el comportamiento del error es *independiente* de la capacidad de cada técnica para modelar a las series de tiempo.

La Figura 8-17, muestra el comportamiento del error BE para el caso de modelado de series de tiempo, los casos extremos corresponden a la técnica Sugimay que muestra el subaprendizaje de la técnica, lo cual deteriora su capacidad de modelar a las series de tiempo de forma adecuada. Para el resto de las técnicas se observa que el error BE es cercano a cero, lo cuál indica un *ajuste óptimo* de los modelos a las series de tiempo, el promedio del error BE para todo el conjunto de técnicas es  $\langle\langle BE_j \rangle_k\rangle = -0.0567$ , que muestra que hubo una *ligera tendencia al subaprendizaje* por parte de la mayoría de los modelos, sin embargo la magnitud de este promedio es dominada por la técnica Sugimay, por lo cual no se considera un valor confiable, si se excluye a dicha

Serie de Tiempo	Técnica de modelado											Total de Técnicas que Modelan a cada Serie de Tiempo
	Inteligencia Artificial				Teoría de Sistemas Dinámicos No Lineales					Estadística		
	FFNBP	RFNBP	Híbridos		Predict	Polynom	rft	Sugimay	ARMA			
Sine	1	1	1	1	1	1	1	1	1	1	1	9
Vanderpol	1	1	1	1	1	1	1	1	1	1	1	9
Qperiodic2	1	1	1	1	0	0	1	1	1	1	1	7
Qperiodic3	1	1	1	1	0	0	0	1	1	1	1	6
Mackey-Glass	1	1	1	1	0	0	1	1	1	1	1	7
Logistic	1	1	1	1	0	1	0	1	1	0	0	6
Lorenz	1	1	1	1	1	0	0	1	1	1	1	7
Rossler	1	1	1	1	1	0	0	1	1	1	1	7
Ikeda	0	0	1	1	0	0	0	0	0	0	0	2
Henon	1	1	0	1	0	0	0	1	1	0	0	4
Cantor	0	0	0	0	0	0	0	0	0	0	0	0
Tent	0	0	1	1	0	1	1	1	1	1	1	6
A1	1	1	1	1	1	0	0	1	1	1	1	7
D1	1	1	1	1	0	0	0	1	1	0	0	5
Laser	0	0	1	1	1	1	1	1	1	1	1	7
Dow Jones	1	0	1	1	1	0	0	1	1	1	1	6
Kobe	0	0	1	1	0	0	0	1	1	1	1	4
EEG	1	0	0	1	0	0	0	0	1	1	1	4
ASCII TXT	0	0	0	0	0	0	0	0	0	0	0	0
El niño	0	0	1	1	0	0	0	1	1	1	1	4
HIV DNA	0	0	0	0	0	0	0	0	0	0	0	0
Human DNA	0	0	1	1	1	0	0	1	1	1	1	5
Lovina	1	1	1	1	0	0	0	1	1	1	1	6
Plasma	0	0	0	1	0	0	0	0	1	1	1	3
Primos	0	0	0	0	0	0	0	0	0	0	0	0
SP500	0	0	0	0	0	0	0	0	0	0	0	0
Star	1	1	1	1	0	0	0	1	1	1	1	6
Brownian Motion	1	1	1	1	1	0	0	1	1	1	1	7
White Noise	0	0	0	0	0	0	0	0	0	0	0	0
Total de Series Modeladas por cada Técnica	16	14	20	23	9	5	6	22	19			

Figure 8-15: Matriz de modelado de series de tiempo

### RMSE Técnicas de Modelado

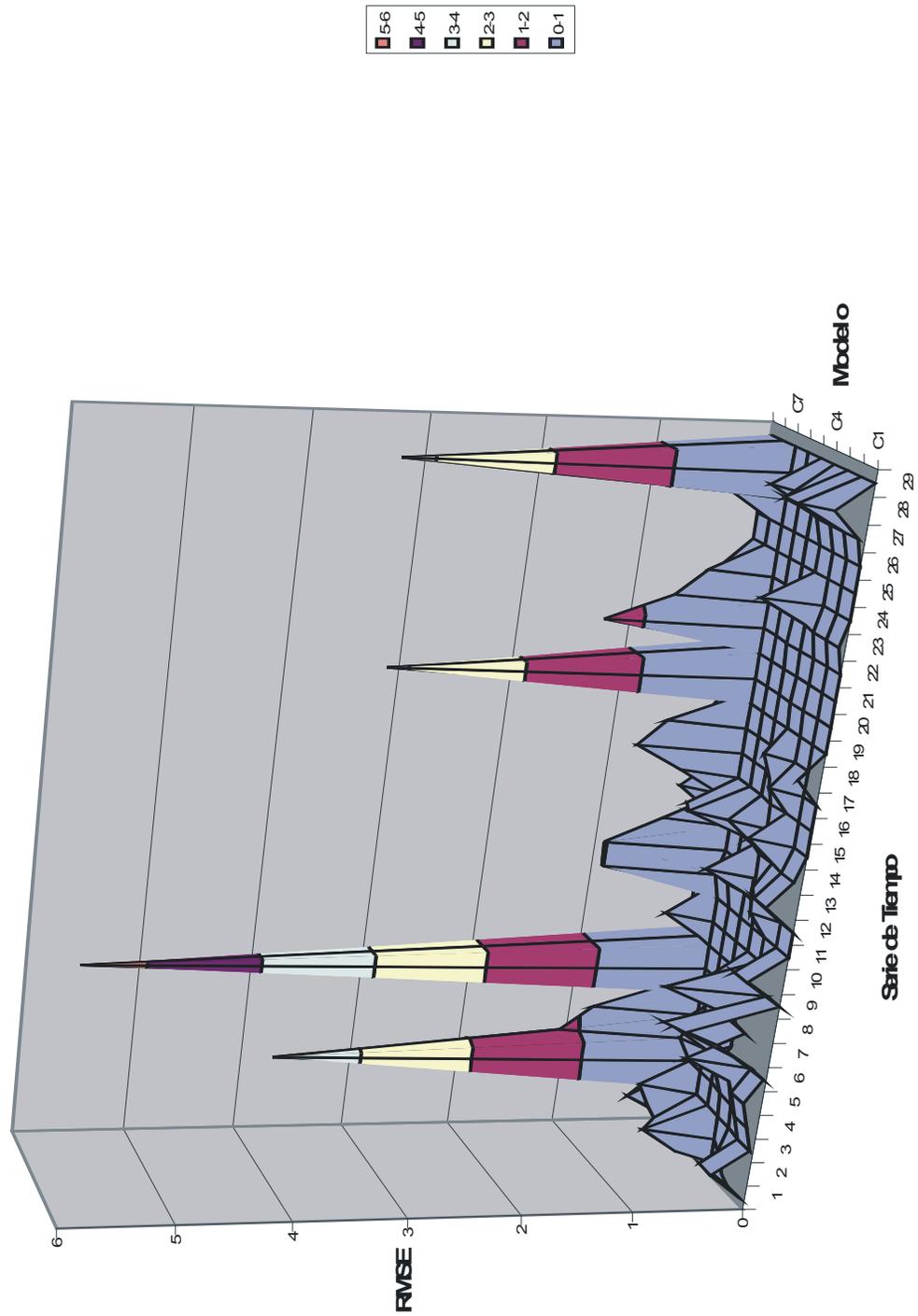


Figura 8-16: Gráfica de Superficie del error RMSE para el modelado de Series de tiempo

técnica se obtiene un valor promedio de  $\langle\langle BE_j \rangle_k\rangle = -0.0015$ . En forma similar al caso de las técnicas de predicción, el comportamiento del error es *independiente* de la capacidad de cada técnica para modelar a las series de tiempo.

La Figura 8-18, muestra el comportamiento del error MAE para el caso de modelado de series de tiempo, los casos extremos que corresponden a la técnica Sugimay, muestran que el error local crece en forma considerable para esta técnica de modelado. Nuevamente considerando al total de técnicas para calcular el promedio del error MAE, la contribución de la técnica Sugimay domina siendo su resultado  $\langle\langle MAE_j \rangle_k\rangle = 0.1497$ . En general, para el resto de las técnicas el error MAE es pequeño, excluyendo a la técnica Sugimay, se tiene un promedio  $\langle\langle MAE_j \rangle_k\rangle = 0.0664$ , lo cual indica que el error local de los modelos *no crece* en forma apreciable para un intervalo de modelado de  $N = 50$  unidades. Finalmente, como en el caso de las técnicas de predicción, el comportamiento del error se observa es *independiente* de la capacidad de cada técnica para modelar a las series de tiempo.

## 8.4 Discusión

En este capítulo se presentó una *metodología para la evaluación y comparación* de dos conjuntos de técnicas que representan respectivamente, diferentes modelos para la predicción y para el modelado de series de tiempo. Esta metodología genera un conjunto de *matrices de datos* que junto con los matrices correspondientes a la caracterización de las series de tiempo (ver capítulo 7), proporcionan información útil para el desarrollo de métricas de capacidad de predicción (o modelado) y para el estudio de la predictibilidad de las series de tiempo.

Para el caso de las técnicas de predicción (*predictores*), se encuentra que las mejores técnicas corresponden a aquellas cuyo origen es la *Teoría de Sistemas Dinámicas No Lineales* y que se basan en la construcción de *modelos locales en el espacio fase*. La magnitud del error de predicción RMSE es pequeño y cercano a cero, como una medida de carácter global, muestra que los *modelos de predicción son buenos*. Por su parte el error BE, muestra que los modelos se *ajustan casi en forma óptima* a la dinámica de las series, y en promedio hay una tendencia de las técnicas al *subaprendizaje*. Por último, el error MAE como medida del comportamiento del error local de predicción, muestra que este *no crece* en forma apreciable para el conjunto

### BE Técnicas de Modelado

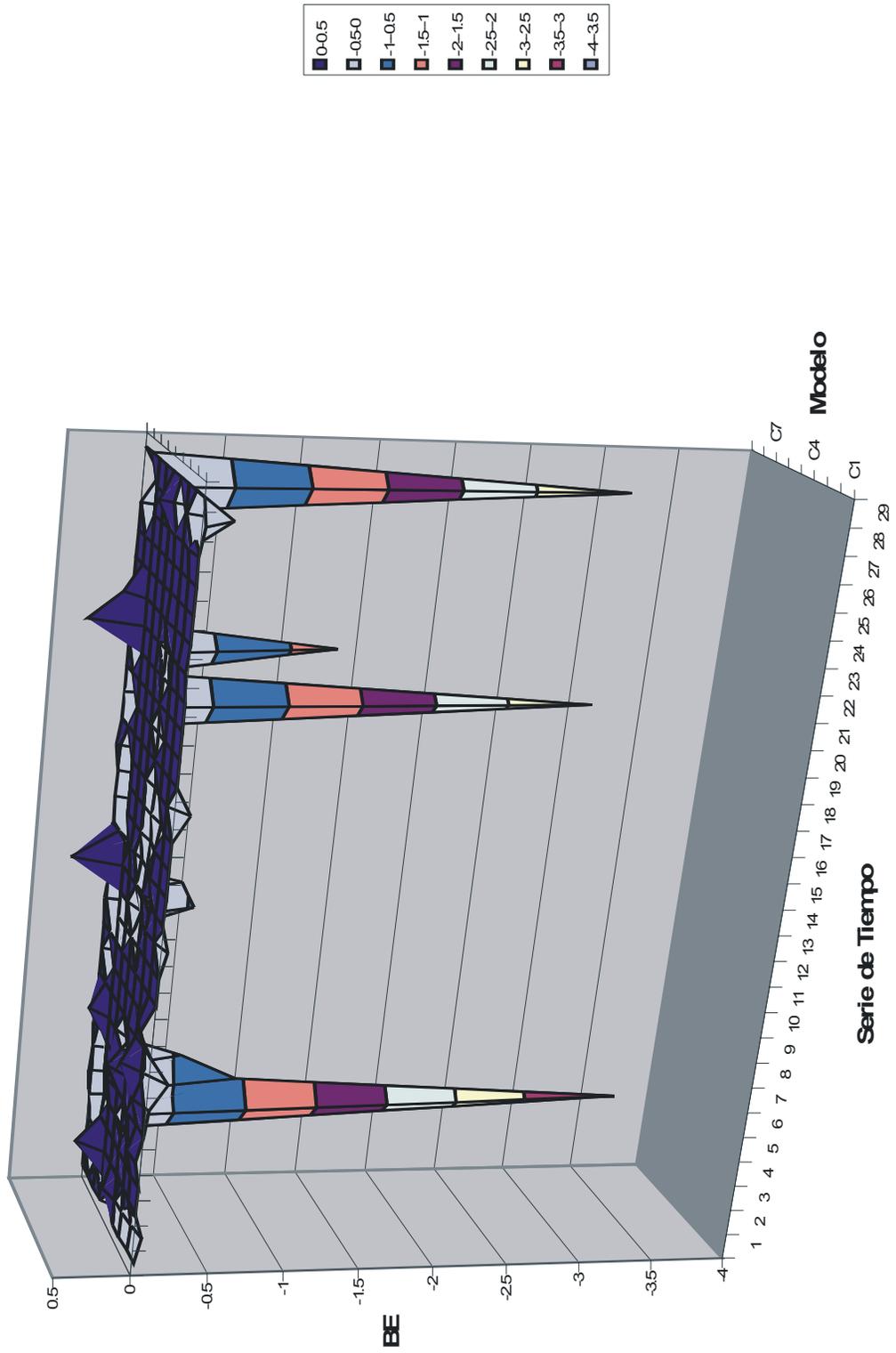


Figura 8-17: Gráfica de superficie del error BE para el modelado de series de tiempo

### MAE Técnicas de Modelado

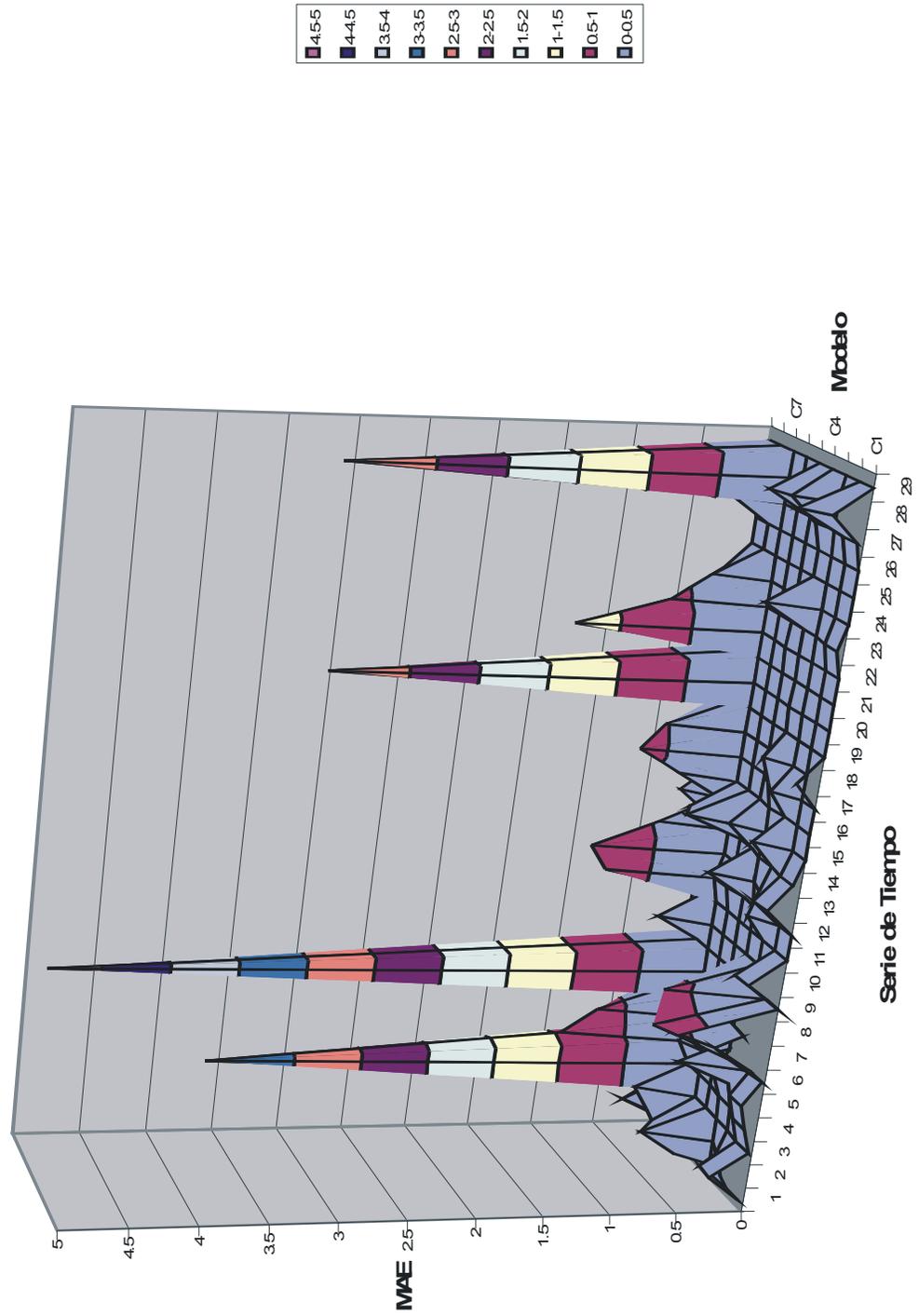


Figura 8-18: Gráfica de superficie del error MAE para el modelado de series de tiempo

de datos predichos. En general, el comportamiento de los tres errores es *independiente* de la capacidad de las técnicas para predecir (i.e. del número de series predichas por cada una de ellas).

Con respecto a las técnicas de modelado, se verifica en forma *empírica* que la predicción y el modelado poseen *diferente dificultad*, ya que aún cuando las técnicas de predicción o de modelado buscan la construcción de un modelo que reproduzca la dinámica de un sistema, en la predicción se genera un error que no es posible compensar como en el caso del modelado durante el proceso de entrenamiento, es decir, *la diferencia en los procesos de expansión de los errores de predicción y de modelado marcan la diferencia entre la capacidad de predecir y modelar a las series de tiempo*. Como se mencionó en la sección 2.6, las diferentes fuentes de error son *limitantes* para la predicción de las series de tiempo *reduciendo la extensión del horizonte de predicción* a cero en el caso de los fracasos en la predicción.

Dentro del conjunto de técnicas de modelado *no se observa una técnica que posea una superioridad clara para el modelado de series de tiempo*. Del subconjunto de técnicas con mejores resultados, la mayoría tienen como origen a la *Inteligencia Artificial*, y dentro de este subconjunto aquellas *que poseen algoritmos de naturaleza híbrida son las mejores*. Las técnicas cuyo origen se basa en la Teoría de Sistemas Dinámicos No Lineales, en general presentan una capacidad de modelado *deficiente* en todos los casos. En cuanto a los errores de modelado, estos *presentan comportamientos similares* a los observados para las técnicas de predicción. El error RMSE es pequeño y cercano a cero, lo que indica la generación de *buenos modelos* por parte de las técnicas, el error BE muestra que los modelos se *ajustan casi en forma óptima* a la dinámica de las series, y en promedio hay una tendencia de las técnicas al *subaprendizaje*, por último el error MAE como medida del comportamiento del error local de modelado, muestra que este *no crece* en forma apreciable para el conjunto de datos modelados. En este caso nuevamente, el comportamiento de los tres errores se observa es *independiente* de la capacidad de las técnicas para modelar (i.e. del número de series modeladas por cada una de ellas).

Por último, los experimentos realizados en este capítulo muestran que tanto para la predicción como en el modelado de series de tiempo, *no hay modelos que sean universales*, es decir capaces de predecir o modelar a todo el conjunto de series de tiempo (una verificación experimental independiente a la desarrollada en la presente tesis se presenta en el capítulo 13 de

apéndices). Sin embargo, *si es posible el construir técnicas de predicción o de modelado con la capacidad de adaptarse a una amplia gama de dinámicas de las series de tiempo.*

## Capítulo 9

# Métricas de la Predictibilidad de Series de Tiempo y Coeficientes de Capacidad para Predicción y Modelado

### 9.1 Introducción

Existen pocos trabajos en los cuales se realice un análisis para caracterizar la *predictibilidad* (o equivalentemente la *dificultad de predicción*) de una serie de tiempo. En el capítulo 3, se presentó un panorama del estado del arte en el estudio de la predictibilidad de series de tiempo. Se encontró que entre los trabajos reportados, varios de ellos utilizan al exponente de Lyapunov como el parámetro que más se identifica con la predictibilidad de series de tiempo [3, 84, 118, 119, 120, 121, 122, 123, 124]. Sin embargo, el considerar un solo parámetro para medir la predictibilidad como el exponente de Lyapunov o la Entropía de Información de Shannon, no proporciona información suficiente sobre la dinámica de las series de tiempo para determinar adecuadamente la predictibilidad.

En este capítulo se desarrollan una serie de métricas para medir la dificultad de predicción de las series de tiempo y en consecuencia su predictibilidad, en base a un conjunto finito de

parámetros ortogonales que caracterizan a las series de tiempo. Dichos parámetros tienen diferentes bases teóricas: Estadística, Análisis de Fourier, Análisis SSA, Teoría de Sistemas de Dinámicos No Lineales y Análisis de Gramáticas. La sección 9.2, presenta el desarrollo de diferentes métricas de predictibilidad de series de tiempo. En la sección 9.3, se presenta un análisis comparativo entre las diferentes métricas desarrolladas. En la sección 9.4, se presenta el desarrollo de coeficientes de capacidad de predicción y de modelado para la evaluación en forma cuantitativa de las técnicas (i. e. los modelos) para predecir o modelar series de tiempo. Finalmente, en la sección 9.5 se discuten los resultados presentados a lo largo del capítulo.

## 9.2 Métricas de la Predictibilidad de Series de Tiempo

### 9.2.1 Ortogonalidad de los Parámetros

Para el desarrollo de una métrica de *predictibilidad* de las series de tiempo, primero es necesario identificar a aquellos parámetros que en base a su definición, se sabe que proporcionan información relevante sobre las características de la *dinámica* de la serie de tiempo. El segundo paso, es verificar si los parámetros seleccionados son *ortogonales* entre sí, como se mencionó en la sección 7.3, la ortogonalidad asegura que cada uno de los parámetros proporcionan información diferente (i.e. complementaria) sobre las series de tiempo. Mediante un análisis de correlación bivariada es posible identificar el grado de ortogonalidad (i.e. independencia) de los parámetros.

En el capítulo 7, se describieron las tablas de datos correspondientes a los diferentes análisis realizados al conjunto de series de tiempo bajo estudio. Las tablas de datos que corresponden a parámetros estadísticos caracterizan la distribución estadística de los datos de las serie de tiempo. Analizando las definiciones de estos parámetros, se encuentra que los parámetros son de tipo *estático*, es decir, no proporcionan información sobre la dinámica de la serie de tiempo y en consecuencia no son útiles para cuantificar la predictibilidad de una serie de tiempo, aunque son útiles en la construcción de modelos de predicción de origen estadístico (ver tablas en las Figuras 7-17 y 7-18).

Para complementar lo anterior, las Figuras 9-1 y 9-2 presentan tablas del análisis de correlación bivariada de los parámetros estadísticos con respecto al número de predicciones exitosas (*NumPred*) para cada serie de tiempo y también con respecto al número de modelaciones

exitosas (*NumMod*) (ver tablas en las Figuras 8-7 y 8-15).

Se observa que los parámetros estadísticos (con excepción del parámetro de auto correlación) no poseen una correlación fuerte con relación al número de éxitos de predicción y modelado para cada serie de tiempo. Por lo anterior, con la excepción del parámetro de auto correlación, los parámetros de tipo estadístico *no son indicadores adecuados para la caracterización de la predictibilidad en series de tiempo*.

Del conjunto de parámetros estadísticos únicamente la correlación de Pearson (auto correlación) es el que posee mayor correlación con respecto a la predicción y modelado de las series de tiempo, además de que proporciona información sobre la dinámica de las mismas, ya que permite cuantificar el grado de auto correlación entre valores a diferentes instantes de tiempo dentro de una serie, por lo anterior, este parámetro fue seleccionado para formar parte de una de las tres métricas desarrolladas en este trabajo.

En relación con el análisis de Fourier, se seleccionó al parámetro de frecuencia dominante, ya que como se mencionó en la sección 4.3, es un valor escalar útil en la diferenciación entre señales periódicas y no periódicas.

Con respecto a la matriz correspondiente a los datos del porcentaje de varianza en las componentes resultantes del análisis PCA, como se mencionó en la sección 4.4, por ser parámetros escalares que caracterizan la *estructura* de la serie de tiempo, se decidió no emplearlo en la construcción de las métricas de predictibilidad, sino en la búsqueda de relaciones entre la predictibilidad de las series y sus características estructurales.

Finalmente, todos los parámetros derivados de la Teoría de Sistemas Dinámicos No Lineales, Teoría de Información y Teoría de la Computación son empleados en la construcción de las métricas de predictibilidad, pues todos ellos cuantifican características relacionadas con la dinámica de la serie de tiempo (ver tabla en la Figura 7-19).

Una vez seleccionados los parámetros para el desarrollo de las métricas. Se realizó el análisis de correlación bivariada para de esta forma, determinar el grado de ortogonalidad de los parámetros y decidir si el conjunto de parámetros se complementan entre sí, o existe redundancia en la información proporcionada por algunos de ellos. La Figura 9-3, presenta la tabla del conjunto de parámetros seleccionados y la abreviatura con la que se identifican en la tabla de análisis de correlación bivariada.

Correlación Bivariada	
Parámetros Estadísticos	Total de Técnicas que Predicen a cada Serie de Tiempo (NumPred)
Rango de Datos	0.117
Resolución	-0.09
Desviación Promedio	0.163
Desviación Estandar	0.162
Skewness	0.068
Kurtosis	-0.11
Moda (Máxima Probabilidad)	-0.003
Correlación de Pearson (Autocorrelación)	0.281
Tiempo de Correlación	0.078
Valor Promedio	0.123
Valor Medio	0.094
Valor Mínimo	-0.095
Valor Máximo	0.133
Cuartil Mínimo	0.014
Cuartil Máximo	0.186

Figura 9-1: Tabla de correlación bivariada de los parámetros estadísticos respecto a la predicción de series de tiempo

Correlación Bivariada	
Parámetros Estadísticos	Total de Técnicas que Modelan a cada Serie de Tiempo (NumMod)
Rango de Datos	0.204
Resolución	-0.164
Desviación Promedio	0.268
Desviación Estandar	0.259
Skewness	0.034
Kurtosis	-0.164
Moda (Máxima Probabilidad)	0.027
Correlación de Pearson (Autocorrelación)	0.583
Tiempo de Correlación	0.16
Valor Promedio	0.098
Valor Medio	0.074
Valor Mínimo	-0.199
Valor Máximo	0.19
Cuartil Mínimo	-0.078
Cuartil Máximo	0.234

Figura 9-2: Tabla de correlación bivariada de los parámetros estadísticos respecto al modelado de series de tiempo

Parámetro de Serie de Tiempo	Abreviatura
Correlación de Pearson (Autocorrelación)	ACO
Exponente de Hurst	EH
Frecuencia Dominante	FD
Exponente de Lyapunov	EL
Dimensión de Correlación	DCO
Dimensión de Capacidad	DCA
Dimension Fractal	DF
Dimensión Embebida	DE
Entropía Espacio Temporal (%)	EET
Recurrencia (%)	REC
Determinismo (%)	DET
Entropía de Información (Shannon)	ES
Información Mutua Promedio	IMP
Complejidad Relativa LZ	CRLZ
Número de Reglas de Producción	RP

Figura 9-3: Tabla de parámetros seleccionados para el cálculo de métricas de predictibilidad

La Figura 9-4, muestra la tabla de los resultados del análisis de correlación para el conjunto de parámetros seleccionados para caracterizar la predictibilidad de las series de tiempo.

Correlación de Pearson entre los Parámetros de las Series de Tiempo															
Parámetros	ACO	EH	FD	EL	DCO	DCA	DF	DE	EET	REC	DET	ES	IMP	CRLZ	RP
ACO	1														
EH	0.728	1													
FD	-0.471	-0.566	1												
EL	-0.161	-0.243	-0.01	1											
DCO	-0.205	-0.299	0.107	-0.306	1										
DCA	0.148	-0.05	0.337	-0.174	0.098	1									
DF	-0.085	-0.123	-0.041	-0.127	-0.115	0.172	1								
DE	0.324	0.057	-0.443	-0.03	0.173	0.016	0.065	1							
EET	-0.533	-0.517	0.268	0.458	0.059	-0.011	0.117	-0.08	1						
REC	0.244	0.297	-0.143	-0.241	-0.211	-0.062	-0.078	0.282	-0.189	1					
DET	0.718	0.645	-0.237	-0.529	-0.145	0.299	-0.176	0.216	-0.695	0.381	1				
ES	0.548	0.496	-0.011	-0.338	-0.182	0.328	-0.152	0.181	-0.414	0.485	0.718	1			
IMP	0.299	0.416	-0.243	-0.264	-0.301	0.051	0.068	0.173	-0.536	0.441	0.527	0.358	1		
CRLZ	-0.577	-0.632	0.175	0.506	0.306	-0.227	-0.013	0.046	0.718	-0.17	-0.834	-0.6	-0.586	1	
RP	0.245	-0.103	-0.008	0.426	-0.154	0.091	0.145	0.235	0.45	0.01	-0.184	-0.14	-0.055	0.335	1

Figura 9-4: Tabla de correlación bivariada del conjunto de parámetros de las métricas de predictibilidad

Entre los parámetros se observa que algunos presentan una fuerte correlación por ejemplo, el Determinismo y la Complejidad Relativa LZ que poseen una correlación inversa entre sí, o el Exponente de Hurst y la Auto Correlación que poseen una correlación directa entre sí. En el caso de los parámetros cuya correlación es débil, esto es indicativo de su mayor ortogonalidad con respecto a los demás parámetros. En general, *ninguno de los parámetros posee una correlación (o dependencia) significativa con todo el conjunto de parámetros*. Cada uno proporciona información diferente sobre la dinámica de las series de tiempo, por lo cuál, se considera importante el combinarlos para cuantificar la predictibilidad de las series de tiempo.

Una vez definido el conjunto de parámetros para caracterizar la predictibilidad de series de tiempo, a continuación se describe el desarrollo de las métricas de predictibilidad propuestas en este trabajo.

### 9.2.2 Métrica Coeficiente de Complejidad de Predicción (CCOP)

La métrica Coeficiente de Complejidad de Predicción (*CCOP*) se desarrolló a partir de una escala de valores discretos que mapea a los valores reales de un conjunto de parámetros característicos de las series de tiempo, los parámetros seleccionados que proporcionan la información sobre las series de tiempo son los siguientes: Frecuencia Dominante (*FD*), Exponente de Hurst (*EH*), Exponente de Lyapunov (*EL*), Dimensión de Correlación (*DCO*), Dimensión de Capacidad (*DCA*), Entropía Espacio Temporal (*EET*), Porcentaje de Recurrencia (*REC*), Porcentaje de Determinismo (*DET*), Entropía de Información de Shannon (*ES*) y Número de Reglas de Producción (*RP*) [10].

Para construir la escala de valores discretos se procede de la siguiente forma:

- Se identifica el valor máximo y mínimo de cada parámetro calculado para el conjunto de series de tiempo.
- Se calcula el rango de valores contenidos entre estos dos valores extremos y se divide dicho rango en tres subintervalos iguales.
- El valor discreto asignado a cada uno de los subintervalos se define bajo el siguiente criterio, si el valor del parámetro corresponde al subintervalo que indica mayor complejidad de predicción su valor es 3, si corresponde al subintervalo intermedio se le asigna el valor 2 y si corresponde al subintervalo de menor complejidad de predicción su valor es 1; adicionalmente se asigna el valor 0 cuando el parámetro posee valores que no son indicadores de la complejidad de predicción (esto en el caso particular del Exponente de Hurst). En la Figura 9-5, se presenta la tabla con las escalas de valores discretos correspondientes a cada parámetro, construidas a partir del conjunto de series de tiempo analizadas en este trabajo.

Finalmente el Coeficiente de Complejidad de Predicción para la serie de tiempo  $i$  –ésima se define como:

$$CCOP(ST_i) = FD_i + EH_i + EL_i + DCO_i + DCA_i + EET_i + REC_i + DET_i + ES_i + RP_i \quad (9.1)$$

Escala de los parámetros	Interpretación
Exponente de Hurst (Parámetro Estadístico)	
EH > 0.5 (0)	La tendencia local sobre el intervalo continuará (persistencia). La mejor predicción se basa en la extrapolación de la tendencia positiva en la serie de tiempo.
EH = 0.5 (1)	Valores sucesivos son independientes y la mejor predicción es el último valor medido.
EH < 0.5 (0)	La tendencia local puede revertirse (anti persistencia). La mejor predicción tiende al valor medio sobre el intervalo. Tendencia negativa en la serie de tiempo.
Dimensión de Capacidad (Parámetro Topológico)	
DCA > 1.776 (3)	Autosimilaridad alta.
0.91 < DCA < 1.776 (2)	Autosimilaridad media.
DCA < 0.91 (1)	Autosimilaridad baja.
Dimensión de Correlación (Parámetro Topológico)	
DCO > 2.442 (3)	Cantidad de información alta.
1.335 < DCO < 2.442 (2)	Cantidad de información media.
DCO < 1.335 (1)	Cantidad de información baja.
Exponente Principal de Lyapunov (Parámetro Topológico)	
EL < 1.909 (1)	Predictibilidad alta.
1.909 < EL < 3.818 (2)	Mediana Predictibilidad.
EL > 3.818 (3)	Predictibilidad baja.
Frecuencia Dominante (Parámetro Temporal)	
Existe (0)	Puede haber componente aleatoria.
No existe (1)	Fuerte componente aleatoria.
Número de Reglas de producción (Parámetro Computacional)	
RP > 64 (3)	Alta complejidad computacional y dificultad de predicción.
37 < RP < 64 (2)	Mediana Complejidad Computacional y dificultad de predicción.
RP < 37 (1)	Baja complejidad computacional y dificultad de predicción.
Entropía (Shannon) (Parámetro Computacional)	
ES > 4.68 (1)	Alto contenido de información en una medición en la serie de tiempo.
2.34 < ES < 4.68 (2)	Mediano contenido de información en una medición en la serie de tiempo.
ES < 2.34 (3)	Bajo contenido de información en una medición en la serie de tiempo.
Entropía Espacio Temporal (%) (Parámetro Espacio-Temporal)	
EET > 58 % (3)	Poco estructurado el Mapa de Recurrencia.
29% < EET < 58% (2)	Medianamente estructurado el Mapa de Recurrencia.
EET < 29% (1)	Altamente estructurado el Mapa de Recurrencia.
Recurrencia (%) (Parámetro Espacial)	
REC > 66% (1)	Alto grado de correlación en los datos.
33% < REC < 66% (2)	Mediano grado de correlación en los datos.
REC < 33% (3)	Bajo grado de correlación en los datos.
Determinismo (%) (Parámetro Espacial)	
DET > 66% (1)	Alto grado de determinismo.
33% < DET < 66% (2)	Mediano grado de determinismo.
DET < 33% (3)	Bajo grado de determinismo.

Figura 19-5: Tabla de escalas de los parámetros de la métrica CCOP

La Figura 9-6, muestra la tabla para los valores discretizados de los parámetros calculados y el correspondiente CCOP para el conjunto de series de tiempo.

Este coeficiente, como primera propuesta de una métrica de predictibilidad, permitió verificar en forma experimental la idea de utilizar un conjunto de parámetros de diversos orígenes para cuantificar la predictibilidad de series de tiempo. Sin embargo, posee dos desventajas:

- La discretización de los parámetros genera *perdida de información* en relación con las series.
- Los intervalos de discretización de los parámetros dependen del conjunto de series de tiempo, si este conjunto crece o disminuye, los intervalos deben de ser recalculados y los valores del CCOP para cada serie ajustarse, lo cuál hace que el coeficiente no sea *invariante*.

### 9.2.3 Métrica de Predictibilidad con 9 Parámetros (CDP1)

En la sección anterior, se ha desarrollado una primera métrica que unifica y expresa cuantitativamente por medio de un coeficiente, a un conjunto de parámetros que caracterizan a las series de tiempo y que permiten cuantificar la dificultad de predicción de las mismas. Sin embargo, este coeficiente posee dos desventajas: pérdida de información y que no es invariante. Para resolver estas desventajas se diseñó una nueva métrica que se denomina Coeficiente de Dificultad de Predicción 1 ( $CDP_1$ ), para construir esta métrica se parte de un conjunto de nueve parámetros ortogonales, además de los siguientes criterios:

- El conjunto de parámetros ortogonales una vez definido es *cerrado*, es decir, no puede aumentar o disminuir.
- No se discretiza a los parámetros, con lo cuál *no se pierde información* sobre las series de tiempo y se logra que *el coeficiente sea invariante*.
- Cada parámetro posee el mismo peso en la construcción del coeficiente.
- Hay dos tipos de relaciones en los parámetros: Una *relación directa*, es decir a mayor valor de un parámetro es mayor la dificultad de predicción de la serie de tiempo; o a

Serie de Tiempo	Exponente de Hurst	Frecuencia Dominante	Exponente de Lyapunov	Dimensión de Correlación	Dimensión de Capacidad	Entropía Espacio Temporal	Reurrencia	Determinismo	Entropía de Información (Shannon)	Número de Reglas de Producción	Coficiente de Complejidad de Predicción (CCOP)
She	0	0	1	1	1	1	3	2	2	1	12
Vanderpol	1	0	1	1	2	1	3	1	1	1	12
Qperiodc2	0	0	1	1	1	1	3	1	2	3	13
Qperiodc3	0	0	1	1	1	1	3	2	2	3	14
Monkey-Glass	0	0	1	1	2	2	3	2	3	3	17
Logistic	0	0	1	1	2	3	3	3	3	2	18
Lorenz	0	0	1	1	2	2	3	1	1	2	13
Rossler	0	0	1	1	2	1	3	1	2	3	14
Ikeda	0	0	1	1	2	2	3	3	3	3	18
Hanon	0	0	2	1	2	2	3	3	3	3	19
Canfor	0	0	3	1	1	3	3	3	3	3	20
Tent	0	0	1	1	2	2	3	3	2	1	15
A1	0	0	1	3	3	2	3	1	1	3	17
D1	0	1	1	2	2	2	3	2	3	3	19
Laser	0	0	1	2	2	2	3	1	2	3	16
Dow Jones	0	1	1	1	1	1	1	1	1	3	11
Kobe	0	0	1	1	1	3	2	2	2	3	15
EEG	0	0	1	2	1	3	3	3	3	3	19
ASCII TXT	0	0	2	2	1	3	3	3	3	3	20
El niño	0	1	2	2	2	3	3	2	2	3	20
HIV DNA	0	1	1	3	2	1	3	3	3	1	18
Human DNA	0	1	1	1	2	1	3	1	1	2	13
Lovaina	1	1	1	1	2	2	3	1	1	3	16
Plasma	0	0	2	1	1	3	3	3	2	3	18
Primos	0	0	1	3	1	3	3	3	3	3	20
SF500	0	0	2	1	1	3	3	3	2	3	18
Star	0	0	2	1	2	2	3	3	2	3	18
Brownian Motion	0	1	2	1	2	1	3	1	2	3	16
Write Noise	0	1	1	2	2	3	3	3	3	3	21

Figure 9-6: Tabla de resultados de la métrica CCOP

menor valor de un parámetro es menor la dificultad de predicción de la serie de tiempo. Una *relación inversa*, es decir a mayor valor de un parámetro es menor la dificultad de predicción de la serie de tiempo; o a menor valor de un parámetro es mayor la dificultad de predicción de la serie de tiempo.

Los parámetros seleccionados que cuantifican las diferentes características de las series de tiempo y su origen, se muestran en la tabla 9.2.

(9.2)

Parámetro	Origen
Exponente de Lyapunov	No Lineal
Exponente de Hurst	Estadística
Dimensión de Capacidad	No Lineal
Dimensión de Correlación	No Lineal
Entropía Espacio Temporal	No Lineal
Entropía de Shannon	Informacional
Porcentaje de Determinismo	No Lineal
Porcentaje de Recurrencia	No Lineal
Reglas de Producción	Computacional

El conjunto de parámetros seleccionados para la construcción de la métrica poseen diferentes unidades de medida, en la tabla 9.3 se muestran los parámetros y sus respectivas unidades de medida.

(9.3)

Parámetro	Abreviatura	Unidades
Exponente de Lyapunov	EL	$\frac{[bit]}{[tiempo]}$
Exponente de Hurst	EH	$\frac{[desplazamiento]}{[tiempo]}$
Dimensión de Capacidad	DCA	[adimensional]
Dimensión de Correlación	DCO	[adimensional]
Entropía Espacio Temporal	EET	[adimensional]
Entropía de Shannon	ES	[bit]
Porcentaje de Determinismo	DET	[adimensional]
Porcentaje de Recurrencia	REC	[adimensional]
Reglas de Producción	RP	[reglas]

A continuación se presenta el desarrollo de la forma funcional de la métrica de dificultad de predicción:

- Análisis dimensional: para que los parámetros no posean unidades de medida particulares, se multiplicaron aquellos que las poseen por el inverso de sus unidades de medida (ver tabla 9.4).

(9.4)

Parámetro	Unidad Adimensionada
Exponente de Lyapunov	$\frac{[bit]}{[tiempo]} \frac{[tiempo]}{[bit]}$
Entropía de Shannon	$[bit] \frac{1}{[bit]}$
Exponente de Hurst	$\frac{[desplazamiento]}{[tiempo]} \frac{[tiempo]}{[desplazamiento]}$
Reglas de Producción	$[reglas] \frac{1}{[reglas]}$

- Forma funcional del coeficiente de dificultad de predicción: el coeficiente es un cociente en el cual el numerador corresponde a la suma de los parámetros que indican la dificultad para predecir la serie de tiempo (*relación directa*), dichos parámetros son: Exponente de Lyapunov, Dimensión de Capacidad, Dimensión de Correlación, Entropía Espacio Temporal y Reglas de Producción. A su vez, el denominador está formado por la suma de los

parámetros que indican la facilidad para predecir la serie de tiempo (*relación inversa*), dichas propiedades son: Exponente de Hurst, Entropía de Shannon, Porcentaje de Determinismo y Porcentaje de Recurrencia. La tabla 9.5 muestra la tendencia de los parámetros respecto de la dificultad de predicción.

(9.5)

Parámetro	Abreviatura	Tendencia
Exponente de Lyapunov	EL	Directa
Exponente de Hurst	EH	Inversa
Dimensión de Capacidad	DCA	Directa
Dimensión de Correlación	DCO	Directa
Entropía Espacio Temporal	EET	Directa
Entropía de Shannon	ES	Inversa
Porcentaje de Determinismo	DET	Inversa
Porcentaje de Recurrencia	REC	Inversa
Reglas de Producción	RP	Directa

Los valores de los parámetros son tomados en su forma original con la excepción del parámetro exponente de Hurst, el cual por presentar tanto valores positivos como negativos o cero, para considerarlo en la métrica se transformo mediante la siguiente expresión:

$$EH^* = |EH - 0.5| \quad (9.6)$$

El cociente que se obtiene con los parámetros de las series de tiempo es de la forma:

$$\frac{EL + DCA + DCO + EET + RP}{EH^* + ES + DET + REC} \quad (9.7)$$

Experimentalmente se observó que los valores de dicho cociente corresponde a reales positivos sin embargo el rango de los mismos es amplio de valores cercanos a cero a valores del orden  $10^2$ . Para realizar una comparación adecuada de dicho cociente para las diferentes series de tiempo se aplico el logaritmo base 10 y a este se le sumo 1, de tal forma que todos los valores fuesen positivos, de esta forma la expresión final del Coeficiente de Dificultad de Predicción 1

$(CDP_1)$  es:

$$CDP_1 = \text{Log}_{10} \left( \frac{EL + DCA + DCO + EET + RP}{EH^* + ES + DET + REC} \right) + 1 \quad (9.8)$$

Finalmente la métrica de predictibilidad de la serie de tiempo es el inverso del coeficiente de dificultad de predicción:

$$\text{Predictibilidad} = \frac{1}{CDP_1} \quad (9.9)$$

La Figura 9-7, muestra la tabla de resultados de la métrica desarrollada para el conjunto experimental de series de tiempo.

#### 9.2.4 Métrica de Predictibilidad con 14 Parámetros (CDP2)

La métrica anterior consistía de 9 parámetros de las series de tiempo, los cuales fueron seleccionados por caracterizar diferentes aspectos de la dificultad de predicción (y por ende de la predictibilidad) de una serie de tiempo, sin embargo, queda la pregunta de si estos parámetros son suficientes para cuantificar la predictibilidad de una serie de tiempo, por ello, se desarrollo una segunda métrica la cual en su forma funcional es idéntica a la anterior pero a la cual se le incorporaron parámetros adicionales de las series de tiempo que también son indicadores de la dificultad de predicción, estos parámetros son: la Dimensión Embebida ( $DE$ ), la Dimensión Fractal ( $DF$ ), la Información Mutua Promedio ( $IMP$ ), la Auto Correlación ( $ACO$ ) y la Complejidad Relativa LZ ( $CRLZ$ ). La tabla 9.10 muestra el origen de estos parámetros,

Parámetro	Origen
Dimensión Embebida	No Lineal
Dimensión Fractal	No Lineal
Información Mutua Promedio	Informacional
Autocorrelación	Estadística
Complejidad Relativa LZ	Computacional

La tabla 9.11, muestra los parámetros y sus respectivas unidades de medida,

Serie de Tiempo	Coficiente de Dificultad de Predicción con 9 parámetros (CDP <sub>1</sub> )	Predictibilidad (1/CDP <sub>1</sub> )
Sine	0.512997545	1.949327066
Vanderpol	0.604397824	1.654539378
Qperiodic2	0.799020327	1.251532615
Qperiodic3	1.131340574	0.88390713
Mackey-Glass	1.522174611	0.656954855
Logistic	2.148616946	0.465415672
Lorenz	0.974519472	1.026146762
Rosler	0.888900691	1.124985063
Ikeda	3.076784726	0.325014614
Henon	2.068058885	0.483545225
Cantor	2.782220495	0.359425143
Tent	1.276447623	0.783424233
A1	1.122688313	0.890719168
D1	1.32138263	0.756783067
Laser	1.136091261	0.880210978
Dow Jones	0.529138344	1.889864932
Kobe	1.180443197	0.847139449
EEG	2.722554831	0.367302061
ASCIITXT	1.989644041	0.502602465
El niño	1.502251915	0.665667316
HIV DNA	0.658902795	1.517674546
Human DNA	0.611454124	1.63544567
Lovaina	1.223627448	0.81724221
Plasma	2.489831953	0.401633531
Primos	2.133683262	0.468673124
SP500	1.789566214	0.558794635
Star	1.679878416	0.595281177
Brownian Motion	1.042379904	0.95934313
White Noise	3.497587263	0.28591138

Figura 9-7: Tabla de resultados sobre la predictibilidad en series de tiempo con la métrica CDP<sub>1</sub>

(9.11)

Parámetro	Abreviatura	Unidades
Dimensión Embebida	DE	[adimensional]
Dimensión Fractal	DF	$\frac{[\text{número de estados}]}{[\text{volumen}]}$
Información Mutua Promedio	IMP	[ <i>bit</i> ]
Autocorrelación	ACO	[adimensional]
Complejidad Relativa LZ	CRLZ	[número de subcadenas]

Para eliminar las dimensiones y que las unidades sean adimensionales se realizó como en el caso de la métrica  $CDP_1$ , un análisis dimensional (ver tabla 9.12).

(9.12)

Parámetro	Unidad Adimensionada
Dimensión Fractal	$\frac{[\text{número de estados}]}{[\text{volumen}]} \frac{[\text{volumen}]}{[\text{número de estados}]}$
Información Mutua Promedio	$[bit] \frac{1}{[bit]}$
Complejidad Relativa LZ	$[\text{número de subcadenas}] \frac{1}{[\text{número de subcadenas}]}$

La tabla 9.13, muestra la tendencia del conjunto de parámetros adicionales, respecto de la dificultad de predicción.

(9.13)

Propiedad	Abreviatura	Tendencia
Dimensión Embebida	DE	Directa
Dimensión Fractal	DF	Directa
Información Mutua Promedio	IMP	Inversa
Autocorrelación	ACO	Inversa
Complejidad Relativa LZ	CRLZ	Directa

La expresión final del Coeficiente de Dificultad de Predicción 2 ( $CDP_2$ ) ahora con 14 parámetros es:

$$CDP_2 = \text{Log}_{10} \left( \frac{EL + DCA + DCO + EET + RP + DE + DF + CRLZ}{EH + ES + DET + REC + IMP + ACO} \right) + 1 \quad (9.14)$$

y la correspondiente expresión para la métrica de predictibilidad es nuevamente de la forma:

$$\text{Predictibilidad} = \frac{1}{CDP_2} \quad (9.15)$$

La Figura 9-8, muestra la tabla de resultados de la métrica desarrollada para el conjunto experimental de series de tiempo.

### 9.3 Comparación de las Métricas de Predictibilidad

En esta sección se presentan los resultados de comparar las tres métricas desarrolladas con la métrica denominada Índice de Dificultad de Modelado (IDM) [162], y con un conjunto de parámetros (Exponente de Lyapunov, Entropía de Información de Shannon, Coeficiente de Complejidad Relativa LZ y Número de Reglas de Producción) representativos de características con diferentes orígenes (No Lineal, Informacional y Computacional) sobre la series de tiempo. El Exponente de Lyapunov es ampliamente usado como una medida de predictibilidad de las series de tiempo [3, 84, 118, 119, 120, 121, 122, 123, 124]. Por su parte, la Entropía de Información de Shannon es usada para determinar la cantidad de información en un mensaje (o señal) en telecomunicaciones [138, 145]. En cuanto al Coeficiente de Complejidad Relativa LZ, es una medida de complejidad algorítmica de una cadena de datos [126, 146]. Por último, el Número de Reglas de Producción es una medida de complejidad computacional que ha sido utilizada en la determinación de la predictibilidad de las series de tiempo [9, 10]. El conjunto de parámetros ha sido descrito con detalle en el capítulo 4, a continuación se describe la métrica Índice de Dificultad de Modelado.

#### 9.3.1 Índice de Dificultad de Modelado (IDM)

Para la construcción de esta métrica se utilizaron tres parámetros de las series de tiempo: Exponente de Lyapunov ( $EL$ ) cuyo origen es la Teoría de Sistemas Dinámicos No Lineales,

Serie de Tiempo	Coefficiente de Dificultad de Predicción con 14 parámetros (CDP <sub>2</sub> )	Predictibilidad (1/CDP <sub>2</sub> )
Sine	0.536188191	1.86501683
Vanderpol	0.604585119	1.654026816
Qperiodic2	0.789694626	1.266312277
Qperiodic3	1.125352046	0.888610816
Mackey-Glass	1.502452625	0.665578391
Logistic	1.87158717	0.534305864
Lorenz	0.941044521	1.062648979
Rosler	0.852835387	1.172559224
Ikeda	2.273844247	0.43978386
Henon	1.662901586	0.601358498
Cantor	2.658143428	0.376202424
Tent	1.175060008	0.851020368
A1	1.125337164	0.888622568
D1	1.29667964	0.77120051
Laser	1.163525516	0.859456872
Dow Jones	0.558037204	1.791995215
Kobe	1.193006503	0.838218398
EEG	2.242075478	0.446015315
ASCIITXT	1.939832694	0.515508375
El niño	1.460186187	0.684844172
HIV DNA	0.809846705	1.234801592
Human DNA	0.638748604	1.565561151
Lovaina	1.195510181	0.836462973
Plasma	2.314944118	0.43197587
Primos	2.04697427	0.488525926
SP500	1.769062943	0.565271012
Star	1.550244553	0.645059515
Brownian Motion	0.966610896	1.034542446
White Noise	3.055508184	0.327277801

Figura 9-8: Tabla de resultados sobre la predictibilidad en series de tiempo con la métrica CDP<sub>2</sub>

Entropía de Información de Shannon ( $ES$ ) basada en la Teoría de Información y Número de Reglas de Producción ( $RP$ ) cuyo origen es la Teoría de Computación. A partir de estos tres parámetros se construyeron las siguientes definiciones:

- El Índice de Exponente de Lyapunov de una serie de tiempo,

$$IEL = \frac{EL_i}{EL_{\max}} \quad (9.16)$$

donde  $EL_i$  es el Exponente de Lyapunov de la serie analizada y  $EL_{\max}$  es el máximo Exponente de Lyapunov registrado dentro del conjunto de series de tiempo analizado.

- El índice de Entropía de Shannon de una serie de tiempo es:

$$IES = \frac{ES_{\min} + 1}{ES_i + 1} \quad (9.17)$$

donde  $ES_i$  es la Entropía de Shannon de la serie analizada y  $ES_{\min}$  es la mínima Entropía de Shannon registrada en el conjunto de series de tiempo analizado.

- El Índice de Reglas de Producción de una serie de tiempo,

$$IRP = \frac{RP_i}{RP_{\max}} \quad (9.18)$$

donde  $RP_i$  es el número de Reglas de Producción de la serie analizada y  $RP_{\max}$  es el máximo número de Reglas de Producción registrado dentro del conjunto de series de tiempo analizado.

Finalmente el Índice de Dificultad de Modelado de una serie de tiempo que pertenece a un conjunto de series analizadas esta dado por la expresión:

$$IDM = IEL + IES + IRP \quad (9.19)$$

La Figura 9-9, muestra la tabla de los valores calculados del IDM para las series de tiempo analizadas en esta tesis.

Serie de Tiempo	Índice de Dificultad de Modelado (IDM)
Sine	0.599
Vanderpol	0.8
Qperiodic2	1.176
Qperiodic3	1.256
Mackey-Glass	1.945
Logistic	1.768
Lorenz	0.969
Rosler	1.263
Ikeda	2.033
Henon	2.139
Cantor	2.791
Tent	0.453
A1	1.031
D1	1.775
Laser	1.145
Dow Jones	0.873
Kobe	1.271
EEG	2.11
ASCII TXT	1.673
El niño	1.538
HIV DNA	0.603
Human DNA	0.737
Lovaina	1.184
Plasma	1.849
Primos	1.396
SP500	1.803
Star	1.806
Brownian Motion	1.558
White Noise	2.082

Figura 9-9: Tabla de resultados de la métrica IDM

### 9.3.2 Comportamiento de las Métricas

La Figura 9-10, presenta la tabla de resultados de las métricas desarrolladas en este trabajo, la métrica IDM y el conjunto de parámetros seleccionados, además de dos variables correspondientes a: el número de predicciones para cada serie de tiempo ( $NumPred$ ) y el número de modelaciones para cada serie de tiempo ( $NumMod$ ). Estas dos variables se consideran la referencia contra la que se comparan las métricas y los parámetros evaluados.

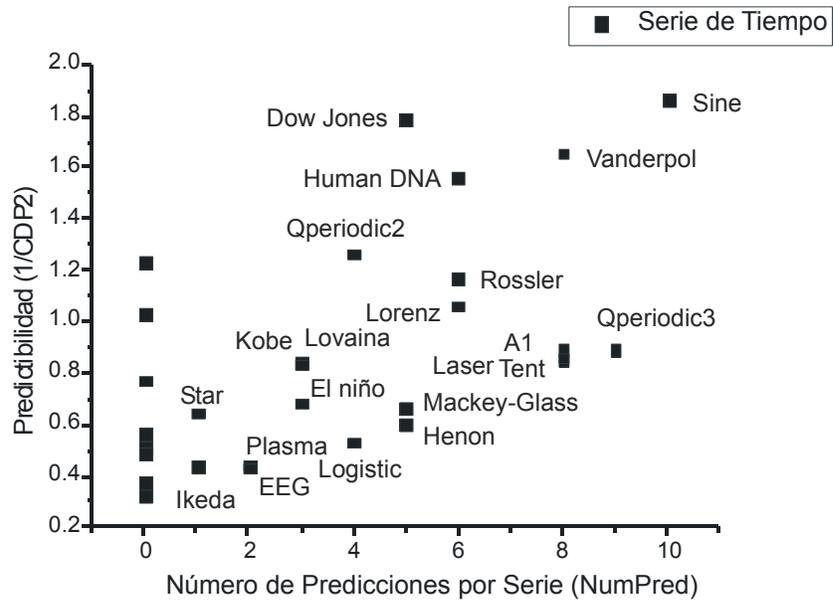
A partir de la tabla que se muestra en la Figura 9-10, se observa que el comportamiento de las métricas y parámetros para medir la predictibilidad de las series de tiempo, es de carácter *no lineal* respecto del número de éxitos para predicción y modelado de cada serie de tiempo [10]. Para ilustrar lo anterior, se presenta a continuación en la Figura 9-11, la gráfica de la predictibilidad correspondiente a la métrica  $CDP_2$  versus el número de predicciones exitosas para cada serie de tiempo ( $NumPred$ ). Además de la no linealidad observada es importante notar la dispersión de los puntos en la gráfica en especial en el caso de cero predicciones, en el cual se observa una acumulación de series con predictibilidad alta pero que no fueron predichas exitosamente, a priori lo único que se puede decir es que *no se tuvo éxito en la búsqueda del modelo capaz de predecir la serie de tiempo*, pero dicho modelo existe aunque no fue localizado, hay que recordar que los teoremas de No Free Lunch imponen *límites sobre la habilidad para la búsqueda de la solución óptima de un problema* [55, 56], y la gráfica muestra de forma indirecta esta limitación. Se espera que al continuar desarrollando en el futuro más estudios sistemáticos con nuevos modelos de predicción y series de tiempo, el comportamiento de la métrica de predictibilidad se defina mejor al ajustarse la distribución de resultados conforme crezca la colección de datos experimentales.

Las métricas de predictibilidad también muestran que el tipo de comportamiento dinámico de las series establecido con la siguiente clasificación: *periódico o regular, cuasi periódico, caótico, complejo y aleatorio o estocástico*; no está relacionado de forma directa (o *lineal*) con la predictibilidad de las series de tiempo. Por ejemplo, una serie de tiempo que tenga un comportamiento de tipo caótico o complejo puede tener una predictibilidad alta (por ejemplo: Rossler o Dow Jones) comparable con la de series cuyo comportamiento sea más *simple* como las de tipo periódico (series Sine y Vanderpol), como se observa en la tabla de la Figura 9-12.

La Figura 9-13, muestra en orden ascendente los valores normalizados (con valor máximo

Serie de Tiempo	Coefficiente de Complejidad de Predicción (CCOP)	Coefficiente de Predicción con 9 parámetros (CDP-1)	Coefficiente de Predicción con 14 parámetros (CDP2)	Índice de Dificultad de Modelado (IDM)	Exponente de Lyapunov (EL)	Entropía de Información de Shannon (ES)	Complejidad Relativa LZ (CRLZ)	Número de Reglas de Producción (RP)	Total de Técnicas que Predicen a cada Serie de Tiempo (NumPred)	Total de Técnicas que Modelan a cada Serie de Tiempo (NumMod)
Sine	12	0.51299755	0.53618819	0.599	0.517	3.7118	0.05979471	27	10	9
Vanderpol	12	0.60439782	0.60458512	0.8	1.864	4.981	0.07972627	28	8	9
Qperiodic2	13	0.79902033	0.78969463	1.176	0.925	2.445885	0.1494868	66	4	7
Qperiodic3	14	1.13134057	1.12535205	1.256	1.383	4.232	0.4384945	75	9	6
Monkey-Glass	17	1.52217461	1.50245263	1.945	1.481	0.15914973	0.4783576	75	5	7
Logistic	18	2.14861695	1.87158717	1.768	0.76	0.03522222	0.7175364	61	4	6
Lorenz	13	0.97451947	0.94104452	0.969	0.601	4.78931065	0.1793841	63	6	7
Rosler	14	0.88890069	0.85283539	1.263	1.049	2.45897942	0.1694183	72	6	7
Ikeda	18	3.07678473	2.27384425	2.033	1.452	0	0.787297	71	1	2
Hanon	19	2.06805889	1.66290159	2.139	2.301	0.0204499	0.6378102	69	5	4
Caritor	20	2.7822205	2.65814343	2.791	5.728	0	1.056373	72	0	0
Tent	15	1.27644762	1.17506001	0.453	0.477	3.401	0.03986314	13	8	6
A1	17	1.12268831	1.12533716	1.031	0.2234006	7.02	0.28900078	79	8	7
D1	19	1.32138263	1.29667964	1.775	1.288	0.8521749	0.229213	92	0	5
Laser	16	1.13609126	1.16362552	1.145	0.949	2.93549693	0.3786998	66	8	7
Dow Jones	11	0.52913834	0.5580372	0.873	0.144	6.43519134	0.1993157	65	5	6
Kobe	15	1.1804432	1.1930065	1.271	1.102	3.74401017	0.787297	79	3	4
EEG	19	2.72255483	2.24207548	2.11	1.199	0	0.7773312	82	2	4
ASCII TXT	20	1.98964404	1.93983269	1.673	2.709	2.107	1.036442	80	0	0
El niño	20	1.50225192	1.46018619	1.538	2.322	3.526	0.2092815	83	3	4
HIV DNA	18	0.6589028	0.80984671	0.603	0	1	0.8072285	9	0	0
Human DNA	13	0.61145412	0.6387486	0.737	0.322	7.28	0.04982892	51	6	5
Lovaina	16	1.22362745	1.19551018	1.184	1.0693	5.574	0.3089393	77	3	6
Plasma	18	2.48983195	2.31494412	1.849	3.383	2.866	0.9766468	91	2	3
Primos	20	2.13368326	2.04697427	1.396	0.594	1.295	0.9766468	78	0	0
SP500	18	1.78956621	1.76906294	1.803	3.569	2.585	1.036442	82	0	0
Star	18	1.67987842	1.55024455	1.806	3.289	3.102	0.5879813	90	1	6
Brownian Motion	16	1.0423799	0.9666109	1.558	2.043	2.74038235	0.1694183	85	0	7
White Noise	21	3.49758726	3.05550818	2.082	1.606	0	1.066339	73	0	0

Figure 9-10: Tabla comparativa de las diferentes métricas de predictibilidad de series de tiempo



Con NumPred=0 en orden ascendente de Predictibilidad: White Noise

- Cantor
- Primos
- ASCIITXT
- SP500
- D1
- Brownian Motion
- HIV DNA

Figura 9-11: Gráfica del comportamiento de la métrica de predictibilidad CDP2

Serie de Tiempo	Comportamiento Dinámico	Coficiente de Complejidad de Predicción (CCOP)	Predictibilidad (1/CDP1)	Predictibilidad (1/CDP2)
Sine	periódica	12	1.94932707	1.86501683
Vanderpol	periodica	12	1.65453938	1.65402682
Qperiodic2	cuasi periódica	13	1.25153262	1.26631228
Qperiodic3	cuasi periódica	14	0.88390713	0.88861082
Mackey-Glass	caótica	17	0.65695486	0.66557839
Logistic	caótica	18	0.46541567	0.53430586
Lorenz	caótica	13	1.02614676	1.06264898
Rosler	caótica	14	1.12498506	1.17255922
Ikeda	caótica	18	0.32501461	0.43978386
Henon	caótica	19	0.48354523	0.6013585
Cantor	caótica	20	0.35942514	0.37620242
Tent	caotica	15	0.78342423	0.85102037
A1	compleja	17	0.89071917	0.88862257
D1	compleja	19	0.75678307	0.77120051
Laser	compleja	16	0.88021098	0.85945687
Dow Jones	compleja	11	1.88986493	1.79199522
Kobe	compleja	15	0.84713945	0.8382184
EEG	compleja	19	0.36730206	0.44601532
ASCIITXT	compleja	20	0.50260247	0.51550838
El niño	compleja	20	0.66566732	0.68484417
HIV DNA	compleja	18	1.51767455	1.23480159
Human DNA	compleja	13	1.63544567	1.56556115
Lovaina	compleja	16	0.81724221	0.83646297
Plasma	compleja	18	0.40163353	0.43197587
Primos	compleja	20	0.46867312	0.48852593
SP500	compleja	18	0.55879464	0.56527101
Star	compleja	18	0.59528118	0.64505952
Brownian Motion	estocástica	16	0.95934313	1.03454245
White Noise	estocástica	21	0.28591138	0.3272778

Figura 9-12: Tabla de métricas de predictibilidad y el comportamiento dinámico de las series de tiempo

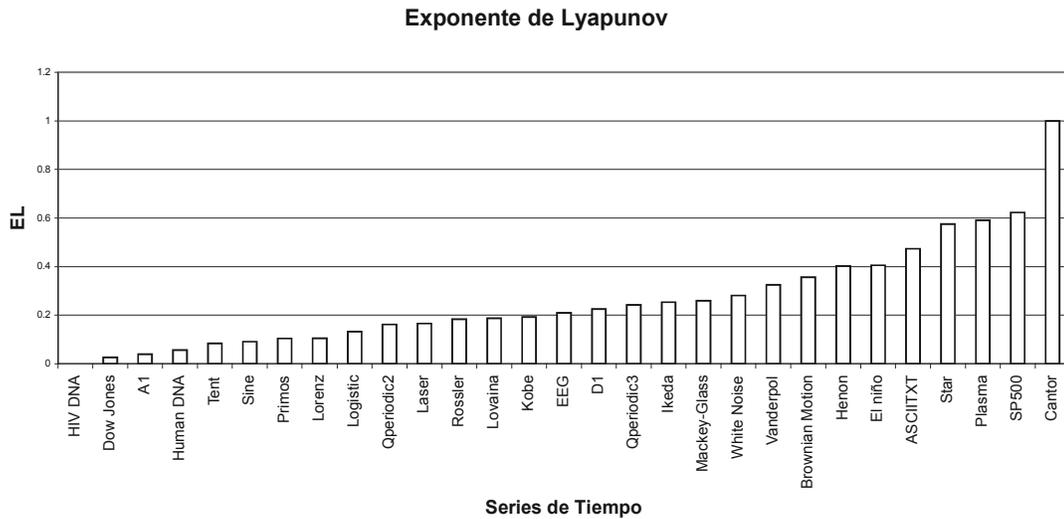


Figura 9-13: Parámetro exponente de Lyapunov normalizado y ordenado en forma ascendente para el conjunto de series de tiempo

uno) del parámetro Exponente de Lyapunov ( $EL$ ) correspondientes al conjunto de las series de tiempo. Se puede observar en la gráfica que si consideramos a las series de tiempo Sine cuyo comportamiento dinámico es periódico y White Noise que presenta un comportamiento aleatorio, como las series representativas de los comportamientos extremos (de menor a mayor complejidad en el comportamiento dinámico), la información proporcionada por este parámetro (horizonte de predicción, ver tabla en la Figura 7-19) no es suficiente para definir con claridad una correspondencia entre el comportamiento dinámico y la predictibilidad de las series de tiempo.

La Figura 9-14, muestra otro ejemplo del ordenamiento de las series de tiempo ahora en base al parámetro Complejidad Relativa LZ ( $CRLZ$ ), este parámetro que proporciona información sobre la estructura y jerarquía de cadenas de datos (ver tabla en la Figura 7-19), al considerar las series de tiempo Sine y White Noise como series de referencia, observamos que presenta una mejor correspondencia en relación con el comportamiento dinámico y la predictibilidad de las series de tiempo, sin embargo se observan zonas donde se presentan saltos en la magnitud de los valores del parámetro, es decir no hay un crecimiento suave de los valores del parámetro

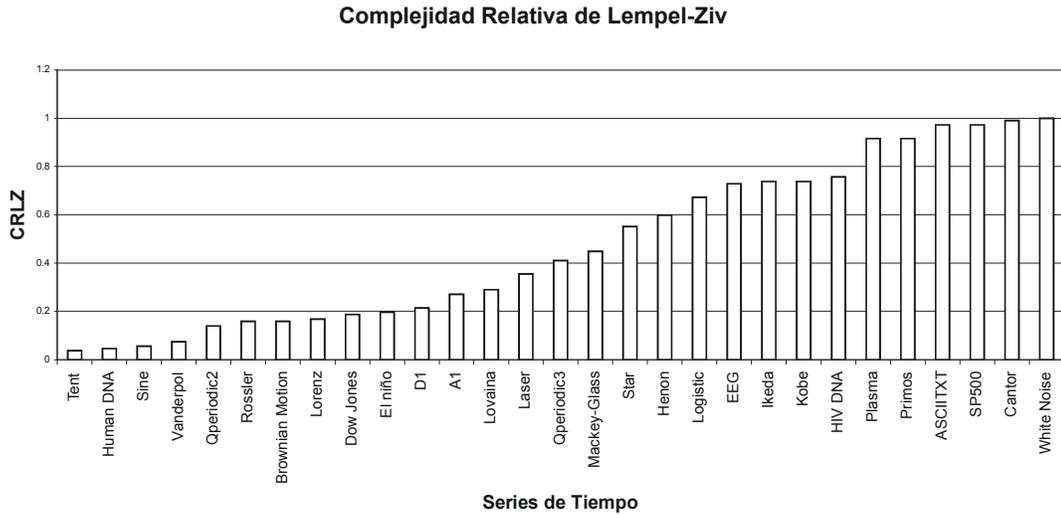


Figura 9-14: Parámetro complejidad relativa LZ normalizado y ordenado en forma ascendente para el conjunto de series de tiempo

correspondientes al conjunto de series de tiempo.

En la Figura 9-15, se muestran en orden ascendente los valores normalizados (con valor máximo uno) de la métrica de predictibilidad  $CDP_2$  en su representación como *coeficiente de dificultad de predicción*, se observa que la relación entre el comportamiento dinámico de las series de tiempo y esta métrica aunque es no lineal (como se explicó en un párrafo anterior), permite obtener una mejor correspondencia entre el comportamiento dinámico y la predictibilidad de las series de tiempo. Las series de tiempo Sine (menor valor del coeficiente de dificultad de predicción) y White Noise (mayor valor del coeficiente de dificultad de predicción) se asocian con los valores extremos de la predictibilidad y el crecimiento de los valores de la métrica presenta un comportamiento más suave, sin los saltos en la magnitud de los valores como ocurre para el parámetro Complejidad Relativa LZ. La Figura 9-16 muestra otra representación en la forma de un eje vertical, del mismo ordenamiento en orden ascendente de los valores normalizados del *coeficiente de dificultad de predicción*. Al usar un conjunto de parámetros que caracterizan con diferentes facetas la dinámica de las series de tiempo obtenemos mayor información para calcular la predictibilidad de las series de tiempo, a diferencia de utilizar como medida de predictibilidad un solo parámetro (por ejemplo: Exponente de Lyapunov o

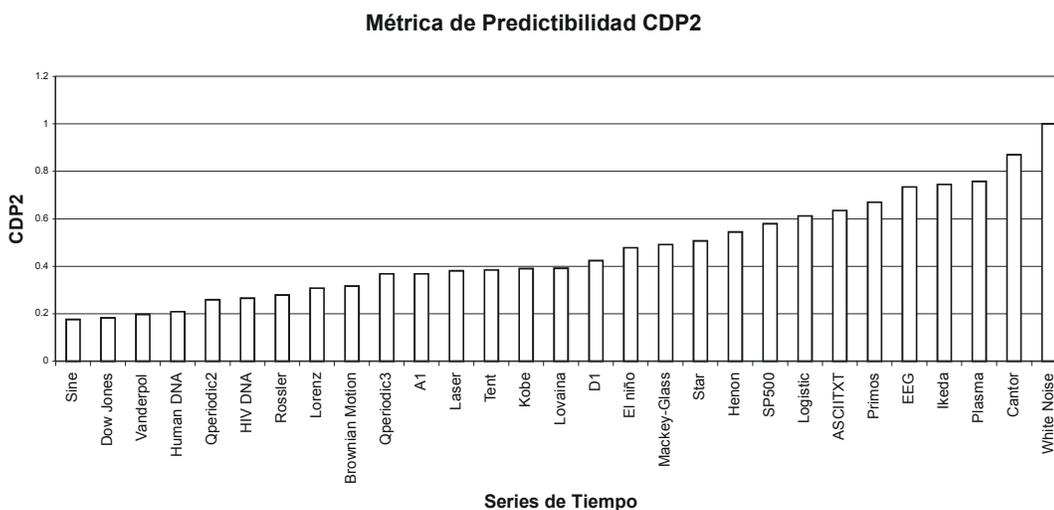


Figura 9-15: Métrica de predictibilidad CDP2 normalizada y ordenada en forma ascendente para el conjunto de series de tiempo

Complejidad Relativa LZ) que proporciona información parcial sobre la dinámica de las series de tiempo y en consecuencia sobre la predictibilidad de las mismas. En el capítulo 13 de apéndices se presentan una serie de gráficas, que ilustran como el conjunto de parámetros de las series de tiempo caracteriza en forma particular la dinámica para cada una de ellas.

### 9.3.3 Análisis de Correlación de las Métricas

La Figura 9-17, presenta la tabla del análisis de correlación correspondiente al conjunto de datos sobre las diferentes métricas (ver tabla en la Figura 9-10).

A partir del análisis de correlación se observa que de las métricas desarrolladas las que poseen una *correlación significativa* con las variables de referencia NumPred y NumMod, correspondientes al número de predicciones y modelaciones exitosas para cada serie, son la *métrica CCOP* y la *métrica CDP<sub>2</sub>*. Con relación a los parámetros individuales, se observa que el parámetro con la *mejor correlación* con respecto a las variables de referencia es la *Complejidad Relativa LZ (CRLZ)*, su correlación es además *la mejor de todo el conjunto evaluado*. Por último, la *mejor correlación* entre este parámetro y el resto del conjunto evaluado corresponde a la *métrica CDP<sub>2</sub>*. Es importante hacer notar que aún cuando la mejor correlación con respecto

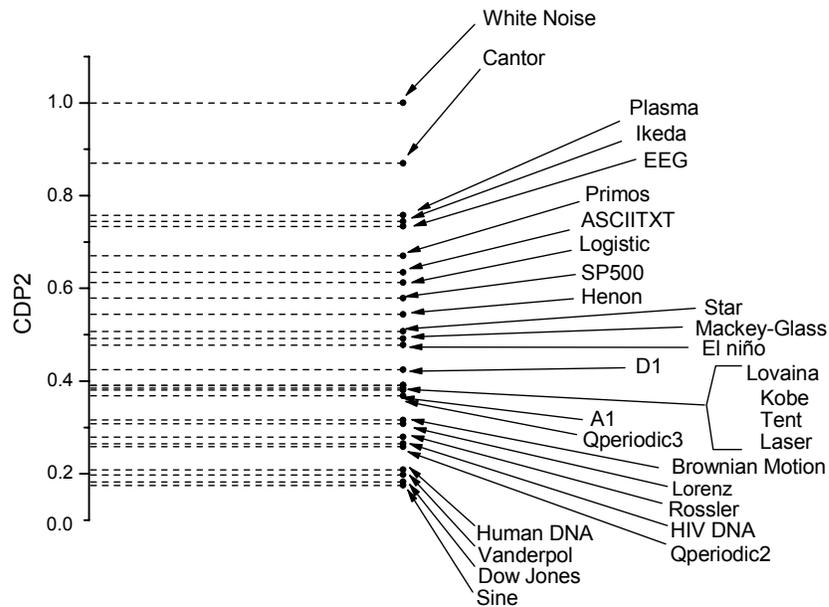


Figura 9-16: Métrica de predictibilidad CDP2 normalizada y ordenada en forma ascendente para el conjunto de series de tiempo

Métricas	CCOP	CDP1	CDP2	IDM	EL	ES	CRLZ	RP	NUM PRED	NUM MOD
CCOP	1									
CDP1	0.783	1								
CDP2	0.826	0.984	1							
IDM	0.724	0.821	0.826	1						
EL	0.473	0.502	0.56	0.726	1					
ES	-0.672	-0.683	-0.667	-0.695	-0.338	1				
CRLZ	0.734	0.774	0.825	0.662	0.506	-0.6	1			
RP	0.435	0.425	0.45	0.664	0.426	-0.14	0.335	1		
NUM PRED	-0.708	-0.563	-0.601	-0.603	-0.456	0.534	-0.674	-0.445	1	
NUM MOD	-0.728	-0.644	-0.7	-0.483	-0.406	0.481	-0.851	-0.185	0.758	1
Número de Parámetros de Métrica	10	9	14	3	1	1	1	1		

Figura 9-17: Tabla de correlación de las métricas de predictibilidad

a las variables de referencia corresponde a el parámetro Complejidad Relativa LZ, la diferencia con la correlación de la métrica  $CDP_2$  es mínima, por otra parte, la métrica  $CDP_2$  utiliza para medir la predictibilidad un conjunto de parámetros que como se vio anteriormente caracterizan en forma global a las series de tiempo, lo cual proporciona una mayor información sobre las mismas y una mejor correspondencia entre el comportamiento dinámico y la predictibilidad de las series, a diferencia de un solo parámetro que únicamente proporciona información sobre una característica particular de las series de tiempo lo que es insuficiente para determinar la predictibilidad.

## 9.4 Coeficientes de Capacidad de Predicción y de Modelado

De forma análoga al desarrollo de una métrica para medir la dificultad de predicción de la serie de tiempo en base a los parámetros que la caracterizan, se desarrollo una métrica para medir la capacidad de predicción (o de modelado) de las técnicas evaluadas, y que además posibilite la comparación y clasificación de los diferentes modelos de predicción y modelado. En la literatura especializada no se encuentran reportadas métricas similares para caracterizar a los modelos, aún cuando hay gran cantidad de trabajos donde se comparan diferentes modelos de predicción [29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42], la métrica que se presenta en este trabajo es producto de la evolución de una primera versión que fue desarrollada por Bautista-Thompson y Figueroa-Nazuno en conjunto con la métrica de predictibilidad CCOP [9, 10]. Para la descripción de esta métrica se toma como ejemplo la correspondiente a las técnicas de predicción o predictores (sin embargo todo el desarrollo es idéntico para las técnicas de modelado). La métrica se expresa mediante un coeficiente que se denomina Coeficiente de Capacidad de Predicción ( $CCAP$ ) o de Modelado ( $CCAM$ ) según sea el caso.

La métrica tiene las siguientes características:

- 1 *No es invariante*, ya que al aumentar el número de casos exitosos de predicción de cada técnica el coeficiente correspondiente se ajusta.
- 2 Es *dependiente* del conjunto de series de tiempo que se analiza.

La métrica se construye a partir de los siguientes parámetros:

- El promedio del error raíz media cuadrada ( $PRMSE$ ) para el conjunto de series predichas por cada predictor. Para poder realizar una comparación equitativa de este parámetro para cada predictor se asigno el valor de 1 para cada una de las series no predichas por el predictor. La expresión correspondiente para este parámetro es:

$$PRMSE = \frac{1}{P} \sum_{i=1}^P RMSE_i + \frac{1}{N-P} \sum_{j=1}^{N-P} (1)^j \quad (9.20)$$

donde  $P$  corresponde al número de series predichas y  $N - P$  corresponde al número de series no predichas por el predictor. Este promedio proporciona información de tipo global sobre el comportamiento del error de la predicción en el rango de la misma, que en todos los casos es de 50 datos.

- El promedio del promedio del error local de predicción ( $PMAE$ ) para el conjunto de series predichas por el predictor. Nuevamente para realizar una comparación equitativa de este parámetro se asigno el valor de 1 para cada una de las series no predichas por el predictor. La expresión correspondiente para este parámetro es:

$$PMAE = \frac{1}{P} \sum_{i=1}^P MAE_i + \frac{1}{N-P} \sum_{j=1}^{N-P} (1)^j \quad (9.21)$$

donde  $P$  corresponde al número de series predichas y  $N - P$  corresponde al número de series no predichas por el predictor. Este promedio condensa la información sobre la expansión del error local de las predicciones.

- La eficiencia de predicción ( $EP$ ) se define como el cociente del número de series predichas entre el número total de series evaluadas para cada predictor:

$$EP = \frac{P}{N} \quad (9.22)$$

donde  $P$  corresponde al número de series predichas y  $N$  corresponde al número total de series evaluadas con los predictores. Este parámetro mide la eficiencia para predecir con relación al conjunto de series de tiempo evaluadas.

- La suma de los coeficientes de dificultad de predicción para las series predichas por el predictor:

$$SCDP_2 = \sum_{i=1}^P CDP_2(i) \quad (9.23)$$

en este caso se seleccionó la métrica  $CDP_2$  la cual contiene la *mayor cantidad de información* sobre las series de tiempo, ya que utiliza a 14 parámetros para cuantificar la dificultad de predicción o equivalentemente la predictibilidad de las mismas, adicionalmente, a partir del análisis de las métricas de predictibilidad (ver la sección 9.3) resulto ser la mejor métrica para cuantificar esta característica de las series de tiempo junto con el parámetro CRLZ. El parámetro  $SCDP_2$  proporciona información sobre el nivel de dificultad de las predicciones exitosas realizadas por cada predictor.

Finalmente la expresión matemática para el coeficiente de capacidad de predicción es:

$$CCAP = \frac{SCDP_2 * EP}{PRMSE + PMAE} \quad (9.24)$$

donde el numerador premia al modelo de predicción mediante el producto de la eficiencia de predicción y la suma de los coeficientes de dificultad de predicción, y el denominador castiga al modelo mediante la suma de los errores de la predicción.

La Figura 9-18, presenta la tabla de resultados del cálculo para el coeficiente de capacidad de predicción y la Figura 9-19, la tabla de resultados correspondientes al coeficiente de capacidad de modelado.

Los coeficientes poseen un comportamiento de tipo *no lineal*, lo cuál es debido a su diseño, sin embargo de esta forma se expresa mejor la capacidad de las técnicas para predecir o modelar a las series, por ejemplo, en el caso de la técnica de modelado Sugimay (ver tabla en la Figura 9-19) se observa que a pesar de ser una de las técnicas que más series logra modelar, su coeficiente  $CCAM$  es menor al de otras técnicas que modelaron menor número de series, como la técnica FFNBP, lo anterior se debe a que el error resultante de las modelaciones con la técnica Sugimay penaliza su capacidad de modelado considerablemente.

Técnica de Predicción	Coficiente de Capacidad de Predicción (CCAP)	Número de Series Predichas
FIRNet	1.920439941	9
PNNGauss	3.648760065	12
PNNReciprocal	1.521040683	9
MLFFN	1.490337069	10
MySVM Multipunto (kernel dot)	0.125357341	3
MySVM Multipunto (kernel radial)	3.414282627	13
Polynomp	3.895233526	11
Nstep	8.339671318	16
K-Nearest-Neighbours	10.94043514	17
ARIMA	1.214445183	7

Figura 9-18: Tabla de resultados de CCAP

Técnica de Modelado	Coficiente de Capacidad de Modelado (CCAM)	Número de Series Modeladas
FFNBP	9.44867589	16
RBFNBP	6.2110583	14
ANFIS	19.3456088	20
RNAP	44.2074516	23
Predict	1.50673888	9
Polynom	0.53189946	5
RBF	0.71278845	6
Sugimay	9.1881005	22
ARMA	16.6776503	19

Figura 9-19: Tabla de resultados del CCAM

## 9.5 Discusión

En este capítulo se ha presentado el desarrollo y evaluación de tres métricas que miden la predictibilidad (o equivalentemente la dificultad de predicción) de series de tiempo. *El comportamiento de este conjunto de métricas es de carácter no lineal*, reflejo del tipo de relaciones implícitas existentes entre los parámetros que caracterizan a las series de tiempo y la predictibilidad de las mismas.

Al utilizar en la construcción de las métricas de predictibilidad *un conjunto de parámetros que caracterizan diferentes aspectos de la dinámica de las series de tiempo*, fue posible introducir información global sobre las series de tiempo, a diferencia de métricas que utilizan un solo parámetro (por ejemplo, el Exponente de Lyapunov) y que en consecuencia miden la predictibilidad utilizando información parcial sobre las series de tiempo.

Las métricas de predictibilidad también muestran que el tipo de comportamiento dinámico de las series: *periódico o regular, cuasi periódico, caótico, complejo y aleatorio o estocástico*; no está relacionado de forma directa (o *lineal*) con la predictibilidad de las series de tiempo.

Del conjunto de métricas desarrolladas la denominada *Coficiente de Dificultad de Predicción 2 (CDP<sub>2</sub>)* presenta *la mejor correlación* con respecto a las variables de referencia, así como también, con respecto al parámetro *Complejidad Relativa LZ (CRLZ)* que presentó *una correlación ligeramente mayor* con respecto de las variables de referencia, para todo el conjunto de métricas evaluadas.

También se presentó el desarrollo de un coeficiente de predicción (o modelado), que permite en forma cuantitativa evaluar la capacidad de los modelos para predecir y modelar a las series de tiempo. Los coeficientes de capacidad de predicción (o modelado) presentan un *comportamiento no lineal*, reflejo no solo de la estructura del coeficiente sino también, del tipo de relaciones implícitas entre las características de las series y los modelos para predicción y modelado.

## Capítulo 10

# Relaciones de la Predictibilidad de Series de Tiempo

### 10.1 Introducción

En el capítulo anterior se desarrollaron y evaluaron diferentes métricas de la predictibilidad de series de tiempo, así como métricas sobre la capacidad de los modelos para la predicción y el modelado de series de tiempo. En este capítulo, a la información generada sobre la predictibilidad de las series de tiempo y sobre los modelos de predicción y modelado; se le aplican diferentes técnicas de análisis multivariado y reconocimiento de patrones (mapas autoorganizados jerárquicos, escalamiento multidimensional y correlación bivariada) para estudiar: *las relaciones que la predictibilidad presenta y las relaciones entre los parámetros de las series de tiempo y los modelos de predicción*. La sección 10.2, presenta la identificación de jerarquías de las series de tiempo, en base al análisis con mapas auto organizados jerárquicos (GHSOM) del conjunto de parámetros dinámicos utilizados para evaluar la predictibilidad. La sección 10.3, presenta el estudio de la relación entre el conjunto de parámetros dinámicos de las series de tiempo y diferentes representaciones de la estructura de las mismas. La sección 10.4, presenta el estudio de la relación entre las características de la arquitectura de los modelos y su capacidad de predicción o de modelado. La sección 10.5, presenta los resultados del estudio de las relaciones entre los parámetros de series de tiempo y los modelos de predicción y de modelado, mediante el uso de mapas auto organizados jerárquicos y el análisis de correlación bivariada.

Finalmente, la sección 10.6 presenta la discusión de los resultados de este capítulo.

## 10.2 Agrupamiento de Series de Tiempo en base a Parámetros de Predictibilidad

En la Figura 10-1, el mapa GHSOM muestra la jerarquía de las series de tiempo en base al conjunto de parámetros formado por: Auto Correlación ( $ACO$ ), Exponente de Hurst ( $EH$ ), Frecuencia Dominante ( $FD$ ), Exponente de Lyapunov ( $EL$ ), Dimensión de Correlación ( $DCO$ ), Dimensión de Capacidad ( $DCA$ ), Dimensión Fractal ( $DF$ ), Dimensión Embebida ( $DE$ ), Entropía Espacio Temporal ( $EET$ ), Porcentaje de Recurrencia ( $REC$ ), Porcentaje de Determinismo ( $DET$ ), Entropía de Información de Shannon ( $ES$ ), Información Mutua Promedio ( $IMP$ ), Complejidad Relativa LZ ( $CRLZ$ ) y el Número de Reglas de Producción ( $RP$ ). Este conjunto de parámetros dinámicos fueron utilizados para la construcción de las diferentes propuestas de métricas de predictibilidad presentadas en el capítulo anterior.

Para entender mejor el mapa GHSOM anterior, la Figura 10-2, muestra el árbol de jerarquías correspondiente a dicho mapa. Se observa que el conjunto de series de tiempo se desglosa en cinco niveles, a partir del nivel dos se forman cuatro grupos o familias de series de tiempo las cuales a su vez se subdividen hasta alcanzar en tres casos el nivel cuatro de profundidad. En el grupo uno, se encuentran las series HIV DNA y Tent, en el grupo dos se agrupan las series: Sine, Vanderpol, Dow Jones, Human DNA, Brownian Motion, Qperiodic2, Rossler y Qperiodic3; en el grupo tres las series D1, El niño, Mackey-Glass, Lorenz, Lovaina, A1, Laser y Kobe; finalmente en el grupo cuatro se tiene a las series Henon, Ikeda, Star, Primos, Plasma, SP500, ASCII TXT, Logistic, EEG, Cantor y White Noise. Las series agrupadas muestran *la jerarquía de similitud entre las series dentro de cada grupo o cluster*, al agruparse nuevamente y formar niveles de mayor profundidad dentro del árbol. A cada serie se le agregó el valor de la predictibilidad ( $1/CDP_2$ ), los grupos uno y dos corresponden a las series con *predictibilidad alta*, el grupo tres a las series con *predictibilidad media* y por último el grupo cuatro corresponde a las series con *predictibilidad baja*.

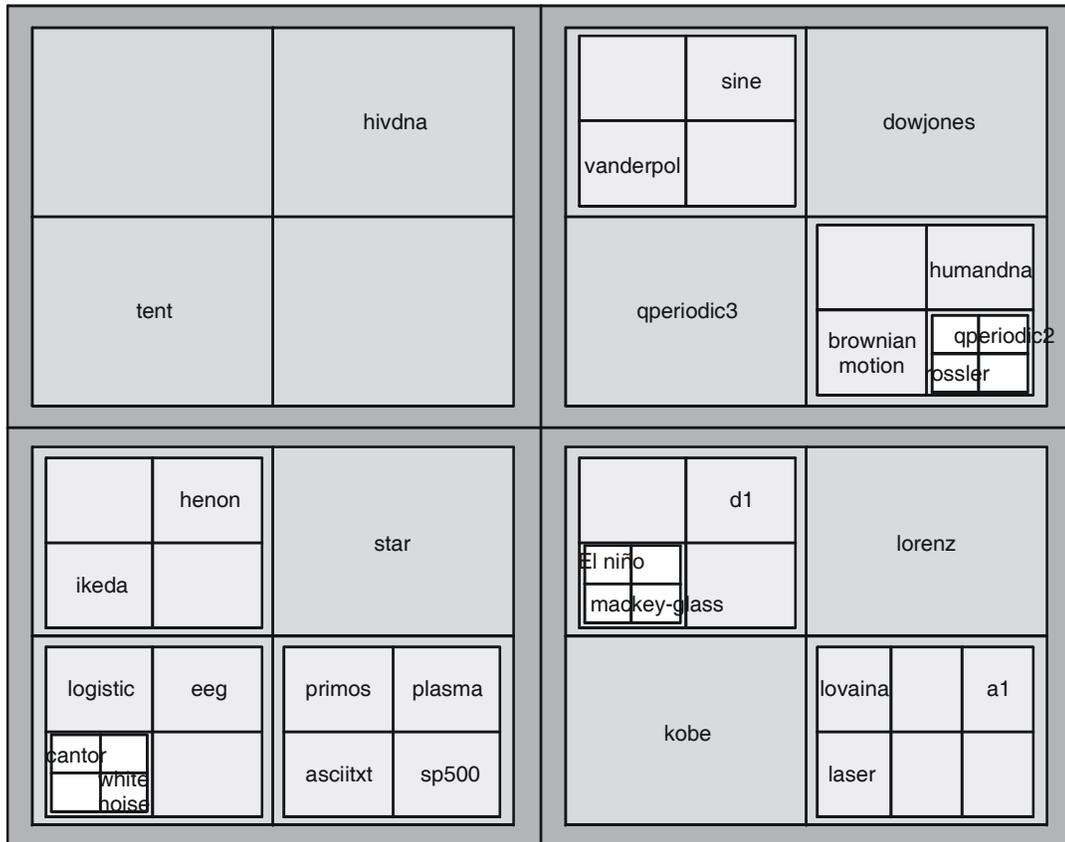


Figura 10-1: Mapa Auto Organizado Jerárquico para las series de tiempo

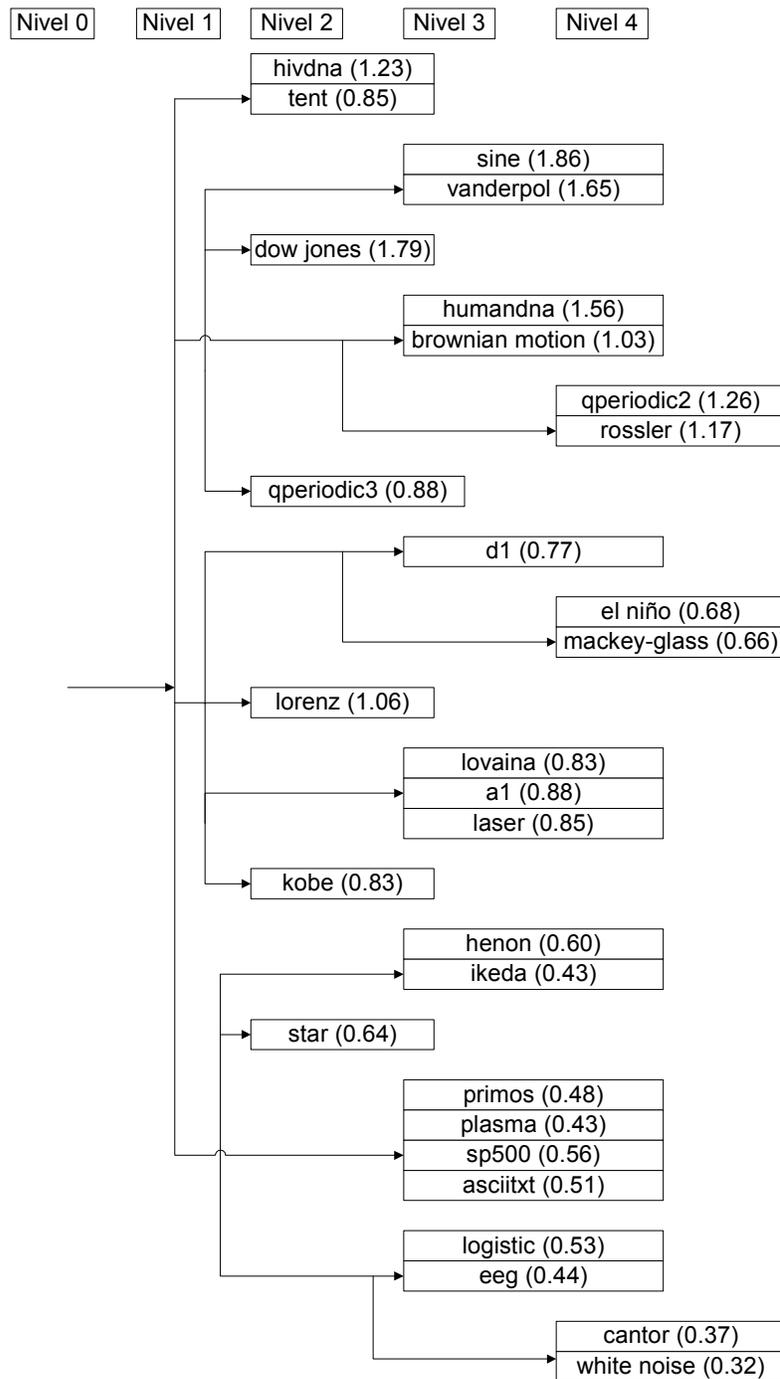


Figura 10-2: Árbol de jerarquías del agrupamiento de las series de tiempo con el Mapa Auto Organizado

### 10.3 Relación Predictibilidad-Estructura de Series de Tiempo

El análisis con escalamiento multidimensional (MDS) se utilizó para estudiar la agrupación de las series de tiempo en base a los 15 parámetros que caracterizan a cada una de ellas. La agrupación de las series no necesariamente coincide con la encontrada con el mapa auto organizado jerárquico, ya que en este caso no se obtienen agrupaciones (*clusters*) con una jerarquía de similitud. Sin embargo, si se compara el árbol de jerarquías (ver la Figura 10-2) con los resultados del escalamiento multidimensional *se observa que ambos presentan agrupamientos similares de las series de tiempo*, lo cual corrobora por dos métodos diferentes de agrupamiento que los grupos de series de tiempo se agrupan por su predictibilidad. La Figura 10-3 muestra los agrupamientos obtenidos, *se identificaron 5 agrupamientos de series de tiempo*. Si se considera el valor de la predictibilidad de las series de tiempo (etiquetas en paréntesis) se observa que *la dimensión correspondiente al eje horizontal corresponde a la predictibilidad de las series de tiempo*, aunque no es estrictamente paralela a dicho eje.

Al mapa anterior se le agregaron las gráficas de las series de tiempo (ver la Figura 10-4), esto con el objetivo de visualizar la estructura de las series, para buscar si existe una relación entre dicha estructura y la agrupación de las series en base a sus parámetros dinámicos, con los cuales se calcula la predictibilidad. Intuitivamente, *se espera que exista dicha relación*, ya que varios de los parámetros calculados se basan en explotar la topología del atractor reconstruido para una serie de tiempo y de esta forma poder caracterizarla, a grosso modo se observa que los grupos (*clusters*) de series de tiempo, poseen cierta *similitud en su estructura*, sin embargo no es del todo claro con esta forma de representación de las series. Con el objetivo de identificar con mayor claridad la posible relación entre la estructura de las series de tiempo y sus agrupamientos en base a sus parámetros dinámicos, se utilizaron los mapas de recurrencia (ver la subsección 4.5.3) los cuales son otra *representación de la estructura de las series de tiempo* que permite obtener patrones característicos mucho más definidos y fáciles de estudiar.

La Figura 10-5, muestra los mapas de recurrencia correspondientes a las series de tiempo, a continuación se describen los resultados producto de las *observaciones experimentales* sobre dichos mapas, para identificar los patrones básicos se compararon los mapas de recurrencia de cada agrupamiento de series de tiempo a diferentes niveles de acercamiento (*zoom*), un ejemplo de la identificación de los patrones se ilustra en la Figura 10-6, una vez identificados los patrones

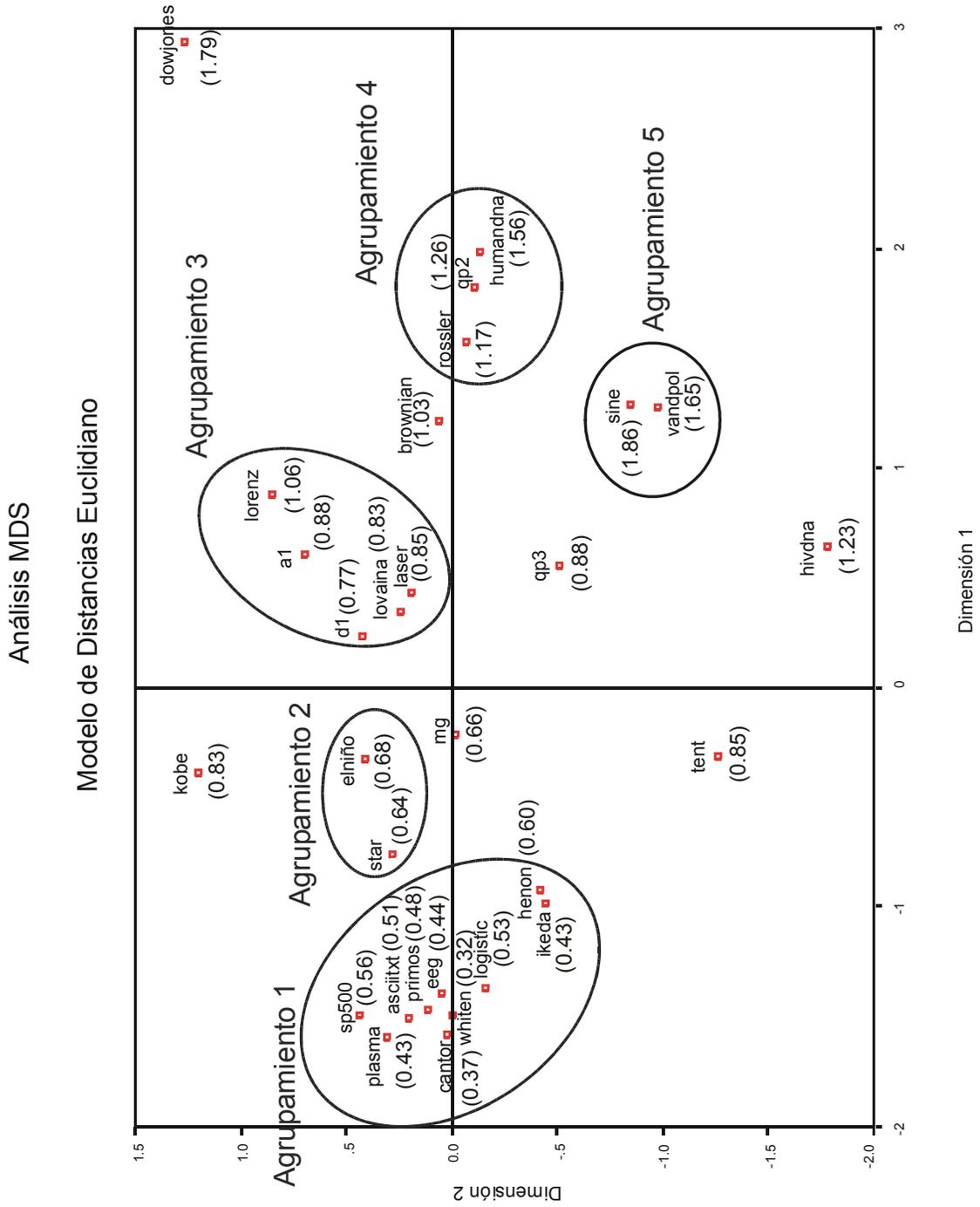


Figura 10-3: Mapa de escalamiento multidimensional de las series de tiempo

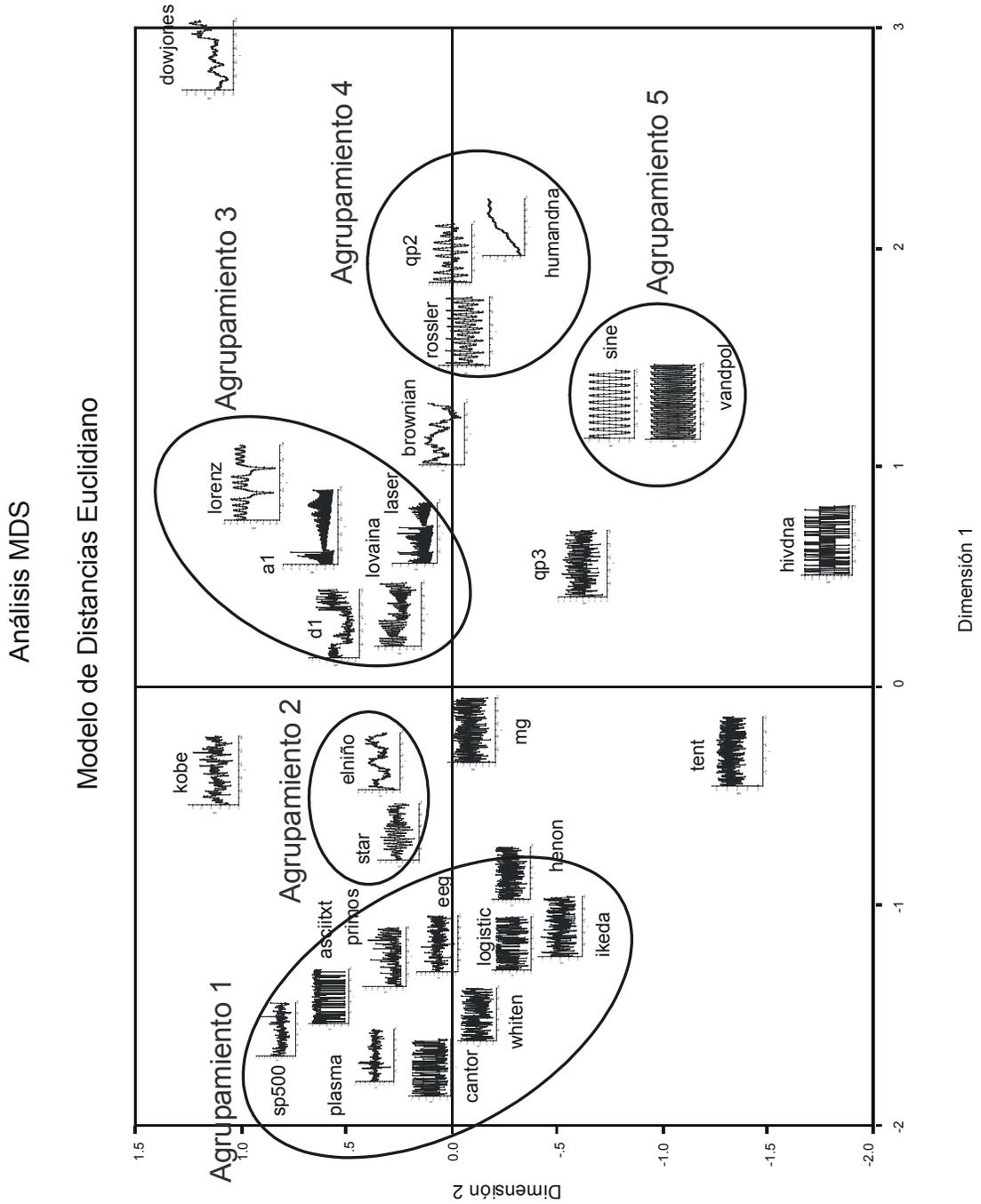


Figura 10-4: Mapa de escalamiento multidimensional con las gráficas de las series de tiempo

base de cada agrupamiento, se compararon estos patrones entre todos los grupos de series de tiempo para comprobar que efectivamente fuesen diferentes entre sí.

En las Figuras 10-7, 10-8, 10-9, 10-10, 10-11, 10-12 y 10-13; se muestran *los patrones de mapas de recurrencia que se identificaron como básicos dentro de las series de tiempo*, cada patrón tiene asociado un agrupamiento de series de tiempo las cuales se observa poseen *una predictibilidad similar*. Existen varias formas en que se relacionan las series agrupadas con los patrones encontrados, se puede presentar el caso en que las series de un mismo grupo poseen *un patrón único* como en los patrones P2 y P5 (Figuras 10-8 y 10-10), se puede presentar *más de un patrón básico* para un grupo de series como en los patrones P1a y P1b (Figura 10-7), se puede presentar que dentro de un grupo de series *un subconjunto de ellas tenga más de un patrón básico* como ocurre con los patrones P3a y P3b (Figuras 10-8 y 10-9), y se puede presentar que un agrupamiento de series posea *una predictibilidad similar pero con diferentes patrones básicos* para las series del grupo como en los patrones P4a y P4b (Figuras 10-9 y 10-10), las relaciones anteriores entre predictibilidad y estructura se resumen en la tabla 10.1, donde la primera columna corresponde a los agrupamientos de las series de tiempo (ver Figura 10-5).

Agrupamiento	Relaciones de Predictibilidad-Estructura
1	Series tienen una mezcla de dos patrones básicos
2 y 5	Series tienen solamente un patrón básico
3	Series tienen uno o más patrones básicos
4	Series con diferentes patrones básicos

(10.1)

En la búsqueda de *otras formas de representación* de las series de tiempo que puedan correlacionarse con la predictibilidad y ser de utilidad en la clasificación de las series de tiempo, se analizaron los resultados de aplicar el análisis espectral singular (SSA), el cual descompone a la serie de tiempo en un conjunto de series cuya combinación lineal genera la serie original. Estas componentes representan *diferentes dinámicas* dentro de la serie de tiempo (tendencias, oscilatorias y ruido), a las componentes se les aplicó el análisis de componentes principales (PCA) para obtener el *comportamiento estadístico* de los datos que las forman.

Para estudiar la información del análisis PCA de una forma visual, se aplicó el análisis con escalamiento multidimensional (MDS). La Figura 10-14 muestra el mapa MDS, a cada serie

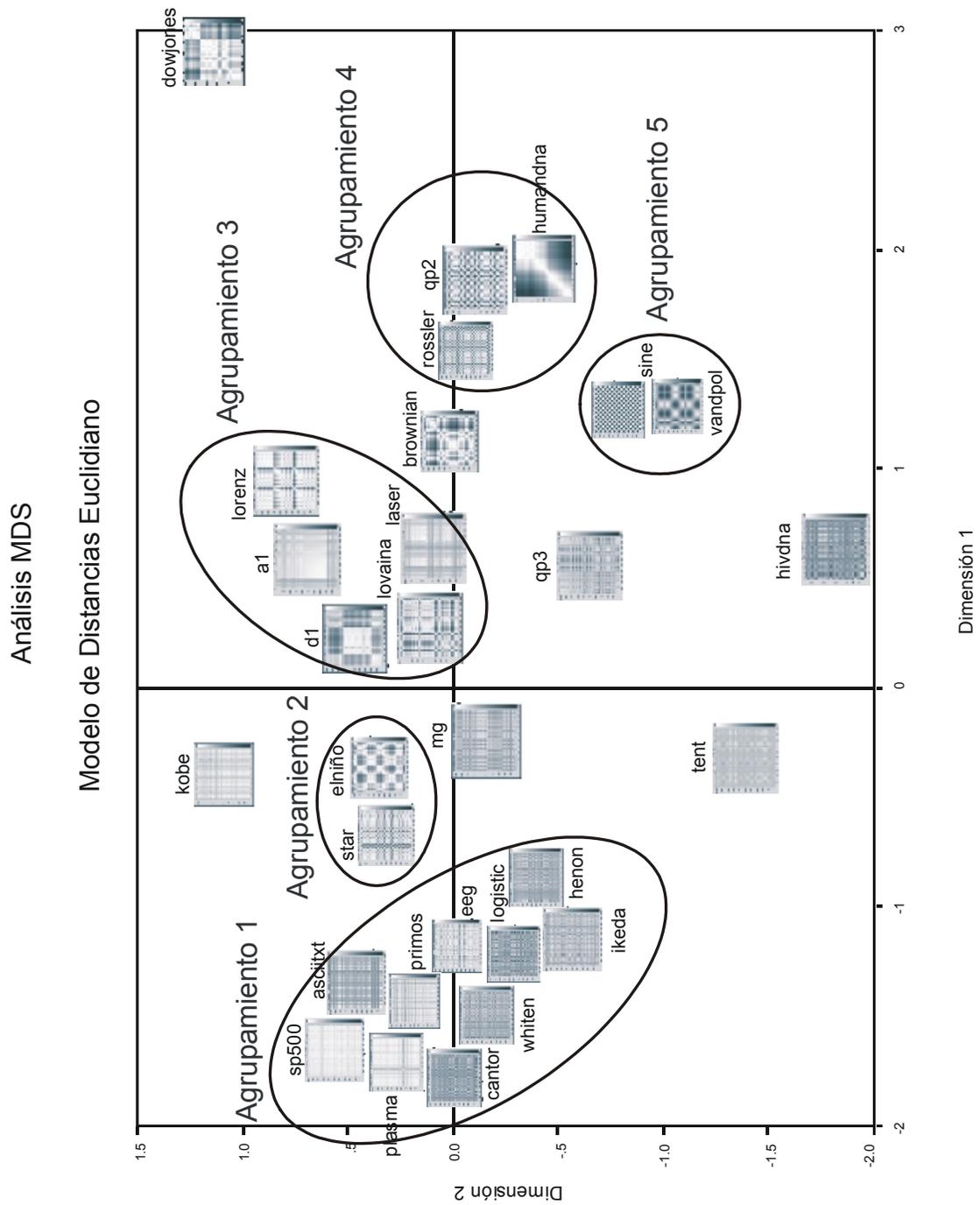


Figura 10-5: Mapa de escalamiento multidimensional con los mapas de recurrencia de las series de tiempo

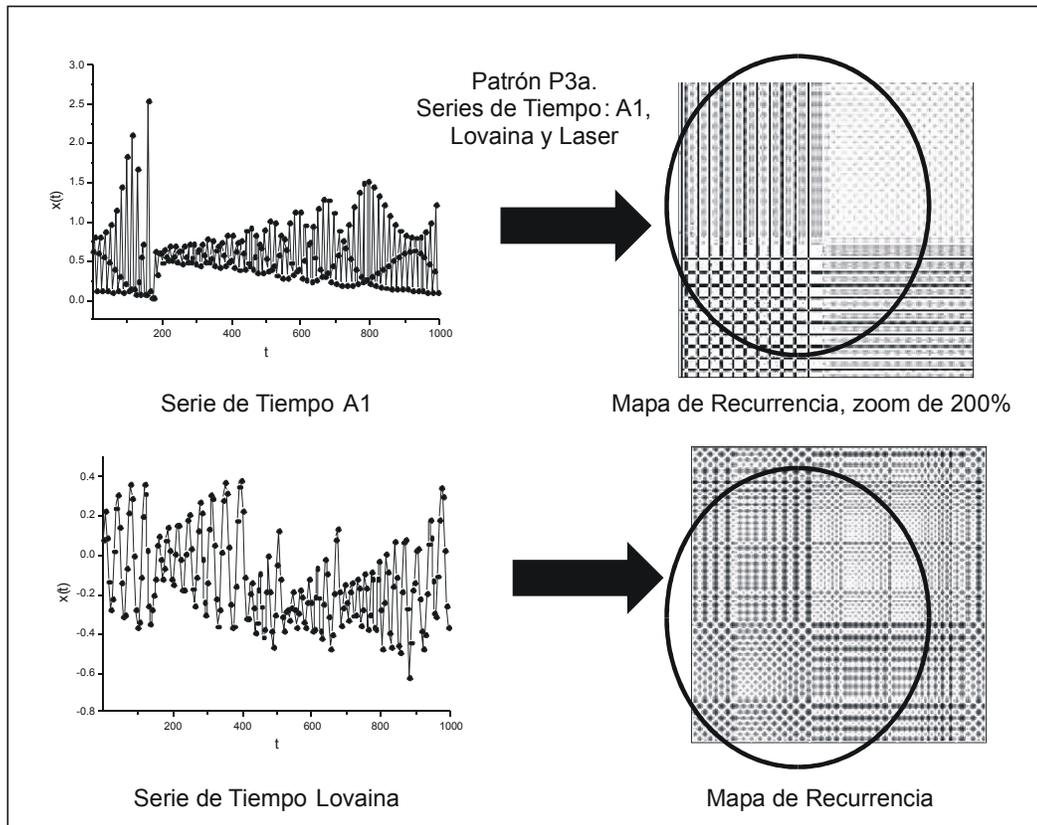


Figura 10-6: Ejemplo de la identificación de patrones básicos

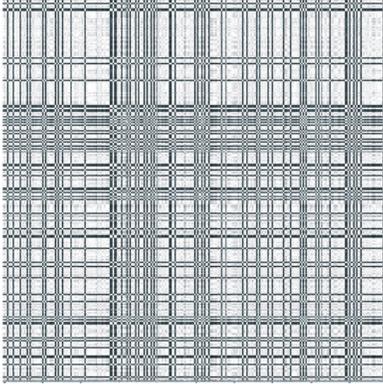
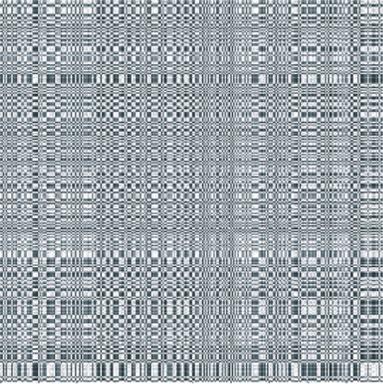
Clase de Patrón de Recurrencia	Mapa de Recurrencia característico	Series de Tiempo que poseen este Patrón y su Predictibilidad (1/CDP2)
<b>P1a</b>		Plasma (0.43) SP500 (0.56) ASCIITXT (0.51) Primos (0.48) EEG (0.44) White Noise (0.32) Cantor (0.37) Logistic (0.53) Henon (0.60) Ikeda (0.43)
<b>P1b</b>		Plasma (0.43) SP500 (0.56) ASCIITXT (0.51) Primos (0.48) EEG (0.44) White Noise (0.32) Cantor (0.37) Logistic (0.53) Henon (0.60) Ikeda (0.43)

Figura 10-7: Patrones básicos de mapas de recurrencia

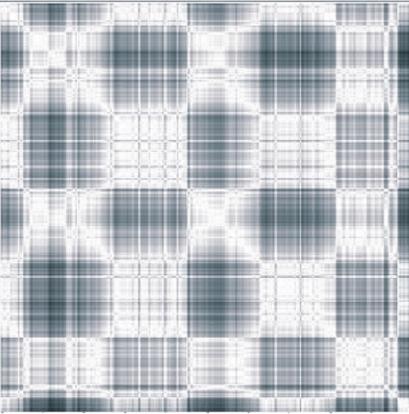
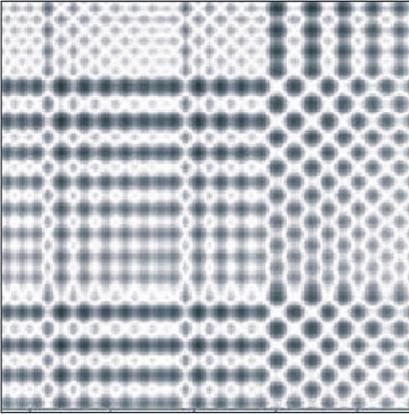
Clase de Patrón de Recurrencia	Mapa de Recurrencia característico	Series de Tiempo que poseen este Patrón y su Predictibilidad (1/CDP2)
P2		Star (0.64) El niño (0.68)
P3a		A1 (0.88) Lovaina (0.83) Laser (0.85)

Figura 10-8: Patrones básicos de mapas de recurrencia

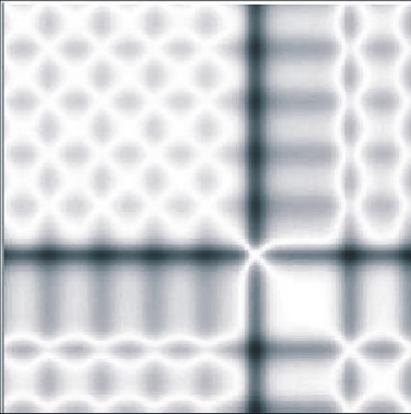
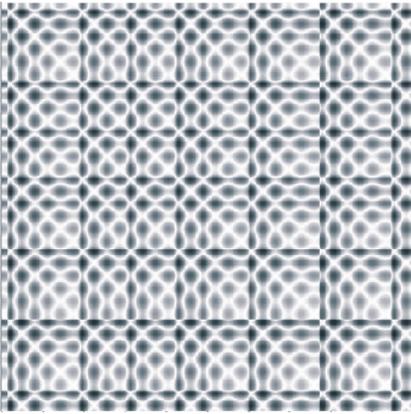
Clase de Patrón de Recurrencia	Mapa de Recurrencia característico	Series de Tiempo que poseen este Patrón y su Predictibilidad (1/CDP2)
<b>P3b</b>		<p style="text-align: center;"> A1 (0.88)  Lovaina (0.83)  Laser (0.85)  Lorenz (1.06)  D1 (0.77) </p>
<b>P4a</b>		<p style="text-align: center;"> Qperiodic2 (1.26)  Rossler (1.17) </p>

Figura 10-9: Patrones básicos de mapas de recurrencia

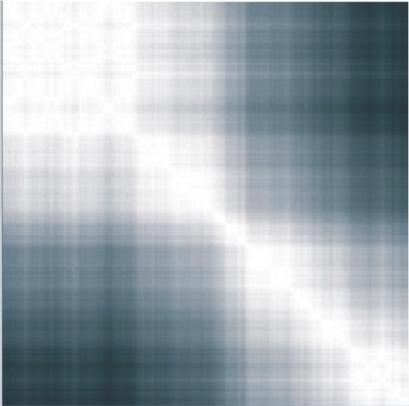
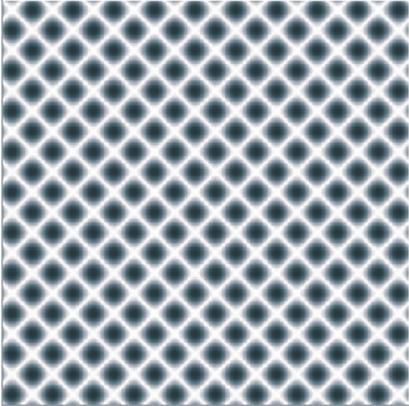
Clase de Patrón de Recurrencia	Mapa de Recurrencia característico	Series de Tiempo que poseen este Patrón y su Predictibilidad (1/CDP2)
P4b		Human DNA (1.56)
P5		Sine (1.86) Vanderpol (1.65)

Figura 10-10: Patrones básicos de mapas de recurrencia

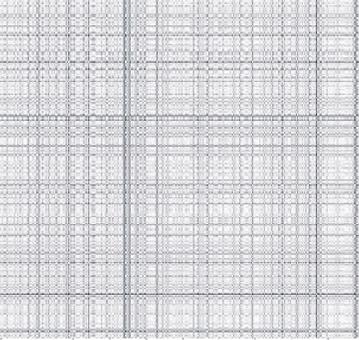
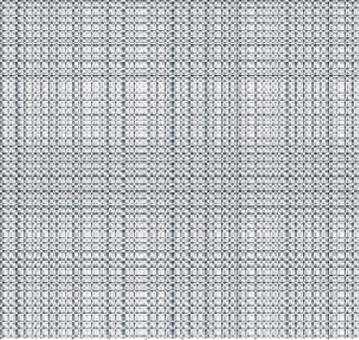
Clase de Patrón de Recurrencia	Mapa de Recurrencia característico	Series de Tiempo que poseen este Patrón y su Predictibilidad (1/CDP2)
P6		Kobe (0.83)
P7		Mackey-Glass (0.66)

Figura 10-11: Patrones básicos de mapas de recurrencia

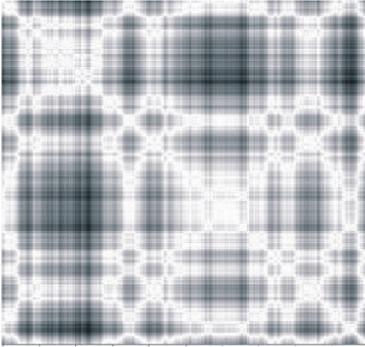
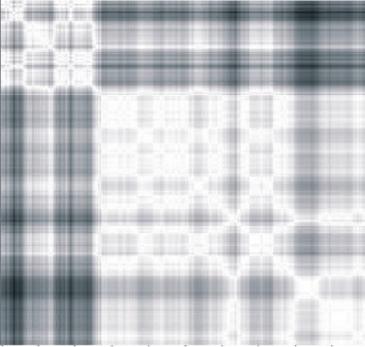
Clase de Patrón de Recurrencia	Mapa de Recurrencia característico	Series de Tiempo que poseen este Patrón y su Predictibilidad (1/CDP2)
P8		Brownian Motion (1.03)
P9		Dow Jones (1.79)

Figura 10-12: Patrones básicos de mapas de recurrencia

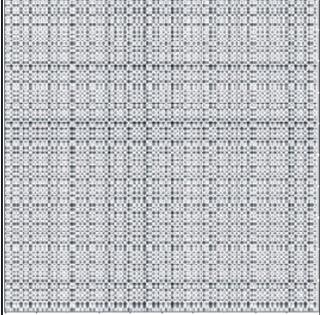
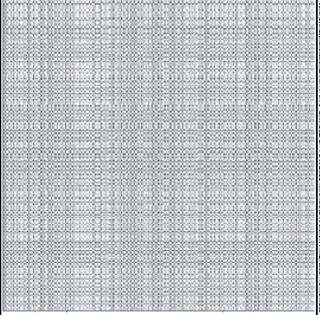
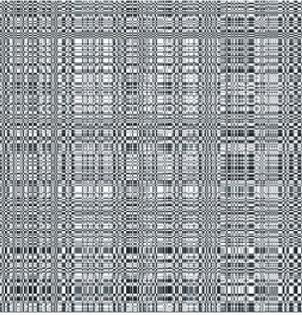
Clase de Patrón de Recurrencia	Mapa de Recurrencia característico	Series de Tiempo que poseen este Patrón y su Predictibilidad (1/CDP2)
P10		Qperiodic3 (0.88)
P11		Tent (0.85)
P12		HIV DNA (1.23)

Figura 10-13: Patrones básicos de mapas de recurrencia

se le asignó posteriormente su predictibilidad correspondiente (valores entre paréntesis). Se identifican *tres grandes agrupaciones* las cuales corresponden a las series de tiempo con *valores de predictibilidad alta, media y baja*. El mapa es de dos dimensiones, donde *la dimensión correspondiente al eje horizontal se asocia con la predictibilidad de las series de tiempo*. Este resultado corrobora los resultados encontrados con el estudio de los mapas de recurrencia, *que las características sobre la estructura de las series de tiempo están relacionadas con la predictibilidad de las mismas*.

## 10.4 Relación entre las Características de los Modelos con su Capacidad de Predicción o de Modelado

Los modelos de predicción se caracterizaron en base a su origen, arquitectura, alcance, espacio donde actúa sobre la serie, equivalencia con otro modelo, método de aprendizaje, su dependencia de parámetros, el número de series predichas y su capacidad de predicción. En la Figura 10-15, se presenta la tabla con la información sobre los modelos de predicción, esta tabla combina información obtenida de la evaluación de las técnicas de predicción (ver capítulo 8) e información derivada de la descripción de los modelos (ver capítulo 5). A partir del análisis de esta tabla se derivan las siguientes afirmaciones: *La capacidad de predicción más alta corresponde a modelos cuyo alcance es local y que actúan en el espacio fase. Los modelos que son de alcance global y que actúan en diversos espacios no logran alcanzar la capacidad de predicción de los modelos locales*.

En relación con los modelos para el modelado de series de tiempo, la Figura 10-16 muestra la tabla correspondiente a estos modelos, y que fue construida en forma similar a la correspondiente a los modelos de predicción, se observa que en este caso *los modelos con alcance global que actúan en el espacio real son sumamente eficientes en modelar las series de tiempo*, y por otra parte, *los modelos que actúan en el espacio fase poseen una menor capacidad de modelado*. Para ambos tipos de modelos (predicción y modelado), las dos características que diferencian a los modelos con mayor capacidad para predecir o modelar del resto de modelos evaluados son: *el alcance del modelo y el espacio donde opera el modelo*.

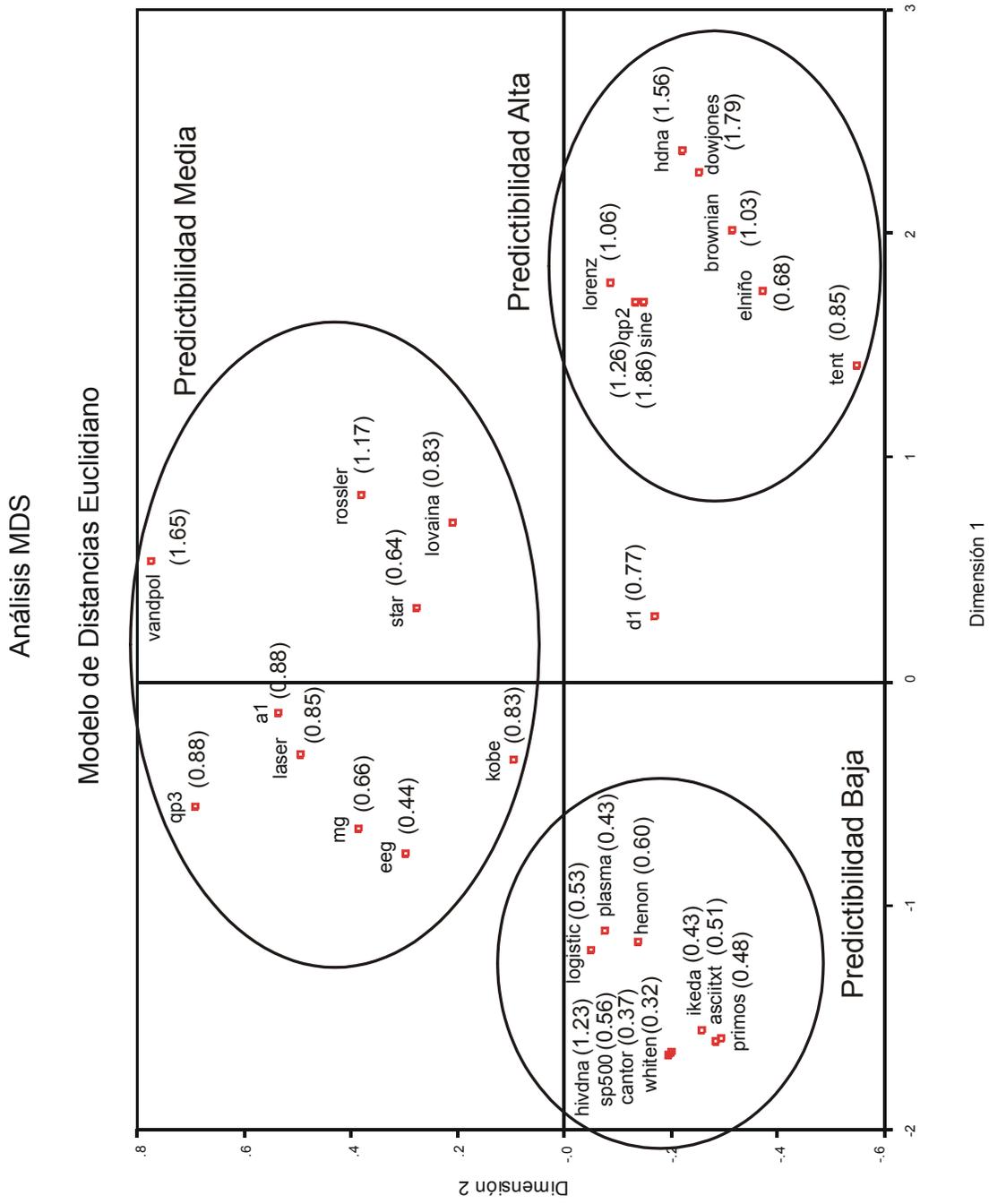


Figura 10-14: Mapa de escalamiento multidimensional del análisis PCA de las series de tiempo

Técnica de Predicción	Origen de Técnica	Arquitectura del Modelo	Alcance de modelo	Espacio donde actúa	Equivalencia	Método de Aprendizaje	Dependencia de parámetros	Número de Series Predichas	Coefficiente de Capacidad de Predicción (CCAP)
FIRNet	Inteligencia Artificial	Red Neuronal.	Global	Espacio Real	MA	Retropropagación temporal minimizando error cuadrático.	Parámétrico	9	1.920439941
PNN Gauss	Inteligencia Artificial	Red Neuronal.	Global	Espacio Real	Estadístico FDP	Método de Parzen aproximación asintótica de la FDP.	Parámétrico	12	3.648760065
PNN Reciproca	Inteligencia Artificial	Red Neuronal.	Global	Espacio Real	Estadístico FDP	Método de Parzen aproximación asintótica de la FDP.	Parámétrico	9	1.521040683
MLFFN	Inteligencia Artificial	Red Neuronal.	Global	Espacio Real	AR	Regresión con refinamiento con gradiente conjugado minimizando el error cuadrático.	Parámétrico	10	1.490337069
MySVM Multipunto (kernel dot)	Inteligencia Artificial	Busqueda de una función de decisión que minimice el riesgo esperado.	Global	Espacio de características	Estadístico FDP	Método de optimización de tipo estadístico, minimizando una función de riesgo.	Parámétrico	3	0.125357341
MySVM Multipunto (kernel radial)	Inteligencia Artificial	Busqueda de una función de decisión que minimice el riesgo esperado.	Global	Espacio de características	Estadístico FDP	Método de optimización de tipo estadístico, minimizando una función de riesgo.	Parámétrico	13	3.414282627
Polynomp	Teoría de Sistemas Dinámicos No Lineales	Modelo que ajusta un polinomio a los puntos del sistema.	Global	Espacio Fase Reconstruido	Ajuste de una Función	Calculo de los coeficientes del polinomio en base a minimizar la varianza del conjunto de puntos.	Parámétrico	11	3.895233526
Nstep	Teoría de Sistemas Dinámicos No Lineales	Modelo lineal, aproximación local con expansión de Taylor.	Local	Espacio Fase Reconstruido	Ajuste de una Función	Calculo de los coeficientes de ecuaciones lineales locales en base a minimizar la varianza del conjunto de puntos.	Parámétrico	16	8.339671318
K-Nearest-Neighbours	Teoría de Sistemas Dinámicos No Lineales	Calculo del siguiente estado con el promedio simple o pesado de los k estados vecinos de trayectorias pasadas.	Local	Espacio Fase Reconstruido	Promedio de estados del sistema similares	Busqueda de k-vecinos cercanos.	No Parámétrico	17	10.94043514
ARIMA	Estadística	Combinación de Modelo Autoregresivo y Modelo de Promedio Móvil para obtener un modelo de combinaciones lineales de valores previos y aleatorios.	Global	Espacio de Derivadas	ARMA	Busqueda de los pesos de las combinaciones lineales minimizando la varianza.	Parámétrico	7	1.214445183

Figure 10-15: Tabla de características de modelos de predicción

Técnica de Modelado	Origen de Técnica	Modelo simple o híbrido	Arquitectura de Modelo	Alcance de modelo	Espacio donde actúa	Equivalencia	Método de Aprendizaje	Dependencia de parámetros	Número de Series Modeladas	Coefficiente de Capacidad de Modelado (CCAM)
FFNBP	Inteligencia Artificial	Simple	Red Neuronal.	Global	Espacio Real	AR	Aprendizaje con retropropagación, minimizando el error cuadrático.	Paramétrico	16	9.448675887
RBFNBP	Inteligencia Artificial	Simple	Red Neuronal.	Global	Espacio Real	Ajuste de una función	Aprendizaje en dos etapas: determinación de centros y determinación de pesos y umbrales en capa de salida con retropropagación.	Paramétrico	14	6.211058301
ANFIS	Inteligencia Artificial	Híbrido	Red Neuronal.	Global	Espacio Real	RBFN	Aprendizaje con gradiente descendente para construir los parámetros de premisa y con estimación de mínimos cuadrados para construir los parámetros de consecuencia.	Paramétrico	20	19.34560877
RNAP	Inteligencia Artificial	Híbrido	Red Neuronal.	Global	Espacio Real	Ajuste de una función	Optimización por medio de un algoritmo genético para obtener la mejor arquitectura de la red (minimizando la función de error para obtener el vector de pesos óptimo).	Paramétrico	23	44.20745158
Predict	Teoría de Sistemas Dinámicos No Lineales	Simple	Modelo lineal, aproximación local con el orden cero (función constante).	Local	Espacio Fase Reconstituido	Ajuste de una Función	No busca minimizar error, sin aprendizaje.	No Paramétrico	9	1.506738879
Polynom	Teoría de Sistemas Dinámicos No Lineales	Simple	Modelo que ajusta un polinomio.	Global	Espacio Fase Reconstituido	Ajuste de una Función	Se determinan los coeficientes con método de algebra lineal, minimizando la varianza.	Paramétrico	5	0.531899455
rbf	Teoría de Sistemas Dinámicos No Lineales	Simple	Modelo que ajusta una función de base radial.	Global	Espacio Fase Reconstituido	Ajuste de una Función	Se determina los coeficientes con método de mínimos cuadrados, minimizando la varianza.	Paramétrico	6	0.712768452
Sugimay	Teoría de Sistemas Dinámicos No Lineales	Simple	Busqueda de los vecinos cercanos, proyecta el dominio del vecindario (simplex) en su rango. Promedio pesado con pesos de tipo exponencial.	Local	Espacio Fase Reconstituido	Promedio Pesado de elementos con pesos de tipo exponencial	Busca vecinos cercanos.	No Paramétrico	22	9.188100504
ARMA	Estadística	Simple	Combinación de Modelo Autorregresivo y Modelo de Promedio Móvil para obtener un Modelo de combinaciones lineales de valores previos y aleatorios.	Global	Espacio Real	ARMA	Busqueda de los pesos de las combinaciones lineales minimizando la varianza (error cuadrático).	Paramétrico	19	16.67765029

Figure 10-16: Tabla de características de modelos para modelado

## 10.5 Relación Parámetros de Predictibilidad-Modelos de Predicción y Modelado

En esta sección se investiga la relación de los modelos de predicción y modelado con los diferentes parámetros dinámicos que caracterizan la predictibilidad de las series de tiempo, para realizar el estudio anterior se utiliza la correlación bivariada y los mapas auto organizados jerárquicos (GHSOM). El conjunto de parámetros consiste de 12 elementos que miden características de la dinámica de la serie de tiempo y con los cuales se construyeron anteriormente las métricas de predictibilidad, la Figura 10-17 muestra la tabla con los parámetros y el tipo de información que proporcionan sobre las series de tiempo.

Para estudiar la sensibilidad de los modelos con relación a los parámetros que caracterizan a las series de tiempo, se construyó una tabla de correlación bivariada. Para obtener los coeficientes de correlación para cada modelo se construyó una tabla de datos como la que se muestra en la Figura 10-18, donde las primeras doce columnas corresponden a parámetros de las series de tiempo y la columna número trece corresponde a cada técnica o modelo a analizar, en caso de que el modelo no hubiera predicho una serie se asignaba el valor 0 y en caso de una predicción exitosa el valor 1. Con tablas como la mostrada se realizó el cálculo de la correlación bivariada para obtener los coeficientes de Pearson correspondientes a cada modelo de predicción.

En la Figura 10-19, se presenta la tabla de correlación bivariada entre los modelos de predicción y los parámetros dinámicos usados para caracterizar la predictibilidad de las series de tiempo. Debido a la gran cantidad de información presente en la tabla, se normalizaron los resultados de la correlación de la siguiente forma, para cada parámetro dinámico se identificó el mayor valor (en valor absoluto) de correlación encontrado y con dicho valor se normalizaron todos los valores de correlación encontrados para dicho parámetro. Una vez normalizada la tabla de correlación se construyó una tabla simbólica para discretizar los resultados de forma que sea más fácil el análisis de los resultados de la correlación. La tabla 10.2, muestra la escala de valores para la discretización.

Parámetros	Alcance de Parámetro	Tipo de Información que proporciona
Exponente de Lyapunov	Global	Horizonte de Predicción
Dimensión de Correlación	Local	Correlación espacial local
Dimensión de Capacidad	Local	Grado de autosimilitud del sistema
Dimensión Fractal	Global	Dimensión promedio en una vecindad de tamaño epsilon sobre el atractor
Dimensión Embebida	Global	Estimado del número grados de libertad del sistema
%Entropía Espacio Temporal	Global/Local	Grado de no correlación espacial
% Recurrencia	Global/Local	Periodicidad y estructuras definidas
% Determinismo	Global/Local	Determinismo en base a patrón de recurrencia
Entropía de Shannon	Global/Local	Cantidad de Información que puede ser extraída al medir el estado del sistema
Información Mutua Promedio	Global	Información mutua contenida en una variable para dos instantes de tiempo diferentes
Complejidad LZ Relativa	Global/Local	Estructuras y jerarquía de cadenas de datos
Reglas de Producción	Global/Local	Número de reglas gramaticales para modelar la serie (complejidad computacional)

Figura 10-17: Tabla de características de parámetros dinámicos de las series de tiempo

Tabla para calcular la correlación de Pearson entre los parámetros de las series de tiempo y la técnica de predicción K-Nearest-Neighbours.

Exponente de Lyapunov	Dimensión de Correlación	Dimensión de Capacidad	Dimensión Fractal	Dimensión Embebida	Entropía Espacio Temporal (%)	Recurrencia (%)	Determinismo (%)	Entropía de Información (Shannon)	Información Mutua Promedio	Complejidad Relativa LZ	Número de Reglas de Producción	Predicción Exitosa = 1, Fracaso = 0
0.517	0.228	0.246	0.85	2	0	19.556	62.19	3.7118	3	0.05979471	27	1
1.864	0.984	0.989	1.14	2	0	7.826	66.355	4.981	7	0.07972627	28	1
0.925	0.923	0.679	1.1	9	0	8.148374	97.760539	2.445885	18	0.1494868	66	1
1.383	0.96	0.605	2.76	5	0	16.133	36.733	4.232	6	0.4384945	75	1
1.481	1.025	0.983	2.25	7	50	0.51034759	37.5178503	0.15914973	4	0.4783576	75	1
0.76	0.93	0.941	1.28	2	78	8.61304444	0.85112222	0.03522222	9	0.7175364	61	1
0.601	1.025	0.965	1.44	5	57	32.8839734	92.0699379	4.78937065	17	0.1793841	63	0
1.049	1.026	0.994	1.82	2	0	6.38852058	88.0434671	2.45897942	13	0.1694183	72	1
1.452	1.019	0.983	4.14	5	53	0.54849383	0	0	6	0.787297	71	1
2.301	0.991	0.997	1.67	2	51	9.43468507	0.63758282	0.0204499	17	0.6378102	69	1
5.728	0.658	0.661	0	5	84	2.193	0	0	1	1.056373	72	0
0.477	1.029	0.971	3.95	1	53	3.079	29.009	3.401	12	0.03986314	13	1
0.2234006	2.5	2.644	1.7	4	51	7.812	86.786	7.02	3	0.2890078	79	1
1.288	2.053	0.986	1.05	7	39	3.30067901	60.2251502	0.8521749	7	0.23213	92	0
0.949	2.096	0.961	1.94	9	47	10.716364	71.4658139	2.93549693	2	0.3786998	66	1
0.144	1.035	0.887	2.31	12	2	98.8534487	98.8812005	6.43519134	26	0.1993157	65	1
1.102	1.053	0.904	1.8	6	78	57.048522	44.3674678	3.74401017	2	0.787297	79	1
1.199	1.876	0.89	3.83	6	73	2.7157191	0	0	6	0.7773312	82	1
2.709	2.011	0.485	0.88	7	82	13.144	1.226	2.107	3	1.036442	80	0
2.322	1.637	0.971	2.59	9	65	0.32	44.212	3.526	8	0.2092815	83	0
0	5.017	0.953	0	8	0	0.004	31.331	1	4	0.8072285	9	0
0.322	1.037	0.983	0.55	12	0	24.915	98.244	7.28	20	0.04982892	51	1
1.0693	1.027	0.958	2.39	5	47	0.547	69.791	5.574	9	0.3089393	77	1
3.383	0.967	0.821	0.52	10	81	1.683	0.748	2.866	3	0.9766468	91	0
0.594	3.55	0.044	0.79	3	80	9.705	0.412	1.295	3	0.9766468	78	0
3.569	1.026	0.859	0.6	3	87	20.868	4.383	2.585	2	1.036442	82	0
3.289	1.144	0.963	3.39	7	53	16.438	11.137	3.102	13	0.5879813	90	0
2.043	1.103	0.971	2.78	5	0	11.7750441	66.2786765	2.74038235	23	0.1694183	85	0
1.606	2.086	0.983	6.09	10	79	0	0	0	1	1.066339	73	0

Figure 10-18: Ejemplo de conjunto de parámetros para series de tiempo predichas por la técnica K-Nearest-Neighbours

Tabla de la correlación de Pearson entre los parámetros de las series y las técnicas de predicción											
Parámetros de Series de Tiempo	Técnica de Predicción						Teoría de Sistemas Dinámicos No				Estadística
	Inteligencia Artificial			Múltiplo (kernel radial)			Polynomp	Nstep	K-Nearest-Neighbours	ARIMA	
	FIRNet	PNN Gauss	PNN Reciprocal	MLFFN	Múltiplo (kernel dot)	Múltiplo (kernel radial)					
Exponente de Lyapunov	-0.247	-0.331	-0.394	-0.328	-0.075	-0.418	-0.209	-0.314	-0.495	-0.247	
Dimensión de Correlación	-0.174	-0.22	-0.171	-0.243	-0.263	-0.298	-0.237	-0.346	-0.365	-0.146	
Dimensión de Capacidad	0.226	0.184	0.206	0.178	-0.248	0.183	0.19	0.245	0.211	-0.089	
Dimensión Fractal	-0.162	-0.032	0.036	-0.068	-0.083	-0.026	-0.049	-0.077	0.148	0.201	
Dimensión Embebida	-0.116	-0.031	-0.262	-0.181	-0.318	-0.39	-0.267	-0.109	-0.198	0.21	
Entropía Espacio Temporal (%)	-0.333	-0.574	-0.44	-0.492	-0.468	-0.252	0.125	-0.279	-0.376	-0.208	
Recurrencia (%)	0.027	0.174	0.3	-0.033	0.015	0.007	0.062	0.108	0.179	0.325	
Determinismo (%)	0.37	0.549	0.533	0.576	0.129	0.469	0.014	0.424	0.358	0.181	
Entropía de Información (Shannon)	0.362	0.506	0.529	0.553	0.247	0.318	0.093	0.502	0.257	0.288	
Información Mutua Promedio	0.022	0.327	0.13	0.225	-0.158	-0.062	-0.165	0.222	0.178	0.258	
Complejidad Relativa LZ	-0.413	-0.666	-0.572	-0.695	-0.3	-0.49	-0.117	-0.564	-0.446	-0.33	
Número de Reglas de Producción	-0.255	-0.373	-0.411	-0.394	-0.38	-0.306	-0.143	-0.268	-0.292	-0.291	

Figure 10-19: Tabla de correlación entre los parámetros dinámicos y los modelos de predicción

Tipo de Correlación	Rango de la Correlación
MUY ALTA	0.81 a 1.0
ALTA	0.61 a 0.809
MEDIA	0.41 a 0.609
BAJA	0.21 a 0.409
MUY BAJA	0.0 a 0.209

(10.2)

A partir de la normalización de los datos de la tabla de correlación y de su posterior discretización, se obtuvo la tabla que se muestra en la Figura 10-20. Los parámetros de mayor interés son aquellos que muestran correlaciones altas y muy altas para la mayoría de los modelos de predicción, para analizarlos con mayor facilidad, se construyó una tabla con solamente dichos parámetros la cual se muestra en la Figura 10-21.

La tabla en la Figura 10-21, muestra los parámetros que *más influyen en los modelos al predecir a una serie de tiempo*, estos parámetros miden a su vez características de la dinámica de las series de tiempo que el modelo trata de reproducir durante el proceso de predicción. Los parámetros son: *Exponente de Lyapunov, Dimensión de Correlación, Dimensión de Capacidad, Entropía Espacio-Temporal, Determinismo, Entropía de Información de Shannon, Complejidad Relativa LZ y Número de Reglas de Producción.*

Con respecto a los modelos para modelado, se realizó el mismo análisis de correlación, la Figura 10-22 muestra la tabla de correlación correspondiente a estos modelos.

Nuevamente, se discretizó la tabla de correlación para analizar la información de la misma, la Figura 10-23 muestra la tabla de correlación con los datos discretizados.

Para el caso de las técnicas de modelado (ver la tabla en la Figura 10-24), *los parámetros para los cuales los modelos presentan una mayor correlación son: Exponente de Lyapunov, Dimensión de Correlación, Dimensión de Capacidad, Entropía Espacio-Temporal, Determinismo, Información Mutua Promedio y Complejidad Relativa LZ*, y tanto en el caso de la predicción como de el modelado, *cada modelo responde en forma particular a los diferentes parámetros y no se observa un patrón de respuesta común*, lo cual indica que *cada modelo se adapta a la dinámica de las series de tiempo de forma diferente durante el proceso de aprendizaje o entrenamiento.*

Para estudiar la forma en que se agrupan y la jerarquía de las agrupaciones de los modelos

Tabla de la correlación de Pearson entre los parámetros de las series y las técnicas de predicción												
Parámetros de Series de Tiempo	Técnica de Predicción											
	Inteligencia Artificial						Teoría de Sistemas Dinámicos			Estadística		
	FINet	PNNGauss	PNNReciprocal	MLFFN	MySVM (kernel dot)	MySVM (kernel radial)	Polynomp	Nstep	K-Nearest Neighbours	ARIMA		
Exponente Lyapunov	MEDIA	ALTA	ALTA	ALTA	MUYBAJA	MUYALTA	MEDIA	ALTA	MUYALTA	MEDIA		
Dimensión de Correlación	MEDIA	MEDIA	MEDIA	ALTA	ALTA	MUYALTA	ALTA	MUYALTA	MUYALTA	BAJA		
Dimensión de Capacidad	MUYALTA	ALTA	MUYALTA	ALTA	MUYALTA	ALTA	ALTA	MUYALTA	MUYALTA	BAJA		
Dimensión Fractal	ALTA	MUYBAJA	MUYBAJA	BAJA	MEDIA	MUYBAJA	BAJA	ALTA	ALTA	MUYALTA		
Dimensión Entrelada	BAJA	MUYBAJA	ALTA	MEDIA	MUYALTA	MUYALTA	ALTA	BAJA	MEDIA	MEDIA		
Entropía Espacio Temporal (%)	MEDIA	MUYALTA	ALTA	MUYALTA	MUYALTA	MEDIA	BAJA	ALTA	BAJA	BAJA		
Recurrencia (%)	MUYBAJA	MEDIA	MUYALTA	MUYBAJA	MUYBAJA	MUYBAJA	MUYBAJA	MEDIA	BAJA	MUYALTA		
Determinismo (%)	ALTA	MUYALTA	MUYALTA	MUYALTA	BAJA	MUYALTA	MUYBAJA	ALTA	ALTA	BAJA		
Entropía de Información (Shannon)	ALTA	MUYALTA	MUYALTA	MUYALTA	MEDIA	MEDIA	MUYBAJA	MEDIA	MUYALTA	MEDIA		
Información Mútua Promedio	MUYBAJA	MUYALTA	BAJA	ALTA	MEDIA	MUYBAJA	MEDIA	ALTA	ALTA	ALTA		
Complejidad Relativa LZ	MEDIA	MUYALTA	MUYALTA	MUYALTA	MEDIA	ALTA	MUYBAJA	ALTA	MUYALTA	MEDIA		
Número de Reglas de Producción	ALTA	MUYALTA	MUYALTA	MUYALTA	MUYALTA	ALTA	BAJA	ALTA	ALTA	ALTA		

Figure 10-20: Tabla discretizada de la correlación entre los parámetros dinámicos y los modelos de predicción

Subconjunto de la tabla de la correlación de Pearson entre los parámetros de las series y las técnicas de predicción										
Parámetros de Series de Tiempo	Inteligencia Artificial						Técnica de Predicción			Estadística
	FIRNet	PNN Gauss	PNN Reciprocal	MLFFN	MYSVM (kernel dot)	MYSVM (kernel radial)	Teoría de Sistemas Dinámicos No Lineales			
							Polynomp	Nstep	K-Nearest-Neighbours	
Exponente de Lyapunov		ALTA	ALTA	ALTA		MUY ALTA		ALTA	MUY ALTA	
Dimensión de Correlación				ALTA	ALTA		ALTA	ALTA	MUY ALTA	
Dimensión de Capacidad	MUY ALTA	ALTA	MUY ALTA	ALTA	MUY ALTA	ALTA	ALTA	ALTA	MUY ALTA	
Entropía Espacio Temporal (%)		MUY ALTA	ALTA	MUY ALTA	MUY ALTA				ALTA	
Determinismo (%)	ALTA	MUY ALTA	MUY ALTA	MUY ALTA		MUY ALTA		ALTA	ALTA	
Entropía de Información (Shannon)	ALTA	MUY ALTA	MUY ALTA	MUY ALTA				MUY ALTA		
Complejidad Relativa LZ		MUY ALTA	MUY ALTA	MUY ALTA		ALTA		MUY ALTA	ALTA	
Número de Reglas de Producción	ALTA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	ALTA		ALTA	ALTA	ALTA

Figure 10-21: Tabla con los parámetros dinámicos cuya correlación es alta o muy alta respecto de los modelos de predicción

Tabla de la correlación de Pearson entre los parámetros de las series y las técnicas de modelado												
Parámetros de Series de Tiempo	Técnica de modelado											
	Inteligencia Artificial				Teoría de Sistemas Dinámicos No Lineales				Estadística			
	FFNBP	RBTFNDP	Antis	Híbridos	Predict	Polynom	rbf	Sugimay	ARMA			
Exponente de Lyapunov	-0.242	-0.145	-0.44	-0.345	-0.363	-0.226	-0.203	-0.321	-0.275			
Dimensión de Correlación	-0.322	-0.321	-0.409	-0.513	-0.159	-0.192	-0.219	-0.449	-0.375			
Dimensión de Capacidad	0.207	0.212	0.272	0.308	0.276	-0.096	-0.129	0.276	0.211			
Dimensión Fractal	0.055	-0.062	0.157	0.196	-0.149	-0.029	-0.017	0.055	0.142			
Dimensión Embebida	-0.267	-0.408	-0.03	-0.023	0.006	-0.397	-0.144	0.001	0.157			
Entropía Espacio Temporal (%)	-0.453	-0.421	-0.5	-0.383	-0.562	-0.126	-0.308	-0.384	-0.423			
Recurrencia (%)	0.114	-0.143	0.234	0.152	0.365	-0.084	-0.136	0.197	0.248			
Determinismo (%)	0.408	0.375	0.688	0.497	0.737	0.058	0.273	0.562	0.632			
Entropía de Información (Shannon)	0.154	0.122	0.507	0.37	0.611	0.059	0.048	0.452	0.652			
Información Mutua Promedio	0.442	0.292	0.398	0.459	0.399	-0.129	-0.065	0.464	0.341			
Complejidad Relativa LZ	-0.514	-0.504	-0.802	-0.706	-0.627	-0.323	-0.444	-0.732	-0.673			
Número de Reglas de Producción	0.09	0.05	-0.103	0.041	-0.244	-0.603	-0.512	0.025	-0.045			

Figure 10-22: Tabla de correlación entre los parámetros dinámicos y los modelos para modelado

Tabla de la correlación de Pearson entre los parámetros de las series y las técnicas de modelado												
Parámetros de Series de Tiempo	Técnicas de modelado											
	Inteligencia Artificial				Teoría de Sistemas Dinámicos No Lineales				Estadística			
	FTNBP	RBFNBP	Antfs	RNADP	Predict	Polynom	trf	SugImay	ARMA			
Exponente de Lyapunov	MEDIA	BAJA	MUY ALTA	ALTA	MUY ALTA	MEDIA	ALTA	ALTA	ALTA	ALTA	ALTA	ALTA
Dimensión de Correlación	ALTA	ALTA	ALTA	MUY ALTA	BAJA	BAJA	MUY ALTA	MUY ALTA	MUY ALTA	ALTA	ALTA	ALTA
Dimensión de Capacidad	ALTA	ALTA	MUY ALTA	MUY ALTA	MUY ALTA	MEDIA	MUY ALTA	MUY ALTA	MUY ALTA	ALTA	ALTA	ALTA
Dimensión Fractal	BAJA	BAJA	ALTA	MUY ALTA	ALTA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	ALTA	ALTA	ALTA
Dimensión Entebida	ALTA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	BAJA	MUY ALTA	BAJA
Entropía Espacio Temporal (%)	ALTA	ALTA	MUY ALTA	ALTA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	MEDIA	ALTA	ALTA
Recurrencia (%)	BAJA	BAJA	ALTA	MEDIA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	BAJA	MEDIA	ALTA
Determinismo (%)	MEDIA	MEDIA	MUY ALTA	ALTA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	ALTA	ALTA	MUY ALTA
Entropía de Información (Shannon)	BAJA	MUY ALTA	ALTA	MEDIA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	ALTA	MUY ALTA
Información Mutua Promedio	MUY ALTA	ALTA	MUY ALTA	MUY ALTA	MUY ALTA	BAJA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	ALTA
Complejidad Relativa LZ	ALTA	ALTA	MUY ALTA	MUY ALTA	ALTA	BAJA	MEDIA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA
Número de Reglas de Producción	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	BAJA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA	MUY ALTA

Figure 10-23: Tabla discretizada de la correlación entre los parámetros dinámicos y los modelos para modelado

Subconjunto de la tabla de la correlación de Pearson entre los parámetros de las series y las técnicas de modelado										
Parámetros de Series de Tiempo	Técnica de modelado									
	Inteligencia Artificial				Teoría de Sistemas Dinámicos No Lineales				Estadística	
	FTNBP	RBTNBP	Antis	Híbridos RNAP	Predict	Polynom	rbf	Sugimay	ARMA	
Exponente de Lyapunov			MUYALTA	ALTA	MUYALTA			ALTA		ALTA
Dimensión de Correlación	ALTA	ALTA	ALTA	MUYALTA				MUYALTA		ALTA
Dimensión de Capacidad	ALTA	ALTA	MUYALTA	MUYALTA	MUYALTA			MUYALTA		ALTA
Entropía Espacio Temporal (%)	ALTA	ALTA	MUYALTA	ALTA	MUYALTA			ALTA		ALTA
Determinismo (%)			MUYALTA	ALTA	MUYALTA			ALTA		MUYALTA
Información Múta	MUYALTA	ALTA	MUYALTA	MUYALTA	MUYALTA			MUYALTA		ALTA
Procedo			MUYALTA	MUYALTA	MUYALTA			ALTA		MUYALTA
Complejidad Relativa LZ	ALTA	ALTA	MUYALTA	MUYALTA	ALTA			MUYALTA		MUYALTA

Figure 10-24: Tabla con los parámetros dinámicos cuya correlación es alta o muy alta respecto de los modelos para modelado

de predicción en base a su correlación con los parámetros dinámicos se construyó un mapa GHSOM. Para construirlo se utilizó una tabla de datos como la que se muestra en la Figura 10-25. En la tabla, las filas corresponden a los modelos de predicción y las columnas a los parámetros de las series de tiempo, en las celdas se colocan los coeficientes de correlación de los parámetros para cada modelo de predicción.

La Figura 10-26, muestra el mapa GHSOM de los modelos de predicción, nuevamente este mapa se desglosa en la forma de un árbol de jerarquías como se muestra en la Figura 10-27.

Se formaron cuatro niveles de jerarquía, con seis grupos de modelos, donde *los modelos se agrupan por la similitud de sus correlaciones con los diferentes parámetros dinámicos, que caracterizan la predictibilidad de las series de tiempo*. Se observa que en el agrupamiento de modelos se presentaron los siguientes casos: *modelos del mismo origen pertenecen a un mismo grupo* como PNNReciprocal, PNNGauss y MLFFN cuyo origen es Inteligencia Artificial; *modelos de origen diferente pertenecen a un mismo grupo* como en el caso de MySVMdot (Inteligencia Artificial) con Polynomp (Teoría de Sistemas Dinámicos No Lineales) y Nstep (Teoría de Sistemas Dinámicos No Lineales) con PNNReciprocal, PNNGauss y MLFFN; *modelos que son el único elemento de un grupo* como en el caso de MySVMradial (Inteligencia Artificial), KNN (Inteligencia Artificial), ARIMA (Estadística) y FIRNet (Inteligencia Artificial). Se observa además que *la capacidad de predicción de los modelos (valores entre paréntesis) es independiente de la forma que se correlacionan con el conjunto de diferentes parámetros dinámicos*.

Para el conjunto de técnicas de modelado, se hizo el mismo análisis con el mapa GHSOM, la Figura 10-28, muestra el mapa correspondiente a las técnicas de modelado. El árbol de jerarquías correspondiente se presenta en la Figura 10-29.

El árbol posee cuatro niveles, pero a diferencia del caso de los modelos de predicción, en el nivel de mayor profundidad se encuentra *el modelo (RNAP) con mayor capacidad para modelar a las series de tiempo*. En el nivel dos se encuentran tres grupos de modelos, dos de ellos agrupan modelos que además son similares en el origen de su arquitectura, un grupo corresponde a Inteligencia Artificial (RBFNBP y FFNBP) y el otro grupo corresponde a modelos de Teoría de Sistemas Dinámicos No Lineales (Polynom y RBF); el tercer grupo de modelos en el nivel dos, se subdivide a su vez en dos grupos, donde uno de ellos forma el nivel tres del árbol de jerarquías. Los agrupamientos muestran (como también se observó para los modelos de

	Exponente de Lyapunov	Dimensión de Correlación	Dimensión de Capacidad	Dimensión Fractal	Dimensión Embebida	Entropía Espacio Temporal (%)	Recurrencia (%)	Determinismo (%)	Entropía de Información (Shannon)	Información Mútua Promedio	Complejidad Relativa LZ	Número de Reglas de Producción
FIRNet	-0.247	-0.174	0.226	-0.162	-0.116	-0.333	0.027	0.37	0.362	0.022	-0.413	-0.255
FNNGauss	-0.331	-0.22	0.184	-0.032	-0.031	-0.574	0.174	0.549	0.506	0.327	-0.666	-0.373
FNNRadprocal	-0.394	-0.171	0.205	0.036	-0.262	-0.44	0.3	0.533	0.529	0.13	-0.572	-0.411
MFRN	-0.328	-0.243	0.178	-0.088	-0.181	-0.482	-0.033	0.576	0.563	0.225	-0.685	-0.394
MFSMI												
Multipunto (kernel ddt)	-0.075	-0.263	-0.248	-0.063	-0.318	-0.468	0.015	0.129	0.247	-0.158	-0.3	-0.38
MFSMI												
Multipunto (kernel radial)	-0.418	-0.298	0.183	-0.026	-0.39	-0.252	0.007	0.469	0.318	-0.062	-0.49	-0.306
Polynomp	-0.209	-0.237	0.19	-0.049	-0.267	0.125	0.062	0.014	0.083	-0.165	-0.117	-0.143
Nstep	-0.314	-0.346	0.245	-0.077	-0.109	-0.279	0.108	0.424	0.502	0.222	-0.564	-0.268
K-Nearest-Neighbours	-0.465	-0.365	0.211	0.148	-0.198	-0.376	0.179	0.358	0.257	0.178	-0.446	-0.292
ARIMA	-0.247	-0.146	-0.089	0.201	0.21	-0.208	0.325	0.181	0.288	0.258	-0.33	-0.291

Figure 10-25: Tabla de datos para construir mapa GHSOM de modelos en función de su correlación con los parámetros de las series de tiempo

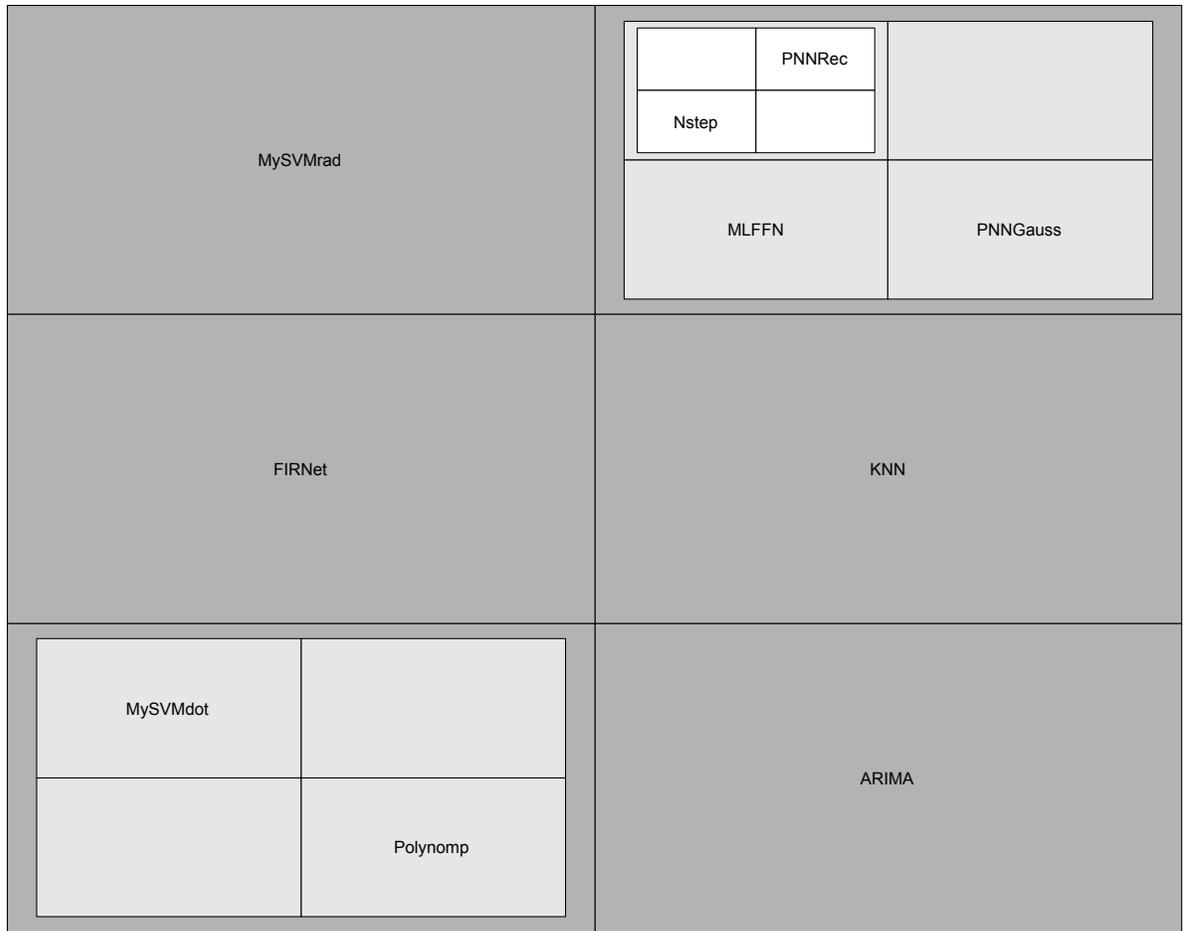


Figura 10-26: Mapa Auto Organizado Jerárquico para los modelos de predicción

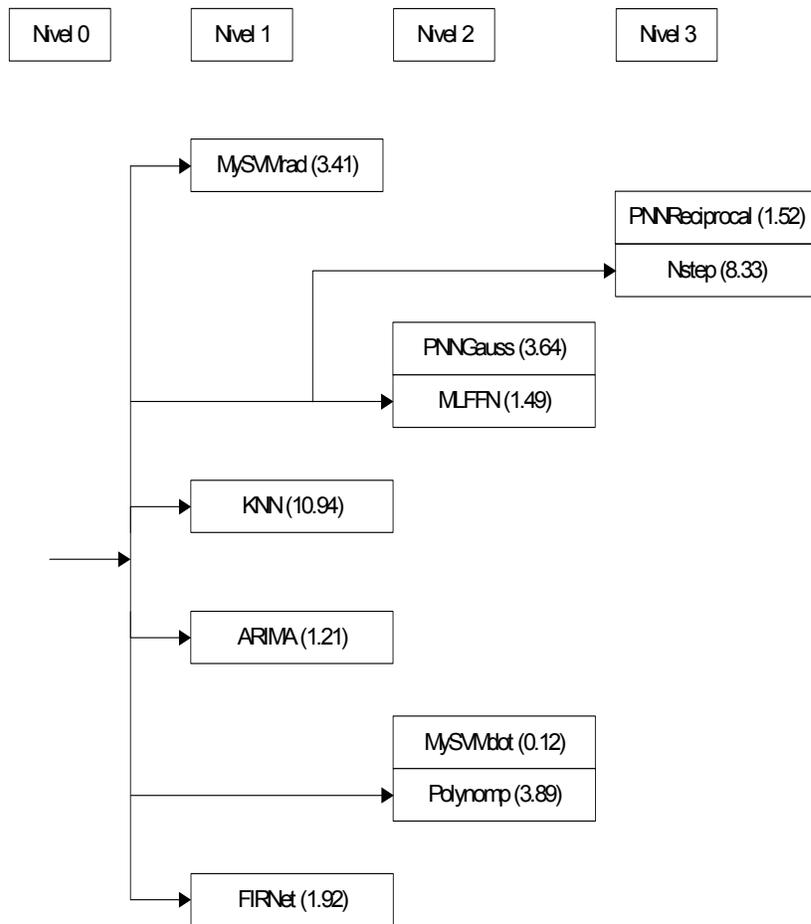


Figura 10-27: Árbol de jerarquías del agrupamiento de los modelos de predicción con el Mapa Auto Organizado

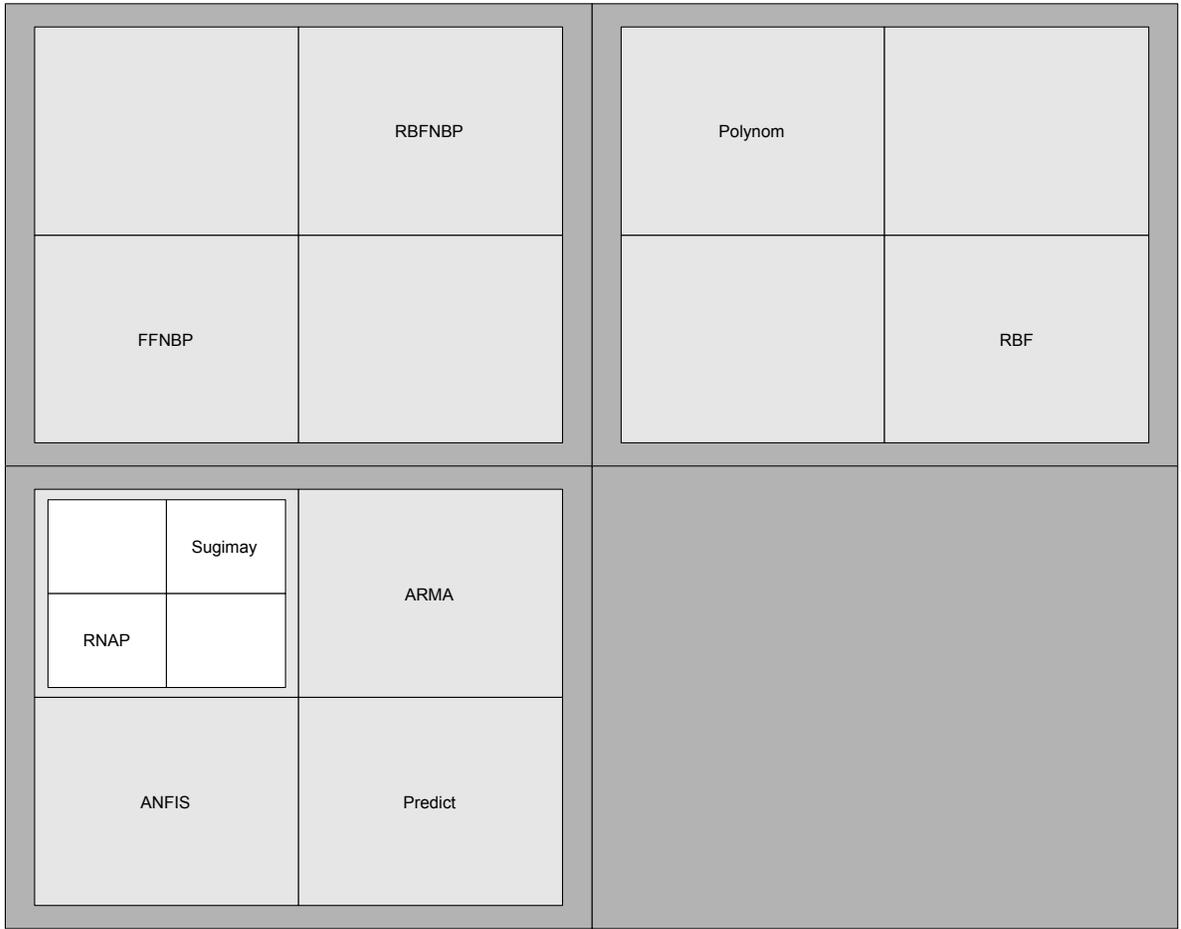


Figura 10-28: Mapa Auto Organizado Jerárquico de los modelos para modelado

predicción) que se puede dar el caso de que *modelos de mismo origen presenten una respuesta similar de su correlación con los parámetros dinámicos de las series* como ocurre en los grupos del nivel dos; pero también es posible, que ocurra el caso de que *modelos de diferente origen presenten respuesta similar de su correlación con los parámetros dinámicos de las series*, como el caso de los grupos del nivel tres donde el modelo RNAP (Inteligencia Artificial) y el modelo Sugimay (Teoría de Sistemas Dinámicos No Lineales) pertenecen al mismo grupo, y lo mismo ocurre para los modelos ANFIS (Inteligencia Artificial), Predict (Teoría de Sistemas Dinámicos No Lineales) y ARMA (Estadística). Por último, a diferencia de los modelos de predicción, se observa que *la capacidad de modelado (valores entre paréntesis) presenta una relación con la forma en que se agrupan los modelos de acuerdo a su correlación con los parámetros de las series de tiempo.*

## 10.6 Discusión

La agrupación de las series de tiempo por sus parámetros dinámicos, permitió *identificar en forma experimental que hay tres grandes grupos de series con diferente predictibilidad, etiquetada como baja, media y alta predictibilidad; y dentro de esas agrupaciones una jerarquía de similitud de las series.* Esto muestra, *la riqueza del comportamiento de la predictibilidad de series de tiempo que es consecuencia de las diferentes dinámicas de los fenómenos que las series representan.*

El haber cuantificado la predictibilidad de las series de tiempo hizo posible *encontrar en forma experimental la existencia de una relación entre la estructura de una serie de tiempo y su predictibilidad.* Esta relación se manifiesta con diferentes formas de representación de las series como son el mapa de recurrencia (VRA) y las componentes de una serie producto del análisis espectral singular (SSA). En particular con los mapas de recurrencia se observó que *existen patrones de estructura básicos presentes en series de tiempo de diferente origen (natural o artificial).* Estos patrones básicos representan *dinámicas comunes para diferentes sistemas (físicos, matemáticos, fisiológicos, económicos, etc.).* La relación predictibilidad-estructura se manifiesta por el agrupamiento de series de tiempo con predictibilidad similar y donde se observó que sus mapas de recurrencia presentaban uno o más patrones comunes o básicos.

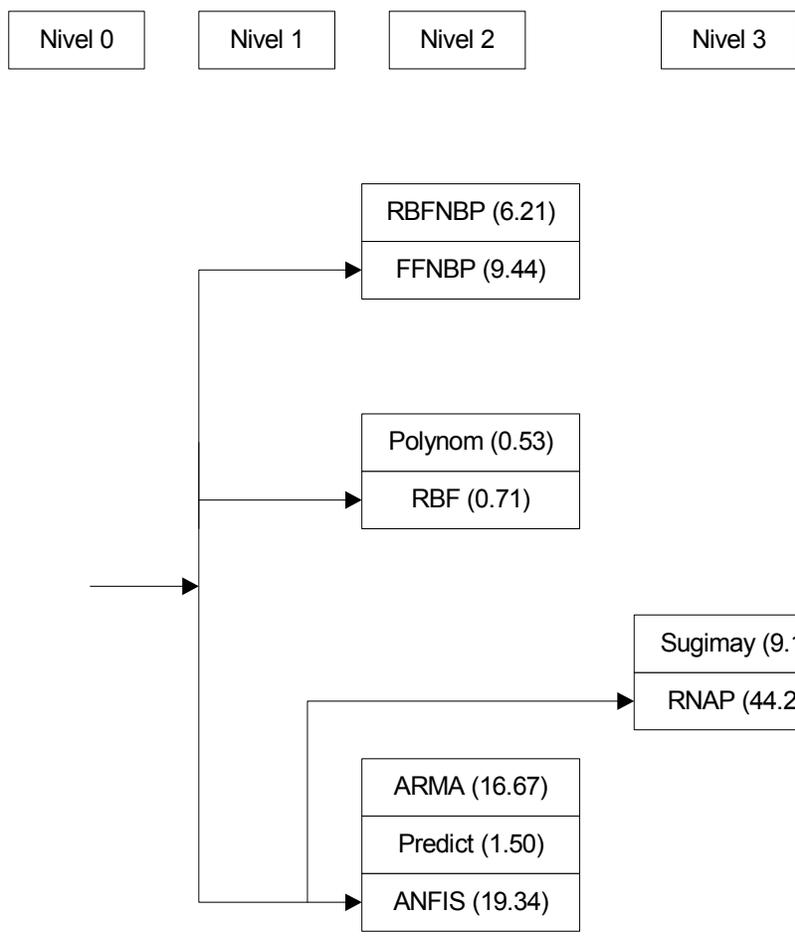


Figura 10-29: Árbol de jerarquías del agrupamiento para los modelos de modelado con el Mapa Auto Organizado

Al analizar las características de la arquitectura de los modelos y relacionarlo con su capacidad de predicción y de modelado se encontraron los siguientes resultados:

- *la capacidad de predicción más alta corresponde a modelos cuyo alcance es local y que actúan en el espacio fase, los modelos de predicción de alcance global y que actúan en diferentes espacios son menos eficientes en la predicción de series de tiempo.*
- *la capacidad de modelado más alta corresponde a modelos cuyo alcance es global y que actúan en el espacio real, los modelos que actúan en el espacio fase son menos eficientes en el modelado de series de tiempo.*

El estudio de la relación entre los parámetros que caracterizan la predictibilidad y los modelos de predicción y de modelado por medio de los análisis de correlación, arrojó los siguientes resultados: *no se observa un patrón o patrones de respuesta comunes de los modelos con respecto a los parámetros, cada modelo responde en forma particular a los diferentes parámetros, lo que indica que cada modelo se adapta a la dinámica de las series de tiempo de forma diferente, lo cual intuitivamente es de esperarse dadas las diferentes arquitecturas de los modelos.* Aun cuando cada modelo responde en forma particular a los diferentes parámetros, se encontró que los parámetros que presentan una mayor correlación con la mayoría de los modelos en el caso de la predicción son: *Exponente de Lyapunov, Dimensión de Correlación, Dimensión de Capacidad, Entropía Espacio-Temporal, Determinismo, Entropía de Información de Shannon, Complejidad Relativa LZ y Número de Reglas de Producción.* Y en el caso del modelado son: *Exponente de Lyapunov, Dimensión de Correlación, Dimensión de Capacidad, Entropía Espacio-Temporal, Determinismo, Información Mutua Promedio y Complejidad Relativa LZ.*

Al analizar con los mapas GHSOM, la forma en que los modelos de predicción y de modelado se agrupan en función de su correlación con los diferentes parámetros, se encontraron dos clases de agrupamientos: *modelos del mismo origen* (por ejemplo, Inteligencia Artificial) *que responden globalmente de la misma forma a los parámetros de las series de tiempo, y modelos de origen diferente que responden globalmente de la misma forma a los parámetros de las series de tiempo.* La jerarquía de los agrupamientos de modelos para el caso de la predicción muestra que *no existe una relación entre la forma en que los modelos se correlacionan con los parámetros y su capacidad de predicción.* En el caso del modelado, se encontró que *existe una relación entre*

*la capacidad de modelado y la agrupación de los modelos en jerarquías de similitud en función de su respuesta a los parámetros de las series de tiempo.*

## Parte III

# Combinación de Predictores

## Capítulo 11

# Combinación de Predictores: Mejorando la Predicción de las Series de Tiempo

### 11.1 Introducción

En esta tesis se han desarrollado estudios sobre la predictibilidad de las series de tiempo y su relación con las características dinámicas y estructurales de las mismas, también sobre la capacidad de predicción (o modelado) de diferentes modelos y su relación con las características de las series de tiempo. Los resultados experimentales del presente trabajo muestran que *cada modelo de predicción presenta una respuesta diferente a los parámetros que caracterizan a una serie de tiempo*, y en consecuencia *no hay un modelo que sea capaz de adaptarse a todo el universo de dinámicas que las series de tiempo pueden presentar*, esto confirma que ante la falta de modelos universales, una de las *mejores estrategias* para predecir una serie de tiempo y a su vez mejorar el horizonte de la predicción es mediante *la combinación de modelos de predicción (o combinación de predictores)*.

Como se mencionó en la sección 2.6, al predecir o modelar una serie de tiempo existe un error que es el resultado de los siguientes factores:

- Las limitaciones inherentes a la simplificación de la dinámica del sistema por el modelo

de predicción.

- Las propiedades del sistema como por ejemplo el caos, que provoca una desviación de tipo exponencial de la trayectoria del sistema para cambios pequeños en las condiciones iniciales del mismo.
- El ruido presente en la serie de tiempo de origen natural, debido al proceso de medición experimental de la misma.

Este error lo podemos expresar formalmente como:

$$E_{predicción} = E_{modelado} + E_{intrínseco} + E_{medición} \quad (11.1)$$

El error de predicción corresponde a la clase de error sistemático [65], el cual no es posible de eliminar o disminuir en forma directa durante el proceso de predicción o de modelado. Sin embargo, es posible disminuir dicho error al post procesar los datos predichos mediante la combinación de los resultados de un conjunto de técnicas de predicción (*predictores*) o en el caso de modelar series de tiempo mediante la combinación de técnicas de modelado. Al mejorar el resultado de la predicción combinando predictores es posible a su vez mejorar el horizonte de predicción de la serie de tiempo.

En este capítulo se describe el desarrollo y evaluación de dos técnicas de combinación de predictores. En la sección 11.2, se presenta una descripción de las técnicas clásicas de combinación de predictores. La sección 11.3, describe la primera técnica de combinación de predictores desarrollada en esta tesis denominada: *GABoost*. Esta técnica hace uso de las ideas de aprendizaje mediante el algoritmo Boosting (voto pesado) y de búsqueda con Algoritmo Genético de dichos pesos. La sección 11.4, presenta la segunda técnica desarrollada denominada: *CombFEC*, la cual utiliza una función de error y correlación para obtener los pesos de la combinación de predictores. Finalmente, la sección 11.5 presenta la discusión del presente capítulo.

## 11.2 Técnicas Clásicas de Combinación de Predictores

Es posible mejorar el resultado de la predicción de una serie de tiempo, mediante la aplicación de operaciones estadísticas simples como por ejemplo, el promedio simple y el promedio pesado.

Gran cantidad de métodos de combinación se han propuesto a partir del análisis estadístico, se dividen en dos grupos, los métodos de varianza-covarianza y los métodos de regresión [11, 12]. A continuación se describen ambos grupos de combinación estadística de modelos de predicción.

Los métodos de varianza-covarianza se construyen de la siguiente forma, supongamos que se tienen dos predicciones sin bias, para las cuales se forma la siguiente predicción compuesta:

$$y_{t+h,t}^c = \omega y_{t+h,t}^a + (1 - \omega) y_{t+h,t}^b \quad (11.2)$$

Debido a que la suma de los pesos es igual a la unidad, las predicciones compuestas no tendrán bias. Además, el error de la predicción combinada satisface la misma relación que las predicciones combinadas,

$$e_{t+h,t}^c = \omega e_{t+h,t}^a + (1 - \omega) e_{t+h,t}^b \quad (11.3)$$

con una varianza

$$\sigma_c^2 = \omega^2 \sigma_{aa}^2 + (1 - \omega)^2 \sigma_{bb}^2 + 2\omega(1 - \omega) \sigma_{ab}^2 \quad (11.4)$$

donde  $\sigma_{aa}^2$  y  $\sigma_{bb}^2$  son las varianzas del error de predicción y  $\sigma_{ab}^2$  es la covarianza. Minimizando la varianza del error de la combinación de modelos con respecto a  $\omega$ , se obtiene el peso de combinación óptimo,

$$\omega^* = \frac{\sigma_{bb}^2 - \sigma_{ab}^2}{\sigma_{bb}^2 + \sigma_{aa}^2 - 2\sigma_{ab}^2} \quad (11.5)$$

La varianza del error de predicción asociado con la predicción producto de la combinación óptima, es menor o igual que la menor varianza de  $\sigma_{aa}^2$  y  $\sigma_{bb}^2$ . En la practica, los pesos son desconocidos, así que son estimadas de la siguiente forma,

$$\hat{\sigma}_{ij}^2 = \frac{1}{T} \sum_{i=1}^T e_{t+h,t}^i e_{t+h,t}^j \quad (11.6)$$

así el peso de combinación estimado es,

$$\hat{\omega}^* = \frac{\hat{\sigma}_{bb}^2 - \hat{\sigma}_{ab}^2}{\hat{\sigma}_{bb}^2 + \hat{\sigma}_{aa}^2 - 2\hat{\sigma}_{ab}^2} \quad (11.7)$$

El otro grupo de métodos se basan en combinar las predicciones mediante la regresión, de hecho este tipo de métodos son un caso particular de los métodos de varianza-covarianza, ya que los pesos óptimos se pueden interpretar como los coeficientes de una proyección lineal de  $y_{t+h}$  sobre la predicción, sujeta a dos restricciones: la suma de pesos es igual a la unidad y la intercepción es excluida.

La expresión general de los métodos de regresión para combinación de predictores es la siguiente,

$$y_{t+h} = \beta_0 + \beta_1 y_{t+h,t}^{p_1} + \beta_2 y_{t+h,t}^{p_2} + \dots + \beta_n y_{t+h,t}^{p_n} + \varepsilon_{t+h,t} \quad (11.8)$$

Hay muchas variaciones y extensiones del mismo por ejemplo: combinación de pesos que varían con el tiempo, regresiones combinadas dinámicamente, reducción de pesos de combinación a la igualdad (es decir promedio aritmético de predicciones), y regresiones de combinaciones no lineales [11, 12].

## 11.3 GABOOST: Búsqueda de Pesos con Algoritmo Genético Canónico para Minimización del Error de Predicción RMSE

### 11.3.1 El Algoritmo Boosting

En el área de Aprendizaje de Máquinas (*Machine Learning*), se han desarrollado técnicas para incrementar la habilidad de aprendizaje durante el entrenamiento de las redes neuronales utilizadas en tareas de clasificación y regresión. Estas técnicas reducen el error de predicción de la red, por medio de la siguiente idea: crear muchas versiones del mismo predictor y combinarlas para mejorar el proceso de aprendizaje de los predictores (por ejemplo, redes neuronales) [180].

El conjunto de predictores se obtiene alterando el conjunto de aprendizaje y usando esta nueva versión para crear al predictor. Existen varios algoritmos para construir conjuntos de predictores, tales como bagging, boosting, arcing, alteración de la estructura interna y la introducción de aleatoriedad en la salida. En particular en este trabajo, se ha utilizado el algoritmo

boosting como una analogía para el desarrollo de métodos de reducción del error de predicción, a continuación se describen a detalle dicho algoritmo.

Boosting es un método general para mejorar la precisión de cualquier algoritmo de aprendizaje. El objetivo del algoritmo boosting es impulsar (*boost*) la precisión del proceso de aprendizaje de predictores, su fundamento teórico proviene del modelo de aprendizaje *PAC* (*Probably Approximately Correct Model*) donde la hipótesis del algoritmo de aprendizaje debe ser correcta en forma aproximada (poseer bajo error) con una alta probabilidad, este modelo es debido a Valiant [181]. Kearns y Valiant, fueron los primeros en proponer la pregunta de si un algoritmo de aprendizaje débil (i.e. de bajo rendimiento) cuyo rendimiento es ligeramente mejor que una adivinanza aleatoria dentro del modelo *PAC* puede ser impulsado a una precisión arbitraria correspondiente a un algoritmo de aprendizaje fuerte (i.e. de alto rendimiento) [182]. Schapire fue el primero en desarrollar un algoritmo boosting en 1989 [183]. Posteriormente Freund desarrolló una versión más eficiente. En 1995 Freund y Schapire introdujeron el algoritmo AdaBoost [182], el cual resolvió muchas de las dificultades prácticas de los primeros algoritmos de boosting.

A continuación se describe el algoritmo para mostrar cual es la forma en que opera. El algoritmo AdaBoost toma como entrada un conjunto de entrenamiento  $(x_1, y_1), \dots, (x_m, y_m)$  donde cada  $x_i$  pertenece a algún dominio o espacio de instancias  $X$ , y cada etiqueta  $y_i$  está dentro de algún conjunto de etiquetas  $Y$ . Se asume en esta descripción en la cual  $Y = \{-1, +1\}$ . AdaBoost llama a un algoritmo de aprendizaje débil o base (por ejemplo, una red neuronal para tareas de clasificación como en este caso, aunque también es posible que el algoritmo de aprendizaje realice tareas de regresión) en forma repetida en una serie de iteraciones  $t = 1, \dots, T$ . Una de las principales ideas del algoritmo es mantener una distribución o conjunto de pesos sobre el conjunto de entrenamiento. El peso de esta distribución en el ejemplo de entrenamiento  $i$  en la iteración  $t$  se denota como  $D_t(i)$ . Inicialmente, todos los pesos son iguales, pero en cada iteración, los pesos de los ejemplos clasificados de forma incorrecta se incrementan de tal forma que el predictor débil (*weak learner*) es forzado a enfocarse en los ejemplos difíciles del conjunto de entrenamiento.

Esta distribución de pesos es de la forma

$$D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t} \quad (11.9)$$

donde  $Z_t$  es un factor de normalización (seleccionado de tal forma que  $D_{t+1}$  sea una distribución).

El algoritmo de aprendizaje tiene que encontrar una hipótesis débil  $h_t : X \rightarrow \{-1, +1\}$  apropiada para la distribución  $D_t$ . La bondad de una hipótesis débil se mide por su error

$$\epsilon_t = \Pr_{i \sim D_t} [h_t(x_i) \neq y_i] = \sum_{i: h_t(x_i) \neq y_i} D_t(i) \quad (11.10)$$

El error es medido con respecto a la distribución  $D_t$  sobre la cual el algoritmo de aprendizaje es entrenado.

En la practica, el algoritmo de aprendizaje debería ser capaz de utilizar los pesos  $D_t$  en los ejemplos de entrenamiento. En forma alternativa, cuando no es posible lo anterior, un subconjunto de los ejemplos de entrenamiento puede ser muestreado de acuerdo a  $D_t$ , y estos ejemplos (no pesados) muestreados pueden ser usados para entrenar al algoritmo.

Una vez que la hipótesis débil  $h_t$  ha sido obtenida, AdaBoost selecciona un parámetro  $\alpha_t$ . De forma intuitiva,  $\alpha_t$  mide la importancia que se asigna a  $h_t$ , y cumple con lo siguiente

$$\alpha_t \geq 0 \text{ si } \epsilon_t \leq \frac{1}{2} \quad (11.11)$$

y  $\alpha_t$  crece conforme  $\epsilon_t$  decrece.

La distribución  $D_t$  es luego actualizada de tal forma que los pesos de los ejemplos mal clasificados por  $h_t$  se incrementa, y a la vez se decrementa el peso de los ejemplos clasificados correctamente. De esta forma, el peso tiende a concentrarse en los ejemplos difíciles.

La hipótesis final  $H$  es un voto de mayoría pesado de las  $T$  hipótesis débiles donde  $\alpha_t$  es el peso asignado a  $h_t$ .

$$H(x) = \text{sign}\left(\sum_{t=1}^T \alpha_t h_t(x)\right) \quad (11.12)$$

### 11.3.2 Descripción del Algoritmo de la Técnica GABOOST

Esta técnica se desarrollo a partir de la idea de boosting utilizada en aprendizaje de máquina (e.g. redes neuronales), en la cual se asigna el mayor peso al peor predictor (*learner*) para reforzar su aprendizaje en la siguiente iteración de la etapa de entrenamiento y donde estos

pesos se obtienen a partir de la construcción de una función de distribución de pesos. Como una *analogía* ahora asignamos el mayor peso al mejor predictor de una serie de tiempo, de tal forma que construimos un predictor de voto pesado, el cual se expresa como la suma pesada de  $M$  predictores:

$$P(x_i)_{combinación} = \sum_{j=1}^M W_j P^j(x_i) \quad (11.13)$$

La búsqueda de los pesos óptimos  $W_j$  se realiza a diferencia de la técnica boosting, que requiere la construcción explícita de una función de distribución de pesos, utilizando un *algoritmo genético canónico* que construye esta distribución de forma *implícita* [184, 185, 186], el criterio de búsqueda de los pesos óptimos se basa en la función objetivo de minimización del error medio raíz cuadrada (*RMSE*),

$$RMSE = \left[ \frac{1}{N} \sum_{i=1}^N \left( x_i^{combinación} - x_i^{original} \right)^2 \right]^{\frac{1}{2}} \quad (11.14)$$

Cada individuo de la población esta formado por un conjunto de pesos  $\{W_j\}$  expresados como números reales, de la siguiente forma:

$$Popmember [k] = W_1 W_2 \dots W_M \quad (11.15)$$

la función de ajuste (*fitness function*) penaliza a los individuos con mayores errores *RMSE* de la forma:

$$fitness = 1 - \frac{RMSE(k)}{Max \{RMSE(k)\}_{k=1}^Q} \quad (11.16)$$

El algoritmo de la técnica *GABoost* para combinación de predictores, se muestra en la Figura 11-1.

### 11.3.3 Técnicas y Series para la Evaluación de GABoost

Las series de tiempo utilizadas para evaluar la técnica de combinación *GABoost* fueron seleccionadas considerando que el número de técnicas de predicción capaces de predecir su compor-

### *Etapa de Entrenamiento*

Búsqueda de pesos para la reducción del error de predicción.

$P_1^{Entrenamiento}, P_2^{Entrenamiento}, \dots, P_M^{Entrenamiento}$  Conjuntos de datos predichos correspondientes al mismo intervalo de predicción, obtenidos a partir de  $M$  modelos de predicción.

continúe

Búsqueda de pesos con Algoritmo Genético:

$W_1, W_2, \dots, W_M$

Salida de la Combinación de Predictores:

$$P_{combinación}(x_i) = \sum_{j=1}^M W_j P_j^{Entrenamiento}(x_i)$$

Evaluación de la salida:

$$RMSE = \left[ \frac{1}{N} \sum_{i=1}^N (x_i^{combinación} - x_i^{original})^2 \right]^{\frac{1}{2}}$$

if ( $RMSE < RMSE_{mejorpredicador}$  o número de iteraciones = MaxNumIterations) then

end program

else

continúe

end program

### *Etapa de Prueba*

Uso del conjunto de pesos para la reducción del error de predicción en el conjunto de prueba.

$P_1^{Prueba}, P_2^{Prueba}, \dots, P_M^{Prueba}$  Conjuntos de datos predichos correspondientes al mismo intervalo de predicción, obtenidos a partir de  $M$  modelos de predicción.

$$P_{combinación}(x_i) = \sum_{j=1}^M W_j P_j^{Prueba}(x_i)$$

Figura 11-1: Algoritmo GABoost

tamiento fuera al menos de dos. Para la selección de las técnicas y de las series, se utilizó la *matriz de predicción* de las series de tiempo que se presentó en el capítulo 8. El conjunto de técnicas de predicción (*predictores*) y sus características generales se presenta en la tabla de la Figura 11-2.

Técnica de Predicción	Descripción
FIRNet	Red Neuronal con Filtros de Respuesta de Impulso Finito.
Red Neuronal Probabilística (PNN-Gauss)	Red Neuronal con una función de pesos Gaussiana, y aprendizaje basado en Estadística Bayesiana.
Red Neuronal Multicapa hacia Adelante (MLFFN)	Perceptron simple con 0 capas ocultas
Polynomp	Modelo de funciones polinomiales en el espacio fase reconstruido.
Nstep	Modelo lineal local en el espacio fase reconstruido.

Figura 11-2: Predictores de series de tiempo

El conjunto de series de tiempo utilizadas para la evaluación de la técnica *GABoost* se presenta en la tabla de la Figura 11-3.

Serie de Tiempo	Comportamiento Dinámico	Predictibilidad
Qperiodic3	Cuasiperiódica	0.888610816
Mackey-Glass	Caótica	0.665578391
Rosler	Caótica	1.172559224
A1	Compleja	0.888622568
Dow Jones	Compleja	1.791995215

Figura 11-3: Series de Tiempo

### 11.3.4 Resultados Experimentales

A continuación se presentan los resultados experimentales de la evaluación de la técnica de combinación de predictores *GABoost*, los datos utilizados para cada serie de tiempo se presentan

en la tabla de la Figura 11-4. Solo en el caso de la serie Mackey-Glass se utilizaron únicamente 50 datos predichos para la construcción del conjunto de entrenamiento (búsqueda de pesos) y del conjunto de prueba para la combinación de los predictores. El uso de solo 50 datos para construir los conjuntos de datos, es por el comportamiento caótico de esta serie que expande el error de predicción de tal forma que la combinación de los predictores, no logra mejorar el resultado de la mejor predicción individual o de promedio simple.

Serie de Tiempo	Datos de entrenamiento o para predicción	Datos predichos	Datos de entrenamiento o para búsqueda de pesos	Datos de prueba para la combinación de los predictores
Qperiodic3	1900	100 (1901 a 2000)	50 (1901 a 1950)	50 (1951 a 2000)
Mackey-Glass	1900	50 (1901 a 1950)	25 (1901 a 1925)	25 (1926 a 1950)
Rosler	4000	100 (4001 a 4100)	50 (4001 a 4050)	50 (4050 a 4100)
A1	1800	100 (1801 a 1900)	50 (1801 a 1850)	50 (1851 a 1900)
Dow Jones	3000	100 (3001 a 3100)	50 (3001 a 3050)	50 (3051 a 3100)

Figura 11-4: Conjuntos de datos para la evaluación de GABoost

En la búsqueda de una combinación de predictores óptima un criterio experimental es el grado de correlación de los predictores, a menor correlación entre predictores (i.e. mayor ortogonalidad entre sus predicciones) la información que cada uno aporta para la combinación de predictores es menos redundante logrando que la predicción resultante posea un menor error de predicción. En la Figura 11-5 se muestra la tabla de los resultados del análisis de correlación. Se observa que en el caso de las series de tiempo Rosler, Mackey-Glass y A1, el grado de correlación entre los predictores es alto, en el caso de la serie Qperiodic3 la correlación entre los predictores es alta en unos casos y en otros es baja lo cual indica que la combinación de los predictores puede ser útil para mejorar la predicción de la serie. Por último en el caso de la serie Dow Jones, la correlación entre los dos predictores a combinar es de magnitud media pero con una relación de tipo inverso.

Una vez obtenidos los datos de la predicción e identificados los predictores para cada serie

Serie de Tiempo	Predictores	Correlación de Pearson
Qperiodic3	pnngauss vs mlffn	0.828
	pnngauss vs polynomp	0.818
	pnngauss vs nstep	0.163
	mlffn vs polynomp	0.725
	mlffn vs nstep	0.191
	polynomp vs nstep	0.097
Rossler	pnngauss vs mlffn	0.897
	pnngauss vs nstep	0.997
	mlffn vs nstep	0.909
Mackey-Glass	firnet vs nstep	0.958
A1	pnngauss vs polynomp	0.949
	pnngauss vs nstep	0.963
	polynomp vs nstep	0.982
Dow Jones	pnngauss vs nstep	-0.666

Figura 11-5: Correlación de los predictores

de tiempo, se realizaron los experimentos para la búsqueda de los pesos óptimos que minimicen el RMSE de la combinación de predictores, tal que sea menor o igual al RMSE de la mejor predicción individual de la serie de tiempo. La Figura 11-6, presenta la tabla de los parámetros experimentales con los cuales se obtuvieron los pesos óptimos para cada caso.

Serie de Tiempo	Parámetros de experimentos exitosos					Número de experimentos para cada serie de tiempo
	Tamaño de población	Máximo número de generaciones	Probabilidad de cruzamiento	Probabilidad de mutación	Número de genes	
Qperiodic3	200	100000	0,25	0,03	4	18
Mackey-Glass	50	400000	0,25	0,01	2	2
Rossler	50	400000	0,25	0,01	3	5
A1	100	400000	0,29	0,025	3	1
Dow Jones	50	400000	0,25	0,01	2	1

Figura 11-6: Parámetros de los experimentos para la búsqueda de los pesos óptimos

En las Figuras 11-7, 11-8 y 11-9; se muestran las gráficas en las cuales se compara la serie original, la predicción GABOOST y la predicción con el mejor predictor individual, los casos

presentados corresponden a las series Mackey-Glass, A1 y Dow Jones. En el caso de la serie Mackey-Glass se observa que debido al comportamiento caótico de esta serie, la propagación del error genera una predicción combinada que solamente mejora la predicción para los primeros 12 valores de la serie. Con relación a la serie A1, se observa que la combinación de los predictores mejora ligeramente la predicción de la amplitud de la serie de tiempo, reduciendo un error de predicción que ya era pequeño debido a las predicciones individuales como la realizada por la técnica Nstep. Por último, en el caso de la serie Dow Jones se observa una mejora considerable en la predicción de la magnitud de los valores de la serie predicha.

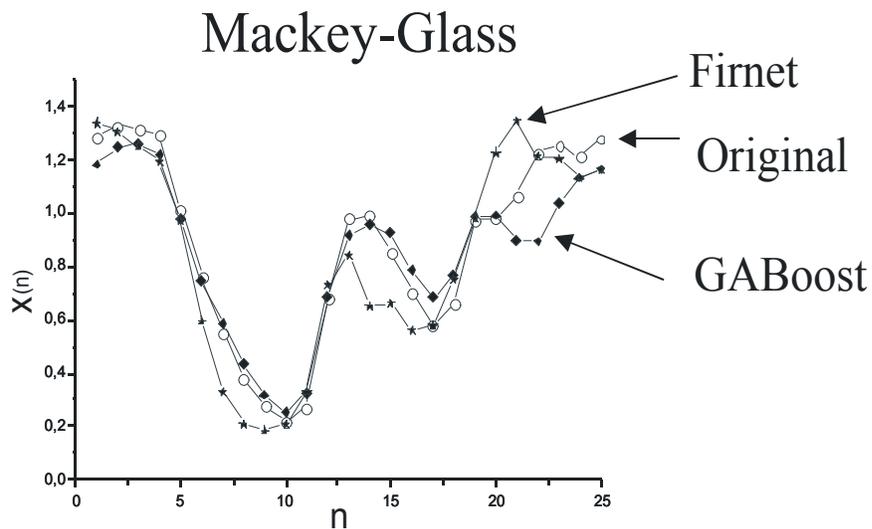


Figura 11-7: Comparación de predicciones de GABoost y mejor predictor individual para la serie Mackey-Glass

Para verificar la eficiencia de la técnica para reducir el error RMSE se compararon los errores de los predictores individuales con el correspondiente a *GABoost*, las Figuras 11-10, 11-11, 11-12, 11-13 y 11-14; muestran las comparaciones para las diferentes series de tiempo. El porcentaje que se muestra en cada figura, corresponde a la reducción del error RMSE de la combinación con respecto al error RMSE del mejor predictor individual. Se observa que en todos los casos se logro reducir el error RMSE con relación al resultado del mejor predictor individual para cada serie de tiempo.

El Error Bias (BE) que permite cuantificar si el modelo de predicción se ajusta a la dinámica

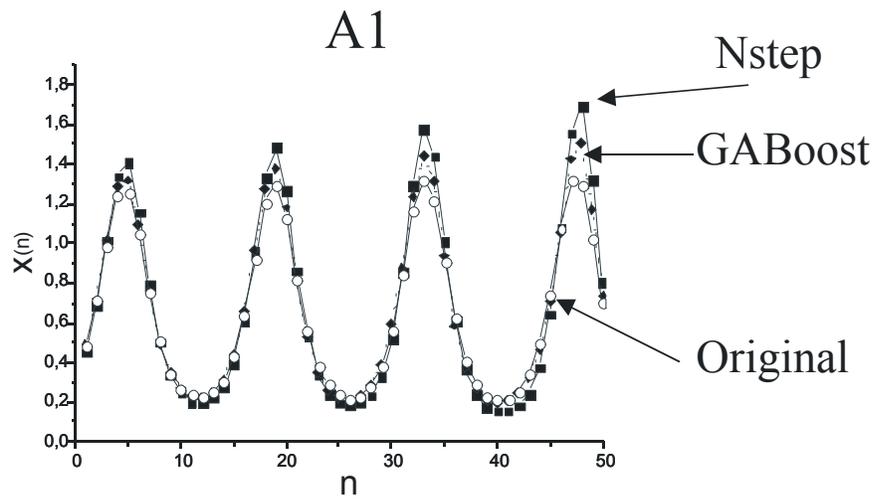


Figura 11-8: Comparación de predicciones de GABoost y mejor predictor individual para la serie A1

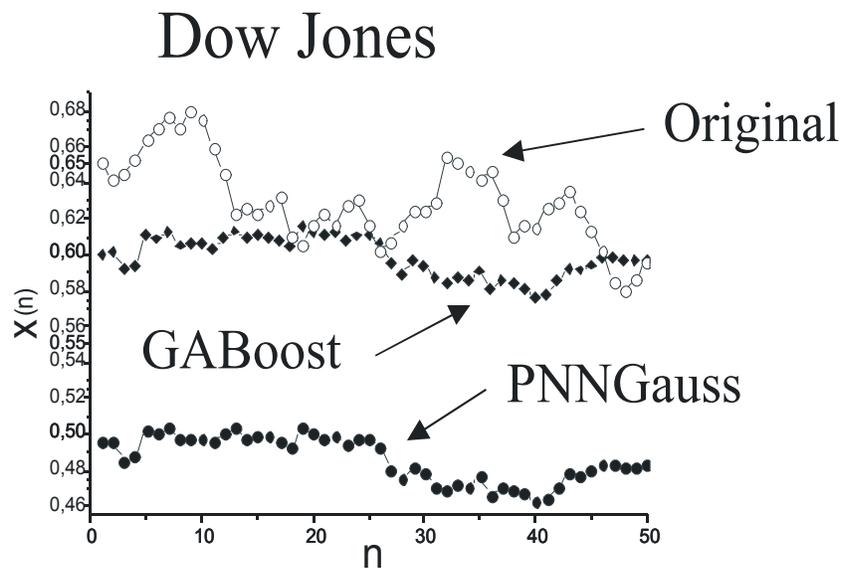


Figura 11-9: Comparación de predicciones de GABoost y mejor predictor individual para la serie Dow Jones

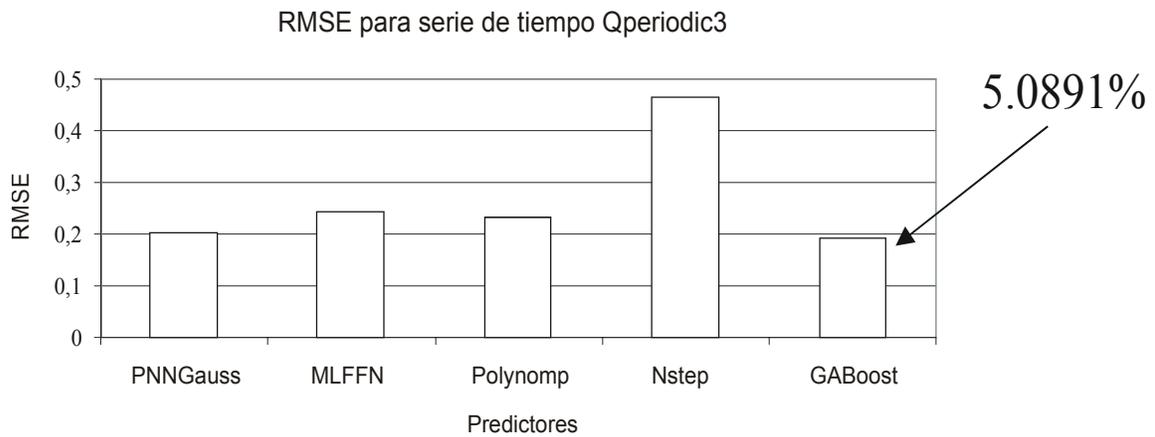


Figura 11-10: Comparación del error RMSE para la serie Qperiodic3

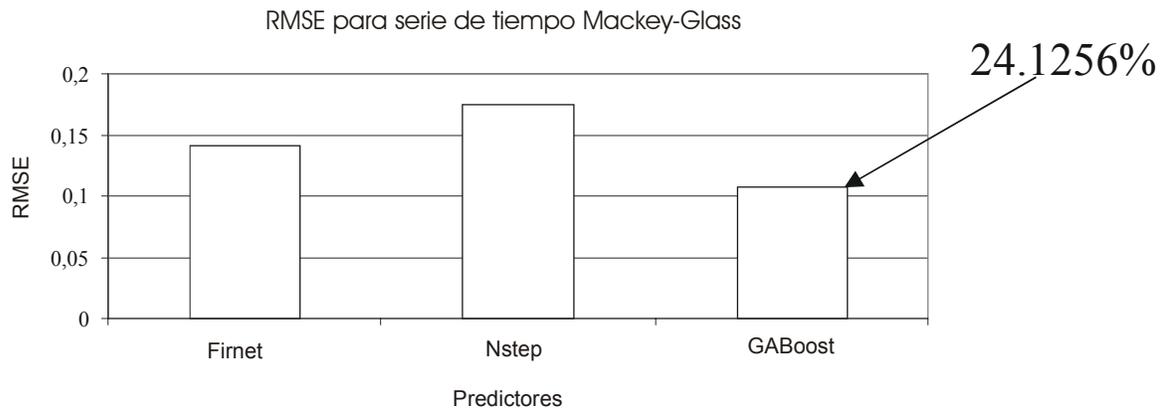


Figura 11-11: Comparación del error RMSE para la serie Mackey-Glass

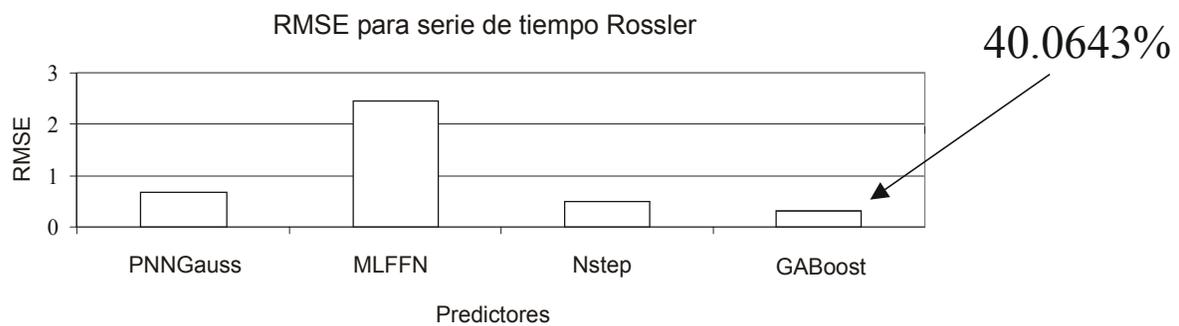


Figura 11-12: Comparación del error RMSE para la serie Rossler

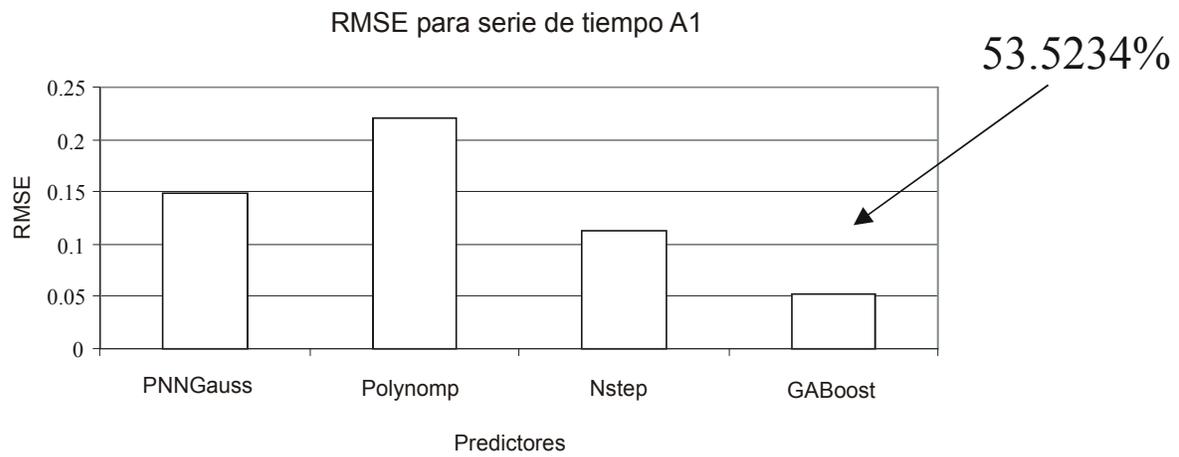


Figura 11-13: Comparación del error RMSE para la serie A1

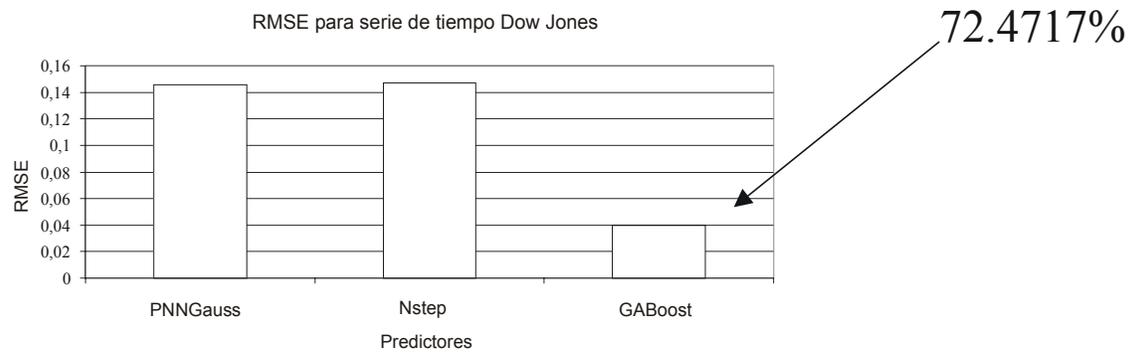


Figura 11-14: Comparación del error RMSE para la serie Dow Jones

de la serie de tiempo, se comparo en forma similar que en el caso del error RMSE, las Figuras 11-15, 11-16, 11-17, 11-18 y 11-19 muestran los resultados de este análisis. Se observa que con excepción de las series Mackey-Glass y A1, el error BE de la combinación de predictores fue menor que el correspondiente a los predictores individuales, lo que indica que la combinación de modelos se adapta mejor a la dinámica de las series de tiempo predichas.

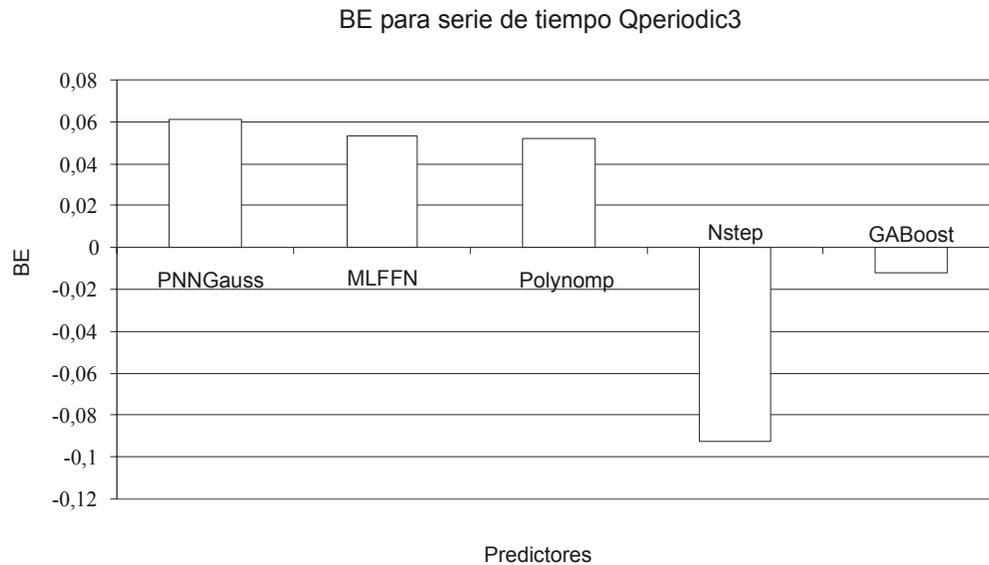


Figura 11-15: Comparación del error BE para la serie Qperiodic3

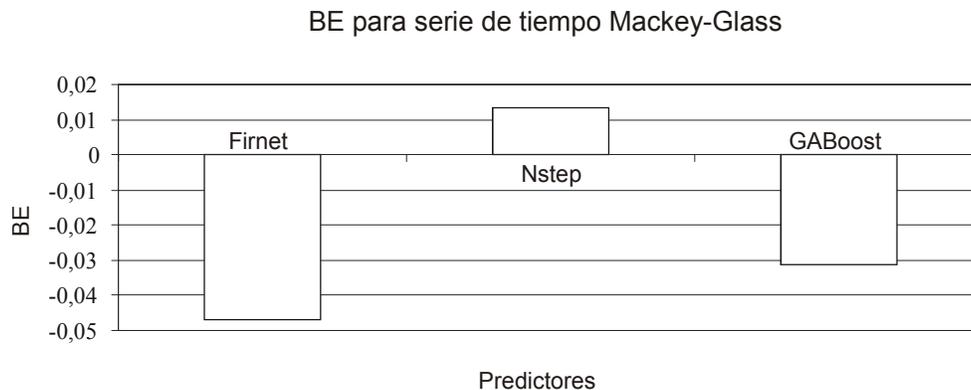


Figura 11-16: Comparación del error BE para la serie Mackey-Glass

Para estudiar el comportamiento del Error Local (LRSE) se gráfico para cada serie de tiempo

### BE para serie de tiempo Rossler

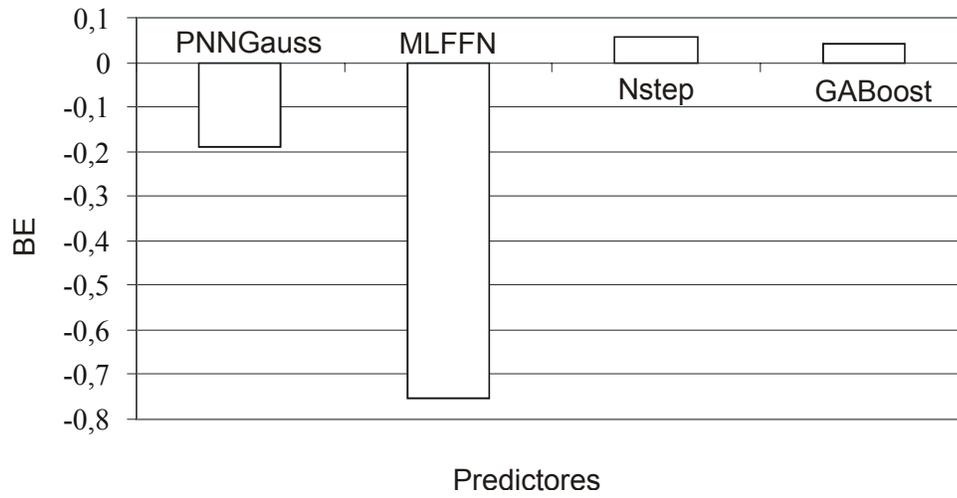


Figura 11-17: Comparación del error BE para la serie Rossler

### BE para serie de tiempo A1

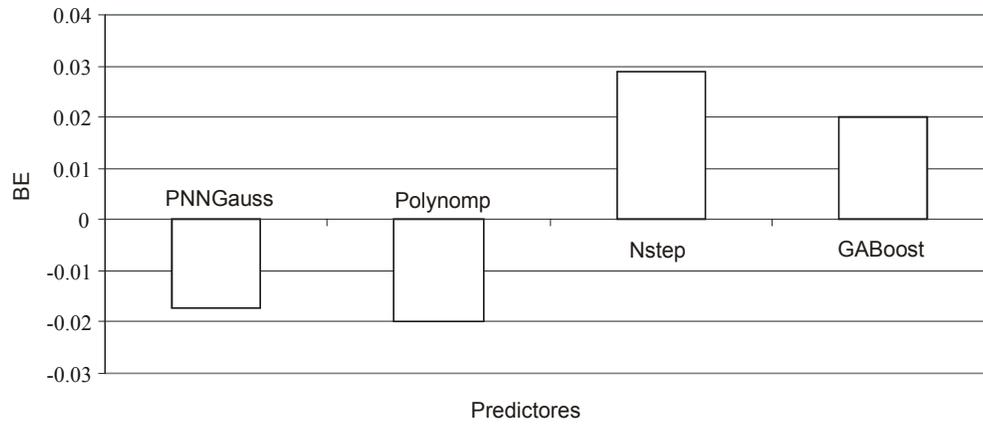


Figura 11-18: Comparación del error BE para la serie A1

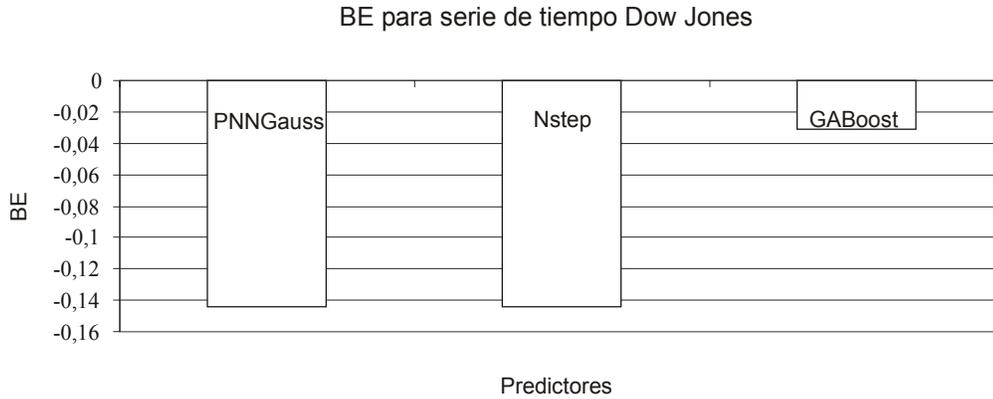


Figura 11-19: Comparación del error BE para la serie Dow Jones

el LRSE de los predictores individuales y se comparo con el correspondiente a la predicción con *GABoost*. Las Figuras 11-20, 11-21, 11-22, 11-23 y 11-24; presentan los resultados del comportamiento del error local de la predicción, en todos los casos se observa que el error local de la combinación de predictores (línea continua en las figuras) se mantiene en el límite inferior respecto a los errores locales de los predictores individuales. Esto muestra que la propagación del error disminuye gracias a la combinación de los predictores.

Finalmente se hizo la comparación del error RMSE para *GABoost* con respecto al error correspondiente de la combinación de los predictores con el promedio simple, el cual es uno de los mejores métodos para mejorar la predicción de series de tiempo y es usado como referencia en la evaluación de técnicas de combinación de predictores. Las Figuras 11-25, 11-26, 11-27, 11-28 y 11-29; muestran los resultados de esta comparación. Se observa que con excepción de la serie caótica Mackey-Glass, el error RMSE fue menor para la predicción utilizando la técnica *GABoost* que la predicción producto del promedio simple, los porcentajes de cambio del error RMSE de *GABoost* respecto del error RMSE del promedio simple se muestran en cada figura.

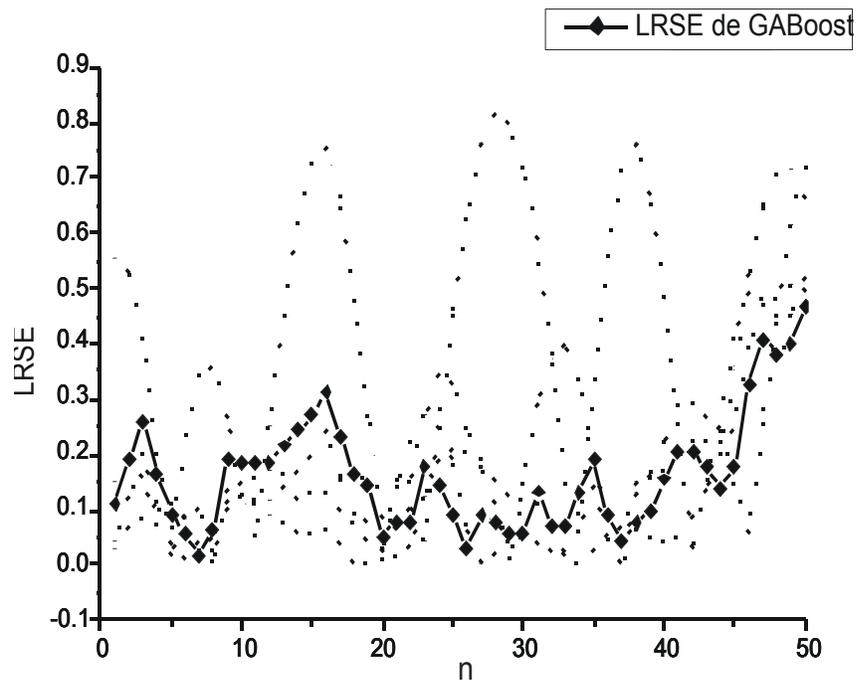


Figura 11-20: Comparación del error local de predicción para la serie Qperiódica de los predictores individuales y la combinación con GABOost

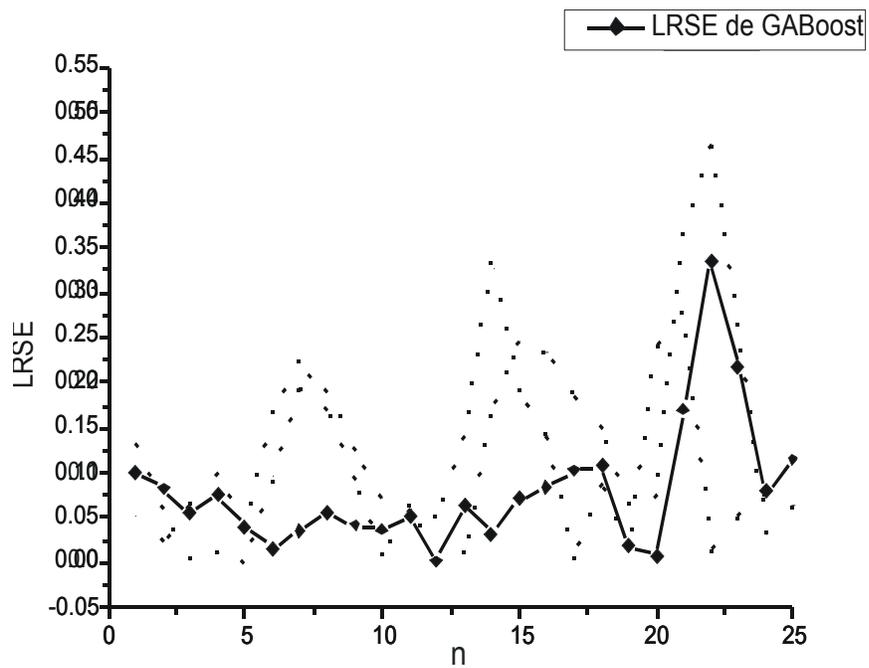


Figura 11-21: Comparación del error local de predicción para la serie Mackey-Glass de los predictores individuales y la combinación con GABoost

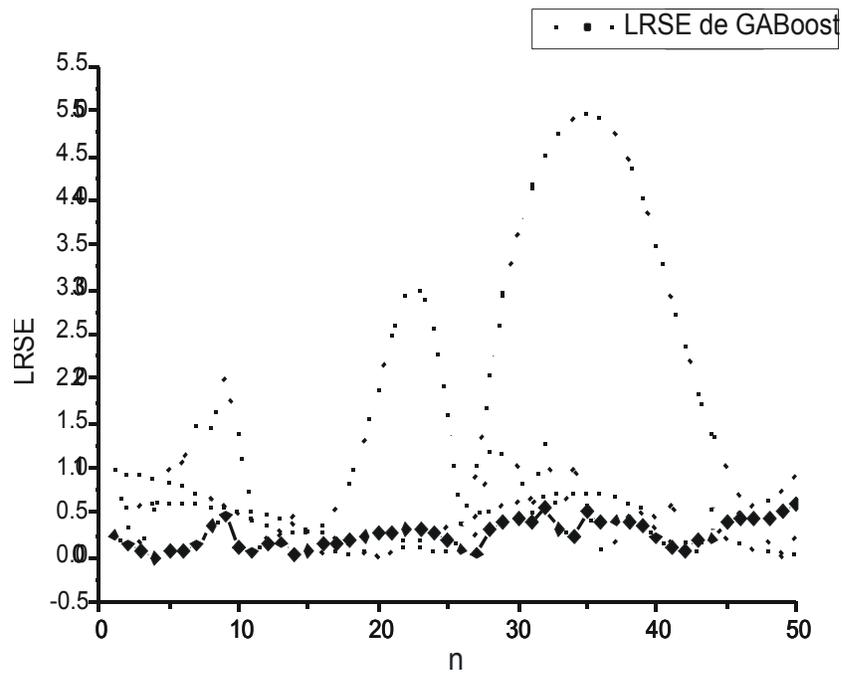


Figura 11-22: Comparación del error local de predicción para la serie Rössler de los predictores individuales y la combinación con GABoost

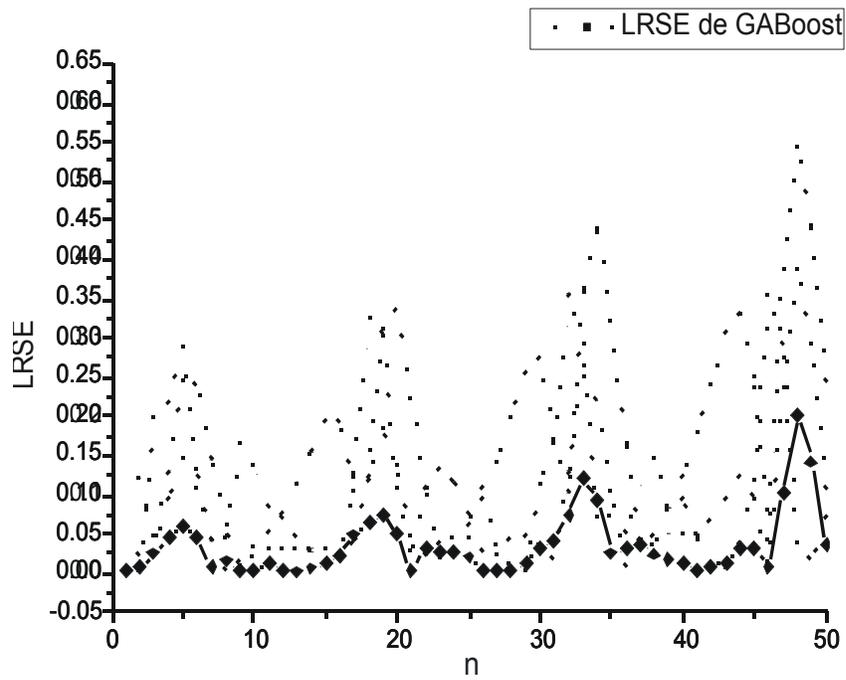


Figura 11-23: Comparación del error local de predicción para la serie A1 de los predictores individuales y la combinación con GABoost

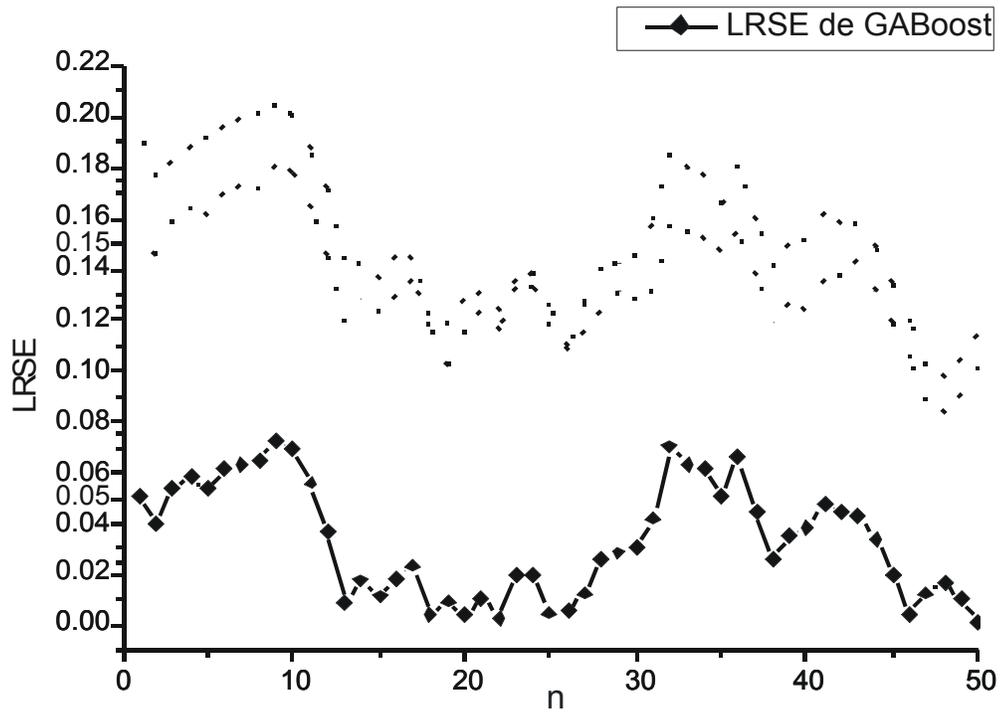


Figura 11-24: Comparación del error local de predicción para la serie Dow Jones de los predictores individuales y la combinación con GABOost

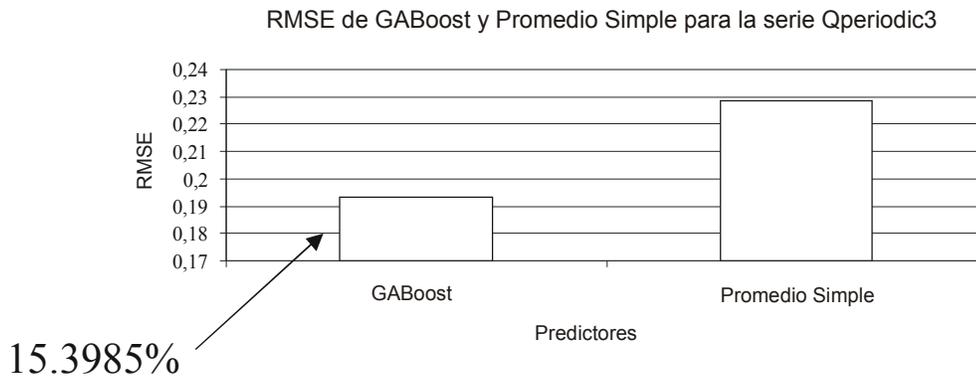


Figura 11-25: Comparación del error RMSE de la combinación de predictores con GABOost y promedio simple para la serie de tiempo Qperiodic3

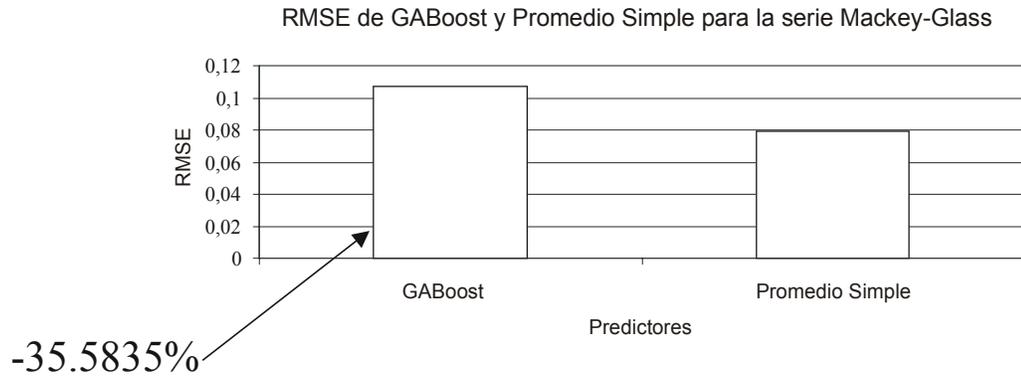


Figura 11-26: Comparación del error RMSE de la combinación de predictores con GABoost y promedio simple para la serie de tiempo Mackey-Glass

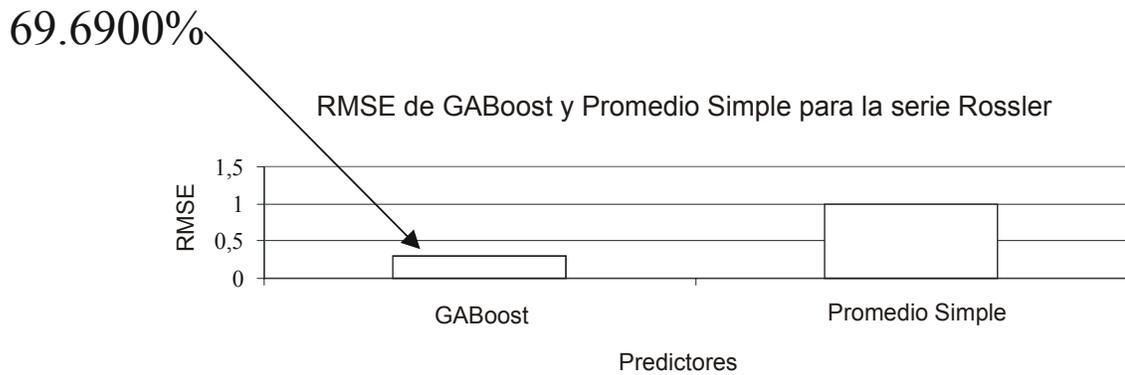


Figura 11-27: Comparación del error RMSE de la combinación de predictores con GABoost y promedio simple para la serie de tiempo Rossler

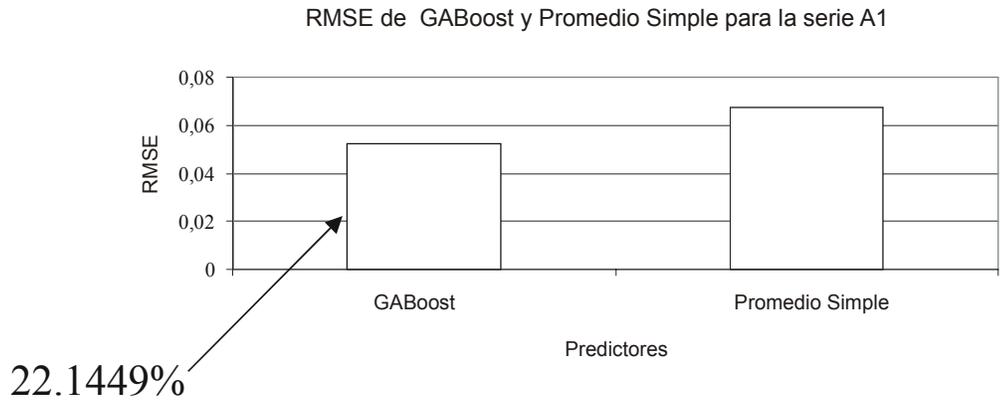


Figura 11-28: Comparación del error RMSE de la combinación de predictores con GABoost y promedio simple para la serie de tiempo A1

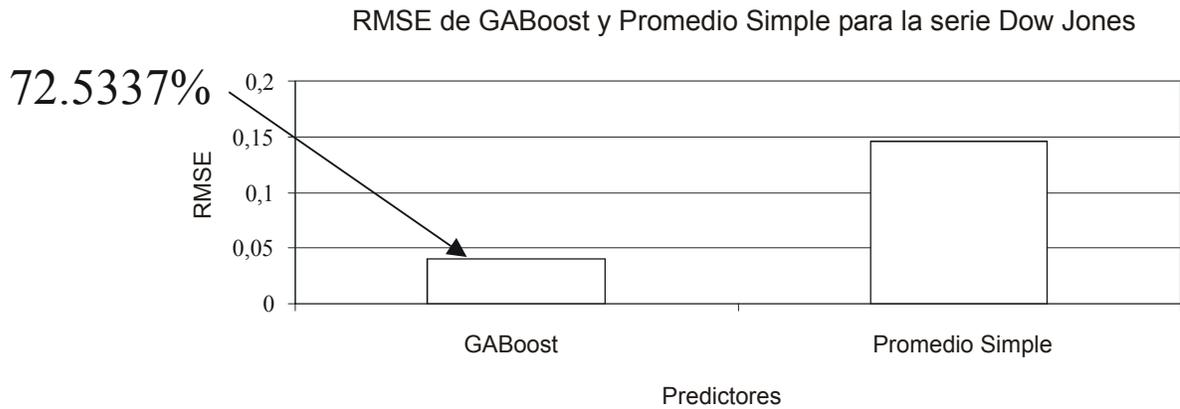


Figura 11-29: Comparación del error RMSE de la combinación de predictores con GABoost y promedio simple para la serie de tiempo Dow Jones

## 11.4 CombFEC: Combinación de Predictores con una Función de Pesos de Error y Correlación

### 11.4.1 La Función de Pesos de Error y Correlación

La segunda técnica denominada *CombFEC* se basa en explotar la información proporcionada por los diferentes errores de predicción y por el grado de correlación entre cada predictor y los datos de la serie original. La función de pesos se basa en modificar la función *SOFTMAX* [187, 188, 189],

$$c_j(x) = \frac{\exp(z_j(x))}{\sum_{k=1}^M \exp(z_k(x))} \quad (11.17)$$

la cual es utilizada en el entrenamiento de redes neuronales, ya que permite construir pesos que cumplan las condiciones:

$$0 \leq c_j(x) \leq 1 \quad \text{y} \quad \sum_{j=1}^M c_j(x) = 1 \quad (11.18)$$

La modificación consiste en la construcción de un exponente que permita introducir la información sobre los errores de la predicción de una serie de tiempo para cada modelo de predicción utilizado y la información sobre el grado de correlación de un modelo con relación a la serie de tiempo predicha. Los errores de predicción (definidos con anterioridad, ver la sección 8.2) aportan información sobre la expansión del error local de predicción por medio del error *MAE*, el rendimiento global del modelo al predecir por medio del error *RMSE* y el nivel de sub o sobreaprendizaje del modelo sobre la dinámica de la serie de tiempo por medio del error *BE*. Por su parte, el grado de correlación del modelo respecto de la serie de tiempo se mide con el coeficiente de correlación de Pearson (definido con anterioridad, ver la sección 6.4) que se identifica como *COPS* o correlación de predictor-serie. Esta correlación se obtiene en la etapa de entrenamiento, a partir de un conjunto de datos predichos por un modelo (modelo de la serie) y del conjunto correspondiente de datos originales. Con estos cuatro elementos se construye la *Función de Error y Correlación (FEC)* de cada modelo de predicción:

$$fec_j = (1 - MAE_j) + (1 - |BE_j|) + (1 - RMSE_j) + COPS_j \quad (11.19)$$

de esta forma la función de pesos para cada modelo de predicción es:

$$W_j = \frac{\exp(fec_j)}{\sum_{k=1}^M \exp(fec_k)} \quad (11.20)$$

### 11.4.2 Descripción del Algoritmo de la Técnica CombFEC

Con la función anterior, se calculan los pesos correspondientes para la combinación de los predictores (o modelos de predicción) de la serie de tiempo a la que se le desea mejorar el resultado de su predicción, en la Figura 11-30 se muestra la descripción del algoritmo completo de la técnica *CombFEC*.

### 11.4.3 Resultados Experimentales

La evaluación de la técnica *CombFEC* se realizó con el mismo conjunto experimental de series de tiempo y de técnicas de predicción utilizados en la evaluación de la técnica *GABOOST* (ver las Figuras 11-2, 11-3, 11-4, y 11-5) de esta forma es posible comparar ambas técnicas, los resultados de dicha comparación se presentan más adelante en la sección 11.5 correspondiente a la discusión de los resultados del presente capítulo.

En las Figuras 11-31, 11-32 y 11-33; se muestran las gráficas en las cuales se compara la serie original, la predicción *CombFEC* y la predicción con el mejor predictor individual, los casos presentados corresponden a las series Mackey-Glass, A1 y Dow Jones.

Para verificar la eficiencia de la técnica para reducir el error RMSE se compararon los errores de los predictores individuales con el correspondiente a *CombFEC*, las Figuras 11-34, 11-35, 11-36, 11-37 y 11-38; muestran las comparaciones para las diferentes series de tiempo. El porcentaje que se muestra en cada figura, corresponde a la reducción o aumento del error RMSE de la combinación con respecto al error RMSE del mejor predictor individual. En el caso de las series Mackey-Glass, Rossler y A1 se logro mejorar el resultado de la predicción con respecto al resultado de la mejor predicción individual, en los otros dos casos: Qperiodic3 y

### *Etapa de Entrenamiento*

Búsqueda de pesos para la reducción del error de predicción.

$P_1^{Entrenamiento}, P_2^{Entrenamiento}, \dots, P_M^{Entrenamiento}$  Conjuntos de datos predichos correspondientes al mismo intervalo de predicción, obtenidos a partir de  $M$  modelos de predicción.

Cálculo de los Errores: Error Local Promedio ( $MAE$ ) Error Bias ( $BE$ ) y Error Global Promedio ( $RMSE$ ), a partir del conjunto de datos de entrenamiento.

Cálculo de la Correlación entre los datos Predichos y los datos Originales: Correlación de Predictor-Serie ( $COPS$ ), a partir del conjunto de datos de entrenamiento.

Cálculo del peso correspondiente a cada Predictor con la expresión:

$$W_j = \frac{e^{fec_j}}{\sum_{k=1}^M e^{fec_k}}$$

donde la función  $fec_j$  (*función de error-correlación*) es de la forma:

$$fec_j = (1 - MAE_j) + (1 - ABS(BE_j)) + (1 - RMSE_j) + COPS_j$$

### *Etapa de Prueba*

Uso del conjunto de pesos para la reducción del error de predicción en el conjunto de prueba.

$P_1^{Prueba}, P_2^{Prueba}, \dots, P_M^{Prueba}$  Conjuntos de datos predichos correspondientes al mismo intervalo de predicción, obtenidos a partir de  $M$  modelos de predicción.

$$P_{combinación}(x_i) = \sum_j^M W_j P_j^{Prueba}(x_i)$$

Figura 11-30: Algoritmo CombFEC

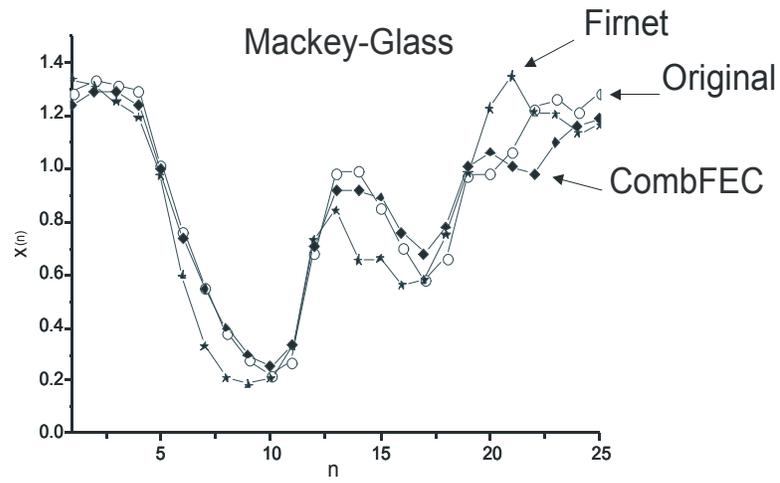


Figura 11-31: Comparación de predicciones de CombFEC y mejor predictor individual para la serie Mackey-Glass

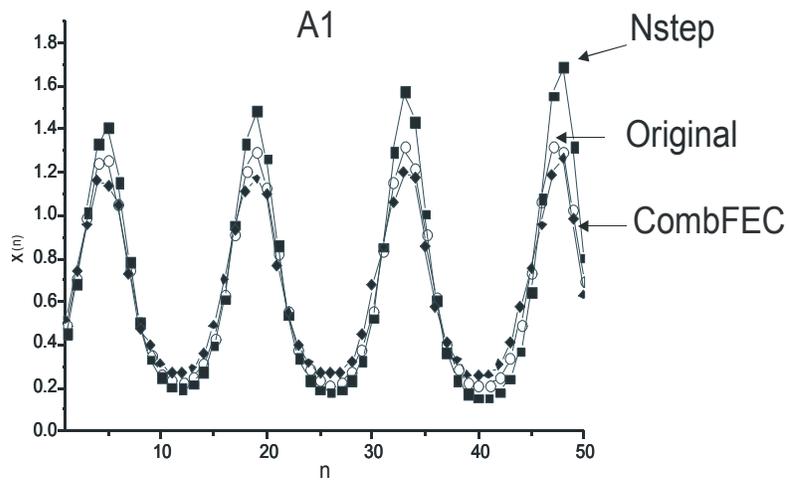


Figura 11-32: Comparación de predicciones de CombFEC y mejor predictor individual para la serie A1

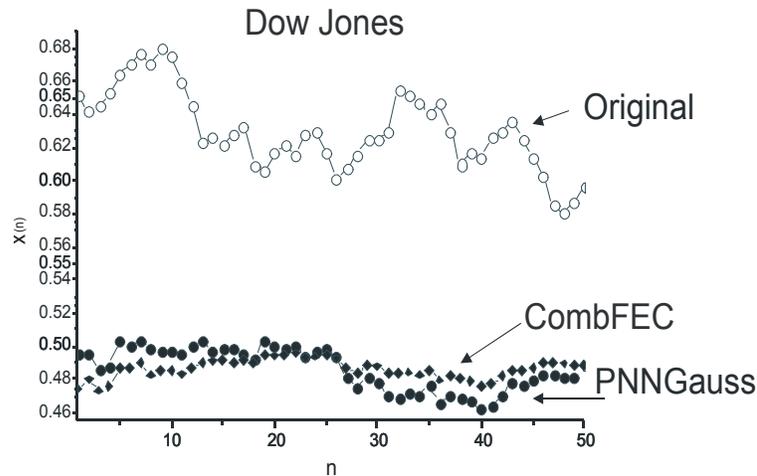


Figura 11-33 Comparación de predicciones de CombFEC y mejor predictor individual para la serie Dow Jones

Dow Jones se observa un pequeño aumento del error de la predicción resultado de combinar los predictores.

El Error Bias (BE) que permite cuantificar si el modelo de predicción se ajusta a la dinámica de la serie de tiempo durante el aprendizaje, se comparo en forma similar que en el caso del error RMSE, las Figuras 11-39, 11-40, 11-41, 11-42 y 11-43; muestran los resultados de este análisis. Con excepción de los casos para las series Mackey-Glass y Dow Jones, en el resto de las series evaluadas se logra reducir el error BE, lo que indica que el modelo resultante de la combinación de los predictores se adapta mejor a la dinámica de la serie de tiempo.

Para estudiar el comportamiento del Error Local (LRSE) se gráfico para cada serie de tiempo el LRSE de los predictores individuales y se comparo con el correspondiente a la predicción con *CombFEC*. Las Figuras 11-44, 11-45, 11-46, 11-47 y 11-48; presentan los resultados del comportamiento del error local de la predicción. Se observa, que con excepción de la serie de tiempo Dow Jones, para el resto de las series de tiempo el error local de la combinación de predictores (línea continua en las figuras), se mantiene en el límite inferior con relación a los errores locales correspondientes a los predictores individuales. La combinación de predictores permite entonces que la expansión del error sea minimizada.

Finalmente se hizo la comparación del error RMSE para *CombFEC* con respecto al error

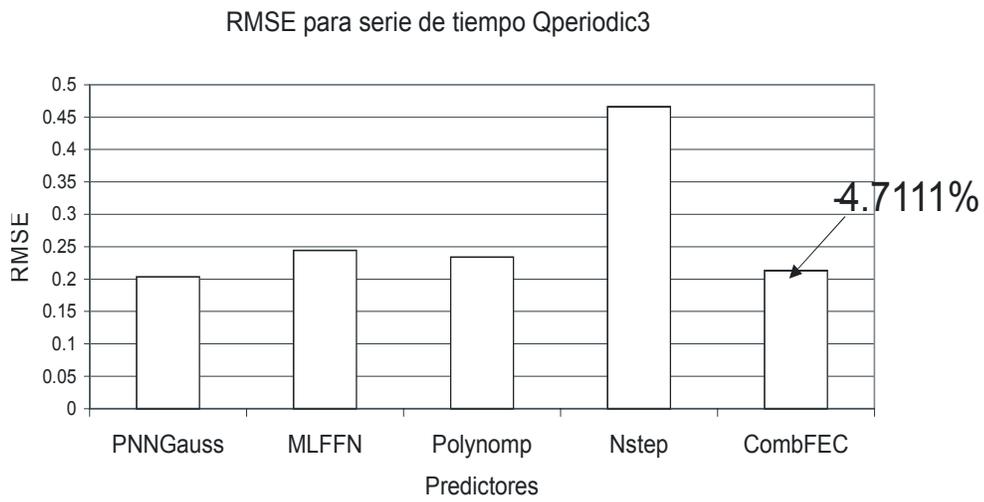


Figura 11-34: Comparación del error RMSE para la serie Qperiodic3

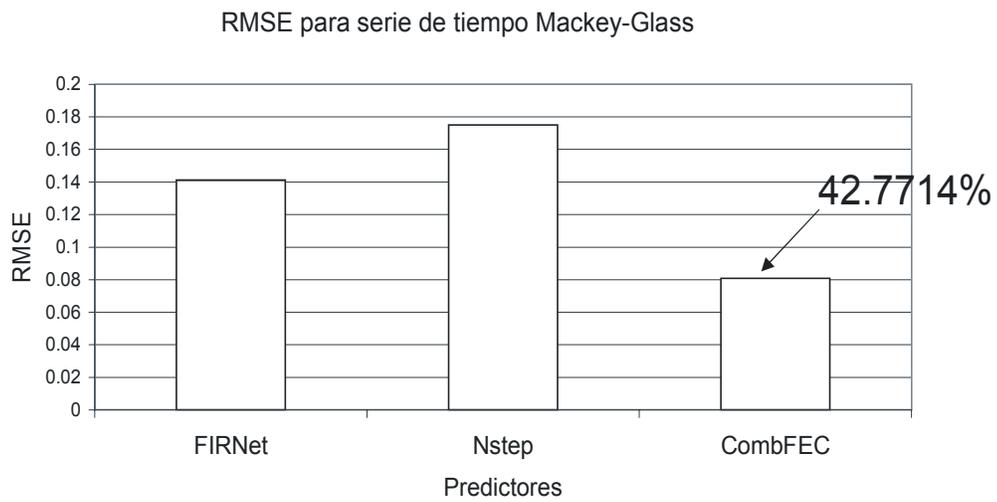


Figura 11-35: Comparación del error RMSE para la serie Mackey-Glass

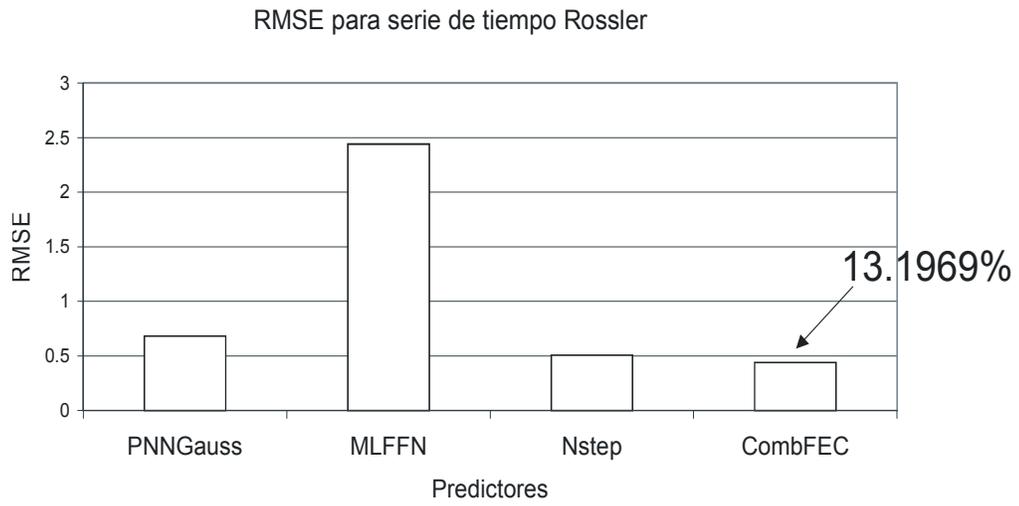


Figura 11-36: Comparación del error RMSE para la serie Rossler

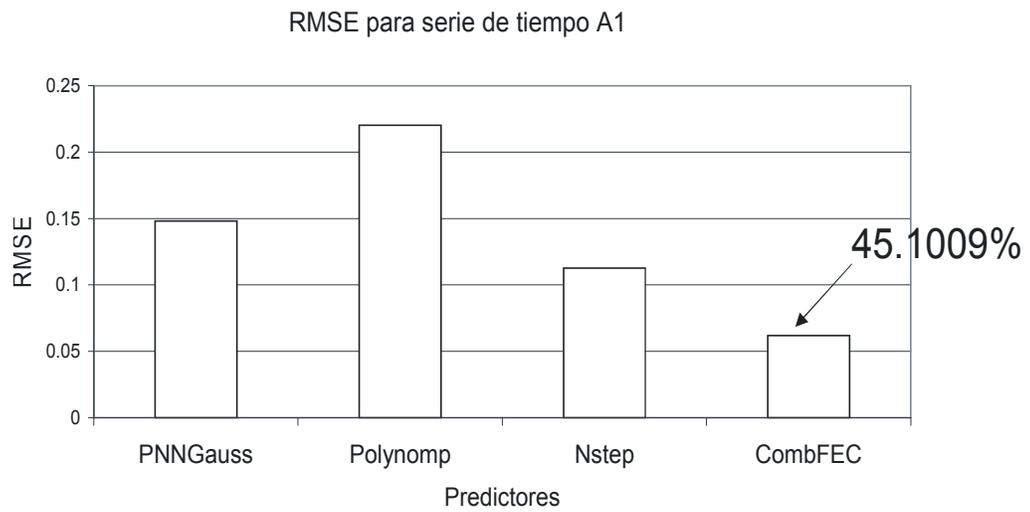


Figura 11-37: Comparación del error RMSE para la serie A1

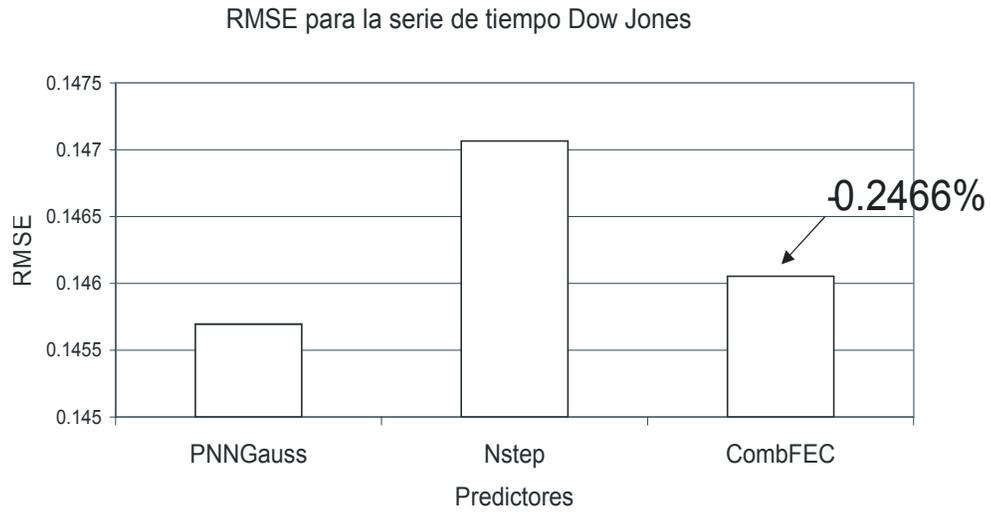


Figura 11-38: Comparación del error RMSE para la serie Dow Jones

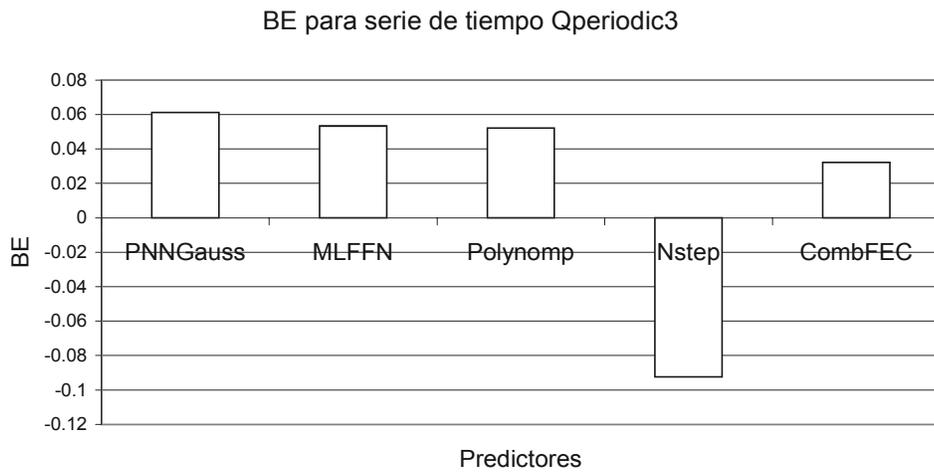


Figura 11-39: Comparación del error BE para la serie Qperiodic3

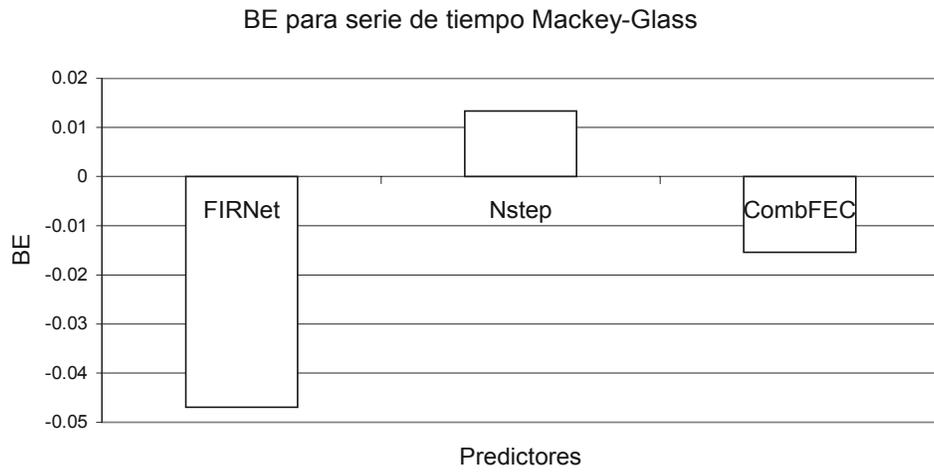


Figura 11-40: Comparación del error BE para la serie Mackey-Glass

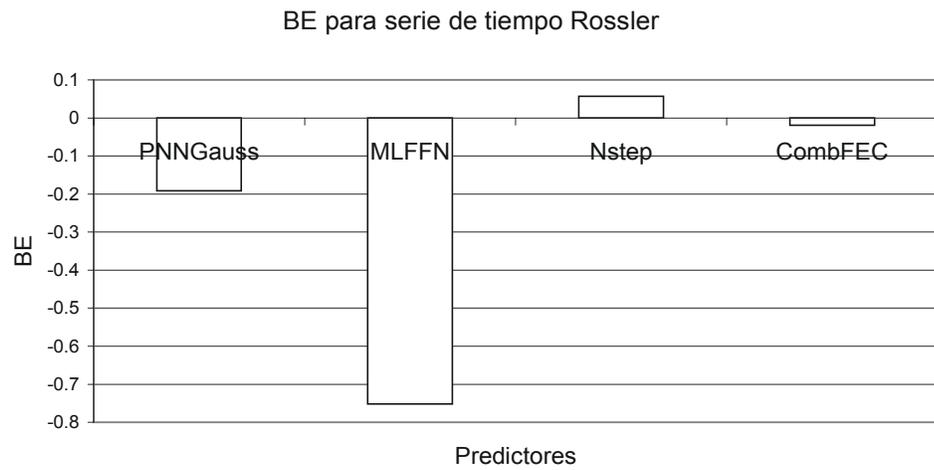


Figura 11-41: Comparación del error BE para la serie Rossler

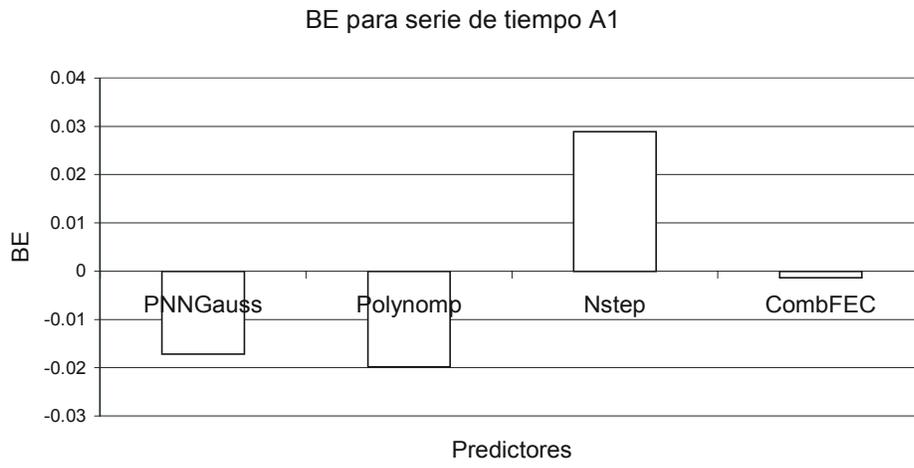


Figura 11-42: Comparación del error BE para la serie A1

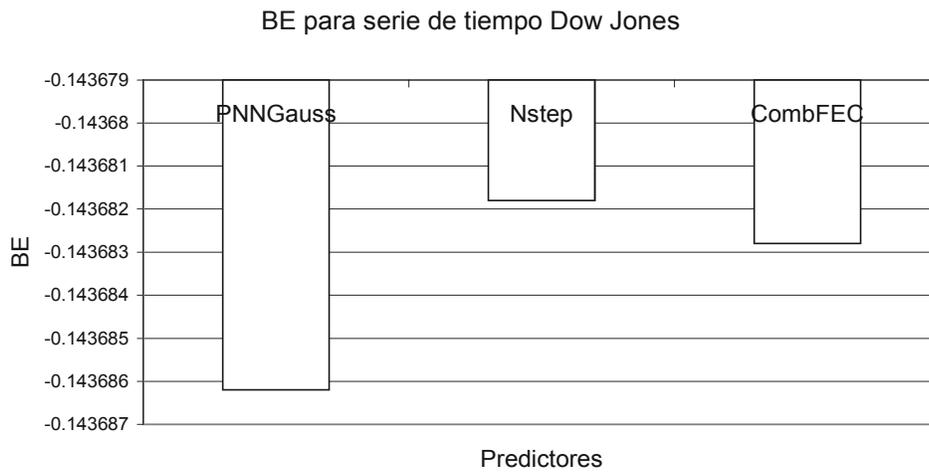


Figura 11-43: Comparación del error BE para la serie Dow Jones

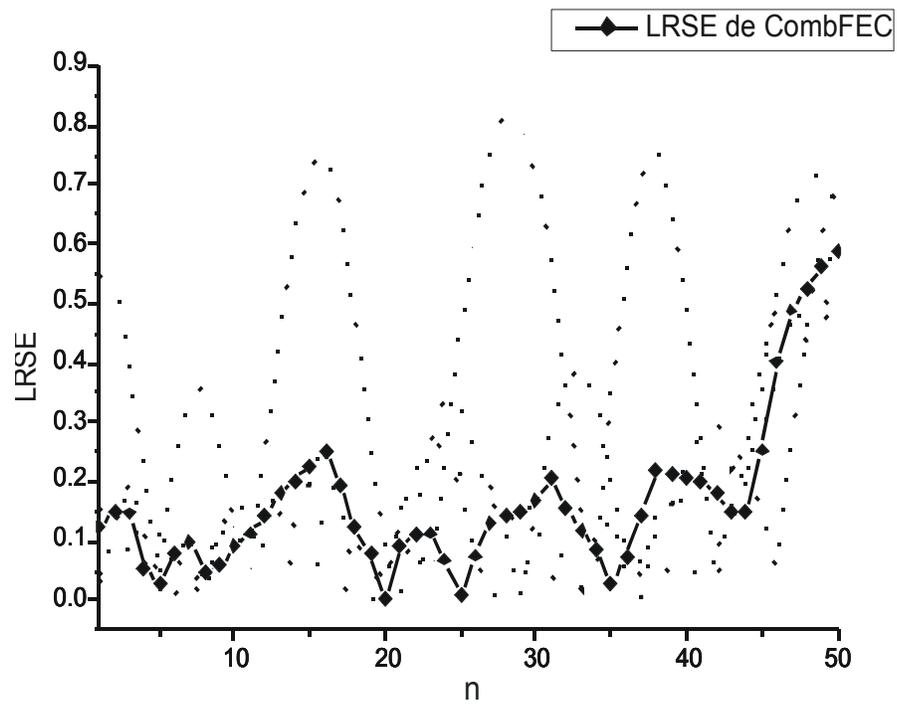


Figura 11-44: Comparación del error local de predicción para la serie Q periódica de los predictores individuales y la combinación con CombFEC

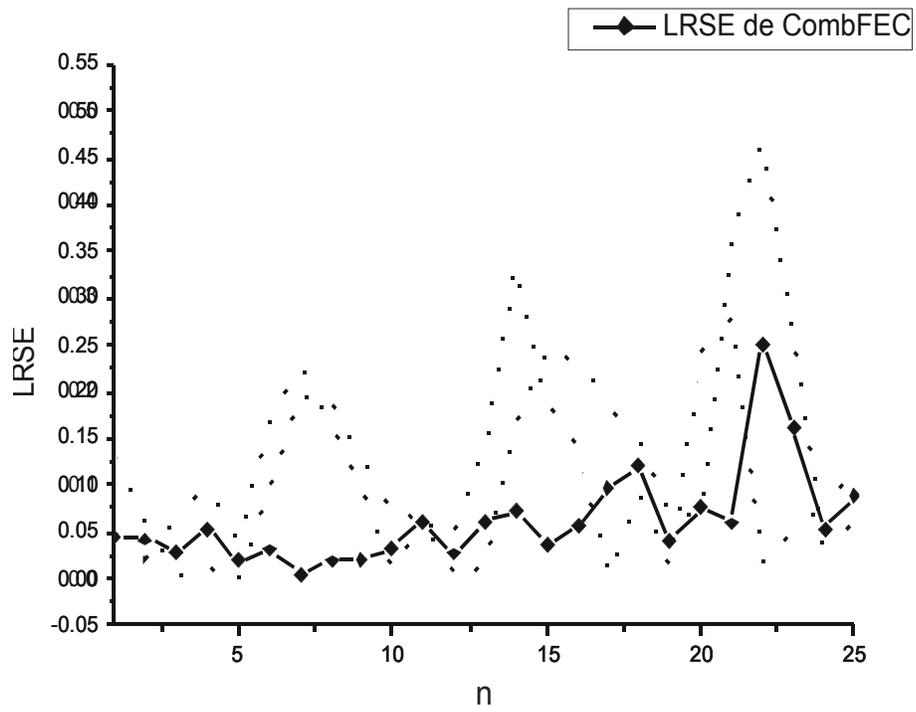


Figura 11-45: Comparación del error local de predicción para la serie Mackey-Glass de los predictores individuales y la combinación con CombFEC

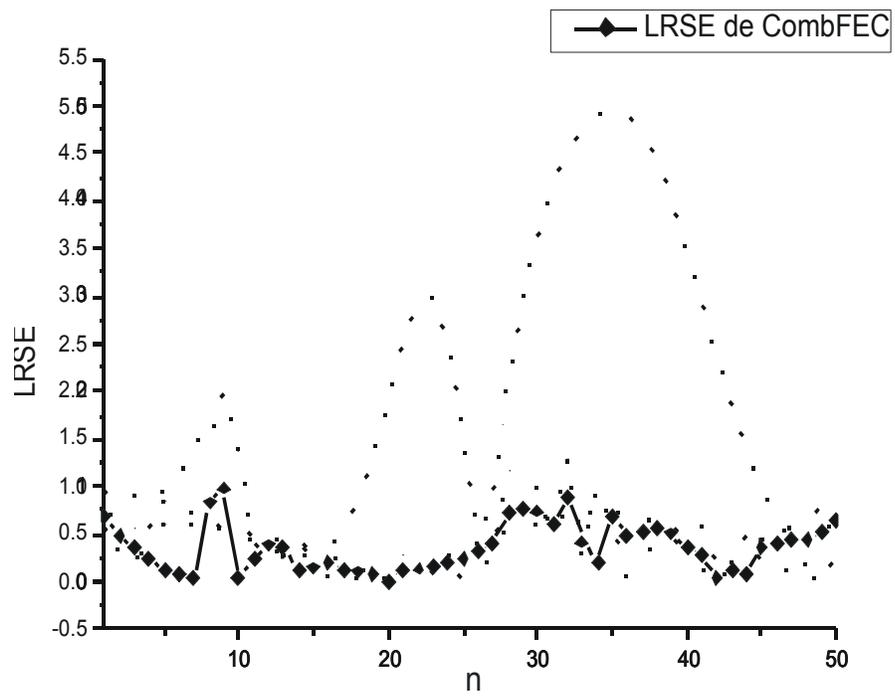


Figura 11-46: Comparación del error local de predicción para la serie Rossler de los predictores individuales y la combinación con CombFEC

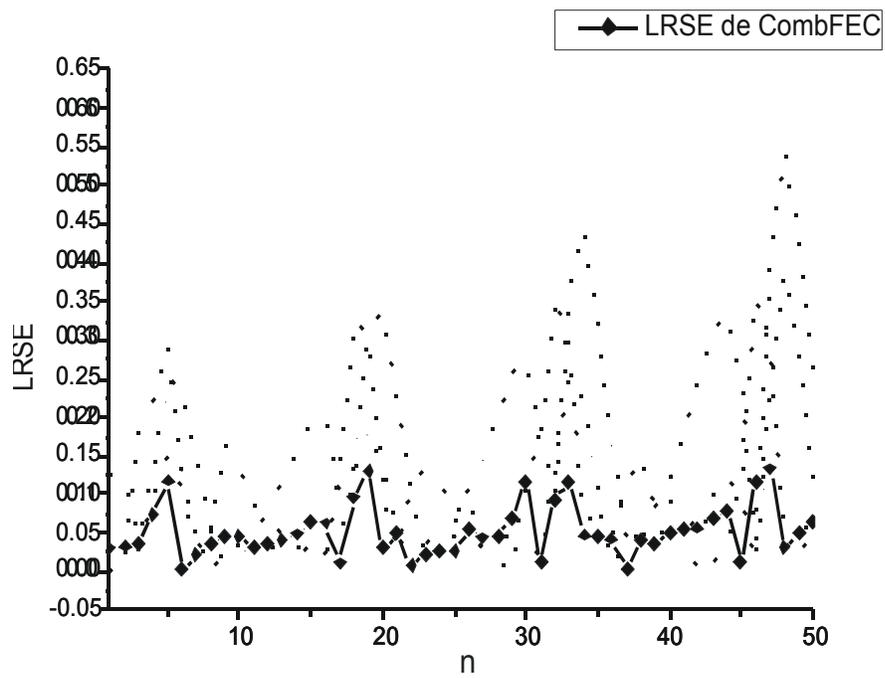


Figura 11-47: Comparación del error local de predicción para la serie A1 de los predictores individuales y la combinación con CombFEC

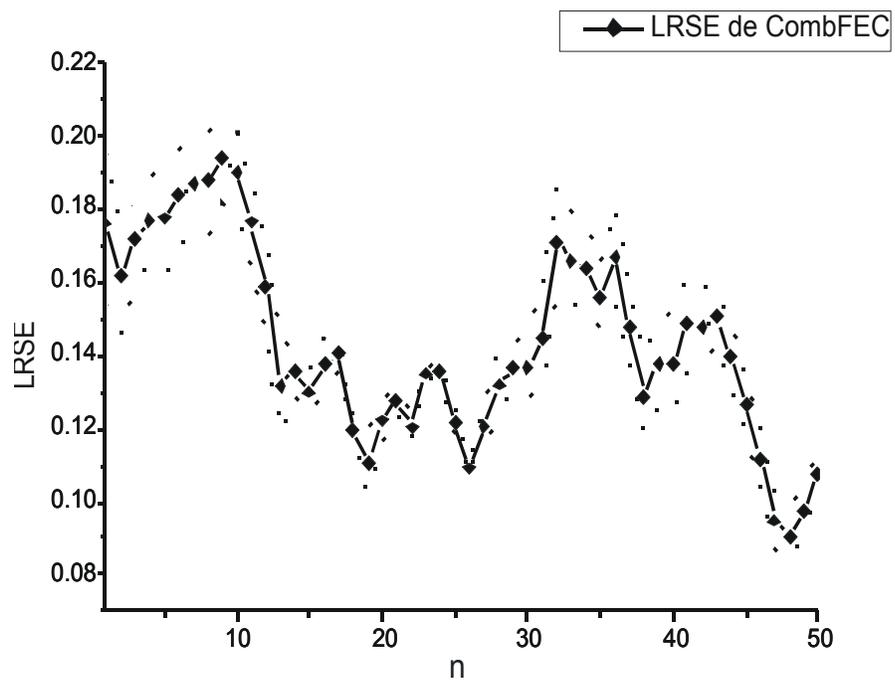


Figura 11-48: Comparación del error local de predicción para la serie Dow Jones de los predictores individuales y la combinación con CombFEC

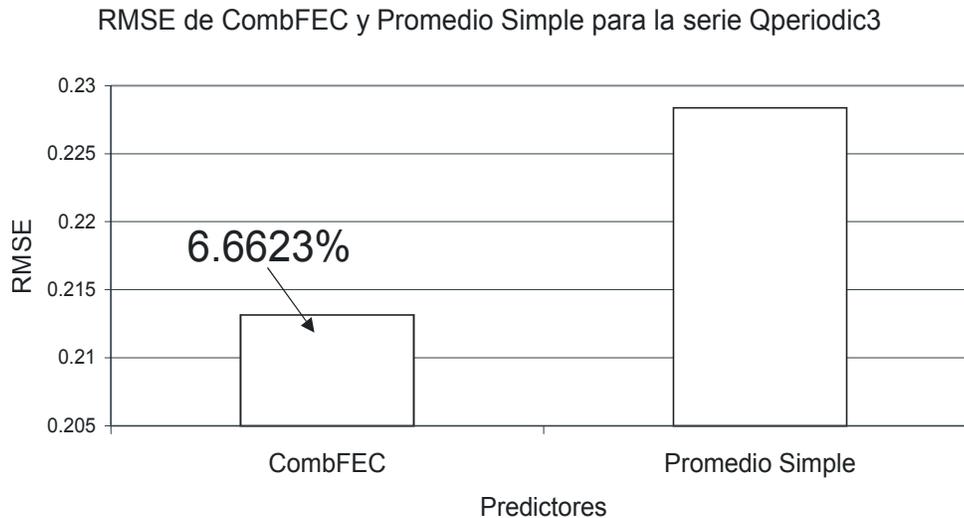


Figura 11-49: Comparación del error RMSE de la combinación de predictores con CombFEC y promedio simple para la serie de tiempo Qperiodic3

correspondiente de la combinación de los predictores con el promedio simple, el cual es uno de los mejores métodos para mejorar la predicción de series de tiempo y es usado como referencia en la evaluación de técnicas de combinación de predictores. Las Figuras 11-49, 11-50, 11-51, 11-52 y 11-53; muestran los resultados de esta comparación. Para las series Qperiodic3, Rossler y A1 se observa que el error RMSE es menor que en el caso de la combinación de predictores usando un promedio simple, en el caso de la serie Mackey-Glass se observa un error RMSE un poco mayor que el correspondiente al promedio simple, por último en el caso de la serie Dow Jones el error RMSE es ligeramente mayor al correspondiente a la combinación de predictores con promedio simple. Los porcentajes de cambio del error RMSE de *CombFEC* respecto del error RMSE del promedio simple se muestran en cada figura.

## 11.5 Discusión

En este capítulo se presentó la descripción de dos técnicas de combinación de predictores denominadas *GABoost* y *CombFEC*. La primera técnica, tiene como base la construcción de una combinación de modelos de predicción a partir de votos pesados, en analogía con el algoritmo de aprendizaje Boosting. Pero a diferencia del algoritmo Boosting, que requiere de la construcción

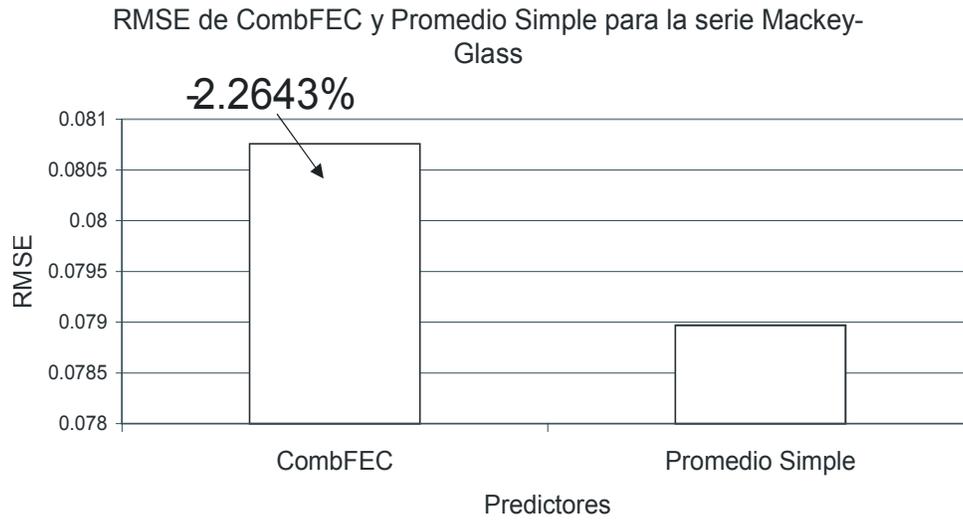


Figura 11-50: Comparación del error RMSE de la combinación de predictores con CombFEC y promedio simple para la serie de tiempo Mackey-Glass

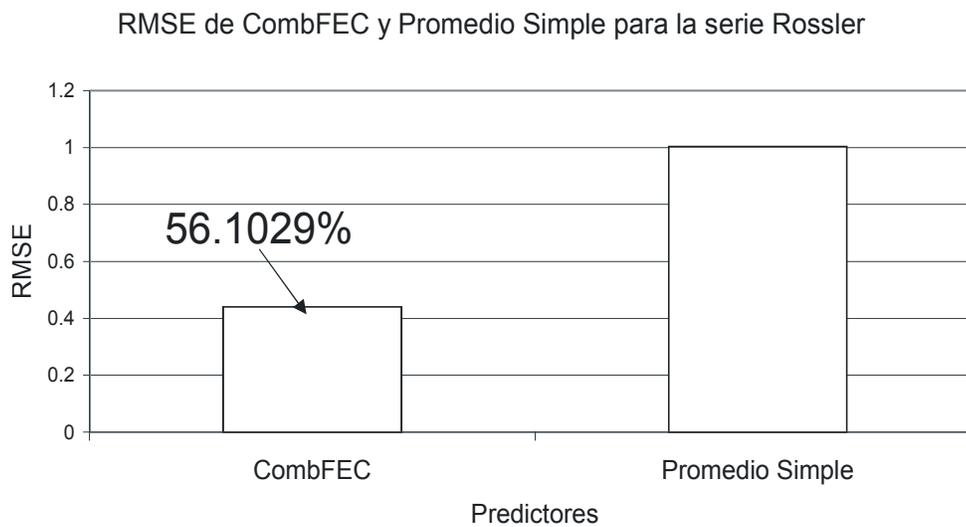


Figura 11-51: Comparación del error RMSE de la combinación de predictores con CombFEC y promedio simple para la serie de tiempo Rossler

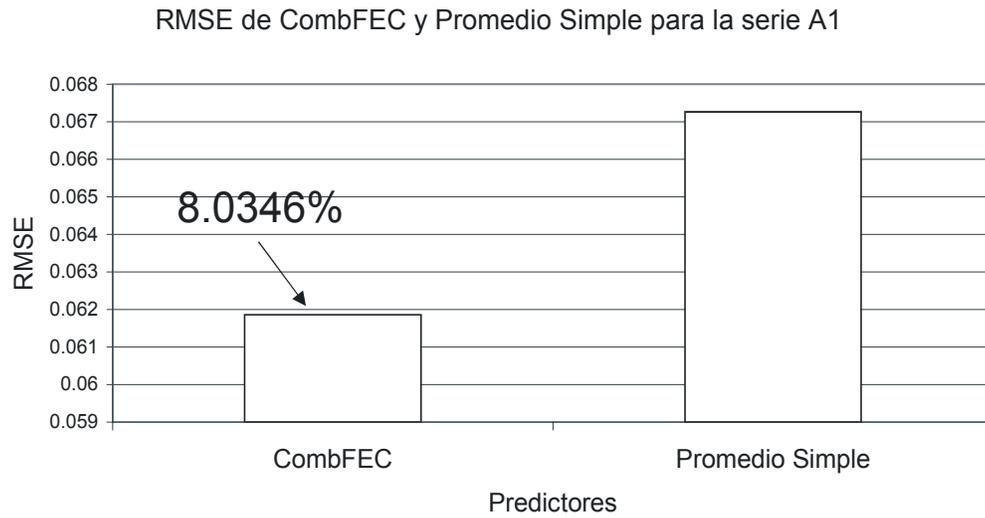


Figura 11-52: Comparación del error RMSE de la combinación de predictores con CombFEC y promedio simple para la serie de tiempo A1

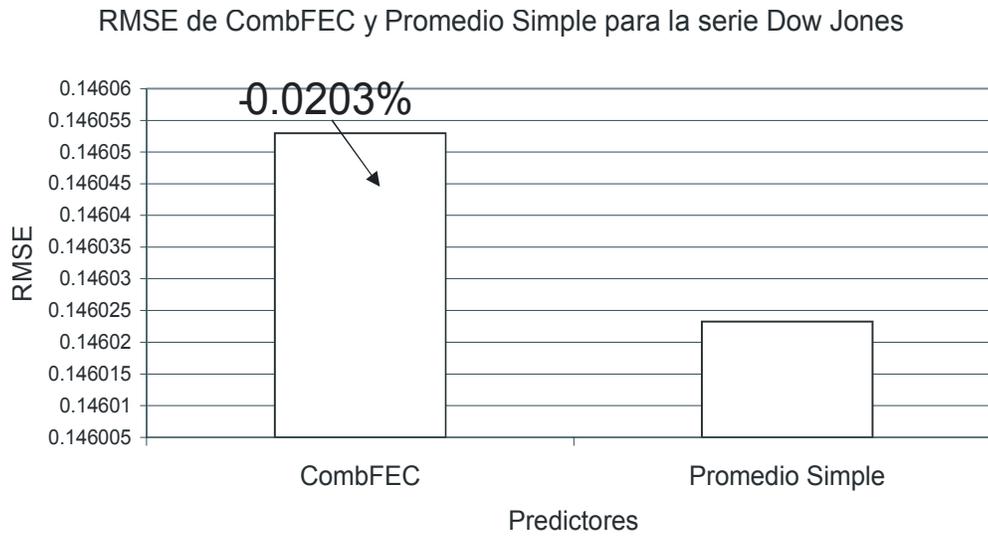


Figura 11-53: Comparación del error RMSE de la combinación de predictores con CombFEC y promedio simple para la serie de tiempo Dow Jones

Técnica de combinación de predictores	Reduce RMSE respecto de mejor predictor individual	Reduce RMSE respecto de combinación con promedio simple	Total de éxitos
GABOOST	Qperiodic3	Qperiodic3	4
	Mackey-Glass		
	Rossler	Rossler	
	A1	A1	
	Dow Jones	Dow Jones	
CombFEC		Qperiodic3	2
	Mackey-Glass		
	Rossler	Rossler	
	A1	A1	

Figura 11-54: Tabla comparativa de las técnicas de combinación de predictores

en forma explícita de la función de distribución de pesos, en el algoritmo desarrollado se hace uso de la búsqueda con algoritmo genético canónico de los pesos óptimos de la combinación. La segunda técnica, hace uso de la función SOFTMAX modificada usando una función de error y correlación (*FEC*) para encontrar los pesos para combinar a los predictores o modelos de predicción. La función *FEC*, introduce información sobre el rendimiento de los predictores individuales a través de tres coeficientes de error y un coeficiente de correlación. Ambas técnicas se evaluaron con un conjunto de series de tiempo de diferente predictibilidad y comportamiento dinámico.

Los resultados muestran que la técnica *GABOOST* es capaz de mejorar la predicción de las series de tiempo, reduciendo el error RMSE de la predicción con respecto a los mejores predictores individuales para cada serie evaluada y también con respecto a la combinación de predictores usando el promedio simple (ver tabla en la Figura 11-54). La combinación de modelos de predicción con *GABOOST* permite además disminuir el error bias (BE), lo cual muestra que el modelo resultante se adapta mejor a la dinámica de la serie de tiempo predicha.

Con respecto a la técnica *CombFEC* aunque logra reducir el error de predicción en comparación con el mejor predictor individual y con la combinación de predictores con el promedio simple, su eficiencia no es tan buena como en el caso de *GABOOST* (ver tabla en la Figura 11-54), sin embargo debe tomarse en cuenta que la influencia que el método usado para combinar los

predictores (i. e. la forma de obtener los pesos óptimos de la combinación) tiene sobre la reducción del error de predicción no se ha estudiado en la literatura, por lo que no es claro como afecta la forma en que se combinan los modelos en el resultado final, esto sin duda es materia de investigaciones futuras.

## Parte IV

# Conclusiones

## Capítulo 12

# Conclusiones y Líneas de Trabajo Futuro

### 12.1 Conclusiones

En este trabajo de tesis se ha desarrollado *un estudio experimental sobre la predictibilidad de series de tiempo*, para llevarlo a cabo fue necesario como primer paso, la identificación de un conjunto de parámetros que proporcionen información sobre la dinámica y estructura de las series de tiempo. La identificación de estos parámetros se hizo sobre la base de una evaluación sistemática de los mismos para un conjunto de series de tiempo de orígenes diversos, y que son utilizadas en forma estándar en la literatura para el estudio del problema de predicción. El segundo paso, fue el desarrollo de una métrica para medir y posteriormente analizar la predictibilidad de las series. Se propusieron varias métricas y a partir de la evaluación y comparación de las mismas, se encontró que la métrica denominada  $CDP_2$  es la más eficiente para medir la predictibilidad, ya que presentó la correlación más alta entre las métricas desarrolladas en este trabajo con respecto a las variables de referencia (número de predicciones para cada serie de tiempo,  $NumPred$  y número de modelaciones para cada serie de tiempo,  $NumMod$ ). En la construcción de esta métrica, se usaron un conjunto de parámetros dinámicos ortogonales que son ampliamente utilizados en la caracterización de las series de tiempo.

Para validar las métricas fue necesario desarrollar un conjunto de experimentos para evaluar la predicción y el modelado de series de tiempo utilizando un conjunto de técnicas basadas en

diferentes modelos de predicción y de modelado. Para la evaluación de las técnicas se desarrolló una metodología experimental para medir el éxito de cada técnica ya sea para predecir o modelar las series de tiempo.

En la Figura 12-1 se presenta un diagrama que sintetiza las metodologías experimentales desarrolladas en la presente tesis:

- 1 Medición y caracterización de la predictibilidad de las series de tiempo.
- 2 Evaluación de técnicas para predicción o modelado.

Las metodologías han sido descritas en detalle a lo largo de los diferentes capítulos de la presente tesis y en el capítulo 13 de apéndices se incluye una guía que describe los pasos de cada metodología.

En relación con los resultados experimentales se encontró que:

- 1 Para el caso de la predicción, los modelos de alcance local y que operan en el espacio fase poseen una mayor capacidad de predicción, estos modelos tienen como origen a la Teoría de Sistemas Dinámicos No Lineales.
- 2 En el caso del modelado los modelos de alcance global y que operan en el espacio real, presentan la mayor capacidad para el modelado de series de tiempo, el origen de estos modelos es diverso: modelos híbridos de Inteligencia Artificial, Teoría de Sistemas Dinámicos No Lineales y Estadística.

En cuanto a los errores de predicción y de modelado de las series de tiempo, el comportamiento general de los errores es independiente de la capacidad de cada modelo para la predicción y el modelado, es decir, el que un modelo sea capaz de predecir o modelar un número alto de series de tiempo no implica, que los errores de la predicción o el modelado sean menores a otros modelos con menor capacidad.

A partir de las matrices de datos correspondientes a la evaluación de las técnicas de predicción y modelado y de la métrica  $CDP_2$  de las series de tiempo, se desarrolló una métrica de capacidad de predicción ( $CCAP$ ) y de modelado ( $CCAM$ ) con la que fue posible medir y evaluar la habilidad de los modelos para la predicción y modelado tomando en cuenta factores

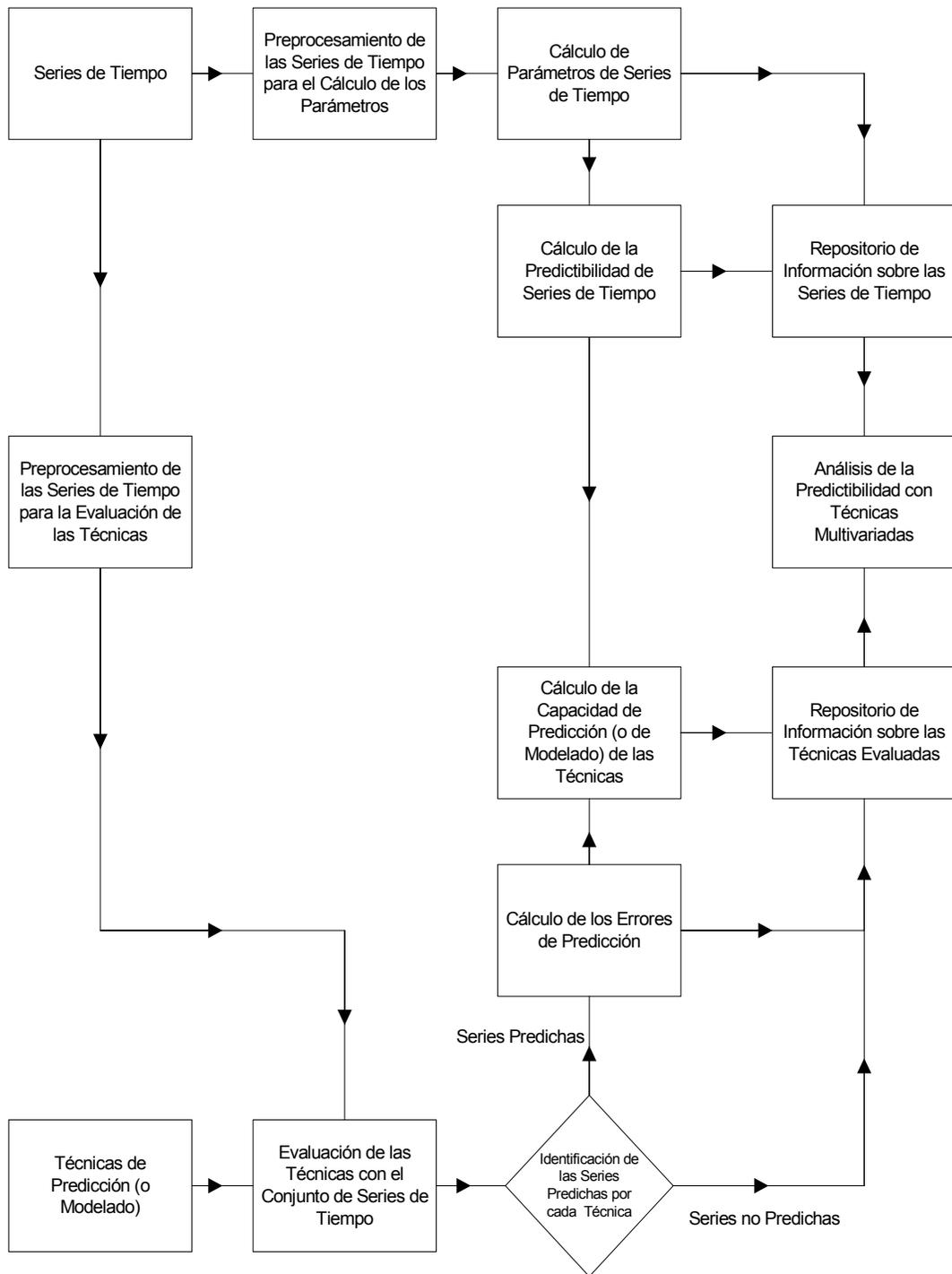


Figura 12-1: Diagrama de las metodologías experimentales desarrolladas para el estudio sobre la predictibilidad de series de tiempo

como la expansión del error de cada modelo, la dificultad de predicción de las series de tiempo y la eficiencia de los modelos para predicción y modelado de conjuntos de series de tiempo.

La medición de la predictibilidad de series de tiempo generó diversos resultados producto de su análisis experimental. La predictibilidad es no lineal con respecto al comportamiento dinámico de las series de tiempo: periódico, cuasi periódico, caótico, complejo y estocástico. La predictibilidad de una serie es independiente del éxito de un modelo o modelos para predecir o modelar dicha serie, el modelo ad hoc para una serie de tiempo puede existir sin embargo no hay la seguridad de poder encontrarlo en el espacio de soluciones, lo que esta métrica proporciona es un valor que permite comparar la predictibilidad entre series de tiempo de orígenes diversos.

Al utilizar diferentes técnicas de análisis multivariado se encontró que: Las series de tiempo se agrupan en tres grandes conjuntos de acuerdo a la magnitud de su predictibilidad (baja, media y alta) y dentro de esos conjuntos existe una jerarquía de similitud entre las series de tiempo.

La predictibilidad presenta una relación con la estructura de las series de tiempo, esta relación se observó utilizando diferentes formas de representación de las series de tiempo, en particular el mapa de recurrencia (VRA) hizo posible observar e identificar patrones de estructura básicos relacionados con diferentes conjuntos de series de tiempo, las series en dichos conjuntos, se agruparon en base a la similitud de sus parámetros dinámicos con los que se calcula su predictibilidad. Dado que el mapa de recurrencia muestra la relación espacial-temporal entre los valores de una serie de tiempo, el mapa muestra en forma visual la dinámica de la estructura de la serie. Entonces, los patrones identificados corresponden a dinámicas básicas las cuales están presentes en series de tiempo de diferente origen.

Al estudiar la relación entre los modelos de predicción y de modelado con respecto a los parámetros dinámicos con los que obtiene la predictibilidad, no se encontró un patrón o patrones comunes de respuesta de los modelos con respecto a los parámetros, cada modelo responde en forma particular a los diferentes parámetros, lo que indica que se adaptan a la dinámica de las series de tiempo de forma diferente, lo cual es de esperarse dada la arquitectura particular de cada modelo. Al considerar la forma en que responden globalmente al conjunto de parámetros se encontró que en el caso de la predicción cada modelo responde en forma diferente, pero en el caso del modelado, los modelos con mayor capacidad de modelar series de tiempo presentan

una respuesta global similar a los parámetros de las series de tiempo.

Por último, dado que no existen modelos de predicción capaces de adaptarse a toda la riqueza de dinámicas presentes en las series de tiempo (*modelos universales*), como lo corroboran los resultados experimentales del presente trabajo, para mejorar la predicción de las series de tiempo se desarrollaron y evaluaron con éxito, dos técnicas de combinación de modelos de predicción: *GABOOST* y *CombFEC*, los cuales aplican ideas sobre el aprendizaje usadas en el área de Aprendizaje de Máquina, para la optimización de la combinación de los modelos de predicción.

Para finalizar, se espera que el conjunto de resultados obtenidos en esta tesis, contribuya a la formación de una base de conocimiento experimental que ayude a una mejor comprensión *del problema de la predicción de series de tiempo*.

## 12.2 Líneas de Trabajo Futuro

A continuación se presentan un conjunto de posibles líneas de trabajo futuro derivadas de esta tesis. Estas líneas de trabajo futuro se presentan ordenadas, de tal forma que cada una de ellas apoya a la siguiente para su desarrollo con éxito.

- Automatizar las metodologías experimentales utilizadas en este trabajo de tesis para mejorar la eficacia en el cálculo de la predictibilidad de series de tiempo.
- Ampliar la base de conocimiento sobre las series de tiempo. Con el análisis de series de tiempo adicionales y el incremento de la base de datos disponibles, se pueden utilizar otras técnicas de extracción de conocimiento para obtener nueva información experimental sobre la predictibilidad.
- Mejorar la métrica de predictibilidad, investigar el uso de nuevos parámetros que caracterizan a las series de tiempo y mejorar la eficiencia de la métrica para cuantificar la predictibilidad de las series de tiempo explorando nuevas formas funcionales de la misma.
- Ampliar la clasificación de estructuras de mapas de recurrencia mediante el análisis de una mayor cantidad de series de tiempo y profundizar en el estudio de la relación estructura-predictibilidad de series de tiempo.

- Estudiar otras formas de representación de series de tiempo por ejemplo: la descomposición empírica de modos [190, 191], y su posible utilidad en la caracterización de la predictibilidad de series de tiempo.
- Ampliar el conjunto de técnicas de predicción evaluadas para extender el alcance del estudio sobre los modelos de predicción y su relación con los parámetros que caracterizan a las series de tiempo.
- Estudiar el proceso de la combinación de diferentes modelos de predicción para tratar de deducir reglas empíricas para la combinación de los mismos.

# Capítulo 13

## Apéndices

### 13.1 Trabajos Derivados de la Tesis (Artículos, Ponencias, Posters, Seminarios)

#### 13.1.1 Listado de Artículos, Ponencias y Posters

1 Título: *Reducción de Ruido y Cálculo de la Dimensión de Información en Series de Tiempo.*

Autores: E. Bautista-Thompson, A. Sánchez-Fernández, E. Gutiérrez-Arias y J. Figueroa-Nazuno.

Presentado en: XLIV Congreso Nacional de Física, Morelia, Michoacán, del 15 al 19 de Octubre del 2001.

Resumen publicado en: *Memorias del XLIV Congreso Nacional de Física*, N. R. Silva González (Editor), Sociedad Mexicana de Física, 2001.

Resumen: El análisis de datos experimentales requiere, en la mayoría de los casos, considerar la influencia del ruido sobre la capacidad para analizar la información generada por el sistema. En particular, en la estimación de la dimensión de información de un sistema no lineal, es necesario aplicar a la serie de tiempo bajo estudio técnicas de filtrado para disminuir el ruido contenido en los datos. En este trabajo, se presentan los resultados de la aplicación a diferentes series de tiempo de un filtro no lineal para la reducción de

ruido basado en proyecciones locales, así como el cálculo de la dimensión de información utilizando el algoritmo de masa fija.

2 Título: *Descubrimiento y Extracción de Subestructuras de Bases de Datos.*

Autores: E. Gutiérrez-Arias, E. Bautista-Thompson y J. Figueroa-Nazuno.

Presentado en: Tercer Congreso Estudiantil de Computación CORE 02, México, D. F. del 13 al 14 de Marzo del 2002.

Artículo publicado en: *Memorias del Tercer Congreso Estudiantil de Computación CORE 02* (formato electrónico), Centro de Investigación en Computación IPN, 2002.

Resumen: Existen varios métodos para la extracción de relaciones de información no conocidas en Bases de Datos. A estos métodos en general se les llama minado de datos (DATA MINING). Este trabajo evalúa un método capaz de descubrir subestructuras repetidas en una base de datos, mejorando la habilidad para interpretar y comprimir los datos. La idea del sistema SUBDUE es encontrar estructuras interesantes y repetitivas en Bases de Datos, a partir de su representación en forma de grafos etiquetados, usando el principio de descripción mínima. Esto es identificar subestructuras capaces de comprimir información por abstracción de eventos y también identificar conceptualmente subestructuras de interés que incrementen las posibles interpretaciones de los datos y tener la posibilidad de crear modelos. La subestructura obtenida por SUBDUE representa nuevas estructuras encontradas en las Bases de Datos y reduce la complejidad de las bases de datos por la abstracción de eventos de las subestructuras. En este trabajo, se presentan resultados de la evaluación de la capacidad de simplificación de conocimiento de las bases de datos por el algoritmo de SUBDUE, utilizando bases obtenidas en Internet y otras disponibles en nuestro laboratorio.

3 Título: *Análisis de Parámetros que Caracterizan la Dificultad de Predicción en Series de Tiempo.*

Autores: E. Bautista-Thompson y J. Figueroa-Nazuno.

Presentado en: XLV Congreso Nacional de Física, León, Guanajuato, del 28 de Octubre al 1 de Noviembre del 2002.

Resumen publicado en: *Memorias del XLV Congreso Nacional de Física*, N. R. Silva González (Editor), Sociedad Mexicana de Física, 2002.

Resumen: Se presentan los resultados del análisis de la relación entre los parámetros que caracterizan a las series de tiempo y la dificultad de diferentes técnicas para la predicción o modelado de las mismas. Se identifican parámetros que en conjunto están correlacionados con la dificultad de las técnicas para predecir o modelar a las series. El análisis hace uso de técnicas de inteligencia artificial (Machine Learning) para la identificación de dichas correlaciones.

4 Título: *Matriz de Conocimiento sobre la Complejidad de Predicción en Series de Tiempo*.

Autores: E. Bautista-Thompson y J. Figueroa-Nazuno.

Presentado en: VII Congreso Iberoamericano en Reconocimiento de Patrones, México, D. F. del 19 al 22 de Noviembre del 2002.

Artículo publicado en: *Reconocimiento de Patrones Avances y Perspectivas*, J. L. Díaz de León Santiago y C. Yáñez Márquez (Editores), Centro de Investigación en Computación IPN, 2002.

Resumen: Se presentan los resultados de la evaluación de la capacidad de diferentes técnicas para la predicción y modelado de series de tiempo. Utilizando mapas autoorganizados (SOM), se identificaron relaciones entre la capacidad de predicción de las técnicas y la complejidad de las series de tiempo. Se construyeron clases de complejidad de las series y clases de capacidad de predicción de las técnicas.

5 Título: *Análisis de Patrones aplicado a Problemas de Teoría de Números: El Problema de Collatz*.

Autores: E. Bautista-Thompson, A. Villanueva-Becerril y J. Figueroa-Nazuno.

Presentado en: VII Congreso Iberoamericano en Reconocimiento de Patrones, México, D. F. del 19 al 22 de Noviembre del 2002.

Artículo publicado en: *Reconocimiento de Patrones Avances y Perspectivas*, J. L. Díaz de León Santiago y C. Yáñez Márquez (Editores), Centro de Investigación en Computación IPN, 2002.

Resumen: El problema de teoría de números planteado por L. Collatz en 1937 genera series numéricas (Series de Hailstone) que presentan comportamiento complejo difícil de caracterizar en forma analítica. Se presentan los resultados de aplicar técnicas para el Análisis de Sistemas Dinámicos No Lineales, a series obtenidas a partir de números enteros aleatorios entre  $1 \times 10^{50}$  y  $1 \times 10^{150}$ , con el objetivo de identificar patrones característicos en dichas series.

6 Título: *Evaluación de Red Neuronal Artificial Polinomial (RNAP) para el Modelado de Series de Tiempo.*

Autores: E. Guzmán-Ramírez, E. Bautista-Thompson y J. Figueroa-Nazuno.

Presentado en: Cuarto Congreso Nacional de Computación CORE 2003, México, D. F. del 6 al 7 de Mayo del 2003.

Artículo publicado en: *Avances en la Ciencia de la Computación en México Vol. 2*, A. Gelbukh y M. Hernández Cruz (Editores), Centro de Investigación en Computación IPN, 2003.

Resumen: Se presenta la evaluación de RNAP (Red Neuronal Artificial Polinomial) para el modelado de un conjunto de treinta series de tiempo de diferente origen. En la literatura existen pocos trabajos reportados sobre el desempeño de técnicas de modelado para un conjunto amplio de series de tiempo, así como la comparación de diferentes técnicas de modelado. El objetivo de este trabajo es cuantificar el desempeño de RNAP para series con diferente Índice de Dificultad de Modelado construido a partir de tres propiedades características: Reglas de Producción, Exponente de Lyapunov y Entropía de Shannon. Se comparan los resultados obtenidos con aquellos reportados en la literatura para otras técnicas de modelado de series de tiempo, mostrándose que la capacidad de modelado de los algoritmos híbridos tales como RNAP es comparable con la de las mejores técnicas reportadas.

7 Título: *Caracterización de Series de Tiempo con Escalamiento Multidimensional.*

Autores: V. Rivera-Mancera, E. Bautista-Thompson y J. Figueroa-Nazuno.

Presentado en: Cuarto Congreso Nacional de Computación CORE 2003, México, D. F.

del 6 al 7 de Mayo del 2003.

Artículo publicado en: *Avances en la Ciencia de la Computación en México Vol. 2*, A. Gelbukh y M. Hernández Cruz (Editores), Centro de Investigación en Computación IPN, 2003.

Resumen: Dentro de las técnicas de análisis multivariadas podemos citar el Escalamiento Multidimensional (Multidimensional Scaling MDS). El MDS es una técnica de interdependencia que trata de representar en un espacio geométrico de pocas dimensiones, las proximidades existentes entre un conjunto de objetos en base a métricas de similitud. El problema de identificar series de tiempo con propiedades similares que son indicativas de su dificultad de predicción, motivo en el presente trabajo, el aplicar la técnica MDS para identificar agrupaciones de series con relaciones similares en sus propiedades. Para ello se caracterizan un conjunto de 30 series de tiempo de diferentes orígenes (natural y artificial), a las cuales se les calculó un conjunto de propiedades representativas de origen computacional, topológico, estadístico, espacial, y temporal.

8 Título: *Boosting Technique with Evolutionary Search of Weights for the Improvement of the Output in Time Series Prediction.*

Autores: E. Bautista-Thompson y J. Figueroa-Nazuno.

Presentado en: 23rd International Symposium on Forecasting, Mérida, Yucatán, del 15 al 18 de Junio del 2003.

Resumen publicado en: *Forecasting in Business, Finance and Economics in the Electronic Era: Abstracts Book*, INEGI, 2003.

Resumen: The Prediction of Time Series has limits imposed by the properties of the Series (e. g. degree of chaos) and by the model characteristics of the Prediction Techniques. Each Prediction Technique is able to deal with a specific set of Time Series for which the Prediction Error is small, but for Time Series outside this specific set the Prediction Technique could generate predicted outputs with considerable error. In Time Series Prediction in order to improve the predicted output several methodologies has been developed that combines a group of predictors of the same nature, for example, the simple average of Statistical Models. In a similar way in Machine Learning, Boosting Algorithms (Weighted

Predictor Vote) have been applied to increase the learning capability of Neural Networks for Classification and Regression Tasks. In this work, an evolutionary approach (Genetic Algorithm) has been developed in order to search for the weighted coefficients of a combination of non linear predictors (Neural Network and Non Linear Dynamics Techniques) that minimizes the Prediction Error (RMSE) of predicted outputs. This variant of a boosting algorithm is evaluated for a set of Time Series from different origins (Economy, Physics and Mathematics).

9 Título: *Predicción de Múltiples Puntos de Series de Tiempo Utilizando Support Vector Machines.*

Autores: E. Guzmán-Ramírez, E. Bautista-Thompson y J. Figueroa-Nazuno.

Presentado en: XII Congreso Internacional de Computación CIC 2003, México, D. F. del 13 al 17 de Octubre del 2003.

Artículo Publicado en: *Avances en Ciencias de la Computación Vol. 3*, J. L. Díaz de León Santiago, G. González Santos y J. Figueroa Nazuno (Editores), Centro de Investigación en Computación IPN, 2003.

Resumen: Se presenta la evaluación de la predicción de múltiples puntos de series de tiempo, mediante un corrimiento de ventana para Support Vector Machines (SVM) con dos funciones de kernel distintas (lineal y con base radial). Para la evaluación se utilizó un conjunto de treinta series de diferente origen y comportamiento dinámico. Se encuentra que la SVM posee una buena capacidad para ajustarse a las diferentes dinámicas de las series de tiempo y presenta un buen desempeño para la predicción de los primeros puntos de las series utilizando la función de kernel radial, a pesar del proceso de expansión del error de predicción.

10 Título: *Análisis de la Tolerancia al Ruido en el Modelado de Series de Tiempo con la Red Neuronal RNAP.*

Autores: H. Jiménez-Hernández, E. Bautista-Thompson y J. Figueroa-Nazuno.

Presentado en: XLVI Congreso Nacional de Física, Mérida, Yucatán, del 27 al 31 de Octubre del 2003.

Resumen Publicado en: *Memorias del XLVI Congreso Nacional de Física*, M. L. Marquina Fábrega (Editor), Sociedad Mexicana de Física, 2003.

Resumen: Las Series de Tiempo son un conjunto de valores que caracterizan a un fenómeno. La discretización es el resultado de la medición experimental (en el caso de fenómenos naturales) o de la generación numérica (en el caso de un modelo analítico) de los datos que la conforman. Es común que se presente ruido en las mediciones experimentales afectando de forma directa la posibilidad de modelado de las series de tiempo. El ruido que presentan no necesariamente está distribuido de una manera uniforme. Una manera de disminuir el efecto del ruido es utilizar métodos de filtrado, este proceso puede generar pérdida de información relevante sobre la serie de tiempo a costa de disminuir el ruido presente en la misma. En este trabajo se presentan los resultados de la influencia de distintos tipos de ruido y del proceso de filtrado de los mismos en modelos de series de tiempo generadas a partir de redes neuronales con la arquitectura RNAP.

11 Título: *Descomposición Empírica de Modos para el Análisis de Series de Tiempo*.

Autores: H. Jiménez-Hernández, E. Bautista-Thompson y J. Figueroa-Nazuno.

Presentado en: Decimocuarta Reunión de Otoño de Comunicaciones, Computación, Electrónica y Exposición Industrial, IEEE ROC&C 2003, Acapulco, Guerrero, del 26 al 30 de Noviembre del 2003.

Artículo Publicado en: *Memorias de la Decimocuarta Reunión de Otoño de Comunicaciones, Computación, Electrónica y Exposición Industrial, IEEE ROC&C 2003* (formato electrónico), IEEE Sección México, 2003.

Resumen: El análisis de descomposición empírica de modos es una nueva técnica desarrollada por Norden E. Huang et. al., para el estudio de señales no lineales y no estacionarias. El uso del concepto de modos, es la principal diferencia con relación a las técnicas derivadas del análisis de Fourier, las cuales caracterizan de forma global a las señales y además requieren que estas cumplan con ser lineales y estacionarias. La capacidad de la técnica para descomponer señales complejas en un conjunto finito de funciones intrínsecas de modos, permite la identificación de frecuencias instantáneas a diferentes escalas de tiempo dentro de la señal. Posibilitando no solo una caracterización global,

sino también local de las señales. El conjunto de funciones intrínsecas sirve además para estudiar las estructuras embebidas (o modos de oscilación) dentro de la señal original, las cuales se relacionan con la dinámica del fenómeno que la señal representa. En este trabajo, se presenta la aplicación de esta técnica a diferentes señales (series de tiempo) correspondientes a diversos fenómenos experimentales y numéricos. El objetivo es mostrar la utilidad de la técnica para la descomposición de una señal en sus modos de oscilación. El análisis de este conjunto de modos permite el estudio de las diferentes dinámicas presentes en el fenómeno representado por la señal original.

12 Título: *Predicción de Múltiples Puntos de Series de Tiempo Utilizando Support Vector Machines.*

Autores: E. Bautista-Thompson, E. Guzmán-Ramírez y J. Figueroa-Nazuno.

Artículo Publicado en: *Revista Computación y Sistemas*, enero-marzo 2004, volumen 7, número 3, páginas 148-155.

Resumen: Se presenta la evaluación de la predicción de múltiples puntos de series de tiempo, mediante un corrimiento de ventana para Support Vector Machines (SVM) con dos funciones de kernel distintas (lineal y con base radial). Para la evaluación se utilizó un conjunto de treinta series de diferente origen y comportamiento dinámico. Se encuentra que la SVM posee una buena capacidad para ajustarse a las diferentes dinámicas de las series de tiempo y presenta un buen desempeño para la predicción de los primeros puntos de las series utilizando la función de kernel radial, a pesar del proceso de expansión del error de predicción.

13 Título: *Optimization of Time Series Forecasting by Combination of Models with Evolutionary Heuristic and Error-Correlation Parameters.*

Autores: E. Bautista-Thompson y J. Figueroa-Nazuno.

Presentado en: Encuentro Internacional de Ciencias de la Computación, ENC 2004, Colima, Colima, del 20 al 24 de Septiembre del 2004.

Artículo Publicado en: *Proceedings of the Fifth Mexican International Conference on Computer Science 2004, ENC 2004*, R. Baeza-Yates, J. L. Marroquin y E. Chávez (Edi-

tores), IEEE Computer Society Press, 2004.

Resumen: Two algorithms for the optimization of time series forecasting by combination of models are proposed and evaluated. The first named GABOOST, exploits the heuristic of genetic algorithm in order to search the optimal weights for the mixing of forecasting models. The second named CombFEC, extracts information provided by the forecast errors (RMSE, BE and MAE) of each model to be combined, and the correlation between each model and the forecasted time series, in order to build an Error-Correlation Function (FEC) used to calculate the weights with a SOFTMAX function. The results show that both algorithms are able to improve the forecasting of different time series, reducing the forecast error (RMSE) and increasing the modeling capability expressed by a reduction of the Bias Error (BE).

14 Título: *Classification of Time Series Based on Their Inner Structures.*

Autores: H. Solís-Estrella, E. Bautista-Thompson y J. Figueroa-Nazuno.

Presentado en: XIII Congreso Internacional de Computación CIC 2004, México, D. F. del 13 al 15 de Octubre del 2004.

Artículo Publicado en: *Advances in: Artificial Intelligence, Computing Science and Computer Engineering, Research on Computing Science Vol. 10*, J. Figueroa-Nazuno, A. Gelbukh-Khan, C. Yáñez-Márquez y O. Camacho-Nieto (Editores), Centro de Investigación en Computación IPN, 2004.

Resumen: There are techniques such as the Singular-Spectrum Analysis (SSA) which jointly with Principal Component Analysis (PCA), help us to analyze the underlying structures of a time series and condense them for their study. Using such information with the Multidimensional Scaling (MDS), we can find a representation that allows locating hidden regularities and thus classify the analyzed data. In this paper we present the study of time series of diverse nature using the three abovementioned techniques and show how they group in a bidimensional plane according to similar patterns within their components disregarding the dynamics of the series.

15 Título: *Identification of Structure-Predictability Relations in Time Series with Pattern Recognition Techniques.*

Autores: E. Bautista-Thompson y J. Figueroa-Nazuno.

Presentado en: XIII Congreso Internacional de Computación CIC 2004, México, D. F. del 13 al 15 de Octubre del 2004.

Artículo Publicado en: *Advances in: Artificial Intelligence, Computing Science and Computer Engineering, Research on Computing Science Vol. 10*, J. Figueroa-Nazuno, A. Gelbukh-Khan, C. Yáñez-Márquez y O. Camacho-Nieto (Editores), Centro de Investigación en Computación IPN, 2004.

Resumen: The predictability deals with the difficulty that can be assigned to a time series in order to be forecasted by a model. In this work, the identification of relations between predictability and time series structure is done by means of two pattern recognition techniques: Multidimensional Scaling and Recurrence Plots. The first technique allows the clustering of the time series by their predictability degree that is associated with a set of time series parameters, the second technique allows the visualization of the spatial and temporal dynamics hidden in the structure of the time series. The results shows that the predictability is related with the structural features of the time series through a set of basic structural patterns, these patterns show different kinds of associations with groups of time series that possess a similar predictability.

16 Título: *Relación entre las Estructuras Subyacentes y el Comportamiento Dinámico de las Series de Tiempo.*

Autores: H. Solís-Estrella, E. Bautista-Thompson y J. Figueroa-Nazuno.

Presentado en: XLVII Congreso Nacional de Física, Hermosillo, Sonora, del 25 al 29 de Octubre del 2004.

Resumen Publicado en: *Memorias del XLVII Congreso Nacional de Física*, M. L. Marquina Fábrega (Editor), Sociedad Mexicana de Física, 2004.

Resumen: Existen técnicas como el análisis espectral singular (SSA) que junto con el análisis de componentes principales (PCA) permiten la extracción y análisis de estructuras independientes de una serie de tiempo; y mediante el uso del escalamiento multidimensional (MDS), que facilita la visualización de relaciones entre pares de objetos se pueden

interpretar las relaciones entre las estructuras subyacentes. En este trabajo, se presentan resultados numéricos utilizando las técnicas anteriores, que muestran cómo series de tiempo de diversos fenómenos, se agrupan en base a la similitud de sus estructuras básicas, observándose una relación que es no lineal entre el agrupamiento y el comportamiento dinámico de las series de tiempo.

Las Figuras 13-1 y 13-2, presentan un cronograma de los trabajos desarrollados a la fecha y la temática de cada uno de ellos.

### **13.1.2 Seminarios por Invitación**

- Título: *Modelado de Oleaje y Computación Social.*

Presentada en: Laboratorio de Computación Paralela y Distribuida CIC-IPN a profesores de la Universidad Nacional de Ingeniería de Nicaragua, México D. F., 15 de Noviembre del 2002.

- Título: *Matriz de Conocimiento para Predicción de Series de Tiempo.*

Presentada en: Centro de Investigaciones en Fijación de Nitrógeno UNAM durante el seminario conjunto entre el CIC-IPN y el CIFN-UNAM. Cuernavaca, Morelos, 11 de Febrero del 2003.

- Título: *El Problema de la Predictibilidad en las Series de Tiempo.*

Presentada en: Centro de Investigación en Computación IPN en el seminario de investigación del ciclo de conferencias: Estudiantes de Doctorado. México D. F., 3 de Octubre del 2003.

## **13.2 Guía sobre las Metodologías Experimentales para la Caracterización de las Series de Tiempo y de las Técnicas de Predicción y Modelado**

En el presente apéndice se presenta una descripción paso a paso de las metodologías usadas para caracterizar las series de tiempo y las técnicas de predicción y modelado. Como un ejemplo se

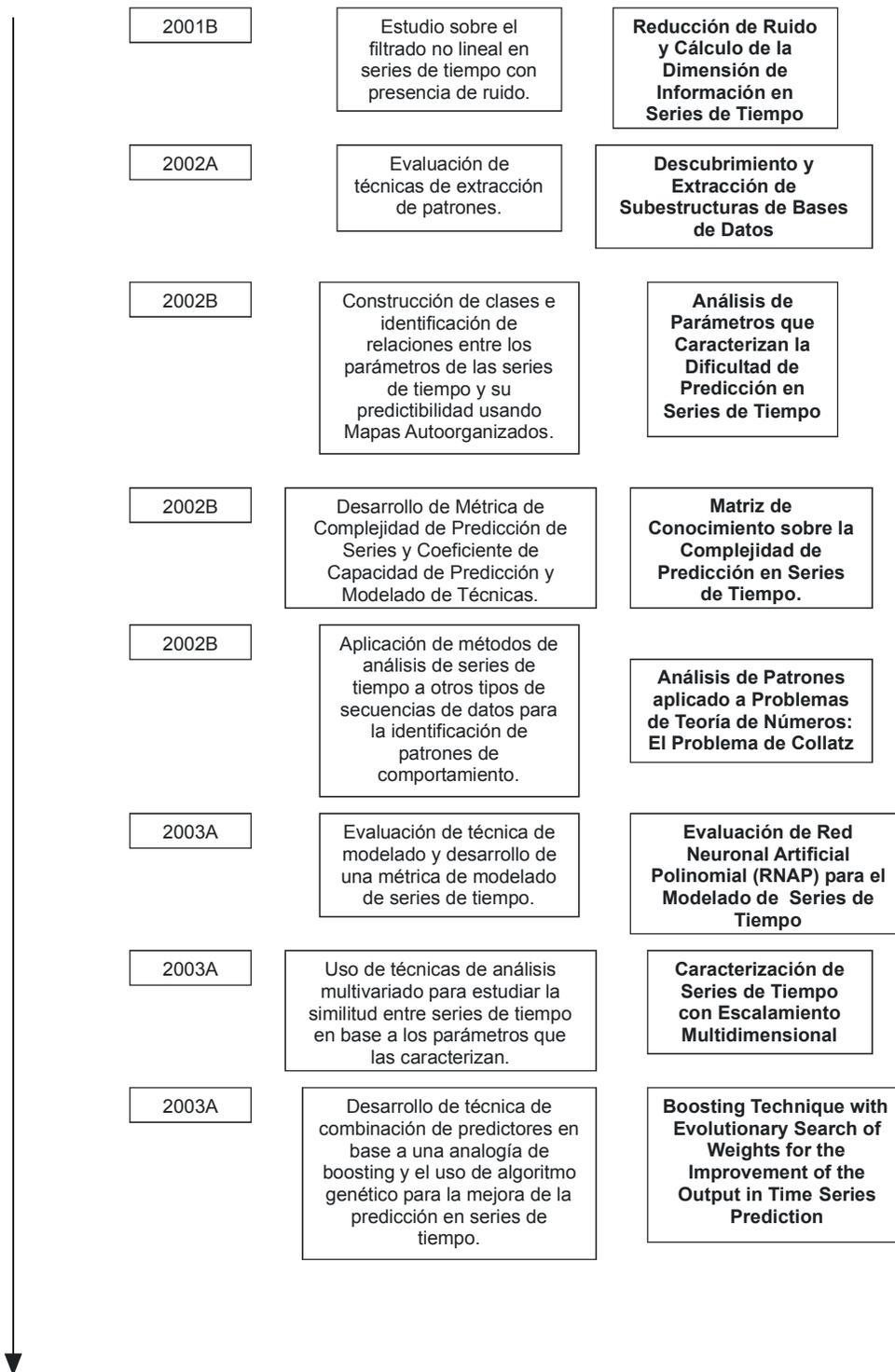


Figura 13-1: Cronograma de la publicación de resultados derivados de la tesis doctoral.

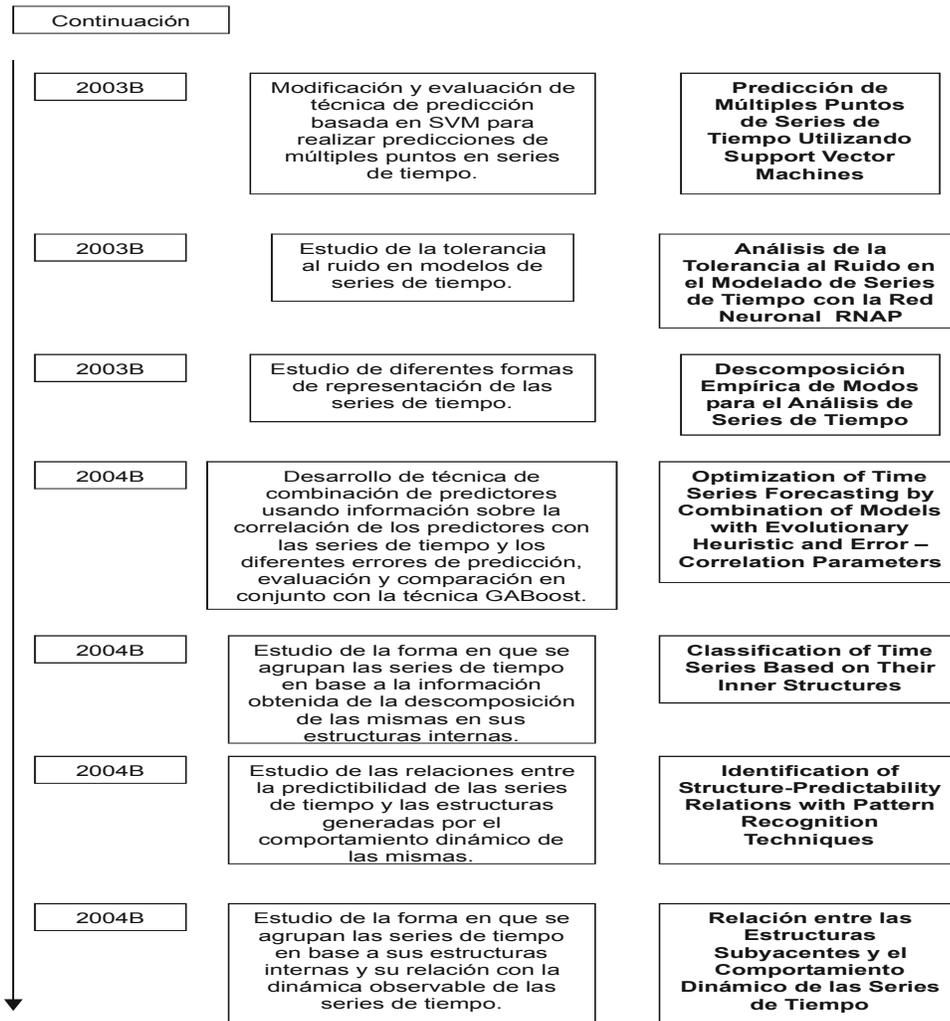


Figura 13-2: Cronograma de la publicación de resultados derivados de la tesis doctoral.

desarrolla la caracterización de la predictibilidad para dos series de tiempo, que no pertenecen al conjunto original de series de tiempo estudiado en la presente tesis, para ilustrar el proceso de agregar nueva información a la base de conocimiento sobre la predictibilidad de series de tiempo.

### 13.2.1 Caracterización de la Predictibilidad de Series de Tiempo

Los pasos para la caracterización de la predictibilidad de las series de tiempo son los siguientes:

- 1 Seleccionar las series de tiempo que se desean estudiar.
- 2 Estandarizar el número de datos que posee cada serie de tiempo por medio de su pre-procesamiento, para obtener un conjunto de series con la misma cantidad de datos.
- 3 Calcular los siguientes parámetros para cada una de las series de tiempo: Exponente de Lyapunov, Dimensión de Correlación, Dimensión de Capacidad, Dimensión Fractal, Dimensión Embebida, Entropía Espacio Temporal, Recurrencia, Determinismo, Entropía de Información de Shannon, Información Mutua Promedio, Complejidad Relativa Lempel-Ziv, Número de Reglas de Producción, Exponente de Hurst y Auto correlación.
- 4 Aplicar la métrica de Coeficiente de Dificultad de Predicción:  $CDP_2$  para calcular la predictibilidad de las series de tiempo bajo estudio, donde la expresión de la métrica como se describió en la sección 9.2 es:

$$CDP_2 = \text{Log}_{10} \left( \frac{EL + DCA + DCO + EET + RP + DE + DF + CRLZ}{EH + ES + DET + REC + IMP + ACO} \right) + 1 \quad (13.1)$$

El valor de la predictibilidad se calcula a partir del inverso de la expresión anterior:

$$\text{Predictibilidad} = \frac{1}{CDP_2} \quad (13.2)$$

- 5 Una vez que se ha calculado la predictibilidad se pueden aplicar diferentes técnicas de análisis multivariado tanto las utilizadas en la presente tesis como otras más para: comparar, discriminar, clasificar, jerarquizar, etc.; los datos obtenidos y realizar un análisis sobre la predictibilidad para el conjunto de series bajo estudio.

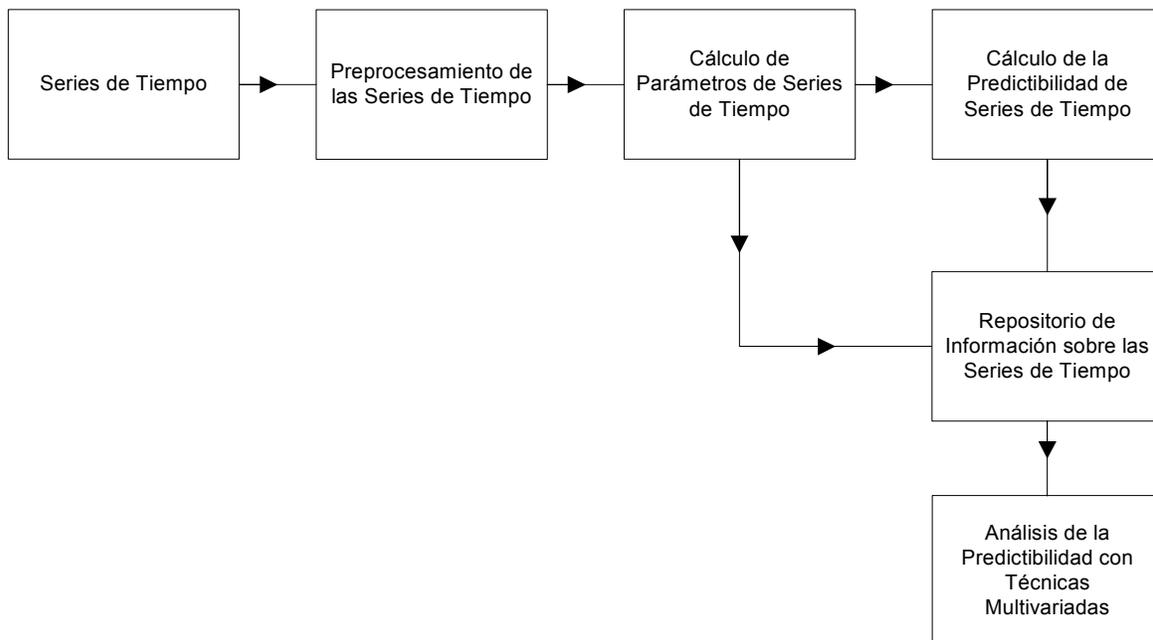


Figura 13-3: Diagrama de la metodología para caracterizar la predictibilidad de series de tiempo

La Figura 13-3, presenta un diagrama de la metodología para la caracterización de la predictibilidad de las series de tiempo.

Como un ejemplo de la aplicación de la metodología, se calculo la predictibilidad de dos series de tiempo adicionales que no pertenecen al conjunto de series usadas en la presente tesis: Doublesine y SPECTA. La serie Doublesine que corresponde a una función de tipo seno modificada presenta un comportamiento dinámico de tipo periódico, mientras que la serie SPECTA que es el resultado de la codificación con números enteros del 1 al 20 de una cadena de aminoácidos (un valor entero por cada aminoácido diferente) de la proteína alpha-spectrin, presenta un comportamiento dinámico de tipo complejo. Las Figuras 13-4 y 13-5 muestran las gráficas de dichas series de tiempo.

El conjunto de parámetros dinámicos correspondientes a las dos series de tiempo se muestra en la tabla de la Figura 13-6, los parámetros calculados se incluyeron en un análisis con escalamiento multidimensional (MDS) para identificar a que agrupamiento corresponden cada una de las dos series en relación con los agrupamientos de series previamente identificados (ver sección 10.3).

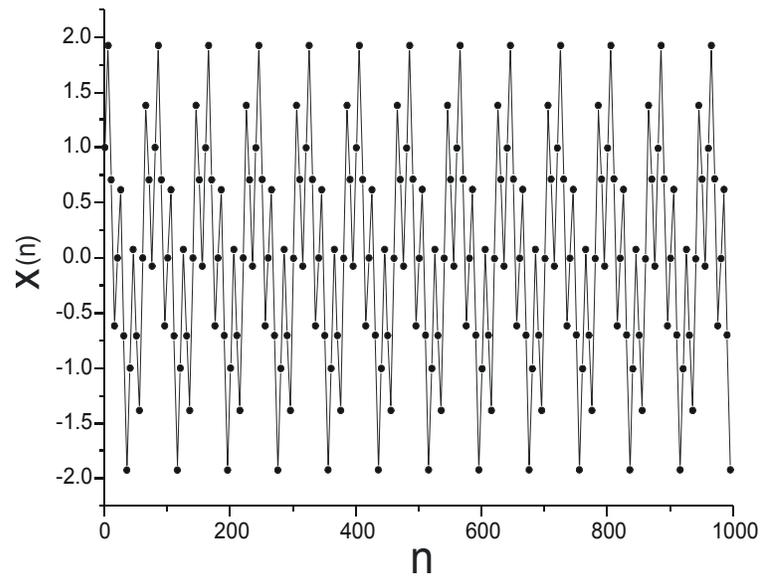


Figura 13-4: Serie de Tiempo DoubleSine

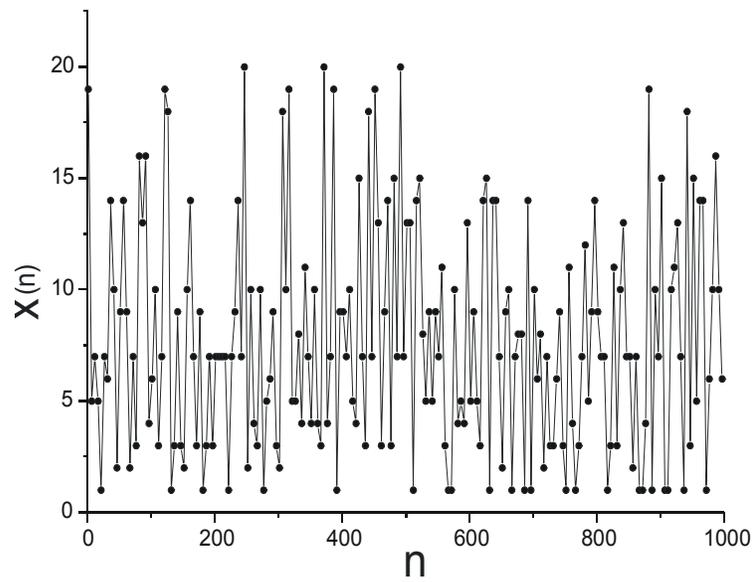


Figura 13-5: Serie de Tiempo SPECTA

Serie de Tiempo	Doublesine	SPECTA
Comportamiento Dinámico	periódica	compleja
Correlación de Pearson (Autocorrelación)	0.9739691	0.04924266
Exponente de Hurst	0.6143159	0.009643
Frecuencia Dominante	0.015625	0.03125
Exponente de Lyapunov	1.027	0.839
Dimensión de Correlación	0.871	3.506
Dimensión de Capacidad	0.604	2.619
Dimension Fractal	0.84	0.8
Dimensión Embebida	2	3
Entropía Espacio Temporal (%)	5	80
Recurrencia (%)	9.911	3.452
Determinismo (%)	47.743	0
Entropía de Información (Shannon)	1.653	0
Información Mutua Promedio	4	2
Complejidad Relativa LZ	0.1295552	1.056373
Número de Reglas de Producción	18	79

Figura 13-6: Tabla de los parámetros dinámicos de las series de tiempo

La Figura 13-7 presenta el mapa de escalamiento multidimensional construido, se observa que la serie Doublesine corresponde al agrupamiento 5 de series de tiempo el valor de su predictibilidad es 1.54 y el patrón básico de mapa de recurrencia asociado con este agrupamiento es el P5, la Figura 13-8 muestra el patrón básico P5 y el mapa de recurrencia correspondiente a la serie Doublesine. Con respecto a la serie SPECTA el valor de su predictibilidad es 0.40, dicha serie corresponde al agrupamiento 1 cuyas series de tiempo poseen los patrones básicos P1a y P1b, la Figura 13-9 muestra los patrones P1a y P1b así como el mapa de recurrencia de la serie SPECTA.

### 13.2.2 Evaluación de la Capacidad de Técnicas para la Predicción o el Modelado

La evaluación de una técnica de predicción o modelado, consta de los siguientes pasos:

- 1 Se selecciona un conjunto de series de tiempo como conjunto de evaluación.
- 2 Se preprocesa a las series de tiempo normalizando los datos de cada una de ellas a la misma escala, por ejemplo que el valor máximo de los datos este en la escala de  $10^1$ .
- 3 Se define el número de datos usados para entrenamiento, por ejemplo los primeros 900 datos de cada serie de tiempo, y el número de datos ha predecir (conjunto de datos de prueba), por ejemplo los siguientes 50 datos para cada serie de tiempo.
- 4 Se define un número máximo de experimentos para intentar la predicción de cada serie por parte de las técnicas evaluadas.
- 5 Se seleccionan los tipos de errores a ser usados en la evaluación del conjunto de técnicas de predicción, en el presente trabajo las medidas de error fueron: RMSE, BE, MAE. Y para el caso de la evaluación global de dichos errores para cada técnica de predicción se usaron los siguientes promedios sobre las medidas anteriores:  $\langle\langle RMSE_j \rangle_k\rangle$ ,  $\langle\langle BE_j \rangle_k\rangle$ ,  $\langle\langle MAE_j \rangle_k\rangle$ .
- 6 El procedimiento para asignar un éxito o un fracaso en la predicción de una serie de tiempo es el siguiente: se grafican los datos del conjunto de prueba de la serie original

Análisis MDS

Modelo de Distancias Euclidianas

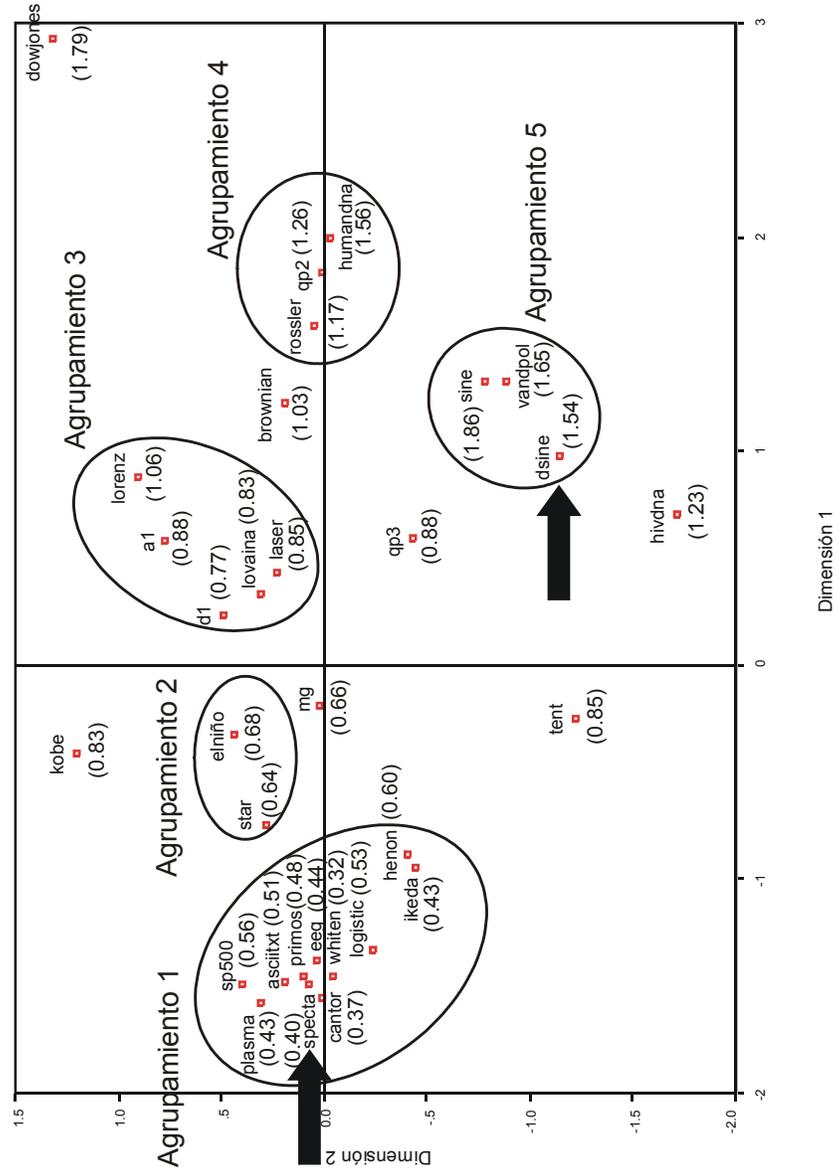


Figura 13-7: Mapa de escalamiento multidimensional de las series de tiempo

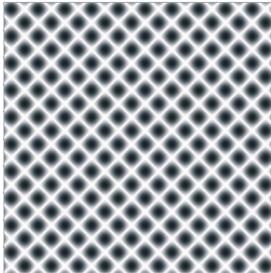
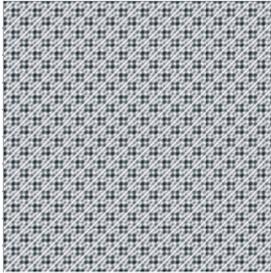
Clase de Patrón de Recurrencia	Mapa de Recurrencia característico	Mapa de Recurrencia de Serie de Tiempo Doublesine
P5		

Figura 13-8: Comparación del mapa de recurrencia de la serie Doublesine con el patrón básico asociado

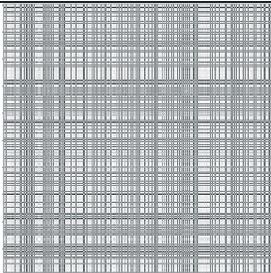
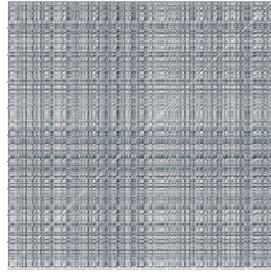
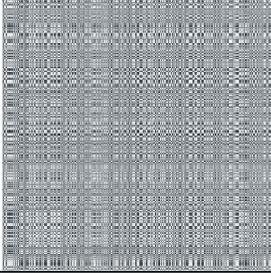
Clase de Patrón de Recurrencia	Mapa de Recurrencia característico	Mapa de Recurrencia de Serie de Tiempo SPECTA
P1a		
P1b		

Figura 13-9: Comparación del mapa de recurrencia de la serie SPECTA con los patrones básicos asociados

y del resultado de la predicción, se comparan ambas gráficas para evaluar si los datos predichos reproducen los originales sujetos a la siguiente regla: si la serie predicha sigue la tendencia (*trend*) de la serie original, así como su amplitud, aún cuando a una escala de mayor resolución no exista una mejor predicción, se considera que la predicción es buena.

7 Ahora se procede a la evaluación de cada técnica de predicción usando el conjunto de series de tiempo.

8 Una vez identificadas las predicciones exitosas para cada técnica, se calculan los errores de la predicción mencionados en el paso 5.

9 Como complemento se calcula la predictibilidad de cada una de las series de tiempo usadas en la evaluación de las técnicas de predicción

10 Para finalmente determinar la capacidad de predicción de cada técnica usando la expresión (ver la sección 9.4 para una descripción más detallada de sus términos):

$$CCAP = \frac{SCDP_2 * EP}{PRMSE + PMAE} \quad (13.3)$$

La Figura 13-10, presenta un diagrama de la metodología para la evaluación de la capacidad de las técnicas para la predicción o el modelado.

### **13.3 Análisis de Gramáticas (Pendiente de la Curva como una Medida de Predictibilidad)**

Otra forma de evaluar la predictibilidad de las series de tiempo es utilizando la pendiente de la curva correspondiente a la evolución de las reglas de producción que expresan en forma simbólica a una serie de tiempo, como *información suplementaria* se presenta en este apéndice resultados del cálculo de esta forma de medir la predictibilidad. Este método es reportado por R. Menchaca-Méndez et. al. [116, 117], la Figura 13-11 muestra ejemplos de las curvas de evolución de las reglas de producción correspondientes a cinco series de tiempo estudiadas en la presente tesis.

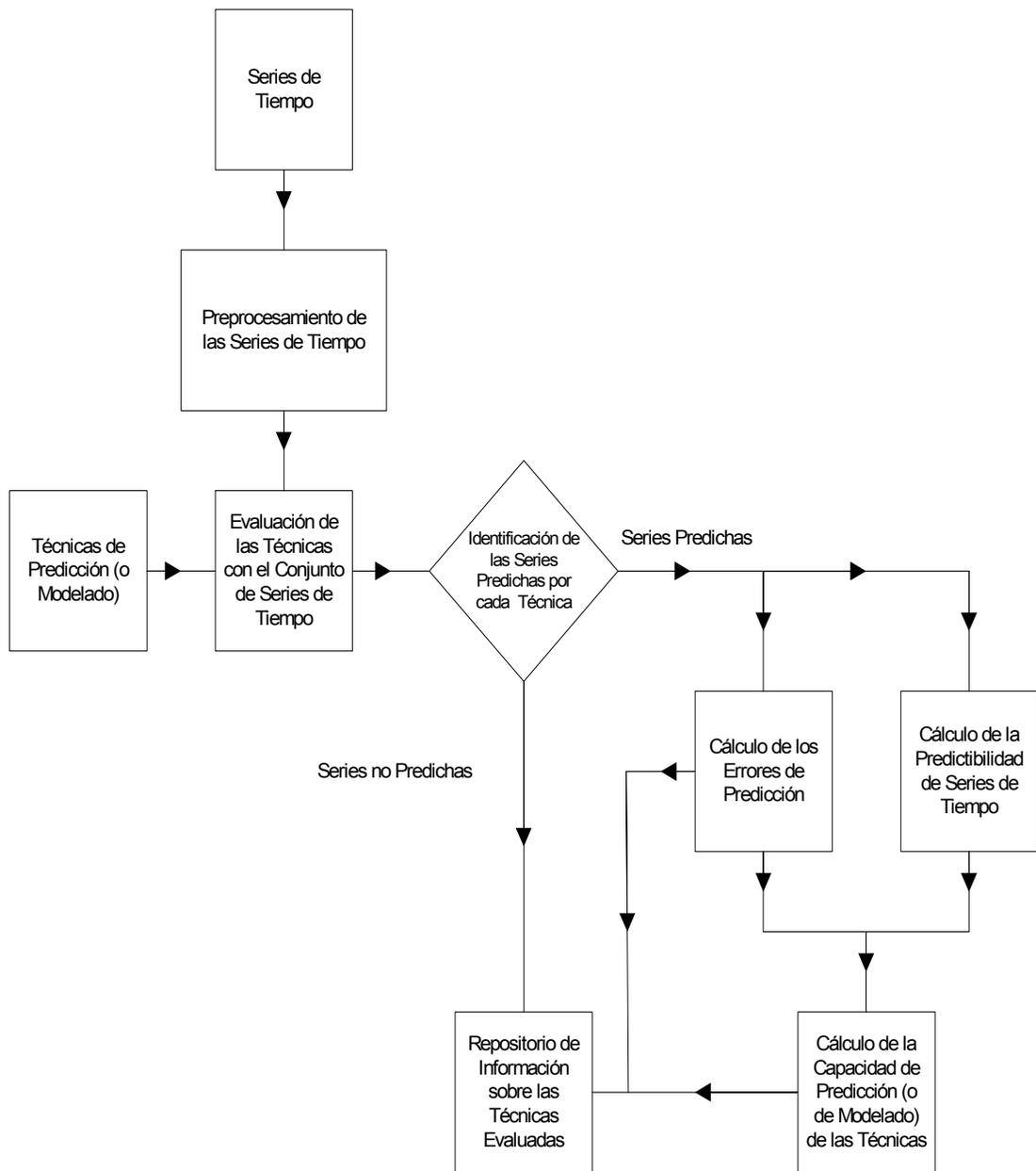


Figura 13-10: Diagrama de la metodología para calcular la capacidad de técnicas para la predicción o el modelado

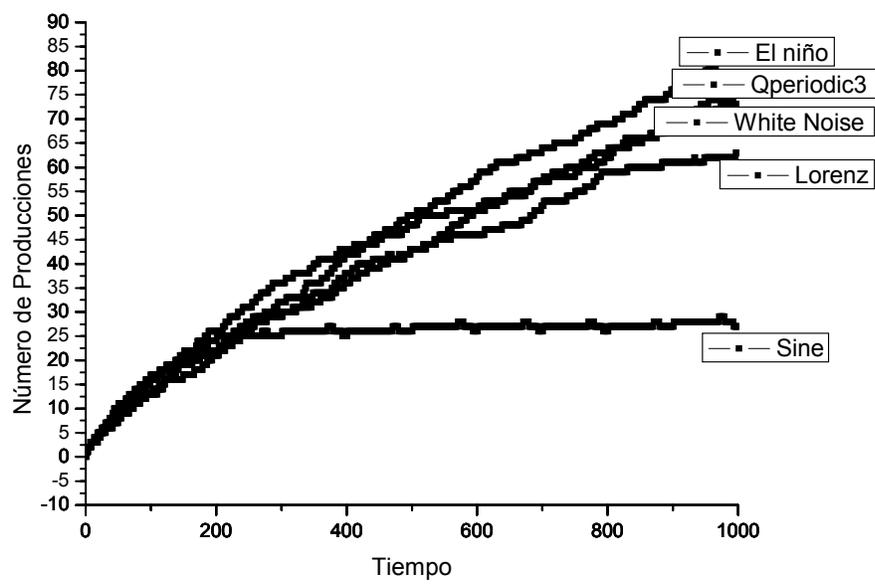


Figura 13-11 Ejemplos de las curvas de evolución de las reglas de producción para las series de tiempo

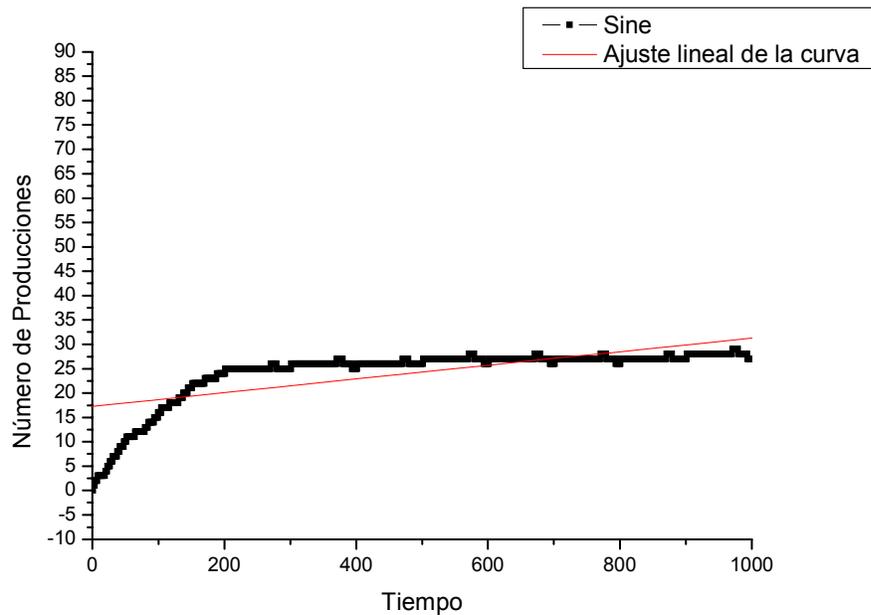


Figura 13-12: Ejemplo del cálculo de la regresión lineal sobre la curva para la serie de tiempo Sine

La Figura 13-12, muestra un ejemplo del cálculo mediante regresión lineal de la pendiente de la curva para la serie de tiempo Sine y la Figura 13-13 muestra el conjunto de parámetros calculados ( $A$  ordenada al origen y  $B$  pendiente de la curva).

En la Figura 13-14, se presenta la tabla con los valores de la pendiente de la curva para cada serie de tiempo usada en la tesis, junto con los datos de las variables de control: número de predicciones para cada serie de tiempo ( $NumPred$ ) y número de modelaciones para cada serie de tiempo ( $NumMod$ ).

La correlación de la pendiente de la curva respecto a las variables de control  $NumPred$  y  $NumMod$  se presenta en la tabla de la Figura 13-15, su correlación con respecto a la predicción de las series de tiempo representada por la variable  $NumPred$  es menor a la correspondiente a la métrica  $CDP_2$  desarrollada en la presente tesis, lo que nuevamente corrobora que *utilizar un solo parámetro para caracterizar la predictibilidad de las series de tiempo*, aunque puede ser útil en caracterizar una faceta de la predictibilidad de las series de tiempo (en este caso de tipo

Regresión lineal de la curva: Y = A + B * X			
Parámetro	Valor	Error	
A	17.25159	0.25331	
B	0.01405	4.38409E-4	
R	SD	N	P
0.71228	4.00211	1000	<0.0001

Figura 13-13: Resultado del cálculo con regresión lineal para la serie de tiempo *Sine*

complejidad computacional mediante el análisis gramáticas), *no proporciona la información suficiente para cuantificarla en forma completa.*

## 13.4 Visualizando los Parámetros Dinámicos de las Series de Tiempo

La importancia de multiplicar la información sobre la dinámica de las series de tiempo, al utilizar un conjunto de parámetros que caracterizan a las series de tiempo desde diferentes puntos de vista, se puede visualizar mediante gráficas de tipo radial, en las cuales cada uno de los parámetros corresponde a un radio de la gráfica. Para que el comportamiento de los parámetros reflejara la misma tendencia: una mayor magnitud del parámetro corresponde a una mayor contribución a la predictibilidad de las series de tiempo, en el caso de los parámetros cuyo comportamiento indica una menor predictibilidad a una mayor magnitud como son: el Exponente de Lyapunov, la Dimensión de Correlación, la Dimensión de Capacidad, la Dimensión Fractal, la Dimensión Embebida, la Entropía Espacio Temporal, la Complejidad Relativa LZ y el Número de Reglas de Producción; se utilizó su valor inverso para construir las gráficas. Además, todos los parámetros fueron normalizados para que su valor máximo sea uno. Cada gráfica que corresponde a una serie de tiempo presenta los catorce parámetros utilizados en el cálculo de la métrica de predictibilidad  $CDP_2$ , los valores de los parámetros están conectados

Serie de Tiempo	Pendiente de la Curva de Reglas de Producción	Número de Predicciones por Serie (NumPred)	Número de Modelaciones por Serie (NumMod)
Sine	0.01405	10	9
Vanderpol	0.01983	8	9
Qperiodic2	0.05915	4	7
Qperiodic3	0.07047	9	6
Mackey-Glass	0.07186	5	7
Logistic	0.05245	4	6
Lorenz	0.05752	6	7
Rossler	0.06375	6	7
Ikeda	0.06334	1	2
Henon	0.06486	5	4
Cantor	0.07048	0	0
Tent	0.00577	8	6
A1	0.07931	8	7
D1	0.0871	0	5
Laser	0.05892	8	7
Dow Jones	0.05899	5	6
Kobe	0.07552	3	4
EEG	0.07241	2	4
ASCIITXT	0.07423	0	0
El niño	0.07525	3	4
HIV DNA	0.00399	0	0
Human DNA	0.03991	6	5
Lovaina	0.07015	3	6
Plasma	0.08136	2	3
Primos	0.06748	0	0
SP500	0.07669	0	0
Star	0.08624	1	6
Brownian Motion	0.07738	0	7
White Noise	0.06791	0	0

Figura 13-14: Tabla de valores de la pendiente de la curva para el conjunto de series de tiempo

Correlación		
	Número de Predicciones por Series (NumPred)	Número de Modelaciones por Serie (NumMod)
Pendiente de la Curva de Reglas de Producción	-0.445	-0.197

Figura 13-15: Tabla de la correlación entre la pendiente de la curva y las variables de control NumPred y NumMod

entre sí por líneas lo cuál forma una figura geométrica característica. La Figura 13-16, muestra la codificación usada en las gráficas radiales para identificar a cada uno de los parámetros, las Figuras 13-17 a 13-21, presentan las gráficas correspondientes al conjunto de series de tiempo.

La gráfica de tipo radial proporciona una visualización de las diferentes facetas de la dinámica de las series de tiempo, y muestra como cada parámetro proporciona diferente información sobre la misma, se observa que el peso de la información proporcionada por cada parámetro es diferente para cada serie de tiempo, y cada serie además posee una figura geométrica característica sobre su dinámica, las series de tiempo cuya predictibilidad es mayor presentan una figura geométrica más abierta, mientras que las series de tiempo con una predictibilidad baja presentan una figura geométrica más cerrada (por ejemplo, EEG, Ikeda, SP500, Cantor, ASCII TXT, Plasma, Primos y White Noise), las gráficas ilustran *la naturaleza multi-dimensional de la predictibilidad de las series de tiempo*, y la riqueza de información que puede ser extraída de una serie de tiempo.

### 13.5 Sobre la No Universalidad de los Modelos de Predicción

La presente tesis corrobora en base a los resultados experimentales encontrados la afirmación sobre la no universalidad de los modelos de predicción, es decir no hay técnica de predicción capaz de predecir todo el continuo de comportamientos dinámicos representados por diferentes series de tiempo. En forma independiente O. Herrera-Alcántara et. al. [192, 193], reportan la

Parámetro	Número asociado
Correlación de Pearson	1
Exponente de Hurst	2
Exponente de Lyapunov	3
Dimensión de Correlación	4
Dimensión de Capacidad	5
Dimension Fractal	6
Dimensión Embebida	7
Entropía Espacio Temporal (%)	8
Recurrencia (%)	9
Determinismo (%)	10
Entropía de Información (Shannon)	11
Información Mutua Promedio	12
Complejidad Relativa LZ	13
Número de Reglas de Producción	14

Figura 13-16: Tabla de simbología de los parámetros en las gráficas radiales

confirmación de esta afirmación por lo que se presenta en este apéndice información sobre dicho estudio, se realizaron una serie de 1045 experimentos en los cuales se investigo en base a un *análisis detallado de los parámetros de cada técnica de predicción*, la habilidad de un conjunto de técnicas (diferentes a las evaluadas en la presente tesis) para predecir al mismo conjunto de series de tiempo estudiado en el presente trabajo. La Figura 13-22, muestra la información sobre los modelos de predicción evaluados en dicho estudio (los modelos están etiquetados para su fácil identificación). Cada letra corresponde a un modelo de predicción combinado con variaciones de funciones y / o distancias, utilizando en el caso de la técnica con modelos de funciones de base radial (Radial Basis), diferentes funciones base (Multicuadráticas, Lineales, Cúbicas, Splines y Gaussianas) combinadas con diferentes medidas de distancia (Euclidiana, Bloque Manhattan, Norma Máxima, etc.), las otras técnicas evaluadas fueron Vecinos Cercanos (Nearest Neighbor), Constante Localmente (Locally Constant) y Lineal Localmente (Locally Linear).

La Figura 13-23, presenta las series de tiempo y los modelos que las predijeron, los modelos

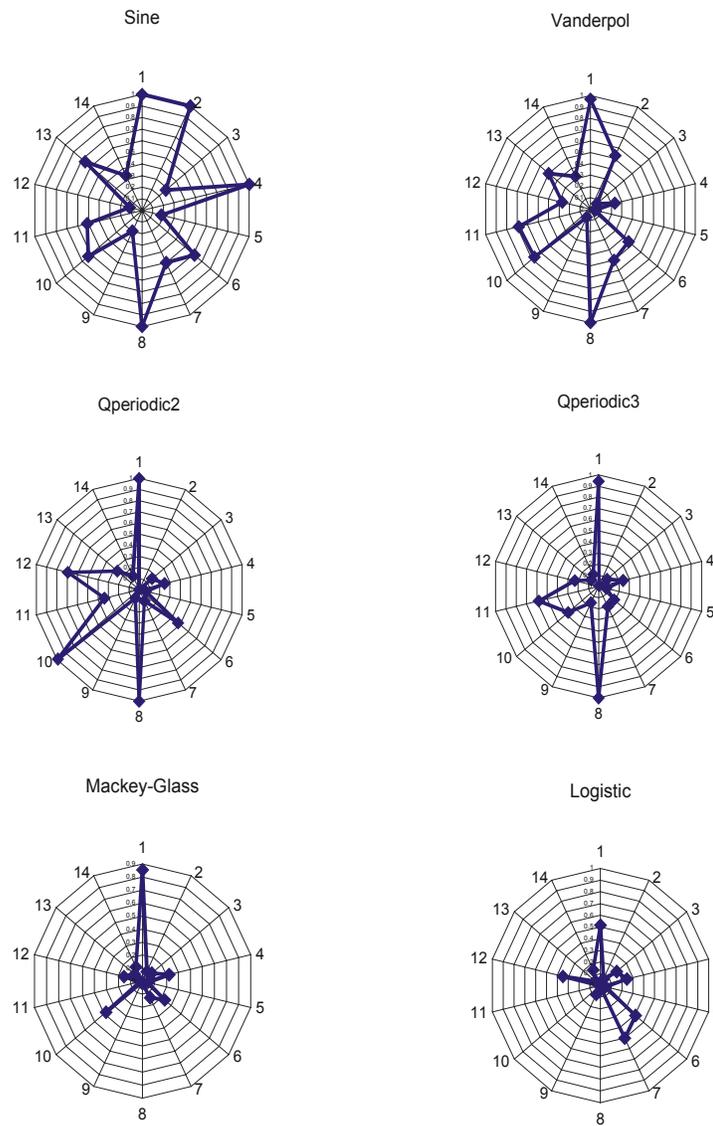


Figura 13-17: Gráficas radiales de los parámetros de las series de tiempo: Sine, Vanderpol, Qperiodic2, Qperiodic3, Mackey-Glass y Logistic

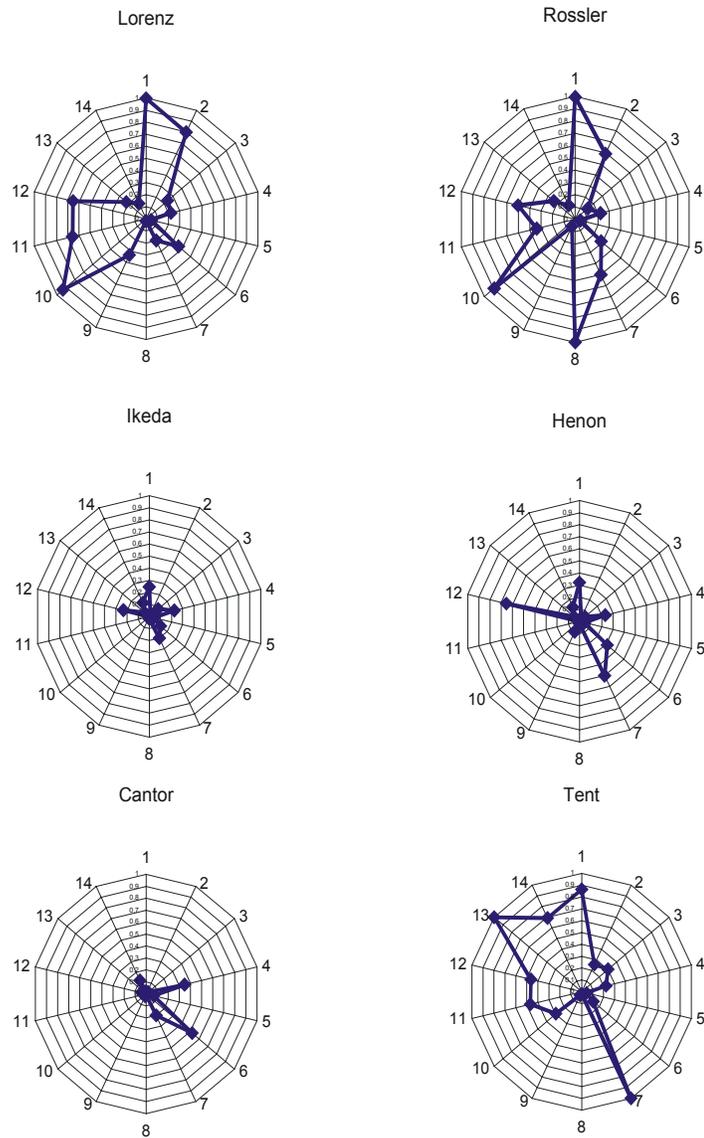


Figura 13-18: Gráficas radiales de los parámetros de las series de tiempo: Lorenz, Rossler, Ikeda, Henon, Cantor y Tent

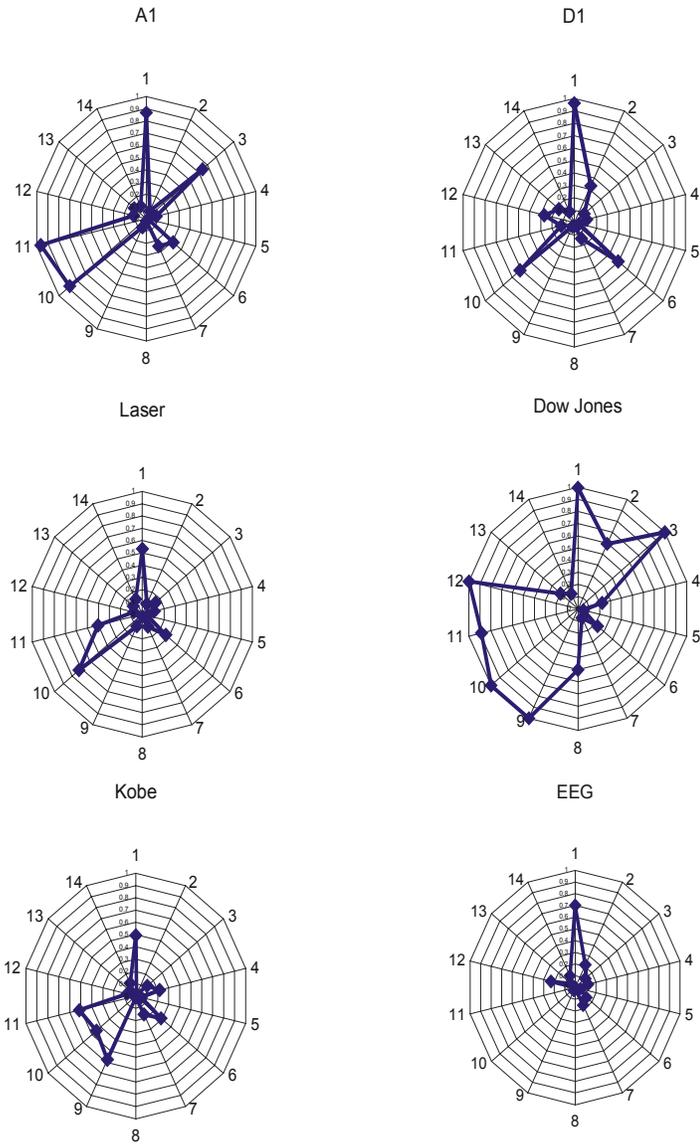


Figura 13-19: Gráficas radiales de los parámetros de las series de tiempo: A1, D1, Laser, Dow Jones, Kobe y EEG

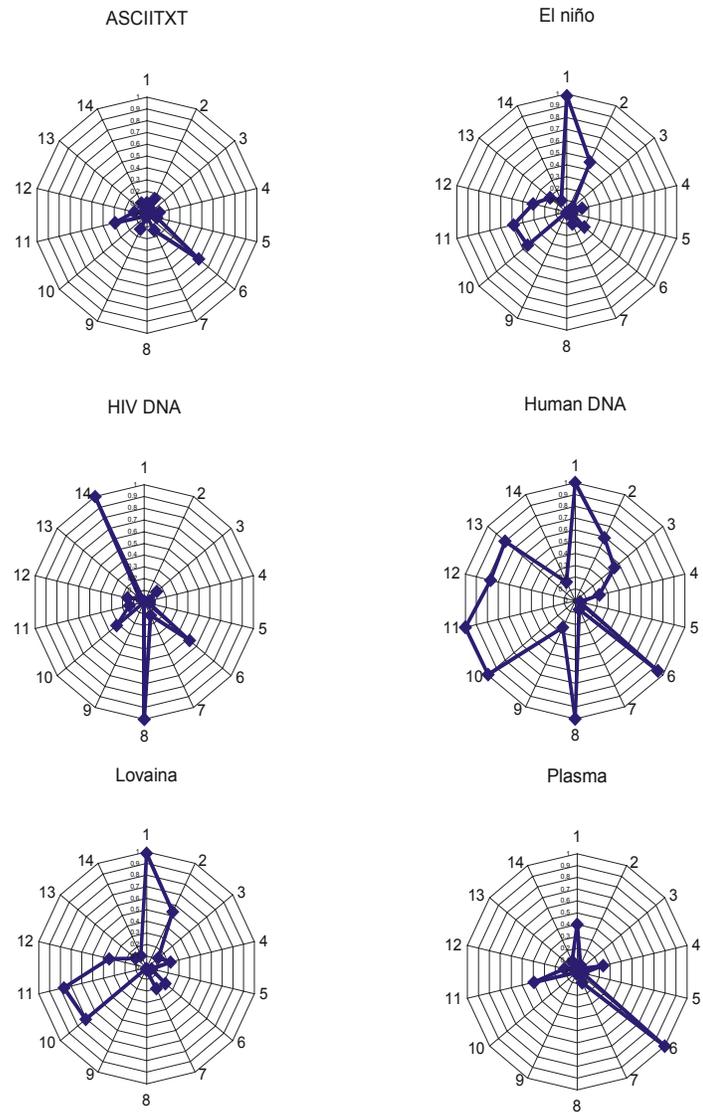


Figura 13-20: Gráficas radiales de los parámetros de las series de tiempo: ASCIITXT, El niño, HIV DNA, Human DNA, Lovaina y Plasma

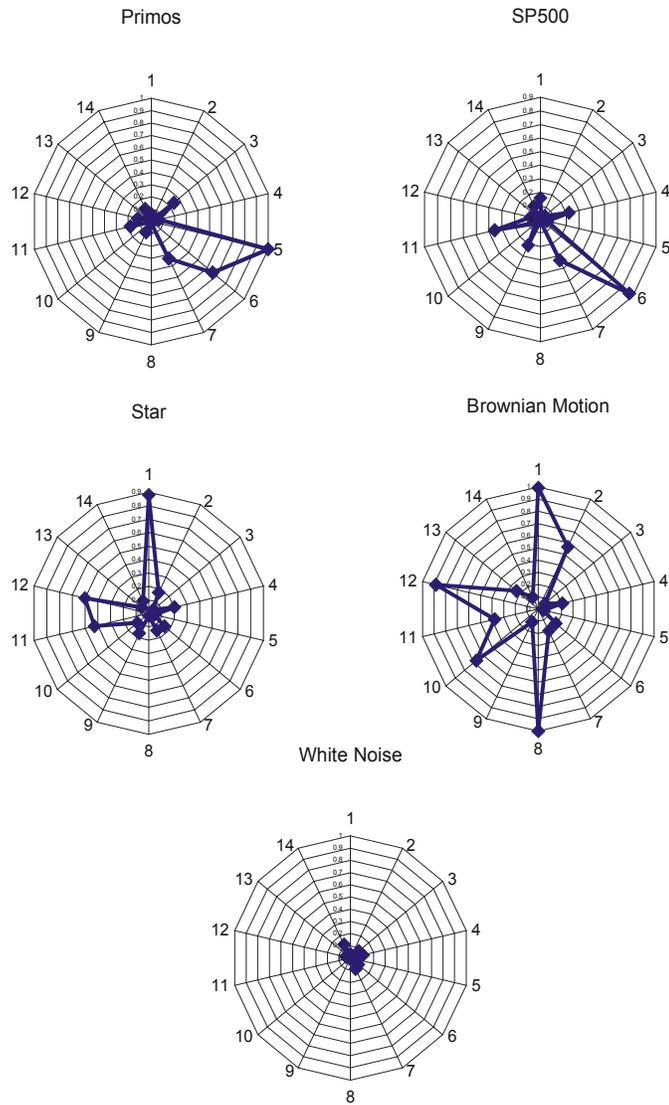


Figura 13-21 Gráficas radiales de los parámetros de las series de tiempo: Primos, SP500, Star, Brownian Motion, White Noise

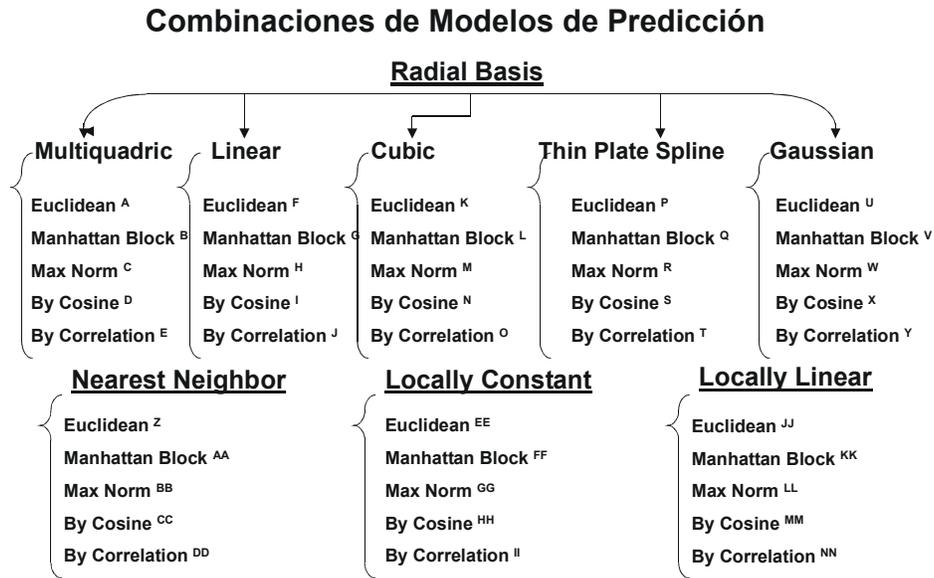


Figura 13-22: Modelos de predicción evaluados en el estudio experimental de O. Herrera-Alcántara et. al.

están codificados por etiquetas (letras mayúsculas) obsérvese que nuevamente se verifica la afirmación de que *no hay modelos capaces de predecir todas las series de tiempo*.

### 13.6 Comentario sobre los Teoremas No Free Lunch

Los teoremas de No Free Lunch (NFL) imponen límites al rendimiento de los algoritmos para búsqueda y optimización. Si un algoritmo es eficiente para una clase particular de problemas entonces su rendimiento para las clases de problemas restantes será pobre. Los teoremas se aplican tanto a algoritmos deterministas como a algoritmos estocásticos y fueron construidos en forma general sin particularizar en algún algoritmo específico, aunque fueron publicados originalmente en revistas del área de computación evolutiva [55, 56], eso no limita su uso a problemáticas similares de otras áreas. En el caso del problema de predicción de series de tiempo, por ejemplo, la búsqueda de los parámetros óptimos para modelar y predecir la dinámica de una serie de tiempo por parte de una red neuronal, cae dentro de la problemática planteada por dichos teoremas. Y en general el problema de predicción de series de tiempo puede verse como un problema de optimización multiobjetivo, en el cuál se buscan los parámetros

### Resultados de la Evaluación de los Modelos de Predicción

min(En)	#Serie-Serie	Tipo	Experimento
• 0.000000	26 - Sine	artificial	H,P,Q,R,Z,AA,BB,CC, EE FF, GG, HH, JJ, KK, LL MM, I, JJ, KK, LL, MM
• 0.000000	29 -Vanderpol	artificial	H,P,Q,R,Z,AA,BB,CC, EE,FF,GG,HH,JJ,KK,LL MM,I,JJ,KK,LL,MM
• 0.000469	17 - Lorenz	artificial	KK
• 0.002099	24 - Rossler	artificial	A
• 0.003361	22 - Qperiodic2	natural	LL
• 0.038376	19 - Mackey-Glass	artificial	U
• 0.166785	28 - Tent	artificial	HH
• 0.233121	23 - Qperiodic3	natural	G
• 0.258252	1 - A1	natural	A
• 0.455643	15 - Laser	natural	V
• 0.617193	18 - Lovaina	artificial	M
• 0.684171	6 - Dow Jones	natural	AA
• 0.833612	5 - D1	artificial	BB
• 0.864338	20 - Plasma	natural	F
• 0.890679	14 - Kobe	natural	II

Figura 13-23: Resultados de la evaluación de los modelos de predicción en el estudio realizado por O. Herrera-Alcántara et. al.

adecuados para modelar y poder predecir la dinámica de una serie de tiempo, por ello los resultados obtenidos en la presente tesis y en trabajos independientes como el comentado en la sección 13.5, al observar en forma experimental que no hay técnicas de predicción universales, manifiestan las limitaciones enunciadas por dichos teoremas. C. Coello-Coello comenta que: *es muy interesante saber que en una problemática diferente a la de algoritmos evolutivos como es la predicción de series de tiempo se presenta una fenomenología similar a la enunciada por los teoremas NFL, y sin duda sería interesante ver como formalizar y encontrar una conexión o una reducción a las condiciones enunciadas por los teoremas NFL, pues sería de gran apoyo para la comunidad que trabaja en predicción de series de tiempo tener teoremas límite de este tipo* [194].

# Bibliografía

- [1] P. E. Rapp, T. I. Schmah y A. I. Mees, "Models of Knowing and the Investigation of Dynamical Systems", *Physica D*, vol. 132, 1999, pág. 133-149.
- [2] C. K. R. T. Jones, "Whiter Applied Nonlinear Dynamics?", *Mathematics Unlimited-2001 and Beyond*, B. Engquist y W. Schmid (Editores), Springer-Verlag, Berlin, 2001, pág. 631-645.
- [3] H. Kantz y T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge, 2000.
- [4] A. Espinosa-Romero, E. Vargas-Medina y J. Figueroa-Nazuno, "Análisis de Señales Aperiódicas por Métodos Gramaticales y Caóticos ", *Memorias del Congreso Internacional de Ingeniería Eléctrica, Electrónica y Computación MEXICON-94*, UDLA, Puebla, 1994.
- [5] J. Figueroa Nazuno, E. Vargas Medina, W. W. Mayol Cuevas y A. Espinosa Romero, "The Analysis and Representation of Complex Aperiodic and Chaotic Signals with Ultrametricity Techniques", *Proceedings of Third SIAM Conference on Applications of Dynamical Systems*, SIAM Society for Industrial and Applied Mathematics, Philadelphia, 1995.
- [6] J. Kurths, U. Schwarz, A. Witt, R. Th. Krampe y M. Abel, "Measures of Complexity in Signal Analysis" *Proceedings of the Third Technical Conference on Nonlinear Dynamics (Chaos) and Full Spectrum Processing*, Naval Undersea Warfare Center, Mystic, 1995.
- [7] M. Palus, "Coarse-Grained Entropy Rates for Characterization of Complex Time Series", *Physica D*, vol. 93, 1996, pág. 64-77.

- [8] A. Espinosa-Contreras y J. Figueroa-Nazuno, "Análisis del Comportamiento de la Pérdida de Paquetes en la Red Internet con Técnicas de la Dinámica No-Lineal", *Memorias del Congreso Internacional de Computación CIC 2000*, Centro de Investigación en Computación IPN, México D. F., 2000, pág. 529-535.
- [9] E. Bautista-Thompson y J. Figueroa-Nazuno, "Análisis de Parámetros que Caracterizan la Dificultad de Predicción en Series de Tiempo", *Memorias del XLV Congreso Nacional de Física*, Sociedad Mexicana de Física, León, 2002, pág. 59.
- [10] E. Bautista-Thompson y J. Figueroa-Nazuno, "Matriz de Conocimiento sobre la Complejidad de Predicción en Series de Tiempo", *Reconocimiento de Patrones Avances y Perspectivas, Research on Computing Science Vol. 1*, J. L. Díaz de León-Santiago y C. Yáñez-Márquez (Editores), VII Congreso Iberoamericano en Reconocimiento de Patrones, Centro de Investigación en Computación IPN, México D. F., 2002, pág. 337-348.
- [11] F. X. Dielbold, *Elements of Forecasting*, South-Western Thomson Learning, Cincinnati, 2001.
- [12] S. A. DeLurgio, *Forecasting Principles and Applications*, McGraw-Hill, Boston, 1998.
- [13] M. Casdagli y S. Eubank (Editores), *Nonlinear Modeling and Forecasting*, SFI Studies in the Sciences of Complexity, Addison-Wesley Longman, Reading, 1992.
- [14] A. Weigend y N. Gershenfeld (Editores), *Time Series Prediction. Forecasting the Future and Understanding the Past*, SFI Studies in the Sciences of Complexity, Proc. Vol. XVII, Addison-Wesley, Reading, 1994.
- [15] D. P. Mandic y J. A. Chambers, *Recurrent Neural Networks for Prediction Learning Algorithms, Architecture and Stability*, John Wiley & Sons, New York, 2001.
- [16] T. Masters, *Neural Novel & Hybrid Algorithms for Time Series Prediction*, John Wiley & Sons, New York, 1995.
- [17] E. A. Wan, *Finite Impulse Response Neural Networks with Applications in Time Series Prediction*, Ph. D. Thesis, Department of Electrical Engineering, Stanford University, 1993.

- [18] C. Sánchez-Rodríguez, *FIRNet con Adaptación de Ventana Temporal para la Predicción de Series de Tiempo*, Tesis de Maestría en Ciencias de la Computación (en proceso), Centro de Investigación en Computación, Instituto Politécnico Nacional, México D. F., 2004.
- [19] T. K. Müller, A. Samola, G. Rätsch, B. Schölkopf, J. Kohlmorgen y V. Vapnik, "Predicting Time Series with Support Vector Machines", *Lecture Notes in Computer Science 1327 Proceedings of ICANN'97*, Springer-Verlag, Berlín, 1997, pág. 999-1004.
- [20] E. Guzmán-Ramírez, E. Bautista-Thompson y J. Figueroa-Nazuno, "Predicción de Múltiples Puntos de Series de Tiempo Utilizando Support Vector Machines", *Memorias del Congreso Internacional de Computación CIC 2003*, J. L. Díaz de León-Santiago, G. González-Santos y J. Figueroa-Nazuno (Editores), Centro de Investigación en Computación IPN, México D. F., 2003, pág. 136-140.
- [21] A. N. Edmonds, *Time Series Prediction using Supervised Learning and Tools from Chaos Theory*, Ph. D. Thesis, School of Science and Computing, University of Luton, 1996.
- [22] D. Kustrin, *Forecasting Financial Time Series with Correlation Memories for Tactical Asset Allocation*, Ph. D. Thesis, Department of Computer Science, University of York, 1998.
- [23] J. López Herrera, *Time Series Prediction Using Inductive Reasoning Techniques*, Ph. D. Thesis, Instituto de Organización y Control de Sistemas Industriales, Universidad Politécnica de Cataluña, 1999.
- [24] J. MacNames, *Innovations in Local Modeling for Time Series Prediction*, Ph. D. Thesis, Department of Electrical Engineering, Stanford University, 1999.
- [25] A. Foka, *Time Series Prediction using Evolving Polynomial Neural Networks*, M. Sc. Thesis, Department of Electrical Engineering and Electronics, University of Manchester Institute of Science and Technology, 1999.

- [26] B. Schoner, *Probabilistic Characterization and Synthesis of Complex Driven Systems*, Ph. D. Thesis, Program in Media, Arts and Sciences, School of Architecture and Planning, MIT, 2000.
- [27] J. A. K. Suykens y J. Vandewalle (Editores), *Nonlinear Modeling Advanced Black-Box Techniques*, Kluwer Academic Press, Amsterdam, 1998.
- [28] S. Makridakis y M. Hibon, "The M3-Competition: Results, Conclusions and Implications", *International Journal of Forecasting*, vol. 16, núm. 4, 2000, pág. 451-476.
- [29] P. Newbold y C. W. J. Granger, "Experience with Forecasting Univariate Time-Series and the Combination of Forecasts (with Discussion)", *J. Roy. Statist. Soc. A*, vol. 137, 1974, pág. 131-165.
- [30] D. J. Reid, "A Review of Short-Term Projection Techniques", *Practical Aspects of Forecasting*, Editor H. A. Gordon, Op. Res. Soc., London, 1975, pág. 8-25.
- [31] S. Makridakis, A. Anderson, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen y R. Winkler, "The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition", *Journal of Forecasting*, vol. 1, 1982, pág. 111-153.
- [32] S. Makridakis, A. Anderson, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen y R. Winkler, *The Forecasting Accuracy of Major Time Series Methods*, John Wiley & Sons, New York, 1984.
- [33] D. H. R. Price y J. A. Sharp, "A Comparison of the Performance of Different Univariate Forecasting Methods in a Model of Capacity Acquisition in UK Electricity Supply", *International Journal of Forecasting*, vol. 2, núm. 3, 1986, pág. 333-348.
- [34] M. Casdagli, "Nonlinear Prediction of Chaotic Time Series", *Physica D*, vol. 35, 1989, pp. 212-233.
- [35] A. García-Ferrer, J. Del Hoyo y A. S. Martín-Arroyo, "On Univariate Forecasting Comparisons: The Case of the Spanish Automobile Industry", *Journal of Forecasting*, vol. 16, núm. 1, 1997, pág. 1-17.

- [36] M. Guo y Y. K. Tseng, "A Comparison Between Linear and Nonlinear Forecasts for Nonlinear AR Models", *Journal of Forecasting*, vol. 16, núm. 7, 1997, pág. 491-508.
- [37] M. Adya y F. Collopy, "How Effective are Neural Networks at Forecasting and Prediction? A Review and Evaluation", *Journal of Forecasting*, vol. 17, núm. 5-6, 1998, pág. 481-495.
- [38] S. Moshiri y N. Cameron, "Neural Networks Versus Econometric Models in Forecasting Inflation", *Journal of Forecasting*, vol. 19, núm. 3, 2000, pág. 201-217.
- [39] S. Gokcan, "Forecasting Volatility of Emerging Stock Markets: Linear Versus Non-Linear GARCH Models", *Journal of Forecasting*, vol. 19, núm. 6, 2000, pág. 499-504.
- [40] G. A. Darbellay y M. Slama, "Forecasting the Short-Term Demand for Electricity, Do Neural Networks Stand a Better Chance?", *International Journal of Forecasting*, vol. 16, núm. 1, 2000, pág. 71-83.
- [41] N. Meade, "A Comparison of the Accuracy of Short Term Foreign Exchange Forecasting Methods", *International Journal of Forecasting*, vol. 18, núm. 1, 2002, pág. 67-83.
- [42] L. R. Weatherford y S. E. Kimes, "A Comparison of Forecasting Methods for Hotel Revenue Management", *International Journal of Forecasting*, vol. 19, núm. 3, 2003, pág. 401-415.
- [43] P. Newbold, J. K. Zumwalt y S. Kannan, "Combining Forecasts to Improve Earnings Per Share Prediction. An Examination of Electric Utilities", *International Journal of Forecasting*, vol. 3, núm. 2, 1987, pág. 229-238.
- [44] D. K. Fan, K-N. Lau y P-L. Leung, "Combining Ordinal Forecasts with an Application in a Financial Market", *Journal of Forecasting*, vol. 15, núm. 1, 1996, pág. 37-48.
- [45] R. G. Donaldson y M. Kamstra, "Forecast Combining with Neural Networks", *Journal of Forecasting*, vol. 15, núm. 1, 1996, pág. 49-61.
- [46] C. T. West, "System-Based Weights Versus Series-Specific Weights in the Combination of Forecasts", *Journal of Forecasting*, vol. 15, núm. 5, 1996, pág. 369-383.

- [47] S. I. Gunter y C. Aksu, "The Usefulness of Heuristic N(E)RLS Algorithms for Combining Forecasts", *Journal of Forecasting*, vol. 16, núm. 6, 1997, pág. 439-462.
- [48] A. Fiordaliso, "A Nonlinear Forecasts Combination Method Based on Takagi-Sugeno Fuzzy Systems", *International Journal of Forecasting*, vol. 14, núm. 3, 1998, pág. 367-379.
- [49] L. M. de Menezes y D. W. Bunn, "The Persistence of Specification Problems in the Distribution of Combined Forecasts Errors", *International Journal of Forecasting*, vol. 14, núm. 3, 1998, pág. 415-426.
- [50] I. Fischer y N. Harvey, "Combining Forecasts: What Information Do Judges Need to Outperform the Simple Average?", *International Journal of Forecasting*, vol. 15, núm. 3, 1999, pág. 227-246.
- [51] J. V. Hansen, *Combining Predictors Meta Machine Learning Methods and Bias-Variance & Ambiguity Decompositions*, Ph. D. Thesis, Department of Computer Science, University of Aarhus, 2000.
- [52] N. R. Swanson y T. Zeng, "Choosing Among Competing Econometric Forecasts: Regression-Based Forecast Combination Using Model Selection", *Journal of Forecasting*, vol. 20, núm. 6, 2001, pág. 425-440.
- [53] J. Shadbolt y J. G. Taylor (Editores), *Neural Networks and the Financial Markets. Predicting, Combining and Portfolio Optimisation*, Springer-Verlag, London, 2002.
- [54] N. Terui y H. K. van Dijk, "Combined Forecasts from Linear and Nonlinear Time Series Models", *International Journal of Forecasting*, vol. 18, núm. 3, 2002, pág. 421-438.
- [55] D. H. Wolpert y W. G. Macready, "No Free Lunch Theorems for Search", *Santa Fe Technical Report Number SFI-TR-95-02-010*, Santa Fe Institute, Santa Fe, 1995.
- [56] D. H. Wolpert y W. G. Macready, "No Free Lunch Theorems for Optimization", *IEEE Transactions on Evolutionary Computation*, vol. 1, núm. 1, 1997, pág. 67-82.
- [57] Y. Bar-Yam, *Dynamics of Complex Systems*, Perseus Books, Reading, 1997.
- [58] S. Wolfram, *Cellular Automata and Complexity*, Westview Press, Reading, 1994.

- [59] Edward Ott, *Chaos in Dynamical Systems*, Cambridge University Press, Cambridge, 2000.
- [60] C. A. Pickover, *Computers, Pattern, Chaos, and Beauty*, Dover Publications, Mineola, 2001.
- [61] E. Bautista-Thompson, A. Villanueva-Becerril y J. Figueroa-Nazuno, "Análisis de Patrones Aplicado a Problemas de Teoría de Números: El Problema de Collatz", *Reconocimiento de Patrones Avances y Perspectivas, Research on Computing Science Vol. 1*, J. L. Díaz de León-Santiago y C. Yáñez-Márquez (Editores), VII Congreso Iberoamericano en Reconocimiento de Patrones, Centro de Investigación en Computación IPN, México D. F., 2002, pág. 349-358.
- [62] H. Whitney, "Differentiable Manifolds", *Annals of Mathematics*, vol. 37, núm. 03, 1934, pág. 645-680.
- [63] F. Takens, "Detecting Strange Attractors in Turbulence", *Lecture Notes in Mathematics*, vol. 898, 1981, pág. 366-381.
- [64] J. M. Medina-Apodaca, *Herramienta para la Extracción de Redes Bayesianas Predictivas a partir de Bases de Datos Temporales*, Tesis de Maestría en Ciencias de la Computación, Centro de Investigación en Computación, Instituto Politécnico Nacional, México D. F., 2005.
- [65] D. C. Baird, *Experimentación: Una Introducción a la Teoría de Mediciones y al Diseño de Experimentos*, Prentice-Hall Hispanoamericana, México, 1991.
- [66] E. Bautista-Thompson, A. Sánchez-Fernández y J. Figueroa-Nazuno, "Reducción de Ruido y Cálculo de la Dimensión de Información en Series de Tiempo", *Memorias del XLIV Congreso Nacional de Física*, Sociedad Mexicana de Física, Morelia, 2001, pág. 99.
- [67] H. Jiménez-Hernández, E. Bautista-Thompson y J. Figueroa-Nazuno, "Análisis de la Tolerancia al Ruido en el Modelado de Series de Tiempo con la Red Neuronal RNAP", *Memorias del XLVI Congreso Nacional de Física*, Sociedad Mexicana de Física, Mérida, 2003, pág. 173.

- [68] R. Rosen, *Fundamentals of Measurement and Representation of Natural Systems*, Elsevier North-Holland, New York, 1978.
- [69] D. H. Krantz, R. D. Luce, P. Suppes y A. Tversky, *Foundations of Measurement, Vol. I*, Academic Press, New York, 1971.
- [70] P. Suppes, D. H. Krantz, R. D. Luce y A. Tversky, *Foundations of Measurement, Vol. II*, Academic Press, San Diego, 1990.
- [71] R. D. Luce, D. H. Krantz, P. Suppes y A. Tversky, *Foundations of Measurement, Vol. III*, Academic Press, San Diego, 1990.
- [72] D. Hand, H. Mannila y P. Smyth, *Principles of Data Mining*, MIT Press, Cambridge, 2001.
- [73] D. S. G. Pollock, *A Handbook of Time Series Analysis, Signal Processing, and Dynamics*, Academic Press, San Diego, 1999.
- [74] G. U. Yule, "On a Method of Investigating Periodicities in Disturbed Series with Special Reference to Wolfer's Sunspot Numbers", *Philos. Trans. Roy. Soc. Series A*, vol. 226, 1927, pág. 267-298.
- [75] E. Slutsky, "The Summation of Random Causes as the Source of Cyclical Processes", *Econometrica*, vol. 5, 1937, pág. 105-146.
- [76] A. N. Kolmogorov, "Interpolation and Extrapolation", *Bulletin de l'academie des sciences de U.S.S.R., Serv. Math*, vol. 5, 1941, pág. 3-14.
- [77] N. Wiener, "Generalised Harmonic Analysis", *Acta Mathematica*, vol. 35, 1930, pág. 117-258.
- [78] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, MIT Press, Cambridge, 1950.
- [79] J. W. Cooley y J. W. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series", *Mathematics of Computation*, vol. 19, 1965, pág. 297-301.

- [80] G. Box y G. Jenkins, *Time-series Analysis, Forecasting and Control*, Prentice Hall, Englewood Cliffs, 1976.
- [81] P. Holmes, "Poincaré, Celestial Mechanics, Dynamical-Systems Theory and Chaos", *Physics Reports*, vol. 193, pág. 137-163, 1990.
- [82] F. X. Diebold y L. Kilian, "Measuring Predictability: Theory and Macroeconomic Applications", *Working Papers Series Number 97-23*, Federal Reserve Bank of Philadelphia, Philadelphia, 1997.
- [83] C. Brooks, "Linear and Non-Linear (Non-)Forecastability of High-Frequency Exchange Rates", *Journal of Forecasting*, vol. 16, núm. 2, 1997, pág. 125-145.
- [84] L. A. Smith, "Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems", *Nonlinear Dynamics and Statistics*, A. Mees (Editor), Birkhäuser, Boston, 2000.
- [85] D. Avramov, "Stock-Return Predictability and Model Uncertainty", *Rodney L. White Center for Financial Research Working Papers Number 12-00*, The Wharton School University of Pennsylvania, Philadelphia, 2000.
- [86] M. Brisson, B. Campbell y J. W. Galbraith, "Forecasting Some Low-Predictability Time Series Using Diffusion Indices", *CIRANO Working Papers Number 2001 s-46*, CIRANO: Centre for Interuniversity Research and Analysis on Organizations, Quebec, 2001.
- [87] U. Feldmann y J. Bhattacharya, "Mixed Predictability as an Asymmetrical Measure of Interdependence in Multivariate Time Series", *6th Experimental Chaos Conference*, M. Spano (Editor), American Institute of Physics, USA, 2002.
- [88] U. Feldmann y J. Bhattacharya, "Predictability Improvement as an Asymmetrical Measure of Interdependence in Bivariate Time Series", *Int. J. Bifurcation and Chaos*, vol. 14, núm. 2, 2004.
- [89] W. Ferson y C. Harvey, "The Risk and Predictability of International Equity Returns", *Review of Financial Studies*, vol. 6, 1993, pág. 527-566.

- [90] W. Ferson y R. Korajczyk, "Do Arbitrage Pricing Models Explain the Predictability of Asset Returns?", *Journal of Business*, vol. 68, 1995, pág. 309-349.
- [91] M. Pesaran y A. Timmermann, "Predictability of Stock Returns: Robustness and Economic Significance", *Journal of Finance*, vol. 50, 1995, pág. 1201-1228.
- [92] S. Kandel y R. Stambaugh, "On the Predictability of Stock Returns: An Asset Allocation Perspective", *Journal of Finance*, vol. 51, 1996, pág. 385-424.
- [93] A. Lo y A. MacKinlay, "Maximizing Predictability in the Stock and Bond Markets", *Macroeconomic Dynamics*, vol. 1, 1997, pág. 102-134.
- [94] M. Cooper, D. H. Downs y G. A. Patterson, "Asymmetric Information and the Predictability of Real State Returns", *The Journal of Real Estate Finance and Economics*, vol. 20, núm. 2, 2000.
- [95] M. Pritamani y V. Singal, "Return Predictability Following Large Price Changes and Informational Releases", *Journal of Banking and Finance*, vol. 25, 2001, pág. 631-656.
- [96] H. Gulen, "Learning About Out-of-Sample Predictability and Its Impact on Real-Time Investments Decisions", *Working Papers Series*, Department of Finance, Pamplin College of Business Virginia Tech, Virginia, 2001.
- [97] A. Ang y G. Bekaert, "Stock Return Predictability: Is it There?", *NBER Working Papers Series Number 8207*, National Bureau of Economic Research, Inc., Cambridge, 2001.
- [98] M. J. Cooper, W. E. Jackson III y G. A. Patterson, "Evidence of Predictability in the Cross-Section of Bank Stock Returns", *Journal of Banking and Finance*, vol. 27, 2003, pág. 817-850.
- [99] V. R. Eleswarapu y M. R. Reinganum, "The Predictability of Aggregate Stock Market Returns: Evidence Based on Glamour Stocks", *The Journal of Business*, vol. 77, núm. 2, 2004.
- [100] N. P. Jewell y P. Bloomfield, "Canonical Correlations of Past and Future for Time Series: Definitions and Theory", *Annals of Statistics*, vol. 11, 1983, pág. 837-847.

- [101] N. P. Jewell, P. Bloomfield y F. C. Bartmann, "Canonical Correlations of Past and Future for Time Series: Bounds and Computation", *Annals of Statistics*, vol. 11, 1983, págs. 848-855.
- [102] E. J. Hannan y D. S. Poskitt, "Unit Canonical Correlations Between Future and Past", *Annals of Statistics*, vol. 16, 1988, págs. 7894-7900.
- [103] C. W. J. Granger y P. Newbold, *Forecasting Economic Time Series*, Academic Press, Orlando, 1986.
- [104] H. Theil, *Applied Economic Forecasting*, North-Holland, Amsterdam, 1966.
- [105] M. A. Kaboudan, "A GP Approach to Distinguish Chaotic from Noisy Signals", *Genetic Programming 1998: Proceedings of the Third Annual Conference*, Morgan Kaufmann, San Francisco, 1998, págs. 187-192.
- [106] M. A. Kaboudan, "A Measure of Time Series Predictability Using Genetic Programming Applied to Stock Returns", *Journal of Forecasting*, vol. 18, núm. 5, 1999, págs. 345-357.
- [107] M. Duan y R. J. Povinelli, "Estimating Time Series Predictability Using Genetic Programming", *Proceedings of Artificial Neural Networks in Engineering*, Artificial Neural Networks in Engineering, 2001, págs. 215-220.
- [108] M. Duan y R. J. Povinelli, "Estimating Stock Price Predictability Using Genetic Programming", *Proceedings of the Genetic and Evolutionary Computation Conference GECCO 2001*, 2001, págs. 174.
- [109] M. Duan, *Time Series Predictability*, M. S. Thesis, Marquette University, Milwaukee, 2002.
- [110] J. W. Galbraith, "Content Horizons for Univariate Time-Series Forecasts", *International Journal of Forecasting*, vol. 19, núm. 1, 2003, págs. 43-55.
- [111] M. J. Cooper y H. Gulen, "Is Time Series Based Predictability Evident in Real-Time?", *Working Papers Series*, Department of Finance, Pamplin College of Business Virginia Tech, Virginia, 2002.

- [112] K. J. M. Cremers, "Stock Return Predictability: A Bayesian Model Selection Perspective", *Review of Financial Studies*, vol. 15, núm. 4, 2002, pág. 1223-1249.
- [113] W. Ebeling, L. Molgedey, J. Kurths y U. Schwarz, "Entropy, Complexity, Predictability and Data Analysis of Time Series and Letter Sequences", *Theory of Disaster*, A. Bundle y H. J. Schellnhuber (Editores), Springer-Verlag, Berlin, 2000.
- [114] L. Molgedey y W. Ebeling, "Intraday Patterns and Local Predictability of High Frequency Financial Time Series", *Physica A: Statistical Mechanics and Its Applications*, vol. 287, núm. 3-4, 2000, pág. 420-428.
- [115] L. Molgedey y W. Ebeling, "Local Order, Entropy and Predictability of Financial Time Series", *Eur. Phys. J. B*, vol. 15, 2000, pág. 733-737.
- [116] R. Menchaca-Méndez, C. Sánchez-Rodríguez., A. Espinosa-Contreras y J. Figueroa-Nazuno, "Modelado y Caracterización de Series de Tiempo por Medio de Análisis Gramatical", *Memorias del XLIII Congreso Nacional de Física*, Sociedad Mexicana de Física, Puebla, 2000.
- [117] R. Menchaca-Méndez, C. Sánchez-Rodríguez y J. Figueroa-Nazuno, "Predicción de Series de Tiempo Mediante Análisis Gramatical", *Memorias del XLIV Congreso Nacional de Física*, Sociedad Mexicana de Física, Morelia, 2001, pág. 134.
- [118] J. C. Sprott, *Strange Attractors: Creating Patterns in Chaos*, Hungry Minds Inc., 1993.
- [119] G. J. Deboeck, *Trading on the Edge. Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets*, John Wiley & Sons, New York, 1994.
- [120] L. W. Salvino, R. Cawley, C. Grebogi y J. Yorke, "Predictability in Time Series", *Physics Letters A*, vol. 209, 1995, pág. 327-332.
- [121] L. A. Smith, "The Maintenance of Uncertainty", *Proc. International School of Physics "Enrico Fermi" Course CXXXIII*, Societa Italiana di Fisica, Bologna, 1997, pág. 177-246.
- [122] P. E. McSharry, *Innovations in Consistent Nonlinear Deterministic Prediction*, Ph. D. Thesis, St. Catherine's College, University of Oxford, 1999.

- [123] L. A. Smith, C. Ziehmann y K. Fraedrich, "Uncertainty Dynamics and Predictability in Chaotic Systems", *Quart. J. Royal Meteorological Soc.*, vol. 125, 1999, pág. 2855-2886.
- [124] C. Ziehmann, L. A. Smith y J. Kurths, "Localized Lyapunov Exponents and the Prediction of Predictability", *Phys. Lett. A*, vol. 271, 2000, pág. 237-251.
- [125] J. J. Zebrowski, W. Poplawska, R. Baranowski y T. Buchner, "Measuring the Complexity of Non-Stationary Time Series Nonlinear Interpretations of Selected Physiological Processes", *Acta Physica Polonica B*, vol. 30, núm. 8, 1999, pág. 2547-2570.
- [126] R. Badii and A. Politi, *Complexity: Hierarchical structures and scaling in physics*, Cambridge University Press, Cambridge, 1999.
- [127] W. H. Zurek (Editor), *Complexity, Entropy and the Physics of Information*, SFI Studies in the Sciences of Complexity, Addison-Wesley Longman, Reading, 1990.
- [128] M. Palus, L. Pecun y D. Pivka, "Estimating Predictability: Redundancy and Surrogate Data Method", *Neural Network World*, vol. 5, núm. 4, 1995, pág. 537-552.
- [129] I. M. Nemenman, *Information Theory and Learning: A physical Approach*, Ph. D. Thesis, Department of Physics, Princeton University, 2000.
- [130] W. Bialek, I. Nemenman y N. Tishby, "Predictability, Complexity and Learning", *Neural Computation*, vol. 13, 2001, pág. 2409-2463.
- [131] E. Acevedo-Mosqueda, C. León-Vega y J. Figueroa-Nazuno, "Medición de Complejidad de Series de Tiempo", *Memorias del XLIV Congreso Nacional de Física*, Sociedad Mexicana de Física, Morelia, 2001, pág. 99.
- [132] V. Rivera-Mancera, E. Bautista-Thompson y J. Figueroa-Nazuno, "Caracterización de Series de Tiempo con Escalamiento Multidimensional", *Avances en la ciencia de la computación en México, Research on Computing Science Vol. 2*, Alexander Gelbukh y Macario Hernández-Cruz (Editores), Cuarto Congreso Nacional de Computación CORE 2003, Centro de Investigación en Computación IPN, México, 2003, pág. 171-180.
- [133] R. Torres León, *Introducción a la Probabilidad y a la Estadística*, Ediciones de la Universidad Autónoma de Yucatán, Mérida, 1985.

- [134] J. Mandel, *The Statistical Analysis of Experimental Data*, Dover Publications Inc., New York, 1964.
- [135] W. Mendenhall, R. L. Scheaffer y D. D. Wackerly, *Estadística Matemática con Aplicaciones*, Grupo Editorial Iberoamérica, México D. F., 1986.
- [136] E. E. Peters, *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics*, John Wiley & Sons, New York, 1994.
- [137] N. Golyandina, V. Nekrutkin y A. Zhigljavsky, *Analysis of Time Series Structure SSA and Related Techniques*, Chapman & Hall / CRC, Boca Ratón, 2001.
- [138] R. M. A. Urbach, *Footprints of Chaos in the Markets*, Financial Times-Prentice Hall, London, 2000.
- [139] G. P. Williams, *Chaos Theory Tamed*, Joseph Henry Press, Washington, 1997.
- [140] R. C. Hilborn, *Chaos and Nonlinear Dynamics*, Oxford University Press, Oxford, 2000.
- [141] J. D. Farmer, E. Ott y J. A. Yorke, "The Dimension of Chaotic Attractors", *Physica D*, vol. 7, 1983, pág. 153-180.
- [142] J. P. Eckmann, S. O. Kamphorst y D. Ruelle, "Recurrence Plots of Dynamical Systems", *Europhys. Lett.*, vol. 4, núm. 9, 1987, pág. 973-977.
- [143] J. P. Zbilut y C. L. Webber, "Embeddings and Delays as Derived from Quantification of Recurrence Plots", *Physics Letters A*, vol. 171, 1992, pág. 199-203.
- [144] J. Gao y H. Cai, "On the Structures and Quantification of Recurrence Plots", *Physics Letters A*, vol. 270, 2000, pág. 75-87.
- [145] J. R. Pierce, *An Introduction to Information Theory: Symbols, Signals and Noise*, Dover Publications, New York, 1980.
- [146] A. Lempel y J. Ziv, "On the Complexity of Finite Sequences", *IEEE Transactions on Information Theory*, vol. 22, núm. 1, 1976.

- [147] W. Stallings, *Data and Computer Communications*, Maxwell Macmillan International, New York, 1991.
- [148] C. G. Nevill-Manning y I. H. Witten, "Identifying Hierarchical Structure in Sequences: A Linear-Time Algorithm", *Journal of Artificial Intelligence Research*, vol. 7, 1997, pág. 67-82.
- [149] E. A. Wan, "Time Series Prediction by Using a Connectionist Network with Internal Delay Lines", *Time Series Prediction. Forecasting the Future and Understanding the Past*, A. Weigend y N. Gershenfeld (Editores), SFI Studies in the Sciences of Complexity, Addison-Wesley, Reading, 1994.
- [150] T. Masters, *Advanced Algorithms for Neural Networks. A C++ Sourcebook*, John Wiley & Sons, New York, 1995.
- [151] B. Martín del Brío y A. Sanz-Molina, *Redes Neuronales y Sistemas Difusos*, Alfaomega Ra-Ma, Madrid, 2002.
- [152] F. M. Ham y I. Kostanic, *Principles of Neurocomputing for Science & Engineering*, McGraw-Hill, Boston, 2001.
- [153] V. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, Chichester, 1998.
- [154] N. Cristianini y J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel Based Methods*, Cambridge University Press, Cambridge, 2000.
- [155] G. Nakhaeizadeh y C. C. Taylor (Editores), *Machine Learning and Statistics The Interface*, John Wiley & Sons, New York, 1997.
- [156] E. Ott, T. Sauer y J. A. Yorke (Editores), *Coping with Chaos. Analysis of Chaotic Data and the Exploitation of Chaotic Systems*, John Wiley & Sons, New York, 1994.
- [157] B. Carnahan y A. E. Smith, *User's Guide to Pittnet Neural Network Educational Software*, Department of Industrial Engineering, University of Pittsburgh, Pittsburgh, 1997.
- [158] J. S. R. Jang, "ANFIS: Adaptive-Network-Based Fuzzy Inference Systems", *IEEE Transactions on Systems, Man and Cybernetics*, vol. 23, núm. 3, 1993, pág. 665-685.

- [159] C. T. Lin y C. S. G. Lee, *Neural Fuzzy Systems: A Neuro-Fuzzy Synergism to Intelligent Systems*, Prentice-Hall, Upper Saddle River, 1996.
- [160] E. Gómez-Rámirez, *Adaptación de la Arquitectura de Redes Neuronales Artificiales y su Aplicación al Control Adaptable*, Tesis de Doctorado, Departamento de Ingeniería Eléctrica, CINVESTAV IPN, México, 1999.
- [161] A. González-Yunes, M. A. Ávila-Alvarez, E. Gómez-Rámirez, O. Mulet, F. Mazzanti y X. Vilasis-Cardona, "Redes Neuronales para Identificación y Predicción de Series de Tiempo", *Rev. Centro Inv.*, vol. 4, núm. 13-14, 2000, pág. 47-65.
- [162] E. Guzmán-Ramírez, E. Bautista-Thompson y J. Figueroa-Nazuno, "Evaluación de Red Neuronal Artificial Polinomial (RNAP) para el Modelado de Series de Tiempo", *Avances en la ciencia de la computación en México, Research on Computing Science Vol. 2*, Alexander Gelbukh y Macario Hernández-Cruz (Editores), Cuarto Congreso Nacional de Computación CORE 2003, Centro de Investigación en Computación IPN, México, 2003, pág. 181-192.
- [163] G. Sugihara y R. M. May, "Non Linear Forecasting as a Way of Distinguishing Chaos from Measurement Error in Time Series", *Nature*, vol. 344, 1990, pág. 734-740.
- [164] E. Gutiérrez-Arias, E. Bautista-Thompson y J. Figueroa-Nazuno, "Descubrimiento y Extracción de Subestructuras de Bases de Datos", *Memorias del Tercer Congreso Estudiantil de Computación CORE 2002*, Centro de Investigación en Computación IPN, México, 2002.
- [165] T. Kohonen, *Self-Organizing Maps*, Springer-Verlag, Berlín, 2001.
- [166] G. Deboeck y T. Kohonen, *Visual Explorations in Finance with Self-Organizing Maps*, Springer-Verlag, Berlín, 1998.
- [167] M. Dittenbach, D. Merkl y A. Rauber, "The Growing Hierarchical Self-Organizing Map", *Proceedings of the International Joint Conference on Neural Networks (IJCNN 2000) Vol. 6*, S. I. Amari, C. L. Giles, M. Gori y V. Piuri (Editores), IEEE Computer Society, Como, 2000, pág. 15-19.

- [168] M. Dittenbach, D. Merkl y A. Rauber, "Using Growing Hierarchical Self-Organizing Maps for Document Classification", *Proceedings of the 8th European Symposium on Artificial Neural Networks (ESANN 2000)*, D-Facto Publications, Bruges, 2000, pág. 7-12.
- [169] I. Borg y P. Groenen, *Modern Multidimensional Scaling*, Springer-Verlag, New York, 1997.
- [170] W. R. Dillon y M. Goldstein, *Multivariate Analysis Methods and Applications*, John Wiley & Sons, New York, 1984.
- [171] S. K. Kachigan, *Multivariate Statistical Analysis A Conceptual Introduction*, Radius Press, New York, 1991.
- [172] A. Espinosa-Contreras y J. Figueroa-Nazuno, "Análisis Caótico y Predictivo de la Contaminación Ambiental en la Ciudad de México", *Memorias del VIII Congreso Internacional de Electrónica, Comunicaciones y Computadoras CONTELECOMP'98*, Puebla, 1998.
- [173] J. C. Sprott, *Chaos and Time-Series Analysis*, Oxford University Press, New York, 2004.
- [174] M. C. Mackey y L. Glass, "Oscillation and Chaos in Physiological Control Systems", *Science*, vol. 197, 1977, pág. 287-289.
- [175] E. N. Lorenz, "Deterministic Non Periodic Flow", *J. Atmos. Sci.*, vol. 20, 1963, pág. 130-141.
- [176] O. E. Rossler, "An Equation for Continuous Chaos", *Physics Letters A*, vol. 57, 1976, pág. 397-398.
- [177] M. Hénon, "A Two-Dimensional Mapping with a Strange Attractor", *Comm. Math. Phys.*, vol. 50, 1976, pág. 69-77.
- [178] J. G. Proakis y D. G. Manolakis, *Tratamiento Digital de Señales: Principios, Algoritmos y Aplicaciones*, Prentice Hall, Madrid, 2000.
- [179] C. Maté y J. Arroyo, "Forecasting with Symbolic Data. Applications to Boxplots", *Proceedings of the 23rd International Symposium on Forecasting ISF 2003*, International Institute of Forecasters, Mérida, 2003, pág. 87.

- [180] A. J. C. Sharkey (Editor), *Combining Artificial Neural Nets*, Springer-Verlag, London, 1999.
- [181] L. G. Valiant, "A Theory of the Learnable", *Communications of the ACM*, vol. 27, 1984, pág. 1134-1142.
- [182] R. E. Schapire, "A Brief Introduction to Boosting", *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence IJCAI 99*, Stockholm, 1999.
- [183] R. E. Schapire, *The Design and Analysis of Efficient Learning Algorithms*, MIT Press, Cambridge, 1992.
- [184] M. Mitchell, *An Introduction to Genetic Algorithms*, MIT Press, Cambridge, 2001.
- [185] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, Berlín, 1999.
- [186] T. Bäck, D. B. Fogel and T. Michalewicz (Editores), *Evolutionary Computation 1: Basic Algorithms and Operators*, Institute of Physics Publishing, Bristol, 2000.
- [187] J. S. Bridle, "Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition", *Neurocomputing: Algorithms, Architectures and Applications*, F.Fogleman-Soulie y J.Herault (Editores), Springer-Verlag, Berlín, 1990, pág. 227-236.
- [188] J. S. Bridle, "Training Stochastic Model Recognition Algorithms as Networks Can Lead to Maximum Mutual Information Estimation of Parameters", *Advances in Neural Information Processing Systems 2*, D.S.Touretzky (Editor), Morgan Kaufmann, San Mateo, 1990, pág. 211-217.
- [189] R. O. Duda, P. E. Hart y D. G. Stork, *Pattern Classification*, John Wiley & Sons, New York, 2001.
- [190] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung y H. H. Liu, "The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis", *Proc. R. Soc. Lond. A*, vol. 454, 1998, pág. 903-995.

- [191] H. Jiménez-Hernández, E. Bautista-Thompson y J. Figueroa-Nazuno, "Descomposición Empírica de Modos para el Análisis de Series de Tiempo", *Memorias de la Decimocuarta Reunión de Otoño de Comunicaciones, Computación, Electrónica y Exposición Industrial IEEE ROC&C 2003*, IEEE Society Sección México, Acapulco, 2003.
- [192] O. Herrera-Alcántara, H. Jiménez-Hernández, P. Escobedo-Gorostizaga y J. Figueroa-Nazuno, "Análisis de la Predictibilidad de Series de Tiempo Usando Mapas Recurrentes", *Memorias del XLVII Congreso Nacional de Física*, Sociedad Mexicana de Física, Hermosillo, 2004, pág. 164.
- [193] O. Herrera-Alcántara, P. Escobedo-Gorostizaga, H. Jiménez-Hernández y J. Figueroa-Nazuno, "Análisis de la Predictibilidad de Series de Tiempo Usando la Matriz de Recurrencia y Espacio de Fase", *Memorias de la Decimoquinta Reunión de Otoño de Comunicaciones, Computación, Electrónica y Exposición Industrial IEEE ROC&C 2004*, IEEE Society Sección México, Acapulco, 2004.
- [194] C. A. Coello-Coello, (*Comunicación Privada*), 2004.