

UNIVERSIDAD NACIONAL DE CÓRDOBA

Facultad de Ciencias Exactas, Físicas y Naturales

Tesis Doctoral



**Modelos no lineales de pronóstico de series temporales  
basados en inteligencia computacional para soporte en la  
toma de decisiones agrícolas**

Autor: Ing. Cristian Rodríguez Rivero

Director: Dr. Julián Antonio Pucheta

Diciembre de 2016

# Modelos no lineales de pronóstico de series temporales basados en inteligencia computacional para soporte en la toma de decisiones agrícolas

por

Ing. Cristian Rodríguez Rivero

**Director**

Dr. Ing. Julián Antonio Pucheta

**Comisión Asesora**

Dr. Ing. Julián Antonio Pucheta

FCEFYN - UNC

Dr. Ing. Víctor Hugo Sauchelli

FCEFYN - UNC

Dr. Oscar Humberto Bustos

FAMAF - UNC

Esta tesis fue enviada a la Facultad de Ciencias Exactas Física y Naturales de la Universidad Nacional de Córdoba, para cumplimentar los requerimientos de obtención del grado académico de Doctor en Ciencias de la Ingeniería.

Córdoba, Argentina  
Diciembre 2016



UNIVERSIDAD NACIONAL DE CORDOBA  
Facultad de Cs. Exactas, Físicas y Naturales

**ACTA DE EXAMENES**

Libro: 00001 Acta: 03756 Hoja 01/01  
LLAMADO: 1 22/12/2016

CATEDRA - MESA:

DI002 TESIS DOCTORADO EN CIENCIAS DE LA INGENIERIA

NUMERO	APELLIDO Y NOMBRE	DOCUMENTO	INGRESO	COND.	NOTA	FIRMA
26480476	RODRIGUEZ RIVERO, Cristian Maxi	DNI: 26480476	2010	T	APROBADO	

CORRAL BRIONES, Graciela - MUT, Vicente - BUSTOS, Oscar - HUEDA, Mario Rafaél - MALDONADO, Ana -

Observaciones:

OSCAR H. BUSTOS

Córdoba, \_\_\_/\_\_\_/\_\_\_.

Certifico que la/s firma/s que ha/n sido puesta/s en la presente Acta pertenece/n a: \_\_\_\_\_

1 0 1 0 1  
Inscriptos Ausentes Examinados Reprobados Aprobados

16/12/2016 12:04:14

Libro/Acta: 0000103756

Hoja: 01/01

A Pauli, a mi bebé Agustín y a mi familia.

A Dios, a Jesús Cristo y a la Virgen María.

# Agradecimientos

Dedicado a todas las personas que hicieron posible esta titulación, mi esposa Paula y a mi bebe Agustín por el aguante todo este tiempo, a mi madre María Margarita Rivero y mis hermanas Gisela, Eliana y Marilúz. A mi director Julián Pucheta por su gran colaboración y soporte, a mi comisión asesora Dr. Víctor Sauchelli (UNC), al Dr. Oscar Bustos (UNC) y al Dr. Héctor Daniel Patiño (INAUT-UNSJ) por su apoyo incondicional a lo largo de la tesis.

A mi compañero Josef Baumgartner, a Sergio Laboret, a Martín Herrera de UNCA por su ayuda en los momentos que hicieron falta. A la Agencia Nacional de Promoción Científica y Tecnológica, a la Secretaria de Ciencia y Tecnología (SECyT) de la UNC y al Dr. Ing. Luis Godoy (UNC) por sus consejos.

Agradezco a las siguientes instituciones que, con su valioso aporte técnico/financiero, posibilitaron el desarrollo de esta Tesis durante estos cinco años de formación en investigación y desarrollo: a la Agencia Nacional de Promoción Científica y Tecnológica, del Ministerio de Ciencia, Tecnológica e Innovación Productiva (MINCyT) de la Nación, a la Secretaria de Ciencia y Tecnología (SECyT) de la Universidad Nacional de Córdoba (UNC), y en especial al Laboratorio de Investigación Matemática aplicada a Control (LIMAC) de la FCEFyN de la UNC.

Agradezco a todas las personas que de una u otra forma apoyaron, aconsejaron, y colaboraron en el desarrollo de esta Tesis.

Finalmente, agradezco a todos los que de alguna u otra forma hicieron posibles la realización de esta Tesis, a Ronald del Águila y a los Ing. Agrónomos Ernesto Carreño, Nicolás Bernáldez Brunt, Cecilia Scalerandi, Mónica Piccardi y Rodrigo Lasa por su contribución con datos históricos de series de lluvia locales y a la CONAE a través del Dr. C. Marcelo Scavuzzo.

## Resumen

### Modelos no lineales de pronóstico de series temporales basados en inteligencia computacional para soporte en la toma de decisiones agrícolas

*Por Cristian Rodríguez Rivero*

Existe una creciente demanda de necesidad de predicciones para variables meteorológicas, principalmente precipitación, por numerosos sectores entre los que se pueden citar el agrícola, ambiental, biológico, hidrológico, sanitario, industrial, energético y transportes. Para encarar esta problemática en el sector agrícola, en la actualidad se usan herramientas tecnológicas integrando información usada en la agricultura de precisión.

Esta tesis se enfoca en modelos predictivos basados en redes neuronales destinados a pronosticar datos históricos de lluvia observados para la toma de decisiones. Estos algoritmos de aprendizaje automático pueden mejorarse en numerosos aspectos y son una herramienta muy promisoría en el ámbito agropecuario. Técnicas avanzadas eficientes como las redes neuronales han sido utilizadas para predecir cualquier tipo de serie temporal real. Los resultados teóricos y empíricos reportados en la literatura, demuestran que estos enfoques producen predicciones más fiables comparadas con los obtenidos a partir de métodos de predicción estadísticos convencionales.

Para saber si esta metodología es válida en el campo del pronóstico de series temporales de lluvia, esta tesis realiza un estudio comparativo con otras técnicas que utilizan modelos lineales y no lineales. Este trabajo supone que la variabilidad de las series de precipitaciones de lluvia puede ser modelada como una realización del Movimiento Browniano fraccionario. El aporte principal es el desarrollo de modelos predictivos causales basados en redes neuronales usando varios métodos. Estos son: métodos de suavización de la serie mediante *Energía asociada a la serie* (EAS) y *Decimación ó Submuestreo* (SUB) y métodos basados en aproximación bayesiana a través del movimiento browniano fraccional, como el *mejorado bayesiano* (BEA) y *mejorado bayesiano modificado* (BEMA). Luego se comparan los resultados obtenidos con otros modelos existentes en la literatura.

Dentro de los aportes se toma como referencia el método de predicción basado en *rugosidad* (NN-Modified) y el *método bayesiano* (BA) desarrollados en esta tesis, adaptando las suposiciones de los nuevos modelos propuestos a las necesidades particulares con un mejor desempeño. Para comparar los resultados de los modelos estudiados se utilizan los índices SMAPE (Error Medio Absoluto Simétrico Porcentual) y RMSE (Raíz del Error Cuadrático Medio), que son medidas de desempeño cuantitativo usados comúnmente para evaluar métodos de pronóstico.

Se observa que, en la gran mayoría de los casos, los algoritmos desarrollados en esta tesis tienen un mejor desempeño respecto de los algoritmos de referencia. Para evaluar los modelos en la predicción de series de lluvia, se tomó como caso de estudio distintos puntos geográficos de la Provincia de Córdoba. Expresado en términos de exactitud y errores en el horizonte del pronóstico se puede constatar que los índices SMAPE y RMSE de los algoritmos BEA-BEMA son más bajos que los algoritmos de referencia.

**Palabras claves:** *series temporales, precipitación de lluvia, algoritmos de aprendizaje automático, modelos predictivos, redes neuronales, algoritmos genéticos, método bayesiano, método de suavización, Método mejorado Bayesiano (BEA), Método mejorado Bayesiano modificado (BEMA), Método Bayesiano (BA), Energía asociada a la serie (EAS), Decimación o Submuestreo (SUB), Red neuronal modificada por rugosidad (NN-Mod.).*



# Abstract

## Nonlinear models of time series forecasting based on computational intelligence to support the agricultural decision-making

by *Cristian Rodríguez Rivero*

There is a growing demand for weather predictions, mainly precipitation, for many sectors among which may be mentioned agricultural, environmental, biological, hydrological, health, industrial, energy and transport. To address this problem in agriculture, in fact, it is currently used technological tools integrating information employed in precision agriculture.

This thesis focuses on predictive models based on neural networks for rainfall forecasting from observed historical data to support the decision-making. These machine-learning algorithms can be improved in many ways and is a very promising tool in the agricultural field. Efficient advanced techniques such as neural networks have been used to predict any real-time series. Theoretical and empirical results reported in the literature, show that these approaches produce more reliable predictions than those obtained from conventional methods.

In order to show that this methodology is valid in the field of rainfall time-series forecasting, this dissertation makes a comparative study with other techniques using linear and nonlinear models. This work assumes that the variability of rainfall series can be modeled as trace path of a fractional Brownian motion. The main contributions are the development of causal predictive models based on neural networks using distinct approaches. These methods are; *Energy associated to series* (EAS) and *decimation or subsampling* (SUB); and methods based on Bayesian approach by means of fractional Brownian motion, *Bayesian Enhanced Approach* (BEA) and *Bayesian Enhanced Modified Approach* (BEMA), used for rough time series prediction, and compared to results obtained by other existing models in the literature.

Within these contributions, reference prediction methods are chosen such as neural networks modified by roughness (NN-Modified) and Bayesian method (BA) developed in this thesis, adapting the assumptions of the new models proposed to the requirements with better performance. To compare the results of the models studied, indexes such as SMAPE (mean

absolute error Symmetrical Percentage) and RMSE (Root Mean Square Error) are utilized by performance measures, which is commonly employed to evaluate forecasting methods in the literature.

It is noted that, in most cases, the algorithms developed in this thesis have better performance against the reference algorithms. To evaluate the models in predicting rainfall time series, studies from different geographical locations of the Province of Cordoba are taken as case studies. Expressed in terms of accuracy and errors in the forecast horizon, it can be seen that the SMAPE and RMSE indexes of BEA-BEMA algorithms are lower than the reference approaches.

**Keywords:** *time series, rainfall, machine learning algorithms, neural networks, predictive models, Bayesian method, heuristic, Bayesian Enhanced approach (BEA), Bayesian Enhanced modified approach (BEMA), Bayesian approach (BA), Energy Associated to Series (EAS), Subsampling (SUB), Neural Network modified by roughness (NN-Mod.).*

## Resumo

### **Modelos não lineares de previsão de séries temporais com base em inteligência computacional para apoiar a tomada de decisões agrícolas**

*Por Cristian Rodríguez Rivero*

Existe uma demanda crescente de necessidades de variáveis meteorológicas, principalmente precipitação, por muitos sectores entre os que podem ser citar o agrícola, ambiental, biológico, hidrológico, sanitário, industrial, energético e transportes. Para encarar esta problemática no setor agrícola, na atualidade usamos ferramentas tecnológicas integrando informação usada na agricultura de precisão.

Esta tesis é enfocada em modelos preditos basados em redes neuronais. Esses algoritmos de aprendizagem automática podem ser melhorados em seu estilo e possuem uma ferramenta muito promissora no âmbito do agropecuário. Técnicas avançadas eficientes como redes neuronais foram utilizadas para predecir qualquer tipo de série temporal real. Os resultados teóricos e empíricos relatados na literatura, demuestran que esses enfoques produzem previsões mais confiáveis comparadas com os obtidos a partir de métodos de previsão estatísticos convencionais.

Para saber se esta metodologia é válida no campo do pronóstico de séries temporais de chuva, esta teses realiza um estudo comparativo com outras técnicas que usam modelos lineares e não lineares. Este trabalho suporta que a variabilidade da série de precipitações de chuva pode ser modelado como uma realização do movimento Browniano fraccionario. O principal é o desenvolvimento de modelos que utilizam diversos métodos. Estes são: métodos de suavização da série através de Energia associada à série (EAS) e Decimação Submuestreo (SUB) e métodos basados na aproximação bayesiana através do movimento browniano fraccional, como o melhor bayesiano (BEA) e melhorado bayesiano modificado BEMA). Então compare os resultados obtidos com outros modelos existentes na literatura.

Dentro dos aportes tome como referência o método da previsão baseado na rugosidade (NN-Modified) e o método bayesiano (BA) desenvolvido em esta teses, adaptando as suposições dos novos modelos projeta as necessidades especiais com um melhor desempenho. Para os resultados

dos modelos estudados utilizam-se os índices SMAPE (Error Médio Absoluto Simétrico Porcentual) e RMSE (Raíz do Erro Cuadrático Médio), que são medidas de desempenho quantitativo usado comumente para avaliar métodos de pronóstico.

Se observa que, na grande maioria dos casos, os algoritmos desenvolvidos em teses têm um melhor desempenho no que diz respeito aos algoritmos de referência. Para avaliar os modelos na previsão da série de chuva, é como um exemplo de estudo geográfico dos pontos geográficos da província de Córdoba. Expresado em termos de exatidão e erros no horizonte do pronombre pode se constatar que os índices SMAPE e RMSE do algoritmo BEA-BEMA filho mais baixos que os algoritmos de referência.

**Palavras-chave:** *séries temporais, precipitação de chuva mensal, algoritmos de aprendizado de máquina, modelagem preditiva, redes neurais, algoritmos genéticos, o método Bayesian, método de alisamento, método aperfeiçoado Bayesiana (BEA), melhorou método modificado Bayesiana (BEMA), método Bayesian (BA) Energia associada com a série (EAS), Dizimação ou Subsampling (SUB), alterada pela neural rugosidade Network (NN-Mod.)*

## Lista de Símbolos

$R^n$	Espacio vectorial
$n$	Dimensión del espacio vectorial $R^n$
$f(x)$	función de suavizado monótona
$Y(t)$	salida de la red neuronal
$y^*(t)$	valor predicho
$h_0$	numero de neuronas iniciales
$H$	Transformada de Fourier
$X^*$	entrada desconocida
$K$	matriz de covarianza
$H$	parámetro de Hurst
$H_A$	parámetro de Hurst medido en el área de serie pronosticada
$H_S$	parámetro de Hurst medido en la serie pronosticada
$K^{-1}$	matriz de covarianza inversa
$l_x$	longitud del vector de entrada
$N_p$	numero de patrones
$i_t$	numero de iteraciones
$I$	Primitiva periódica
$\theta$	parámetro
$x_s$	serie orginial
<i>i.i.d.</i>	igual e idénticamente distribuido
$H_L$	entropía de Renyi
$M$	modelo de datos
$(t_n, t_p)$	partición del intervalo de integración
$Z^{-1}$	operador de retardo
$F_p$	operador del filtro predictor
$\{\{x_n\}, \{x_e\}\}$	serie aumentada, original más predicha
$e_n$	error de predicción de la red neuronal
$e_H$	error de predicción de la red neuronal considerando la rugosidad

$(x_i, y_i)$	par entrada-salida
$\mathbf{w}$	vector de pesos de la red neuronal
$y^D$	serie de dato a decimada
$t_n$	serie de dato con valores objetivos
$\Omega$	espacio muestral
$n$	cantidad de datos
$\tanh$	tangente hiperbólica
$\Gamma$	Función Gama
$B$	Movimiento Browniano
cov	Covarianza
var	Varianza
$B^H$	Movimiento Browniano fraccional
$X_t$	Procesos estocásticos
$e$	Numero neperiano
$V$	Constante positiva
$r$	parámetros de la red neuronal
$f_n$	función de activación sigmoial

## Lista de Siglas

<b>AADECA</b>	Asociación Argentina de Control Automático
<b>ACIATSF</b>	Advances in Computational Intelligence for Applied Time Series Forecasting
<b>AR</b>	Modelo autorregresivo
<b>ARCH</b>	Modelos de heterocedasticidad condicional autorregresivos
<b>ARD</b>	Determinación de relevancia automática
<b>ARIMA</b>	Modelo autorregresivo integrado de media móvil
<b>ARMA</b>	Modelo autorregresivo de media móvil
<b>ASMADI</b>	Asociación Argentina de Matemática Aplicada, Computacional e Industrial
<b>BA</b>	Método Bayesiano
<b>BEA</b>	Método Bayesiano mejorado
<b>BEMA</b>	Método Bayesiano mejorado modificado
<b>BNN</b>	Redes neuronales bayesianas
<b>CADI</b>	Congreso Argentino de Ingeniería
<b>CAI</b>	Congreso Argentino de Agroinformática
<b>CBIC</b>	Congresso Brasileiro de Inteligência Computacional
<b>CERMA</b>	Conference on Electronics, Robotics and Automotive Mechanics
<b>CIF</b>	Computational Intelligence in Forecasting
<b>CIS</b>	Sociedad de Inteligencia Computacional
<b>CONAE</b>	Comisión Nacional de Actividades Espaciales
<b>CONAGUA</b>	Congreso Nacional del Agua
<b>CONEISC</b>	Congreso Nacional de Estudiantes de Ingeniería de Sistemas y Computación
<b>CONICET</b>	Consejo Nacional de Investigaciones Científicas y Tecnológicas
<b>CSCI</b>	International Conference on Computational Science and Computational Intelligence
<b>CSCS</b>	International Conference on Control Systems and Computer Science
<b>EAS</b>	Método basado en la energía de la serie
<b>EM</b>	Máxima expectación

<b>ESTSP</b>	European Time Series Symposium
<b>EUSPN</b>	International Conference on Emerging Ubiquitous Systems and Pervasive Networks
<b>FAMAF</b>	Facultad de Astronomía, Matemática y Física
<b>FARIMA</b>	Modelo autorregresivo integrado de media móvil fraccional
<b>FCEfyN</b>	Facultad de Ciencias Exactas, Físicas y Naturales
<b>FFNN</b>	Red neuronal de retroalimentación hacia adelante
<b>FGN</b>	Ruido Gaussiano fraccional
<b>FL</b>	Lógica Difusa o Fuzzy Logic
<b>GA</b>	Algoritmos Genéticos
<b>GARCH</b>	Modelos de heterocedasticidad condicional autorregresivos generalizado
<b>GP</b>	Procesos Gaussianos
<b>H</b>	Parámetro de Hurst
<b>HEN</b>	Serie temporal de Henon
<b>HM</b>	Markov Oculto
<b>ICANCM</b>	International Conference on Applied, Numerical and Computational Mathematics
<b>IEEE</b>	Institute of Electrical and Electronics Engineers (Instituto de Ingenieros Electricistas Electrónicos)
<b>IJACSA</b>	International Journal of Advance Computer Science and Applications
<b>IJCNN</b>	International Joint Conference on Neural Networks (Conferencia Conjunta Internacional sobre Redes Neuronales)
<b>IK</b>	Serie temporal de Ikeda
<b>INNS</b>	International Neural Network Society (Sociedad Internacional de Redes Neuronales)
<b>INTERCON</b>	Congreso Internacional de Ingeniería Electrónica, Eléctrica y Computación
<b>IJICA</b>	International Journal of Innovative Computing and Applications
<b>KNN</b>	Vecino más cercano
<b>LA-CCI</b>	Latin American Conference on Computational Intelligence (Conferencia latinoamericana de Inteligencia Computacional)



<b>LOG</b>	Serie temporal Logística
<b>Loglog</b>	Gráfico Logarítmico Bidimensional
<b>MA</b>	Modelo de medias móviles
<b>MACI</b>	Congreso de Matemática Aplicada, Computacional e Industrial
<b>MATLAB</b>	MATrix LABoratory, "laboratorio de matrices"
<b><i>mB</i></b>	Movimiento Browniano
<b><i>mBf</i></b>	Movimiento Browniano fraccional
<b><i>mBG</i></b>	Movimiento Browniano Gaussiano
<b>MCMC</b>	Cadena de Markov Monte Carlo
<b>MG</b>	Mackey-Glass
<b>MLP</b>	Perceptrón multipaca
<b>MSP</b>	Método de suavizado promedio
<b>NAR</b>	Procesos autorregresivo no lineal
<b>NARMA</b>	Procesos autorregresivo de medias móviles no lineal
<b>NN-Mod.</b>	Red neuronal modificada por rugosidad
<b>NN3</b>	Neural Networks & Computational Intelligence Competition
<b>NN5</b>	Neural Networks & Computational Intelligence Competition
<b>NSGA</b>	Algoritmo genético multiobjetivo
<b>QUA</b>	Serie temporal cuadrática
<b>PDF</b>	Función de densidad de probabilidad
<b>PI</b>	Intervalo de Predicción
<b>PID</b>	Controlador Proporcional, Integrador y Derivativo
<b>PSO</b>	Optimización por enjambre de Partículas
<b>RBF</b>	Funciones de base radial
<b>RMSE</b>	Raíz del Error Cuadrático Medio
<b>RNA</b>	Red neuronal artificial
<b>RPIC</b>	Reunión para el Procesamiento de la Información y Control
<b>SARIMA</b>	Modelo autorregresivo integrado de media móvil estacional
<b>SECyT</b>	Secretaría de Ciencia y Tecnología
<b>SEEMI</b>	Simpósio en Estadística Espacial y Modelamiento de Imágenes

<b>SIAM</b>	Society for Industrial and Applied Mathematics (Sociedad para Industria y las Matemáticas Aplicadas)
<b>SMAPE</b>	Error Medio Absoluto Simétrico Porcentual
<b>ST-TSF</b>	Pronóstico de semporales de corto alcance
<b>SUB</b>	Método de Decimación o Subsampling
<b>SVM</b>	Máquinas de Soporte Vectorial
<i><b>tanh</b></i>	Tangente hiperbólica
<b>TAR</b>	Modelo autorregresivo por umbral
<b>TDNN</b>	Red neuronal con retardo de tiempo
<b>TIC</b>	Tecnología de la Información en Córdoba
<b>UAV</b>	Vehículo aéreo no tripulado
<b>UTN-FRC</b>	Universidad Tecnológica Nacional – Facultad Regional Córdoba
<b>VAR</b>	Vector autorregresivo
<b>VARIMA</b>	Modelo vectorial autorregresivo integrado de media móvil
<b>WESCIS</b>	Workshop Internacional IEEE Computational Intelligence & Education Society
<b>WCCI</b>	World Congress on Computational Intelligence
<b>wfbmesti</b>	Función de MATLAB para la estimación del parámetro de Hurst del movimiento Browniano fraccionario.

## **Índice General**

Resumen .....	1
Abstract .....	3
Resumo .....	5
Lista de Símbolos .....	7
Lista de Siglas.....	9
Índice General.....	13
1. Introducción .....	17
1.1. Motivación .....	19
1.2. Objetivos de la Tesis .....	21
1.3. Metodología.....	22
1.4. Contribuciones.....	23
1.5. Organización.....	25
2. Marco Teórico y Fundamentos .....	27
2.1. Series Temporales .....	29
2.2. Series temporales como realización del movimiento browniano fraccionario .....	31
2.2.1. Representación de las series temporales de lluvia .....	32
2.2.2. Selección del modelo para las series de lluvia .....	33
2.2.3. Limitaciones del mB al comportamiento de las series de lluvia .....	35
2.2.4. Estimación de la rugosidad de la serie a través del exponente de Hurst.....	35
2.3. El problema del pronóstico.....	36
2.4. Arquitecturas para la predicción de series temporales .....	38
2.5. Redes Neuronales Artificiales.....	39
2.6. Métodos de Aprendizaje.....	42
2.6.1. El algoritmo Levenberg-Marquardt .....	43

2.7. Criterios de medidas de exactitud y errores en el pronóstico.....	44
2.8. Métricas en el pronóstico.....	45
2.9. Filtros Predictores basados en redes neuronales .....	46
2.9.1. Predictor basado en redes neuronales ARMA lineal.....	47
2.9.2. Ajuste heurístico propuesto .....	49
2.9.3. Evaluación para ajuste de estructura .....	50
2.10. Filtro predictor basado en algoritmos genéticos .....	51
2.10.1. Proceso de ajuste .....	52
2.10.2. Entrenamiento y predicción del modelo basado en Procesos Gaussianos.....	53
2.11. Comentarios finales .....	54
3. Modelos basados en suavización de la serie.....	55
3.1. Introducción.....	55
3.2. Predictor NAR basado en rugosidad de la serie o NN-Mod. ....	56
3.2.1. Proceso de aprendizaje basado en rugosidad .....	57
3.3. Predictor basado en la Energía de la Serie.....	59
3.3.1. Método de aproximación.....	60
3.3.2. Configuración del modelo y proceso de aprendizaje .....	62
3.4. Método de Decimación o Submuestreo .....	63
3.4.1. Descripción general del submuestreo.....	63
3.4.2. Metodología Propuesta.....	65
3.4.3. Configuración del modelo y proceso de ajuste .....	66
3.4.4. Ejemplificación .....	67
3.5. Comentarios finales .....	69
4. Métodos basados en Aproximación Bayesiana.....	71
4.1. Introducción.....	71

4.2. Teoría de Aproximación Bayesiana .....	72
4.3. Método Bayesiano de aprendizaje .....	74
4.3.1. Predictor basado en el Método Bayesiano .....	75
4.3.2. Entrenamiento Bayesiano de las RNAs.....	77
4.4. Método Bayesiano mejorado por movimiento Browniano Fraccional .....	78
4.4.1. Motivación .....	78
4.4.1. Método mejorado usando movimiento Browniano Fraccionario .....	79
4.5. Método mejorado Bayesiano modificado usando la entropía Renyi en la serie .....	81
4.5.1. Motivación .....	81
4.5.2. Introducción.....	81
4.5.3. Entropía de Renyi.....	83
4.5.4. Entrenamiento .....	84
4.6. Comentarios finales .....	85
5. Resultados experimentales.....	86
5.1. Introducción.....	87
5.2. Series caóticas artificiales .....	88
5.2.1. Mackey-Glass .....	88
5.2.2. Lorenz.....	90
5.2.3. Hénon .....	91
5.2.4. Ikeda.. .....	92
5.2.5. Mapeo Logístico .....	93
5.2.6. Mapeo Cuadrático .....	94
5.3. Resultados experimentales en series caóticas artificiales.....	95
5.4. Resultados experimentales en series cortas con incompletitud de datos.....	101
5.4.1. Introducción.....	101

5.4.2. Mecanismos de pérdidas de datos.....	103
5.4.3. Método de suavización promedio (MSP).....	103
5.4.4. Performance y errores de predicción.....	104
5.5. Resultados de predicción en series de lluvia mensual acumulada .....	108
5.5.1. Datos meteorológicos .....	109
5.5.2. Agregado de ruido Gaussiano fraccionario.....	110
5.5.3. Configuración de los parámetros de los filtros predictores.....	111
5.5.4. Series de lluvia mensual acumulada .....	112
5.5.5. Performance y errores de predicción.....	115
5.6. Discusión y Comentarios finales .....	122
6. Conclusiones finales .....	124
6.1. Trabajos futuros .....	125
Referencias.....	127
Movimiento Browniano Fraccional .....	162

# Capítulo 1

*“Podemos saber poco del futuro, pero lo suficiente para darnos cuenta de que hay mucho que hacer”*

Alan Turing

## 1. Introducción

En la actualidad, con los avances en el procesamiento de información en el campo de la agricultura, la necesidad de contar con herramientas orientadas a la toma de decisiones [1] [2] [3], ha conducido en la última década a un mayor interés en el desarrollo de modelos de pronóstico de series temporales encaminados a superar las limitaciones teóricas, conceptuales y computacionales que presentan muchas de las técnicas disponibles [4].

Existe una creciente demanda de este tipo de predicciones para variables de superficie, principalmente precipitación y temperatura por numerosos sectores, entre los que se pueden citar el agrícola, industrial, energético, ambiental, biológico, hidrológico, sanitario, transportes, etc.

La disponibilidad que se tiene de grandes cantidades de datos históricos en la actualidad y la necesidad de realizar pronósticos precisos del comportamiento futuro en varios campos de la ciencia demandan la definición de sistemas robustos y eficientes capaces de inferir de las observaciones, la relación de correspondencia entre el pasado y futuro. El estudio de pronósticos ha sido influenciado, a partir de los años sesenta por métodos estadísticos clásicos [5]. Recientemente han surgido los métodos de modelos ajustables mediante técnicas heurísticas, como por ejemplo aprendizaje automático [49] [116] [135] [143], los cuales se han convertido en alternativas promisorias a los modelos clásicos en la disciplina de pronósticos [47].

La clasificación de los métodos de predicción desde el punto de vista estadístico, en general, presenta dos corrientes: la primera está orientada a los métodos causales, como los modelos de regresión [46], y la otra enfocada a las series temporales, donde se incluyen procesos ARMA [49], suavizado exponencial [45], modelos ARIMA [45] y redes neuronales [44].

Desde la estadística, el pronóstico está orientado a la tarea de predecir el comportamiento futuro basado solamente en los datos disponibles obviando el proceso que los generó [50]. Por otro lado, los modelos basados en aprendizaje automático se orientan a encontrar la estructura del proceso generador de los datos con el fin de pronosticar tendencias en el comportamiento a corto, mediano o largo plazo, así como para entender el pasado [33].

Si bien, en la literatura se han propuesto diversos modelos orientados a la representación de series temporales [48], su utilidad depende del grado de similitud entre la dinámica del proceso generador de la serie temporal y la formulación matemática del modelo con que se represente. Por otro lado, las Redes Neuronales Artificiales (RNAs) se usan en problemas reales donde las series temporales suelen tener irregularidades o cambios abruptos o estructurales provocados por datos atípicos, inconsistentes o faltantes que no pueden ser tratados por los métodos clásicos [5]. Por tal motivo, las RNAs constituyen una herramienta muy útil y versátil a los fines de obtener un mejor pronóstico [39].

Muchos autores concuerdan en afirmar que las series de datos reales presentan comportamientos complejos, caóticos, no lineales y la aproximación que pueda hacerse con modelos lineales resulta inadecuada [6] [30]. Para modelos no lineales hay estudios que tratan estos inconvenientes de cambios abruptos o estructurales y reconocen las bondades de las RNAs en el modelado de datos atípicos con perturbaciones [110]. En la actualidad, se continúa en la búsqueda de aplicaciones orientadas a patrones complejos de no linealidad sobre las series temporales, tal como medir la dimensión fractal a través de la estimación del exponente o parámetro de Hurst [7], la cual cuantifica la rugosidad o escasez de suavidad en el sentido estadístico a fin estimar el grado de predictibilidad.

Esto pone en evidencia la necesidad de contar con modelos adecuados para la representación de las características que aproximen al proceso que genera las series temporales afectadas por cambios en su estructura y así obtener un pronóstico más útil para la toma de decisiones. Esta problemática ha llevado al surgimiento de una amplia gama de modelos donde las



RNAs, las cuales son vistas como una técnica no paramétrica de regresión no lineal, han demostrado tener un gran potencial. Zhang *et al.* [8] realizan una profunda revisión de la literatura en el pronóstico de series de tiempo con redes neuronales, incluyendo las publicaciones más relevantes y los tópicos de investigación hasta la penúltima década.

Las RNAs se emplean actualmente como alternativas atractivas ante el problema planteado, ya que permiten extraer las dinámicas no lineales desconocidas presentes entre las variables significativas y la serie temporal para hacer predicciones sin necesidad de realizar ningún tipo de supuestos. A partir de esta última afirmación, muchas conclusiones acerca del error de desempeño en el horizonte de pronóstico de las RNAs son obtenidas a partir de estudios empíricos, presentando así resultados limitados que a menudo no pueden ser extendidos de forma general. En la bibliografía aparecen muchas publicaciones que indican un mejor desempeño de las RNAs con respecto a otras metodologías [32] [27]. Estas son debidas a que están relacionadas con paradigmas novedosos o extensiones de métodos existentes, arquitecturas y algoritmos de entrenamiento, pero carecen de una evaluación confiable y válida de la evidencia empírica de su error de desempeño en la predicción [9].

Durante las últimas décadas se han propuesto distintas métricas para evaluar la performance en el horizonte de pronóstico [10], donde se muestra que los métodos para medir el error en el pronóstico han mejorado [11], debido a características específicas de las series temporales de datos. Por otro lado, debido a las debilidades conocidas de las medidas existentes [147], es conveniente el uso tanto de una medida simétrica (SMAPE - Error Medio Absoluto Simétrico Porcentual) como de una medida cuadrática (RMSE - Raíz del Error Cuadrático Medio). Estas tienen la ventaja de ser más robustas a la hora de cuantificar el error de desempeño en el horizonte de pronóstico [12].

### 1.1. Motivación

La motivación de esta investigación surge de la necesidad de disponer de un método científico para la predicción de series temporales de lluvias confiable. Así, el propósito de la metodología es obtener, mediante modelos no lineales basados en RNAs, un horizonte de pronóstico a corto o largo plazo apto para tomar decisiones agrícolas.

Aún con los grandes avances de los modelos estadísticos, los desafíos matemáticos y computacionales apuntan a mejorar los algoritmos de predicción [13] [14] [15] [16] para ser usados en el ámbito agrícola. Las variables incluyen temperatura, humedad y precipitación [17] [18], como así también la incorporación de datos masivos usando nuevas tecnologías, como el Big Data [19].

Las nuevas herramientas tecnológicas [20] se orientan a predecir variables principalmente relacionadas con cantidad de lluvia y humedad de suelo, como se muestra en la Fig. 1-1. Esta potencial aplicación apunta a la toma de decisiones en los siguientes escenarios:

- Producción agrícola. La predicción de cantidad de agua y humedad de suelo es fundamental para la toma de decisiones en cuanto al tipo, calidad, cantidad y área de semillas a ser sembradas.
- Análisis crediticios destinados a la industria agropecuaria. La predicción de ciertas variables agrícolas puede inferir en cuanto al rendimiento real a ser obtenido, dándole un sólido soporte tecnológico al análisis de riesgo que realiza la entidad prestadora de dinero.
- Construcción de grandes obras de infraestructuras. La predicción de cantidad de agua puede evitar su falta o exceso, dando soporte a la toma de decisiones al momento de la planificación territorial.

Por lo tanto, esta tesis considera de gran importancia estudiar los métodos actuales, sus alcances, limitaciones y compararlos con nuevos métodos aquí propuestos de predicción [16] [21], [22] [23] [24] [25] [26]. En particular, aquí se utilizan redes neuronales [27] [28] [29] teniendo en cuenta los errores de desempeño que presentan en el horizonte de pronóstico [30].

El desarrollo local de esta tecnología es de gran relevancia [31], dado que su aplicación abarca muchas de las áreas prioritarias contempladas en el Plan Estratégico Nacional en Ciencia, Tecnología e Innovación de la Nación Argentina.

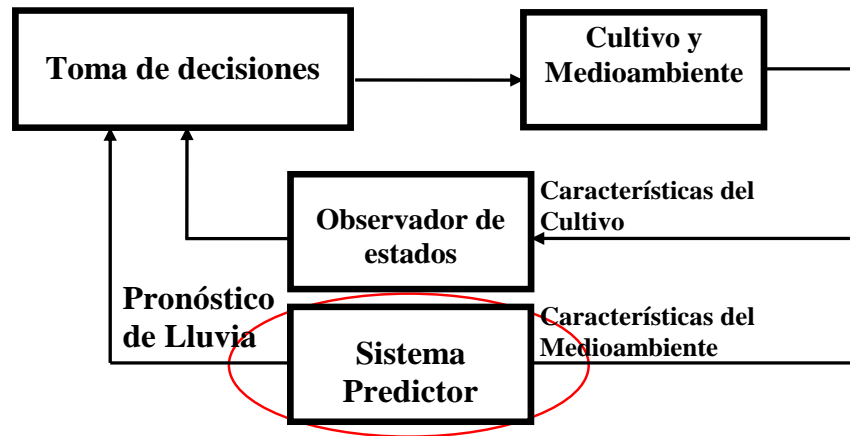


Fig. 1-1. Sistema de control a lazo cerrado que considera el modelado de variables asociadas al cultivo para la predicción, series de lluvia.

## 1.2. Objetivos de la Tesis

El objetivo general de esta tesis es desarrollar una metodología para pronosticar series temporales de alta rugosidad en el sentido estadístico. Es esencial que la misma sea capaz de pronosticar series temporales de lluvia mensual acumulada con mayor precisión y eficacia en el horizonte de pronóstico para ser usado como soporte en la toma de decisiones agrícolas. El objetivo de esta tesis se puede desglosar en los siguientes objetivos particulares:

- Proponer un modelo matemático generador del proceso subyacente de las series temporales;
- Desarrollar métodos de predicción efectivos y prácticos basado en:
  - Modelos de suavización de las series temporales de alta rugosidad;
  - Definición de la teoría de aproximación bayesiana;
  - Completar datos faltantes en la serie temporal;
- Mejorar el diseño de los métodos de predicción de acuerdo a los propuestos actualmente en la literatura y encontrar combinaciones de selección de parámetros para los modelos de redes neuronales;

- Aplicar los métodos de predicción propuesto y comparar su error de desempeño medido a través los índices SMAPE y RMSE con los resultados reportados en la literatura en series temporales del tipo:
  - Caóticas artificiales bien conocidas;
  - De lluvia mensual acumulada.

Como se demostrará en los siguientes capítulos, los objetivos planteados al inicio de esta Tesis Doctoral pudieron ser alcanzados a partir de la publicación de diferentes trabajos científicos.

### 1.3. Metodología

La primera parte de la Tesis se concentra en estudiar esquemas eficientes basados en suavización de las series para el entrenamiento y evaluación de las RNAs. Dado que las series temporales reales presentan datos atípicos, la predicción resulta dificultosa con los modelos tradicionales. Para cumplir con este requerimiento:

- Se realizó un análisis detallado del estado del arte, que incluye los algoritmos existentes en la literatura con los problemas puntuales mencionados.
- Se plantearon posibles soluciones basadas en estrategias simples y confiables. Aquí se desarrollaron las propuestas de suavización de la serie temporal, que consisten en suavizar la serie y modificar heurísticamente la topología de la RNA, tomando como nuevas entradas las series suavizadas para hacer pronóstico.
- Se emplearon modelos que permitieron estudiar las estrategias propuestas. Se utilizaron algoritmos tradicionales de referencia, como algoritmos ARMA, NARMA, los cuales son ampliamente conocidos.

La segunda parte de la Tesis se enfoca en estudiar métodos alternativos derivados de la teoría de aproximación bayesiana. Para cumplir con estos requisitos:

- Se desarrollaron herramientas de pronóstico que posibiliten caracterizar el comportamiento de las series temporales a fin de ser usado para modelar los parámetros de una RNA.
- Se plantearon soluciones basadas en estrategias bayesianas para el entrenamiento de RNAs. Aquí se desarrollaron las propuestas de método mejorado bayesiano y

mejorado bayesiano modificado, que consisten en modelar los pesos y parámetros de la RNA usando inferencia bayesiana, y modificando heurísticamente la topología.

- Se usó la entropía de información medida en la serie para ser aplicados a la predicción de series temporales, en particular, cortas.
- Se utilizaron algoritmos tradicionales de referencia, como el algoritmo BA, los cuales son ampliamente conocidos.

Para llevar adelante un estudio empírico del desempeño de estos métodos de pronóstico se realizaron los siguientes pasos:

- Se emplearon series de lluvia reales obtenidas de diferentes puntos geográficos de la provincia de Córdoba, Argentina. Además, se utilizaron otras series de referencia obtenidas por métodos matemáticos como Mackey Glass, Lorenz, Hénon, Ikeda, Logístico y Cuadrático.
- La implementación se realizó en un entorno de simulación mediante software de cálculo numérico.
- Se estudiaron los métodos para completar los datos faltantes en una serie temporal a fin de proponer un método eficiente y práctico para solucionar este problema.
- La evaluación de los resultados obtenidos se realizó mediante la comparación con métodos lineales y no-lineales. Para medir el error de desempeño en el horizonte de pronóstico, se usaron los índices SMAPE y RMSE.

### 1.4. Contribuciones

En esta tesis se establece un marco teórico general que permite la comparación de los resultados de los distintos métodos de predicción de series temporales, en particular de lluvia, lo cual se puede resumir en las siguientes contribuciones.

A partir de los modelos lineales y no lineales en la literatura, se propuso un esquema eficiente para el entrenamiento y evaluación de las RNAs basados en modelos no lineales NARMA para pronosticar series rugosas. Esta contribución se basa en métodos de suavización de la serie [33] desarrollados en el Capítulo 3.

- El primero, llamado *rugosidad de la serie* o NN-Mod. [34] propone una ley heurística para establecer el proceso de ajuste y de la topología de la RNA, asumiendo que la serie temporal pronosticada tenga la misma rugosidad que la serie dato original.
- El segundo, llamado *Energía asociada a la serie* (EAS) [35] reproduce los valores de la serie temporal a partir de las áreas de la serie temporal pronosticada para generar una primitiva calculada como una nueva entrada a la RNA.
- El último, llamado *Submuestreo* (SUB) [36] genera series temporales mediante la técnica de dividir la serie original en subseries para ser usadas como nuevas entradas a la RNA y generando pronósticos que luego es promediado a la salida del predictor.

El segundo aporte se basa en el desarrollo de métodos de aproximación bayesianos desarrollados en el Capítulo 4.

- El primero, llamado *método bayesiano* (BA) [37], modela los parámetros y la estructura de una RNA mediante la inferencia bayesiana en el supuesto que los parámetros siguen una regla estocástica para el ajuste de los pesos sinápticos.
- El segundo, llamado *método mejorado bayesiano* (BEA) [38], usa la función de densidad del *movimiento Browniano fraccionario* (*mBf*) para modelar los parámetros de las redes neuronales, que es una alternativa del modelo propuesto como referencia BA.
- El tercero, un *método mejorado bayesiano modificado* (BEMA) [39] que usa la propiedad de la entropía Renyi de la información en la serie dato, combinado con el método BEA para la predicción de series cortas de lluvia. Aquí se aborda la complejidad inherente de los datos de las series temporales en el mundo real, la cual resulta de la dimensión del modelo y el hecho que el proceso de generación sea no lineal y no estacionario; además de problemas de ruido, el cual esta correlacionado con la predicción en series con baja cantidad de datos.

En el último aporte, el problema de datos incompletos se estudió experimentalmente en la Sección 5.4. Allí se plantea la dificultad para el análisis de series temporales incompletas. Para la sustitución de los datos faltantes dentro de la serie temporal se propone el método de suavizado promedio (MSP) [40] [41]. Una gran ventaja de la solución propuesta es que al aumentar la cantidad de datos que faltan, la complejidad no se incrementa en igual medida.

## 1.5. Organización

La tesis se encuentra estructurada de la siguiente manera: en el Capítulo 2 se brinda el marco teórico y contexto que da origen a los modelos y métodos propuestos en los capítulos subsiguientes donde se hace una revisión del estado de arte en los modelos para hacer pronóstico de series temporales. Luego se hace una descripción de las diferentes técnicas existentes para analizar series temporales no lineales con el predictor basado en RNAs. Además, se introducen los problemas de pronóstico a los que se ha enfrentado esta tesis como son, por una parte, la identificación del modelo generador de los datos de la serie asumido como una traza del  $mBf$ , y por otra la estimación de densidades predictivas y el cálculo para distintos horizontes de pronóstico empleando modelos estocásticos. En los Capítulo 3 y Capítulo 4 se presentan nuevos métodos basados en suavización de la serie y aproximación Bayesiana como principales aportes de esta tesis. En el Capítulo 3 se proponen dos métodos, uno basado en la Energía asociada a la serie (EAS) y el método de Decimación o Submuestreo (SUB), donde se toma como referencia al método basado en rugosidad de la serie (NN-Mod.), que dan lugar a mejoras sustanciales en los rendimientos para los distintos horizontes de predicción.

En el Capítulo 4 se propone y evalúa un modelo basado en inferencia bayesiana que permite abordar un conjunto de problemas que se enmarcan en el pronóstico de series temporales dependiendo de la cantidad de datos (con o sin un gran número de muestras). Se analiza la representación y desarrollo de estos modelos en modelos predictores basados en RNAs empleando el método bayesiano mejorado modificado (BEMA) y el método mejorado bayesiano (BEA), donde se toma como referencia el método bayesiano (BA). En el Capítulo 5 se presentan los principales resultados de esta Tesis Doctoral basados en un estudio experimental respecto de la complejidad de la representación de las series temporales de lluvia y su influencia en problemas de conceptos, en conjunto de datos de diversos tamaños, completos e incompletos presentados a través de series benchmark de la literatura. Como nuevo aporte experimental, en la Sección 5.4 se propone un método de suavizado promedio (MSP) para completar los datos faltantes en una serie incompleta. Se evalúan los distintos modelos predictores con esquemas de completitud e incompletitud en los datos observados. Para comparar los resultados de los modelos estudiados se utilizan los índices de desempeño SMAPE (Error Medio Absoluto Simétrico Porcentual) y RMSE (Raíz del Error Cuadrático Medio) comúnmente usados para evaluar métodos de pronóstico. Los

## *Capítulo 1. Introducción*

---

resultados obtenidos con los modelos de pronóstico propuestos para series de lluvia mensual acumulada de diferentes puntos geográficos y series caóticas artificiales de referencia usados en la bibliografía para describir sistemas caóticos, se puede constatar que los índices SMAPE y RMSE son más bajo que los algoritmos de referencia. Finalmente, en el Capítulo 6 se presentan conclusiones y trabajos a futuro.



# Capítulo 2

*"Hacer predicción es muy difícil, especialmente cuando se trata del futuro."*

Niels Bohr

## 2. Marco Teórico y Fundamentos

En este capítulo se presenta el marco de soporte de esta tesis donde se describen los modelos usados en la literatura para representar tanto el modelo matemático subyacente de las series temporales como los modelos usados en el pronóstico. Se hace una revisión de los trabajos de investigaciones anteriores y actuales en el pronóstico de series temporales existentes en la literatura. Para la identificación del modelo generador de los datos de la serie, se suponen que las series temporales se comportan y ajustan al caso de la realización de una traza del movimiento Browniano fraccionario de largo alcance, tanto las series de lluvia acumulada mensual como así también las series caóticas artificiales de referencia usadas en la literatura para predecir fenómenos meteorológicos y físicos. Luego se procede a establecer el modelo NAR de partida basado en redes neuronales para hacer pronóstico a corto y largo plazo, en contraste con otras técnicas usadas en esta Tesis como referencia, tales como las redes neuronales ARMA y algoritmos genéticos. A fin de una mejor descripción, se muestran en la Fig. 2-1 los métodos más usados en la literatura para pronóstico de series temporales.

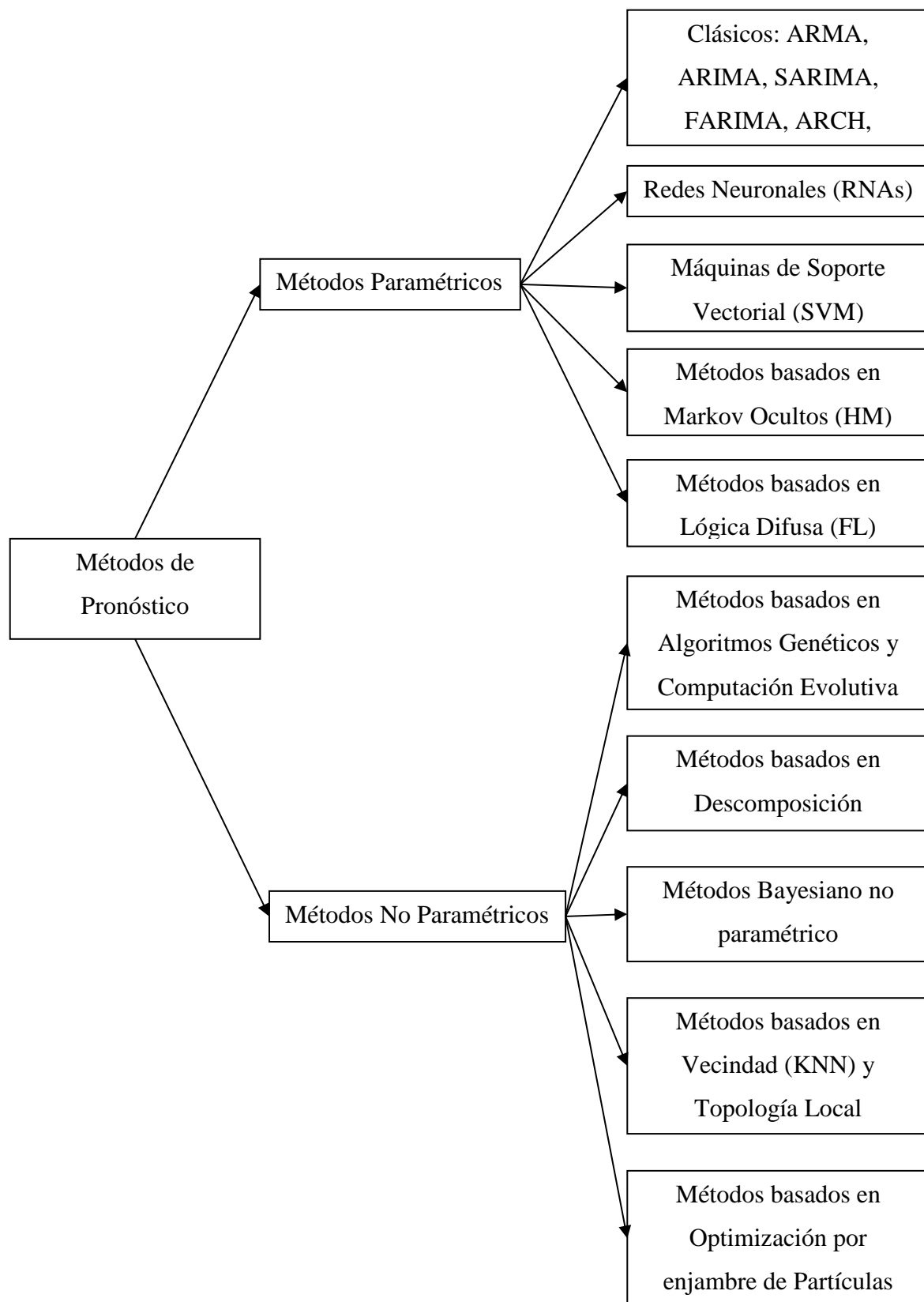


Fig. 2-1. Clasificación de los modelos de pronóstico de series temporales.

## 2.1. Series Temporales

En esta sección, además de las definiciones de serie temporal y proceso estocástico, se hace especial hincapié en las series temporales como una realización de un proceso estocástico dado [42]. Si bien el objetivo de esta Tesis se centra en el desarrollo de modelos no lineales de pronóstico, el modelo matemático subyacente de la serie temporal [43] es fundamental a la hora de preprocesar los datos históricos, ya que, al ser sometidas por los modelos propuestos, se extraen características que serán usadas en los modelos para hacer pronóstico.

En la actualidad, el análisis de series temporales es imprescindible en muchos campos de la ciencia, tales como la ingeniería [44] y la economía [45], esto es investigar cómo ha evolucionado una variable de interés hasta el momento puede ser de gran utilidad para predecir su comportamiento futuro. Sin embargo, suponer que las series temporales tienen un comportamiento análogo reviste de gran interés, en particular para la toma de decisiones en el área de la agricultura de precisión para su modelamiento.

Se define a la serie temporal como un conjunto de observaciones sobre valores que toma una variable (cuantitativa) en diferentes momentos del tiempo. Los datos se pueden comportar de diferentes formas a través del tiempo [46], esto es que presente una tendencia, un ciclo; no tener una forma definida o aleatoria, variaciones estacionales (anual, mensual, etc.).

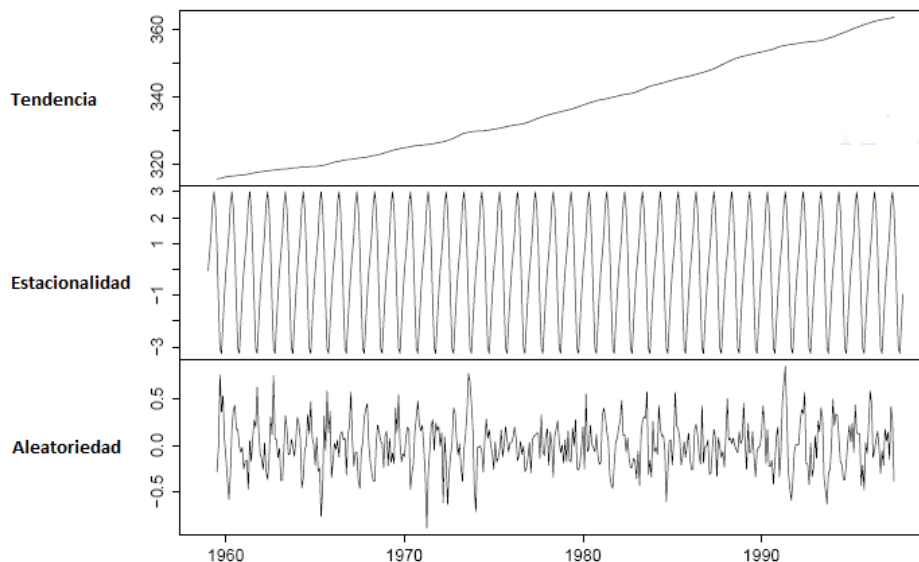
Una serie temporal puede estar constituida solo por eventos determinísticos, estocásticos o una combinación de ambos [47]. Se sabe que muchas series temporales presentan comportamientos dinámicos no lineales, cuya complejidad hacen que sea imposible formular un modelo matemático basado en leyes físicas o económicas que representen su evolución adecuadamente. El problema de la formulación del modelo se ve agravado por la presencia de observaciones atípicas y cambios estructurales, para los cuales no existen modelos matemáticos que permitan su representación en el caso no lineal [48].

Los métodos tradicionales para el análisis de series temporales [49] se hacen a través de la descomposición de las mismas en varias partes. Se dice que una serie temporal puede descomponerse en tres componentes que no son directamente observables, de los cuales únicamente se pueden obtener estimaciones. Estos tres componentes son:

**Tendencia:** representa el comportamiento predominante de la serie. Esta puede ser definida no formalmente como el cambio de la media a lo largo de un extenso período de tiempo.

**Estacionalidad:** es un movimiento periódico que se producen dentro de un periodo corto y conocido. Este componente está determinado, por ejemplo, por factores climáticos.

**Aleatorio:** son movimientos erráticos que no siguen un patrón específico y que obedecen a causas diversas. Este componente es prácticamente impredecible. Estos comportamientos representan todos los tipos de movimientos de una serie temporal que no son tendencia ni variaciones estacionales ni fluctuaciones cíclicas.



**Fig. 2-2.** Descomposición de una serie temporal en sus componentes principales.

Está claro que una cantidad importante de datos muestrales es necesaria para que el análisis sea representativo en la población general a la que pertenece la serie. Por otro lado, si se conoce suficiente información que fundamenten las causas del comportamiento de una serie, el análisis de las mismas se torna accesorio, pero no indispensable [50]. En general, mediante el análisis de series temporales se persigue la posibilidad de efectuar una predicción. Por predicción se entiende la estimación de los valores futuros de una serie, en función del comportamiento pasado de la misma.

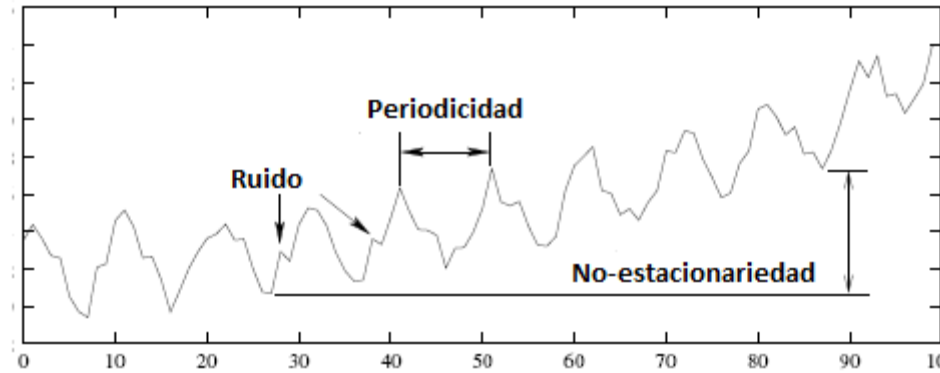


Fig. 2-3. Ejemplo de una serie temporal no lineal, no estacionaria y ruidosa.

## 2.2. Series temporales como realización del movimiento browniano fraccionario

Aquí se supone que las series temporales se comportan como una realización de una traza del movimiento Browniano fraccionario (*mBf*), cuya definición es la dada por los primeros trabajos de [51] [52].

El *mBf* se define según [71] mediante,

$$B_H(t) = \frac{1}{\Gamma\left(H + \frac{1}{2}\right)} \left( \int_{-\infty}^0 \left( (t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}} \right) dB(s) + \int_0^t (t-s)^{H-\frac{1}{2}} dB(s) \right) \quad (2-1)$$

La estimación del exponente de Hurst  $H$  fue desarrollada originalmente en la hidrología. Sin embargo, las técnicas modernas para estimar el exponente de Hurst vienen de las matemáticas fractales. Este parámetro surge de la generalización de la definición del Movimiento Browniano, pasando la integral de orden 1 a la de orden  $H$  fraccionario, resultando que para un  $H=0,5$  se tiene el Movimiento Browniano Gaussiano (*mBG*), mientras que para valores diferentes se tienen varianzas positivas o negativas, donde  $H$  está en el conjunto  $[0,1]$ .

El integrador  $B$  es un proceso estocástico, comúnmente movimiento Browniano, donde  $B$  se recupera con  $H=1/2$ . Aquí se asume a  $B$  definido en un espacio de probabilidad  $(\Omega, F, P)$ ,  $H$  es el parámetro de Hurst y  $\Gamma(\cdot)$  es la función Gamma.

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx, \quad (2-2)$$

Este tema se amplía en el APENDICE. El  $mBf$  es autosimilar en distribución y la varianza de los incrementos está dado por

$$Var(B_H(t) - B_H(s)) = \nu |t - s|^{2H} \quad (2-3)$$

donde  $\nu$  es una constante positiva.

Esta forma especial de la varianza de los incrementos sugiere diversas formas de estimar el parámetro  $H$ . Así, hay diversos métodos para calcular  $H$  asociado a un  $mBf$  [53] [54] [55] [56]. En la Fig. 2-4 se muestran distintas realizaciones del  $mBf$ . Mediante el parámetro  $H$  se tiene una idea de la suavidad de la función. Para  $H = \{0,2; 0,5; 0,8\}$  se tiene las siguientes trayectorias:

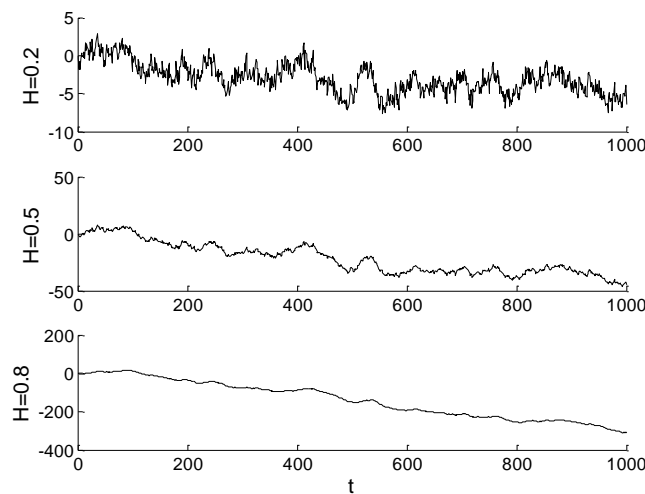


Fig. 2-4. Distintas trazas del  $mBf$ .

Así,  $H$  mide la rugosidad en sentido estadístico de la señal, determinando si la señal es nerviosa o no respecto del  $mB$  normal. Con  $H > 0,5$  se tiene una señal más suave que con  $H < 0,5$ , que además da idea de dependencia estocástica de los eventos de corto y largo alcance, respectivamente [57] [58] [59].

### 2.2.1. Representación de las series temporales de lluvia

El cambio climático durante los últimos 30 años [60] [61] parece ser bastante peculiar en el sentido de que uno es testigo de las sequías [62] y las inundaciones [63] en lugares donde era difícil esperar de ellos. Tal evolución de la variabilidad hidrológica [64] es diametralmente opuesta a las necesidades de la comunidad [65], desde que la precipitación es mucho más útil cuando se

produce durante la temporada de siembra, que cuando ocurre en otros momentos del año. De hecho, una previsibilidad confiable de los cambios diarios y estacionales [66], así como las inclemencias del tiempo, son importantes para determinar el comienzo de siembra de un cultivo. En particular, el problema de las reservas de agua potable y la generación de energía en abundancia hablan de la importancia de los modelos de distribución de las precipitaciones [67]. Sin embargo, la cuestión de hacer una predicción confiable de la distribución estacional de las lluvias [68] [69] para cualquier lote o zona agrícola es ahora mucho más difícil de resolver. Esto es así porque no es fácil identificar las causas naturales y antropogénicas del cambio climático en el último siglo [70] con la certeza científica. Es por ello que se utilizan los conceptos de la geometría fractal [71] para la descripción y predicción de procesos climáticos. Por otro lado, para el mejor conocimiento de los lectores, la aproximación mediante modelos fractales para las series históricas de lluvia [72] [73] [74] [75] no ha tenido la debida importancia en el análisis de los sistemas climáticos, por lo cual la pregunta sigue siendo abierta [76]. Por lo tanto, esta Tesis propone el modelo de distribución del  $mBf$  como representante del proceso generador de la naturaleza fractal [77] y caótica [78] en las precipitaciones mediante el cálculo del exponente de Hurst  $H$  [79].

### 2.2.2. Selección del modelo para las series de lluvia

Se puede inferir que las series temporales provenientes de las precipitaciones de lluvia poseen un comportamiento no determinista [80] y fractal [81]. Es por ello que asumimos que estas series se comportan como una realización de la traza unidimensional del movimiento browniano fraccionario, es decir, que posee *autosimilaridad* y tiene *memoria o dependencia de largo alcance* [82] [83] [84].

Una serie temporal de lluvia generalmente está integrada por una componente estocástica y una determinista. La componente determinística puede ser clasificada como, tendencia, saltos, periodicidad o combinación de éstas [85]. Los intervalos de tiempo en las series de tiempo de lluvia son generalmente del orden de horas, días, semanas, meses o años.

Los cambios estructurales en las series históricas de lluvia pueden ser debidos a diferentes factores. Un tipo de cambio puede ser gradual, esto es conocido como tendencia o de forma abrupta o salto y puede adoptar formas más complejas [86] [87] [88]. La Fig. 2-5 muestra una serie de

lluvia mensual y la Fig. 2-6 una serie de lluvia anual proveniente de localidades del sur de la Provincia de Córdoba.

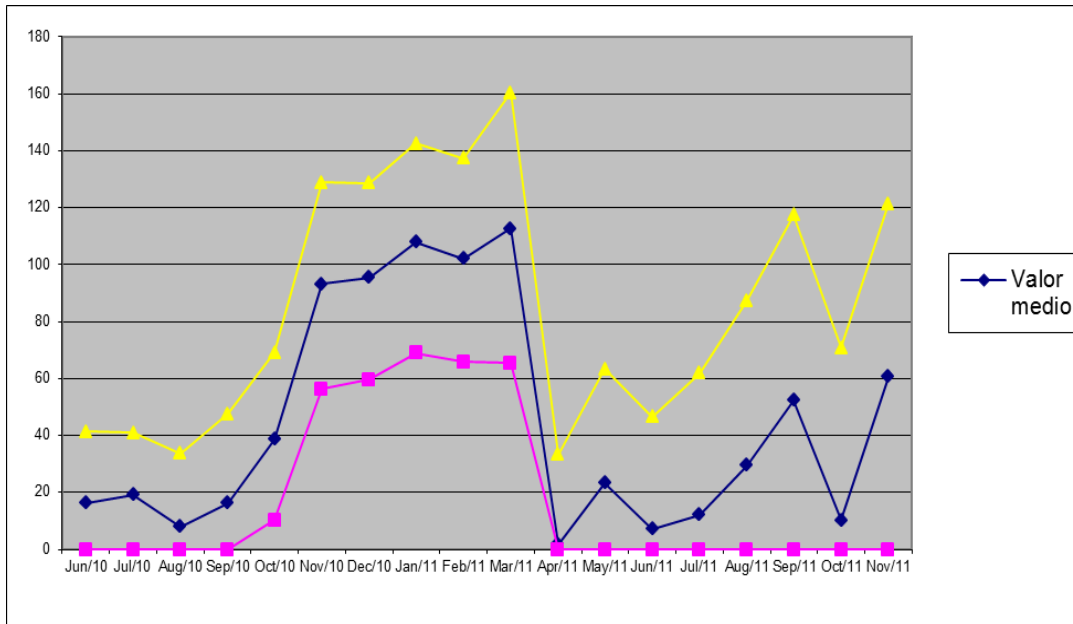


Fig. 2-5. Lluvia mensual de la localidad Despeñaderos, Establecimiento Santa Francisca, Provincia de Córdoba entre los años 2010-2011.

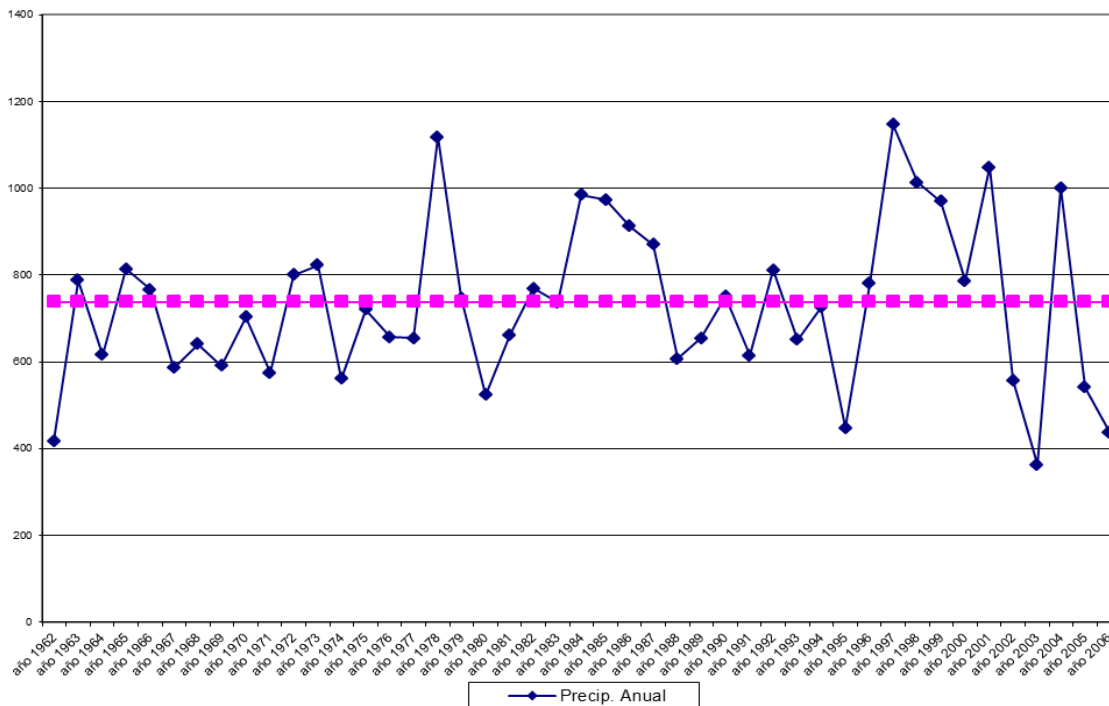


Fig. 2-6. Lluvia anual de la localidad La Perla, Provincia de Córdoba entre los años 1962-2006.



### 2.2.3. Limitaciones del $mB$ al comportamiento de las series de lluvia

A pesar de la fácil manejabilidad que ha demostrado tener el movimiento Browniano con memoria de largo alcance [89] [90], lo que ha propiciado el desarrollo de todas estas teorías, que son de suma utilidad y aplicados hoy día por doquier. Dicho proceso estocástico representa las características de una serie de lluvia con las características del movimiento browniano cuando:

- No hay estacionariedad aparente de los datos de series de lluvia. Las variaciones parecen tener una distribución no estacionaria, es decir, se ven periodos de mucha variación seguido de periodos muchos más tranquilos [91].
- Hay concentración de la variabilidad. La variabilidad no sólo no es constante, sino que además se concentran los periodos de alta variabilidad, es decir los periodos en los que la variación de precipitación de lluvia mensual es elevada no se encuentran aislados, sino concentrados temporalmente [92].

Por ende, el método de “*dimensión fractal*” requiere computación intensiva y aumenta la precisión en series temporales más cortas [93] [94].

### 2.2.4. Estimación de la rugosidad de la serie a través del exponente de Hurst

El análisis de wavelets y el método de transformación de wavelets [55] ha sido implementado para la estimación del exponente o parámetro de Hurst [95].

En esta investigación propone que el modelo matemático subyacente del comportamiento de las series de lluvia manifiesta auto-similitud en ausencia de una longitud natural, es decir, posee dependencia o memoria de largo alcance.

El análisis de las irregularidades de los datos de las series, su comportamiento temporal y espectral, implica la necesidad de estimar el parámetro Hurst  $H$  para este modelo. Los principales métodos paramétricos para estimar la auto-similitud son los siguientes:

- Los métodos espectrales: log-periodograma, una variante del método de Lobato y de Robinson [96].
- Máxima verosimilitud: estimador de Whittle [97].
- Métodos Tiempo escala: La descomposición wavelet del  $mBf$  [79].
- Los métodos temporales: número de pasos a nivel, variaciones discretas [98].

Por ejemplo, el valor de Hurst medido en la serie temporal de lluvia La Perla (1962-2001) es de 0.017. Se busca conocer el comportamiento fractal por el crecimiento local del método de la varianza, es decir, mediante el uso de la función *wfbmesti del toolbox Wavelet de MATLAB* para la estimación del parámetro H de movimiento browniano fraccional a partir de una realización en una dimensión.

La herramienta *wfbmesti* [99] ofrece tres estimaciones diferentes. Más precisamente, la primera, debido a [53], se basa en la derivada de segundo orden discreto. La segunda es una adaptación basada en ondas y tiene propiedades similares [55]. El tercero, propuesto por [100], estima H utilizando la pendiente de la gráfica logarítmica bidimensional (loglog). De esta forma, hay diferentes métodos para computar el parámetro H asociado a un movimiento Browniano, por lo tanto este trabajo adopta para estimar el exponente H de una traza de *mBf* por el método Wavelets [55].

### 2.3. El problema del pronóstico

**Definición de pronóstico:** predicción de la evolución de un proceso o de un hecho futuro a partir de criterios lógicos o científicos.

El término **predicción** puede referirse tanto a la acción y al efecto de predecir como a las palabras que manifiestan aquello que se predice; en este sentido, predecir algo es anunciar por revelación, ciencia o conjetura algo que ha de suceder.

En términos estadísticos, el pronóstico es el proceso de estimación en situaciones de incertidumbre [46]. El término predicción es similar, pero más general, y usualmente se refiere a la estimación de series temporales o datos instantáneos.

La predicción es un *problema importante* que abarca muchos campos, incluyendo todas las ciencias, los negocios, la industria, el gobierno, la economía, la política y las finanzas. Los problemas de predicción a menudo se clasifican en corto, mediano plazo y largo plazo. Los problemas a corto plazo implican la predicción de eventos sólo en períodos de tiempo (días, semanas, meses) dependiendo del problema [5]. El término mediano plazo se extiende de uno a dos años en el futuro, y la predicción a largo plazo puede extenderse más allá. Por lo general pensamos en un pronóstico como un simple número que representa nuestra mejor estimación del valor futuro de la

variable de interés. Los estadísticos lo llaman robustez en el pronóstico estadístico [95] [101]. Ahora, este pronóstico es casi siempre erróneo; es decir, experimentamos un *error de pronóstico*. Por consiguiente, una buena práctica es acompañar a un pronóstico con una estimación del error. Una forma de hacer esto es proporcionar un intervalo de predicción (PI) para acompañar el pronóstico puntual [102] [103]. Otras características importantes del problema de pronóstico son el horizonte de proyección y el intervalo de pronóstico. El horizonte de predicción es el número de períodos futuros para los cuales deben hacerse las predicciones, y a menudo es dictado por la naturaleza del problema.

Encontramos dos interpretaciones del problema de pronóstico que existen sobre la base de datos históricos. La teoría del pronóstico estadístico [104] asume que una secuencia observada es una realización específica de un proceso aleatorio, donde surge la aleatoriedad de muchos grados de libertad independientes que interactúan linealmente [5] [105]. Sin embargo, según el punto de vista de la teoría de sistemas dinámicos [106], es que el comportamiento aparentemente errático puede ser generado por los sistemas deterministas con sólo un pequeño número de grados de libertad, la interacción no lineal. Este comportamiento complicado y aperiódico también se llama caos determinista [107].

El problema básico de predicción puede ser formulado como sigue. Dados los valores pasados de un proceso uniformemente espaciados en el tiempo tales como  $x(n-T)$ ,  $x(n-2T)$ , ...,  $x(n-mT)$ , con un intervalo de muestreo  $T$  y un orden de predicción  $m$ , se desea predecir el valor presente  $x(n)$  de ese proceso. De esta manera se busca obtener la mejor predicción (en algún sentido) de los valores presentes de una señal (aleatoria o pseudo-aleatoria).

Algunos métodos causales de pronóstico asumen que es posible identificar los factores subyacentes que pueden tener influencia sobre la variable a pronosticar [108]. Si estas causas se entienden, se pueden hacer proyecciones de las variables que influyen, para utilizarlas en la predicción.

Algunos de estos métodos son [109]:

- Análisis de la regresión, que puede ser lineal o no lineal.
- Modelo autorregresivo de media móvil (ARMA)
- Modelo ARIMA
- Econometría

Sin embargo, todos los intentos de pronosticar y reproducir los datos históricos de fenómenos naturales, tal como la precipitación de lluvia, son procesos necesarios para ganar la comprensión de los mecanismos pertinentes que muestran que el clima sólo puede ser interpretado como un sistema estocástico. En concreto, en vista de las altas relaciones no lineales que rigen el fenómeno de lluvia, el pronóstico a largo plazo puede pensarse de una manera estocástica. Este es un intento de examinar la naturaleza de los procesos (estocástico/fractal) como modelos que pueden representar esta variabilidad en hidrometeorología, como se presentó anteriormente en la Sección 2.2.2.

Esta Tesis emplea una estimación del modelado fractal de las series de lluvia y series caóticas artificiales medido a través del parámetro de Hurst, el cual es usado en los algoritmos para modificar los modelos, los parámetros y topología de las RNAs propuestas aquí.

### 2.4. Arquitecturas para la predicción de series temporales

Un aspecto importante de la predicción está representado por el tamaño del horizonte. Si la predicción a un paso de una serie temporal ya es un reto, la realización del pronóstico a pasos múltiples es más difícil [103] debido a condiciones adicionales tales como acumulación de errores, la reducción de la precisión, y el aumento de la incertidumbre [110] [111]. El dominio de la predicción ha sido influenciado por un largo tiempo por métodos estadísticos lineales tales como los modelos ARIMA [5]. Sin embargo, a finales de 1970 y principios 1980, se hizo cada vez más claro que los modelos lineales no se adaptan a muchas aplicaciones reales [32]. En el mismo período, varios métodos no lineales se propusieron como el modelo bilineal [112], el modelo umbral autorregresivo TAR [113] y el modelo heterocedástico condicional autorregresivo (ARCH) [45]. Sin embargo, el estudio analítico no lineal de las series temporales y la predicción está todavía en pañales en comparación con series temporales lineales [114].

En las tres últimas décadas, los modelos basados en aprendizaje automático han llamado la atención y se han establecido como modelos contrincantes a los modelos estadísticos clásicos en la comunidad de pronóstico [44]. Estos modelos, son también llamados de caja negra o modelos basados en datos [115], y son ejemplos de modelos no lineales, que utilizan únicamente los datos históricos para aprender la dependencia estocástica entre el pasado y el futuro. Uno de los primeros fue Werbos [116] quien encontró que las RNAs superaban a los métodos estadísticos clásicos como

la regresión lineal y Box-Jenkins [8]. Más tarde un estudio similar lo llevó a cabo Lapedes y Farber [117] que concluyeron que las RNAs pueden utilizarse con éxito para el modelado y la predicción de series temporales no lineales. Ante esto aparecieron otros modelos tales como arboles de decisiones, máquinas de soporte vectorial y regresión del vecino más cercano [109], algoritmos genéticos y basados en Lógica Difusa o Fuzzy (FL) [118]. Además, la exactitud empírica de estos modelos de aprendizaje automático han sido evaluados en competiciones internacionales de pronósticos bajo diferentes condiciones de los datos, por ejemplo las competiciones internacionales de pronóstico propuestas por Makridakis M3 [119], Santa Fe Competición [110], las Competiciones NN3 [120] y NN5 [121], y las ESTSP anuales [122] creando debates científicos interesante en el área de minería de datos y la predicción [123], así como sesiones especiales organizadas en eventos internacionales, tal como el *International Joint Conferences on Neural Networks* [125] [126] e *IEEE World Congress on Computational Intelligence* [127] [128].

En las siguientes subsecciones se presentan las técnicas de aprendizaje automático basadas en RNAs para pronóstico de series temporales, centrándose en tres aspectos: la formalización de los problemas de predicción de un solo paso como tareas de aprendizaje supervisado, la discusión de técnicas de aprendizaje locales como una herramienta eficaz para tratar los datos temporales y el papel que juega en la estrategia de predicción cuando se pasa de un solo paso a múltiples pasos usando técnicas dentro del *softcomputing* [129] [130].

### 2.5. Redes Neuronales Artificiales

Los RNAs estudiadas en esta tesis son los perceptrones multicapa (MLP), principalmente las redes neuronales con retardo de tiempo (TDNN). Se sabe bien que las redes neuronales artificiales (RNA) son aproximadores universales de funciones, y que no requieren un conocimiento a priori sobre el proceso en cuestión. Las RNAs también son bien conocidas por su capacidad para modelar sistemas no lineales.

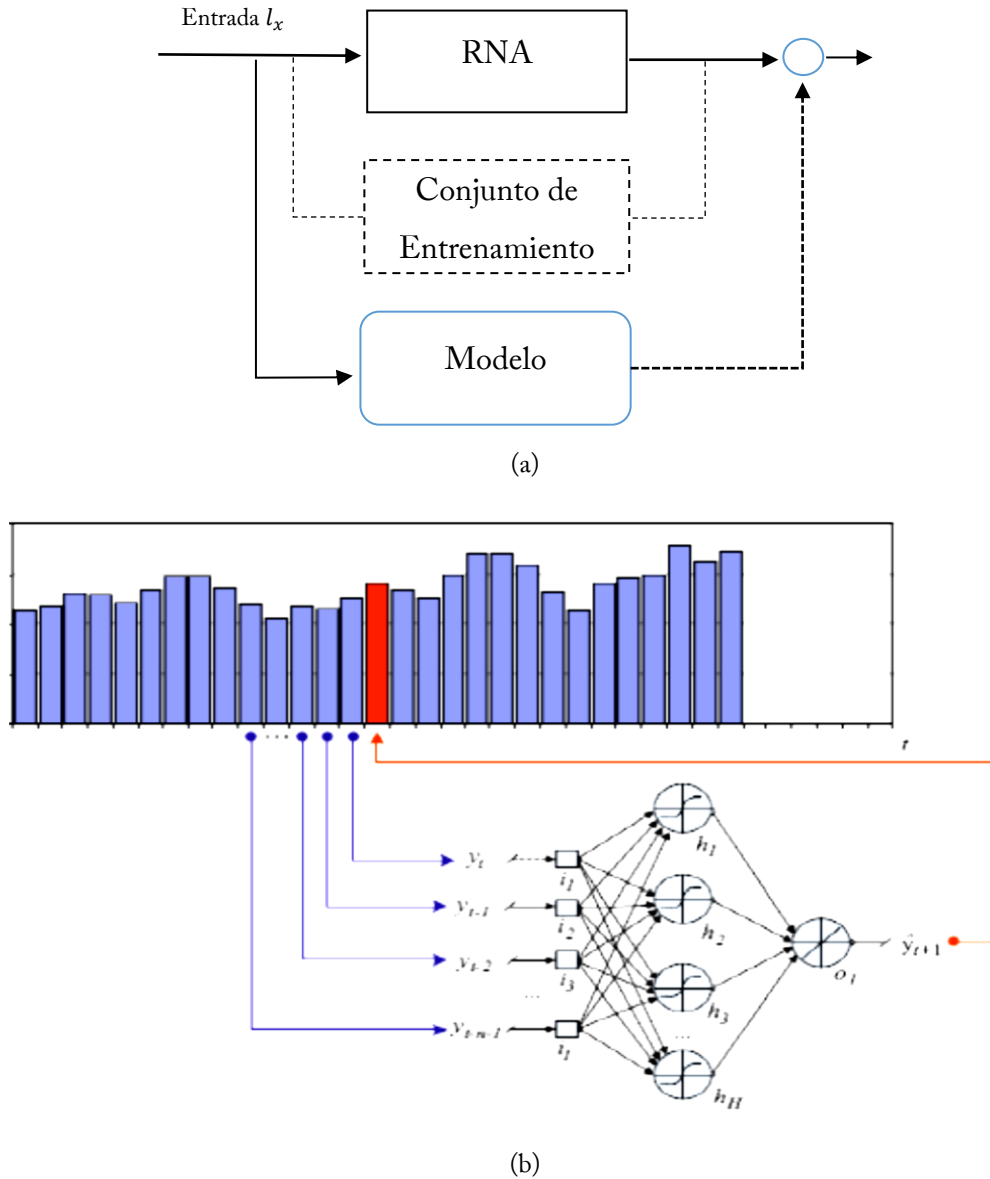
Estas propiedades son exactamente por las que son usadas en esta tesis en la modelización de series temporales no lineales con naturaleza desconocida o muy compleja, es decir, son muy atractivas como herramientas para predicciones de series temporales. La aplicación de las RNAs a la predicción con series temporales no es nueva. Desde hace décadas existen numerosos trabajos al

respecto, siendo probablemente los más conocidos los de Werbos [116] [131], Lapedes [117], Weigend [132]. Desde la estadística, Box y Jenkins [5] desarrollaron la metodología de los modelos autorregresivos integrados de promedios móviles (ARIMA) para ajustar una clase de modelos lineales para series temporales.

Posteriormente surgieron versiones robustas de modelos ARIMA y de series temporales no lineales [133] tendientes a resolver los problemas que introducen la presencia de valores aberrantes o extremos en los datos. Más recientemente, las RNAs han sido consideradas como una alternativa para modelar series temporales no lineales. Los modelos de RNAs se ajustan tradicionalmente por mínimos cuadrados y por lo tanto carecen de robustez en presencia de valores extremos o aberrantes (*outliers*). Como algunos de los procedimientos que tratan con RNAs surgen como una generalización natural de los modelos estadísticos lineales AR y ARMA al caso no lineal NAR y NARMA, los procedimientos para ajustar las RNAs suelen estar relacionados con los procedimientos empleados para modelar series temporales robustas [134].

En su forma más simple, una RNA FFNN relaciona una respuesta o variable de salida (output) con  $I_x$  predictores o variables de entrada (input) mediante una relación no-lineal representada como una transformación no-lineal aplicada a una combinación lineal de sus entradas o inputs [135]. En el contexto de las series temporales, es común relacionar una serie dada con su pasado (lags o retardos) mediante un proceso auto-regresivo lineal, que puede ser extendido a una forma no-lineal (modelo NAR).

La especificación de la arquitectura de la red implica, la definición de las variables de entrada como ser el número  $p$  de retrasos u orden del modelo NAR, el número de nodos en la capa intermedia y la transformación no-lineal aplicada en la capa intermedia llamada también función de activación. La activación de la salida, usualmente empleada en series temporales, es la función identidad.



**Fig. 2-7.** Red neuronal artificial; a) modelo de predicción; b) RNA en adelante basada en modelo NAR para pronosticar series de tiempo.

Una vez que la arquitectura de la red está completamente definida, el objetivo consiste en estimar las ponderaciones a partir de los datos. Una generalización natural del modelo lineal AR(p) al caso no lineal podría ser el siguiente modelo autorregresivo no lineal (NAR):

$$x_t = f(x_{t-1}, x_{t-1}, \dots, x_{t-k}) + e(t) \quad (2-4)$$

donde  $f(x)$  es una función de suavizado monótona de tipo *sigmoide* y la función de activación de la salida es supuesta la identidad. Los  $w_{ij}$  representan los parámetros de las conexiones entre las

entradas y las neuronas de la capa oculta,  $W_j$  representa los parámetros de las conexiones entre las neuronas ocultas y la salida y  $n$  es el número de neuronas en la capa oculta. Los parámetros  $W_j$  y  $w_{ij}$  son estimados a partir de una muestra  $x_t^0, x_{t-1}^0, \dots, x_{t-p}^0$  permitiendo obtener una estimación  $\hat{f}$  de  $f$ . Las estimaciones se obtienen minimizando la suma de cuadrados residuales  $\sum (x_t - \hat{x}_t)^2$ , la que al ser no lineal en sus parámetros se puede minimizar mediante un procedimiento numérico como el del gradiente descendente, conocido también como retropropagación (BP) [136] o por un método de segundo orden [137].

### 2.6. Métodos de Aprendizaje

El aprendizaje es un proceso fundamental de las RNAs y de forma general, consiste en el ajuste de todos los parámetros de la red, en base a la actividad para la que se quiera utilizar dicha red. Partiendo de un conjunto de pesos aleatorios, el aprendizaje busca un conjunto de pesos que permita a la RNA, desarrollar una tarea determinada. El aprendizaje es un proceso iterativo, en el que la red, va refinando sus parámetros, para alcanzar el objetivo de la aplicación concreta que se intenta abordar. Las redes perceptrón multicapa (MLP) utilizan una función de error que mide su rendimiento actual, en función de sus pesos. El aprendizaje se convierte en un proceso de búsqueda de aquellos pesos que hagan mínima dicha función [138], [139].

El aprendizaje no supervisado es un método donde un modelo es ajustado a las observaciones. Se distingue del aprendizaje supervisado dado que no hay un conocimiento a priori. En el aprendizaje no supervisado se hace un tratamiento de los datos de entrada como un conjunto de variables aleatorias, siendo construido un modelo de densidad para el conjunto de datos. El aprendizaje no supervisado puede ser usado en conjunto con la Inferencia bayesiana para producir probabilidades condicionales (es decir, aprendizaje supervisado) para cualquiera de las variables aleatorias dadas [140]. En cambio, el aprendizaje supervisado normalmente funciona mucho mejor cuando los datos iniciales son primero traducidos en un código factorial. El aprendizaje no supervisado también es útil para la compresión de datos: fundamentalmente, todos los algoritmos de compresión dependen tanto explícita como implícitamente de una distribución de probabilidad sobre un conjunto de entrada.



Por otro lado, el aprendizaje supervisado consiste en el modelado sobre la base de un conjunto finito de observaciones, la relación entre un conjunto de variables de entrada y una o más variables de salida, que se consideran algo dependiente de las entradas [105]. Una vez que un modelo se ajusta y está disponible, se puede utilizar para la predicción de un solo paso. En la predicción a un solo paso, los  $n$  valores anteriores de la serie temporal están disponibles y el problema de la predicción se puede convertir en un problema de regresión. El enfoque general para modelar un fenómeno de entrada/salida, con un escalar de salida y una entrada vectorial, se basa en la disponibilidad de una colección de pares observados que son típicamente el conjunto de entrenamiento [141]. A continuación, se describen ventajas y desventajas del algoritmo de aprendizaje supervisado Levenberg-Marquardt para entrenar a las redes neuronales.

### 2.6.1. El algoritmo Levenberg-Marquardt

El algoritmo Levenberg-Marquardt [142] resuelve los problemas existentes tanto en el método de descenso de gradiente como en el método de Gauss-Newton para el entrenamiento de las redes neuronales[143], por la combinación de estos dos algoritmos. Es considerado como uno de los algoritmos de entrenamiento más eficaces [144] [145]. Sin embargo, el algoritmo de Levenberg-Marquardt tiene sus defectos. Un problema es que la inversión de la matriz Hessiana del método de descenso más rápido debe ser calculada cada vez, por la actualización de los pesos y puede haber varias actualizaciones en cada iteración. Para el entrenamiento de redes de tamaño pequeño, el cálculo es eficiente, pero para redes de gran tamaño, como problemas de reconocimiento de imágenes, este cálculo de inversión puede ser un desastre y la velocidad adquirida por aproximación de segundo orden puede ser totalmente ineficaz. En ese caso, el algoritmo de Levenberg-Marquardt puede ser incluso más lento que el algoritmo de máxima pendiente [332]. Otro problema es que la matriz jacobiana tiene que ser almacenada para el cálculo, y su tamaño es  $P \times M \times N$ , donde  $P$  es el número de patrones,  $M$  es el número de salidas, y  $N$  es el número de pesos. Para los patrones de entrenamiento de gran tamaño, el costo de memoria para el almacenamiento de matriz jacobiana puede ser demasiado grande para ser práctico. Además, el algoritmo de Levenberg-Marquardt fue implementado sólo para redes perceptrón multicapa [333]. A pesar de que todavía hay algunos problemas no resueltos en el entrenamiento de este algoritmo

para redes y patrones de pequeños y medianos tamaños, el algoritmo de Levenberg-Marquardt es notablemente eficiente y altamente recomendado para el entrenamiento de redes neuronales. En comparación con los algoritmos de gradiente descendiente, el algoritmo de Levenberg-Marquardt requiere un mayor esfuerzo computacional, pero su convergencia es más rápida. Aquí se toma este algoritmo para el entrenamiento de los filtros predictores propuestos en esta tesis.

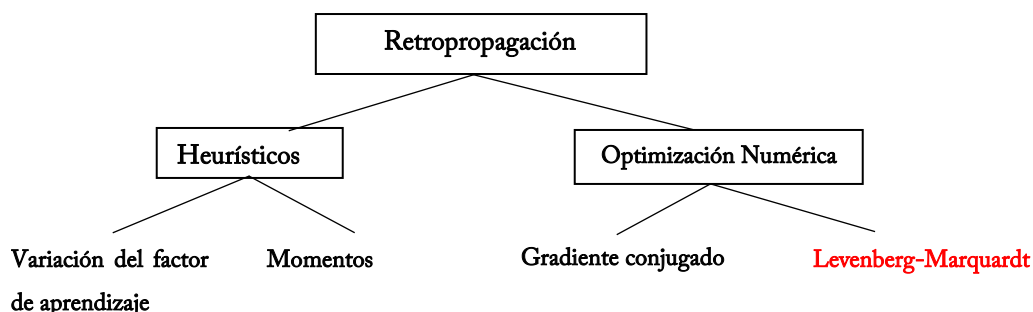


Fig. 2-8. Clasificación del algoritmo de retropropagación BP y el método Levenberg-Marquardt

### 2.7. Criterios de medidas de exactitud y errores en el pronóstico

Es importante evaluar la exactitud del pronóstico utilizando las predicciones originales [146]. Es decir, que no es válido evaluar cuan bien un modelo se ajusta a los datos históricos; la exactitud de las predicciones sólo se puede determinar teniendo en cuenta lo bien que un modelo predice en nuevos datos que no fueron utilizados en el montaje del modelo [147].

Al elegir modelos, es común el uso de una porción de los datos disponibles para las pruebas, y utilizar el resto de los datos para ajustar el modelo. Los datos de las pruebas se pueden utilizar para medir qué tan bien el modelo es probable que pronosticar nuevos datos. El tema de la medición de la exactitud de las predicciones con diferentes métodos ha sido el tema de mucha atención en estos últimos años [148]. Una discusión más a fondo es dada por [11]. También hay otros métodos disponibles para evaluar predicciones de intervalo [149] y de densidad [150] que solo son nombrados en esta tesis.

## 2.8. Métricas en el pronóstico

Hay algunas medidas de precisión de uso común cuya escala depende de la escala de los datos. Estos son útiles cuando se comparan diferentes métodos en el mismo conjunto de datos, pero no debe ser utilizado, por ejemplo, al comparar a través de conjuntos de datos que tienen diferentes escalas. Las medidas dependientes de la escala más comúnmente utilizadas se basan en el error absoluto o al cuadrado [151]. Una de las cuestiones más difíciles es la elección de las medidas más apropiadas fuera de estos grupos mencionados anteriormente, error absoluto o al cuadrado. Debido al hecho de que cada medida de error tiene la desventaja que pueden conducir a la evaluación inexacta de los resultados de predicción, es imposible elegir sólo una medida. Las medidas se dividen en grupos de acuerdo con el método de calcular un valor de error para cierto tiempo  $t$  [152].

Los resultados computacionales que se muestran en esta tesis usan diferentes aproximaciones de orden y períodos de tiempo para las métricas en los horizontes de pronósticos de 6-9-15-18, y el rendimiento de los modelos propuestos durante la calibración y validación, se miden a través del error porcentual absoluto medio simétrico (SMAPE) propuesto en la mayor parte de la evaluación métrica [153] [154], definido por

$$SMAPE_s = \frac{1}{n} \sum_{t=1}^n \frac{|X_t - F_t|}{(|X_t| + |F_t|)/2} \cdot 100 \quad (2-5)$$

donde  $t$  es el tiempo de observación,  $n$  es el tamaño del conjunto de Test o prueba,  $s$  es cada serie temporal,  $X_t$  y  $F_t$  son los valores de las series temporales actual y pronosticada en el tiempo  $t$ , respectivamente. El índice SMAPE de cada serie  $s$  calcula el error absoluto simétrico en porcentaje entre el  $X_t$  real y su valor de pronóstico correspondiente  $F_t$ , en todas las observaciones  $t$  del conjunto de prueba de tamaño  $n$  para cada serie temporal  $s$ . La raíz del error cuadrático medio (RMSE) se define de la siguiente manera para cada serie  $s$ ,

$$RMSE_s = \sqrt{\frac{1}{n} \sum_{t=1}^n (X_t - F_t)^2} \quad (2-6)$$

El valor porcentual que arroja el índice SMAPE se autolimita a una razón del 200%, es decir que el límite inferior y superior está comprendido entre 0% and 200%. Hay muchas maneras de medir la exactitud de los pronósticos, y las respuestas a estas preguntas depende de lo que se está

pronosticando, la medida de precisión que se utiliza, y qué conjunto de datos se utiliza para calcular la medida de precisión [155] [156] [157] [158].

## 2.9. Filtros Predictores basados en redes neuronales

El verdadero poder de las redes neuronales se debe a su capacidad de imitar algunos de los procesos creativos del cerebro [159]. Las RNAs son una de las aplicaciones inteligentes dominante en la predicción de series temporales. Con el éxito del algoritmo de retropropagación de [160], nuevas líneas de investigación en redes neuronales han crecido rápidamente y adaptado con éxito para el problema de la predicción de series temporales. Algunos de esos resultados han demostrado que las redes neuronales son capaces de superar a los métodos estadísticos de predicción como el análisis de regresión y predicción de Box-Jenkins [5]. Trabajos como [27] [28] muestran los avances en los últimos años usando redes neuronales para pronóstico de series temporales con énfasis en los distintos métodos usados en la literatura mostrando evidencia de la buena performance de las RNAs.

La capacidad de las redes neuronales para capturar las relaciones no lineales en los datos, las hacen uno de los mejores candidatos ideales para la predicción de series temporales no lineales [29]. La mayor parte de estas redes se basan en el aprendizaje supervisado y el perceptrón multicapa (MLP) [161]. La arquitectura general de un RNA para ser usado como predictor se da en el siguiente diagrama.

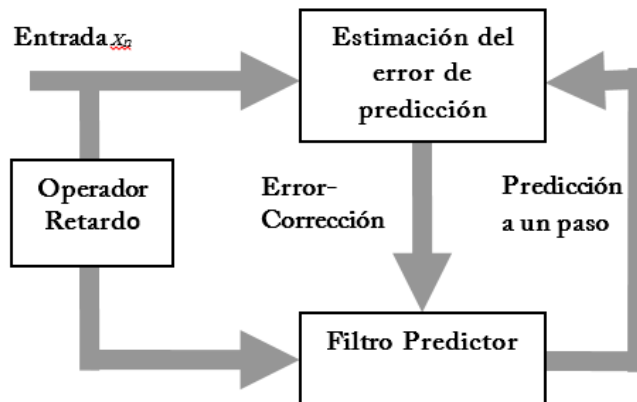


Fig. 2-9. Estructura del filtro predictor basado en RNA para predicción

Dados los valores pasados uniformemente espaciados  $x(n-1), x(n-2), \dots, x(n-m)$ , con un intervalo de muestreo  $T$  y un orden de predicción  $m$ . Se desea predecir el valor presente  $x(n)$ . Para órdenes de predicción superior, se usan las salidas del filtro predictor  $x(n+1), x(n+2), \dots, x(n+m)$ . Por lo tanto, el objetivo se reduce a diseñar un predictor que aproxime a su desempeño a un conjunto de datos, y sea capaz de predecir datos futuros usando su propia salida.

La predicción clásica puede ser formulada como sigue. Dados los valores pasados de un proceso uniformemente espaciados  $x(n-1), x(n-2), \dots, x(n-m)$ , con un intervalo de muestreo  $T$  y un orden de predicción  $m$ , se desea predecir el valor presente  $x(n)$  para tales procesos. Es decir, los valores se ajustan al filtro con la predicción de un paso, y el vector de entrada se obtiene usando como entrada los valores que el mismo filtro genera para la selección del modelo [162].

### 2.9.1. Predictor basado en redes neuronales ARMA lineal

Uno de los objetivos de esta tesis se encuentra en los modelos, el proceso de ajuste y estructura del filtro predictor basado en *rugosidad* tomando como base el trabajo presentado por Pucheta *et. al.* [15], que emplea la regla del gradiente descendiente y considera la dependencia estocástica el corto y largo alcance de los valores pasados de la serie temporal para ajustar en cada época, el número de patrones, el número de iteraciones y la longitud del vector de entrada.

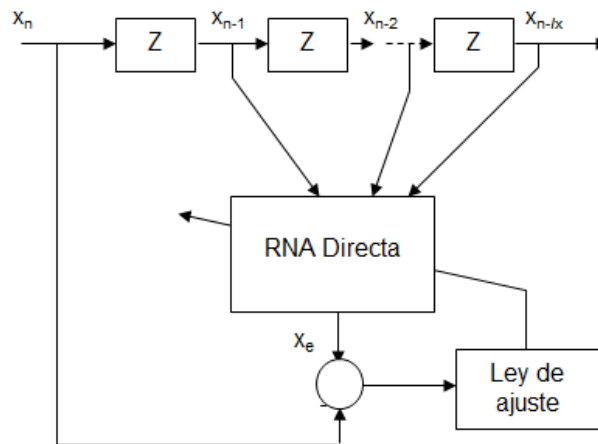


Fig. 2-10. Estructura del filtro predictor

El operador retardo  $Z^{-1}$ , sobre el arreglo de la serie de entrada es:

$$x_e(n+1) = F_p(Z^{-1}I(\{x_n\})) \quad (2-7)$$

donde,  $F_p$  es el operador del filtro predictor, y  $x_e(n+1)$  su salida en  $n$  con  $I \in \mathfrak{R}^k \times \mathfrak{R}^k$ . El filtro no causal está basado en el esquema ARMA. Sin embargo, la función será lineal o no lineal, usando RNAs.

El sistema predictor va a ser implementado usando un modelo auto-regresivo lineal adaptable usando RNAs. Con el objeto de realizar una predicción confiable se debe de analizar los datos disponibles, esto para identificar algún patrón que pueda utilizarse en la descripción de los mismos, en este caso el parámetro de Hurst medido en la serie.

El patrón obtenido se extrapola, es decir, se amplía hacia el futuro y con esto obtener un pronóstico, aclarando que el patrón identificado anteriormente no sufra algún cambio en el futuro. Si el patrón que se identificó sufre algún cambio, entonces la técnica de predicción obtenida no generara un buen pronóstico, sin embargo, se puede realizar cambios apropiados en el patrón de modo tal que se evite la inexactitud en el pronóstico.

Las mejoras del pronóstico se deben a; mejoras de los modelos hechos [162]; mejoras de las observaciones o mejoras de la asimilación de datos que se detallan en los siguientes capítulos como aportes de esta tesis que dieron lugar a varias publicaciones que se describe a continuación.

El objetivo de esta subsección es mostrar las RNAs usando el modelo ARMA, esto es, dada la serie  $\{x_n\}$  y la serie  $\{\{x_n\}, \{x_e\}\}$  donde  $\{x_e\}$  tiene distintos valores según el horizonte de pronóstico [43].

El método propuesto es, dado el vector de entrada  $x_i \in \mathfrak{R}^k$  aplicando  $Z^{-1} \cdot I$  a  $\{x_n\}$ , entonces, se define el error de predicción como

$$e_n = x_n - f_n\left(\left[x_{n-1}, x_{n-2}, \dots, x_{n-k}\right]^T, r\right) \quad (2-8)$$

donde  $r$  son los parámetros de la red neuronal,  $f_n$  con función de activación *sigmoidal*, y se ajustan mediante el algoritmo Levenberg-Marquardt. Para cada  $x_n$  se define una secuencia de entrada, para implementar la iteración por lotes. Aquí se propone hacer variable a la cantidad de iteraciones  $i_t$  y la longitud de cada lote  $l_k$ . A partir de la secuencia  $\{x_n\}$  en un instante  $n$ , para obtener la mejor predicción  $\{x_e\}$  de los valores a pronosticar a futuro. Se define un vector de entrada de tamaño  $l_k$ , obtenido mediante el operador retardo,  $Z^{-1}$ , de la secuencia  $\{x_n\}$ . El filtro siempre tiene como entradas valores de la secuencia  $\{x_n\}$  que no incluyen al valor  $x_n$ . Entonces se define el error:

$$e_n = x_n - x_e \quad (2-9)$$

el cual es usado para la regla de ajuste de los coeficientes del filtro. Se realiza un ajuste en lotes, mediante el algoritmo Levenberg-Marquardt. El ajuste cambia en línea, modificando el número de patrones, el número de iteraciones, y la longitud del vector de entrada en función del parámetro de Hurst  $H$ .

Dado que la serie  $\{x_n\}$ , y la serie aumentada  $\{\{x_n\}, \{x_e\}\}$  deberían presentar la misma rugosidad, se propone cambiar los parámetros del algoritmo según la diferencia de sus  $H$  asociados, debido a que  $H$  es la medida de rugosidad, la serie aumentada debería tener la misma rugosidad que la serie dato.

Sean  $N_p$  pares entrada salida  $(x_i, y_i)$  con  $i = 1, 2, \dots, N_p$  donde  $x_i \in \mathfrak{R}^k$  e  $y \in \mathfrak{R}$  son las entradas y las salidas respectivamente.

Éstas se definen como  $x_i = Z^{(-1)} \cdot I(x_i)$ ,  $y_i = x_i$ . El número de pares  $N_p$  está definido en términos de la dimensión del vector de entrada  $I_x$ , como  $I_x \leq N_p \leq 4I_x$ . Además, el número de iteraciones  $i_t$  está dado por  $I_x \leq i_t \leq 4I_x$ .

La estrategia de ajuste de los parámetros del filtro predictor está implementada en lotes (BATCH). Cada lote tiene  $N_p$  pares de entrada-salida, o patrones. Con cada lote de  $N_p$  patrones el algoritmo realiza  $i_t$  iteraciones para ajustar los parámetros  $r$  del filtro predictor implementado mediante una red neuronal.

### 2.9.2. Ajuste heurístico propuesto

Se propone modificar el par  $(i_t, N_p)$  según la dependencia estocástica de corto o largo alcance asociado a la serie temporal  $\{x_n\}$  suponiendo que es una realización de un *mBf*. El  $H$  se estima mediante el método de wavelets [55]. Se implementa el siguiente ajuste heurístico

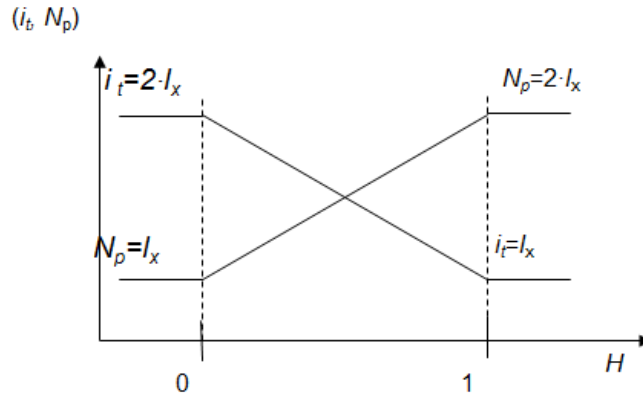


Fig. 2-11. Ajuste heurístico de  $(i_t, N_p)$  en términos de  $H$  después de cada época.

### 2.9.3. Evaluación para ajuste de estructura

Finalmente, terminada la predicción de la serie, se define el error como de época como:

$$e_H = H(\{x_n\}) - H(\{\{x_n\}, \{x_c\}\}) \quad (2-10)$$

Si  $e_H > \theta$  se ajusta el vector de entrada  $l_x$  mediante el conocido esquema de control proporcional-integrador-derivativo (PID) citado en [15],

$$l_x = l_x + \text{round}(K_p \cdot e_H + K_D \cdot \dot{e}_H) \quad (2-11)$$

donde se fija que el umbral  $\theta$  sea del 1%. Las condiciones iniciales que implementa el filtro se muestran en la Tabla 2-1.

Tabla 2-1. Condiciones iniciales de la red neuronal ARMA lineal

Variable	Condición inicial
$l_x$	12
$i_t$	$5 \cdot l_x$
$H$	0.5
$\eta$	0.1

El objetivo es pronosticar series que presenten el mismo  $H$  que las series de dato. El error  $e_H$  se calcula comparando el  $H$  asociado a la serie completa ( $x_n$  seguida de  $x_c$ ) y a la serie que es dato  $x_n$ . Se cambia la longitud  $l_x$ , que significa cambiar la dimensión de  $I$  en la implementación del operador retardo  $Z^{-1} I$ . Finalmente, después de cada paso, el número de entradas del filtro no-lineal son ajustadas, de acuerdo al siguiente criterio heurístico. Luego que se completa el proceso de aprendizaje, ambas secuencias —  $\{x_n\}$  y  $\{\{x_n\}, \{x_c\}\}$  — deberían tener el mismo parámetro  $H$ . Si



el error entre  $H(\{x_n\})$  y  $H(\{\{x_n\}, \{x_c\}\})$  es mayor que el umbral del parámetro  $\theta$ , el valor de  $I_x$  es incrementado o decrementado en  $I_x \pm 1$ .

## 2.10. Filtro predictor basado en algoritmos genéticos

Los *Algoritmos Evolutivos* constituyen un método práctico de optimización de una función objetivo. La tarea de optimización consiste en, dados los parámetros  $(x_1, \dots, x_n)$  de la función objetivo  $f$ , encontrar los valores  $(\bar{x}_1, \dots, \bar{x}_n)$  tales que hacen mínimo (o máximo, según sea el objetivo) el valor de la función en dicho punto,  $f(\bar{x}_1, \dots, \bar{x}_n)$ . Los Algoritmos Evolutivos tratan de imitar los procesos relacionados con la evolución natural, y principalmente la selección y la generación de nuevos genotipos para los descendientes de los individuos de la población. Como métodos para afrontar la predicción de series temporales se han usado principalmente regresiones, sistemas dinámicos no lineales, redes neuronales, y en menor medida, algoritmos evolutivos.

El objetivo de esta sección es presentar un algoritmo propuesto [163] [164] para la predicción de series temporales basado en métodos kernel [143], específicamente en los *Procesos Gaussianos* (GP) [165]. Los GP son filtros no paramétricos, por lo tanto, el único objetivo antes de entrenar el modelo de filtro autorregresivo es descubrir un conjunto óptimo de los retardos de tiempo (lags) y una función de covarianza. Teniendo en cuenta el hecho de que cada función de covarianza puede requerir un conjunto diferente de entradas de tiempo (lags), ambas tareas de optimización se resuelven simultáneamente mediante un algoritmo genético [166]. Estos algoritmos se basan en un método para seleccionar las entradas y la función de covarianza de un filtro paramétrico realizado como un proceso gaussiano. Ambos problemas, el encontrar una función de covarianza y la determinación de las entradas del modelo del filtro son discretos, ya que la función de covarianza se elige de un conjunto discreto de posibles funciones y las entradas están definidas por tiempo discreto se retrasa. Para hacer frente a estos dos problemas de optimización se utiliza un algoritmo de ordenación genética, no dominada llamada NSGA-II [167]. El NSGA-II es un algoritmo genético multiobjetivo que mostró buenos resultados para varios problemas de optimización [168]. Se parte de un conjunto elegido al azar de individuos cuyos valores de ajustes (*fitness*) son evaluados. A continuación, se crea la siguiente generación de individuos a través de operaciones de cruce y

mutación de los individuos con los mejores valores de fitness. Continuar con este procedimiento por varias generaciones conduce a los individuos con valores de fitness optimizados. En el trabajo [164] cada individuo representa un GP con una función de covarianza y ciertos desfases de tiempo. Para validar los resultados de este algoritmo, más de un inicio de múltiples puede ser ejecutado. Esto significa que el algoritmo se inicia varias veces a partir de una población aleatoria.

Para aplicar el NSGA-II en este caso especial los valores de fitness, que sirven como criterios de optimización, se deben definir en primer lugar. Así dos objetivos deben ser tenidos en cuenta, por un lado, el modelo de filtro tiene que ser lo más preciso posible y por otra parte el número de entradas del proceso gaussiano debe ser tan pequeño como sea posible para evitar sobreajuste (*overfitting*).

### 2.10.1. Proceso de ajuste

Existen varios parámetros que tienen que ser definidos cuando se ejecuta el NSGA-II. Además de los parámetros internos del algoritmo genético hay varios parámetros específicos del problema que tienen influencia en las soluciones encontradas.

En el trabajo propuesto [164], la población de la NSGA-II consta de 200 individuos que son evaluados sobre 200 generaciones. Por lo tanto, los valores de fitness de 40.000 individuos se calculan en cada arranque múltiple. Después de la evaluación de una generación, nuevos individuos se crean a través de la mutación con una probabilidad del 90% y a través de operaciones de cruce con una probabilidad del 10%.

Cada individuo debe tener entre 1 y 20 entradas a partir del pasado y ninguna de estas entradas puede tener un intervalo de tiempo menor que 1 o mayor de 30. En otras palabras  $1 \leq i \leq 20$ ,  $\min(m_1, \dots, m_i) \geq 1$  y  $\max(m_1, \dots, m_i) \leq 30$  para

$$y^*(t) = GP(x(t - m_1 T), \dots, x(t - m_i T)) \quad (2-12)$$

donde  $GP(\cdot)$  es el proceso de Gauss,  $y^*(t)$  es el valor predicho en el tiempo  $t$  y  $T$  es el periodo de muestreo. Para esta configuración hay más de  $6 \times 10^8$  posibles combinaciones de entradas a partir del pasado. Como se muestra en la siguiente subsección, sin embargo, las evaluaciones de cada 40.000 MultiStart son suficientes para obtener una buena aproximación de la frente de Pareto. Los parámetros mencionados describen el proceso de optimización que busca para los conjuntos

óptimos de entradas, en combinación con una función de covarianza. De esta manera la función de covarianza se toma de un conjunto de posibles funciones que cada función tiene propiedades especiales que podría conducir a filtros precisa de ciertos insumos. Las funciones de covarianza utilizadas son:

- Función de covarianza lineal con detección de relevancia automática (ARD);
- Cuadrado función de covarianza exponencial con ARD;
- La función neural red covarianza;
- Isotrópica función de covarianza cuadrática racional;
- Función de covarianza Matern con  $\nu = 3/2$  y  $\nu = 5/2$  [169];
- La suma de una función de covarianza lineal con ARD y una función de covarianza de ruido blanco;
- La suma de una covarianza exponencial cuadrado función con ARD y una covarianza de ruido blanco función;
- La suma de una función de covarianza de la red neural y una función de covarianza de ruido blanco;
- La suma de una cuadrática racional isotrópica función de covarianza y una covarianza de ruido blanco función.

De este modo ARD significa determinación de relevancia automática [170]. Una descripción detallada de estas funciones de covarianza se puede encontrar en [165]. En total siete multistarts se llevan a cabo con la configuración descrita para validar las propiedades de convergencia de la NSGA-II.

### 2.10.2. Entrenamiento y predicción del modelo basado en Procesos Gaussianos

Cada individuo que se evalúa por el NSGA-II consta de ciertos retardos de tiempo y una función de covarianza. Por lo tanto, las entradas de entrenamiento del GP tienen que ser construido a partir de los datos proporcionados de acuerdo con los lapsos de tiempo de los individuos antes de afinar el modelo de filtro.

En el caso de los GP, el modelo está sintonizado variando los hiperparámetros de la función de covarianza. Dependiendo de la función de covarianza, hay varios hiperparámetros que necesitan

ser ajustados para adaptarse a los datos de entrenamiento. En otras palabras, uno está interesado en la búsqueda de un máximo del logaritmo de la verosimilitud. Sin entrar en detalles el marco presentado en [165], se utiliza para optimizar los hiperparámetros en este trabajo. Una vez que se encuentran los hiperparámetros, el proceso de entrenamiento ha terminado.

Para evaluar un modelo GP uno tiene que calcular la matriz de covarianza  $K$  y su inversa  $K^{-1}$ . Para  $n$  dado puntos de entrenamiento,  $K$  tiene el tamaño  $(n \times n)$ . Sus entradas son las covarianzas por pares de las entradas de entrenamiento que hace que  $K$  sea una matriz simétrica. Suponiendo que las variables tienen una distribución de Gauss conjunta con media cero, la predicción media de una entrada desconocida  $f^*$  viene dada por

$$f^* = K(X^*, X)K(X, X)^{-1}f \quad (2-13)$$

donde  $X^*$  es la entrada desconocida,  $X$  son las entradas de entrenamiento y  $f$  son los productos del entrenamiento. Si la media de los datos no es cero, se puede transformar sencillamente para adaptarse a las condiciones.

### 2.11. Comentarios finales

En este Capítulo, se han presentado dos métodos, uno basado en RNAs usando modelos ARMA lineal y otro basado en algoritmos genéticos para la predicción de series temporales. El algoritmo basado en RNAs usa una ley heurística para sintonizar sus parámetros. La regla de aprendizaje propuesto para ajustar los pesos sinápticos de la RNA se basa en el método de Levenberg-Marquardt. Además, en función de la dependencia estocástica de corto o largo alcance de la serie medida por el parámetro de Hurst  $H$  para el ajuste en línea del algoritmo, se propuso una ley heurística adaptativa para actualizar la topología de la RNA en cada época, el número de entrada  $I_s$ , el número de patrones  $N_p$  y el número de iteraciones  $i_t$ . En cambio, el algoritmo genético usa el algoritmo NSA-II para sintonizar los parámetros. En el siguiente Capítulo estos enfoques se mejoraron y se proponen nuevos métodos para su estudio.

# Capítulo 3

*“El futuro tiene muchos nombres. Para los débiles es lo inalcanzable. Para los temerosos, lo desconocido. Para los valientes es la oportunidad.”*

Víctor Hugo

## 3. Modelos basados en suavización de la serie

En el capítulo anterior se propusieron modelos de pronósticos no lineales basados en RNA ARMA y en Procesos Gaussianos (GP), que tienen en cuenta la rugosidad de la serie temporal para modelar los parámetros de los filtros y para el caso de las RNAs, una ley heurística para el sintonizado de los mismos. Además, se mostraron que estos esquemas son una alternativa atractiva para mejorar el desempeño de los sistemas predictores. En este Capítulo se derivan tres nuevos métodos para el diseño de filtros predictores, que además de tener en cuenta la rugosidad se suaviza la serie temporal de datos, lo cual es crítico cuando la misma sufre cambios abruptos y se emplea para hacer pronóstico. El nuevo análisis que se presenta permite obtener nuevos criterios de diseño con buenos rendimientos en términos de error de desempeño de predicción.

### 3.1. Introducción

Este capítulo presenta tres casos de estudio como aportes para hacer pronósticos, en el primer caso se presenta una metodología de pronóstico clásica en modelo de redes neuronales NAR basado en *rugosidad* llamado NN-Mod. En el segundo y tercer caso se procesa a la serie dato generando una serie más suave donde se facilita el proceso de ajuste del filtro basado en RNA, para finalmente comparar el desempeño y resultados en el Capítulo 5. Se presentan tres nuevos métodos para el diseño de algoritmos basados en RNAs [33], uno basado en la rugosidad (NN-Mod.) [179] [34],

otro llamado Energía asociada a la serie (EAS) [35] [180] [181] y por último llamado Decimación o Subsampling (SUB) [36] [182] [183], para predecir series temporales de alta rugosidad como así también series de precipitación mensual de agua acumulada, donde se toma como referencia al filtro predictor basado en rugosidad (NN-Mod.), mejorado del propuesto en la Sección 2.9.1. Dados que existen una variedad de modelos basadas en RNAs para las predicciones, específicamente de lluvias [171] y tomando un conjunto de puntos de medición [172], otros métodos de pronóstico basados en inteligencia computacional, tal como [173] [174] [175] [176] [177] [178] se encuentran en la literatura como modelos para pronosticar series de lluvia.

### 3.2. Predictor NAR basado en rugosidad de la serie o NN-Mod.

El enfoque se basa en RNA usando el modelo NAR y la hipótesis de predicción es que la rugosidad de la serie dato debe coincidir con la rugosidad de la serie dato más la serie pronosticada [34]. En función de la rugosidad de la serie temporal, se propone una ley heurística para establecer el proceso de ajuste de la topología de la RNA, asumiendo que la serie temporal pronosticada tenga el mismo parámetro de Hurst que la serie dato original. La regla de ajuste utilizada en el proceso de aprendizaje se basa en el método Levenberg-Marquardt.

El sistema predictor es implementado usando un modelo auto-regresivo adaptativo no lineal. Las RNAs son usadas para el desarrollo de modelos no lineales, en el sentido que a menor error de predicción (en un sentido estadístico), mayor será la utilidad de la RNA como modelo de los procesos subyacentes responsables de generar los datos. El modelo predictor se diseña de tal manera que a partir de una determinada secuencia  $\{x_n\}$  en un tiempo  $n$  correspondiente a una serie temporal, se puede obtener la mejor predicción  $\{x_c\}$  para los siguientes valores de secuencia, según el horizonte de pronóstico. Por lo tanto, se propone un filtro predictor con un vector de entrada  $I_x$ , el cual es obtenido por la aplicación de un operador de retardo  $Z^{-1}$  a la secuencia  $\{x_n\}$  como se muestra en la Fig. 2-10.

Entonces, la salida del filtro puede generar  $x_c$  como el valor próximo que será igual al valor presente  $x_n$ . Así, el error de predicción en un tiempo  $k$ ,  $0 < k < n$ , puede ser evaluado como

$$e(k) = x_n(k) - x_c(k) \quad (3-1)$$

el cual es usado para regla de aprendizaje para ajuste de los coeficientes de la RNA. El parámetro de Hurst es usado durante el proceso de aprendizaje para modificar en tiempo real el número de patrones, el número de iteraciones y el número de entradas del filtro. El parámetro H sirve para tener en cuenta la rugosidad de la serie y para determinar su dependencia estocástica.

Para hacer predicción con este filtro basado en rugosidad se emplearon series temporales que han sido utilizadas comúnmente como benchmark muy difundidas en la literatura para sistemas caóticos, como son la ecuación de Mackey-Glass [184] [185] usada en el modelado de procesos meteorológicos. La definición de la característica estocástica que tiene relación con la rugosidad, la dimensionalidad fractal de un conjunto de datos [71], y el enfoque clásico del pronóstico de series temporales basado en inteligencia artificial [135].

### 3.2.1. Proceso de aprendizaje basado en rugosidad

Los coeficientes de las RNAs son ajustados mediante la regla Levenberg-Marquardt, la cual considera la dependencia estocástica de corto o largo alcance de la serie temporal medida por el parámetro de Hurst (H). El aprendizaje propuesto consiste en cambiar el número de patrones  $N_p$ , el vector de entrada  $I_x$  y el número de iteraciones  $i_t$  en función del parámetro de Hurst para cada serie temporal.

El proceso de aprendizaje se lleva a cabo utilizando un modelo de proceso por lotes. En este caso la actualización de los pesos de la RNA se realiza después de la presentación de todos los datos de entrenamiento, que forma una época. Los pares de los patrones de entrada-salida que se utilizan son

$$(x_i, y_i) \quad i=1,2,\dots,N_p \quad (3-2)$$

donde,  $x_i$  y  $y_i$  son el patrón de entrada y de salida correspondiente, respectivamente, y  $N_p$  es el número de patrones de entrada-salida presentados en cada época. Aquí, el vector de entrada se define como

$$X_i = Z^{-1} \cdot I(\{x_i\}) \quad (3-3)$$

y su correspondiente vector de salida como

$$Y_i = x_i \quad (3-4)$$

Por otra parte, el índice  $i$  está dentro de la gama de  $N_p$  propuesta por

$$I_x \leq N_p \leq 2 \cdot I_x \quad (3-5)$$

donde  $I_x$  es la dimensión del vector de entrada. Además, a través de cada época el número de iteraciones realizadas se da por

$$I_x \leq i_t \leq 2 \cdot I_x \quad (3-6)$$

El criterio propuesto para modificar el par  $(i_t, N_p)$  está dado por la dependencia estadística de la serie temporal  $\{x_n\}$ , suponiendo que es una realización del *mBf*. La dependencia es evaluada por el parámetro de Hurst  $H$ , que se calcula utilizando un método basado en wavelets [55]. Por lo tanto, se propone un ajuste heurístico para el par  $(i_t, N_p)$  en función de  $H$  de acuerdo con las funciones de pertenencia mostradas en la Fig. 3-1. Finalmente, después de cada época el número de entradas del filtro no lineal se sintoniza, es decir la longitud de línea de retardo con derivaciones, de acuerdo con el siguiente criterio heurístico. Una vez completado el proceso de entrenamiento, ambas secuencias -  $\{x_n\}$  y  $\{\{x_n\}, \{x_e\}\}$ , deben tener el mismo parámetro  $H$ . Si el error entre  $H(\{x_n\})$  y  $H(\{\{x_n\}, \{x_e\}\})$  es mayor que un parámetro de umbral  $\theta$ , entonces se incrementa el valor de  $I_x$  (o se disminuye), de acuerdo con  $I_x \pm 1$ . Explícitamente,

$$I_x = I_x + \text{sign}(\theta) \quad (3-7)$$

Esta dependencia es evaluada por el parámetro de Hurst, el cual es obtenido usando el método basado en wavelets. En la siguiente figura se muestra un ajuste heurístico para el par  $(i_t, N_p)$  en función de  $H$ , acorde a una función de pertenencia.

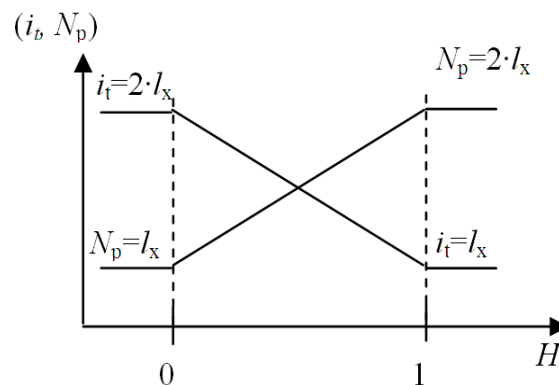


Fig. 3-1. Ajuste Heurístico de  $(i_t, N_p)$  en términos de  $H$  después de cada iteración.

Las condiciones iniciales del algoritmo de aprendizaje de la RNA se ajustan en función de la cantidad de entradas. Las condiciones iniciales usadas para algoritmo de aprendizaje de la serie



temporal a pronosticar, se basan en los resultados experimentales que se presentan en el Capítulo 5.

**Tabla 3-1.** Condiciones iniciales de la red neuronal NAR modificado por rugosidad

Variable	Condición inicial
$l_x$	7
$i_t$	105
$h_0$	9
H	0.5

### 3.3. Predictor basado en la Energía de la Serie

En esta sección se presenta un nuevo algoritmo para pronóstico de series temporales que ajusta los parámetros de un modelo auto-regresivo no lineal (NAR) basado en RNA en función de la energía de las series (EAS) [35] [180] [181]. El criterio de ajuste consiste en producir los valores de la serie temporal a partir del cálculo del área de la primitiva de la serie temporal pronosticada. Estos valores son aproximados por la RNA que genera una primitiva calculada como un área por el filtro predictor lineal, es decir se basa en la aproximación por primitivas de los valores de la serie temporal, donde se considera tomar la superficie resultante de la integración de los datos de la serie temporal  $\{x_n\}$ . A la primitiva la obtenemos considerando la derivada de cada valor de la serie temporal como

$$y_{t_k} = \int_{t_k}^{t_{k+1}} x_s ds \quad (3-8)$$

donde  $x_s$  es el valor correspondiente a la serie original, que requiere de una aproximación por ser una serie temporal. El área aproximada  $I$  se supone la primitiva periódica

$$I_{t_n} = \int_{t_n}^{t_{n+p}} x_s ds = x_t \Big|_{t_n}^{t_{n+p}}, n = 1, 2, \dots, N \quad (3-9)$$

Durante el proceso de aprendizaje, se calculan las primitivas y se utilizan como nuevas entradas a la RNA. La predicción intenta igualar el área de la serie pronosticada al área real de la primitiva pronosticada.

El criterio de ajuste en el pronóstico consiste en generar una serie de valores a partir del área de la serie temporal. A esos valores lo aproxima una RNA para generar una primitiva. Mediante

una RNA se genera la predicción del área debajo de la función correspondiente a los valores pronosticados. El valor de obtenido de la integral debe coincidir con el obtenido de evaluar a la función primitiva entre el intervalo de pronóstico. Teniendo la misma rugosidad que la serie dato, indicada por el parámetro de Hurst H. A continuación, se ejemplifica este método.

### 3.3.1. Método de aproximación

Puede introducirse una partición del intervalo de integración  $(t_n, t_{n+p})$  en un punto intermedio  $k$ , si se dispone de un valor intermedio del pronóstico de áreas,

$$I_{t_n} = I_{t_n} + I_{t_k} = \int_{t_n}^{t_k} y_t dt + \int_{t_k}^{t_{n+p}} y_t dt = Y_t|_{t_n}^{t_k} + Y_t|_{t_k}^{t_{n+p}}. \quad (3-10)$$

El filtro pronosticador generará un valor para  $I_t$  correspondiente al área bajo la curva pronosticada, que es la derivada de la primitiva.

$$y_{t_{n+1}} = Y_{t_{n+1}} - Y_{t_n} \quad (3-11)$$

Durante el proceso de aprendizaje podrán calcularse estas primitivas, y ser una entrada extra a la RNA, para que en el pronóstico se intente igualar al área de la serie pronosticada y a las primitivas de los valores de la serie.

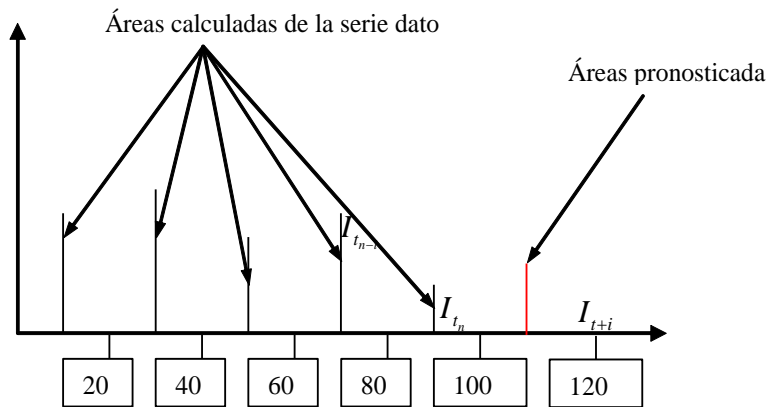


Fig. 3-2. Pronóstico del área de la serie predicha según el área de la serie dato, en  $i$  intervalos de tiempo.

Se pretende calcular las áreas y colocarlas en el centro del periodo tomado de manera que en la predicción quede en el centro o dentro del periodo formado por  $i$  intervalos de tiempo, de integración de la serie.

Por ejemplo, estas áreas se calculan con un regresor lineal con un periodo de 20 muestras,  $I_{t+1} = .w1 * I_t + w2 * I_{t-1} + w2 * I_{t-2} + w2 * I_{t-3} + \dots + .$

En la Fig. 3-3 se muestran dos series temporales MG en el que se representa la energía asociada a cada una, en conjunto con el parámetro H. Nótese que analizando la rugosidad de ambas series se observa una mayor rugosidad en la serie que se encuentra en la parte superior.

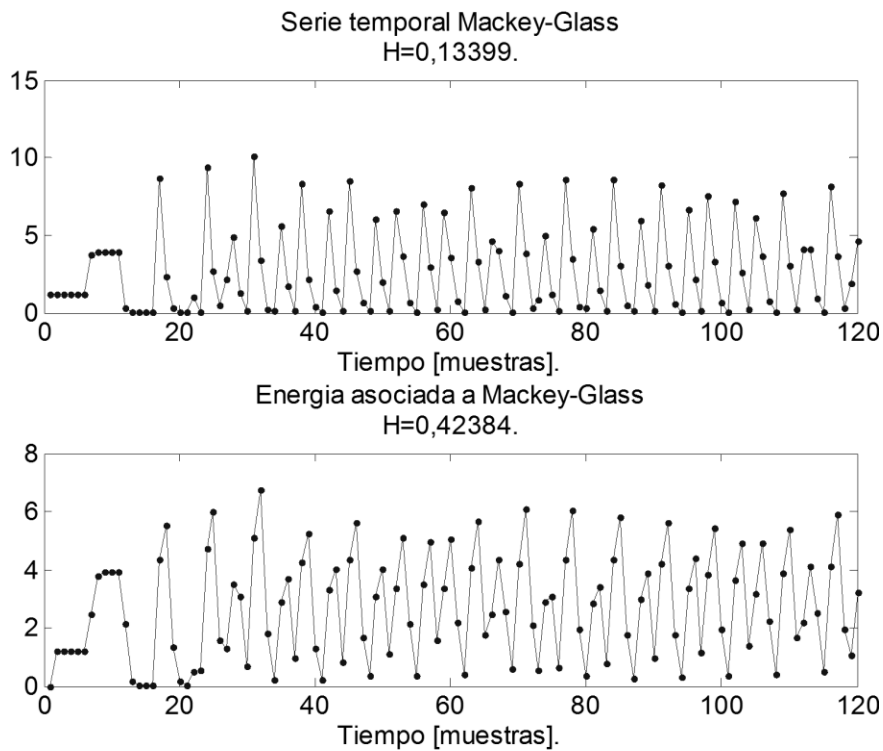


Fig. 3-3. Serie temporal de Mackey-Glass; arriba: la serie temporal original, abajo la energía asociada a la serie.

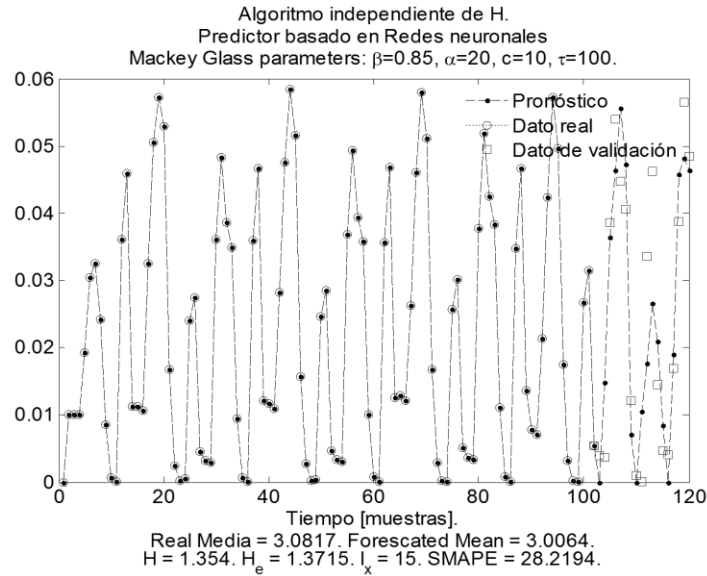


Fig. 3-4. Pronóstico obtenido de la serie temporal de Makey-Glass, donde el algoritmo de ajuste considera la energía asociada de la serie.

El área real es usada en dos instancias. En primer lugar, desde la serie temporal real un área es obtenida y es usada por el algoritmo predictivo. El algoritmo se describe en [180]. El parámetro  $H$  de esta serie es llamado  $H_A$ . Por otro lado, los datos pronosticados por el algoritmo tienen asociado un parámetro  $H$  llamado  $H_S$ . Finalmente, de acuerdo con el criterio heurístico, después de cada paso es ajustado el número de entradas del filtro no lineal. Cuando el proceso de aprendizaje se completa, se analizan las dos secuencias  $\{\int\{\{x_n\},\{x_e\}\}\}$  y  $\{\int\{x_n\}\}$ , que de acuerdo con la hipótesis deben tener el mismo  $H$ . Si el error entre  $H_A$  y  $H_S$  es mayor que un parámetro umbral  $\theta$ , el valor de  $l_x$  es entonces incrementado o decrementado de acuerdo a  $l_x \pm 1$ .

### 3.3.2. Configuración del modelo y proceso de aprendizaje

Las condiciones iniciales para el filtro y para el algoritmo de aprendizaje se muestran en la Tabla 3-2. Nótese que el primer número de neuronas ocultas y la iteración se establecen en función del número de entradas. Estas condiciones iniciales del algoritmo de aprendizaje se utilizan para pronosticar la primitiva de las series temporales, cuya longitud depende de la cantidad de datos en la serie.

Tabla 3-2. Condiciones Iniciales de los parámetros

Variable	Condición inicial
$l_x$	15
$i_t$	$l_x$
$h_0$	$l_x/2$
H	0.5

### 3.4. Método de Decimación o Submuestreo

Dado que las series de alta rugosidad son muy complejas de predecir, se propone una metodología para suavizarla, donde se realiza un submuestreo. El número de entradas del filtro se pone en función de la cantidad de datos disponibles, de tal manera que cuando los datos son submuestreados la cantidad de entradas es menor. La metodología consiste en generar series temporales a partir de submuestrear a la serie original, partiendo del último valor disponible hacia el más antiguo. Para cada una de las series generadas por el submuestreo, se ajustó a un filtro diferente basado en RNAs [36], [186], y cada uno de ellos genera un pronóstico que luego es promediado en su conjunto. Éste promedio, que consiste en un valor para cada uno de instantes futuros, está compuesto por el tamaño del horizonte de pronóstico. El desempeño del filtro propuesto se evalúa a través del índice SMAPE. Se puede concluir que es mejor suavizar a las series temporales de la clase rugosa, muy frecuente en las series de lluvia acumulada mensualmente. Por ello, es motivante generar algún criterio para fusionar las salidas de los filtros predictores en lugar de simplemente promediarlas.

#### 3.4.1. Descripción general del submuestreo

En esta Tesis se implementa un muestreo de la serie temporal de datos, para formar varias series diferentes. En cada una de éstas se pretende utilizar la misma estructura del filtro predictor para determinar los valores futuros  $\alpha k$  a un paso, donde  $\alpha$  es la decimación de la serie y  $\alpha=1$  representa la serie original. En la Fig. 3-5 y Fig. 3-6 se muestra un ejemplo de una serie temporal, originadas mediante la decimación de una serie temporal de Mackey Glass y de lluvia acumulada mensual, donde  $\alpha k$  varía de 0 a la longitud de la serie, y en la serie original  $\alpha=1$ .

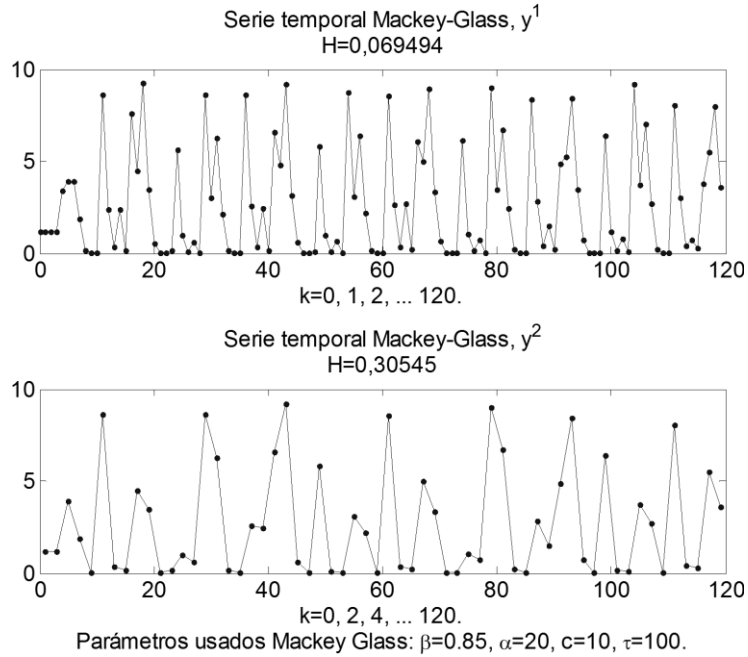


Fig. 3-5. Serie original  $y^D$ , series sub muestreadas con Decimación de D de 1, 2 y 3.

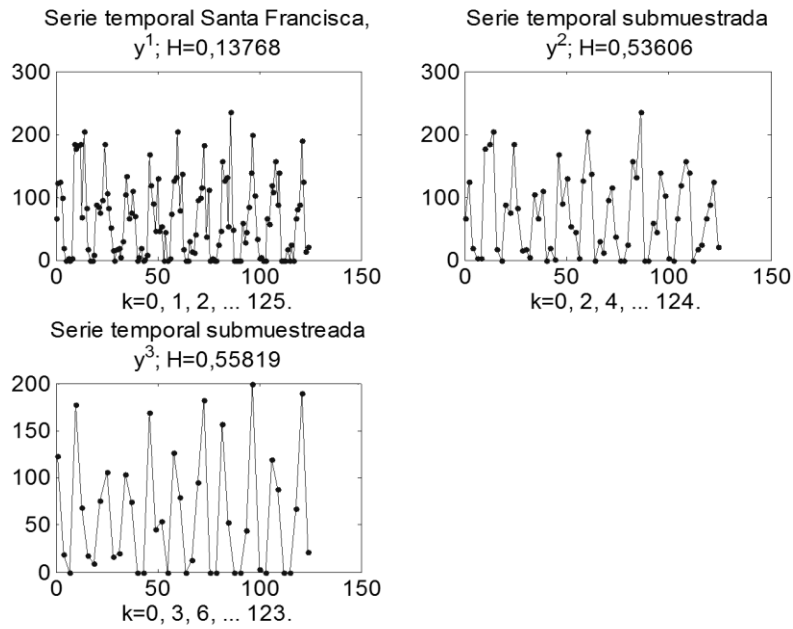


Fig. 3-6. Serie temporal original de Santa Francisca  $y^D$ , submuestreada con decimación D de 1, 2 y 3.

A partir de la serie de la ecuación

$$y_k = R(\alpha k), \quad k = 0, 1, 2, \dots, 120 \quad (3-12)$$

se genera una segunda serie muestreando a cada  $k$  impar, lo que resulta

$$y_k^D = y_k, \quad D = 2, \quad K = 0, 2, 4, 6 \dots 120 \quad (3-13)$$

siendo la cantidad de pasos que se sub muestrea de D, y en esta tesis varía de 2 a 7.

No obstante, en la Fig. 3-5. se muestran las series obtenidas cuando D varía de 1 a 2 para la serie de MG. Nótese que una serie submuestreada tiene su valor futuro a un paso, coincidente con la serie original en cada punto sub muestreado, por lo que el  $\Delta t$  pasa de 1 a 2 partiendo de  $\Delta t$  de la serie temporal original. Nótese que el parámetro H asociado a cada serie pasa 0,069494 para la serie original a 0,30545 para un sub muestreo de donde se destaca que las series obtenidas son más suaves que la original. Algo similar se observa en la serie temporal de Santa Francisca como muestra la Fig. 3-6.

### 3.4.2. Metodología Propuesta

Algunos resultados han sido obtenidos desde una aproximación lineal autorregresiva la cual se detalla en [183]. Ahora, se propone un modelo de filtro NAR [8] donde la RNA usada es realimentada con retardos en el tiempo. La topología de la RNA consiste en  $I_x$  entradas, una capa oculta de  $h_o$  neuronas, y una neurona de salida. El algoritmo de aprendizaje utilizado es el método de Levenberg-Marquardt [143]. Con el fin de predecir la secuencia  $\{x_c\}$  un paso adelante, el primer valor sin retardo que se toma desde la línea es  $x_n$ , que se toma como entrada. De esta manera, la predicción de salida puede ser expresada como

$$x_e(n+1) = F_p \left( Z^{-1} I(\{x_n\}) \right) \quad (3-14)$$

donde,  $F_p$  es el operador del filtro predictor no-lineal, y  $x_e(n+1)$  la salida del predictor en  $n$ . Para lograr la serie pronosticada, el valor generado de  $x_e(n+1)$  es utilizado como entrada del filtro  $F_p$  con el retardo adecuado,

$$x_e(n+2) = F_p \left( Z^{-1} I(\{x_n, x_e(n+1)\}) \right) \quad (3-15)$$

y se logra el valor  $x_e(n+2)$ , que a su vez es retrasado e ingresado como entrada para obtener mediante la Ecuación (3-14) el valor  $x_e(n+3)$ ,

$$x_e(n+3) = F_p \left( Z^{-1} I(\{x_n, x_e(n+1), x_e(n+2)\}) \right) \quad (3-16)$$

y así sucesivamente hasta un valor  $x_e(n+De)$ ,

$$x_e(n+D) = F_p \left( Z^{-1} I(\{x_n, x_e(n+1), \dots, x_e(n+D_e-1)\}) \right) \quad (3-17)$$

que en el caso de la serie original aquí se pretende llegar hasta 18.

Con cada una de las secuencias  $\{x_n\}^D$  se implementa un filtro  $F_p$  específico, que generará predicciones para cada serie según Ecuación (3-18). De esta manera que, si existen  $s$  series correspondientes a la secuencia  $\{x_n\}$ , existirán  $s$  filtros  $F_p$  que implementarán la predicción de la Ecuación (3-17).

Los valores que se obtienen se toman como una matriz P de entrada a una red neuronal Adaline [187], (elemento combinado adaptativo lineal que recibe todas las entradas, las suma ponderadamente y produce una salida) en la que se obtiene un conjunto de parámetros  $\{w_j\}$ , para ponderar la salida de la red principal

$$x'_e(n+1) = \sum w_i F_p^s \left( Z^{-1} I(\{x_n\}^s) \right) \quad (3-18)$$

donde  $x'_e$  es la variable a predecir cuyo argumento  $n+1$  se ubica en la serie con una decimación D según indique el índice  $s$ , la secuencia  $\{x_n\}^s$  se corresponde con  $\{x_n\}^D$  con el cambio adecuado de índices, y  $F_p^s$  se refiere al filtro predictor para cada secuencia  $\{x_n\}^s$ .

Cabe aclarar que cada vez que se implementa la expresión de la Ecuación (3-18) según el procedimiento de Ecuación (3-15), Ecuación (3-16), Ecuación (3-17) en las series submuestradas, el valor de  $D_e(n+1)$  va a cambiar, ya que va a disminuir según lo implique D. Así, para  $s=1$   $D_e=18$ , para  $s=2$   $D_e=9$ , para  $s=3$   $D_e=6$ , y así sucesivamente.

### 3.4.3. Configuración del modelo y proceso de ajuste

Las condiciones iniciales del filtro y el algoritmo de aprendizaje son mostradas en la Tabla 3-3.

Tabla 3-3. Parámetros de ajuste del algoritmo

Variable	Símbolo	Valor
Número de entradas	$l_x$	$D \cdot D_e$
Número de patrones	$N_p$	$l_x$
Número de neuronas en la capa oculta	$h_o$	$l_x + 3$
Número de iteraciones	$i_t$	$4 \cdot l_x + 10$



Nótese que el número de neuronas en la capa oculta y de iteraciones son ajustadas en función del número de entradas. Las condiciones iniciales del algoritmo de aprendizaje son usadas para el pronóstico de series temporales, cuya longitud depende de la cantidad de datos de la serie de datos.

#### 3.4.4. Ejemplificación

Lo que se hace es una decimación de una serie de datos, de la cual formamos a partir de esta, 18 series diferentes, para las cuales utilizamos el mismo filtro predictor para determinar los valores futuros de a un paso, dos pasos, tres pasos,... hasta 18 pasos. La primera serie será la original.

Tomando el ejemplo de una señal de  $y(t)=sen(t)$ ,  $t=0, 1, 2, \dots, 102$ .

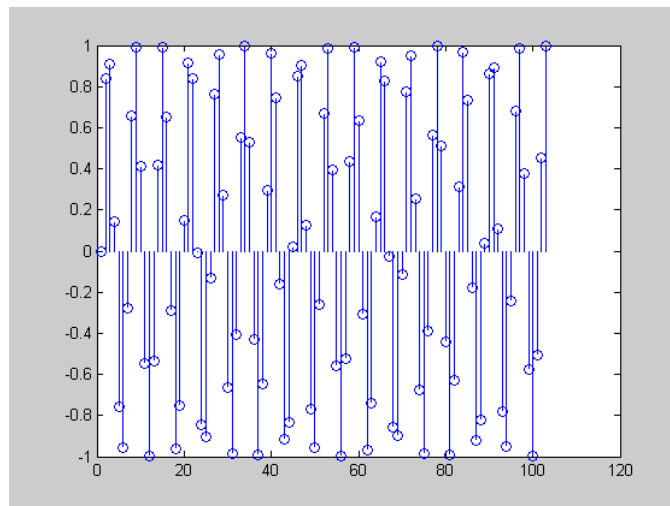


Fig. 3-7. Serie original,  $y(t)$ .

A partir de esta serie formamos la segunda serie tomando pasos de a dos. Nos queda:  $t$  va de 0 hasta 102 con paso de a 2. La serie para  $\alpha=2$  es  $y^2(t)=sen(t)$ ,  $t=0, 2, 4, \dots, 102$ .

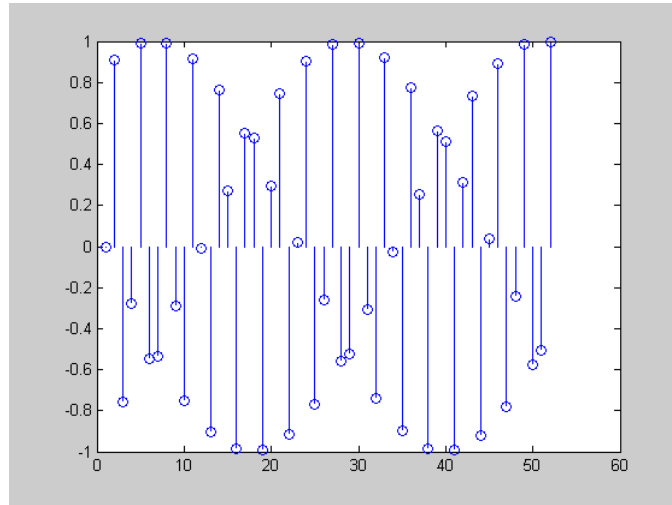


Fig. 3-8. Serie  $y^2(t)$ .

La tercera serie será con  $\alpha=3$ ,  $y^3(t)=\text{sen}(t)$ ,  $t=0, 3, 6, \dots, 102$ .

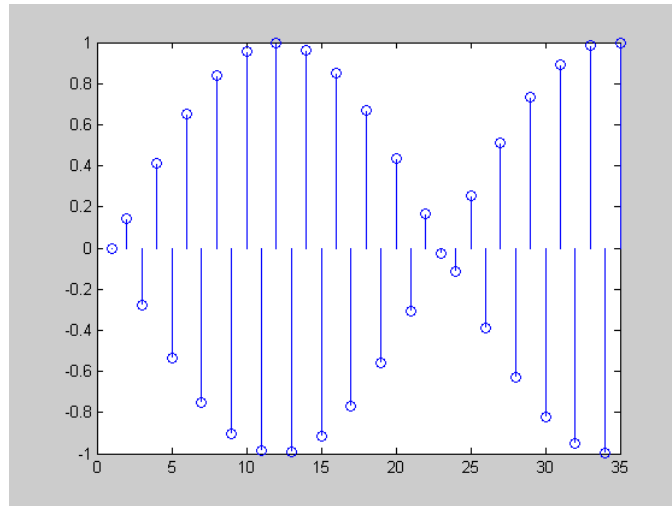


Fig. 3-9. Serie  $y^3(t)$ .

y así sucesivamente hasta hacer una supresión de 17 puntos cada 18.

Se entrena la red para los puntos de la serie original en la forma que lo venimos haciendo

$$\hat{y}^1(t_n) = w_1 y(t_{n-1}) + w_2 y(t_{n-2}) + \dots + w_k y(t_{n-k})$$

se hace el ajuste para la predicción a un paso, obteniendo una matriz  $W^1$ ,

$$\hat{y}^2(t_n) = w_1 y^2(t_{n-1}) + w_2 y^2(t_{n-2}) + \dots + w_k y^2(t_{n-k})$$

se hace el ajuste para la predicción a un paso, obteniendo una matriz  $W^2$ ,

$$\hat{y}^3(t_n) = w_1 y^3(t_{n-1}) + w_2 y^3(t_{n-2}) + \dots + w_k y^3(t_{n-k})$$

se hace el ajuste para la predicción a un paso, obteniendo una matriz  $W^3$ ,

$$\hat{y}^{18}(t_n) = w_1 y^{18}(t_{n-1}) + w_2 y^{18}(t_{n-2}) + \dots + w_k y^{18}(t_{n-k})$$

se hace el ajuste para la predicción a un paso, obteniendo una matriz  $W^{18}$ .

Luego tomamos los valores de la serie 2, entrenamos la red para predecir los segundos puntos, o sea, 2 pasos adelantados que serían de la segunda serie con una matriz de  $W^2$ , lo mismo para el valor (tercera serie), siempre utilizando los mismos puntos. Así obtendremos diferentes matrices  $W$  para cada predicción, 18 en total. Lo que vamos a ver es como influyen los puntos a 18 valores futuros. Luego observamos el error existente entre la predicción del predictor a un paso y del predictor a 2 pasos, y así hasta el paso 18, o sea que para cada punto vamos a obtener 18 predicciones diferentes, en el cual se puede hacer un promedio o ver la desviación o el tipo de comportamiento de la distribución de los puntos alrededor de los valores reales, de esta manera crear un matriz virtual de  $W$  donde cada  $W^i$  multiplique a cada predicción de cada punto (18 en total), para generar el valor real.

Durante el entrenamiento se computa el valor medio de las predicciones en cada punto.

$$\bar{\hat{y}}(t_n) = \left[ \sum_{i=1}^{18} \hat{y}^i(t_n) \right] \frac{1}{18} \quad (3-19)$$

El error en cada punto de predicción con respecto al valor medio de las predicciones, se mide como  $Error = y(t) - \hat{y}(t_n)$  y la varianza como

$$S_n = \frac{1}{18} \sum_{i=1}^{18} (y(t_n) - \hat{y}^i(t_n))^2 \quad (3-20)$$

### 3.5. Comentarios finales

En este Capítulo se han presentado tres modelos de filtros no-lineales para el pronóstico de series temporales de alta rugosidad donde se emplea la técnica suavización de la serie. Estos métodos son: energía asociada a las series y submuestreo, en contraste con el método basado en rugosidad de la serie. La principal contribución reside en generar una nueva serie suavizada, ya sea integrando la misma o generando una subserie a partir de la serie original de datos, lo cual permite ajustar los parámetros de acuerdo a un criterio tomado.

El conjunto de las series temporales ha sido submuestreada uniformemente, por valores enteros, este procedimiento demostró ser simple y sencillo. La rugosidad de las series temporales

resultantes pronosticadas ha sido evaluada mediante el exponente o parámetro de Hurst (H). Estos algoritmos fueron comparados en [188], [189] y [190] con los propuestos en la Sección 2.9.1 y Sección 2.10, obteniéndose una buena performance en los basados en RNAs. Estos métodos se mejoraron usando nuevos mecanismos de aprendizajes, como se muestran en el siguiente Capítulo.

# Capítulo 4

*“El único elemento que puede sustituir la dependencia del pasado es la dependencia del futuro.”*

John Dos Passos

## 4. Métodos basados en Aproximación

### Bayesiana

Este capítulo presenta tres nuevos algoritmos basados en la teoría de aproximación bayesiana para el diseño de filtros predictores de series temporales rugosas cortas o largas.

Estos enfoques usan RNAs de retroalimentación hacia adelante (FFNN) y se establecen los parámetros de un modelo NAR en función de modelos probabilísticos y heurísticos. Dependiendo de la rugosidad de la serie temporal, se propone modelar los parámetros y la estructura mediante la inferencia bayesiana en el supuesto que los parámetros siguen una regla estocástica para el ajuste de los pesos de la red neuronal. La regla de sintonización utilizada en el proceso de ajuste se basa en el método de Levenberg-Marquardt.

#### 4.1. Introducción

Las redes neuronales han demostrado ser uno de las mejores herramientas para el pronóstico de series temporales. Sin embargo, dos principales dificultades deben ser bien tratadas: uno es el control de la complejidad del modelo

[191] y la otra es que los modelos de redes neuronales convencionales carecen de herramientas para analizar los resultados de salida, por ejemplo, intervalos de confianza, nivel de confidencialidad

(95% o 5%) y cuantíles [192]. Buntine [193] y Mackay [194] fueron los pioneros en usar el enfoque bayesiano para el entrenamiento de las redes neuronales bayesianas (BNN). MacKay [195] introdujo el enfoque bayesiano basado en la aproximación de Gauss y Neal [170] [196] adoptó el método de Monte Carlo Híbrido que facilita el aprendizaje bayesiano para redes neuronales sin aproximaciones. En la siguiente sección se presenta un enfoque para ajustar los parámetros de un filtro autorregresivo no lineal basado en RNAs usando el método de inferencia bayesiana para el pronóstico de series temporales, ya sean estas provenientes de lluvia acumulada mensual o series caóticas artificiales. Las RNAs que componen al filtro predictor tienen un número incierto de parámetros, el cual es ajustado mediante la inferencia Bayesiana [143]. Los parámetros de las RNAs son considerados como variables aleatorias cuya distribución posterior de probabilidad es inferido desde los datos basándose en la teoría de aproximación bayesiana.

### 4.2. Teoría de Aproximación Bayesiana

Cuando se toma a las máquinas de aprendizaje como modelos de ajustes en el sentido paramétrico o no paramétrico para los datos con el fin de hacer predicción, es en este sentido que el enfoque bayesiano se ha basado en verlos como modelos probabilísticos [197]. Así cuando tomamos el modelo probabilístico, literalmente, como un modelo del proceso que genera los datos; entonces podemos usar la inferencia bayesiana para implementar el aprendizaje y la predicción [198]. Esta perspectiva ha permitido en las últimas dos décadas producir mejores predicciones utilizando el concepto bayesiano de marginación [199] [200] [201]. Todas estas mejoras en el aprendizaje dependen literalmente de la creencia del modelo propuesto. Una alternativa, sin embargo, es pensar en él como una caja negra no teniendo relación con el proceso de la generación de los datos, o nuestras creencias acerca de qué proceso lo generó [39] [38]. La mayoría de los trabajos en la teoría de aproximación bayesiana han utilizado una teoría de muestreo [105], la entrega de los límites en la generalización de error que es probable en el sentido de la teoría del muestreo para ser verdad [202].

En el enfoque bayesiano se especifica un modelo de muestreo  $P(M/\theta)$  para los datos dados los parámetros y una distribución previa para los parámetros  $P(\theta)$ , que refleja el conocimiento sobre  $\theta$  antes de ver los datos. A continuación, el cálculo de la distribución posterior

$$P(\theta / M) = \frac{P(M / \theta)P(\theta)}{\int P(M / \theta)P(\theta)d\theta} \quad (4-1)$$

representa el conocimiento actualizado sobre  $\theta$  después de observar los datos [203]. Para entender esta distribución posterior, se puede extraer muestras de ella o calculando su media o modo. El método bayesiano difiere del método estándar ("frecuentista") en el uso de una distribución a priori para expresar la incertidumbre presente antes de ver los datos, y permitir que la incertidumbre restante después de los datos observados, se exprese en forma de una distribución posterior.

La distribución posterior también proporciona la base para predecir valores futuros a través de la distribución predictiva

$$P(x^{nuevo} / M) = \int P(x / \theta)P(\theta / M)d\theta. \quad (4-2)$$

La teoría de aproximación Bayesiana permite la propagación de incertidumbre en las cantidades que son desconocidos para otros supuestos en el modelo, que puede ser generalmente válido o más fáciles de adivinar que en el problema. La principal dificultad en la construcción de modelos es controlar la complejidad [204].

Es bien sabido que el número óptimo de grados de libertad de un modelo depende del número de muestras para entrenar, cantidad de ruido en las muestras y la complejidad de la función que se calcula. El procedimiento para determinar la densidad a priori y las funciones de máxima verosimilitud asociadas con la incertidumbre en las series temporales de lluvia acumulada es complicado y no es un requisito para asumir la linealidad y distribución normal en el marco de los parámetros propuestos [205].

El problema de la selección del modelo suele estar dividida en descubrir la organización de parámetros ajustan perfectamente, como la topología de la red, por ejemplo, así como el número de patrones, numero de capas, neuronas ocultas por capa, los cuales resultan en la mejor generalización del rendimiento. Un resultado común en el entrenamiento de las RNAs con demasiados parámetros libres es que se tiende a un *sobrepasamiento* usando los datos de entrenamiento y, por lo tanto, se tiene un escaso rendimiento [206]. En este contexto, sólo los parámetros útiles deben ser utilizados como entradas a la red neuronal [207]. Una RNA que recibe entradas válidas y entradas "con ruido" contendrá demasiados parámetros libres y, por lo tanto, ser propensa al *sobrepasamiento* de los datos de entrenamiento dando lugar a una mala generalización [208].

### 4.3. Método Bayesiano de aprendizaje

El enfoque de máxima verosimilitud para el entrenamiento de las RNAs encuentra un único vector de peso que minimiza una función de error; por el contrario, el esquema Bayesiano considera una distribución de probabilidad sobre los pesos  $\mathbf{w}$ . Esto se describe por una distribución a priori  $P(\mathbf{w})$  que es modificado cuando se observan los datos  $D$ . Este proceso se puede expresar con el Teorema de Bayes

$$P(\mathbf{w} / D) = \frac{P(D / \mathbf{w})P(\mathbf{w})}{P(D)} \quad (4-3)$$

Para evaluar la distribución posterior, necesitamos expresiones para el  $P(\mathbf{w})$  y la probabilidad  $P(D / \mathbf{w})$ . En las siguientes secciones mostrarán cómo elegir estas funciones de probabilidad de densidad (PDF) para el problema en cuestión.

En el sentido bayesiano, el producto del aprendizaje se da como un conjunto plausible de los valores de los parámetros, en tal sentido, solo se evalúa la probabilidad posterior que se obtiene multiplicando la máxima verosimilitud por la distribución a priori elegida sobre el espacio  $\mathbf{w}$ .

Es por ello que el enfoque bayesiano [209] [210] provee un marco unificado para el modelado de los parámetros de la red neuronal. La meta es desarrollar modelos probabilísticos que logren la mejor aproximación entre los datos y hacer predicciones óptimas [211].

En este sentido se encuentran algunas ventajas; la primera es que el método bayesiano es mecánico. Una vez que se tiene el modelo de espacio bien definido, las reglas que lo definen dan una única respuesta en base a los datos que se tienen, contradiciendo los métodos estadísticos tradicionales, los cuales se deben inventar estimadores de interés y luego elegir esos estimadores usando algún criterio midiendo sus propiedades.

La inferencia bayesiana, por el contrario [212] satisface el principio de probabilidad, donde las inferencias hechas solo dependen de las probabilidades asignadas a los datos recibidos y no en las propiedades de otros conjuntos de datos que pudieron haber ocurrido, pero no lo hicieron. Por consiguiente, el modelado probabilístico maneja la incertidumbre de una manera más natural. El inconveniente que se tiene es la marginalización de las probabilidades para incorporar incertidumbre en los parámetros dentro de nuestras predicciones de otras variables [213].

Por otro lado, el modelado Bayesiano personifica el principio de la “*Navaja de Occam*”, según el cual la solución más simple es siempre preferible. En esta tesis se presenta una justificación



diferente de este principio que involucra inferencias coherentes para expresar el principio de *Occam* en forma de predicción cuantitativa [214].

Por último, un tipo útil de distribuciones a priori son los conjugados [215], los cuales cuando se combinan producen un modelo de distribución posterior en la misma familia. Hay que tener bien en cuenta que la idea de la conjugación es la probabilidad dependiente, en que una distribución a priori es conjugada para una probabilidad particular.

Es por ello que las distribuciones a priori conjugadas son ampliamente utilizadas por conveniencia, ya que ellos llevan a distribuciones posteriores analíticamente manejables. En muchos casos, tales como una red neural, no se conoce un conjugado a priori. Las distribuciones a priori para los parámetros individuales pueden ser elegidas para ser *condicionalmente conjugada*, en el sentido que cuando todos los demás parámetros están condicionados, la distribución posterior es condicional y pertenece a la misma familia que la condicional a priori. Casi todas las distribuciones previas presentadas por las redes neuronales ponen una distribución normal de los parámetros pesos sinápticos (posiblemente condicionado a los otros parámetros), que es condicionalmente conjugado. Sin embargo, los parámetros de una red neural no son típicamente interpretables. Esto hace que la elección de una distribución informativa a priori resulte difícil [216].

### 4.3.1. Predictor basado en el Método Bayesiano

En esta sección se presenta el algoritmo basado en el *método Bayesiano* (BA) [37] [217], que es usado como algoritmo de referencia para los nuevos modelos propuestos basados en inferencia bayesiana.

Uno de los principios más importantes del modelado Bayesiano es construir las distribuciones de probabilidad posterior para todos los parámetros desconocidos en un modelo, dado la muestra de datos  $D$ . Para utilizar el modelo, las distribuciones marginales se construyen para todos esos parámetros en los que se están interesados, es decir, en este caso son las salidas de las RNAs. Estos pueden ser los parámetros en los modelos paramétricos, o las predicciones en regresiones (no-paramétrico) o tareas de clasificación. El uso de las probabilidades a posteriori requiere una definición explícita de las probabilidades a priori para los parámetros. La probabilidad posterior de los parámetros en un modelo  $M$  dado los datos  $D$  es, según la regla de Bayes,

$$P(\theta / D, M) = \frac{P(\theta / D, M) \cdot P(\theta / M)}{P(D / M)} \quad (4-4)$$

donde  $P(\theta / D, M)$  es la verosimilitud de los parámetros  $\theta$ ,  $P(\theta / M)$  es la probabilidad a priori  $\theta$  dado el modelo  $M$  y  $P(D / M)$  es una normalización constante, llamado la evidencia del modelo  $M$ . El término  $M$  denota todas las hipótesis y supuestos que son realizados en la definición del modelo. Todos los resultados están condicionados por estos supuestos, y para hacer más claro esto preferimos tener el término  $M$  explícitamente en las ecuaciones. En esta notación término de normalización  $P(\theta / M)$  es directamente entendible como la probabilidad marginal de los datos, condicionado a  $M$ , integrado sobre todas las hipótesis elegidas  $M$  y a priori  $P(\theta / M)$  comprenden

$$p(x_e / x_n) = \int p(x_e / x_n, \theta) p(\theta) d\theta \quad (4-5)$$

Cuando se tienen varios modelos,  $P(D / M)$  es la verosimilitud del modelo, que pueden ser utilizadas para comparar las probabilidades de los modelos, de ahí el término "*evidencia del modelo*". Uno de los métodos más utilizados de elección del modelo Bayesiano entre dos modelos se basa en el factor de Bayes,  $P(D / M1) / P(D / M2)$ . La notación más común de la fórmula de Bayes, con  $M$  dado, causa fácilmente la malinterpretación del denominador  $P(\theta)$  como una especie de probabilidad de obtener datos  $D$  en el problema estudiado (o probabilidad a priori de los datos antes de adoptar el modelo  $M$ ). La distribución a priori expresa nuestras creencias acerca de los valores de los parámetros antes de que cualquier dato es observado. Después de que nuevos datos se observan  $D = \{(x^{(1)}, x_e^{(1)}), \dots, (x^{(n)}, x_e^{(n)})\}$ , la distribución a priori será actualizada a la distribución posterior utilizando regla de Bayes donde  $P(\theta / M)$  es función de verosimilitud de los parámetros desconocido del modelo a los datos observados.

En caso de datos independientes e intercambiables, la función de verosimilitud es

$$P(\theta / D) = \prod_{i=1}^n P(x_e^{(i)} / x^{(i)}, \theta) \quad (4-6)$$

donde  $n$  es el número de datos. Para predecir la nueva salida  $x_e(n+1)$  para la nueva entrada  $x(n+1)$ , se obtiene la distribución predictiva de integrar las predicciones del modelo con respecto a la distribución posterior de los parámetros del modelo

$$\begin{aligned} P(\theta / D) &= P(x_e^{(n+1)} / x^{(n+1)}, D) \\ &= \int_{\Omega} P(x_e^{(n+1)} / x^{(n+1)}, D) \cdot P(\theta / D, M) d\theta, \end{aligned} \quad (4-7)$$

donde  $\Omega$  es el espacio de todos los posibles parámetros. Nótese que la distribución predictiva para  $x_e(n+1)$  está condicionado implícitamente en las hipótesis que se mantienen en todo y para ser más explícitos como el siguiente [213]

$$P(x_e^{(n+1)} / D, M) = \int_{\Omega} P(x_e^{(n+1)} / x^{(n+1)}, \theta, D) \cdot P(\theta / D, M) d\theta, \quad (4-8)$$

donde  $M$  hace referencia al conjunto de hipótesis o supuestos utilizados para definir el modelo. Prácticamente, la distribución posterior en los parámetros de Ecuación (4-7) es muy compleja y con muchos modos. Como resultado, evaluar la integral es una tarea difícil. Por ende, para resolver este problema hay varios métodos, Neal [196] presentó la Cadena de Markov Monte Carlo (MCMC) para realizar este tipo de integraciones difíciles. A continuación, el método MCMC ha sido utilizado por muchos otros autores [188] [189] [192] [194] [200] [211] [218] [219] [220].

### 4.3.2. Entrenamiento Bayesiano de las RNAs

Cuando se analizan las series con memoria de corto o largo alcance, es importante hacer uso del modelo más simple posible. Específicamente, el número de parámetros desconocidos deben mantenerse al mínimo. La distribución *Gamma* ha sido considerada en la literatura para este propósito. Cuando un análisis Bayesiano se lleva a cabo, las inferencias sobre los parámetros desconocidos se derivan de la distribución posterior. Este modelo de probabilidad describe los conocimientos adquiridos después de observar un conjunto de datos. La aplicación del problema de regresión afecta al entrenamiento correspondiente de la RNA. Tales elementos son los  $N$  pares de entrada-salida, el vector de entrada  $I_x$  y el vector de datos objetivos  $t_n$  ( $n=1, \dots, N$ ). Este último representa la salida deseada ó conjunto de validación usado para ajustar el entrenamiento dado el vector  $I_x$  de entrada. Asumiendo ruido gaussiano en la salida de la RNA, la función probabilidad tiene la siguiente forma:

$$P(D / w, M) = \left( \frac{\beta}{2\pi} \right)^{N/2} \exp \left\{ -\frac{\beta}{2} \sum_{n=1}^N \|y(x_n; w) - t_n\|^2 \right\} \quad (4-9)$$

donde  $\beta$  es un hiperparámetro que representa la inversa de la varianza del ruido. Este trabajo asume que las RNB poseen una sola capa oculta de neuronas de unidades “*tanh*” y una neurona de salida lineal. Para completar el enfoque Bayesiano, la información previa de la red es necesaria. Se propone utilizar, análoga a los términos de penalización, la siguiente ecuación

$$P(w) = (2\pi w^2)^{-N/2} \exp\left(-\frac{|w|^2}{2w^2}\right) \quad (4-10)$$

suponiendo que el escalado numérico de los pesos  $w$  es dada a mano. Esto se lleva a cabo teniendo en cuenta que la función de red  $f(x_n+1, w)$  es aproximadamente lineal con respecto a  $w$  en las proximidades de este modo, de hecho, la distribución predictiva de  $x_n+1$  será otra gaussiana multivariada.

## 4.4. Método Bayesiano mejorado por movimiento Browniano Fraccional

### 4.4.1. Motivación

En el contexto del pronóstico, el futuro nunca es exacto como el pasado, lo que significa que la exactitud de las predicciones nunca puede ser asegurada. La pregunta crucial es la extensión de la inexactitud de tales predicciones.

La mayoría de los pronósticos de series temporales son influenciados por sucesos aleatorios y a menudo se comportan no muy lejos de los caminos aleatorios ó “*random walks*”, favoreciendo a los métodos simples que son capaces de suavizar dicha aleatoriedad.

En el pronóstico largo alcance, la exactitud de las predicciones cae abruptamente mientras la incertidumbre aumenta. Es por ello que se busca la reducción del error en el pronóstico promediando más de un modelo. Por ende, los resultados no son solo de gran exactitud, pero también hay una reducción en la dimensión de los errores pronosticados, con el simple uso del promedio siendo el mejor camino para combinar distintos pronósticos. Rossi *et. al* [221] argumentan el hecho que no hay otros métodos excepto el Bayesiano, el cual provee un tratamiento modificado de inferencia y decisión como así también contando propiamente los parámetros de la incertidumbre del modelo. Sin embargo, la inferencia Bayesiana provee respuestas condicionadas

sobre los datos observados en vez de basarse en la distribución de los test estadísticos sobre muestras imaginarias no observadas [222]. Aunque el método Bayesiano tiene un importante beneficio, este incluye costos triviales en la formulación del modelo a priori de los datos, requiriendo de una función de verosimilitud y la computación de varias integrales requeridas en el paradigma Bayesiano. Con el advenimiento de los avances computacionales se hace posible la resolución de estas integrales. Además, la elección de un apropiado modelo a priori de los datos ha sido un problema en el modelado Bayesiano [223]. En la siguiente sección se propone un modelo de distribución adoptado para las series temporales.

#### 4.4.1. Método mejorado usando movimiento Browniano Fraccionario

En esta subsección se presenta un nuevo algoritmo basado en el método Bayesiano mejorado (BEA) [38] usando la función de probabilidad del movimiento Browniano fraccionario [224], con un modelo beta de distribución como distribución a priori cuando el proceso es observado en tiempo discreto. La inferencia bayesiana es utilizada para actualizar la distribución a priori en la distribución posteriori incorporando la información provista como una función de probabilidad del modelo Browniano fraccionario [225],

$$f(x/H) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{R}|^{\frac{1}{2}}} \exp\left\{\frac{1}{2} x^T \mathbf{R}^{-1} x\right\} \quad (4-11)$$

dado por los pesos de las redes neuronales desde los datos observados para generar pronósticos de puntos e intervalos mediante la combinación de toda la información y las fuentes de incertidumbre en una distribución predictiva para los futuros valores.

Se sabe que el vector de peso  $w$  define un mapeo desde un vector de entrada  $x$  hacia un vector de salida predictor  $\hat{y}$  dado por  $\hat{y} = f(x, w)$ . Asumiendo la función de distribución del movimiento Browniano fraccional como modelo, la distribución de probabilidad condicional para la salida viene dado por el vector de entrada  $l_x$  como sigue:

$$P(D/w, \mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{R}|^{\frac{1}{2}}} \exp\left(-\frac{(y(x_n; w) - t_n)^T \mathbf{R} (y(x_n; w) - t_n)}{2}\right), \quad (4-12)$$

$$P(t / H) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{R}|^{\frac{1}{2}}} \exp\left(-\frac{t^T \mathbf{R} t}{2}\right), \quad (4-13)$$

La aplicación al problema de regresión involucra la correspondiente función de la red neuronal  $y(x, w)$ , tomada con  $\tanh$  y el conjunto de datos que consiste de  $N$  pares, el vector de entrada  $l_x$  y los valores a pronosticar  $t_n = (t_1, t_2, \dots, t_n)$  (esto es el conjunto de validación usado para ajustar el entrenamiento). Para completar el método mejorado Bayesiano, se requiere información a priori para los pesos de la red neuronal. La distribución *beta* es elegida para este propósito. La función de densidad beta es muy versátil para representar las salidas como proporciones o probabilidades. Esta está definida en el intervalo continuo  $[0,1]$ . Hay dos parámetros  $\alpha$  y  $\beta$  los cuales trabajan conjuntamente para determinar si la distribución tiene un modo en el intervalo unitario y si es simétrica. Este es un modelo de probabilidad el cual describe el conocimiento ganado después de observar el conjunto de datos observados. Es por ello que se propone el uso del modelo de la función de densidad del movimiento Browniano fraccional, donde  $H$  es el parámetro de Hurst,  $t_n = (t_1, t_2, \dots, t_n)$  es el conjunto de validación que sigue al *mBf* estándar y la siguiente ecuación describe la relación entre el parámetro  $H$  y la estimación de los datos a priori,

$$\mathbf{R}_{ij} = \frac{1}{2} \left( i^{2h} + j^{2h} - |i - j|^{2h} \right) \quad (4-14)$$

asumiendo que la escala esperada de los pesos de las redes neuronales está dada por  $w$  a mano.

La distribución beta está dado por

$$P(h) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} h^{\alpha-1} (1-h)^{\beta-1}, \quad (4-15)$$

donde  $\alpha$  y  $\beta$  son los parámetros de dicha distribución que varían entre 0 y 1.

El modelo de probabilidad completo se deriva del producto de la Ecuación (4-13) y Ecuación (4-15) como sigue

$$P(w / h) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{R}|^{\frac{1}{2}}} \exp\left(-\frac{w^T \mathbf{R} w}{2}\right) \frac{\beta^\alpha}{\Gamma(\alpha)} h^{-\alpha-1} \exp\left(-\frac{\beta}{h}\right), \quad (4-16)$$

Entonces, la distribución posteriori se forma de la siguiente manera

$$P(h/w) = \frac{\frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{R}|^{\frac{1}{2}}} \exp\left(-\frac{w^T \mathbf{R} w}{2}\right) P(h)}{\int_0^1 \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{R}|^{\frac{1}{2}}} \exp\left(-\frac{w^T R w}{2}\right) P(h) dh}, \quad (4-17)$$

El algoritmo Metrópolis-Hasting fue usado para la computación con un valor inicial de 0.1. El número de iteraciones fue iniciado en 10.000. El error Monte Carlo fue usado para examinar la convergencia, esto fue llevado a cabo considerando que la función de la red  $f(x_n + 1, w)$  es aproximadamente lineal con respecto a  $w$  en la vecindad de este modo, de hecho, la distribución predictiva  $y_{n+1}$  será otra Gaussiana multivariada.

## 4.5. Método mejorado Bayesiano modificado usando la entropía Renyi en la serie

### 4.5.1. Motivación

En esta sección se presenta un nuevo algoritmo llamado método mejorado bayesiano modificado (BEMA) [39] [226], basado en el modelo anterior BEA. El entrenamiento de los pesos de la RNA se modifica usando un método heurístico mediante la entropía Rényi de permutación en las series de datos, el cual asume que para actualizar una distribución a priori en una distribución posterior mediante la incorporación de la función de probabilidad del modelo de distribución del movimiento Browniano fraccional, a partir de datos observados se pueden generar pronósticos mediante la combinación de todos los datos y fuentes de incertidumbre en una distribución predictiva para los valores futuros [227].

### 4.5.2. Introducción

Se conjetura que la complejidad inherente de los datos de las series temporales del mundo real resulta de la no linealidad del proceso de generación, como así también de ruido, que están directamente correlacionado con la predicción [228]. La mayoría de los métodos se han ideado con

el fin de extraer la mayor cantidad de información posible a partir de los datos observados [229]. Algunos aspectos importantes de este análisis se refieren a, por ejemplo, si los datos son indicativos de la dinámica caótica subyacentes, si surgen éstas de un atractor de baja dimensionalidad [48], si este atractor tiene una dimensión fractal que se puede deducir de la serie temporal, por ejemplo de las series de lluvia [230] [231]. Muchas cuestiones se han planteado para estimar la dimensión fractal [232] [233] o información entrópica a partir de datos de series temporales [234]. Esto conduce naturalmente a una pregunta importante y desafiante: dada una serie temporal real ruidosa de un sistema desconocido, *¿existe un modelo de pronóstico el cual pueda inferir la información (si los hay) que está siendo transmitida hacia delante en el tiempo dado por el proceso generador subyacente?* En el tratamiento de señales complejas, el conocimiento del pasado no da ninguna idea del futuro, independientemente de cuál es el modelo que se elija para su uso.

El interés de esta metodología se centra en proponer un modelo para estas series con el fin de llevar a cabo predicciones a corto plazo. Estos horizontes de pronóstico son importantes para muchos problemas prácticos, especialmente en la predicción de lluvia local.

Por consiguiente, se propone calcular la entropía de los datos de las series temporales para modificar la estructura de los parámetros de redes neuronales asumiendo el modelo BEA, mencionado en la sección anterior. Las condiciones iniciales de patrón ordinal y la entropía de permutación demuestran ser muy prometedor en la cuantificación y el análisis del comportamiento dinámico de series temporales caóticas y otras series en la literatura. La entropía de permutación se introdujo por Bandt y Pompe [235] en 2002 como una nueva medida de la complejidad de las series temporales, la cual extrae la información cualitativa de las series temporales no lineales. Estudios recientes [236] han abordado esta cuestión, sobre la base de cálculo de la entropía de permutación de secuencias simbólicas, derivado de la serie temporal original mediante patrón ordinal [237] [238] [239]. La serie temporal se convierte en una secuencia simbólica señalando el orden en que la señal pasa a través de las ventanas sucesivas. Si se conoce la dinámica exacta, una partición puede ser ideada con el fin de obtener una representación simbólica fiel de la señal a estudiar.



### 4.5.3. Entropía de Renyi

La estimación de la entropía de una serie temporal real arbitraria es un desafío laborioso [240]. En teoría de la información, la entropía o entropía de información de Shannon es una medida de la incertidumbre asociada con una variable aleatoria. El concepto fue introducido por Claude E. Shannon [241].

Bandt y Pompe [235] sugirieron un enfoque para el análisis de series temporales en el que se insertan en una serie temporal continua como una secuencia simbólica en otro espacio, un proceso que llamaron "*entropía de permutación*". Uno de los ingredientes principales de la entropía de permutación es el patrón ordinal. El patrón ordinal de una secuencia de elementos  $x_1, \dots, x_n$  es la permutación (reorganización)  $\pi = (i_1, i_2, \dots, i_n)$  que ordena los valores de amplitud en orden ascendente de manera que  $x_{i1} \leq x_{i2} \leq \dots \leq x_{in}$ . El orden de la entropía de permutación  $L$  de una serie temporal  $x_1, \dots, x_N$  se calcula como sigue.

Sea  $\pi_t$  el patrón ordinal para el segmento de la serie temporal bajo la ventana deslizante de longitud  $L$  que termina en  $x_t$ , es decir, la subsecuencia  $x_{t-L+1}, \dots, x_t$ . Sea  $S_L = \{\pi_k\}$  el conjunto de la totalidad de los patrones únicos  $\pi_t$  ordinales. Para la serie temporal  $x_1, \dots, x_N$  corresponde la secuencia  $\langle \pi_t : t = L, \dots, N - L + 1 \rangle$  de  $N - L + 1$  patrones ordinales del alfabeto  $S_L$ . La entropía de esta secuencia de patrones ordinales es la entropía de permutación de la serie temporal  $x_{1..N}$ . Por ejemplo, la entropía de permutación Shannon se define como,

$$H_L = - \sum_{k \in S_L} P(\pi_k) \log(P(\pi_k)). \quad (4-18)$$

donde  $P(\pi_k)$  es la frecuencia de  $\pi_k$  en la secuencia  $\langle \pi_t \rangle$ . Aquí se presenta y se utiliza la entropía Rényi con parámetro  $\alpha$  de la secuencia  $\langle \pi_t \rangle$  definido como

$$R_L^{(\alpha)} = \frac{1}{1-\alpha} \log \left( \sum_{k \in S_L} P(\pi_k)^\alpha \right). \quad (4-19)$$

El parámetro  $\alpha$  en la entropía de Rényi actúa como un selector de probabilidades. Se asigna casi el mismo peso a cada posible probabilidad, cuando  $\alpha$  es suficientemente cercano a cero. Cuando  $\alpha$  es más grande, se pone más peso en las probabilidades más altas. Con esta propiedad, la entropía de Rényi puede filtrar los pequeños acontecimientos de probabilidad, y captar mejor la esencia del sistema. Por lo tanto, una serie temporal que consta de variables aleatorias *i.i.d.*, tales como ruido

blanco, tienen tasas de máxima entropía, mientras que las series de tiempo muy estructuradas, por ejemplo, periódicas, tienen muy baja (o cero) las tasas de entropía. Una serie temporal con un alto índice de entropía es casi impredecible; al contrario, uno con tasa baja de entropía es a menudo bastante predecible.

#### 4.5.4. Entrenamiento

Se propone un método heurístico para entrenar las RNAs a través de la información proporcionada por la entropía de Rényi [242] con parámetros  $\alpha$  de la secuencia de entrada de datos  $I_x$  para ajustar la ley de aprendizaje. El método de aprendizaje propuesto consiste en cambiar el número de patrones  $N_p$ , la longitud del filtro  $l_x$  y el número de iteraciones  $i_t$  en función de la entropía Rényi para cada serie temporal correspondiente. El proceso de aprendizaje se lleva a cabo utilizando un modelo de proceso por lotes. En este caso la actualización de los pesos de la red neuronal se realiza después de la presentación de todos los datos de entrenamiento, que forma una época. Además, el criterio propuesto para modificar el par viene dada por el número de iteraciones realizadas por cada época se da por

$$l_x \leq i_t \leq R_L^{(\alpha)} \cdot l_x. \quad (4-20)$$

$$l_x \leq N_p \leq R_L^{(\alpha)} \cdot l_x. \quad (4-21)$$

donde  $l_x$  es la dimensión del vector de entrada. A continuación, se propone un ajuste heurístico para el par  $(i_t, N_p)$  en la función de acuerdo con las funciones de pertenencia se muestran en la Fig. 4-1.

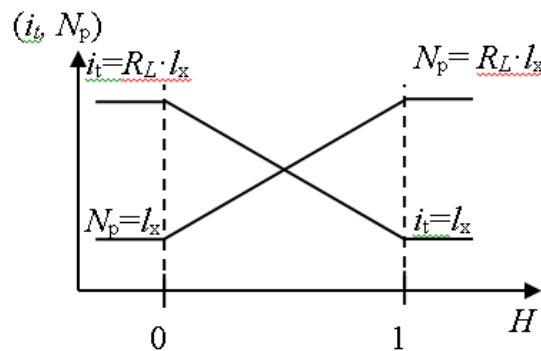


Fig. 4-1 Ajuste heurístico del par  $(i_t, N_p)$  en termino of  $R_L^{(\alpha)}$  después de cada época.

Con respecto a la estimación entrópica  $H_L$  de una serie temporal dada, un buen punto de partida es el diagrama  $n$ -gramas (bloque) de la entropía de Shannon como se sugiere en [47] y [243]. Para obtener la entropía de series temporales más complejas también se puede considerar el uso de RNAs no lineales multicapas [135]. Los caminos simulados se añaden al final de la secuencia original  $\{x_j\}$ . Suponiendo que la última entropía observada disponible es  $H_L$ , para cada traza se puede observar la secuencia de la entropía que corresponde  $H_L \rightarrow H_{L+1}^k \rightarrow H_{L+2}^k \rightarrow \dots \rightarrow H_{L+N}^k$ . Para los errores de predicción en el algoritmo, se consideró que éstas están delimitadas por  $|\hat{H}_{L+i}^k - H_{L+i}^k| \leq 1$ .

### 4.6. Comentarios finales

En este capítulo, se han presentado tres nuevos métodos basados en redes neuronales bayesianas para la predicción de series temporales, BEMA, BEA y BA tomado como algoritmo de referencia. El método BEMA utiliza la combinación de la entropía de permutación Renyi y enfoque BEA para ser aplicados a la predicción de series temporales, en particular, series temporales cortas. Las dificultades de estos algoritmos para pronosticar de series no lineales, no estacionarias son atribuibles a calcular la entropía de los datos de las series temporales, por un lado, para modificar la estructura de los parámetros de redes neuronales y por otro asumiendo el modelo BEA que infiere a partir de las observaciones en el modelo de datos, la dependencia estocástica entre el pasado y el futuro. Estos métodos demuestran ser muy prometedor en la cuantificación y el análisis del comportamiento dinámico de series de temporales caóticas, como las series de lluvia.

# Capítulo 5

*“Nosotros siempre sobrestimamos el cambio que ocurrirá en los próximos 2 años y subestimamos los cambios que ocurrirán en los próximos 10 años.”*

Bill Gates

## 5. Resultados experimentales

El objetivo principal de este capítulo es evaluar y comparar los algoritmos de la literatura con los algoritmos desarrollados en este trabajo en series caóticas artificiales y de lluvia acumulada mensual. A nivel de métodos empleados, los algoritmos de referencia son las redes neuronales basadas en rugosidad (NN-Mod.) con los competidores EAS (basado en la energía de la serie) y SUB (basado en la decimación de la serie). Para el método de aproximación bayesiana, el BA se toma como referencia para los algoritmos competidores BEA (basado en modelos bayesiano mejorado) y BEMA (basado en la combinación entropía de Renyi y BEA). Los algoritmos competidores en esta tesis son EAS y SUB (suavizados) y BEA (basado en el modelo del  $mB\hat{f}$ ) y BEMA como una extensión de BEA. Además, para el estudio se tuvieron en cuenta series cortas y largas de lluvia mensual acumulada con distintos horizontes de pronóstico 6-9-15-18 pasos hacia adelante. Las series artificiales caóticas incluyendo Mackey-Glass-17 (MG17), Mackey-Glass-30 (MG30), el mapa Cuadrático (QUA), Logístico (LOG) y de Henon (HEN), el sistema atractor de Lorenz (LOR) e Ikeda (IK) son estudiadas mediante diferentes arquitecturas de filtros predictores basados en redes neuronales y algoritmos genéticos. El rendimiento se mide por los índices SMAPE y RMSE tanto en el entrenamiento como en los distintos horizontes de pronóstico.

## **5.1. Introducción**

Mucho sistema en el mundo natural ahora se sabe que presentan caos o comportamiento no lineal, cuya complejidad es tan grande que eran previamente considerados aleatorios. El desenlace de estos sistemas ha sido ayudado por el descubrimiento, sobre todo en este siglo, de expresiones matemáticas que exhiben tendencias similares. Mediante el análisis de estas expresiones técnicas se han desarrollado y aplicado al mundo real. El más famoso y más antigua de ellas es la ecuación logística, concebido originalmente como un modelo de crecimiento de la población y como último ejemplo está el sistema de ecuaciones utilizado por primera vez por Lorenz [244] en sus intentos de modelar las interacciones atmosféricas.

El estado futuro de un sistema complejo es imposible de conocer; sin embargo, el intentar predecir aproximadamente el estado futuro es beneficioso para quien deba tomar decisiones. En las últimas décadas, algunas técnicas no lineales se han introducido en la literatura para pronosticar el estado futuro de los sistemas caóticos.

Antes de la década de 1980, los tradicionales modelos autoregresivos lineales de media móvil (ARMA) introducidos por Box y Jenkins fueron los más populares en el área de la predicción. La desventaja que tenían los modelos ARMA es que no eran capaces de pronosticar series temporales no lineales, no estacionarias y caóticas.

En la actualidad, dependiendo del tipo de aplicación y sus características, se han desarrollado diversos modelos predictivos basados en inteligencia computacional, que se aplican para la predicción sistemas caóticos de variables climáticas [245] [246]. El estudio y utilización de las RNA para la predicción de fenómenos relacionados con el cambio climático [247] [248] y en hidrología para la predicción de la precipitación de lluvia locales, usando diferentes técnicas, las cuales pueden encontrarse en investigaciones recientes [249] [250] [251] [252] [253] [254], entre otros. Es por ello que es de suma importancia conocer cómo va a evolucionar esta variable para la toma de decisiones en un lote agrícola [255]. Para realizar una agricultura sostenible, es necesario prevenir problemas y solucionar aquellos factibles, tratando de anticiparse a hechos que inevitablemente ocurrirán y que en muchos casos se pueden prevenir [256]. Esto hace factible y adecuada la utilización de una RNA para el pronóstico de las mismas [257] [258].

Esta tesis aporta a la literatura nuevos modelos para hacer predicciones de series temporales de baja y alta rugosidad mencionadas en los Capítulos 3 y 4. Además, se alcanza mejoras en las predicciones con los algoritmos propuestos, que explotan la característica estocástica para predecir un resultado determinístico. Las evaluaciones de los resultados fueron obtenidas mediante la comparación de los resultados de los modelos aquí propuestos respecto a los de referencias y otros usados en la literatura.

Los índices SMAPE y RMSE como medida del error en el pronóstico [259] [260] [261] [262] [263] que se obtienen por los predictores propuestos, mejoran hasta un orden cien veces menos para una clase de series temporales de alta rugosidad en la señal medido a través del parámetro de Hurst, en este caso con  $H < 0,1$  que es una de las peores condiciones para pronosticar la señal.

### 5.2. Series caóticas artificiales

En este apartado, un conjunto de series temporales y sus características que son consagradas en la literatura, se discuten y se toman como referencia para la comparativa de los distintos modelos propuestos. Si bien, el tamaño de una serie temporal puede ser pequeña o grande, en esta tesis se considera un tamaño de 102 muestras para las series artificiales caóticas largas y 35 para las cortas, usadas para el entrenamiento de los filtros predictores frecuentemente en competencias internacionales. El tamaño original de las series es de 120 y 50 datos, los cuales los últimos valores se usan para validar la predicción. Se comparan los resultados de los modelos predictivos basados en redes neuronales propuestos de los Capítulos 3 y 4 con los algoritmos de referencia donde se estudian seis conjuntos de serie temporales caóticas benchmark: *Mackey-Glass (MG)*, *Lorenz (LOR)*, *Henon (LOR)*, *Ikeda (IK)*, *Logística (LOG)* y *Cuadrática (QUA)* que se obtienen de mapeo de sistemas dinámicos.

#### 5.2.1. Mackey-Glass

La ecuación MG sirve para modelar fenómenos naturales y ha sido utilizado por diferentes autores para realizar comparaciones entre diferentes técnicas para los modelos de predicción y de

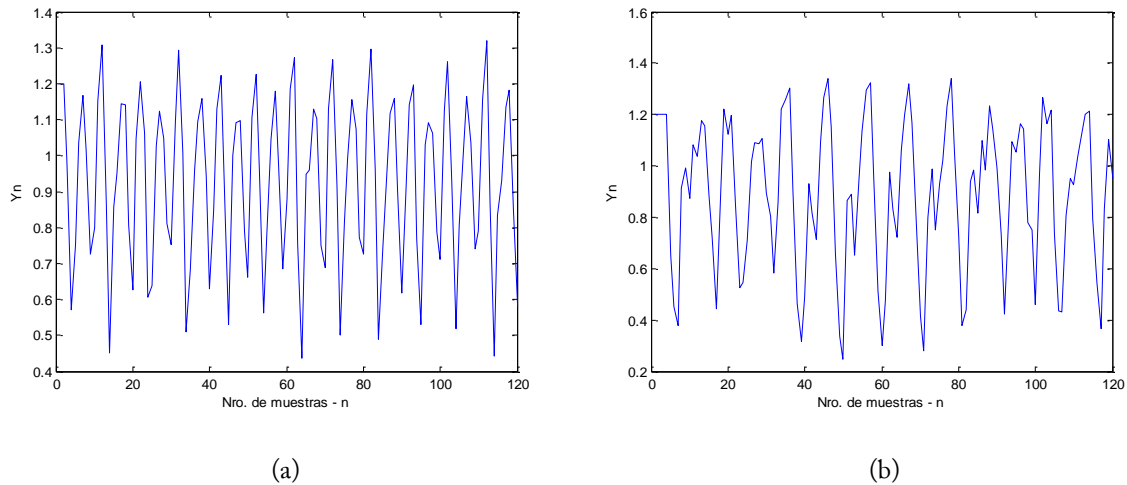
regresión [264] [265] [266]. Aquí proponemos un algoritmo para predecir los valores de serie temporales tomado de la solución de la ecuación MG [184].

La ecuación de MG se describe por la ecuación diferencial de retardo de tiempo definido como

$$\dot{y}(t) = \frac{\alpha y(t-\tau)}{1+x^c(t-\tau)} - \beta y(t) \quad (5-1)$$

donde,  $y$ ,  $c$  son parámetros y  $\tau$  es el tiempo de retardo. De acuerdo al valor que tome  $\tau$ , la solución es una serie que puede transformarse de periódica en caótica. La Ecuación (5-1) se resuelve por integración estándar Runge-Kutta de cuarto orden, y la serie se forma mediante el muestreo de los valores con un intervalo de tiempo dado.

Las series temporales a pronosticar están compuestas por muestras a partir de un punto de referencia de la solución de la ecuación de MG y la serie temporal determinista así obtenida son de diversas rugosidades. Por otra parte, al establecer el parámetro  $\beta$  entre 0,1 y 0,8, la dependencia estocástica de la serie temporal determinista así obtenida varía de acuerdo con su rugosidad.



**Fig. 5-1.** Muestras de series temporales caóticas  $\{x_n\}$  derivado del mapa de Mackey-Glass-17; a) serie MG17-1 con parámetros iniciales  $\alpha=0.2$ ,  $\beta=0.1$ ,  $c=10$  y  $\tau=17$ ; b) serie MG30-1 con parámetros iniciales  $\alpha=0.2$ ,  $\beta=0.1$ ,  $c=10$  y  $\tau=17$ .

**Tabla 5-1.** Parámetros para generar Series MG largas

Series No.	N	$\beta$	$\alpha$	c	$\tau$	H
MG17-1	120	0.2	0.1	10	17	0.98
MG30-1	120	0.2	0.1	10	30	1

**Tabla 5-2.** Parámetros para generar Series MG cortas

Series No.	N	$\beta$	$\alpha$	c	$\tau$	H
MG17-2	50	0.2	0.1	10	17	0.94
MG30-2	50	0.2	0.1	10	30	0.98

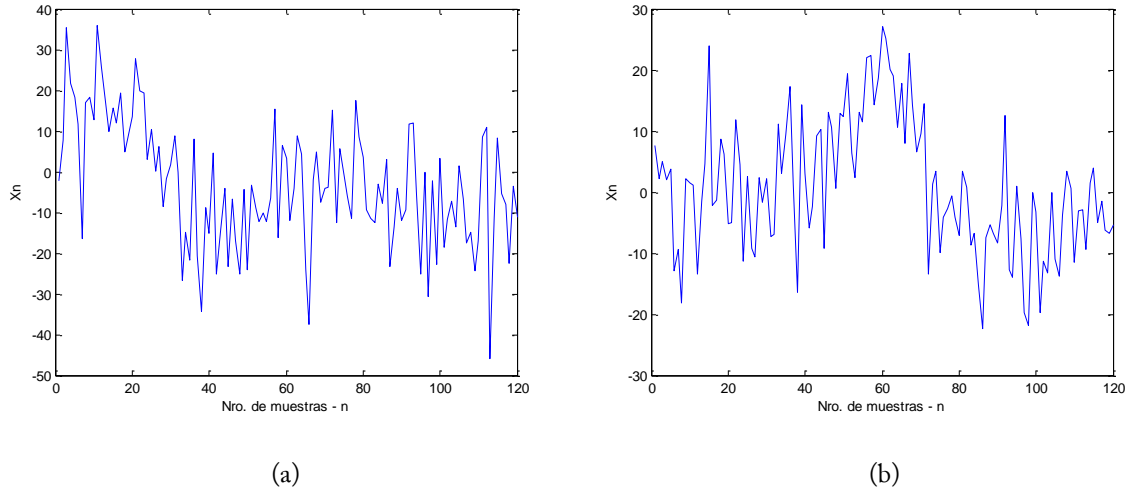
### 5.2.2. Lorenz

Lorenz encontró tres ecuaciones diferenciales ordinarias que se aproximan mucho a un modelo para la convección térmica [78]. Estas ecuaciones también se han convertido en un punto de referencia popular para probar predictores no lineales [267]. El modelo de Lorenz está dado por las Ecuaciones (5-2), los datos se derivan del sistema de Lorenz, que viene dada por tres sistemas diferenciales con retardo

$$\begin{cases} \frac{dx}{dt} = a(y - x), \\ \frac{dy}{dt} = bx - y - xz, \\ \frac{dz}{dt} = xy - cz \end{cases} \quad (5-2)$$

Una opción típica de los valores de los parámetros es  $a = 10$ ,  $b = 28$ ,  $c = 8/3$ . En este caso, el sistema es caótico. El conjunto de datos se construye mediante el uso de cuatro órdenes método de Runge-Kutta con el valor inicial como se muestra en la Tabla 5-3 para LOR01 y la serie LOR02. El tamaño del paso se elige como 0,01, respectivamente. Estos conjuntos de parámetros se utilizan comúnmente en la generación del sistema de Lorenz porque exhibe caos determinístico.





**Fig. 5-2.** Muestras de series temporales caóticas  $\{x_n\}$  derivado del mapa de Lorenz; a) con condiciones iniciales  $x(0)=12$  ,  $y(0)=9$  ,  $z(0)=2$ ; b) con condiciones iniciales  $x(0)=0.1$  ,  $y(0)=0.1$  ,  $z(0)=2$ .

**Tabla 5-3.** Parámetros para generar Series LOR

Series No.	N	X(0)	Y(0)	Z(0)	H
LOR01	120	12	9	2	0.158
LOR03	120	0.1	0.1	2	0.080

### 5.2.3. Hénon

El mapa de Hénon es un sistema dinámico discreto en el tiempo. Es uno de los ejemplos de sistemas dinámicos más estudiado que muestra comportamiento caótico. El mapa de Hénon toma un punto  $(x_n, y_n)$  en el plano y lo mapea a un punto nuevo según

$$\begin{cases} x_{n+1} = 1 - ax_n^2 = 1 - ax_n^2 \\ y_{n+1} = bx_n^2 \end{cases} \quad (5-3)$$

Las ecuaciones fueron introducidas por primera vez por Hénon y fue sugerida en el artículo [268]. Sin embargo, presenta muchos aspectos del comportamiento dinámico de los sistemas caóticos más complicados [269].

La representación de dos series caóticas se da en la Fig. 5-3, que se obtuvieron mediante la observación de la coordenada x del mapa Hénon, donde se usó la función de MATLAB *henon* ( $n$ ,

$level, a, b, x_0, y_0$ ) como muestra en la Fig. 5-3 (a) y (b) tanto para series cortas,  $n=120$  muestras como para series largas  $n=120$  muestras.

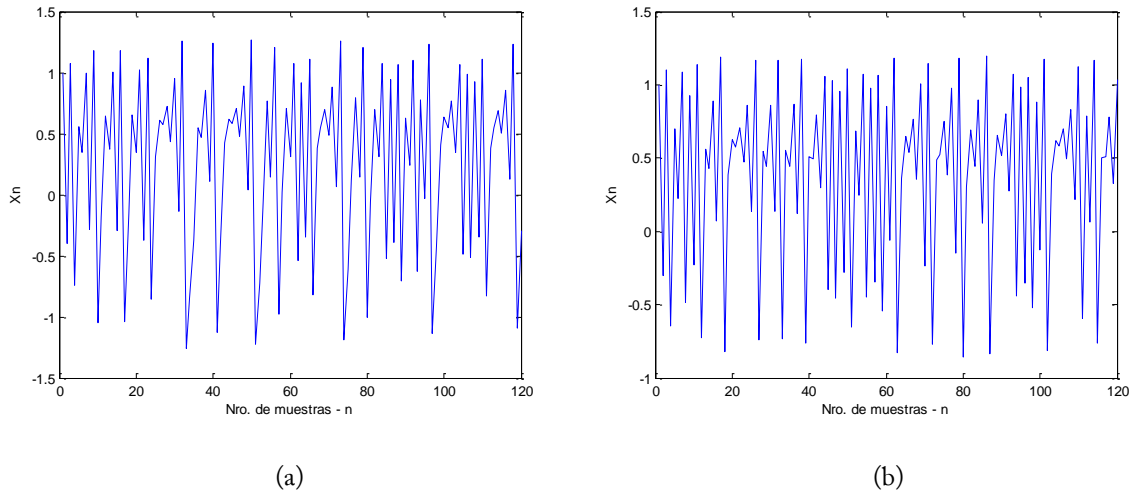


Fig. 5-3. Muestras de series temporales caóticas  $\{x_n\}$  derivado del mapa Hénon; a) con parámetro  $a=1.4$  y  $b=0.3$ ; b) con parámetro  $a=1.3$  y  $b=0.22$ .

Tabla 5-4. Parámetros para generar Series HEN

Series No.	N	a	b	$X_0$	$Y_0$	H
HEN01	120	1.4	0.3	0	0	0.187
HEN02	120	1.3	0.22	0	0	0.833

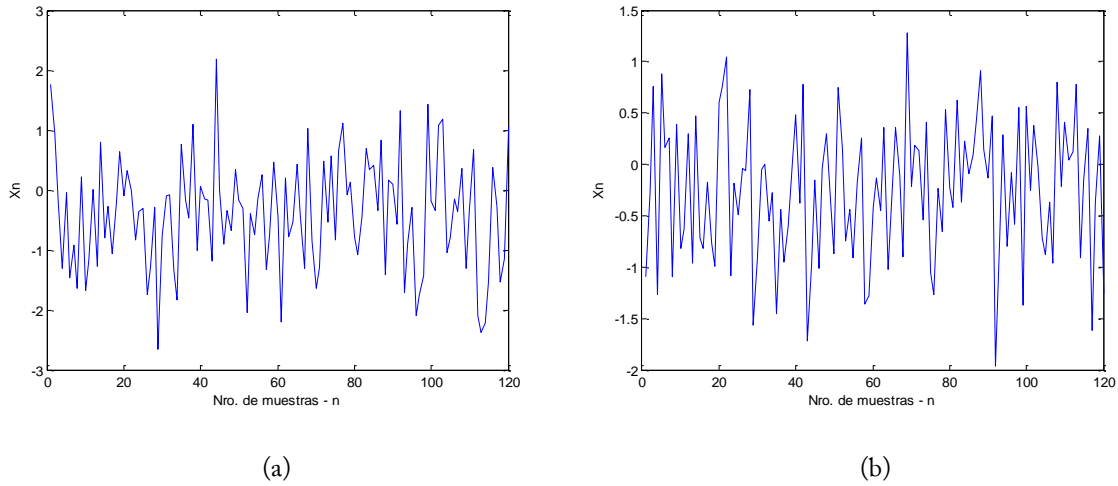
#### 5.2.4. Ikeda

Antes de describir la reconstrucción, presento el sistema que se utilizará para generar la mayor parte de las series de tiempo se describe en el presente documento, a saber, el mapa Ikeda [270] se da como sigue:

$$\begin{aligned} \dot{x} &= 1 - \mu[x \cos(t) - y \sin(t)], \\ \dot{y} &= \mu[x \cos(t) + y \sin(t)] \end{aligned} \quad (5-4)$$

donde  $t = 1/(1+x^2+y^2)$ .

Este sistema muestra un comportamiento caótico en un rango de valores para el parámetro  $\mu$  incluyendo los valores elegidos aquí. Para generar las series se usó la función de MATLAB  $ikeda(n, level, \mu, x_0, y_0)$  como muestra la Fig. 5-4 (a) y (b).



**Fig. 5-4.** Muestras de series temporales caóticas  $\{x_n\}$  derivado del mapa Ikeda; a) con parámetro  $\mu = 0.9$  y condiciones iniciales  $x(0)=0.5$  ,  $y(0)=0.5$ .; b) con parámetro  $\mu = 0.8$  y condiciones iniciales  $x(0)=0.5$  ,  $y(0)=0.5$ .

**Tabla 5-5.** Parámetros para generar Series IK

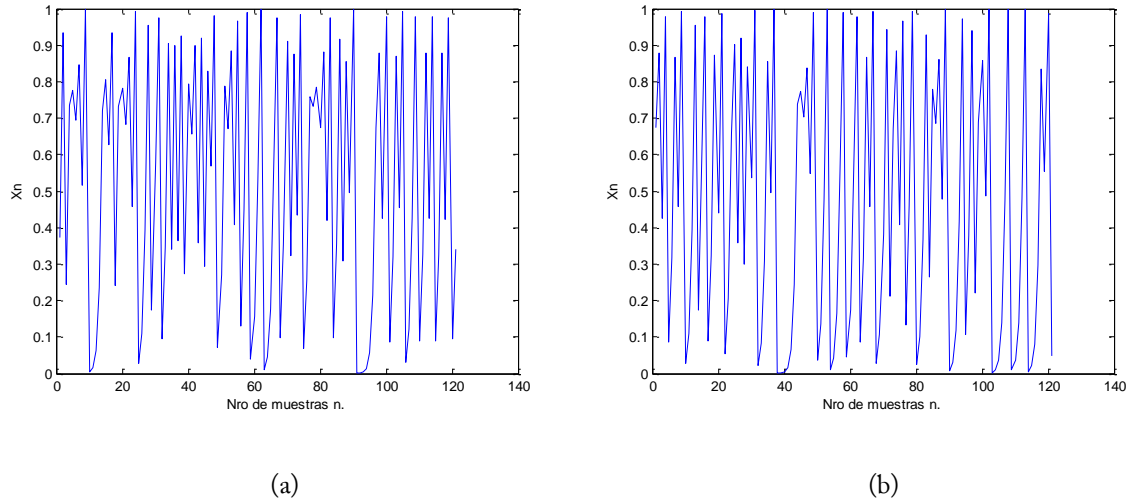
Series No.	n	$\mu$	X(0)	Y(0)	H
IK01	120	0.9	0.5	0.5	0.029
IK03	120	0.8	0.9	0.6	0.060

### 5.2.5. Mapeo Logístico

La serie logística (LOG) se define por

$$x(t+1) = ax(t)[1 - x(t)] \tag{5-5}$$

Cuando  $a=4$ , las iteraciones de la Ecuación (5-5) realizan una serie temporal caótica [271].



**Fig. 5-5.** Muestras de series temporales caóticas  $\{x_n\}$  derivado del mapa Logístico; a) serie LOG01 con parámetro  $a=0.1$  y condiciones iniciales  $x(0)=0.1$ ; b) serie LOG02 con parámetro  $a=0.3$  y condiciones iniciales  $x(0)=0.1$ .

**Tabla 5-6.** Parámetros para generar Series LOG largas

Series No.	N	a	$X_0$	H
LOG01	120	4	0.1	0.151
LOG03	120	4	0.3	0.100

**Tabla 5-7.** Parámetros para generar Series LOG cortas

Series No.	N	a	$X_0$	H
LOG03	50	4	0.1	0.14
LOG04	50	4	0.3	0.3

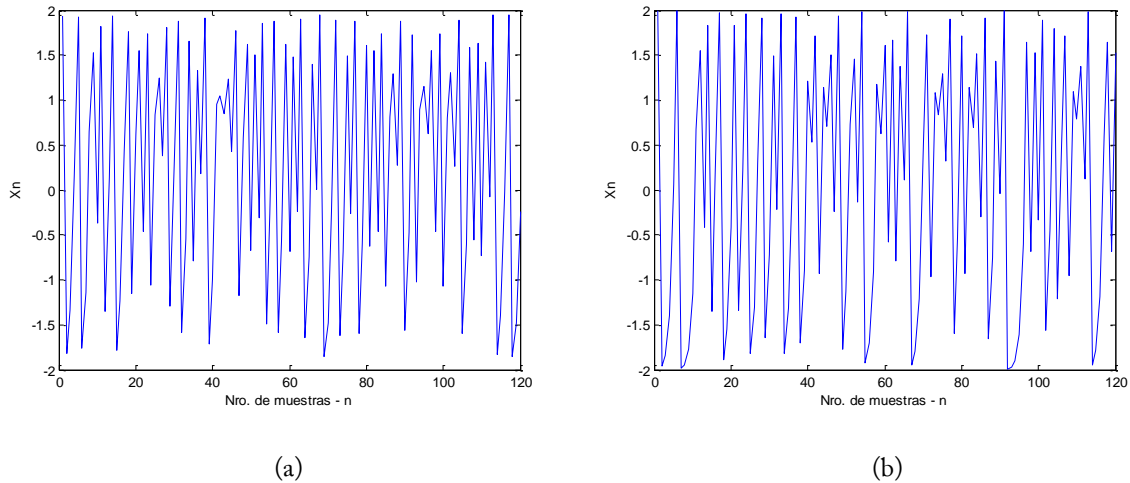
### 5.2.6. Mapeo Cuadrático

El mapa cuadrático se define por la ecuación

$$x(t+1) = \mu x(t)(1-x(t)). \tag{5-6}$$

Esta asignación se itera por  $\mu = 4$ , a partir de un número aleatorio en el intervalo entre 0 y 1, a continuación, se obtiene un comportamiento dramáticamente diferente dependiendo del valor inicial de  $x$ . Los valores iniciales de  $x$  que son muy próximos entre sí pueden tener drásticamente

diferentes iteraciones. Esta imprevisibilidad o dependencia sensible de las condiciones iniciales es una propiedad familiar que muestra un comportamiento caótico [272] en un rango de valores para el parámetro que incluye los valores elegidos aquí. Para generar las series se usó la función de MATLAB *quadratic* ( $n, level, c, x0$ ) como muestra la Fig. 5-6 (a) y (b).



**Fig. 5-6.** Muestras de series temporales caóticas  $\{x_n\}$  derivado del mapa cuadrático; a) con parámetro  $\mu = 1.95$  y condiciones iniciales  $x(0)=0.1$ ; b) con parámetro  $\mu = 2$  y condiciones iniciales  $x(0)=0.1$ .

**Tabla 5-8.** Parámetros para generar Series QUA

Series No.	N	c	X(0)	H
QUA01	120	1.95	0.5	0.31
QUA03	120	0.8	0.9	0.011

Las series anteriores se introdujeron no solo porque ilustran la complejidad del comportamiento que puede ser fácilmente producido por ecuaciones simples con elementos no lineales, sino también porque posteriormente se utilizan como series de referencia.

### 5.3. Resultados experimentales en series caóticas artificiales

Para mostrar los resultados de las predicciones con los modelos propuestos, 120 muestras se seleccionaron para las series temporales, los primeros 102 valores se utilizan para el entrenamiento y los 18 valores restantes se mantienen para los datos de validación y prueba.

El comportamiento a largo plazo cambia completamente debido a las condiciones iniciales para obtener la dependencia estocástica de la serie temporal determinista según su rugosidad evaluada por el parámetro  $H$ . A continuación, los 18 datos extras de prueba se utilizan para medir el rendimiento de predicción.

Los resultados de la simulación y la performance de predicción de la comparación entre los algoritmos propuestos se miden por los índices SMAPE y RMSE propuesto en la mayoría de las evaluaciones métricas [11] que se muestran en la Tabla 5-9.

**Tabla 5-9.** Comparación de los métodos de pronóstico a 18 pasos adelante

Series	Método	SMAPE	RMSE
MG17-1	BEMA	0.021	0.016
	BEA	0.045	0.032
	BA	0.085	0.036
	NN-Mod.	0.32	0.047
	EAS	0.12	0.90
	SUB <sub>2-3-6</sub>	0.050	0.388
	NN-ARMA	4.58	0.84
MG30-1	BEMA	0.013	0.015
	BEA	0.027	0.092
	BA	0.11	0.017
	NN-Mod.	0.32	0.038
	EAS	0.11	0.89
	SUB <sub>2-3-6</sub>	0.29	0.18
	NN-ARMA	0.14	0.042
LOR01	BEMA	0.24	1.50
	BEA	0.12	7.61
	BA	1.42	10.01
	NN-Mod.	5.06	32.41
	EAS	14.12	102.48
	SUB <sub>2-3-6</sub>	5.19	37.26
	NN-ARMA	11.94	45.30

Capítulo 5. Resultados Experimentales

LOR02	BEMA	1.26	18.81
	BEA	2.84	18.30
	BA	3.15	20.60
	NN-Mod.	4.95	30.08
	EAS	14.12	98.10
	SUB <sub>2-3-6</sub>	3.32	19.74
	NN-ARMA	19.75	120.36
HEN01	BEMA	0.0007	0.005
	BEA	0.002	0.012
	BA	0.0025	0.014
	NN-Mod.	0.018	0.117
	EAS	0.044	0.239
	SUB <sub>2-3-6</sub>	0.009	0.063
	NN-ARMA	0.047	0.326
HEN02	BEMA	0.0005	0.003
	BEA	0.0001	0.007
	BA	0.0019	0.012
	NN-Mod.	0.009	0.061
	EAS	0.019	0.131
	SUB <sub>2-3-6</sub>	0.011	0.065
	NN-ARMA	0.061	0.103
IK01	BEMA	0.078	0.53
	BEA	0.017	0.125
	BA	0.105	0.29
	NN-Mod.	0.136	0.99
	EAS	1.04	5.80
	SUB <sub>2-3-6</sub>	0.11	0.65
	NN-ARMA	3.86	9.16
IK02	BEMA	0.067	0.53
	BEA	0.016	0.101
	BA	0.021	0.121
	NN-Mod.	0.159	0.93
	EAS	0.52	2.77
	SUB <sub>2-3-6</sub>	0.091	0.65

*Capítulo 5. Resultados Experimentales*

---

	NN-ARMA	3.33	8.52
LOG01	BEMA	0.056	0.018
	BEA	0.059	0.017
	BA	0.12	0.043
	NN-Mod.	0.35	0.063
	EAS	0.14	0.91
	SUB <sub>2-3-6</sub>	0.082	0.460
	NN-ARMA	6.88	12.68
LOG02	BEMA	0.04	0.025
	BEA	0.07	0.034
	BA	0.12	0.057
	NN-Mod.	0.40	0.072
	EAS	0.46	2.42
	SUB <sub>2-3-6</sub>	0.51	0.31
	NN-ARMA	1.42	6.35
QUA01	BEMA	0.043	0.07
	BEA	0.022	0.128
	BA	0.027	0.169
	NN-Mod.	0.22	1.32
	EAS	0.96	5.34
	SUB <sub>2-3-6</sub>	0.39	2.29
	NN-ARMA	6.42	15.65
QUA02	BEMA	0.012	0.08
	BEA	0.031	0.186
	BA	0.034	0.231
	NN-Mod.	0.27	1.51
	EAS	0.78	4.73
	SUB <sub>2-3</sub>	0.49	2.96
	NN-ARMA	6.29	18.32



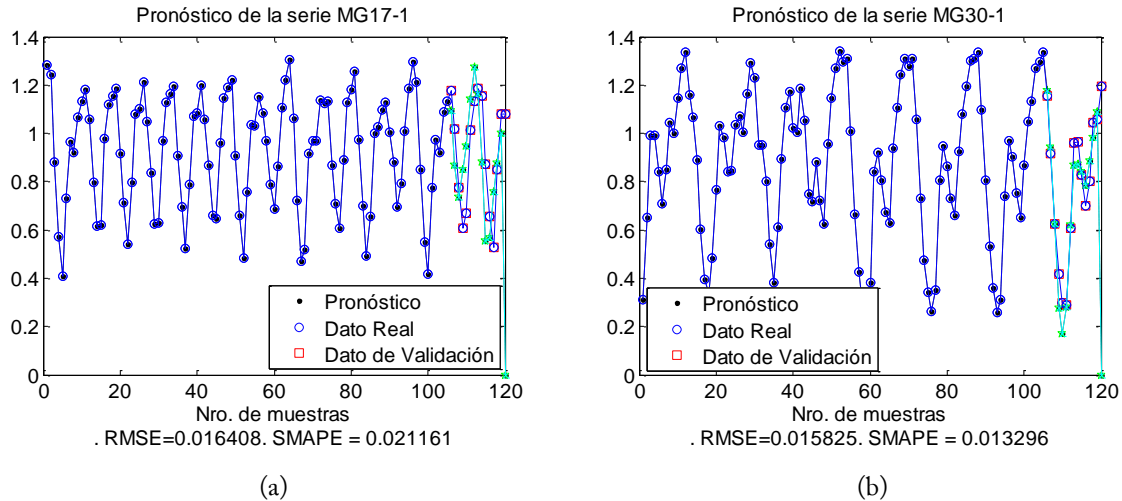


Fig. 5-7. Pronósticos a 18 pasos de la serie MG usando el modelo BEMA; a) MG17-1; b) MG30-1.

La similitud de la tendencia de predicción de rendimiento entre ellos es clara, BEMA es mejor sobre todo en las series temporales de alta rugosidad, con referencia a BEA y enfoques BA. Los resultados de la Tabla 5-9 muestran que los métodos basados en aproximación Bayesiana BEMA y BEA son superiores para series temporales de alta rugosidad, con un SMAPE y RMSE aproximadamente de 10 veces respecto de sus competidores EAS, SUB, y NN-Mod. Los resultados muestran que los desempeños de los predictores BEMA, BEA y BA son mejores que los basados en suavización de la serie en término de SMAPE y RMSE, debido a la existencia de los valores atípicos. Con serie largas, BEMA y BEA pueden detectar adecuadamente la relación subyacente entre las variables correlacionadas. Estos resultados se ven en las figuras Fig. 5-8 y Fig. 5-9.

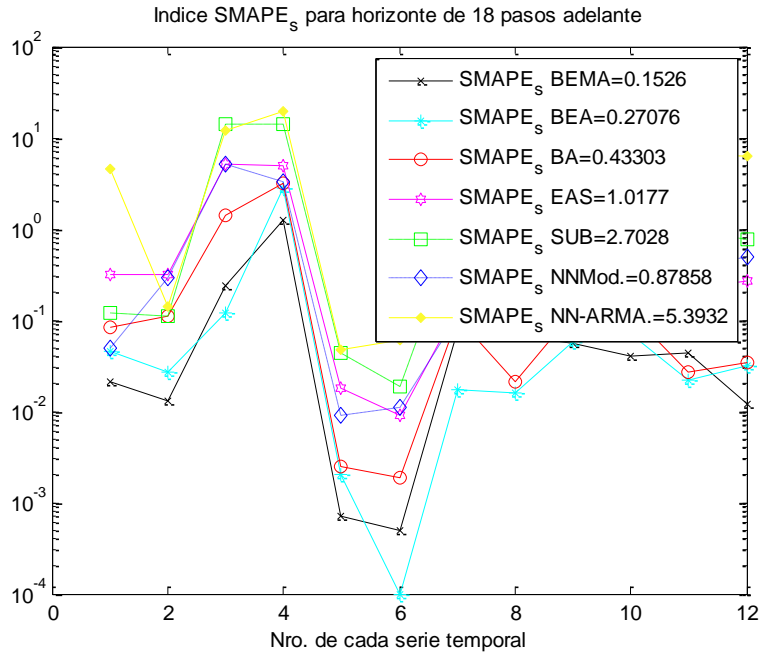


Fig. 5-8. - Medición del rendimiento a través del índice SMAPE para 18 valores pronosticados fuera de la muestra.

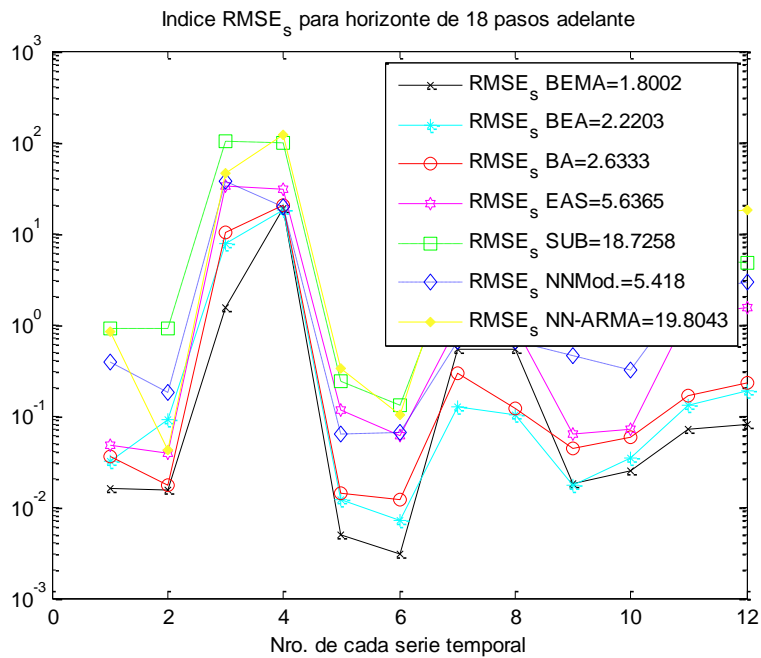


Fig. 5-9. Medición del rendimiento a través del índice RMSE para 18 valores pronosticados fuera de la muestra.

Este estudio, la discusión de cómo las RNA FFNN pueden aproximar con éxito los cambios cuantitativos en la dinámica de los datos de series de tiempo debido a cambios en los valores de los parámetros de las variables exógenas permanece abierta para el estudio, principalmente.

Aunque la comparación se realizó sobre filtros basados en RNA, los resultados experimentales confirman que los métodos basados en aproximación Bayesiana mejorada (BEMA y BEA) pueden predecir series temporales caóticas de manera más eficaz en términos de SMAPE y los índices de RMSE cuando se compara con otros métodos de pronóstico existentes en la literatura [273] [274] [275] [276] [277] [278]. Sin embargo, el deseo de conservar las dependencias estocásticas restringe todos los horizontes de predicción con la misma estructura del modelo, ya sea en su configuración y condiciones iniciales. Dado que esta restricción podría reducir la flexibilidad del enfoque de pronóstico, la variante del método BEA, BEMA explota el uso de la información compartida con las observaciones pasadas- por ejemplo, la información mutua entre el pasado y presente- y la información en el presente que no está contenido en el pasado (la entropía condicional de la presente dado el pasado) para los que los valores son óptimos para el tamaño de ventana y el rango del parámetro  $\alpha$  de la entropía de permutación de Renyi.

## 5.4. Resultados experimentales en series cortas con incompletitud de datos

### 5.4.1. Introducción

A pesar de los avances en las técnicas de imputación para falta de datos en las últimas tres décadas, el problema de la falta de datos permanece en gran parte sin resolver [279]. El problema de los datos incompletos plantea una dificultad para el análisis de series temporales y los procesos de toma de decisiones [280], que dependen de estos datos, que requieren métodos de estimación más precisa y eficiente para los sistemas de predicción [281]. Existen diversas técnicas como una solución a este problema, que van desde la eliminación de datos [282] hasta los métodos que emplean técnicas estadísticas [283] [284] e inteligencia artificial [285] [286] para imputar para las variables que faltan.

Kidson *et. al.* [287] evaluaron el impacto de los datos faltantes en las estadísticas generales de circulación al disminuir sistemáticamente la cantidad de datos de entrenamiento disponibles. Se

determinó que el RMSE en la media mensual respecto de la desviación estándar diaria, era de dos a tres veces mayor cuando los datos que faltan estaban espaciados al azar en comparación con el espacio por igual, y el RMSE aumenta hasta un factor de dos cuando los datos perdidos ocurrían en un bloque. En [288] se encontró que en vecindades altamente correlacionadas pueden ser usadas para interpolar los datos que faltan en los conjuntos de datos de temperatura canadienses.

Sin embargo, algunos métodos, como técnica de media suavizado [40] pueden simplificar la probabilidad de producir estimaciones sesgadas o hacer suposiciones acerca de los datos que puede no ser cierto y se puede utilizar como un buen estimador de la calidad en la toma de decisiones sobre la base de estos datos. La sustitución de los datos que faltan dentro de las series temporales dentro del rango de datos conocido es crucial [351] para diseños más precisos y también en la evaluación de desempeño. En la literatura estadística, los datos faltantes son generalmente imputados utilizando estimadores de máxima verosimilitud que corresponden a un modelo subyacente específico. Muy a menudo, estos estimadores no se calculan de manera eficiente, lo que motiva el uso de algoritmos de máxima expectación (EM) [289].

Más recientemente, se han propuesto métodos de minería de datos basados en Inteligencia Computacional [290] para la predicción de series temporales a corto plazo (ST-TSF). Ejemplos aplicados a ST-TSF incluyen: RNA [291] [292] [293] [294] [295], la computación evolutiva [296], máquinas de soporte vectorial (SVM) [297], usando Lógica Fuzzy (FL) [298], o sus combinaciones [299] [300] [301] [302].

La motivación de este problema surge de la predicción con información incompleta y faltante [40], que es aplicable a una amplia clase de algoritmos de aprendizaje [303] [304] que incluye a las RNA. Una gran ventaja de la solución propuesta es que la complejidad no aumenta con un número creciente de entradas que faltan. Estas soluciones pueden generalizarse al problema de datos inciertos (ruidosos) [305]. La estimación de datos incompletos en elementos del vector para aplicaciones de procesamiento en tiempo real requiere de un sistema que posea el conocimiento de determinadas características, tales como las correlaciones entre las variables, que son inherentes en el espacio de entrada [306]. El benchmark de series elegidas para la experimentación se eligen de algunas provistas en la Sección 5.2, MG17-2 y MG30-2, LOG03 y LOG04, de series incompleta de lluvia provenientes de la estancia La Sevillana (SEV), cuyo pronóstico es simulado por el método Monte Carlo empleando RNAs.

### 5.4.2. Mecanismos de pérdidas de datos

Little y Rubin [279] distinguen entre los tres mecanismos de datos faltantes: falta al azar, falta por completo al azar y no perdidos al azar. En esta tesis, se toma la perdida no al azar que implica que el mecanismo de los datos que faltan se relaciona con los valores que faltan. También se conoce como caso no ignorado [307] [308] mientras la observación que falta depende del resultado de interés. Por lo tanto, la metodología de esta investigación sigue el mecanismo por perdida no al azar para contribuir con una técnica basada en suavizado promedio para la falta de imputación de datos.

### 5.4.3. Método de suavización promedio (MSP)

El principal problema al pronosticar una serie temporal es cómo recuperar el máximo de información de los datos disponibles. Con el fin de predecir un paso adelante, se asume el enfoque suavizado promedio propuesto para completar los datos faltantes. Se propone para llenar estos valores vacíos, el dividir el conjunto de datos en subconjuntos tomados de a 12 valores. A continuación, una matriz se forma por 12 columnas y el número de las filas dependerá del tamaño de conjunto de datos. Para completar la información que falta, se utiliza la técnica de suavizado promedio de la siguiente manera: tres conjuntos de datos se construyen, el primer conjunto es el original que contiene datos faltantes llamados  $x$  marcados en color rojo. El segundo conjunto se completa con ceros, y el tercero utilizando la fila de arriba y de abajo a fin de que promedia los datos de la fila anterior y posterior como se muestra en la Fig. 5-12. La misma analogía se utiliza para construir las series descritas más adelante.

$x_n$	1	2	3	4	5	6	7	8	9	10	11	12
	$x_1$	$x_2$	<b>x</b>	<b>x</b>	<b>x</b>	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$	$x_{11}$	$x_{12}$
	$x_{13}$	$x_{14}$	$x_{15}$	$x_{16}$	$x_{17}$	<b>x</b>	<b>x</b>	<b>x</b>	$x_{21}$	$x_{22}$	$x_{23}$	$x_{24}$
	$x_{25}$	$x_{26}$	$x_{27}$	$x_{28}$	$x_{29}$	$x_{30}$	$x_{31}$	$x_{32}$	$x_{33}$	$x_{34}$	$x_{35}$	$x_{36}$
	$x_{37}$	$x_{38}$	$x_{39}$	$x_{40}$	$x_{41}$	$x_{42}$	$x_{43}$	$x_{44}$	$x_{45}$	$x_{46}$	$x_{47}$	$x_{48}$
	$x_{49}$	$x_{50}$	..	..	..	..	..	..	..	..	..	..

Fig. 5-10. Método de suavizado promedio: los datos faltantes se marcan en rojo.

$x_n$	1	2	3	4	5	6	7	8	9	10	11	12
	$x_1$	$x_2$	0	0	0	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$	$x_{11}$	$x_{12}$
	$x_{13}$	$x_{14}$	$x_{15}$	$x_{16}$	$x_{17}$	0	0	0	$x_{21}$	$x_{22}$	$x_{23}$	$x_{24}$
	$x_{25}$	$x_{26}$	$x_{27}$	$x_{28}$	$x_{29}$	$x_{30}$	$x_{31}$	$x_{32}$	$x_{33}$	$x_{34}$	$x_{35}$	$x_{36}$
	$x_{37}$	$x_{38}$	$x_{39}$	$x_{40}$	$x_{41}$	$x_{42}$	$x_{43}$	$x_{44}$	$x_{45}$	$x_{46}$	$x_{47}$	$x_{48}$
	$x_{49}$	$x_{50}$	..	..	..	..	..	..	..	..	..	..

Fig. 5-11. Método de suavizado promedio: llenado de datos faltantes con ceros.

$x_n$	1	2	3	4	5	6	7	8	9	10	11	12
	$x_1$	$x_2$	$x_{15}/2$	$x_6/2$	$x_{17}/2$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$	$x_{11}$	$x_{12}$
	$x_{13}$	$x_{14}$	$x_{15}$	$x_{16}$	$x_{17}$	$(x_6 + x_{30})/2$	$(x_7 + x_{31})/2$	$(x_8 + x_{32})/2$	$x_{21}$	$x_{22}$	$x_{23}$	$x_{24}$
	$x_{25}$	$x_{26}$	$x_{27}$	$x_{28}$	$x_{29}$	$x_{30}$	$x_{31}$	$x_{32}$	$x_{33}$	$x_{34}$	$x_{35}$	$x_{36}$
	$x_{37}$	$x_{38}$	$x_{39}$	$x_{40}$	$x_{41}$	$x_{42}$	$x_{43}$	$x_{44}$	$x_{45}$	$x_{46}$	$x_{47}$	$x_{48}$
	$x_{49}$	$x_{50}$	..	..	..	..	..	..	..	..	..	..

Fig. 5-12. Método de suavizado promedio: se completa promediando el dato de la fila anterior y posterior.

#### 5.4.4. Performance y errores de predicción

En este apartado, el objetivo del experimento es la comparación para evaluar la performance de predicción de los modelos predictivos propuestos para el pronóstico de series temporales con datos faltantes de alta rugosidad. Estos experimentos se han realizado en función de cada tipo de series de datos y la misma metodología utilizada en la Sección 5.4.3, donde el método MSP se aplica a todas las series con la diferencia de que los últimos 15 datos observados se reservaron como criterio para ser usados como validación para las series MG, LOG y la serie de lluvia acumulada mensual mostrada en la Fig. 5-13. Cinco series temporales se utilizan en total para comparar el rendimiento de los predictores propuestos. Las series MG17-2, MG30-2, LOG03 y LOG04 constan de 50 muestras cada una, los primeros 35 valores se utilizan para el entrenamiento y los 15 valores restantes se mantienen para los datos de validación y prueba. Para la serie de lluvia SEV de la Fig. 5-13 que consta de 79 valores, los últimos 15 datos son tomados como validación para el

pronóstico. El método MSP se aplica a la serie SEV como se muestra en las Tabla 5-10 y Tabla 5-11.

El comportamiento de las series a corto plazo cambia completamente por cambio en las condiciones iniciales para obtener la dependencia estocástica de la serie temporal determinista según su rugosidad evaluada por el parámetro H.

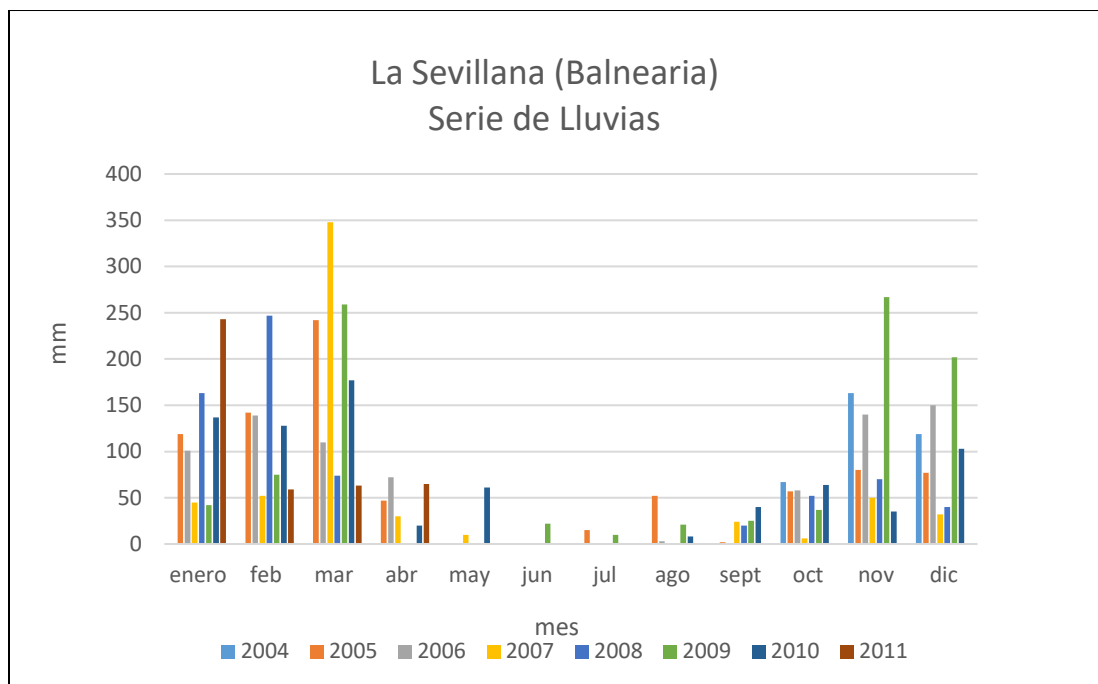


Fig. 5-13. Lluvia mensual acumulada del Establecimiento La Sevillana (SEV).

Tabla 5-10. Método de suavizado promedio aplicado a la serie SEV: se completa promediando el dato de la fila anterior y posterior.

$x_t$	En.	Feb.	Mar.	Ab.	May.	Jun.	Jul.	Ag.	Sep.	Oct.	Nov.	Dic.
2004										67	163	119
2005	119	142	242	47	0	0	15	52	2	57	80	77
2006	101	139	110	72	0	0	0	3	0	58	140	150
2007	45	52	348	30	10	0	0	0	24	6	50	32
2008	163	247	74	x	x	x	x	x	20	52	70	40
2009	42	75	259	0	0	22	10	21	25	37	267	202
2010	137	128	177	20	61	0	0	8	40	64	35	103
2011	243	59	63	65								

**Tabla 5-11.** Método de suavizado promedio aplicado a la serie SEV: se completa promediando el dato de la fila anterior y posterior.

$x_t$	En.	Feb.	Mar.	Ab.	May.	Jun.	Jul.	Ag.	Sep.	Oct.	Nov.	Dic.
2004										67	163	119
2005	119	142	242	47	0	0	15	52	2	57	80	77
2006	101	139	110	72	0	0	0	3	0	58	140	150
2007	45	52	348	30	10	0	0	0	24	6	50	32
2008	163	247	74	15	5	11	5	11	20	52	70	40
2009	42	75	259	0	0	22	10	21	25	37	267	202
2010	137	128	177	20	61	0	0	8	40	64	35	103
2011	243	59	63	65								

La Tabla 5-12 muestra la cantidad de datos disponibles en el conjunto total. Para el experimento, se extrajeron datos de las series MG y LOG para formar las series MG17-2, MG30-2, LOG03 y LOG04 usando el método MSP. Los resultados de la comparación para cada tipo de serie se describen en la Tabla 5-13.

El último de ellos se utiliza para comparar si el pronóstico es aceptable o no, en el que los últimos 15 observaciones se pueden utilizar para validar el funcionamiento del sistema de predicción.

**Tabla 5-12.** Porcentaje de datos disponibles en las series temporales cortas

Serie incompletas	Porcentaje de disponibilidad de los datos
MG17-2	91%
MG30-2	90%
LOG03	90%
LOG04	87%
SEV	93%

**Tabla 5-13.** Performance y resultados de comparación para las series temporales cortas

Series	Método	SMAPE		RMSE	
		Serie Completa	Serie Incompleta	Serie Completa	Serie Incompleta
MG17-2	BEMA	0.036	0.043	0.22	0.52
	BEA	0.036	0.056	0.23	0.66
	BA	0.056	0.085	0.33	0.59
	EAS	0.054	0.148	0.43	0.56
	NN-Mod.	0.090	0.140	0.33	0.59
MG30-2	BEMA	0.036	0.046	0.21	0.42



	BEA	0.038	0.059	0.23	0.55
	BA	0.045	0.062	0.32	0.69
	EAS	0.059	0.146	0.68	0.78
	NN-Mod.	0.033	0.148	0.30	0.54
LOG03	BEMA	0.202	0.430	0.309	0.870
	BEA	0.210	0.445	0.325	0.898
	BA	0.523	0.721	0.632	1.36
	EAS	0.202	0.430	0.82	0.112
	NN-Mod.	0.962	1.02	0.35	0.95
LOG04	BEMA	0.548	0.764	0.647	0.98
	BEA	0.560	0.789	0.765	1.03
	BA	1.63	0.904	1.98	2.32
	EAS	0.548	0.764	1.15	2.45
	NN-Mod.	0.905	1.11	0.77	1.32
SEV	BEMA	4.26	5.92	29.95	35.48
	BEA	5.17	6.08	34.88	39.16
	BA	7.98	8.17	44.68	53.20
	EAS	7.31	11.63	25.42	29.74
	NN-Mod.	5.72	9.48	42.84	58.50

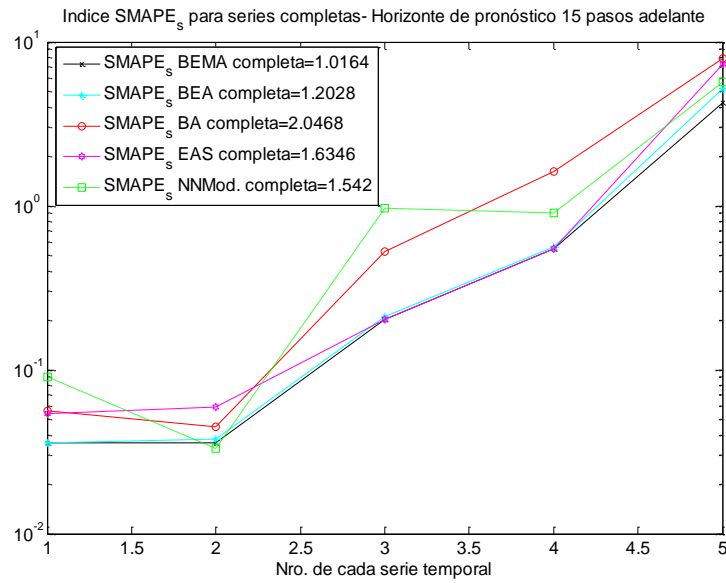


Fig. 5-14. Índice SMAPE para series completas con horizonte de predicción 15 pasos adelante.

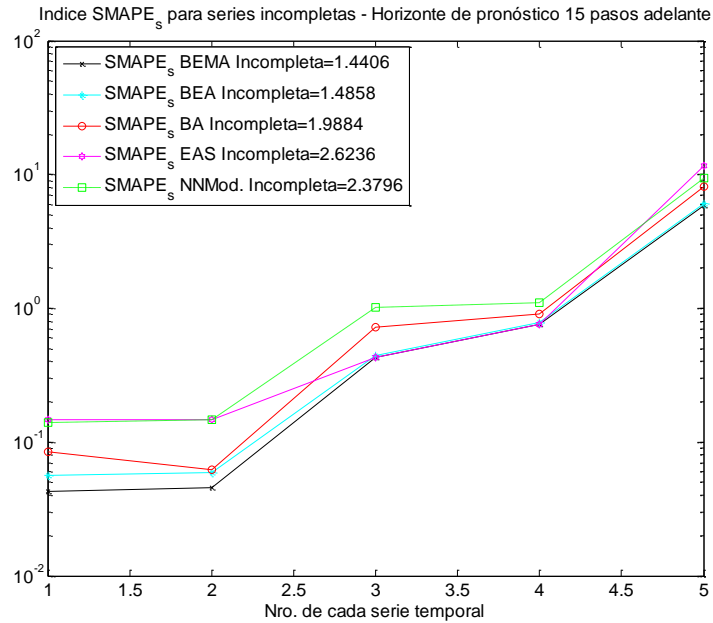


Fig. 5-15. Índice SMAPE para series incompletas con horizonte de predicción 15 pasos adelante.

Para la selección de los parámetros del filtro predictor BEMA, se tomó como base las siguientes consideraciones; se seleccionaron aquellos conjuntos de datos con el menor porcentaje de puntos que faltan y los datos que faltan para aumentar las posibilidades de conservar más patrones que se muestran en la Tabla 5-12. Las variables dependientes correlacionadas pueden estar incluidas en el conjunto de datos para la búsqueda de patrones ordinales. Sin embargo, no es redundante incluir esas variables cuando la relación entre las variables correlacionadas no preserva el orden. Por lo tanto, se seleccionó el rango para el parámetro  $\alpha$  de la Renyi entropía de 0,1 a 2,0 con un tamaño de paso de 0,01.

Los resultados de la Tabla 5-13 muestran que los métodos basados en aproximación Bayesiana (BEMA, BEA y BA) respecto de sus competidores EAS y NN-Mod. son superiores para las series temporales cortas completas e incompletas según el índice SMAPE como se muestran en las Fig. 5-14 y Fig. 5-15.

## 5.5. Resultados de predicción en series de lluvia mensual acumulada

En esta sección, las series de lluvia mensual de agua acumulada provenientes de distintos lugares de la provincia de Córdoba son estudiadas mediante diferentes arquitecturas de filtro

predictores basados en los modelos predictivos propuestos combinando pronósticos a corto, mediano y largo alcance. El rendimiento se mide por los índices SMAPE y RMSE tanto en el entrenamiento como en los distintos horizontes de pronóstico.

### **5.5.1. Datos meteorológicos**

Los datos de entrada de la RNA se obtuvieron de la estaciones pluviométricas provistas por distintos Ingenieros Agrónomos y agricultores medianos de campos tomados de estaciones que pertenecen de la Provincia de Córdoba, del Depto. Santa María, Despeñaderos y San Agustín, del Depto. Calamuchita (-31.914218, -64.310961), provistas por el Ing. Agrónomo Ernesto Carreño del establecimiento Santa Francisca (-31.8670, -64.3655), de Alta Gracia provistas por el Ing. Agrónomo Nicolás Bernáldez Brunt (-31.679583, -64.430465), de Balnearia provistas por la Ing. Agrónoma Eugenia Scalerandi del establecimiento “La Sevillana” (-31.016841, -62.643746), de San Bartolomé y Las Varillas (-31.869773, -62.749611) provistas por Mónica Piccardi y Rodrigo Lasa. El clima de la región de Córdoba, Argentina, se clasifica, según Köppen, como Clima húmedo subtrópico, 383 (Contar), Cwa con inviernos secos.

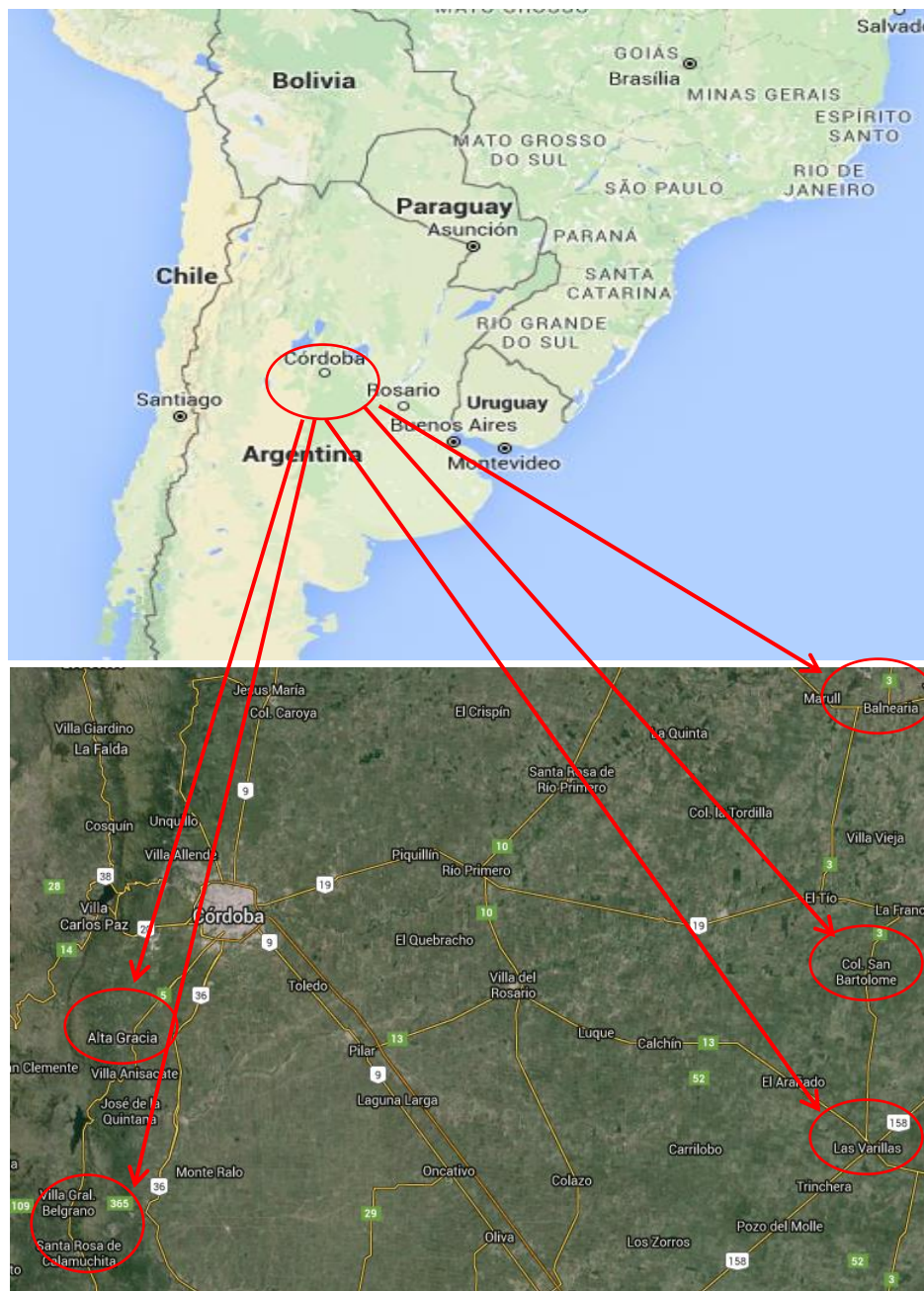


Fig. 5-16. Datos experimentales provenientes de distintas localidades de la Provincia de Córdoba.

### 5.5.2. Agregado de ruido Gaussiano fraccionario

El método Hosking es un algoritmo para calcular ruido gaussiano fraccional (FGN) [309]. Este método se basa en el hecho de que una muestra particular puede ser calculada por completo dado su pasado. En otras palabras, el método genera  $x_{n+1}$  dado de forma recursiva. A partir de este

resultado, a continuación, sólo tenemos que aplicar el algoritmo de forma recursiva hasta que producimos suficientes muestras.

El rendimiento del enfoque presentado se muestra mediante la predicción de los 6 valores futuros de cada serie temporal simulada por un Monte Carlo de 500 realizaciones con ruido gaussiano fraccional para especificar la varianza. El ruido fraccionario se ha generado por el método Hosking [57] con el parámetro  $H$  estimada a partir de la serie temporal de datos [54]. La media y la varianza de 500 ensayos de los valores pronosticados horizonte se muestran en cada serie temporal de lluvia. Estos resultados se dan para un (30%) y dos (69%) sigma como se muestran en la figuras Fig. 5-18, Fig. 5-19, Fig. 5-20 y Fig. 5-21 en la siguiente sección.

### 5.5.3. Configuración de los parámetros de los filtros predictores

Se ha dividido los datos en dos partes: el conjunto de entrenamiento y el conjunto de prueba. En la fase de entrenamiento, cada uno de los modelos predictores se entrenó con los parámetros de optimización dada por cada filtro. Esto significa que cada modelo se construye con sus valores óptimos de los parámetros respectivos.

El diseño de los métodos propuestos basados en RNAs consiste en el ajuste de los valores de parámetros. En el caso de MLP con una sola capa oculta y el algoritmo Levenberg-Marquard según los parámetros: número de entradas nodos, número de neuronas ocultas, y todo el conjunto de pesos sinápticos de conexión (tasa de aprendizaje, error mínimo para alcanzar la convergencia).

Los resultados computacionales en diferentes aproximaciones de orden y períodos de tiempo se presentan en la Tabla 5-14 con la adición de ruido gaussiano fraccional (FGN), las Tabla 5-15 y Tabla 5-16 muestran las métricas para 6-9-15-18 horizonte de pronóstico. El rendimiento de los modelos durante ajuste y validación se midieron por el porcentaje de error absoluto (SMAPE) y el error cuadrático medio (RMSE) propuesto en todas las evaluaciones de esta tesis. Las limitaciones de los algoritmos de aprendizaje propuestos residen en la resolución espacial de la serie temporal de lluvia que se suponen proviene de la realización de una traza del  $mBf$  medida por el parámetro  $H$ , en la que se necesita el uso del re-escalado a fin de poder ser usados por los algoritmos propuestos. Esta alternativa fue adoptada en las series temporales a fin de evitar preprocesamiento de la serie

temporal, que típicamente incluye dos etapas: identificación y tratamiento patrón, que es un trabajo intensivo y se limita a un pequeño número de serie.

### 5.5.4. Series de lluvia mensual acumulada

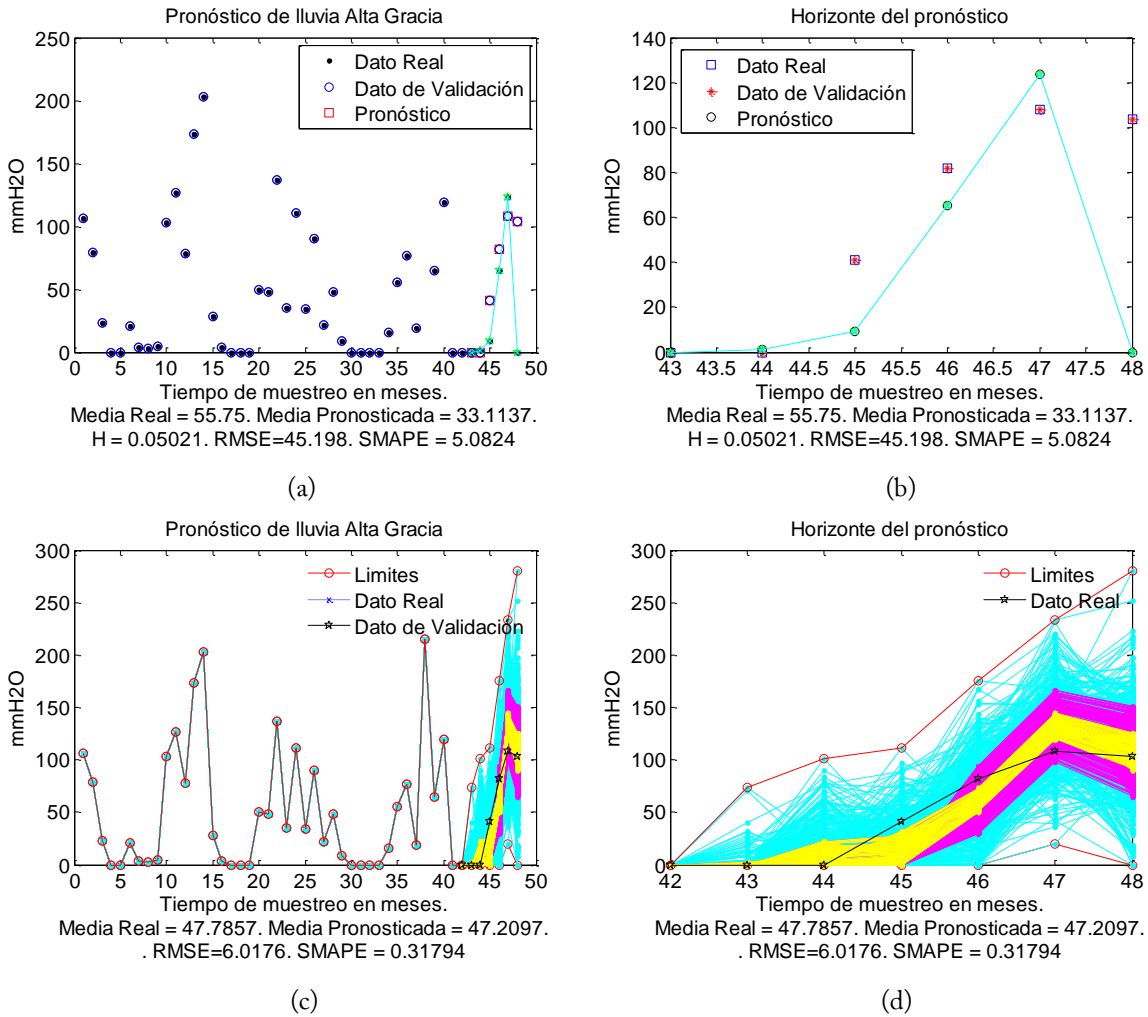
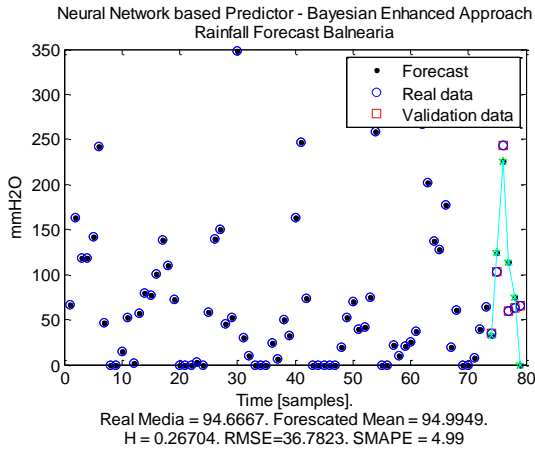
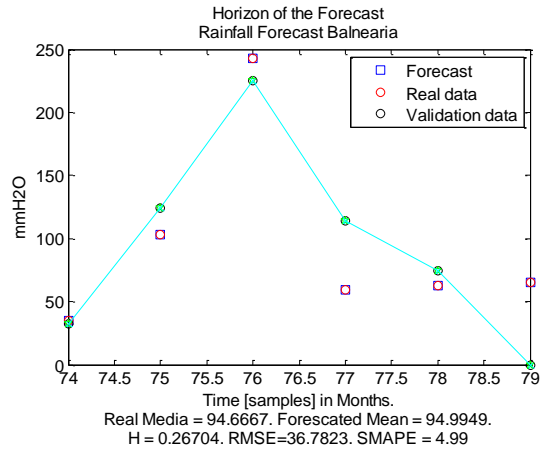


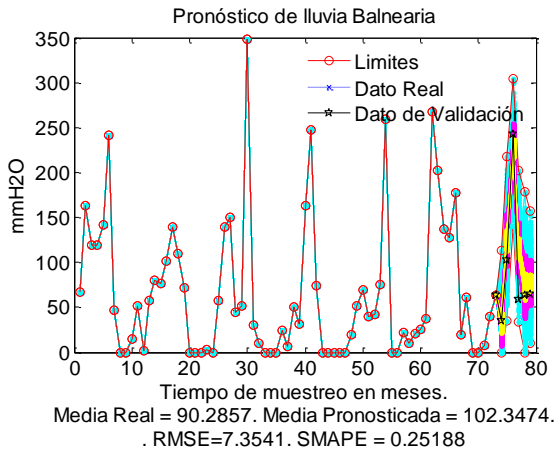
Fig. 5-17. Horizonte de pronóstico para la Serie AG a 6 meses; a) método BEA, b) BEA con ruido gaussiano fraccional.



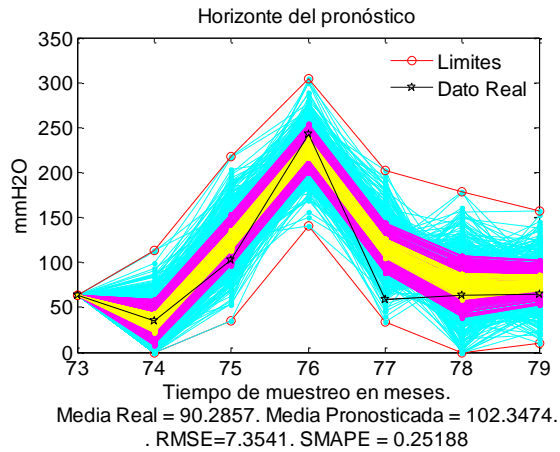
(a)



(b)

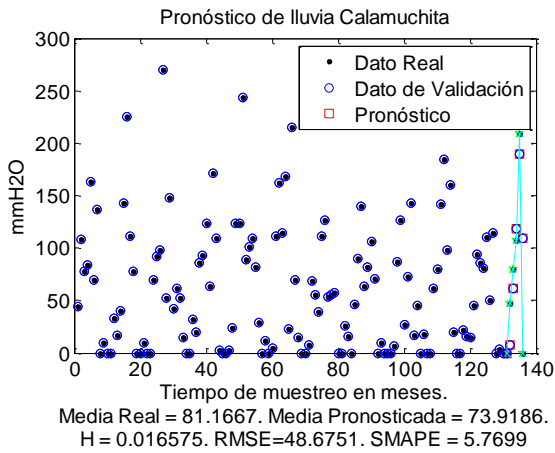


(c)

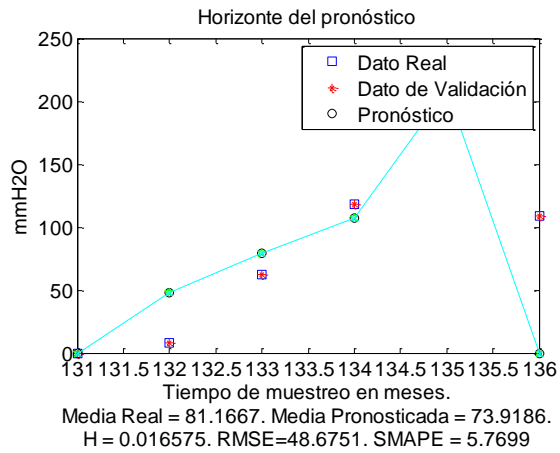


(d)

Fig 5-18. Horizonte de pronóstico para la Serie BA a 6 meses; a) método BEA, b) BEA con ruido gaussiano fraccional.



(a)



(b)

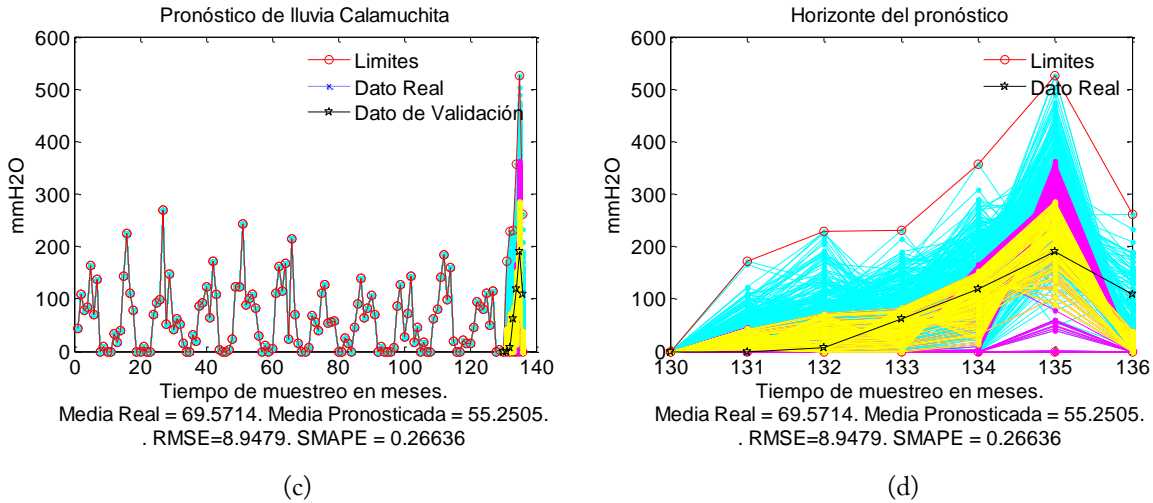


Fig. 5-19. Horizonte de pronóstico para la Serie CA a 6 meses; a) método BEA, b) BEA con ruido gaussiano fraccional

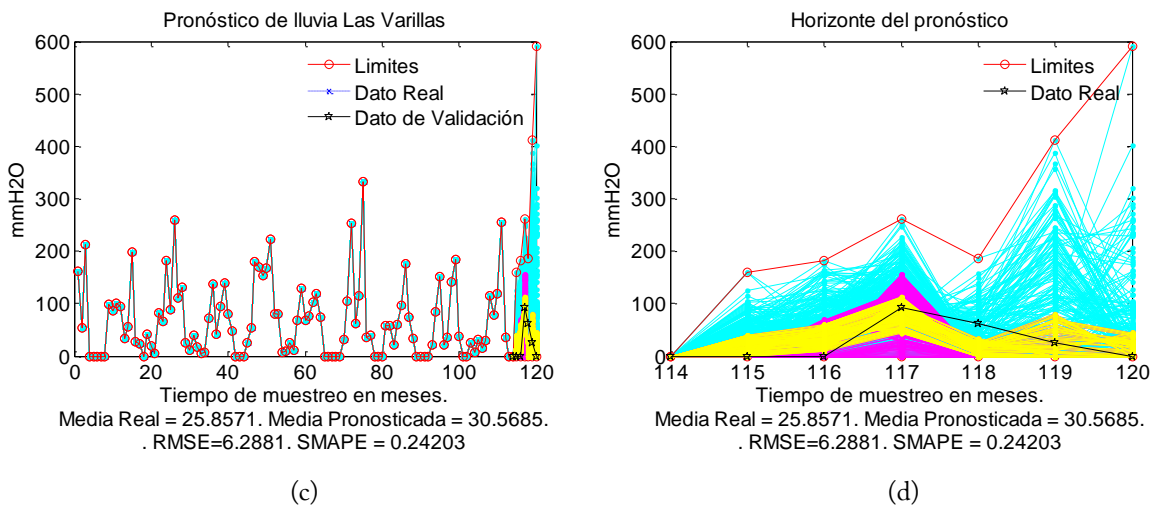
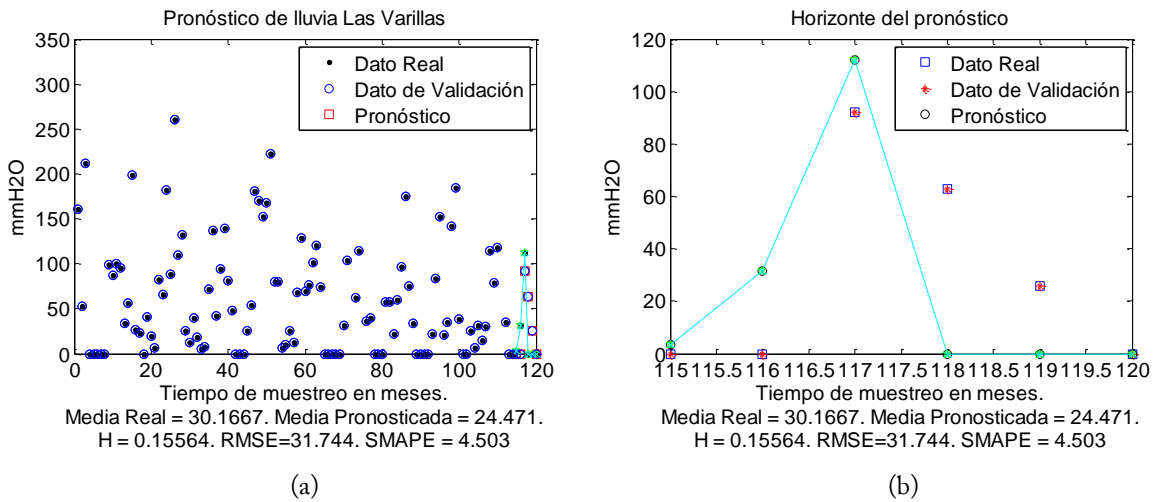


Fig. 5-20. Horizonte de pronóstico para la Serie LV a 6 meses; a) método BEA, b) BEA con ruido gaussiano fraccional.



La evaluación de los resultados obtenidos al comparar el rendimiento de los filtros propuestos muestra una mejora significativa medida por los índices SMAPE y RMSE de los métodos basados en aproximación Bayesiana sobre todas las series y diferentes horizontes de pronóstico, en particular con la adición de FGN.

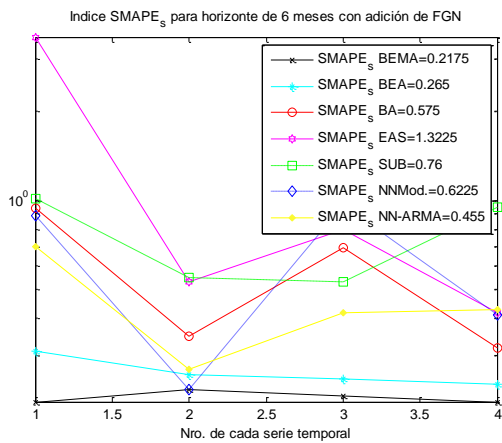
Aunque la diferencia entre los filtros reside sólo en los modelos asumidos, los coeficientes que cada filtro tiene, cada uno realiza diferentes comportamientos. Se puede observar que incluso los conjuntos de entrenamiento son demasiado cortos para el proceso de aprendizaje, el comportamiento de los filtros predictores propuesto tiene un buen desempeño predecir series temporales a corto y largo plazo. Las series de precipitaciones CA y LV presentan más rugosidad que las series AG y BA, por lo que el método BEMA y BEA demuestran una mejora como se ve en las figuras Fig. 5-22, Fig. 5-23, Fig. 5-24 y Fig. 5-25, en el que el modelo de distribución elegida para los modelos Bayesiano mejorados se puede utilizar para el ajuste de los parámetros y salidas del predictor para predecir series de lluvia.

### 5.5.5. Performance y errores de predicción

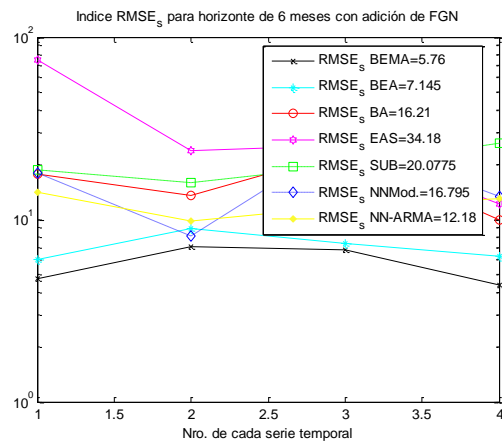
Tabla 5-14. Comparación de los métodos de pronóstico usando FGN en las series de lluvia a 6 meses

Series con FGN	FILTRO	MEDIA REAL	MEDIA PRONOSTICADA	SMAPE	RMSE
AG	BEMA	47.78	47.55	0.21	4.71
	BEA	47.78	47.20	0.31	6.01
	BA	47.78	67.21	0.94	17.76
	EAS	246.76	221.08	3.54	75.21
	SUB <sub>2-3</sub>	47.78	32.73	1.01	18.59
	NN-Mod.	47.78	13.84	0.89	17.91
	NN-ARMA	47.78	32.36	0.70	14.03
CA	BEMA	69.57	58.64	0.23	7.12
	BEA	69.57	55.25	0.26	8.94
	BA	69.57	41.64	0.35	13.50
	EAS	525.56	424.39	0.53	23.80

	SUB <sub>2-3</sub>	69.57	63.08	0.55	15.82
	NN-Mod.	69.57	52.90	0.23	8.12
	NN-ARMA	69.57	56.00	0.27	9.82
BA	BEMA	90.28	98.56	0.22	6.85
	BEA	90.28	102.34	0.25	7.35
	BA	90.28	48.54	0.69	23.65
	EAS	332.14	338.20	0.80	25.59
	SUB <sub>2-3</sub>	90.28	58.64	0.53	19.63
	NN-Mod.	90.28	79.94	0.96	27.81
	NN-ARMA	90.28	101.86	0.42	11.82
LV	BEMA	25.85	22.65	0.21	4.36
	BEA	25.85	30.56	0.24	6.28
	BA	25.85	20.93	0.32	9.93
	EAS	187.34	174.47	0.42	12.12
	SUB <sub>2-3-6</sub>	25.85	35.63	0.95	26.27
	NN-Mod.	25.85	32.33	0.41	13.34
	NN-ARMA	25.85	36.33	0.43	13.05



(a)



(b)

Fig. 5-21. Horizonte de pronóstico a 6 meses; a) y c) método BEA, b) y d) con ruido gaussiano fraccional.

Tabla 5-15. Comparación de los métodos de pronósticos a 15-18 meses

Series	MÉTODO	HORIZONTE DE PRONÓSTICO A 15 PASOS				HORIZONTE DE PRONÓSTICO A 18 PASOS			
		MEDIA REAL	MEDIA PRONOSTICADA	SMAPE	RMSE	MEDIA REAL	MEDIA PRONOSTICADA	SMAPE	RMSE
AG	BEMA	60.06	51.09	7.76	55.80	50.05	47.97	8.54	61.36
	BEA	60.06	60.24	7.45	56.85	50.05	52.69	11.61	86.42
	BA	60.06	37.22	5.79	50.96	50.05	35.29	13.11	52.43
	EAS	313.05	113.54	23.81	209.10	278.19	287.22	24.63	211.66
	SUB <sub>2-3</sub>	60.06	56.31	9.59	63.53	50.05	52.94	18.97	93.07
	NN-Mod.	60.06	46.75	6.10	50.38	50.05	67.50	10.44	67.21
	NN-ARMA	60.06	25.05	7.81	63.14	50.05	59.06	12.88	74.72
CA	BEMA	68.4	65.93	4.16	34.17	61.22	55.78	4.34	39.10
	BEA	68.4	61.71	5.52	44.91	61.22	58.56	5.07	41.88
	BA	68.4	95.54	7.97	55.13	61.22	86.79	7.07	53.10
	EAS	308.83	297.95	9.64	107.50	308.83	329.82	17.59	187.19
	SUB <sub>2-3-6</sub>	68.4	60.32	11.54	65.26	61.62	56.49	8.37	66.47
	NN-Mod.	68.4	92.02	8.88	74.26	61.62	72.82	6.98	58.32
	NN-ARMA	68.4	53.24	6.35	48.23	61.22	89.89	6.84	60.03
BA	BEMA	71.06	70.18	3.62	27.20	92.88	83.71	5.30	50.62
	BEA	71.06	78.92	7.08	55.80	92.88	99.85	12.91	94.87
	BA	71.06	86.21	11.90	96.09	92.88	75.55	13.40	97.02
	EAS	332.14	397.69	32.80	334.24	332.14	309.90	20.98	253.30
	SUB <sub>2-3</sub>	71.06	85.18	25.63	58.79	92.88	55.62	16.57	129.42
	NN-Mod.	71.06	98.64	16.30	139.89	92.88	100.30	13.54	115.98
	NN-ARMA	71.06	135.45	10.81	83.57	92.88	42.05	13.85	119.73
LV	BEMA	55.2	55.65	3.51	30.36	49.55	53.80	3.47	30.13
	BEA	55.2	52.36	4.50	31.74	49.55	51.59	6.85	51.93
	BA	55.2	57.80	9.75	66.68	49.55	66.52	11.63	86.08
	EAS	280.40	469.19	20.73	250.53	260.33	447.53	18.58	228.62
	SUB <sub>2-3-6</sub>	55.2	74.70	15.30	95.76	49.55	65.69	13.38	81.08
	NN-Mod.	55.2	57.38	5.95	48.40	49.55	51.75	5.98	45.44
	NN-ARMA	55.2	93.01	10.30	89.99	49.55	78.07	9.23	85.50

Tabla 5-16. Comparación de los métodos de pronósticos a 6-9 meses

Series	MÉTODO	HORIZONTE DE PRONOSTICO A 6 PASOS				HORIZONTE DE PRONOSTICO A 9 PASOS			
		MEDIA REAL	MEDIA PRONOSTICADA	SMAPE	RMSE	MEDIA REAL	MEDIA PRONOSTICADA	SMAPE	RMSE
AG	BEMA	55.75	22.78	4.86	31.42	50.38	40.88	4.12	28.47
	BEA	55.75	15.44	7.37	56.55	50.38	47.88	8.26	54.10
	BA	55.75	50.34	6.01	51.33	50.38	83.72	10.43	65.24
	EAS	248.18	222.43	21.86	168.77	266.00	194.44	23.64	188.66
	SUB <sub>2-3</sub>	55.75	63.34	8.43	66.32	50.38	54.39	14.69	79.3
	NN-Mod.	55.75	12.62	7.87	57.96	50.38	39.49	8.93	58.34
	NN-ARMA	55.75	32.36	7.70	64.03	50.38	51.71	8.58	57.67
CA	BEMA	81.16	73.91	2.88	24.33	54.55	61.09	2.78	22.09
	BEA	81.16	42.52	6.55	57.47	54.55	30.62	5.50	50.69
	BA	81.16	33.83	8.19	63.61	54.55	57.86	3.61	31.04
	EAS	525.53	411.83	9.07	98.04	376.93	269.92	20.12	177.95
	SUB <sub>2-3-6</sub>	81.16	69.57	7.76	62.04	54.55	61.67	7.93	55.44
	NN-Mod.	81.16	43.43	8.13	57.93	54.55	37.70	8.82	60.45
	NN-ARMA	81.16	61.01	9.61	66.08	54.55	46.98	6.31	45.10
BA	BEA	94.66	96.03	5.94	42.90	75.55	81.39	7.50	48.91
	BA	94.66	85.51	8.70	79.86	75.55	65.83	8.02	70.05
	EAS	475.27	517.28	4.10	45.54	434.68	424.00	2.31	24.67
	SUB <sub>2-3</sub>	94.66	44.93	10.66	59.97	75.55	46.76	19.67	121.01
	NN-Mod.	94.66	82.10	15.50	108.91	75.55	47.99	12.44	85.61
	NN-ARMA	94.66	108.12	5.96	43.12	75.55	72.61	50.4	37.34
LV	BEMA	30.16	25.81	3.16	25.06	24.00	18.01	3.03	25.67
	BEA	30.16	24.47	4.53	47.27	24.00	37.76	6.26	47.28
	BA	30.16	22.49	6.63	50.74	24.00	34.88	6.15	50.63
	EAS	187.34	143.98	7.66	56.17	163.78	141.08	18.36	149.55
	SUB <sub>2-3-6</sub>	30.16	25.85	4.98	48.12	24.00	47.20	10.06	53.02
	NN-Mod.	30.16	16.79	6.86	48.63	24.00	28.16	7.24	52.63
	NN-ARMA	30.16	37.05	5.96	44.13	24.00	41.86	8.70	60.37

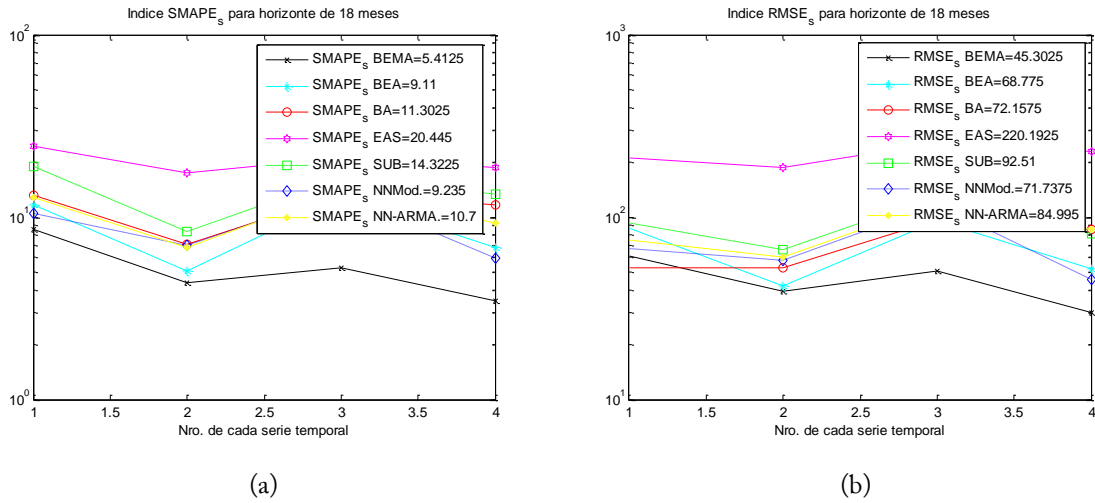


Fig. 5-22. Performance en el horizonte de pronóstico a 18 meses; a) SMAPE para las cuatro series de lluvias; b) RMSE para las cuatro series de lluvias.

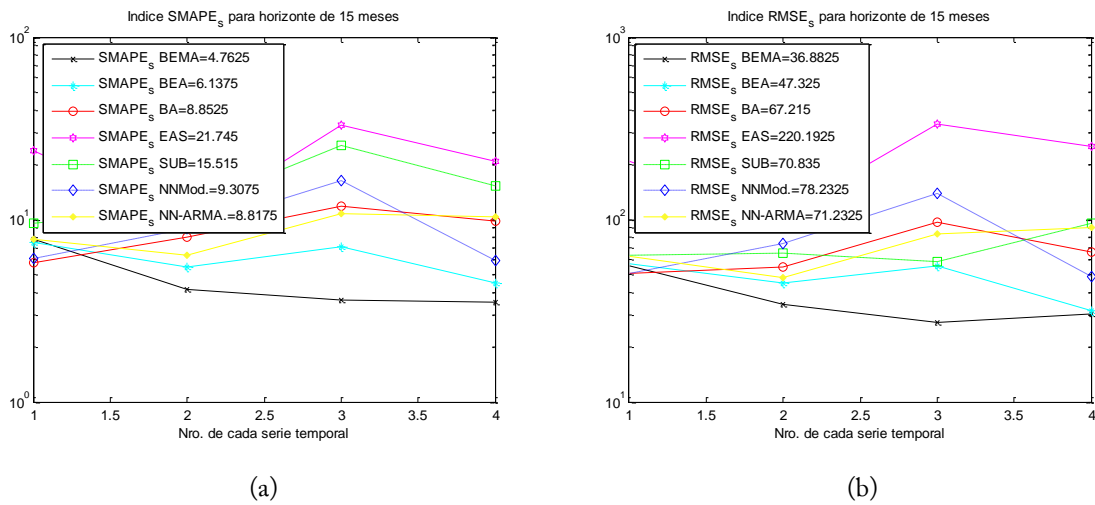
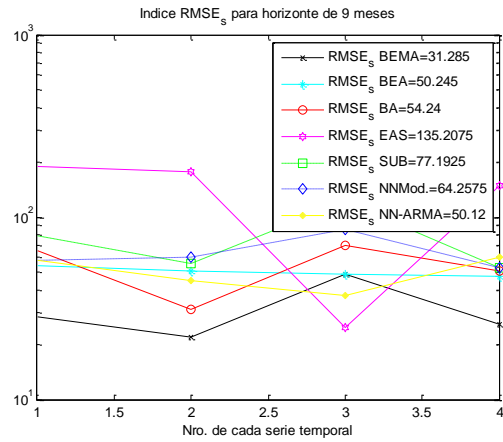
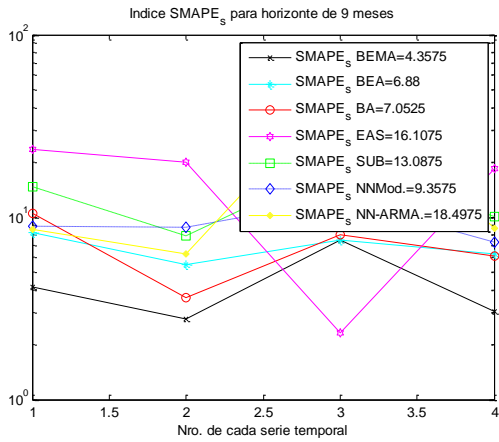


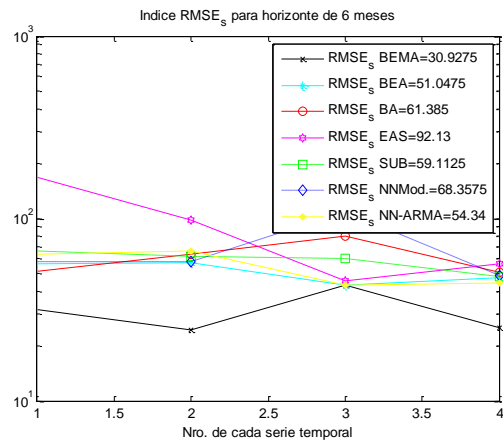
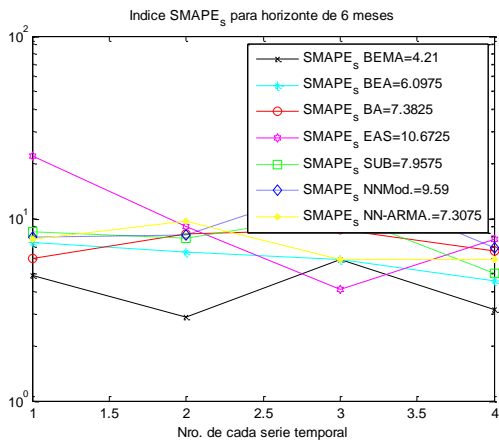
Fig. 5-23. Performance en el horizonte de pronóstico a 15 meses; a) SMAPE para las cuatro series de lluvias; b) RMSE para las cuatro series de lluvias.



(a)

(b)

Fig. 5-24. Performance en el horizonte de pronóstico a 9 meses; a) SMAPE para las cuatro series de lluvias; b) RMSE para las cuatro series de lluvias.



(a)

(b)

Fig. 5-25 Performance en el horizonte de pronóstico a 6 meses; a) SMAPE para las cuatro series de lluvias; b) RMSE para las cuatro series de lluvias.

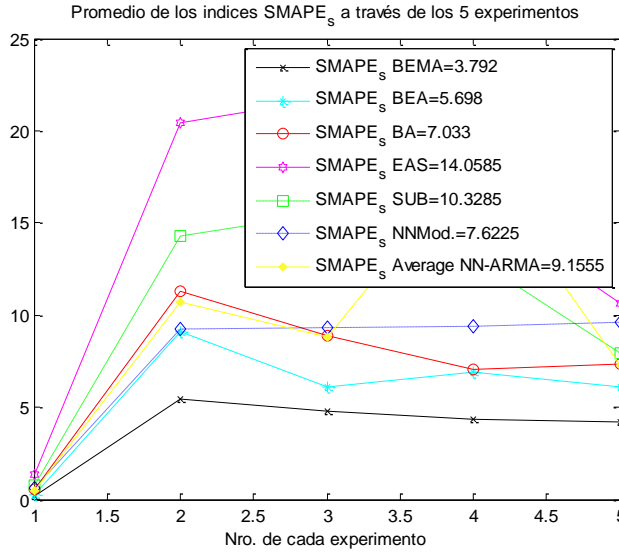


Fig. 5-26. Promedio de todos los índices SMAPE a través de los distintos horizontes de pronóstico.

El método de Monte Carlo se utilizó para predecir los próximos 6-9-15-18 valores de cada serie temporales de lluvia. Los resultados mostrados en la Tabla 5-14 a 6 meses se utilizó el método FGN con varianza 0,11 y 500 pruebas como se muestran en las figuras Fig. 5-17, Fig. 5-18, Fig. 5-19 y Fig. 5-20. La evaluación de los resultados a través de los cuatro casos analizados se realizó por los mismos parámetros iniciales para cada filtro. Los parámetros y la estructura del filtro están sintonizados al considerar su dependencia estocástica del modelo asumido.

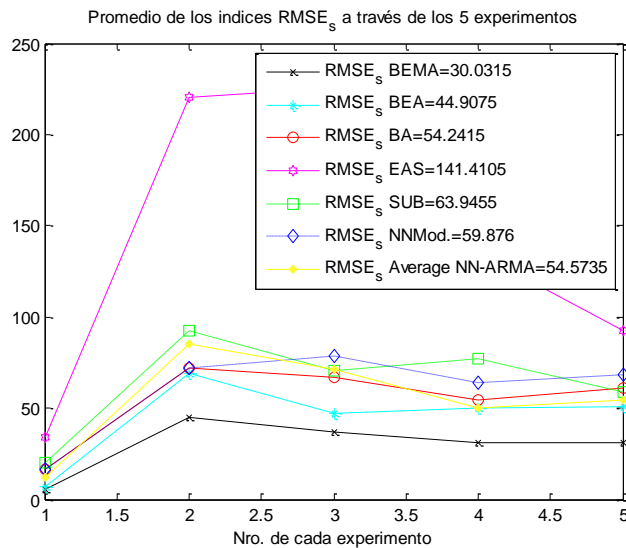


Fig. 5-27. Promedio de todos los índices RMSE a través de los distintos horizontes de pronóstico.

Puede observarse que en las figuras Fig. 5-22, Fig. 5-23, Fig. 5-24, Fig. 5-25 los índices SMAPE y RMSE se calculan entre los horizontes a 6-9-15-18 meses adelante (que incluye la validación y prueba de serie horizonte) para cada serie. Tenga en cuenta que hay una pequeña mejora del pronóstico dado para los modelos basados en aproximación Bayesiana, particularmente BEMA y BEA con sus competidores, que resulta de la utilización de una característica estocástica para generar un resultado determinista para la predicción a corto o largo plazo, mostrados en promedio para todos los horizontes de pronóstico de las cuatro series de lluvia, en las figuras Fig. 5-26 y Fig. 5-27.

Sin embargo, esta tesis sostiene que mantener las dependencias estocásticas de las series temporales restringe todos los horizontes de predicción con la misma estructura del modelo, la similitud de tendencia cíclica del rendimiento de las predicciones entre ellos es clara.

Los resultados relativos de la comparación muestran que las incertidumbres que existen en los datos pueden ser modelados asumiendo una función a priori de la no linealidad de la serie temporal, como el modelo del *mBf*. Los continuos desafíos para la predicción de series temporales incluyen la no linealidad, no estacionariedad, el ruido, y, en particular, la cantidad límite de datos. Esta tesis reconoce que al aplicar RNAs para el modelado de series temporales no lineales [310], temas tan importantes como la selección de variables de entrada adecuada y modelos previos asumidos en los datos de la serie temporal [311], [312], y el costo del equilibrio de bias/varianza en el modelo de la RNA [313]. Un gran número de simulaciones empíricas usando RNAs han demostrado que el uso como técnica de predicción por sí sola no es suficiente, se necesita de una combinación de modelos y técnicas seleccionadas para conseguir mejores resultados [314], como los modelos propuestos en esta tesis.

### 5.6. Discusión y Comentarios finales

Este capítulo presentó los resultados de la comparación de diferentes métodos de predicción para una clase de series temporales de alta rugosidad de corto y largo alcance en el horizonte del pronóstico medido por el parámetro de Hurst. Las predicciones de lluvia obtenidas entre los algoritmos propuestos se contrastan con reconocidos algoritmos, en particular para series caóticas de referencia en la literatura, y para casos de estudio sobre distintos puntos geográficos de la



Provincia de Córdoba, Argentina. Los algoritmos basados en suavización de la serie y teoría de aproximación bayesiana de este trabajo se han evaluado en distintos escenarios con datos reales y simulados. Puede observarse que en las figuras Fig. 5-26 y Fig. 5-27 donde se promedian todos los índices SMAPE y RMSE entre los horizontes de tiempo 6-9-15-18 pasos fuera de la muestra (que incluye la validación y la serie Test) para cada serie, se mejora el pronóstico dado por los métodos BEMA y BEA con sus competidores, que resulta de la utilización de la característica estocástica para generar un resultado determinista para la predicción a corto o largo plazo.

Sin embargo, el intento de mantener las dependencias estocásticas en la estructura del modelo restringe todos los horizontes de predicción. La discusión aparece de cómo los modelos basados en redes neuronales FFNN [317] pueden aproximar mejor la no linealidad, el proceso no estacionario y cambios cuantitativos en la dinámica de los datos de lluvia debido a cambios abruptos en los valores de series temporales permanece abierta para su estudio.

A pesar de que la comparación se realiza en filtros basados en RNAs, los resultados experimentales confirman que los métodos de BEMA y BEA, como así también EAS y SUB pueden ser usados para pronosticar algunas clases de series temporales caóticas de manera más efectiva en términos de SMAPE y RMSE, cuando se los compara con otros métodos de predicción existentes. Estas herramientas proporcionan un gran soporte para la toma de decisiones para su uso en el ámbito agrícola.

# Capítulo 6

*“Si usted tiene razones para pensar que el pronóstico de ayer salió mal, no hay gloria en ajustarse a él”*

Nate Silvers

## 6. Conclusiones finales

Esta tesis ha propuesto nuevos modelos no lineales de pronóstico de series temporales que resuelven la problemática de encontrar mejores predicciones para las series de lluvia, apto para la toma de decisiones agrícolas.

Particularmente, las series de lluvia admiten ser representadas por modelos no lineales, caóticos incluso estocástico. Por lo tanto, aquí se propuso que estas series se comportan como la realización de una traza del *mBf*, es decir, poseen *autosimilaridad* y tienen *memoria o dependencia de corto o largo alcance*. Los problemas y las limitaciones aparecen especialmente si la serie histórica pasada de la información tiene grandes cambios abruptos.

A partir de los resultados de simulación, se concluye que suavizar las series temporales tiene una mejora significativa en términos del error de desempeño en las predicciones para series temporales de alta rugosidad. Estos nuevos modelos, llamados EAS y SUB, permiten obtener nuevos criterios de diseño con buenos rendimientos en términos de error de desempeño de predicción. Sin embargo, dado que estos modelos presentan buenos desempeños en el horizonte para series largas, no poseen buen rendimiento cuando las series son cortas.

Se propuso la combinación de la entropía de permutación Renyi y el enfoque BEA de la teoría de aproximación bayesiana, llamado BEMA, para ser aplicados a la predicción de series temporales, en particular cortas. Los aportes de estos algoritmos para pronosticar de series no

lineales, no estacionarias son atribuibles a calcular la entropía de los datos de las series temporales, por un lado, para modificar la estructura de los parámetros de redes neuronales asumiendo el modelo BEA y por otro, que infiere a partir de las observaciones en el modelo de datos, la dependencia estocástica entre el pasado y el futuro. En general, la teoría de aproximación bayesiana muestra la importancia de emplear modelos estocásticos para abordar la complejidad inherente tanto de los pesos sinápticos como de los datos de las series temporales en el mundo real, que resulta directamente correlacionado con la predicción en series con baja cantidad de datos. Por el contrario, cuando la complejidad es relativamente baja, es más efectivo tomar en cuenta el método de suavización de la serie.

Por último, se concluye que, para encarar el problema de datos incompletos o faltantes dentro de las series temporales, se puede usar el método de suavizado promedio (MSP) descrito en la Sección 5.4 para completar los datos en una serie. Una gran ventaja de la solución propuesta, dado en la Sección 5.4.4, es que la complejidad del problema no aumenta, ya sea adicionando ruido gaussiano como un número creciente de datos que faltan en series cortas. Por lo tanto, esta contribución puede generalizarse al problema de datos inciertos o ruidosos.

### 6.1. Trabajos futuros

Con el advenimiento de los avances en técnicas basadas en inteligencia computacional para pronóstico de series temporales, tanto en las RNAs como el “*aprendizaje profundo ó Deep Learning*” [318] y sus combinaciones con otras técnicas, más los esfuerzo de investigación para representar la selección del modelo de los datos, con el fin de optimizar en entornos complejos en los datos, es que se propone analizar, diseñar y desarrollar herramientas computacionales para modelar, predecir y controlar en forma óptima el comportamiento dinámico del desarrollo de cultivos agrícolas considerados como modelos estocásticos; como así también, desarrollar herramientas computacionales avanzadas para el pronóstico de variables climáticas, que afectan directa o indirectamente el desarrollo normal de los cultivos para ser incorporado en las estrategias, estructuras y algoritmos de control usando neuroprogramación dinámica [319]. Para la estimación de la predicción en el campo de la agricultura [320] y demanda de energía [219] se propondrán

algoritmos de remuestreo de las series temporales de las variables para lograr robustez estadística dado que se trabajará con series de baja/alta cantidad de datos.

# Referencias

- [1] V. Rossi, F. Salinari, S. Poni, T. Caffi, and T. Bettati. *Addressing the implementation problem in agricultural decision support systems: The example of vite.net@*. Computers and Electronics in Agriculture, 100:88–99, 2014.
- [2] Papajorgji, Petraq J., Pardalos, Panos M. “*Advances in Modeling Agriculture Systems: optimizations and its applications*”. Volume 25. ISBN: 978-0-387-75181-8, Springer (2009).
- [3] Toni Klemm and A. McPherson Renee. “*The development of seasonal climate forecasting for agricultural producers*”. Agricultural and Forest Meteorology, Volume 232, 15 January 2017, Pages 384–399. [doi:10.1016/j.agrformet.2016.09.005](https://doi.org/10.1016/j.agrformet.2016.09.005).
- [4] Sanjay Khajure and S.W.Mohod (2016), “*Future Weather Forecasting Using Soft Computing Techniques*”, Procedia Computer Science 78, pp. 402-407. [doi:10.1016/j.procs.2016.02.081](https://doi.org/10.1016/j.procs.2016.02.081).
- [5] Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. Time Series Analysis: Forecasting and Control, 5th ed. Wiley, 2015.
- [6] Zhang, G.P., B.E. Patuwo, and M.Y. Hu. 2001. “*A simulation study of artificial neural networks for nonlinear time-series forecasting*”. Computers & Operations Research. 28(4), pp.381-396.
- [7] J.-M. Bardet, G. Lang, G. Oppenheim, A. Phillipe, and M. S. Taqqu (2003a). “*Semi-parametric estimation of the long-range dependence parameter: A survey*”. In P. Doukhan, G. Oppenheim, and M. S. Taqqu, editors, Theory and Applications of Long-Range Dependence, pages 557–577. Birkhäuser.
- [8] Zhang, G., Eddy Patuwo, B., Hu, M. Y.: “*Forecasting with artificial neural networks: The state of the art*”. International Journal of Forecasting 14(1), 35–62 (1998) [doi:10.1016/S0169-2070\(97\)00044-7](https://doi.org/10.1016/S0169-2070(97)00044-7).
- [9] Paola Andrea Sanchez. *Una nueva metodología de entrenamiento de redes neuronales y sus implicaciones en la selección de modelos*. PhD thesis, Universidad Nacional de Colombia; Facultad Nacional de Minas, Escuela de Sistemas, 2012.

- [10] Davydenko, A., & Fildes, R. (2014). "*Measuring Forecasting Accuracy: Problems and Recommendations (by the Example of SKU-Level Judgmental Adjustments)*". In *Intelligent Fashion Forecasting Systems: Models and Applications* (pp. 43-70). Springer Berlin Heidelberg.
- [11] Hyndman, R. J. & Koehler, A. B. (2006), "*Another look at measures of forecast accuracy*", *International Journal of Forecasting* 22(4), 679–688. [doi:10.1016/j.ijforecast.2006.03.001](https://doi.org/10.1016/j.ijforecast.2006.03.001).
- [12] Davydenko, A., & Fildes, R. (2016). *Forecast Error Measures: Critical Review and Practical Recommendations*. In *Business Forecasting: Practical Problems and Solutions*. John Wiley & Sons Inc.
- [13] J. Kajornrit, K. W. Wong, C. C. Fung and Y. S. Ong, "*An integrated intelligent technique for monthly rainfall time series prediction*," 2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Beijing, 2014, pp. 1632-1639. [doi:10.1109/FUZZ-IEEE.2014.6891619](https://doi.org/10.1109/FUZZ-IEEE.2014.6891619).
- [14] N. Hasan, N. C. Nath and R. I. Rasel, "*A support vector regression model for forecasting rainfall*," 2015 2nd International Conference on Electrical Information and Communication Technologies (EICT), Khulna, 2015, pp. 554-559. [doi:10.1109/EICT.2015.7392014](https://doi.org/10.1109/EICT.2015.7392014).
- [15] Pucheta, J., Patiño, D. and Kuchen, B. "*A Statistically Dependent Approach For The Monthly Rainfall Forecast from One Point Observations*". In *IFIP International Federation for Information Processing Volume 294, Computer and Computing Technologies in Agriculture II, Volume 2*, eds. D. Li, Z. Chunjiang, (Boston: Springer), pp. 787–798. (2009). <http://www.springerlink.com/content/v4651015hn60t7g8/?p=c2668d5d24fc41ef914d439a9ec86142>.
- [16] M. P. Darji, V. K. Dabhi and H. B. Prajapati, "*Rainfall forecasting using neural network: A survey*," *Computer Engineering and Applications (ICACEA)*, 2015 International Conference on Advances in, Ghaziabad, 2015, pp. 706-713. [doi:10.1109/ICACEA.2015.7164782](https://doi.org/10.1109/ICACEA.2015.7164782).
- [17] Instituto Nacional de Tecnología Agropecuaria (INTA). *Sistemas de Información Clima y Agua*. Fuente: <http://climayagua.inta.gob.ar/pronosticos>.

- [18] Senthold Assenga, Peter C. McIntoshb, Geoff Thomas, Elizabeth E. Ebertd, Nirav Khimashiae (2016). *Is a 10-day rainfall forecast of value in dry-land wheat cropping?*. Agricultural and Forest Meteorology Volume 216, 15 January 2016, Pages 170–176. [doi:10.1016/j.agrformet.2015.10.012](https://doi.org/10.1016/j.agrformet.2015.10.012).
- [19] M. R. Bendre, R. C. Thool and V. R. Thool, "Big data in precision agriculture: Weather forecasting for future farming," 2015 1st International Conference on Next Generation Computing Technologies (NGCT), Dehradun, 2015, pp. 744-750. [doi:10.1109/NGCT.2015.7375220](https://doi.org/10.1109/NGCT.2015.7375220).
- [20] Farkas I., 2005. "Modelling and control in agricultural processes", Computers and Electronics in Agriculture, Volume 49, Issue 3, December 2005, Pages 315-316. [doi:10.1016/j.compag.2005.08.002](https://doi.org/10.1016/j.compag.2005.08.002).
- [21] B. Kavitha Rani and A.Govardhan, "Rainfall prediction using data mining techniques - a survey", Natarajan Meghanathan et al. (Eds): ITCSE, ICDIP, ICAIT - 2013 pp. 23–30, 2013. © CS & IT-CSCP 2013. [doi:10.5121/csit.2013.3903](https://doi.org/10.5121/csit.2013.3903).
- [22] Wu, J., J. Long, and M. Liu, "Evolving RBF neural networks for rainfall prediction using hybrid particle swarm optimization and genetic algorithm", Neurocomputing, 2015, (148), 136–142. [doi:10.1016/j.neucom.2012.10.043](https://doi.org/10.1016/j.neucom.2012.10.043).
- [23] Xinguang Hea, Huade Guana, b, c, Jianxin Qina, "A hybrid wavelet neural network model with mutual information and particle swarm optimization for forecasting monthly rainfall", Journal of Hydrology Volume 527, August 2015, Pages 88–100. [doi:10.1016/j.jhydrol.2015.04.047](https://doi.org/10.1016/j.jhydrol.2015.04.047).
- [24] Labraga, Juan C.; Brandizi, Laura D. y Lopez, Mónica A. *Avances en el pronóstico climático de las anomalías de lluvia en la Región Pampeana. Meteorologica* [online]. 2011, vol.36, n.2 [citado 2016-03-27], pp. 59-71. Disponible en: [http://www.scielo.org.ar/scielo.php?script=sci\\_arttext&pid=S1850-468X2011000200001&lng=es&nrm=iso](http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1850-468X2011000200001&lng=es&nrm=iso). ISSN 1850-468X
- [25] S. S. Monira, Z. M. Faisal and H. Hirose, "Comparison of artificially intelligent methods in short term rainfall forecast," Computer and Information Technology (ICCIT), 2010 13th

- International Conference on, Dhaka, 2010, pp. 39-44.  
[doi:10.1109/ICCITECHN.2010.5723826](https://doi.org/10.1109/ICCITECHN.2010.5723826).
- [26] J. Kaiornrit, K. W. Wong and C. C. Fung, "A modular technique for monthly rainfall time series prediction," *Computational Intelligence in Dynamic and Uncertain Environments (CIDUE)*, 2013 IEEE Symposium on, Singapore, 2013, pp. 76-83.  
[doi:10.1109/CIDUE.2013.6595775](https://doi.org/10.1109/CIDUE.2013.6595775).
- [27] Crone, S. F., Hibon, M., Nikolopoulos, K., 2011. *Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction*. *International Journal of Forecasting* 27 (3), 635–660. [doi:10.1016/j.ijforecast.2011.04.001](https://doi.org/10.1016/j.ijforecast.2011.04.001)
- [28] M. R. Cogollo and J. D. Velázquez. *Methodological Advances in Artificial Neural Networks for Time Series Forecasting*, *IEEE Latin America Transactions*, vol. 12. nro. 4, June 2014.  
[doi:10.1109/TLA.2014.6868881](https://doi.org/10.1109/TLA.2014.6868881)
- [29] M. R. Cogollo and J. D. Velasquez, "Are Neural Networks Able To Forecast Nonlinear Time Series With Moving Average Components?," *IEEE Latin America Transactions*, vol. 13, no. 7, pp. 2292 – 2300, July 2015. [doi:10.1109/TLA.2015.7273790](https://doi.org/10.1109/TLA.2015.7273790).
- [30] Juan David Velásquez Henao. *Acotación del error de modelos de redes neuronales aplicados al pronóstico de series de tiempo*. *UIS Ingenierías*, Volumen 10, No. 1, pags. 65 – 71. Junio 2011; Facultad de Ingenierías Fisicomecánicas, UIS.
- [31] Secretaria de Ciencia y Tecnología de la Universidad Nacional de Córdoba (SECyT-UNC). Investigación 2016: Primeros proyectos valorizados. Fuente:  
<http://www.unc.edu.ar/investigacion/programas/valorizacion-del-conocimiento/primeros-proyectos-valorizados>.
- [32] De Gooijer, J.G., Hyndman, R.J., *25 years of time series forecasting*. *International Journal of Forecasting* 22(3), 443–473 (2006).
- [33] Julián Pucheta, C. Rodríguez Rivero, Martín Herrera, Carlos Salas, Víctor Sauchelli, H. Daniel Patiño, Capítulo titulado, *Non-Parametric Methods for Forecasting Time Series from Cumulative Monthly Rainfall* del libro *Rainfall: Behavior, Forecasting and Distribution*. Editors: Olga E. Martín and Tricia M. Roberts, Nova Science Publishers, Inc. ISBN: 978-1-62081-551-9 (2012). Disponible en



[https://www.novapublishers.com/catalog/product\\_info.php?products\\_id=33335&osCsid=f6e42dd9afd32cd8edc67c0e8c48a6aa](https://www.novapublishers.com/catalog/product_info.php?products_id=33335&osCsid=f6e42dd9afd32cd8edc67c0e8c48a6aa).

- [34] Julián A. Pucheta, Cristian M. Rodríguez Rivero, Martín R. Herrera, Carlos A. Salas, H. Daniel Patiño y Benjamín R. Kuchen, “*A NN approach for cumulative monthly rainfall time series forecasting tuned by roughness*”, International Journal of Physical Sciences Vol. 7(24), pp. 3018 -3031, 22 June, 2012. [doi:10.5897/IJPS11.857](https://doi.org/10.5897/IJPS11.857).
- [35] C. Rodríguez Rivero, M. Herrera, J. Pucheta, J. Baumgartner, D. Patiño and V. Sauchelli, “*High roughness time series forecasting based on energy associated of series*”, Journal of Communication and Computer, Vol. 9, No. 5, 2012, pp 576-586, ISSN 1548-7709, USA, David Publishing Company.
- [36] J. Pucheta, C. Rodríguez Rivero, M. Herrera, C. Salas and V. Sauchelli, “*Rainfall Forecasting Using Sub sampling Nonparametric Methods*,” in IEEE Latin America Transactions, vol. 11, no. 1, pp. 646-650, Feb. 2013. [doi:10.1109/TLA.2013.6502878](https://doi.org/10.1109/TLA.2013.6502878).
- [37] C. Rodríguez Rivero, J. Pucheta, M. Herrera, V. Sauchelli and S. Laboret, “*Time Series Forecasting Using Bayesian Method: Application to Cumulative Rainfall*,” in IEEE Latin America Transactions, vol. 11, no. 1, pp. 359-364, Feb. 2013. [doi:10.1109/TLA.2013.6502830](https://doi.org/10.1109/TLA.2013.6502830).
- [38] Cristian Rodríguez Rivero, Daniel Patiño, Julian Pucheta and Victor Sauchelli, “*A New Approach for Time Series Forecasting: Bayesian Enhanced by Fractional Brownian Motion with Application to Rainfall Series*”, International Journal of Advanced Computer Science and Applications (IJACSA), 7(3), 2016. [doi:10.14569/IJACSA.2016](https://doi.org/10.14569/IJACSA.2016).
- [39] Cristian Rodríguez Rivero, Julián A. Pucheta, Daniel Patiño, Víctor Sauchelli. “*Short time series prediction: Bayesian Enhanced modified Approach with application to cumulative rainfall series*”, Int. J. of Innovative Computing and Applications, 2016 Vol.7, No.3, pp.153 - 162. [doi:10.1504/IJICA.2016.078730](https://doi.org/10.1504/IJICA.2016.078730).
- [40] C. R. Rivero, H. D. Patiño and J. A. Pucheta, “*Short-term rainfall time series prediction with incomplete data*,” 2015 International Joint Conference on Neural Networks (IJCNN), Killarney, 2015, pp. 1-6. [doi:10.1109/IJCNN.2015.7280315](https://doi.org/10.1109/IJCNN.2015.7280315).

- [41] C. Rodríguez Rivero, J. Pucheta, S. Laboret, H. Patiño, V. Sauchelli, “*Energy associated tuning method for short-term series forecasting by complete and incomplete datasets*”, JAISCR, 2017, Vol. 7, No. 1, pp. 5-16. [doi: 10.1515/jaisrc-2017-0001](https://doi.org/10.1515/jaisrc-2017-0001).
- [42] Paul G. Hoel, Sidney C. Port, Charles J. Stone, *Introduction to Stochastic Processes*, Waveland Press, 1986.
- [43] C. Rodríguez Rivero, J. Pucheta, J. Baumgartner, H.D. Patiño and B. Kuchen, “*An approach for time series forecasting by simulating stochastic processes through time-lagged feed-forward neural network*”, Proceedings of the 2010 International Conference on Data Mining, DMIN 2010, July 12-15, 2010, Las Vegas, Nevada, USA, p.p 278-293. CSREA Press 2010, ISBN 1-60132-138-4.
- [44] Ajoy K. Palit, Dobrivoje Popovic *Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications*, (Advances in Industrial Control) (Kindle Edition), Springer, 2005.
- [45] Tsay, R.S. (2010). *Analysis of Financial Time Series*. Third Edition. Canada: John Wiley & Son Inc.
- [46] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, Springer Verlag, New York, 2010.
- [47] C. Chatfield, *The Analysis of Time Series: An Introduction*, 6th ed. Boston, MA: Chapman & Hall, 2004.
- [48] H. Kantz, T. Schreiber, *Nonlinear Time Series Analysis*, second ed., Cambridge University Press, Cambridge, 2006.
- [49] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd edition. Springer Series in Statistics, 2009.
- [50] Kitagawa G. (2010), *Introduction to time series modeling*, Monographs on Statistics and Applied Probability, Chapman&Hall/CRC, Boca Ration.
- [51] Decreusefond, L. & Ustunel, A. (1958), *Fractional brownian motion: Theory and applications*. <http://www.emath.fr/proc/vol5/>.
- [52] Mandelbrot, B. & Ness, J. V. (1968), “*Fractional brownian motion, fraccional noises and applications*”, SIAM Review 10(4), 422–437. [doi:10.1137/1010093](https://doi.org/10.1137/1010093).

- [53] Istas, J. and Lang, G. (1997), “*Quadratic variations and estimation of the Hölder index of a Gaussian process*”, Ann. Inst. H. Poincaré Probab. Statist. 33 407–436. [doi:10.1016/S0246-0203\(97\)80099-4](https://doi.org/10.1016/S0246-0203(97)80099-4).
- [54] M.S. Taqqu, V. Teverovsky, and W. Willinger, “*Estimators for long-range dependence: an empirical study*”, Fractals, 3(4):785–788, 1995. [doi:10.1016/S0895-7177\(99\)00104-1](https://doi.org/10.1016/S0895-7177(99)00104-1).
- [55] Abry, P., Flandrin, P., Taqqu, M.S. & Veitch, D. (2003). *Self-similarity and long-range dependence through the wavelet lens*. In P. Doukhan, G. Oppenheim & M. Taqqu (Eds.), Theory and applications of long-range dependence (527-556). Boston: Birkhäuser.
- [56] Coeurjolly, J.-F. (2007). *Hurst exponent estimation of locally self-similar Gaussian processes using sample quantiles*. Available at <http://fr.arxiv.org/abs/math.ST/0506290>.
- [57] Dieker, T. (2004). *Simulation of fractional Brownian motion*. MSc thesis, University of Twente, Amsterdam, The Netherlands.
- [58] J.-M. Bardet, G. Lang, G. Oppenheim, A. Phillippe, and M. S. Taqqu (2003b). *Generators of long-range dependent processes: A survey*. In P. Doukhan, G. Oppenheim, and M. S. Taqqu, editors, Theory and Applications of Long-Range Dependence, pages 579–623. Birkhäuser.
- [59] A. Montanari, M. S. Taqqu, and V. Teverovsky, *Estimating long-range dependence in the presence of periodicity: an empirical study*. Math. Comput. Modell. 29, 217 (1999). [doi:10.1016/S0895-7177\(99\)00104-1](https://doi.org/10.1016/S0895-7177(99)00104-1).
- [60] Diana Analía Dominguez, Marcela Hebe Gonzalez, “*Variabilidad de la Precipitación en el Centro Oeste de Argentina y un Modelo de Predicción Estadística*”, Meteorológica, Vol. 38, No.2 (2013), pp 105-120.
- [61] Penalba, O. C., and W. M. Vargas (2004), “*Interdecadal and Interannual variations of annual and extreme precipitation over central-northeastern Argentina*”, Int. J. Climatology, 24, 12, 1565-1580. [doi:10.1002/joc.1069](https://doi.org/10.1002/joc.1069).
- [62] Ravelo, A. C. (2000). *Caracterización agroclimática de las sequías extremas en la región pampeana argentina*. Rev. Arg. de Agrometeorología, 20(2): 187-192.
- [63] Barbeito, O y Ambrosino, S. (2004). “*Inundaciones Repentinas en las Sierras de Córdoba*”. Capítulo 8. Pp. 205-215. En “*Inundaciones en Argentina*”. World Meteorological Organization. Global Water Partnership. Programa asociado de gestión de crecidas América

- del Sur. Proyecto: Mejoramiento de las Prácticas de Gestión de Crecidas en Sudamérica. Editor Juan Carlos Bertoni. Autores: S. Ambrosino, O. Barbeito, J.C. Bertoni; A. Daniele, J.A. Maza, C. Paoli y J.J. Serra. ISBN 987-9406-82-6.
- [64] Almeida, G.; Ciappesoni, H. y Goniadzki, D., 2009. *Algunos aspectos de la precipitación regionalizada en el centro-norte de Argentina*. Preprint Congremet X, 5 - 9 de octubre, Buenos Aires, Argentina.
- [65] Rivarola, A., M. Vinocur y R. Seiler (2003). *Coping with climate variability in central Argentina: an analysis of farmers' attitudes toward risk*. 2003 Open Meeting of the Human Dimensions of Global Environmental Change Research Community, Montreal, Canadá, 16-19 de Octubre de 2003. Disponible <http://sedac.ciesin.columbia.edu/openmeeting/downloads.html>.
- [66] Ravelo, A., R. Irastorza y L. Luque (2002). *Variabilidad de la tendencia en series pluviométricas de la provincia de Córdoba (Argentina)*. Rev. Arg. de Agrometeorología, 2(1): 69-74.
- [67] Grimm, A. M., 2011: “*Interannual climate variability in South America: impacts on seasonal precipitation, extreme events and possible effects of climate change*”. Stochastic Environmental Research and Risk Assessment, 25, 537-554, [doi:10.1007/s00477-010-0420-1](https://doi.org/10.1007/s00477-010-0420-1).
- [68] Domínguez D. y Gonzalez M. (2011). *Tendencias de precipitación en el centro de Argentina*. Preprint XIV Congreso Latino-Americano e Ibérico de Meteorología, Setubal, Portugal.
- [69] De la Casa, A. y Nasello, O. (2012). *Tendencias parciales de los días de lluvia y la intensidad media anual en la provincia de Córdoba, Argentina*. Meteorológica 37, 2, 67-77.
- [70] Seiler, R. y M. Vinocur (2006). *Characterization of current climate and scenarios of future climate change: Argentine-South of Córdoba case study region*. p. 38-69. In Gay, C. Vulnerability and adaptation to climate variability and change: the case of farmers in Mexico and Argentina. A Final Report submitted to Assessment of Impacts and Adaptations to Climate Change (AIACC), project No. LA 29. Disponible en [http://www.aiaccproject.org/Final%20Reports/final\\_reports.html](http://www.aiaccproject.org/Final%20Reports/final_reports.html).
- [71] Mandelbrot, B. B., (1983), *The Fractal Geometry of Nature*, Freeman, San Francisco, CA. 1983.

- [72] H. Ghanmi, Z. Bargaoui, C. Mallet, “*Investigation of the fractal dimension of rainfall occurrence in a semi-arid Mediterranean climate*”, *Hydrological Sciences Journal*, 2013, 58, 3, 483. [doi:10.1080/02626667.2013.775446](https://doi.org/10.1080/02626667.2013.775446).
- [73] Beecham, S. and R. K. Chowdhury (2010). *Temporal characteristics and variability of point rainfall: a statistical and wavelet analysis*. *International Journal of Climatology*, 30, 458-473. [doi:10.1002/joc.1901](https://doi.org/10.1002/joc.1901).
- [74] Venugopal V., Stéphane G. Roux, Efi Foufoula-Georgiou, Alain Arnéodo, “*Scaling behavior of high resolution temporal rainfall: New insights from a wavelet-based cumulant analysis*”, *Physics Letters A*, 2006, 348, 3-6, 335
- [75] Monjo R. (2016), “*Measure of rainfall time structure using the dimensionless n-index*”, *Clim Res* 67:71-86.
- [76] Bellie Sivakumar, *Is a chaotic multi-fractal approach for rainfall possible?*, *Hydrological Processes*, 2001, 15, 6, 943. [doi:10.1002/hyp.260](https://doi.org/10.1002/hyp.260).
- [77] J. Feder, *Fractals*, Plenum Press, New York, 1988.
- [78] E. N. Lorenz, *The Essence of Chaos*, University of Washington Press, 1993.
- [79] Abry P. and Sellan F. (1996), “*The wavelet-based synthesis for fractional Brownian motion*”, *Applied and computational harmonic analysis*, Vol.3, p.377-383. [doi:10.1006/acha.1996.0030](https://doi.org/10.1006/acha.1996.0030).
- [80] A. Ramachandra Rao, Khaled H. Hamed, Huey-Long Chen (2003). *Nonstationarities in Hydrologic and Environmental Time Series*, Springer. ISBN 978-94-010-3979-6 ISBN 978-94-010-0117-5 (eBook). [doi:10.1007/978-94-010-0117-5](https://doi.org/10.1007/978-94-010-0117-5).
- [81] Breslin MC, Belward JA (1999): *Fractal dimension for rainfall time series*. *Mathematics and computers in simulation*, 48:437-446. [doi:10.1016/S0378-4754\(99\)00023-3](https://doi.org/10.1016/S0378-4754(99)00023-3).
- [82] *Theory and Applications of Long-range Dependence*, Paul Doukhan, Georges. Oppenheim and Murad S. Taquq editors (718 pages) Birkhäuser, Boston (2003).
- [83] Annette Witt and Bruce D. Malamud (2013), *Quantification of Long-Range Persistence in Geophysical Time Series: Conventional and Benchmark-Based Improvement Techniques*. *Surv. Geophys*, 34:541–651 [doi:10.1007/s10712-012-9217-8](https://doi.org/10.1007/s10712-012-9217-8).

- [84] Hong Quian, *Fractional Brownian Motion and Fractional Gaussian Noise, in Processes with long range correlation: Theory and Applications*. Lecture Notes in Physics, vol. 621, G. Rangarajan & M. Z. Ding, Eds., pp. 22-33, Springer (2003). [doi:10.1029/97WR01982](https://doi.org/10.1029/97WR01982).
- [85] Haan, C.T. (2002). *Statistical Methods in Hydrology*. 2nd edition, Iowa State University Press, Iowa, 496 pp.
- [86] Deepesh Machiwal, Madan Kumar Jha, Hydrologic (2012). *Time Series Analysis: Theory and Practic*. 1Ed. Springer. ISBN 978-94-007-1860-9 (HB). ISBN 978-94-007-1861-6 (e-book).
- [87] Mudelsee M. (2014), *Climate Time Series Analysis: Classical Statistical and Bootstrap Methods*. Second edition. Springer, Cham Heidelberg New York Dordrecht London. [doi:10.1007/978-3-319-04450-7](https://doi.org/10.1007/978-3-319-04450-7)
- [88] Luis A. Gil-Alana, U.K. Rainfall Data: *A Long-Term Persistence Approach*, Journal of Applied Meteorology and Climatology, 2012, 51, 10, 1904. [doi:10.1175/JAMC-D-11-0201.1](https://doi.org/10.1175/JAMC-D-11-0201.1).
- [89] G. Rangarajan and M. Ding (Eds.), *Processes with long-range correlations: theory and applications* (lecture notes in physics, 621), Springer-Verlag, 2003.
- [90] Robinson, P. M. (2003). *Time Series with Long Memory*, eds., Oxford University Press: New York.
- [91] Penalba O. and Robledo F., 2004: “*Frequency of precipitation in the humid Pampa of Argentina*”. The 15th Conference on Applied Climatology/13th Symposium on Meteorological Observations and Instrumentation, 20-24 Junio 2005, American Meteorology Society (AMS) (Savannah, Georgia, USA) [http://ams.confex.com/ams/15AppClimate/techprogram/meeting\\_15AppClimate.htm](http://ams.confex.com/ams/15AppClimate/techprogram/meeting_15AppClimate.htm).
- [92] Caamaño Nelli, G. y García, C. M.; 1999. “*Relación Intensidad-Duración-Recurrencia de Lluvias Máximas: Enfoque a través del Factor de Frecuencia, Caso Lognormal*”. Ingeniería Hidráulica En México (ISSN 0186-4076). Vol. XIV. N° 3, septiembre-diciembre. 37-44. D. F., México.
- [93] Taqqu, M. S., and V. Teverovsky (1998). *Estimating long-range dependence in finite and infinite variance series*, in R. Adler, R. Feldman, and M. S. Taqqu, eds., *A Practical Guide to Heavy Tails*, Birkhäuser: Boston.

- [94] Taqqu, M. S., and W. Willinger (1995). *Estimators for long-range dependence: an empirical study*, Fractals, vol. 3, No. 4, pp. 785-798.
- [95] Kharin Yu. *Robustness in Statistical Forecasting*. New York: Springer, 2013. pp. 55-72.
- [96] Lobato I., Robinson P.M. (1996), *Averaged periodogram estimation of long memory*, Journal of Econometrics, Vol. 73, p.303-324. [doi:10.1016/0304-4076\(95\)01742-9](https://doi.org/10.1016/0304-4076(95)01742-9).
- [97] M. S. Taqqu and V. Teverovsky. *Robustness of Whittle type estimators for time series with long-range dependence*. Stochastic Models, 13:723-757, 1997.
- [98] Coeurjolly, J.-F. (2001). *Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths*. Statist. Inference Stoch. Process. 4 199-227. [doi:10.1023/A:1017507306245](https://doi.org/10.1023/A:1017507306245).
- [99] Wfbmesti – Matlab toolbox Signal Analysis. Disponible en: <http://www.mathworks.com/help/wavelet/ref/wfbmesti.html?requestedDomain=www.mathworks.com>.
- [100] Flandrin, P. (1992). *Wavelet analysis and synthesis of fractional Brownian motion*. IEEE Transactions on Information Theory, 38(2), 910-917. [doi:10.1109/18.119751](https://doi.org/10.1109/18.119751).
- [101] Kharin, Yu, *Optimality and robustness in statistical forecasting*. - In: International Encyclopedia of Statistical Science. Springer: N.Y., 2011, p. 1034-1037.
- [102] Anestis Antoniadis, Xavier Brossat, Jairo Cugliari, Jean-Michel Poggi (2014). *A prediction interval for a function-valued forecast model*, arXiv:1412.4222 [stat.ME].
- [103] Chatfield, C. (2001). *Time-Series Forecasting*. Chapman & Hall/CRC Press.
- [104] Bosq, Denis and Blanke, Delphine. *Inference and prediction in large dimensions*. Wiley Series in Probability and Statistics, John Wiley & Sons Ltd., Dunod. 2007, 316 pages. ISBN: 978-0-470-01761-6.
- [105] Vapnik, V. N. *The Nature of Statistical Learning Theory* (2nd Ed.), Springer Verlag, 2000.
- [106] J. C. Sprott. *Chaos and Time-Series Analysis*, Oxford University Press, 2003.
- [107] Heinz Georg Schuster, Wolfram Just. *Deterministic Chaos: An Introduction*, 4th, revised and Enlarged Edition, 312 pages, Wiley (2006). ISBN: 978-3-527-60641-2.
- [108] Shumway R. H., Stoffer (2011), *Time Series Analysis and its Applications*, Springer.
- [109] Hastie, T., Tibshirani, R., Friedman, J.: *The elements of statistical learning: data mining, inference and prediction*, 2nd edn. Springer (2009).

- [110] Weigend, A.S., Gershenfeld, N.A. *Time Series Prediction: forecasting the future and understanding the past*. Addison Wesley, Harlow (1994).
- [111] Sorjamaa, A., Hao, J., Reyhani, N., Ji, Y., Lendasse, A. *Methodology for long-term prediction of time series*. Neurocomputing 70(16-18), 2861–2869 (2007).
- [112] W. K. Kim, L. Billard, and I. V. Basawa, "Estimation for the First-Order Diagonal Bilinear Time Series Model," Journal of Time Series Analysis, Vol. 11, No.3, pp. 215–229: 1990. [doi:10.1111/j.1467-9892.1990.tb00053.x](https://doi.org/10.1111/j.1467-9892.1990.tb00053.x).
- [113] Tong, H., 2011, "Threshold Models in Time Series Analysis—30 Years On," Statistics & Its Interface, 4:107–118.
- [114] William W S Wei (2006). *Time Series Analysis, Univariate And Multivariate Methods*, 2Ed. Addison Wesley.
- [115] Mitchell, T.M.: *Machine Learning*. McGraw-Hill, New York (1997).
- [116] Werbos, P.J.: *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. PhD thesis, Harvard University, Cambridge, MA (1974).
- [117] Lapedes, A., Farber, R.: *Nonlinear signal processing using neural networks: prediction and system modelling*. Technical Report LA-UR-87-2662, Los Alamos National Laboratory, Los Alamos, NM (1987).
- [118] Advances in time series forecasting (2015), Bentham Science Publishers Ltd., eISBN: 978-1-60805-373-5.
- [119] Spyros Makridakis, Michele Hibon. *The M3-Competition: results, conclusions and implications*, International Journal of Forecasting 16 (2000) 451–476. [doi:10.1016/S0169-2070\(00\)00057-1](https://doi.org/10.1016/S0169-2070(00)00057-1).
- [120] Crone, S.F.: *NN3 Forecasting Competition*, <http://www.neural-forecasting-competition.com/NN3/index.html> (last update May 26, 2009)
- [121] Crone, S.F.: *NN5 Forecasting Competition*, <http://www.neural-forecasting-competition.com/NN5/index.html> (last update May 27, 2009)
- [122] Lendasse, A. (ed.): ESTSP 2008: Proceedings. Multiprint Oy/Otamedia (2008) ISBN: 978-951-22-9544-9.



- [123] Crone, S.F.: *Mining the past to determine the future: Comments*. International Journal of Forecasting 5(3), 456–460 (2009); Special Section: Time Series Monitoring [doi:10.1016/j.ijforecast.2009.05.022](https://doi.org/10.1016/j.ijforecast.2009.05.022).
- [124] C. Rodríguez Rivero, Session Chair Organizer - Special Issue: *S19 Advances and Applications in Forecasting*, for the International Joint Conference on Neural Networks (IJCNN 2013). Disponible en [http://www.inns.org/assets/docs/2013\\_ijcnn\\_program\\_final.pdf](http://www.inns.org/assets/docs/2013_ijcnn_program_final.pdf).
- [125] C. Rodríguez Rivero, Session Chair Organizer - Special Issue: *SS37 Advances in Computational Intelligence for applied Time Series Forecasting*. International Joint Conference on Neural Networks (IJCNN 2015). Disponible en <http://uaf46365.ddns.uark.edu/proposals.html>.
- [126] C. Rodríguez Rivero, Session Chair Organizer - Special Issue: *Advances in Computational intelligence fo Applied Time Series Forecasting (ACIATSF)*, for the International Joint Conference on Neural Networks (IJCNN 2017). Disponible en <http://www.ijcnn.org>.
- [127] C. Rodríguez Rivero (2016). Session Chair Organizer - Special Issue: *IJCNN-13 Advances in Computational Intelligence for Applied Time Series Forecasting (ACIATSF)*, 2016 IEEE World Congress on Computational Intelligence - IEEE WCCI 2016. Disponible en <http://wcci2016.org/spsessions.php>.
- [128] Martin Stepnika, (2016). International Time Series Forecasting Competition - *Computational Intelligence in Forecasting (CIF)*. Disponible en <http://irafm.osu.cz/cif/main.php>.
- [129] S.N.Sivanandam and S.N.Deepa. *Principles of Soft Computing*, Wiley India (P) Ltd, 2010 Edition.
- [130] Andries P. Engelbrecht, *Computational Intelligence: An Introduction*, 2nd Edition, Wiley, 2007.
- [131] Werbos, P.J. *Generalization of backpropagation with application to a recurrent gas market model*. Neural Networks 1(4), 339–356 (1988) [doi:10.1016/0893-6080\(88\)90007-X](https://doi.org/10.1016/0893-6080(88)90007-X).
- [132] A. S. Weigend, B. A. Huberman, and D. E. Rumelhart. *Predicting the future: a connectionist approach*. International Journal of Neural Systems, 1(3):193-209, 1990.

- [133] Zhang, G. P. (2003). *Time series forecasting using a hybrid ARIMA and neural network model*. Neurocomputing, 50, 159-175. [doi:10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)
- [134] Zhang, G.; B.E. Patuwo, and M. Y. Hu. “*Forecasting with artificial neural networks: The state of art*”. J. Int. Forecasting, vol. 14, pp. 35-62. 1998. [doi:10.1016/S0169-2070\(97\)00044-7](https://doi.org/10.1016/S0169-2070(97)00044-7)
- [135] Haykin, S. (1999). *Neural Networks: a comprehensive foundation* (2nd ed.). Upper Saddle River, NJ, USA: Prentice Hall PTR.
- [136] Werbos, P. J. (1988). “*Generalization of Backpropagation with applications to a recurrent Gas-Market model*”. Neur. Net. Vol. 1, pp. 339-356. [doi:10.1016/0893-6080\(88\)90007-X](https://doi.org/10.1016/0893-6080(88)90007-X)
- [137] J. D. Velásquez Henao, C. J. Franco Cardona, and Y. Olaya Morales, “*A review of DAN2 (Dynamic Architecture for Artificial Neural Networks) model in time series forecasting*,” Ingeniería y universidad, vol. 16, no. 1, pp. 135–146, 2012.
- [138] X. Yu, M. O. Efe, and O. Kaynak, “*A General Backpropagation Algorithm for Feedforward Neural Networks Learning*”, IEEE Trans. On Neural Networks, Vol. 13, No. 1, 2002, pp. 251 – 254. [doi:10.1109/72.977323](https://doi.org/10.1109/72.977323).
- [139] M. K. S. Alsmadi, K. B. Omar, and S. A. Noah, “*Back Propagation Algorithm: The Best Algorithm Among the Multi-layer Perceptron Algorithm*”, International Journal of Computer Science and Network Security, Vol.9, No.4, 2009, pp. 378 – 383.
- [140] G. Hinton and T. Sejnowski, *Unsupervised Learning: Foundations of Neural Computation*, A Bradford Book. MIT Press, 1999.
- [141] R. Rojas, *The Backpropagation Algorithm*, Chapter 7: Neural Networks, Springer-Verlag, Berlin, pp. 151-184, 1996.
- [142] Nocedal, Jorge; Wright, Stephen J. (2006). Numerical Optimization (2nd ed.). Springer. ISBN 0-387-30303-0.
- [143] Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.
- [144] M. T. Hagan and M. Menhaj, *Training feedforward networks with the Marquardt algorithm*, IEEE Transactions on Neural Networks, 5(6), 989–993, 1994. [doi:10.1109/72.329697](https://doi.org/10.1109/72.329697).
- [145] B. M. Wilamowski, *Neural network architectures and learning algorithms*, IEEE Industrial Electronics Magazine, 3(4), 56–63, 2009. [doi:10.1109/MIE.2009.934790](https://doi.org/10.1109/MIE.2009.934790).

- [146] Armstrong J.S. (Ed.) *Principles of forecasting: Handbook for researchers and practitioners*. Kluwer, 2001.
- [147] Armstrong, J. S. (2006), *Findings from evidence-based forecasting: methods for reducing forecast error*, *International Journal of Forecasting* 22, 583–598.
- [148] Armstrong, J. S. & R. Fildes (1995), “*On the selection of error measures for comparisons among forecasting methods*,” *Journal of Forecasting*, 14, 67-71. Full text at [hops.wharton.upenn.edu/forecast](http://hops.wharton.upenn.edu/forecast).
- [149] A. Khosravi, S. Nahavandi, D. Srinivasan and R. Khosravi, “*Constructing Optimal Prediction Intervals by Using Neural Networks and Bootstrap Method*,” in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 8, pp. 1810-1815, Aug. 2015. [doi:10.1109/TNNLS.2014.2354418](https://doi.org/10.1109/TNNLS.2014.2354418).
- [150] Corradi, V. & Swanson, N. R. (2006), *Predictive density evaluation*, in ‘*Handbook of Economic Forecasting*, North-Holland, Amsterdam, The Netherlands.
- [151] Armstrong, J. S. (2001), “*Evaluating forecasting methods*,” in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Publishers.
- [152] Maxim Vladimirovich, Shcherbakov, Adriaan Brebels, Nataliya Lvovna, Shcherbakova, Anton Pavlovich Tyukov, Timur Alexandrovich Janovsky and Valeriy Anatolevich Kamaev (2013). *A Survey of Forecast Error Measures*. *World Applied Sciences Journal* 24 (Information Technologies in Modern Industry, Education & Society): 171-176. [doi:10.5829/idosi.wasj.2013.24.itmies.80032](https://doi.org/10.5829/idosi.wasj.2013.24.itmies.80032).
- [153] Armstrong, J. S. (1985), *Long-range forecasting: from crystal ball to computer*, Wiley.
- [154] Chris Tofallis. *A better measure of relative prediction accuracy for model selection and model estimation*. *Journal of the Operational Research Society* (2015) 66, 1352–1362. [doi:10.1057/jors.2014.103](https://doi.org/10.1057/jors.2014.103)
- [155] Tashman, L. J., 2000. *Out-of-sample tests of forecasting accuracy: an analysis and review*. *International Journal of Forecasting* 16 (4), 437–450. [doi:10.1016/S0169-2070\(00\)00065-0](https://doi.org/10.1016/S0169-2070(00)00065-0).
- [156] T. Chai and R. R. Draxler (2014). *Root mean square error (RMSE) or mean absolute error (MAE)? –Arguments against avoiding RMSE in the literature*. *Geosci. Model Dev.*, 7, 1247–1250. [doi:10.5194/gmd-7-1247-2014](https://doi.org/10.5194/gmd-7-1247-2014).

- [157] Green, K and Tashman, L. (2009). *Percentage error: What denominator?* Foresight, issue 12, winter 2009.
- [158] Hoover, J. H. (2009). *How to track forecast accuracy to guide forecast process improvement*. Foresight: The International Journal of Applied Forecasting 14 (Summer), 17–23.
- [159] Gurney, K. (1997). *An Introduction to Neural Networks*. London: Routledge. [ISBN 1-85728-673-1](#).
- [160] Rumelhart, D. E.; Hinton, G. E. y Willams, R. J. (1986.b) “*Learning Representations by Back-Propagating Errors*”. Nature 323. pp. 533-536. [doi:10.1038/323533a0](#).
- [161] Haykin, S., 2009. *Neural Networks and Learning Machines*. Pearson Education, Inc.
- [162] Curry, B., and Morgan, P. H. (2006). “*Model selection in neural networks: some difficulties*”, European Journal of Operational Research, 170(2), 567–577. [doi:10.1016/j.ejor.2004.05.026](#).
- [163] J. Baumgartner, C. Rodríguez Rivero and J. Pucheta. “*Pronóstico de lluvia en un punto desde diversos puntos geográficos de observación mediante procesos gaussianos*.” Anales del Congreso CONAGUA (Congreso Nacional del Agua), Chaco, ISSN 1853-7685, Argentina, realizado del 22 al 25 de Junio, 2011. <http://www.conagua2011.com.ar/dsite/actas>.
- [164] Josef Baumgartner, Cristian Rodríguez Rivero, Julián Pucheta, “*A genetic algorithm based design approach for the properties of a Gaussian process for time series forecasting*”, Anales del XXII Congreso Argentino de Control Automático (AADECA 2010) realizado del 31 de Agosto al 2 de Septiembre de 2010 – Buenos Aires, Argentina.
- [165] Rasmussen, C. E. and Williams C. K. I. (2006). *Gaussian Processes for Machine Learning*. The MIT Press.
- [166] Mitchell, M. (1998). *An Introduction to Genetic Algorithms*. The MIT Press.
- [167] Deb K., Pratap A., Agarwal S. y Meyarivan T. (2002). *A Fast and Elitist Multiobjective Genetic Algorithm*. In: IEEE Transactions on evolutionary computation, Vol.35, No.2. doi: [doi:10.1109/4235.996017](#).
- [168] Salim Fettaka, Jules Thibault, Yash Gupta (2015). *A new algorithm using front prediction and NSGA-II for solving two and three-objective optimization problems*. Optimization and Engineering, vol. 16, Issue 4, pp 713-736. [doi:10.1007/s11081-014-9271-9](#).

- [169] Matern, B. (1960). *Spatial Variation*. Meddelanden fran Statens Skogsforskningsinstitut, 49, No.5. Almaanna Forlaget, Stockholm. Second edition, Springer-Verlag, Berlin, pp. 85, 87, 89.
- [170] Neal, R. M. (1996). *Bayesian Learning for Neural Networks*, Lecture Notes in Statistics No. 118, New York: Springer-Verlag.
- [171] Deepak Ranjan Nayak, Amitav Mahapatra, Pranati Mishra. *A Survey on Rainfall Prediction using Artificial Neural Network*. International Journal of Computer Applications (0975 – 8887), Vol. 72– No.16, June 2013 [doi:10.5120/12580-9217](https://doi.org/10.5120/12580-9217).
- [172] J. N. K. Liu and R. S. T. Lee, "Rainfall forecasting from multiple point sources using neural networks," *Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings. 1999 IEEE International Conference on*, Tokyo, 1999, pp. 429-434 vol.3. [doi:10.1109/ICSMC.1999.823243](https://doi.org/10.1109/ICSMC.1999.823243).
- [173] P. M. de Lima and E. B. Guedes, "Rainfall prediction for Manaus, Amazonas with artificial neural networks," 2015 Latin America Congress on Computational Intelligence (LA-CCI), Curitiba, 2015, pp. 1-6. [doi:10.1109/LA-CCI.2015.7435934](https://doi.org/10.1109/LA-CCI.2015.7435934).
- [174] V. B. Nikam and B. B. Meshram, "Modeling Rainfall Prediction Using Data Mining Method: A Bayesian Approach," 2013 Fifth International Conference on Computational Intelligence, Modelling and Simulation, Seoul, 2013, pp. 132-136. [doi:10.1109/CIMSim.2013.29](https://doi.org/10.1109/CIMSim.2013.29).
- [175] L. Jin, Y. Huang and H. s. Zhao, "Ensemble prediction of monthly mean rainfall with a Particle Swarm Optimization-neural network model," Information Reuse and Integration (IRI), 2012 IEEE 13th International Conference on, Las Vegas, NV, 2012, pp. 287-294. [doi:10.1109/IRI.2012.6303022](https://doi.org/10.1109/IRI.2012.6303022).
- [176] J. Du, B. Zhao and S. Miao, "An Application on the Immune Evolutionary Algorithm Based on Back Propagation in the Rainfall Prediction," Computer Science and Electronics Engineering (ICCSEE), 2012 International Conference on, Hangzhou, 2012, pp. 313-317. [doi:10.1109/ICCSEE.2012.146](https://doi.org/10.1109/ICCSEE.2012.146).
- [177] F. Nhita and Adiwijaya, "A rainfall forecasting using fuzzy system based on genetic algorithm," Information and Communication Technology (ICoICT), 2013 International Conference of, Bandung, 2013, pp. 111-115. [doi:10.1109/ICoICT.2013.6574557](https://doi.org/10.1109/ICoICT.2013.6574557).

- [178] A. Sharma and M. K. Goyal, "*Bayesian network model for monthly rainfall forecast*," 2015 IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kolkata, 2015, pp. 241-246. [doi:10.1109/ICRCICN.2015.7434243](https://doi.org/10.1109/ICRCICN.2015.7434243).
- [179] J. A. Pucheta; C. Rodríguez Rivero; Herrera, M.; Carlos Salas; Londero, O.; Sauchelli, V.; H. D. Patiño; Cano, J. E.; Savio, M. E. *Método no paramétrico basado en rugosidad para el diseño de filtros predictores de series temporales*. Producción Científica de la Facultad de Tecnología y Ciencias Aplicadas III. Catamarca: Editorial Científica Universitaria de la Universidad Nacional de Catamarca. 2012. p117 - 122. ISBN 978-987-661-116-9.
- [180] C. Rodríguez Rivero, J. Pucheta, J. Baumgartner, M. Herrera, D. Patiño y B. Kuchen, "*A NN-based model for time series forecasting in function of energy associated of series*", Proceedings of the International Conference on Applied, Numerical and Computational Mathematics (ICANCM11), Barcelona, Spain, September 15-17, 2011, ISBN 978-1-61804-030-5, Pp. 80-86. (2011). Disponible en: <http://www.wseas.us/e-library/conferences/2011/Barcelona/ICICIC/ICICIC-13.pdf>.
- [181] Pucheta, J.; Rodríguez Rivero, C.; Herrera, M.; Carlos Salas; Londero, O.; Sauchelli, V.; H. D. Patiño; Cano, J. E.; Savio, M. E. *Método no paramétrico basado en energía para el diseño de filtros predictores de series temporales*. Producción Científica de la Facultad de Tecnología y Ciencias Aplicadas III. Catamarca: Editorial Científica Universitaria de la Universidad Nacional de Catamarca. 2012. p.129 - 134. ISBN 978-987-661-116-9.
- [182] J. Pucheta, M. Herrera, C. Rivero Rodríguez and J. Baumgartner. *Filtro no lineal para pronóstico de series temporales de Alta rugosidad empleando submuestreo*. Anales del Congreso CONAGUA (Congreso Nacional del Agua), Chaco, ISSN 1853-7685, Argentina, realizado del 22 al 25 de Junio, 2011. <http://www.conagua2011.com.ar/dsite/actas>.
- [183] J. A. Pucheta, M. R. Herrera, C. M. Rodríguez Rivero, J. S. Baumgartner, V. H. Sauchelli, "*Filtro predictor basado en redes neuronales para pronóstico de series temporales de lluvia acumulada empleando submuestreo*", Anales del CAI 2011 - Congreso Argentino de Agroinformática, 40 Jornadas Argentinas de Informática, 29 de Agosto al 2 de Septiembre

de 2011, Universidad Tecnológica Nacional Facultad Regional Córdoba, Argentina (2011).  
<http://www.40jaiio.org.ar/node/162>.

- [184] Glass, L. and M. C. Mackey. *From Clocks to Chaos, The Rhythms of Life*. Princeton University Press, Princeton, NJ, 1988.
- [185] M. Mackey and L. Glass, "Oscillation and chaos in physiological control systems" *Science*, vol. 197, p. 287, 1977. [doi:10.1126/science.267326](https://doi.org/10.1126/science.267326).
- [186] Pucheta, J. A.; C. Rodríguez Rivero; Herrera, M.; Carlos Salas; Londero, O.; Sauchelli, V.; H. D. Patiño; Cano, J. E.; Savio, M. E. *Método no paramétrico basado en submuestreo para el diseño de filtros predictores de series temporales*. Producción Científica de la Facultad de Tecnología y Ciencias Aplicadas III. Catamarca: Editorial Científica Universitaria de la Universidad Nacional de Catamarca. 2012. Pp. 123 - 128. ISBN 978-987-661-116-9.
- [187] Patrick K. Simpson. "Foundation of Neural Networks," IEEE Technology Update Series, vol. Neural Networks Theory, Technologies and Applications, 1996, pp. 1-22.
- [188] Cristian M. Rodríguez Rivero, Julián A. Pucheta, "Forecasting rainfall time series with stochastic output approximated by neural networks Bayesian approach", *International Journal of Advanced Computer Science and Applications (IJACSA)*. West Yorkshire: The Science and Information (SAI) Organization. 2014 vol.5 n°6. p 145 - 150. ISSN 2156-5570. [doi:10.14569/IJACSA.2014.050623](https://doi.org/10.14569/IJACSA.2014.050623).
- [189] C. R. Rivero, J. Pucheta, H. Patiño, J. Baumgartner, S. Laboret and V. Sauchelli, "Analysis of a Gaussian process and feed-forward neural networks based filter for forecasting short rainfall time series," *Neural Networks (IJCNN)*, The 2013 International Joint Conference on, Dallas, TX, 2013, pp. 1-6. [doi:10.1109/IJCNN.2013.6706741](https://doi.org/10.1109/IJCNN.2013.6706741).
- [190] Pucheta, J., Rodríguez Rivero, C., Herrera, M., Sauchelli, V. and J. Baumgartner, "Time Series Forecasting using Kernel and Feed-Forward Neural", *Anales del XIV Reunión de Trabajo en Procesamiento de la Información y Control RPIC 2011*, 16 al 18 de Noviembre de 2011 Oro Verde, Entre Ríos, Argentina. ISBN 978-950-698-280-5. (2011).
- [191] Henrique S. Hippert, James W, Taylor A, (2010). *Evaluation of Bayesian techniques for controlling model complexity and selecting inputs in a neural network for short-term load forecasting*. *Neural Networks* (23), 386-395.

- [192] Jouko Lampinen, Aki Vehtari (2001). *Bayesian approach for neural networks - review and case studies*. Neural Networks (14), 257-274. [doi:10.1016/S0893-6080\(00\)00098-8](https://doi.org/10.1016/S0893-6080(00)00098-8).
- [193] Buntine, W. L., & Weigend, A. S. (1991). *Bayesian back-propagation*. Complex Systems, 05(6), 603.
- [194] D. J. C. MacKay, "A Practical Bayesian Framework for Backpropagation Networks," in Neural Computation, vol. 4, no. 3, pp. 448-472, May 1992. [doi:10.1162/neco.1992.4.3.448](https://doi.org/10.1162/neco.1992.4.3.448).
- [195] MacKay, David J.C. (1992) *Bayesian methods for adaptive models*. Dissertation (Ph.D.), California Institute of Technology.
- [196] Neal, R. M. (1992). *Bayesian training of backpropagation networks by the hybrid Monte Carlo method*. Technical report CRG-TR-92-1, Department of Computer Science, University of Toronto.
- [197] David J.C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.
- [198] MacKay, D. J. C. (1995). *Probable networks and plausible predictions - a review of practical Bayesian methods for supervised neural networks*. Network: Computation in Neural Systems, 6 (3), 469-505.
- [199] Ozan Kocadağlı, Barış Aşıkgil (2014). *Nonlinear time series forecasting with Bayesian neural networks*. Expert Systems with Applications, vol. 41, Issue 15, pp. 6596–6610. [doi:10.1016/j.eswa.2014.04.035](https://doi.org/10.1016/j.eswa.2014.04.035).
- [200] Faming Liang.(2005). *Bayesian neural networks for nonlinear time series forecasting*. Statistics and Computing, Vol. 15, Issue 1, pp 13-29. [doi:10.1007/s11222-005-4786-8](https://doi.org/10.1007/s11222-005-4786-8).
- [201] T. Zhang and A. Fukushige, "Forecasting time series by Bayesian neural networks," Neural Networks, 2002. IJCNN '02. Proceedings of the 2002 International Joint Conference on, Honolulu, HI, 2002, pp. 382-387. [doi:10.1109/IJCNN.2002.1005502](https://doi.org/10.1109/IJCNN.2002.1005502).
- [202] Tobias Scheffer and Thorsten Joachims (1998), *Estimating the expected error of empirical minimizers for model selection*. TR-98-9, TU-Berlin.
- [203] W. Buntine, "A guide to the literature on learning probabilistic networks from data," in IEEE Transactions on Knowledge and Data Engineering, vol. 8, no. 2, pp. 195-210, Apr 1996. [doi:10.1109/69.494161](https://doi.org/10.1109/69.494161).



- [204] Tanner, M. A. (1996). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. 3rd Edn., New York.
- [205] D. J. C. MacKay. *Hyperparameters: optimise or integrate out?*, Maximum Entropy and Bayesian Methods, Santa Barbara, Dordrecht, 1993.
- [206] Hippert HS, Taylor JW (2010). *An evaluation of Bayesian techniques for controlling model complexity and selecting inputs in a neural network for short-term load forecasting*. Neural Netw. (3):386-95. [doi:10.1016/j.neunet.2009.11.016](https://doi.org/10.1016/j.neunet.2009.11.016).
- [207] Neapolitan, R. (2003). *Learning Bayesian Networks*. Prentice-Hall, Inc.
- [208] D.J.C. MacKay, *A practical Bayesian framework for backpropagation networks*, Neural Comput. 4 (1992) 448-472.
- [209] Wasserman, L. (2004), *All of Statistics: A Concise Course in Statistical Inference*, New York: Springer-Verlag.
- [210] Congdon, P. (2003), *Applied Bayesian Modeling*, John Wiley & Sons.
- [211] D. M. Titterington (2004). *Bayesian Methods for Neural Networks and Related Models*. Statistical Science, Vol. 19, No. 1, 128–139. [doi:10.1214/088342304000000099](https://doi.org/10.1214/088342304000000099).
- [212] Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis*, Second Edition, New York: Springer-Verlag.
- [213] S. Marra, F. C. Morabito and M. Versaci, "*Design of neural predictors using tools of chaos theory and Bayesian learning*," Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005, Montreal, Que., 2005, pp. 2222-2227 vol. 4. [doi:10.1109/IJCNN.2005.1556246](https://doi.org/10.1109/IJCNN.2005.1556246).
- [214] Congdon, P. (2001), *Bayesian Statistical Modeling*, John Wiley & Sons.
- [215] Lee, H.K.H. (2004). *Bayesian Nonparametrics via Neural Networks*, SIAM. [doi:10.1137/1.9780898718423](https://doi.org/10.1137/1.9780898718423).
- [216] Kevin B. Korb and Ann E. Nicholson, *Bayesian Artificial Intelligence*, Second edition, 2010, CRC press.
- [217] C. Rodríguez Rivero, J. Pucheta, J. Baumgartner, M. Herrera, H.D. Patiño and V. Sauchelli "*Modelado bayesiano de un filtro autorregresivo no lineal basado en redes neuronales para el pronóstico de series temporales de lluvia acumulada mensual*", Anales del Congreso

Nacional del Agua (CONAGUA 2011), 22 al 25 de Junio de 2011, Chaco, Argentina. ISSN 1853-7685.

- [218] O. Kocadagli and B. Asikgil, "Nonlinear time series forecasting with Bayesian neural networks", *Expert Syst. with Applications*, vol. 41, no. 15, pp. 6596–6610, 2014. [doi:10.1016/j.eswa.2014.04.035](https://doi.org/10.1016/j.eswa.2014.04.035).
- [219] C. R. Rivero, V. Sauchelli, H. Daniel Patiño, J. A. Pucheta and S. Laboret, "Long-term power consumption demand prediction: A comparison of energy associated and Bayesian modeling approach," 2015 Latin America Congress on Computational Intelligence (LA-CCI), Curitiba, 2015, pp. 1-6. [doi: 10.1109/LA-CCI.2015.7435938](https://doi.org/10.1109/LA-CCI.2015.7435938).
- [220] Liang Faming, Chuanhai Liu, Raymond J. Carroll. *Advanced Markov chain Monte Carlo methods: learning from past samples*. Hoboken, NJ: Wiley, 2010.
- [221] Rossi, P. E., Allenby G. M., & McCulloch, R. (2006). *Bayesian statistics and marketing*. Hoboken, NJ: John Wiley & Sons.
- [222] Chaohong Song, Qiang Luo and Feng Shi (2008). *A Bayesian Dynamic Forecast Model Based on Neural Network*. In proc. of International Symposium on Intelligent Information Technology Application Workshops, IITAW '08., pp 130–132. [doi:10.1109/IITA.Workshops.2008.57](https://doi.org/10.1109/IITA.Workshops.2008.57).
- [223] M. Ghofrani, K. West and M. Ghayekhloo, "Hybrid time series-bayesian neural network short-term load forecasting with a new input selection method," 2015 IEEE Power & Energy Society General Meeting, Denver, CO, 2015, pp. 1-5. [doi:10.1109/PESGM.2015.7286140](https://doi.org/10.1109/PESGM.2015.7286140).
- [224] T. Lundahl, W. J. Ohley, S. M. Kay and R. Siffert, "Fractional Brownian Motion: A Maximum Likelihood Estimator and Its Application to Image Texture," in *IEEE Transactions on Medical Imaging*, vol. 5, no. 3, pp. 152-161, Sept. 1986. [doi:10.1109/TMI.1986.4307764](https://doi.org/10.1109/TMI.1986.4307764).
- [225] Rao P., B. L. S. (2004). *Self-similar process, fractional Brownian motion and statistical inference*. Lecture Notes-Monograph Series, 45, 98-125.
- [226] C. R. Rivero, J. Pucheta, D. Patiño, S. Laboret, G. Juárez and V. Sauchelli, "A combined approach for long-term series prediction: Renyi permutation entropy with BEA predictor

- filter*," 2016 IEEE Biennial Congress of Argentina (ARGENCON), Buenos Aires, Argentina, 2016, pp. 1-5. doi: [10.1109/ARGENCON.2016.7585299](https://doi.org/10.1109/ARGENCON.2016.7585299).
- [227] D. Niu, H. Shi and Z. D. D. Wu, "Short-term load forecasting using Bayesian neural networks learned by hybrid Monte Carlo algorithm", Applied Soft Computing Journal, vol. 12, no. 6, pp. 1822-1827, 2012.
- [228] Akaike H, Kitagawa G, editors. *The practice of time series analysis: statistics for engineering and physical science*. Berlin: Springer; 1996.
- [229] Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). New York, NY: Chapman & Hall/CRC.
- [230] BELLIE SIVAKUMAR (2000). *Fractal analysis of rainfall observed in two different climatic regions*, Hydrological Sciences Journal, 45:5, 727-738, [doi:10.1080/02626660009492373](https://doi.org/10.1080/02626660009492373).
- [231] Rodriguez-Iturbe, I., D. R. Cox, and V. Isham (1987), *Some models for rainfall based on stochastic point processes*, Proc. R. Soc. London, Ser. A, 410, 269–288. [doi:10.1098/rspa.1987.0039](https://doi.org/10.1098/rspa.1987.0039).
- [232] Kurchenko, O. O. (2003). *A consistent estimator of the Hurst parameter for a fractional Brownian motion*, Theory of Probability and Mathematical Statistics, 67, 97-106.
- [233] Hu, Y., Nualart, D., Xiao, W., & Zhang, W. (2011). *Exact maximum likelihood estimator for drift fractional Brownian motion at discrete observation*. Acta Mathematica. [doi:10.1016/S0252-9602\(11\)60365-2](https://doi.org/10.1016/S0252-9602(11)60365-2).
- [234] Tian Y, Gao F. *Extraction of delay information from chaotic series based on information entropy*. Physica D 1997; 108:113–8. [doi:10.1016/S0167-2789\(97\)82008-8](https://doi.org/10.1016/S0167-2789(97)82008-8).
- [235] Bandt, C. and Pompe, B. (2002), "Permutation entropy – a natural complexity measure for time series", Phys. Rev. Lett., April, Vol. 88, No. 17. p. 174102
- [236] J. Garland and E. Bradley, *On the Importance of Nonlinear Modeling in Computer Performance Prediction*, Chapter in Advances in Intelligent Data Analysis XII, Lecture Notes in Computer Science, Vol. 8207, edited by A. Tucker, F. Hoppner, A. Siebes, and S. Swift (Springer Berlin/Heidelberg, 2013) pp. 210-222.
- [237] K. Kalpakis, S. Yang, P. F. Hu, C. F. Mackenzie, L. G. Stansbury, D. M. Stein, T. M. Scalea (2015). *Permutation entropy analysis of vital signs data for outcome prediction of*

- patients with severe traumatic brain injury*. Computers in Biology and Medicine, 56:167-74. [doi:10.1016/j.combiomed.2014.11.007](https://doi.org/10.1016/j.combiomed.2014.11.007).
- [238] Joshua Garland, Ryan James and Elizabeth Bradley (2014). *Model-free quantification of time-series predictability*. Physical Review E 90 (5), 052910. [doi:10.1103/PhysRevE.90.052910](https://doi.org/10.1103/PhysRevE.90.052910).
- [239] Cina Aghamohammadia, Mehran Ebrahimianb, Hamed Tahmooresic (2014). *Permutation approach, high frequency trading and variety of micro patterns in financial time series*. Physica A: Statistical Mechanics and its Applications, vol. 413, pp. 25–30. [doi:10.1016/j.physa.2014.06.027](https://doi.org/10.1016/j.physa.2014.06.027).
- [240] G. A. Darbellay and D. Wuertz. *The entropy as a tool for analysing statistical dependences in financial time series*. Physica A, 287:429– 439, June 2000.
- [241] Shannon, C. E. "A Mathematical Theory of Communication." The Bell System Technical J. 27, 379-423 and 623-656, July and Oct. 1948.
- [242] A Renyi. *On measures of entropy and information*. In Proc. Fourth Berkeley Symp. Math. Stat. Prob., 1960, volume 1, page 547, Berkeley, 1961. University of California Press.
- [243] L Molgedey and W Ebeling. *Intraday patterns and local predictability of high-frequency in financial time series*. Physica A, 287:420–428, June 2000. [doi:10.1016/S0378-4371\(00\)00381-2](https://doi.org/10.1016/S0378-4371(00)00381-2).
- [244] Lorenz E., *Deterministic non-periodic flow*, Journal of Atmospheric Science, 20, pp. 130-146. [doi:10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2).
- [245] Cano, R., Sordo, C., and Gutierrez, J. M. (2004). *Bayesian networks in meteorology*. In J. A. Gamez, S. Moral, and A. Salmeron, eds., Advances in Bayesian Networks, pp. 309-327. Springer Verlag.
- [246] Nazim Osman Bushara, Ajith Abraham (2013). *Computational Intelligence in Weather Forecasting: A Review*. Journal of Network and Innovative Computing ISSN 2160-2174, vol. 1, pp. 320-331.
- [247] Fildes, R., Kourentzes, N., 2011. *Validation and forecasting accuracy in models of climate change*. International Journal of Forecasting 27 (4), 968–995. [doi:10.1016/j.ijforecast.2011.03.008](https://doi.org/10.1016/j.ijforecast.2011.03.008).

- [248] Center for Studies of Variability and Climate Change (CEVARCAM) - Facultad de Ingeniería y Ciencias Hídricas (FICH), Universidad Nacional del Litoral (UNL).
- [249] J. Abbot, J. Marohasy. *Application of artificial neural networks to rainfall forecasting in Queensland*, Australia. *Advances in Atmospheric Sciences*, vol. 29(4), pp. 717-730, 2012. [doi:10.1007/s00376-012-1259-9](https://doi.org/10.1007/s00376-012-1259-9).
- [250] C.L. Wu and K.W.Chau (2013). *Prediction of rainfall time series using modular soft computing methods*. *Engineering Applications of Artificial Intelligence*, 26, 997-1007. [doi:10.1016/j.engappai.2012.05.023](https://doi.org/10.1016/j.engappai.2012.05.023).
- [251] Chattopadhyay, S., Chattopadhyay, G., 2008b. *Comparative study among different neural net learning algorithms applied to rainfall time series*. *Meteorol. Appl.* 15, 273-280. [doi:10.1002/met.71](https://doi.org/10.1002/met.71).
- [252] Chattopadhyay, S., Chattopadhyay, G., 2010. *Univariate modelling of summer-monsoon rainfall time series: comparison between ARIMA and ARNN*. *CR Geoscience* 342, 100-107. [doi:10.1016/j.crte.2009.10.016](https://doi.org/10.1016/j.crte.2009.10.016).
- [253] R. Venkata Ramana, B. Krishna, S. R. Kumar and N. G. Pandey (2013). *Monthly Rainfall Prediction Using Wavelet Neural Network Analysis*. *Water Resour Manage* 27:3697-3711. [doi:10.1007/s11269-013-0374-4](https://doi.org/10.1007/s11269-013-0374-4).
- [254] Changsam Jeong, Ju-Young Shin, Taesoon Kim and Jun-Haneg Heo, “*Monthly Precipitation Forecasting with neuro-fuzzy model*”, *Water Resour Manage*, 26:4467-4483, 2012. [doi:10.1007/s11269-012-0157-3](https://doi.org/10.1007/s11269-012-0157-3).
- [255] Silvina M. Cabrini, Carlos P. Calcaterra (2016). *Modeling economic-environmental decision making for agricultural land use in Argentinean Pampas*. *Agricultural Systems*, vol. 143, March 2016, Pp. 183-194. [doi:10.1016/j.agry.2015.12.016](https://doi.org/10.1016/j.agry.2015.12.016).
- [256] Diana Analia Dominguez, Marcela Hebe Gonzalez (2013), “*Variabilidad de la precipitacion en el centro oeste de argentina y un modelo de prediccion estadística*”, *Meteorologica*, Vol. 38 (2), 105-120.
- [257] ASCE Task Committee on Application of Artificial Neural Network in Hydrology (2000a), *Artificial neural network in hydrology. I: Preliminary concepts*, *J. Hydrol. Eng.*, 5, 115-123. [doi:10.1061/\(ASCE\)1084-0699\(2000\)5:2\(115\)](https://doi.org/10.1061/(ASCE)1084-0699(2000)5:2(115)).

- [258] ASCE Task Committee on Application of Artificial Neural Network in Hydrology (2000b), *Artificial neural network in hydrology. II: Hydrologic applications*, J. Hydrol. Eng., 5, 124–137. [doi:10.1061/\(ASCE\)1084-0699\(2000\)5:2\(124\)](https://doi.org/10.1061/(ASCE)1084-0699(2000)5:2(124)).
- [259] Souhaib Ben Taieb, Gianluca Bontempi, Amir Atiya and Antti Sorjamaa, *A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition*, 2011.
- [260] Koehler, AB. (2001). *The asymmetry of the sMAPE measure and other comments on the M3-Competition*. Int. J. Forecasting, 17, 570-574.
- [261] Elliott, G., Timmermann, A., 2004. *Optimal forecast combinations under general loss functions and forecast error distributions*. Journal of Econometrics 122 (1), 47–79.
- [262] B. G. Flores, “*A pragmatic view of accuracy measurement in forecasting*”, Omega (Oxford), 1986, Vol. 19, No. 2, pp. 93–98.
- [263] Gneiting, T. (2011). *Making and Evaluating Point Forecasts*. J. Amer. Stat. Assoc. 106,746-762. [doi:10.1198/jasa.2011.r10138](https://doi.org/10.1198/jasa.2011.r10138)
- [264] M. Han and X. Wang, “*Robust neural predictor for noisy chaotic time series prediction*,” Neural Networks (IJCNN), The 2013 International Joint Conference on, Dallas, TX, 2013, pp. 1-5. [doi:10.1109/IJCNN.2013.6706996](https://doi.org/10.1109/IJCNN.2013.6706996).
- [265] Espinoza Contreras, Adriana Eliza. *El Caos y la caracterización de series de tiempo a través de técnicas de la dinámica no lineal*. Universidad Autónoma de México. Campus Aragón. 2004.
- [266] Velásquez Henao, Juan David, Dyna, Red. *Pronóstico de la serie de Mackey glass usando modelos de regresión no-lineal*. Universidad Autónoma de Mexico. Campus Aragón. 2004.
- [267] A. Gholipour, B. Araabi, and C. Lucas, “*Predicting Chaotic Time Series Using Neural and Neurofuzzy Models: A Comparative Study*,” Neural Processing Letters, vol. 24, pp. 217-239, 2006. [doi:10.1007/s11063-006-9021-x](https://doi.org/10.1007/s11063-006-9021-x).
- [268] Hénon, M. (1976) *A two-dimensional mapping with a strange attractor*. Communications in Mathematical Physics. Vol. 50, pp. 69–77. [doi:10.1007/BF01608556](https://doi.org/10.1007/BF01608556).
- [269] Davies, B. (1999). *Exploring Chaos, theory and experiment*. Perseus Books, Reading, Massachusetts.
- [270] K. Ikeda. Opt . Commun. 30, 257 (1979).

- [271] Robert M. May *"Simple mathematical models with very complicated dynamics,"* Nature, vol. 261, 459-467 (1976). [doi:10.1038/261459a0](https://doi.org/10.1038/261459a0).
- [272] Sprott, J. C. *"Automatic Generation of Strange Attractors."* Comput. & Graphics 17, 325-332, 1993. Reprinted in Chaos and Fractals, A Computer Graphical Journey: Ten Year Compilation of Advanced Research (Ed. C. A. Pickover). Amsterdam, Netherlands: Elsevier, pp. 53-60, 1998.
- [273] D. Li, M. Han and J. Wang, *"Chaotic Time Series Prediction Based on a Novel Robust Echo State Network,"* in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 5, pp. 787-799, May 2012. [doi:10.1109/TNNLS.2012.2188414](https://doi.org/10.1109/TNNLS.2012.2188414).
- [274] V. Tresp and R. Hofmann, *"Missing and noisy data in nonlinear time-series prediction,"* *Neural Networks for Signal Processing [1995] V. Proceedings of the 1995 IEEE Workshop, Cambridge, MA, 1995*, pp. 1-10. [doi:10.1109/NNSP.1995.514873](https://doi.org/10.1109/NNSP.1995.514873).
- [275] S. Bhardwaj, S. Srivastava, S. Vaishnavi and J. R. P. Gupta, *"Chaotic time series prediction using combination of Hidden Markov Model and Neural Nets,"* Computer Information Systems and Industrial Management Applications (CISIM), 2010 International Conference on, Krackow, 2010, pp. 585-589.
- [276] Nhat-Duc Hoang, Anh-Duc Pham, and Minh-Tu Cao, *"A Novel Time Series Prediction Approach Based on a Hybridization of Least Squares Support Vector Regression and Swarm Intelligence,"* Applied Computational Intelligence and Soft Computing, vol. 2014, Article ID 754809, 8 pages, 2014. [doi:10.1155/2014/754809](https://doi.org/10.1155/2014/754809).
- [277] Y. Dong and J. Zhang, *"An improved boosting scheme based ensemble of Fuzzy Neural Networks for nonlinear time series prediction,"* 2014 International Joint Conference on Neural Networks (IJCNN), Beijing, 2014, pp. 157-164. [doi:10.1109/IJCNN.2014.6889431](https://doi.org/10.1109/IJCNN.2014.6889431).
- [278] Havaluddina, Ahmad Jawahir, *Comparing of ARIMA and RBFNN for short-term forecasting,* International Journal of Advances in Intelligent Informatics, Vol. 1, No 1, March 2015, pp. 15-22. [doi:10.12928/ijain.v1i1.10](https://doi.org/10.12928/ijain.v1i1.10).
- [279] Little, R.J.A. and D.B. Rubin, 2002. *Statistical Analysis with Missing Data.* John Wiley Publishers Company.

- [280] Tsirikiktsis, N.: 2005, *A review of techniques for treating missing data in OM survey research*, Journal of Operations Management 24, 53–62. [doi:10.1016/j.jom.2005.03.001](https://doi.org/10.1016/j.jom.2005.03.001)
- [281] Maravall A, Peña D. (1986) *Missing observations and additive outliers in time series models*. In: Mariano RS (ed) *Advances in statistical analysis and statistical computing*. JAI Press, Stanford.
- [282] Mitat Uysal, *Reconstruction of Time Series Data with Missing Values*. Journal of Applied Sciences, 7 (6): 922-925, 2007. ISSN 1812-5654. [doi:10.3923/jas.2007.922.925](https://doi.org/10.3923/jas.2007.922.925)
- [283] Kornelsen, K., & Coulibaly, P. (2012). *Comparison of Interpolation, Statistical, and Data-Driven Methods for Imputation of Missing Values in a Distributed Soil Moisture Dataset*. Journal of Hydrologic Engineering, 19(1), 26-43. [doi:10.1061/\(ASCE\)HE.1943-5584.0000767](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000767).
- [284] Kharin, Yu.S. and Huryin, A.S. (2003). *Statistical analysis and forecasting of autoregressive time series under missing values*. Bulletin of the International Statistic Institute, 1, 612-613.
- [285] Hong, B. and CH. Chen, 2003. *Radial basis function neural network-based nonparametric estimation approach for missing data reconstruction of nonstationary series*. IEEE Int. Conf. Neural Networks and Signal Processing Nanjing, China, December, 14-17, pp: 75-78. [doi:10.1109/ICNNSP.2003.1279216](https://doi.org/10.1109/ICNNSP.2003.1279216).
- [286] Coulibaly P. *Comparison of neural network methods for infilling missing daily weather records*. Journal of Hydrology. v. 341, p. 27-41, 2007. [doi:10.1016/j.jhydrol.2007.04.020](https://doi.org/10.1016/j.jhydrol.2007.04.020).
- [287] Kidson, J.W., and K.E. Trenberth, 1988: *Effects of missing data on estimates of monthly mean general circulation statistics*, J. Climate, 1, 1261–1275. [doi:10.1175/1520-0442\(1988\)001<1261:EOMDOE>2.0.CO;2](https://doi.org/10.1175/1520-0442(1988)001<1261:EOMDOE>2.0.CO;2).
- [288] Vincent, L. A., and D. W. Gullet., 1999: *Canadian historical and homogeneous temperature datasets for climate change analyses*. International Journal of Climatology. 19, 1375-1388. [doi:10.1002/\(SICI\)1097-0088\(199910\)19:12<1375::AID-JOC427>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1097-0088(199910)19:12<1375::AID-JOC427>3.0.CO;2-0).
- [289] Nelwamondo, F. V., Mohamed, S. and Marwala, T.: n.d., *Missing data: A comparison of neural networks and expectation maximization techniques*, Current Science 93(12).
- [290] L.P. Wang and X.J. Fu, *Data Mining with Computational Intelligence*, Springer, Berlin, 2005.



- [291] C. Zecchin, A. Facchinetti, G. Sparacino, G. De Nicolao and C. Cobelli, "*A new neural network approach for short-term glucose prediction using continuous glucose monitoring time-series and meal information*," 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, 2011, pp. 5653-5656. [doi:10.1109/IEMBS.2011.6091368](https://doi.org/10.1109/IEMBS.2011.6091368).
- [292] L. P. Wang and Shekhar Gupta, "*Neural networks and wavelet de-noising for stock trading and prediction*", Time Series Analysis, Modeling and Applications, Witold Pedrycz and Shyi Ming Chen (eds.), Springer, pp.229-247, 2013.
- [293] Lipo Wang, Kok Keong Teo and Zhiping Lin, "*Predicting time series with wavelet packet neural networks*," Neural Networks, 2001. Proceedings. IJCNN '01. International Joint Conference on, Washington, DC, 2001, pp. 1593-1597 vol.3. [doi:10.1109/IJCNN.2001.938397](https://doi.org/10.1109/IJCNN.2001.938397).
- [294] D. C. Park, "*A Time Series Data Prediction Scheme Using Bilinear Recurrent Neural Network*," 2010 International Conference on Information Science and Applications, Seoul, 2010, pp. 1-7. [doi:10.1109/ICISA.2010.5480383](https://doi.org/10.1109/ICISA.2010.5480383).
- [295] K.K. Teo, L.P. Wang, Z.P. Lin, "*Wavelet packet multi-layer perceptron for chaotic time series prediction: effects of weight initialization*," Computational Science – ICCS 2001, Proceedings Pt 2, Volume: 2074, pp. 310-317, 2001. [doi:10.1007/3-540-45718-635](https://doi.org/10.1007/3-540-45718-635).
- [296] Araujo, R. (2010), "*A quantum-inspired evolutionary hybrid intelligent approach for stock market prediction*", International Journal of Intelligent Computing and Cybernetics, Vol. 3, pp. 24-54.
- [297] C.-N. Ko, C.-M. Lee, "*Short-term load forecasting using SVR (support vector regression)-based radial basis function neural network with dual extended kalman filter*, Energy 49 (2013) 413–422. [doi:10.1016/j.energy.2012.11.015](https://doi.org/10.1016/j.energy.2012.11.015).
- [298] Kamal S. Selim and Gihan A. Elanany, "*A New Method for Short Multivariate Fuzzy Time Series Based on Genetic Algorithm and Fuzzy Clustering*," Advances in Fuzzy Systems, vol. 2013, Article ID 494239, 10 pages, 2013. [doi:10.1155/2013/494239](https://doi.org/10.1155/2013/494239).
- [299] Y. Ren, P. N. Suganthan, N. Srikanth, G. Amaratunga, "*Random Vector Functional Link Network for Short-term Electricity Load Demand Forecasting*", Information Sciences, 2016.

- [300] M. Zhu and L. Wang, "Intelligent trading using support vector regression and multilayer perceptrons optimized with genetic algorithms," The 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, 2010, pp. 1-5. [doi:10.1109/IJCNN.2010.5596301f](https://doi.org/10.1109/IJCNN.2010.5596301f).
- [301] R.-A. Hooshmand, H. Amooshahi, M. Parastegari, *A hybrid intelligent algorithm based short-term load forecasting approach*, International Journal of Electrical Power & Energy Systems 45 (2013) 313–324. [doi:10.1016/j.ijepes.2012.09.002](https://doi.org/10.1016/j.ijepes.2012.09.002).
- [302] Amjady, Nima, and Farshid Keynia. "Short-term load forecasting of power systems by combination of wavelet transform and neuroevolutionary algorithm." Energy, vol.34, Issue 1, pp. 46-57, January 2009. [doi:10.1016/j.energy.2008.09.020](https://doi.org/10.1016/j.energy.2008.09.020).
- [303] Richman, M. B.; Trafalis, T. B.; Adrianto, I. 2009. *Missing data imputation through machine learning algorithms*, in Artificial Intelligence Methods in the Environmental Sciences. Ed. by H. Sue Ellen, P. Antonello, M. Caren. Springer Netherlands Press, pp.153–169. [doi:10.1007/978-1-4020-9119-3\\_7](https://doi.org/10.1007/978-1-4020-9119-3_7).
- [304] Havaluddina, Ahmad Jawahir, *Comparing of ARIMA and RBFNN for short-term forecasting*, International Journal of Advances in Intelligent Informatics, Vol. 1, No 1, March 2015, pp. 15-22. [doi:10.12928/ijain.v1i1.10](https://doi.org/10.12928/ijain.v1i1.10).
- [305] Tresp, V., & Hofmann, R. (1998). *Nonlinear Time-Series Prediction with Missing and Noisy Data*. Neural Computation, 10, 731-747.
- [306] Schneider, T., 2001: *Analysis of incomplete climate data: estimation of mean values and covariance matrices and imputation of missing values*. J. Climate, 14, 853–871. [doi:10.1175/1520-0442\(2001\)014<0853:AOICDE>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<0853:AOICDE>2.0.CO;2).
- [307] Rubin, D. B.: 1987, *Multiple Imputation for Nonresponse in Surveys*, New York: Wiley.
- [308] Schafer, J.: 1997, *Analysis of Incomplete Multivariate Data*, Chapman & Hall.
- [309] Hosking, J. R. M. 1984. *Modeling persistence in hydrological time series using fractional differencing*. Water Resources Res. 20, 12, 1898–1908. [doi:10.1029/WR020i012p01898](https://doi.org/10.1029/WR020i012p01898).
- [310] Zhang, G. P., Patuwo, B. E., Hu, M. Y., 2001. *A simulation study of artificial neural networks for nonlinear time-series forecasting*. Computers and Operations Research 28 (4), 381–396. [doi:10.1016/S0305-0548\(99\)00123-9](https://doi.org/10.1016/S0305-0548(99)00123-9).

- [311] S. F. Crone and K. Nikolopoulos, "Input variable selection for time series prediction with neural networks-an evaluation of visual, autocorrelation and spectral analysis for varying seasonality," in Proc. 1st Eur. Symp. Time Series Predict, Helsinki, Finland, 2007, pp. 1–11.
- [312] D. K. Barrow, S. F. Crone and N. Kourentzes, "An evaluation of neural network ensembles and model selection for time series prediction," The 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, 2010, pp. 1-8. [doi:10.1109/IJCNN.2010.5596686](https://doi.org/10.1109/IJCNN.2010.5596686).
- [313] Crone, S. F., Kourentzes, N., 2010. *Feature selection for time series prediction - a combined filter and wrapper approach for neural networks*. Neurocomputing 73 (10-12), 1923–1936. [doi:10.1016/j.neucom.2010.01.017](https://doi.org/10.1016/j.neucom.2010.01.017).
- [314] Abdusselam Altunkaynak, Tewodros Assefa Nigussie (2015), *Prediction of daily rainfall by a hybrid wavelet-season-neuro technique*, Journal of Hydrology (529), 287–301. [doi:10.1016/j.jhydrol.2015.07.046](https://doi.org/10.1016/j.jhydrol.2015.07.046).
- [315] Hyndman, R. J. and Y. Khandakar (2008). *Automatic time series forecasting: the forecast package for R*. Journal of Statistical Software 26(3), 1–22.
- [316] W. Yan, "Toward Automatic Time-Series Forecasting Using Neural Networks," in IEEE Transactions on Neural Networks and Learning Systems, vol. 23, no. 7, pp. 1028-1039, July 2012. [doi:10.1109/TNNLS.2012.2198074](https://doi.org/10.1109/TNNLS.2012.2198074).
- [317] Remus, W. and M. O'Connor. 2001. *Neural networks for time-series forecasting*. In: Principles of Forecasting: A Handbook for Researchers and Practitioners, Norwell, MA: Kluwer Academic Publishers, pp.245-256.
- [318] P. Jiang, C. Chen and X. Liu, "Time series prediction for evolutions of complex systems: A deep learning approach," 2016 IEEE International Conference on Control and Robotics Engineering (ICCRE), Singapore, 2016, pp. 1-6. [doi:10.1109/ICCRE.2016.7476150](https://doi.org/10.1109/ICCRE.2016.7476150).
- [319] C. R. Rivero, J. Pucheta, D. Patiño, S. Laboret, G. Juárez and V. Sauchelli, "On the approximate suboptimal control by neural network - rainfall observer," 2016 IEEE Biennial Congress of Argentina (ARGENCON), Buenos Aires, Argentina, 2016, pp. 1-8. [doi:10.1109/ARGENCON.2016.7585302](https://doi.org/10.1109/ARGENCON.2016.7585302).
- [320] C. Rodríguez Rivero, J. Pucheta, J. Baumgartner, S. Laboret and V. Sauchelli, "Short-series Prediction with BEMA Approach: application to short rainfall series," in IEEE Latin

- America Transactions, vol. 14, no. 8, pp. 3892-3899, Aug. 2016. [doi:10.1109/TLA.2016.7786377](https://doi.org/10.1109/TLA.2016.7786377).
- [321] Biagini, F., Hu, Y., Øksendal, B., & Zhang, T. (2010). *Stochastic calculus for fractional Brownian motion and applications*. Berlin: Springer. [doi:10.1007/978-1-84628-797-8](https://doi.org/10.1007/978-1-84628-797-8).
- [322] Felix Höfling and Thomas Franosch, *Anomalous transport in the crowded world of biological cells*, arXiv:1301.6990v1 [cond-mat.soft], 29 Jan 2013.
- [323] E. Zhao, S. Jiang, D. Luo, Y. Bao and Z. Han, "Scaling Behavior of Time Series and an Empirical Indication to Financial Prediction," Computer and Information Science (ICIS), 2011 IEEE/ACIS 10th International Conference on, Sanya, China, 2011, pp. 126-131. [doi:10.1109/ICIS.2011.28](https://doi.org/10.1109/ICIS.2011.28).
- [324] M. E. Haque, M. M. Rahman, A. Rahman and K. M. Imtiaz-Ud-Din, "Self organized sensor deployment with Brownian motion in wireless sensor and robot networks," 2015 18th International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2015, pp. 196-201. [doi:10.1109/ICCITechn.2015.7488067](https://doi.org/10.1109/ICCITechn.2015.7488067).
- [325] Nualart, D. (2003), *Stochastic integration with respect to fractional Brownian motion an applications*. Preprint.
- [326] Lamperti, J. (1962), 'Semi-stable stochastic processes', Amer. Math. Soc. Trans.104, 62-78.
- [327] T. E. Duncan, Y. Hu and B. Pasik-Duncan: *Stochastic calculus for fractional Brownian motion I. Theory*. SIAM J. Control Optim. 38 (2000), 582-612. [doi:10.1137/S036301299834171X](https://doi.org/10.1137/S036301299834171X).
- [328] Embrechts, P. & Maejima, M. (2002), *Self-similar Stochastic Processes*, Princeton University Press.
- [329] Vervaat, W. (1985), 'Sample path properties of self-similar processes with stationary increments', Annals of Probability 13, 1-27.
- [330] Lin, S. (1995), 'Stochastic analysis of fractional brownian motion', Stochastic and Stochastic Reports 55(1-2), 121-140. [doi:10.1080/17442509508834021](https://doi.org/10.1080/17442509508834021).
- [331] K. J. Falconer. *Fractal Geometry*. Mathematical Foundations and Applications. John Wiley & Sons, 1990.
- [332] B. M. Wilamowski. Y. Chen, and A. Malinowski, *Efficient algorithm for training neural networks with one hidden layer*, in 1999 International Joint Conference on Neural Networks

- (IJCNN'99), pp. 1725–1728, Washington, DC, July 10–16, 1999. [doi:10.1109/IJCNN.1999.832636](https://doi.org/10.1109/IJCNN.1999.832636).
- [333] B. M. Wilamowski, N. J. Cotton, O. Kaynak, and G. Dunder, *Computing gradient vector and Jacobian matrix in arbitrarily connected neural networks*, IEEE Transactions on Industrial Electronics, 55(10), 3784–3790, October 2008. [doi:10.1109/TIE.2008.2003319](https://doi.org/10.1109/TIE.2008.2003319).
- [334] William W. Hsieh, *Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels*, Cambridge University Press, New York, NY, 2009.
- [335] Ozan Kocadağlı (2015). *A novel hybrid learning algorithm for full Bayesian approach of artificial neural networks*. Vol. 35, pp. 52–65. [doi:10.1016/j.asoc.2015.06.003](https://doi.org/10.1016/j.asoc.2015.06.003).
- [336] D. MacKay, “*Bayesian neural networks and density networks*”, Nuclear Instrum. and Methods in Phys. Res. Section A, vol. 354, no. 1, pp. 73–80, 1995.
- [337] Kreinovich, Vladik; Nguyen, Hung T.; and Ouncharoen, Rujira, “*How to Estimate Forecasting Quality: A System-Motivated Derivation of Symmetric Mean Absolute Percentage Error (SMAPE) and Other Similar Characteristics*” (2014). Departmental Technical Reports (CS). Paper 865 [http://digitalcommons.utep.edu/cs\\_techrep/865](http://digitalcommons.utep.edu/cs_techrep/865).
- [338] Cover, T. M. and Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons.
- [339] J. Baumgartner, J. Giménez, M. Scavuzzo and J. Pucheta, “*A New Approach to Segmentation of Multispectral Remote Sensing Images Based on MRF*,” in IEEE Geoscience and Remote Sensing Letters, vol. 12, no. 8, pp. 1720-1724, Aug. 2015. [doi:10.1109/LGRS.2015.2421736](https://doi.org/10.1109/LGRS.2015.2421736).
- [340] Lorna R. Nayagam, Rajesh J. and H.S. Ram Mohan, *Chapter: Empirical Approaches in Long-term Rainfall Forecasting*, in book: Rainfall Forecasting, vol. 1, pp. 39- 128, Nova Science Publisher, (2012).
- [341] L. Sergio, R. R. Cristian, P. Julián and S. Victor, “*Robust Fractional Digital Control of a First Order Plus Integrator Process*,” Electronics, Robotics and Automotive Mechanics Conference (CERMA), 2012 IEEE Ninth, Cuernavaca, 2012, pp. 225-230. [doi:10.1109/CERMA.2012.43](https://doi.org/10.1109/CERMA.2012.43).

- [342] N. I. Sapankevych and R. Sankar, "Time Series Prediction Using Support Vector Machines: A Survey," in IEEE Computational Intelligence Magazine, vol. 4, no. 2, pp. 24-38, May 2009. [doi:10.1109/MCI.2009.932254](https://doi.org/10.1109/MCI.2009.932254).
- [343] C. P. Zhao, C. Liang and H. w. Guo, "Research on Hydrology Time Series Prediction Based on Grey Theory and epsilon-Support Vector Regression," Digital Manufacturing and Automation (ICDMA), 2011 Second International Conference on, Zhangjiajie, Hunan, 2011, pp. 968-971. [doi:10.1109/ICDMA.2011.240](https://doi.org/10.1109/ICDMA.2011.240).
- [344] Z. Yuelong, Q. Jihong, F. Qingsong, W. Dingsheng and L. Shijin, "Study on Hydrology Time Series Prediction Based on Wavelet-neural Networks," Computer and Information Science, 2009. ICIS 2009. Eighth IEEE/ACIS International Conference on, Shanghai, 2009, pp. 411-415. [doi:10.1109/ICIS.2009.116](https://doi.org/10.1109/ICIS.2009.116).
- [345] Sophie Achard, Irène Gannaz (2015), *Multivariate wavelet Whittle estimation in long-range dependence*, in arXiv:1412.0391 [math.ST].
- [346] Raquel Prado and Mike West: *Time series: modelling, computation and inference*. Chapman & Hall/CRC, 2010, 368 pp.
- [347] P. S. Addison, *Fractals and Chaos: An Illustrated Course*, Institute of Physics, Bristol, PA, 1997.
- [348] Gamerman, D. (1997). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*. London: Chapman & Hall.
- [349] Zhang, G. P., Patuwo, B. E., Hu, M. Y., 2001. *A simulation study of artificial neural networks for nonlinear time-series forecasting*. Computers and Operations Research 28 (4), 381-396. [doi:10.1016/S0305-0548\(99\)00123-9](https://doi.org/10.1016/S0305-0548(99)00123-9).
- [350] L. Chun-Ying, Y. You-Long and Z. Heng-Wei, "Prediction of chaotic time series based on incremental method for Bayesian network learning," Control Conference (CCC), 2012 31st Chinese, Hefei, 2012, pp. 4245-4249.
- [351] Wasito, I.: 2003, *Least Squares Algorithms with Nearest Neighbour Techniques for Imputing Missing Data Values*, PhD thesis, University of London.
- [352] Berger, J. O. (2006), "The Case for Objective Bayesian Analysis," Bayesian Analysis, 3, 385-402. <http://ba.stat.cmu.edu/journal/2006/vol01/issue03/berger.pdf>.

- [353] Chen, Jianhua, *"Neural Network Applications in Agricultural Economics"*(2005). University of Kentucky Doctoral Dissertations. Paper 228. [http://uknowledge.uky.edu/gradschool\\_diss/228](http://uknowledge.uky.edu/gradschool_diss/228).
- [354] Bin Wu, Yu Cui, Ding Xiao and Cunyong Zhang, *"Prediction of energy consumption time series using Neural Networks combined with exogenous series,"* Natural Computation (ICNC), 2015 11th International Conference on, Zhangjiajie, 2015, pp. 37-41. [doi:10.1109/ICNC.2015.7377962](https://doi.org/10.1109/ICNC.2015.7377962).
- [355] Cristian Rodríguez Rivero, Julián Pucheta, Efrén Gorrostieta, H. Daniel Patiño, Sergio Laboret and Víctor Sauchelli. *"Bayesian enhanced modified forecasting approach: application to wind power series"*. Anales del 25° Congreso Argentino de Control Automático. 1 al 3 de Noviembre de 2016 – Buenos Aires, Argentina. ISBN 978-950-99994-9-7. 2016.

# Apéndice A

## Movimiento Browniano Fraccional

El movimiento browniano fraccional ( $mBf$ ) es un proceso estocástico que ha sido estudiado intensamente en los últimos años [321]. Su auge se debe a sus múltiples aplicaciones en diferentes campos, tales como la meteorología , biología [322], finanzas [323] y telecomunicaciones [324]; desde el punto de vista teórico el  $mBf$  es interesante, pues no es proceso de Markov ni una semimartingala y por lo tanto el cálculo desarrollado por Itó no se puede aplicar [51].

**Definición A.1** Un proceso gaussiano centrado

$$B_t^H = \{B_s^H : 0 \leq t \leq \infty\} \quad (\text{A.1})$$

Con  $B_0^H = 0$  es un movimiento browniano fraccional con parámetro (o constante de Hurst)  $H \in (0, 1)$  si su función de covarianza está dada por:

$$\text{cov}(B_t^H, B_s^H) = \frac{1}{2} (t^{2H} + s^{2H} - |t - s|^{2H}) \text{var } B_1^H \quad (\text{A.2})$$

para todo  $t, s \geq 0$ .

En [325] se muestra que si  $H \in (0, 1]$  entonces la Ecuación (A.1) es definida no negativa y por lo tanto el  $mBf$  es el único proceso gaussiano, con incrementos estacionarios de segundo orden y con función de covarianza de la Ecuación (A.2). El  $mBf$  es un proceso *autosimilar*, es decir, es invariante en distribución bajo un adecuado cambio de escala de tiempo y espacio. Este tipo de proceso es usado para modelar fenómenos aleatorios con dependencia a gran distancia. El primer tratamiento riguroso del tema fue desarrollado por [326], en donde el concepto de autosimilaridad es un caso especial de lo que llama proceso semiestable.

Rigurosamente hablando un proceso estocástico  $(X_t)_{t \in \mathbb{R}^+}$  es autosimilar,  $H_{-ss}$ , si existe  $H > 0$  tal que para todo  $a > 0$  se tiene:

$$X_{at} \stackrel{\Delta}{=} a^H X_t \quad (\text{A.3})$$



donde  $\stackrel{\Delta}{=}$  denota la igualdad de las distribuciones finito dimensionales. Intuitivamente la relación (9) indica que las trayectorias de los procesos  $X_{at}$  y  $a^H X_t$  aun cuando no son idénticas son visualmente similares. Si  $(X_t)_{t \in \mathfrak{R}^+}$  es un proceso no trivial, H–ss, con incrementos estacionarios y  $H > 0$ , se tienen las siguientes propiedades [327], [328] y [329]:

- i)  $X_0 = 0$  casi siempre.
- ii) Si  $E[(X_1)^2] < \infty$  entonces:  $E[(X_1 X_s)^2] = \frac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H})E[(X_1)^2]$
- iii) Si  $E[(|X_1|)^\alpha] < \infty$  para algún  $0 < \alpha \leq 1$ , entonces  $\alpha H \leq 1$ .
- iv) Si  $X_t$  tiene varianza finita y  $H \neq 1$  entonces  $E[X_t] = 0$ .
- v) Las trayectorias son casi siempre no diferenciales.

Por ser el *mBf* un proceso gaussiano, generalmente se define en términos de su covarianza, sin embargo, existen varias definiciones equivalentes. El siguiente teorema resume las equivalencias de la definición del *mBf* presentadas en [330].

**Teorema A.1.** Un proceso gaussiano centrado  $B_t^H = \{B_s^H : 0 \leq t \leq \infty\}$  con  $B_0^H = 0$  es un movimiento browniano fraccional con parámetro  $H \in (0,1)$ , si y solo si, se cumple alguna de las siguientes condiciones:

- a)  $\text{cov}(B_t^H, B_s^H) = \frac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H}) \text{var} B_1^H$   $t, s \geq 0$ .
- b)  $(B_{t_2}^H - B_{t_1}^H, B_{s_2}^H - B_{s_1}^H) = (B_{t_2+h}^H - B_{t_1+h}^H, B_{s_2+h}^H - B_{s_1+h}^H)$ , para todo  $t_1, t_2, s_1, s_2$  y  $h \geq 0$ , y si existe un  $H \in (0,1)$  tal que:  $(B_{t+\tau}^H - B_t^H) \stackrel{\Delta}{=} (B_{t+\tau}^H - B_t^H)$  para todo  $t, \tau, h \geq 0$ .
- c)  $\text{Var}(B_{t+\tau}^H - B_t^H) = (|t-s|^{2H}) \text{var} B_1^H$  para todo  $t, s \geq 0$ .

Como consecuencia del teorema anterior se tiene que el *mBf* posee incrementos estacionarios, es H–ss y se tiene que:

$$\text{cov}(B_{t_2}^H - B_{t_1}^H, B_{s_2}^H - B_{s_1}^H) = \frac{1}{2}[(s_2 - t_1)^{2H} + (s_1 - t_2)^{2H} - (s_2 - t_2)^{2H} (s_1 - t_1)^{2H}] \quad (\text{A.4})$$

De (A.4) se tiene que si  $H = \frac{1}{2}$ , el  $mBf$  está negativamente correlacionado, si  $H > \frac{1}{2}$  el  $mBf$  está positivamente correlacionado y cuando  $H = \frac{1}{2}$  la covarianza es cero y por ser el  $mBf$  un proceso gaussiano, se tiene la independencia de los incrementos, esto es, el movimiento browniano estándar. De la definición de covarianza del  $mBf$  se puede ver que el  $mBf$  es un proceso de Markov, si y sólo si  $H = \frac{1}{2}$ .

Acerca de las trayectorias del  $mBf$ , el criterio de continuidad de Kolmogorov garantiza que estas son casi siempre Hölder continuas de orden menor que  $H$ . Además, por ser el  $mBf$  un proceso autosimilar se tiene que sus trayectorias son diferenciables en ninguna parte casi seguramente, una demostración alternativa se puede encontrar en [52]. De la variación de las trayectorias se puede decir que tienen  $p$ -variación no acotada casi seguramente, para  $H < 1$ , y que su variación cuadrática es cero si  $H = \frac{1}{2}$ , como [325]. Como una consecuencia de lo anterior se tiene que el  $mBf$  con parámetro de Hurst  $H \neq \frac{1}{2}$  no es una semimartingala [52].

La dimensión no entera de su grafo es otra característica del movimiento browniano que permite incluirlo dentro de los procesos fractales [331].

La característica fractal de invarianza de la distribución bajo un cambio adecuado de escala en el tiempo y el espacio generaliza el resultado obtenido para el movimiento browniano clásico [328].

El  $mBf$  se define según [71] mediante,

$$B_H(t) = \frac{1}{r\left(H + \frac{1}{2}\right)} \left( \int_{-\infty}^0 \left( (t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}} \right) dB(s) + \int_0^t (t-s)^{H-\frac{1}{2}} dB(s) \right) \quad (A.5)$$

donde  $B$  es un  $mB$  en  $(\Omega, F, P)$ ,  $H$  es el parámetro de Hurst y  $\Gamma(\cdot)$  representa la función con

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx.$$

El  $mBf$  tiene varianza de sus incrementos  $Var(B_H(t) - B_H(s)) = \mathcal{G}|t-s|^{2H}$  donde  $\mathcal{G}$  es una constante positiva.